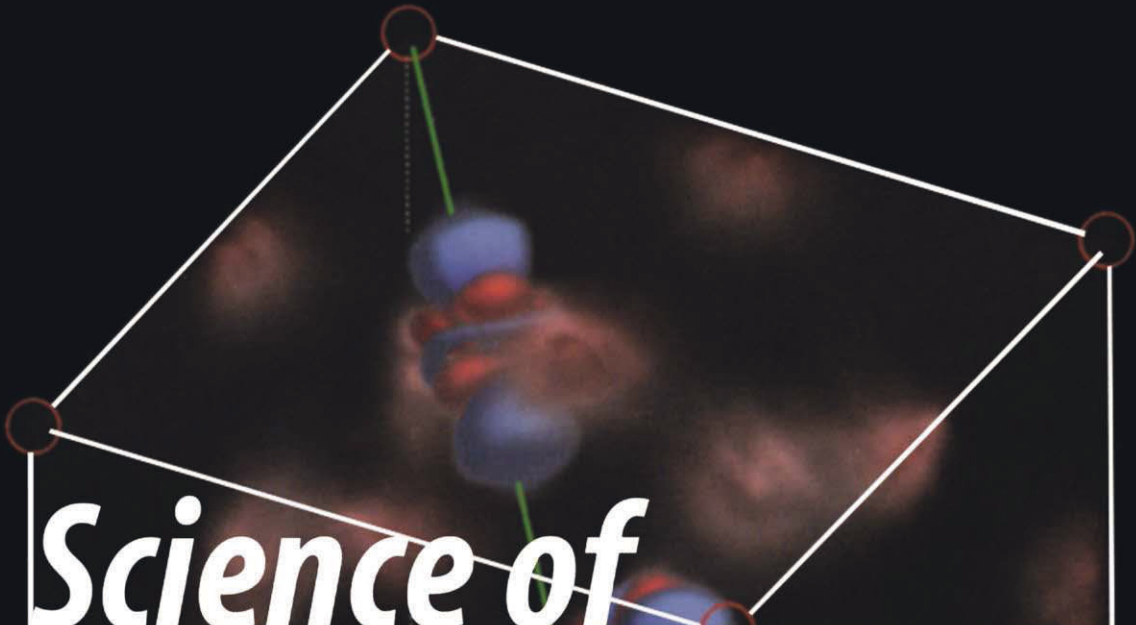
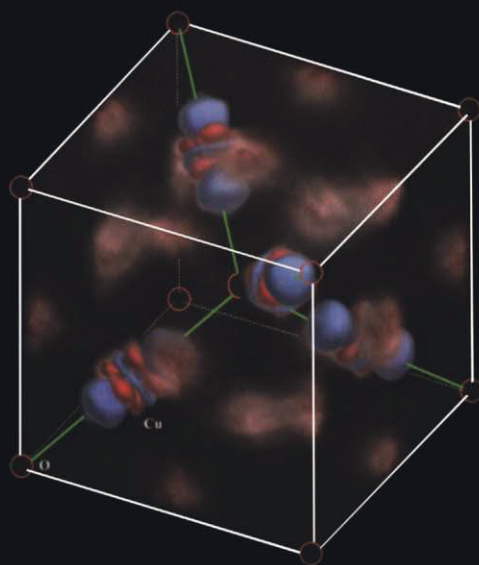




Science of Microscopy

Volume I



Peter W. Hawkes
John C.H. Spence
Editors

Science of Microscopy

Science of Microscopy

Edited by

Peter W. Hawkes
John C.H. Spence

Volume I

Peter W. Hawkes
CEMES CNRS
Toulouse
France

John C.H. Spence
Department of Physics
Arizona State University
Tempe, AZ
and
Lawrence Berkeley Laboratory
Berkeley, CA
USA

Library of Congress Control Number: 2005927385

ISBN 10: 0-387-25296-7
ISBN 13: 978-0387-25296-4

Printed on acid-free paper.

© 2007 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 (corrected printing, 2008)

springer.com

Preface

In these two volumes, we have asked many of the leaders in the field of modern microscopy to summarize the latest approaches to the imaging of atoms or molecular structures, and, more especially, the way in which this aids our understanding of atomic processes and interactions in the organic and inorganic worlds.

Man's curiosity to examine the nanoworld is as at least as old as the Greeks. Aristophanes, in a fourth-century BC play, refers to a burning glass; the Roman rhetorician Seneca describes hollow spheres of glass filled with water being used as magnifiers, while Marco Polo in the thirteenth century remarks on the Chinese habit of wearing spectacles. Throughout this time it would have been common knowledge that a drop of water over a particle on glass will provide a magnified image, while a droplet within a small hole does even better as a biconvex lens. By the sixteenth century magnifying glasses were common in Europe, but it was Anthony van Leeuwenhoek (1632–1723) who first succeeded in grinding lenses accurately enough to produce a better image with his single-lens instrument than with the primitive compound microscopes then available. His 112 papers, published in *Philosophical Transactions of the Royal Society*, brought the microworld to the general scientific community for the first time, covering everything from sperm to the internal structure of the flea. Robert Hooke (1635–1703) developed the compound microscope, publishing his results in careful drawings of what he saw in his *Micrographia* (1665). The copy of this book in the University of Bristol library shows remarkable sketches of faceted crystallites, below which he has drawn piles of cannon balls, whose faces make corresponding angles. This strongly suggests that Hooke believed that matter consists of atoms and had made this discovery long before its official rediscovery by the first modern chemists, notably Dalton in 1803. (Greeks such as Leucippus (450 BC) had long before convinced themselves that a stone, cut repeatedly, would eventually lead to “a smallest fragment” or fundamental particle; Democritus once said that “nothing exists except atoms and empty space. All else is opinion” (!))

This atomic hypothesis itself has a fascinating history, and is intimately connected with the history of microscopy. It was Brown's observation in 1827 of the motion of pollen in water by optical micro-

scopy which laid the basis for the modern theory of matter based on atoms. As late as 1900 many chemists and physicists did not believe in atoms, despite the many independent estimates that could be made of their size. These were summarized by Kelvin and Tait in an appendix to their *Treatise on Natural Philosophy*, together with an erroneous and rather superficial estimate of the age of the earth, to be used against Darwin. (This text was the standard English language physics text of the late nineteenth century, despite its failure to cover much of Maxwell's work.) Einstein's 1905 theory of Brownian motion, and Perrin's (1909) more accurate repetition of Brown's experiment, using microscope observations to estimate Avogadro's number, finally settled the matter regarding the existence of atoms. Einstein does not reference Brown's paper, but indicates that he had been told about it. But as Archie Howie has commented, it is interesting to speculate how different the history of science would be if Maxwell had read the Brown paper and applied his early statistical mechanics to it. By the time of Perrin's paper, Bohr, Thomson, Rutherford and others were well committed to atomic and even subatomic physics.

In biology, the optical microscope remained an indispensable tool from van Leeuwenhoek's time with many incremental improvements, able to identify bacteria and their role in disease, but not viruses, which were first seen with the transmission electron microscope (TEM) in 1938. With Zernike's phase contrast theory in the thirties a major step forward was taken, but the really dramatic and spectacular modern advances had to await the widespread use of the TEM, the invention of the laser and the CCD detector, the introduction of the scanning mode, computer control and data acquisition, and the production of fluorescent proteins.

The importance of this early history should not be underestimated—in the words of Feynman “If in some cataclysm, all scientific knowledge were to be destroyed and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis—that all things are made of atoms.”

Images of individual atoms were first provided by Muller's field-ion microscope in the early 1950s, soon to be followed by Albert Crewe's Scanning Transmission Electron Microscope (STEM) images of heavy atoms on thin-film surfaces in 1970. With its subångström resolution, the modern transmission electron microscopes (TEM) can now routinely image atomic columns in thin crystals. For favorable surface structures, the scanning tunneling microscope has provided us with images of individual surface atoms since its invention in 1982, and resulted in a rich spin-off of related techniques.

Probes of condensed and biological matter must possess a long lifetime if they are to be used as free-particle beams. For the most part this has limited investigators to the use of light, X-rays, neutrons and electrons. The major techniques can then often be classified as imaging, diffraction, and spectroscopy. These may be used in both the transmission and reflection geometries, giving bulk and surface information respectively. Chapter 8 (Bauer) reviews both the low-energy electron

microscope (LEEM) and spin-polarized LEEM methods which, using reflected electrons, have recently revolutionized surface science and thin-film magnetism. Here the high cross-section allows movies to be made of surface processes at submicrometer resolution, while Auger electron spectroscopy is conveniently incorporated. Chapter 9 (Feng and Scholl) deals with the closely related photoelectron microscopy, where a LEEM instrument is used to image the photoelectrons excited by a synchrotron beam. Here the superb energy selectivity of optical excitation can be used to great advantage. Chapter 3 (Reichelt) describes advances in scanning electron microscope (SEM) research, where the lower-energy secondary electrons provide images with large depth of focus for the most versatile of all electron-optical instruments. The numerous modes of operation include X-ray analysis, cathodoluminescence, low-voltage modes for insulators and the controlled-atmosphere environmental SEM (ESEM). Turning now to the transmission geometry, we review the latest work in atomic-resolution TEM in Chapter 1 (Kirkland, Chang and Hutchison), the technique which has transformed our understanding of defect processes in the bulk of solids such as oxides, and the STEM mode in Chapter 2 (Nellist). STEM provides an additional powerful analytical capability, which, like the STM, can provide spectroscopy with atomic-scale spatial resolution. An entire chapter (Chapter 4, Botton) is then devoted to analytical TEM (AEM), with a detailed analysis of the physics and performance of its two main detectors, for characteristic X-ray emission and energy-loss spectroscopy. The remarkable recent achievements of *in-situ* TEM are surveyed in Chapter 6 (Ross), including transmission imaging of liquid cell electrolysis, observations of the earliest stages of crystal and nanotube growth, phase transitions and catalysts, superconductors, magnetic and ferroelectric domains and plastic deformation in thin films, all at nanometer resolution or better. Again, the large scattering cross-section of electron probes provides plenty of signal even from individual atoms, so that movies can be made. Chapter 5 (King, Armstrong, Bostanjoglo and Reed) summarizes the dramatic recent revival of time-resolved electron microscope imaging, which uses laser-pulses to excite processes in a sample. The excited state may then be imaged by passing the delayed pulse to the photocathode of the TEM in this “pump-probe” mode. Single-shot transmission electron diffraction patterns have now been obtained using electron pulses as short as a picosecond. Most of these techniques are undergoing a quiet revolution as electron-optical aberration corrector devices are being fitted to microscopes. The dramatic discovery, that, after 60 years of effort, aberration correction is now a reality, was made about ten years ago, and we review the relevant electron-optical theoretical background in Chapter 10 (Hawkes). Finally, in biology, potentially the largest scientific payoff of all is occurring in the field of cryo-electron microscopy, where single-particle images of macromolecules embedded in thin films of ice are imaged, and a three-dimensional reconstruction is made. While the structure of the ribosome and purple membrane protein (among many others) have already been determined in this way, the grand challenge of locating every protein and molecular machine in a single cell remains

to be completed. We summarize this exciting field in Chapter 7 (Plitzko and Baumeister).

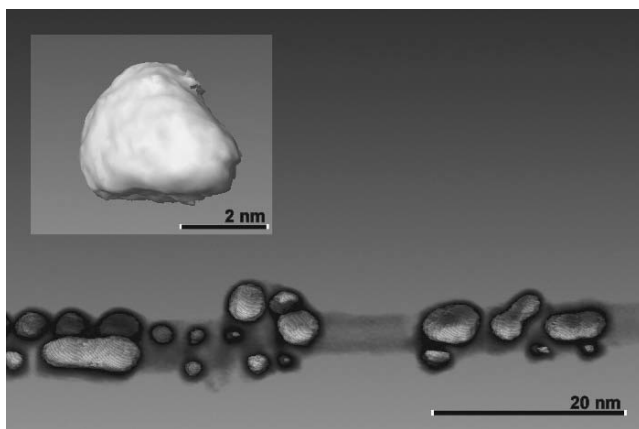
Electrons, with the largest cross-section and a source brighter than current generation synchrotrons, provide the strongest signal and hence the best resolution. They do this in a manner that can conveniently be combined with spectroscopy, and we now have aberration-corrected lenses for them. But multiple scattering and inelastic background scattering often complicate interpretation. X-ray imaging of nanostructures, even at synchrotrons, involves much longer data acquisition times but the absence of background and multiple scattering effects greatly improves quantification of data, and thicker samples can be examined. (It is easy to show that the small magnitude of the fine-structure constant will almost certainly never permit imaging of individual atoms using X-rays. We should also recall that in protein crystallography, about 98% of the X-ray beam hits the beam-stop and does not interact with the sample. Of the remaining 2%, 84% is annihilated in production of photoelectrons, and 8% in Compton scattering, while only the remainder produces Bragg diffraction. For light elements the inelastic cross-section for kilovolt electrons is about three times the elastic.) Success with X-rays has thus come mainly through the use of crystallographic redundancy to reduce radiation damage in protein crystallography. However, soft X-ray imaging with zone-plate lenses provides about 30 nm resolution in the “water window” with the advantages of thick samples and an aqueous environment. Applications have also been found in environmental science, materials science and magnetic materials. In addition, the equivalent of the STEM has been developed for soft X-rays: the scanning transmission x-ray microscope (STXM), which uses a zone-plate to focus X-rays onto a sample that can be translated by piezo motors. This arrangement can then provide spatially-resolved X-ray absorption spectroscopy. This work is reviewed in Chapter 13 (Howells, Jacobsen and Warwick).

Both X-ray and electron-beam imaging methods are limited in biology by the radiation damage they create, unlike microscopy with visible light, which also allows observations in the natural state. Optical microscopy is undergoing a revolution, with the development of super-resolution, two-photon, fluorescent labeling and scanning confocal methods. These methods are reviewed in Chapters 11 and 12. Chapter 11 (Diaspro, Schneider, Bianchini, Caorsi, Mazza, Pesce, Testa, Vicidomini, and Usai) discusses two-photon confocal microscopy, in which the spot-scanning mode is adopted, and a symmetrical lens beyond the sample collects light predominantly from the excitation region, thereby eliminating most of the “out-of-focus” background produced in the normal full-field “optical sectioning” mode. Three-dimensional image reconstruction is then possible. Two-photon microscopy combines this with a fluorescence process in which two low-energy incident photons are required to excite a detectable photon emitted at the sum of their energies. This has several advantages, by reducing radiation damage and background, and allowing observation of thicker samples. The method can also be used to initiate photochemical reactions for study. Chapter 12 (Hell and Schönle) describes the

latest super-resolution schemes for optical microscopy, which have now brought the lateral resolution to about 28 nm and, by the symmetrical lens arrangement (4π confocal), increased resolution measured along the optic axis by a factor of up to seven. The lateral resolution can be improved by modulating the illumination field or by using the stimulated emission depletion microscopy mode (STED), in which saturated excitation of a fluorophor produces nonlinear effects allowing the diffraction barrier to resolution to be broken.

For the scanning near-field probes new possibilities arise. Although restricted to the surface (the site of most chemical activity) and requiring in some cases complex image interpretation, damage is reduced, while the subångström resolution normal to the surface is unparalleled. The method is also conveniently combined with spectroscopy. Early work was challenged by problems of reproducibility and tip artifacts, but Chapters 14–17 in this book show the truly remarkable recent progress in surface science, materials science and biology. Chapter 14 (Nikiforov and Bonnell) describes the various modes of atomic force microscopy which can be used to extract atomic-scale information from the surfaces of modern materials, including oxides and semiconductors. Work-functions can be mapped out (a Kelvin probe with good spatial resolution) and a variety of useful signals obtained by modulation spectroscopy methods. In this way maps of magnetic force, local dopant density, resistivity, contact potential and topography may be obtained. Chapter 15 (Sutter) describes applications of the scanning tunneling microscope (STM) in materials science, including inelastic tunneling, surface structure analysis in surface science, the information on electronic structure which may be extracted, atomic manipulation, quantum size and sub-surface effects, and high temperature imaging. Weierstall, in Chapter 16, reviews STM research at low temperatures, including a thorough analysis of instrumental design considerations and applications. These include measurements of local density-of-states oscillations, energy dispersion measurements, electron confinement, lifetime measurements, the Stark and Kondo effects, atomic manipulation, local inelastic tunneling spectroscopy, photon emission, superconductivity and spin-polarized tunneling microscopy. Finally, in Chapter 16, Amrein reviews the special problems that arise when the atomic force microscope (AFM) is applied to the imaging of biomolecules; much practical information on instrumentation and sample preparation is provided, and many striking examples of cell and macromolecule images are shown.

We include two chapters on unconventional “lensless” imaging methods—Chapter 18 (Dunin-Borkowski, Kasama, McCartney and Smith) deals with electron holography and Chapter 19 with diffractive imaging. Gabor’s original 1948 proposal for holography was intended to improve the resolution of electron microscopes, and only recently have his plans been realized. Meanwhile, electron holography using Möllenstedt’s biprism and the Lorentz mode has proved an extremely powerful method of imaging the magnetic and electrostatic fields within matter. Dramatic examples have included TEM movies of superconducting vortices as temperature and applied fields are varied, and ferroelectric and magnetic domain images, all within thin self-supporting films. Chapter 19 (Spence) describes the recent develop-



A projection from a three-dimensional image of a carbon nanotube with gold clusters attached. This was reconstructed by taking a series of projected STEM-ADF images at different tilt angles. A faceted gold cluster is shown in the inset. Electron tomography makes it possible to study the three-dimensional nanotube–metal contact geometry which determines the electrical contact resistance to the nanotubes (courtesy of J. Cha, M. Weyland, and D. Muller, 2006).

ment of new iterative solutions to the non-crystallographic phase problem, which now allows diffraction-limited images to be reconstructed from the far-field scattered intensity distribution. This has produced lensless atomic-resolution images of carbon nanotubes (reconstructed from electron microdiffraction patterns) and phase contrast images from both neutron and soft X-ray Fraunhofer diffraction patterns of isolated, non-periodic objects. In this work, lenses are replaced by computers, so that images may now be formed with any radiation for which no lens exists, free of aberrations. Our volumes end with a chapter by van Aert, den Dekker, van Dyck and van den Bos on the definition of resolution in all its forms.

Coverage has been limited to high-resolution methods, with the result that some important new microscopies have been omitted (such as magnetic resonance imaging (MRI), projection X-ray tomography, acoustic imaging etc.). Field-ion microscopy and near-field optical microscopy are also absent. Conversely, although there is no chapter on tomography in materials science, we must mention the rapid progress of these techniques, which has culminated in a remarkable near-atomic reconstruction by J. Cha, M. Weyland and D. Muller of a carbon nanotube to which gold clusters are attached (see figure). For further information on this branch of tomography, see Midgley and Weyland (2003), Midgley (2005), Weyland et al. (2006) and Kawase et al. (2007).

The ingenuity and creativity of the microscopy community as recorded in these pages are remarkable, as is the spectacular nature of the images presented. Neither shows any signs of abating. As in the past, we fully expect major advances in science to continue to result from breakthroughs in the development of new microscopies.

*Peter W. Hawkes
John C.H. Spence*

References

- Kawase, N., Kato, M., Nishioka, H. and Jinnai, H. (2007). Transmission electron microtomography without the “missing wedge” for quantitative structural analysis. *Ultramicroscopy* **107**, 8–15.
- Midgley, P.A. (2005). Tomography using the transmission electron microscope. In *Handbook of Microscopy for Nanotechnology* (Yao, N. and Wang, Z.L., Eds) 601–627 (Kluwer, Boston).
- Midgley, P.A. and Weyland, M. (2003). 3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography. *Ultramicroscopy* **96**, 413–431.
- Weyland, M., Yates, T.J.V., Dunin-Borkowski, R.E., Laffont, L. and Midgley, P.A. (2006). Nanoscale analysis of three-dimensional structures by electron tomography. *Scripta Mater.* **55**, 29–33.

Note on the second printing

Barely a year has passed since we wrote the above Preface but very considerable progress has been made in many forms of microscopy. These are indicated in the additional comments and references at the ends of the chapters. One technique that was not accorded a chapter in the first printing has come of age in 2007. This is scanning ion microscopy. Although attempts to use ions in a scanning instrument go back to the 1960s (e.g., Drummond and Long, 1967; Martin, 1973), it is only very recently that technical progress has resulted in a high-resolution commercial instrument (the ORION helium ion microscope released by Carl Zeiss in 2007). The secret of this advance lies in the atomic-level ion source (ALIS): the tip is precisely shaped with either an atomic triad or a single atom at its apex and operates at liquid-nitrogen temperature. A resolution of 0.7 nm with 45 keV helium ions has been obtained with a prototype; the quoted energy spread is only 0.3 eV and the brightness is of the order of 10^9 A/cm² sr. The current is, however, very low, in the femtoampère or picoampère range and thus considerably less than that in a STEM. Earlier scanning ion microscopes used liquid-metal ion sources and their resolution rarely exceeded 10 nm. For the background to this development, see Ishitani and Tsuboi (1997), Sakai et al. (1999), Ishitani and Ohya (2003) and Maclaren et al. (2003) and for recent progress, Ward et al. (2006), Griffin and Joy (2007) and Joy et al. (2007).

References

- Drummond, I.W. and Long, J.V.P. (1967). Scanning ion microscopy and ion beam machining. *Nature* **215**, 950–952.
- Griffin, B.J. and Joy, D.C. (2007). Imaging with the He scanning ion microscope and with low-voltage SEM – a comparison using carbon nanotube, platinum thin film, cleaved molybdenum disulphide samples and metal standards. *Acta Microscopica* **16**, Suppl. 2, 3–4.

- Ishitani, T. and Ohya, K. (2003). Comparison in spatial spreads of secondary electron information between scanning ion and scanning electron microscopy. *Scanning* **25**, 201–205.
- Ishitani, T. and Tsuboi, H. (1997). Objective comparison of scanning ion and scanning electron microscope images. *Scanning* **19**, 489–497.
- Joy, D.C., Griffin, B.J., Notte, J. and Fenner, C. (2007). Device metrology with high-performance scanning ion beams. *Proc. SPIE* **6518**, to be published.
- Maclaren, D.A., Holst, B., Riley, D.J. and Allison, W. (2003). Focusing elements and design considerations for a scanning helium microscope (SheM). *Surface Rev. Lett.* **10**, 249–255.
- Martin, F.W. (1973). Is a scanning ion microscope feasible? *Science* **179**, 173–175.
- Sakai, Y., Yamada, T., Suzuki, T. and Ichinokawa, T. (1999). Contrast mechanisms of secondary electron images in scanning electron and ion microscopy. *Appl. Surface Sci.* **144**, 96–100.
- Ward, B.W., Notte, A. and Economou, N.P. (2006). Helium ion microscope: a new tool for nanoscale microscopy and metrology. *J. Vac. Sci. Technol.* **B24**, 2871–2874.

PWH and JCES

Contents

VOLUME I

PART I IMAGING WITH ELECTRONS	1
1 Atomic Resolution Transmission Electron Microscopy <i>Angus I. Kirkland, Shery L.-Y. Chang and John L. Hutchison</i>	3
2 Scanning Transmission Electron Microscopy <i>Peter D. Nellist</i>	65
3 Scanning Electron Microscopy <i>Rudolf Reichelt</i>	133
4 Analytical Electron Microscopy <i>Gianluigi Botton</i>	273
5 High-Speed Electron Microscopy <i>Wayne E. King, Michael R. Armstrong, Oleg Bostanjoglo, and Bryan W. Reed</i>	406
6 <i>In Situ</i> Transmission Electron Microscopy <i>Frances M. Ross</i>	445
7 Cryoelectron Tomography (CET) <i>Juergen M. Plitzko and Wolfgang Baumeister</i>	535
8 LEEM and SPLEEM <i>Ernst Bauer</i>	605
9 Photoemission Electron Microscopy (PEEM) <i>Jun Feng and Andreas Scholl</i>	657
10 Aberration Correction <i>Peter W. Hawkes</i>	696
Index	II

VOLUME II

PART II IMAGING WITH PHOTONS	749
11 Two-Photon Excitation Fluorescence Microscopy <i>Alberto Diaspro, Marc Schneider, Paolo Bianchini, Valentina Caorsi, Davide Mazza, Mattia Pesce, Ilaria Testa, Giuseppe Vicidomini, and Cesare Usai</i>	751
12 Nanoscale Resolution in Far-Field Fluorescence Microscopy <i>Stefan W. Hell and Andreas Schönle</i>	790
13 Principles and Applications of Zone Plate X-Ray Microscopes <i>Malcolm Howells, Christopher Jacobsen, and Tony Warwick</i>	835
PART III NEAR-FIELD SCANNING PROBES	927
14 Scanning Probe Microscopy in Materials Science <i>Maxim P. Nikiforov and Dawn A. Bonnell</i>	929
15 Scanning Tunneling Microscopy in Surface Science <i>Peter Sutter</i>	969
16 Atomic Force Microscopy in the Life Sciences <i>Matthias Amrein</i>	1025
17 Low-Temperature Scanning Tunneling Microscopy <i>Uwe Weierstall</i>	1070
PART IV HOLOGRAPHIC AND LENSLESS MODES	1139
18 Electron Holography <i>Rafal E. Dunin-Borkowski, Takeshi Kasama, Martha R. McCartney, and David J. Smith</i>	1141
19 Diffractive (Lensless) Imaging <i>John C.H. Spence</i>	1196
20 The Notion of Resolution <i>S. Van Aert, Arnold J. den Dekker, D. Van Dyck, and A. Van den Bos</i>	1228
Index	II

Contributors

Matthias Amrein

Microscopy and Imaging Facility, Faculty of Medicine, Department of Cell Biology and Anatomy, University of Calgary, Calgary, Canada

Michael R. Armstrong

University of California Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA

Ernst Bauer

Department of Physics, Arizona State University, Tempe, AZ, USA

Wolfgang Baumeister

Max Planck Institute of Biochemistry, Martinsried, Germany

Paolo Bianchini

Department of Physics, LAMBS-IFOM MicroScoBIO Research Centre, University of Genoa, Genoa, Italy

Dawn A. Bonnell

Department of Materials Science and Engineering, Nano-Bio Interface Center, University of Pennsylvania, Philadelphia, PA, USA

Oleg Bostanjoglo

Optisches Institut, Sekr. P1-1, Technische Universität Berlin, Berlin, Germany

Gianluigi Botton

Department of Materials Science and Engineering, McMaster University, Hamilton, Canada

Valentina Caorsi

Department of Physics, LAMBS-IFOM MicroScoBIO Research Centre, University of Genoa, Genoa, Italy

Shery L.-Y. Chang

Department of Materials, University of Oxford, Oxford, UK

Arnold J. den Dekker

Delft Centre for Systems and Control, Delft University of Technology,
Delft, The Netherlands

Alberto Diaspro

Department of Physics, LAMBS-IFOM MicroScoBIO Research Centre,
University of Genoa, Genoa, Italy

Rafal E. Dunin-Borkowski

Department of Materials Science and Metallurgy, University of
Cambridge, Cambridge, UK; also at Centre for Electron Nanoscopy,
Technical University of Denmark, Kongens Lyngby, Denmark

Jun Feng

Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Peter W. Hawkes

CEMES-CNRS, Toulouse, France

Stefan W. Hell

Department for NanoBiophotonics, Max Planck Institute of
Biophysical Chemistry, Göttingen, Germany

Malcolm Howells

Advanced Light Source, Lawrence Livermore National Laboratory,
Livermore, CA, USA

John L. Hutchison

Department of Materials, University of Oxford, Oxford, UK

Christopher Jacobsen

Department of Physics and Astronomy, Stony Brook University, Stony
Brook, NY, USA

Takeshi Kasama

Frontier Research System, Institute of Physical and Chemical
Research, Hatoyama, Japan, and Department of Materials
Science and Metallurgy, University of Cambridge, Cambridge,
UK

Wayne King

Chemistry and Materials Science Directorate, Lawrence Livermore
National Laboratory, Livermore, CA, USA

Angus I. Kirkland

Department of Materials, University of Oxford, Oxford, UK

Davide Mazza

Department of Physics, LAMBS-IFOM MicroScoBIO Research Centre,
University of Genoa, Genoa, Italy

Martha R. McCartney

Department of Physics and Astronomy and Center for Solid-State
Science, Arizona State University, Tempe, AZ, USA

Peter D. Nellist

Department of Physics, Trinity College, Dublin, Ireland; Department of
Materials, University of Oxford, Oxford, UK

Maxim P. Nikiforov

Nano-Bio Interface Center, University of Pennsylvania, Philadelphia,
PA, USA

Mattia Pesce

Department of Physics, LAMBS-IFOM MicroScoBIO Research Centre,
University of Genoa, Genoa, Italy

Juergen Plitzko

Max Planck Institute of Biochemistry, Martinsried, Germany

Bryan W. Reed

Chemistry and Materials Science Directorate, Lawrence Livermore
National Laboratory, Livermore, CA, USA

Rudolf Reichelt

Institut für Medizinische Physik und Biophysik, Westfälische
Wilhelms-Universität, Münster, Germany

Frances M. Ross

IBM Research Division T. J. Watson Research Center, Yorktown
Heights, NY, USA

Marc Schneider

Biopharmaceutics and Pharmaceutical Technology, Campus Saarbrücken,
Germany

Andreas Scholl

Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Andreas Schönle

Department of NanoBiophotonics, Max Planck Institute of
Biophysical Chemistry, Göttingen, Germany

David J. Smith

Department of Physics and Astronomy and Center for Solid-State Science, Arizona State University, Tempe, AZ, USA

John C.H. Spence

Department of Physics, Arizona State University Tempe, AZ, and Lawrence Berkeley Laboratory, Berkeley, CA, USA

Peter Sutter

Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY, USA

Ilaria Testa

Department of Physics, LAMBS-IFOM MicroScoBIO Research Centre, University of Genoa, Genoa, Italy

Cesare Usai

National Research Council Institute of Biophysics, Genoa, Italy

Sandra Van Aert

University of Antwerp, Antwerp, Belgium

A. Van den Bos

Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands

D. Van Dyck

University of Antwerp, Antwerp, Belgium

Giuseppe Vicidomini

Department of Physics, LAMBS-IFOM MicroScoBIO Research Centre, University of Genoa, Genoa, Italy

Tony Warwick

Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Uwe Weierstall

Department of Physics, Arizona State University, Tempe, AZ, USA

Part I

IMAGING WITH ELECTRONS

Atomic Resolution Transmission Electron Microscopy

Angus I. Kirkland, Shery L.-Y. Chang and John L. Hutchison

1 Introduction and Historical Context

1.1 Introduction

High-Resolution Transmission Electron Microscopy (HRTEM) uses a self-supporting thin sample (typically tens of nanometres) illuminated by a highly collimated kilovolt electron beam. A series of magnetic electron lenses image the electron wavefield across the exit face of the sample onto a detector at high magnification. HRTEM has evolved from initial instrumentation constructed by Knoll and Ruska (1932a–c) to its current state where individual atom columns in a wide range of materials can be routinely imaged (Smith, 1997; Krakow et al., 1984) using sophisticated computer-controlled microscopes (Figure 1–1). For this reason HRTEM now occupies a central place in many characterization laboratories worldwide and has made a substantial contribution to key areas of materials science, physics, and chemistry [for key examples showing its wide ranging influence see the frontispiece in the book by Spence (2002)]. Instrument development for HRTEM also supports a substantial commercial industry of manufacturers (Hall, 1966; Hawkes, 1985; Fujita, 1986).¹

Numerous HRTEM studies of bulk semiconductors (Smith and Lu, 1991; Smith et al., 1989), defects (Figure 1–2) (Olsen and Spence, 1981), and interface structures (Figure 1–3), (Bourret et al., 1988; Gutekunst et al., 1994) in these materials, of metals and alloys (Penisson et al., 1988; Krakow, 1990; Ishida et al., 1983; Amelinckx et al., 1993; Thomas, 1962), and of ceramics, particularly oxides (Lundberg et al., 1989; Buseck et al., 1989), have been reported in a vast literature spanning four

¹ We note that HRTEM has also made substantial contributions to structural biology (see, Burge, 1973; Unwin and Henderson, 1975; Henderson, 1995; Koehler, 1973, 1978, 1986 for reviews of selected representative examples from this field; see also Chapter 7 by Plitzko and Baumeister in this volume). However, due to limitations of space we will not consider this aspect further.

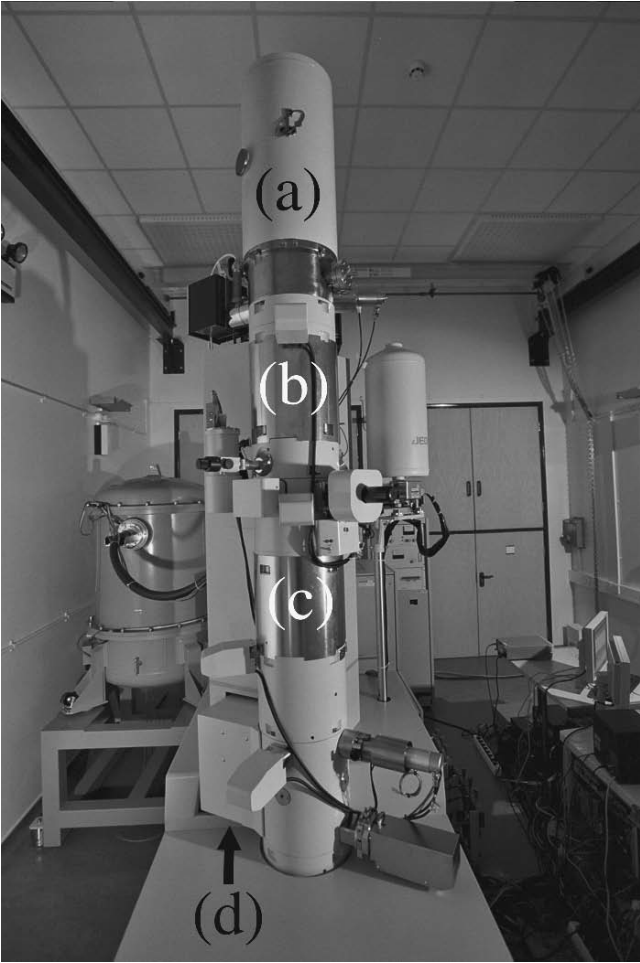


Figure 1-1. A modern 200-kV HRTEM fitted with a (a) field emission gun, (b) probe and (c) image forming aberration correctors, and (d) an omega geometry energy filter.

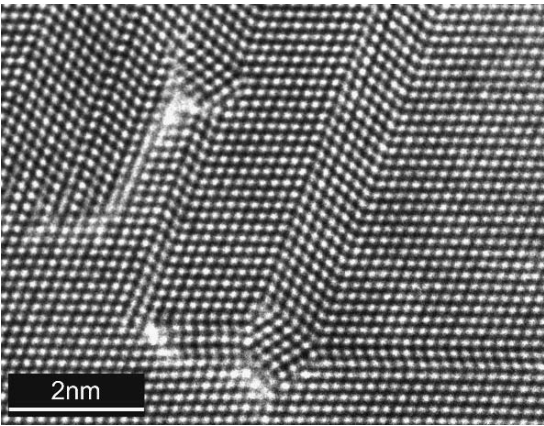
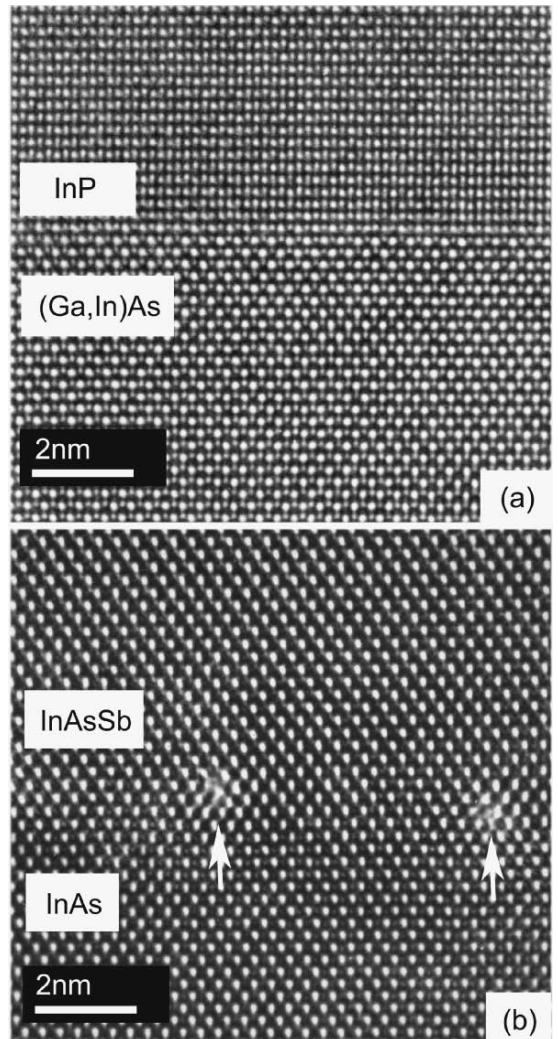


Figure 1-2. HRTEM image of a [110] oriented CVD deposited diamond film showing twins, stacking faults, and nanograins. Note the local disorder at the intersection of the stacking faults and twins.

Figure 1–3. (a) HRTEM image of a lattice matched heterojunction between InP and (Ga, In)As. At the defocus used and for this particular foil thickness differences in scattering between these two isostructural materials allows them to be distinguished. (b) HRTEM image of a heterojunction between InAs and (In, As)Sb that have a significant lattice mismatch. In this case the lattice misfit is accommodated as a regular array of Lomer dislocations marked.



decades (for additional general reviews see Buseck et al., 1989; Smith, 1997; van Tendeloo, 1998; Spence, 1999). An excellent collection of representative HRTEM images can be found in Shindo and Hiraga (1998). More recently HRTEM has become an essential tool in the characterization and discovery of nanoscale materials (Figure 1–4) (see, for example, Iijima, 1991; Yao and Wang, 2005). Of crucial importance in all of these areas is the ability of HRTEM to provide real-space images of the atomic configuration at localized structural irregularities and defects in materials, that are inaccessible to broad-beam bulk diffraction methods and that largely control their properties.

Advances in instrumentation for HRTEM over the same timescale have enabled this information to be recorded with increasing resolution and precision leading to improvements in the quantification of the data obtained.

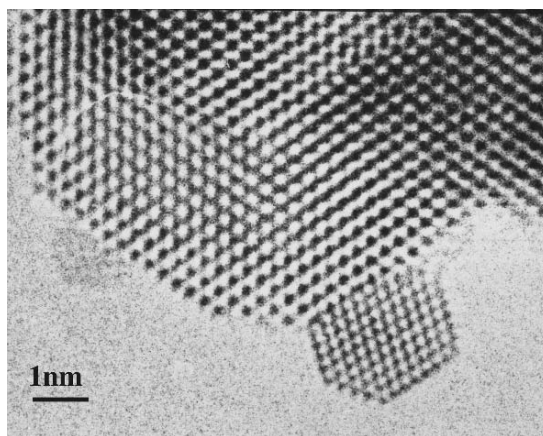


Figure 1–4. HRTEM image of a nanocrystalline gold particle supported on a {111} cerium oxide surface. The gold particle shows an almost perfect cubeoctahedral morphology and both particle and substrate are in an epitaxial [110] orientation despite the large lattice mismatch.

This chapter concentrates on HRTEM at atomic resolution. Following a brief historical overview of the development of HRTEM (for a more detailed article outlining key events in the history of electron microscopy see Haguenau et al., 2003) we begin by outlining some of the theory pertinent to image formation at high resolution and the effects on recorded images of the aberrations introduced by imperfect objective optics. We also provide various definitions of resolution.

The second section surveys the key instrumental components affecting HRTEM and provides an outline of currently available solutions. The final section describes computational approaches to both the recovery of the specimen exit-plane wavefunction (coherent detection) from a series of images and methods available for HRTEM image simulation.

1.2 Historical Summary

Historically, the resolving power of the electron microscope rapidly matched and then exceeded that of the optical microscope in 1934 (Ruska, 1934). However, further improvements in resolution proved relatively slow due to the need to identify and overcome various instrumental limitations (see later).

The first subnanometer lattice-fringe images were obtained in the 1950s (Menter, 1956) from phthalocyanine crystals and this was later extended to lattice images of metal foils showing crossed fringe patterns (Komoda, 1966).

Concurrently, the first published HRTEM images of complex transition metal oxides provided preliminary evidence that many of these (specifically those of Mo, W, Ti, and V) were not perfect structures (Allpress et al., 1969; see also Buseck et al., 1989, for a review). These observations of planar faults in oxides possibly represent the first occa-

sion in which useful information at the atomic level was provided by HRTEM. This work created much interest among solid-state chemists, who for the first time saw a new scientific tool that would enable them to overcome the barriers to structural determinations of these materials imposed by their large unit cells and often extensive disorder. It also immediately provided an explanation for nonstoichiometry in these materials and entirely changed the way in which thermodynamic properties of oxides were modeled.

The work summarized above was possible with the typical instrumental resolutions available in most laboratories at that time. However, it was not until this improved that it became possible to resolve individual cation columns in these and certain other classes of material. In the 1970s the first images showing the component octahedra were published (Cowley and Iijima, 1972) with a resolution of 0.3 nm for a series of mixed Ti–Nb structures that demonstrated a direct correspondence between the lattice image and the projected crystal structure.

The typical spatial resolution (slightly better than ca. 0.5 nm) provided by most commercial instruments in the 1960s and 1970s was largely limited by mechanical and electrical instabilities. Subsequent improvements in instrument design and construction led to a generation of microscopes becoming available in the mid 1970s with point resolutions of less than 0.3 nm operating at intermediate voltages around 200 kV (Uyeda et al., 1972). Toward the end of this period the dedicated 600 kV Cambridge HREM (Cosslett et al., 1979) and several other high-voltage instruments also became operational (Hirabayashi et al., 1982), providing a resolution somewhat better than 0.2 nm.

The following two decades saw further significant improvements in microscope design with dedicated high-resolution instruments being produced by several manufacturers. One outcome of these developments was the installation of commercial high-voltage HRTEMs (operating at ca. 1 MV) in several laboratories worldwide (Gronsky and Thomas, 1983; Matsui et al., 1991). These machines were capable of point resolutions of ca. 0.12 nm, significantly higher than that available in intermediate voltage instruments. Concurrently, commercial instrument development also started to concentrate on improved intermediate voltage instrumentation (at up to 400 kV) (Hutchison et al., 1986) with interpretable resolutions between 0.2 and 0.15 nm.

In the 1990s further progress was made in improving resolution through a combination of key instrumental and theoretical developments. For the former the successful design and construction of improved high-voltage instrumentation (Phillip et al., 1994; Allen and Dornig, 1998) demonstrated interpretable resolutions close to the long sought after goal of 0.1 nm. Perhaps more significantly, field emission sources became widely available on intermediate voltage microscopes (Honda et al., 1994; Otten and Coene, 1993) improving the absolute information limits of these machines to values close to the point resolutions achievable at high voltage.

This new generation of instruments also led to renewed theoretical and computational efforts aimed at reconstructing the complex specimen exit wavefunction using either electron holograms (Lichte, 1991;

Orchowski et al., 1995) or extended focal or tilt series of HRTEM images (see later). These latter “indirect” approaches extended the interpretable resolution beyond conventional axial image limits and provided both the phase and modulus of the specimen exit wavefunction, free from effects due to the objective lens rather than the aberrated intensity available in a conventional HRTEM image.

The latest instrumental developments have seen the successful completion and testing of the necessary electron optical components for direct correction of the spherical aberration present in all electromagnetic round lenses and these corrected instruments are now capable of directly interpretable resolutions close to or below 0.1 nm at intermediate voltages.

2 Essential Theory

In this section we outline some of the essential theory required for understanding HRTEM image contrast. Many of the topics described here are treated in more detail elsewhere (see, for example, Spence, 2002; Reimer, 1984, 1997; Hawkes and Kasper, 1996; Ernst and Rühle, 2003) (for the latter see Chapter 2 in particular) and only selected frameworks directly relevant to HRTEM imaging using modern instrumentation are discussed further.

We begin with a general review of the essentials of the HRTEM image formation process and subsequently treat the key areas of resolution, and the effects of the objective lens and source in more detail.

2.1 Image Formation

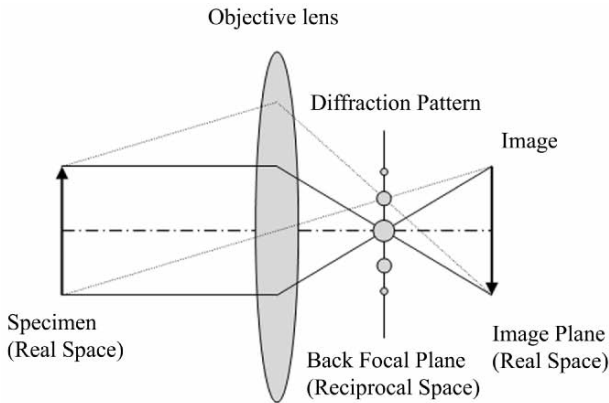
As shown schematically in Figure 1–5 (from a simplified ray optical perspective and from a wave optical perspective as described subsequently) the overall process in the formation of an HRTEM image involves three steps.

1. Electron scattering in the specimen.
2. Formation of a diffraction pattern in the back focal plane of the objective lens.
3. Formation of an image in the image plane.²

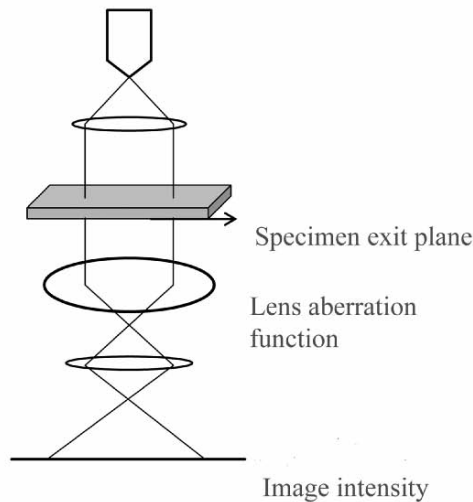
To understand the relationship between contrast recorded in an HRTEM image and the atomic arrangement within the specimen it is essential to develop theoretical frameworks describing each of these steps.

To describe the general case (of arbitrary specimens) each of the above steps requires a complex mathematical and computational treatment that is outside the scope of the section (comprehensive accounts can be found elsewhere, e.g., Spence, 2002; Buseck et al., 1989; Cowley,

² Although not formally derived here it can easily be shown that the specimen, back focal, and image planes are mathematically related by Fourier transform operations.



(a)



(b)

Figure 1–5. (a) Schematic optical ray diagram showing the principles of the imaging process in HRTEM and indicating the reciprocal relationships between specimen, diffraction and image spaces. (b) Schematic diagram illustrating the wave optical relationship between the recorded image intensity in HRTEM and the specimen exit-wave through the objective lens aberration function.

1975; Reimer, 1997; Ernst and Rühle, 2003; Hawkes and Kasper, 1996). We therefore restrict ourselves to treatment of the simplest cases for illustrative purposes and subsequently follow the nomenclature and sign conventions given in Spence (2002).

For thin specimens, neglecting absorption, the effect of the specimen on an incident electron wave is to alter only its phase leaving the amplitude unchanged. Under this *phase object approximation* (POA), which ignores Fresnel diffraction within the specimen but includes the

effects of multiple scattering, the specimen exit-wave complex amplitude can be written as

$$\Psi_e(x, y) = \exp\{-i\sigma\phi_p(x, y)\} \tag{1}$$

where σ is an interaction constant given by

$$\sigma = 2\pi m e \lambda_r / h^2 \tag{2}$$

in which both m and λ_r are relativistically corrected values of the electron mass and wavelength and $\phi_p = \int_{-t/2}^{t/2} \phi(x, y, z) dz$ is the two-dimensional projection of the specimen potential along the beam direction. The interaction constant decreases with accelerating voltage (with values of $0.00729 \text{ V}^{-1} \text{ nm}^{-1}$ at 200 kV decreasing to $0.00539 \text{ V}^{-1} \text{ nm}^{-1}$ at 1000 kV), whereas the specimen inner potential generally increases with atomic number, although this also depends on the density (Shindo and Hiraga, 1998) (Table 1–1).

Equation (1) shows that within this model the effect of the specimen is to advance the phase of the electron wave by $\sigma\phi_p(x, y)$ over the wave in vacuum. For suitably thin specimens of low atomic number the values of the mean inner potential are such that this phase advance is small and hence the exponential term in Eq. (1) can be expanded and approximated as³

$$\Psi_e(x, y) \approx 1 - i\sigma\phi_p(x, y) \tag{3}$$

in the *weak phase object approximation* (WPOA), which assumes kinematic scattering within the specimen requiring that the intensity of the central unscattered beam is significantly stronger than that of the diffracted beams.

It is important to note that both of the above formulations are projection approximations such that atoms within the specimen can be moved along the incident beam direction without affecting the exit wavefunction.

Table 1–1. Mean inner potential of representative materials in volts.

Element	Z (atomic number)	Mean inner potential (V)
C	6	7.8 ± 0.6
Al	13	13 ± 0.4
Si	14	11.5
Cu	29	23.5 ± 0.6
Ge	32	15.6 ± 0.8
Au	79	21.1 ± 2

³ Strictly it is the variation in the phase change produced by different parts of the specimen that is important, which supports this approximation.

The complex amplitude of the scattered wave in the back focal plane of the objective lens is given by the Fourier transform of Eq. (3). With $\phi_p(x, y)$ real this gives

$$\Psi_d(u, v) = \delta(u, v) - i\sigma F\{\phi_p(x, y)\} \quad (4)$$

Equation (4) is subsequently modified by the presence of a limiting objective aperture and by phase shifts introduced by the objective lens.

The former can be included through the simple function

$$\begin{aligned} P(\mathbf{k}_x, \mathbf{k}_y) &= 1 & |\mathbf{k}| &\leq r. \\ P(\mathbf{k}_x, \mathbf{k}_y) &= 0 & |\mathbf{k}| &\geq r. \end{aligned} \quad (5)$$

The phase shifts introduced by the objective lens are parameterized by the coefficients of a wave aberration function, $W(u, v)$, which is treated in detail in a subsequent section.

Thus, including aperture and lens effects the complex amplitude, under the WPOA, is given by

$$\psi'_d(\mathbf{k}_x, \mathbf{k}_y) = \delta(\mathbf{k}_x, \mathbf{k}_y) - i\sigma F\{\phi_p(x, y)\} P(\mathbf{k}_x, \mathbf{k}_y) \exp[iW(\mathbf{k}_x, \mathbf{k}_y)]. \quad (6)$$

A further Fourier transform of Eq. (6) finally gives the image amplitude (in the image plane) as

$$\psi_i(x, y) = 1 - i\sigma \{\phi_p(-x, y)\} F\{P(\mathbf{k}_x, \mathbf{k}_y) \exp[iW(\mathbf{k}_x, \mathbf{k}_y)]\}. \quad (7)$$

Since the cosine terms in the expansion of equation (7) cancel, the recorded image intensity, to first order, is

$$\begin{aligned} I(x, y) &= \psi_i(x, y) \psi_i^*(x, y) \\ &\approx 1 + 2\sigma \phi_p(-x, -y) * \mathcal{F}\{\sin[W(\mathbf{k}_x, \mathbf{k}_y)] P(\mathbf{k}_x, \mathbf{k}_y)\}. \end{aligned} \quad (8)$$

The above expression shows that for this simplest theory the image contrast is proportional to the projection of the specimen potential convolved with an impulse response function arising from the instrument.

Detailed treatment of the latter requires the inclusion of the effects due to the partial coherence of the electron source, which acts to damp higher spatial frequencies (see later), and of the detector, which also modifies the recorded contrast through its modulation transfer function (see later).

A useful modification to the above treatment makes the potential $\phi_p(x, y)$ complex.

This complex projected specimen potential (Cowley and Pogany, 1968) provides a description of the attenuation of the image wavefield through either scattering outside a limiting aperture or more usefully for unfiltered HRTEM by the depletion of the elastic wavefield by inelastic processes (Yoshioka, 1957).

A number of further extensions to this basic treatment have previously been proposed to overcome the limitations of a projection approximation in the thick phase grating approximation (Cowley and Moodie, 1962). We do not give a detailed derivation here but note that this approximation successfully accounts for multiple scattering and a degree of curvature of the Ewald sphere (Fresnel diffraction within the specimen) and is thus more generally applicable to HRTEM imaging under less restrictive conditions than the WPOA.

The *projected charge density* (PCD) approximation is an alternative extension that provides a tractable analytic expression for the image intensity including the effects of multiple scattering (unlike the weak phase object) but retaining the restriction of a projection approximation.

Starting from the expression for the specimen exit-wave complex amplitude given by the POA in the absence of an objective aperture and with no wave aberration function we can write

$$\Psi_e(x, y) = \exp\{-i\sigma\phi_p(x, y)\} \quad (9)$$

The amplitude in the back focal plane of the objective lens is given by Fourier transformation of the above as

$$\begin{aligned} \psi_d(\mathbf{k}_x, \mathbf{k}_y) &= F\{\exp[-i\sigma\phi_p(x, y)]\} \exp[i\pi\Delta C_1\lambda(u^2 + v^2)] \\ &\approx \Phi(\mathbf{k}_x, \mathbf{k}_y)[1 + i\pi\Delta C_1\lambda|\mathbf{k}|^2]. \end{aligned} \quad (10)$$

if only a small defocus, ΔC_1 , is allowed and where $\Phi(\mathbf{k}_x, \mathbf{k}_y)$ represents the Fourier transform of $\exp[i\sigma\phi_p(x, y)]$.

Thus the image amplitude (in the image plane) is given by

$$\psi_i(x, y) = \exp(-i\sigma\phi_p(x, y)) + i\pi\Delta C_1\lambda F^{-1}\{(\mathbf{k}_x^2 + \mathbf{k}_y^2)\Phi(\mathbf{k}_x, \mathbf{k}_y)\}. \quad (11)$$

A standard theorem from Fourier analysis is now used (Bracewell, 1965), which states that if $f(x, y)$ and $\Phi(u, v)$ are a Fourier transform pair then

$$F^{-1}\{(\mathbf{k}_x^2 + \mathbf{k}_y^2)\Phi(\mathbf{k}_x, \mathbf{k}_y)\} = -\frac{1}{4}\pi^2[\nabla^2 f(x, y)]. \quad (12)$$

Applying this result the image amplitude is given by

$$\begin{aligned} \psi_i(x, y) &= \exp[-i\sigma\phi_p(x, y)] - i\Delta C_1\lambda/4\pi\nabla^2\{\exp[-i\sigma\phi_p(x, y)]\} \\ &= \exp[-i\sigma\phi_p(x, y)] + (i\Delta C_1\lambda\sigma/4\pi)\exp[-i\sigma\phi_p(x, y)] \\ &\quad \times \{\sigma\nabla\phi_p(x, y) + i\nabla^2\phi_p(x, y)\} \end{aligned} \quad (13)$$

which yields an image intensity to first order as

$$I(x, y) \approx 1 - (\Delta C_1\lambda\sigma/2\pi)\nabla^2\phi_p(x, y) \quad (14)$$

From Poisson's equation $\nabla^2\phi_p(x, y) = -\rho_p(x, y)/\epsilon_0\epsilon$ we finally obtain

$$I(x, y) = 1 + (\Delta C_1\lambda\sigma/2\pi\epsilon_0\epsilon)\rho_p(x, y). \quad (15)$$

in which $\rho_p(x, y)$ is the projected total charge density including the nuclear contribution.

Historically, the restriction of limited defocus and no spherical aberration (or other uncorrected aberrations) meant that the application of the PCD approximation was restricted to relatively low resolution imaging. However, this approximation would now seem to be ideal for the interpretation of aberration corrected images (see later) in which these restrictions can be experimentally achieved at high resolution (O’Keefe, 2000). Finally, we note that this theory has also been modified (Lynch et al., 1975; Chang, 2000) to include the effects due to higher order lens aberrations and the presence of a limiting objective aperture.

Further extensions to the models outlined above require solution of the dynamic electron diffraction problem using one of several possible computational algorithms, a description of which is outside the scope of this section (Goodman and Moodie, 1974; Cowley, 1975; Self et al., 1983; Jap and Glaeser, 1978; van Dyck, 1983; van Dyck and Coene, 1984; Coene and van Dyck, 1984a,b; Hirsch et al., 1965; Stadelmann, 1987, 1991; Spence and Zuo, 1992; Chen and van Dyck, 1997; Kirkland, 1998; Ernst and Rühle, 2003). However, for generalized HRTEM image simulation the multislice algorithm (Cowley, 1959a–c, 1975; Goodman and Moodie, 1974) has been most commonly employed for the simulation of HRTEM images, to a large extent due to its computational efficiency compared to alternative methods (van Dyck and Op de Beeck, 1994) such as Bloch wave calculations, and this is therefore outlined in a subsequent section.

2.2 Resolution Limits

Unlike their optical equivalents there is no simple measure of resolution for the electron microscope, as the resolution depends on both the instrument and also on the scattering properties of the sample used.⁴

The ultimate resolution of any optical system is the diffraction limit imposed by the wavelength of the radiation, λ , and the aperture angle of the objective lens, α , and the refractive index, n , which can be formalized through Abbe’s equation as⁵

$$r_d = k\lambda/n\sin(\alpha) \quad (16)$$

However, due to imperfections in the objective lens and limited coherence (as discussed subsequently) experimental resolution limits are far lower than that set by Eq. (16) and hence a “single figure” definition of resolution for HRTEM is not possible.

Two independent definitions of attainable resolution are commonly used, defined, respectively, by the key optical properties of the objective lens and those of the source.

⁴ For a more detailed treatment of resolution see Chapter 20 by van Aert et al. in this volume.

⁵ The value of the constant k lies between 0.6 and 0.8 depending on the coherence of the illumination.

The first of these is the “directly interpretable” or Scherzer limit (Scherzer, 1949) or point resolution, which defines the maximum width of a pass band transferring all spatial frequencies from zero, without phase reversal, and is determined by the coefficients of the wave aberration function of the objective lens (see later).

Ignoring the phase shifts due to higher order aberrations (see later) the *phase contrast transfer function* (PCTF) (Hawkes and Kasper, 1996) due to defocus and spherical aberration is given by

$$\sin W(\mathbf{k}) = \sin\{\pi C_1 \lambda |\mathbf{k}|^2 + \frac{\pi}{2} C_3 \lambda^3 |\mathbf{k}|^4\}. \quad (17)$$

For HRTEM this defines a focus setting (Scherzer, 1949) that offsets the phase shift due to spherical aberration, C_3 through a suitable choice of defocus:⁶

$$C_{1,\text{Scherzer}} = -1.2(C_3 \lambda)^{1/2} \quad (18)$$

which leads to a broad band of phase contrast transfer without zero crossings (Figure 1–6) up to a frequency of $\mathbf{k}_{\text{max}} = 1.6(C_3 \lambda^3)^{-1/4}$.

The reciprocal gives the point resolution as

$$d_1 = 0.625(C_3 \lambda^3)^{1/4} \quad (19)$$

Thus, HRTEM images of thin specimens recorded at the Scherzer defocus (or its extended variant) will have components that are directly proportional to the (negative of) the *projected potential* of the specimen extending to spatial frequencies equal to the interpretable resolution limit (Cowley and Iijima, 1972; Hanßen, 1971). For higher spatial frequencies up to that defined by the information limit the contrast is partially reversed as the PCTF starts to oscillate (Figure 1–6).⁷

Given the form of Eq. (19) it is evident that a decrease in electron wavelength has a greater effect than an equivalent decrease in the spherical aberration, and for this reason high-voltage instrumentation (see earlier) has, until recently, been the preferred route to achieving higher interpretable resolutions.

The higher resolution, the *information limit*, defines the highest spatial frequency transferred from the specimen exit wavefunction to the image intensity. This is determined by the effects of spatial and temporal coherence (see later) in the illumination and by mechanical instabilities and acoustic noise that also act to damp the transfer of higher spatial frequencies.

⁶ In this definition of the Scherzer focus, the passband in the CTF contains a local minimum = 0.7. The original definition, $C_{1,\text{Scherzer}} = (C_3 \lambda)^{1/2}$, avoids a local minimum in the passband at the cost of a slightly poorer point resolution $\rho_s = 0.707(C_3 \lambda^3)^{1/4}$.

⁷ The above definitions of point resolution assume a fixed positive C_3 . Variable C_3 in corrected microscopes modifies these results as detailed in a subsequent section.

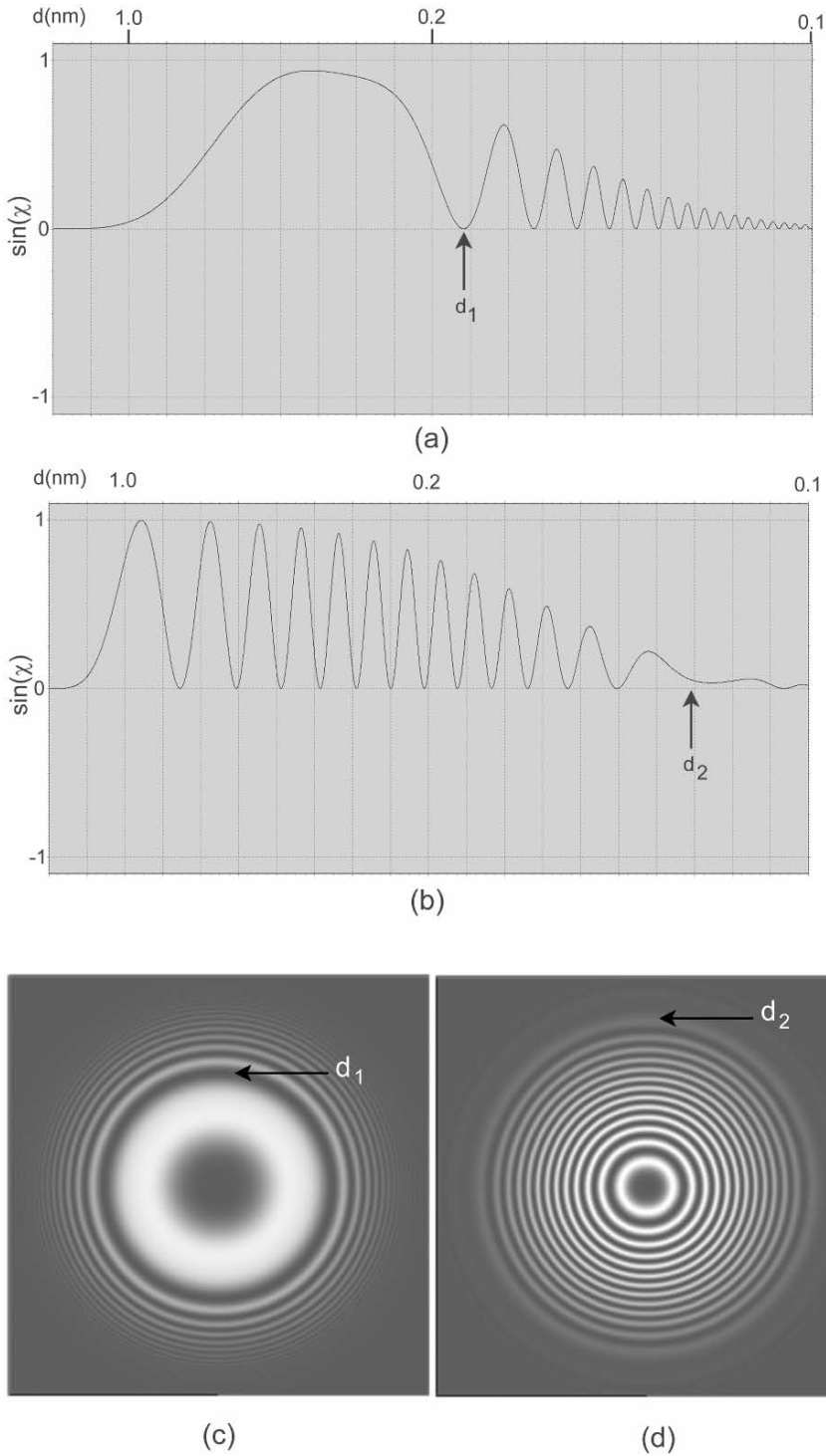


Figure 1-6. Phase contrast transfer functions (PCTFs) plotted in one dimension (a and b) and two dimensions (c and d) calculated for a modern 300-kV FEGTEM illustrating the interpretable resolution (d_1) and information limits (d_2). (a and c) calculated at the Scherzer defocus (-34.4 nm) (b and d) calculated at the higher Lichte underfocus (-174.4 nm). In all cases CTFs are calculated for 300 kV, $C_3 = 0.6$ mm, $\Delta E = 0.8$ eV, beam divergence = 0.1 mrad and the interpretable and information limits are indicated.

As will be discussed in detail in a subsequent section the effects of temporal and spatial coherence (see also Hawkes and Kasper, 1996) can be treated through respective envelope functions of form

$$\begin{aligned} E_t(\mathbf{k}) &= \exp\{-0.5\pi^2\Delta^2(\lambda\mathbf{k}^2)^2\} \\ E_s(\mathbf{k}) &= \exp\{-\pi^2q_0^2(C_1\lambda\mathbf{k} + C_3\lambda^3\mathbf{k}^3)^2\} \end{aligned} \quad (20)$$

in which the expression for the spatial envelope includes only defocus and spherical aberration, with q_0 , the standard deviation of a Gaussian modeling the convergent cone of illumination at the specimen surface, and where

$$\Delta = C_c \left[\frac{\sigma^2(V_0)}{V_0^2} + \frac{4\sigma^2(I_0)}{I_0^2} + \frac{\sigma^2(E_0)}{V_0^2} \right]^{1/2} \quad (21)$$

In Eq. (21) C_c is the chromatic aberration, E_0 is the spread in electron energies arising from the source, V_0 is the accelerating voltage, and I_0 is the objective lens current, which affects the objective lens magnetic field (O'Keefe, 1992).^{8,9}

The form of these distributions is far less important than their width and in general, as above, Gaussian distribution functions are used, as this makes any further derivation analytically tractable.^{10,11}

The form of these two envelope functions leads to definitions of information limits in which the information transfer drops to a level of $\exp(-2)$ or 13.5% given by (O'Keefe and Pitt, 1980).

$$\begin{aligned} d_2 &= \sqrt{\pi\lambda d/2} \\ d_3 &= S_+^{1/3} + S_-^{1/3} \\ S_{\pm} &= \left\{ \frac{3.3}{4\pi q_0} \pm \left[\frac{C_1^3}{27C_3\lambda^2} + \left(\frac{3.3}{4\pi q_0} \right)^2 \right]^{1/2} \right\} / C_3\lambda^2 \end{aligned} \quad (22)$$

However, it should be noted that the limit defined by d_3 increases with defocus and does not therefore define an absolute information limit

⁸ The last term is written here in terms of the objective lens current I_0 following convention. However, this is not strictly correct, particularly if the lens is operated close to saturation where the magnetic field is not proportional to the current.

⁹ It may seem surprising that the objective lens current influences the variation of focus only in the presence of chromatic aberration. However, this is due to the general scaling rule that electron trajectories are identical when energy and magnetic field are changed according to $E' = k^2E$ and $B' = kB$. Hence, in a microscope corrected for chromatic aberration, the focus cannot be changed by changing the currents in all lens elements by the same factor.

¹⁰ For the defocus spread, the terms due to voltage and lens current instabilities are well described by Gaussian function, whereas the intrinsic source energy spread would be more accurately described by a Maxwellian distribution for thermionic emitters or the Fowler–Nordheim equation (Fowler and Nordheim, 1928) for field emitters.

¹¹ The above definition of δ assumes that the fluctuations are independent of each other.

per se. In particular, for the case of field emission sources where the source size is small, the spatial coherence is not limiting and thus the information limit due to temporal coherence determines the information limit for all defoci.

2.3 The Wave Aberration Function

The key optical component affecting HRTEM image formation is the objective lens and in this section we review its influence in terms of the wave aberration function (see also Hawkes and Kasper, 1996).

For an ideal lens, a point object at a position (x, y) in the object plane leads to a spherical wavefront in the diffraction plane, contracting to a conjugate point in the image plane.¹² However, all electromagnetic lenses suffer from aberrations causing deviations from this ideal spherical wavefront thus reducing the sharpness of an image point much more severely than the diffraction limit.

For HRTEM, a wave aberration function $W(u, v)$ is therefore defined that describes the distance between the ideal and actual wavefronts in the diffraction plane as a function of the position of the point object in the diffraction plane, (u, v) (Saxton, 1995) (Figure 1-7).¹³

The wave aberration function, $W(\omega)$, written in terms of a complex position variable $\omega = u + iv$ can be Taylor expanded to 3rd order in terms of the axial aberrations as

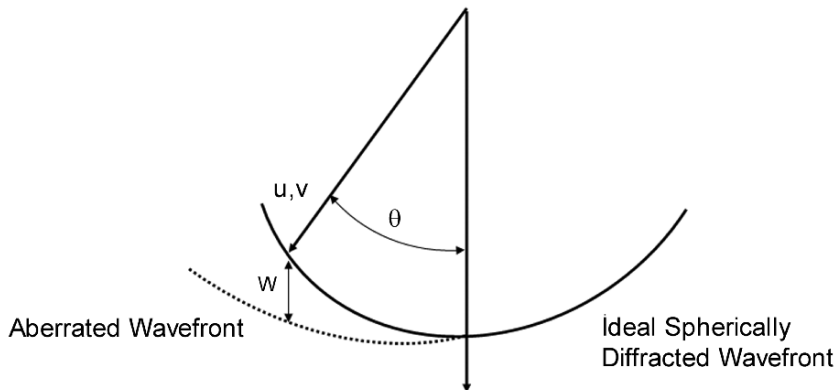


Figure 1-7. Schematic diagram showing the origin of the wave aberration function, $W(u, v)$, which describes the complex deviation from an ideal spherically diffracted wave.

¹² For a pure phase object this condition would, however, lead to zero contrast at the Gaussian focus.

¹³ More rigorously W is a function of both position of the point object in the diffraction and image planes, i.e., $W(x, y, u, v)$. However, HRTEM experiments are carried out at high magnification and hence the field of view is small and the dependence of W on (x, y) can be neglected in the *isoplanatic* approximation.

$$\begin{aligned}
W(\omega) = \Re\left\{\frac{2\pi}{\lambda}\left\{A_0\lambda\omega^* + \frac{1}{2}A_1\lambda^2\omega^{*2} + \frac{1}{2}C_1\lambda^2\omega^*\omega\right.\right. \\
+ \frac{1}{3}A_2\lambda^3\omega^{*3} + \frac{1}{3}B_2\lambda^3\omega^{*2}\omega \\
+ \frac{1}{4}A_3\lambda^4\omega^{*4} + \frac{1}{4}S_3\lambda^4\omega^{*3}\omega + \frac{1}{4}C_3\lambda^4\omega^{*2}\omega^2 \\
\left.\left. + \dots\right\}\right\}
\end{aligned} \tag{23}$$

Converting to polar notation, with $\omega = \mathbf{k}e^{i\phi}$, $A_n = |A_n|e^{i\alpha_n}$ and $B_n = |B_n|e^{i\beta_n}$, Eq. (23) can be more conveniently rewritten as

$$\begin{aligned}
W(\mathbf{k}, \phi) = \frac{2\pi}{\lambda}\{ & |A_0|\lambda k \cos(\phi - \alpha_0) \\
& + \frac{1}{2}|A_1|\lambda^2 k^2 \cos 2(\phi - \alpha_1) + \frac{1}{2}C_1\lambda^2 k^2 \\
& + \frac{1}{3}|A_2|\lambda^3 k^3 \cos 3(\phi - \alpha_2) + \frac{1}{3}|B_2|\lambda^3 k^3 \cos(\phi - \beta_2) \\
& + \frac{1}{4}|A_3|\lambda^4 k^4 \cos 4(\phi - \alpha_3) + \frac{1}{4}|S_3|\lambda^4 k^4 \cos 2(\phi - \sigma_3) \\
& + \frac{1}{4}C_3\lambda^4 k^4 + \dots\}
\end{aligned} \tag{24}$$

which makes the azimuthal and radial dependence of the various aberration terms more apparent.

Table 1–2 lists the aberration coefficients important in HRTEM ignoring those with a chromatic dependence.¹⁴

All of the aberration coefficients listed in Table 1–2 except C_1 and C_3 are due to lens imperfections (either mechanical or electrical) and would not appear in a perfect round lens (Scherzer, 1949). However, they can be corrected using combinations of multipole fields of appropriate symmetry and in recent years this approach has enabled the

Table 1–2. Isoplanatic aberration coefficients important for HRTEM.

Aberration	Order in ω	Azimuthal symmetry	Name and description
A_0	1	1	Image shift
A_1	2	2	Two-fold astigmatism
A_2	3	3	Three-fold astigmatism
A_3	4	4	Four-fold astigmatism
B_2	3	1	Axial coma
S_3	4	2	Axial star
C_1	2	∞	Defocus
C_3	4	∞	Spherical aberration

¹⁴ The seemingly counterintuitive notation (e.g., C_1 in Table 1–2 for a **second-order** term in the wave aberration function stems from the ray-optical theory of Seidel aberrations (Hawkes and Kasper, 1989), which are described in terms of displacements of ray-path intersections with the image plane as a function of (u, v) . These displacements are proportional to the *gradient* of the wave aberration function and hence an n th-order Seidel aberration corresponds to a term of order $n + 1$ in the wave aberration function.

Table 1–3. Accuracy of aberration coefficients.^a

Resolution		Accuracy			
$K_{\max}$ (/nm ⁻¹)	$d_{\min}$ (nm)	A_1, C_1 (nm)	A_2, B_2 (nm)	A_3, S_3, C_3 (μm)	Tilt (μrad)
11	0.09	0.5	35	2.1	21
10	0.1	0.6	47	3.1	27
9	0.11	0.8	64	4.8	38
8	0.125	1.0	92	7.6	54
7	0.14	1.3	137	13	80

^a Accuracy to which the aberration coefficients need to be determined such that each of them causes a maximum RMS error in the wave aberration function of less than $\lambda/16$, i.e., a phase change of less than $\pi/8$ for a given target resolution. The values are calculated for an accelerating voltage of 300 kV ($\lambda \approx 2$ pm). The necessary accuracy for the beam tilt τ is calculated from that for B_2 using $\Delta\tau = -\Delta B_2/(3C_3)$ with $C_3 = 0.57$ mm.

correction of C_3 using long hexapole fields (see later) (Rose, 1981, 1990; Haider et al., 1998c; Urban et al., 1999).

For HRTEM imaging the effects of the coefficients of the wave aberration function are to introduce phase shifts in the image given by multiplying each term in W by $2\pi/\lambda$. These phase shifts can also be used to estimate the maximum tolerable value in any coefficient for a particular target resolution (Table 1–3).

Few of the aberration coefficients defined above are directly observable under axial illumination and their determination therefore relies on measurements taken as a function of known injected beam tilts.

If the illumination is tilted by an angle τ the new aberration coefficients up to C_3 (marked with a prime) are given by

$$\begin{aligned}
 A'_0 &= A_0 + A_1\tau^* + A_2\tau^{*2} + C_1\tau + \frac{1}{3}B_2^*\tau^2 + \frac{2}{3}B_2\tau^*\tau + C_3\tau^*\tau^2 \\
 A'_1 &= A_1 + 2A_2\tau^{*2}\tau + \frac{2}{3}B_2\tau + C_3\tau^2 \\
 C'_1 &= C_1 + \mathcal{R}\left(\frac{4}{3}B_2\tau^*\right) + 2C_3\tau^*\tau \\
 A'_2 &= A_2 \\
 B'_2 &= B_2 + 3C_3\tau \\
 C'_3 &= C_3
 \end{aligned} \tag{25}$$

From Eq. (25) it is immediately apparent that the shift $A'_0 - A_0$ between two images taken at beam tilts differing by τ depends on all the other aberration coefficients. Hence, measuring the image shifts induced by a suitable set of beam tilts provides a method for measuring all coefficients.

Alternatively, the tilt-induced changes, $C'_1 - C_1$ and $A'_1 - A_1$ in defocus or two-fold astigmatism measured from diffractograms (Figure 1–8), depend on A_1, C_1, A_2, B_2 , and C_3 , thus providing an alternative measure of these coefficients. If the spherical aberration is independently determined then measurement of the *orientations only* of diffractograms also provides another simple measure of A_1, A_2 , and B_2 (Saxton, 2000).

Measurements of diffractograms (Krivanek, 1976; Krivanek and Leber, 1994; Zemlin et al., 1978; Zemlin, 1979; Typke and Dierksen, 1995; Pan, 1998; Saxton, 1995) or image shifts (Typke and Dierksen, 1995;

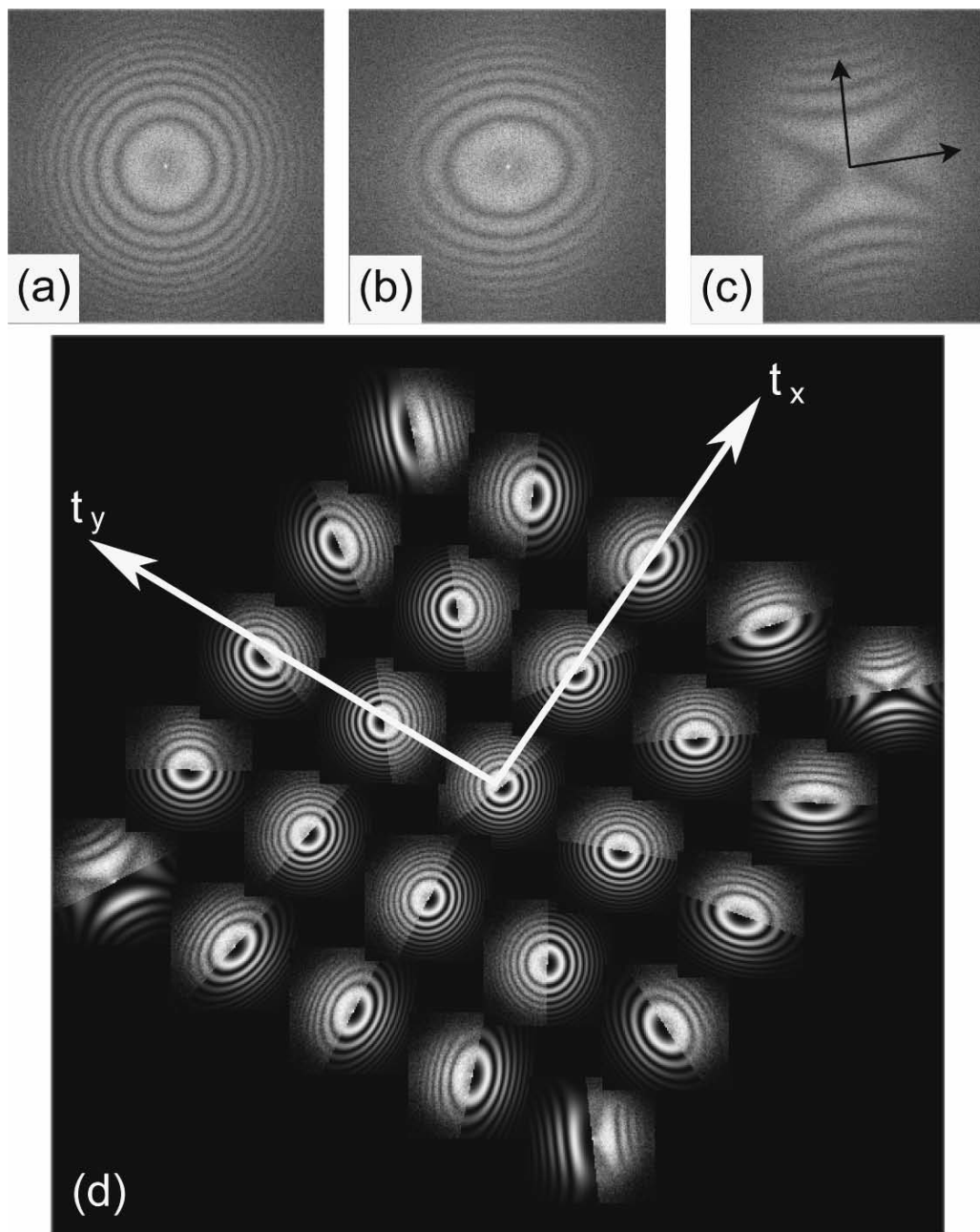


Figure 1-8. Individual axial diffractograms calculated from HRTEM images of a thin amorphous germanium foil showing the effects of defocus (C_1) and two-fold astigmatism (A_1). (a) Axial image recorded at large underfocus with the two-fold astigmatism corrected. (b) Axial image in the presence of a small amount of two-fold astigmatism giving rise to an elliptical diffractogram. (c) As for (b) but with a larger two-fold astigmatism giving rise to a “Maltese Cross”-shaped diffractogram showing overfocus and underfocus along the axes indicated. (d) Representative tableau of diffractograms recorded for a set of differently tilted illumination conditions illustrating the variation in A_1 and C_1 with illumination tilt. Each experimental diffractogram is shown merged with a simulated diffractogram calculated from the fitted values of A_1 and C_1 . The illumination tilt axes are indicated.

Koster et al., 1987, 1989; Koster, 1989; Koster and de Ruijter, 1992; Saxton, 1995), however, have their own particular experimental advantages and disadvantages.

A practical difficulty with the application of tilt-induced displacements arises from the measurement of the image shifts using the peak position in the cross-correlation function defined as

$$\text{XCF} = FT^{-1}[c_1^* c_2] \quad (26)$$

The tilt-induced change in A_0 introduces a linear phase variation in the cross spectrum $c_1^* c_2$, which leads to a displacement of the XCF peak to a position given by the shift vector between the images. As already shown, the other imaging parameters also change as the beam is tilted and this causes the phase variation to become nonlinear at higher spatial frequencies and leads to distorted cross-correlation peaks. However, when the imaging conditions in both images are approximately known, these nonlinear phase shifts can be compensated and a sharp XCF peak can be recovered (Saxton, 1994; Kirkland et al., 1995). Shift measurements also fail for clean perfectly periodic specimens in which image positions differing by any integer multiple of a lattice vector cannot be distinguished. A more severe problem associated with the tilt-induced shift method is that any displacement due to specimen drift is indistinguishable from the tilt-induced displacement required and hence this approach is most frequently used at low resolution or for initial coarse alignment at high resolution.

Conversely, diffractogram measurements require the presence of an area of thin disordered material and are thus less generally applicable. However, they are insensitive to specimen drift and their measurement is relatively straightforward even under tilted illumination conditions. Hence defocus and astigmatism measurements are best suited to fine adjustment at high resolution (Pan, 1998).

Historically the use of diffractograms for the determination of aberrations was first suggested by Thon (1966) for the measurement of defocus from ring positions in an optically generated diffractogram. This method was later extended through the use of diffractogram tableaux acquired with different beam tilt azimuths (Figure 1–8) to the measurement of the spherical aberration C_3 , axial coma B_2 , and three-fold astigmatism A_2 by Krivanek (1976) and Zemlin et al. (1978).

However, the diffractogram tableau method was computationally too demanding for routine use at this time and was used only to demonstrate that the alignment achieved from current reversal or voltage centering was inadequate as an alignment for HRTEM (Saxton et al., 1983).¹⁵

¹⁵ The *current reversal center* alignment involves reversing the current of the objective lens and is no longer practical with the strong lenses used in modern instruments. It should not be confused with the *current center* alignment, where the objective lens current is oscillated by a small amount. Similarly, in the *voltage center* alignment, the high tension is oscillated. Generally, the axes found by the three methods are distinct and the voltage center provides a workable approximation to the *coma-free axis*, but for HRTEM coma-free alignment is preferred.

Fitting C_1 and A_1 to match simulated diffractogram patterns with experimental ones can be done manually to relatively high accuracy (Figure 1–8d; Chand, 1997). This, however, is a lengthy and tedious process and automation of this task is highly desirable. This problem has been addressed by Baba et al. (1987) and Fan and Krivanek (1990) in an algorithm where the diffractogram is divided into 32 sectors and the defocus along the directions in each sector is determined by cross-correlating the rotational sector average with an array of theoretical diffractograms. The sinusoidal focus variation as a function of the azimuth angle obtained is subsequently fitted to the measured focus values to determine C_1 and A_1 and automatic alignment can be achieved using a tableau with as few as four tilt azimuths.

More generally, the major challenges for implementing a robust automated diffractogram fitting procedure are as follows:

1. The most abundant amorphous material in the microscope, carbon, is a weak scatterer. Hence, particularly at high spatial frequencies, the signal is weak compared to the noise background affecting overall diffractogram quality.
2. The strength of the observed signal depends on the object as well as on the phase contrast transfer function.
3. Diffractograms taken close to Scherzer or Gaussian defocus show few rings and are difficult to fit.
4. In the presence of large two-fold astigmatism and at close to Gaussian defocus the diffractograms are cross- rather than ring-shaped, which leads to difficulties for algorithms based on the evaluation of rotationally averaged sectors of the diffractogram.
5. It is difficult to distinguish between overfocus and underfocus.
6. Most automated algorithms fail when a significant amount of crystalline material is present, leading to strong reflections at positions unrelated to the ring pattern.

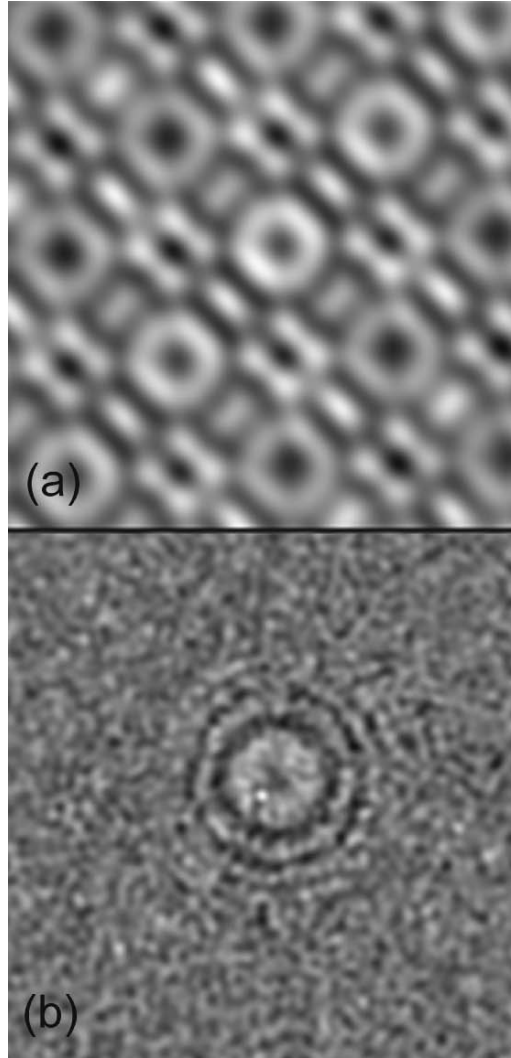
More recently the advent of C_3 -corrected microscopes (see later) has made automated aberration measurement more important because the nonround lens elements introduce a multitude of high-order aberrations (up to six-fold astigmatism) that require correction in an elaborate alignment procedure. Uhlemann and Haider (1998) have implemented an algorithm that can evaluate the apparent defocus and astigmatism from a diffractogram in less than 400 ms based on a comparison of the experimental diffractogram with an extensive library of precalculated diffractograms and is robust to disturbances from the presence of crystalline material.

However, given the limitations of both of these traditional methods an alternative method has also been proposed, which is both applicable to a wide range of specimens and sufficiently accurate for use in atomic resolution imaging (Meyer, 2002; Meyer et al., 2002, 2004).

In this approach a *phase correlation function* (PCF) (Figure 1–9) (Kuglin and Hines, 1975) is initially calculated (defined as the conventional XCF with the modulus set to unity):

$$PCF(\mathbf{x}) = FT^{-1} \left[F(\mathbf{k}) \frac{c_1(\mathbf{k})^* c_2(\mathbf{k})}{|c_1(\mathbf{k})^* c_2(\mathbf{k})|} \right] \quad (27)$$

Figure 1–9. Comparison of a (a) cross- and (b) phase correlation functions (XCF and PCF) between two images of a crystalline material with a defocus difference of 69 nm. Due to the periodicity of the crystalline specimen, the XCF peak repeats periodically but the PCF does not show this repetitive pattern and consists of a single peak broadened into a concentric ring pattern due to the defocus difference.



where $c_i(\mathbf{k})$ are the image Fourier transforms and $F(\mathbf{k})$ is a rotationally symmetric weighting factor used to suppress high-frequency noise. The modulus normalization is essential as it suppresses the crystal reflections that make the conventional XCF periodic, thus allowing aberration measurement in the absence of amorphous material.

The PCF calculated between two images recorded at different defocus levels consists of a centrosymmetric ring pattern, the exact form of which depends on the *relative* defocus between them (assuming that the other aberration coefficients remain constant between the two images) (Figure 1–9). It is possible to compensate for the phase shifts giving rise to this ring pattern by the application of a phase factor dependent on the defocus difference, thus defining a phase compensated PCF

$$PCF_w(\mathbf{x}) = FT^{-1} \left\{ F(\mathbf{k}) \frac{\cos[W_D(\mathbf{k})]c_1^*(\mathbf{k})c_2(\mathbf{k})}{|\cos[W_D(\mathbf{k})]c_1^*(\mathbf{k})c_2(\mathbf{k}) + h|} \right\} \quad (28)$$

in which the very small positive number, h , prevents a zero denominator and where $W_D(\mathbf{k}) = \pi\Delta C_1\lambda|\mathbf{k}|^2$ describes the propagation from the first to the second image in the presence of *only* a defocus difference ΔC_1 . When the value of ΔC_1 matches the actual focus difference the phase compensated PCF collapses to a sharp localized correlation peak. Hence the relative defocus difference between two images can be determined by simply maximizing peak height in the phase compensated PCF as a function of the compensated defocus difference (Figure 1–10), which with subsequent refinement can yield relative defoci to an accuracy better than 1 nm (Meyer et al., 2002).

In the second step an initial restored image wavefunction (see later for details of the restoration process), $\psi_i(\mathbf{k})$, in the plane of a reference image is restored and the *absolute* values of A_1 and C_1 are determined using a *phase contrast index function* (PCI) f_{PCI} given by

$$f_{\text{PCI}}(\mathbf{k}, C_1, A_1) = -\cos[\arg(\psi_{\text{si}}(\mathbf{k})) + \arg[\psi_{\text{si}}(-\mathbf{k})] + 2W(\mathbf{k}, C_1, A_1)] \quad (29)$$

where

$$W(\mathbf{k}, C_1, A_1) = \frac{1}{2}|A_1|\lambda\mathbf{k}^2 \cos 2(\phi - \alpha_1) + \frac{1}{2}C_1\lambda\mathbf{k}^2 + \frac{1}{2}C_3\lambda^3\mathbf{k}^4. \quad (30)$$

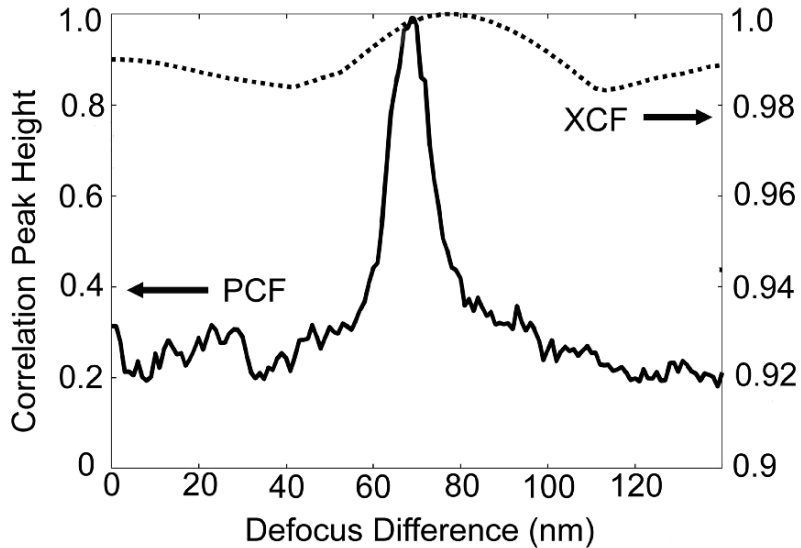


Figure 1–10. Peak height of the PCF (solid line) and XCF (dashed line) (shown against the left hand side and right hand side scales) between two images as a function of compensated focus difference showing a sharp maximum at the correct compensating focus difference in the PCF.

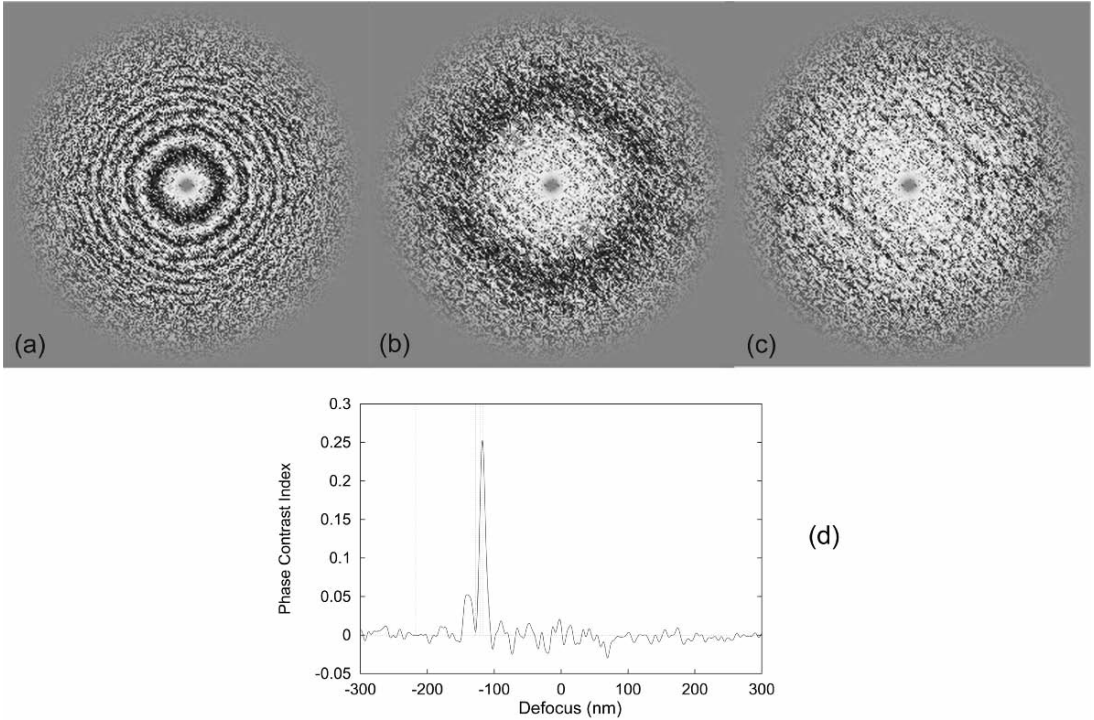


Figure 1-11. Phase contrast index function f_{PCI} . For mismatched values of C_1 (a) -50 nm , (b) -10 nm , f_{PCI} shows dark rings, whereas at the correct values of C_1 , f_{PCI} is close to one (white) at all spatial frequencies (c). (d) The f_{PCI} averaged over $\mathbf{k}$ and plotted as a function of C_1 showing a sharp maximum at the correct focus value of -128 nm . The values ΔC_1 given in (a) to (c) are relative to this focus.

is the symmetric part of the wave aberration function defined earlier with trial parameters C_1 , A_1 and fixed spherical aberration C_3 . This function measures the conjugate symmetry in the corrected image wavefunction and has a value of $+1$ when this is conjugate antisymmetric and -1 when conjugate symmetric. Therefore under the weak phase object approximation when the trial values of C_1 and A_1 are correct the f_{PCI} tends to 1 for all spatial frequencies, whereas for mismatched parameters, f_{PCI} shows dark bands (Figure 1-11a–c). In practice, the correct values of C_1 and A_1 are readily determined from a plot of the f_{PCI} averaged over $\mathbf{k}$ (Figure 1-11(d)).

This general method is best implemented experimentally using a combined tilt-defocus dataset geometry as illustrated in Figure 1-12, which comprises a set of short focal series recorded at a number of tilt azimuths. This dataset is overredundant and can be used to calculate values of C_1 and A_1 for each of the recorded tilt azimuths using the PCF/PCI approach.

For all three measurement methods described, data recorded for a suitable number of defined illumination directions and magnitudes provide the full set of aberration coefficients (Meyer, 2002) and these

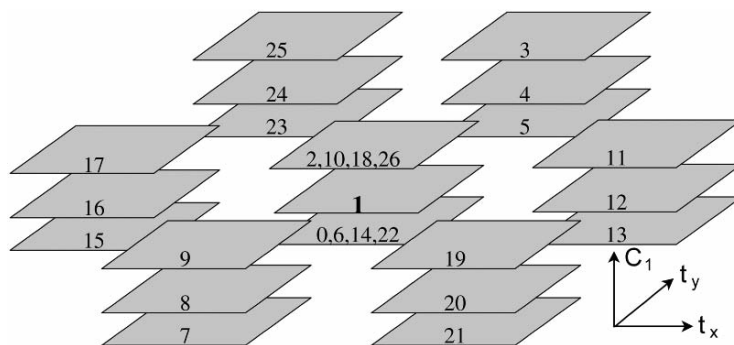


Figure 1-12. Schematic representation of the combined tilt/defocus dataset used for both aberration determination using the PCF/PCI approach and tilt azimuth reconstruction described in a subsequent section. A three member focal series is recorded for each of six different tilt azimuth angles and also for axial illumination. Image numbers refer to the experimental recording sequence designed to minimize hysteresis in the objective lens and beam tilt coils and to enable measurement of focus drift during acquisition.

can be conveniently and reliably determined in practice by least-squares fitting of the parameters to be determined to the observations available. Where the injected tilt is known accurately the dependence of the observables on the aberration coefficients is linear yielding an easily found, unique solution (Saxton, 1995; Meyer et al., 2004). Alternatively, if the calibrations of the beam tilt coil magnitudes and orientations are not known, these unknowns can be fitted in an additional nonlinear iterative minimization of the residual misfit from the linear fitting of the aberration coefficients (Saxton, 1995).

2.4 Coherence

As already described, contrast in a high-resolution image formed in an electron microscope arises from the interference of electron waves that have been scattered by the specimen. The degree to which these waves can interfere is often referred to as the degree of coherence (Born and Wolf, 1999). Hence, the degree of coherence is vital in determining high-resolution image contrast (for a review see Hawkes, 1978).

For all electron sources used in HRTEM imaging (Reimer, 1997), the source is only partially coherent (Hawkes and Kasper, 1996; Hawkes, 1978). There are two aspects to this partial coherence: partial *spatial* coherence, due to the finite dimensions of the electron source (Born and Wolf, 1999), and partial *temporal* coherence, which is associated with the finite energy distribution of the source and fluctuations in both the accelerating voltage and the objective lens current (Hanßen and Trepte, 1971). The effect of this partial coherence of the source is to restrict the amount of information that can be extracted from high-resolution images (Frank, 1973; Wade and Frank, 1977; Hawkes, 1978). Thus, the source and the associated condenser lens system represent the second major optical component that directly affects HRTEM image contrast.

Frequently, the illumination system of the electron microscope is treated as an incoherently filled effective source (Hopkins, 1953, 1951) with an intensity distribution $S(q)$. Within this model the effects of

partial spatial coherence on the HRTEM image intensity can be calculated by incoherent summation of intensities over all the incident angles from the effective source.¹⁶

For this source model the temporal partial coherence effect can also be treated incoherently (Hanßen and Trepte, 1971) and it is generally assumed that the spatial and temporal distributions of the source are not correlated (for a detailed treatment of the case in which this is not valid see Hawkes and Kasper, 1996). Under these conditions the image intensity can be written as

$$I(\mathbf{x}) = \sum_{\mathbf{q}} \sum_f S(\mathbf{q}) F(f) |\mathcal{F}^{-1}\{\psi_d(\mathbf{k}; \mathbf{q}) \exp\{-iW(\mathbf{k}, C_1 + f, \mathbf{q})\}\}|^2. \quad (31)$$

in which the incoherent summations of the coherent image intensities at incident angles $\mathbf{q}$ and focus spread levels, f are weighted by the spatial and temporal intensity distributions $S(\mathbf{q})$, $F(f)$, respectively.

The individual coherent image intensities are thus given as the modulus square of the convolution of the Fourier transform of specimen exit wavefunction $\psi(\mathbf{k}, \mathbf{q})$ at an incident angle $\mathbf{q}$ and the objective lens aberration function $W(\mathbf{k}, C_1 + f, \mathbf{q})$ with corresponding beam tilt and defocus. General computation of the image intensity therefore involves the full dynamical calculation of coherent image intensities over an equally spaced mesh of beam tilts, $\mathbf{q}$ and focal spread values, f . This is optimally tackled through the use of numerical Monte Carlo integration (Chang et al., 2005). A number of approximations may also be made that make this calculation more tractable and that provide analytical solutions.

For small beam tilts of the incident wave and for modest specimen thickness, the specimen exit wavefunction can be approximated as a single exit wavefunction at the mean incident illumination angle. A further approximation can also be made by expanding the aberration function to only first order which is valid if the beam tilt and focal spread are small as typified by many modern instruments. Using these approximations and assuming that the focal spread and beam divergence distributions are Gaussian with e^{-1} values of q_0 and C_1 respectively the image intensity can be written as (Ishizuka, 1980)

$$I(\mathbf{k}) = \sum_{\mathbf{k}'} \psi(\mathbf{k} + \mathbf{k}') \psi^*(\mathbf{k}') T(\mathbf{k} + \mathbf{k}', \mathbf{k}'). \quad (32)$$

with

$$\begin{aligned} T(\mathbf{k}_1, \mathbf{k}_2) = & \exp\{-i[W(\mathbf{k}_1; C_1) - W(\mathbf{k}_2; C_1)]\} \exp\{-\pi^2 q_0^2 [W'_{\mathbf{k}_1} - W'_{\mathbf{k}_2}]^2\} \\ & \times \exp\{-\pi^2 \Delta^2 [\frac{\partial W}{\partial C_1}|_{\mathbf{k}=\mathbf{k}_1} - \frac{\partial W}{\partial C_1}|_{\mathbf{k}=\mathbf{k}_2}]^2\}. \end{aligned} \quad (33)$$

¹⁶ For cases in which this basic assumption does not hold the mathematical description of coherence is best treated in terms of the mutual dynamic object spectrum and associated mutual coherence function. This treatment is outside the scope of this section and an excellent overview can be found in Ernst and Rühle (2003).

The Fourier transform of Eq. (33) is given by the integral over all pairs of diffracted beams with T representing a *transmission cross coefficient* (TCC) (Figure 1–13). Within this expression for the TCC the first exponential term describes the spatial coherence for a Gaussian spatial intensity distribution with an even aberration function. This term is a

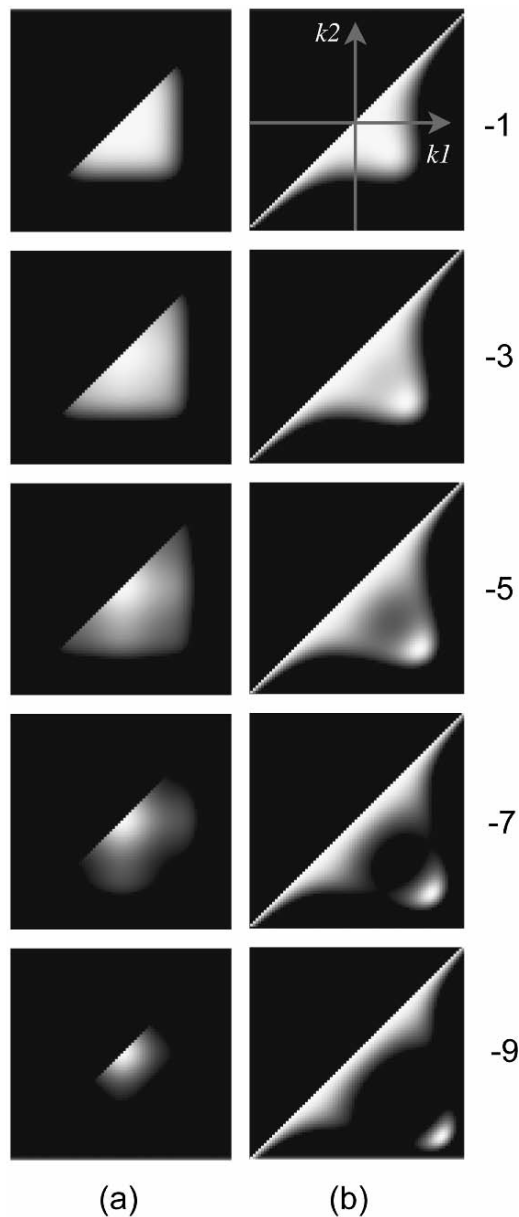


Figure 1–13. Moduli of the transmission cross-coefficient (TCC) functions defined in the text for combined spatial and temporal coherence and calculated under (a) the weak object approximation and (b) for a strong object. In both case the functions are shown for defoci ranging from -1 to -9 Scherzer. All calculations were carried out at 300 kV with $C_3 = 0.6$ nm and with e^{-1} values of Gaussian beam divergence and focal spread distributions of 0.28 mrad and 4.7 nm, respectively (Chang, 2004).

function of the square difference of the gradient of the aberration function at two spatial frequencies and allows maximum transfer when the slopes of the aberration function are identical at k_1 and k_2 . The second exponential term describes the temporal coherence, and is a function of the square difference of the derivative of the aberration function with respect to the focus at two spatial frequencies. This function has maximum transfer when $|k_1| = |k_2|$.

If a further assumption of a weak scattering object is made, such that only interference between the transmitted beam and diffracted beams is considered, the TCC can be further simplified as (Wade and Frank, 1977)

$$T(k) = \exp[-iW(k)] \exp\{-\pi^2 q_0^2 (C_1 \lambda k + C_3 \lambda^3 k^3)^2\} \exp[-\pi^2 \Delta^2 (\lambda k^2 / 2)^2]. \quad (34)$$

In this limit both spatial and temporal coherence functions for weak objects decrease with increasing spatial frequencies (Figure 1-13), and can be recognized as the simple damping envelopes described in an earlier section defining the information limit.

3 Instrumentation

This section describes certain aspects of instrumental design that are important for HRTEM in that they directly contribute to the resolution attainable. Inevitably, due to limitations of space, not all aspects are covered fully and the reader is referred to one of several dedicated texts on HRTEM and electron optics for more detailed treatments of individual areas (Spence, 2002; Hawkes, 1972, 1982; Reimer, 1984, 1995; Hawkes and Kasper, 1989).

3.1 Electron Sources

For HRTEM electron sources have to fulfill a number of key requirements summarized as follows (Spence, 2002; Hawkes and Kasper, 1989):

1. High brightness and coherence.
2. High current efficiency.
3. Long life under available vacuum conditions.
4. Stable emission characteristics.
5. Low energy spread.

An ideal source would therefore provide independent control of the illumination intensity and coherence for any given illuminated area, a situation that is currently best approximated by field emission sources.

For any source, the current, I , passing through an area, A , is proportional to both A and to the solid angle subtended by the illumination aperture at the source (Reimer, 1984). A beam brightness can be defined as the constant of proportionality, β , in the above relationship and as the area and angle tend to zero this is given by

$$\beta = I / \pi A \theta^2 \quad (35)$$

for an aperture semiangle θ .

The effect of an ideal lens (Hawkes, 1982; Reimer, 1984; Hawkes and Kasper, 1989) below the source forming the illumination system is to reduce the current density by M^2 but to increase the angular aperture by $1/M^2$. Thus, in the absence of any lens aberrations, the brightness remains constant at all conjugate planes in the microscope.

To compare different source types a theoretical brightness for a particular source can be expressed in terms of the emission current density at the cathode (filament), ρ , the cathode temperature, T , and the relativistic high voltage, V , as

$$\beta_m = \rho eV / \pi kT \tag{36}$$

although this requires optimal operating conditions and in practice is often reduced to lower values for thermionic sources of between 0.1 and $0.5\beta_m$. Table 1–4 lists values of β_m together with other physical and operational characteristics for the four sources commonly used in HRTEM instruments.

Tungsten hairpin and LaB₆ cathodes rely on thermionic emission of electrons, where at suitably high temperatures electrons in the tail of the Fermi distribution acquire sufficient kinetic energy to overcome the workfunction of the cathode material. Tungsten hairpin sources are the oldest electron sources in commercial use but are only marginally useful for HRTEM and have largely been replaced by LaB₆ cathodes (Ahmed and Broers, 1972) and more recently by field emission sources (Honda et al., 1994; Otten and Coene, 1993). The choice of single crystal LaB₆ cathodes as thermionic emitters for HRTEM is largely due to their higher brightness while still being operable under relatively poor vacuum conditions. They also have the advantage of operating at a lower temperature than W filaments (due to a lower workfunction), thus providing lower energy spreads. For these reasons until the relatively recent availability of commercial field emission sources LaB₆ cathodes were most widely deployed for HRTEM.

Table 1–4. Physical properties of modern electron sources important for HRTEM.^a

Spource	Virtual source diameter	Measured brightness at 100 kV (A cm ⁻² sr ⁻¹)	Energy FWHM (eV)	Melting point (°C)	Vacuum required (torr)	Emission current (μA)
Heated field emission ZrO/W	100 nm	10 ⁷ –10 ⁸	0.8	3370	10 ⁻⁸ 10 ⁻⁹	50–100
Room temperature field emission	2 nm	2 × 10 ⁹	0.3	3370	10 ⁻¹⁰	10
Hair-pin filament	30 μm	5 × 10 ⁵	0.8	3370	10 ⁻⁵	100
LaB ₆	5–10 μm	7 × 10 ⁶ (at 75 kV)	1	2200	10 ⁻⁶	50

^a From Spence (2002).

The detailed physics of field emission and the design of field emission guns for HRTEM are beyond the scope of this section and the article by Crewe (1971) and the chapter in Reimer (1997) provide detailed reviews. These sources are now relatively common in HRTEM instruments and consist of either a heated or, more rarely, an unheated single crystal of tungsten in a $\langle 111 \rangle$ or $\langle 311 \rangle$ orientation. In operation the tip of the source is held in a region of high electrostatic field enabling electrons to tunnel through the lowered potential energy barrier at the surface. In many HRTEM instruments the alternative Schottky emitter comprising a ZrO-coated W tip (Tuggle and Swanson, 1985; Swanson and Schwind in Orloff, 1997), in which the workfunction is lowered to 2.8 eV (from 4.6 eV), is often used allowing electrons to overcome the potential barrier at a lower temperature than for conventional thermionic sources leading to a reduced energy spread. True field emission sources operating at room temperature have been less widely employed due to the need for stringent ultrahigh vacuum (UHV) conditions at the tip region to avoid contamination (Von Harrach, 1995). However, these sources provide the highest brightness and lowest energy spread.

The above physical requirements of the source with respect to brightness and energy spread affect HRTEM images via image recording times and the resolution limit imposed by temporal coherence.

The detail in an HRTEM image is also strongly affected by the spatial coherence of the electrons emitted from the source and certain physical parameters contribute to this. A theoretical treatment of coherence and its influence on HRTEM images has been given in a previous section and we therefore confine discussion here to the relevant physical characteristics of the source itself.

As already described an effective source can be defined as an imaginary electron emitter filling the illuminating aperture (Hopkins, 1951, 1953). In this case each point within the aperture represents a point source of electrons giving an emergent spherical wave, becoming approximately plane at the specimen. Thus, at the specimen plane each electron can be specified by the direction of an incident plane wave. Hence increasing the size of the illuminating aperture increases the size of the central diffraction spot and decreases the spatial coherence. This can also be described as a coherence width, which defines the transverse distance at the object plane over which the illuminating radiation may be treated coherently. Thus waves scattered from atoms separated by less than this distance will interfere coherently and their complex amplitudes must be added, whereas atoms separated by more than this distance scatter incoherently and the intensities of the scattered radiation are added. For sources to fulfill the commonly held assumption that the illuminating source can be replaced by an incoherently filled aperture (see earlier) the coherence width in the plane of this aperture must therefore be small compared to its size, thus making the degree of coherence dependent only on the size of the illuminating aperture. The effect of source size on coherence hence becomes important only for sources smaller than ca. 1 μm , e.g., the smallest LaB₆

or pointed tungsten sources available. Field emission sources do not conform to this model and for FEG HRTEM instruments the degree of coherence is critically dependent on both the source size and the excitation of the condenser and gun lens system.

3.2 Optics

The optical column of a modern HRTEM instrument typically consists of two or three condenser lenses, an objective lens, and up to six imaging lenses below the specimen. The science of magnetic lens design dates back to the late 1920s (Busch, 1926) when it was realized that rotationally symmetric magnetic fields could be used to focus electrons and that to produce a lens of high refractive power the magnetic field along the axis of rotational symmetry needed to be confined to a small region in which the field strength was sufficiently high. This and other elegant early work using algebraic expressions for the lens field has been extensively reviewed elsewhere (Hall, 1966; Hawkes, 1972, 1982; Hawkes and Kasper, 1989; Orloff, 1997) and due to limitations of space is not considered further. Modern lens design is guided by computed solutions of the Laplace equation (see, for example, Septier, 1967, for a review) and subsequent numerical solution of the ray equation (Mulvey and Wallington, 1973). The practical design of electron optics has also been extensively reviewed elsewhere (Hawkes, 1982; Grivet, 1965; Tsuno, 1999; Hawkes and Kasper, 1989) (see also the chapter by Tsuno in Orloff, 1997) including not only the design of the lens polepiece itself but also that of the overall magnetic circuit. For these reasons we provide here only a brief non-mathematically rigorous summary of selected aspects of electron optics relevant to HRTEM and further restrict our discussion to the properties of the objective lens only.¹⁷

For all lenses the aberrations important for HRTEM increase with angle and it is largely the optical properties of the objective lens that dominate the final quality of the image formed. These can be treated analytically through either the Eikonal (Sturrock, 1955) or the Trajectory methods (for collected references to these see Hawkes, 1982, Chapter 1). Alternatively, a numerical approach has been described by (Munro, 1973; Orloff, 1997) based on finite element calculation (FEM) of the magnetic flux density distribution for a trial model lens geometry (see also Lencova in Orloff, 1997, for details). Subsequently this can be used to calculate the axial magnetic field distributions and the paraxial rays, which can be finally combined to calculate the aberration integrals (for further details see Tsuno, 1999, and Orloff, 1997).

The most important lens aberrations for HRTEM are spherical and chromatic aberration and the effects of these are illustrated in Figure 1–14.

¹⁷ Lenses which use the image formed by a preceding lens as an object are classed as projector lenses and include “intermediate” lenses in modern instruments. These are thus distinguished from the objective lens which uses the physical specimen as its object.

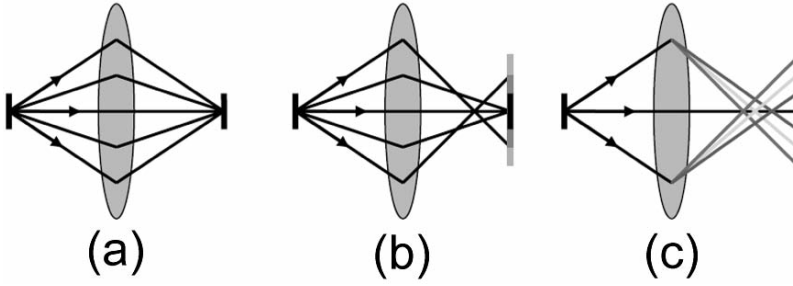


Figure 1-14. Illustration of certain lens aberrations. (a) A perfect lens focuses a point source to a single image point. (b) Spherical aberration causes rays at higher angles to be overfocused. (c) Chromatic aberration causes rays at different energies (indicated by color) to be focused differently. (See color plate.)

Spherical aberration occurs when rays leaving an object at large angle, θ_0 , are refracted too strongly in the outer regions of the lens and are brought to a focus before the Gaussian image plane. If all rays from such an object are considered then the radius of the circle of least confusion due to the spherical aberration is $r = MC_3\theta_0^3$ in the Gaussian image plane, with M the magnification. The value of C_3 depends on the lens geometry and excitation and on the object position within the lens field and typically takes values between 0.5 and 2 mm for modern HRTEM instruments and sets the limit to the interpretable resolution (see earlier) in the absence of corrector elements.

For a known lens geometry with a defined magnetic field distribution values for the spherical aberration can be obtained from either computed solutions of the full (nonparaxial) ray equation (Liebmann, 1949) or through the use of the expression below due to Glaser (1956):

$$C_3 = \frac{e}{16m_0V_r} \int_{z_1}^{z_2} \left[\left\{ \frac{dB_z(z)}{dz} \right\}^2 + \frac{3e}{8m_0V_r} B_z^4(z) - B_z^2(z) \left\{ \frac{h'(z)}{h(z)} \right\}^2 \right] h^4(z) dz \quad (37)$$

where $h(z)$ is the paraxial trajectory of a ray that leaves an axial object with unit slope and $B_z(z)$ is the magnetic field.¹⁸ The presence of derivatives in Eq. (37) explains why the value of C_3 is sensitive to the shape of the field. For this reason the detailed geometry of the polepiece (Munro, 1973) and the location of the specimen within the polepiece gap are key design parameters of the objective lens (Tsuno, 1999).

Chromatic aberration arises from variations in the lens focal length with wavelength and hence electron energy. As all electron sources are polychromatic to a greater or lesser extent (see earlier) a series of in-focus images are formed on a set of planes normal to the optic axis, one for each wavelength present in the illumination. A geometrical optics treatment, as a first approximation, can be used to determine the effects of this focal length variation, which leads to an extended disk

¹⁸ The spherical aberration can be measured using either diffractograms of image shifts as detailed earlier.

in the Gaussian image plane (of a point object) with radius, $r = M\theta_0 C_c[(\Delta V_0/V_0) - (2\Delta I/I)]$. Typical values for C_c in modern instruments are similar to those for C_3 and contribute to setting the absolute information limit for HRTEM (see earlier).

In a similar fashion to the spherical aberration the value of the chromatic aberration can also be evaluated if the field distribution is known (Glaser, 1952) as

$$C_c = \frac{e}{8m_0 V_r} \int_{z_1}^{z_2} B_z^2(z) h(z)^2 dz \quad (38)$$

For applications to HRTEM both of these aberrations can be reduced by introducing the specimen into the field of an immersion objective lens for which the magnetic field on the illumination side of the electron optical column (the prefield) does not directly affect image formation. This symmetrical (condenser objective) geometry also has the significant practical advantage of enabling switching from HRTEM using a relatively broad parallel beam to probe-forming modes for analysis or convergent beam electron diffraction from the same area of the sample and has largely replaced earlier asymmetric “top entry” lens designs.

The spherical and chromatic aberration are sensitive to the overall geometry of the polepiece, which is often described in terms of the key dimensions defining the upper and lower polepiece faces and bores and the gap between the two polepieces. However, the behavior of both C_3 and C_c (and of focal length) can be readily calculated for a given geometry as a function of the lens excitation (Spence, 2002; Tsuno, 1999; Hawkes, 1982) and specimen position enabling these coefficients to be optimized, subject to constraints imposed by machining tolerances and requirements for specimen movement and tilt.

3.3 Specimen Stages

The most important purely mechanical components of modern HRTEM instruments are the specimen holder and the goniometer into which it is fitted. A number of complex designs for these, retaining maximum flexibility in specimen movement and access to peripheral devices under constraints of extremely high mechanical and thermal stability, have been proposed and are reviewed elsewhere (Valdrè, 1979; Valdrè and Goringe, 1971; Watt, 1986; Turner et al., 2005). Most recently facilities for nanoscale specimen manipulation and electrical measurement have also been included (Wang, 2003).

The original designs for instruments used for high resolution were based on “top entry” goniometers into which a conical cartridge was inserted fitting into the upper tapered bore of the objective lens polepiece with the specimen supported in a cup at the base of the cartridge. However, this arrangement leads to restrictions in objective lens design as the polepiece bore must be grossly asymmetric with an upper bore of sufficient diameter to allow the specimen to pass (Tsuno, 1999). Translate and tilt movements also require complex, high precision mechanical mechanisms as the specimen is located at the base of the

cartridge surrounded by the polepiece. The geometry of top entry stages also precludes any EDX analysis and the provision of specimen heating or cooling is extremely difficult (Watt, 1986). Despite these limitations this design has inherent advantages in that the specimen cartridge is mechanically and acoustically isolated from the external environment and its cylindrical symmetry leads to high thermal stability.

The more common side-entry design in which the specimen is attached to a rod that is inserted into the goniometer allows far greater flexibility but introduces potential mechanical and thermal instability. However, this is outweighed for many applications by the flexibility of this design (for example, in providing specimen heating and cooling) and modern commercial implementations are sufficiently stable for atomic resolution imaging. In the majority of cases commercial side-entry goniometers provide high precision computer control of all five axes of specimen movement with stepper motors or donator drives (Emile et al., 1993). This design aspect has recently been further improved to subnanometer precision in movement using additional piezo controls for translation (Kondo et al., 1994a).

It is also worth noting that several designs for fully bakable UHV HRTEM stages have been proposed in which a side-entry mechanism is employed with a detachable tip that locates within the goniometer allowing the transfer rod to be withdrawn (Kondo et al., 1994b).

3.4 Energy Filters

In many applications of HRTEM, energy filtering is of great importance, particularly in improving quantification and interpretability of images by removal of inelastically scattered electrons (Stobbs and Saxton, 1987; Krivanek et al., 1990; Boothroyd, 1998, 2000; see Reimer, 1995, for reviews). At this point we note that the energy filters outlined here also find widespread application as spectrometers for electron energy loss spectroscopy (EELS) and in energy filtered imaging at medium resolution (Reimer, 1995; Tsuno, 1999; Egerton, 1996). However, this section is restricted to briefly reviewing their use for imaging.

Imaging filters are best differentiated by the nature and geometry of their constituent optical elements and can be classified as purely electric, purely magnetic, and combined electric/magnetic (Uhlemann and Rose, 1996). To separate electrons according to their energy loss dipole fields are required and hence all filters have a curved optic axis with the notable exception of the Wien filter (Tsuno, 1993). For HRTEM purely electric (see, for example, Henkelman and Ottensmeyer, 1979) and combined electric/magnetic filters have the disadvantage of being limited to operation at relatively low accelerating voltages due to difficulties associated with the large electric field strengths required at higher voltage. However, as discussed later, these have found recent alternative application as monochrometers in which the electrons have an energy of only a few electronvolts (Rose, 1990).

Magnetic filters have been constructed as both straight vision (in-column) types (Rose and Plies, 1974; Tsuno et al., 1997) and in a

postcolumn geometry, which can be retrofitted to conventional instruments (Krivanek et al., 1992).

The former can be classified further as either Ω geometries in which the deflection angles of the sector magnets cancels to zero and α filters in which the sum of the deflection angles equals 2π (Lanio, 1986). For routine operation and ease of alignment of in-column filters simple filter geometries are required while any residual aberrations should not appreciably affect the quality of the final HRTEM image. This can be achieved through careful choice of suitable symmetries of the magnetic deflection field and of the paraxial rays about the midplane of the system by which second-order axial aberrations and distortions can be canceled making these designs suitable for energy-filtered HRTEM. Based on this principle several commercial filters have been successfully produced (Rose and Plies, 1974; Tsuno et al., 1997). A fully corrected Ω geometry filter has also been proposed and constructed, initially using curved polepiece faces (Rose and Pejas, 1979) but subsequently revised to use straight polepiece faces and additional sextupole elements (Lanio et al., 1986). Finally we note that a corrected in column filter, with an extremely high dispersion achieved through the use of inhomogeneous magnetic fields created by conical magnets has also been described (Uhlemann and Rose, 1994).

The alternative postcolumn filter geometry based on a single 90° sector magnet with curved polepiece faces has also been produced commercially and has found widespread application (Krivanek et al., 1992). For energy-filtered HRTEM this design has the disadvantage of nonzero second-order distortions and chromatic aberration of magnification. These filters therefore operate at a large intermediate magnification so that these do not appreciably affect the image quality. However, due to the large magnifications required the third-order distortions and higher order chromatic aberrations limit the number of clearly resolved object elements. For these reasons a sequence of quadrupole and hexapole elements is incorporated after the sector magnet and alignment to third order is carried out by automated computer control. A significant practical advantage, however, of this geometry is that it can be retrofitted to conventional microscopes.

Finally, we also note that investigations into the contribution of phonon scattering to the Stobbs factor (Boothroyd, 2000) in off-axis electron holograms provide compelling evidence for intrinsic elastic energy filtering (within experimental limitations) in this mode. This issue has been rigorously treated (van Dyck et al., 2000) in a quantum mechanical framework using a global Hamiltonian describing the electron, source and object to prove that although inelastic interference is possible it is too small to be observed. Hence, electron holography provides *de facto* perfect energy filtering, which therefore suggests that electron-optical filters may be redundant in this geometry.

3.5 Detectors

Image recording should not impair the overall resolution of HRTEM data. However, unless the detector used is carefully optimized with respect to microscope operating conditions this component can have a negative effect on overall performance. Historically, photographic

plates using one of several possible specialist emulsions were universally used for image recording (Kuo and Glaeser, 1975; Zeitler, 1992) but these have now been largely superseded by digital detectors of which slow-scan charge-coupled device (CCD) cameras are most commonly used (Spence and Zuo, 1988).

Particular benefits associated with these detectors (Zuo, 2000; Fan and Ellisman, 2000; de Ruijter, 1995; Spence and Zuo, 1988) are that the images are instantly available, quantitatively in digital form, and the camera response is practically linear over a large dynamic range. Their sensitivity is also far higher than that of most photographic emulsions, making single electron detection possible.

Though limited early experiments using CCD chips as direct TEM electron detectors have been performed (Roberts et al., 1982), these were not viable due to the sensitivity of the gate insulator to radiation damage. For low energies (<10 keV), this damage can be avoided using back-thinned CCDs (Ravel and Reinheimer, 1991), but higher energy electrons penetrate through to the sensitive gate oxide on the front side. Moreover, a primary electron of energy E generates $E/3.64$ eV electron-hole pairs in silicon (Fiebigler and Müllner, 1972) leading to saturation of the CCD well capacity after the detection of only a few electrons at typical energies used in HRTEM.

For this reason indirect detection is employed in all cameras currently used for HRTEM by which the electrons impinge on a YAG single crystal¹⁹ or phosphor powder scintillator²⁰ and the generated light is relayed to the CCD chip via a lens- or fiber optical coupling (Figure 1–15). Within this complex coupling, scattering of both the primary incident electrons and the emitted photons in the scintillator material occurs. These processes blur the image, attenuating its high spatial frequencies leading to relatively poor *Modulation Transfer Functions* (MTFs) (Meyer and Kirkland, 1998, 2000; Meyer et al., 2000b).

Electron-sensitive imaging plates have also found application as an alternative digital recording media (Zuo, 2000; Zuo et al., 1996). These consist of a thin embedded layer (ca. 40 μm thick) of a photostimulable phosphor. Luminescence is activated postexposure by a scanning laser in a separate processing system that includes a photomultiplier to convert the output light into a digitized electronic signal. The exposed plate can be subsequently erased by exposure to a suitable light source. These systems provide excellent recording linearity over a wide dynamic range with a sensitivity about three times that of photographic emulsions, but which is voltage dependent. However, their performance also depends strongly on electron dose and drops rapidly at low dose levels. Hence, the most promising application of this technology is in recording quantitative diffraction information (Zuo, 2000).

Very recent experiments by several groups (Faruqi et al., 1994, 2005a,b; Faruqi and Cattermole, 2005; Fan et al., 1998; Milazzo et al.,

¹⁹ YAG: yttrium-aluminum-garnet $\text{Y}_3\text{Al}_5\text{O}_{12}$, a transparent crystal that is made scintillating by doping with impurity atoms, usually europium (YAG:Eu) or cerium (YAG:Ce).

²⁰ A wide range of materials is available for powder scintillators. Popular choices in CCD cameras include P22 ($\text{Y}_2\text{O}_3\text{:S:Eu}$) and P43 ($\text{Gd}_2\text{O}_3\text{:S:Tb}$).

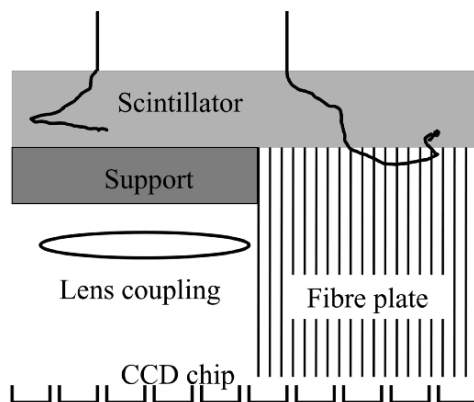


Figure 1–15. Schematic diagram of a CCD detector used for HRTEM. The incoming high-energy electrons are scattered and along their trajectory within the scintillator where they give rise to photon emission. Electrons that penetrate through the scintillator may be back-scattered into the scintillator from the fiber plate or a mechanical support layer. A lens- or fiber-optical coupling system conveys the generated light to the CCD chip where the photons generate electron-hole pairs. The electrons of these pairs are collected in pixelated potential wells. (Taken from Meyer et al., 2002.)

2005) suggest that detectors for HRTEM are on the verge of a revolution in which *direct* injection of beam electrons into solid-state devices may be achieved leading to substantial increases in both resolution and sensitivity. At the time of writing a number of pixellated CMOS-based devices have been fabricated (albeit with initially limited array sizes) including active pixel sensors (Milazzo et al., 2005) in which the readout electronics and amplification are integrated at each pixel. These have demonstrated direct electron imaging with performance characteristics exceeding indirect CCD devices in cryoelectron microscopy and diffraction from radiation-sensitive biological material (Faruqi et al., 1994, 2005a,b; Faruqi and Cattermole, 2005).

3.5.1 Detector Characterization

All digital detectors can be characterized in terms of three key parameters that, in combination, describe their overall performance. These involve measurement of the gain, resolution, and detector quantum efficiency (DQE). The linearity and uniformity of the response may also be important, but in currently available CCD devices are generally extremely well controlled.

Detector resolution is expressed through the MTF, which describes the spatial-frequency-dependent signal transfer, which is affected by both electron and photon scattering within the detection chain. This parameter can be evaluated by integration (over the scintillator area) of the ratio of the output to the input signal in reciprocal space (u, v), normalized to unity at zero spatial frequency (Meyer and Kirkland, 1998).

More formally

$$MTF(u, v) = 1/G \int g_{\mu}(u, v) d\mu \quad (39)$$

where the Fourier transform of the light intensity (number of photons per unit area) collected at a particular position on the detector in its

focal plane is given by $\hat{g}_\mu(u, v)$, $d\mu$ defines the probability of a particular electron hitting the scintillator at this point, and G is the total gain, i.e., the average number of detectable CCD well electrons per primary electron given by

$$G = \int \hat{g}_\mu(0, 0) d\mu \quad (40)$$

The inverse Fourier transform of this MTF is the point spread function (PSF)

$$\text{PSF}(x, y) = 1/G \int \hat{g}_\mu(x, y) d\mu \quad (41)$$

The DQE measures the statistical performance of radiation detectors (Hermann and Krah, 1982) generally defined as the quotient of the squared signal-to-noise ratio (SNR) at the output and input of the detector [in general as a function of spatial frequency (de Ruijter, 1995)].

This can also be expressed in terms of experimentally accessible quantities as

$$\text{DQE}(u, v) = \frac{\hat{p}(0, 0)G^2 [\text{MTF}(u, v)]^2}{V(u, v)} \quad (42)$$

where $\hat{p}(0, 0)$ is the total number of electrons recorded in each exposure, $V(u, v)$ is the variance, (in units of “number of CCD well electrons squared”) and G is the gain defined previously. It should be noted that unless the dose is very low, in which case dose-independent noise sources such as readout noise become important, the observed variance $V(u, v)$ is proportional to the electron dose and therefore the DQE is dose independent.

The above expression can be conveniently rewritten through a simple rescaling of units into the digital numbers (DN) that are read out directly from the camera to give the expression

$$\text{DQE}(u, v) = \frac{I_{\text{DN}}G_{\text{DN}} [\text{MTF}(u, v)]^2}{V_{\text{DN}}(u, v)} \quad (43)$$

where G_{DN} is the gain in digital numbers per incident electron, $I_{\text{DN}} = \hat{p}(0, 0)G_{\text{DN}}$ is the total intensity in digital numbers, and $V_{\text{DN}}(u, v)$ is the variance in (DN)².

Experimentally these performance characteristics can be readily measured from a series of controlled exposures of a deterministic input signal (such as a knife edge) (Meyer et al., 2000b) for a range of experimental conditions (Figure 1–16).

3.6 Aberration Correctors

One of the long standing goals in electron optics has been the correction of the positive spherical aberration that is present in all round lenses (Scherzer, 1949). The essential optical elements used in all correctors are nonround (multipole) lenses as originally proposed by Scherzer (1947).²¹ Subsequently designs for various correctors have

²¹ For a more detailed treatment of spherical aberration correctors see Chapter 10 by Hawkes in this volume.

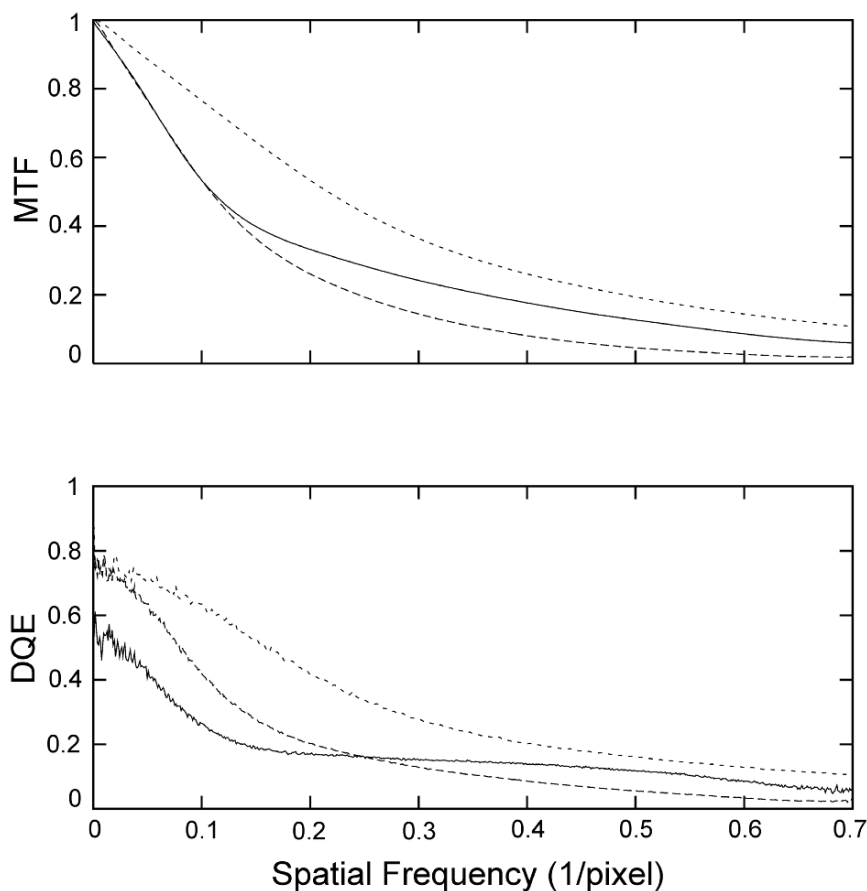


Figure 1-16. (a) MTFs measured for 1024×1024 and 2048×2048 CCD cameras. Solid line, 1024×1024 ; short dashed line, 2048×2048 ; long dashed line, 2048×2048 with 2×2 pixel binning. (b) DQEs corresponding to (a). In all cases plots extend from 0 spatial frequency to the spatial frequency at the corner of Fourier space. (Adapted from Meyer et al., 2002.)

been proposed (for reviews see Rose, 1990; Hawkes and Kasper, 1989). However, it is only relatively recently that these components have been developed to a level at which improvements to the highest performance uncorrected instruments can be made. For the most part this has been due to the extreme mechanical precision and electrical stability required in the corrector elements and the need for sophisticated computer control of the corrector alignment. However, in the 1990s significant improvements in resolution for both HRTEM (Haider et al., 1998a,b) and Scanning Transmission Electron Microscopy (STEM) (Krivanek et al., 1999) were demonstrated, although for the purpose of this section we will not consider the latter further. We also note that very recently proposals for correction of both spherical and chromatic aberration have also been detailed (Rose, 2005) based on a combination of electromagnetic and electrostatic multipole elements.

Correctors for HRTEM are based on a pair of strong electromagnetic hexapole elements together with four additional lenses (Haider et al., 1995). Correction is achieved due to the fact that the primary, non-rotationally symmetric second-order aberrations of the first hexapole (a strong three-fold astigmatism) are exactly compensated by the second hexapole element. Because of their nonlinear diffraction power, the two hexapoles additionally induce a residual secondary, third-order spherical aberration that is rotationally symmetric (Beck, 1979) and proportional to the square of the hexapole strength. This aberration has a negative sign, thus canceling the positive spherical aberration of the objective lens. For HRTEM applications it is essential that the corrector is aplanatic to provide a sufficiently large field of view, which is achieved by matching the coma-free plane of the objective lens to that of the corrector using a round transfer lens doublet. To reduce the azimuthal (anisotropic) component of the off-axial coma the current direction of the first transfer lens doublet is opposite to that in the objective lens. In addition to these primary optical elements commercial correctors (Haider et al., 1998a; Hutchison et al., 2005) contain a number of additional multipole elements for alignment and correction of any residual parasitic aberrations.

The main difficulties associated with the practical operation of these complex electron optical components are the requirements for sophisticated computer control of the various elements and a systematic alignment procedure providing rapid, accurate measurement of the coefficients of the wave aberration function. In practice, this is achieved using measurements of the two-fold astigmatism, A_1 , and defocus, C_1 , from tableaus of diffractograms recorded at known beam tilts (see earlier). These measured values are subsequently used to calculate the aberrations present in the corrector and the appropriate currents to the optical elements are then applied under direct computer control.

3.6.1 Aberration-Corrected Imaging Conditions for HRTEM

The optimization of imaging conditions for corrected instruments and the interpretation of the images obtained have also required renewed analysis. It is possible to identify three conditions for HRTEM imaging when spherical aberration is a variable parameter (Lentzen, 2004).

If the spherical aberration of the objective lens is exactly corrected and the defocus is also set to zero, then the phase contrast transfer function equals 0 for all spatial frequencies up to the information limit set by partial coherence, while the amplitude contrast transfer function equals 1.²² Under these conditions HRTEM can be carried out under pure amplitude contrast conditions in a mode that is not available in a standard uncorrected TEM (Lentzen, 2004).

Alternatively, phase contrast HRTEM images can be obtained using several conditions in which the spherical aberration and defocus are balanced against either C_5 or the chromatic aberration coefficient, C_c (Chang et al., 2003; Chang et al., 2006).

²² The above treatment ignores the effects of higher order aberrations such as C_5 , which also contribute to high resolution phase contrast.

For the first of these conditions, in the presence of C_5 , restricting the phase shifts due to the wave aberration function to lie between 0 and $\pi/2$ gives optimum values for C_1 and C_3 as

$$\begin{aligned} C_1 &= 1.56(C_5\lambda^2)^{1/3} \\ C_3 &= -2.88(C_5^2\lambda)^{1/3} \end{aligned} \quad (44)$$

with a corresponding point resolution, limited by C_5 , given by

$$d = (C_5\lambda^5)^{1/6}/1.47 \quad (45)$$

Under circumstances in which $C_5 = 0$ the choice of C_1 reduces to the conventional Scherzer defocus ($C_1 = -(C_3\lambda)^{1/2}$) (for finite positive C_3) with a corresponding point resolution (as defined previously) of $d = \frac{(C_3\lambda^3)^{1/4}}{\sqrt{2}}$.

Matching the first zero of the phase contrast transfer function to the information limit of the instrument, determined by C_σ and again constraining the phase shifts due to the aberration function to lie between 0 and $\pi/2$, yields alternative optimum values of

$$\begin{aligned} C_1 &= 1.7\pi\Delta \\ C_3 &= -3.4 \frac{(\pi\Delta)^2}{\lambda} \\ C_5 &= 1.3 \frac{(\pi\Delta)^3}{\lambda^2} \end{aligned} \quad (46)$$

where Δ is the focal spread defined in a previous section.

By analogy with the C_5 limited condition when C_5 is zero the above reduce to (O'Keefe, 2000; Chang et al., 2006)

$$\begin{aligned} C_1 &= -\pi\Delta \\ C_3 &= \frac{(\pi\Delta)^2}{\lambda} \end{aligned} \quad (47)$$

For either of these conditions when C_5 is zero or has a negligible effect at the experimentally available resolution, the signs of both C_3 and C_1 can be inverted, and this has been shown to differentially affect the transfer of the linear and nonlinear components of the image intensity in images of complex oxides, providing higher contrast at the weakly scattering anion sites (Lentzen et al., 2002).

3.7 Monochromators

As already discussed, one of the limiting factors in determining the information limit for HRTEM is the energy spread of the source.²³

One possible route to reducing the energy spread is in the use of high brightness sources with intrinsically low energy spreads such as nanometer-sized cold field emitters (Morin and Fink, 1994; Purcell et al., 1995) or more ambitiously, ballistic emission sources (Borgonjen et al., 1997), photocathodes (Baum et al., 1995a,b), and metal ion sources (Purcell and Bihn, 2001). However, none of these has currently been

²³ In addition to the benefits of a reduced energy spread for HRTEM there are also substantial benefits in EELS and related techniques where a lower energy spread improves the spectral energy resolution.

manufactured with sufficient brightness and stability for routine use in HRTEM.

For this reason monochromation of the electron beam produced from conventional Schottky or cold field emission sources using dedicated electron optical components provides an attractive alternative that has now been realized experimentally using several different designs operating at 120 kV and 200 kV (Mook and Kruit, 1999; Rose, 1999; Batson, 1986).

The number of possible monochromator geometries that have been proposed and tested can broadly be divided into those with straight optical axes (typically Wien filters) (Mook and Kruit, 1999; Rose, 1987; Tsuno, 1999) and those with curved axes (typified by Ω geometries) (Rose, 1999; Kahl and Rose, 1998). The former require both electromagnetic and electrostatic elements to satisfy the Wien condition and to achieve the required energy dispersion, whereas the latter can be constructed from purely electromagnetic or electrostatic elements. Further classification of monochromators can be made by considering whether the monochromator and/or the energy selecting slit are operated at high potential or at ground and whether the monochromator is a single dispersion or double dispersion design. Finally, monochromators can be operated in either short field (deflector) mode, in which the monochromator field length is short compared to the distance to the beam focus, or in long field mode, in which the monochromator also acts as a focusing element. Within these various classes there are advantages and disadvantages associated with each configuration that have been extensively reviewed by Mook and Kruit (1999).

For all of the above design possibilities the monochromator must provide sufficient current density for HRTEM at a given energy resolution determined by the energy selecting slit width. Monochromation inevitably reduces brightness as it filters electrons with unwanted energies and thereby reduces the current. However, the optical design of the monochromator itself also affects the brightness through the aberrations introduced by the filter and effects due to Coulomb interactions within the monochromator (Rose, 1999; Tsuno, 1999; Mook and Kruit, 1999).

4 Exit-Wave Reconstruction

Although HRTEM now routinely achieves atomic resolution, the relationship of the image contrast to the underlying object structure for a single image is complex due to the nature of the scattering and imaging processes, as already outlined. A number of alternative approaches to inversion of the imaging process to recover the specimen exit-plane wavefunction, including holography, first proposed by Gabor (1948) and more recently realized at atomic resolution (Lichte, 1991; Orchowski et al., 1995; Lehmann et al., 1999; Tonamura, 1987, 1999). The particular approach described in this section combines the complementary information available from images taken under different conditions in indirect exit-wave reconstruction, which is directly related to the more general problem of coherent detection (Spence, 2002).

As already discussed, the wave aberration function of the objective lens defines a passband of well transmitted spatial frequencies up to the instrument point resolution (at the Scherzer defocus) and beyond this limit the transfer function oscillates. At higher values of defocus this passband extends to higher frequencies, which suggests, in principle, that the addition of several images recorded at different defoci should enable the deconvolution of the effects of the objective lens transfer function and provide a wider passband extending to the axial information limit. This forms the basis of focal series restoration (reconstruction). A simpler alternative using only a single image has also been proposed (O’Keefe et al., 2001) but suffers from difficulties associated with the positions of the zero crossings in the transfer function that prevent deconvolution and from amplification of noise.

4.1 Theory

The recorded image plane intensity can be written as²⁴

$$I(\mathbf{x}) = |\psi_i(\mathbf{x})|^2 = 1 + \psi_i(\mathbf{x}) + \psi_i^*(\mathbf{x}) + |\psi_i(\mathbf{x})|^2 \quad (48)$$

For conditions in which the quadratic term is small then the Fourier transform of the image contrast (the fractional intensity deviation) is given as

$$c(\mathbf{k}) = \psi_i(\mathbf{k}) + \psi_i^*(-\mathbf{k}) \quad (49)$$

The Fourier transforms of the object wave (in the back focal plane of the objective lens) and image waves are related by the previously described phase shifting wave aberration function and thus

$$\psi_i(\mathbf{k}) = \psi_a(\mathbf{k})w(\mathbf{k}) \quad (50)$$

where

$$w(\mathbf{k}) = \exp \{-iW(\mathbf{k})\} \quad (51)$$

Hence in terms of the object wave (in the back focal plane of the objective lens) the Fourier transform of the image contrast can be rewritten as

$$c(\mathbf{k}) = \psi_a(\mathbf{k})w(\mathbf{k}) + \psi_a^*(-\mathbf{k})w(-\mathbf{k}) + n(\mathbf{k}) \quad (52)$$

in which the term $n(\mathbf{k})$ represents the observational noise in the image.

The essence of all reconstructions is now to find an estimate $\psi'_a(\mathbf{k})$ of $\psi_a(\mathbf{k})$ given a set of observed image contrast Fourier transforms $c(\mathbf{k})$ and measurements of the individual $w(\mathbf{k})$.

Given data from several differently aberrated images (obtainable from focal or tilt azimuth dataset geometries as described later), an

²⁴ Throughout this section we use $\mathbf{k}$ and $\mathbf{x}$ as Fourier space and real-space variables rather than (u, v) and (x, y) as in Section 2.

optimum solution for $\psi'_d(\mathbf{k})$ can be defined in various ways (Saxton, 1988). In particular, a Wiener filter applied to a series of images (Schiske, 1973) in the presence of noise gives an optimal estimate of the reconstructed wavefunction, expressed in the form of a weighted superposition of the individual transforms as

$$\psi'_d(\mathbf{k}) = \sum_i r_i(\mathbf{k}) c_i(\mathbf{k}) \quad (53)$$

in which the *restoring filters*, $r_i(\mathbf{k})$, depend on the $w_i(\mathbf{k})$ for the set of images as

$$\begin{aligned} r_i(\mathbf{k}) &= \frac{W(-\mathbf{k})w_i^*(\mathbf{k}) - C^*(\mathbf{k})w_i(-\mathbf{k})}{W(-\mathbf{k})W(\mathbf{k}) - |C(\mathbf{k})|^2 + v(\mathbf{k})} \\ W(\mathbf{k}) &= \sum_i |w_i(\mathbf{k})|^2 \quad C(\mathbf{k}) = \sum_i w_i(\mathbf{k})w_i(-\mathbf{k}) \end{aligned} \quad (54)$$

From Eq. (55) above it can be seen that the effect of the restoring filters on a Fourier component transmitted in only a single image is simply to retain it after division by the corresponding transfer function and for Fourier components present in multiple images to average the estimates. For a Fourier component not transferred in any image the value of the filter tends to zero due to the inclusion of the noise-to-object power ratio, $v(\mathbf{k})$.²⁵ In the final step of the overall reconstruction process the exit-plane wavefunction itself is obtained simply by inverse transformation.

For completeness we also provide an alternative formulation of through focal series restoration by Spence (2002), which gives useful insight into the behavior of the various terms and clearly illustrates the general power of coherent detection methods.

The image amplitude under the WPOA for a single image at a defocus, C_{1n} is given as

$$\begin{aligned} \psi_i(\mathbf{x}, C_{1n}) &= 1 - i\sigma\phi_p(-\mathbf{x}) \cdot F\{P(\mathbf{k})\exp[iW(\mathbf{k})]\} \\ &= \phi_p(\mathbf{x}) \cdot F\{P(\mathbf{k})\exp[iW(\mathbf{k}, C_{1n})]\} \end{aligned} \quad (55)$$

with an image intensity

$$\begin{aligned} I(\mathbf{x}, C_{1n}) &= |1 + \psi_i(\mathbf{x}, C_{1n})|^2 \\ &= 1 + \psi_i(\mathbf{x}, C_{1n}) + \psi_i^*(\mathbf{x}, C_{1n}) + h \end{aligned} \quad (56)$$

where h are the higher order, nonlinear terms.

We are now able to extract the wanted second term in Eq. (56) and discriminate against the conjugate and the higher order terms.

To effect this we first multiply the transform of the intensity by the conjugate of the transfer function to deconvolve it and then sum over N images recorded at different defoci.

This gives

²⁵ This is the major advantage of the Wiener filter in preventing noise amplification where $w_i(\mathbf{k})$ is close to zero.

$$\begin{aligned}
S(\mathbf{k}) &= \sum_n I(\mathbf{k}, C_{1n}) \exp[-iW(\mathbf{k}, C_{1n})] \\
&= \delta(\mathbf{k}) + \phi_p(\mathbf{k}) \sum_n 1 + \phi_p^*(-\mathbf{k}) \sum_n \exp[-2iW(\mathbf{k}, C_{1n})] \\
&\quad + \sum_n H_n(C_{1n}) \exp[-iW(\mathbf{k}, C_{1n})]
\end{aligned} \tag{57}$$

where $\phi(\mathbf{k}) = 1 - i\sigma F\{\phi_p(\mathbf{x})\}$ is the image we wish to restore.

Equation (57) now clearly shows the power of coherent detection. The second term sums as N times the wanted image, whereas the unwanted third and fourth terms sum as $N^{1/2}$ times their initial values. The final term behaves as a two-dimensional random walk. Overall the effects of the coefficients of the wave aberration function appear to have been eliminated and the only remaining resolution limiting effect is the function $P(\mathbf{k})$.²⁶

A further independent formulation of a linear restoration scheme similar to that described above, known as the *paraboloid method*, has also been reported (Op de Beeck et al., 1996; Thust et al., 1996a,b). This is based on a three-dimensional Fourier transform of a through focal series of images with the third variable conjugate to the defocus. The exit-wave sought is localized on a parabolic shell in this three-dimensional space and can readily be separated from the conjugate wave and the nonlinear terms. This approach has the attraction of providing intuitive insight into how the information from a series of images is gathered to yield the complex exit-wave, although unlike the Wiener filter it is nonoptimal in its suppression of the conjugate wave and experimental noise (Saxton, 1994).

The approaches for exit-plane wavefunction restoration described above assume linear imaging. An alternative method has also been developed for a more general case, including the nonlinear contributions to the image intensity. In the original implementation of this latter approach (Kirkland et al., 1985; Kirkland, 1984, 1988) the nonlinear image reconstruction is accommodated by matching the intensities calculated from the restored wave to the measured intensities of images in a focal series through minimization of a least-squares functional [the multiple input maximum a posteriori (MIMAP)]. The more recent improved maximum likelihood (MAL) description (Coene et al., 1992, 1996; Thust et al., 1996a) provides a computationally efficient and numerically optimized recursive solution and explicitly includes the coupling between the exit wave and its complex conjugate.

An alternative imaging geometry for use in reconstruction uses a dataset recorded for several different illumination tilts (of the same magnitude but with different azimuths) (Kirkland et al., 1995, 1997) in a super resolution scheme also referred to as aperture synthesis. The basis of this approach relies on the fact that for an incident beam tilt

²⁶ We note that the above describes a general approach of integration against a kernel to provide a stationary phase condition for a wanted signal and as such finds widespread application throughout physics.

of a Bragg angle Θ_b the optic axis bisects the angle $\Theta = 2\Theta_b$ between the incident beam and a first-order Bragg beam. For this geometry the two beams define the diameter of an achromatic circle (for positive C_3) on which all even-order terms on the wave aberration function cancel. However, only one sector of the diffraction pattern contributes to each image and it is therefore necessary to record images at several (minimally four) different tilt azimuths. The method is also restricted to relatively thin samples due to the effects of parallax (Kirkland et al., 1997). The restoring filters required are of a form similar to those used for focal series reconstruction but with the transfer functions modified so as to be appropriate to tilted illumination (Saxton, 1988). The advantage of this geometry is in the improved flat information transfer, albeit in only one direction in a single image, with the disadvantage of poorer low-frequency transfer in the restored wave. In this way spatial frequencies up to twice the conventional Scherzer limit can be recovered from a dataset composed of four tilted illumination images recorded with orthogonal tilt directions (Kirkland et al., 1995).

It is also worth noting that although the reconstructed wavefunction is free of artefacts due to the lens aberrations, there still exists a significant discrepancy between absolute values of the experimental and simulated exit wavefunctions (Meyer et al., 2000a). The cause of this discrepancy (frequently known as the Stobbs factor; Boothroyd, 1998, 2000), which is also observed in comparisons between conventional experimental HRTEM images and simulations, remains an active research area and recent holographic data (Lehmann and Lichte, 2002) suggest that phonon scattering may make a substantial contribution.

4.2 Experimental Geometries

The most widely applied procedure for exit-wave reconstruction is, as already described, to record several (or many) images at different focus levels for which the positions of weak transfer are different. This provides almost continuous transfer in the recovered exit-wave extending to the ultimate limits set by the axial coherence envelopes (Figure 1–17). In practice the dataset for a focal series reconstruction is relatively easily obtained and typically some 20 or more images separated by close focal increments can conveniently be recorded on a CCD detector utilizing external computer control of the microscope. The subsequent processing involves image registration across the series (to account for specimen drift) and the determination of the individual imaging conditions (using one of the methods described earlier).

The basic methodology for recording data for use in tilt series reconstruction is the same as that for focal series but particular attention must be paid to the image registration (due to the distorted form of cross-correlation functions involving tilted illumination images) and to the initial defocus condition. In general, for this geometry optimal transfer is achieved in the tilted illumination mode (with positive C_3) when $C_1 = \theta^2$ with C_1 and θ being the defocus and illumination tilt measured in reduced units of $(\sqrt{C_3\lambda})$ and $(\lambda/C_3)^{1/4}$, respectively (Hawkes, 1980) (Figure 1–17). With the development of aberration-

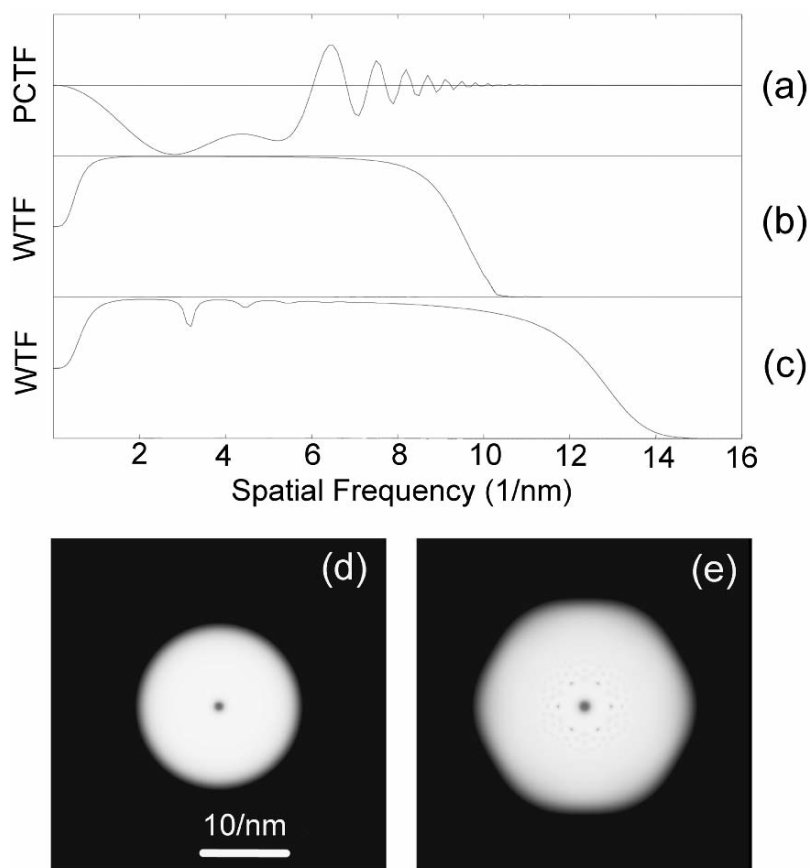


Figure 1-17. A comparison of transfer functions. (a) Conventional phase contrast transfer function (PCTF) for a single Scherzer defocus image. (b) Effective wave transfer function (WTF) for a 21-member focal series with a defocus step between images of 5 nm. (c) Effective wave transfer function for a 6 member tilt azimuth series with a tilt of 7.7 mrad. (d) Two-dimensional representation of the effective wave transfer function in (b). (e) Two-dimensional representation of the effective wave transfer function in (c). In all cases electron optical parameters appropriate to a 300-kV FEGTEM were used (300 kV, $C_s = 0.57$ nm, focal spread = 4 nm, beam divergence = 0.1 mrad). (a) Scaled -1 to 1; (b and c) scaled 0 to 1; (d and e) scaled 0 to 1 (black to white).

corrected instruments, this method is further enhanced by the removal of these restrictions in the absence of C_3 . In practice a combined defocus/tilt geometry provides optimal transfer and has the additional benefit of providing the required data for the determination of the aberration coefficients (see earlier).

4.3 Exit-Wave Restoration of Complex Oxides

To illustrate the benefits of exit-wave restoration over conventional HRTEM imaging we have chosen two examples of complex oxides. Figure 1-18a shows a structural model of $\text{Nb}_{16}\text{W}_{18}\text{O}_{94}$ projected along

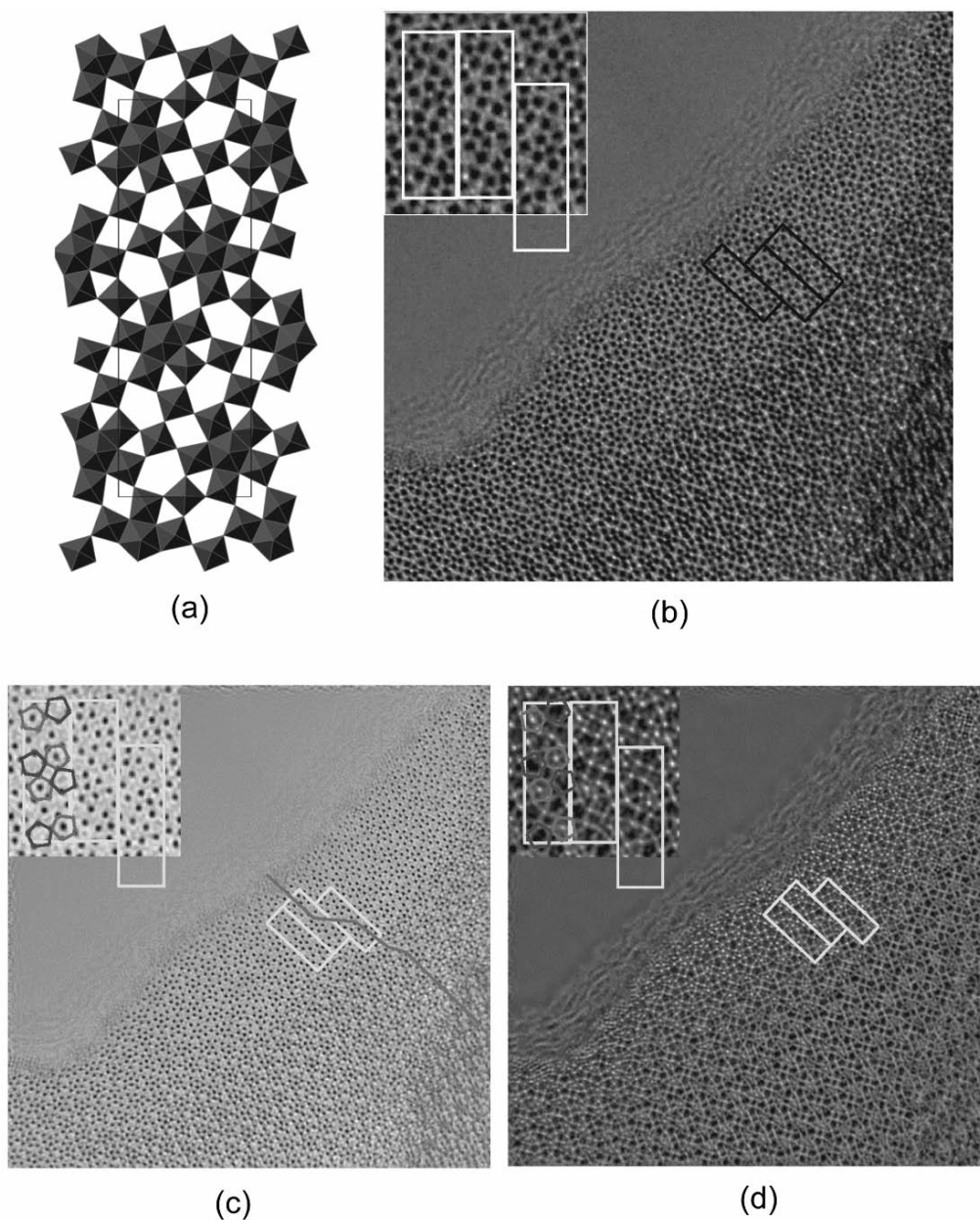


Figure 1-18. (a) Structural model of the complex oxide $\text{Nb}_{16}\text{W}_{18}\text{O}_{94}$ projected along [001]. (b) Conventional axial HRTEM image recorded at the Scherzer defocus of a thin crystal. (c) Reconstructed modulus of the exit-plane wavefunction of $\text{Nb}_{16}\text{W}_{18}\text{O}_{94}$ with the marked area enlarged (inset), which directly shows the cation positions (black) with improved resolution compared to the axial image. The line indicates a stacking fault with a shift of a third of a unit cell along [010]. (d) Reconstructed phase of the exit-plane wavefunction with the marked area enlarged (inset). The cation sites in the phase are recovered with positive (white) contrast and additional weak between the cation atomic columns which indicate the positions of the oxygen anions are also resolved. The reconstructed phase and modulus are shown at the same scale. (See color plate.)

[001] together with an axial image recorded at the Scherzer defocus (Figure 1–18b) and the restored exit-plane wavefunction (Figures 1-18c and d) recovered to a resolution of ca. 0.11 nm. It is apparent that although the basic cation lattice can be determined from the axial image the restored modulus shows the positions of the cation columns in projection at substantially higher resolution. Moreover, the modulus remains directly interpretable to a greater specimen thickness than the axial image. The restored phase shows a more complex contrast changing rapidly with specimen thickness. In addition to the strong positive contrast (white, corresponding to a phase advance) located at the cation sites and corresponding directly to the positions of strong negative (black) contrast in the modulus there is additional weak contrast at positions between the cations at the anion sublattice sites.

Our second example illustrates an experimental tilt azimuth series restoration of an inorganic perovskite with 0.1 nm information transfer.

The basic perovskite structure is cubic with general formula ABO_3 and can be considered as an array of corner-sharing BO_6 octahedra where B is typically a small transition metal cation (Figure 1–19). The cuboctahedral interstices generated by this basic lattice are occupied by larger A cations, typically alkali earth or rare earth metals. Few

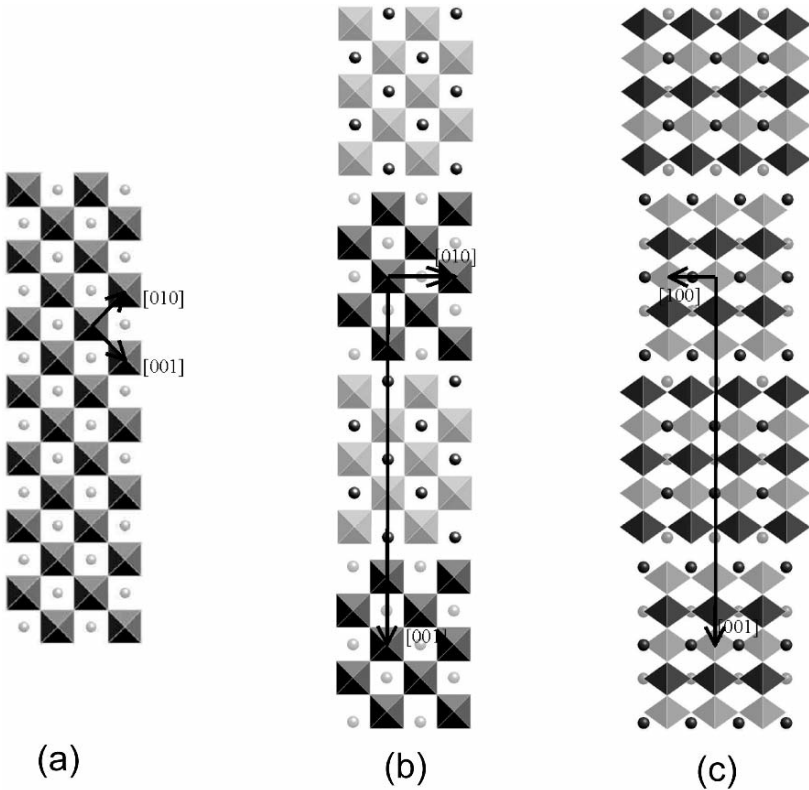


Figure 1–19. Simplified structural models of $A_nB_nO_{3n+2}$ compounds represented by corner-sharing BO_6 octahedra and isolated A cations. Octahedra and cations drawn with light gray shading are half a lattice plane below those with darker shading. (a) The basic perovskite ABO_3 . (b and c) The $n = 5$ structure $A_5B_5O_{17}$ with layers five octahedra wide, projected in the [100] and [010] directions.

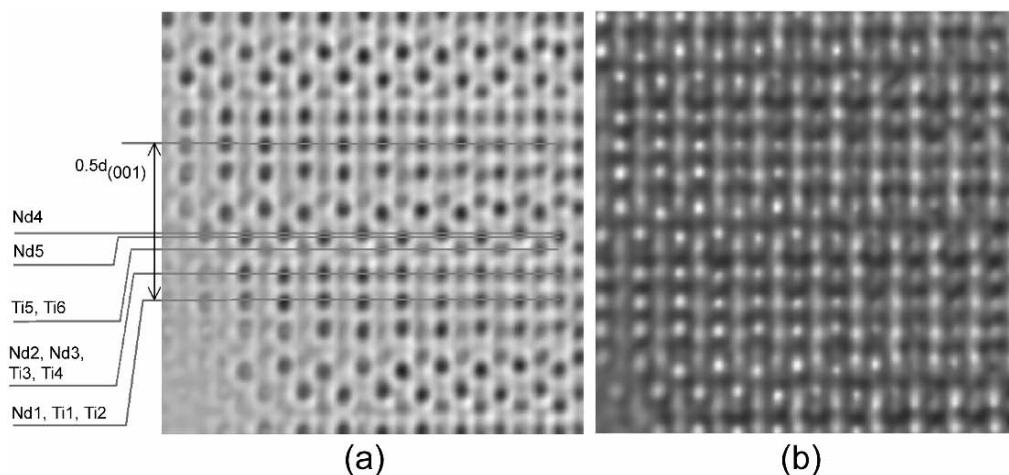


Figure 1-20. (a) Enlarged region taken from the reconstructed modulus calculated from a tilt azimuth dataset of an $\text{Nd}_4\text{SrTi}_5\text{O}_{17}$ crystal edge in the [010] projection showing a small difference in positions of the Nd(4) and Nd(5) cations at the interface between adjacent perovskite slabs. (b) Reconstructed phase of the same region as (a) showing details of the oxygen anion sublattice between the Ti sites. The experimental data were recorded using the tilt series geometry described in the text with a JEOL JEM-3000F FEGTEM, 300 kV, $C_3 = 0.57$ mm with an injected tilt of 1.9 mrad. (See color plate.)

perovskites are cubic and most distort to a lower symmetry to stabilize the structure.

The layered structure described here (of general formula $A_nB_nO_{3n+2}$) is composed of slabs of perovskite sliced along the [110] plane with the two excess oxygen anions accommodated at the interface region between two slabs (Figure 1-19). These structures show similar distortions to bulk perovskites but with additional degrees of freedom at the interface.

The restored exit-plane wavefunction for $\text{Nd}_5\text{Ti}_5\text{O}_{17}$ in the [010] projection is shown in Figure 1-20. The restoration again shows improved resolution in both modulus and phase compared to the conventional axial image and an enhanced sensitivity to the weakly scattering oxygen sublattice in the restored phase. Comparison with the structural model shown in Figure 1-19 shows that the Ti are bridged by O anions and correspondingly in the restored phase, the Nd rows show distinct cations whereas the Ti rows also show weak contrast between the cation sites (Figure 1-20) corresponding to the anion positions. The reconstruction also reveals local distortions present in this material by which the Nd cations at the outside of the perovskite slabs are displaced by small amounts in alternate directions (Figure 1-20).

5 Image Simulation

A key step in the process of obtaining quantitative structural information from HRTEM images is the calculation of image simulations based on defined structural models and imaging conditions for comparison with experimental images. As already discussed, simple models for

electron scattering and imaging at high resolution suffer from limitations and in general, computationally tractable N-beam dynamic calculations are essential to this process.

For this purpose, the multislice method is most commonly used to compute the electron wave at the exit-plane of a specimen with known atomic structure. This approach was first suggested by Cowley and Moodie (1957) and has found extensive use in HRTEM simulation (Goodman and Moodie, 1974). In particular, the availability of efficient Fast Fourier Transform (FFT) algorithms and the general increase in readily available computing power have made earlier constraints on its accuracy due to limitations in the number of slices or the number of diffracted beams immaterial.

An alternative to the multislice method is the Bloch wave approach, first introduced by Bethe (1928) and described in detail, for instance, in Buseck et al. (1989). By analogy with the Bloch theorem in solid-state physics, the solution of the Schrödinger equation in a periodic crystal potential is written as a product of a plane wave and a function that has the same periodicity as the crystal. The latter function is then expanded into its Fourier components from which the Schrödinger equation reduces to a matrix equation in these Fourier coefficients. For simple crystals, relatively accurate solutions can be obtained using only a few of these Bloch waves, with the minimal case requiring only two (the two beam approximation). This approach has the advantage of providing valuable insight into the working of dynamic electron diffraction and explains phenomena such as thickness fringes. However, for complex crystals with large unit cells, the method becomes impractical, as a large number of beams has to be used in the calculation and the computation time for the matrix solution scales with N^3 , where N is the number of beams included. For this reason it has only rarely been applied to calculations of HRTEM images.

Finally we note that for the majority of cases matching of experimental and simulated images is carried out purely visually with limited attempts to define quantitative figures of merit (FOM) describing the differences between experimental and simulated images (or restored exit wavefunctions). In part this is due to the large number of parameters involved in HRTEM simulations (including atomic coordinates and those describing imaging conditions) giving rise to a large multiparameter optimization problem with many local minima. However, a number of possible FOMs have been reported (Mobus et al., 1998; Tang et al., 1993, 1994; Saxton, 1997) based on comparisons in both image and Fourier space.

5.1 The Multislice Formalism

The formal basis of the multislice method (for an extensive treatment see Kirkland, 1998) is the division of the specimen into a number of thin slices perpendicular to the direction of the incident beam. The effects of the specimen potential (transmission, in real space) and of Fresnel diffraction (propagation, in Fourier space) are then treated separately for each slice (Figure 1–21). It is a requirement that the individual slices used in the simulation must be thin enough to be weak

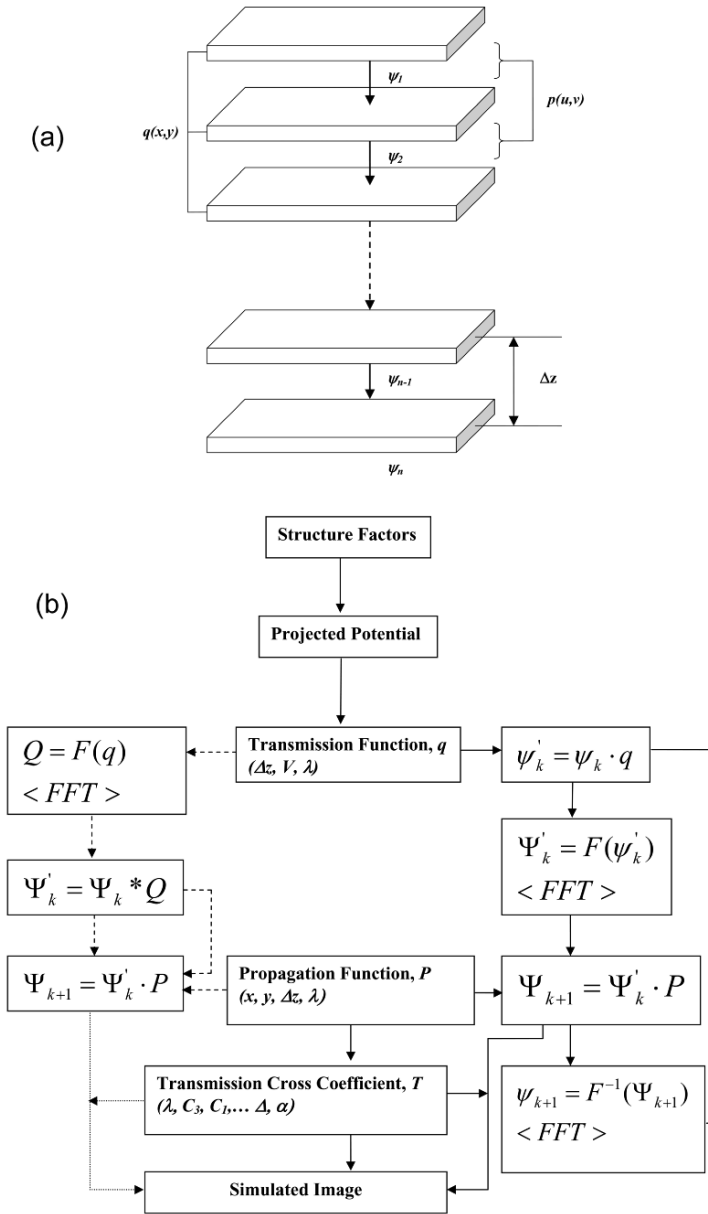


Figure 1–21. (a) Schematic diagram showing the real space transmission and fourier space propagation steps. (b) Flow chart illustrating the steps involved in a multislice simulation.

phase objects and that they obey periodic boundary conditions perpendicular to the incident beam direction.

The equations central to the multislice algorithm can be formally derived starting from the Schrödinger equation for the wavefunction ψ_i of an electron in the electrostatic potential $V(x, y, z)$ of the specimen:

$$\left[-\frac{\hbar^2}{2m} \nabla^2 - eV(x, y, z) \right] \psi_i(x, y, z) = E \psi_i(x, y, z) \quad (58)$$

where m is the relativistically corrected electron mass. For high-energy electrons, the motion is predominantly in the z direction. Hence, it is convenient to separate the full wavefunction, ψ_i , into a product of the solution of the free Schrödinger equation (with $V \equiv 0$), which is a plane wave propagating in the z direction, and a wavefunction ψ that represents the effects due to the specimen and varies much more slowly with z :

$$\psi_i(x, y, z) = e^{2\pi k z} \cdot \psi(x, y, z) \quad (59)$$

where $k = 1/\lambda$ is the inverse wavelength of the free electron. Substituting this into Eq. (58) and using $E = \hbar^2 k^2 / 2m$ yields²⁷

$$-\frac{\hbar^2}{2m} \left[\nabla_{xy}^2 + \frac{\partial^2}{\partial z^2} + 4\pi k \frac{\partial}{\partial z} + \frac{2meV(x, y, z)}{\hbar^2} \right] \psi(x, y, z) = 0 \quad (60)$$

In an elastic scattering process away from the z direction, k_x and k_y are proportional to the scattering angle, while the change, Δk_z , is proportional to its square. For small scattering angles therefore the term $|\partial^2 \psi / \partial z^2| \ll |\nabla_{xy}^2 \psi|$ and can be neglected in the paraxial approximation.

What remains is a first-order differential equation in z

$$\frac{\partial \psi(x, y, z)}{\partial z} = [A + B] \psi(x, y, z) \quad (61)$$

with the operators

$$A = \frac{i\lambda}{4\pi} \nabla_{xy}^2 \quad (62)$$

$$B = i\sigma V(x, y, z) \quad (63)$$

and the interaction parameter σ defined here as

$$\sigma = \frac{2\pi m e \lambda}{\hbar^2} \quad (64)$$

This differential equation has the formal solution

$$\psi(x, y, z + \Delta z) = \exp \left\{ \int_z^{z+\Delta z} [A(z') + B(z')] dz' \right\} \psi(x, y, z) \quad (65)$$

When Δz is small, this reduces to

²⁷ It should be noted that k is defined here as $1/\lambda$ rather than $2\pi/\lambda$ and therefore h rather than $\hbar$ is used in the expression for E .

$$\psi(x, y, z + \Delta z) = \exp\left[\frac{i\lambda}{4\pi}\Delta z \nabla_{xy}^2 + i\sigma V_{\Delta z}(x, y, z)\right]\psi(x, y, z) \quad (66)$$

where $V_{\Delta z}$ is the projected specimen potential between z and $z + \Delta z$:

$$V_{\Delta z}(x, y, z) = \int_z^{z+\Delta z} V(x, y, z') dz' \quad (67)$$

The two operators in the exponent in Eq. (66) do not commute and therefore this function cannot be rewritten as a product of two exponential functions. However, as both operators are small (to the order of Δz), the approximation

$$e^{\epsilon A + \epsilon B} = e^{\epsilon A} e^{\epsilon B} + \frac{\epsilon^2}{2}[B, A] + \mathcal{O}(\epsilon^3) = e^{\epsilon A} e^{\epsilon B} + \mathcal{O}(\epsilon^2) \quad (68)$$

can be used.

This leaves the expression

$$\psi(x, y, z + \Delta z) = \exp\left(\frac{i\lambda\Delta z}{4\pi}\nabla_{xy}^2\right)t(x, y, z)\psi(x, y, z) \quad (69)$$

where $t(x, y, z)$ is the transmission function for the specimen slice between z and $z + \Delta z$:

$$t(x, y, z) = \exp[i\sigma V_{\Delta z}(x, y, z)] \quad (70)$$

The first operator can be further separated into components for x and y coordinates as

$$\exp\left(\frac{i\lambda\Delta z}{4\pi}\nabla_{xy}^2\right) = \exp\left(i\alpha\frac{\partial^2}{\partial x^2}\right)\exp\left(i\alpha\frac{\partial^2}{\partial y^2}\right) \quad (71)$$

where $\alpha = \lambda\Delta z/4\pi$.

Considering the Fourier transform of the operator for the x direction applied to a function $f(x)$:

$$FT\left[\exp\left(i\alpha\frac{\partial^2}{\partial x^2}\right)f(x)\right] = \int e^{-2\pi kx} \exp\left(i\alpha\frac{\partial^2}{\partial x^2}\right)f(x) dx \quad (72)$$

$$= \sum \frac{(i\alpha)^n}{n!} \int e^{-2\pi kx} f^{(2n)}(x) dx \quad (73)$$

$$= \sum \frac{(i\alpha)^n}{n!} (-2\pi k)^{2n} \int e^{-2\pi kx} f(x) dx \quad (74)$$

$$= e^{-4i\pi^2 k^2 \alpha} FT[f(x)] \quad (75)$$

The above can be repeated for the y direction, finally yielding

$$FT\left[\exp\left(\frac{i\lambda\Delta z}{4\pi}\nabla_{xy}^2\right)f(x, y)\right] = \exp[-i\pi\lambda\Delta z(k_x^2 + k_y^2)]FT[f(x, y)] \quad (76)$$

Using this result a single multislice step can be written as

$$\psi(x, y, z + \Delta z) = FT^{-1}\{P(k_x, k_y)FT[t(x, y, z)\psi(x, y, z)]\} \quad (77)$$

where the real-space transmission function, $t(x, y, z)$, and the Fourier space propagation function, $P(k_x, k_y)$, are given by

$$t(x, y, z) = \exp[i\sigma V_{\Delta z}(x, y, z)] \quad (78)$$

$$P(k_x, k_y) = \exp[-i\pi\lambda\Delta z(k_x^2 + k_y^2)] \quad (79)$$

It should be noted that the propagator has the same form as a wave aberration with a defocus of Δz .

The algorithm can be implemented very efficiently using FFTs and is particularly suitable for simulations of crystals, as only one unit cell has to be calculated and the periodic boundary conditions necessary for the Fourier transforms are automatically fulfilled. The density of sampling points in the unit cell determines the maximum spatial frequency in the Fourier transform and therefore the maximum number of diffracted beams included in the simulation. To avoid aliasing artifacts, it is also necessary to exclude all beams above a limit less than the Nyquist frequency in the propagation step (Kirkland, 1998).

Although the multislice method provides an accurate and computationally tractable method for calculating HRTEM images the approximations made in the above derivation can, under certain conditions, limit the accuracy achieved.

The number of beams included in the simulation is ultimately limited by the discrete sampling of the unit cell. Unlike the Bloch wave approach (Bethe, 1928), which gives self-consistent results even with a relatively small number of beams, the multislice method is reliable only when scattering into beams that are not included in the simulation is negligible. For any given simulation, this latter effect can be tested by verifying that the total intensity of the wavefunction does not decrease significantly as the wave propagates through the specimen. A decrease of 5% over the complete specimen is considered acceptable (Kirkland, 1998), but with modern computers it is possible to choose a sufficiently large number of beams to give a loss of intensity smaller than 0.1%.

The paraxial approximation implies the neglect of the term

$$\frac{i\lambda}{4\pi} \frac{\partial^2 \psi}{\partial z^2} \quad (80)$$

A wave scattered by an angle, θ , toward the x direction can be written as

$$\psi = e^{2\pi k[\sin \theta x + (\cos \theta - 1)z]} \quad (81)$$

with $k = 1/\lambda$, hence

$$\left| \frac{\partial^2 \psi}{\partial z^2} \right| < \pi^2 k^2 \theta_{\max}^4 \quad (82)$$

Therefore, the error due to the paraxial approximation accumulated over a specimen thickness, t , can be estimated as

$$\Delta\psi < \frac{\lambda t}{4\pi} \left| \frac{\partial^2 \psi}{\partial z^2} \right|_{\max} < \frac{\pi t}{4\lambda} \theta_{\max}^4 \quad (83)$$

This term is vanishingly small for typical situations in HRTEM (e.g., $E = 300 \text{ keV}$, $\lambda = 2 \text{ pm}$, $t = 20 \text{ nm}$, $\theta_{\max} = 20 \text{ mrad}$ yields $\Delta\psi =$

0.16%) and therefore any error introduced by the paraxial approximation can generally be neglected.

In principle, the error due to the paraxial approximation can be made arbitrarily small by making the slices sufficiently thin. However, as the potential of each atom is usually projected into the slice that contains the atom center, the accuracy is still limited when propagation over the range of the atomic potential has a noticeable effect. In this case, the atomic potential has to be divided across several slices.

A useful feature of the exit-plane wavefunction calculated using the multislice method is that it depends qualitatively on the local potential of the crystal within a small cylinder whose axis is aligned to the incident beam direction with a diameter of typically a few nanometers for HRTEM imaging conditions. This suggests that HRTEM images of defects within a crystalline supercell can be simulated without perfectly smooth periodic continuation of the crystal potential at the supercell boundaries, as any abrupt changes in the potential will influence the exit-plane wavefunction only laterally, within the radius of the above cylinder. Thus, simulated images of large scale defects can be separately computed in a montage of “patches” and subsequently joined (Olsen and Spence, 1981) provided the small allowance described is made at the borders for discontinuities in the potential. Significantly this is not the case for simulations of dynamic electron diffraction patterns from equivalent structures and these cannot be calculated using this “patching” approach.

Finally we observe that the derivation of the multislice algorithm described here is based on the Schrödinger equation, whereas a fully relativistic treatment would require the use of the Dirac equation. However, it has been shown (Fujiwara, 1962; Moodie et al., 2001) that provided relativistically correct expressions are used for both electron mass and wavelength, the expressions derived from the Schrödinger equation are an extremely good approximation of those derived from the Dirac equation.

6 Conclusions and Future Prospects

The current status of HRTEM is such that instrumentation, theory, and computational tools are available for structural determination of a wide range of technologically important materials at atomic resolution. However, there still remain a number of significant experimental and theoretical challenges within this field.

From a theoretical standpoint the majority of HRTEM images are still matched with simulations qualitatively, and even in instances in which quantitative metrics are applied a relatively poorly understood contrast mismatch remains (Boothroyd, 1998). There are also the continuing requirements for refinement and extension of current theoretical frameworks to accommodate the increasing resolution and precision in the data available. Similarly, almost all HRTEM data are recorded for a limited number of projections and therefore access to full three-dimensional information is restricted.

Instrumentation for HRTEM continues to develop and the relatively recent development of monochromators, aberration correctors, and energy filters represent substantial advances. However, there remains considerable scope for improvements in detectors and the correction of chromatic aberration has not been realized experimentally at the time of writing.

Therefore, at present, the “grand challenge” is still the fully quantitative determination of an unknown structure from HRTEM data, quantitatively matched to accurate image simulations and with sufficient resolution to determine the coordinates of all atom species in three dimensions.

References

- Ahmed, H. and Broers, A. (1972). *J. Appl. Phys.* **43**, 2185–2192.
- Allen, C. and Dorignac, D. (1998). In *Proceedings of the 14th ICEM* (H.A.C. Benavidez and M.J. Yacaman, Eds.), Vol. 1 of *Electron Microscopy 1998*, 275–276 (IoP, Cancun).
- Allpress, J., Sanders, J. and Wadsley, A. (1969). *Acta Crystallogr.* **B25**, 1156–1164.
- Amelinckx, S., Milat, O. and van Tendeloo, G. (1993). *Ultramicroscopy* **51**, 90–108.
- Baba, N., Oho, E. and Kanaya, K. (1987). *Scanning Microsc.* **1**, 1507–1514.
- Batson, P. (1986). *Rev. Sci. Instrum.* **57**, 43–48.
- Baum, A., Spicer, W., Pease, R., Costello, K. and Aebi, V. (1995a). *SPIE* **2522**, 208–219.
- Baum, A., Spicer, W., Pease, R., Costello, K. and Aebi, V. (1995b). *SPIE* **2550**, 189–196.
- Beck, V. (1979). *Optik* **53**, 241–255.
- Bethe, H. (1928). *Ann. Phys.* **87**, 55–129.
- Boothroyd, C. (1998). *J. Microsc.* **190**, 99–108.
- Boothroyd, C. (2000). *Ultramicroscopy* **83**, 159–168.
- Borgonjen, E., van Bakel, G., Hagen, C. and Kruit, P. (1997). *Appl. Surf. Sci.* **111**, 165–169.
- Born, M. and Wolf, E. (1999). *Principles of Optics*, 7th ed. (Cambridge University Press, Cambridge).
- Bourret, A., Rouvière, J. and Penisson, J. (1988). *Acta Crystallogr.* **A44**, 838–847.
- Bracewell, R. (1965). *The Fourier Transform and Its Applications* (McGraw-Hill, New York).
- Burge, R. (1973). *J. Microsc.* **98**, 251–285.
- Busch, H. (1926). *Arch. Electrotech.* **18**, 583–594.
- Buseck, P., Cowley, J. and Eyring, L. (1989). *High Resolution Transmission Electron Microscopy and Related Techniques* (Oxford University Press, Oxford).
- Chand, G. (1997). Aberration Determination and Compensation in High Resolution Electron Microscopy. Ph.D. Thesis, University of Cambridge.
- *Chang, L.Y. (2004). PhD thesis, Cambridge University. Reference to Figure 1–13.
- *Chang, L.Y., Chen, F.R., Kai, J.J. and Kirkland, A.I. (2000). *Proc. 12th European Congress on Electron Microscopy* **3**, 89–90.
- *Chang, L.Y., Chen, F.R., Kai, J.J. and Kirkland, A.I. (2003). *J. Elec. Microsc.* **52**, 359–364.
- Chang, L., Meyer, R. and Kirkland, A. (2005). *Ultramicroscopy* **104**, 271–280.
- *Chang, L.Y., Kirkland, A.I. and Titchmarsh, J. (2006). *Ultramicroscopy* **106**, 301–306.
- Chen, J. and van Dyck, D. (1997). *Ultramicroscopy* **70**, 29–44.

- Coene, W. and van Dyck, D. (1984a). *Ultramicroscopy* **15**, 41–50.
- Coene, W. and van Dyck, D. (1984b). *Ultramicroscopy* **15**, 287–300.
- Coene, W., Janssen, G., Op de Beeck, M. and van Dyck, D. (1992). *Phys. Rev. Lett.* **69**, 3743–3746.
- Coene, W., Thust, A., Op de Beeck, M. and van Dyck, D. (1996). *Ultramicroscopy* **64**, 109–135.
- Cosslett, V., Camps, R., Saxton, O., Smith, D., Nixon, W., Ahmed, H., Catto, J., Cleaver, J., Smith, K., Timbs, A., Turner, P. and Ross, P. (1979). *Nature* **281**, 49–51.
- Cowley, J. (1959a). *Acta Crystallogr.* **12**, 367–375.
- Cowley, J. (1959b). *Acta Crystallogr.* **12**, 353–359.
- Cowley, J. (1959c). *Acta Crystallogr.* **12**, 360–367.
- Cowley, J. (1975). *Diffraction Physics* (North Holland, Amsterdam).
- Cowley, J. and Iijima, S. (1972). *Z. Naturforsch* **A27**, 445–451.
- Cowley, J. and Moodie, A. (1957). *Acta Crystallogr.* **10**, 609–619.
- Cowley, J. and Moodie, A. (1962). *J. Phys. Soc. Jpn.* **17** (Suppl B. II), 86–91.
- Cowley, J. and Pogany, A. (1968). *Acta Crystallogr.* **A24**, 109–116.
- Crewe, A. (1971). In *Electron Microscopy in Materials Science* (U. Valdre, Ed.), 162–207 (Academic Press, New York).
- de Ruijter, W. (1995). *Micron* **26**, 247–275.
- Egerton, R. (Ed.) (1996). *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd ed. (Plenum, New York).
- Emile, P., Asselbergs, S., Brock, J., de Jong, M. and Otten, M. (1993). In *Institute of Physics Conference Series, EMAG 95* (A. Craven, Ed.), Vol. 138, 503–506. Institute of Physics, (London).
- Ernst, F. and Rühle, M. (Eds.) (2003). *High Resolution Imaging and Spectrometry of Materials* (Springer, Berlin).
- Fan, G. and Ellisman, M. (2000). *J. Microsc.* **200**, 1–13.
- Fan, G. and Krivanek, O. (1990). In *Electron Microscopy 1990* (L. Peachey and D. Williams, Eds.), Vol. 1, 532–533 (San Francisco Press, Seattle).
- Fan, G., Datte, P., Beuville, E., Beche, J., Millaud, J., Downing, K., Burkard, F., Ellisman, M. and Xuong, N. (1998). *Ultramicroscopy* **70**, 107–113.
- Faruqi, A. and Cattermole, D. (2005). *Nucl. Instrum. Methods* **A549**, 192–198.
- Faruqi, A., Cattermole, D., Henderson, R., Mikulec, B. and Raeburn, C. (1994). *Ultramicroscopy* **94**, 263–276.
- Faruqi, A., Henderson, R., Pryddetch, M., Allport, P. and Evans, A. (2005a). *Nucl. Instrum. Methods* **A546**, 170–175.
- Faruqi, A., Henderson, R. and Tlustos, L. (2005b). *Nucl. Instrum. Methods* **A546**, 160–163.
- Fiebiger, J. and Müller R. (1972). *J. Appl. Phys.* **43**, 3202–3207.
- Fowler, R. and Nordheim, L. (1928). *Proc. R. Soc. London A* **119**, 173–181.
- Frank, J. (1973). *Optik* **38**, 519–536.
- Fujita, H. (1986). *History of Electron Microscopes* (Komiya Printing Co., Tokyo).
- Fujiwara, K. (1962). *J. Phys. Soc. Jpn* **17**(Supplement B2), 118–123.
- Gabor, D. (1948). *Nature* **161**, 777–778.
- Glaser, W. (1952). *Grundlagen der Elektronenoptik* (Springer, Berlin).
- Glaser, W. (Ed.) (1956). In *Handbuch der Physik* (S. Flügge, Ed.). Vol. 33, 123–395 (Springer, Berlin).
- Goodman, P. and Moodie, A.F. (1974). *Acta Crystallogr.* **A30**, 280–290.
- Grivet, P. (Ed.) (1965). *Electron Optics* (Pergamon, London).
- Gronsky, R. and Thomas, G. (1983). In *Proceedings of the 41st Annual Meeting of the Electron Microscopy Society of America* (G. Bailey, Ed.) (San Francisco Press, San Francisco, CA).
- Gutkunst, G., Mayer, J. and Rühle, M. (1994). *Scripta. Metall.* **31**, 1097–1102.

- Haguenau, F., Hawkes, P., Hutchison, J., Satiat-Jeunemaître, B., Simon, G. and Williams, D. (2003). *Microsc. Microanal.* **9**, 96–138.
- Haider, M., Braunshausen, G. and Schwan, E. (1995). *Optik* **99**, 167–179.
- Haider, M., Rose, H., Uhlemann, S., Kabius, B. and Urban, K. (1998a). *Nature* **392**, 768–769.
- Haider, M., Rose, H., Uhlemann, S., Kabius, B. and Urban, K. (1998b). *J. Electron Microsc.* **47**, 395–405.
- Haider, M., Rose, H., Uhlemann, S., Schwan, E., Kabius, B. and Urban, K. (1998c). *Ultramicroscopy* **75**, 53–60.
- Hall, C. (1966). *Introduction to Electron Microscopy* (McGraw-Hill, New York).
- Hanßen, K. (1971). *Adv. Opt. Electron Microsc.* **4**, 1–84.
- Hanßen, K. and Trepte, L. (1971). *Optik* **33**, 166–181.
- Hawkes, P. (1972). *Electron Optics and Electron Microscopy* (Taylor and Francis, London).
- Hawkes, P. (1978). *Adv. Opt. Electron Microsc.* **7**, 101–184.
- Hawkes, P. (1980). *Ultramicroscopy* **5**, 67–70.
- Hawkes, P. (Ed.) (1982). *Magnetic Electron Lenses* (Springer, Berlin).
- Hawkes, P. (1985). *Adv. Electron. Electron Phys.* **16(Supplement)**.
- Hawkes, P. and Kasper, E. (1989). *Principles of Electron Optics: Basic Geometrical Optics* (Academic Press, London).
- Hawkes, P. and Kasper, E. (1996). *Principles of Electron Optics: Wave Optics* (Academic Press, London).
- Henderson, R. (1995). *Q. Rev. Biophys.* **28**, 171–193.
- Henkelman, R. and Ottensmeyer, F. (1979). *J. Microsc.* **102**, 79–95.
- Hermann, K. and Krah, D. (1982). *J. Microsc.* **127**, 17–29.
- Hirabayashi, M., Hiraga, K. and Shindo, D. (1982). *Ultramicroscopy* **9**, 197–202.
- Hirsch, P., Howie, A., Nicholson, R., Pashley, D. and Whelan, M. (1965). *Electron Microscopy of Thin Crystals* (Butterworths, London).
- Honda, T., Tomita, T., Kaneyama, T. and Ishida, Y. (1994). *Ultramicroscopy* **54**, 132–144.
- Hopkins, H. (1951). *Proc. Roy. Soc.* **A208**, 263–277.
- Hopkins, H. (1953). *Proc. Roy. Soc.* **A217**, 408–432.
- Hutchison, J., Honda, T. and Boyes, E. (1986). *JEOL News* **E24**, 9–13.
- Hutchison, J., Titchmarsh, J., Cockayne, D., Doole, R., Hetherington, C., Kirkland, A. and Sawada, H. (2005). *Ultramicroscopy* **103**, 7–15.
- Iijima, S. (1991). *Nature* **354**, 56–58.
- Ishida, Y., Ichinose, H., Mori, M. and Hashimoto, M. (1983). *Trans. Jpn. Inst. Metals* **24**, 349–359.
- Ishizuka, K. (1980). *Ultramicroscopy* **5**, 55–65.
- Jap, B. and Glaeser, R. (1978). *Acta Crystallogr.* **A34**, 94–102.
- Kahl, F. and Rose, H. (1998). In *Electron Microscopy 1998*, Vol. 1, 71–72 (Institute of Physics, Bristol).
- Kirkland, E. (1984). *Ultramicroscopy* **15**, 151–172.
- Kirkland, E. (1988). In *Image and Signal Processing in Electron Microscopy, Proceedings of the 6th Pfefferkorn Conference, Niagara* (P. Hawkes, F. Ottensmeyer, O. Saxton and A. Rosenfeld, Eds.), 139 (Scanning Microscopy International, Chicago).
- Kirkland, E. (1998). *Advanced Computing in Electron Microscopy* (Plenum, New York).
- Kirkland, E., Siegel, B., Uyeda, N. and Fujiyoshi, Y. (1985). *Ultramicroscopy* **17**, 87–103.
- Kirkland, A., Saxton, O., Chau, K-L., Tsuno, K. and Kawasaki, M. (1995). *Ultramicroscopy* **57**, 355–374.
- Kirkland, A., Saxton, O. and Chand, G. (1997). *J. Electron Microsc.* **1**, 11–22.

- Knoll, M. and Ruska, E. (1932a). *Z. Phys.* **78**, 318–339.
- Knoll, M. and Ruska, E. (1932b). *Ann. Phys.* **12**, 607–640.
- Knoll, M. and Ruska, E. (1932c). *Ann. Phys.* **12**, 641–661.
- Koehler, J. (Ed.) (1973). *Biological Electron Microscopy Advanced Techniques I* (Springer, Berlin).
- Koehler, J. (Ed.) (1978). *Biological Electron Microscopy Advanced Techniques II* (Springer, Berlin).
- Koehler, J. (Ed.) (1986). *Biological Electron Microscopy Advanced Techniques III* (Springer, Berlin).
- Komoda, T. (1966). *Jpn. J. Appl. Phys.* **5**, 603–607.
- Kondo, Y., Hosokawa, F., Okhura, M., Hamochi, M., Nakagawa, A., Kirkland, A. and Honda, T. (1994a). In *Electron Microscopy 1994*, (B. Jouffrey and C. Colliex, Eds.), Vol. 1, 275–276 (Les Editions de Physique, Les Ulis, France).
- Kondo, Y., Kobayashi, H., Kasai, T., Nunome, H., Kirkland, A. and Honda, T. (1994b). In *Electron Microscopy 1994* (B. Jouffrey and C. Colliex, Eds.), Vol. 1, 269–270 (Les Editions de Physique, Les Ulis, France).
- Koster, A. (1989). *Ultramicroscopy* **31**, 473–474.
- Koster, A. and de Ruijter, W. (1992). *Ultramicroscopy* **40**, 89–107.
- Koster, A., van den Bos, A. and van der Mast, K. (1987). *Ultramicroscopy* **21**, 209–221.
- Koster, A., de Ruijter, W., van den Bos, A. and van der Mast, K. (1989). *Ultramicroscopy* **27**, 251–272.
- Krakow, W. (1990). *J. Mater. Res.* **5**, 2658–2662.
- Krakow, W., Smith, D. and Hobbs, L. (Eds.) (1984). *Electron Microscopy of Materials* (North Holland, Amsterdam).
- Krivanek, O. (1976). *Optik* **45**, 97–101.
- Krivanek, O. and Leber, M. (1994). In *Proceedings of the 13th ICEM* (B. Jouffrey and C. Coliex, Eds.), Vol. 1 of *Electron Microscopy 1994*, 157–158. (les Editions de Physique, Paris).
- Krivanek, O., Ahn, C. and Wood, G. (1990). *Ultramicroscopy* **33**, 177–185.
- Krivanek, O., Gubbens, A., Dellby, N. and Meyer, C. (1992). *Microsc. Microanal. Microstruct.* **3**, 187–199.
- Krivanek, O., Delby, N. and Lupini, A. (1999). *Ultramicroscopy* **78**, 1–11.
- Kuglin, C. and Hines, D. (1975). *Proc. IEEE Int. Conf. Cybernet. Soc.* 163–165.
- Kuo, I. and Glaeser, R. (1975). *Ultramicroscopy* **1**, 53–66.
- Lanio, S. (1986). *Optik* **73**, 99–107.
- Lanio, S., Rose, H. and Krahll, D. (1986). *Optik* **73**, 56–68.
- Lehmann, M. and Lichte, H. (2002). *Microsc. Microanal.* **8**, 447–466.
- Lehmann, M., Lichte, H., Geiger, D., Lang, G. and Schweda, E. (1999). *Materials Character.* **42**, 249–263.
- Lentzen, M. (2004). *Ultramicroscopy* **99**, 211–220.
- Lentzen, M., Jahnen, B., Jia, C., Thust, A., Tillmann, K. and Urban, K. (2002). *Ultramicroscopy* **92**, 233–242.
- Lichte, H. (1991). In *Advances in Optical and Electron Microscopy*, Vol. 12, 25 (Academic Press, London).
- Liebmann, G. (1949). *Proc. Phys. Soc.* **B62**, 753–772.
- Lundberg, M., Hutchison, J. and Smith, D. (1989). *J. Solid State Chem.* **80**, 178.
- Lynch, D., Moodie, A. and O’Keefe, M. (1975). *Acta Crystallogr.* **A31**, 300–307.
- Matsui, Y., Horiuchi, S., Bando, Y., Kitami, Y., Yokoyama, M. and Suehara, S. (1991). *Ultramicroscopy* **39**, 8–20.
- Menter, J. (1956). *Proc. Roy. Soc.* **A236**, 119–135.
- Meyer, R. (2002). Quantitative Automated Object Wave Restoration in High Resolution Electron Microscopy, Ph.D. Thesis. Dresden Technical University.
- Meyer, R. and Kirkland, A. (1998). *Ultramicroscopy* **75**, 23–33.
- Meyer, R. and Kirkland, A. (2000). *Microsc. Res. Technol.* **49**, 269–280.
- Meyer, R., Sloan, J., Dunin-Borkowski, R., Kirkland, A., Novotny, M., Bailey, S., Hutchison, J. and Green, M. (2000a). *Science* **289**, 1324–1326.

- Meyer, R., Kirkland, A., Dunin-Borkowski, R. and Hutchison, J. (2000b). *Ultramicroscopy* **85**, 9–13.
- Meyer, R., Kirkland, A. and Saxton, O. (2002). *Ultramicroscopy* **92**, 89–109.
- Meyer, R., Kirkland, A. and Saxton, O. (2004). *Ultramicroscopy* **99**, 115–123.
- Milazzo, A., Leblanc, P., Duttweiler, F., Jin, L., Bouwer, J., Peltier, S., Ellisman, M., Bieser, F., Matis, H., Wieman, H., Denes, P., Kleinfelder, S. and Xuong, N. (2005). *Ultramicroscopy* **104**, 152–159.
- Mobus, G., Schweinfest, R., Gemming, T., Wagner, T. and Rühle, M. (1998). *J. Microsc.* **190**, 109–130.
- Moodie, A., Cowley, J. and Goodman, P. (2001). In *International Tables of Crystallography*, Vol. B: *Reciprocal Space*, 2nd ed. (U. Shmueli, Ed.), (Kluwer Academic Publishers, Dordrecht).
- Mook, H. and Kruit, P. (1999). *Ultramicroscopy* **78**, 43–51.
- Morin, R. and Fink, H. (1994). *Appl. Phys. Lett.* **65**, 2362–2364.
- Mulvey, T. and Wallington, M. (1973). *Rep. Prog. Phys.* **36**, 347–421.
- Munro, E. (1973). In *Image Processing and Computer Aided Design in Electron Optics* (P. Hawkes, Ed.), 284 (Academic Press, London).
- O’Keefe, M. (1992). *Ultramicroscopy* **47**, 282–297.
- O’Keefe, M. (2000). In *Proceedings Microscopy and Microanalysis 2000* (Cambridge, Philadelphia).
- O’Keefe, M. and Pitt, A. (1980). In *Proceedings of the Seventh European Conference on Electron Microscopy* (P. Brederoo and G. Boom, Eds.), 122–123 (The Hague).
- *O’Keefe, M.A. and Sanders, J.V. (1973). *Acta Crystallogr.* **A29**, 381. Reference to PCD approximation.
- O’Keefe, M., Hetherington, C., Wang, Y., Nelson, E., Turner, J., Kisielowski, C., Malm, O., Mueller, R., Ringnalda, J., Pan, M. and Thust, A. (2001). *Ultramicroscopy* **89**, 215–241.
- Olsen, A. and Spence, J. (1981). *Phil. Mag.* **A43**, 945–965.
- Op de Beeck, M., van Dyck, D. and Coene, W. (1996). *Ultramicroscopy* **64**, 167–183.
- Orchowski, A., Rau, W. and Lichte, H. (1995). *Phys. Rev. Lett.* **74**, 399–402.
- Orloff, J. (Ed.) (1997). *Handbook of Charged Particle Optics*. (CRC Press, New York).
- Otten, M. and Coene, W. (1993). *Ultramicroscopy* **48**, 77–91.
- Pan, M. (1998). In *Proceedings of the 14th ICEM* (H.A.C. Benavidez and M.J. Yacamán, Eds.), Vol. 1 of *Electron Microscopy 1998*, 263–264 (IoP, Cancun).
- Penisson, J., Nowicki, T. and Biscondi, M. (1988). *Phil. Mag.* **A58**, 947–956.
- Phillip, F., Hoschen, R., Osaki, M., Möbus, G. and Rühle, M. (1994). *Ultramicroscopy* **56**, 1–10.
- Purcell, S. and Bihn, V. (2001). *J. Vac. Sci. Technol.* **B19**, 79–86.
- Purcell, S., Binh, V. and Garcia, N. (1995). *Appl. Phys. Lett.* **67**, 436–438.
- Ravel, M. and Reinheimer, A.L. (1991). *SPIE* **1447**, 109–122.
- Reimer, L. (1984). *Transmission Electron Microscopy* (Springer, Berlin).
- Reimer, L. (Ed.) (1995). *Energy Filtering Transmission Electron Microscopy* (Springer, Berlin).
- Reimer, L. (1997). *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*, 4th ed. (Springer, Berlin).
- Roberts, P., Chapman, J. and Macleod, A. (1982). *Ultramicroscopy* **8**, 385–396.
- Rose, H. (1981). *Nucl. Instrum. Methods* **187**, 187–199.
- Rose, H. (1987). *Optik* **77**, 26–34.
- Rose, H. (1990). *Optik* **85**, 19–24.
- Rose, H. (1999). *Ultramicroscopy* **78**, 13–25.
- Rose, H. (2005). *Ultramicroscopy* **103**, 1–6.
- Rose, H. and Pejas, W. (1979). *Optik* **54**, 235–250.
- Rose, H. and Plies, E. (1974). *Optik* **40**, 336–341.

- Ruska, E. (1934). *Z. Phys.* **87**, 580–602.
- Saxton, O. (1988). In *Image and Signal Processing in Electron Microscopy, Proceedings of the 6th Pfefferkorn Conference, Niagara* (P. Hawkes, F. Ottensmeyer, O. Saxton and A. Rosenfeld, Eds.), 213 (Scanning Microscopy International, Chicago).
- Saxton, O. (1994). *J. Microsc.* **174**, 61–68.
- Saxton, O. (1995). *J. Microsc.* **179**, 201–213.
- Saxton, O. (1997). *J. Microsc.* **190**, 52–60.
- Saxton, O. (2000). *Ultramicroscopy* **81**, 41–45.
- Saxton, O., Smith, D.J. and Erasmus, S. (1983). *J. Microsc.* **130**, 187–201.
- Scherzer, O. (1947). *Optik* **2**, 114–132.
- Scherzer, O. (1949). *J. Appl. Phys.* **20**, 20–29.
- Schiske, P. (1973). In *Image Processing and Computer Aided Design in Electron Optics* (P. Hawkes, Ed.), 82 (Academic Press, London).
- Self, P., O’Keefe, M., Buseck, P. and Spargo, A. (1983). *Ultramicroscopy* **11**, 35–52.
- Septier, A. (Ed.) (1967). *Focusing of Charged Particles* (Academic Press, New York).
- Shindo, D. and Hiraga, K. (1998). *High Resolution Electron Microscopy for Materials Science* (Springer, Tokyo).
- Smith, D. (1997). *Rep. Progr. Phys.* **60**, 1513–1580.
- Smith, D. and Lu, P. (1991). *Inst. Phys. Conf. Ser. Microsc. Semiconducting Mater.* **117**, 1–10.
- Smith, D., Glaisher, R. and Lu, P. (1989). *Phil. Mag. Lett.* **59**, 69–75.
- Spence, J. (1999). *Mater. Sci. Eng.* **R26**, 1–49.
- Spence, J. (2002). *High Resolution Electron Microscopy*, 3rd ed. (Oxford University Press, Oxford).
- Spence, J. and Zuo, J. (1988). *Rev. Sci. Instrum.* **59**, 2102–2105.
- Spence, J. and Zuo, J. (1992). *Electron Microdiffraction* (Plenum, New York).
- Stadelmann, P. (1987). *Ultramicroscopy* **21**, 131–146.
- Stadelmann, P. (1991). *Micron Microsc. Acta* **22**, 175–176.
- Stobbs, M. and Saxton, O. (1987). *J. Microsc.* **151**, 171–184.
- Sturrock, P. (Ed.) (1955). *Static and Dynamic Electron Optics* (Cambridge University Press, Cambridge).
- Tang, D., Kirkland, A. and Jefferson, D. (1993). *Ultramicroscopy* **48**, 321–331.
- Tang, D., Kirkland, A. and Jefferson, D. (1994). *Ultramicroscopy* **53**, 137–146.
- Thomas, G. (Ed.) (1962). *Transmission Electron Microscopy of Metals* (Wiley, New York).
- Thust, A., Coene, W., Op de Beeck, M. and van Dyck, D. (1996a). *Ultramicroscopy* **64**, 211–230.
- Thust, A., Overwijk, M., Coene, W. and Lentzen, M. (1996b). *Ultramicroscopy* **64**, 249–264.
- Tonamura, A. (1987). *Rev. Mod. Phys.* **59**, 639–669.
- Tonamura, A. (Ed.) (1999). *Electron Holography*, 2nd ed. (Springer, Heidelberg).
- Tsuno, K. (1993). *Rev. Sci. Instrum.* **64**, 659–666.
- Tsuno, K. (1999). *J. Electron Microsc.* **48**, 801–820.
- Tsuno, K., Kaneyama, T., Honda, T., Tsuda, K., Terauchi, M. and Tanaka, M. (1997). *J. Electron Microsc.* **46**, 357–368.
- Tuggle, D. and Swanson, L. (1985). *J. Vac. Sci. Technol.* **B3(1)**, 220–223.
- Turner, J., Valdre, U. and Fukami, A. (2005). *J. Electron Microsc. Technol.* **11**, 258–271.
- Typke, D. and Dierksen, K. (1995). *Optik* **99**, 155–166.
- Uhlemann, S. and Haider, M. (1998). *Ultramicroscopy* **72**, 109–119.

- Uhlemann, S. and Rose, H. (1994). *Optik* **96**, 163–178.
- Uhlemann, S. and Rose, H. (1996). *Ultramicroscopy* **63**, 161–167.
- Unwin, N. and Henderson, R. (1975). *J. Mol. Biol.* **94**, 425–432.
- Urban, K., Kabius, B., Haider, M. and Rose, H. (1999). *J. Electron Microsc.* **48**, 821–826.
- Uyeda, N., Kobayashi, E., Suito, Y., Harada, M. and Watanabe, J. (1972). *J. Appl. Phys.* **43**, 5181–5189.
- Valdrè, U. (1979). *J. Microsc.* **117**, 55–77.
- Valdrè, U. and Goringe, M. (1971). In *Electron Microscopy in Materials Science* (U. Valdre, Ed.), 208 (Academic Press, New York).
- van Dyck, D. (1983). *J. Microsc.* **132**, 31–42.
- van Dyck, D. and Coene, W. (1984). *Ultramicroscopy* **15**, 29–40.
- van Dyck, D. and Op de Beeck, M. (1994). *Ultramicroscopy* **55**, 435–457.
- van Dyck, D., Lichte, H. and Spence, J. (2000). *Ultramicroscopy* **81**, 187–194.
- van Tendeloo, G. (1998). *J. Mater. Chem.* **8**, 797–808.
- Von Harrach, H. (1995). *Ultramicroscopy* **58**, 1–5.
- Wade, R. and Frank, J. (1977). *Optik* **49**, 81–92.
- Wang, Z. (2003). *Adv. Mater.* **15**, 1497–1514.
- Watt, I. (1986). *J. Phys. E. Sci. Instrum.* **19**, 668–678.
- Yao, N. and Wang, Z. (Eds.) (2005). *Handbook of Microscopy for Nanotechnology* (Springer, Berlin).
- Yoshioka, H. (1957). *J. Phys. Soc. Jpn.* **12**, 618–628.
- Zeitler, E. (1992). *Ultramicroscopy* **46**, 405–416.
- Zemlin, F. (1979). *Ultramicroscopy* **4**, 241–245.
- Zemlin, F., Weiss, K., Schiske, P., Kunath, W. and Herrmann, K. (1978). *Ultramicroscopy* **3**, 49–60.
- Zuo, J. (2000). *Microsc. Res. Technol.* **49**, 245–268.
- Zuo, J., McCartney, M. and Spence, J. (1996). *Ultramicroscopy* **66**, 35–47.

Scanning Transmission Electron Microscopy

Peter D. Nellist

1. Introduction

The scanning transmission electron microscope (STEM) is a very powerful and highly versatile instrument capable of atomic resolution imaging and nanoscale analysis. The purpose of this chapter is to describe what STEM is, to highlight some of the types of experiments that can be performed using a STEM, to explain the principles behind the common modes of operation, to illustrate the features of typical STEM instrumentation, and to discuss some of the limiting factors in its performance.

1.1 The Principle of Operation of a STEM

Figure 2-1 shows a schematic of the essential elements of a STEM. Most dedicated STEM instruments have their electron gun at the bottom of the column with the electrons traveling upward, which is how Figure 2-1 has been drawn. Figure 2-2 shows a photograph of a dedicated STEM instrument.

More commonly available at the time of writing are combined conventional transmission electron microscope (CTEM)/STEM instruments. These can be operated in both the CTEM mode, where the imaging and magnification optics are placed after the sample to provide a highly magnified image of the exit wave from the sample, or the STEM mode as described in Section 8. Combined CTEM/STEM instruments are derived from conventional transmission electron microscopy (TEM) columns and have their gun at the top of the column. The pertinent optical elements are identical, and for a TEM/STEM Figure 2-1 should be regarded as being inverted.

In many ways, the STEM is similar to the more widely known scanning electron microscope (SEM). An electron gun generates a beam of electrons that is focused by a series of lenses to form an image of the electron source at a specimen. The electron spot, or *probe*, can be scanned over the sample in a raster pattern by exciting scanning deflection coils, and scattered electrons are detected and their intensity

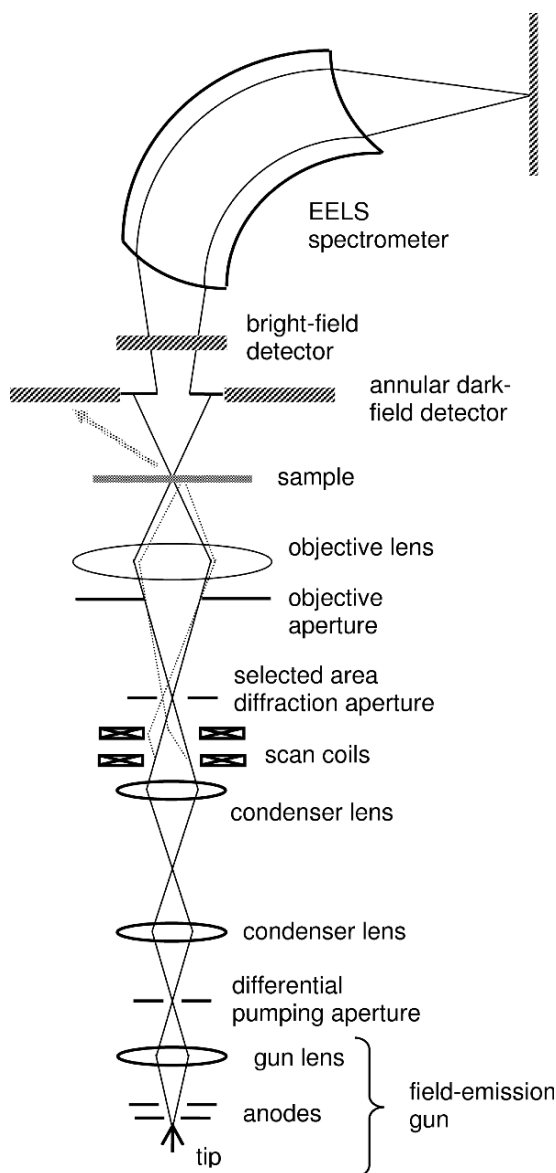


Figure 2-1. A schematic of the essential elements of a dedicated STEM instrument showing the most common detectors.

plotted as a function of probe position to form an image. In contrast to an SEM, where a bulk sample is typically used, the STEM requires a thinned, electron transparent specimen. The most commonly used STEM detectors are therefore placed after the sample, and detect transmitted electrons.

Since a thin sample is used (typically less than 50nm thick), the probe spreading within the sample is relatively small, and the spatial resolution of the STEM is predominantly controlled by the size of the probe. The crucial image forming optics are therefore those before the

sample that are forming the probe. Indeed the short-focal-length lens that finally focuses the beam to form the probe is referred to as the *objective* lens. Other *condenser* lenses are usually placed before the objective to control the degree to which the electron source is demagnified to form the probe. The electron lenses used are comparable to those in a conventional TEM, as are the electron accelerating voltages used (typically 100–300 kV). Probe sizes below the interatomic spacings in many materials are often possible, which is the great strength of STEM. Atomic resolution images can be readily formed, and the probe can then be stopped over a region of interest for spectroscopic analysis at or near atomic resolution.

To form a small, intense probe we clearly need a correspondingly small, intense electron source. Indeed, the development of the cold field emission gun by Albert Crewe and co-workers nearly 40 years ago (Crewe et al., 1968a) was a necessary step in their subsequent construction of a complete STEM instrument (Crewe et al., 1968b). The quantity of interest for an electron gun is actually the source *brightness*, which will be discussed in Section 9. Field-emission guns are almost always

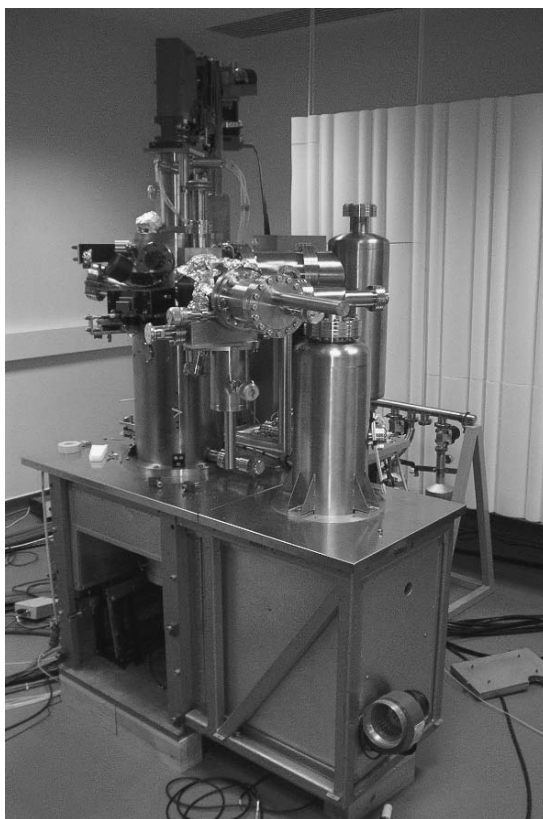


Figure 2–2. A photograph of a dedicated STEM instrument (VG Microscopes HB501). The gun is below the table level, with most of the electron optics above the table. At the top of the column can be seen a magnetic prism spectrometer for electron energy-loss spectroscopy.

used for STEM, either a cold field emission gun (CFEG) or a Schottky thermally assisted field emission gun. In the case of a CFEG, the source size is typically around 5 nm, so the probe-forming optics must be capable of demagnifying its image of the order of 100 times if an atomic sized probe is to be achieved. In a Schottky gun the demagnification must be even greater.

The size of the image of the source is not the only probe size defining factor. Electron lenses suffer from inherent aberrations, in particular spherical and chromatic aberrations. The aberrations of the objective lens generally have greatest effect, and limit the width of the beam that may pass through the objective lens and still contribute to a small probe. Aberrated beams will not be focused at the correct probe position, and will lead to large diffuse illumination thereby destroying the spatial resolution. To prevent the higher angle aberrated beams from illuminating the sample, an objective aperture is used, and is typically a few tens of microns in diameter. The existence of an objective aperture in the column has two major implications: (1) As with any apertured optical system, there will be a diffraction limit to the smallest probe that can be formed, and this diffraction limit may well be larger than the source image. (2) The current in the probe will be limited by the amount of current that can pass through the aperture, and much current will be lost as it is blocked by the aperture.

Because the STEM resembles the more commonly found SEM in many ways, several of the detectors that can be used are common to both instruments, such as the secondary electron (SE) detector and the energy-dispersive X-ray (EDX) spectrometer. The highest spatial resolution in STEM is obtained by using the transmitted electrons, however. Typical imaging detectors used are the bright-field (BF) detector and the annular dark-field (ADF) detector. Both these detectors sum the electron intensity over some region of the far field beyond the sample, and the result is displayed as a function of probe position to generate an image. The BF detector usually collects over a disc of scattering angles centered on the optic axis of the microscope, whereas the ADF detector collects over an annulus at higher angle where only scattered electrons are detected. The ADF imaging mode is important and unique to STEM in that it provides *incoherent* images of materials and has a strong sensitivity to atomic number allowing different elements to show up with different intensities in the image.

Two further detectors are often used with the STEM probe stationary over a particular spot: (1) A Ronchigram camera can detect the intensity as a function of position in the far field, and shows a mixture of real-space and reciprocal-space information. It is mainly used for microscope diagnostics and alignment rather than for investigation of the sample. (2) A spectrometer can be used to disperse the transmitted electrons as a function of energy to form an electron energy-loss (EEL) spectrum. The EEL spectrum carries information about the composition of the material being illuminated by the probe, and even can show changes in local electron structure through, for example, bonding changes.

1.2 Outline of Chapter

The crucial aspect of STEM is the ability to focus a small probe at a thin sample, so we start by describing the form of the STEM probe and how it is computed. To understand how images are formed by the BF and ADF detectors, we need to know the electron intensity distribution in the far field after the probe has been scattered by the sample, which is the intensity that would be observed by a Ronchigram camera. This allows us to go on and consider BF and ADF imaging.

Moving on to the analytical detectors, there is a section on the EEL spectrum that emphasizes some aspects of the spatial localization of the EEL spectrum signal. Other detectors, such as EDX and SE, that are also found on SEM instruments are briefly discussed.

Having described STEM imaging and analysis we return to some instrumental aspects of STEM. We discuss typical column design, and then go on to analyze the requirements for the electron gun in STEM. Consideration of the effect of the finite gun brightness brings us to a discussion of the resolution limiting factors in STEM where we also consider spherical and chromatic aberrations. We finish that section with a discussion of spherical aberration correction in STEM, which is arguably having the greatest contribution in the field of STEM and is producing a revolution in performance.

There have been several review articles previously published on STEM (for example, Cowley, 1976; Crewe, 1980; Brown, 1981). More recently, instrumental improvements have increased the emphasis on *atomic resolution* imaging and analysis. In this chapter we tend to focus on the principles and interpretation of STEM data when it is operating close to the limit of its spatial resolution.

2. The STEM Probe

The crucial aspect of STEM performance is the ability to focus a sub-nanometer-sized probe at the sample, so we start by examining the form of that probe. We will initially assume that the electron source is infinitesimal, and that the beam is perfectly monochromatic. The effects of these assumptions not holding are explored in more detail in Section 10.

The probe is formed by a strong imaging lens, known as the *objective* lens, that focuses the electron beam down to form the crossover that is the probe. Typical electron wavelengths in the STEM range from 3.7 pm (for 100-keV electrons) to 1.9 pm (for 300-keV electrons), so we might expect the probe size to be close to these values. Unfortunately, all circularly symmetric electron lenses suffer from inherent spherical aberration, as first shown by Scherzer (1936), and for most TEMs this has typically limited the resolution to about 100 times worse than the wavelength limit.

The effect of spherical aberration from a geometric optics standpoint is shown in Figure 2-3. Spherical aberration causes an overfocusing of the higher angle rays of the convergent beam so that they are brought to a premature focus. The Gaussian focus plane is defined as the plane

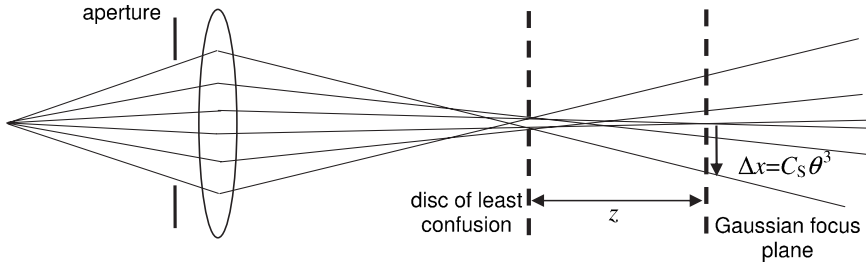


Figure 2-3. A geometric optics view of the effect of spherical aberration. At the Gaussian focus plane the aberrated rays are displaced by a distance proportional to the cube of the ray angle, θ . The minimum beam diameter is at the disc of least confusion, defocused from the Gaussian focus plane by a distance, z .

at which the beams would have been focused had they been unaberrated. At the Gaussian plane, spherical aberration causes the beams to miss their correct point by a distance proportional to the cube of the angle of ray. Spherical aberration is therefore described as being a third-order aberration, and the constant of proportionality is given the symbol, C_s , such that

$$\Delta x = C_s \theta^3 \quad (2.1)$$

If the convergence angle of the electron beam is limited, then it can be seen in Figure 2-3 that the minimum beam waist, or *disc of least confusion*, is located closer to the lens than the Gaussian plane, and that the best resolution in a STEM is therefore achieved by weakening or *underfocusing* the lens relative to its nominal setting. Underfocusing the lens compensates to some degree for the overfocusing effects of spherical aberration.

The above analysis is based upon geometric optics, and ignores the wave nature of the electron. A more quantitative approach is through wave optics. Because the lens aberrations affect the rays converging to form the probe as a function of angle, they can be incorporated as a phase shift in the front-focal plane (FFP) of the objective lens. The FFP and the specimen plane are related by a Fourier transform, as per the Abbe theory of imaging (Born and Wolf, 1980). A point in the front-focal plane corresponds to one partial-plane wave within the ensemble of plane waves converging to form the probe. The deflection of the ray by a certain distance at the sample corresponds to a phase gradient in the FFP aberration function, and the phase shift due to aberration in the FFP is given by

$$\chi(\mathbf{K}) = \left(\pi z \lambda |\mathbf{K}|^2 + \frac{1}{2} \pi C_s \lambda^3 |\mathbf{K}|^4 \right) \quad (2.2)$$

where we have also included the defocus of the lens, z , and $\mathbf{K}$ is a reciprocal space wavevector that is related to the angle of convergence at the sample by

$$\mathbf{K} = \frac{\theta}{\lambda} \quad (2.3)$$

Thus the point $\mathbf{K}$ in the front-focal plane of the objective lens corresponds to a partial plane wave converging at an angle θ at the sample. Once the peak-to-peak phase change of the rays converging to form the probe is greater than $\pi/2$, there will be an element of destructive interference, which we wish to avoid to form a sharp probe. Equation (2.2) is a quartic function, but we can use negative defocus (underfocus) to minimize the excursion of χ beyond a peak-to-peak change of $\pi/2$ over as wide a range of angles as possible (Figure 2-4). Beyond a critical angle, α , we use a beam-limiting aperture, known as the objective aperture, to prevent the more aberrated rays contributing to the probe. This aperture can be represented in the FFP by a two-dimensional top-hat function, $H_\alpha(\mathbf{K})$. Now we can define a so-called *aperture function*, $A(\mathbf{K})$, that represents the complex wavefunction in the FFP,

$$A(\mathbf{K}) = H_\alpha(\mathbf{K})\exp[i\chi(\mathbf{K})] \quad (2.4)$$

Finally we can compute the wave function of the probe at the sample, or *probe function*, by taking the inverse Fourier transform of (2.4) to give

$$P(\mathbf{R}) = \int A(\mathbf{K})\exp(-i2\pi\mathbf{K} \cdot \mathbf{R})d\mathbf{K} \quad (2.5)$$

To express the ability of the STEM to move the probe over the sample, we can include a shift term in (2.5) to give

$$P(\mathbf{R} - \mathbf{R}_0) = \int A(\mathbf{K})\exp(-i2\pi\mathbf{K} \cdot \mathbf{R})\exp(i2\pi\mathbf{K} \cdot \mathbf{R}_0)d\mathbf{K} \quad (2.6)$$

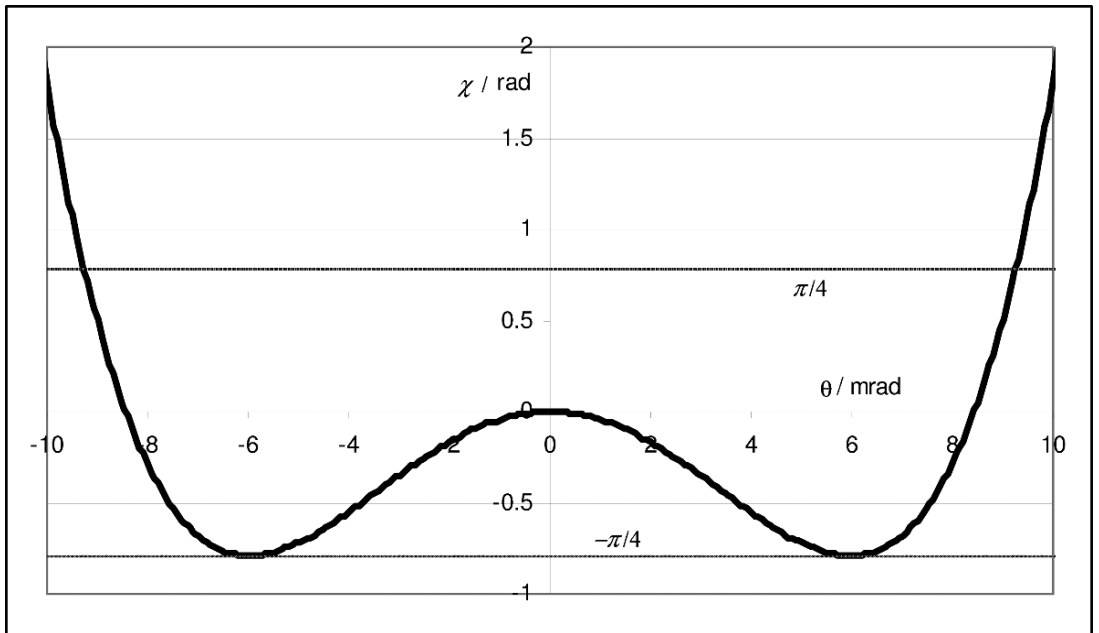


Figure 2-4. The aberration phase shift, χ , in the front-focal, or aperture, plane plotted as a function of convergence angle, θ , for an accelerating voltage of 200 kV, $C_s = 1$ mm and defocus $z = -35.5$ nm. The darker lines indicate the $\pi/4$ limits giving a peak-to-peak variation of $\pi/2$.

Moving the probe is therefore equivalent to adding a linear ramp to the phase variation across the FFP.

The intensity of the probe function is found by taking the modulus squared of $P(\mathbf{R})$, as is plotted for some typical values in Figure 2–5. Note that this so-called *diffraction limited probe* has subsidiary maxima sometimes known as Airy rings, as would be expected from the use of an aperture with a sharp cut-off. These subsidiary maxima can result in weak features observed in images (see Section 5.3) that are image artifacts and not related to the specimen structure.

Let us examine the defocus and aperture size that should be used to provide an optimally small probe. Different ways of measuring probe size lead to various criteria for determining the optimal defocus (see, for example, Mory et al., 1987), but they all lead to similar results. We can again use the criterion of constraining the excursions of χ so that they are no more than $\pi/4$ away from zero. For a given objective lens spherical aberration, the optimal defocus is then given by

$$z = -0.71\lambda^{1/2}C_s^{1/2} \quad (2.7)$$

allowing an objective aperture with radius

$$\alpha = 1.3\lambda^{1/4}C_s^{-1/4} \quad (2.8)$$

to be used. A useful measure of STEM resolution is the full-width at half-maximum (FWHM) of the probe intensity profile. At optimum

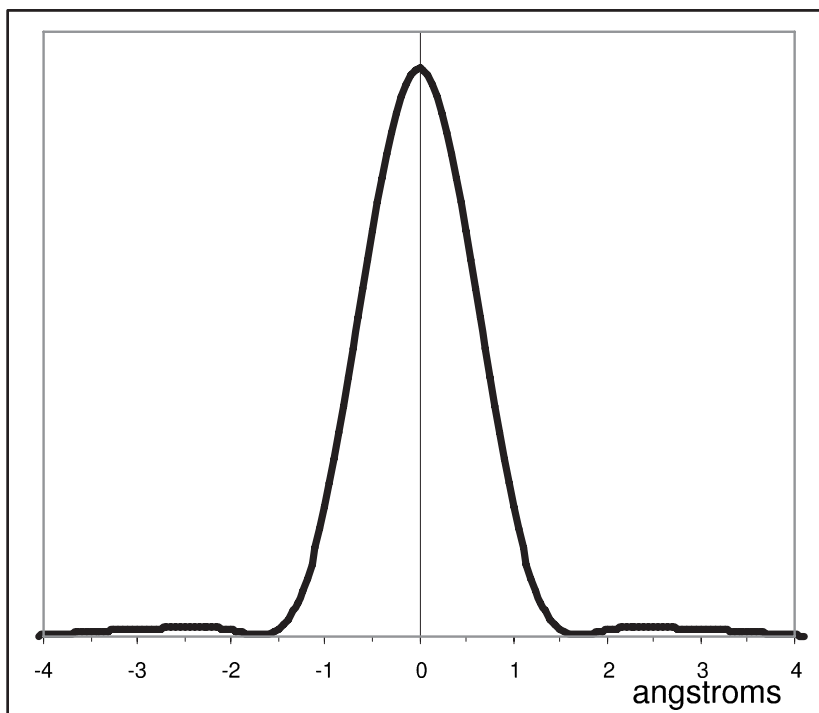


Figure 2–5. The intensity of a diffraction-limited STEM probe for the illumination conditions given in Figure 2–4. An objective aperture of radius 9.3 mrad has been used.

defocus and with the correct aperture size, the probe FWHM is given by

$$d = 0.4\lambda^{3/4}C_s^{1/4} \quad (2.9)$$

Note that the use of increased underfocusing can lead to a reduction in the probe FWHM at the expense of increased intensity in the subsidiary maxima, thereby reducing the useful current in the central maximum and leading to image artifacts. Along with other ways of quoting resolution, the FWHM must be interpreted carefully in terms of the image resolution.

3. Coherent CBED and Ronchigrams

Most STEM detectors are located beyond the specimen and detect the electron intensity in the far field. To interpret STEM images, it is therefore first necessary to understand the intensity found in the far field. In combination CTEM/STEM instruments, the far-field intensity can be observed on the fluorescent screen at the bottom of the column when the instrument is operated in STEM mode with the lower column set to diffraction mode. In dedicated STEM instruments it is usual to have a camera consisting of a scintillator coupled to a CCD array in order to observe this intensity.

In conventional electron diffraction, a sample is illuminated with a highly parallelized plane wave illumination. Electron scattering occurs, and the intensity observed in the far field is given by the modulus squared of the Fourier transform of the wavefunction, $\psi(\mathbf{R})$, at the exit surface of the sample,

$$I(\mathbf{K}) = |\Psi(\mathbf{K})|^2 = \left| \int \psi(\mathbf{R}) \exp[i2\pi\mathbf{K} \cdot \mathbf{R}] d\mathbf{R} \right|^2 \quad (3.1)$$

The scattering wavevector in the detector plane, $\mathbf{K}$, is related to the scattering angle, θ , by

$$\mathbf{K} = \theta/\lambda \quad (3.2)$$

A detailed discussion of electron diffraction is in general beyond the scope of this text, but the reader is referred to the many excellent textbooks on this subject (Hirsch et al., 1977; Cowley, 1990, 1992). In STEM, the sample is illuminated by a probe that is formed from a collapsing convergent spherical wavefront. The electron diffraction pattern is therefore broadened by the range of illumination angles in the convergent beam. In the case of a crystalline sample where one might expect to observe diffracted Bragg spots, in the STEM the spots are broadened into *discs* that may even overlap with their neighbors. Such a pattern is known as a convergent beam electron diffraction (CBED) or microdiffraction pattern because the convergent beam leads to a small illumination spot. See Spence and Zuo (1992) for a textbook covering aspects of microdiffraction and CBED and Cowley (1978) for a review of microdiffraction.

3.1 Ronchigrams of Crystalline Materials

If the electron source image at the sample is much smaller than the diffraction limited probe, then the convergent beam forming the probe can be regarded as being *coherent*. A crystalline sample diffracts electrons into discrete Bragg beams, and in a STEM these are broadened to give discs. The high coherence of the beam means that if the discs overlap then interference features can be seen, such as the fringes in Figure 2–6. Such coherent CBED patterns are also known as coherent microdiffraction patterns or even nanodiffraction patterns. Their observation in the STEM has been described extensively by Cowley (1979b, 1981) and Cowley and Disko (1980) and reviewed by Spence (1992).

To understand the form of these interference fringes, let us first consider a thin crystalline sample that can be described by a simple transmittance function, $\phi(\mathbf{R})$. The exit-surface wavefunction will be given by,

$$\psi(\mathbf{R}, \mathbf{R}_0) = P(\mathbf{R} - \mathbf{R}_0)\phi(\mathbf{R}) \quad (3.3)$$

Where $\mathbf{R}_0$ represents the probe position. Because Eq. 3.3 is a product of two functions, taking its Fourier transform [inserting into Eq. (3.1)] results in a convolution between the Fourier transform of $P(\mathbf{R})$ and the Fourier transform of $\phi(\mathbf{R})$. Taking the Fourier transform of $P(\mathbf{R})$, from Eq. (2.5) simply gives $A(\mathbf{K})$. For a crystalline sample, the Fourier transform of $\phi(\mathbf{R})$ will consist of discrete Dirac δ -functions, which correspond to the Bragg spots, at values of $\mathbf{K}$ corresponding to the reciprocal lattice points. We can therefore write the far field wavefunction, $\Psi(\mathbf{K})$, as a sum of multiple aperture functions centered on the Bragg spots,

$$\Psi(\mathbf{K}, \mathbf{R}_0) = \sum_{\mathbf{g}} \phi_{\mathbf{g}} A(\mathbf{K} - \mathbf{g}) \exp[i2\pi(\mathbf{K} - \mathbf{g}) \cdot \mathbf{R}_0] \quad (3.4)$$

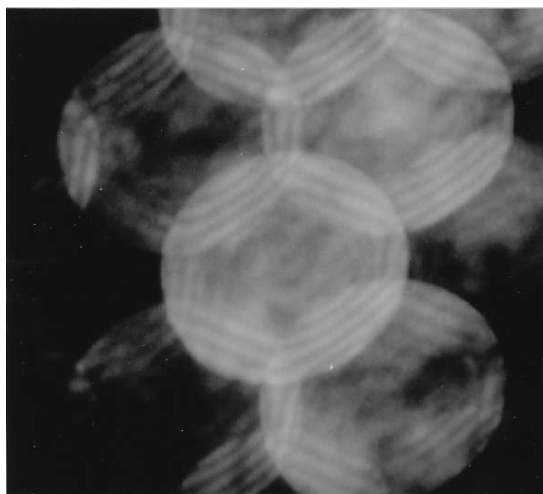


Figure 2–6. A coherent CBED pattern of Si<110>. Note the interference fringes in the overlap region that show that the probe is defocused from the sample.

where ϕ_g is a complex quantity expressing the amplitude and phase of the $\mathbf{g}$ diffracted beam. Equation 3.4 is simply expressing the array of discs seen in Figure 2–6.

To examine just the overlap region between the $\mathbf{g}$ and $\mathbf{h}$ diffracted beam, let us expand (3.4) using (2.4). Since we are just interested in the overlap region we will neglect to include the top-hat function, $H(\mathbf{K})$, which denotes the physical objective aperture, leaving

$$\Psi(\mathbf{K}, \mathbf{R}_0) = \phi_g \exp[i\chi(\mathbf{K} - \mathbf{g}) + i2\pi(\mathbf{K} - \mathbf{g}) \cdot \mathbf{R}_0] + \phi_h \exp[i\chi(\mathbf{K} - \mathbf{h}) + i2\pi(\mathbf{K} - \mathbf{h}) \cdot \mathbf{R}_0] \quad (3.5)$$

and we find the intensity by taking the modulus squared of Eq. (3.5),

$$I(\mathbf{K}, \mathbf{R}_0) = |\phi_g|^2 + |\phi_h|^2 + 2|\phi_g||\phi_h|\cos[\chi(\mathbf{K} - \mathbf{g}) - \chi(\mathbf{K} - \mathbf{h}) + 2\pi(\mathbf{h} - \mathbf{g}) \cdot \mathbf{R}_0 + \angle\phi_g - \angle\phi_h] \quad (3.6)$$

where $\angle\phi_g$ denotes the phase of the $\mathbf{g}$ diffracted beam. The cosine term shows that the disc overlap region contains interference features, and that these features depend on the lens aberrations, the position of the probe, and the phase difference between the two diffracted beams.

If we assume that the only aberration present is defocus, then the terms including χ in (3.6) become

$$\begin{aligned} \chi(\mathbf{K} - \mathbf{g}) - \chi(\mathbf{K} - \mathbf{h}) &= \pi z \lambda [(\mathbf{K} - \mathbf{g})^2 - (\mathbf{K} - \mathbf{h})^2] \\ &= \pi z \lambda [2\mathbf{K} \cdot (\mathbf{h} - \mathbf{g}) + |\mathbf{g}|^2 - |\mathbf{h}|^2] \end{aligned} \quad (3.7)$$

Because Eq. (3.7) is linear in $\mathbf{K}$, a uniform set of fringes will be observed aligned perpendicular to the line joining the centers of the corresponding discs, as seen in Figure 2–6. For interference involving the central, or bright-field, disc we can set $\mathbf{g} = \mathbf{0}$. The spacing of fringes in the microdiffraction pattern from interference between the BF disc and the $\mathbf{h}$ diffracted beam is $(z\lambda|\mathbf{h}|)^{-1}$, which is exactly what would be expected if the interference fringes were a shadow of the lattice planes corresponding to the $\mathbf{h}$ diffracted beam projected using a point source a distance z from the sample (Figure 2–7). When the objective aperture is removed, or if a very large aperture is used, then the intensity in the detector plane is referred to as a shadow image. If the sample is crystalline, then the shadow image consists of many crossed sets of fringes distorted by the lens aberrations. These crystalline shadow images are often referred to as Ronchigrams, deriving from the use of similar images in light optics for the measurement of lens aberrations (Ronchi, 1964). It is common in STEM for shadow images of both crystalline and nonperiodic samples to be referred to as Ronchigrams, however.

The term containing $\mathbf{R}_0$ in the cosine argument in Eq. (3.6) shows that these fringes move as the probe is moved. Just as we might expect for a shadow, we need to move the probe one lattice spacing for the fringes all to move one fringe spacing in the Ronchigram. The idea of the Ronchigram as a shadow image is particularly useful when considering Ronchigrams of amorphous samples (see Section 3.2). Other aberrations, such as astigmatism or spherical aberration, will distort

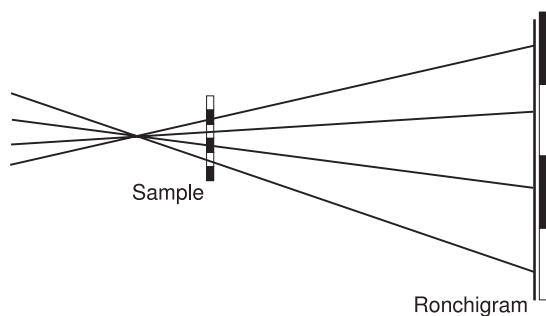


Figure 2-7. If the probe is defocused from the sample plane, the probe crossover can be thought of as a point source located distant from the sample. In the geometric optics approximation, the STEM detector plane is a shadow image of the sample, with the shadow magnification given by the ratio of the probe-detector and probe-sample distances. If the sample is crystalline, then the shadow image is referred to as a Ronchigram.

the fringes so that they are no longer uniform. These distortions may be a useful method of measuring lens aberrations, though the analysis of shadow images for determining lens aberrations is more straightforward with nonperiodic samples (Dellby et al., 2001).

The argument of the cosine in Eq. (3.6) also contains the phase difference between the $\mathbf{g}$ and $\mathbf{h}$ diffracted beams. By measuring the position of the fringes in all the available disc overlap regions, the phase difference between pairs of adjacent diffracted beams can be determined. It is then straightforward to solve for the phase of all the diffracted beams, thereby solving the phase problem in electron diffraction. Knowledge of the phase of the diffracted beams allows immediate inversion to the real-space exit-surface wavefunction. The spatial resolution of such an inversion is limited only by the largest angle diffracted beam that can give rise to observable fringes in the microdiffraction pattern, which will typically be much larger than the largest angle that can be passed through the objective lens (i.e., the radius of the BF disc in the microdiffraction pattern). The method was first suggested by Hoppe (1969a,b, 1982) who gave it the name *ptychography*. Using this approach, Nellist et al. (1995; Nellist and Rodenburg, 1998) were able to form an image of the atomic columns in Si(110) in a STEM that conventionally would be unable to image them. Ptychography has not become a common method in STEM, mainly because the phasing method described above works only for thin samples. In thicker samples, for which dynamic diffraction theory is applicable, the phase of the diffracted beams can depend on the angle of the incident beam. The inherent phase of a diffracted beam may therefore vary across its disc in a microdiffraction pattern, making the simple phasing approach discussed above fail. Spence (1998a,b) has discussed in principle how a crystalline microdiffraction pattern data set can be inverted to the scattering potential for dynamically scattering samples, though as yet there has not been an experimental demonstration.

3.2 Ronchigrams of Noncrystalline Materials

When observing a noncrystalline sample in a Ronchigram, it is generally sufficient to assume that most of the scattering in the sample is at angles much smaller than the illumination convergence angles, and that we can broadly ignore the effects of diffraction. In this case only the BF disc is observable to any significance, but it contains an image of the sample that resembles a conventional bright-field image that would be observed in a conventional TEM at the defocus used to record the Ronchigram (Cowley, 1979b). The magnification of the image is again given by assuming that it is a shadow projected by a point source a distance z (the lens defocus) from the sample. As the defocus is reduced, the magnification increases (Figure 2–8) until it passes through an infinite magnification condition when the probe is focused exactly at the sample. For a quantitative discussion of how Eq. (3.6) reduces to a simple shadow image in the case of predominantly low angle scattering, see Cowley (1979b) and Lupini (2001).

Aberrations of the objective lens will cause the distance from the sample to the crossover point of the illuminating beam to vary as a function of angle within the beam (Figure 2–3), and therefore the apparent magnification will vary within the Ronchigram. Where crossovers occur at the sample plane, infinite magnification regions will be seen. For example, positive spherical aberration combined with negative defocus can give rise to rings of infinite magnification (Figure 2–8). Two infinite magnification rings occur, one corresponding to infinite magnification in the radial direction and one in the azimuthal direction (Cowley, 1986; Lupini, 2001).

Measuring the local magnification within a noncrystalline Ronchigram can readily be done by moving the probe a known distance and measuring the distance features move in the Ronchigram. The local magnifications from different places in the Ronchigram can then be inverted to values for aberration coefficients. This is the method invented by Krivanek et al. (Dellby et al., 2001) for autotuning of a STEM aberration corrector. Even for a non-aberration corrected machine, the Ronchigram of a nonperiodic sample is typically used to align the instrument (Cowley, 1979a). The coma free axis is immediately obvious in a Ronchigram, and astigmatism and focus can be carefully adjusted by observation of the magnification of the speckle contrast. Thicker crystalline samples also show Kikuchi lines in the shadow image, which allows the crystal to be carefully tilted and aligned with the microscope coma-free axis simply by observation of the Ronchigram.

Finally it is worth noting that an electron shadow image for a weakly scattering sample is actually an in-line hologram (Lin and Cowley, 1986) as first proposed by Gabor (1948) for the correction of lens aberrations. The extension of resolution through the ptychographical reconstruction described in Section (3.1) can be extended to nonperiodic samples (Rodenburg and Bates, 1992), and has been demonstrated experimentally (Rodenburg et al., 1993).

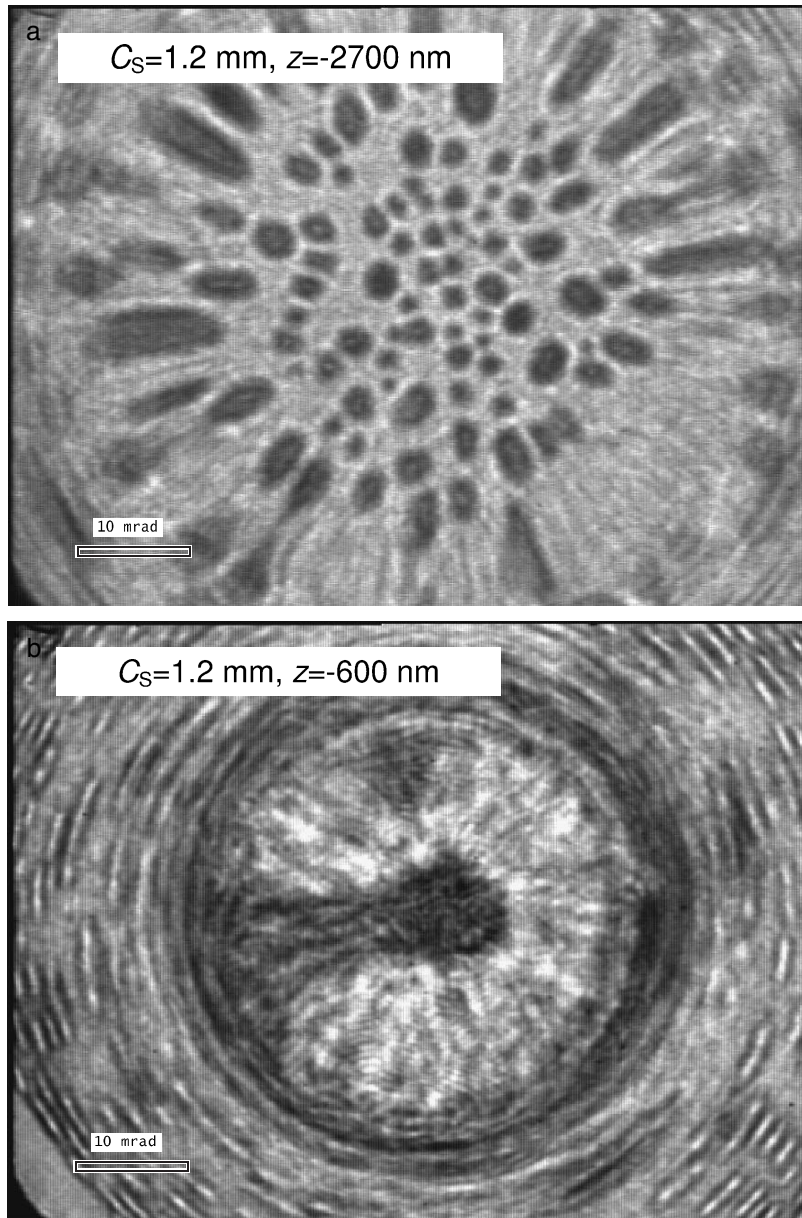


Figure 2-8. Ronchigrams of Au nanoparticles on a thin C film recorded at different defocus values (a and b). Notice the change in image magnification, and the radial and azimuthal rings of infinite magnification.

4. Bright-Field Imaging and Reciprocity

In Section 3 we examined the form of the electron intensity that would be observed in the detector plane of the instrument using an area detector, such as a CCD. In STEM imaging we detect only a single signal, not a two-dimensional array, and plot it as a function of the

probe position. An example of such an image is a STEM BF image, for which we detect some or all of the BF disc in the Ronchigram. Typically the detector will consist of a small scintillator, from which the light generated is directed into a photomultiplier tube. Since the BF detector will just be summing the intensity over a region of the Ronchigram, we can use the Ronchigram formulation in Section 3 to analyze the contrast in a BF image.

4.1 Lattice Imaging in BF STEM

In Section 3.1 we saw that if the diffracted discs in the Ronchigram overlap then coherent interference can occur, and that the intensity in the disc overlap regions will depend on the probe position, $\mathbf{R}_0$. If the discs do not overlap, then there will be no interference and no dependence on probe position. In this latter case, no matter where we place a detector in the Ronchigram, there will be no change in intensity as the probe is moved and therefore no contrast in an image.

The theory of STEM lattice imaging has been described (Spence and Cowley, 1978). Let us first consider the case of an infinitesimal detector right on the axis, which corresponds to the center of the Ronchigram. From Figure 2-9 it is clear that we will see contrast only if the diffracted beams are less than an objective aperture radius from the optic axis. The discs from three beams now interfere in the region detected. From (3.5), the wavefunction at the point detected will be

$$\begin{aligned} \Psi(\mathbf{K} = 0, \mathbf{R}_0) = & 1 + \phi_g \exp[i\chi(-\mathbf{g}) - i2\pi\mathbf{g} \cdot \mathbf{R}_0] \\ & + \phi_{-\mathbf{g}} \exp[i\chi(\mathbf{g}) + i2\pi\mathbf{g} \cdot \mathbf{R}_0] \end{aligned} \quad (4.1)$$

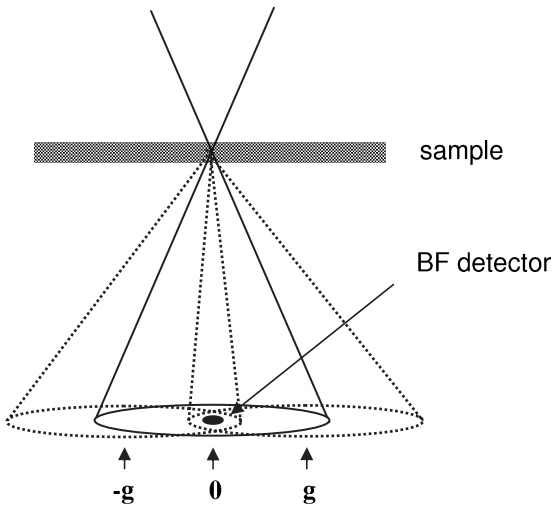


Figure 2-9. A schematic diagram showing that for a crystalline sample, a small, axial bright-field (BF) STEM detector will record changes in intensity due to interference between three beams: the 0 unscattered beam and the $+\mathbf{g}$ and $-\mathbf{g}$ Bragg reflections.

which can also be written as the Fourier transform of the product of the diffraction spots of the sample and the phase shift due to the lens aberrations,

$$\Psi(\mathbf{K} = \mathbf{0}, \mathbf{R}_0) = \int [\delta(\mathbf{K}') + \phi_g \delta(\mathbf{K}' + \mathbf{g}) + \phi_{-g} \delta(\mathbf{K}' - \mathbf{g})] \exp[i\chi(\mathbf{K}')] \exp(i2\pi\mathbf{K}' \cdot \mathbf{R}_0) d\mathbf{K}' \quad (4.2)$$

Equations (4.1) and (4.2) are identical to those for the wavefunction in the image plane of a CTEM when forming an image of a crystalline sample. In the simplest model of a CTEM (Spence, 1988), the sample is illuminated with plane wave illumination. In the back focal plane of the objective lens we could observe a diffraction pattern, and the wavefunction for this plane corresponds to the first bracket in the integrand of (4.2). The effect of the aberrations of the objective lens can then be accommodated in the model by multiplying the wavefunction in the back focal plane by the usual aberration phase shift term, and this can also be seen in (4.2). The image plane wavefunction is then obtained by taking the Fourier transform of this product. Image formation in a STEM can be thought of as being equivalent to a CTEM with the beam trajectories reversed in direction.

What we have shown here, for the specific case of BF imaging of a crystalline sample, is the *principle of reciprocity* in action. When the electrons are purely elastically scattered, and there is no energy loss, the propagation of the electrons is time reversible. The implication for STEM is that the source plane of a STEM is equivalent to the detector plane of a CTEM and vice versa (Cowley, 1969; Zeitler and Thomson, 1970). Condenser lenses are used in a STEM to demagnify the source, which corresponds to projector lenses being used in a CTEM for magnifying the image. The objective lens of a STEM (often used with an objective aperture) focuses the beam down to form the probe. In a CTEM, the objective lens collects the scattered electrons and focuses them to form a magnified image. Confusion can arise with combined CTEM/STEM instruments, in which the probe-forming optics are distinct from the image-forming optics. For example, the term *objective aperture* is usually used to refer to the aperture after the objective lens used in CTEM image formation. In STEM mode, the beam convergence is controlled by an aperture that is usually referred to as the *condenser aperture*, although by reciprocity this aperture is acting optically as an objective aperture. The correspondence by reciprocity between CTEM and STEM can be extended to include the effects of partial coherence. Finite energy spread of the illumination beam in CTEM has an effect on the image similar to that in STEM for the equivalent imaging mode. The finite size of the BF detector in a STEM gives rise to limited spatial coherence in the image (Nellist and Rodenburg, 1994), and corresponds to having a finite divergence of the illuminating beam in a STEM. In STEM, the loss of the spatial coherence can easily be understood as the averaging out of interference effects in the Ronchigram over the area of the BF detector. At the other end of the column there is also a correspondence between the source size in STEM and the detector pixel size in a CTEM. Moving the position of the BF STEM

detector is equivalent to tilting the illumination in CTEM. In this way dark-field images can be recorded. A carefully chosen position for a BF detector could also be used to detect the interference between just two diffracted discs in the microdiffraction pattern, allowing interference between the $\mathbf{0}$ beam and a beam scattered by up to the aperture *diameter* to be detected. In this way higher-spatial resolution information can be recorded, in an equivalent way to using a tilt sequence in CTEM (Kirkland et al., 1995).

Although reciprocity ensures that there is an equivalence in the image contrast between CTEM and STEM, it does not imply that the efficiency of image formation is identical. Bright-field imaging in a CTEM is efficient with electrons because most of the scattered electrons are collected by the objective lens and used in image formation. In STEM, a large range of angles illuminates the sample and these are scattered further to give an extensive Ronchigram. A BF detector detects only a small fraction of the electrons in the Ronchigram, and is therefore inefficient. Note that this comparison applies only for BF imaging. There are other imaging modes, such as annular dark-field (Section 5), for which STEM is more efficient.

4.2 Phase Contrast Imaging in BF STEM

Thin weakly scattering samples are often approximated as being weak phase objects (see, for example, Cowley, 1992). Weak phase objects simply shift the phase of the transmitted wave such that the specimen transmittance function can be written

$$\phi(\mathbf{R}_0) = 1 + i\sigma V(\mathbf{R}_0) \quad (4.3)$$

where σ is known as the interaction constant and has a value given by

$$\sigma = 2\pi m e \lambda / h^2 \quad (4.4)$$

where the electron mass, m , and the wavelength, λ , are relativistically corrected, and V is the projected potential of the sample. Equation (4.3) is simply the expansion of $\exp[i\sigma V(\mathbf{R}_0)]$ to first order, and therefore requires that the product $\sigma V(\mathbf{R}_0)$ is much smaller than unity. The Fourier transform of (4.3) is

$$\Phi(\mathbf{K}') = \delta(\mathbf{K}') + i\sigma \tilde{V}(\mathbf{K}') \quad (4.5)$$

and can be substituted for the first bracket in the integrand of (4.2)

$$\Psi(\mathbf{K} = \mathbf{0}, \mathbf{R}_0) = \int [\delta(\mathbf{K}') + i\sigma \tilde{V}(\mathbf{K}')] \exp[i\chi(\mathbf{K}')] \exp(i2\pi \mathbf{K}' \cdot \mathbf{R}_0) d\mathbf{K}' \quad (4.6)$$

Noticing that (4.6) is the Fourier transform of a product of functions, it can be written as a convolution in $\mathbf{R}_0$.

$$\Psi(\mathbf{K} = \mathbf{0}, \mathbf{R}_0) = 1 + i\sigma V(\mathbf{R}_0) \otimes \text{FT}\{\cos[\chi(\mathbf{K}')] + i\sin[\chi(\mathbf{K}')] \} \quad (4.7)$$

Taking the intensity of (4.7) gives the BF image

$$I(\mathbf{R}_0) = 1 - 2\sigma V(\mathbf{R}_0) \otimes \text{FT}\{\sin[\chi(\mathbf{R}_0)] \} \quad (4.8)$$

where we have neglected terms greater than first order in the potential, and made use of the fact that the sine and cosine of χ are even and therefore their Fourier transforms are real.

Not surprisingly, we have found that imaging a weak-phase object using an axial BF detector results in a phase contrast transfer function (PCTF) (Spence, 1988) identical to that in CTEM, as expected from reciprocity. Lens aberrations are acting as a phase plate to generate phase contrast. In the absence of lens aberrations, there will be no contrast. We can also interpret this result in terms of the Ronchigram in a STEM, remembering that axial BF imaging requires an area of triple overlap of discs (Figure 2–9). In the absence of lens aberrations, the interference between the BF disc and a scattered disc will be in antiphase to that between the BF disc and the opposite, conjugate diffracted disc, and there will be no intensity changes as the probe is moved. Lens aberrations will shift the phase of the interference fringes to give rise to image contrast. In regions of two disc overlap, the intensity will always vary as the probe is moved. Moving the detector to such two beam conditions will then give contrast, just as two-beam tilted illumination in CTEM will give fringes in the image. In such conditions, the diffracted beams may be separated by up to the objective aperture diameter, and still the fringes resolved.

4.3 Large Detector Incoherent BF STEM

Increasing the size of the BF detector reduces the degree of spatial coherence in the image, as already discussed in Section 4.1. One explanation for this is the increasing degree to which interference features in the Ronchigram are being averaged out. Eventually the BF detector can be large enough that the image can be described as being incoherent. Such a large detector will be the complement of an annular dark-field detector: the BF detector corresponding to the hole in the ADF detector. Electron absorption in samples of thicknesses usually used for high-resolution microscopy is small compared to the transmittance, which means that the large detector BF intensity will be

$$I_{\text{BF}}(\mathbf{R}_0) = 1 - I_{\text{ADF}}(\mathbf{R}_0) \quad (4.9)$$

We will defer discussion of incoherent imaging to Section 5. It is, however, worth noting that because I_{ADF} is a small fraction of the incident intensity (typically just a few percent), the contrast in I_{BF} will be small compared to the total intensity. The image noise will scale with the total intensity, and therefore it is likely that a large detector BF image will have worse signal to noise than the complimentary ADF image.

5. Annular Dark-Field Imaging

Annular dark-field (ADF) imaging is by far the most ubiquitous STEM imaging mode [see Nellist and Pennycook (2000) for a review of ADF STEM]. It provides images that are relatively insensitive to focusing

errors, in which compositional changes are obvious in the contrast, and atomic resolution images that are much easier to interpret in terms of atomic structure than their high-resolution TEM (HRTEM) counterparts. Indeed, the ability of a STEM to perform ADF imaging is one of the major strengths of STEM and is partly responsible for the growth of interest in STEM over the past two decades.

The ADF detector is an annulus of scintillator material coupled to a photomultiplier tube in a way similar to the BF detector. It therefore measures the total electron signal scattered in angle between an inner and an outer radius. These radii can both vary over a large range, but typically the inner radius would be in the range of 30–100 mrad and the outer radius 100–200 mrad. Often the center of the detector is a hole, and electrons below the inner radius can pass through the detector for use either to form a BF image, or more commonly to be energy analyzed to form an electron energy-loss spectrum. By combining more than one mode in this way, the STEM makes highly efficient use of the transmitted electrons.

Annular dark-field imaging was introduced in the first STEMs built in Crewe's laboratory (Crewe, 1980). Initially their idea was that the high angle elastic scattering from an atom would be proportional to the product of the number of atoms illuminated and $Z^{3/2}$, where Z is the atomic number of the atoms, and this scattering would be detected using the ADF detector. Using an energy analyzer on the lower-angle scattering they could also separate the inelastic scattering, which was expected to vary as the product of the number of atoms and $Z^{1/2}$. By forming the ratio of the two signals, it was hoped that changes in specimen thickness would cancel, leaving a signal purely dependent on composition, and given the name Z contrast. Such an approach ignores diffraction effects within the sample, which we will see later is crucial for quantitative analysis. Nonetheless, the high-angle elastic scattering incident on an ADF detector is highly sensitive to atomic number. As the scattering angle increases, the scattered intensity from an atom approaches the Z^2 dependence that would be expected for Rutherford scattering from an unscreened Coulomb potential. In practice this limit is not reached, and the Z exponent falls to values typically around 1.7 (see, for example, Hartel et al., 1996) due to the screening effect of the atom core electrons. This sensitivity to atomic number results in images in which composition changes are more strongly visible in the image contrast than would be the case for high-resolution phase-contrast imaging. It is for this reason that using the first STEM operating at 30 kV (Crewe et al., 1970), it was possible to image single atoms of Th on a carbon support.

Once STEM instruments became commercially available in the 1970s, attention turned to using ADF imaging to study heterogeneous catalyst materials (Treacy et al., 1978). Often a heterogeneous catalyst consists of highly dispersed precious metal clusters distributed on a lighter inorganic support such as alumina, silica, or graphite. A system consisting of light and heavy atomic species such as this is an ideal subject for study using ADF STEM. Attempts were made to quantify the number of atoms in the metal clusters using ADF intensities. Howie

(1979) pointed out that if the inner radius was high enough, the thermal diffuse scattering (TDS) of the electrons would dominate. Because TDS is an incoherent scattering process, it was assumed that ensembles of atoms would scatter in proportion to the number of atoms present. It was shown, however, that diffraction effects can still have a large impact on the intensity (Donald and Craven, 1979). Specifically, when a cluster is aligned so that one of the low order crystallographic directions is aligned with the beam, a cluster is observed to be considerably brighter in the ADF image.

An alternative approach to understanding the incoherence of ADF imaging invokes the principle of reciprocity. Phase contrast imaging in an HREM is an imaging mode that relies on a high degree of coherence in order to form contrast. The specimen illumination is arranged to be as plane wave as possible to maximize the coherence. By reciprocity, an ADF detector in a STEM corresponds hypothetically to a large, annular, incoherent illumination source in a CTEM. This type of source is not really viable for a CTEM, but illumination of this sort is extremely incoherent, and renders the specimen effectively self-luminous as the scattering from spatially separated parts of the specimen are unable to interfere coherently. Images formed from such a sample are simpler to interpret as they lack the complicating interference features observed in coherent images. A light-optical analogue is to consider viewing an object with illumination from either a laser or an incandescent light bulb. Laser beam illumination would result in strong interference features such as fringes and speckle. Illumination with a light bulb gives a view much easier to interpret.

Although ADF STEM imaging is very widely used, there are still many discrepancies between the theoretical approaches taken, which can be very confusing when reviewing the literature. A picture of the imaging process that bridges the gap between thinking of the incoherence as arising from integration over a large detector to thinking of it as arising from detecting predominantly incoherent TDS has yet to emerge. Here we will present both approaches, and attempt to discuss the limitations and advantages of each.

5.1 Incoherent Imaging

To highlight the difference between coherent and incoherent imaging, we start by reexamining coherent imaging in a CTEM for a thin sample. Consider plane wave illumination of a thin sample with a transmittance function, $\phi(\mathbf{R}_0)$. The wavefunction in the back focal plane is given by the Fourier transform of the transmittance function, and we can incorporate the effect of the objective aperture and lens aberrations by multiplying the back focal plane by the aperture function to give

$$\Phi(\mathbf{K}')A(\mathbf{K}') \quad (5.1)$$

which can be inverse Fourier transformed to the image wavefunction, which is then a convolution between $\phi(\mathbf{R}_0)$ and the Fourier transform of $A(\mathbf{K}')$, which from Section 2 is $P(\mathbf{R}_0)$. The image intensity is then

$$I(\mathbf{R}_0) = |\phi(\mathbf{R}_0) \otimes P(\mathbf{R}_0)|^2 \quad (5.2)$$

Although for simplicity we have derived (5.2) from the CTEM standpoint, by reciprocity (5.2) applies equally well to BF imaging in STEM with a small axial detector.

For the ADF case we follow the argument first presented by Loane et al. (1992). Similar analyses have been performed by Jesson and Pennycook (1993), Nellist and Pennycook (1998a), and Hartel et al. (1996). Following the STEM configuration, the exit-surface wavefunction is given by the product of the sample transmittance and the probe function,

$$\phi(\mathbf{R}) P(\mathbf{R}-\mathbf{R}_0) \quad (5.3)$$

We can find the wavefunction in the Ronchigram plane by Fourier transforming (5.3), which results in a convolution between the Fourier transform of ϕ and the Fourier transform of P [given in Eq. (2.6)]. Taking the intensity in the Ronchigram and integrating over an annular detector function gives the image intensity

$$I_{\text{ADF}}(\mathbf{R}_0) = \int D_{\text{ADF}}(\mathbf{K}) \left| \int \Phi(\mathbf{K}-\mathbf{K}') A(\mathbf{K}') \exp(i2\pi\mathbf{K}' \cdot \mathbf{R}_0) d\mathbf{K}' \right|^2 d\mathbf{K} \quad (5.4)$$

Taking the Fourier transform of the image allows simplification after expanding the modulus squared to give two convolution integrals

$$\begin{aligned} \tilde{I}_{\text{ADF}}(\mathbf{Q}) = \int \exp(i2\pi\mathbf{Q} \cdot \mathbf{R}_0) \int D_{\text{ADF}}(\mathbf{K}) \left\{ \int \Phi(\mathbf{K}-\mathbf{K}') A(\mathbf{K}') \right. \\ \left. \exp(i2\pi\mathbf{K}' \cdot \mathbf{R}_0) d\mathbf{K}' \right\} \times \left\{ \int \Phi^*(\mathbf{K}-\mathbf{K}'') A^*(\mathbf{K}'') \right. \\ \left. \exp(-i2\pi\mathbf{K}'' \cdot \mathbf{R}_0) d\mathbf{K}'' \right\} d\mathbf{K} d\mathbf{R}_0 \end{aligned} \quad (5.5)$$

Performing the $\mathbf{R}_0$ integral first results in a Dirac δ -function,

$$\begin{aligned} \tilde{I}_{\text{ADF}}(\mathbf{Q}) = \iiint D_{\text{ADF}}(\mathbf{K}) \Phi(\mathbf{K}-\mathbf{K}') A(\mathbf{K}') \Phi^*(\mathbf{K}-\mathbf{K}'') \\ A^*(\mathbf{K}'') \delta(\mathbf{Q} + \mathbf{K}' - \mathbf{K}'') d\mathbf{K} d\mathbf{K}' d\mathbf{K}'' \end{aligned} \quad (5.6)$$

which allows simplification by performing the $\mathbf{K}''$ integral,

$$\begin{aligned} \tilde{I}_{\text{ADF}}(\mathbf{Q}) = \iint D_{\text{ADF}}(\mathbf{K}) A(\mathbf{K}') A^*(\mathbf{K}' + \mathbf{Q}) \Phi(\mathbf{K}-\mathbf{K}') \\ \Phi^*(\mathbf{K}-\mathbf{K}'-\mathbf{Q}) d\mathbf{K} d\mathbf{K}' \end{aligned} \quad (5.7)$$

Equation (5.7) is straightforward to interpret in terms of interference between diffracted discs in the Ronchigram (Figure 2-10). The integral over $\mathbf{K}'$ is a convolution, so that (5.7) could be written,

$$\begin{aligned} \tilde{I}_{\text{ADF}}(\mathbf{Q}) = \int D_{\text{ADF}}(\mathbf{K}) \{ [A(\mathbf{K}) A^*(\mathbf{K} + \mathbf{Q})] \otimes_{\mathbf{K}} [\Phi(\mathbf{K}) \Phi^* \\ (\mathbf{K} - \mathbf{Q})] \} d\mathbf{K} \end{aligned} \quad (5.8)$$

The first bracket of the convolution is the overlap product of two apertures, and this is then convolved with a term that encodes the interference between scattered waves separated by the image spatial frequency $\mathbf{Q}$. For a crystalline sample, $\Phi(\mathbf{K})$ will have values only for discrete $\mathbf{K}$ values corresponding to the diffracted spots. In this case (5.8) is easily

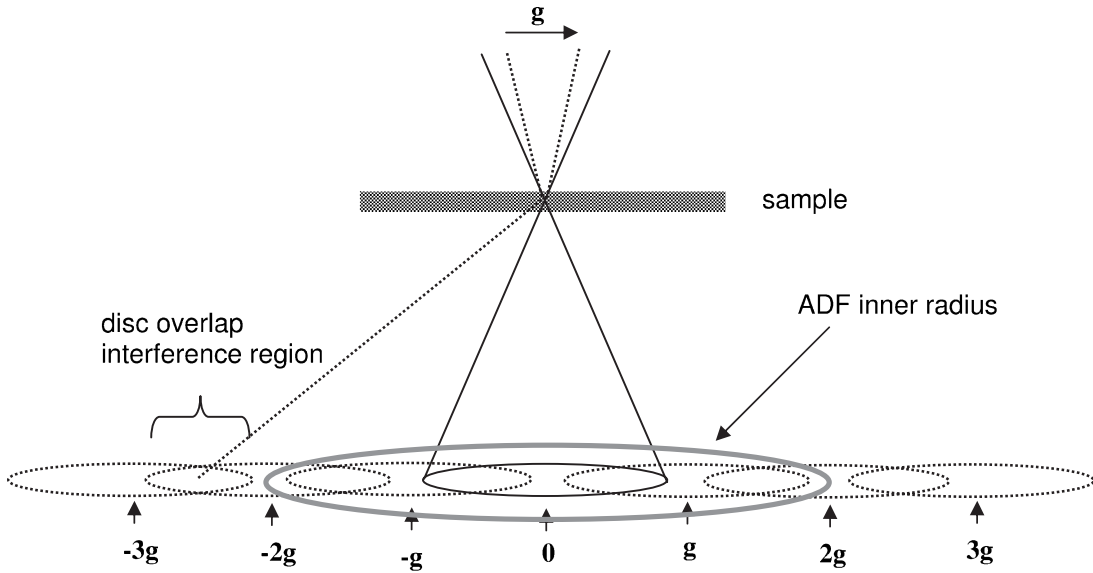


Figure 2-10. A schematic diagram showing the detection of interference in disc overlap regions by the ADF detector. Imaging of a g lattice spacing involves the interference of pairs of beams in the convergent beam that are separated by g . The ADF detector then sums over many overlap interference regions.

interpretable as the sum over many different disc overlap features that are within the detector function. An alternative, but equivalent, interpretation of (5.8) is that for a spatial frequency, $\mathbf{Q}$, to show up in the image, two beams incident on the sample separated by $\mathbf{Q}$ must be scattered by the sample so that they end up in the same final wavevector $\mathbf{K}$ where they can interfere (Figure 2-10). This model of STEM imaging is applicable to any imaging mode, even when TDS or inelastic scattering is included. It was immediately concluded that STEM is unable to resolve any spacing smaller than that allowed by the diameter of the objective aperture, no matter which imaging mode is used.

Figure 2-10 shows that we can expect that the aperture overlap region is small compared with the physical size of the ADF detector. In terms of Eq. (5.7) we can say the domain of the $\mathbf{K}'$ integral (limited to the disc overlap region) is small compared with the domain of the $\mathbf{K}$ integral, and we can make the approximation,

$$\tilde{I}_{\text{ADF}}(\mathbf{Q}) = \int A(\mathbf{K}') A^*(\mathbf{K}' + \mathbf{Q}) d\mathbf{K}' \times \int D_{\text{ADF}}(\mathbf{K}) \Phi(\mathbf{K} - \mathbf{K}') \Phi^*(\mathbf{K} - \mathbf{K}' - \mathbf{Q}) d\mathbf{K} \quad (5.9)$$

In making this approximation we have assumed that the contribution of any overlap regions that are partially detected by the ADF detector is small compared with the total signal detected. The integral containing the aperture functions is actually the autocorrelation of the aperture function. The Fourier transform of the probe intensity is the autocorrelation of A , thus Fourier transforming (5.9) to give the image results in

$$I(\mathbf{R}_0) = |P(\mathbf{R}_0)| \otimes O(\mathbf{R}_0) \quad (5.10)$$

where $O(\mathbf{R}_0)$ is the inverse Fourier transform of the integral over $\mathbf{K}$ in (5.9).

Equation (5.10) is essentially the definition of incoherent imaging. An incoherent image can be written as the convolution between the intensity of the point-spread function of the image (which in STEM is the intensity of the probe) and an object function. Compare this with the equivalent expression for coherent imaging, (5.2), which is the intensity of a convolution between the complex probe function and the specimen function. We will see later that $O(\mathbf{R}_0)$ is a function that is sharply peaked at the atom sites. The ADF image is therefore a sharply peaked object function convolved (or blurred) with a simple, real point-spread function that is simply the intensity of the STEM probe. Such an image is much simpler to interpret than a coherent image, in which both phase and amplitude contrast effects can appear. The difference between coherent and incoherent imaging was discussed at length by Lord Rayleigh in his classic paper discussing the resolution limit of the microscope (Rayleigh, 1896).

A simple picture of the origins of the incoherence can be seen schematically by considering the imaging of two atoms (Figure 2–11). The scattering from the atoms will give rise to interference features in the detector plane. If the detector is small compared with these fringes, then the image contrast will depend critically on the position of the

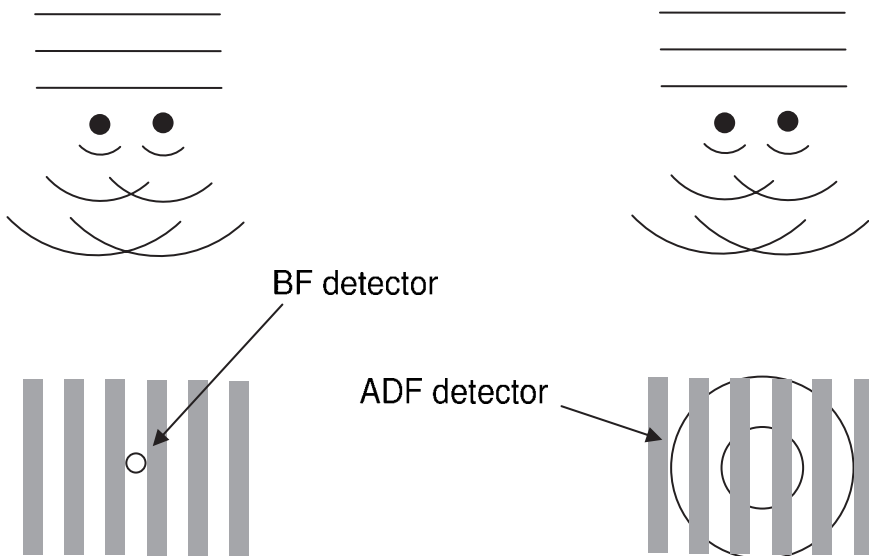


Figure 2–11. The scattering from a pair of atoms will result in interference features such as the fringes shown here. A small detector, such as a BF, will be sensitive to the position of the fringes, and therefore sensitive to the relative phase of the scattered waves and phase changes across the illuminating wave. A larger detector, such as an ADF, will average over many fringes and will therefore be sensitive only to the *intensity* of the scattering and not the phase of the waves.

fringes, and therefore on the relative phases of the scattering from the two atoms, which means that complex phase effects will be seen. A large detector will average over the fringes, destroying any sensitivity to coherence effects and the relative phases of the scattering. By reciprocity, use of the ADF detector can be compared to illuminating the sample with large angle incoherent illumination. In optics, the Van Cittert–Zernicke theorem (Born and Wolf, 1980) describes how an extended source gives rise to a coherent envelope that is the Fourier transform of the source intensity function. An equivalent coherence envelope exists for ADF imaging, and is the Fourier transform of the detector function, $D(\mathbf{K})$. As long as this coherence envelope is significantly smaller than the probe function, the image can be written in the form of (5.10) as being incoherent. This condition is the real-space equivalent of the approximation that allowed us to go from (5.7) to (5.9).

The strength at which a particular spatial frequency in the object is transferred to the image is known, for incoherent imaging, as the optical transfer function (OTF). The OTF for incoherent imaging, $T(\mathbf{Q})$, is simply the Fourier transform of the probe intensity function. In general it is a positive, monotonically decaying function (see Black and Linfoot (1957) for examples under various conditions), which compares favorably with the phase contrast transfer function for the same lens parameters (Figure 2–12).

It can also be seen in Figure 2–12 that the interpretable resolution of incoherent imaging extends to almost twice that of phase-contrast imaging. This was also noted by Rayleigh (1896) for light optics. The explanation can be seen by comparing the disc overlap detection in Figure 2–9 and Figure 2–10. For ADF imaging single overlap regions can be detected, so the transfer continues to twice the aperture radius. The BF detector will detect spatial frequencies only to the aperture radius.

An important consequence of (5.10) is that the phase problem has disappeared. Because the resolution of the electron microscope has always been limited by instrumental factors, primarily the spherical aberration of the objective lens, it has been desirable to be able to deconvolve the transfer function of the microscope. A prerequisite to doing this for coherent imaging is the need to find the phase of the image plane. The modulus-squared in (5.2) loses the phase information, and this must be restored before any deconvolution can be performed. Finding the phase of the image plane in the electron microscope was the motivation behind the invention of holography (Gabor, 1948). There is no phase problem for incoherent imaging, and the intensity of the probe may be immediately deconvolved. Various methods have been applied to this deconvolution problem (Nellist and Pennycook, 1998a, 2000) including Bayesian methods (McGibbon et al., 1994, 1995). As always with deconvolution, care must be taken not to introduce artifacts through noise amplification. The ultimate goal of such methods, though, must be the full quantitative analysis of an ADF image, along with a measure of certainty; for example, the positions of

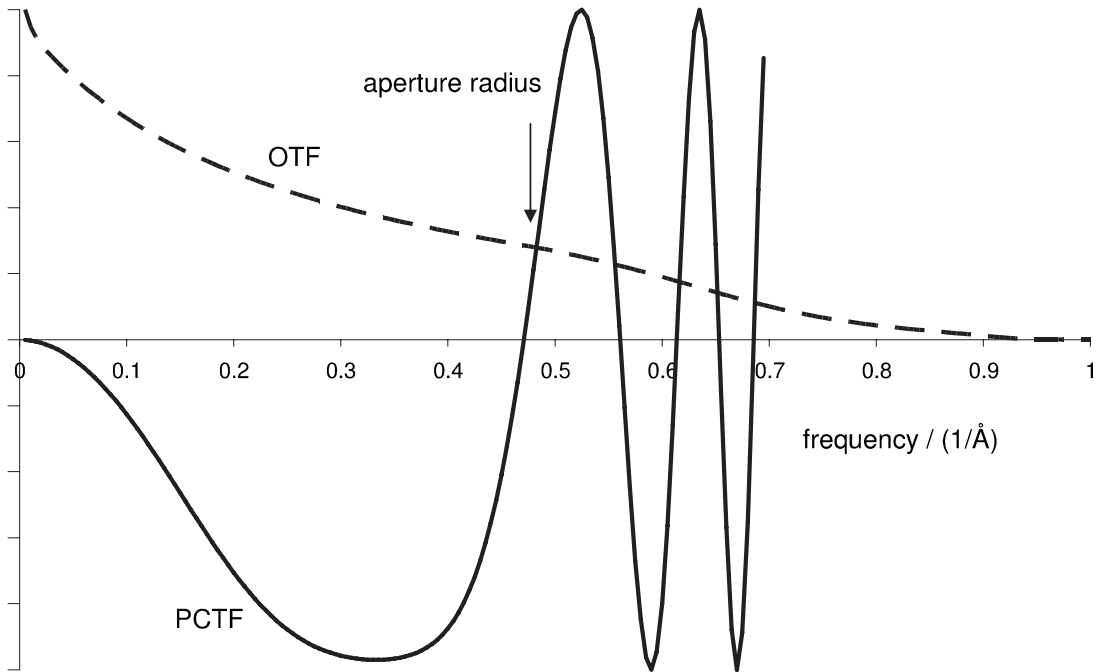


Figure 2-12. A comparison of the incoherent object transfer function (OTF) and the coherent phase-contrast transfer function (PCTF) for identical imaging conditions ($V = 300\text{ kV}$, $C_s = 1\text{ mm}$, $z = -40\text{ nm}$).

atomic columns in an image along with a measure of confidence in the data. Such a goal is yet to be achieved, and the interpretation of most images is still very much qualitative.

The object function, $O(\mathbf{R}_0)$, can also be examined in real space. By assuming that the maximum $\mathbf{Q}$ vector is small compared to the geometry of the detector, and noting that the detector function is either unity or zero, we can write the Fourier transform of the object function as

$$\tilde{O}(\mathbf{Q}) = \int D_{\text{ADF}}(\mathbf{K}) \Phi(\mathbf{K}) D_{\text{ADF}}(\mathbf{K} - \mathbf{Q}) \Phi^*(\mathbf{K} - \mathbf{Q}) d\mathbf{K} \quad (5.11)$$

This equation is just the autocorrelation of $D_{\text{ADF}}(\mathbf{K})\phi(\mathbf{K})$, and so the object function is

$$O(\mathbf{R}_0) = |\tilde{D}_{\text{ADF}}(\mathbf{R}_0) \otimes \phi(\mathbf{R}_0)|^2 \quad (5.12)$$

Neglecting the outer radius of the detector, where we can assume the strength of the scattering has become negligible, $D_{\text{ADF}}(\mathbf{K})$ can be thought of as a sharp high-pass filter. The object function is therefore the modulus-squared of the high-pass filtered specimen transmission function. Nellist and Pennycook (2000) have taken this analysis further by making the weak-phase object approximation, under which condition the object function becomes

$$O(\mathbf{R}_0) = \int_{\text{half plane}} \frac{J_1(2\pi k_{\text{inner}}|\mathbf{R}|)}{2\pi|\mathbf{R}|} [\sigma V(\mathbf{R}_0 + \mathbf{R}/2) - \sigma V(\mathbf{R}_0 - \mathbf{R}/2)]^2 d\mathbf{R} \quad (5.13)$$

where k_{inner} is the spatial frequency corresponding to the inner radius of the ADF detector, and J_1 is a first-order Bessel function of the first kind. This is essentially the result derived by Jesson and Pennycook (1993). The coherence envelope expected from the Van Cittert–Zernicke theorem is now seen in (5.13) as the Airy function involving the Bessel function. If the potential is slowly varying within this coherence envelope, the value of $O(\mathbf{R}_0)$ is small. For $O(\mathbf{R}_0)$ to have significant value, the potential must vary quickly within the coherence envelope. A coherence envelope that is broad enough to include more than one atom in the sample (arising from a small hole in the ADF), however, will show unwanted interference effects between the atoms. Making the coherence envelope too narrow by increasing the inner radius, on the other hand, will lead to too small a variation in the potential within the envelope, and therefore no signal. If there is no hole in the ADF detector, then $D(\mathbf{K}) = 1$ everywhere, and its Fourier transform will be a delta-function. Eq. (5.12) then becomes the modulus-squared of ϕ , and there will be no contrast. To get signal in an ADF image, we require a hole in the detector leading to a coherence envelope that is narrow enough to destroy coherence from neighboring atoms, but broad enough to allow enough interference in the scattering from a single atom. In practice, there are further factors that can influence the choice of inner radius, as discussed in later sections. A typical choice for incoherent imaging is that the ADF inner radius should be about three times the objective aperture radius.

5.2 ADF Images of Thicker Samples

One of the great strengths of atomic resolution ADF images is that they appear to faithfully represent the true atomic structure of the sample even when the thickness is changing over ranges of tens of nanometers. Phase contrast imaging in a CTEM is comparatively very sensitive to changes in thickness, and displays the well-known contrast reversals (Spence, 1988). An important factor in the simplicity of the images is the incoherent nature of ADF images, as we have seen in Section 5.1. The thin object approximation made in Section 5.1, however, is not applicable to the thickness of samples that are typically used, and we need to include the effects of the multiple scattering and propagation of the electrons within the sample. There are several such *dynamical* models of electron diffraction (see Cowley, 1992). The two most common are the Bloch wave approach and the multislice approach. At the angles of scatter typically collected by an ADF detector, the majority of the electrons are likely to be thermal diffuse scattering, having also undergone a phonon scattering event. A comprehensive model of ADF imaging therefore requires both the multiple scattering and the thermal scattering to be included. As discussed earlier, some approaches assume that the ADF signal is dominated by the TDS, and this is assumed to be inco-

herent with respect to the scattering between different atoms. The demonstration of transverse incoherence through the detector geometry and the Van Cittert–Zernicke theorem is therefore ignored by this approach. For lower inner radii, or increased convergence angle (arising from aberration correction, for example) a greater amount of coherent scatter is likely to reach the detector, and the destruction of coherence through the detector geometry will be important for the coherent scatter. As yet, a unifying picture has yet to emerge, and the literature is somewhat confusing. Here we will present the most important approaches currently used.

Initially let us neglect the phonon scattering. By assuming a completely stationary lattice with no absorption, Nellist and Pennycook (1999) were able to use Bloch waves to extend the approach taken in Section 5.1 to include dynamical scattering. It could be seen that the narrow detector coherence function acted to filter the states that could contribute to the image so that the highly bound 1s-type states dominated. Because these states are highly nondispersive, spreading of the probe wavefunction into neighboring column 1s states is unlikely (Rafferty et al., 2001), although spreading into less bound states on neighboring columns is possible. Although this analysis is useful in understanding how an incoherent image can arise under dynamical scattering conditions, its neglect of absorption and phonon scattering effects means that it is not effective as a quantitative method of simulating ADF images.

Early analyses of ADF imaging took the approach that at high enough scattering angles, the TDS arising from phonons would dominate the image contrast. In the Einstein approximation, this scattering is completely uncorrelated between atoms, and therefore there could be no coherent interference effects between the scattering from different atoms. In this approach the intensity of the wavefunction at each site needs to be computed using a dynamical elastic scattering model and then the TDS from each atom summed (Pennycook and Jesson, 1990). When the probe is located over an atomic column in the crystal, the most bound, least dispersive states (usually 1s- or 2s-like) are predominantly excited and the electron intensity “channels” down the column. When the probe is not located over a column, it excites more dispersive, less bound states and spreads leading to reduced intensity at the atom sites and a lower ADF signal. Both the Bloch wave (for example, Pennycook, 1989; Amali and Rez, 1997; Mitsuishi et al., 2001; Findlay et al., 2003b) and multislice (for example, Dinges et al., 1995; Allen et al., 2003) methods have been used for simulating the TDS scattering to the ADF detector. Typically, a dynamic calculation using the standard phenomenological approach to absorption is used to compute the electron wavefunction in the crystal. The absorption is incorporated through an absorptive complex potential that can be included in the calculation simultaneously with the real potential. This method makes the approximation that the absorption at a given point in the crystal is proportional to the product of the absorptive potential and the intensity of the electron wavefunction at that point. Of course, much of the absorption is TDS, which is likely to be detected by the ADF detector.

It is therefore necessary to estimate the fraction of the scattering that is likely to arrive at the detector, and this estimation can cause difficulties. Many estimates of the scattering to the detector, however, make the approximation that the TDS absorption computed for electron scattering in the kinematic approximation to a given angle will end up being at the same angle after phonon scattering. The cross section for the signal arriving at the ADF detector can then be approximated by integrating this absorption over the detector (Pennycook, 1989; Mitsuishi et al., 2001),

$$\sigma_{\text{ADF}} = (4\pi m/m_0)(2\pi/\lambda) \int_{\text{ADF}} |f(s)[1 - \exp(-Ms^2)]|^2 d^2s \quad (5.14)$$

where $s = \theta/2\lambda$ and the $f(s)$ is the electron scattering factor for the atom in question. Other estimates have also been made, some including TDS in a more sophisticated way (Allen et al., 2003b). Caution must be exercised, though. Because this approach is two step—first electrons are absorbed, then a fraction is reintroduced to compute the ADF signal—a wrong estimation in the nature of the scattering can lead to more electrons being reintroduced than were absorbed, thus violating conservation laws.

Making the approximation that all the electrons incident on the detector are TDS neglects any elastic scattering that might be present at the detection angles, which might become significant for lower inner radii. In most cases, including the elastic component is straightforward because it is always computed in order to find the electron intensity within the crystal, but this is not always done in the literature.

Note that the approach outlined above for incoherent TDS scatterers is a fundamentally different approach to understanding ADF imaging, and does not invoke the principles of reciprocity or the Van Zittert–Zernicke theorem. It does not rely on the large geometry of the detector, but just on the fact that it detects only at high angles at which the TDS dominates.

The use of TDS cross sections as outlined above also neglects the further elastic scattering of the electrons after they have been scattered by a phonon. The familiar Kikuchi lines visible in the TDS are manifestations of this elastic scattering. Such scattering occurs only for electrons traveling near Bragg angles, and the major effect is to redistribute the TDS in an angle. It may be reasonably assumed that an ADF detector is so large that the TDS is not redistributed off the detector, and that the electrons are still detected. In general, therefore, the effect of elastic scattering after phonon scattering is usually neglected.

A type of multislice formulation that does include phonon scattering and postphonon elastic scattering has been developed specifically for the simulation of ADF images, and is known as the frozen phonon method (Kirkland et al., 1987; Loane et al., 1991, 1992). An electron accelerated to a typical energy of 100 keV is traveling at about half the speed of light. It therefore transits a sample of thickness, say, 10 nm in 3×10^{-17} s, which is much smaller than the typical period of a lattice vibration ($\sim 10^{-13}$ s). Each electron that transits the sample will see a

lattice in which the thermal vibrations are frozen in some configuration, with each electron seeing a different configuration. Multiple multislice calculations can be performed for different thermal displacements of the atoms, and the resultant intensity in the detector plane is summed over the different configurations. The frozen phonon multislice method is therefore not limited to calculations for STEM; it can be used for many different electron scattering experiments. In STEM, it will give the intensity at any point in the detector plane for a given illuminating probe position. The calculations faithfully reproduce the TDS, Kikuchi lines, and higher-order Laue zone (HOLZ) reflections (Loane et al., 1991). To compute the ADF image, the intensity in the detector plane must be summed over the detector geometry, and this calculation repeated for all the probe positions in the image. The frozen phonon method can be argued to be the most complete method for the computation of ADF images and has been used to compute contrast changes due to composition and thickness changes (Hillyard et al., 1993; Hillyard and Silcox, 1993). Its major disadvantage is that it is computationally expensive. For most multislice simulations of STEM, one calculation is performed for each probe position. In a frozen phonon calculation, several multislice calculations are required for each probe position in order to average effectively over the thermal lattice displacements.

Most of the approaches discussed so far have assumed an Einstein phonon dispersion in which the vibrations of neighboring atoms are assumed to be uncorrelated, and thus the TDS scattering from neighboring atoms incoherent. Jesson and Pennycook (1995) have considered the case for a more realistic phonon dispersion, and showed that a coherence envelope parallel to the beam direction can be defined. The intensity of a column can therefore be highly dependent on the destruction of the longitudinal coherence by the phonon lattice displacements. Consider two atoms, A and B, aligned with the beam direction, and let us assume that the scattering intensity to the ADF detector goes as the square of the atomic number (as for Rutherford scattering from an unscreened Coulomb potential). If the longitudinal coherence has been completely destroyed, the intensity from each atom will be independent and the image intensity will be $Z_A^2 + Z_B^2$. Conversely, if there is perfect longitudinal coherence the image intensity will be $(Z_A + Z_B)^2$. A partial degree of coherence with a finite coherence envelope will result in scattering somewhere between these two extremes. However, frozen phonon calculations by Muller et al. (2001) suggest that for a real phonon dispersion, the ADF image is not significantly changed from the Einstein approximation.

Lattice displacements due to strain in a crystal can be regarded as an ensemble of static phonons, and therefore strain can have a large effect on an ADF image (Perovic et al., 1993), giving rise to so-called strain contrast. The degree of strain contrast that shows up in an image is dependent on the inner radius of the ADF detector. As the inner radius is increased, the effect of strain is reduced and the contrast from compositional changes increases. Changing the inner radius of the detector and comparing the two images can often be used to distinguish between

strain and composition changes. A further similar application is the observation of thermal anomalies in quasicrystal lattices (Abe et al., 2003).

It is often found in the literature that the veracity of a particular method is justified by comparing a calculation with an experimental image of a perfect crystal lattice. An image of a crystal contains little information: it can be expressed by a handful of Fourier components and is not a good test of a model. Much more interesting is the interpretation of defects, such as impurity or dopant atoms in a lattice, and particularly their contribution to image when they are at different depths in the sample. Of particular interest is the effect of probe dechanneling. In the Bloch wave formulation, the excitation of the various Bloch states is given by matching the wavefunctions at the entrance surface of a crystal. When a small probe is located over an atomic column, it is likely that the most excited state will be the tightly bound 1s-type state. This state has high transverse momentum, and is peaked at the atom site leading to strong absorption. Whichever model of ADF image formation is used, it may be expected that this will lead to high intensity on the ADF detector and that there will be a peak in the image at the column site. The 1s states are highly nondispersive, which means that the electrons will be trapped in the potential well and will propagate mostly along the column. This channeling effect is well known from many particle scattering experiments, and is important in reducing thickness effects in ADF imaging. The 1s state will not be the only state excited, however, and the other states will be more dispersive, leading to intensity spreading in the crystal (Fertig and Rose, 1981; Rossouw et al., 2003). Spreading of the probe in the crystal is similar to what would happen in a vacuum. The relatively high probe convergence angle means that the focus depth of field is low, and beyond that the probe will spread. Calculations suggest that this dechanneling can lead to artifacts in the image whereby the effect of a heavy impurity atom substitutional in a column can be seen in the intensity of neighboring columns. The degree to which this occurs, however, is dependent on the model of ADF imaging used, and the literature is still far from agreement on this issue.

5.3 Examples of Structure Determination Using ADF Images

Despite the complications in understanding ADF image formation, it is clear that atomic resolution ADF images do provide direct images of structures. An atomic resolution image that is correctly focused will have peaks in intensity located at the atomic columns in the crystal from which the atomic structure can be simply determined. The use of ADF imaging for structure determination is now widespread (Pennycook, 2002).

The subsidiary maxima of the probe intensity (see Section 2) will give rise to a weak artifactual maxima in the image (Figure 2–13) [see also Yamazaki et al. (2001)], but these will be small compared with the primary peaks, and often below the noise level. The ADF image is somewhat “fail-safe” in that incorrect focusing leads to very low con-

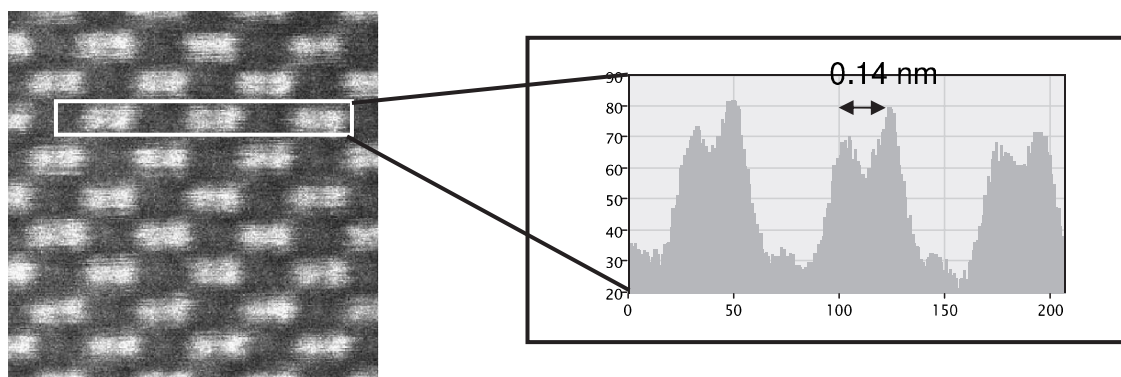


Figure 2-13. An ADF image of GaAs<110> taken using a VG Microscopes HB603U instrument (300 kV, $C_s = 1$ mm). The 1.4-Å spacing between the “dumbbell” pairs of atomic columns is well resolved. An intensity profile shows the polarity of the lattice with the As columns giving greater intensity. The weak subsidiary maxima of the probe can be seen between the columns.

trast, and it is obvious to an operator when the image is correctly focused, unlike phase contrast CTEM for which focus changes do not reduce the contrast so quickly, and just lead to contrast reversals.

There are now many examples in the literature of structure determination by atomic resolution ADF STEM. An excellent recent example is the three-dimensional structural determination of a $\text{NiS}_2/\text{Si}(001)$ interface (Falke et al., 2004) (Figure 2-14). The ability to immediately interpret intensity peaks in the image as atomic columns allowed this structure to be determined, and to correct an earlier erroneous structure determination from HRTEM data.

A disadvantage of scanned images such as an ADF image compared to a conventional TEM image that can be recorded in one shot is that instabilities such as specimen drift manifest themselves as apparent lattice distortions. There have been various attempts to correct for this by using the known structure of the surrounding matrix to correct for the image distortions before analyzing the lattice defect of interest (see, for example, Nakanishi et al., 2002).

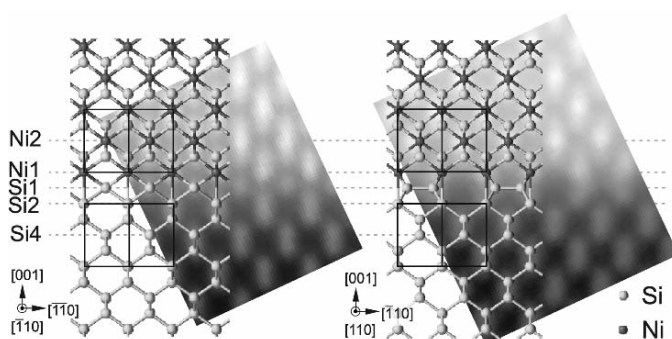


Figure 2-14. An ADF image of an $\text{NiS}_2/\text{Si}(001)$ interface with the structure determined from the image overlaid. [Reprinted with permission from Falke et al. (2004). Copyright (2004) by the American Physical Society.] (See color plate.)

5.4 Examples of Compositionally Sensitive Imaging

The ability of ADF STEM to provide images with high composition sensitivity enabled the very first STEM, operating at 30 kV, to image individual atoms of Th on a carbon support (Crewe et al., 1970). In such a system, the heavy supported atoms are obvious in the image, and little is required in the way of image interpretation. A useful application of this kind of imaging is in the study of ultradispersed supported heterogeneous catalysts (Nellist and Pennycook, 1996). Figure 2–15 shows individual Pt atoms on the surface of a grain of a powered γ -alumina support. Dimers and trimers of Pt may be seen, and their interatomic distances measured. The simultaneously recorded BF image shows fringes from the alumina lattice, from which its orientation can be determined. By relating the BF and ADF images, information on the configuration of the Pt relative to the alumina support may be determined. The exact locations of the Pt atoms were later confirmed from calculations (Sohlberg et al., 2004).

When imaging larger nanoparticles, it is found that the intensity of the particles in the image increases dramatically when one of the particle's low-order crystallographic axes is aligned with the beam. In such a situation, quantitative analysis of the image intensity becomes more difficult.

A more complex situation occurs for atoms substitutional in a lattice, such as dopant atoms. Modern machines have shown themselves to be capable of detecting both Bi (Lupini and Pennycook, 2003) and even Sb dopants (Voyles et al., 2002) in an Si lattice (Figure 2–16). In Voyles

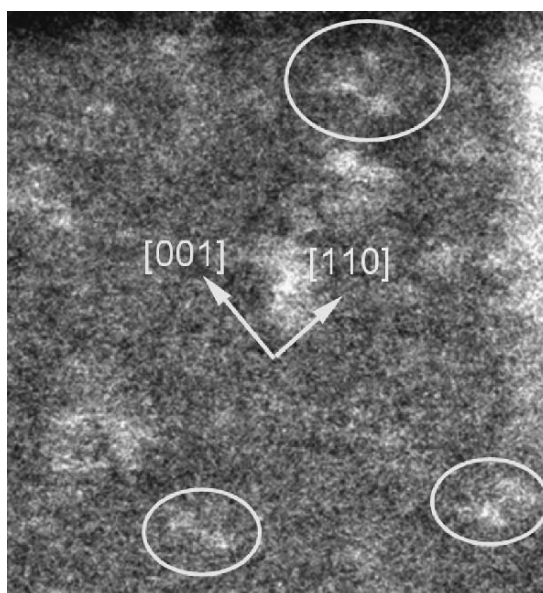


Figure 2–15. An ADF image of individual atoms of Pt on a γ - Al_2O_3 support material. The BF image collected simultaneously showed fringes that allowed the orientation of the γ - Al_2O_3 to be determined. Subsequent theory calculations (see text) confirmed the likely locations of the Pt atoms.

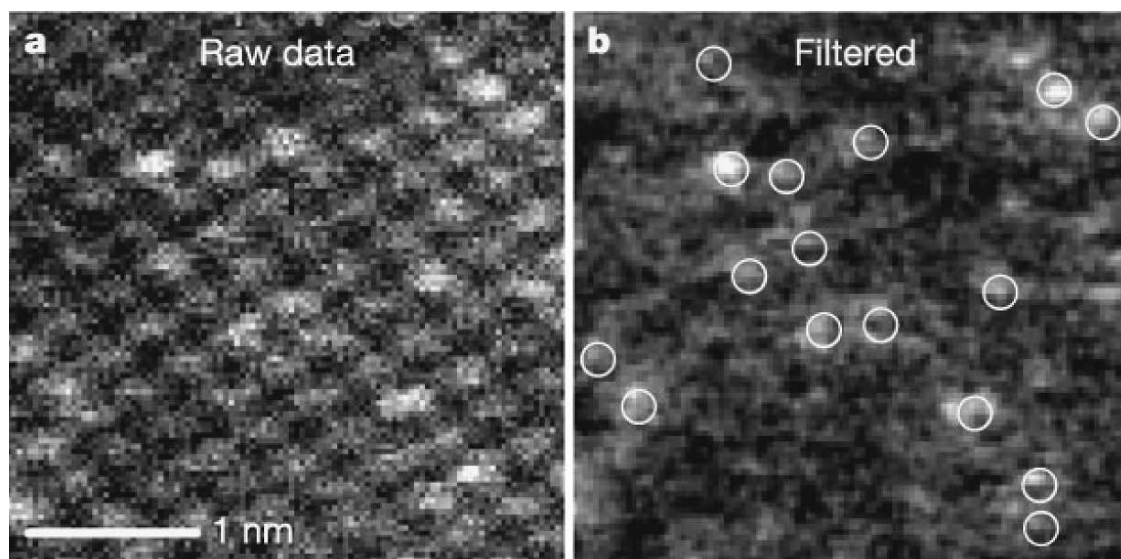


Figure 2-16. An ADF image (left) of Si<110> with visible Sb dopant atoms. On the right, the lattice image has been removed by Fourier filtering leaving the intensity changes due to the dopant atoms visible. (From Voyles et al. (2002), reprinted with permission of Nature Publishing Group.)

et al. (2004) it was noted that the probe channeling then dechanneling effects can change the intensity contribution of the dopant atom depending on its depth in the crystal. Indeed there is some overlap in the range of possible intensities for either one or two dopant atoms in a single column. Another similar example is the observation of As segregation at a grain boundary in Si (Chisholm et al., 1998).

Naturally, ADF STEM is powerful when applied to multilayer structures in which composition sensitivity is desirable. There have been several examples of the application to AlGaAs quantum well structures (see, for example, Anderson et al., 1997). Simulations have been used to enable the image intensity to be interpreted in terms of the fractional content of Al, where it has been assumed that the Al is uniformly distributed throughout the sample.

6. Electron Energy Loss Spectroscopy

So far we have considered the imaging modes of STEM, which predominantly detect elastic or quasielastic scattering of the incident electrons. An equally important aspect of STEM, however, is that it is an extremely powerful analytical instrument. Signals arising from inelastic scattering processes within the sample contain much information about the chemistry and electronic structure of the sample. The small, bright illuminating probe combined with the use of a thin sample means that the interaction volume is small and that analytical information can be gained from a spatially highly localized region of the sample.

Electron energy-loss spectroscopy (EELS) involves dispersing in energy the transmitted electrons through the sample and forming a spectrum of the number of electrons inelastically scattered by a given energy loss versus the energy loss itself. Typically, inelastic scattering events with energy losses up to around 2 keV are intense enough to be useful experimentally.

The energy resolution of EELS spectra can be dictated by both the aberrations of the spectrometer and the energy spread of the incident electron beam. By using a small enough entrance aperture to the spectrometer the effect of spectrometer aberrations will be minimized, albeit with loss of signal. In such a case, the incident beam spread will dominate, and energy resolutions of 0.3 eV with a CFEG source and of about 1 eV with a Schottky source are possible. Inelastic scattering tends to be low angled compared to elastic scattering, with the characteristic scattering angle for EELS being (for example, Brydson, 2001)

$$\theta_E = \Delta E / 2E_0 \quad (6.1)$$

For 100-keV incident electrons, θ_E has a value of 1 mrad for a 200 eV energy loss ranging up to 10 mrad for a 2 keV energy loss. The EELS spectrometer should therefore have a collection aperture that accepts the forward scattered electrons, and should be arranged axially about the optic axis. Such a detector arrangement still allows the use of an ADF detector simultaneously with an EELS spectrometer (see Figure 2-1), and this is one of the important strengths of STEM: an ADF image of a region of the sample can be taken, and spectra can be taken from sites of interest without any change in the detector configuration of the microscope.

There are reviews and books on the EELS technique in both TEM and STEM (see Egerton, 1996; Brydson, 2001; Botton, this volume). In the context of this chapter on STEM, we will mostly focus on aspects of the spatial localization of EELS.

6.1 The EELS Spectrometer

A number of spectrometer designs have emerged over the years, but the most commonly found today, especially with STEM instruments, is the magnetic sector prism, such as the Gatan Enfina system. An important reason for their popularity is that they are not designed to be *in-column*, but can be added as a peripheral to an existing column. Here we will limit our discussion to the magnetic sector prism.

A typical prism consists of a region of homogeneous magnetic field perpendicular to the electron beam (see, for example, Egerton, 1996). In the field region, the electron trajectories follow arcs of circles (Figure 2-1) whose radii depend on the energy of the electrons. Slower electrons are deflected into smaller radii circles. The electrons are therefore dispersed in energy. An additional property of the prism is that it has a focusing action, and will therefore focus the beam to form a line spectrum in the so-called *dispersion* plane. In this plane, the electrons are typically dispersed by around 2 $\mu\text{m}/\text{eV}$. Some spectrometers are fitted with a mechanical slit at this plane that can be used to select part of the

spectrum. A scintillator–photomultiplier combination allows detection of the intensity of the selected part of the spectrum. Using this arrangement, a spectrum can be recorded by varying the strength of the magnetic field, thus sweeping the spectrum over the slit and recording the spectrum serially. Alternatively, the magnetic field can be held constant, selecting just a single energy window, and the probe scanned to form an energy-filtered image.

If there is no slit, or the slit is maximally widened, the spectrum may be recorded in parallel, a technique known as parallel EELS (PEELS). The dispersion plane then needs to be magnified in order that the detector channels allow suitable sampling of the spectrum. This is normally achieved by a series of quadrupoles (normally four) that allows both the dispersion and the width of the spectrum to be controlled at the detector. Detection is usually performed either by a parallel photodiode array, or more commonly now using a scintillator–CCD combination.

Like all electron-optical elements, magnetic prisms suffer from aberrations, and these aberrations can limit the energy resolution of the spectrometer. In general, a prism is designed such that the second-order aberrations are corrected for a given object distance before the prism. Prisms are often labeled with their nominal object distances, which is typically around 70 cm. Small adjustments can be made using sextupoles near the prism and by adjusting the mechanical tilt of the prism. It is important, though, that care is taken to arrange that the sample plane is optically coupled to the prism at the correct working distance to ensure correction of the second-order spectrometer aberrations. More recently, spectrometers with higher order correction (Brink et al., 2003) have been developed. Alternatively, it has been shown to be possible to correct spectrometer aberrations with a specially designed coupling module that can be fitted immediately prior to the spectrometer (see Section 8.1).

Aberrations worsen the ability of the prism to focus the spectrum as the width of the beam entering the prism increases. Collector apertures are therefore used at the entrance of the prism to limit the beam width, but they also limit the number of electrons entering the prism and therefore the efficiency of the spectrum detection. The trade-off between signal strength and energy resolution can be adjusted to the particular experiment being performed by changing the collector aperture size. Aperture sizes in the range of 0.5–5 mm are typically provided.

6.2 Inelastic Scattering of Electrons

The different types of inelastic scattering event that can lead to an EELS signal have been discussed many times in the literature (for example, Egerton, 1996; Brydson, 2001; Botton, this volume), so we will restrict ourselves to a brief description here. A schematic diagram of a typical EEL spectrum is shown in Figure 2-17.

The samples typically used for high-resolution STEM are usually thinner than the mean free path for inelastic scattering (around 100 nm

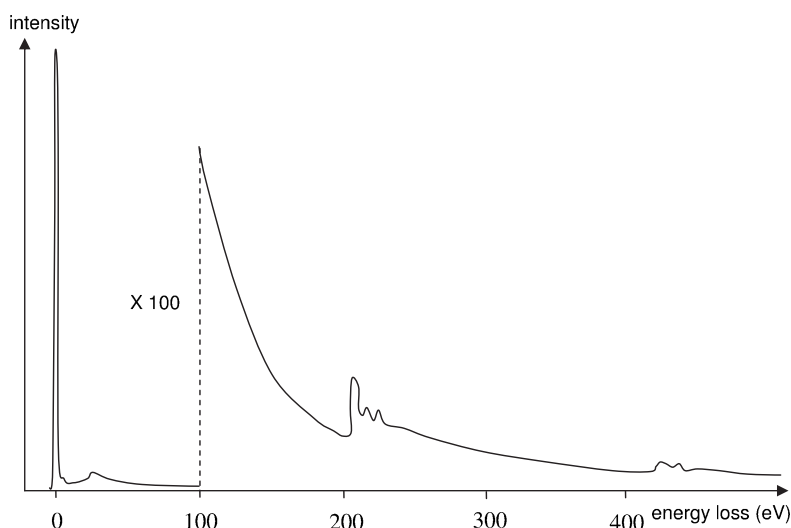


Figure 2–17. A schematic EEL spectrum.

at 100 keV), so the dominant feature in the spectrum is the zero-loss (ZL) peak. When using a spectrometer for high energy resolution, the width of the ZL is usually limited by the energy width of the incident beam. Because STEM instruments require a field-emission gun, this spread is usually small. In a Schottky gun this spread is around 0.8 eV, whereas a CFEG can achieve 0.3 eV or better. The lowest energy losses in the sample will arise from the creation and destruction of phonons, which have energies in the range of 10–100 meV. This range is smaller than the width of the ZL, so such losses will not be resolvable.

The *low-loss* region extends from 0 to 50 eV and corresponds to excitations of electrons in the outermost atomic orbitals. These orbitals can often extend over several atomic sites, and so are delocalized. Both collective and single electron excitations are possible. Collective excitations result in the formation of a *plasmon* or resonant oscillation of the electron gas. Plasmon excitations have the largest cross section of all the inelastic excitations, so the plasmon peak dominates an EEL spectrum, and can complicate the interpretation of other inelastic signals due to multiple scattering effects. Single electron excitations from states in the valence band to empty states in the conduction band can also give rise to low-loss features allowing measurements similar to those in optical spectroscopy, such as band-gap measurements. Further information, for example, distinguishing a direct gap from an indirect gap is available (Rafferty and Brown, 1998). Detailed interpretation of low-loss features involves careful removal of the ZL, however. More commonly, the low-loss region is used as a measure of specimen thickness by comparing the inelastically scattered intensity with the intensity in the ZL. The frequency of inelastic scattering events follows a Poisson distribution, and it can be shown that the sample thickness can be estimated from

$$t = \Lambda \ln(I_T/I_{ZL}) \quad (6.2)$$

where I_T and I_{ZL} are the intensities in the spectrum and zero loss, respectively, and Λ is the inelastic mean-free path, which has been tabulated for some common materials (Egerton, 1996).

From 50 eV up to several thousand eV of energy loss, the inelastic excitations involve electrons in the localized core orbitals on atom sites. Superimposed on a monotonically decreasing background in this *high-loss* region are a series of steps or *core-loss* edges arising from excitations from the core orbitals to just above the Fermi level of the material. The energy loss at which the edge occurs is given by the binding energy of the core orbital, which is characteristic of the atomic species. Measurement of the edge energies therefore allows chemical identification of the material under study. The intensity under the edge is proportional to the number of atoms present of that particular species, so that quantitative chemical analysis can be performed. In a solid sample the bonding in the sample can lead to a significant modification to the density of unoccupied states near the Fermi level, which manifests itself as a fine structure (energy loss near-edge structure, ELNES) in the EEL spectrum in the first 30–40 eV beyond the edge threshold. Although the interpretation of the ELNES can be somewhat complicated, it does contain a wealth of information about the local bonding and structure associated with a particular atomic species. For example, Batson (2000) has used STEM EELS to observe gap states in Si L-edges that are associated with defects observed by ADF. Beyond the near edge region can be seen weaker, extended oscillations (extended energy loss far-edge structure, EXELFS) superimposed on the decaying background. Being further from the edge onset, these excitations correspond to the ejection of a higher kinetic energy electron from the core shell. This higher energy electron generally suffers single scattering from neighboring atoms leading to the observed oscillations and thereby information on the local structural configuration of the atoms such as nearest-neighbor distances.

Clearly EELS has much in common with X-ray absorption studies, with the advantage for EELS being that spectra can be recorded from highly spatially localized regions of the sample. The X-ray counterpart of ELNES is XANES (X-ray absorption near-edge structure), and EXELFS corresponds to EXAFS (extended X-ray absorption fine structure). There are many examples in the literature (for a recent example see Ziegler et al., 2004) in which STEM has been used to record spectra at a defect and the core-loss fine structure used to understand the bonding at the defect.

6.3 The Spatial Localization of EELS Signals and Inelastic Imaging

The strength of EELS in a STEM is that the spectra can be recorded with a high spatial resolution, so the question of the spatial resolution of an EELS signal is an important one. The literature contains several papers demonstrating atomic resolution EELS (Batson, 1993; Browning et al., 1993) and even showing sensitivity to a single impurity atom

(Varela et al., 2004). The lower the energy loss, however, the more the EELS excitation will be delocalized, and an important question is for what excitations is atomic resolution possible.

In addition to the inherent size of the excitation, we must also consider the beam spreading as the probe propagates through the sample. A simple approximation for the beam spreading is given by (Reed, 1982),

$$b = 0.198(\rho/A)^{1/2} (Z/E_0)t^{3/2} \quad (6.3)$$

where b is in nanometers, ρ is the density (g cm^{-3}), A is the atomic weight, Z is the atomic number, E_0 is the incident beam energy in keV, and t is the thickness. At the highest spatial resolutions, especially for a zone-axis oriented sample, a detailed analysis of diffraction and channeling effects (Allen et al., 2003a) is required to model the propagation of the probe through the sample. The calculations are similar to those outlined in Section 5.

Having computed the wavefunction of the illuminating beam within the sample, we now need to consider the spatial extent of the inelastic excitation. This subject has been covered extensively in the literature. Initial studies first considered an isolated atom using a semiclassical model (Ritchie and Howie, 1988). A more detailed study requires a wave optical approach. For a given energy-loss excitation, there will be multiple final states for the excited core electron. The excitations to these various states will be mutually incoherent, leading to a degree of incoherence in the overall inelastic scattering, unlike elastic scattering, which can be regarded as coherent. Inelastic scattering can therefore not be described by a simple multiplicative scattering function, rather we must use a mixed dynamic form factor (MDFF), as described by Kohl and Rose (1985). The formulation used for ADF imaging in Section 5.1 can be adapted for inelastic imaging. Combining the notation of Kohl and Rose (1985) with (5.7) allows us to replace the product of transmission functions with the MDFF,

$$\tilde{I}_{\text{inel}}(\mathbf{Q}) \propto \iint D_{\text{spect}}(\mathbf{K}) A(\mathbf{K}') A^*(\mathbf{K}' + \mathbf{Q}) \frac{S(\mathbf{k}, \mathbf{k} + \mathbf{Q})}{|\mathbf{k}|^2 |\mathbf{k} + \mathbf{Q}|^2} d\mathbf{K} d\mathbf{K}' \quad (6.4)$$

where some prefactors have been neglected for clarity and D now refers to the spectrometer entrance aperture. The inelastic scattering vector, $\mathbf{k}$, can be written as the sum of the transverse scattering vector coupling the incoming wave to the outgoing wave, and the change in wavevector due to the energy loss,

$$\mathbf{k} = \frac{\theta_E \mathbf{e}_z}{\lambda} + \mathbf{K} - \mathbf{K}' \quad (6.5)$$

where $\mathbf{e}_z$ is a unit vector parallel to the beam central axis.

Equations (6.4) and (6.5) show that for a given spatial frequency $\mathbf{Q}$ in the image, the inelastic image can be thought of arising from the sum over pairs of incoming plane waves in the convergent beam separated by $\mathbf{Q}$. Each pair is combined through the MDFF into a final wavevector that is collected by the detector. This is analo-

gous to the model for ADF imaging (see Figure 2–10), except that the product of elastic scattering functions has been replaced with the more general MDFF allowing intrinsic incoherence of the scattering process.

In Section 5.1 we found that under certain conditions, (5.7) could be split into the product of two integrals. This allowed the image to be written as the convolution of the probe intensity and an object function, a type of imaging known as *incoherent imaging*. Let us examine whether (6.4) can be similarly separated. In a similar fashion to the ADF incoherent imaging derivation, if the spectrometer entrance aperture is much larger than the probe convergence angle, then the domain of the integral over $\mathbf{K}$ is much larger than that over $\mathbf{K}'$, and the latter can be performed first. The integral can be then separated thus,

$$\tilde{I}_{inel}(\mathbf{Q}) \propto \int A(\mathbf{K}') A^*(\mathbf{K}' - \mathbf{Q}) d\mathbf{K}' \int D_{spect}(\mathbf{K}) \frac{S(\mathbf{k}, \mathbf{k} + \mathbf{Q})}{|\mathbf{k}|^2 |\mathbf{k} + \mathbf{Q}|^2} d\mathbf{K} \quad (6.6)$$

where the $\mathbf{K}'$ term in $\mathbf{k}$ is now neglected. Since this is a product in reciprocal space, it can be written as a convolution in real space,

$$I_{inel}(\mathbf{R}_0) \propto |P(\mathbf{R}_0)| \otimes O(\mathbf{R}_0) \quad (6.7)$$

where the object function $O(\mathbf{R})$ is the Fourier transform of the integral over $\mathbf{K}$ in (6.5). For spectrometer geometries, $D_{spect}(\mathbf{K})$, that collect only high angles of scatter, it has been shown that this can lead to narrower objects for inelastic imaging (Muller and Silcox, 1995; Rafferty and Pennycook, 1999). Such an effect has not been demonstrated because at such a high angle the scattering is likely to be dominated by combination elastic–inelastic scattering events, and any apparent localization is likely to be due to the elastic contrast.

For inelastic imaging, however, there is another condition for which the integrals can be separated. If the MDFF, S , is slowly varying in $\mathbf{k}$, then the integral in $\mathbf{K}'$ over the disc overlaps will have a negligible effect on S , and the integrals can be separated. Physically, this is equivalent to asserting that the inelastic scattering real-space extent is much smaller than the probe, and therefore the phase variation over the probe sampled by the inelastic scattering event is negligible and the image can be written as a convolution with the probe *intensity*.

We have described the transition from coherent to incoherent imaging for inelastic scattering events in STEM. Note that these terms simply refer to whether the probe can be separated in the manner described above, and does not refer to the scattering process itself. Incoherent imaging can arise with coherent elastic scattering, as described in Section 5.1. The inelastic scattering process is not coherent, hence the need for the MDFF. However, certain conditions still need to be satisfied for the imaging process to be described as incoherent, as described above. An interesting effect occurs for small collector apertures. Because dipole excitations will dominate (Egerton, 1996), a probe located exactly over an atom will not be able to excite transverse excitations because it will not apply a transverse dipole. A slight displacement of the probe is required for such

an excitation. Consequently a dip in the inelastic image is shown to be possible, leading to a donut type of image, demonstrated by Kohl and Rose (1985) and more recently by Cosgriff et al. (2005). This can be thought of as arising from an antisymmetric inelastic object function for a transverse dipole interaction. With a larger collector aperture, the transition to incoherent imaging renders the object function symmetric, removing the dip on the axis.

The width of an inelastic excitation as observed by STEM is therefore a complicated function of the probe, the energy, and the initial wave-function of the core electron and the spectrometer collector aperture geometry. Various calculations have been published exploring this parameter-space. See, for example, Rafferty and Pennycook (1999) and Cosgriff et al. (2005) for some recent examples.

6.4 Spectrum Imaging in the STEM

Historically, the majority of EELS studies in the STEM have been performed in spot mode, in which the probe is stopped over the region of interest in the sample and a spectrum is collected. Of course, the STEM is a scanning instrument, and it is possible to collect a spectrum from every pixel of a scanned image, to form a *spectrum image*. The image may be a one-dimensional line scan, or a two-dimensional image. In the latter case, the data set will be a three-dimensional data cube: two of the dimensions being real-space imaging dimensions and one being the energy loss in the spectra (Figure 2–18).

The spectrum-image data cube naturally contains a wealth of information. Individual spectra can be viewed from any real-space location, or energy-filtered images formed by extracting slices at a given energy loss (Figure 2–18). Selecting energy losses corresponding to the characteristic core edges of the atomic species present in the sample allows

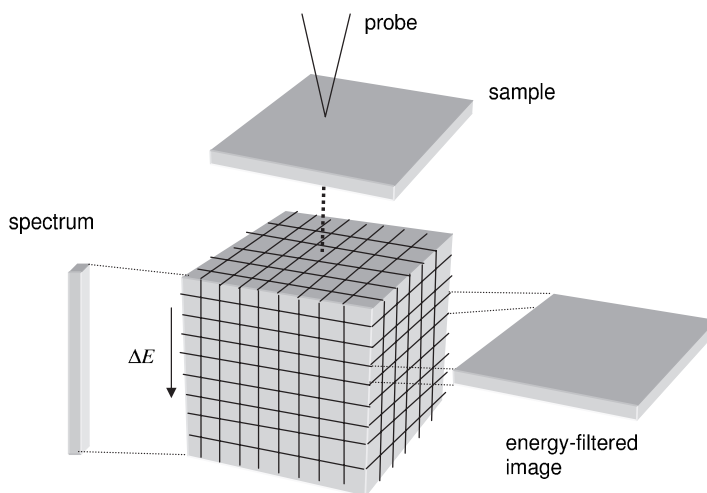


Figure 2–18. A schematic diagram showing how collecting a spectrum at every probe position leads to a data cube from which can be extracted individual spectra or images filtered for a specific energy.

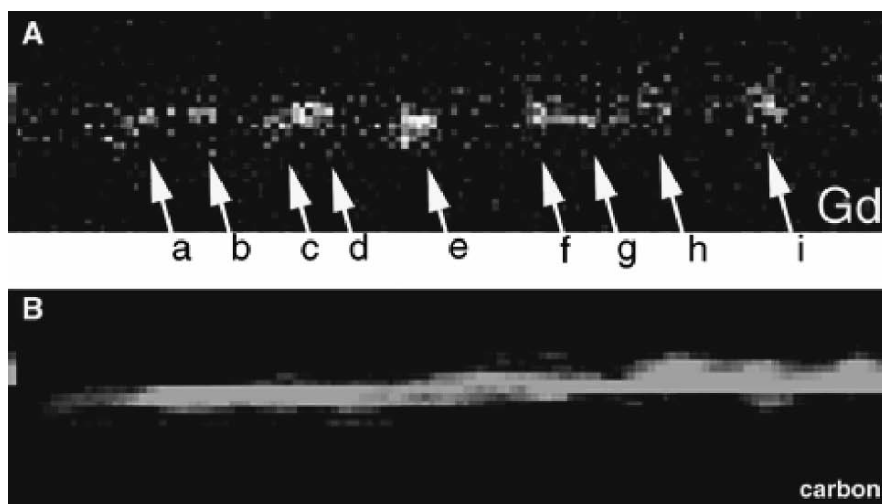


Figure 2–19. A spectrum image filtered for Gd (A) and C (B). Individual atoms of Gd inside a carbon nanotube can be observed. [Reprinted from Suenaga et al. (2000), with copyright permission from AAAS.]

elemental mapping, which, given the inelastic cross sections of the core-loss events, can be calibrated in terms of composition. Using this approach, individual atoms of Gd have been observed inside a carbon nanotube structure (Suenaga et al., 2000) (Figure 2–19). A more sophisticated approach is to use multivariate statistical (MSI) methods (Bonnet et al., 1999) to analyze the compositional maps. With this approach, the existence of phases of certain stoichiometry can be identified, and maps of the phase locations within the sample can be created. Even the fine structure of core-loss edges can be used to form maps in which only the bonding, not the composition, within the sample has changed. An example of this is the mapping of the sp^2 and sp^3 bonding states of carbon at the interface of chemical vapor deposition diamond grown on a silicon substrate (Muller et al., 1993) (Figure 2–20). The sp^2 signal shows the presence of an amorphous carbon layer at the interface.

A similar three-dimensional data cube may also be recorded by conventional TEM fitted with an imaging filter. In this case, the image is recorded in parallel while varying the energy loss being filtered for. Both methods have advantages and disadvantages, and the choice can depend on the desired sampling in either the energy or image dimensions. The STEM does have one important advantage, however. In a CTEM, all of the imaging optics occur after the sample, and these optics suffer significant chromatic aberration. Adjusting the system to change the energy loss being recorded can be done by changing the energy of the incident electrons, thus keeping the energy of the desired inelastically scattered electrons constant within the imaging system. However, to obtain a useful signal-to-noise ratio in energy-filtered transmission electron microscopy (EFTEM), it is necessary to use a selecting energy window that is several electronvolts in width, and even this energy spread in the imaging system is enough to worsen

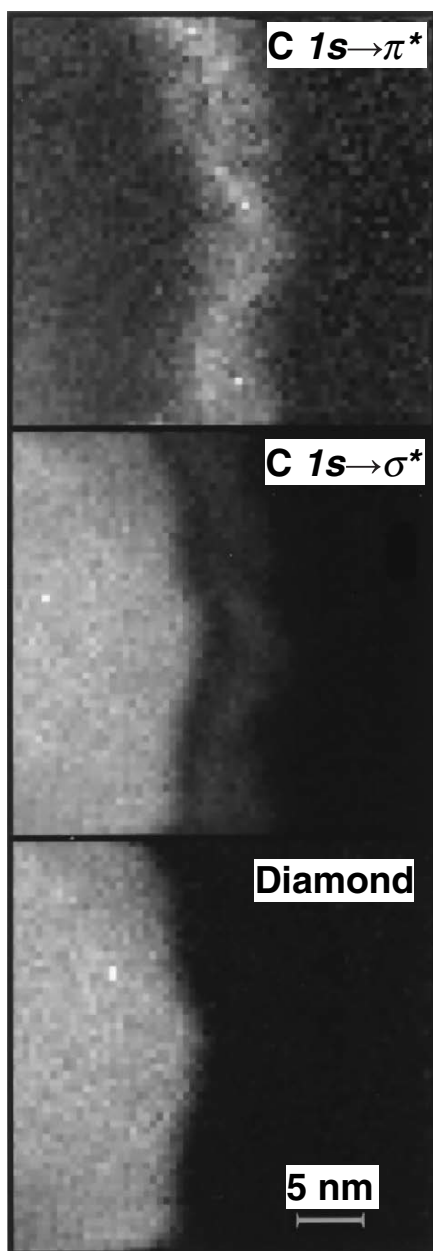


Figure 2–20. By filtering for specific peaks in the fine structure of the carbon K-edge, maps of π and σ bonded carbon can be formed. The presence of an amorphous sp^2 bonded carbon layer at the interface of a chemical vapor deposition (CVD)-grown diamond on an Si substrate can be seen. The diamond signal is derived by a weighted subtraction of the π bonding image from the σ bonding image. [Reprinted from Muller et al. (1993), with permission of Nature Publishing Group.]

the spatial resolution significantly. In STEM, all of the image-forming optics are before the specimen, and the spatial resolution is not compromised.

Inelastic scattering processes, especially single electron excitations, have a scattering cross section that can be an order of magnitude smaller than for elastic scattering. To obtain sufficient signal, EELS acquisition times may be of the order of 1 s. Collection of a spectrum image with a large number of pixels can therefore be very slow, with

the associated problems of both sample drift, and drift of the energy zero point due to power supplies warming up. In practice, spectrum image acquisition software often compensates for these drifts. Sample drift can be monitored using cross-correlations on a sharp feature in the image. Monitoring the position of the zero-loss peak allows the energy drift to be corrected. The advent of aberration correction will have a major impact in this regard. Perhaps one of the most important consequences of aberration correction is that it will increase the current in a given sized probe by more than an order of magnitude (see Section 10.3). Fast elemental mapping through spectrum imaging will then become a much more routine application of EELS. However, to achieve this improvement in performance, there will have to be corresponding improvements in the associated hardware. In general, commercially available systems can achieve around 200 spectra per second. Some laboratories with custom instrumentation have reported reaching 1000 spectra per second (Tencé, personal communication). Further improvement will be necessary to fully make use of spectrum imaging in an aberration corrected STEM.

7. X-Ray Analysis and Other Detected Signals in the STEM

It is obvious that the STEM bears many resemblances to the SEM: a focused probe is formed at a specimen and scanned in a raster while signals are detected as a function of probe position. So far we have discussed BF imaging, ADF imaging, and EELS. All of these methods are unique to the STEM because they involve detection of the fast transmitted electron through a thin sample; bulk samples are typically used in an SEM. There are of course, a multitude of other signals that can be detected in STEM, and many of these are also found in SEM machines.

7.1 Energy Dispersive X-Ray Analysis

When a core electron in the sample is excited by the fast electron traversing the sample, the excited system will subsequently decay with the core hole being refilled. This decay will release energy in the form of an X-ray photon or an Auger electron. The energy of the particle released will be characteristic of the core electron energy levels in the system, and allows compositional analysis to be performed.

The analysis of the emitted X-ray photons is known as energy-dispersive X-ray (EDX) analysis, or sometimes energy-dispersive spectroscopy (EDS) or X-ray EDS (XEDS). It is a ubiquitous technique for SEM instruments and electron-probe microanalyzers. The technique of EDX microanalysis in CTEM and STEM has been extensively covered (Williams and Carter, 1996), and we will review here only the specific features of EDX in a STEM.

The key difference between performing EDX analysis in the STEM as opposed to the SEM is the improvement in spatial resolution (see Figure 2–21). The increased accelerating voltage and the thinner sample used

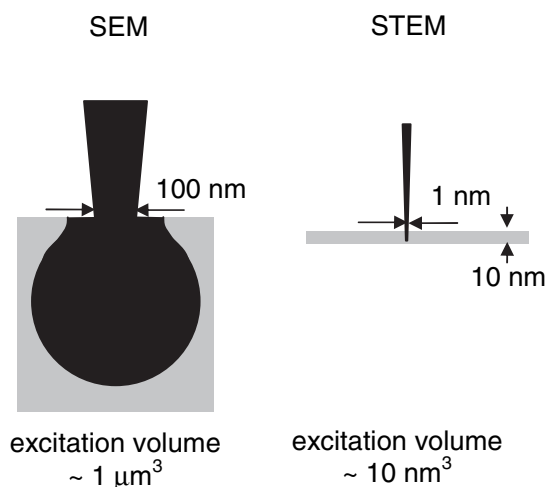


Figure 2-21. A schematic diagram comparing the beam interaction volumes for an SEM and a STEM. The higher accelerating voltage and thinner samples in STEM lead to much higher spatial resolution for analysis, with an associated loss in signal.

in STEM lead to an interaction volume that is some 10^8 times smaller than for an SEM. Beam broadening effects will still be significant for EDX in STEM, and Eq. (6.2) provides a useful approximation in this case. For a given fraction of the element of interest, however, the total X-ray signal will be correspondingly smaller. For a discussion of detection limits for EDX in STEM see Watanabe and Williams (1999). A further limitation for high-resolution STEM instruments is the geometry of the objective lens pole pieces between which the sample is placed. For high resolution the pole piece gap must be small, and this limits both the solid angle subtended by the EDX detector and the maximum take-off angle. This imposes a further reduction on the X-ray signal strength. A high probe current of around 1 nA is typically required for EDX analysis, and this means that the probe size must be increased to greater than 1 nm (see Section 10), thus losing atomic resolution sensitivity. A further concern is the mounting of a large liquid nitrogen dewar on the column for the necessary cooling of the detector. It is often suspected that the boiling of the liquid nitrogen and the unbalancing of the column can lead to mechanical instabilities. A positive benefit of EDX in STEM, however, is that windowless EDX detectors may commonly be used. The vacuum around the sample in STEM is typically higher than for other electron microscopes to reduce sample contamination during imaging and to reduce the gas load on the ultrahigh vacuum of the gun. A consequence is that contamination or icing of a windowless detector is less common.

For the reasons described above, EDX analysis capabilities are sometimes omitted from ultrahigh resolution dedicated STEM instruments, but are common on combination CTEM/STEM instruments. A notable exception has been the development of a 300-kV STEM instrument with the ultimate aim of single-atom EDX detection (Lyman et al., 1994).

It is worth making a comparison between EDX and EELS for STEM analysis. The collection efficiency of EELS can reach 50%, compared to around 1% for EDX, because the X-rays are emitted isotropically. EELS is also more sensitive for light element analysis ($Z < 11$), and for many transition metals and rare-earth elements that show strong spectral features in EELS. The energy resolution in EELS is typically better than 1 eV, compared to 100–150 eV for EDX. The spectral range of EDX, however, is higher with excitations up to 20 keV detectable, compared with around 2 keV for EELS. Detection of a much wider range of elements is therefore possible.

7.2 Secondary Electrons, Auger Electrons, and Cathodoluminescence

Other methods commonly found on an SEM have also been seen on STEM instruments. The usual imaging detector in an SEM is the secondary electron (SE) detector, and these are also found on some STEM instruments. The fast electron incident upon the sample can excite electrons so that they are ejected from the sample. These relatively slow moving electrons can escape only if they are generated relatively close to the surface of the material, and can therefore generate topographical maps of the sample. Once again, because the interaction volume is smaller, the use of SE in STEM can generate high-resolution topographical images of the sample surface. An intriguing experiment involving secondary electrons has been the observation of coincidence between secondary electron emission and primary beam energy-loss events (Mullejans et al., 1993).

Auger electrons are ejected as an alternative to X-ray photon emission in the decay of a core-electron excitation, and spectra can be formed and analyzed just as for X-ray photons. The main difference, however, is that whereas X-ray photons can escape relatively easily from a sample, Auger electrons can escape only when they are created close to the sample surface. It is therefore a surface technique, and is sensitive to the state of the sample surface. Ultrahigh vacuum conditions are therefore required, and Auger in STEM is not commonly found.

Electron-hole pairs generated in the sample by the fast electron can decay by way of photon emission. For many semiconducting samples, these photons will be in or near the visible spectrum and will appear as light, known as cathodoluminescence. Although rarely used in STEM, there has been the occasional investigation (see, for example, Pennycook et al., 1980).

8. Electron Optics and Column Design

Having explored some of the theory and applications of the various imaging and analytical modes in STEM, it is a good time to return to the details of the instrument itself. The dedicated STEM instrument provides a nice model to show the degrees of freedom in the STEM optics, and then we go on to look at the added complexity of a hybrid CTEM/STEM instrument.

8.1 The Dedicated STEM Instrument

We will start by looking at the presample or probe-forming optics of a dedicated STEM, though it should be emphasized that most of the comments in this section also apply to TEM/STEM instruments. In addition to the objective lens, there are usually two condenser lenses (Figure 2-1). The condenser lenses can be used to provide additional demagnification of the source, and thereby control the trade-off between probe size and probe current (see Section 10.1). In principle, only one condenser lens is required because movement of the crossover between the condenser and objective lens (OL) either further or nearer to the OL can be compensated by relatively small adjustments to the OL excitation to maintain the sample focus. The inclusion of two condenser lenses allows the demagnification to be adjusted while maintaining a crossover at the plane of the selected area diffraction aperture. The OL is then set such that the selected area diffraction (SAD) aperture plane is optically conjugate to that of the sample.

In a conventional TEM instrument, the SAD aperture is placed after the OL, and the OL is set to make it optically conjugate to the sample plane. The SAD aperture then selects a region of the sample, and the post-OL lenses are used to focus and magnify the diffraction pattern in the back-focal plane of the OL to the viewing screen. By reciprocity, an equivalent SAD mode can be established in a dedicated STEM (Figure 2-22). With the condenser lenses set to place a crossover at the

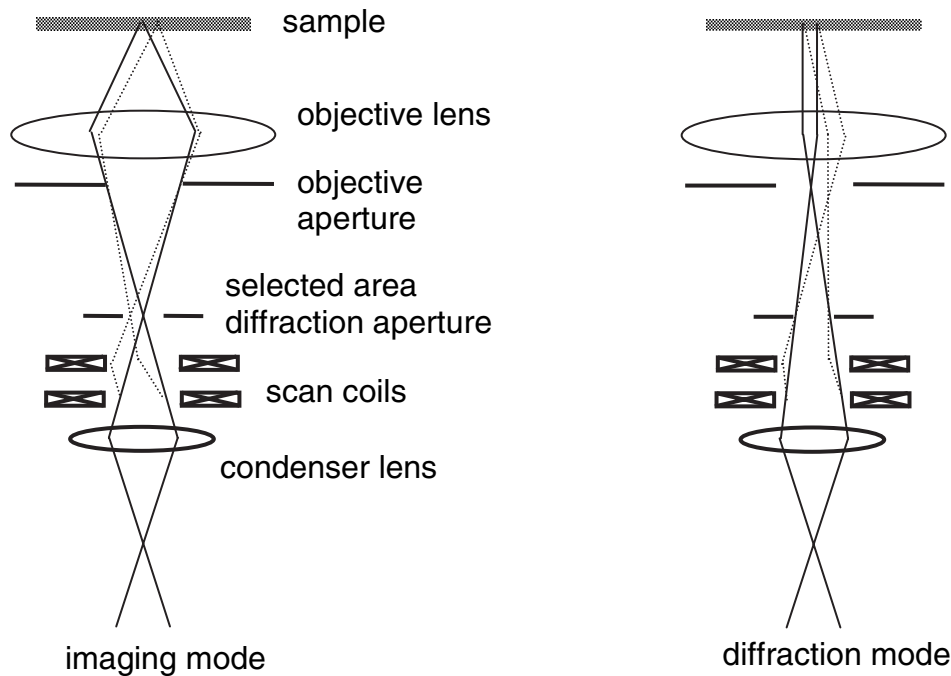


Figure 2-22. The change from imaging to diffraction mode is shown in this schematic of part of a STEM column. By refocusing the condenser lens on the objective lens FFP rather than the SAD aperture plane, the objective lens generates a parallel beam at the sample rather than a focused probe. The SAD aperture is now the beam-limiting aperture, and defines the illumination region on the sample.

SAD, an image can be formed with the SAD selecting a region of interest in the sample. The condenser lenses are then adjusted to place a crossover at the front focal plane of the OL, and the scan coils are set to scan the crossover over the front focal plane. The OL then generates a parallel *pencil* beam that is rocked in angle at the sample plane. In the detector plane is therefore seen a conventional diffraction pattern that is swept across the detector by the scan. By using a small BF detector, a scanned diffraction pattern will be formed. If a Ronchigram camera is available in the detector plane, then the diffraction pattern can be viewed directly and scanning is unnecessary. In practice, SAD mode in a STEM is more commonly used for measuring the angular range of BF and ADF detectors rather than diffraction studies of samples. It is also often used for tilting a crystalline sample to a zone axis if a Ronchigram camera is not available.

To avoid having to mutually align the two condenser lenses, many users employ only one condenser at a time. Both are set to focus a crossover at the SAD aperture plane, but the different distance between the lenses and the SAD plane means that the overall demagnification of the source will differ. Often the two discrete probe current settings then available are suitable for the majority of experiments. Alternatively, many users, especially those with a Ronchigram camera, need an SAD mode very infrequently. In this case, there is no requirement for a crossover in the SAD plane, and one condenser lens can be adjusted freely.

In more modern STEM instruments, a further gun lens is provided in the gun acceleration area. The purpose of this lens is to focus a crossover in the vicinity of the differential pumping aperture that is necessary between the ultrahigh vacuum gun region and the rest of the column. The result is that a higher total current is available for very high current modes. For lower current, higher resolution modes, a gun lens is not found to be necessary.

Let us now turn our attention to the objective lens and the postspecimen optics. The main purpose of the OL is to focus the beam to form a small spot. Just like a conventional TEM, the OL of a STEM is designed to minimize the spherical and chromatic aberration, while leaving a large enough gap for sample rotation and providing a sufficient solid angle for X-ray detection.

An important parameter in STEM is the postsample compression. The field of the objective lens that acts on the electrons after they exit the sample also has a focusing effect on the electrons. The result is that the scattering angles are compressed and the virtual crossover position moves down. Most of the VG dedicated STEM instruments have top-entry OLs, which are consequently asymmetric in shape. The bore on the probe forming (lower) side of the OL is smaller than on the upper side, and therefore the field is more concentrated on the lower side. The typical postsample compression for these asymmetric lenses, typically a factor of around 3, is comparatively low. The entrance to the EELS spectrometer will often be up to 60 cm or more after the sample, to allow room for deflection coils and other detectors. A 2-mm-diameter EELS entrance aperture then subtends a geometric entrance semiangle of

1.7 mrad. Including the factor of 3 compression from the OL gives a typical collection semiangle of 5 mrad. The probe convergence angle of an uncorrected STEM will be around 9 mrad, so the total collection efficiency of the EELS system will be poor, being below 25% after accounting for further angular scattering from the inelastic scattering process. After the correction of spherical aberration, the probe convergence semiangle will rise to 20 mrad or more, and the coupling of this beam into the EELS system will become even more inefficient.

A postspecimen lens would in principle allow improved coupling into the EELS by providing further compression after the beam has left the objective lens. However, there needs to be enough space for deflection coils and lens windings between the lenses, so it is hard to position a postspecimen lens closer than about 100 mm after the OL. By the time the beam has propagated to this lens, it will be of the order of 1 mm in diameter. This is a large diameter beam to be handled by an electron lens, in the lower column typical widths are 50 μm or less, and large aberrations will be introduced that will obviate the benefit of the extra compression. In many dedicated STEMs, therefore, postspecimen lenses are rarely used. A more common work around solution is to mount the sample as low in the OL as possible and to excite the OL as hard as possible to provide the maximum compression possible, though it is difficult to do this and to maintain the tilt capabilities.

A novel solution demonstrated by the Nion Co. is to use a four-quadrupole four-octupole system to couple the postspecimen beam to the spectrometer and provide increased compression. The four-quadrupole system has enough degrees of freedom to provide compression while also ensuring that the virtual crossover as seen by the spectrometer is at the correct object distance. As with any postspecimen lens system in a top entry STEM, the beam is so wide at the lens system that large third-order aberrations are introduced. The presence of the octupoles allows for correction of these aberrations and additionally the third-order aberrations of the spectrometer, which in turn allows a larger physical spectrometer entrance aperture to be used. Collection semiangles up to 20 mrad have been demonstrated with this system (Nellist et al., 2003).

8.2 CTEM/STEM Instruments

At the time of writing, dedicated STEM columns are available from JEOL and Hitachi. Nion Co. has a prototype aberration-corrected dedicated STEM column under test, and this will soon be added to the array of available machines. However, many researchers prefer to use a hybrid CTEM/STEM instrument, which is supplied from all the main manufacturers. As their name suggests, CTEM/STEM instruments offer the capabilities of both modes in the same column.

A CTEM/STEM is essentially a CTEM column with very little modification apart from the addition of STEM detectors. When field-emission guns (FEGs) were introduced onto CTEM columns, it was found that the beam could be focused onto the sample with spot sizes down to 0.2 nm or better (for example, James and Browning, 1999). The

addition of a suitable scanning system and detectors thus created a STEM. The key is that modern CTEM instruments with a side-entry stage tend to make use of the condenser-objective lens (Figure 2–23). In the condenser-objective lens, the field is symmetric about the sample plane, and therefore the lens is just as strong in focusing the beam to a probe presample as it is in focusing the postsample scattered electrons as it would do in conventional TEM mode. The condenser lenses and gun lens play the same roles as those in the dedicated STEM. The main difference in terminology is that what would be referred to as the *objective aperture* in a CTEM/STEM is referred to as the *condenser aperture* in a TEM/STEM. The reason for this is that the aperture in question is usually in or near the condenser lens closest to the OL, and this is the condenser aperture when the column is used in CTEM mode.

An important feature of the CTEM/STEM when operating in the STEM mode is that there are a comparatively large number of post-specimen lenses available. The condenser-objective lens ensures that the beam is narrow when entering these lenses, and so coupling with high compression to an EELS spectrometer does not incur the large aberrations discussed earlier. Further pitfalls associated with high compression should be borne in mind, however. The chromatic aberration of the coupling to the EELS will increase as the compression is increased, leading to edges being out of focus at different energies. Also, the scan of the probe will be magnified in the dispersion plane of the prism, so careful descans needs to be done postsample. A final feature of the extensive postsample optics is that a high magnification image of the probe can be formed in the image plane. This is not as useful for diagnosing aberrations in the probe as one might expect because the aberrations might well be arising from aberrations in the TEM imaging system. Nonetheless, potential applications for such a confocal arrangement have been discussed (see, for example, Möbus and Nufer, 2003).

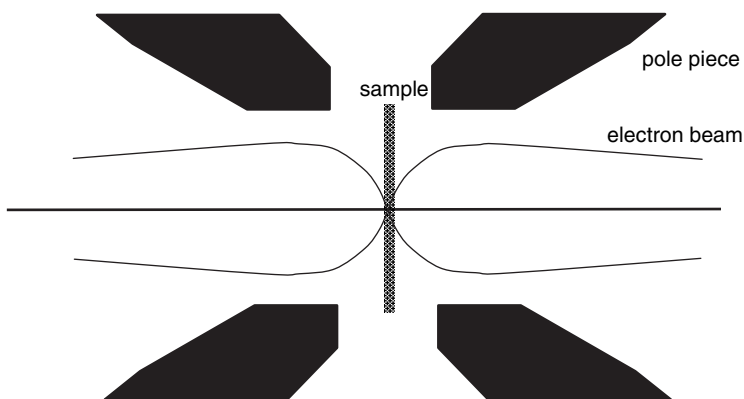


Figure 2–23. A condenser-objective lens provides symmetrical focusing on either side of the central plane. It can therefore be used to provide postsample imaging, as in a CTEM, or to focus a probe at the sample, as in a STEM, or even to provide both simultaneously if direct imaging of the STEM probe is required.

9. Electron Sources

9.1 The Need for Sufficient Brightness

Naively one might expect that the size of the electron source is not critical to the operation of a STEM because we have condenser lenses available in the column to increase the demagnification of the source at will, and thereby still be able to form an image of the source that is below the diffraction limit. We will see, however, that increasing the demagnification decreases the current available in the probe, and the performance of a STEM relies on focusing a significant current into a small spot. In fact, the crucial parameter of interest is that of *brightness* (see, for example, Born and Wolf, 1980). The brightness is defined at the source as

$$B = I/A\Omega \quad (9.1)$$

where I is the total current emitted, A is the area of the source over which the electrons are emitted, and Ω is the solid angle into which the electrons are emitted. Brightness is a useful quantity because at any plane conjugate to the image source (which means any plane where there is a beam crossover), brightness is conserved. This statement holds as long as we consider only geometric optics, which means that we neglect the effects of diffraction. Figure 2–24 shows schematically how the conservation of brightness operates. As the demagnification of an electron source is increased, reducing the area A of the image, the solid angle Ω increases in proportion. Introduction of a beam-limiting aperture forces Ω to be constant, and therefore the total beam current, I , decreases in proportion to the decrease in the area of the source image.

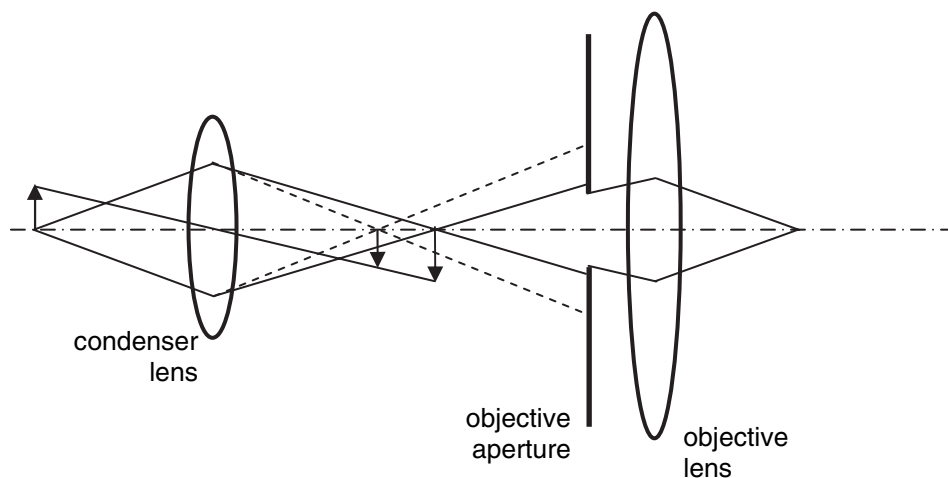


Figure 2–24. A schematic diagram showing how beam current is lost as the source demagnification increased. Reducing the focal length of the condenser lens further demagnifies the image of the source, but the solid angle of the beam correspondingly increases (dashed lines). At a fixed aperture, such as an objective aperture, more current is lost when the beam solid angle increases.

Conservation of brightness is extremely powerful when applied to the STEM. At the probe, the solid angle of illumination is defined by the angle subtended by the objective aperture, α . The maximum value of α is dictated primarily by the spherical aberration of the microscope, and can therefore be regarded as a constant. Given the brightness of the source, we can immediately infer the beam current given the desired size of the source image, or vice versa. Knowledge of the source size is important in determining the resolution of the instrument for a given source size. We can now ask what the necessary source brightness for a viable STEM instrument is. In an order-of-magnitude estimation, we can assume that we need about 25 pA focused into a probe diameter, d_{src} , of 0.1 nm. In an uncorrected machine, the spherical aberration of the objective lens limits α to about 10 mrad. The corresponding brightness can then be computed from

$$B = \frac{I}{\left(\pi d_{\text{src}}^2 / 4\right)(\pi \alpha^2)} \quad (9.2)$$

which gives $B \sim 10^9 \text{ A cm}^{-2} \text{ sr}^{-1}$, expressed in its conventional units.

Having determined the order of brightness required for a STEM we should now compare this number with commonly available electron sources. A tungsten filament thermionic emitter operating at 100 kV has a brightness B of around $10^6 \text{ A cm}^{-2} \text{ sr}^{-1}$, and even an LaB₆ thermionic emitter improves this by only a factor of 10 or so. The only electron sources currently developed that can reach the desired brightness are field-emission sources.

9.2 The Cold Field-Emission Gun

In developing a STEM in their laboratory, a prerequisite for Crewe and co-workers was to develop a field emission gun (Crewe et al., 1968a). The gun they developed was a CFEG, shown schematically in Figure 2-25. The principle is shown in Figure 2-26. A tip is formed by electrochemically etching a short length of single crystal tungsten wire (a typical crystallographic orientation is [310]) to form a point with a typical radius of 50–100 nm. When a voltage is applied to the extraction anode, an intense electron field is applied to the sharp tip. The potential in the vacuum immediately outside the tip therefore has a large gradient, resulting in a potential barrier small enough for conduction electrons to tunnel out of the tungsten into the vacuum. An extraction potential of around 3 kV is usually required. A second anode, or multiple anodes, is then provided to accelerate the electrons to the desired total accelerating voltage.

Although the total current emitted by a CFEG (typically 5 μA) is small compared to other electron sources (a W hairpin filament can reach 100 μA), the brightness of a 100-kV CFEG can reach $2 \times 10^9 \text{ A cm}^{-2} \text{ sr}^{-1}$. The explanation lies in the small area of emission ($\sim 5 \text{ nm}$) and the small solid angle cone into which the electrons are emitted (semiangle of 4°). Electrons are likely to tunnel into the vacuum only over the small area in which the extraction field is high enough or where a surface with a suit-

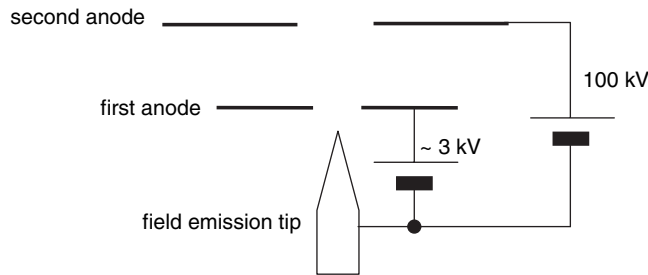


Figure 2–25. A schematic diagram of a 100-kV cold field-emission gun. The proximity of the first anode combined with the sharpness of the tip leads to an intense electric field at the tip, thus extracting the electrons. The first anode is sometime referred to as the *extraction* anode. The second anode provides the further acceleration up to the full beam energy.

ably low workfunction is presented, leading to a small emission area. Only electrons near the Fermi level in the tip are likely to tunnel, and only those whose Fermi velocity is directed perpendicular to the surface, leading to a small emission cone. In addition, the energy spread of the beam from a CFEG is much lower than for other sources, and can be less than 0.3 eV FWHM.

A consequence of the large electrostatic field required for cold field emission is that ultrahigh vacuum conditions are required. Any gas molecules in the gun that become positively ionized by the electron beam will be accelerated and focused directly on the sharp tip. Sputtering of the tip by these ions will rapidly degrade and blunt the tip until its radius of curvature is too large to generate the high fields required for emission. Pressures in the low 10^{-11} Torr are usually maintained in a CFEG. Achieving this kind of pressure requires that the gun be bakable to greater than 200°C, which imposes constraints on the materials and methods of gun construction. Nonetheless, the tip will slowly become contaminated during operation leading to a decay in the beam current. Regular “flashing” is required, whereby a current

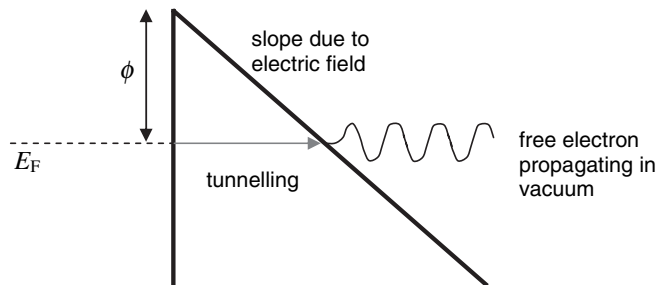


Figure 2–26. A schematic diagram showing the principle of cold field-emission. The vacuum energy level is pulled down into a steep gradient by the application of a strong electric field, producing a triangular energy barrier of height given by the work function, ϕ . Electrons close to the Fermi energy, E_F , can tunnel through the barrier to become free electrons propagating in the vacuum.

is passed through the tip support wire to heat the tip and to desorb the contamination. This is typically necessary once every few hours.

9.3 The Schottky FEG

Cold FEGs have until now been found commercially only in dedicated STEM instruments of VG Microscopes (no longer manufactured) and in some instruments manufactured by Hitachi, although the manufacturers' ranges are always changing. More common is the thermally assisted Schottky field-emission source, introduced by Swanson and co-workers (Swanson and Crouser, 1967).

The principle of operation of the Schottky source is similar to the CFEG, with two major differences: the workfunction of the tungsten tip is lowered by the addition of a zirconia layer, and the tip is heated to around 1700 K. Lowering the workfunction reduces the potential barrier through which electrons have to tunnel to reach the vacuum. Heating the tip promotes the energy at which the electrons are incident on the potential barrier, increasing their probability of tunneling. Heating the tip is also necessary to maintain the zirconia layer on the tip. A reservoir of zirconium metal is provided in the form of a donut on the shank of the tip. The heating of the tip allows zirconium metal to surface migrate under the influence of the electrostatic field toward the sharpened end, oxidizing as it does so as to form a zirconia layer.

Compared to the CFEG, the Schottky source has some advantages and disadvantages. Among the advantages are the fact that the vacuum requirements for the tip are much less strict since the zirconia layer is reformed as soon as it is sputtered away. The Schottky source also has a much greater emission current (around $100\mu\text{A}$) than the CFEG. This makes it a useful source for combination CTEM/STEM instruments with sufficient current for parallel illumination for CTEM work. Disadvantages include a lower brightness (around $2 \times 10^8 \text{ A cm}^{-2} \text{ sr}^{-1}$) and a large emission area, which requires greater demagnification for forming atomic sized probes. For applications involving high-energy resolution spectroscopy, a more serious drawback is the energy spread of the Schottky source at about 0.8 eV.

10. Resolution Limits and Aberration Correction

Having reviewed the STEM instrument and its applications, we finish by reviewing the factors that limit the resolution of the machine. In practice there can be many reasons for a loss in resolution, for example, microscope instabilities or problems with the sample. Here we will review the most fundamental resolution-limiting factors: the finite source brightness, spherical aberration, and chromatic aberration. Round electron lenses suffer from inherent spherical and chromatic aberrations (Scherzer, 1936), and these aberrations dominate the ultimate resolution of STEM. For a field-emission gun, in particular a cold FEG, the energy width of the beam is small, and the effect of C_c is usually smaller than for C_s . The effect of spherical aberration on the resolution and the need for an objective aperture to limit the higher-angle more

aberrated beams have been discussed in Section 2, so here we focus on the effect of the finite brightness and chromatic aberration. Finally we describe the benefits that arise from spherical aberration correction in STEM, and show further applications of aberration correction.

10.1 The Effect of the Finite Source Size

In Section 1 it was mentioned that the probe size in a STEM can be either source size or diffraction limited. In both regimes, the performance of the STEM is limited by the aberrations of the lenses. The aberrations of the OL usually dominate, but in certain modes, such as particularly high current modes, the aberrations of the condenser lenses and even the gun optics might start to have an effect. The lens aberrations limit the maximum size of the beam that may pass through the OL to be focused into the probe. A physical aperture prevents higher angle, more aberrated rays from contributing.

The size of the diffraction-limited probe was described in Section 2. When the probe is diffraction limited, the aperture defines the size of the probe. The resolution of the STEM can be defined in many different ways, and will be different for different modes of imaging. For incoherent imaging we are concerned with the probe intensity, and the full-width at half-maximum may be used given by Eq. (2.9), and repeated here,

$$d_{\text{diff}} = 0.4\lambda^{3/4}C_S^{1/4} \quad (10.1)$$

In the diffraction-limited regime, there is no dependence of the probe size on the probe current.

Once the image of the demagnified source is larger than the diffraction limit, though, the probe will be source size limited. Now the probe size may be traded against the probe current through the source brightness, by rearranging Eq. (9.2) to give

$$d_{\text{src}} = \sqrt{\frac{4I}{B\pi^2\alpha^2}} \quad (10.2)$$

Note that the probe current is limited by the size of the objective aperture, α , and is therefore still limited by the lens aberrations.

The effect of the finite source size will depend on the data being acquired. It can be thought of as an incoherent sum (i.e., a sum in intensity) of many diffraction-limited probes displaced over the source image at the sample. To explain the effect of the finite source size on an experiment, the measurement made for a diffraction-limited probe arising from an infinitesimal source should be summed in intensity with the probe shifted over the source distribution.

The effect on a Ronchigram is to blur the fringes in the disc overlap regions. Remember that the fringes in a disc overlap region correspond to a sample spacing whose spatial frequency is given by the difference of the $\mathbf{g}$ -vectors of the overlapping discs. Once the source size as imaged at the sample is larger than the relevant spacing, the fringes will disappear. This is a very different effect to increasing the probe size through a coherent aberration, such as by defocusing the probe. Defocusing the

probe will lead to changes in the fringe geometry in the Ronchigram, but not in their visibility. The finite source size, however, will reduce the visibility of the fringes. The Ronchigram is therefore an excellent method for measuring the source size of a microscope.

The effect of the finite source size on a BF image is a simple blurring of the image intensity, as would be expected from reciprocity. Once again the image should be computed for a diffraction limited probe arising from an infinitesimal source, and then the image *intensity* blurred over the profile of the source as imaged at the sample. Because BF is a coherent imaging mode, the effect of a finite source size is different to simply increasing the probe size.

The effect of the finite source size on incoherent imaging, such as ADF, is simplest. Because the image is already incoherent, the effect of the finite source size can be thought of as simply increasing the probe size in the experiment. Assuming that both the probe profile and the source image profile are approximately gaussian in form, the combined probe size can be approximated by adding in quadrature,

$$d_{\text{probe}}^2 = d_{\text{diff}}^2 + d_{\text{src}}^2 \quad (10.3)$$

This allows us now to generate a plot of the probe size for incoherent imaging versus the probe current (Figure 2–27).

10.2 Chromatic Aberration

It is not surprising that electrons of higher energies will be less strongly deflected by a magnetic field than those of lower energy. The result of

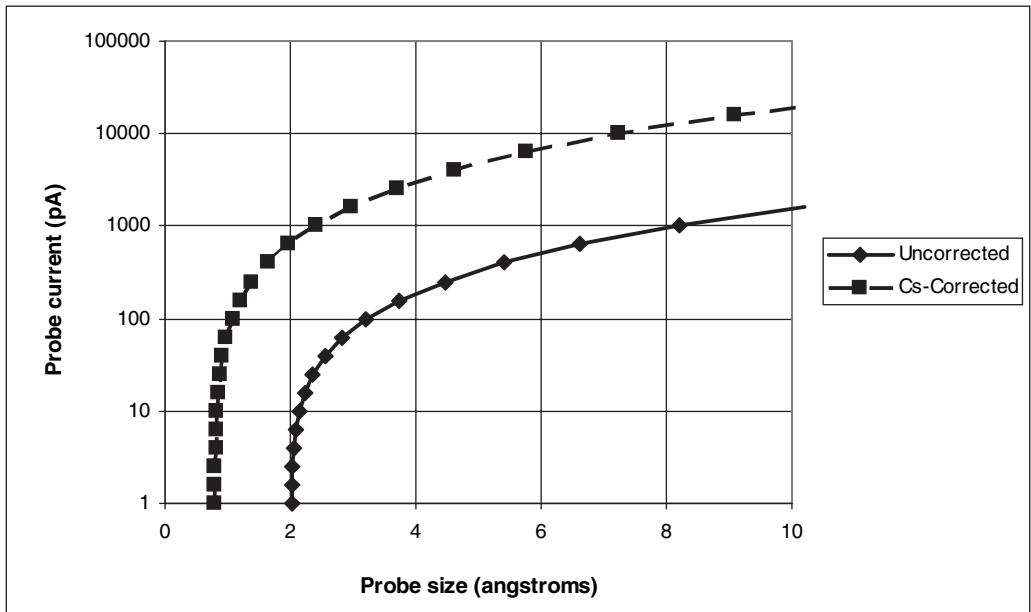


Figure 2–27. A plot of probe size for incoherent imaging versus beam current for both a C_s -afflicted and C_s -corrected machine. The parameters used are 100 kV CFEG with $C_s = 1.3$ mm. Note the diffraction-limited regime where the probe size is independent of current, changing over to a source-size-limited regime at large currents.

this is that the energy spread of the beam will manifest itself as a spread of focal lengths when focused by a lens. In fact, the intrinsic energy spread, instabilities in the high-voltage supply, and instabilities in the lens supply currents will all give rise to a defocus spread through the formula

$$\Delta z = C_c \left(\frac{\Delta E}{V_0} + \frac{2\Delta I}{I_0} + \frac{\Delta V}{V_0} \right) \quad (10.4)$$

where ΔE is the intrinsic energy spread of the beam, ΔV is the variation in accelerating voltage supply, ΔV_0 , and I is the fluctuation in the lens current supply, I_0 . In a modern instrument, the first term should dominate, even with the low energy spread of a CFEG. A typical defocus spread for a 100-kV CFEG instrument will be around 5 nm.

Chromatic aberration is an incoherent aberration, and behaves in a way somewhat similar to the finite source size as described above. The effect of the aberration again depends on the data being acquired. The effect of the defocus spread can be thought of as an incoherent sum (i.e., a sum in intensity) of many experiments performed at a range of defocus values integrated over the defocus spread.

The effect of chromatic aberration on a Ronchigram has been described in detail by Nellist and Rodenburg (1994). Briefly, the perpendicular bisector of the line joining the center of two overlapping discs is achromatic, which means that the intensity does not depend on the defocus value. This is because defocus causes a symmetric phase shift in the incoming beam, and beams equidistant from the center of a disc will therefore suffer the same phase shift resulting in no change to the interference pattern. Away from the achromatic lines, the visibility of the interference fringes will start to reduce.

The effect of C_c on phase contrast imaging has been extensively described in the literature (see, for example, Wade, 1992; Spence, 1988). Here we simply note that in the weak-phase regime, C_c gives rise to a damping envelope in reciprocal space,

$$E_{Cc}(\mathbf{Q}) = \exp\left[-\frac{1}{2}\pi^2\lambda^2(\Delta z)^2|\mathbf{Q}|^4\right] \quad (10.5)$$

where $\mathbf{Q}$ is the spatial frequency in the image. Clearly Eq. (10.5) shows that the Q^4 dependence in the exponential means that C_c imposes a sharp truncation on the maximum spatial frequency of the image transfer.

In contrast, the effect of C_c on incoherent imaging is much less severe. Once again, the effect for incoherent imaging can simply be incorporated by changing the probe intensity profile, $P_{chr}(\mathbf{R})$, through the expression

$$P_{chr}(\mathbf{R}) = \int f(z) |P(\mathbf{R}, z)|^2 dz \quad (10.6)$$

where $f(z)$ is the distribution function of the defocus values.

Nellist and Pennycook (1998b) have derived the effect of C_c on the optical transfer function (OTF). Rather than imposing a multiplicative envelope function, the chromatic spread leads to an upper limit on the OTF that goes as $1/|\mathbf{Q}|$. A plot of the effects of C_c on the incoherent optical transfer function is shown in Figure 2–28. An interesting feature

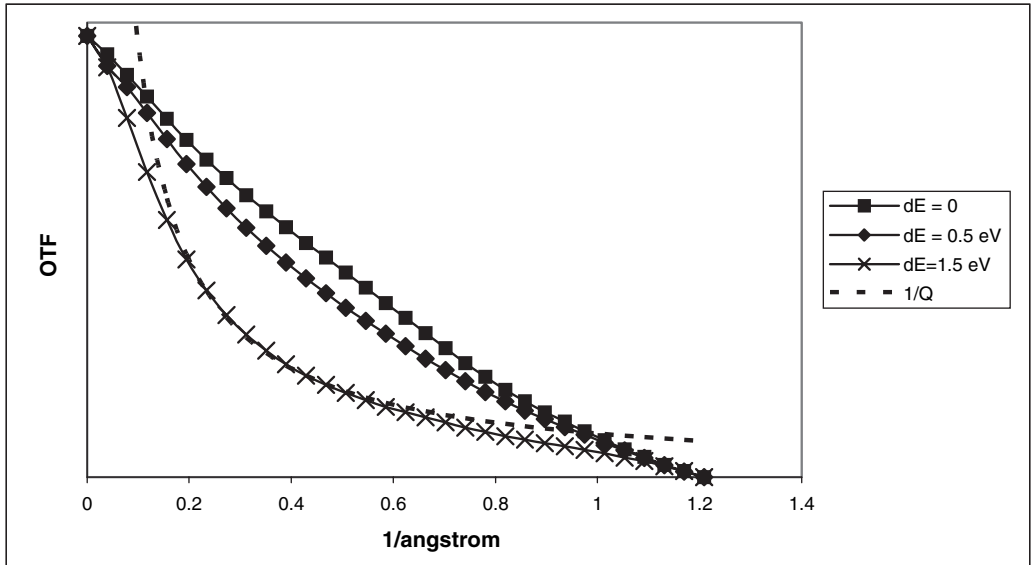


Figure 2-28. A plot of the incoherent optical transfer functions (OTFs) for various defocus spread FWHM values. The microscope parameters are 100 kV with C_s corrected but $C_s = 0.1$ m. Note how the effect is to limit the magnitude of the OTF by a value proportional to the reciprocal of spatial frequency. Such a limit mostly affects the midrange frequencies and not the highest spatial frequencies.

of the effect of C_c on the incoherent transfer function is that the highest spatial frequencies transferred are little affected, explaining the ability of incoherent imaging to reach high spatial resolutions despite any effects of C_c , as shown in Nellist and Pennycook (1998b).

An intuitive explanation of this phenomenon can be found in both real and reciprocal space approaches. In reciprocal space, STEM incoherent imaging can be considered as arising from separate partial plane wave components in the convergent beam that are scattered into the same final wavevector and thereby interfere (see Section 5). The highest spatial frequencies arise from plane wave components on the convergent beam that are separated maximally, which, since the aperture is round, is when they are close to being diametrically opposite. The interference between such beams is often described as being achromatic because the phase shift due to changes in defocus will be identical for both beams, with no resulting effect on the interference. Coherent phase contrast imaging, however, relies on interference between a strong axial beam and scattered beams near the aperture edge, resulting in a high sensitivity to chromatic defocus spread.

The real-space explanation is perhaps simpler. Coherent imaging, as formulated by (5.2), is sensitive to the phase of the probe wavefunction, and the phase will change rapidly as a function of defocus. Summing the image intensities over the chromatic defocus spread will then wash out the high resolution contrast. Incoherent imaging is sensitive only to the intensity of the probe, which is a much more slowly varying

function of defocus. Summing probe intensities over a range of defocus values (see Figure 2–29) shows the effect. The central peak of the probe intensity remains narrow, but intensity is lost to a skirt that extends some distance. Analytical studies will be particularly affected by the skirt, but for a CFEG gun, the effect of C_c will show up only at the highest resolutions, and typically is only seen after the correction of C_s . Krivanek (private communication) has given a simple formula for the fraction of the probe intensity that is shifted away from the probe maximum,

$$f_s = (1 - w)^2 \quad (10.7)$$

where

$$w = 2d_g^2 E_0 / (\Delta E C_c \lambda) \text{ or } w = 1, \text{ whichever is smaller,} \quad (10.8)$$

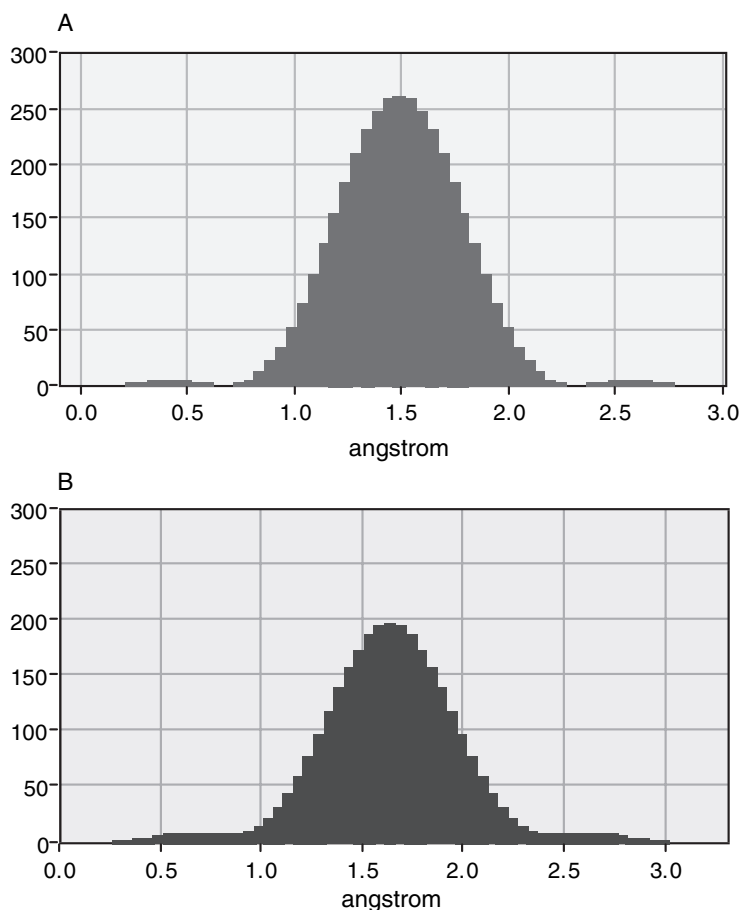


Figure 2–29. Probe profile plots with (A) and without (B) a chromatic defocus spread of 7.5 nm FWHM. The microscope parameters are 100 kV with C_s corrected but $C_s = 0.1$ m. Note that the width of the main peak of the probe is not greatly affected, but intensity is lost from the central maximum into diffuse tails around the probe.

and d_g is the resolution in the absence of chromatic aberration. At a resolution $d_g = 0.8 \text{ \AA}$, energy spread $\Delta E = 0.5 \text{ eV}$, coefficient of chromatic aberration $C_c = 1.5 \text{ mm}$, and primary energy $E_0 = 100 \text{ keV}$, the above gives $f_s = 30\%$ as the fraction of the electron flux shifted out of the probe maximum into the probe tail. This shows that with the low energy spread of a cold field emission gun, the present-day 100 kV performance is not strongly limited by chromatic aberration.

10.3 Aberration Correction

We have spent a lot of time discussing the effects of lens aberrations on STEM performance. Except for some specific circumstances, round electron lenses always suffer positive spherical and chromatic aberrations. This essential fact was first proved by Scherzer in 1936 (Scherzer, 1936), and until recently lens aberrations were the resolution-limiting factor. Scherzer also pointed out that nonround lenses could be arranged to provide negative aberrations (Scherzer, 1947), thereby providing correction of the round lens aberrations. He also proposed a corrector design, but it is only within the last decade that aberration correctors have started to improve microscope resolution over those of uncorrected machines [see, for example, Zach and Haider (1995) for SEM, Haider et al. (1998b) for TEM, and Batson et al. (2002) and Nellist et al. (2004) for STEM]. The key has been the control of parasitic aberrations. Aberration correctors consist of multiple layers of nonround lenses. Unless the lenses are machined perfectly and aligned to each other and the round lenses they are correcting perfectly, nonround parasitic aberrations, such as coma and three-fold astigmatism, will arise and negate the beneficial effects of correction. Recent aberration correctors have been machined to extremely high tolerances, and additional windings and multipoles have been provided to enable correction of the parasitic aberrations. Perhaps even more crucial has been the development of computers and algorithms that can measure and diagnose aberrations fast enough to feed back to the multipole power supplies to correct the parasitic aberrations. A particularly powerful way of measuring the lens aberrations is through the local apparent magnification of the Ronchigram of a nonperiodic object (Dellby et al., 2001) (see Section 3.2).

The key benefits of spherical aberration correction in STEM are illustrated by Figure 2-27. Correction of spherical aberration allows a larger objective aperture to be used because it is no longer necessary to exclude beams that previously would have been highly aberrated. A larger objective aperture has two results: First, the diffraction-limited probe size is smaller so the spatial resolution of the microscope is increased. Second, in the regime in which the electron source size is dominant, the larger objective aperture allows a greater current in the same size probe. Figure 2-27 shows both effects clearly. For low currents the diffraction-limited probe decreases in size by almost a factor of two. In the source size-limited regime, for a given probe size, spherical aberration correction increases the current available by more than an order of magnitude. The increased current available in a C_s cor-

rected STEM is very important for fast elemental mapping or even mapping of subtle changes in fine structure using spectrum imaging (Nellist et al., 2003) (see Section 6).

So far, the impact of spherical aberration correction on resolution has probably been greater in STEM than in CTEM. Part of the reason lies in the robustness of STEM incoherent imaging to C_C . Correction of C_C is more difficult than for C_s , and at the time of writing a commercial C_C corrector for high-resolution TEM instruments is not available. We saw in Section 10.2 that compared to HRTEM, the resolution of STEM incoherent imaging is not severely limited by C_C . Furthermore, the dedicated STEM instruments that have given the highest resolutions have all used cold field emission guns with a low intrinsic energy spread. A second reason for the superior C_s -corrected performance of STEM instruments lies in the fact that they are scanning instruments. In a STEM, the scan coils are usually placed close to the objective lens and certainly there are no optical elements between the scan coils and the objective lens. This means that in most of the electron optics, in particular the corrector, the beam is fixed and its position does not depend on the position of the probe in the image, unlike the case for CTEM. In STEM therefore, only the so-called *axial* aberrations need to be measured and corrected, a much reduced number compared to CTEM for which off-axial aberrations must also be monitored.

Commercially available C_s correctors are currently available from Nion Co. in the United States and CEOS GmbH in Germany. The existing Nion corrector is a quadrupole–octupole design, and is retrofitted into existing VG Microscopes dedicated STEM instruments. Because the field strength in an octupole varies as the cube of the radial distance, it is clear that an octupole should provide a third-order deflection to the beam. However, the four-fold rotational symmetry of the octupole means that a single octupole acting on a round beam will simply introduce third-order four-fold astigmatism. A series of four quadrupoles is therefore used to focus line crossovers in two octupoles, while allowing a round beam to be acted on by the third (central) octupole (see figures in Krivanek et al., 1999). The line crossovers in the outer two octupoles give rise to third-order correction in two perpendicular directions, which provides the necessary negative spherical aberration, but also leaves some residual four-fold astigmatism that is corrected by the third central round-beam octupole. This design is loosely based on Scherzer's original design that used cylindrical lenses (Scherzer, 1947). Although this design corrects the third-order C_s , it actually worsens the fifth-order aberrations. Nonetheless, it has been extremely successful and productive scientifically. A more recent corrector design from Nion (Krivanek et al., 2003) allows correction of the fifth-order aberrations also. Again it is based on third-order correction by three octupoles, but with a greater number of quadrupole layers, which can provide control of the fifth-order aberrations. This more complicated corrector is being incorporated into an entirely new STEM column designed to optimize performance with aberration correction.

An alternative corrector design that is suitable for both HRTEM and STEM use has been developed by CEOS (Haider et al., 1998a). It is based on a design by Shao (1988) and further developed by Rose (1990).

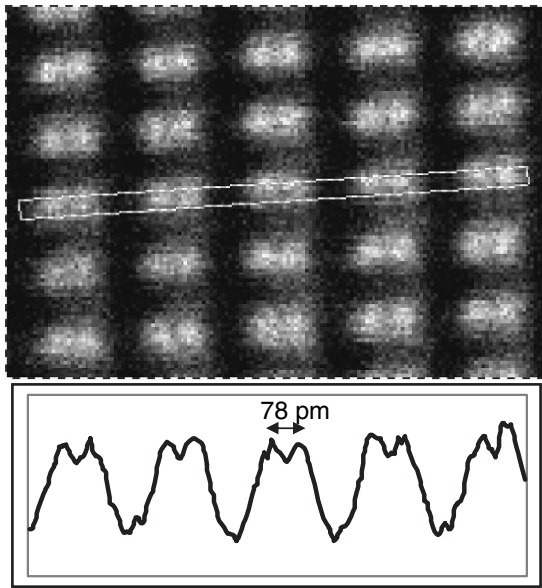


Figure 2–30. An ADF STEM image of Si<112> recorded using a 300-kV VG Microscopes HB603U STEM fitted with a Nion aberration corrector. The 78 pm spacing of the atomic columns in this projection is well resolved, as can be seen in the intensity profile plot from the region indicated.

It includes two sextupole lenses with four additional round lens coupling lenses. The primary aberration of a sextupole is three-fold astigmatism, but if the sextupole is extended in length it can also generate negative, round spherical aberration. If two sextupoles are used and suitably coupled by round lenses, the three-fold astigmatism from each of them can cancel, resulting in pure, negative spherical aberration. The optical coupling between the sextupole layers and the objective lens means that the off-axial aberrations are also canceled, which allows the use of this kind of corrector for HRTEM imaging in addition to STEM imaging.

Aberration correction in STEM has already produced high impact results. The improvement in resolution has been dramatic with a resolution as high as $0.78 \text{ \AA}$ and information transfer to $0.6 \text{ \AA}$ being demonstrated (Figure 2–30) (Nellist et al., 2004). The ability to image at atomic resolution along different orientations has allowed a full, three-dimensional reconstruction of a heterointerface to be determined (Falke et al., 2004). Spectroscopy of single atoms of impurities in a doped crystalline matrix has been demonstrated (Varela et al., 2004). Clearly, aberration correction in STEM is now well established and will become more commonplace.

11. Conclusions

In this chapter we have tried to describe the range of techniques available in a STEM, the principles behind those techniques, and some examples of applications. Naturally there are many similarities

between the CTEM and the STEM, and some of the imaging modes are equivalent. Certain techniques in STEM, however, are unique, and have particular strengths. In particular, STEM is being used for ADF and electron energy-loss spectroscopy. The ADF imaging mode is important because it is an incoherent imaging mode and shows atomic number (Z) contrast. The incoherent nature of ADF imaging makes the images simpler to interpret in terms of the atomic structure under observation, and we have described how it has been used to determine atomic structures at interfaces, even correcting earlier structural analyses by HRTEM. The CTEM cannot efficiently provide an incoherent imaging mode. The spatial resolution of STEM can also be applied to composition analysis through EELS, and atomic resolution and single-atom sensitivity are both now being demonstrated. Not only can EELS provide compositional information, but analysis of the fine structure of spectra can reveal information on the bonding between materials.

The capabilities listed above, combined with the availability of combination CTEM/STEM instruments, has dramatically increased the popularity of STEM. For many years, the only high-resolution STEM instruments available were dedicated STEM instruments with a CFEG. These machines were designed as high-end research machines and they tended to be operated by experts who could devote time to their operation and maintenance. Modern CTEM/STEM instruments are much more user friendly, and the Schottky gun system usually found on such machines is easier to operate.

We have also discussed some of the technical details of the electron optics and resolution-limiting factors, which raises the question of where the development of STEM instrumentation is likely to go in the future. Clearly spherical aberration correction is already having a major impact on STEM performance, and the fraction of STEM instruments fitted with correctors is bound to increase. The benefits of aberration correction are not only the increased spatial resolution, but also the dramatically improved beam current and also the possibility of creating more room around the sample for *in situ* experiments. The increased beam current already allows fast mapping of spectrum images with sufficient signal to noise for fitting of fine-structure changes. Much faster elemental mapping should become possible, with acquisition rates perhaps reaching 1000 spectra/s, which would allow a 256 by 256 pixel spectrum image to be recorded in around 1 min. To achieve this goal, however, requires further development of the spectra acquisition instrumentation, such as the CCD camera and probe scan controller. With aberration correction now available it is often found that the STEM performance is being limited by other aspects of the instrumentation. It is now an excellent time for a reevaluation of the design of electron-optical columns to be used for aberration-corrected STEM. Already a number of manufacturers are launching new columns and the STEM community is eagerly awaiting new data demonstrating their performance. New columns also allow the inclusion of *in situ* experiments, and we are likely to see columns fitted with scanning probe systems, nanomanipulators, or environmental cells. Environ-

mental cells, for example, would add to the STEM's existing strengths in the imaging of dispersed catalysts by allowing samples to be viewed while being dosed with active gases.

The other important technical development currently being introduced into STEM instruments is monochromation. There are two motivations for this development. Obviously a more monochromated beam will lead to improved energy resolution in EELS. Defect states in band gaps would become visible in the low-loss spectrum and core-loss fine structure would show greater detail. Furthermore, Schottky guns have a greater energy spread in the beam (about 0.8 eV) compared to a CFEG (about 0.3 eV), so there is a strong motivation to fit Schottky systems with a monochromator to improve their energy resolution. With a spherical aberration-corrected machine, the spatial resolution is then limited by chromatic aberration, which will be worse for a Schottky gun, hence a spatial resolution benefit from monochromation. An important consequence of monochromation, however, is that it reduces the brightness of the electron gun. Nonetheless, it is possible to produce monochromated atomic-scale probes (Mayer et al., 2007). Starting with a gun that is brighter and has an intrinsically narrower energy spread, such as a CFEG, obviously has strong benefits for STEM. Time will tell whether the CFEG will become more popular again. Nevertheless, it is clear that STEM itself has a very strong future in the imaging and analysis of materials.

References

- Abe, E., Pennycook, S. J. and Tsai, A. P. (2003). Direct observation of a local thermal vibration anomaly in a quasicrystal. *Nature* **421**, 347.
- Allen, L. J., Findlay, S. D., Lupini, A. R., Oxley, M. P. and Pennycook, S. J. (2003a). Atomic-resolution electron energy loss spectroscopy imaging in aberration corrected scanning transmission electron microscopy. *Phys. Rev. Lett.* **91**, 105503.
- Allen, L. J., Findlay, S. D., Oxley, M. P. and Rossouw, C. J. (2003b). Lattice-resolution contrast from a focused coherent electron probe. Part I. *Ultramicroscopy* **96**, 47.
- Amali, A. and Rez, P. (1997). Theory of lattice resolution in high-angle annular dark-field images. *Microscopy and Microanalysis* **3**, 28.
- Anderson, S. C., Birkeland, C. R., Anstis, G. R. and Cockayne, D. J. H. (1997). An approach to quantitative compositional profiling at near-atomic resolution using high-angle annular dark field imaging. *Ultramicroscopy* **69**, 83.
- Batson, P. E. (1993). Simultaneous STEM imaging and electron energy-loss spectroscopy with atomic-column sensitivity. *Nature* **366**, 727.
- Batson, P. E. (2000). Structural and electron characterisation of a dissociated 60 degrees dislocation in GeSi. *Phys. Rev. B* **61**, 16633.
- Batson, P. E., Dellby, N. and Krivanek, O. L. (2002). Sub-ångstrom resolution using aberration corrected electron optics. *Nature* **418**, 617.
- Black, G. and Linfoot, E. H. (1957). Spherical aberration and the information limit of optical images. *Proc. R. Soc. Lond. A* **239**, 522.
- Bonnet, N., Brun, N. and Colliex, C. (1999). Extracting information from sequences of spatially resolved EELS spectra using multivariate statistical analysis. *Ultramicroscopy* **77**, 97.
- Born, M. and Wolf, E. (1980). *Principles of Optics* (Oxford, Pergamon Press).

- Brink, H. A., Barfels, M. M. G., Burgner, R. P. and Edwards, B. N. (2003). A sub-50 meV spectrometer and energy-filter for use in combination with 200 kV monochromated (S)TEMs. *Ultramicroscopy* **96**, 367.
- Brown, L. M. (1981). Scanning transmission electron microscopy: microanalysis for the microelectronic age. *J. Phys. F* **11**, 1.
- Browning, N. D., Chisholm, M. F. and Pennycook, S. J. (1993). Atomic-resolution chemical analysis using a scanning transmission electron microscope. *Nature* **366**, 143.
- Brydson, R. (2001). *Electron Energy Loss Spectroscopy* (Oxford, BIOS).
- Chisholm, M. F., Maiti, A., Pennycook, S. J. and Pantelides, S. T. (1998). Atomic configuration and energetics of arsenic impurities in a silicon grain boundary. *Phys. Rev. Lett.* **81**, 132.
- Cosgriff, E. C., Oxley, M. P., Allen, L. J. and Pennycook, S. J. (2005). The spatial resolution of imaging using core-loss spectroscopy in the scanning transmission electron microscope. *Ultramicroscopy* **102**, 317.
- Cowley, J. M. (1969). Image contrast in a transmission scanning electron microscope. *Appl. Phys. Lett.* **15**, 58.
- Cowley, J. M. (1976). Scanning transmission electron microscopy of thin specimens. *Ultramicroscopy* **2**, 3.
- Cowley, J. M. (1978). Electron microdiffraction. *Adv. Electronics & Electron Phys.* **4**, 6, 1.
- Cowley, J. M. (1979a). Adjustment of a STEM instrument by Use of Shadow Images. *Ultramicroscopy* **4**, 413.
- Cowley, J. M. (1979b). Coherent interference in convergent-beam electron diffraction & shadow imaging. *Ultramicroscopy* **4**, 435.
- Cowley, J. M. (1981). Coherent interference effects in SIEM and CBED. *Ultramicroscopy* **7**, 19.
- Cowley, J. M. (1986). Electron-diffraction phenomena observed with a high-resolution STEM Instrument. *J. Electron Microsc. Tech.* **3**, 25.
- Cowley, J. M. (1990). *Diffraction Physics* (Amsterdam, North-Holland).
- Cowley, J. M. (1992). *Electron Diffraction Techniques* (Oxford, OUP).
- Cowley, J. M. and Disko, M. M. (1980). Fresnel diffraction in a coherent convergent electron beam. *Ultramicroscopy* **5**, 469.
- Crewe, A. V. (1980). The physics of the high-resolution STEM. *Rep. Prog. Phys.* **43**, 621.
- Crewe, A. V., Eggenberger, D. N., Wall, J. and Welter, L. M. (1968a). Electron gun using a field emission source. *Rev. Sci. Instr.* **39**, 576.
- Crewe, A. V., Wall, J. and Welter, L. M. (1968b). A high-resolution scanning transmission electron microscope. *J. Appl. Phys.* **39**, 5861.
- Crewe, A. V., Wall, J. and Langmore, J. (1970). Visibility of single atoms. *Science* **168**, 1338.
- Dellby, N., Krivanek, O. L., Nellist, P. D., Batson, P. E. and Lupini, A. R. (2001). Progress in aberration-corrected scanning transmission electron microscopy. *J. Electron Microsc.* **50**, 177.
- Dinges, C., Berger, A. and Rose, H. (1995). Simulation of TEM images considering phonon and electron excitations. *Ultramicroscopy* **60**, 49.
- Donald, A. M. and Craven, A. J. (1979). A study of grain boundary segregation in Cu-Bi alloys using STEM. *Philosophical Magazine* **A39**, 1.
- Egerton, R. F. (1996). *Electron Energy-Loss Spectroscopy in the Electron Microscope* (New York, Plenum Press).
- Falke, U., Bleloch, A. L., Falke, M. and Teichert, S. (2004). Atomic structure of a (2x1) Reconstructed NiSi₂/Si(001) Interface. *Phys. Rev. Lett.* **92**, 116103.
- Fertig, J. and Rose, H. (1981). Resolution and contrast of crystalline objects in high-resolution scanning transmission electron microscopy. *Optik* **59**, 407.

- Findlay, S. D., Allen, L. J., Oxley, M. P. and Rossouw, C. J. (2003). Lattice-resolution contrast from a focused coherent electron probe. Part II. *Ultramicroscopy* **96**, 65.
- Gabor, D. (1948). A new microscope principle. *Nature* **161**, 777.
- Haider, M., Rose, H., Uhlemann, S., Schwan, E., Kabius, B. and Urban, K. (1998a). A spherical-aberration-corrected 200 kV transmission electron microscope. *Ultramicroscopy* **75**, 53.
- Haider, M., Uhlemann, S., Schwan, E., Rose, H., Kabius, B. and Urban, K. (1998b). Electron microscopy image enhanced. *Nature* **392**, 768.
- Hartel, P., Rose, H. and Dinges, C. (1996). Conditions and reasons for incoherent imaging in STEM. *Ultramicroscopy* **63**, 93.
- Hillyard, S., Loane, R. F. and Silcox, J. (1993). Annular dark-field imaging: resolution and thickness effects. *Ultramicroscopy* **49**, 14.
- Hillyard, S. and Silcox, J. (1993). Thickness effects in ADF STEM zone axis images. *Ultramicroscopy* **52**, 325.
- Hirsch, P., Howie, A., Nicholson, R., Pashley, D. W. and Whelan, M. J. (1977). *Electron Microscopy of Thin Crystals* (Malabar, Krieger).
- Hoppe, W. (1969a). Beugung im inhomogenen primärstrahlwellenfeld. I. Prinzip einer Phasenmessung von Elektronenbeugungsinterferenzen. *Acta Cryst. A* **25**, 495.
- Hoppe, W. (1969b). Beugung im inhomogenen primärstrahlwellenfeld. III. Amplituden- und Phasenbestimmung bei unperiodischen Objekten. *Acta Cryst. A* **25**, 508.
- Hoppe, W. (1982). Trace structure analysis, ptychography, phase tomography. *Ultramicroscopy* **10**, 187.
- Howie, A. (1979). Image contrast and localised signal selection techniques. *J. Microsc.* **117**, 11.
- James, E. M. and Browning, N. D. (1999). Practical aspects of atomic resolution imaging and analysis in STEM. *Ultramicroscopy* **78**, 125.
- Jesson, D. E. and Pennycook, S. J. (1993). Incoherent imaging of thin specimens using coherently scattered electrons. *Proc. R. Soc. Lond. A* **441**, 261.
- Jesson, D. E. and Pennycook, S. J. (1995). Incoherent imaging of crystals using thermally scattered electrons. *Proc. R. Soc. Lond. A* **449**, 273.
- Kirkland, A. I., Saxton, W. O., Chau, K. L., Tsuno, K. and Kawasaki, M. (1995). Superresolution by aperture synthesis—Tilt series reconstruction in CTEM. *Ultramicroscopy* **57**, 355.
- Kirkland, E. J., Loane, R. F. and Silcox, J. (1987). Simulation of annular dark field STEM images using a modified multislice method. *Ultramicroscopy* **23**, 77.
- Kohl, H. and Rose, H. (1985). Theory of image formation by inelastically scattered electrons in the electron microscope. *Adv. Electronics & Electron Phys.* **65**, 173.
- Krivanek, O. L., Dellby, N. and Lupini, A. R. (1999). Towards sub-Å electron beams. *Ultramicroscopy* **78**, 1.
- Krivanek, O. L., Nellist, P. D., Dellby, N., Murfitt, M. F. and Szilagy, Z. (2003). Towards sub-0.5 Å electron beams. *Ultramicroscopy* **96**, 229.
- Lin, J. A. and Cowley, J. M. (1986). Reconstruction from In-Line Electron Holograms by Digital Processing. *Ultramicroscopy* **19**, 179.
- Loane, R. F., Xu, P. and Silcox, J. (1991). Thermal vibrations in convergent-beam electron diffraction. *Acta Cryst. A* **47**, 267.
- Loane, R. F., Xu, P. and Silcox, J. (1992). Incoherent imaging of zone axis crystals with ADF STEM. *Ultramicroscopy* **40**, 121.
- Lupini, A. R. (2001). Cavendish laboratory, The University of Cambridge. PhD Thesis.

- Lupini, A. R. and Pennycook, S. J. (2003). Localisation in elastic and inelastic scattering. *Ultramicroscopy* **96**, 313.
- Lyman, C. E., Goldstein, J. I., Williams, D. B., Ackland, D. W., Von Harrach, S., Nicholls, A. W. and Statham, P. J. (1994). High performance X-ray detection in a new analytical electron microscope. *J. Microsc.* **176**, 85.
- *Mayer, J., Houben, L., Lopatin, S., Luysberg, M., and Thust, A. (2007). Performance and applications of the aberration corrected TEM and STEM instruments at the Ernst Ruska-Centre. In Proceedings of microscopy and microanalysis 2007, *Microscopy and Microanalysis* **13(Suppl 2)**, 128.
- McGibbon, A. J., Pennycook, S. J. and Angelo, J. E. (1995). Direct observation of dislocation core structures in CdTe/GaAs(001). *Science* **269**, 519.
- McGibbon, M. M., Browning, N. D., Chisholm, M. F., McGibbon, A. J., Pennycook, S. J., Ravikumar, V. and Dravid, V. P. (1994). Direct determination of grain boundary atomic structures in SrTiO₃. *Science* **266**, 102.
- Mitsuishi, K., Takeguchi, M., Yasuda, H. and Furuya, K. (2001). New scheme for calculation of annular dark-field STEM image including both elastically diffracted and TDS wave. *J. Electron Microsc.* **50**, 157.
- Möbus, G. and Nufer, S. (2003). Nanobeam propagation and imaging in a FEGTEM/STEM. *Ultramicroscopy* **96**, 285.
- Mory, C., Colliex, C. and Cowley, J. M. (1987). Optimum defocus for STEM imaging and microanalysis. *Ultramicroscopy* **21**, 171.
- Mullejans, H., Bleloch, A. L., Howie, A. and Tomita, M. (1993). Secondary-electron coincidence detection and time-of-flight spectroscopy. *Ultramicroscopy* **52**, 360.
- Muller, D. A., Edwards, B., Kirkland, E. J. and Silcox, J. (2001). Simulation of thermal diffuse scattering including a detailed phonon dispersion curve. *Ultramicroscopy* **86**, 371.
- Muller, D. A. and Silcox, J. (1995). Delocalisation in inelastic imaging. *Ultramicroscopy* **59**, 195.
- Muller, D. A., Tzou, Y., Raj, R. and Silcox, J. (1993). Mapping sp² and sp³ states of carbon at sub-nanometre spatial resolution. *Nature* **366**, 725.
- Nakanishi, N., Yamazaki, T., Rečnik, A., Čeh, M., Kawasaki, M., Watanabe, K. and Shiojiri, M. (2002). Retrieval process of high-resolution HAADF-STEM images. *J. Electron Microsc.* **51**, 383.
- Nellist, P. D., McCallum, B. C. and Rodenburg, J. M. (1995). Resolution beyond the “information limit” in transmission electron microscopy. *Nature* **374**, 630.
- Nellist, P. D., Dellby, N., Krivanek, O. L., Murfitt, M. F., Szilagy, Z., Lupini, A. R. and Pennycook, S. J. (2003). Towards sub-0.5 angstrom beams through aberration corrected STEM. Proceedings of EMAG2003, Oxford, (IOP Conf. Ser. 179).
- Nellist, P. D., Chisholm, M. F., Dellby, N., Krivanek, O. L., Murfitt, M. F., Szilagy, Z., Lupini, A. R., Borisevich, A., Sides, W. H. J. and Pennycook, S. J. (2004). Direct sub-angstrom imaging of a crystal lattice. *Science* **305**, 1741.
- Nellist, P. D. and Pennycook, S. J. (1996). Direct imaging of the atomic configuration of ultradispersed catalysts. *Science* **274**, 413.
- Nellist, P. D. and Pennycook, S. J. (1998a). Accurate structure determination from image reconstruction in ADF STEM. *J. Microsc.* **190**, 159.
- Nellist, P. D. and Pennycook, S. J. (1998b). Subangstrom Resolution by Underfocussed Incoherent Transmission Electron Microscopy. *Phys. Rev. Lett.* **81**, 4156.
- Nellist, P. D. and Pennycook, S. J. (1999). Incoherent imaging using dynamically scattered coherent electrons. *Ultramicroscopy* **78**, 111.

- Nellist, P. D. and Pennycook, S. J. (2000). The principles and interpretation of annular dark-field Z-contrast imaging. *Adv. Imaging & Electron Phys.* **113**, 148.
- Nellist, P. D. and Rodenburg, J. M. (1994). Beyond the conventional information limit: the relevant coherence function. *Ultramicroscopy* **54**, 61.
- Nellist, P. D. and Rodenburg, J. M. (1998). Electron ptychography I: experimental demonstration beyond the conventional resolution limits. *Acta Cryst. A* **54**, 49.
- Pennycook, S. J. (1989). Z-contrast STEM for materials science. *Ultramicroscopy* **30**, 58.
- Pennycook, S. J. (2002). Structure determination through Z-contrast microscopy. *Adv. Imaging Electron Phys.* **123**, 173.
- Pennycook, S. J., Brown, L. M. and Craven, A. J. (1980). Observation of cathodoluminescence at single dislocations by STEM. *Phil. Mag. A* **41**, 589.
- Pennycook, S. J. and Jesson, D. E. (1990). High-resolution incoherent imaging of crystals. *Phys. Rev. Lett.* **64**, 938.
- Perovic, D. D., Rossouw, C. J. and Howie, A. (1993). Imaging elastic strain in high-angle annular dark-field scanning transmission electron microscopy. *Ultramicroscopy* **52**, 353.
- Rafferty, B. and Brown, L. M. (1998). Direct and indirect transitions in the region of the band gap using EELS. *Phys. Rev. B* **58**, 10326.
- Rafferty, B., Nellist, P. D. and Pennycook, S. J. (2001). On the origin of transverse incoherence in Z-contrast STEM. *J. Electron Microsc.* **50**, 227.
- Rafferty, B. and Pennycook, S. J. (1999). Towards column-by-column spectroscopy. *Ultramicroscopy* **78**, 141.
- Rayleigh, Lord (1896). On the theory of optical images with special reference to the microscope. *Philosophical Magazine* (5) **42**, 167.
- Reed, S. J. B. (1982). The single-scattering model and spatial-resolution in X-ray analysis of thin foils. *Ultramicroscopy* **7**, 405.
- Ritchie, R. H. and Howie, A. (1988). Inelastic scattering probabilities in scanning transmission electron microscopy. *Phil. Mag. A* **58**, 753.
- Rodenburg, J. M. and Bates, R. H. T. (1992). The theory of super-resolution electron microscopy via wigner-distribution deconvolution. *Phil. Trans. R. Soc. Lond. A* **339**, 521.
- Rodenburg, J. M., McCallum, B. C. and Nellist, P. D. (1993). Experimental tests on double-resolution coherent imaging via STEM. *Ultramicroscopy* **48**, 303.
- Ronchi, V. (1964). 40 Years of history of grating interferometer. *Applied Optics* **3**, 437.
- Rose, H. (1990). Outline of a spherically corrected semiaplanatic medium-voltage transmission electron microscope. *Optik* **85**, 19.
- Rossouw, C. J., Allen, L. J., Findlay, S. D. and Oxley, M. P. (2003). Channelling effects in atomic resolution STEM. *Ultramicroscopy* **96**, 299.
- Scherzer, O. (1936). Über einige Fehler von Elektronenlinsen. *Zeit Physik* **101**, 593.
- Scherzer, O. (1947). Sphärische und chromatische korrektur von elektronenlinsen. *Optik* **2**, 114.
- Shao, Z. (1988). On the fifth order aberration in a sextupole corrected probe forming system. *Rev. Sci. Instr.* **59**, 2429.
- Sohlberg, K., Rashkeev, S., Borisevich, A. Y., Pennycook, S. J. and Pantelides, S. T. (2004). Origin of anomalous Pt-Pt distances in the Pt/alumina catalytic system. *ChemPhysChem* **5**, 1893.
- Spence, J. C. H. (1988). *Experimental High-Resolution Electron Microscopy* (New York, OUP).

- Spence, J. C. H. (1992). Convergent-beam nanodiffraction, in-line holography and coherent shadow imaging. *Optik* **92**, 57.
- Spence, J. C. H. (1998a). Crystal structure determination by direct inversion of dynamical microdiffraction patterns. *J. Microsc.* **190**, 214.
- Spence, J. C. H. (1998b). Direct inversion of dynamical electron diffraction patterns to structure factors. *Acta Cryst. A* **54**, 7.
- Spence, J. C. H. and Cowley, J. M. (1978). Lattice imaging in STEM. *Optik* **50**, 129.
- Spence, J. C. H. and Zuo, J. M. (1992). *Electron Microdiffraction* (New York, Plenum Press).
- Suenaga, K., Tencé, M., Mory, C., Colliex, C., Kato, H., Okazaki, T., Shinohara, H., Hirahara, K., Bandow, S. and Iijima, S. (2000). Element-selective single atom imaging. *Science* **290**, 2280.
- Swanson, L. W. and Crouser, L. C. (1967). Total energy distribution of field-emitted electrons and single-plane work functions for tungsten. *Phys. Rev.* **163**, 622.
- Treacy, M. M. J., Howie, A. and Wilson, C. J. (1978). Z contrast imaging of platinum and palladium catalysts. *Phil. Mag. A* **38**, 569.
- Varela, M., Findlay, S. D., Lupini, A. R., Christen, H. M., Borisevich, A. Y., Dellby, N., Krivanek, O. L., Nellist, P. D., Oxley, M. P., Allen, L. J. and Pennycook, S. J. (2004). Spectroscopic imaging of single atoms within a bulk solid. *Phys. Rev. Lett.* **92**, 095502.
- Voyles, P. M., Muller, D. A., Grazul, J. L., Citrin, P. H. and Gossmann, H. J. L. (2002). Atomic-scale imaging of individual dopant atoms and clusters in highly n-type bulk Si. *Nature* **416**, 826.
- Voyles, P. M., Muller, D. A. and Kirkland, E. J. (2004). Depth-dependent imaging of individual dopant atoms in silicon. *Microscopy and Microanalysis* **10**, 291.
- Wade, R. H. (1992). A brief look at imaging and contrast theory. *Ultramicroscopy* **46**, 145.
- Watanabe, M. and Williams, D. B. (1999). Atomic-level detection by X-ray microanalysis in the analytical electron microscope. *Ultramicroscopy* **78**, 89.
- Williams, D. B. and Carter, C. B. (1996). *Transmission Electron Microscopy* (New York, Plenum Press).
- Yamazaki, T., Kawasaki, M., Watanabe, K., Hashimoto, I. and Shiojiri, M. (2001). Artificial bright spots in atomic-resolution high-angle annular dark-field STEM images. *J. Electron Microsc.* **50**, 517.
- Zach, J. and Haider, M. (1995). Correction of spherical and chromatic aberration in a low-voltage SE. *Optik* **98**, 112.
- Zeitler, E. and Thomson, M. G. R. (1970). Scanning transmission electron microscopy. *Optik* **31**, 258.
- Ziegler, A., Idrobo, J. C., Cinibulk, M. K., Kisielowski, C., Browning, N. D. and Ritchie, R. O. (2004). Interface structure and atomic bonding characteristics in silicon nitride ceramics. *Science* **306**, 1768.

Scanning Electron Microscopy

Rudolf Reichelt

1 Introduction

The earliest historical contribution to the idea of a scanning electron microscope (SEM) was probably made by H. Stintzing in 1927 in a German patent application (Stintzing, 1927). In his patent he proposed irradiating a sample with a narrowly collimated beam (light, X-ray, corpuscles) and moving the sample transversely to the beam. The magnitude of interaction between beam and sample was to be measured by a sufficiently sensitive recording device, to be amplified and then displayed on an electron tube. This idea aimed to determine the size of small particles not accessible to light microscopy. However, the method proposed was unable to generate a magnified image.

The first electron beam scanner capable of producing an image of the surface of a bulk sample with the emitted secondary electrons (SE) was developed by H. Knoll in 1935 (Knoll, 1935). In this instrument a focused electron beam was scanned across the sample surface by magnetic deflection line by line. The generated SE were converted into an electronic signal, which was amplified and then used to modulate the brightness of an electron tube. Both the electron beam of the scanner and of the display tube were scanned synchronously across the same distance by perpendicular pairs of deflection coils, thus the electron beam scanner at that time possessed unit magnification. The contrast in the SE image was due to the varying yield of SE caused by the different local chemical composition of the sample surface. However, sub-microscopic resolution with an SEM was first obtained by M. von Ardenne using the transmission mode [so-called scanning transmission electron microscope (STEM)] (von Ardenne, 1938). The early development of electron microscopy up to 1940 has been described extensively by E. Ruska (1979). While the development of SEM in Europe was interrupted by World War II, the idea of SEM was used by Zworykin, Hillier, and Snyder in the United States for the construction of such an

instrument (Zworykin et al., 1942). After World War II experiments with scanning electron microscopes started in England and France. Since 1948 C.W. Oatley from the Engineering Laboratory of the University of Cambridge directed intensive development in that field, which led to the first commercial scanning electron microscope available in 1965. Pioneering work in the improvement of instrumentation (e.g., electron sources, electron optics, detectors, signal processing), the investigation of electron-specimen interaction, elucidation of fundamental contrast mechanisms, and development of methods for the preparation of samples were done in the 1960s and 1970s (for books, proceedings of SEM conferences, and reviews see, e.g., Oatley, 1972; Ohnsorge and Holm, 1973; Holt et al., 1974; Wells, 1974, 1975; Hayat, 1974; Goldstein and Yakowitz, 1975; Reimer and Pfefferkorn, 1973, 1977; Pfefferkorn, 1968; Johari, 1968; Reimer, 1978). At the same time the SEM was applied extensively in very different fields for *imaging of surfaces*, for the local *crystallographic characterization* of polycrystalline materials by electron backscattered diffraction (EBSD), and in combination with X-ray microanalytical equipment such as an energy dispersive spectrometer (EDX) or a wavelength dispersive spectrometer (WDX) for the *local element analysis* of specimens (for books or proceedings of SEM conferences see, e.g., Heywood, 1971; Fujita et al., 1971; Thornton, 1972; Troughton and Donaldson, 1972; Siegel and Beaman, 1975; Chandler, 1978; Revel et al., 1983; Newbury et al., 1987; Wetzig and Schulze, 1995; Reed, 1996; Goldstein et al., 1984, 2003; Pfefferkorn, 1968; Johari, 1968).

A significant improvement in SEM instrumentation was made by the development of the field emission scanning electron microscope (FESEM) that became commercially available in the 1980s. In the FESEM instead of a thermionic gun a field emission gun (FEG) is used for electron beam generation, which allows the formation of an electron probe with a diameter of about 0.5 nm. Together with further improvements in electron optics and electron detectors on the one hand and in specimen preparation on the other hand, high-resolution imaging by FESEM became feasible. Additionally, the FESEM enables specimens to be imaged at low acceleration voltages, i.e., below 5 kV, also at high resolution. The operation of the FESEM at low acceleration voltages opens new avenues for interesting applications in the characterization of surfaces (for reviews of specific aspects of the instrumentation, image formation, and application of FESEM at conventional or low acceleration voltages, respectively, see, e.g., Lyman et al., 1990; Reimer, 1993; Joy, 1995; Sawyer and Grubb, 1996; Müllerova and Frank, 2003).

A very interesting step forward in the instrumentation was the development of a so-called environmental scanning electron microscope (ESEM) pushed in the 1980s in particular by G.D. Danilatos. Investigations of specimens using secondary or backscattered electrons for imaging, in SEM and FESEM restricted to high vacuum, can be performed in the ESEM in a low vacuum at a pressure of about 10 up to a few thousands Pa. Obviously, this is of great interest for samples that consist of materials or may contain dirt or fluids, respectively,

having a partial pressure in the low vacuum range mentioned. Typical examples are water or oil containing natural specimens. Moreover, electric insulators can be imaged without prior conductive coating by ESEM in low vacuum without significant electric charging artifacts. The ESEM (ESEM is used as a trade name by the manufacturers Electroscan/Philips/FEI) became commercially available in 1987 (for reviews of specific aspects of the instrumentation, image formation, and application of ESEM see, e.g., Danilatos, 1988, 1990). Since about the second half of the 1990s other manufacturers also offer commercial SEMs for low vacuum operation, usually referred to as low vacuum SEM, variable pressure SEM, natural SEM, etc.; however, these instruments are restricted typically to a maximum pressure of about 300 Pa and allow imaging only with backscattered electrons (BSE).

Modern high-resolution FESEMs have at an electron energy of 30 keV a specified resolution power in the SE mode in the range of 0.5–1 nm, which corresponds to about the size of a small molecule. It marks the smallest size of a structure accessible on one hand. Working at the smallest magnification of an SEM on the other hand allows imaging of visible structures as large as about 1 mm, i.e., high-resolution SEMs cover a wide range of six orders of magnitude for the structural characterization of surfaces. A further advantage of the SEM is the simultaneous acquisition of different signals generated by the local interaction of the beam electrons with the specimen. Each of these signals, e.g., SE and BSE, carries different information about the sample, thus an extensive multidimensional data set about an area of interest can be obtained by one scan line-by-line across this area. The recording time may vary from a few seconds only to about 1 min depending on the strength of the signal and the signal-to-noise ratio required.

The scanning electron microscopy is now a well-established method for the characterization of surfaces in ultrahigh vacuum (UHV), high vacuum (HV), and low vacuum (LV) in many different fields. Clearly, it is not possible to mention all of them, however, the main fields are certainly the materials (metals, alloys, ceramics, glasses) and surface sciences, semiconductor research and industry, life sciences, and miscellaneous sciences such as polymer and food research, mineralogy, geology, the oil industry, and archaeology.

In addition to scanning electron microscopy some other surface-sensitive methods such as atomic force microscopy, scanning tunneling microscopy, and photoelectron microscopy are described in the book "Science of Microscopy." It is certainly of particular interest to see in detail how these surface-sensitive methods complement each other, what specific advantages they offer, and how they compare with SEM.

1.1 Abbreviations

ADC	Analog-to-digital-converter
AE	Auger electrons
AES	Auger electron spectroscopy

AFM	Atomic force microscope/microscopy
ASEM	Atmospheric scanning electron microscope/ microscopy
CC	Charge collection
CE	Collection efficiency
CFEG	Cold field emission gun
CL	Cathodoluminescence
CRT	Cathode ray tube
CSEM	Conventional high vacuum scanning electron microscope/microscopy
EBIC	Electron beam-induced current
EBSD	Electron backscattered diffraction
ECP	Electron channeling pattern
EDX	Energy dispersive X-ray
ESEM	Environmental scanning electron microscope/microscopy
ET	Everhart–Thornley
ETD	Everhart–Thornley detector
FEG	Field emission gun
FESEM	Field emission scanning electron microscope/microscopy
FLM	Fluorescence light microscopy
HFEG	Hot field emission gun
HRSEM	High-resolution scanning electron microscope/microscopy
HV	High vacuum
LCD	Liquid crystal display
LEED	Low-energy electron diffraction
LP	Light pipe
LV	Low vacuum
LVSEM	Low-voltage scanning electron microscope/microscopy
MCP	Microchannel plate
MEMS	Microelectromechanical systems
NEMS	Nanoelectromechanical systems
OIM	Orientation imaging microscopy
PC	Personal computer

SAM	Scanning Auger microscopy
SE	Secondary electrons
SEC	Schottky emission cathode
SEM	Scanning electron microscopes
SLEEM	Scanning low-energy electron microscope/microscopy
S/N	Signal to noise
SNR	Signal-to-noise ratio
STEM	Scanning transmission electron microscope/ microscopy
STM	Scanning tunneling microscope/microscopy
UHV	Ultrahigh vacuum
UV	Ultraviolet
VLVSEM	Very low-voltage scanning electron microscopy
WD	Work distance
WDX	Wavelength dispersive X-ray
YAG	Yttrium-aluminum-garnet
YAP	Yttrium-aluminum-perovskite
3D	Three dimensional

1.2 Symbols

A	Atomic weight
C_c	Chromatic aberration constant
C_s	Spherical aberration constant
d	Lattice-plane spacing
d_p	Final electron probe diameter d_p at the specimen
d_{pe}	Effective final electron probe diameter
$d_{p,min}$	Minimum effective electron probe diameter
d_c	Diameter of the error disc of chromatic aberration
d_f	Diameter of the diffraction caused Airy disc
d_s	Diameter of the error disc of spherical aberration
d_o	Diameter d_o of the first crossover
E	Electron energy
E_{AE}	Energy of Auger electrons
E_o	Electron energy of the beam electrons

E_{BSE}	Energy of backscattered electrons
E_2	Beam electron energy where the number of incoming and emitted electrons is equal
E_{X}	X-ray energy
E_{exm}	Mean energy per excitation of an electron hole pair
F	Flux of gas molecules
f	Focal length of a lens
I_{CC}	Charge-collection current
I_{p}	Electron probe current
$I_{\text{p,max}}$	Maximum electron probe current
j_{c}	Current density for thermionic emission
J	Mean ionization potential
L	Electron path length in gas
N	Number of atoms per unit volume
N_{A}	Avogadro's number
$N \cdot I$	Ampere-windings of an electromagnetic lens
n_{m}	Mean number of electron hole pairs
M	Magnification
p	Object distance to the center of the lens
p_{g}	Pressure in the specimen chamber of the SEM
p_{x}	Parallax
q	Image distance to the center of the lens
R	Electron range
R_{B}	Bethe range
T	Temperature
t	Specimen thickness
Z	Atomic number
α_{p}	Semiangle of the convergent impinging electron probe at the specimen
α_{opt}	Optimum value of α_{p}
ΔE	Energy spread of the beam electrons
Φ	Work function
θ	Angle of incidence of the electron beam
Θ	Tilt angle of the specimen
φ	Scattering angle

ϑ	Bragg angle
α_o	Semiangle of divergence of the first crossover
δ	Secondary electron yield
η	Backscattered electron coefficient
λ	Wavelength of the beam electrons
λ_x	Wavelength of X-ray
Λ	Mean free path for electron scattering in gas
Λ_{el}	Mean free path for elastic scattering
Λ_{in}	Mean free path for inelastic scattering
σ	Electron scattering cross section
σ_{el}	Elastic electron scattering cross section
σ_{in}	Inelastic electron scattering cross section
σ_g	Total electron scattering cross section of gas molecule
ε	Complex dielectric constant
ε_0	Dielectric constant
ν	Frequency of radiation
ω	Fluorescence yield of X-rays

2 Conventional Scanning Electron Microscopy

The principle of a scanning electron microscope is shown schematically in Figure 3-1. The two major parts are the microscope column and the electronics console. The microscope column consists of the electron gun (with the components cathode, Wehnelt cylinder, anode), one or two condenser lenses, two pairs of beam deflection coils (scan coils for X, Y deflection), the objective lens, and some apertures. In the specimen chamber at the lower end of the microscope column are located the specimen stage and the detectors for the different signals generated by the electron-specimen interaction. The microscope column and the specimen chamber are evacuated using a combination of prevacuum and high vacuum pumps (usually oil diffusion pumps). The pressure in the specimen chamber typically amounts to about 10^{-4} Pa, allowing the beam electrons to travel from the cathode to the specimen with little interaction with the residual gas molecules. The electronics console consists of the electric power supplies for the acceleration voltage (usual range about 0.5–30 kV) as well as the condenser and objective lenses, the scan generator, and electronic amplifiers for the different signals acquired. Moreover, the console also houses one or more monitors [cathode ray tube (CRT) or liquid crystal display (LCD)] for displaying the micrograph(s), a photo-CRT for analogous image recording, and numerous knobs and a computer keyboard to control the electron

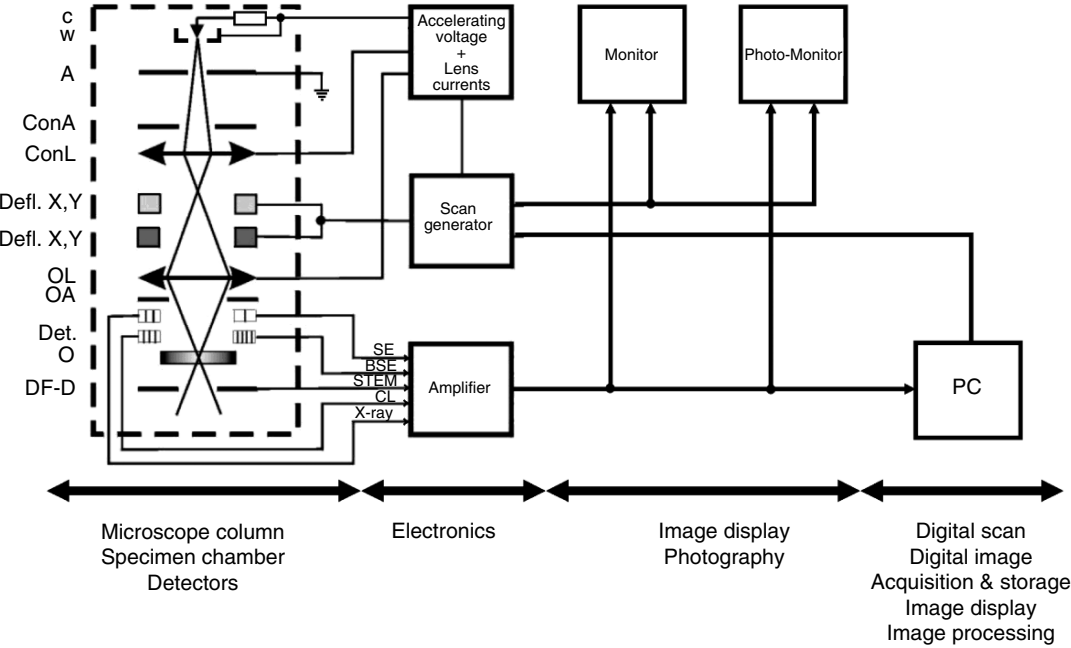


Figure 3–1. Schematic drawing of a conventional SEM. The evacuated microscope column (inside the bold dashed frame) contains the electron gun, electromagnetic lenses, electromagnetic deflection coils, apertures, the specimen stage, and the detectors. The electronics console houses the power supplies for the acceleration voltage and the electromagnetic lenses, the scan generator, amplifiers for the signals, and monitors for display and recording of images. Modern SEMs are controlled by a PC. A, anode; BSE, backscattered electrons; C, cathode; ConA, condenser aperture; ConL, condenser lens; CL, cathodoluminescence; Defl. X, pair of beam deflection coils in the X direction; Defl. Y, pair of beam deflection coils in the Y direction; Det., detectors; DF-D, dark-field detector; O, specimen; OA, objective aperture; OL, objective lens; PC, personal computer; SE, secondary electrons; STEM, scanning transmission electron microscope signal; W, Wehnelt cylinder; X-ray, X-ray signal.

beam, the signals selected, and the image recording. Now modern SEMs mostly use a PC to control the electron beam, to select the signals, and to record as well as to store the digital image(s). In that case the numerous knobs are obsolete and are replaced by a mouse-controlled interactive program running on the PC.

How does the SEM work? The beam electrons are emitted from the cathode and accelerated by a voltage of 0.5–30 kV between the cathode and anode forming a smallest beam cross section—the crossover—near the anode with a diameter of about 10–50 μm. This spot size is too large to produce a sharp image. Therefore the crossover is demagnified by the lens system consisting of one or two condenser lenses and one objective lens and focused on the specimen surface. Most SEMs can produce an electron beam having a smallest spot size of about 5–10 nm and an electron probe current in the range of 10⁻¹²–10⁻¹⁰ A, which is sufficient to form an image with a reasonable signal-to-noise (S/N) ratio. For higher probe currents required for some modes of operation the smallest probe spot size increases to 100 nm or more. The objective lens has a variable relatively long focal length that allows a large

working distance (WD; it corresponds to the distance between the specimen and lower pole piece) in the range of about 5–30 mm. This ensures that the various signals generated by the impinging beam electrons (Figure 3–2) in the small specimen interaction volume can be collected by detectors located lateral above the specimen with sufficient efficiency. Pairs of beam deflection coils located in front of the objective lens and controlled by a scan generator scan the electron probe line by line across a small area of the specimen. Simultaneously, the scan generator controls the deflection coil system of a monitor. The intensity of the monitor is modulated by the amplified signal selected by the operator.

The signals may vary from one location to another as the electron–specimen interaction changes due to, e.g., topography and specimen composition. The magnification of the image is given by the ratio of the length of the scan on the monitor and the corresponding length of the scan on the specimen. For example, an increase in magnification can simply be achieved by decreasing the current of the deflection coils in the microscope column (i.e., lowering the length of the scan on the specimen) and keeping the image size on the monitor constant. It should be mentioned that the magnification also depends on the WD, however, modern SEMs compensate automatically for each WD, thus keeping the displayed magnification correct. Figure 3–3 shows a series of images recorded with increasing magnifications over a range of almost three orders of magnitude.

For crystal structure analysis (cf. Section 7) basically two strategies exist: (1) The mode of beam deflection changes from scanning line by line to rocking of the electron beam when the probe is at rest on a chosen location and the angle of incidence is scanned within a select-

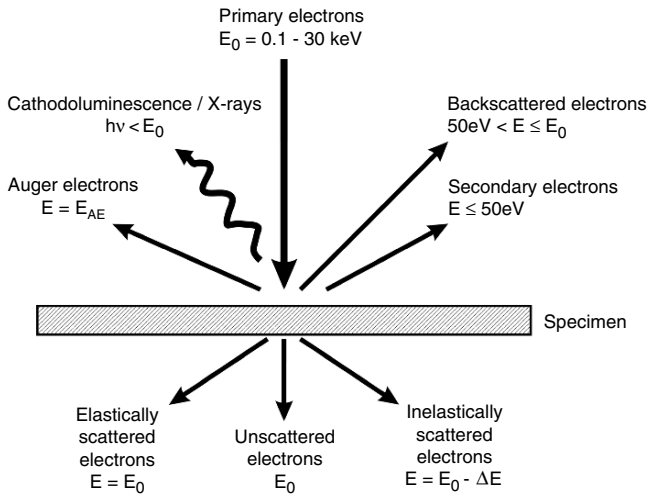


Figure 3–2. Schematic drawing of signals for a thin sample generated by the impinging electrons. E_0 , energy of beam electrons; E , energy of signal electrons; E_{AE} , energy of Auger electrons; ΔE , energy loss of inelastically scattered electrons; $h\nu$, energy of radiation.

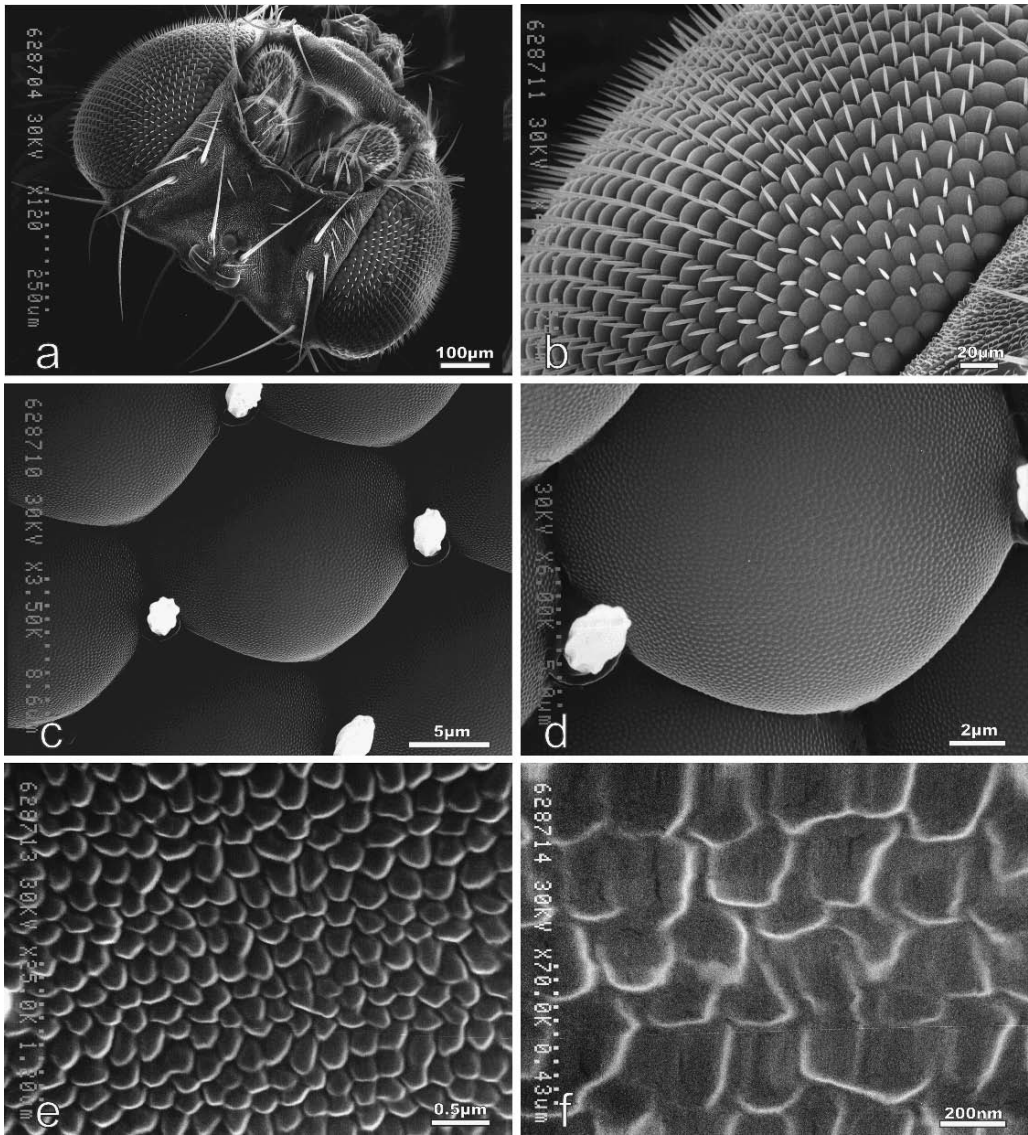


Figure 3-3. Micrograph series of increasing magnification of the head and the eye of a fly (*Drosophila melanogaster*), recorded with secondary electrons at 30kV. The specimen was air dried and sputter coated with about 15nm gold. Note the large depth of focus. The last magnification step from (e) to (f) barely reveals further fine details of the specimen because of the preparation used. The scale of dimensions of the micrographs covers about three orders of magnitude.

able angular range to form an electron channeling pattern. (2) The electron backscattered diffraction pattern from the point of electron beam impact is recorded by means of a position-sensitive detector and analyzed revealing information about the local crystal structure.

One of the greatest strengths of the SEM is the tremendous depth of focus, i.e., the range of heights of the specimen being simultaneously in focus (cf. Figure 3-3). Due to the small objective aperture diaphragm (about 50–100µm) and the large WD the semiangle α_p of the convergent

impinging electron probe is in the order of 10 mrad only. At magnifications that are comparable to those of light microscopy (e.g., 1000 $\times$) the SEM has a depth of focus that is about 100 times greater than that of an optical microscope, obviously because the semiangle of convergence is much larger in the latter case.

To take full advantage of all the information that SEM can provide, an understanding of its operation modes and the influence of electron beam parameters on the image resolution, the image contrast, the signal strength, and the S/N ratio as well as the electron-specimen interaction is mandatory. The remarkable success of scanning electron microscopy over several decades is mainly due to the tremendous depth of focus, the brilliant image contrast, and the relatively straightforward sample preparation for imaging of surfaces, and, in combination with X-ray microanalytical equipment, its capability of local quantitative element analysis of specimens.

2.1 Electron Guns, Electron Lenses, Detectors, and Stages

2.1.1 Electron Guns

The electron gun provides the SEM with an electron beam of adjustable current and energy. The most classic electron gun is the triode gun based on thermionic emission from a tungsten filament heated to about $T_c = 2700$ K (cf. Figure 3–4). The filament has a diameter of about 0.1 mm and is bent in the shape of a V hairpin to localize the emission area on the tip. The size of this area is around $100 \times 150 \mu\text{m}$. By thermionic excitation the electrons overcome the work function Φ of the tungsten tip and a current with the density j_c is emitted according to the Richardson law

$$j_c = AT_c^2 \exp(-\Phi/kT_c) \quad (2.1)$$

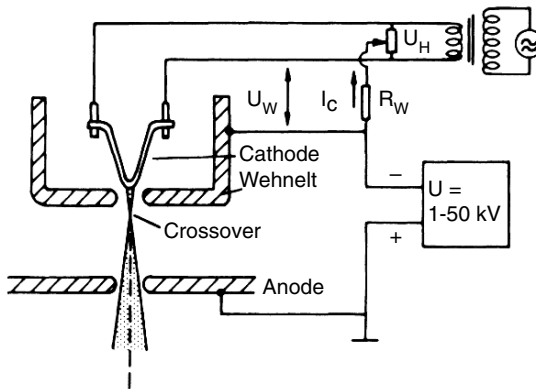


Figure 3–4. Schematic drawing of the thermionic emission triode gun with a tungsten hairpin filament. The filament is heated by the applied voltage U_H . R_w , variable resistor to adjust the potential U_W between the Wehnelt cylinder and cathode; U , acceleration voltage. [Adapted from Reimer (1993); with kind permission of the International Society of Optical Engineering (SPIE), Bellingham, WA.]

where A represents a constant depending on the cathode material and k is the Boltzmann constant; $\Phi = 4.5\text{eV}$ for tungsten. The density j_c depends strongly on the temperature: j_c is about 1.8A cm^{-2} for $T_c = 2700\text{K}$ and about 3A cm^{-2} for $T_c = 2800\text{K}$.

The emitted electrons are accelerated from the filament at a high negative potential (e.g., -30kV) toward the anode at ground potential (0V). Central holes in the Wehnelt cylinder and in the anode enable a fraction of the accelerated electrons (e.g., $E_o = 30\text{keV}$) to move toward the lenses in the microscope column. The emission current is typically in the order of $100\mu\text{A}$ and can be controlled by the bias of the Wehnelt cylinder, which surrounds the filament. The negative Wehnelt bias is provided by a voltage drop caused by the emission current through the resistor R_w . The electrostatic field distribution inside the triode gun has a focusing action to the emitted electrons generating a crossover that is located between the Wehnelt cylinder and the anode. This crossover can be characterized by the diameter d_o and the semiangle α_o of the divergence. d_o is usually in the order of $50\mu\text{m}$. As we will see later the condenser and objective lenses produce a demagnified image of that crossover on the specimen surface representing the final electron probe (diameter d_p).

An important parameter of an electron gun is its axial brightness β , which is defined as the beam current per area (equal to current density) into a solid angle $\pi\alpha^2$ (Reimer, 1985)

$$\beta = j/\pi\alpha^2 = \text{const.} \quad (2.2)$$

It is important to note that the brightness remains constant for all points along the electron optical axis from the cathode through the microscope column to the specimen. This means that the brightness of the final electron probe on the specimen surface is equal to the brightness of the gun regardless of apertures in the microscope column, i.e.,

$$\beta = 4I_o/\pi^2 d_o^2 \alpha_o^2 = 4I_p/\pi^2 d_p^2 \alpha_p^2 \quad (2.3)$$

where I_o is the beam current at the crossover inside the electron gun. Equation (2.3) shows that the characteristic illumination parameters I_p , d_p and α_p cannot be changed independently. For example, an increase of β for given d_p and α_p clearly requires an increase of I_p .

The work function of tungsten is relatively high. Lanthanum hexaboride (LaB_6) has a significantly lower work function ($\Phi = 2.7\text{eV}$) and can therefore emit greater current densities at lower temperature ($T_c = 1900\text{K}$). At the same time the brightness of the electron probe is also increased since the maximum brightness of an electron gun (Reimer, 1985) is given as

$$\beta = j_c E_o / \pi k T_c \quad (2.4)$$

i.e., the brightness β is inversely proportional to the temperature of the cathode. The LaB_6 cathode consists of a small piece of an LaB_6 single crystal with a tip radius typically of about $1\mu\text{m}$. The single crystal is supported by a nonreactive material and is resistively heated. It seems worth mentioning that (1) in the cathode chamber the operation of an

LaB₆ requires a vacuum better than 10⁻⁴ Pa to avoid cathode contamination (tungsten cathode: about 10⁻³ Pa) and (2) its alignment is critical.

Characteristic values of the triode gun with thermionic tungsten and the LaB₆ cathode are summarized in Table 3–1.

2.1.2 Electron Lenses

As discussed in Section 2.1.1 the electrons emerge from the electron gun as a divergent beam. Two or three electromagnetic lenses and apertures in the microscope column (cf. Figure 3–1) reconverge and focus the beam into a demagnified image of the first crossover generated by the gun. The final lens—the objective lens—focuses the beam into the smallest possible spot of 4–10 nm on the sample surface, i.e., the total demagnification is about 5000×

Rotationally symmetric electromagnetic lenses consist of a coil with $N \cdot I$ ampere windings inside an iron pole piece. Typically, $N \cdot I$ is in the order of 10³ A for the condenser and objective lenses. The iron pole piece has a small gap in its axial bore. The current in the coil generates a magnetic field carried by the iron, which also appears at the gap forming a bell-shaped stray field distribution on the optical axis with a radial and axial field component. Off-axis electrons move due to the Lorentz force along screw trajectories because the radial component of the field results in a rotation around the optical axis. Electrons emerging divergently from a point in front of the lens are focused in an image point behind the lens. The lenses of SEMs can usually be considered weak lenses (because the pole piece is not saturated). In this case, the principal planes of the lens coincide with its optical center and the formulas for thin light optical lenses can be used. In close analogy to light optics the strength of an electromagnetic lens can be characterized by its focal length f . Using the thin lens formulas we can write

Table 3–1. Characteristic parameters of different electron guns.

Parameters	Source				Schottky emission cathode ^c
	Thermionic W-cathode ^a	LaB ₆ -cathode ^{a,b}	Cold FEG ^c	Hot FEG ^c	
Brightness (A cm ⁻² sr ⁻¹)	10 ⁵	10 ⁶	10 ⁷ –10 ⁸	10 ⁷ –10 ⁸	10 ⁷ –10 ⁸
Energy spread (eV)	1–3	0.5–2	0.2–0.4	0.5–0.7	0.8
Vacuum (Pa)	~10 ⁻³	~10 ⁻⁴	10 ⁻⁸ –10 ⁻⁹	10 ⁻⁸ –10 ⁻⁹	10 ⁻⁸
Emission current (μA)	~100	1–>50	~10	~10	30–70
Life time	~40 h ~150 h ^d	~200 h	>1 year ~8 years ^d	>1 year	>1 year

^a Reimer (1985).

^b DeVore and Berger (1996).

^c Reimer (1993).

^d Reichelt (unpublished).

$$1/f = 1/p + 1/q \quad (2.5)$$

where p is the distance from the object (= crossover) and q is the distance to the image. Both, p and q are related to the center of the lens. The magnification M is given simply by

$$M = q/p \quad (2.6)$$

where $M < 1$ for $p > 2f$, i.e., a demagnified image of an object is obtained at these imaging conditions. Therefore, a strong demagnification of about $5000\times$ of the first crossover can be obtained for $p \gg 2f$ for each of the two or three lenses by successive demagnification of each intermediate crossover. In case of two condenser lenses they usually are combined and adjusted by one control only.

The pole pieces of condenser lenses are symmetrical, i.e., the diameters of the axial bores in the upper and lower half of the pole piece are identical. In contrast to that the pole piece of the objective lens is very asymmetric (1) to limit the magnetic field at the specimen level and (2) to house the beam deflections coils, the adjustable objective aperture, and the stigmator (not shown in Figure 3–1). The asymmetric objective lens (called the pinhole or conical lens) adapts for the wide range of the WD of about 5–30mm by an adjustable focal length. However, working at a large WD inevitably degrades the electron optical properties of the objective lens and enlarges the final spot size d_p . For a detailed description of the electron optical properties of electromagnetic lenses and deflection coils the reader is referred to books about electron optics (e.g., Glaser, 1952; Grivet, 1972; Klemperer, 1971; Reimer, 1998).

All electromagnetic lenses involved in successive demagnification suffer from an imperfect rotational symmetry and aberrations, which degrade their electron optical performance. The effects of lens aberrations cannot be compensated, however, they can be minimized, which is most effective for the final—the objective—lens. Let us consider briefly the three significant effects.

1. Spherical Aberration. The spherical aberration constant C_s causes an error disc of the diameter (Cosslett, 1972)

$$d_s = 1/2 C_s \alpha_p^3 \quad (2.7)$$

2. Chromatic Aberration. The chromatic aberration caused mainly by the energy spread of the electrons from the gun is characterized by the constant C_c causes an error disc of the diameter

$$d_c = C_c \cdot \Delta E/E_0 \cdot \alpha_p \quad (2.8)$$

where $\Delta E/E_0$ represents the relative energy spread of the beam electrons.

3. Diffraction. The diffraction of electrons on the objective aperture results in a further error disc—the Airy disc—of diameter

$$d_f = 0.6\lambda/\alpha_p \quad (2.9)$$

where λ is the wavelength of the electrons.

In a first approximation it is possible to superpose the squared diameters of the individual discs to estimate the effective electron probe diameter

$$d_{pe}^2 = d_p^2 + d_s^2 + d_c^2 + d_f^2 \quad (2.10)$$

d_p^2 is given by Eq. (2.3) as $d_p^2 = (4I_p/\pi^2\beta)\alpha_p^{-2}$. More precise, but at the same time more complicated relations for the effective probe diameter were derived by Barth and Kruit (1996) and Kolarik and Lenc (1997).

Under the conditions normally used in conventional SEM (i.e., $E_o = 10\text{--}30\text{ keV}$) the chromatic aberration as well as the effect of the diffraction are relatively small compared to the remaining contributions and can be neglected (Reimer, 1985). The optimum aperture α_{opt} , which allows the smallest effective electron probe diameter d_{min} , can be obtained by the first derivative $\partial d_{pe}/\partial \alpha_p = 0$ and is given as

$$\alpha_{opt} = (4/3)^{1/8} [(4I_p/\pi^2\beta)^{1/2}/C_s]^{1/4} \quad (2.11)$$

By using the approach mentioned above, i.e., $d_{pe}^2 = d_p^2 + d_s^2$, and Eqs. (2.3), (2.7), and (2.11), the minimum effective electron probe diameter is

$$d_{p,min} = (4/3)^{3/8} [(4I_p/\pi^2\beta)^{3/2}/C_s]^{1/4} \quad (2.12)$$

It is obvious that $d_{p,min}$ increases as I_p increases or β decreases. Both, I_p and β are parameters depending on the performance of the electron gun [cf. Eq. (2.3)]. C_s is a parameter characterizing the performance of the objective lens and should be as small as possible. As previously mentioned, the operation of the SEM at a large WD inevitably degrades the electron optical properties of the objective lens, i.e., C_s increases as the WD increases. Just to provide a rough idea about values for $d_{p,min}$ and α_{opt} at usual electron energies (10–30 keV), a moderate WD and a probe current I_p of about 10^{-11} A , which gives a sufficient S/N ratio, $d_{p,min}$ typically amounts to approximately 10 nm and α_{opt} to 5–10 mrad.

It is also of interest to know the maximum probe current $I_{p,max}$ under these conditions. Using the Eqs. (2.12) and (2.3) one obtains

$$I_{p,max} = 3\pi^2/16 \cdot \beta C_s^{-2/3} d_{p,min}^{8/3} \quad (2.13)$$

Interestingly, it becomes obvious from Eq. (2.13) that including the effect of the spherical aberration, I_p is now proportional to $d_p^{8/3}$ instead of d_p^2 as before [cf. Eq. (2.3)].

The electron probe current in a SEM equipped with a thermionic gun can be increased several orders of magnitude above 10^{-11} A as required, e.g., for microanalytical studies (cf. Section 6). It is clear from the considerations above that an increase in the probe current inevitably increases the probe size. A rough estimate for 30 keV electrons shows that an increase of I_p to 10^{-9} A requires a probe size of about 60 nm. However, due to the electron–specimen interaction the lateral resolution of X-ray microanalysis is limited to about $1\text{ }\mu\text{m}$ for thick samples. Therefore, a probe diameter of 100 nm or even several hundred nanometers can be tolerated without disadvantage for X-ray microanalysis in this case.

When considering the effective electron probe diameter the chromatic aberration of the objective lens could be neglected for energies >10 keV. Because d_c is inversely proportional to E_0 [cf. Eq. (2.8)] there is a significant increase for energies below 10 keV, in particular for the low-voltage range below 5 keV. For example, for 1 keV electrons the diameter of the chromatic error disc increases by a factor of 30 compared to 30 keV! When using a thermionic cathode with a tungsten filament and a probe current of about 10^{-11} A the energy spread is about 2 eV (cf. Table 3–1) and d_c contributes dominantly to the enlargement of the probe diameter [cf. Eq. (2.10)]. Therefore, the thermionic source is inappropriate for imaging in the low-voltage range. As we shall see in Section 3, field emission guns with a one order of magnitude smaller energy spread and about five orders of magnitude larger brightness are very well suited for low-voltage SEM (LVSEM).

In the context of the objective lens the existence of a stigmator was mentioned, which usually is located near the pole-piece gap. Due to imperfect rotational symmetry of the pole-piece bores, magnetic inhomogeneities of the pole piece, or some charging effects in the bore or at the objective aperture, the magnetic field in the objective lens becomes asymmetric. This causes different focal lengths in the sagittal and meridional planes, which leads to low image quality degraded by astigmatism. The astigmatism can be compensated for by adding a cylinder lens adjustable in its strength and azimuth. The effect of a cylinder lens is realized by the stigmator consisting of a pair of quadrupole lenses.

2.1.3 Detectors and Detection Strategies

Electron detectors specifically collect the signals emerging from the specimen as a result of electron–specimen interaction. The efficiency of the signal collection depends on the type of the detector, its performance, and its detection geometry, i.e., its position related to location of the signal emitting area. For an understanding of the recorded signals, knowledge of the influence of these parameters is critical.

2.1.3.1 Detectors

To detect electrons in SEM three different principles are commonly used. One principle is based on the conversion of signal electrons to photons by a scintillation material. Then, the photons are converted into an electric signal by a photomultiplier, which is proportional to the number of electrons impinging on the scintillator. The second principle is based on the conversion of electrons to electron hole pairs by a semiconductor, which can be separated before recombination causing an external charge collection current. This current is proportional to the number of electrons impinging on the semiconductor. While the principle of scintillation detection is used for secondary, backscattered, and transmitted electrons (in case of thin specimens), the semiconductor detector is mostly used for backscattered electrons only. Finally, the third principle is based on the electron channel multiplier tube, which converts the signal electrons by direct impact at its

input to secondary electrons and multiplies them inside the tube. The output signal is proportional to the number of impinging signal electrons.

As we shall see later, Auger electrons (AE) have a characteristic energy, which is related to the atomic number of the element involved in their generation. For recording of AE in the energy range from 50 eV to several keV, spectrometers with a high energy resolution and high angular collection efficiency are needed in combination with the scintillation detection used, e.g., for secondary and backscattered electrons. The most widely used spectrometer for AE spectroscopy is the cylindrical mirror analyzer. Only those electrons, which pass the energy-selecting diaphragm of the spectrometer, can impinge on the scintillator, thus contributing to the signal.

Besides of electrons the electron-specimen interaction can also produce electromagnetic radiation, namely cathodoluminescence (CL) and X-rays (cf. Figure 3–2). Cathodoluminescence shows a close analogy to optical fluorescence light microscopy (FLM) where light emission is stimulated by irradiation with ultraviolet light (photoluminescence). In principle, for the detection of emitted light, which has a wavelength in the range of about 0.3–1.2 μm , a photomultiplier is very well suited (see above) and therefore most often used. However, the commonly low intensity of the CL signal requires, for a sufficient S/N ratio, a high collection efficiency of the emitted light. Table 3–2 presents the most common detector types for SE, BSE, and CL. The detectors for X-rays will be described in Section 6 of this chapter.

Scintillation Detector. The scintillation detector for SE—the Everhart–Thornley (ET) detector (Everhart and Thornley, 1960)—is shown schematically in Figure 3–5. The generated SE are collected by a positively biased collector grid, then they pass the grid and are accelerated by about 10 kV to the conductive coated scintillator. The scintillation material converts electrons to photons, which are guided by a metal-coated quartz glass to the photocathode of a photomultiplier where photoelectrons are generated and amplified by a factor of about 10^6 . Usually the electronic signal at the output of the photomultiplier is further amplified. Several scintillator materials, such as plastic scintillators, lithium-activated glass, P-47 powder, or YAG and YAP single crystals, are in use, which differ in their performance (for details see, e.g., Reimer, 1985; Autrata, 1990; Autrata and Hejna, 1991; Autrata et al., 1992a,b; Schauer and Autrata, 2004).

When the collector grid of the ET detector is negatively biased by $<-50\text{ V}$ SE are not collected. In this case only the BSE can reach the scintillator on almost straight trajectories because of their higher energies. The detected fraction of BSE is very low because of the small solid angle of collection, i.e., small angular collection efficiency (CE). However, for an efficient detection of BSE the solid angle of collection of BSE detectors is significantly larger by using a larger scintillator and at the same time a shorter distance to the specimen. The BSE detector does not require the collector grid used for the SE (cf. Figure 3–5).

Table 3–2. Most common electron detectors for SEM.^a

Signal	Type of detector	Principles	Specifications	References
SE	Everhart–Thornley	Scintillator–LP–PM	High CE; positively biased collector grid	Everhart and Thornley (1960); Reimer (1985)
	Solid state	Electron hole pair generation	SE are accelerated to >10 keV before detection	Crewe (1970); Reimer (1985)
	MCP	Electron multiplier tube	Positively biased from plate	Postek and Keery (1990); Reimer (1985)
BSE	Everhart–Thornley	Scintillator–LP–PM	Very low CE; negatively biased collector grid	Everhart and Thornley (1960); Reimer (1985)
	Autrata	Scintillator–LP–PM	High CE; $E_{\text{BSE}} \geq 0.8 \text{ keV}$	Autrata et al. (1992)
	Robinson	Scintillator–LP–PM	High CE; $E_{\text{BSE}} \geq 0.9 \text{ keV}$	Robinson (1990)
	Solid state	Electron hole pair generation	High CE; $E_{\text{BSE}} \geq 1.5 \text{ keV}$ bandwidth about $\leq 2 \text{ MHz}$	Stephen et al. (1975)
	MCP	Electron multiplier tube	High CE; $E_{\text{BSE}} \geq 1 \text{ keV}$; negatively biased front plate	Postek and Keery (1990)
CL				
	Ellipsoidal or parabolic mirror with parallel or focused light output or coupled to an LP	Mirror–PM; mirror–spectrometer–PM; mirror–LP–PM	High CE; normally simultaneous BSE detection is not possible (for exception see Autrata et al., 1992)	Autrata et al. (1992); Bond et al. (1974); Rasul and Davidson (1977); Reimer (1985)

^a MCP, Microchannel plate; LP, light pipe; PM, photomultiplier; CE, collection efficiency; E_{BSE} , energy of backscattered electrons.

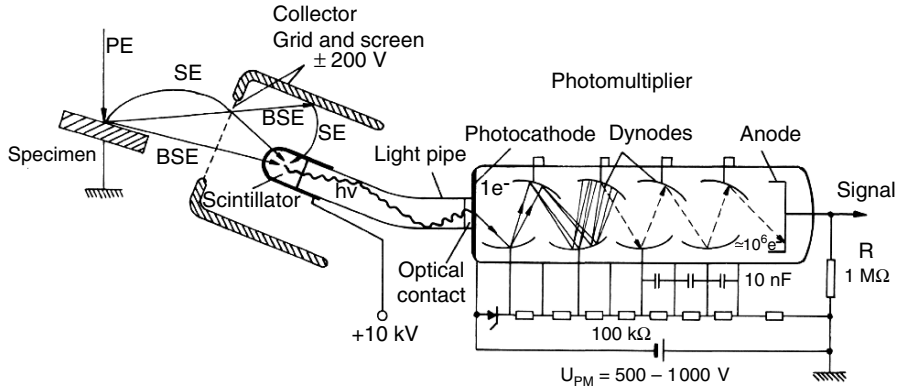


Figure 3–5. Schematic drawing of Everhart–Thornley detector (scintillator–photomultiplier combination) for recording secondary electrons (SE). BSE, backscattered electrons; PE, primary electrons; PM, photomultiplier; $h\nu$, energy of photons. [From Reimer (1985); with kind permission of Springer-Verlag GmbH, Heidelberg, Germany.]

Semiconductor Detector. The semiconductor detector—often denoted as a solid state detector—generates from an impinging electron with the energy E a mean number of electron hole pairs given by

$$n_m = E/E_{\text{exm}} \quad (2.14)$$

where $E_{\text{exm}} = 3.6 \text{ eV}$ is the mean energy per excitation in silicon (Wu and Wittry, 1974). The electron hole pairs can be separated before recombination, in this way generating an external charge collection current, which is proportional to the number of impinging electrons. Because of the energy dependence of n_m the BSE with higher energy contribute with a larger weight to the signal than the BSE having low energies. The semiconductor detector can be used only for the direct detection of BSE because impinging SE are absorbed in its thin electrical conductive layer. However, a special detector design for accelerating the SE to energies above 10 keV also allows for detection of SE (Crewe et al., 1970).

Microchannel Plate Detector. A microchannel plate (MCP) consists of a large number of parallel very small electron multiplier tubes (diameter about 10–20 μm , length of a few millimeters) covering an area of about 25 mm in diameter (e.g., Postek and Keery, 1990). Thus this detector is thin and, when placed between objective pole piece and specimen, enlarges the work distance by only about 3.5 mm. The MCP detector system is efficient at both high and low accelerating voltages, and is capable of both secondary electron and backscattered electron detection. The MCP becomes of increasing interest for studies with low currents and in low-voltage scanning electron microscopy (Russel and Manusco, 1985). However, as yet the MCP detector is not as common as the other detector types described above.

Cathodoluminescence Detectors. In the few cases of strongly luminescent specimens a lens or a concave mirror is sufficient for light collection (Judge et al., 1974). As mentioned above, mostly the intensity of the CL signal is low, thus a high collection efficiency of the emitted light is

indispensable. This requires a solid angle of collection as large as possible, an optimum transfer of the collected light to a monochromator or directly to the photomultiplier, and a photomultiplier with a high quantum efficiency in the spectral range of the CL (Boyde and Reid, 1983). Commercial CL collector and imaging systems allow for investigations with a wavelength from less than 200 nm to about 1800 nm in the imaging and spectroscopy mode. The following are the most commonly used collection systems.

1. Parabolic or elliptic mirrors. The light-emitting area of the specimen is located at the focus of the mirror and is formed into a parallel beam for a parabolic mirror (Bond et al., 1974) or focused to a slit of a spectrometer for an elliptic mirror (e.g., McKinney and Hough, 1977). The solid angle of collection is in the order of π sr but SE detection with an ET detector is still feasible.

2. Rotational ellipsoidal mirror. The light-emitting area of the specimen is located at one focus of the half of the ellipsoid of rotation (Hörl, 1972). The emitted light is focused to a light pipe or to the focal point of an optical microscope objective at the second focal point of the ellipsoid. Although the ellipsoidal mirror has the largest collection angle, the effective collection angle is limited by the acceptance angle of the light pipe or the optical microscope objective, respectively, to about 0.75π sr. The limitation by the acceptance angle can be avoided by placing a parabolic mirror below the second focal point of the ellipsoid (Hörl, 1975). Very recently Rau et al. (2004) proposed an ellipsoidal confocal system collecting the emitted light, which enables CL microtomography in SEM. In principle, the proposed system allows for CL studies at high resolution, which is well below the size of the light-emitting volume.

3. Optical microscope objective. The CL of an optically transparent specimen can be studied by an optical microscope objective positioned below the specimen. The collection angle of this setup amounts to about 1.4π sr (Ishikawa et al., 1973).

2.1.3.2 Detection Strategies

Generally, the detectors for the various signals can be combined and each of them should have an optimum position to make the best use of the electron–specimen interaction. The use of electron spectrometers for AE, BSE, and SE can provide supplementary information about the specimen surface but additional space is needed for a spectrometer. As a matter of fact, the space for detectors is limited in particular with a short WD or with an in-lens position of the specimen for higher resolution. Very recently a proposal was made to improve this situation in scanning electron microscopes with a new design (Khursheed and Osterberg, 2004). The suggested arrangement allows for the efficient collection, detection, and spectral analysis of the scattered electrons on a hemispherical surface that is located well away from the rest of the SEM column.

A conventional SEM commonly is equipped with an ET detector located laterally above the specimen and a BSE detector (for different types see Table 3–2) located centrally above the specimen (top position). Additional ports at the specimen chamber of the SEM enable

additional detectors to be installed. Because of limited space not all of the installed detectors may be used simultaneously, however, there are retractable detectors (e.g., BSE detectors) available, which can be kept in the retracted position when not needed (providing space for another detector or allowing for a shorter WD) and can readily be moved into working position if required for signal recording. Numerous multidetector systems have been proposed for BSE and SE (for review see Reimer, 1984a, 1985). In the top position, e.g., two semiannular semiconductor detectors (Kimoto et al., 1966; Hejna and Reimer, 1987) allow for separation of topographic and material contrast; with a four-quadrant semiconductor detector (Lebiedzik, 1979; Kaczmarek, 1997; Kaczmarek and Domaradzki, 2002) the surface profile can be reconstructed and the distinction between elements with different atomic numbers is improved. Even a six-segment semiconductor detector is of interest (Müllerová et al., 1989).

A combination of two opposite ET detectors, A and B, allows two SE signals, S_A and S_B , to be recorded simultaneously. The difference signal $S_A - S_B$ illustrates the topographic contrast whereas the sum $S_A + S_B$ signal illustrates the material contrast (Volbert and Reimer, 1980; Volbert, 1982). The mixing of the analog electronic signals at that time was performed by electronic circuitry. After analog signal mixing the two original signals were lost. Today, modern SEMs usually record digital images, which are stored in a PC. Thus the mixing of images (their raw data are stored in a memory) can be performed readily after image recording by means of image processing software available from numerous software companies.

For high-resolution and LVSEM the work distance should be as short as possible (say below 5 mm) because both the focal length and the aberrations of the objective lens increase with the WD (see also Sections 2.1.2 and 3). In contrast to the asymmetric objective lens (large focal length) where the region above the specimen is a magnetic field free space, the specimen is immersed in the field of the objective lens with a short focal length. In this case the specimen is very close to the lower objective pole piece or is placed directly inside the pole-piece gap [as in a transmission electron microscope (TEM); see Chapters 1, 2, 6, and 7]. For the latter lens type—the specimen has an “in-lens” position and is limited in size to a few millimeter only—the collection of SE takes advantage of the fact that they can spiral upward in the magnetic field of the objective lens due to their axial velocity component. The SE have to be deflected off the axis to be recorded by an ET detector located laterally above the lens (cf. Figure 3–6).

The separation of the downward moving beam electrons and the upward moving secondary electrons can be done most efficiently by an $\mathbf{E} \times \mathbf{B}$ system, which employs crossed electric and magnetic fields. The forces of these fields compensate each other for the beam electrons, but add for the opposite moving secondary electrons. This magnetic “through-the-lens” detection (for review see Kruit, 1991) of SE has several advantages: (1) SE are separated from BSE, which do not reach the detector because their higher kinetic energy causes different trajectories; (2) very high collection efficiency for real SE emerging from the specimen and a suppression of SE created on the walls of the

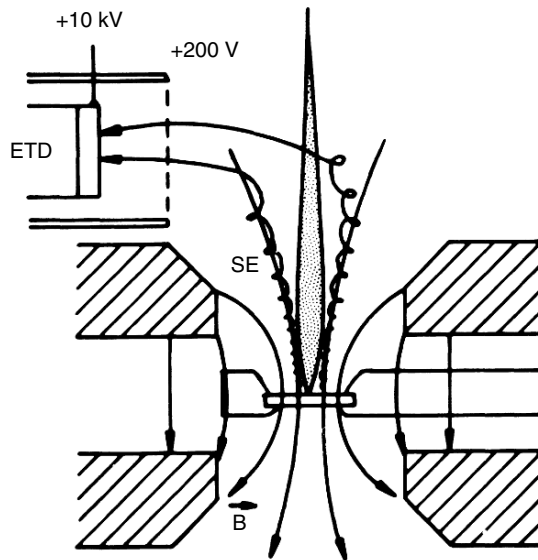


Figure 3-6. Schematic drawing of the magnetic “through-the-lens” detection of secondary electrons (SE) for the “in-lens” position of the specimen. B, magnetic field lines, ETD, Everhart-Thornley detector. [From Reimer (1993); with kind permission of the International Society of Optical Engineering (SPIE), Bellingham, WA.]

system by BSE; (3) improved collection efficiency from inside a porous specimen (in particular cavities or holes facing the electron beam) (Lukianov et al., 1972); and (4) loss of directionality in the image because the SE are detected irrespective of the direction of emission (in contrast to the lateral position of the ET detector; cf. Figure 3-7a). However, a combination of two opposite ET detectors to illustrate the topographic or material contrast has not been tried as yet.

It seems worth mentioning that a real “through-the-lens” detection system was incorporated in one of the early SEMs (Zworykin et al., 1942). The magnetic “through-the-lens” detection of SE was established by Koike et al. (1970) using a TEM with scanning attachment. Finally, magnetic “through-the-lens” detection can be combined with any electron spectrometer as done in SE and Auger spectroscopy (for review see Kruit, 1991).

Another type of “in-lens” detection of SE and BSE is used in the electrostatic detector objective lens (Zach and Rose, 1988a,b; Zach, 1989). The detector is of the annular type and possesses a high collection efficiency of SE of about 75%. Replacing the annular detector by a combination of two semiannular detectors A and B (Figure 3-8) could be used to illustrate the topographic or material contrast, respectively (Reimer, 1993). Similarly, “in-lens” annular type detection of SE and BSE is also used in combined magnetic-electrostatic objective lenses (Frosien et al., 1989), known under the trade name of “Gemini lens.” Both types of lens are advantageous for low-voltage SEM (see Section 3.2) because they provide excellent image resolution at low electron energies.

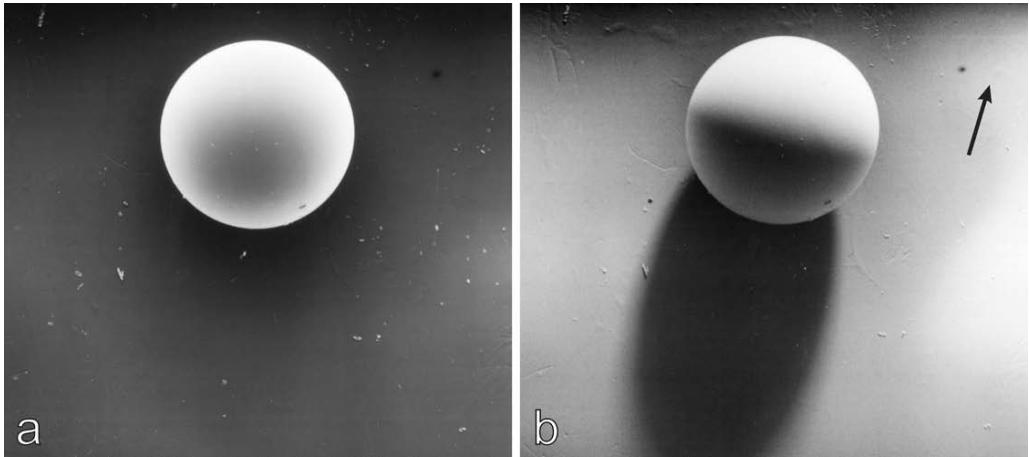


Figure 3-7. Secondary (a) and backscattered electron (b) micrograph of a 1-mm steel ball. The arrow in (b) indicates the direction of the backscattered electron (negatively biased ETD) and the secondary electron detector.

2.1.4 Specimen Stages and Attached Equipment

A conventional SEM is equipped with a specimen stage. The stage commonly can be loaded with the specimen via a specimen-exchange chamber without breaking the high vacuum in the specimen chamber.

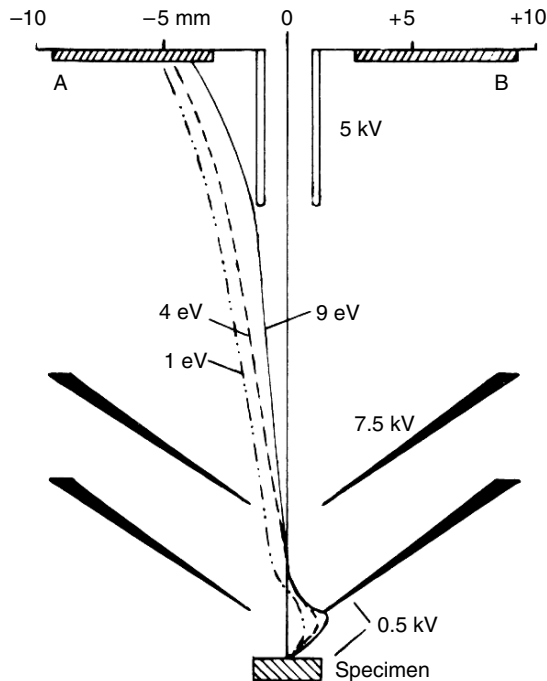


Figure 3-8. Schematic drawing of the “in-lens” detection of secondary electrons (SE) with the electrostatic detector-objective lens (Zach and Rose, 1988a,b; Zach, 1989). [From Reimer (1993); with kind permission of the International Society of Optical Engineering (SPIE), Bellingham, WA.]

The stage allows X, Y, Z movements, rotation around 360° , and tilting (the tilting range depends on the type of the stage, e.g., -15° to $+75^\circ$) of the specimen. The movements, rotation, and tilt are usually motorized in modern scanning electron microscopes. The specimen stage is eucentric if the observation point does not vary during tilting and rotation, however, some stages have this property only for tilting (semieucentric specimen stage) or do not have it (goniometric specimen stage). If the specimen stage is eucentric or semieucentric, the WD and therefore the magnification do not change during X, Y movement or movement along the tilt axis, respectively. Usually, the specimen is at ground potential (0 V), however, the wiring allows also the recording of the specimen current or absorbed current. It is obvious that the higher the electron optical performance of the SEM the better the quality of the specimen stage in terms of mechanical and thermal stability.

The manufacturers of SEMs as well as small companies supplying special attachments offer optionally specimen stages for specific investigations. For example, there are commercial hot stages available for *in situ* surface investigations at elevated temperatures. Depending on the type of heating device, it is possible to reach specimen temperatures up to about 1370 K with a maximum heating rate of about 200 K/min. A hot specimen stage in an environmental scanning electron microscope (see Section 4) is, among other things, very useful in studying the surface modifications caused by chemical reactions due to the exposure of samples to gases. For specific *in situ* heating experiments, e.g., local heating with rapid thermal loads, irradiation heating by a high-power laser coupled to an SEM can be used (see, e.g., Menzel et al., 1992; Wetzig and Schulze, 1995).

Mainly for investigations of organic materials and, in particular, of biological specimens, cold stages are of great interest for low temperature studies. At low temperature the electron beam damage of the sample due to electron-specimen interaction is smaller than at room temperature (see, e.g., Craven et al., 1978; Isaacson, 1977, 1979a; Reimer and Schmidt, 1985; Egerton et al., 2004) and specimens can be investigated in the frozen-hydrated stage (see, e.g., Bastachy et al., 1988; Read and Jeffree, 1990; Walther et al., 1990). However, cold stages are also of significant interest for materials science to investigate the low temperature behavior of materials such as changes in mechanical properties or variations in electrical conductivity. In most cases liquid nitrogen or liquid helium is used as the cooling medium. In particular the temperature range around 4 K and below down to about 1.5 K allows for the investigation of typical low temperature phenomena such as superconductivity and low temperature devices used in cryoelectronics. Furthermore, experiments can be performed in which the temperature range of liquid He is required by the measuring principle, e.g., the ballistic phonon signal represents an example. Here the small specimen volume locally heated by the electron beam acts as a source of phonons, which propagate ballistically (i.e., without scattering) to the opposite side of the crystal where the photon detector is located. Both, the specimen and the detector have to be kept in the temperature range of liquid He. The SEM at very low temperatures was reviewed by Huebener (1988).

Further, deformation stages are used in materials science to study static and dynamic specimen deformation-related phenomena *in situ* (e.g., Wetzig and Schulze, 1995). In more detail, different types of sample deformation such as tension and compression, unidirectional bending, bending fatigue, materials machining (e.g., study of the microscopic mechanisms of abrasive wear), and microhardness testing can be performed with a microhardness tester mounted on the stages in the SEM specimen chamber. This allows for very precise positioning of the indentations generated with a very low force and their subsequent viewing/measuring. In combination with a surface displacement transducer for the detection of acoustic emission signals, the quantitative acoustic emission due to crack coalescence can be measured (Lawson, 1995). There are also stages in different laboratories that combine, e.g., deformation and heating capabilities.

With high-precision stages based on laser interferometer technology, a new field of applications is opened up in the area of SEM/FIB-based e-beam lithography, metrology, and semiconductor failure analysis. The fine positioning of the stage is made with piezoelements, which, according to the manufacturer's specification, allow a positioning reproducibility of better than 50 nm.

To obtain ultralow magnification SEM images an SEM equipped with a motor drive specimen stage fully controlled with a personal computer (PC) has been utilized (Oho and Miyamoto, 2004). This motor drive stage works as a mechanical scanning device. To produce ultralow magnification SEM images, a combination of the mechanical scanning, electronic scanning, and digital image processing techniques is used. This is a time-saving method for ultralow magnification and wide-area observation.

The stage in the SEM specimen chamber can integrate not only tools such as a microhardness tester but also other types of high-resolution microscopes, e.g., a scanning tunneling (Gerber et al., 1986; Stemmer et al., 1994; Troyon et al., 1992), scanning force (Joachimsthaler et al., 2003), or scanning near-field optical microscope (Heiderhoff et al., 2000), thus combining two different microscopic techniques with their specific advantages in one hybrid microscope.

As mentioned in the previous section, the specimen in "in-lens" SEMs must be small because it has to be placed in the gap of the objective lens. This requires specimen stages and holders almost identical to the ones used in TEM (side-entry sample exchange system). The specimen is mounted in a specimen holder, which commonly allows for tilting the specimen around one axis by $\pm 30^\circ$ or $\pm 40^\circ$. Optionally, there are, for example, double-tilt specimen holders with two tilt axes as well as hot and cold/cryoholders available. The specimen holders also normally allow the use of support grids 3 mm in diameter, which is of interest for studies in transmission mode.

2.1.5 Special Topics

2.1.5.1 Digital Image Recording

As mentioned briefly in Section 2.1.3.1, the SEM generates analog electronic image signals. In older SEMs the electron beam must be scanned incessantly across the surface area of interest while viewing the image

on the monitor. The information obtained by each scan is just used to refresh the image on the monitor unless the image is recorded photographically. The electron irradiation dose of each scan accumulates to a high total dose. Because radiation damaging as well as beam-induced contamination scale with the total electron dose applied, irradiation-sensitive materials in particular are possibly already damaged during the visual inspection, i.e., before recording an image of the viewed area.

Modern SEMs allow recording of multichannel (e.g., SE and BSE) digital images, which are stored pixel by pixel (pixel: picture element) in a PC. Digital images usually have a square size of 512×512 , 1024×1024 pixels or larger (Postek and Vladár, 1996), however, rectangular image formats are also in use, e.g., 3000×2000 pixels. For each pixel the analog signal arriving from the detector is integrated during the pixel time. The value obtained represents the pixel intensity, which is digitized by an analog-to-digital-converter (ADC) usually into a range of 10 or 12 bits. In practice, a sufficient lateral and signal resolution can be obtained with 1024×1024 pixels and 10 bits, respectively, which requires a storage capacity over 10^7 bits, i.e., 2 MB. Twice the number of pixels in the X and Y direction requires a 4-fold storage capacity, i.e., 8 MB. A modern PC presently has roughly 200 GB mass storage on the harddisk, which corresponds to about 10^5 images of 2 MB or 2.5×10^4 images of 8 MB, respectively. The fast processors presently used in PCs allow the display of an 8-MB image within milliseconds. For final storage, the digital images can be transferred from the PC via fast data transfer to external mass storage devices or big computers. However, access to images stored in an external mass storage device is slower than for direct access to the harddisk of the PC.

2.1.5.2 Specimen Tilting and Stereo Imaging

Specimen stages allow the sample to be tilted, which is of interest for several special applications such as stereo imaging, reconstruction of the topography, three-dimensional morphometry, and possibly contrast enhancement. In case of a flat object aligned normal to the beam, i.e., the tilt angle Θ amounts to 0° , there is no distortion of the projected shape of structures. For example, circular holes in a flat specimen appear circular in the image (Figure 3–9a). After tilting the flat object (the tilt axis has a horizontal direction in the micrograph) two effects become obvious in the image (especially visible at high tilt angles, e.g., $\Theta = 45^\circ$) (Figure 3–9b): (1) the circular holes have an elliptic shape with an axis ratio of approximately 0.7, i.e., the shape is distorted, and (2) the upper and lower rim of the upper and lower hole appears unsharp whereas the rim of the central hole appears sharp.

The first effect is caused by the fact that due to the tilt the scanned range on the specimen surface perpendicular to the tilt axis is enlarged by a factor $1/\cos\Theta$ (Figure 3–9e). That corresponds to a reduced magnification

$$M' = M \cos \Theta \quad (2.15)$$

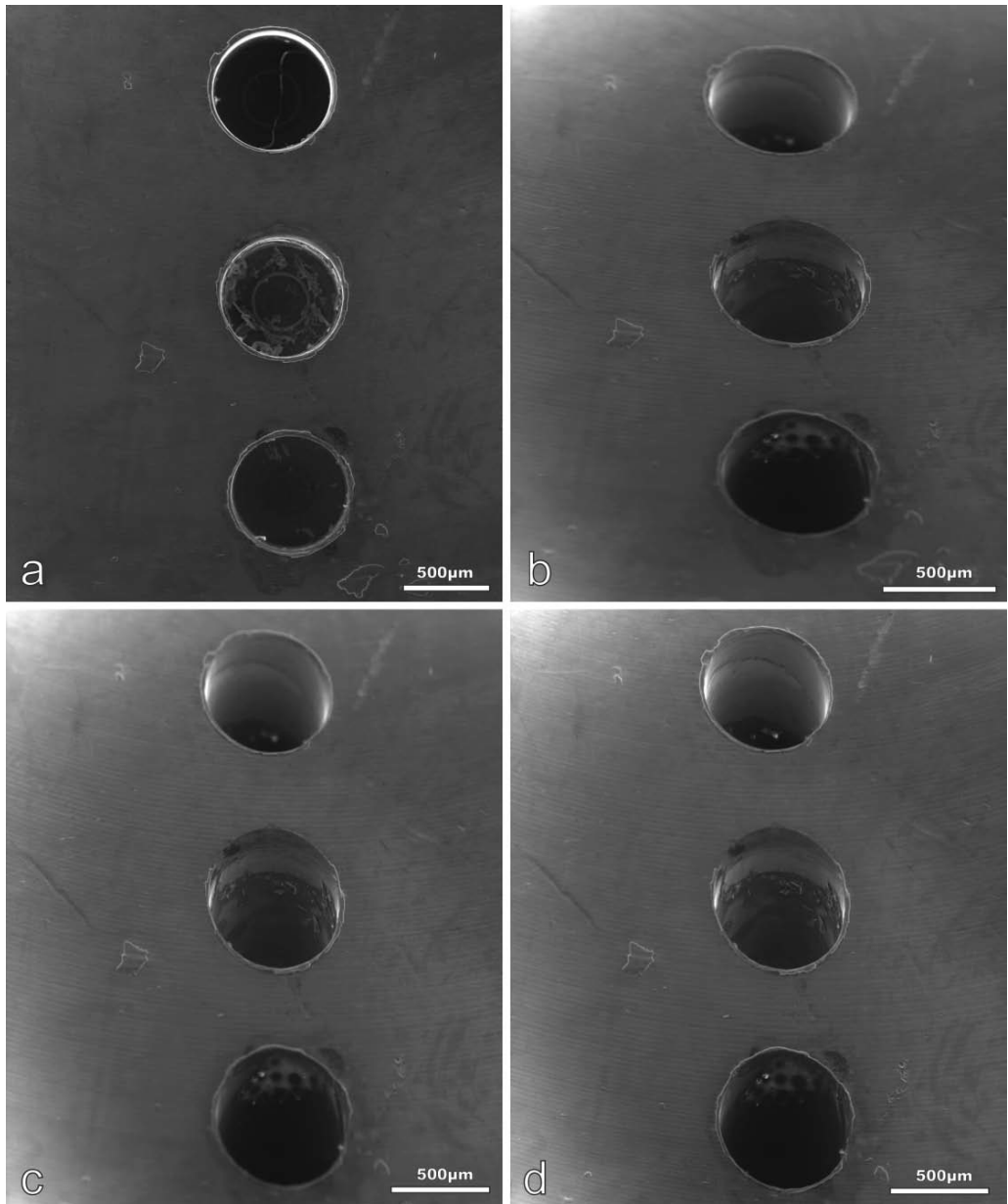


Figure 3–9. Effect of tilt compensation and dynamic focusing in the SEM. Secondary electron micrographs of the holes in a flat aluminum specimen. (a) Tilt angle $\Theta = 0^\circ$. (b) Tilting $\Theta = 45^\circ$ around the horizontal axis (the focus of the beam is located in the center of the micrograph). (c) Tilt compensation is “ON.” (d) Tilt compensation and dynamic focusing are “ON.” The visible wall of the bore of the holes proves that the specimen is still tilted. (e) Schematic illustration of the effects caused by tilting the sample. The position of the optimum focus plane, the depth of focus D , and the height range Δz ($\Delta z > D$) are shown for a constant focus of the beam (solid lines). In that case, only a central region along the tilt axis is within the depth of focus (i.e., sharp image) whereas the lower and the upper range are outside D (unsharp region of the image). In case of the dynamic focus, three positions of the beam (dashed lines) are drawn indicating that the whole scan range will be in focus, thus being imaged sharply.

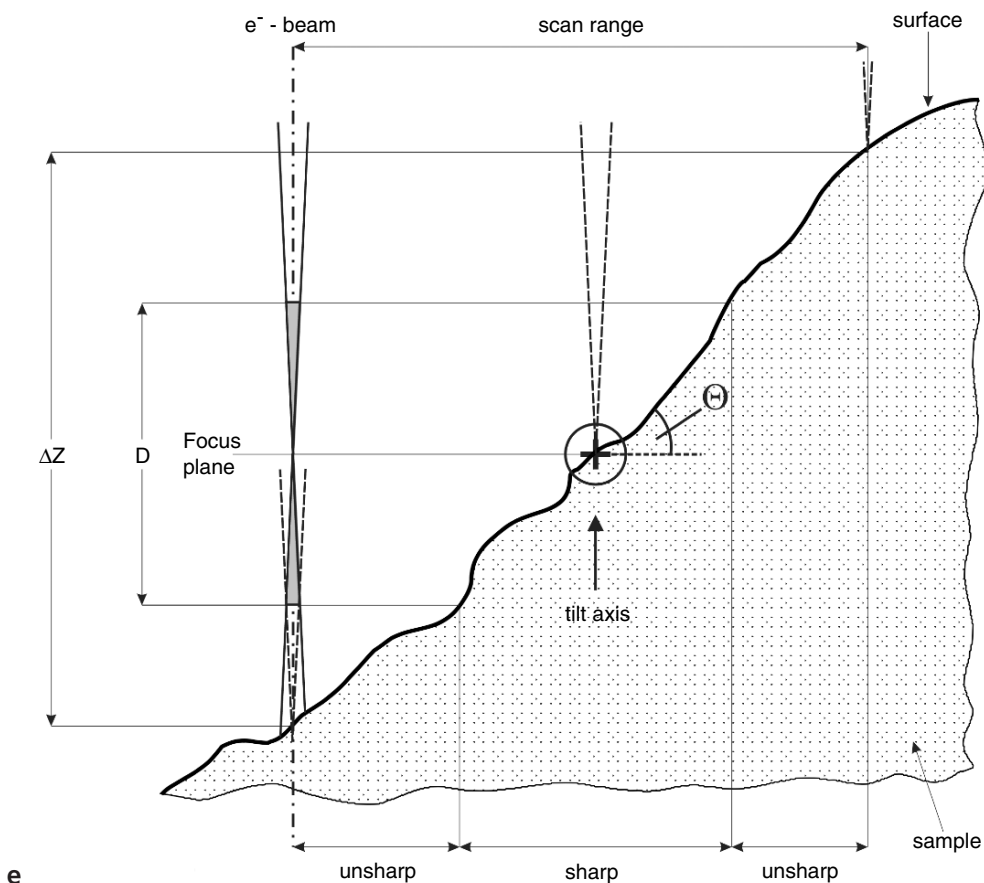


Figure 3-9. Continued

in the direction of the short axis whereas along the tilt axis the magnification M is not affected. The effect can be fully compensated by enlarging the reduced magnification by $1/\cos \Theta$, which restores the magnification M (Figure 3-9c and e). The tilt compensation can be performed directly by electronic means (hardware; the unit is called "tilt compensation") during scanning or posterior by digital image processing on condition that the directions of the tilt axis and the tilt angle are known.

The second effect is caused by the fact that due to the tilt the height range of the tilted specimen extends the depth of focus, thus the image is not sharp in regions outside the depth of focus. This effect can be compensated by "dynamic focusing" (Yew, 1971), i.e., by adjusting the strength of the objective lens as a function of the scan position perpendicular to the tilt axis. This adjustment brings the optimum focus position in coincidence with the surface at all working distances in the scanned range (Figure 3-9d and e). Both effects mentioned can be compensated completely only for planar specimens and a known tilt angle.

The SEM forms in imaging mode a two-dimensional image of a three-dimensional specimen with each of the signals generated by

electron–specimen interaction (cf. Figure 3–2). Although these images contain a wealth of information about the specimen, there is no solid information about the third dimension, which is parallel to the optical axis. Stereo imaging is one possibility to obtain information about the third dimension. It takes advantage of the fact that depth perception is obtained when viewing an object from two separate directions. In an SEM stereo imaging is performed by taking two images—the stereo pair—at two different tilt angles of the specimen. A good stereo effect is obtained when the angles differ from each other by about 6° . A significantly larger difference in Θ overemphasizes the stereo effect whereas a smaller difference in Θ shows a softened stereo effect. Usually, a stereo pair is viewed through a stereo viewer. A simple version of a stereo viewer consists of two short focal lenses on a stand at the correct distance from the stereo pair. There are more sophisticated versions with lens–mirror combinations, which allow for a larger field of view. For each type of viewer it is mandatory to place and to align both images precisely to obtain correct depth perception. Figure 3–10 shows a stereo pair of SEM micrographs.

A further method for viewing the stereo images is the anaglyph technique (Judge, 1950), which can now readily be performed by PC (cf. Figure 3–56). In this technique both images are superimposed in

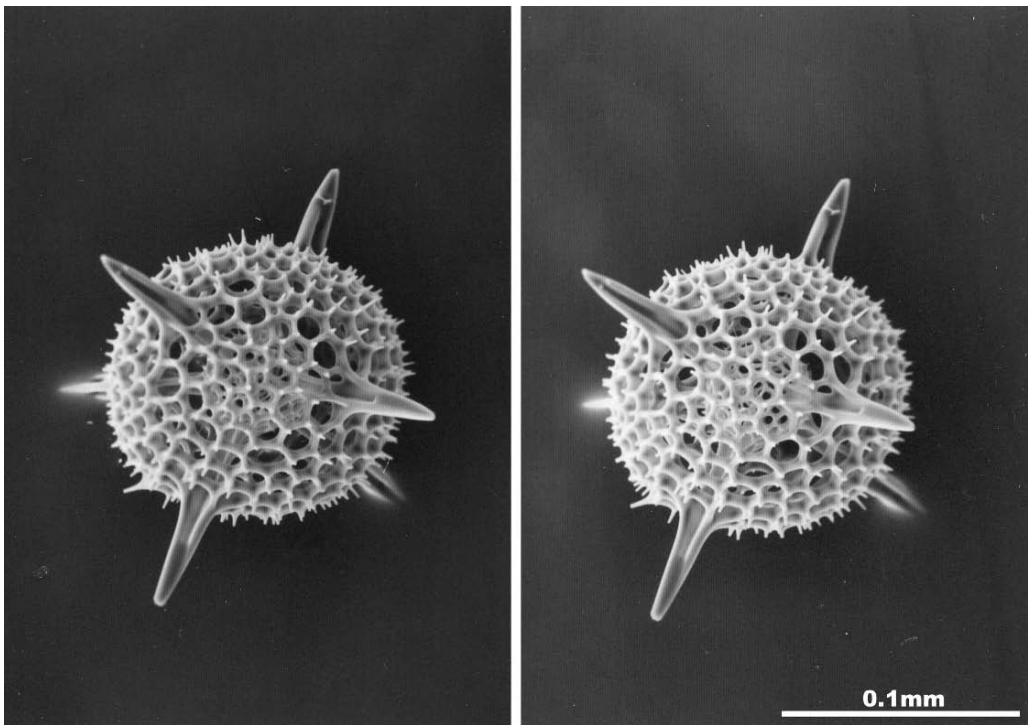


Figure 3–10. Stereopair of a radiolarian with radiating threadlike pseudopodia and a siliceous skeleton re-recorded in SE mode at 15 keV. The specimen was sputter coated with gold for sufficient electrical conductivity. The difference in tilt angles of the micrographs amounts to $(\Theta_r - \Theta_l) = 6^\circ$. (The stereopair was kindly provided by Rudolf Göcke, Institut für Medizinische Physik und Biophysik, Münster, Germany.)

different colors. A red–green stereo anaglyph coding can be obtained readily by commercial software, e.g., *AnalySIS Pro 3.1* [Soft Image System (SIS), Münster, Germany]. This allows for a quick and simple qualitative assessment. Usually, red is used for the “left eye” and green for the “right eye” image. The mixed colored image has to be viewed by colored glasses using red for the left and green for the right eye. However, the stereo pair can also be used to calculate the height difference Δh between two image points 1 and 2 by measuring the parallax p_x given as

$$p_x = (x_2^l - x_1^l) - (x_2^r - x_1^r) \quad (2.16)$$

and

$$\Delta h = p_x / 2M \sin(\Theta_r - \Theta_l) \quad (2.17)$$

where $(x_2^l - x_1^l)$ corresponds to the distance between the two points in the left and $(x_2^r - x_1^r)$ to the distance between the two points in the right image. Θ_r and Θ_l correspond to the tilt angle of the specimen used for the right and left image, respectively. Equation (2.17) holds for magnifications $M > 100$, i.e., in case of parallel projection. The successful application of the latter two formulas requires (1) distinct surface structures to measure p_x with sufficient accuracy, and (2) the magnification and the tilt angle must be known exactly. On the basis of the relations (2.16) and (2.17) and data analysis software [e.g., *AnalySIS Pro 3.1* (SIS, Münster, Germany), *3-D Morphometry in SEM* (ComServ, Salzburg, Austria), and *MeX* (Alicona Imaging GmbH, Grambach, Austria)] quantitative dimensional and angular measurements, the reconstruction of the specimen topography and three-dimensional morphometry can be achieved. The latter is very useful to analyze microstructures such as blood capillaries, which have diameters in the range of a few micrometers (Malkusch et al., 1995; Minnich et al., 1999, 2003). Using the image pair imported for anaglyph generation, a point of reference has to be selected in one of the images that would also be clearly visible in the second image. The software (e.g., *AnalySIS Pro 3.1*, see above) uses this reference point to define a small region and then uses correlation to accurately align the images. Once aligned, the parallax difference allowed generation of a new 8-bit grayscale image in which depth differences are encoded as different gray values. Quantitative measurements, e.g., the volume of depressions, can be performed using an SIS Macro (macroinstruction, i.e., application of a specific programming language), based on the SIS-Stereo imaging module.

2.1.5.3 Magnification Calibration

The actual magnification of the SEM is indicated numerically and by a scale bar on the monitor with a precision of about $\pm 3\%$. However, if exact measurements have to be made (e.g., see Section 2.1.5.2), the magnification should be verified using an external standard. Calibrated gratings with known spacings are commercially available from different suppliers of electron microscopy accessories (e.g., Agar Scientific, <http://www.agarscientific.com>). For a magnification

range up to about 100,000 $\times$ crossed gratings with spacings of 1200 and 2160 lines/mm (cf. Figure 3–11) are recommended.

Latex spheres of defined diameter (different diameters in the range from about 0.1 to 1 μm are commercially available) can also be used. However, the size of the latex spheres varies to some extent (e.g., small diameter, 0.112 μm ; standard deviation, $10^{-3} \mu\text{m}$; large diameter, 1.036 μm ; standard deviation, $16.1 \times 10^{-3} \mu\text{m}$). Moreover, the latex spheres are sensitive to electron radiation, thus their size may change caused by electron dose-induced damages. To calibrate the magnification range above 100,000 $\times$ negatively stained catalase (periodic lattice spacings: 8.75 and 6.85 nm) can be used as standard (preferentially in the STEM mode).

2.2 Electron–Specimen Interaction and Signal Generation

As the beam electrons enter the specimen, they interact with the atoms of the specimen. This interaction either results in elastic or inelastic scattering of the impinging electrons.

The elastic scattering of the electron is caused by its interaction with the electrical field of the positively charged nucleus and results only in a deflection of the beam electron, i.e., after the scattering event the electron trajectory has a different direction than before scattering. There is almost no loss of kinetic energy of the electron scattered elastically. For scanning electron microscopy it is necessary to know the elastic electron scattering through large angles between 0° and 180°. The scattering can be described quantitatively by the scattering cross section σ . The exact elastic scattering cross sections for large-angle

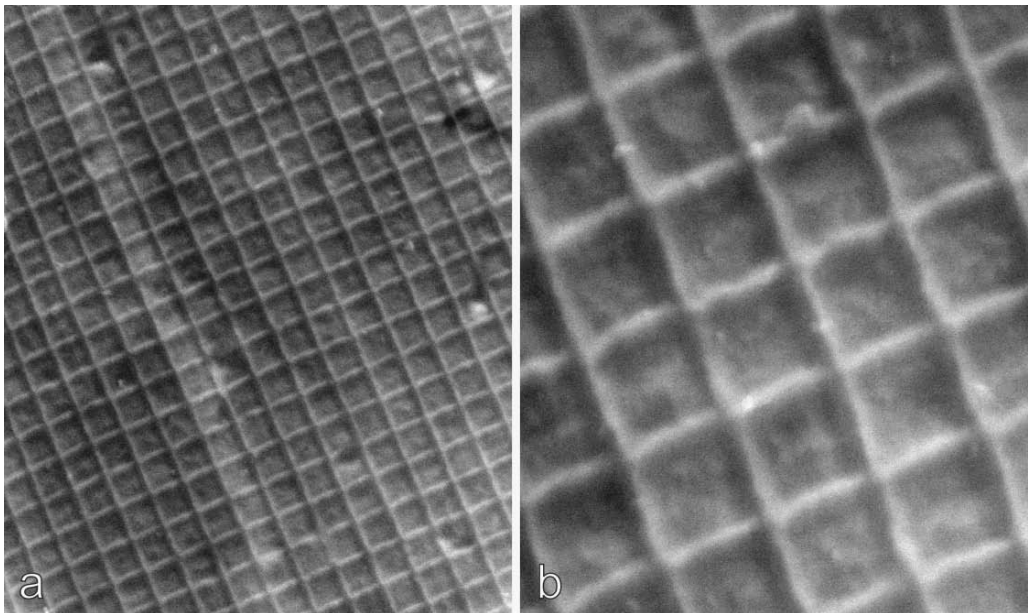


Figure 3–11. SE micrographs of a commercial cross-grating replica with 2160 lines/mm at low (a) and medium magnification (b).

scattering are the Mott cross sections $\sigma_{M,el}$, which, in contrast to Rutherford scattering, consider the electron spin and spin-orbit coupling during scattering (for details see, e.g., Reimer and Krefting, 1976; Rez, 1984; Reimer, 1985). The easy to calculate unscreened differential Rutherford cross section $d\sigma_{R,el}/d\Omega$ are given as (Reimer, 1985)

$$d\sigma_{R,el}/d\Omega = e^4 Z^2 / [4(4\pi\epsilon_0)^2 m^2 v^4 \sin^4(\varphi/2)] \quad (2.18)$$

where $d\Omega$ is the cone of the solid angle, e is the electric charge ($e = 1.602 \times 10^{-19}$ C) and m the mass of the electron, v is the velocity of the electron, Z is the atomic number, ϵ_0 is the dielectric constant ($\epsilon_0 = 8.85 \times 10^{-12}$ C/Vm), and φ is the scattering angle. The comparison of the differential Mott cross sections $d\sigma_{M,el}/d\Omega$ (Reimer and Lödding, 1984) with the unscreened differential Rutherford cross sections $d\sigma_{R,el}/d\Omega$ for electron energies between 1 and 100 keV shows that there are strong deviations of the Rutherford cross section, particularly for high Z . Mott cross sections for electron energies below 1 keV (energy range 20 eV to 20 keV) were calculated by Czyzewski et al. (1990; see also <http://web.utk.edu/~srcutk/Mott/mott.htm>). There is very reasonable agreement between both cross sections for low atomic numbers and electron energies above 5 keV. However, for low Z and energies below 5 keV, the Rutherford cross sections are larger than the Mott cross sections for scattering angles below 70° – 80° and are smaller than the Mott cross sections for scattering angles above 70° – 80° . The probability for elastic scattering is approximately proportional to Z^2 and inversely proportional to E^2 (with $E = mv^2/2$), i.e., the scattering cross section strongly increases with the atomic number and decreases for increasing electron energy E . The total elastic scattering cross section σ_{el} can be obtained by integration

$$\sigma_{el} = 2\pi \int_0^\pi (d\sigma_{el}/d\Omega) \sin \varphi d\varphi \quad (2.19)$$

σ_{el} can be used to calculate the mean free path for elastic scattering Λ_{el} , i.e., the free path between two consecutive elastic scattering events in a specimen consisting of many atoms, which is given as

$$\Lambda_{el} = 1/N\sigma_{el} \quad (2.20)$$

N represents the number of atoms per unit volume and can be calculated simply by

$$N = N_A \rho / A \quad (2.21)$$

where ρ is the density, N_A is Avogadro's number ($N_A = 6.0221 \times 10^{23}$ mol $^{-1}$), and A is the atomic weight (g/mol). Much more detailed data of elastic electron scattering cross sections were recently published (Jablonski et al., 2003).

As we shall see later, the mean free path is an important quantity for describing plural (mean number of collisions <25) and multiple electron scattering (mean number of collisions $>25 \pm 5$).

The inelastic scattering of the electron is caused by its interaction with the electrical field of the electrons in the solid, i.e., either with the electrons in the valence or conduction band and with atomic electrons of

inner shells, respectively. After an inelastic scattering event the electron trajectory has a slightly different direction than before scattering (typically the inelastic scattering angles are of the order of a few milliradians only) and less kinetic energy. If the lost energy was transferred to electrons in the valence or conduction band then the excitation of plasmons (a plasmon is a longitudinal charge-density wave of the valence or conduction electrons) or inter- and intraband transitions may occur. Both the energy of plasmons and the energy differences of inter- and intraband transitions are in the order of about 5–50 eV. The physics of the latter processes is reviewed by Raether (1980). If the lost energy was transferred to atomic electrons of inner shells then, for example, K-, L-, or M-shell ionization may occur. In this case the energy loss typically is higher than 50 eV. The differential inelastic electron scattering cross section with a free electron (which is an approximation for an electron in the valence or conduction band) is given (Reimer, 1985) as

$$d\sigma_{\text{in}}/dW = \pi e^4 / [(4\pi\epsilon_0)^2 E W^2] \quad (2.22)$$

where W is the energy loss. The equation shows that the differential inelastic scattering cross section is inversely proportional to electron energy E and to W^2 and that small energy losses occur with a larger probability. In a more complex approach for the differential inelastic scattering cross section the energy loss function $\text{Im}(-1/\epsilon)$ is used taking the dielectric properties of the material into account (e.g., Powell, 1984). An impinging electron can be inelastically scattered passing the atom even in a distance of a few nanometers, thus the inelastic scattering is delocalized to a certain extent (Isaacson and Langmore, 1974; Zeitler, 1978; Reichelt and Engel, 1986; Müller and Silcox, 1994, 1995).

In the case of inner shell excitation the electron interaction is localized to an electron shell. The corresponding inelastic scattering, also called ionization cross section, is the probability of bringing a scattering atom to a given excited state through an inelastic process. The related cross sections are typically at least two orders of magnitude smaller than those for the electrons in the valence or conduction band. Calculations of the ionization cross sections of the K-, L-, and M-shell have been published (see, e.g., Leapman et al., 1980; Inokuti and Manson, 1984; Egerton, 1986).

The total inelastic scattering cross section σ_{in} can be obtained by integration

$$\sigma_{\text{in}} = 2\pi \int_0^\pi (d\sigma_{\text{in}}/d\Omega) \sin\varphi d\varphi \quad (2.23)$$

The mean free path for inelastic scattering Λ_{in} is given analogous to Eq. (2.20) as

$$\Lambda_{\text{in}} = 1/N\sigma_{\text{in}} \quad (2.24)$$

(for detailed data and calculation of the electron inelastic mean free path see Powell and Jablonski, 2000).

The total scattering cross section then is given as

$$\sigma = \sigma_{\text{el}} + \sigma_{\text{in}} \quad (2.25)$$

In bulk specimens multiple scattering of the impinging electrons takes place. Mainly the multiple elastic scattering causes a successive broadening of their angular distribution and can, after numerous scattering events, result in beam electrons leaving the specimen. The beam electrons, which leave the specimen, are designated as backscattered electrons and carry an important class of information about the local specimen volume through which they have been passing. Multiple inelastic scattering along the electron trajectories results in a slowing down and the beam electron can come to a standstill if it cannot leave the specimen as BSE. The majority of beam electrons are scattered both elastically as well as inelastically. Therefore, the majority of BSE have energies smaller than E_0 (cf. Figure 3–12). The broadening of the angular distribution can be calculated analytically using the autoconvolution of the single scattering distribution expanded in terms of Legendre polynomials (Goudsmit and Saunderson, 1940).

Another method to treat multiple scattering is the simulation of the successive scattering events by Monte Carlo calculations for about 10^3 – 10^5 electron trajectories (cf. Figure 3–13; for Monte Carlo simulations of electron scattering see, e.g., Reimer and Krefling, 1976; Kyser, 1984; Reimer and Stelter, 1986; Joy, 1987b; Reimer, 1968, 1996; Drouin et al., 1997; Hovington et al., 1997a,b).

In this method, the most important scattering parameters, such as scattering angle, mean free electron path, and energy loss, are simulated for each individual scattering event along the trajectory by a computer using random numbers and probability functions of the scattering parameters. The energy loss along the trajectory can be described by the Bethe continuous-slowness-down approximation (Bethe, 1930)

$$dE/ds[\text{eV/cm}] = -7.8 \times 10^{10} (Z\rho/AE)\ln(1.166 E/I) \quad (2.26)$$

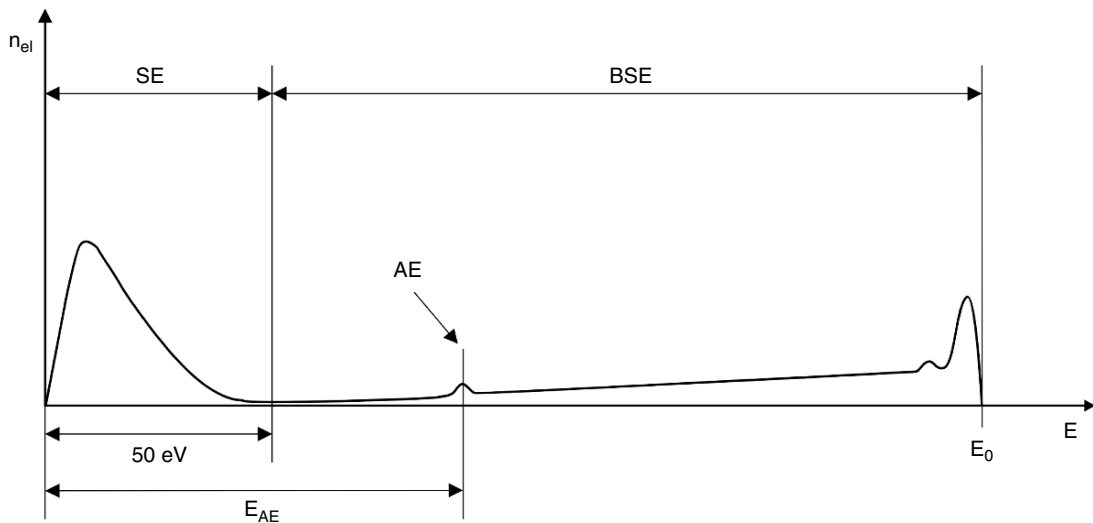


Figure 3–12. Schematic energy distribution of electrons emitted from a surface as a result of its bombardment with fast electrons with energy E_0 . AE, Auger electrons; BSE, backscattered electrons; SE, secondary electrons; E_{AE} , energy of AE; n_{el} , energy-dependent number of emitted electrons.

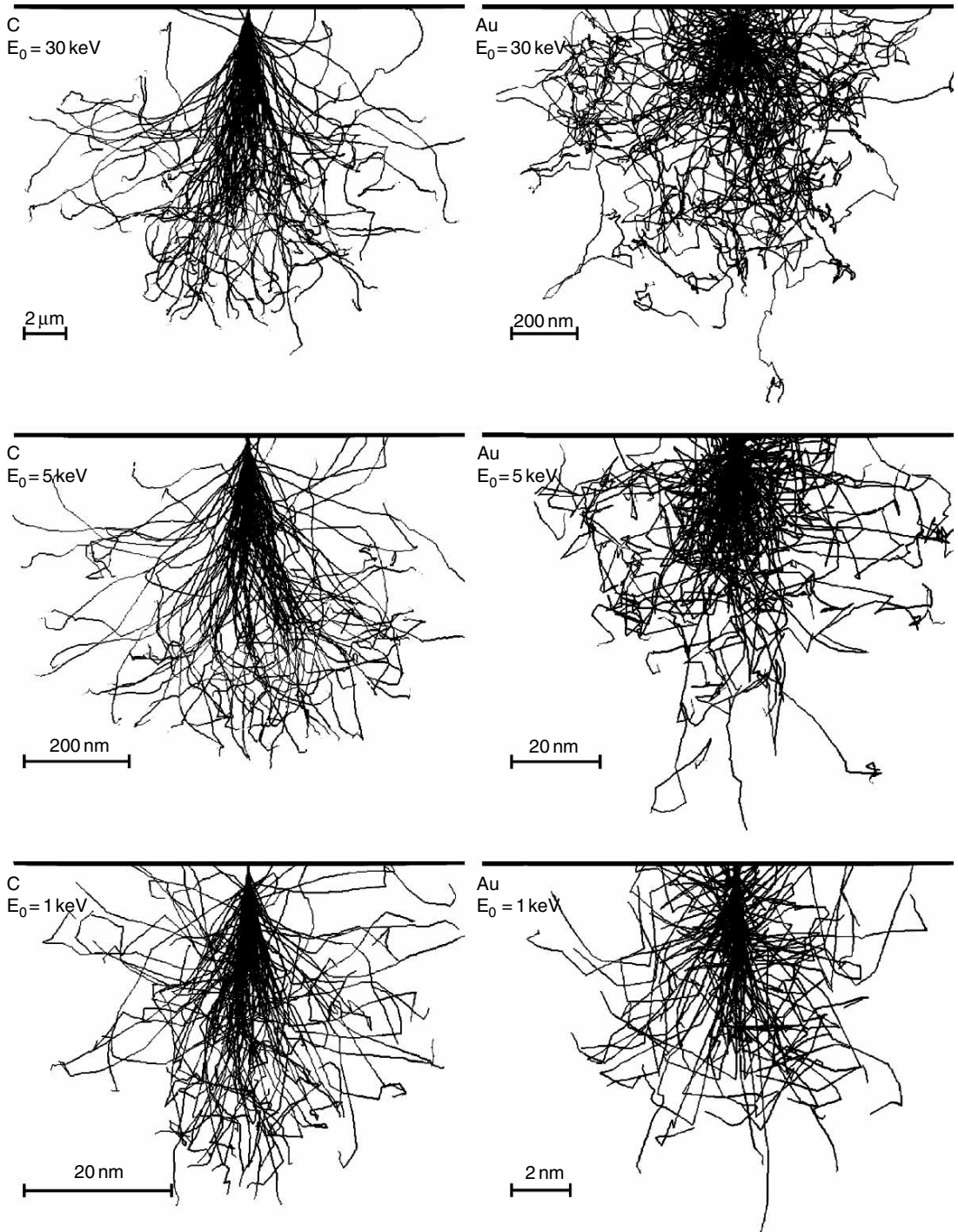


Figure 3-13. Monte Carlo simulation of the trajectories of 100 electrons for carbon (atomic number $Z = 6$) and gold ($Z = 79$) for electron energies $E_0 = 30, 5$, and 1 keV . For simulation of the electron trajectories the Monte Carlo program MOCASIM (Reimer, 1996) was used. Note the different scales across three orders of magnitude indicated by bars and the variation of the shape of the local volume where electron scattering takes place. That local volume is usually denominated as the excitation volume.

where ρ is the density (g/cm^3), E the electron energy (eV), and J the mean ionization potential (Berger and Seltzer, 1964) given by

$$J[\text{eV}] = 9.76Z + 58.8Z^{-0.19} \quad (2.27)$$

The limitations of the Bethe expression at low electron energy can be overcome by using an energy-dependent value J^* for the mean ionization potential (Joy and Luo, 1989)

$$J^* = J/(1 + kJ/E) \quad (2.28)$$

where k varies between 0.77 (carbon) and 0.85 (gold). The total traveling distance of a beam electron in the specimen—the Bethe range R_B —can be obtained by integration over the energy range from E_0 to a small threshold energy and extrapolation to $E = 0$.

The practical electron range R (cf. Figure 3–14) obtained by fitting experimental data of specimens with different Z over a wide energy range is given by the power law

$$R = aE_0^n \quad (2.29)$$

where n is in the range of about 1.3–1.7 and the parameter a depends on the material (Reimer, 1985). Characteristic values for R , σ_{el} , σ_{in} , Λ_{el} ,

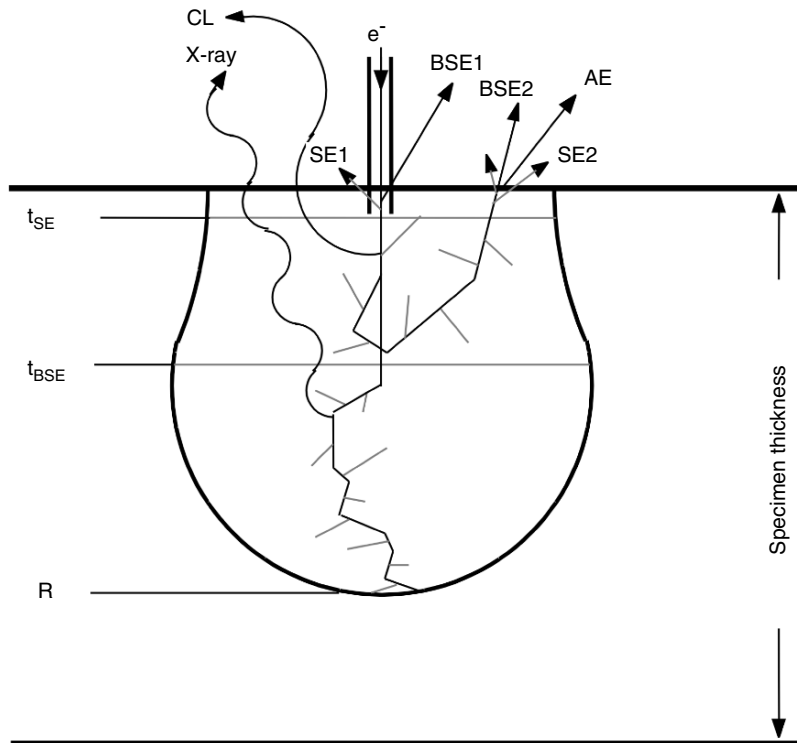


Figure 3–14. Schematic illustration of the generation of secondary electrons SE1 and SE2, backscattered electrons BSE1 and BSE2, Auger electrons AE, cathodoluminescence CL, and X-rays in a bulky specimen. t_{SE} and t_{BSE} indicate the escape depth for SE and BSE, respectively. R is the electron range.

Table 3–3. Characteristic values for R , σ_{el} , σ_{in} , Λ_{el} , and Λ_{in} .^a

Element	Parameter	$E_0 = 1 \text{ keV}$	$E_0 = 5 \text{ keV}$	$E_0 = 10 \text{ keV}$	$E_0 = 30 \text{ keV}$
Carbon $Z = 6$	$\sigma_{el} (\text{nm}^2) \times 10^2$	0.65	0.11	0.055	0.018
	$\sigma_{in} (\text{nm}^2) \times 10^2$	1.95	0.33	0.165	0.054
	$\Lambda_{el} (\text{nm})$	1.5	9.0	18	55
	$\Lambda_{in} (\text{nm})$	0.5	3.0	6	18
	$R (\mu\text{m})$	0.033	0.49	1.55	9.7
Copper $Z = 29$	$\sigma_{el} (\text{nm}^2) \times 10^2$	1.84	0.64	0.37	0.15
	$\sigma_{in} (\text{nm}^2) \times 10^2$	1.10	0.38	0.22	0.09
	$\Lambda_{el} (\text{nm})$	0.64	1.8	3.2	7.8
	$\Lambda_{in} (\text{nm})$	1.07	3.0	5.3	13
	$R (\mu\text{m})$	0.007	0.11	0.35	2.26
Gold $Z = 79$	$\sigma_{el} (\text{nm}^2) \times 10^2$	3.93	1.6	1.05	0.52
	$\sigma_{in} (\text{nm}^2) \times 10^2$	0.79	0.32	0.21	0.10
	$\Lambda_{el} (\text{nm})$	0.43	1.0	1.6	3.3
	$\Lambda_{in} (\text{nm})$	2.15	5.0	8.0	16.5
	$R (\mu\text{m})$	0.003	0.05	0.17	1.0

^a Values are listed for four different electron energies between 1 and 30 keV and three elements having a low (C), medium (Cu), and high atomic number (Au), respectively. For calculation, the following densities were used: C, $\rho = 2 \text{ g cm}^{-3}$; Cu, $\rho = 8.9 \text{ g cm}^{-3}$; Au, $\rho = 19.3 \text{ g cm}^{-3}$.

and I_{in} are shown in Table 3–3. It shows that independent on the electron energy the electron range for carbon is about one order of magnitude larger than for gold. The decrease of the electron energy from 30 to 1 keV, i.e., a factor of 30, reduces the electron range by a significantly higher factor of roughly 300.

The mean free path lengths indicate after which traveling distance on average elastic and inelastic collisions will occur. For example, in a thin organic specimen having a thickness of 50–100 nm only a few collisions on average will take place with 30-keV electrons but about seven times more with 5-keV electrons. Specimens, which have thicknesses of about $t \leq 10[\Lambda_{el}\Lambda_{in}/(\Lambda_{el} + \Lambda_{in})]$ can also be imaged in the transmission mode (cf. Figures 3–1 and 3–2) using unscattered, elastically or inelastically scattered electrons, respectively. The angular and energy distribution of the scattered electrons can be calculated by Monte Carlo simulations if the elemental composition and the density of the specimen are known (Reichelt and Engel, 1984; Krzyzanek et al., 2003).

The inelastic electron scattering events in the specimen cause secondary electrons, Auger electrons, cathodoluminescence, and X-rays, which carry a wealth of local information about the topography, the electronic structure, and the composition of the specimen. The signals, resulting from inelastic electron scattering, can also be calculated by Monte Carlo simulations.

2.2.1 Secondary Electrons

The energy spectrum of the electrons emitted from a specimen irradiated with fast electrons consists of secondary electrons, backscattered

electrons, and Auger electrons (cf. Figure 3–12). The SE show a peak at low energies with a most probable energy of 2–5 eV. By definition the maximum energy of SE amounts 50 eV. Secondary electrons are generated by inelastic scattering of the beam electrons along their trajectories within the specimen (Figure 3–14). The physics of secondary electron emission is reviewed by Kollath (1956) and Dekker (1958) but is beyond the scope of this chapter. Due to the low energy of the SE only those SE are observable that are generated within the escape depth from the surface. The actual escape depth of SE for pure elements varies with their atomic number (Kanaya and Ono, 1984). A general rule for their escape depth is $t_{\text{SE}} = 5 \Lambda_{\text{SE}}$ (Seiler, 1967), where Λ_{SE} is the mean free path of the SE. t_{SE} amounts to about 5 nm for metals and up to about 75 nm for insulators (Seiler, 1984). The angular distribution of SE follows Lambert's Law, i.e., is a $\cos \zeta$ distribution, where ζ represents the SE emission angle relative to the surface normal (Jonker, 1957; Oppel and Jahrreiss, 1972). The angular distribution of the SE is not important for the image contrast in SEM because the extraction field of the ET detector normally collects the emitted SE. The situation, however, is different in case of magnetic "through-the-lens" detection where no electric extraction field is applied (cf. Figure 3–6).

Figure 3–15 shows schematically the SE yield δ versus the energy of the beam electrons, which is the number of SE produced by one beam electron. δ increases with E_0 , reaches its maximum δ^m at $E_{0,m}$, and then decreases with further increasing E_0 . Typical values for metals are $0.35 \leq \delta^m \leq 1.6$ and $100 \text{ eV} \leq E_{0,m} \leq 800 \text{ eV}$ and for insulators $1.0 \leq \delta^m \leq 10$ and $300 \text{ eV} \leq E_{0,m} \leq 2000 \text{ eV}$ (Seiler, 1984). For $E_0 \gg E_{0,m}$ δ is proportional to $E_0^{-0.8}$ (Drescher et al., 1970), which indicates that δ is significantly smaller at 30 than at 5 keV. Both parameters, δ^m at $E_{0,m}$ depend on the ionization energy of the surface atoms (Ono and Kanaya, 1979). Figure 3–16 shows the SE yield δ versus the energy E_0 for the element copper.

There is no monotonic relation between δ and the atomic number as shown in Figure 3–17. However, published data of δ scatter which

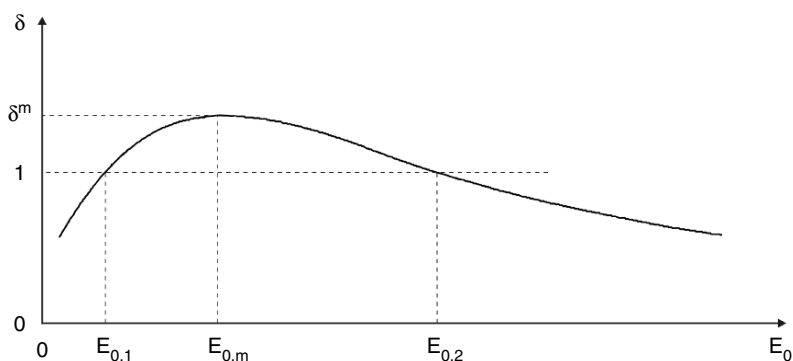


Figure 3–15. Schematic representation of the SE yield δ vs. the energy E_0 of beam electrons.

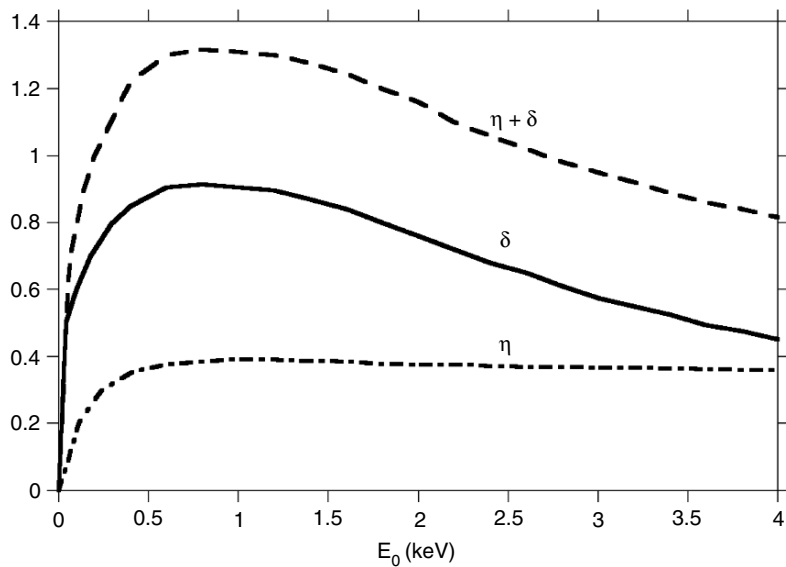


Figure 3-16. SE yield δ , BSE yield η , and $\delta + \eta$ versus the energy E_0 for polycrystalline copper at $\theta = 0^\circ$. (Data from Bauer and Seiler, 1984.)

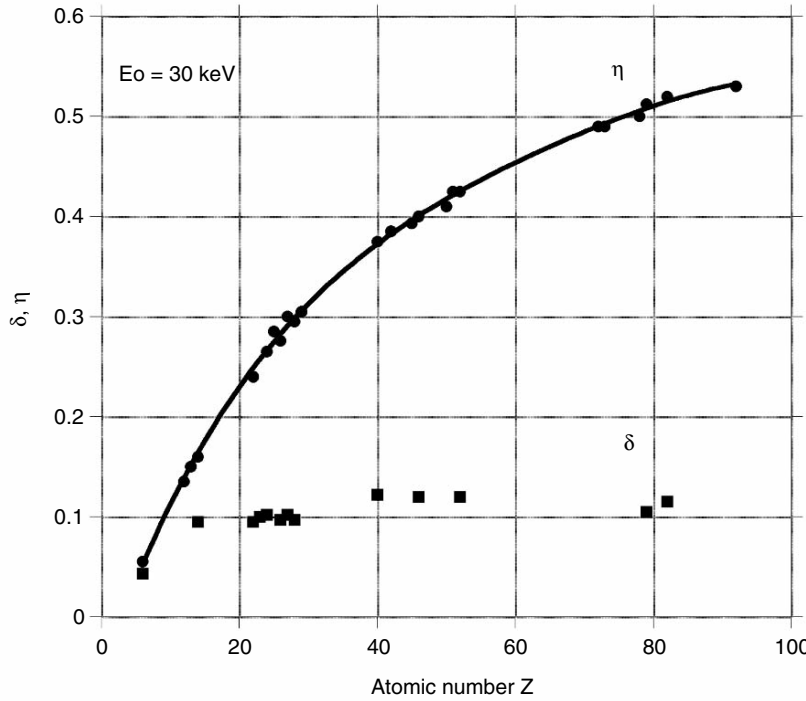


Figure 3-17. SE yield δ and BSE yield η versus atomic number Z at $E_0 = 30$ keV and $\theta = 0^\circ$. (Data from Heinrich, 1966; Wittry, 1966.)

indicates that the specimen surface conditions and the quality of the vacuum can significantly affect the secondary yield (cf. "Data Base on Elector-Solid Interactions" by Joy, 2001).

Secondary electrons generated by the incident beam electrons are designated SE1 (Drescher et al., 1970). The SE1 carry local information about the small cylindrical volume that is given approximately by the cross section of the beam $(\pi/4)d_{pe}^2$ and the escape depth t_{SE} . For a beam diameter about ≤ 1 nm the SE1 deliver high-resolution information. Those beam electrons, which are multiply scattered and emerge from the specimen as BSE, also generate secondary electrons within the escape depth. These secondary electrons are designated SE2 (Drescher et al., 1970). Their origin is far from the point of incidence of the beam caused by the spatial distribution of BSE. Changes of the amount of SE2 correlate with corresponding changes of BSE, thus SE2 carry information about the volume from which the BSE originate. The size of the volume depends on the electron range R and is much larger than the excitation volume of the SE1 for electron energies $E_0 > 1$ keV (cf. Figure 3–14 and Table 3–3); thus SE2 deliver low-resolution information. The SE yield δ consists of the contributions of SE1 and SE2 given as

$$\delta = \delta_{SE1} + \eta\delta_{SE2} \quad (2.30)$$

where η is the BSE coefficient and δ_{SE2} the SE2 yield, i.e., the number of SE2 generated per BSE. For $E_{0,m} < E_0 < 5$ keV the ratio $\delta_{SE2}/\delta_{SE1}$ amounts to about 4 and for $E_0 \geq 10$ keV about 2 (Seiler, 1967). For an increasing angle of incidence θ , this ratio decreases (Seiler, 1968).

The SE yield increases with increasing angle of incidence θ according to

$$\delta(\theta) = \delta_0/\cos\theta; \quad \delta_0 = \delta(\theta = 0) \quad (2.31)$$

(Figure 3–18). This relation is valid for a specimen with a mean atomic number, for $E_0 \geq 5$ keV, and θ up to a few degrees below 90° . The increase of δ with θ is greater for specimens with a low atomic number and smaller for samples with high Z (Reimer and Pfefferkorn, 1977). For crystalline objects, the increase of δ with θ is superimposed by electron channeling and crystalline orientation contrast (see Section 2.3). The distinct dependence of the SE yield on θ provides the basis for the topographic contrast in secondary electron micrographs.

2.2.2 Backscattered Electrons

The majority of BSE is due to multiple scattering of the beam electrons within the specimen (Figure 3–14). The energy spectrum of the backscattered electrons is shown schematically in Figure 3–12. By definition the energy of BSE is in the range $50\text{ eV} < E_{BSE} \leq E_0$. The BSE spectrum has a small peak consisting of elastically scattered electrons at E_0 (this peak is not visible in Figure 3–12). Toward energies lower than E_0 there is a broad peak, which covers the range down to about $0.7E_0$ for high atomic numbers and further down to about $0.4E_0$ for low atomic numbers. The majority of BSE are within this broad peak. For high atomic number elements such as gold, the maximum of the distinct peak is at about $0.9E_0$, whereas for low atomic numbers, e.g., carbon, the maximum of the less distinct peak is located at about $(0.5\text{--}0.6)E_0$.

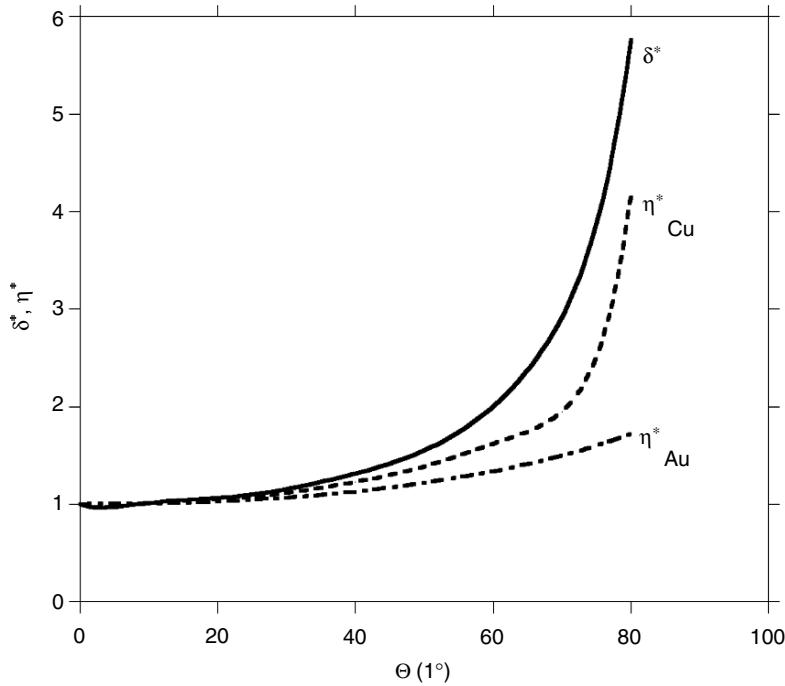


Figure 3–18. Normalized SE yield δ^* and BSE yield η^* versus the angle of incidence θ of the electron beam. $\delta^* = \delta(\theta)/\delta_0$, $\eta^* = \eta(\theta)/\eta_0$. η was calculated for gold and copper according to Eq. (2.34).

The cumulative fraction of 50% of BSE is reached for carbon at $E_{\text{BSE}}/E_0 = 0.55$ and for gold at 0.84, respectively (Goldstein et al., 2003). It seems worth mentioning that the energy distribution of BSE is shifted toward higher energy if the angle of incident electrons is larger than 70° (Wells, 1974).

As shown in Figure 3–14, the BSE can originate either from the small area directly irradiated by the electron beam—they are denoted as BSE1—or after multiple elastic and inelastic scattering events from a significantly larger circular area around the beam impact point, which are designated BSE2. The lateral distribution of BSE2 has been calculated by Monte Carlo simulation for different materials (see, e.g., Murata, 1974, 1984). It shows that the BSE-emitting surface area increases with electron energy E_0 . For a given energy E_0 the size of the BSE-emitting area increases with descending atomic number. As with SE1, the BSE1 carry local information about the small volume and deliver high-resolution information for a beam diameter of about $\leq 1\text{ nm}$. As a consequence of lateral spreading the BSE2 carry information about a much larger region, thus fine structural details on the scale of the beam diameter cannot be resolved.

Figure 3–14 also shows that the beam electrons travel in a small subsurface volume before they return to the surface to escape as BSE2. The escape depth of BSE is much larger than t_{SE} and depends on—in

contrast to t_{SE} —the electron energy E_0 . Experimental data for different materials show that t_{BSE} amounts to about half of the electron range R (Drescher et al., 1970; Seiler, 1976).

Knowledge of the angular distribution of BSE is of great importance for understanding and optimization of BSE detection geometry. For normal beam incidence the angular distribution can be approximated by a $\cos\zeta$ distribution (Drescher et al., 1970), where ζ represents the BSE emission angle relative to the surface normal. Due to the fact that the emitted BSE move on nearly straight trajectories, the angular detector position has a strong influence on the collection efficiency of the detector. For nonnormal beam incidence the distribution is asymmetric and a reflection-like emission maximum is observed. The angular distribution consists for large angles of incidence θ of cosine distribution approximately directed to $-\theta$ and a superimposed fraction at smaller emission angles (Drescher et al., 1970).

The BSE coefficient η is defined by

$$\eta = n_{\text{BSE}}/n_{\text{b}} \quad (2.32)$$

where n_{BSE} is the number of BSE and n_{b} is the number of incident electrons. η is approximately independent of the electron energy E_0 in the range of about 10–30 keV. For low atomic numbers and beam energies below 5 keV η increases as E_0 decreases, whereas for medium and high atomic numbers η decreases with E_0 (cf. Figure 3–16) (Reimer and Tollkamp, 1980). However, at low energies η depends in a complex manner on the atomic number (Heinrich, 1966; cf. “Data Base on Electron-Solid Interactions” by Joy, 2001).

The BSE coefficient monotonically increases with the atomic number as shown for 30 keV in Figure 3–17. Because of the approximate independence of the electron energy E_0 , the graph of the BSE coefficient is valid for beam energy ranging from 30 down to about 5 keV. The graph of η versus Z can be approximated by a polynomial (Reuter, 1972)

$$\eta(Z) = 0.0254 + 0.016Z - 1.86 \times 10^{-4}Z^2 + 8.31 \times 10^{-7}Z^3 \quad (2.33)$$

For energies below 5 keV the dependence of η on Z is more complicated (for details see Hunger and Kuchler, 1979; Joy, 1991; Zadrazil et al., 1997). The distinct dependence of the BSE coefficient on the atomic number Z provides the basis for the atomic number contrast (see Section 2.3).

Like the SE yield, the backscattering coefficient also increases monotonically with increasing angle of incidence θ according to (Arnal et al., 1969)

$$\eta(\theta) = (1 + \cos\theta)^{-9/\sqrt{Z}} \quad (2.34)$$

Figure 3–18 shows the graphs $\eta(\theta)$ versus θ for Cu ($Z = 29$) and Au ($Z = 79$). The graphs indicate the strong influence of the atomic number, in particular for $\theta > 50^\circ$. The monotonic increase of η with θ provides the basis for the topographic contrast in BSE micrographs. For the sake of completeness it should be mentioned that Drescher et al. (1970) derived from experimental data at 25 keV an analytical expression for $\eta(\theta, Z)$ other than the one given by Eq. (2.34).

The backscattering coefficient of a single crystal depends sensitively on the direction of the incident electrons related to the crystal lattice (Reimer et al., 1971; Seiler, 1976). This dependence is caused by the regular three-dimensional arrangement of the atoms in the lattice, whose atomic density depends on the direction. The backscattering coefficient is lower along directions of low atomic density, which permits a fraction of the incident electrons to penetrate deeper than in amorphous material before being scattered. Those electrons have a reduced probability of returning to the specimen surface and leaving the sample as BSE. The maximum relative variation of the backscattering coefficient is in the order of 5%.

2.2.3 *Transmitted Electrons*

When the thickness of a specimen approaches the electron range R or becomes even smaller than R , an increasing fraction of beam electrons is transmitted. Those transmitted electrons interact with the support, thus generating non-specimen-specific signals, which superimpose the specimen-specific signals. The spurious contribution of the support to the signal, originating from the specimen, can be reduced significantly by replacing the solid support by a very thin (about 5–15 nm thick) amorphous carbon film. Such thin carbon films supported by a metallic mesh grid are commonly used in TEM and STEM as electron-transparent support for thin specimens. To improve the stability of the 5-nm-thick carbon film, the film is placed onto a holey thick carbon film supported by a mesh grid. In contrast to a solid support, a 5-nm-thick carbon film contributes only insignificantly to the SE and BSE signal (cf. Figure 3–19), thus particles deposited onto a thin support film can be imaged in the normal manner using SE and BSE, respectively.

In addition to the reduction of spurious signal, the transparent support film enables use to be made of the transmitted electrons, which carry information about the interior of the specimen (in some SEM the specimen stage must be altered to make the transmitted electrons accessible). As a result of electron–specimen interaction the transmitted electrons can be unscattered or elastically or inelastically scattered (cf. Figure 3–2). Due to their characteristic angular and energy distribution, the transmitted electrons can be separated by placing suitable detectors (preferentially combined with an electron spectrometer) below the specimen. Frequently, a rather simple and inexpensive device for observing an STEM image (Oho et al., 1986)—sometimes called “poor man’s STEM in SEM detector”—is used. The transmitted electrons are passing through an angle-limiting aperture, strike a tilted gold-coated surface, and thus create a high SE and BSE signal, which can then be collected by a conventional ET detector. The angle-limiting aperture cuts off the transmitted, scattered electrons. In this case the “poor man’s STEM in SEM detector” acquires those electrons, which represent the bright-field signal. The “poor man’s STEM in SEM detector” just cuts the transmitted scattered electrons without making use of their inherent information. Both the elastically and the inelastically scattered electrons are signals, which very sensitively depend on the

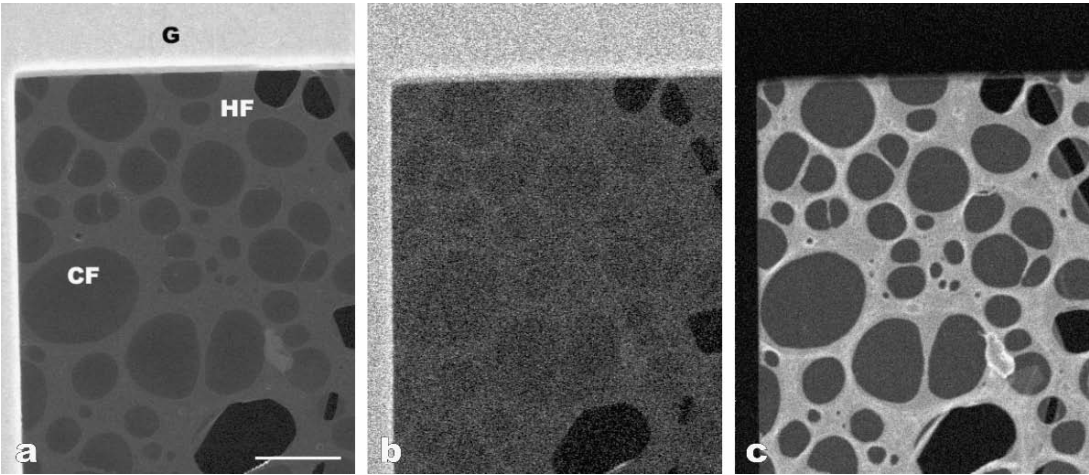


Figure 3–19. SE and annular dark-field micrographs showing a particle on an ~4-nm-thick amorphous carbon film (CF) attached to a holey carbon film of ~20 nm in thickness (HF), which is supported by a Cu mesh grid (G). (a) SE micrograph recorded with a probe current of about 30 pA at 30 kV. A very low probe current of ~3 pA (pixeltime: 23 μ s. i.e., ~430 incident electrons/pixel) was used for simultaneously recording the SE micrograph (b) and the dark-field micrograph with scattered transmitted electrons (annular dark-field mode) (c). The image intensity and the signal-to-noise ratio in (c) is much higher than the intensity in (b) because the number of SE (almost only SE1 are generated) is significantly smaller than the number of the collected transmitted scattered electrons. No usable BSE signal can be detected from the thick and thin carbon films. The scale bar corresponds to 20 μ m.

mass thickness ρt if the specimen thickness $t \leq [\Lambda_{el}\Lambda_{in}/(\Lambda_{el} + \Lambda_{in})]$ (Reichelt and Engel, 1984, 1985) (Figure 3–20) can be used, e.g., for mass determination of biomolecules and assemblies thereof (Zeitler and Bahr, 1962; Lamvik, 1977; Engel, 1978; Wall, 1979; Feja et al., 1997). Although most of the mass determination studies were made with dedicated STEM, high-resolution FESEM equipped with an efficient

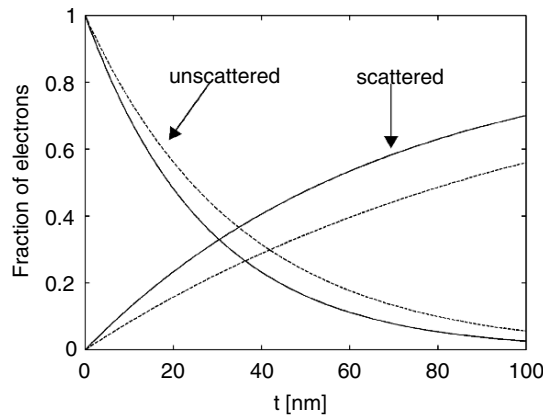


Figure 3–20. Fraction of transmitted electrons scattered into an angular range of 25–300 mrad for carbon ($\rho = 2 \text{ g cm}^{-3}$; solid line) and protein ($\rho = 1.35 \text{ g cm}^{-3}$; dashed line). Parameters: $E_0 = 30 \text{ keV}$, $\alpha_p = 10 \text{ mrad}$. The graphs show an increasing fraction of scattered and a decreasing fraction of unscattered electrons with increasing thickness. (Calculation according to Krzyzanek and Reichelt, 2003.)

annular dark-field detector capable of single-electron counting and MHz counting rates would allow for such quantitative studies (Reichelt et al., 1988; Krzyzanek and Reichelt, 2003; Krzyzanek et al., 2004).

Another important application of STEM imaging is the measurement of the physical probe size of the SEM using a thin carbon film (thickness below 10 nm, preferably containing nanoparticles of gold or gold-palladium for better contrast). In this case the broadening of the electron beam in the film is negligible and so the resolution of the STEM image is equal to the probe diameter. The image resolution can be determined either by analysis of the diffractogram (power spectrum) of the STEM micrograph (Frank et al., 1970; Reimer, 1985; Joy, 2002) or by cross-correlation function analysis (Frank, 1980) of the phase noise in the bright-field STEM image of the carbon film. The latter directly yields the probe diameter of the SEM (Joy, 2002).

Combining the SE and BSE detectors above as well as the bright-field and dark-field detectors below the specimen, its surface as well as its internal structure can be observed simultaneously in the SEM.

2.2.4 Cathodoluminescence

Cathodoluminescence (CL) is the emission of light generated by the electron bombardment of semiconductors and insulators (Muir and Grant, 1974; cf. Figure 3–14). Those materials have an electronic band structure characterized by a filled valence band and an empty conduction band separated by an energy gap $\Delta E_{CV} = E_C - E_V$. Electrons from the valence band can interact inelastically with a beam electron and can be excited to an unoccupied state in the conduction band. The excess energy of the excited electron will be lost by a cascade of nonradiative phonon and electron excitations. Most of the recombination processes of excited electrons with holes in the valence band are nonradiative processes, which elevate the sample temperature. There are different radiative processes, which take place in inorganic materials, semiconductors, and organic molecules.

In inorganic materials intrinsic and extrinsic transitions can take place. The intrinsic emission is due to direct recombination of electron hole pairs. Extrinsic emission is caused by the recombination of trapped electrons and holes at the donor and acceptor level, respectively. The trapping increases the probability of recombination. The extrinsically emitted photons have a lower energy than intrinsically emitted photons.

In semiconductors the radiative recombination can be due to the direct collision of an electron with a hole with the emission of a phonon. Depending on the nature of the band structure of the material, the recombination can be either direct or indirect. In the latter case the recombination must occur by simultaneous emission of a photon. Indirect recombination is less likely than direct recombination. If the material contains impurities, the process of recombination via impurity level becomes important. The modification of CL efficiency as a function of the purity and the perfection of the material is the most important aspect of the use of this method in scanning electron

microscopy. It is because of such modifications that a contrast is generated (for details see, e.g., Holt and Yacobi, 1989; Yacobi, 1990). It was shown in some cases that the sensitivity of CL analyses can be at least 10^4 times higher than that obtainable by X-ray microanalysis, i.e., an impurity concentration as low as 10^{14}cm^{-3} (Holt and Saba, 1985).

In organic materials the excitation is inside an individual molecule. Electrons go from a ground state to a singlet state at least two states above. Then the deexcitation to the ground state is radiationless up to the singlet state directly above the ground level and from this state the deexcitation can be either radiationless or radiative with decay times larger than 10^{-7}s (fluorescence). The CL spectra depend on the chemical structure of the molecule (DeMets et al., 1974; DeMets, 1975). Cathodoluminescence of organic matter also can be caused by selective staining with luminescent molecules (fluorochromes). Typical fluorochromes are, e.g., fluoresceine, fluoresceine isothiocyanate (FITC), and acridine orange.

Independent on the material the light generated by CL inside the specimen has to pass the surface according to the Snell law (Bröcker et al., 1977). The critical angle θ_i of total internal reflection is given as

$$\sin \theta_i = n_1/n \quad (2.35)$$

where $n_1 = 1$ (vacuum) and n is the refractive index of the specimen ($1 < n < 5$). As shown by Eq. (2.35) the fraction of emitted light is significantly reduced by total internal reflection for $n > 2$ (semiconductors).

2.2.5 X-Rays

The X-ray spectrum is considered to be that part of the electromagnetic spectrum that covers the wavelengths λ_x from approximately 0.01 to 10 nm. The energy of X-rays is given as

$$E_x = h\nu = hc/\lambda_x \quad (2.36)$$

where $h = 6.6256 \times 10^{-34}\text{Js}$ is Planck's constant, $c = 2.99793 \times 10^8\text{m/s}$ is the speed of light, and ν is the frequency of X-rays. The X-rays are generated by deceleration of electrons (X-ray continuum or bremsstrahlung) or by electron transition from a filled higher state to a vacancy in a lower electron shell (characteristic X-ray lines) (Figure 3-21).

The X-ray continuum is made up of a continuous distribution of intensity as a function of energy whereas the characteristic spectrum represents a series of peaks of variable intensity at discrete element-specific energies. As the electron energy increases the intensity of the continuous spectrum also increases and the maximum of the distribution is shifted toward higher energies. The general appearance of the continuous spectrum is independent of the atomic number of the specimen, however, the absolute intensity values are dependent on the atomic number. The maximum possible energy E_x is given by the electron energy E_0 , which corresponds to instantaneous stopping of an electron at a single collision (Duane-Hunt limit). According to Kramers

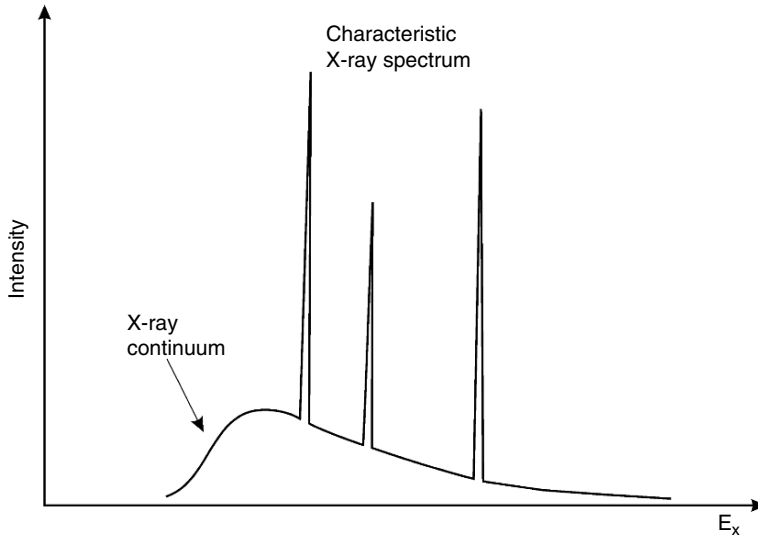


Figure 3-21. Schematic representation of the X-ray spectrum emitted from a specimen bombarded with fast electrons.

(1923) the intensity of the continuous spectrum I_c emitted in an energy interval with the width dE_x is given as

$$I_c(E_x)dE_x = kZ(E_0 - E_x)/E_x \cdot dE_x \quad (2.37)$$

k represents the Kramers constant, which varies slightly with the atomic number (Reimer, 1985). A detailed treatment of the continuous X-ray emission is given by Stephenson (1957).

The characteristic X-ray spectrum consisting of peaks at discrete energies is superimposed on the continuous X-ray spectrum (cf. Figure 3-21). Their positions are independent of the energy of the incident electrons. The peaks occur only if the corresponding atomic energy level is excited. The generation of characteristic X-rays consists of three different steps. First, a beam electron interacts with an inner shell electron of an atom and ejects this inner shell electron leaving that atom in an excited state, i.e., with a vacancy on the electron shell. Second, subsequently the excited atom relaxes to the ground state by transition of an electron from an outer to an inner shell vacancy. The energy difference ΔE_{ch} between the involved shells is characteristic for the atomic number. Third, this element-specific energy difference is expressed either by the emission of an electron of an outer shell with a characteristic energy (Auger electron) or by the emission of a characteristic X-ray with energy $E_x = \Delta E_{ch}$. The fraction of characteristic X-rays emitted when an electron transition occurs is given by the fluorescence yield ω . This quantity increases with the atomic number and depends on the inner electron shell involved (Figure 3-22). The complement, $1 - \omega$ represents the Auger electron yield, which gives the corresponding fraction of Auger electrons produced. The fluorescence yield for the different shells and subshells can be calculated (for details see Bambynek et al., 1972).

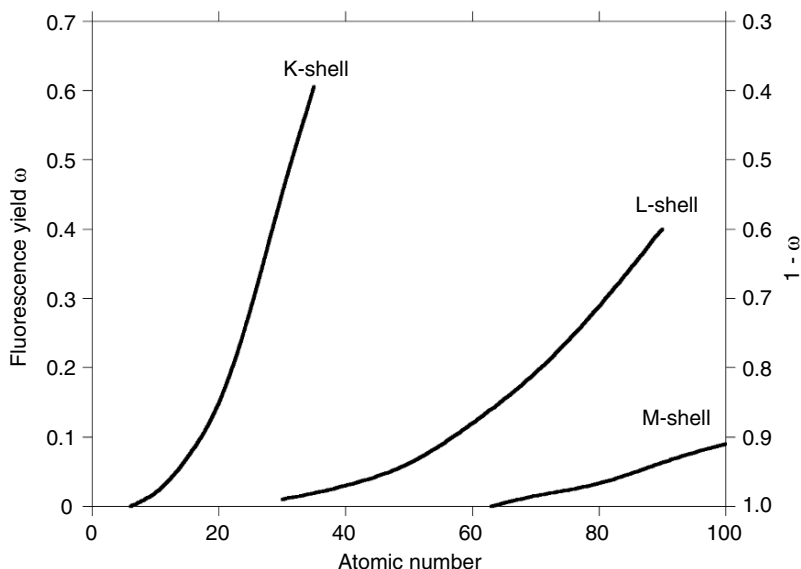


Figure 3-22. Dependence of the X-ray fluorescence yield ω and its complement ($1 - \omega$) of the K-, L-, and M-shell from the atomic number. The complement ($1 - \omega$) corresponds to the Auger electron yield.

Moseley studied the line spectra in detail and found that the general appearance of the X-ray spectrum is the same for all elements. The energy of a characteristic X-ray line depends on the atomic shells involved in the transition resulting in the emission of this line. The X-ray lines can be classified in series according to the shell where the ionization took place, e.g., K-, L-, M-shell, etc. The quantum energies of a series are given by Moseley's law

$$E_x = A(Z - B)^2 \quad (2.38)$$

where A and B are parameters that depend on the series to which the line belongs. The characteristic X-ray energy E_x is denoted by symbols that identify the transition that produced it. The first letter, e.g., K, L, identifies the original excited level, whereas the second letter, e.g., α , β , designates the type of transition occurring. For example, K_α denotes the excitation energy between the K- and L-shell, whereas K_β denotes the excitation energy between the K- and M-shell. Transitions between subshells are designated by a number, e.g., a transition from the subshell L_{III} to K is denoted as $K_{\alpha 1}$ and from the subshell L_{II} to K is denoted as $K_{\alpha 2}$, respectively. The transition from the subshell L_I to K is forbidden. The characteristic X-ray energies and X-ray atomic energy levels for the K-, L-, and M-shells are listed in tables (Bearden, 1967a,b). Fortunately, the atomic energy levels are not strongly influenced by the type and strength of the chemical bonds. However, chemical effects on X-ray emission are observed for transitions from the valence electron states, which are involved in chemical bonds. In such cases the narrow lines show changes of their shape and their position (energy shift $< 1\text{ eV}$) as well (Baun, 1969).

2.2.6 Auger Electrons

As mentioned in Section 2.2.5, when an excited atom relaxes to the ground state by transition of an electron from an outer to an inner shell the energy difference ΔE_{ch} between the involved shells can be expressed by the emission of an electron of an outer shell with a characteristic energy E_{AE} . The emission is due to the Auger effect (Burhop, 1952; Åberg and Howart, 1982) and the emitted electron is designated as an Auger electron (AE). Its energy E_{AE} is given by

$$E_{\text{AE}} = \Delta E_{\text{ch}} - E_{\text{ionr}} \quad (2.39)$$

where the term E_{ionr} contains the ionization energy of the AE emitting outer shell and also considers relaxation effects. The shape and the position of the AE peaks are influenced by the type and strength of the chemical bonds (Madden, 1981). The AE peaks have an energy width of a few electronvolts and are superimposed on the low-energy range of the BSE spectrum up to energies of about 2.5 keV (Figure 3–12). The identification of the AE peaks on the BSE background (Bishop, 1984) can be improved by differentiation of the electron energy spectrum.

The Auger electrons are generated within the excitation volume (Figure 3–14). Due to their low energies only AE with a short pathway to the specimen surface can escape. However, energy losses caused by inelastic scattering on the pathway to the surface remove AE from the AE peaks. The decrease of the AE peak is proportional to $\exp(-x/\Lambda_{\text{AE}})$, where x denotes the length of the path inside the specimen and Λ_{AE} the mean free path of the AE. Depending on their energy and the atomic number of the specimen the mean free path Λ_{AE} can have values in the range of about 0.4 nm to a few nanometers (Palmberg, 1973; Seah and Dench, 1979). If AEs are inelastically scattered on their path to the surface, then they cannot be identified as an AE in the BSE background. Therefore, only atoms within a depth of about Λ_{AE} can contribute to the AE peaks. Since Auger electrons yield information on element concentrations very near the surface, the specimen must be in an ultra-high vacuum environment and special sample preparations are required (e.g., ion sputtering *in situ*, cleavage *in situ*) to obtain clean surfaces.

Because the AE yield of the K-shell $1 - \omega_{\text{K}}$ is much larger than ω_{K} for light elements (cf. Figure 3–22) AE spectroscopy is advantageous for elements with atomic number $Z = 4 (\omega_{\text{K}} = 4.5 \times 10^{-4})$ up to $Z \sim 30 (\omega_{\text{K}} = 4.8 \times 10^{-1})$ (Bauer and Teliëps, 1988). Similar to SE, which are also emitted only from a very thin surface layer, the AE can be generated directly by the beam electrons and by BSE within a larger circular area around the beam impact point (cf. Figure 3–14). Like the SE yield, which increases with increasing angle of incidence θ according to $\delta(\theta) = \delta_0/\cos\theta$ [cf. Eq. (2.31)], the integral AE peak intensity is proportional to $1/\cos\theta$ (Bishop and Riviere, 1969; Kirschner, 1976; Reimer, 1985).

Recent developments in scanning Auger microscopy and AE spectroscopy are described by Jacka (2001).

2.2.7 Others

The incident electron beam bombards the specimen with electrons thereby introducing a negative electric charge. A certain amount of negative electric charge is leaving the specimen as secondary (I_{SE}), backscattered (I_{BSE}) and Auger electrons (I_{AE}). To avoid an accumulation of charges a specimen current I_{sp} must flow from the specimen to the ground. The conservation equation for the electric charge is

$$I_p = I_{SE} + I_{BSE} + I_{AE} + I_{sp} \quad (2.40)$$

where I_p is the probe current. The specimen current changes the sign when $I_{SE} + I_{BSE} + I_{AE} > I_p$. Because $I_{SE} + I_{BSE} \gg I_{AE}$ this means basically that $\delta + \eta > 1$ (cf. Figure 3–16). I_{sp} depends on the angle of beam incidence θ and the electron energy E_0 as expected from $\delta(\theta, E_0)$ and $\eta(\theta, E_0)$ (Reimer, 1985). The resolution of specimen current images is comparable to that of BSE images. One advantage of the specimen current mode is that the contrast is independent on the detector position. A critical review of this mode was published by Newbury (1976).

The electron–specimen interaction also generates acoustic waves. The frequencies of these waves depend on the imaging conditions of the SEM and the specimen studied. They cover a range from low sound up to very high ultrasound frequencies. The electron acoustic mode, usually denoted as scanning electron acoustic microscopy (SEAM), was introduced by Brandis and Rosencwaig (1980) and Cargill (1980). SEAM uses a periodic beam modulation or short electron beam pulses that allow for analysis of the SEAM frequency response (Balk and Kultscher, 1984; Balk, 1986; Kultscher and Balk, 1986). The SEAM was reviewed by Balk (1989) and applications of this method in semiconductor research (Balk, 1989) and for the investigation of magnetic structures were shown (Balk et al., 1984).

2.3 Contrast Formation and Resolution

Since the image formation is due to the image signal fluctuation ΔS from one point to another point, the contrast C is designated as in television to be

$$C = (S - S_{av})/S = \Delta S/S \quad (2.41)$$

S_{av} is the average value of the signal and S represents the signal of the considered point ($S > S_{av}$ i.e., C is always positive). The signal fluctuation may be caused by local differences in the specimen topography, composition, lattice orientation, surface potential, magnetic or electric domains, and electrical conductivity. The minimum contrast is obtained if $S = S_{av}$, whereas the maximum contrast is obtained for $S_{sv} = 0$. This is the case, e.g., when the signal S from a feature is surrounded by a background with $S_{av} = 0$. The contrast will be visible if C exceeds the threshold value of about 5×10^{-2} .

According to the point-resolution criterion two image points separated by some horizontal distance (i.e., within the x–y plane perpendicular to the optical axis) are resolved when the minimum intensity

between them is 75% or less of the maximum intensity. The point resolution therefore corresponds to the minimum distance of two object points, those superimposed image intensity distributions drop to 75% of their maximum intensity between them. Due to the inherent noise of each signal of the SEM characterized by its signal-to-noise ratio (SNR) the drop to 75% of the maximum intensity will not be reliably defined at the minimum distance. Consequently, at low SNR two image points can be resolved only if their distance is larger than the minimum distance reliably defined for noiseless signals.

As opposed to the light or transmission electron microscope the resolution of the SEM cannot be defined by Rayleigh's criterion. The resolution obtained in the SEM image depends in a complex manner on the electron beam diameter, the electron energy, the electron-specimen interaction, the selected signal, the detection, as well as the electronic amplification and electronic processing. An object "point" corresponds to the size of a small local excitation volume (cf. Figures 3-13 and 3-14) designated as the spatial detection limit from which a sufficient signal can be obtained. Obviously, the point resolution cannot be less than the spatial detection limit. It becomes clear from Figure 3-14 that the spatial resolution of an SEM is different for each signal since the size of the signal emitting volume as well as the signal intensity depends on the type of signal selected.

The important "quality parameters" such as spatial resolution, astigmatism, and SNR of SEM images, as well as drift and other instabilities that occur during imaging, can be determined most reliably and objectively by Fourier analysis of the recorded micrographs (Frank et al., 1970; Reimer, 1985). Recently, a program SMART (Scanning Microscope Analysis and Resolution Testing) became freely available, which allows the SEM resolution and the imaging performance to be measured in an automated manner (Joy, 2002). It should be mentioned in the context of resolution that the ultimate resolution of an SEM specified by the manufacturers is determined by imaging a test sample (gold or gold-palladium-coated substrate with a low atomic number). However, such samples are quite atypical compared to those usually investigated, thus the resolution determined in this way is not properly representative of the routine performance of the SEM. At present, conventional SEMs using a heated tungsten or lanthanum hexaboride emitter as the electron source have a specified resolution in the range of 3–5 nm at an acceleration voltage of 30 kV.

2.3.1 Topographic Contrast

Presumably the SEM is most frequently used to visualize the topography of three-dimensional objects. The specimen topography gives rise to a marked topographic contrast obtained in secondary and back-scattered images. This contrast has a complex origin and is formed in SE images by the following mechanisms:

1. Dependence of the SE yield δ on the angle of incidence θ of the electron beam at the local surface element [cf. Eq. (2.31)]. The tilt angle of the local surface elements is given by the topography of the sample.

2. Dependence of the detected signal on the angular orientation of the local surface element related to the ET detector (see Section 2.1.3). SE generated “behind” local elevations, in holes, in fissures, or in cavities reach the ET detector incomplete. This causes a more or less pronounced shadow contrast (cf. Figure 3–7a).

3. Increase of the SE signal when diffusely scattered electrons pass through an increased surface area. This is the case at edges or at protruding surface features, which are smaller than the excitation volume. Electron diffusion leads to overbrightening of edges and small surface protrusions in the micrograph and is known as an edge effect.

4. Charging artifacts with objects of low electric conductivity.

Contributions (1) to (3) are illustrated by SE micrographs of different specimens shown in Figures 3–23a and 3–24a as well as schematically by profiles of the topography and the related SE signals in Figure 3–25. In these figures the direction to the ET detector is indicated. The ball in Figure 3–23a shows a contrast, which is mainly due to the varying angle θ of beam incidence across the ball [cf. (1) above] and the angular orientation of the local surface elements related to the ETD [cf. (2) above]. The collection efficiency of the ETD is significantly higher for surface elements facing the detector than for those on the back (shadow region). Whereas the intensity of emitted secondary electrons of the ball reveals radial symmetry, the effect of detection geometry causes the nonradial symmetric image intensity distribution of the ball (cf. Figure 3–25a–c). The rim of the ball is bright in the SE image because of the enhanced SE emission due to an incidence angle $\theta \approx 90^\circ$ and the effect of diffusely scattered electrons passing through an increased surface area [cf. (3) above]. The radius of the ball is larger than the electron range R (cf. Figure 3–14) therefore the latter effect occurs just near the rim of the ball. If the mean radius of ball-like particles becomes comparable or smaller than the electron range, diffusely scattered electrons generate more SE over the whole particle surface, thus the SE emission typically is distinctly enhanced (small particles are marked by small arrowheads in Figures 3–23a and 3–24a). If the shadow contrast is visible in the image and the direction toward the ETD is known then elevations and depressions clearly can be readily identified (cf. Figure 3–23a). Another way to distinguish elevations and depressions is to record and to analyze SE stereopairs.

The SE micrograph of large crystal-like particles (Figure 3–24a) basically shows the same contrast mechanisms as discussed above but with a more complex structured sample than the ball. The individual flat surface planes of the crystal-like particles occur with almost constant brightness because of the constant angle of beam incidence and the constant detection geometry (provided that there is no shadow effect from other large particles). Some surface planes possess fissures of different size, which typically appear rather dark because just a minor fraction of the generated SE can escape from inside the fissures. In such cases SE can be extracted either by a positively biased grid in front of the specimen (Hindermann and Davis, 1974) or by a superimposed magnetic field in which the SE follow spiral trajectories around the

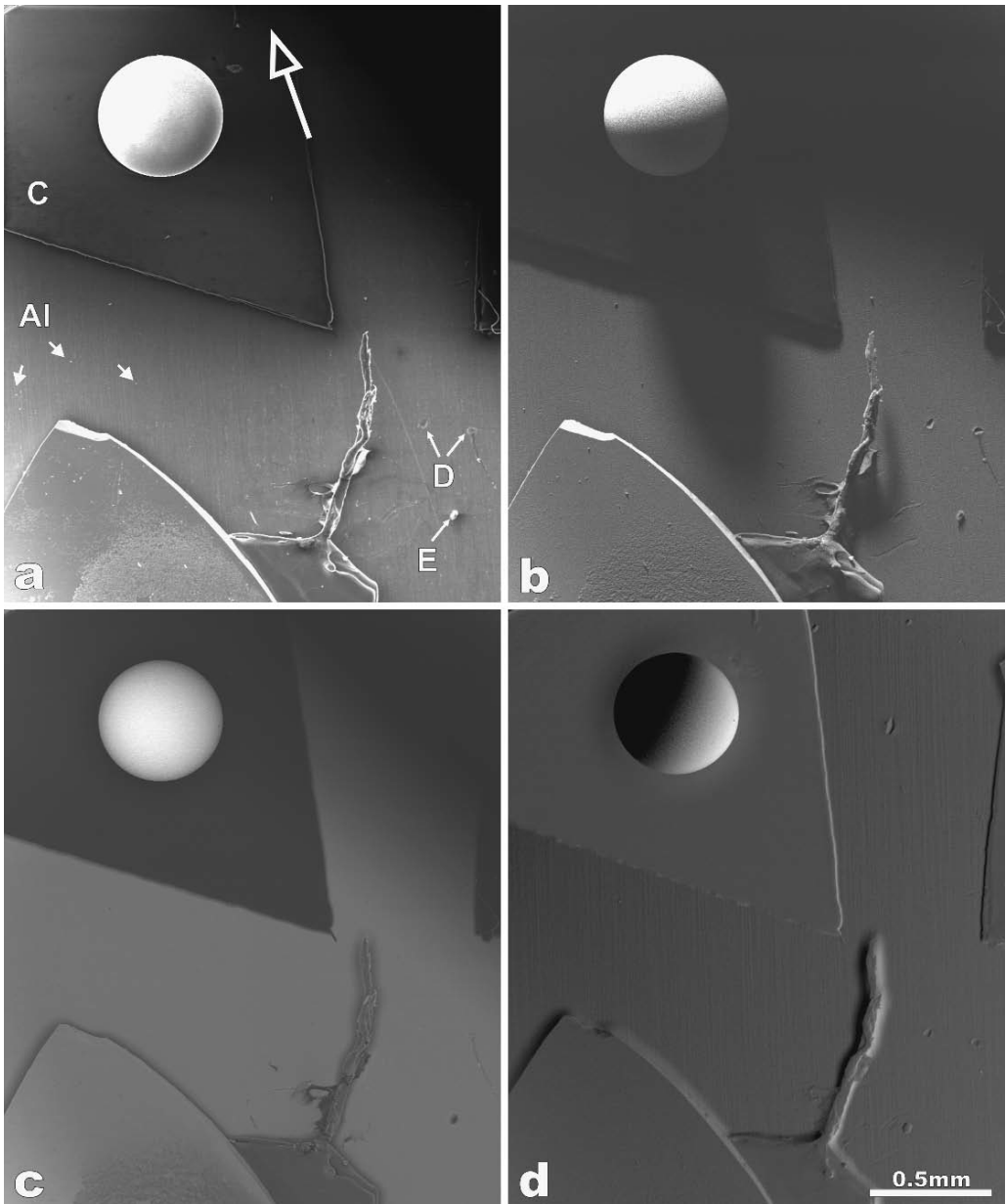


Figure 3-23. Secondary (a) and backscattered electron micro graphs (b–d) of a steel ball recorded at 30kV and normal beam incidence. The arrow in (a) indicates the direction of the laterally located ET detector. The BSE micrographs shown in (c and d) were acquired using a four-quadrant semiconductor detector mounted below the objective pole piece, which records BSE over a large solid angle. The steel ball is mounted on carbon (marked by C), which is supported by aluminum (marked by Al). The small arrowheads in (a) indicate small particles with enhanced SE emission (bright blobs in the SE image). Elevations (E) and depressions (D) are also marked by small arrowheads.

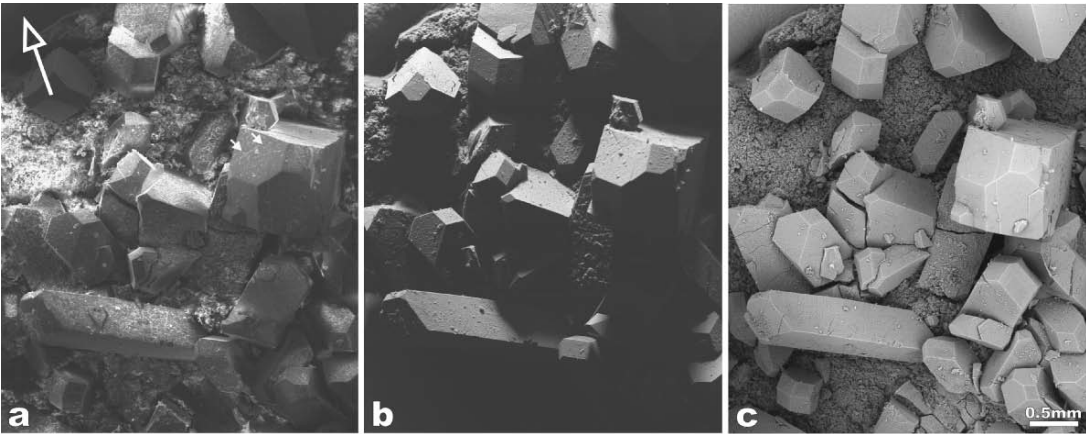


Figure 3–24. Secondary (a) and backscattered electron micrographs (b and c) of 10-nm gold-coated crystal-like tartar (tartar contains mostly potassium–hydrogen–tartrate and calcium–tartrate) recorded at 30kV and normal beam incidence. The arrow in (a) indicates the direction of the laterally located ET detector. The BSE micrograph shown in (c) was acquired using a four-quadrant semiconductor detector mounted below the objective pole piece, which records BSE over a large solid angle. The small arrowheads in (a) indicate small particles with enhanced SE emission (bright blobs in the SE image).

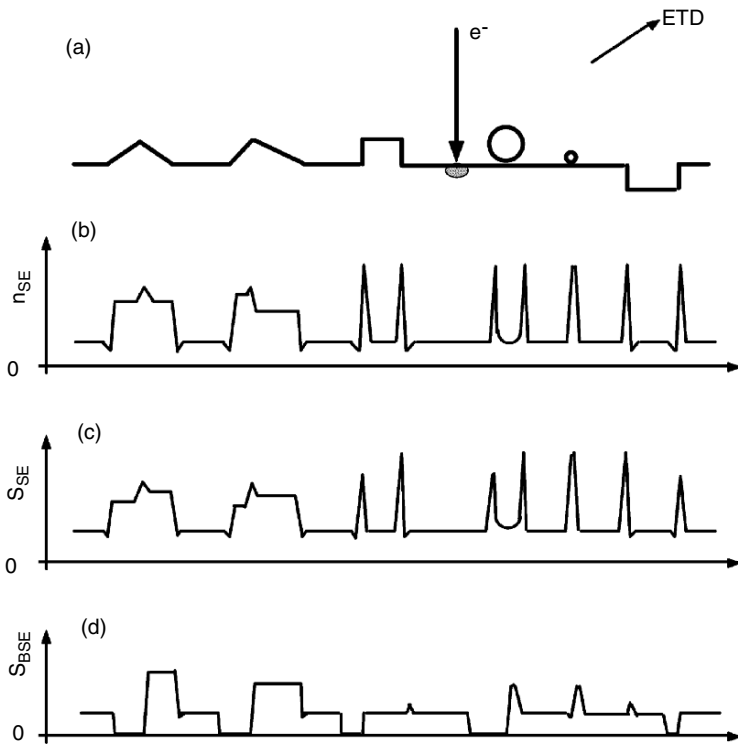


Figure 3–25. Schematic specimen surface profile of an assumed topography having elementally shaped elevations and a depression (a). Those elemental shapes are present in the samples shown in Figures 3–23 and 3–24. The size of the excitation volume of the electron beam is drawn in relation to the local topographic structures. The amount n_{SE} of locally emitted SE is shown qualitatively in (b) and the corresponding SE signal S_{SE} in (c). The BSE signal S_{BSE} collected by the negatively biased ET detector is schematically presented by the graph in (d). ETD, Everhart–Thornley detector.

lines of magnetic flux until they reach the collecting field of the ETD (Lukianov et al., 1972).

It should be mentioned that the laterally located ETD also register those BSE, which are within the small solid angle of collection defined by the scintillator area and the specimen–scintillator distance. The BSE contribute in the order of 10–20% to the SE signal (Reimer, 1985) and are the same as those collected by the negatively biased ET detector. Furthermore, BSE that are not intercepted by the detector strike the pole piece of the objective lens and the walls of the specimen chamber. These stray BSE generate so-called SE3 emitted from the interior surfaces of the specimen chamber. The SE3 carry BSE information and form a significant fraction of the SE signal for specimens with an intermediate and high atomic number (Peters, 1984).

The contrast in BSE images is formed by the following mechanisms:

1. Dependence of the BSE coefficient η on the angle of incidence θ of the electron beam at the local surface element [cf. Eq. (2.34)].
2. Dependence of the detected signal on the angular orientation of the local surface element related to the BSE detector (see Section 2.1.3). BSE emitted “behind” local elevations, in holes, or in cavities, which do not reach the BSE detector on nearly straight trajectories, are not acquired. This causes a pronounced shadow contrast (cf. Figure 3–7b).
3. Increase of the BSE signal when diffusely scattered electrons pass through an increased surface area. This is the case at edges or at protruding surface features, which are smaller than the excitation volume.

The BSE leave the specimen on almost straight trajectories and only those within the solid angle of collection of the BSE detector can be recorded. Thus, dedicated BSE detectors have a large solid angle of collection to record a significant fraction of the BSE and to generate a signal with a sufficient SNR. The larger the solid angle of collection the less pronounced the shadow effects.

Contributions (1) to (3) mentioned above are illustrated by BSE micrographs from two specimens used for SE imaging and are shown in Figures 3–23b and c and 3–24b and c as well as schematically by profiles of the topography and the related BSE signals in Figure 3–25. Two different types of BSE images are shown: the highly directional image recorded with the negatively biased ETD (Figures 3–23b and 3–24b) and the “top-view” image recorded with the four-quadrant semiconductor detector mounted below the objective pole piece (Figures 3–23c and 3–24c). The ball in Figure 3–23b shows a contrast, which is mainly due to the varying angle θ of beam incidence across the ball [cf. (1) above] and the angular position of the local surface elements related to the negatively biased ETD [cf. (2) above]. A pronounced sharp shadow occurs at the back of the ball and behind the ball (shadowed oblong area of the support). Whereas the intensity of the BSE of the ball reveals radial symmetry, the effect of detection geometry causes the nonradial symmetric image intensity distribution of the ball (cf. Figure 3–25a and d). The fade contour of the ball at its back is due

to BSE redirected toward the negatively biased ETD by scattering on some interior surfaces of the specimen chamber. The pronounced directional shadow contrast in the image allows for unambiguous identification of elevations and depressions (cf. Figure 3–23b). Moreover, if the detection geometry of the BSE is known, the length of the shadow can be used in some cases to obtain a rough estimate of the height of elevations or depth of depressions.

The BSE micrograph of the ball recorded with the four-quadrant semiconductor detector (Figure 3–23c) shows an almost radial symmetric image intensity distribution. It is obvious that the increase of the BSE coefficient η with the increasing angle of incidence θ (see Figure 3–18) toward the rim of the ball is superimposed by the stronger counteracting effect of the directed asymmetric distribution, for a large θ reflection-like angular distribution of BSE for nonnormal beam incidence (cf. Section 2.2.2). The shadow-like hem along the contour of the elevations reflects the fact that BSE emitted from the lower surrounding areas toward elevations can be absorbed or redirected; thus those BSE do not reach the BSE detector. In case of depressions there is also a shadow-like hem but it is located inside the contour of the depression. A comparison of the different types of BSE images in Figure 3–23b and c clearly shows that the topography of the sample is pronounced in Figure 3–23b while—as we shall see later—the atomic number contrast is pronounced in Figure 3–23c.

The BSE micrographs of large crystal-like particles (Figure 3–24b and c) basically show the same contrast mechanisms as discussed previously [no orientation anisotropy of the electron backscattering and SE emission (Reimer et al., 1971; Seiler and Kuhnle, 1970) is involved]. Figure 3–24b recorded with the negatively biased ETD shows large shadowed regions (containing almost no information) and some highlighted individual flat surface planes of the crystal-like particles that occur with almost constant brightness because of the constant angle of beam incidence of the constant detection geometry. Figure 3–24b demonstrates that the detection geometry used for recording was not optimum. The micrograph obtained with the four-quadrant semiconductor detector is shown in Figure 3–24c, which depicts exactly the same area as Figure 3–24b. Due to the large solid angle of collection of this BSE detector the effects mentioned above in (1) and (2) do generate just small differences in the image intensity of differently oriented surface planes of the crystal-like particles. The effect of shadowing is not substantial in that micrograph. The increase of the BSE signal at edges, at surface steps, and small protruding particles [cf. (3) above] located on the flat surface planes due to enhanced BSE emission is significant. The fissures on some surface planes of the crystal-like particles (cf. Figure 3–24a) occur in the BSE micrograph also as rather dark features because just a minor fraction of the BSE can escape from inside the fissures.

The SEM micrographs are closely analogous to viewing a macroscopic specimen by eye. In the light optical analogy the specimen is illuminated with light from the side of the detector and viewed from the position of the electron beam (see, e.g., Reimer et al., 1984). When

a rather diffuse illumination is used then all surface elements are illuminated but those directed to the light source are highlighted. This corresponds to the situation for the positively biased ETD. The light optical analogy shows a pronounced shadow contrast if a directional light source illuminates the specimen surface from a suitable direction. This situation closely resembles BSE images recorded with a positively biased ETD. The strong light optical analogy very likely explains the fact that SEM micrographs of objects with a distinct topography can be readily interpreted even without extensive knowledge of the physics “behind” the imaging process.

As briefly mentioned in Section 2.1.3.2, the topographic and the material contrast can be pronounced or suppressed, respectively, by a combination of the signals of two oppositely placed detectors, A and B. Two BSE semiconductor detectors were first used by Kimoto et al. (1966) and they showed that the sum $A + B$ results in material and the difference $A - B$ in topographic contrast. Volbert and Reimer (1980) used a BSE/SE converter system and two opposite ET detectors for that kind of contrast separation in the SEM. The four-quadrant semiconductor detector used for recording Figures 3–23c and 3–24c allows for signal mixing of the four signals acquired simultaneously. Figures 3–23c and 3–24c represent the sum of the four signals ($S_{Q1}, \dots, S_{Q4}$), thus both micrographs show a pronounced material contrast. By addition of the signals of two adjacent quadrants at a time (i.e., $S_{Q1} + S_{Q2} = A$; $S_{Q3} + S_{Q4} = B$) the effect of two semiannular detectors A and B is obtained. The difference image $A - B$ shows a pronounced topographic contrast (Figure 3–23d). The directionality in Figure 3–23d can be varied readily by using a different combination of the individual signals of the quadrants, e.g., $A = S_{Q2} + S_{Q3}$ and $B = S_{Q1} + S_{Q4}$. Difference SE and BSE images recorded at exactly defined detection geometry allow for the reconstruction of the surface topography (Lebiedzik, 1979; Reimer, 1984a; Kaczmarek, 1997; Kaczmarek and Domaradzki, 2002; see also Sections 2.1.3.2 and 2.1.5.2). However, special care is required for the reconstruction of the surface topography using BSE images because of artifacts in the reconstructed image (Reimer, 1984a). To demonstrate the effect of directionality for four different detection Figure 3–26 shows four individual BSE micrographs each recorded with another quadrant of the semiconductor detector. The individual BSE images contain superimposed topographic and compositional contrast components.

2.3.2 Material Contrast

The material or compositional contrast arises from local differences in chemical composition of the object investigated. As shown in Figure 3–17, the SE yield δ increases weakly with increasing atomic number but the increase is significantly less than that of the BSE coefficient η . Experimental values of δ (see, e.g., the data collection by Joy 2001) scatter strongly around a mean curve. The increase of $\delta(Z)$ with Z is mainly due to SE [cf. Eq. (2.30)] generated by emitted BSE near the specimen surface (SE2). At electron energies larger than 5 keV the SE images usually show the same compositional contrast as the

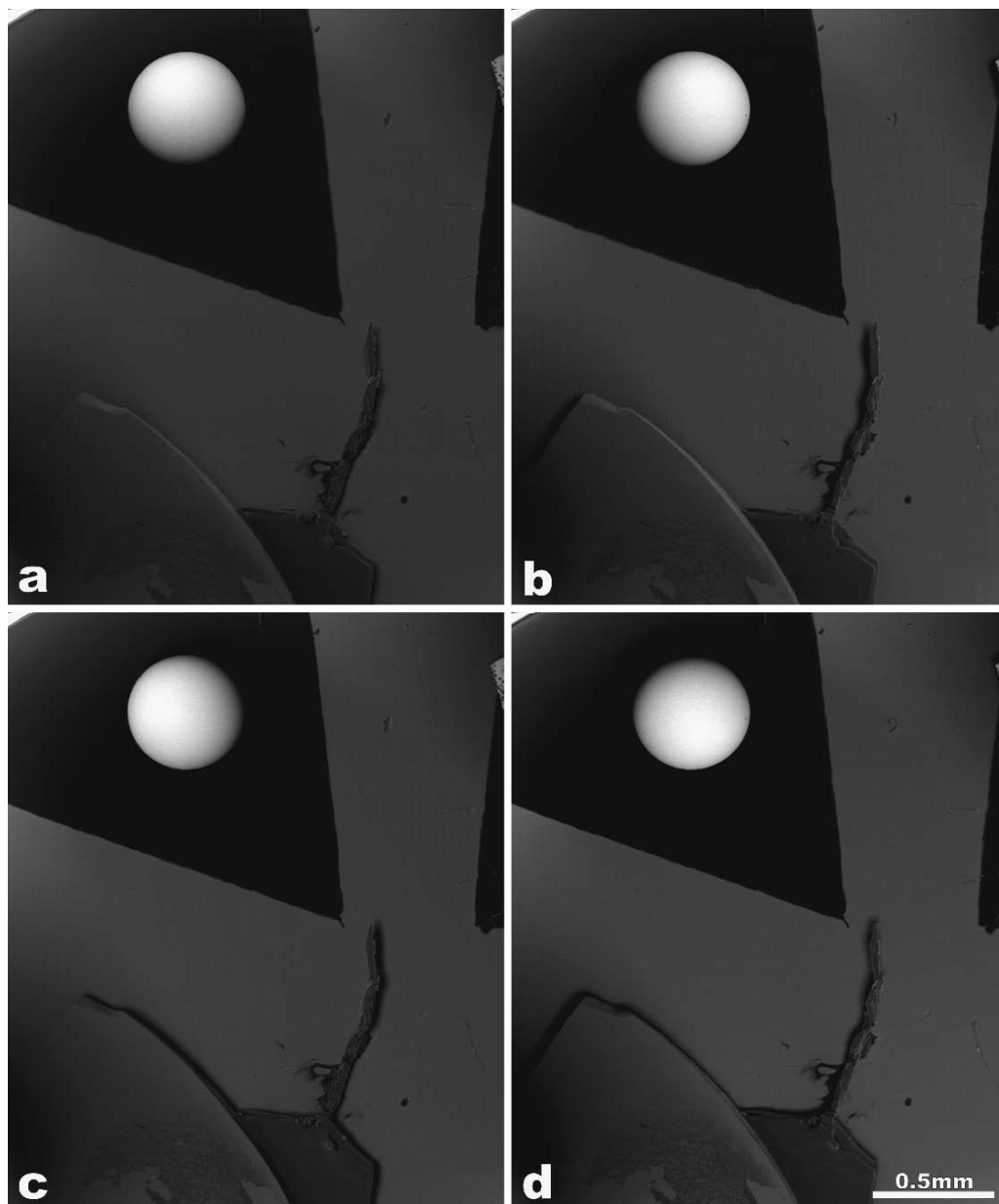


Figure 3–26. BSE micrographs of a steel ball on carbon (cf. Figure 3–23) each recorded with another individual quadrant of the four-quadrant semiconductor detector. (a) (–X)-quadrant, (b) (+Y)-quadrant, (c) (–Y)-quadrant, (d) (+X)-quadrant. The micrographs were recorded at 30kV and normal beam incidence. The angular position of the four-quadrant semiconductor detector is rotated clockwise against the x – y coordinates of the images by 34° . Shadows of the surface step of the feature at the bottom left of each image help to identify the position of the active quadrant visually.

Table 3–4. Compositional contrast calculated according to Eq. (2.33) for normal beam incidence $\theta = 0$ for the elements C, Al, and Fe.^a

Element 1 (Z)	Element 2 (Z)	η_1	η_2	$C = (\eta_1 - \eta_2)/\eta_2$
Aluminum (13)	Carbon (6)	0.1530	0.0641	0.581
Iron (26)	Carbon (6)	0.2794	0.0641	0.771
Iron (26)	Aluminum (13)	0.2794	0.1530	0.452

^a Compare Figures 3–23b and c and 3–26. $E_0 = 30$ keV.

corresponding BSE image. This situation is illustrated in Figure 3–23a–c where at normal beam incidence carbon ($Z = 6$) is darker than aluminum ($Z = 13$) in both the SE and BSE image. Table 3–4 gives some numerical values for the compositional contrast for carbon, aluminum, and iron calculated with the related BSE coefficients for normal beam incidence, which qualitatively agrees with the contrasts obtained in Figure 3–23c.

2.3.3 Other Contrasts

2.3.3.1 Voltage Contrast

The secondary electron image intensity varies if the potential of the specimen is positively or negatively biased with respect to the ground. In principle, a positively biased surface area shows decreased image intensity because low-energy SE are attracted back to the specimen by the electric field. Conversely, a negatively biased surface area shows enhanced image intensity because all SE are repulsed from the specimen. This voltage-dependent variation in contrast is designated as voltage contrast and dates back to the late 1950s (Rappaport, 1954; Oatley and Everhart, 1957; Everhart et al., 1959). Strictly speaking all emitted electrons are influenced to some extent by the potential of the sample, but only the SE and, in principle, the Auger electrons can be used for voltage contrast studies (Werner et al., 1998). The use of Auger electrons is more difficult than that of SE because of the very low yield of AE and the ultrahigh vacuum requirements of AE analysis.

The voltage contrast depends on the energy of the beam electrons and on the properties of the specimen, being most pronounced in the low electron energy region where the SE yield is highest (cf. Figure 3–12). Biasing the specimen positively or negatively by a few volts not only affects the amount of emitted SE but also their trajectories. This is caused by the fact that the majority of the SE have energies of a few electron volts in contrast to BSE and Auger electrons. The effect of specimen voltage on the SE trajectories is rather complex because it depends on the SE detection geometry, the sample position in the specimen chamber, the properties of the sample, and the operation conditions of the SEM. However, voltage contrast is a valuable tool for the investigation of a wide range of simple faults in microelectronic devices or studies of the potential distribution at grain boundaries obtained on the cross section of varistors at applied low DV voltage and their breakdown behavior at elevated voltage (Edelmann and Wetzig, 1995). The voltage contrast is also used to characterize the surface charge distribution of ferroelectrics (Uchikawa and Ikeda, 1981;

Hesse and Meyer, 1982; Roshchupkin and Brunel, 1993) and piezoelectrics (Bahadur and Parshad, 1980).

Voltage contrast measurements can also be performed in a dynamic mode on semiconductor devices by pulsing the electron beam [called electron stroboscopy (Spivak, 1966)] synchronously with the device signal as shown by Plows and Nixon (1968). This dynamic mode allows for quantitative voltage contrast measurements on semiconductor devices at high-frequency operation conditions known as electron beam testing widely used by the electronics industry for the development, fault diagnosis, and debugging of innovative integrated circuits. High-frequency electron stroboscopy requires high-speed electrostatic beam blanking systems with subnanosecond time resolution. For very high-frequency electron stroboscopy in the gigahertz range a special transverse-longitudinal combination gate system (Hosokawa et al., 1978) or microwave structure-based beam blanking techniques have been employed (Fujioka and Ura, 1981). A comprehensive treatment of the fundamentals of voltage contrast and stroboscopy has been published by Davidson (1989) and the state of the art of voltage contrast has been reviewed by Girard (1991). Furthermore, improvements of voltage contrast detectors as well as of detection strategies are discussed in detail by Dubbeldam (1991). Voltage contrast is now of a mature age, but the extension to future microelectronics also presupposes an extension in the domain of *in situ* testing methods and techniques.

2.3.3.2 Electron Beam-Induced Current

The electron beam generates a variety of signals emitted from the specimen as shown in Figures 3–2 and 3–14. In semiconductors the primary electrons generate electron hole pairs or minority carriers within the excitation volume. The mean number of electron hole pairs is given by E_0/E_{exm} [cf. Eq. (2.14)], where E_{exm} is the mean energy per electron hole pair forming event. For example, E_{exm} amounts to 3.6 eV for Si and 2.84 eV for Ge, i.e., one 10-keV electron generates on average approximately 2.7×10^3 electron hole pairs in Si and 3.5×10^3 in Ge (McKenzie and Bromely, 1959). The charge collection (CC) signal is detected between two electric contacts; one of these contacts collects the electrons and the other one collects the holes. If electromotive forces caused by electron voltaic effects are generated by the beam electrons in the specimen then a charge collection current I_{CC} designated as an electron beam-induced current (EBIC) flows through the ohmic contacts. If no electron voltaic effects occur, the beam electrons cause local β -conductivity, where the separation of charge carriers results in an electron beam-induced voltage (EBIV). The most important type of signal of the two charge-collecting modes is EBIC. A detailed treatment of the basic physical mechanisms and applications of the charge collection mode is given by Holt (1974, 1989), Deamy (1982), Reimer (1985), Shea et al. (1978), Alexander (1994), and Yakimov (2002).

EBIC can be observed in SEM simply by connecting a high-gain large bandwidth amplifier across the specimen using the amplified EBIC signal as a video signal. The input impedance of the amplifier must be

very low relative to that of the specimen to measure the true EBIC. For usual electron probe currents of some nanoamperes the charge collection currents are in the order of microamperes since for many materials the mean energy per electron hole pair is between approximately 1 and 13 eV (Holt, 1989). In contrast to EBIC, for the measurement of the true EBIV an amplifier with a very high input resistance is necessary.

The resolution obtained in the charge-collecting modes depends on the size of the excitation volume within the specimen, which readily can be extracted from Monte Carlo simulation data (see Section 2.2). For the CC mode, a depth and a lateral resolution have to be defined. The depth-dose function, which represents the energy loss per unit depth in the electron beam direction, determines the depth resolution. The lateral-dose function, which represents the energy loss per unit distance perpendicular to the electron beam direction, determines the lateral resolution. There are also empirical (Grün, 1957) and semiempirical expressions (Everhart and Hoff, 1971) as well as several analytical models (Bishop, 1974; Leamy, 1982) for the depth-dose and for the lateral-dose function as well (Bishop, 1974; Leamy, 1982).

Electron beam chopping and time-resolved EBIC can enhance the accuracy of measurements in several cases, e.g., for the estimation of the depth of p - n junction parallel to the surface (Georges et al., 1982) or allows for quantitative analysis of electrical properties of defects in semiconductors (Sekiguchi and Sumino, 1995) and interesting applications for the failure analysis of VLSI circuits (Chan et al., 2000).

2.3.3.3 *Crystal Orientation Contrast*

As previously mentioned, the backscattering coefficient η of a single crystal varies with the direction of the incident beam electrons related to the crystallographic orientation (cf. Section 2.2). This effect is caused by the variation of the atomic density, which the incident electrons encounter when penetrating into the crystal. In certain crystallographic directions the beam electrons penetrate more deeply. Those directions represent “channels” for the incident electrons. Changing the direction of the incident electrons relative to the crystallographic orientation causes the so-called crystal orientation or channeling contrast of the BSE image, which amounts to a maximum of approximately 5%. Crystal orientation contrast arises if a large single crystal is imaged at very low magnification using a small electron probe aperture of about 1 mrad. Scanning at low magnification both moves the electron probe and changes the angle of incidence across the field, thereby generating an electron channeling pattern (ECP). At higher magnification the angle of beam incidence varies just insignificantly across the small scanned field and channeling contrast is obtained in polycrystalline samples from small grains with different crystal orientations (Figure 3–27). The information depth of the crystal orientation contrast is in the order of a few nanometers only (Reimer, 1985) and therefore the contrast is very sensitive to distortions of the crystal at the surface. The channeling contrast reaches the maximum at energies between 10 and 20 keV (Reimer et al., 1971; Drescher et al., 1974).

An orientation anisotropy also occurs for the secondary yield (Reimer et al., 1971), which gives rise to an SE orientation contrast.

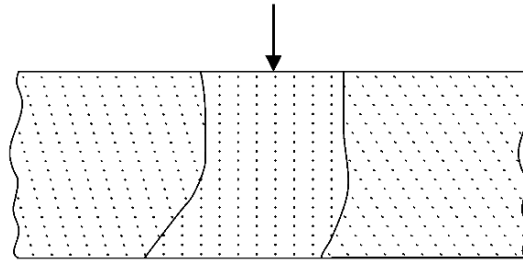


Figure 3-27. Cross section of a polycrystalline sample having grains with different crystal lattice orientation relative to the electron beam.

2.3.3.4 Magnetic Contrasts

Basically, two different types of magnetic contrast can arise from the interaction of the emitted electrons with the magnetic field of small domains of the specimen.

Type-1 Magnetic Contrast. Secondary electrons are deflected after emission by an external magnetic field, thus generating a magnetic contrast (Dorsey, 1969). External magnetic fields can exist in natural or synthetic engineered ferromagnetic materials such as magnetic tape, magnetic cards, and computer disks. The fringe fields near the surface are highly inhomogeneous and the SE trajectories are affected by the Lorentz force, which is proportional to $\mathbf{v} \times \mathbf{B}$ where $\mathbf{v}$ is the velocity vector of the SE and $\mathbf{B}$ the magnetic field. The most probable velocity corresponds to the electron energy of a few electronvolts (see Section 2.2.1). The acting Lorentz force deflects the trajectories of the SE and the resulting effect can be approximated to a tilt of Lambert's angular SE emission characteristics of the SE (see Section 2.2.1; Reimer, 1985). Figure 3-28 illustrates this effect for two domains in the specimen having oppositely directed external magnetic fields. To observe type-1 magnetic contrast in case of weak magnetic fields an ETD with a high angular sensitivity, a two-detector system (Dorsey, 1969; Wardly, 1971) or digital image processing (Szmaja, 2000, 2002) has been employed. For the type-1 magnetic contrast low beam electron energies are favorable because of the enhanced SE yield and therefore an increased signal-to-noise ratio (SNR). The actual problems related to the complicated mechanism of type-I magnetic contrast and its relatively low resolution were discussed by Szmaja (2002).

Type-2 Magnetic Contrast. This type of contrast arises from the deflection of backscattered electrons by the internal magnetic field within the specimen (Philibert and Tixier, 1969; Fathers et al., 1973). Depending on the direction of the magnetic field inside the sample, the BSE are bent toward or away from the surface between consecutive scattering events, i.e., the BSE coefficient is increased in domains where trajectories are bent toward the surface and decrease when bending the BSE trajectories in an opposite direction. To observe a sufficient type-2 magnetic contrast the beam electrons need an energy of at least 30 keV and a relatively high beam current. The type-2 contrast is maximized if the

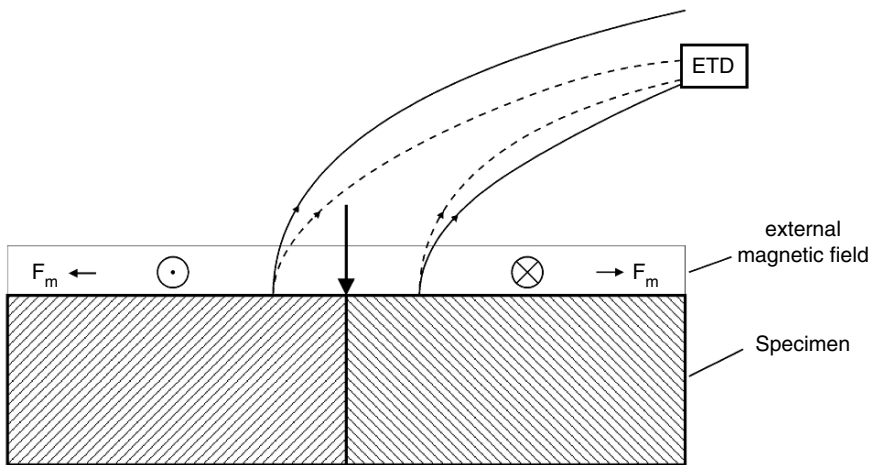


Figure 3-28. Scheme of type-1 magnetic contrast formation between two domains having oppositely directed external magnetic fields. The dashed lines indicate the SE trajectories for the most probable SE energy to the positively biased Everhart-Thornley detector (ETD) without magnetic field and the solid lines the trajectories with magnetic fields. The effect of the magnetic force F_m on the SE tilts the trajectories by a small angle toward or away from the ETD, respectively.

specimen is tilted by approximately $40\text{--}60^\circ$, however, the maximum contrast also depends on the takeoff angle of the BSE detector (Wells, 1978; Yamamoto et al., 1976). The BSE signal modulation due to magnetic fields inside the specimen is typically less than 1% of the collected current and unwanted topographic contrasts can be reduced in comparison with this magnetic contrast by a lock-in technique (Wells, 1979).

2.4 Specimen Preparation

The specimen preparation procedures required for optimum results of scanning electron microscopic investigations are of crucial importance. The dedicated preparation of the specimen under study is an essential prerequisite for the reliability of the experimental data obtained and has a significance comparable to the performance of the SEM used for the investigation. Unfortunately, the importance of specimen preparation is often underestimated. In principle, the preparation required depends significantly on the properties of the specimen to be investigated as well as on the type of SEM study, i.e., whether imaging of the surface or of cross sections of the sample (cf. Sections 2.2, 2.3, and 3-6), crystallographic characterization by electron diffraction techniques (cf. Section 7), or X-ray microanalytical investigations (cf. Section 6) are considered. Bearing in mind the variety of specimens having unknown properties on the one hand and the multitude of possible investigation techniques on the other hand, it is obvious that the choice of the most promising preparation procedures can be a rather complex matter. Although the preparation techniques are described in a variety of books (e.g., Reimer, 1967; Hayat, 1974-1976, 1978; Revel et al., 1983; Polak

and Varndell, 1984; Müller, 1985; Steinbrecht and Zierold, 1987; Albrecht and Ornberg, 1988; Edelmann and Roomans, 1990; Grasenick et al., 1991; Echlin, 1992; Malecki and Romans, 1996), collections of methods (e.g., Schimmel and Vogell, 1970; Robards and Wilson, 1993) with updates, and publications, the successful preparation still also depends in many cases on experience and skillful hands.

It is beyond the scope of this chapter to discuss the wide field of preparation techniques. Therefore, a brief rather general outline of specimen preparation with reference to specific literature will be given. Figure 3–29 schematically outlines some important preparation procedures used for inorganic and organic materials. As a general rule, a successful investigation by SEM requires specimens, which have clean surfaces, sufficient electrical conductivity, are not wet or oily, and possess a certain radiation stability, to resist electron irradiation during imaging. An exception of this rule is allowed only for SEMs working at ambient pressure (say at low vacuum; see Section 4), which permits direct imaging of dirty, wet, or oily samples, although radiation damage occurs with radiation-sensitive specimens. The goal of an ideal prepa-

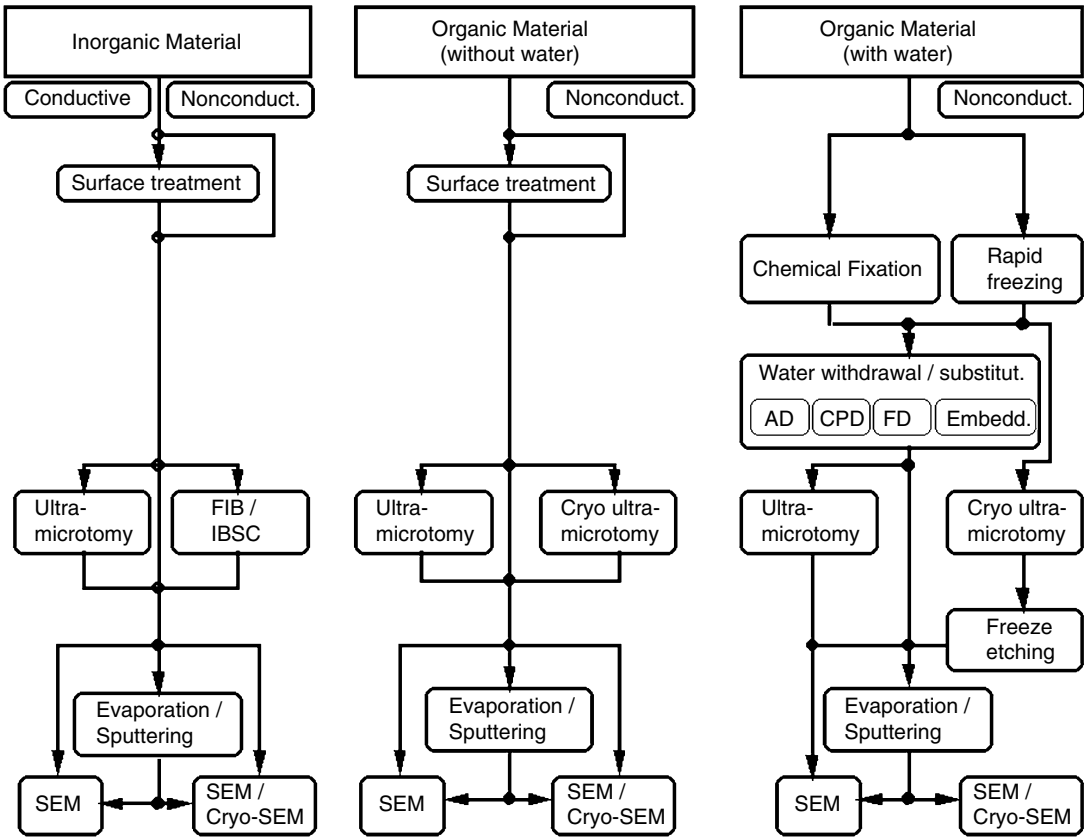


Figure 3–29. Schematic drawing of important preparation procedures for SEM used for inorganic and organic materials with and without water. AD, air drying; CPD, critical point drying; FD, freeze drying; FIB, focused ion beam; IBSC, ion beam slope cutting.

ration consists in making specimens accessible for high vacuum SEM studies without changing the relevant properties under investigation.

Many inorganic samples with sufficient electrical conductivity, such as metals, alloys, or semiconductors, can be imaged directly with little or no specimen preparation (Figure 3–29). This is one very useful feature of scanning electron microscopy. In some cases a surface treatment may be required, e.g., to clean the specimen surface with an appropriate solvent, possibly in an ultrasonic cleaner, and with low-energy reactive gas plasma for the removal of hydrocarbon contamination (Isabell et al., 1999). The cleanings are suitable to prepare electrically conductive specimens for surface imaging in SEM. In case of nonconductive samples, such as ceramics, minerals, or glass, a conductive coating (e.g., Willison and Rowe, 1980) with a thin metal film (e.g., gold, platinum, tungsten, chromium) or a mixed conductive film (e.g., gold/palladium, platinum/carbon, platinum/iridium/carbon) is required for good-quality imaging. For X-ray microanalysis carbon coating is preferred because of its minimum effect on the X-ray spectrum. The coating can be performed by evaporation (e.g., Reimer, 1967; Shibata et al., 1984; Hermann et al., 1988; Robards and Wilson, 1993), by diode sputtering (Apkarian and Curtis, 1986), or by planar magnetron sputtering (Nagatani and Saito, 1989; Müller et al., 1990). High-quality conductive thin-film coating for high-resolution SEM (see Section 3) can be performed in an oil-free high vacuum by both evaporation, using e.g., tungsten, tantalum/tungsten, platinum/carbon, or platinum/iridium/carbon, and rotary shadowing methods (Gross et al., 1985; Hermann et al., 1988; Wepf and Gross, 1990; Wepf et al., 1991) as well as by ion beam and by penning sputtering with, e.g., chromium, tantalum, and niobium (Peters, 1980).

For the study of microstructural features (see Section 7) and for microanalytical investigations (see Section 6) a flat surface is required, therefore rough specimen surfaces have to be flattened by careful grinding and subsequent polishing according to standard metallographic methods (Glauert, 1973). To remove mechanical deformations caused by grinding and mechanical polishing, a final treatment with electrochemical polishing or ion beam polishing may be necessary. In case of polycrystalline and heterogeneous material, selective etching by ion bombardment may be used, which generates a surface profile caused by locally different sputtering yields, thus giving rise to topographic contrast of grains and the individual materials (Hauffe, 1971, 1995).

Often, specimens need to be characterized and analyzed both above and below the surface, e.g., if the subsurface composition of the material, process diagnosis, failure analysis, *in situ* testing, or three-dimensional reconstruction of the spatial microstructure is required. Flat cross sections through the specimen can be obtained by ultramicrotomy (Reid and Beesley, 1991; Sitte, 1984, 1996; the block face can be used for SEM imaging), ion beam slope cutting (Hauffe, 1990; Hauffe et al., 2002), an FIB technique (Kirk et al., 1988; Madl et al., 1988; Ishitani and Yaguchi, 1996; Shibata, 2004; Giannuzzi and Stevie, 2005), or by a combination of an FIB system with a field emission SEM (SEM/FIB),

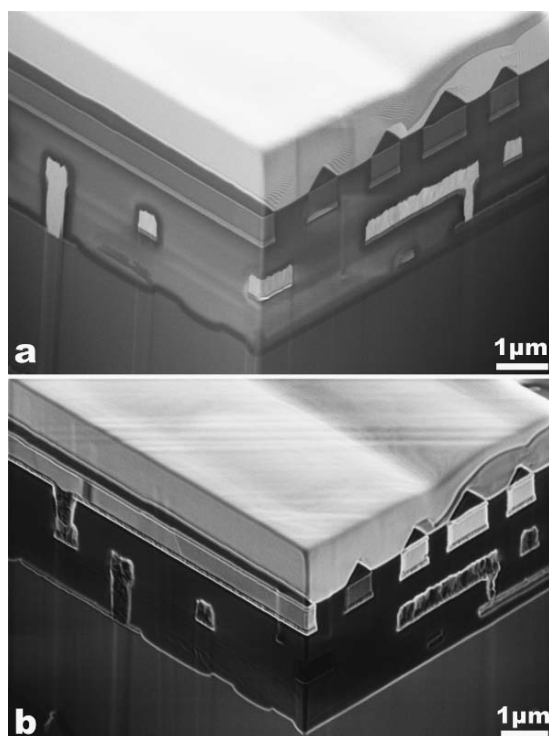


Figure 3–30. Integrated circuit with two perpendicular vertical cross sections into the interior. FIB sectioning was performed with the CrossBeam® tool from Carl Zeiss NTS. The secondary electron images using the EDT were obtained by the field emission SEM (a) at 3 kV and by the FIB system (b) at 5 kV. Both micrographs reveal the site-specific internal structure of the integrated circuit, although some features occur with different contrast caused by different mechanisms of SE generation by electrons (a) and ions (b) (Courtesy of Carl Zeiss NTS, Oberkochen, Germany.)

which allows the precise positioning of the cross section and, most importantly, realtime high-resolution SEM imaging of the cutting process, which enables, among other things, the examination of the spatial structure (Sudraud et al., 1987; Gnauck et al., 2003a, b; McGuinness, 2003; Holzer et al., 2004; Sennhauser et al., 2004). An example for two perpendicular vertical cross sections into an integrated circuit is shown in Figure 3–30. The combined SEM/FIB additionally can be equipped with analytical techniques such as energy-dispersive X-ray spectroscopy, wavelength-dispersive X-ray spectroscopy, Auger electron spectroscopy, and secondary ion mass spectrometry allowing for three-dimensional elemental analysis of the interior of the specimen.

The other class of samples consists of organic material, which usually has an insufficient electrical conductivity for scanning electron microscopy. Although biological specimens contain water—the water content ranges in human tissues from approximately 4 to 99% (Flindt, 2000)—many other organic materials do not, e.g., numerous polymers. The preparation strategies to be applied to specimens with and without water differ (cf. Figure 3–29), although there are also some similarities between them.

The surface treatment of organic specimens without water, such as cleaning, grinding, polishing, and etching by dissolution, chemical attack, or ion bombardment, has many similarities to the surface treatment of inorganic materials. A detailed discussion of and the recipes for specific preparation procedures for polymers are given in the chapter "Specimen preparation methods" in the book by Sawyer and Grubb (1996). Analogous to nonconductive inorganic materials, conductive coating with a thin metal film (e.g., gold, platinum, tungsten, chromium) or a mixed conductive film (e.g., gold/palladium, platinum/carbon, platinum/iridium/carbon) is required for good-quality imaging. If the subsurface structure of the material has to be studied, flat cross sections usually are prepared by ultramicrotomy or cryoultramicrotomy, depending on the cutting behavior of the specimen under study. In principle, cutting with ions and imaging and analysis with electrons by using a combined FIB/SEM tool seem possible also with polymers. It was recently shown that ion milling is possible, e.g., with rubber (Milani et al., 2004; cf. also Figure 3–31) and with plastic material (cf. Figure 3–32). As yet, the application of FIB for cutting and milling of organic specimens is rare.

Most of the organic specimens that contain water are biological samples. A small fraction of water-containing specimens is nonbiological, e.g., hydrogels. Caused by the high vacuum in the SEM the water-containing specimens cannot be investigated in the wet state. In principle, three different preparation strategies exist to make wet specimens accessible to SEM investigations:

1. Withdrawal of the water;
2. replacement of the water by some vacuum-resistant material such as resins or freeze substitution (Feder and Sidman, 1958; Hess, 2003) of the ice of the rapidly frozen specimen by some organic solvent; and
3. rapid freezing of the water.

Irrespective of the preparation strategy used the native spatial structure of the specimen should be maintained. Air drying, which is the most simple method of drying, is not suitable for drying soft specimens because the surface tension induces remarkable forces during the process of air drying, deforming the specimen irreversibly (Kellenberger and Kistler, 1979; Kellenberger et al., 1982). Figure 3–29 shows different paths, which can be used, even though the degree of structural preservation depends on the preparation procedures applied. The different preparation procedures have been described in detail (Kellenberger and Kistler, 1979; Robards and Sleytr, 1985; Steinbrecht and Zierold, 1987; Dykstra, 1992; Echlin, 1992; Kellenberger et al., 1992; Robards and Wilson, 1993). Among the different preparation methods rapid freezing is the method of choice for preparing biological specimens in a defined physiological state (Echlin, 1992). In case of chemical fixation, which may create artifacts (Kellenberger et al., 1992), the water of the sample has to be withdrawn or replaced afterward. If the surface structure of the specimen has to be studied, then the specimen surface has to be coated with a thin conductive film prior to SEM investigation. If the interior of the specimen has to be studied, the sample has to be

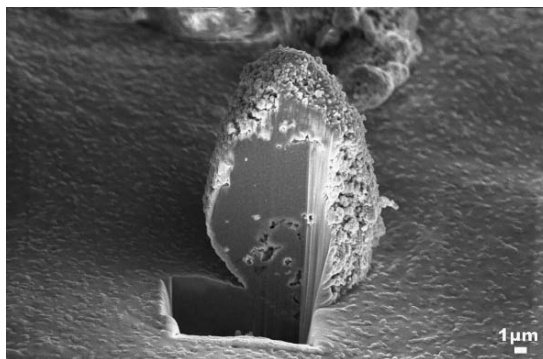


Figure 3–31. Secondary electron micrograph (“through-the-lens” detection) of a site-specific FIB cross-sectioned abrasive wear particle of a tyre supported onto carbon. The FIB sectioning and imaging at 5kV was performed with the CrossBeam® tool combining a focused ion beam system with a field emission SEM (cf. Gnauck et al., 2003a). The cross section reveals the interior features of the rubber particle. (Courtesy of Carl Zeiss NTS, Oberkochen, Germany.)

opened by sectioning with the ultramicrotome or possibly FIB and subsequently coated. In case of physical fixation, i.e., rapid freezing, the specimen has to be opened by freeze fracturing (for review see Severs and Shotton, 1995; Walther, 2003), cryosectioning, or now possibly by ion milling the frozen-hydrated sample [ion milling in ice is possible (McGuinness, 2003)]. After short partial freeze drying (also called freeze etching), the fracture face or block face have to be properly coated by a conductive film and then can be directly analyzed in the cryo-SEM (Echlin, 1971; Hermann and Müller, 1993; Walther and

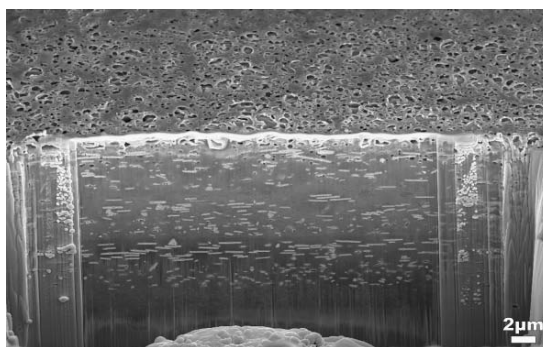


Figure 3–32. Secondary electron micrograph of an FIB cross-sectioned color film. The SE image is composed of the signals of two SE detectors, where the “through-the-lens” detection contributes a fraction of 60% and the positively biased ETD a fraction of 40%. The FIB sectioning and imaging at 5kV were performed with the CrossBeam® tool combining a focused ion beam system with a field emission SEM (cf. Gnauck et al., 2003a). The cross section in the lower half of the micrograph reveals the interior of the film (three color layers corresponding to red, green, and blue and the submicrometer features as well), whereas the upper half of the micrograph shows the porous outer surface of the film. (Courtesy of Carl Zeiss NTS, Oberkochen, Germany.)

Müller, 1999; Walther, 2003). Another possible path is complete freeze-drying and subsequent conductive coating of the sample, which then can be analyzed at room temperature in the SEM.

As yet, FIB sample preparation and subsequent FIB or SEM imaging are in early stages of application in the life sciences. Very recently, *in situ* FIB sectioning was successfully performed with critical-point-dried hepatopancreatic cells (Drobne et al., 2004) and some epithelium cells (Drobne et al., 2005).

2.5 Radiation Damage and Contamination

The inelastic electron–specimen interaction inevitably damages the irradiated specimen and can induce contamination at the specimen surface. Once radiation damage, in particular of organic specimens, has been extensively investigated for thin films in transmission electron microscopy, comparatively little is systematically studied for irradiation-sensitive samples in SEM. This may be due to the fact that the interpretation of radiation damage in TEM is easier because of the uniform ionization density through thin specimens. In bulk specimens, however, the ionization density is a function of the depth (for a detailed treatment of the depth dose function see, e.g., Shea, 1984) and a layer below the surface at the maximum ionization density will be damaged faster than others within the electron range R (cf. Figures 3–13 and 3–14). According to the Bethe stopping power [see Eq. (2.26)], the damage will be proportional to $1/E \ln(1.166 E/J)$. Table 3–5 gives values of the stopping power for carbon and protein for electron energies from 0.1 to 30 keV, which show the increase of the stopping power with decreasing electron energy. It is commonly assumed that the shape of the depth-dose curve is not a function of either the primary electron energy or the material when normalized to the electron range (Shea, 1984). That means that the layer with the maximum ionization density approaches the surface as the electron energy decreases.

In organics, the radiation breaks due to the transfer of typically tens of electron volts to an electron at the site of the interaction of many intra- and intermolecular bonds, which generates free radicals (e.g., Bolt and Carroll, 1963; Dole, 1973; Baumeister et al., 1976). Many excited

Table 3–5. Mean ionization potential J [Eqs. (2.27) and (2.28), respectively] and the Bethe stopping power dE/ds [Eq. (2.26)] for carbon and protein at different electron energies.^a

Sample	E (keV)					
	Parameter	0.1	1.0	5.0	10	30
Carbon	J (eV)	56.5	92.8	98.5	99.2	100.4
	dE/ds (eV/cm)	–56.4	–19.7	–6.4	–3.7	–1.5
Protein	J (eV)	50.6	78.0	82.0	83.0	83.0
	dE/ds (eV/cm)	–43.8	–14.2	–4.5	–2.6	–1.1

^a The values listed for dE/ds have to be multiplied by 10^7 . The following values were used for the calculation (Reichelt and Engel, 1984): carbon: $Z = 6$; $A = 12$; $\rho = 2 \text{ g/cm}^3$; protein: mean atomic number $\langle Z \rangle = 3.836$; $A = 7.7$; $\rho = 1.35 \text{ g/cm}^3$.

species will very rapidly recombine in 10^{-9} to 10^{-8} s and will reform the original chemical structure dissipating the absorbed energy as heat. Some recombinations will form new structures, breaking chemical bonds and forming others. If the material was initially crystalline, defects will form and gradually it will become amorphous. In addition to these structural changes the generated free radicals will rapidly diffuse to and across the surface or can evaporate, i.e., loss of mass and composition change will occur (e.g., Egerton, 1989, 1999; Egerton et al. 1987; Engel, 1983; Isaacson, 1977, 1979a; Reimer, 1984b; Reichelt et al., 1985). Bubbles may form at high dose rates when volatile products are trapped. Not only the beam electrons damage the organic sample but also fast secondary electrons ($E_{SE} > 50$ eV) can produce damages outside the directly irradiated specimen area (Siangchaew and Libera, 2000). Furthermore, beam-induced electrostatic charging and heating can also damage organic samples. Conductive coating of the organic specimen, as suggested for inorganic materials by Strane et al. (1988), can keep trapped free radicals as well as reduce beam-induced temperature rise or electrostatic charging (Salih and Cosslett, 1977). Lowering of the temperature of the specimen is a further measure to reduce the sensitivity of an organic specimen to structural damage and mass loss. However, the reduction factor depends considerably on the material of the specimen.

The radiation damage mechanisms in semiconductors are different from those described above. As mentioned in Sections 2.1.3.1 and 2.3.3.2, the incident electrons generate electron hole pairs, which will be trapped in the Si_2O layer due to their decreased mobility. This can generate space charges, which in turn can affect the electronic properties of the semiconductor.

Beam-induced contamination is mass gain, which occurs when hydrocarbon molecules on the specimen surface are polymerized by the beam electrons. The polymerized molecules have a low surface mobility, i.e., the amount of polymerized molecules increases in the surface region where polymerization takes place. There are two main sources for hydrocarbon contamination: (1) gaseous hydrocarbons arising from oil pumps, vacuum grease, and possibly O-rings, and (2) residual hydrocarbons on the specimen. Several countermeasures exist to reduce the contamination to a tolerable level (see, e.g., Fourie, 1979; Wall, 1980; Postek, 1996). The amount of gaseous hydrocarbons is substantially reduced when the SEM is operated with an oil-free pumping system and a so-called cold finger located above the specimen. Further, the contamination rate falls more rapidly as the specimen temperature is lowered, and below -20°C contamination is difficult to measure (Wall, 1980). This is caused by the reduced diffusion of hydrocarbons on the specimen. In some cases, preirradiation of a large surface area with the electron beam is helpful, which immobilizes (polymerizes) hydrocarbons around the field of view to be imaged. Finally, specimens are mostly exposed to the atmosphere before transfer into the specimen chamber. Weakly bound molecules (e.g., hydrocarbons) can be completely eliminated by gently heating the sample in the specimen exchange chamber (low vacuum) to 40 – 50°C for several minutes by a spot lamp (Isaacson et al., 1979b). A detailed topical review on the

radiation damage and contamination in electron microscopy is given by Egerton et al. (2004).

2.6 Applications

Scanning electron microscopy is an indispensable tool for investigations of a tremendous variety of specimens from very different fields such as materials science, mineralogy, geology, semiconductor research, microelectronics, industry, polymer research, ecology, archeology, art, and life sciences. Although the investigations are not restricted just to imaging of surface structures, the majority of SEM studies apply the imaging modes. As mentioned previously, considerable additional information about the local elemental composition, electronic and magnetic properties, crystal structure, etc. can be acquired when the SEM is combined with supplementary equipment such as electron and X-ray spectrometers to take advantage of the energy spectra of the emitted electrons and X-rays. Table 3–6 surveys the information, which can be obtained from inorganic and organic specimens not containing

Table 3–6. SEM applications on specimens from materials science, mineralogy, geology, polymer science, semiconductors, and microelectronics.^a

Specimen	Information
Metals, alloys, and intermetallics	<i>At the specimen surface:</i> Topography (three-dimensional); microroughness; cracks; fissures; fractures; grain size and shape; texture; phase identification; localization of magnetic domains; size and shape of small particles; elemental composition; elemental map <i>Inside the specimen:</i> Grain and phase structures; three-dimensional microstructure; cracks; fissures; material inclusions; elemental composition
Ceramics, minerals, glasses	<i>At the specimen surface:</i> Topography (three-dimensional); microroughness; cracks; fissures; fractures; grain size and shape; pores; phase identification; size and shape of small particles; elemental composition <i>Inside the specimen:</i> Grain and phase structures; three-dimensional microstructure; cracks; fissures; material inclusions; pores; elemental composition
Polymers, wood	<i>At the specimen surface:</i> Morphology; topography (three-dimensional); microroughness; cracks; fissures; fractures; pores; size and shape of small fibers and particles; fiber assemblage in woven fabrics; elemental composition <i>Inside the specimen:</i> Cracks; fissures; fractures; pores; composite structure; elemental composition
Semiconductors, integrated circuits, microelectronic devices	Dislocation studies (with CL); metallization and passivation integrity; quality of wire bonds; electrical performance; design validation; fault diagnosis; testing

^a State-of-the-art preparation and image analysis techniques are required to take full advantage of the capabilities of SEM.

water. Further, *in situ* scanning electron microscopy allows for different specific specimen treatments in the specimen chamber (see, e.g., Wetzig and Schulze, 1995), which serves as a microlaboratory, and the simultaneous observation of the specimen response (cf. Table 3–7).

The advancement of nanoscale science and technology demands the manipulation of nanoobjects at the molecular level and ultimately the manufacture of things via a bottom-up approach. Very recently, a four nanoprobe system has been installed inside a field emission SEM, which may be used for gripping, moving, and manipulating nanoobjects, e.g., carbon nanotubes, setting up electric contacts for electronic measurements, tailoring the structure of the nanoobject by cutting, etc. and for making nanostructures (Peng et al., 2004). The SEM in this setup allows for visualization of the four nanoprobes operating inside the specimen chamber as well as the process of formation of microstructures.

Less spectacular, but nevertheless important, are applications of scanning electron microscopy to image macroscopic samples in the millimeter range at very low magnification (about 10× to 100×), which cannot be seen clearly by the eye or by the light microscope for some reasons. Two examples from very different fields are shown in Figures 3–33 and 3–34 taking advantage of the large depth of focus as well as distinct topographic and material contrast.

Working in the low magnification range, the depth of focus limit in the SEM (see Section 2.1.5.2) can be overcome by recording stacks of through-focus images (as in conventional and confocal optical microscopy), which are digitally postprocessed to generate an all-in-focus image (Boyde, 2004). The application of the technique is advantageous when BSE imaging of spongy specimens is required, as demonstrated with examples from the study of human osteoporotic bone (Boyde, 2004).

Table 3–7. *In situ* treatments in SEM and available information about specimens from materials science.

Treatment	Information
Static and dynamic deformation, e.g., by tension, compression, bending, machining	Kinematic processes during deformation; submicrometer cracks visible only under load; localized deformation centers, e.g., slip bands, crack nucleation; deformation-induced acoustic emission
Laser irradiation, e.g., in pulse mode, Q-switch mode	Phase transformations; structural modifications; crack formation due to thermal shock; diffusion processes; laser-induced surface melting and evaporation processes; vapor deposition on substrates; cumulative effects of multiple laser pulses
Ion beam irradiation	Depth profile/cross section; grain boundaries; spatial microstructure; internal grain and phase structures
Electrical and magnetic effects	Reversible and irreversible breakdown of voltage barriers; size distribution of magnetic and ferroelectric domains; orientation distribution of magnetic and ferroelectric domains; effects accessible by EBIC, EBIV, and CL

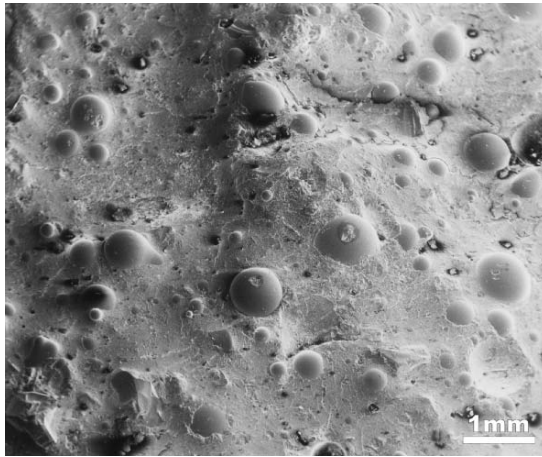


Figure 3–33. Secondary electron micrograph of the material of white lines, dividing the opposite lanes on roads, used to reflect the light from the headlights of cars. Some of the light-reflecting glass spheres are damaged or snatched off (their original positions can be recognized by sphere-like indentations). For safety reasons protruding sharp-edged particles are embedded in the material (e.g., at the bottom left corner) to generate a high friction between the tire and the white line. The sample is sputter coated with gold. (Micrograph kindly provided by Rudolf Göcke, Institut für Medizinische Physik und Biophysik, Münster, Germany.)

In life sciences, application of SEM mainly for morphological studies is also widespread and started when commercial SEMs became available in the late 1960s (see, e.g., Pfefferkorn and Pfautsch, 1971). However, in life sciences SEM is used less than TEM. Table 3–8 provides a survey of the specimens and information that can be obtained by SEM. Engineered biomaterials and tissues are becoming increasingly important

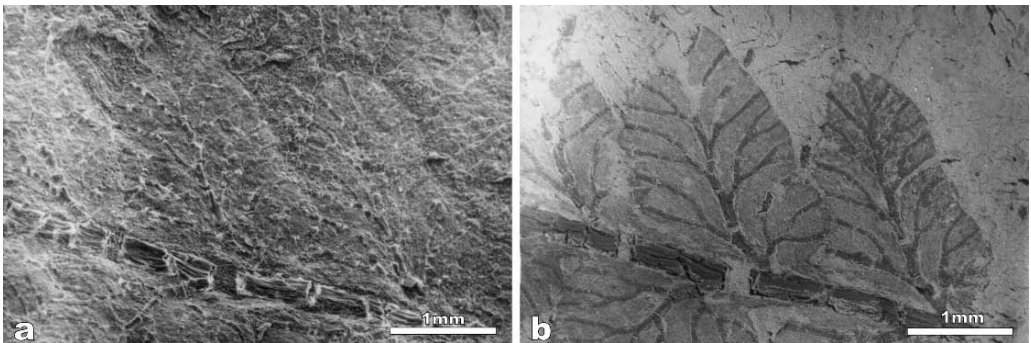


Figure 3–34. Secondary electron (a) and backscattered electron (b) micrograph of carbonized fossil remains of a fern bot embedded in clay. The sample is coated with a very thin carbon film. The contrast in the SE image is caused by the topography, whereas the BSE image shows a distinct atomic number contrast. The carbonized fossil remains appear dark due to the lower mean atomic number surrounded by bright areas of clay. (Micrographs kindly provided by Rudolf Göcke, Institut für Medizinische Physik und Biophysik, Münster, Germany.)

Table 3–8. SEM applications on specimens from life sciences.^a

Specimen	Information
Bones, teeth, dentin, cartilage, hairs, fingernails, toenails	<i>At the specimen surface:</i> Morphology; ultrastructure, pathological alterations of ultrastructure, microstructure, roughness; cracks; fissures; fractures; elemental composition <i>Inside the specimen:</i> Three-dimensional microstructure; cracks; fissures; elemental composition
Biominerals, e.g., gallstone, kidney stone, tartar, calcification	<i>At the specimen surface:</i> Morphology; microstructure, cracks; fissures; fractures; grain size and shape; size and shape of small particles; elemental composition <i>Inside the specimen:</i> Grain size and shape; microstructure; cracks, fissures; material inclusions; cavities; elemental composition
Soft tissues cells, bacteria	<i>At the specimen surface:</i> Morphology; topography (three-dimensional); roughness; ultrastructure; pathological alterations of ultrastructure; size and shape of cells and bacteria; elemental composition <i>Inside the specimen:</i> Ultrastructure; pathological alterations of ultrastructure; elemental composition
Biomaterials, implants, prostheses	Morphology; biocompatibility; biostability; ultrastructure of and degradation mechanisms at the bone–implant interface; mineral apposition; cell and tissue apposition; adsorption behaviors of fibrinogen, albumin, and fresh plasma on implants for the cardiac–vascular systems; fault diagnosis of prostheses; failure analysis after loading tests in simulator; wear of prostheses; surface erosion of prostheses after use

^a State-of-the-art preparation and image analysis techniques are required to take full advantage of the capabilities of SEM.

in biomedical practice and it has become clear that cellular responses to materials depend on structural properties of the material at both the micrometer and nanometer scale. SEM is one of several methods for controlling material properties on both of these scales and thus it is increasingly used to study those materials.

Scanning electron microscopy can be used for comparative morphological studies of tissues as demonstrated by the application in cardiovascular surgery to detect endothelial damage caused by skeletonization (Rukosujew et al., 2004). In cardiovascular surgery, the radial artery is increasingly used for myocardial revascularization because of its presumed advantageous long-term patency rates. The vessel can be harvested as a pedicle or skeletonized. The SEM reveals the endothelial morphology (cf. Figure 3–35), and thus allows comparison of the skeletonization technique with pedicle preparation using either an ultrasonic scalpel or scissors.

Investigation of implants is a strongly growing field, where scanning electron microscopy is an indispensable tool. Figure 3–36 shows an example from ophthalmology where a technique for knotting a suture to the haptic of an intraocular lens is used for its firm fixation onto the sclera to avoid postoperative dislocation. Because of the large depth of focus and the distinct topographic contrast in the secondary electron micrograph, which conveys a pseudo three-dimensional and detailed view, the SEM allows checking as to the extent to which the haptic and suture can be damaged by knotting. Moreover, the quality of knots can be studied.

Microtechnology and “microelectromechanical systems” (MEMS) are additional fields in which SEM is used as a tool for monitoring processes, detecting defects, or measuring sizes and distances, e.g., in micromachines and micromechanical or micromachining processes (see, e.g., Ishikawa et al., 1993; Aoyagi, 2002; Hernandez-Lopez et al., 2003; Khamsehpour and Davies, 2004).

The acquisition of quantitative data about the third dimension (stereo, 3D) of surfaces and interior specimen structures was previously mentioned (see Sections 2.14, 2.15, and 2.4). In general, this requires digital image analysis, specific instrumentation for the SEM (e.g., specimen stage, detectors), and special specimen preparation (e.g., ultramicrotome, IBSC, FIB). Recently, interesting applications of 3D morphometry for accurate dimensional and angular measurements of microstructures (Minnich et al., 1999, 2000) and of volumetric measurements (Chan et al., 2004) were shown using stereopaired images and digital image analysis.

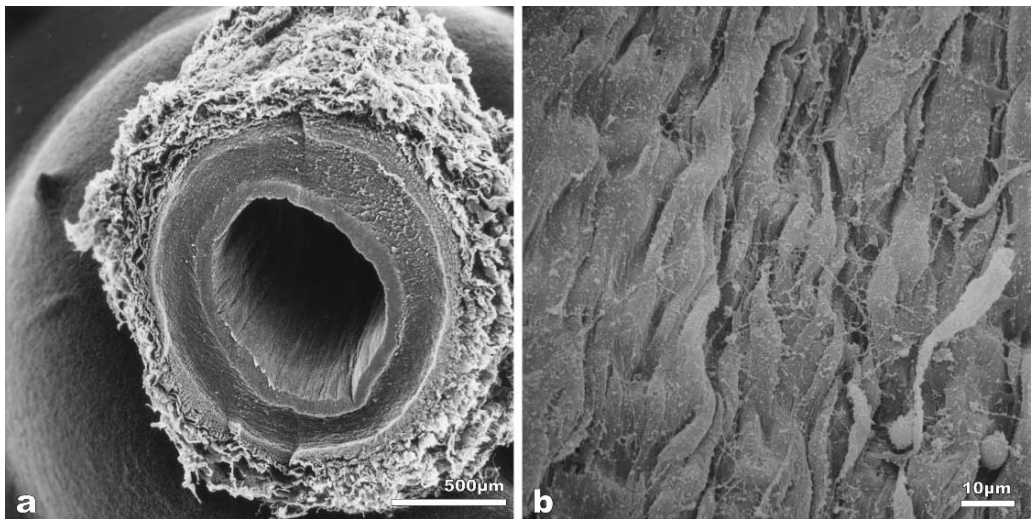


Figure 3–35. Secondary electron micrograph of the cross-sectioned radial artery at low magnification (a) and of the endothelial cells at medium magnification (b). The vessel was critical point dried and sputter coated with gold.

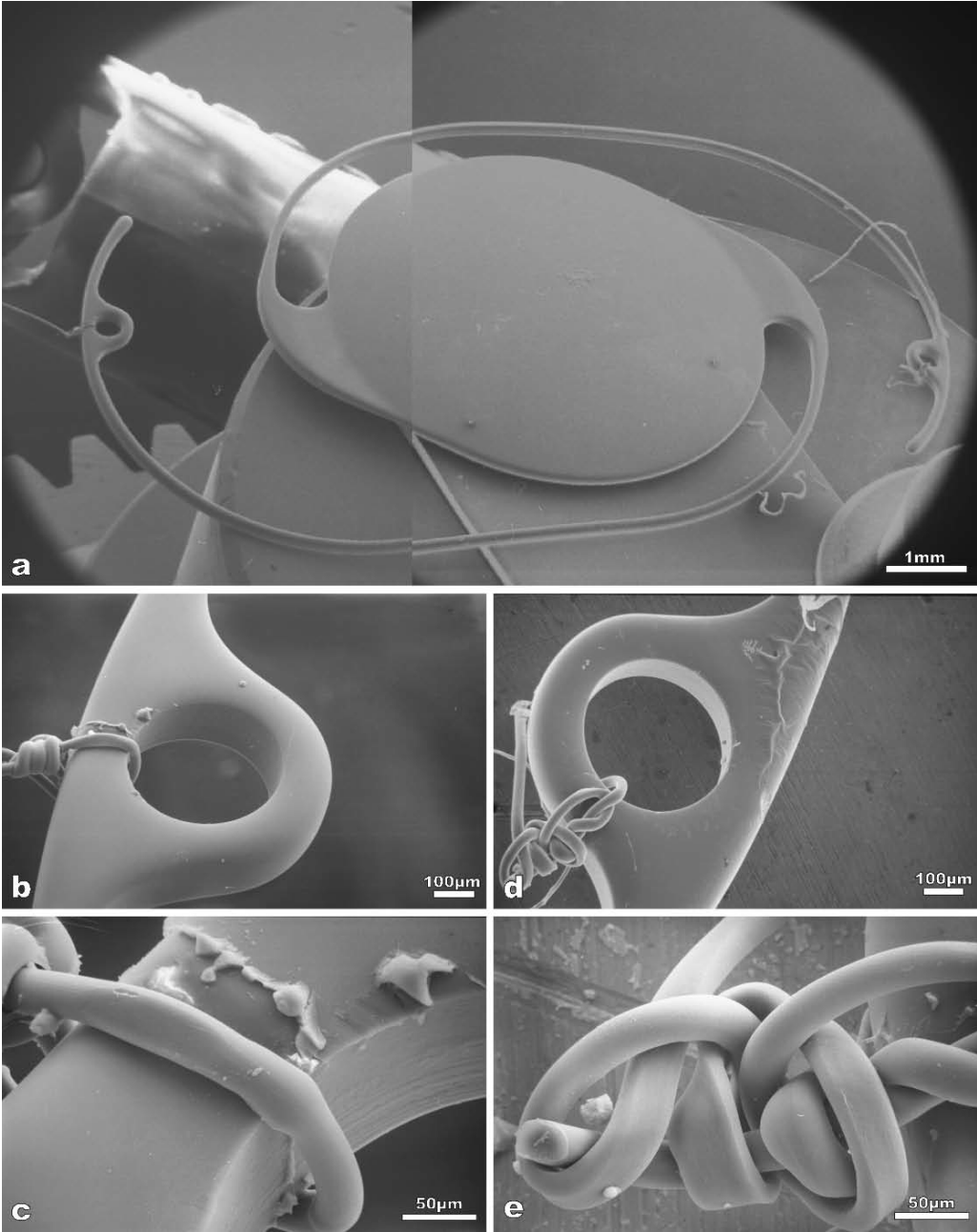


Figure 3–36. Secondary electron micrographs of a poly(methyl methacrylate) (PMMA) intraocular lens (IOL) with knots from 10-0 polypropylene (Prolene) on the haptic recorded at different magnifications. (a) Whole IOL with the haptic and the fixated sutures; (b–e) Detailed views to the haptic and knots, respectively, showing some minor damage. The IOL is sputter coated with gold and imaged at 25kV.

3 Field Emission Scanning Electron Microscopy

The diameter of the electron beam at the specimen surface sets a fundamental lower limit to the signal localization and, therefore, also to the resolution, which can potentially be obtained. As discussed in Section 2.2 and shown in Figures 3–13 and 3–14, the SE and BSE are emitted from a surface area, which commonly is much larger than the beam diameter at the specimen surface. The large emitting area is caused by multiple elastic and inelastic electron scattering events within the excitation volume, whose size depends on the specimen composition and energy of the beam electrons. Only the SE1 and BSE1 generated as the beam enters the specimen carry local information, while the SE2 and BSE2 carry information about the larger region surrounding the point of beam entrance (cf. Figure 3–14). High-resolution information can be obtained from SE1 and BSE1 generated by an electron probe with a diameter at the specimen surface of about 1 nm or even less. A probe of that small size can be achieved by using field emission electron sources, electromagnetic lenses with low aberration coefficients (cf. Eqs. 2.7, 2.8, and 2.10), and both highly stabilized acceleration voltage (cf. Eq. 2.8) and objective lens current. High-resolution scanning electron microscopy at conventional acceleration voltages—that is 5–30 kV—will be treated in Section 3.1. Alternatively, high-resolution information, in principle, can also be achieved when the excitation volume is reduced to a size similar to the SE1 and BSE1 emitting area by using low-energy beam electrons. By definition, electrons below 5 keV are considered low-energy beam electrons and, consequently, scanning electron microscopy at low energies is called scanning low-energy electron microscopy or low (acceleration)-voltage scanning electron microscopy (LVSEM). This type of scanning electron microscopy will be treated in Section 3.2. However, the majority of commercial high-resolution SEMs are capable of operation at both conventional energies, i.e., from 5 to usually 30 keV, and at low energies, i.e., below 5 keV down to usually 0.5 keV.

3.1 High-Resolution Scanning Electron Microscopy

3.1.1 Electron Guns

Three different types of electron guns are suitable sources for high-resolution SEM: the cold field emission gun (FEG), the hot FEG, and the so-called Schottky emission cathode (SEC). The characteristic parameters of the different electron guns are listed in Table 3–1.

Schottky emission cathodes are of the ZrO/W(100) type—also called ZrO/W(100) thermal field emitter (TFE)—and have a tip radius of 0.6–1 μm (Tuggle and Swanson, 1985). The work function of the TFE is lowered to about 2.8 eV. In operation the SEC is heated to about 1800 K and electrons are extracted by a high electric field, which lowers the potential barrier (Schottky effect). The SEC brightness is about three orders of magnitude higher and the energy spread of the emitted electrons is about a factor of 2 lower than those for the thermionic W-

cathode. Presently, the SEC in commercial high-resolution SEM is less frequently an electron source than the FEG.

The FEG usually consists of a very sharp [100] or [321] oriented tungsten single crystal and two anodes in front, which extract (first anode) and accelerate or decelerate (second anode) the electrons by the electric field to a final energy $E_0 = eU$ (Figure 3–37). Caused by the small tip radius r , which is in the range of 10 to about 50 nm, the electric field strength amounts to at least 10^8 V/cm with an extraction voltage of approximately 4–5 kV applied between the first anode and the tip. Due to the high field strength at the tip the width of the potential barrier is significantly reduced and field emissions take place. The field emission current density j_c is described by the Fowler–Nordheim equation

$$j_c = c_1 |E|^2 / \Phi \exp(-c_2 \Phi^{3/2} / |E|) \quad (3.1)$$

where $|E| \approx U_1/r$, c_1 and c_2 depend weakly on $|E|$, and Φ is the work function of tungsten. The density j_c depends strongly on $|E|$, and E can be varied by U_1 . The so-called cold FEG (CFEG) is operated at room temperature and generates a current density of typically 2×10^5 A cm⁻². However, after several hours of work adsorbed gas layers have to be removed by short heating to about 2500 K (flashing), otherwise the emission current becomes very unstable. The distinct advantage of the cold FEG is the low-energy spread.

The hot FEG (HFEG) is operated at approximately 1800 K, which increases the energy spread to about twice that from the cold FEG. The current density is higher than for the cold FEG and typically amounts to 5×10^6 A cm⁻². The advantage of the hot FEG is the less noisy emission current.

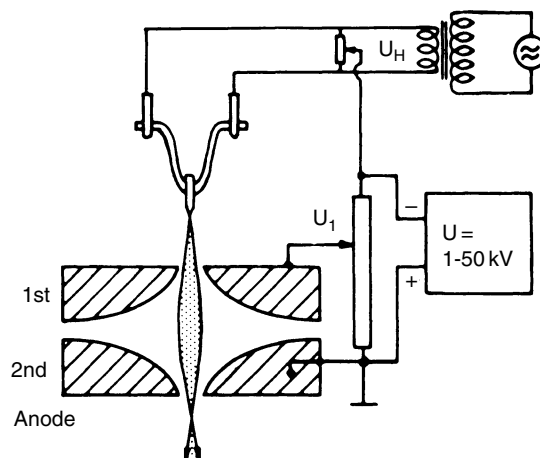


Figure 3–37. Schematic drawing of the field emission gun with an electrolytically polished sharp monocrystalline tungsten tip. The hot FEG operates the tip at high temperature heated by the applied voltage U_H . U , acceleration voltage; U_1 , extraction voltage. [Adapted from Reimer (1993); with kind permission of the International Society of Optical Engineering (SPIE), Bellingham, WA.]

Field emission guns require ultrahigh vacuum in the order of 10^{-8} – 10^{-9} Pa in the gun chamber, which is generated by ion getter pumps. This means that SEMs equipped with an FEG need a sophisticated and consequently cost-intensive vacuum system. Another disadvantage of FEGs compared to the thermionic W-cathode is their significantly lower short- and long-time beam current stability.

3.1.2 Electron Lenses

Electron lenses are used to demagnify the virtual source size, which amounts to 3–5 nm for both the cold and hot FEG, and about 20–30 nm for the SEG. To obtain an electron beam diameter of about 1 nm or less a demagnification of only 10–100 $\times$ is required in contrast to up to about 5000 $\times$ for the thermionic emission triode gun (cf. Section 2.1.2). To achieve the smallest effective electron probe diameter, the spherical and the chromatic aberration constants have to be as small as possible [see Eqs. (2.7), (2.8), and (2.10)]. In the conventional SEM usually large working distances ranging from about 10 to 40 mm are used. Typical values of the spherical aberration constant C_s are 10–20 mm. Since C_s increases strongly with increasing WD ($C_s \sim WD^3$) sufficiently small values of $C_s \sim 1$ –2 mm can be achieved only with very short WD, i.e., the focus of the electron beam has to be inside (so-called “in-lens” type) or very close to the objective lens [frequently called “semi-in-lens” with a snorkel-type conical objective lens (Mulvey, 1974)]. The chromatic aberration constant C_c corresponds approximately to the focal length of the objective lens for large WD, i.e., also the chromatic aberration is strongly lowered at a very short WD. The shortest WD of the “in-lens” type SEM is about 2.5 mm in order to secure a specimen traverse in the x and y direction perpendicular to the optical axis as well as specimen tilt angles up to a maximum of $|\pm 15^\circ|$. Larger tilt angles obviously require a larger work distance. To obtain the minimum effective electron probe diameter under these conditions, the optimum aperture α_{opt} has to be used [see Eq. (2.11)]. Presently, the highest resolution obtained with the “in-lens” type FESEM at 30 keV using a test sample amounts to 0.4 nm (Hitachi, 2001).

3.1.3 Detectors and Detection Geometries

The detectors used in field emission scanning electron microscopes (FESEM) have been described in Section 2.1.3.1. The detection geometry depends on the particular type of the FESEM. The instruments using the conventional specimen position outside the objective lens (“out-lens”), i.e., the WD is in the range of about 5–30 mm, are commonly equipped with an ET detector located laterally above the specimen and a BSE detector located centrally above the specimen. The “semi-in-lens” instruments, where the specimen is outside but immersed in the field of the objective lens, usually have both the detector arrangement of the “out-lens” type SEM and the “through-the-lens” detection, thus combining the advantages of both detection geometries. The “in-lens” type SEM is restricted to “through-the-lens” detection (cf. Section 2.1.3.2).

3.1.4 Specimen Stages

The purpose of the specimen stage in high-resolution scanning electron microscopes is of course the same as in conventional SEM, i.e., the stage has to allow for precise backlash-free movement, tilting, and possibly rotation of the sample during the investigation. As for conventional SEM, there are optionally special specimen stages available that allow investigations of the specimen at elevated temperature, during different types of mechanical deformation, at positive or negative bias, and last at low temperature. Independent on the special type of specimen stage, a higher stability in terms of mechanical vibrations as well as mechanical or thermal drift is required to avoid any deterioration of the performance of the high-resolution SEM. The “in-lens”-type SEMs use side-entry specimen holders, which are almost identical to the ones used in TEMs (cf. Section 2.1.4). However, the limited space available in this type of SEM places some restrictions on the specimen stage for the ultimate resolution of “in-lens”-type FESEM.

3.1.5 Contrast Formation and Resolution

At high beam energy, e.g., 30 keV, the lateral extension of the excitation volume in the specimen is for carbon approximately $10\mu\text{m}$ and for a high atomic number element such as gold about $1\mu\text{m}$ (cf. Figure 3–13). Secondary and backscattered electrons are emitted from a surface area of the specimen, which corresponds in size to about the lateral extension of the excitation volume (cf. Figure 3–14). As discussed in Sections 2.2.1 and 2.2.2, the SE2 and BSE2 represent the majority of the SE and BSE, respectively, whereas the SE1 and BSE1, both carrying high-resolution information, represent the minority. Assuming for simplicity an electron beam diameter of 1 nm, the ratio of the lateral size of the excitation volume and the beam diameter amounts to approximately 10^4 for carbon and 10^3 for gold. By choosing the magnification such that the field of view at the specimen surface approaches the lateral size of the excitation volume, i.e., related to a 100-mm image size about $10,000\times$ for carbon and $100,000\times$ for gold, both the SE2 and the BSE2 contributions will change in response to the features of the field of view on the size scale of the excitation volume. In contrast to this the SE1 and BSE1 contributions will change in response to the features of the field of view approximately on a size scale of the electron beam diameter. That means that in the course of scanning the electron beam across the field of view, the SE2/BSE2 contribution only insignificantly varies from pixel to pixel whereas the SE1/BSE1 contribution depends sensitively on local features as small as the beam diameter. With a further increase of magnification the field of view becomes significantly smaller than the lateral size of the excitation volume, consequently the SE2/BSE2 contribution is almost constant over the image. The changes in the total SE/BSE signal are almost exclusively due to the SE1/BSE1 component and correspond to the changes in the very tiny volume where SE1/BSE1 are generated. Figure 3–38 shows an example of a high-resolution SE micrograph recorded from a test sample at a magnification of $500,000\times$. The distinct changes in image intensity reflect the variation of the SE1 component, which is due to

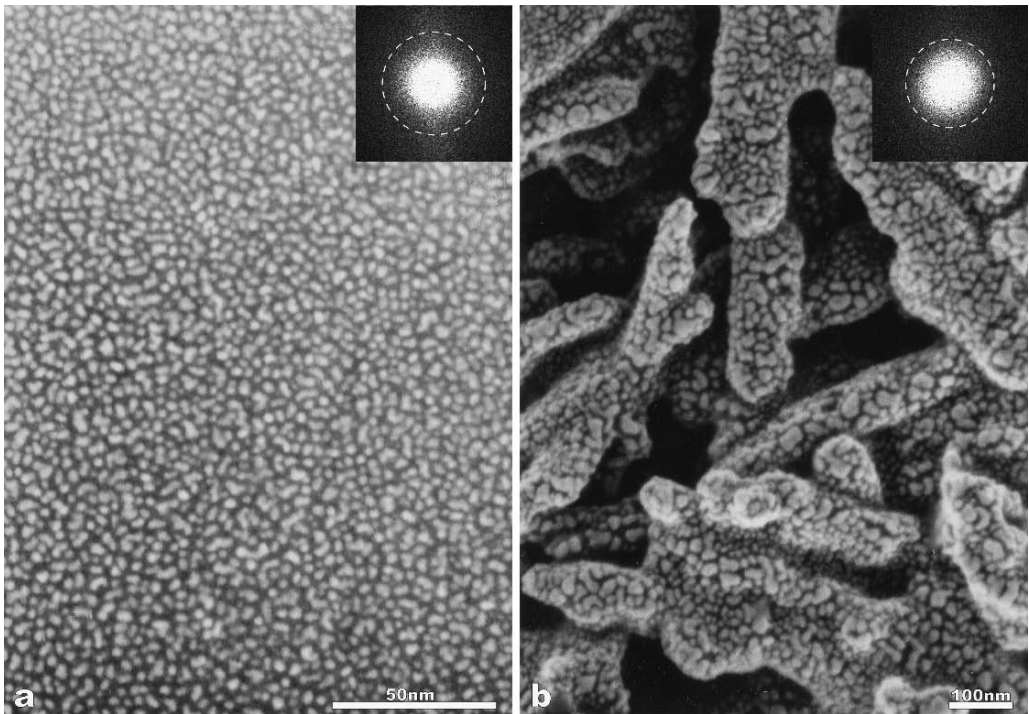


Figure 3–38. High-resolution secondary electron micrographs of test specimens recorded with the “in-lens” field emission scanning electron microscope S-5000 (Hitachi Ltd., Tokyo, Japan) at 30 kV (a) and 1 kV (b). The test specimens were Au–Pd particles on carbon (a) and magnetic tape evaporated with gold (b). The related power spectra are inserted at the top right. The dashed circles correspond to $(0.6\text{ nm})^{-1}$ (a) and $(3.5\text{ nm})^{-1}$ (b), respectively.

the large differences in the atomic number between the carbon and the Au–Pd particles. This type of test sample is usually used to demonstrate the performance of SEMs.

The low SE yield of low atomic number specimens (cf. Figure 3–17) such as soft biological objects and polymers limits the resolution due to the poor SNR. However, the SNR can be improved significantly by coating the specimen surface with an ultrathin very fine-grain metal film (Peters, 1982) by Penning sputtering or by evaporation in oil-free high vacuum (cf. Section 2.4). The thickness of such films can be as small as 1 nm and, as we shall see later, such ultrathin films do not mask fine surface structures. In addition to improving the SNR the ultrathin coating plays an important role in contrast formation and the image resolution obtainable. As mentioned earlier, the SE1 arise from the area directly irradiated by the electron beam and its immediate vicinity caused by the delocalization of the inelastic scattering in the order of a very few nanometers (cf. Section 2.2). In the case of the specimen coated with an ultrathin metal film the SE1 generation is confined almost exclusively to the film. Figure 3–39a shows schematically the cross section of an object coated with a continuous metal film of constant thickness. As the electron beam is scanned across the object the projected film thickness will vary between the nominal film thickness

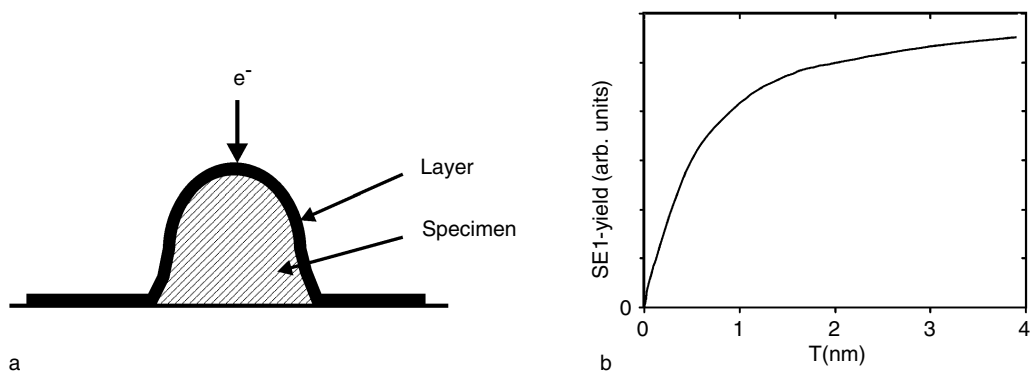


Figure 3-39. Schematic cross section of a specimen coated with an ultrathin continuous metal layer of constant thickness (a). The projected mass thickness of the metal layer varies as the electron beam is scanned across the specimen. (b) Graph of the SE1 yield versus the thickness of the coating film.

and the maximum, which is several times greater than the nominal thickness. As shown by Monte Carlo calculations the SE1 yield increases very quickly with the thickness of the metal film (Joy, 1984). For example, the Monte Carlo calculations by Joy (1984) reveal for chromium and 20-keV electrons that half of the maximum SE1 yield is reached for a thickness of 1–1.5 nm only. The dependence of the SE1 yield versus the thickness of a coating film is shown schematically in Figure 3-39b. It indicates that the increase of the SE1 yield with the thickness slows down at twice the thickness at half of the maximum SE1 yield, i.e., the continuous film should be as thin as possible. Monte Carlo calculations of the SE1 yield for some of the metals suitable for preparing ultrathin very fine-grain metal films show a monotonic increase with the atomic number (Joy, 1991); thus some further improvement of the SNR may be expected with high atomic number metals.

The ultrathin very fine-grain metal film on the sample surface also improves the BSE1 component significantly, thus improving the SNR in high-magnification BSE micrographs. The BSE1 are very important for high-resolution SEM because the elastic electron scattering is strongly localized. The intensity of the BSE1 component increases with the projected film thickness, i.e., increases with the number of atomic scattering centers. Since the BSE coefficient strongly increases with the atomic number (cf. Figure 3-17), the BSE1 component of the metal film is significantly larger than the contribution from the coated low-atomic number specimen. The same is also true for small metal clusters or small particles at the specimen surface, e.g., such as colloidal gold down to a minimum diameter of 0.8 nm (Hermann et al., 1991), which can be identified unambiguously in the high-resolution BSE micrograph.

3.1.6 Selected Applications

Since the achievable resolution is the main difference between the high-resolution field emission SEM and the conventional SEM, it is obvious that the high-resolution SEM (HRSEM) can readily handle almost all of the applications mentioned in Section 2.6. Exceptions are

possibly, such as applications in which high emission currents (cf. Table 3-1) or high electron beam current stability are indispensable. Because vacuum conditions in FESEMs are more strict concerning the pressure in the specimen chamber (at least one order of magnitude less than in CSEM) as well as the content of gaseous hydrocarbons and hydrocarbons at the specimen, some specimens may not meet the requirements for cleanliness and very low partial pressure. However, if high-resolution FESEM is applied instead of CSEM, more information about the specimen will be obtained due to the higher resolution as soon as the magnification used exceeds approximately $10,000\times$ to $20,000\times$. That means that lateral resolutions requiring magnifications clearly beyond about $20,000\times$ belong to the dedicated domain of high-resolution SEM.

The following few applications selected from an almost unlimited quantity should demonstrate the strength of HRSEM in several fields of research. It is clearly beyond the scope of this section to discuss in this context specific details about the specimens and imaging techniques.

Figure 3-40a shows the secondary electron micrograph of a regular protein surface layer of a bacterial cell envelope. The specimen was unidirectional shadowed with an ultrathin tungsten layer leaving an uncoated region behind the latex bead. Comparison of the regular structure of the HPI layer in the coated and the uncoated region shows that the contrast in the uncoated area is significantly lower than in the coated region, though the resolution of structural details is very similar as verified by the related power spectra. This example also demonstrates that coating with the ultrathin very fine-grain metal film does not mask fine structural features. On principle, a similar resolution can also be obtained with nonregular organic specimens, however, it remains more difficult to quantify unambiguously the resolution obtained.

An extremely important application of HRSEM, as yet unrivaled by other surface imaging techniques, is the localization of molecules on surfaces by immunolabeling techniques (for review see, e.g., Griffith, 1993; Hayat, 1989/1991, 2002; Polak and Varndell, 1984; Verkleij and Leunissen, 1989). The use of HRSEM for immunoelectron microscopy started more than 20 years ago (de Harven et al., 1984; Gamliel and Polliack, 1983; Hicks and Molday, 1984; Molday and Maher, 1980; Walther and Müller, 1985, 1986; Ushiki et al., 1986). Since then efforts were made to optimize the technique of immuno-scanning electron microscopy in terms of localization precision, contrast, and SNR (e.g., Albrecht et al., 1988; Hermann et al., 1991; Hirsch et al., 1993; Simmons et al., 1990). While the colloidal gold can be localized directly in the BSE image, the precision of the indirect localization of the antigen depends on the type of labeling and the size of the colloidal gold and ranges from less than 5 to about 30 nm (Baschong and Wrigley, 1990; Müller and Hermann, 1990). Figure 3-41b demonstrates the unambiguous detection of immunogold-labeled calcium-binding birch pollen allergen Bet v4 in birch pollen using 10 nm colloidal gold (Grote et al., 1999b). The BSE micrograph shows the topographic features and the

small colloidal gold beads as sharp bright spots due to the material contrast. However, for more than a decade immuno-scanning electron microscopy has been established as a trusted technique and, with the commercial availability of high-quality gold probes (available in sizes ranging from 1 to 40 nm), is used in many electron microscopic laboratories for various studies (e.g., Apkarian and Joy, 1988; Erlandsen et al., 1995; Grote et al., 2000; Müller and Hermann, 1992; Ris and Malecki, 1993; Yamaguchi et al., 1994).

The interior structure of biological specimens is accessible by HRSEM, if samples are rapidly frozen and opened by cryofracturing or cryoul-

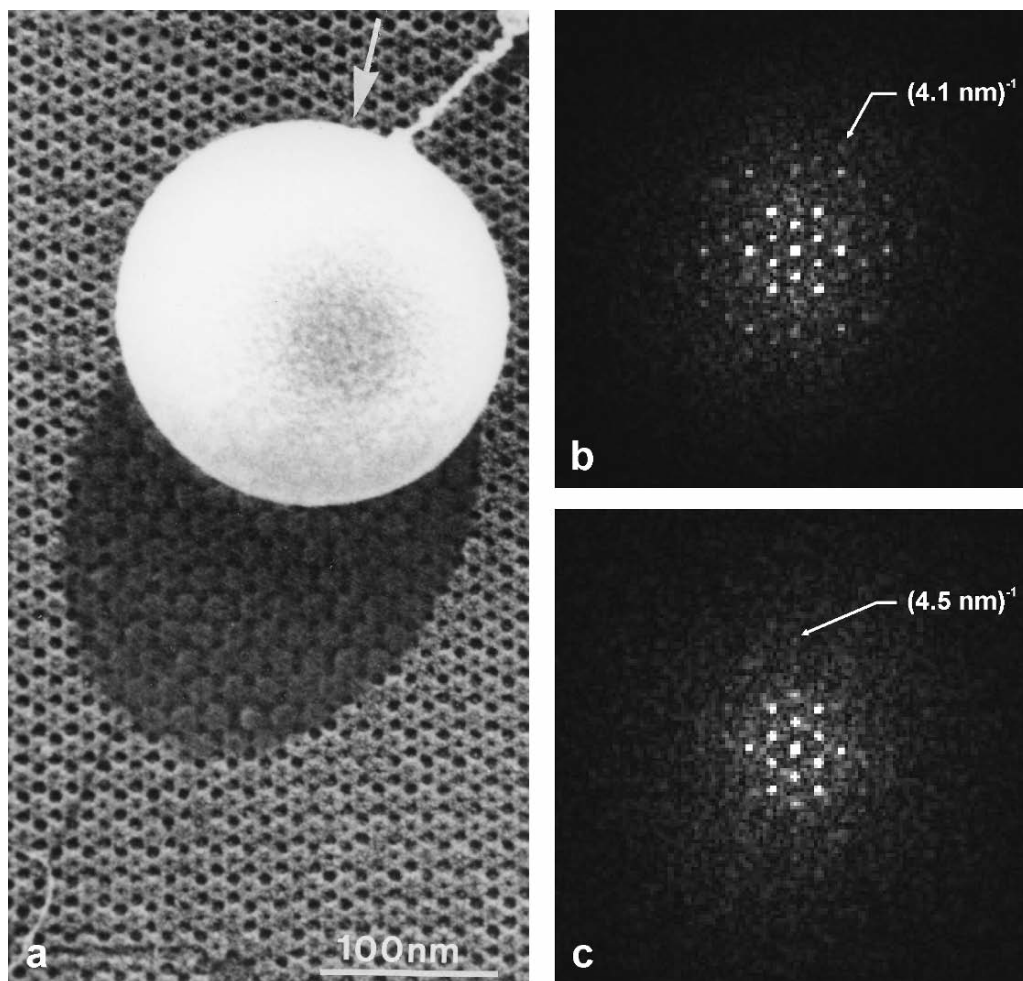


Figure 3-40. Secondary electron micrograph of a regular protein surface layer [hexagonally packed intermediate (HPI) layer (Baumeister et al., 1982)] of *Deinococcus radiodurans* recorded with an “in-lens” FESEM at 30 kV (a). The specimen was unidirectional shadowed (see arrow) at an elevation angle of 45° with a 0.7-nm-thick tungsten layer leaving an uncoated region behind the latex bead. The power spectra of a coated (b) and an uncoated (c) region of the HPI layer reveal the resolution obtained (outermost diffraction spots are indicated and the corresponding reciprocal values of resolution are given). The contrast in the uncoated region is about 15–20% of that from the coated region. [Micrograph kindly provided by Dr. R. Wepf; from Reichelt (1995); with kind permission of GIT Verlag, Darmstadt, Germany.]

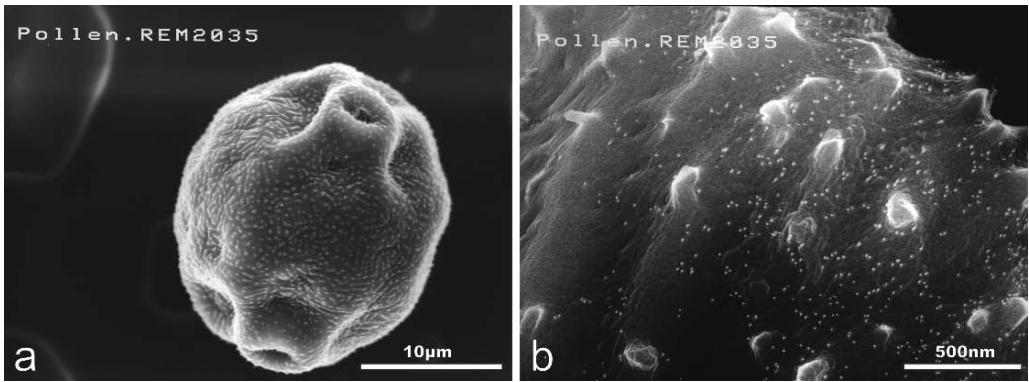


Figure 3-41. Secondary electron micrograph from a birch pollen at low magnification (a) and the backscattered electron micrograph at high magnification recorded with an “in-lens” FESEM at 10 kV. Immunolabeling of the calcium-binding birch pollen allergen Bet v4 in dry and rehydrated birch pollen was performed using 10-nm colloidal gold (for more details see Grote et al., 1999a,b). The BSE image shows superimposed topographic and material contrast. The colloidal gold beads are unambiguously detected in the BSE micrograph as tiny bright spots.

tramicrotomy. After partial freeze-drying and double-layer coating of the block face, the specimen can be directly analyzed in the cryo-SEM (Echlin, 1971; Hermann and Müller, 1993; Walther et al., 1995). Figure 3-42 shows for comparison the BSE micrograph of the cryosectioned and the cryofractured block face of high-pressure frozen yeast cells (Walther and Müller, 1999). At low magnification the cryosectioned block face appeared very flat in the image (Figure 3-42a), whereas the cryofractured face exhibits the typical rough fracture pattern (Figure 3-42c). At high magnification (Figure 3-42b and d) the cytoplasm appeared densely packed with different classes of particle. Particles as small as 25 nm can be visualized clearly. It is important to mention that the typical artifacts of cryosections such as compression and crevasses are not visible on the block face.

Both strengths of the HRSEM, namely the high resolution and the high depth of focus, are required to resolve surface structures at the nanometer scale on tilted surfaces randomly oriented. One typical example with submicrometer-sized crystalline zeolite particles is shown in Figure 3-43. The HRSEM is the tool most suited to characterize the habit of the individual particles as well as to visualize the fine surface structure such as growth steps of terraces (see, e.g., González et al., 2004). The growth step-edge height, usually in direct relation to the unit cell dimension and important in understanding the crystal growth mechanism of this novel microporous material, cannot be measured precisely by HRSEM, but the atomic force microscope (AFM) enables measurement of direct height with subnanometer resolution. Thus, the combination of the two high-resolution surface-imaging methods HRSEM and AFM is strongly advisable to obtain more complete information as demonstrated by different applications (e.g., González et al., 2004; Huang et al., 2004; Keller and Chih-Chung, 1992; Lian et al., 2005; Stracke et al., 2003; Wang et al., 2004; Z. X. Zhao et al., 2005).

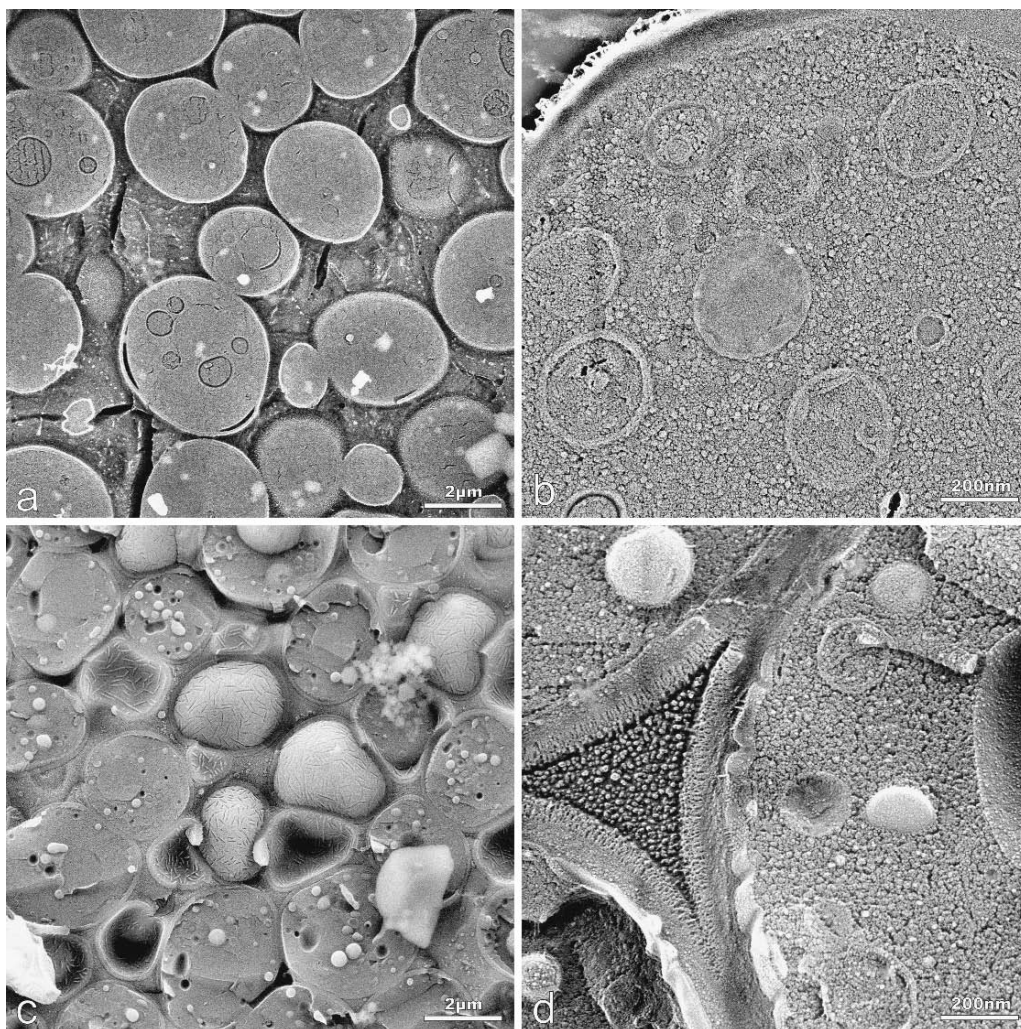


Figure 3-42. BSE micrographs of high-pressure frozen yeast cells recorded in an “in-lens” cryo-FESEM at 10 kV. (a and b) Block face after cryosectioning and (c and d) block face after freeze fracturing. The block faces were double-layer coated with 2.5 nm platinum/carbon and subsequently with 5–10 nm carbon. The cytoplasm from both the sectioned (b) and the fractured (d) sample appears to consist of particles with variable size. [From Walther and Müller (1999); with kind permission of Blackwell Publishing Ltd., Oxford, U.K.]

HRSEM is also a very valuable tool for the evaluation of mechanical properties of structural materials. For example, most structural materials are strengthened by fine particles of second phases usually having diameters less than 500 nm. The strengthening effect is primarily governed by the mean size, the size distribution, and the volume fraction of the particles. Both HRSEM and AFM allow for the precise determination of the mean size, size distribution, and volume fraction of the particles as demonstrated by Fruhstorfer et al. (2002). Figure 3-44 shows the SEM micrograph (Figure 3-44a) and the AFM topograph (Figure 3-44c) of the electrolytically polished surface of the superalloy NIMONIC PE16 with the protruding caps of the second phase parti-

cles. In contrast to AFM, where corrections were necessary to take into account the exact tip radius, corrections for the very small electron probe diameter are not urgently required in HRSEM. The size distribution function and mean radius of the second phase particles calculated from HRSEM (Figure 3–44b) and AFM (Figure 3–44d) data are in excellent agreement with those gained earlier by TEM (Nembach, 1996). The distinct advantages of HRSEM in this application are that micrographs are readily recorded and the data can be processed without additional correction procedures.

The characterization of porous materials such as porous silicon or porous aluminum oxide gains increasing attention because of important potential applications (see, e.g., Anglin et al., 2004; Galca et al., 2003; Pan et al., 2004; Yamazaki, 2004; Z. X. Zhao et al., 2005; Y. C. Zhao et al., 2005). Among others, HRSEM is an indispensable tool for structural characterization of porous materials taking advantage of the large depth of focus and the high resolution obtainable. Figure 3–45 shows high-resolution SE and BSE micrographs of the surface and cross section of porous aluminum oxide, which exhibits a network with randomly distributed, but almost perfectly aligned cylindrical pores perpendicular to the substrate. The simultaneous imaging of the surface and the cross section reveals information about the three-dimensional specimen structure. Under the conditions given the SE mode yields higher resolution than the BSE mode.

However, the BSE mode is of significant importance if greater information depth and material differentiation are required. Figure 3–46 shows SE and BSE micrographs of temperature-sensitive hydrogels, based on poly (vinylmethyl ether) (PVME), with ferromagnetic properties due to incorporated nickel particles used as ferromagnetic filler. The contrast in the SE micrograph (Figure 3–46a) is mainly caused by

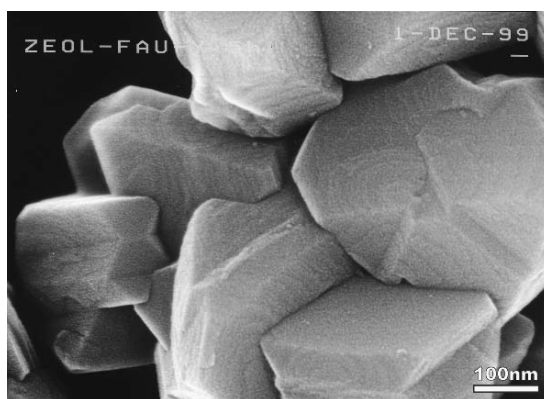


Figure 3–43. Secondary electron micrograph of zeolite FAU (faujasite) particles recorded with an “in-lens” field emission SEM at 10 kV. The particles are adsorbed to a thin hydrophilic amorphous carbon film and rotary shadowed with 1.5 nm platinum/carbon (Pt/C) at an elevation angle of 65° and, additionally, unidirectional shadowed with 2 nm Pt/C at an elevation angle of 10°. The habit, intergrowth of particles, and growth steps at the surface are clearly visible. (Specimen kindly provided by Dr. G. González, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela.)

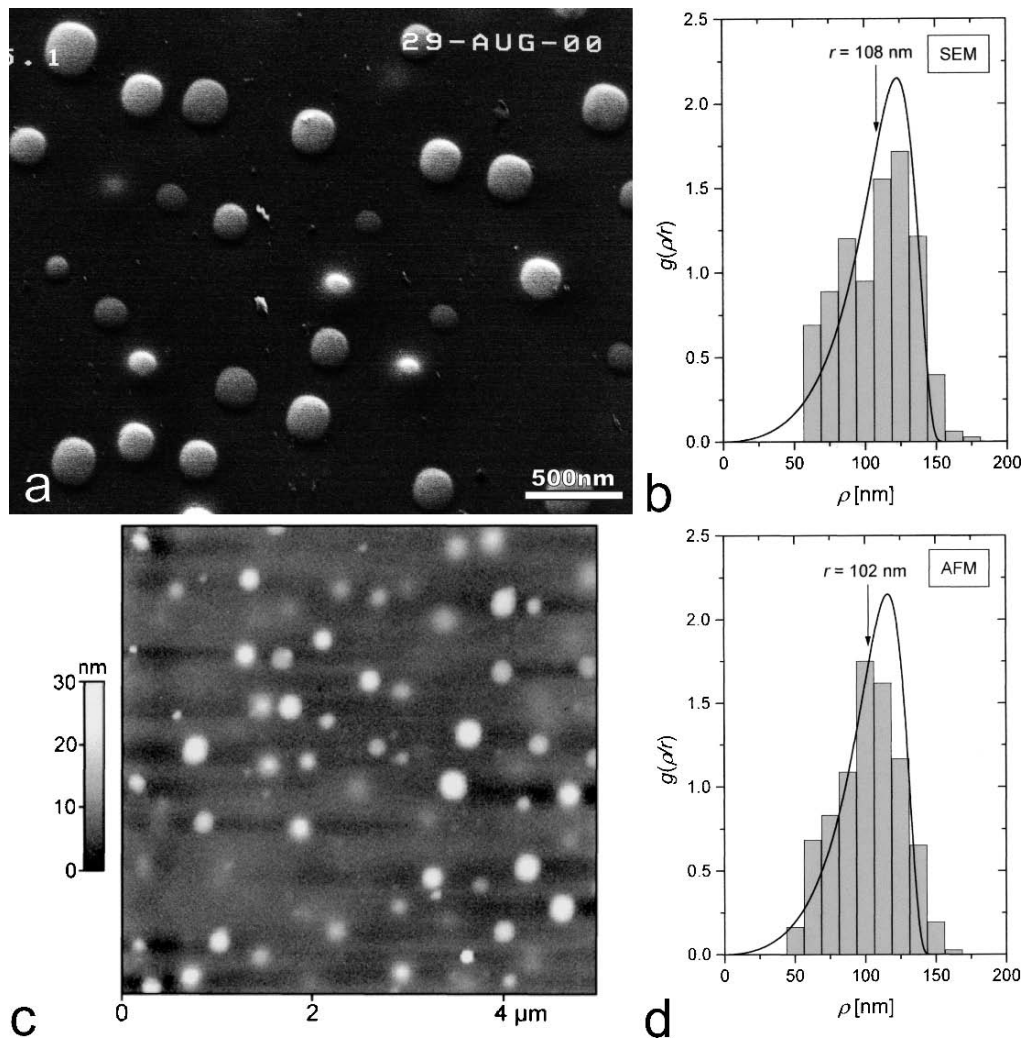


Figure 3-44. Surface of electrolytically polished superalloy NIMONIC PE16. Secondary electron micrograph recorded with an “in-lens” FESEM at 10kV (a) and AFM topograph (c). The related distribution functions g of the true radii ρ are shown for the HRSEM in (b) and for the AFM in (c). [Adapted from Fruhstorfer et al. (2002); with kind permission of Taylor & Francis Ltd., <http://www.tandf.co.uk/journals>.

the very thin membrane-like PVME, which envelops the nickel particles, whereas the BSE image (Figure 3-46b) has a strong material contrast component due to the nickel particles underneath the PVME membrane. This new class of hydrogels is of great interest for delivery of materials at the micro- and nanometer scale.

As mentioned in Section 2.2.3, the high-resolution “in-lens” FESEM equipped with an annular dark-field detector is capable of mass measurements on thin specimens (Engel, 1978; Wall, 1979) at a resolution close to that of a dedicated STEM (Reichelt et al., 1988; Krzyzanek and Reichelt, 2003; Krzyzanek et al., 2004). Mass measurement of molecules

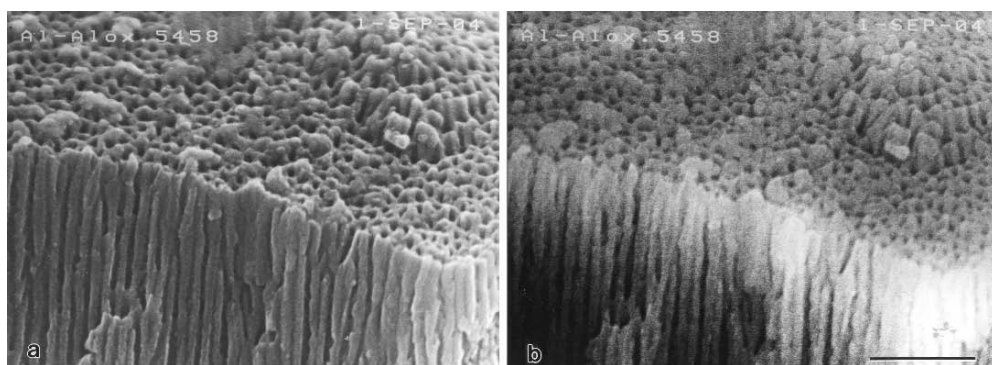


Figure 3–45. High-resolution micrographs of the surface (upper part of image) and the cryosectioned cross section (lower part of image) of porous aluminum oxide recorded with an “in-lens” FESEM at 8 kV with SE (a) and BSE (b). The specimen is rotary shadowed with 1.5 nm platinum/carbon. The bar corresponds to 200 nm. (Specimen kindly provided by Drs. C. Blank and R. Frenzel, Institut für Polymerforschung Dresden e.V., Dresden, Germany.)

and molecular assemblies are of great importance in biophysics and structural biology (for review see, e.g., Müller and Engel, 2001).

Finally, nanotechnology and “nanoelectromechanical systems” (NEMS) are additional fields in which HRSEM is used as a tool for monitoring processes, detecting defects, or measuring sizes and distances, e.g., in nanodevices, which will contain nanotubes, nanoparticles, nanowires, and other particles (see, e.g., Aoyagi, 2002; Nagase and Kurihara, 2000; Nagase and Namatsu, 2004).

3.2 Low- and Very-Low-Voltage Scanning Electron Microscopy

Scanning electron microscopy with electron energies below 5 keV is usually designated as scanning low-energy electron microscopy

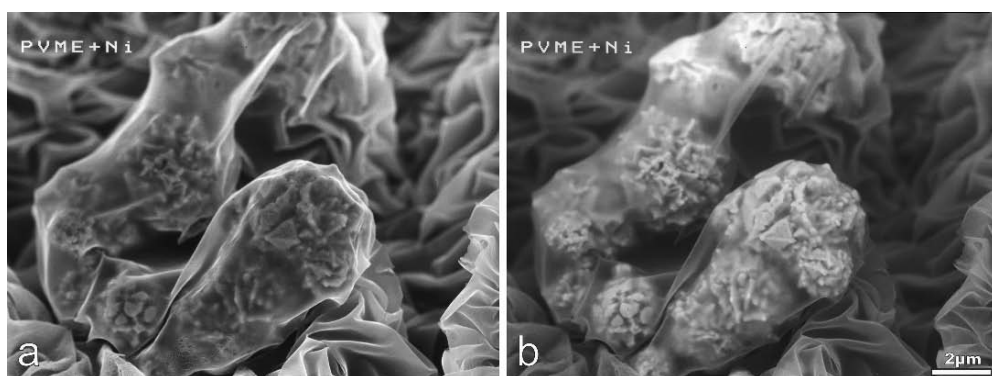


Figure 3–46. High-resolution micrographs of poly(vinyl methyl ether) (PVME) hydrogel with ferromagnetic properties filled with submicrometer nickel particles in the swollen state. The hydrogel was rapidly frozen, freeze dried, and rotary shadowed with an ultrathin layer of platinum/carbon. The SE (a) and BSE (b) micrograph were recorded with an “in-lens” FESEM at 10 kV, (Specimen kindly provided by Dr. K.-F. Arndt, Institut für Physikalische Chemie und Elektrochemie, Technische Universität Dresden, Dresden, Germany.)

(SLEEM) or, related to the acceleration voltage, LVSEM. The energy of 5 keV can be considered as some threshold energy because the monotonic dependence of the BSE coefficient on the atomic number breaks below this (cf. Section 2.2.2). A second prominent energy threshold is at about 50 eV, which corresponds to the electron energy with minimum inelastic mean free path of electrons in matter (Seah and Dench, 1979). Therefore, scanning microscopy with electron energies below 50 eV is designated as scanning very low-energy electron microscopy (Müllerova and Frank, 2003) or, related to the acceleration voltage in the scanning mode, very low-voltage scanning electron microscopy (VLVSEM).

What is the motivation for low electron energy operation in SEM? What are the advantages expected at low energies and what are the inherent disadvantages?

Clearly, almost all of the advantages for working at low energy derive directly from the energy dependence of the electron–specimen interaction (see Section 2.2). The advantages include the following:

1. The penetration depth of the impinging electrons decreases with decreasing energy due to the reduced electron range R [Eq. (2.29)], i.e., the excitation volume in the specimen shrinks (cf. Figure 3–13) and the volume emitting SE2 and BSE2 approaches the volume emitting SE1 and BSE1 (cf. Figure 3–14). As a result the edge effect, i.e., overbrightening of edges, is strongly reduced or even suppressed completely.

2. The SE yield δ increases because of the reduced electron range and the SE are generated near the surface, where they can escape (cf. Figure 3–15). As a result, the SNR of the SE signal increases with decreasing energy as low as $E_{0,m}$.

3. As the SE yield increases, the total amount of emitted electrons approaches unity (cf. Figure 3–16). Because of the conservation of electric charge [Eq. (2.40)] the amount of incoming and emitted charges is balanced and, consequently, the specimen current equals zero. That means that at this particular electron energy E_2 no electric conductivity of the specimen is required. Ideally, imaging of electric insulators without conductive coating becomes possible. For normal incidence, E_2 is within the range 0.5–5 keV for most of the materials. E_2 increases with the increasing angle of beam incidence θ according to

$$E_2(\theta) = E_2(0)/\cos^2\theta \quad (3.2)$$

where $E_2(0) = E_2(\theta = 0)$ (Joy, 1989), i.e., increases as θ increases.

4. As mentioned above, the monotonic dependence of the BSE coefficient on the atomic number breaks below 5 keV (Reimer and Tollkamp, 1980; Schmid et al., 1983). This behavior enables the material contrast in the BSE image to be finetuned by choosing the most suitable electron energy (Müllerova, 2001).

5. There is a reduced depth of specimen radiation damage (see Section 2.5). At very low electron energies, say less than 30 eV, the elastic scattering dominates and radiation damage becomes negligible (Müllerova and Frank, 2003).

The problems and disadvantages inherent to microscopy at low electron energy concern both the instrumentation and the specimen include the following:

1. Reduced resolution due to chromatic aberration and diffraction [see Eqs. (2.8)–(2.10)].
2. Stronger sensitivity of the electron beam to electromagnetic stray fields.
3. Special detector strategies required for SE and BSE.
4. Enhanced contamination rate, which can be counteracted by ultrahigh vacuum.
5. Reduced topographic contrast in SE and BSE micrographs. For electron energies below 5 keV, the increase of SE yield $\delta(\theta)$ with increasing θ [cf. Eq. (2.31)] drops way down to 0.5 keV as shown for different metals experimentally and by Monte Carlo simulation (Joy, 1987a; Böngeler et al., 1993). Similarly, the backscattering coefficient $\eta(\theta)$ shows less increase with θ than given by Eq. (2.34), which is more pronounced at low electron energies (Böngeler et al., 1993).
6. Reduced material contrast, because the differences of the backscattering coefficient between low and high atomic number material become smaller (Darlington and Cosslett, 1972; Lödding and Reimer, 1981; Reimer and Tollkamp, 1980).

3.2.1 Electron Lenses

Modern commercial field emission scanning electron microscopes can operate usually from 30 keV down to 1 keV or even 0.5 keV, i.e., that energy range covers conventional electron energies and most of the low-energy region. Improved computer-aided methods enable electron optical systems to be designed that have high performance within the whole energy range mentioned above. Compared with the old-fashioned thermionic gun scanning electron microscopes the aberration coefficients of the objective lens were improved dramatically for modern field emission instruments commercially available: C_s was reduced by a factor of about 30 down to $C_s = 1.6$ mm (Uno et al., 2004), and C_c was reduced by a factor of about 10. With the ultrahigh-resolution objective lens, the CFEG, improved electrical and mechanical stability, as well as strongly reduced specimen contamination rate, the resolution obtained with test specimens amounts presently to 0.5 nm (at 30 keV) and 1.8 nm (at 1 keV) (Sato et al., 2000). Those values are exemplary for the high performance of commercial FESEM over an energy range from 1 to 30 keV, though obtained with special test specimens. Very recently, a new commercial FESEM became available equipped with a spherical and chromatic aberration corrector, which consists of four sets of a 12-pole component that corrects the spherical and chromatic aberration simultaneously (Kazumori et al., 2004). Using the spherical and chromatic correction, the resolution obtained with the test specimen amounts to 1.0 nm (at 1 keV) and 0.6 nm (at 5 keV) (Kazumori et al., 2004).

Electrostatic as well as combined magnetic and electrostatic lenses in LVSEM are a very interesting alternative to the magnetic lenses mentioned above. Microscopes equipped with this type of objective lens permit nonconstant beam energy along the column, i.e., the beam electrons pass the column with high energy and are decelerated to low energy in the immersion electrostatic lens. First, the magnitude of the

aberrations of immersion electrostatic lenses corresponds to the high energy at the entrance side. A more detailed treatment of the estimation of electrostatic lenses is beyond the scope of this section (see, e.g., Lencová and Lenc, 1994; Lencová, 1997). Second, the high electron energy in the column is advantageous because the gun brightness increases with electron energy [see Eqs. (2.4) and (3.1)] and electromagnetic stray fields result in less deterioration of the electron beam at high energy. The combined magnetic–electrostatic objective lens (Frosien et al., 1989) has aberration coefficients as low as $C_s = 3.7$ mm and $C_c = 1.8$ mm. Martin et al. (1994) achieved with this lens a resolution of 2.5 nm at 5 keV, 4.0 nm at 1 keV, and 5.0 nm at 0.5 keV.

Very low landing energies of the electrons can be realized with a retarding-field SEM. There are several retarding-field configurations described in the literature but basically in all of them the specimen is connected to the adjustable bias supply U_{sp} (e.g., Zworykin et al., 1942; Paden and Nixon, 1968; Zach and Rose, 1988a,b; Munro et al., 1988; Müllerova and Lenc, 1992). The landing energy of the beam electrons simply is given by the difference $E_0 - eU_{sp}$. Using retarding-field SEM, landing energies of a few electronvolts are achievable and recently micrographs with reflected electrons even at 0.5 eV were obtained (Müllerova et al., 2001).

With the availability of magnetic materials having high coercive force permanent rare-earth-metal magnets attract attention as replacements for magnetic lens coils (Adamec et al., 1995). Khursheed (1998) proposed a portable SEM column design, which makes use of permanent magnets. The column of this miniature SEM amounts to a height of less than 12 cm and is designed to be modular, so that it can fit onto different specimen chamber types, and can also be readily replaced. Focusing of the electron beam onto the specimen can be achieved by varying the specimen height or by an outer magnetic slip ring on the objective lens, which controls the strength of the magnetic field on the axis. Scanning of the beam is performed by deflection coils, which are located above and within the permanent magnet objective lens. A high-resolution miniature SEM with a total height of less than 5.5 cm, proposed by Khursheed (2000), uses a permanent magnet objective lens that lies outside the vacuum with spherical and chromatic aberration coefficients (parameters: $E_0 = 1$ keV, $WD = 7.5$ mm) of 0.36 and 0.6 mm, respectively. These aberration coefficients are about an order of magnitude smaller than those for conventional SEMs with comparable working distance conditions.

Miniaturization of the SEM column has advantages such as microlenses with small aberration coefficients, reducing the influence of electromagnetic stray fields and of the electron–electron interaction, improving the mechanical stability, and reducing the demands on space for the microscope. Chang et al. (1990) proposed a miniaturized electron optical system consisting of a field emission microsource and an electrostatic microlens for probe forming with performance, exceeding that of a conventional system over a wide range of potentials (0.1–10 kV) and working distances (up to 10 mm). Liu et al. (1996) proposed another design that has a column length of only 3.5 mm and can be

operated over a wide retarding range of potentials (0.1–10 kV). The instrument has an optimized design (microcoincident lens followed by a retarding region) to minimize the primary beam diameter and to maximize secondary electron collection (approximately 50% of SE are collected).

3.2.2 Detectors and Detection Strategies

As mentioned previously, modern commercial FESEMs can operate usually from 30 keV down to 1 keV or even 0.5 keV; the commonly used detectors and detection strategies of these instruments were discussed in Sections 2.1.3 and 3.1.3. It is clear from Figure 3–12 that the lower the energy E_0 of the beam electrons the lower the energy difference between the secondary and backscattered electrons. The lower the difference of the different signal electrons the more difficult is their separation. At very low electron landing energies SE and BSE are almost indistinguishable, thus the total emission is detected.

Very recently, Kazumori (2002) proposed two new SE detection systems in a commercial FESEM, the so-called “r-filter” and the “Gentle Beam,” for high-resolution observation of tiny regions of uncoated specimens that include domains with low or no electrical conductivity at low electron energies. The “r-filter” is an energy-selective SE detection system, and the “Gentle Beam” system consists of a strongly excited conical lens (semi-in-lens type) that can retard the beam electrons and works together with the 1–2 kV negatively biased specimen. The “Gentle Beam” system improves the obtainable resolution significantly for acceleration voltages below 3 keV. SE micrographs of good quality down to 100 V can be recorded by the system (Kazumori, 2002).

In the past few years there have been many attempts to improve the noise and time characteristics of semiconductor detectors, to improve the properties of microchannel plate detectors, and to increase the light output as well as lower the energy threshold of scintillator-based BSE detectors below 2 keV (Autrata and Schauer, 2004; Schauer and Autrata, 2004). Compared with the semiconductor detectors and the MCP, the scintillator-photomultiplier still possesses the best SNR and bandwidth characteristics.

3.2.3 Contrast Formation

The contrast formation in LVSEM and VLVSEM is controlled—as at conventional electron energies—by the electron specimen interaction at the electron energy used, the specific signal considered, the detector, and the detection geometry. However, the contrast formation at low energies is much more complex than for electron energies ranging from 5 to 30 keV. A variety of reasons accounts for this complexity. For example, the BSE coefficient is for a given material almost constant and the SE yield depends just weakly on the electron energy for $E_0 \geq 5$ keV. In contrast to this, the monotonic increase of the BSE coefficient with rising atomic number breaks below 5 keV as previously mentioned and, additionally, the BSE coefficient becomes dependent on the electron energy for many chemical elements. Furthermore, the signals obtained at low electron energies are affected more strongly on electron beam-

induced contamination or other thin layers on the surface, which is caused by the strongly reduced electron range.

Nevertheless, the main types of contrast, such as topographic, compositional, voltage, electron channeling, crystal orientation, and type-1 and type-2 magnetic and mass-thickness contrast, are also observed in LVSEM, although it is in many respects different from that obtained at conventional energies. There are also several observations that evidently show some "chemical" or "electronic" contrast, i.e., contrast that does not result from an increase in the mean atomic number of the specimen (e.g., Wollman et al., 1993; Bleloch et al., 1994; Perovic et al., 1994). Although these effects may also be visible at conventional energies they are most readily observed at low energies where the SE yield is higher.

The thickness contrast described in Section 3.1.5 also plays an important role in LVSEM of electric insulators. Though direct imaging of electrical insulators without electric charge-up should be feasible at electron energy E_2 , where incoming and emitted charges are balanced, in practice it often does not work for various reasons. Therefore, coating the specimen surface with an ultrathin very fine-grain metal film (Peters, 1982) by Penning sputtering or by evaporation in oil-free high vacuum is often done. As in high-resolution SEM with conventional beam energies, the film plays an important role in contrast formation, in image resolution obtainable, and in the improvement of the SNR. The image contrast of coated specimens essentially depends on the projected film thickness, which will vary between the nominal film thickness and the maximum film thickness, which is several times greater than the nominal thickness in tilted regions (cf. Figure 3–39a). Monte Carlo calculations of the SE yield of a film of chromium at 2 keV also prove for low electron energy a monotonic increase with film thickness (Joy, 1987a).

3.2.4 *Selected Applications*

The application of LVSEM and VLVSEM logically seems likely in cases in which SEM at conventional acceleration voltages obviously would fail, e.g., the investigation of uncoated insulating materials and radiation-sensitive semiconductors. Another compelling reason is the necessity of a reduced electron range, e.g., with specimens having one or more very thin surface layers and samples possessing a spongy- or foam-like fine structure. SEM studies of these types of specimens aim at information restricted to the surface-near zone. With ever decreasing device dimension and film thickness this issue becomes more and more crucial. There are also noncompelling, but still for good reasons, which may aim at optimum imaging conditions at low electron energy, or LVSEM may be part of a series of increasing or decreasing electron energies over a wide energy range as used for depth profiling. Finally, there are also applications of LVSEM that may also work at conventional energies but are most readily obtained at low energies.

The LVSEM is widely applied to semiconductor structures relating to an examination of their geometry, critical dimensions, and local

voltages or currents, which may be either biased or induced by the electron beam. One example of an integrated circuit was previously shown in Figure 3–30. Figure 3–47 shows the cross-fractured semiconductor structure with Schottky barrier on tungsten contacts. A nanostructured two-dimensional lattice of 100-nm spaced inverted square pyramids in silicon used as standard for scanning probe microscopy is shown in Figure 3–48. Imaging of the uncoated lattice is necessary to avoid modifications of the standard by thin film coating, thus LVSEM is most appropriate.

Another challenging application for LVSEM is the quantitative characterization of the geometry and radius of very sharp tips for atomic force microscopy, which are necessary for many quantitative measurements with the AFM (e.g., Fruhsdorfer et al., 2002; Matzelle et al., 2000, 2003). Figure 3–49 shows two extremely sharp commercial tips. The tip radius at the very end amounts typically to 2–3 nm, thus only SE1 contribute to the signal. An optimum quality of SE imaging in terms of sharpness, contrast, and SNR can be obtained with electron energies ranging from about 3 to 10 keV. It seems worth mentioning that very sharp tips are interesting samples with which to study experimentally the delocalization of the secondary electrons.

The characterization of organic mono- and multilayers on solids is especially valuable in technology development, such as bio- and chemosensors, since detailed information on the film surface and its morphology is obtained. Figures 3–50 to 3–53 demonstrate with different mono- and multilayered ultrathin uncoated and coated organic films how direct information about the film thickness, step heights of the film, and differences in the “chemistry” and molecular packing density can be obtained. As shown by Figure 3–50a, upward and downward steps with height differences of a few nanometers can be readily

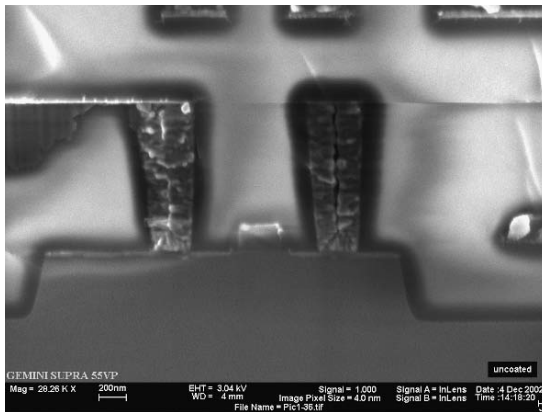


Figure 3–47. Secondary electron micrograph (“through-the-lens” detection) of a cross-fractured semiconductor structure with Schottky barrier on tungsten contacts. The image was recorded at 3.04 kV. The cross-fracture reveals the interior features and the potential barrier of 0.6 eV (dark due to voltage contrast). (Courtesy of Carl Zeiss NTS, Oberkochen, Germany.)

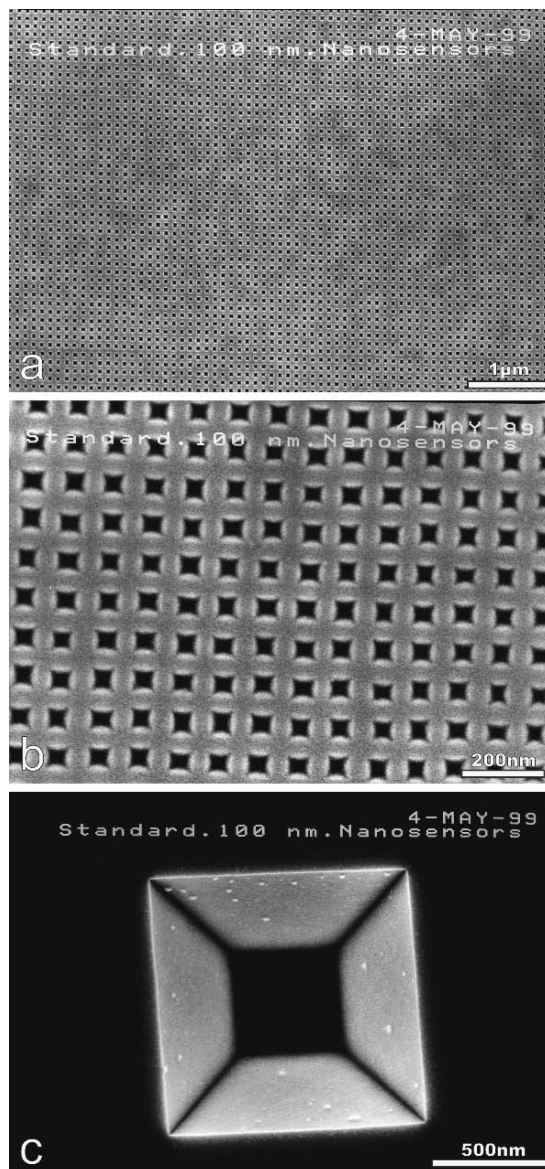


Figure 3-48. Secondary electron micrographs of an uncoated 100-nm calibration standard made from silicon for scanning probe microscopy (NANO-WORLD, Neuchatel, Switzerland) recorded with an “in-lens” field emission SEM at 3 kV. The calibration standard consists of a two-dimensional lattice (lattice constant = 100 nm) of inverted pyramids shown at different magnifications (a and b). (c) Structural details of a large pyramidal pit.

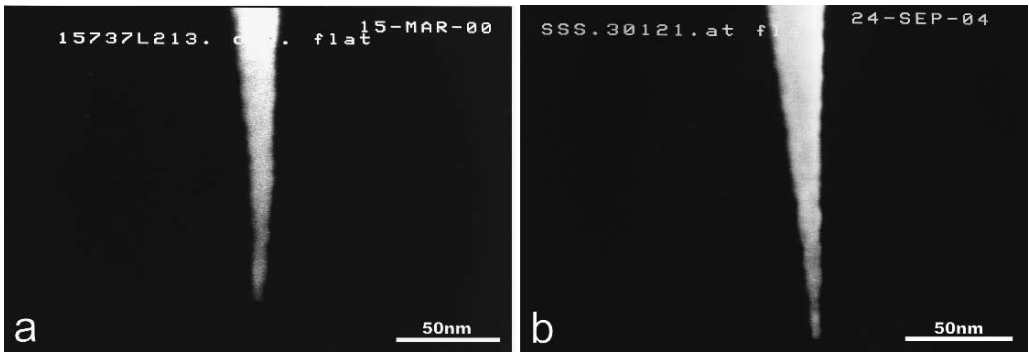


Figure 3-49. Secondary electron micrographs of uncoated SuperSharpSilicon AFM Probe silicon cantilevers for noncontact/tapping mode (NANOWORLD, Neuchatel/Switzerland) in atomic force microscopy recorded with an “in-lens” field emission SEM at 3 kV (a) and at 10 kV (b). The tip radius of both tips amounts to about 2–3 nm.

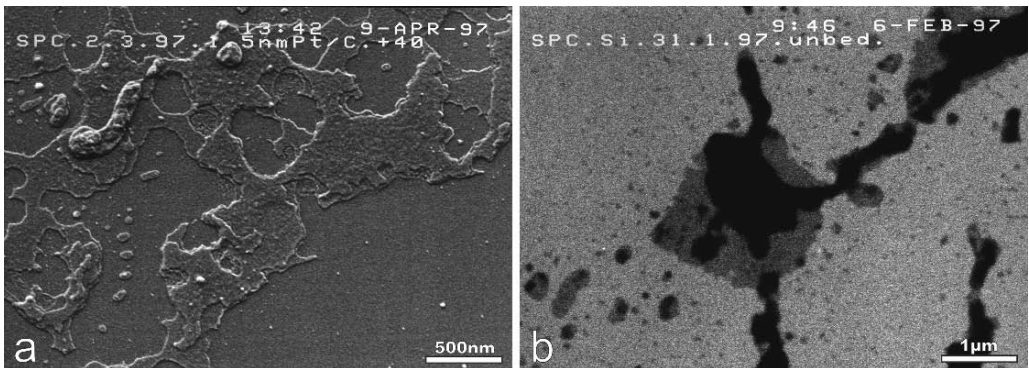


Figure 3-50. Secondary electron micrographs of a phospholipid/protein film [dipalmitoylphosphatidylcholine (DPPC):dipalmitoylphosphatidylglycerol (DPPG) (ratio = 4:1)/pulmonary surfactant protein C (SP-C; 0.4 mol%)] supported by a silicon wafer. The organic film has terrace-like regions of different thickness (height differences between terraces are between 5.5 and 6.5 nm (von Nahmen et al., 1997). Micrographs were recorded with an “in-lens” FESEM at 2 keV from the ultrathin platinum/carbon-coated film (tilted 40° around the horizontal axis) (a) and at 1.8 keV from the uncoated film (b). [Specimens kindly provided by Dr. H.-J. Galla and Dr. M. Siebert, Institut für Biochemie, University of Münster, Münster, Germany.]

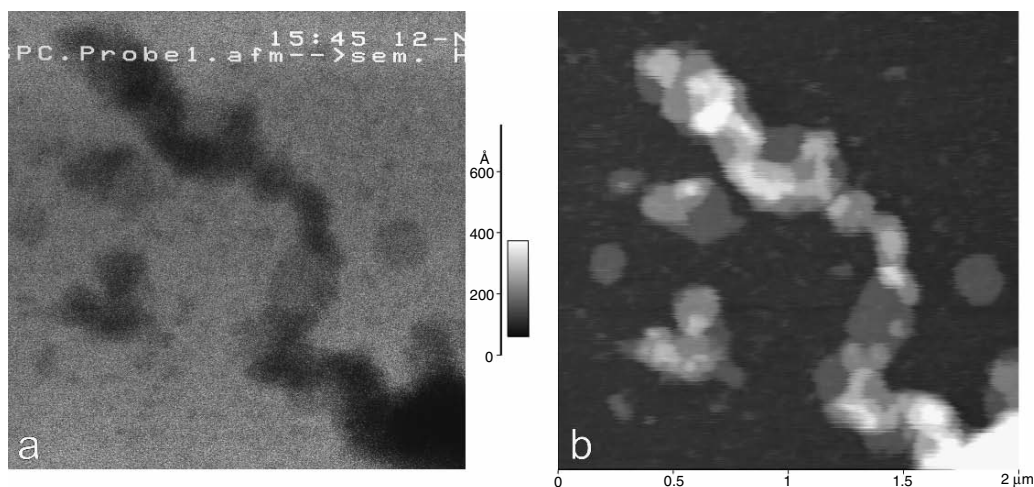


Figure 3-51. Secondary electron micrograph (a) and AFM topography (b) of the same area of an uncoated phospholipid/protein film [dipalmitoylphosphatidylcholine (DPPC):dipalmitoylphosphatidylglycerol (DPPG) (ratio = 4:1)/pulmonary surfactant protein C (SP-C; 0.4 mol%)] supported by a silicon wafer. The organic film has terrace-like regions of different thickness [height differences between terraces are between 5.5 and 6.5 nm (von Nahmen et al., 1997)]. The micrograph was recorded with an “in-lens” FESEM at 2 keV. The scale inbetween (a) and (b) represents the coding of brightness relative to the height used in the topograph (b). (Specimens kindly provided by Dr. H.-J. Galla & Dr. M. Siebert, Institut für Biochemie, University of Münster, Münster, Germany.)

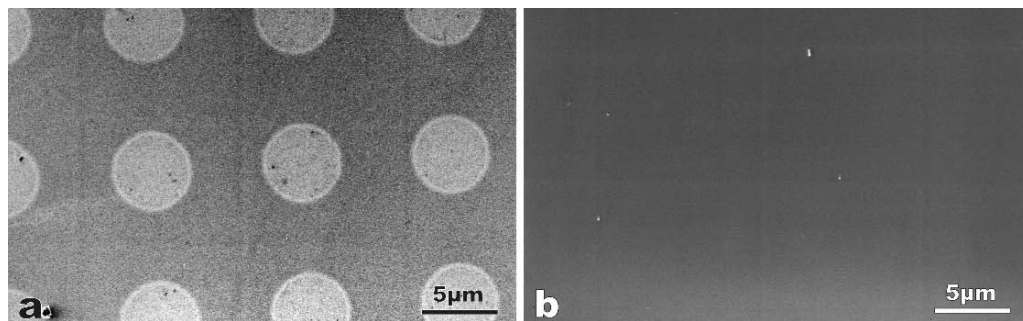


Figure 3-52. Secondary electron micrographs of a patterned self-assembled thiol monolayer on polycrystalline gold recorded at 2 keV with the “in-lens” FESEM. (a) Uncoated monolayer. The circular domains consist of $\text{—S(CH}_2\text{)}_{15}\text{CH}_3$ molecules (hydrophobic), which are surrounded by $\text{—S(CH}_2\text{)}_{12}\text{OH}$ molecules (hydrophilic). The contrast is due to the different end groups rather than to the small difference in chain length. (b) Monolayer coated with an ultrathin platinum/carbon film. (Specimen kindly provided by Dr. G. Bar, Freiburger Material Forschungszentrum, Freiburg, Germany.)

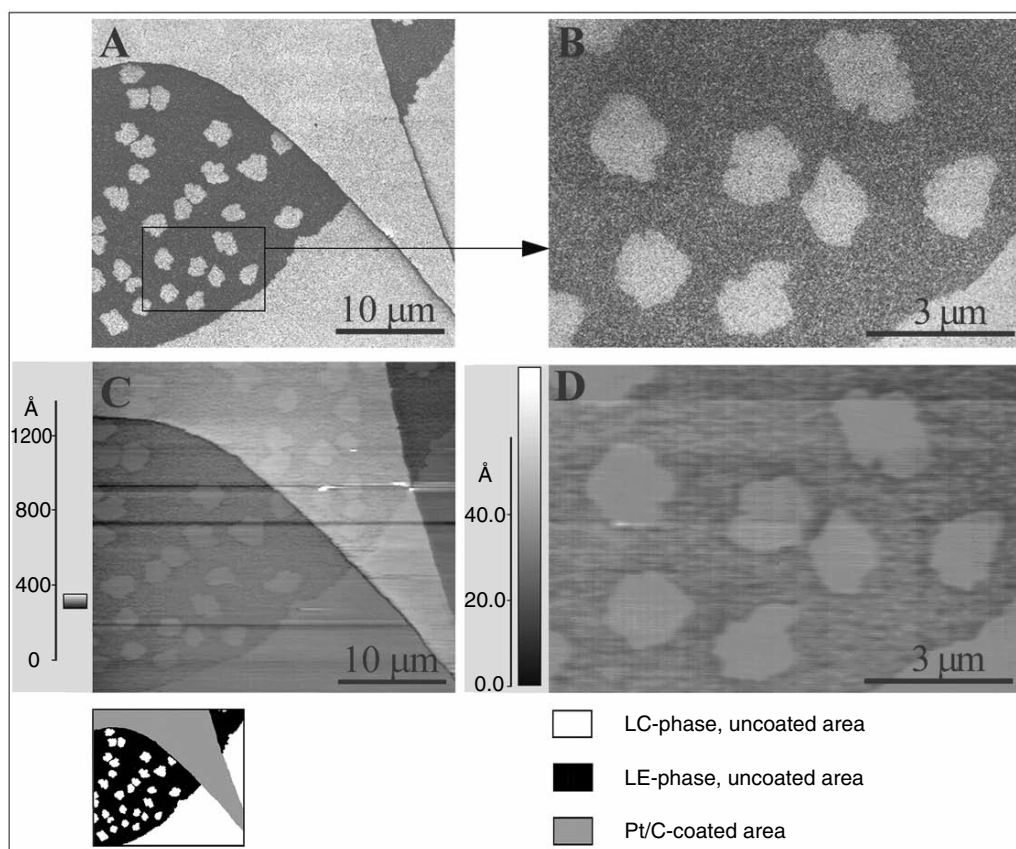


Figure 3-53. Secondary electron micrographs (A and B) and AFM topographs (C and D) of the same area of a 1,2-dipalmitoyl-*sn*-glycero-3-phosphothioethanol (DPPTE) monolayer on a silicon wafer having domains with densely [liquid condensed (LC) phase] and loosely [liquid expanded (LE) phase] packed molecules. The specimen was masked by a TEM finder grid and then coated with an ultrathin platinum/carbon film to obtain neighboring coated and uncoated areas on the specimen (for details see Bittermann et al., 2001). The SE micrographs were recorded with an “in-lens” FESEM at 5 keV. The brighter regions of the SE micrograph (B) correlate with the elevated domains (LC phase) in the AFM topograph (D), whereas the darker regions correlate with the LE phase. In contrast to the SE micrograph (A) the height differences in coated areas of the film, which are related to its molecular packing density, are still visible in the AFM topograph (C). [From Bittermann et al. (2001); with kind permission of the American Chemical Society, Columbus, OH.]

identified on a tilted sample coated with an ultrathin conductive film. Whereas the step of constant height reveals in an “in-lens” SEM a constant intensity at normal electron beam incidence, tilting causes an asymmetry such that steps can face upward or downward, which leads to an increase or decrease of their image intensities, respectively. Uncoated organic layers on solids usually reduce the SE yield as shown in Figure 3–50b. As demonstrated in Figure 3–51 by a comparison of an SE micrograph with an AFM topograph of the same area, the SE intensity decreases with increasing thickness of the organic layer (Reichelt, 1997). The monotonic dependence of SE intensity and the thickness of the organic film enables its thickness to be mapped without destruction of the film. The influence of organic film thickness on the SE yield vanishes after ultrathin coating of the organic film as proven by Figure 3–50a.

Figure 3–52a demonstrates that the SE yield also depends on the chemical nature of the molecules assembling an organic film. For example, differences in the terminal group of molecules obviously cause a significant difference in the SE yield, which creates a sufficient chemical contrast in the micrograph. This chemical contrast vanishes after ultrathin coating of the organic film (Figure 3–52b). Finally, Figure 3–53 shows that the SE yield is sensitive to the molecular packing density of the organic film, i.e., the number of organic molecules per area (Bittermann et al., 2001).

It is easy to understand that the BSE signal is not sensitive to the film thickness and differences in the “chemistry” and molecular packing density, because the backscattering of thin low atomic number films is negligible compared with those of the substrate having a significantly higher atomic number.

Figures 3–54 and 3–55 show secondary electron micrographs of an uncoated glass micropipette and a microtome glass knife, which are almost free of electric charging. However, at higher magnifications the typical signs of charging occur.

The characterization of sponge-like microstructures, such as hydrogels and microgels, is a further challenging application of LVSEM, where a large depth of focus, high resolution, and low penetration power (i.e., small electron range) of the electron beam are required. Figure 3–56 shows a stereopair of highly magnified SE micrographs of a hydrogel. The optimum imaging quality of fine structural details well below 10 nm was obtained with electron energies around 2 keV.

Figures 3–57 and 3–58 show a set of secondary electron micrographs recorded from biological samples at low magnification with different electron energies. The micrographs demonstrate to what extent the contrast and information depth vary with the electron energy in a range from 0.4 to 30 keV, which corresponds to about the accessible energy range of commercial FESEMs. As yet, not all of the contributing contrast mechanisms are fully understood, thus the interpretation of micrographs recorded at a specific selected energy requires great care.

Finally, LVSEM is also a promising and efficient alternative to conventional approaches for micromorphological and microstructural

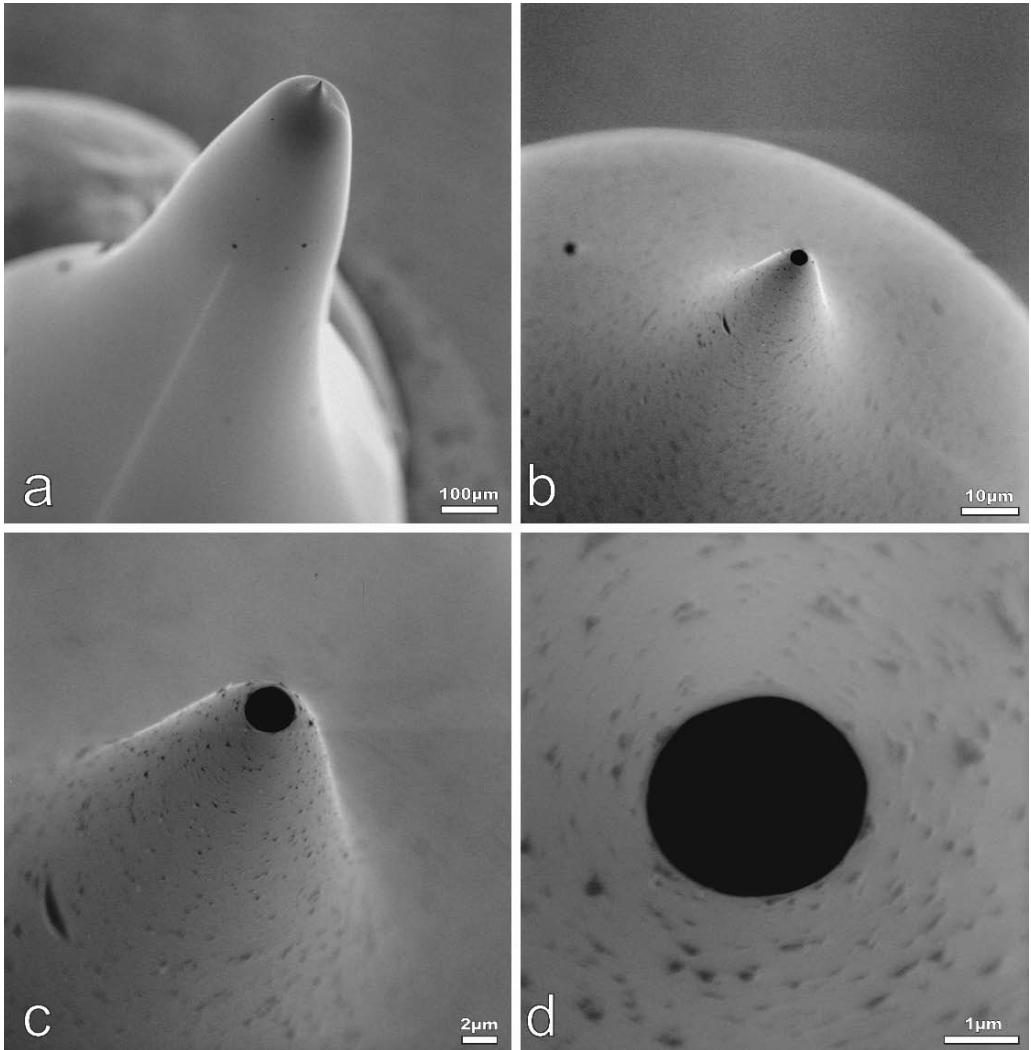


Figure 3-54. Secondary electron micrograph series (a-d) of increasing magnification of an uncoated glass micropipette recorded at 2 kV with an “out-lens” FESEM. The uppermost part of the tip of the micropipette is within the depth of focus. The lower part is out of focus.

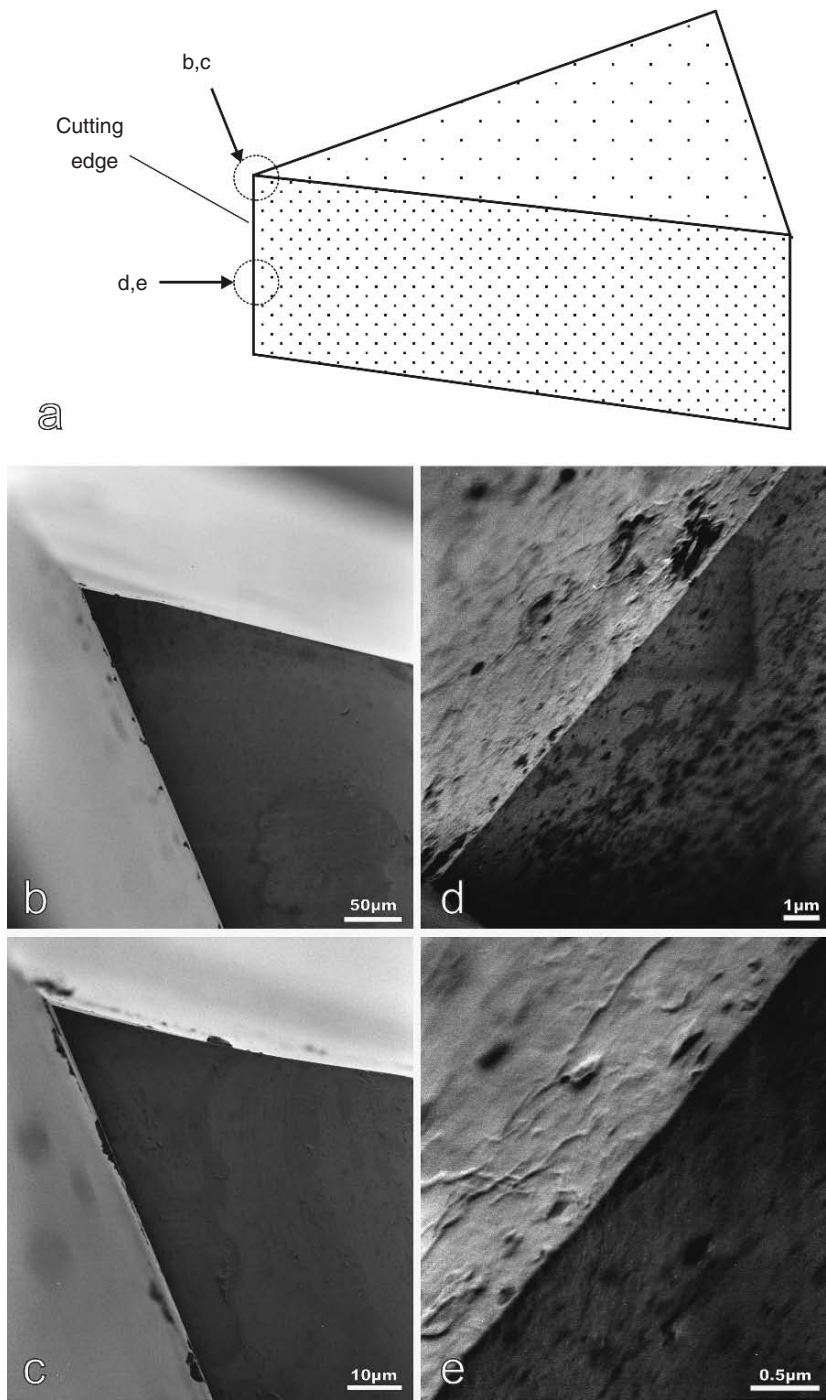


Figure 3-55. Scheme and SE micrographs of an uncoated microtome glass knife recorded at 1 kV with an “out-lens” FESEM. The arrows in (a) indicate the two directions of the electron beam related to the glass knife, which were used for imaging. (b and c) The electron beam has a shallow angle against the cutting edge. Only the uppermost part of the cutting edge is within the depth of focus. (d and e) The electron beam impinges perpendicularly onto the cutting edge. The different mean brightness of the clearance angle side and backside of the knife is due to the effect of the detection geometry of the ET detector.

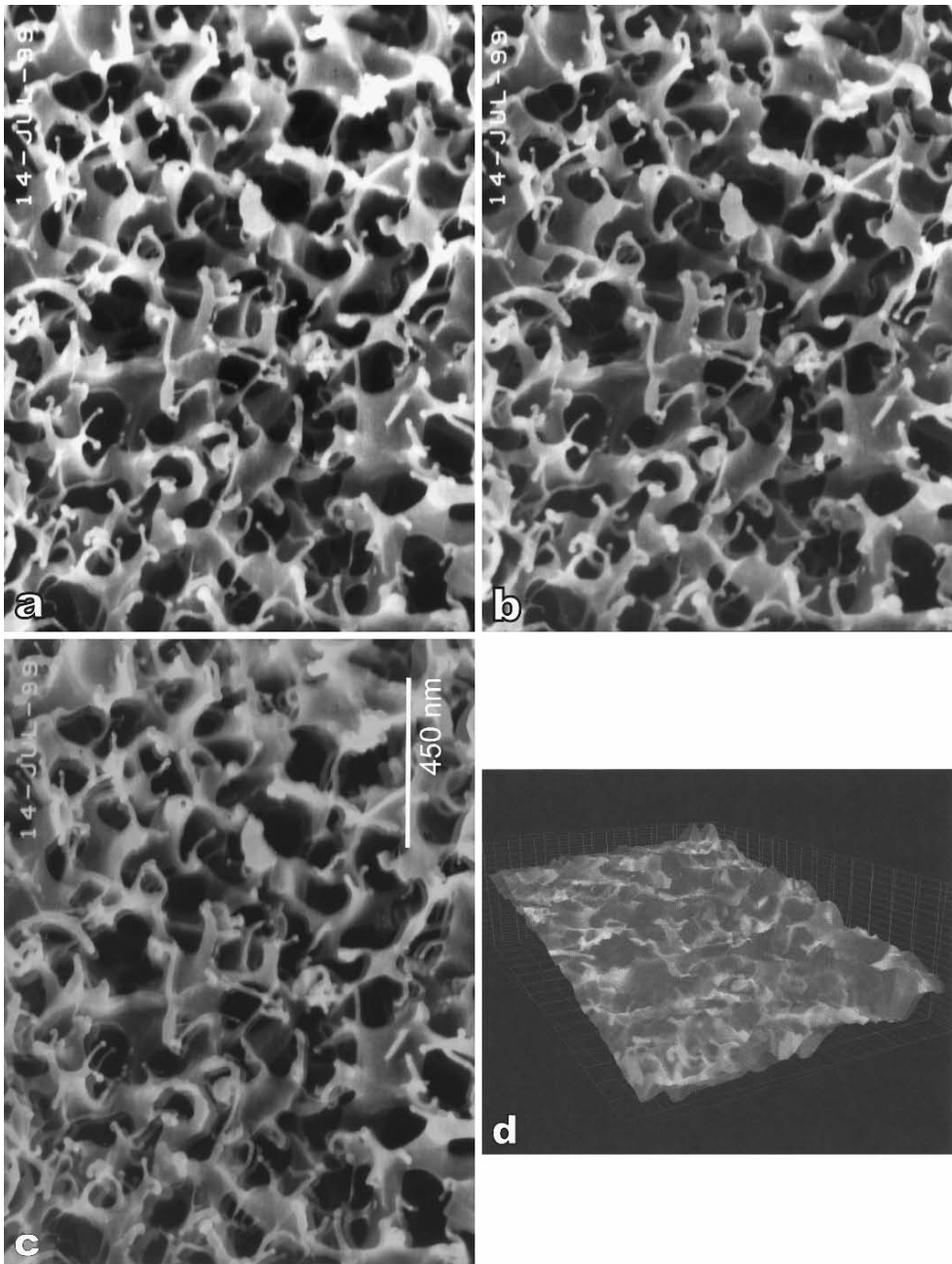


Figure 3-56. Stereo pair of SE micrographs (a and b) of the hydrogel poly-(*N*-isopropylacrylamide) (PNIPAAm) in the swollen state recorded at 2 keV with the “in-lens” FESEM. The specimen was rapidly frozen, freeze dried, and ultrathin rotary shadowed with platinum/carbon (for details see Matzelle et al., 2002). (c) Red–green stereo anaglyph prepared from (a and b). The tilt axis has a vertical direction. (d) Red–green stereo anaglyph in a “bird view.” (For parts c and d, see color plate.)

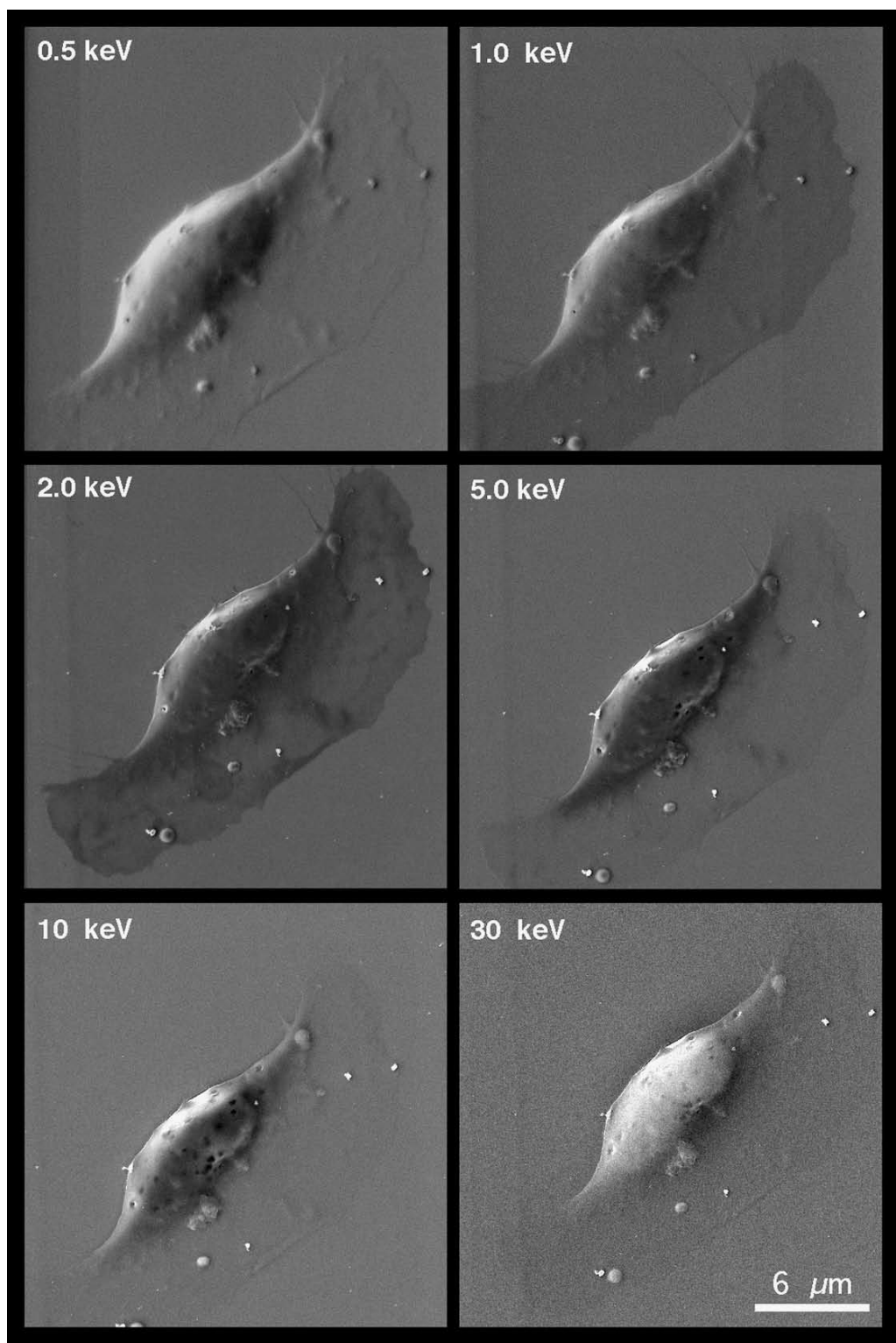


Figure 3–57. Secondary electron micrograph series of increasing electron energies from 0.5 to 30 keV from a keratinocyte. The micrographs are recorded with an “in-lens” FESEM. The image contrast varies significantly with the electron energy. Inhomogeneities in the “leading edge” of the keratinocyte, which has a thickness of about 200–400 nm, are most clearly visible at 2 keV. (Micrographs kindly provided by Dr. R. Wepf, Beiersdorf AG, Hamburg, Germany.)

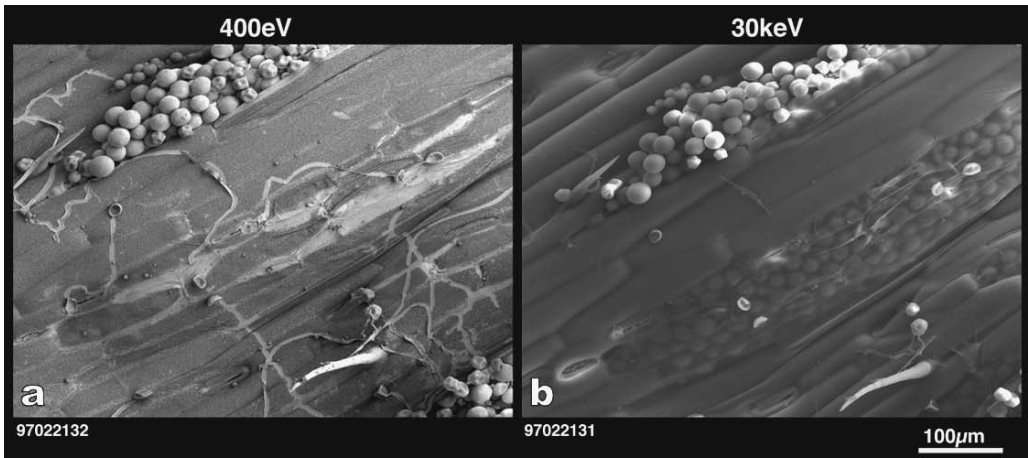


Figure 3–58. Secondary electron micrograph pair of the cuticula of a leaf recorded at electron energies of 0.4 (a) and 30 keV (b) with an “in-lens” SEM. The low-energy image contains information only from the surface whereas the 30-keV image also reveals information about structural features below the surface, e.g., new spores, which are not visible in (a). (Micrographs kindly provided by Dr. R. Wepf, Beiersdorf AG, Hamburg, Germany.)

characterization of polymers (Berry, 1988; Butler et al., 1995; Brown and Butler, 1997; Sawyer and Grubb, 1996).

4 Scanning Electron Microscopy at Elevated Pressure

The scanning electron microscopic investigation of specimens must meet several requirements, which were mentioned in previous sections. To sum it up, it can be said that specimens (1) have to be compatible with the low pressure in the specimen chamber ($\sim 10^{-3}$ Pa in conventional SEM and 10^{-5} – 10^{-4} Pa in field emission SEM), (2) have to be clean, i.e., the region of interest has to give free access to the primary beam, (3) need sufficient electrical conductivity, (4) need to be resistant to some extent to electron radiation, and (5) have to provide a sufficient contrast. In a narrower sense, only metals, alloys, and metallic compounds fulfill those requirements. Numerous preparation procedures mentioned in Section 2.4 were developed in the past and are still in the process of improvement, to provide a sufficient electrical conductivity to nonconductive specimens, to remove the water in samples, and to replace it or to rapidly freeze it in a structure-conserving manner. Nevertheless, there was and still is enormous interest in investigating specimens in their genuine state.

Thirty years ago Robinson (1975) proposed examining any uncoated insulating specimen in the SEM at high accelerating voltages in the specimen chamber, which had been modified to contain a small residual water vapor environment. It appeared that the presence of the water vapor sufficiently reduced the resistance of the insulator so that no charging effects were detected in backscattered electron micrographs. Danilatos (1980) developed an “atmospheric scanning electron microscope” (ASEM), which later was called an “environmental scanning

electron microscope” (ESEM®) (Danilatos, 1981) and is now a registered trademark. To enable the investigation of water and water-containing specimens in their native state at stationary conditions a minimum pressure of water vapor of about 612 Pa is required at 0°C (cf. Figure 3–59). Stationary conditions in the specimen chamber of the SEM can be accomplished by controlling the water vapor pressure p in close vicinity of the specimen as well as the specimen temperature T such that the p – T values always correspond to points on the solid p – T graph in Figure 3–59. For example, at 20°C a water vapor pressure as large as about 2330 Pa is required for stationary conditions. p – T values below the solid graph, e.g., 300 Pa at 0°C (Figure 3–59), corresponds to a relative humidity of less than 100%, thus representing nonstationary conditions.

How can stationary conditions be reached during imaging of a wet sample in the specimen chamber of an SEM? Figure 3–60 shows the cross section of the ESEM, which permits investigations at pressures sufficient for stationary conditions. Basically, the electron beam propagates in the column as in a conventional SEM until it reaches the final aperture. Then, since the pressure increases gradually as the electrons proceed toward the specimen, the electrons undergo significant scattering on gas molecules until they reach the specimen surface.

The electron–gas interaction is discussed in detail by Danilatos (1988). According to this study the average number of scattering events per electron n can be approximated by

$$n = \sigma_g p_g L / kT \tag{4.1}$$

where σ_g represents the total scattering cross section of the gas molecule for electrons, L is the electron path length in gas, and k is the Boltzmann constant. These approximations hold for $\Lambda \gg L$, where Λ represents the

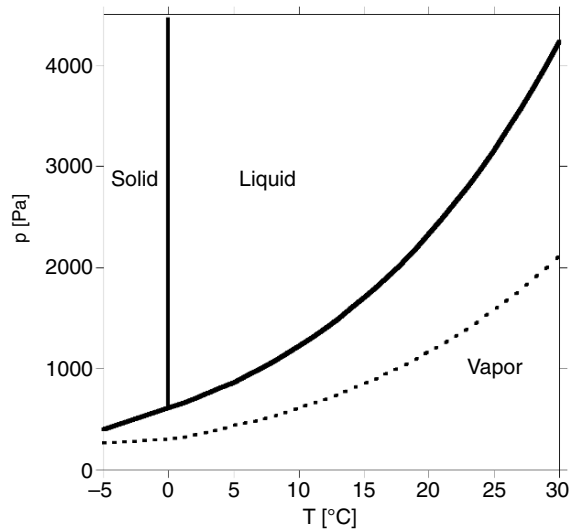


Figure 3–59. Phase diagram of water. Solid line, 100% relative humidity (saturated vapor conditions); dashed line, 50% relative humidity. (Data from Lax, 1967.)

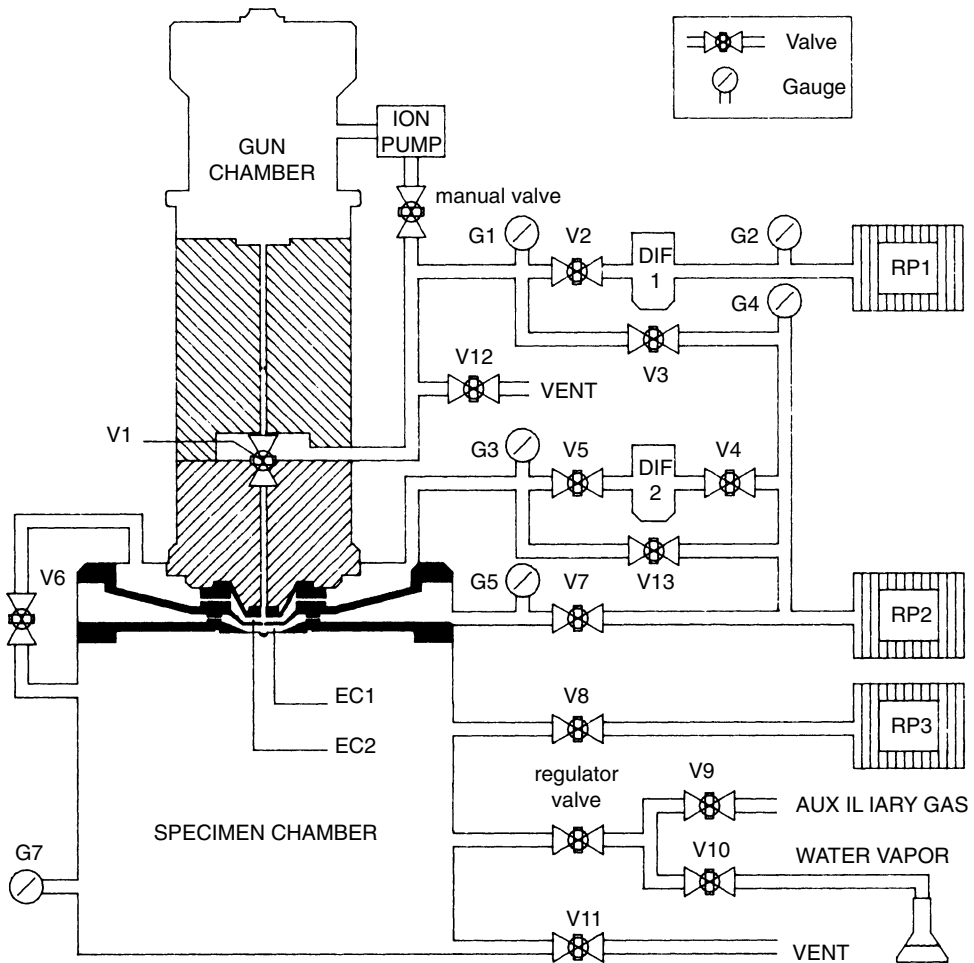


Figure 3–60. Schematic cross section of the first commercial Electroscan Environmental SEM (ESEM®) showing the vacuum and pumping system. Two pressure-limiting apertures separate the electron optical column from the specimen chamber. Differential pumping of the stage above and between the two pressure-limiting apertures ensures the separation of high vacuum in the column from low vacuum in the specimen chamber. The differential pumping of two stages and optimum arrangement of the pressure-limiting apertures can work successfully to achieve pressures up to 10^{-3} Pa in the specimen chamber. [From Danilatos, 1991; with kind permission of Blackwell Publishing Ltd., Oxford, U.K.]

mean free path of a beam electron in the gas. According to Eq. (4.1) the average number of collisions increases linearly with the gas pressure p_g and the path length in the specimen chamber. Furthermore, n depends via the scattering cross section on the type of gas molecules and on the temperature. When the beam electrons start to be scattered by the gas molecules, the fraction of scattered electrons is removed from the focused beam and hit the specimen somewhere in a large area around the point of incidence of the focused beam. The scattered electrons form a “skirt” around the focused beam, which has a radius of $100\text{ }\mu\text{m}$ for a pathlength of 5 mm (conditions: $E_0 = 10\text{ keV}$, water vapor pressure = 10^{-3} Pa) (Danilatos, 1988). Using a phosphor imaging plate, the distribution of unscat-

tered beam electrons and the scattered “skirt” electrons was directly imaged by exposure to the electron beam for a specified time (Wight and Zeissler, 2000). Related to the electron beam intensity within $25\text{ }\mu\text{m}$, the “skirt” intensity as a function of the distance from the center drops to 15% at $100\text{ }\mu\text{m}$, 5% at $200\text{ }\mu\text{m}$, and 1% at $500\text{ }\mu\text{m}$ (conditions: $E_0 = 20\text{ keV}$, water vapor pressure = 266 Pa , $l = 10\text{ mm}$) (Wight and Zeissler, 2000). The signals generated by the electrons of the skirt originate from a large area, which contributes to the background, whereas the unscattered beam remains focused to a small spot on the specimen surface, although its intensity is reduced by the fraction of electrons removed by scattering. The resolution obtainable depends on the beam diameter and the size of the interaction volume in the specimen, which is analogous to the situation in conventional and high-resolution SEM, i.e., the resolving power of ESEM can be maintained in the presence of gas.

The detection of BSE, CL, and X-rays is to a great extent analogous to the detection in a conventional SEM, because these signals can penetrate the gas sufficiently (Danilatos, 1985, 1986). However, the situation is completely different for the detection of SE. The conventional Everhart–Thornley detector would break down at elevated pressure in the specimen chamber. However, the gas itself can be used as an amplifier in a fashion similar to that used in ionization chambers and gas proportional counters. An attractive positive voltage on a detector will make all the secondary electrons drift toward it. If the attractive field is sufficiently large, each drifting electron will be accelerated, thus gaining enough energy to cause ionization of gas molecules, which can create more than one electron. This process repeating itself results in a significant avalanche amplification of the secondary electron current, which arrives at the central electrode of the environmental secondary electron detector (ESD) (Danilatos, 1988). The avalanche amplification works best only in a limited pressure range and can amplify the SE signal up to three orders of magnitude (Thiel et al., 1997). Too high pressure in the specimen chamber makes the mean free path of the electrons very small and a high electric field between specimen and detector is required to accelerate them sufficiently. Too low pressure in the chamber results in a large mean free electron path, i.e., only a few ionization events take place along the electron path from the specimen to the detector, thus the avalanche amplification factor is low. The new generation of ESD, the gaseous secondary electron detector (GSED), which consists of a 3-mm-diameter metallic ring placed above the specimen, provides better discrimination against parasitic electron signals. Both the ESD and GSED are patented and are available only in the ESEM.

However, the ionization of gas molecules creates not only electrons but also ions and gaseous scintillation. The latter can be used to make images (Danilatos, 1986), i.e., in that case the imaging gas acts as a detector. This principle is used in the patented variable pressure secondary electron (VPSE) detector. Nonconductive samples attract positive gas ions to their surface as negative charge accumulates from the electron beam, thus effectively suppressing or at least strongly reducing charging artifacts (Cazaux, 2004; Ji et al., 2005; Tang and Joy, 2003; Thiel et al., 2004; Robertson et al., 2004). The gas ions can affect or even reverse the contrast in the GSED image under specific conditions, e.g.,

at specimen regions of enhanced electron emission, where the rate of electron-ion pairs increases (Thiel et al., 1997). The highly mobile electrons generated by electron-gas interaction are removed from the gas by rapid sweeping to the GSED, which in turn causes an increased concentration of positive ions during image acquisition due to different electric field-induced drift velocities of negative and positive charge carriers in the imaging gas (Toth and Phillips, 2000).

However, imaging of wet, soft specimens can be hampered by the effect of surface tension (Kellenberger and Kistler, 1979), which may flatten and hereby deform the specimen. Obviously, this is a misleading situation demonstrating that “environmental conditions” do not necessarily guarantee structural preservation.

As mentioned above, about 612 Pa is the crucial minimum pressure for wet specimens. In addition to the ESEM, which enables imaging with SE at pressures up to about 6500 Pa, numerous variable pressure SEM (VPSEM), high pressure SEM, and low vacuum SEM (sometimes the abbreviation LVSEM is used, which cannot be distinguished from the low-voltage SEM) became commercially available. The water vapor pressure in the specimen chamber of those SEM is typically at maximum 300 Pa, i.e., below the crucial value of 612 Pa, which is not sufficient for imaging of wet specimens at stationary conditions. To separate the specimen pressure of maximum 300 Pa from the high vacuum in the column only one pressure-limiting aperture is sufficient. For imaging at pressures in the range from 250 to 300 Pa backscattered electrons are utilized.

Very recently, Thiberge et al. (2004) demonstrated scanning electron microscopy of cells and tissues under fully hydrated atmospheric conditions using a small chamber with a polyimide membrane (145 nm in thickness) that is transparent to beam and backscattered electrons. The membrane protects the fully hydrated sample from the vacuum. BSE imaging at acceleration voltages in the range of 12–30 kV revealed structures inside cultured cells and colloidal gold particles having diameters of 20 and 40 nm, respectively. Another interesting experimental setup is the habitat chamber designed to keep living cells under fully hydrated atmospheric conditions as long as possible and to reduce the exposure time to the lower pressure in the ESEM below 2 min (Cismak et al., 2003).

Scanning electron microscopy at elevated pressure is increasingly used in very different fields. Apart from variations in the pressure and chamber gas a heating stage (maximum temperature about 1500°C) allows changes in the specimen temperature. For example, chemical reactions such as corrosion of metals, electrolyte-solid interactions, alloy formation, and the degradation of the space shuttle ceramic shields by increasing oxygen partial pressures at high temperatures are possible with micrometer resolution. The onset of chemical reactions that depend on various parameters can be studied in detail. Insulators, including oil and oily specimens, can be directly imaged. Water can also be imaged directly in the ESEM, which allows studies of wetting and drying surfaces (e.g., de la Parra, 1993; Stelmashenko et al., 2001; Liukkonen, 1997) and direct visualization of the dynamic behavior of a water meniscus (Schenk et al., 1998; Rossi et al., 2004).

Figure 3–61 shows an example of dynamic studies of a water meniscus between the scanning tunneling tip and a support when the tip is

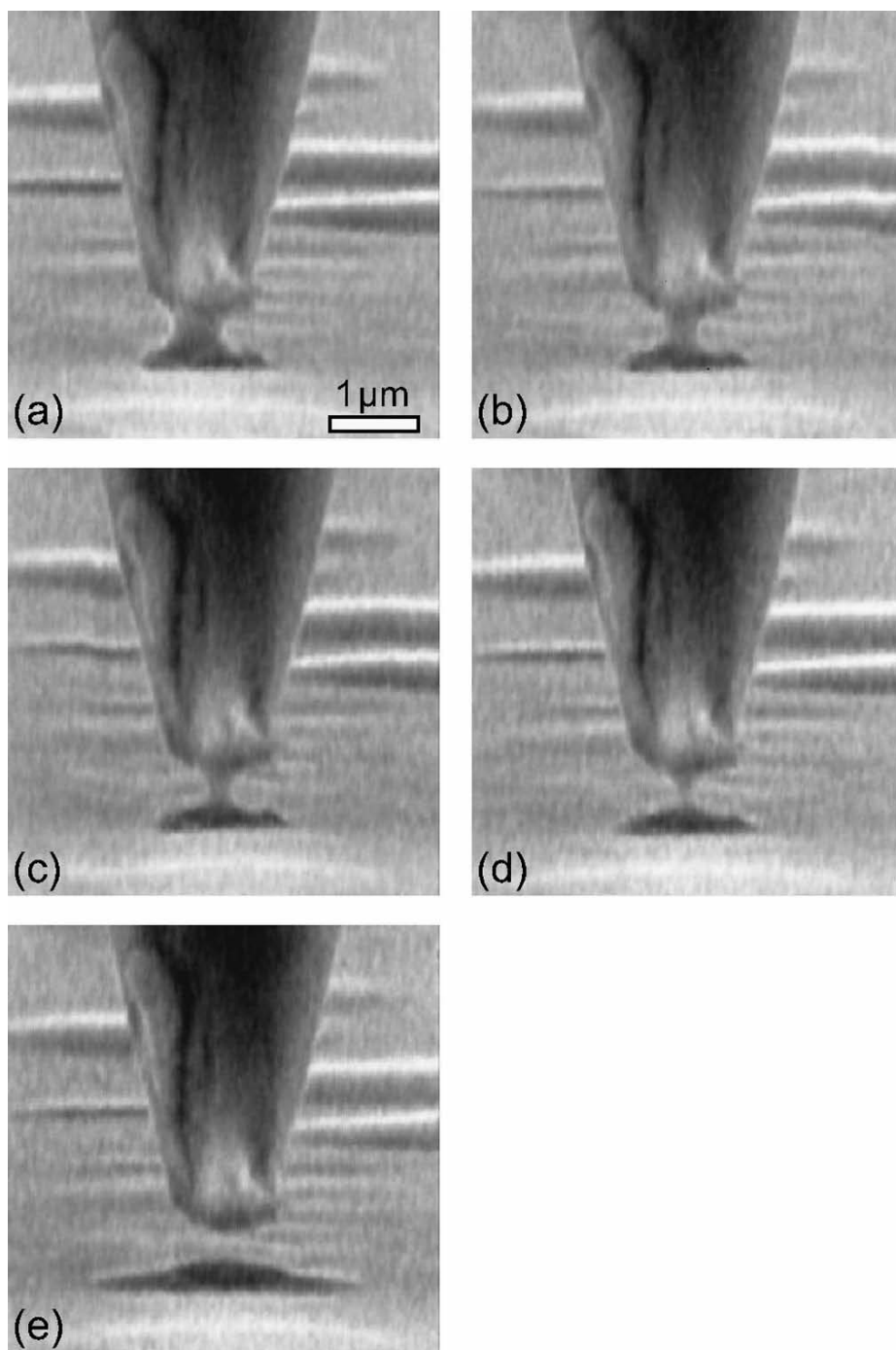


Figure 3-61. Time-resolved sequence of secondary electron images recorded with an ESEM®-E3 (ElectroScan Corp., Wilmington, MA). The water meniscus between the hydrophilic tungsten tip (normal electron beam incidence) and the Pt/C-coated mica (incidence angle of 85°) is clearly visible (a–d). Due to locally decreasing relative humidity the meniscus becomes gradually smaller until it snaps off (e). The absence of the meniscus leads to a significant change of shape of the water bead below the tip [cf. (d) and (e)]. Some water drops are located on the sample in front of and behind the tip. The sequence was recorded within 11 s and each image was acquired within about 2 s. Experimental conditions: $E_0 = 30$ keV, $I_p = 200$ pA, $p_g = 1.2$ kPa. (From Schenk et al., 1998; with kind permission of the American Institute of Physics, Woodbury, NY.)

moved across the sample. The wetting of the tip indicates a hydrophilic surface, whereas Figure 3–62 clearly indicates a hydrophobic tip surface.

ESEM studies of the wettability alteration due to aging in crude oil/brine/rock systems that are initially water wet are of significant importance in the petroleum industry in understanding the water condensation behavior on freshly exposed core chips. Surface active compounds are rapidly removed from the migrating petroleum, thus changing the wettability and subsequently allowing larger hydrophobic molecules to sorb (Bennett et al., 2004; Kowalewski et al., 2003; Robin, 2001). Furthermore, the ESEM is a powerful tool with which study the influence of salt, alcohol, and alkali on the interfacial activity of novel polymeric surfactants that exhibit excellent surface activity due to their unique structure (Cao and Li, 2002).

Environmental scanning electron microscopy disseminates rapidly among scientific and engineering disciplines. Applications range widely over diverse technologies such as pharmaceutical formulations, personal care and household products, paper fibers and coatings, cement-based materials, boron particle combustion, hydrogen sulfide corrosion of Ni–Fe, micromechanical fabrication, stone preservation, and biodeterioration. In spite of the broad applications, numerous contrast phenomena are not fully understood as yet. This is illustrated in Figure 3–63 by a series of SE micrographs recorded at different electron energies, but otherwise identical conditions. In addition, ESEM investigations of polymeric and biological specimens, which are known from conventional electron microscopy to be highly irradiation sensitive, are more difficult because water acts as a source of small, highly mobile free radicals, which accelerate specimen degradation (Kitching and Donald, 1996; Royall et al., 2001).

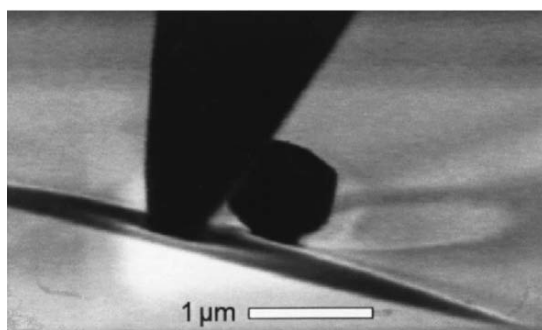


Figure 3–62. Secondary electron image recorded with an ESEM-E3 (ElectroScan Corp., Wilmington, MA) from a hydrophobic tungsten tip (normal electron beam incidence) and a water bead on Pt/C-coated mica (incidence angle of 85°). The shape of the deformed water surface in the submicrometer vicinity of the tip clearly indicates its hydrophobic surface. The spherical object (black) at the right of the tip in the back is probably a polystyrene sphere and any resemblance is purely coincidental. (From Schenk et al., 1998; with kind permission of the American Institute of Physics, Woodbury, NY.)

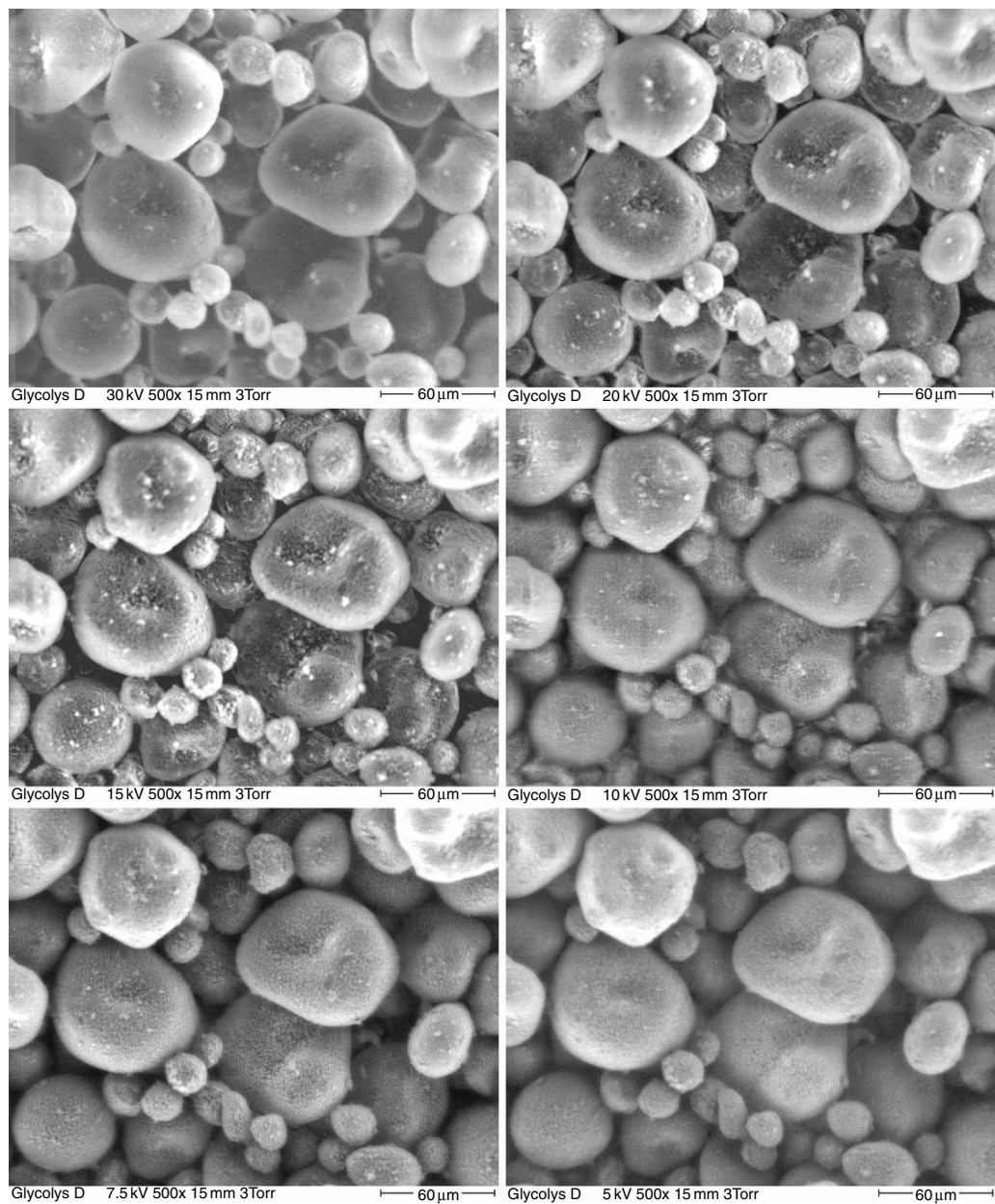


Figure 3–63. Secondary electron micrograph series of the starch glycolys D recorded at electron energies from 30 keV down to 5 keV (see the individual legends below each micrograph) with an ESEM demonstrating the effect of the electron energy on image contrast. (Micrographs kindly provided by Fraunhofer-Institut für Werkstoffmechanik, Halle, Germany; the project was supported by the State Sachsen-Anhalt, FKZ 3075A/0029B.)

5 Ultrahigh Vacuum Scanning Electron Microscopy in Surface Science

Ultrahigh vacuum (UHV) scanning electron microscopy is from the point of view of the pressure inside the specimen chamber the opposite of SEM at elevated pressure. Since the conventional SEMs typically work at high vacuum with a pressure of about 10^{-4} Pa inside the specimen chamber, the elevated pressure SEMs are operated at pressures six to seven orders of magnitude higher and the UHV SEMs about three to four orders of magnitude lower than 10^{-4} Pa. UHV is required for most surface science experiments for two principal reasons:

1. To obtain atomically clean surfaces for studies and to maintain such clean surfaces in a contamination-free state for the duration of the experiment.
2. To permit the use of a low-energy electron technique and, in addition to that, ion-based and scanning probe techniques without undue interference from gas phase scattering.

One crucial factor in determining how long it takes to build up a certain surface concentration of adsorbed species is the incident molecular flux F of gas molecules on surfaces given by the Hertz–Knudsen equation

$$F = p_g / \sqrt{(2\pi mkT)} \quad (5.1)$$

where m is the molecular mass, k is the Boltzmann constant, and T is the temperature (K). To obtain the minimum estimate of time it will take for a clean surface to become covered with a complete monolayer of adsorbate, a sticking probability $S = 1$ and a monolayer coverage typically in the order of 10^{15} – 10^{19} molecules/m² are assumed. Then the minimum time per monolayer simply equals $10^{19} \cdot \sqrt{(2\pi mkT)} / p_g$, i.e., the lower the pressure the longer the time for coverage. At a pressure of about 10^{-4} Pa forming a monolayer takes about 1 s.

Clean surfaces and UHV are required to apply very surface-sensitive methods such as Auger electron spectroscopy (AES) and spectromicroscopy, X-ray- and UV-induced photoelectron spectroscopies, and LVSEM and VLVSEM. SEM in UHV can be done with customized FESEMs equipped with an energy analyzer for AES, chemical state analysis, or trace element detection, respectively. These UHV-SEMs enable highly sensitive surface area observation using high-resolution SE imaging and are very well suited for EBSD observation (see Section 7) of electron channeling patterns without contamination as well as low accelerating voltage EDX analysis (see Section 6). An attached neutralizing ion gun allows analysis of insulating bulk materials and thin films without electrical charge ups.

SEM in UHV can also be performed with commercial surface science UHV chamber systems and a UHV SEM column attached to it. Such chamber systems offer an *in situ* combination of several imaging methods (e.g., LVSEM, SAM, AFM, and STM), methods for chemical analysis such as AES, and diffraction methods such as low-energy

electron diffraction (LEED). Dual-chamber systems have a large chamber for the imaging and analysis techniques mentioned above and a second chamber for specimen preparation such as ion sputtering, heating, and thin film growth.

SEM in UHV combined with other imaging, spectroscopic, and diffraction methods enables the true surface structure and chemistry to be studied. In particular, LVSEM and VLVSEM are most powerful in UHV because of their high surface sensitivity, which is not noticeably degraded by irradiation-induced contamination.

6 Microanalysis in Scanning Electron Microscopy

The generation of X-rays due to electron-specimen interactions was discussed in Section 2.2.5. The characteristic X-rays emitted from the specimen carry information about its local element composition, which is utilized as a powerful microanalytical tool combining SEM with EDX and WDX spectrometers. X-Ray microanalysis is by far the most widely used method combined with SEM, which enables in various modes qualitative and quantitative element analysis from a point or area of interest as well as mapping of the distribution of various elements simultaneously with SE and BSE imaging. The size of the interaction volume emitting X-rays is significantly larger than the ones for AE, SE, and BSE because of the weaker absorption of X-rays inside the specimen (cf. Figure 3–14) and the secondary emission by X-ray fluorescence outside the electron interaction volume. The secondary X-ray emission volume is much larger than that for primary X-ray emission since X-rays are more penetrating than electrons having the same energy. For electron-excited X-ray spectrometry performed on thick specimens in the SEM, the range R for X-ray excitation is given according to Kanaya and Kayama (1972) by

$$R [\mu\text{m}] = (0.0276 A/\rho Z^{0.89}) \cdot (E_0^{1.67} - E_x^{1.67}) \quad (6.1)$$

where A is the atomic weight (g/mol), ρ is the density (g/cm³), Z is the atomic number, E_0 is the incident electron energy (keV), and E_x is the X-ray energy (keV).

For reliable quantitative X-ray microanalytical studies, X-ray absorption, X-ray fluorescence, and the fraction of backscattered electrons, all of which depend on the composition of the specimen, have to be taken into account and are performed by the so-called ZAF correction. Z stands for atomic number, which affects the penetration of incident electrons into the material, A for absorption of X-rays in the specimen on the path to the detector, and F for fluorescence caused by other X-rays generated in the specimen.

Two different types of detectors are used to measure the emitted X-ray intensity as a function of the energy or wavelength. In an EDX system (Figure 3–64) the X-rays enter the solid-state semiconductor detector and create electron hole pairs that cause a pulse of current to flow through the detector circuit. The number of pairs produced by each X-ray photon is proportional to its energy (see Section 2.1.3.1). In a WDX spectrometer (Figure 3–65) the X-rays fall on a bent crystal and

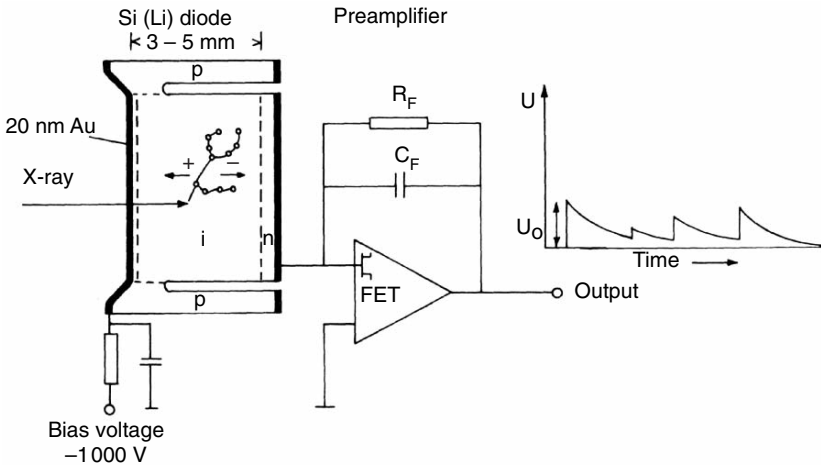


Figure 3-64. Scheme of Si(Li) X-ray diode coupled to a field effect transistor (FET) with a resistive feedback loop (R_F , C_F). The shape of the output signal is shown in the output voltage vs. time diagram. Typically, this principle is used in energy-dispersive X-ray (EDX) detectors. (From Reimer, 1985; with kind permission of Springer-Verlag GmbH, Heidelberg, Germany.)

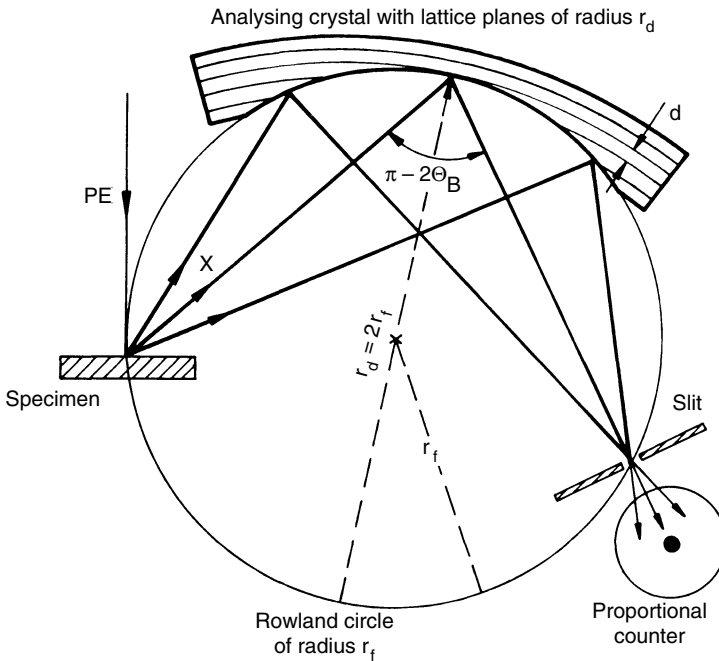


Figure 3-65. Principle of a wavelength-dispersive X-ray (WDX) spectrometer. Generated X-rays that hit the analysing crystal are focused and due to Bragg reflection directed to a slit in front of the proportional counter lying on a Rowland circle with radius r_f . The lattice planes of the crystal are bent to a radius of $2r_f$. (From Reimer, 1985; with kind permission of Springer-Verlag GmbH, Heidelberg, Germany.)

are reflected only if they satisfy Bragg's law. The crystal bending is such that it focuses X-rays of one specific wavelength onto a proportional counter and rotates to scan the wavelength detected. Some important features of both types of X-ray spectrometer are listed in Table 3–9. However, since instrumentation and analysis of data in X-ray microanalysis are usually considered a separate discipline, no further details are discussed in this section [see, e.g., Heinrich (1982), Heinrich and Newbury (1991), Reimer (1998), Goldstein et al. (1984, 2003), and Newbury and Bright (2005)].

A selected application of the very powerful combination of SEM imaging, X-ray microanalysis, and element mapping—the latter was invented almost exactly 50 years ago by Cosslett and Duncumb (1956)—is illustrated in Figure 3–66. The selected specimen is a Cr–Fe alloy with an Si phase, which has a locally varying composition as clearly indicated by the energy-dispersive spectra in Figure 3–66a and b recorded at different locations (Figure 3–66c).

The area under each characteristic peak represents the amount of X-ray counts, which is—after subtraction of the unspecific background below the peak and ZAF correction—a direct quantitative measure of the number of atoms of the specific element belonging to that peak. However, a simple visual inspection of the spectra shows, e.g., that the location “Punkt1” (see Figure 3–66a) contains significantly more chromium and less iron than location “Punkt2” (see the spectrum in Figure 3–66b). In addition, a strong silicon peak emerges in the spectrum of location “Punkt2” not present in the spectrum of “Punkt1” (see the spectrum in Figure 3–66a). The element distribution maps of four important chemical elements in the specimen, namely iron, chromium, silicon, and titanium, are shown in Figure 3–66d–g. Comparing the information given by the four element distribution maps on the one hand and the two spectra on the other hand immediately makes clear why the titanium peak does not emerge in the spectra and the chromium peak is dominant at location “Punkt1” but not at “Punkt2”. By means of simple image processing procedures the SE micrograph (Figure 3–66c) and the element distribution maps (Figure 3–66d–g) can be superimposed in one image (Figure 3–66h) presenting information for five individual images.

The most powerful tool in electron beam microanalysis is the ability to depict the elemental compositional heterogeneity of matter with micrometer to nanometer lateral resolution. Many developments have occurred since the intervening years to advance this critical method. We are now on the leading edge of extraordinary new X-ray mapping performance: the emergence of the silicon drift detector (SDD) that permits recording of X-ray spectrum images (XSI) in an energy-dispersive operating mode with output count rates of 600 kHz and even higher (cf. Table 3–9). Further, computer-controlled SEM in connection with image processing and EDX spectrometry enables the unattended and automated determination of both the geometric parameters and the chemical composition of thousands of individual particles down to a size of 50–100 nm (see, e.g., Poelt et al., 2002). Consequently, correlations between particle size, chemical composition, the number of

Table 3–9. Characteristic features of energy-dispersive X-ray (EDX) and wavelength-dispersive X-ray (WDX) detectors.^a

Features	Detector				
	Energy dispersive Si(Li) X-ray detector ^b	Energy dispersive HPGe X-ray detector ^b	Energy dispersive Si drift (SD) X-ray detector ^{c,d,e}	Energy dispersive microcalorimeter X-ray detector ^f	Wavelength dispersive X-ray detector ^b
Geometric collection efficiency	<2%	<2%	<2%	$\approx 8 \times 10^{-3}\%$	<0.2%
Quantum efficiency	$\approx 100\%$	$\approx 100\%$	>90%	60–80% (within 1–2 keV)	$\leq 30\%$
Element detection	$Z \geq 11$ (Be window) $Z \geq 4$ (windowless)	$Z \geq 4$	$Z \geq 5$	$Z \geq 6^g$	$Z \geq 4$
Energy resolution	150 eV (at 5.9 keV)	133 eV (at 10^4 cps) 115 eV (at 10^3 cps)	150 eV (at 10^5 cps) 230 eV (at 6×10^5 cps)	12 eV (at 5.9 keV) 3.1 eV (at 1.49 keV)	≈ 5 eV
Maximum counting rate	3×10^3 cps	3×10^4 cps	$\approx 6 \times 10^5$ cps	8×10^2 cps– 10^3 cps	10^5 cps
Spectrum acquisition	All energies simultaneously	All energies simultaneously	All energies simultaneously	All energies simultaneously	One wavelength at a time
Probe current (nA)	0.1–20	0.1–10	0.1–10	0.1–10	1–100

^a Si(Li), Lithium drifted silicon; HPGe, high-purity germanium; cps: counts per second.

^b Reichelt (1995).

^c Strüder et al. (1998a).

^d Strüder et al. (1998b).

^e Lechner et al. (2001).

^f Newbury et al. (1999).

^g With BN/Moxtek window, thermal isolation, and Au-absorber.

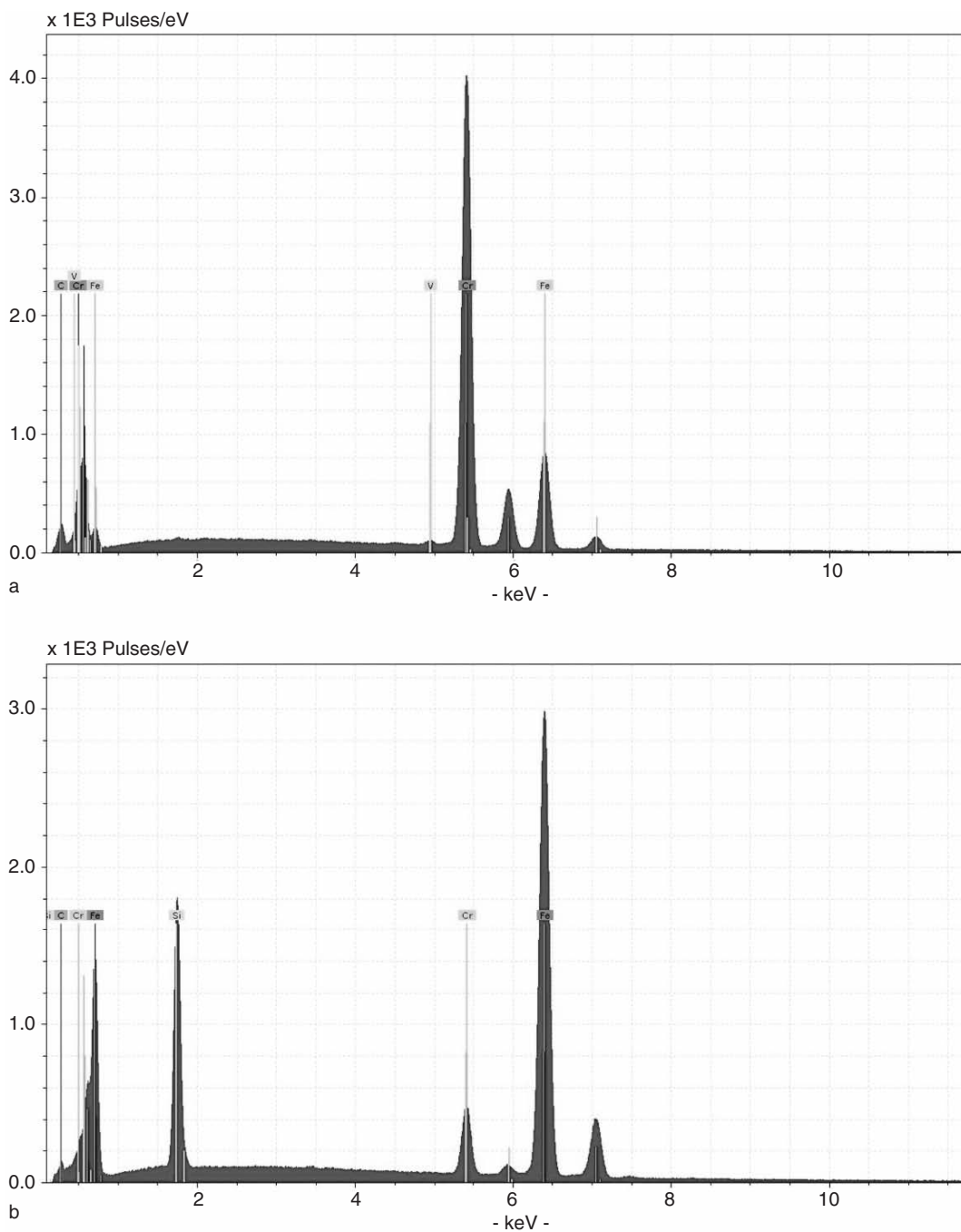


Figure 3–66. X-Ray microanalysis of a Cr-Fe-alloy with a Si phase. The EDX spectra (a and b) were recorded with the Röntec XFlah3001 from locations “Punkt1” and “Punkt2” marked in the SE micrograph of the specimen (c). The positions of the characteristic X-ray energies for the various elements emerging in the spectra are indicated by thin colored lines, which are labeled with the chemical symbol of the corresponding chemical element. The elements iron, chromium, and vanadium occur with one K_{α} peak each in the energy range from 4.95 to 6.40 keV and with one less intense L_{α} peak each in the energy range from 0.51 to 0.71 keV. Elemental distribution maps of Fe (d), Cr (e), Si (f), and Ti (g) were recorded using the K_{α} lines. (h) Mixed micrograph obtained by superimposition of the SE image and the maps of the distribution of Fe, Cr, Si, and Ti within the field of view. Experimental conditions: SEM, LEO 438VP. For recording spectra: $E_0 = 20$ keV; count rate, $\approx 3 \times 10^3$ cps; acquisition time, 300 s. For recording maps: $E_0 = 25$ keV; count rate $\approx 1.5 \times 10^5$ cps; acquisition time 600 s. (EDX spectra, SE micrograph, and elemental distribution maps were kindly provided by Röntec GmbH, Berlin, Germany.) (See color plate.)

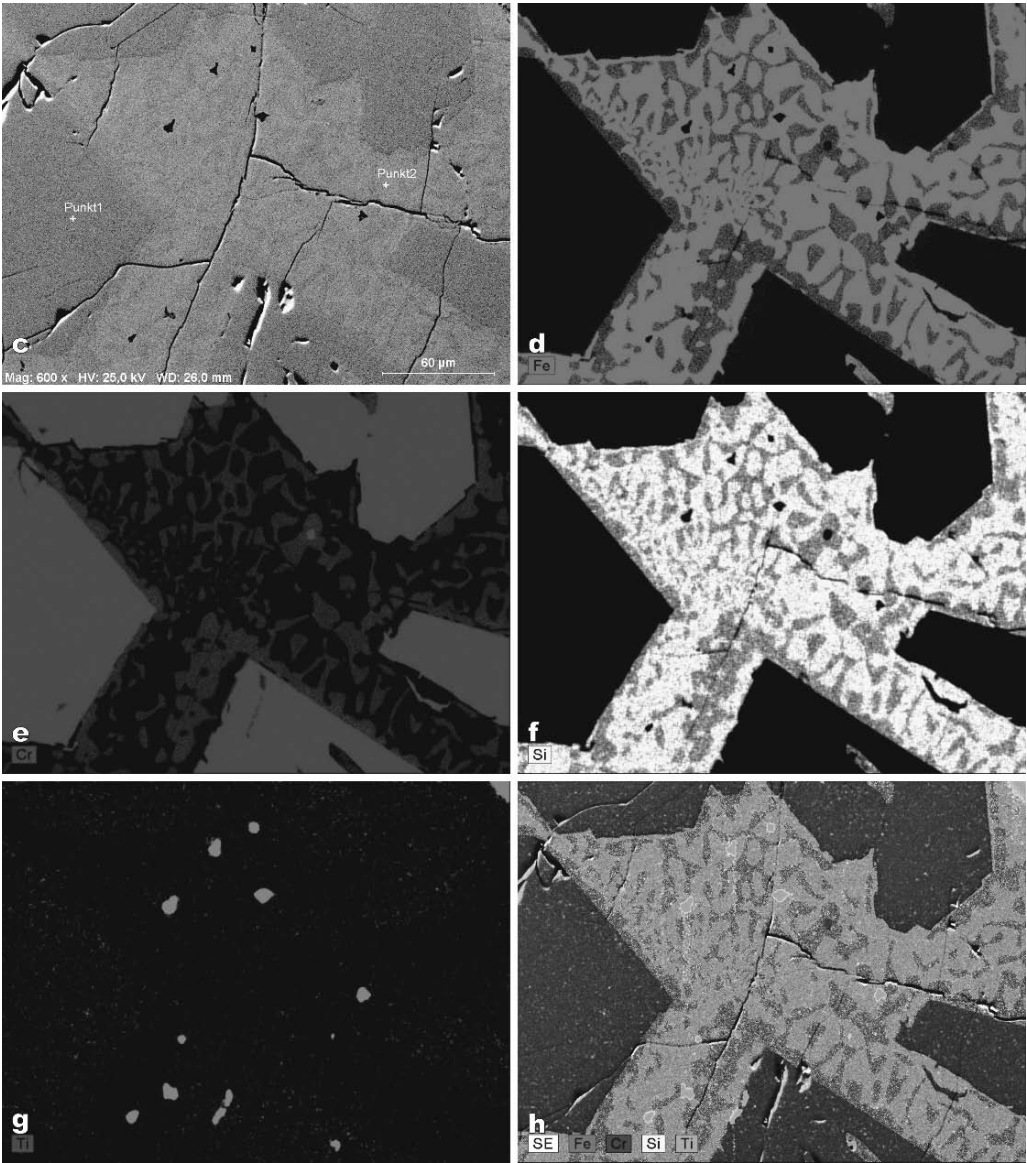


Figure 3–66. *Continued*

different compounds, and their contribution to the overall concentration can be established. Problems may arise in connection with specimen preparation, optimization of the image contrast, the sometimes nonhomogeneous composition of particles, shadowing of the X-rays in case of large particles, and the lack of a rigorous ZAF correction procedure for particles of arbitrary shape.

Another increasingly important application is multilayer analysis, i.e., the nondestructive measurement of the thickness and composition of thin films both unsupported and on substrates, which can be performed with high accuracy down to a thickness of 2 nm by a combination SEM and EDX (see, e.g., Hu and Pan, 2001; Rickerby and Thiot, 1994; Rickerby et al., 1998; Smith et al., 1995). For this purpose the ratio between the X-ray intensity of the film and the intensity of the same element of a bulk standard is used.

Low-voltage scanning electron microscopy ($E_0 < 5 \text{ keV}$) offers both high spatial resolution and a significantly reduced X-ray generation depth. This enables the composition of thin layers on substrates to be determined without the need to use a dedicated thin film analysis program. The analysis of small phases down to a size of 50 nm is also possible (see, e.g., Poelt, 2000; Wurster, 1997). However, the disadvantages are that the X-ray intensity produced is small because of low fluorescence yield ω (cf. Figure 3–22) and not all elements can be analyzed. In principle elements of higher atomic number can be identified using L- and M-shell X-rays, but these are much more complicated than the rather simple K-shell X-ray emissions. Although the enhanced surface sensitivity may be useful sometimes, in many cases it is more likely to be a problem. If thin film coating is required this film may be a significant fraction of the electron range, and without it the specimen may charge up. Working at $E_0 = E_2$, where incoming and emitted electrons are balanced, is usually not sufficient. Furthermore, having the electron beam stationary on the specimen for a long time to integrate the weak signals is exactly the way to produce contamination, which can stop the electrons before they reach the real specimen surface. Then X-ray microanalysis does not relate to the specimen at all.

7 Crystal Structure Analysis by Electron Backscatter Diffraction

In crystalline specimens electrons are diffracted at lattice planes according to Bragg's law given as

$$2d \sin \vartheta = n\lambda \quad (7.1)$$

where d is the lattice-plane spacing, ϑ is Bragg's angle, and λ is the electron wavelength. Bragg's law requires that the incident and emergent angles should be equal ϑ . As previously mentioned, the backscattering coefficient sensitively depends on the tilt of the incident electron beam relative to the lattice and Bragg position. Changing the tilt of the

incident beam relative to the lattice, e.g., by rocking the electron beam or tilting the specimen, affects the backscattering coefficient, which results in an electron channeling pattern (ECP). To obtain from an ECP information about the crystal structure, e.g., the crystal orientation and the lattice plane spacings, a so-called panorama diagram recorded by successively tilting the specimen over a large angular range is required (Joy et al., 1982; van Essen et al., 1971; Reimer, 1998). However, establishing a panorama diagram is a somewhat difficult task.

Another way of obtaining information about the crystal structure of the specimen is the use of electron diffraction effects associated with the scattered electrons. As described in Section 2.2, the beam electrons are scattered elastically and inelastically due to the electron-specimen interaction, thus scattered beam electrons travel within the excitation volume in all directions (cf. Figure 3–13). This scattering process can be considered as a small electron source inside the crystalline specimen emitting electrons in all directions as shown in Figure 3–67. These electrons may be diffracted at sets of parallel lattice planes according

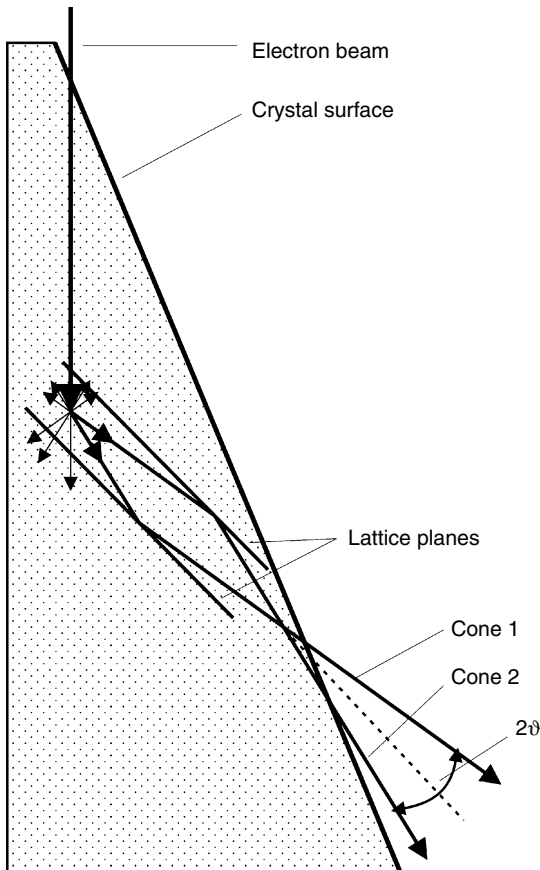


Figure 3–67. Scheme of the formation of one pair of cones from diffraction of scattered electrons at one set of parallel lattice planes.

to Bragg's law [(Eq. (7.1)] as first observed by Venables and Harland (1973). Figure 3–67 illustrates that the electrons emitted from one point and diffracted at lattice planes will form pairs of cones centered with respect to the normal vector of these lattice planes. The opening angle of the cones is $180^\circ - 2\vartheta$ and the angle between them is 2ϑ . Since the Bragg angle is in the order of 1° for 15–30 keV electrons and lattice-plane spacings of 0.2–0.3 nm, the intersections of the cones with a flat observation screen positioned at a distance much larger than d (usually 2–4 cm from the point of beam–specimen intersection) and tangential to the propagation sphere of the scattered electrons are almost straight and parallel pairs of lines. They are called Kikuchi lines. The angle between parallel pairs of Kikuchi lines is 2ϑ . The whole EBSD pattern with the Kikuchi lines reveals local information about the crystalline structure within the individual excitation volume. A detailed treatment of EBSD, which is beyond the scope of this section, is given by Wilkinson and Hirsch (1997). Nevertheless, it is worth mentioning that the relationship between EBSD and the previously introduced ECP is characterized by the reciprocity of their ray diagrams (Reimer, 1985).

The recording of the EBSD pattern, which originally was performed by exposure of a photographic film, is now done by position-sensitive detectors such as a phosphorous screen coupled to a TV camera or a scintillation window with a CCD camera attached. To allow the diffracted electrons to escape from the specimen its surface is usually tilted approximately 60° or 70° toward the screen. The EBSD patterns are digitally acquired by computer, which also controls the positioning of the beam. In case of automated crystal orientation mapping, the computer scans the electron beam stepwise across the specimen and controls the dynamic focusing of the beam (cf. Figure 3–9e and Section 2.1.5.2). Since special software algorithms for the automated detection and indexing of Kikuchi lines in EBSD pattern were introduced by Krieger-Lassen et al. (1992), Adams et al. (1993) developed a new scanning technique called “orientation imaging microscopy” (OIM). In OIM, which is also called automated crystal orientation mapping, the EBSD pattern from each individual point at the specimen surface radiated by the electron beam is recorded and analyzed. The current state of the art of OIM is reviewed by Schwarzer (1997) and Schwartz et al. (2000).

In the past decade EBSD has become a powerful tool for crystallographic analysis such as the determination of the orientation of individual crystallites of polycrystalline materials in the SEM, phase identification, and characterization of grain boundaries, which is illustrated by the following examples. Figure 3–68 shows two different EBSD patterns from an as-cast niobium sample recorded with high (Figure 3–68a) and low resolution (Figure 3–68b), respectively (Zaefferer, 2004), at a stationary beam position. Figure 3–68c shows the colored pairs of Kikuchi lines generated by automatic indexing and overlaid to the EBSD monitored in Figure 3–68a.

Figure 3–69a shows an SE micrograph of polycrystalline austenite with the related color-coded orientation map of polycrystalline austenite monitored in Figure 3–69b (Zaefferer et al., 2004). The spatial resolu-

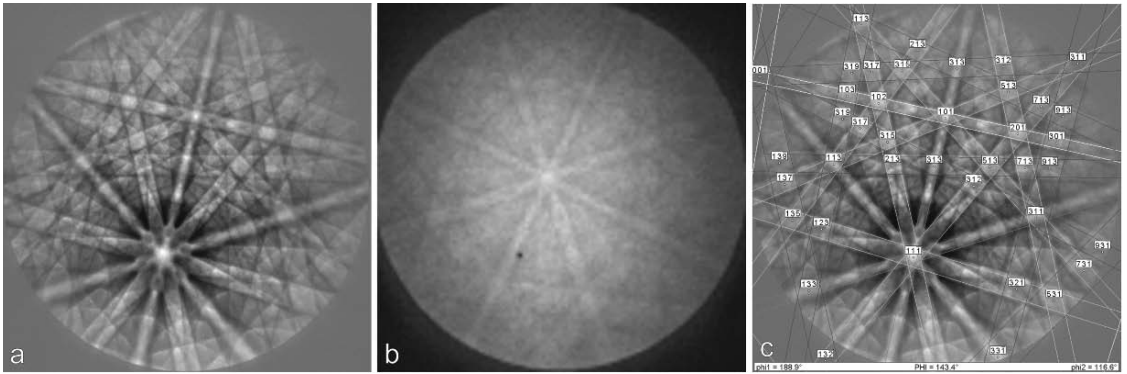


Figure 3–68. EBSD patterns from an as-cast niobium specimen. (a) High-resolution EBSD pattern with background subtraction. Exposure time, 7s. (b) Raw data EBSD pattern for high-speed mapping. Exposure time, 15 ms. EBSD patterns are recorded with thermal FESEM JSM-6500F. (c) EBSD pattern from (a) with colored pairs of Kikuchi lines generated by automatic indexing. [EBSD patterns were kindly provided by Dr. S. Zaefferer, Max-Planck-Institut für Eisenforschung, Düsseldorf, Germany. (a and b) From Zaefferer, 2004; with kind permission from JEOL (Germany) GmbH, München, Germany.] (For part c, see color plate.)

tion obtained amounts to about 50 nm (at $E_0 = 15$ keV), i.e., smaller grains or precipitates cannot be detected in the orientation map. The map shown in Figure 3–69c provides information about the grain boundary character and the orientation variations inside of each grain. The latter is shown for an individual grain in more detail in Figure 3–70.

The OIM developed into a powerful technique providing a wealth of information about the type and distribution of different phases, the size, shape, and defects of grains, the type of grain boundaries, the local crystal orientation, and the preferential orientation (texture). To take full advantage of this new imaging technique in terms of spatial and orientation accuracy the thermal field emission SEM providing a high beam current is most suitable. Furthermore, the SEM needs a very high mechanical and electronic stability because EBSD measurements require very long recording time, e.g., up to 12 h for very large orientation maps (Zaefferer, 2004). Finally, the preparation of specimens is a very delicate task since EBSD is a very surface-sensitive technique with an information depth of less than 10 nm up to a few tens of nanometers. Thus, the surface-near zone must be free of any deformation and, in addition, the surface has to be rather smooth because any surface relief may affect data acquisition.

The application of OIM is not restricted to metals and alloys. Crystalline materials such as semiconductors (e.g., Wilkinson, 2000), ceramics (e.g., Katrakova and Mucklich, 2002; Koblischka and Koblischka-Veneva, 2003), and minerals (e.g., Mauler et al., 2001; Prior et al., 1999) can be investigated.

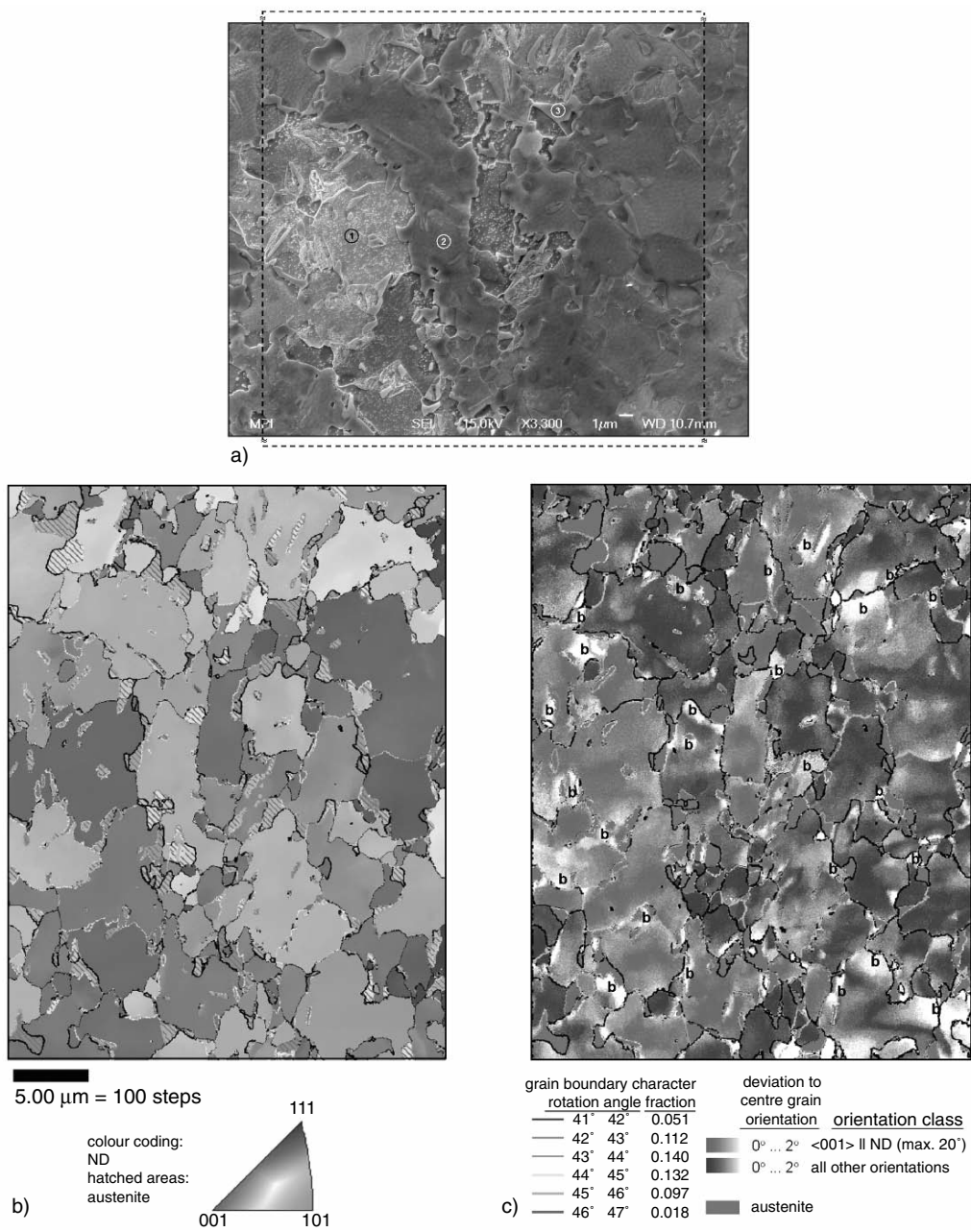


Figure 3–69. Secondary electron micrograph of austenite (a) with rough (1), relatively smooth (2), and smooth (3) surface areas. (b) Orientation map of (a) measured by automated crystal orientation mapping and color coded for the crystal direction parallel to the normal direction (ND) of the sheet. Hatched areas correspond to austenite grains. (c) The boundary character between γ - and α -grains, different orientation components [$\{001\}||ND$, red; all others blue] and the orientation variations inside of each grain (color shading; b, bainite). The micrograph and the maps are recorded with thermal FESEM JSM-6500F. Note: the extension toward the top and bottom of the measured area in (b) and (c) is larger than the area marked in (a). (Reprinted from Zaefferer et al., 2004; copyright 2004, with permission from Elsevier.) (For parts b and c, see color plate.)

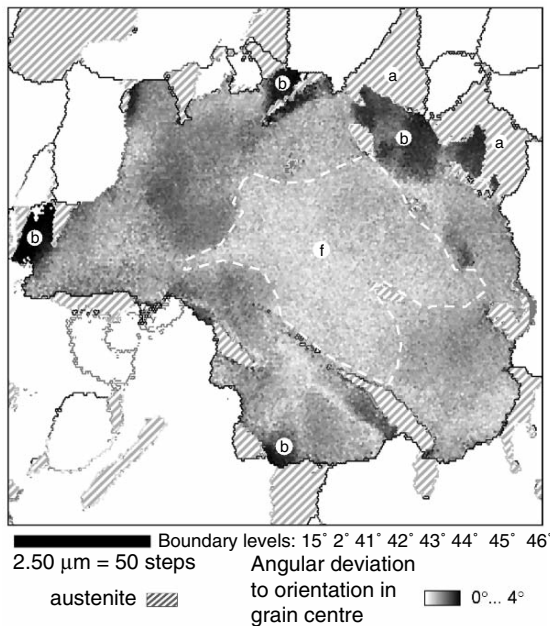


Figure 3–70. Orientation map of one grain from the microstructure in Figure 3–69a. Color code: angular deviation of every mapping point to one orientation in the center of the grain. Bainite appears in conjunction with a steep orientation gradient in ferrite. The white line marks the maximum extension of austenite at austenization temperature. f, ferrite; a, austenite (hatched); b, bainite. [Reprinted from Zaefferer et al., 2004; copyright, 2004, with permission from Elsevier.] (See color plate.)

Acknowledgments. I would like to acknowledge the very valuable help of Dipl.-Ing. Harald Nüsse [Institut für Medizinische Physik und Biophysik (IMPB)] in preparing numerous schematic drawings and figures composed of individual micrographs. Dr. Vladislav Krzyzanek (IMPB) kindly contributed data from Monte Carlo simulations of the electron scattering in thin and in bulky specimens. I am greatly indebted to Mrs. Ulrike Keller (IMPB) for expert technical assistance in the preparation and scanning electron microscopic investigation of numerous specimens. Finally, the expert photographic work of Mrs. Gudrun Kieffermann (IMPB) with those micrographs, which were recorded in the Institut für Medizinische Physik und Biophysik, is greatly acknowledged.

References

- Åberg, T. and Howat, G. (1982). In *Handbuch der Physik*, vol 31 (Mehlhorn, W. and Flügge, S., eds) 469–619. (Springer, Berlin).
- Adamec, P., Delong, A. and Lencova, B. (1995). *J Microsc* **179**, 129.
- Adams, B.L., Wright, S.I. and Kunze, K. (1993). *Metall Trans* **24A**, 819.
- Albrecht, R.M. and Ornberg, R.L. (eds.) (1988). *The Science of Biological Specimen Preparations for Microscopy and Microanalysis*. (Scanning Microscopy International, Chicago).

- Albrecht, R.M., Simmons, S.R., Prudent, J.R. and Erickson, C.M. (1988). In *46th Ann Meeting Electron Microscopy Soc America* (Bailey, G.W., ed.) 214–217. (San Francisco Press, San Francisco).
- Alexander, H. (1994). *Materials Sci Engin B: Solid State Mat Adv Technol* **24**, 1.
- Anglin, E.J., Schwartz, M.P., Ng, V.P., Perelman, L.A. and Sailor, M.J. (2004). *Langmuir* **20**, 11264.
- Aoyagi, S. (2002). *JEOL News* **37**, 70.
- Apkarian, R.P. and Curtis, J.C. (1986). *Scann Electr Microsc* **4**, 1381.
- Apkarian, R.P. and Joy, D.C. (1988). In *Microbeam Analysis* (Newbury, D.E., ed.) 459–462. (San Francisco Press, San Francisco).
- Arnal, F., Verdier, P. and Vincensini, P.-D. (1969). *Comp rend Acad Sci (Paris)* **268**, 1526.
- Autrata, R. (1990). *Scann* **12**, 119.
- Autrata, R. and Hejna, J. (1991). *Scann* **13**, 275.
- Autrata, R. and Schauer, P. (2004). In *13th Eur Congr Microscopy*, vol **1** (Schryvers, D., Timmermans, J.-P., van Dyck, D. and van Oostveldt, eds.) 75–76. (Belgian Society for Microscopy, Liège).
- Autrata, R., Hermann, R. and Müller, M. (1992a). *Scann* **14**, 127.
- Autrata, R., Jiráč, J., Spinka, J. and Hutar, O. (1992b). *Scann Microsc* **6**, 69.
- Bahadur, H. and Parshad, R. (1980). *Scann Electr Microsc* **1**, 509.
- Balk, L.J. (1986). *Can J Phys* **64**, 1238.
- Balk, L.J. (1989). In *SEM Characterization of Semiconductors* (Holt, D.B. and Joy, D.C., eds.) 425–445. (Academic Press, New York).
- Balk, L.J. and Kultscher, N. (1984). *J Phys (Paris)*, Colloque **C2**, C2-873.
- Balk, L.J., Davies, D.G. and Kultscher, N. (1984). *IEEE Transact Magnet* **20**, 1466.
- Bambynek, W., Crasemann, B., Fink, R.W., Freund, H.U., Mark, H., Swift, C.D., Price, R.E. and Rao, P.V. (1972). *Rev Mod Phys* **44**, 716.
- Barth, J.E. and Kruit, P. (1996). *Optik* **101**, 101.
- Baschong, W. and Wrigley, N.G. (1990). *J Electr Microsc Techn* **14**, 313.
- Bastacky, J., Wodley, C., Labrie, R. and Backhus, C. (1988). *Scann* **10**, 37.
- Bauer, E. and Teliéps, W. (1988). In *Surface and Interface Characterization by Electron Optical Methods* (Howie, A. and Valdre, A., eds.) 195–233. (Plenum, New York).
- Bauer, H.E. and Seiler, H. (1984). In *Scanning Electron Microscopy*, vol **3** (Johari, O., ed.) 1081–1088. (Scanning Electron Microscopy, Chicago).
- Baumeister, W., Hahn, M., Seredynski, J. and Herberzt, L.M. (1976). *Ultramicroscopy* **1**, 377.
- Baumeister, W., Karrenberg, F., Rachel, R., Engel, A., ten Heggeler, B. and Saxton, W.O. (1982). *Eur J Biochem* **125**, 535.
- Baun, W.L. (1969). *Adv Electr Phys Suppl.* **6**, 155.
- Bearden, J.A (1967a). *Rev Mod Phys* **39**, 78.
- Bearden, J.A. (1967b). *Rev Mod Phys* **39**, 125.
- Bennett, B., Buckman, J.O., Bowler, B.F. and Larter, S.R. (2004). *Petrol Geosci* **10**, 271.
- Berger, M.J. and Seltzer, S.M. (1964). In *Penetration of Charged Particles in Matter* (Fano, U., ed.) 205–268. National Academy of Sciences–National Research Council Publication 1133.
- Berry, V.K. (1988). *Scann* **10**, 19.
- Bethe, H. (1930). *Ann Phys* **5**, 325.
- Bishop, H.E. (1974). In *Quantitative Scanning Electron Microscopy* (Holt, D.B., Muir, M.D., Grant, P.R. and Boswarva, I.M., eds.) 41–64. (Academic Press, London).

- Bishop, H.E. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 259–269. (Scanning Electron Microscopy, Chicago).
- Bishop, H.E. and Riviere, J.C. (1969). *J Appl Phys* **40**, 1740.
- Bittermann, A.G., Jacobi, S., Chi, L.F., Fuchs, H. and Reichelt, R. (2001). *Langmuir* **17**, 1872.
- Bleloch, A.L., Castell, M.R., Howie, A. and Walsh, C.A. (1994). *Ultramicroscopy* **54**, 107.
- Bolt, R.O. and Carroll, J.G. (eds.) (1963). *Radiation Effects on Organic Materials*. (Academic Press, New York).
- Bond, E.F., Beresford, D. and Haggis, H.H. (1974). *J Microsc* **100**, 271.
- Böngeler, R., Golla, U., Kässens, M., Reimer, L., Schindler, B., Senkel, R. and Spranck, M. (1993). *Scann* **15**, 1.
- Boyde, A. (2004). *Scann* **26**, 265.
- Boyde, A. and Reid, S.A. (1983). *Scann Electr Microsc* **IV**, 1803.
- Brandis, E. and Rosencwaig, A. (1980). *Appl Phys Lett* **37**, 98.
- Bröcker, W., Krefting, E.-R. and Reimer, L. (1977). *Beitr Elektronenmikr Direktabb Analyse Oberfl* **10**, 647.
- Brown, G.M. and Butler, J.H. (1997). *Polymer* **38**, 3937.
- Burhop, E.H.S. (1952). *The Auger Effect*. (Cambridge University Press, Cambridge).
- Butler, J.H., Joy, D.C., Bradley, G.F. and Krause, S.J. (1995). *Polymer* **36**, 1781.
- Cao, Y. and Li, H.L. (2002). *Eur Polym J* **38**, 1457.
- Cargill, G.S. III (1980). *Nature* **286**, 691.
- Cazaux, J. (2004). *Microsc Microanal* **10**, 670.
- Chan, D.S.H., Phang, J.C.H., Chin, J.M. and Kolachina, S. (2000). *Beam Inject Assess Microstruct Semiconduct* **78–79**, 11.
- Chan, E.K.M., Darendeliler, M.A., Petocz, P. and Jones, A.S. (2004). *Eur J Oral Sci* **112**, 134.
- Chandler, J.A. (1978). *X-Ray Microanalysis in the Electron Microscope*. (North-Holland Publ., Amsterdam).
- Chang, T.H.P., Kern, D.P. and Muray, L.P. (1990). *J Vac Sci Technol B* **8**, 1698.
- Cismak, A., Schwanecke, M., Fütting, M. and Heilmann, A. (2003). *Microsc Microanal* **9** (Suppl 3), 480.
- Cosslett, V.E. (1972). *Optik* **36**, 85.
- Cosslett, V.E. and Duncumb, P. (1956). *Nature* **177**, 1172.
- Craven, A.J., Gibons, J.M., Howie, A. and Spalding, D.R. (1978). *Phil Mag* **38**, 519.
- Crewe, A.V., Isaacson, M. and Johnson, P. (1970). *Rev Sci Instrum* **41**, 20.
- Czyzewski, Z., O'Neill MacCallum, D., Romig, A. and Joy, D.C. (1990). *J Appl Phys* **68**, 3066.
- Danilatos, G.D. (1980). *Scann* **3**, 215.
- Danilatos, G.D. (1981). *J Microsc* **121**, 235.
- Danilatos, G.D. (1985). *Scann* **7**, 26.
- Danilatos, G.D. (1986). *Scann* **8**, 279.
- Danilatos, G.D. (1988). In *Advances in Electronics and Electron Physics* (Hawkes, P.W., ed.) 109–250. (Academic Press, London).
- Danilatos, G.D. (1990). In *Advances in Electronics and Electron Physics* (Hawkes, P.W., ed.) 1–102. (Academic Press, New York).
- Danilatos, G.D. (1991). *J Microsc* **162**, 391.
- Darlington, E.H. and Cosslett, V.E. (1972). *J Phys D* **5**, 1969.
- Davidson, S.M. (1989). In *SEM Microcharacterization of Semiconductors* (Holt, D.B. and Joy, D.C., eds.) 153–240. (Academic Press, New York).
- Deamy, H.J. (1982). *J Appl Phys* **53**, R51.

- de Harven, E., Leung, R. and Christensen, H. (1984). *J Cell Biol* **99**, 53.
- Dekker, A.J. (1958). *Solid State Phys* **6**, 251.
- de la Parra, R.E. (1993). *Microsc Res Techn* **25**, 362.
- DeMets, M. (1975). *Microsc Acta* **76**, 405.
- DeMets, M., Howlett, K.J. and Yoffe, A.O. (1974). *J Microsc* **102**, 125.
- DeVore, W. and Berger, S.D. (1996). *J Vac Sci Technol B* **14**, 3764.
- Dole, M. (ed.) (1973). *The Radiation Chemistry of Macromolecules*. (Academic Press, New York).
- Dorsey, J.R. (1969). *Adv Electr Phys Suppl* **6**, 291.
- Drescher, H., Reimer, L. and Seidel, H. (1970). *Zeit Angew Phys* **29**, 332.
- Drescher, H., Krefting, E.-R., Reimer, L. and Seidel, H. (1974). *Zeit Naturforsch* **29a**, 833.
- Drobne, D., Milani, M., Ballerini, M., Zrimec, A., Zrimec, M.B., Tatti, F. and Draslar, K. (2004). *J Biomed Opt* **9**, 1238.
- Drobne, D., Milani, M., Zrimec, A., Zrimec, M.B., Tatti, F. and Draslar, K. (2005). *Scann* **27**, 30.
- Drouin, D., Hovington, P. and Gauvin, R. (1997). *Scann* **19**, 20.
- Dubbeldam, L. (1991). In *Advances in Optical and Electron Microscopy*, vol **12** (Mulvey, T. and Sheppard, C.J.R., eds.) 139–242. (Academic Press, London).
- Dykstra, M.J. (1992). *Biological Electron Microscopy*. (Plenum, New York).
- Echlin, P. (1971). *Scann Electr Micros* **1**, 225.
- Echlin, P. (1992). *Low-Temperature Microscopy and Analysis*. (Plenum, New York).
- Edelmann, J. and Wetzig, K. (1995). In *In Situ Scanning Electron Microscopy in Materials Research* (Wetzig, K. and Schulze, D., eds.) 109–125. (Akademie Verlag, Berlin).
- Edelmann, L. and Roomans, G.M. (eds.) (1990). *The Science of Biological Specimen Preparations for Microscopy and Microanalysis*, vol **5**, Chicago: Scanning Microscopy International.
- Egerton, R.F. (1986). *Electron-Energy-Loss Spectroscopy in the Electron Microscope*. (Plenum, New York).
- Egerton, R.F. (1989). *Electron-Energy-Loss Spectroscopy in the Electron Microscope*. 2nd ed. (Plenum, New York).
- Egerton, R.F. (1999). *Ultramicroscopy* **80**, 247.
- Egerton, R.F., Crozier, P.A. and Rice, P. (1987). *Ultramicroscopy* **23**, 305.
- Egerton, R.F., Li, P. and Malac, M. (2004). *Micron* **35**, 399.
- Engel, A. (1978). *Ultramicroscopy* **3**, 273.
- Engel, A. (1983). In *Microscopie Électronique en Science des Matériaux Bombannes* (Jouffrey, B., Bourret, A. and Colliex, C., eds.) 185–192. (Éditions du CNRS, Toulouse).
- Erlandsen, S.L., Nelson, R.D., Hasslen, S.R., Dunney, G.M., Olmsted, S.B., Frethem, C. and Wells, C.L. (1995). *Hitachi Instrum News* **27**, 10.
- Everhart, T.E. and Hoff, P.H. (1971). *J Appl Phys* **42**, 5837.
- Everhart, T.E. and Thornley, R.F.M. (1960). *J Sci Instrum* **37**, 246.
- Everhart, T.E., Wells, O.C. and Oatley, C.W. (1959). *J Electron Control* **7**, 97.
- Fathers, D.J., Jacobovics, J.P., Joy, D.C., Newbury, D.E. and Yakowitz, H. (1973). *Phys Stat Sol (a)* **20**, 535.
- Feder, N. and Sidman, R.L. (1958). *J Biophys Biochem Cytol* **4**, 593.
- Feja, B., Dürrenberger, M., Müller, S., Reichelt, R. and Aebi, U. (1997). *J Struct Biol* **119**, 72.
- Flindt, R. (2000). *Biologie in Zahlen*, 5th ed. (Spektrum Akademischer Verlag, Berlin).
- Fourie, J.T. (1979). *Scann Electr Microsc* **2**, 87.
- Frank, J. (1980). In *Computer Processing of Electron Microscope Images* (Hawkes, P.W., ed.) 187–222. (Springer, New York).

- Frank, J., Bussler, P., Langer, R. and Hoppe, W. (1970). *Ber Bunsengesellsch Phys Chem* **74**, 1105.
- Frosien, J., Plies, E. and Anger, K. (1989). *J Vac Sci Techn B* **7**, 1874.
- Fruhstorfer, B., Mohles, V., Reichelt, R. and Nembach, E. (2002). *Phil Mag A* **82**, 2575.
- Fujioka, H. and Ura, K. (1981). *Appl Phys Lett* **39**, 81.
- Fujita, T., Tokunaga, M.D.J. and Inoue, H. (1971). *Atlas of Scanning Electron Microscopy in Medicine*. (Elsevier, Amsterdam).
- Galca, A.C., Kooij, E.S., Wormeester, H., Salm, C., Leca, V., Rector, J.H. and Poelsema, B. (2003). *J Appl Phys* **94**, 4296.
- Gamliel, H. and Polliack, A. (1983). *Scann Electr Microsc* **2**, 929.
- Georges, A., Fournier, J.M. and Bois, D. (1982). *Scann Electr Microsc* **1**, 147.
- Gerber, C., Binnig, G., Fuchs, H., Marti, O. and Rohrer, H. (1986). *Rev Sci Instrum* **57**, 221.
- Giannuzzi, L.A. and Stevie, F.A. (2005). *Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice*. (Springer, New York).
- Girard, P. (1991). *J Phys IV* **C6**, 259.
- Glaser, W. (1952). *Grundlagen der Elektronenoptik*. Wien: Springer.
- Glauert, A.M. (ed.) (1973). *Practical Methods in Electron Microscopy*. (North-Holland Publishing Comp., Amsterdam).
- Gnauck, P., Hoffrogge, P. and Greiser, J. (2003a). *Eur Microsc Anal* **82**, 7.
- Gnauck, P., Zeile, U., Rau, W. and Schuhmann, M. (2003b). *Microsc Microanal* **9** (Suppl 3), 524.
- Goldstein, J. and Yakowitz, H. (1975). *Practical Scanning Electron Microscopy*. (Plenum, New York).
- Goldstein, J.J., Newbury, D.E., Echlin, P., Joy, D.C., Fiori, C. and Lifshin, E. (1984). *Scanning Electron Microscopy and X-Ray Microanalysis. A Text for Biologists Material Scientists and Geologists*. (Plenum, New York).
- Goldstein, J.J., Newbury, D.E., Joy, D.C., Lyman, C.E., Echlin, P., Lifshin, E., Sawyer, L.C. and Michael, J.R. (2003). *Scanning Electron Microscopy and X-Ray Microanalysis*. (Kluwer Academic/Plenum, New York).
- González, G., Stracke, W., Lopez, Z., Keller, U., Ricker, A. and Reichelt, R. (2004). *Microsc Microanal* **10**, 224.
- Goudsmit, S.A. and Saunderson, J.L. (1940). *Phys Rev* **58**, 36.
- Grasenick, F., Aldrian, A., Bauer, R., Bangert, H., Essl, R., Haefer, R.H., Hagemann, P., Hermann, K.-H., Hörl, E.M., Karthaler, P., Knappek, E., Nobiling, R., Peters, K.-R. and Weber, G. (1991). *Elektronenmikroskopie. Erweiterte Einsatzmöglichkeiten und spezielle Abbildungs- und Präparationsmethoden*. Ehnlingen bei Böblingen: Expert Verlag.
- Griffith, G. (1993). *Fine Structure Immunocytochemistry*. (Springer, Berlin).
- Grivet, P. (1972). *Electron Optics*. (Pergamon Press, Oxford).
- Gross, H., Müller, T., Wildhaber, I. and Winkler, H. (1985). *Ultramicroscopy* **16**, 287.
- Grote, M., Reichelt, R. and Wiermann, R. (1999a) *Micron* **30**, 65.
- Grote, M., Hayek, B., Reichelt, R., Kraft, D. and Valenta, R. (1999b) *Int Arch Allergy Immunol* **120**, 287.
- Grote, M., Mahler, V., Spitzauer, S., Fuchs, T., Valenta, R. and Reichelt, R. (2000). *J Allergy Clin Immunol* **105**, 561.
- Grün, A.E. (1957). *Zeit Naturforsch* **12a**, 89.
- Hauffe, W. (1971). *Phys Stat Sol (a)* **4**, 111.
- Hauffe, W. (1990). *Beitr Elektronenmikr Direktabb Analyse Oberfl* **23**, 305.
- Hauffe, W. (1995). In *In Situ Scanning Electron Microscopy in Materials Research* (Wetzig, K. and Schulze, D., eds.) 195–218. (Akademie Verlag, Berlin).
- Hauffe, W., Pannicke, S., Däbritz, S. and Schade, P. (2002). *Surf Interface Anal* **34**, 768.

- Hayat, M.A. (ed.). (1974–1976). *Principles and Techniques of Scanning Electron Microscopy*. Vol. 1–6. (Van Nostrand Reinhold, New York).
- Hayat, M.A. (ed.). (1978). *Principles and Techniques of Biological Scanning Electron Microscopy*. (University Park Press, Baltimore).
- Hayat, M.A. (ed.). (1989/91). *Colloidal Gold. Principles, Methods, and Applications*, vol 1–3. (Academic Press, London).
- Hayat, M.A. (ed.) (2002). *Microscopy, Immunohistochemistry, and Antigen Retrieval Methods: For Light and Electron Microscopy*. (Kluwer Academic/Plenum, New York).
- Heiderhoff, R., Sergeev, O.V., Liu, Y.Y., Phang, J.C.H. and Balk, L.J. (2000). *J Cryst Growth* **210**, 303.
- Heinrich, K.F.J. (1966). In *Fourth International Congress on X-Ray Optics and Microanalysis* (Castaing, R., Deschamps, P. and Philibert, J., eds.) 159–167. (Hermann, Paris).
- Heinrich, K.F.J. (1982). *Microbeam Analysis*. (San Francisco Press, San Francisco).
- Heinrich, K.F.J. and Newbury, D.E. (eds.) (1991). *Electron Probe Quantification*. (Plenum, New York).
- Hejna, J. and Reimer, L. (1987). *Scann* **9**, 162.
- Hermann, R. and Müller, M. (1993). *Scann Microsc* **7**, 343.
- Hermann, R., Pawley, J., Nagatani, T. and Müller, M. (1988). *Scann Microsc* **2**, 1215.
- Hermann, R., Schwarz, H. and Müller, M. (1991). *J Struct Biol* **107**, 38.
- Hernandez-Lopez, J.L., Bauer, R.E., Chang, W.S., Glasser, G., Grebel-Koehler, D., Klapper, M., Kreiter, M., Leclair, J., Majoral, J.P., Mittler, S., Mullen, K., Vasilev, K., Weil, T., Wu, J., Zhu, T. and Knoll, W. (2003). *Mater Sci Eng C* **23**, 267.
- Hess, M.W. (2003). *J Microsc* **212**, 44.
- Hesse, D. and Meyer, K.-P. (1982). In *Elektronenmikroskopie in der Festkörperphysik* (Bethge, H. and Heydenreich, J. eds.) 473–507. (VEB Deutscher Verlag der Wissenschaften, Berlin).
- Heywood, V.H. (1971). *Scanning Electron Microscopy. Systematic and Evolutionary Applications*. (Academic Press, London).
- Hicks, D. and Molday, R.S. (1984). In *The Science of Biological Specimen Preparation for Microscopy and Microanalysis* (Johari, O., ed.) 203–220. (Scanning Electron Microscopy, Chicago).
- Hindermann, D.K. and Davis, R.H. (1974). *Scann Electr Microsc* **1**, 183.
- Hirsch, P., Kässens, M., Reimer, L., Senkel, R. and Spranck, M. (1993). *Ultramicroscopy* **50**, 263.
- Hitachi (2001). Technical Data. Sheet No. **98**, HTD-E050-01 2001.11.
- Holt, D.B. (1974). In *Quantitative Scanning Electron Microscopy* (Holt, D.B., Muir, M.D., Grant, P.R. and Boswarva, I.M., eds.) 213–286. (Academic Press, London).
- Holt, D.B. (1989). In *SEM Characterization of Semiconductors* (Holt, D.B. and Joy, D.C., eds.) 241–338. (Academic Press, London).
- Holt, D.B. and Saba, F.M. (1985). *Scann Electr Microsc* **3**, 1023.
- Holt, D.B. and Yacobi, B.G. (1989). In *SEM Microcharacterization of Semiconductors* (Holt, D.B. and Joy, D.C., eds.) vol **12**. London: Academic Press.
- Holt, D.B., Muir, M.D., Grant, P.R. and Boswarva, I.M. (eds.) (1974). *Quantitative Scanning Electron Microscopy*. (Academic Press, London).
- Holzer, L., Indutnyi, F., Gasser, P.H., Munch, B. and Wegmann, M. (2004). *J Microsc* **216**, 84.
- Hörl, E.M. (1972). *Micron* **3**, 540.
- Hörl, E.M. (1975). *Beitr Elektronenmikroskop Direktabb Analyse Oberfl (BEDO)* **8**, 369.

- Hosokawa, T., Fujioka, H. and Ura, K. (1978). *Rev Sci Instrum* **49**, 624.
- Hovington, P., Drouin, D. and Gauvin, R. (1997a). *Scann* **19**, 1.
- Hovington, P., Drouin, D., Gauvin, R., Joy, D.C. and Evans, N. (1997b). *Scann* **19**, 29.
- Hu, Y. and Pan, Y. (2001). *X-Ray Spectrom* **30**, 110.
- Huang, L.M., Cui, X.D., Dukovic, G. and O'Brien, S.P. (2004). *Nanotechnology* **15**, 1450.
- Huebener, R.P. (1988). In *Advances in Electronics and Electron Physics*, vol **70** (Hawkes, P.W., ed.) 1–78. (Academic Press, New York).
- Hunger, H.-J. and K  chler, L. (1979). *Phys Stat Sol* **56**, K45.
- Inokuti, M. and Manson, S.T. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 1–17. (Scann Electron Microscopy, Chicago).
- Isaacson, M. (1979a). *Ultramicroscopy* **4**, 193.
- Isaacson, M. (1979b). *Ultramicroscopy* **4**, 97.
- Isaacson, M. and Langmore, J.P. (1974). *Optik* **41**, 92.
- Isaacson, M.S. (1977). In *Principles and Techniques of Electron Microscopy*, vol **7** (Hayat, M.A., ed.) 1–78. (Van-Nostrand Reinhold, New York).
- Isabell, T.C., Fischione, P.E., O'Keefe, C., Guruz, M.U. and Dravid, V.P. (1999). *Microsc Microanal* **5**, 126.
- Ishikawa, A., Mizuno, F., Uchikawa, Y. and Maruse, S. (1973). *Jpn J Appl Phys* **12**, 286.
- Ishikawa, H., Dobashi, H., Kodama, T., Furuhashi, T. and Uchikawa, Y. (1993). *J Electr Microsc* **42**, 35.
- Ishitani, T. and Yaguchi, T. (1996). *Microsc Res Techn* **35**, 320.
- Jablonski, A., Salvat, F. and Powell, C.J. (2003). *NIST Electron Elastic-Scattering Cross Section Database #64 Version 3.1*. (National Institute of Standards and Technology, Gaithersburg, MD).
- Jacka, M. (2001). *J Electr Spectrosc Rel Phenom* **114–116**, 277.
- Ji, Y., Guo, H.S., Zhong, T.X., Zhang, H., Quan, X.L., Zhang, Y.Q. and Xu, X.D. (2005). *Ultramicroscopy*, **103**, 191.
- Joachimsthaler, I., Heiderhoff, R. and Balk, L.J. (2003). *Measure Sci Technol* **14**, 87.
- Johari, O. (ed.) (1968). *Proceedings of the Annual Scanning Electron Microscopy Symposium*. Annual Conferences since 1968. (IIT Research Institute; since 1987 *Scanning Microscopy*. Scanning Microsc. Intl., Chicago).
- Jonker, J.L.H. (1957). *Philips Res Rep* **12**, 249.
- Joy, D.C. (1984). In *Microbeam Analysis* (Romig, A.D. Jr. and Goldstein, J.I., eds.) 81–86. (San Francisco Press, San Francisco).
- Joy, D.C. (1987a). In *Electron Microscopy Microanalysis*, vol **90** (Chapman, J.N. and Craven, A.J., eds.) 175–180. Bristol: Institute of Physics Conf Series.
- Joy D.C. (1987b). *J Microsc* **147**, 51.
- Joy D.C. (1989). *Scann* **11**, 1.
- Joy D.C. (1991). *J Microsc* **161**, 343.
- Joy D.C. (1995). *Monte Carlo Modeling for Electron Microscopy and Microanalysis*. (Oxford University Press, New York).
- Joy D.C. (2001). *A Data Base on Electron-Solid Interactions*. This compilation is    David C Joy.
- Joy D.C. (2002). *J Microsc* **208**, 24.
- Joy D.C. and Luo, S. (1989). *Scann* **11**, 176.
- Joy D.C. Newbury, D.E. and Davidson, D.L. (1982). *J Appl Phys* **53**, R81.
- Judge, A.W. (1950). *Stereographic Photography*. (Chapman and Hall, London).
- Judge, F.J., Stubbs, J.M. and Philp, J. (1974). *J Phys E* **7**, 173.

- Kaczmarek, D. (1997). *Scann* **19**, 310.
- Kaczmarek, D. and Domaradzki, J. (2002). *Scann* **24**, 65.
- Kanaya, K. and Kayama, S.O. (1972). *J Appl Phys* **5**, 43.
- Kanaya, K. and Ono, S. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 69–98. (Scanning Electron Microscopy, Chicago).
- Katrakova, D. and Mucklich, F. (2002). *Prakt Metallogr* **39**, 644.
- Kazumori, H. (2002). *JEOL News* **37E**, 44.
- Kazumori, H., Honda, K., Matuya, M. and Date, M. (2004). In *8th Asia-Pacific Conference on Electron Microscopy* (8APEM Publication Committee, ed.) 52–53. (Yoshida Printing Inc., Kanazawa).
- Kellenberger, E. and Kistler, J. (1979). In *Advances in Structure Research by Diffraction Methods*, vol **3** (Hoppe, W. and Mason, R., eds.) 49–79. (Friedrich Vieweg, Wiesbaden).
- Kellenberger, E., Häner, M. and Wurtz, M. (1982). *Ultramicroscopy* **9**, 139.
- Kellenberger, E., Johansen, R., Maeder, M., Bohrmann, B., Stauffer, E. and Viliger, W. (1992). *J Microsc* **168**, 181.
- Keller, D.J. and Chih-Chung, C. (1992). *Surf Sci* **268**, 333.
- Khamsehpour, B. and Davies, S.T. (2004). *Vacuum* **45**, 1169.
- Khursheed, A. (2000). *Rev Sci Instrum* **71**, 1712.
- Khursheed, A. and Osterberg, M. (2004). *Scann* **26**, 296.
- Khursheed, A., Phang, J.C. and Thong, J.T.L. (1998). *Scann* **20**, 87.
- Kimoto, S., Hashimoto, H. and Suganama, T. (1966). In *The Electron Microprobe* (McKinley, T.D., Heinrich, K.F.J.H. and Wittrey, D.B., eds.) 480–489. (John Wiley, New York).
- Kirk, E.C.G., McMahon, R.A., Cleaver, J.R.A. and Ahmed, H. (1988). *J Vac Sci Technol B* **6**, 1940.
- Kirschner, J. (1976). In *Scanning Electron Microscopy*, vol. **1** (Johari, O., ed.) 215–220. (Scanning Electron Microscopy, Chicago).
- Kitching, S. and Donald, A.M. (1996). In *11th Europ Conf Electron Microscopy*, vol **1** (Comm. Europ. Soc. Microscopy, ed.) 138–139. Dublin: Comm. Europ. Soc. Microscopy.
- Klemperer, O. (1971). *Electron Optics*. (Cambridge University Press, Cambridge).
- Knoll, M. (1935). *Zeit Techn Phys* **11**, 467.
- Koblichka, M.R. and Koblichka-Veneva, A. (2003). *Phys C-Superconduct Appl* **392**, 545.
- Koike, H., Ueno, K. and Watanabe, N. (1970). In *Proceedings of the 7th International Congress on Electron Microscopy*, vol **1** (Favard, P., ed.) 241–242. (Societe Française de Microscopie Electronique, Grenoble).
- Kolarik, R. and Lenc, M. (1997). *Optik* **106**, 135.
- Kollath, R. (1956). In *Handbuch der Physik*, vol **XXI** (Flügge, S., ed.) 232–302. (Springer, Berlin).
- Kowalewski, E., Boassen, T. and Torsaeter, O. (2003). *J Petrol Sci Eng* **39**, 377.
- Kramers, H.A. (1923). *Phil Mag* **46**, 836.
- Krieger-Lassen, N.C., Juul-Jensen, D. and Conradsen, K. (1992). *Scann Microsc* **6**, 115.
- Kruit, P. (1991). In *Advances in Optical and Electron Microscopy*, vol **12** (Mulvey, T. and Sheppard, C.J.R., eds.) 93–137. (Academic Press, London).
- Krzyzanek, V. and Reichelt, R. (2003). *Microsc Microanal* **9** (Suppl 3), 110.
- Krzyzanek, V., Nüsse, H. and Reichelt, R. (2004). In *Proc 13th Europ Microscopy Cong*, vol **1** (van Dyck, D. and van Oostveldt, P., eds.) 391–392. (Belgian Society for Microscopy, Antwerp).
- Kultscher, N. and Balk, L.J. (1986). *Scann Electr Microsc* **1**, 33.

- Kyser, D.F. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 119–135. Chicago: Scanning Electron Microscopy, Inc.
- Lamvik, M.K. (1977). *Scann Electr Microsc* **1**, 401.
- Lawson, L. (1995). *Scann* **17**, 322.
- Lax, E. (ed.) (1967). D'ans. Lax. *Taschenbuch für Chemiker und Physiker*. (Springer, Berlin).
- Leamy, H.J. (1982). *J Appl Phys* **53**, R51.
- Leapman, R.D., Rez, P. and Mayers, D.F. (1980). *J Chem Phys* **72**, 1232.
- Lebiedzik, J. (1979). *Scann* **2**, 230.
- Lechner, P., Fiorini, C., Hartmann, R., Kemmer, J., Krause, N., Leutenegger, P., Longoni, A., Soltan, H., Stotter, D., Stotter, R., Strüder, L. and Weber, U. (2001). *Nucl Instrum Methods Phys Res Sect A* **458**, 281.
- Lencová, B. (1997). In *Handbook of Charged Particle Optics* (Orloff, J., ed.) 177–221. (CRC Press, New York).
- Lencová, B. and Lenc, M. (1994). *Optik* **97**, 121.
- Lian, J.S., Dong, Q.Z., Guo, Z.X., Xu, Q., Yang, J., Hu, J.D., Guan, Q.F. and Chen, B. (2005). *Mater Sci Eng A* **391**, 210.
- Liu, W., Ambe, T. and Pease, R.F. (1996). *J Vac Sci Technol B* **14**, 3738.
- Liukkonen, A. (1997). *Scann* **19**, 411.
- Lödding, B. and Reimer, L. (1981). *Beitr Elektronenmikr Direktabb Analyse Oberfl (BEDO)* **14**, 315.
- Lukianov, A.E., Spivak, G.V., Rau, E.I. and Gorodsky, D.D. (1972). In *Proc 5th Europ Congr Electron Microscopy* (Cosslett, V.E., ed.) 186–187. (The Institute of Physics, London).
- Lyman, C.E., Newbury, D.E., Goldstein, J.I., Williams, D.B., Romig, A.D., Armstrong, J.T., Echlin, P., Fiori, C.E., Joy, D.C., Lifshin, E. and Peters, K.-R. (1990). *Scanning Electron Microscopy, X-Ray Microanalysis and Analytical Electron Microscopy*. (Plenum, New York).
- Madden, H.H. (1981). *J Vac Sci Technol* **18**, 677.
- Madl, K., Toth, A.L. and Barna, A. (1988). *Inst Phys Conf Ser* **93**, 65.
- Malecki, M. and Roomans, G.M. (eds.) (1996). *The Science of Biological Specimen Preparations for Microscopy*. (Scanning Microscopy International, Chicago).
- Malkusch, W., Konerding, M.A., Klapthor, B. and Bruch, J. (1995). *Anal Cell Pathol* **9**, 69.
- Martin, J.P., Weimer, E., Frosien, J. and Lanio, S. (1994). *Eur Microsc Analysis* **28**, 43.
- Matzelle, T.R., Kruse, N. and Reichelt, R. (2000). *J Microsc* **199**, 239.
- Matzelle, T.R., Ivanov, D., Landwehr, D., Heinrich, L.A., Herkt-Bruns, C., Reichelt, R. and Kruse, N. (2002). *J Phys Chem B* **106**, 2861.
- Matzelle, T.R., Gnaegi, H., Ricker, A. and Reichelt, R. (2003). *J Microsc* **209**, 113.
- Mauler, A., Godard, G. and Kunze, K. (2001). *Tectonophysics* **342**, 81.
- McGuinness, P.E. (2003). *Scann* **25**, 221.
- McKenzie, J.M. and Bromely, D.A. (1959). *Phys Rev Lett* **2**, 303.
- McKinney, W.R. and Hough, P.V.C. (1977). *Scann Electr Microsc* **1**, 251.
- Menzel, S. and Wetzig, K. (1992). *J Mat Sci: Mat Electron* **3**, 5.
- Milani, M., Pucillo, F.P., Ballerini, M., Camatini, M., Gualtieri, M. and Martino, S. (2004). *Mat Character* **52**, 283.
- Minnich, B. and Lametschwandtner, A. (2000). *Scann* **22**, 173.
- Minnich, B., Leeb, H., Bernroider, E.W.N. and Lametschwandtner, A. (1999). *J Microsc* **195**, 23.
- Minnich, B., Krautgartner, W.-D. and Lametschwandtner, A. (2003). *Microsc Microanal* **9** (Suppl. 3), 118.

- Molday, R.S. and Maher, P. (1980). *Histochem J* **12**, 273.
- Muir, M.D. and Grant, P.R. (1974). In *Quantitative Electron Microscopy* (Holt, D.B., Muir, M.D., Grant, P.R. and Boswarva, I.M., eds.) 287–334. (Academic Press, London).
- Müller, D.A. and Silcox, J. (1994). In *13th Intern Congr Electron Microscopy*, vol 1. (Jouffrey, B. and Colliex, C., eds.) 741–742. (Les Editions de Physique Les Ulis, Paris).
- Müller, D.A. and Silcox, J. (1995). *Ultramicroscopy* **59**, 195.
- Müller, M. (ed.) (1985). *The Science of Biological Specimen Preparations for Microscopy and Microanalysis*. (Scanning Microscopy International, Chicago).
- Müller, M. and Hermann, R. (1990). *Hitachi Instr News* **19**, 50.
- Müller, M. and Hermann, R. (1992). In *10th Europ Congr Electron Microsc*, vol 3. (Megías-Megías, L., Rodríguez-García, M.I., Ríos, A. and Arias, J.M., eds.) 741–742. (Secretariado de Publicaciones de la Universidad de Granada, Granada).
- Müller, S.A. and Engel, A. (2001). *Micron* **32**, 21.
- Müller, T., Walther, P., Scheidegger, C., Reichelt, R., Müller, S. and Guggenheim, R. (1990). *Scann Micr* **4**, 863.
- Müllerová, I. (2001). *Scann* **23**, 379.
- Müllerová, I. and Frank, L. (2003). In *Advances in Imaging and Electron Physics* (Hawkes, P.W., ed.) 310–443. (Academic Press, Amsterdam).
- Müllerová, I. and Lenc, M. (1992). *Mikrochim Acta* **12** (Suppl), 173.
- Müllerová, I., Lenc, M. and Florián, M. (1989). *Scann Micr* **3**, 419.
- Müllerová, I., Frank, L. and Hutar, O. (2001). *Scann* **23**, 115.
- Mulvey, T. (1974). *Scann Electr Microsc* **1**, 43.
- Munro, E., Orloff, J., Rutherford, R. and Wallmark, J. (1988). *J Vac Sci Technol B* **6**, 1971.
- Murata, K. (1974). *J Appl Phys* **45**, 4110.
- Murata, K. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 311–329. (Scanning Electron Microscopy, Chicago).
- Nagase, M. and Kurihara, K. (2000). *Microelectron Engin* **53**, 57.
- Nagase, M. and Namatsu, H. (2004). *Jpn J Appl Phys Part 1* **43**, 4624.
- Nagatani, T. and Saito, S. (1989). *Inst Phys Conf Ser No* **98**, 519.
- Nembach, E. (1996). *Particle Strengthening of Metals and Alloys*. (Wiley, New York).
- Newbury, D., Wollman, D., Irwin, K., Hilton, G. and Martinis, J. (1999). *Ultramicroscopy* **78**, 73.
- Newbury, D.E. (1976). *Scann Electr Microsc* **1**, 111.
- Newbury, D.E. and Bright, D.S. (2005). *Scann* **27**, 15.
- Newbury, D.E., Joy, D.C., Echlin, P., Fiori, C.E. and Goldstein, J.I. (1987). *Advanced Scanning Electron Microscopy and X-Ray Microanalysis*. (Plenum, New York).
- Oatley, C.W. (1972). *The Scanning Electron Microscope—Part 1: The Instrument*. (Cambridge University Press, Cambridge).
- Oatley, C.W. and Everhart, T.E. (1957). *J Electron* **2**, 568.
- Ohnsorge, J. and Holm, R. (1973). *Rasterelektronenmikroskopie—Eine Einführung für Mediziner und Biologen*. Stuttgart: Georg Thieme Publishers.
- Oho, E. and Miyamoto, M. (2004). *Scann* **26**, 250.
- Oho, E., Sasaki, T., Adachi, K., Muranaka, Y. and Kanaya, K. (1986). In *11th International Congress on Electron Microscopy*, vol 1. (T. Imura, S. Maruse and T. Suzuki, eds.) 421–422. Kyoto: Japanese Society of Electron Microscopy.
- Ono, S. and Kanaya, K. (1979). *J Phys D: Appl Phys* **12**, 619.
- Oppel, W. and Jahrreiss, H. (1972). *Zeit Phys* **252**, 107.

- Paden, R.S. and Nixon, W.C. (1968). *J Phys E* **2**, 1073.
- Palmberg, P.W. (1973). *Anal Chem* **45**, 549A.
- Pan, H., Gao, H., Lim, S.H., Feng, Y.P. and Lin, J. (2004). *J Nanosci Nanotechnol* **4**, 1014.
- Peng, L.-M., Chen, Q., Liang, X.L., Gao, S., Wang, J.Y., Kleindiek, S. and Tai, S.W. (2004). *Micron* **35**, 495.
- Perovic, D.D., Castell, M.R., Howie, A., Lavoie, C., Tiedje, T. and Cole, J.S.W. (1994). In *13th International Congress on Electron Microscopy*, vol **1** (Jouffrey, B. and Colliex, C., eds.) 91–92. (Les Editions de Physique Les Ulis, Paris).
- Peters, K.-R. (1980). *Scann Electr Microsc* **I**, 143.
- Peters, K.-R. (1982). *Scann Electr Microsc* **4**, 1359.
- Peters, K.-R. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 363–372. (Scanning Electron Microscopy, Chicago).
- Pfefferkorn, G. (ed.) (1968). *Beiträge zur elektronenmikroskopischen Direktabbildung und Analyse von Oberflächen (BEDO)* (Engl.: Proceedings of Direct Electron Microscopic Imaging and Analysis of Surfaces). Annual Conferences since 1968 (vol **1**) until 1995 (vol **28**). (R.A. Remy, Münster).
- Pfefferkorn, G. and Pfautsch, M. (1971). *Beitr Elektronenmikr Direktabb Analyse Oberfl (BEDO)* **4**, 137.
- Philibert, J. and Tixier, R. (1969). *Micron* **1**, 174.
- Plows, G.S. and Nixon, W.C. (1968). *J Phys E* **1**, 595.
- Poelt, P. (2000). *Mikrochim Acta* **132**, 129.
- Poelt, P., Schmied, M., Obernberger, I., Brunner, T. and Dahl, J. (2002). *Scann* **24**, 92.
- Polak, J.M. and Varndell, J.M. (eds.) (1984). *Immunolabelling for Electron Microscopy*. (Elsevier Science Publ., Amsterdam).
- Postek, M.T. (1996). *Scann* **18**, 269.
- Postek, M.T. and Keery, W.J. (1990). *Rev Sci Instrum* **61**, 1648.
- Postek, M.T. and Vladár, A.E. (1996). *Scann* **18**, 1.
- Powell, C.J. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 19–31. (Scanning Electron Microscopy, Chicago).
- Powell, C.J. and Jablonski, A. (2000). *NIST Standard Reference Database #71 Electron Inelastic-Mean-Free-Path Database Version 1.1*. (National Institute of Standards and Technology, Gaithersburg, MD).
- Prior, D.J., Boyle, A.P., Brenker, F., Cheadle, M.C., Day, A., Lopez, G., Peruzzo, L., Potts, G.J., Reddy, S., Spiess, R., Timms, N.E., Trimby, P., Wheeler, J. and Zeterstrom, L. (1999). *Am Mineral* **84**, 1741.
- Raether, H. (1980). In *Springer Tracts in Modern Physics*, vol **88**. (Höhler, G., ed.). (Springer, Berlin).
- Rappaport, P. (1954). *Phys Rev* **93**, 246.
- Rasul, A. and Davidson, S.M. (1977). *Scann Electr Microsc* **I**, 233.
- Rau, E.I., Sennov, R.A., Chan, D.S.H. and Phang, J.C.H. (2004). In *Proc 13th Europ Congr Microscopy*, vol **1** (Timmermans, J.-P., Schryvers, D., van Dyck, D. and van Oostveldt, P., eds.) 411–412. (Belgian Society for Microscopy, Liège).
- Read, N.D. and Jeffree, C.E. (1990). *J Microsc* **161**, 59.
- Reed, S.J.B. (1996). *Electron Microprobe Analysis and Scanning Electron Microscopy in Geology*. (Cambridge University Press, Cambridge).
- Reichelt, R. (1995). In *Mikroskopie in Forschung und Praxis* (Robenek, H., ed.) 185–217. (GIT Verlag, Darmstadt).
- Reichelt, R. (1997). Unpublished results.
- Reichelt, R. and Engel, A. (1984). *Ultramicroscopy* **13**, 279.

- Reichelt, R. and Engel, A. (1985). *J Microsc Spectrosc Electr* **10**, 491.
- Reichelt, R. and Engel, A. (1986). *Ultramicroscopy* **19**, 43.
- Reichelt, R., Carlemalm, E., Villiger, W. and Engel, A. (1985). *Ultramicroscopy* **16**, 69.
- Reichelt, R. and Engel, A. and Aebi, U. (1988). In *Proc 9th Europ Cong Electron Microscopy*, vol **1** (Dickinson, H.G. and Goodhew, P.J., eds.) 33–34. (IOP Publishing, York).
- Reid, N. and Beesely, J.E. (1991). *Sectioning and Cryosectioning for Electron Microscopy*, vol **13**, (Elsevier Science Publ., Amsterdam).
- Reimer, L. (1967). *Elektronenmikroskopische Untersuchungs- und Präparationsmethoden*, 2nd ed. (Springer, Berlin).
- Reimer, L. (1968). *Optik* **27**, 86.
- Reimer, L. (1978). *Scann* **1**, 3.
- Reimer, L. (1984a). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 299–310. (SEM, Chicago).
- Reimer, L. (1984b). *Ultramicroscopy* **14**, 291.
- Reimer, L. (1985). *Scanning Electron Microscopy*. (Springer, Berlin).
- Reimer, L. (1993). *Image Formation in Low-Voltage Scanning Electron Microscopy*, vol **TT12**. (O'Shea, D.C., ed.). (SPIE Optical Engineering Press, Bellingham & Washington).
- Reimer, L. (1996). *Beitr Elektronenmikroskop Direktabb Analyse Oberfl (BEDO)* **29**, 1.
- Reimer, L. (1998). *Scanning Electron Microscopy*, 2nd ed. (Springer, Berlin).
- Reimer, L. and Krefting, E.-R. (1976). In *NBS Spec Publ 460* (Heinrich, K.F.J., Newbury, D.E. and Yakowitz, H., eds.) 45–60. (U.S. Department of Commerce, Washington, DC).
- Reimer, L. and Lödding, B. (1984). *Scann* **6**, 128.
- Reimer, L. and Pfefferkorn, G. (1973). *Raster-Elektronenmikroskopie*. (Springer, Berlin).
- Reimer, L. and Pfefferkorn, G. (1977). *Raster-Elektronenmikroskopie*. (Springer, Berlin).
- Reimer, L. and Schmidt, A. (1985). *Scann* **7**, 47.
- Reimer, L. and Stelter, D. (1986). *Scann* **8**, 265.
- Reimer, L. and Tollkamp, C. (1980). *Scann* **3**, 35.
- Reimer, L., Badde, H.G. and Seidel, H. (1971). *Zeit Angew Phys* **31**, 145.
- Reimer, L., Riepenhausen, N. and Tollkamp, C. (1984). *Scann* **6**, 155.
- Reuter, W. (1972). In *Sixth Intern Cong on X-ray Optics and Microanalysis* (Shinoda, G., Kohra, K. and Ichinokawa, T., eds.) 121–130. (University of Tokyo Press, Tokyo).
- Revel, J.-P., Haggis, G.H. and Barnard, T. (eds.) (1983). *The Science of Biological Specimen Preparation for Microscopy and Microanalysis*. (Scann Electron Microscopy, Chicago).
- Rez, P. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 43–49. (SEM, Chicago).
- Rickerby, D.G. and Thiot, J.-F. (1994). *Mikrochim Acta* **114/115**, 421.
- Rickerby, D.G., Wächter, N. and Reichelt, R. (1998). *Mikrochim Acta Suppl* **15**, 149.
- Ris, H. and Malecki, M. (1993). *J Struct Biol* **111**, 148.
- Robards, A.W. and Sleytr, U.B. (1985). *Low Temperature Methods in Biological Electron Microscopy*, vol **10**. (North-Holland, Amsterdam).
- Robards, A.W. and Wilson, A.J. (eds.) (1993). *Procedures in Electron Microscopy*. (John Wiley, Chichester).

- Robertson, K., Gauvin, R. and Finch, J. (2004). *Microsc Microanal* **10**, 721.
- Robin, M. (2001). *Oil Gas Sci Technol—Revue Institut Francais Petrole* **56**, 55.
- Robinson, V.N.E. (1975). *J Phys E* **8**, 638.
- Robinson, V.N.E. (1990). *Hitachi Instrum News* **19**, 32.
- Roshchupkin, D.V. and Brunel, M. (1993). *Scann Microsc* **7**, 543.
- Rossi, M.P., Ye, H.H., Gogotsi, Y., Babu, S., Ndungu, P. and Bradley, J.C. (2004). *Nano Lett* **4**, 989.
- Royall, C.P., Thiel, B.L. and Donald, A.M. (2001). *J Microsc* **204**, 185.
- Rukosujew, A., Reichelt, R., Fabricius, A.M., Drees, G., Tjan, T.T.D., Rothenburger, M., Hoffmeier, A., Scheld, H.H. and Schmid, C. (2004). *Ann Thorac Surg* **77**, 120.
- Ruska, E. (1979). *Die frühe Entwicklung der Elektronenlinsen und der Elektronenmikroskopie*. (Deutsche Akademie der Naturforscher Leopoldina, Halle).
- Russel, P.E. and Mancuso, J.F. (1985). *J Microsc* **140**, 323.
- Salih, S.M. and Cosslett, V.E. (1974). *Phil Mag* **30**, 225.
- Sato, M., Kageyama, K. and Nakagawa, M. (2000). *Hitachi Intrum News* **36**, 9.
- Sawyer, L.C. and Grubb, D.T. (1996). *Polymer Microscopy*. (Chapman & Hall, London).
- Schauer, P. and Autrata, R. (2004). In *13th Europ Congr Microscopy*, vol **1** (Schryvers, D., Timmermans, J.-P., van Dyck, D. and van Oostveldt, eds.) 69–70. (Belgian Society for Microscopy, Liège).
- Schenk, M., Fütting, M. and Reichelt, R. (1998). *J Appl Phys* **84**, 4880.
- Schimmel, G. and Vogell, W. (1970). *Methodensammlung der Elektronenmikroskopie*. vol **1** and **2**. (Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart).
- Schmid, R., Gaukler, K.H. and Seiler, H. (1983). *Scann Electr Microsc* **2**, 501.
- Schwartz, A.J., Kumar, M. and Adams, B.L. (eds.) (2000). *Electron Backscatter Diffraction in Materials Science*. (Kluwer Academic/Plenum Publ., New York).
- Schwarzer, R.A. (1997). *Micron* **28**, 249.
- Seah, M.P. and Dench, W.A. (1979). *Surf Interface Anal* **1**, 2.
- Seiler, H. (1967). *Zeit Angew Phys* **22**, 249.
- Seiler, H. (1968). *Beitr Elektronenmikr Direktabb Analyse Oberfl (BEDO)* **1**, 27.
- Seiler, H. (1976). *Scann Electr Microsc* **1**, 9.
- Seiler, H. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 33–42. (Scanning Electron Microscopy, Chicago).
- Seiler, H. and Kuhnle, G. (1970). *Zeit Angew Phys* **29**, 254.
- Sekiguchi, T. and Sumino, K. (1995). *Rev Sci Instrum* **66**, 4277.
- Sennhauser, U., Jacob, P. and Gasser, P. (2004). *Pract Metallogr* **41**, 199.
- Severs, N.J. and Shotton, D.M. (eds.) (1995). *Rapid Freezing, Freeze Fracture and Deep Etching*. (John Wiley, Chichester).
- Shea, S.P. (1984). In *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Kyser, D.F., Niedrig, H., Newbury, D.E. and Shimizu, R., eds.) 145–151. (Scanning Electron Microscopy, Chicago).
- Shea, S.P., Partain, L.D. and Warter, P.J. (1978). *Scann Electr Microsc* **1**, 435.
- Shibata, M. (2004). *JEOL News* **39**, 28.
- Shibata, Y., Arima, T. and Yamamoto, T. (1984). *J Microsc* **136**, 121.
- Siangchaew, K. and Libera, M. (2000). *Phil Mag A* **80**, 1001.
- Siegel, B.M. and Beaman, D.R. (1975). *Physical Aspects of Electron Microscopy and Microbeam Analysis*. (John Wiley, New York).
- Simmons, S.R., Pawley, J.B. and Albrecht, R.M. (1990). *J Histochem Cytochem* **38**, 1781.

- Sitte, H. (1984). In *The Science of Biological Specimen Preparation for Microscopy and Microanalysis* (Johari, O., ed.) 97–104. (Scanning Electron Microscopy, Chicago).
- Sitte, H. (1996). In *The Science of Biological Specimen Preparation for Microscopy and Microanalysis* (Malecki, M. and Roomans, G.M., eds.) 387–466. (Scanning Microscopy International, Chicago).
- Smith, G.C., Park, D. and Cochonneau, O. (1995). *J Microsc* **178**, 48.
- Spivak, G.V. (1996). *Izv Akad Nauk SSSR, Ser Fiz* **30**, 749.
- Steinbrecht, R.A. and Zierold, K. (eds.) (1987). *Cryotechniques in Biological Electron Microscopy*. (Springer, Berlin).
- Stelmashenko, N.A., Craven, J.P., Donald, A.M., Terentjev, E.M. and Thiel, B.L. (2001). *J Microsc* **204**, 172.
- Stemmer, A., Reichelt, R. Wyss, R. and Engel, A. (1991). *Ultramicrosc* **35**, 255.
- Stephen, J., Smith, B.J., Marshall, D.C. and Wittam, E.M. (1975). *J Phys E* **8**, 607.
- Stephenson, S.T. (1957). In *Handbuch der Physik*, vol **30** (Flügge, S., ed.) 337–370. (Springer, Berlin).
- Stintzing, H. (1927). Verfahren und Einrichtung zum automatischen Nachweis, Messung und Zählung von Einzelteilchen beliebiger Art, Form und Grösse. German Patent No. 485155, patent applied for on 14-06-1927, patent granted on 10-10-1929.
- Stracke, W., González, G., Keller, U., González, S., Ricker, A. and Reichelt, R. (2003) *Microsc Microanal* **9**(Suppl 3), 126.
- Strane, J., Marks, L.D., Luzzi, D.E., Buckett, M.I., Zhang, J.P. and Wessels, B.W. (1988). *Ultramicroscopy* **25**, 253.
- Strüder, L., Lechner, P. and Leutenegger, P. (1998a). *Naturwissenschaften* **85**, 539.
- Strüder, L., Meidinger, N., Stotter, D., Kemmer, J., Lechner, P., Leutenegger, P., Soltan, H., Eggert, F., Rohde, M. and Schulein, T. (1998b). *Microsc Microanal* **4**, 622.
- Sudraud, P., Ballongue, P., Varoquaux, E. and Avenel, O. (1987). *J Appl Phys* **62**, 2163.
- Szmaja, V. (2002). *Czechoslovak J Phys* **52**, A145.
- Szmaja, W. (2000). *J Magnet Magnet Mater* **219**, 281.
- Tang, X. and Joy, D.C. (2003). *Scann* **25**, 194.
- Thiberge, S., Nechushtan, A., Sprinzak, D., Gileadi, O., Behar, V., Zik, O., Chowars, Y., Michaeli, S., Schlessinger, J. and Moses, E. (2004). *Proc Natl Acad Sci USA* **101**, 3346.
- Thiel, B.L., Bache, I.C., Fletcher, A.L., Meredith, P. and Donald, A.M. (1997). *J Microsc* **187**, 143.
- Thiel, B.L., Toth, M. and Craven, J.P. (2004). *Microsc Microanal* **10**, 711.
- Thornton, P.R. (1972). *Scanning Electron Microscopy. Application to Materials and Device Science*. (Chapman & Hall, London).
- Toth, M. and Phillips, M.R. (2000). *Scann* **22**, 370.
- Troughton, P.R. and Donaldson, L.A. (1972). *Probing Plant Structure*. (Chapman & Hall, London).
- Troyon, M., Lei, H.N. and Bourhettar, A. (1992). *Ultramicroscopy* **42–44**, 1564.
- Tuggle, D.W. and Swanson, L.W. (1985). *J Vac Sci Technol B* **3**, 220.
- Uchikawa, Y. and Ikeda, S. (1981). *Scann Electr Microsc* **1**, 209.
- Uno, S., Honda, K., Nakamura, N., Matsuya, M. and Zach, J. (2004). In *8th Asia-Pacific Conf Electron Microscopy* (8APEM Publication Committee, ed.) 46–47. (Yoshida Printing Inc., Kanazawa).
- Ushiki, T., Yui, R. and Fujita, T. (1986). *J Electr Microsc* **35**, 157.
- van Essen, C.G., Schulson, E.M. and Donaghay, R.H. (1971). *J Mater Sci* **6**, 213.

- Venables, J.A. and Harland, C.J. (1973). *Phil Mag* **27**, 74.
- Verkleij, A. and Leunissen, J. (1989). *Immunogold Labelling in Cell Biology*. (CRC Press, Boca Raton, FL).
- Volbert, B. (1982). *Scann Electr Microsc* **3**, 897.
- Volbert, B. and Reimer, L. (1980). *Scann Electr Microsc* **4**, 1.
- von Ardenne, M. (1938). *Zeit Techn Phys* **19**, 407.
- von Nahmen, A., Schenk, M., Sieber, M. and Amrein, M. (1997). *Biophys J* **72**, 463.
- Wall, J.S. (1979). *Scann Electr Microsc* **2**, 291.
- Wall, J.S. (1980). *Scann Electr Microsc* **1**, 99.
- Walther, P. (2003). *J Microsc* **212**, 34.
- Walther, P. and Müller, M. (1985). In *43rd Annual Meeting of the Electron Microscopy Society of America* (Bailey, G.W., ed.) 538–541. (San Francisco Press, San Francisco).
- Walther, P. and Müller, M. (1986). In *Science of Biological Specimen Preparation* (Müller, M., Becker, R.P., Boyde A. and Wolosewick, J.J., eds.) 195–201. (Scanning Electron Microscopy, Chicago).
- Walther, P. and Müller, M. (1999). *J Microsc* **196**, 279.
- Walther, P., Hentschel, J., Herter, P., Müller, T. and Zierold, K. (1990). *Scann* **12**, 300.
- Walther, P., Wehrli, E., Hermann, R. and Müller, M. (1995). *J Microsc* **179**, 229.
- Wang, C., Krausch, G., Radlein, E., Tratzky, S., Schramm, M. and Weber, A. (2004). *Glass Sci Technol* **77**, 273.
- Wardly, G.A. (1971). *J Appl Phys* **42**, 376.
- Wells, O.C. (1974). *Scanning Electron Microscopy*. (McGraw-Hill, New York).
- Wells, O.C. (1975). *Scanning Electron Microscopy*. (McGraw-Hill, New York).
- Wells, O.C. (1978). *Scann Electr Microsc* **1**, 293.
- Wells, O.C. (1979). *Appl Phys Lett* **35**, 644.
- Wepf, R. and Gross, H. (1990). In *12th Intern Congr Electron Microscopy* (Peachy, L.D. and Williams, D.B., eds.) 6–7. (San Francisco Press, Seattle).
- Wepf, R., Amrein, M., Bürkli, U. and Gross, H. (1991). *J Microsc* **163**, 51.
- Werner, W.S.M., Lakatha, H., Smith, H.E., LeTarte, L., Ambrose, V. and Baker, J. (1998). *J Vac Sci Technol B: Microelectron Nanometer Struct* **16**, 420.
- Wetzig, K. and Schulze, D. (eds.) (1995). *In Situ Scanning Electron Microscopy in Materials Research*. (Akademie Verlag, Berlin).
- Wight, S.A. and Zeissler, C.J. (2000). *Scann* **22**, 167.
- Wilkinson, A.J. (2000). *J Electr Microsc* **49**, 299.
- Wilkinson, A.J. and Hirsch, P.B. (1997). *Micron* **28**, 279.
- Willison, J.H.M. and Rowe, A.J. (1980). *Replica, Shadowing and Freeze-Etching Techniques*. (North-Holland, Amsterdam).
- Wittry, D.B. (1966). In *Fourth Intern Cong on X-ray Optics and Microanalysis* (Castaing, R., Deschamps, P. and Philibert, J., eds.) 168–180. (Hermann, Paris).
- Wollman, E.W., Frisbie, C.D. and Wrighton, M.S. (1993). *Langmuir* **9**, 1517.
- Wu, C.H. and Wittry, D.B. (1974). *J Appl Phys* **49**, 2827.
- Wurster, R. (1997). *J Trace Microprobe Techn* **15**, 467.
- Yakimov, E.B. (2002). *J Phys Cond Matter* **14**, 13069.
- Yakobi, B.G. and Holt, D.B. (eds.) (1990). *Cathodoluminescence Microscopy of Inorganic Solids*. (Plenum, New York).
- Yamaguchi, J., Shibano, M. and Saito, T. (1994). In *13th Intern Congr Electron Microscopy*, vol **3A** (Jouffrey, B. and Colliex, C., eds.) 43–44. (Les Editions de Physique Les Ulis, Paris).
- Yamamoto, T., Nishizawa, H. and Tsuno, K. (1976). *Phil Mag* **34**, 311.
- Yamazaki, Y. (2004). *Electrochim Acta* **50**, 663.
- Yew, N.C. (1971). *Scann Electr Microsc* **33**.

- Zach, J. (1989). *Optik* **83**, 30.
- Zach, J. and Rose, H. (1988a). In *9th Europ Congr Electron Microscopy*, vol **1** (Dickinson, H.G. and Goodhew, P.J., eds.) 81–82. (IOP Publishing Ltd., York)
- Zach, J. and Rose, H. (1988b). *Scann* **8**, 285.
- Zadrazil, M., El-Gomati, M.M. and Walker, A. (1997). *J Comput Assist Microsc* **9**, 123.
- Zaefferer, S. (2004). *JEOL News* **39**, 10.
- Zaefferer, S., Ohlert, J. and Bleck, W. (2004). *Acta Mater* **52**, 2765.
- Zeitler, E. (1978). *Ann NY Acad Sci* **306**, 62.
- Zeitler, E. and Bahr, G.F. (1962). *J Appl Phys* **33**, 847.
- Zhao, Y.C., Chen, M., Zhang, Y.N., Xu, T. and Liu, W.M. (2005). *Mater Lett* **59**, 40.
- Zhao, Z.X., Cui, R.Q., Meng, F.Y., Zhou, Z.B., Yu, H.C. and Sun, T.T. (2005). *Solar Energy Mater Solar Cells* **86**, 135.
- Zworykin, V.K., Hillier, J. and Snyder, R.L. (1942). *ASTM Bull* **117**, 15.

Analytical Electron Microscopy

Gianluigi Botton

1 Introduction

The term analytical electron microscopy (AEM) refers to the collection of spectroscopic data in the transmission electron microscope (TEM) based on various signals generated following the inelastic interaction of the incident electron beam with the sample. These signals can be used to identify and quantify the concentration of the elements present in the analyzed area, map their distribution in the sample with high spatial resolution (down to 1 nm or better), and even determine their chemical state. Although there are several signals generated by the interaction of primary incident electrons on a sample (Figure 4–1), the two main techniques at the core of AEM are based on the detection of X-ray signals generated in the sample by the primary incident electrons with the technique called energy dispersive X-ray spectroscopy (EDXS) and the measurement of the energy lost by the incident electrons with electron energy loss spectroscopy (EELS). These techniques are used within a TEM (Figure 4–2) equipped with EDXS, EELS detectors, and other detectors to record TEM images, diffraction patterns, and other signals that can be combined to form other images containing additional information on the chemical nature and structure of the sample. Other techniques such as cathodoluminescence, Auger spectroscopy, and electron beam-induced current imaging are also part of the arsenal of techniques available in the TEM but are less frequently implemented in commercial TEMs or have been used only as proof of concept. Convergent beam electron diffraction has traditionally been considered one of the AEM techniques, but given the breadth of this topic, the chapter will focus on the principles of EDXS and EELS, instrumentation requirements, the quantitative aspects of microanalysis, and the respective advantages and limitations. Throughout this chapter, several examples of application of these techniques in practical problems will be given. A comparison of these AEM methods with other analytical techniques is presented in a summary. Since this topic has been extensively covered in key detailed early textbook references (Joy et al., 1986; Williams and Carter, 1996) an update on the current status of the tech-

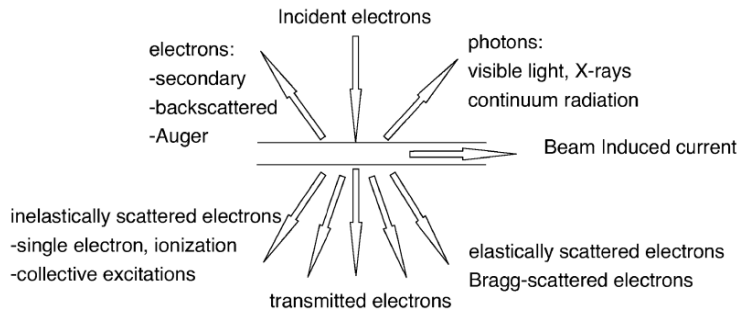


Figure 4–1. Signals generated when high-energy incident electrons interact with a thin sample.

niques will be given based on recent published literature and examples of work.

1.1 Overview of EDXS

As incident electrons generated by the electron gun in the TEM interact with atoms and their electrons in the solid, various inelastic processes are generated. When the primary electron excites bound electrons on a given shell (Figure 4–3), a core hole is created. This excited state is created only temporarily and the empty state is therefore filled by electrons from higher energy levels through a deexcitation process by generation of photons (of sufficient energy to be considered in the X-ray

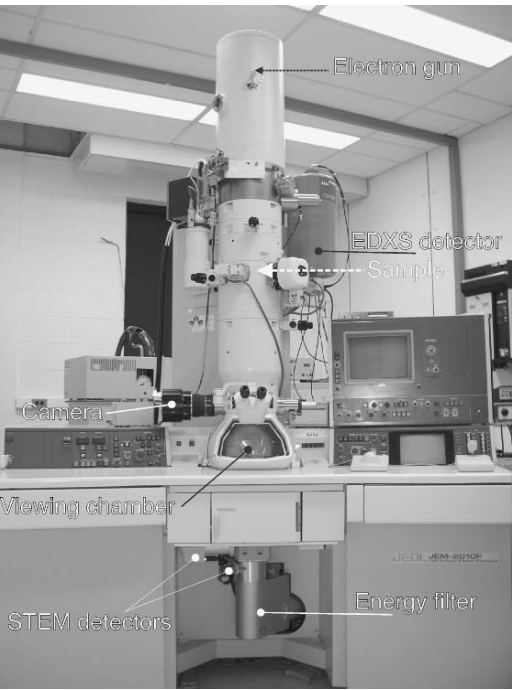


Figure 4–2. Example of a commercially available analytical TEM. Visible are the energy dispersive X-ray detector, the energy loss spectrometer, and digital cameras and detectors.

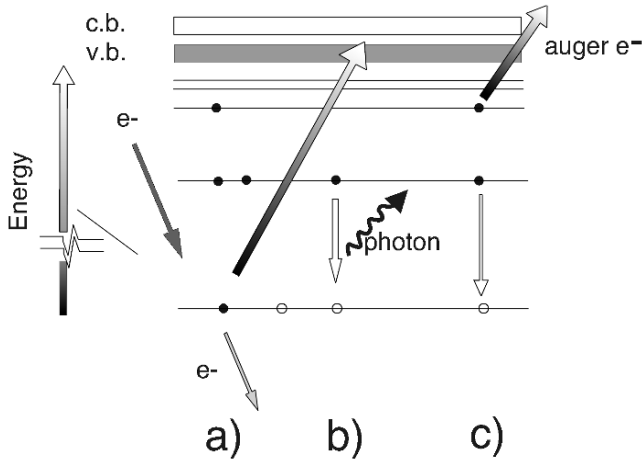


Figure 4–3. Energy level diagrams showing the transitions required for the generation of X-rays and Auger electrons following the excitation of core electrons by the primary incident electrons.

part of the spectrum) or Auger electrons. The energy of these X-rays, typically in the range of few hundred electronvolts to 20–40 keV, is also characteristic of the energy differences between the levels involved in the excitation and deexcitation process (Figures 4–4 and 4–5). Because various energy shells can be excited, peaks in the spectra are labeled according to the corresponding quantum levels involved in the transitions based on the nomenclature illustrated in Figure 4–4. The dependence on characteristic energy levels of the bound electrons makes it possible to identify the atomic number of the elements that have been involved in the excitation process. However, not all transitions between

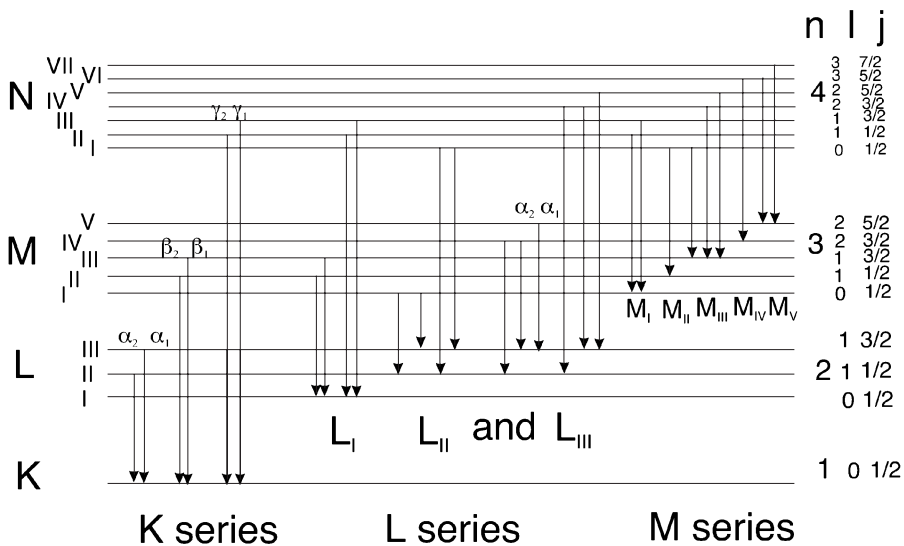


Figure 4–4. Detailed energy levels, their associated quantum number n, l , and j , and associated families with the transitions respecting the selection rules described in the text.

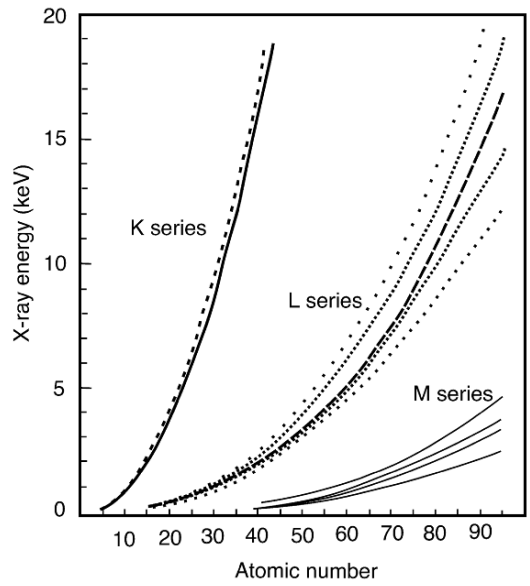


Figure 4–5. Characteristic X-ray energy for different families of lines as a function of atomic number.

the energy levels are allowed: there must be a change in the angular momentum quantum numbers ℓ and j for transitions to be observed according to the rule $\Delta\ell = \pm 1$, $\Delta j = -1, 0, +1$ (Figure 4–4). The probability of these excitations and generation of X-rays or Auger electrons varies with atomic number based on cross sections and fluorescence yield. The efficiency of the detection of the X-rays also varies according to their energy due to absorption in the detector material. In addition, the X-rays generated can be absorbed in the sample itself before reaching the detector. The spectrum itself consists of the characteristic X-ray peaks for the excited atoms present in the sample superimposed on a continuum noncharacteristic background (Figure 4–6).

To relate the intensity detected in spectra to the concentration of the elements, several effects must therefore be considered: the intensity of the peaks in the spectra must be corrected using cross sections and fluorescence corrections and absorption in the sample and in the detector. Figure 4–7 summarizes the process of signal generation, collection, display, and quantification in a flow chart containing various steps of signal correction. Based on the quantification of signals, it is possible

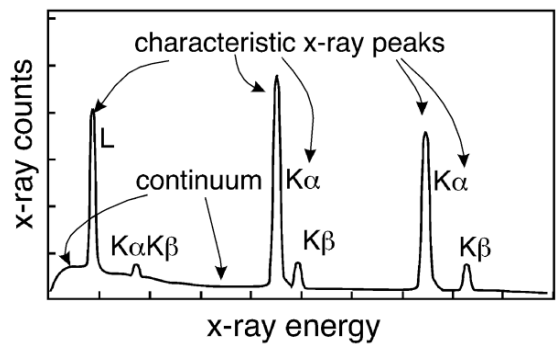


Figure 4–6. EDS spectrum demonstrating the characteristic lines (with energies presented in Figure 4–5) and the continuum background.

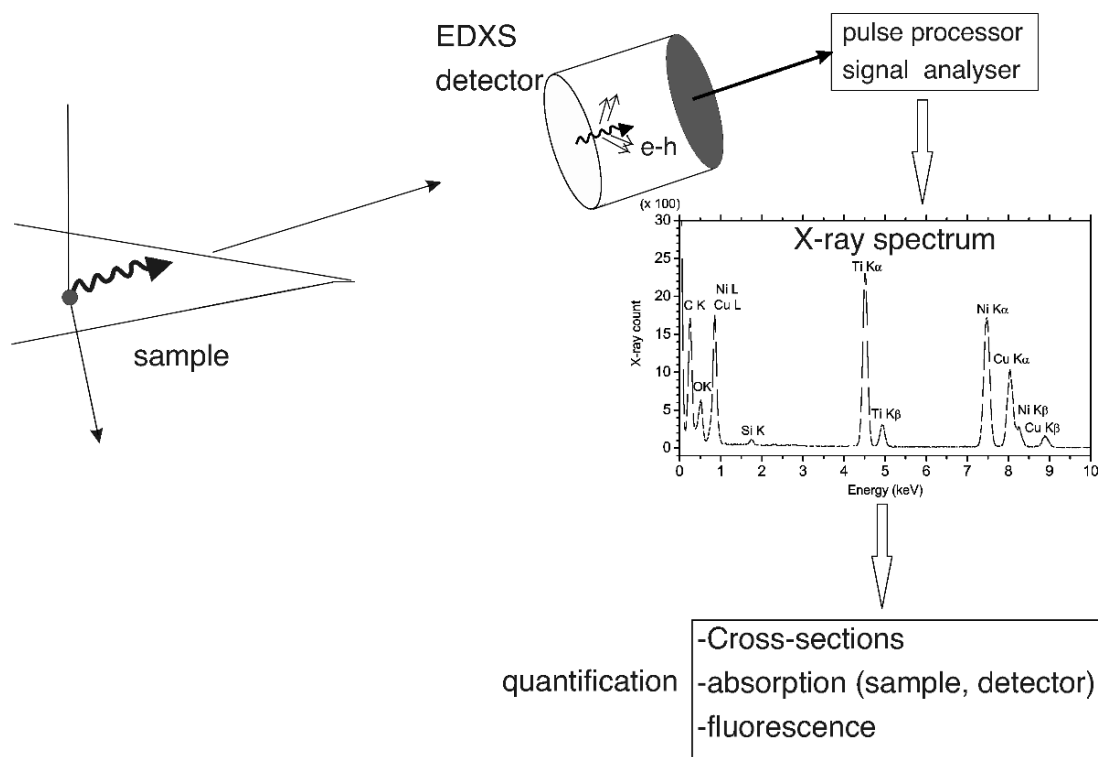


Figure 4–7. Schematic diagram for the generation of X-rays in the sample, detection of the X-rays in the detector (by generation of electron and hole pairs), pulse analysis, and quantification process of the spectrum to derive the composition of the sample. Detailed procedures of quantification are described in Section 4.

to measure the local concentration to an accuracy limited by the statistic uncertainty of the spectrum and the errors in the cross sections.

Due to the limited signal collection efficiency of the detector, the finite acquisition time and the presence of a noncharacteristic background under the element-specific peaks, the technique is generally not suitable for the detection of trace elements in samples but it can provide rapid quantitative data on elements to within a few atomic percent accuracy, with detection limits typically of a few percent. Improvements in collection efficiency and long acquisition times, however, have led to detection limits of fractions of 1% (see Section 7 of this chapter).

Using software that can control the electron beam position on the sample and the data collection in a sequential manner, generated signals can be collected over an area of the sample so that the intensity of characteristic signals, as a function of position, represents the local composition variations in the sample as displayed in an elemental map (Figure 4–8). Further processing of spatially resolved spectra can also be carried out so that quantitative maps and statistical analysis of the concentration and element distribution can be displayed. Details of the performance and limitations of EDXS as well as the approaches used to quantify the data are given in subsequent sections.

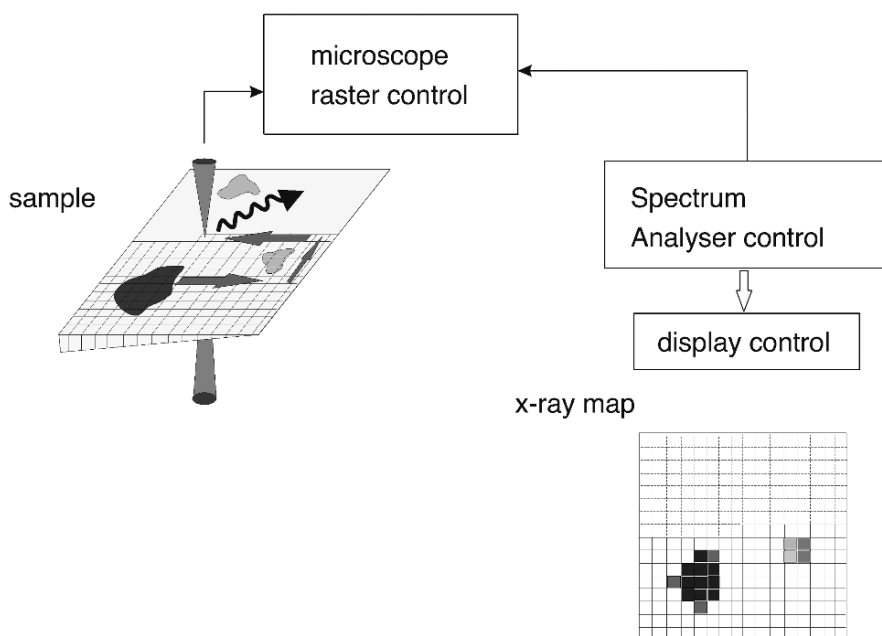


Figure 4–8. Schematic diagram describing the processes necessary to record an elemental map with EDXS. The synchronous scan of the beam is associated with a pixel position where the recording of the signal takes place. The peak intensity is measured for each element at each pixel position and is plotted in a two-dimensional elemental map.

1.2 Overview of Electron Energy Loss Spectroscopy

EELS is based on the measurement of the energy that the primary incident electrons have lost while causing various inelastic processes in the sample. The excitation of electrons from core energy levels that precedes the generation of X-rays or Auger electrons is only one of the mechanisms by which the primary electrons can lose some of their energy. For this particular case, the energy loss process gives rise to signals known as “core edges” with characteristic energies closely related to the binding energy of the excited electrons (Figure 4–9) in the atoms. Excitation of valence electrons into the conduction band and collective excitation of weakly bound electrons are also potential energy loss processes (Figure 4–9) called plasmons. These losses contribute to the low loss part of the spectrum (Figure 4–10) from a few to about 50–100 eV. Although characteristic core edges can appear at relatively low energies, strong low-loss signals contain information about the materials dielectric properties. With reference spectra and databases, it is therefore possible to identify particular compounds based on the shape of the low-loss spectrum. Identification of the chemical state and the compound is also possible through the analysis of fine modulations appearing in the first few electronvolts from the core edges threshold. These modulations are known as electron energy loss near-edge structure (ELNES) and contain information about the electronic structure and bonding environment of the excited atom. This information is now frequently used in the study of electronic structure and chemical state

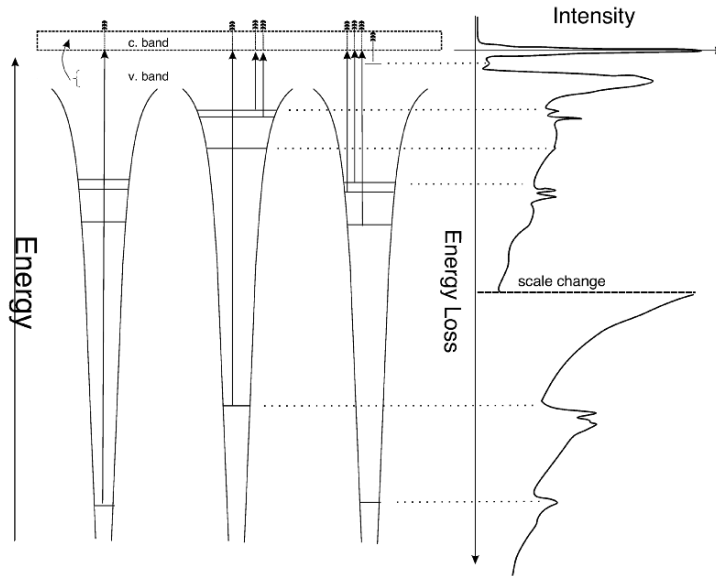


Figure 4-9. Schematic diagram of the associated features in a spectrum. The core edges arise from transitions from deep core levels to the first unoccupied states above the valence band (and Fermi energy) and the continuum. Excitation from defect states in the gap are also shown as well as collective excitations of valence electrons giving rise to broad features called plasmon peaks.

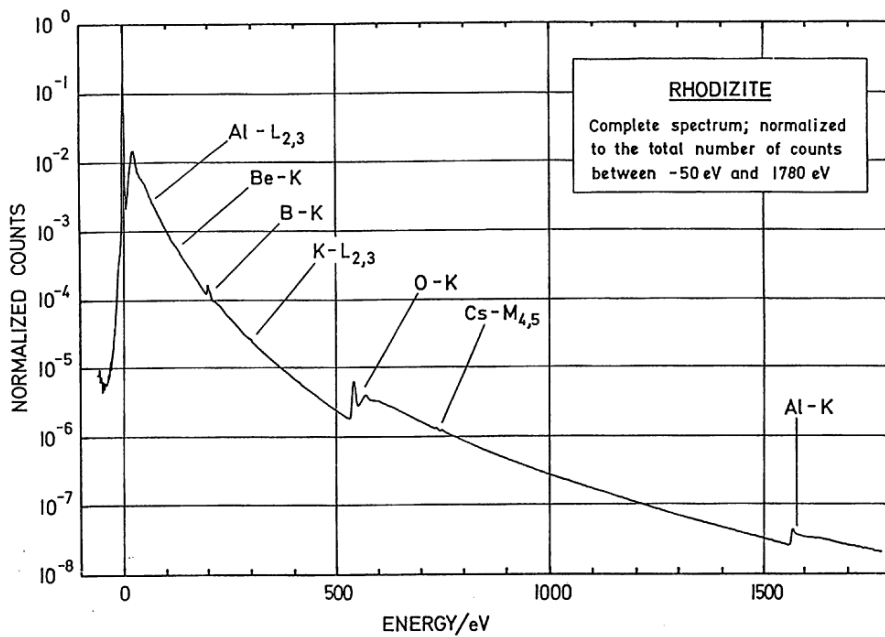


Figure 4-10. Full energy loss spectrum recorded over a large energy range demonstrating the large dynamic range of the recorded intensities and the relative intensity of core losses as compared with the background. (Courtesy of H. Sauer, Fritz-Haber Institut/MPG, Berlin.)

of materials with applications ranging from semiconductor devices to the study of minerals whereas low-loss structures have been used in the study as diverse as biological structures to superconductors.

Similar to the case of EDXS microanalysis, the intensity of core edges is related to the probability of excitation and thus to cross-section values and the concentration of elements. The intensity of edges relative to the background, however, is strongly dependent on the thickness of the analyzed area and edges can remain simply undetected in the case of thick samples. As in the case of EDXS analysis, this technique is not ideal for routine detection of trace elements due to the very intense background typically dominating the signal at the edges and the overall small recorded signal of edges with respect to the total recorded signal (Figure 4-10), although acquisition conditions can be optimized for the detection of minor constituents (discussed in Section 7.1).

EELS signals offer the advantage of being generated by a primary event: the loss of energy. As compared to EDXS, the intensity of recorded signals is therefore not linked to the secondary process of fluorescence resulting in the deexcitation via X-ray emission. For light elements such as O, N, C, B, this is a remarkable advantage because the fluorescence yield (the probability of X-ray relative to Auger electrons generation, see Section 4.1) decreases by orders of magnitude as compared to higher atomic number elements such as transition metals. Therefore, EELS analysis is generally considered to be more appropriate for the detection of light elements than EDXS analysis.

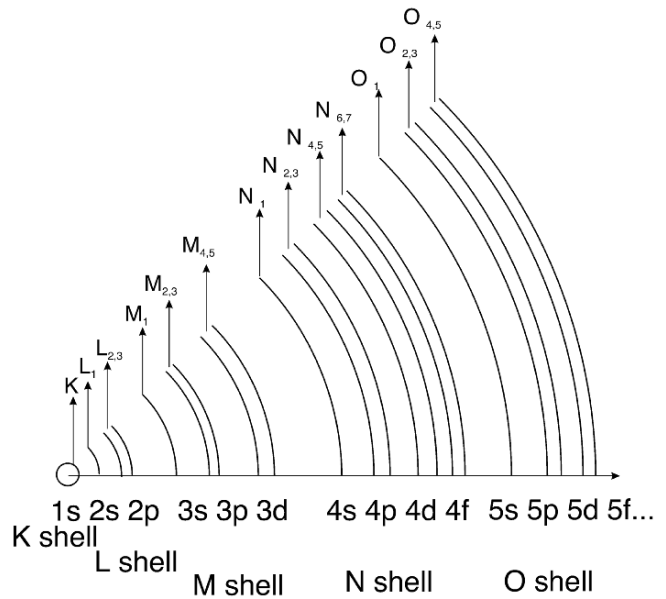
The core edges can be identified and labeled according to the energy levels of the ejected electron and the respective quantum numbers. K, L, M, N, O edges are related to the transitions involving $n = 1, 2, 3, 4, 5$ principal quantum numbers, respectively. The angular momentum quantum numbers ℓ (s,p,d,f) and j lead to sublabels as indicated in Figure 4-11. A summary of the information that can be retrieved from EELS spectra is shown in Table 4-1 (Colliex, 1996).

The collection of EELS spectra is carried out with an energy loss spectrometer either attached at the bottom of the TEM column (post-column filter) or within the projector lens system (in-column filter) (Section 2.4.1). In both cases, the electron energy distribution is analyzed with one or a series of dispersing elements that separate the electrons according to their energy. The dispersion will result in the generation of a spectrum that will be recorded on a detector system. Depending on the filter electron optical configuration and detector system, spectra, images and diffraction patterns corresponding to specific energy losses can be recorded as discussed in Section 2.4.1. When images or diffraction patterns are obtained using electrons with specific energy losses or with electrons having lost no energy, the technique is called energy-filtered microscopy.

1.3 Comparison with Other Spectroscopies

EDXS and EELS offer information complementary to other techniques that yield compositional or spectroscopic data typically available in

Figure 4–11. Diagram demonstrating the origin of the spectroscopic labels of energy loss spectra and the associated core levels. (Adapted from EELS Atlas, C.C. Ahn and O.L. Krivanek.)



surface analysis instruments such as X-ray photoelectron spectroscopy (XPS), Auger spectroscopy, X-ray absorption spectroscopy, inverse photoemission, etc. XPS provides information on the binding energy (E_b) of electrons in atomic core levels as ejected by incident photons. These photoelectrons with kinetic energy E_k are detected in vacuum as they

Table 4–1. Information from EELS spectra.

Spectral region	Type of information	Application
Full spectrum	Thickness, inelastic mean free path	All analytical methods of quantification, volume fraction
Low-loss	Average electron density	Microanalysis of alloys, H content, identification of phases
Low-loss	Joint density of states	Optical properties of solids, electronic structure, correlation effects, bandgap measurement
Low-loss	Dielectric properties/interfaces	Relativistic effects, interface excitation effects/modes
Core-loss	Edges intensity	Quantification of concentration of elements
Core-loss	Near edge structures	Chemical state, coordination, ionicity/valence, phase identification
Core-loss	Extended fine structure	Determination of radial distribution functions
Core-loss	White lines: density of holes in the d-band	Formal charge, charge transfer

escape the sample surfaces and the system workfunction (ϕ). This process typically detects photoelectrons with very low kinetic energy as the incident X-rays photons (E_v) are typically a few kiloelectronvolts and $E_k = E_v - (E_b + \phi)$. The technique thus provides essentially information from the topmost few atomic layers and is used to analyze ultra-thin layers and quantify composition of thin films deposited on surfaces, surface contaminants etc. Although the lateral resolution is typically about a few tens of micrometers in commercial instruments, near micrometer resolution can be achieved in synchrotron facilities and in imaging XPS instruments. Using ion beams to sputter the sample surface, depth profiling can be carried out with a depth resolution of 2–5 nm due to the (energy-dependent) escape depth of the electrons and, when sputtering is used, the ballistic mixing induced by the incident ions. Angular resolved XPS methods can reach a depth resolution of about 1–2 nm as the escape angle can be tuned with the spectrometer. The technique can therefore provide information on the composition of surface layers and changes in the chemical state of atoms as reflected in the changes in binding energy.

In Auger spectroscopy, the energy of the Auger electrons typically ranges from a few tens of electronvolts to 1–2 keV and, as in the case of XPS, the escape depth from the sample surface is also limited to the topmost few nanometers. The Auger electron energy between energy levels A (initial core level) and levels B and C (secondary levels) (Figure 4–3) is determined by the energy levels involved in the transition as $E_{ABC} = E_A - E_B - E_C - \phi$. The changes in bonding due to changes in oxidation state or structure are therefore reflected in the energy of Auger peaks as the energy levels would be affected by the changes in bonding. The technique therefore provides both information on the elemental composition and the chemical state.

X-Ray absorption spectroscopy provides information on the absorption process of incident photons caused by transitions from inner-shell energy levels to the unoccupied states just above the Fermi energy and the continuum free states. The technique is therefore complementary to EELS as unoccupied states are probed but it offers the advantage of giving access to higher energy edges and the possibility of analysis in a controlled nonvacuum environment. With zone plate focusing of incident photons in third-generation synchrotrons, it is possible to obtain spot sizes of 15–30 nm. The X-ray absorption process can be directly observed in transmission mode or via indirect yield of electrons generated via the absorption process (total electron yield or fluorescence yield). In the latter case, the technique becomes surface sensitive as the escape depth of detected electrons is limited by their energy. Detection of elements in ppm concentration is possible due to the low background of edges as compared to EELS.

As compared with these spectroscopic techniques, EDXS, carried out with typical commercial detectors, can be considered a “bulk” analysis technique yielding elemental information through the thickness of the thin TEM foil with practically no content on the chemical state of the detected elements. EELS in the analytical microscope also provides information on the chemical composition through the thin foil thick-

Table 4–2. Comparison of spectroscopy techniques.

Technique	Lateral resolution limits	Depth resolution	Detection limit	Elemental information	Spectroscopic information
EDXS	1–2 nm	No	Minor	Yes	No
EELS	0.5–1 nm	No	Minor	Yes	Yes
Auger	10–50 nm	Yes (2–5 nm)	Minor	Yes	Yes
XPS	1–10 μ m	Yes (1–5 nm)	Minor	Yes	Yes
XAS	20–100 nm	No	Trace	Yes	Yes

ness, but offers the clear advantage of providing spectroscopic information on the chemical state.

Table 4–2 summarizes the general applications of the techniques, limitations, resolution, etc. Characteristic details on resolution vary depending on the acquisition conditions, energy of the elements of interest, and efficiency of the detection system.

2 Instrumentation

The ultimate aims of AEM are to analyze materials with high spatial resolution. These goals require the use of electrons source and electron optic components capable of producing intense beam currents into small electron probes, and detectors to collect the various analytical signals generated from the interaction of incident electrons with atoms in the solid. These requirements are met in a TEM configured for analytical work offering bright electron sources, flexible condensing optics, clean vacuum, and a range of detectors for imaging and spectroscopy. Thin and very clean samples are a necessity for achieving the ultimate performance expected from the high-spatial resolution techniques and these are as important as the quality of the microscope.

2.1 Electron Sources and Probes

TEMs for conventional TEM, high-resolution TEM, and AEMs can be equipped with two types of electron sources—thermionic and field emission. Thermionic sources such as W hairpin filaments and refractory crystals such as LaB₆ operate at high temperature and emit electrons that are subsequently accelerated by the anode potential (100–200 kV or more). The thermionic sources are heated either by a flow of current through the emitting material itself (for the W hairpin cathode filament) or by thermal contact of a low-workfunction emitter material such as LaB₆ and resistive heating of a W wire supporting material. The field-emission gun (FEG) source operates on the principles of electron tunneling from the tip to vacuum following the application of a strong electric field ($\approx 10^9$ V/m) generating a very large electric field gradient at the tip of the cathode. This high field results in a very narrow potential barrier allowing tunneling of the electrons from a low-workfunction metallic tip to vacuum. This emission is generated

from a very small area (in the order of 10 nm or less) of a single crystal tip resulting in high current density. There are variants to this gun configuration. Cold FEGs emit at room temperature and have the disadvantage of requiring ultrahigh vacuum (10^{-8} – 10^{-9} Pa) to prevent the adsorption of gas molecules on the surface of the tip. This leads to a reduction of the emission current and to instabilities caused by an increase of the workfunction due to the surface contaminants. To reduce this sensitivity to adsorption, thermally assisted field emission and Schottky emission sources have been introduced. Thermally assisted FEGs are based on the application of a high electric field to W single crystal tips heated to about 1600 K. The emission also occurs through tunneling from a small area of the tip and the characteristics are similar to cold field emission. The benefits of high temperature operation is the increased stability due to a cleaner tip at the expense of a higher energy spread of the emitted electrons as compared to cold FEG. The last type of source also considered in the class of FEGs is the Schottky gun. These guns are based on the Schottky emission principle that causes a reduction of the energy barrier for simple thermionic emission with a combination of electron image forces and strong electric fields applied to the tip (e.g., see Reimer, 1984; Solymar and Walsh, 2004). Strictly speaking, the emission is therefore not due to tunneling as it combines thermal emission with high electric fields. Due to the high coherence and brightness, however, Schottky sources are also considered in the class of field emission sources. To further decrease the workfunction, the W tip is generally coated with ZrO to increase the emission current. Optimal current stability is obtained with W crystals oriented so as to expose (100) crystalline facets and ZrO coating. The high operating temperature reduces the sensitivity to adsorption but it has the drawback that the material is sensitive to reactions with the gases present in the gun area. Ultrahigh vacuum is therefore required but at less strict levels than what is necessary for cold FEG. High current densities can be achieved with Schottky sources due to the very small emission area. High total emission currents can also be generated by controlling the extraction voltage and the gun lens operation parameters. In some implementations of the Schottky guns, total beam currents in the order of 100–300 nA can be achieved.

The previous description has been very qualitative and further analysis is required to effectively compare the performance of the various types of sources and understand the requirements for AEM. A key quantity characterizing the gun performance is the brightness of the source B defined as the current per unit area and solid angle

$$B = \frac{i_e}{\pi(d/2)^2 \pi\alpha^2} = \frac{4i_e}{(\pi d\alpha)^2} \quad (1)$$

where i_e is the current emission, d is the beam diameter, α is the convergence angle of the cone containing the electrons, and $\pi\alpha^2$ is the solid angle corresponding to the cone. The units of B are A/(m²sr) and the values scale with the accelerating voltage (B values are typically given at 100 keV) and are maintained throughout the optical system (from the emission source to the sample).

The emission current density J for thermal sources is related to the material and the operation temperature T and follows the Richardson's law

$$J = AT^2 \exp(-\phi/KT) \quad (2)$$

where A is a constant of the material (the Richardson constant given in units of $\text{A}/\text{cm}^2\text{K}^2$), and ϕ is the workfunction. The field emission current is related to the tunneling process and thus to the electric field at the tip. For Schottky emission, the workfunction is effectively reduced by a factor $\Delta\phi$ due to the strong electric field applied to the tip and becomes $\phi_{\text{eff}} = \phi - \Delta\phi = \phi - e\sqrt{eE/4\pi\epsilon_0}$ where E is the electric field (in the order of 10^8 V/m), ϵ_0 is the vacuum permittivity, and e is the electron charge. This reduction of the workfunction arises from the well-established Schottky effect (e.g., Solymar and Walsh, 2004) that effectively increases the emission current by a factor $\exp(\Delta\phi/kT)$ (Egerton, 2005) due to the negative term in the exponential factor in Richardson's law [Eq. (2)]. The temperature of the source will affect the temporal coherence of the source and the energy spread of the electrons due to thermal energy. The source size and the angular spread α will affect the spatial coherence. Considering the source size, the temperature of the source, the vacuum requirements, and the brightness, the sources can be compared (Table 4–3).

It is possible to comment on the merits of the various electron sources for AEM. The high brightness of the cold FEG makes it the ideal tool for analytical work requiring a small probe. Minimal source demagnification is required as the source size is in the order of 10 nm although the ultimate probe size and shape will depend on electron optical considerations related to aberration of the probe forming optics (see below). An additional advantage of the cold FEG is the low energy

Table 4–3. Comparison of sources.

Characteristic	Thermionic W hairpin	Thermionic LaB ₆	Schottky emission FEG (ZrO on W crystal)	Thermal FEG W (100) orientation	Cold FEG (W single crystal)
ϕ (eV)	4.5	2.7	2.8	4.5	4.5
J_c (A/m^2)	$\approx 10^4$	$\approx 10^6$	$\approx 10^7$	$\approx 10^7$ – 10^8	$\approx 10^9$
α (rad)	$\approx 10^{-2}$	$\approx 10^{-2}$	$\approx 10^{-3}$	$\approx 10^{-3}$ – 10^{-4}	$\approx 10^{-4}$
d (μm) ^a	≈ 50	≈ 10	≈ 0.02 – 0.03	≈ 0.01 – 0.1	≈ 0.01
Brightness [$\text{A}/(\text{m}^2\text{sr})$] ^b	$\approx 10^9$	$\approx 10^{10}$	$\approx 10^{11}$ – 10^{12}	$\approx 10^{12}$	$\approx 10^{12}$ – 10^{13}
Temperature (K)	2700	1800	1800	1600	300
Energy spread (eV)	1–3 ^c	0.6–2 ^c	0.6–1.5 ^c	0.6–0.8	0.3
Vacuum (Pa)	10^{-2}	10^{-4}	10^{-7}	10^{-7}	10^{-8}

^a Specific values depend on the exact tip configuration.

^b The brightness values are given for 100-keV electrons and scale linearly with accelerating voltage. The reduced brightness is given by B/V (where V is the accelerating voltage in volts).

^c The energy spread is dependent on the operation condition of the gun (the bias of the Wehnelt, filament heating current, or the extraction voltage of the Schottky FEG).

spread of the electron source allowing, in principle, detailed analysis of the energy loss near-edge structures in EELS spectra. Although the ultimate brightness of the Schottky sources is lower by about one order of magnitude, the practicalities of the less stringent vacuum and demonstrated reliability have made the Schottky guns very common within the AEM field. Worth mentioning is the fact that the total current might be a more important parameter for some AEM analysis techniques such as energy-filtered imaging when a large field of view needs to be illuminated (with enough current per individual pixel in a map). In that case, the thermionic emission sources (LaB₆ in particular) provide very large total currents. New research is in progress to develop alternate even brighter sources by exploiting low workfunction materials, carbon nanotube electronic properties (Fransen et al., 2005), and semiconductor p–n junctions.

Once the electrons are accelerated by the anode stack to the high-voltage potential of the TEM, the condensing optics demagnifies (Figure 4–12) the source image in the plane of the sample. In this chapter we will not describe the various aberrations and defects of lenses but refer the reader to other chapters in this book. We remind the reader, however, that all round electromagnetic lenses have positive spherical aberration coefficients C_s (in the order of $C_s = 0.5\text{--}1.0\text{ mm}$ for high-resolution microscopes offering limited sample tilt and a small polepiece gap and around $C_s = 1.5\text{--}3\text{ mm}$ for large tilt and large polepiece gaps). Electromagnetic lenses also suffer from chromatic aberrations C_c in the order of $C_c = 1\text{--}2\text{ mm}$ and that, just as in light optics, the wave nature of elec-

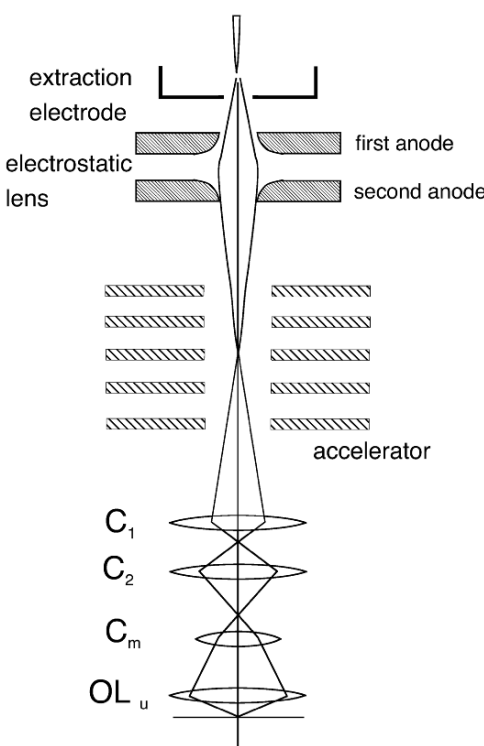


Figure 4–12. Schematic diagram of a field emission electron gun with gun anodes (also called gun lens) and the accelerating anodes (accelerating the electrons up to the accelerating voltage of the TEM). The illumination system consists of condenser lenses (C_1 , C_2 , C_m) and the objective lens (OL). See the text for function details.

trons imposes a diffraction limit to the resolution as imposed by the Rayleigh criterion. Given these imperfections of the lenses and wave nature of electrons, the probe size at the specimen level d_0 is limited not only by the demagnified image source size d_s but also by a combination of the aberration of the illumination system (spherical and chromatic) and the diffraction limit of the probe-forming aperture all added in quadrature (Spence and Zuo, 1992; Reimer, 1984)

$$d_0^2 = d_s^2 + d_d^2 + d_{sa}^2 + d_c^2 + d_f^2 \quad (3)$$

where d_s is the demagnified source image, $d_d = 0.6\lambda/\alpha$ is the diffraction limit broadening, $d_{sa} = 0.5 C_s \alpha^3$ is the spherical aberration broadening, $d_c = (\Delta E/E_0) C_c \alpha$ is the chromatic aberration contribution of the probe forming lens, $d_f = 2\alpha \Delta f$ is the defocus contribution to a small defocus error where α is the convergence half-angle at the specimen determined by the condenser C_2 aperture limiting the probe and Δf is the defocus of the probe forming lens. The size of the unaberrated probe source image d_s is related to the definition of brightness that, as mentioned above, is constant through the illumination system (from the gun to the sample) and is determined as $d_s = (2/\pi\alpha)\sqrt{I_b/B}$ where I_b is the beam current. Calculations of the unaberrated source size therefore assume that the current at the sample is known (or can be measured).

Equation (3) assumes incoherent illumination, i.e., all electrons at the aperture are not in phase and every point within the aperture can be considered as an independent source. Strictly speaking, this assumption is not valid for high brightness field emission sources (and for particular operating conditions of thermionic electron sources) and it is necessary to define the conditions where incoherence in illumination applies to determine the conditions of validity of Eq. (3) for the calculation of the probe size. Due to the various aberration terms in Eq. (3) there is an optimum probe size as a function of the illuminating angle. Plots for two microscope configurations and sources are shown in Figure 4–13 for a high-resolution microscope configuration—not dedicated to analytical work but for designed optimum imaging resolution (low tilt angle in the specimen area: ± 20 – 24°)—and for a high-tilt analytical configuration (± 35 – 40°). It is clear that the use of FEG allows smaller probes. Based on the equations above and typical gun brightnesses for Schottky guns, it is possible to calculate the probe current dependence with probe size, calculated based on the incoherent illumination condition, for an FEG and an LaB₆ source (Figure 4–14). We also demonstrate the effect of reducing the spherical aberrations in the probe-forming lens using the recently developed correctors in a cold FEG-scanning transmission electron microscope (STEM) operating at 100 kV based on the data reported in the literature (Dellby et al., 2001) (Figure 4–15). For the latter case, Eq. (3) needs to be modified to account for the limited residual aberrations (see Chapter 2, this volume; Dellby et al., 2001).

As discussed above, the coherence of the source is a key concept that needs to be considered for high-brightness field emission sources and analytical microscopy. In conditions of coherent illumination all electrons can be considered to be emitted by a single point source and all

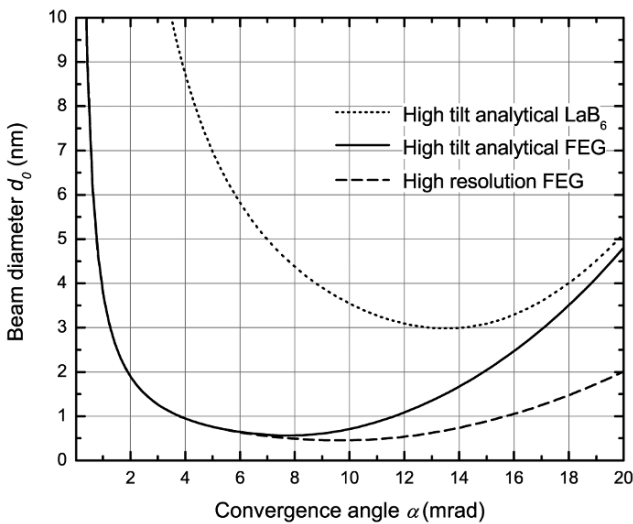


Figure 4-13. Variation of the electron probe size for analytical TEM work as a function of the illumination aperture convergence for instruments equipped with different electron sources and lenses.

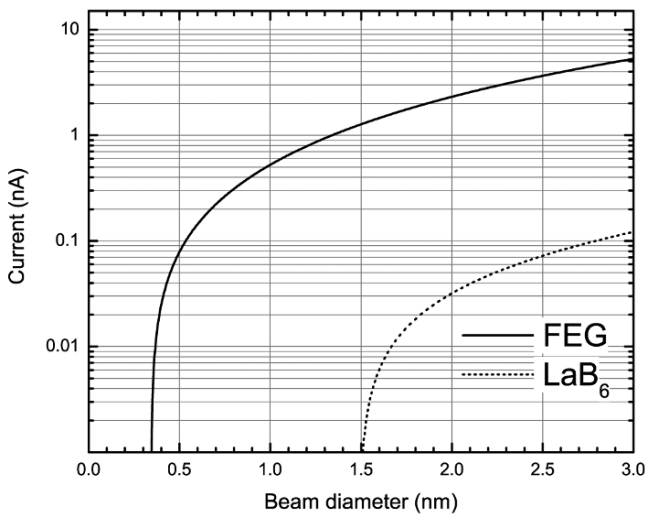


Figure 4-14. Variation of the electron beam current with electron probe size for two instruments equipped with a thermionic source and Schottky FEG.

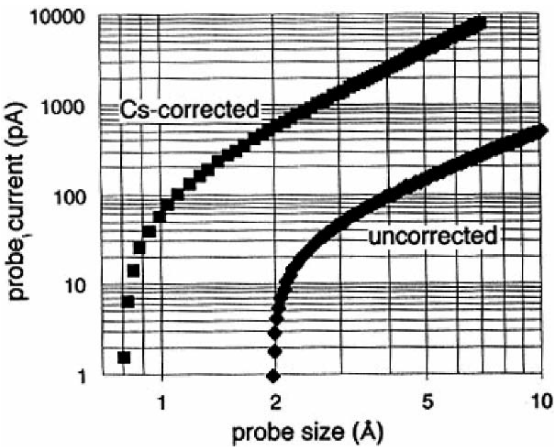


Figure 4-15. Variation of the probe current as a function of probe size for a conventional cold field emission source dedicated scanning transmission electron microscope and an aberration corrected cold-FEG microscope. (From Dellby et al., © 2001. Reprinted with permission from the Japanese Microscopy Society.)

are in phase in the illuminating aperture and on the electron wavefront at the sample. This condition implies that interference effects between electrons can occur. A detailed discussion on coherence and sources can be found in Spence and Zuo (1992) and we summarize the key elements of coherence relevant to the discussion on the probe size for the purpose of discussing the ultimate limits of the instrumentation. Following the concepts developed in classical optics, it is necessary to consider the transverse coherence width X_a concept related to the width of the electron beam in the plane of the condenser aperture $X_a = \lambda/(2\pi\theta_s) = f_{C_2}\lambda/(\pi d_s)$ where θ_s is the angular width of the probe of size d_s (i.e., not the convergence angle) as subtended at the condenser aperture (Figure 4-16). This angle can be determined geometrically given a probe size d_s and focal length f_{C_2} of the probe-forming lens. Source points in the aperture plane closer than X_a can interfere as in a slit experiment with coherent light. This transverse coherence term must be compared with the diameter of the probe-forming aperture D_{C_2} to determine if the illumination is coherent or not. If $X_a \ll D_{C_2}$, the aperture is incoherently illuminated and all points in the illumination aperture can be considered as emitting independently. Still, for a given aperture of convergence angle α incoherently filled, the coherence width will be $X_s = \lambda/(2\pi\alpha)$. Object points at the sample separated by a distance smaller than X_s will be illuminated coherently and will give rise to interference effects.

In the other extreme, if $X_a > D_{C_2}$, the aperture is coherently filled and the electron wavefront illuminating the sample will be in phase. For example, for the smallest probe in an FEG (say 0.2 nm), a C_2 aperture of 5 μm would be coherently filled. At 200 keV, for a LaB₆ emitter and a probe size of about 3 nm, one would need an aperture of 0.5 μm for the coherence criterion to be fulfilled. By changing the illumination conditions (demagnification of the source and reducing the accelerating voltage) and the condenser aperture, the coherence illumination condition could be achieved even for thermionic sources as demonstrated by Dowell and Goodman (1973).

In coherent illumination conditions, the terms contributing to the determination of the probe size cannot be added incoherently as inde-

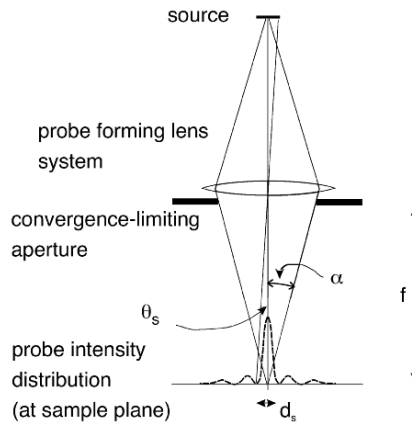


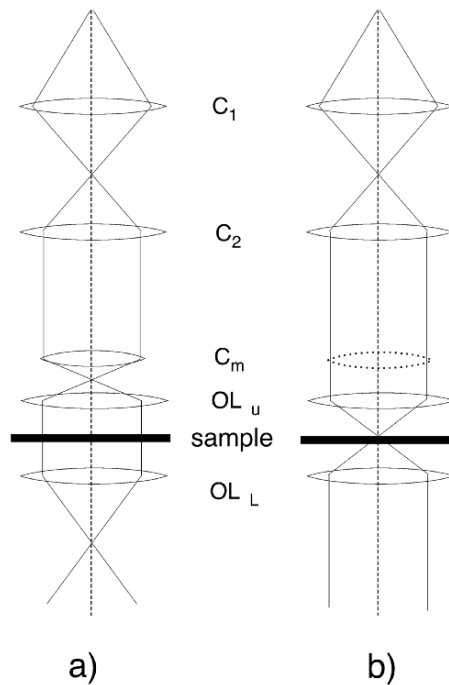
Figure 4-16. Factors entering the description of the source coherence for illumination of samples. (Adapted from Spence and Zuo, 1992.)

pendent contributions [as done in Eq. (3)] and detailed calculations based on the incident electron wavefunction and the transfer function of the objective lens must be carried out. It has been demonstrated that with the judicious choice of defocus of the probe forming lens, sub-Ångstrom electron probes can be obtained even without aberration correctors (Nellist and Pennycook, 1998) to improve imaging resolution albeit with very small currents and tails to the intensity distribution containing up to 50% of the beam current. Such beams, although relevant for optimal imaging resolution, are therefore not suitable for analytical work. Optimum probe sizes (Mory *et al.*, 1987) for analytical work are achieved with convergence α_C and defocus Δf conditions determined by $\alpha_C = 1.27 C_s^{-1/4} \lambda^{1/4}$ and $\Delta f = -0.75 C_s^{1/2} \lambda^{1/2}$. In these optimal conditions, the probe size containing 80% of the intensity is $d_{(80\%)} = 0.4 C_s^{1/4} \lambda^{3/4}$. For aberration corrected instruments these terms need to be revised to correctly treat the contributions of the residual aberrations as discussed in Dellby *et al.* (2001) and in Chapter 2 (this volume).

2.2 Electron Optic Configuration

In the discussion so far we have made abstraction of the technical aspects necessary to achieve the demagnification of the source required for small probe analysis and the need to provide, with the same system, illumination at the sample for conventional imaging and energy-filtered microscopy. Modern instruments capable of conventional TEM and STEM are based on the double condensing optic system ($C_1 + C_2$) with the addition of a supplementary weaker condenser lens C_m (called “condenser minilens” or “minicondenser”) and a strong magnetic field before the sample (i.e., a prefield) generated by an objective lens (OL) composed of two parts: an upper lens and a lower lens surrounding the sample (Figure 4–17). “Parallel” illumination at the sample plane is achieved by a combination of C_2 and C_m yielding, at the focal point of the objective lens, a convergent beam that is then made parallel by the strong upper objective lens field (Figure 4–17a). A small probe required for analytical work or for STEM imaging and analysis in STEM mode is obtained by effectively optically switching the C_m off (using various schemes depending of the exact location of the C_m lens) and by making use of the strong OL prefield (of the upper objective lens) to achieve large convergence angles and large source demagnification (Figure 4–17b). Practically speaking, this mode of operation is called the nanoprobe or energy-dispersive spectrometry (EDS) mode. STEM operation for these microscopes is achieved in exactly the same electron-optical configuration. FEGs (cold and Schottky) have an additional electrostatic gun lens (Figure 4–12) leading to additional demagnification of the source and more flexibility in the gun operation. The strength of the C_1 lens determines the fraction of electrons (hence the current) that will enter the C_2 aperture and the demagnification of the source. A strongly excited C_1 lens will give rise to larger demagnification and, conversely, a weakly excited C_1 lens will result in a smaller demagnification and more current entering the C_2 aperture. In this fixed operation mode, changes in the convergence angle are achieved

Figure 4–17. Electron optical configuration of the illumination system of a double-objective lens system. Configuration (a) describes the formation of parallel illumination while configuration (b) describes the conditions needed to achieve a highly convergent and small focused electron beam.



by changing the physical aperture size (using a strip aperture containing four to eight apertures, including a top hat Pt aperture for analytical work—see Sections 2.3.6 and 7.2 on instrumental contributions in EDXS analysis). Continuous change in convergence is achieved by controlling the strength of the C_m lens and OL prefield or/and addition of a third condenser lens.

The electron optical configuration required for STEM imaging and analysis is achieved, in a TEM-STEM instrument capable of both operation modes and, as discussed above, with the combination of the C_m lens optically switched off and the subsequent focus of the nearly parallel beam into a small source image by the upper OL field. The scanning operation of the beam over the sample is carried out by deflection coils located before the specimen (optically before the upper objective lens) so that the beam is shifted (but not tilted) on the specimen plane (Figure 4–18). As the beam is rastered pixel by pixel over the area of sample of interest, various signals (including analytical information) can be recorded sequentially at each position to form images and elemental maps based on analog or digital signals recorded synchronously (e.g., bright-field, secondary electrons, backscattered electrons, annular dark-field electrons, and EELS, EDXS, etc). Dedicated commercial STEM instruments (offering no TEM operation mode) built by the company Vacuum Generator in the 1980s and 1990s have traditionally been equipped with cold FEG and, in the later models, operate with a gun lens, two condenser lenses, and an asymmetric objective lens (no imaging lenses are necessary). Their design is based on the gun located at the “bottom” of the microscope column for stability reasons with the detectors at the “top.” New commercial instruments

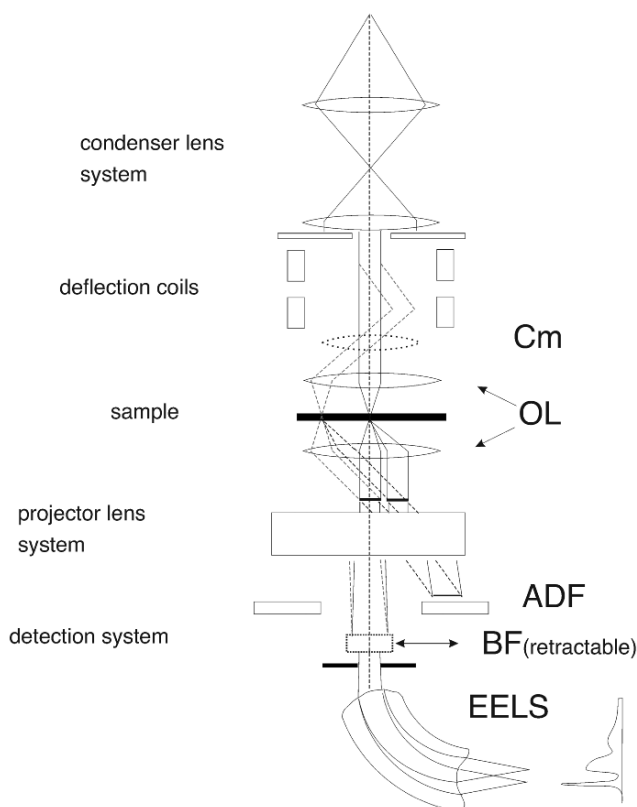


Figure 4–18. Electron optical configuration for STEM operation showing the role of the deflection coils in shifting the electron beam over the sample (without tilt of the beam) and the detection of signals on the dark-field STEM detectors and bright-field or energy-loss spectrometer. (Adapted from Gross et al., 1987.)

developed based on the same approach and equipped with aberration correctors are in development (Krivanek, 2005). Dedicated STEM instruments based on the upper portion of a conventional TEM column (but in this case without imaging lenses) are also available. These instruments are designed for quick and simple operation particularly popular for routine analysis in device fabrication environments.

As supplementary information, we should briefly note that the processes of image formation and interpretation in STEM mode and TEM mode are closely linked by the reciprocity theorem further discussed in Chapter 2 (this volume) and earlier references (Cowley, 1986; Humphreys, 1979). This principle links the electron source in STEM to a detector point in the TEM image (pixel on a camera or negative plate) and the detector in STEM to the source of electrons in TEM. Effectively, the principle states that for identical optical components, sources, and detectors, STEM and TEM images will show the same resolution, and contrast. Given the respective strengths and weaknesses of these techniques (related to the field of view, recording time, dose, sequential recording, and analytical signals) the techniques should be considered

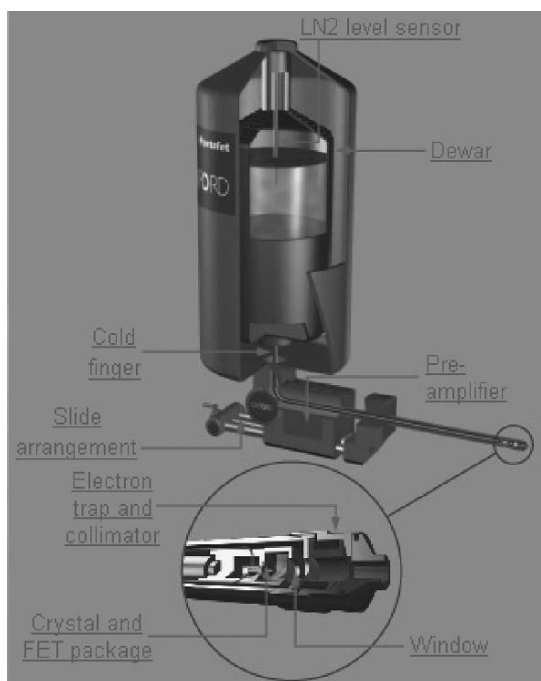
as complementary and not competitive (Cowley, 1986). Practically, this implies that materials analysis, and AEM in particular, should involve all possible optimized techniques and instrumentation both for TEM and STEM approaches. As demonstrated in the case of EELS analysis, energy-filtered imaging carried out in the TEM mode is highly complementary to the scanning EELS imaging method with both approaches offering advantages and presenting limitations (Section 6).

2.3 EDXS Detector Systems

2.3.1 The EDXS Detector

In AEM experiments, photons are emitted in the sample following ionization of atoms by the primary incident electrons and subsequent deexcitation process. The energy of these photons is in the X-ray part of the electromagnetic spectrum (a few hundred electronvolts to a few tens of kiloelectronvolts) and we therefore refer to them as X-rays. In the analytical TEM, the most common and effective tool for the detection of X-rays is the energy-dispersive detector that is attached to the microscope column (refer to Figures 4-2 and 4-19) with the active component detecting the X-rays located as close as possible to the sample. Alternative approaches to detect X-ray signals with wavelength dispersive detectors have been attempted in prototype systems but have not been commercially implemented due to low efficiency (serial analysis of X-rays energies) and compatibility with the geometry of the microscope column (solid angle, vacuum etc.). The key component of the energy-dispersive X-ray detector is a semiconductor material (Si or Ge) that absorbs incident photons by generating electron hole (e-h)

Figure 4-19. Schematics of a commercial EDXS detector showing the detector front, the Dewar system to cool the detector, and various components that are interfaced in the electron microscope. (Courtesy of N. Rowlands, Oxford Instruments.) (See color plate.)



pairs through photoelectric effect. When this process occurs in an electric field, a current pulse is generated and subsequently measured by low-noise electronic components. Detectors based on Si are more commonly used in AEM because of the lower cost than Ge-based detectors (described below). Due to the fact that very high purity Si crystals are not available, Li additions in Si are used to compensate for residual impurity dopants that would create undesirable recombination sites for electron-hole pairs and uncharacteristic current related to the existence of impurity-generated acceptor levels (thus holes in the valence band). The Li role in the so-called Si(Li) detectors is therefore to effectively create an intrinsic semiconductor region where e-h pairs generated by the incident photons produce a current flow that can be measured. Incident photons generate e-h pairs at the rate of 1 pair/3.8eV. The number of e-h pairs generated in the “active layer” of the detector is thus proportional to the incident photon energy (which is, in turn, related to the energy level involved in the transitions following ionization of the atom by the incident electron beam). For Ge detectors, high-purity crystals can be obtained and Li additions are not necessary.

The electric field in the detector necessary to cause the current flow is generated by metallic contacts (Au) (Figure 4-20) on the front and back surfaces of the crystals (typically 3 mm thick) and the application of a bias (0.5–1 kV) between the front and back of the detector with electrons traveling to the positive electrode and holes to the negative electrode. Some loss of charges occurs, however, in proximity of the contacts due to recombination in the charges near the metallic contact (about 200 nm in width). These regions are typically referred to as “dead-layers” in the literature as they do not contribute to the generation of signal and transfer of charge. The thickness of the crystal is normally large enough to convert all the photons to e-h pairs but for

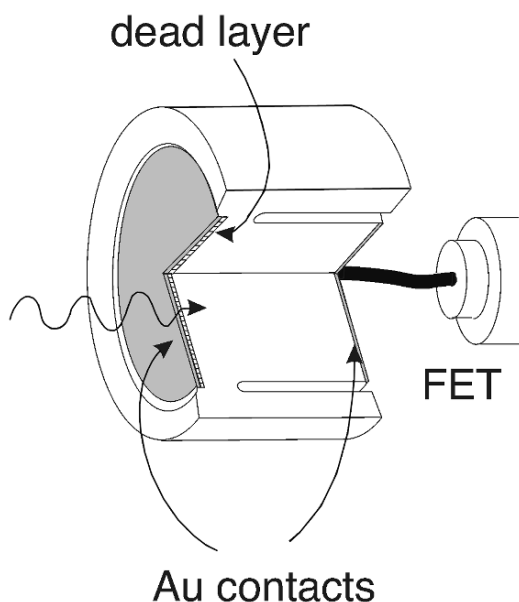


Figure 4-20. Diagram of an EDXS detector demonstrating the various elements of the detector contacts, the dead layer, and the preamplifier. (Adapted from Woldseth, 1973.)

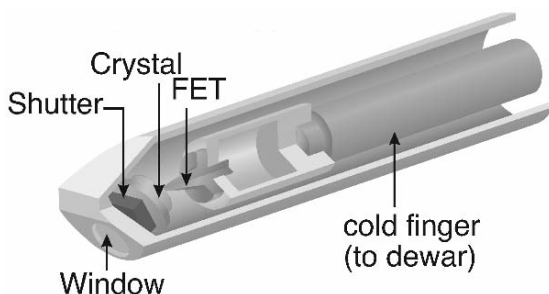


Figure 4–21. Schematics of the detector front of a TEM with components inserted into the TEM column detector window. (Courtesy of N. Rowlands and A. Kirk, Oxford Instruments.)

high-energy X-rays, transmission through the detector is possible. Measurement of the current is carried out from the back electrode, which is connected to a field effect transistor (FET) that acts as a first stage of amplification.

The detector and FET are cooled to low temperature by a Cu rod connected to a liquid nitrogen reservoir (Figure 4–19) to prevent the diffusion of Li in the electric field and to reduce the thermal noise of the e–h generation and the FET electronics. The mechanical design of the detector assembly is crucial to reduce the transfer of mechanical vibrations into noise in the spectra. For example, bubbling in the Dewar arising from floating ice crystals and vibrations from the microscope frame or other components (such as fans) touching the detector can result in additional noise. The detector should therefore be supported by the same support mechanism as the frame of the microscope column and be isolated from the rest of the microscope components.

2.3.2 Detector Windows

Due to the low operating temperature, the detector can act as a cold trap attracting contaminants to the detector front. For this reason, detector windows are used to isolate the detector from the vacuum of the microscope (Figure 4–21). The window technology has evolved in recent decades with the initial technology based on Be windows (7–8 μm in thickness) allowing X-rays for elements down to Na to be detected due to the absorption of lower energy photons in this material. This limitation for the detection of light elements has led to the development of polymer-based thin windows and some implementation of systems without windows at all (called windowless detectors). Various polymer-based window materials are available in ultrathin windows (UTW). These use proprietary technology combining light element composites (polymers, diamond, nitrides, etc.) some of which are strong enough to withstand atmospheric pressure (labeled as atmospheric thin windows—ATW). Due to the combination of various X-ray absorbing materials the sensitivity of the detector system for light elements strongly varies at low energy due to the absorption the “soft” X-rays into the window material. UTWs, for example, allow detection of elements down to B but with reduced sensitivity as compared to

windowless detectors that are actually able to detect X-rays down to the BeK α line. The latter technology, however, normally requires the use of ultrahigh vacuum in the microscope to prevent rapid contamination of the crystal by the microscope environment and the related absorption of soft X-rays by the ice contamination layer. Whether windowless or UTW, most detectors are equipped with crystal heaters that gently warm up the crystal surface causing desorption of the ice layer built up on the crystal surface. Ice contamination is revealed by a reduction of the intensity of soft X-ray lines as compared to higher-energy lines such as demonstrated by the NiL/NiK ratios measured as a function of time. Since the Ni-L line at 0.85 keV is much more strongly absorbed than the NiK (at about 7 keV) set of lines the ratio is an effective means to appreciate the contamination effect (L'Espérance *et al.*, 1990).

The detector efficiency (Figure 4–22) accounts for all these effects and represents the fraction of X-rays that is transmitted through the window system as compared to the incident intensity as a function of energy. Windowless detectors still show an important drop in efficiency at low energy due to the presence of the dead layer and metallization in front of the intrinsic active portion of the detector. Even for UTW and ATW detectors, the efficiency drops significantly for low-energy X-rays resulting in difficulties for the analysis of light elements in low concentration. Discontinuities in the detector efficiency are visible at the energies corresponding to the absorption edges of the elements contained in the window material (e.g., C, O, B, N, for example), the detector, and metallization. As discussed above, metallization is required for the application of a bias on the semiconductor crystal but it is also necessary on the window material to prevent light (generated by some samples by cathodoluminescence) from entering the detector system and resetting the signal amplification system.

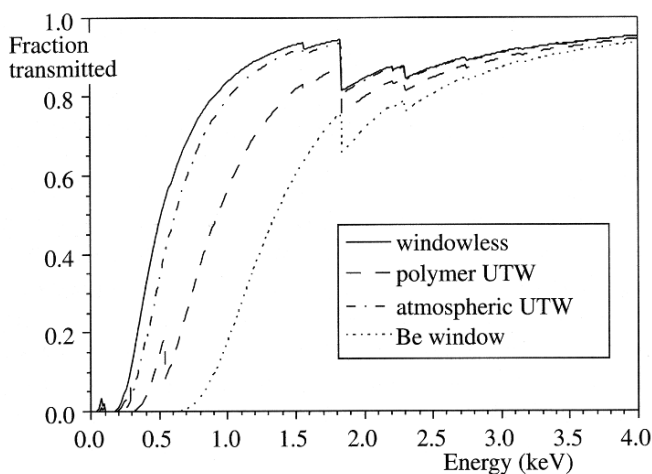


Figure 4–22. Detector efficiency curves for various materials used as detector windows. (From Williams and Carter, © 1996, with permission from Springer Science+Business Media.)

In addition to the fact that high-purity Ge (HPGe) crystals can be produced and no Li additions are required, Ge-based detectors offer the advantage that the absorption of X-ray is stronger (1 e-h pair/2.9 eV) and higher-energy lines can be analyzed as the related high-energy X-rays are not transmitted through the crystal. The width of the X-ray peaks is narrower for Ge detectors than Si(Li) detectors leading to better sensitivity and less overlap in measurements. Some AEM systems are therefore equipped with a combination of both Si(Li) and HPGe detectors for a more efficient analysis of X-rays from a larger range of elements.

2.3.3 Signal Processing

The e-h pair-generated current is detected by the FET as a pulse signal that is subsequently fed into a main amplifier system as a voltage. The sequence of pulses, separated by a time interval, generates a staircase signal where each step represents a photon arrival and the height is linked to the energy of the photon. After the integrated signal reaches a threshold level, the FET must be reset to a base value by the means of a light pulse generated by a light-emitting diode in a "optoelectronic feedback" system. This is necessary to avoid saturation of the signals. Each step rise lasts in the order of 150 ns. Pulses can be amplified and shaped for subsequent analysis (to determine exact height and thus photon energy) with analog technology. Analog systems give the user flexibility on the process time of the pulse and thus accuracy in the signal analysis. High processing speeds of pulses (in the order of few microseconds process time per pulse) result in low energy resolution of the peaks due to the uncertainty in the pulse height and thus the energy of the photon. Low processing speed (about 50 μ s/pulse) results in more accurate determination of the pulse height and more accurate determination of the X-ray energy. During analysis of the pulses, the detector is effectively not able to process more photons entering the detector resulting in analysis dead time, which represents the time the detector is not processing signals. Due to this limitation, the output count rate is not linear with input count rate at high X-ray fluxes. Count rates imposing detector dead times in the order of 60% are acceptable for modern systems and exhibit a nearly linear response. Above 60% dead time, there is a drop in the output rate with an increase in the input rate. Therefore, with thick samples and thus large photon fluxes, high processing speeds are required to reduce the process time and the resulting dead time. Recent developments have allowed much faster pulse processing with digital technology resulting in higher throughputs of signals and linear response of the system with respect to the input count rates. With digital technology, the voltage rise output from the FET is directly digitized and can subsequently be processed with numerical pulse processing techniques leading to reduced noise and better high-count rate responses. Details of the various tests and procedures to determine linearity of the system response and examples are given in Williams and Carter (1996) and Goldstein *et al.* (2003).

2.3.4 Peak Shapes

The intrinsic width of an X-ray emission line is in the order of 1–2 eV. The width of the X-ray peaks as processed by the detector system,

however, depends on the generation of e-hole pairs and the noise introduced during the measurement process. A peak in the spectrum represent the distribution of X-rays “detected” at a given energy with each incident photon generating a variable number of e-h pairs due to statistical fluctuations in the generation process. The standard deviation of the number of e-h pairs produced is one of the factors affecting the width of the peaks. The noise of the detector system and detector collection artifacts, however, also contribute to the broadening. The broadening due to the statistical generation of e-h pairs is given by $\Delta E_s = 2.35\sqrt{\epsilon FE}$ with ϵ representing the energy for e-h pair creation (3.8 eV in Si, 2.9 eV in Ge), E the energy of the X-ray peak, and F a parameter representing the statistical correlation in the e-h pair generation process known as the Fano factor, which varies between 0 and 1. $F = 1$ if there is no correlation between the e-h generation events and $F = 0$ if the processes are completely deterministic (the process is completely reproducible and yields the same result time after time). Noise and artifacts also contribute to the broadening ΔE_N yielding a total broadening of a Gaussian peak distribution $\Delta E = \text{FWHM}$ (full width at half maximum) of the X-ray peaks as

$$\Delta E^2 = (\Delta E_s)^2 + (\Delta E_N)^2 = (2.35)^2 \epsilon FE + (\Delta E_N)^2 \quad (4)$$

For Si(Li), $F = 0.12$, and the FWHM of the MnK α line (used as a reference for resolution because of availability of radioactive standards producing MnK α peaks) is around 138–140 eV. If the noise contributions were completely removed, the theoretical resolution is solely based on the e-h generation process and would be around 110 eV for the MnK α peak. The energy width decreases at lower energy with FWHM below 100 eV for light elements. Since the noise term is not linear with energy and depends on the processing time of the amplification system, a full prediction of the peak resolution depends on the operating conditions and energy. Reference values given for the MnK α lines are used to compare the electronics and detector performance and are usually given at optimum process time with typical values around 140 eV FWHM at the MnK α line. Improved energy resolution is achieved even for high count rates on digital detectors and with HPGe detectors due to the lower ϵ values (FWHM for MnK α is around 120 eV). Noise introduced by vibrations (e.g., mechanical coupling with environment and/or ice crystals floating in the Dewar) can also contribute to peak broadening, hence lowering the spectral resolution, and should be minimized.

2.3.5 Detector and Signal Processing Artifacts

A summary of detection artifacts is presented here with further details given in more extended reviews (e.g., Goldstein *et al.*, 2003; Williams and Carter, 1996). In perfect detector conditions, the peak shape is expected to be Gaussian but small distortions can arise if the generation of e-h pairs is perturbed. For example, recombination of e-h pairs in the dead layer or at lattice defects generated by high-energy incident electrons accidentally entering the crystal can give rise to a phenomenon known as *incomplete charge collection*. This effect gives rise to

low-energy tails in the peak distribution as not all the e-h pairs are collected.

Incident X-rays can cause fluorescence of the $\text{SiK}\alpha$ line (or a Ge line). If $\text{SiK}\alpha$ photons are not absorbed within the detector and exit the active area, incident photons will have lost a fraction of their energy in this process equivalent to the $\text{SiK}\alpha$ ionization energy (1.74 keV). This will cause an “escape peak” in the spectrum at an energy $E_{\text{es}} = E - E_{\text{SiK}\alpha}$. This effect is particularly important when small trace elements are investigated since there is potential overlap between E_{es} and X-ray lines (for example Fe overlaps with the escape peak of the $\text{CuK}\alpha$ line).

If the count rate is high, there is the possibility that two incident photons of energy E will be perceived by the pulse counter as one single photon of energy $2E$. This effect, known as a “sum peak,” is visible when count rates are above the reliable limit of the system (which of course varies depending on the processing technology). High count rates, leading to dead times greater than around 60%, are likely to lead to sum peaks.

Internal fluorescence peaks can be also detected if the incident photons generate a Si (or Ge) $\text{K}\alpha$ peak in the dead layer of the detector, which is subsequently detected in the active area of the detector. This effect is small but can, once again, be significant for trace analysis.

High-energy incident electrons can also generate spurious signals and damage the semiconductor crystals. The location of the detector and the operation of the microscope should be such that these contributions are minimized (for example, objective apertures must be removed during acquisition).

2.3.6 Geometry of the EDXS Detector in the AEM

To optimize the solid angle and thus the collection of the X-ray radiation generated by the incident electron, the detector is placed as close as possible to the sample area (Figure 4–23). The Si(Li) detector active area A is typically 10 mm^2 with recent systems as large as 30 mm^2 . For a detector positioned at a distance R with respect to the optic axis (and the origin point of the emission) the solid angle $\Omega = A/R^2$ (measured

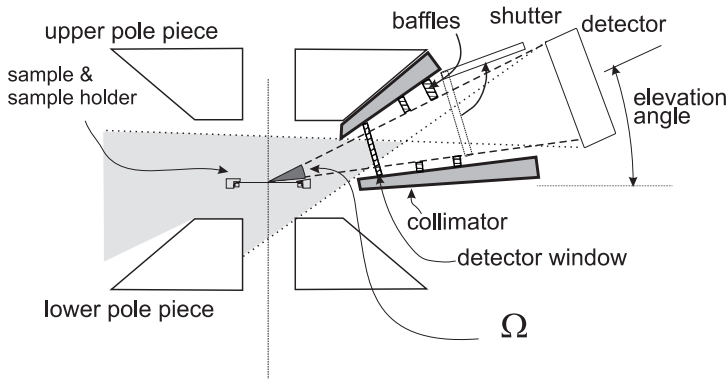


Figure 4–23. Interface of the detector with the microscope sample area. (Adapted from Otten, 1996.)

in steradians) is the key parameter determining how effective the system collects emitted X-rays. For optimal solid angle, the detector normal is in direct line of sight to the emission point and not tilted away from it. Typical solid angles in current AEM are 0.13sr but with combinations of large detector areas and effective coupling with the microscope specimen area, solid angles in the order of 0.3sr have been achieved. For $\Omega = 0.13\text{sr}$ the fraction of collected X-rays with respect to the full emission solid angle is only 1%! The detection of X-rays is therefore a very inefficient process considering that the X-ray emission is fully isotropic.

The elevation angle (also known as the “take-off” angle in the literature) is an important parameter affecting the quantification of data through the absorption correction and the quality of the spectra. A large elevation angle minimizes the path length of X-rays into the sample (see the quantification section) and also reduces the continuum background emission, which is forward peaked. High detector elevation angles, however, are impractical in the TEM due to the fact that the detector would need to be above or within the objective lens at a large distance from the sample, resulting in even lower collection efficiency. In addition, backscattered electrons have direct sight to the detector and can cause significant contributions and potential damage to the detector. Lower elevation angles ($0\text{--}20^\circ$) allow larger solid angles and lead to an effective shielding of the backscattered electrons by the objective lens magnetic fields. This shielding is not as effective for high elevation angles.

The interest in large solid angles and the proximity of the detector to the sample lead to significant drawbacks in terms of spurious signal collection. The field of view of the detector is much larger than the sample area and X-rays generated by backscattered electrons or by fluorescence of hard X-rays generated in upper parts of the illumination area of the microscope easily enter into the detector (see Section 7.2). High-energy backscattered electrons can also enter the detector and generate additional secondary electrons/X-rays while low-energy electrons would spiral away from the detector due to the high magnetic field of the objective lens or the presence of a magnetic trap in the detector system (Figure 4–19). To reduce these effects, detectors are equipped with collimators that limit the field of view to the smallest possible area, thus preventing hard X-rays generated in the illumination system from directly hitting the detector, and contain baffles that reduce the effects of potential incident backscattered electrons that might enter the collimation system.

Many other contributions arising from stray electrons hitting the microscope components such as apertures, cold traps, the polepieces, etc. lead to increased noncharacteristic signals resulting in weak detection limits. As demonstrated in the work of Nicholson *et al.* (1982), many of these contributions can be reduced by improving the microscope and detector chamber using coatings to cover the microscope components with low atomic number materials and by improving the collimation system. These effects can be minimized in systems using the precautions discussed in Section 7.2.

2.4 EELS

2.4.1 Spectrometers

The measurement of energy losses suffered by the incident electrons as they exit the sample is carried out with energy loss spectrometers. These devices also make it possible to select electrons with a particular energy loss (or no loss at all) with the use of energy-selecting slits. The ability to select electrons with a particular energy loss is called energy-filtered microscopy. The technique also allows the operator to obtain images and diffraction patterns where parts of the inelastically scattered electrons are filtered out so that information deriving from the elastically scattered electrons only is used.

These instruments are based on the use of a magnetic field that modifies the trajectory of the electron according to the electron energy. The radius of curvature R_e of the electron trajectory is related to their velocity v and magnetic field strength B_f as

$$R_e = \frac{\gamma m_0}{e B_f} v \quad (5)$$

where $\gamma = 1/\sqrt{1-v^2/c^2}$ is the relativistic factor and m_0 is the rest mass of the electron. Slower electrons will follow trajectories with a smaller radius of curvature and will be dispersed on a detector plane located after the spectrometer. The dispersion refers to the separation of electron energies in space and is typically in the order of $1\text{--}2\mu\text{m}/\text{eV}$ at 100 keV . The exact location of this detector depends on the implementation of the spectrometer and its coupling with the microscope column (see below). The most common electron optical component generating the magnetic field is a magnetic sector (used in various configurations, whether the spectrometer is implemented within the microscope column or after the viewing chamber of standard TEMs). Current flow in the prism generates the required field B_f that disperses the electrons (Figure 4–24). Other approaches to filtering have been implemented in

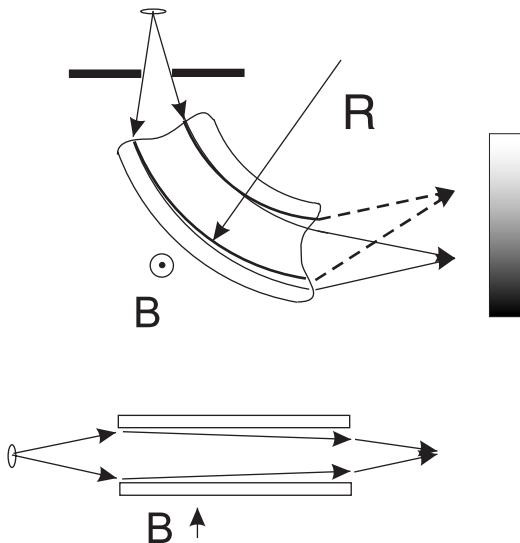


Figure 4–24. Prism spectrometer system showing the bending of electrons as they travel through the spectrometer. Dispersion of electrons according to their energy is achieved by the spectrometer in one direction and focusing is achieved in the other direction of travel.

prism-mirror spectrometers (see below) and Wien spectrometers, which combine magnetic and electrostatic fields in different configurations (Egerton, 1996).

Two types of energy filtering spectroscopy approaches should be distinguished (Figure 4–25). The so-called in-column filter/spectrometers are located within the projector lens system/postspecimen area of the microscope column. These spectrometers generate electron trajectories and dispersion that will result in the transfer of the electrons into the projector lens system and viewing chamber of the microscope. The alternative approach is realized with the postcolumn spectrometers/filters attached at the bottom of the microscope column. In the case of in-column filters, energy loss spectra and/or energy filtered images (obtained by selection of electrons of a particular value of energy loss using a slit) are realized. For postcolumn spectrometers, dedicated imaging lenses are required to generate energy-filtered images after selection of electrons of a particular energy loss. Two implementations of postcolumn spectrometers therefore exist. For acquisition of spectra only, the magnetic prism is followed by a series of optical components dedicated to increase/vary the dispersion at the detector system. For energy-filtered imaging, a more elaborated series of nonround lenses (i.e., based on multipoles) and a removable energy selecting slit are used to provide both spectroscopy and imaging capabilities.

There are various implementations of in-column filters. The earliest commercial applications were based on the electrostatic mirror-prism system initially proposed by Castaing and Henry (1962) implemented in the Zeiss microscope (Figure 4–26a). These instruments were developed on 80 keV microscopes and thus remained very popular for biological applications, although excellent fundamental electron scattering experiments were carried out on such instruments (Mayer *et al.*, 1995). The first portion of the prism is used for an initial dispersion and the

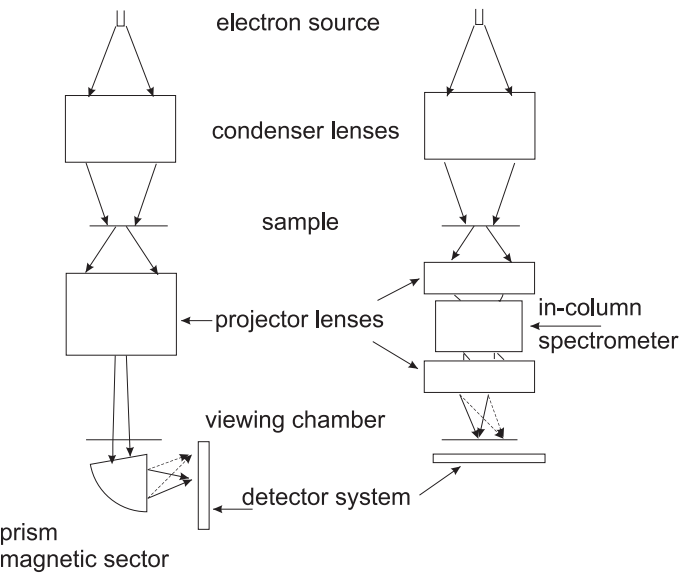


Figure 4–25. Schematic diagrams of energy-filtered electron microscopes. The postcolumn configuration (left) is based on the simple prism attached at the bottom of the microscope while the in-column configuration (right) is achieved by various components inserted in the projector lens system.

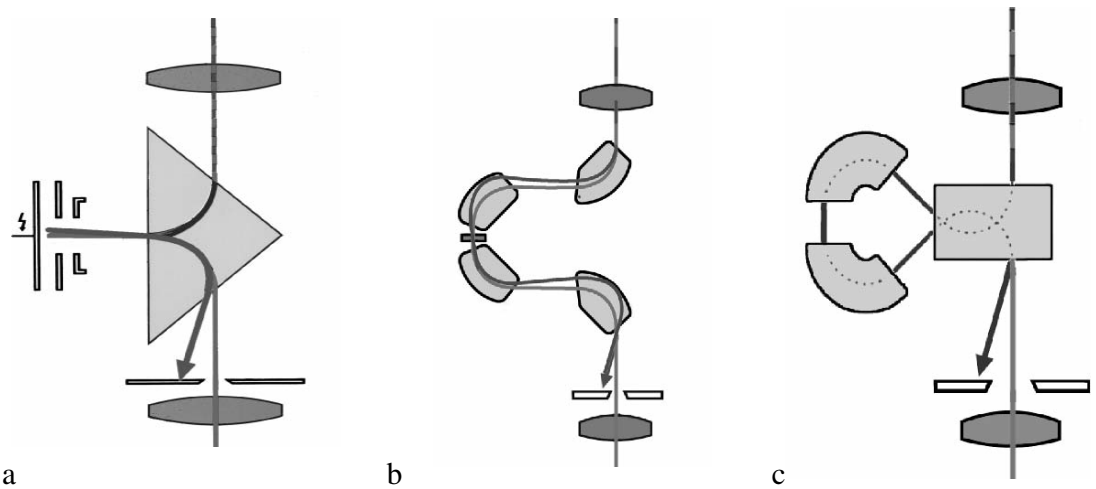


Figure 4-26. Various in-column spectrometer configurations. (a) Mirror-prism spectrometer, (b) OMEGA filter, and (c) Mandolin filter. (Courtesy of P. Schlossmacher, Zeiss SMT.) (See color plate.)

electrostatic mirror's function is to deflect the electrons back on the second section of the prism and then the optic axis. The mirror voltage must be close to the accelerating voltage of the microscope. After the mirror-prism, electrons are dispersed and continue to travel down the optic axis of the microscope. Energy-filtered images are obtained by allowing electrons to pass through the slit and the projector lens system. Energy loss spectra, angular resolved energy-scattering diagrams (showing the energy loss distribution as a function of scattering angle), and filtered images and diffraction patterns can be obtained by careful selection of the operating conditions of the microscope and crossover points. This can be achieved by selecting, with the microscope postspecimen lenses (objective, intermediate), the object point entering the spectrometer and the transfer of the crossover points on the viewing screen. Subsequent implementations of the in-column filters in higher voltage instruments (100, 200 keV microscopes) are based on OMEGA-type spectrometers that use four magnetic sectors (Figure 4-26b) generating the dispersion and transfer of electrons back onto the optic axis of the microscope. Aberrations of the spectrometer that lead to loss of resolution in spectra and generate nonuniformities in the energy distribution of electrons in images are reduced by a combination of design of the magnetic sectors entrance and exit faces, the symmetry of the configuration (the fact that the aberration of the first two sectors is compensated by the aberrations in the opposite direction of the third and fourth sectors), and the use of a series of multipoles within the path of the electrons. These filters are introduced within the projector lens system (Figure 4-27). The last implementation of in-column spectrometers is the Mandolin filter recently developed (Essers and Benners, 2006) (Figure 4-26c). This filter generates larger dispersions (a factor 3 larger than OMEGA) and is optimized for lower aberrations. This system is ideally suited for energy filtering with very narrow energy windows and large fields of view.

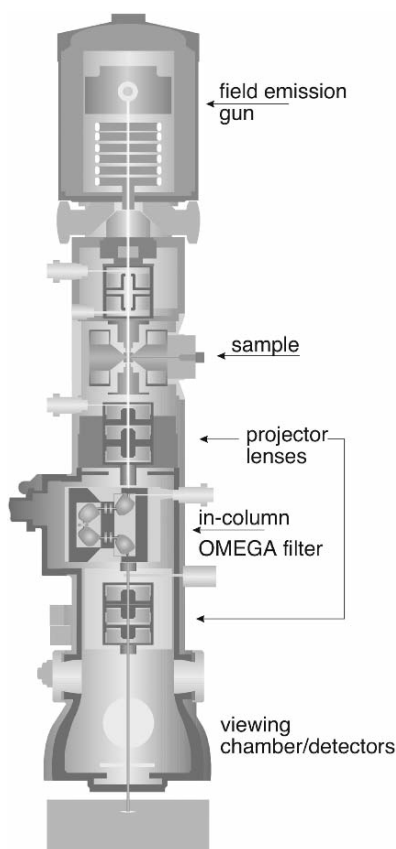


Figure 4–27. Schematic diagram of the in-column energy-filtered microscope (Zeiss-Libra 200). (Courtesy of P. Schlossmacher, Zeiss SMT.) (See color plate.)

Images, diffraction patterns (filtered/nonfiltered), and spectra for in-column filters can be observed directly on the viewing screen of the microscope and recorded with analog techniques (on negatives), imaging plates, or a digital camera following conversion of the incident electrons to photons using scintillator materials (YAG, phosphor, etc.).

For postcolumn spectrometers/imaging filters, the technology is based on the magnetic sector (Figure 4–24). Aberrations of the prism can be minimized through design of the spectrometer entrance and exit faces and the use of multipole correcting elements before the prism. Only one prism generates the required dispersion to form a spectrum in the dispersion plane. The early spectrometers were used to generate energy loss spectra by making use of a serial detection system where the spectrum is scanned, using an electrostatic field, in front of an energy slit. The number of electrons (or the current) entering the slit is subsequently measured with a scintillator detector and photomultiplier with pulse counting or current measurement methods. This serial detection process (one energy recorded at a time) is extremely inefficient for recording large energy ranges (several seconds/minutes) and

therefore parallel detectors were developed in the late 1980s (Krivanek *et al.*, 1987) to record a portion of the spectrum of 1024 energy channels. For this early technology, the slit is replaced by a series of multipole lenses (three quadrupoles) to focus and magnify the spectrum (Figure 4–28). These optical elements change the prism dispersion (from about $1.8\mu\text{m}/\text{eV}$ at 100 keV) to up to $1\text{ mm}/\text{eV}$ at the detector plane (Krivanek *et al.*, 1987). Incident electrons are converted to photons using a scintillator material (e.g., YAG) and a photodiode array (with 1024 channels) is used as the recording system with acquisition times as short as 25 ms. Recent commercial spectrometers have more complex optics (better focusing capability and aberration correction of the magnification lenses), new scintillator materials with improved transfer function, and improved transfer of the generated light between the scintillator and detector. Fast-readout two-dimensional arrays (now about 100×1200 pixels) allow acquisition of more than 100 spectra per second in commercial systems (the ENFINA spectrometer from Gatan). Noncommercial systems have achieved the same readout performance and used optical lenses to transfer the signal generated in the scintillator to two-dimensional (2D) detectors (Tencé, 2002).

For the acquisition of energy-filtered images using postcolumn filters, a series of multipole lenses (Figure 4–29) is used to transform the spectrum at the slit plane back to an image (or diffraction pattern) at the detector plane and to correct the image distortions and aberration. This “transformation” is possible because both the in-column filter and the postcolumn prism are electron optical components that produce, at the dispersion plane, spectra that contain information on the object that

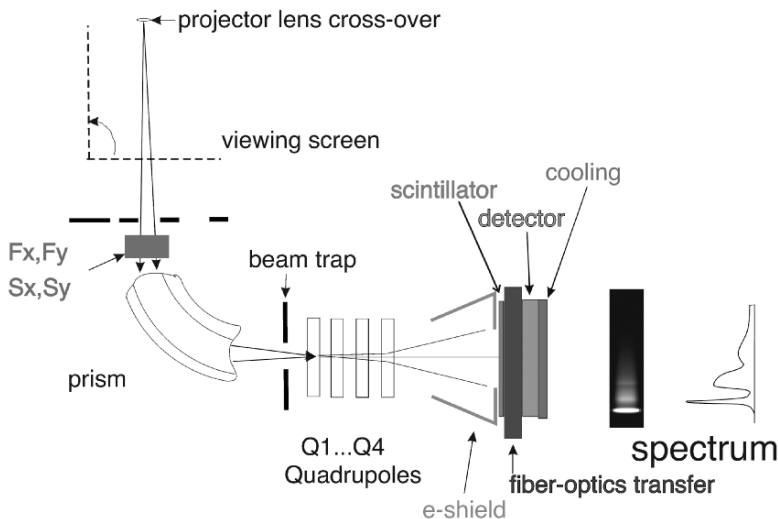


Figure 4–28. Schematic diagram of the parallel EELS spectrometer. (Adapted from Krivanek *et al.*, 1987.)

enters the spectrometer. The slit allows selection of the electrons with the energy loss of interest, which is subsequently used to form an image at the detector plane. As for the in-column filters, both images and diffraction patterns can be obtained in the detector plane of post-column filters depending on whether an image or diffraction pattern is projected at the entrance plane of the spectrometer. Worth mentioning is the use of filters to remove inelastically scattered electrons from diffraction patterns (Figure 4–30) of thick samples to enhance the visibility of the dynamic structure in the convergent beam disks and to retrieve quantitative information on structure factors (Saunders, 2003). Further details of the optical function, aberration of the spectrometers, and filters have been described extensively (Rose and Krah1, 1995; Krivanek *et al.*, 1991a, 1995a; Egerton, 1996).

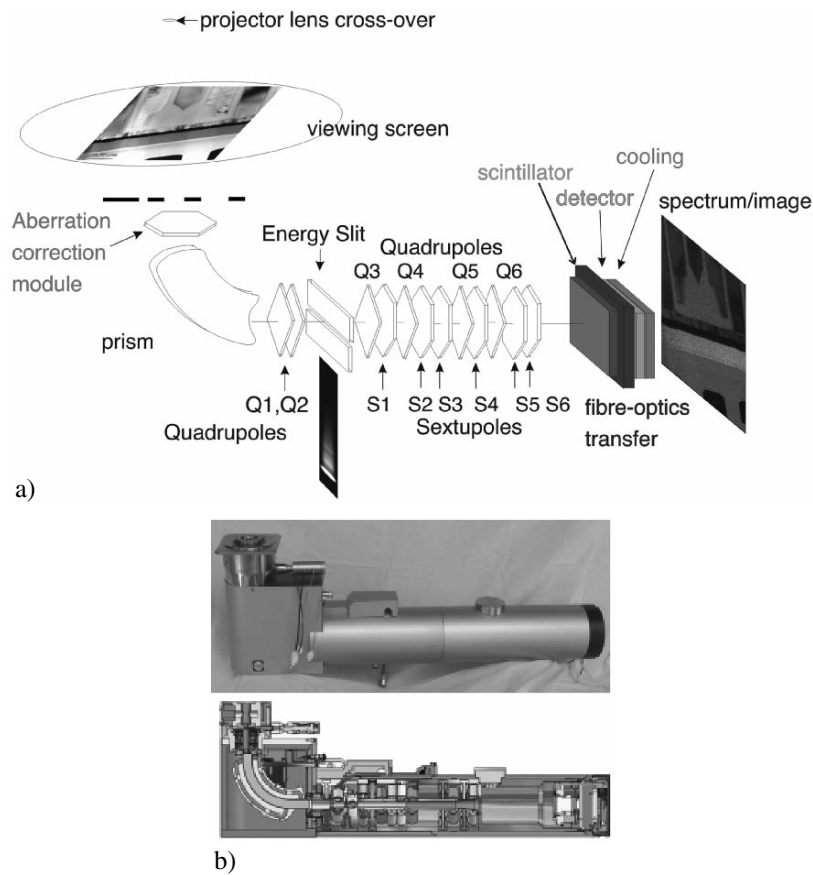


Figure 4–29. (a) Postcolumn imaging filter (Gatan Imaging Filter GIF); schematic diagram of the electron optics components and detection system (top diagram). (Adapted from Krivanek, Gubbens *et al.*, 1991a.) (b) Actual spectrometer [Gatan’s GIF 2000 series spectrometer (Tridiem model)] and components (bottom diagram). (Courtesy of M. Kundman, Gatan.) (See color plate.)

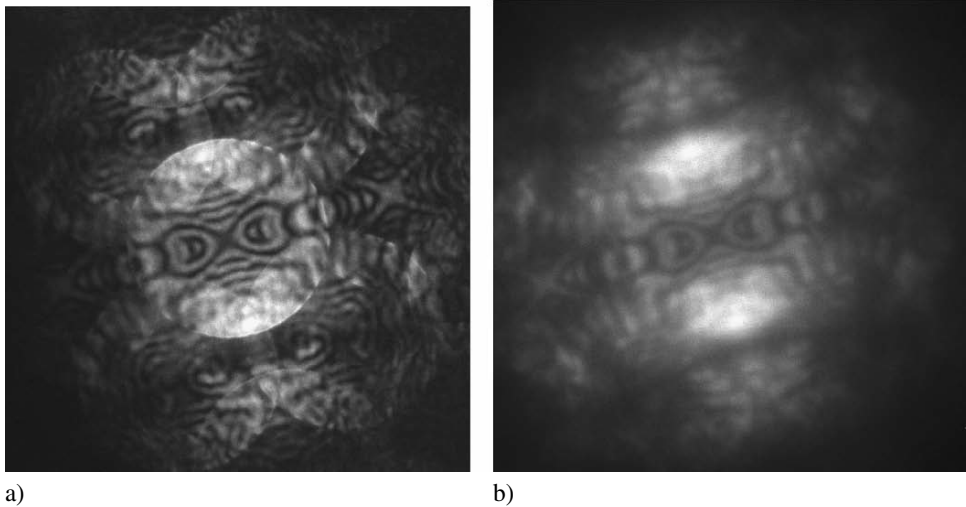


Figure 4-30. Energy-filtered electron diffraction pattern of Si (110) orientation. (a) Pattern recorded by selecting only the electrons that have lost no energy (Zero-loss filtering) and (b) pattern recorded without energy filtering.

Alignments in postcolumn filters are carried out using automated routines tuning the multipole lenses functions so as to minimize distortions, aberrations, and the uniformity of the energy within the field of view so that every point in an image corresponds to the same energy loss (the isochromaticity). The detection of images in postcolumn filters can only be carried out using scintillators and CCD cameras.

From a practical point of view, in-column energy filters offer the advantage that energy-filtered images can be directly observed on the microscope viewing screen while the postcolumn filters offer the advantage of being added to a microscope column as an optional attachment. As various implementations and models of in-column and post-columns exist, a detailed comparison of the two filtering techniques should be carried out cautiously based on the type of application of interest and thus the specific relevant parameters for that application.

2.4.2 New Developments: Monochromators

Based on the interest in energy loss near edge structures and for the purpose of increasing the temporal coherence terms of the imaging transfer function, the energy spread of the electron energy source must be decreased. One approach is to use electron guns with narrower energy distribution (e.g., see Table 4-1 in Section 2.1). This approach is limited to about 0.3 eV energy distribution with cold FEG as measured at the FWHM of the energy distribution. For cold FEG, the distribution is nonsymmetrical due to the nature of the tunneling process, which follows a Fowler–Nordheim distribution. Although this energy width allows the acquisition of good quality energy loss spectra where the intrinsic broadening of the features is in the order of 0.3 eV or greater, there are cases when spectra obtained with electrons having a narrower energy distribution is of interest (Terauchi *et al.*, 1999; Egerton,

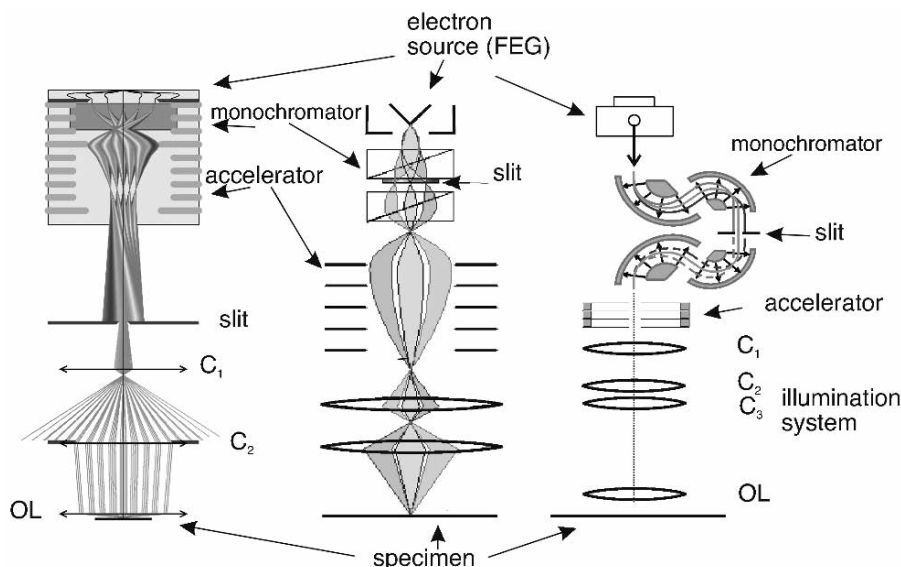


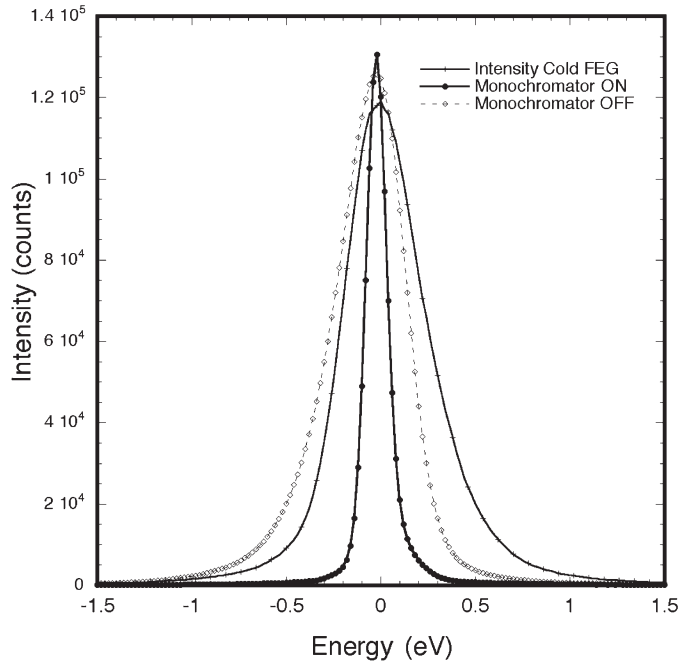
Figure 4-31. Various implementations of monochromators in commercially available instruments. Left: The FEI monochromator, single Wien filter. (Courtesy of P. Tiemeijer, FEI Company.) Center: The JEOL monochromator double-Wien filter. (Courtesy of JEOL Ltd.) Right: The electrostatic omega filter implemented in the Libra Zeiss microscope. (Courtesy of M. Haider, CEOS GmbH.) (See color plate.)

2003; Lazar *et al.*, 2003; Mitterbauer *et al.*, 2003). This can be achieved with gun monochromators and improvements in microscope stability and spectrometer resolution (Barfels *et al.*, 2002). Improvements in energy resolution to 0.1 eV have been obtained in recent commercial instruments based on various approaches (Figure 4-31). Three types of monochromators have recently been developed commercially based on the Wien spectrometer (Tiemeijer *et al.*, 2001), the double focusing Wien spectrometer (Tanaka *et al.*, 2002), and the omega-type filter (Uhlemann and Haider, 2002). The effect of monochromation on the energy distribution of the incident electrons is shown in Figure 4-32 and on spectra in Figure 4-33.

2.5 Sample Preparation Requirements

Although the quality of the microscope and of the vacuum are key elements of good AEM work, sample preparation and cleanliness are certainly factors affecting the quality of the data. To avoid sample contamination buildup under the electron beam, plasma cleaning of the samples with dedicated systems prior to insertion into the TEM column has become a routine practice. This technique gently burns off all hydrocarbons built up on the surface of the sample so that diffusion of the species during analysis in the TEM is no longer possible. An alternative to this approach is to gently heat the samples prior to insertion on the TEM with a halogen lamp in vacuum (it is possible to build such a system with off-the-shelf vacuum components) at temperatures of about 70–80°C. Although this technique does not replace the plasma cleaning approach, it is a solution for samples that might be sensitive

Figure 4–32. Zero-loss peaks obtained with a monochromator switched on using a Schottky FEG source (with stabilized electronics); the same instrument with the monochromator switched off and with a cold-field emission source (VG-STEM).



to the reactive gases used in the plasma process and often solves contamination problems due to the sample. Samples permitting, alternatives are also low-energy milling (a few hundred electronvolt ions) prior to sample insertion (with regular mills or dedicated low-energy systems). Cooling samples in a dedicated cryogenic analytical holder in the TEM also reduces the contamination buildup during

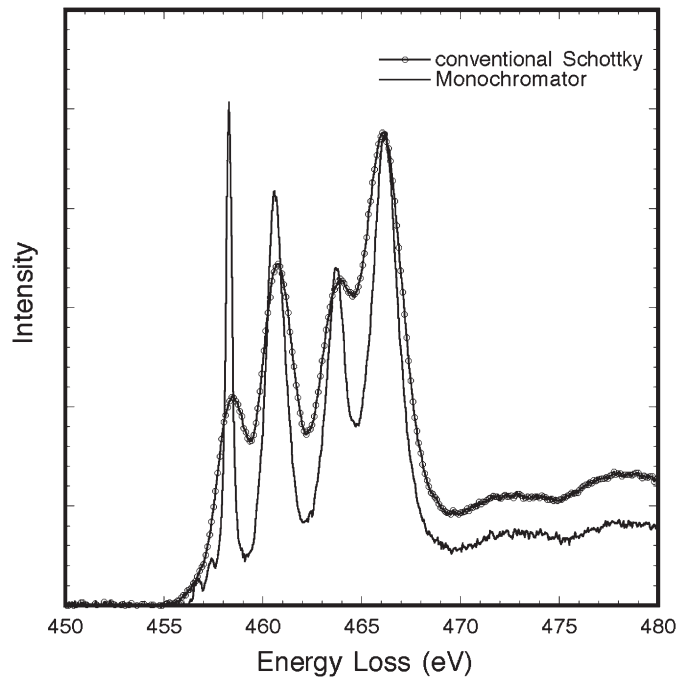


Figure 4–33. Effects of a monochromator on the visibility of peaks of the Ti L_{23} edge in CaTiO_3 . Clearly resolved are the small triplet states in front of the first strong peak.

analysis but at the expense of increased drift rate during very long analyses.

Removal of the amorphized layer produced by ion milling has become an increasingly common approach as ever thinner samples are used to achieve the ultimate spatial resolution in analysis. A combination of mechanical polishing and low-energy milling has proved excellent for obtaining very clean samples and minimal amorphous damage. Several manufacturers offer these new approaches to improve the quality of the samples. Often these methods make the difference between successful work, major scientific breakthrough, and plain disaster!

2.6 New Developments in Electron Optics

In recent years, there have been significant developments in electron optics that have led to a new generation of instruments with much finer probe-forming capabilities making use of aberration correctors (Section 2.1 and this volume, Chapter 2). Although there is emerging literature making use of these capabilities, particularly with EELS systems (Varela *et al.*, 2004; Arslan *et al.*, 2005), results on EDXS are very limited at this moment in the open literature (e.g., Watanabe and Williams, 2005b; Watanabe *et al.*, 2005). The main advantage of the aberration corrector of the probe-forming lens for AEM purposes is not the ability to form smaller probes (although a positive advantage) but rather the significant increase in electron beam current for a given spot size (see Figure 4–15). It is expected that the launch of commercial instruments equipped with such correctors will result in a dramatic improvement in the analytical performance of new AEMs. There will be new limitations, however, facing the users as electron beam damage will undoubtedly be the ultimate barrier in the analysis of most interesting materials.

3 Fundamentals

3.1 Fundamental Processes of Elastic and Inelastic Scattering

The interaction of primary electrons with electrons and nuclei in the solid results in various scattering processes and the generation of signals that can be detected in different analysis tools such as the TEM, the scanning electron microscopes, or surface analysis instruments. We can subdivide these processes into elastic and inelastic based on the energy changes of the primary incident electrons following the scattering event. Elastic scattering causes no detectable change in the energy of the primary electrons within the resolution of the measurement system typically available in the TEM. These processes do not give rise to “analytical” signals in the strict sense of the term but are nevertheless very relevant to the understanding of signal generation, imaging, and all discussions on spatial resolution as they significantly affect the angular distribution of the incident electrons, their propagation in the solid, and consequently the spatial spread of these electrons as they travel through the sample. Elastic scattering is at the basis of contrast mechanisms in TEM and STEM imaging and can directly

affect the interpretation of energy-filtered images. Furthermore, the powerful technique of Z-contrast imaging (see Chapter 2, this volume) is based on signals generated by electrons elastically scattered at high angles. The technique is commonly used in combination with analytical measurements using EELS or X-ray microanalysis and provides indirect information on changes in average atomic number within the area illuminated by the electron beam. Although traditionally this is not considered an analytical technique, this imaging method combined with spectroscopic measurements such as EDXS and EELS provides the most powerful ability to analyze materials with the highest spatial resolution.

Inelastic processes are the key to all analytical measurements as they directly and indirectly give rise to the signals detected. Figure 4–34 summarizes the various energy loss processes generating the excitation and the subsequent signal generation processes by deexcitation. Analysis techniques focus, on the one hand, on the detection of the primary event of energy loss where excitation of single electrons from strongly bound core energy levels (inner-shell ionization) and weakly bound electrons (from the valence band) occurs. These excitations cause transitions of electrons from the deep bound states to levels just above the Fermi energy and the continuum. Energy losses for the primary electrons can also occur through collective excitation of weakly bound valence electrons in the solid behaving as an electron gas (known as plasmon excitations) (Figure 4–34) or from defect states within the gap of a material. In the case of EELS, the energy losses of the primary electrons range from a few electronvolts to tens of elec-

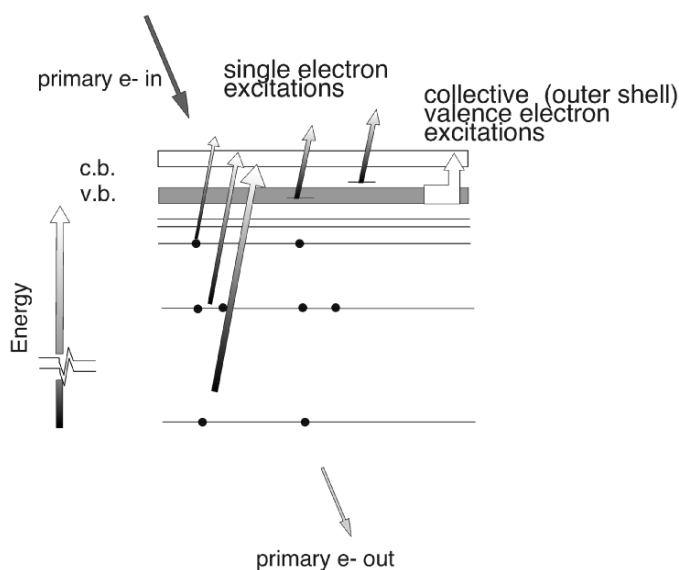


Figure 4–34. Various processes of inelastic scattering events from core levels, valence states, and defect states. Inelastic processes via collective excitation of valence electrons is also possible. Measurement of the energy losses of the primary electrons is possible with EELS. The deexcitation processes give rise to X-ray and Auger signals.

tronvolts (typically up to 50–60 eV) in the case of single electron valence excitation and plasmon losses while core losses range from a few tens of electronvolts (typically down to 30 eV) up to a few kiloelectronvolts for the core losses with significant overlap between these regions. Other processes of inelastic scattering leading to detectable signals in the spectra include the excitation of quasi-free single electrons (also known as Compton scattering) and losses due to radiative phenomena (Cerenkov effect).

On the other hand, secondary processes give rise to other signals (Figure 4–3) used in analytical techniques. The deexcitation process leads to the generation of photons detected with EDXS or Auger electrons with related characteristic energies.

3.2 Elastic Cross Sections

As discussed above, some theoretical background on elastic scattering is useful to understand electron propagation in the sample and the related contributions on inelastic signal generation. A good understanding of elastic scattering is particularly useful to simulate electron trajectories and to estimate the electron beam broadening and thus the spatial resolution in analytical measurements. We will follow the description given in detail by Reimer (1995) and Egerton (1996) using the terminology adopted by R.F. Egerton.

The differential elastic cross section $d\sigma/d\Omega$ representing the probability of scattering per unit solid angle $d\Omega$ is given by

$$\frac{d\sigma}{d\Omega} = |f|^2 \quad (6)$$

where the amplitude of the scattering factor f is directly proportional to the Fourier transform of the potential of the atom. Various models and approximations of the potential can be used. In its simplest form, the potential considers the unscreened electrostatic Coulomb potential of type

$$V(r) = \frac{Ze}{4\pi\epsilon_0 r} \quad (7)$$

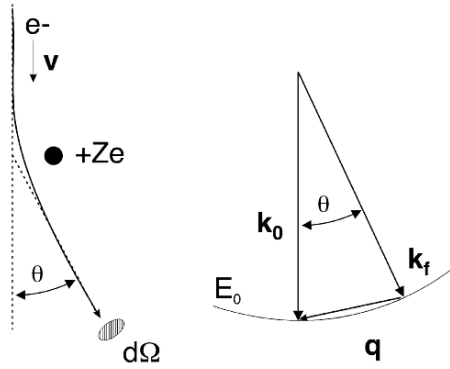
for a free atom yielding the Rutherford cross section

$$\frac{d\sigma}{d\Omega} = \frac{4\gamma^2 Z^2}{a_0^2 q^4} \quad (8)$$

where a_0 is the Bohr radius, γ is the relativistic factor [$\gamma^2 = (1 - v^2/c^2)^{-1}$], $q = 2k_0 \sin(\theta/2)$ represents the amplitude of the scattering vector $\mathbf{q}$ shown in Figure 4–35, $\hbar\mathbf{k}_0 = \gamma m_0 \mathbf{v}$ represents the momentum of the incident electron, $\mathbf{q}$ is the scattering vector defined in Figure 4–35, and $\hbar\mathbf{q}_0$ is the momentum given to the nucleus.

This first approximation of the unscreened electrostatic Coulomb potential satisfactorily describes scattering for light atoms, for large scattering angles and high incident electron energy (typical of TEM). The model, however, needs refinements for elements of high atomic number, for small scattering angles, and low incident energies (in low

Figure 4–35. Diagram for an elastic scattering process.



voltage SEM, for example) due to a singularity in the cross sections at $\theta = 0$ arising from neglect of the screening. The Rutherford model can be subsequently refined by incorporating an exponential screening factor used in the Wentzel atomic model

$$V(r) = \frac{Ze}{4\pi\epsilon_0 r} \exp(-r/r_0) \quad (9)$$

where r_0 is the screening radius that can be estimated by $r_0 = a_0 Z^{-1/3}$.

This potential and the screening radius lead to a Lenz model of differential cross section

$$\frac{d\sigma}{d\Omega} = \frac{4\gamma^2}{a_0^2} \left(\frac{Z}{q^2 + r_0^{-2}} \right) \approx \frac{4\gamma^2 Z^2}{a_0^2 k_0^4} \frac{1}{(\theta^2 + \theta_0^2)^2} \quad (10)$$

where θ_0 is the characteristic elastic scattering angle representing the width of the angular distribution of elastic scattering.

The total scattering cross section is obtained by integrating Eq. (10) over all scattering angles

$$\sigma_e = \int_0^\pi \frac{d\sigma}{d\Omega} 2\pi \sin \theta d\theta = \frac{4\pi\gamma^2}{k_0^2} Z^{4/3} \quad (11)$$

A more accurate model of the potential considering electron spin and relativistic effects leads to the Mott cross section applicable to a large range of atomic numbers, low voltages, and low scattering angles.

Based on the elastic cross sections and the number of atoms per unit volume n_a we can define a quantity called the elastic mean free path

$$\lambda_e = (\sigma_e n_a)^{-1} \quad (12)$$

which represents the mean distance between elastic scattering events in a solid assumed, for simplicity, to be amorphous.

In crystals, the elastic scattering must account for diffraction effects and the treatment of the electron distribution in the solid follows the diffraction theory developed elsewhere in this volume. With the development of Z-contrast imaging as one of the tools available in the analytical electron microscope and use of this technique with electron probes of diameter smaller than the unit cell, the characteristics of electron beam propagation in the sample become important. Detailed calculations with the multislice technique have shown that when a

small probe is positioned on an atomic column, electrons are channeled along the column for samples of thicknesses up to a few nanometers. When the beam is positioned in between the columns it strongly disperses and broadens while propagating in the sample (Dwyer & Etheridge, 2003) (Figure 4–36). This effect is probe size dependent with the channeling conditions being more relaxed with “larger” probes of 0.2 nm. Calculations show that the electron intensity builds up in the adjacent columns as the electron beam propagates (Figure 4–36b). The effect is demonstrated in other calculations (Voyles and Muller, 2004; Mobus and Nufer, 2003) showing that the beam can channel on an adjacent column and give rise to a signal even if not nominally on the initial position (Figures 4–36b, 4–37b). Although these effects are not part of the traditional AEM literature, they are becoming increasingly relevant when the very smallest probes produced by aberration-corrected instruments are used to achieve the ultimate spatial resolution in very thin samples (Section 5.1).

3.3 Inelastic Scattering Cross Sections

Before describing the cross sections of various inelastic processes, it is useful to review the concept of the total inelastic cross sections so that a comparison of elastic and inelastic scattering distributions can be made. The total inelastic cross section is also relevant to understand the process of energy losses that slow down the electrons in thin foils or in bulk samples. This concept is useful in the modeling of electron propagation in samples using Monte Carlo methods.

The differential inelastic cross section can be calculated by

$$\frac{d\sigma_i}{d\Omega} = \frac{4\gamma^2 Z}{a_0^2 q^4} \left\{ 1 - \frac{1}{[1 + (qr_0)^2]^2} \right\} \quad (13)$$

where γ , a_0 , and r_0 are as described in Section 3.2. As the scattering vector q is energy loss dependent (Section 3.4), it is approximated for the purpose of the evaluation of the total cross sections by

$$q^2 \approx k_0^2(\theta^2 + \bar{\theta}_E^2) \quad (14)$$

where $k_0 = 2\pi/\lambda = \gamma m_0 v/\hbar$ is the magnitude of incident wavevector, θ is the scattering angle, and $\bar{\theta}_E = \bar{E}/(\gamma m_0 v^2)$ is the characteristic scattering angle corresponding to the mean energy loss $\bar{E}$.

As pointed out in Egerton (1996), the first term of the expression of the total inelastic cross section [Eq. (13)] is similar to the Rutherford elastic cross section with a second term being described by the inelastic form factor.

Substitution of the scattering vector by the scattering angle leads to the angular dependence of total inelastic scattering (Colliex & Mory, 1984)

$$\frac{d\sigma_i}{d\Omega} = \frac{4\gamma^2 Z}{a_0^2 k_0^4} \frac{1}{(\theta^2 + \bar{\theta}_E^2)^2} \left[1 - \frac{\theta_0^4}{(\theta^2 + \bar{\theta}_E^2 + \theta_0^2)^2} \right] \quad (15)$$

where $\theta_0 = (k_0 r_0)^{-1}$ is related to the elastic scattering distribution defined in Section 3.2. As demonstrated in Figure 4–38, the expressions for the

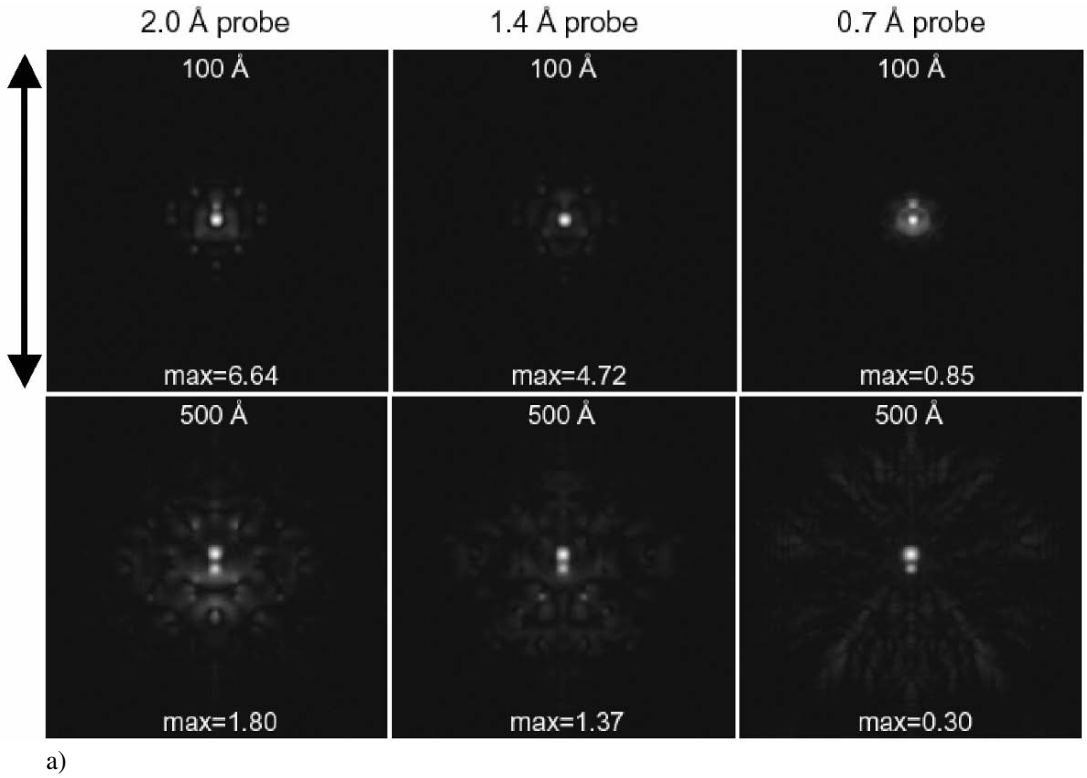
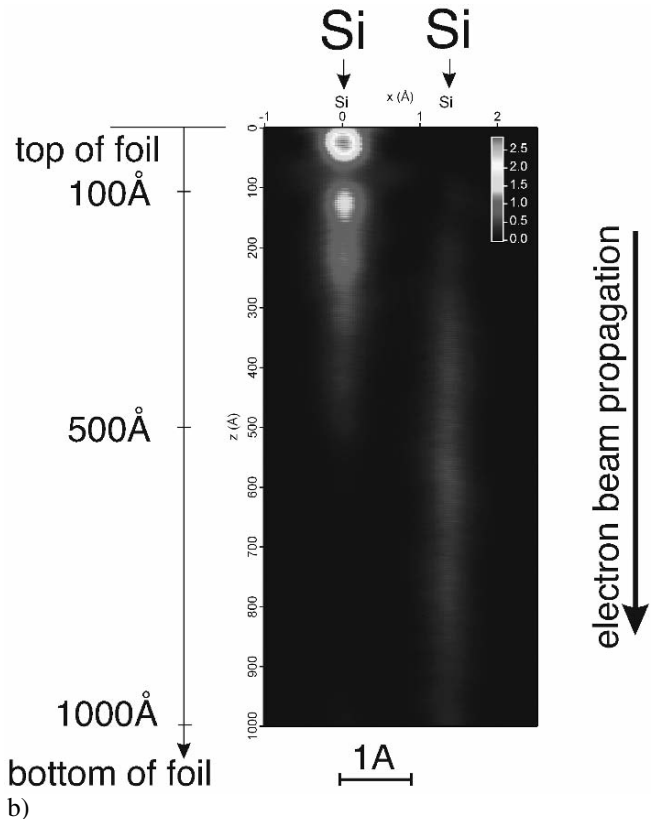


Figure 4–36. Real space intensity distribution of the probe electron density in the sample as it propagates through the thickness of the foil. Plots of the intensity distribution at two depths (100 and 500 Å) are shown for incident probe sizes of 2, 1.4, and 0.7 Å (as obtained with an aberration-corrected microscope) when the electron beam is positioned on the atomic column down the 110 orientation of the crystal (arrow length is 16.3 Å). The electron beam is channeled onto the atomic column, but the intensity maxima moves from one atomic column to the adjacent one (Dwyer and Etheridge, 2003). (Images courtesy of C. Dwyer and J. Etheridge.) (b) Alternate visualization of the process of channeling viewed as a function of thickness. The beam intensity clearly channels from one atomic column to the adjacent one as the electrons propagate in the sample (Voyles and Muller, © 2004. Reprinted with permission from Cambridge University Press.)



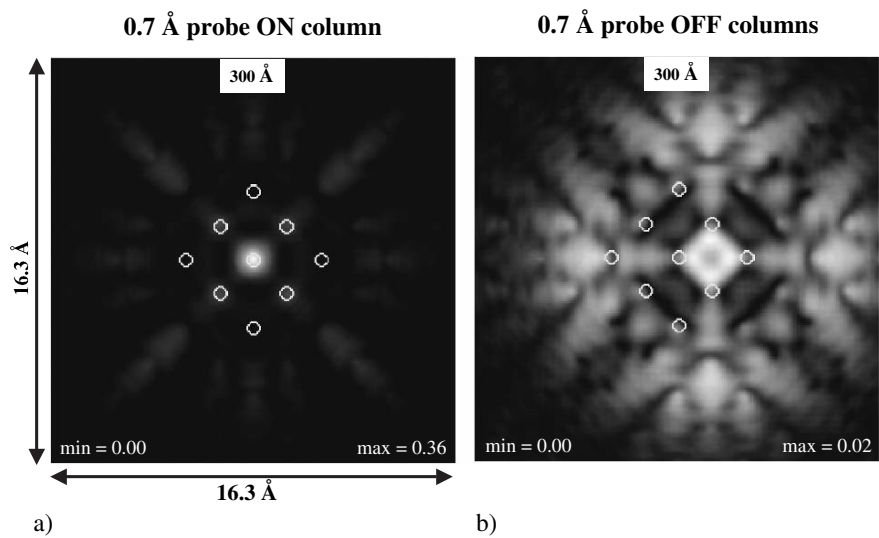


Figure 4-37. Real space intensity plots demonstrating the dispersion of the electron intensity when the electron beam is located on top of the atomic column and when it is located between two atomic columns. The bright empty circles indicate the position of the atoms in the cell closest to the point of impact of the electron beam. Channeling is observed when the beam is positioned on the atomic column (a) while much stronger dispersion is observed when the electron beam is not on the atomic column (b). (Data courtesy of C. Dywer and J. Etheridge.) (For part A, see color plate.)

total inelastic and elastic scattering cross sections make it possible to compare the width of the respective angular distributions. The angular distributions are Lorentzian with an angular width of θ_E for the inelastic distributions and θ_0 for the elastic distribution. It is also possible to note that the angular width of the elastic distribution (for scattering from free atoms) is larger than the inelastic distribution.

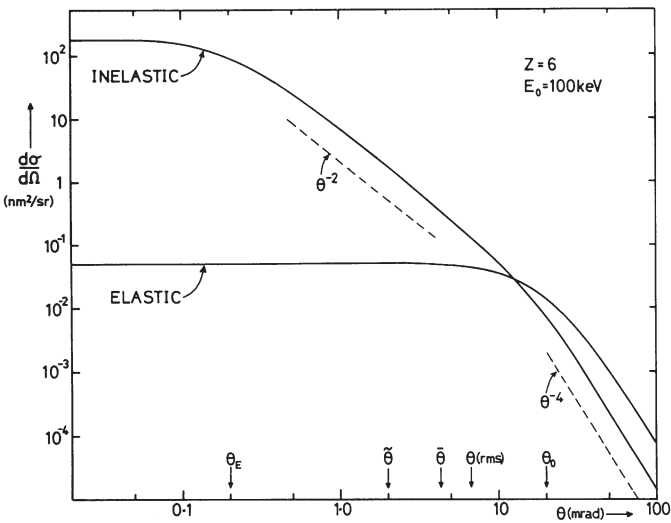


Figure 4-38. Comparison of the elastic and total inelastic angular scattering distributions for the C atoms at 100 keV. The characteristic angles for the inelastic and elastic distributions (θ_E and θ_0), the mean angle θ , the median angle $\bar{\theta}$, and the root-mean-square angle θ_{rms} are shown. (From Egerton, © 1996, with permission from Springer Science+ Business Media.)

The total cross section integrated up to a scattering angle β is relevant when calculating the inelastic free path or the stopping power (see below)

$$\sigma_i(\beta) = \frac{8\pi\gamma^2 Z^{1/3}}{k_0^2} \ln \left[\frac{(\beta^2 + \theta_E^2)(\theta_0^2 + \theta_E^2)}{\theta_E^2(\beta^2 + \theta_0^2 + \theta_E^2)} \right] \quad (16)$$

The total inelastic scattering cross section integrated over all scattering angles is approximated by

$$\sigma_i \approx 16\pi\gamma^2 Z^{1/3} \ln(\theta_0/\theta_E) \approx 8\pi\gamma^2 Z^{1/3} \ln(2/\theta_E) \quad (17)$$

by replacing the cutoff angle θ_0 with the Bethe-ridge angle $(2\theta_E)^{1/2}$. This expression leads to a comparison of the relative magnitude of the elastic and inelastic cross sections as described by

$$\sigma_i/\sigma_e \approx 2 \ln(2/\bar{\theta}_E)/Z = C/Z \quad (18)$$

where the coefficient C (around 20) does not vary significantly with atomic number and incident electron energy.

This expression can be used to calculate the scattering contrast as defined in Reimer (1995) and to interpret the contrast in STEM images obtained by calculating the ratio of inelastic and elastic signals.

These calculations are a good first approximation of the behavior of the inelastic cross sections with further refinements, accounting for the outer-shell electrons, leading to systematic variations in the total cross sections related to the filling of the periodic table. Minima in the inelastic cross sections occur for atoms with closed shell while maxima occur for atoms filling the s shell due to strong effects of valence excitations (Egerton, 1996) (Figure 4–39).

The importance of the total inelastic cross section becomes apparent in AEM as it forms the basis of calculations of the stopping power of the electrons in the solid. This quantity is therefore relevant to understanding the propagation of the electrons and simulation of electron trajectories in Monte Carlo simulations (Section 5.1):

$$S = \frac{dE}{dz} = n_a \bar{E} \sigma_i \quad (19)$$

where E is the energy loss, z is the distance traveled in the sample, $\bar{E}$ is the mean energy loss for the inelastic event, and n_a is the number of atoms in the solid per unit volume.

The total inelastic cross section considers all possible events giving rise to energy losses represented by an average energy loss $\bar{E}$ and does not consider the individual interactions of the incident electrons with the inner shells or outer shell atomic electrons. Predictions of the details of a spectrum and the intensity at a given energy loss, however, must take into account these various inelastic processes, their energy dependence, and the angular distribution of scattering. To do so, we must consider Bethe's theory to predict the probability of transitions of electrons from an initial state wavefunction ψ_0 to final state wavefunction ψ_n following interaction with incident fast electrons and the related cross section. If the energy losses and the momentum transfer are small compared to the momentum of the incident electron and there

is only one scattering event during the interaction (the first Born approximation) the cross section can be described by

$$\frac{d\sigma_n}{d\Omega} = \left(\frac{4\gamma^2}{a_0^2 q^4} \right) \frac{k_1}{k_0} |\epsilon_n(q)|^2 \quad (20)$$

where k_0 and k_1 are the magnitudes of the wave vectors of the incident electron before and after scattering respectively and $\mathbf{q}$ is the scattering vector related to the momentum transfer $\hbar\mathbf{q} = \hbar(\mathbf{k}_0 - \mathbf{k}_1)$ (Figure 4-40). The first term has already been encountered in the description of Rutherford scattering for a single charge [Eq. (8)] and constitutes the amplitude factor in the cross section. This term is modified by the inelastic form factor related to the transition matrix element defined as

$$\epsilon_n = \int \Psi_n^* \sum_j \exp(i\theta \cdot \rho_j) \Psi_0 d\tau = \left\langle \Psi_0 \left| \sum_j \exp(i\theta \cdot \rho_j) \right| \Psi_0 \right\rangle \quad (21)$$

which expresses the interaction of the incident electron and the atomic electron via an operator $\exp(i\mathbf{q} \cdot \mathbf{r}_j)$ where $\mathbf{r}_j$ is the coordinate position of the fast electron treated as a plane wave and the sum is carried out over the atomic electrons from $j = 1$ to $j = Z$. This form factor contains the information related to the properties of the material through the wavefunctions of the electrons in the solid (see Section 8.1). $|\epsilon_n(q)|^2$ is independent of the electron energy and solely dependent of the atom and its environment.

From the form factor, the generalized oscillator strength (GOS) can be defined as

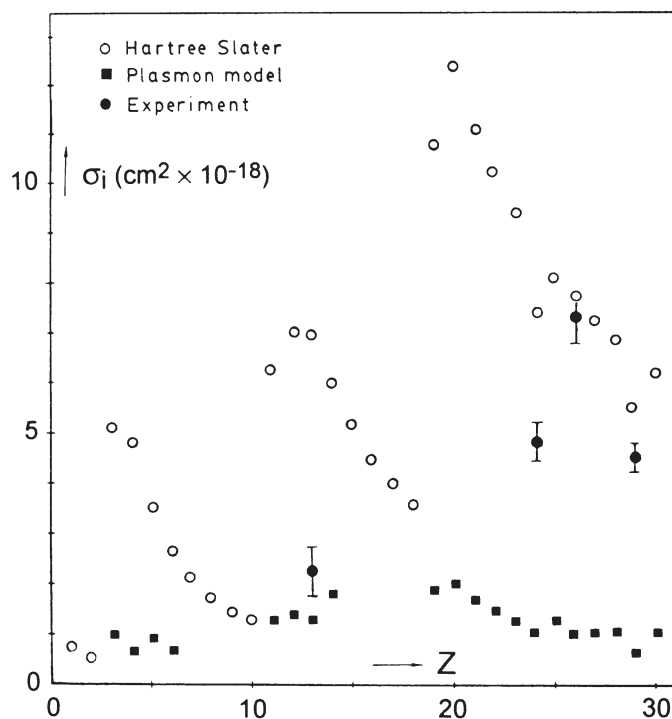
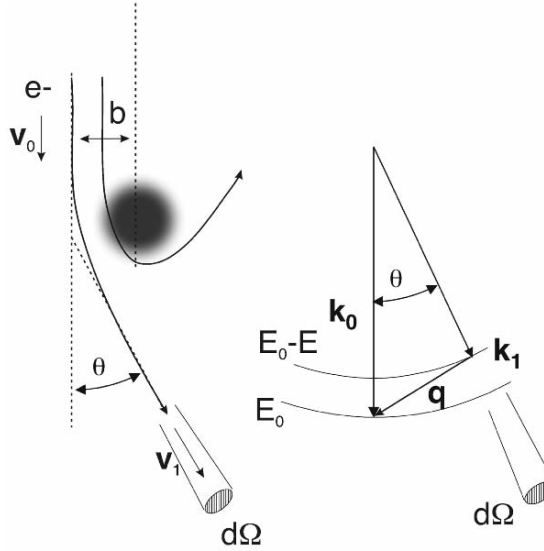


Figure 4-39. Total inelastic cross sections as a function of atomic number (80 keV electrons). Open circles are calculations based on the Hartree-Slater models, solid squares are based on calculations accounting for plasmon losses, and the solid circles are experimental data. (From Egerton, © 1996, with permission from Springer Science+Business Media.)

Figure 4–40. Inelastic scattering diagram showing the scattering vectors, the energy loss, and the effect of the classical impact parameter b on the scattering angle (also refer to discussions on spatial resolution in Section 6).



$$f_n(q) = \frac{E_n}{R} \frac{|\epsilon_n(q)|^2}{(qa_0)^2} \quad (22)$$

where R is the Rydberg energy (13.6 eV) and E_n is the energy loss of the transition. This term contains also information on the probability of the transition from an initial state wavefunction ψ_0 to a final state wavefunction ψ_n .

Using the concept of the generalized oscillator strength, the cross section can be expressed as

$$\frac{d\sigma_n}{d\Omega} = \frac{4\gamma^2 R}{E_n q^2} \frac{k_1}{k_0} f_n(q) \quad (23)$$

The full angular and energy dependence of the scattering can be described by the double differential cross section

$$\frac{d^2\sigma}{d\Omega dE} = \frac{4\gamma^2 R}{E q^2} \frac{k_1}{k_0} \frac{df}{dE}(q, E) \quad (24)$$

where the scattering vector $\mathbf{q}$ is expressed in term of scattering angle θ (Figure 4–40) and the initial and final wavevectors $\mathbf{k}_0$ and $\mathbf{k}_1$. For small scattering angles and small energy losses relative to the incident energy (as in typical TEM experiments) $k_1/k_0 \approx 1$ and

$$q^2 \approx k_0^2(\theta^2 + \theta_E^2) \quad (25)$$

with θ_E the relativistically corrected characteristic scattering angle

$$\theta_E = \frac{E}{\gamma m_0 v^2} = \frac{E}{(E_0 + m_0 c^2)(v/c)^2} = \frac{E}{E_0} \left(\frac{E_0 + m_0 c^2}{E_0 + 2m_0 c^2} \right) \quad (26)$$

and the cross section becomes

$$\frac{d^2\sigma}{d\Omega dE} \approx \frac{4\gamma^2 R}{E k_0^2} \left(\frac{1}{\theta^2 + \theta_E^2} \right) \frac{df}{dE} = \frac{8a_0^2 R^2}{E m_0 v^2} \left(\frac{1}{\theta^2 + \theta_E^2} \right) \frac{df}{dE} \quad (27)$$

The first two terms of Eq. (27) represent a kinematic term of the scattering while the df/dE term provides the information on the initial and final wavefunctions of the electrons (including changes in the bonding state) either via inner shell or valence excitations. The second term imposes a Lorentzian angular distribution of scattering at small scattering angles as it contains the term θ_E representing the half-width of the distribution. At low energy losses and small scattering angles (i.e., $q \rightarrow 0$), df/dE does not vary with θ and it can therefore be considered as a dipole oscillator strength. In this region of scattering (achieved by limiting the maximum scattering angle θ with a small aperture β) the response to the excitation from an electron would be equivalent to the excitation by a photon.

3.3.1 Outer Shell Excitations

The expression of df/dE has been derived in Bethe's theory using atomic models. To treat inner shell excitations, the initial and final wavefunctions of the electrons must be known. These wavefunctions vary with the chemical state in a solid but the atomic models are still a good approximation for the purpose of quantification of the atomic concentration based on intensity of edges. In the case of a solid, however, wavefunctions related to outer shell excitations are much more strongly modified by the interaction between atoms and the collective behavior of the electrons. Similarly, to model fine modulations in the edges simple atomic models are not sufficient (Section 8.1). To consider these effects it is more convenient to consider the dielectric response of the medium $\epsilon(q, E)$ to the incident electron treated as a point charge perturbing the solid.

The dielectric formulation of scattering relates the double differential scattering cross section to

$$\frac{d^2\sigma}{d\Omega dE} \approx \frac{\text{Im}[-1/\epsilon(q, E)]}{\pi^2 a_0 m_0 v^2 n_a} \left(\frac{1}{\theta^2 + \theta_E^2} \right) \quad (28)$$

thus the GOS relates to the dielectric function as

$$\frac{df}{dE}(q, E) = \frac{2E}{\pi E_a^2} \text{Im} \left[\frac{-1}{\epsilon(q, E)} \right] \quad (29)$$

where $E_a^2 = \hbar^2 n_a e^2 / (\epsilon_0 m_0)$, n_a is the number of electrons per unit volume, and $\text{Im}[-1/\epsilon(q, E)]$ is the loss function that contains all the material dependence via the complex dielectric response of the solid $\epsilon = \epsilon_1 + i\epsilon_2$ to electromagnetic radiation. The real part of the dielectric function $\text{Re}[\epsilon(q, E) - 1] = \epsilon_1(q, E) - 1$ is related to the polarizability of the medium and the imaginary part $\text{Im}[\epsilon(q, E)] = \epsilon_2$ is related to the absorption. Maxima in the loss function $\text{Im}(-1/\epsilon) = \epsilon_2 / (\epsilon_1^2 + \epsilon_2^2)$ that result in strong peaks in the energy loss spectra (Figure 4–7) occur when $\epsilon_1 = 0$ and ϵ_2 is small (the damping is weak). The condition $\epsilon_1 = 0$ suggests that a condition of resonance is met and that the medium is unstable. This instability corresponds to creation of a quasiparticle called plasmon of energy E_a . Further discussion on the derivation of the loss function and applications can be found in Section 8.2.

3.4 Calculations of Cross Sections

Quantification of EELS and EDXS spectra based on signals recorded from edges and X-ray peaks (Section 4) is based on the knowledge of cross sections. For inner shell excitations in EELS (and the related EDXS peaks) it is assumed, as a first approximation, that the initial and final wavefunctions are not affected by collective electron behavior and the cross sections can be calculated using various models based either on the simple hydrogenic description of the atomic electrons or the more accurate Hartree–Slater method. For an atom of atomic number Z , the hydrogenic model uses the simplification of the electrostatic potential arising from the treatment of the nuclear charge Ze and the screening due to remaining inner nonexcited electrons. Different expressions of the effective charges are used for K and L shells and consideration is given to the presence of outer electrons in higher energy levels that modify the binding energy of the inner shell electrons. Solving the Schrödinger equation in the revised simplified potential leads to analytical solutions that can be easily calculated for K shells using programs developed by Egerton (1996). The treatment of L shells, although initially unsuccessful due to the simplifications of the hydrogenic model, was revised by considering experimental optical and energy loss data with built-in corrections. These modifications have led to improved accuracy in the treatment of L shells for transition metals; programs and models for L shells are also available (Egerton, 1996) in the literature and in commercial EELS analysis programs (the Digital Micrograph software from Gatan). Empirical modifications of the hydrogenic models for M edges have also been developed (Luo & Zeitler, 1991). The Hartree–Slater (HS) approach requires iterative solutions that lead to more accurate cross sections that consider a potential calculated based on the charge density of the electrons in a self-consistent manner. Cross section tabulations have been developed by Leapman et al. (1980) and Rez (1982) and the results have been implemented within commercial EELS analysis software for K, L, and M shells within Digital Micrograph. The HS model predicts more realistic shapes of edges and includes corrections for sharp features related to unoccupied bound states present on transition metal L edges. Recent developments allow the combination of solid-state effects near the edge threshold, calculated from band structure techniques, with the atomic models at higher energy losses into more accurate cross section models (Potapov et al., 2004).

3.4.1 Angular and Energy Dependence of the GOS

Some physical insight into the scattering process is given by the analysis of the angular and energy dependence of the GOS $df(q,E)/dE$, known as the Bethe surface, as plotted for the C K edge (Figure 4–41a). This plot also provides information related to the optimization of the acquisition conditions as it shows the dependence of the differential cross section on angle and energy loss. There are two regions of key importance on the Bethe surface. At energy losses just above the threshold, the distribution is peaked at small scattering angles ($\theta = 0, q \rightarrow 0$) and the angular dependence in the cross section $d^2\sigma/d\Omega dE$ is controlled by

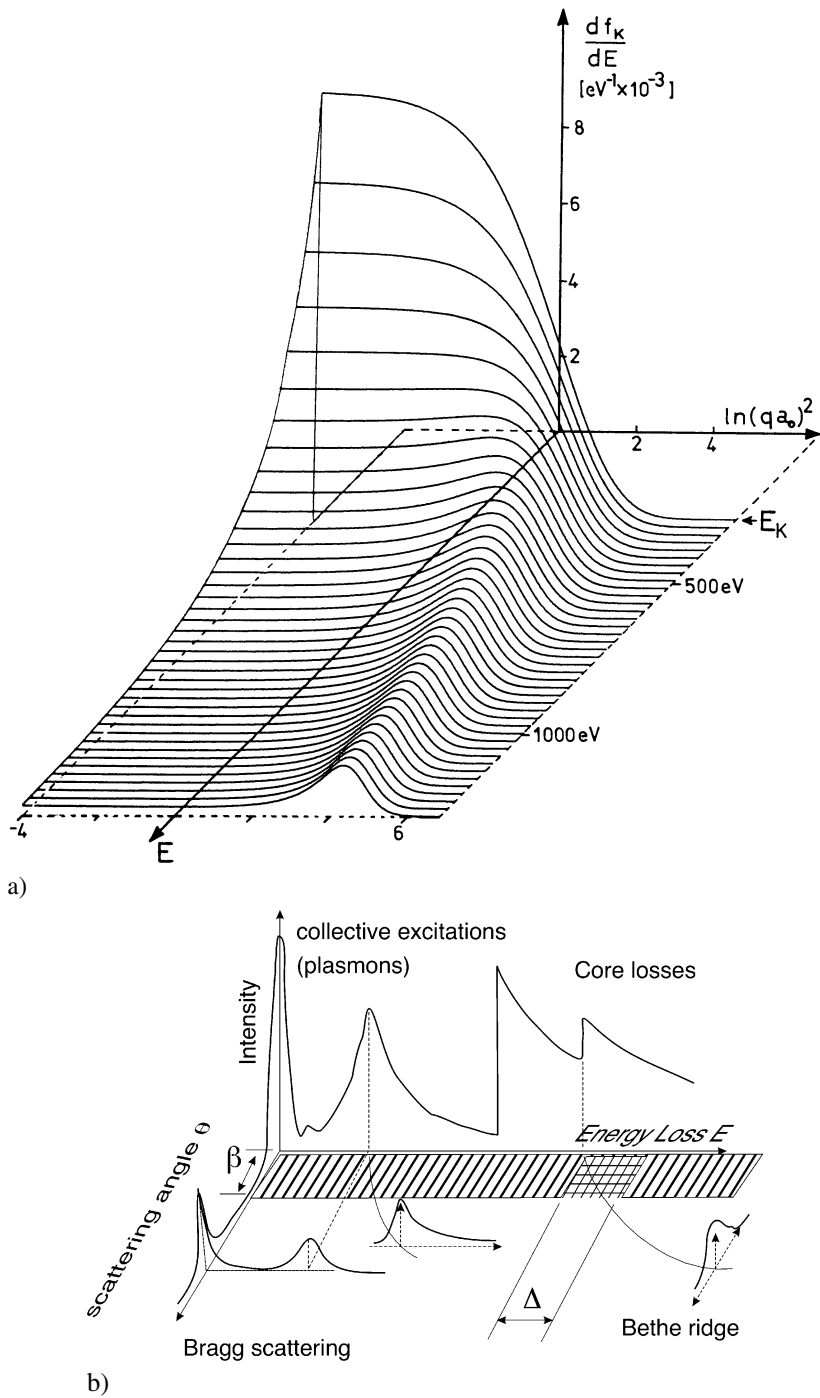


Figure 4-41. (a) Two-dimensional angular and energy distribution of the general oscillator strength per unit energy loss (df/dE) for the CK edge (also described as the Bethe surface of scattering). The “hump” at high energy from the edge onset and large scattering vector q is the Bethe ridge. (From Egerton, © 1996, with permission from Springer Science+Business Media.) (b) Two-dimensional schematic plot of the full energy and scattering angle distribution of inelastic losses including plasmon losses, core losses, as well as the Bethe ridge and the plasmon dispersions. The diagram shows the elastic scattering as part of Bragg peaks (as in a crystalline sample). Scattering from atoms or amorphous samples would give rise to a broad peak centered at zero scattering angle. The angle (β) and energy (Δ) integration windows for the calculations of the partial cross sections are identified. (Adapted from Schattschneider and Werner, 2005.)

the kinematic term of the cross section since the GOS does not vary strongly with q at small scattering angles. This region of the GOS is known as the dipole region and represents collisions with a large impact parameter b (see Figure 4–40 and Section 5.2) and little momentum transfer. Spectra acquired with small scattering angles in electron energy loss experiments are thus equivalent to optical (X-ray) absorption spectra where only dipole transitions are allowed. Based on this equivalence, photoabsorption spectra can be used to improve GOS models for use in cross-sectional calculations (Egerton & Leapman, 1995; Egerton, 1993). The dipole transitions also imply that the change of angular momentum quantum number ℓ when the electron is initially excited from a core level “i” to a final level “f” is $\Delta\ell = \pm 1$.

At larger energy losses from threshold, corresponding to collision with nearly-free electrons, the GOS peaks at large scattering angles $\theta_C = (E/E_0)^2$ corresponding to what is known as the Bethe ridge. Experimentally, this peak can be readily visualized by collecting energy loss spectra in thin samples with a spectrometer aperture centered few tens mrad off the transmitted beam collection angle (thus away from the forward scattered direction). This acquisition condition is equivalent to Compton scattering experiments between an incident electron and a free electron. As discussed in a review by Schattschneider and Exner (1995), a quantitative analysis of the width of this experimental peak provides information on the momentum distribution of atomic electrons. Similarly, by recording diffraction patterns of thin specimens with an energy filter (see Section 2.4.1) at high energy loss (few hundred electronvolts) it is possible to see a ring corresponding to the Compton scattering peak (Mayer et al., 1995; Egerton, 1996). The full energy and scattering angle dependence of the entire spectrum includes therefore the various processes of low-loss scattering (including the dispersion of plasmon losses, the elastic Bragg scattering (in crystalline specimens) and the core losses (with the Bethe ridge visible) (Figure 4–41b).

3.4.2 Partial and Total Ionization Cross Sections

Whether we have to compute cross sections for quantification of EDXS or energy loss spectra, the differential cross sections have to be integrated with respect to angle and energy loss to yield partial or total cross sections. In the case of energy loss experiments, the spectra are recorded with a fixed collection angle β and the partial cross section is therefore related to the double differential cross section through integration up to an angle β

$$\frac{d\sigma}{dE} \equiv \frac{4R\hbar^2}{Em_0^2v^2} \int_0^\beta \frac{df(q, E)}{dE} 2\pi \frac{\theta}{(\theta^2 + \theta_E^2)} d\theta \quad (30)$$

where the GOS is expressed in terms of the scattering vector $\mathbf{q}$ and the scattering angle is θ . By considering the kinematics of scattering (Figure 4–40), $q^2 \equiv k_0^2(\theta^2 + \theta_E^2)$ and the integration can be either carried out over θ or, after transformation of variables and limits of integration, over q (Egerton, 1996). The partial cross section follows a simple trend as $d\sigma/dE \propto E^{-s}$ where s is the slope of the function.

For the purpose of quantification in energy loss spectroscopy (see Section 4.2), core shell edge spectra acquired with a collection angle β are integrated over an energy window Δ . The integrated signal for a given collection angle β and energy window is thus related to the partial cross section as

$$I_k^1(\beta,\Delta) = NI_0\sigma_k(\beta,\Delta) \tag{31}$$

where N is the number of atoms per unit area, I_0 is the incident intensity (approximated by the number of counts under the zero-loss peak), and

$$\sigma_k(\beta,\Delta) = \int_{E_k}^{E_k+\Delta} \frac{d\sigma}{dE} dE \tag{32}$$

Examples of the values of partial cross section for light elements as a function of collection angle β are shown in Figure 4-42. These plots

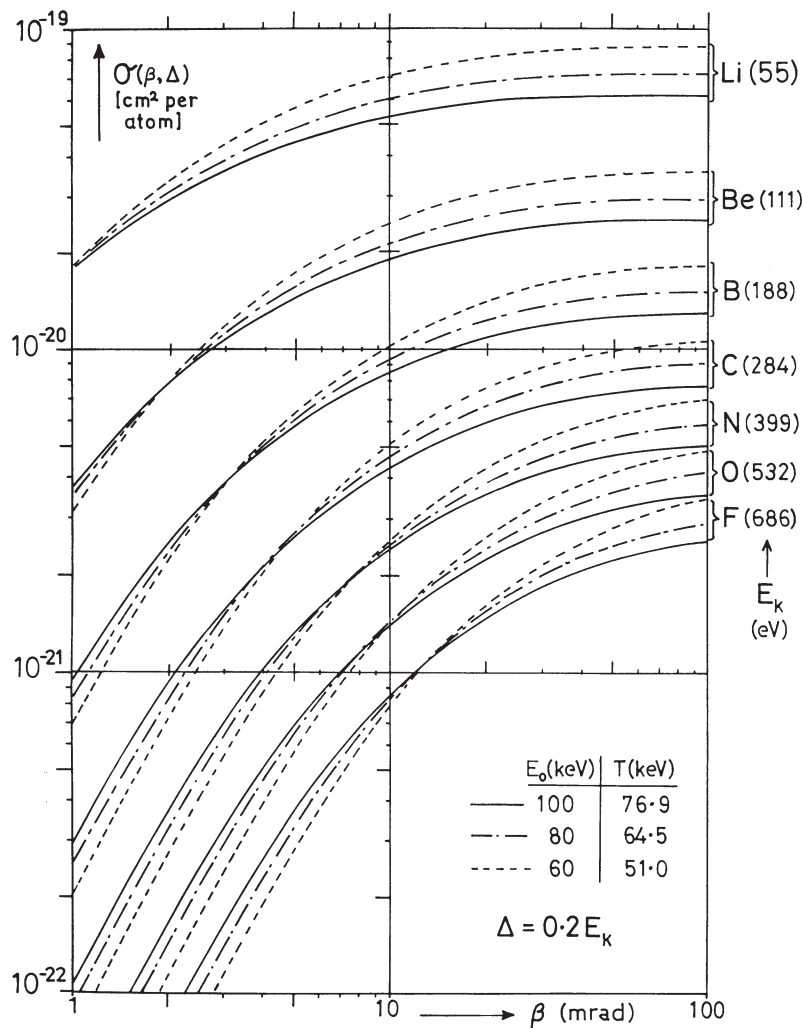
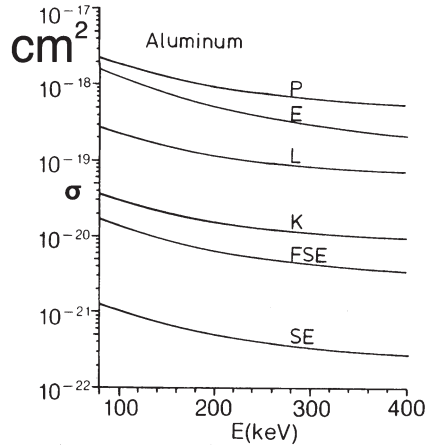


Figure 4-42. Angular dependence of the partial cross sections for lighter elements. The energy integration is carried out over an energy window Δ equivalent to $0.2 E_k$ where E_k is the threshold energy. (From Egerton, © 1996, with permission from Springer Science+Business Media.)

Figure 4-43. Comparison of cross section for various scattering processes for Al as a function of incident electron energy. The comparison shows the strongest inelastic scattering (dominated by plasmon losses P), followed by elastic scattering (E), L shell ionization (L), K shell ionization (K), fast secondary electrons (FSE), and secondary electrons (SE). The generations of FSE and SE are not discussed in this chapter. (From Joy et al., © 1986, with permission from Springer Science+Business Media.)



show the increase of expected signal with increases of collection angle and the variation with atomic number. Note the asymptotic behavior of the cross sections for large scattering angles approaching 100 mrad. This effect suggests that by using large collection angles above a few tens of millirad, the collected intensity does not increase significantly. This asymptotic behavior affects the optimization of signals and signal-to-background ratios as discussed in relation to the detection limits (Section 7). This integration over scattering angles and energy window is demonstrated in the hatched area in Figure 4-41b.

For the calculation of the probability of X-ray signal generation following ionization, the collection angle of the primary electrons is irrelevant and it is necessary to consider all possible scattering angles of the incident electron. The integration must be carried out up to $\beta = \pi$, which includes possible backscattering. This integration leads to the total inelastic cross section for core shell ionization that is used in X-ray quantification (Section 4.1.1)

$$\sigma_k \equiv 4\pi a_0^2 N_k b_k (R/T) (R/E_k) \ln(c_k T/E_k) \quad (33)$$

where N_k is a number related to the occupancy in a particular shell (2, 8, 18 for K, L, M shells, respectively) and b_k and c_k are factors deduced from theory or experiments [(Inokuti, 1971, 1978; Powell, 1976), also tabulated in Goldstein et al. (1986)] and T is the kinetic energy of the electron.

It is useful to compare the magnitude of the cross sections for elastic and various inelastic events such as K and L shell excitation, outer shell (including excitations of plasmons), and generation of secondary electrons (Figure 4-43) as a function of energy. Based on the inelastic cross sections it is possible to define an inelastic mean free path (mean distance between inelastic scattering of events) in a fashion similar to the elastic mean free path (Section 3.1). For an inelastic scattering event i ,

$$\lambda_i = (n_a \sigma_i)^{-1} \quad (34)$$

where n_a is the number of atoms per unit volume and σ_i is the cross section for an inelastic event i (either total inelastic scattering or K, L shell excitation).

3.5 Shape of EELS Edges

There are two factors that contribute to the shape of inner-shell absorption edges observed in energy loss spectra. The atomic calculations of the cross sections (hydrogenic of Hartree–Slater) give, to a first good approximation, the expected shape of the edge related to transitions from a core state to free continuum states. Although the general shape is reproduced particularly well at high energies from the threshold, details arising from transitions to discrete unoccupied bound states can be important. These effects are visible both in terms of the integrated edge intensity (particularly for transition metal L edges) and in terms of the fine details arising from electronic structure and the bonding environment in the solid state. These latter modulations are called ELNES. The dipole selection rules for the transitions from an initial core level to a final state (see Section 8.1) imply that there must be a change in angular momentum quantum number $\Delta\ell = \pm 1$. This implies that the angular character of the final states (i.e., the symmetry) is determined by the core state quantum numbers. Making use of the nomenclature presented in Figure 4–11, the following rules can be summarized. For K edges, the principal quantum number is $n = 1$, the angular momentum quantum number of the core state is s (i.e., $\ell = 0$), and the final state character is p ($\ell = 1$). For L_{23} edges, the core level is $2p$ and the final states probed are s ($\ell = 0$) and d ($\ell = 2$). For M_{45} edges, the core level is $3d$ and the final states will be f or p . Spectroscopically the K, L, M, N, and O edges therefore reflect the main quantum number of the core level $n = 1, 2, 3, 4, 5$, respectively, and each edge presents sublevels due to the angular momentum quantum number ℓ and spin-orbit coupling due to the j quantum number (Figures 4–11 and 4–4). L edges thus present transitions to three sublevels for $2s$, $2p_{3/2}$, and $2p_{1/2}$ giving rise to the L_1 , L_2 , and L_3 edges, respectively, while M edges have five sublevels and so forth as summarized in figure 4–11.

A summary of the types of shapes is reflected in Figures 4–44 and 4–45 along with a table (Table 4.4) showing the expected edges visible for energy losses up to 2000 eV. These edges can be detected in typical TEM experiments. K edges are observed for light elements up to Si

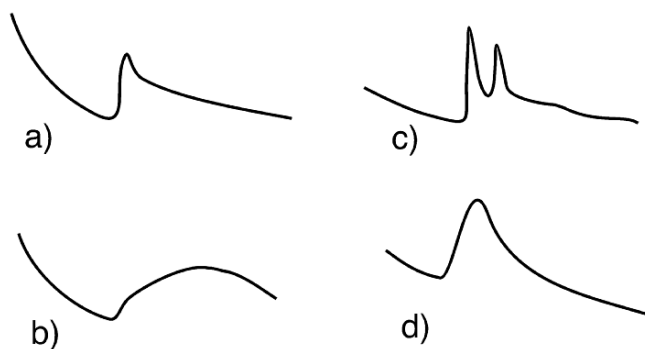


Figure 4–44. Schematic diagram of edge shapes. (Adapted from Colliex, 1984, and Hofer, 1995.)

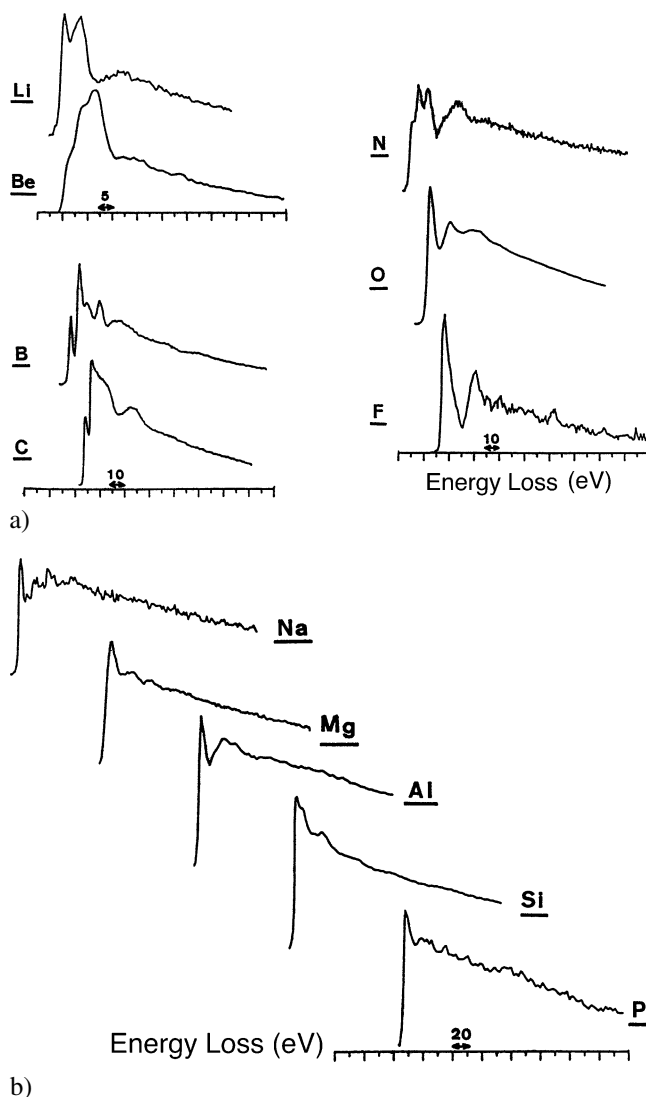


Figure 4-45. (a) Trends in K edges for light elements and (b) *sp* electron metals.

(1800 eV) and generally present a sawtooth shape. The overall edge shapes are relatively well predicted by the hydrogenic and Hartree–Slater cross section models although, within the first 30 eV from the edge threshold, fine variations due to the bound final states of *p*-character in the ELNES are present. L edges are observed for a larger range of elements and the shape of these varies with the atomic number. The L_2 and L_3 edges (also labeled as L_{23} edges) are the strongest in the series due to the large cross sections and the selection rules imposing transitions to unoccupied levels with large density of states above the Fermi energy (if there are 3*d* empty states). The L_1 edges are hardly visible in typical acquisition conditions favoring the dipole transitions (i.e., small collection angle). The edges of elements Na to Ar present

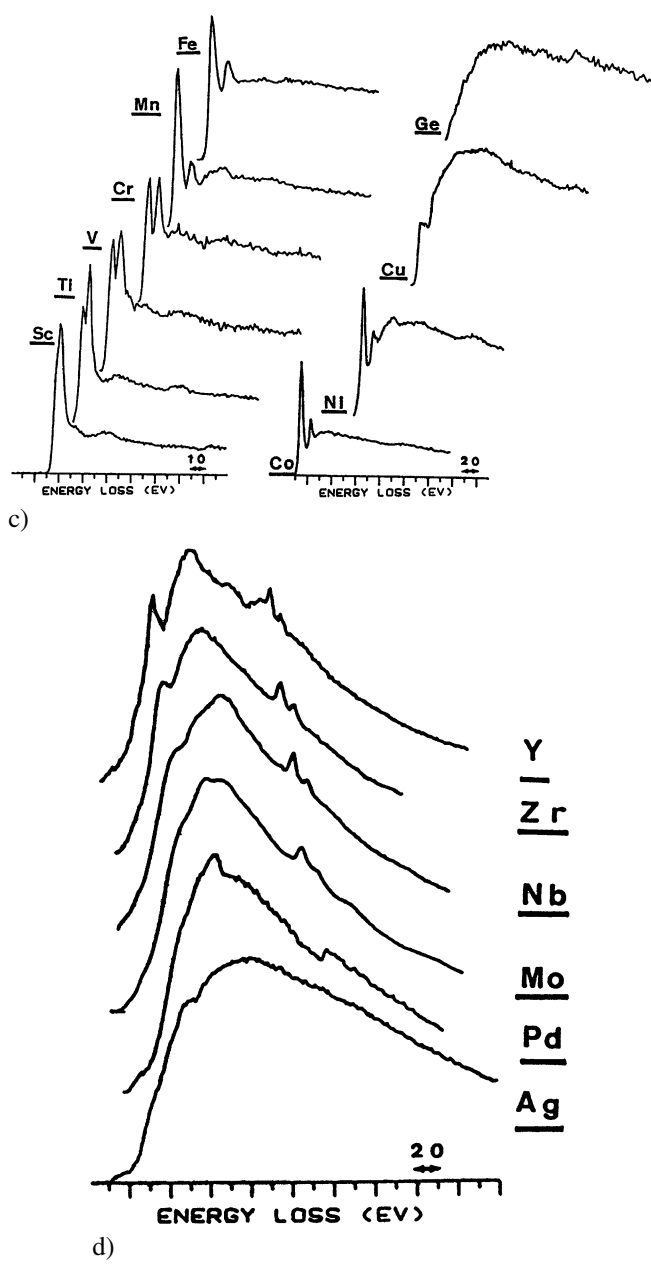


Figure 4-45. *Continued* (c) Trends in L₂₃ edge intensities (called white lines) for the transition metal series. It is clearly visible that the strong peaks at the edge threshold decrease in intensity with respect to the continuum after the edge threshold. This effect reflects the filling of the 3d transition metal band with electrons. (d) Shape of delayed M edges (Zaluzec, 1982, 1984). (Courtesy of N. Zaluzec.)

the so-called “delayed edges” (Figure 4-44b) with a broad maximum of the edge intensity 10–20 eV above the edge threshold. This delayed shape is predicted by cross-sectional calculations and is due to a maximum in the effective atomic potential resulting in a “centrifugal barrier” that prevents the overlap between the core and the broad final

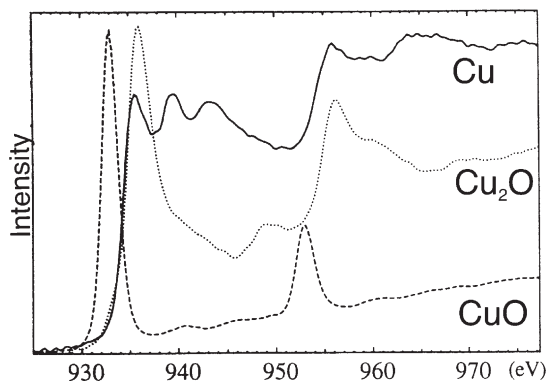
Table 4-4. Summary of Expected Edges.

Edge	Level	Elements	Shapes
K	1s	Li-Si	a
L ₂₃	2p _{1/2} , 2p _{3/2}	Mg-Ar	b
		K-Ni	c
		Cu-B	b
		K-Cu	d
M ₂₃	3p _{1/2} , 3p _{3/2}	K-Cu	d
M ₄₅	3d _{3/2} , 3d _{5/2}	Se-Kr	d
		Rb-I	b
		Cs-Yb	c
		Lu-Au	b
N ₄₅	4d _{3/2} , 4d _{5/2}	Cs-Yb	d
O ₄₅	5d _{3/2} , 5d _{5/2}	U,Th	d

state wavefunctions at low energy from the threshold. This poor overlap results in low values of the matrix element near the edge threshold (see Section 8.1). When the final state wavefunctions are very localized, as in the case of the $3d$ electrons, they are contained within the barrier and there is therefore a strong overlap with the core wavefunctions. This results in strong features at the edge threshold for the L₂₃ edges in the series K to Ni where the $3d$ band is being filled and transitions to empty $3d$ electron states are possible. In these cases, sharp features, known as “white-lines,” are clearly visible leading to noticeable trends in the transition metal series (Figure 4-45). For the Cu L₂₃ edge, where the $3d$ band is full with 10 d electrons, the delayed maximum is again observed without strong peaks at the edge threshold. When solid-state effects result in changes in the electronic configuration, however, for example, when Cu loses $3d$ electrons when bonding to O (such as CuO or Cu₂O), white lines do appear in the spectra (Figure 4-46). In Zn and its oxide ZnO, however, the electrons affected by bonding are of s character and the changes with bonding are not visible in the L edge of Zn. White line features in L₂₃ edges are observed for the $4d$ transition metals (the series Rb to Pd) at higher energy losses.

The M₄₅ and M₂₃ edges are the strongest in the M series with large cross sections. These edges exhibit the same trends with delayed

Figure 4-46. Cu L₂₃ edge in metallic Cu, CuO, Cu₂O. The presence of the white line in CuO and Cu₂O (as compared to metallic Cu) reflects the transfer of electrons from the $3d$ band to O atoms in the oxidation.



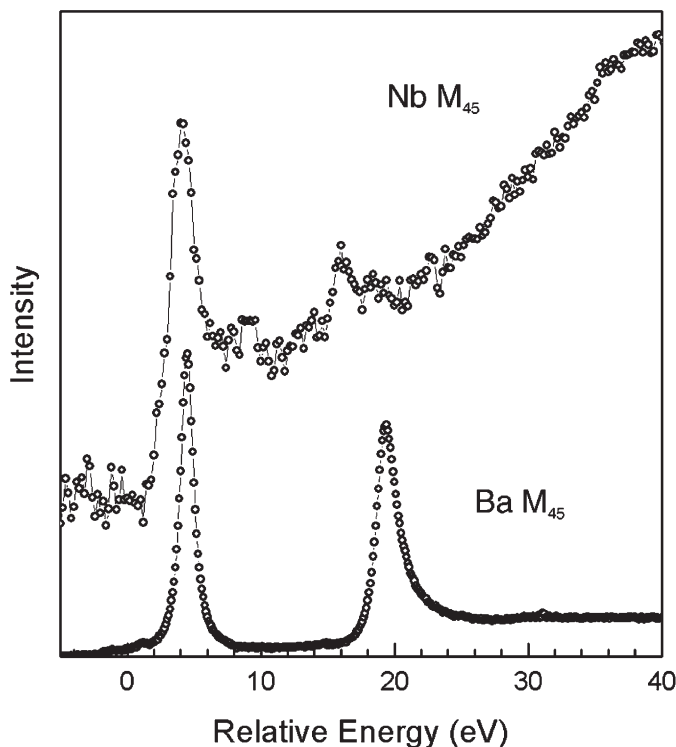


Figure 4-47. Nb M₄₅ and Ba M₄₅ edges. While the Ba edge shows strong peaks and a lower continuum past the edge onset, the Nb M₄₅ edge shows a strong rising intensity at the postedge threshold.

maxima and strong white lines depending on the atomic number (Figure 4-47). For the series Rb to In, a delayed maximum is observed due to transitions to delocalized empty *f* states well above the Fermi level while strong white lines are visible for Cs to Yb since the *f* states are bound and are being filled in the series. The very strong white lines make these edges very suitable for detection of trace quantities of elements in samples. For Lu to Au, as the *f* band is completely filled, very broad delayed edges are present and their detection, due to the background in the spectra, is often difficult even for elements present in samples in large concentrations. M₂₃ edges are detected for the K to Zn series in the energy range between 30 and 100 eV with plasmon-like features represented in Figure 4-44d. These edges overlap with the strongly varying background in the low portion of the energy loss spectrum and are thus difficult to extract from the spectra. Due to the large cross sections, however, elemental maps can be easily obtained with very high spatial resolution (Freitag & Mader, 1999).

3.6 The Background in EELS and EDXS

The background in core level energy loss spectra (Figure 4-10) is dominated by lower energy loss processes such as lower core losses and plasmon/outer shell losses and the characteristic signal of edges must be extracted with procedures described in Section 4.2.1. In the case of

EDXS, the continuum background depicted in Figure 4–6c arises from the “bremsstrahlung” radiation (or braking radiation) as the electrons are subjected to centripetal acceleration forces in the Coulomb potential of the atoms in the solid as their trajectories are modified. The X-rays are generated over a continuous energy range from 0 eV up to the energy of the incident electrons. The angular distribution of the emitted X-ray due to the bremsstrahlung radiation is not isotropic as in the case of the characteristic radiation emitted following ionization and deexcitation. The distribution is almost forward-peaked with maxima in intensity a few degrees from the forward direction and very little intensity emitted in the backscattered direction. Hence the peak over background ratio is dependent on the location of the EDXS detector (lowest peak-over-background with higher elevation angles) although additional difficulties in the acquisition of spectra arise due to the solid angle limitations and the instrumental contribution of the spectra (see Section 7.2). Expressions describing the continuum spectrum intensity (I_{cm}) demonstrate the dependence on the incident current and the average atomic number $\bar{Z}$ of the sample irradiated by the electron beam (Kramers, 1923)

$$I_{\text{cm}}(E_v) \propto i_b \bar{Z} \left(\frac{E_0 - E_v}{E_v} \right) \quad (35)$$

where i_b is the beam current, E_0 is the incident energy of the electrons, and E_v is the energy of the emitted radiation. At lower energy, the recorded spectrum does not show the increase in the intensity expected from Eq. (35) due to the absorption in the sample and in the detector. Although the generated continuum spectrum does not show sharp features, in crystalline solids well-defined peaks can occur due to coherent bremsstrahlung reported by Reese et al. (1984) when electrons travel close to a zone axis. The coherent bremsstrahlung energy peak positions E_{CB} are related to the spacing of the atomic planes in the direction of the electron beam (L), the ratio of the velocity of the electrons/velocity of light (v/c), and the elevation angle α_E (Figure 4–23) as described by Spence et al. (1983) and Reese et al. (1984)

$$E_{\text{CB}}(\text{keV}) = \frac{12.4(v/c)}{L[1 - (v/c)\cos(90 + \alpha_E)]} \quad (36)$$

These peaks can be mistaken for minor constituent phases when they overlap in energy with characteristic peaks and can be minimized by carrying out analyses when the samples are oriented far from major zone axes.

4 Quantification

4.1 EDXS Microanalysis

4.1.1 Quantification in EDXS

As indicated in Figure 4–7, there are several processes that need to be considered in the measurements of concentration based on the detected signals. First, X-rays generated by the excited atoms must travel through the specimen and are therefore subjected to absorption. These X-rays

can also cause possible fluorescence of secondary X-ray radiation within the sample, which generates additional signals further away from the site irradiated by the primary electrons. The subsequent detection process includes transmission through a detector window, various detector layers, and generation of electron-hole pairs in the solid-state detector device. All these effects need to be considered in the quantification procedure. We shall consider the different processes starting with the generation of X-rays in the sample assuming that the characteristic signals in the spectrum can be extracted from the continuum background by simple interpolation of the background intensity under the characteristic peak from the background on the low and high energy sides of the characteristic peak.

As electrons travel through a sample of thickness t and are scattered at various angles according to the elastic cross section (Sections 3.1 and 3.2) losing energy as dictated by the inelastic cross sections and the stopping power (Sections 3.1, 3.3, and 3.4) they generate photons in the X-ray portion of the electromagnetic spectrum. The generated X-ray intensity for element A is

$$I_A^* \propto (C_A \omega_A \sigma_A a_A t) / A_A \quad (37)$$

where

C_A = weight fraction of element A,

ω_A = fluorescence yield for the K, L, or M line,

a_A = fraction of the total K, L, or M line intensity that is measured,

A_A = atomic weight of element A,

σ_A = total ionization cross section for a shell K, L, or M for atom A in the specimen, and

t = the sample thickness.

The total ionization cross section derived in Section 3.4 can be expressed in terms of the overvoltage $U = (E_0/E_C)$ relative to the ionization energy E_C and incident energy E_0 commonly used in the AEM literature where the constants have been included into one single term and N , b , c are the same constants found in Equation (33) with subscripts related to the shell s , i.e.,

$$\sigma = Q = \frac{6.51 \times 10^{-20}}{E_C^2} N_s b_s \ln(c_s U) \quad (38)$$

When the thickness of the sample satisfies the thin-film criterion (Section 4.1.2), the absorption and fluorescence in the sample can be neglected. In this condition, the X-ray intensity still remains to be corrected for the detection process that includes absorption in the detector's various layers and for the efficiency of the detector active layer in generating electron-hole pairs from the impinging X-rays (Sections 2.3.1 and 2.3.2). For each particular X-ray peak, the intensity detected must therefore include a detector efficiency term that describes how the intensity is absorbed

$$I_A = I_A^* \epsilon_A \quad (39)$$

where

$$\epsilon_A = \{ \exp[-(\mu/\rho)_W^A \rho_W X_W - (\mu/\rho)_{Au}^A \rho_{Au} X_{Au} - (\mu/\rho)_{Si}^A \rho_{Si} X_{Si}] \} \\ \{ 1 - \exp[-(\mu/\rho)_{Si}^A \rho_{Si} Y_{Si}] \} \quad (40)$$

and

(μ/ρ) = the mass absorption coefficients for elements A in the window material W, the Au contact layer, and the Si,

ρ = the densities of the window material W, Au surface contact layer, and Si dead layer,

X = the thickness of the window W, the Au layer, and the Si dead layer, and

Y_{Si} = the EDXS detector active layer thickness,

The second term accounts for the detector thickness that generates the signal based on the electron-hole pairs created by the incident photons (the active layer thickness).

Although these equations can, in principle, be easily calculated, exact values of the various constants, sample thickness t , and detector layer thicknesses are not known to sufficient accuracy and alternative approaches have been suggested to eliminate the influence of these parameters on quantification. The calculation of the relative concentration of elements (C_A/C_B) makes it possible to include all the constant terms both in the cross sections and detector response into one single term and thus eliminates the need to know the sample thickness

$$\frac{C_A}{C_B} = \left[\frac{(\sigma\omega a/A)_B \epsilon_B}{(\sigma\omega a/A)_A \epsilon_A} \right] \frac{I_A}{I_B} \quad (41)$$

The term within brackets is a constant for a given set of X-ray peaks for elements A and B, detector, and accelerating voltage. This term allows the user to relate the intensity ratio for two peaks to the relative concentration of the elements in the sample. The k_{AB} factor approach suggested by Cliff and Lorimer (1975) makes it possible to simplify the calculation of relative concentration of elements provided that standards can be obtained so that

$$\frac{C_A}{C_B} = k_{AB} \frac{I_A}{I_B} \quad (42)$$

For binary systems, the concentration can be retrieved assuming $C_A + C_B = 1$ and for ternary systems a similar assumption can be made. In principle, the k_{AB} term in Eq. (41) can be calculated based on knowledge of the terms within the bracket but the significant advantage of the Cliff-Lorimer approach resides in the fact that the k_{AB} factors can be obtained from a standard of known composition C_A and C_B related to one reference element. As a convention, k factors can be expressed in terms of a standard with respect to Si and Fe (i.e., k_{ASi} , k_{BSi} or k_{AFer} , k_{BFe} , etc.) so that any set of k_{AB} factors can be retrieved by

$$k_{AB} = \frac{k_{ASi}}{k_{BSi}} \quad (43)$$

Reference k factors can be developed with respect to any elements but references to Si and Fe were historically developed for users in the

geological science and metallurgy communities. Examples of k_{AFe} factors are shown in Figure 4–48 and can be used as a good approximation to quantify measurements, although the exact values are dependent on a particular detector (and in particular detector efficiency) and accelerating voltage. Some tabulated values for a range of elements can be found in Williams and Carter (1996). Users of EDXS in AEM typically develop their own standards based on samples of known composition C_{A} , C_{B} to quantify unknown samples containing the same elements. These k_{AB} factors would therefore account for the specifics of the microscope-detector system. Standardless k_{AB} factors can be calculated based on the knowledge of all the terms expressed in Eq. (41) making use of constants tabulated in the literature.

The fluorescence term can be derived based on the parameterization

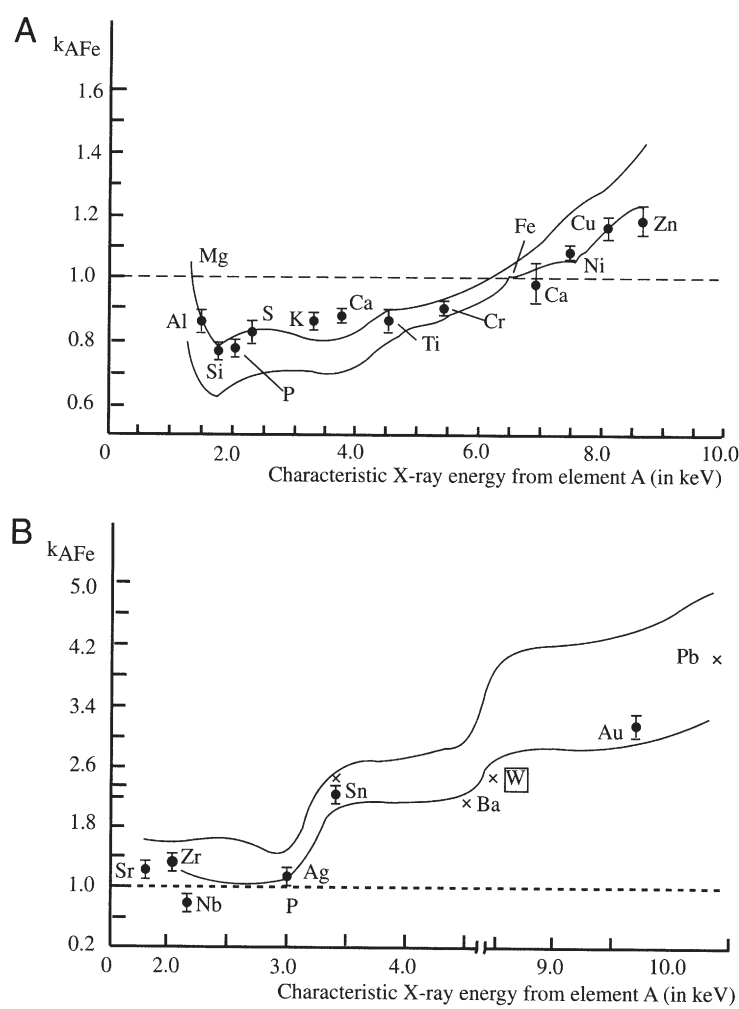
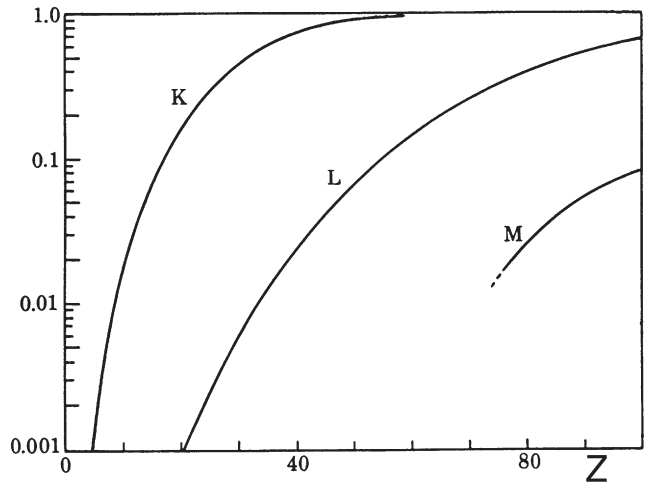


Figure 4–48. Experimental k_{AFe} factors for a range of elements (top) for K α lines and (bottom) for L α lines. The limits reflect the range of theoretical values obtained with different cross-sectional parameterizations. (From Williams and Carter, © 1996, with permission from Springer Science+Business Media.)

Figure 4–49. Fluorescence coefficient for families of lines as a function of atomic number. (From Egerton, © 1996, with permission from Springer Science+Business Media.)



$$[\omega/(1 - \omega)]^{1/4} = A + BZ + CZ^3 \quad (44)$$

where the constants for K, L, and M edges have been summarized in Goldstein et al. (1986) and they give rise to the dependence shown in Figure 4–49.

Parameterization of the cross sections and the constants found in Eqs. (37) and (41) have been discussed in detail in Goldstein et al. (1986) and show that relative errors in k_{AB} factors in the range of 1–5% for K lines and higher for L lines (10–15%) can be expected based on the various sources of data and estimates of detector parameters as compared with experimental data. Quantification packages in modern AEM systems provide accurate quantification based on more precise knowledge of the internal parameters for the detector often not accessible to the end user. Verification of the quantification procedure is, however, suggested based on standards of known composition to validate the built-in procedures of the EDXS software.

4.1.2 Absorption

The simple expression for the quantification of EDXS spectra with k_{AB} factors based on standards, or with the standardless approach discussed above, assumes that the sample is thin enough that the absorption and fluorescence terms can be neglected, i.e., the thin-film criterion is applicable. Photons generated by the incident electrons, however, are absorbed by atoms in the sample via electron excitation processes while traveling from their origin point to the detector. As in the case of absorption of X-rays occurring in the EDXS detector, the strength of the absorption process is related to the X-ray absorption coefficient (tabulated in Goldstein et al., 1981) for a given X-ray line and a specific absorbing element. The intensity of the incident radiation (and not the energy) will decrease with an exponential decay (Figure 4–50) while it travels through a sample. Large absorption coefficients occur when low-energy X-ray lines are just above the absorption thresholds of the absorbing element, i.e., when the photoabsorption cross sections are high. At the other extreme, small absorption coefficients occur for high-

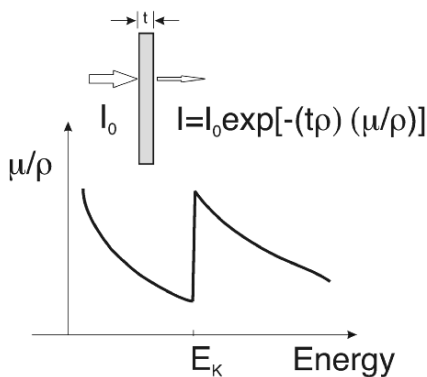


Figure 4–50. Absorption process of X-rays and variation of the absorption coefficient with energy.

energy lines far away from the absorption threshold of the absorbing element. Considering these effects, the thin film criterion stipulates that the k_{AB} factors should not change more than the arbitrary value of 3% (or less stringent 10%) due to absorption.

If the generation of X-rays is uniform throughout the sample thickness, the k_{AB} factor is modified by the absorption correction factor (ACF) term:

$$k_{AB} = k_{AB}^{\text{ThinFilm}} (\text{ACF})$$

$$k_{AB} = k_{AB}^{\text{ThinFilm}} \left[\frac{(\mu/\rho)_{\text{spec}}^A}{(\mu/\rho)_{\text{spec}}^B} \right] \left[\frac{1 - \exp\left\{ -(\mu/\rho)_{\text{spec}}^B \cos \alpha (pt) \right\}}{1 - \exp\left\{ -(\mu/\rho)_{\text{spec}}^A \cos \alpha (pt) \right\}} \right] \quad (45)$$

where k_{AB}^{ThinFilm} is the k_{AB} factor for the thin film without the absorption term, μ/ρ is the absorption coefficient for element A or B in the sample, and α is the X-ray elevation (or take-off) angle required to calculate the pathlength of X-rays in the sample (Figure 4–51).

In practical cases, it is clear that the 3% limit to the absorption correction is very stringent and severely limits the practical applications of the Cliff and Lorimer approach, thus forcing the use of the correction term and the measurement of the sample thickness. Examples of calculations of the thickness satisfying the thin film criterion [Eq. (45)] are given in Table 4–4 for the 10% and 3% criteria (Goldstein et al., 1986). By inspection, it is possible to note that to avoid absorption corrections, extremely thin samples are required when low-energy X-ray lines are generated in the sample and when there are elements with absorption thresholds in proximity of the X-ray lines. For higher energy photons (and larger separation between the photon energy and the

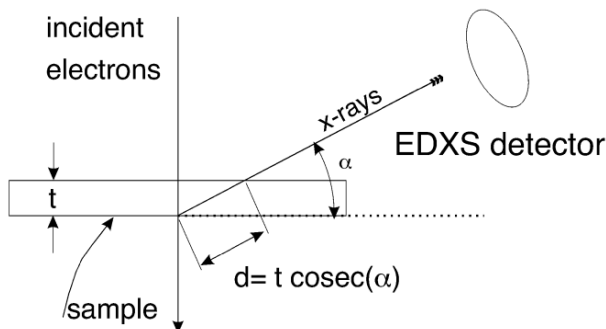


Figure 4–51. Absorption length dependence on elevation angle and foil thickness.

Table 4–5. Thickness criteria for the 10% and 3% absorption limit.

Material	Thickness for		Absorbed X-ray line
	10% limit	3% limit	
Fe–5% Ni	322	89	NiK α
MgO	304	25	MgK α , OK α
NiAl	32	9	AlK α
SiC	13	3	SiK α , CK α
Si ₃ N ₄	413	6	SiK α , NK α

absorption threshold), absorption can be neglected in typical AEM samples. To illustrate this effect, Table 4–5 indicates that an NiAl specimen of 9 nm thickness would satisfy the 3% criterion, while a Fe–5% Ni alloy would require an $\approx$ 90-nm-thick foil.

The calculation of the absorption correction factor implies that absorption coefficients are known, the density of the sample is known, and that the thickness of the sample can be measured at the point of analysis so that the pathlength of the X-rays can be determined. Although tabulated values for the mass absorption coefficients for pure elements can be easily found in the literature (e.g., Goldstein et al., 1986), the knowledge of the actual sample mass absorption coefficient and sample density means that the composition of the sample should be calculated in an iterative process where initial input of composition, and thus density and mass absorption coefficient values, is refined at each iteration step. An additional complication resides in the requirement to calculate the mass absorption coefficients based on all elements present in the sample even if only relative concentrations are of interest. The mass absorption coefficient of the sample is determined from

$$(\mu/\rho)_{\text{spec}}^A = \sum_i C_i (\mu/\rho)_i^A \quad (46)$$

where the composition C_i for each element is used. All elements where absorption is potentially significant should be considered, particularly for low-energy peaks in the presence of light elements where absorption is important. Similar difficulties arise with the determination of the sample density that must be deduced following a weighed average based on the concentration of the pure elements (whether the elements are detected/analyzed or not). Given the complexity of novel synthetic structures produced by chemists and physicists and materials scientists, extrapolating the density of materials is not a trivial task for some samples and the best approach to quantification is to avoid thick regions where absorption correction is required.

An additional parameter required for absorption correction is the sample thickness that must be determined accurately. Various approaches based on convergent beam electron diffraction (Kelly et al., 1975; Williams and Carter, 1996), contamination spot measurements (Goldstein et al., 1986; Williams and Carter, 1996), and electron energy loss spectroscopy (Section 8.2 in this chapter) exist to retrieve the sample thickness. Their accuracy, however, is approximately 5–10% at

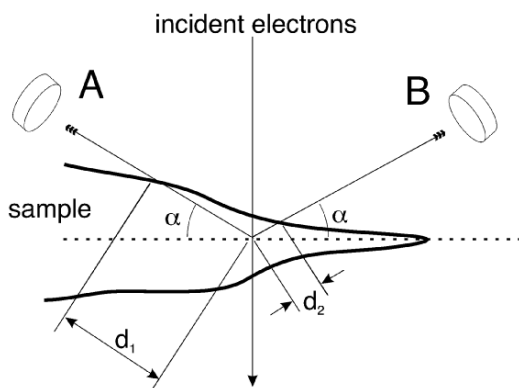


Figure 4-52. Absorption path for two configurations of sample with respect to the detector. When the sample is rotated so that the detector is in position A (relative to the sample), the absorption length is significantly longer than when the detector is in position B (relative to the sample). (Adapted from Goldstein et al., 1986.)

best. An additional difficulty in the evaluation of the absorption path-length, from the point of origin to the exit point of the sample, is the knowledge of the sample geometry and the position of the detector relative to the point of analysis. This is demonstrated in Figure 4-52 where the sample is tilted for analysis or when the area of analysis is not suitably positioned with respect to the detector (this might be necessary to achieve a suitable contrast condition in the image or to align an interface so that the electron beam is parallel to the interface plane). Significant differences in absorption corrections, and therefore relative concentration, can occur if careful consideration is not given to these issues during experiments (Glitz et al., 1981; Goldstein et al., 1986). Errors in concentration up to a factor of two have been demonstrated if the detector position is not known during the analysis (Glitz et al., 1981). Similarly, quantitative analysis of interfaces and grain boundaries should be carried out so that the X-ray path from the analysis point to the detector and the electron beam are contained in the same plane. This approach simplifies all correction procedures since the absorption is considered within the boundary itself and the X-rays are generated by the electron beam traveling within the same material (Figure 4-53).

These difficulties in accurately determining the correction coefficients, density, exact geometry of analysis, and the detector efficiency, particularly at low energies, suggest that quantitative work on carbides, nitrides, and oxides is particularly prone to inaccuracies and error propagation so that complementary quantitative measurements with EELS should be carried out whenever possible. For higher energy peaks (such as transition metal K lines) absorption is significantly lower and

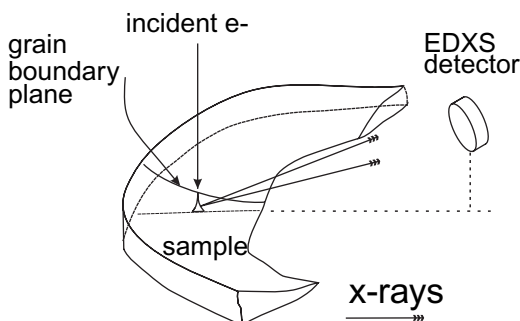


Figure 4-53. Ideal geometry for analysis of interfaces and grain boundaries. The electron beam and the absorption path of X-rays from the generated area to the detector are all contained within the plane of the boundary.

can more easily be neglected in TEM measurements on typical thin foils.

Alternative approaches to retrieve the composition of TEM samples while considering the absorption effects are given by the extrapolation techniques developed over the past decade. These approaches involve sequences of measurements at increasing thicknesses as discussed in Section 4.1.4.

4.1.3 Fluorescence

Fluorescence occurs when X-rays have sufficient energy to ionize atoms and new photons are generated by the deexcitation process. For example, the FeK α (6.4 keV) can generate fluorescence of Cr (5.9 keV) X-rays but cannot excite MnK (6.5 keV) lines. However, the FeK β (7.0 keV) line can generate the MnK X-rays (although the contribution would be small since the intensity of the K β is about 14% of the K α). At low energies, fluorescence can be important given the proximity of several lines. Although it is possible to generate fluorescence far away from the point of origin, fluorescence in thin specimens is not as important as in bulk samples for the simple reason that there is less material to fluoresce. The original developments required to correct for fluorescence effects are due to Philibert and Tixier (1975). Their model considered that the X-rays are generated from the middle of the tin foil and predicted small changes in intensities due to fluorescence. Improvements were later proposed by Nocholds et al. (1979) to consider uniform emission of X-rays in the sample including tilted sample geometries when element B causes fluorescence of element A:

$$\frac{I^A}{I_A} = C_B \omega_B [(r_A - 1)/r_A] \frac{A_A}{A_B} (\mu/\rho)_A^B \frac{E_{CA}}{E_{CB}} \left(\frac{\ln E_0/E_{CB}}{\ln E_0/E_{CA}} \right) \frac{\rho t}{2} \left\{ 0.932 - \ln [(\mu/\rho)_{\text{spec}}^B \rho t] \right\} \sec \alpha \quad (47)$$

where

- I^A/I_A = the ratio of fluorescence intensity to primary intensity,
- ω_B = fluorescence yield of element B,
- r_A = absorption edge jump ratio of element A,
- $(\mu/\rho)_A^B, (\mu/\rho)_{\text{spec}}^B$ = mass absorption coefficients of X-rays from element B in element A, and the specimen,
- A_A, A_B = atomic weights of elements A and B, and
- E_{CA}, E_{CB} = critical excitation energy for the characteristic radiation of A and B.

As in the case of the absorption corrections, prior knowledge of the sample composition (in this case of element B) is required for the calculations, thus resulting in an iterative process to estimate the correction term. This expression proved effective in correcting for fluorescence in very thick samples by modern AEM standards (Figure 4–54). Additional corrections for the fluorescence caused by the continuum radiation are presently considered to be negligible.

4.1.4 Correction Techniques

The k factor expression [Eq. (42)] assumes that the sample is infinitesimally thin and corrections are not required. The thin-film criterion

discussed above states that corrections are not required if the k_{AB} factors do not change by more than about 3%. This implies that estimates of the absorption correction factors using Eq. (45) are necessary if the thickness does not satisfy this criterion and that estimates of the sample thickness for the points of analysis are required. If the ACF is greater than 1.03 or smaller than 0.97, corrections must be carried out and exact measurements of the thickness are necessary.

Similarly, fluorescence corrections are required if the X-rays of one of the elements generate significant X-ray emission from other elements present in the sample. The fluorescence correction term proposed by Nocholds et al. (1979) with estimates of the thickness is used to calculate the ratio of fluoresced intensity to primary intensity I^A/I_A for element A (and for the other elements present). When the correction term is greater than 5%, the thin-film criterion is not met and the correction is required. If both the thin-film criteria for absorption and fluorescence are not satisfied, quantification requires the related corrections

$$\frac{C_A}{C_B} = k_{AB} \frac{I_A}{I_B} (ACF) \frac{1}{(1 + I^A/I_A)} \quad (48)$$

where it is assumed that element A is fluoresced by element B and the thickness is calculated at each analysis point. As discussed in Sections 4.1.2 and 4.1.3, the calculations require “prior” knowledge of the sample composition, thus leading to the use of iterative techniques with initial assumptions of the composition and refinement of the compositions until there is agreement between the input and output concentrations.

Various approaches have been proposed over the past few years to simplify the quantitative analysis of samples without prior knowledge of the sample thickness. These methods rely on the extrapolation of the composition for a “zero-thickness” sample based on several measurements obtained from areas of different (but unknown) thicknesses. These measurements would require a constant acquisition time and electron beam current during the series of measurements and a uniform sample concentration in the areas of analysis. The original approach was proposed by Horita et al. (1987) who suggested the use of measure-

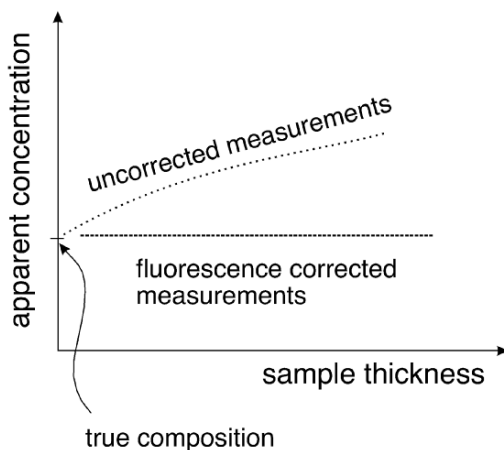
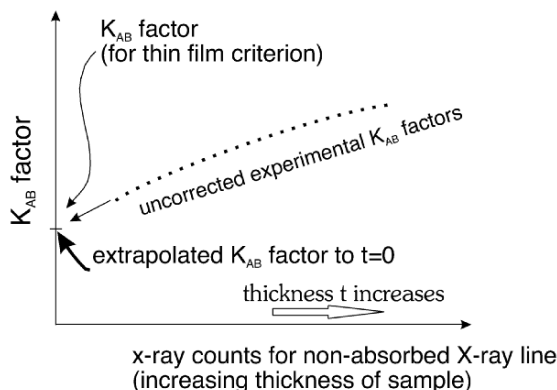


Figure 4-54. Effect of fluorescence correction on the quantification of spectra as a function of thickness. The raw data show strong thickness dependence while corrected data are shown with Eq. (47) is independent of thickness.

Figure 4-55. Extrapolation procedure using the method proposed by Horita et al. (1987). The k_{AB} factor is plotted as a function of the integrated number of counts of an X-ray line (not absorbed in the sample) as an indirect indicator of the thickness of the sample. Extrapolation to zero counts leads to the determination of the k_{AB} factor of a thin sample. The nonabsorbed X-ray line must be of sufficiently high energy and far from any absorption edge.



ments of k_{AB} factors at different sample thickness. By plotting the k_{AB} factors as a function of the intensity of one high-energy X-ray line in the spectrum and extrapolating the curve to zero intensity, an “absorption and fluorescence-free” k_{AB} factor can be obtained (Figure 4-55). This approach is based on the assumption that the high-energy X-ray line is not strongly absorbed and that its intensity is therefore directly related to the sample thickness [Eq. (37)]. This method can be used to determine “zero-thickness” k_{AB} factors of standard samples and the composition of unknown samples following the extrapolation of X-ray line intensity ratios to zero thickness. Although extremely useful for the quantification of homogeneous samples, the extrapolation techniques are not suitable for the determination of the composition of structures where the variations in thickness cause overlap of phases throughout the analyzed area. For example, when the composition of precipitates in a thin foil is required, only areas where the precipitates do not overlap with the matrix can be analyzed.

Various modifications to the Horita technique were proposed later by Van Cappellen (1990) for work on alloys, Eibl (1993) and Van Cappellen and Doukhan (1994) for work on oxides using the charge neutrality concept, and in light element analysis demonstrated by Westwood et al. (1992).

4.2 Quantification in EELS

4.2.1 Quantification Procedures

The atomic cross sections (Section 3) describe the probability of excitation of an atom by a fast incident electron. The energy dependence of these cross sections provides a basic description of the shape of the edges (without any solid-state and bonding effects) while the partial cross section integrated for a given collection angle and energy window allows the intensity recorded in the spectrum to be related to the incident beam current and the number of atoms excited by the primary electron beam. The recorded signal for a particular edge, however, is superimposed on a large background due to lower energy excitations (including collective excitations and lower energy ionization edges). Therefore, even for elements present in the sample in large concentrations, the background can constitute the major contribution to the

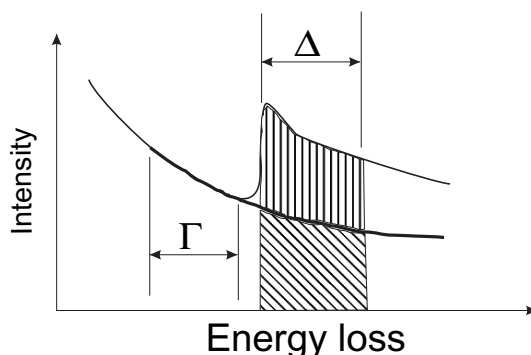


Figure 4-56. Background extrapolation of EELS edges. Fitting of the background is carried over a window Γ and integration is carried out over an energy window Δ .

intensity at a given energy loss (Figure 4-10) and must be removed, prior to the integration of the signal of the characteristic edge, using an extrapolation technique (Figure 4-56). The extrapolation is based on a power-law dependence of the background derived from atomic physics (Section 3.4.2)

$$I_b(E) = AE^{-r} \quad (49)$$

where A and r can be considered as constants within a small region of the spectrum and their values can be retrieved based on a fit of the spectrum over an energy window prior to the edge of interest. The A parameter is strongly dependent on the intensity of the spectrum (hence, the current, recording time, collection efficiency, etc.) whereas the r parameter (related to the slope of the spectrum) varies between the values of 2 and 6 and is mostly dependent on energy loss and angular collection conditions. The fit and extrapolation of the background are carried out over energy windows (Figure 4-56) using least-squares fitting after transformation of the intensity and energy to a logarithmic scale so that the intercept of the fit relates to the A parameter and the slope to the r parameter. Alternatively, nonlinear least squares can also be used. Once the A and r parameters are found in the fitting region, the extrapolation makes it possible to subtract the background and to extract the characteristic edge (Figure 4-56). This power-law extrapolation assumes that the contributions from multiple inelastic losses to the spectrum in the region of interest are small. For typical samples having thicknesses in the order of 50–100 nm, this requirement is relatively easy to meet at high energy losses (a few hundred electronvolts and higher) but low energy edges (<80 eV) such as AlL_{23} and LiK are particularly difficult to analyze due to the rapid variation of r and the contributions of multiple scattering (see Section 8.2). Very thin samples are thus required and more advanced approaches able to deal with the steep background and rapid variation of the r parameter for extrapolation are necessary.

Polynomial-based extrapolation and/or constraint power-law (forcing the background to cross the spectrum at a given energy) can be used to deal with numerical instabilities arising from the extrapolation. To remove the complications due to multiple scattering when extracting low-energy edges, deconvolution of the spectrum using the Fourier-log technique is necessary (see Section 8.2). Fitting approaches

using reference spectra and multiple scattering convolution with low-loss spectra have been developed recently (Verbeeck and Van Aert, 2004) to deal very successfully with weak and/or overlapping edges (Figure 4–57). Signal extraction using difference techniques (Egerton and Leapman, 1995) has also been used to analyze very weak signals for elements present in trace concentrations (Leapman and Newbury, 1993; Shuman and Somlyo, 1987). A variant of this latter method uses a numerical difference technique (Zaluzec, 1985). This technique is relatively insensitive to the sample thickness. Following the extraction of the background, integration of the signal over an energy window Δ must be carried out (Figure 4–56) to relate the signal to the number of atoms present under the electron beam.

Assuming that single scattering is dominant (i.e., one single loss event occurs when the electron travels through the sample), the integrated edge intensity, obtained for a given collection angle β and integration window Δ , is related to the number of atoms per unit area N , the unscattered intensity I_0 (the zero-loss peak is assumed to be nearly equal to the incident beam in a single scattering first approximation), and the partial cross section σ_k for the same collection and integration conditions

$$I_k^l(\beta, \Delta) = NI_0\sigma_k(\beta, \Delta) \quad (50)$$

Ga/As ratio of a cleaved wedge of GaAs

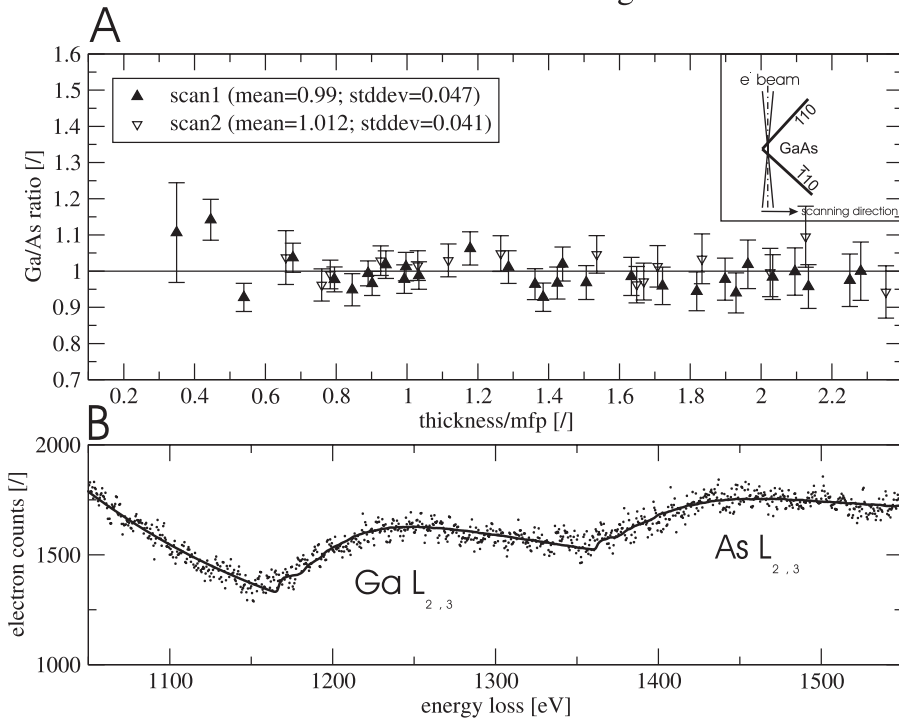


Figure 4–57. Model-based quantification of EELS of edges using maximum likelihood estimators of parameters of spectra and variation of quantification with respect to thickness (bottom panel). Models include thickness effects and cross sections. The top panel shows that quantification with this technique appears to be less sensitive to thickness variations based on a series of measurements with increasing thickness (relative to the inelastic mean free path) on a wedge sample (Reprinted from Microscopy, 101, J. Verbeeck and S. Van Aert, Model based quantification of EELS spectra, p. 207–224 © (2004), with permission from Elsevier.)

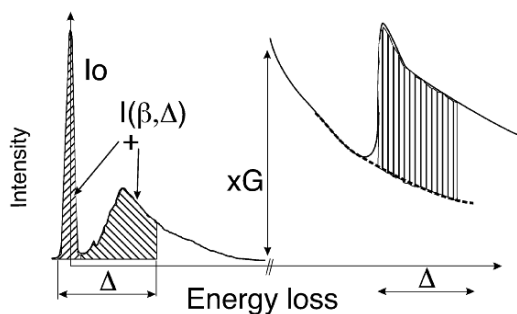


Figure 4-58. Extrapolation of edge signals and the integration windows used for quantification. The integration of the signal over the window Δ is carried out for the signal at the edge and for the low-loss part of the spectrum [based on Eq. (51)]. G is the value of the gain change (scale change factor) to visualize the high-energy part of the spectrum.

where the index refers to the single scattering assumption. More realistically, we must consider that plasmon interactions redistribute the intensity in the spectrum causing multiple inelastic processes visible both at low energy and at the core edges (i.e., core loss and plasmon scattering occurs as discussed in Section 8.2) (Figure 4-58). Furthermore, elastic scattering reduces the collected intensity by scattering electrons outside the collection aperture both for the apparent zero-loss intensity, relative to the incident beam intensity, and for the edge. In these conditions, we must integrate the spectrum intensity above the zero-loss peak so as to include inelastic scattering up to an energy loss Δ (Figure 4-58).

$$I_k(\beta, \Delta) = NI(\beta, \Delta)\sigma_k(\beta, \Delta) \quad (51)$$

Although this equation allows us to measure the absolute number of atoms per unit area, the more typical approach in EELS quantification is to use intensity ratios to deduce the relative concentration using

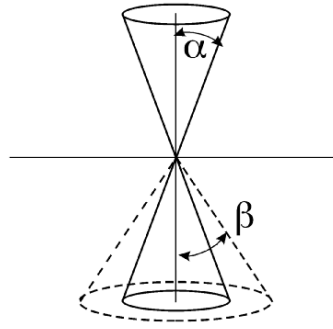
$$\frac{N_a}{N_b} = \frac{I_{ka}(\beta, \Delta)}{I_{jb}(\beta, \Delta)} \frac{\sigma_{jb}(\beta, \Delta)}{\sigma_{ka}(\beta, \Delta)} \quad (52)$$

In this case, the indices k and j relate to the fact that different types of edges can be used for quantification (K, L, or M). When multiple scattering is removed from the spectrum through the use of deconvolution techniques (Section 8.2) the expression remains the same but the extracted intensity is obtained from extrapolation of the deconvoluted spectrum.

4.2.2 Correction for Convergence

Equation (52) assumes that the incident beam is parallel and the angular distribution of scattering (represented by the Bethe surface in Figure 4-41) is not altered by any convergence of the incident beam. Corrections to account for convergence of the incident beam onto the sample, however, become necessary when the electron beam is convergent ($\alpha \geq 0.3\beta$) as in the case of analysis in STEM mode. In STEM configuration the convergence angle α typically exceeds this criterion (Figure 4-59). The relative concentration must be calculated by including convergence

Figure 4–59. Relation between the convergence and collection angles for quantification of spectra accounting for the incident beam convergence.



correction factors into Eq. (52) that can be deduced analytically (Egerton, 1996) for each particular edge. The relative concentration becomes

$$\frac{N_a}{N_b} = \frac{I_{ka}(\beta, \Delta) \sigma_{jb}(\beta, \Delta) F_{1b}}{I_{jb}(\beta, \Delta) \sigma_{ka}(\beta, \Delta) F_{1a}} \quad (53)$$

where $F_{1a,b}$ are the convergence correction factors that are dependent on the angular scattering distribution for a particular edge via the characteristic scattering angle θ_E . Correction factors must therefore be calculated for each edge of interest in the quantification and can be deduced from Figure 4–60 or from simple programs (Egerton, 1996). It is possible to note that for edges of very similar intrinsic angular distributions (i.e., two edges close in energy will have very similar θ_E) the

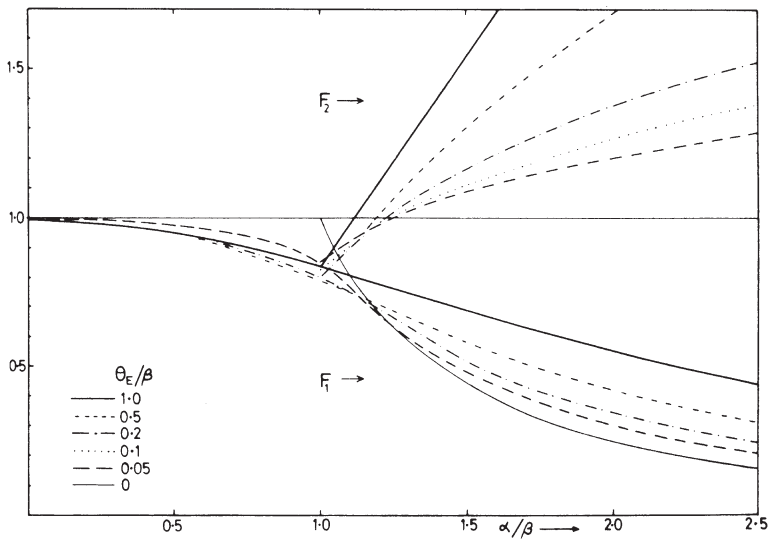


Figure 4–60. Correction factors F_1 useful to quantify spectra accounting for convergence of the incident beam α . The convergence factors F_1 must be determined for each edge based on the convergence angle α relative to the collection angle β . For each edge, the characteristic angle θ_E relative to the collection angle β must be used to identify the appropriate curve to be used in the quantification. The F_2 curves are used only for absolute quantification. (From Egerton © 1996, with permission from Springer Science+Business Media.)

ratio F_{1b}/F_{1a} is close to one and no effect of convergence is visible irrespective of the convergence angle.

For absolute quantification, the correction of the expression relating the number of atoms and the recorded intensity under the edge [Eq. (51)] becomes necessary irrespective of the intrinsic angular distribution

$$I_k(\beta, \Delta, \alpha) \approx F_2 NI(\beta, \Delta) \sigma_k(\beta, \Delta \alpha) \quad (54)$$

where the correction factor $F_2 \approx F_1$ for $\alpha \leq \beta$ and $F_2 = (\alpha/\beta)^2 F_1$ for $\alpha \geq \beta$ and can be deduced from Figure 4–60.

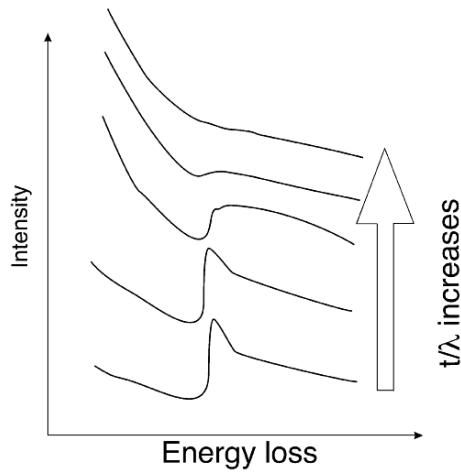
4.2.3 *k-Factor Approach*

All the expressions for relative quantification in EELS are similar to the k_{AB} factor equations in EDXS microanalysis. In a similar manner, a k -factor approach has been proposed (Malis and Titchmarsh, 1985) to quantify EELS spectra, although the experimental and sample conditions (convergence, collection, sample thickness) must be very strictly controlled and reproducible. Once the k -factor has been determined on the reference sample accounting for the cross-sectional ratios and possible convergence effects, the same k -factor can be used to deduce the composition of the unknown samples as in EDXS microanalysis. Cross section calculations using the currently available models (Hartree–Slater or hydrogenic, see Section 3.4) and convergence corrections are already built-in in commercial programs that control the acquisition functions of the spectrometers. An additional calculation approach to deduce the cross sections necessary for the quantification makes use of tabulations of the oscillator strengths obtained from optical data (see Section 3.4.1). These methods require the acquisition of spectra in dipole conditions (i.e., with small collection angles) and within the energy range where the angular distribution of scattering is still dominated by the Lorentzian term of Eq. (27) (i.e., near the edge threshold). The methodology is described in Egerton (1993) and summarized in Egerton and Leapman (1995).

4.2.4 *Limitations in Analysis and Quantification*

Although EELS quantification is not affected by absorption or fluorescence, there are major drawbacks in quantitative and even qualitative EELS analysis. The major difficulty is the strong effect of the sample thickness on the detection of EELS edges. For example, spectra from specimens of very simple composition a few tens of nanometers in thickness can reveal well-resolved edges while spectra from “thicker” areas (as thin as 100 nm) might show no edges (Figure 4–61). This severe limitation to the visibility of edges is due to the contribution of multiple inelastic scattering that increases the background under the edge. The contribution of multiple scattering, however, is not uniform as a function of energy loss and the quantification of spectra for increasingly thick samples demonstrates a variation of the apparent concentration with thickness. Systematic measurements of the ratio of two elements, for example, show that samples with thickness relative to the mean inelastic free path (Sections 3.4 and 8.2) $t/\lambda > 0.5$ are unreliable (Figure 4–62). Even when multiple scattering effects are removed with

Figure 4–61. Variation of the visibility of spectra with increasing thickness of the sample. Thick samples will not necessarily show EELS edges even for major constituents.



deconvolution techniques (Section 8.2), the effects persist (Egerton, 1996), suggesting that additional contributions due to the angular distributions of losses are present and cannot be neglected in correction approaches. Calculations that include the contributions of elastic scattering (Cheng and Egerton, 1993) and convolutions of the energy and angular distributions of the scattering angle (Su et al., 1995) demonstrate trends that are consistent with the experimental variation of the composition with thickness. Correction programs have been developed to account for these effects (Wong and Egerton, 1995). The lack of visibility of edges in spectra, even for pure elements, implies that users should be particularly cautious about drawing conclusions on the absence of elements during the analysis of point spectra and particularly with the application of energy-filtered imaging techniques. In the latter condition, since spectra are usually disregarded and the acquisition is done in an automatic procedure, erroneous results can often be obtained if care is not taken to check the sample thickness prior to the analysis of spectra. Relative thickness maps should therefore be used as a routine check prior to elemental imaging to verify the applicability of the $t/\lambda > 0.5$ condition. The procedure for the measurement of the relative thickness is described in detail in Section 8.2.

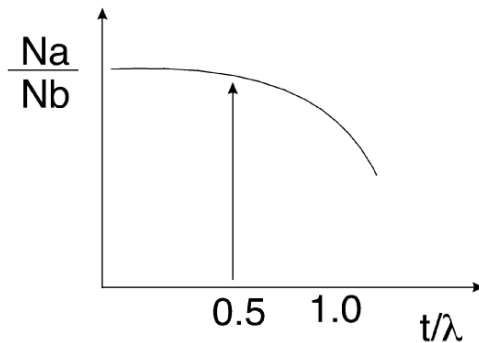


Figure 4–62. Variation in the quantification of two elements as a function of thickness relative to the total inelastic mean free path.

5 Resolution in Microanalysis

5.1 EDXS Microanalysis

Two contributions determine the spatial resolution in X-ray microanalysis. The first contribution arises from the electron beam diameter and the second from the electron beam broadening generated when electrons travel in the sample. The early analytical TEMs provided beam diameters of the order of tens to a hundred nanometers and the beam broadening within the sample was not a significant issue. Measurements using contamination spots demonstrated that the illuminated area on the top of the sample was essentially the same as the exit surface of the electron beam (e.g., see Stenton et al., 1981). The situation changed radically with the development of analytical instruments equipped with field emission sources capable of achieving probe sizes in the order of a nanometer or smaller (Section 2.1) whereby the ultimate limits in spatial resolution due to the intrinsic beam broadening could be probed.

To determine electron beam broadening, we must consider that the trajectory of the incident beam is controlled by an elastic scattering process that causes the deviation of the incident electrons as they travel through the sample. Transport equations (Rez, 1984) and detailed multislice calculations (e.g., Loan et al., 1988; Mobus and Nufer, 2003; Voyles and Muller, 2004; Dwyer and Etheridge, 2003; and Section 3.2) have been developed to describe electron beam propagation [including the impact of sub-Angstrom beams (Dwyer and Etheridge, 2003)] but more extensive work has been carried out in the field of AEM using Monte Carlo simulations that consider the individual trajectories of the incident electrons, the elastic scattering cross sections that modify the electron trajectories, and the inelastic scattering that causes the slowdown of the electrons (Figure 4–63). Monte Carlo approaches are easily applicable in complex geometries of samples (for example, interfaces and particles), although they neglect the effect of channeling of electrons in crystal and therefore assume the sample is amorphous or tilted away from a zone axis. The Monte Carlo technique is based on the generation of random numbers that are used to calculate the scattering angles [via equations of the elastic cross sections: Eq. (11)], the path-length between scattering events [using the elastic mean free path of Eq. (12)], and the energy loss between the scattering events [using the stopping power derived from the inelastic cross sections: Eq. (19)].

Electron trajectories simulated with the Monte Carlo method for thin films show the dependence of beam broadening on the accelerating voltage (Figure 4–64). The effect of the average atomic number and sample geometry can also be easily determined. With these simulations, it is possible, in principle, to evaluate the exit area and the volume containing an arbitrary fraction of electrons that will contribute to the generation of the X-rays signal and thus the spatial resolution. For example, the interaction volume containing 90% of scattered electrons is typically used as a reference in the AEM literature to determine the resolution, although more stringent criteria [with 95% of the electrons

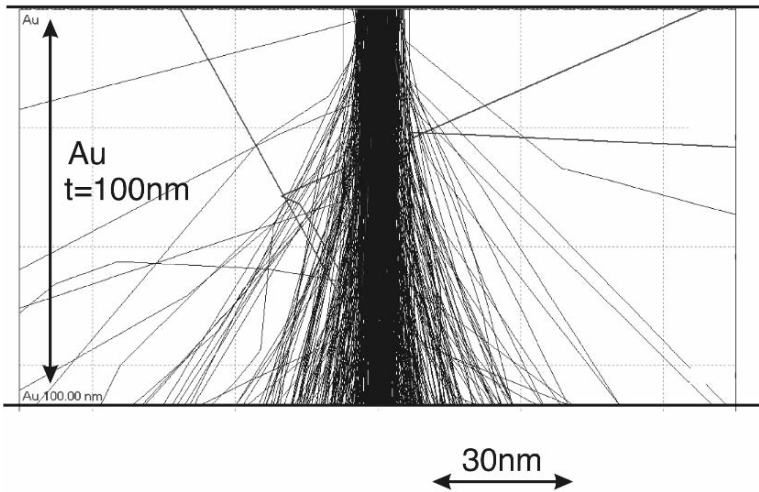


Figure 4-63. Simulation of the electron trajectories in a thin sample of Au (100 nm thickness) using the Monte Carlo method for 100-keV incident electrons. Calculations of electron trajectories are carried out using the public-domain code CASINO developed in the Gauvin group (Hovington et al., 1997) available on the web (<http://www.montecarlomodeling.mcgill.ca>).

(Faulkner and Norrgard, 1978) and 99% of the electrons (Reed, 1966)] have also been proposed. Although broadening values can be retrieved from the simulated trajectories in a few minutes of computation even on laptop computers or through web-based programs (Hovington et al., 1997), quick estimates of the electron beam broadening are necessary to evaluate the approximate loss of spatial resolution in microanalysis. Goldstein et al. (1977) developed a simple analytical model assuming a single scattering event in the middle of the foil thickness

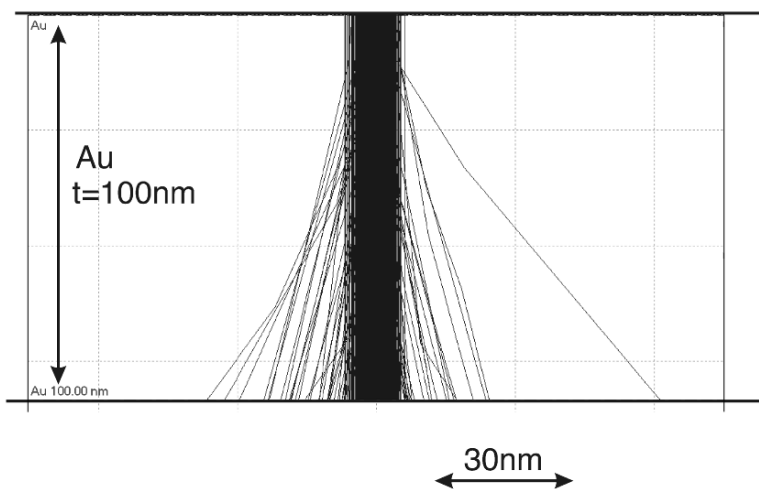


Figure 4-64. Electron trajectories in a 100-nm-thick Au Monte Carlo calculations for 300-keV incident electrons.

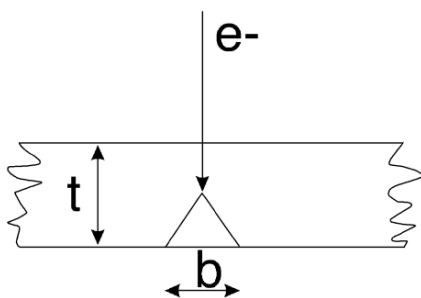


Figure 4-65. Simplified diagram describing the broadening of electronic samples. The parameter b describes the broadening at the bottom of the foil using Eq. (55) and the single scattering assumption.

(Figure 4-65) to calculate the broadening of the electron beam b that contains 90% of the scattered electrons:

$$b = 7.21 \times 10^5 \frac{Z}{E_0} \left(\frac{\rho}{A} \right)^{1/2} t^{3/2} \quad (55)$$

where b is the broadening in centimeters, Z is the atomic number, A is the atomic weight, E_0 is the accelerating voltage (electronvolts), ρ is the density (grams/cubic centimeter), and t is the specimen thickness (centimeters). This simple equation is the most common method to estimate the broadening and is most suitable for foils of thickness approximately equal to the elastic mean free path (see Section 3.2), although refinements were proposed to account for multiple scattering in thicker foils (Cliff and Lorimer, 1981; Goldstein et al., 1986).

The spatial resolution can be determined based on the contribution of both the electron beam size (d) and the electron beam broadening (b). Early models proposed the addition of the two terms in quadrature

$$R = (b^2 + d^2)^{1/2} \quad (56)$$

to yield the resolution R , but the currently adopted definition is based on experimental evidence suggesting the use of a less stringent, but still arbitrary, average sum of the two terms leading to a definition of spatial resolution based on the broadening in the middle of the foil thickness as suggested in Figure 4-66:

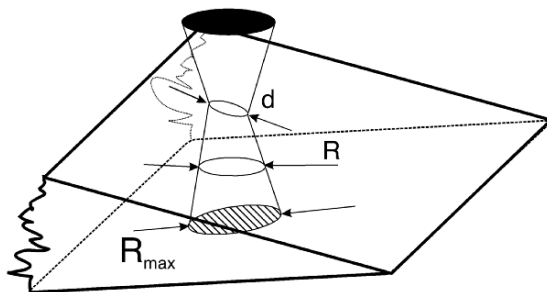


Figure 4-66. Definition of parameters related to the calculation of the resolution with Eqs. (57) and (56) [where $R_{\max}$ is estimated with Eq. (56) as the maximum broadening at the bottom of the foil] and R is the broadening in the middle of the foil. d is the diameter of the incident electron beam.

$$R = \frac{d + R_{\max}}{2} \quad (57)$$

where $R_{\max}$ is determined from Eq. (56) as the maximum broadening at the bottom of the foil. If 90% of the beam current distribution is considered, the resolution becomes (Keast and Williams, 2000)

$$R = \sqrt{d^2 + bd + b^2/3} \quad (58)$$

Recent measurements with sample geometries that probe the electron beam broadening at the bottom of the foil have been published and have led to the conclusion that although the single scattering models predict the right beam broadening magnitude, these quantitatively overestimate the broadening observed experimentally (Nakata et al., 2001). More accurate models (Keast and Williams, 2000; Doigt and Flewitt, 1982) consider a Gaussian probe distribution of standard deviation σ and hence a probe with an FWHM of 4.29σ propagating the foil with a distribution accounting for beam spreading and the incident probe distribution (Keast and Williams, 2000)

$$I(x, y, t) = \frac{i_b}{\pi(2\sigma^2 + \beta t^3)} \exp\left[\frac{-(x^2 + y^2)}{2\sigma^2 + \beta t^3}\right] \quad (59)$$

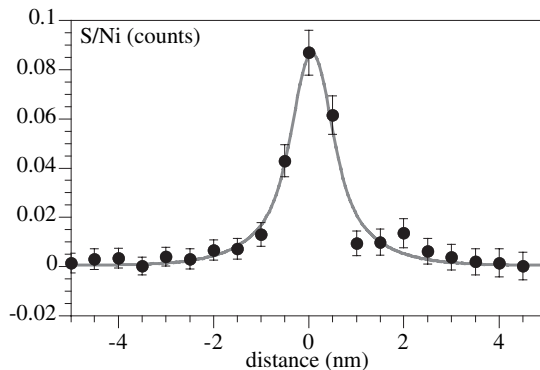
where $\beta = 500(4Z/E_0)^2(\rho/A)$ and i_b is the incident beam current.

Using this more detailed electron distribution in the probe (as well as more simple models) as it travels through the sample, Keast and Williams (1999, 2000) determined the equilibrium segregation profile in grain boundaries based on composition line scans and two-dimensional map measurements (Figure 4-67). Based on the Gaussian intensity distribution of incident electrons the resolution, for a given fraction of electrons Q , as a function of thickness, can be defined as

$$Q = \frac{1}{h} \int_0^h \left\{ 1 - \exp\left[\frac{-R^2}{4(2\sigma^2 + \beta t^3)}\right] \right\} dt \quad (60)$$

where the integration is carried out over the thickness up to $t = h$ where R is the diameter of the cone defining the resolution. The distribution is assumed circularly symmetric (since there is no dependence of the azimuth angle in the equation) and applicable to a two-dimensional case. By defining the resolution criteria based on the fraction of

Figure 4-67. Composition profile of S segregation at an Ni grain boundary. The profile is based on the modeling of the beam profile using a Gaussian probe distribution and Eq. (59) (Keast and Williams, 2000). (Profile courtesy of V. Keast, University of Sydney.)



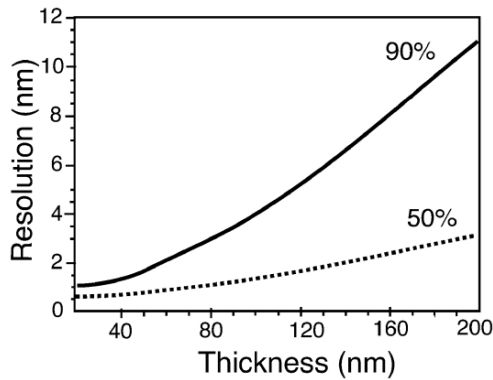


Figure 4-68. Resolution variation as a function of thickness for the 50% and 90% fraction of electrons within the beam. (Adapted from Keast and Williams, 2000.)

electrons within a given radius, it is possible to plot the expected resolution (Figure 4-68) for a diameter containing 50% and 90% of electrons. A comparison of the predictions with experimental profiles obtained on equilibrium segregation in Cu (for FWHM and at 10% maximum) (Keast and Williams, 2000) shows excellent agreement for thicknesses up to the elastic mean free path.

With a dedicated STEM, recent work on grain boundary segregation and multilayer composite materials has shown that it is possible to detect submonolayer segregation at grain boundaries (Keast et al., 1998; Keast and Williams, 1999, 2000) and map the composition of quantum wells in semiconductor materials (Figure 4-69). EDXS analysis with

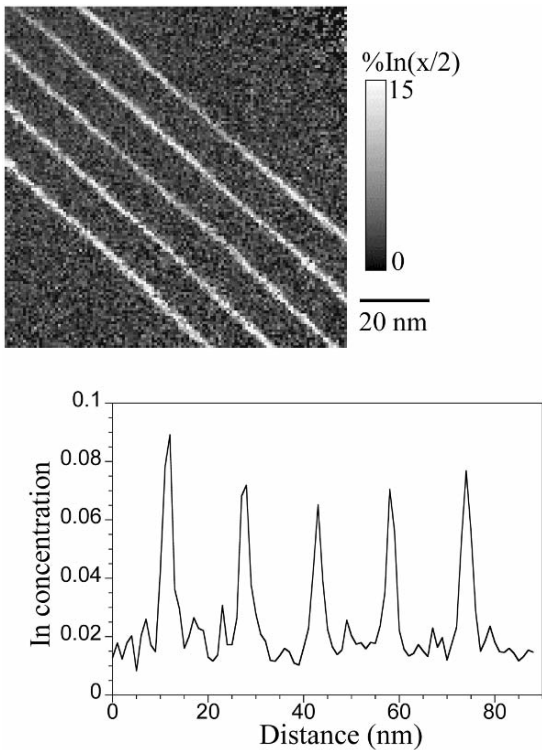


Figure 4-69. Quantitative EDXS elemental maps of In-rich quantum wells in semiconductors and related intensity profiles across the quantum wells. (Courtesy of V. Keast, University of Sydney.)

aberration-corrected instruments has been recently reported by Watanabe et al. (2005) and Watanabe and Williams (2005a).

5.2 Energy-Filtered Microscopy and EELS Microanalysis

The spatial resolution in energy loss spectroscopy measurements is strongly dependent on the operating mode of the microscope during EELS analysis due to the illumination conditions of the area of interest, the effects of lens aberration, and the spectrometer coupling conditions. The contribution from the microscope lens aberrations comes essentially from the objective lens and can be calculated as follows. With respect to an incident beam of primary energy E_0 propagating along the optic axis, the inelastic interactions that induce an energy loss E and scattering at an angle θ will cause a blur of an object point (and the image on the viewing screen of the microscope) due to chromatic aberrations $R = M_0\theta\Delta f$ where M_0 is the magnification of the objective lens and $\Delta f = C_c(E/E_0)$. According to the Rayleigh criterion, two object points with blurred independent distributions can be resolved if they are separated by a distance giving rise to an intensity drop of 20% between the maxima of the summed distributions (Figure 4-70).

The distance between these two points is defined as the resolution and is obtained as (Egerton, 1996)

$$r_i \approx 2\theta_E\Delta f \approx C_c(E/E_0)^2 \quad (61)$$

This resolution is expected for energy-filtered images acquired at an energy loss E when the incident electron energy is kept constant, the image is initially focused for electrons that have lost no energy (which is also known as a zero-loss image, Section 2.4.1), and the scattering angles are limited only by the intrinsic scattering distribution at the energy loss E (i.e., no limiting objective aperture is used). For images of thick samples, it should be pointed out that the inelastic losses can constitute the most important part of the signal resulting in significant blurring of images. These contributions can be removed by energy filtering so that the inelastically scattered electrons are removed with the

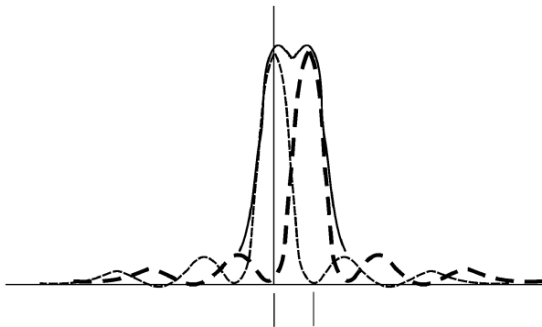


Figure 4-70. Rayleigh definition of resolution based on the overlap of two diffracting-limited functions placed in close proximity to each other. If the sum of the two intensity profiles shows a dip of 20% of the maximum intensity, the peaks are considered to be resolvable.

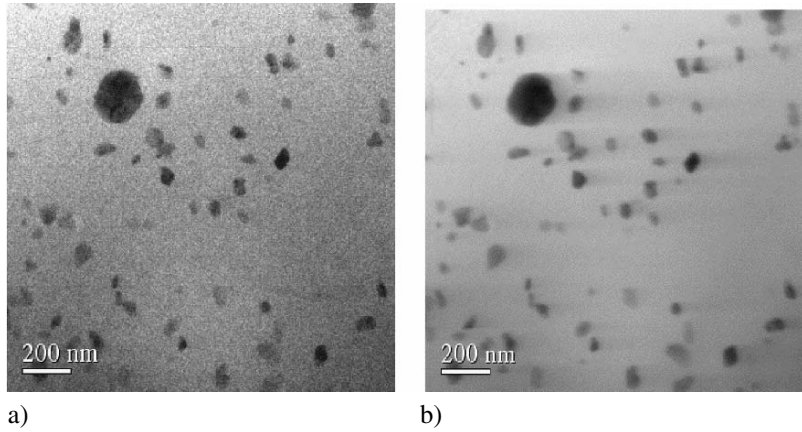


Figure 4-71. (a) Energy-filtered image of an Al alloy with strengthening precipitates. The image has been obtained by selecting only the electrons that have not lost significant amounts of energy (i.e., zero-loss image) and (b) unfiltered image (containing all the inelastically scattered electrons). Although the total intensity is reduced, the precipitates are more clearly defined with sharper edges in the zero-loss image.

energy-selecting slit and the images appear sharp even for thick samples (Figure 4-71). In typical conditions for optimum energy loss imaging at an energy loss E , however, the incident electron energy is readjusted so that the spectrometer is focused at the energy loss E (i.e., by raising the high tension of the microscope so that the primary energy is now $E_0 + E$) and the electrons of the related loss are in focus. If an energy-selecting slit ΔE wide is used, the resolution becomes (Egerton, 1996)

$$r_i \approx 20\theta_E \Delta f \approx (C_C/4)(\Delta E/E_0)^2 \quad (62)$$

When the angular distribution of scattering is limited by an objective aperture β smaller than the angular distribution of scattering, namely θ_E , other contributions become dominant (diffraction limit, spherical aberration, delocalization of scattering) and the spatial resolution in elemental maps is derived from a combination of all factors as described in Section 6.2.3.

For point analyses carried out in TEM image mode, the spectrometer entrance aperture is used to select electrons from the regions of interest so that they enter the spectrometer for analysis. In these conditions, the chromatic aberration will also cause contributions from areas further away from the spectrometer aperture-delimited area due to the significant broadening of the angular distributions and high energy losses. If no angular-selecting aperture (the objective aperture in this case) is used, the total broadening due to angular scattering up to the Bethe-ridge ($\theta_r = \theta_C$, which includes most of the scattering intensity, see Section 3.4.1) is

$$r_c \approx \theta_r \Delta f \quad (63)$$

while the use of an angular-selecting aperture (in the objective back-focal plane when the TEM is operated in image mode) will limit the

broadening to $r_c(\beta) \approx \beta \Delta f$. Once again, two cases can be distinguished: $\Delta f = C_c(E/E_0)$ if the image is focused at the primary energy E_0 and $\Delta f = C_c(\Delta E/4E_0)$ if the electrons are focused at an energy loss of interest and an energy selecting slit ΔE wide is used (Egerton, 1996).

The broadening due to aberrations is a significant contribution that cannot be reduced in current instrumentation, although work is in progress to develop chromatic aberration correctors that would significantly impact the ultimate resolution in energy-filtered imaging. Thus, at the present time, the ultimate resolution is achieved by limiting the illuminated area with the electron beam. Current commercial Schottky-type FEG instruments make it possible to achieve a probe size of about 0.2 nm with typically 10 pA of current (more with a cold field emission source). Spherical aberration correctors can be used to improve the probe-forming capability. These instruments have already been developed and implemented on different platforms by major manufacturers on both dedicated scanning transmission microscopes and STEM-TEM instruments yielding sub-Ångstrom probes with 10–20 pA current or about a few hundred picoamperes for 0.2-nm-diameter probes [Dellby et al., 2001; Krivanek et al., 2003; Batson et al., 2002; Krivanek et al., 2003; and Chapter 2 (this volume)]. With such small probes, however, it becomes increasingly important to be aware of the detailed electron propagation within the sample as discussed in Section 3.2.

Since the electron propagation process in the sample is the same irrespective of the microanalysis technique probing various signals, elastic scattering does affect the broadening of the electron beam not only in EDXS but also in EELS measurements. The impact of this broadening on the degradation of the spatial resolution, however, can be somewhat controlled with the use of a collection aperture that limits the scattering angles entering the spectrometer. As shown in Figure 4–72a, electrons scattered at high angles (thus away from the forward direction and the incident probe distribution) can be eliminated with the use of an angle-limiting collection aperture (either the objective aperture if the spectra are acquired in image mode or the spectrometer aperture if the spectra are acquired in diffraction and STEM mode). Assuming that the elastic scattering distribution is large compared to the collection angle and that there is no strong Bragg scattering (essentially an amorphous sample) the fraction of electrons contained within a radius r for a given sample thickness t and collection aperture β can be estimated geometrically (Figure 4–72b and c). Based on Figure 4–72b we can estimate this geometric broadening contribution for a parallel incident electron beam. For a sample thickness of 100 nm and a collection angle of 10 mrad, the fraction of electrons contained within 0.5 nm would be 85% and nearly 100% for 1 nm. These contributions can be small compared to the intrinsic broadening due to the convergence of the electron beam required for STEM imaging and optimal probe size (in the range from a few millirad to a few tens of millirad in the case of aberration-corrected STEM instruments (see Section 2.1). The latter geometric broadening contribution for a 100-nm-thick sample assuming optimal probe size with 30 mrad convergence in an aberration corrected STEM would be 3 nm! Although these numbers repre-

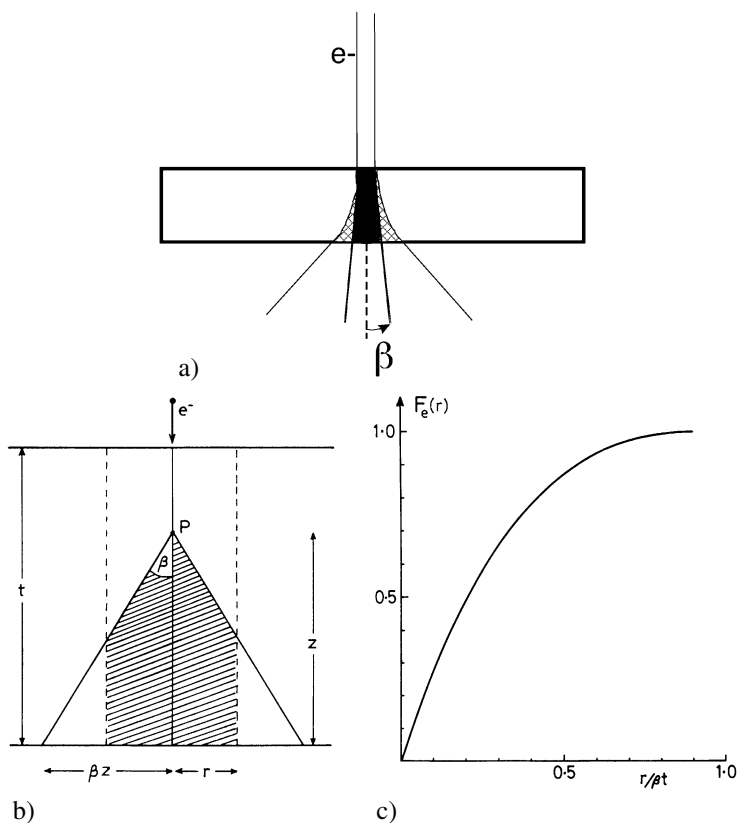


Figure 4-72. Geometric contribution to the broadening of the electron beam. (a) The effective broadening can be reduced by using an angle-limiting aperture even if the real broadening still occurs in the sample as larger scattering angles (causing the broadening) are cut off by the aperture. (b) Diagram defining the broadening terms necessary to calculate the fraction of electrons scattered outside a radius r for a given combination of thickness and collection aperture. (c) Geometric evaluation of the fraction of electrons contained within a radius r . (From Egerton©1996, with permission from Springer Science+Business Media.)

sent an extreme effect due to the sample thickness, typical samples used to demonstrate the ultimate resolution tend to be in the order of 5–10 nm thick as required to limit the propagation of the electron beam within a single atomic column (Section 3.2). In these cases, the geometric broadening would be of the order of 0.1–0.3 nm.

When the electron beam size is in the order of a fraction of a nanometer, a significant contribution affecting the spatial resolution is the delocalization of inelastic scattering, i.e., the excitation and energy loss can occur even for electrons traveling at a finite distance b from the target atoms and not only when the incident electron is “directly on the atom” in the classical particle point of view. This factor ultimately limits the spatial resolution in EELS analysis with aberration-corrected electron microscopes capable of achieving sub-Ångstrom beam sizes.

As expected from the wave-mechanical perspective, the origins of the delocalization effect are quantum mechanical in nature and relate

to the uncertainty relation, but a simple classical treatment makes it possible to identify the significant elements related to spatial resolution. Based on classical scattering, the impact parameter b is related to the scattering angle of electrons (Figure 4–40): large scattering angles are indicative of a smaller impact parameter and thus interactions closer to the atom. Hence, measurement of energy losses carried out on electrons scattered at a large scattering angle would imply higher apparent resolution. The simplest treatment of the spatial resolution involving the uncertainty principle (e.g., Brown, 1999) invokes the time of interaction ($\Delta\tau$) of an electron traveling at a speed v and at a distance b from the atom. The interaction time is $\Delta\tau = b/v$. Applying the uncertainty relation $\Delta E \Delta\tau \leq h$ yields $b_{\max} = hv/\Delta E$ with b considered the ultimate spatial resolution and ΔE the energy loss. For energy losses of about 20 eV (at 100 keV) the delocalization is of the order of 2 nm while for energy losses of 200 eV it would be 0.2 nm. A more detailed analysis (Pennycook, 1982; Pennycook et al., 1996) based on the same principles yields a delocalization with the form

$$b = \frac{hv\beta}{\Delta E} \left[(\beta^2 + \theta_E^2) \ln \left(1 + \frac{\beta^2}{\theta_E^2} \right) \right]^{-1/2} \quad (64)$$

where β is the limiting collection angle (the maximum scattering angle entering the spectrometer) and θ_E is the characteristic scattering angle at the energy loss ΔE ($\theta_E = \Delta E/2E_0$).

The detailed quantum mechanical treatments by Muller and Silcox (1995) and Kohl and Rose (1985) as well as the work of Pennycook (1982) led to results (in terms of trends and order of magnitude) similar to the simpler analysis presented by Egerton (1996, 1999), who considered a spatial resolution related to the diffraction limit imposed by the width of the energy-dependent scattering distribution and the collection aperture. This approach can be deduced as follows. For high energy losses, the angular distribution can be broader than the collection aperture and thus the diffraction limit imposes a resolution determined by the collection aperture β similar to the Rayleigh criterion, i.e., $0.6\lambda/\beta$.

When the angular scattering distribution is not limited by the collection aperture (typically for low energy losses), the scattering distribution is limited by a cut-off angle $\theta_c = (2\theta_E)^{1/2}$ related to the Bethe-ridge maximum scattering angle (Sections 3.3 and 3.4.1). In this case, the median scattering $\tilde{\theta}$ angle containing 50% of the electrons is $\tilde{\theta} \approx (\theta_E \theta_c)^{1/2}$ and thus $\tilde{\theta} \approx 1.2(\theta_E)^{3/4}$. By considering that the limiting aperture containing 50% of the electrons is effectively given by $\tilde{\theta}$, this diffraction limit contribution will be $0.6\lambda/\tilde{\theta}$, i.e., $0.5\lambda/(\theta_E)^{3/4}$. Combining the two limiting terms in quadrature, the delocalization contribution to the resolution, related to inelastic scattering of 50% of the intensity, is

$$d_{50} \approx \sqrt{\left[0.5\lambda/(\theta_E)^{3/4} \right]^2 + (0.6\lambda/\beta)^2} \quad (65)$$

Experimental measurements roughly agree with these estimates and more detailed quantum mechanical treatments (Kohl and Rose, 1985; Muller and Silcox, 1995) (Figure 4–73). More recent work deals with the impact of aberration correction and ultimate spatial resolution by combining the effects of inelastic scattering and sub-Ångstrom beam prop-

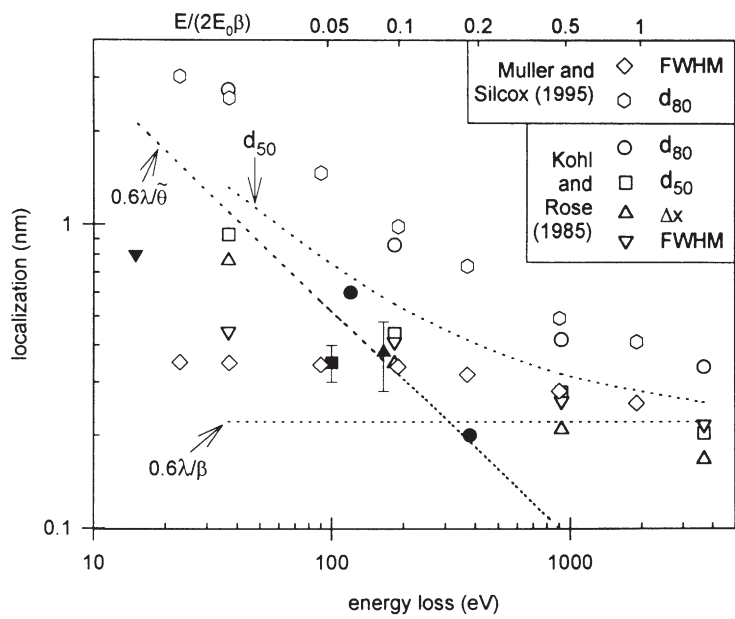


Figure 4-73. Localization of inelastic scattering as a function of energy loss for a collection angle β of 10 mrad. Curves are based on a Rayleigh approach to the evaluation of the localization. Various experimental measurements based on the literature are presented (filled symbols). (From Egerton, 1996.)

agation in the sample (Oxley et al., 2005; Cosgriff et al., 2005; Dwyer, 2005). This delocalization contribution enters the definition of resolution for energy-filtered elemental mapping (Section 6.2.3). It is important to note that elastic scattering can affect the apparent resolution shown in inelastic images due to the modulation of the signal entering the spectrometer. If strong elastic contrast is present (for example, lattice fringes, larger atomic number elements, or diffraction contrast), inelastic images might show features with apparent resolution that is not simply related to the inelastic scattering distribution but rather to the much more localized elastic scattering (Spence and Lynch, 1982). Lattice images obtained with inelastically scattered electrons can be obtained, therefore, at energy losses where the localization contribution is, in principle, larger than the interatomic spacing due to these effects. As discussed in Section 6.2, approaches have been devised to circumvent these effects by selecting the collection aperture so that the phase contrast, giving rise to lattice images, disappears and by normalizing the inelastic images with the background prior to the edge threshold (Freitag and Mader, 1999; Hofer et al., 2000).

6 Elemental Mapping

6.1 Elemental Mapping EDXS

With the control of scanning coils of the electron microscope (Section 2.2) and by recording X-ray counts within a particular energy window

of interest under a characteristic peak or even the full spectrum at each pixel of a rastered area, it is possible to combine the intensity recorded for each element into a map to display the distribution of elements in the sample with a resolution appropriate to the experimental conditions and sample thickness. Due to the long recording time at each pixel necessary to obtain a statistically significant number of counts at each pixel and still maintain a small probe size compatible with the spatial resolution of the technique, the acquisition of elemental maps with EDXS can take several minutes to a few hours. Examples of such maps in the case of semiconductor materials are shown in Figures 4-69 and 4-74. Although simple unprocessed elemental maps provide a wealth of information on the distribution of elements, it is also possible to extract quantitative information on the concentration of elements based on the use of ratios of images and the use of k -factor analysis. Relative concentration maps for elements A and B can thus be obtained by processing the entire image rather than just integrated intensities under an energy window. The quantitative concentration map can be obtained as $I_{C_A/C_B}(x,y) = k_{AB}[I_A(x,y)/I_B(x,y)]$ where k_{AB} is the Cliff-Lorimer factor discussed in Section 4.1.1 and $I_{A,B}(x,y)$ are the elemental maps for elements A and B, respectively. Each elemental map has a background image (obtained by recording the intensity within a window where no characteristic peak is visible) subtracted prior to the quantification. The division is carried out pixel by pixel with the constant k_{AB} factor.

If absorption is not significant, relative concentration maps are insensitive to changes in thickness. Therefore, although raw elemental maps show changes in intensity related to variations in the projected density

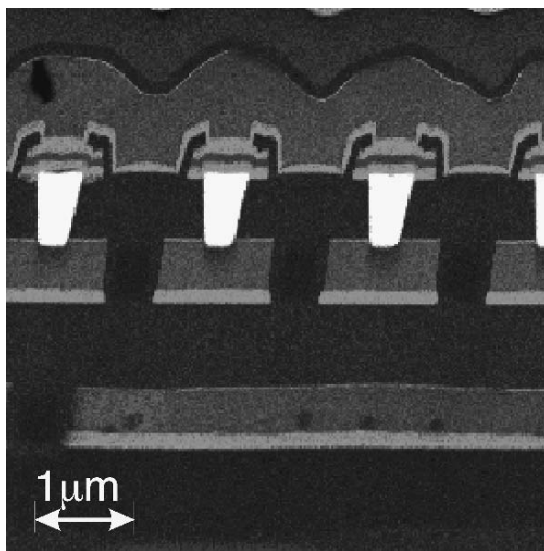


Figure 4-74. Color-coded elemental map of a device showing the distribution of elements in the rastered area: red, Al-rich area; blue, Si-rich area; green, Ti-rich area. White, W. Interdiffusion of Si into the Al is noted through the bottom barrier layer containing Ti. (See color plate.)

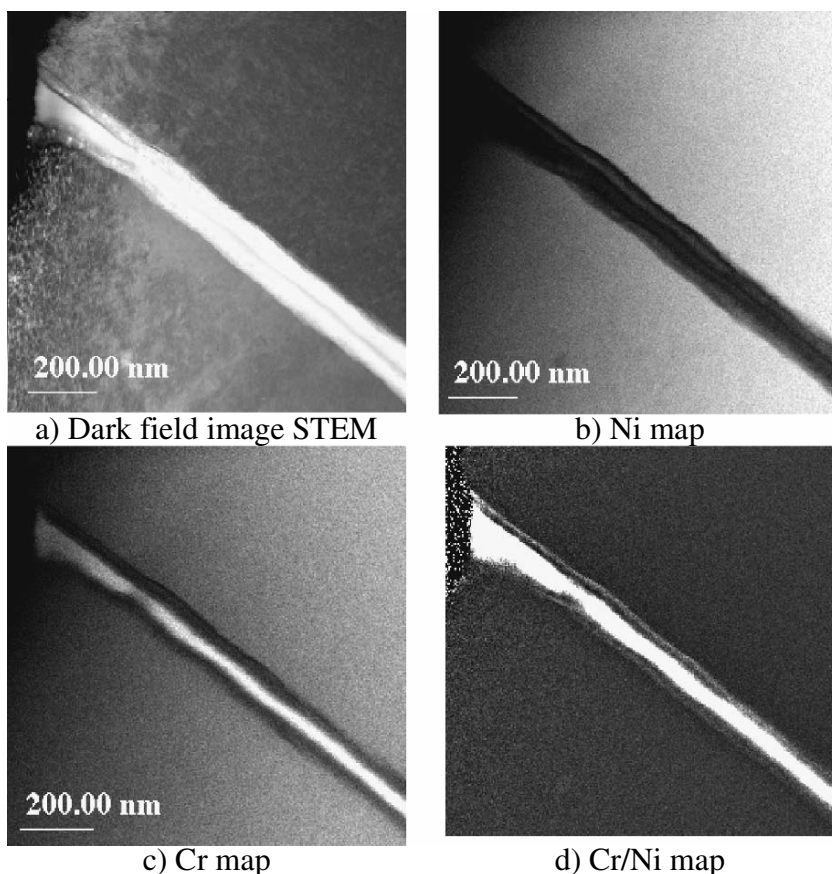


Figure 4-75. Elemental maps of Ni-based alloy with intergranular product. (a) STEM image; (b) Ni map showing the variation of intensity due to thickness variations (from the top left corner to the bottom right corner); (c) Cr elemental map with similar variations in intensity due to thickness; (d) elemental ratio map for Ni/Cr demonstrating the correction of the thickness variations in the maps. The intergranular product contains a Cr-rich carbide phase surrounded by oxidation product. (See color plate.)

of atoms with changes in thickness, the ratio map does not (Figure 4-75). When absorption is significant, this simple approach fails and absorption correction is required. For this correction, the sample thickness must be determined. To deal with strong absorption cases without calculation of the absorption correction factor (Section 4.1.2), quantification of maps through the ζ -factor approach have been proposed (Williams et al., 2003; Watanabe et al., 1996; Watanabe and Williams, 1999).

The technique makes it possible to circumvent the problem of having to determine the sample thickness independently (with intrinsic large errors) by calculating the absorption correction composition and the sample thickness simultaneously. The ζ -factor relates the intensity and composition of a standard sample to the mass thickness for both elements A and B as

$$\rho t = \zeta_A \frac{I_A}{C_A} \quad \text{and} \quad \rho t = \zeta_B \frac{I_B}{C_B} \quad (66)$$

where $I_{A,B}$ are the intensities recorded for elements A and B and $C_{A,B}$ are the concentrations of the same elements. If the ζ -factors are determined through measurements of standards of known composition and thickness via Eq. (66) the concentrations are deduced as (Williams et al., 2003)

$$C_A = \frac{I_A \zeta_A}{I_A \zeta_A + I_B \zeta_B}, \quad C_B = \frac{I_B \zeta_B}{I_A \zeta_A + I_B \zeta_B}, \quad \rho t = I_A \zeta_A + I_B \zeta_B \quad (67)$$

Quantitative concentration and thickness maps accounting for this effect have been demonstrated (Figure 4–76).

Advanced analysis techniques based on the full processing of the spectrum and multivariate analysis make it possible to extract the occurrence of the various phases without prior knowledge of the individual component phases within the samples. Examples of multivariate analysis at high spatial resolution have been demonstrated in AEM (Kotula et al., 2003) following initial developments in SEM (Kotula et al., 2001) and earlier work of line profiles by Titchmarsh and Dumbill (1996) and Chevalier and Botton (1999). These techniques are particu-

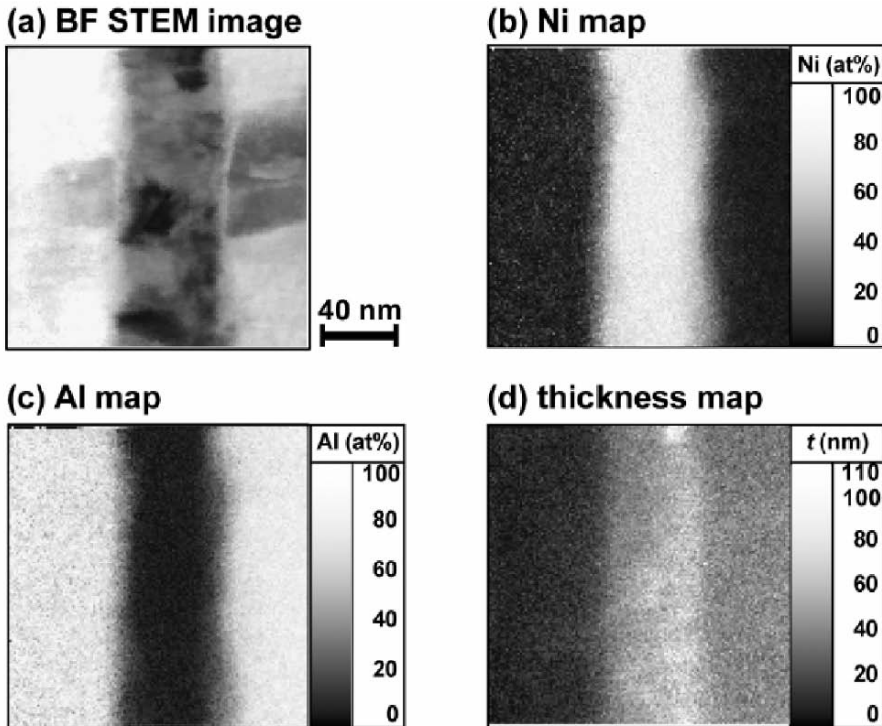


Figure 4–76. Quantitative elemental maps of an NiAl multilayer obtained with the ζ -factor approach. The gray scale represents the quantitative information on sample composition including correction for absorption. (Reprinted from Williams et al. © 2003, with permission from Elsevier.)

larly useful for the analysis of segregation at grain boundaries and interfacial phases.

6.2 EELS Mapping

6.2.1 Energy-Filtered TEM Mapping

With both postcolumn and in-column energy filters described in Section 2.4.1, it is possible to retrieve images at specific energy losses by taking advantage of an energy selected slit. These images can show the distribution of elements following the extraction of the background from the images. Recent reviews of the technique and the applications have been given in Hofer and Warbichler (2004), Verbeeck et al. (2004), and Hofer et al. (2000), and we will summarize the approaches here. In the basic approach of energy-filtered elemental mapping, this extraction of element-specific images can be achieved through two different methods: the three-window technique and the jump-ratio technique. In both cases, energy-selected images at selected energy losses must be taken before the ionization edge and at the edge of interest (Figure 4-77).

In the three-window method, two energy-filtered images (called preedge images) acquired prior to the edge onset are used to extrapolate the noncharacteristic background under the edge where the third energy-filtered image (called postedge image) has been acquired. The same power-law extrapolation model used for quantitative analysis is used [namely, $I_b(E) = AE^{-r}$] using the two preedge windows. The elemental map is then obtained by subtraction of the background image from the postedge image. Because there are only two images for this extrapolation, the resulting error in the intensity of the postedge background image, and thus the subtraction, can be large. If the intensity in the recorded images is low and the background varies steeply at the energy loss of interest, spurious images with negative extracted intensity can often be obtained because the variance in the extrapolation can be larger than the intensity of the ionization edge. Since the signal in the ionization edge is related to the ionization cross section and the number of atoms (Section 4.2.1) images represent the quantitative distribution of elements within the field of view.

The second approach to extract the distribution of elements is through jump ratio images. In this case, the postedge image is divided by one of the preedge images (generally obtained just before the edge threshold). Although these images are not directly quantifiable, they have the advantage that the noise is lower (no extrapolation is involved), the elastic contrast (due to diffraction, high atomic number, and phase contrast) is canceled as it affects both the preedge and postedge images in a similar way, and finally there is less drift involved as only two images are used (and thus a shorter acquisition time) (Verbeeck et al., 2004). This latter point significantly affects the spatial resolution of images as discussed below. Both methods can be applied with energy filters (irrespective of whether they are in-column or postcolumn) and also in STEM instruments equipped with serial spectrometers allowing the acquisition of energy-filtered images through an energy-selecting slit.

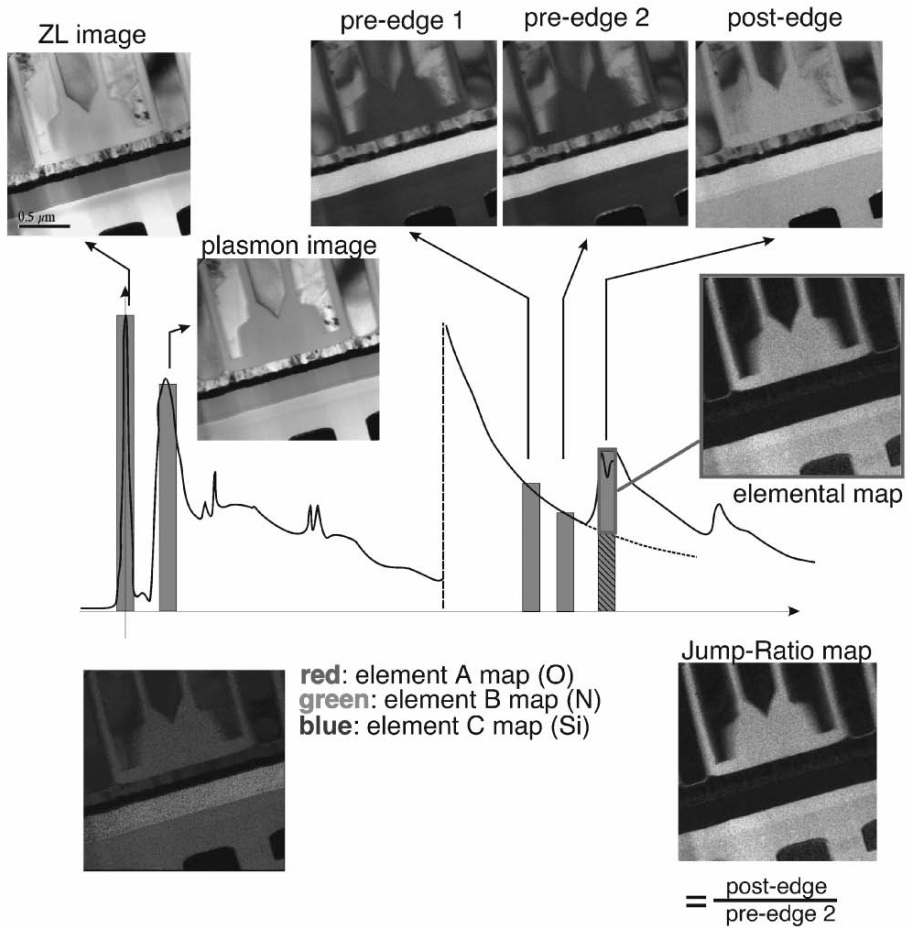


Figure 4-77. Various approaches to EFTEM imaging. Zero-loss (ZL) filtered imaging (selecting only electrons that have lost no significant amounts of energy), plasmon imaging (selecting only electrons that have lost energy in the 10–30 eV range), and core loss imaging with the three-windows technique (extrapolation of the background under the edge) and the jump-ratio technique. (See color plate.)

Early serial spectrometers on dedicated VG-STEM and the first serial spectrometers from Gatan both equipped with scintillators and photomultipliers served this purpose. Parallel spectrometers and imaging filters equipped with single-channel photomultiplier detectors (Krivanek et al., 1994) or a fast array of photomultipliers (developed by the Ottensmeyer group in Toronto) (Ottensmeyer, 2004) were also developed in prototype systems but were superseded by the development of much faster detectors allowing the acquisition of full spectra at each pixel of the rastered area as discussed in Section 6.2.2.

In addition to these basic energy-filtered TEM (EFTEM) acquisition techniques, there have been variants of the three-window and jump-ratio methods for EELS imaging based on the development of more advanced acquisition software controlling the spectrometer, the microscope, and increased storage capabilities in desktop computers. One variant is the EFTEM-spectrum imaging method initially proposed by

Lavergne et al. (1992) and fully developed as very powerful tools for mapping and quantitative analysis by Mayer et al. (1997) and Thomas and Midgley (2001a,b). This technique is now implemented in commercial packages from energy filter vendors. The method is based on the acquisition of an information data cube consisting of a large number of energy-filtered images with narrow energy windows (from 1–2 to 5–10 eV) and small step intervals making it possible to cover large energy ranges around the edges [in some cases even from 0 eV to the characteristic core losses (Thomas and Midgley, 2001b)] so that the full information from the spatial distribution and energy loss can be retrieved with high spatial sampling (Figure 4–78). If the spatial drift of the sample is well corrected in the image stack using automated

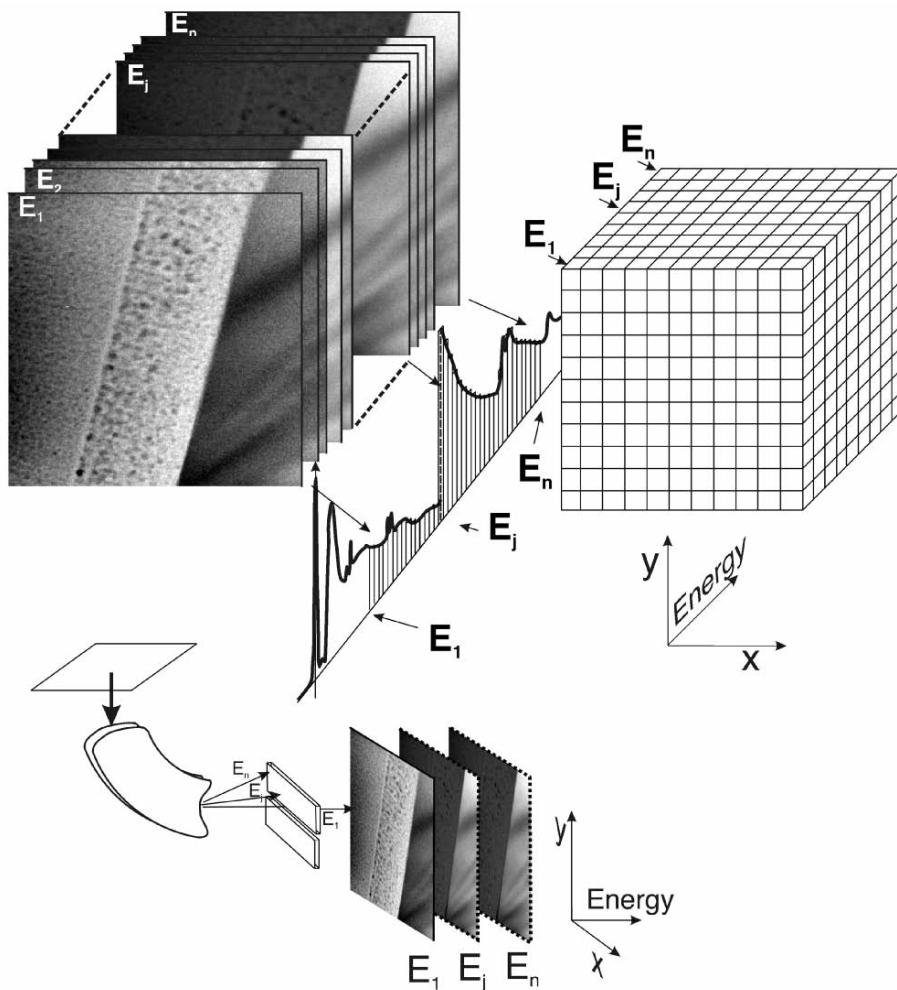


Figure 4–78. Schematic description the energy-filtered TEM spectrum imaging technique (EFTEM-SI). Each image obtained at a given energy loss is part of a three-dimensional data cube containing information on the distribution of elements. The technique provides detailed sampling of the spatial information with little sampling of the energy loss distribution (energy window widths can vary from less than 1 to 10 eV).

procedures (Schaffer et al., 2004), these datasets have the advantage that the background extrapolation can be carried out with much higher precision than in the three-window technique. Energy windows as small as 0.1 eV have been used on instruments equipped with monochromators and high-resolution filters corrected up to third-order aberrations (Hofer et al., 2005). Quantitative analysis to extract the absolute atomic density and deconvolution of the effects of multiple scattering are also possible since the full spectrum is retrievable at each pixel (Thomas and Midgley, 2001b). In fact, for each pixel of the image sequence the intensity can be measured and a spectrum of energy resolution equivalent to the width of the energy-selecting slit is deduced. Based on limited sequences of images with narrow energy windows (1–2 eV) and selection of spectra at the edge threshold, this approach has demonstrated that changes in the near-edge structure due to variations in the chemical bonding state of elements are visible (Muller et al., 1993; Botton and Phaneuf, 1999; Bayle-Guillemaud et al., 2003; Hofer et al., 2000). This capability suggests that similar to X-ray absorption scanning transmission microspectroscopy (e.g., Hitchcock et al., 2005), bonding changes can be visualized with EELS mapping, albeit with a spatial resolution typical of EFTEM images (see Section 6.2.3).

6.2.2 STEM-EELS Mapping

The other variant of the more advanced EELS imaging technique makes use of developments in fast detectors, large storage capacity, and fast computers. The approach is based on the use of STEM instruments equipped with parallel or dedicated 2D fast detectors. The original idea of this technique was proposed by the Orsay group (Jeanguillaume and Colliex, 1989) and implemented in subsequent years by various groups (Hunt and Williams, 1991; Botton and L'Espérance, 1994; Colliex et al., 1994). As for EFTEM spectrum imaging, the STEM-EELS imaging is also commercially available from the spectrometer manufacturer Gatan. The technique involves the sequential acquisition of an energy loss spectrum (acquired with the photodiode array or a 2D detector) at each pixel of a rastered area. The filling of the data cube (Figure 4–79) is thus achieved by scanning the beam over each pixel of the area of interest (pixel by pixel in two dimensions or over a line across interfaces) with the third (and fourth) dimension in the data cube being the energy loss spectrum (energy and intensity).

The advantage of this approach is the availability of the full spectrum at each pixel making it easy to implement various data-mining approaches. Signals can be extracted with advanced methods, including multiple least-squares techniques, and the detailed shapes of the near-edge structures can be fitted with reference standards from different phases. This technique can therefore be used to map changes in the chemical bonding environment of atoms in nanoscale structures as in EFTEM imaging but with much higher spectral sampling. Through the analysis of a single edge and with reference standards, it is thus possible to extract the distribution of the individual phases rather than just the chemical composition. This powerful technique can be imple-

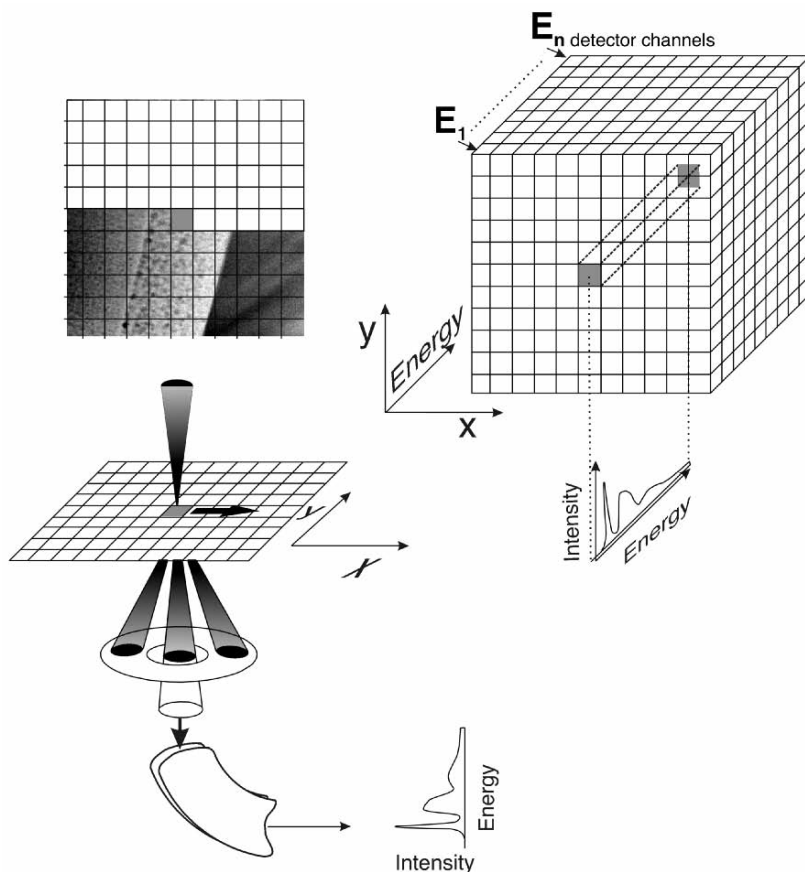


Figure 4-79. Schematic description of the STEM spectrum imaging technique (STEM-SI). For each pixel of the rastered area an energy loss spectrum is acquired. Although the spatial sampling is typically lower than in the EFTEM-SI technique, the spectral sampling is higher with easy recording of the near-edge structure features at each pixel.

mented in one or two dimensions (Figure 4-80) for core loss near-edge structure features and also for low-loss features. Using the differences in low-loss spectra of various biological structures and ice, Sun et al. (1995) were able to map the distribution of the various functional components in cells (Figure 4-81). Similarly, for the analysis of polymer-based materials, the use of low-loss features related to the presence of π and $\pi + \sigma$ plasmons has been used to distinguish polystyrene and polyethylene in composite blends (Figure 4-82) (Oikawa, 2006; Hunt et al., 1995).

6.2.3 Quantitative EELS Imaging

If core-loss spectra are combined with low-loss spectra acquired from the same pixel, it is possible to retrieve more quantitative information on the sample. Hence, deconvolution techniques (Section 8.2) can be applied to retrieve the single scattering distribution for accurate quantification (Section 4.2), dielectric function measurements (Section 8.2),

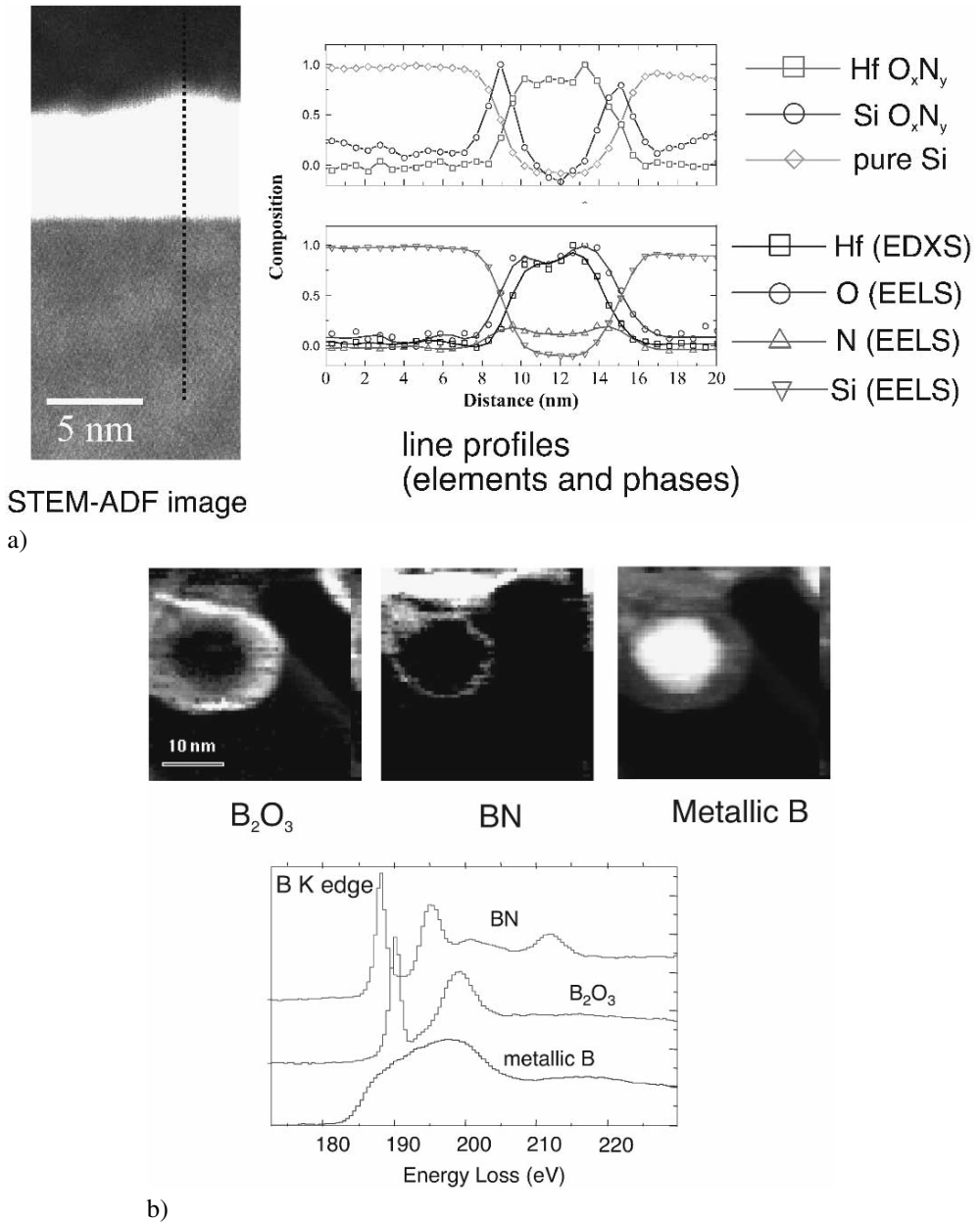


Figure 4-80. (a) Application of the STEM-SI technique to determine the distribution of elements across interfaces between a high dielectric constant material (Hf-O-N) and Si (STEM annular dark-field image, left). EDXS spectra and EELS spectra were recorded simultaneously to extract the elemental composition (bottom right profile). By analysis of the near-edge structure shape of the Si L_{23} edge and O K edge, it is possible to distinguish and map the contribution of pure Si, Si-O-N, and Hf-O-N. (Courtesy of M. Couillard, McMaster University.) (b) Two-dimensional phase maps of B nanostructure with BN, B_2O_3 , and metallic B separated according to the shape of the near-edge structure spectra of the B K edges using the STEM-Spectrum imaging technique and least-squares fitting of spectra. The structure consists of a metallic B core with a thin BN shell and outer thick shell of B_2O_3 . (Courtesy of O. Stephan and C. Colliex, University Paris-Sud.)

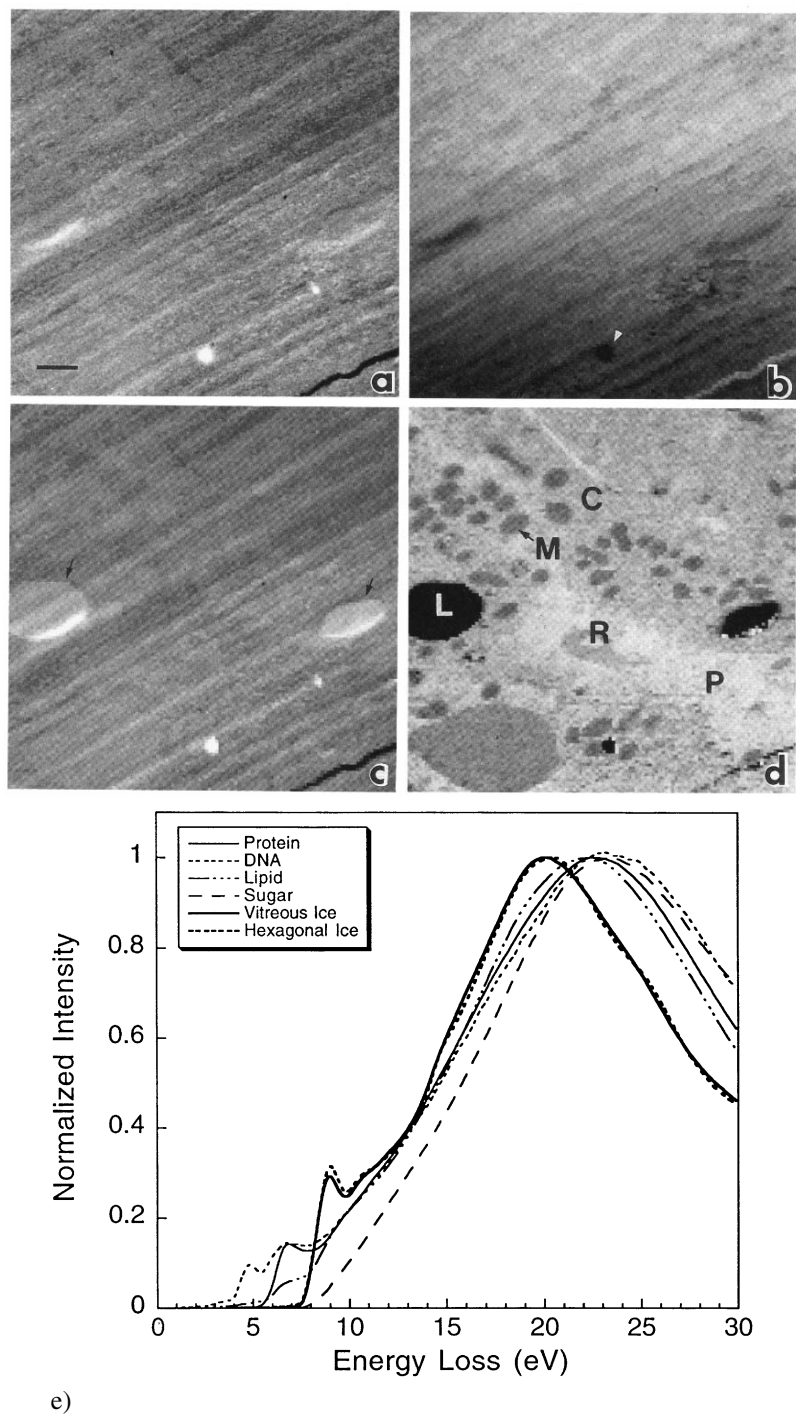


Figure 4-81. STEM images and phase component maps in frozen hydrated liver tissue. (a) Low-dose dark-field STEM image sectioned sample showing no contrast; (b) bright-field image of the same region based on integration of all signals on the spectrum at each pixel of the STEM-SI (the intensity drop is due to the variation of beam current during the acquisition); (c) relative thickness map (t/λ), contrast is visible on the lipid droplets; (d) water maps and identification of the various biological components based on the multiple least-squares fit of the low-loss spectra with water and protein reference compounds and the amount of water within the structure: L, lipid droplet (zero water content); P, plasma (91% water); R, erythrocyte (65% water); M, mitochondria (57% water); (e) low-energy loss spectra of the various components differentiating the phases. Bar marker is 1 μm . (From Sun et al. © 1995, reprinted with permission from Blackwell Publishing.)

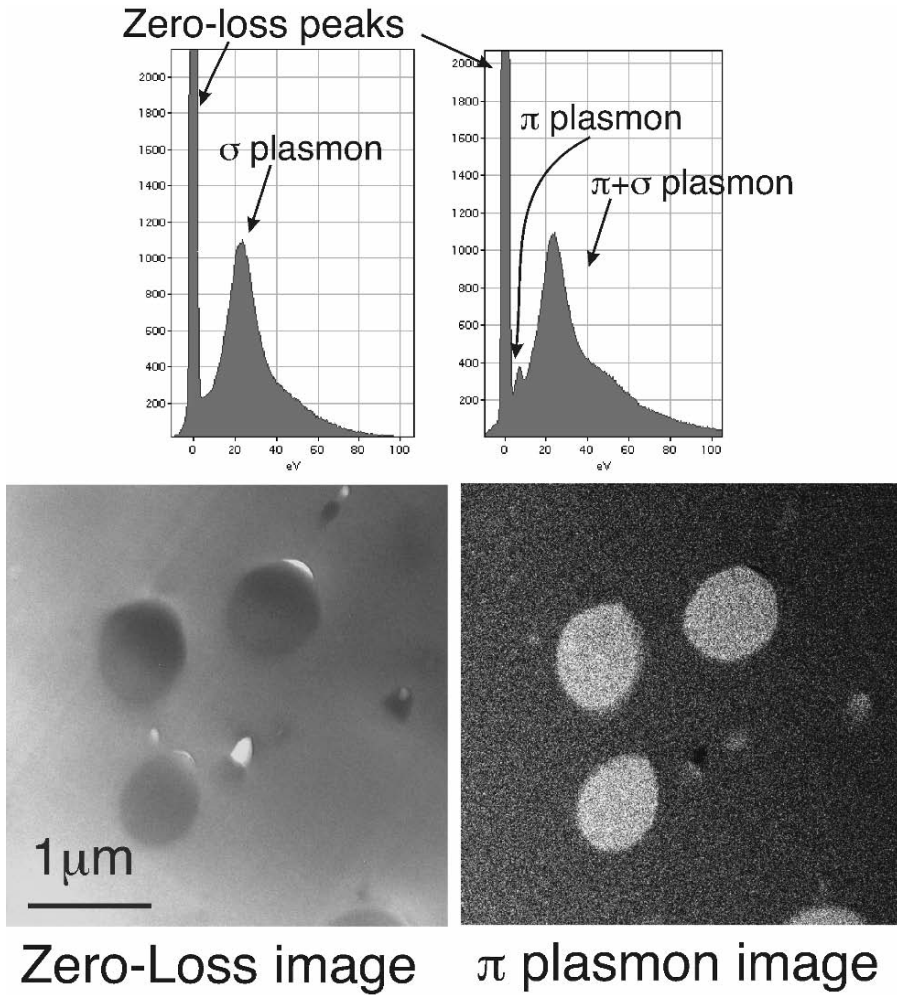


Figure 4-82. Section of polyfin-polycarbonate polymer composite. The low-loss spectrum of the polyfin (top left) does not show the π plasmon resonance visible on the spectrum of the polycarbonate material (top right). The zero-loss image (bottom left) does not show strong contrast between the two phases while the π plasmon energy-filtered image (bottom right) clearly makes it possible to distinguish the regions where the polycarbonate is present. (Courtesy of Dr. T. Oikawa, JEOL.)

and the thickness relative to the inelastic mean free path of the sample (Section 8.2), and also to implement quantitative statistical analysis (Bonnet et al., 1999) to retrieve significant spectral components and analyze changes in bonding at interfaces in materials. Thickness maps, relative to the mean free path (Section 8.2), can be obtained with the EFTEM and STEM imaging approaches by acquiring a zero-loss image $I_0(x,y)$ and an unfiltered image $I_t(x,y)$. Following the approach discussed in Section 8.2 for the analysis of individual spectra to obtain relative thickness (t/λ) values, the ratio of the two images can be combined to give the relative thickness map $I_{t/\lambda}(x,y) = \ln[I_t(x,y)/I_0(x,y)]$. The variant of this EFTEM method is to acquire the low-loss spectrum at each pixel and process the individual spectra to deduce t/λ at each

point of an image (see Section 8.2). This information on the sample thickness is valuable for the determination of the volume of the sample under analysis (hence the volume fraction of particular phases or defects), or to determine whether the changes in thickness affect the apparent intensity of elemental maps. The thickness information is also useful to verify whether the thickness of the sample is beyond the critical thickness at which accurate extraction and quantification can be carried out (Section 4.2.4). Finally, the thickness information, if combined with EDXS maps, can lead to fully quantitative X-ray maps accounting for X-ray and absorption corrections.

As in the case of EDXS imaging, elemental maps can also be combined to retrieve fully quantitative concentration maps and deduce phase analysis histograms using experimental k -factors or cross sections (Hofer et al., 1997; Kothleitner and Hofer, 2003). The advantage of concentration maps is, as in the case of EDXS mapping, the fact that within a range of relative thickness $t/\lambda < 0.5$, images are independent of thickness as discussed in Section 4.2.4. Diffraction effects due to elastic scattering of electrons outside the objective aperture can also lead to apparent variations in the intensity of elemental maps and can be canceled out using the jump-ratio imaging technique. Concentration maps based on the single scattering distribution of energy losses obtained after deconvolution of the full spectrum at each pixel show that reliable quantitative images can be obtained for thickness up to $t/\lambda \cong 2$ (Thomas and Midgley, 2001a).

Another useful technique demonstrating the removal of diffraction effects is the use of the rocking beam method during the acquisition of energy-filtered images. In this approach, the incident electron beam is tilted over a cone of angles (of the order of the Bragg angle) so as to average out the local diffraction effects including deviations of the scattering due to dislocations. Energy-filtered images with little diffraction contrast can thus be obtained even in bent and highly deformed samples (Hofer and Warbichler, 1996; Hofer et al., 2000) (Figure 4–83). Removal of diffraction contrast for qualitative imaging and visualization of precipitates in highly deformed samples has also been demonstrated using ratios of plasmon images obtained at different energies (Carpenter, 2004).

Various aspects of optimization of signals for EFTEM and STEM-EELS maps, including the position of energy windows, automatic detection of edges, illumination conditions, and magnification, are discussed in the work of Kothleitner and Hofer (1998, 2003), Grogger et al. (2003), Berger and Kohl (1992, 1993), and Berger et al. (1994). Through image analysis of the quantitative maps, it is also possible to segment images based on the composition and relative fraction of elements (Hofer et al., 1997, 2000). Algorithms to allow automatic detection of edges, for quantitative analysis of phase distributions, for the determination of thresholds for phase detection and problematic zones in the samples have also been developed with the use of full spectra recorded in STEM mode (Kothleitner and Hofer, 2003). Corrections of drifts in EFTEM images for quantitative analysis have been discussed in detail in Schaffer et al. (2004).

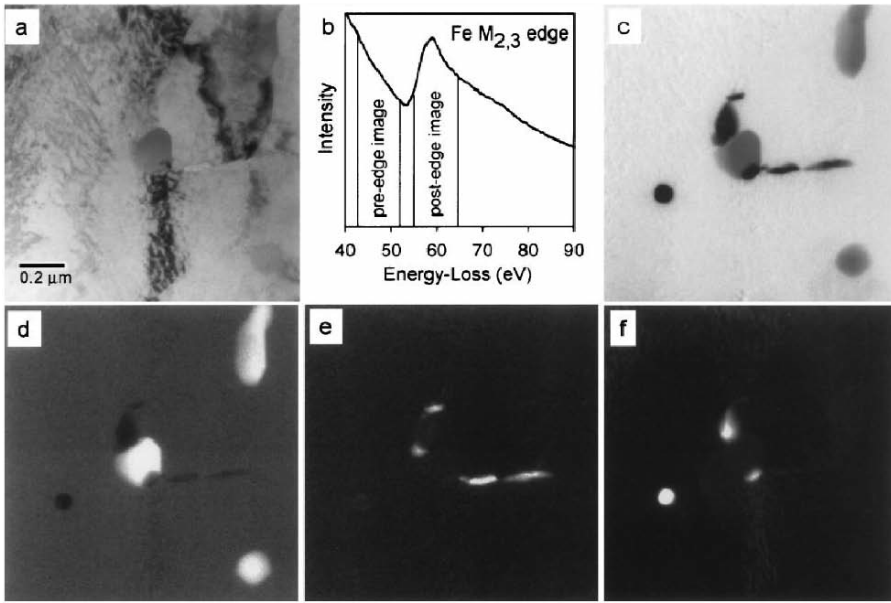


Figure 4–83. Imaging of precipitates in a steel sample. (a) TEM bright-field image of the 10% Cr steel with Cr_{23}C_6 -VN and $\text{Nb}(\text{C},\text{N})$ precipitates. (b) EELS spectrum with the Fe- $\text{M}_{2,3}$ edge. (c) Fe- $\text{M}_{2,3}$ jump-ratio image recorded with the rocking beam illumination. (d) Cr- $\text{L}_{2,3}$ jump-ratio image. (e) V- $\text{L}_{2,3}$ jump-ratio image. (f) Nb- $\text{M}_{4,5}$ jump-ratio image. (From Hofer et al. © 2000, with permission from Springer Science+Business Media.)

6.2.4 Spatial Resolution in EFTEM Elemental Mapping

The resolution in EFTEM elemental mapping depends on several factors related to the operation parameters of the microscope and the energy loss of interest so that further discussion of this topic is required. When no angular limiting aperture is present (in imaging the objective aperture would be limiting the scattering angles) the dominant factor is related to the chromatic aberration term discussed in Section 5.2. For general conditions, however, we must summarize the contributions that need to be added in quadrature to retrieve the total broadening of an object point (Krivanek et al., 1995b).

1. Following the discussion in Section 5.2, the chromatic aberration broadening term when a limiting aperture is used can be described as

$$d_c = C_c \frac{\Delta E}{E_0} \beta \quad (68)$$

where β is the collection angle (limited by the objective aperture in imaging mode), ΔE is the width of the energy window used to acquire the image, and E_0 is the incident energy. This expression assumes that images are focused at the energy loss E where the energy window is located (rather than at the elastic image) and that the aperture is filled with the electrons. This assumption is important since, as discussed in Section 5.2, the width of angular distribution of scattering can be smaller and the contribution to the chromatic aberration term would

be modified. As experimental conditions often impose some convergence in the illumination and the chromatic aberration term is small compared to the subsequent broadening terms for low scattering angles (either limited by β or θ_E) this assumption is often assumed to be valid.

2. The delocalization of inelastic scattering term b [introduced in Section 5.2, Eq. (64)] contributes to the broadening with an energy and angular-dependent term increasing at low energy losses and low scattering angles (large scattering angles imply a small impact parameter).

3. The diffraction limit contribution arises from use of the objective aperture and dominates for small angles due to the denominator term

$$d_d = 0.6 \lambda / \beta \quad (69)$$

4. The spherical aberration term $d_s = 2C_s\beta^3$ strongly varies with β and is considered to contribute to a uniform background in the image and a decrease of contrast when the conditions are optimized for minimal chromatic aberration contributions (Krivanek et al., 1995b). Egerton's work (1999) showed that this term is smaller than the chromatic aberration term for typical energy windows used for EFTEM mapping, but for small energy windows (a few electronvolts wide as used, for example, in EFTEM spectrum imaging) the term will dominate the resolution at high collection angles and will need to be included in the analysis. Experimental evidence suggests that the term should be neglected given the good spatial resolution of energy-filtered images with relatively large collection angles.

Additional terms affecting the resolution depend on the noise in the images (requiring averaging of signals and increase of acquisition time), radiation damage of the specimen due to the high doses required for imaging at core losses, and instabilities of the sample and microscope leading to drift of the area under analysis during acquisition.

Considering the significant terms added in quadrature it is possible to determine the ultimate physical limits to EFTEM mapping resolution (Krivanek et al., 1995b)

$$d_{\text{tot}}^2 = d_c^2 + (2b)^2 + d_d^2 \quad (70)$$

The trends, as a function of collection aperture, suggest that there are optimal operating conditions for a given energy loss, energy window ΔE , and microscope characteristics (Figure 4–84). When energy losses of the order of a few hundred electronvolts are analyzed, for small collection angles the diffraction limit term dominates, while for large angles the chromatic aberration term is most important. The width of the energy window has a significant effect on the resolution as it is linearly related to the chromatic contribution. Quite often, large energy windows are used to increase the counting statistics and reduce the noise in the images. When the energy window is large, the choice of the optimal collection aperture is more crucial as the chromatic aberration term rises steeply. For modern analytical TEMs oper-

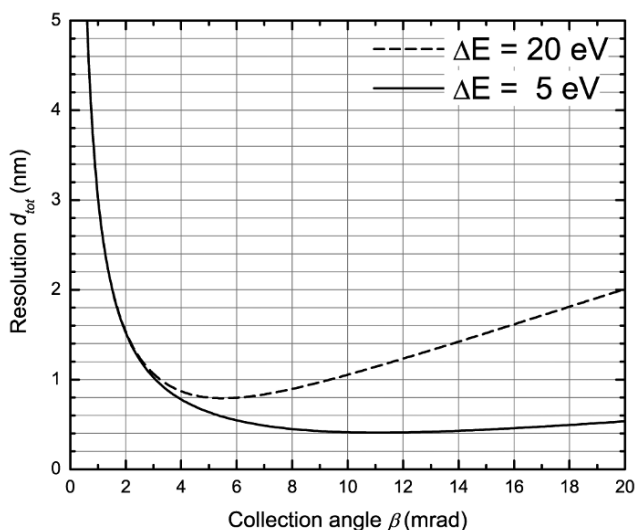


Figure 4-84. Calculations of the expected resolution in EFTEM elemental maps at the oxygen K edge (530 eV) as a function of the collection angle for two energy-filtering windows of width $\Delta E = 5$ and 20 eV and $C_s = C_c = 1$ mm (at 200 keV).

ating at or above 200 keV, and for lenses with chromatic aberration parameters around 1 mm, the delocalization term [Eq. (64)] is not a limiting factor for core losses above around 500 eV. For lower energy losses, in the order of 100 eV or less, this term can dominate the resolution limit in theory. Experimental results in the literature, however, suggest that the low losses delocalization terms based on Eq. (64) are overestimated (Muller and Silcox, 1995; Grogger et al., 2005). For high core losses, elemental maps show resolution limits consistent with the calculations of optimal values given above (e.g., below 1 nm based on Figure 4-84) while for low energy losses (including plasmon losses around 10–20 eV), images with a resolution around 1 nm (significantly better than the prediction of 4–5 nm) have been obtained (Grogger et al., 2005).

Based on the significant contributions of chromatic aberrations for EFTEM imaging, the optimal approach to achieve the ultimate spatial resolution limits in EELS mapping, as imposed by unavoidable delocalization, is to use the STEM approach with small and high-intensity probes achievable today on modern analytical electron microscopes. With current technology making use of bright electron sources and aberration correctors it is possible to focus several hundred picoamperes of current into a near 2 Å probe [Sections 2.1 and 2.2 and Chapter 2 (this volume)].

Examples of chemical analysis with a noncorrected STEM demonstrate the ability to resolve individual atomic planes of Ca in the $\text{Bi}_2\text{Sr}_2\text{Ca}_1\text{Cu}_2\text{O}_{8+\delta}$ superconductor (Figure 4-85) and analyze half-unit cell defects where the Ca planes are not present, suggesting the existence of a $\text{Bi}_2\text{Sr}_2\text{Ca}_0\text{Cu}_1\text{O}_{8+\delta}$ subunit cell intergrowth.

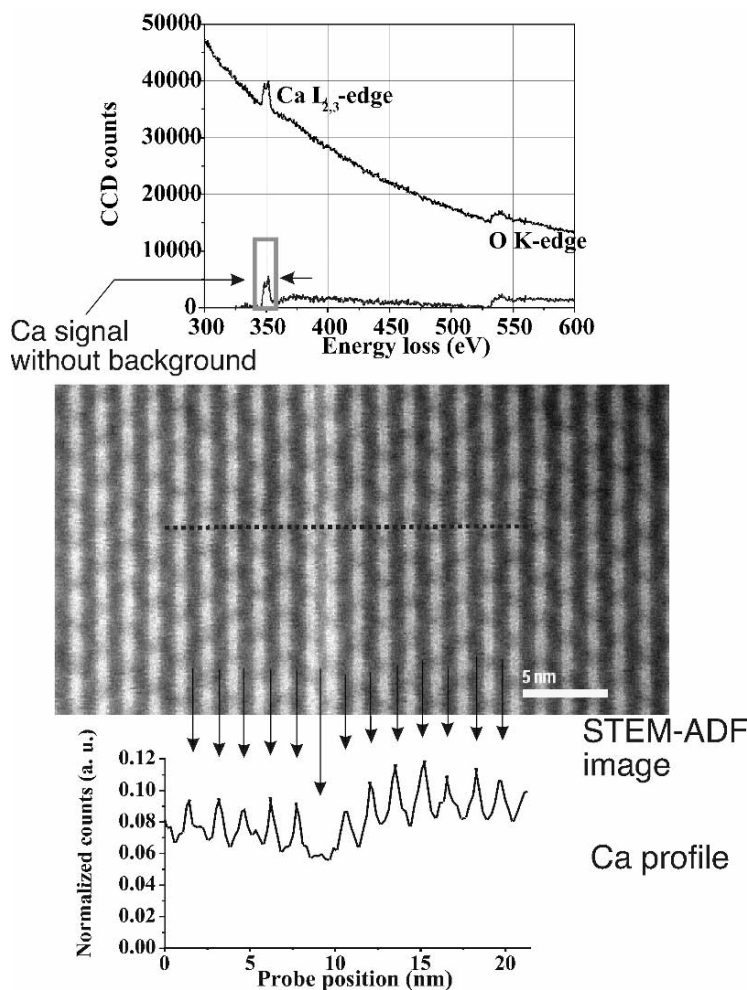


Figure 4-85. High-resolution EELS profile of Ca in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8.8}$ based on the Ca $L_{2,3}$ edge showing the detection of a single plane of Ca in the half-unit cell of the structure. The central panel shows a high-angle annular dark-field image of the sample. The profile shows the presence of a half-unit cell defect of the $\text{Bi}_2\text{Sr}_2\text{Ca}_0\text{Cu}_2\text{O}_{8.8}$ phase (where there is a missing plane of Ca) in the structure. (Courtesy of Y. Zhu, McMaster University.)

7 Detection Limits in Microanalysis

The very high spatial resolution of both EDXS and EELS results in relatively low detection limits for most elements as compared to “bulk” analysis methods. This effect is caused by a combination of factors including the small analyzed volumes, the low signals resulting from poor efficiency in signal collection, or/and low incident beam current, high background, short acquisition time, and instrumental contributions. Two quantities characterize the detection limits for EDXS and EELS depending on information of interest. The minimum detectable fraction (MDF) refers to a dilute element uniformly distributed in the

analysis area and represents the lowest concentration that can be detected. The minimum detectable number (MDN) refers to the smallest number of atoms that can be detected. This applies to analyses in which these atoms are clustered (such as biological molecules containing few atoms, atomic clusters, etc.) or very small probes are used and only a few atoms of impurity are being analyzed in the interaction volume.

Predictions of the MDF are related to the process of signal generation (incident number of electrons, cross sections, fluorescence yield), collection efficiency of the detector system (solid angle, detector quantum efficiency, noise), sample characteristics (thickness, scattering outside apertures leading to loss of signal, background signal, and signal extraction variance), and potential sources of noncharacteristic signals arising from the instrument.

The overall principle for estimating the MDF is based on the statistical certainty of detecting a signal above a background containing some noise due to the variance in the background intensity. This detection is based on the Rose visibility criterion (Rose, 1970) which states that the signal should be three times the standard deviation of the background (the noise) for reliable identification of the signal as genuine in a 98% confidence level. This criterion ensures that the signal can be clearly distinguishable from a simple statistical variation of the background with a high degree of confidence. This general criterion is applicable in both EELS and EDXS measurements (as in any other signal processing method).

7.1 Detection Limits for EDXS

We can describe the evaluation of the MDF for EDXS as follows. Assuming that the noise in the background signal follows a Poisson statistical distribution, the minimum detectable signal for element B ($I_B^{\min}$) in EDXS measurements is $I_B^{\min} \geq 3\sqrt{2 \cdot I_B^b}$ where the square root term represents the variance of the background signal under the peak of element B (I_B^b), i.e., the noise (Figure 4–86). Experimentally, the detection limit of an element B in a matrix A can easily be determined by using a standard of known composition C_B and the detected signal I_B . The MDF is the concentration resulting in a minimum signal-to-noise ratio (SNR) = 3, hence $SNR_{\text{std}}^{C_B} = (I_B - I_B^b) / \sqrt{2 \cdot I_B^b}$ and $SNR_{\text{MDF}} = 3$ (the SNR corresponding to the Rose criterion and the MDF). This resulting concentration yielding an SNR = 3 is the minimum detectable fraction:

$$C_B^{\text{MMF}} = \frac{C_B \cdot 3\sqrt{2 \cdot I_B^b}}{I_B - I_B^b} \quad (71)$$

If the standard containing element B is not available, we can still retrieve the detectable fraction based on a pure element A standard using k_{AB} factors (see Section 4) and by assuming that the background under the peak of element B (if it was present) is not significantly altered when this element is in trace concentration in a sample. In this case, we obtain

$$C_B^{\text{MMF}} = \frac{C_A \cdot 3\sqrt{2 \cdot I_B^b}}{k_{AB}(I_A - I_A^b)} \quad (72)$$

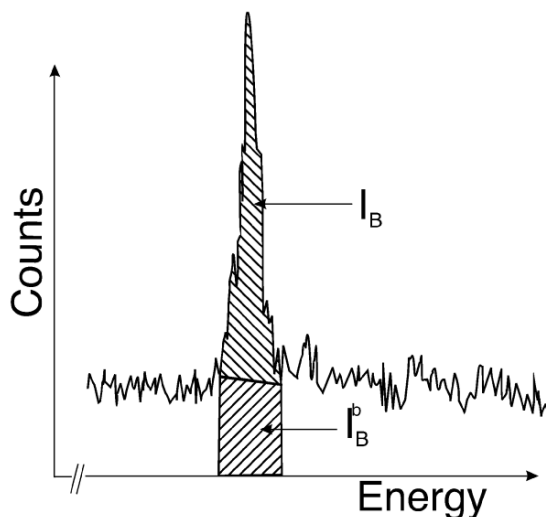


Figure 4-86. Definition of peak and background intensities in EDXS spectra. The peak intensity is defined with respect to intensities just beside the peak.

where I_A^b is the background under the peak of element A and C_A is the concentration of element A (assumed to be 1 if the standard is pure).

From this simple empirical treatment it can be seen that improvements in detection limits (i.e., lower values) can be achieved by increasing the peak intensity (through a longer acquisition time and/or larger probe current) and by reducing the background. This latter contribution is affected by sample thickness, instrument operating conditions, and instrumental contributions (discussed in Section 7.2). The general treatment required to understand the trends in detection limits is due to Ziebold (1967), who related the MMF to microanalysis conditions (Goldstein et al., 1986)

$$MMF \propto \frac{1}{\sqrt{P(P/B)n\tau}} \quad (73)$$

where P is the peak above the background count rate, P/B is the peak-to-background ratio (with integration of the background defined over the same energy window as the peak), n is the number of analyses, and τ is the acquisition time for each analysis carried out. It is possible to increase the peak counts by increasing the electron beam current, the thickness of the sample, and the collection efficiency (larger solid angle of the detector) while increasing the P/B ratio by increasing the accelerating voltage of the microscope and reducing the instrumental contribution leading to noncharacteristic signals. Although a large fraction of the background arises from bremsstrahlung radiation in the sample, significant contributions can come from spurious signals arising from electrons scattered at high angles (including backscattering), which, in turn, generate X-rays within the column. An additional contribution to the background arises from hard X-rays generated in the condenser lens system—when electrons hit apertures in the optic path—that fluo-

resce X-rays in the specimen area, including the sample holder. Such spurious signals are known as instrumental contributions.

As discussed above, increased count rates for a given spot size can be achieved through brighter electron sources (Section 2.1) and thicker samples, although the latter will also lead to increased undesired contributions (increase in background) and loss of spatial resolution (Section 5). Increased total analysis time $n\tau$ can be achieved only if the sample is stable under the electron beam (due to electron beam damage, contamination, etc.) and sample drift is minimal or can be corrected via alignment algorithms. A clean vacuum system and clean samples are of the utmost importance. Dry pumping systems, bakeable columns, clean sample holders always kept in vacuum, and plasma cleaning of the samples prior to the TEM sessions are key components of improved analytical performance of the microscope and are ultimately as important as the quality of the EDXS detector and the microscope. With the use of analytical electronic microscopes equipped with small electron beams, thin samples, and more recently aberration correctors, it has been possible to achieve fractions of 1% detection with subnanometer spatial resolution (Figure 4–87) (Watanabe and Williams, 2005a).

7.2 Instrumental Contributions in EDXS

These spurious signals can be minimized by improvements in the column, detector, and sample holder designs. For example, a significant reduction in hard X-rays generated in the upper part of the column can be achieved by using thick top-hat-shaped Pt apertures. These are now available in most modern analytical microscopes as part of the selection of C_2 apertures (some might just be Mo or lighter elements allowing transmission of hard X-rays). Good collimation of the detector is also important to minimize the line of sight between the specimen and the active area of the detector so that X-rays generated elsewhere in the speci-

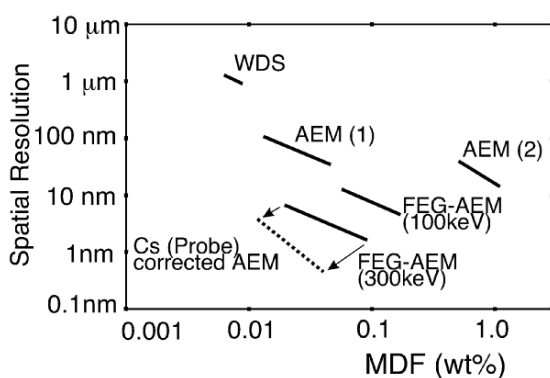


Figure 4–87. Detection limits reached in modern AEM and various examples of performance reached with different instruments. WDS is data obtained with a dedicated microprobe (30 keV), AEM (1) 120 keV, AEM (2) 100 keV, LaB₆ instruments, 100 keV FEG, 300 keV (Watanabe and Williams, 1999), and aberration-corrected FEG. (Adapted from Watanabe and Williams, 2005a, and Williams and Carter, 1996.)

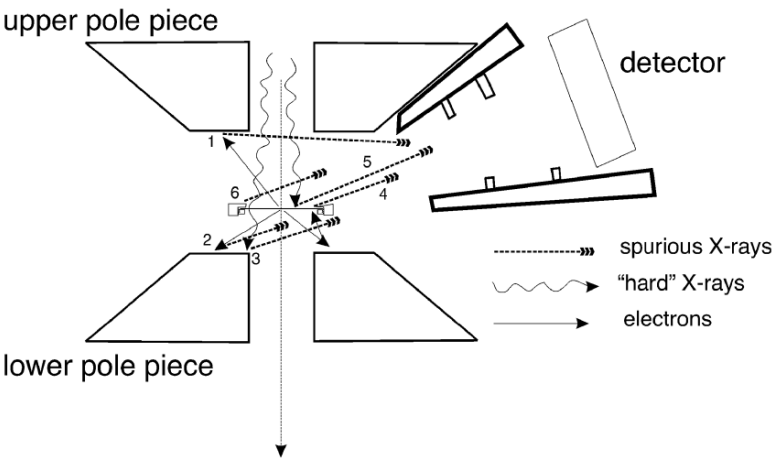


Figure 4–88. Instrumental contributions to the characteristic signals detected from the sample. Spurious X-rays are produced by the scattered electrons interacting with the microscope components (labeled 1, 2, 4). Spurious signals are also produced by hard X-rays generated on the top part of the column hitting the sample holder, grid, and other parts of the specimen chamber (labeled 5, 6, and 3). Nonoptimal collimation allows X-rays generated other than the sample area to reach the detector (compare with Figure 4–23).

men chamber area by scattered electrons reaching apertures, the pole-piece of the objective lens, do not reach the detector (Figures 4–23 and 4–88). Using small samples (rather than the standard 3-mm disks with bulk edges) also minimizes the contributions due to fluorescence and electron scattering (electrons returning back to the sample after hitting parts of the column). Finally, the sample holder must be designed for analytical purposes (these are available commercially), namely, built with light element materials, ideally Be as one of the main components in the specimen supporting area, with a Be ring to tighten the sample and a design geometry allowing the acquisition of spectra without significant tilt of the sample (Figure 4–89). This configuration leads to minimal instrumental contributions so that X-rays generated in the sample have a

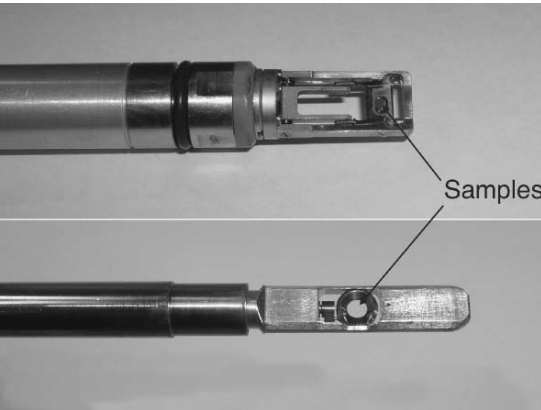


Figure 4–89. Sample holders for AEM (top) JEOL 2010F microscope holder and (bottom) Philips/FEI CM-Tecni series holder with machined groove for minimal tilt of sample.

direct line of sight to the detector without being absorbed (and causing fluorescence) in the sample holder. Sample grids made of Cu (or other metals such as Ni or Mo) to support continuous, lacey or holey C films where the sample is distributed (in the form of dispersed particles, replicas, or lift-out focused ion beam samples) generate significant spurious signals as shown by the presence of peaks of the grid material even if the electrons are not directly illuminating the grid. Recently, diamond grids have been developed to minimize these contributions when these problems affect the analysis. In many cases, the selection of the grid material can be made judiciously to avoid quantification problems given the commercial availability of several grid materials.

Based on these instrumental considerations, the overall quality of the microscope analytical performance can be evaluated with measurements of the “hole-count” signals. By measuring the signal generated when the electron beam is directed into the hole of the sample, contributions from the grid, the microscope chamber, and holder can potentially be observed (and identified from the element present) and should be, in a good analytical TEM, less than 1% of the signal generated on the sample. Quantitative evaluation of the performance can be carried out using standard samples of Cr (100 nm thick) and NiO and well-defined tests based on *P/B* ratios (Williams and Carter, 1996; Egerton and Cheng, 1994). Overall, low instrumental contributions lead to lower peak-to-background values and improved detection limits. Well-designed analytical microscopes (Section 2.3.6) and the use of analytical conditions (correct analytical apertures, sample geometry, tilt) lead to significantly improved detection limits.

7.3 EELS Detection Limits

The basic statistical principle for detection of signals used in EDXS (the Rose criterion) is also applicable in the case of EELS. The empirical approach to estimate the detection limit based on experiments and known standards is the same as in EDXS but with additional complications due to the determination of the noise component arising from the extrapolation of the background rather than the simple interpolation. The noise, and thus the SNR, cannot be determined based on the simple variance of the number of counts at a given energy loss. The noise must be estimated directly from a detailed statistical analysis of the spectra and the errors related to the determination of the extrapolation parameters (Trebbia, 1988) or simpler approximations of the extrapolation error based on the width of the fitting and extrapolation windows (Egerton, 1996).

From first principles, it is possible to estimate the detection limits accounting for these statistical effects and physical principles as described in detail by Egerton (1996) and Egerton and Leapman (1995). The variance of the signal must take into account the error due to the extrapolation of the background under the edge and the introduction of noise related to the detective quantum efficiency (DQE) of the spectrometer. The DQE is defined (Krivanek et al., 1987; Egerton, 1996) as $SNR_{\text{output}}^2 / SNR_{\text{input}}^2$ (where the indices “input” and “output” are based on the

incident counting statistic and measured statistic, respectively) and varies with the number of incident electrons, but, for the purpose of our estimations, is assumed constant and not equal to one as generally assumed in the case of EDXS spectra. The determination of the EELS detection limits must also include the effect of elastic scattering reducing the efficiency of collection of the signals due to scattering outside the aperture. Accounting for the detector response, the SNR is (Egerton, 1996),

$$SNR \approx \sqrt{DQE} \frac{I_k}{\sqrt{h \cdot I^b}} \quad (74)$$

where I^b is the background under the edge, h is an error parameter due to the extrapolation, and I_k is the integrated edge intensity. Large fitting intervals close to the edge threshold and small integration windows lead to small h parameters (around 5–10), while small fitting intervals further from the edge and large integration windows result in large h values (as large as 20–30) (Egerton, 1996). The number of electrons causing a signal is related to the dose D [coulomb/(unit area)] and the area illuminated by a probe with diameter d as

$$I(\beta, \Delta) \approx (\pi/4)d^2(D/e)\exp(-t/\lambda_e) \quad (75)$$

where the exponential term represents the loss of electrons following elastic scattering outside the collector aperture β in a specimen of thickness t relative to the elastic mean free path λ_e . The background signal (I^b) related to the number of atoms of the matrix N_t and the number of atoms of the trace element N giving rise to edge signals I_k are related to the respective cross sections and the number of incident electrons and number of atoms in the volume as

$$I^b \approx N_t I(\beta, \Delta) \sigma_b(\beta, \Delta) \text{ and } I_k \approx N I(\beta, \Delta) \sigma_k(\beta, \Delta) \quad (76)$$

Since the fraction of trace element $f = N/N_t$ we obtain

$$f = \frac{SNR}{\sigma_k(\beta, \Delta)} \left[\frac{h \sigma_b(\beta, \Delta)}{N_t I(\beta, \Delta)} \right]^{1/2} (DQE)^{-1/2} \quad (77)$$

The MDF is the fraction $f_{\min}$ corresponding to an SNR = 3 (as in the case of EDXS following the Rose criterion). This yields (Egerton, 1996)

$$MDF = f_{\min} \approx \frac{3}{\sigma_k(\beta, \Delta)} \left(\frac{1.1}{d} \right) \left[\frac{h \sigma_b(\beta, \Delta)}{(DQE)(D/e)N_t} \right]^{1/2} \exp\left(\frac{t}{2\lambda_e}\right) \quad (78)$$

The minimum detectable number quantity is related to the number of atoms within the analyzed area corresponding to the minimum detectable fraction determined above:

$$MDN = \frac{\pi}{4} d^2 f_{\min} N_t \approx \frac{2.7d}{\sigma_k(\beta, \Delta)} \left[\frac{N_t h \sigma_b(\beta, \Delta)}{(DQE)(D/e)} \right]^{1/2} \exp\left(\frac{t}{2\lambda_e}\right) \quad (79)$$

By calculating the various cross sections (Section 3), detection limits can be estimated. Assuming a DQE ≈ 0.5 typical of parallel spectrometers, $h = 9$, $\lambda_e = 200$ nm, a carbon matrix of 30 nm thickness, based on the

expression given above (Egerton, 1996) calculated the detection of Ca atoms for a range of doses using 100 keV electrons and hydrogenic cross sections. The detection of a few atoms within a 1-nm probe is possible assuming the sample can withstand doses of the order of 10^6 C/cm^2 . For elements with large cross sections (such as L and M lines) detection of a few ppm has been demonstrated (Leapman and Newbury, 1993) while higher detection limits are possible for K edges and delayed edges of elements such as Ag, Au, and S. A significant improvement in the detection limits is achieved if the spectra are processed using multiple least-squares analysis of data acquired in spectral-difference mode (see Section 4.2.1) where reference data are fitted to the experimental spectrum. This approach results in a much lower signal extraction error with h values approaching 1. Detection of single atoms of thorium was demonstrated by Krivanek et al. (1991b) and individual Gd atoms were detected by Suenaga et al. (2000). More recently, systematic work on Ca and Fe containing molecules demonstrated elemental maps with the detection of just a few (7–8) base pairs of a DNA molecule under the electron beam (within the pixel size) containing 14–16 P atoms and 4 atoms of Fe in the one single hemoglobin molecule (Leapman and Rizzo, 1999) and single Ca atom detection (Leapman, 2003). Experiments demonstrating detection of a few ppm in standard reference materials of known composition have been reported by Leapman and Newbury (1993), Newbury (1998), and Newbury et al. (2000) (Figure 4–90). The work assumes detection of the edges based on spectral difference technique (Section 4.2.1), parallel detection spectrometers, and very long acquisition times. These detection limits also assume very thin samples, typically $t/\lambda < 0.3\text{--}0.5$ (see Section 4.2.4). If thicker samples are used, the detection limits degrade significantly up to the point at which even pure elements would not be detected in samples as a result of the increase in the background due to multiple scattering (Section 8.2) that masks the edges. A summary of the detection limits based on the work of Leapman and Newbury (1993; Newbury, 1998; Newbury et al., 2000) is presented in Table 4–6.

A detailed simulation package predicting the detection limits including the effects of increased background, multiple scattering, and angular collection has been developed commercially (available by Gatan as the “EELS advisor”) based on the initial work of Natusch et al. (1999) and Menon and Krivanek (2002). Applications of these simulations are extremely useful in predicting whether elements present at low concentration levels in any matrix can be detected and in suggesting experimental conditions to detect these elements.

7.4 Comparison of EELS and EDXS Detection Limits

An expression for the detected X-ray signal based on the incident current I and acquisition time T can be deduced so that the relative merits of EDXS and EELS can be compared (Egerton, 1996).

Following the description given for X-ray signal quantification (Section 4.1) we describe the intensity

$$I_x = N(I/e)T\omega_k\sigma_k\eta_x \quad (80)$$

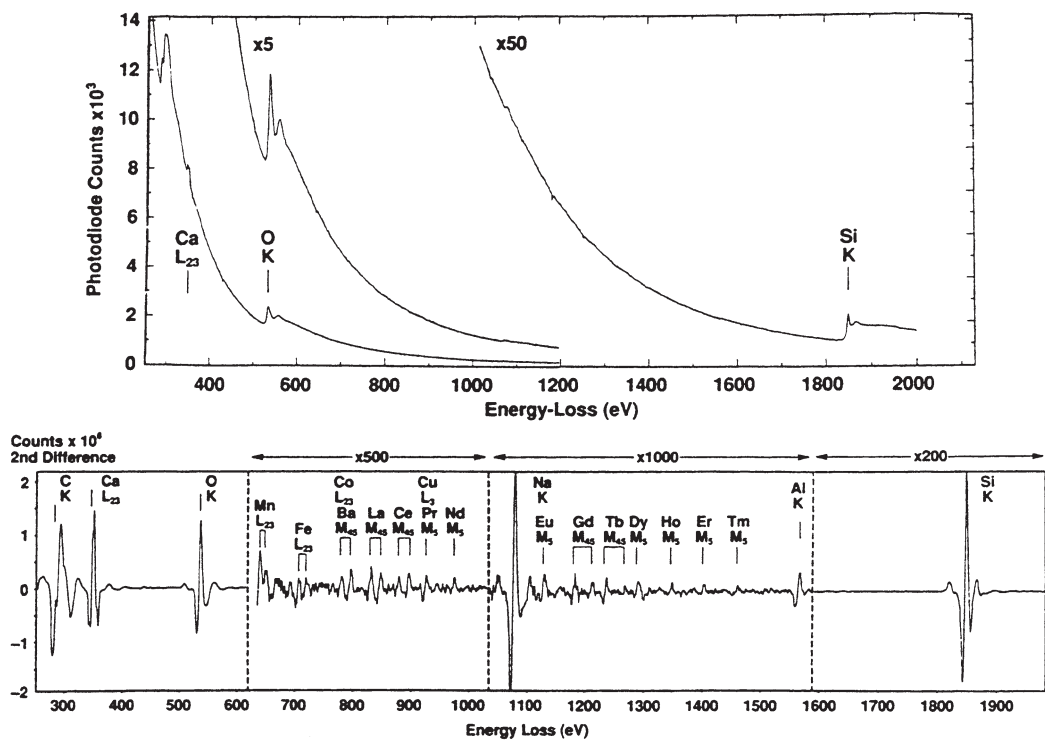


Figure 4-90. Extraction of signals using the second-difference technique for the detection of trace concentration of reference materials in a standard sample (SRM 610 glass from NIST). (Reprinted from Leapman and Newbury © 1993, American Chemical Society.) In the raw spectrum nearly all edges of trace constituents are not visible. In the second difference spectrum (bottom) the edges of trace elements are well resolved. Elements Ba, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, and Yn are present in the 60–80 ppm range. Sc, Ti, V, Cr, Mn, Fe, and Co are present in the 150–240 ppm range.

Table 4-6. Edge shapes and labels for elements of the periodic table as detectable in EELS experiments.

Detection limit	Element and edges	Edge type (following the Gatan chart convention)
10–100 ppm	3d transition metals and preceding alkaline earth elements: L edges (Ca, Sc, . . . Ni)	
	4f lanthanides and preceding alkaline earth elements M45 edges (Ba, La, . . . Yb)	
100–1000 ppm (0.1%)	Light elements: K edges (Li, Be, B . . . P)	
	L23 edges (Mg, Al, Si, P, S, Cl)	
	4d transition metal elements and preceding alkaline earth elements L23 edges (Sr, Y . . . Rh)	
>1000 ppm (0.1%)	Ga, Ge, As, Se, Ag, Cd, In, Sn, Sb, Te, I, W . . .	

where ω_k is the fluorescence yield for the particular line, σ_k is the total cross section (integrated over all the angles; see Section 3.4.2), and η_k is the collection efficiency including all the detector components and any absorption in the sample.

The relative sensitivity of EELS and EDXS signals can be calculated by discussing the ratio of EELS and EDXS intensities

$$\frac{I_k}{I_x} = \frac{\sigma_k(\beta, \Delta)}{\omega_k \sigma_k} \frac{1}{\eta_x} \exp\left(\frac{-t}{\lambda_e}\right) \quad (81)$$

The ratio of the partial and total cross-sectional term is typically in the order of 0.1 (Egerton, 1996), although large scattering angles and energy windows ($\Delta > 50$ eV) would tend to include most of the inelastic distribution given the asymptotic behavior of the scattering distribution and the cross sections (Section 3.4). The fluorescence term is the major contributor to the greater effectiveness of EELS for the detection of a large number of elements in "ideal" samples. For K edges of heavy elements with $Z > 40$ (difficult to reach in TEM-EELS experiments given their high energy loss), ω approaches 1 (Figure 4-49), but for light elements or L edges, it drops very sharply (for K edges it is around 0.1 for Al, 0.02 for Na, down to 0.001 for B, while for L edges it is less than 0.001 for $Z < 20$). The detector efficiency term also has a significant effect on the relative merit of EELS. The small solid angle of the EDXS detector implies that only 1% of the emitted X-rays are collected by the detector. Furthermore, low-energy X-rays can be absorbed in the thin window and the dead layer of the detector (Section 2.3.2). Due to the combined terms of fluorescence and detector efficiency, the EELS signals are three to four orders of magnitude stronger than EDXS signals for light elements while they are slightly stronger for most heavy elements. The relative merits are summarized in Figure 4-91. The drawback with EELS, however, is not fully accounted for in Eq.

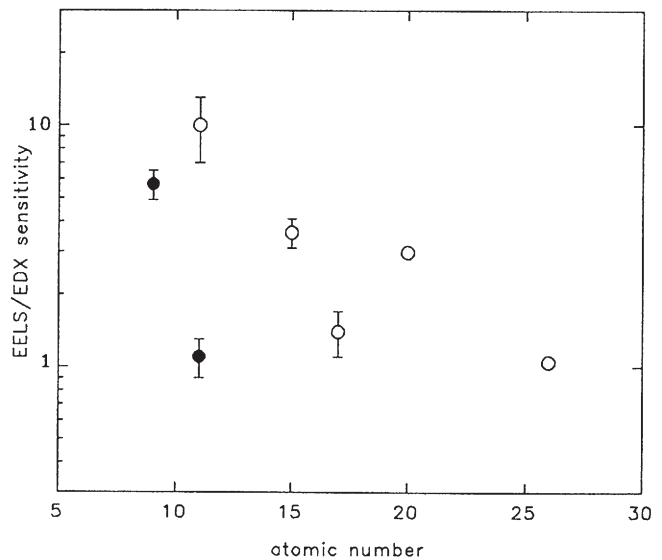


Figure 4-91. Comparison of the relative sensitivity of EELS/EDXS detection limits. Note that the calculations assume samples are very thin. (Reprinted from Egerton and Leapman © 1995, with permission from Springer Science+Business Media.)

(81) due to the strong increase of background with thickness arising from multiple scattering. For thick samples, even of pure elements, EELS would not show any edge irrespective of the atomic number, while EDXS spectra would still show peaks (even for light elements, albeit small)! The absence of EELS edges in a spectrum from just a sample does not necessarily imply the absence of any elements.

8 Energy Loss Fine Structures

As discussed in the introduction of this chapter, there are fine modulations in the structure of spectra that yield useful information on the chemical environment of the atoms and the dielectric properties of the material. These fine structures can be subdivided into three parts. The ELNES are modulations appearing in the first 10–20 eV from the ionization edge threshold (Figure 4–92). These are now used almost routinely to characterize the chemical environment of atoms, including the type of phases and valence state.

At higher energy losses (from about 30–50 eV of the threshold up to several hundred electronvolts), the extended energy loss fine structures (EXELFS) provide information on the radial distribution function of the material [similar to X-ray absorption fine structures (XAFS)] (Figure 4–92). These modulations arise from the backscattering of the ejected electron in the solid and the creation of interference between the ejected and backscattered wavefunctions and are particularly useful in providing the bond distances in amorphous solids at a nanometer scale. Since current applications of the technique in the AEM literature are limited we will refer the reader to a good overview of this technique in Egerton (1996) and to XAFS literature describing the principles of the analysis method.

Finally, the fine modulations in the low-loss part of the EEL spectra (from 0 to 50–100 eV) also provide a wealth of information on the dielectric properties of materials. Quantitative analysis makes it possible to compare optical spectroscopy measurements to low-loss energy loss data and measure electron density using some simple approximations for metals. Qualitative analysis allows us to use the differences in spectra for various materials to map the distribution of phases. Given the impact of the ELNES and low-loss spectra in AEM, details of these two techniques are presented below.

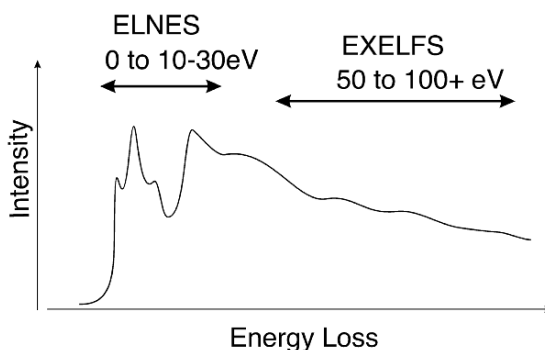


Figure 4–92. Regions and energy ranges for the energy loss near-edge structures (ELNES) and extended energy loss fine structures (EXELFS) of core edges.

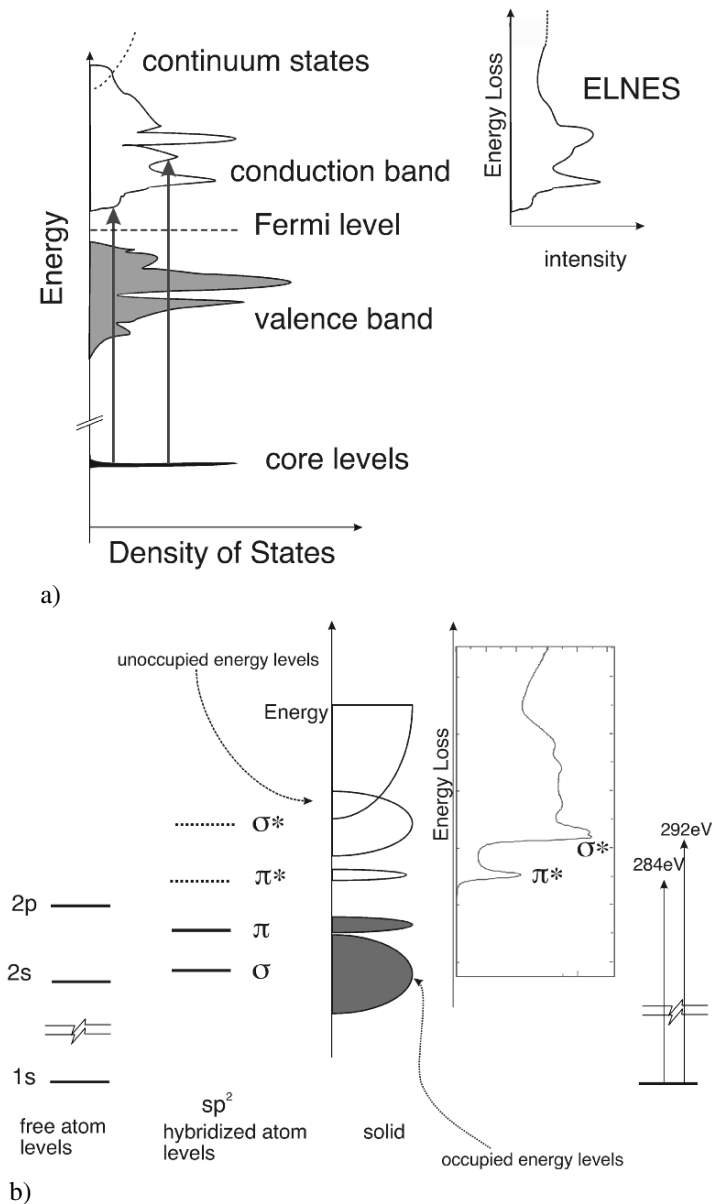


Figure 4-93. Relationship between the near-edge structure observed on the EELS edges and the unoccupied electronic states. (a) Transitions are observed from core levels to unoccupied electronic states above the Fermi level. (b) Example of near-edge structure (experimental spectrum) for the C K edge in graphite with the relationship between the π and σ orbitals (and the antibonding orbitals π^* and σ^*) in the hybridized atoms and the related bands in the solid.

8.1 Energy Loss Near Edge Structure

As discussed in the introductory section of this chapter, ELNES provide information on the electronic structure and bonding environment of the atoms probed by the incident fast electrons. An example of the information is demonstrated in Figure 4-93 showing the relationship between

the spectrum and the energy states along with examples of near edge structures for different compounds (Figure 4–94). The features visible in the near edge structure represent the unoccupied energy states as modified with respect to a free atom by effects such as hybridization, coordination changes, and solid-state effects. The technique therefore provides data equivalent to the well-established X-ray absorption near-edge structure (XANES) spectroscopy carried out in synchrotrons. As such, reference data and literature from XANES can often be used to identify compounds and understand trends and electronic structure effects in ELNES. Examples of changes in the ELNES derived from changes in the electronic structure demonstrate the sensitivity to the structural environment and the chemical state (Figure 4–94). In addition to the changes in the shape of near edge structures, the energy position of the edges can vary with the oxidation state in a manner similar to X-ray photoelectron spectroscopy. Changes in the core energy level and the position of unoccupied states can result from charge transfer effects due to oxidation, bonding, and coordination changes. Systematic trends

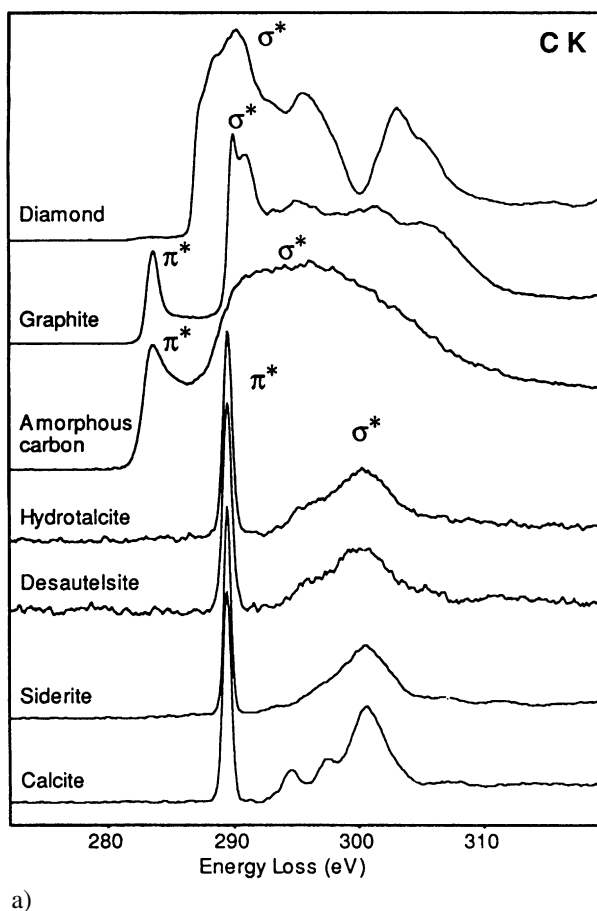
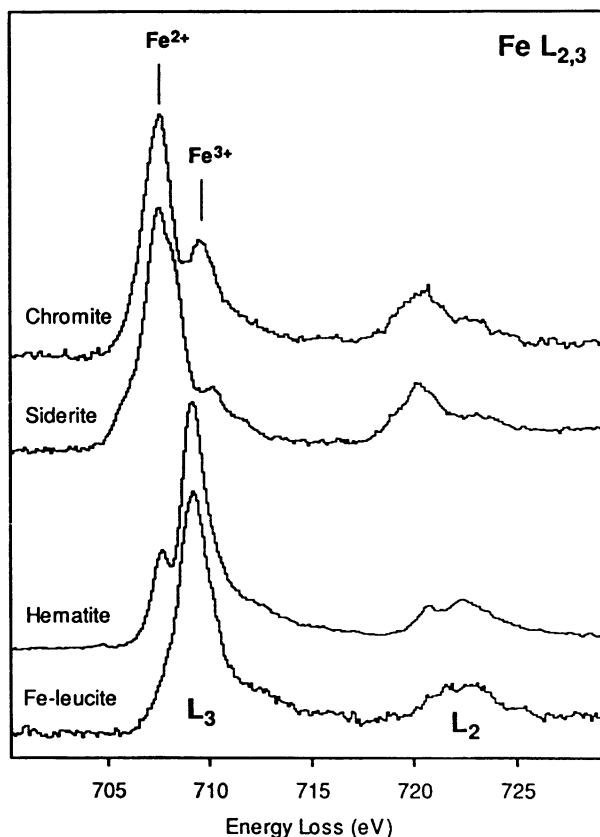


Figure 4–94. Examples of near-edge structures in (a) various carbon-based compounds showing the sensitivity to the structural environment and hybridization and in (b) Fe-based compounds. (From Garvie et al., 1994.)



b)

Figure 4-94. *Continued*

are therefore observed for several metals with oxidation states for both X-ray absorption spectroscopy (Chen, 1997) and for EELS (e.g., Mansot et al., 1994). Detailed reviews on applications of ELNES can be found in Botton (1999), Keast et al. (2001), and Garvie et al. (1994). Examples demonstrating the application of ELNES at high spatial resolution are given in Batson (1993), Muller et al. (1999, 2004) and Spence (2005).

At a more quantitative level, the sensitivity of the ELNES to bonding can be explained by the general oscillator strength and the form factor terms of the partial cross sections (Section 3.3). These two terms are dependent on the initial and final state wavefunctions of the interacting electrons and thus contain information on the chemical state and electronic structure of the probed atoms as modified in the solid. In the dielectric formulation of the cross sections (Section 3.3.1), it is also possible to understand how the energy loss spectrum relates to other spectroscopy measurements:

$$\frac{d^2\sigma}{d\Omega dE} \propto \left(\frac{1}{\theta^2 + \theta_E^2} \right) \text{Im}[-1/\epsilon(q, E)] \quad (82)$$

where ϵ is the dielectric function of the material (Section 3.3.1), which can be expressed by its real part ϵ_1 (related to the screening process of

the electrons) and its imaginary part ε_2 (related to the absorption process and thus to optical and X-ray absorption measurements). The first term in Eq. (82) is general and is related to the kinematics of scattering. It imposes a simple Lorentzian angular distribution to the scattering and the rapid drop in intensity with increasing scattering angle. The second term is the loss function, which is related to the dielectric response of the solid to an electromagnetic radiation and is therefore ultimately linked to the intrinsic properties of the solid. At high energy losses (from about 50–100 eV and above) where the screening of the incident electron charge by collective effects is not important, $\varepsilon_1 \rightarrow 1$ and ε_2 is small ($\varepsilon_2 \ll \varepsilon_1$) so that the loss function is reduced to $\text{Im}(-1/\varepsilon) = \varepsilon_2/(\varepsilon_1^2 + \varepsilon_2^2) = \varepsilon_2$ and is thus directly related to the absorption part of the dielectric function and therefore to absorption measurements. In this particular condition and in the one electron approximation, the cross section can be calculated based on Fermi's golden rule describing the transition rate of an electron from an initial energy level i described by an initial wavefunction ψ_i of energy E_i to a final level f described by a wavefunction ψ_f of energy E_f

$$\frac{d^2\sigma}{d\Omega dE} \propto \sum_{if} |\langle \Psi_f | \exp(i\theta \cdot \rho) | \Psi_i \rangle|^2 \delta(E - E_f + E_i) \quad (83)$$

where we have omitted the first (kinematic) term from Eq. (82) because it is independent of the solid state effects. As defined in Section 3.3, the scattering vector $\mathbf{q} = \mathbf{k}_0 - \mathbf{k}_1$ where $\mathbf{k}_0$ and $\mathbf{k}_1$ are the incident and final wavevectors of the incident electron. The exponential operator is derived from the Hamiltonian describing the interaction between the incident electron, the nucleus, and the atomic electrons in the solid [see Paxton (2005) for a detailed derivation of this term]. The sum is carried out over all possible final energy states limited by the δ function to ensure the conservation of energy in the scattering event so that the energy loss E corresponds to the difference in the energies of the final and initial states. The exponential operator can be expanded as a Taylor series

$$\exp(i\mathbf{q} \cdot \mathbf{r}) = 1 + i\mathbf{q} \cdot \mathbf{r} + \frac{1}{2}(i\mathbf{q} \cdot \mathbf{r})^2 \dots \quad (84)$$

The first term of this expansion is zero due to the orthogonality of the initial and final wavefunctions (i.e., $\langle \psi_f | \psi_i \rangle = 0$). For most conditions used in EELS experiments (i.e., small spectrometer collection angles β and edges above around 100 eV), only the term related to dipole excitation $\mathbf{q} \cdot \mathbf{r}$ is retained because the scattering angles are small and the core wavefunctions are very localized in typical EELS experiments. In these conditions, $\mathbf{q} \cdot \mathbf{r} \ll 1$ leads to what is known as the dipole approximation and the second term in the Taylor expansion can be neglected (the treatment is thus similar to X-ray absorption spectroscopy). For larger scattering angles or/and for very low core losses (the core state wavefunctions are less and less localized as the core energy decreases, i.e., the mean radius of the core wavefunction increases), this approximation is, strictly speaking, no longer valid but it gives a good representation of the spectra [nonetheless, full calculations with nondipole terms are possible with current methodologies (Blaha et al., 2001)]. In the dipole approximation, we obtain

$$\frac{d^2\sigma}{d\Omega dE} \propto \sum_{i,f} |\langle \Psi_f | q \cdot r | \Psi_i \rangle|^2 \delta(E - E_f + E_i) \quad (85)$$

The squared term is the dipole matrix element representing the transition rate from a core state to a final state. In this condition, the sum in Eq. (85) is simplified as

$$\frac{d^2\sigma}{d\Omega dE} \propto |M_{\ell,\ell-1}(E)|^2 \rho_{\ell-1}(E) + |M_{\ell,\ell+1}(E)|^2 \rho_{\ell+1}(E) \quad (86)$$

where $\rho(E)$ is the density of states (DOS) resolved in angular momentum components ℓ (s, p, d, f). In the development of the matrix elements of Eq. (85), the decomposition of the wavefunctions in radial and angular terms imposes two very interesting effects summarized by Eq. (86). The radial part of the matrix elements reveals that for a transition to be observed (i.e., $M \neq 0$), there must be an overlap between the core states and the final state wavefunctions. This implies that the energy loss spectra probe the site-dependent electronic structure projected on the excited atom. Practically speaking, since the core states are very localized, the information is also very local and is not limited just to the electron beam illuminated area but is specific to the atom type excited by the electron beam: different atoms in the same area illuminated by the electron beam will have a different “local” electronic structure environment. Second, from the development of the angular parts of Eq. (85), the matrix element of a transition from a state of angular momentum component ℓ is zero unless the final state is of angular momentum $\ell \pm 1$ as suggested by the subscripts in Eq. (86). Hence, transitions will therefore occur for states with a change in angular momentum component $\Delta\ell = \pm 1$. This implies that if the initial state is of s character ($\ell = 0$), the transition will occur only to states of p character ($\ell = 1$). This is observed for K edges where the core state is 1s. For an initial state of p character ($\ell = 1$), transitions will be observed to states of s ($\ell = 0$) and d character ($\ell = 2$). This applies to L_{23} edges arising from the $2p_{3/2}$ and $2p_{1/2}$ initial states.

The matrix elements therefore allow us to probe the local DOS of each element *separately* by selecting the edge corresponding to the atomic number of the element of interest and the edge type (K, L, etc.) according to different principal and angular momentum quantum numbers. This latter sensitivity is very illuminating since EELS experiments allow us to probe not only the unoccupied states but also the atom-site-specific and symmetry-projected density of unoccupied states.

The calculations of near-edge structures using Eq. (86) have been demonstrated to be equivalent for XANES and ELNES in most experimental collection conditions respecting the dipole approximation. The only difference in the formulation between XANES and ELNES derives from the fact that the scattering vector $\mathbf{q}$ is replaced by the electric field $\mathbf{E}$ in Eqs. (84) and (85). The ELNES is therefore an extremely powerful probe of the electronic structure of solids with high spatial resolution as compared to other valence or conduction band spectroscopies in which the entire bands (valence or conduction) are probed irrespective of the angular momentum character.

The calculations based on Eq. (86) with density of states determined by first principle methods assume an infinite lifetime of the excited state. To account for more realistic conditions where the decay occurs via deexcitation processes, a finite lifetime must be considered. This can be achieved artificially by broadening the δ function with a Lorentzian distribution to account for the finite lifetime of the excited state, the lifetime of the core state, as well as the instrumental function accounting for the energy spread of the incident electrons and the resolution of the spectrometer.

The derivation of Eqs. (83)–(86) has been described here very briefly and we refer the interested reader to the work of Fink (1992), Vvedensky (1992), Saldin (1987), and Schattschneider and Jouffrey (1995) and more recently a review by Paxton (2005) for further details. It is important to mention, however, some of the limitations of this description to give an idea of what can be expected from first principle calculations. First, the one electron derivation of the transition probability is based on the “single particle” approach. This simplification assumes that the excited state (where there is a hole in the core state and an ejected electron) can be represented by ground state wavefunctions (no excitation effects are accounted for). This most important approximation can be improved on, in principle, by considering the final state rule proposed by von Barth and Grossmann (1982), which considers the electronic structure of the system in the potential probed by the ejected electron, i.e., with a core hole in the initial level. In spite of this approximation, the single particle approach is a first useful step in understanding general features in the spectra and in assessing the need for more refined models accounting for the more realistic final state. For many systems including metallic materials and even some insulators where the screening of the core hole is effective, this description is successful, whereas in others, the interactions between the core hole and the ejected electron significantly modify the ground state electron wavefunctions of the solid.

A second limitation is related to the specific approaches used to describe the electronic structure of the solid. The predictions of the DOS are based on the use of density functional theory and the different implementations to calculate the electron wavefunctions (for a review, see Hébert, 2006). For the most part, electronic structure calculations for solids have focused on the description of occupied states and low-lying unoccupied states. This presents a limitation for the calculations of energy loss spectra that probe unoccupied states 10–30 eV above the threshold (and thus above the Fermi energy). To simplify the computation, linear band structure methods are often used (such as linear muffin tin orbital methods and linear augmented plane wave), only a limited energy range will be reproduced (5–20 eV from the edge threshold). Alternatives to these techniques are the multiple scattering-based techniques such as the Korringa–Kohn–Rostoker method [(used for transition metal edges (Botton et al., 1997)], the real space multiple scattering technique (Ankudinov et al., 1998), and the pseudopotential technique based on the use of plane waves that has produced impressive results at about 40–50 eV above the threshold in diamond (Pickard and Payne, 1997).

In addition to these calculations of ELNES based on band structure techniques or real space methods, other approaches must be used to deal with systems demonstrating strong electron–electron interactions due to the localization of the electrons. For these systems the band structure and real space methods fail to describe the spectra and a different scheme must be used. For transition metals L edges and rare-earth M edges, a description based on atomic multiplet theory is much more effective in describing the spectra. These techniques include solid-state effects by including crystal field and charge transfer effects to describe the structural and bonding environment. A full description of the methods is given in the excellent reviews by deGroot (1994, 2005). Examples of calculations with this technique and a range of other methods are shown in Figure 4–95. A description of the hierarchy of these methods starting from the molecular orbital approach and multiple scattering methods can be found in Rez et al. (1995).

In addition, the flexibility of EELS experiments in the TEM makes it possible to tune very effectively and elegantly the scattering vector of

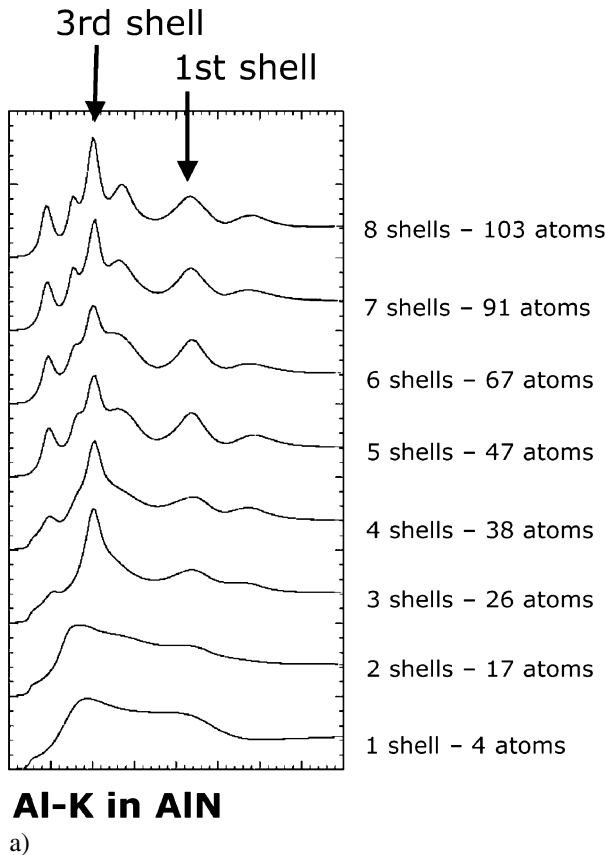


Figure 4–95. Examples of calculations of ELNES with (a) the multiple scattering method for the Al K edge in AlN and (b) bandstructure technique: O K edge in rutile experiments (dots) and calculations (full line) with *abinitio* code Wien2K. (Spectrum courtesy of P. Tiemeijer, FEI.) (c) Atomic multiplet calculations of the MnL₂₃ edge in MnTiO₃. (Courtesy of G. Radtke, McMaster University.)

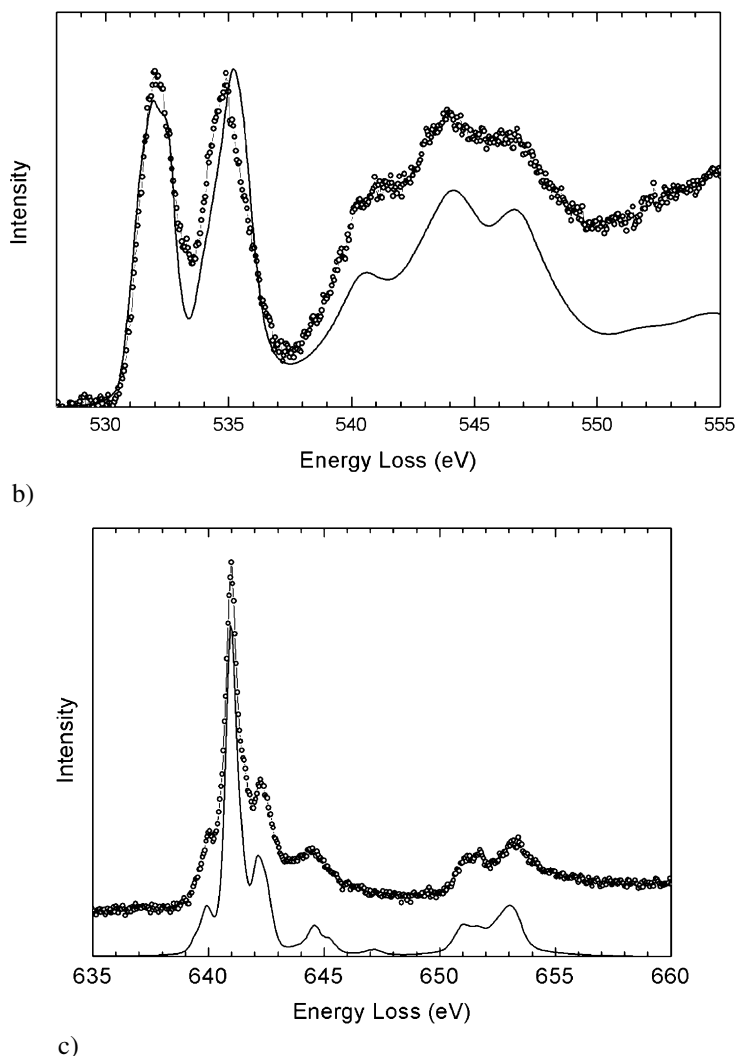


Figure 4-95. *Continued*

the incident electron using momentum-resolved energy loss experiments to study anisotropic materials. Experiments demonstrating this effect in low-loss spectra (Fink, 1989, 1992) and core losses (Radtke et al., 2003; Botton et al., 1995a; Jouffrey et al., 2004) and to record inelastic scattering distributions (Botton, 2005) can be found in the literature.

Various tools are therefore available to materials scientists and microscopists with which to understand and model near edge structures should the need arise. In many cases of AEM applications, however, a detailed comparison of various edges with reference materials is sufficient to understand the trends and associate the spectral features to bonding bands and materials properties. Databases are currently in development allowing users to access libraries of spectra from various compounds: <http://www.cemes.fr/~eelsdb> and <http://people.ccmr.cornell.edu/~davidm/WEELS>.

8.2 Low-Loss Spectroscopy

8.2.1 Fundamentals

At low energy losses (0 to 50–100 eV), the most intense part of the spectrum is directly related to the collective response of the electrons in the solid to the fast incident electron. The incident charges with their associated electric field polarize the medium and set up oscillations of the weakly bound electrons of the solid at particular eigenfrequencies related to the mass of the electrons and their density in the solid. This is analogous to the resonance frequency of an object of mass m attached to a spring with spring constant k of classical mechanics. The collective behavior of these electrons in a solid is best characterized by treating the collective response as an effective particle called *plasmon* oscillating at a frequency ω_p that results in an energy loss of the primary electron equivalent to the energy $E_p = \hbar\omega_p$.

The link between the spectrum, the general oscillator strength, and the loss function presented in Section 3.3.1 can be discussed in the context of low energy losses. Whereas at high energies (Section 8.1) the loss function is essentially determined by ϵ_2 , at low energies the screening of the electrons is very effective and $|\epsilon_1|$ becomes much greater (in simple metals) than (or is about the same order of) ϵ_2 . The loss function $\text{Im}(-1/\epsilon) = \epsilon_2/(\epsilon_1^2 + \epsilon_2^2)$ must then consider both the polarizability and the absorption terms essentially describing the dielectric response of the solid to electromagnetic radiation.

Detailed derivations of the models describing the formulation of the dielectric response have been reviewed by Schattschneider and Jouffrey (1995) and Raether (1980) and only a brief summary is given here to explain the most significant features in the spectra and the relation to the dielectric response of the medium.

The Drude model is the simplest case that deals with free-electron metals and the dielectric function of the material. This model considers electrons in the solid as free particles interacting with the medium via a simple damping term τ describing the relaxation time due to friction in the electron gas. The dielectric function then becomes

$$\epsilon(\omega) = 1 - \frac{n_a e^2}{m\epsilon_0} \cdot \frac{1}{\omega^2 + (i\omega/\tau)} \quad (87)$$

where we define the term

$$\omega_p = \sqrt{\frac{n_a e^2}{m\epsilon_0}} \quad (88)$$

as the eigenfrequency of the electron plasma oscillation with

$$\epsilon_1 = \text{Re}\{\epsilon\} = 1 - \frac{\omega_p^2}{\omega^2 + (1/\tau^2)} \quad \text{and} \quad \epsilon_2 = \text{Im}\{\epsilon\} = \frac{1}{\omega\tau} - \frac{\omega_p^2}{\omega^2 + (1/\tau^2)} \quad (89)$$

τ is related to the FWHM of the plasmon peaks $\Delta E_p = \hbar/\tau$, m is the effective mass of the electrons, ϵ_0 is the vacuum dielectric constant, and n_a is the free electron density and the plasmon energy

$$E_p = \hbar\omega_p = \hbar\sqrt{\frac{n_a e^2}{m\epsilon_0}} \quad (90)$$

The free-electron model, although very applicable for many *sp* metals, is not realistic for all materials or even all metals. To treat the full range of transition metals, insulators, and semiconductors more realistically, there must be provision for transitions from various occupied to unoccupied bands in addition to contributions from the free electrons. This is done in the Drude–Lorentz model where discrete transitions are added by considering independent oscillators having eigenfrequencies equal to the transition energies (for example, to account for band-to-band transitions related to *d* electrons and interband transitions from the valence to conduction band). The expression contains the free oscillators (such as the Drude oscillators) and the “bound” oscillators. The dielectric function is therefore the sum over all these oscillators “*j*”:

$$\epsilon(\omega) = 1 + \frac{e^2}{m\epsilon_0} \sum_j \frac{n_j}{(\omega_j^2 - \omega^2 - i\omega/\tau)} \quad (91)$$

where n_j is the number of electrons able to oscillate at the eigenfrequency ω_j .

The presence of oscillators due to interband transitions or low-energy edges causes the peaks in the absorption part of the dielectric function and generates shifts in the position of the plasmon peaks. Similarly, the effect of the plasmon peaks is to shift the apparent energy of the interband transitions in the energy loss spectra with respect to the energy in the absorption part of the dielectric function. Such shifts therefore imply that caution should be taken when interpreting peaks in spectra: the position of a peak at a given energy loss E does not imply that there is an interband transition with same exact energy E .

8.2.2 Applications

The effects discussed in the previous section, even if not always fully quantified in analytical electron microscopy work, can be exploited in energy-filtered “plasmon” images to identify the presence of phases with different electron densities (hence plasmon position) or dielectric function. Systematic variations of the plasmon energy as a function of the atomic number have been demonstrated in the early work of Colliex (1984) (Figure 4–96). Plasmon energies in alloys and metallic hydrides as a function of alloying element concentration and hydrogen content have been tabulated from various sources in Egerton (1996). This technique of plasmon measurement has also been applied to study the ratio of *sp*²/*sp*³ hybridization bonding in C films and the water content in biological structures (Figure 4–81) (Sun et al., 1995).

Although a detailed analysis of the low-loss spectra is not yet a routine AEM techniques, detailed study of spectra can be extremely useful to understand some of the functional properties of materials. For examples, the relationship between the spectrum and the dielectric properties of solids can help elucidate some local variations of optical properties of materials (Turowski and Kelly, 1992; Schamm and Zanchi, 2003; Mullejans and French, 2000). Such studies make use of some key

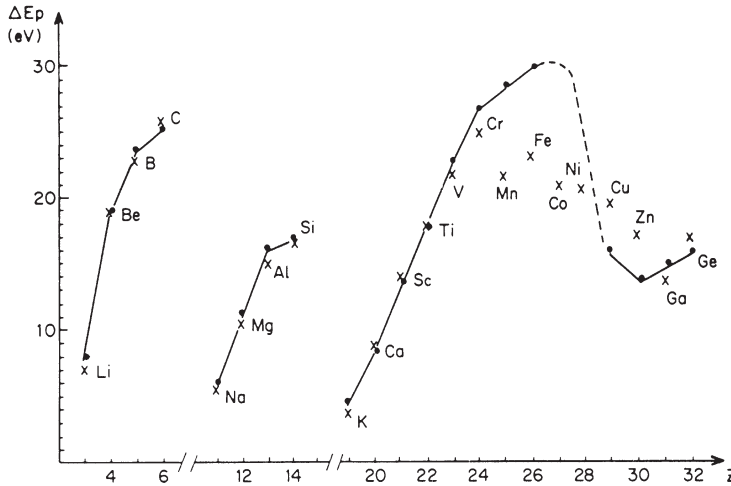


Figure 4-96. Variation of the plasmon energy for a series of pure elements. "x" symbols are experimental values and full circles are theoretical values calculated with the free electron model of Eq. (90). (Reprinted from Colliex © 1984, with permission from Springer Science+Business Media.)

properties of the dielectric function and, in particular, the Kramers–Kronig relations (Wooten, 1972) that link the real and imaginary parts of the dielectric function with the help of the Kramers–Kronig analysis (KKA): if the loss function $\text{Im}\{-1/\epsilon\}$ is known from the measurement of a spectrum,¹ then $\text{Re}\{1/\epsilon\}$ can be determined by KKA. Based on this relationship it is possible to retrieve both ϵ_1 and ϵ_2 and gain full knowledge of the dielectric function of a material (Figure 4-97). Once the

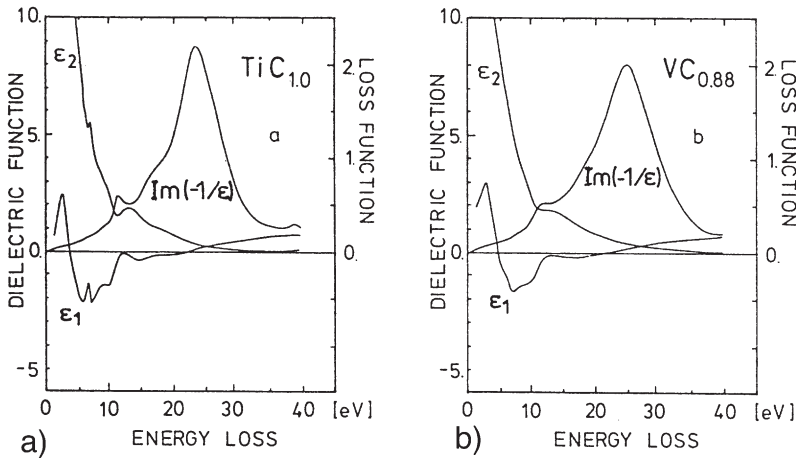


Figure 4-97. Real and imaginary part of the dielectric function based on Kramers–Kronig analysis of energy loss spectra for (a) TiN and (b) VC. (Reprinted from Schattschneider and Jouffrey © 1995, with permission from Springer Science+Business Media.)

¹ To retrieve the loss function from a spectrum, the single scattering distribution must be retrieved as demonstrated below.

dielectric function is known, the results can be compared to optical measurements at low energies and can be used to associate peaks to features in electronic transitions. One important advantage of the dielectric function measurement with EELS (as compared to optical techniques) is in the energy range of the measurements. With EELS, it is possible to analyze energies up to few hundred electronvolts covering, with the use of monochromator energies as low as the infrared region of the spectrum (down to about 0.5 eV), the visible and the UV up to the soft X-ray region (few 100 eV). Examples of KKA with EELS to study the dielectric function of material is given in Moreau and Cheynet (2003) and Launay et al. (2004). Systematic work on the modeling of low-loss spectra using first principles and a discussion of the limitation of this approach has been presented by Keast (2005).

Another very useful application of low-energy losses is related to the strong dependence of the spectra on the thickness of the samples (Figure 4–98). As the sample thickness increases, the probability of multiple inelastic losses increases. For example, the incident electron can excite one or more plasmon losses when traveling in the sample. If the mean distance between inelastic losses is λ , the probability of multiple loss events n is given by

$$P_n = \frac{1}{n!} \left(\frac{t}{\lambda} \right)^n \exp\left(-\frac{t}{\lambda}\right) \quad (92)$$

If we consider the intensity of the zero-loss peak I_0 (corresponding to zero plasmon loss events, i.e., $n = 0$) with respect to the total intensity in the spectrum I_t , we can deduce the probability P_0 for $n = 0$ and

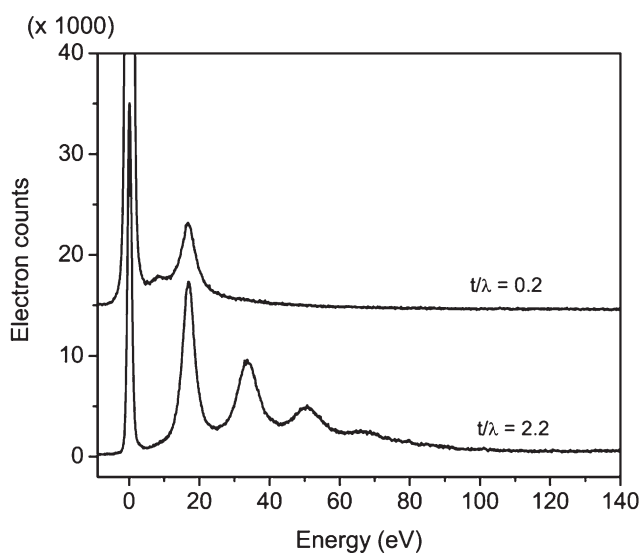


Figure 4–98. Example of the variation of the low-loss spectrum in Al for two different relative thicknesses. The increase in thickness (from $t/\lambda = 0.2$ to $t/\lambda = 2.2$) causes multiple plasmon losses visible at energy values of multiple single losses (i.e., n times E_p). For broad plasmon peaks, multiple peaks are not well resolved. (Courtesy of N. Braidy, McMaster University.)

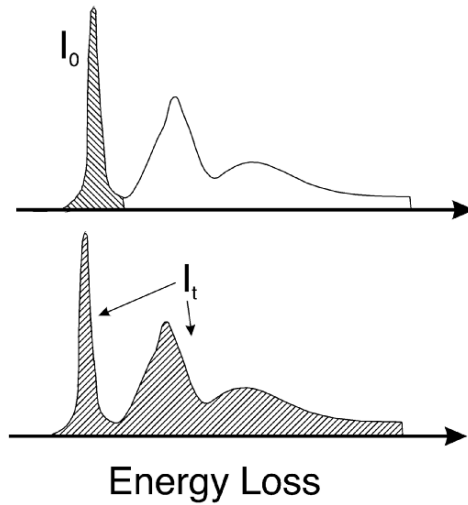


Figure 4–99. Features used for the determination of the relative thickness t/λ according to Eq. (94).

$$\frac{I_0}{I_t} = P_0 = \exp\left(-\frac{t}{\lambda}\right) \quad (93)$$

This expression is typically used as

$$\frac{t}{\lambda} = \ln\left(\frac{I_0}{I_t}\right) \quad (94)$$

with the mean free path λ values either tabulated in the literature (for a standard collection condition) or based on measurements with standards of known thickness or local measurements using complementary techniques (such as convergent beam or contamination spot measurements) (Williams and Carter, 1996; Kelly et al., 1975). Measurements of the intensities for I_t and I_0 are based on the integration of the spectra over windows as demonstrated in Figure 4–99. The measurements of thickness have been demonstrated to be reliable for thickness up to $t/\lambda = 4$ (Botton et al., 1995b). Values of the inelastic mean free path are dependent on the energy of the primary electrons, the material, and the angular collection conditions. Estimates of λ can be obtained based on the work of Malis et al. (1988).

The use of a finite collection angle β results in scattering outside the collection aperture and affects the thickness measurements. For multiple inelastic losses, the total angular distribution emerging from the sample is broader as compared to the single scattering distribution due to multiple scattering events and the convolution of the inelastic scattering distributions. A finite collection angle therefore results in loss in collection efficiency of electrons having lost higher energies by multiple events and leads to changes in the apparent mean free path. Approaches to convert mean free paths from one aperture collection condition to another have been presented by Botton et al. (1995b), while iterative methods and computer programs have been developed by

Egerton (1996). Mean free paths for 100 keV electrons and $\beta = 10$ mrad are in the range of 115 nm for C, 100 nm for Al, 110 for Si, and 75 nm for Fe and have been tabulated in Egerton (1996) and Malis et al. (1988). The mean free path can be estimated with some basic assumptions on the sample composition if $\beta \ll (E/E_0)$ where E is the energy loss and E_0 is the incident energy (so as to satisfy the dipole condition). If these conditions are respected, values of mean free paths can be calculated using (Malis et al., 1988)

$$\lambda(\beta) \approx \frac{106F(E_0/E_m)}{\ln(2\beta E_0/E_m)} \quad (95)$$

with E_0 values given in kiloelectronvolts, λ in nanometers, and β in millirad. F is a relativistic factor that can be calculated as

$$F = \frac{1 + E_0/1022}{(1 + E_0/511)^2} \quad (96)$$

and E_m is the mean energy loss in electronvolts. E_m can be approximated as $E_m \approx 7.5 Z^{0.36}$ without taking into account any contributions from changes in density or the crystallographic structure. For compounds containing elements of known atomic fraction f_i , a mean effective atomic number can be calculated by summing over each element i as

$$Z_{\text{eff}} = \frac{\sum_i f_i Z_i^{1.3}}{\sum_i f_i Z_i^{0.3}} \quad (97)$$

so that the effective atomic number Z_{eff} can be used to determine the mean energy loss E_m . Using the Kramers–Kronig analysis of the scattering distribution, the absolute sample thickness can also be calculated (Yang and Egerton, 1995) without prior knowledge of the sample.

From the recorded spectra containing multiple inelastic losses (e.g., Figure 4–98), the single scattering distribution can be retrieved using deconvolution methods. These techniques are founded on the principle that the recorded spectrum is based on the sum of the single scattering distribution, the double scattering distribution, . . . , the i scattering distribution, up to the n th scattering distribution. Each i distribution is obtained by the convolution of the single scattering distribution with itself i times. This sum of convolutions can be analyzed with Fourier techniques to retrieve, from the experimental spectrum containing multiple inelastic losses, the original single scattering distribution. These methods are available in commercial EELS analysis packages as standard functions. In this context, the Fourier-log technique is used to process the full spectrum (Egerton, 1996) from zero electronvolts to few hundred electronvolts. The analysis of the spectrum to retrieve the distribution representative of single scattering is necessary for all quantitative work related to the Kramers–Kronig analysis and the measurement of absolute thickness discussed in Egerton (1996) and Egerton and Cheng (1987). The additional application of the deconvolution method is to remove multiple scattering effects from core loss spectra.

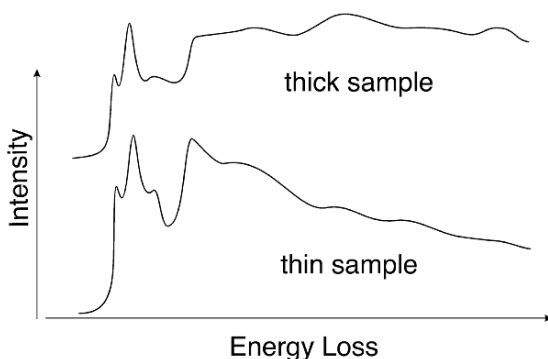


Figure 4-100. Effect of multiple inelastic losses on the shape of the ELNES.

The probability of multiple scattering including one core loss and plasmon losses increases in thick samples and the effect is visible as an increased intensity further away from the edge threshold: the spectra appear as a convolution of the low-loss spectra with the expected edge profile (Figure 4-100). These multiple scattering effects must be removed by deconvolution methods (in this context using the Fourier ratio method) to correctly interpret modulations in the spectra at high energy from the threshold. Structures within a few electronvolts from the edge threshold, however, are less sensitive to these multiple scattering effects.

Acknowledgments. I am grateful for the patience and understanding of Profs. Peter Hawkes and John Spence who allowed me to complete the work with on-going academic commitments, conference organization, and setting up of a new facility. I am indebted to members of my group for providing some figures and for feedback, in particular, N. Braidy, G. Radtke, M. Couillard, C. Maunders, and Y. Zhu. I want to thank several collaborators and friends who have provided, over the years, interesting samples, motivating discussions, and moral support.

References

- Ankudinov, A.L., Ravel, B., Rehr, J.J. and Conradson, S.D. (1998). *Phys. Rev. B* **58**, 7565.
- Arslan, I., Bleloch, A., Stach, E.A. and Browning, N.D. (2005). *Phys. Rev. Lett* **94**, 025504.
- Barfels, M.M.G., Burgner, P., Edwards, R. and Brink, H.A. (2002). In *Microscopy and Microanalysis 2002* (E. Voelkl, D. Piston, R. Gauvin, A.J. Lockley, G.W. Bailey and S. McKernan, Eds.), p. 614CD (Cambridge University Press, Cambridge).
- Batson, P.E. (1993). *Nature* **366**, 727.
- Batson, P.E., Dellby, N. and Krivanek, O.L. (2002). *Nature* **418**, 617.
- Bayle-Guillemaud, P., Radtke, G. and Sennour, M. (2003). *J. Microsc.* **210**, 66.
- Berger, A. and Kohl, H. (1992). *Microsc. Microanal. Microstruct.* **3**, 159.
- Berger, A. and Kohl, H. (1993). *Optik* **92**, 175.
- Berger, A., Mayer, J. and Kohl, H. (1994). *Ultramicroscopy* **55**, 101.

- Blaha, P., Schwarz, K., Madsen, G., Kvasnicka, D. and Luitz, J. (2001). Wien2k, An augmented Plane Wave Plus Local Orbitals Program for Calculating Crystal Properties. Vienna University of Technology, Institute of Physical and Theoretical Chemistry.
- Bonnet, N., Brun, N. and Colliex, C. (1999). *Ultramicroscopy* **77**, 97.
- Botton, G.A. (1999). In *Impact of Electron and Scanning Probe Microscopy on Materials Research* (D.G. Rickerby, G. Valdré and U. Valdré, Eds.), p. 265 (Kluwer Academic Publishers, London).
- Botton, G.A. (2005). *J. El. Spect. Rel. Phen.* **143**(2–3), 129–137.
- Botton, G.A. and L'Espérance, G. (1994). *J. Microsc.* **173**, 9.
- Botton, G.A. and Phaneuf, M.W. (1999). *Micron* **30**, 109.
- Botton, G.A., Boothroyd, C.B. and Stobbs, W.M. (1995a). *Ultramicroscopy* **59**(1–4), 93–107.
- Botton, G.A., L'Espérance, G., Gallerneault, C.E. and Ball, M.D. (1995b). *J. Microsc.* **180**, 217.
- Botton, G.A., Guo, G.Y., Temmerman, W.M. and Humphreys, C.J. (1997). In *Properties of Complex Inorganic Solids* (A. Gonis, A. Meike and P. Turchi, Eds.), p. 175 (Plenum Press, New York).
- Brown, L.M. (1999). In *Impact of Electron and Scanning Probe Microscopy on Materials Research* (D.G. Rickerby, G. Valdré and U. Valdré, Eds.), p. 231 (Kluwer Academic Publishers, London).
- Carpenter, G.J.C. (2004). *Microsc. Microanal.* **10**, 435.
- Castaing, R. and Henry, L. (1962). *C.R. Acad. Sci. Paris* **B255**, 76.
- Chen, J.G. (1997). *Surface Sci. Rep.* **30**, 1.
- Cheng, S.C. and Egerton, R.F. (1993). *Micron* **24**, 251.
- Chevalier, E.J.A. and Botton, G.A. (1999). *Inst. Phys. Conf. Ser.* **161 EMAG 99**, p. 175.
- Cliff, G. and Lorimer, G.W. (1975). *J. Microsc.* **103**, 203.
- Cliff, G. and Lorimer, G.W. (1981). In *Quantitative Microanalysis with High Spatial Resolution* (G.W. Lorimer, M.H. Jacobs and P. Doigt, Eds.), p. 41 (The Metals Society, London).
- Colliex, C. (1984). In *Advances in Optical and Electron Microscopy* (V.E. Cosslett and R. Barer, Eds.), Vol. 9, p. 65 (Academic Press, London).
- Colliex, C. (1996). *J. Electron Microsc.* **45**, 44.
- Colliex, C. and Mory, C. (1984). In *Quantitative Electron Microscopy* (J.N. Chapman and A.J. Craven, Eds.), p. 149 (SUSSP Publications, Edinburgh).
- Colliex, C., Tencé, M., Mory, C., Gu, H., Bouchet, D. and Jeanguillaume, C. (1994). *Mikrochim. Acta* **114**, 71.
- Cosgriff, E.C., Oxley, M.P., Allen, L.J. and Pennycook, S.J. (2005). *Ultramicroscopy* **102**, 317.
- Cowley, J. (1986). In *Principles of Analytical Electron Microscopy* (D.C. Joy, A.D.J. Romig and J.I. Goldstein, Eds.) (Plenum, New York).
- deGroot, F.M.F. (1994). *J. El. Spect. Rel. Phen.* **64**, 529.
- deGroot, F.M.F. (2005). *Coord. Chem. Rev.* **249**, 31.
- Dellby, N., Krivanek, O.L., Nellist, P.D., Batson, P.E. and Lupini, A.R. (2001). *J. Elect. Microsc.* **50**, 177.
- Doigt, P. and Flewitt, P.E.J. (1982). *Met. Trans. A* **13**, 1397.
- Dowell, W.C.T. and Goodman, P. (1973). *Phil. Mag.* **28**, 471.
- Dwyer, C. (2005). *Ultramicroscopy* **104**, 141.
- Dwyer, C. and Etheridge, J. (2003). *Ultramicroscopy* **96**, 343.
- Egerton, R.F. (1993). *Ultramicroscopy* **50**, 13.
- Egerton, R.F. (1996). *Electron Energy Loss Spectroscopy in the Electron Microscope*, 2nd ed. (Plenum, New York).
- Egerton, R.F. (1999). *J. Electron Microsc.* **48**, 711.

- Egerton, R.F. (2003). *Micron* **34**, 127.
- Egerton, R.F. (2005). *Physical Principles of Electron Microscopy* (Plenum, New York).
- Egerton, R.F. and Cheng, S.C. (1987). *Ultramicroscopy* **21**, 231.
- Egerton, R.F. and Cheng, S.C. (1994). *Ultramicroscopy* **55**, 43.
- Egerton, R.F. and Leapman, R.D. (1995). In *Quantitative Electron Energy Loss Spectroscopy, Energy Filtering Transmission Electron Microscopy* (L. Reimer, Ed.), p. 269 (Springer, Heidelberg).
- Eibl, O. (1993). *Ultramicroscopy* **50**, 179.
- Essers, E. and Benners, G. (2006). *Ultramicroscopy* (submitted).
- Faulkner, R.G. and Norrgard, K. (1978). *X-ray Spectrom.* **7**, 184.
- Fink, J. (1989). *Adv. Electr. Electron Phys.* **75**, 121.
- Fink, J. (1992). In *Unoccupied Electronic States* (J.C. Fuggle and J.E. Inglesfield, Eds.), p. 139 (Springer-Verlag, Berlin).
- Fransen, M.J., van Rooy, T.L. and Kruit, P. (2005). *Appl. Surf. Sci.* **146**, 312.
- Freitag, B. and Mader, W. (1999). *J. Microsc.* **194**, 42.
- Garvie, L.A.J., Craven, A.J. and Brydson, R. (1994). *Am. Mineral.* **79**, 411.
- Glitz, R.W., Notis, M.R. and Williams, D.B. (1981). *Microbeam Analysis—1981* (R.H. Geiss, Ed.), p. 309 (San Francisco Press, San Francisco).
- Goldstein, J.I., Costley, J.L., Lorimer, G.W. and Reed, S.J.B. (1977). In *SEM/77* (O. Johari, Ed.), p. 315 (IITRI, Chicago).
- Goldstein, J.I., Newbury, D.E., Echlin, P., Joy, D.C., Fiori, C.E. and Lifshin, E. (1981). *Scanning Electron Microscopy and X-ray Microanalysis: A Text for Biologists, Materials Scientists and Geologists* (Plenum, New York).
- Goldstein, J.I., Williams, D.B. and Cliff, G. (1986). In *Quantitative X-ray Analysis, Principles of Analytical Electron Microscopy* (D.C. Joy, A.D.J. Romig and J.I. Goldstein, Eds.), p. 155 (Plenum, New York).
- Goldstein, J.I., Newbury, D.E., Echlin, P., Joy, D.C., Lyman, C.E., Lifshin, E., Sawyer, L. and Michael, J.R. (2003). *Scanning Electron Microscopy and X-ray Microanalysis* (Kluwer Academic, New York).
- Grogger, W., Schaffer, B., Krishnan, K.M. and Hofer, F. (2003). *Ultramicroscopy* **96**, 481.
- Grogger, W., Varela, M., Ristau, R., Schaffer, B., Hofer, F. and Krishnan, K.M. (2005). *J. El. Spect. Rel. Phen.* **143**, 139.
- Gross, U., Mescher, F.J.M. and Tiemeijer, J.C. (1987). *Philips Tech. Rev.* **43**, 273.
- *Gu, L., Srot, V., Sigle, W., Koch, C., van Aken, P., Scholz, F., Thapa, S.B., Kirchner, C., Jetter, M. and Rühle, M. (2007). Band-gap measurements of direct and indirect semiconductors using monochromated electrons. *Physical Review B* **75**: Art. No. 195214.
- Hébert, C. (2006). *Micron* (in press).
- Hitchcock, A.P., Morin, C., Zhang, X.R., Araki, T., Dynes, J., Stover, H., Brash, J., Lawrence, J.R. and Leppard, G.R. (2005). *J. El. Spect. Rel. Phen.* **144**, 259.
- Hofer, F. (1995). In *Inner Shell Ionization, Energy-Filtering Transmission Electron Microscopy* (L. Reimer, Ed.), p. 225 (Springer, Heidelberg).
- Hofer, F. and Warbichler, P. (1996). *Ultramicrosc.* **63**, 21.
- Hofer, F. and Warbichler, P. (2004). In *Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS Atlas* (C.C. Ahn, Ed.), p. 195 (Wiley-VCH, New York).
- Hofer, F., Grogger, W., Kothleitner, G. and Warbichler, P. (1997). *Ultramicroscopy* **67**, 83.
- Hofer, F., Grogger, W., Warbichler, P. and Papst, I. (2000). *Mikrochim. Acta* **132**, 273.
- Hofer, F., Schaffer, B., Grogger, W. and Kothleithner, G. (2005). In *Microscopy and Microanalysis 2005* (R. Price, P. Kotula, M. Marko, J.H. Scott, G.F. Van der Voort, M. Mah Lee Ng, K. Smith, B. Griffin, P. Smith and S. McKernan, Eds.), Vol. 11, Suppl. 2, p. 48 (Cambridge University Press, Honolulu, HI).
- Horita, Z., Sano, T. and Nemoto, M. (1987). *Ultramicroscopy* **21**, 271.

- Hovington, P., Drouin, D. and Gauvin, R. (1997). *Scanning* **19**, 1.
- Humphreys, C.J. (1979). In *Principles of STEM, Introduction to Analytical Electron Microscopy* (D.C. Joy, A.D.J. Romig and J.I. Goldstein, Eds.) (Plenum, New York).
- Hunt, J.A. and Williams, D.B. (1991). *Ultramicroscopy* **38**, 47.
- Hunt, J.A., Disko, M.M., Behal, S.K. and Leapman, R.D. (1995). *Ultramicroscopy* **58**, 55.
- Inokuti, M. (1971). *Rev. Mod. Phys.* **43**, 297.
- Inokuti, M. (1978). *Rev. Mod. Phys.* **50**, 23.
- Jeanguillaume, C. and Colliex, C. (1989). *Ultramicroscopy* **28**, 252.
- Jouffrey, B., Schattschneider, P. and Hebert, C. (2004). *Ultramicroscopy* **102**, 61.
- Joy, D.C., Romig, A.D.J. and Goldstein, J.I. (1986). *Principles of Analytical Electron Microscopy* (Plenum, New York).
- Keast, V.J. (2005). *J. El. Spect. Rel. Phen.* **143**, 97.
- Keast, V.J. and Williams, D.B. (1999). *Acta Mater.* **47**, 3999.
- Keast, V.J. and Williams, D.B. (2000). *J. Microsc.* **199**, 45.
- Keast, V.J., Bruley, J., Rez, P., MacLaren, J.M. and Williams, D.B. (1998). *Acta Mater.* **46**, 481.
- Keast, V.J., Scott, A.J., Brydson, R., Williams, D.B. and Bruley, J. (2001). *J. Microsc.* **203**, 135.
- Kelly, P.M., Jostson, A., Blake, R.G. and Napier, J.G. (1975). *Phys. Stat. Sol. (a)* **31**.
- Kohl, H. and Rose, H. (1985). *Adv. Electron. Electron Phys.* **65**, 175.
- Kothleitner, G. and Hofer, F. (1998). *Micron* **29**, 349.
- Kothleitner, G. and Hofer, F. (2003). *Ultramicroscopy* **96**, 491.
- Kotula, P., Michael, J.R. and Kennan, M.R. (2001). *Microsc. Microanal.* **7**, 198.
- Kotula, P., Kennan, M.R. and Michael, J.R. (2003). *Microsc. Microanal.* **9**, 1.
- Kramers, M.A. (1923). *Phil. Mag.* **46**, 836.
- Krivanek, O.L., Bacon, N.J., Corbin, G.J., Dellby, N., Hincirik, P., Krivanek, O.L., McNamara-Smith, A., Murfitt, M.F. and Szilagy, Z.S. (2005). *Microsc. Microanal.* **11** (Supplement 2), 1422.
- Krivanek, O.L., Ahn, C.C. and Keeney, R.B. (1987). *Ultramicroscopy* **22**, 103.
- Krivanek, O.L., Gubbens, A.J. and Dellby, N. (1991a). *Microsc. Microanal. Microstruct.* **2**, 315–332.
- Krivanek, O.L., Mory, C., Tencé, M. and Colliex, C. (1991b). *Microsc. Microanal. Microstruct.* **2**, 257.
- Krivanek, O.L., Bui, D.N., Ray, D.A., Boothroyd, C.B. and Humphreys, C.J. (1994). In *13th International Electron Microscopy Congress*, p. 167 (Les Editions de Physique, Les Ulis, Paris).
- Krivanek, O.L., Friedman, S.L., Gubbens, A.J. and Kraus, B. (1995a). *Ultramicroscopy* **59**, 267.
- Krivanek, O.L., Kundmann, M.K. and Kimoto, K. (1995b). *J. Microsc.* **180**, 277.
- Krivanek, O.L., Nellist, P.D., Dellby, N., Murfitt, M.F. and Z., S. (2003). *Ultramicroscopy* **96**, 229.
- L'Espérance, G., Botton, G.A. and Caron, M. (1990). *Proceedings of the Microbeam Analysis Society 1990* (J.R. Michael and P. Ingram, Eds.), p. 284. (San Francisco Press, San Francisco).
- Launay, M., Boucher, F. and Moreau, P. (2004). *Phys. Rev. B* **69**, 03101.
- Lavergne, J.-L., Martin, J.-M. and Belin, M. (1992). *Microsc. Microanal. Microstruct.* **3**, 517.
- Lazar, S., Botton, G.A., Wu, M.-Y., Tichelaar, F.D. and Zandbergen, H.W. (2003). *Ultramicroscopy* **96**, 535.
- Leapman, R.D. (2003). *J. Microsc.* **210**, 5.
- Leapman, R.D. and Newbury, D.E. (1993). *Anal. Chem.* **65**, 2409.
- Leapman, R.D. and Rizzo, N.W. (1999). *Ultramicroscopy* **78**, 251.

- Leapman, R.D., Rez, P. and Mayers, D.F. (1980). *J. Chem. Phys.* **72**, 1232.
- Loan, R.F., Kirkland, E.J. and Silcox, J. (1988). *Acta Crystallogr. A* **44**.
- Luo, B.P. and Zeitler, E. (1991). *J. El. Spect. Rel. Phen.* **57**, 285.
- Malis, T. and Titchmarsh, J.M. (1985). In *Electron Microscopy and Analysis 1985*, p. 181 (Institute of Physics Conference Series, Institute of Physics, Bristol).
- Malis, T., Cheng, S.C. and Egerton, R.F. (1988). *J. Electron Microsc. Tech.* **8**, 193.
- Mansot, J.L., Leone, P., Euzen, P. and Pavaldeau, P. (1994). *Microsc. Microanal. Microstruct.* **5**, 79.
- Mayer, J., Deininger, C. and Reimer, L. (1995). In *Energy Filtering Transmission Electron Microscopy* (L. Reimer, Ed.) (Springer, Heidelberg).
- Mayer, J., Eigenthaler, U., Plitzko, J.M. and Dettenwanger, F. (1997). *Micron* **28**, 361.
- Menon, N. and Krivanek, O.L. (2002). *Microsc. Microanal.* **8**, 203.
- Mitterbauer, C., Kothleithner, G., Hofer, F. and Zandbergen, H. (2003). *Ultramicroscopy* **96**, 469.
- Mobus, G. and Nufer, S. (2003). *Ultramicroscopy* **96**, 285.
- Moreau, P. and Cheynet, M.C. (2003). *Ultramicroscopy* **94**, 293.
- Mory, C., Colliex, C. and Cowley, J. (1987). *Ultramicroscopy* **21**, 171.
- Mullejans, H. and French, R.H. (2000). *Microsc. Microanal.* **6**, 297.
- Muller, D.A. and Silcox, J. (1995). *Ultramicroscopy* **59**, 195.
- Muller, D.A., Tzou, Y., Raj, R. and Silcox, J. (1993). *Nature* **366**, 725.
- Muller, D.A., Sorch, T., Moccio, S., Baumann, F.H., Evans-Luttertoldt, K. and Timp, G. (1999). *Nature* **399**, 758.
- Muller, D.A., Nakagawa, N., Ohtomo, A., Grazul, J.L. and Hwang, H.Y. (2004). *Nature* **430**, 657.
- Nakata, K., Okada, O. and Ueki, Y. (2001). *J. Electron Microsc.* **50**, 89.
- Natusch, M.K.H., Humphreys, C.J., Menon, N. and Krivanek, O.L. (1999). *Micron* **30**, 173.
- Nellist, P. and Pennycook, S.J. (1998). *Phys. Rev. Lett.* **81**, 4156.
- Newbury, D.E. (1998). *J. Electron Microsc.* **47**, 407.
- Newbury, D.E., Wollman, D.A., Hilton, G.C., Irwin, K.D., Bergren, N.F., Rudman, D.A., and Martinis, J.M. (2000). *J. Radioanal. Nucl. Chem.* **244**, 627.
- Nicholson, W.A.P., Gray, C.C., Chapman, J.N. and Robertson, B.W. (1982). *J. Microsc.* **125**, 25.
- Nocholds, C., Nasir, M.J., Cliff, G. and Liorimer, G.W. (1979). In *Electron Microscopy and Analysis* (T. Mulvey, Ed.), p. 417 (Institute of Physics Conference Series No. 52, Institute of Physics, London).
- Oikawa, T. (2006). Private communication.
- Otten, M.T. (1996). *The CM Series Transmission Electron Microscopes in Materials Research*. (Electron Optics, Eindhoven, The Netherlands).
- Ottensmeyer, F.P. (2004). Private communication.
- Oxley, M.P., Cosgriff, E.C. and Allen, L.J. (2005). *Phys. Rev. Lett.* **94**, 203906.
- Paxton, A.T. (2005). *J. El. Spect. Rel. Phen.* **143**, 51.
- Pennycook, S.J. (1982). *Contemp. Phys.* **23**, 371.
- Pennycook, S.J., Jesson, D.E., McGibbon, A.J. and Nellist, P.D. (1996). *J. Electron Microsc.* **45**, 36.
- Philibert, J. and Tixier, R. (1975). In *Physical Aspects of Electron Microscopy and Microbeam Analysis* (D.R. Beaman and B.M. Siegel, Eds.), p. 333 (Wiley, New York).
- Pickard, C.J. and Payne, M.C. (1997). In *Institute of Physics Conference Series 153 EMAG 97*, p. 179 (Institute of Physics, London).
- Potapov, P.L., Jorissen, K., Schryvers, D. and Lamoen, D. (2004). In *European Microscopy Congress* (D. Schryvers and P. Van Oostveldt, Eds.), p. 247 (Belgian Society for Microscopy, Antwerp).
- Powell, C.J. (1976). *Rev. Mod. Phys.* **48**, 33.

- Radtke, G., Epicier, T., Bayle-Guillemaud, P. and Le Bosse, J.C. (2003). *J. Microsc.* **210**, 60.
- Raether, H. (1980). *Excitations of Plasmons and Interband Transitions by Electrons* (Springer-Verlag, New York).
- Reed, S.J.B. (1966). *Proceedings of the 5th International Congress on X-ray Optics and Microanalysis* (R. Castaing, R. Deschamps and J. Philibert, Eds.), p. 339 (Herman, Paris).
- Reese, G.M., Spence, J.C.H. and Yamamoto, Y. (1984). *Phil. Mag. A* **49**, 697.
- Reimer, L. (1984). *Transmission Electron Microscopy* (Springer-Verlag, Heidelberg).
- Reimer, L. (1995). In *Introduction, Energy Filtering Transmission Electron Microscopy* (L. Reimer, Ed.), p. 1 (Springer, Heidelberg).
- Rez, P. (1982). *Ultramicroscopy* **9**, 283.
- Rez, P. (1984). *Ultramicroscopy* **12**, 29–38.
- Rez, P., Bruley, J., Brohan, P., Payne, M. and Garvie, L.A.J. (1995). *Ultramicroscopy* **59**, 159.
- Rose, A. (1970). *Image Technol.* **12**, 1315.
- Rose, H. and Krah, D. (1995). In *Electron Optics of Imaging Filters, Energy Filtering Electron Microscopy* (L. Reimer, Ed.), p. 43 (Springer, Heidelberg).
- Saldin, D.K. (1987). *Phil. Mag. B* **25**, 515.
- Saunders, M. (2003). *Microsc. Microanal.* **9**, 411.
- Schaffer, B., Grogger, W. and Kothleithner, G. (2004). *Ultramicroscopy* **102**, 27.
- Schamm, S. and Zanchi, G. (2003). *Ultramicroscopy* **96**, 559.
- Schattschneider, P. and Exner, A. (1995). *Ultramicroscopy* **59**, 241.
- Schattschneider, P. and Jouffrey, B. (1995). In *Energy Filtering Transmission Electron Microscopy* (L. Reimer, Ed.), p. 151 (Springer, Heidelberg).
- Schattschneider, P. and Werner, W.S.M. (2005). *J. El. Spect. Rel. Phen.* **143**, 81.
- Shuman, H. and Somlyo, A.P. (1987). *Ultramicroscopy* **21**, 23.
- *Sigle, W. (2005). Analytical transmission electron microscopy. *Annual Review of Materials Research* **35**, 239–314.
- *Sigle, W., Kramer, S., Varshney, V., Zern, A., Eigenthaler, U. and Rühle, M. (2003). Plasmon energy mapping in energy-filtering electron microscopy. *Ultramicroscopy* **96**, 565.
- Solymar, L. and Walsh, D. (2004). *Lectures on the Electronic Properties of Materials*, 6th ed. (Oxford University Press, Oxford).
- Spence, J.C.H. (2006). *Rep. Prog. Phys.* **69**, 725.
- Spence, J.C.H. and Lynch, J. (1982). *Ultramicroscopy* **9**, 267.
- Spence, J.C.H. and Zuo, J.M. (1992). *Electron Microdiffraction* (Plenum, New York).
- Spence, J.C.H., Reese, G.M., Yamamoto, Y. and Kurizki, G. (1983). *Phil. Mag. B* **48**, L39.
- Stenton, N., Notis, M.R., Goldstein, J.I. and Williams, D.B. (1981). In *Quantitative Analysis with High Spatial Resolution* (G.W. Lorimer, M.H. Jacobs and P. Doigt, Eds.), Book 277, p. 35 (The Metals Society, London).
- *Stöger-Pollach, M., Franco, H., Schattschneider, P., Lazar, B., Schaffer, B., Grogger, W. and Zandbergen, H.W. (2006). Cerenkov losses: a limit for bandgap determination and Kramers–Kronig analysis. *Micron* **37**, 396–402.
- Su, D.S., Wang, H.F. and Zeitler, E. (1995). *Ultramicroscopy* **59**, 181.
- Suenaga, K., Tencé, M., Mory, C., Colliex, C., Okazaki, T., Shinohara, H., Hirahara, K., Bandow, S. and Iijima, S. (2000). *Science* **290**, 2280.
- Sun, S.Q., Shi, S.L., Hunt, J.A. and Leapman, R.D. (1995). *J. Microscopy* **177**, 18.
- Tanaka, M., Terauchi, M., Tsuda, K., Saitoh, K., Mukai, M., Kaneyama, T., Tomita, T., Tsuno, K., Kersker, M., Naruse, M. and Honda, T. (2002). In *Microscopy and Microanalysis 2002* (E. Voelkl, D. Piston, R. Gauvin,

- A.J. Lockley, G.W. Bailey and S. McKernan, Eds.), p. 68 (Cambridge University Press, Cambridge).
- Tencé, M. (2002). Private communication.
- Terauchi, M., Tanaka, M., Tsuno, K. and Ishida, M. (1999). *J. Microsc.* **194**, 203.
- Thomas, P.J. and Midgley, P.A. (2001a). *Ultramicroscopy* **88**, 187.
- Thomas, P.J. and Midgley, P.A. (2001b). *Ultramicroscopy* **88**, 179.
- Tiemeijer, P.C., van Lin, J.H.A. and de Jong, A.F. (2001). *Microsc. Microanal.* **7**, 1130.
- Titchmarsh, J.M. and Dumbill, S. (1996). *J. Microsc.* **184**, 195.
- Trebbia, P. (1988). *Ultramicroscopy* **24**, 399.
- Turowski, M.A. and Kelly, T.F. (1992). *Ultramicroscopy* **41**, 41.
- Uhlemann, S. and Haider, M. (2002). In *Microscopy and Microanalysis 2002* (E. Voelkl, D. Piston, R. Gauvin, A.J. Lockley, G.W. Bailey and S. McKernan, Eds.), p. 584CD (Cambridge University Press, Cambridge).
- Van Cappellen, E. (1990). *Microsc. Microanal. Microstruct.* **1**, 1.
- Van Cappellen, E. and Doukhan, J.C. (1994). *Ultramicroscopy* **53**, 343.
- Varela, M., Findlay, S.D., Lupini, A.R., Christen, H.M., Borisevich, A.Y., Dellby, N., Krivanek, O.L., Nellist, P.D., Oxley, M.P., Allen, L.J. and Pennycook, S.J. (2004). *Phys. Rev. Lett.* **92**, 095502.
- Verbeeck, J. and Van Aert, S. (2004). *Ultramicroscopy* **101**, 207.
- Verbeeck, J., Van Dyck, D. and Van Tandeloo, G. (2004). *Spectrochim. Acta B* **59**, 1529.
- von Barth, U. and Grossman, G. (1982). *Phys. Rev. B* **25**(8), 5, 50.
- Voyles, P. and Muller, D.A. (2004). *Microsc. Microanal.* **10**, 291.
- Vvedensky, D.D. (1992). In *Theory of X-Ray Absorption Fine Structure, Unoccupied Electronic States* (J.C. Fuggle and J.E. Inglesfield, Eds.), p. 138 (Springer-Verlag, Berlin).
- Watanabe, M. and Williams, D.B. (1999). *Microsc. Microanal.* **5**, 88.
- Watanabe, M. and Williams, D.B. (2005a). *Microscopy and Microanalysis 2005, Vol. 11, Suppl. 2*, edited by R. Price, P. Kotula, M. Marko, J.H. Scott, G.F. Van der Voort, M. Mah Lee Ng, K. Smith, B. Griffin, P. Smith and S. McKernan, p. 1362. Honolulu, HI: Cambridge University Press.
- Watanabe, M. and Williams, D.B. (2005b). *Scanning* **27**, 94.
- Watanabe, M., Horita, Z. and Nemoto, M. (1996). *Ultramicroscopy* **65**, 187–198.
- Watanabe, M., Ackland, D.W., Burrows, A., Kiely, C.J., Williams, D.B., Kanno, M. and Hynes, R. (2005). In *Microscopy and Microanalysis 2005* (R. Price, P. Kotula, M. Marko, J.H. Scott, G.F. Van der Voort, M. Mah Lee Ng, K. Smith, B. Griffin, P. Smith and S. McKernan, Eds.), Vol. 11, Suppl. 2, p. 2132 (Cambridge University Press, Honolulu, HI).
- Westwood, A.D., Michael, J.R. and Notis, M.R. (1992). *J. Microsc.* **167**, 287.
- Williams, D.B. and Carter, C.B. (1996). *Transmission Electron Microscopy, A Textbook for Materials Science* (Plenum, New York).
- Williams, D.B., Watanabe, M., Papworth, A.J. and Li, J.C. (2003). *Thin Solid Films* **424**, 50.
- Woldseth, R. (1973). *X-Ray Energy Spectrometry* (Kevex Instruments).
- Wong, K. and Egerton, R.F. (1995). *J. Microsc.* **178**, 198.
- Wooten, F. (1972). *Optical properties of solids*. (Academic Press, New York).
- Yang, Y.Y. and Egerton, R.F. (1995). *Micron* **26**, 1.
- *Yurtserver, A., Weyland, M. and Muller, D.A. (2006). Three-dimensional imaging of nonspherical silicon nanoparticles embedded in silicon oxide by plasmon tomography. *Applied Physics Letters* **89**, 151920.
- Zaluzec, N.J. (1982). *Ultramicroscopy* **9**, 319.
- Zaluzec, N.J. (1984). *J. Phys.* **45** (Suppl. 2), C2-429.
- Zaluzec, N. (1985). *Ultramicroscopy* **18**, 185.
- Ziebold, T.O. (1967). *Anal. Chem.* **39**, 858.

5

High-Speed Electron Microscopy

Wayne E. King, Michael R. Armstrong, Oleg Bostanjoglo, and
Bryan W. Reed

1 What Is High-Speed Electron Microscopy?

Historically, transmission electron microscope (TEM) images have been recorded on film using exposure times of a few seconds. The use of the TEM, and especially the high-voltage electron microscope, for *in situ* experiments stimulated interfacing of video cameras with the microscope column. This enabled capture of frames with millisecond exposure times and interframe times of milliseconds and proved to be very useful for the study of dynamic phenomena with characteristic times that were consistent with the time resolution of the detectors, such as dislocation motion in deforming metals.

Many phenomena in chemistry, biology, and materials science proceed at rates that are significantly faster than could be captured with standard video methods. Figure 5–1 attempts to represent some of those phenomena in terms of spatial resolution and time resolution. Experimental techniques exist for probing most of this space, but the quadrant at high space and time resolutions has proved to be the most difficult to access.

The chemistry community, which studies phenomena with characteristic time scales of the order of the atomic vibrational period, has employed short pulse lasers and optical techniques to study ultrafast phenomena. Likewise, X-rays are becoming an important tool for ultrafast studies. Recently, ultrafast electron diffraction (UED) has been developed with time resolutions approaching that of optical and X-ray methods.^{1–9}

Although optical and X-ray methods are powerful, the widespread use of TEMs and the accompanying extensive literature demonstrate that there are many scientific investigations that benefit from high spatial resolution imaging. In materials science, these include observation of defects in crystalline materials (e.g., dislocations, grain boundaries, and heterophase boundaries) in which imaging affords a more direct interpretation of the defect structure than does diffraction.

One group has demonstrated the potential of high-speed TEM or dynamic (DTEM).^{10–16} This group has produced an imaging instrument

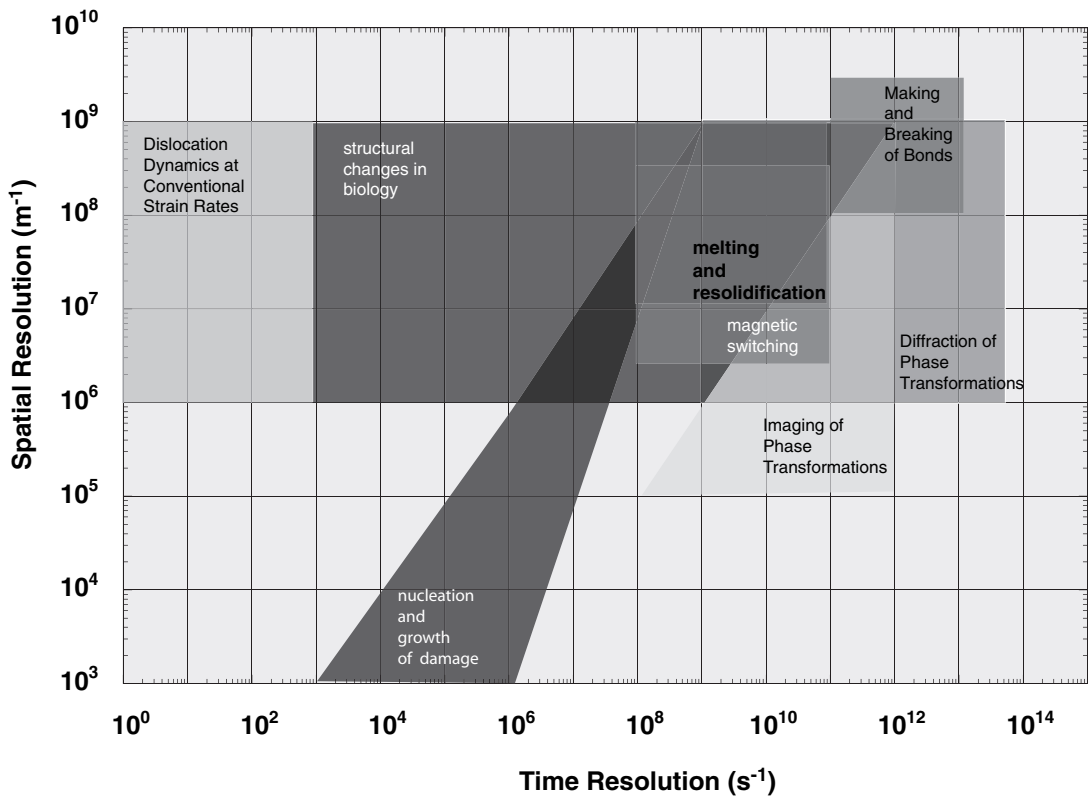


Figure 5–1. Phenomena classified by spatial and temporal resolution. Spatial resolution, is defined as follows: (1) if the technique is an imaging method, the low end of the bounding box would be defined by the smallest resolvable feature and the high end by the typical field of view or (2) for a nonimaging technique, the bounding box would be defined by the range of probe or spot sizes. Time resolution is defined as that for a single-shot investigation of irreversible processes. So time resolution is defined as the single-shot exposure time to obtain data that demonstrate a particular spatial resolution. (See color plate.)

with time resolution $\sim 3\text{--}10 \text{ ns}$ ^{10,11} and has demonstrated pump/probe-train “movie” or nanosecond multiframe operation.^{17,18} In this chapter we review the events leading to the development of the DTEM. We then describe the technologies employed in DTEM. Finally, we discuss the performance, limitations, and applications and conclude with a look to the future of these instruments.

2 Technologies of DTEM

2.1 Anatomy of a DTEM

Figure 5–2 shows a photo of the DTEM that was developed at TU Berlin. The instrument in Berlin is based on a Siemens Elmiskop 1A constructed in 1961. Several modifications, shown in Figure 5–3, were required to convert the instrument from a conventional continuous-wave (CW) mode to an instrument that could carry out pump-probe

experiments. These included the development of a photoelectron gun to replace the conventional thermionic tungsten filament. This was realized by interfacing a laser to the electron-optical column via a vacuum feed-through and directing that laser to the filament using a turning mirror. The photocathode is driven by a *Q*-switched frequency-quadrupled Nd:YAG laser (wavelength 266 nm). Using the photoelectric effect, a short laser pulse stimulates the emission of a low-energy electron bunch that is accelerated by the high-voltage stage of the electron microscope. In CW mode, the self-biasing Wehnelt brings the electrons emitted at the filament to a focus, thus defining the object plane for the condenser lens system. In the pulsed mode, circuit time constants inhibit the self-biasing effect of the Wehnelt, no crossover is formed, and the optics are a bit different than CW operation. The photocathode in the Berlin instrument has been designed so that it can operate in either CW or pulsed modes. This facilitates easy electron-optical alignment of the instrument. Since the electrons are emitted from slightly different regions of the cathode in thermionic operation compared with pulsed operation, it is essential to be able to align the gun in pulsed mode. This is achieved by pulsing the photocathode periodically at a rate of several pulses per second, observing the image on the digital image capture system (discussed later), and carrying out a gun alignment.

Accelerators for TEMs are designed to give smooth electric field gradients and stable potentials. The TU Berlin instrument typically operates at 80 kV with an accelerating gap of 1 cm. This corresponds to



Figure 5–2. The DTEM at TU Berlin. Cathode and sample drive lasers are at left. (See color plate.)

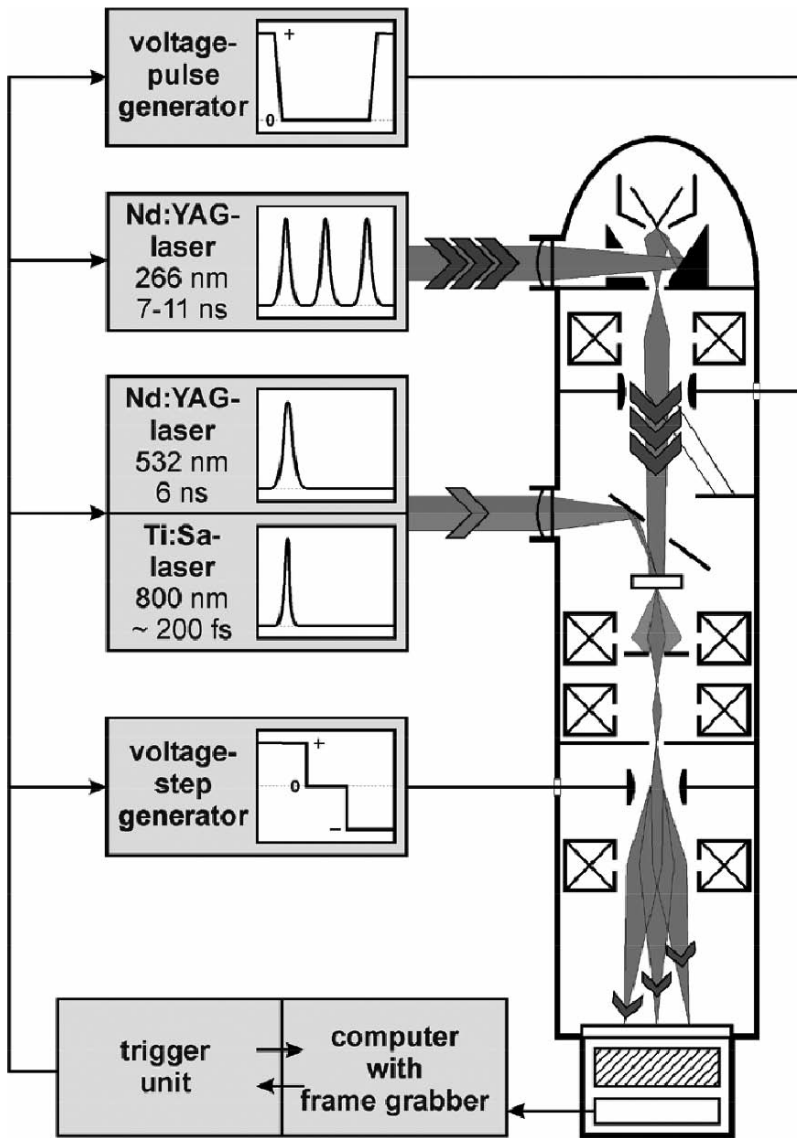


Figure 5–3. TU Berlin dynamic transmission electron microscope. (See color plate.)

a field gradient of $8 \times 10^4 \text{ V/cm}$. When electrons are photoemitted at low energy from the cathode, they are most vulnerable to space charge effects. It is therefore advantageous to optimize the extraction potential at the cathode and the field gradient to minimize space charge effects. This may be achieved by using an extraction electrode similar to those used in field emission electron guns. The steady-state current limit due to space charge in a parallel-plate accelerating gap is dictated by the Child–Langmuir law,¹⁹ and varies as $V^{3/2}/d^2$, with V the voltage and d the distance between the cathode and the extraction electrode. Thus it is most effective to have both a large electric field (possibly enhanced by surface curvature of the cathode) and a short gap.

Since the photoemitted electrons will be incoherent, it is not essential to have a very small source size. Therefore, it may be possible to spread the laser beam across a relatively large area to ensure that sufficient electrons can be generated without damaging the cathode with the laser. Of course, the size of the source will control the size of the smallest electron beam spot on the sample. Therefore, there will be a trade-off between the region of interest and source size. The source size should be adjusted to achieve the desired viewing area and no larger (to minimize space charge effects).

Apertures will play an important role in limiting space charge effects in DTEM. The condenser aperture should be removed to maximize the electron fluence on the sample. The objective aperture can be used to produce contrast by blocking scattered electrons as is routinely done in diffraction contrast imaging. This will decrease the number of electrons in the beam as the beam passes through the projector lens system. Alternatively, it is possible to block the main beam and allow the scattered electrons to pass.

In a double condensor lens system common to most electron microscopes, the first condensor will likely be operated at low or zero excitation to allow as many electrons as possible to pass to the second condensor lens. This lens can then be used in conjunction with the source size to optimize the illumination area on the sample.

Because of the energy spread in pulsed beams, chromatic aberration will limit the obtainable resolution. Thus constraints on the design of the objective lens can be relaxed until a need for high resolution arises. At that point, an aberration correction may be possible.

The projector lens has the potential to introduce distortion in the image if space charge and trajectory displacement effects are significant at the numerous crossovers in the projector lens system. For sufficiently short pulses, a lateral space charge effect could become significant at the crossovers.

The time resolution of the TU Berlin instrument is currently 3–10 ns. At this resolution, the electron pulse length is controlled primarily by the laser pulse length and by the number of electrons in an electron bunch required to obtain an image. By making a few simplifications and assumptions, the number of electrons needed to form an image can be estimated. The current state of the art in low-dose TEM applications in the biological sciences requires an average of about 100 electrons per pixel.²⁰ For a 1 k × 1 k pixel image, this requirement leads to 10^7 – 10^8 total electrons at the image plane (see Figure 5–4). This number of electrons per pulse is currently achieved by the nanosecond electron microscope at TU Berlin.¹¹ However, many electrons are lost in the TEM condenser lenses. Thus, the number of electrons that is required at the cathode may be substantially higher. Other factors that will influence the number of electrons arriving at the image plane include specimen thickness, diffracting conditions chosen for image formation, and the size and placement of the objective aperture.

The number of electrons needed to form a diffraction pattern can be substantially less than that required for an image. As few as 10^4 electrons may be needed for the qualitative assessment of a diffraction

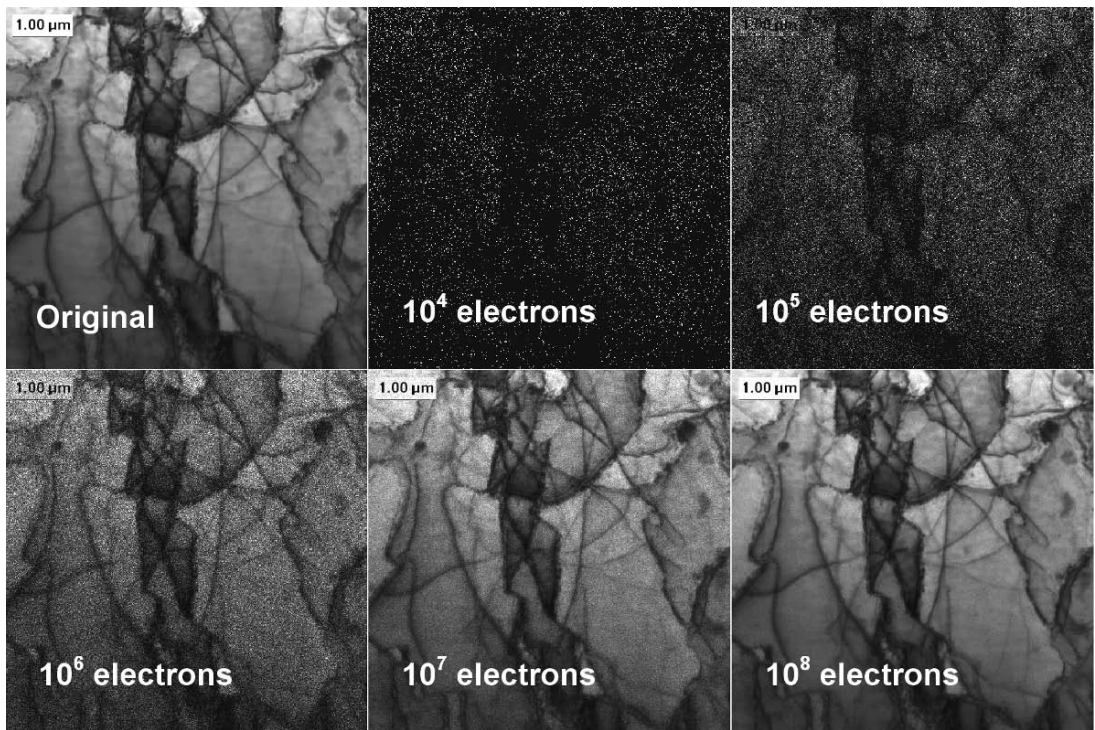


Figure 5-4. Simulation demonstrating the number of electrons required for a given image quality. The original 512×512 8-bit grayscale image was normalized and taken as the probability distribution for the detected position of each independent electron. Binning to a smaller number of pixels would reduce the required number of electrons, but at the expense of resolution. Various methods for estimating this (including the one shown here) consistently indicate that 10^6 – 10^8 electrons are needed at the detector, depending on the number of pixels and the required contrast. Crystalline diffraction patterns require ~10–100 times fewer electrons because of the concentration of intensity into relatively few pixels.

pattern.²⁰ However, for quantitative assessment, more (as many as 10^8) could be required for improved signal-to-noise ratio (SNR).

It is desirable to have the ability to induce physical changes in the sample that can be easily synchronized with the probe electron pulse. This has been accomplished at TU Berlin by interfacing a pump laser with the DTEM instrument. The pump beam is fed into the column at a position above the objective lens and directed to the sample by a turning mirror. Using this configuration, it is possible to bring an ~20-μm-diameter pump beam into coincidence with the probe electron beam. In the TU Berlin instrument, two different lasers are used for pump and probe. This is reasonable for systems with time resolutions on a nanosecond scale. At higher time resolutions, unacceptable jitter between the lasers will require the use of a single seed-laser with beam split into pump and probe beams.

Because of the limited number of electrons per pulse, it is essential to have a highly efficient detector, the so-called “every-electron camera.” Space-charge limits on the number of electrons in a short pulse put a premium on detection efficiency. However, because of the large spec-

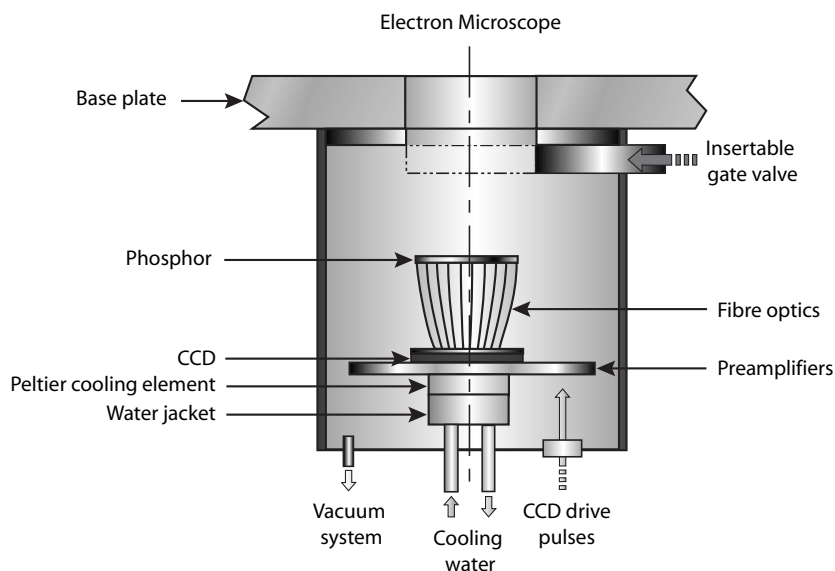


Figure 5–5. Directly coupled CCD camera system.

trum of desired electron energies (keV–MeV) and the disparate requirements of imaging and diffraction, it is clear that no single camera is ideal for every application. At high energies (>100 keV), phosphor scintillators are preferred (Figure 5–5). Phosphor thickness can be optimized to achieve an energy conversion efficiency of ~ 0.12 (with intensification not required).²¹ Light-scattering effects within the phosphor and the various optical interfaces limit system resolution.²² The point spread function (PSF) for a detector system composed of phosphor on a fiberoptic faceplate is determined by electron and photon scattering as well as the numerical aperture (NA) of the fiberoptic plate. The next step is to couple the phosphor to the charge-couple device (CCD). This is typically achieved using a fiberoptic bundle or lens coupling. Fiberoptic coupling is preferred for energies below 400 keV because of the high-photon-collection solid angle from the phosphor. Because the PSF of the phosphor is greater than the pixel size of typical CCDs, taper-fiber bundles are often employed to improve the PSF/pixel size match. However, tapered bundles can exhibit spatial distortions of $\sim 2\text{--}3\%$ that need to be corrected before quantitative measurements can be made from images or diffraction patterns.²¹ With the availability of large CCD chips, 1:1 coupling becomes practical. Also, binning of adjacent pixels is used to boost the detective quantum efficiency (DQE).

Finally, to achieve images that follow the evolution of dynamic processes, a multiframe movie mode is desirable. Interframe times less than a nanosecond are desirable to observe dynamic processes. Currently, no technology exists that will enable such rapid filling and readout of CCD cameras. At TU Berlin, movie mode is achieved through the use of an electrostatic beam shifter (Figure 5–3). An image-shifting capacitor with low inductance leads is used to produce successive frames by displacing an image patch across the detector between the

electron pulses. For multiframe imaging of the specimen, the image is confined by an adjustable slit aperture in the second-intermediate image plane, exploiting the space between the intermediate and projector lenses. For multiframe diffraction, the diffraction pattern in the back focal plane of the objective lens is confined with a suitable aperture, and this pattern is as usual transferred with the intermediate and projector lenses to the detector.¹⁰

The beam shifter is a low-capacitance rectangular plate capacitor mounted on a side-entry retractable holder. It was designed to fit into the space between the intermediate and projector lenses. Two blades glued onto the plate electrodes at the beam entrance serve as the image field-confining aperture. The plates of the capacitor are tilted to prevent the deflected electrons from hitting the attracting electrode. The capacitor can be axially moved and the distance between the plates varied, allowing alignment and field selection. The deflecting unit is inserted at such a location that the second intermediate image can be positioned near the field-confining slit aperture.

Three-frame imaging is achieved by applying a two-step voltage waveform across the capacitor (Figure 5-6). Because the deflecting field is short due to limited space and since the deflection occurs near an image plane, high voltages of several hundred volts are required to shift 100-kV electron images. The shifting steps are produced by shorting voltages with fast transistor-based switches (Figure 5-7). When both switches are open the image is on the left-hand side (Figure 5-6). As switch S1 closes, the image is moved to the center. Finally, switch S2 closes transferring a positive voltage step via the 6.8-nF capacitor to the right electrode, thereby displacing the image to the right side.

A high-voltage radio frequency *n*-channel metal oxide semiconductor field-effect transistor and cascaded bipolar transistors (Motorola 2N5551) driven in the avalanche mode were used as switches. The voltage fall times were 2–3 ns. Overshoot and ringing were minimized by careful grounding, shielding, and placing of leads avoiding stray capacitance, parasitic inductance, and closed loops. In addition, ferrite-damping beads were suitably distributed along the leads. The switches were mounted directly on the holder of the shifting capacitor to avoid voltage distortions caused by reflections and to minimize electromagnetic interference.

Electromagnetic interference easily was kept below a deleterious level using the MOSFET-based switch. But it is rather inconvenient to

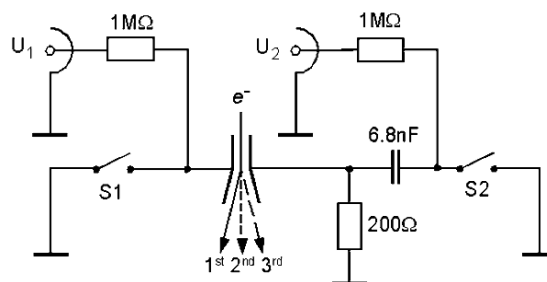


Figure 5-6. Generation of voltage steps across the image shifting capacitor for three-frames imaging.

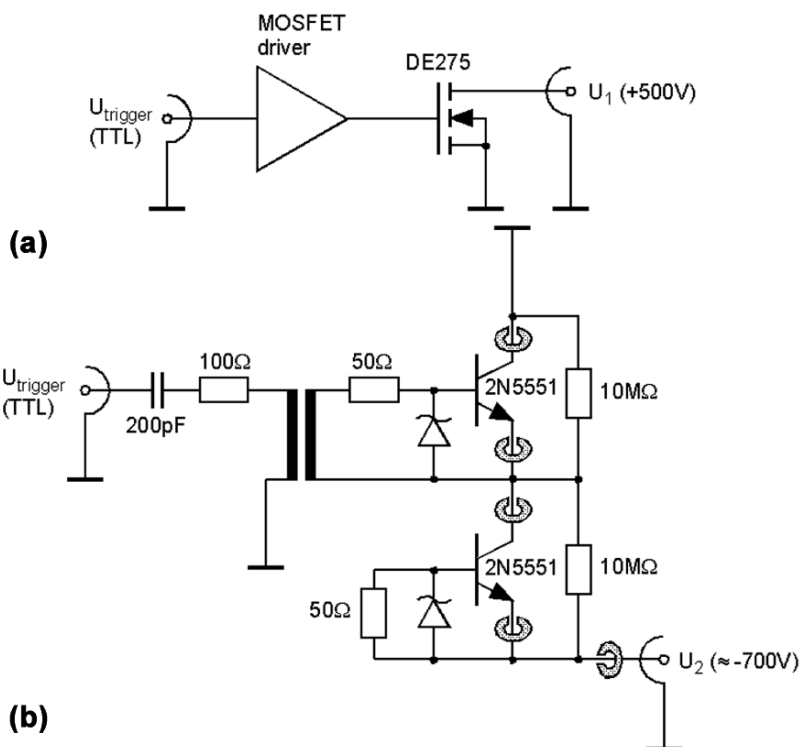


Figure 5-7. Fast high-voltage switches. (a) High-power radiofrequency *n*-channel MOSFET for shutting down a positive voltage below 500 V. DE275 is supplied by Directed Energy Inc. (b) Cascaded avalanche npn transistors for switching off a negative voltage of twice the collector-base breakdown voltage (about -700 V). The transformer is a high-frequency ferrite ring (40–400 MHz) with five windings of strand wire for each coil. The Schottky diodes and the 100- Ω resistor protect the trigger circuit against voltage steps passed by the gain-multiplied collector-base capacitance during switching. The 200-pF capacitor holds off the DC high voltage in case of isolation failure.

supply the required large gate currents if a high negative voltage is to be switched, since gate and source are then at a high potential. In contrast, the voltage-triggered avalanche transistor cascades can readily be used as switches for voltages of either polarity. Unfortunately, electromagnetic interference is pronounced, and it proved problematic to prevent two cascades from firing simultaneously. For these reasons switch 1, handling a positive voltage, is based on the fast MOSFET, and the avalanche transistor cascade is used to switch the negative voltage.

2.2 Photoelectron Guns

DTEM requires an extremely bright pulsed electron source. A 512 $\times$ 512 pixel image on the border of usefulness requires $\sim 10^6$ electrons at the detector, while a good image requires $\sim 10^8$ (Figure 5-4). To obtain a usable image in a single shot, this number of electrons must pass through the sample region of interest in ~ 1 –10 ns, which implies a

current of order 0.01–10 mA. Further, to obtain good results the illumination spot must be significantly larger than the region of interest, and some electrons will have to be thrown away at various apertures. So even these enormously high currents may not be enough for some applications. To squeeze 10 mA of current into a 1- μm radius spot with a 1 milliradian convergence semiangle would require a normalized brightness of order $4 \times 10^9 \text{ A cm}^{-2} \text{ steradian}^{-1}$ (assuming a typical TEM voltage of 200 keV). This is too high a brightness for a thermionic cathode (pulsed or otherwise) and too much current for a typical field emission gun.²³ This brightness is just within reach for a pulsed photocathode-based system^{24,25} and the required current can be attained by using a somewhat larger emission area than is typical for TEM. Beyond this point, signal levels must be traded off against spot size and/or convergence angle—or invent a brighter electron source, possibly using pulse compression (discussed in a later section). Phenomena such as charge trapping²⁶ and space-charge repulsion from previously emitted electrons (i.e., the effect that leads to Child’s law in the steady-state case)²⁷ may also limit the number of electrons in a pulse.

Bostanjoglo et al. report a maximum current of 20 mA in a 20-ns pulse at 80 kV with a brightness of $\sim 2 \times 10^6 \text{ A cm}^{-2} \text{ steradian}^{-1}$, achieving a factor of 50 enhancement in brightness by operating in a laser-driven pulsed mode as compared to a DC thermionic mode.^{27–29} The energy spread was reported as 8.7 eV. Brightness is roughly proportional to accelerating voltage in this regime; most of the reported increase of brightness with voltage may be attributed to the conservation of normalized brightness, which is $\sim 6 \times 10^6 \text{ A cm}^{-2} \text{ steradian}^{-1}$ for this case. Pulsed high-voltage RF photoelectron guns can significantly exceed this brightness and current,^{24,25,30} but at the cost of very high energy spreads $\sim 1\%$ or more, which may limit their suitability for DTEM applications. An interesting near-future challenge would be to determine whether and how to borrow design concepts from the RF guns to improve the DTEM sources.³⁰

At any rate, in the DTEM every electron is precious. In static TEM it is common practice to block most of the electrons at the condenser apertures in order to attain good spatial coherence and small lateral emittance [emittance is, to within definition-dependent factors of order unity,³¹ the product of the spot size and convergence angle and is a conserved purely geometric quantity in a linear electron-optical system at constant voltage with no apertures (Reiser,¹⁹ pp. 56ff)]. This allows the production of very fine spots limited in size primarily by diffraction and spherical aberration, or of wider spots with very small incident angular spreads. This enables the familiar “nanoanalytical” and “parallel-beam” modes of operation. In the DTEM one will typically be electron-starved and will not have the luxury of using a small condenser aperture; the instrument performance will necessarily suffer. A typical TEM that can attain atomic resolution lattice images and subnanometer spot sizes in a static mode will simply not be able to do these things in nanosecond pulses with any foreseeable electron source.

2.3 Materials for Photoelectron Guns

The ideal photo cathode should (1) be able to produce high current densities of about 1 kA/cm^2 , (2) be resistant to residual gases in a high vacuum (oxygen, water, hydrocarbons, carbon monoxide, and carbon dioxide), (3) be tolerant to ion bombardment and thermal spikes due to arcing, and (4) exhibit a low work function to relax requirements on the driving photons.³²

To shift the negative space-charge-induced saturation of the electron emission to higher current densities, the accelerating electric field should be as high as possible. This increases the chance for an electric breakdown between the cathode and the accelerating anode. An ion bombardment of the cathode results until the charge of the capacitance between the cathode (usually at high negative potential) and ground is exhausted. The ions are produced by ionization of evaporated anode material and desorbed/disintegrated layers (oxides, hydrocarbons, water, air molecules) covering all surfaces in an ordinary high vacuum system. If the capacitance between cathode and ground is not small enough a high current arc can develop between cathode and anode. The energy of the impinging ion pulse is large enough to sputter and heat the cathode substantially. A weak but continuous ion bombardment derives from electron beam-induced ionization of the residual gas. The sputtering and chemical reaction caused by these ions are negligible, when the cathode is operated in ultrahigh vacuum.

Materials suitable for photocathodes can be divided in semiconducting and metallic photoelectron emitters in view of their band structure and associated current limitation.

2.3.1 Semiconducting Emitters

In semiconducting emitters valence electrons are photoexcited into the conduction band. If the energy of the photons exceeds the sum of band gap and electron affinity, some excited electrons tunnel through the surface potential barrier and escape to the vacuum. The rest are trapped in surface states, increasing the surface barrier. Fortunately, they are neutralized by thermionic and tunneling holes from the valence band, which restore the equilibrium surface barrier. However, if the photocathode is driven with nanosecond and shorter light pulses of high intensity, the number of excited electrons that are trapped can become too large to be compensated by the restoring hole currents. The surface barrier is then increased, and the extractable charge saturates at a value lower than the space charge limit. This surface charge limit effect is a characteristic of nonmetallic emitters.²⁶

Cesium-based antimonides, such as Cs_2Sb and K_2CsSb , are among the materials with low work function (about 2 eV).³³ But they are extremely sensitive to contamination, thermal spikes, and ion bombardment. For a higher work function, about 3.5 eV, there are materials that are much more resistant to contamination by residual gases such as Cs_2Te and GaAs-based emitters.^{33–36} Their thermal stability is poor, however. Providing heavily p-doped GaAs or other III–V semiconductors with Cs-suboxide surface layers produces materials with a nega-

tive electron affinity.^{34–37} These materials have large quantum efficiencies for spin-polarized electrons, 0.3% for wavelengths of 810–830 nm. The heavy surface doping also enhances the restoring hole currents that compensate trapped photoexcited electrons, avoiding the surface charge limit phenomenon.^{35–37} Unfortunately, the doped surface layers make the negative electron affinity cathodes again vulnerable to contamination, ion bombardment, and thermal stressing.

In contrast to all the above semiconducting materials, p-doped diamond and fullerene films³⁸ are extremely wear and heat resistant. The quantum efficiencies, however, are distinctly lower (about 3×10^{-4} for 217-nm photons).

2.3.2 *Metallic Emitters*

Metallic emitters can be subdivided into compounds with metallic conductivity and metals.

Rare earth monosulfides³⁹ and hexaborides,^{40–42} borides and carbides of refractory metals, are examples of compounds with a low work function and high conductivity.⁴⁰ The rare earth sulfides, e.g., CeS, NdS, and PrS, are distinguished by very low work functions of 1.1–1.4 eV, high melting temperatures of 2400–2700 K, and high resistance to ion bombardment and contamination. Thin films of these materials are readily produced by vacuum evaporation. However, their vapor pressure at elevated temperatures is substantial, which prohibits thermionic operation.

The rare earth hexaborides LaB₆ and CeB₆ are widely used in thermal cathodes. They are thermally more stable than the monosulfides, but they are disintegrated by ion bombardment and have a higher work function of about 2.6 eV.

Borides and carbides of refractory metals have even higher work functions of 3.0–3.5 eV. But since their melting temperatures are extremely high, e.g., 3800 K in the case of ZrC, these emitters readily tolerate arcing. In addition, they are not sensitive to contamination and can therefore be operated in ordinary high vacuum.

Metals exhibit a wide range of work functions as well as thermal and mechanical strengths. Candidates for high-performance photocathodes are reviewed by Travier.⁴³ Alkali and alkali earth metals are at the low end of the work functions. They are chemically extremely sensitive and mechanically and thermally weak. The mechanical strength and tolerance to arcing of cathodes based on these high quantum efficiency metals can be substantially increased by implanting their atoms into tungsten or other high melting metals.⁴⁴ Nonetheless, these cathodes remain sensitive to contamination and require ultrahigh vacuum.

Rare earth metals have medium work functions of about 3.5 eV. They are mechanically strong and high melting metals. But they vigorously react with oxygen producing highly isolating oxide layers. Positive ions attracted by the negative cathode settle on the oxide layer where they produce a high electric field, which in turn triggers a breakdown (Malter effect). Accordingly, rare earth cathodes suffer from perpetual electric breakdowns when they are operated in high vacuum.

Most metals with a high thermal and electrical stability and tolerating high vacuum have inconveniently large work functions. Zirconium is a compromise. This metal combines a medium work function (3.8–4.05 eV depending on impurities) with a high melting temperature (2130 K), and it forms semiconducting suboxides in high vacuum, which neither deteriorate the quantum efficiency nor favor an electric breakdown.

The influence of impurities on photoemission is very complicated. Surface contamination by water, carbon monoxide, and carbon dioxide, and hydrocarbons usually poisons photoemitters by producing hydroxide or carbonaceous layers. Oxygen, for instance, can either increase or reduce the work function, depending on the material, the emitting crystal lattice plane, and the surface coverage. Oxidation of materials with a low or medium work function usually decreases photoemission. Zirconium is an exception, if oxidation is limited so that only a few layers of the suboxide ZrO are formed.⁴⁵ Water, carbon monoxide, carbon dioxide, and hydrocarbon molecules can poison photoemitters by producing covering oxides or carbonaceous layers. Alkali and alkali earth metals are chemically most reactive and therefore very sensitive to the vacuum atmosphere. Cs₂Tb is a favorable exception³³ as it is relatively unaffected by carbon compounds abundant in conventional high vacuum, such as CO, CO₂, and CH₄. If contamination is an issue, a satisfactory lifetime (several hours) can be achieved only at pressures below 10⁻⁹ mbar.

The choice of a photocathode is a compromise between high quantum efficiency and long stable operation. Having selected the material and the driving photon source, being almost exclusively a laser, the generation of electron pulses with high current densities can be improved in several ways:

1. Using a photoelectron emitter with a rough surface enhances light absorption, makes use of field-assisted photoemission (Schottky effect), and photo-assisted field emission, increases the accelerating field and thus shifts the space charge-induced saturation to higher current densities, and provides the correct orientation of many surface elements to the electric field vector of the light for a maximum electron yield. Depending on the nature and degree of roughness the quantum efficiencies can be increased by a factor of three⁴⁶ and more.⁴⁷
2. Producing a small plasma cloud with the driving laser pulse, the saturation of the electron current due to negative space charge is avoided, and the electron pulse charge is augmented by secondary electrons.²⁷ However, the pulse is stretched as trapped electrons are released when the plasma decays.
3. Ultrashort driving laser pulses provide additional benefits. First, if the emitted ultrashort electron bunch is shorter than the accelerating gap, its space charge field is approximately $j\tau/2\epsilon_0$, with j the electron current density, τ the pulse duration, and ϵ_0 the vacuum permittivity. The current density saturates as this obstructing field approaches the accelerating field, at values that may exceed the Child–Langmuir limit of stationary space charge by orders of magnitude.⁴⁸ Second, if the laser

pulse duration is shorter than the electron–phonon relaxation time, about 10 ps for metals, the electrons can be excited to a high temperature and be thermally emitted before losing substantial energy to the lattice.⁴⁹

If photoemitters on substrates are to be subjected to elevated temperatures, e.g., for continuous thermionic emission to simplify alignment and for cleaning and conditioning as in the case of dispenser cathodes,⁵⁰ the supporting material should neither disintegrate the photoemitter nor produce a low melting eutectic. Substrate atoms rapidly penetrate along grain boundaries and eutectic channels to the surface and deteriorate the quantum efficiency if their work function is high.⁵¹ In view of this, for example, Zr on W or Re is a bad choice, as eutectics exist for these systems. However, Zr on Ta works well since these metals have a miscibility gap confining Ta that reaches the surface predominantly to Zr grain boundaries (Figure 5–8). A Zr-covered Ta hairpin emitter can be used as a conventional thermionic electron source and as a pulsed high-current photocathode, both in a normal high vacuum ($<10^{-4}$ mbar). Electron pulses with a current density of 700 A/cm^2 were produced by a laser radiation with a wavelength of 266 nm.⁵¹

In summary, the quantum efficiencies (QE) of different less sensitive metals are listed with references in Table 5–1 with the exciting wavelength. Photocathodes used at TU Berlin had a brightness of $\sim 6\% \cdot 10^7 \text{ A/cm}^2\text{sr}$ at 100 kV, a quantum efficiency of $1.2\% \cdot 10^{-5}$, and emitted photoelectron pulses with current densities of up to 700 A/cm^2 from an area of $20 \mu\text{m}$ diameter.¹⁰

2.4 Lasers

Two lasers are required in the typical DTEM pump-probe experiment—one each for the cathode and the sample. At TU Berlin, these lasers must produce $\sim 6\text{--}10 \text{ ns}$ pulses with spot diameters of $20\text{--}40 \mu\text{m}$. This is achieved with Q-switched Nd:YAG and Ti:Sapphire lasers. The

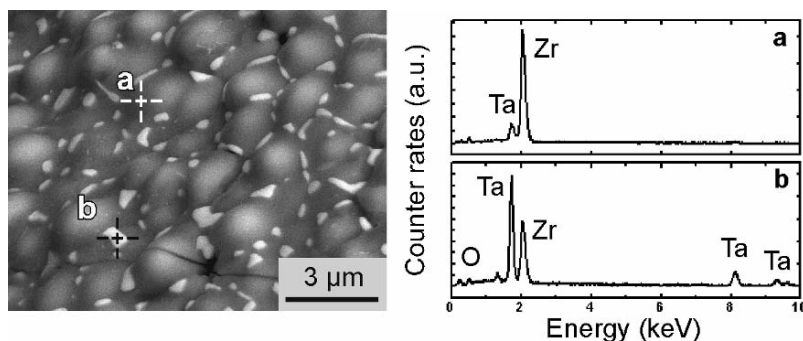


Figure 5–8. A Zr on Ta cathode after 6 h of operation as a thermionic cathode, imaged by scanning electron microscopy with energy dispersive X-ray spectroscopy. The bright islands between the Zr grains (dark) are Ta, which has penetrated the Zr layer. The quantum efficiency ($4\% \cdot 10^{-5}$ for light with a wavelength of 266 nm) was not affected.

Table 5–1. Quantum efficiencies of less sensitive metals.

	λ (nm)							
	193	211	213	248	263	266	308	355
	QE (10^{-5})							
Ag		28.4 ^a			2.6 ^a	2.0 ^b		
Al		35.9 ^a	84 ^c		2.6 ^a	3.2 ^c		0.034 ^c
Au		16.2 ^a	40 ^c		1.1 ^a	1.3, ^c 4.7 ^b		
Cu	20 ^c	28.6 ^a	15 ^c		0.84 ^a	0.22, ^c 14 ^{a,b}	0.016 ^c	
Mg						62 ^b		
Mo		14.4 ^a			0.49 ^a			
Nb	45 ^d			0.3 ^d				
Pd						1.2 ^b		
Sm		6.8 ^a			0.47 ^a	72 ^b	0.016 ^c	
Ta		12.5 ^a			0.94 ^a	1.0 ^b		
Tb		8.9 ^a			0.61 ^a	23 ^b		
Y		13.9 ^a			1.4 ^a	0.27, ^c 50 ^b	0.11 ^c	
W(K+)							1.2 ^c	
W(BaO)						35 ^e	14 ^e	2.3 ^e
W(Ba-Ca-Sr-O)						230 ^d		
Zn		24.7 ^a			1.4 ^a	1.4 ^b		
Zr						1.0 ^b		
Li-Al						60 ^f		10 ^f

^a Anderson T., Tomov I.V., Rentzepis P.M. Laser-driven metal photocathodes for picosecond electron and X-ray pulse generation. J Appl Phys 1992;71(10):5161–5167.

^b Srinivasan-Rao T., Fischer J., Tsang T. Photoemission studies on metals using picosecond ultraviolet-laser pulses. J Appl Phys 1991;69(5):3291–3296.

^c Chevallay E., Durand J., Hutchins S., Suberlucq G., Wurgel M. Photocathodes tested in the Dc gun of the Cern Photoemission Laboratory. Nucl Instrum Meth A 1994;340(1):146–156.

^d Leblond B., Rajanoera G. Photoemission in the picosecond regime from a coated trioxide cathode. Nucl Instrum Meth A 1994;340(1):195–198.

^e Leblond B. Short pulse photoemission from a dispenser cathode under the 2nd, 3rd and 4th harmonics of a picosecond Nd-Yag laser. Nucl Instrum Meth A 1992;317(1–2):365–372.

^f Septier A., Sabary F., Dudek J.C., Bergeret H., Leblond B. A binary Al/Li alloy as a new material for the realization of high-intensity pulsed photocathodes. Nucl Instrum Meth A 1991;304(1–3):392–395.

required pulse energy depends significantly on the application and can be as high as ~200μJ for driving a pulsed thermionic cathode.²⁹ Frequency doubling and quadrupling is employed to more effectively drive the required material transitions. This is especially important for the photocathode guns, which require ultraviolet light and are driven with a frequency-quadrupled Nd:YAG laser at 266 nm wavelength. The laser spots have Gaussian profiles, which affects the spatial uniformity of both the sample drive and the cathode emission.

The need for multiple-frame movies (with each frame illuminated by its own cathode pulse) complicates the laser situation somewhat, especially since different applications may require very different pulse delays. The TU Berlin pulsed photocathode can produce single 7-ns

electron pulses, but uniformity suffers as the number of pulses is increased. This is because of the difficulty of producing multiple closely spaced identical pulses in a Q-switched nanosecond laser system. The second and third pulses have durations of 10 and 11 ns, respectively, with spacings ranging from 20 ns to 10 μ s.⁵² Producing more than three nearly uniform pulses becomes progressively more difficult. The development of many-frame movies will require that this challenge be fully addressed.

3 Limitations

3.1 Time Resolution

The duration of the electron pulse at the sample determines the time resolution of the DTEM in a normal imaging or diffraction mode. This will be our primary concern in this section. Higher time resolution has been demonstrated in a streak mode,^{53,54} at the cost of having only a one-dimensional image available at any given point in time. The great majority of the electrons penetrating the sample are lost in these cases, but the higher time resolution may be worthwhile for many applications.

The electron pulse length is essentially equal to the cathode laser pulse length in current instruments. The cathode laser pulses are several nanoseconds long, whereas the mechanisms that would further degrade the time resolution typically occur on a picosecond scale. The time delay between a photon striking the cathode material and a photoelectron exiting is much faster than a nanosecond (streak cameras and ultrafast electron diffraction systems would not work otherwise), provided the emission is purely via the photoelectric effect. Schafer and Bostanjoglo²⁷ reported a pulse duration of 20 ns due to heating and cooling of the filament and plasma relaxation processes, but this was in a laser-driven thermionic source rather than a true photocathode; in later work this limitation is bypassed.⁵² Temporal expansion due to the energy spread in the pulse in the initial acceleration region is of order

$$\Delta t = \frac{(m\Delta E/2)^{1/2}d}{eV} \quad (1)$$

with m and e the electron mass and charge, ΔE the energy spread, d the cathode–anode distance, and V the accelerating voltage.⁵⁵ This amounts to a few picoseconds or less for typical DTEM parameters. After initial acceleration, the velocity spread is less than one part per thousand and again is a negligible source of temporal spread. The pulse of electrons is much longer than it is wide—a 1-ns, 200-keV pulse will be ~15 cm in length but a small fraction of a millimeter in diameter (depending on lens settings and position within the column). In this high-aspect-ratio limit, longitudinal space charge effects can be practically neglected except near the very leading and trailing edges of the pulse. Image charges in the metal walls of the column may interact with nonuniformities in the linear charge density to produce broadening and instabilities,^{56–58} but it is fairly easy to estimate from the gov-

erning equations that this phenomenon will have a negligible effect over the few nanoseconds that the electrons are in the DTEM column. The effect is far more important for large-scale particle accelerators. Space charge thus might spread out the tails of the temporal response curve but will have very little effect on its FWHM; the pulse is for the most part steady state. Timing jitter in laser pump-probe experiments is similarly on the picosecond scale or faster. Thus until DTEM pulses get into the picosecond range (which is likely to be very difficult for the above reasons), the time resolution will be controlled entirely by the photocathode pulse length.

Picosecond DTEM will also be extremely challenging purely on the basis of signal levels. As we have seen, the DTEM is already electron starved in the nanosecond regime, so that the optimal time resolution is dictated by (1) the required fluence at the sample, (2) the maximum tolerable convergence angle, and (3) the brightness of the electron source. Pushing the time resolution to the picosecond range will only be possible either with an extremely bright electron source or with very unchallenging imaging requirements. It will likely be some time before this is accomplished.

In summary, the time resolution for a single shot is dictated by the electron fluence requirements and directly controlled by the photocathode laser pulse duration. What then determines the interframe time resolution for a multiple-shot experiment? Single-electron-sensitive high-resolution high-dynamic-range cameras with nanosecond readout times are somewhat beyond current technology. This means multiple shots must be fit into a single CCD image, which is read out once after all exposures are complete. This is achieved with the previously described "movie mode," implemented with a stepped electrostatic deflector system. The minimum interframe time is ~ 20 ns, dictated by the ring-down time of the high-voltage switching circuitry.

With the principle proven, the problem of producing better movies (with more frames and/or reduced interframe times) is a nontrivial matter of high-voltage/high-speed electrical engineering. The driving system must produce a series of preprogrammed voltage plateaus at very precise times with as little ring-down as possible. It needs to be flexible, precise, and as free of parasitic inductance and capacitance as possible. Some advantage may be gained by using a more recent TEM column design with more than one intermediate lens. Such columns offer more degrees of freedom to optimize the "leverage" (displacement at the camera plane per unit deflection voltage). It may also be possible to increase the frame count by using a two-dimensional deflector, although this greatly complicates the electrical design problem. Ongoing developments in multiframe camera technology^{59–61} may improve the frame counts, particularly if combined with an advanced electrostatic frame shifter. Each buffered CCD frame could thereby contain multiple images, and the CCD's interframe time could be much longer than the DTEM's interframe time. This would unfortunately force the high-voltage switching system to repeat its performance multiple times with very little delay, not to mention the challenges in producing the required cathode laser pulses. If the engineering problems

turn out to be surmountable, we may (although certainly not very soon) see DTEM movies with hundreds of frames and interframe times perhaps as short as 1 ns.

It is instructive to compare the time resolution in DTEM with that of conventional *in situ* TEM. Conventional systems are typically limited to video rate imaging, ~30 ms per frame. DTEM thus opens up a regime of more than six orders of magnitude in time resolution. This advantage is currently offset by the stringent limitations on multiframe DTEM movies. DTEM is therefore complementary to conventional *in situ* techniques. Video-rate TEM is too slow to capture many nanoscale solid state processes—often a significant part of the dynamics occurs between two frames of video.⁶² DTEM offers an opportunity to help fill in these gaps of time, provided issues of timing and frame count limitations can be addressed.

3.2 Image Resolution

Our present understanding of the spatial resolution limits in DTEM is limited, due to a lack of systematic studies on both empirical and theoretical fronts. This section must therefore be considered as a sort of snapshot of our current state of understanding in a very young field. Still, we can make some statements on general principles and identify future areas of research.

The stated resolution in the TU Berlin instruments in pulsed mode is ~200 nm⁵² (in some of the images features as small as ~100 nm seem to be coming through⁶³). This resolution limit is attributed to a combination of the point spread function of the camera and the SNR limiting the pixel size for a given fluence⁵² (recall Figure 5–4). Both limits could in principle be improved by increasing magnification and reducing the illuminated area, but these will incur various tradeoffs. For example, time-dependent radiation sensitivity of the sample may limit the current density in specific cases, so that the practical resolution limit may be very strongly dependent on the robustness of the sample under the electron beam. Also, the uniformity of the illumination area may suffer due to the large beam emittance coupled with condenser lens aberrations and nonuniform emission at the cathode.

In any case, if we assume that the sample can survive the required fluence, then the fundamental resolution limit of the machine must lie in the electron optics. The ability of static TEM to attain Angstrom-scale resolution is dependent on several factors including brightness, energy spread, objective lens aberrations, the quality of the sample, and the faithful magnification of the image by the intermediate/projector lens system. All of these may have to be compromised in the DTEM. Further, electron–electron interactions including space charge, Boersch effects, and trajectory displacement effects^{64–66} will be much more important in the DTEM, with its extremely high currents compared to static TEM. The DTEM has at least one advantage, in that low-frequency vibrations and high-voltage drift that would harm the resolution in the conventional TEM will have no effect on a nanosecond-exposure image.

Let us consider the electron-optical factors in some detail. First, consider the need for a bright electron source. As we have discussed, high brightness is needed to attain high fluence with nearly parallel illumination. Any angular spread in the condensed beam will tend to wash out the high-spatial-frequency information.⁶⁷ The necessarily large emittance of the DTEM electron source will tend to compromise this performance. Also, the energy spread ΔE of the electrons will be quite large in the DTEM, due most likely to an enhanced Boersch effect at high currents.⁵² ΔE is about 8 eV for the TU Berlin instruments.²⁸ The energy spread will interact with the objective lens chromatic aberration C_C to limit resolution to

$$d_C = C_C \theta \frac{\Delta E}{E} \quad (2)$$

where θ is the angular deviation of the electron path from the optic axis.⁶⁷ If we make the simplifying assumption of normal incidence, then θ is related to the feature spacing d via the usual formula

$$d = \lambda / \theta \quad (3)$$

(assuming coherent single scattering and small angles), and this implies a chromatic resolution cutoff of order

$$d_C = \sqrt{C_C \lambda \frac{\Delta E}{E}} \quad (4)$$

In addition, spherical aberration limits resolution to a value of order⁶⁷

$$d_S = 0.66(C_S \lambda^3)^{1/4} \quad (5)$$

This rounds out the list of electron-optical effects that usually limit resolution in a conventional TEM.

These effects all interact with one another, and it is best to use TEM simulation software (such as Java EMS,⁶⁸ used here) to understand their combined effects. The square root dependence in Eq. (4), for example, is very much complicated by beam convergence effects. Consider Figure 5–9, which presents amplitude and phase contrast transfer functions for a set of imaging parameters that are similar to those that occur in a DTEM. Parameters comparable to the ones used in this calculation were measured²⁸ near, at, or below the sample position and thus include all effects that take place in the gun, accelerator, and condenser systems. It would appear that atomic resolution (via phase contrast, at spatial frequencies ~ 3 cycles/nm or higher) will be extremely difficult, but that it may well be possible to achieve usable contrast up to a resolution limit of ~ 1 – 2 cycles/nm, particularly through amplitude contrast mechanisms. The rough square root dependence on energy spread helps to minimize the impact of chromatic aberration; Eq. (4) predicts a cutoff of 1.4 cycles/nm. Spherical aberration [Eq. (5)] predicts a roughly comparable 2.1 cycles/nm. The lack of spatial coherence (represented in the simulation as an incident angular spread) interacts with the other parameters in a complicated way, broadening the CTF pass band but significantly reducing its amplitude. Thus it may be

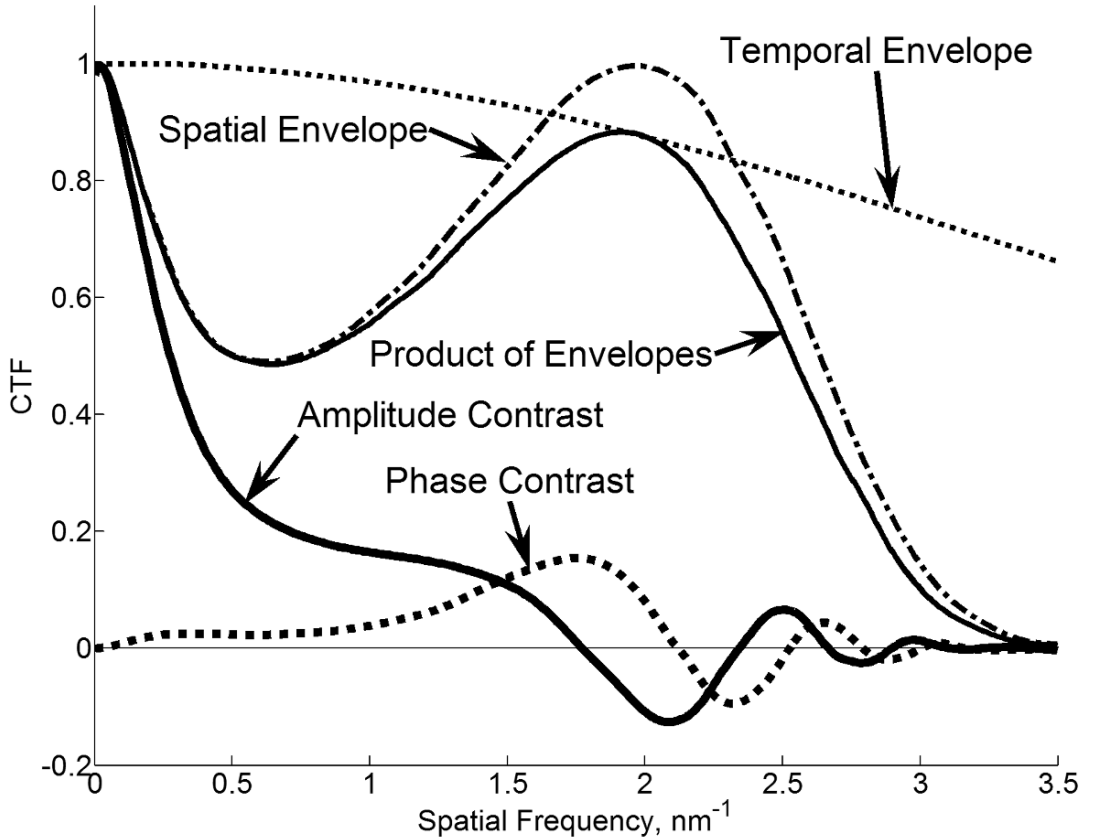


Figure 5-9. Phase and amplitude contrast transfer functions (CTF's) as simulated by the commercial software package Java EMS [68] for the following parameters: $E = 100$ keV, $DE = 5$ eV (the maximum allowed by the software), $CS = CC = 2$ mm, convergence semi-angle 10 mrad, defocus 105 nm, no objective aperture. The cross-interaction between convergence angle and aberrations is primarily responsible for the low contrast. Phase contrast will be very weak, however amplitude contrast mechanisms should produce reasonable images with better than 1 nm resolution under the assumptions in this model. Therefore the effects that limit DTEM resolution must be among those not included in this simulation, which was designed for conventional TEM.

very important to control the convergence angle to obtain a good DTEM image. Reasonably high-quality selected area diffraction patterns have been achieved in DTEM over areas of some tens of micrometers,⁶⁹ indicating that the convergence angles can be kept to reasonable levels.

In short, the estimated resolution limit accounting for all known (and measured) effects in the gun, accelerator, condenser system, and objective lens would appear to be ~ 1 nm, dramatically better than is achieved experimentally. This suggests that the limitation is in one of the effects not included in conventional TEM simulations; this also suggests that if the limiting factor can be identified and eliminated, a great improvement in resolution should be possible. The above argument indicates that the culprit should be either in the sample, in the postsample lens systems, or (as indicated in Domar and Bostanjoglo⁵²) in the camera (including detection statistics). The balance of this section will discuss some of these effects.

Consider the interaction of a very intense beam with a solid sample. In conventional TEM, $1\text{ }\mu\text{A}$ would be an extremely high current at the sample; this corresponds to roughly 1 electron per 160 fs, with an axial spacing of $\sim 33\text{ }\mu\text{m}$ at 200 keV. Plasmon lifetimes are in the femtosecond range, while TEM samples are usually less than $1\text{ }\mu\text{m}$ thick, so each electron encounters a sample that has had time to relax since the last electron passed through it (apart from heating and radiation damage). DTEM currents may be four orders of magnitude larger. More than one electron will be in the sample at a time, and at high fluences each electron may encounter multiple previously excited plasmons on its way through the material. The theory of electron-sample interactions in the electron microscope always makes the one-electron-at-a-time approximation. This theory may have to be modified for the DTEM. Further, a great deal of energy will be deposited in a region highly localized in time and space, which raises the possibility of new radiation damage mechanisms. It is not clear at this time whether and how these effects might worsen the image resolution.

The abnormally high currents in the DTEM may also affect the behavior of the intermediate lens system, specifically through electron-electron interactions in the intense crossovers. We have performed rough estimates of the various effects involved (space charge, Boersch, and stochastic particle displacement), following the formalism of Kruit and Jansen.⁶⁵

Space charge effects can change the effective focal lengths and aberration coefficients of each of the lenses. Space charge defocusing would merely require a readjustment of the lens currents and a possible recalibration of the microscope in pulsed mode. The experience at TU Berlin suggests that this is not a major issue. Estimating the effects of space-charge-induced spherical aberration is somewhat more difficult and has not yet been performed in detail. Part of the problem is that this effect vanishes to first order for a circular beam of uniform intensity,⁶⁵ so any associated distortions would have to be driven by nonstatistical inhomogeneities in the beam intensity (including those due to

the image contrast itself). Little more can be said without extensive modeling, but fortunately the relevant models are already well developed in other contexts.⁶⁶

The Boersch effect acts to increase the energy spread of a charged particle beam by coupling the lateral and longitudinal degrees of freedom via statistical electron–electron interactions.^{64–66} This process will saturate when all degrees of freedom are, in some sense, in thermal equilibrium. Saturation energy spreads depend on accelerating voltages and convergence angles.⁶⁶ Rough estimates suggest that this effect has adequate time to saturate in the early part of the column, and that it may well be responsible for much of the energy spread measured by Bostanjoglo et al.²⁸ [a very rough estimate of the saturation energy spread based on assumed parameters yields ~10 eV, compared to the measured 8.7 eV, while the original authors used Loeffler’s formalism (validated to within ~20%)⁷⁰ to estimate a value of 7.6 eV]. In any case it would seem that the maximal effect of chromatic spread would already have occurred in the objective lens (with a resolution limit ~1 nm), so that it seems doubtful that chromatic aberration in the intermediate lenses is to blame for the observed resolution limit of ~200 nm.

This brings us to the trajectory displacement effect,^{65,66} which slightly and randomly deflects electrons in and near a crossover. Suppose we take an intermediate lens with focal length 24 mm forming an image of radius 0.1 mm at a distance of 60 mm, with a crossover at 27 mm. The divergence angle α_0 is therefore 3 mrad. These values might come up between the first and second intermediate lenses of a typical TEM in a moderate-magnification imaging mode, with an illuminated area of 1 μm at the sample. Suppose 10^8 electrons are in a pulse of duration 1 ns, for a current of 16 mA at 200 keV. Using the formulas quoted by Kruit and Jansen,⁶⁵ we find that we are in the Gaussian regime, with an average random trajectory displacement of 2.2 μm at the crossover and an estimated 5 μm at the image plane. The result will be an image blur much like what happens in projection electron lithography systems. Referencing this displacement back to the sample plane yields a resolution limit of ~50 nm due to this crossover alone, assuming that the rest of the optical system faithfully transfers this image plane to the screen. This is a very significant limitation to the resolution (comparable to the quoted resolution limit of ~200 nm in current instruments⁵²), yet it must be taken as only a very rough estimate of the effect in a real TEM column. This resolution limit depends strongly on all relevant parameters, including the positions, intensities, and convergence angles of all crossovers, the positions and degrees of magnification at every image plane, and the beam current and voltage. This means that (1) much more involved calculations and experiments are required if we are to understand if this is truly the resolution limit, and (2) if it is the resolution limit, then it may be possible to dramatically reduce its effects by changing the lens excitations in the intermediate/projector lens system. The idea would be to avoid intense crossovers with small convergence angles at high-leverage parts of the column. A simple way to test the

hypothesis would be to take the same pulsed images with and without a small selected-area aperture. If trajectory displacement is the resolution limit, then when the aperture is inserted the electrons from outside the region of interest will no longer be present at the intermediate lens crossovers, and the effect should be greatly reduced.

We should also remember that the sample may well be moving, rapidly, during the exposure. A phase front or shock wave moving at roughly the speed of sound in the material may move several micrometers during a few-nanosecond electron pulse. The sample may also be distorting from local heating or moving vertically (out of focus) due to the effects of the pump laser. These effects must be considered in the design of experiments. However this does not appear to be the resolution limit in the experiments to date; roughly the same resolution is obtained in test shots on a static sample.

In summary, the observed resolution limit in DTEM may arise from a combination of intermediate crossover interactions, motion blurring, and low signal levels, while the effects that limit resolution in conventional TEM would suggest a pulsed image resolution ~ 1 nm. Various experiments are proposed that may identify the true culprits, while some methods of mitigating their effects are suggested. This is the present state of knowledge at the time of writing; the near-future evolution of the understanding of these effects is likely to be rapid.

3.3 Damage

Due to the inelastic scattering of electrons in TEM, a large amount of energy (typically some 10s of electron-volts per primary electron) is deposited in the sample, and in many cases damage due to energy deposition is a limiting factor in obtaining high-resolution images. For continuous electron exposure used in standard TEM, damage is linearly related to the total radiation dose,^{71,72} but to fully understand the impact of damage on image quality in DTEM, the effects of higher dose rate and the time-dependent nature of radiation damage must also be considered. At this point, we do not know enough about either of these subjects to confidently predict the damage-related resolution degradation in a DTEM image. Nevertheless, here we briefly review what is known as a baseline for further investigation.

The consequences of energy deposition for TEM images depend strongly on the nature of the sample. Although a wide variety of inorganic and metallic samples can be readily imaged, sometimes with very high resolution and contrast,⁷³ damage is the most significant limiting factor for imaging less robust samples such as biomaterials. Although damage in solid-state material due to direct atomic displacement can be significant at high beam currents,⁷⁴ typically this is appreciable only for a large integrated number of primary electrons incident on the sample and can be avoided (in standard TEM) through the use of a primary electron energy that is less than a sample-dependent threshold. For a current density of 10^5 A/cm² through a carbon film sample at between 100 and 200 keV, the sputtering rate is roughly 10 nm/s.⁷⁴ Although the instantaneous current density in DTEM could

be considerably higher than this (as high as 10^7 A/cm² in future instruments), this current density will be sustained only over nanoseconds. Assuming the sputtering rate is linearly related to the current density (which may not be the case for DTEM), sputtering would be insignificant for less than millions of shots.

Since the time scale of the electron pulse can be comparable to the thermal diffusion time, sample heating is a more significant issue for DTEM than standard TEM, but this depends strongly on the sample and beam properties. The deposited energy density can be very large. For water irradiated by ~200 keV electrons at a density of 10 electrons/Å² (a typical density for cryoEM) in a single pulse, straightforward estimates of the inelastic cross section⁷¹ result in a deposited energy density of about 300 eV/nm³, about 10 eV per water molecule! For a Gaussian beam with $1/e$ radius of 100 nm and $1/e$ half width of 1 ns in water, naively assuming only standard thermal diffusion using room temperature constants, the deposited energy density will reach 90% of the value it would have with no diffusion. The situation is significantly better with the lower heat capacity and larger thermal conductivity found in metals. With the same beam properties, the energy density in an aluminum sample will reach only 5% of the zero diffusion energy density. These results depend strongly on the pulse duration. A pulse of sufficiently short duration (less than 10s of picoseconds for a 100-nm radius beam in aluminum) will still overwhelm diffusion within this model. At a sufficiently short pulse duration and electron density, sample heating may result in sample melting and lowering the atomic displacement energy threshold. Nonetheless, questions remain as to the actual state of the material at these early times and high temperatures—it is unlikely that any standard assumptions about thermal transport or material stability would still apply.

Biomaterials (such as proteins) are especially ill suited to examination via TEM due to their native aqueous environment, caustic species created in the irradiation of water, low contrast, intrinsic radiolytic susceptibility, and the relatively weakly bound tertiary and quaternary structures of proteins. The “lifetime” of irradiated protein crystals (i.e., the amount of time a crystal may be irradiated until its diffraction pattern fades) is a well-documented obstacle⁷⁵ to high-resolution protein structural characterization. Yet, biology would benefit immensely from the ability to efficiently image proteins at high resolution. Knowledge of the detailed functioning of biomolecules and biomolecular complexes is essential to our understanding of the functioning and evolution of life.⁷⁶ Here we focus on the imaging of biomaterials, since this application has great scientific benefit as well as being very challenging. Damage mechanisms would be simpler for just about any other condensed-state sample.

Specifically, how much will damage degrade the resolution of a DTEM image over the time scale of the electron pulse? More optimistically, is it possible to effectively mitigate the effects of damage with respect to a high-resolution image by using a short electron pulse? Such an idea has already been proposed for imaging with X-rays,⁷⁷ where single particle diffractive protein imaging is thought to be

possible using a 10-fs duration X-ray pulse. Although pulsed imaging ultimately results in the destruction of the sample, the idea is to obtain the image before damage has significantly distorted the structure being imaged. An incidental, yet highly beneficial product of this scheme is the ability to time-resolve structural changes, enabling the direct visualization of nanobiostructural dynamics—a molecular movie.⁹

An analogous imaging scheme using electron pulses presents both advantages and challenges. Electrons deposit much less energy per inelastic scattering event than X-rays: less than 50 eV per event for electrons⁷¹ versus approximately 10 keV for X-rays, and the ratio of inelastic to elastic scattering events is roughly a factor of three smaller for electrons.⁹ This results in about 600–1000 times less energy deposited in the sample per unit signal for electrons compared to X-rays. Furthermore, electrons scatter with 10^4 greater probability than X-rays, allowing imaging with a much lower electron density. Yet, electrons are electrically charged fermions, which (as discussed in the sections on photoelectron guns and resolution) limits the fluence at the sample and both the spatial and temporal resolution. The challenge of pulsed electron microscopy will be to obtain a pulse of sufficient quality and electron density to obtain a high-resolution image, which is also short enough to avoid damage to the sample during the pulse.

Hydrodynamic modeling of single particle diffractive imaging with electrons (in analogy with the X-ray method above)⁷⁸ indicates that 4 Å resolution is possible for a molecule of 10 nm radius with 100 keV electrons at a density of 10^7 electrons/(100 nm)², with a pulse width of 2 ps. Using the same model for the X-ray case and sufficient photon density to achieve the same resolution, an X-ray pulse of 2 fs is required. These results are reasonably consistent with expectations, giving a difference in the maximum acceptable pulse length of 1000, exactly the ratio of deposited energy per unit signal between the X-ray and electron cases. These single particle models consider “naked” molecules only where the damage mechanism is Coulomb explosion. For the X-ray case, damage by this mechanism is the result of ionized electrons exiting the sample and leaving a net positive charge, resulting in rapid sample distortion due to electrostatic repulsion. Charging is much less severe for an electron probe due to the much lower kinetic energy of ionized secondary electrons. Also, it is likely that damage will occur on a slower time scale in solution, where the dispersion of molecules will be impeded by the presence of solution and secondary electron transit will be significantly reduced, resulting in less charging. In this case, the primary damage mechanism may be radical chemistry and diffusion, rather than Coulomb expansion. Although a 2-ps electron pulse width is probably beyond the capability of DTEM using standard electron guns, sub-100-ps pulse widths are not inconceivable at the required electron density. Attaining the necessary spatial and temporal coherence for diffractive imaging using such a pulse will probably be very difficult but perhaps within the reach of foreseeable improvements in electron gun technology.

Although atomic resolution may be challenging, electron gun modifications to shorten the pulses into the 10s of picoseconds may well allow very efficient imaging of protein complexes at sufficient resolution to significantly contribute to our understanding of their structure and dynamics. Furthermore, the RF accelerator community is aware of the potential for ultrashort, high brightness electron guns in electron microscopy.³⁰ Such sources do not currently exist, but if these were developed, they could very significantly improve our ability to rapidly image biomolecules.

Currently, there is very little information directly concerning biosample radiolysis damage at the very high current densities that will be found in DTEM. Although simulations such as those described above can provide us with some insight into the damage process, they do not consider the effect of an aqueous solution surrounding the sample. It is likely that a surrounding solution will significantly alter the damage process and may impede the physical diffusion of a damaged biosample long enough to obtain a high-resolution image with an electron pulse that is longer than 2ps. Given the available information, it is difficult to speculate. Nonetheless, here we attempt to approach the problem based on what is known.

First, we consider standard TEM. In some cases, the difficulty in imaging biomaterials may be overcome with standard TEM via any of three strategies, which all depend on the mitigation of damage: increasing sample contrast (via, for example, staining), increasing the sample damage threshold (as in cryoelectron microscopy⁷⁵), or reducing the dose per unit sample (as in diffraction and cryoEM) with signal averaging over many samples. Of these techniques, cryoEM most closely approximates imaging of biosamples in a native environment, since this method does not require contrast enhancement. In cryoEM, a sample is flash frozen such that the embedding ice is vitreous rather than crystalline.⁷⁹ This minimally distorts the sample and reduces diffractive background signal in the image. In addition, cooling of the sample to cryogenic temperatures increases the radiation tolerance by factors of thousands⁷⁵ with respect to image-manifested damage. Although irradiation may have caused the conditions under which sample damage would occur, the physical diffusion that would make this apparent in an image does not occur when the sample is frozen in place.⁷⁴ This result indicates that molecular diffusion, and temperature and phase-dependent chemical reactions play a role in damage,^{74,75} and that damage resulting from direct atomic displacement is insignificant at the doses (1–10 electrons/Å²) used in cryoEM.

To further establish the relevance of diffusion in damage, an early time diffusional dependence is indicated by electron energy loss experiments that demonstrate an orders of magnitude increase in the damage threshold with *reduced* excitation volume at equivalent or greater irradiation density. Varlot et al.⁸⁰ observed a change in damage threshold of 10⁴ between a 100 nm and 0.7 nm radius cylindrical excitation volume irradiating polymers. Their analysis did not attribute the change in damage threshold to temperature effects and they speculated that dif-

fusion of damaged molecules (e.g., free radicals) might explain the effect. Siangchaew et al.⁸¹ also observed a substantial increase in the damage threshold of polyethylene using a small electron probe. They speculate that the increase in damage threshold is due to fast secondary electrons (of energy >50 eV) exiting the 100-nm-diameter probe region. These electrons create damage collateral to the probe region, but leave the probe region relatively pristine. Egerton et al.⁷⁴ disagree with this interpretation, asserting that (for polyethylene) the number, range, and energy of fast secondary electrons (FSEs) should result in 75% of the FSE energy being deposited within 2 nm of the probe beam. Slow electrons have longer mean free paths and may exit a small probe region,⁸² but it is not known how damaging these low-velocity electrons would be to organic samples.

Diffusion also plays a primary role in radiolysis damage. In an aqueous environment, radiolysis damage is complex and involves many disparate processes. The physical excitation of secondary electrons through inelastic scattering is followed by electron thermalization and diffusional transport, and trapping occurring over a time scale on the order of 1 ps, along with the initial generation of radicals. This is followed by equilibrium with nuclei, which occurs in a few picoseconds. At longer time scales, chemistry begins to occur. Ionization and energy deposition in the first stage result in a transient distribution of radicals and other species (including H^+ , OH , e_{aq} , and H_2O_2 in water), which then diffuse and react with solutes and each other, reaching a homogeneous chemistry limit within micro- to milliseconds.^{83,84} With respect to solute interactions, pulse radiolysis studies indicate that the primary damage mechanism in DNA is attack by radiation-created hydroxyl radicals⁸⁵ and based on simulations, the time scale of this process is thought to be hundreds of femtoseconds,⁸⁶ assuming low deposited energy density.

These data seem to indicate that the time scale for image manifested damage of a biomolecule in aqueous solution would be no faster than a few picoseconds, possibly much longer. Nonetheless, these data do not directly address the question. Although pulse radiolysis studies have provided a significant amount of information,⁸⁷ this information generally concerns the time domain behavior of irradiated water (cf. time domain spectroscopy) and stoichiometric chemistry. Such data and corroborating simulations^{83,84} can be used to aid in the understanding of the diffusion of energy and radiolytic species in water, but ultimately they do not tell us about structural damage in solutes, particularly within a few picosecond time scale. More specific experimental studies of biomolecules such as DNA address damage, but again, this typically provides information regarding the chemistry of damage (typically via radiation-induced radicals) and the functional viability of irradiated molecules without reference to detailed structure.^{85,88} Furthermore, the energy deposition used in pulse radiolysis studies is much lower (~ 10 Gr)⁸⁹ than what is used in TEM ($>10^7$ Gr),⁷⁵ and the excitation volume in TEM is correspondingly much smaller. As a result, transport and relaxation of secondary electrons⁷¹ and high-dose rate effects⁹⁰ will probably play a significant role in the

damage process in DTEM. These issues are not as significant for continuous irradiation in conventional TEM since thermal diffusion occurs on a faster time scale than the average time between electrons transiting the sample. The instantaneous energy density in the DTEM will be much higher, and these data do not provide information in this regime.

Finally, we note that there is actually some empirical evidence that damage will be significantly reduced by the use of pulsed electron exposure. Fryer⁹¹ observed damage thresholds of $90\text{--}100 e/\text{\AA}^2$ using 10- to 100-ms pulses to image monolayer films of aromatic hydrocarbons. Using continuous exposure, the damage threshold for these samples was $3\text{--}4 e/\text{\AA}^2$, comparable to the damage threshold for proteins in cryoEM. This 22–25% improvement was observed for much longer pulses than we will use in the DTEM and points in the direction of trying even shorter pulses.

Unfortunately, no currently available data directly address the question of resolution smearing in DTEM due to damage from short pulses of electrons. As far as damage is concerned, currently available simulations and empirical results do not unambiguously exclude the possibility of molecular bioimaging with short electron pulses, but they do not clearly indicate that it is possible, either. As with many aspects of DTEM, the development and use of the instrument will ultimately provide these answers.

4 DTEM Applications

We include a brief historical review that attempts to put electron imaging in the context of work on electron diffraction. For the purpose of this discussion, we limit the review to time resolutions of microseconds or less. Figure 5–10 provides a summary. In the mid-1970s, Bostanjoglo et al.^{16,92} developed a technique to produce stroboscopic images in TEM using beam blanking. In these pump-probe studies, ultrasonics were used to pump samples that were studied stroboscopically. This led to a number of papers studying fast transitions.^{93–101}

Stroboscopic diffraction studies were carried out in the early 1980s.^{102–104} Real-time gas-diffraction experiments were reported in 1984.¹⁰⁵ The first picosecond time-resolved structural studies in the solid state occurred in the early 1980s with a streak camera used as the basis of a diffraction instrument.^{1,3} This led to the first picosecond time-resolution observation of a phase transition in Al.² In the late 1980s, pulsed photocathode,¹⁰⁶ multiframe movies,¹⁰⁷ and dynamic imaging were demonstrated at TU Berlin.¹⁰⁸ In the 1990s there was work in gas-phase energy transfer,¹⁰⁹ superheating of lead, and structural dynamics.¹¹⁰ The techniques were applied to complex transient phenomena including dynamics of cyclohexadiene,⁶ photo-dissociation,¹¹¹ direct observation of ultrafast thermal lattice expansion (Ag) on picosecond time scales,¹¹² melting of amorphous Si,⁷ phase explosion in metal films,¹¹³ and melting of Al.¹¹⁴

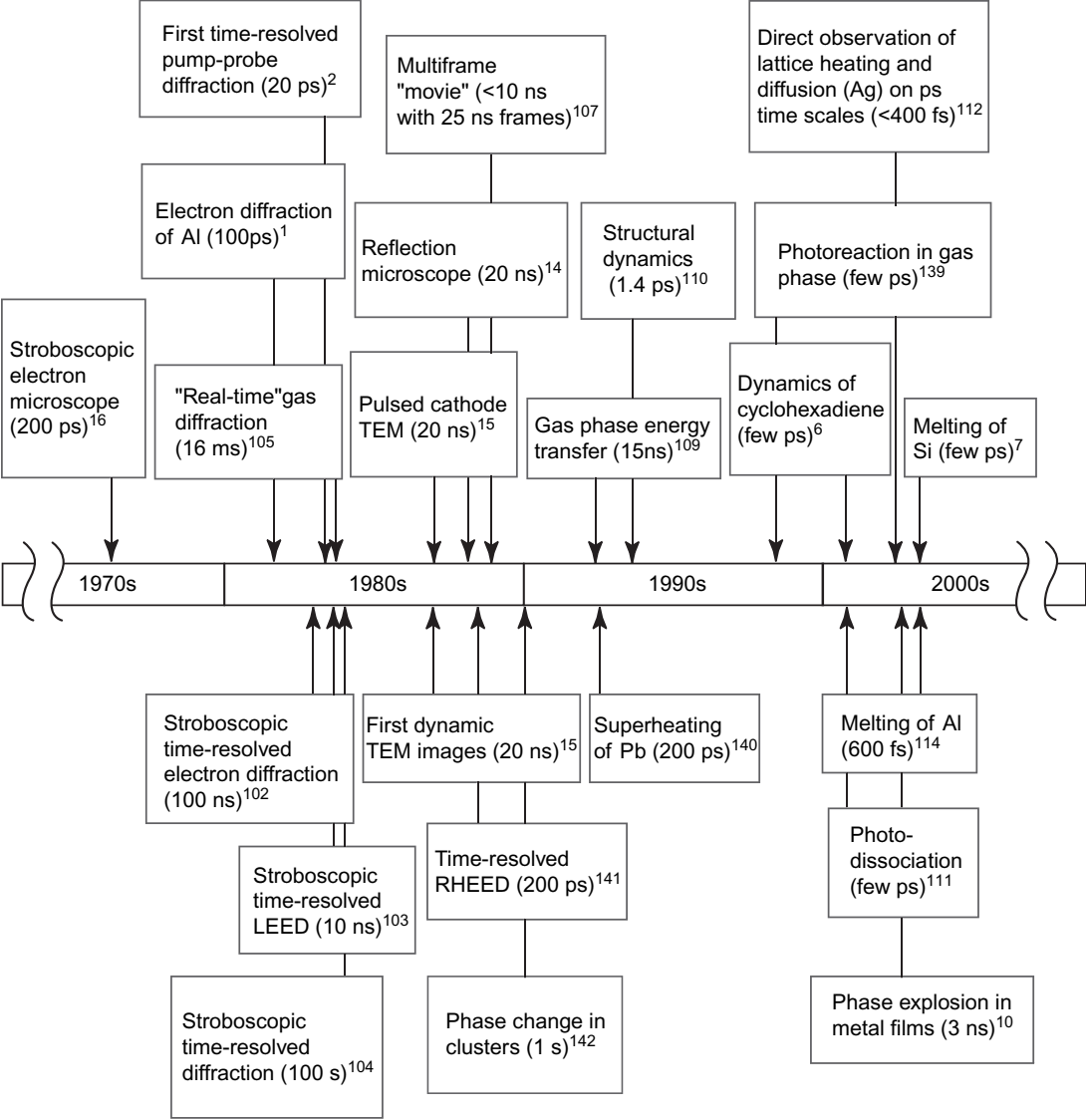


Figure 5–10. Timeline showing the parallel development of ultrafast electron diffraction and DTEM technologies.

TEM with high time resolution began at TU Berlin in the 1970s with stroboscopic imaging of periodic processes using a periodically deflected electron beam.^{16,115} In this configuration, time resolutions of 200ps were achieved for periodic processes up to 100MHz.¹¹⁶ This technique was applied to ultrasonically driven disruption of crystals and magnetoelastic effects^{92,115} and magnetic-field-induced oscillations of the domain magnetization, of domain walls, and of their substructures.^{98,116,117} After 1980, the instrument was modified to study nonperiodic processes. That instrument operated in three modes:

1. Selected-area image-intensity streaking provided time resolution of 3 ns and selected areas ≥ 500 nm.⁹⁴ This was applied to studies of electron^{94,95,99} and laser-induced^{118,119} crystallization of metals and semiconductors, thermocapillary oscillations,^{53,120} and solidification of laser-induced melts.^{53,69,120}

2. Streaked imaging was used to study solidification,⁵³ flow,^{53,121,122} and evaporation^{121,122} of laser-pulsed samples.

3. High-speed imaging and three-frame movies with temporal/spatial resolutions of 10 ns/200 nm, interpulse spacing >20 ns,¹⁰ and diffraction with time resolution of 10 ns were demonstrated.¹⁰

Several techniques were employed to achieve the high time resolution, including gating the electron detector,^{108,122,123} rapid beam blanking,¹²⁴ laser-induced thermal emission,^{27–29,69,125} and finally laser-induced photoemission.^{10,126} Nanosecond time resolution studies, using the photoemission source, were carried out on hydrodynamic instabilities caused by chemocapillary, thermocapillary,¹²⁷ and mechanical stresses.¹²⁸ Phase explosion and plasma formation of highly superheated metals were studied at the same time resolution.^{113,129}

5 Future

5.1 Improving Time Resolution

5.1.1 Pulse Compression

To recap, the electron pulse duration at the sample is equal to the single-shot time resolution of the DTEM. At present this is essentially equal to the cathode laser pulse duration and is dictated by the experiment's fluence requirements and the electron gun brightness. However, many methods exist for longitudinally compressing an electron pulse, many of them having been developed by the accelerator community for high-voltage RF electron guns. These involve passing the pulse through specialized elements such as a microwave cavity operating at a carefully chosen phase,^{58,130,131} or static magnetic fields in a chicane,^{132–134} α ,¹³⁵ or other pulse-compressing¹³⁶ arrangement. These devices are routinely used to compress multimegavolt pulses of $\sim 10^{10}$ electrons down to picosecond or shorter durations, with normalized emittance values ~ 1 μ m radian and $\sim 1\%$ energy spreads. The desired DTEM parameter space is very different, of order 200 keV, 10^8 electrons, less than 0.1- μ m radian emittance, and an energy spread of order one part in 10^5 . Significant work will be required to adapt the technology, possibly involving fundamental redesigns and entirely new ideas for pulse generation and compression. The outcome of research in the infant field of DTEM pulse compression is quite impossible to predict at this time; future DTEM electron guns might look almost nothing like the current designs. Pulse compression may be the only way to dig deeply into the subnanosecond regime, however. Without increasing the effective brightness of the electron source by crowding the

electrons into a shorter pulse, there are many measurements (requiring a combination of high fluence, small convergence angle, and high time resolution) that will simply be impossible.

Even then, there will be fundamental limits. Electrons are fermions, and can be compressed to a finite density only in six-dimensional phase space.⁶⁴ The best electron pulses produced to date are still roughly four orders of magnitude short of this quantum degeneracy limit, as they have been for decades.⁶⁴ Passive transformations to the electron pulse (e.g., acceleration, lens action, and most compression methods) can change only the shape of the phase space envelope, and not its density⁶⁴ (also see Reiser,¹⁹ pp. 62ff). Thus every compression method that does not actually refrigerate the electron pulse will involve trading off one design parameter against another.

5.2 Relativistic Beams

An alternate method to achieve high time resolution is to increase the accelerating voltage of the electrons. RF guns with thermionic cathodes have been used for many years for injection into storage rings⁸³ and RF guns with photocathodes as sources for free-electron lasers (FELs).⁸⁴ Currently photocathode rf guns are the brightest electron gun sources available.⁸⁵ Bright beams can be generated due to the high rf fields (typically 100 MV/m or more) used to accelerate the beam over a short distance. The beams become relativistic in a few centimeters, quickly reducing the space-charge force repulsion. (Space-charge force is reduced for high-energy beams. Several books derive the space-charge equations for both longitudinal and transverse space charge showing the force varies as the Lorentz factor γ^{-2} .^{137,138}) The particles are typically accelerated to ~5–6 MeV depending on the rf amplitude and the electron beam launch phase.

It is conceivable that such a gun could be fitted to an electron optical column suitably engineered for operation with 5–6 MeV electrons. This would, however, require large-scale installations similar to previous HVEM facilities. Two significant differences would be that the accelerator would be less than 10% of the size of previous HVEMs and that there would be no need for the elaborate and costly vibration isolation schemes that were needed. Further, because the duty cycle of the pulsed instruments is small compared with CW instruments, it is possible that some of the radiation shielding requirements may be relaxed in pulsed machines.

6 Conclusions

It appears that the growing interest in ultrafast science using optical and X-ray methods will spawn a new group of investigations using electrons. The potential for doing ultrafast diffraction experiments is already well established. The scientific case for extending this to ultrafast imaging with electrons is compelling. Technical hurdles remain to be overcome but it is clear that imaging will play a role certainly for experiments in which nanometer spatial resolution and nanosecond

time resolution are required. It is also clear that experiments at time resolutions previously inaccessible (nanosecond to millisecond) in which space charge limitations are not a factor will also benefit from this technology. Extension of this technology to femtosecond time resolution and atomic spatial resolution is a worthy goal and conceptually achievable. The benefits that accrue from going to relativistic accelerating voltages may revive interest in high-voltage electron microscopes with pulsed electron sources.

Acknowledgments. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory and supported by the Office of Science, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, of the U.S. Department of Energy under contract No. W-7405-Eng-48.

References

1. Mourou, G. and Williamson, S. (1982). Picosecond electron-diffraction. *Appl. Phys. Lett.* **41**, 44–45.
2. Williamson, S., Mourou, G. and Li, J.C.M. (1984). Time-resolved laser-induced phase-transformation in aluminum. *Phys. Rev. Lett.* **52**, 2364–2367.
3. Williamson, S. and Mourou, G. (1982). Electron-diffraction in the picosecond domain. *Appl. Phys. B-Photo.* **28**, 249–250.
4. Ewbank, J.D., Schafer, L., Paul, D.W. and Benston, J. (1984). Real-time data acquisition for gas electron-diffraction patterns. *Abstr. Pap. Am. Chem. S.* **188**, 97-Anyl.
5. Elsayed-Ali, H.E. and Herman, J.W. (1990). Picosecond time-resolved surface-lattice temperature probe. *Appl. Phys. Lett.* **57**, 1508–1510.
6. Dudek, R.C. and Weber, P.M. (2001). Ultrafast diffraction imaging of the electrocyclic ring-opening reaction of 1,3-cyclohexadiene. *J. Phys. Chem. A.* **105**, 4167–4171.
7. Ruan, C.Y., Vigliotti, F., Lobastov, V.A., Chen, S.Y. and Zewail, A.H. (2004). Ultrafast electron crystallography: Transient structures of molecules, surfaces, and phase transitions. *Proc. Natl. Acad. Sci. USA* **101**, 1123–1128.
8. Ruan, C.Y., Lobastov, V.A., Vigliotti, F., Chen, S.Y. and Zewail, A.H. (2004). Ultrafast electron crystallography of interfacial water. *Science* **304**, 80–84.
9. Siwick, B.J., Dwyer, J.R., Jordan, R.E. and Miller, R.J.D. (2004). Femtosecond electron diffraction studies of strongly driven structural phase transitions. *Chem. Phys.* **299**, 285–305.
10. Dömer, H. and Bostanjoglo, O. (2003). High-speed transmission electron microscope. *Rev. Sci. Instrum.* **74**, 4369–4372.
11. Bostanjoglo, O. (2002). High-speed electron microscopy. *Adv. Imag. Elect. Phys.* **121**, 1–51.
12. Bostanjoglo, O., Elschner, R., Mao, Z., Nink, T. and Weingartner, M. (2000). Nanosecond electron microscopes. *Ultramicroscopy* **81**, 141–147.
13. Bostanjoglo, O. and Weingartner, M. (1997). Pulsed photoelectron microscope for imaging laser-induced nanosecond processes. *Rev. Sci. Instrum.* **68**, 2456–2460.
14. Bostanjoglo, O. and Heinrich, F. (1988). A laser pulsed high emission thermal electron-gun. *Inst. Phys. Conf. Ser.* 105–106.

15. Bostanjoglo, O., Tornow, R.P. and Tornow, W. (1987). Nanosecond transmission electron-microscopy and diffraction. *J. Phys. E. Sci. Instrum.* **20**, 556–557.
16. Bostanjoglo, O. and Rosin, T. (1976). Stroboscopic study on ultrasonic activity in electron-microscope. *Mikroskopie*. **32**, 190.
17. Dömer, H. and Bostanjoglo, O. (2003). High-speed transmission electron microscopy of laser-pulsed metal films, splashing by a hydrodynamic instability. *Microsc. Microanal.* **9**, 358–359.
18. Dömer, H. and Bostanjoglo, O. (2002). Stress effects in laser-pulsed chromium films tracked by nanosecond transmission electron microscopy. *Adv. Eng. Mater.* **4**, 623–625.
19. Reiser, M. (1994). Theory and Design of Charged Particle Beams. New York: Wiley, pp. xix, 607.
20. Reed, B. (2004). Propagation dynamics of femtosecond electron packets and relativistic effects in ultrafast electron microscopy (UEM). (First National Lab and University Alliance Workshop on Ultrafast Electron Microscopies, Pleasanton, CA).
21. Faruqi, A.R. and Subramaniam, S. (2000). CCD detectors in high-resolution biological electron microscopy. *Q. Rev. Biophys.* **33**, 1–27.
22. Faruqi, A.R. and Andrews, H.N. (1997). Cooled CCD camera with tapered fibre optics for electron microscopy. *Nucl. Instrum. Meth. A.* **392**, 233–236.
23. Qian, W., Scheinfein, M.R. and Spence, J.C.H. (1993). Brightness measurements of nanometer-sized field-emission-electron sources. *J. Appl. Phys.* **73**, 7041–7045.
24. Bluem, H., Todd, A.M.M., Cole, M.D., Rathke, J. and Schultheiss, T. (2003). Recent advances in high-brightness electron guns at AES. *Nucl. Instrum. Meth. A.* **507**, 215–219.
25. Dowell, D.H., Bolton, P.R. and Clendenin, J.E. (2003). Slice emittance measurements at the SLAC gun test facility. *Nucl. Instrum. Meth. A.* **507**, 327–330.
26. HerreraGomez, A., Vergara, G. and Spicer, W.E. (1996). Physics of high-intensity nanosecond electron source: Charge limit phenomenon in GaAs photocathodes. *J. Appl. Phys.* **79**, 7318–7323.
27. Schafer, B. and Bostanjoglo, O. (1992). Laser driven thermionic electron-gun. *Optik*. **92**, 9–13.
28. Bostanjoglo, O., Heinrich, F. and Wunsch, F. (1990). Performance of a Laser-Pulsed Thermal Electron Gun. Proc. 12th Intl Congr Electron Microsc. San Francisco, pp. 124–125.
29. Bostanjoglo, O. and Heinrich, F. (1987). Producing high-current nanosecond electron pulses with a standard tungsten hairpin gun. *J. Phys. E. Sci. Instrum.* **20**, 1491–1493.
30. Lewellen, J.W. (2004). Overview of High-Brightness Proceedings of LINAC 842–846 (JACOW, Lübeck).
31. Brau, C.A. (2000). What brightness means. In: Rosenzweig J, Travish G, Serafini L, eds. The Physics and Applications of High Brightness Electron Beams. Proceedings of the ICFA Workshop. World Scientific, pp. 1–27.
32. Michaelson, H.B. (1977). Work function of elements and its periodicity. *J. Appl. Phys.* **48**, 4729–4733.
33. Michelato, P. (1997). Photocathodes for RF photoinjectors. *Nucl. Instrum. Meth. A.* **393**, 455–459.
34. Baum, A.W., Spicer, W.E., Pease, R.F., Castello, K.A. and Aebi, V.W. (1995). Negative electron affinity photocathodes as high performance electron sources. *SPIE* **2522**, 208–212.

35. Togawa, K., Nakanishi, T. and Baba, T. (1998). Surface charge limit in NEA superlattice photocathodes of polarized electron source. *Nucl. Instrum. Meth. A* **414**, 431–445.
36. Saka, T., Kato, T. and Nakanishi, T. (2000). Spinpolarized electron sources with GaAs-GaAsP superlattices for surface analyses. *Surf. Sci.* **454**, 1042–1045.
37. Luh, D.-A., Brachmann, A., Clendenin, J.E., Desikan, T., Garwin, E.L., Harvey, S., Kirby, R.E., Maruyama, T., Prescott, C.Y. and Prepost, R. (2003). Recent polarized photocathode R&D at SLAC. *AIP Conf. Proc.* **675**, 1029–1033.
38. Muggli, P., Brogle, R., Jou, S., Doerr, H.J., Bunshah, R.F. and Joshi, C. (1996). Photoemission from diamond and fullerene films for advanced accelerator applications. *IEEE T. Plasma. Sci.* **24**, 428–438.
39. Modukuru, Y., Thachery, J., Tang, H., Malhotra, A., Cahay, M. and Boolchand, P. (2001). Growth and characterization of rare-earth monosulfides for cold cathode applications. *J. Vac. Sci. Technol. B* **19**, 1958–1961.
40. Yada, K. (1986). Researches of cathode materials for thermionic emission. In: Imura T, Maruse S, Suzuki T, eds. Proc 11th Int Congr Electron Microscopy, Vol. **1**, pp. 227–228.
41. Watari, F. and Yada, K. (1986). Photoemission from LaB₆ cathode using an excimer laser. In: Imura T, Maruse S, Suzuki T, eds. Proc 11th Int Congr Electron Microscopy, Vol. **1**, pp. 261–262.
42. May, P.G., Petkie, R.R., Harper, J.M.E. and Yee, D.S. (1990). Photoemission from thin-film lanthanum hexaboride. *Appl. Phys. Lett.* **57**, 1584–1585.
43. Travier, C. (1994). An introduction to photo-injector design. *Nucl. Instrum. Meth. A* **340**, 26–39.
44. Girardeau-Montaut, J.P., Girardeau-Montaut, C., Afif, M., Perez, A. and Moustazis, S.D. (1995). Enhancement of photoelectric-emission sensitivity of tungsten by potassium-ion implantation. *Appl. Phys. Lett.* **66**, 1886–1888.
45. Orloff, J. (1986). Development of the ZrO/W thermal field cathode. In: Imura T, Maruse S, Suzuki T, eds. Proc 11th Int Congr Electron Microscopy, Vol. **1**, pp. 211–214.
46. Nassisi, V. and Perrone, M.R. (1999). Generation and characterization of high intensity electron beams generated from rough photocathodes. *Rev. Sci. Instrum.* **70**, 4221–4224.
47. Boussoukaya, M., Bergeret, H., Chehab, R., Leblond, B. and Leduff, J. (1989). High quantum yield from photofield emitters. *Nucl. Instrum. Meth. A* **279**, 405–409.
48. Kawamura, Y., Jeong, Y.U., Akiyama, Y., Kubodera, S., Midorikawa, K. and Toyoda, K. (1993). Generation of high-current photoelectrons using a sub-picosecond ultraviolet-laser. *Jpn. J. Appl. Phys. Part 2-Lett* **32**, L297–L299.
49. Korte, F., Serbin, J. and Koch, J. (2003). Towards nanostructuring with femtosecond laser pulses. *Appl. Phys. A.—Materials Sci. Process* **77**, 229–235.
50. Bergeret, H., Boussoukaya, M., Chehab, R., Leblond, B. and Leduff, J. (1991). Short pulse photoemission from a dispenser cathode. *Nucl. Instrum. Meth. A* **301**, 389–394.
51. Höbel, F., Dömer, H., Gartenschläger, C., Gernert, U., Nink, T. and Bostanjoglo, O. (2003). A rugged cathode for photoelectron and DC thermionic operation. In: Gemming T, Lehmann M, Lichte H, Wetzig K, eds. Proc Microsc Conf Microsc Microanal. Volume 9 Suppl. 3, 150–151.
52. Dömer, H. and Bostanjoglo, O. (2003). High-speed transmission electron microscope. *Rev. Sci. Instrum.* **74**, 4369–4372.
53. Bostanjoglo, O. and Nink, T. (1996). Hydrodynamic instabilities in laser pulse-produced melts of metal films. *J. Appl. Phys.* **79**, 8725–8729.

54. Bostanjoglo, O. and Nink, T. (1997). Liquid motion in laser pulsed Al, Co and Au films. *Appl. Surf. Sci.* **110**, 101–105.
55. Bradley, D.J., Roddie, A.G. and Sibbett, W. (1975). Picosecond X-ray chronoscopy. *Opt. Commun.* **15**, 231–236.
56. Hansen, S., Hereward, H.G., Hofmann, A., Hubner, K. and Myers, S. (1975). Effects of space-charge and reactive wall impedance on bunched beams. *IEEE T. Nucl. Sci.* **NS22**, 1381–1384.
57. Neuffer, D. (1979). Longitudinal motion in high-current ion-beams—self-consistent phase-space distribution with an envelope equation. *IEEE T. Nucl. Sci.* **26**, 3031–3033.
58. Qin, H. and Davidson, R.C. (2002). Longitudinal drift compression and pulse shaping for high-intensity particle beams. *Phys. Rev. Spec. Top-Ac* **5**, 034401.
59. Ben-Yakar, A. and Hanson, R.K. (2002). Ultra-fast-framing schlieren system for studies of the time evolution of jets in supersonic crossflows. *Exp. Fluids* **32**, 652–666.
60. Kosonocky, W.F., Yang, G. and Kabra, R.K. (1997). 360×360 element three-phase very high frame rate burst image sensor: Design, operation, and performance. *IEEE T. Electron. Dev.* **44**, 1617–1624.
61. Reich, R.K., Rathman, D.D. and O'Mara, D.M. (2003). High-speed, electronically shuttered solid-state imager technology (invited). *Rev. Sci. Instrum.* **74**, 2027–2031.
62. Minor, A.M., Lilleodden, E.T., Stach, E.A. and Morris, J.W. (2004). Direct observations of incipient plasticity during nanoindentation of Al. *J. Mater. Res.* **19**, 176–182.
63. Domer, H. and Bostanjoglo, O. (2003). Phase explosion in laser-pulsed metal films. *Appl. Surf. Sci.* **208**, 442–446.
64. Rose, H. and Spehr, R. (1983). Energy broadening in high-density electron and ion-beams—the Boersch effect. *Adv. Electron. Electron. Phys. Suppl* **13C**, 475–530.
65. Kruit, P. and Jansen, G.H. (1997). Space charge and statistical Coulomb effects. In: Orloff J, ed. *Handbook of Charged Particle Optics*, pp. 275–318. (CRC Press, Boca Raton, FL).
66. Jansen, G.H. (1990). *Coulomb Interactions in Particle Beams*. (Academic Press, San Diego).
67. Cowley, J.M. (1992). Imaging. In: Buseck P, Cowley JM, Eyring LE, eds. *High-Resolution Transmission Electron Microscopy and Associated Techniques*, pp. 3–37. (Oxford University Press, Oxford).
68. Stadelmann, P. and Java, E.M.S. (2001). <http://cimewww.epfl.ch/people/stadelmann/jemswebsite/jems.html>. (Institution, Lausanne).
69. Bostanjoglo, O. and Otte, D. (1993). High-speed transmission electron-microscopy of laser quenching. *Mat. Sci. Eng. A.—Struct* **173**, 407–411.
70. Loeffler, K.H. (1969). Energy-spread generation in electron-optical instruments. *Z. Angew. Phys.* **27**, 145–149.
71. Egerton, R.F. (1986). *Electron Energy Loss Spectroscopy in the Electron Microscope*. (Plenum Press, New York).
72. Henderson, R. (2004). Realizing the potential of electron cryo-microscopy. *Q. Rev. Biophys.* **37**, 3–13.
73. Sun, K., Liu, J. and Browning, N.D. (2002). Correlated atomic resolution microscopy and spectroscopy studies of Sn(Sb)O₂ nanophase catalysts. *J. Catal.* **205**, 266–277.
74. Egerton, R.F., Li, P. and Malac, M. (2004). Radiation damage in the TEM and SEM. *Micron.* **35**, 399–409.
75. Henderson, R. (1990). Cryoprotection of protein crystals against radiation-damage in electron and X-ray-diffraction. *Proc. R. Soc. London Ser. B. Biol. Sci.* **241**, 6–8.

76. Schotte, F., Lim, M.H., Jackson, T.A., Smirnov, A.V., Soman, J., Olson, J.S., Phillips, G.N., Wulff, M. and Anfinrud, P.A. (2003). Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. *Science* **300**, 1944–1947.
77. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. and Hajdu, J. (2000). Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757.
78. Hau-Riege, S.P., London, R.A. and Chapman, H.N. (2004). Pulse requirements for electron diffraction imaging of single biological molecules: LLNL internal report.
79. Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, T.J., McDowell, A.W. and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228.
80. Varlot, K., Martin, J.M. and Quet, C. (1998). Physical and chemical changes in polystyrene during electron irradiation using EELS in the TEM: Contribution of the dielectric function. *J. Microsc. Oxford* **191**, 187–194.
81. Siangchaew, K. and Libera, M. (2000). The influence of fast secondary electrons on the aromatic structure of polystyrene. *Phil. Mag. A.* **80**, 1001–1016.
82. Emfietzoglou, D. and Moscovitch, M. (2002). Inelastic collision characteristics of electrons in liquid water. *Nucl. Instrum. Methods Phys. Res. B.* **193**, 71–78.
83. Cobut, V., Frongillo, Y., Patau, J.P., Goulet, T., Fraser, M.J. and Jay-Gerin, J.P. (1998). Monte Carlo simulation of fast electron and proton tracks in liquid water—I. Physical and physicochemical aspects. *Radiat. Phys. Chem.* **51**, 229–243.
84. Frongillo, Y., Goulet, T., Fraser, M.J., Cobut, V., Patau, J.P. and Jay-Gerin, J.P. (1998). Monte Carlo simulation of fast electron and proton tracks in liquid water—II. Nonhomogeneous chemistry. *Radiat. Phys. Chem.* **51**, 245–254.
85. Dizdaroglu, M. (1992). Measurement of radiation-induced damage to DNA at the molecular-level. *Int. J. Radiat. Biol.* **61**, 175–183.
86. Wu, Y.D., Mundy, C.J., Colvin, M.E. and Car, R. (2004). On the mechanisms of OH radical induced DNA-base damage: A comparative quantum chemical and Car-Parrinello molecular dynamics study. *J. Phys. Chem. A.* **108**, 2922–2929.
87. Zimbrick, J.D. (2002). Radiation chemistry and the radiation research society: A history from the beginning. *Radiat. Res.* **158**, 127–140.
88. Douki, T., Spinelli, S., Ravanat, J.L. and Cadet, J. (1999). Hydroxyl radical-induced degradation of 2'-deoxyguanosine under reducing conditions. 1875–1880.
89. Wardman, P. (1978). Application of pulse-radiolysis methods to study reactions and structure of biomolecules. *Rep. Prog. Phys.* **41**, 259–302.
90. Jay-Gerin, J.-P. (2004). Personal communication.
91. Fryer, J.R. (1987). The effect of dose-rate on imaging aromatic organic-crystals. *Ultramicroscopy* **23**, 321–327.
92. Bostanjoglo, O. and Rosin, T. (1977). Ultrasonically induced magnetic reversals observed by stroboscopic electron-microscopy. *Opt. Acta.* **24**, 657–664.
93. Bostanjoglo, O. and Horinek, W.R. (1983). Pulsed Tem—a new method to detect transient structures in fast phase-transitions. *Optik.* **65**, 361–367.
94. Bostanjoglo, O., Schlotzhauer, G. and Schade, S. (1982). Shaping trigger pulses from noisy signals and time-resolved TEM of fast phase-transitions. *Optik.* **61**, 91–98.

95. Bostanjoglo, O. (1982). Time-resolved TEM of pulsed crystallization of amorphous Si and Ge films. *Phys. Status. Solidi. A.* **70**, 473–481.
96. Bostanjoglo, O. and Hoffmann, G. (1982). Time-resolved TEM of transient effects in pulse annealing of Ge and Ge-Te films. *Phys. Status. Solidi. A.* **73**, 95–105.
97. Bostanjoglo, O. and Schlotzhauer, G. (1981). Impulse stimulated crystallization of Sb films investigated by time resolved TEM. *Phys. Status. Solidi. A.* **68**, 555–560.
98. Bostanjoglo, O. and Rosin, T. (1980). Resonance oscillations of magnetic domain-walls and Bloch lines observed by stroboscopic electron-microscopy. *Phys. Status. Solidi. A.* **57**, 561–568.
99. Bostanjoglo, O. and Liedtke, R. (1980). Tracing fast phase-transitions by electron-microscopy. *Phys. Status. Solidi. A.* **60**, 451–455.
100. Bostanjoglo, O. and Rosin, T. (1980). Stroboscopic TEM studies of high-frequency magnetic test materials. *Mikroskopie* **36**, 344.
101. Bostanjoglo, O. and Gemund, H.P. (1978). Low-temperature specimen holder for magnetization processes by means of short field pulse. *Mikroskopie* **34**, 188.
102. Ischenko, A.A., Golubkov, V.V., Spiridonov, V.P., Zgurskii, A.V., Akhmanov, A.S., Vabischevich, M.G. and Bagratashvili, N. (1983). A stroboscopical gas-electron diffraction method for the investigation of short-lived molecular-species. *Appl. Phys. B-Photo* **32**, 161–163.
103. Becker, R.S., Higashi, G.S. and Golovchenko, J.A. (1984). Low-energy electron-diffraction during pulsed laser annealing—a time-resolved surface structural study. *Phys. Rev. Lett.* **52**, 307–310.
104. Rood, A.P. and Milledge, J. (1984). Combined flash-photolysis and gas-phase electron-diffraction studies of small molecules. *J. Chem. Soc. Faraday Trans. 2* **80**, 1145.
105. Ewbank, J.D., Schafer, L., Paul, D.W., Benston, O.J. and Lennox, J.C. (1984). Real-time data acquisition for gas electron-diffraction. *Rev. Sci. Instrum.* **55**, 1598–1603.
106. Bostanjoglo, O., Endruschat, E., Heinrich, F., Tornow, R.P. and Tornow, W. (1987). Short-interval electron-microscopy and pulsed lasers. *Eur. J. Cell Biol.* **44**, 10.
107. Bostanjoglo, O., Kornitzky, J. and Tornow, R.P. (1989). Nanosecond double-frame electron-microscopy of fast phase-transitions. *J. Phys. E. Sci. Instrum.* **22**, 1008–1011.
108. Bostanjoglo, O., Tornow, R.P. and Tornow, W. (1987). Nanosecond-exposure electron-microscopy of laser-induced phase-transformations. *Ultramicroscopy* **21**, 367–372.
109. Ewbank, J.D., Luo, J.Y., English, J.T., Liu, R.F., Faust, W.L. and Schafer, L. (1993). Time-resolved gas electron-diffraction study of the 193-nm photolysis of 1,2-dichloroethenes. *J. Phys. Chem-Us* **97**, 8745–8751.
110. Dantus, M., Kim, S.B., Williamson, J.C. and Zewail, A.H. (1994). Ultrafast electron-diffraction. 5. Experimental time resolution and applications. *J. Phys. Chem-Us* **98**, 2782–2796.
111. Ihee, H., Lobastov, V.A., Gomez, U.M., Goodson, B.M., Srinivasan, R., Ruan, C.Y. and Zewail, A.H. (2001). Direct imaging of transient molecular structures with ultrafast diffraction. *Science* **291**, 458–462.
112. Cao, J., Hao, Z., Park, H., Tao, C., Kau, D. and Blaszczyk, L. (2003). Femtosecond electron diffraction for direct measurement of ultrafast atomic motions. *Appl. Phys. Lett.* **83**, 1044–1046.
113. Dömer, H. and Bostanjoglo, O. (2003). Phase explosion in laser-pulsed metal films. *Appl. Surf. Sci.* **208**, 442–446.

114. Siwick, B.J., Dwyer, J.R., Jordan, R.E. and Miller, R.J.D. (2003). An atomic-level view of melting using femtosecond electron diffraction. *Science* **302**, 1382–1385.
115. Bostanjoglo, O. and Rosin, T. (1976). Ultrasonically induced magnetoelastic effects observed by stroboscopic electron microscopy. In: Brandon DG, ed. Proc 6th Eur Congr Electr Microscopy, Vol. 1. Jerusalem, p. 360.
116. Bostanjoglo, O. and Rosin, T. (1980). Stroboscopic Lorentz TEM at 100 kV up to 100 MHz. In: Brederoo P, Boom G, eds. Proc 7th Eur Congr Electr Microscopy. The Hague, p. 88.
117. Bostanjoglo, O. and Rosin, T. (1980). Mass and relaxation time of domain walls in thin NiFe films from forced oscillations. *J. Magnetism. Magnetic Mat.* **15–18**, 1529–1539.
118. Bostanjoglo, O., Endruschat, E. and Tornow, W. (1986). Time-resolved TEM of laser-induced phase transitions in a-Ge and a-Si/Al films. *Mat. Res. Soc. Symp. Proc.* **71**, 345–350.
119. Bostanjoglo, O. and Endruschat, E. (1985). Kinetics of laser-induced crystallization of amorphous germanium films. *Phys. Status. Solidi. A.* **91**, 17–28.
120. Bostanjoglo, O. and Nink, T. (1996). Liquid motion in laser-pulsed Al, Co, and Au films. *Appl. Surf. Sci.* **109–110**, 101–105.
121. Bostanjoglo, O., Kornitzky, J. and Tornow, R.P. (1991). High-speed electron-microscopy of laser-induced vaporization of thin-films. *J. Appl. Phys.* **69**, 2581–2583.
122. Bostanjoglo, O. and Kornitzky, J. (1990). Nanoseconds double-frame and streak transmission electron microscopy. Proc 12th Int Congr Electr Microscopy. San Francisco, p. 180.
123. Bostanjoglo, O., Tornow, R.P. and Tornow, W. (1987). A pulsed image converter for nanosecond electron-microscopy. *Scan. Microsc.* 197–203.
124. Bostanjoglo, O., Niedrig, R. and Wedel, B. (1994). Ablation of metal-films by picosecond laser-pulses imaged with high-speed electron-microscopy. *J. Appl. Phys.* **76**, 3045–3048.
125. Bostanjoglo, O. and Otte, D. (1995). High-speed electron-microscopy of nanocrystallization in Al-Ni films by nanosecond laser-pulses. *Phys. Status. Solidi. A.* **150**, 163–169.
126. Nink, T., Galbert, F., Mao, Z.L. and Bostanjoglo, O. (1999). Dynamics of laser pulse-induced melts in Ni-P visualized by high-speed transmission electron microscopy. *Appl. Surf. Sci.* **139**, 439–443.
127. Nink, T., Mao, Z.L., and Bostanjoglo, O. (2000). Melt instability and crystallization in thin amorphous Ni-P films. *Appl. Surf. Sci.* **154**, 140–145.
128. Dömer, H. and Bostanjoglo, O. (2002). Laser ablation of thin films with very high induced stresses. *J. Appl. Phys.* **91**, 5462–5467.
129. Dömer, H. and Bostanjoglo, O. (2003). Relaxing melt and plasma bubbles in laser-pulsed metals. *J. Appl. Phys.* **94**, 6280–6284.
130. Steinhauer, L.C. and Kimura, W.D. (1999). Longitudinal space charge debunching and compensation in high-frequency accelerators. *Phys. Rev. ST-AB* **2**, 081301.
131. Serafini, L. (1996). Micro-bunch production with radio frequency photo-injectors. *IEEE T. Plasma. Sci.* **24**, 421–427.
132. Anderson, S.G., Rosenzweig, J.B., Musumeci, P. and Thompson, M.C. (2003). Horizontal phase-space distortions arising from magnetic pulse compression of an intense, relativistic electron beam. *Phys. Rev. Lett.* **91**, 074803.
133. Dowell, D.H., Adamski, J.L., Hayward, T.D., Johnson, P.E., Parazzoli, C.D. and Vetter, A.M. (1997). Results of the Boeing pulse compression and energy recovery. *Nucl. Instrum. Meth. A.* **393**, 184–187.

134. Zhou, F., Cline, D.B. and Kimura, W.D. (2003). Beam dynamics analysis of femtosecond microbunches produced by the staged electron laser acceleration experiment. *Phys. Rev. Spec. Top-Ac.* **6**, 054201.
135. Kung, P., Lihn, H.C., Wiedemann, H. and Bocek, D. (1994). Generation and measurement of 50-Fs (Rms) electron pulses. *Phys. Rev. Lett.* **73**, 967–970.
136. Kozawa, T., Kobayashi, T., Ueda, T. and Uesaka, M. (1997). Generation of high-current (1kA) subpicosecond electron single pulse. *Nucl. Instrum. Meth. A.* **399**, 180–184.
137. Chao, A. (1993). *Physics of Collective Beam Instabilities in High Energy Accelerators*. New York: Wiley. pp. x, 371.
138. Edwards, D.A. and Syphers, M.J. (1993). *An Introduction to the Physics of High Energy Accelerators*. New York: Wiley, pp. xii, 292.
139. Williamson, J.C., Cao, J.M., Ihee, H., Frey, H. and Zewail, A.H. (1997). Clocking transient chemical changes by Ultrafast electron diffraction. *Nature* **386**, 159–162.
140. Herman, J.W. and Elsayedali, H.E. (1992). Superheating of Pb(111). *Phys. Rev. Lett.* **69**, 1228–1231.
141. Elsayedali, H.E. and Mourou, G.A. (1988). Pico second reflection high-energy electron-diffraction. *Rev. Phys. Lett.* **52**, 103–104.
142. Bartell, L.S. and Dibble, T.S. (1990). Observation of the time evolution phase—changes in clusters. *J. Am. Chem. Soc.* **112**, 890–891.

In Situ Transmission Electron Microscopy

Frances M. Ross

1 A Working Definition of *in situ* Transmission Electron Microscopy

Since the earliest days of transmission electron microscopy, microscopists have realized the potential of microscopy for studying dynamic processes. Images recorded sequentially can be used to track the changes caused by deliberate actions, such as heating or straining, or uncontrolled processes, such as beam damage. The class of experiments where a specimen is changed or acted on while it remains under observation (i.e., *in situ* in the polepiece) is referred to as *in situ* microscopy. In a sense every TEM observation is an *in situ* experiment, since every specimen is affected by the electron beam to some extent. But the *in situ* experimenter aims to modify the specimen in a deliberate way and learn something from the results.

In the best *in situ* experiments, a controlled change is made to a specimen's environment, and this is correlated with the resulting change in its structure, measured using any of the imaging, analytical, or diffraction techniques available, or its electronic or mechanical properties, which can also be measured *in situ*. Preferably both the "input," in other words the change in sample environment, and the "output," or consequent change in structure or properties, are recorded simultaneously and quantitatively. Given sufficient care with artifacts, a quantitative understanding of a fundamental physical process can be obtained.

There are numerous advantages to performing experiments *in situ*. A single *in situ* experiment gives a continuous view of a process, so may take the place of multiple post-mortem measurements. A single heating experiment, for example, can provide information that would otherwise have to be extracted by examination of many samples which had been annealed to different temperatures or for different times. Because an *in situ* experiment is continuously recorded, it is easier to catch a transient phase or observe a nucleation event. *In situ* experiments can yield specific and detailed kinetic information, measuring

for example the motion of individual dislocations under known stress, or the growth rates of individual nanocrystals. Properties can be determined for well-characterized nanostructures, such as the conductivity of single nanotubes or the melting point of precipitates. Finally, growth experiments in particular provide a window into the behavior of materials under real processing conditions, since significant changes can occur if we remove a material from the growth chamber and perform analysis post-mortem.

Although *in situ* experiments provide unique information, it is at a cost of increased experimental complexity. Careful design of the specimen is necessary to minimize thin foil effects. Tests must be carried out to understand beam effects, and calibration of the applied stimulus is critical.

Given the complexity of some *in situ* experiments, what is remarkable and inspirational is the variety of materials and properties that have been studied *in situ*. In the vast majority of experiments, the "input" to the specimen ranges from simple beam heating to controlled sample heating, cooling, or straining; application of a voltage or a magnetic field; or even modification with a scanning probe tip. A specially designed sample holder may be used which could include a heater, electrical contacts, a tip, or a mechanical straining stage. Such experiments can be carried out in most microscopes, apart from those with the tiniest polepiece gaps, although a side entry design is most convenient for experiments involving feedthroughs. A second, less common class of experiments is based on changing the sample's environment by, for example, exposing it to a reactive gas or depositing another material onto it. For such studies, two experimental strategies are possible. Firstly, one can use a conventional microscope, but achieve environmental control in a modified sample holder in which the sample and reactive environment are enclosed in a region between two electron-transparent membranes. Alternatively, the microscope itself may be modified, for example by adding gas feedthroughs to the specimen area. The sample can then be exposed to the desired environment in a standard holder. For certain experiments involving reactive surfaces, a clean environment is important and the entire microscope must then be designed for ultra high vacuum (UHV). A UHV specimen region allows clean surfaces to be prepared (for example by heating) and then modified controllably in the polepiece. True UHV systems are relatively rare as they represent a large investment. They often include side chambers attached to the microscope in which other preparation or deposition treatments can be carried out *ex situ*.

In terms of collecting the "output" data, some *in situ* experiments require atomic resolution, while others use lower resolution strain or defect imaging, diffraction analysis, and occasionally elemental mapping. Data may be collected onto videotape or hard drive or as a series of still images, and in some cases high speed image acquisition may be necessary. Commonly, other measurements, such as sample temperature, gas pressure, electrical conductivity or applied force, are collected simultaneously and may be written onto the video tape or stored electronically with the images.

The following sections will describe some of the accomplishments of *in situ* microscopy in improving our understanding of the bulk properties and surface physics of materials. *In situ* microscopy has a rich history, but here we will focus on more recent experiments, mainly within the last decade, with the hope that this emphasis will capture the ongoing excitement of this quickly developing field. Along the way we will discuss experimental requirements for *in situ* experiments and the recognition and elimination of artifacts. We hope to show how widely *in situ* microscopy has enhanced our understanding of phenomena associated with phase transformations, crystal growth, electrical and mechanical properties, magnetism and ferroelectricity, implantation and beam effects, and even processes which take place in the liquid phase. Improvements such as the larger polepiece gap made possible by aberration correction, more sophisticated data analysis techniques, and enhanced abilities to fabricate specimens of controlled geometry, promise to extend the possibilities of *in situ* techniques even further.

2 Phase Transformations

The largest body of work accomplished using *in situ* TEM techniques has been in the area of phase transformations: melting, crystallization, transformations between crystal structures, and the formation of new phases by solid state diffusion. An understanding of such transformations is scientifically interesting and technologically essential in, for example, the processing of alloys or the development of new materials having extreme hardness or superplastic, magnetic or shape memory properties. *In situ* TEM has provided detailed information on the mechanism, kinetics, and structures produced during many phase transformations, both in the bulk and in nanoscale volumes. Microscopy is well suited for such studies because its high resolution allows atomic motions to be visualized, and diffraction can identify the phases present under changing conditions. Small precipitates or nuclei can be characterized and their evolution followed, and complex or incommensurate structures can be analyzed.

The requirements for these studies are usually simple, consisting of time-resolved imaging and a heating stage, although some experiments involve cooling, straining, or deposition. Images are commonly recorded at video rate (30 images per second) and the temperature can often be chosen to give a convenient reaction velocity. Heat is applied using the electron beam, or, more controllably, by heating a furnace or a resistive wire close to the sample. Alternatively, a heating current may be passed through the sample or the sample may be stuck onto a resistively heated wire. For qualitative results, such as determining reaction mechanisms, structural changes are observed during heating and cooling. For quantitative measurements the sample temperature must be measured accurately. This is a challenge since the temperature measured at the furnace by a thermocouple, say, may not be the same as the temperature at the region under observation. However, with some effort, careful calibrations have been achieved.

We first discuss transformations in bulk materials, then examine transformations in small volumes of material. These small volumes may be free-standing nanostructures or nanoparticles encapsulated in a matrix. A recurring theme will be the finding that small volumes do not transform under the same conditions as larger volumes, which is extremely important for the development of complex materials.

2.1 Crystallization, Melting and Grain Growth in Bulk Materials

2.1.1 Amorphization and Crystallization

The crystallization of amorphous materials is an interesting and important process which is uniquely suited to TEM analysis. Early, elegantly simple experiments involved the recrystallization of silicon, deposited as an amorphous thin film and then heated in cross section in a high resolution TEM. These experiments (Parker et al., 1986; Sinclair et al., 1987) showed the power of high resolution imaging at high temperature. The nucleation of crystallites was visualized, allowing estimation of the critical nucleus size, and the irregular progress of the reaction front was demonstrated, even though macroscopically the kinetics were consistent with a more continuous ledge mechanism. This pioneering work provided an atomic scale view of a bulk phase transformation, showing the start-stop motion we now expect for atomic scale processes. Using multilayer specimens to extend this approach to metal-mediated crystallization of Si, Ge, or C clearly demonstrated the mechanism for these reactions as well (Figure 6-1, Konno and Sinclair, 1992, 1995a, b, c; Sinclair et al., 2002). Si crystallization has now been so well studied, both *in situ* and *ex situ*, that it has actually been used as a calibration tool to measure the temperature in thin specimens (Hull and Bean, 1994; Stach et al., 1998a). An accurately calibrated temperature is essential in obtaining quantitative information, such as activation energy, for reactions carried out *in situ*. More recent crystallization studies have used plan view rather than cross sectional geometry, allowing many individual grains to be imaged. For example, the nucleation and growth rates of individual NiTi crystals during heating were found to be in agreement with the Johnson-Mehl-Avrami-Kolmogorov model (Lee et al., 2005), allowing grain size distributions to be predicted in this shape-memory alloy.

An interesting industrial application of this type of experiment, relevant to new types of information storage, is shown in Figure 6-2. Phase change materials such as GeSb and GeSbTe can store bits of information as amorphous areas embedded in crystalline regions. A high laser power is used to write amorphous spots, a medium power erases by recrystallizing, and a low power (or other measurement) reads the bits. *In situ* observations of crystallization have been made in films deposited on SiN membranes (Kooi et al., 2004; Kooi and De Hosson, 2004), free-standing films (Petford-Long et al., 1995) and actual compact disc materials (Kaiser et al., 2004). Beam heating shows nucleation and growth kinetics (Figure 6-2B), while more controlled experiments using a heating stage measure activation energies (Figure 6-2A). SbO_x, another potential phase change material, has been examined *in*

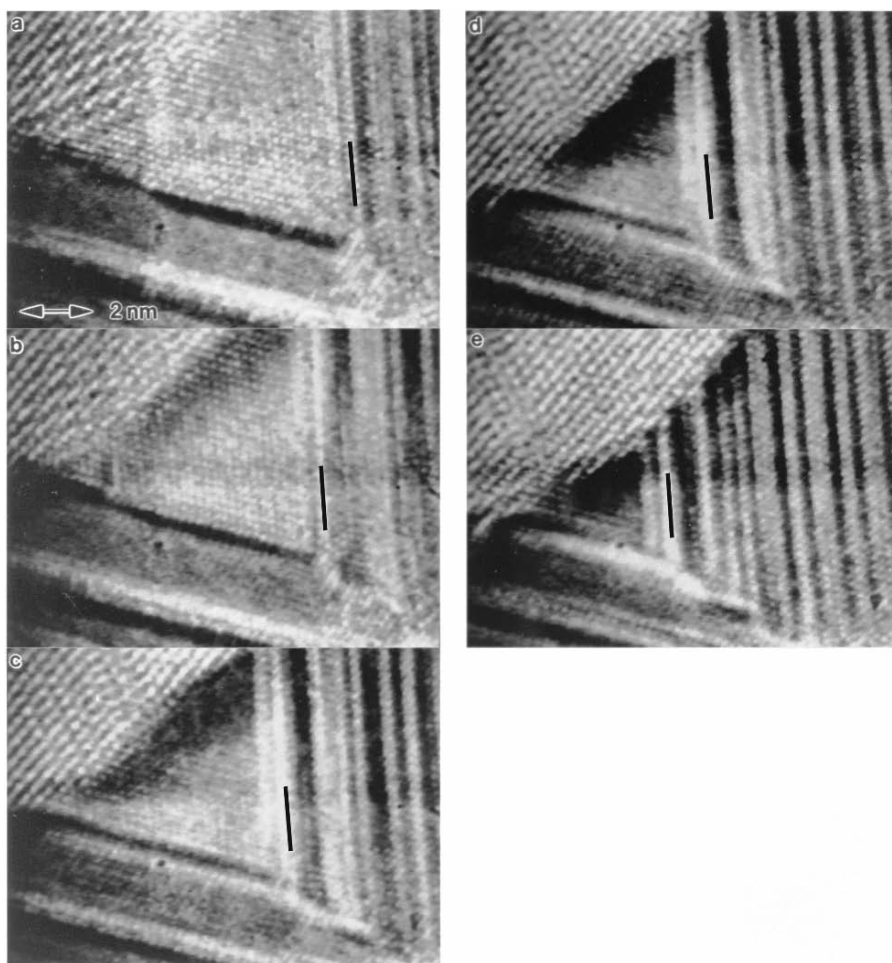


Figure 6-1. Metal-induced crystallization. *In situ* high-resolution images recorded during the Ag-mediated crystallization of Ge at 250°C. The time between frames is 8 s. The Ag is the faulted region in the center and the crystalline Ge is in the upper left. The Ag crystal appears to migrate toward the amorphous Ge region but the faults remain fixed (one fault is indicated by a line). The inference is that Ge is supplied by diffusion through the Ag lattice, and the net motion of the Ag is caused by counterdiffusion of Ag atoms. The lack of any amorphous eutectic is clearly demonstrated. (From Konno and Sinclair, 1995a, with kind permission of Taylor and Francis Ltd.)

situ for different stoichiometries x , again determining activation energy (Missana et al., 1999). Although stress effects may change the kinetics in electron transparent foils, these experiments are useful in allowing transformation parameters to be measured and structural changes examined.

The reverse process of solid state amorphization is hard to measure using other experimental techniques. *In situ* heating of systems such as Ti-Si, Zr-Si, Pt-GaAs (Sinclair, 1994), and Al-Pt (Blanpain et al., 1990) allow nucleation locations to be determined and diffusion processes to be characterized. Amorphization can also be induced by the electron beam, as will be discussed in Section 8.

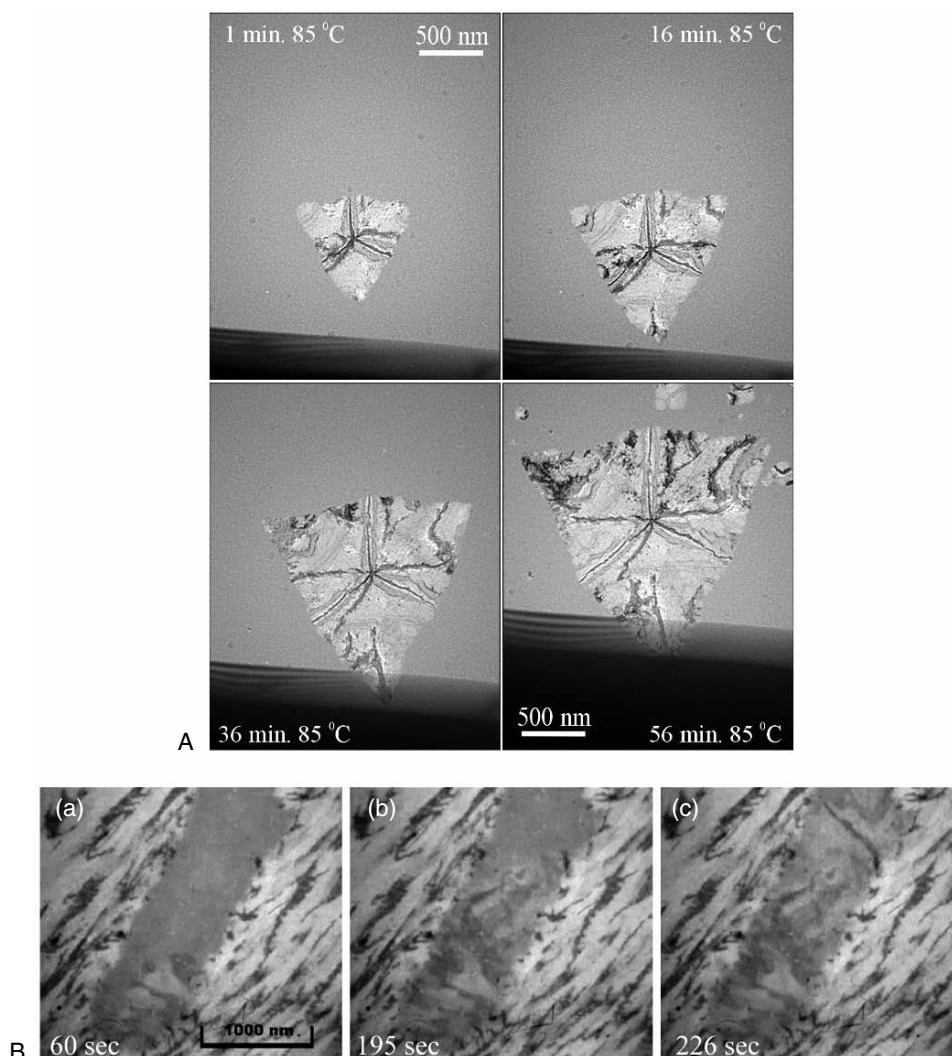


Figure 6–2. The amorphous to crystalline transformation in phase change materials. (A) Bright-field images recorded during crystallization of a 40-nm $\text{Sb}_{3.6}\text{Te}$ film at 85°C. The film was deposited on an SiN membrane. Note that the growing crystal was prenucleated by heating for 5 min at 95°C. (Reprinted with permission from Kooi and de Hosson, © 2004. American Institute of Physics) (B) Bright-field images displaying stepwise electron irradiation-induced crystallization of an amorphous data mark in 14-nm-thick $\text{Ga}_{15}\text{Sb}_{85}$ after irradiation for the times indicated at a current density of 1.5 nA mm^{-2} . The specimen was made from a conventional CD-RW/DVD1RW disk consisting of a $\text{ZnS:SiO}_2/\text{GaSb}/\text{ZnS:SiO}_2/\text{SiN}/\text{Ag}/\text{SiN}$ layered stack on a polycarbonate substrate, with all layers removed except for the GaSb and surrounding dielectric layers. The phase-change layer was crystallized using a broad laser beam then amorphous data marks were “written” using a home-built recorder. (Reprinted with permission from Kaiser et al., © 2004. American Institute of Physics)

2.1.2 The Solid-to-Liquid Transformation and the Structure of the Solid-Liquid Interface

Melting and freezing can be observed *in situ* by diffraction or imaging. Perhaps the ultimate example is the “nanothermometer” shown in Figure 6–3, fabricated by enclosing Ga in a large diameter carbon

nanotube (Gao and Bando, 2002; Z. Liu et al., 2004). This structure was used to measure the expansion coefficient of liquid Ga and observe different structures on freezing. Several other transformations involving liquids have also been studied *in situ*. It may at first appear surprising that liquids can be examined at all. However, liquids with low vapor pressure, such as Ga, In, Si, or Al, may be treated in the same way as solids, provided they do not move around too much, while liquids with high vapor pressure can be observed if they are naturally encapsulated, for example as inclusions, or are cooled. High vapor pressure liquids which are not in the form of inclusions require special techniques which will be described in Section 7. Liquids which are not encapsulated and are too mobile may be studied by coating the thin foil with a polymer to maintain shape (Kato et al., 2000; Senda et al., 2004). Such studies show that melting points in thin films differ from bulk (Senda et al., 2004).

TEM studies have provided information about the structure of the solid-liquid interface and the transformation between solid and liquid. In spite of extensive theoretical work on solid-liquid interface structure and the transient ordering in liquids just before solidification, such phenomena have proven difficult to study experimentally using other techniques. Even with TEM, relatively few systems have been examined. The nature of the solid to liquid transition in Xe films has been determined using diffraction in an environmental cooling cell (Zerrouk et al., 1994). Imaging studies have shown the persistence of order into liquids, for example at the (211) interface in PdSi (Howe, 1996) and at the interfaces of liquid Xe inclusions in Al (Donnelly et al., 2002). A combination of energy filtered imaging and diffraction contrast has been used to examine the interface between Al-Si eutectic and solid Al-Si alloy, to determine that the interface is quite abrupt and that changes in crystallinity correlate with composition (Storaska et al., 2004). Finally, the Si crystal-liquid interface has been imaged at high

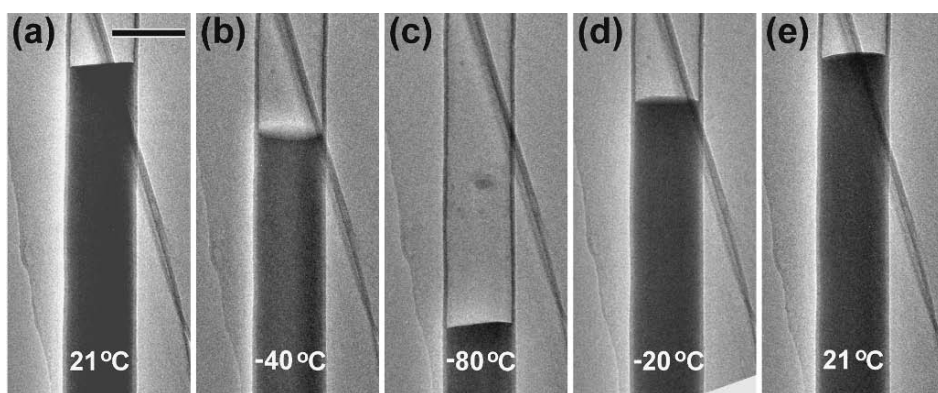


Figure 6-3. The Ga thermometer. Ga contraction and expansion inside a carbon nanotube upon cooling and heating. The background feature is part of the carbon support film. Scale bar = 100 nm. (a) At room temperature, 21°C. (b) At -40°C. (c) At -80°C, when solidification occurred. (d) The crystallized Ga was melted at -20°C. (e) Reheated to room temperature, 21°C. (Reprinted with permission from Z. Liu et al., © 2004 by the American Physical Society)

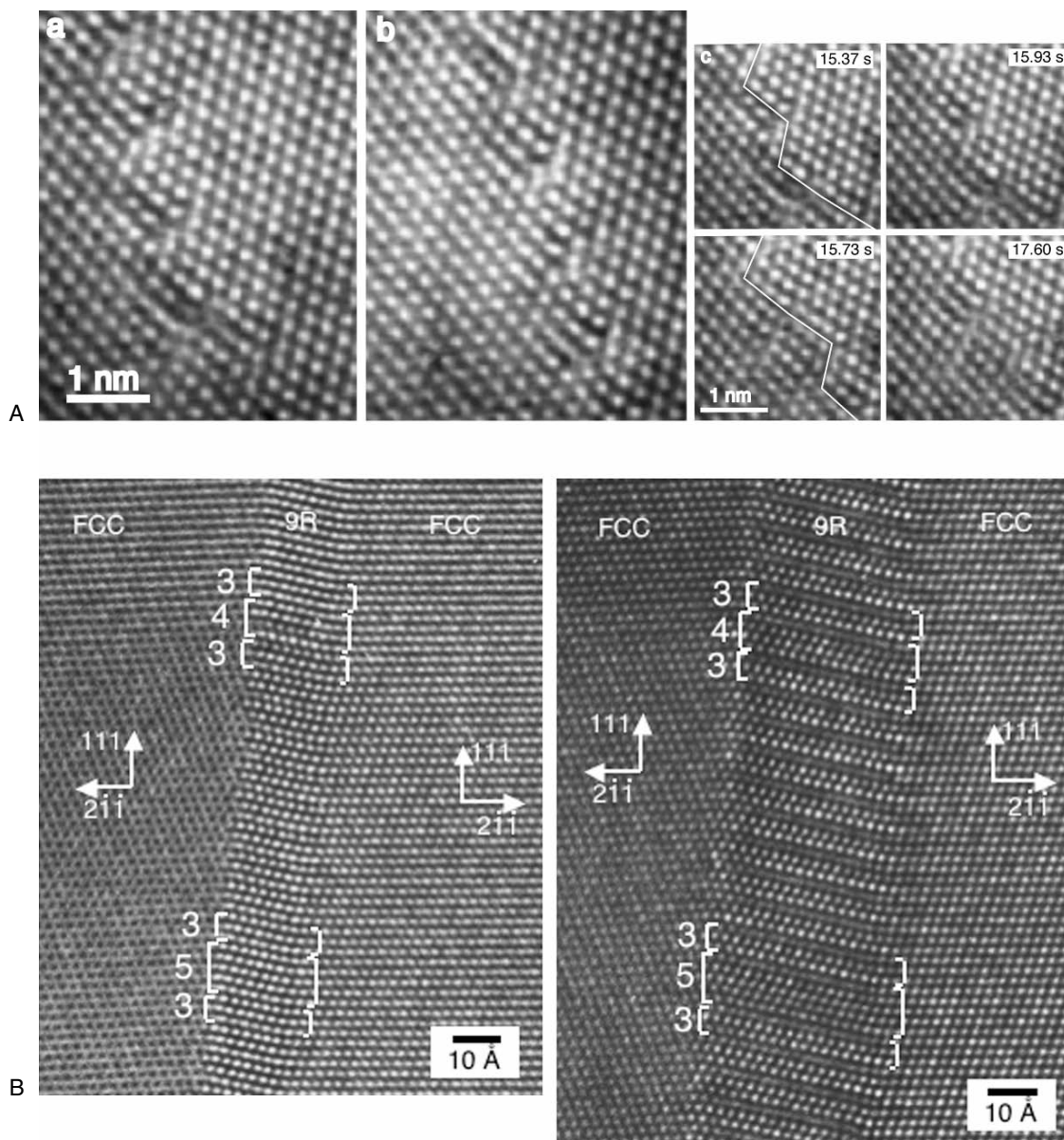


Figure 6-4. Grain boundary dynamics. (A) High-resolution electron microscopy images of a $[110]$ $\theta = 14^\circ$ tilt grain boundary in Au at 893K. Individual frames are shown from a video sequence recorded near optimum defocus. (a) GB at $t = 15.37$ s. (b) GB has moved to the right at $t = 44.93$ s and is near the $(6, -6, 1)$ symmetric orientation. (c) Detail of (a) depicted at various times in the four panels. A small region composed of eight atomic columns switches orientation between the two grains. Note the stacking disorder and misfit localization at the dislocation cores. Reprinted with permission from Merkle et al., © 2002b by the American Physical Society. (B) High-resolution image of the Cu $\Sigma = 3$ interface imaged along $[01\bar{1}]$. The grain boundary is dissociated into a narrow slab of 9R stacked material (fcc stacking but with an intrinsic stacking fault inserted every three planes). The two images were recorded 5 min apart after 400 kV electron beam irradiation and the 9R stacked region has expanded due to changes in the internal stress state induced by the beam. Stacking defects in the 9R structure can be related to the presence of secondary grain boundary dislocations at the interface. (From Medlin et al., 1998, with permission from Elsevier.)

temperature (Nishizawa et al., 2002) to determine the planarity of different index interfaces.

As well as static configurations at the solid-liquid interface, interface dynamics during solidification have been examined in several bulk materials. Arai et al. (2000, 2003) observed Si growth in the Al-Si system, while Sasaki and Saka (1996), Kamino et al. (1997a) and Saka et al. (1999) imaged surface melting in Al and Al₂O₃, observing ledge motion during solidification. Other studies have involved nanoparticles and inclusions and will be discussed below.

2.1.3 Grain Growth and Grain Boundary Motion

Grain boundaries in polycrystalline materials can move on annealing or mechanical deformation. Just as we have seen for crystallization studies, *in situ* observation of grain boundary dynamics can supply information on the growth mechanism, and even measure the effects of impurities and gas atmosphere. Grain boundary motion during heating has been observed for metals such as Cu, Au, and Al (Keller et al., 1997; Dannenberg et al., 2000; Kaouache et al., 2003). For nanocrystalline Ag thin films, the measurement of kinetic parameters demonstrated that grain growth is dominated by surface diffusion mass transport (Dannenberg et al., 2000). Grain boundary motion during mechanical stressing will be discussed in Sections 5.1 and 5.3.

More complex grain boundary dynamics can also be studied. An interesting example is the penetration of liquid Ga along grain boundaries in Al, relevant to the important embrittlement process (Hugo and Hoagland, 1998, 1999). The structure and strain field during penetration, the kinetics in different grain orientations, and the effects of dislocations were observed.

While dark field imaging works well for grain boundary dynamics in polycrystalline films or at low symmetry boundaries, high resolution heating experiments are very useful for understanding high symmetry boundaries. High resolution experiments on bicrystals having engineered boundaries with high symmetry, particularly in Au and Cu (Medlin et al., 1998; Merkle and Thompson, 2001; Merkle et al., 2002a, b), enable very detailed measurements at the atomic level and the determination of grain boundary migration mechanisms. *In situ* experiments show that collective mechanisms operate during migration, and that unusual structures may form and grow at boundaries (Figure 6–5). Dislocations may also be emitted, and the details of their structure and relationship with the boundaries can be measured (Lucadamo and Medlin, 2002).

2.2 Structural Phase Transitions

As well as the phenomena of melting, solidification and grain boundary motion, *in situ* techniques have been applied to understand transformations between different crystal structures and solid state reactions involving diffusion. These experiments have mostly relied on *in situ* sample heating, although transformations have also been initiated by straining, electron beam heating, electric and magnetic fields, and a gas environment. High resolution imaging and analysis, diffraction, or

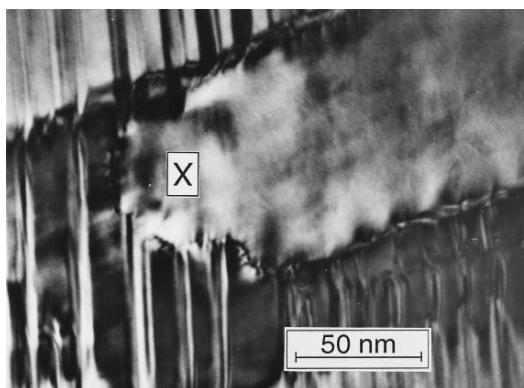


Figure 6–5. Phase stability in NiAl. When a martensitic Ni-rich $\text{Ni}_x\text{Al}_{1-x}$ sample, with $x > 63\text{at.}\%$, is annealed at moderate temperatures (550°C) it transforms into Ni_5Al_3 . On further heating to 780°C , the Ni_5Al_3 phase itself decomposes, forming B2 grains in a twinned L1_2 matrix. This image is part of a sequence obtained during heating that shows a smaller B2 grain growing by forming a small extension (marked as X) into the L1_2 matrix, consuming some twins, then rapidly expanding laterally. The Ni_5Al_3 phase is undesirable as it degrades the shape memory properties by inhibiting the transformation back to austenite, and its formation and stability are therefore important. (From Schryvers and Ma, 1995 with permission from Elsevier.)

low resolution weak beam or bright field imaging provide information that is complementary to that obtained from other *in situ* techniques, such as x-ray diffraction, which average over larger volumes.

Subtle changes in symmetry can be detected using the sensitive combination of diffraction and high resolution imaging. These combined techniques show the transformations between orthogonal, tetragonal, and cubic phases in oxides such as SrRuO_3 (Jiang and Pan, 2000). They also work well for transformations involving charge ordering and incommensurate phases, which will be discussed in Sections 4.1 and 4.3. Ordering can be investigated directly in the STEM, using its capabilities for elemental analysis of individual atomic columns. For example, Klie and Browning (2001) heated LaSrFeO_3 in the STEM, using the column environment, which is low in oxygen, to reduce the material. EELS showed that the resulting change in symmetry was due to ordering of oxygen vacancies.

When planning to study these sorts of phase transformations, it is important to consider the same experimental artifacts that we have already mentioned for melting and solidification. The results described below illustrate both the advantages and some pitfalls of *in situ* TEM.

We first consider transformations in intermetallic alloys such as shape memory alloys, which provide an excellent opportunity for *in situ* microscopy to display its strength. For example, for TiNi, the orientation relation between the different phases can be determined, and the dynamics of the emergence of martensite plates during straining can be observed *in situ* (Gao et al., 1996; Ma and Komvopoulos, 2005). *In situ* heating of NiAl alloys (Schryvers and Ma, 1995, Schryvers et al., 1998) showed how the texture and defect structure in the high tem-

perature phase are derived from the lower temperature phase, and illustrated the processes occurring during phase decomposition with low resolution imaging and diffraction (Figure 6–5). Higher resolution imaging showed details of the formation of the gamma phase in TiAl which would have been difficult to ascertain otherwise, for example, the ledge motion of interfaces (Howe et al., 2004). Other materials examined include the shape memory alloys CuAlMn (Dutkiewicz et al., 1995), TiNiHf (Han et al., 1997), and FeMnSi (Jiang et al., 1997). Crystallographic relationships, the interaction of dislocations with the transformation front, and the morphology of phases produced on heating or straining were studied. The presence and significance of incommensurate reflections in related materials have been examined using imaging plates and an in-column filter (Tamiya et al., 1998; Cai et al., 1999; Murakami and Shindo, 2001; Li et al., 2003). Cooling stages allow an even larger range of transformations to be accessed (Tanner et al., 1990).

However, thin foil effects are important in these transformations. Kuninori et al. (1996) showed that the foil thickness influences transformation temperatures—indeed, transformations do not occur at all at some thicknesses—and Ma and Komvopoulos (2005) showed that thickness can affect the sequence of phases. Electron irradiation effects are also important, for example in inducing some transformations in NiMnTi (Schryvers et al., 1996), and in changing the formation kinetics of the omega phase in beta phase Ti-Mo alloys on cooling (Matsumoto et al., 1999), a transformation which is important in understanding the anomalous electrical conductivity of these alloys.

Both beam and thin foil effects, of course, must be considered in any *in situ* transformation. Beam effects should always be evaluated by examining unirradiated areas after the transformation. Thin foil effects can be minimized in some cases by depositing the material of interest onto an electron transparent membrane. This reduces buckling and provides a more uniform temperature than a conventional specimen of varying thickness, advantageous for quantitative studies. An example is the transformation between beta and alpha phases of tungsten, for which the change in stress state is important in lithographic mask applications. Deposition of a uniform W film on a silicon nitride membrane (Ross et al., 1994a) allowed the transformation dynamics to be measured and the presence of voids to be related to the initial grain structure. Membrane specimens have been used successfully for many types of material, for example, by Morkved et al. (1998), Dannenberg et al. (2000), Kooi and de Hosson (2004), Grant et al. (2004), and Lee et al. (2005).

2.2.1 A Solid-State Diffusion Reaction: Silicide Formation

Reactions which occur at a planar interface between two materials have provided fruitful subjects for *in situ* experiments. *In situ* observations may allow determination of the diffusing species, the nature of nucleation sites, the sequence of phases, and the relationship between the crystal structures of the initial and final phases. However, the complexity of these reactions, compared to the crystallization and

melting reactions described previously, means that we have to be particularly careful to avoid artifacts. For example, if the sample dimensions are comparable to or less than the diffusion lengths of the moving species, then surface diffusion may affect the kinetics. Surface nucleation sites may be significant, and beam effects and stress relaxation in thin regions of the foil must be recognized. In spite of these issues, a successful body of work has been carried out on these transformations. We illustrate this by discussing silicide formation, a popular “test system” of great relevance to the microelectronics industry that has been examined using a range of *in situ* TEM techniques.

In situ silicidation was initially studied in cross section by heating a metal film such as Ti, Zr, or Cr that had previously been deposited on Si (H. Tanaka et al., 1995, 1996, 1998; Sidorov et al., 1998a; Figure 6–6A). Plan view experiments later provided the opportunity to examine silicidation on patterned substrates to study, for example, nucleation-limited transformations in small areas (Teodorescu et al., 2001; Ghica et al., 2001; Gignac et al., 1997; Figure 6–6B). These *in situ* experiments were very helpful in showing the sequence of phases, some of which are short-lived or hard to see otherwise; as mentioned previously, a single *in situ* experiment can replace a whole series of *ex situ* preparations (Wang and Chen, 1992). It is interesting to consider the sample geometry, however, as it illustrates some limitations of the *in situ* studies. In a cross sectional experiment, quantitative results are only obtained if surface diffusion pathways are suppressed (perhaps by coating the sample) and nucleation sites on the milled surface are minimized. It can only be assumed that the *in situ* experiment is an accurate representation of the bulk situation if both the activation energy of the reaction, and the final structure produced, are comparable with bulk experiments (Sinclair et al., 1988). In plan view, surface effects are not as significant, but thin film buckling must be considered.

Specialized *in situ* deposition techniques offer an interesting alternative way of looking at silicide formation. Rather than depositing a metal on the Si (or Ge) substrate *ex situ*, the substrate can be cleaned *in situ* in a UHV TEM and the metal then deposited *in situ*. The metal may be deposited onto a cool substrate which is then heated (Gibson et al., 1987; Ross, 2000) or it may be deposited at high temperature, where silicide phases form at once as islands (Ross et al., 1999a; H.P. Sun et al., 2005; Nath et al., 2005). To study the structure of such 3D islands in more detail, a combined UHV system allowing sequential TEM and STM imaging has been used to determine surface reconstruction as well as the sequence of phases (Tanaka et al., 2004). *In situ* deposition has also been used to study more complex silicide reactions, such as oxide and nitride mediated epitaxy (Kleinschmit et al., 1999; Chong et al., 2003). The experimental complexities of carrying out deposition *in situ* will be discussed in more detail in Section 3. But the advantages are clear in terms of avoiding contamination or oxidation (or evaluating their effects; see Figure 6–6C), discovering changes in the phase sequence as a function of film thickness, and looking at kinetic effects during deposition, such as coarsening.

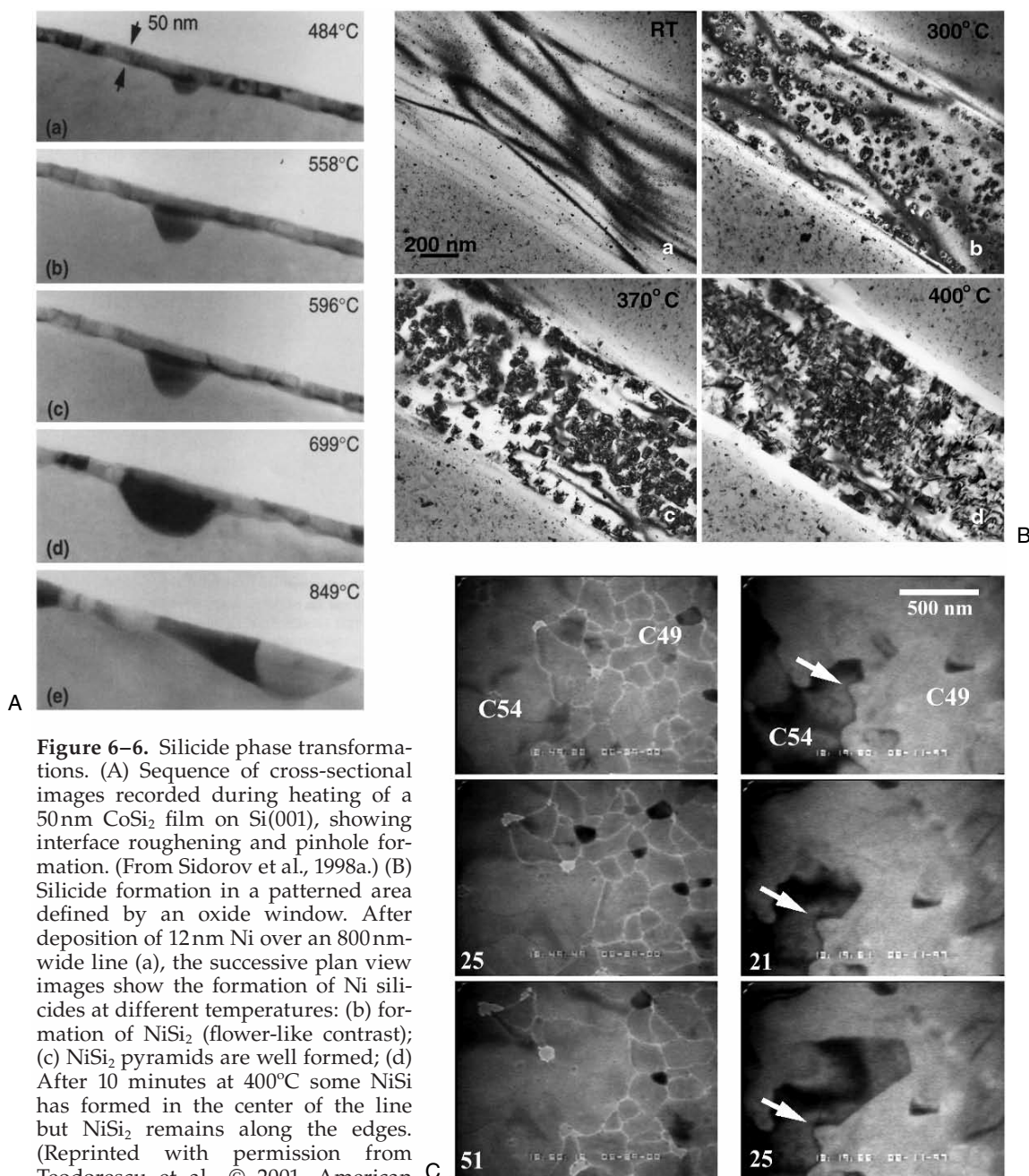


Figure 6-6. Silicide phase transformations. (A) Sequence of cross-sectional images recorded during heating of a 50 nm CoSi_2 film on Si(001), showing interface roughening and pinhole formation. (From Sidorov et al., 1998a.) (B) Silicide formation in a patterned area defined by an oxide window. After deposition of 12 nm Ni over an 800 nm-wide line (a), the successive plan view images show the formation of Ni silicides at different temperatures: (b) formation of NiSi_2 (flower-like contrast); (c) NiSi_2 pyramids are well formed; (d) After 10 minutes at 400°C some NiSi has formed in the center of the line but NiSi_2 remains along the edges. (Reprinted with permission from Teodorescu et al., © 2001. American Institute of Physics.) (C) The C49–C54 phase transformation in TiSi_2 which had been deposited and annealed *in situ* (left hand column), compared to *ex situ* deposited TiSi_2 , where Auger spectroscopy showed the presence of oxygen (right hand column). Times are shown in seconds and the arrow marks a fixed point on the specimen. The phase transition occurs smoothly in the clean film while it is strongly pinned at grain boundaries in oxidized films. (Reprinted with permission from Ross, ©2000 AAAS.)

2.3 Phase Transformations in Nanostructured Materials

So far we have discussed phase transformations in bulk materials or thin films, where *in situ* microscopy allows us to see changes in structure, how phases nucleate, and how growth fronts propagate. But TEM is particularly good at imaging nanostructures; in other words, resolving the shape, structure, and composition of small regions of a specimen. Thus, the techniques applied to study transformations in bulk materials can naturally be extended to transformations in small volumes, either embedded in a matrix or free standing. The stability of phases and the mechanisms of transformations in small volumes have been determined for several cases, confirming the important general conclusion that small particles show different phase diagrams compared to larger volumes of the same material. This is especially important given the many applications of nanostructured materials, for example, in high strength metal alloys, and individual, free-standing nanoparticles, for example, as catalysts or components in advanced electronic devices.

2.3.1 Size-Dependent Transformations in Embedded Nanostructures

By focusing on an individual inclusion or precipitate, *in situ* microscopy provides precise information about phase transformations and stability in nanoscale volumes. Excellent quantitative work in several systems shows the potential of the technique for future studies on a wider range of materials.

Pb in Al is a model system, since the lack of solubility of Pb in Al means that Pb spontaneously forms small cuboctahedral inclusions with a cube-on-cube orientation relation. Heating experiments allow strain, melting, and diffusion phenomena to be studied. A fascinating range of size-dependent properties has been seen (Figure 6–7). Melting of the Pb particles is size-dependent with huge supercooling possible, and there is a hysteresis on solidification due to the difficulty of nucleating ledges (Gabrisch et al., 2000). The decay of strain during solidification and melting provides information on the diffusion of point defects (Zhang et al., 2004). In particles at grain boundaries, which have complex structures, the melting of each interface at a different temperature can be seen (Bhattacharya et al., 1999; Dahmen et al., 2004). Co-implantation of different materials into Al, such as Cd/Pb, Sn/Pb, or Tl/Pb, allows phenomena associated with phase separation, melting, and interface structure to be examined (Johnson et al., 2002) and binary phase diagrams determined as a function of size.

Phase transformations involving precipitate growth have provided equally interesting information. In cases where precipitates are pinned on dislocations, diffusion parameters can be measured from their coarsening (Legros et al., 2005) or motion (Johnson et al., 2004). The kinetics of ledge motion and kink nucleation on precipitates can be observed during high resolution heating experiments (Howe, 1998). For example, for precipitate plates in Al-Cu-Mg-Ag alloys, imaging parallel and perpendicular to the interface demonstrated that precipitates grow by the terrace-ledge-kink mechanism (Benson and Howe, 1997) and even

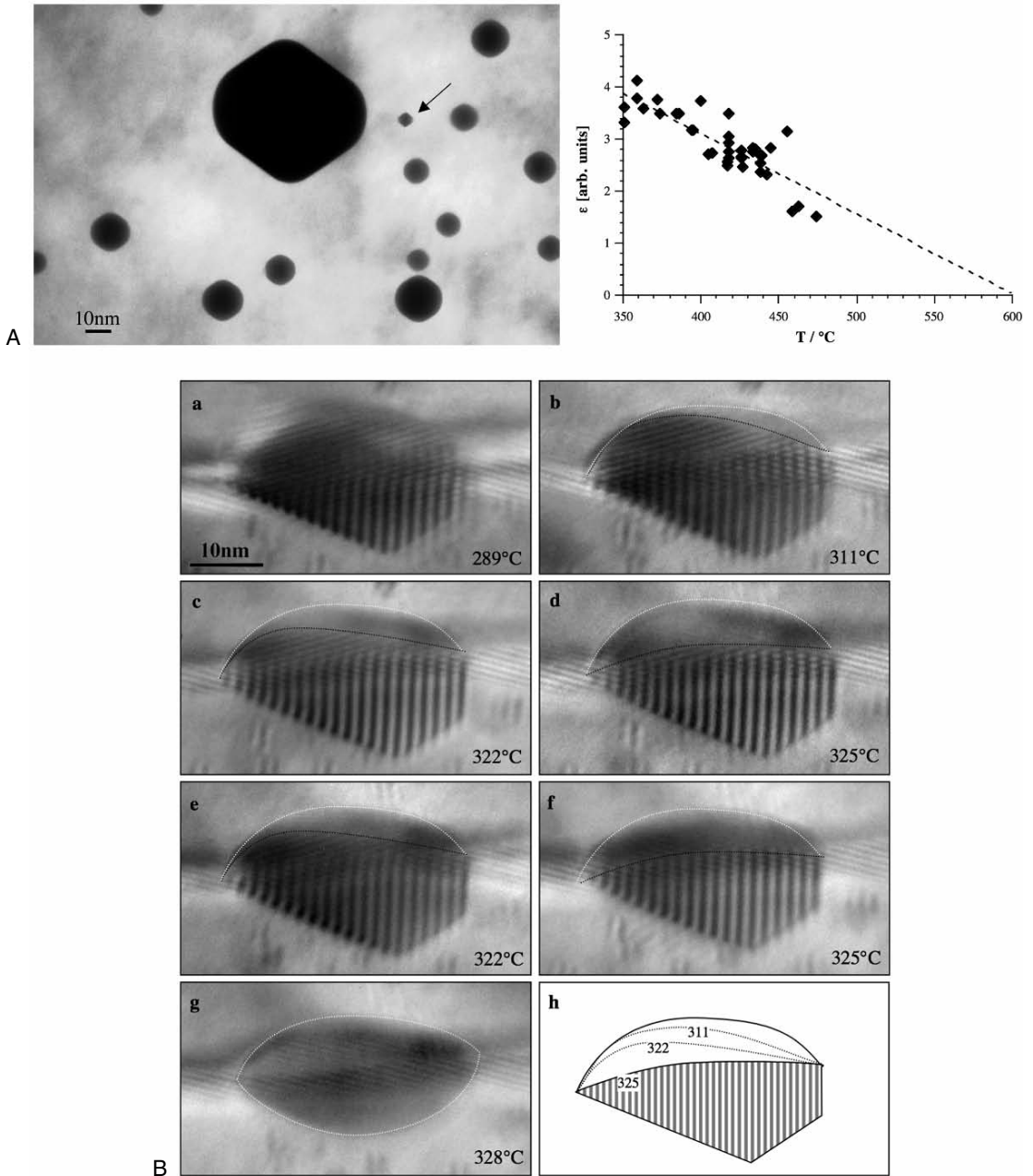


Figure 6–7. Nanoparticle melting phenomena. (A) Size-dependent melting of Pb inclusions in Al. The sample was produced by rapid solidification of an Al–0.5% Pb alloy and the image shows an array of particles at 423°C, which is 96°C above the bulk melting point. The rounding of most particles indicates their liquid state, while the smallest particles (arrow) are still faceted and solid. By measuring the dependence of inclusion shape on temperature and considering the inclusion shape change kinetics, the step energy as a function of temperature for steps on the inclusion surface can be calculated (lower graph). The least-squares fit indicates a roughening transition at about 600°C. (From Gabrisch et al., 2001 with permission from Elsevier.) (B) Reversible melting of 25-nm Pb inclusion at a grain boundary in Al (a–h). This particle has two different interfaces with two different grains and the two interfaces melt at different temperatures. The thin black line indicates the solid–liquid interface at different temperatures. (From Dahmen et al., 2004 with kind permission of Taylor and Francis Ltd.)

allowed the rate limiting steps and thermodynamic parameters of kink nucleation to be determined (Figure 6–8). Analytical techniques provide complementary information on the relationship between composition and structure at these growing interfaces, and the development of simulations for dynamic high-resolution imaging promises to make these studies even more quantitative (Howe et al., 1998). Other reactions, such as oxidation and reduction, can also be observed in precipitates (for example Isshiki et al., 1995; Kooi and De Hosson, 2001).

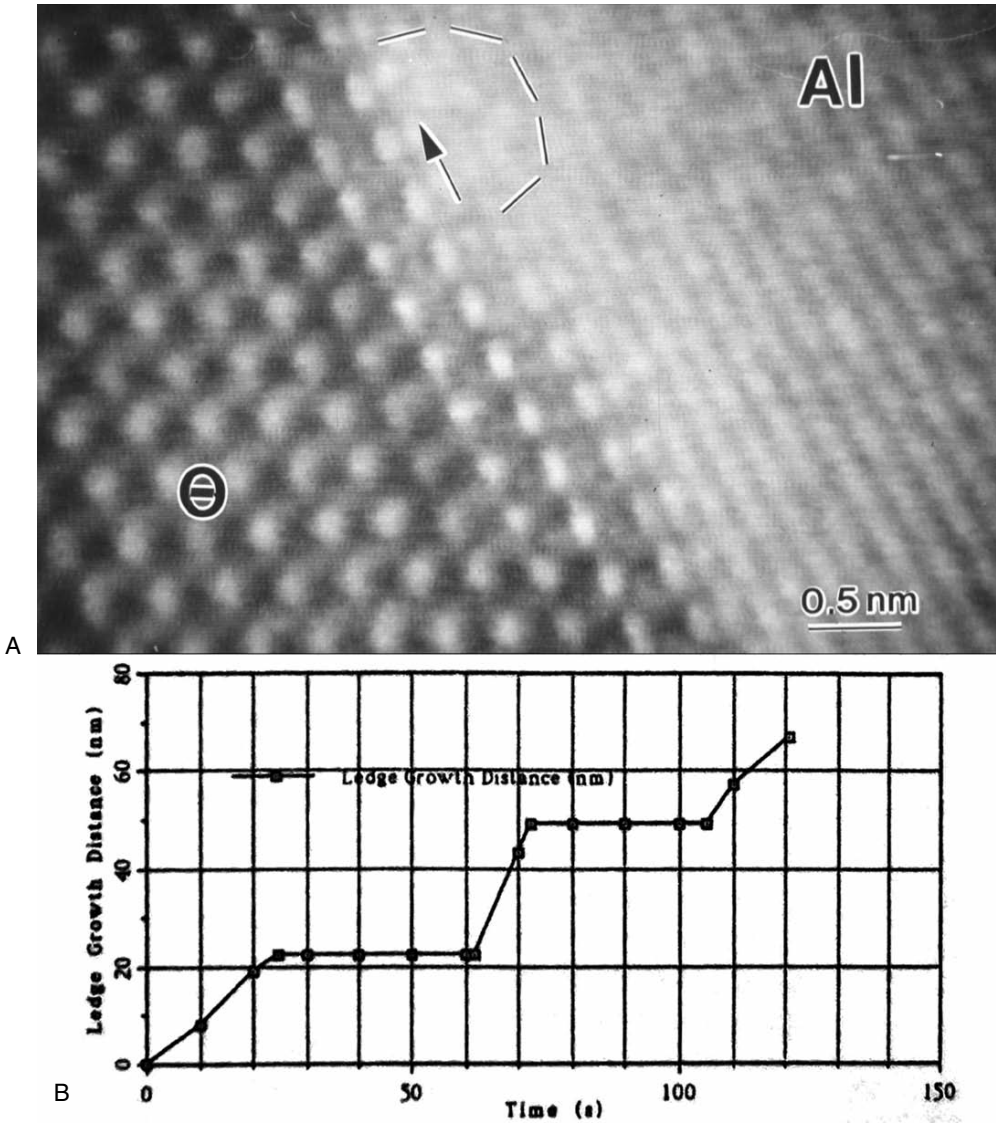


Figure 6–8. Precipitate growth mechanism. High-resolution image of a single ledge on a 111 θ plate in an Al–Cu–Mg–Ag alloy during growth at about 220°C, and a graph showing the position of the ledge as a function of time demonstrating growth by irregular motion of ledges. The ledge height is two 111 matrix planes high (half a unit cell). The circled area appears blurred in videos due to enhanced atomic motion there. (From Howe et al., © 1998 Courtesy of Cambridge University Press.)

2.3.2 Phase Transformations and Sintering in Free-Standing Nanoparticles

Isolated nanoparticles are often used as catalysts, and this has generated interest in understanding the factors determining their shape, phase stability, and sintering. For precipitates, we have seen properties quite different from the bulk material. Similar results are found for free-standing nanostructures when studied *in situ*.

The earliest *in situ* studies of free-standing particles demonstrated the dynamic nature of the atomic arrangement (Smith et al., 1986; Iijima and Ichihashi, 1986), and the large fraction of atoms on or near the surface indeed leads to unusual behavior. TEM has shown that phase transformations in free-standing particles are different from those in bulk, for example in observations of size-dependent melting (Howe, 1997) and changes in phase stability (Chatterjee et al., 2004). In this context, binary systems such as Au-Sn, Pb-Sn, Bi-Sn, and In-Sn have been extensively studied. For this, a two-source evaporator is used to form mixed composition clusters *in situ* (Figure 6–9). The binary phase diagram is found to depend strongly on size, with changes in the eutectic temperature (Yasuda et al., 2001; Lee et al., 2002a). Melting behavior, phase separation, and mixing also depend on the composition and size (Yasuda et al., 2000, Lee and Mori, 2004a, b). These effects reflect a change in solubility or the relatively high cost of forming phase boundaries.

Unusual structures may occur in certain free-standing particles on melting. In Al-Si, a solid Al particle inside a molten Al-Si sphere can form, moving with fractional Brownian motion (Yokota et al., 2004). In GaSb, particles decompose into a crystalline Sb core surrounded by liquid Ga (Yasuda et al., 2004). Stress may also play an important role in particle reactions. Metals encapsulated within multiwalled carbon onions have changed melting points due to the pressure (Banhart et al., 2003; Schaper et al., 2005), and the metal can even migrate through the graphitic covering (Schaper et al., 2005). When there is a solid oxide layer covering a nanoparticle, stress relief can cause cracking (Storaska and Howe, 2004).

Sintering of free-standing particles is particularly important in materials processing and has been examined *in situ*. Ceramic particles such as SiN can be made to sinter in a conventional microscope provided that a very high temperature stage is used (Kamino et al., 1995). For metals, of course, the surface oxide strongly influences sintering. For example, the degree of sintering of Fe and Nb nanoparticles on membranes, prepared *ex situ* but observed at high resolution during annealing in high vacuum (Vystavel et al., 2003a, b), was found to depend on surface oxidation. To solve this problem, an integrated system may be used, where particles are created and imaged in the same high vacuum. Yeadon et al. (1997) connected a sputtering chamber to a UHV TEM to carry out successful studies of metal sintering. Sintering of Cu on Cu foils proceeded by neck growth and grain boundary motion, whereas Co particles on Cu and Ag foils “burrowed” beneath the surface to minimize surface energy (Zimmermann et al., 1999). Sintering of metal particles on oxide substrates in a controlled environment is of

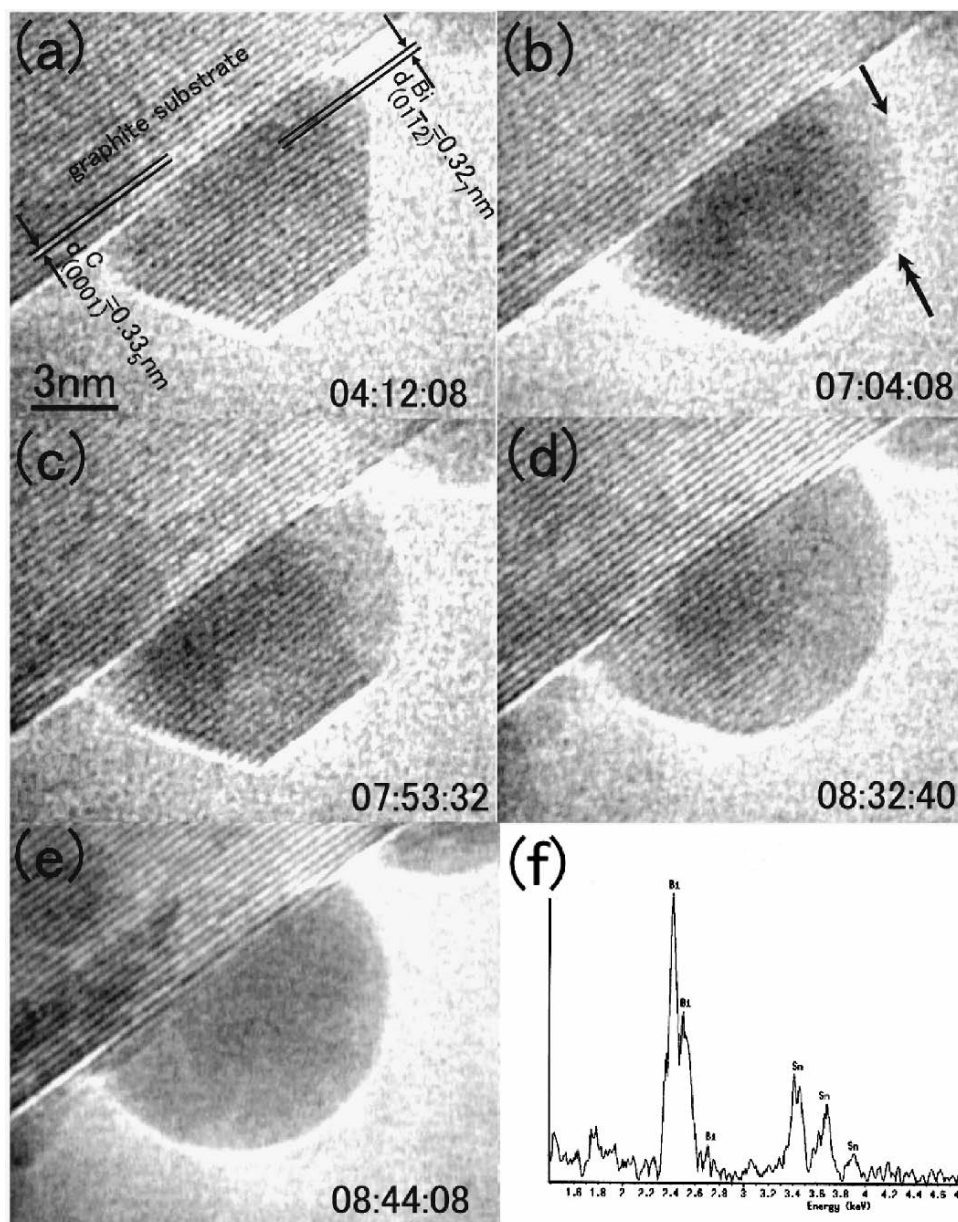


Figure 6-9. Interdiffusion in small particles. Sequence of images showing the alloying of Sn into Bi at 350K. (a) As-evaporated Bi particle; (b–e) the same particle during *in situ* Sn deposition. First a crystalline–liquid interface forms [arrows in (b)] and this then propagates through the crystal until the whole particle becomes liquid. EDX shows a composition of 50% Sn at this point (f). Not shown is the asymmetric behavior of Sn particles during Bi deposition; these become liquid at once without forming a phase boundary, an abrupt crystalline to liquid transition that is not expected from the bulk phase diagram. (From Lee and Mori, 2004a with kind permission of Taylor and Francis Ltd.)

particular relevance to catalysis, and will be discussed in that context in Section 3.2.

2.4 Summary

In situ microscopy of phase transformations has touched on a wide range of materials and addressed important problems related to crystallization and melting, diffusion, structural transformations, and grain boundary dynamics. Bulk crystals, embedded nanostructures, and free-standing nanoparticles have been studied, yielding quantitative information on reaction mechanisms and on the relationship of structure to dynamics. From this survey of results, it is clear that the *in situ* techniques we have described could be applied to many currently unstudied systems. However, for proper interpretation of results, care must be taken with thin foil effects, such as strain, surface diffusion and surface nucleation, and with beam effects. High voltage microscopes have significant advantages in minimizing thin foil artifacts, though at the cost of increased beam damage. But even at intermediate voltage, careful accounting for these effects can lead to quantitative results relevant to materials development.

3 Surface Reactions and Crystal Growth

A unique application of *in situ* microscopy, building on some of the techniques we have discussed above, is the examination of surface reactions and crystal growth. Rather than looking at bulk transformations as in Section 2, here we are more concerned with changes to the specimen surface. These changes may be initiated by heating or by exposure to a reactive environment or deposition flux. It is possible to study atomic scale processes on surfaces, including step dynamics and surface phase formation, as well as the growth of thin films and nanostructures. As we might expect from the discussion in Section 2, these *in situ* surface studies allow transient structures to be seen and kinetics to be measured. We will show that such experiments indeed contribute to an understanding of both surface reactions and growth, in some cases leading to improved control of surface structure or crystal size and shape.

These studies usually take place in a controlled environment TEM. The column of a standard TEM contains a mildly reducing atmosphere of 10^{-6} to 10^{-7} mbar and may also be contaminated with hydrocarbons. By controlling the environment, the specimen can be exposed to, for example, clearly oxidizing or reducing conditions, a solvent rich atmosphere to prevent dehydration (see Section 7), or an environment that allows vapor phase growth. Many such experiments can be carried out by leaking gases into the specimen area of a conventional TEM. However, some specimens require a microscope capable of UHV base pressure to avoid any background contamination. Such microscopes can be complex and expensive, but they enable experiments which can not be realized otherwise, especially if adjacent chambers are available for sample preparation. Major advances in surface science have been

achieved using UHV microscopy, and the equipment and the science have recently been reviewed by Poppa (2004).

3.1 Measurement and Modification of Surface Structure

Step flow and the development and stability of surface structures such as reconstructions may be studied by controlled heating, beam irradiation, or environmental stimuli such as deposition or exposure to a reactive environment. It may appear surprising that TEM is appropriate for surface studies at all, since the issues of temperature nonuniformity and strain relaxation associated with the preparation of thin foils could be avoided completely by using techniques such as LEEM and UHV SEM. But TEM has a wide variety of imaging and analytical modes sensitive to different aspects of a surface, and can be highly quantitative in terms of image analysis. Scanning probe microscopy also provides sensitive imaging of surface structures but lacks the time resolution necessary for surface dynamics, requiring typically 30 seconds to acquire each frame.

In situ TEM initially gained attention as a surface science tool with the successful determination of the Si (111) 7×7 reconstruction (Takayanagi et al., 1985), an accomplishment which STM and LEED had not been able to achieve at that time. A clean Si surface was prepared by heating in a UHV TEM and diffraction patterns were obtained and analyzed. Since then, many other static and dynamic surface structures have been determined after *in situ* preparation.

Surface structures may be prepared in the UHV microscope column by heating or deposition onto a thin foil (e.g., Kamino et al., 1997a; Oshima et al., 2000; Liu et al., 2001), or may be prepared in an adjacent chamber connected to the microscope by UHV (Marks et al., 1998). Every possible mode of the TEM has been used to analyze these surface structures. In plan view, diffraction techniques have solved reconstructions of metals on Si (Collazo-Davila et al., 1998; Oshima et al., 2000). Reflection electron microscopy (REM) has been used extensively to examine surface step dynamics due to electromigration, and the effect of metal adsorption on surface structure (Aoki et al., 1998; Minoda and Yagi, 1999; Minoda et al., 1999; Latyshev et al., 2000; Figure 6–10A). REM has also provided useful information on polar surface structures in oxides (Gajdardziska-Josifovska et al., 2002) and decomposition of the InP surface on heating (Gajdardziska-Josifovska et al., 1997). Information from REM and plan view TEM is complementary to that obtained from *in situ* SEM (Homma and Finnie 2002). Profile imaging, in which a surface parallel to the beam is imaged at high resolution, shows directly the periodicity and corrugations associated with surface reconstructions. This was first recognized early on (Marks, 1983; Bovin et al., 1985; Mitome et al., 1990; Smith et al., 1993), and has recently helped to determine complex structures like Si (5 5 12) (Liu et al., 2001) as well as faceting, reconstructions and dynamics of Au-decorated Si surfaces (Kamino et al., 1997b; Figure 6–10B) and beam-induced changes in surface structure and stoichiometry (Ning et al., 1996).

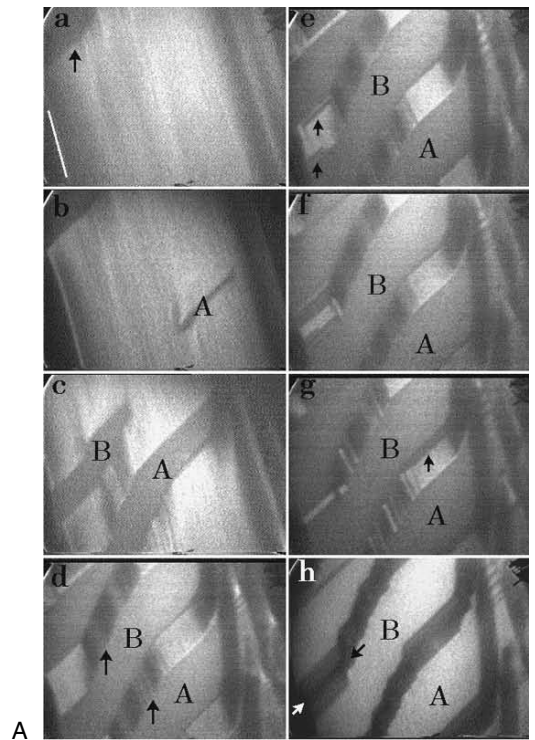
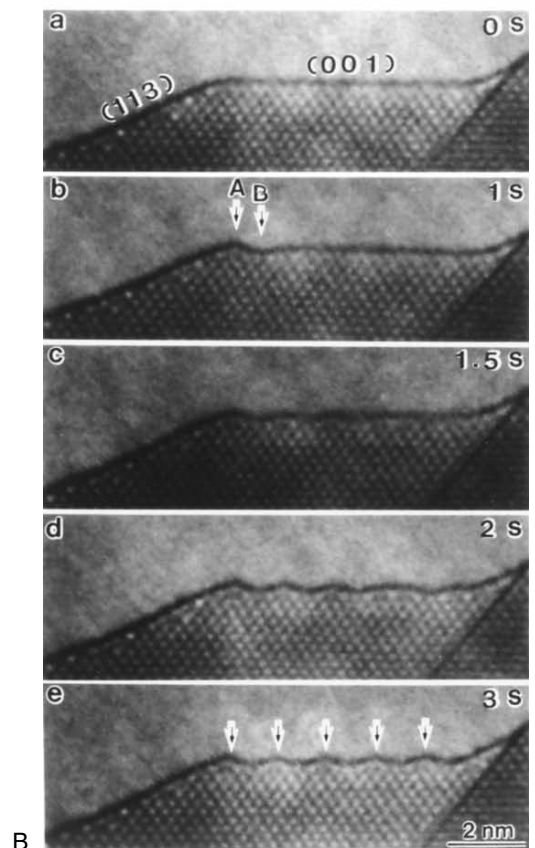


Figure 6-10. Deposition of Au on Si observed by different *in situ* imaging modes. (A) Step bunching on vicinal Si(001). A series of REM images obtained during the deposition of Au at 800°C onto a vicinal (4° miscut) Si(001) surface shows first the formation of (001) terraces in (a) and (b). Step bunching and formation of step bands are shown in (c) and (d); facet nucleation on the step bands is shown in (f) and (g). Finally there is a complete transformation of the entire vicinal surface into a hill-and-valley structure of (001) superterraces and (119) and (117) facets (h). (From Minoda et al., 1999 with kind permission of Taylor and Francis Ltd.) (B) The Au-induced reconstruction of a flat Si(001) surface on heating. The sample was prepared by depositing Au onto a rough, oxidized Si particle. The irregular sample geometry means that the temperature is not accurately known. However, on heating, the Au first agglomerated; further heating caused the oxide layer to disappear and the Au to spread over the surface (dark line along the sample edge), causing localized faceting into (001) and other terraces. The (001) facets then reconstruct over the timeframe shown, starting from the terrace boundary (arrows A and B). (Reprinted with permission from Kamino et al., © 1997 by the American Physical Society.)



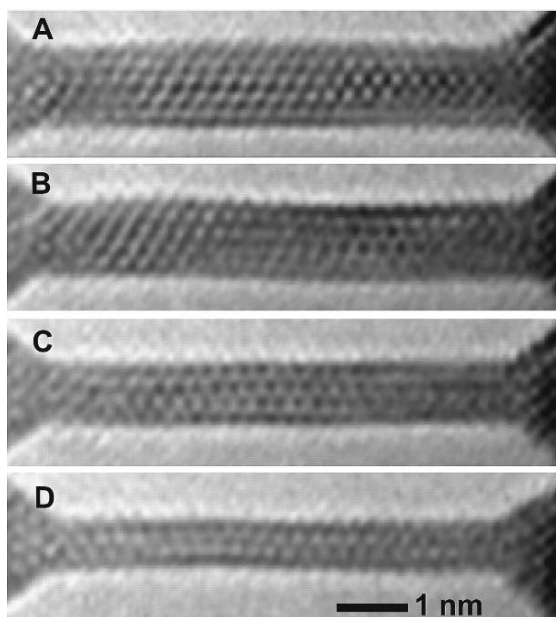


Figure 6-11. Nanostructures formed by surface mobility. Images of stable Au nanowires observed during the electron beam thinning of a specimen. The diameters of the wires are (A–D) 1.3, 1.1, 0.8, and 0.6 nm, respectively. The wire images are wavy, especially the thinnest, and can be modeled with helical structures. (Reprinted with permission from Kondo and Takayanagi, © 2000, AAAS.)

Surface diffusion under beam irradiation is evident during profile imaging. This has been used to advantage to study narrow wires *in situ*. Two adjacent holes are formed in a thin foil and the neck between them observed as it thins and breaks due to inbuilt stresses. Such experiments show a variety of interesting non-bulklike structures in Au (Figure 6–11), such as single chains (Koizumi et al., 2001) and helical multishell wires (Kondo and Takayanagi, 2000; Oshima et al., 2003a), with some of the structures related to shear (Oshima et al., 2003b). Beam-induced atom migration reduces the dimensions one layer at a time (Oshima et al., 2002). Unusual wire structures also form in Pt (Oshima et al., 2002). The surface diffusion of single atoms, for example W on MgO, can be imaged in TEM (N. Tanaka et al., 1998).

Controlled environment and UHV TEM thus has an excellent track record for creating and observing surface reconstructions and observing step motion and surface diffusion. We now focus on some specific applications where the ability to characterize and modify surfaces *in situ* has led to particularly interesting and significant advances.

3.2 Catalysis

Controlled environment TEM has enabled pioneering studies of the structural details of reactions during heterogeneous catalysis (Gai, 2002a; Sharma, 2001, 2005). High resolution imaging can be carried out at high temperature and under a controlled gas environment, up to several millibar of, say, H₂, CO, or O₂. Catalyst surface structure has been related to reactivity, intermediate phases have been determined, and changes in catalyst structure have been visualized during activation and poisoning. Several controlled environment TEMs are in opera-

tion in industrial laboratories, and we suspect that the many important results remain proprietary!

Most studies have focused on catalysts consisting of metal particles on an oxide substrate (Figure 6–12). Gai et al. (1990) first examined shape changes in Cu particles in oxidizing and reducing environments, and observed the stability of Pt and Ru particles on TiO₂ (Gai 1998; Gai et al., 2000). Hansen et al. examined the effect of Ba promoters on surface structures in Ru particles (2001), and used the shapes of Cu particles to determine relative surface energies under oxidizing and reducing conditions (2002; Figure 6–12B). Regeneration processes are important in such catalysts, and here too *in situ* studies have proven useful. For example, during regeneration of Pd/Al₂O₃ catalyst, sintering also takes place. The sintering behavior of “used” Pd particles, studied using analytical TEM as well as controlled environment heating, is affected by hydrocarbons which build up during use of the catalyst (R.-J. Liu et al., 2004; Figure 6–12A).

A wide variety of other materials and reactions have also been investigated under a controlled environment. Gai and Kourtakakis (1995) observed a glide shear rearrangement in vanadium pyrophosphate during reduction, and developed a model for the surface activity of this material, which is important in butane catalysis. Sharma et al. (2004a) examined structural changes during reduction of CeO₂ catalyst, and interestingly used *in situ* energy loss spectroscopy to determine oxidation states as a function of temperature. Even the growth of polymeric reaction products *in situ* has been observed (Crozier et al., 2001; Gai, 2002b). Other reactions include intercalation in some interesting layered structures (Diebolt et al., 1995; Sidorov et al., 1998b), the nitridation of zirconia (Sharma et al., 2001), and de- and rehydroxylation of the lamellar material Mg(OH)₂, which is important in CO₂ sequestration (McKelvy et al., 2001). The reaction of MgO with water vapor has also been observed *in situ* (Sharma et al., 2004b; Gajdardziska-Josifovska et al., 2005).

Finally, photocatalysis can be studied *in situ* if a UV light source is brought into the specimen area. High resolution *in situ* imaging of the decomposition of hydrocarbons deposited on a TiO₂ film (Yoshida et al., 2005) provides information on the mechanism of the process. These exciting results suggest that *in situ* studies will continue to have an important impact in the future development of catalysts and other functional materials.

3.3 Oxidation of Surfaces

Corrosion of metals has significant impact on industry, so it is important to gain a fundamental understanding of oxidation and reduction by comparing controlled observations with theoretical predictions. For copper, oxidation was found to proceed via nucleation, growth, and coalescence of oxide islands (Figure 6–13). This result allowed the development of oxidation theories beyond simple models that had assumed a continuous oxide film (Yang et al., 1998, 2002; Zhou and Yang, 2002, 2003). Other than Cu and its alloys (Wang and Yang, 2005),

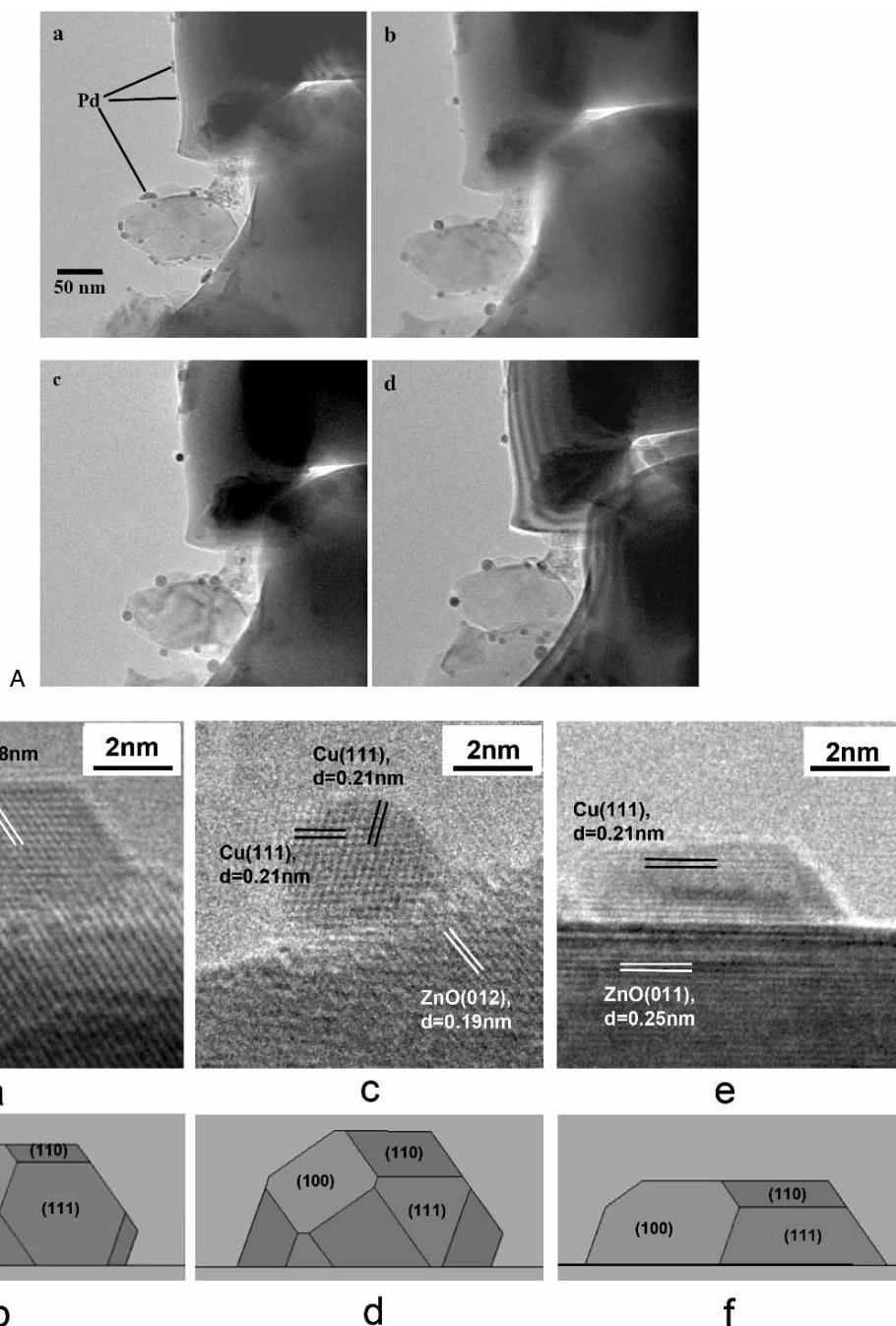


Figure 6–12. *In situ* imaging of catalysts. (A) Fresh Pd/Al₂O₃ catalyst (used for hydrogenation of acetylene) (a) in the as-received condition (room temperature), and (b–d) after heating in 500 m Torr steam at 700°C for 1, 4, and 7 hours. Catalysts are regenerated by heating in steam to remove hydrocarbon buildup, but this causes sintering of the metal particles, reducing activity. *In situ* experiments show that, for fresh catalysts, sintering is by conventional Ostwald ripening, while movement and coalescence occurs for used catalysts. (Reprinted with permission from R.-J. Liu et al., © 2004 AAAS.) (B) Images of a Cu/ZnO catalyst (the methanol synthesis catalyst) in various environments at 220°C, together with the corresponding Wulff constructions of the Cu nanocrystals. (a, b) 1.5 mbar H₂ at 220°C; (c, d) in a mixture of H₂ and H₂O with ratio 3:1 at a total pressure of 1.5 mbar; (e, f) in a mixture of 95% H₂, 5% CO at a total pressure of 5 mbar. These images allow the relative surface energies to be determined as a function of environment. (From Hansen et al., © 2002. Courtesy of Cambridge University Press.)

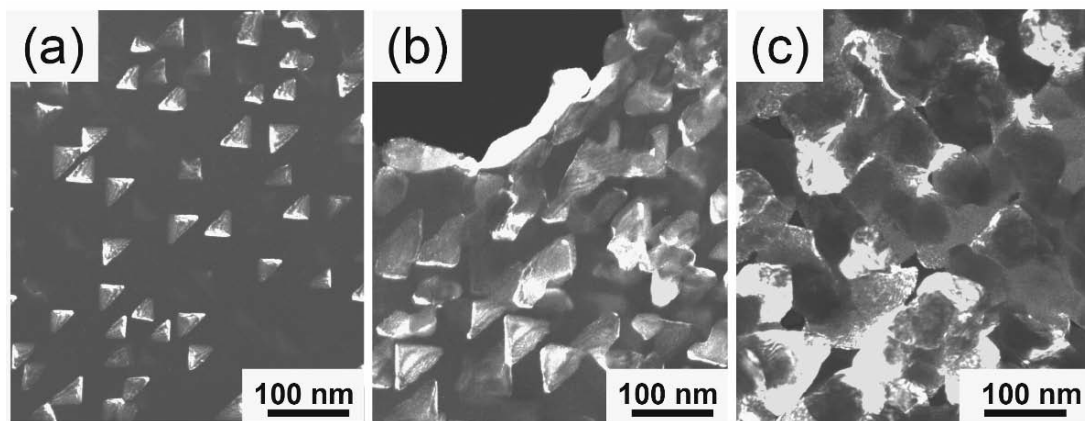


Figure 6-13. Mechanism of copper oxidation. Dark-field images obtained during oxidation of Cu at 0.1 Torr and 350°C in a UHV TEM. Imaging using the Cu_2O {110} reflection showed that oxidation is not planar, but takes place by Cu_2O island (a) nucleation (5 min), (b) growth (15 min), and (c) coalescence (25 min). The specimen was prepared by floating a 60-nm Cu film onto a support and then removing the surface oxide *in situ* by annealing at 350°C in methanol vapor for 15–30 min. The area and number density of the islands grown both at low pressure and at high pressure (shown here) were modeled using Johnson–Mehl–Avrami–Kolmogorov theory to give surface diffusion parameters. (Reprinted with permission from Yang et al., © 2002. American Institute of Physics.)

metal oxidation remains largely unstudied, but *in situ* experiments could clearly offer the possibility of improving corrosion resistance through alloying or surface processing.

Silicon oxidation is another industrially significant process, as a defect-free Si/SiO₂ interface is fundamental to transistor operation. For Si(111) (Figure 6-14A), forbidden-reflection imaging showed that steps do not move during oxidation, meaning that any surface roughness remains during processing (Ross and Gibson, 1991; Ross et al., 1994b). For Si(001), the same result is seen (Figure 6-14B). Forbidden reflection experiments are useful in their ability to probe the buried Si/SiO₂ interface, and are complementary to results obtained by *in situ* scanning reflection electron microscopy (Ichikawa, 1999).

High resolution imaging provides insights into reactions in more complex materials. By oxidizing and reducing niobium oxides, Sayagues and Hutchison (1996, 2002) showed the formation of a series of block structures with changing stoichiometry (Figure 6-15). The combination of analytical techniques with imaging would undoubtedly lead to further useful insights into oxidation and related reactions.

3.4 Growth of Carbon Nanostructures

The growth of carbon nanostructures has been extensively studied *in situ* since the discovery of these interesting materials by TEM. It is relatively easy to form carbon structures *in situ* with a controlled environment plus a catalyst or a graphitic precursor (Figure 6-16). Beam effects provide an important ingredient in the synthesis, especially if catalysts are not used. The beam interacts with the atmosphere above the specimen producing a plasma that can generate fullerenes (Burden and Hutchison, 1998). Alternatively, irradiation of graphitic materials

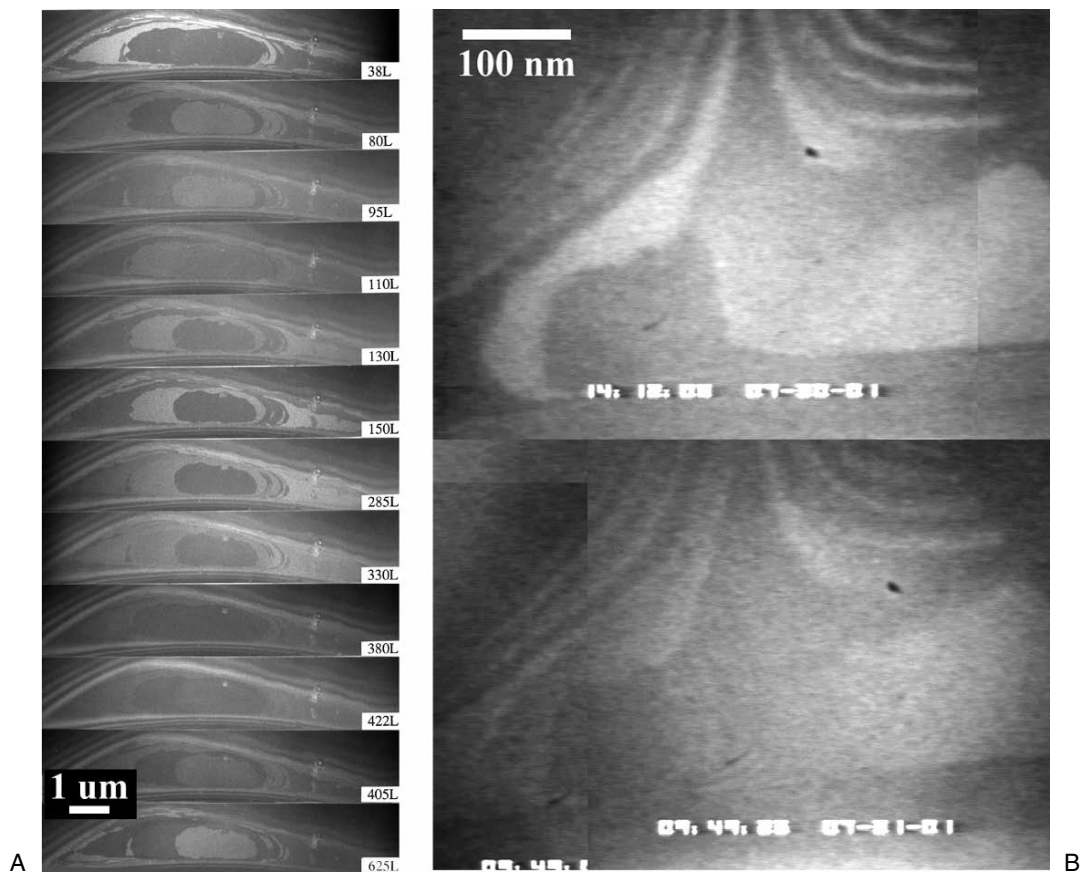


Figure 6–14. Mechanism of silicon oxidation. (A) Series of images of an Si(111) sample observed in plan view during oxidation in 2×10^{-6} Torr water vapor at 400°C . The time of each frame is given. A $1/3$ (422) forbidden reflection was used to form the images so that the gray levels correspond to terraces, with intensities repeating every unit cell (three steps). Steps do not move during oxidation of several layers showing that step sites are no more reactive than terrace sites. (Reprinted with permission from Ross and Gibson, © 1991 by the American Physical Society.) (B) Two images of an Si(001) sample observed before and after oxidation in air at room temperature. A $1/4$ (220) forbidden reflection was used to form these images. The steps do not move on this surface either.

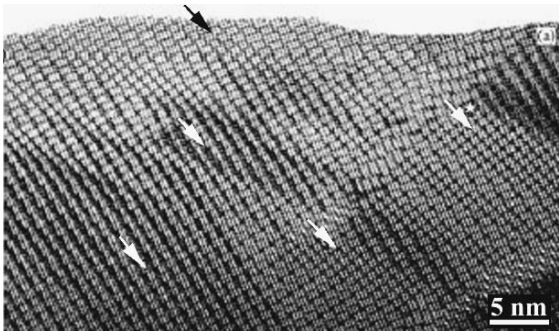


Figure 6–15. Oxidation of a block oxide structure. High-resolution image of an $\text{Nb}_{22}\text{O}_{54}$ crystal after heating by the electron beam and exposure to 15 mbar oxygen. The partly oxidized structure consists of microdomains of $\text{Nb}_{10}\text{O}_{25}$ (arrowed) in an $\text{Nb}_{22}\text{O}_{54}$ matrix. These images have been used to identify the structure of the $\text{Nb}_{10}\text{O}_{25}$ phase and the complete oxidation sequence from $\text{Nb}_{12}\text{O}_{29}$ to $\text{Nb}_{10}\text{O}_{25}$ has been determined using *in situ* experiments. (From Sayagues and Hutchison, 2002 with permission from Elsevier.)

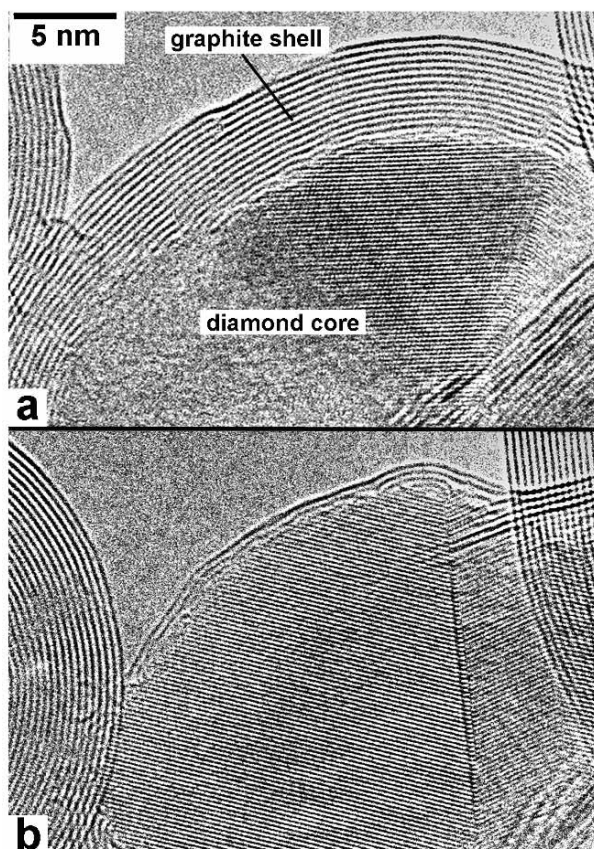
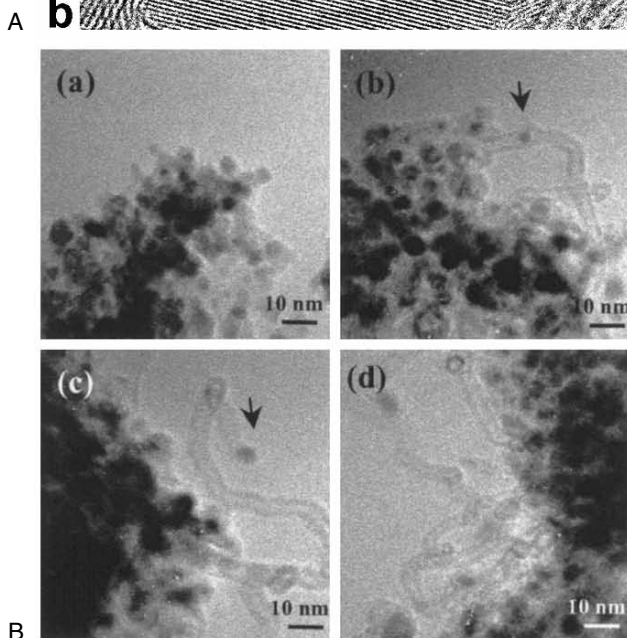


Figure 6-16. Carbon nanostructure growth *in situ*. (A) Formation of diamond *in situ*. Polyhedral graphitic particles were produced by arc-discharge and were transformed into perfectly spherical onions by electron beam irradiation above 600K. The decreasing distance between shells toward the center showed that the onions are in a state of high self-compression. The nucleation of cubic diamond occurs in the centers of the onions during irradiation above 900K, and the diamond grows until, surprisingly, almost all the onion shells are consumed. (a) After 2h of irradiation at 1.25MeV and 20mAcm⁻². (b) After a further hour of irradiation. A typical twin is visible in the diamond. High pressure appears necessary to nucleate the diamond, but further growth is by beam-induced defects. (Reprinted with Permission from Banhart, © 1997. American Institute of Physics.) (B) Growth of multiwalled carbon nanotubes by catalytic chemical vapor deposition. Bright-field images showing an Ni-SiO₂ catalyst (a) exposed to H₂ at 450°C; (b) after exposure to acetylene (300mTorr of an H₂-C₂H₂ mixture), and (d) another part of the sample after 3min. The arrows show an individual Ni particle. At higher temperatures single-walled tubes formed but with catalysts present at the bases rather than the tips of the tubes. (Reprinted with permission from Sharma and Iqbal, © 2004. American Institute of Physics.)



generates point defects which deform the graphitic sheets, forming new structures (Ugarte, 1992; Banhart, 1997), while irradiation of materials such as Cu implanted with C causes graphitic onions to grow (Abe et al., 2002). Familiar or new C and BN structures, and even formation of diamond from graphite, can thus be observed *in situ* (Ugarte, 1992; Ru et al., 1996; Niihara et al., 1996; Banhart, 1997, 1999, 2003; Bengu and Marks, 2000; Roddatis et al., 2002; Troiani et al., 2003; Gloter et al., 2004; Wang et al., 2005).

Carbon nanotubes are of particular interest, and can be grown *in situ* by introducing a precursor gas such as methane, propylene, or acetylene over a catalyst (Sharma and Iqbal, 2004; Sharma et al., 2005). During growth, individual catalyst particles change their shape, and nucleation sites can be identified (Helveg et al., 2004). Once grown, carbon nanotubes can be modified with the beam (Terrones et al., 2000, 2002) to produce more complex structures.

3.5 Epitaxial and Polycrystalline Thin Film Growth

The experiments in Section 3.4 have shown the exciting possibilities for controlled environment growth of nanostructures. Continuous thin films can also be grown *in situ*, and this allows important processes such as nucleation, development of surface morphology, and relaxation to be observed. Although some studies describe polycrystalline film growth (for example Al; Drucker et al., 1995), most systems examined *in situ* have been epitaxial. These include Au on MgO (Kizuka and Tanaka, 1997a, b), Ge on Si (see below), and silicides on Si (Section 2). These experiments can provide detailed and quantitative information if growth conditions such as flux and temperature are calibrated carefully.

The most detailed studies have examined Ge and SiGe epitaxy on Si. This is a “test system” for studying epitaxial growth phenomena which also has great relevance to the development of microelectronic devices. A true UHV environment is required for the experiments, as the Si substrate foil is cleaned by heating in UHV to above the oxide desorption temperature. Growth is then carried out by UHV-CVD using gases such as disilane or digermane. Growth was observed by Krishnamurthy et al. (1991) in STEM, and by Minoda and Yagi (1996) and Ichikawa et al. (1998) in REM. But most work in this system has been carried out using conventional weak beam imaging in plan view, giving the highest sensitivity to strain fields (Figure 6–17). LeGoues et al. (1996) and Hammar et al. (1996) grew Ge on Si(111) and (001) *in situ*, clearly imaging the initial surface reconstruction, the nucleation of Ge islands, and later their growth and coalescence. The structures produced depend strongly on growth conditions. By varying the parameters, a range of fascinating phenomena now known to be common in other epitaxial systems was observed, such as the change in island shape during the introduction of stress-relieving dislocations (LeGoues et al., 1994, 1995). The range of structures observed in these studies would have been tedious to capture *ex situ*, and dynamic phenomena

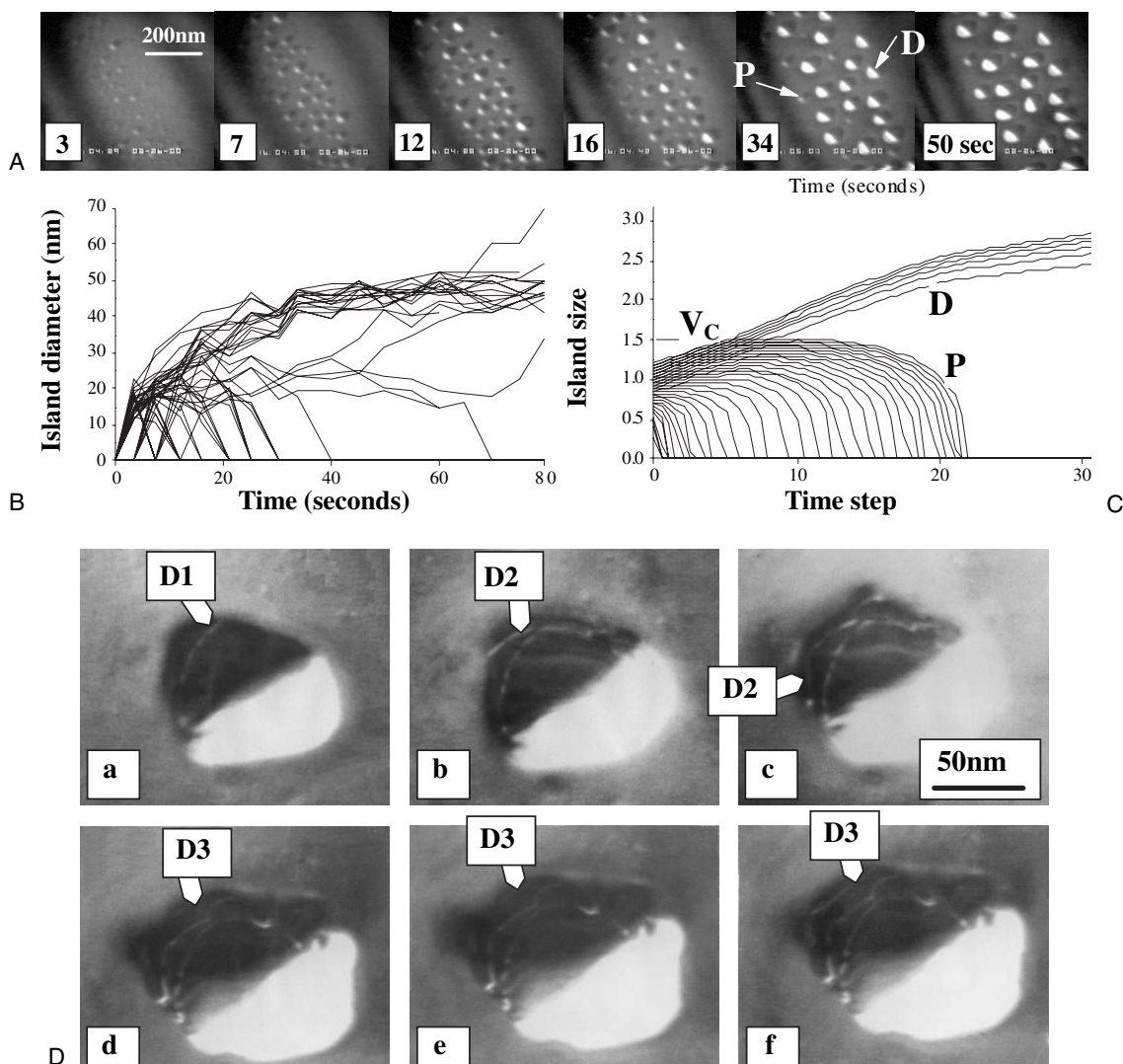


Figure 6-17. Ge island nucleation, growth, and relaxation. (A) Island coarsening. A series of video frames showing stages of island growth within a $0.5\mu\text{m} \times 0.5\mu\text{m}$ area. Growth conditions were 5×10^{-7} Torr Ge_2H_6 and 650°C . Weak beam images were acquired in a (g , $3g$) condition with $g = 220$. The time of each frame after “nucleation” (first appearance of islands) is given. (B) The time evolution of every island within the area. (C) A simulation showing the fate of islands with different initial sizes, based on a modified Ostwald ripening process in which islands coarsen but also undergo a shape transition at V_c . The scale of the plot is chosen to match the data shown in (b). (Reprinted with permission from Ross et al., © 1998 by the American Physical Society.) (D) The introduction of dislocations into a partially relaxed island. Images were obtained 59, 94, 96, and 140 min after the beginning of growth, with the last three images taken within 1 s of each other. Growth conditions were 650°C and 10^{-6} Torr of 10:1 He: GeH_4 . Only dislocations in the dark part of the image can be seen (D1–D3). (From Hammar et al., 1996 with permission from Elsevier.)

such as island shape changes and the essential role of Ostwald ripening (Ross et al., 1998) would not have been detected without the use of *in situ* TEM. Changes caused by the presence of surfactants during growth have also been examined *in situ* (Maruno et al., 1996;

Portavoce et al., 2004). It is interesting to note that complementary studies using *in situ* LEEM have been important in fully understanding growth in the SiGe system, as LEEM is more sensitive to surface structure (e.g., Ross et al., 1999b).

When low Ge content (say 15%) SiGe is grown on Si, islands are not seen, but instead a continuous flat layer forms. When sufficiently thick, this film relaxes by introduction of dislocations. The motion of these dislocations can be measured during film deposition and compared with dislocation dynamics during post-growth annealing, which will be discussed in Section 5. Interestingly, the parameters governing dislocation motion are different during growth, where the surface is H terminated, versus during annealing, where the surface is oxidized (Stach et al., 1998b, 2000). A higher kink nucleation rate under the oxidized surface, perhaps due to surface stress or increased point defects, is the probable cause. This is important for modeling relaxation during device processing, and shows once again the unique information that can be obtained when materials are observed during growth rather than *ex situ*.

3.6 Crystal Growth on Patterned Surfaces

An interesting extension of the growth studies described above is the study of crystal growth on a non-uniform substrate. By carrying out growth on substrates which have been patterned to create areas of different reactivity, strain, or topographic contrast, the effect of these parameters on growth may be visualized directly and even quantitatively, if kinetic data is obtained.

Again, most work has been done in the SiGe system, motivated by an interest in controlling island self-assembly for fabricating novel electronic devices. It is well known that if Ge is deposited on patterned Si, islands form at positions aligned with the topography. *In situ* measurement of nucleation times at different locations (Ross et al., 2004) showed that this is controlled by competition between edge adsorption of adatoms and terrace nucleation. Patterns may also be created with a focused ion beam. At low doses, this forms a shallow topography (Figure 6–18A) that is sufficient to control nucleation and alter wetting layer thickness (Kammler et al., 2003; Portavoce et al., 2006). In these experiments the focused ion beam gun was installed in a chamber connected to the TEM by UHV. The inclusion of surface processing tools that are not *in situ* (i.e., in the polepiece) but are within the microscope's vacuum system enables a wider range of processes to be carried out controllably.

A second example also relates to semiconductor nanostructures: vapor-liquid solid growth of nanowires. Here, a droplet of liquid eutectic catalyzes growth to form elongated wirelike structures. This process has been imaged *in situ* in plan view, allowing wire growth to be observed qualitatively, in Si/Au (Wu and Yang, 2001), Si/Fe (Zhou et al., 2002), and GaN (Stach et al., 2003). In GaAs/Au (Persson et al., 2004), post-growth heating of wires was used to deduce an alternative mechanism, vapor-solid-solid growth. If experiments are performed in

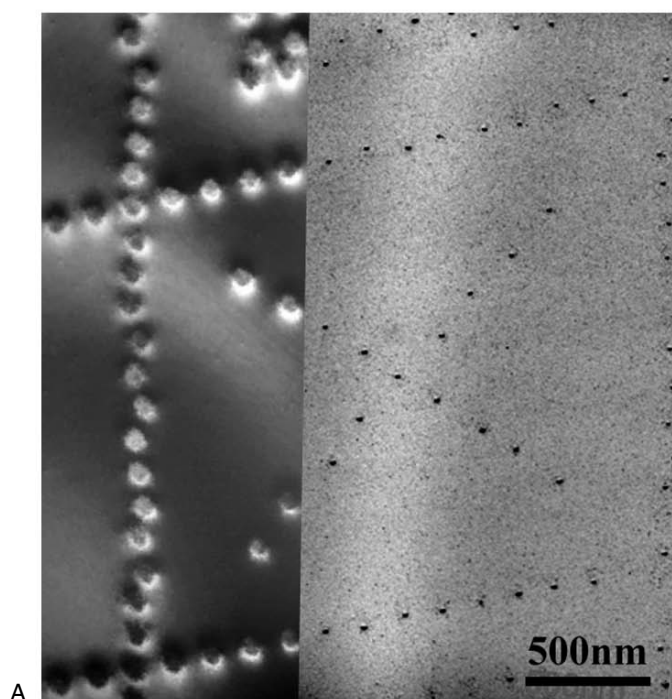
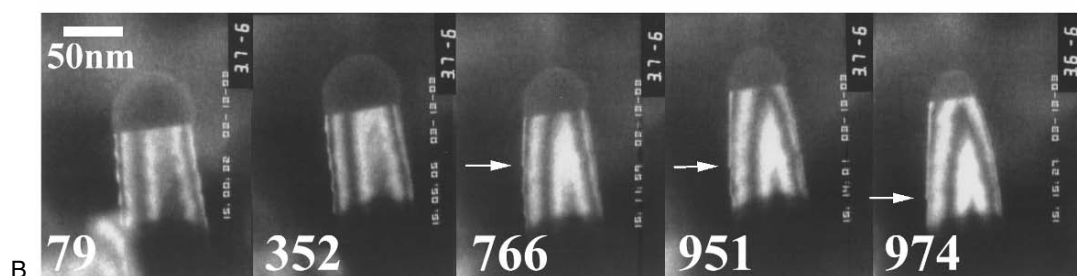


Figure 6–18. Growth on patterned substrates. (A) Bright field image of a Si substrate with $g = 220$, recorded directly after focused ion beam patterning using an irradiation time of 0.1 ms per feature (left). After annealing followed by deposition of Ge at 650°C and 5×10^{-8} Torr; ($g, 3g$) weak beam image with $g = 220$ (right), showing islands (small dots) on the irradiated areas. (Reprinted with permission from Kammler et al., 2003. American Institute of Physics.) (B) Growth of an epitaxial Si nanowire by the vapour-liquid-solid mechanism, at 620°C and 3×10^{-6} Torr Si_2H_6 with times in seconds indicated. The wire direction is (111) and the viewing direction is near (01-1) in this dark field $g = 220$ series. The Au-Si droplet is visible on the end of the wire. As growth continues, the droplet shrinks due to diffusion of Au down the wire surface, causing the wire diameter to decrease with it.



reflection geometry, however, growth kinetics can be measured quantitatively (Figure 6–18B). Such studies demonstrate effects that are not expected from the basic growth model, such as surface faceting (Ross et al., 2005) and catalyst diffusion during growth (Hannon et al., 2006).

3.7 Summary

In this section we have highlighted insights in surface physics and crystal growth derived from *in situ* TEM. Experiments in this field, while often requiring dedicated equipment with complex additional preparation chambers, provide information that is difficult or impossible to obtain using other techniques. Continuous and direct observation *in situ* avoids artifacts arising from *ex situ* observation, which may be especially significant for growth studies. It is particularly encouraging that realistic growth conditions can be accessed in the TEM, enabling *in situ* studies of, for example, catalysis or CVD to be related

to their counterparts in the outside world. Indeed, the relevance of *in situ* TEM to the catalyst and microelectronics industries is shown by the presence of environmental microscopes in several industrial laboratories.

Opportunities clearly exist for further study. Growth on patterned surfaces, prepared if necessary with integrated processing tools in the TEM system, seems a particularly exciting way to examine the fundamental processes controlling growth and to fabricate nanoscale objects with particular properties. Another potentially significant area is the correlation of structural or kinetic *in situ* measurements with macroscopic properties such as stress. Stress may be measured by depositing onto a crystalline membrane whose curvature can be observed. This may provide insight into the important processes associated with coalescence and grain boundary motion during polycrystalline film growth, as well as stress relaxation mechanisms in epitaxial films.

4 Magnetic, Ferroelectric, and Superconducting Materials

It is happily convenient that magnetic and ferroelectric domain walls and magnetic flux vortices can be visualised relatively easily in the TEM using several different imaging modes. This provides us with the opportunity to study the dynamics of ferromagnetic and ferroelectric domain switching, as well as interesting phase transformations associated with magnetic, ferroelectric, and superconducting materials. Holography or Lorentz imaging, either in Fresnel or Foucault mode, can be used to visualise magnetic domain walls and flux vortices, while ferroelectric domains can be observed through their strain field or defect structure. By applying a varying magnetic or electric field to the sample, or by changing its temperature, we can modify its magnetic or ferroelectric structure and correlate the change with the physical microstructure. Such experiments have provided insight into domain boundary dynamics, useful in modeling the workings of storage media or memory elements, as well as providing compelling illustrations of the motion of flux vortices through superconductors.

4.1 Magnetic Domain Motion in Ferromagnetic Materials

The interesting physics and industrially important applications of magnetic materials have generated a strong *in situ* experimental effort in this area. Studies have been motivated by a desire for a basic understanding of magnetic phenomena, as well as by the use of magnetic materials in storage, sensor, and other applications, where the magnetization of small regions of a material has to change controllably and reversibly many times. The most powerful experiments have related *in situ* observation of, say, the micromagnetic changes during a hysteresis cycle to the macroscopic magnetic properties. This allows us to understand, for example, the volume fraction of ferromagnetically or antiferromagnetically coupled regions in a material, or the structural reasons for an imperfect coupling of adjacent layers.

To carry out *in situ* magnetic experiments, two components are required: the ability to apply a controlled field to the specimen, and appropriate imaging capabilities for resolving its magnetic structure. Also useful is the ability to heat or cool the sample so that phenomena can be studied around the Curie temperature.

For applying a controlled magnetic field, it is preferable if the microscope is designed so that the sample sits in a field-free or low-field region. However, if a conventional microscope is used, the objective lens can be switched off, and the magnification this lens usually supplies can be achieved with the other lenses. The controlled magnetic field can then be applied using coils placed in the column or on the specimen holder. Several different designs of specimen holders incorporating coils have been developed to generate in-plane and out of plane fields, preferably without shifting the beam (for recent examples see Uhlig et al., 2003 and Yi et al., 2004). An alternative holder geometry (Park et al., 2005) uses a sharp needle made of a permanent magnet to produce a strong field near the specimen. A simpler solution that does not involve modification of the sample holder is to tilt the sample in an excited objective lens, thereby changing the field it experiences. This has been used to good effect in the experiments on superconductors described in Section 4.2.

Several imaging techniques are available for magnetic structures. These are described in detail elsewhere (Zhu and De Graef, 2001), and include holography, Lorentz imaging in the Fresnel (easier) or Foucault (higher resolution) modes, and the differential phase contrast technique. These techniques have been used extensively to examine static magnetic structures, but here we discuss only experiments where the magnetic structure is deliberately changed. Such experiments have been carried out on single crystals, thin films (often of complex multilayers), and patterned magnetic elements.

4.1.1 *Thin Film Magnetic Materials*

The motion of domain walls in polycrystalline magnetic materials has been studied for over 30 years to determine how domain walls move in perfect materials, how they interact, and how grain boundaries and other defects alter their motion by pinning. For example, macroscopic hysteresis loops have been correlated with pinning and domain size in Co/Pt (Donnet et al., 1993), and during magnetization reversals in Nd-Fe-B, the grain structure has been shown to affect the mechanism and motion of the domain boundaries (Thompson et al., 1997; Volkov and Zhu, 2000).

Interesting and complex functionalities can be obtained by fabricating multilayered materials where each layer has different magnetic properties. For example, two ferromagnetic materials separated by a conducting layer can show giant magnetoresistance properties which have applications in nonvolatile magnetic random access memory and magnetoresistive read head technologies. When one ferromagnetic layer is “pinned” by an adjacent antiferromagnetic layer, multistep hysteresis loops make the structures suitable for spin valves or spin tunnel junctions. Magnetic tunnel junctions made of two (or more)

magnetic layers separated by an insulating tunnel barrier can also be used for storage or sensing. Details of the magnetization reversal mechanism (for example, whether it occurs by domain nucleation and growth or by spin rotation), and the relationship of the hysteresis loop to the microstructure (interface roughness and grain size) are open questions in these interesting multilayers. For NiFe/Al-oxide/Co and Co/NiFe/Al-oxide/NiFe multilayers, Lorentz imaging in Fresnel mode (Yu et al., 2002a, b) showed the process of magnetization reversal by wall motion and rotation (Figure 6–19), and the relationship of grain size and texture to domain wall motion. Lim et al. (2002, 2004) examined reversal in the complex structures such as PtMn/CoFe/Ru/CoFe/Cu/CoFe/NiFe used in advanced spin valves, while Portier et al. (1999a) studied the effect of heating on Ta/NiFe/Cu/Co/MnFe/Ta films to model thermal damage in spin valves. In IrMn/CoFe films, *in situ* experiments showed that magnetization reversal must overcome two energy barriers, explaining the asymmetry in the reversal behavior and other features of the hysteresis loop, as well as the dependence of the properties on the layer thicknesses (Y.G. Wang et al., 2002). It is clear that *in situ* experiments provide a unique way of studying the properties of the ever more complex multilayers being developed for modern read heads and other applications.

4.1.2 Small Magnetic Elements

Since most applications of magnetic materials require the material to be patterned into small elements, it is naturally useful to examine the behavior of individual elements *in situ*. Typically an electron transparent substrate such as a SiN layer is used, on which magnetic elements of different shapes and sizes are patterned using electron beam lithography. The elements may either be far apart so that they can be studied in isolation, or near their neighbours so that crosstalk can be observed. It is possible to perform micromagnetic simulations of the complete structures, to provide a way to understand the microstructural observations and relate them to macroscopic measurements of hysteresis loops.

A switching experiment carried out on arrays of Co elements is shown in Figure 6–20 (Volkov et al., 2004). Such experiments have allowed the switching field of several materials to be measured (e.g., K.J. Kirk et al., 1999). Permalloy (NiFe) elements have been particularly closely examined (Johnston et al., 1996, McVitie and Chapman, 1997, Schneider et al., 2001, 2002, 2003, X.X. Liu et al., 2004, Lau et al., 2005). These studies showed how the different configurations of remanent magnetization and the process of magnetization reversal depend on the aspect ratio and symmetry of the elements, as well as the temperature, and how the reversal rate of the field affects the final domain configuration. Multilayer elements can also be studied, and by examining spin valve elements, Portier and Petford-Long (2000) showed that certain end shapes allowed easier switching by enabling 360° domain walls to form. Analogous phenomena occur when arrays of holes (antidots) are patterned in an otherwise continuous film. In Permalloy, antidots alter the domain structure and reversal mechanism (Toporov et al., 2000).

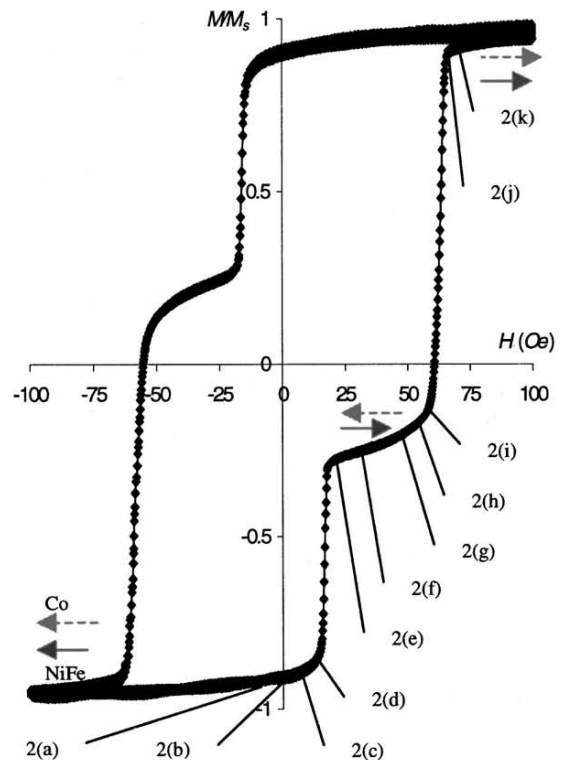
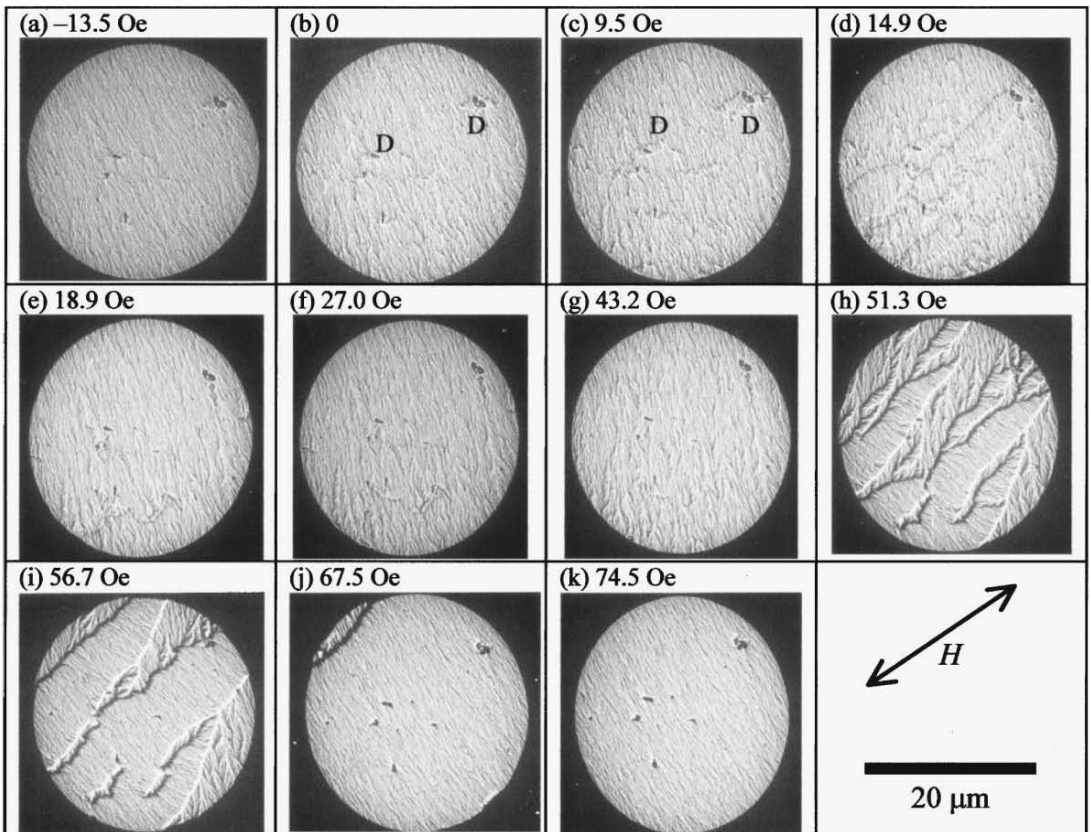


Figure 6-19. Magnetization of a multilayered film *in situ*. Lorentz Fresnel images (a–k) of the magnetization process for an NiFe/Al-oxide/Co junction film. The direction of the applied field H is indicated. All images are of the same area. Also shown is the normalized magnetization vs applied field for the same film. The corresponding domain structure at different field values along the hysteresis loop is shown. (Reprinted with permission from Yu et al., © 2002a. American Institute of Physics.)

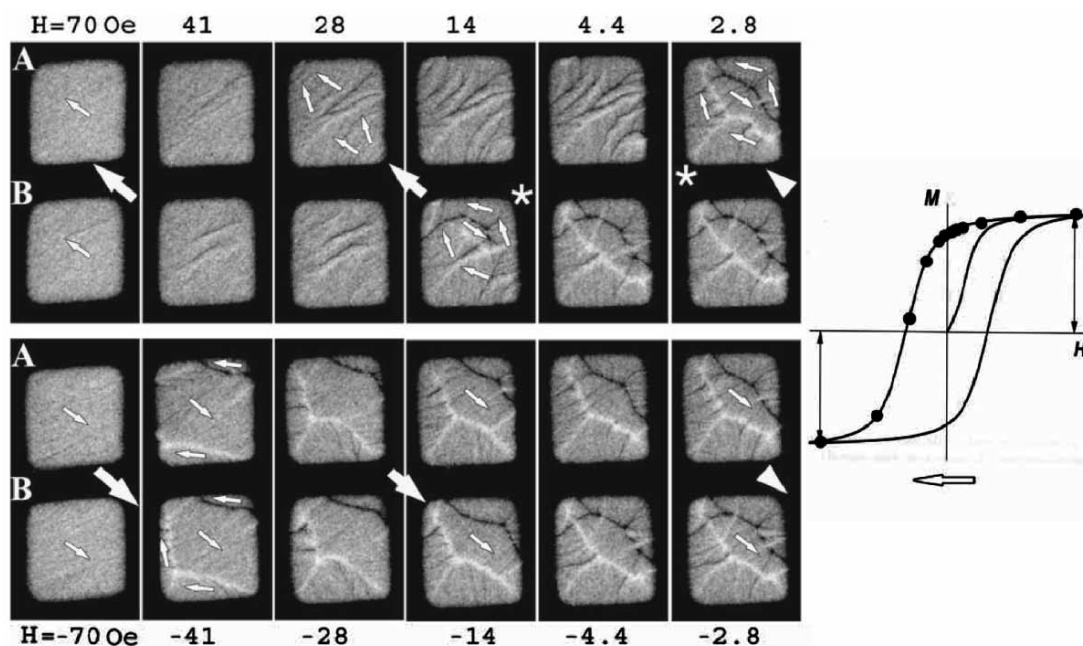


Figure 6-20. Magnetization of small elements. (A, B) *In situ* magnetization of two Co elements 25 nm thick and 6 μ m on a side on an SiN membrane. Images are shown as a function of decreasing applied field, achieved by tilting the sample in a fixed normal field excited by the objective lens (160 Oe), and images are obtained using defocused Fresnel mode. Large and small arrows show the direction of the applied field and local magnetization, respectively. The nucleation of the reverse domain is indicated with asterisks. Only the upper semiloop of the hysteresis curve is shown; the reverse process is somewhat alike but differs in the reverse switching field. The processes observed in these images are coarsening of ripples, due to coherent spin rotation; nucleation and expansion of reverse domains; wall motion and spin rotation in the remaining domains; expulsion of unfavorable boundary domains; and edge annihilation. (From Volkov et al., 2004 with kind permission of Taylor and Francis Ltd.)

An interesting extension of these *in situ* studies on magnetic elements has been the simultaneous measurement of resistivity. Using a holder in which the resistivity of a single lithographically patterned element can be measured while its magnetization is changed, Portier et al. (1997, 1998, 1999b) were able to correlate magnetoresistance with the magnetic domain structure. Studies on single spin valves made of multilayers such as NiFe/Cu/Co/NiFe/MnNi allowed magnetoresistance to be correlated with the angle between the magnetization directions in the ferromagnetic layers. It was also possible to show the mechanism of magnetization reversal, and to demonstrate the effect of stray-field coupling, which introduces edge domains, on the reversal mechanism.

4.1.3 Phase Transitions in Magnetic Materials

The industrial application of materials with giant magnetoresistance has stimulated study of an unusual class of materials which show colossal magnetoresistance. These are certain manganites with the perovskite structure, such as LaSrMnO_3 , which have both ferromag-

netic and paramagnetic phases. The phase transition and mixed phase regions in these materials can be studied *in situ* in a cooling stage using electron holography. The nucleation of ferromagnetic domains can be observed on cooling (Yoo et al., 2004a), while in the mixed phase region, application of a magnetic field creates channels connecting ferromagnetic regions, thereby changing the conductivity (Yoo et al., 2004b). These studies help to explain details of the mechanism of colossal magnetoresistance.

Another interesting class of magnetic materials is ferromagnetic shape memory alloys such as Ni_2MnGa and CoNiAl . In these materials, the shape change may be induced by applying a magnetic field. Again, *in situ* TEM using Lorentz imaging and a cooling stage (Murakami et al., 2002, Park et al., 2003) allows the correlation of the magnetic domain structure with the grain structure in these materials. Phase transformations in Ni_2MnGa are shown in Figure 6–21. In experiments like this on bulk materials, the effects of varying sample thickness on domain motion should be considered in order to obtain the most quantitative results.

4.2 Superconducting Materials

In situ TEM has provided a fascinating glimpse into the dynamics of superconducting materials. Over the last several years, several groups, most notably that of Tonomura and coworkers, have observed the presence and dynamics of vortices in superconductors using *in situ* techniques. Single vortices, with their magnetic flux of $h/2e$, have an observable effect on the phase of the imaging electrons. Thus Lorentz microscopy or holographic techniques can be used to determine their positions and characteristics. A medium or high voltage TEM with a cooling stage is used for these studies and the magnetic field can be conveniently applied to the sample by tilting the sample in the existing field of the objective lens.

These experiments have provided unique insights into the behavior of superconducting materials, in particular concerning the formation of vortex lattices, as well as vortex pinning, which must be controlled for practical applications of superconductors. Real time imaging has allowed vortex pinning and dynamics to be related to microstructural features for several superconducting materials (Figure 6–22). The motion of vortices was first imaged in Nb foils below 5 K (Figure 6–22A). The effects of grain boundaries on vortex motion were immediately visible (Harada et al., 1992, Bonevich et al., 1993; Tonomura, 2002). The role of dislocations in pinning vortices and nucleating locally ordered regions of vortices (the Abrikosov lattice) was demonstrated (Horiuchi et al., 1998). Nb specimens which had been irradiated with an FIB to produce artificial pinning centers showed fascinating vortex dynamics in which local regions of Abrikosov lattice formed and most motion took place at the boundaries of such lattices (Matsuda et al., 1996). Regular arrays of vortices could be formed with period matching the pinning point period (Harada et al., 1996a). Vortex annihilation was also observed (Harada et al., 1997) and the interactions between vortices

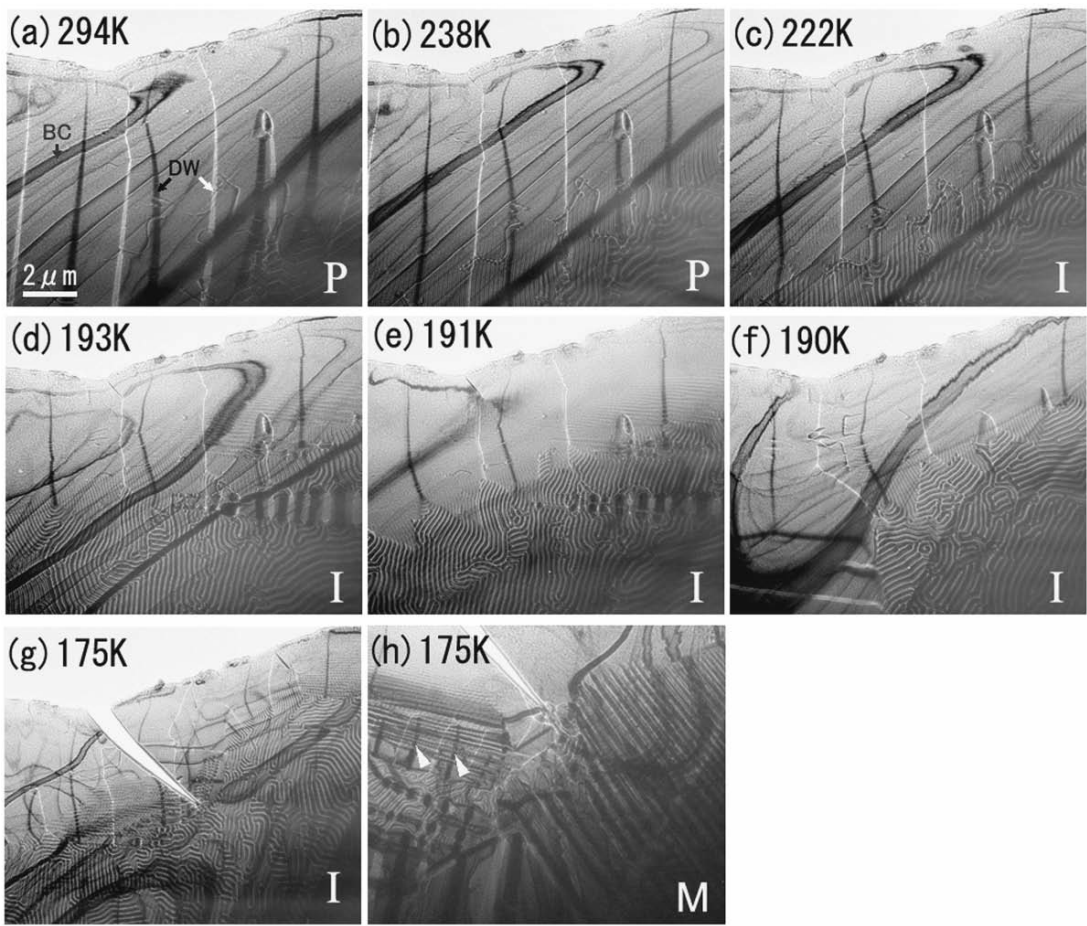


Figure 6-21. Phase transformation in a magnetic material. (a–h) Bright-field images and electron diffraction patterns of the three phases of Ni_2MnGa : the parent phase at 294K, intermediate phase at 207K, and martensite at 170K (P, I, and M, respectively). Both the crystal structure (via diffraction) and the magnetic structure (magnetic domains observed by Lorentz microscopy) are visible. BC and DW represent the bend contours and the domain walls, respectively. (Reprinted with permission from Park et al., © 2003. American Institute of Physics.)

quantified (Sow et al., 1998) by analyzing their motion through the foil. Most recently, asymmetric (one-way) motion of vortices has been controlled by FIB-patterning asymmetric channels (Togawa et al., 2005).

For high temperature superconductors, with their potentially wide range of applications, the pinning of vortices is weak and therefore particularly important to understand and control. Harada et al. (1996b) showed that, in these materials, vortex dynamics also depend on the defects present. The effects of artificial pinning centers are highly temperature dependent, giving useful insight into the different mechanisms active (Tonomura et al., 2001). Furthermore, the vortices adopt unusual chainlike arrangements in these superconductors (Matsuda et al., 2001, Tonomura et al., 2002), as shown in Figure 6-22B. It is worth noting that a high voltage (1 MeV) TEM was required to obtain the necessary resolution for studying these materials (Tonomura, 2003).

4.3 Ferroelectric Phenomena

Ferroelectric domain boundary motion due to an applied electric field or stress has applications in information storage, and the piezoelectric properties of these materials make them interesting as sensors, actuators and transducers. Two key issues in the development of ferroelectric devices are fatigue, in other words the change in boundary dynamics resulting from repeated cycling, and the effects of film thickness and electrode material on boundary dynamics. Polarized optical

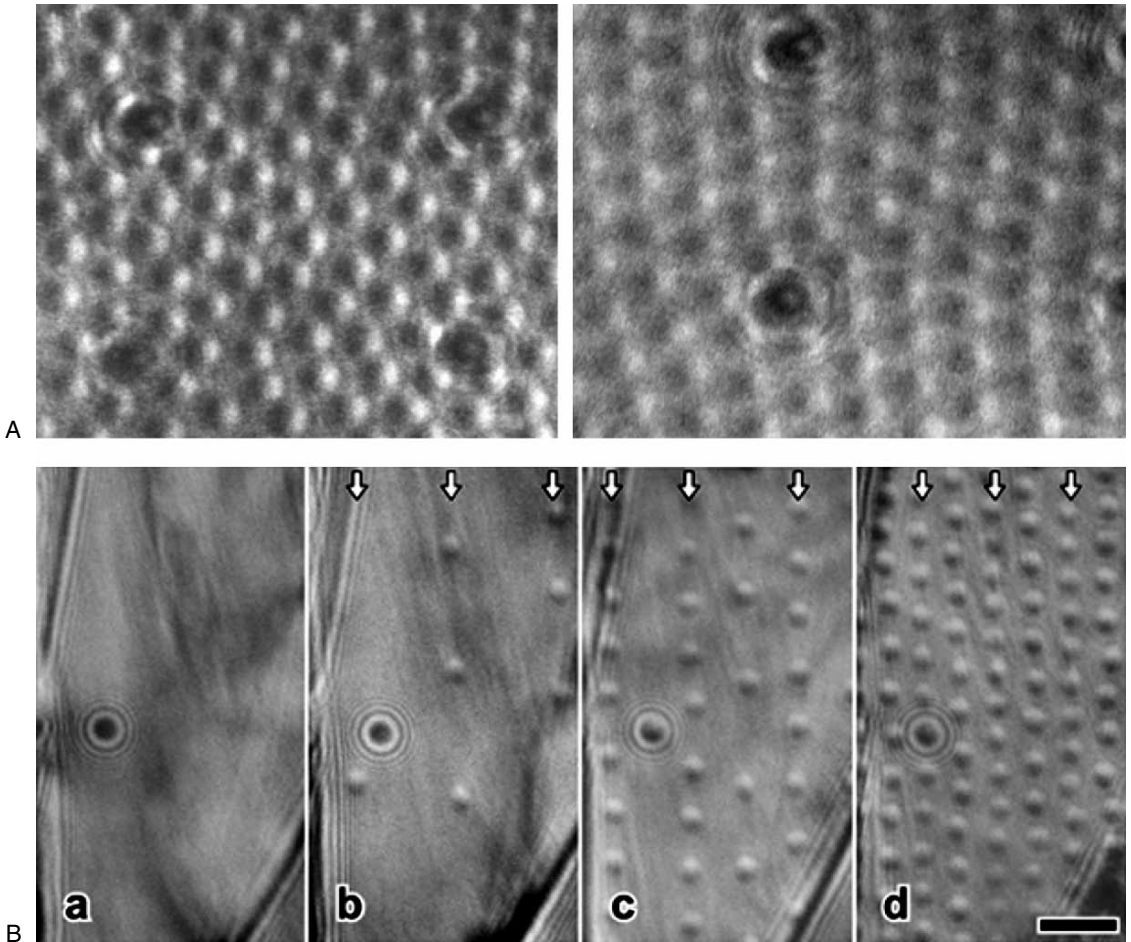


Figure 6-22. Vortex motion in superconductors. (A) Pinning of vortices in Nb. Typical interaction between the vortex lattice and an array of defects produced in an Nb thin foil by irradiation with a Ga⁺ focused ion beam. In the first image, obtained at 4.5K, the pinning site acts as a domain boundary, while at higher temperatures (8K, lower image) pinning points typically act as cores of edge dislocations in the lattice. (From Harada et al., 1997b with permission from Elsevier.) (B) A series of Lorentz micrographs of vortices in field-cooled Bi-2212 film at 50 K when a magnetic field B_z perpendicular to the layer plane is applied. The in-plane magnetic field B_x is fixed at 5 mT. (a) $B_z = 0$ (b) $B_z = 0.02$ mT. (c) $B_z = 0.1$ mT. (d) $B_z = 0.17$ mT. The arrangement of the vertical vortices in chains is marked with arrows. These chains are caused by interactions with horizontal vortices produced by the in-plane field. (Reprinted with permission from Tonomura et al., © 2002 by the American Physical Society.)

microscopy and AFM have been used successfully to investigate the overall features of boundary motion, but naturally TEM is unparalleled in its ability to relate the microstructural features within the material to boundary motion. *In situ* experiments are usually carried out by using a specimen holder with electrical connections to apply an electric field, although domain motion during heating and straining has also been studied. A heating biasing holder is preferable as it allows domain dynamics to be studied at different distances from the transition temperature.

Ferroelectric biasing has been carried out both on mechanically thinned polycrystalline or single crystal samples, and on thin films deposited on a substrate. Depending on the material geometry, the electrical contacts may either both be on top of the sample or one on each surface. As in other experiments, for quantitative analysis a well controlled specimen and field geometry is important. Changes in sample thickness may change the area of domain walls and therefore influence kinetics. Furthermore, as we show below, defects affect wall motion, so the defects introduced during sample preparation must be minimized. For all these reasons, experiments on bulk materials may provide more qualitative information, whereas thin films, especially on substrates which can be made into electron transparent membranes with a controlled electrode geometry and with minimal processing, provide the best opportunity for quantitative results. For example, thin film studies provide the opportunity to understand the “dead layer,” in which surface pinning retards domain motion.

Domain motion has been observed under electron beam heating, for example in $\text{K}(\text{Ta},\text{Nb})\text{O}_3$ (Xu et al., 1993), and during straining, for example in ferroelastic zirconia (Baufeld et al., 1997). Most studies, however, have used *in situ* biasing or controlled heating to achieve domain motion. The most detailed results have come from studies of BaTiO_3 and related materials. For example, Ren et al. (1994) observed domain wall motion in PbTiO_3 , while Snoeck et al. (1994) observed domain growth in BaTiO_3 by tip motion and then by lateral wall motion, and noted defects along the domain boundaries. Krishnan et al. (1999, 2000, 2002; Figure 6–23) observed different modes of domain wall motion in BaTiO_3 and KNbO_3 under heating, biasing, and UV irradiation. These studies showed that the motion of 90° boundaries depends on their curvature and on locking interactions with neighboring domains, and that motion may occur by rippling rather than rigidly. Interestingly, images showed the presence of trapped charge at curved or tilted boundaries or at domain tips. The buildup of charge at boundaries observed *in situ* may be important in fatigue.

In relaxor ferroelectrics such as $\text{Pb}(\text{Mg},\text{Nb})\text{O}_3\text{-PbTiO}_3$, cracking is important in piezoelectric applications. The structures of domain wall intersections has been characterized in these materials (Tan and Shang, 2004a, b; Tan et al., 2005). Cycling the electric field causes cracking in the TEM specimens, thereby providing information on crack propagation pathways along domain walls (Xu et al., 2000; Tan et al., 2000, 2005).

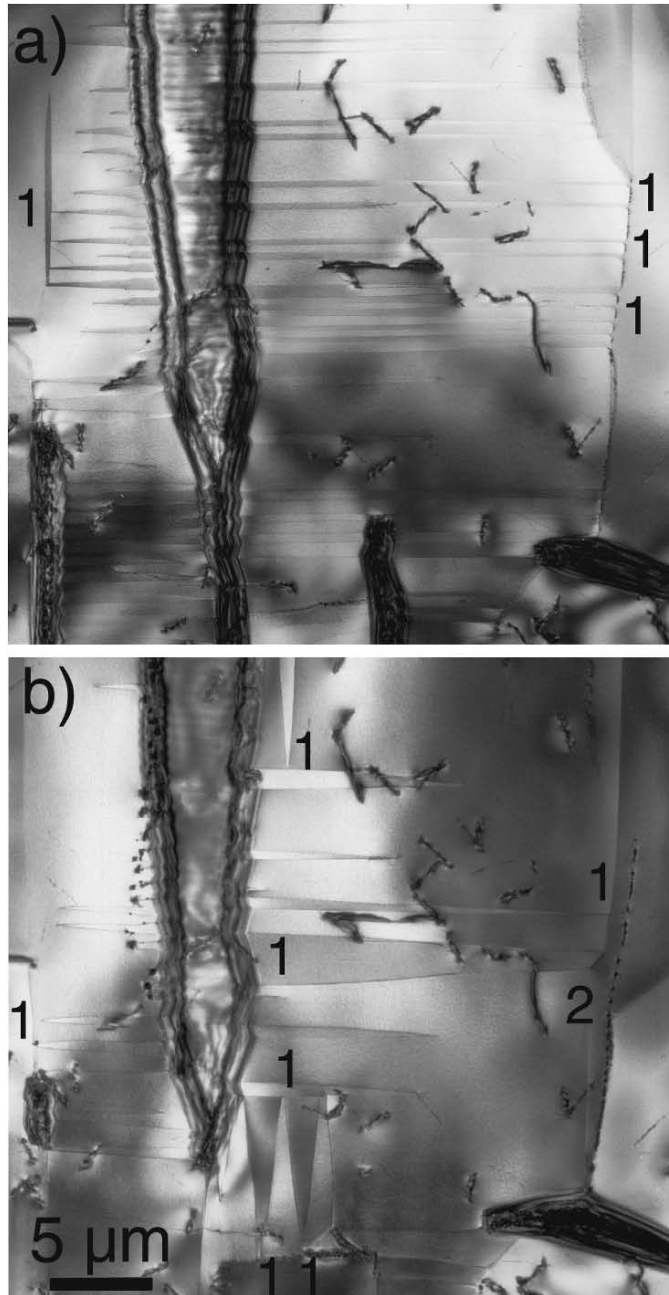


Figure 6–23. Domain wall motion in ferroelectrics under biasing. Bright field images of single crystal KNbO_3 viewed close to the $[010]$ direction. (a) Before applying an E-field. (b) After applying the field. Initially, numerous needle-like 90° domain walls as well as dislocations (dark wiggly features) and other domain wall geometries (vertical stripes) exists. On application of the field, the needle like walls move readily, coarsening the domain structure and forming straighter walls which then move less easily. This is because curved and tilted 90° domain walls (indicated by 1) must support a geometrically required Maxwellian displacement charge hence experience a direct force from an applied field, while charge-neutral domain wall regions (indicated by 2) do not experience a direct force. This suggests an intrinsic mechanism of fatigue. (Reprinted with permission from Krishnan et al., © 2000. American Institute of Physics.)

By analogy with the magnetic phase transitions described in Section 4.1.3, ferroelectric phase transitions may be examined by applying an electric field *in situ*. An example is the electric field-induced transformation of incommensurate modulations in Sn-modified $\text{Pb}(\text{Zr,Ti})\text{O}_3$ (He and Tan, 2004, 2005). Other ferroelectric materials, such as $\text{Ba}_2\text{NaNb}_5\text{O}_{15}$, show interesting incommensurate phases on heating or cooling, and these have been studied *in situ* for several decades. For $\text{Ba}_2\text{NaNb}_5\text{O}_{15}$, the incommensurate to tetragonal phase transition has been examined in diffraction (Pan et al., 1985; S. Mori et al., 1997). The nucleation of the incommensurate phase can be investigated (Verwerft et al., 1988) as well as the effect of beam-induced defects on the phase transformation (Barre et al., 1991).

We have seen that many ferroelectric phenomena have been examined using electric field and/or hot stage TEM. But it is worth noting that, in comparison with magnetic studies of reversal phenomena, the study of ferroelectric materials has not been as quantitative. As with the magnetic studies, it may be possible to combine simulations with *in situ* studies on patterned ferroelectric elements (Lin et al., 1998), perhaps prepared using focused ion beam processing, to obtain a more detailed understanding of switching and phenomena such as surface effects and fatigue.

4.4 Summary

In situ experiments have provided a persuasive demonstration of the dynamic processes within superconducting, magnetic, and ferroelectric materials: domain motion, flux vortex arrays, and pinning. In the future, where applications of nanostructured ferroelectric and ferromagnetic materials in storage and microelectronic devices may become even more important, *in situ* techniques are ideally suited to analyze switching and fatigue effects. This will be especially useful in combination with nanoscale patterning to create well-defined geometries. We envisage interesting results if such studies are combined with high speed imaging techniques as well as 3D analysis techniques, such as tomography, for boundary or vortex configurations. However, it is also worth noting that real materials are buried within structures, giving boundary conditions for the electric, magnetic, and strain fields which are different from those found in a TEM foil. Care must be taken that this does not limit the utility of *in situ* TEM. For modeling real life systems, these effects must be included, perhaps by imaging at higher voltage or using tomographic or other analytical techniques to pick out the material of interest from within a complex structure.

5 Elastic and Plastic Deformation

TEM is well suited for studying the mechanical properties of materials. Under appropriate conditions, TEM is very sensitive to lattice distortion, allowing observation of both elastic and plastic deformation via strain fields. *In situ* deformation studies aim to impose a known stress

on a sample and measure the response quantitatively. This may be achieved using a straining stage to pull the specimen uniaxially or biaxially with piezo or mechanical linkages. Alternatively, stress can be imposed by heating samples with a thermal expansion mismatch or an inbuilt stress, such as epitaxial films. Straining can even be carried out in a controlled atmosphere to simulate phenomena such as hydrogen embrittlement. The sample itself may be a thin film, a nanostructure or a bulk material, perhaps notched to initiate cracks. The information from *in situ* experiments is important because deformation is a bulk process and so requires a technique that can see deep into the specimen; surface techniques can not give the whole picture, even if the signature of dislocations can sometimes be seen on the surface. Of course TEM samples are thin in the beam direction so have free surfaces nearby, a factor that must be considered in interpreting results. This can actually be put to good use, though, in studies of tribology or the deformation of individual nanostructures *in situ*.

5.1 Microscopic Phenomena during Deformation of Bulk Materials

Structural changes during deformation, such as grain boundary motion, dislocation motion, and cracking have been studied *in situ* for just about every class of bulk material. These experiments have yielded useful information on dislocation interactions, pinning, the transfer of strain across boundaries, and the effects of temperature. This area has a history going right back to the start of high voltage microscopy, and a review of the first, pioneering work can be found in Butler and Hale (1981). Here we discuss some recent studies on metals, ceramics, semiconductors and alloys. Most experiments require straining stages which are also capable of heating, up to very high temperatures in some cases. Design rules for such extreme heating and straining holders are discussed by Messerschmidt et al. (1998) and Komatsu et al. (1994). Other studies examine the mechanical response of materials to irradiation and will be discussed in Section 8.

5.1.1 Deformation Phenomena in Single Crystals and Polycrystalline Materials

For single crystals, it is possible to relate an applied stress to the motion and interactions of specific types of dislocations. *In situ* observations provide detailed measurements of dislocation generation and multiplication mechanisms, as well as geometry, dissociation, interactions, and slip systems (Figure 6–24A). Studies on polycrystalline materials, on the other hand, allow us to understand the important process of strain transmission across grain boundaries. Heating, straining, and heating/straining experiments have been the focus of several groups, and comprehensive reviews are available (Messerschmidt et al., 1997; Mori, 1998; Pettinari et al., 2001; Vanderschaeve et al., 2001; Messerschmidt, 2003).

We first consider metals and alloys. Dislocation motion has been studied in single crystals such as Mg, Cu, and Al at moderate temperature, while refractory metals require high temperatures (Mori, 1998).

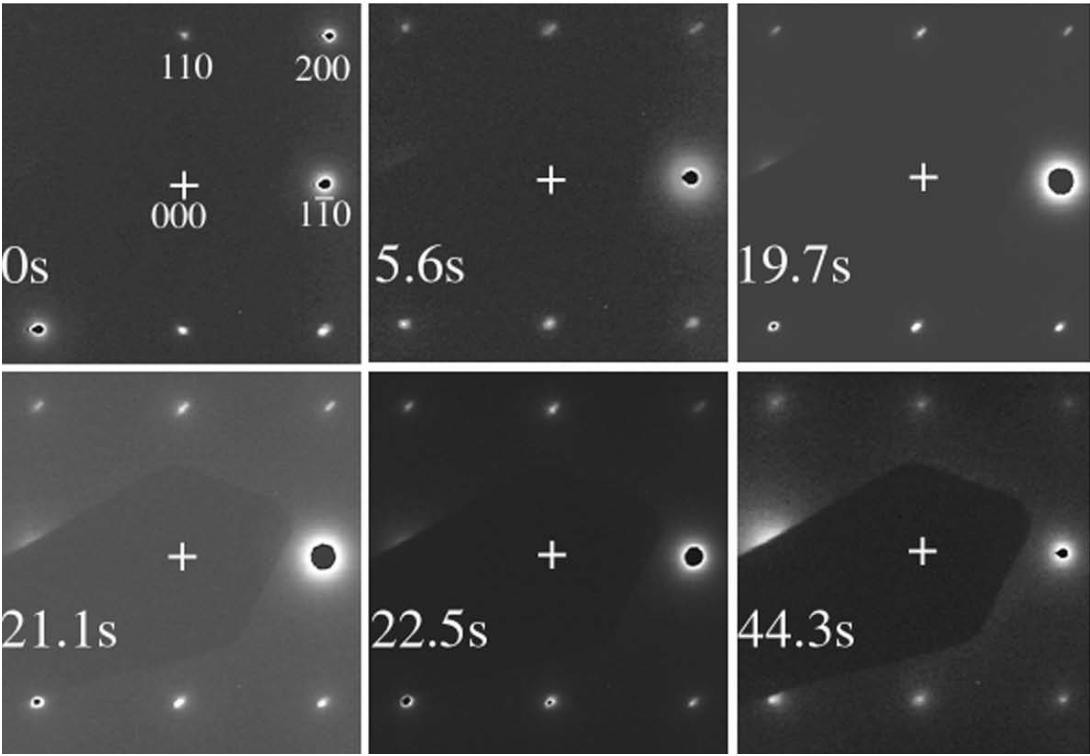
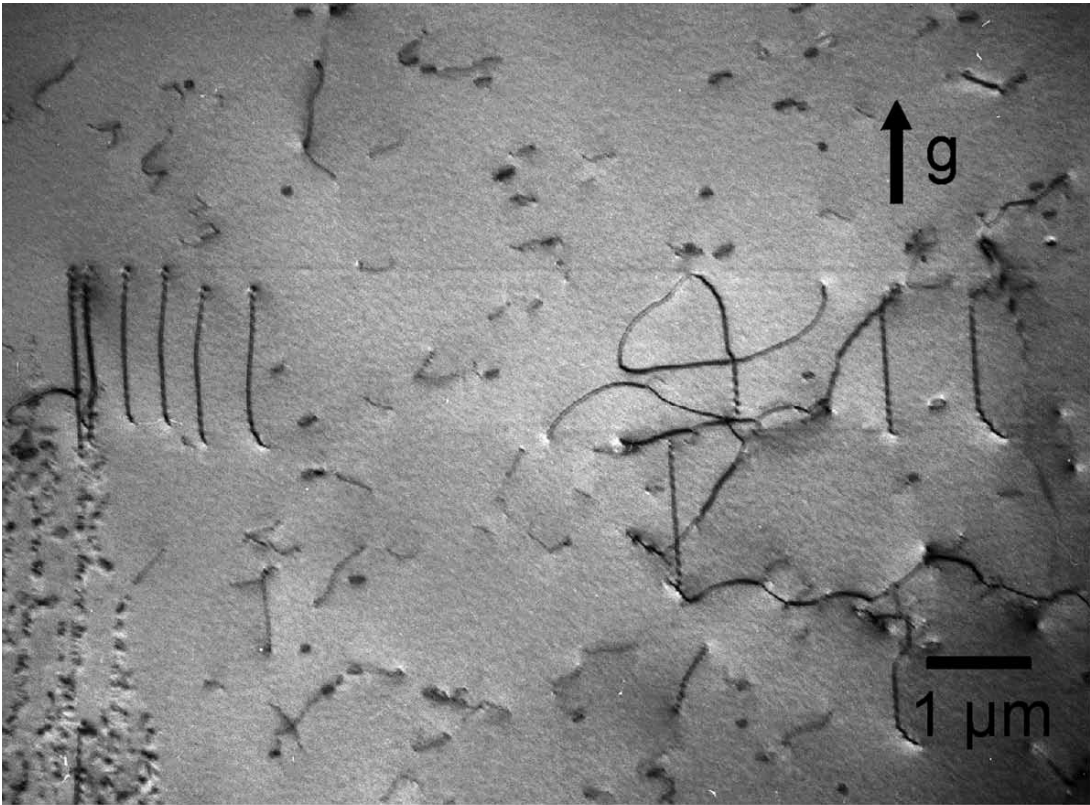


Figure 6–24. Deformation of bulk crystals. (A) Dislocations moving away from a localized source during *in situ* deformation of a MoSi_2 single crystal at $\sim 900^\circ\text{C}$. The loading axis is $[201]$, which is a soft orientation, and dislocations of Burgers vectors $1/2 \langle 111 \rangle$ are activated on $\{110\}$ planes. The traces of these slip planes run horizontally. Owing to the generation of these dislocations in localized sources, as shown here, planar slip dominates. The foil plane is (010) and $g = [220]$. (From Messerschmidt et al., © 1998. Courtesy of Cambridge University Press.) (B) Deformation of a refractory metal, V, without formation of dislocations. Diffraction patterns are obtained near the crack tip during deformation of a notched foil. As the internal stresses increase, the spots expand perpendicularly to the crack so appear elongated. (The central spot is masked to prevent overexposure.) Images do not show clusters of point defects in V, unlike the case of Au strained under the same conditions. (From Komatsu et al., 2003 with permission from Elsevier.)

Most studies apply a uniaxial stress, although, for example, cyclic shearing of Al single crystals has been described (Kassner et al., 1997). Plastic deformation does not always require dislocation generation or motion, and in V, Mo, and other bcc metals, deformation may occur by formation of point defects rather than dislocations. This too can be observed *in situ* with appropriate diffraction and imaging techniques (Komatsu et al., 2003; Figure 6–24B). The challenge in all these straining experiments is to make them quantitative. One interesting technique for the measurement of local stress is through the curvature of dislocations between pinning points (Pettinari et al., 2001).

As an intermediate case between single crystal and polycrystalline metals, bicrystals provide the opportunity to apply the precision of single crystal experiments to studies of strain transfer across boundaries. For example, in symmetric $\Sigma 3$ Fe-4 at.% Si bicrystals, Gemperlova et al. (2002) determined the primary slip system and studied the particular slip systems which can transmit stress across the boundary.

For polycrystalline metals, deformation can be investigated for thin films by either straining or heating. TEM is well suited to these thin film studies, although the sample geometry must still be controlled (for example with respect to thickness uniformity) to avoid artifacts, and surface passivation can affect the results. Straining experiments have shown, for example, that the failure mode of polycrystalline Ni films depends on grain boundary structure (Hugo et al., 2003). Thermal cycling experiments show grain boundary motion (see Section 2.1.3) and the interactions of dislocations with grain boundaries and precipitates (e.g., Kaouache et al., 2003). For thin Cu and Al films, the higher yield stress known for these materials has been related to particular types of dislocation motion (Dehm and Arzt, 2000; Balk et al., 2003). Threading dislocation motion and the effect of the passivation layer have also been studied (Keller-Flaig et al., 1999; Legros et al., 2002). Recent progress with nanoindentation stages (Section 5.3.1) and the development of specimens incorporating microelectromechanical (MEMS) free-standing films (Section 5.1.2) promise further developments in understanding the deformation of polycrystalline films in the future.

We now consider intermetallics, which have been examined *in situ* to understand the mechanism of their high temperature toughness and phase transformations. Generally, dislocation dynamics change as a

function of temperature in these materials (e.g., Haussler et al., 1999). Dislocation interactions, cross slip, and nucleation of loops control work hardening, ductility, and the critical resolved shear stress (Legros et al., 1996, 1997; Legros and Caillard, 2001). Dislocation motion during creep has also been examined (Malaplate et al., 2004).

Polysynthetically twinned TiAl crystals have lamellar structures with well-defined boundaries. By analogy with the bicrystal experiments, strain propagation across these boundaries can be quantified *in situ* (Zghal et al., 2001; Pyo and Kim, 2005). Transformations in shape memory alloys have already been mentioned in Section 2, and *in situ* straining experiments relate the microstructural changes during deformation to the stress-strain response (Jiang et al., 1997; Dutkiewicz et al., 1995; Gao et al., 1996). The cracking of intermetallics is also important in applications. Dislocations emitted at a crack tip can be analyzed (in NiAl; Caillard et al., 1999), and amorphization can be detected at crack tips (in NiTi ordered alloys; Watanabe et al., 2002).

Quasicrystals have interesting mechanical properties and dislocation geometry. *In situ* heating and straining experiments have led to the identification of shear systems and models for dislocation motion in these materials (Messerschmidt et al., 1999; Messerschmidt, 2001; Caillard et al., 2002; Momprou et al., 2004; Bartsch et al., 2005).

Ceramics have been important subjects of study since the 1970s, with an interest in comparing dislocation motion at lower and higher temperatures (e.g., in zirconia, Messerschmidt et al., 1997; alumina, Komatsu et al., 1994; MoSi₂, Guder et al., 2002). In quartz, strain-induced phase transformations have been studied by combined straining/heating experiments (Snoeck and Roucau, 1992).

We finally consider semiconductors. Kinetic studies show that dislocation motion is consistent with glide governed by the Peierls mechanism (Vanderschaeve et al., 2001). From dislocation motion and pinning, the kink mean free path and formation and migration energies can be determined (Gottschalk et al., 1993; Vanderschaeve et al., 2000; Kruml et al., 2002). Radiation enhances the glide of dislocations by changing some of these parameters (Section 8). Most studies have used low resolution, dark field imaging to record dislocation motion, but it is actually possible to observe directly the thermal motion of kinks in Si using a high resolution forbidden reflection imaging technique (Kolar et al., 1996). This generates fascinating information. The kink formation energy and unpinning barrier can be derived from the distribution and pinning of individual kinks, and videos show directly that kink migration is the rate limiting step in dislocation motion. This technique should be applicable to other materials and is expected to lead to further advances in our understanding of dislocation motion (Spence et al., 2006).

5.1.2 Deformation of Multiphase, Composite, or Layered Materials

Deformation experiments in multiphase materials, such as dispersion strengthened alloys, are particularly important in materials development, since they show how strain is transmitted between components. Dispersion strengthened Al alloys have been examined extensively, to

characterize the nucleation of dislocations at precipitates and their motion past precipitates (Vetrano et al., 1997; Hattenhauer, 1994). These phenomena are critical in high temperature deformation of such alloys. Other phenomena, such as grain boundary migration, coalescence, and elimination of subboundaries, occur during dynamic continuous crystallization on loading in the TEM (Vetrano et al., 1995; Dougherty et al., 2003), with *in situ* studies helping to understand the processes. Steel

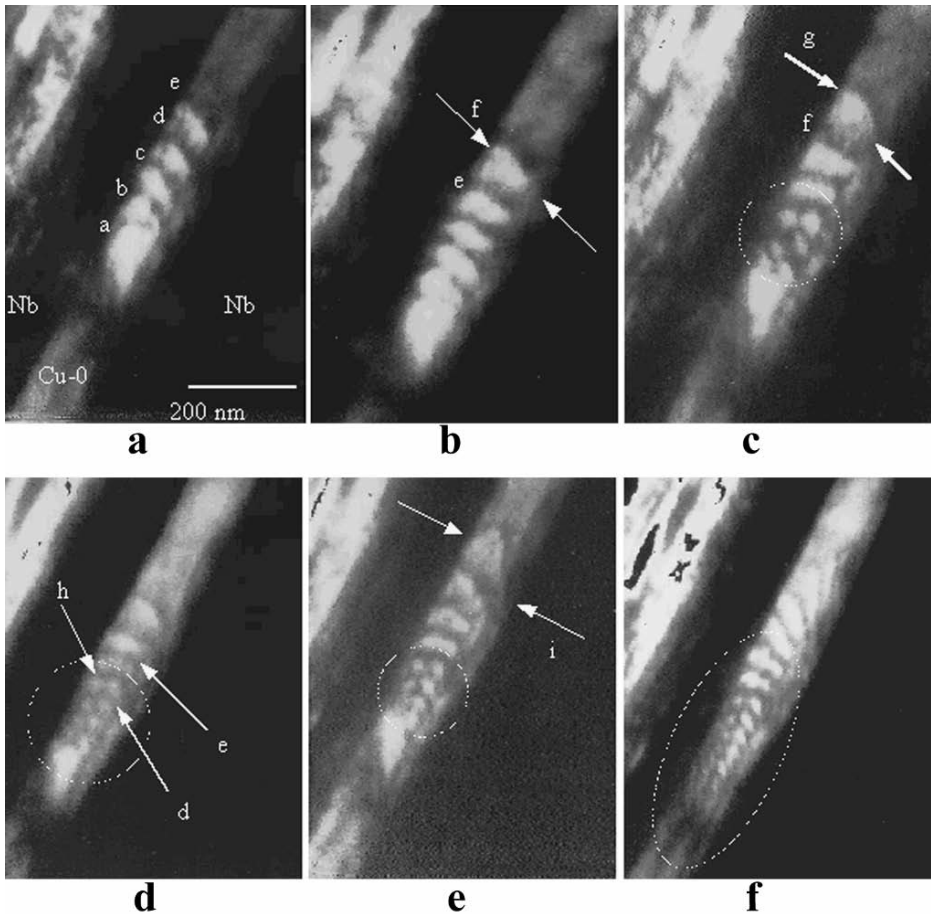


Figure 6-25. Deformation in a Cu-Nb multifilamentary composite. Nanocomposite materials are used in wires requiring high strength and high conductivity. This sample was prepared by cold drawing an Nb rod inside a Cu jacket several times until a composite was formed consisting of Cu-17%vol-Nb with 10^7 Nb filaments 40 nm across in Cu. The Cu channels have $\langle 111 \rangle$ texture, the Nb filaments have $\langle 110 \rangle$ texture, and the Cu-Nb interfaces are semicoherent. The sequence of images shows the introduction of dislocations under an applied force of 8 N. At this point Cu is plastically deformed while Nb is still elastically deforming. The area shown is a Cu channel of width about 100 nm (white) between two Nb filaments (black). At the beginning of the sequence (image a), five dislocation loops are present on parallel planes. A few seconds later the sixth and seventh loops appear (images b and c). By the end of the sequence 13 loops are present, and they have interacted to produce a honeycomb pattern (circles in images c-f). This dislocation behavior is used in a plastic flow model, which explains the very high ultimate tensile stress of fcc-bcc nanocomposite structures. (From Thilly et al., 2002, with kind permission of Taylor and Francis Ltd.)

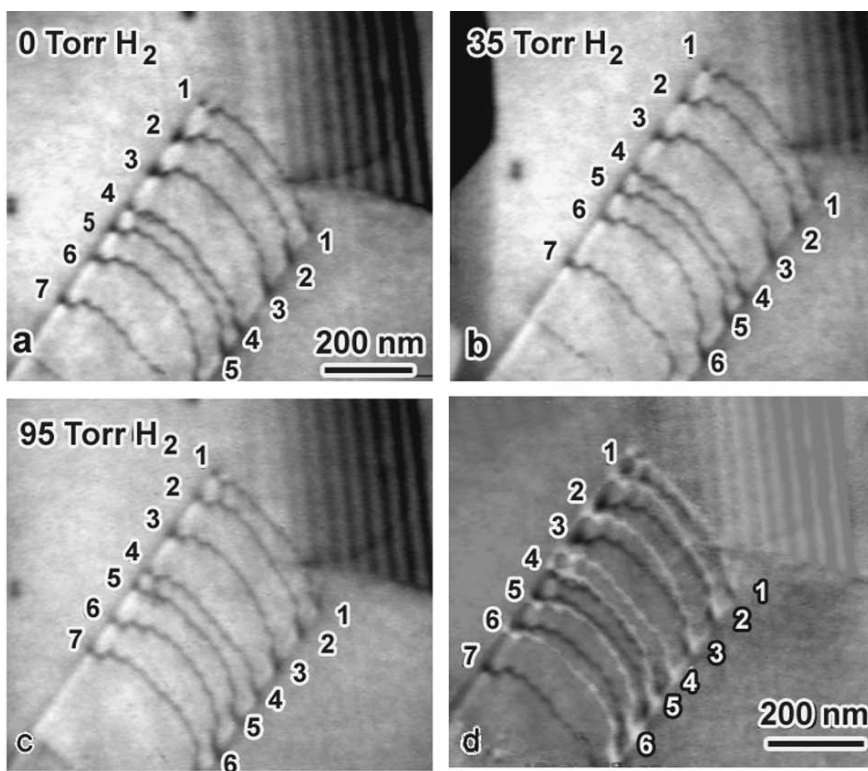


Figure 6-26. Dislocation motion in H atmosphere. Showing the effect of H on the mobility of dislocations in 310S stainless steel. (a) dislocation configuration in vacuum under constant load; (b) 35 Torr; (c) 90 Torr of hydrogen; (d) composite image formed by subtracting (c) from (a). (Reprinted with permission from Robertson and Teter, © 1998 by the American Physical Society.)

is of course another multiphase material whose behavior *in situ* has been studied for some time (Caillard et al., 1980); industrially relevant studies of dislocation motion and slip transmission through interfaces in steels continue (Verhaeghe et al., 1997; Janecek et al., 2000; Zielinski et al., 2003). Many other nanocomposite materials provide interesting examples of dislocation-interface interactions. One such is shown in Figure 6-26 (Thilly et al., 2002) and other work is reviewed by Louchet et al. (2001).

Techniques developed for multiphase materials extend easily to engineered multilayers, where the interaction of dislocations or cracks with interfaces is of interest. Cross sections of multilayers may be prepared using a FIB to form a thin section at which cracks will initiate (Wall et al., 1995; Wall and Barbee, 1997). Alternatively, multilayers (or thin films) may be integrated into a mechanical testing stage. One example is a piezo stage for cyclic loading of solder thin films, use to show that cavitation is the dominant fatigue damage mechanism (Tan et al., 2002). In this stage, a multilayered structure is fabricated which includes the material of interest as well as a piezoelectric layer.

The future for thin film and multilayer studies undoubtedly lies in extending this sort of approach, achieving controlled deformation using MEMS-based techniques. MEMS force sensors and displacement

measurement systems can be integrated with the film of interest, allowing for quantitative experiments. Stages for uniaxial tensile testing of free-standing thin films have already been demonstrated (Haque and Saif, 2003, 2004, 2005; Hattar et al., 2005), with initial results on polycrystalline Al thin films, and other functionalities such as heating are under development (Zhang et al., 2005). We anticipate further exciting progress in this area.

5.1.3 Deformation in a Controlled Atmosphere

By combining mechanical testing with controlled environment TEM, a very important objective, understanding the mechanism of hydrogen embrittlement of steels, has been achieved (Figure 6–27). *In situ* straining experiments carried out in an environmental cell showed that solute hydrogen can increase the speed at which dislocations move through the material, as well as increasing the rate of crack propagation (Robertson and Teter, 1998; Teter et al., 2001; Sofronis and Robertson, 2002). By quantitatively analyzing dislocation dynamics on adding and removing hydrogen, competing mechanisms for embrittlement could be distinguished. The hydrogen shielding model was confirmed, in which the hydrogen-enhanced localized plasticity mechanism is the shielding of the dislocation from interactions with other strain centers such as pinning points, other dislocations, or solutes. Few other materials have been strained in a controlled atmosphere, although Maeda et al. (2000) showed that the incorporation of hydrogen from a plasma increases the motion of dislocations in semiconductors. Further studies could provide a sensitive probe of dislocation properties and interactions.

5.2 Relaxation of Epitaxially Strained Materials

In this section, we consider another method of straining a material: growth of the material epitaxially on a lattice mismatched substrate. As the layer is deposited a high intrinsic stress builds up, and, once above a “critical thickness,” the layer may relax by forming a network of dislocations at the layer-substrate interface (Figure 6–27A, B). This phenomenon can be examined *in situ*. First, a layer thicker than the critical thickness is grown, but at a temperature that is too low to allow dislocation nucleation or propagation. Then a plan view TEM specimen is prepared and heated *in situ*, allowing relaxation to be observed as a function of temperature. The advantage of this type of experiment is that the stress state is very well defined, making this a situation where dislocation dynamics can be measured quantitatively. Such studies are motivated by their importance to the microelectronics industry, since the positions of individual dislocations, the density of threading arms, and the motion of dislocations during processing are known to affect device performance.

This work was carried out for the system $\text{Si}_{1-x}\text{Ge}_x$ on Si, with x ranging from about 0.05 to 0.3 (Hull et al., 1989, 1991). Measurements of dislocation velocity as a function of layer parameters are shown in Figure 6–27C. For capped layers, the results supported the diffusive kink pair model of dislocation propagation, and for uncapped layers they showed that propagation is through motion of single kinks.

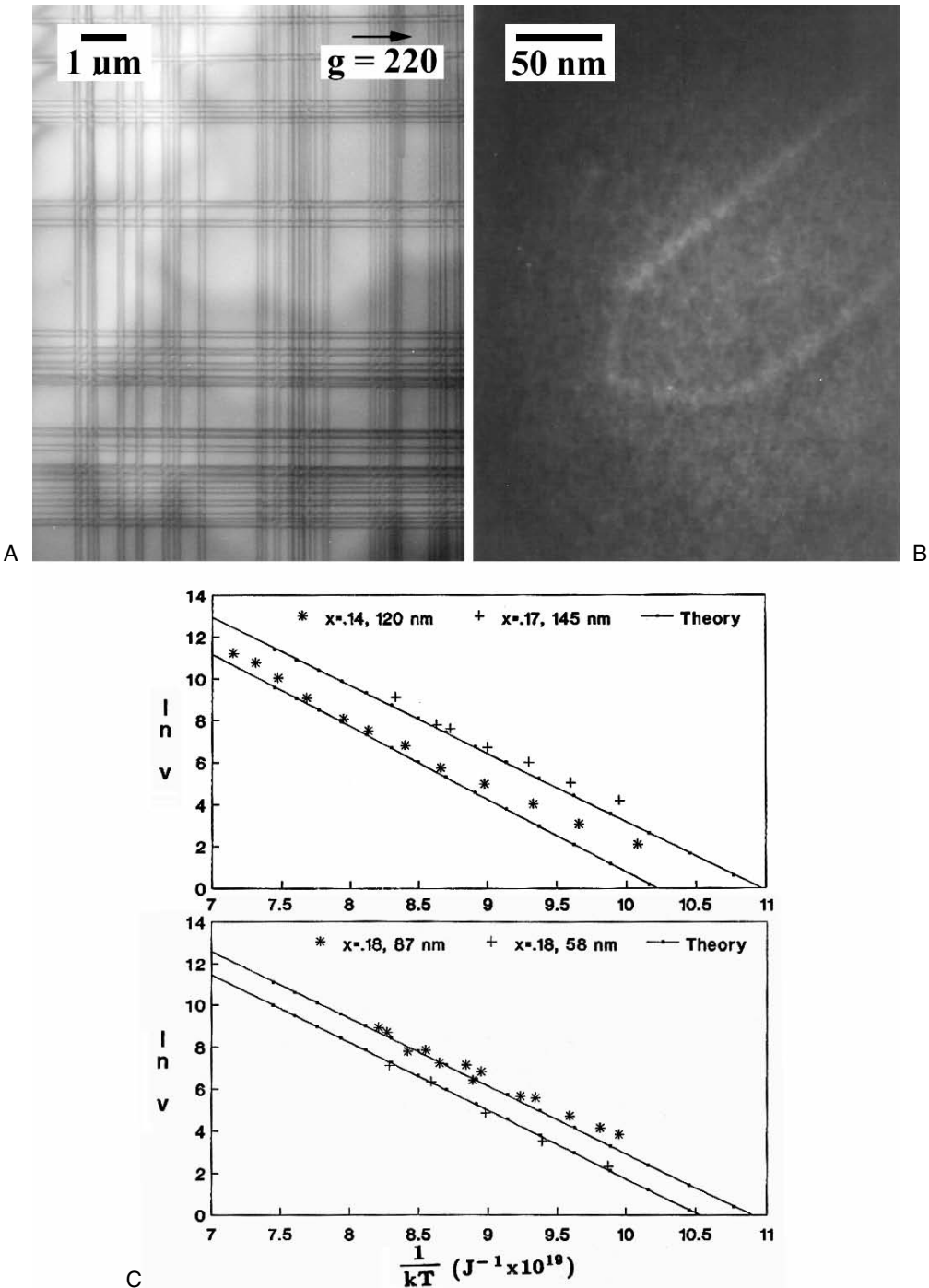


Figure 6-27. Dislocation dynamics in SiGe heterostructures. (A) Image showing the typical dislocation structure in SiGe heterostructures. Plan-view TEM image of a relaxed 300 nm Si/150 nm Si₈₀Ge₂₀/Si(001) heterostructure. Closely spaced interfacial dislocation pairs are segments of the same dislocation loop at the top and bottom SiGe/Si interfaces. (From Stach et al., © 1998b Courtesy of Cambridge University Press.) (B) 220g, 3g weak-beam dark-field micrograph of a threading dislocation segment connecting two interfacial dislocation segments. (From Stach et al., 1998b.) (C) Measured dislocation velocities compared with predictions of the double-kink theory for SiGe layers buried beneath a 300 nm Si cap, as in (A). The Ge concentrations, x , and epilayer thicknesses are given on the graphs while the dislocation velocities are given in Angstrom s^{-1} . The fitting parameter used in the model is half the double-kink nucleation energy, taken as 1.0 eV. (Reprinted with permission from Hull et al., 1991. American Institute of Physics.)

Activation energies and prefactors were measured for dislocation nucleation and propagation. Note that a well calibrated system is necessary if TEM-derived parameters such as activation energy are to be meaningful. In this case, the sample temperature was calibrated using Si regrowth (Section 2.1.1), and finite element analysis was used to determine the conditions under which sample bending could be neglected (Hull and Bean, 1994). The parameters determined *in situ* were eventually used in a processing model for microelectronic device design (Hull et al., 1989).

A similar experimental approach allows other features of the relaxation process to be examined as well. Dislocation nucleation has been studied in SiGe/Si implanted with F (Stach et al., 1998a) and with He, which forms platelets and bubbles (Follstaedt et al., 1997; Hueging et al., 2005). Although most work has been in the SiGe/Si system, a few other epitaxial systems have been studied as well, such as BaTiO₃ on SrTiO₃ (Sun et al., 2004), ZnSe on GaAs (Lavagne et al., 2001), and Al on Al₂O₃ (Dehm et al., 2002). Clearly, many other materials systems could yield useful information with this technique.

We finally note that dislocation configurations in thin films, such as in Figure 6-27A, are somewhat random. However, in well controlled geometries—relaxed epitaxial islands, or surfaces with modulated stress fields—dislocation locations can be predicted accurately (Liu et al., 2000; Kammler et al., 2005). This suggests opportunities for *in situ* TEM to provide quantitative information on dislocation nucleation and growth in finite nanostructures with built in strain, such as epitaxial islands or composite wires or ribbons.

5.3 Mechanical Properties of Nanostructures, Thin Films, and Surfaces

TEM is particularly appropriate for studying the mechanical properties of small volumes of material. In the work discussed so far, the same stimulus (strain, heat) is applied to the entire specimen. By applying the stimulus to only a small part of the specimen, the correlation between mechanical input and structural response can hopefully be made more precise. In this section we discuss experiments in which small volumes of a specimen are mechanically deformed (nanoindentation), the surface is probed, or nanostructures are deformed. These experiments make use of scanning probe techniques to bring a tip near the sample for controlled deformation of an area under observation.

5.3.1 *In Situ Indentation of Thin Films*

By integrating a nanoindenter into a specimen holder, it is possible to examine localized deformation of single crystals, polycrystalline materials, and layered structures. Nanoindentation holders have been described by Wall and Dahmen (1998), Stach et al. (2001), Minor et al. (2001), Bobji et al. (2003) and Li et al. (2004). The design must include some way of making sure the indent occurs in an electron transparent region of the specimen. In the design of Wall and Dahmen, for example, a Si sample (mounted vertically) is etched to form a ridge which projects out into the beam. Elastic deformation followed by dislocation

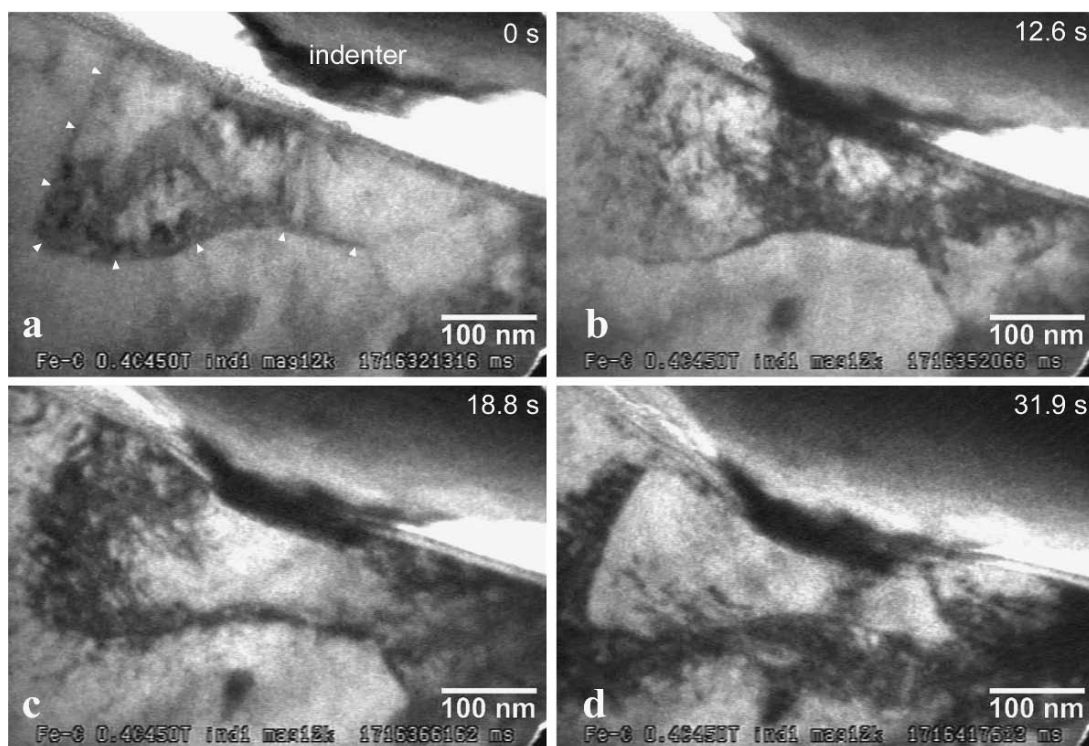


Figure 6–28. Nanoindentation of steel. Series of images of an Fe-C martensite, including a low-angle grain boundary. (a) Before indentation, (b) at 21 nm penetration depth, showing dislocation emission beneath the indenter, (c) at 46 nm penetration depth, showing dislocation pileup at the grain boundary, and (d) at 84 nm penetration depth, demonstrating dislocation emission at the far side of the grain boundary. (From Ohmura et al., ©2004. Courtesy of Institute of Materials Research Society.)

formation is visible in this ridge during indentation. Using this sample geometry, the motion of dislocations and grain boundaries under the indenter tip has been studied in a variety of materials deposited onto the substrate. In ultra fine grained Al, for example, grain boundary motion occurs under the tip (Jin et al., 2004). This does not occur in Al-Mg, suggesting that solute drag is important (Soer et al., 2004). Similarly, in Ni, deformation proceeds by diffusion-controlled grain boundary-mediated processes rather than dislocation motion within grains (Shan et al., 2004). Si shows dislocation plasticity on room temperature nanoindentation (Minor et al., 2005). Nanoindentation of martensitic steel is shown in Figure 6–28 (Ohmura et al., 2004), and the formation of dislocations and the strain transfer across grain boundaries are visible. As well as dislocation formation and motion, nanoindentation can cause crack propagation. The nature and pathway for such cracks can be determined (Li et al., 2004; Matsunaga et al., 2004). In the future, we expect to see exciting developments in the area of *in situ* nanoindentation, driven by the incorporation of sensors that allow stress-strain curves to be measured during deformation. This suggests exciting possibilities for quantitative analysis of nanoscale mechanical properties.

5.3.2 Tribology and Nanomanipulation

Nanoindentation is only one of several interesting experiments that can be carried out in holders incorporating a scanning tip with piezo-electric actuators. For example, the tip can be used for fundamental studies of tribology by scraping it across a surface. Tip-substrate interactions were observed in reflection mode geometry as a graphite specimen was imaged using STM (Spence et al., 1990; Lo and Spence, 1992), and changes in the surface were attributed to shearing and abrasion. Using a similar holder in a UHV TEM, Naitoh et al. (2000) correlated the atomic configuration at a tip surface with the resolution of STM images it produced. Fujisawa and Kizuka (2003a) and Ohnishi et al. (1998a) observed the motion of a tip across a stepped surface and determined the effects of rastering and surface topography on lateral displacement. This sort of information is naturally important in interpreting scanning probe images.

Alternatively, an *in situ* STM tip can be used to form small necks or grain boundaries by touching it to the surface, deforming both materials (Figure 6–29). The structure of these necks can then be observed at

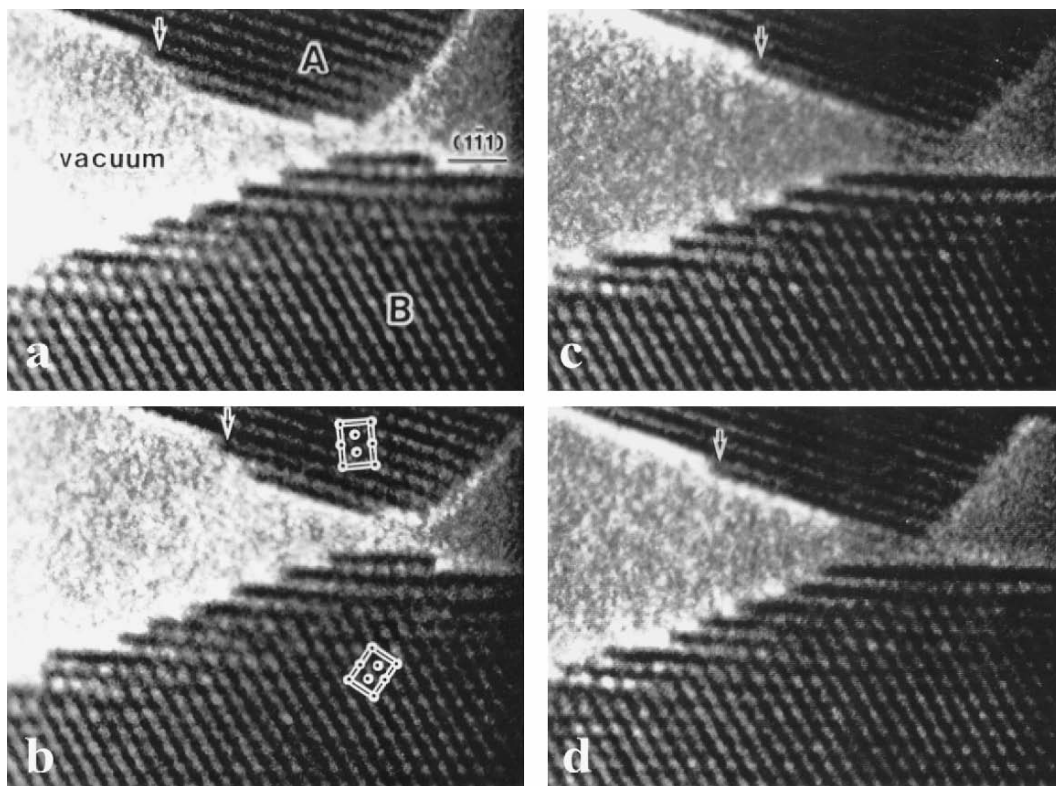


Figure 6–29. STM surface manipulation. Series of high-resolution images of atomic-scale removal type mechanical processing of an Au surface. A region six atomic columns wide on the fixed side (B) is removed by the Au tip on the mobile side (A). On both sides the beam is parallel to the [110] axis. The time is (a) 0s, (b) 1.2s, (c) 3.5s, and (d) 5.8s. Boxed circles show the unit cell of Au. These images give a visual impression of events that may occur during STM operation as well as phenomena associated with friction. (From Kizuka et al., © 1998c. Courtesy of Cambridge University Press.)

high resolution. Kizuka and Hosoki (1999) recorded the interactions as two oxidized Si tips came into contact, observing the strength of the boundaries formed, and Naitoh et al. (2000) observed twins in necks of Si formed between clean Si and a W tip. Most work, however, has involved Au. Kizuka (1998a, b) showed formation of a neck between Au contacts and observed compression, shear, deformation, slip, and twinning. These videos give a stunning visual impression of the interaction of the tip with a surface during scanning, and allow a study of friction at the nanoscale (Fujisawa and Kizuka, 2003b). Naturally, the electrical properties of such necks can also be studied, and this will be described in Section 6.

5.3.3 Mechanical Properties of Nanostructures

The mechanical properties of elongated nanostructures are a natural subject for *in situ* studies. For carbon nanotubes, mechanical parameters such as Young's modulus were first measured by observing the vibration of tubes which extend out as cantilevers (Krishnan et al., 1998). These tubes vibrate because of coupling to motion in the stage. However, the mechanical properties of tubes can also be measured by bending them with an STM tip. Individual CNTs can be bent and broken (Kuzumaki et al., 2001) and stress and strain can be measured during deformation (Asaka and Kizuka, 2005). The welding of CNTs to a tip can be seen (Hirayama et al., 2001; Kuzumaki et al., 2004). A tip can even be used to operate a bearing made of telescoping CNTs (Cumings and Zettl, 2000; Figure 6–30). Such experiments can use the

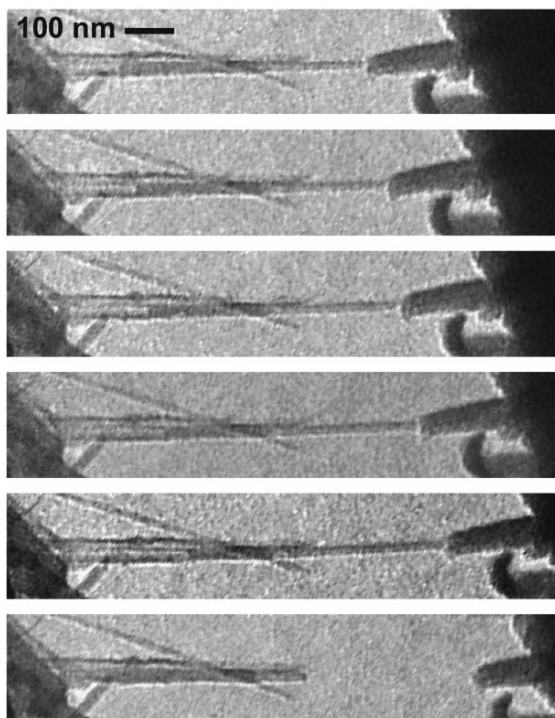


Figure 6–30. Mechanical properties of a carbon nanotube. Selected frames of a video recording of the *in situ* telescoping of a multi-walled nanotube. In the first five frames, the core nanotubes are slowly withdrawn to the right. In the sixth image, which occurred one video frame after the core was released, the core has fully retracted into the outer nanotube housing as a result of the attractive van der Waals force. (Reprinted with permission from Cumings and Zettl, © 2000, AAAS.)

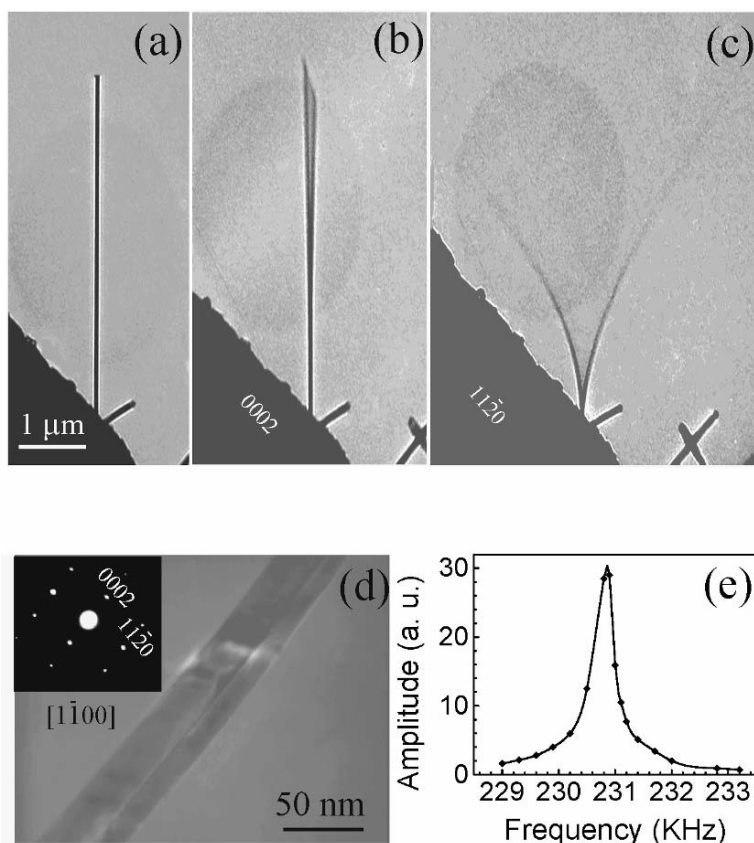


Figure 6–31. Mechanical properties of a nanobelt. A ZnO nanobelt at (a) stationary, (b) the first-harmonic resonance in the x direction, frequency 622 kHz, and (c) the first-harmonic resonance in the y direction, frequency 691 kHz. (d) An enlarged image of the nanobelt and its electron diffraction pattern (inset). The projected shape of the nanobelt is apparent. (e) The FWHM of the resonance peak measured from another ZnO nanobelt. The resonance occurs at 230.9 kHz. (Reprinted with permission from Bai et al., © 2003. American Institute of Physics.)

TEM for more than just imaging; for example, Suenaga et al. (2001) were able, using EELS, to probe changes in electronic structure during bending of multiwalled and bundled single walled CNTs, and correlate these changes with deformation.

By using a stage with piezo drives to induce an alternating electrical field between a nanowire and an electrode, controlled frequency vibrations can be set up in the nanowire (Wang et al., 2001, 2002a). For example, in ZnO wires of rectangular cross section, each direction of vibration has its own resonances from which the modulus and time constant or Q factor can be derived (Bai et al., 2003a; Figure 6–31). Such measurements of bending modulus can be related to defects in individual structures (Gao et al., 2000). In fact, the mechanical resonance depends so sensitively on the structure that modulus measurements have potential use for measuring small masses. A less quantitative application (at present) is a stage developed for stretching chains of nanoparticles, of interest for their use as reinforcing fillers and their

presence in diesel exhausts (Suh et al., 2004). It will be exciting to see these mechanical techniques applied to more complex structures such as composite nanowires or filled CNTs.

5.4 Summary

Mechanical deformation has been successfully examined *in situ* using several innovative techniques. The future challenge in mechanical testing is to make its results even more quantitative and to minimize or account for thin foil effects. Customised design of specimens using MEMS technology, integrating for example tensile testers, tweezers, notches, or other structures, will certainly become more widespread for thin film experiments. Measurement of stress-strain curves during nanoindentation will provide accurate information on the mechanical response of a system, for example as single dislocations are introduced. Deposition of a film of interest, perhaps with an engineered boundary, onto a crystalline membrane may become useful in some cases, since the substrate deformation can be measured accurately from its strain contrast. Integration of such experiments with controlled atmosphere TEM will allow materials to be tested under the most realistic conditions. Elegant results on nanostructures have already been achieved using nanomanipulation with a tip. There is no doubt that these types of experiments are already bringing our understanding of the mechanical properties of small volumes of materials to a new level of precision.

6 Correlation of Structural and Electrical Properties of Materials

We have shown that fascinating data can result from the correlation between an applied stimulus, such as temperature, environment, or strain, and a microstructural change. Extending this approach, if we measure the electrical properties of a material while also observing or changing its structure, we can investigate structure-electrical property relationships in a unique way. The material under study can have a large volume, perhaps most or all of the TEM sample. Alternatively, structure-property relationships can be derived for individual nanostructures.

These interesting experiments all require electrical biasing holders, but not necessarily video rate recording, since the sample itself may change only slowly, or not at all. Examples include measuring the conductivity of individual nanowires as a function of their diameter, measuring changes in resistance as voids form during electromigration, measuring the potential distribution across p-n junctions as a function of bias, or measuring conductivity after dislocations are introduced.

6.1 Electrical Measurements on TEM Samples: Complete Samples as Devices

Many important processes can be studied by applying a voltage across a bulk or patterned specimen *in situ*. We have already described the

use of an applied voltage to generate an electric field for ferroelectric switching (Section 4.3), and now consider other possibilities. One interesting example is the measurement of electric fields across p-n junctions. This has been a favorite topic from the early days (Darlington and Valdre, 1975). More recently, a varying bias has been applied *in situ* across a FIB-prepared p-n junction, using holography to measure the potential distribution (Twitchett et al., 2002, 2004). The eventual aim of such studies is to understand the complex fields within real transistors, although quantitative image matching shows that surface effects must first be understood. Similar biasing experiments can be useful probes of other materials too. For example, the potential distribution across the boundary in a ferroelectric bicrystal was mapped by holography (Johnson and Dravid, 1999, 2000), showing breakdown and the presence of trap states associated with dopants.

In situ biasing also allows resistivity to be measured, which can then be correlated with a structural transformation. Thus, the amorphous to crystalline transformation for phase-change memory materials (Section 2.1) can be correlated with resistivity, as shown in Figure 6–32A (Verheijen et al., 2004). Similarly, phase transformations in TiNi shape memory alloys can be related to resistivity (Ma and Komvopoulos, 2005). Another interesting application of *in situ* resistivity measurement is the observation of conductivity changes during ion or electron beam irradiation to determine the cross section and threshold energy for Frenkel pair production (King et al., 1981; Haga et al., 1986).

In situ resistivity measurements have also been used to probe the electrical properties of dislocations (Ross et al., 1993). In this experiment, dislocations are formed progressively in a p-n junction diode by heating a metastable SiGe layer (Section 5.2) *in situ*. Measurement of the diode reverse leakage current as a function of dislocation density yielded the leakage current per length of dislocation, useful in device modeling. In this experiment, however, leakage was measured through the whole device, while dislocation density was measured only in the electron transparent area. Imaging of the whole device would have been more precise. Mathes et al. (2003) were, in fact, able to thin the entire active region of a laser structure by FIB, and observed degradation by dislocation formation during device operation. Such biasing studies could be combined with laser excitation *in situ* to examine degradation mechanisms and photoplasticity.

Current flow through narrow metal lines results in electromigration. Due to its importance and its strong dependence on microstructure, electromigration has been studied intensively *in situ* (Figure 6–32B). Current is passed through an Al or Cu line patterned onto an oxide substrate, with the aim of observing void formation and dynamics at microstructural features such as grain boundaries and precipitates. There are two important experimental concerns in these experiments. Passivation of the metal surface is known from SEM studies to alter failure lifetimes (Doan et al., 2001); passivating the lines makes the experimental results more relevant to real life, but it is then harder to see void shapes and dynamics. Joule heating in the lines can be another problem (Shih and Green, 1995), especially since elevated temperatures

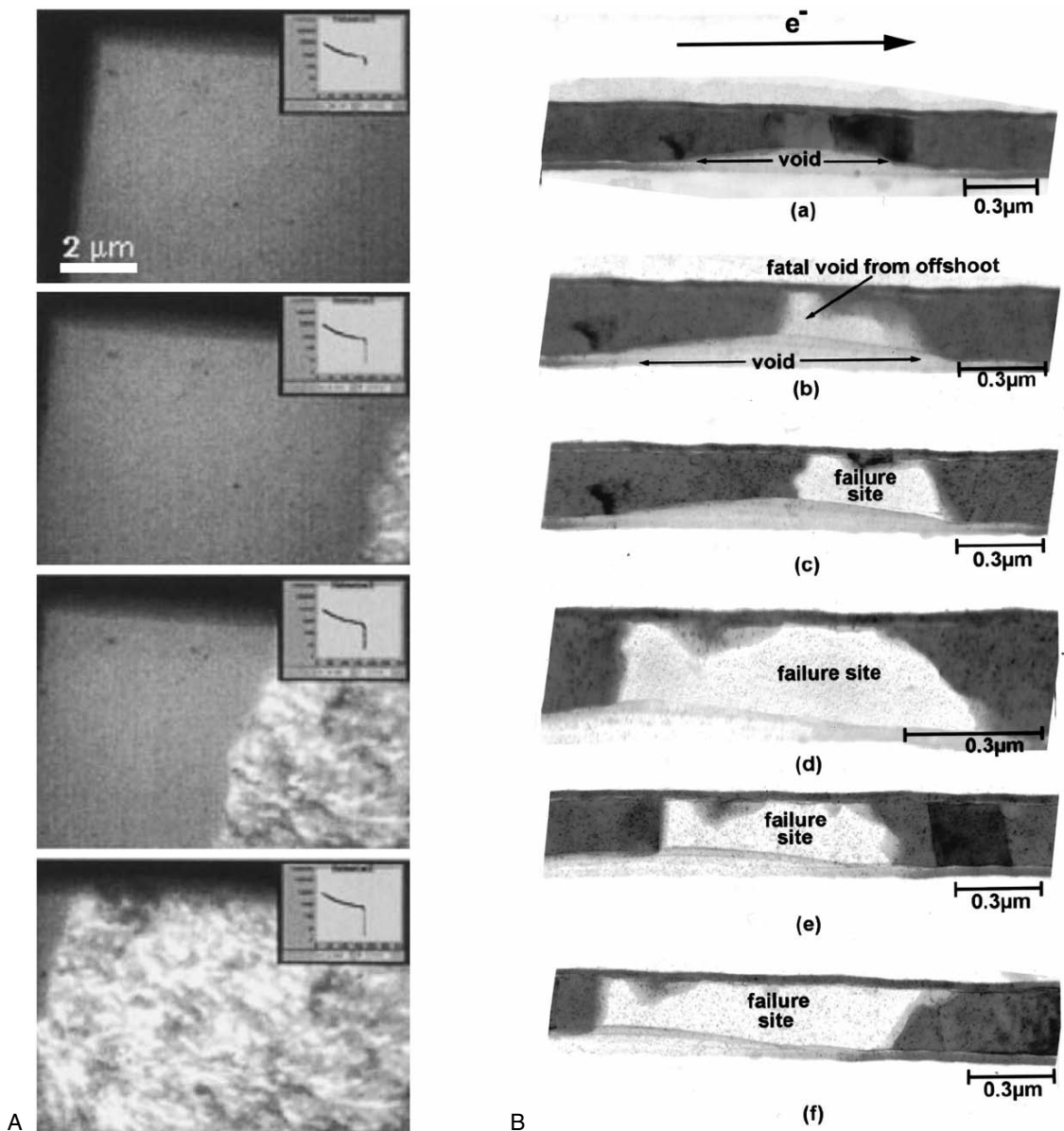


Figure 6-32. Correlation of resistivity with structure. (A) Video images taken at 15 s time intervals simultaneously displaying the crystallization of an Al-Ge film at 450K and the accompanying drop in electrical resistance (insets). The crystallization front can be seen to proceed from the right- to the left-hand side. (Reprinted with permission from Verheijen et al., © 2004. American Institute of Physics.) (B) Evolution of a 300 nm wide Al (0.5 wt% Cu) interconnect line, 400 nm thick and 150 μm long on a 20 nm TiN underlayer. Lines are deposited on a SiN/SiO₂ bilayer membrane to minimise changes in the stress state on heating. 100 nm oxide passivation is added, and the lines are then annealed to form the bamboo structure (i.e., all grain boundaries run perpendicularly across the line). In situ stressing is carried out at 320°C and a current density of $2 \times 10^6 \text{ Acm}^{-2}$, for (a) 3.5, (b) 4.0, (c) 4.1, (d) 4.5, (e) 4.6 and (f) 5.5 hours. (Reprinted with permission from Lau et al., © 1998 by the American Physical Society.)

and current densities above 10^6 Acm^{-2} are used for accelerated testing, so in some experiments, on-chip heat sink structures are integrated into the samples (Prybyla et al., 1998).

These studies show that electromigration-induced voids nucleate well before complete failure of the lines (Riege et al., 1996), and void dynamics depend on the local grain boundary structure. The failure mechanism depends on temperature, since at high temperature voids unpin from grain boundaries while at lower temperatures they grow at their nucleation site (Prybyla et al., 1998). Interestingly, when TiN barrier layers are present, voids do not migrate, presumably due to changes in surface diffusion (Lau et al., 2000). Even thin oxide films alter electromigration kinetics. Surface diffusion is in fact the dominant failure mechanism if grain boundary motion is hindered (Vook, 1994; Chang and Vook, 1995).

Since a thick passivation layer may be essential for meaningful results, in some studies the sample is mounted vertically and imaged in cross section at high voltage (Okabayashi et al., 1996; H. Mori et al., 1997). Such experiments show mass transport through Al and TiN layers, as well as vertical void and whisker growth. The measurement of local strain by CBED during electromigration (Nucci et al., 2005) is an exciting recent development that promises to relate these *in situ* results more closely to models.

6.2 Electrical Measurements on Individual Nanostructures

The experiments described above, where electrical biasing is applied to a relatively large volume of the specimen, probe several important phenomena. At the nanoscale, equally interesting information is provided by biasing individual nanostructures using STM technology. Commercial TEM/STM holders are in fact becoming increasingly common for both electrical and mechanical applications (www.nanofactory.com, www.gatan.com).

As with the mechanical experiments in Sections 5.3.2 and 5.3.3, carbon nanotubes and Au wires are favored for electrical studies. We firstly discuss CNTs. One can bias a tip on which a tube has been placed or grown, or equivalently bias the tip and approach a tube on the substrate. This allows observation of the electric field distribution at the tip of a biased tube (Cumings et al., 2002), and structural changes during field emission (Wang et al., 2002b; Kuzumaki et al., 2004; Jin et al., 2005). The work function of CNTs can be measured and related to structure (Gao et al., 2001; Bai et al., 2003b; Xu et al., 2005a, b) and CNT growth can be observed in the gap between a biased tip and the substrate (Yamashita et al., 1999).

If an individual tube is contacted, its conductivity can of course be measured. Kociak et al. (2002) were even able to correlate the chiral indices of double walled tubes, measured using diffraction, with their transport properties (Figure 6–33). Conductivity measurements show that CNTs are ballistic conductors at room temperature (Poncharal et al., 2002), and that the telescoping multiwalled CNTs mentioned in Section 5.3.3 behave as near-ideal rheostats (Cumings and Zettl, 2004).

Structural changes taking place during current transport have also been observed (Huang et al., 2005).

The other common material for electrical studies is Au. When thin wires are formed by touching Au tips to a substrate and then pulling away, the conductance can be measured and correlated with the structure. Figure 6–34 demonstrates quantized conductance through single and double Au chains (Ohnishi et al., 1998b). Single chains in fact show a metal-insulator transition (Kizuka et al., 2001a). Conductance and structure change together, showing the dynamic nature of the system (Kizuka et al., 2001b; Oshima et al., 2003c). Catalytically grown nanowires (Section 3.6) provide another interesting subject for study. Although Larsson et al. (2004) measured the conductivity of GaAs nanowires, nanowire electrical properties have generally not been examined in as much detail as Au wires or CNTs.

For any nanostructure studied *in situ*, the nature of the contact, for example how the STM tip is cleaned, is important in ensuring that the measurements relate to the structure rather than the contact. A drop of Hg can be used (Poncharal et al., 1999; Wang et al., 2001; Kociak et al., 2002) to measure properties like work function and quantum conductance, but there are still real issues in creating ohmic contacts to nanostructures (Larsson et al., 2004) which must be solved before extending these experiments to a wider range of materials systems.

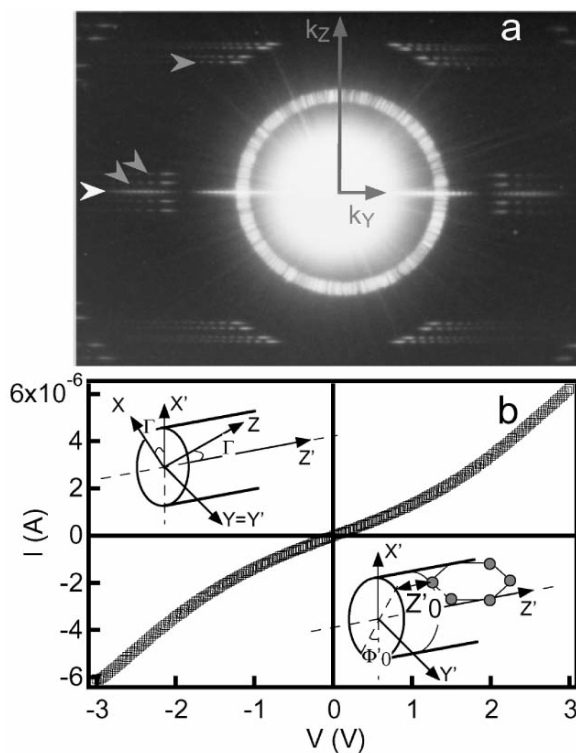
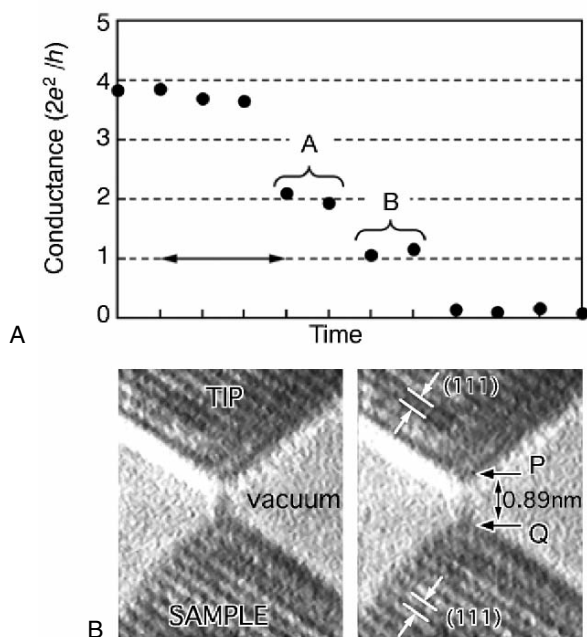


Figure 6–33. Simultaneous structure analysis and measurement of transport properties for a double-walled nanotube (DWNT). (A) Experimental diffraction pattern of a DWNT. The iris-like ring is an artifact. Note the layered line structure of the diffraction pattern. The white arrow indicates the equatorial line, and the gray ones indicate some other intensity lines. (B) Current-voltage characteristic of the same tube. Ohmic behavior is visible up to 0.5 V. Insets: Diffraction geometry. The upper inset shows the electron beam oriented along X with (X', Y', Z') the frame of reference of the MWNT. The diffraction pattern is recorded in the YZ plane. In the lower inset, a carbon atom (gray dot) is characterized by a translation distance Z'_0 and a rotation angle Φ'_0 with respect to the tube axis. (Reprinted with permission from Kociak et al., © 2002 by the American Physical Society.)

Figure 6–34. Electrical properties of nanowires. Quantized conductance of a single and a double strand of gold atoms prepared by contacting an Au surface with an STM tip in a UHV TEM. (a) Conductance change of the contact while withdrawing the tip. Conductance is shown in units of $G_0 = 2e^2/h = (13\text{k}\Omega)^{-1}$. (b) Electron microscope images of gold bridges obtained simultaneously with the conductance measurements in (a). Left, bridge at step A; right, bridge at step B. The bridge at step A has two rows of atoms; the bridge at step B has only one row of atoms. The distance from P to Q [(see (b))] is about 0.89 nm, wide enough to have two gold atoms in a bridge if the gold atoms have the nearest-neighbor spacing of the bulk crystal (0.288 nm). (From Ohnishi et al., © 1998. Courtesy of Nature/NPG.)



6.3 Outlook

In situ electrical experiments have touched a range of phenomena, providing detailed and unique information on both bulk and nanoscale materials. With nanostructured materials appearing in more and more electronic applications, we anticipate increased demand for the sort of information that only these experiments can provide. We also anticipate that other properties may be amenable to an analogous approach. Perhaps nanoscale optical response could be studied by feeding a laser beam into the sample area, or thermoelectric properties could be measured in nanoscale structures. No doubt equally fascinating behavior would be revealed.

7 Liquid Phase Processes

We now consider another unusual application of *in situ* TEM: studying the important class of processes that takes place in the liquid phase. For the vapor phase crystal growth and surface reaction processes of Section 3, real time observations clearly improved our understanding of nanostructure formation and thin film growth. Is it possible to carry out the same sort of quantitative studies for crystal growth from the liquid phase, or study other liquid processes? The answer, perhaps surprisingly, is yes, and the presence of a liquid is not incompatible with the vacuum requirements of the TEM. In this section we describe several studies involving liquids, which range from electrochemical deposition to the hydration of minerals and the growth of polymers.

In Sections 2.1 and 2.3 we presented examples of experiments involving melting and crystallization. These succeeded in the TEM either because the liquid was naturally trapped as an inclusion in a solid matrix, or because it had a low vapor pressure. Studies of vapor-liquid-solid growth (Section 3.6) also succeeded because of the low vapor pressure of the eutectic. For example, the vapor pressure above the Si-Au droplets in Figure 6–18 is only 10^{-10} Torr at the growth temperature.

But for liquids with higher vapor pressure, most importantly water, TEM can present some challenges. The liquid must be artificially encapsulated or differentially pumped to keep the pressure at the gun low enough for operation. Thus it is necessary either to use a special “window cell,” in which the liquid is hermetically sealed between electron transparent membranes (windows), or to modify the microscope to increase pumping to the specimen area. Both approaches have advantages and disadvantages.

If a hermetically sealed cell is used, the windows cause undesirable background in the images. Furthermore, since the liquid scatters just as strongly as a solid, it must be confined within a short path length. Either it must be present as a thin film covering a solid support, or (if it is to fill the cell) the windows must be very close together, which can be an engineering challenge. Even for thin liquid films on a solid support, the window separation should be minimized, as a long path length of vapor above the liquid leads to undesirable electron-gas interactions, especially for water vapor (Shah, 2004). Window cells are complex and do not usually allow heating or two-axis tilting, although, as we will see below, electrical connections are possible. In spite of the experimental limitations, a hermetically sealed cell can allow liquid-covered specimens to be examined in a close to natural state, or can be used to maintain a solvent-rich atmosphere.

In the alternative approach, no windows are used, but instead additional apertures are incorporated into the TEM and the pumping in the specimen area is modified to maintain a high pressure at the sample while leaving the rest of the column at low pressure. Microscopes with differential pumping have already been described in the context of gas phase reactions, but they can also be used to observe samples on which a liquid is present. The liquid can be introduced by injecting it through a valve. This approach requires a dedicated microscope, but has the potential to provide higher resolution images than are possible using window cells, and is also compatible with conventional heating, cooling, and tilting holders.

Several studies involving liquids have overcome the experimental challenges described above and have yielded useful and unique information. Liquid phase TEM is still uncommon, however, and this is an area offering many opportunities.

7.1 Imaging in a Saturated Environment or Under a Liquid Film

Window cells or differentially pumped microscopes have been used in studies of a range of phenomena including hydration and dehydration of minerals, the growth of nanostructures, and the action of cata-

lysts. Early work was carried out using window cells to provide a water saturated atmosphere, for example in studying the hydration of Portland cement (see Butler and Hale, 1981, for a review). The use of high voltage microscopes meant that the window separation was not too critical. Since then, several groups have used window cells to image biological and non-biological samples. Such cells are often based on the design of Fukami et al. (1991). The windows consist of 20 nm amorphous C membranes attached over a perforated metal disc. This material can withstand a pressure differential of 250 Torr and does not produce much background in the images. Imaging in an atmosphere of ~100 Torr air saturated with water vapor allowed Sugi et al. (1997) to examine an interesting biological mechanism: the movement of the myosin head in living muscle filaments. It also allowed Daulton et al. (2001) to determine the oxidation state of Cr in particles produced by an important Cr-reducing bacterium. Use of a window cell for biological materials avoids cryofixation with its attendant artifacts, so that for example bacterial cells do not burst, although beam damage still occurs at high doses. By injecting liquid droplets into a window cell, Foo et al. (2001) and Chiou et al. (2002) examined the arrangement and shape of small particles suspended in solvent, such as fume silica, smectite, and Au modified with DNA. Use of the window cell avoided the problem of agglomeration which usually occurs on evaporation. Although not strictly *in situ* experiments, this work illustrates the possibilities for examining liquid phase dynamics of small particles.

As mentioned above, the use of a differentially pumped electron microscope allows for greater flexibility in the sample holder during liquid experiments. This approach has been successful for examining melting of Xe films (Zerrouk et al., 1994), liquid phase synthesis of Au nanorods (Gai and Harmer, 2002), and several important catalytic reactions, including the sequence used to form nylon (Gai, 2002b). The first reaction in this sequence, shown in Figure 6–35, was observed by injecting the precursor, in a methanol solvent, over a metal/TiO₂ catalyst specimen, while simultaneously heating the catalyst and flowing hydrogen; the second reaction was observed at a higher temperature on introduction of adipic acid. Since many commercial polymerization reactions take place from solution, these types of experiments can have significant impact on the development of industrial processes, in the same way that gas phase experiments have supported industrial development of gas reaction catalysts. Recently, imaging in an environment of varying relative humidity using differential pumping and a cooling stage has allowed the deliquescence of salt and aerosol particles to be correlated with their structure and composition (Wise et al., 2005). The importance of these particles in the Earth's atmosphere makes this an urgent area for further investigation.

7.2 Electrochemical Deposition

Electrochemical deposition is an important crystal growth process which takes place from the liquid phase. Among its many applications

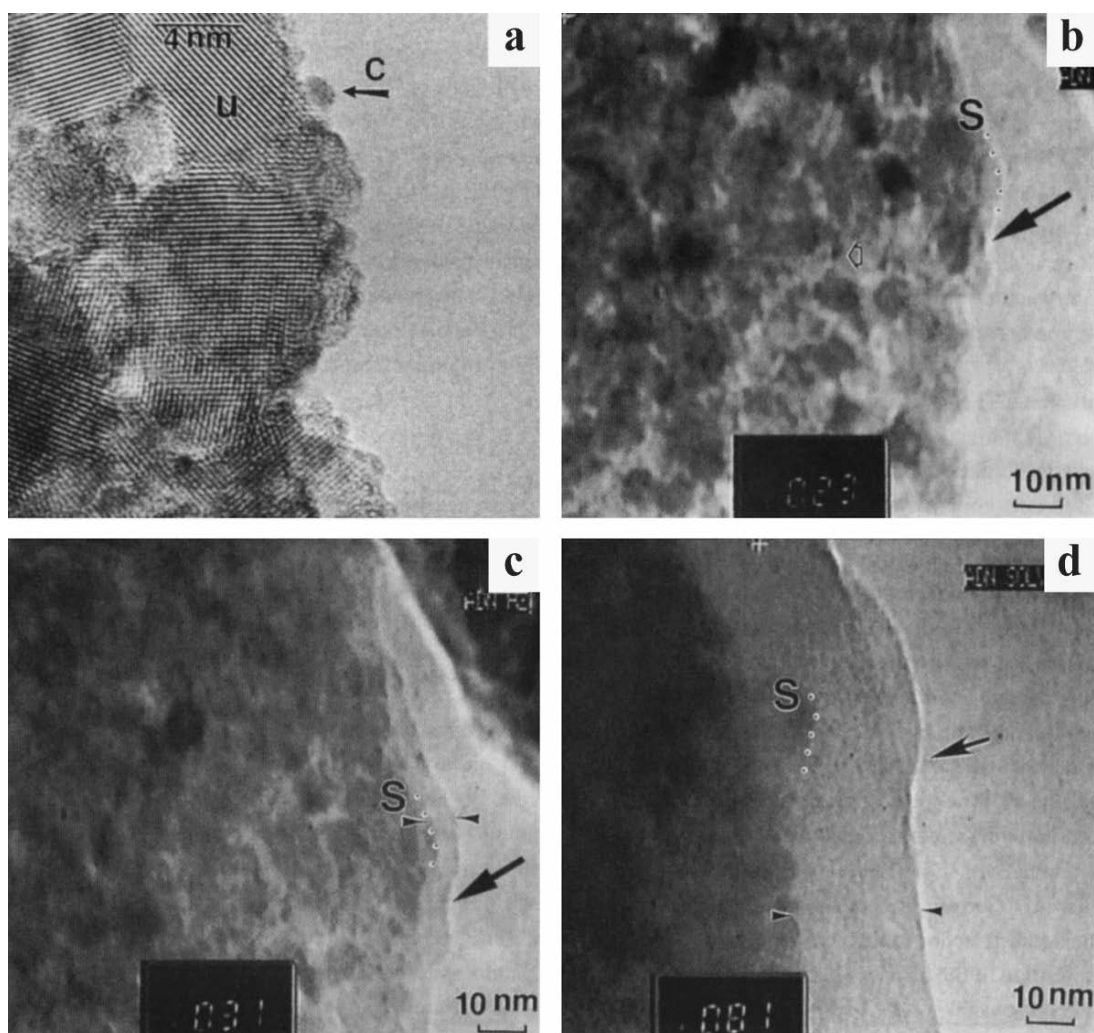


Figure 6-35. A liquid phase reaction *in situ*. The hydrogenation of adiponitrile (ADN) forms hexamethylene diamine (HMD), an important material since it in turn is reacted with adipic acid and polymerized to produce nylon 6,6. The ADN to HMD reaction is carried out with ADN in a methanol solvent under gaseous hydrogen over a metallic catalyst. (a) High-resolution image of the catalyst, nanoclusters of Co–Ru (arrowed, c) over rutile titania support (with larger grains, u), at RT. (b) After immersion in adiponitrile in solvent and H₂ gas, at RT. (c) The formation of layers of the product HMD at 31°C. The width is indicated by double arrows. (d) Thicker layers form at 81°C (arrowed) after 3 min. (S with dots indicated the original catalyst profile.) (From Gai, 2002. Courtesy of Cambridge University Press.)

is the formation of the copper interconnects in integrated circuits, and a detailed understanding of nucleation and growth is important in optimizing this process. In order to study electrochemical deposition, not only must imaging be carried out in a stable liquid environment, but it must also be done under electrochemical control. Thus the experiment uses a window cell, as described above, with electrical connections, as discussed in Section 6.

The electrochemical window cell (Williamson et al., 2003; Radisic et al., 2006) includes three electrodes connected externally to a potentiostat, allowing the same degree of control as a standard electrochemical cell. The entire volume between the windows is filled with the electrolyte, and the current passes between the working and counter electrodes through this liquid. The working electrode, on which the material of interest is deposited, is patterned over one of the windows. The counter electrode and reference electrodes are thin wires extending into the liquid reservoir. Since the cell volume is completely filled, unlike the experiments described above, it is particularly important to restrict the window spacing to below 1–2 μm to minimize multiple scattering. An imaging filter may be used to improve the contrast.

Typical results, given in Figure 6–36, show the nucleation and growth of copper clusters on a gold electrode in real time, with simultaneous measurement of the voltage applied and the current flowing in the cell. If care is taken to verify that the behavior of the small area of electrode under observation is typical of the entire electrode, and that the small volume cell behaves like a standard electrochemical cell, quantitative data on individual clusters can be compared with electrochemical models. In the experiment shown here, conventional electrochemical models could not explain the nucleation density and growth rates observed. Conventional models include only direct attachment of copper ions onto existing clusters, and to explain the results it was necessary to modify these models to include a parallel pathway, direct attachment of Cu onto the electrode followed by surface diffusion to clusters (Radisic et al., 2006).

Although complex, the electrochemical liquid cell developed here may be adapted for a variety of electrochemical growth and corrosion processes. In the future, this type of experiment could be enhanced by combining the cell with microfluidics technology to allow flowing rather than stationary liquid, and to enable controlled mixing or heating of liquids.

7.3 Outlook

Dynamic processes which take place in liquids and at liquid/solid interfaces have great importance across broad areas of science and technology. The results described in this section show that processes involving water or other liquids may indeed be observed in real time and with reasonable spatial resolution *in situ*, providing information which is difficult to obtain using techniques such as scanning probe microscopy or SEM. Further experiments could improve our understanding of, for example, changes in electrode surfaces during battery charging or underwater corrosion of metals such as steel. We anticipate that future liquid studies in the TEM will provide important information relevant to microelectronics fabrication, catalyst design, and certain biological processes, making this a key area of *in situ* TEM for further development.

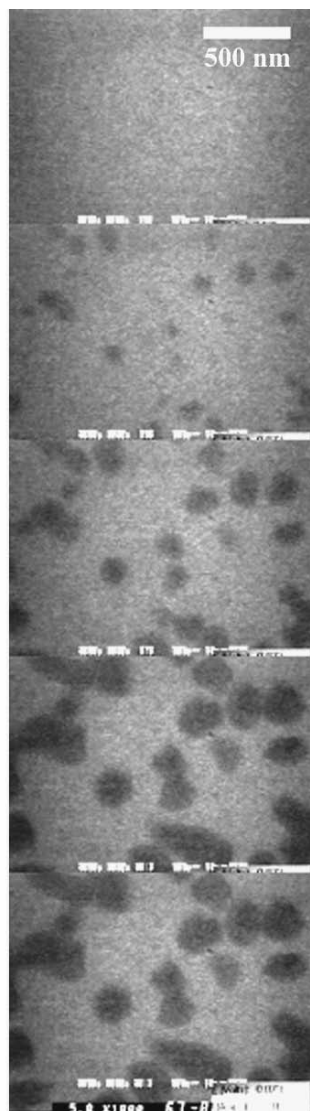
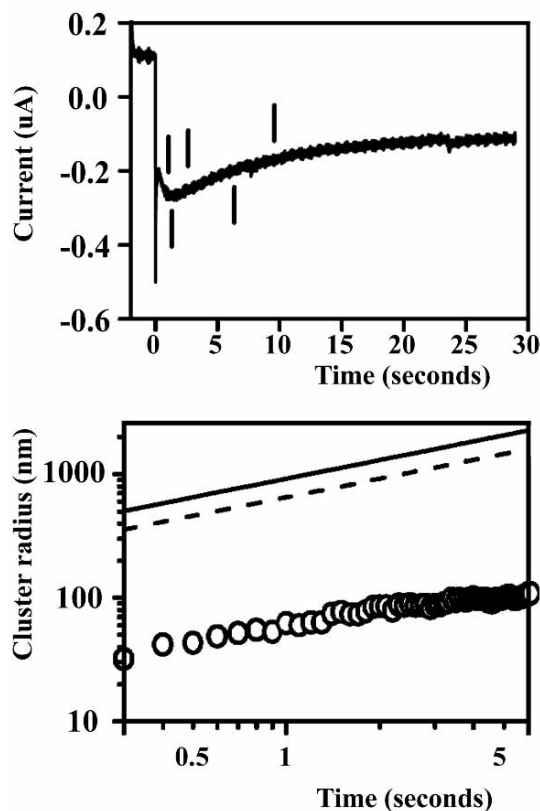


Figure 6-36. Electrochemical deposition of copper. Deposition was carried out *in situ* onto a gold electrode from acidified CuSO_4 solution. First a reverse potential was applied for 2 seconds to clean the electrode, then deposition was carried out at a constant potential of -0.07 V , measured with respect to a Cu reference electrode. The upper graph shows the current flowing through the TEM cell. After a sharp transient, the current curve shows a characteristic maximum ($-0.26\mu\text{A}$ after 1.5 seconds in this case) which can be fitted with a model of cluster nucleation and growth. The images provide a direct view of cluster evolution on part of the electrode, with Cu appearing dark. The lower graph shows the growth of a single cluster (radius vs. time) compared with the prediction of two models. The growth rate is much slower than predicted, and the cluster density much higher, suggesting that the models needs to be modified by the inclusion of surface diffusion effects. (From Radisic et al., © 2006 with permission of the American Chemical Society.)

8 Ion and Electron Beam-Induced Processes

Ion and electron beams have multiple effects on a specimen, depending on the material, and atomic displacements, implantation of atoms, or ionization may all be seen during *in situ* experiments. Irradiation

effects are certainly worthy of study both as basic and applied science, and are of interest far beyond simply understanding TEM artifacts. Historically, irradiation damage by ions, electrons, or both simultaneously was studied *in situ* in attempts to model neutron damage in the context of nuclear electric power generation and the development of host materials for long term nuclear waste storage. The high dose rates possible make TEM useful for accelerated experiments, and irradiation-induced defects and the effects of irradiation on phase equilibria and electrical and mechanical properties were studied. However, several other areas have recently become interesting. These include materials processing (ion beam assisted deposition, ion beam modification of interfaces, and ion implantation for microelectronics processing), development of semiconductor devices for aerospace applications, photoplasticity, and the study of radiation damage in minerals in order to improve the dating of radioisotope-bearing minerals. A recent review can be found in Birtcher et al. (2005).

8.1 Electron Beam-Induced Phenomena

In this section we discuss several effects of the electron beam on the specimen. Electron beam effects may be obvious, such as heating that induces or accelerates phase transformations, or more subtle, such as an increased point defect concentration that may affect deformation. Beam effects must of course be considered when interpreting any TEM observation, but can also be used to probe defect dynamics or growth processes.

8.1.1 Interaction with the Atmosphere Above the Specimen

As discussed in Section 3.4, the interaction of the electron beam with the column atmosphere produces a plasma in which structures such as fullerenes can grow, or certain components of oxides can be etched. Water vapor has a particularly strong interaction with the beam, as mentioned in Section 7. Beam-atmosphere interactions must be considered when using a microscope without a controlled environment, because they can lead to uncontrolled effects. But even when the environment is controlled, interactions will of course still occur. For example, in the hydrogen embrittlement studies described in Section 5.1, the electron beam significantly increases the “fugacity” (an effective partial pressure) of the hydrogen (Bond et al., 1986).

Beam effects in a reactive atmosphere can be used to advantage for deposition. For example, tungsten can be deposited via electron beam stimulated decomposition of $W(CO)_6$ (Furuya et al., 2003) to grow structures such as that shown in Figure 6–37. Similarly, iron can be deposited from $Fe(CO)_5$ (Tanaka et al., 2005; Takeguchi et al., 2005). Three dimensional structures may even be formed using a programmed STEM (Shimojo et al., 2004).

8.1.2 Formation and Dynamics of Point Defects

Once the beam hits the sample it creates point defects, and both point defect motion and the dynamics of phenomena induced by point defects can be examined *in situ*. In general, individual point defects are

difficult to see. However, if the specimen is thin and has a well-defined structure, such as a carbon nanotube, it may be possible to image point defects directly. Thus, point defects caused by irradiating graphene sheets may be imaged at high resolution (Hashimoto et al., 2004), and individual vacancy-interstitial pairs can be observed in double walled CNTs (Urita et al., 2005). Their relaxation can be measured, as can the defect-induced migration of metal atoms (Urita et al., 2004).

When point defects cluster into extended defects, the extended defect dynamics can be measured to obtain indirect information about the point defects such as their diffusion parameters. The fundamentals of point defect motion have been studied in many materials in this way using high voltage microscopy and are reviewed by Kiritani (1999). For example, by measuring the growth and shrinking of interstitial loops during intermittent irradiation, it is possible to obtain activation energies for vacancy migration and self diffusion (Arai et al., 1995). The density of point defects can be measured by examining the formation of jogs in dislocation loops (Arakawa et al., 2000) while their diffusion parameters can be measured by examining extended defect dynamics far from the irradiated area (Arai et al., 2004). Loop growth and shrinkage can even be used to investigate defects produced by *ex situ* neutron irradiation (Horiki et al., 1998; M.A. Kirk et al., 1999).

Extended defect dynamics also provides information on interactions between point defects and extended defects. In Cu, irradiation causes growth of stacking fault tetrahedra, and an analysis of size fluctuations showed that growth is by ledge motion after capture of defects (Arakawa et al., 2002). In Si, irradiation forms interstitial clusters, and Fedina et al. (1998) and Vanhellemont et al. (1995) observed their structure and formation kinetics (Figure 6–38). Strain fields influence point

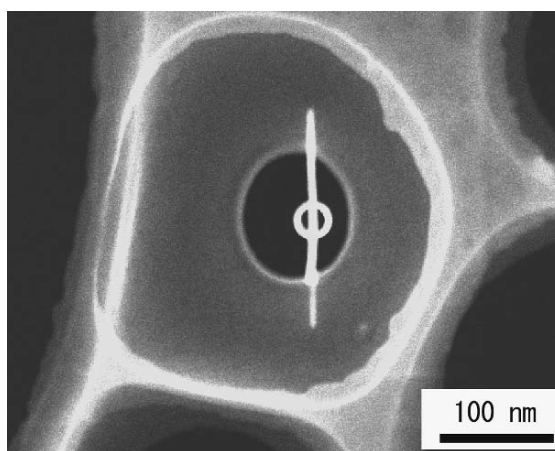


Figure 6–37. Fabrication of nanostructures by beam-vapor interaction. Nanoscale fabrication carried out on a carbon grid, using a JEM 2500SE STEM operated at 200 kV with a beam size of 0.8 nm and a beam current of 0.5 nA. $W(CO)_6$ gas was used at a flux of approximately $2 \times 10^{-4} \text{ Pa L s}^{-1}$. A line across a hole was produced followed by fabrication of a circle. (From Shimojo et al., 2004; with kind permission of Blackwell Publishing, Ltd., Oxford, U.K.)

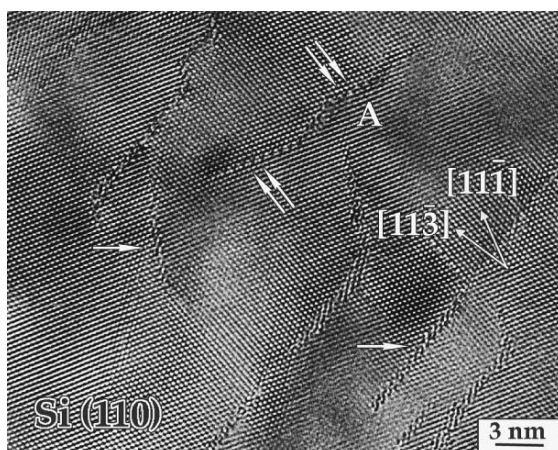


Figure 6–38. Irradiation damage. High-resolution electron micrograph image of point defect aggregates created in Si during electron irradiation at room temperature. Irradiation was carried out at 400 kV and 10^{20} electrons $\text{cm}^{-2} \text{s}^{-1}$ for 35 min. The {113} and {111} defects are marked with single and double arrows, respectively. (From Fedina et al., 1998 with kind permission of Taylor and Francis Ltd.)

defect clustering (Vanhellemont et al., 1995; Fedina et al., 1997), so presumably clustering could be used to measure strain locally. However, imaging of defect clusters can be challenging and special weak beam imaging techniques may be necessary (M.A. Kirk et al., 1999).

High voltage electron irradiation can be used to move atoms, say from the surface into the bulk of a substrate, via elastic collisions of the electrons with the heavy atoms. Systems studied include Au implanted into Si (Mori et al., 1992), Hf into SiC (Yasuda et al., 1992), and Au into Al (Lee et al., 2002b), which forms Al_2Au phases that move downstream. These experiments are useful in studying the phenomena taking place during ion beam processing.

8.1.3 Beam-Induced Phase Transformations, Surface Reactions and Growth

We have seen examples throughout this chapter where the beam has induced a phase transformation or caused a growth process. To understand these phenomena it is important to separate heating effects from those caused by knock-on damage or electronic excitations. This may require detailed hot stage measurements.

Amorphization is the most commonly observed beam-induced transformation, taking place for example in SiC (Inui et al., 1992), Si (Takeda and Yamasaki, 1999) and GaAs (Yasuda and Mori, 1999). In InGa_N, beam-induced amorphization looks just like compositional fluctuations (Smeeton et al., 2003) which may of course cause problems in image interpretation. Interestingly, in Zr_2Ni and Zr_3Al , the amorphization kinetics are changed by the presence of hydrogen (Tappin et al., 1995), which alters the stability of the defective structure. Thus the microscope atmosphere should be considered in these transformations as well. The reverse process of beam-induced recrystallization also

occurs, for example in semiconductors (e.g., Jencic et al., 1995). It may be seen at voltages below those required to create point defects, and may be caused by the formation of dangling bonds at the crystalline/amorphous interface. Other beam effects include phase separation and the formation of non-equilibrium phases (Section 2). For example, in borosilicate glasses, which have applications in nuclear waste storage, B-rich phases separate under the beam (K. Sun et al., 2005).

Beam-induced growth processes have interesting applications in nanofabrication. Apart from the beam-induced deposition described in Section 8.1.1, nanostructures can be built by defect generation in suitable substrates (see Section 3.4 for C examples) and wires can be formed by beam-enhanced surface diffusion (see Section 3.1 for an Au example). Nanostructured material can also be formed by beam-induced decomposition. This has been examined particularly in SiO_2 , where a combination of sputtering and desorption produces Si rich regions (Fujita et al., 1996; Chen et al., 1998; Du et al., 2003; Furuya et al., 2003). Beam-induced decomposition occurs in many materials (Al_2O_3 , MgO , AlF , etc.) and may proceed from either or both surfaces. Several processes are active, including interactions with the atmosphere, changes in surface diffusion (Mera et al., 2003) and surface roughening (Grozea et al., 1997). If the beam is finely focused, such as in a STEM, hole drilling may occur (e.g., Berger et al., 1987; Walsh, 1989; Kizuka and Tanaka, 1997a, b), another possible method of fabricating nanostructures.

8.1.4 Radiation-Enhanced Dislocation Motion

In many materials, the electron beam influences dislocation motion. Radiation-enhanced dislocation glide during heating or straining has been studied quantitatively by recording the motion of individual dislocations (Figure 6–39). Such measurements suggest a mechanism based on an enhancement in the creation rate of kink pairs, due to energy that is released by nonradiative recombination of electron-hole pairs at electronic levels associated with dislocations (Levade et al., 1994; Werner et al., 1995; Yonenaga et al., 1999; Vanderschaeve et al., 2000, 2001; Maeda et al., 2000). The radiation-enhanced motion of individual kinks can actually be observed directly in plan view (Inoue et al., 1998). A radiation-enhanced climb process can also occur, due to absorption of interstitials by dislocations (Yonenaga et al., 1998). These studies are important in interpreting *in situ* deformation experiments, but also have a close relationship to the phenomenon of photoplasticity which is relevant to optoelectronic device degradation.

8.2 Ion Implantation

It is possible to irradiate a sample with ions while simultaneously imaging or irradiating it with electrons. This area of research requires complex, expensive instruments with tandem accelerators. However, because of the importance of the results, such equipment has been funded by several institutions around the world. A discussion of the microscopes capable of simultaneous ion and electron irradiation can be found in Allen and Ryan (1998) and some of the research has been summarized by Ruault et al. (2005). Our understanding of sputtering, ion

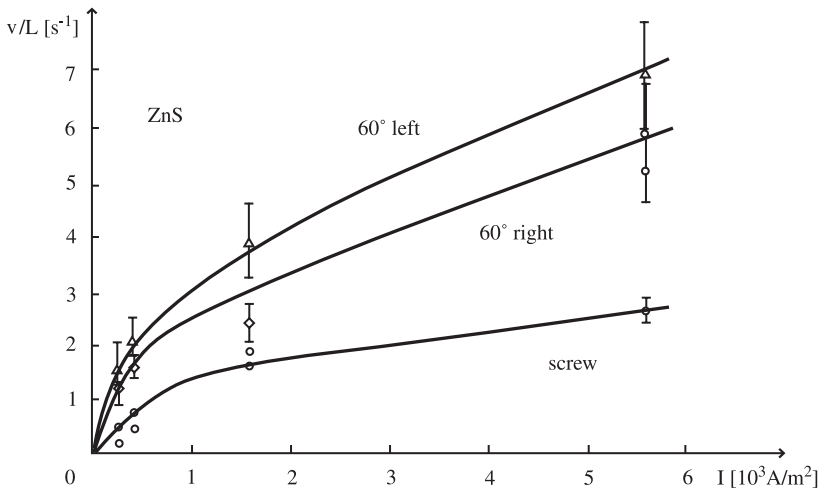


Figure 6–39. Radiation-induced dislocation glide. Velocity of dislocations in ZnS of different types per unit length, plotted as a function of electron beam intensity. Velocities were measured during *in situ* straining at 40 ± 10 MPa and 390K. The motion remains slow suggesting that the Peierls mechanism still controls motion, and velocity is proportional to dislocation length showing that kinks form individually and do not collide. The linear regime at low intensity suggests a change (reduction) in apparent activation energy due to nonradiative recombination of carriers at dislocations, assisting kink formation. At high intensities the recombination rate saturates. (From Levade et al., 1994 with kind permission of Taylor and Francis Ltd.)

implantation, and materials processing, for example by impurity doping, may be developed using tandem microscopes. However, the driving force for these microscopes is a need to understand the effects of radiation on materials which may be used in reactor walls or for storage of radioactive waste. Simultaneous ion and electron irradiation simulates the defect types and concentrations created by irradiation with high energy neutrons from nuclear reactions. Neutrons induce displacement damage and cause the formation of helium as a byproduct. The helium forms bubbles and causes swelling, embrittlement, and radiation-induced segregation, and phase stability may change as a result.

Because of their industrial importance, many irradiation studies have been carried out in steels. These include determining the stability of different phases after Xe irradiation (Chu et al., 2002) and measuring parameters of the motion of bubbles formed by He implantation (Ono et al., 2002). Quantitative studies provide information on the key defect formation and diffusion processes taking place. For example, in Cu, the kinetics of displacement collision cascade formation and the recombination of point defects into clusters were measured using Kr irradiation (Daulton et al., 2000), while in Si, irradiation with Si shrinks the bubbles caused by H implantation (Ruault et al., 2002). Bubble motion has been studied in most detail in a simple system, however. This model system consists of insoluble noble gases such as Ar, Ne, or Xe implanted into an Al or similar matrix, forming small precipitates (Figure 6–40A). Implantation can be carried out *ex situ* (Birtcher et al.,

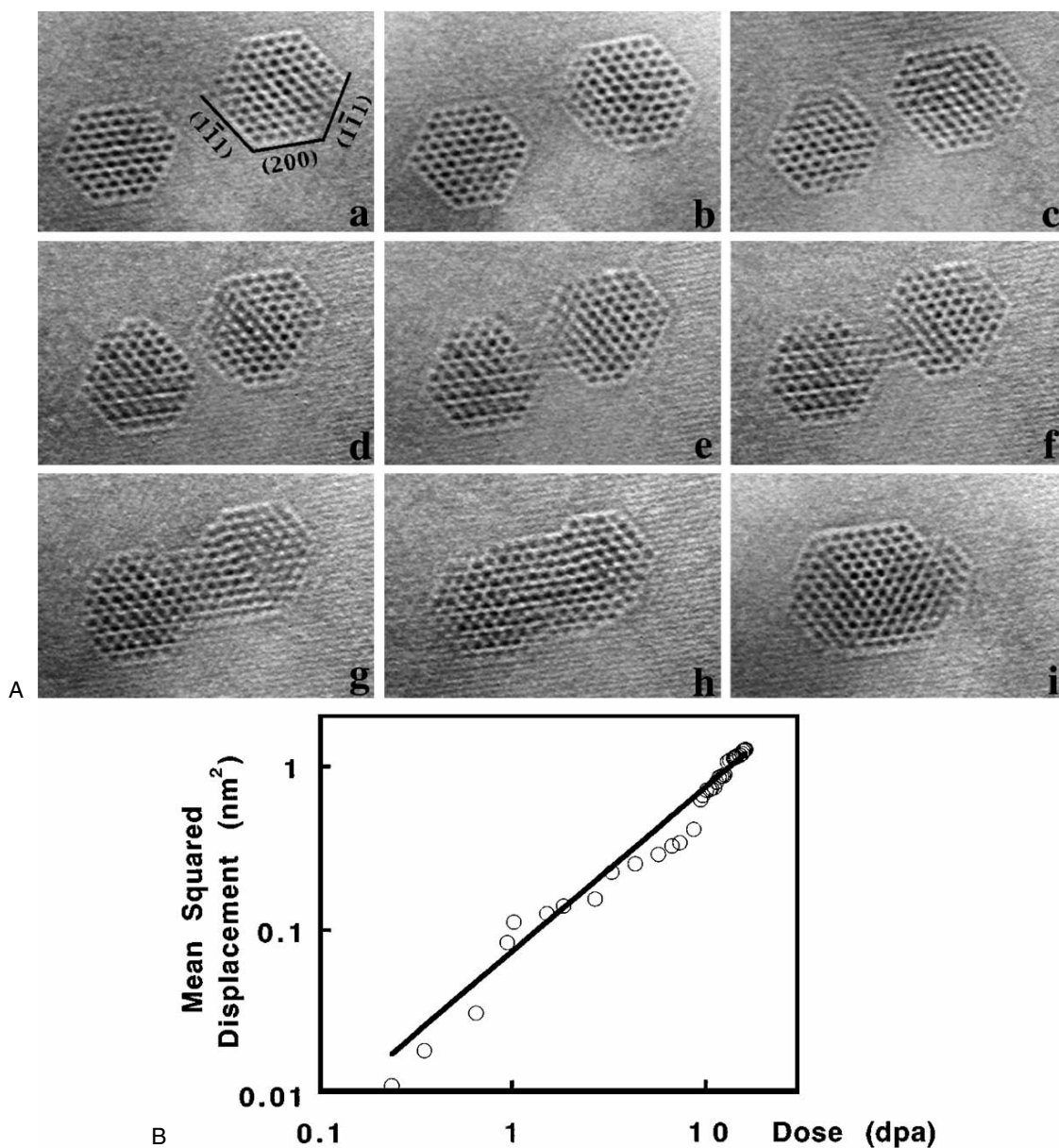


Figure 6-40. Dynamics of Xe precipitates. (A) Motion and coalescence of two isolated crystalline Xe precipitates (Xe implanted *ex situ*) during continuous 1 MeV electron irradiation at room temperature. The irradiation causes a damage rate in Al of approximately 3.9×10^{-2} displacements per atom (dpa) per second. Measured from the first image, the elapsed times at which video frames were recorded are (a) 0, (b) 101, (c) 418, (d) 549, (e) 550, (f) 551, (g) 561, (h) 584, and (i) 727 s. Traces of crystallographic planes are indicated in frame (a). Analysis of such data shows that surface diffusion of Al is responsible for motion and shape changes, while the Xe deforms by shear in response to the reshaping of its cavity. Since the total volume, not the surface area, was conserved during coalescence, cavity pressure depends on the gas structure and not just the interface structure. (Reprinted with permission from Birtcher et al., © 1999 by the American Physical Society.) (B) Mean square displacement of a Xe precipitate containing approximately 38 Xe atoms (in a volume equivalent to that of approximately 128 Al atoms), as a function of damage in the Al matrix under 1 MeV electron irradiation. The precipitate moves because of Al jumps on its surface, and these data yield values for the diffusivity and an average jump frequency of about 5600 jumps/dpa. (Reprinted with permission from Allen et al., © 1999. American Institute of Physics.)

1999; Allen et al., 1999; Ono et al., 2002) or *in situ* (Song et al., 2002), so that the coalescence of defects to form the precipitates can be studied as a function of dose. Striking dynamic changes occur in these precipitates under 1 MeV electron irradiation (Birtcher et al., 1999; Allen et al., 1999) or on heating (Ono et al., 2002): migration, shape changes, faulting, melting, crystallization, and coalescence. The shape, nature, and motion of the enclosed phase provide information on interface energies, structure, and diffusion pathways (Donnelly et al., 2002; Allen et al., 2003; Figure 6–40B).

We conclude with a related but lower voltage ion irradiation project: the integration of a 20 kV focused ion beam into a TEM (M. Tanaka et al., 1998; Gnauck et al., 1998). This tandem microscope allows the damage produced by the FIB to be observed directly, which is useful for understanding artifacts and low voltage damage effects. It also allows sample thinning to be carried out *in situ* so that the area imaged has not been exposed to the air. Studies of focused ion beam damage are increasingly important since the FIB is used for materials processing, surface patterning, growth, and fabrication of nanostructures.

9 Outlook

In this chapter we have attempted to show some of the outstanding recent results achieved using *in situ* microscopy and to illustrate the huge variety of experiments possible. In spite of experimental complexity, many successful and informative experiments have mimicked the real world inside the microscope column.

It has become clear that there is no one “*in situ* microscope.” Interesting experiments can be done in conventional machines, using standard holders, just by heating the sample with the electron beam. More complex experiments can be carried out by purchasing commercial *in situ* holders (heating, cooling, straining) or by developing customized ones (nanoindenters, electrical biasing holders, window cells). Many experiments also require the ability to measure some property of the sample, such as resistivity, simultaneously with imaging its structure. The most complex *in situ* experiments take place in expensive microscopes which are modified to achieve a controlled specimen environment or are designed for UHV.

This is an exciting time for *in situ* microscopy. Specimen design is developing rapidly with innovative use of FIB preparation or integration of thin film specimens into micromachined substrates. Furthermore, as we have emphasized throughout this chapter, the increased interest in nanostructures fits perfectly with the ability of TEM to analyze small volumes of material with minimal sample preparation. Apart from thin foil effects and beam-induced artifacts, the main limitation of *in situ* microscopy has been the small space available in the polepiece. In the future this problem too will be solved by the ongoing developments in aberration correction. Corrected microscopes have an increased polepiece gap for the same resolution. The extra space will help experiments requiring combined stimuli (heating and irradiation,

straining in a gas environment, simultaneous E and B fields) or new stimuli (lasers, micromanipulators, microfluidics). An increased pole-piece gap will also improve calibration of the sample environment. It is hard to overemphasize the importance of this for quantitative results. For example, one could include a fiber optic coupled to a radiation thermometer (Isshiki et al., 1998), a thickness monitor for growth experiments, a pressure gauge, or MEMS-based sensors. Aberration correction would also allow low voltage microscopes to achieve a resolution comparable with today's high voltage microscopes. This will reduce beam damage, although sample thickness, or more generally sample design, will remain a constraint to decreasing the voltage.

However, there are many areas where *in situ* techniques could be extended even without aberration correction. Relatively few *in situ* experiments have used analytical techniques, such as energy filtered imaging or EELS, to examine time-resolved chemical changes. This is probably because rapid acquisition limits the signal to noise ratio. Thus, the ongoing improvements in detectors are promising for *in situ* analytical studies. Similarly, convergent beam electron diffraction has not been widely used *in situ*, but could give useful information on changes in local strain. And other advanced TEM techniques such as tomography, holography, or imaging of amorphous materials could also be used *in situ*, providing the phenomenon of interest is slow enough to make image acquisition feasible. High speed imaging, discussed elsewhere in this book, will of course open a new set of phenomena for study. And in terms of data collection and analysis, direct storage of images and videos on disc is improving the searchability of data, although most data collection still leaves something to be desired in terms of usability. Intelligent analysis of video data, such as object detection and tracking, is currently a limitation, as custom software often has to be developed for each experiment.

A final comment arises from the complexity of controlled environment *in situ* microscopes, which has increased since the development of the first such TEM by Poppa (1965). Although some processes, such as sample heating, must take place in the polepiece, it is an advantage to keep others out of the polepiece though still in the same vacuum system. Examples include ion guns or deposition tools such as evaporators. Moving some capabilities *ex situ*, but within the same vacuum system, improves the reliability of complex microscope systems and provides a more flexible approach to experiments, especially in a multi-user environment where time is a constraint.

The field of *in situ* transmission electron microscopy has advanced tremendously since its beginning in the 1960s. We anticipate even more exciting results over the next few decades.

Acknowledgments. I would like to acknowledge the inspirational scientists with whom I have had the pleasure of working. In particular, my Ph.D. advisor the late Mike Stobbs, who emphasized quantitative information from TEM; my colleagues Murray Gibson, Robert Hull, Uli Dahmen, and Ruud Tromp; and Mike McDonald, Mark Reuter, and

Arthur Ellis, whose outstanding technical help made out *in situ* experiments possible. I would like to acknowledge the support of IBM's T.J. Watson Research Center with its far sighted view of basic research. I would also like to thank my husband and baby daughter, without whom writing this chapter would have been much easier.

References

- Abe, H., Yamamoto, S. and Miyashita, A. (2002). *J. Electron Microsc.* **51** (Supplement), S183–187.
- Allen, C.W. and Ryan, E.A. (1998). *Microsc. Res. Tech.* **42**, 255–259.
- Allen, C.W., Birtcher, R.C., Donnelly, S.E., Furuya, K., Ishikawa, N. and Song, M. (1999). *App. Phys. Lett.* **74**, 2611–2613.
- Allen, C.W., Birtcher, R.C., Donnelly, S.E., Song, M., Mitsuishi, K., Furuya, K. and Dahmen, U. (2003). *Philos. Mag. Lett.* **83**, 57–64.
- Aoki, K., Minoda, H., Tanishiro, Y. and Yagi, K. (1998). *Surface Rev. Lett.* **5**, 653–663.
- Arai, S., Morita, C., Arakawa, K. and Kiritani, M. (1995). *J. Electron Microsc.* **44**, 1–7.
- Arai, S., Tsukimoto, S., Muto, S. and Saka, H. (2000). *Microsc. Microanal.* **6**, 358–361.
- Arai, S., Tsukimoto, S. and Saka, H. (2003). *J. Electron Microsc.* **52**, 79–84.
- Arai, S., Satoh, Y., Arakawa, K., Morita, C. and Kiritani, M. (2004). *J. Electron Microsc.* **53**, 21–27.
- Arakawa, K., Satoh, Y., Arai, S. and Kiritani, M. (2000). *Philos. Mag. A* **80**, 2041–2055.
- Arakawa, A., Arai, S., Orihara, H., Ono, K. and Kiritani, M. (2002). *J. Electron Microsc.* **51** (Suppl), S225–229.
- Asaka, K. and Kizuka, T. (2005). *Phys. Rev. B* **72**, 115431.
- Bai, X.D., Gao, P.X., Wang, Z.L. and Wang, E.G. (2003). *App. Phys. Lett.* **82**, 4806–4808.
- Bai, X.D., Wang, E.G., Gao, P.X. and Wang, Z.L. (2003b). *Nano Lett.* **3**, 1147–1150.
- Balk, T.J., Dehm, G. and Arzt, E. (2003). *Acta Mat.* **51**, 4471–4485.
- Banhart, F. (1997). *J. App. Phys.* **81**, 3440–3445.
- Banhart, F. (1999). *Rep. Prog. Phys.* **62**, 1181–1221.
- Banhart, F., Hernandez, E. and Terrones, M. (2003). *Phys. Rev. Lett.* **90**, 185502/1–4.
- Barre, S., Mutka, H., Roucau, C., Litzler, A., Schneck, J., Toledano, J.C., Bouffard, S. and Rullier-Albenque, F. (1991). *Phys. Rev. B* **43**, 11154–11161.
- Bartsch, M., Schall, P., Feuerbacher, M. and Messerschmidt, U. (2005). *J. Mat. Res.* **20**, 1814–1824.
- Baufeld, B., Baither, D., Messerschmidt, U., Bartsch, M., Foitzik, A.H. and Rühle, M. (1997). *J. Amer. Ceramic Soc.* **80**, 1699–1705.
- Bengu, E. and Marks, L.D. (2001). *Phy. Rev. Lett.* **86**, 2385–2387.
- Benson, W.E. and Howe, J.M. (1997). *Philos. Mag. A* **75**, 1641–1663.
- Berger, S.D., Salisbury, I.G., Milne, R.H., Imeson, D. and Humphreys, C.J. (1987). *Philos. Mag. B* **55**, 341–358.
- Bhattacharya, P., Bhattacharya, V. and Chattopadhyay, K. (1999). *J. Electron Microsc.* **48** (Supplement), S1047–1054.
- Birtcher, R.C., Donnelly, S.E., Song, M., Furuya, K., Mitsuishi, K. and Allen, C.W. (1999). *Phys. Rev. Lett.* **83**, 1617–1620.
- Birtcher, R.C., Kirk, M.A., Furuya, K., Lumpkin, G.R. and Ruault, M.O. (2005). *J. Mat. Res.* **20**, 1654–1683.

- Blanpain, B., Legresy, J.-M. and Mayer, J.W. (1990). *Colloque Phys.* **C-4**, 131–137.
- Bobji, M.S., Ramanujan, C.S., Doole, R.C., Pethica, J.B. and Inkson, B.J. (2003). *Mat. Res. Soc. Sympos. Pro.* **778**, 105–110.
- Bond, G.M., Robertson, I.M. and Birnbaum, H.K. (1986). *Scripta Metallurgica* **20**, 653–658.
- Bonevich, J.E., Harada, K., Matsuda, T., Kasai, H., Yoshida, T., Pozzi, G. and Tonomura, A. (1993). *Phys. Rev. Lett.* **70**, 2952–2955.
- Bovin, J.-O., Wallenberg, R. and Smith, D.J. (1985). *Nature* **317**, 47–49.
- Burden, A.P. and Hutchison, J.L. (1998). *Carbon* **36**, 1167–1173.
- Butler, E.P. and Hale, K.F. (1981). Dynamic Experiments in the Electron Microscope, Vol. 9. In *Practical Methods of Electron Microscopy* (North-Holland, Amsterdam).
- Cai, W., Murakami, Y. and Otsuka, K. (1999). *Mat. Sci. Eng. A* **273–275**, 186–189.
- Caillard, D., Martin, J.L. and Jouffrey, B. (1980). *Acta Metallurgica* **28**, 1059–1067.
- Caillard, D., Vailhe, C. and Farkas, D. (1999). *Philos. Mag. A* **79**, 723–739.
- Caillard, D., Roucau, C., Bresson, L. and Gratias, D. (2002). *Acta Materialia* **50**, 4499–4509.
- Cantoni, M., Uchida, M., Matsui, Y., Takekawa, S., Tsuruta, T. and Horiuchi, S. (1998). *J. Electron Microsc.* **47**, 443–448.
- Chang, C.Y. and Vook, R.W. (1993). *Thin Solid Films* **223**, 23–30.
- Chatterjee, K., Howe, J.M., Johnson, W.C. and Murayama, M. (2004). *Acta Mat.* **52**, 2923–2935.
- Chen, G.S., Boothroyd, C.B. and Humphreys, C.J. (1998). *Philos. Mag. A* **78**, 491–506.
- Chiou, W.-A., Ishikawa, A., Foo, K.H., Miyata, H., Tsukada, H. and Fukushima, K. (2002). *Proc. ICEM-15* **1**, 937–938.
- Chong, R.K.K., Yeadon, M., Choi, W.K., Stach, E.A. and Boothroyd, C.B. (2003). *App. Phys. Lett.* **82**, 1833–1835.
- Chu, F., Song, M., Mitsuishi, K., Yasuda, H. and Furuya, K. (2002). *J. Electron Microsc.* **51** (Supplement), S231–234.
- Collazo-Davila, C., Grozea, D. and Marks, L.D. (1998). *Phys. Rev. Lett.* **80**, 1678–1681.
- Crozier, P.A., Oleshko, V.P., Westwood, A.D. and Cantrell, R.D. (2001). *Electron Microsc. Anal.* 2001. Proceedings 393–396.
- Cumings, J. and Zettl, A. (2000). *Science* **289**, 602–604.
- Cumings, J. and Zettl, A. (2004). *Phys. Rev. Lett.* **93**, 086801/1–4.
- Cumings, J., Zettl, A., McCartney, M.R. and Spence, J.C.H. (2002). *Phys. Rev. Lett.* **88**, 056804/1–4.
- Dahmen, U., Hagege, S., Faudot, F., Radetic, T. and Johnson, E. (2004). *Philos. Mag.* **84**, 2651–2662.
- Dannenbergh, R., Stach, E., Grozac, J.R. and Dresser, B.J. (2000). *Thin Solid Films* **379**, 133–138.
- Darlington, E.H. and Valdre, U. (1975). *J. Phys. E* **8**, 321–324.
- Daulton, T.L., Kirk, M.A. and Rehn, L.E. (2000). *Philos. Mag. A* **80**, 809–842.
- Daulton, T.L., Little, B.J., Lowe, K. and Jones-Meehan, J. (2001). *Microsc. Microanal.* **7**, 470–485.
- Dehm, G. and Arzt, E. (2000). *App. Phys. Lett.* **77**, 1126–1128.
- Dehm, G., Wagner, T., Balk, T.J., Arzt, E. and Inkson, B.J. (2002). *J. Mat. Sci. Tech.* **18**, 113–317.
- Diebolt, L., Sharma, R., McKelvy, M. and Glaunsinger, W.S. (1995). *Mat. Res. Soc. Proc.* **404**, 183–188.

- Doan, J.C., Lee, S., Lee, S.-H., Flinn, P.A., Bravman, J.C. and Marieb, T.N. (2001). *J. App. Phys.* **89**, 7797–7808.
- Donnelly, S.E., Birtcher, R.C., Allen, C.W., Morrison, I., Furuya, K., Song, M., Mitsuishi, K. and Dahmen, U. (2002). *Science* **296**, 507–510.
- Donnet, D.M., Lewis, V.G., Chapman, J.N., O’Grady, K. and van Kesteren, H.W. (1993). *J. Phys. D* **26**, 1741–1745.
- Dougherty, L.M., Robertson, I.M. and Vetrano, J.S. (2003). *Acta Mat.* **51**, 4367–4378.
- Drucker, J., Sharma, R., Weiss, K. and Kouvetakis, J. (1995). *J. App. Phys.* **77**, 2846–2848.
- Du, X.W., Takeguchi, M., Tanaka, M. and Furuya, K. (2003). *App. Phys. Lett.* **82**, 1108–1110.
- Dutkiewicz, J., Bartsch, M., Miura, S., Messerschmidt, U. and Kato, H. (1995). *J. Phys. IV (Colloque)* **5(C2)**, 275–280.
- Fedina, L., Lebedev, O., Van Tendeloo, G. and Van Landuyt, J. (1997). *Proc. Microsc. Semicond. Mat.* **10**, 55–58.
- Fedina, L., Gutakovskii, A., Aseev, A., van Landuyt, J. and Vanhellemont, J. (1998). *Philos. Mag. A* **77**, 423–435.
- Follstaedt, D.M., Myers, S.M., Floro, J.A. and Lee, S.R. (1997). *Nucl. Inst. Methods Phys. Res. B* **127–128**, 375–378.
- Foo, K.H., Chiou, W.-A., Ishikawa, A., Fukushima, K. and Meshii, M. (2001). *Microsc. Microanal.* **7** (Supplement 2), 1216–1217.
- Fujisawa, S. and Kizuka, T. (2003a). *J. J. App. Phys. Part 2* **42**, L1182–L1184.
- Fujisawa, S. and Kizuka, T. (2003b). *Tribol. Lett.* **15**, 163.
- Fujita, S., Maruno, S., Watanabe, H. and Ichikawa, M. (1996). *App. Phys. Lett.* **69**, 638–640.
- Fukami, A., Fukushima, K. and Kohyama, N. (1991). In *Microstructure of Fine Grained Sediments from Mud to Shale* (Bennett, R.H., Bryant, W.R., Hulbert, M.H., Eds.) 321–331 (Springer-Verlag, New York).
- Furuya, K., Mitsuishi, K., Shimojo, M. and Takeguchi, M. (2003). *Rev. Adv. Mat. Sci.* **5**, 381–386.
- Gabrisch, H., Kjeldgaard, L., Vaernholt Olesen, A., Johnson, E. and Dahmen, U. (2000). *Mat. Res. Soc. Sympos. Proc.* **580**, 257–264.
- Gabrisch, H., Kjeldgaard, L., Johnson, E. and Dahmen, U. (2001). *Acta Mat.* **49**, 4259.
- Gai, P.L. (1998). *Adv. Mat.* **10**, 1259–1263.
- Gai, P.L., Kourtakis, K. and Ziemecki, S. (2000). *Microsc. Microanal.* **6**, 335–342.
- Gai, P.L. (2002a). *Top. Catalysis* **21**, 161–173.
- Gai, P.L. (2002b). *Microsc. Microanal.* **8**, 21–28.
- Gai, P.L. and Harmer, M.A. (2002). *Nano Lett.* **2**, 771–774.
- Gai, P.L. and Kourtakis, K. (1995). *Science* **267**, 661.
- Gai, P.L., Smith, B.C. and Owen, G. (1990). *Nature* **348**, 430–432.
- Gajdardziska-Josifovska, M., Malay, M.H. and Smith, D.J. (1997). *Surface Rev. Lett.* **4**, 655–669.
- Gajdardziska-Josifovska, M., Plass, R., Schofield, M.A., Giese, D.R. and Sharma, R. (2002). *J. Electron Microsc.* **51** (Supplement), S13–S26.
- Gajdardziska-Josifovska, M. and Sharma, R. (2005). *Microsc. Microanal.* **11**, 524–533.
- Gao, R., Wang, Z.L., Bai, Z., de Heer, W.A., Dai, L. and Gao, M. (2000). *Phys. Rev. Lett.* **85**, 622–625.
- Gao, R., Pan, Z. and Wang, Z.L. (2001). *App. Phys. Lett.* **78**, 1757–1759.
- Gao, Y. and Bando, Y. (2002). *Nature (London)* **415**, 599.
- Gao, Y., Liu, Y.Q. and Wu, K.H. (1996). *Mat. Res. Soc. Sympos. Proc.* **404**, 171–176.

- Gemperlova, J., Jacques, A., Gemperle, A., Vystavel, T., Zarubova, N. and Janecek, M. (2002). *Mat. Sci. Eng. A* **324**, 183–189.
- Ghica, C., Nistor, L., Bender, H., Steegen, A., Lauwers, A., Maex, K. and Van Landuyt, J. (2001). *J. Mat. Res.* **16**, 701–708.
- Gibson, J.M., Batstone, J.L. and Tung, R.T. (1987). *App. Phys. Lett.* **51**, 45–47.
- Gignac, L.M., Svilan, V., Clevenger, L.A., Cabral, C. and Lavoie, C. (1997). *Mat. Res. Soc. Sympos. Proc.* **441**, 225–230.
- Gloter, A., Suenaga, K., Kataura, H., Fujii, R., Kodama, T., Nishikawa, H., Ikemoto, I., Kikuchi, K., Suzuki, S., Achiba, Y. and Iijima, S. (2004). *Chem. Phys. Lett.* **390**, 462–466.
- Gnauck, P., Nisch, W. and Piles, E. (1998). *Proc. 14th Internat. Cong. Electron Microsc.* **3**, 515–516.
- Gottschalk, H., Hiller, N., Sauerland, S., Specht, P. and Alexander, H. (1993). *Phys. Status Solidi A* **138**, 547–555.
- Grant, A.W., Hu, Q.-H. and Kasemo, B. (2004). *Nanotechnology* **15**, 1175–1181.
- Grozea, D., Landree, E. and Marks, L.D. (1997). *App. Phys. Lett.* **71**, 2301–2303.
- Guder, S., Bartsch, M. and Messerschmidt, U. (2002). *Philos. Mag. A* **82**, 2737–2754.
- Haga, K., King, W.E., Merkle, K.L. and Meshii, M. (1986). *Nuc. Inst. Methods Phys. Res. B* **16**, 134–142.
- Hammar, M., LeGoues, F.K., Tersoff, J., Reuter, M.C. and Tromp, R.M. (1996). *Surface Sci.* **349**, 129–144.
- Han, X.D., Wang, R., Zhang, Z. and Yang, D.Z. (1997). *Mat. Lett.* **30**, 23–28.
- Hannon, J.B., Kodambaka, S., Ross, F.M. and Tromp, R.M. (2006). Nature, advanced online publication, 29 January (DOI 10.1038/nature04574).
- Hansen, P.L., Wagner, J.B., Helveg, S., Rostrup-Nielsen, J.R., Clausen, B.S. and Topsøe, H. (2002). *Science* **295**, 2053–2055.
- Hansen, T.W., Wagner, J.B., Hansen, P.L., Dahl, S., Topsøe, H. and Jacobsen, C.J.H. (2001). *Science* **294**, 1508–1510.
- Haque, M.A. and Saif, M.T.A. (2003). *Exp. Mech.* **43**, 248–255.
- Haque, M.A. and Saif, M.T.A. (2005). *J. Mat. Res.* **20**, 1769–1777.
- Haque, M.A. and Saif, M.T.A. (2004). *Proc. Nat. Acad. Sci.* **101**, 6335–6340.
- Harada, K., Matsuda, T., Bonevich, J., Igarashi, M., Kondo, S., Pozzi, G., Kawabe, U. and Tonomura, A. (1992). *Nature* **360**, 51–53.
- Harada, K., Kamimura, O., Kasai, H., Matsuda, F., Tonomura, A. and Moshchalkov, V.V. (1996a). *Science* **274**, 1167–1170.
- Harada, K., Kasai, H., Kamimura, O., Matsuda, T., Tonomura, A., Okayasu, S. and Kazumata, Y. (1996b). *Phys. Rev. B* **53**, 9400–9405.
- Harada, I., Kasai, H., Matsuda, T., Yamasaki, M. and Tonomura, A. (1997a). *J. Electron Microsc.* **46**, 227–232.
- Harada, K., Matsuda, T., Kasai, H., Kamimura, O. and Tonomura, A. (1997b). *Ultramicroscopy* **67**, 133–141.
- Hashimoto, A., Suenaga, K., Gloter, A., Urita, K. and Iijima, S. (2004). *Nature* **430**, 870–873.
- Hattar, K., Han, J., Saif, M.T.A. and Robertson, I.M. (2005). *J. Mat. Res.* **20**, 1869–1877.
- Hattenhauer, R., Bartsch, M., Messerschmidt, U., Haasen, P. and Wilbrandt, P.-J. (1994). *Philos. Mag. A* **70**, 447–461.
- Haussler, D., Bartsch, M., Aindow, M., Jones, I.P. and Messerschmidt, U. (1999). *Philos. Mag. A* **79**, 1045–1071.
- He, H. and Tan, X. (2004). *App. Phys. Lett.* **85**, 3187–3189.
- He, H. and Tan, X. (2005). *Phys. Rev. B* **72**, 24102.

- Helveg, S., Lopez-Cartes, C., Sehested, J., Hansen, P.L., Clausen, B.S., Rostrup-Nielsen, J.R., Abild-Pedersen, F. and Norskov, J.K. (2004). *Nature* **427**, 426–429.
- Hirayama, H., Kawamoto, Y., Ohshima, Y. and Takayanagi, K. (2001). *App. Phys. Lett.* **79**, 1169–1171.
- Homma, Y. and Finnie, P. (2002). *J. Crystal Growth* **237–239**, 28–34.
- Horiki, M., Arai, S., Satoh, Y. and Kiritani, M. (1998). *J. Nucl. Mat.* **255**, 165–173.
- Horiuchi, S., Cantoni, M., Uchida, M., Tsuruta, T. and Matsui, Y. (1998). *App. Phys. Lett.* **73**, 1293–1294.
- Howe, J.M. (1996). *Philos. Mag. A* **74**, 761–775.
- Howe, J.M. (1997). *Interf. Mat.* (Wiley, New York).
- Howe, J.M. (1998). *Mat. Trans, JIM* **39**, 3–23.
- Howe, J.M., Murray, T.M., Moore, K.T., Csontos, A.A., Tsai, M.M., Garg, A. and Benson, W.E. (1998). *Microsc. Microanal.* **4**, 235–247.
- Howe, J.M., Reynolds, W.T., Jr. and Vasudevan, V.K. (2004). *Z. Met.* **95**, 275–278.
- Huang, J.Y., Chen, S., Jo, S.H., Wang, Z., Han, D.X., Chen, G., Dresselhaus, M.S. and Ren, Z.F. (2005). *Phys. Rev. Lett.* **94**, 236802/1–4.
- Hueging, N., Luysberg, M., Urban, K., Buca, D. and Mantl, S. (2005). *App. Phys. Lett.* **86**, 42112–42113.
- Hugo, R.C. and Hoagland, R.G. (1998). *Scripta Mat.* **38**, 523–529.
- Hugo, R.C. and Hoagland, R.G. (1999). *Scripta Mat.* **41**, 1341–1946.
- Hugo, R.C., Kung, H., Weertman, J.R., Mitra, R., Knapp, J.A. and Follstaedt, D.M. (2003). *Acta Mat.* **51**, 1937–1943.
- Hull, R. and Bean, J. (1994). *Mat. Res. Soc. Bull.* **19** (6), 32–37.
- Hull, R., Bean, J.C. and Buescher, C. (1989). *J. App. Phys.* **66**, 5837–5843.
- Hull, R., Bean, J.C., Bahnck, D., Peticolas, L.J., Jr., Short, K.T. and Unterwald, F.C. (1991). *J. App. Phys.* **70**, 2052–2065.
- Ichikawa, M., Fujita, S. and Maruno, S. (1998). *Proc. 14th Internat. Cong. Electron Microsc.* **13**, 147–148.
- Ichikawa, M. (1999). *J. Phys. Cond. Matt.* **11**, 9861–9870.
- Ii, S., Nishida, M., Murakami, Y. and Shindo, D. (2003). *J. Phys. IV (Proceedings)* **112**, 1035–1038.
- Ii, S., Iwamoto, C., Matsunaga, K., Yamamoto, T., Yoshiya, M. and Ikuhara, Y. (2004). *Philos. Mag.* **84**, 2767–2775.
- Iijima, S. and Ichihashi, T. (1986). *Phys. Rev. Lett.* **56**, 616–619.
- Inoue, M., Suzuki, K., Amasuga, H., Mera, Y. and Maeda, K. (1998). *J. App. Phys.* **83**, 1953–1957.
- Inui, H., Mori, H., Suzuki, A. and Fujita, H. (1992). *Philos. Mag. B* **65**, 1–14.
- Isshiki, T., Ohta, M., Nishio, K. and Shiojiri, M. (1995). *Thin Solid Fims* **271**, 84–87.
- Isshiki, T., Nishio, K., Deguchi, Y., Yamamoto, H. and Shiojiri, M. (1998). *Proc. 14th Internat. Cong. Electron Microsc.* **3**, 521–522.
- Janecek, M., Louchet, F., Doisneau-Cottignies, B. and Brechet, Y. (2000). *Philos. Mag. A* **80**, 1605–1619.
- Jencic, I., Bench, M.W., Robertson, I.M. and Kirk, M.A. (1995). *J. App. Phys.* **78**, 974–982.
- Jiang, B., Tadakit, T., Mori, H. and Hsu, T.Y. (1997). *Mat. Trans. JIM* **38**, 1072–1077 and 1078–1082.
- Jiang, J.C. and Pan, X.Q. (2000). *Philos. Mag. Lett.* **80**, 271–279.
- Jin, M., Minor, A.M., Stach, E.A. and Morris, J.W., Jr. (2004). *Acta Mat.* **52**, 5381–5387.
- Jin, C.H., Wang, J.Y., Wang, M.S., Su, J. and Peng, L.M. (2005). *Carbon* **43**, 1026–1031.

- Johnson, E., Johansen, A., Nelson, C. and Dahmen, U. (2002). *J. Electron Microsc.* **51** (Supplement), S201–209.
- Johnson, E., Levinsen, M.T., Steenstrup, S., Prokofjev, S., Zhilin, V., Dahmen, U. and Radetic, T. (2004). *Philos. Mag.* **84**, 2663–2673.
- Johnson, K.D. and Dravid, V.P. (1999). *Microsc. Microanal.* **5**, 428–436.
- Johnson, K.D. and Dravid, V.P. (2000). *Interface Sci.* **8**, 189–198.
- Johnston, A.B., Chapman, J.N., Khamsehpour, B. and Wilkinson, C.D.W. (1996). *J. Phys. D* **29**, 1419–1427.
- Kaiser, M., van Pieterse, L. and Verheijen, M.A. (2004). *J. App. Phys.* **96**, 3193–3198.
- Kamino, T., Yaguchi, T., Ukiana, M., Yasutomi, Y. and Saka, H. (1995). *Mat. Transac, JIM* **36**, 73–75.
- Kamino, T., Sasaki, K. and Saka, H. (1997a). *Microsc. Microanal.* **3**, 393–408.
- Kamino, T., Yaguchi, T., Tomita, M. and Saka, H. (1997b). *Philos. Mag. A* **75**, 105–114.
- Kammler, M., Hull, R., Reuter, M.C. and Ross, F.M. (2003). *App. Phys. Lett.* **82**, 1093–1095.
- Kammler, M., Chidambarrao, D., Schwarz, K.W., Black, C.T. and Ross, F.M. (2005). *App. Phys. Lett.* **87**, 133116.
- Kaouache, B., Gergaud, P., Thomas, O., Bostrom, O. and Legros, M. (2003). *Microelectr. Eng.* **70**, 447–454.
- Kassner, M.E., Wall, M.A. and Delos-Reyes, M.A. (1997). *Metallurg. Mat. Transac. A* **28**, 595–609.
- Kato, T., Nunome, K., Morimoto, Y., Nishimura, K. and Saka, H. (2000). *Philos. Mag. Lett.* **80**, 187–192.
- Keller, R.-M., Sigle, W., Baker, S.P., Kraft, O. and Arzt, E. (1997). *Mat. Res. Soc. Sympos. Proc.* **436**, 221–226.
- Keller-Flaig, R.-M., Legros, M., Sigle, W., Gouldstone, A., Hemker, K.J., Suresh, S. and Arzt, E. (1999). *J. Mat. Res.* **14**, 4673–4676.
- King, W.E., Merkle, K.L. and Meshii, M. (1981). *Phys. Rev. B* **23**, 6319–6334.
- Kiritani, M. (1999). *Radiat. Eff. Def. Sol.* **148**, 233–267.
- Kirk, K. J., Chapman, J.N., McVitie, S., Aitchison, P.R. and Wilkinson, C.D.W. (1999a). *App. Phys. Lett.* **75**, 3683–3685.
- Kirk, M.A., Jenkins, M.L. and Fukushima, H. (1999b). *Mat. Res. Soc. Sympo. Proc.* **540**, 521–526.
- Kizuka, T. (1998a). *Phys. Rev. Lett.* **81**, 4448–4451.
- Kizuka, T. (1998b). *Phys. Rev. B* **57**, 11158–11163.
- Kizuka, T. and Hosoki, K. (1999). *App. Phys. Lett.* **75**, 2743–2745.
- Kizuka, T. and Tanaka, N. (1997a). *Phys. Rev. B* **56**, R10079–10082.
- Kizuka, T. and Tanaka, N. (1997b). *Philos. Mag. Lett.* **76**, 289–297.
- Kizuka, T., Uhmehara, S. and Fujisawa, S. (2001a). *J. J. App. Phys.*, Part 2 **40**, L71–74.
- Kizuka, T., Ohmi, H., Sumi, T., Kumazawa, K., Deguchi, S., Naruse, M., Fujisawa, S., Sasaki, S., Yabe, A. and Enomoto, Y. (2001b). *J. J. App. Phys.* **40**, L170–L173.
- Kizuka, T., Tanaka, N., Deguchi, S. and Naruse, M. (1998c). *Microsc. Microanal.* **4**, 218–225.
- Kleinschmit, M.W., Yeadon, M. and Gibson, J.M. (1999). *App. Phys. Lett.* **75**, 3288–3290.
- Klie, R.F. and Browning, N. (2001). *J. Electron Microsc.* **50** (Supplement), S59.
- Kociak, M., Suenaga, K., Hirahara, K., Saito, Y., Nakahira, T. and Iijima, S. (2002). *Phys. Rev. Lett.* **89**, 155501.
- Koizumi, H., Oshima, Y., Kondo, Y. and Takayanagi, K. (2001). *Ultramicroscopy* **88**, 17–24.

- Kolar, H.R., Spence, J.C.H. and Alexander, H. (1996). *Phys. Rev. Lett.* **77**, 4031–4034.
- Komatsu, M., Mori, H. and Iwasaki, K. (1994). *J. Amer. Ceramic Soc.* **77**, 839–842.
- Komatsu, M., Matsukawa, Y., Yasunaga, K. and Kiritani, M. (2003). *Mat. Sci. Eng. A* **350**, 25–29.
- Kondo, Y. and Takayanagi, K. (2000). *Science* **289**, 606–608.
- Konno, T.J. and Sinclair, R. (1992). *Philos. Mag. B* **66**, 749–765.
- Konno, T.J. and Sinclair, R. (1995a). *Philos. Mag. B* **71**, 179–199.
- Konno, T.J. and Sinclair, R. (1995b). *Philos. Mag. B* **71**, 163–178.
- Konno, T.J. and Sinclair, R. (1995c). *Acta Met. Mat.* **43**, 471–484.
- Kooi, B.J. and De Hosson, J.T.M. (2001). *Mat. Res. Soc. Sympos. Proc.* **654**, AA4.1.1–12.
- Kooi, B.J. and De Hosson, J.T.M. (2004). *J. App. Phys.* **95**, 4714–4721.
- Kooi, B.J., Groot, W.M.G. and De Hosson, J.T.M. (2004). *J. App. Phys.* **95**, 924–932.
- Krishnamurthy, M., Drucker, J.S. and Venables, J.A. (1991). *Mat. Res. Soc. Sympos. Proc.* **202**, 77–82.
- Krishnan, A., Dujardin, E., Ebbesen, T.W., Yianilos, P.N. and Treacy, M.M.J. (1998). *Phys. Rev. B* **58**, 14013–14019.
- Krishnan, A., Bisher, M.E. and Treacy, M.M.J. (1999). *Mat. Res. Soc. Sympos. Proc.* **541**, 475–480.
- Krishnan, A., Treacy, M.M.J., Bisher, M.E., Chandra, P. and Littlewood, P.B. (2000). *Ameri. Inst. Phys. Conf. Proc.* **535**, 191–200.
- Krishnan, A., Treacy, M.M.J., Bisher, M.E., Chandra, P. and Littlewood, P.B. (2002). *Integ. Ferroelectrics* **43**, 31–49.
- Kruml, T., Caillard, D., Dupas, C. and Martin, J.-L. (2002). *J. Phys. Cond. Matt.* **14**, 12897–12902.
- Kuninori, T., Sukedai, E. and Hashimoto, H. (1996). *Mat. Trans, JIM* **37**, 1404–1407.
- Kuzumaki, T., Sawada, H., Ichinose, H., Horiike, Y. and Kizuka, T. (2001). *App. Phys. Lett.* **79**, 4580–4582.
- Kuzumaki, T., Horiike, Y., Kizuka, T., Kona, T., Oshima, C. and Mitsuda, Y. (2004). *Diamond Rel. Mat.* **13**, 1907–1913.
- Larsson, M.W., Wallenberg, L.R., Persson, A.I. and Samuelson, L. (2004). *Microsc. Microanal.* **10**, 41–46.
- Latyshev, A.V., Nasimov, D.A., Savenko, V.N. and Aseev, A.L. (2000). *Thin Sol. Films* **367**, 142–148.
- Lau, J.T., Prybyla, J.A. and Theiss, S.K. (2000). *App. Phys. Lett.* **76**, 164–166.
- Lau, J.W., Beleggia, M., Schofield, M.A., Neumark, G.F. and Zhu, Y. J. (2005). *App. Phys.* **97**, 10E702/1–3.
- Lavagne, S., Levade, C. and Vanderschaeve, G. (2001). *Proc. Microsc. Semicond. Mat.* **2001**, 219–224.
- Lee, H.-J., Ni, H., Wu, D.T. and Ramirez, A.G. (2005). *App. Phys. Lett.* **87**, 124102.
- Lee, J.G. and Mori, H. (2003). *J. Electron Microsc.* **52**, 57–62.
- Lee, J.G. and Mori, H. (2004a). *Philos. Mag.* **84**, 2675–2686.
- Lee, J.G. and Mori, H. (2004b). *Phys. Rev. Lett.* **93**, 235501/1–4.
- Lee, J.G., Mori, H. and Yasuda, H. (2002a). *Phys. Rev. B* **65**, 132106/1–4.
- Lee, J.G., Taguchi, E. and Mori, H. (2002b). *J. Electron Microsc.* **51** (Suppl), S195–200.
- LeGoues, F.K., Reuter, M.C., Tersoff, J., Hammar, M. and Tromp, R.M. (1994). *Phys. Rev. Lett.* **73**, 300–303.
- LeGoues, F.K., Tersoff, J., Reuter, M.C., Hammar, M. and Tromp, R. (1995). *App. Phys. Lett.* **67**, 2317–2319.

- LeGoues, F.K., Hammar, M., Reuter, M.C. and Tromp, R.M. (1996). *Surface Sci.* **349**, 249–266.
- Legros, M. and Caillard, D. (2001). *J. Microsc.* **203**, 90–98.
- Legros, M., Couret, A. and Caillard, D. (1996). *Philos. Mag.* **73**, 61–80 81–99.
- Legros, M., Minonishi, Y. and Caillard, D. (1997). *Philos. Mag. A* **76**, 995–1011 and 1013–1032.
- Legros, M., Hemker, K.J., Gouldstone, A., Suresh, S., Keller-Flaig, R.M. and Arzt, E. (2002). *Acta Mat.* **50**, 3435–3452.
- Legros, M., Kaouache, B., Gergaud, P., Thomas, O., Dehm, G., Balk, T.J. and Arzt, E. (2005). *Philos. Mag.* **85**, 3541–3552.
- Levade, C., Faress, A. and Vanderschaeve, G. (1994). *Philos. Mag. A* **69**, 855–870.
- Lim, C.K., Chapman, J.N., Rahman, M., Johnston, A.B. and O'Donnell, D.O. (2002). *J. Phys. D* **35**, 2344–2351.
- Lim, C.K., Chapman, J.N., Rahman, M., Johnston, A.B. and O'Donnell, D.O. (2004). *J. App. Phys.* **95**, 1510–1515.
- Lin, X., Murray, C. and Dravid, V.P. (1998). *Proc. Microsc. Microanal.* **1998**, 552.
- Liu, J., Takeguchi, M., Tanaka, M., Yasuda, H. and Furuya, K. (2001). *J. Electron Microsc.* **50**, 541–544.
- Liu, R.-J., Crozier, P.A., Smith, C.M., Hucul, D.A., Blackson, J. and Salaita, G. (2004). *Microsc. Microanal.* **10**, 77–85.
- Liu, X.H., Ross, F.M. and Schwarz, K.W. (2000). *Phys. Rev. Lett.* **85**, 4088–4091.
- Liu, Z., Bando, Y., Mitome, M. and Zhan, J. (2004). *Phys. Rev. Lett.* **93**, 095504/1–4.
- Liu, X.X., Chapman, J.N., McVitie, S. and Wilkinson, C.D.W. (2004). *App. Phys. Lett.* **84**, 4406–4408.
- Lo, W. and Spence, J. (1992). *Ultramicroscopy* **48**, 433.
- Louchet, F., Blandin, J.-J. and Veron, M. (2001). *Adv. Eng. Mat.* **3**, 608–612.
- Lucadamo, G. and Medlin, D.L. (2002). *Acta Mat.* **50**, 3045–3055.
- Ma, X.-G. and Komvopoulos, K. (2005). *J. Mat. Res.* **20**, 1808–1813.
- Maeda, K., Suzuki, K., Yamashita, Y. and Mera, Y. (2000). *J. Phys. Cond. Matt.* **12**, 10079–10091.
- Malaplate, J., Caillard, D. and Couret, A. (2004). *Philos. Mag.* **84**, 3671–3687.
- Marks, L.D. (1983). *Phys. Rev. Lett.* **51**, 1000.
- Marks, L.D., Bengu, E., Collazo-Davila, C., Grozea, D., Landree, E., Leslie, C. and Sinkler, W. (1998). *Surface Rev. Lett.* **5**, 1087–1106.
- Maruno, S., Fujita, S., Watanabe, H., Kusumi, Y. and Ichikawa, M. (1996). *App. Phys. Lett.* **68**, 2213–2215.
- Mathes, D.T., Hull, R., Choquette, K.D., Geib, K.M., Allerman, A.A., Guenter, J.K., Hawkins, B. and Hawthorne, R.A. (2003). *Proc. SPIE* **4994**, 67–82.
- Matsuda, T., Harada, K., Kasai, H., Kamimura, O. and Tonomura, A. (1996). *Science* **271**, 1393–1395.
- Matsuda, T., Kamimura, O., Kasai, H., Harada, K., Yoshida, T., Akashi, T., Tonomura, A., Nakayama, Y., Shimoyama, J., Kishio, K., Hanaguri, T. and Kitazawa, K. (2001). *Science* **294**, 2136–2138.
- Matsumoto, H., Sukedai, E. and Hashimoto, H. (1999). *Mat. Trans. JIM* **40**, 1436–1443.
- Matsunaga, K., Ii, S., Iwamoto, C., Yamamoto, T. and Ikuhara, Y. (2004). *Nano-technology* **15**, S376–S381.
- McKelvy, M.J., Sharma, R., Chizmeshya, A.V.G., Carpenter, R.W. and Streib, K. (2001). *Chem. Mat.* **13**, 921–926.
- McVitie, S. and Chapman, J.N. (1997). *Microsc. Microanal.* **3**, 146–153.

- Medlin, D.L., Campbell, G.H. and Carter, C.R. (1998). *Acta Mat.* **46**, 5135–5142.
- Mera, Y., Suzuki, K. and Maeda, K. (2003). *Phys. B* **340–342**, 488–491.
- Merkle, K.L. and Thompson, L.J. (2001). *Mat. Lett.* **48**, 188–193.
- Merkle, K.L., Thompson, L.J. and Phillipp, F. (2002a). *Philos. Mag. Lett.* **82**, 589–597.
- Merkle, K.L., Thompson, L.J. and Phillipp, F. (2002b). *Phys. Rev. Lett.* **88**, 225501/1–4.
- Messerschmidt, U. (2001). *J. Microsc.* **203**, 68–71.
- Messerschmidt, U. and Bartsch, M. (2003). *Mat. Chem. Phys.* **81**, 518–523.
- Messerschmidt, U., Baither, D., Baufeld, B. and Bartsch, M. (1997). *Mat. Sci. Eng. A* **233**, 61–74.
- Messerschmidt, U., Baither, D., Bartsch, M., Baufeld, B., Geyer, B., Guder, S., Wasilkowska, A., Czyrska-Filemonowicz, A., Yamaguchi, M., Feuerbacher, M. and Urban, K. (1998). *Microsc. Microanal.* **4**, 226–234.
- Messerschmidt, U., Bartsch, M., Feuerbacher, M., Geyer, B. and Urban, K. (1999). *Philos. Mag. A* **79**, 2123–2135.
- Minoda, H. and Yagi, K. (1996). *J. Crystal Growth* **163**, 48–53.
- Minoda, H. and Yagi, K. (1999). *Surface Sci.* **437**, L761–766.
- Minoda, H., Yagi, K., Meyer zu Heringdorf, F.-J., Meier, A., Kahler, D. and Horn von Hoegen, M. (1999). *Phys. Rev. B* **59**, 2363–2375.
- Minor, A.M., Morris, J.W. and Stach, E.A. (2001). *App. Phys. Lett.* **79**, 1625–1627.
- Minor, A.M., Lilleodden, E.T., Jin, M., Stach, E.A., Chrzan, D.C. and Morris, J.W. (2005). *Philos. Mag.* **85**, S323–S330.
- Missana, T., Afonso, C.N., Petford-Long, A.K. and Doole, R.C. (1999). *Philos. Mag. A* **79**, 2577–2590.
- Mitome, M., Takayanagi, K. and Tanishiro, Y. (1990). *Phys. Rev. B* **42**, 7238–7241.
- Mompious, F., Caillard, D. and Feuerbacher, M. (2004). *Philos. Mag.* **84**, 2777.
- Mori, H., Yasuda, H., Koyama, H., Sumida, N. and Fujita, H. (1992). *Nucl. Inst. Methods Phys. Res. B* **63**, 291–296.
- Mori, H., Okabayashi, H. and Komatsu, M. (1997). *Thin Sol. Films* **300**, 25–29.
- Mori, H. (1998). *Mat. Trans. JIM* **39**, 959–966.
- Mori, S., Yamamoto, N., Koyama, Y. and Uesu, Y. (1997). *Phys. Rev. B* **55**, 11212–11217.
- Morkved, T.L., Lopes, W.A., Hahm, J., Sibener, S.J. and Jaeger, H.M. (1998). *Polymer* **39**, 3871–3875.
- Murakami, Y. and Shindo, D. (2001). *Philos. Mag. Lett.* **81**, 631–638.
- Murakami, Y., Shindo, D., Oikawa, K., Kainuma, R. and Ishida, K. (2002). *Acta Mat.* **50**, 2173–2184.
- Naitoh, Y., Takayanagi, K., Oshima, Y. and Hirayama, H. (2000). *J. Electron Microsc.* **49**, 211–216.
- Nath, R., Soo, C.W., Boothroyd, C.B., Yeadon, M., Chi, D.Z., Sun, H.P., Chen, Y.B., Pan, X.Q. and Foo, Y.L. (2005). *App. Phys. Lett.* **86**, 201908–1–3.
- Niihara, K., Sekino, T., Nakahira, A., Kamino, T. and Yaguchi, T. (1996). *Mat. Lett.* **27**, 121–124.
- Ning, X.G., Avalos-Borja, M. and Ye, H.Q. (1996). *J. App. Phys.* **79**, 125–128.
- Nishizawa, H., Hori, F. and Oshima, R. (2002). *J. Crystal Growth* **236**, 51–58.
- Nucci, J., Krämer, S., Arzt, E. and Volkert, C.A. (2005). *J. Mat. Res.* **20**, 1851–1859.
- Ohmura, T., Minor, A.M., Stach, E.A. and Morris, J.W., Jr. (2004). *J. Mat. Res.* **19**, 3626–3632.

- Ohnishi, H., Kondo, Y. and Takayanagi, K. (1998a). *Surface Sci.* **415**, L1061–1064.
- Ohnishi, H., Kondo, Y. and Takayanagi, K. (1998b). *Nature* **395**, 780–783.
- Okabayashi, H., Kitamura, H., Komatsu, M. and Mori, H. (1996). *App. Phys. Lett.* **68**, 1066–1068.
- Ono, K., Arakawa, K., Hojou, K., Oohasi, M., Birtcher, R.C. and Donnelly, S.E. (2002). *J. Electron Microsc.* **51** (Supplement), S245–251.
- Oshima, Y., Hirata, T., Yokoyama, T., Hirayama, H. and Takayanagi, K. (2000). *Surface Sci.* **465**, 81–89.
- Oshima, Y., Koizumi, H., Mouri, K., Hirayama, H., Takayanagi, K. and Kondo, Y. (2002). *Phys. Rev. B* **65**, 121401/1–4.
- Oshima, Y., Kondo, Y. and Takayanagi, K. (2003a). *J. Electron Microsc.* **52**, 49–55.
- Oshima, Y., Onga, A. and Takayanagi, K. (2003b). *Phys. Rev. Lett.* **91**, 205503/1–4.
- Oshima, Y., Mouri, K., Hirayama, H. and Takayanagi, K. (2003c). *Surface Sci.* **531**, 209–216.
- Pan, X.-Q., Hu, M.-S., Yao, M.-H. and Feng, D. (1985). *Phys. Stat. Sol. A* **92**, 57–68.
- Park, H.S., Murakami, Y., Shindo, D., Chernenko, V.A. and Kanomata, T. (2003). *App. Phys. Lett.* **83**, 3752–3754.
- Park, H.S., Park, Y.-G., Gao, Y., Shindo, D. and Inoue, M. (2005). *J. App. Phys.* **97**, 33908-1–4.
- Parker, M.A., Sigmon, T.W. and Sinclair, R. (1986). *Mat. Res. Soc. Sympos. Proc.* **62**, 311–322.
- Parker, M.A. and Sinclair, R. (1986). *Nature* **322**, 531–533.
- Persson, A.I., Larsson, M.W., Stenstrom, S., Ohlsson, B.J., Samuelson, L. and Wallenberg, L.R. (2004). *Nat. Mat.* **3**, 677–681.
- Petford-Long, A.K., Doole, R.C., Afonso, C.N. and Solis, J. (1995). *J. App. Phys.* **77**, 607–613.
- Pettinari, F., Couret, A., Caillard, D., Molenat, G., Clement, N. and Coujou, A. (2001). *J. Microsc.* **203**, 47–56.
- Poncharal, P., Frank, S., Wang, Z.L. and de Heer, W.A. (1999). *Eur. Phys. J. D* **9**, 77–79.
- Poncharal, P., Berger, C., Yan Yi, Wang, Z.L. and de Heer, W.A. (2002). *J. Phys. Chem. B* **106**, 12104–12118.
- Poppa, H. (1965). *J. Vac. Sci. Tech.* **2**, 42–48.
- Poppa, H. (2004). *J. Vac. Sci. Tech. A* **22**, 1931–1947.
- Portavoce, A., Kammler, M., Hull, R., Reuter, M.C., Copel, M. and Ross, F.M. (2004). *Phys. Rev. B* **70**, 195306-1–9.
- Portavoce, A., Hull, R. and Ross, F.M. (2006). In *Lateral Alignment of Epitaxial Quantum Dots* (O. Schmidt, Ed.). Springer, New York.
- Portier, X. and Petford-Long, A.K. (2000). *App. Phys. Lett.* **76**, 754–756.
- Portier, X., Petford-Long, A.K., Doole, R.C., Anthony, T.C. and Brug, J.A. (1997). *App. Phys. Lett.* **71**, 2032–2034.
- Portier, X., Tsymbal, E. Yu., Petford-Long, A.K., Anthony, T.C. and Brug, J.A. (1998). *Phys. Rev. B* **58**, R591–594.
- Portier, X., Petford-Long, A.K., Anthony, T.C. and Brug, J.A. (1999a). *App. Phys. Lett.* **75**, 1290–1292.
- Portier, X., Petford-Long, A.K., Anthony, T.C. and Brug, J.A. (1999b). *J. App. Phys.* **85**, 4120–4126.
- Prybyla, J.A., Riege, S.P., Grabowski, S.P. and Hunt, A.W. (1998). *App. Phys. Lett.* **73**, 1083–1085.
- Pyo, S.G. and Kim, N.J. (2005). *J. Mat. Res.* **20**, 1888–1901.
- Radisic, A. (2004). Ph.D. Thesis, The Johns Hopkins University.

- Radisic, A., Vereecken, P.M., Hannon, J.B., Searson, P.C. and Ross, F.M. (2006). *Nano Lett.* **6**, 238–242.
- Ren, S.B., Lu, C.J., Shen, H.M. and Wang, Y.N. (1994). *Mat. Res. Soc. Sympos. Proc.* **404**, 95–100.
- Riege, S.P., Prybyla, J.A. and Hunt, A.W. (1996). *App. Phys. Lett.* **69**, 2367–2369.
- Robertson, I.M. and Teter, D. (1998). *Microsc. Res. Tech.* **42**, 260–269.
- Roddatis, V.V., Kuznetsov, V.L., Butenko, Yu, V., Su, D.S. and Schlogl, R. (2002). *Phys. Chem. Chem. Phys.* **4**, 1964–1967.
- Ross, F.M. (2000). *IBM J. Res. Dev.* **44**, 489–501.
- Ross, F.M. and Gibson, J.M. (1991). *Phys. Rev. Lett.* **68**, 1782–1785.
- Ross, F.M., Hull, R., Bahnck, D., Bean, J.C., Peticolas, L.J. and King, C.A. (1993). *App. Phys. Lett.* **62**, 1426–1428.
- Ross, F.M., Kola, R.R., Hull, R. and Bean, J.C. (1994a). *Mat. Res. Soc. Sympo. Proc.* **318**, 697–702.
- Ross, F.M., Gibson, J.M. and Twesten, R.D. (1994b). *Surface Sci.* **310**, 243–266.
- Ross, F.M., Tersoff, J. and Tromp, R.M. (1998). *Phys. Rev. Lett.* **80**, 984–987.
- Ross, F.M., Bennett, P.A., Tromp, R.M., Tersoff, J. and Reuter, M. (1999a). *Micron* **30**, 21–32.
- Ross, F.M., Tromp, R.M. and Reuter, M.C. (1999b). *Science* **286**, 1931–1934.
- Ross, F.M., Kammler, M., Walsh, M.E. and Reuter, M.C. (2004). *Microsc. Microanal.* **10**, 105–111.
- Ross, F.M., Tersoff, J., Kodambaka, S. and Reuter, M.C. (2006). *Springer Proceedings in Physics* **107**, 283–286.
- Ross, F.M., Tersoff, J. and Reuter, M.C. (2005). *Phys. Rev. Lett.* **95**, 146104/1–4.
- Ru, Q., Okamoto, M., Kondo, Y. and Takayanagi, K. (1996). *Chem. Phys. Lett.* **259**, 425–431.
- Ruault, M.O., Fortuna, F., Bernas, H., Chaumont, J., Kaitasov, O. and Borodin, V.A. (2005). *J. Mat. Res.* **20**, 1758–1768.
- Ruault, M.O., Fortuna, F., Bernas, H., Ridgway, M.C. and Williams, J.S. (2002). *App. Phys. Lett.* **81**, 2617–2619.
- Saka, H., Arai, S., Tsukimoto, S. and Sasaki, K. (1999). *Mat. Sci. Forum* 294–296, 617–620.
- Sasaki, K. and Saka, H. (1997). *Mat. Res. Soc. Sympos. Proc.* **466**, 185–190.
- Sayagues, M.J. and Hutchison, J.L. (1996). *J. Sol. State Chem.* **124**, 116–122.
- Sayagues, M.J. and Hutchison, J.L. (2002). *J. Sol. State Chem.* **163**, 137–143.
- Schaper, A.K., Phillipp, F. and Hou, H. (2005). *J. Mat. Res.* **20**, 1844–1850.
- Schneider, A., Hoffmann, H. and Zweck, J. (2003). *J. Magn. Magn. Mat.* **257**, 1–10.
- Schneider, M., Hoffmann, H. and Zweck, J. (2001). *App. Phys. Lett.* **79**, 3113–3115.
- Schneider, M., Hoffmann, H., Otto, S., Haug, Th. and Zweck, J. (2002). *J. App. Phys.* **92**, 1466–1472.
- Schryvers, D. and Ma, Y. (1995). *Mat. Lett.* **23**, 105–111.
- Schryvers, D., Lahjouji, D.E., Slootmaekers, B. and Potapov, P.L. (1996). *Scripta Mat.* **35**, 1235–1241.
- Schryvers, D., Yandouzi, M. and Toth, L. (1998). *Thin Sol. Films* **326**, 126–133.
- Senda, Y., Sasaki, K. and Saka, H. (2004). *Philos. Mag.* **84**, 2635–2649.
- Shah, J. (2004). *Philos. Mag.* **84**, 2749–2765.
- Shan, Z., Stach, E.A., Wiezorek, J.M.K., Knapp, J.A., Forstaeedt, D.M. and Mao, S.X. (2004). *Science* **305**, 654–657.
- Sharma, R. (2001). *Microsc. Microanal.* **7**, 494–506.

- Sharma, R. (2005). *J. Mat. Res.* **20**, 1695–1707.
- Sharma, R. and Iqbal, Z. (2004). *App. Phys. Lett.* **84**, 990–992.
- Sharma, R., Crozier, P.A., Kang, Z.C. and Eyring, L. (2004). *Philos. Mag.* **84**, 2731–2747.
- Sharma, R., McKelvy, M.J., Bearat, H., Chizmeshya, A.V.G. and Carpenter, R. W. (2004b). *Philos. Mag.* **84**, 2711–2729.
- Sharma, R., Naedele, D. and Schweda, E. (2001). *Chem. Mat.* **13**, 4014–4018.
- Sharma, R., Rez, P., Treacy, M.M.J. and Stuart, J. (2005). *J. Electron Microsc.* **54**, 231–237.
- Shih, W.C. and Green, A.L. (1995). *Mat. Res. Soc. Sympo. Proc.* **356**, 417.
- Shimojo, M., Mitsuishi, K., Tanaka, M., Han, M. and Furuya, K. (2004). *J. Microsc.* **214**, 76–79.
- Sidorov, M.V., Kardynal, B. and Smith, D.J. (1998a). *Microsc. Microanal.* **4**, 317–324.
- Sidorov, M.V., McKelvy, M.J., Sharma, R. and Glaunsinger, W.S. (1998b). *J. Sol. State Chem.* **141**, 330–337.
- Sinclair, R. (1994). *Mat. Res. Soc. Bull.* **19**(6), 26–31.
- Sinclair, R., Parker, M.A. and Kim, K.B. (1987). *Ultramicroscopy* **23**, 383–395.
- Sinclair, R., Yamashita, T., Parker, M.A., Kim, K.B., Holloway, K. and Schwartzman, A.F. (1988). *Acta Cryst. A* **44**, 965–975.
- Sinclair, R., Itoh, T. and Chin, R. (2002). *Microsc. Microanal.* **8**, 288–304.
- Smeeton, T.M., Kappers, M.J., Barnard, J.S., Vickers, M.E. and Humphreys, C. J. (2003). *App. Phys. Lett.* **83**, 5419–5421.
- Smith, D.J., Petford-Long, A.K., Wallenberg, L.R. and Bovin, J.-O. (1986). *Science* **233**, 872–875.
- Smith, D.J., Gajdardziska-Josifovska, M., Lu, P., McCartney, M.R., Podbrdsky, J., Swann, P.R. and Jones, J.S. (1993). *Ultramicroscopy* **49**, 26–36.
- Snoeck, E. and Roucau, C. (1992). *Phys. Rev. B* **45**, 12720–12724.
- Snoeck, E., Normand, L., Thorel, A. and Roucau, C. (1994). *Phase Trans. A* **46**, 77–88.
- Soer, W.A., De Hosson, J.T.M., Minor, A.M., Morris, J.W., Jr. and Stach, E.A. (2004). *Acta Mat.* **52**, 5783–5790.
- Sofronis, P. and Robertson, I.M. (2002). *Philos. Mag. A* **82**, 3405–3413.
- Song, M., Mitsuishi, K., Yasuda, H. and Furuya, K. (2002). *J. Electron Microsc.* **51** (Supplement), S211–214.
- Sow, C.-H., Harada, K., Tonomura, A., Crabtree, G. and Grier, D.G. (1998). *Phys. Rev. Lett.* **80**, 2693–2696.
- Spence, J.C.H., Lo, W. and Kuwabara, M. (1990). *Ultramicroscopy* **33**, 69–82.
- Spence, J.C.H., Kolar, H.R., Hembree, H.R., Humphreys, C.J., Barnard, J., Datta, R., Koch, C., Ross, F.M. and Justo, J.F. (2006). *Philosophical Magazine*, in press.
- Stach, E.A., Hull, R., Bean, J.C., Jones, K.S. and Nejim, A. (1998a). *Microsc. Microanal.* **4**, 294–307.
- Stach, E.A., Hull, R., Tromp, R.M., Reuter, M.C., Copel, M., LeGoues, F.K. and Bean, J.C. (1998b). *J. App. Phys.* **83**, 1931–1937.
- Stach, E.A., Hull, R., Tromp, R.M., Ross, F.M., Reuter, M.C. and Bean, J.C. (2000). *Philos. Mag. A* **80**, 2159–2200.
- Stach, E.A., Freeman, T., Minor, A.M., Owen, D.K., Cumings, J., Wall, M.A., Chraska, T., Hull, R., Morris, Jr., J.W., Zettl, A. and Dahmen, U. (2001). *Microsc. Microanal.* **7**, 507–517.
- Stach, E.A., Pauzauskie, P.J., Kuykendall, T., Goldberger, J., He, R. and Yang, P. (2003). *Nano Lett.* **3**, 867.
- Storaska, G.A. and Howe, J.M. (2004). *Mat. Sci. Eng. A* **368**, 183–190.
- Storaska, G.A., Moore, K.T. and Howe, J.M. (2004). *Philos. Mag.* **84**, 2619–2634.

- Suenaga, K., Colliex, C. and Iijima, S. (2001). *App. Phys. Lett.* **78**, 70–72.
- Sugi, H., Akimoto, T., Sutoh, K., Chaen, S., Oishi, N. and Suzuki, S. (1997). *Proc. Nat. Acad. Sci. USA* **94**, 4378–4382.
- Suh, Y.J., Prikhodko, S.V. and Friedlander, A.K. (2004). *Microsc. Microanal.* **8**, 497.
- Sun, H.P., Pan, X.Q., Haeni, J.H. and Schlom, D.G. (2004). *App. Phys. Lett.* **85**, 1967–1969.
- Sun, H.P., Chen, Y.B., Pan, X.Q., Chi, D.Z., Nath, R. and Foo, Y.L. (2005a). *App. Phys. Lett.* **86**, 71904–1–3.
- Sun, K., Wang, L.M., Ewing, R.C. and Weber, W.J. (2005b). *Philos. Mag.* **85**, 597–608.
- Takayanagi, K., Tanishiro, Y., Takahashi, S. and Takahashi, M. (1985). *Surface Sci.* **164**, 367–392.
- Takeda, S. and Yamasaki, J. (1999). *Phys. Rev. Lett.* **83**, 320–323.
- Takeguchi, M., Shimojo, M. and Furuya, K. (2005). *Nanotechnology* **16**, 1321–1325.
- Tamiya, T., Shindo, D., Murakami, Y., Bando, Y. and Otsuka, K. (1998). *Mat. Trans. JIM* **39**, 714–723.
- Tan, X., Xu, Z., Shang, J.K. and Han, P. (2000). *App. Phys. Lett.* **77**, 1529–1531.
- Tan, X., Du, T. and Shang, J.K. (2002). *App. Phys. Lett.* **80**, 3946–3948.
- Tan, X. and Shang, J.K. (2004a). *J. App. Phys.* **95**, 2805–2810.
- Tan, X. and Shang, J.K. (2004b). *J. App. Phys.* **96**, 635–639.
- Tan, X.L., He, H. and Shang, J.K. (2005). *J. Mat. Res.* **20**, 1641–1653.
- Tanaka, H., Konno, T.J., Sinclair, R. and Hirashita, N. (1995). *J. App. Phys.* **78**, 4982–4987.
- Tanaka, H., Hirashita, N. and Sinclair, R. (1996). *Jap. J. Appl. Phys. Part 2 (Letters)* **35**, L479–481.
- Tanaka, H., Hirashita, N. and Sinclair, R. (1998a). *Jap. J. Appl. Phys. Part 1*, **37**, 4284–4287.
- Tanaka, N., Kimata, H., Takashima, T. and Kizuka, T. (1998b). *Surface Rev. Lett.* **5**, 723–729.
- Tanaka, M., Furuya, K. and Saito, T. (1998c). *Microsc. Microanal.* **4**, 207–217.
- Tanaka, M., Han, M., Takeguchi, M. and Furuya, K. (2004). *Philos. Mag.* **84**, 2699–2709.
- Tanaka, M., Chu, F., Shimojo, M., Takeguchi, M., Mitsuishi, K. and Furuya, K. (2005). *App. Phys. Lett.* **86**, 183104.
- Tanner, L.E., Pelton, A.R., Van Tendeloo, G., Schryvers, D. and Wall, M.E. (1990). *Scripta Met. Mat.* **24**, 1731–1736.
- Tappin, D.K., Robertson, I.M. and Birnbaum, H.K. (1995). *Phys. Rev. B* **51**, 14854–14860.
- Teodorescu, V., Nistor, L., Bender, H., Steegen, A., Lauwers, A., Maex, K. and Van Landuyt, J. (2001). *J. App. Phys.* **90**, 167–174.
- Terrones, M., Terrones, H., Banhart, F., Charlier, J.-C. and Ajayan, P.M. (2000). *Sci.* **288**, 1226–1229.
- Terrones, M., Banhart, F., Grobert, N., Charlier, J.-C., Terrones, H. and Ajayan, P.M. (2002). *Phys. Rev. Lett.* **89**, 075505/1–4.
- Teter, D.F., Robertson, I.M. and Birnbaum, H.K. (2001). *Acta Mat.* **49**, 4313–4323.
- Thilly, L., Veron, M., Ludwig, O., Lecouturier, F., Peyrade, J.P. and Askenazy, S. (2002). *Philos. Mag. A* **82**, 925–942.
- Thompson, P., Gutfleisch, O., Chapman, J.N. and Harris, I.R. (1997). *J. Phys. D* **30**, 1854–1860.
- Togawa, Y., Harada, K., Akashi, T., Kasai, H., Matsuda, T., Nori, F., Maeda, A. and Tonomura, A. (2005). *Phys. Rev. Lett.* **95**, 87002.
- Tonomura, A. (2002). *J. Electron Microsc.* **51** (supplement), S3–S11.

- Tomomura, A. (2003). *J. Electron Microsc.* **52**, 11–19.
- Tomomura, A., Kasai, H., Kamimura, O., Matsuda, T., Harada, K., Nakayama, Y., Shimoyama, J., Kishio, K., Hanaguri, T., Kitazawa, K., Sasase, M. and Okayasu, S. (2001). *Nature* **412**, 620–622.
- Tomomura, A., Kasai, H., Kamimura, O., Matsuda, T., Harada, K., Yoshida, T., Akashi, T., Shimoyama, J., Kishio, K., Hanaguri, T., Kitazawa, K., Masui, T., Tajima, S., Koshizuka, N., Gammel, P.L., Bishop, D., Sasase, M. and Okayasu, S. (2002). *Phys. Rev. Lett.* **88**, 237001/1–4.
- Toporov, A. Yu., Langford, R.M. and Petford-Long, A.K. (2000). *App. Phys. Lett.* **77**, 3063–3065.
- Troiani, H.E., Camacho-Bragado, A., Armendariz, V., Gardea Torresday, J.L. and Yacaman, M.J. (2003). *Chem. Mat.* **15**, 1029–1031.
- Twitchett, A.C., Dunin-Borkowski, R.E. and Midgley, P.A. (2002). *Phys. Rev. Lett.* **88**, 238302/1–4.
- Twitchett, A.C., Dunin-Borkowski, R.E., Broom, R.F. and Midgley, P.A. (2004). *J. Phys. Cond. Matt.* **16**, S181–192.
- Ugarte, D. (1992). *Nature* **359**, 707–709.
- Uhlig, T., Heumann, M. and Zweck, J. (2003). *Ultramicroscopy* **94**, 193–196.
- Urita, K., Sato, Y., Suenaga, K., Gloter, A., Hashimoto, A., Ishida, M., Shimada, T., Shinohara, H. and Iijima, S. (2004). *Nano Lett.* **4**, 2451–2454.
- Urita, K., Suenaga, K., Sugai, T., Shinohara, H. and Iijima, S. (2005). *Phys. Rev. Lett.* **94**, 155502.
- Vanderschaeve, G., Levade, C. and Caillard, D. (2000). *J. Phys. Cond. Matt.* **12**, 10093–10103.
- Vanderschaeve, G., Levade, C. and Caillard, D. (2001). *J. Microsc.* **203**, 72–83.
- Vanhellemont, J., Romano-Rodriguez, A., Fedina, L., Van Landuyt, J. and Aseev, A. (1995). *Mat. Sci. Tech.* **11**, 1194–1202.
- Venables, J. (1981). *Ultramicroscopy* **7**, 81–98.
- Verhaeghe, B., Louchet, F., Doisneau-Cottignies, B., Brechet, Y. and Massoud, J.-P. (1997). *Philos. Mag. A* **76**, 1079–1091.
- Verheijen, M.A., Donkers, J.J.T.M., Thomassen, J.F.P., van den Broek, J.J., van der Rijt, R.A.F., Dona, M.J.J. and Smit, C.M. (2004). *Rev. Sci. Inst.* **75**, 426–429.
- Verwerft, M., Van Tendeloo, G., Van Landuyt, J. and Amelinckx, S. (1988). *Ferroelectrics* **88**, 27–36.
- Vetrano, J.S., Bruemmer, S.M. and Robertson, I.M. (1995). *Mat. Res. Soc. Sympos. Proc.* **404**, 177–182.
- Vetrano, J.S., Bruemmer, S.M., Pawlowski, L.M. and Robertson, I.M. (1997). *Mat. Sci. Eng. A* **238**, 101–107.
- Volkov, V.V. and Zhu, Y. (2000). *J. Magn. Magn. Mat.* **214**, 204–216.
- Volkov, V.V., Zhu, Y. and Malac, M. (2004). *Philos. Mag.* **84**, 2607–2617.
- Vook, R.W. (1994). *Mat. Chem. Phys.* **36**, 199–216.
- Vystavel, T., Palasantzas, G., Koch, S.A. and De Hosson, J.T.M. (2003a). *App. Phys. Lett.* **82**, 197–199.
- Vystavel, T., Palasantzas, G., Koch, S.A. and De Hosson, J.T.M. (2003b). *App. Phys. Lett.* **83**, 3909–3911.
- Wall, M.A. and Barbee, T.W., Jr. (1997). *Mat. Res. Soc. Sympos. Proc.* **480**, 245–250.
- Wall, M.A. and Dahmen, U. (1998). *Microsc. Res. Tech.* **42**, 248–254.
- Wall, M.A., Barbee, T.W. and Weihs, T.P. (1995). *Mat. Res. Soc. Sympos. Proc.* **404**, 43–48.
- Walsh, C.A. (1989). *Philos. Mag. A* **59**, 227–246.
- Wang, M.H. and Chen, L.J. (1992). *J. App. Phys.* **71**, 5918–5925.
- Wang, Y.G. and Petford-Long, A.K. (2002). *J. App. Phys.* **92**, 6699–6707.

- Wang, Z.L., Gao, R.P., Pan, Z.W. and Dai, Z.R. (2001). *Adv. Eng. Mat.* **3**, 657–661.
- Wang, Z.L., Dai, Z.R., Gao, R. and Gole, J.L. (2002a). *J. Electron Microsc.* **51** (Supplement), S79–85.
- Wang, Z.L., Gao, R.P., de Heer, W.A. and Poncharal, P. (2002b). *App. Phys. Lett.* **80**, 856–858.
- Wang, L. and Yang, J.C. (2005). *J. Mat. Res.* **20**, 1902–1909.
- Wang, M.S., Wang, J.Y., Chen, Q. and Peng, L.M. (2005). *Adv. Funct. Mat.* **15**, 1825–1831.
- Watanabe, S., Haishi, Y., Suda, T., Ohnuki, S., Takahashi, H. and Kiritani, M. (2002). *Radia. Eff. Def. Sol.* **157**, 101–108.
- Werner, M., Weber, E.R., Bartsch, M. and Messerschmidt, U. (1995). *Phys. Stat. Sol. A* **150**, 337–341.
- Williamson, M.J., Tromp, R.M., Vereecken, P.M., Hull, R. and Ross, F.M. (2003). *Nature Mat.* **2**, 532–536.
- Wise, M.E., Biskos, G., Martin, S.T., Russell, L.M. and Buseck, P.R. (2005). *Aerosol Sci. Tech. B*, 849–856.
- Wu, Y. and Yang, P. (2001). *J. Amer. Chem. Soc.* **123**, 3165.
- Xu, H.P., Li, Q., Feng, D., Shen, G.J., Guan, Q.C. and Wang, J.Y. (1993). *Ferro-electrics* **146**, 137–144.
- Xu, Z., Tan, X., Han, P. and Shang, J.K. (2000). *App. Phys. Lett.* **76**, 3732–3734.
- Xu, Z., Bai, X.D., Wang, E.G. and Wang, Z.L. (2005b). *App. Phys. Lett.* **87**, 163106.
- Xu, Z., Bai, X.D., Wang, E.G. and Wang, Z.L. (2005a). *J. Phys.—Cond. Matter.* **17**, L507–L512.
- Yamashita, J., Hirayama, H., Ohshima, Y. and Takayanagi, K. (1999). *App. Phys. Lett.* **74**, 2450–2452.
- Yang, J.C., Kolasa, A.B., Gibson, J.M. and Yeadon, M. (1998). *App. Phys. Lett.* **73**, 2841–2843.
- Yang, J.C., Evan, D. and Tropia, L. (2002). *App. Phys. Lett.* **81**, 241–243.
- Yasuda, H. and Mori, H. (1999). *J. Electron Microsc.* **48**, 581–584.
- Yasuda, H., Mori, H., Sakata, T., Naka, M. and Fujita, H. (1992). *Philos. Mag. B* **65**, 1001–1009.
- Yasuda, H., Mori, H. and Furuya, K. (2000). *Philos. Mag. Lett.* **80**, 181–186.
- Yasuda, H., Mitsuishi, K. and Mori, H. (2001). *Phys. Rev. B* **64**, 094101/1–6.
- Yasuda, H., Mori, H. and Lee, J.G. (2004). *Phys. Rev. B* **70**, 214105/1–6.
- Yeadon, M., Yang, J.C., Averbach, R.S., Bullard, J.W., Olynick, D.L. and Gibson, J.M. (1997). *App. Phys. Lett.* **71**, 1631–1633.
- Yi, G., Nicholson, W.A.P., Lim, C.K., Chapman, J.N., McVitie, S. and Wilkinson, C.D.W. (2004). *Ultramicroscopy* **99**, 65–72.
- Yokota, T., Howe, J.M., Jesser, W.A. and Murayama, M. (2004). *J. App. Phys.* **95**, 5756–5761.
- Yonenaga, I., Brown, P.D. and Humphreys, C.J. (1998). *Mat. Sci. Eng. A* **253**, 148–150.
- Yonenaga, I., Werner, M., Bartsch, M., Messerschmidt, U. and Weber, E.R. (1999). *Phys. Stat. Sol. A* **171**, 35–40.
- Yoo, J.H., Murakami, Y., Shindo, D. and Ohsuna, T. (2004a). *Philos. Mag.* **84**, 2597–2606.
- Yoo, J.H., Murakami, Y., Shindo, D., Atou, T. and Kikuchi, M. (2004b). *Phys. Rev. Lett.* **93**, 047204/1–4.
- Yoshida, K., Yamasaki, J. and Tanaka, N. (2004). *App. Phys. Lett.* **84**, 2542–2544.
- Yu, A.C.C., Lo, C.C.H., Petford-Long, A.K., Jiles, D.C. and Miyazaki, T. (2002a). *J. App. Phys.* **91**, 780–784.

- Yu, A.C.C., Petford-Long, A.K., O'Grady, K. and Miyazaki, T. (2002b). *J. App. Phys.* **91**, 5234–5239.
- Zerrouk, T.E.A., Hamichi, M., Pilkington, J.D.H. and Venables, J.A. (1994). *Phys. Rev. B* **50**, 8946–8949.
- Zghal, S., Coujou, A. and Couret, A. (2001). *Philos. Mag. A* **81**, 345.
- Zhang, L.H., Johnson, E. and Dahmen, U. (2004). *Mat. Res. Soc. Sympos. Proc.* **821**, 181–186.
- Zhang, M., Olson, E.A., Twesten, R.D., Wen, J.G., Allen, L.H., Robertson, I.M. and Petrov, I. (2005). *J. Mat. Res.* **20**, 1802–1807.
- Zhou, G. and Yang, J.C. (2002). *Phys. Rev. Lett.* **89**, 106101/1–4.
- Zhou, G. and Yang, J.C. (2003). *Surface Sci.* **531**, 359–367.
- Zhou, G., Yang, J.C., Xu, F., Barnard, J.A. and Zhang, Z. (2002). *Mat. Res. Soc. Sympos. Proc.* **737**, F6.3.
- Zhu, Y. and De Graef, M., Eds. (2001). *Magnetic Imaging and Its Applications to Materials*. (Academic Press, San Diego).
- Zielinski, W., Swiatnicki, W., Barstch, M. and Messerschmidt, U. (2003). *Mat. Chem. Phys.* **81**, 476–479.
- Zimmermann, C.G., Yeadon, M., Nordlund, K., Gibson, J.M., Averbach, R.S., Herr, U. and Samwer, K. (1999). *Phys. Rev. Lett.* **83**, 1163–1166.

Cryoelectron Tomography (CET)

Juergen M. Plitzko and Wolfgang Baumeister

Clearly, if recent improvements in electron microscopes are to be fully utilized in biology, contrast must be enhanced without drastic molecular alteration of the specimen or obscuration by extraneous material. . . . Thus the *polytropic montage* seems to offer a means of determining the three-dimensional structures of low-contrast biological specimens at a resolution of 3 Å, or the best resolution attainable with existing electron microscopes. (Hart, 1968)

1 Introduction

More than 70 years have now passed since the invention of the first transmission electron microscope (Knoll and Ruska, 1932; Ruska, 1987). Many reports and publications have been written covering the technological achievements in electron microscopy (EM) and more important the scientific breakthroughs related to the information disclosed by EM studies. EM, in its various “flavors” (see below), is now a well-established method in life as well as in material science. Therefore, almost every laboratory in the field of structural analysis is equipped with one or several low or intermediate voltage microscopes for routine use and in some instances for high-end applications. However, compared to light microscopy, EM is still a “young” method but with great potential for further improvements. This has been shown, especially in the past decade: among the highlights are the incorporation of energy filters, monochromators, aberration correctors, and above all the almost complete replacement of negative plates by charge coupled device (CCD) cameras and imaging plates. By all means, these technical achievements have been supported by the development of the computer and its availability throughout science. The various techniques in EM, such as bright field (BF) or dark field (DF) imaging, weak beam imaging, conventional transmission EM (CTEM), high-resolution TEM or EM (HRTEM or HREM), scanning transmission electron microscopy (STEM), energy-filtered TEM (EFTEM), and many others, have been improved and, moreover, expanded with the computer power on hand. Today’s most modern techniques are all based on automated acquisition

and alignment procedures, reconstruction algorithms, and especially on elaborate image processing routines.

Here, we address the advances made in biological EM and especially in cryo-EM. Biological structures can be, by and large, characterized as pleiomorphic. Just as every human being has a different face, cells, proteins, and macromolecular complexes have different shapes and forms, designed for a higher functional purpose. It is far from random and instead of addressing them as "amorphic" or "amorphous," as in physics for randomly ordered solids, they are called "pleiomorphic" or "pleiomorph." Moreover, the inside of a cellular structure resembles the image of a giant factory, where the single constituents act together, building highly specific molecular machines and, if necessary, change their purpose (Alberts, 1998). Therefore it is a highly variable and dynamic environment. Every intrusion into this fragile system can lead to changes. The suitable preparation for a final characterization with the EM is a major challenge. And it is likewise a challenge in terms of the environment inside EMs: an ultrahigh vacuum and electron radiation.

Despite the fact that biologists were impressed by the resolving power of the EM, they remained very sceptical about the usefulness of the EM in structural biology. The major drawback was and still is the sensitivity to radiation of the biological samples. After a few minutes exposure to the electron beam the biological substance in question is literally "incinerated." However, the present knowledge about the ultrastructure of cells, viruses, and other biological substances was accumulated by EM investigations with, at that time, suitable preparation techniques. Higher vacuum resistance was achieved by dehydration and water-substitution methods, beam resistance was increased by staining the samples with heavy metals (Brenner and Horne, 1959), and, in addition, enhanced contrast and transparency were obtained by sectioning, e.g., "big" cells in slices about 200 nm thick, with ultramicrotomes (Porter and Blum, 1953). These preparation techniques enabled biologists to establish the basis of a common image of the cell's interior. Nevertheless, this "image" remained incomplete and was still at a resolution far above that potentially available with the EM. Additionally the samples were altered by these preparation methods, thus complicating and limiting image interpretation. In some instances the artifacts introduced even misled scientists. Although these preparation techniques are still used in various ways, but there was an obvious need for far more reliable and less harmful procedures. In 1981 Dubochet and McDowell introduced a new means of sample preservation for EM investigations: the cryotechnique. Without doubt, cryopreparation is one of if not the greatest development in biological EM. Instead of replacing the water or dehydrating the whole system, the biological substance is embedded within its original buffer solution or simply water by rapidly freezing at very low temperatures, namely the temperature of liquid nitrogen (~ 90 K). This way the water was transferred into an amorphous state, inhibiting crystallization, and thus disruption of the cell due to the volume increase of crystalline water. This plunge-freezing technique revolutionized the field of structural

biology, because for the first time it was possible to investigate biological samples in their native state, without introducing any artifacts, corroborating image interpretation.

Excluding artifacts was clearly one achievement in obtaining higher resolved images of biological structures, but since these structures are pleiomorphic, single two-dimensional (2D) images are clearly insufficient for a complete structural characterization, which can only be done three-dimensionally (3D). Additionally, due to the large depth of focus of the EM, images represent only 2D projections of the specimen, in which almost all information about the third dimension is lost. As mentioned, sectioning can help, and especially serial sectioning, where the structure of interest, e.g., big cells, is serially sliced. However, serial sectioning was previously done only on epoxy resin-embedded samples, limited in resolution to the section slice thickness (~50 nm) and having lost any information about the molecular organization. But it was evident that information retrieval down to the supramolecular level for all three dimensions is absolutely crucial for an in-depth structural analysis of native biological samples.

In the late 1960s and early 1970s three groups independently experimented with the possibility of 3D microscopy, namely electron tomography (Hart, 1968; DeRosier and Klug, 1968; Hoppe et al., 1968; Hoppe, 1974). Perhaps inspired by the tomographic methods invented for medical examinations (Cormack, 1963, 1980; Hounsfield, 1972) they developed acquisition schemes to access the 3D information by recording images from different orientations or from differently oriented particles. DeRosier and Klug (1968) studied the helical tails of T4-bacteriophages and with a single EM they were able to present a complete 3D reconstruction. By selecting an object with the *a priori* knowledge of its helical symmetry they avoided a complicated data acquisition procedure. In the same year Hart (1968) reported very precisely on his idea: the “polytropic montage.” He investigated the rod-shaped tobacco mosaic virus (TMV). Hart and later Hoppe have chosen a more complicated and cumbersome acquisition process: the manual rotation of the object perpendicular to the incoming electron beam and the acquisition of single projections from different viewing angles. They reported on the 3D representation from a series of images acquired over a large angular regime. In the late 1960s, when the average electron microscopist used photographic plates, these procedures were inevitably very cumbersome and time consuming and of course reconstruction calculations were possible only with large supercomputers. Nevertheless, with the introduction of CCD cameras for image recording in EM in the late 1980s (Mochel and Mochel, 1986) and the introduction of computer-controlled EMs attempts (Figure 7–1) were made to improve the pioneering work with the prospect of implementing it for routine and fast use (Typke et al., 1991; Koster et al., 1992, 1997). With the addition of suitable preparation techniques such as cryofixation of biological specimens, cryo-EM promised to be the method of choice for further structural work in molecular biology. Nevertheless, it was not until the mid-1990s that the first reports on applications on automated cryo-electron tomography appeared (Horowitz et al., 1994;

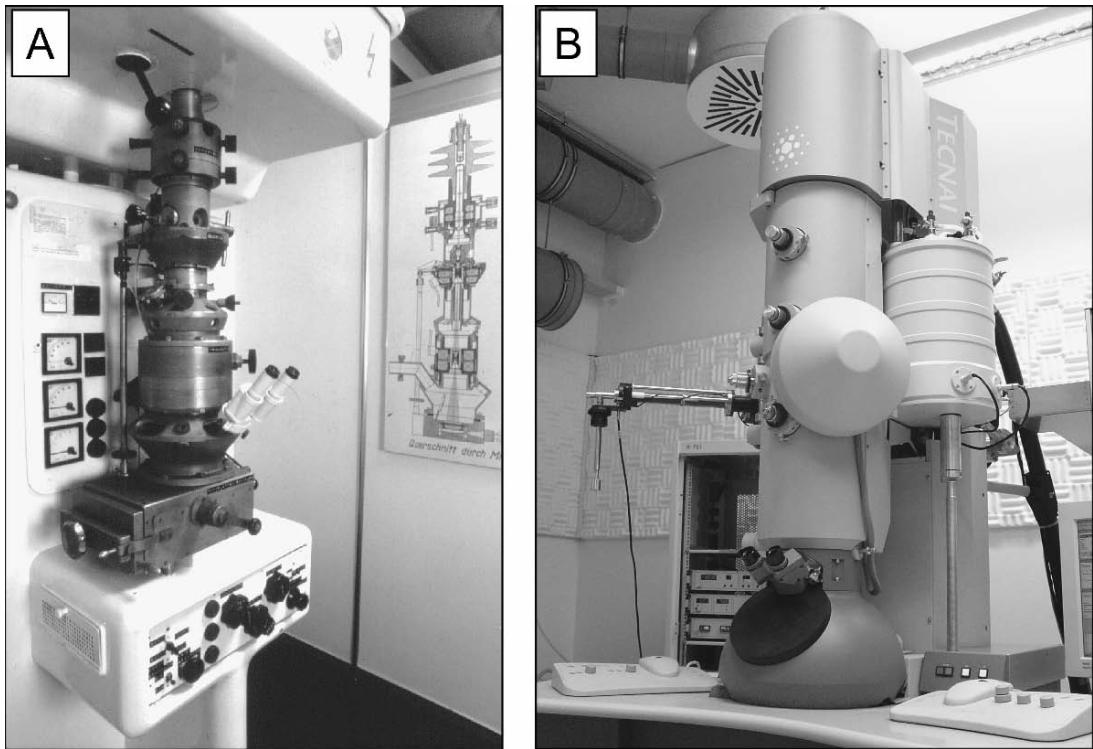


Figure 7-1. Transmission electron microscopes (TEM) in the past and today. a) 1939; First commercial TEM ('Siemens Super Microscope') from Siemens & Halske Ltd., Berlin, Germany. b) 2002; Tecnai F30 Helium ('Polara') from FEI company, Eindhoven, The Netherlands as installed at the MPI of Biochemistry in Martinsried (near Munich), Germany. A computer controlled TEM dedicated for cryo-electron microscopy, which allows cooling to liquid nitrogen (90 K) and helium temperature (~10 K).

Dierksen et al., 1995; Grimm et al., 1996a; Walz et al., 1997a; Nitsch et al., 1998), now abbreviated cryo-ET or, for the extended community, three-dimensional electron microscopy (3DEM).

Today, cryo-EM is established in the field of structural biology and three different techniques are building the base for 3D characterization: electron crystallography (cryo-EC), single-particle EM (cryo-EM), and electron tomography (cryo-ET). In the following we will describe and explain one of these tomographic approaches, namely cryo-ET (Frank, 1992, 1996), in detail.

2 Three-Dimensional Cryoelectron Microscopy

2.1 Principles of 3D Imaging

The main contrast mechanism in cryo-EM is phase contrast. In first-order approximation, the micrograph is an interference pattern of the primary beam and the focused scattered beam. In this so-called weak phase approximation, the electrostatic potential of the sample is the quantity that is imaged. To create a phase delay between the primary and scattered beams, the objective lens has to be defocused. The

imaging process of weakly scattering objects in a TEM can be summarized elegantly by introducing a contrast transfer function (CTF) in Fourier space (Reimer, 1993). In this notation, the image I_{xy} of an 3D object with an electrostatic potential V_{xyz} can be written as

$$I_{xy} = \mathcal{F}^{-1} \left[\text{CTF}_{k_x, k_y} \cdot \mathcal{F} \left(\int dz V_{xyz} \right) \right].$$

Here, $\mathcal{F}$ denotes the Fourier transform and k_x and k_y denote the reciprocal vectors. The CTF is an oscillating function of the defocus value Δz , which is damped by an envelope function. The envelope function finally limits the resolving power of a TEM. To achieve high-resolution data in biological EM, the use of a field emission gun (FEG) instrument with a high coherence and brightness is mandatory (Baumeister and Typke, 1993), since the increased temporal and spatial coherence reduces the damping of the envelope function (Frank, 1973). The oscillatory behavior of the CTF leads to contrast inversion of I_{xy} for certain frequency bands. Provided the signal-to-noise ratio (SNR) of I_{xy} is sufficient, I_{xy} can be corrected for these CTF effects, but CTF correction will not be feasible close to the zeros of the CTF where hardly any signal is transferred. In single particle analysis or cryoelectron crystallography, images of different focus levels are recorded to avoid that problem; the data of different focus levels can then be combined in a so-called focus series (Schiske, 1968, 1973; Typke, 1992), which ensures data coverage in Fourier space without gaps around CTF zeros. Through-focus series reconstruction is particularly suitable in material science studies where high magnifications combined with a high dose can be utilized almost without restrictions to determine the atomic positions even for very light elements in crystalline specimen (Thust et al., 1994, 1996; Coene et al., 1996; Kisielowski et al., 2001). However, in low-dose EM, the CTF is often not easily accessible due to the low SNR, in particular if a CCD camera is used as a detecting device. Therefore, the first zero crossing of the CTF often limits the attainable resolution in a cryoelectron micrograph, for example, in cryo-ET.

Since an EM image is essentially a projection, features from different z -levels within the object overlap in the resulting micrograph and cannot be separated. The traditional way to extract meaningful information from an electron micrograph is to reduce the z -dimension of the object, i.e., to image almost 2D objects. In biological imaging of cells this was accomplished by sectioning the previously fixed object using an ultramicrotome (Porter and Blum, 1953). By means of serial sectioning even 3D data could be obtained by imaging successive sections belonging to the same cell. In this case, the achievable resolution in z depends on the thickness of the sections; the best resolution attained was in the range of 50 nm. However, the combination of serial sectioning with tomographic data acquisition (see below) offers the possibility of improving the resolution (Soto et al., 1994; Mueller-Reichert et al., 2003).

A different approach is to combine different views of the specimen to derive 3D information (Figure 7–2). The cryo-EM methods for obtaining 3D data can be summarized as tomographic or at least quasitomographic. They all rely on the fact that the parallel projection of a 3D object corresponds to a slice in the 3D Fourier space of the object. This

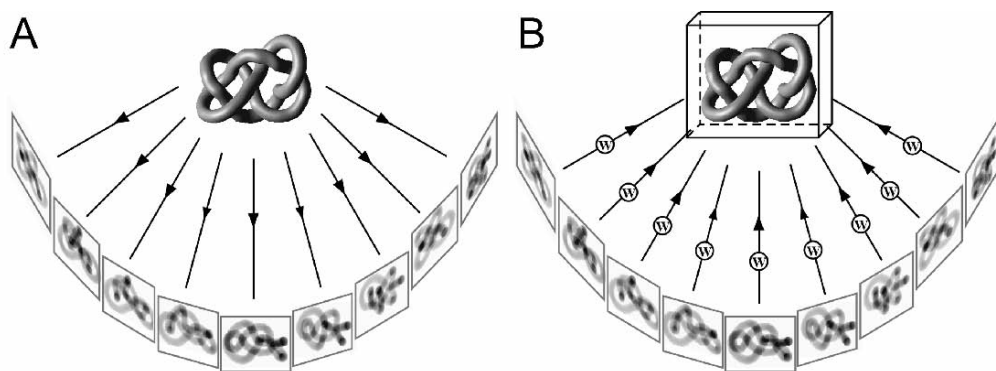


Figure 7-2. The Figure 7-shows a single axis tilt series data acquisition scheme. The object is represented by a pleiomorphic object (in this case a knot) to emphasize the fact that electron tomography can retrieve 3D information from non-repetitive structures. a) A set of 2D projection images is recorded while the specimen is tilted incrementally around an axis perpendicular to the electron beam, and projection images of the same area are recorded on a CCD camera at each tilt angle. Tilt range is typically $\pm 70^\circ$ with tilt increments between $1.5\text{--}3^\circ$. b) The backprojection method explains the principle of 3D reconstruction in a fairly intuitive manner. For each weighted projection, a backprojection body is calculated, and the sum of all projection bodies represents the density distribution of the original object—the tomogram.

“projection-slice theorem” and the possibility of restoring the 3D data using their projections were first discovered by Radon (1917). An English translation of this paper can be found in Deans (1983). To obtain the 3D information, different slices of the Fourier space, i.e., projections in real space, have to be gathered to sample the entire information. Different orientations of the sample can be realized by changing the orientation of the specimen. For this purpose the specimen has to be rotated, as is the case in electron crystallography or ET, or identical copies of a specimen that occur in different orientations can be reconstructed in a 3D model as in “single particle” analysis. In the field of EM, those tomographic methods for 3D structure determination were proposed in the late 1960s (Hart, 1968; Hoppe et al., 1968) and realized in the pioneering work of DeRosier and Klug (1968).

2.2 The Single-Particle Approach

One powerful advantage of EM is the fact that it produces projection images, unlike X-ray crystallography, which yields diffraction data with no information on the position; this is known as the “phase problem.” Thus, images may be treated in real space as well as in Fourier space and there is no difference between those that contain information about a single molecule without any internal symmetry and those generated by a 2D crystalline array. This fact is exploited in the so-called single-particle approach, which enables the electron microscopist to determine 3D density maps of individual macromolecules (Frank, 1996). Depending on a variety of conditions the 3D models show near-atomic, or, in the more characteristic case, intermediate resolution ($\geq 7\text{ \AA}$) (Jiang et al., 2003a; Spahn et al., 2001).

Projection images from macromolecules contain information about the 3D structure as 2D density profiles, characteristic of the particular orientation of the particles in the electron beam (Figure 7–3). To disentangle molecular substructures in the third dimension a number of different images of known projection geometry have to be combined. Protein complexes not too large in size are usually oriented arbitrarily in vitrified ice, so that even one EM image contains many different projections of the isolated protein species (Figure 7–3A). Once the relative orientations of the particles are known, i.e., the three Eulerian angles (φ , Ψ , θ) defining the rotation of a particle around the axes in 3D space, their projections can be placed on the surface of a sphere centered on the origin of a common coordinate system. The Eulerian angles exactly define the positions of the projections that can now be backprojected centrally in such a way as to superimpose the 2D densities into a common 3D volume of the particle. The most popular and utilized reconstruction algorithm in real space is the so-called weighted backprojection (Radermacher, 1992; see Section 3.5.2).

Very large macromolecular complexes, filamentous aggregates, or proteins embedded in biological membranes usually occur in preferred orientations in frozen samples. Here, and in cases where nonredundant structures such as singular protein assemblies are to be investigated, the specimen is tilted around a fixed axis at a certain angular incre-

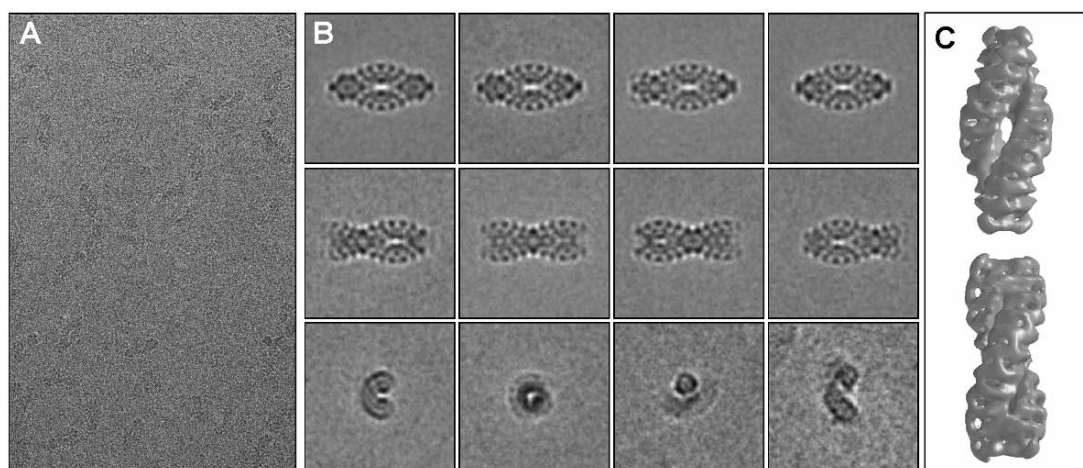


Figure 7–3. Single particle investigation of the giant protein complex TPP II from *Drosophila melanogaster* embedded in vitrified ice. In eukaryotes, tripeptidyl peptidase II (TPPII) is a crucial component of the protein degradation pathway. The 150-kDa subunits of *Drosophila* TPPII assemble into a giant proteolytic complex of 6MDa with a remarkable architecture consisting of two segmented and twisted strands that form a spindle-shaped structure (length 56nm, width 24nm). a) Cryo-electron micrograph of isolated TPP II complexes illustrating the very weak image contrast and the high level of noise. b) Averaging and classification of a large number of equivalent projections of separate molecules. Once a large set of views is available, a preliminary 3D reconstruction can be computed and refined iteratively. c) The 3D model obtained by cryo-electron microscopy, reveals details of the molecular architecture and, in conjunction with biochemical data, provides insight into the assembly mechanism. The building blocks of this complex are apparently dimers, within which the 150kDa monomers are oriented head to head. Stacking of these dimers leads to the formation of twisted single strands, two of which comprise the fully assembled spindle (Rockel et al., 2002 and 2005). (See color plate.)

ment in the EM to obtain a series of projections. This approach of electron tomography is applicable almost universally if the specimen is not too thick for the electrons to penetrate.

Since the electron exposure must be restricted to a value low enough not to destroy the sensitive biological material, the images contain a high level of so-called "shot noise" (statistical variation in the number of electrons detected at each pixel in the image). Additionally electron scattering from elements low in atomic number (Z), like the constituents of most biological materials, e.g., carbon, oxygen, and hydrogen, produces very weak image contrast. Based on these facts biological structures are barely visible in single cryo-EM images. Furthermore the exact alignment of the particles becomes a challenge, which is tackled by averaging a large number of equivalent projections of separate molecules, thus reducing the noise level. For this purpose, the images of some 10^3 to 10^5 particles are sorted into distinct classes of views by multivariate statistical analysis (Van Heel and Frank, 1981). Once a large set of views is available, a preliminary 3D reconstruction can be computed and refined iteratively. Typically, molecular complexes must be large in size (>200 kDa; $1 \text{ Da} = 1 \text{ U} = 1.66 \times 10^{-24} \text{ g}$) to provide sufficient signal for the alignment at high resolution. If the particles possess a high internal symmetry, e.g., virus capsids or regular structures like 2D crystals, the averaging and reconstruction process is simplified and the number of images necessary and the computation time are reduced.

To determine the structural basis of the working mechanism of molecular machines we have to study the structures, positions, and interactions of their building blocks, i.e., their distinct subunits. Since the structure of protein-protein contacts is resolved only at near-atomic resolution it is hardly possible to identify molecular borders in 3D models of cryo-EM data by visual inspection. But the ability to identify elements of secondary structure, and α -helices in particular, at a resolution of about $6\text{--}8 \text{ \AA}$, makes it possible to fit a previously determined (atomic) model of protein subunits or subcomplexes into 3D density maps of larger assemblies obtained by cryo-EM (Li et al., 2002; Bottcher et al., 1997). This approach, called "docking," enables us to unravel the organization of macromolecular machines that cannot be structurally characterized otherwise. Right now researchers continue to develop quantitative criteria that can guide this operation (Volkman and Hanein, 1999).

Single-particle cryo-EM is a powerful but still slow method compared to X-ray crystallography, provided that for the latter well-diffracting crystals are available. The data collection and processing can extend over several months or longer even for specialists in the field. However, with the advent of computer-controlled electron microscopes data acquisition and analysis have become subject to the development of elaborate automation procedures that promise to reduce the time needed to create high-resolution 3D density maps of large macromolecular complexes considerably.

The real bottleneck for investigations of known or still unknown molecular machines arose from another experimental task. Purification procedures commonly used in biochemistry tend to select stable

and abundant protein complexes only. Very large molecular machines, rare or transient macromolecular assemblies, complexes consisting of membrane spanning and soluble components, and all those held together by forces too weak to withstand the purification procedures, escape isolation or even detection. The lesson is twofold: We need new approaches for isolating or reconstituting fragile assemblies, and cryo-EM of isolated single particles will probably not give us the whole truth. To address the full complexity of functional macromolecular assemblies noninvasive 3D imaging techniques have to take over to fulfill the task of studying the supramolecular architecture in its native, cellular context (Plitzko et al., 2002). The technique of cryo-ET is a “nouvelle route” with a great potential in the field of structural biology that promises to provide noninvasive and deep insight into the functional organization of the cellular proteome (Sali et al., 2003).

2.3 Cryoelectron Tomography

CET is unique in the sense that it can image nonpurified macromolecules three-dimensionally. However, in terms of resolution, cryo-EC and single-particle analysis are currently superior, which makes cryo-ET a complementary method to the aforementioned. In contrast to these methods, cryo-ET does not involve implicit or explicit averaging of the specimen, which makes it a suitable method for imaging of pleiomorphic objects. Therefore it is suitable to image entire cells (Grimm et al., 1998; Medalia et al., 2002; Kuerner et al., 2005; Nicastro et al., 2000, 2005; Scheffel et al., 2005) or nonsymmetric viruses (Gruenewald et al., 2003) *in situ* (Figure 7–4). Furthermore it has the

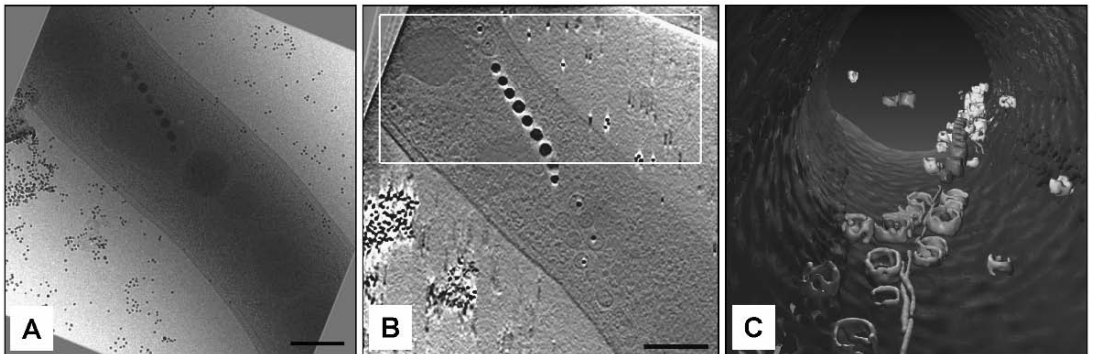


Figure 7–4. Cellular cryo-electron tomography of the magnetotactic microorganism *Magnetospirillum griphiswaldense*. The entire bacterium is oriented like a compass needle inside the magnetic field in its search for optimal living conditions. The miniature cellular compass is made by a chain of single nano-magnets, called magnetosomes (the scale bar represents 200nm). a) The two-dimensional image represents one projection (at 0°) from an angular tilt-series. b) *x–y* slices along the *z* axis through a typical three-dimensional reconstruction (tomogram). c) Surface-rendered representation of the inside of the cell showing the membrane (blue), vesicles (yellow), magnetite crystals (red) and a filamentous structure (green). Until now, it was not clear how the cells organise magnetosomes into a stable chain, against their physical tendency to collapse by magnetic attraction. However, the biochemical analysis revealed a protein responsible for the chain formation and the 3D investigation a cytoskeletal structure, which aligns the magnetosomes like pearls on a string (Scheffel et al., 2005). (See color plate.)

potential to unravel not only the structure but also the molecular interactions of the various macromolecules inside a cell (Beck et al., 2004). Whereas tomography of plastic embedded specimens (Marsh et al., 2001; Murk et al., 2003) reveals primarily data on the ultrastructure and morphology of a cell cryo-ET aims to resolve the cell at a molecular level. Although the SNR of a tomogram of a frozen hydrated specimen is generally worse than that of a plastic embedded specimen, more biologically significant information is contained in the high spatial frequencies since the preparation technique does not limit the interpretable information.

A cryo-electron tomogram essentially depicts the entire proteome of a cell. Therefore, cryo-ET can make significant contributions to the larger enterprise of structural proteomics (Sali et al., 2003). It has the potential to unravel the interactions of the various proteins that reside not only in the cytoplasm but also in the cell membrane. Moreover, it can provide medium resolution structural characterization of complexes that are largely undescribed to date. A prerequisite for those contributions is sufficient resolution to recognize macromolecules in a cellular tomogram.

The first published cryo-electron tomogram of a eukaryotic cell (Figure 7-5), the slime mould *Dictyostelium discoideum*, demonstrated that cryo-ET is able to resolve individual macromolecular complexes such as ribosomes or the 26S proteasome in the context of a cell (Medalia et al., 2002). Currently, cellular cryo-ET is effectively confined to relatively thin samples. Cells of a thickness beyond 1 μm are not transparent to the electron beam at a voltage of 300 kV. These cells are typically prokaryotic cells or unusually thin eukaryotic cells. Electron tomography of frozen hydrated sections of cells (Hsieh et al., 2002; Leis et al., 2005) can possibly extend the specimen class to thicker cells and even tissue cells. Unfortunately, the sectioning process still introduces severe artifacts; in particular, the forces that arise during the sectioning process tend to compress the cells strongly. The use of vibrating cryoknives can possibly offer some remedy from these shortcomings (Al-Amoudi et al., 2003). The technique is promising, but it is far from being a routine preparation technique, as yet.

To date, cryo-ET is effectively the only 3D imaging technique that can image cells or organelles in a close-to-native state at molecular resolution. It is of vital importance to have an imaging method that can image biological macromolecules since many of the complexes present in a cell tend to be fragile or transient. The detection of statistically significant distribution patterns can provide unprecedented information about the interaction of macromolecules, in transient complexes, for example. The basic problem that cryo-ET has to face is to derive statistically significant information from the initially very noisy tomograms. Interpretation of the raw data can be performed only by low-pass filtering the data to a resolution that is regarded as significant. Application of so-called "denoising" algorithms can extend this limit gradually and greatly improve the visualization of tomograms. These techniques mainly perform nonlinear filtering of the data (Frangakis

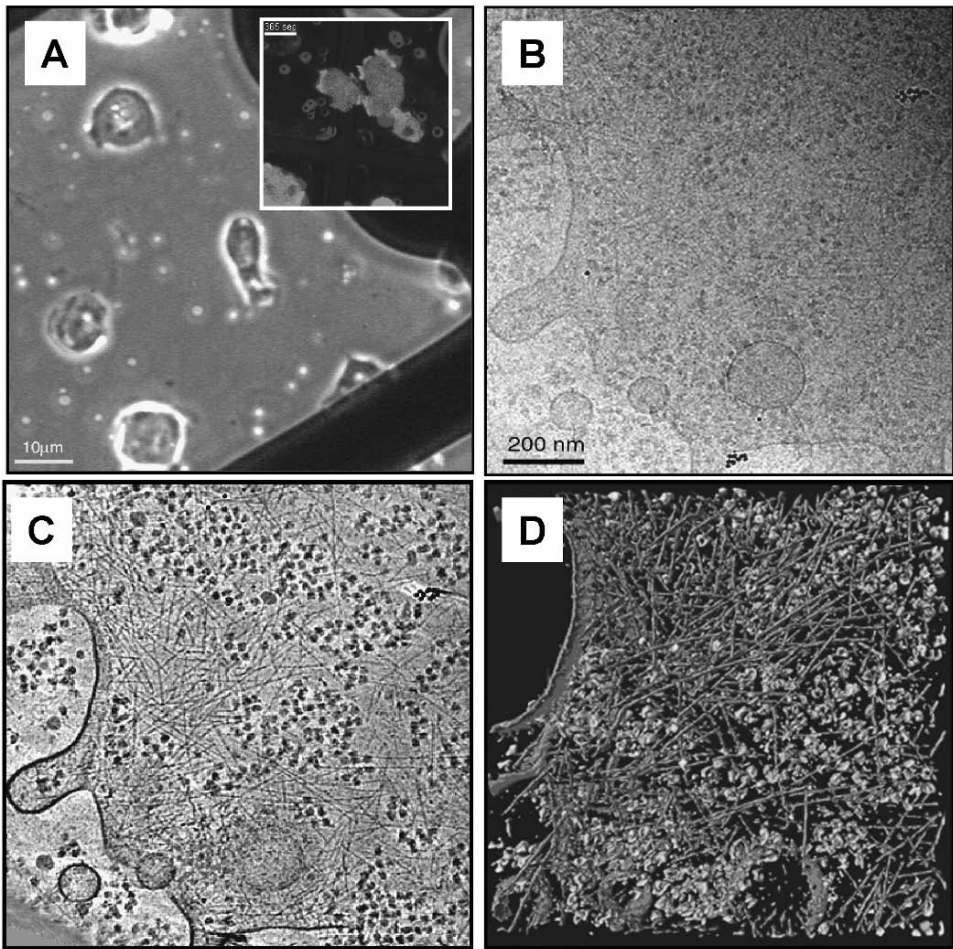


Figure 7-5. First electron tomographic investigation of a eukaryotic cell; the slime mould *Dictyostelium discoideum* embedded in vitrified ice. a) Phase contrast image and corresponding fluorescence image (inset) of cells on TEM grids. b) Cryo-electron micrograph at 0° tilt (conventional 2D projection) of a ~200nm thin peripheral region of the cell. c) Tomographic reconstruction from a complete tilt-series (120 images) and d) visualization by segmentation. Large macromolecular complexes, e.g. Ribosomes are shown in a green color, the actin filament network in orange-red and the cells' membrane in blue. Cryo-tomograms of *Dictyostelium discoideum* cells grown directly on carbon support films have provided unprecedented insights into the organization of actin filaments in an unperturbed cellular environment. The tomograms show, on the level of individual filaments, their modes of interaction (isotropic networks, bundles, etc.), they allow us to determine the branching angles precisely (in 3-D), and they reveal the structure of membrane attachment sites (Medalia et al., 2002). (See color plate.)

and Hegerl, 2001; Jiang et al., 2003b). The main purpose of filtering/“denoising” is to improve the visualization of the raw data and to facilitate interpretation primarily at the level of the ultrastructure. However, interpretation at the molecular level requires more quantitative methods.

A prerequisite for understanding the interaction of proteins is the localization of macromolecules in the cell. Practical ways to achieve this goal are labeling of macromolecules by specific and electron-dense markers (Koster and Klumperman, 2003) or recognition of individual particles by their known structural signature. The former is very hard to realize under *in vivo* conditions, whereas the latter requires sufficiently high resolution. Moreover, labeling techniques can be used for only a few macromolecules simultaneously since the different labels need to be distinguishable and specific. In the context of structural proteomics this makes recognition by means of the structural signature favorable since simultaneous recognition of molecules under at least close-to-native conditions is mandatory.

It has been shown previously that the currently achieved resolution of cryo-ET should suffice to identify large protein complexes in the range of 1MDa (Boehm et al., 2000). Moreover, detection of proteasomes (Lowe et al., 1995; Zwickl, 2000) and thermosomes (Klumpp et al. 1997; Klumpp and Baumeister, 1998) in phantom cells, which mimic real cells in size and composition with the advantage of being biochemically well defined, showed that this approach is also feasible experimentally (Frangakis et al., 2002). The approach used in these works requires structural knowledge from other sources; matched filtering of the tomograms with a structural template yields the locations and orientations of these complexes in the cell (see Section 3.5.4). Therefore, cryo-ET requires structural data from other techniques such as X-ray crystallography or single-particle analysis to obtain molecular information. The principal dose limitation of cryo-ET does not permit high-resolution structure determination on the basis of the tomogram alone. The dose fractionation theorem (Hegerl and Hoppe, 1976; McEwen et al., 1995) states that tomography images each volume element with the same statistical significance as a 2D projection acquired with the same overall dose. This makes tomographic data highly advantageous for location of features since the information is distributed over three dimensions. However, the statistical significance of each volume element is generally not sufficient to resolve macromolecules at the level of tertiary or quaternary structural elements initially. However, the combination of cryo-ET and single-particle averaging can offer a remedy for this; by averaging identical copies of complexes, they can potentially be resolved to at least a few nanometers. First realizations of the combination of cryo-ET and single-particle averaging have been reported for isolated proteins and yielded structures of 800-kDa protein complexes to ~2nm resolution (Nitsch et al., 1998); averaging of complexes directly from tomograms of organelles or viruses is a novel approach (Figure 7–6). First 2D results on the envelope complexes of dried and stained retroviruses indicated the potential of this approach (Zhu et al., 2003). Recent 3D results on the structure of the nuclear pore complex, a membrane complex that is difficult to assess by other techniques, taken from tomograms of entire nuclei, showed that this approach can yield structural characterization of unprecedented resolution (Beck et al., 2004, 2007).

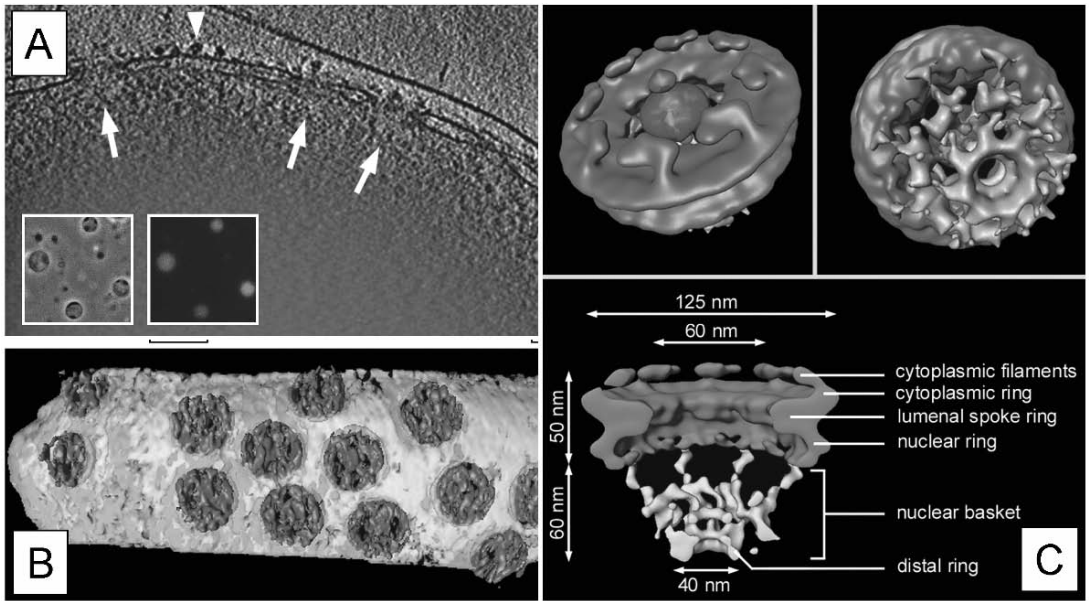


Figure 7-6. CET in combination with the single particle approach of transport-competent *Dictyostelium discoideum* nuclei. a) Three-dimensional reconstruction of the peripheral rim of an intact nucleus. X-y slice of 10 nm thickness along the z axis through a typical tomogram. Side views of nuclear pore complexes (NPCs) are indicated by arrows. Ribosomes connected to the outer nuclear membrane are visible (arrowheads). Inset displays a phase-contrast image and the corresponding fluorescence image. b) Surface rendered representation of a segment of nuclear envelope (NPCs in blue, membranes in yellow). c) Structure of the *Dictyostelium* NPC after classification and averaging of subtomograms. Cytoplasmic face of the NPC (upper left); the cytoplasmic filaments are arranged around the central channel. Nuclear face of the NPC (upper right); the distal ring of the basket is connected to the nuclear ring by the nuclear filaments. Cross sectional view of the NPC (bottom). The dimensions of the main features are indicated. All views are surface-rendered (nuclear basket in brown; Beck et al. 2004). (See color plate.)

3 Major Difficulties in Cryoelectron Tomography

Despite the fact that the acquisition process for ET sounds very simple and straightforward, the acquisition of a tomographic tilt series, especially from biological samples is, even today, a major technical challenge. The explanation will be given in the following sections addressing the five major obstacles (“core problems”) of ET: (1) sample preparation and preservation, (2) radiation sensitivity, (3) tilting geometry, (4) instrumentation (EM, energy filters, CCD detectors, etc.) and automation, and (5) alignment, reconstruction, and visualization.

3.1 Sample Preparation

3.1.1 Basics

Biological objects such as cells consist mainly of water (more than 70%). Built up from carbon-based molecules, they form the essential framework of life but for a closer structural investigation with the EM they have to withstand a constant “bombardment” of fast electrons under ultrahigh vacuum conditions: clearly not the best starting points

and clearly conditions that are far from natural and extremely inhospitable.

Indeed, the central goal in any EM study is the observation of structures down to the atomic level. Unlike material science studies, where the samples remain steady, biological specimens tend to be more “resistant” to an investigation with the EM. First they are exposed to ionizing radiation, namely electrons with speeds usually above 200,000 km/s, according to the acceleration voltage used. Second, the ultrahigh vacuum generates an ambient pressure far below the atmospheric pressure, and thus literally “sucks” out every shred of liquid, subsequently resulting in the implosion of the biological structure. Third, they are build up from carbon-based compounds plus elements with very low atomic numbers like hydrogen, oxygen, and nitrogen, and minute amounts of other low Z-elements, all very weak scatterers, in regard to their interaction with electrons, resulting in low-contrast images. Fourth, if we regard cellular structures of higher organisms, such as mammalian cells, they can easily reach sizes of tenth of micrometers in all three dimensions, and are therefore almost or effectively impenetrable to the electron beam. Fifth, the secrets of these structures are hidden in a highly crowded environment, the cytoplasm, where functional units like proteins, protein complexes, or even molecular machines are “densely packed,” literally touching each other. This “molecular crowding” (Ellis, 2001; Ellis and Minton, 2003) makes it difficult to separate them for structural characterization. Sixth and last, every cell is different, just as every human has a different face. Thus, single 2D projections will not provide a complete 3D characterization, and averaging techniques as in the so-called single-particle approach are therefore impractical. All six of these facts have to be addressed in any life science study and especially when investigating biological structures in a close-to-life state, quasi *in vivo*.

The electron beam represents a form of ionizing radiation that is harmful to the health of any living organism. Clearly, we cannot shield the biological substance but we can try to extend its “life-span” during investigation before structural damage occurs. Researchers in the past have been ingenious in inventing preparation techniques for EM studies of biological samples. In the late 1950s staining techniques were introduced that addressed the first three issues at once (Brenner and Horne, 1959). Staining with salts from heavy metals (usually osmium or uranium salts) envelopes the structure of interest and after insertion into the EM the negative imprint of the structure can be observed. Moreover, it increases the contrast dramatically, according to the high atomic numbers of the metals used. However, the biological structure desiccates in the vacuum environment of the microscope and illumination by the electron beam literally incinerates whatever is left of the biological substance, leaving behind the imprint of the structure outlined by the staining substance. Staining and dehumidification alter the structure and, because of the nanocrystalline nature of the staining salts, limit the resolution, in the best cases, to approximately 2 nm. The major advantage of this technique is its simplicity (no special instrumentation is required), its speed, and the fact that

the negative “shadow” of the sample is almost unaffected by radiation damage. The obvious disadvantage is the highly “unnatural” state in which the investigation is performed, leading to artifacts and possible misinterpretations.

To study larger objects such as cells or organelles staining has been combined with embedding in epoxy resins and subsequent sectioning with the help of ultramicrotomes (Porter and Blum, 1953). Despite the fact that almost all our ultrastructural knowledge of the cell interior is based on this technique, there are some obvious limitations. First the substitution material, e.g., epoxy resin, shrinks during hardening, thus deforming the original structure. Second the stain applied after sectioning reduces the resolution (as described above) and third a second shrinkage step of the section is observed after initial illumination in the EM. This effect can be as large as 30–40% in the direction of the electron beam and 5–10% in the plane perpendicular to it. Shrinkage of the initial section dimensions is caused by outgassing of the solid polymer in the ultrahigh vacuum and to a larger extent by the loss of mass due to radiation damage (Luther, 1992). However, shrinkage occurs early in the illumination process, so that preexposing the sample until the contraction is complete can help to stabilize the section. The total electron dose is therefore increased dramatically and can lead to dose-induced stain aggregation. After preexposure the sample remains steady for further investigation of the modified structures and can be reused almost indefinitely.

3.1.2 Cryopreparation

In the 1970s researchers experimented on 2D protein crystals with the possibility of preserving specimen hydration in the EM (Taylor and Glaeser, 1973, 1974). However, In 1981 Dubochet and McDowell published their work on the vitrification of pure water for EM and in 1984 Adrian et al. reported the first successful investigation of a shock-frozen biological object, namely viruses, embedded in vitrified ice, now known as the cryopreparation technique. There is no doubt that this invention is one of the important milestones in life science EM studies, if not the most important. With this technique, scientists were, for the first time, able to investigate biological objects in a close-to-life state, quasi *in vivo*.

The object-holders are commercially available copper grids with a diameter of 3 mm (Figure 7–7A). They can be purchased in different geometries, mesh formats, and with different coatings (Plano GmbH, Wetzlar, Germany; Ted Pella, Inc., Redding, CA; Quantifoil Micro Tools GmbH, Jena, Germany; Ermantraut et al., 1998). Typically these grids are coated with a very thin support foil, usually amorphous carbon with thicknesses ranging between 4 and 20 nm. Shock-freezing has to be done very fast and therefore a special apparatus is necessary. These “guillotine”-like instruments, so called plungers, permit freezing of the sample in milliseconds to a temperature of liquid nitrogen (~90 K). To ensure a direct phase transformation from liquid to an amorphous (vitrified) solid, cryogenics with very high cooling rates, like ethane or less frequently propane, are used (Figure 7–7B and C).

The grid is clamped in an inverse tweezer and vertically mounted into the plunging instrument. A droplet of a couple of microliters of the sample solution, either isolated protein complexes or whole cells in their buffer solution, is applied to the surface of the copper grid. Since only thin objects in the range of a couple of hundred nanometers ($t \leq 500$ nm) are suitable for investigation with the EM, the sample volume, typically in the range of 1–4 μ l, has to be reduced. This is done with a filter paper carefully brought either to one side or to both sides of the grid to remove excess liquid. This process is called blotting. To prevent any mechanically induced disruption of, e.g., cellular structures, blotting can be done from the backside of the grid (Figure 7–8). Depending on the duration of the blotting process, the viscosity of the sample solution, and the size of the object, the thickness of the resulting ice layer can be adjusted. It guarantees a complete vitrification, without the formation of unwanted and harmful ice crystals, which can already be observed with fluid layers exceeding a couple of micrometers.

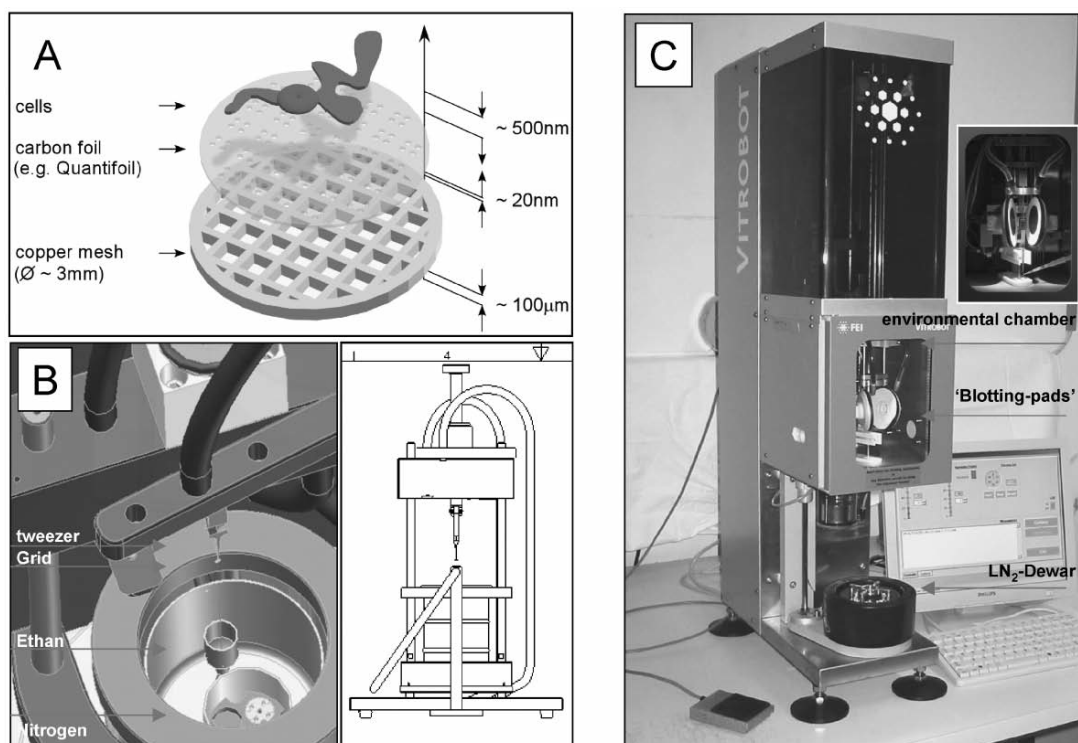


Figure 7–7. Plunge freezing instrumentation. a) Typical arrangement of cells cultured on TEM grid just prior to plunging. The schematic indicates the dimensions. b) CAD (computer aided design) image of a home-built plunge freezing apparatus (Images courtesy of R. Gatz, MPI of Biochemistry, Martinsried (near Munich), Germany). The small reservoir (yellow) in the middle of the dewar (green) contains the liquid ethane, which is chilled by a surrounding bath of liquid nitrogen. The forceps, which hold the grid are attached to a weighted arm. When the arm is released by means of a foot-trigger, the grid is gravity-plunged into the ethane. The cross-sectional scetch (left part of image B) illustrates the ‘guillotine-like’ arrangement. c) The VitroBOT (FEI company, Eindhoven, The Netherlands), a ‘robot’ for vitrification, allows the control of environmental and all necessary processing parameters. (See color plate.)

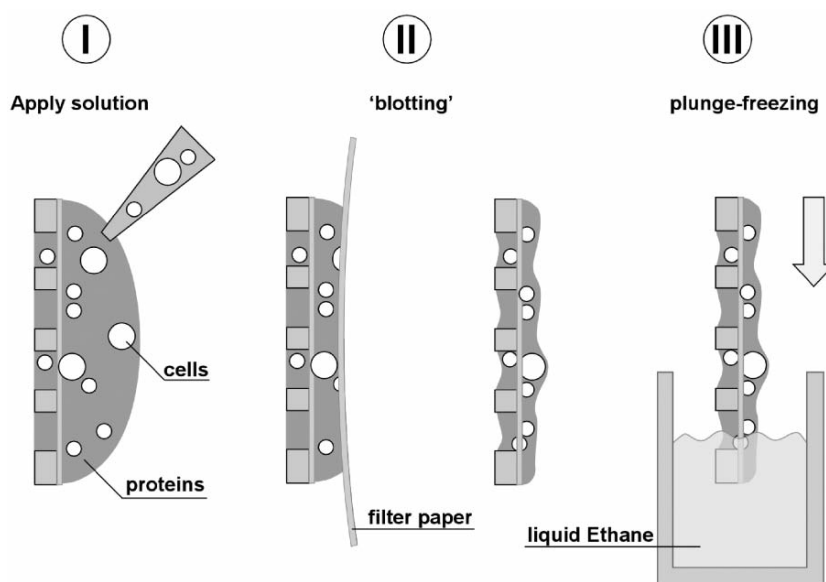


Figure 7–8. Steps to be taken in the plunge freezing process. I Apply the solution containing protein complexes or cells (TEM grid illustrated in a cross sectional way). To remove the excess solution and thus adjust the ice layer thickness, II blotting is done with a filter paper, carefully brought to the front- or backside of the TEM grid. After blotting the remaining solution on the grid is rapidly plunged III into liquid ethane, and afterwards kept at all times at liquid nitrogen temperature to prevent contamination, crystallization or smelting of the amorphous (vitrified) ice layer. (See color plate.)

Existing cryoplungers are mainly home-built and vary in design. Quite recently commercially available instruments have been introduced with the additional ability to adjust and control some of the preparation parameters, such as humidity, temperature, and air pressure, in specially designed environmental chambers. Moreover the whole process, except for the application of the sample, which has to be done manually, is computer controlled and automated, to facilitate the sample preparation process and expand its reproducibility (Vitrobot, FEI company, Eindhoven, Netherlands).

After vitrification, the sample is transferred in a cryoholder, specially designed to ensure that the sample remains in the frozen state at approximately liquid nitrogen temperature. If, by any chance, the temperature drops below -160°C the possibility that the vitrified ice will start to crystallize increases. Therefore during transfer and during the subsequent investigation, frozen-hydrated samples have to be kept at all times at liquid nitrogen temperature and below -160°C , which resulted in the name cryoelectron microscopy.

Compared to the commonly used techniques, cryopreparation allows us to preserve the native structure of the cytoplasm and the whole arrangement within the cell and even the physiological state at the moment of plunging. We thus circumvent alterations or modifications as seen with staining or epoxy resin embedding. The technique is

much more sophisticated and the result of the preparation is not known until a look is taken through the EM. The task of determining suitable processing parameters for the material at hand, to get, in the end, the best preparation result for an in-depth investigation of the biological structure, can therefore be very tedious.

3.1.3 Cryosectioning

While plunge-freezing is now almost routinely used for the preparation of purified and isolated macromolecular complexes as well as for prokaryotic and very small eukaryotic cells, larger bulky structures, such as big mammalian cells or larger organelles, cannot be addressed as a whole for EM investigations. However, the technique of sectioning using ultramicrotomes, likewise known from epoxy resin-embedded samples, has been adapted for the use in combination with high-pressure frozen samples. To obtain a vitrified sample of a specimen tenths of micrometers in dimension, plunge-freezing as described above is not sufficient, because of its limited "depth of vitrification," explained by the freezing rate, which is, for example 10^6C/s for pure water. If the sample dimensions exceed approximately $10\mu\text{m}$, vitrification is incomplete, which will result in the formation of cubic ice crystals. Therefore, high-pressure freezing is the method of choice, because the increase in pressure will lower the freezing rate (Moore, 1987). The lowest values are observed around 2000 bar, where the depth of vitrification increases up to 10 times as compared to freezing at ambient pressures (Sartori et al., 1993). The sample material, e.g., a cell suspension, is forced into a 15-mm-long copper tube with an inner diameter of approximately 0.3 mm. Afterward a pressure of 2000 bar builds up in a few milliseconds and, as soon as the pressure is established, cooling of the specimen, e.g., with a jet of pressurized liquid nitrogen, takes place. Before cutting the sections the copper jacket has to be removed using the cryo-ultramicrotome, designed to enable cutting at liquid nitrogen temperature. This is normally done with a diamond trimming device to form a square face of about $150\text{--}200\mu\text{m}$. Subsequent sectioning can be carried out using a diamond or a good glass knife (Figure 7–9). Cutting conditions such as knife-angle, cutting speed and stroke time, and temperature of the specimen, of the knife, and of the surrounding liquid nitrogen gas can be adjusted according to the specimen properties and the desired section thickness. However, the optimum sectioning conditions have to be found empirically in each experiment.

In contrast to conventional room temperature sectioning, where the sections are floated on water, cryosections have to be transferred manually to an EM grid with an eyelash, a tedious procedure that requires patience and great skill. The sections tend to attract contaminating ice crystals in the preparation chamber of the cryoultramicrotome and during transfer to the EM. Furthermore they are generally not flat and exhibit cutting artifacts, e.g., deformation, crevasses, chatter, and knife marks (for a complete description of cryosectioning and related cutting artifacts see Dubochet et al., 1988; Frederik et al., 1991; Michel et al., 1991; Sitte, 1996; Al-Amoudi et al., 2005).

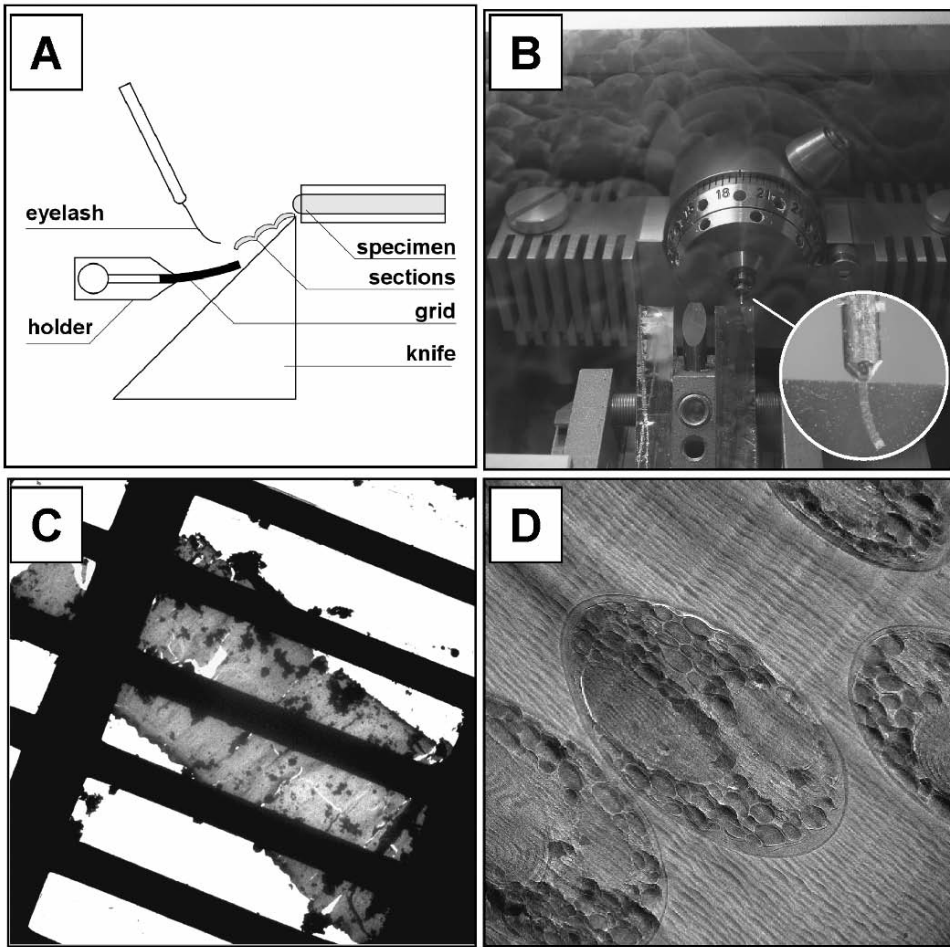


Figure 7-9. Cryo-sectioning. a) Cross-sectional schematic of the cryo-sectioning process with a cryo-ultramicrotome. b) View into the cryo-ultramicrotome during sectioning. The inset shows a magnification of the trimmed copper tube and the 'ribbon' formed after several cutting cycles. c) Low-magnification image of a cryo-section placed on a TEM grid. d) Cryo-electron micrograph at higher magnification of the sectioned cells (unicellular red algae *Cyanidioschyzon merolae*; images courtesy of A. Leis and L. Andrees, MPI of Biochemistry, Martinsried, Germany). The cutting artefacts are clearly visible.

While the idea of high-pressure freezing and cryosectioning is relatively old (Moore and Riehle, 1968; Bernhard, 1965) and commercially available instruments have been available for almost 20 years, its impact on biological and biomedical research has been comparatively small, owing to the fact that its use in combination with cryoultramicrotomy is still technically difficult, the yield of good micrographs is low, and the cutting artefacts in most of the sections complicate subsequent image analysis. However, in recent years cryosectioning was somehow rediscovered, and improvements have been made to facilitate the freezing and cutting procedure and to minimize the influence of cutting artefacts (Studer and Gnaegi, 2000; Al-Amoudi et al., 2003). Cryosec-

tioning to fabricate thicker sections of 150–200 nm for cryo-ET studies or even serial sectioning, likewise used in room temperature experiments (Soto et al., 1994), in combination with cryo-ET would definitely help in the 3D investigation of larger structures. The production of undistorted vitrified sections remains a specialized task and it is by no means a routinely used technique. Nevertheless the prospects for the future of cryosectioning in combination with cryo-ET are promising (Hsieh et al., 2002; Leis et al., 2005).

3.2 Radiation Sensitivity—Beam Damage

When biological specimens are irradiated by the electron beam in the EM, the specimen structure is damaged as a result of ionization and subsequent chemical reactions. Ionization occurs if energy is deposited in the specimen as a result of inelastic scattering events. This can primarily induce the heating of the sample in the irradiated area, radiochemical decomposition of water (radiolysis), as well as secondary chemical reactions (breaking of chemical bonds, formation of new molecules or even radicals). Unlike materials science, where beam damage is almost negligible, radiation damage in life sciences is the fundamental limitation in any cryo-EM investigation.

The damage that occurs depends on the number of electrons transmitted through the sample. Therefore the current density per unit area j (A/cm²) multiplied by the exposure time t (s) has been chosen as an appropriate measure for the electron dose (C/cm², or $e^-/\text{\AA}^2$, e^- = electrons). This is not the same as the definition in radiochemistry, where the dose is defined in units of gray; the energy adsorbed per unit weight (Gy = J/kg). According to the last definition, a carbon sample exposed to $50 e^-/\text{\AA}^2$ generated by a 300 kV source would correspond to a dose of 1.6×10^8 Gy, which, e.g., can be observed in the vicinity of a nuclear reaction!

Since the amount of damage is proportional to the applied dose, and the amount of exposure received by the specimen escalates with the square of the magnification, it is clear that this is the fundamental limitation in HREM of biological materials. However, several measures can be taken to partially evade this dilemma. Although the ionization processes are not temperature dependent, the resulting beam damage, e.g., bubbling, mass loss, diffusion of free radicals, is. Therefore, by lowering the temperature to, for example, liquid nitrogen temperature (~190°C), radiation damage effects can be drastically reduced. Cooling the sample from room temperature to 90 K, the “lifetime” (the time the sample can be investigated before structural damage is observed) can be increased by a factor of approximately nine, the so called cryoprotection factor C_p (Knapek, 1982; Conway et al., 1993; Stark et al., 1996). This cryoprotection factor is given by the ratio of the critical dose N_c at different temperatures (N_c is defined as the dose at which the diffraction intensity has fallen to $1/e$ (~37%) of its original value (Hayward and Glaeser, 1979). Further cooling to liquid helium temperature (~4 K) has been proven to be advantageous for very thin specimens as used for electron crystallography or even in selected cases for single particle

work [$C_p \sim 2$, if compared to liquid nitrogen studies (Fujiyoshi, 1998)]. However, the benefit of liquid helium cooling is still a subject of discussion, especially for cryo-ET, where cells embedded in thick ice layers are typically used. Unfortunately, the cryoprotection can be determined quantitatively only by measuring the fading of the electron diffraction patterns using 2D protein crystals. The transfer of these findings to frozen-hydrated specimens of, for example, cellular structures, where in most cases no crystalline arrays are present, remains difficult and therefore these values can be regarded as only a rough estimate. Moreover, the experiments made at helium temperature indicate a change in the density of the vitrified ice (after an initial illumination of a couple of $e^-/\text{\AA}^2$), from a low-density amorphous phase at liquid nitrogen temperature to a high-density form at liquid helium temperature, which could be responsible for the decreased contrast observed at around 9K (Heide, 1984; Schweikert et al., 2007; Plitzko et al., 2004; Comolli and Downing, 2005; Wright, 2006).

Regarding the complexity of the biological system at hand, every sample material will show a slightly different behavior during the investigation at low temperatures. High amounts of sugars, for example, either as ingredients of the buffer solution (DeCarlo et al., 1999) or the RNA content of the macromolecules, tend to increase the acceptable dose. Overall it can be said that electron damage at lower temperatures induces a gradual loss of high-resolution features, much slower than in room temperature experiments, and a reduced mass loss. However, at “very high doses” ($>500 e^-/\text{\AA}^2$) extensive mass loss, the formation of bubbles, and various other dimensional changes can be seen in all cases (Figure 7–10).

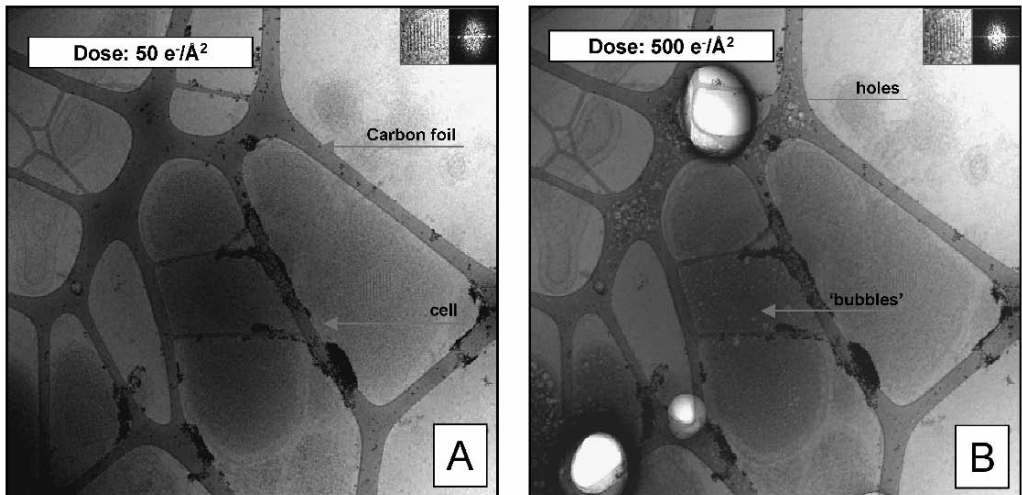


Figure 7–10. Electron dose and radiation damage shown for vitrified samples. a) Cryo-electron micrograph of a ice embedded prokaryotic cell on lacey carbon film exposed to $50 e^-/\text{\AA}^2$. b) The same cell after an exposure to $500 e^-/\text{\AA}^2$. The adverse effect of increased electron exposure manifest at the carbon foil and within the cellular structure in the formation of bubbles and furthermore in the smelting of the ice (Images courtesy of S. Nickell, MPI of Biochemistry, Martinsried, Germany).

A second measure among the physical methods to improve the stability of the sample in the electron beam is the use of intermediate voltage instruments. The ionization probability is, in a first-order approximation, inversely proportional to the acceleration voltage. Thus increasing the acceleration voltage will increase the penetration depth and thus reduce the influence of multiple scattering. Unfortunately, transmission electron microscopes with, e.g., 1250 kV acceleration voltage are huge instruments extending over almost three stories of a regular building. The cost for the instrument and for all subsequent resources is immense, making it almost impossible to place it in a normal laboratory environment. Furthermore, the higher the voltage used the lower the recorded image contrast, which is undesirable in the case of ice-embedded biological samples, where scattering occurs mainly from low Z-elements. Therefore, only a few high-voltage EMs have been built and installed around the world at selected places. It is more practical to operate EMs in the intermediate voltage range, between 200 and 400 kV. Instruments with 300-kV guns are the most common and they represent a good compromise in costs and resulting performance.

A last measure to reduce the exposure to electron radiation is the application of so-called "low-dose" acquisition schemes (Dierksen et al., 1992; Rath and Marko, 1997; see Section 3.4.7). Only during the time of recording electron micrographs is the electron beam allowed to illuminate the specimen. In all other cases the beam remains blanked. Additionally, the microscope can be preset in different states, to reduce the amount of time needed for changing the magnification or the adjustment of other electron optical parameters, during the process of screening, focusing, and final image acquisition. It is obvious that this "low-dose" approach became possible only with the introduction of computer-controlled instruments with the necessary stability and reproducibility in the illumination system.

During the acquisition of a tilt series in electron tomography, where typically 100–200 images are recorded, it is essential that the total applied dose stays below a critical value, which is given by the specific sample at hand. Most biological materials can tolerate an exposure of no more than $10 e^-/\text{\AA}^2$ to $100 e^-/\text{\AA}^2$, at which point major changes will have become evident in the structure of the specimen. According to the dose-fractionation theorem (Hegerl and Hoppe, 1976), the integrated dose of a conventional 2D image is likewise sufficient for a 3D reconstruction, if the resolution and the statistical significance between the two are identical. It is therefore feasible to distribute the total applied dose among as many statistically "noisy" 2D images as possible. A total applied dose of $50 e^-/\text{\AA}^2$ distributed among 100 2D images over an angular range would correspond to $0.5 e^-/\text{\AA}^2$ per image. Thus the resulting single images will be of a very noisy quality. Depending on the detection device utilized the statistical fluctuations from one picture element (pixel) to another can be much greater than the inherent change in image intensity from one pixel to another, i.e., the statistical fluctuations may exceed the inherent contrast. While the structure of interest is well preserved, it is impossible to observe the structural

detail because of the noisy quality of the image, associated with the very low SNR. Therefore it is very important that the quality and the performance of the detection device are sufficient to record images, even at very low doses (see Section 3.4.5). While the radiation sensitivity of biological samples restricts the exposure to the electron beam, the detection device fundamentally limits the resolution of the recorded tomogram. Investigations on ice-embedded biological samples under low-dose conditions can literally be described as “cloak-and-dagger operations” and only by means of appropriate data acquisition schemes and image analysis methods, i.e., tomographic methods, can the structural “secret” of the biological entity be disclosed.

3.3 Tilting Geometry

3.3.1 *Single-Axis Tilting*

The quality of a tomographic reconstruction is furthermore dependent on the tilting geometry within an electron microscope. In contrast to computer-assisted tomography (CT), tilting the specimen through the complete angular regime of 180° within the EM is not possible. For the collection of a single-axis tilt series, the specimen is typically tilted around $\pm 70^\circ$ in very small increments between 0.5° and 3° and an image of the same object area is recorded for every tilt angle. This restricted tilt range is due to the limited spacing within the objective lens polepieces, the slab geometry of the holder, and moreover the planar geometry of the object itself. At higher tilt angles (e.g., $\pm 90^\circ$) the electron beam is blocked by the holder and eventually by the bars of the copper grid, depending on the location of the structure within a mesh of the grid. This angular gap can be nicely illustrated in reciprocal space (respectively, Fourier space), where a wedge-shaped “blind” region is formed, the so-called “missing wedge” (Figure 7-11A). However, the requirement for a distortion-free 3D reconstruction, all projections of the sample over the complete tilt range ($\pm 90^\circ$) (Barnard et al., 1992), is not fulfilled, thus leading to imperfections in the reconstructed object. There are mainly two types of distortions that can be distinguished. First the resolution of the 3D reconstruction will be direction dependent. The resolution parallel to the tilt axis d_x will be determined by the resolution of the instrument itself. Perpendicular to the direction of the tilt axis, the resolution d_y will be directly related to the angular increment α_0 and thus to the number of projections N . The relationship between resolution d_y , increment α_0 , and the amount of projections N for a spherical object with diameter D is given by the Crowther theorem (Hoppe, 1969; Crowther et al., 1970), based on the projection slice theorem (Radon, 1917):

$$d_y = \pi \cdot \frac{D}{N}$$

The reconstruction of a 200 nm-thick cell (assuming that the cell is spherical), with a resolution of 2 nm, requires 320 projections with an angular increment of 0.6° . The Crowther expression assumes that the N projections cover the full angular space of $\pm 90^\circ$. Therefore the angular

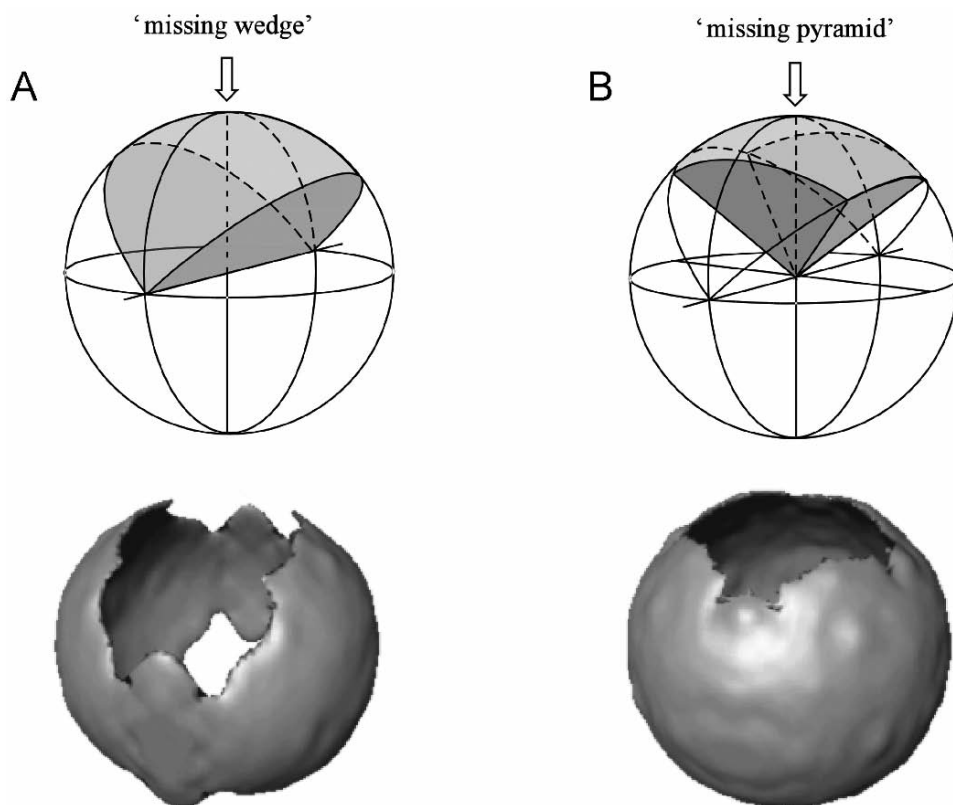


Figure 7-11. The diagrams show schematically the sectors in the Fourier domain, which remain unsampled owing to the limited tilt range (here $\pm 60^\circ$). a) In single-axis tilting there is a 'missing wedge'. b) In dual-axis tilting a 'missing pyramid'. The artefacts resulting from the missing information are shown below for an ideal thin-walled vesicle. In the single-axis case, the vesicle walls perpendicular to the tilt axis and the poles of the vesicle are poorly represented in the reconstruction. In the two-axis case, only the poles of the vesicle are affected by the missing information. In both cases, the same threshold level has been applied.

gap has to be included to determine the maximum achievable resolution in a tomogram, which directly implicates the second type of distortion that can be observed in ET: features will become elongated in the direction of the missing wedge. A spherical object, for example, a gold bead, will be elongated in its 3D reconstruction. This elongation can be described with the elongation factor e_{yz} , which includes the maximum tilt angle α (Radermacher and Hoppe, 1980):

$$e_{yz} = \sqrt{\frac{\alpha + \sin \alpha \cdot \cos \alpha}{\alpha - \sin \alpha \cdot \cos \alpha}}$$

The resolution in the third direction d_z , the "depth direction," is therefore further degraded to

$$d_z = d_y \cdot e_{yz}$$

To provide the maximum amount of 3D information, as many projections as possible should be acquired over as wide a tilt range as pos-

sible. In practice the molecular resolution is not limited by the Crowther criterion but by the structural preservation of the specimen and the signal recording and the noise (see above). Moreover, the size and the shape of the recorded object are poorly defined, especially in cellular tomography, where the cellular structure itself consists of objects of various sizes. However, it remains a useful guide for the expected resolution for a 3D reconstruction.

3.3.2 Dual-Axis Tilting

To reduce the missing information space in single-axis tomography, the acquisition of a second single-axis tilt series, where the tilt axis is rotated in-plane by 90° , can offer some remedy. Dual-axis tilting will reduce the “missing wedge” to a “missing pyramid” and thus increase the amount of information up to almost 20% (Figure 7–11B). Although the maximum resolution is not necessarily increased, the achievable resolution is definitely more isotropic. Moreover, structures that are “hidden” in the single-axis case, because of their orientation relative to the tilt axis and their location in regard to the missing wedge, will emerge in a dual-axis acquisition scheme.

However, dual-axis tilting for cryoapplications is demanding in two ways. First, special cryoholders have to be built, to enable the physical 90° rotation at liquid nitrogen temperatures within the EM (Nickell et al., 2003; Gatan, Pleasanton, USA). Second, for the 3D reconstruction both tilt series have to be combined as accurately as possible, which demands refined alignment routines (Penczek et al., 1995; Mastronarde, 1997; Nickell et al., 2007). It is clear that acquisition of a dual-axis tilt series requires either lowering the dose further or reducing the number of projections to preserve the structure under investigation, to comply with the requirements stated above. Dual-axis tilt cryoholders are now commercially available, which allows either a discrete or semicontinuous in-plane rotation by 90° (Gatan, Pleasanton, CA). The continuous rotation, implemented in side-entry cryoholders, simplifies the “tracking” (see Section 3.5.7) at lower magnifications of the sample during the in-plane rotation because the sample stays inside the objective. The discrete rotation is done “outside” the objective lens system but within the vacuum of the system, making it sometimes tedious to recenter the region of interest after the in-plane rotation.

Despite the missing information in Fourier space (wedge or pyramid), a more fundamental limitation in a tilting experiment within the EM is the increasing specimen thickness at larger tilt angles. Because of the planar (slab) geometry of the sample–holder arrangement, the transmission path of the electrons will increase. Simple geometric calculations show that the transmission path of the electrons for a specimen initially 100 nm thick will double at 60° (200 nm), almost triple at 70° (290 nm), and the projected thickness will be increased more than five times at 80° (590 nm). Thus at very high tilt angles the image contrast is dramatically reduced because of the increase in multiple scattering. For specimens with an initial thickness of more than 500 nm the projected thickness at high tilt angles would even prevent the trans-

mission of electrons in commonly used 300 kV instruments! However, to preserve the contrast at even higher tilts for imaging thick specimens such as vitrified cells, which can easily reach dimensions of a couple of hundred nanometers, energy filtering can offer some remedy (see Section 3.4.6).

3.4 Instrumentation and Automation

Among the keys to the successful investigation of biological structures at cryotemperatures and especially for the application of cryo-ET are certainly the transmission electron microscope (TEM), the preparation of the specimens (see Section 3.1), and sophisticated acquisition and automation procedures (see Section 3.4.7). We therefore have dedicated one section of this chapter to the equipment and the technical improvements made, and to a description of the necessary technical requirements for 3D structural analysis with the EM.

3.4.1 Acceleration Voltage and Emission

TEMs are now operated typically in an acceleration voltage regime between 100 and 1250 kV. With increasing voltage the penetration depth of the electrons increases, allowing us to image samples a couple of micrometers thick; at the same time the inelastic cross section is reduced, thus directly resulting in a longer resistance against beam damage (Wilson et al., 1992; Martone et al., 1999, 2000; O'Toole et al., 1999). However, high-voltage EMs are less frequently used for biological studies, mainly because of the decreased image contrast due to the reduced cross section. The gain in penetration power is limited, because the elastic σ_{el} as well as the inelastic cross section σ_{in} are in a good approximation proportional to $1/\beta^2$ (Scherzer, 1970), where β is the ratio of electron velocity to the velocity of light ($\beta = v/c$). Thus, increasing the acceleration voltage from 100 to 300 kV results in an improvement of a factor of 2, while an additional increase from 300 kV to 1.2 MV will result in only a 1.5 times increase (Koster et al., 1997). As mentioned earlier, it is more practical to operate EMs in the intermediate voltage range, between 200 and 300 kV, because this represents a good compromise in cost and resulting performance and, additionally, EMs are available in combination with FEGs.

FEGs were implemented in the late 1980s. Compared to thermionic emitters FEGs possess several advantages, notably a very small energy spread, usually in the range of ~ 0.8 eV, combined with an increased brightness. Both properties lead to an improved envelope of the CTF, since the increased temporal and spatial coherence reduces the damping of the envelope function (Frank, 1973). This is especially advantageous for ET of ice-embedded specimens at intermediate resolution (~ 1 – 2 nm) because a relatively large defocus (several micrometers) can be selected without sacrificing good contrast transfer. FEG instruments at 200 and 300 kV are today's "workhorses" in many different laboratories from life and material sciences. But LaB₆ 120-kV systems are widespread as well and are adequate whenever high resolution is not of major interest, e.g., for sample screening and assessment.

3.4.2 Electron Optics

Like regular light microscopes the EM is built up from three major lens systems: the condenser, the objective, and the projective lenses. The performance of these electromagnetic lenses almost exclusively determines the quality of the recorded image. Most important are the properties of the objective lens system, likewise known from light microscopy, especially the value of C_s , the spherical aberration coefficient. The objective lens influences the transfer of electrons, and together with the illumination system (FEG or LaB₆) and the chosen focus, is described by the CTF.

The opening in the center of the objective lens system, where the electron beam passes, affects the magnitude of C_s . In simple terms, the bigger this gap the higher C_s , the higher the contrast, but the lower the resulting resolution. However, the sample holder with the specimen is located inside the objective lens system and therefore the gap dimensions cannot be too small. In particular, to tilt the holder, e.g., for an angular acquisition, a considerable amount of space is required to allow tilts to higher angles, or the sample holder has to be designed in a way to enable tilting even within very small gaps (Fischione Instruments, Inc., Export, PA; Gatan, Inc., Pleasanton, CA). The objective lens spacing in current EM systems is in the range of a couple of millimeters, thus resulting in C_s values from 0.7 mm (lowest) up to 6 mm (highest). For life science applications typically objectives with a C_s of ~2 mm are used, because they offer enough space for tilting the specimen and at the same time the possibility of recording high-resolution images.

Ongoing instrumental improvements include, for example, the use of C_s correctors, which, especially in materials science, already revealed new possibilities and real image improvement (Haider et al., 1998; Lentzen et al., 2002; Jia et al., 2003; Freitag et al., 2005). However, in biological EM we have to cope with very weak scatterers and so, for reasonable image contrast, we have to defocus in the range of a couple of micrometers even for very thin objects. Since this low-frequency information is essential for cryo-EM investigations, the advantage of C_s correction is almost canceled (Plitzko et al., 2005). However, there is reason to believe that C_s correctors might be beneficial for biological EM and especially cryo-EM if used in combination with phase plates (Unwin, 1972; Danev and Nagayama, 2001; Majorovits and Schroeder, 2002; Lentzen, 2004; Marko, 2004).

Another important fact is the change in the direction and angle of the tilting axis at different magnifications due to a rotation of the image within the electron optics of the microscope. Previously this change was obvious at every magnification step. It is now partly compensated by so-called “rotation-free lens series,” which guarantee an almost “fixed” (in angle and direction) tilt axis in distinct magnification ranges. This is crucial, especially in low-dose acquisition schemes, where a fast transition and reproducible change between states (see Section 3.4.7) and sample areas at different magnifications are necessary, e.g., in dual-axis experiments to reposition the area of interest after the in-plane rotation.

3.4.3 Tilting Device: Goniometer Tilt Stage

As already mentioned, tilting the holder inside the microscope is a fundamental requirement for ET. The tilting device, normally located outside the EM column, is called the goniometer tilt stage and in some instances the “compustage,” because of its power to be controlled by computers. However, tilting is now done in an exclusively mechanical way, thus resulting in a major practical difficulty: the mechanical imperfections and the inability to set the eucentric height of the specimen precisely. The eucentric plane is normal to the optic axis; this way a point on the optic axis should not move laterally when tilted around the holder axis. As a result, specimens can experience significant shifts in the x , y , and z directions during the course of angular acquisition. Shifts in the z direction cause severe focus changes, and in an uncorrected case, complicate the 3D reconstruction or even render a reconstruction meaningless.

Typically, side entry holders are used—at the rear part of which a small Dewar for liquid nitrogen is located (Figure 7–12). The Dewar is connected to a temperature conduction rod, transferring the temperature of $\sim 180^{\circ}\text{C}$ to the specimen mounted at the outermost end of the holder tip. The cryoholder, almost 30 cm in length, is tilted by the goniometer around its vertical axis. It is easy to imagine that tilting must be very inaccurate. Especially at high tilt angles above $\pm 45^{\circ}$ imprecise tilting will add up to displacements in the micrometer range. However, at a resolution of $10\text{ \AA}$ a tolerable deviation would be in the range of $3\text{ \AA}$, thus resulting in a discrepancy of three to four orders of magnitude between the desired resolution and the actual precision of the tilting device. The best current stages can tilt and move the sample within an accuracy of $0.5\text{ }\mu\text{m}$. To partially compensate for the still large displacements, automation is mandatory during the acquisition of a tilt series composed typically out of 100 micrographs. Therefore sophisticated acquisition and correction schemes have been developed, to keep the feature of interest at all times focused and centered (see Section 3.4.7).

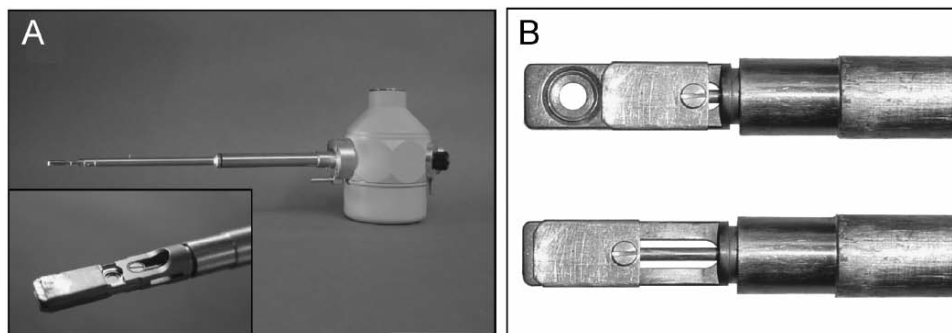


Figure 7–12. Photographs of a standard side entry cryo-holder (Model 626 Gatan, Pleasanton, CA, USA). a) The holder is almost 30 cm in length and equipped with a nitrogen dewar in the back. b) Images of the tip of the holder with open (top) and closed ‘shutter’ (bottom). The stationary cryoshield (‘shutter’) protects the grid from mechanical damage and from frost accumulation during transfer from the specimen loading workstation (not shown) into the TEM.

Developments have already been made to test and implement piezo-controlled stages into the EM environment to allow nanopositioning in the order of ~ 1 nm, which might enable accuracy during tomographic acquisition to be improved dramatically (Lengweiler et al., 2004; JEOL Ltd., Tokyo, Japan).

3.4.4 Cooling System

Since all investigations in cryo-EM have to be made under cryo conditions, i.e., under constant cooling with liquid nitrogen, special holders and/or special cryostages have to be designed. Typically, side entry holders are used—fitted with a small Dewar for liquid nitrogen (Gatan Inc., Pleasanton, CA). The requirements for mechanical and temperature stability are severe and it is quite clear that the dimensions of the holder tip must be smaller than the polepiece gap of the objective. Moreover, special cryoshields have to be installed to guarantee contamination-free investigations. They are seated next to the tip of the holder inside the microscope and connected to the cooling trap of the microscope for constant cooling with liquid nitrogen.

Side-entry cryoholders are routinely used, but recently stages incorporating the cooling system within the electron microscope have been designed (“Polaris” Tecnai F30 by FEI Company, Eindhoven, Netherlands, and Jeol 3200FSC by JEOL Ltd., Tokyo, Japan). By including the cooling system for the sample holder into the microscope the mechanical stability and the temperature stability of the holder are improved enormously, and additionally these systems allow cooling to very low temperatures, even to liquid helium temperature (~ 10 K). Furthermore the temperature can be kept stable over longer time periods; instead of hours in the side-entry case, days or even weeks are now possible. Moreover the stages designed for these cooling systems can hold several samples, unlike the side-entry holder where only one individual sample can be mounted at a time (Figure 7–13). Additionally, prototypes have been developed and tested holding up to 100 different samples, to enable high-throughput applications in the future (Gatan Inc., Pleasanton, CA). The exchange of samples is faster and the transfer of the samples to the microscope is done under vacuum conditions, to minimize the chance of contamination.

3.4.5 Detector

Temperature and mechanical stability of the stages, holders, and goniometers as well as the overall performance of the electron optics are crucial for high-resolution data acquisition, however, the quality of the detector system used is of vital importance, especially in cryoapplications under low-dose conditions (Downing and Hendrickson, 1999; Fan and Ellisman, 2000). Previously photographic plates were utilized, making a fast assessment of the collected data impossible; however, today microscopes are equipped with TV rate systems and digital cameras (Krivanek and Mooney, 1993; Faruqi and Subramanian, 2000), so-called CCD cameras (Janesick, 2001). With the help of the CCD camera it is possible to easily calculate online the Fourier transformation of the image, for example, to determine the focus or to readjust the microscope, e.g., astigmatism and coma correction. In principle they

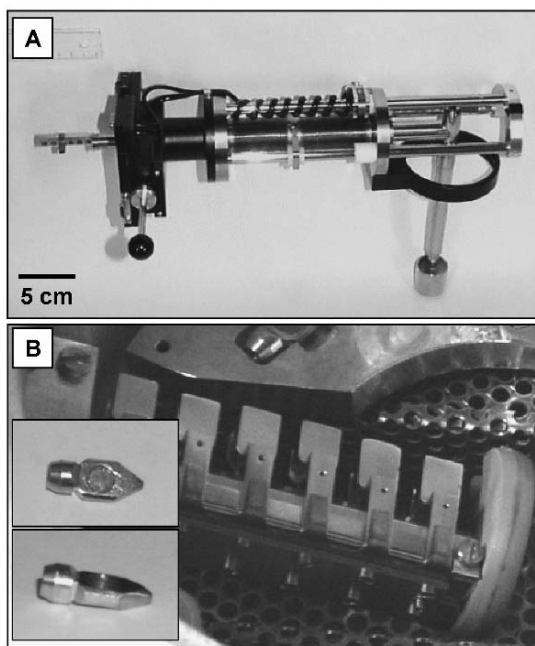


Figure 7-13. Multispecimen chamber and cartridges for the Polara TecnaiF30 (FEI company, Eindhoven, The Netherlands). a) The multispecimen chamber allows the storage of up to 6 samples, which are all protected by stationary cryoshields. b) The samples are clamped or screwed into cartridges (insets) and the transfer of the complete system is done under cryo- and vacuum conditions.

are built up from a scintillator (either a single crystal or a polycrystalline coating), “decelerating” the electrons and thus creating photons by cathodoluminescence. The scintillator is coupled to a fiber optic array, which will transfer the light optical signal to the CCD sensor, which is built up from a series of metal oxide semiconductor (MOS) capacitors, where the light induces the creation of electron hole pairs in the active area.

CCD cameras are now available in dimensions between 1024×1024 and 4096×4096 pixels with a pixel size ranging from 15 to $30 \mu\text{m}$. Thus they are still smaller than the typical photographic plate dimensions ($\sim 10,000 \times 8000$ pixels assuming a pixel size of $10 \mu\text{m}$ for a typical plate). However, CCDs possess an excellent linearity and a large dynamic range [i.e., ratio between maximum signal and the root mean square (RMS) of the noise level]. Because of the excellent linearity the intensity distribution of the recorded image is directly proportional to the number of primary electrons hitting the detection device. However, the lateral resolution is worse than in photographic plates, because of the signal spreading within the whole CCD assembly. This spreading is mainly caused by multiple light scattering of the photons created and by electron backscattering within the scintillator. Thus a point-like input signal produces an output that is spread over several pixels of the CCD chip. In real space this will result in a “blurring” of the recorded image, which is similar to a Gaussian smoothing of the image. In Fourier space it can be observed as a damping of the high spatial frequencies. The extent of this spread can be quantitatively analyzed by inspection of the modulation transfer function (MTF), which is the Fourier transform of the point spread function (PSF), to measure the spatial frequency response (Weickenmeier et al., 1995; Ruijter, 1995).

The second major characteristic of a CCD camera is the sensitivity, quantitatively expressed in the detection quantum efficiency (DQE). In combination with the conversion efficiency (number of created pixel electrons per primary electron) the DQE describes the influence of the detector noise on the SNR of a transferred image and is defined as the ratio of the square of the SNR at the output to the same quantity at the input of the detection device (Herrmann and Krahle, 1982; Fan and Ellismann, 2000). There are essentially four main sources of noise: (1) variations in the electron conversion of electrons into photons, (2) fluctuations in the generation of electron hole pairs on the chip, (3) dark current noise, and (4) preamplifier noise. To reduce the dark current noise, which mainly stems from the electronics, the whole array is cooled down to a temperature of -30°C .

Depending on the scintillator material and thickness, the primary energy of the electrons, and the coupling of the scintillator to the capacitors the detector will show slightly different characteristics (Fan and Ellisman, 1996). The scintillator has the key role in the performance of an image converter since it primarily determines sensitivity, resolution, and noise contribution. Typically, single crystals ($\text{YAG} = \text{Y}_3\text{Al}_5\text{O}_{12}:\text{Ce}^{2+}$) or polycrystalline phosphor scintillators (e.g., $\text{P43} = \text{Gd}_2\text{O}_2\text{S}:\text{Tb}^{3+}$) are used in commercially available CCDs.

Since high resolution is of major interest in materials science, CCDs for this purpose are normally equipped with very thin scintillator coatings to minimize the point spread function and to obtain the best transfer, especially in the high-frequency domain. On the other hand, the detection sensitivity and the amplification of the incoming signal are of minor interest since the specimens are not subject to restricted illumination conditions. This is in contrast to biological applications, where low-dose exposure schemes are typically used, demanding a very high sensitivity of the detection device and thus a somewhat greater scintillator thickness. The optimum thickness of the scintillator depends on the application and on the acceleration voltage used. While the performance for CCDs in the voltage range below 200 kV is satisfactory, it drops dramatically for intermediate or high acceleration voltages.

However, in principle both are needed; the highest sensitivity combined with the maximum transfer of low and high frequencies. But these two requirements are in conflict with each other; the higher the sensitivity, the thicker the scintillator layer, and the greater the possibility of electron backscattering and multiple light scattering, resulting in a severe damping of high-frequency information, especially at primary energies at and above 300 keV.

Clearly, the ideal detector for cryo-ET has not yet been invented, but there are possible ways to overcome the shortcomings of "classical" CCDs, at least to some extent: modified CCDs (Fan et al., 2000), lens-coupling (Mooney and Krivanek, 1994; Mooney, 2004) instead of fiber coupling, "decelerators" in combination with CCDs (Downing et al., 2000; Downing and Mooney, 2004), and active pixel detectors (APS; Fan et al., 1998; Faruqi et al., 2003a,b; Evans et al., 2005). In lens-coupled systems the glass fiber array, transferring the light without magnifica-

tion, has been replaced by huge optical lenses focusing the photons arising from one electron onto the CCD chip. A free-standing scintillator foil generates the light, which is then deflected by a 45° mirror into the lens-coupling system. In this way scattering from the interface between scintillator and glass fibers, and from within the glass fibers is eliminated and moreover the detection of X-ray quanta is completely bypassed, since they are not reflected in the optical mirror. The setup provides a definite gain in resolution, while the detection efficiency is more or less kept constant if compared to highly sensitive fiber-coupled CCDs (Agard; Mooney, 2004). The main technical difficulty in lens coupling is the quality and thus the performance of the optical lens system to guarantee an undistorted transfer (mainly geometric distortions) of the light onto the full array of the CCD chip. Deceleration of, e.g., 300 keV electrons to 100 keV is technically difficult, because it is necessary to slow down the electrons with a “countervoltage” of, e.g., 200 kV (Downing and Mooney, 2004) before detection. This countervoltage is generated with an electrostatic deflector stack floated at a voltage comparable to the acceleration voltage of the microscope. The CCD, the associated electronics, and the electrostatic deflector are enclosed within an insulated high-voltage tank. However, the shielding of the CCD against AC magnetic fields and against possible discharges is still a major technical difficulty, which will have to be addressed in the future.

While lens-coupled systems and decelerators have been designed and tested for 300 kV instruments and represent bulky apparatuses, detectors based on the active or hybrid pixel sensor designs [mainly based on complementary metal oxide semiconductors (CMOS)] come in sizes comparable to “classical” CCDs, with the possibility of detecting the incoming electrons directly. Pixel detectors have been successfully used in the field of X-ray crystallography and condensed matter physics and they do possess some distinct advantages; instead of indirect detection of the incoming electrons, pixel detectors detect the signal directly, the lateral resolution is dramatically increased, and moreover dark noise (mainly from the electronic readout) is quasi-nonexistent (Turchetta et al., 2001). The signal produced by pixel detectors is intrinsically digital, thus allowing very fast readouts, but one major handicap is their “radiation hardness”; while CCDs can continue to be used, pixel detectors do have a somewhat limited lifetime. However, recent developments based on the latest CMOS technology have shown that the radiation hardness can be increased, e.g., if the size of the CMOS capacitors is reduced (Evans et al., 2005).

All three of these different detection “philosophies” are currently being tested in prototypes in combination especially with 300-kV instruments to find an optimum recording device and a long-term solution for digital TEM imaging.

3.4.6 *Energy Filter*

To image thick specimens such as vitrified cells, the development and the subsequent integration of energy filters (Krivanek et al., 1991, 1995) have been a major breakthrough for biological imaging and especially

for ET. Electrons interacting with the sample atoms can be scattered elastically (no loss in energy but change in direction) or inelastically (loss of energy due to interaction with core shell electrons). The mean free path Λ_{in} of inelastically scattered electrons (the path length after an inelastic scattered electron has been scattered once on average) is about 280 nm in vitrified ice at 300 kV at liquid nitrogen temperature and around 210 nm at liquid helium temperature (Schweikert et al., 2007; Plitzko et al., 2004; Wright et al., 2006). Thus for specimens with a thickness in the range above 200 nm multiple scattering is inevitable (Figure 7–14). For biological specimens and vitrified ice, which are almost exclusively built up of light atoms, inelastic scattering is much stronger than elastic scattering. However, the elastically scattered electrons carry the main portion of structural information of the specimen, whereas inelastically scattered electrons form a strong inelastic background and thus contribute significantly to the shot noise of the image background, obscuring fine details. Because of the different energies of inelastically scattered electrons, they are slightly “out of focus” and will thus contribute to a “blurring” of the final image and hence they decrease the SNR.

Energy filters are in principle electron energy loss spectrometers with the additional ability to image the sample at a specific kinetic energy. Introduced in the early 1990s (Krivanek et al., 1992; Probst et al., 1993; Gubbens et al., 1993) they are widespread in materials science for the

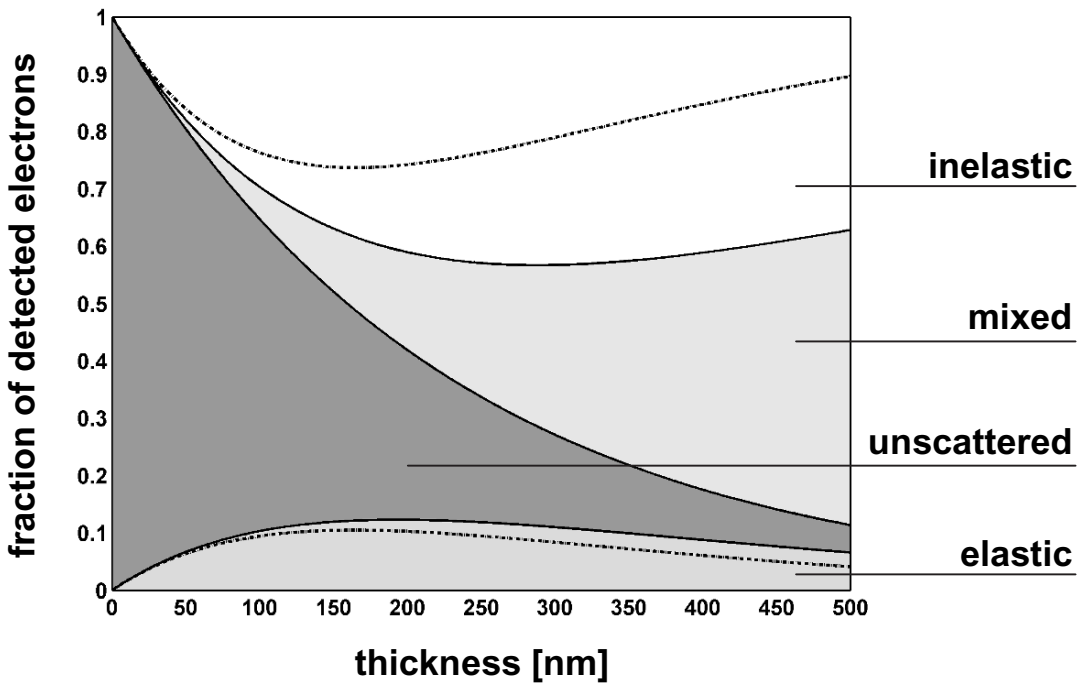


Figure 7–14. Distribution of scattered electrons for vitreous ice. Distribution (without an aperture) over the elastic (lower intermediate grey area), inelastic (upper dark grey area), and mixed (light grey area) scattering channels for vitreous ice as a function of thickness for 300 kV, in nm). The dashed lines mark fractions of single (elastic or inelastic) scattering.

analysis of the chemical composition, either by high resolution electron energy loss spectroscopy (EELS) or electron spectroscopic imaging (ESI). For a complete description of energy filters see Reimer (1995) and for electron energy loss spectroscopy see Egerton (1996). Energy losses can be selected that are specific to certain elements to obtain so-called elemental maps, visualizing the elemental distribution in one single 2D image (Figure 7–15). However, the dose typically used for an elemental map is in the order of $10,000 e^-/\text{\AA}^2$, thus orders of magnitude higher than what would be allowed for the investigation of biological structures. Nevertheless, instead of using the specific energy loss of elements (characteristic ionization edges) for “inelastic filtering” it is always possible to operate the energy filter for “elastic” filtering or to use a more common expression for “zero-loss” filtering. The energy-selecting slit aperture is positioned around the energy of the primary beam, selecting only the unscattered and elastically scattered electrons for imaging. The energy window is typically in the range of 20 eV or slightly larger, but it has to be smaller than the smallest relevant energy loss in the sample (in most cases the 20–25 eV single plasmon loss peak).

Basically the energy filter is built up from one or several magnetic prisms. If an electron travels through a homogeneous magnetic field it will move in circles whose radius is proportional to the kinetic energy of the electron, and in this way electrons of different energy are separated. In the so-called energy-selective plane (or plane of energy dispersion) they are located in different positions, and produce a spectrum, which is called the EELS. A mechanical slit aperture can be inserted in this plane to select electrons with a specific energy. There are basically two types of energy filters: postcolumn energy filters (Krivanek et al., 1991) and in-column energy filters (Rose and Plies, 1974; Lanio, 1986; Lanio et al., 1986). For the postcolumn filter only one 90° magnetic prism is utilized and placed at the lower end of an EM behind the

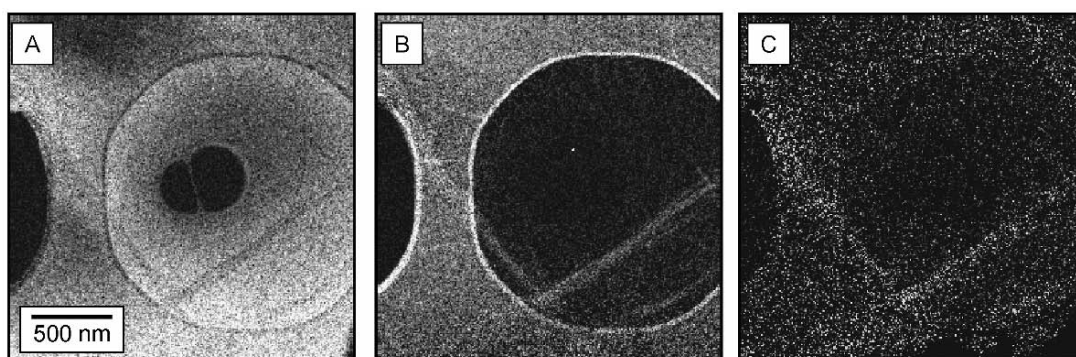


Figure 7–15. Electron spectroscopic imaging of microtubulins embedded in vitrified ice. With the help of an energy filter one can select energy losses specific to certain elements to obtain so-called elemental maps. The maps shown here are created with the three-window technique. Two images have been acquired in front of the ionization edge (pre-edge images) and one directly after it (post-edge image). The two pre-edge images are used to calculate the background intensity, which is extrapolated and subtracted from the post-edge image, resulting in the net-signal of the element under scrutiny. a) Shows the elemental distribution of oxygen, b) of carbon and c) of nitrogen.

projective lens system. In-column energy filters, mainly omega (Ω) filters, are placed above the projective lens system of the microscope within the column. In omega filters the beam has to pass through four magnetic prisms in a symmetric path corresponding to the greek letter omega.

Energy filters are additional optical elements and therefore they will add optical distortions of different kinds (spectral aberrations both transmissivity and nonisochromaticity as well as geometric distortions; Uhlemann and Rose, 1996) to the final image, which have to be compensated by adding additional multipole lenses to the system. In the case of the postcolumn energy filter (Gatan Imaging Filter, GIF, Gatan, Pleasanton, CA) several quadrupoles and hexapole lenses are placed behind the magnetic prism to correct these distortions up to a certain degree. These lenses have to be aligned on a day-to-day basis, to ensure the best performance of the system and, to make life easier, the necessary alignment routines have to be computer controlled and automated.

Because of the symmetry of the in-column omega filter, a number of aberrations are canceled out in this type of energy filter, simplifying the alignment procedure. Furthermore, the pre- and postfilter electron trajectories are parallel, which makes the in-column filter design an obvious choice. However, in-column filters cannot be included in an existing TEM column, in contrast to postcolumn energy filters, which lack the symmetry, but which can be attached to almost every existing microscope column.

Moreover, with the addition of energy filters the whole electron optical path will be increased, thus naturally resulting in an additional magnification factor. In the case of the in-column filter this magnification factor is quite small, while for the postcolumn filter the magnification factor ranges between 20 and 40 depending on the acceleration voltage used. Obviously, this postmagnification is well suited for high-resolution work, as typically used in materials science. However, to apply the filter to biological specimens, the magnification range, in which the objective lens is excited (to allow focus adjustments), has to be extended to much lower magnifications. Therefore special lens series have to be utilized, called EFTEM series, which demagnify the image, before entering through the spectrometer entrance aperture into the energy filter. In this way, the magnification factor is reduced to a value in the range of 2–3 if compared to images observed on the viewing screen. Recently, larger entrance apertures (5 mm in diameter) for the energy filter have been introduced (Brink et al., 2003) to increase the collection angle and thus allow the magnification factor to be further reduced. Especially in biological applications in cryoconditions where limitations include the dose and thus the applied magnification, this has proven to be beneficial because the field of view (FOV) can be considerably increased.

The application of zero-loss filtering for collecting high-resolution structural data from ice-embedded biological macromolecules, which are normally small in size compared to cellular structures, provides only a minor improvement, and therefore it is rarely used (Langmore

and Smith, 1992; Schroeder et al., 1993; Frank et al., 1995; Zhu et al., 1997; Angert et al. 2000). Typically these ice-embedded isolated particles are thinner and thus smaller than the inelastic mean free path and therefore the influence of inelastically scattered electrons is comparably small. However, zero-loss filtering allows the use of smaller defocus values without impairing the accuracy of the subsequent alignment of the particles as required for averaging (see Section 2.2). In electron crystallography, energy filtering has its advantages even for thin specimens, as shown by Yonekura et al. (2002). Overall the effect of energy filtering (especially zero-loss filtering) for thick specimens in the thickness range above the mean free path (~ 200 nm for 300-keV electrons) is more obvious, because details can be observed with a higher contrast and higher resolution than in the unfiltered case (Figure 7–16). Moreover in very thick specimens, in which most of the electrons have been inelastically scattered, it is possible to choose an energy-selective window close to the most probable, instead of the zero-loss value (Olins et al., 1989; Han et al., 1996; Bouwer et al., 2004). Moreover, energy filtering allows us to estimate or, if the mean free path is known, to determine the absolute specimen thickness quickly and simply by means of the so-called log-ratio technique (Malis et al., 1988; Grimm et al., 1996b); by taking the logarithm of the ratio between an unfiltered image and a zero-loss image the relative thickness of the specimen can be obtained in units of the inelastic mean free path. If the inelastic mean free path is known this value can be easily converted to the absolute thickness. This simple procedure can be additionally utilized for the determination of the inelastic mean free path, if the thickness of the specimen has been determined by other techniques (Grimm et al., 1996b; Feyta and Aebi, 1999). Thickness determination is advantageous to ascertain suitable sample areas for follow-up in-depth structural investigation and thus represents a quality criteria, especially for automated acquisition schemes.

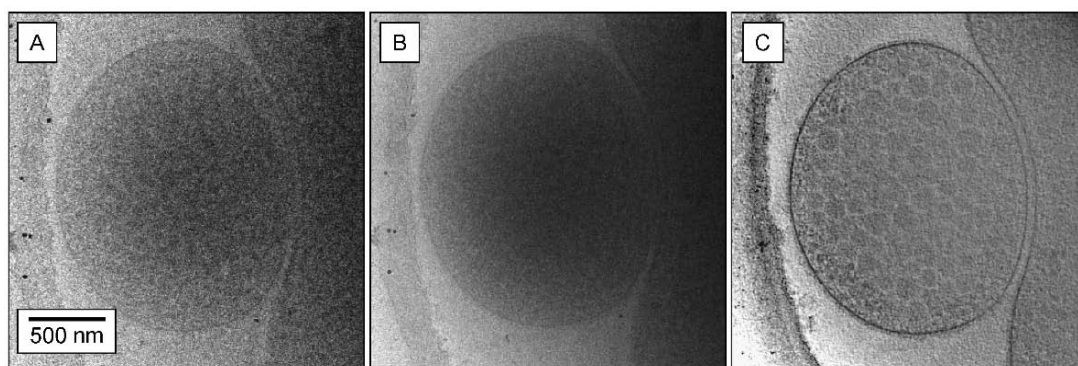


Figure 7–16. Influence of zero-loss filtering on the investigation of ice-embedded thick specimens (*Thermoplasma acidophilum*, images courtesy of C. Kofler, MPI of Biochemistry, Martinsried, Germany). Unfiltered (a) and zero-loss (b) filtered images of a *Thermoplasma acidophilum* cell. The specimen thickness is almost 500 nm, which is greater than the mean free path of 300 kV electrons in ice at liquid nitrogen temperature. The slit-width was 20 eV. c) shows a section through a tomographic reconstruction based on the filtered image set.

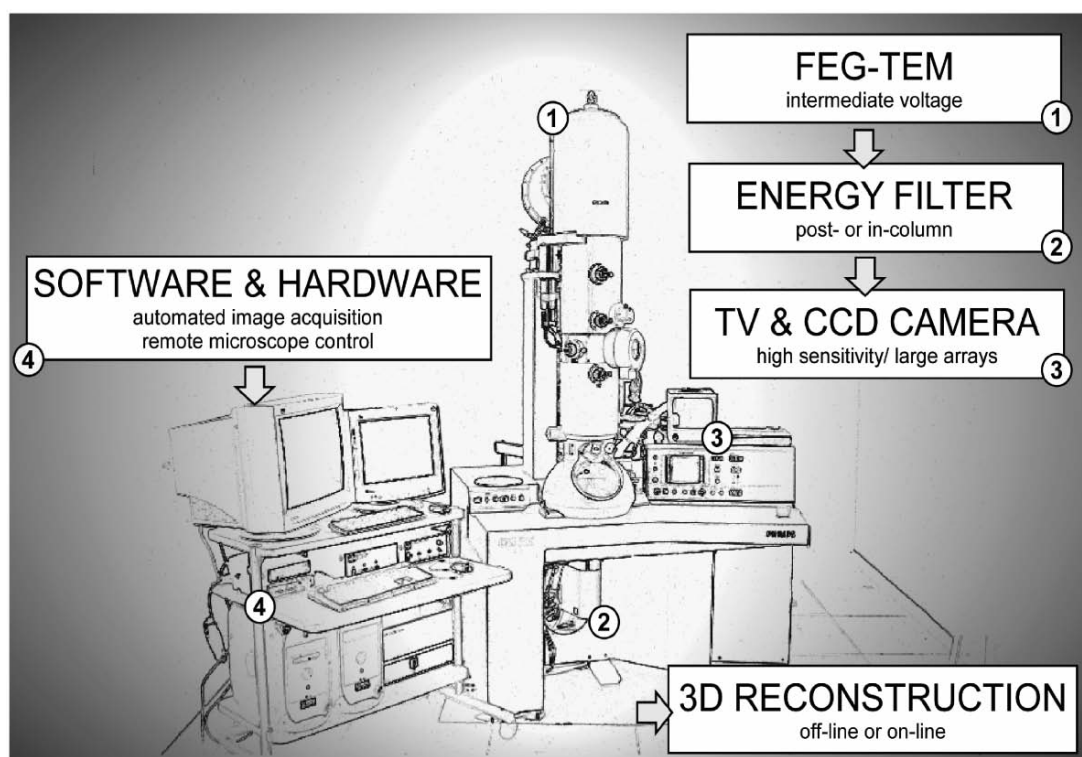


Figure 7-17. Experimental setup for automated tomography based on an intermediate voltage (300kV) instrument (1) equipped with an energy filter (2). The CCD camera (3) makes the image formation available instantaneously and, in conjunction with a computer (4), allows measurement of deviations from a perfectly eucentric tilt geometry. This information is then used, in a feedback loop to the microscope, to activate auto-tuning functions that compensate for the mechanical imperfections of the system.

3.4.7 Automation

Experimentally, the acquisition of a tomogram requires recording of a “tilt series,” which is a sequence of micrographs from different angles containing the feature of interest. However, it took more than a decade after the pioneering work of Hart and Hoppe to establish angular acquisition schemes, especially under cryoconditions as a routine and robust method in structural biology (Hart, 1968; Hoppe et al., 1968; Figure 7-17). This is due to the fact that the specimen, or the feature of interest, inevitably moves out of the field of view during physical tilting of the specimen holder as a result of the mechanical inaccuracies of the tilting device (see Section 3.4.3; Figure 7-18A). The adjustment of the microscope optics, the recentering of the specimen in the electron beam, focusing, and the final recording of the electron micrograph were exclusively done by hand in the first experiments. However, the additional exposure time required for these steps resulted in the overexposure of the specimen and thus in extensive beam damage, which finally culminated in the loss of the sample. Manual recording of a tilt series is therefore inappropriate and the implementation of automated acquisition, which includes “tracking” (recentering) of the specimen, refocusing with a minimum additional exposure (Dierksen et al., 1992;

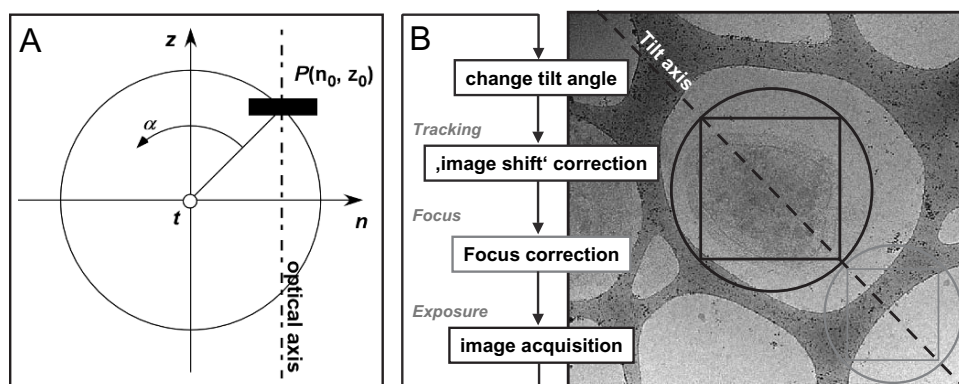


Figure 7-18. a) Geometric model of specimen rotation around a tilt axis. b) Automated data collection scheme for the 'full-tracking/ full-focusing' case. For tracking, focusing and the final image acquisition the same magnification is used and the beam is adjusted in a way that the areas where tracking, focusing and exposure is done do not overlap (blue and green circle). Using very short exposure times in combination with high binning values (4×8 or even 8×8) the exposure to the sample in auto-tracking and auto-focusing can be minimized. (See color plate.)

Koster et al., 1992; Braunfeld et al., 1994; Rath et al., 1997), and image acquisition, was essential for the success of ET in structural biology. Clearly, only with the introduction of computer-controlled electron microscopes was automation within reach.

Automated tomography includes three major steps for every tilt angle: tracking, focusing, and final image acquisition (Figure 7-18B). Under strict low-dose conditions, the tracking and focusing steps are normally carried out adjacent to the area of interest, to reduce the dose that is delivered to the specimen area. These areas are positioned along the tilt axis with respect to the actual acquisition area to determine the correct shift and focus levels prior to final image recording. During the tracking step a micrograph with a very low exposure time at a higher binning mode of the CCD camera is acquired, which is correlated with a micrograph from the previous tilt. Utilizing cross-correlation or by a manual inspection of prominent features (e.g., gold beads), the shift relative to the preceding micrograph can be determined. Depending on the SNR of the image, the cross-correlation algorithm might fail and must be supported by suitable image processing routines, e.g., band-pass filtering or the use of a Hanning window. This is especially the case at higher tilt angles where the projected thickness increases. The shifts are typically compensated by moving the beam with the image deflection coils back to the region of interest. To obtain a higher accuracy (better than 1%) with autotracking the procedure can be repeated several times. The offset in z between the object and the tilt axis z_0 can be minimized by setting the eucentric height. Procedures have been developed to adjust the sample z height automatically until minimum image movement is detected while the goniometer is wobbling within a specified angular range ("Explore3D," automated tomography software of FEI Company, Eindhoven, Netherlands; Schoenmakers et al., 2005). However, due to mechanical imperfections of the goniometer, it

is very difficult to tune z_0 to zero. As a result, the uncorrected residue of z_0 can still be severe enough to make the object disappear from the field of view at high magnification and lead to significant focus changes during the acquisition of a tilt series. Furthermore, it is observed that this offset in z depends not only on the actual sample being used but also varies from point to point within a sample. This can be attributed to many factors including nonuniform thickness and flatness of the specimen. Therefore z_0 has to be determined during data collection for each selected tilt angle.

Automated focusing or autofocus is achieved by acquiring images with different beam tilts (Koster et al., 1992; Koster and de Ruijter, 1989; Ziese et al., 2003) prior to the final exposure. It is based on the fact that tilting the illuminating beam will result in an image displacement, if the image is out of focus. The amount of this displacement, measured by means of crosscorrelation, is proportional to the focus change and can be used to readjust the objective lens current, rather than using the z adjustment of the stage. The amount of shift and focus change that occurs during data collection depends on the goniometer and specimen holder. In most experiments, accumulated shifts are in the order of 1–3 μm during the acquisition of a complete tilt series, provided that the specimen was set as accurately as possible to the eucentric height.

Image shifts are measured by determination of the peak position of the cross-correlation function (XCF) of an image recorded after setting a new tilt angle with a reference image that was recorded previously. Since the two correlated images are acquired at different tilt angles, they will not only be shifted relative to each other but they will also differ to some extent as a result of their different projection angles. Moreover, a tilted image (assuming the tilt axis is parallel to the y -axis and in the center, of the image) will exhibit a defocus difference between the left side of the image, the center, and the right side of the image. Thus defocus determination by the beam-tilt method for tilted samples is impaired by an elongation of the cross-correlation peak, which has contributions from areas with different defocus values. Therefore, prior to the computation of the cross-correlation function, “stretching” of the image at a higher modulus can be applied (Guckenberger, 1982; Ziese et al. 2003). “Stretching” is typically done perpendicular to the tilt axis by a factor of $\cos(\alpha - \alpha_0)/\cos \alpha$ or a similar expression depending on the sign of the actual tilt angle α and the tilt angle increment α_0 (Koster et al., 1997). Instead of stretching the image, the focus determination can also be done in the center region of the image provided that beam tilting is done along the tilt axis to minimize errors in the autofocus procedure. The accuracy of the beam-tilt autofocus procedure without stretching can be as good as $\pm 50 \text{ nm}$ (Dierksen et al., 1993).

Since the effective thickness of the specimen in the beam direction varies during tilting, the portion of “absorbed” electrons will increase with the tilt angle. There are two possible ways to account for this effect: a tilting scheme with a finer sampling at higher tilt angles and an increase in exposure time for higher tilt angles to obtain a similar SNR

within the projections of a tilt series. A tilting scheme according to Saxton has been widely adapted ("Saxton scheme"; Saxton et al., 1984). Since this scheme was proposed for crystalline specimens a slightly different expression is typically used in electron tomography:

$$\alpha_{n+1} = \alpha_n + \arcsin(\sin \alpha_0 \cdot \cos \alpha_n).$$

To achieve a uniform SNR in the single projections of a tilt series and to prevent an insufficient SNR at high tilt angles, and an unnecessarily high SNR at low tilts, the exposure time can be varied in different ways. Instead of using the constant exposure scheme ($t = t_0$) it is possible to use an exponential scheme or a cosine scheme to allow for the thickness increase at high tilts:

$$t = t_0 \exp\left[T \cdot \left(\frac{1}{\cos \alpha - 1}\right)\right]$$

or

$$t = t_0 \cdot \frac{1}{\cos \alpha}$$

or

$$t = t_0 \cdot \frac{1}{\exp[\cos \alpha]}$$

where T is the relative specimen thickness in units of the mean free path (Grimm et al., 1998; see Section 3.4.6). For the reconstruction, the images have to be normalized with respect to the exposure time, so that the incident beam dose will stay nominally constant for all tilt angles. Since the thickness of the sample under investigation is normally not known beforehand, the cosine scheme is favored and is primarily used for the acquisition of a tilt series.

To minimize the electron dose that is delivered to the specimen, the images required for these automated procedures have to be recorded on areas that do not coincide with the chosen area of interest. This way almost all of the total dose (97%) can be employed for the acquisition of the single projections and only a fraction of it (3%) is used for the autoadjustment procedures. In less strict low-dose schemes, these necessary steps can be performed directly on the region of interest, but with different illumination and exposure settings, which will then sum up to 10–15% of the total applied dose.

Clearly, tracking and autofocusing prior to the final exposure are time consuming. However, with the improvements made in electron optics, stage designs, and specimen cooling, even of the already computer-controlled microscopes, it became feasible to use calibration and prediction schemes to dramatically reduce the amount of time typically used for the acquisition of a tilt series (Zheng, 2004; Ziese, 2002).

The acquisition schemes and software packages of today can be grouped in roughly three major domains: the almost classical "full-tracking/full-focusing," the "precalibrated," and the "on-the-fly or dynamic prediction" procedure. Naturally, there are various deriva-

tives and hybrids of these three different acquisition models, especially modified to serve specific purposes and applications. The full-tracking full-focusing scheme (described above) is the most common, due to the fact that it can be literally used universally, e.g., with almost any kind of sample and microscope and holder system. However, the time required for the recording of a tilt series can be very long, up to several hours, depending on the acquisition parameters.

The precalibration model was developed originally for side-entry systems. It mainly includes the calibration of the tilting stage synchronized with the specimen holder. The basic idea is that the image movement is determined prior to data collection. The movement of the stage is measured with a calibration sample (typically fiducial markers on a carbon foil) both in the xy plane (image shifts) and the z direction (defocus change) for the range of tilt angles needed to acquire a tilt series. Typically, the calibration is started at a high tilt angle (e.g., either -65° or $+65^\circ$, depending on the characteristics of the stage) and continues to the outermost tilt on the opposite side. For every tilt angle an image is acquired and the image feature is recentered automatically (as described above) while the defocus change is measured (not corrected) only with the beam-tilt method. In this way, the absolute image shifts in xy and z (defocus) can be determined and stored in a calibration file, which then provides the basis for the xy and z shift compensation during actual data acquisition. Moreover, it is also possible to mathematically model (least-squares fit) the displacements during tilting, based on the calibration curves now available, to predict the behavior of the stage. However, there are some restraints on this course of action: the calibration strongly depends on the initial setup of the microscope (illumination settings, image shift settings, distance from eucentricity, initial xy position of the stage), sample being used, and thermal stability of the cryoholder utilized. Thermal drift of the specimen is not included in the calibration. For higher magnifications, especially, it has to be addressed separately, for example, with an additional tracking step by cross-correlation. Additionally, a significant misalignment of the optic axis relative to the tilt axis can necessitate large corrections in both translation and focus, which in turn can lead to image rotation and magnification changes. Since the image rotation and a possible change in magnification are not covered by the pre-calibration procedure, the acquisition is severely hampered. However, prealigning the optic axis to the tilt axis by invoking an appropriate amount of image shift can circumvent this possible source of error (Ziese et al., 2002). Based on the measured calibration curve this “optimized position” (where the optic axis coincides with the tilt axis) can be determined and adjusted. Overall, it can be said that even in case of thermal instabilities, which will lead to rather small displacements, not every change in tilt angle will require a tracking step and thus the acquisition of a tilt series by precalibration is definitely sped up enormously (five times faster than the full-tracking method).

The prediction-based scheme assumes that the sample follows a simple geometric rotation and that the optical system can be characterized in terms of an offset between the optic and mechanical axes of

the microscope (Zheng et al., 2004). In contrast to the precalibration and full-tracking procedures where the acquisition begins at one end of the tilt range, the prediction scheme starts at zero tilt. In this way, it is possible to model and compensate for any errors in sample geometry simultaneously. The image movement in the xy and z direction due to a stage tilt can be dynamically predicted without precalibration. Mathematically, one data point, characterized by the tilt angle and the x translation, is sufficient to calculate the offset between the object and tilt axis (z_0). However, least-squares fitting from multiple data points is typically used to minimize the consequences of errors introduced when measuring the x translation. In this way, it is possible to estimate and refine “on-the-fly” the focus offset z_0 during the course of data collection, with an accuracy of 100 nm.

The microscope optical system (beam/image shift and focus) is automatically adjusted to compensate for the predicted and refined image movement prior to taking the projected image at each tilt angle. As a consequence, it is not necessary either to record additional images for tracking and focusing during the course of data collection or to spend valuable setup time in a lengthy precalibration of stage motions. However, thermal drift is not taken into account in the dynamic prediction procedure, which will affect the precision of the prediction and thus the quality of the recorded tilt series.

Nevertheless, the acquisition of tilt series of cryosamples even with the given advances in automation is not yet performed on a high throughput basis. This is mainly due to the fact that side-entry cryo-holders are typically used, which are still subject to thermal instabilities, and that the accuracy and reproducibility of mechanical tilting still require compensation with elaborate automation procedures (see above). However, recent advances in instrumentation, like the integration of the cooling system into the microscope column, the simplification of the sample transfer and mounting procedure, multispecimen holders, and the development of piezo-controlled stages will widen the prospects of high-throughput approaches in the near future.

3.5 Alignment, Reconstruction, and Visualization

3.5.1 Alignment

To obtain the 3D image from a set of acquired projections two initial steps have to be carried out: first the individual micrographs need to be aligned to a common coordinate system. The second step is the actual 3D reconstruction of the tomographic volume. The compensation of the specimen movement during the data acquisition process will keep the feature focused in most cases but the compensation for the xy displacement is not sufficient for a subsequent reconstruction, which makes a second, more precise alignment necessary. Primarily, the alignment procedure has to determine the angle of the tilt axis and the lateral shifts. Other changes, such as magnification changes or image rotation, due to large defocus changes during the acquisition, have to be determined and compensated for as well, if present in the recorded tilt series.

Alignment of the individual micrographs in cellular tomography is normally done by high-density markers, so-called fiducial markers. Typically, spherical gold beads with a diameter of ~10 nm are added to the sample solution or directly onto the carbon foil prior to the vitrification process. They can be easily recognized within a single 2D micrograph as “black dots,” due to their high Z number and thus their pronounced elastic scattering. The coordinates of the markers on each projection are determined manually or automatically. To minimize the alignment error as a function of the lateral translations and the tilt axis angle, an alignment model can be calculated based on least-squares procedures (Lawrence, 1992). Their locations are then adjusted to a 3D coordinate system. Apart from translations, this procedure for the determination of a common origin often accounts for possible image rotations or magnification changes (Lawrence, 1992; Luther, 1988). However, the larger the number of determined parameters, the more gold beads need to be located to achieve a sufficient significance level during the minimization of the residual. Therefore, another alignment approach is by means of cross-correlating the entire image (Taylor et al., 1997) or part of the images. Sequentially the images are compared and compensated for image shifts throughout the series. This procedure can be repeated iteratively to achieve a higher accuracy of the alignment. However, this approach inevitably requires some approximation of the shape of the objects to interpolate data from different tilt angles. Moreover, distinct features will dominate the outcome of the cross-correlation, which might not be suitable for proper alignment. Therefore it is often useful to apply image filters (Fourier filter, Sobel filter, etc.) prior to the cross-correlation, to enhance features that are expected to correlate well. The major problem in using cross-correlation-based alignment procedures in combination with low-dose cryoimages of vitrified samples is the very low SNR of these images and the very weak contrast, since cross-correlation methods are typically very noise sensitive. Therefore they work well only for a limited class of specimens, e.g., high-contrast or paracrystalline objects, or in cases where fiducial markers are not present, or difficult to add, as seen in tilt series from cryosections. The marker alignment method is definitely more general because it can even cope with very low SNRs and is therefore more common in cryo-ET applications.

3.5.2 Reconstruction

Although the first practical formulation for applied tomography was achieved in the 1950s (Bracewell, 1956), Johan Radon first outlined the mathematical principles behind the technique in 1917 (Radon, 1917; English translation in Deans, 1983). The paper defines the Radon transform R as the mapping of a function $f(x, y)$, describing a real space object D , by the projection, or line integral, through f along all possible lines L :

$$Rf = \int_L f(x, y) ds$$

where ds is the unit length of L . A discrete sampling of the Radon transform is geometrically equivalent to the sampling of an experimen-

tal object by some form of transmitted signal: a projection. The consequence of such equivalence is that the reconstruction of an object $f(x, y)$ from projections Rf can be achieved by implementation of the inverse Radon transform. All conventional reconstruction algorithms are approximations, with varying accuracy, to this inverse transform. It is mathematically convenient to use polar coordinates (r, ϕ) , related to Cartesian axes for an explicit description of the Radon transform. The Radon transform operation converts the coordinates of the experimental data into "Radon space," where l is a line perpendicular to the projection direction and θ denotes the angle of the projection. Thus a point in real space (r, ϕ) becomes a line in Radon space (l, θ) with the relation $l = r \cdot \cos(\theta - \phi)$. This relationship between real space and Radon space allows a more explicit examination of the experimental situation. A single projection of a one-dimensional (1D) object, for example, a point in real space (a discrete sampling of the Radon transform) is a 2D line at constant θ in Radon space. Thus a series of projections at different angles will provide an increased sampling of the Radon space. Given a sufficient number of $(n - 1)$ -dimensional projections (of an n -dimensional object) from different views, an inverse Radon transform should reconstruct the n -dimensional object. However, in ET the experimental sampling (l, θ) is discrete (limited amount of projections in a restricted angular range), therefore the inversion will be imperfect and the problem of an undistorted and complete reconstruction, especially in ET, becomes evident: to achieve the best reconstruction of the object it is necessary to obtain as many projections as possible in an angular regime as large as possible.

In practice the reconstruction from projections is aided by the understanding of the relationship of a projection in real space and Fourier space. The "projection-slice theorem" states that a 2D projection at a given angle is a central section through the 3D Fourier transform of this object (Figure 7–19). If a series of projections is acquired at different tilt angles, each projection will correspond to part of an object's Fourier transform, thus sampling the object over the full range of frequencies in a central section. The shape of most objects will be described only partially by the frequencies in one section, but by taking multiple projections at different angles many sections will be sampled in Fourier space. This will describe the Fourier transform of an object in many directions, allowing a fuller description of an object in real space. In principle a sufficiently large number of projections taken over all angles will provide a complete description of the object. Therefore tomographic reconstruction is possible from an inverse Fourier transform of the superposition of a set of Fourier transformed projections: an approach known as direct Fourier reconstruction. This was the approach formulated by Bracewell (1956) and was used for the first tomographic reconstruction from electron micrographs (DeRosier and Klug, 1968). It is also known as "Fourier synthesis" and used for atomic structure determination by electron (Henderson et al., 1990) and X-ray crystallography (Perutz et al., 1960).

The direct interpretation of the projection-slice theorem would suggest a reconstruction algorithm based in Fourier space (Hoppe and

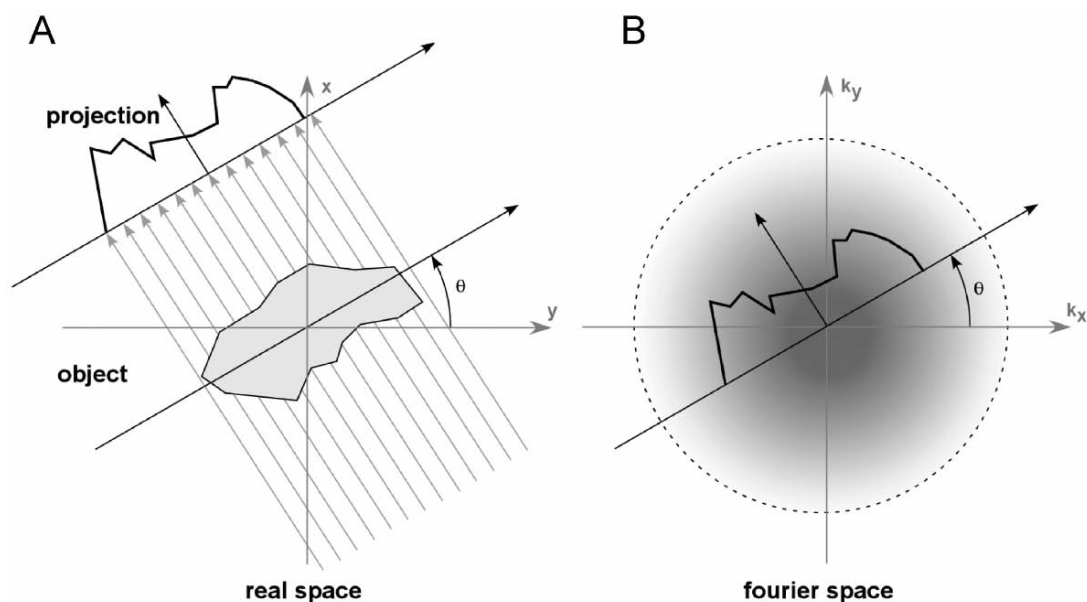


Figure 7-19. 'Projection-slice theorem'. a) An object is shown in real space (x , y) at the origin, and one of its projection images is presented formed by tilted parallel beams. b) The Fourier transform of the projection is a section through the origin of the Fourier space (k_x , k_y) tilted by θ .

Hegerl, 1980). Despite the fact that the first 3D reconstruction by DeRosier and Klug (1968) was carried out in Fourier space, it is common to use real-space-based reconstruction algorithms, because the practical implementation of Fourier-space reconstruction is not as simple as an inverse transform. The projection data are always sampled at discrete angles leaving regular gaps in Fourier space (Figure 7-20). But the inverse transform intrinsically requires a continuous function and therefore interpolation is required to fill the gaps in Fourier space (Crowther et al., 1970). However, the quality of a Fourier-space reconstruction is greatly affected by the type of interpolation implemented,

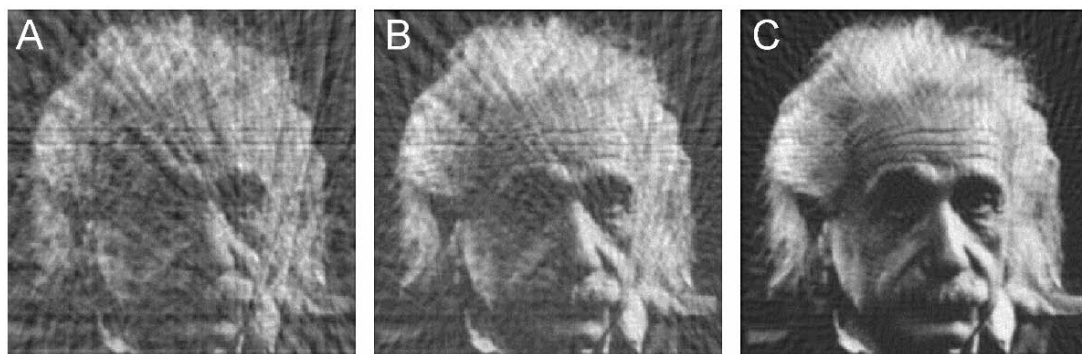


Figure 7-20. Illustration of the effect of sampling at discrete angles in a 2 dimensional example. a) For 18 projections over the full angular range at 10° increment. b) For 36 projections at 5° increment and c) for 72 projections at 2.5° increment which is almost identical to the original image.

as examined by Smith et al. (1973). Although elegant, Fourier reconstruction methods have the disadvantage of being computationally intensive and rather difficult to implement. Therefore they have been superseded by faster real-space alternatives since interpolations are easier to implement there.

By far the most widely utilized algorithm is weighted backprojection (Radermacher, 1992). Algebraic reconstruction techniques (ART) are also well established, despite the fact that they were subject to criticism in their beginnings (Crowther and Klug, 1971; for a detailed review of reconstruction algorithms we refer to Frank, 1996). The theory of backprojection relies on a simple principle: a point in space may be uniquely described by any three "rays" passing through that point. However, as the object increases in complexity more "rays" are required to describe it uniquely. A projection of an object is the inverse of such a "ray," and will describe some of the complexity of the object at hand. Therefore by inverting the projection, smearing out the projection into an object space at the angle of projection, a "ray" is generated that will uniquely describe an object in the projection direction, a process known as backprojection. With sufficient projections, from different angles, the superposition of all the backprojected "rays" will reproduce the shape of the original object—a reconstruction technique known as direct backprojection. Direct backprojection is used for reconstruction in classical computer-assisted tomography (CT; Herman, 1980), and it was the technique used by Hart (1968) for the "polytropic montage." It was also mentioned as an alternative to Fourier methods by Crowther et al. (1970).

However, reconstructions made by direct backprojection are exceptionally blurred, showing distinct enhancement of low frequencies, while fine spatial details are reconstructed poorly. This problem is an effect of the uneven sampling of spatial frequencies in the series of original projections (Figure 7-21). In 2D each of the acquired projections is a line intersecting the center of the Fourier space. Assuming a regular sampling of Fourier space in each projection, far more sampling points are located in the center of Fourier space than in the periphery. The outcome of this is an "undersampling" of the high spatial frequencies and an "oversampling" of the low spatial frequencies of the object, which will subsequently result in a "blurred" reconstruction. Therefore, to remove the blurring in real space and to restore the correct "frequency balance" in Fourier space, it is necessary to apply weighting schemes. The weighted backprojection consists of two steps: First the aligned projections are weighted in Fourier space by a function that characterizes the different sampling density in Fourier space. Assuming an infinitely small tilt increment the weighting has to be proportional to the amount of the spatial frequency orthogonal to the tilt axis ("analytical" weighting); more precise weighting ("exact" weighting) has to consider the shape function of the object to approximate the sampling density in Fourier space, normally using a *sine* function, which is retained only up to its first zero crossing (Harauz and van Heel, 1986; Hoppe and Hegerl, 1980). The second step is the backprojection of the weighted micrographs into the reconstruction body, most frequently by trilinear interpolation.

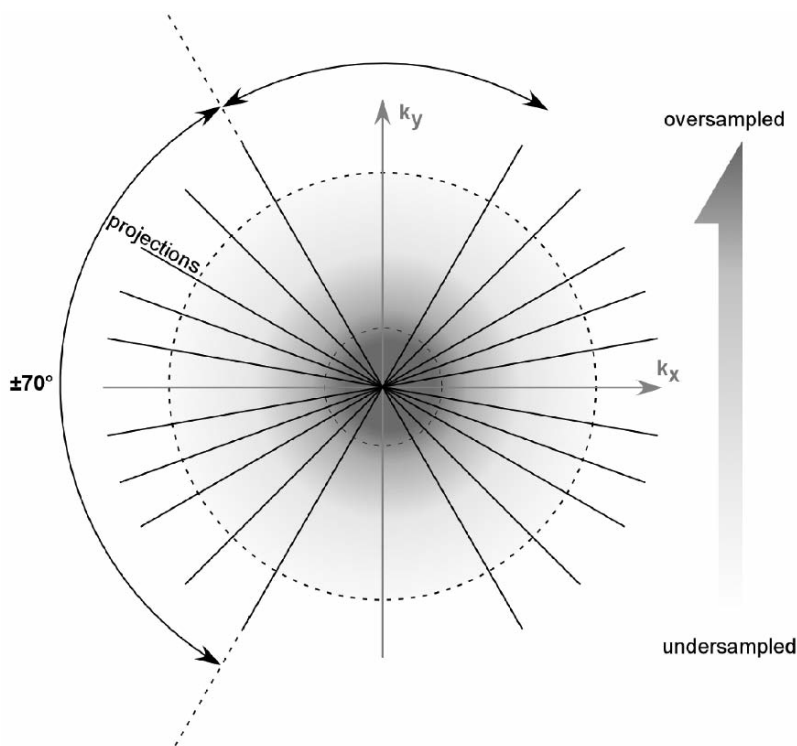


Figure 7-21. Illustration of the sampling 'density' problem in Fourier space. The large number of sampling points at low frequencies (darker area) are in contrast to the periphery (high frequencies), very only few sampling points are included in the single projections (brighter area). This sampling imbalance will result in 'blurred' reconstructions, which can be overcome by weighting schemes.

ART formulates the Radon transform as a system of algebraic equations. The inversion of this algebraic system is performed iteratively until self-consistency is achieved. The solutions of both reconstruction methods should be similar; however, ART determines the correct weighting parameters automatically whereas weighted backprojection requires a priori knowledge about the specimen to provide a correct weighting parameter for exact weighting. Despite this fact weighted backprojection is used in most software packages for 3D reconstruction due to its computational speed. However, the development of parallelized implementations as described in Fernandez et al. (2002) might establish ART as an alternative.

Several methods have been introduced to increase the quality of tomographic reconstructions, i.e., improving the SNR of the data and filling unsampled regions of the object's Fourier space with consistent data. All of these methods incorporate additional constraints to restrict the solution of the reconstruction. Application of solvent flattening in single particle analysis (van Heel, 2000), projection onto convex sets (Carazo and Carrascosa, 1987), or the Gerchberg-Fienup algorithm (Spence et al., 2003) in electron crystallography can improve the

resulting reconstructions considerably. Despite this fact, in most cases collection of sufficient data is preferred since confining the basis set of the reconstruction can cause artifacts. In cryo-ET there is typically no a priori information available; furthermore the data are not fully consistent because some features occurring at high tilt angles are not present in low tilt angles, i.e., the “unit cell is only partially defined” (Hoppe and Hegerl, 1980). Therefore, refinement techniques are commonly not used in cryo-ET, despite unproven claims of resolution improvement close to the subnanometer regime (Sandin et al., 2004). To extend the interpretability of tomograms beyond the first zero of the CTF it would be necessary to correct for the effects of the CTF as is done in single-particle analysis. Corrections based on exit-wave reconstruction, which rely on very few projections, have been proposed to extend the achievable resolution of tomograms (Han et al., 1996). For practical implementation of a CTF correction for tomography the lateral focus gradient, particularly for high tilt angles, needs to be incorporated, which make exit-front reconstructions considerably more complicated. A simpler restoration method, which requires only one micrograph per tilt and incorporates the lateral focus gradient, has been realized recently (Winkler and Taylor, 2003). However, this restoration method is designed primarily for thin specimens. In cryo-ET, CTF corrections have not been established yet. This is primarily due to the fact that the SNR of the individual micrographs is still too low, particularly because of the poor MTF of the CCD cameras, which prohibits a precise determination of the CTF.

3.5.3 *Visualization and Image Analysis*

The interpretation of tomograms at the ultrastructural level requires decomposition of a tomogram into its structural components, e.g., the segmentation of intracellular membranes or the assignment of organelles. Currently, a manual assignment of features is commonly used because human anticipation is still superior in most cases to available segmentation algorithms, although machine-based segmentation is in principle more objective (Frangakis and Hegerl, 2002; Volkmann, 2002). Instead of addressing the ultrastructure (Ladinsky et al., 1999), cryo-ET provides the basis for interpreting tomograms even at the molecular level. However, the analysis and 3D visualization are hampered by a very low SNR. To increase the SNR, so-called denoising algorithms have been developed (reviewed in Frangakis and Foerster, 2004). These algorithms aim to identify noise and remove it from the tomogram, but in practice, they also remove a certain fraction of the signal, resulting in data with reduced information but higher SNR. The simplest denoising techniques used diverse linear filtering operations, such as a simple low-pass filtering in Fourier space. Better signal preservation can be achieved by nonlinear filtering algorithms, such as nonlinear anisotropic diffusion (NAD) (Frangakis and Hegerl, 2001; Fernandez and Li, 2003) or bilateral filtering (Jiang et al., 2003b). NAD is particularly useful for the visualization of the ultrastructural features because it can enhance features such as membranes. In addition, these filters preserve the signal, without major alterations, which is especially suit-

able for visualization at the molecular resolution level. Denoising based on the wavelet transformations (Stoschek and Hegerl, 1997) maintains the high-resolution content, but requires immense computational efforts, thus currently favoring the former approaches.

CET enables us to resolve large macromolecular complexes such as the 26S proteasome inside intact cells (Medalia et al., 2002). High-resolution structures of numerous macromolecules are available from X-ray crystallography or nuclear magnetic resonance (NMR). The purpose of many ongoing structural genomics projects is to generate a comprehensive library of protein structures. Therefore, molecular recognition is the task of locating a priori known structures in the context of cells or other complex biological samples rather than clarifying structurally novel features, e.g., as intended with denoising methods. Mapping of macromolecules on the basis of their structural signature requires the quantitative comparison of tomogram data with a library of macromolecular structures. Ideally, the search of tomograms should be exhaustive and would reveal a cellular atlas of resolvable macromolecules. Simulations and experiments with “phantom cells,” i.e., liposomes encapsulating macromolecules, indicated that such an approach is feasible (Boehm et al., 2000; Frangakis et al., 2002). The information addressing the spatial relationship of different complexes fosters and complements other proteomic methods and will be indispensable for structural proteomic approaches (Sali et al., 2003). The most common molecular detection algorithm is a locally normalized, matched filter, introduced in a different context (Roseman, 2000, 2003). It was modified to account for the missing-wedge effect (Frangakis et al., 2002) and applied to cryotomograms (Rath et al., 2003). These applications demonstrated that it is feasible to identify large macromolecular complexes (>500 kDa) within cryotomograms with high fidelity. Furthermore, the high computational demand of template matching is significantly reduced by parallelization. The information an individually resolved macromolecule contains is limited because of the low electron dose and the missing-wedge effect. Averaging techniques aim to overcome the dose limitation of resolution by explicit averaging of different reconstructions from the same particle. Iterative algorithms are used to align subtomograms of arbitrarily oriented copies of a particle and obtain a consistent average. Medium resolution (2–3 nm) structures of the thermosome and tricorn could be obtained from cryotomograms of purified complexes (Walz et al., 1997b; Nitsch et al., 1998). Generally, the resolution from cryo-ET is inferior to conventional single-particle reconstructions, which use 2D projections of particles recorded on film to obtain 3D structures. Considering the aforementioned dose fractionation theorem (see Section 3.2), the higher information content of tomograms should favor averaging of tomograms over the averaging from projections. However, the low resolution of CCD cameras compared with that of film, the alignment error of the micrographs prior to 3D reconstruction, and the current inability to correct reliably for the CTF argue against averaging of tomograms. Nevertheless, cryo-ET can provide medium-resolution structures of protein complexes without using extensive purification

procedures (Foerster et al., 2005). Furthermore, structures obtained by cryo-ET can be used as starting models that are refined by single-particle techniques or can be combined with other higher resolution structural techniques to provide comprehensive descriptions of molecular complexes.

3.5.4 Identification Strategies of Macromolecular Complexes

A very common approach to the task of locating features of known structure in an input scene is “matched filtering.” Matched filtering is basically the computation of a suitable cross-correlation function (CCF) of the template, a macromolecule of known structure, and the input scene, the cryo-electron tomogram (Figure 7-22). Cross-correlation functions are straightforward to compute as a function of spatial parameters r , since they can be calculated using the fast Fourier transform (FFT) (F):

$$CCF_r = \sum_{r'} I_{r+r'} \cdot T_{r'} = F^{-1}[F(I) \cdot F(T)^*].$$

The gray values of the input scene are denoted I_r and the template's gray values are T_r . The maxima of this cross-correlation function should

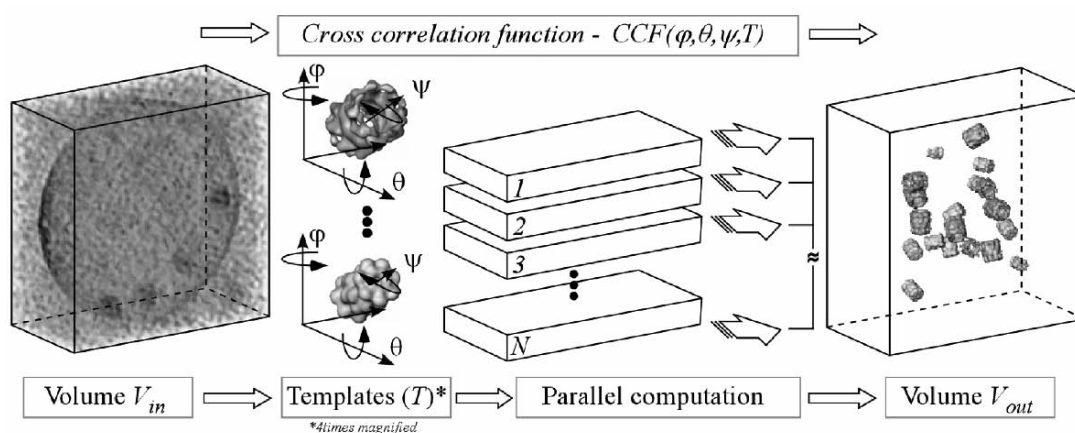


Figure 7-22. Detection and identification of individual macromolecules in cellular tomograms is based on their structural signature. Because of the crowded nature of the cytoplasm and ‘contamination’ with noise, an interactive segmentation and feature extraction is not feasible. It requires sophisticated pattern recognition techniques to exploit the information contained in the tomograms. A volume rendered presentation of a 3D image is presented on the left. Even though some high-density features may be visible, an unambiguous identification of individual structures would be difficult if not impossible given the residual noise. An approach, which has proven to work, is based on template matching. Templates of the macromolecules under scrutiny are obtained by a high- or medium resolution technique (X-ray crystallography, NMR, electron crystallography or single particle analysis). These templates (4 times magnified in this figure; 20S proteasome and thermosome) are then used to search the entire volume of the tomogram systematically for matching structures by three-dimensional cross-correlation and the result is refined by multivariate statistical analysis. In principle the 3D image has to be scanned for all possible Eulerian angles ϕ, ψ and θ around three different axes, with templates of all the different protein structures one is interested in e.g. the thermosome (blue) and proteasome (yellow). The search procedure is computationally very demanding but can be parallelized with respect to the different angular combinations in a highly efficient manner. Finally, the position and orientation of the different complexes can be mapped directly in the 3D image. (See color plate.)

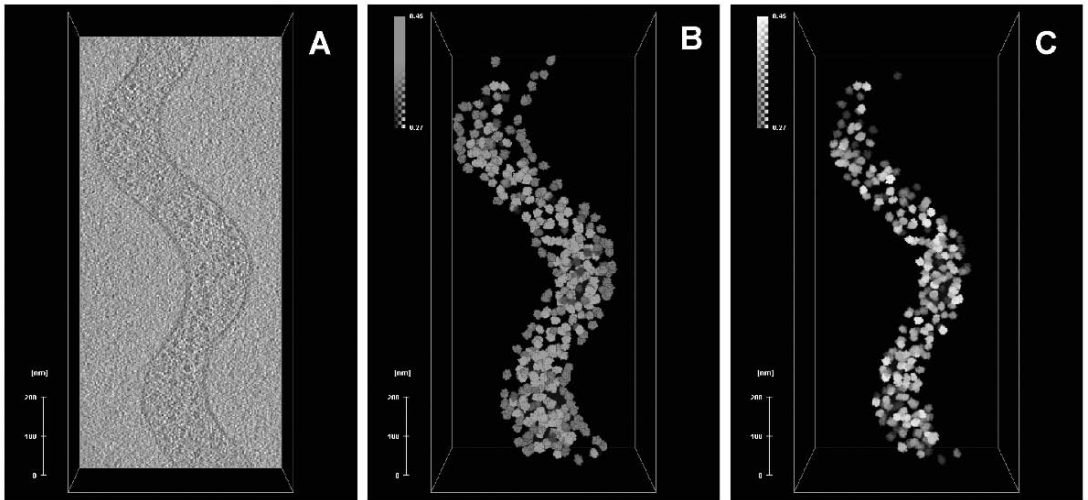


Figure 7-23. Mapping ribosomes in an intact *S. melliferum* cell. a) x - y slice from the corresponding tomogram. Image analysis of this portion of the cell is displayed in subsequent panels. b) Locations and orientation of all ribosomes detected by template matching. Each 70S ribosome is represented by the averaged density (see Fig. 26) derived from the tomogram. The colour coding indicates the detection fidelity; green is high, yellow is intermediate, red is low and probably represents false positives. The brightness of the colour corresponds to correlation peaks heights. c) Final ribosome atlas after removal of putative false positives (images courtesy of J. Ortiz, MPI of Biochemistry, Martinsried, Germany; Ortiz et al., 2006). (See color plate.)

mark the positions of the features in the input scene. Three different CCFs have been introduced in the field of 2D electron microscopy: conventional XCF, phase-only filtering (POF), (Horner and Gianino, 1964), and mutual cross-correlation function (MCF; van Heel et al., 1992). In all those methods a cross-correlation function is computed: The XCF computes the unmodified cross-correlation function of template and micrograph whereas the POF and MCF vary the moduli of the template to sharpen the cross-correlation peaks. In electron microscopy especially, the low frequencies carry an extraordinarily strong signal, which can be misleading (van Heel et al., 1992). The strength of the low-frequency contributions can differ considerably within electron micrographs of amorphous specimens, such as vitrified specimens, since these are typically nonhomogeneous. The POF sets the moduli of template and input scene to one while the MCF replaces the moduli by their square roots to reduce the influence of the very low-frequency amplitudes systematically.

Another way of taking into account the strong local variations of the image contrast has been introduced recently (Roseman, 2000). By normalizing the XCF to the local cross-correlation function (LCF) variance, the efficiency of locating local features such as macromolecules within a micrograph can be increased significantly. Typically, the normalization region should correspond to the shape of the molecule. This calculation of the local variance does not lengthen the computing time unreasonably, since it can be calculated very efficiently in Fourier space too (Figures 7-23 and 7-24).

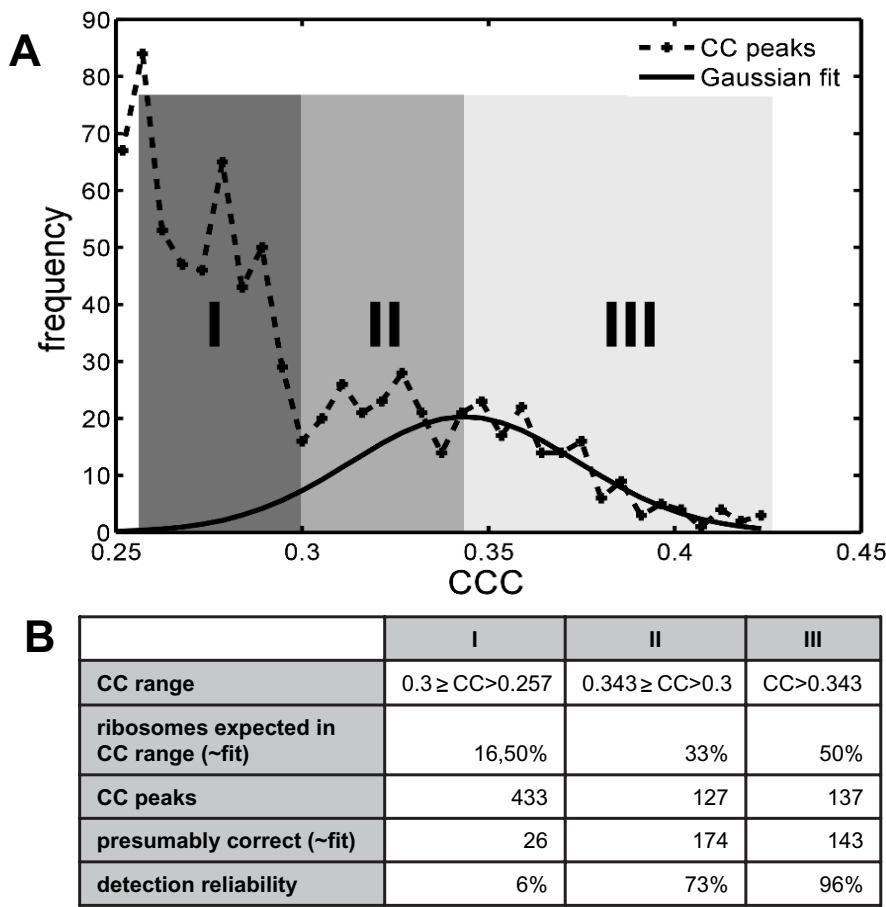


Figure 7–24. a) Histogram of experimentally obtained CCCs (dashed line). Ideally, the histogram should be Gaussian, reflecting the variances in SNR of individual ribosomes in the tomogram, and thresholding of the CCC should allow discriminating between correctly identified ones and false positives. Non-specific correlation with high contrast features such as the cell membrane, leads to a relative large number of false positives. For an objective setting of thresholds, the Gaussian fit (continuous line) was divided into three probability sectors larger than the mean CCC value (I, dark grey); lower than the mean value up to one standard deviation (II, intermediate grey); lower than the mean value and between one and two standard deviations (III, light grey). The spreadsheet in b) illustrates the Gaussian fit of the experimental distribution of CCCs and the fidelity of the experimental detection of 70S ribosomes. About 80% of the particles are detected with more than 85% reliability (images courtesy of J. Ortiz, MPI of Biochemistry, Martinsried, Germany; Ortiz et al., 2006).

To evaluate which filter is best for the task of locating macromolecules, some theoretical considerations are useful. The conventional XCF has been shown to be optimal with respect to minimization of the location error (Kumar et al., 1992) and high SNR of the resulting function (VanderLugt, 1964). However, use of the XCF for locating small features within micrographs or cryoelectron tomograms yields intolerably high rates of false positives; the XCF tends to locate high-contrast features rather than low-contrast macromolecules. The same typically holds for the MCF, whereas the POF overamplifies noise, which makes

it unfavorable. Therefore, local normalization is appropriate. However, regarding the SNR of the resulting correlation function, the LCF is inevitably worse than XCF or MCF; confinement of the correlation to N voxels raises the noise level, which should be proportional to $N^{-1/2}$. The size of the normalization region should correspond to the envelope of the template under scrutiny. Choosing the local region larger will increase the SNR of the filter output at the cost of increasing the chance of false positives simultaneously, in particular if particles tend to be very dense within the input scene. Whenever justified circular (2D) or spherical (3D) normalization regions (Θ) will be used because this accelerates the computation considerably compared to approaches using asymmetric regions. If the template has one rotational degree of freedom perpendicular to the specimen plane within the input scene, the problem of matching a 2D template in a 2D image can be formulated as determining the maxima of a cross-correlation function as a function of spatial variable r and rotation angle φ :

$$XCF_{r,\varphi} = \sum_{r'} I_{r+r'} \cdot R_{\varphi}(T_{r'}) \cdot$$

Here R_{φ} denotes the rotation operator. By expanding the template in circular harmonics, exhaustive scanning of the polar angle φ can be avoided (Kunath and Sack-Kongehl, 1989). In 3D, the problem of finding a macromolecule in a tomogram that matches a 3D template by cross-correlation becomes six-dimensional: Using the Euler angles φ , ψ and ϑ to describe the orientation, the maxima of the $CCF_{r,\varphi\psi\vartheta}$ are to be determined. Again, the translational dependence can be covered using FFT, but it is normally necessary to perform an explicit scanning in angular space to cover those parameters. An exhaustive search is computationally very time demanding. However, because of the emerging facilities of parallel computing this task can easily be faced: Since the CCF can be computed individually for different orientations, the entire calculation can be distributed over different processors with hardly any administrative overlap (Frangakis and Hegerl, 2002; Rath et al., 2003).

In cryo-ET the partial sampling of the object's Fourier space, the "missing wedge," has to be considered further. This can be done by confining the correlation to the sampled region in Fourier space or equivalently by convoluting the object with the point-spread function (Frangakis and Hegerl, 2002). In the framework of the locally normalized CCF the missing wedge affects the local normalization. The missing wedge can be taken into account in Fourier space by multiplication of the template with a suitable binary function or equivalently by convoluting with a suitable PSF in real space. In this notation, the confined cross-correlation function can be defined as

$$CCF_{r,\varphi\psi\vartheta} = \frac{\sum_{r'} I_{r+r'} \cdot (T_{r',\varphi\psi\vartheta} \otimes PSF_{\varphi\psi\vartheta} \cdot \Theta_{r'})}{\sum_{r'} \sqrt{\Theta_{r'} \cdot (I_{r+r'} - \bar{I}_{r'})^2}}$$

where $\bar{I}_{r'}$ represents the local mean value within Θ . A natural way of avoiding the limitation of beam damage in tomography is averaging of copies of structural elements occurring in a tomogram. Saxberg and

Saxton (1981) already concluded from their results on resolution limitation in electron tomography due to beam damage that averaging is a practical way of increasing the effective electron doses and should be considered to achieve high resolutions. The importance of those techniques to derive information on the molecular level was reemphasized recently (Steven and Aebi, 2003).

According to the dose fractionation theorem a tomogram describes a volume element with the same statistical significance as a projection with the same dose (Hegerl and Hoppe, 1976; McEwen et al., 1995). Therefore, tomograms should be just as well suited to posterior averaging procedures for obtaining high-resolution information as compared to single-particle averaging. In practice, both approaches have their pros and cons, which are discussed below.

Averaging of particles from tomograms can be performed in the same way as single-particle averaging. Projection matching is crucial in these techniques: The individual particle projections are iteratively aligned to projections of a reference from different angles by means of cross-correlation. The reference is thus the average of the particles using the orientations and displacements of the previous iteration. The iterative refinement of orientation and displacements improves the resolution of the average until convergence is achieved. The analogous procedure in cryo-ET could be termed “tomogram matching.” Subtomograms containing the roughly centered particles of interest need to be extracted from the entire tomogram and aligned using procedures as sketched in Figure 7–25. The aim of those tomogram matching procedures is to optimize the orientations, i.e., the Euler angles ϕ , ψ , and ϑ , and the displacement shifts of the individual particles. A first realization of this approach was reported by Walz et al. (1997). Application of this implementation on tomograms of purified protein complexes resulted in the generation of a pseudo-atomic map of the thermosome in the open conformation (Nitsch et al., 1998).

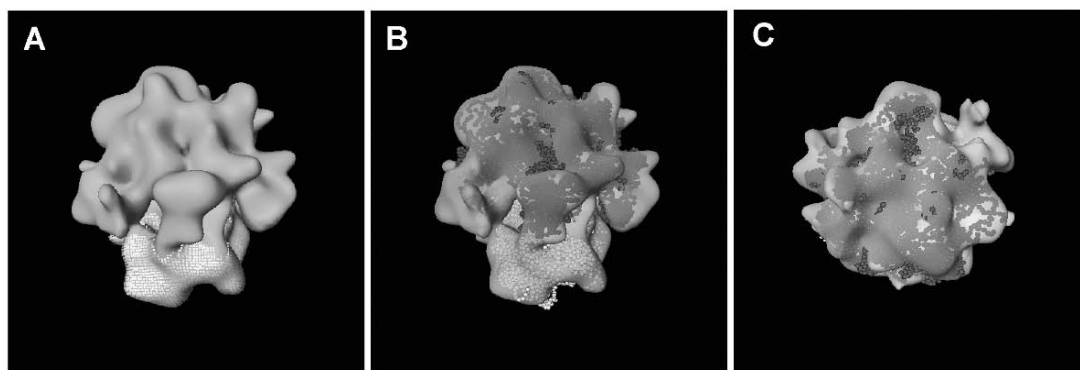


Figure 7–25. a) Averaged structure of the 70S ribosome derived from the dataset above (average from 300 individual particles to ~4nm resolution). The map highlights the 30S subunit in yellow. b) Docking of high resolution structures of the 70S ribosome into the map shown in A. c) “Crown view” of B (images courtesy of J. Ortiz, MPI of Biochemistry, Martinsried, Germany). (See color plate.)

However, cryo-ET is conventionally not used for this purpose for several major reasons: (1) It is much easier to acquire the data by spending the entire dose on only one projection; this makes acquisition of large amounts of individual molecules much easier. (2) The SNR of a projection of a particle is better than the SNR of a tomogram of the same particle acquired with the same electron dose. (3) There is only one level of alignment: In tomography the projections need to be aligned to a common origin prior to aligning the individual particles with respect to each other. Both steps inevitably introduce errors that sum up.

But averaging of cryo-ET data can offer some advantages that are worth the additional efforts in some cases. The most important reason is the ability to image nonpurified samples (Figure 7–26). CET is currently the only technique that can be used for quaternary structure determination of fragile or even transient complexes. The degree of alteration that biological macromolecules undergo under physiological conditions is largely undetermined since there was no imaging technique that could resolve them in the context of a cell. Apart from this, averaging of subtomograms also offers one principal advantage compared to averaging from projections: It is fundamentally easier to determine the orientations of an individual copy from 3D data than from projection data.

4 Perspectives: New Strategies and Developments

The achievements made by cryo-ET over the past decade have proven the possibilities and the feasibility of this technique for quasi *in vivo* studies of the ultrastructure and larger supramolecular assemblies within whole cells. However, based on the initial developments in cryo-ET, major improvements in instrumentation and sample preparation have to be made in the future to exploit its full potential.

The ultimate goal in structural biology is to investigate the structure–function relationship of molecular complexes and supramolecular assemblies in their native environment, e.g., in large cells or even tissue, across several dimensions, from the micrometer level to the subnanometer level, and if possible within one single experiment, to realize the complete view into the inner space of cells and its constituents. Even at the present practical level cryoelectron tomograms of organelles and cells contain an imposing amount of information. They are, essentially, 3D images of entire proteomes, and they should ultimately enable us to map the spatial relationships of the full complement of macromolecules in an unperturbed cellular context. However, it is obvious that new strategies in sample preparation, advancements in instrumentation, and innovative image analysis techniques are needed to make this dream come true, at least to some extent.

In principle there are three major routes to increase the image quality, and especially the 3D image quality, which are intrinsically linked to each other: (1) the technological developments comprising

advanced electron optics, highly resolved high-sensitive detection, improved tilting devices, and the geometry of the sample holder, (2) the sample preparation and thus the quality of the sample, and (3) the design of new software algorithms and acquisition schemes to guarantee a quantitative, objective, and exhaustive analysis of the recorded data.

TEM or EM is still a very young method if compared to light microscopy. While the latter was developed and improved over centuries, EM has a history of less than 80 years. Microscopes can now easily reach a resolution in 2D of less than 1 Å, clearly orders of magnitude better than what light microscopy can offer. However, the electron optical system, determined mainly by the objective lens system and characterized by the spherical and chromatic aberration coefficients, is far from being ideal and is still inferior to light optical devices. Simplified, the performance of present objective lens systems in EM can be described as an attempt to focus through the bottom of a champagne bottle. This way, the quality of the recorded micrograph in EM suffers greatly from the severe shortcomings of the objective lens system, e.g., the higher the spherical aberration coefficient C_s , the better the resulting image contrast, but the lower the resolution. However, the spherical aberration coefficient is linked directly to the “spacing” inside the objective lens system and thus determines the maximum achievable tilt range for an angular acquisition. Recent developments in C_s correction technology for TEM (and even for SEM) demonstrated the possibilities and the final image improvement, especially in material science studies. However, EM investigations of biological samples at cryotemperatures are characterized by a very low image contrast, due to the very weak scatterers, mainly low Z elements, such as carbon, oxygen, and hydrogen.

Thus high defocus values are more or less mandatory, to increase the image contrast. Unfortunately, by tuning the objective lens to very low defocus values of a couple of micrometers, low-frequency information is enhanced, while high-frequency information is almost completely obscured if not totally lost. Thus C_s correction without any additional equipment is not an option in any cryo-EM study, because the actual improvement in image quality is restricted to a defocus regime in the range of a couple of tenths of nanometers, thus orders of magnitude higher than what would be needed for a high-contrast (low-dose) cryoimage. Nevertheless, it might be possible to increase the contrast without defocusing the objective lens, by using so-called phase plates in combination with C_s correctors. The benefit of C_s correction would then become clearly visible, because even high-frequency information would be accessible. Clearly the design and the application of phase plates are not new, but the technological problems in former times could not be addressed to enable their routine use. Two of the major problems were contamination and stability. While placed in the back focal plane, very close to the sample, contamination caused charging and drift problems. Moreover, because the stability was low, it was mandatory to readjust the phase plate in respect to the central beam, which, in terms of required accuracy, is

possible today only with piezo-controlled stages. However, present achievements and possibilities in nanostructuring and nanotechnology made it possible to utilize the already existing designs of phase plates, addressing the aforementioned problems, and place them in a normal EM environment.

The prospects are good and in the near future experiments will be done to show not only the feasibility of this approach, but also the final image improvement if, for example, used in combination with C_s correction. However, the total sum of elastically scattered electrons will be reduced and thus the large fraction of inelastically scattered electrons has to be removed or even utilized for final image formation. "Removal" can be done with imaging energy filters, which are now routinely incorporated into the EM setup. They are especially helpful in studies in which thick cells or sections are used, because the larger the material layer to be penetrated by the electron beam, the larger the fraction of inelastically and multiple-scattered electrons. They will have a different energy, and thus they will be slightly out of focus, blurring the recorded image. However, if the chromatic electrons are removed, the image contrast can be definitely improved, but the number of elastically scattered electrons (forming the image) will be just a fraction of the total sum of electrons leaving the sample. In most cases the amount is only a half or a third of the total applied electron dose to the sample, which contributes to final image contrast. This way, the detection has to be very sensitive, so that literally every electron will count. Today, typically CCD cameras are used, which detect the signal indirectly and which add noise to the recorded image. Moreover the signal is spread over several pixels, decreasing the lateral resolution and thus the transfer of the recorded image. Direct detection would be beneficial and favorable, because every electron would contribute to the final signal, with no readout noise and the best possible lateral resolution. Developments will utilize detectors based on CMOS technology, which will guarantee a noise-free detection and, if back-thinned, a highly resolved recording. However, the radiation hardness of these devices, for 300-keV electrons, is at the moment insufficient for routine use in EM, but improvements are being made to increase the lifetime of these detection devices, so that they can be used almost continuously.

Since we cannot increase the total applied dose to the sample, not even by cooling to liquid helium temperature, the only option we do have is to increase the performance of the detection device and the electron optics. If, for example, an almost ideal detector, could be utilized, the dose in a recorded image can be as low as possible and thus the number of acquired projections in an angular regime can be as large as possible. This way the final resolution would be definitely improved. Moreover, if high-end electron optical systems are used, e.g., C_s or even C_c correctors (which would even utilize the fraction of inelastically scattered electrons for image formation) in combination with phase plates and energy filters, not only the low-frequency information, responsible for the final image contrast, but also the high-frequency information, containing the fine struc-

tural details, could be harvested. Moreover, with just one single 2D image, it would be possible to obtain phase and amplitude information separately.

In addition to the improved quality of the single 2D projection, in phase plate C_s -corrected imaging, ET is still hampered by an uneven sampling in Fourier space and thus by the fact that information in the direction of the tilt axis cannot be accessed. This missing information (missing wedge), due to the limited tilt range, can be decreased by utilizing dual-axis acquisition schemes to a missing pyramid. Clearly, with the 90° in-plane rotation of the sample and the acquisition of two-tilt series of the same sample area, it is possible to increase the amount of information by more than 20% and thus reduce the resulting artifacts related to single-axis-acquisition schemes. The resolution is not improved, however, it is definitely more isotropic. Especially for subsequent automatic detection of macromolecules and proteins within the tomographic volume, e.g., by template matching, errors (false-positives) can be minimized, and thus the quality of the detection can be vastly improved.

At the present resolution of 4–5 nm only very large complexes (ribosomes, 26S proteasomes) can be detected reliably within cryoelectron tomograms of whole cells. However, an improvement in resolution to 2 nm will allow us to detect medium sized complexes in the range of 200–400 kDa. Some of the problems in the interpretation of tomograms will disappear once a resolution of 2 nm is obtained. At the moment, with a resolution of 4–6 nm, docking of high-resolution structures to yield pseudo-atomic maps of molecular complexes is computationally demanding and time consuming. However, this procedure of template matching and docking will become easier if a resolution of 2 nm can be routinely obtained. The computational methods described will aid and complement the information provided by other approaches and information, such as localization, labeling studies, and the binding properties of molecules. While tomograms with a resolution of 2 nm are a realistic prospect, major technical innovations (see above) will be required to go beyond this.

CET of whole cells allows us to investigate the structure–function relationship of molecular complexes and supramolecular assemblies in their native environment. It thus results in a fundamental change in the way we approach biochemical processes that underlie and orchestrate higher cellular functions. In the past, molecular interactions were studied mostly in a collective manner, whereas now we have the tools to visualize the interactions between individual molecules in their unperturbed functional environments. Although they share common underlying principles, no two cells or organelles are identical, owing to the inherent stochasticity of biochemical processes in cells as well as their functional diversity. Therefore, it will be a major challenge to extract generic features from the maps, such as the modes of interaction between molecular species. The ultimate goal, the discovery of general rules that underlie cellular processes, has to go beyond observing qualitative features and has to be based on stringent analytical criteria (Lucic et al., 2005).

References

- Al-Amoudi, A., Dubochet, J., Gnaegi, H., Lueth, W. and Studer, D. (2003). An oscillating cryo-knife reduces cutting-induced deformation of vitreous ultra-thin sections. *J. Microsc.* **212**(1), 26–33.
- Al-Amoudi, A., Studer, D. and Dubochet, J. (2005). Cutting artefacts and cutting process in vitreous sections for cryo-electron microscopy. *J. Struct. Biol.* **150**(1), 109–121.
- Alberts, B. (1998). The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell* **92**, 291–294.
- Adrian, M., Dubochet, J., Lepault, J. and McDowell, A.W. (1984). Cryo-electron microscopy of viruses. *Nature* **308**, 32–36.
- Agard, D.A. (2005). Private communications.
- Angert, I., Majorovits, E. and Schroeder, R.R. (2000). Zero-loss image formation and modified contrast transfer theory in EFTEM. *Ultramicroscopy* **81**(3–4), 203–222.
- Barnard, D.P., Turner, J.N., Frank, J. and McEwen, B.F. (1992). A 360° single-axis tilt stage for the high-voltage electron microscope. *J. Microsc.* **167**, 39–48.
- Baumeister, W. and Typke, D. (1993). Electron crystallography of proteins: State of the art strategies for the future. *MSA Bull.* **23**, 11–19.
- Beck, M., Forster, F., Ecke, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baumeister, W. and Medalia, O. (2004). Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* **306**(5700), 1387–1390.
- *Beck, M., Lucic, V., Förster, F., Baumeister, W. and Medalia, O. (2007). Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* **449**, 611–615.
- Bernhard, W. (1965). Ultramicrotomie à basse température. *Ann. Biol.* **4**, 5–19.
- Boehm, J., Frangakis, A.S., Hegerl, R., Nickell, S., Typke, D. and Baumeister, W. (2000). Toward detecting and identifying macromolecules in a cellular context: Template matching applied to electron tomograms. *Proc. Natl. Acad. Sci. USA* **97**, 14245–14250.
- Boehm, J., Lamber, O., Frangakis, A.S., Letellier, L., Baumeister, W. and Rigaud, J.L. (2001). PhuA-mediated phage genome transfer into liposomes: A cryo-electron tomography study. *Curr. Biol.* **11**, 1168–1175.
- Bottcher, B., Wynne, S.A. and Crowther, R.A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**, 88–91.
- Bouwer, J.C., Mackey, M.R., Lawrence, A., Deerinck, T.J., Jones, Y.Z., Terada, M., Martone, M.E., Peltier, S. and Ellisman, M.H. (2004). Automated most-probable loss tomography of thick selectively stained biological specimens with quantitative measurement of resolution improvement. *J. Struc. Biol.* **148**(3), 297–306.
- Bracewell, R.N. (1956). Strip integration in radio astronomy. *Aust. J. Phys.* **9**, 198–217.
- Braunfeld, M.B., Koster, A.J., Sedat, J.W. and Agard, D.A. (1994). Cryo automated electron tomography—towards high resolution reconstructions of plastic embedded structures. *J. Microsc.* **172**(2), 75–84.
- Brenner, S. and Horne, R.W. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochim. Biophys. Acta* **34**, 60–71.
- Brink, H.A., Barfels, M.M.G., Burgner, R.P. and Edwards, B.N. (2003). A sub-50 meV spectrometer and energy filter for use in combination with 200 kV monochromated (S)TEMs. *Ultramicroscopy* **96**(3–4), 367–384.
- Carazo, J.M. and Carrascosa, J.L. (1987). Restoration of direct Fourier three-dimensional reconstructions of crystalline specimens by the method of convex projections. *J. Microsc.* **145**(2), 159–177.

- Coene, W.M.J., Thust, A., Op de Beeck, M. and Van Dyck, D. (1996). Maximum-likelihood method for focus-variation image reconstruction in high resolution transmission electron microscopy. *Ultramicroscopy* **64**(1–4), 109–135.
- Comolli, L.R. and Downing, K. (2005). Dose tolerance at helium and nitrogen temperature for whole cell electron tomography. *J. Struct. Biol.* **152**, 149–156.
- Conway, J.F., Trus, B.L., Booy, F.P., Newcomb, W.W., Brown, J.C. and Steven, A.C. (1993). The effects of radiation damage on the structure of frozen hydrated HSV-1 capsids. *J. Struct. Biol.* **111**(3), 222–233.
- Cormack, A.M. (1963). Representation of a function by its line integrals, with some radiological applications. *J. Appl. Phys.* **34**(1530), 2722–2727.
- Cormack, A.M. (1980). Early two-dimensional reconstruction and recent topics stemming from it. Nobel lecture 1979. *Medi. Phys.* **7**(4), 277–282.
- Crowther, R.A. and Klug, A. (1971). ART and science or conditions for three-dimensional reconstruction from electron microscope images. *J. Theor. Biol.* **32**, 199–203.
- Crowther, R.A., DeRosier, D.J. and Klug, A. (1970). The reconstruction of a three-dimensional structure from its projections and its applications to electron microscopy. *Proc. R. Soc. London* **317**, 319–340.
- Danev, R. and Nagayama, K. (2001). Complex observation in electron microscopy. II. Direct visualization of phases and amplitudes of exit wave functions. *J. Phys. Soc. Jpn.* **70**(3), 696–702.
- Deans, S.R. (1983). *The Radon Transform and Some of Its Applications* (John Wiley & Sons, New York).
- De Carlo, S., Adrian, M., Kalin, P., Mayer, J.M. and Dubochet, J. (1999). Unexpected property of trehalose as observed by cryo-electron microscopy. *J. Microsc.* **196**(1), 40–45.
- DeRosier, D.J. and Klug, A. (1968). Reconstruction of three dimensional structures from electron micrographs. *Nature* **217**, 130–134.
- Dierksen, K., Typke, D., Hegerl, R., Koster, A.J. and Baumeister, W. (1992). Towards automated electron tomography. *Ultramicroscopy* **40**, 71–87.
- Dierksen, K., Typke, D., Hegerl, R. and Baumeister, W. (1993). Towards automatic electron tomography. II. Implementation of autofocus and low-dose procedures. *Ultramicroscopy* **49**(1–4), 109–120.
- Dierksen, K., Typke, D., Hegerl, R., Walz, J., Sackmann, E. and Baumeister, W. (1995). Three-dimensional structure of lipid vesicles embedded in vitreous ice and investigated by automated electron tomography. *Biophys. J.* **68**, 1416–1422.
- Downing, K.H. and Hendrickson, F.M. (1999). Performance of a 2k CCD camera designed for electron crystallography at 400 kV. *Ultramicroscopy* **75**(4), 215–233.
- Downing, K.H. and Mooney, P.E. (2004). CCD camera with electron decelerator for intermediate voltage electron microscopy. *Microsc. Microanal.* **10**(Suppl. 2), 1378–1379.
- Downing, K.H., Favia, P. and Mooney, P.E. (2000). Electron decelerator for improved CCD performance in intermediate voltage electron microscopy. *Proceedings of the 58th Annual Meeting of the Microscopy Society of America*, 734–735.
- Dubochet, J. and McDowell, A.W. (1981). Vitrification of pure water for electron microscopy. *J. Microsc.* **124**, RP3–4.
- Dubochet, J., Adrian, M., Chang, J.J., Homo, J.-C., Lepault, J., McDowell, A.W. and Schultz, P. (1988). Cryo-electron microscopy of vitrified sections. *Quart. Rev. Biophys.* **21**(2), 129–228.
- Egerton, R.F. (1996). *Electron Energy-Loss Spectroscopy in the Electron Microscope* (Plenum Press, New York).

- Ellis, R.J. (2001). Macromolecular crowding: Obvious but underappreciated. *Trends Biochem. Sci.* **26**(10), 597–604.
- Ellis, R.J. and Minton, A.P. (2003). Join the crowd. *Nature* **425**, 27–28.
- Ermantraut, E., Wohlfart, K. and Tichelaar, W. (1998). Perforated support foils with pre-defined hole size, shape and arrangement. *Ultramicroscopy* **74**(1–2), 75–81.
- Evans, D.A., Allport, P.P., Casse, G., Faruqi, A.R., Gallop, B., Henderson, R., Prydderch, M., Turchetta, R., Tyndel, M., Velthuis, J., Villani, G. and Waltham, N. (2005). CMOS active pixel sensors for ionising radiation. *Nucl. Instrum. Methods Phys. Res. Sect. A—Accel. Spectrom. Detect. Assoc. Equip.* **546**(1–2), 281–285.
- Fan, G.Y. and Ellisman, M.H. (1996). Optimization of thin-foil based phosphor screens for CCD imaging in TEM in the voltage range of 80–400 kV. *Ultramicroscopy* **66**(1–2), 11–19.
- Fan, G.Y. and Ellismann, M.H. (2000). Digital imaging in transmission electron microscopy. *J. Microsc.* **200**(1), 1–13.
- Fan, G.Y., Datte, P., Beuville, E., Beche, J.F., Millaud, J., Downing, K.H., Burkard, F.T., Ellisman M.H. and Xuong, N.H. (1998). ASIC-based event-driven 2D digital electron counter for TEM imaging. *Ultramicroscopy* **70**(3), 107–113.
- Fan, G.Y., Peltier, S., Lamont, S., Dunkelberger, D.G., Burke, B.E. and Ellisman, M.H. (2000). Multiport-readout frame-transfer 5 megapixel CCD imaging system for TEM applications. *Ultramicroscopy* **84**(1–2), 75–84.
- Faruqi, A.R. and Subramaniam, S. (2000). CCD detectors in high-resolution biological electron microscopy. *Quart. Rev. Biophys.* **33**(1), 1–27.
- Faruqi, A.R., Cattermole, D.M. and Raeburn, C. (2003a). Direct electron detection methods in electron microscopy. *Nucl. Instrum. Methods Phys. Res. Sect. A—Accel. Spectrom. Detect. Assoc. Equip.* **513**(1–2), 317–321.
- Faruqi, A.R., Cattermole, D.M., Henderson, R., Mikulec, B. and Raeburn, C. (2003b). Evaluation of a hybrid pixel detector for electron microscopy. *Ultramicroscopy* **94**(3–4), 263–276.
- FEI Company, Hillsboro, OR and Eindhoven, Netherlands (www.feicompany.com).
- Fernandez, J.J. and Li, S. (2003). An improved algorithm for anisotropic nonlinear diffusion for denoising cryo-tomograms. *J. Struct. Biol.* **144**(1–2), 152–161.
- Fernandez, J.J., Lawrence, A.F., Roca, J., Garcia, I., Ellisman, M.H. and Carazo, J.M. (2002). High-performance electron tomography of complex biological specimens. *J. Struct. Biol.* **138**(1–2), 6–20.
- Feya, B. and Aeby, U. (1999). Determination of the inelastic mean free path of electrons in vitrified ice for on-line thickness measurements by zero-loss imaging. *J. Microsc.* **193**, 15–19.
- Fischione Instruments, Inc., Export, PA (www.fischione.com).
- Forster, F., Medalia, O., Zauberman, N., Baumeister, W. and Fass, D. (2005). Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proc. Natl. Acad. Sci. USA* **102**(13), 4729–4734.
- Frangakis, A.S. and Hegerl, R. (2001). Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. *J. Struct. Biol.* **135**(3), 239–250.
- Frangakis, A.S. and Hegerl, R. (2002). Segmentation of two- and three-dimensional data from electron microscopy using eigenvector analysis. *J. Struct. Biol.* **138**(1–2), 105–113.
- Frangakis, A.S. and Foerster, F. (2004). Computational exploration of structural information from cryo-electron tomograms. *Curr. Opin. Struct. Biol.* **14**, 1–7.
- Frangakis, A.S., Böhm, J., Förster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R. and Baumeister, W. (2002). Identification of macromolecular complexes

- in cryoelectron tomograms of phantom cells. *Proc. Natl. Acad. Sci. USA* **99**(22), 14153–14158.
- Frank, J. (1973). The envelope of electron-microscopic transfer functions for partially coherent illumination. *Optik* **38**, 519–539.
- Frank, J. (1992). *Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope* (Plenum Press, New York).
- Frank, J. (1996). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies* (Academic Press, London).
- Frank, J., Zhu, J., Penczeck, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R.K. and Agrawal, R.K. (1995). A model of protein synthesis based on cryo-electron microscopy of the E. Coli ribosome. *Nature* **376**, 441–444.
- Frederik, P.M., Bomans, P.H.H. and Stuart, M.C.A. (1991). The ultrastructure of cryo-sections and intact vitrified cells—The effect of cryoprotectants and acceleration voltage on beam induced bubbling. *Scanning Microsc.* (Suppl. 5), S43–S52.
- Freitag, B., Kujawa, S., Mul, P.M., Ringnalda, J. and Tiemeijer, P.C. (2005). Breaking the spherical and chromatic aberration barrier in transmission electron microscopy. *Ultramicroscopy* **102**(3), 209–214.
- Fujiyoshi, Y. (1998). The structural study of membrane proteins by electron crystallography. *Adv. Biophys.* **35**, 25–80.
- Gatan, Inc., Pleasanton, CA (www.gatan.com).
- Gilbert, P.F.C. (1972). The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. II. Direct methods. *Proc. R. Soc. Lond. B Biol. Sci.* **182**, 89–102.
- Grimm, R., Baermann, M., Haeckl, W., Typke, D., Sackmann, E. and Baumeister, W. (1996a). Energy-filtered electron tomography of ice-embedded actin and vesicles. *Biophys. J.* **72**, 482–489.
- Grimm, R., Barmann, M., Typke, D. and Baumeister, W. (1996b). Determination of the inelastic mean free path by examination of tilted vesicles and most probable loss imaging. *Ultramicroscopy* **63**(3–4), 169–179.
- Grimm, R., Singh, H., Rachel, R., Typke, D., Zillig, W. and Baumeister, W. (1998). Electron tomography of ice-embedded prokaryotic cells. *Biophys. J.* **74**, 1031–1042.
- Gruenewald, K., Desai, P., Winkler, D.C., Heymann, J.B., Belnap, D.M., Baumeister, W. and Steven, A.C. (2003). Three-dimensional structure of herpes simplex virus from cryo-electron tomography. *Science* **302**, 1396–1398.
- Gubbens, A.J. and Krivanek, O.L. (1993). Applications of a post-column imaging filter in biology and materials science. *Ultramicroscopy* **51**(1–4), 146–159.
- Gubbens, A.J., Kraus, B., Krivanek, O.L. and Mooney, P.E. (1995). An imaging filter for high voltage electron microscopy. *Ultramicroscopy* **59**(1–4), 255–265.
- Guckenberger, R. (1982). Determination of a common origin in the micrographs of tilt series in three-dimensional electron microscopy. *Ultramicroscopy* **9**, 167–174.
- Haider, M., Uhlemann, S., Schwan, E., Rose, H., Kabius, B. and Urban, K. (1998). Electron microscopy image enhanced. *Nature* **392**, 768–769.
- Han, K.F., Gubbens, A.J., Sedat, J.W. and Agard, D.A. (1996). Optimal strategies for imaging thick biological specimens—exit wave-front reconstruction and energy filtered imaging. *J. Microsc.* **183**, 124–132.
- Harauz, G. and van Heel, M. (1986). Exact filter for general geomer three-dimensional reconstruction. *Optik* **73**, 146–156.
- Hart, R.G. (1968). Electron microscopy of unstained biological material: The polytropic montage. *Science* **159**, 1464–1467.
- Hayward, S.B. and Glaeser, R.M. (1979). Radiation damage of purple membrane at low temperature. *Ultramicroscopy* **4**(2), 201–210.

- Hegerl, R. and Hoppe, W. (1976). Influence of electron noise on three-dimensional image reconstruction. *Z. Naturforschung*. **31a**, 1717–1721.
- Heide, H.G. (1984). Observation on ice layers. *Ultramicroscopy* **14**(3), 271–278.
- Henderson, R., Baldwin, J.M., Ceska, T.A., Beckman, E., Zemlin, F. and Downing, K. (1990). A model for the structure of bacteriorhodopsin based on high resolution electron cryomicroscopy. *J. Mol. Biol.* **213**, 899–929.
- Herman, G.T. (1980). On the noise in images produced by computed tomography. *Comput. Graphics Image Process.* **12**(3), 271–285.
- Herrmann, K.-H. and Krahle, D. (1982). The detection quantum efficiency of electronic image recording systems. *J. Microsc.* **127**(1), 17–28.
- Hoppe, W. (1969). Das Endlichkeitspostulat und das Interpolationstheorem der dreidimensionalen elektronenmikroskopischen Analyse aperiodischer Strukturen. *Optik* **29**, 617–621.
- Hoppe, W. (1974). Towards three-dimensional electron microscopy at atomic resolution. *Z. Naturwissenschaft.* **61**, 239–249.
- Hoppe, W. and Hegerl, R. (1980). Three-dimensional structure determination by electron microscopy (nonperiodic specimens). In *Computer Processing of Electron Microscopic Images* (P.W. Hawkes, Ed.), 127–185 (Springer-Verlag, Berlin).
- Hoppe, W., Langer, R., Knesch, G. and Poppe, C. (1968). Proteinkristallstrukturanalyse mit Elektronenstrahlen. *Z. Naturwissenschaft.* **55**, 333–336.
- Hoppe, W., Gassmann, J., Hunsmann, N., Schramm, H.J. and Strum, M. (1974). Three dimensional reconstruction of individual negatively stained yeast fatty-acid synthetase molecules from tilt series in the electron microscope. *Hoppe-Seyler's Z. Physiol. Chem.* **335**, 1483–1487.
- Horner, J.L. and Gianino, P.D. (1964). Phase-only matched filtering. *Appl. Opt.* **23**, 812–816.
- Horowitz, R.A., Agard, D., Sedat, J.W. and Woodcock, C.L. (1994). The three-dimensional architecture of chromatin in situ: Electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *J. Cell Biol.* **125**, 1–10.
- Hounsfield, G.N. (1972). A method and apparatus for examination of a body by radiation such as X- or gamma radiation. The Patent Office, London.
- Hsieh, C.E., Marko, M., Frank, J. and Mannella, C.A. (2002). Electron tomographic analysis of frozen-hydrated tissue sections. *J. Struct. Biol.* **138**(1–2), 63–73.
- Janesick, J.R. (2001). *Scientific Charge-Coupled Devices* (SPIE Press, Bellingham, WA).
- JEOL Ltd., Tokyo, Japan (www.jeol.com).
- Jia, C.L., Lentzen, M. and Urban, K. (2003). Atomic-resolution imaging of oxygen in perovskite ceramics. *Science* **299**, 870–873.
- Jiang, W., Li, Z., Zhang, Z., Baker, M.L., Prevelige Jr, P.E. and Chiu, W. (2003a). Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions, *Nature Struct. Biol.* **10**, 131–135.
- Jiang, W., Baker, M.L., Wu, Q., Bajaj, C. and Chiu, W. (2003b). Applications of a bilateral denoising filter in biological electron microscopy. *J. Struct. Biol.* **144**(1–2), 114–122.
- Kisielowski, C., Hetherington, C.J.D., Wang, Y.C., Kilaas, R., O'Keefe, M.A. and Thust, A. (2001). Imaging columns of the light elements carbon, nitrogen and oxygen with sub Angstrom resolution. *Ultramicroscopy* **89**(4), 243–263.
- Klumpp, M. and Baumeister, W. (1998). The thermosome—Archetype of group II chaperonins. *FEBS Lett.* **430**(1–2 Special Issue SI), 73–77.
- Klumpp, M., Baumeister, W. and Essen, L.O. (1997). Structure of the substrate binding domain of the thermosome, an archeal group II chaperonin. *Cell* **91**(2), 263–270.

- Knappek, E. (1982). Properties of organic specimens and their supports at 4K under irradiation in an electron microscope. *Ultramicroscopy* **10**(1–2), 71–86.
- Koster, A.J. and de Ruijter, W.J. (1992). Practical autoalignment of transmission electron microscopes. *Ultramicroscopy* **40**, 89–107.
- Koster, A.J. and Klumperman, J. (2003). Electron microscopy in cell biology: Integrating structure and function. *Nature Cell Biol.* (Suppl. S), SS6–SS10.
- Koster, A.J., de Ruijter, W.J., van den Bos, A. and van der Mast, K.D. (1989). Autotuning of a TEM using minimum electron dose. *Ultramicroscopy* **27**(3), 251–272.
- Koster, A.J., Chen, H., Sedat, J.W. and Agard, D. (1992). Automated microscopy for electron tomography. *Ultramicroscopy* **46**, 207–227.
- Koster, A.J., Grimm, R., Typke, D., Hegerl, R., Stoschek, A., Walz, J. and Baumeister, W. (1997). Perspectives of molecular and cellular electron tomography. *J. Struct. Biol.* **120**(3), 276–308.
- Knoll, M. and Ruska, E. (1932). Das Elektronenmikroskop. *Z. Phys.* **78**, 318–339.
- Kremer, J.R., Mastronarde, D.N. and McIntosh, J.R. (1996). Computer visualization of three-dimensional data using IMOD. *J. Struct. Biol.* **116**(1), 71–76.
- Krivanek, O.L. and Mooney, P.E. (1993). Applications of slow-scan CCD cameras in transmission electron microscopy. *Ultramicroscopy* **49**(1–4), 95–108.
- Krivanek, O.L., Gubbens, A.J. and Dellby, N. (1991). Development in EELS instrumentation for spectroscopy and imaging. *Microsc. Microanal. Microstruct.* **2**, 315–332.
- Krivanek, O.L., Gubbens, A.J., Dellby, N. and Meyer, C.E. (1992). Design and first applications of a post-column imaging filter. *Microsc. Microanal. Microstruct.* **3**(2–3), 187–199.
- Krivanek, O.L., Friedman, S.L., Gubbens, A.J. and Kraus, B. (1995). An imaging filter for biological applications. *Ultramicroscopy* **59**(1–4), 267–282.
- Kumar Vijaya, B.V.K., Dickey, F.M. and DeLaurentis, J.M. (1992). Correlation filters minimizing peak location errors. *J. Opt. Soc. Am. A—Opt. Image Sci.* **9**(5), 678–682.
- Kunath, W. and Sack-Kongehl, H. (1989). Circular harmonic averaging of noisy single-molecule images. *Ultramicroscopy* **27**(2), 171–184.
- Ladinsky, M.S., Mastronarde, D.N., McIntosh, J.R., Howell, K.E. and Staehelin, L.A. (1999). Golgi structure in three dimensions: Functional insights from the normal rat kidney cell. *J. Cell Biol.* **144**(6), 1135–1149.
- Langmore, J.P. and Smith, M.F. (1992). Quantitative energy filtered electron microscopy of biological molecules in ice. *Ultramicroscopy* **46**, 349–373.
- Lanio, S. (1986). High-resolution imaging magnetic energy filters with simple structure. *Optik* **73**, 99–107.
- Lanio, S., Rose, H. and Krahel, D. (1986). Test and improved design of a corrected imaging magnetic energy filter. *Optik* **73**(2), 56–68.
- Lawrence, M.C. (1992). Least-squares method of alignment using markers. In *Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope* (J. Frank, Ed.), (Plenum Press, New York).
- Leis, A., Andrees, L., Gruska, M., Al-Amoudi, A., Sartori, A., Dubochet, J. and Baumeister, W. (2005). Cryo-electron tomography and fluorescence microscopy of unicellular algae in vitreous cryosections. *Proceedings of the 63rd Annual Meeting MSA. Microsc. Microanal.* **11**(Suppl. 2), 330–331.
- Lengweiler, S., Birrer, G., Müller, M. and Stemmer, A. (2004). Ideal eucentric goniometer for electron tomography. *Proceedings of the 13th European Microscopy Congress*, Antwerp, Belgium, Vol. 1, IM07.O4 CD.
- Lentzen, M. (2004). The tuning of a Zernike phase plate with defocus and variable spherical aberration and its use in HRTEM imaging. *Ultramicroscopy* **99**(4), 211–220.

- Lentzen, M., Jahnen, B., Jia, C.L., Thust, A., Tillmann, K. and Urban, K. (2002). High-resolution imaging with an aberration-corrected transmission electron microscope. *Ultramicroscopy* **92**(3–4), 233–242.
- Li, H.L., DeRosier, D.J., Nicholson, W.V., Nogales, E. and Downing, K.H. (2002). Microtubule structure at 8 Ångstrom resolution. *Structure* **10**(10), 1317–1328.
- Lowe, J., Stock, D., Jap, R., Zwickl, P., Baumeister, W. and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**(5210), 533–539.
- Lucic, V., Foerster, F. and Baumeister, W. (2005). Structural studies by electron tomography: From cells to molecules. *Annu. Rev. Biochem.* **74**, 833–865.
- Luther, P.K. (1992). Sample shrinkage and radiation damage. In *Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope* (J. Frank, Ed.), 39 (Plenum Press, New York).
- Luther, P.K., Lawrence, M.C. and Crowther, R.A. (1988). A method for monitoring the collapse of plastic sections as a function of electron dose. *Ultramicroscopy* **24**, 7–18.
- Majorovits, E. and Schroeder, R.R. (2002). Improved information recovery in phase contrast EM for non-two-fold symmetric Boersch phase plate geometry. *Microsc. Microanal.* **8**(Suppl. 2), 540 CD.
- Malis, T., Cheng, S.C. and Egerton, R. (1988). EELS log-ratio technique for specimen-thickness measurement in the TEM. *J. Electron Microsc. Tech.* **8**, 193–200.
- Marko, M. (2004). Techniques for electron tomography in biological and material science. *Microsc. Microanal.* **10**(Suppl. 2), 154–155.
- Marsh, B.J., Mastronarde, D.N., Buttle, K.F., Howell K.E. and McIntosh, J.R. (2001). Organellar relationships in the Golgi region of the pancreatic beta cell line, HIT-T15, visualized by high resolution electron tomography. *Proc. Natl. Acad. Sci. USA* **98**(5), 2399–2406.
- Martone, M.E., Deerinck, T.J., Young, S.J. and Ellisman, M.H. (1999). Three dimensional protein localization using high voltage electron microscopy. *Acta Histochem. Cytochem.* **32**(1), 35–43.
- Martone, M.E., Deerinck, T.J., Yamada, N., Bushong, E. and Ellisman, M.H. (2000). Correlated 3D light and electron microscopy: Use of high voltage microscopy tomography for imaging large biological structures. *J. Histotechnol.* **23**, 261–270.
- Mastronarde, D.N. (1997). Dual-axis tomography: An approach with alignment methods that preserve resolution. *J. Struct. Biol.* **120**(3), 343–352.
- Mcewen, B.F., Downing, K.H. and Glaeser, R.M. (1995). The relevance of dose-fractionation in tomography of radiation-sensitive specimens. *Ultramicroscopy* **60**(3), 357–373.
- Medalia, O., Weber, I., Frangakis, A.S., Gerisch, G. and Baumeister, W. (2002). Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* **298**, 1209–1213.
- Michel, M., Hillmann, T. and Mueller, M. (1991). Cryosectioning of plant material frozen at high pressure. *J. Microsc.* **193**, 3–18.
- Mochel, M.E. and Mochel, J.M. (1986). A CCD imaging and analysis system for the VG HB5 STEM. In *Proceedings of the 51st Annual Meeting of the Microscopy Society of America*, San Francisco Press, 262–263.
- Mooney, P.E. (2004). Private communications.
- Mooney, P.E. and Krivanek, O.L. (1994). Image-coupling methods in CCD cameras for electron microscopy. In *Proceeding of the 52nd Annual Meeting of the Microscopy Society of America*, San Francisco Press, 406–407.

- Moore, H. (1987). Theory and practice of high-pressure freezing. In *Cryotechniques in Biological Electron Microscopy* (R.A. Steinbrecht, and K. Zierold, Eds.), 175–191 (Springer, Berlin).
- Moore, H. and Riehle, U. (1968). Snap freezing under high pressure: A new fixation technique for freeze-etching. In *Electron Microscopy 1968* (D.S. Bocciarelli, Ed.), Vol. 2. *Proceedings of the 4th European Regional Conference on Electron Microscopy Rome*, 33–34.
- Mueller-Reichert, T., Hohenberg, H., O'Toole, E.T. and McDonald, K. (2003). Cryoimmobilization and three-dimensional visualization of C-elegans ultrastructure. *J. Microsc.* **212**(1), 71–80.
- Murk, J.L., Humbel, M., Ziese, U., Griffith, J.M., Posthuma, G., Slot, J.W., Koster, A.J., Verkleij, A.J., Geuze, H.J. and Kleijmeer, M.J. (2003). Endosomal compartmentalization in three dimensions: implications for membrane fusion. *Proc. Natl. Acad. Sci. USA* **100**(23), 13332–13337.
- Nicastro, D., Frangakis, A.S., Typke, D. and Baumeister, W. (2000). Cryo-electron tomography of *Neurospora* mitochondria. *J. Struct. Biol.* **129**, 48–56.
- Nicastro, D., McIntosh, J.R. and Baumeister, W. (2005). 3D structure of eukaryotic flagella in a quiescent state revealed by cryo-electron tomography. *Proc. Natl. Acad. Sci. USA* **102**, 15889–15894.
- Nickell, S., Hegerl, R., Baumeister, W. and Rachel, R. (2003). Pyrodictium canulae enter the periplasmic space but do not enter the cytoplasm, as revealed by cryo-electron tomography. *J. Struct. Biol.* **141**, 34–42.
- Nickell, S., Hegerl, R., Sillje, H. and Baumeister, W. (2007). Dual-axis cryo-electron tomography. *J. Struct. Biol.*, submitted.
- Nitsch, M., Walz, J., Typke, D., Klumpp, M., Essen, L.O. and Baumeister, W. (1998). Group II chaperonin in an open conformation examined by electron tomography. *Nature Struct. Biol.* **5**(10), 855–857.
- Olins, A.L., Olins, D.E., Levy, H.A., Margle, S.M., Tinnel, E.P. and Dufree, R.C. (1989). Tomographic reconstruction from energy-filtered images of thick biological sections. *J. Microsc.* **154**, 257–265.
- *Ortiz, J., Förster, F., Kürner, J., Linaroudis, A.A. and Baumeister, W. (2006). Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *Journal of Structural Biology* **156**, 334–341.
- O'Toole, E.T., Winey, M. and McIntosh, J.R. (1999). High-voltage electron tomography of spindle pole bodies and early mitotic spindles in the yeast *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **10**(6), 2017–2031.
- Penczek, P., Marko, M., Buttle, K. and Frank, J. (1995). Double-tilt electron tomography. *Ultramicroscopy* **60**(3), 393–410.
- Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H. and North, A.C.T. (1960). Structure of haemoglobin: A three-dimensional Fourier synthesis at 5.5 resolution, obtained by x-ray analysis. *Nature* **185**, 416–422.
- Plano GmbH, Wetzlar, Germany (www.plano-em.de).
- Plitzko, J.M., Frangakis, A.S., Nickell, S., Förster, F., Gross, A. and Baumeister, W. (2002). In vivo veritas: Electron cryotomography of cells. *Trends Biotech.* **20**(Suppl. 8), S40–S44.
- Plitzko, J.M., G. Schweikert, G., Pfeifer, G., Nickell, S., Luecken, U. and Baumeister, W. (2004). The benefit of liquid Helium cooling for cryo-electron tomography: a quantitative comparative study. *Proc. of the 13th European Microscopy Congress, Antwerp, Belgium Vol 2*, IM07.04 CD.
- Plitzko, J.M., Freitag, B., Hegerl, R. and Luecken, U. (2005). Aberration free microscopy for life science applications. *Proceedings of the 63rd Annual Meeting of the Microscopy Society of America. Microsc. Microanal.* **11**(Suppl. 2), 2140–2141.
- Porter, K.R. and Blum, J. (1953). A study in microtomy in electron microscopy. *Anat. Rec.* **117**, 685–710.
- Probst, W., Benner, G., Bihr, J. and Weimer, E. (1993). An omega energy filtering TEM—principles and applications. *Adv. Mater.* **5**, 297–300.

- Quantifoil Micro Tools GmbH, Jena, Germany (www.quantifoil.com).
- Radermacher, M. (1992). Weighted backprojection methods. In *Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope* (J. Frank, Ed.), 91–115 (Plenum Press, New York).
- Radermacher, M. and Hoppe, W. (1980). Properties of 3-D reconstruction from projections by conical tilting compared to single-axis tilting. *7th European Congress on Electron Microscopy*, Den Haag, 132–133.
- Radon, J. (1917). Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Math. Phys. Klasse* **69**, 262–277.
- Rath, B.K., Marko, M., Radermacher, M. and Frank, J. (1997). Low-dose automated electron tomography: A recent implementation. *J. Struct. Biol.* **120**, 210–218.
- Rath, B.K., Hegerl, R., Leith, A., Shaikh, T.R., Wagenknecht, T. and Frank, J. (2003). Fast 3D motif search of EM density maps using a locally normalized cross-correlation function. *J. Struct. Biol.* **144**(1–2), 95–103.
- Reimer, L. (1993). *Transmission Electron Microscopy* (Volume 36 of Springer series in Optical Sciences), 3rd Ed. (Springer, Berlin).
- Reimer, L. (1995). *Energy-Filtering Transmission Electron Microscopy* (Volume 71 of Springer series in Optical Sciences) (Springer, Berlin).
- Rockel, B., Peters, J., Kuhlmorgen, B., Glaeser, R.M. and Baumeister, W. (2002). A giant protease with a twist: The TPP II complex from *Drosophila melanogaster* studied by electron microscopy. *EMBO J.* **22**, 5979–5984.
- Rockel, B., Peters, J., Müller, S.A., Gönül, S., Ringler, P., Hegerl, R., Glaeser, R.M. and Baumeister, W. (2005). Molecular architecture and assembly of *Drosophila* tripeptidyl peptidase II. *PNAS* **102**(29), 10135–10140.
- Rose, H. and Plies, E. (1974). Entwurf eines fehlerarmen magnetischen Energie-Analysators. *Optik* **40**, 336–341.
- Roseman, A.M. (2000). Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr. Sect. D—Biol. Crystallogr.* **56**(10), 1332–1340.
- Roseman, A.M. (2003). Particle finding in electron micrographs using a fast local correlation algorithm. *Ultramicroscopy* **94**(3–4), 225–236.
- Ruijter, W.J. (1995). Imaging properties and applications of slow-scan charge coupled device cameras suitable for electron microscopy. *Micron* **26**(3), 247–275.
- Ruska, E. (1987). The development of the electronmicroscope and electron microscopy. Nobel lecture 1986. *Rev. Modern Phys.* **59**(3), 627–638.
- Sali, A., Glaeser, R., Earnest, T. and Baumeister, W. (2003). From words to literature in structural proteomics. *Nature* **422**, 216–225.
- Sandin, S., Ofverstedt, G., Wikstrom, A.C., Wrangé, O. and Skoglund, U. (2004). Structure and flexibility of individual immunoglobulin g molecules in solution. *Structure (Camb.)*, **12**(3), 409–415.
- Sartori, N., Richter, K. and Dubochet, J. (1993). Vitrification depth can be increased more than 10-fold by high-pressure freezing. *J. Microsc.* **172**, 55–61.
- Saxberg, B.E.H. and Saxton, W.O. (1981). Quantum noise in 2D projections and 3D reconstructions. *Ultramicroscopy* **6**(1), 85–89.
- Saxton, W.O., Baumeister, W. and Hahn, M. (1984). Three-dimensional reconstruction of imperfect two-dimensional crystals. *Ultramicroscopy* **13**(1–2), 57–70.
- Scheffel et al. (2005). Advance online publication, *Nature* doi:10.1038/nature04382 and Scheffel, A., Gruska, M., Faivre, D., Linaroudis, A., Plitzko, J.M. and Schüler, D. (2006) An acidic protein aligns magnetosomes along a filamentous structure in magnetotactic bacteria. *Nature* **440**(7080), 110–114.

- Scherzer, O. (1970). Die Strahlenschädigung der Objekte als Grenze für die hochauflösende Elektronenmikroskopie (Radiation damage of objects as limiting factor of high-resolution electron microscopy). *Ber. Bunsenges. Phys. Chem.* **74**, 1154–1167.
- Schiske, P. (1968). Zur Frage der Bildrekonstruktion durch Fokusserien. *Electron Microscopy 1968* (D.S. Bocciarelli, Ed.), *Proceedings of the 4th European Regional Conference on Electron Microscopy*, Rome, 145–146.
- Schiske, P. (1973). *Image Processing and Computer Aided Design in Electron Optics* (Academic Press, New York).
- Schoenmakers, R., Perquin, R., Fliervoet, T., Voorhout, W. and Schirmacher, H. (2005). New software for high-resolution, high-throughput electron tomography. *Microsc. Anal.* **96**, 13–14.
- Schroeder, R.R., Manstein, D.J., Jahn, W., Holden, H., Rayment, I., Holmes, K.C. and Spudich, J.A. (1993). Three-dimensional atomic model of F-actin decorated with Dictyostelium myosin S1. *Nature* **364**, 171–174.
- Schweikert, G., Pfeifer, G., Nickell, S., Luecken, U., Baumeister, W. and Plitzko, J.M. (2007). Helium vs. Nitrogen cooling of biological samples: A quantitative comparative study. *J. Struct. Biol.*, submitted.
- Sitte, H. (1996). Advanced instrumentation and methodology related to cryo-ultramicrotomy: A review. *Scanning Microsc.* (Suppl. **10**), 387–466.
- Smith, P.R., Peters, T.M. and Bates, R.H.T. (1973). Image reconstruction from finite numbers of projections. *J. Phys. A: Math. Nucl. Gen.* **6**, 361–382.
- Soto, G.E., Young, S.J., Martone, M.E., Deerinck, T.J., Lamont, S., Carragher, B.O., Hama, K. and Ellisman, M.H. (1994). Serial section electron tomography, a method for three-dimensional reconstruction of large structures. *Neuroimage* **1**, 230–243.
- Spahn, C.M.T., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G. and Frank, J. (2001). Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit-subunit interactions. *Cell* **107**(3), 373–386.
- Spence, J.C.H., Weierstall, U., Fricke, T.T., Glaeser, R.M. and Downing, K.H. (2003). Three-dimensional diffractive imaging for crystalline monolayers with one-dimensional compact support. *J. Struct. Biol.* **144**(1–2), 209–218.
- Steven, A.C. and Aebi, U. (2003). The next ice age: Cryo-electron tomography of intact cells. *Trends Cell Biol.* **13**(3), 107–110.
- Stoschek, A. and Hegerl, R. (1997). Denoising of electron tomographic reconstructions using multiscale transformations. *J. Struct. Biol.* **120**(3), 257–265.
- Studer, D. and Gnägi, H. (2000). Minimal compression of ultrathin sections with use of an oscillating diamond. *J. Microsc.* **197**, 94–100.
- Taylor, K.A. and Glaeser, R.M. (1974). Electron diffraction of frozen, hydrated protein crystals. *Science* **186**(4168), 1036–1037.
- Taylor, K.A. and Glaeser, R.M. (1973). Hydrophilic support films of controlled thickness and composition. *Rev. Scient. Instrum.* **44**(10), 1546–1547.
- Taylor, K.A., Tang, J., Cheng, Y. and Winkler, H. (1997). The use of electron tomography for structural analysis of disordered protein arrays. *J. Struct. Biol.* **120**(3), 372–386.
- Ted Pella, Inc., Redding, CA (www.tedpella.com).
- Thust, A., Lentzen, M. and Urban, K. (1994). Non-linear reconstruction of the exit plane wave function from periodic high-resolution electron microscopy images. *Ultramicroscopy* **53**(2), 101–120.
- Thust, A., Coene, W.M.J., Op de Beeck, M. and Van Dyck, D. (1996). Focal-series reconstruction in HRTEM: Simulation studies on non-periodic objects. *Ultramicroscopy* **64**, 211–230.
- Turchetta, R., Bertst, J.D., Casadei, B., Claus, G., Colledani, C., Dulinski, W., Hu, Y., Husson, D., Le Normand, J.P., Riester, J.L., Deptuch, G., Goerlach,

- U., Higuieret, S. and Winter, M. (2001). A monolithic active pixel sensor for charged particle tracking and imaging using standard VLSI CMOS technology. *Nucl. Instrum. Methods Phys. Res. Sect. A—Accel. Spectrom. Detect. Associ. Equip.* **458**(3), 677–689.
- Typke, D., Dierksen, K. and Baumeister, W. (1991). Automatic electron tomography. *Proceedings of the 49th Annual Meeting of the Microscopy Society of America*, San Francisco Press, 544–545.
- Typke, D., Hegerl, R. and Kleinz, J. (1992). Image restoration for biological objects using external TEM control and electronic image recording. *Ultramicroscopy* **46**, 157–173.
- Uhlemann, S. and Rose, H. (1996). Acceptance of imaging energy filters. *Ultramicroscopy* **63**(3–4), 161–167.
- Unwin, P.N.T. (1972). Electron microscopy of biological specimens by means of an electrostatic phase plate. *Proc. R. Soc. Lond. A.* **329**, 327–359.
- VanderLugt, A. (1964). Signal detection by complex spatial filtering. *IEEE Trans. Inf. Theory, IT* **10**, 139–145.
- Van Heel, M. and Frank, J. (1981). Use of multivariate statistics in analyzing the images of biological macromolecules. *Ultramicroscopy* **6**, 187–194.
- Van Heel, M., Schatz, M. and Orlova, E.V. (1992). Correlation function revisited. *Ultramicroscopy* **46**, 307–316.
- Van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M. and Patwardhan, A. (2000). Single-particle electron cryo-microscopy: Towards atomic resolution. *Q. Rev. Biophys.* **33**, 307–369.
- Volkman, N. (2002). A novel three-dimensional variant of the watershed transform for segmentation of electron density maps. *J. Struct. Biol.* **138**(1–2), 123–129.
- Volkman, N. and Hanein, D. (1999). Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J. Struct. Biol.* **125**(2–3), 176–184.
- Walz, J., Tamura, T., Tamura, N., Grimm, R., Baumeister, W. and Koster, A.J. (1997a). Tricorn protease exists as an icosahedral supermolecule in vivo. *Mol. Cell* **1**(1), 59–65.
- Walz, J., Typke, D., Nitsch, M., Koster, A.J., Hegerl, R. and Baumeister, W. (1997b). Electron tomography of single ice-embedded macromolecules: Three-dimensional alignment and classification. *J. Struct. Biol.* **120**, 387–395.
- Weickenmeier, A.L., Nuechter, W. and Mayer, J. (1995). Quantitative characterization of point spread function and detection quantum efficiency for a YAG scintillator slow scan CCD camera. *Optik* **99**(4), 147–154.
- Wilson, C.J., Mastronarde, D.N., McEwen, B. and Frank, J. (1992). Measurement of neuronal surface area using high-voltage electron microscope tomography. *Neuroimage* **1**, 11–22.
- Winkler, H. and Taylor, K.A. (2003). Focus gradient correction applied to tilt series image data used in electron tomography. *J. Struct. Biol.* **143**(1), 24–32.
- Wright, E.R., Iancu, C.V., Tivol, W.F. and Jensen, G.J. (2006). Observation on the behaviour of vitreous ice at ~82 and ~12 K. *J. Struct. Biol.* **153**, 241–252.
- Yonekura, K., Yonekura, S.M. and Namba, K. (2002). Quantitative comparison between zero-loss and conventional electron diffraction from 2 dimensional and thin three dimensional protein crystals. *Biophys. J.* **82**, 2784–2797.
- Zheng, Q.S., Braumfeld, M.B., Sedat, J. and Agard, D.A. (2004). An improved strategy for automated electron microscopic tomography. *J. Struct. Biol.* **147**(2), 91–101.

- Zhu, P., Penczeck, P., Schroeder, R.R. and Frank, J. (1997). 3-dimensional reconstruction with contrast transfer-function correction from energy filtered cryoelectron micrographs—Procedure and application to the 70s E. Coli ribosome. *J. Struct. Biol.* **118**, 197–219.
- Zhu, P., Chertova, E., Bess, J., Lifson, J.D., Arthur, L.O., Liu, J., Taylor, K.A. and Roux, K.H. (2003). Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc. Natl. Acad. Sci. USA* **100**(26), 15812–15817.
- Ziese, U., Janssen, A.H., Murk, J.-L., Geerts, W.J.C., van der Krift, T., Verkleij, A.J. and Koster, A.J. (2002). Automated high-throughput electron tomography by pre-calibration of image shifts. *J. Microsc.* **205**(2), 187–200.
- Ziese, U., Geerts, W.J.C., Van Der Krift, T.P., Verkleij, A.J. and Koster A.J. (2003). Correction of autofocus errors due to specimen tilt for automated electron tomography. *J. Microsc.* **211**(2), 179–185.
- Zwickl, P., Goldberg, A.L. and Baumeister, W. (2000). Proteasomes in prokaryotes. In *Proteasomes: The World of Regulatory Proteolysis*, 8–20 (Landes Bioscience, Georgetown, TX).

LEEM and SPLEEM

Ernst Bauer

1 Introduction

Low-energy electron microscopy (LEEM: <http://www.leem-user.com>) is an imaging method that makes use of elastically backscattered electrons with energies below about 100 eV, frequently with less than 10 eV. In contrast to transmission electron microscopy (TEM), which generally works with electrons in the 100 keV range where backscattering is negligible, the backscattering cross sections for low-energy electrons are large enough to make them useful for surface imaging. This was already evident in the classical diffraction experiments of Davission and Germer,¹ but it took 35 years before the use of slow diffracted electrons for surface imaging was suggested² and another 23 years before convincing images could be published.³ Thus, although diffraction of slow electrons and imaging with slow emitted electrons with resolution in the micrometer range were demonstrated⁴ before TEM reached submicron resolution,⁵ LEEM became a viable imaging method only much later.

The reasons for this late appearance of LEEM in electron microscopy are 2-fold: (1) for LEEM well-defined surfaces are necessary, which in general requires ultrahigh vacuum (UHV) and efficient surface cleaning procedures. Although these have been available for some time in glass systems such as in Farnsworth's low-energy electron diffraction (LEED) systems,^{6–8} glass systems are not very user friendly and metal UHV technology did not come into widespread use until the beginning of the 1960s. In fact, the first display type LEED system that started the revival of LEED was a glass system^{9,10} as was the first unsuccessful model of an LEEM system.² (2) There was the widespread belief in the electron microscope community based on the fundamental theoretical work of Recknagel on emission electron microscopy¹¹ that the chromatic aberration of the objective lens would limit the resolution to such an extent that LEEM would be unattractive. This of course was a misunderstanding as already pointed out in the early phase of the development of LEEM.^{12,13} In the decades since then LEEM has developed slowly into a powerful surface imaging technique as recounted elsewhere.¹⁴

In the past 10 years developments concentrated on the combination of LEEM with X-ray induced photoemission electron microscopy (XPEEM), which resulted in the spectroscopic photoemission and low-energy electron microscope (SPELEEM),^{15,16} and on the correction of the aberrations of the objective lens.¹⁷ Despite these efforts LEEM is still far behind the technological state that TEM has now reached.

This chapter reviews the basics of LEEM, its present state of art, and its applications. An associated imaging method, spin-polarized LEEM (SPLEEM), which gives magnetic information, will also be discussed. Other methods, such as mirror electron microscopy (MEM), which gives information mainly on the local surface potential, Auger electron emission microscopy (AEEM), which gives chemical information, electron energy loss microscopy (EELM), which gives some electronic information, and secondary electron emission microscopy (SEEM), will be mentioned only briefly because they have been used much less frequently, although they are also useful. Photoelectron emission microscopy (PEEM), in particular XPEEM, is included only in connection with the discussion of SPELEEM because it is the subject of another chapter.

2 Electron Beam–Specimen Interactions

To understand the possibilities and limitations of LEEM and of the associated techniques, a fundamental understanding of the interactions of slow electrons with condensed matter is necessary. The following interactions have to be taken into account: elastic scattering, inelastic scattering, and quasielastic scattering (phonon and in magnetic materials magnon scattering). The interactions of slow electrons with matter are vastly different from those of fast electrons. In ferromagnetic materials they depend in addition upon the relative orientation of the spin of the incident electrons and the electrons in the matter.

Consider first elastic scattering. Because of its low velocity $v = \sqrt{2E/m}$ the interaction time of a slow electron is much longer than of the fast electrons used in TEM and an n -electron atom may no longer be considered as undisturbed but becomes an $n + 1$ electron system during the interaction. Therefore the incident electron experiences the temporary excitations of the n -electron atom. This can be taken into account by adding a correlation potential to the potential of the ground state n -electron atom. Similarly the repulsive interactions between electrons with the same spin due to the Pauli principle cause a spin-dependent potential that also has to be added. As a consequence, the scattering of slow electrons by the atoms that constitute the condensed matter can no longer be described by the first Born approximation, which assumes a static atom in the ground state and which is a good approximation at high energies. Instead, a partial wave analysis is necessary, taking into account the exchange and correlation potentials.^{18,19} No calculations of this type for condensed atoms, whose potentials are truncated by overlap with the neighbor atoms, are available. Fortunately the magnitudes of the correlation and exchange potential decrease rapidly with energy so

that they may be neglected in the energy range of conventional LEED studies (usually above 30 eV). In LEEM, however, they should be taken into account. This is a formidable task that has not been mastered up to now. Therefore, only some results of partial wave analysis calculations for truncated ground state potentials will be given here.

In the partial wave analysis²⁰ the incident plane wave and the outgoing scattered wave are expanded into spherical harmonics centered at the atom, and the phase differences η_l between the incident and outgoing partial waves are calculated. In the nonrelativistic case the scattering amplitude is given by

$$f(\theta, k) = (1/2ik)\Sigma(2l + 1)[\exp(2i\eta_l) - 1]P_l(\cos\theta) \quad (1)$$

with the sum over l extending from zero to infinity. k is the wave number and the P_l 's are Legendre polynomials. The intensity distribution of the scattered electrons as a function of scattering angle θ and energy $E \sim \sqrt{k}$ is then simply $\sim |f(\theta, k)|^2$. Figure 8-1 shows the angular distribution of the scattered intensity of 50 eV electrons, calculated in this manner for realistic solid state Ag, Al, and Cu atomic potentials.¹⁸ It shows that not only is the total scattering cross section of Cu ($Z = 29$) smaller than that of Al ($Z = 13$), but also its back-scattering cross section, and that Al scatters nearly as much as the much heavier Ag atom ($Z = 47$). The scattering is, however, strongly energy dependent. This is illustrated for the scattering into a 30° cone around the backward direction in Figure 8-2,¹⁹ which shows that at very low energies Cu scatters nearly as strongly as W ($Z = 74$) while Al and Ag scatter much more weakly in the backward direction. It should be noted that the zero of the energy is the inner potential resulting from the overlap of the free atom potentials so that the maxima of the W and Cu backscattering cross sections are just around the vacuum level.

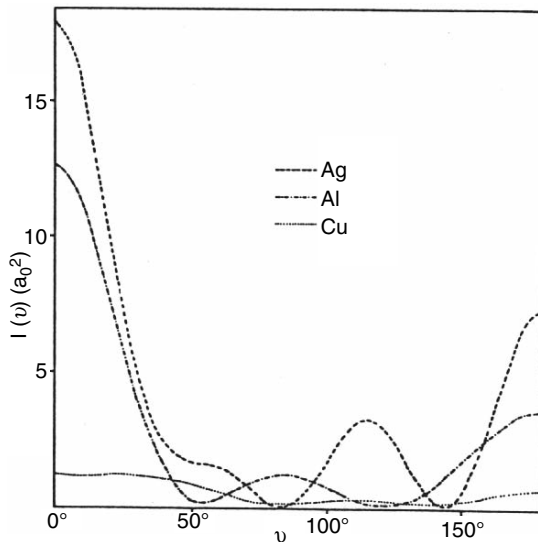


Figure 8-1. Angular distribution of 50 eV electrons elastically scattered from Ag, Al, and Cu atoms in the solid state.¹⁸

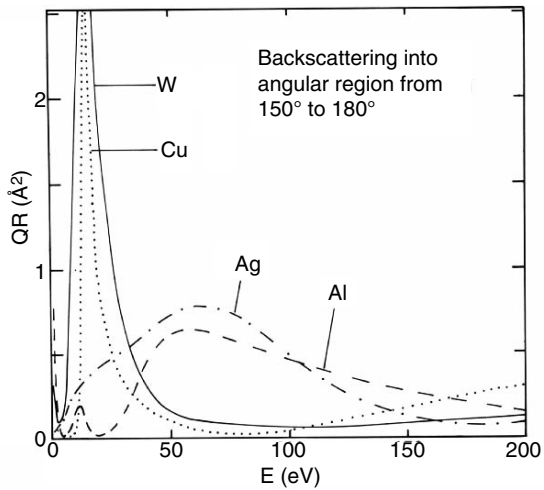


Figure 8-2. Energy dependence of the backscattering into a 30° cone around the backward direction for Ag, Al, Cu, and W atoms in the solid state.¹⁹

In condensed matter the electrons are, of course, not only scattered within one atom but also by the atoms surrounding it, causing strong multiple scattering. This is taken into account in LEED in the dynamic theory of electron diffraction.^{21,22} Another way to look at the problem of scattering in a periodic system is in terms of the band structure theory if we assume that the $n + 1$ -electrons system (n crystal electrons + incident electron) does not differ significantly from the n -electron system (Koopman' theorem²³). Then the 180° backscattering from a single crystal surface is determined by the band structure $E(\mathbf{k})$ perpendicular to the surface. This is illustrated in Figure 8-3 for the W(110) surface.²⁴ The band structure in the $[110]$ (ΓN) direction has a wide

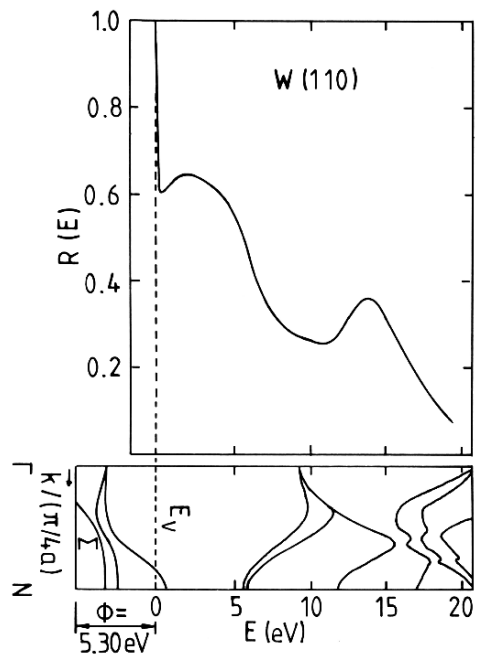


Figure 8-3. Normal incidence specular reflectivity of a W(110) surface (top) and band structure along the surface normal (bottom).²⁴

band gap between about 1 and 6 eV above the vacuum level. An electron incident in this direction, therefore, does not find allowed states in the crystal and forms an evanescent wave. The extinction length of this electron wave in the crystal is quite short in the center of the gap, only about two monolayers,²⁵ so that the electron is reflected before it is attenuated significantly by inelastic scattering. This, together with the strong backscattering cross section, causes the high reflectivity at about 2–3 eV. The second reflectivity peak is due to the low density of states in the crystal as indicated by the steep bands. The band structure influence is strongly orientation dependent. For example, on the W(100) surface the band gap is located between 3 and 5 eV above the vacuum level,^{24,26} which causes a pronounced reflectivity peak at about 4 eV. This is preceded by a deep reflectivity minimum, which is caused by the strong inelastic scattering of the electron that could otherwise penetrate deeply into the crystal. A second reflectivity peak occurs around 8 eV where the density of state in the crystal is small. This simple picture neglects the influence of surface effects. For quantitative agreement between experiment and theory the surface barrier,²⁷ surface resonances,²⁴ and reconstruction have to be taken into account. For LEEM these details are not important, at least at the present state of art, because they determine mainly the reflected intensity and have little influence on the contrast.

The main factor that determines the high surface sensitivity of LEEM is in general not the influence of the band structure and of elastic scattering but the strong attenuation of slow electrons by inelastic scattering. Inelastic scattering is due to single electron excitations (electron hole pair creation) and collective electron excitations (plasmon creation). In the energy range of LEEM single electron excitations mainly involve valence band and weakly bound outer shell core electrons. The universal inelastic mean free path (IMFP) curves usually found in the literature are of rather limited value at the low energies used because they do not take into account the differences in the electronic structure of the various materials. Therefore, only some general features will be discussed and some specific examples will be given. In materials that may be described approximately by a free electron gas imbedded in a homogeneous background of equal charge ("jellium model") the IMFP is a function of k/k_F (k_F Fermi wave number) with the electron density as parameter.^{18,19,28} As an example, the attenuation length $\mu = (\text{IMFP})^{-1}$ of Al, for which the free electron approximation is good, is shown in Figure 8–4 together with the attenuation coefficient ν due to elastic backscattering, assuming a random distribution of Al atoms with bulk density ("randium model").^{29,30} The initial rise of μ until the volume plasmon creation threshold at $E_T = 17.5$ eV (above vacuum level) is due to single electron excitations. The maximum of μ and the corresponding minimum of about 0.3 nm of the IMPF at about 37 eV is mainly due to plasmon losses. Figure 8–4 also shows that attenuation by elastic backscattering is much weaker above E_T than that by inelastic scattering.

For most metals the jellium approximation is not useful, in particular for transition and noble metals. For example, in contrast to jellium, transition metals have a high density of unoccupied states just above

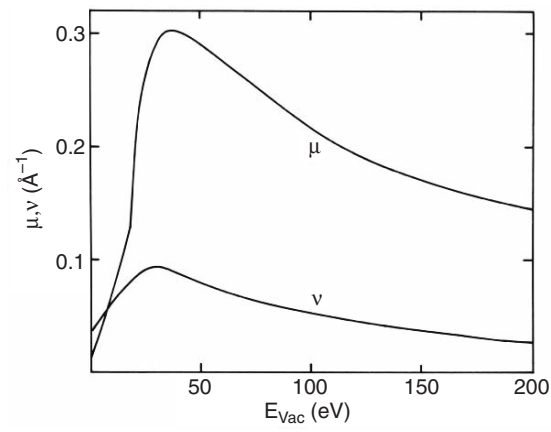


Figure 8–4. Energy dependence of the attenuation coefficients μ, ν of slow electrons in Al by inelastic scattering and elastic backscattering, respectively.²⁹

the Fermi level into which excitations can occur. The deviation from jellium can be partially taken into account by replacing the ω dependence in the Lindhard dielectric function $\epsilon_L(\omega, q)$, which is used in the jellium calculations, by that obtained in the experiment for zero momentum q transfer, that is by optical data for which $q = 0$. Results of such calculations³¹ give typically minimum IMFPs in metals of 0.3–0.5 nm at energies between 30 and 120 eV for Mg and Au, respectively, with a rapid increase at low energies to values as high as 2.4 nm in Si and 3.5 nm in W at 10 eV, for example. Figure 8–5 illustrates the agreement between theory and experiment that can be obtained in this approximation.³² The deviations below $E - E_F = 5$ eV are irrelevant for LEEM because the work function of Au is about 5 eV; those above 50 eV are probably due to inaccurate 5p and 4f ionization cross

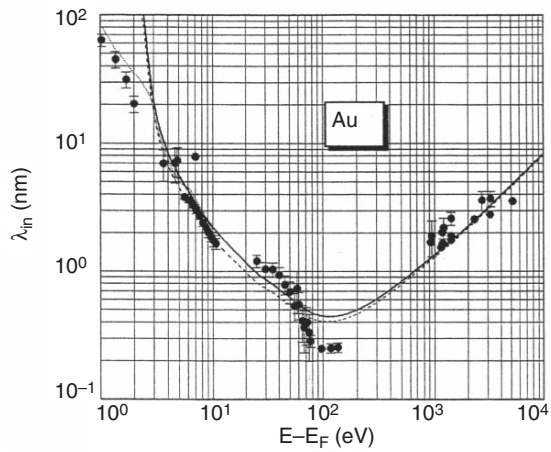


Figure 8–5. Energy dependence of the inelastic mean free path of electrons in Au. The points and dots are experimental data and the dashed and solid lines are theoretical data using different approximations.³²

sections. The agreement is surprisingly good considering that the q dependence of ϵ has been approximated by simple expressions and that correlation and exchange have not been taken into account. At energies below several 10eV these are of comparable importance for inelastic scattering as for elastic scattering, in particular the influence of the detailed band structure and of nondirect ($q \neq 0$) transitions.³³ For example, inclusion of exchange in the dielectric model of the IMFP gives IMFP values that are by a factor of 1.3 and more larger than without exchange.³⁴

The IMFPs calculated in this approximation for insulators are even larger, such as 6nm at 10eV for KCl.³¹ For several groups of insulators with large band gaps (condensed noble gases, N_2 , and organic dielectrics such as benzene or methane) no electronic excitations are possible at low energies. Here the (quasi)elastic mean free path (EMFP) determines the sampling depth. EMFP measurements in the energy range from 2 to 15eV give EMFPs up to 10nm.³⁵ An example is shown for solid Xe in Figure 8–6.³⁶ Thus the MFP may be very long at low energies (≤ 10 eV) while at energies between about 30eV and 100eV, depending upon the material, it may be only a few tenths of a nm. The large mean free paths (MFPs) at very low energies are, however, not general. They depend strongly on the density of unoccupied states into which bound electrons can be excited as clearly evident in an insulator with wide band gaps.³⁵ These are extreme cases inasmuch as their band gaps are so large that the density of unoccupied states becomes significant only at several electronvolts above the vacuum level. At the other extreme are transition metals with their unfilled d bands with high density of states just above the Fermi level. Here the IMFPs are very short as seen in Figure 8–7³⁷ in which the reciprocal values of the IMFPs are plotted as a function of the number of d holes. The values shown are for electrons with energies between 5 and 10eV above the Fermi energy. For Fe the IMFP is only about 0.5nm and for Gd only

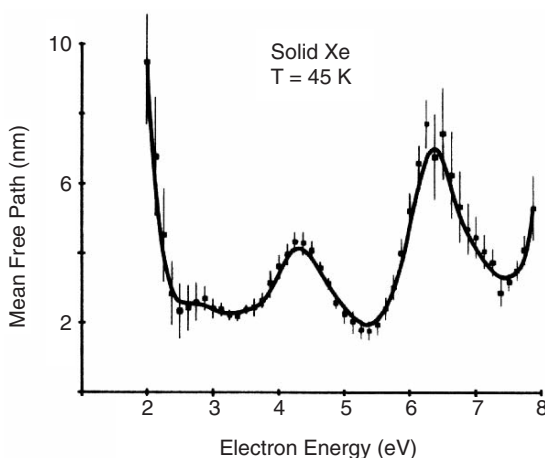


Figure 8–6. Energy dependence of the elastic mean free path of slow electrons in solid Xe.³⁶

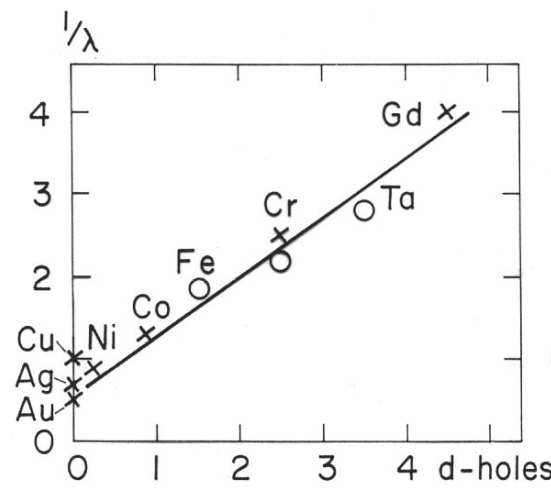


Figure 8-7. Reciprocal inelastic mean free path in nm⁻¹ of electrons with energies between 5 and 10 eV above the Fermi level as function of the number of holes in the 3d and 4d shell.³⁷

0.25 nm. In ferromagnetic materials the density of unoccupied states differs between majority and minority spin states, which causes corresponding differences in the excitation probabilities. Calculations that take this into account show a significant difference between the IMFPs of incident electrons with majority and minority spin as seen in Figure 8-8.³⁸ The zero of the energy is the Fermi energy; the work

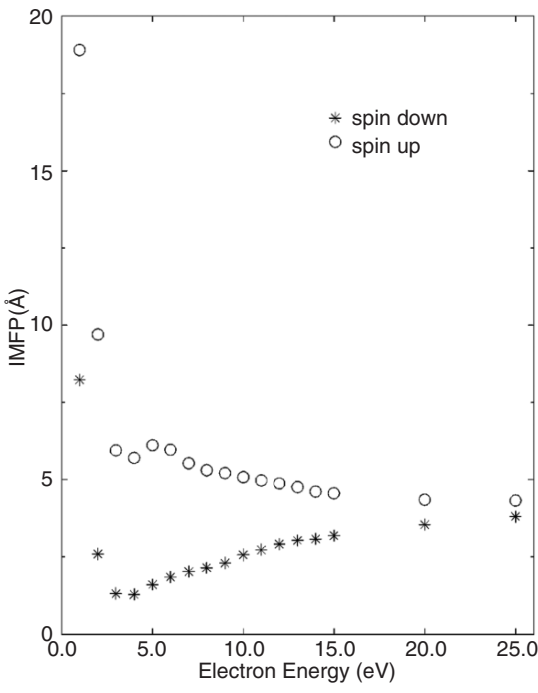


Figure 8-8. Energy dependence of the inelastic mean free path of majority and minority spin electrons in Fe.³⁸

function of Fe is 4.5 eV so that 1 eV above the vacuum level the IMFPs are only 0.6 and 0.2 nm for majority and minority spin electrons, respectively. At 10 eV above the vacuum level the corresponding values are 0.45 and 0.3 nm, which are much lower than the 1.6 nm obtained from the dielectric theory discussed above. For more information see Bauer.³⁹

While the knowledge of reliable absolute numbers for elastic and inelastic mean free paths at low energy is still limited, the influence of phonon and magnon excitation on the effective sampling depth is much less understood. Both processes involve only small energy losses up to several 100 meV but can occur with large momentum transfer. In LEEM only the electrons in a diffraction spot and its immediate environment contribute to the image formation. Therefore energy losses with momentum transfer larger than that determined by the radius of the contrast aperture cause an attenuation of the intensity contributing to the image formation. At not too high temperatures this can be taken into account by the Debye–Waller factor. At higher temperatures multiphonon and -magnon excitations occur that cause an increased attenuation, which may be described by an anharmonic Debye–Waller factor.⁴⁰ In this conventional description of the influence of phonons on the scattering from surfaces there is no thickness dependence. However, in thin films the number of electrons scattered outside the contrast aperture increases with increasing thickness so that an effective attenuation coefficient could be defined. This has not been done to date for several reasons: (1) there are no numbers for the cross sections for phonon and magnon scattering that could be compared with those for inelastic scattering, (2) with increasing temperature there is frequently atomic disordering that causes diffuse scattering, and (3) at high temperatures thin films usually break up into three-dimensional crystals before attenuation by these processes becomes significant. A rough idea of the influence of thermal vibrations on the attenuation length may be obtained from the analysis of LEED patterns from Cu single crystal surfaces. At 50 eV the total attenuation length, which includes elastic backscattering, inelastic scattering, and phonon scattering, from the (111) surface is 0.33 nm at 300 K versus 0.34 nm at 0 K.⁴¹ The difference is well within the limits of error of the values so that at least at this energy phonon scattering does not limit the sampling depth.

According to the present state of understanding the sampling depth of LEEM and SPLEEM is primarily determined by inelastic and elastic backscattering. Depending upon the energy and electronic structure of the material, the sampling depth may be as small as a few tenths of a nanometer, for example, around the plasmon excitation maximum, in transition metals with a high density of unoccupied states above the Fermi level, or in band gaps along the $\mathbf{k}$ direction normal to the surface. On the other hand, sampling depths as large as several nanometers can occur at very low energies, for example, in insulators and free electron like metals. In many cases this allows tuning of the sampling depth by proper choice of the energy, which makes LEEM and SPLEEM ideal for imaging of surfaces and thin films.

3 Instrumentation

The electron optics of an LEEM/SPLEEM instrument in the imaging section is essentially the same as in emission electron microscopes, which date back to the 1930s. In these microscopes the specimen is the cathode of a so-called cathode lens in which the slow emitted electrons are accelerated in a high field to the first of several image-forming electrodes of an electrostatic lens or to the entrance of a magnetic lens. This lens is the objective lens of the microscope, which produces a primary image with fast electrons. The subsequent electron optics is basically the same as in TEM. In fact, the first objective lens used in LEEM was a modified version of an electrostatic triode lens developed for PEEM.⁴² To do LEEM with such a system fast electrons have to be injected from the high-energy side of the objective lens along its optical axis. In the cathode lens they are decelerated to the desired low energy at the specimen. To be able to produce an image, this incident beam has to be separated from the reflected beam by a beam divider. As a consequence an LEEM or SPLEEM system has a bent optical axis. A schematic of the first instrument is shown in Figure 8–9.³ The beam separator (1) deflects the incident beam from a field emission gun (2) that is focused by two quadrupoles (3), the deflection field (1), and optionally by a collimator lens (4) into the back focal plane of the objective lens (5). They reach the specimen (7) on parallel trajectories with an energy that is determined by the adjustable potential difference between field emitter and specimen. The specimen is imaged by the elastically reflected electrons into the center of the beam separator, its diffraction pattern into the back focal plane of the objective lens where the angle-limiting contrast aperture is located. The astigmatism of the objective lens is corrected with

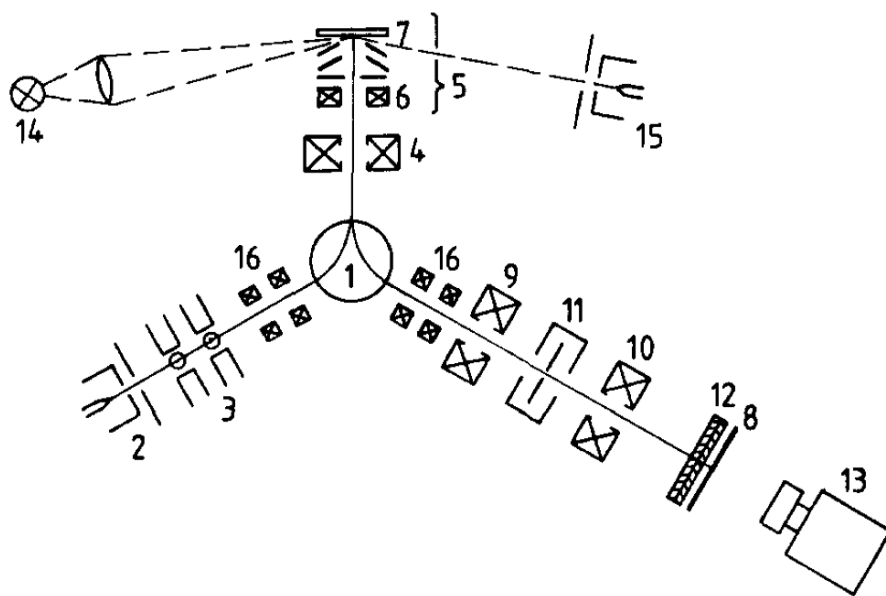


Figure 8–9. Schematic of the first LEEM instrument. For explanation see text.³

a magnetic stigmator (6). The primary image in the center of the beam separator is imaged with a magnetic intermediate lens (9) and a projective lens (10) onto the final screen, and the diffraction pattern by adjusting the focal length of the intermediate lens. The purpose of the electrostatic filter lens (11) was to filter out secondary and inelastically scattered electrons but was later removed because it was found that the dispersive properties of the magnetic beam separator (1) were sufficient to eliminate them from the image. A pair of multichannel plates (12) enhances the image intensity on the fluorescent screen (8), allowing observation and image recording with a video camera (13) at very low beam currents. Both illumination and imaging columns are equipped with deflectors and stigmators, some of which are indicated (16). Emission microscopy is possible with thermionic emission by heating the specimen, with photoelectrons generated by a 100-W high-pressure Hg arc lamp (14) and with secondary electrons using an auxiliary electron gun (15).

While Figure 8–9 shows the principle of the LEEM system, more recent instruments differ considerably in detail. For example, a transfer lens, which transfers the diffraction pattern from the back focal plane of the objective lens toward the front of the intermediate lens, is now inserted just behind the beam separator, so that the illuminating beam does not have to pass through the contrast aperture as in the original design. The electrostatic triode lens is now replaced by lenses with better resolution such as the electrostatic tetrode lens, the magnetic diode lens, or the magnetic triode lens, which will be briefly discussed below. The beam separator has been improved in a variety of ways, such as by close-packed or separated multiple magnetic prisms, concentric square or round pole pieces, or a Wien filter, resulting in deflection angles ranging from 16° to 90° compared to the original 60° . In addition, energy filters have been added to instruments so that they can also be used for AEEM, low electron energy loss microscopy (LEELM), energy-filtered SEEM, and spectroscopic PEEM.

Before discussing these components of an LEEM instrument, a short account of the various designs is appropriate. The first major development after the original instrument and a similar one⁴³ also used a beam separator with 60° deflection but with close-packed multiple magnetic prisms and only magnetic lenses, including the objective lens, a LaB_6 cathode, and a transfer lens so that the contrast aperture could be placed behind the beam separator.⁴⁴ The magnetic prism on the illumination side of the instrument can be excited differently from that on the exit side so that a higher beam energy is possible than on the imaging side, which allows AEEM when an energy filter is added to the instrument. The dense packing of the magnetic lenses together with the shielding of the beam separator and the specimen region makes the magnetic shielding used in the more open earlier instruments unnecessary. The addition of an energy filter allows not only AEEM but in combination with synchrotron radiation also spectroscopic X-ray PEEM.^{13,45,46} From this instrument the presently most widespread commercial instrument shown in Figure 8–10 developed later. The version with energy filter, the SPELEEM, which is presently the

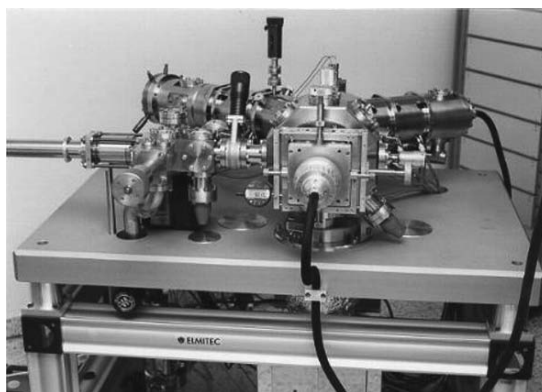


Figure 8–10. Commercial LEEM instrument. The illumination column is on the right side and the imaging column on the left side. The 60° beam separator is hidden behind the specimen chamber in the center, which also contains the objective lens. (Courtesy of ELMITEC GmbH.)

most versatile instrument, has been described repeatedly⁴⁷ in connection with PEEM. Therefore, a cross section of the presently best home-built pure LEEM instrument (Figure 8–11)⁴⁸ is shown here. Similar to the previous instrument (Figure 8–10) it has densely packed magnetic lenses both in the illumination (top) and in the imaging (bottom) section. However, it uses a beam separator with 90° deflection that is incorporated in the vacuum system (center). The objective lens (right

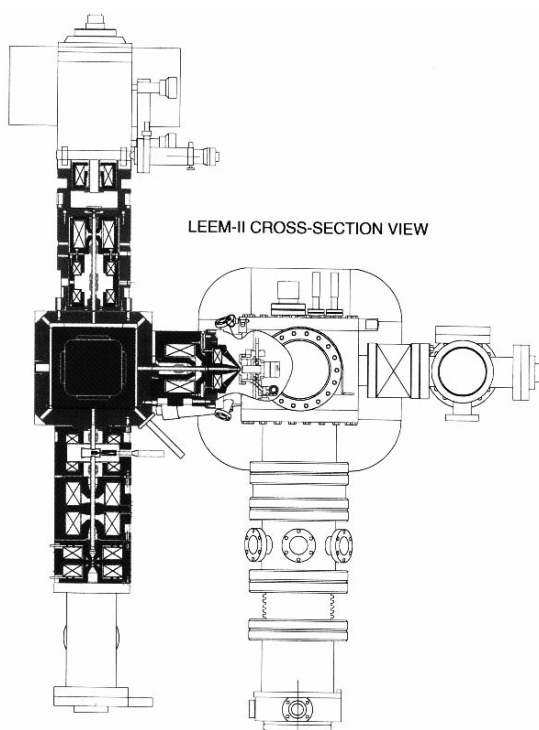


Figure 8–11. Cross section of a LEEM instrument with a 90° separator. The illumination column with the field emission gun is on top and the imaging column is at the bottom. The right side shows the specimen chamber, airlock, and part of the pumping system.⁴⁸

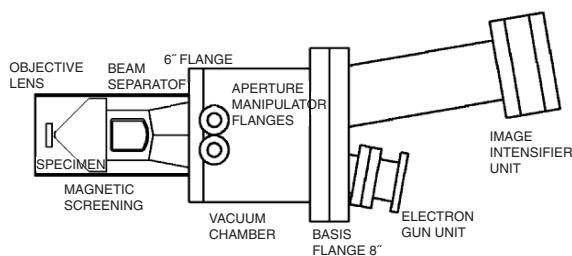


Figure 8–12. Schematic of the mechanical configuration of a flange-on LEEM system.⁴⁹

side) is a magnetic diode and consists of two sections with opposing image rotation. A commercial cold field emission electron gun (top) with an energy spread of about 0.25 eV allows a theoretical resolution of 0.4 nm at 10 eV. The right side is the specimen chamber and a preparation chamber plus airlock. This instrument presently holds the resolution record (0.5 nm) among the instruments without aberration correction.

The instruments described above are free-standing instruments. There has long been the wish to add a LEEM instrument to existing UHV systems. To be practical, such LEEM systems must be much smaller and have small deflection angles. Two solutions were chosen: one uses a simple beam separator with a small deflection angle (10°)⁴⁹ and the other one uses three deflections by 45° .⁵⁰ In both cases the instrument is mounted on an 8-in.-diameter UHV flange and can be attached to a UHV system via a 6-in.-diameter UHV flange. All lenses are electrostatic, including the electrostatic tetrode lens. In contrast to magnetic lenses, electrostatic lenses can be easily floated at high voltage. Therefore the complete electron optics can be at high voltage so that the specimen can be near ground potential, whereas in the magnetic lens systems the specimen is at high voltage. Because of the compact design, high-voltage insulation requirements allow final energies of only 5 keV, in contrast to the 15–20 keV used in the larger systems. Only the extraction electrode of the tetrode lens can be increased to 15 kV. External fields are screened by internal μ -metal screening. Figure 8–12 shows the mechanical configuration of one of these flange-on LEEM instruments.⁴⁹ Its overall length is 60 cm and its weight is about 20 kg.

Several other LEEM instrument designs have been proposed and in part realized. One design⁵¹ included some interesting features such as a combined magnetic-electrostatic beam separator that allows different energies to be used in the illumination and imaging beams, an electrostatic tetrode combined with a Schwarzschild-type optical mirror objective, which focuses UV radiation onto the specimen for PEEM, and a 70° spherical condenser for electron energy filtering. Unfortunately it never came into operation. In another design that is used in a commercial instrument⁵² the beam separation is achieved by a Wien filter into which the illumination beam enters at an angle of 36° while the imaging beam runs along its optical axis. A second Wien filter is used as an energy filter. This instrument has been used mainly for PEEM, MEM, and

metastable impact electron emission microscopy (MIEEM). Apparently the difficulty of aligning the incident beam normal to the surface and keeping the imaging beam on axis makes systematic LEEM studies with acceptable resolution difficult. A similar problem has to be overcome in an instrument in which the beam separator is replaced by a W single crystal.⁵³ This crystal is tilted 45° against the optical axis and reflects a low-energy electron beam from a side-mounted gun along the optical axis. Exact normal incidence on the sample requires that the reflector is on the optical axis, which would obstruct the imaging beam. This instrument, called a double reflection electron emission microscope (DREEM), is also commercially available. Other LEEM instruments have been built as well but their design has not been published.

All instruments discussed up to now suffer from the large chromatic and spherical aberration of the objective lens. That these aberrations can be corrected with an electron mirror has been known for some time but only during the past decade have efforts been made to develop aberration-corrected instruments. Simultaneously, however, the aberrations of the beam separator have to be corrected. This has led to the design of a complex system^{54,55} that has already been realized in the so-called SMART (Spectromicroscope for All Relevant Techniques) instrument,^{17,56,57} schematically shown in Figure 8-13. Similar to the SPELEEM system, the instrument is designed for a wide range of operation modes,⁵⁸ one of which is LEEM. The centerpiece is the beam separa-

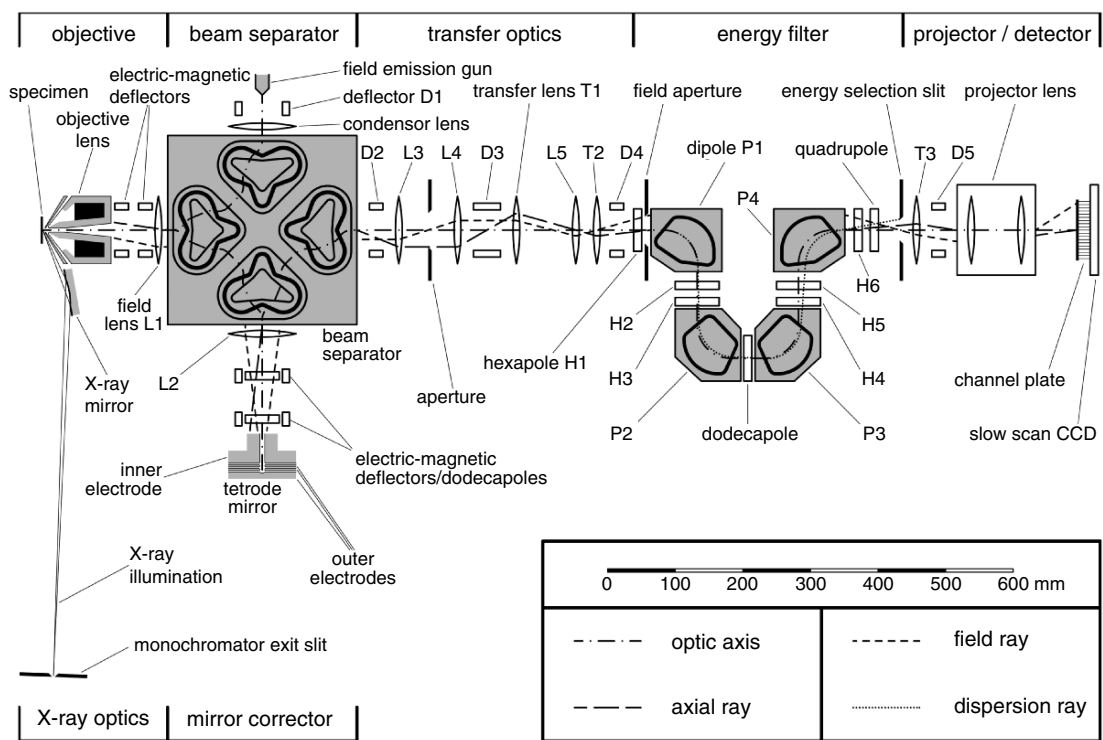


Figure 8-13. Schematic of the electron-optical configuration of the SMART system.⁵⁶

tor to which the illumination system, a field emission gun, the objective lens, and the mirror corrector are attached via field lenses (L). Beyond the beam separator five electrostatic lenses transfer either the diffraction pattern in the plane of the contrast aperture between L_3 and L_4 or the primary image at various magnifications into the entrance of the Ω -type energy filter where the field limiting aperture is located. With the energy selection slit inserted, the subsequent projective lens system produces an energy filtered image or diffraction pattern on the Channel plate-fluorescent screen unit, which is fiberoptically coupled to the CCD camera. Numerous deflectors (D_i) allow precise alignment and several n -pole elements ($n = 2, 6, 12$) are used for the correction of residual aberrations. The instrument operates at 15 kV and is expected to have a resolution in LEEM of 1 nm at 10 eV.

4 Electron Optics

Unless aberrations of the beam separator, the energy filter, other optical components or vibrations, electromagnetic fields, charging, or other disturbances are resolution limiting the chromatic and spherical aberrations of the objective lens determine the resolution. The aberrations of the accelerating field of the cathode lens cause much higher limiting values of the resolution than in transmission microscopy. These values can be calculated analytically by assuming that the lens may be separated into a homogeneous field in front of the specimen, which produces a virtual image behind the specimen, and an einzel lens, which produces a real image of the virtual image. Realistic calculations have to consider the cathode lens as a unit and have been made by many authors. A comparison of the resolution obtainable with an electrostatic triode, electrostatic tetrode, and magnetic triode lens (Figure 8-14) shows that the electrostatic tetrode and the magnetic triode lenses are much better than the original triode lens because in the latter the field strength at the sample is low under focusing conditions (1.18–0.52 kV/mm vs. 10 kV/mm in the other two lens types) (Figure 8-15).⁵⁹ The data are for the optimum aperture, which is determined by minimizing the contributions of the aberration discs due to diffraction at the angle-limiting aperture, chromatic and spherical aberrations. These contributions can be seen in Figure 8-16 together with the radius of the optimum aperture. From these figures it is clear that the magnetic triode is superior to the electrostatic tetrode both in resolution and transmission. Transmission does not play an important role in LEEM but is important in XPEEM with photoelectrons and secondary electrons, which have a wide angular distribution. Today the magnetic diode, which differs from the triode only in that both pole shoes are at the same potential, is the standard in the best LEEM instruments without aberration correction while the electrostatic tetrode is the domain of the smaller, purely electrostatic LEEM instruments. In the multimethod instruments^{15-17,44,56,57} combined electrostatic-magnetic cathode lenses, such as the magnetic triode, are useful because they allow the field strength at the specimen to be varied. Such a lens with a design somewhat different from that shown in Figure 8-14 is used in the

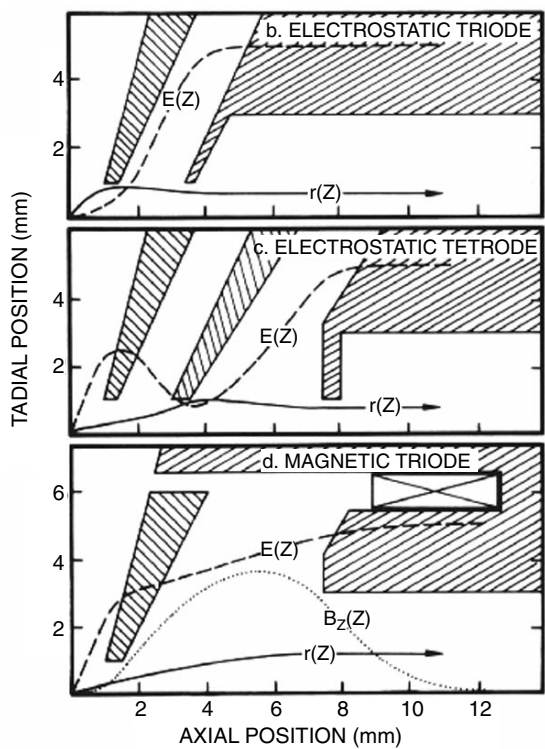


Figure 8-14. Schematic configurations of LEEM cathode lenses. One-half of the electrodes/pole pieces, the electron energy $E(z)$, and the electron ray path $r(z)$ is shown.⁵⁹

SMART instrument. Because the chromatic and spherical aberrations of the objective lens are corrected in this instrument by the mirror corrector, the largest aberrations are now those of the energy filter. The resolution improvement obtained by correction is shown in Figure 8-17 as a function of angular acceptance for a start energy of 10 eV, an energy width of

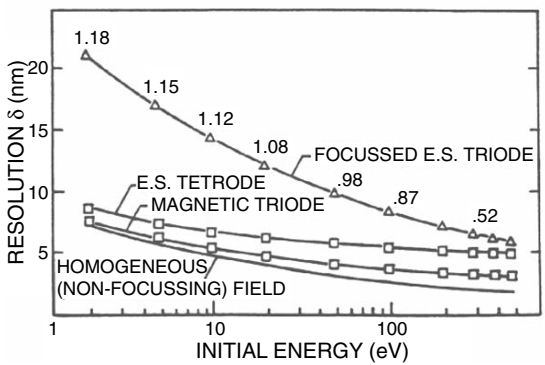
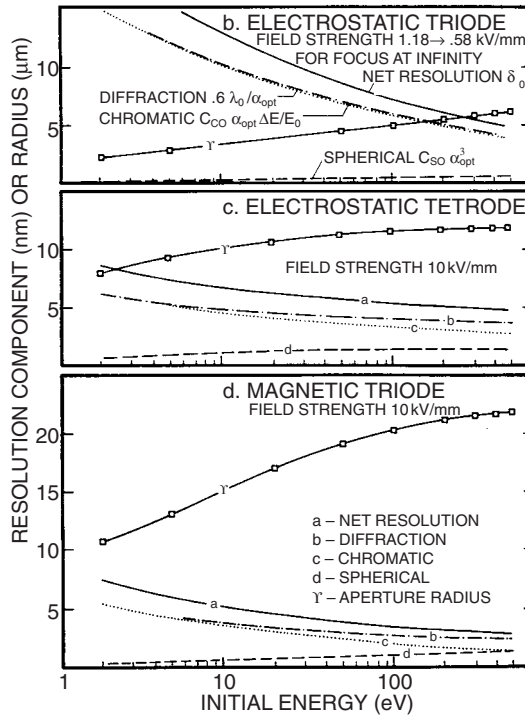


Figure 8-15. Comparison of the resolution of the lenses shown in Figure 8-14 for a final energy of 20 keV, an energy spread of 0.5 eV, and optimized aperture.⁵⁹

Figure 8-16. Energy dependence of the contributions of the spherical, chromatic, and diffraction aberrations (dashed, dotted, and dash-dotted lines) to the resolution (solid line) at the optimum aperture with radius r for the lenses shown in Figure 8-14. In the triode the field strength has to be changed for forming a real image; in the other two lenses 10 kV/mm has been chosen. Energy spread 0.5 eV.⁵⁹



2 eV, and a final energy of 15 keV.⁵⁷ The dashed lines show the contributions of the aberrations of the uncorrected lens and the thin solid lines those of the energy filter. For LEEM instruments with LaB₆ or field emission guns that have lower energy widths, the resolution is still improved by about a factor of 5. The main advantage of aberration correction, the large increase in transmission, however, comes to bear only in emission microscopy, in particular in AEEM and XPEEM.

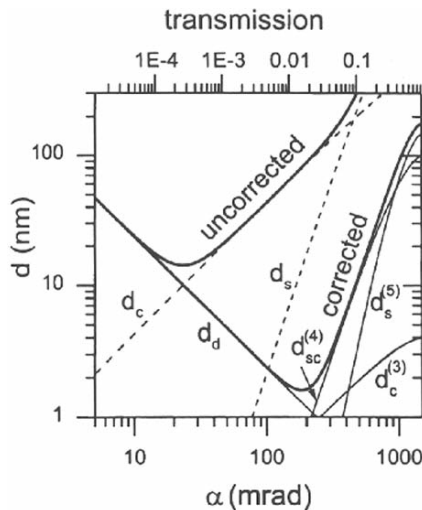


Figure 8-17. Resolution limit d as a function of the acceptance angle α without and with correction of the spherical and chromatic aberrations. Data for the system shown in Figure 8-13 for 10 eV start energy and 2 eV energy spread. In the uncorrected system d is limited by the chromatic and spherical aberrations (dashed lines) and in the corrected system by higher order aberrations (solid lines).⁵⁷

The next critical component of a LEEM instrument is the beam separator. Early beam separators^{3,43} aimed only at the reduction of the unidirectional beam dispersion in the deflection plane. This was achieved with a D-shaped cutout in the round pole pieces.⁶⁰ The focusing action of the fringing field of the magnet was compensated by magnetic quadrupoles. In contrast to these first separators, which tried to eliminate its focusing action, later separators made use of it in order to obtain optimum image and diffraction pattern transfer in close-packed magnetic prism arrays. They consist of an array of inner pole pieces surrounded by a single outer pole piece with different relative excitations. Such arrays act almost like round lenses. They transfer image and diffraction planes stigmatically and distortion free to corresponding planes behind the separator but at different settings.^{61,62} A 90° deflector with four inner prisms was suggested for the addition of a mirror corrector⁶² and a 60° deflector with three inner prisms was realized in the first fully magnetic LEEM instrument.⁴⁴ If illumination and imaging beam have the same energy as in LEEM—in contrast to the more versatile instruments—then a single inner pole surrounded by an outer pole ring are sufficient. Both square^{48,54,63–65} and round 90°^{64,66} separators have been proposed and built. With proper geometry and excitation ratio, astigmatic and distortion-free imaging can be achieved for image and diffraction plane with the same settings. For an aberration-corrected microscope the aberrations of these separators would limit the resolution and have to be corrected. This is achieved in the highly symmetric beam separator seen in Figure 8–13.^{67–69} In small flange-on LEEM instruments the design of the beam separator is determined less by minimizing aberrations than by geometry and space considerations.^{49,50} For example, to achieve parallel illumination and imaging beams within a minimum of space, the beams were translated achromatically⁷⁰ by magnets with field boundaries perpendicular to the beam and electrostatic cylinder lenses in order to achieve stigmatic focusing.⁵⁰ Finally a Wien filter may also be used as a beam separator⁷¹ though with some alignment difficulties.

Although the use of electron mirrors for the correction of lens aberrations was proposed many years ago^{72,73} and suggested for the correction of the cathode lens aberrations in LEEM and PEEM instruments in the early 1990s^{54,61,74} serious efforts have been made only in the past 10 years^{75–77} and only one instrument, the SMART, is presently equipped with a mirror corrector and is in its testing stage.

An energy filter is not needed in simple LEEM instruments with sufficiently large separator deflection angles because at low energies, at which the energy of the secondary electrons differs little from that of the elastically reflected electrons used for the LEEM image, the secondary electron intensity is small compared to the diffracted beam intensity. With increasing energy, when the elastically backscattered intensity decreases and the secondary electron intensity increases, the dispersive action of the sector field deflects the secondary electrons sufficiently so that only a small fraction passes through the contrast aperture. Likewise, inelastic scattering in the high LEEM energy range occurs mainly in the forward direction so that it can be seen in the backward direction

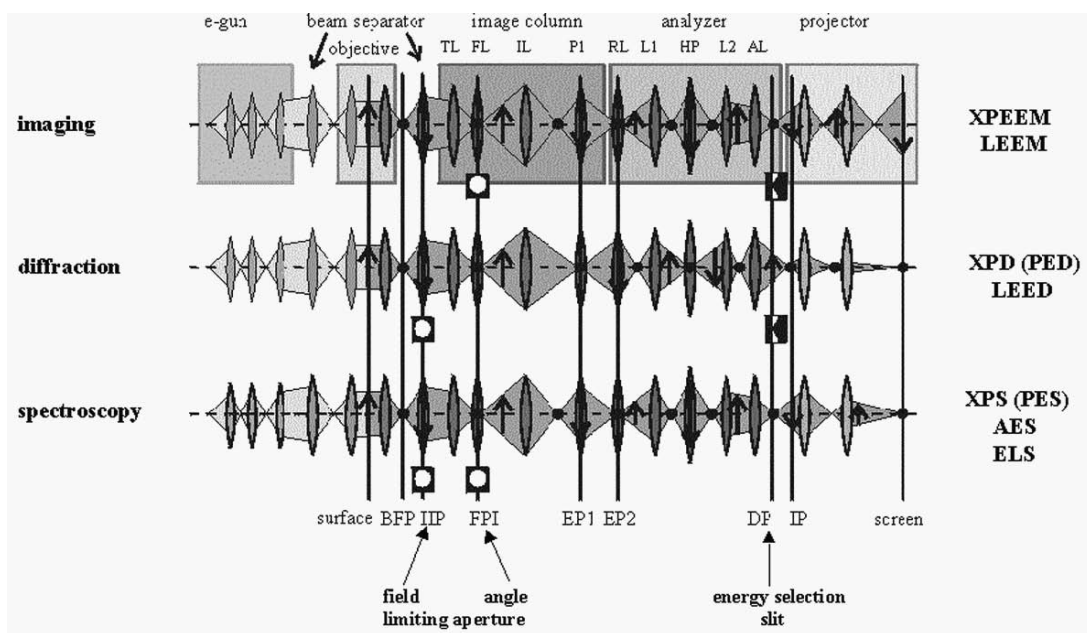


Figure 8-18. The three fundamental operation modes of a SPELEEM system. The various sections of the instrument are shown folded into one plane. In imaging and diffraction the energy selection slit is inserted in the dispersive plane DP and the image/diffraction pattern behind DP is imaged with the projector. The intermediate lens IL is used to switch between imaging and diffraction, simultaneously with the exchange of the contrast aperture in FPI and the field-limiting aperture in IIP. For fast spectroscopy both apertures are inserted, the energy selection slit is removed, and the dispersive plane is imaged by the projector.¹⁵ (See color plate.)

only through diffraction, either via energy loss before or after diffraction. Therefore it is a second-order effect. Furthermore, the dispersion of the beam separator and the momentum transfer in the energy loss ensure that most inelastically scattered electrons do not pass through the contrast aperture. In the low LEEM energy range inelastic scattering is less in the forward direction but in general is weak, so that it does not contribute noticeably to image formation. The main advantage of having an energy filter in a LEEM instrument is that it eliminates the secondary and inelastically scattered electrons in the LEED pattern. In many cases this is not important but in materials with high secondary electron yield it improves the LEED pattern dramatically. Also, for quantitative analysis of the background in LEED patterns an energy filter is useful.

The main motivation for adding an energy filter to a LEEM instrument is its importance in multimethod instruments such as the SPELEEM or the SMART. It allows LEEM to be combined with low-electron energy loss spectroscopy and microscopy. Furthermore, when higher beam energy can be used in the illumination beam than in the imaging beam, Auger electron emission spectroscopy (AEES) and AEEM are possible too. These various operation modes of such an instrument are schemetically shown in Figure 8-18.^{15,47} In SEEM, whether excited by electrons, X-ray photons, or energetic ion/neutrals, it allows selection of a narrow energy window from the wide

secondary electron energy distribution. This leads to a noticeable resolution improvement, also in the aberration-corrected systems, in which the correction rapidly deteriorates at low energies with decreasing energy.⁵⁷ A number of different energy filters are used in these multi-method instruments: one hemispherical analyzer, two hemispherical analyzers, a Wien filter, or an Omega filter. As they are not essential for LEEM they will not be discussed here.

5 Contrast

For imaging with LEEM several contrast mechanisms are available, depending upon the specimen to be imaged. The fundamental contrast is diffraction contrast in crystalline samples or backscattering contrast in amorphous or fine-grained crystalline materials. The origin of the backscattering contrast is evident from Figure 8–1. An example of its consequences is shown in Figure 8–19. Because of the higher backscattering cross section of Co at the selected energy the fine-grained polycrystalline Co squares appear bright compared to the Si surrounding, which is covered with native oxide. A slight preferred orientation of the Co layer enhances the contrast.⁷⁸ In general, however, larger crystals or single crystalline layers with a well-defined orientation are studied. In this case not only the specular beam [(00) beam] but other diffracted beams may be used for imaging. An atomically flat single crystal surface without steps and other defects produces contrast only when regions with different crystal structure are present. A standard example is the Si(111) surface when the unreconstructed (1×1) and reconstructed (7×7) structure coexist. Here, both normal and lateral periodicity differ and produce strong contrast (Figure 8–20a).⁷⁹ The Si(100) surface reconstructs with the formation of dimer rows whose orientation rotates from terrace to terrace by 90° , with constant normal periodicity. Here, the two resulting (2×1) domains are equivalent at normal incidence and well-centered aperture. Contrast is obtained by either tilting the incident beam or shifting the aperture somewhat in the direction of one of the rows. Using the $\frac{1}{2}$ order spot of one of the domains gives maximum contrast (Figure 8–20b).⁸⁰ Similar domain contrast can be obtained on all

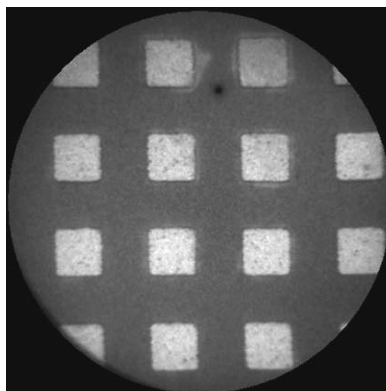


Figure 8–19. Backscattering contrast from 20-nm-thick Co squares on an Si substrate. The electron energy is 5.1 eV and the diameter of field of view is $10\mu\text{m}$.⁷⁸

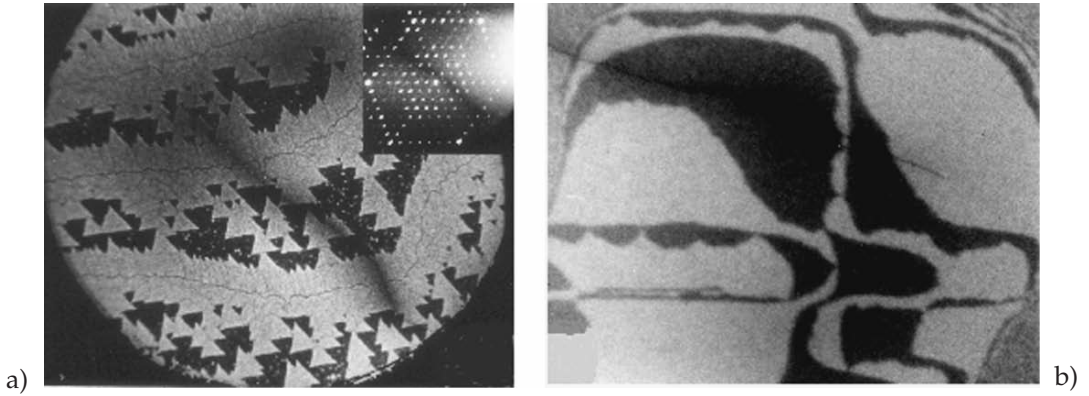


Figure 8-20. Diffraction contrast from Si surfaces. a) Si(111). (Normal incidence) contrast due to different normal periodicity of coexisting (1×1) (dark) and (7×7) (bright) structure. Electron energy 10 eV, diameter of field of view $6 \mu\text{m}$. Inset: LEED pattern.⁷⁹ b) (Oblique incidence) contrast due to different azimuthal orientation of coexisting (2×1) and (1×2) domains. Electron energy 6 eV, image width $3 \mu\text{m}$.⁸⁰

reconstructed surfaces on which reconstruction domains exist with different azimuthal orientations. All surfaces have steps or step bunches that will produce another contrast to be described below.

In general surfaces are heterogeneous not only in crystallography but also in composition and topography. Composition differences are usually connected with crystallographic differences and produce, together with backscattering differences, diffraction contrast in the (00) beam. Also here imaging with nonspecular beams is useful for identifying different coexisting phases as illustrated in Figure 8-21,⁸¹ which is from an Si(111) surface covered with a submonolayer of Au. The three LEEM images are taken with the (00) beam and with nonspecular beams ($1/5$ -order beams) of (5×2) superstructure domains. This allows the identification of the dark regions between the bright $(\sqrt{3} \times \sqrt{3})$ -R30° structure regions in the specular image with different (5×2) domains.

Topography distorts the electric field distribution on the surface. This causes the usual topographic contrast, which is most evident near zero electron energy. Topography also produces diffraction contrast. This happens when surface elements are inclined against the average surface, for example, in small crystals on an otherwise flat surface. In a LEEM instrument the positions of the LEED spots from a flat surface do not change with energy as they do in an ordinary LEED system. This is due to the fact that the observed LEED pattern is a magnified image of the LEED pattern in the back focal plane of the objective lens, where the electrons have a constant high energy $E = (\hbar^2/2m)k^2$ independent of their start energy. Because the wave number k is proportional to the refractive index n , we have

$$k \sin \theta = k_0 \sin \theta_0 \quad (2)$$

where θ and θ_0 are the angles of the electron trajectories with the optical axis in the back focal plane and at the surface, respectively. For two-dimensional diffraction

$$k \sin \theta_o = 2\pi h \quad (3)$$

where h is the distance of the LEED spot $h = (h_1, h_2)$ from the (00) beam, which is on the optical axis at normal incidence. Combining the two equations leads to

$$k \sin \theta = 2\pi h \quad (4)$$

Thus the angular distance θ of the LEED spot h in the back focal plane is independent of the start energy $E_o = (\hbar^2/2m)k_o^2$ and depends only on the final energy $E = (\hbar^2/2m)k^2$. This is no longer true when the surface normal is inclined against the optical axis. Then the specular beam is off-axis and the diffracted beams h move toward the specular beam with increasing energy. The simple geometric relations at normal incidence, which lead to E_o -independent spot positions, are no longer valid so that the spots move now in the back focal plane. An example of these spot movements is shown in Figure 8–22.⁸² A simple geometric analysis allows deduction of the inclination of the surface.

Faceted surfaces, that is surfaces on which all surface elements are tilted, so that no specular beam is on the optical axis, can be imaged either by selecting an energy at which one of the diffracted spots passes

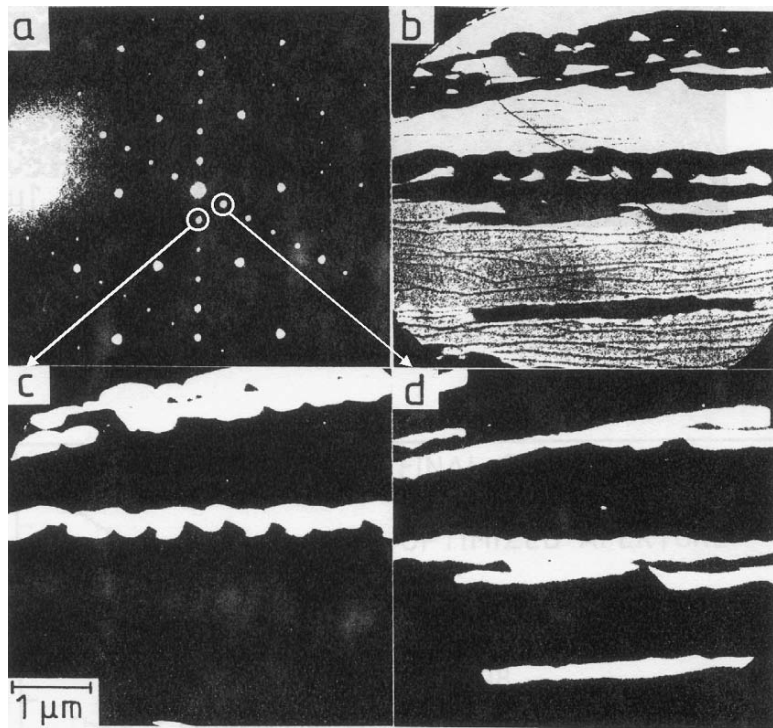


Figure 8–21. Phase identification by dark-field imaging. The LEED pattern (a) of the Au submonolayer on Si(111) shows a hexagonal pattern from the $(\sqrt{3} \times \sqrt{3})$ -R30° phase and two linear patterns from the (5×2) phase. The bright-field image (b) shows dark regions that are identified as (5×2) regions by imaging with $(1/5, 0)$, spots (c and d). Electron energy in (a) 30 eV and in (b–d) 6 eV.⁸¹

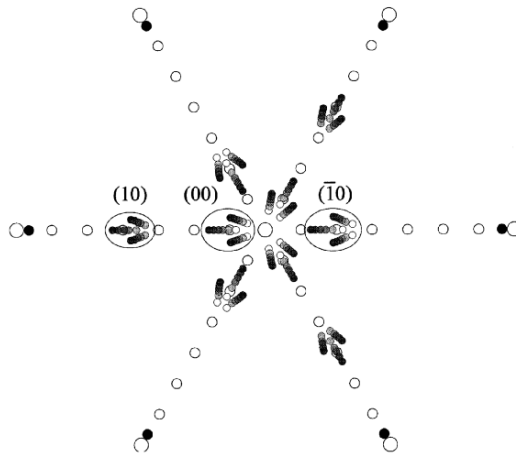


Figure 8-22. Drawing of the movements of LEED spots from faceted Cu silicide crystallites on a Si(111) surface. The open circles are from the Si(111)- $\delta(7 \times 7)$ structure. The small solid and shaded circles are from the crystallites. Increasing shading shows the movement of the spots with energy increasing from 3 to 10.5 eV.⁸²

through the optical axis or by tilting the illuminating beam or by shifting the contrast aperture off axis into one of the specular beams. For large tilt angles only the first mode is practical.

Another important contrast mechanism is the interference contrast on flat surfaces with height differences such as atomic steps. The step contrast was already observed in the early studies (Figure 8-23¹) and attributed to destructive interference between the wave fields reflected from the adjoining terraces within the lateral coherence length (Figure

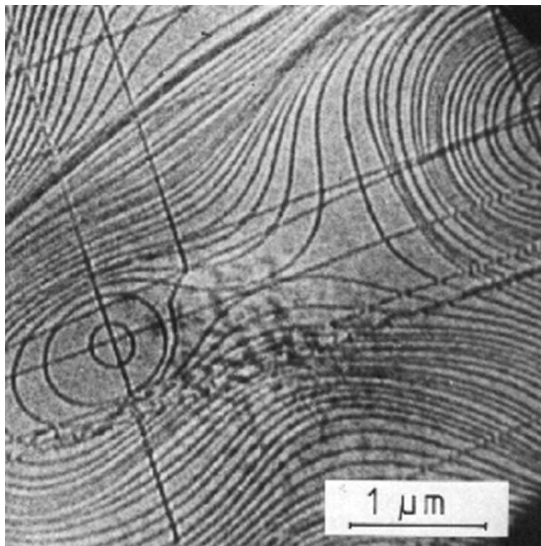


Figure 8-23. Monatomic steps on an MO(110) surface. Electron energy 14 eV.³

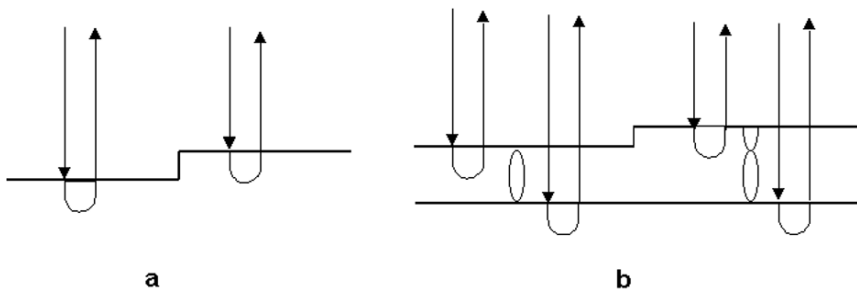


Figure 8–24. Conditions for phase contrast in LEEM. (a) Step contrast. (b) Quantum size contrast. The penetration of the electron wave upon reflection is indicated.

8–24a). Detailed model calculations based on Fresnel diffraction from two adjoining straight edges shifted relative to each other by the step height produce all salient features of the step contrast.^{83,84} Here some results of the general theory of image formation by a typical magnetic cathode lens will be given.⁸⁵ In the absence of aberrations the reflection of slow electrons from a point source would produce an interference pattern that extends far out from the step. This would make image interpretation in the presence of several steps difficult. The spherical aberration reduces the range of the interference pattern significantly and the chromatic aberration reduces it to one intensity maximum next to the step at energy spreads as low as 0.5 eV. The intensity distribution around the step depends upon the phase difference between the waves reflected from both sides of the step, that is upon the step height and the wave length, and upon the defocus. This is illustrated in Figure 8–25⁸⁵ for two phase shifts $\Delta\phi = n\pi$ ($n = 0.5, 1$) and several defocus values $\Delta z^* = \Delta z(C_s\lambda)^{-1/2}$, where Δz is the geometric defocus, C_s the spherical aberration constant, and λ the wavelength. $\Delta z^* = 0, 1$ corresponds to the Gaussian image and to the Scherzer focus, respectively. For integer n the step contrast is symmetric and optimum at $\Delta z^* = 0$; for noninteger n it is asymmetric, with the bright edge changing from one side of the step to the other when the sign of the defocus changes. Optimum contrast is achieved for slight defocus.

A third contrast mechanism, the quantum size contrast, is also based on wave interference, which not necessarily requires crystal periodicity. In a thin film bounded by two parallel surfaces the wave reflected from the bottom surface can interfere constructively or destructively with that reflected from the top surface similar to a Fabry–Perot interferometer, depending upon the wavelength λ , the thickness t , and the phase shifts ϕ upon reflection at the surfaces (Figure 8–24b). Constructive interference and therefore enhanced reflectivity occurs whenever $n(\lambda/2) + \phi = t$, where n is an integer and λ is the wavelength in the film, which differs from the vacuum wavelength by the inner potential. As a consequence, regions with different thickness appear in the image with different brightness. This was first observed in Cu films on Mo(110)⁸⁶ and has been studied since in detail in several other systems with the goal to determine the band structure $k(E)$ above the vacuum

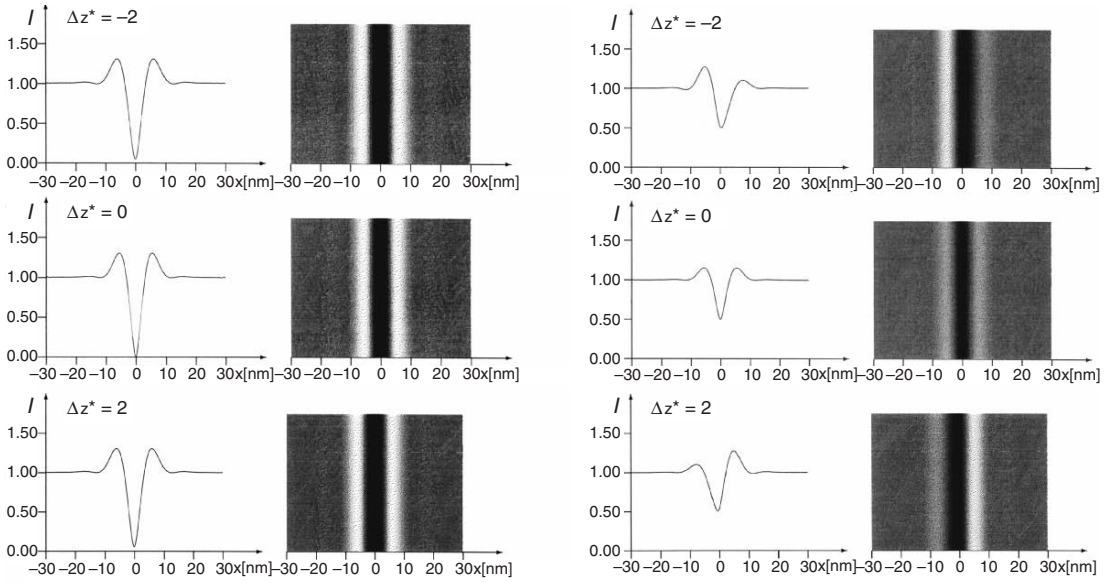


Figure 8–25. Step contrast for the phase differences $\Delta\varphi = \pi$ (left) and $\Delta\varphi = 0.5\pi$ (right) between the waves reflected from the terraces next to the step for zero defocus and small positive and negative defocus.⁸⁵

level,^{83,87,88,89} spin-dependent electron reflectivity effects,⁹⁰ or to understand specific features in thin film growth.^{91,92,93}

To determine the band structure, the constructive interference condition above is rewritten by replacing λ by $k = 2\pi/\lambda$, which leads to $k(E)t - k(E)\phi(E) = n\pi$. The energy-dependent phase term can be eliminated by choosing film thickness pairs t_1, t_2 for which this condition is fulfilled (with different n), which gives a set of $k(E)$ values. With proper growth conditions regions with different thickness can be obtained (Figure 8–26⁸⁷) and analyzed quasimultaneously. After subtraction of

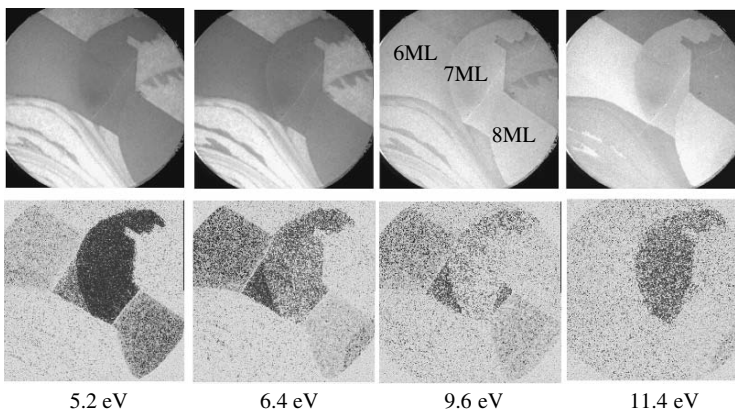


Figure 8–26. Quantum size contrast between regions with different thickness of an Fe film on W(110), taken with different electron energies, that is wavelengths. The images in the top row show the intensity and those at the bottom the magnetic signal (exchange asymmetry). Blue and red correspond to opposite magnetization directions.⁸⁸ (For the bottom row, see color plate.)

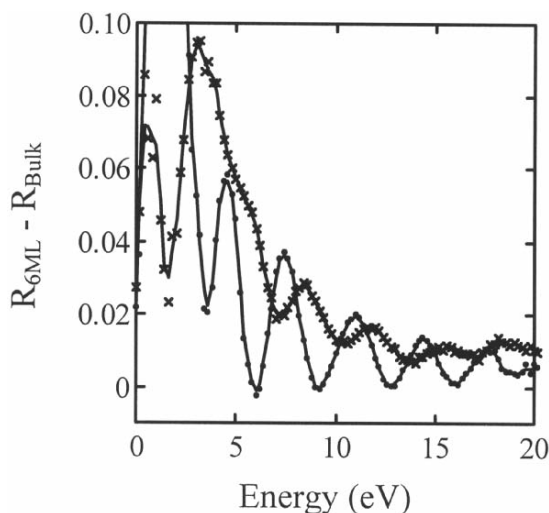


Figure 8-27. Quantum size oscillations of the reflectivity of spin-up (•) and spin-down (*) electrons in a six-monolayer-thick Fe film on W(110) as a function of energy.⁸⁸

the reflectivity of a thick film, which does not show quantum size effects, the oscillations of the reflectivity due to constructive and destructive interference can clearly be seen. This is illustrated in Figure 8-27⁹¹ for a ferromagnetic film in which the band structures of the majority and minority spin electrons differ by the exchange splitting. The reflectivity curve for the minority spin electrons is shifted relative to that of the majority electrons due to this splitting and is also damped more strongly due to the shorter IMFP of the minority electrons mentioned in Section 2.

6 Applications

6.1 General Comments

The high intensity available in LEEM studies of single-crystal surfaces, which allows rapid image acquisition, the high surface sensitivity, which strongly accentuates the topmost layer in imaging, both discussed in Section 2, and the various contrast mechanisms described in Section 5 have made LEEM one of the most powerful methods for surface studies, in particular of the thermodynamics of surfaces and of the kinetics of surface processes. While most of this information came from detailed studies of semiconductor surfaces, mainly from Si(111) and Si(100) surfaces, important insight into surface processes have also been obtained from various metal surfaces and oxide surfaces such as the TiO₂(110) surface. The information obtained from such studies ranges from the chemical potential of adatoms, diffusion across terraces and steps, anisotropic step free energy, step stiffness, step mobility, step-step interactions, surface free energy and surface stress, vacancy exchange between the bulk and the surface to nucleation, growth, phase transi-

tions, self-organization, faceting, segregation, oxidation, and other surface and thin film phenomena. Only some examples can be mentioned in the following subsections. These are organized according to the material, but the references will provide access to most of the relevant work done up to now. For illustrations results of the early, exploratory work will be used because the later quantitative studies require much more discussion.

6.2 The Si(111) Surface

The Si(111) surface is probably the surface most studied with LEEM, mainly because of its phase transition from the reconstructed (7×7) to the disordered “ (1×1) ” at 1100 K or 1135 K, depending upon author. In precise LEED diffractometer measurements,⁹⁴ see also⁹⁵ the superstructure spots disappeared at 1120 K and an intensity fit assuming a continuous transition gave a critical temperature of $1100\text{ K} \pm 1\text{ K}$. However, no critical scattering was observed, which put into question earlier conclusions that the phase transition was second order. The first LEEM measurements^{79,80,96} demonstrated without doubt that the transition was first order as seen in the nucleation and growth of the (7×7) structure (Figure 8–28). The growth rate of the (7×7) domains was found to increase linearly with undercooling ΔT and for $\Delta T > 12\text{ K}$ nucleation also occurred on the terraces.

The transition was found to be strongly influenced by impurities.⁹⁷ In particular, the apparent discrepancy between LEEM and the preceding LEED studies could be attributed to near-surface contamination during the long measurement time near the phase transition needed in the quantitative LEED studies. This is illustrated in Figure 8–29,⁸⁰ which shows that long annealing near the transition temperature

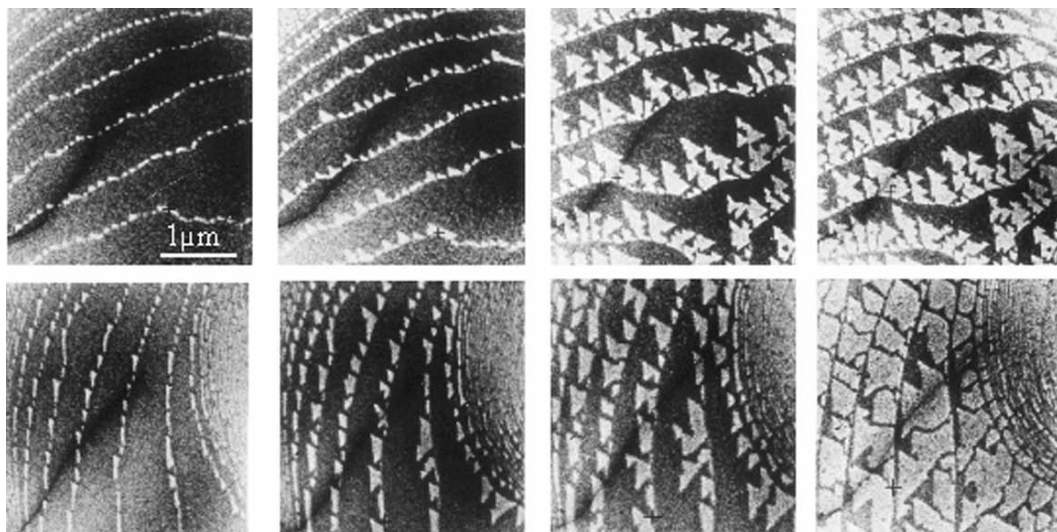


Figure 8–28. Nucleation and growth of the (7×7) structure at surface steps with different orientations at low supersaturation. Electron energy: top 10.5 eV and bottom 1.5 eV.⁹⁶

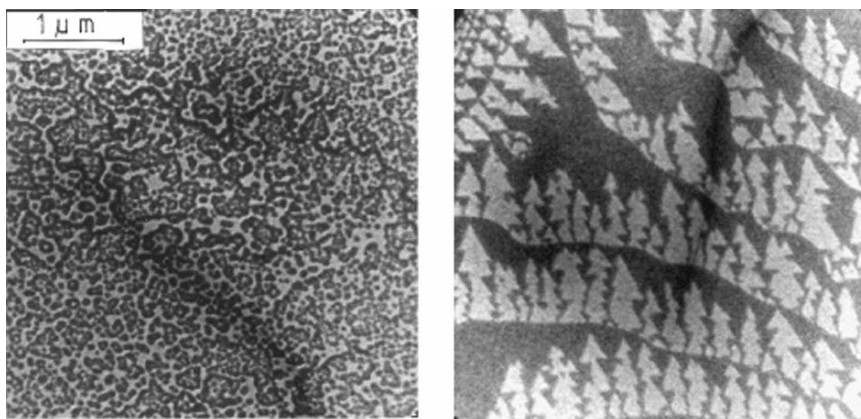


Figure 8-29. Comparison of a surface that had been annealed for a long time around the transition temperature (a) and one that has been cooled rapidly from 1450 K to this temperature (b). Electron energy 10.5 eV.⁸⁰

completely destroys the regular domain structure. The LEED patterns differ only by a slightly higher background in the annealed sample but the transition range is now much wider, similar to that in the LEED studies. On clean surfaces that have been quenched rapidly and have converted completely into the (7×7) structure many domains with various sizes form. Upon subsequent annealing they coarsen without preference of certain boundary orientations and number of bounding domain walls.⁹⁸

Recent, more detailed studies⁹⁹⁻¹⁰⁹ have shed considerable light on the forces and processes involved in the phase transition, such as adatom diffusion,^{99,100,104} the influence of the surface stress difference between the (7×7) and (1×1) phases and of long-range interactions on phase coexistence,¹⁰¹ shape,^{105,109} and distribution of the (7×7) domains¹⁰⁶ and other aspects. There are excellent reviews on these subjects^{110,111} where details may be found. Another phenomenon that has been studied with LEEM and that is closely related to the (1×1) to (7×7) phase transition upon cooling is the faceting of vicinal (stepped) Si(111) surfaces.¹¹²⁻¹¹⁴ Other studies have been concerned with the conditions for step flow growth instead of two-dimensional nucleation from which the parameters that determine the growth kinetics can be derived.¹¹⁵⁻¹¹⁷ In and Sb surfactants that form a $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure were found to either enhance or to suppress step flow, respectively.¹¹⁸ An apparently similar boron-induced surface structure [$(\sqrt{3} \times \sqrt{3})R30^\circ\text{-B}$] has a quite different effect: it causes twinning.¹¹⁵

6.3 Si(100)

This surface is the basis of semiconductor technology and has, therefore, attracted particular attention. It was studied qualitatively in the early years of LEEM⁸⁰ (see Figure 8-20b) and showed convincingly the inequivalence of the A and B type steps, the lower energy S_A steps being smooth while the higher energy S_B steps were rough. Step migra-

tion during sublimation¹¹⁹ and interaction with dislocations formed by plastic deformation during cooling¹²⁰ were studied as well as the enhancement of one domain upon elastic deformation.^{80,120} Other processes included consecutive “Lochkeim” formation during sublimation of flat regions¹²¹ and homoepitaxial growth.^{80,121–123} From the terrace shape during growth close to equilibrium (Figure 8–30) a lower limit of the ratio of the step free energies of S_B and S_A of $\beta_B/\beta_A \geq 2.6$ at about 800 K was obtained. In other qualitative work^{124,125} the step morphology was studied in more detail as a function of miscut angle, which led to a step “phase diagram” ranging from a “hilly” phase near zero miscut via single height wavy steps, straight steps to double height steps at a miscut of about 0.1° .

In subsequent quantitative studies^{126–137} comprehensive information could be deduced from step and island shapes and distributions. Many of the results can also be found in the reviews mentioned above.^{110,111} Some of them are the determination of the mobility and stiffness of the S_A and S_B steps,^{126,130} of their free energy,¹²⁶ and of the anisotropy of the surface stress;¹³⁷ the extraction of the chemical potential, formation energy, and diffusion coefficients of adatoms from number, area, and distribution of two-dimensional islands.^{127,131,134} On the more qualitative side, the fabrication of large step-free regions^{129,136} and of periodic gratings^{132,133,135} by e-beam lithography, reactive ion beam etching, and high-temperature annealing has also contributed considerably to the understanding of the surface properties. Other methods of surface modifications have been studied as well. Oxygen etches the surface at high temperature and produces vacancy islands.^{138,139,140} Arsenic was found to displace Si even on large terraces, driven by surface stress,

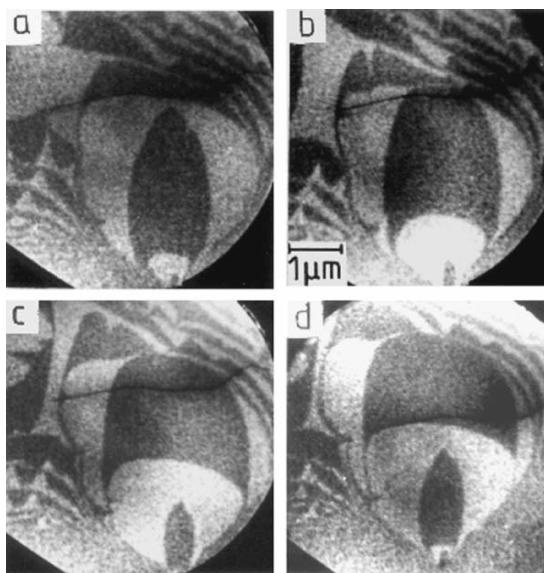


Figure 8–30. Images from a video taken during the homoepitaxial growth of Si(100) at low supersaturation so that growth can occur only from a defect at the lower edge of the image. Electron energy 5 eV.¹²²

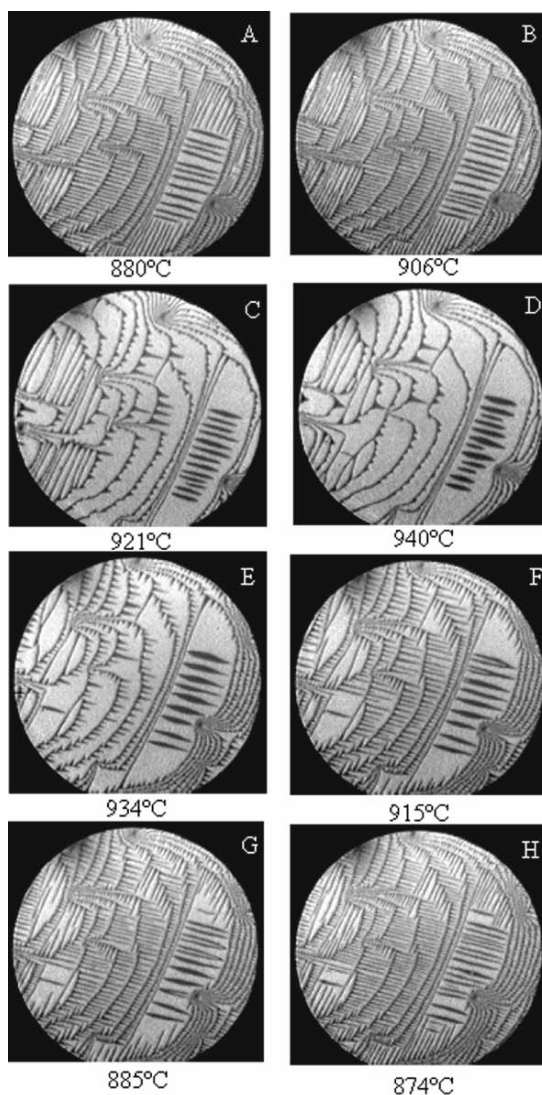


Figure 8-31. Images from a video taken during the segregation and desegregation of B from B-doped Si(100) during cooling and heating. Electron energy 4.2 eV; diameter of field of view 7 μm .

causing two-dimensional island formation.¹⁴¹ Boron segregation leads to strong temperature-dependent surface roughening at the monolayer level, forming a striped (Figure 8-31) or triangular-tiled surface structure.¹⁴²⁻¹⁴⁵ Originally believed to be driven mainly by surface stress relaxation¹⁴² it was shown later that a strong reduction of the step free energy of the S_A steps was the main driving force.^{143,144}

6.4 Other Elemental Semiconductor Surfaces

On the silicon-on-insulator (SOI) (100) surface LEEM was used to determine the dislocation-induced strain.¹⁴⁶ On the Si(311) surface the LEEM image intensity fluctuations in small surface regions were measured during the continuous disordering transition of the (3×1) reconstruction around 965 K, in order to determine the critical parameters.¹⁴⁷

The first-order transition at about 1005 K from the high-temperature “ (1×1) ” to the low-temperature “ (16×2) ” structure of the Si(110) surface and the stress-stabilized coexistence of the two phases was the subject of another study.¹⁴⁸ Similar to the work on Si(111) and Si(100), various surface thermodynamic data have also been obtained for the Si(110) surface from island decay measurements.¹⁴⁹ Finally, an LEEM study of the (2×1) to (1×1) phase transition above 925 K on Ge(100) led to the conclusion that the transition was due to dimer breakup and roughening.¹⁵⁰

6.5 Thin Films on Semiconductors

Because of its importance in the semiconductor industry Ge or more precisely SiGe on Si(100) has been the most-studied system.^{151–161} This system initially forms a several monolayer thick layer that is highly strained depending upon alloy composition. From this layer three-dimensional islands develop with facets that depend upon size and composition, leading to a rich variety of phenomena well suited for LEEM studies. The transition from the initial layer to the three-dimensional islands is strain driven and does not require three-dimensional nucleation.^{156,157,159} Surface steps play a critical role in the alloying of Ge with the Si(100) surface.¹⁶⁰ The results up to 2000 are summarized in an excellent review.¹⁶¹ The growth of Ge on other Si surfaces has been studied in much less detail. On the (111) surface the influence of the surfactant Sb on the growth was studied,¹⁶² on the (311) surface the transition from the two-dimensional layer to three-dimensional islands.¹⁶³ These showed a much more complicated facet structure than on Si(100). Finally, the growth of Ge on GaAs(100) has also been studied briefly.¹⁶⁴

The growth of metals on semiconductor surfaces can also be studied very well with LEEM. Most of the work was done on Si(111) surfaces, some also on Si(100), using video recording of the growth, diffusion, ordering, and disordering processes. Au on Si(111) is a good example: after several two-dimensional superstructures have formed with increasing coverage, three-dimensional particles grow that show an interesting temperature dependence due to the formation of an Au–Si eutectic.^{165,166} On vicinal Si(100) surfaces with a miscut of 4° , adsorption of a monolayer of Au causes pronounced faceting at elevated temperatures.^{167–170} The growth of Ag on Si(111) has also been studied extensively,^{162,171–174} in particular the growth shape,^{172,174} and has been used to demonstrate that buried interfaces can be imaged via their strain fields.¹⁷³ In the various studies of Ag on Si(100)^{175–177} a strain-induced shape transition of Ag crystals into “quantum wires”¹⁷⁵ and bamboo-like growth¹⁷⁸ have been observed. Substrate material incorporation into the two-dimensional superstructure¹⁷⁷ and the influence of steps on the domain ordering in it¹⁷⁶ were found. On vicinal Si(100) self-assembled Ag “quantum wires” form at high temperatures.¹⁷⁹ The growth of Cu on Si(111) has been studied both on the clean and on the hydrogen-passivated surface.^{180–184} Hydrogen termination strongly influences the formation of the two-dimensional “ (5×5) ” superstruc-

ture and the three-dimensional nucleation of Cu silicide. At higher temperatures at which hydrogen is desorbed large Cu_3Si crystals in a variety of shapes and facets form.

In Al films on Si(111) the two-dimensional phase diagram and the phase transitions between the phases have been studied.¹⁸⁵ The growth of Pb on Si(111) has been used to obtain an understanding of how interfactants, that is substrate surface layers that remain at the interface during growth, produce quasi-monolayer-by-monolayer growth, using Au and Ag as interfactants.^{186–188} Figure 8–32 illustrates the influence of an Au interfactant layer on the growth of Pb. On the Si(100) surface Pb grows in $\langle 111 \rangle$ -oriented crystals on the two-dimensional initial layer. The crystal, have frequently “ashtray” shape due to the large supply of Pb atoms by diffusion on the initial layer.¹⁸⁹ On vicinal and high index surfaces of Si, which have been faceted by Au adsorption, Pb grows in mesoscopic wires.^{190,191} In was found not to wet the Si(111)- (7×7) surface but to grow at low temperature monolayer-by-monolayer on the Si(111)- $(\sqrt{3} \times \sqrt{3})$ -R30° surface with a variety of superstructures. Three-dimensional crystals that grow at somewhat higher temperatures have predominantly (100) orientation with a surface reconstruction similar to that of the underlying two-dimensional layer.^{192,193} The Si(100) surface is etched at high temperatures by In similar to the situation mentioned above for Ag, at lower temperatures a superstructure forms.¹⁹⁴

Transition metals are highly reactive with Si and form silicides that frequently are more stable at high temperatures than Si. An example is Co. When deposited or annealed at high temperatures Co layers form more or less hexagonal CoSi_2 crystals that show misfit dislocation contrast when thin enough. When heated to temperatures at which Si sublimates CoSi_2 -topped hillocks form because of the lower vapor pressure of CoSi_2 (Figure 8–33).^{97,195} At lower temperatures the crystals

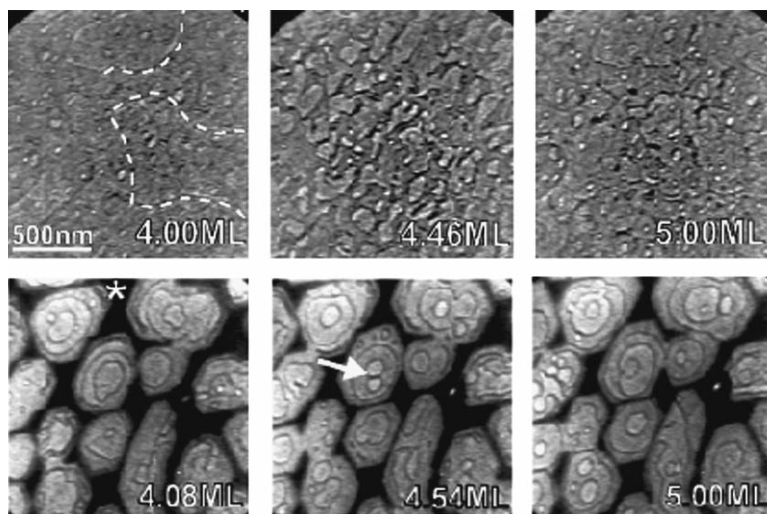
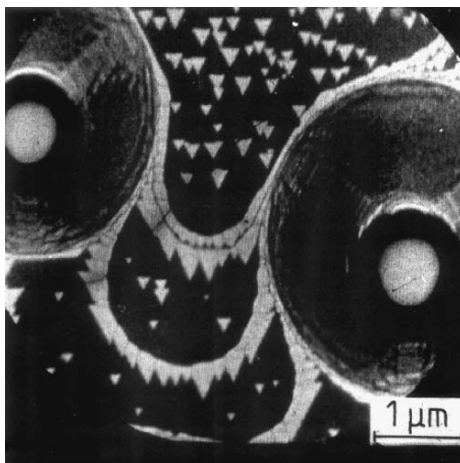


Figure 8–32. Images taken during the growth of Pb on Si(111) at 290 K with (top) and without Au surfactant (bottom). Electron energy 8 eV.¹⁸⁶

Figure 8-33. Hillocks on an Si(111) surface caused by CoSi_2 crystals during the sublimation of Si.⁹⁷



are triangular and act as Co scavengers cleaning the surrounding surface from Co so that it develops the (7×7) structure of the clean surface. At very low Co coverages an interesting “ring cluster” (RC) phase forms, which has been studied in considerable detail.^{196–198} Ni forms a similar RC phase.¹⁹⁹ Another transition metal, Ti, spontaneously forms Ti silicide “nanowires” when deposited at about 1120 K. Their formation process and stability have been studied in detail.²⁰⁰

Only a few nonmetal films on Si have been studied with LEEM: CaF_2 and Si nitride. In CaF_2 films both the complexities of the formation of the first two layers have been studied²⁰¹ as well the formation mechanism of the interfacial dislocation network with increasing film thickness.²⁰² The Si nitride work was concerned mainly with the formation of the initial epitaxial layer by reaction with NH_3 at high temperatures, which occurs via nucleation and growth of nitride domains similar to the (7×7) domains on the clean surface.²⁰³

6.6 Wide-Band Semiconductors

Very little work has been with LEEM on these materials. There is a cursory study of the temperature dependence of the structure of the 6H-SiC(0001) surface used as a substrate in the growth of GaN layers.²⁰⁴ The structure of the GaN(0001) surface was studied with dark-field imaging, which allowed determination of the surface termination,²⁰⁵ similar to that of the SiC(0001) surface.²⁰⁴ The importance of the Ga/N ratio in homoepitaxial growth of GaN was demonstrated in several papers,^{204,206,207} in particular the necessity of a Ga double layer on top of the GaN surface for the growth of films with (0001) surfaces.^{204,207,208} The double layer shows an interesting phase transition.²⁰⁴ Without an adsorbed Ga layer the GaN films grow rough, with {10-11} and {101-2} facets. GaN growth on 6H-Si(0001) is similar²⁰⁹ and not as reported in another paper,²¹⁰ except that initially three-dimensional crystals form instead of the spiral and step flow growth in homoepitaxial growth on GaN.

6.7 Metal Surfaces

Refractory metal surfaces have been the most popular subject in LEEM because of their high melting point and low vapor pressure that allow experiments over a wide temperature range. W and Mo can be cleaned easily by heating in oxygen. The chemisorbed oxygen is then flashed off at high temperatures, a procedure that is not successful in Nb and Ta, in which oxygen goes into solid solution and can be partially removed only by lengthy sputter and annealing cycles. The imaging of the step structure on the Mo(110) surface (Figure 8–23) was one of the first demonstrations of the power of LEEM and step contrast has been one of the mayor tools in the study of surface processes on clean surfaces. While W(110) and Mo(110) have been used frequently as substrates for thin films, and to a much lesser extent also W(100) and W(111), little has usually been published about the clean surface, except for a brief study of the Mo(110) surface.²¹¹ Extensive work was done, however, on epitaxial Mo(110)^{212–217} and Nb(110)^{215,218–220} layers grown on sapphire (11–20) surfaces at high temperatures, which after proper cleaning produces surfaces that are comparable in quality to single crystal surfaces. A complication is the interfacial strain and the dislocations introduced upon cooling and thermal cycling by the different thermal expansion coefficients between film and substrate. Nevertheless, pure surface quantities such as the step stiffness could be extracted from such films. In the case of the Nb(110) films the situation is complicated by the residual oxygen. This causes reconstruction and faceting,^{215,218,219} which in themselves are interesting processes and are useful for the study of extended defects.²²⁰ Some work on Ta(110) films was also done.²¹⁵

Noble metal surfaces have also been the subject of several LEEM studies. For Pt(111) single crystal surfaces step stiffness, step–step interactions, step free energy,^{215,217,221} and bulk-surface vacancy exchange^{222,223} have been determined. For Pd(111) surfaces the step stiffness²¹⁷ has been obtained and sputter erosion processes²²⁴ have been observed. Studies of the island decay on Rh(100) indicated a new surface diffusion process.²²⁵ Step fluctuation spectroscopy of Au(111) yielded the surface mass diffusion coefficient and the orientation-dependent step stiffness.²²⁶ Dark-field imaging of the reconstructed Au(100) surface was used to establish the connection between the reconstruction domains and the step orientations (Figure 8–34).^{80,227,228} On the Ag(111) surface a critical island size for layer-by-layer growth was found.²²⁹ From an investigation of the homoepitaxial growth of Cu on Cu(100) the Ehrlich–Schwoebel barrier, the energy barrier for diffusion across a step, was deduced.²³⁰ Similar to other fcc(110) surfaces, the Pb(110) surface reconstructs. A LEEM study of the various reconstructions, some of them alkali induced, revealed the topography of the various phases and the influence of surface defects on the transitions between them.^{231,232} Finally, studies of the surface morphology of the NiAl(110) surface demonstrated the importance of bulk diffusion for surface smoothing.²³³ Most experiments on clean surfaces rely on the step contrast discussed in Section 5, which illustrates its usefulness.

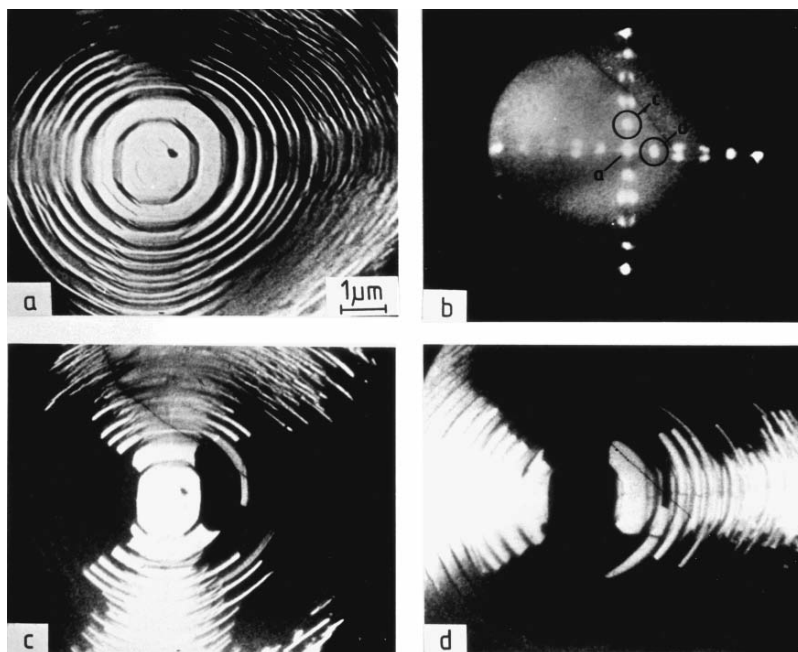


Figure 8-34. Au(100) surface. Bright-field image (a) and dark-field images (c and d) taken with the “(5 × 1)” superstructure reflections indicated in the LEED pattern (b). Electron energy in the images 16 eV.²²⁷

6.8 Metal Layers on Metals

Although the growth of many metals on W(110) and Mo(110) has been studied with other methods, LEEM has been used infrequently, often only in a very cursory manner. Cu is the most investigated layer material, both on Mo^{86,123,166,234,235} and W. The growth on both substrates is similar and has been studied in detail on W(110).²³⁶ Layer spacings have been determined using the quantum size effect discussed in Section 5.²³⁷ The more qualitative work on Mo revealed an interesting striped phase upon annealing at high temperatures and details in the structural phase transition in the double layer. Layer spacings have also been obtained from the quantum size effect for Ag on W(110).⁹² Au has been studied briefly on Mo(110) in the submonolayer range where it forms needle-like crystals.^{123,235} In Pd layers on W(110) the transition from two- to three-dimensional growth was the subject of a combined LEEM-XPEEM study.²³⁸ The growth of Ag and Cu on the Ru(0001) surface served as a demonstration of the influence of substrate steps on the growth of three-dimensional crystals.²³⁹ In the metal-on-metal systems discussed up to now the substrate surface is not modified or is modified only a little by the growing film. This is not the case on less densely packed surfaces such as the W(111) surface that facets upon deposition of certain metals at high temperatures. Pt growth on W(111) is a case study for this process.^{240,241} The growth of ferromagnetic layers will be discussed in Section 7 in connection with SPLEEM.

In many cases the deposited metal forms a surface alloy with the substrate. Pd on Mo(100) is an example that was studied in detail by LEEM up to several monolayers.²⁴² Clear alloying was observed up to a monolayer followed by two-dimensional Pd growth without faceting. Alloying of Sn with Cu(111) is strikingly different: at very low coverages large two-dimensional islands form that travel across the surface, leaving alloy behind and, thereby, decreasing in size.²⁴³ Pb forms on the Cu(100) surface initially several two-dimensional structures followed by the growth of three-dimensional crystals. LEEM clearly shows the correlation between the crystals and the initial layer,²⁴⁴ evidence for the surface alloying in the initial layer at low coverages, followed by dealloying.^{245,246} The nature of the disordering transitions of these phases was also determined.²⁴⁶ One of the most impressive results that LEEM has produced up to now is the self-assembly of stress domain patterns in the two-dimensional Pb–Cu alloy on Cu(111), which consist of domains of a Pb-rich and Pb-poor phase. This process has been studied in great detail, which has produced a wealth of information on stress-induced ordering phenomena^{247–251} and has been well reviewed.²⁵² Finally, the influence of the interface between Pb droplets and the Cu(111) surface on the shape and melting of the droplets has been the subject of an LEEM study.^{253,254}

6.9 Reactions on Metal Surfaces

Although LEEM is well-suited for the study of reactions on surfaces with gases or impurities from the bulk, very little work has been done up to now. Segregation of impurities from the bulk and the formation of precipitation products on the surface are routinely observed before the crystal has been cleaned completely. Surface carbide formation on W and Mo surfaces is an example. Images of such carbides and of carbides formed by CO dissociation have sometimes been published,^{123,213,234} but the precipitation process was never studied in detail. As far as oxidation is concerned, only three systems have been studied: the initial oxidation of W(100),^{255,256} the low oxygen coverage region on Nb(100),²⁵⁷ and the growth of oxide domains on NiAl(110).²⁵⁸ All three studies show interesting unexpected phenomena.

More work has been done on surface reactions in which the reaction products desorb, that is in heterogeneous catalysis. The first studies looked with very limited resolution at reactions of CO with O₂,^{259,260} with NO²⁶¹ and at the reaction of NO and H₂,²⁶² mainly at the propagation of the reaction front and its pinning by defects. On Pt(110) pattern formation in the CO oxidation has been studied^{263,264} and on Rh(110) the reaction of NO²⁶⁵ and of O₂ with H₂.^{266,267} Most of this work has been done with low resolution, but the possibilities of LEEM in this particular field are evident in some studies.^{265,267} The contrast is due to the fact that surface regions with different reactant composition have different structures, which produce characteristic LEED patterns. These can be used for dark-field imaging of these regions.²⁶⁵ Such dark-field images are shown in Figure 8–35 of the NO + H₂ reaction on Rh(110). The propagation of a spiral wave reaction is imaged with the diffraction

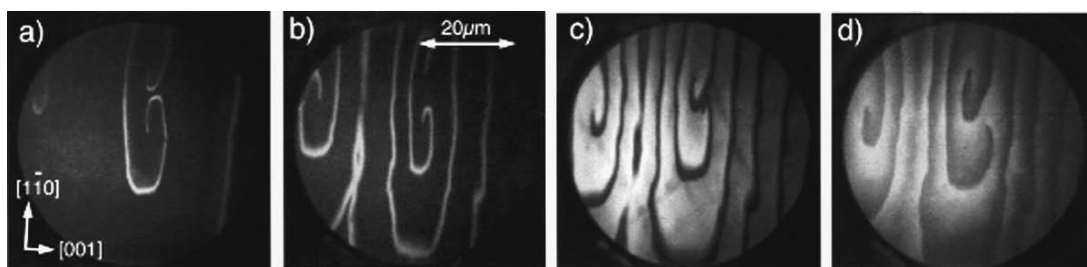


Figure 8-35. Dark-field images taken during the reaction of NO with H₂ on Rh(110) with LEED spots characteristic for the various phases in the oscillatory reaction.²⁶⁵

spots of pure N phases (a and b), a mixed N + O phase (c), and a pure O phase (d), which occur during the oscillatory reaction. Because of the high brightness, LEEM is preferable to PEEM because it allows following the kinetics of the reaction wave propagation and it is superior to MEM because of its better resolution.

6.10 Oxides and Nitrides

The only LEEM work on oxides published up to now is that on the TiO₂(110) surface.²⁶⁸⁻²⁷¹ TiO₂ is a particularly good example of the influence of the exchange of defects between the bulk and the surface on its structure because its stoichiometry can vary considerably. On the nonstoichiometric surface a (1 × 1) to (1 × 2) phase transition occurs as a function of temperature due to vacancy exchange between the bulk and the surface, which was studied thoroughly.²⁶⁸⁻²⁷⁰ When exposed to oxygen, the nonstoichiometric surface grows via step flow at high temperatures but via two-dimensional nucleation at lower temperatures. With suitable growth conditions the surface topography can be changed and the oxygen content of the bulk can be increased.²⁷¹ The only nitride studied to date—silicon nitride and GaN, which have been discussed in Sections 6.5 and 6.6, respectively, excepted—is the (111) surface of TiN layers. These layers form mounds with spiral steps or stacks of two-dimensional islands. The mass transport at high temperatures can be studied well by following the step motion and the diffusion processes and constants can be derived from it.^{272,273}

6.11 Concluding Remarks

The applications discussed in this subsection clearly show the many possibilities of LEEM in the study of surfaces and thin films of a wide variety of materials. Important information on the physical and chemical properties of surfaces and thin films has been extracted from these studies. These can be found in the references cited. Diffraction contrast, step contrast, and quantum size contrast, together with the real time capability, are the essential features that make LEEM so powerful in these studies. It should also be emphasized that LEEM is primarily an imaging method for *in situ* studies. Sample exchange is more time-consuming than in standard transmission and scanning electron

microscopy because the sample has to be transferred into ultrahigh vacuum, usually after cleaning in a preparation chamber, and then aligned in the microscope. These steps are necessary because of the high surface sensitivity of the method and because the sample is part of the objective lens so that its surface has to be exactly perpendicular to the optical axis.

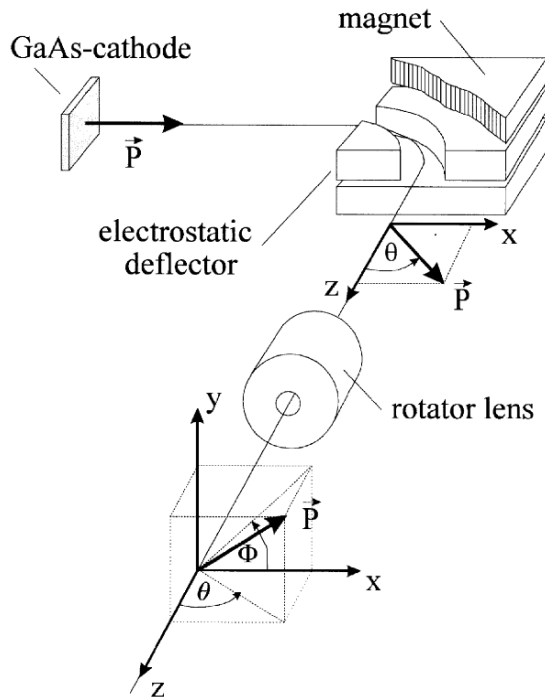
An important aspect of LEEM is that it can be easily combined with other, in part complementary surface imaging techniques. MEM is useful whenever no strong diffracted beam is available for imaging, for example, on fine-grained polycrystalline or amorphous samples. In this case the sample potential is chosen such that the electrons cannot penetrate into the sample. Contrast is then determined by surface topography, surface potential, and work function differences. MEM has been used, for example, in the study of chemical reactions.^{259,260} Ultraviolet light-excited photoemission electron microscopy (UVPEEM) is probably the most popular auxiliary imaging method in LEEM instruments. It is used when fields of view larger than those possible in LEEM have to be imaged or in samples in which it produces better contrast than LEEM. A good example is the study of the growth of pentacene films on oxidized Si.^{274,275} The application range of UVPEEM is the same as that of MEM, but the resolution is in general much better. Synchrotron radiation-excited XPEEM provides chemical information and chemically specific magnetic information. It can be combined easily with LEEM, in particular in instruments equipped with an energy filter, the SPELEEM mentioned in Section 3. XPEEM is discussed elsewhere in the book.

Other imaging methods that are possible in LEEM instruments are thermionic emission electron microscopy (TEEM). TEEM is useful in general only for samples with locally varying work functions and that can be heated high enough for thermionic emission. Metastable impact electron emission microscopy (MIEEM) is another imaging method with limited application range. In this method²⁷⁶ deexcitation of metastable He* atoms at the surface causes electron emission up to energies of 15–20 eV. Because of the chromatic aberration resolutions of 100 nm or less can be achieved only with a band pass energy filter.²⁷⁷ The main application of MIEEM is in the study of adsorbates, which consist of regions with different deexcitation processes. In LEEM instruments that allow higher energies in the illumination system than in the imaging system SEEM and in particular AEEM are possible. AEEM is useful for chemical analysis but inferior to XPEEM because of the larger characteristic peak width and the larger background. As in XPEEM, a band pass energy filter is indispensable for selection of a narrow energy band at the Auger electron peaks. The energy filter is also useful for SEEM whose application range is similar to that of UVPEEM and MEM.

7 Spin-Polarized LEEM (SPLEEM)

SPLEEM is a version of LEEM that requires a separate treatment because it does not give structural but magnetic information. It differs from LEEM only in that the illumination system produces a partially

Figure 8–36. Schematic of the spin manipulator. For explanation see text.²⁸⁵



spin-polarized electron beam. The polarization is achieved by illuminating a GaAs (100) surface with circular polarized light with the wavelength corresponding to the band gap of GaAs. The surface is activated by Cs and O₂ exposure to negative electron affinity so that electrons that have been excited to the bottom of the conduction band can escape the surface. Optical selection rules produce a spin selection in the excitation process such that the spin of the emitted electrons points normal to the surface either inward or outward, depending upon the helicity (right or left) of the exciting light. With specially designed photocathodes spin polarizations up to 80% have been achieved, but the ordinary cathodes used in SPLEEM usually have a polarization of only about 20–30%. Details of this kind of cathode can be found in Pierce.²⁷⁸

After extraction from the cathode, the electron beam is deflected 90° by a combined electrostatic–magnetic sector field. In pure electrostatic deflection the direction of the spin polarization $\mathbf{P}$ is unchanged, in pure magnetic deflection $\mathbf{P}$ is deflected 90°, and if both fields are used for 90° deflection $\mathbf{P}$ can be rotated in any direction in the plane indicated in Figure 8–36. In the early SPLEEM studies^{279–284} only electrostatic deflection was available. Later the magnetic rotator lens indicated in this figure was added, which now allows $\mathbf{P}$ to rotate in any direction in space.²⁸⁵ Usually, however, only three directions are selected, one normal to the surface of the crystal and the other two in preferred directions in the surface plane (easy and hard magnetic axes). The “spin manipulator” shown in Figure 8–36 is incorporated in the original LEEM described in Section 3 (Figure 8–9) and in a second instrument.^{50,286}

The magnetic contrast is due to the fact that the 180° backscattering from magnetic materials depends upon the relative orientation of the spin of the incident electrons and of the electrons in the material, or in other words upon the relative orientation of the polarization $\mathbf{P}$ and the magnetization $\mathbf{M}$. The intensity in the image is then given by $I = I_{\text{str}} + c \mathbf{P} \cdot \mathbf{M}$ where I_{str} is determined by the structure and topography of the surface, c is a small proportionality constant, and the dot indicates the scalar product. Pure magnetic contrast is obtained by subtracting two images taken with opposite $\mathbf{P}$ direction pixel by pixel. Maximum contrast occurs obviously for $\mathbf{P} \parallel \pm \mathbf{M}$. For $\mathbf{P} \perp \mathbf{M}$ the contrast between magnetic domains with opposite magnetization vanishes and only the domain walls produce contrast. Maximum contrast requires imaging at very low energies, typically a few electronvolts. At these energies the exchange splitting of the band structure and the difference between the inelastic mean free paths of the spin-up and spin-down electrons, which were mentioned in Section 2, cause the strongest difference between the backscattering for the two spin directions. Because the magnetic signal is only a small fraction of the total signal, the signal-to-noise ratio in the difference image is small and frequently limits the resolution.²⁸⁷ Addition of two images with opposite $\mathbf{P}$ direction produces only structural contrast, which makes SPLEEM ideal for the correlation between magnetism and structure. More information on magnetic contrast formation may be found in recent SPLEEM reviews.^{288,289}

In most of the SPLEEM studies the ferromagnetic samples are prepared *in situ* while observing the magnetic structure together with the crystal structure and topography via several contrast mechanisms. However, *ex situ* prepared samples can also be studied after proper surface cleaning as illustrated in Figure 8–37 for a Co(0001) surface that had been sputter-cleaned and annealed.²⁷⁹ *Ex situ* prepared samples can also be studied when passivated with a thin layer that is sufficiently transparent for slow electrons such as noble metals.²⁹⁰ In this method it has to be taken into account that such overlayers can change the magnetization in the film below.^{291,292} The *in situ* studies cover single



Figure 8–37. SPLEEM image of the closure domains on a Co(0001) surface. Electron energy 2 eV.²⁷⁹

ferromagnetic layers, ferromagnetic layers covered with nonmagnetic overlayers, ferromagnetic–nonmagnetic–ferromagnetic sandwiches, and small ferromagnetic crystals.

Under certain conditions large regions of a thin Fe film can be grown with constant thickness and atomically flat surfaces. These show pronounced quantum size effects (Figure 8–26).⁸⁸ The intensity reflected from regions with different thickness shows pronounced spin-dependent quantum size oscillations as a function of energy (Figure 8–27) from which the exchange-split band structure above the vacuum level can be derived.^{87,88}

The early SPLEEM work concentrated on Co films on W(110) without^{279–281} and with²⁸² nonmagnetic overlayers, on Co on Au films on W(110),²⁸³ and on initial studies of Co/Cu/Co sandwiches.²⁸⁰ Some of the more interesting results were the dependence of the spin reorientation transition from out-of-plane to in-plane magnetization—deduced from the disappearance of the in-plane magnetic contrast—upon the state of the Au film and the strong influence of the substrate topography;²⁸³ the large difference in the damping of the magnetic signal by Cu and Pd overlayers due to the different inelastic mean free paths in these materials; and the quantum size effect in the magnetic signal caused by Cu overlayers,²⁸² which was studied later in considerable detail in Cu films on fcc Co(100).⁹⁰

With the spin manipulator all three **M** components can be measured. Co on W(110) was found to have up to about ten monolayers an interesting “wrinkled,” predominantly in-plane magnetization²⁹³. The mechanism of the spin-reorientation transition in Co layers on Au(111) layers on W(110) could be elucidated in great detail²⁹⁴. Because of the long inelastic mean free path in Cu and Au and the short inelastic mean free path in Co, the interlayer exchange coupling between Co layers through Cu and Au layers could be shown to be much more complex than deduced from macroscopic measurements. In particular, the dependence of interface roughness-induced biquadratic coupling upon film thickness and growth conditions could be demonstrated^{295,296}. A good example of the sensitivity and flexibility of SPLEEM is the study of the spin reorientation transition of Fe–Co alloy layers on Au(111) layers on W(110)^{297,298}. As seen in Figure 8–38 the magnetic contrast is already quite strong at 1.22 monolayers, increases only slightly with thickness, and then decreases when film approaches the spin reorientation transition at about 2.7 monolayers. During the approach of the transition a pronounced striped phase develops and the magnetization tilts increasingly but converts abruptly into large, predominantly in-plane magnetized domains. The surface has large step bunches pointing toward and away from the vapor source, which causes the transition to occur earlier or later. In lateral averaging studies this would smear out the transition so that it would appear more continuous. Many other details can be extracted from the SPLEEM images, which indicate a rather complex transition.

The growth and magnetic domain structure on other W surfaces shows the crucial role that the substrate orientation plays. On a well-oriented W(111) surface many small domains in various orientations

are observed, indicating weak in-plane anisotropy in the Co(0001)-like layer that forms after the initial pseudomorphic growth. On a slightly misoriented surface only large domains with opposite **M** are observed, presumably induced by steps²⁹⁹. On the W(100) surface Co grows with (11-20) plane parallel to the substrate and forms long narrow and tall crystals. No magnetic signal could be obtained because the top surface of the crystals is not parallel to the substrate so that no electrons are reflected along the optical axis in the energy range in which the magnetic signal is strong³⁰⁰. This shows one of the limitations of SPLEEM.

Another interesting and much discussed magnetic thin film system is Fe on Cu(100). This system passes through several magnetic transitions with increasing film thickness and has been studied thoroughly with SPLEEM as a function of thickness, deposition rate, and temperature^{301,302}. Quantitative measurements allowed fitting of the data to theoretical models of the thickness dependence of the Curie temperature and of the critical behavior. The deposition rate was found to have major influence on initial film growth and magnetism. The same system, Fe/Cu(100), was used to demonstrate that SPLEEM imaging is also possible in a magnetic field perpendicular to the surface. From the images taken in fields up to ± 20 Oe the hysteresis curve of the layer could be determined³⁰³. In Ni films on Cu(100) very large domains were

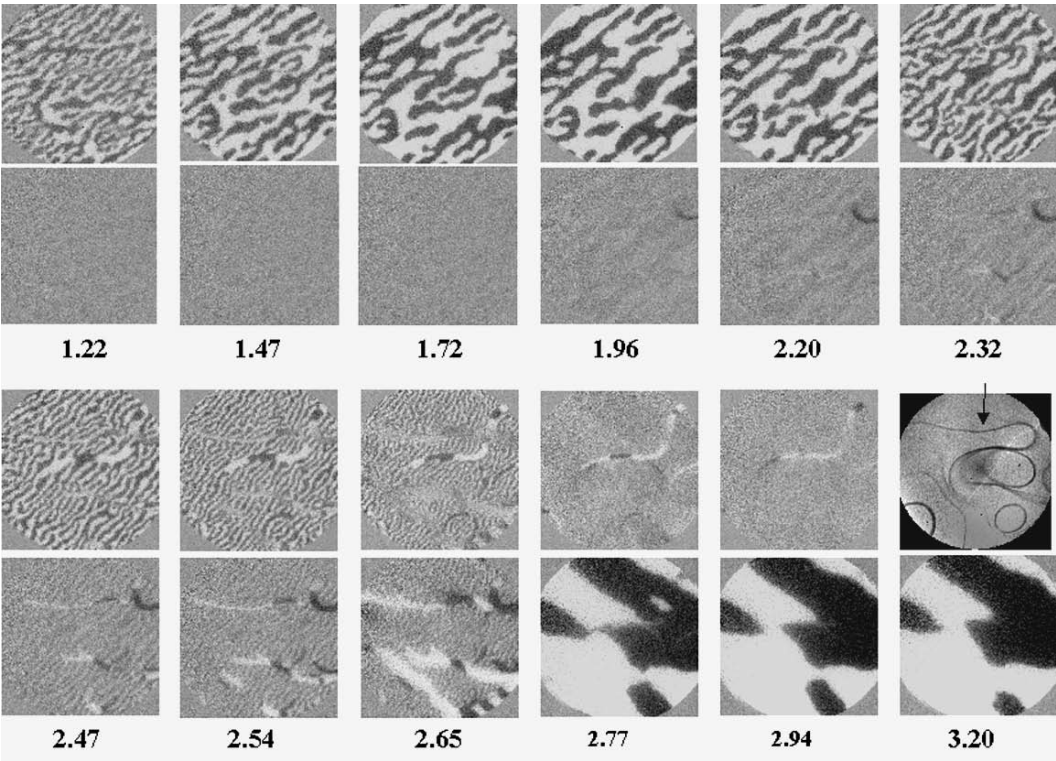


Figure 8–38. The spin reorientation transition in a Fe–Co alloy layer on Au(111). The top row shows the out-of-plane component of the magnetization and the bottom row the in-plane component. Electron energy 2.5 eV and diameter of field of view 10 μm .²⁹⁸

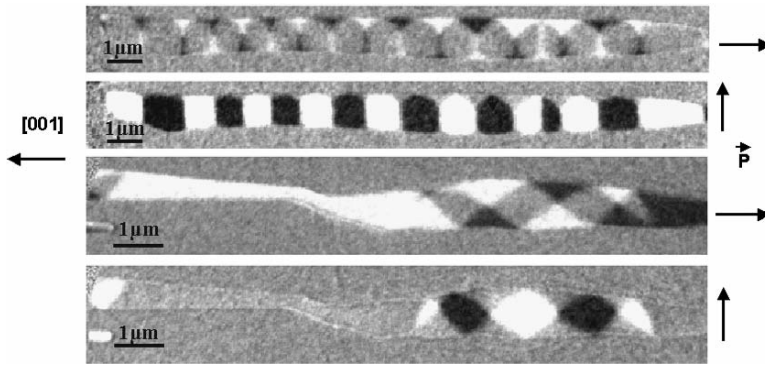


Figure 8-39. Epitaxial Fe ribbon crystals on W(110).³⁰⁵

observed whose domain walls were wide enough (400nm) to allow determination of the profile of the Néel walls³⁰⁴.

SPLEEM can be used not only for the study of the magnetization in thin films but also for that in small particles whose domain structure is of interest in magnetic recording. Small particles with a wide range of shapes can be grown in a SPLEEM instrument on single crystal substrates either by deposition or by annealing at elevated temperatures. The shapes range from nearly isometric crystals with lateral dimensions of several 100nm to long wires with widths down to less than 10nm.³⁰⁵ Figure 8-39 illustrates an intermediate case, wide Fe ribbons on W(110). The domain structure in these crystals is determined mainly by shape and misfit strain anisotropy.

Concluding this section, the strengths of SPLEEM should be emphasized once more: correlation with the microstructure, fast image acquisition compared to other magnetic imaging methods, high surface sensitivity, and easy access to all three magnetization components.

References

1. Davisson, C.J. and Germer, L.H. (1927). *Phys. Rev.* **30**, 705.
2. Bauer, E. (1962). In *Fifth International Congress for Electron Microscopy* (Breese, S.S., Jr. ed.), pp. D-11-12. (Academic Press, New York).
3. Telieps, W. and Bauer, E. (1985). *Ultramicroscopy* **17**, 57.
4. Brüche, E. and Knecht, M. (1934). *Z. Phys.* **92**, 462.
5. Mahl, H. (1939). *Z. Techn. Phys.* **20**, 316.
6. Farnsworth, H.E. (1929). *Phys. Rev.* **34**, 679.
7. Farnsworth, H.E. (1950). *Rev. Sci. Instrum.* **21**, 102.
8. Schlier, R.E. and Farnsworth, H.E. (1954). *J. Appl. Phys.* **25**, 1333 and references therein.
9. Scheibner, E.J., Germer, L.H. and Hartmann, C.D. (1960). *Rev. Sci. Instrum.* **31**, 112.
10. Germer, L.H. and Hartmann, C.D. (1960). *Rev. Sci. Instrum.* **31**, 784.
11. Recknagel, A. (1941). *Z. Phys.* **117**, 689.
12. Bauer, E. (1964). *J. Appl. Phys.* **35**, 3079.
13. Cruise, D.R. and Bauer, E. (1964). *J. Appl. Phys.* **35**, 3080.
14. Bauer, E. (1994). *Surf. Sci.* **299/300**, 102.

15. Schmidt, T., Heun, S., Slezak, J., Diaz, J., Prince, K.C., Lilienkamp, G. and Bauer, E. (1998). *Surf. Rev. Lett.* **5**, 1287.
16. Bauer, E. and Schmidt, T. (2003). In *High Resolution Imaging and Spectroscopy of Materials* (Ernst, F. and Rühle M., eds.), pp. 363–390. (Springer, Berlin).
17. Wichtendahl, R., Fink, R., Kuhlenbeck, H., Preikszas, D., Rose, H., Spehr, R., Hartel, P., Engel, W., Schlögl, R., Freund, H.-J., Bradshaw, A.M., Lilienkamp, G., Schmidt, Th., Bauer, E., Benner, G. and Umbach, E. (1998). *Surf. Rev. Lett.* **5**, 1249.
18. Bauer, E. (1970). *J. Vac. Sci. Technol.* **7**, 3.
19. Bauer, E. (1972). In *Les interactions des électrons avec la Matière Condensée. Applications à l'étude du Solide*, pp. 42–96. (AVCP, Verbier).
20. Wu, T.-Y. and Ohmura, T. (1962). *Quantum Theory of Scattering*, p. 4. (Prentice Hall, Englewood Cliffs, NJ).
21. Pendry, J.B. (1974). *Low-Energy Electron Diffraction*. (Academic Press, London).
22. Van Hove, M.A., Weinberg, W.H. and Chan, C.-M. (1986). *Low-Energy Electron Diffraction*. (Springer, Berlin).
23. Seitz, F. (1940). *The Modern Theory of Solids*, p. 313. (McGraw-Hill, New York).
24. Herlt, H.-J. (1982). Ph.D. thesis, TU Clausthal.
25. Feibelman, P.J. and Eastman, D.E. (1974). *Phys. Rev. B* **10**, 4932.
26. Bauer, E. (1988). In *Studies in Surface Science and Catalysis* (Koukal, J., ed.), pp. 26–36. (Elsevier, Amsterdam).
27. Herlt, H.-J., Feder, R., Meister, G. and Bauer, E.G. (1981). *Solid State Commun.* **38**, 973.
28. Lundquist, B.I. (1969). *Phys. Stat. Sol.* **32**, 273 and references therein.
29. Bauer, E. (1972). *Z. Metallkde.* **63**, 437.
30. Bauer, E. (1975). In *Interactions on Metal Surfaces* (Gomer, R., ed.), pp. 225–274. (Springer, Berlin).
31. Tanuma, S., Powell, C.J. and Penn, D.R. (1991). *Surf. Interface Anal.* **17**, 911, 927.
32. Ding, Z.-J. and Shimizu, R. (1996). *Scanning* **18**, 92.
33. Bauer, E. (1969). *Z. Phys.* **224**, 19.
34. Powell, C.J. and Jablonski, A. (1999). *J. Phys. Chem. Ref. Data* **28**, 19.
35. Goulet, T., Jung, J.-M., Michaud, M., Jay-Gerin, J.-P. and Sanche, L. (1994). *Phys. Rev. B* **50**, 5101.
36. Keszei, E., Goulet, T. and Jay-Gerin, J.-P. (1988). *Phys. Rev. A* **37**, 2183.
37. Siegmann, H.C. (1994). *Surf. Sci.* **307**, 1076.
38. Hong, J. and Mills, D.L. (2000). *Phys. Rev. B* **62**, 5589.
39. Bauer, E. (2005). In *Magnetic Microscopy of Nanostructures* (Hopster, H. and Oepen H.P., eds.), pp. 111–136. (Springer, Berlin).
40. Georgiev, N., Pavlovska, A. and Bauer, E. (1995). *Phys. Rev. B* **52**, 2878 and references therein.
41. Rundgren, J. (1999). *Phys. Rev. B* **59**, 5106.
42. Bartz, G. (1960). In *4th International Congress on Electron Microscopy*, Vol. 1, pp. 201–207. (Springer, Berlin).
43. Tromp, R.M. and Reuter, M.C. (1991). *Ultramicroscopy* **36**, 99.
44. Veneklasen, L.H. (1991). *Ultramicroscopy* **36**, 76.
45. Bauer, E., Franz, T., Koziol, C., Lilienkamp, G. and Schmidt, T. (1997). In *Chemical, Structural and Electronic Analysis of Heterogenous Surfaces on Nanometer Scale* (Rosei, R., ed.), pp. 73–84. (Kluwer, Dordrecht).
46. Bauer, E., Koziol, C., Lilienkamp, G. and Schmidt, T. (1997). *J. Electron Spectrosc. Relat. Phenom.* **84**, 201.
47. Bauer, E. (2001). *J. Electron Spectrosc. Relat. Phenom.* **114**, 975; *J. Phys. Condens. Matter* **13**, 11391.

48. Tromp, R.M., Mankos, M., Reuter, M.C., Ellis, A.W. and Copel, M. (1998). *Surf. Rev. Lett.* **5**, 1189.
49. Adamec, P., Bauer, E. and Lencová, B. (1998). *Rev. Sci. Instrum.* **69**, 3583.
50. Grzelakowski, K. and Bauer, E. (1996). *Rev. Sci. Instrum.* **67**, 742.
51. Liebl, H. and Senftinger, B. (1991). *Ultramicroscopy* **36**, 91.
52. Sakai, Y., Kato, M., Masuda, S., Harada, Y. and Ichinokawa, T. (1998). *Surf. Rev. Lett.* **5**, 1199.
53. Grzelakowski, K. (1999). *Rev. Sci. Instrum.* **70**, 3346; *Surf. Sci.* **454**, 1094.
54. Rose, H. and Preikszas, D. (1992). *Optik* **92**, 31.
55. Preikszas, D., Hartel, P., Spehr, R. and Rose, H. (2000). In *EUREM 12* (Čiampor, P. and Frank, L., eds.), pp. I81–I84. (Czechoslovak Society of Electron Microscopy, Brno).
56. Fink, R., Weiss, M.R., Umbach, E., Preikszas, D., Rose, H., Spehr, R., Hartel, P., Engel, W., Degenhardt, R., Wichtendahl, R., Erlebach, W., Ihmann, K., Schlögl, R., Freund, H.-J., Bradshaw, A.M., Lilienkamp, G., Schmidt, Th., Bauer, E. and Benner, G. (1997). *J. Electron Spectrosc. Relat. Phenom.* **84**, 231.
57. Schmidt, Th., Groh, U., Fink, R., Umbach, E., Schaff, O., Engel, W., Richter, B., Kuhlenbeck, H., Schlögl, R., Freund, H.-J., Bradshaw, A.M., Preikszas, D., Hartel, P., Spehr, R., Rose, H., Lilienkamp, G., Bauer, E. and Benner, G. (2002). *Surf. Rev. Lett.* **9**, 223.
58. Veneklasen, L.H. (1992). *Rev. Sci. Instrum.* **63**, 5513.
59. Chmelik, J., Veneklasen, L. and Marx, G. (1989). *Optik* **83**, 155.
60. Archard, G.D. and Mulvey, T. (1958). *J. Sci. Instrum.* **35**, 279.
61. Maňkoš, M., Kolařík, V. and Veneklasen, L.H. (1990). *Nuclear Instrum. Methods Phys. Res.* **A289**, 189.
62. Kolařík, V., Maňkoš, M. and Veneklasen, L.H. (1991). *Optik* **87**, 1.
63. Degenhardt, R. (1992). Ph.D. thesis, TU Darmstadt.
64. Satchenko, V. (1999). Unpublished.
65. Kan, H.-C., Auerbach, D. and Phaneuf, R.J. (2002). *Rev. Sci. Instrum.* **74**, 1008.
66. Kan, H.-C., Dürkop, T. and Phaneuf, R.J. (2003). *J. Vac. Sci. Technol.* **B20**, 2519.
67. Müller, H., Preikszas, D. and Rose, H. (1999). *J. Electron Microsc.* **48**, 191.
68. Spehr, R., Hartel, P., Müller, H., Preikszas, D. and Rose, H. (2000). In *EUREM 12* (Čiampor, P. and Frank, L., eds.), pp. I85–I86. (Czechoslovak Society of Electron Microscopy, Brno).
69. Wu, Y.K., Robin, D.S., Forest, E., Schlueter, R., Anders, S., Feng, J., Padmore, H. and Wei, D.H. (2003). *Nuclear Instrum. Methods phys. Res.* **A519**, 230.
70. Swenson, D.A. (1964). *Rev. Sci. Instrum.* **35**, 608.
71. Tsuno, K. (1994). *Ultramicroscopy* **55**, 127.
72. Zworykin, V.K., Morton, G.A., Ramberg, E.G., Hillier, J. and Vance, A.W. (1946). *Electron Optics and the electron Microscope*, pp. 29, 630, 645. (John Wiley, New York).
73. Ramberg, E.G. (1949). *J. Appl. Phys.* **20**, 183.
74. Skoczylas, W.P., Rempfer, G.F. and Griffith, O.H. (1991). *Ultramicroscopy* **36**, 252.
75. Preikszas, D. and Rose, H. (1992). *J. Electron Microsc.* **46**, 1.
76. Hartel, P., Preikszas, D., Spehr, R. and Rose, H. (2000). In *EUREM 12* (Čiampor, P. and Frank, L., eds.), pp. I153–I154. (Czechoslovak Society of Electron Microscopy, Brno).
77. Wan, W., Feng, J., Padmore, H.A. and Robin, D.S. (2004). *Nuclear Instrum. Methods Phys. Res. A* **519**, 222.
78. Cherifi, S., Hertel, R., Wang, H., Kirschner, J., Belkhou, R., Locatelli, A., Heun, S., Pavlovskaya, A. and Bauer, E. (2005). *Appl. Phys.* **98**, 843–901.

79. Telieps, W. and Bauer, E. (1985). *Surf. Sci.* **162**, 163.
80. Telieps, W. (1987). *Appl. Phys. A* **44**, 55.
81. Swiech, W., Bauer, E. and Mundschau, M. (1991). *Surf. Sci.* **253**, 283.
82. Yasue, T., Koshikawa, T., Jalochoowski, M. and Bauer, E. (2001). *Surf. Sci.* **493**, 381.
83. Altman, M.S., Chung, W.F. and Liu, C.H. (1998). *Surf. Rev. Lett.* **5**, 1129.
84. Chung, W.F. and Altman, M.S. (1998). *Ultramicroscopy* **74**, 237.
85. Müller, T. (1995). M.S. Thesis, TU Clausthal.
86. Mundschau, M., Bauer, E. and Swiech, W. (1989). *J. Appl. Phys.* **65**, 581.
87. Zdyb, R. and Bauer, E. (2002). *Phys. Rev. Lett.* **88**, 1666403.
88. Zdyb, R. and Bauer, E. (2002). *Surf. Rev. Lett.* **9**, 1485.
89. Altman, M.S. (2005). *J. Phys.: Condens. Matter* **17**, S1305.
90. Wu, Y.Z., Schmid, A.K., Altman, M.S., Jin, X.F. and Qiu, Z.Q. (2005). *Phys. Rev. Lett.* **94**, 027201.
91. Altman, M.S., Chung, W.F., He, Z.Q., Poon, H.C. and Tong, S.Y. (2001). *Appl. Surf. Sci.* **169/170**, 82.
92. Chung, W.F., Feng, Y.J., Poon, H.C., Chan, C.T., Tong, S.Y. and Altman, M.S. (2003). *Phys. Rev. Lett.* **90**, 216105.
93. Man, K.L., Qiu, Z.Q. and Altman, M.S. (2004). *Phys. Rev. Lett.* **93**, 236104.
94. Witt, W. (1984). Ph.D. thesis, TU Clausthal.
95. Bauer, E. (1987). In *Structure and Dynamics of Surfaces II* (Schommers, W. and von Blanckenhagen, P. eds.), pp. 115–179. (Springer, Berlin).
96. Telieps, W. and Bauer, E. (1986). *Ber. Bunsenges. Phys. Chem.* **90**, 19.
97. Bauer, E., Mundschau, M., Swiech, W. and Telieps, W. (1991). *J. Vac. Sci. Technol. A* **9**, 1007.
98. Phaneuf, R.J., Bartelt, N.C., Williams, E.D., Swiech, W. and Bauer, E. (1992). *Surf. Sci.* **268**, 227.
99. Hannon, J.B., Hibino, H., Bartelt, N.C., Swartzentruber, B.S., Ogino, T. and Kellogg, G.L. (2000). *Nature* **405**, 552.
100. Hibino, H., Hu, C.-W., Ogino, T. and Tsong, I.S.T. (2001). *Phys. Rev. B* **63**, 245402.
101. Hannon, J.B., Meyer zu Heringdorf, F.-J., Tersoff, J. and Tromp, R.M. (2001). *Phys. Rev. Lett.* **86**, 4871.
102. Hu, C.-W., Hibino, H., Ogino, T. and Tsong, I.S.T. (2001). *Surf. Sci.* **487**, 191.
103. Hannon, J.B. and Tromp, R.M. (2001). *J. Vac. Sci. Technol. A* **19**, 2596.
104. Hibino, H., Hu, C.-W., Ogino, T. and Tsong, I.S.T. (2001). *Phys. Rev. B* **64**, 245401.
105. Hannon, J.B., Tersoff, J. and Tromp, R.M. (2002). *Science* **295**, 299.
106. Hannon, J.B., Tersoff, J. and Tromp, R.M. (2002). *J. Crystal Growth* **237**, 181.
107. Hannon, J.B., Tersoff, J., Reuter, M.C. and Tromp, R.M. (2002). *Phys. Rev. Lett.* **89**, 266103.
108. Hibino, H., Homma, Y., Hu, C.-W., Uwaha, M., Ogino, T. and Tsong, I.S.T. (2004). *Appl. Surf. Sci.* **237**, 51.
109. Thayer, G.E., Hannon, J.B. and Tromp, R.M. (2004). *Nature Mat.* **3**, 95.
110. Tromp, R.M. and Hannon, J.B. (2002). *Surf. Rev. Lett.* **9**, 1565.
111. Hannon, J.B. and Tromp, R.M. (2003). *Annu. Rev. Mat. Res.* **33**, 263.
112. Phaneuf, R.J., Swiech, W., Bartelt, N.C., Williams, E.D. and Bauer, E. (1991). *Phys. Rev. Lett.* **67**, 2986.
113. Williams, E.D., Phaneuf, R.J., Bartelt, N.C., Swiech, W. and Bauer, E. (1992). *Mat. Res. Soc. Proc.* **238**, 219.
114. Phaneuf, R.J., Bartelt, N.C., Williams, E.D., Swiech, W. and Bauer, E. (1993). *Phys. Rev. Lett.* **71**, 2284.
115. Altman, M.S., Chung, W.F. and Franz, T. (1998). *Surf. Rev. Lett.* **5**, 27.
116. Chung, W.F. and Altman, M.S. (2002). *Phys. Rev. B* **66**, 075338.

117. Chung, W.F., Bromann, K. and Altman, M.S. (2002). *Intern. J. Mod. Phys. B* **16**, 4353.
118. Hibino, H. and Watanabe, Y. (2005). *Jpn. J. Appl. Phys.* **44**, 358.
119. Mundschau, M., Bauer, E., Telieps, W. and Swiech, W. (1989). *Surf. Sci.* **223**, 413.
120. Mundschau, M., Bauer, E., Telieps, W. and Swiech, W. (1990). *Phil. Mag. A* **61**, 257.
121. Bauer, E., Mundschau, M., Swiech, W. and Telieps, W. (1990). *Vacuum* **41**, 5.
122. Swiech, W. and Bauer, E. (1991). *Surf. Sci.* **255**, 219.
123. Bauer, E. (1992). *Appl. Surf. Sci.* **60/61**, 350.
124. Tromp, R.M. and Reuter, M.C. (1992). *Phys. Rev. Lett.* **68**, 820.
125. Tromp, R.M. and Reuter, M.C. (1993). *Phys. Rev. B* **47**, 7598.
126. Bartelt, N.C., Tromp, R.M. and Williams, E.D. (1994). *Phys. Rev. Lett.* **73**, 1656.
127. Theis, W., Bartelt, N.C. and Tromp, R.M. (1995). *Phys. Rev. Lett.* **75**, 3328.
128. Theis, W. and Tromp, R.M. (1996). *Phys. Rev. Lett.* **76**, 2770.
129. Tanaka, S., Umbach, C.C. and Blakely, J.M. (1996). *Appl. Phys. Lett.* **69**, 1235.
130. Bartelt, N.C. and Tromp, R.M. (1996). *Phys. Rev. B* **54**, 11731.
131. Bartelt, N.C., Theis, W. and Tromp, R.M. (1996). *Phys. Rev. B* **54**, 11741.
132. Tanaka, S., Bartelt, N.C., Umbach, C.C., Tromp, R.M. and Blakely, J.M. (1997). *Phys. Rev. Lett.* **78**, 3342.
133. Tanaka, S., Umbach, C.C., Blakely, J.M., Tromp, R.M. and Mankos, M. (1998). *J. Vac. Sci. Technol. A* **15**, 1345.
134. Tromp, R.M. and Mankos, M. (1998). *Phys. Rev. Lett.* **81**, 1050.
135. Blakely, J.M., Tanaka, S. and Tromp, R.M. (1999). *J. Electron Microsc.* **48**, 747.
136. Nielsen, J.-F., Pelz, J.P., Hibino, H., Hu, C.-W. and Tsong, I.S.T. (2001). *Phys. Rev. Lett.* **87**, 136103.
137. Zielasek, V., Liu, F., Zhao, Y., Maxson, J.B. and Lagally, M.G. (2001). *Phys. Rev. B* **64**, 201320.
138. Wurm, K., Kliese, R., Hong, Y., Röttger, B., Wei, Y., Neddermeyer, H. and Tsong, I.S.T. (1994). *Phys. Rev. B* **50**, 1567.
139. Hannon, J.B., Bartelt, M.C., Bartelt, N.C. and Kellogg, G.L. (1998). *Phys. Rev. Lett.* **81**, 4676.
140. Hannon, J.B., Kellogg, G.L., Bartelt, M.C. and Bartelt, N.C. (1998). *Surf. Rev. Lett.* **5**, 1151.
141. Tromp, R.M. and Reuter, M.C. (1992). *Phys. Rev. Lett.* **68**, 2313.
142. Jones, D.E., Pelz, J.P., Hong, Y., Bauer, E. and Tsong, I.S.T. (1996). *Phys. Rev. Lett.* **77**, 330.
143. Hannon, J.B., Bartelt, N.C., Swartzentruber, B.S., Hamilton, J.C. and Kellogg, G.L. (1997). *Phys. Rev. Lett.* **79**, 4226.
144. Hannon, J.B., Swartzentruber, B.S., Kellogg, G.L. and Bartelt, N.C. (1998), *Surf. Rev. Lett.* **5**, 1159.
145. Nielsen, J.-F., Pelz, J.P., Hibino, H., Hu, C.-W., Tsong, I.S.T. and Kouvetakis, J. (2001). *Appl. Phys. Lett.* **79**, 3857.
146. Sutter, P. and Lagally, M.G. (1999). *Phys. Rev. Lett.* **82**, 1490.
147. Tromp, R.M., Theis, W. and Bartelt, N.C. (1996). *Phys. Rev. Lett.* **77**, 2522.
148. Swiech, W., Schwarz-Selinger, Th. and Cahill, D.G. (2002). *Surf. Sci.* **519**, L599.
149. Watanabe, F., Kodambaka, S., Swiech, W., Greene, J.E. and Cahill, D.G. (2004). *Surf. Sci.* **572**, 425.
150. van Vroonhoven, E., Zandvliet, H.J.W. and Poelsema, B. (2003). *Phys. Rev. Lett.* **91**, 116102; **93**, 159902.

- 151 Tromp, R.M. and Reuter, M.C. (1992). *Phys. Rev. Lett.* **68**, 954.
152. Sutter, P., Mateeva, E. and Lagally, M.G. (1998). *J. Vac. Sci. Technol. B* **16**, 1560.
153. Sutter, P. and Lagally, M.G. (1998). *Phys. Rev. Lett.* **81**, 3471.
154. Sutter, P., Mateeva, E., Sullivan, J.S. and Lagally, M.G. *Thin Solid Films* **336**, 262.
155. Ross, F.M., Tromp, R.M. and Reuter, M.C. (1999). *Science* **286**, 1931.
156. Sutter, P. and Lagally, M.G. (2000). *Phys. Rev. Lett.* **84**, 4637.
157. Tromp, R.M., Ross, F.M. and Reuter, M.C. (2000). *Phys. Rev. Lett.* **84**, 4641.
158. Maxson, J.B., Savage, D.E., Liu, F., Tromp, R.M., Reuter, M.C. and Lagally, M.G. (2000). *Phys. Rev. Lett.* **85**, 2152.
159. Sutter, P. and Lagally, M.G. (2002). *Mater. Sci. Engineer. B* **89**, 45.
160. Hannon, J.B., Copel, M., Stumpf, R., Reuter, M.C. and Tromp, R.M. (2004). *Phys. Rev. Lett.* **92**, 216104.
161. Tromp, R.M. and Ross, F.M. (2000). *Annu. Rev. Mater. Sci.* **30**, 431.
162. Denier van der Gon, A.W., Tromp, R.M. and Reuter, M.C. (1993). *Thin Solid Films* **236**, 140.
163. Yasue, T., Koshikawa, T. and Bauer, E. (2002). *J. Vac. Sci. Technol. B* **20**, 2496.
164. Falta, J., Reuter, M.C. and Tromp, R.M. (1994). *Appl. Phys. Lett.* **65**, 1680.
165. Swiech, W., Bauer, E. and Mundschau, M. (1991). *Surf. Sci.* **253**, 283.
166. Bauer, E., Mundschau, M. and Swiech, W. (1991). *J. Vac. Sci. Technol. B* **9**, 403.
167. Horn-von Hoegen, M., Meyer zu Heringdorf, F.-J., Kähler, D., Schmidt, Th. and Bauer, E. (1998). *Thin Solid Films* **336**, 16.
168. Meyer zu Heringdorf, F.-J., Kähler, D., Horn-von Hoegen, M., Schmidt, Th., Bauer, E., Copel, M. and Minoda, H. (1998). *Surf. Rev. Lett.* **5**, 1167.
169. Horn-von Hoegen, M., Meyer zu Heringdorf, F.-J., Hild, R., Zahl, P., Schmidt, Th. and Bauer, E. (1999). *Surf. Sci.* **433**, 475.
170. Meyer zu Heringdorf, F.-J., Hild, R., Zahl, P., Schmidt, Th., Ressel, B., Heun, S., Bauer, E. and Horn-von Hoegen, M. (2001). *Surf. Sci.* **480**, 103; **496**, 151.
171. Denier van der Gon, A.W. and Tromp, R.M. (1992). *Phys. Rev. Lett.* **69**, 3519.
172. Tersoff, J., Denier van der Gon, A.W. and Tromp, R.M. (1993). *Phys. Rev. Lett.* **70**, 1143.
173. Tromp, R.M., Denier van der Gon, A.W., LeGoues, F.K. and Reuter, M.C. (1993). *Phys. Rev. Lett.* **71**, 3299.
174. Tang, W.X., Man, K.L., Huang, H., Woo, C.H. and Altman, M.S. (2002). *J. Vac. Sci. Technol. B* **20**, 2492.
175. Tersoff, J. and Tromp, R.M. (1993). *Phys. Rev. Lett.* **70**, 2782.
176. Michely, T., Reuter, M.C., Copel, M. and Tromp, R.M. (1994). *Phys. Rev. Lett.* **73**, 2095.
177. Tromp, R.M. and Michely, T. (1995). *Nature* **373**, 499.
178. Li, B.Q., Swiech, W., Venables, J.A. and Zuo, J.M. (2004). *Surf. Sci.* **569**, 142.
179. Roos, K.R., Roos, K.L., Horn-von Hoegen, M. and Meyer zu Heringdorf, F.-J. (2005). *J. Phys.: Condens. Matter* **17**, S1407.
180. Mundschau, M., Bauer, E., Teliaps, W. and Swiech, W. (1989). *J. Appl. Phys.* **65**, 4747.
181. Yasue, T., Koshikawa, T., Jatochowski, M. and Bauer, E. (2000). *Surf. Rev. Lett.* **7**, 595.
182. Yasue, T., Koshikawa, T., Jatochowski, M. and Bauer, E. (2001). *Surf. Sci.* **480**, 118.
183. Yasue, T., Koshikawa, T., Jatochowski, M. and Bauer, E. (2001). *Surf. Sci.* **493**, 381.

184. Kuroiwa, N., Fukushima, Y., Rajasekar, P., Neddermeyer, H., Jatochowski, M., Bauer, E., Yasue, T. and Koshikawa, T. (2003). *Surf. Interface Anal.* **35**, 24.
185. Michely, T., Reuter, M.C. and Tromp, R.M. (1996). *Phys. Rev. B* **53**, 4105.
186. Schmidt, Th. and Bauer, E. (2000). *Phys. Rev. B* **62**, 15815.
187. Schmidt, Th., Ressel, B., Heun, S., Prince, K.C. and Bauer, E. (2000). In *X-Ray Microscopy: Proceedings of the 6th International Conference* (Meyer-Else, W., Warwick, T. and Attwood, D., eds.). pp. 27–32. (AIP, Melville, NY).
188. Schmidt, T. and Bauer, E. (2001). *Surf. Sci.* **480**, 137.
189. Li, L., Koziol, C., Wurm, K., Hong, Y., Bauer, E. and Tsong, I.S.T. (1994). *Phys. Rev. B* **50**, 10834.
190. Jatochowski, M. and Bauer, E. (2001). *Progr. Surf. Sci.* **67**, 79.
191. Jatochowski, M. and Bauer, E. (2001). *Surf. Sci.* **480**, 109.
192. Pavlovskaya, A., Bauer, E. and Giessen, M. (2002). *J. Vac. Sci. Technol. B* **20**, 2478.
193. Pavlovskaya, A. and Bauer, E. (2005). *Surf. Interface Anal.* **37**, 110.
194. McKay, H.A. and Feenstra, R.M. (2003). *Surf. Sci.* **547**, 127.
195. Bauer, E. (1994). *Scanning Microsc.* **8**, 765.
196. Phaneuf, R.J., Hong, Y., Horch, S. and Bennett, P.A. (1997). *Phys. Rev. Lett.* **78**, 4605.
197. Phaneuf, R.J., Hong, Y., Horch, S. and Bennett, P.A. (1999). *Micron* **30**, 13.
198. Phaneuf, R.J., Bennett, P.A., Marsi, M., Günther, S., Gregoratti, L., Casalis, L. and Kiskinova, M. (1999). *Surf. Sci.* **431**, 232.
199. Bennett, P.A., Lee, M.Y., Parikh, S.A., Wurm, K. and Phaneuf, R.J. (1995). *J. Vac. Sci. Technol. A* **13**, 1728.
200. Bennett, P.A., Ashcroft, B., He, Z. and Tromp, R.M. (2002). *J. Vac. Soc. Technol. B* **20**, 2500.
201. Tromp, R.M. and Reuter, M.C. (1994). *Phys. Rev. Lett.* **73**, 1236.
202. Tromp, R.M., LeGoues, F.K. and Reuter, M.C. (1995). *Phys. Rev. Lett.* **74**, 2706.
203. Bauer, E., Wei, Y., Müller, T., Pavlovskaya, A. and Tsong, I.S.T. (1995). *Phys. Rev. B* **51**, 17891.
204. Pavlovskaya, A., Bauer, E. and Smith, D.J. (2002). *Surf. Sci.* **496**, 160.
205. Maxson, J.B., Perkins, N., Savage, D.E., Woll, A.R., Zhang, L., Kuech, T.F. and Lagally, M.G. (2000). *Surf. Sci.* **464**, 217.
206. Pavlovskaya, A., Torres, V.M., Edwards, J.L., Bauer, E., Smith, D.J., Doak, R. B., Tsong, I.S.T., Thomson, D.B. and Davis, R.F. (1999). *Phys. Stat. Sol. (a)* **176**, 469.
207. Pavlovskaya, A. and Bauer, E. (2001). *Surf. Sci.* **480**, 128.
208. Pavlovskaya, A., Bauer, E., Torres, V.M., Edwards, J.L., Doak, R.B., Tsong, I.S.T., Ramachandran, V. and Feenstra, R.M. (1998). *J. Crystal Growth* **189/190**, 310.
209. Pavlovskaya, A. and Bauer, E. (2001). *Surf. Rev. Lett.* **8**, 337.
210. Hu, Ch., Smith, D.J., Doak, R.B. and Tsong, I.S.T. (2000). *Surf. Rev. Lett.* **7**, 565.
211. Mundschau, M., Bauer, E. and Swiech, W. (1989). *Phil. Mag. A* **59**, 217.
212. Swiech, W., Mundschau, M. and Flynn, C.P. (1999). *Appl. Phys. Lett.* **74**, 2626.
213. Swiech, W., Mundschau, M. and Flynn, C.P. (1999). *Surf. Sci.* **437**, 61.
214. Mundschau, M., Swiech, W., Durfee, C.S. and Flynn, C.P. (1999). *Surf. Sci.* **440**, L831.
215. Ondrejcek, M., Swiech, W., Yang, G. and Flynn, C.P. (2002). *J. Vac. Sci. Technol. B* **20**, 2473.
216. Ondrejcek, M., Swiech, W., Durfee, C.S. and Flynn, C.P. (2003). *Surf. Sci.* **541**, 31.

217. Ondrejcek, M., Swiech, W. and Flynn, C.P. (2004). *Surf. Sci.* **566**, 160.
218. Flynn, C.P. and Swiech, W. (1999). *Phys. Rev. Lett.* **83**, 3482.
219. Flynn, C.P., Swiech, W., Appleton, R.S. and Ondrejcek, M. (2000). *Phys. Rev. B* **62**, 2096.
220. Appleton, R.S., Swiech, W., Ondrejcek, M. and Flynn, C.P. (2003). *Phil. Mag.* **83**, 1639.
221. Flynn, C.P., Ondrejcek, M. and Swiech, W. (2003). *Chem. Phys. Lett.* **378**, 161.
222. Ondrejcek, M., Swiech, W., Yang, G. and Flynn, C.P. (2004). *Phil. Mag. Lett.* **84**, 69.
223. Poelsema, B., Hannon, J.B., Bartelt, N.C. and Kellogg, G.L. (2004). *Appl. Phys. Lett.* **84**, 2551.
224. Ondrejcek, M., Swiech, W., Rajappan, M. and Flynn, C.P. (2003). *J. Phys.: Condens. Matter* **15**, L735.
225. Kellogg, G.L. and Bartelt, N.C. (2005). *Surf. Sci.* **577**, 151.
226. Ondrejcek, M., Rajappan, M., Swiech, W. and Flynn, C.P. (2005). *Surf. Sci.* **574**, 111.
227. Teliëps, W., Mundschauf, M. and Bauer, E. (1987). *Optik* **77**, 93.
228. Teliëps, W., Mundschauf, M. and Bauer, E. (1990). *Surf. Sci.* **225**, 87.
229. Tersoff, J., Denier van der Gon, A.W. and Tromp, R.M. (1994). *Phys. Rev. Lett.* **72**, 266.
230. Gerlach, R., Maroutian, T., Douillard, L., Martinotti, D. and Ernst, H.-J. (2001). *Surf. Sci.* **480**, 97.
231. Altman, M.S., Chiang, S., Statiris, P., Gustafsson, T. and Bauer, E. (1994). In *The Structure of Surfaces IV* (Xide, X., Tong, S.Y. and van Hove, M.A., eds.), pp. 183–191. (World Scientific, Singapore).
232. Altman, M.S. and Bauer, E. (1995). *Surf. Sci.* **344**, 51.
233. McCarty, K.F., Nobel, J.A. and Bartelt, N.C. (2001). *Nature* **412**, 622.
234. Mundschauf, M., Bauer, E. and Swiech, W. (1988). *Catal. Lett.* **1**, 405.
235. Bauer, E. (1991). *Ber. Bunsenges. Phys. Chem.* **95**, 1311.
236. Lou, E.Z., Cai, Q., Chung, W.F., Orr, B.G. and Altman, M.S. (1996). *Phys. Rev. B* **54**, 14673.
237. Man, I.K.H. and Altman, M.S. (2005). *Surf. Interface Anal.* **37**, 235.
238. Aballe, L., Barinov, A., Locatelli, A., Heun, S., Cherifi, S. and Kiskinova, M. (2004). *Appl. Surf. Sci.* **238**, 138.
239. Ling, W.L., Giessel, T., Thürmer, K., Hwang, R.Q., Bartelt, N.C. and McCarty, K.F. (2004). *Surf. Sci.* **570**, L297.
240. Pelhos, K., Hannon, J.B., Kellogg, G.L. and Madey, T.E. (1999). *Surf. Sci.* **432**, 115.
241. Pelhos, K., Hannon, J.B., Kellogg, G.L. and Madey, T.E. (1999). *Surf. Rev. Lett.* **6**, 767.
242. Wu, D., Lau, W.K., He, Z.Q., Feng, Y.J., Altman, M.S. and Chan, C.T. (2000). *Phys. Rev. B* **62**, 8366.
243. Schmid, A.K., Bartelt, N.C. and Hwang, R.Q. (2000). *Science* **290**, 1561.
244. Kellogg, G.L. and Plass, R. (2000). *Surf. Sci.* **465**, L777.
245. Plass, R. and Kellogg, G.L. (2000). *Surf. Sci.* **470**, 106.
246. Kellogg, G.L. and Plass, R.A. (2000). *Surf. Rev. Lett.* **7**, 649.
247. Plass, R., Last, J.A., Bartelt, N.C. and Kellogg, G.L. (2001). *Nature* **412**, 875.
248. Plass, R., Bartelt, N.C. and Kellogg, G.L. (2002). *J. Phys.: Condens. Matter* **14**, 4227.
249. van Gastel, R., Plass, R., Bartelt, N.C. and Kellogg, G.L. (2003). *Phys. Rev. Lett.* **91**, 055503.
250. van Gastel, R., Bartelt, N.C., Feibelman, P.J., Léonard, F. and Kellogg, G.L. (2004). *Phys. Rev. B* **70**, 245413.

251. Léonard, F., Bartelt, N.C. and Kellogg, G.L. (2005). *Phys. Rev. B* **71**, 045416.
252. Kellogg, G.L. (2002). In *The Chemical Physics of Solid Surfaces*, Vol. 10 (Woodruff, D.P., ed.), pp. 152–183. (Elsevier, Amsterdam).
253. Altman, M.S., Cai, Q., Chung, W.F., Luo, E.Z., Pinkvos, H. and Bauer, E. (1995). *Mat. Res. Soc. Symp. Proc.* **355**, 235.
254. Luo, E.Z., Cai, Q., Chung, W.F. and Altman, M.S. (1996). *Appl. Surf. Sci.* **92**, 331.
255. Altman, M.S. and Bauer, E. (1991). *J. Vac. Sci. Technol. A* **9**, 659.
256. Altman, M.S. and Bauer, E. (1996). *Surf. Sci.* **347**, 265.
257. Ondrejcek, M., Appleton, R.S., Swiech, W., Petrova, V.L. and Flynn, C.P. (2001). *Phys. Rev. Lett.* **87**, 116102.
258. McCarty, K.F. (2001). *Surf. Sci.* **474**, L165.
259. Rausenberger, B., Swiech, W., Engel, W., Bradshaw, A.M. and Zeitler, E. (1993). *Surf. Sci.* **287/288**, 235.
260. Rausenberger, B., Swiech, W., Rastomjee, C.S., Mundschauf, M., Engel, W., Zeitler, E. and Bradshaw, A.M. (1993). *Chem. Phys. Lett.* **215**, 109.
261. Schmid, A.K., Swiech, W., Rastomjee, C.S., Rausenberger, B., Engel, W., Zeitler, E. and Bradshaw, A.M. (1995). *Surf. Sci.* **331**, 225.
262. Rausenberger, B., Swiech, W., Schmid, A.K., Rastomjee, C.S., Engel, W. and Bradshaw, A.M. (1998). *J. Chem. Soc., Faraday Trans.* **94**, 963.
263. Rose, K.C., Berton, B., Imbihl, R., Engel, W. and Bradshaw, A.M. (1997). *Phys. Rev. Lett.* **79**, 3427.
264. Wei, H., Lilienkamp, G. and Imbihl, R. (2004). *Chem. Phys. Lett.* **389**, 284.
265. Schmidt, Th., Schaak, A., Günther, S., Ressel, B., Bauer, E. and Imbihl, R. (2000). *Chem. Phys. Lett.* **318**, 549.
266. Marbach, H., Lilienkamp, G., Wei, H., Günther, S., Suchorski, Y. and Imbihl, R. (2003). *Phys. Chem. Chem. Phys.* **5**, 2730.
267. Locatelli, A., Heun, S. and Kiskinova, M. (2004). *Surf. Sci.* **566**, 1130.
268. McCarty, K.F. and Bartelt, N.C. (2003). *Phys. Rev. Lett.* **90**, 046104.
269. McCarty, K.F. and Bartelt, N.C. (2003). *Surf. Sci.* **527**, L203.
270. McCarty, K.F. and Bartelt, N.C. (2003). *Surf. Sci.* **540**, 157.
271. McCarty, K.F. (2003). *Surf. Sci.* **543**, 185.
272. Kodambaka, S., Khare, S.V., Swiech, W., Ohmori, K., Petrov, I. and Greene, J.E. (2004). *Nature* **429**, 49.
273. Kodambaka, S., Israeli, N., Bareño, J., Swiech, W., Ohmori, K., Petrov, I. and Greene, J.E. (2004). *Surf. Sci.* **560**, 53.
274. Meyer zu Heringdorf, F.-J., Reuter, M.C. and Tromp, R.M. (2001). *Nature* **412**, 517; *Appl. Phys. A* **78**, 787.
275. Harada, Y., Yamamoto, S., Aoki, M., Masuda, S., Ichinokawa, T., Kato, M. and Sakai, Y. (1994). *Nature* **372**, 657.
276. Yamamoto, S., Masuda, S., Yasufuku, H., Ueno, N., Harada, Y., Ichinokawa, T., Kato, M. and Sakai, Y. (1997). *J. Appl. Phys.* **82**, 2954.
277. Lilienkamp, G., Wei, H., Maus-Friedrichs, W., Kempter, V., Marbach, H., Günther, S. and Suchorski, Y. (2003). *Surf. Sci.* **532**, 132.
278. Pierce, D.T. (1995). In *Atomic, Molecular and Optical Physics* (Vol. 29A of *Experimental Methods in the Physical Sciences*) (Dunning, F.B. and Hulet, R.G., eds.), pp. 1–38. (Academic Press, San Diego).
279. Altman, M.S., Pinkvos, H., Hurst, J., Poppa, H., Marx, G. and Bauer, E. (1991). *Mat. Res. Soc. Symp. Proc.* **232**, 125.
280. Pinkvos, H., Poppa, H., Bauer, E. and Hurst, J. (1992). *Ultramicroscopy* **47**, 339.
281. Pinkvos, H., Poppa, H., Bauer, E. and Kim, G.-M. (1993). In *Magnetism and Structure in Systems of Reduced Dimensions* (Farrow, R.F.C., Dieny, B., Donath, M., Fert, A. and Hermsmeier, B.D., eds.), pp. 25–31. (Plenum Press, New York).

282. Poppa, H., Pinkvos, H., Wurm, K. and Bauer, E. (1993). *Mat. Res. Soc. Symp. Proc.* **313**, 219.
283. Altman, M.S., Pinkvos, H. and Bauer, E. (1995). *J. Magn. Soc. Jpn.* **19**(Suppl. S1), 129.
284. Bauer, E., Duden, T., Pinkvos, H., Poppa, H. and Wurm, K. (1996). *J. Magn. Magn. Mater.* **156**, 1.
285. Duden, T. and Bauer, E. (1995). *Rev. Sci. Instrum.* **66**, 2861.
286. Grzelakowski, K., Duden, T., Bauer, E., Poppa, H. and Chiang, S. (1994). *IEEE Trans. Magnet.* **30**, 4500.
287. Duden, Th. and Bauer, E. (1998). *Surf. Rev. Lett.* **5**, 1213.
288. Bauer, E. (2005). In *Magnetic Microscopy of Nanostructures* (Hopster, H. and Oepen, H.P., eds.), pp. 111–136. (Springer, Berlin).
289. Bauer, E. (2005). In *Novel Techniques for Characterizing Magnetic Materials* (Zhu, Y., ed.), pp. 361–379. (Springer, Berlin).
290. Tober, E.D., Witte, G. and Poppa, H. (2000). *J. Vac. Sci. Technol. A* **18**, 1845.
291. Duden, T. and Bauer, E. (1998). *J. Electron Microsc.* **47**, 379.
292. Duden, T. and Bauer, E. (1999). *Phys. Rev. B* **59**, 468.
293. Duden, T. and Bauer, E. (1996). *Phys. Rev. Lett.* **77**, 2308.
294. Duden, T. and Bauer, E. (1997). *Mat. Res. Soc. Symp. Proc.* **475**, 283.
295. Duden, T. and Bauer, E. (1999). *J. Magn. Magn. Mat.* **191**, 301.
296. Duden, T. and Bauer, E. (1999). *Phys. Rev. B* **59**, 474.
297. Zdyb, R. and Bauer, E. (2003). *Phys. Rev. B* **67**, 134420.
298. Zdyb, R., Locatelli, A., Heun, S., Cherifi, S., Belkhou, R. and Bauer, E. (2005). *Surf. Interface Anal.* **37**, 239.
299. Man, K.L., Zdyb, R., Huang, S.F., Leung, T.C., Chan, C.T., Bauer, E. and Altman, M.S. (2003). *Phys. Rev. B* **67**, 184402.
300. Duden, Th., Zdyb, R., Altman, M.S. and Bauer, E. (2001). *Surf. Sci.* **480**, 145.
301. Man, K.L., Altman, M.S. and Poppa, H. (2001). *Surf. Sci.* **480**, 163.
302. Man, K.L., Ling, W.L., Paik, S.Y., Poppa, H., Altman, M.S. and Qiu, Z.Q. (2001). *Phys. Rev. B* **65**, 024409.
303. Poppa, H., Tober, E.D. and Schmid, A.K. (2002). *J. Appl. Phys.* **91**, 6932.
304. Ramchal, R., Schmid, A.K., Farle, M. and Poppa, H. (2003). *Phys. Rev. B* **68**, 054418.
305. Zdyb, R., Pavlovskaya, A., Jalochofski, M. and Bauer, E. (2006). *Surf. Sci.* **600**, 1586.
- *306. Mankos, M., Adler, D., Veneklasen, L. and Munro, E. (2007). *Surf. Sci.* **601**, 4733 (Section 4).
- *307. de la Figuera, J., Bartelt, N.C. and McCarty, K.F. (2006). *Surf. Sci.* **600**, 4062 (Section 5).
- *308. Sadowski, J.T., Sazaki, G., Nishikata, S., Al-Mahboob, A., Fujikawa, Y., Nakajima, K., Tromp, R.M. and Sakurai, T. (2007). *Phys. Rev. Lett.* **98**, 046104 (Section 6.5).
- *309. Tang, S.-J., Kodambaka, S., Swiech, W., Petrov, I., Flynn, C.P. and Chiang, T.-C. (2006). *Phys. Rev. Lett.* **96**, 126106 (Section 6.7).
- *310. van Gastel, R., Bartelt, N.C. and Kellogg, G.L. (2006). *Phys. Rev. Lett.* **96**, 036106 (Section 6.8).
- *311. Man, K.L., Feng, Y.J., Chan, C.T. and Altman, M.S. (2007). *Surf. Sci.* **601**, L95 (Section 6.8).
- *312. Aballe, L., Barinov, A., Locatelli, A., Montes, T.O. and Kiskinova, M. (2007). *Phys. Rev. B* **75**, 115411 (Section 6.8).
- *313. El Gabaly, F., Gallego, S., Muñoz, Szunyogh, L., Weinberger, P., Klein, Ch., Schmid, K.A., McCarty, K.F. and de la Figuera, J. (2006). *Phys. Rev. Lett.* **96**, 147202 (Section 7).

Photoemission Electron Microscopy (PEEM)

Jun Feng and Andreas Scholl

1 Introduction

1.1 A Brief History of Photoemission Electron Microscopy

Photoemission electron microscopy (PEEM) has proven to be a powerful tool in material science, surface physics and chemistry, thin film magnetism, polymer science, and biology. PEEM, also called photoelectron microscopy (PEM), is the most widely used type of emission microscopy and is closely related to the more recently developed low-energy electron microscopy (LEEM) (Bauer, 1994). In 1905, Einstein proposed that light quanta are quantized and explained the photoelectric effect in terms of a photon's energy to release an electron from a surface (Einstein, 1905). A photoemission electron microscope forms an image of a solid surface based on the spatial distribution of electrons emitted by the absorption of photons with energy that exceeds the sample work function. Historically, the invention of PEEM dates to the early 1930s shortly after the introduction of electron lenses. The first working photoemission electron microscope was built by Brüche in 1932 using ultraviolet (UV) light to image photoelectrons emitted from a metal (Brüche, 1932; 1933) (Knoll et al., 1932). The principal design of his PEEM is still used. In Brüche's PEEM, UV light from a mercury lamp was focused by a quartz lens onto the sample cathode. The emitted electrons were accelerated by a potential difference of 10–30 kV between the cathode and the anode of the PEEM and focused onto a phosphor screen. At about the same time, Knoll, Hontermans, and Schulzer constructed an emission microscope using two magnetic lenses (Langmuir, 1937). The first calculation of the resolution in emission microscopy was published by Langmuir in 1937, and subsequent treatments were published by Reckangel (1941). PEEM techniques were refined and improved by Engel (1966). Engel's PEEM achieved an order-of-magnitude increase in resolution over previous instruments and served as a prototype for the KE-3 emission microscopes developed by Wegmann in the 1970s (1972). At this time it was recognized that PEEM is a surface technique, which requires proper attention to the condition of the sample surface on an atomic scale (Engel, 1966; Griffith and Rempfer, 1987). A detailed

description of the early history of electron lenses and electron microscopy can be found in Griffith and Engel (1991) and in Bauer (1994).

In the development of PEEM, the Oregon microscope project, lead by Griffith and Rempfer, played an important role. Griffith and Rempfer built an ultrahigh vacuum UV-PEEM with a spatial resolution of 10 nm and produced a steady stream of beautiful micrographs of biological samples (Griffith and Rempfer, 1987; Rempfer and Griffith, 1989). Rempfer also investigated the spatial resolution of PEEM and the properties of electron lenses of different shape and using different operation voltages, both in experiment and in theory (Rempfer, 1985; Rempfer and Griffith, 1989). Bauer and Teliëps built an electron microscope with both LEEM and PEEM modes (Teliëps and Bauer, 1987). Instead of photoelectrons, the LEEM employs electron diffraction of low-energy electrons and excels at the investigation of crystalline surfaces, epitaxial films, and film growth.

A relatively recent development is the use of X-rays instead of UV radiation. X-ray photoemission electron microscopy (X-PEEM) was demonstrated for the first time in 1988 by Tonner and Harp (1988). In 1993 Stöhr et al. (1993) showed that X-PEEM can image magnetic domains at high resolution. X-PEEM instrumentation has evolved rapidly during the past decade, and almost every major synchrotron facility is now home to a PEEM instrument (Anders et al., 1999; Heyderman et al., 2003; Kuch et al., 2000; Schneider and Schönhense, 2002; Wei et al., 2003).

PEEM is limited in resolution by the chromatic and spherical aberrations of the electron lenses. It has been shown that an aberration corrector can improve the resolution down to 1 nm (Fink et al., 1997). Currently, two aberration-corrected X-PEEMs are under construction, the “SMART” PEEM at BESSY II (Hartel et al., 2002) and the PEEM-3 at the Advanced Light Source (Wan et al., 2004). In Section 2 we will discuss the basic layout of an X-PEEM experiment. We will also describe the image generation due to X-ray absorption and discuss contrast mechanisms with the aid of some examples. Section 3 will address the electron optics of uncorrected PEEM, and Section 4 will extend the discussion to aberration correction. Section 5 will deal with a very important application of X-PEEM: imaging of magnetic domains. The last section will introduce time-resolved X-PEEM, sometimes called TR-PEEM, which is a very recent development. TR-PEEM is used to image magnetic processes on the nanoscale with picosecond time resolution.

2 X-Ray PEEM

2.1 X-Ray Sources

Electron storage rings, also called synchrotrons, are today’s premier sources of polarized, intense X-rays. Synchrotron radiation is generated when relativistic electrons are deflected, for example, by a dipole magnet. The radiation, which is polarized, spans a wide range of wavelengths from micrometers (infrared) to Ångströms (hard X-rays), and energies from millielectronvolts to kiloelectronvolts (Attwood, 2000).

Bending magnets or so-called insertion devices are used as X-ray sources. Radiation from a dipole or bending magnet is linearly polarized in the storage ring plane. A few degrees above and below, the radiation is right and left elliptically polarized with a high degree of circular polarization of typically 80%. In a bending magnet beamline, the polarization is chosen by a movable aperture, which masks out a vertical slice of the beam. Insertion devices, in particular undulators or elliptically polarizing undulators (EPUs), provide X-rays of much higher intensity. Undulators are multipole magnet structures that repeatedly deflect the electron beam, generating X-rays at each bend that are coherently superimposed. The magnetic structure in an EPU can be adjusted to generate radiation of variable polarization: left and right circularly polarized, linearly polarized in and perpendicular to the ring plane, and often also linearly polarized at intermediate angles. The X-rays from the storage ring are brought through an evacuated beamline to the experiment (Figure 9–1). A grating monochromator in the beamline generates a monochromatic beam with an energy resolution of typically several 100 meV. X-ray photoemission electron microscopes are mostly used in an energy range between about 100 and 2000 eV, called the soft X-ray regime, which spans the absorption edges of many important elements, in particular of the transition metal ferromagnets Fe, Co, and Ni. The X-rays are focused onto the sample using X-ray mirrors to match the X-ray spot size to the field of view of the microscope, usually a few 10s or 100s of micrometers. The electron optics of the X-PEEM experiment image the photoelectron yield of the sample with magnification onto a detector, usually a phosphor and a charge-coupled device (CCD) camera. The CCD digitizes the signal, which can later be processed by a PC.

2.2 The X-Ray Absorption and Photoemission Process

When X-rays are absorbed by matter, electrons are excited from core levels into unoccupied states, leaving empty core states (Figure 9–2). Secondary electrons are excited by the decay of the core hole. Auger processes and inelastic electron scattering create a cascade of low-energy electrons, of which some penetrate the sample surface, escape

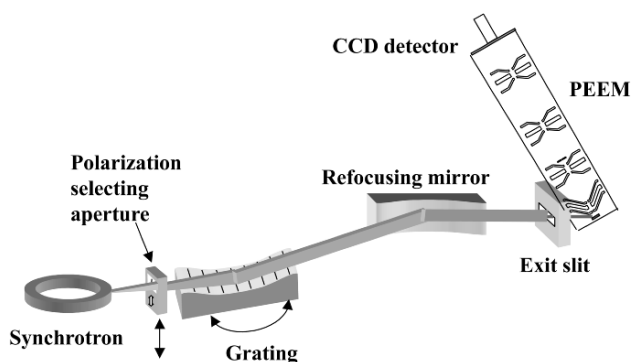


Figure 9–1. Layout of X-ray source, beamline, and PEEM of the Advanced Light Source PEEM-2 microscope.

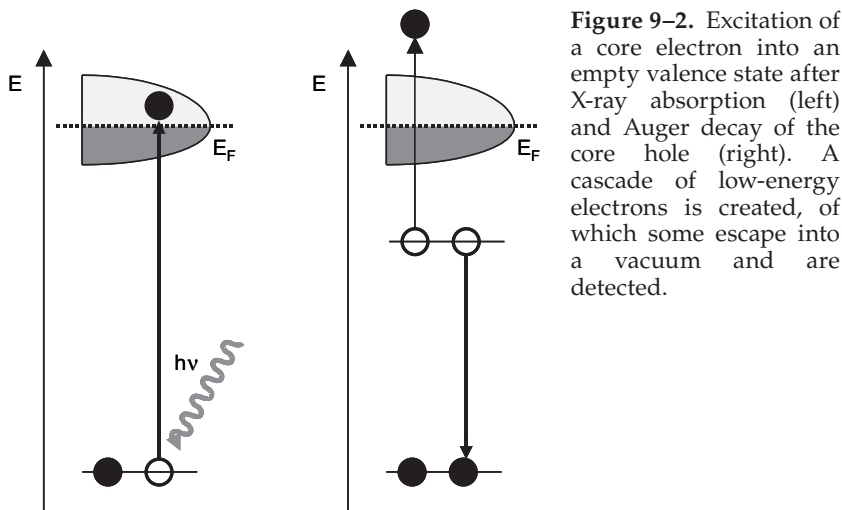


Figure 9–2. Excitation of a core electron into an empty valence state after X-ray absorption (left) and Auger decay of the core hole (right). A cascade of low-energy electrons is created, of which some escape into a vacuum and are detected.

into vacuum, and are collected by the PEEM optics. A wide spectrum of electrons is emitted, with energies between the energy of the illumination and the work function of the sample. This wide electron distribution is the principal source of image aberration in the microscope since electron lenses are chromatic. In the case of X-ray illumination the distribution of the low-energy secondary electrons can be approximated by $n_e \sim E/(E + W)^4$, where W is the work function of the material and E is kinetic energy (Tonner and Dunham, 1994). In Figure 9–3 the electron distributions after X-ray and UV illumination are compared. Note that for X-PEEM in contrast to UV-PEEM a much wider range of energies contribute to the image. Therefore X-PEEM has significantly

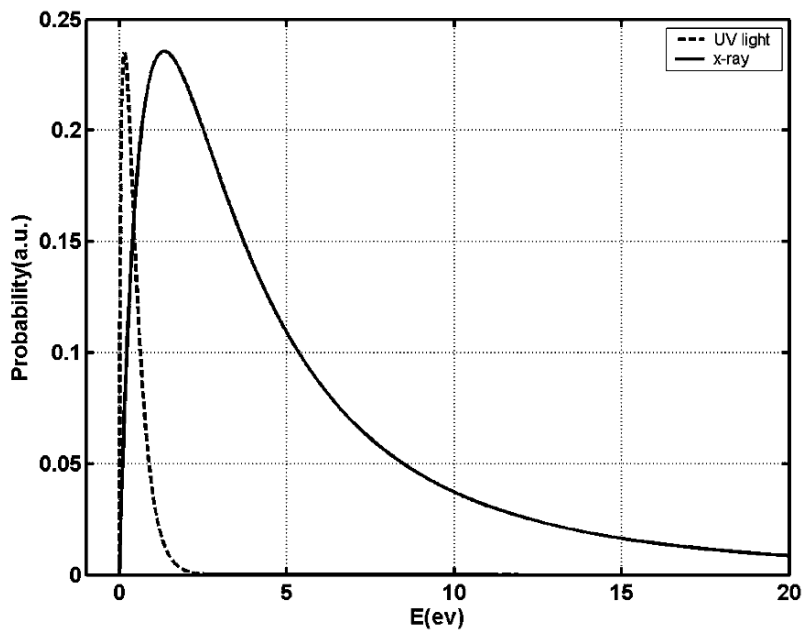


Figure 9–3. Energy distribution of the emitted secondary electrons excited by X-rays (solid line) and UV light (dashed line).

lower resolution than UV-PEEM, if no aberration corrector is used. The effective mean free path λ_{el} of the low-energy secondary electrons sets a limit to the probing depth of the technique. In metals a λ_{el} of 1.5–2 nm is typical (Nakajima et al., 1999). In polymers the effective mean free path is at least twice as large (Ohara et al., 1999; Zharnikov et al., 2002). Image contrast can still be obtained through cover layers with a thickness of up to about three times the effective mean free path.

The total electron emission of the sample is proportional to the X-ray absorption coefficient averaged over the probed thickness. PEEM images the lateral electron emission and thereby maps the local X-ray absorption of the sample as a function of X-ray energy. Local X-ray absorption spectra are obtained by monitoring the local image intensity while scanning the X-ray energy. Therefore, X-PEEM is called a spectromicroscopy technique. PEEM is also a full-field imaging technique, and spectra from many image points are taken in parallel.

2.3 Image Contrast in X-PEEM

Contrast in X-PEEM arises from the element-, chemistry-, and magnetism-specific absorption of different materials. Magnetic contrast will be discussed in Section 5. Two other contrast mechanisms, work function contrast and topographic contrast, appear using both X-ray and UV radiation. Work function contrast is particularly strong if a UV source is used with energy between the work function of two materials that are present at the sample surface. Only the area with low work function emits electrons and appears bright. Topographic contrast arises from rough or structured surfaces. Topographic features shape the electrical potential close to the sample and lead to electron deflection, electron focusing, and defocusing. Usually, experimenters seek to minimize topographic contrast by preparing smooth samples, e.g., single crystals, polished surfaces, or films grown on flat substrates, because sample roughness can lower the image resolution.

2.3.1 Elemental Contrast

Elemental contrast is a consequence of the large enhancement (by factors of 10 and more) of the X-ray absorption right above core-level absorption edges, compared to the absorption below the edge. Resonant absorption into valence states can further increase the absorption at particular photon energies. By tuning the photon energy to a specific absorption edge, areas that predominantly contain that particular element light up in the X-PEEM image. K edges of light elements from C to Si, L edges of transition metals from Ti to Zn and of main group elements from Ca to Br, and M edges of rare-earth metals from La to Yb can be studied with soft X-rays. The technologically most important magnetic elements, the transition metals Cr, Mn, Fe, Co, and Ni and the rare-earth metals Sm, Eu, Gd, and Tb, have absorption edges between 570 and 1300 eV, at which magnetism can be measured easily and quantitatively, as will be discussed later.

The ultimate frontier of PEEM is the imaging and spectroscopy of nanoscale objects. Although current instruments do not yet possess the ability to spatially resolve a nanometer object, the high sensitivity of PEEM allows us to detect the signal from a single nanoobject, if it is

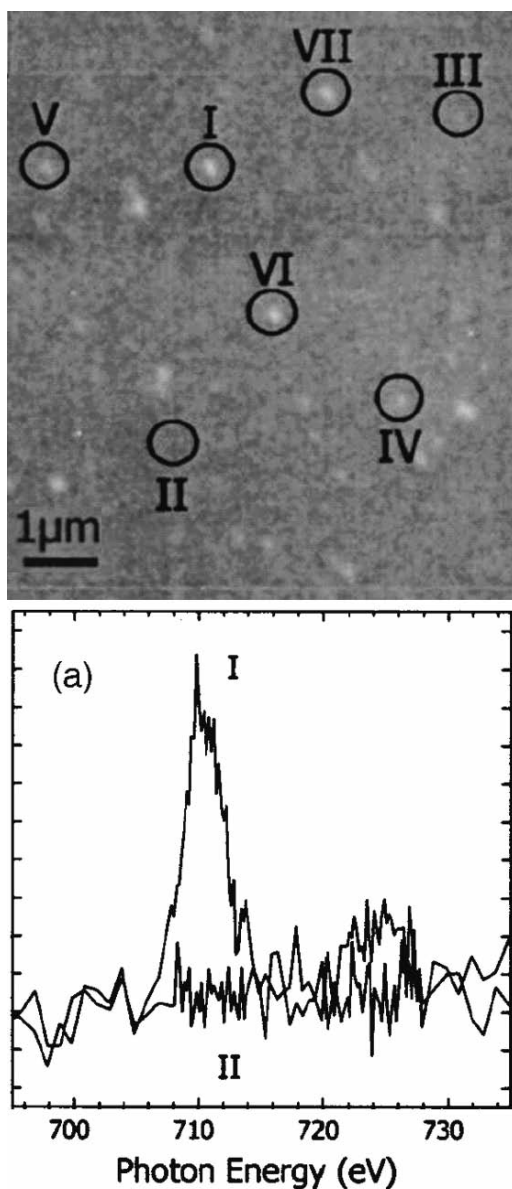


Figure 9-4. High-resolution micrograph of 10-nm iron oxide clusters ($\gamma\text{-Fe}_2\text{O}_3$) imaged by X-PEEM at the Fe edge. Below: local absorption spectra acquired at a cluster position (I) and on the substrate (II). [Reprinted with permission from Rockenberger et al. (2002). Copyright 2002 by the American Institute of Physics.]

sufficiently separated in space from neighboring objects. As an example for the sensitivity of X-PEEM and for elemental contrast imaging, measurements of 10-nm $\gamma\text{-Fe}_2\text{O}_3$ clusters are shown in Figure 9-4 (Rockenberger et al., 2002). Clusters that were brought onto HF-etched Si wafers from solution were studied. Images of these clusters were obtained by subtracting a background image acquired before the Fe L edge from an image acquired on the Fe L₃ peak (at 710 eV). This often-used difference method augments the signal originating from a particular element. Individual clusters are visible as light dots. Local spectra acquired from a single cluster (I) and the clean substrate (II) show a clear Fe signal from a single cluster but only background from the surrounding area.

2.3.2 Chemical Contrast

The fine structure, which is superimposed onto the step-like increase in the absorption at photon energies close to the absorption edge, is called near edge X-ray absorption fine structure, or NEXAFS. NEXAFS is selective regarding the electronic structure and chemical environment of the atom that absorbed the X-ray photon (Stöhr, 1992). This chemical and electronic sensitivity is due to the fact that close to the absorption edge photoexcited core electrons are excited into and therefore probe valence states. The density of these states is determined by the electronic band structure and therefore the electronic, chemical, and magnetic properties of the material. In hydrocarbons spectral features can often be associated with particular bonds and molecular orbitals, allowing a fingerprint analysis of spectra (Stöhr, 1992; Urquhart et al., 1999). In inorganic compounds, the NEXAFS spectrum is often used to determine oxidation states. Figure 9–5 shows Fe and Mn images and spectra of a manganese nodule found on the floor of the deep ocean. Regions that are rich in a material containing Mn and Fe ions appear brighter, when imaged at energies that are specific for the oxidation states of ions. Local spectra acquired in selected regions distinguish the materials that dominate in the core and rim region of the studied growth band, providing information about the genesis of the material (Smith et al., 2003).

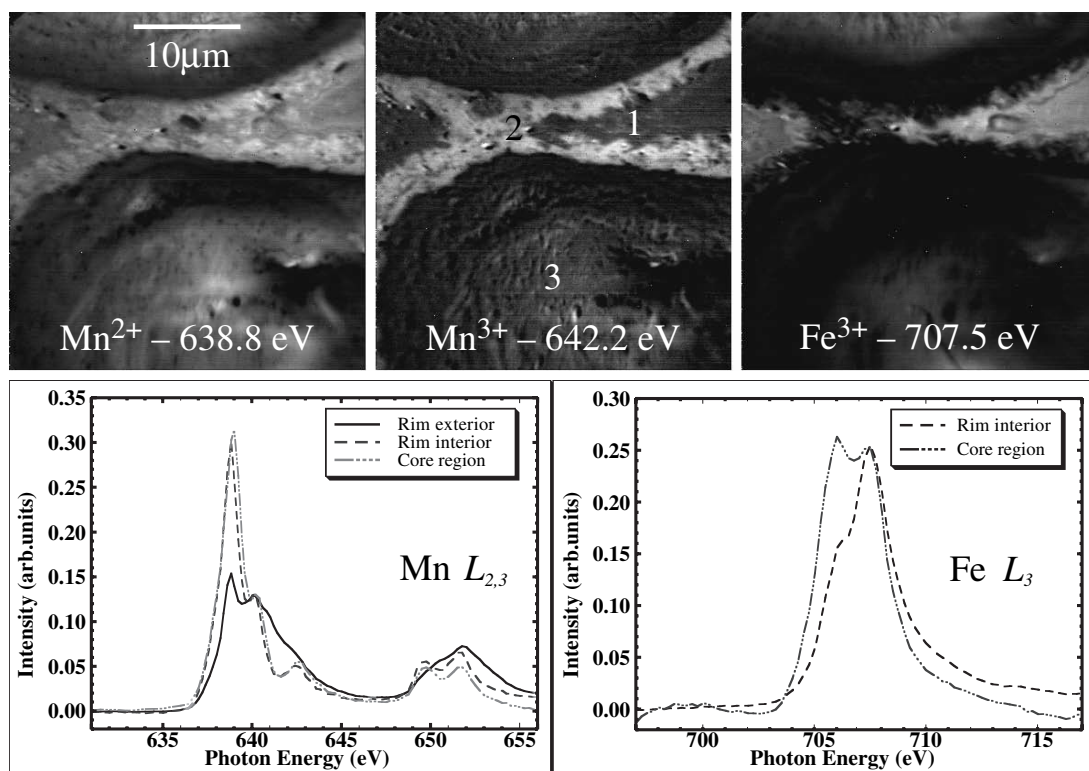


Figure 9–5. Chemical imaging of a manganese nodule from the ocean floor. X-PEEM micrographs and spectra show areas with different Fe and Mn oxidation states. [Reprinted with permission from Smith et al. (2003).³⁵ Copyright 2003 by EDP Sciences.]

3 Uncorrected PEEM Microscopes

Most photoemission electron microscopes have some common characteristics. Modern microscopes use at least three lenses—an objective, an intermediate, and a projective lens—to achieve optimum resolution and magnification. The outline of the electron optics is sketched in Figure 9–6. Some microscopes use magnetic or compound magnetic–electrostatic lenses, but here a microscope based on simple electrostatic lenses will be described. The sample, which is illuminated by UV or X-ray radiation, is located in front of an immersion objective lens. In contrast to light microscopes, the sample is an integral part of the first electrostatic lens of the imaging system. The design of the sample stage in a PEEM is critical. Any vibration or drift is magnified by the optics and reduces the microscope resolution. Significant misalignment of the sample causes coma and also a loss of resolution. The sample–objective lens optics consist of two parts: an acceleration field and a unipotential lens. A strong acceleration field of about 20 kV over 2 mm accelerates the electrons between the sample and the first electrode of the objective lens. The strong acceleration limits the relative energy spread and the angular spread of the electrons that are transported into the microscope. The objective lens together with a contrast aperture in the back focal plane determines the resolution and transmission of the system. By choosing different aperture sizes, resolution can be traded for microscope

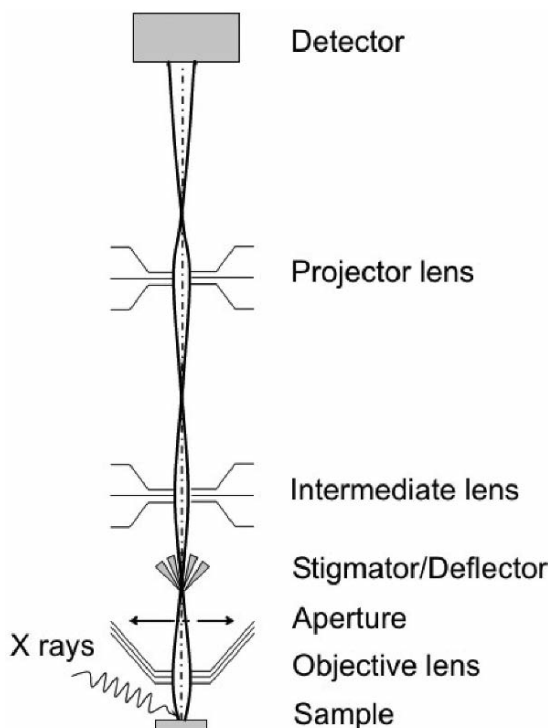


Figure 9–6. Schematic electron optical layout of a photoemission electron microscope.

transmission. A multipole stigmator–deflector is located close to the back focal plane of the objective lens to compensate astigmatism and alignment errors. The intermediate lens and projective lens transfer and magnify the image onto a microchannel plate or a phosphor screen. This phosphor transforms the electron into a visible image, which is captured by a CCD camera.

3.1 The Objective Lens

The objective lens is the most important component of the electron optical system. Triode and tetrode cathode lenses have been used. Chmelik et al. compared several types of cathode lenses (Chmelik et al., 1989). Rempfer et al. analyzed different objective lens types based on extensive electron optical bench measurements (Rempfer et al., 1999; Rempfer and Griffith, 1989).

The lens evaluation occurs in four steps: (1) calculation of the field distribution of the lens, (2) calculation of the electron trajectories and the electron optical properties, (3) evaluation of the interaction of the lens with other parts of the electron optics, and (4) tolerance analysis in regard to mechanical errors. Often an analytical treatment is not possible and numerical methods must be used, which benefit from the rapid progress in computer technology. A charged particle optics (CPO) program was developed by Read for solving charged particle optics problems using the surface charge method (Harting and Read, 1976).^{*} Munro has developed a series of electron optics programs using the finite element method (Hawkes, 1973; Orloff, 1997). Perhaps the most popular program used in electron optics is SIMION 3D (<http://www.simion.com>) which was developed by Dahl (Dahl, 2000). SIMION is an electrostatic and magnetic field modeling program that solves the field using a finite element method and traces the motion of electrons using a fourth-order Range-Kutta integrator. Figure 9–7 shows the equipotential contours of a tetrode objective lens, calculated using SIMION. The acceleration field is approximately uniform in the vicinity of the optical axis except close to the anode aperture. The aperture acts as a thin diverging lens (Rempfer and Griffith, 1989; Davisson and Calbrick, 1932).

The objective lens can be characterized by the trajectory of the principal rays, the field ray and the axial ray (Figure 9–8). The calculation treats the electrodes as a single lens because the fields overlap, but it is instructive to describe the objective lens optics as being composed of (1) the accelerating field, (2) a weak diverging lens, and (3) a unipotential lens. The acceleration field forms a virtual image at unit magnification at twice the sample distance l . The weak diverging lens created by the aperture of the first electrode forms another virtual image at distance $3/l$ behind the sample. It is demagnified by $M = 2/3$. Finally, the unipotential lens forms a real image with a typical magnification of $M = 10\text{--}40$ (Anders et al., 1999; Schneider and Schöhense, 2002; Watts et al., 1997).

^{*} Student versions of the program are available at no cost from <http://www.electronoptics.com/>.

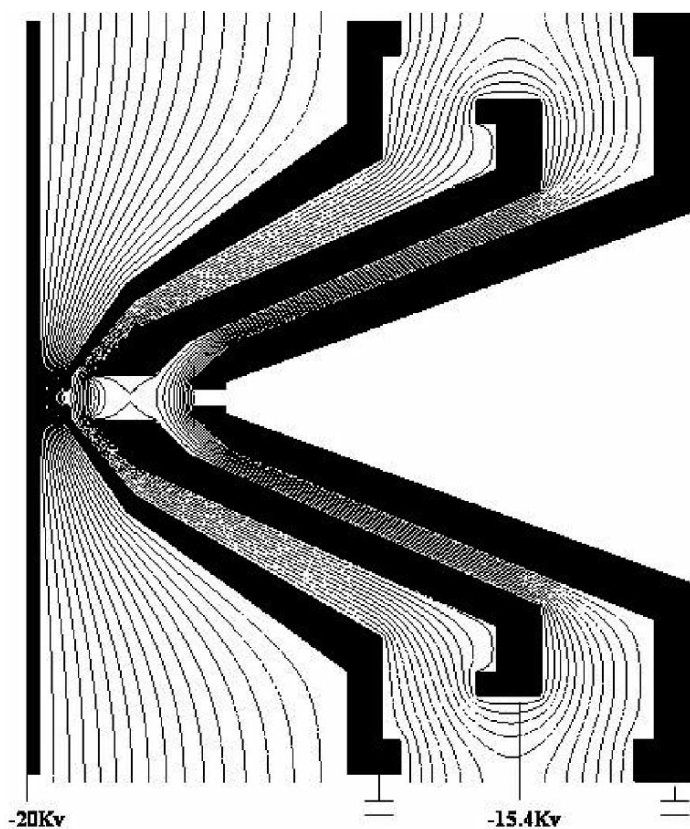


Figure 9-7. Equipotential contours of the electric field of an electrostatic PEEM objective lens at 20 kV sample voltage.

3.2 The Intermediate and Projective Lens

The intermediate and projective lenses magnify the image of the objective lens without distortion to reach a final magnification typically between 30 and 4000. The angle and energy-dependent aberrations of the intermediate and projective lenses are negligible because the energy and angular spread of the electron beam are small here. However, the distortion aberrations need to be minimal because the object size after magnification by the objective lens is significant, fractions of a millimeter or more. Therefore, the intermediate and projective lenses are designed to work at minimum focal length. As the excitation of the electrostatic lens is increased, its focal length decreases and the object side focal point enters the lens field. The distortion coefficient can reach zero at minimum focal length. This type of lens has been examined in detail by Rempfer et al. (1991) who studied rotationally symmetric, three electrode, unipotential lenses (Figure 9-9). The term unipotential means that the image and object side electrodes are on the same potential. Important lens parameters are the thickness of the center electrode t , the interelectrode spacing S , the diameter of the center electrode D , the diameter of the end electrode aperture A , and the radius of the center electrode R . Starting with parameters chosen from diagrams

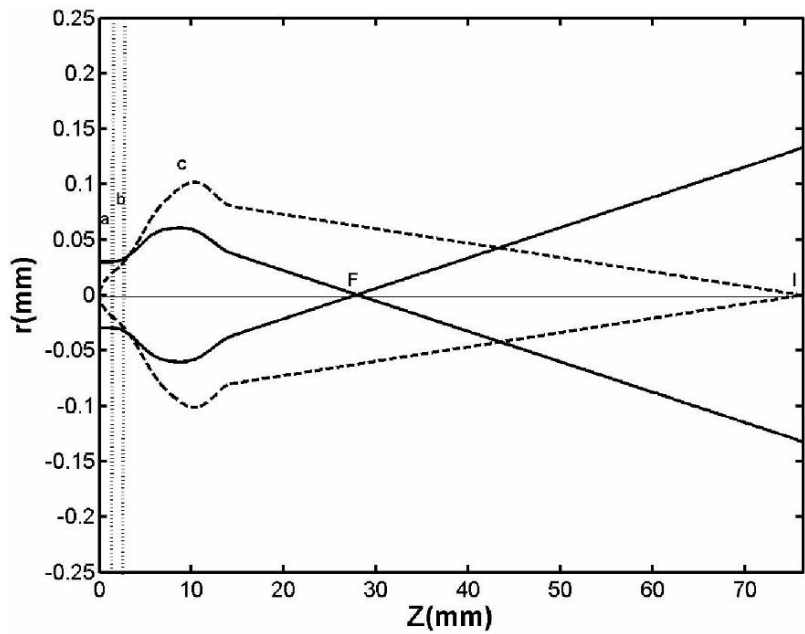


Figure 9–8. Principal electron rays in an objective lens: dashed line, axial ray; solid line, field ray; F, back focal plane; I, image plane.

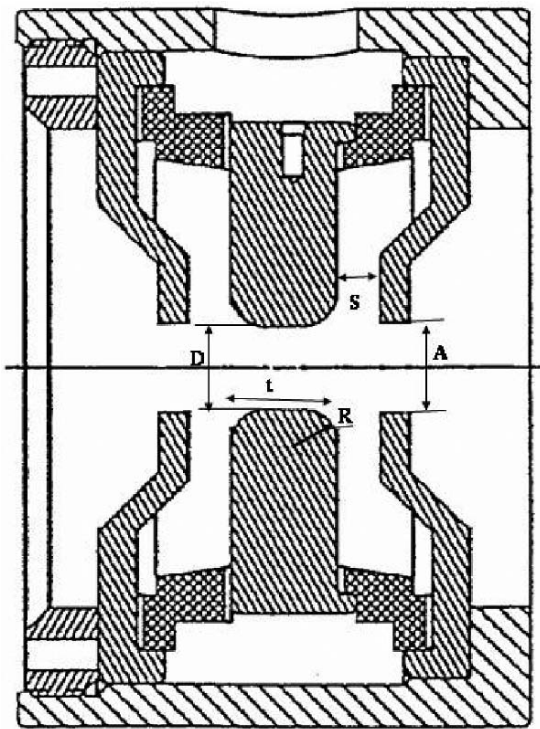


Figure 9–9. Electrode geometry of a symmetric unipotential lens.

published by Rempfer (1985), the lens properties can be optimized for a particular application using numerical methods.

3.3 Transmission and Spatial Resolution of PEEM

The microscope transmission and spatial resolution determine the quality of a PEEM. Bauer (1991) has used these two quantities to define a quality factor. The spatial resolution determines how well spatial details are preserved, and the transmission determines the noise level in an image and the acquisition speed (Jacobsen et al., 1992; Schneider, 1998). The sample is an integral part of PEEM optics, and the optimum resolution can be achieved only on ideally flat surfaces. The effects of roughness and sample tilt have been previously discussed (Marcus, 2001; Nepijko et al., 2000).

3.3.1 Transmission

The microscope transmission is defined as the fraction of emitted electrons that reaches the detector. To reduce aberrations most microscopes contain several apertures of different size that can be moved into the electron beam. Thus transmission can be traded for resolution. The aperture confines the electron beam to the paraxial region and can also limit the energy acceptance of the microscope. Using $E/(E + w_t)^4$ for the electron energy distribution for a work function w_t and using Lambert's law for the electron angular distribution, the transmission of PEEM can be evaluated in a simple closed form

$$T = 1 - \frac{1}{[1 + (a/f_i^*)^2 (U/W_f)]^2} \quad (1)$$

Here, a is the aperture radius, U is the sample potential, and f_i^* is the image side focal length of the objective lens. A high sample potential and a large aperture improve transmission, whereas a long focal length and large work function reduce transmission (Feng et al., 2002).

Figure 9–10 shows SIMION simulations of the angle and energy distribution of the electrons transmitted through apertures of different size. Every point represents one electron. Both the energy spread and the transmission decrease with decreasing aperture size. Note that high-energy, paraxial electrons still pass through the aperture. The resolution is dominated by electron diffraction when the aperture size is lowered beyond a certain minimum value. Therefore, apertures with a diameter larger than 10 μm are usually used.

3.3.2 Spatial Resolution

The spatial resolution of microscopes is characterized by a spread function. The spread function describes the microscope image of a chromatic point source and a diverging electron beam. Diffraction, spherical, and chromatic aberrations are regarded as independent contributions and are added quadratically. The total resolution is given by

$$r_{\text{tot}} = \sqrt{r_s^2 + r_c^2 + r_d^2} \quad (2)$$

where the third-order spherical aberration is defined as

$$r_s = C_s \alpha^3 \quad (3)$$

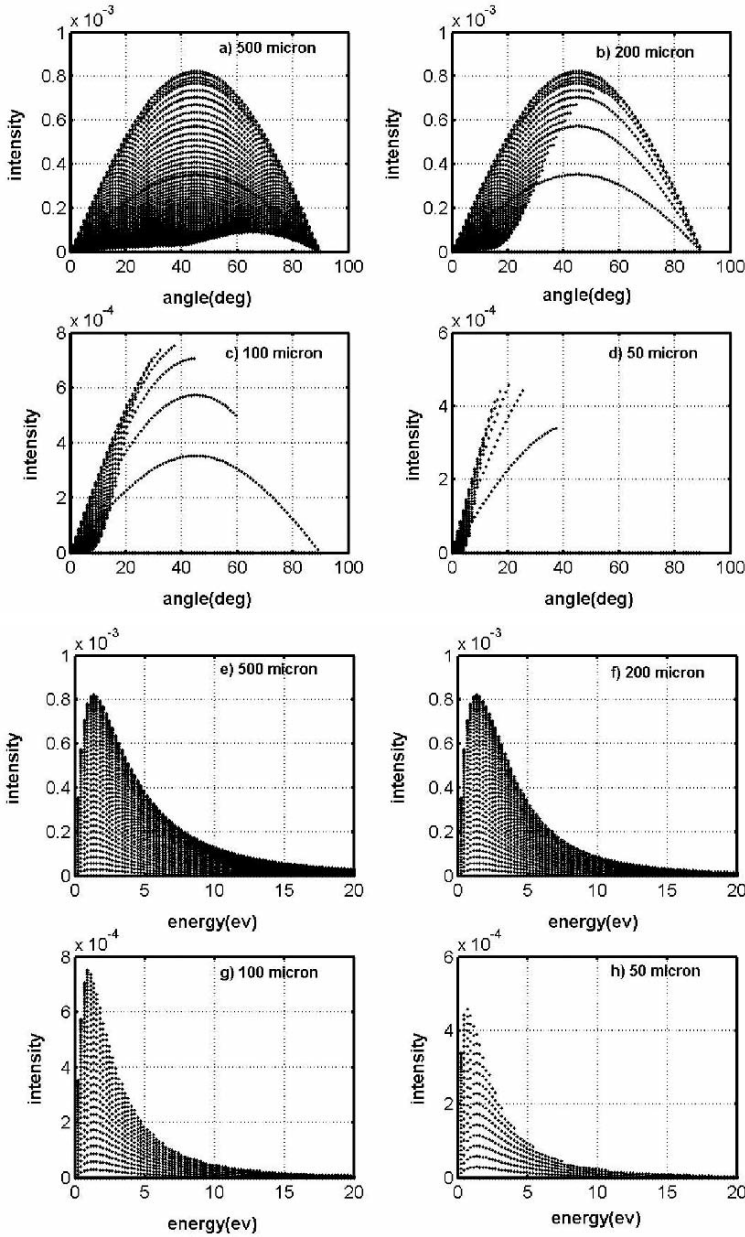


Figure 9-10. SIMION simulations of the angle and energy distribution of the emitted electrons that are accepted by X-PEEM when apertures of different size are used. (a–d) The angle; (e–h) the energy distribution.

the first-order chromatic aberration is defined as

$$r_c = C_c \alpha \delta E / (E) \quad (4)$$

and the diffraction disk is defined as

$$r_d = 0.61 \lambda / \alpha \quad (5)$$

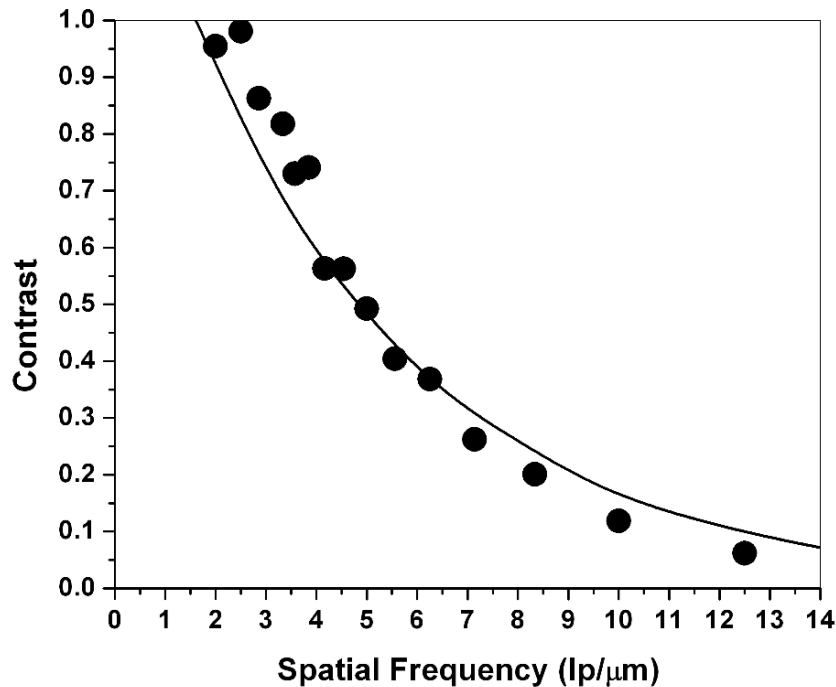


Figure 9-11. Measured (black dots) and calculated (solid line) modulation transfer function (MTF) of PEEM-2 at the Advanced Light Source.

Here, C_s and C_c are the spherical and chromatic aberration coefficients, λ and E are the electron wavelength and the electron energy, δE is the electron energy spread, and α is the half angle of the pencil beam passing the back focal plane aperture. Though Eqs. (2)–(5) can be easily solved, the results are often misleading because the contributions of aberrations and diffraction are not Gaussian and therefore the equations give only a rough estimate of the microscope resolution.

An alternate way of characterizing the spatial resolution is by determining the spatial frequency transfer of the microscope. The intensity distribution of an image produced by a point object under incoherent illumination is called the impulse response function $h(x, y)$. The object intensity distribution $f(x, y)$ and the image intensity distribution $g(x, y)$ are related by the convolution equation

$$g(x,y) = \iint f(x_1,y_1)h(x-x_1,y-y_1)dx_1dy_1 \tag{6}$$

After a Fourier transformation one obtains

$$G(f_x,f_y) = F(f_x,f_y) \cdot H(f_x,f_y) \tag{7}$$

G , F , and H are Fourier transforms of $g(x, y)$, $f(x, y)$, and $h(x, y)$. The function $H(f_x, f_y)$ is the complex optical transfer function (OTF), with real and imaginary parts

$$OTF(f) = |H(f_x,f_y)|e^{i\phi(f)} \tag{8}$$

The real part $|H(f_x, f_y)|$ is called the modulation transfer function or MTF, and the function $\phi(f)$ is called the phase transfer function.

Experimentally, the MTF is measured as the contrast with which spatial frequencies of an object are transferred to an image. Higher frequencies correspond to smaller object features. Figure 9–11 shows the measured and calculated MTF of PEEM-2 at the Advanced Light Source using an extraction field of 20 kV and a 12- μm aperture (Doran et al., in preparation). It starts with unity modulation transfer at low frequencies and gradually drops at higher spatial frequencies. A transfer of 1 indicates that the spatial detail of an object at that frequency is perfectly transferred, while zero transfer indicates a complete loss of information. Values between 0 and 1 indicate partial transfer of spatial frequencies and a reduced contrast. The Rayleigh criterion for spatial resolution is equivalent to a 9% contrast transfer, which is reached for a frequency corresponding to a feature size of 35 nm. Often a higher contrast is required in experiments. At about 80 nm PEEM-2 reaches 40% contrast transfer. The agreement between experiment and simulation over a wide spatial frequency range confirms the accuracy of the simulation.

4 Aberration-Corrected PEEM Microscopes

The improvements in resolution of light and electron microscopes over the past two centuries are compared in Figure 9–12 (Rose, 1994). At the beginning of the twentieth century the light microscope had reached its best resolution after the invention of aberration correction by Abbe, using a combination of a concave and a convex lens, called an achromat lens. The electron microscope soon surpassed the resolution of optical microscopes because of the much smaller wavelength of kilovolt elec-

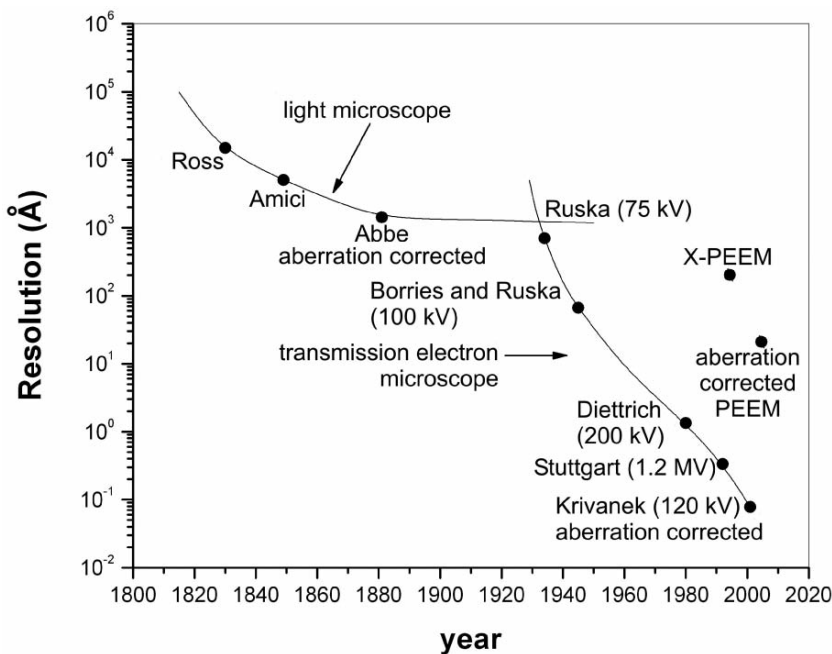


Figure 9–12. History of microscopy techniques and evolution of spatial resolution. Shown are significant steps in the development of optical microscopy, electron microscopy, and X-PEEM. (From Rose, 1994).

trons, effectively removing the effect of diffraction, if electrons of sufficiently high energy are used. Today, the electron microscope has not reached the diffraction limit and lens aberrations still dominate. Among these aberrations, astigmatism, coma, and field distortions can either be easily corrected or they are small close to the optical axis of the microscope*. Early in the development of electron optics theory, Scherzer showed that simple electron lenses always suffer from two other important aberrations: chromatic and spherical aberrations (Scherzer, 1936). The electron-optical analogue to a light-optical achromat does not exist. Different approaches have been tried to remove these aberrations by relinquishing one of the conditions of Scherzer's theorem, which applies only to simple lenses (Hawkes et al., 1996). Multipole aberration correctors, which use a nonrotationally symmetric field, have been successfully used in a low-voltage scanning electron microscope (SEM) (Zach and Haider, 1998), in transmission electron microscopes (TEMs) (Haider et al., 1995, 1998), and in scanning transmission electron microscopes (STEMs) (Dellby et al., 2001; Batson et al., 2002). X-PEEM has much larger aberration coefficients than SEM or TEM, because of the large energy width of the secondary electron distribution, the low initial electron energy, and the large field of view. Therefore, the aberrations of X-PEEM are corrected using an electron mirror, which has approximately opposite chromatic and spherical aberrations of the objective lens.

4.1 Aberrations of an Electron Lens

The electron trajectories in an electron lens are affected by spherical and chromatic aberrations (Figure 9–13). The Gaussian image plane is defined as the intersection of the paraxial ray with the optic axis. The spherical aberration refers to the variation of the focal distance with the angle of a ray from the paraxial ray. An electron passing the lens far from its center experiences a higher field and is focused more strongly. The focus lies before the image plane. The chromatic aberration refers to the variation of the focal distance with the electron energy. Slower electrons experience the field for a longer time and are focused more strongly. The focus lies before (slow electrons) or behind (fast electrons) the image plane.

Figure 9–14 shows the resolution of a typical X-PEEM with a 2-mm acceleration gap and 20-kV acceleration field for a 4-eV work function, broken down according to the source of the aberration. Dominant are the aberration caused by the accelerating field and the chromatic aberration of the lens. Although the resolution can be improved by using a very small aperture, the usefulness of this approach is limited because of the severe reduction in microscope transmission and the onset of diffraction at very small aperture sizes. An ultimate resolution of close to 20 nm has been achieved in several instruments (Anders et al., 1999; De Stassio et al., 1999; Bauer, 1997; Schönhense and Spiecker, 2002).

* A simple electrostatic lens is rotational symmetric, uses static fields, is space charge free, and the electronic lenses does not reverse its direction.

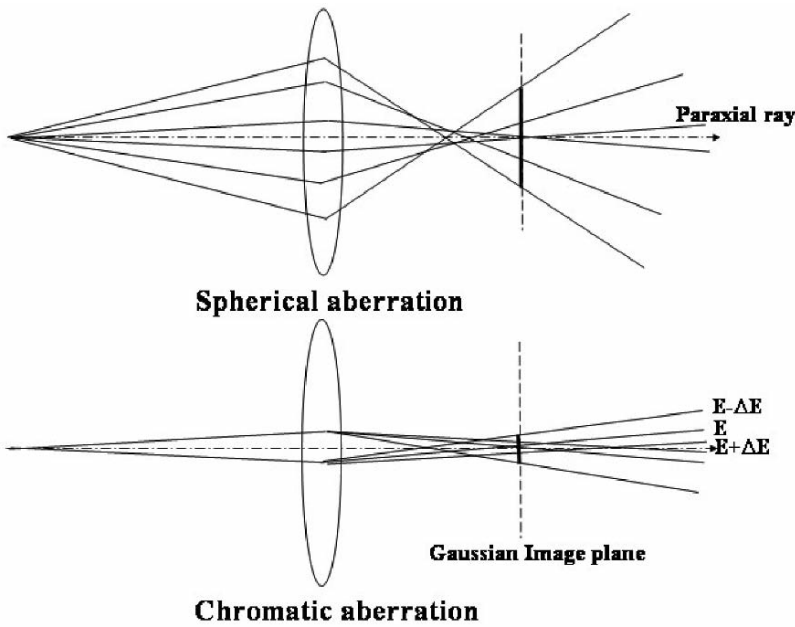


Figure 9-13. Illustration of spherical (top) and chromatic (bottom) aberrations of an electron lens.

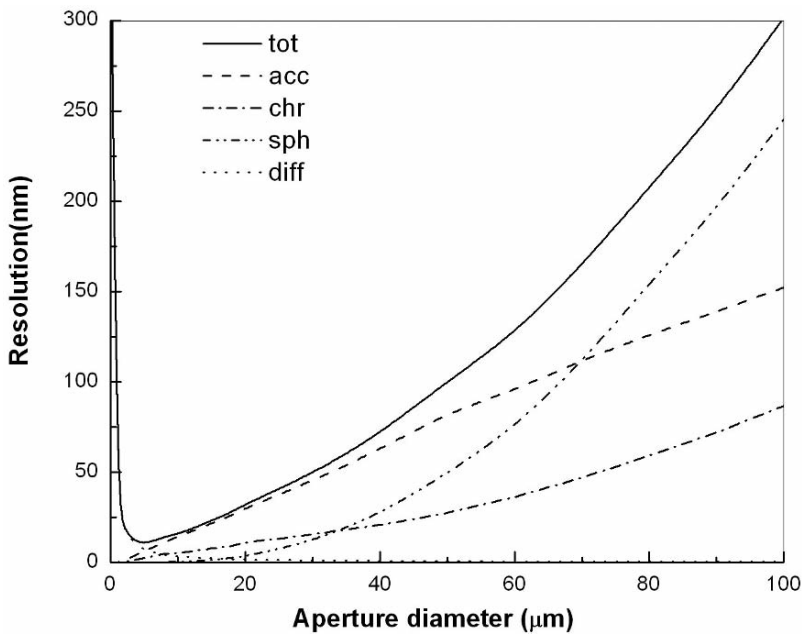


Figure 9-14. Contributions of various aberrations to the total resolution limit on uncorrected PEEM. Sample voltage 20 kV, gap = 2 mm, work function $w_i = 4$ eV. tot, total aberration; acc, aberration of accelerating field; chr, chromatic aberration of objective lens; sph, spherical aberration of objective lens; diff, contribution of diffraction.

Aberration correction promises to significantly improve the ultimate resolution beyond this limit and also to increase the transmission of X-PEEM at moderate resolution by a factor of 100 (Bauer, 2001).

4.2 An Aberration-Corrected Microscope Using an Electron Mirror

The idea of using an electron mirror to correct the chromatic and spherical aberration of a round lens dates back more than half a century (Zworykin and Kosma, 1945; Ramberg, 1949). Extensive studies of electron mirrors have been performed by Kel'man (1973, 1974), Dodin (1981), Rempfer (1990, 1992, 1997), Shao (1990a, 1990b), and Rose (1995, 1997), and co-workers. An electron mirror uses a reflecting electrode with a sufficiently high negative potential to reverse the propagation direction of the electron beam. In 1990, Shao and Wu proposed using an electron mirror with more than two electrodes to be able to freely adjust the aberration coefficients.⁷³ Through numerical analysis of a four-electrode electric mirror, they showed that spherical and chromatic aberrations can be adjusted electrically without changing the image distance of the mirror. Later, a more sophisticated theoretical model using a time-dependent perturbation method was developed by Rose and Preikszas (1995) to fully understand the mirror system. Integral expressions for the aberration coefficients of an electron mirror were derived and higher order aberrations were studied. Currently, two aberration-corrected PEEM microscopes using electron mirrors are being built. The first is called SMART (SpectroMicroscope for All Relevant Techniques) in Germany and has been designed as an ultra-high-resolution spectromicroscope for BESSY II, Germany (Fink et al., 2003). The second is called PEEM-3 at the Advanced Light Source, USA, and has been designed for the study of magnetic materials and polymers. Both microscopes rely on the correction of chromatic and spherical aberrations using a hyperbolic electron mirror, which was pioneered by Rempfer and co-workers (1997). Here, we will focus on the electron optical system of the aberration-corrected PEEM-3 experiment, although many aspects of the two aberration-corrected microscopes are similar and are based on the Rose design.

Figure 9-15 shows a schematic overview of the electron optics of PEEM-3. The electrons travel through the microscope along the dashed lines. The objective lens, together with the field lens, forms a telescopic round lens system. This allows the mirror to run in the so-called symmetric mode in which first-order chromatic distortion and third-order coma vanish and curvature of field effect is reduced (Rose and Preikszas, 1992). A set of two electric dodecapoles steers the beam into the magnetic separator and corrects astigmatism. For a single deflection of 90°, the beam separator images its entrance plane 1:1 onto its exit plane, without introducing aberrations of second order or dispersion of the first or second degree. The electron mirror images the mirror-side exit plane of the beam separator back onto itself. To cancel coma generated by the mirror, the magnification of the mirror is chosen to be -1 (Wan et al., 2004). The separator then transfers the aberration-corrected image to the projector optics, which magnify the image. A pair of electromagnetic dodecapoles in the mirror arm steers the beam into the mirror and back into the separator. A CCD camera behind the mirror is

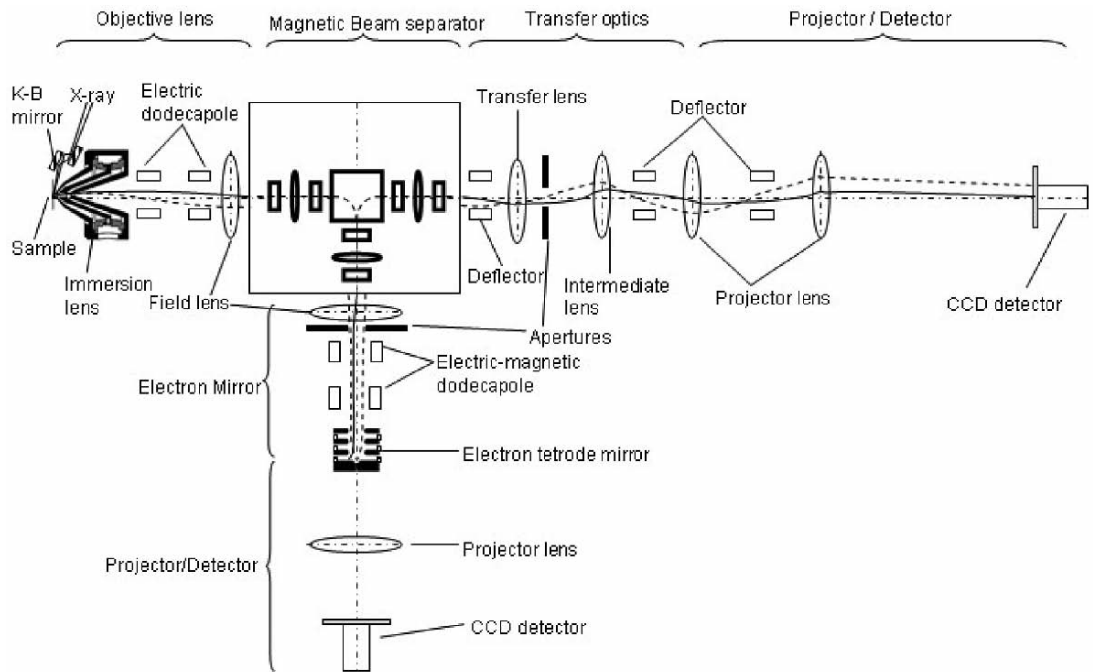


Figure 9–15. Layout of the aberration-corrected PEEM-3 at the Advanced Light Source. An electron mirror is used as an aberration corrector.

used as a diagnostic to test and optimize the objective lens and the first half of the beam separator. In this operation mode, the mirror acts as a unipotential lens and the electron beam passes through a 500- μm hole in the reflecting electrode.

The PEEM-3 electron mirror has four rotationally symmetric electrodes, and the reflecting electrode is a segment of a sphere with a radius of 5.6 mm. The inner electrode is held at ground voltage, while the potentials of the other electrodes provide three free knobs to set the focal length, the chromatic aberration correction, and the spherical aberration correction of the mirror. Figure 9–16 shows the equipotential distribution and correction region of the mirror. This region is sufficiently large to correct the microscope aberrations in various modes of operation.

4.3 The Beam Separator

The task of the magnetic beam separator is to separate the aberration-corrected beam leaving the mirror from the beam that enters the mirror. A separator with very small aberrations is critical for the performance of an aberration-corrected PEEM microscope. Two different separator designs have been considered and are currently being implemented. The SMART microscope uses a triple-bend achromat optic in each of the four sectors of the separator magnet, which has a side length of 280 mm. The route to this type of separator was first identified by Rose and Preikszas⁷⁷ and further developed in the work of Müller et al. (1999) and Wu et al. (2004). The design uses a highly symmetric beam path to cancel a maximum number of aberrations. It is very compact but

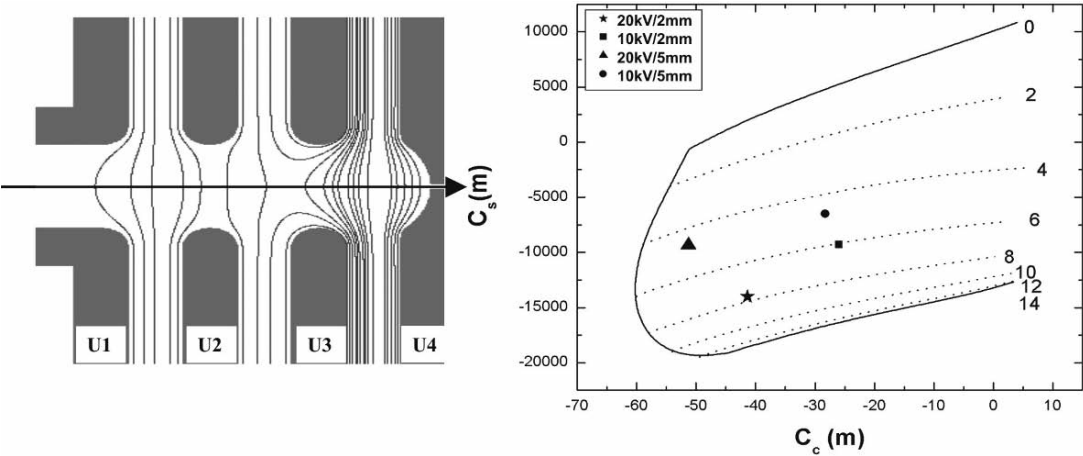


Figure 9-16. Aberration correction by a tetrode electron mirror. Left panel, equipotential contours. Right panel, aberration region that can be corrected by the mirror.

poses high demands on tolerances, and on mechanical and electrical stability (Hartel et al., 2002).

Figure 9-17 shows the 90° beam separator that will be used in the Advanced Light Source PEEM-3. The separator consists of three round

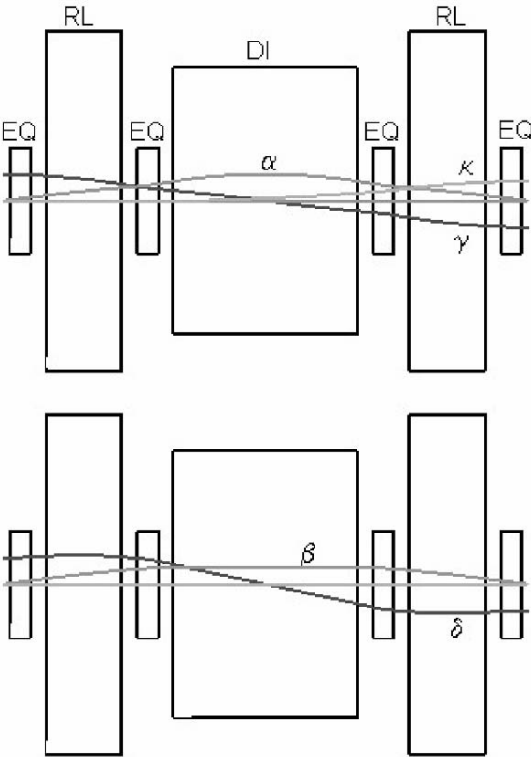


Figure 9-17. Schematic layout and principal rays in the x - z (top) and y - z (bottom) plane of one 90° sector of the PEEM-3 separator. Two 90° bends transport the beam into the mirror and back to the projector optics. One dipole (DI), two round lenses (RL), and four quadrupoles (EQ) are used.

lenses, one dipole magnet, and six electrostatic quadrupoles. Round lenses are chosen to be the main focusing optics. Weak electrostatic quadrupoles are used to form a stigmatic image after each 90° bend. Each 90° bend is mirror symmetric about the center and acts as a telescope in the x - z and the y - z plane. The linear dispersion and the second-order axial aberrations, which are not corrected within each 90° bend, cancel to zero after the second pass through the separator, because of the negative unity magnification of the mirror. Calculations show that the chromatic and spherical aberrations of PEEM-3 and the SMART separator are of the same order (Wan et al., in preparation). The PEEM-3 beam separator uses electrostatic lenses instead of fringe fields for image focusing and is robust regarding the details of the magnetic field in the separator. A simple dipole magnet can be used, and the demands on the power supply stability and mechanical tolerances are relatively low.

4.4 Resolution and Transmission of an Aberration-Corrected PEEM

Aberration correction significantly improves both the spatial resolution and the transmission. The resolution of PEEM-3 was modeled in a way similar as that of PEEM-2. An ensemble of electrons with different initial energy and emission angles was sent through a model of the electron optics, intensity weighted with the probability of each emission energy and angle. The diffraction effect was calculated for each electron in the ensemble and the result was summed up to yield the point-spread function. The resolution was defined as the distance between the 15th percentile and 85th percentile of the point-spread function. The resolution and transmission of PEEM-3 are shown in Figure 9–18 in comparison with the uncorrected PEEM-2 (Feng et al., 2004). PEEM-3 is designed to have two main operation modes, a high-resolution mode and a high-flux mode. Operating at 20kV and 2 mm working distance, the resolu-

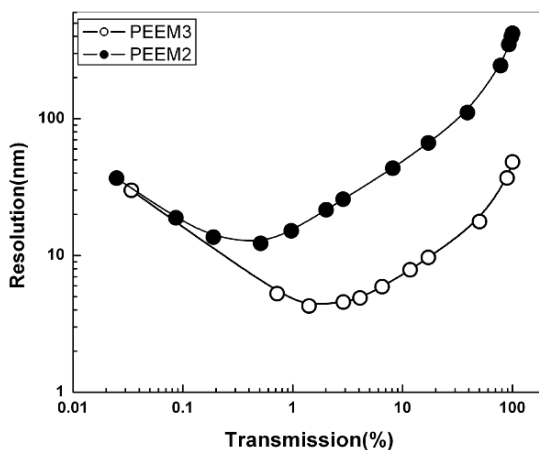


Figure 9–18. Comparison of calculated resolution versus transmission of the Advanced Light Source PEEM-2 and PEEM-3. The acceleration potential is 20kV and the sample–lens distance is 2mm.

tion for 100% transmission (no aperture used) reaches 50 nm using the mirror corrector, a significant improvement over PEEM-2, which reaches only 440 nm. Using a small aperture the optimum resolution will be 5 nm at 2% transmission, as opposed to 20 nm at 1% transmission for PEEM-2. For any chosen spatial resolution, a significantly larger aperture can be used, resulting in a large increase in efficiency over PEEM-2. An aberration-corrected X-PEEM such PEEM-3 will have significantly improved ultimate resolution and transmission, key attributes for the imaging of nanostructures (Bauer, 2001a, 2001b).

5 Application: Magnetic Domain Imaging

Magnetic materials have applications in key areas of information technology, mainly because of low cost and nonvolatility, the ability to retain information without a power source. Magnetic hard disks are the dominant technology in data storage. New applications for magnetic materials emerge when magnetic and electronic properties are concurrently utilized (Wolf et al., 2001), for example, in magnetic random access memory (MRAM) chips. Magnetic storage technology relies on structures with lateral dimensions of well below $1\text{ }\mu\text{m}^2$. The storage density on commercial hard disks in 2004 was about 60 Gbit/in², corresponding to an area of $0.01\text{ }\mu\text{m}^2$ per bit. At a typical bit aspect ratio of about 10:1 this corresponds to a bit length of 30 nm and a track width (or bit width) of 300 nm. In demonstrations and in some products even smaller dimensions have been reached. Progress in device development, and also in the fundamental understanding of the properties of magnetic material, depends on our ability to characterize magnetic materials on a nanoscale.

5.1 Magnetic Circular and Linear Dichroism

X-ray magnetic circular dichroism (XMCD) is a standard method to study magnetic thin films and surfaces (Rortright et al., 1999). The wide availability of polarized and tunable X-rays from synchrotron sources has played an important role in the success of X-ray dichroism techniques. In the past 15 years there has been great progress following the first XMCD spectroscopy measurements at the important transition metal L edges in 1987 by Schütz et al. (1987), the first imaging of ferromagnetic domains in 1993 by Stöhr et al. (1993), and the first imaging of antiferromagnetic domains by Stöhr et al. (1999) and Scholl et al. (2000).

Circularly polarized X-rays measure the direction of the atomic magnetic moment of a ferromagnet relative to the polarization vector of the X-rays. In the presence of spin orbit interaction the photon angular momentum is transferred to the angular momentum of the photoexcited electron, in particular the electron spin, which senses the different density of empty states in each spin channel. An imbalance of the occupation of the majority and minority states is characteristic for a ferromagnet. Strong XMCD effects (up to 40%) appear at the L edges ($2p \rightarrow 3d$ transition) of the transition metal ferromagnets Fe, Co, and Ni, since the d states carry most of the magnetic moment. The XMCD spectrum is defined as the difference spectrum of two absorption

spectra acquired with either opposite polarization or opposite magnetization. The XMCD effect is opposite in sign at the L_3 and L_2 edge because of the opposite sign of the spin-orbit coupling in the $2p$ states: $l + s$ for $2p_{3/2}$, and $l - s$ for $2p_{1/2}$. The different coupling gives rise to a unique feature of XMCD, its ability to separate spin and orbital moment. The spin momentum is proportional to the difference of the integrated XMCD intensity at the L_3 and the L_2 edge and the orbital momentum is proportional to the sum. Sum rules have been developed that are used to quantitatively determine the spin and orbital magnetic moment per atom (Thole et al., 1992; Wu et al., 1992; Carra et al., 1993). The angle and magnetization dependence of XMCD in the total absorption signal is approximately given by $I_{\text{XMCD}} \sim \cos \alpha \langle M \rangle_T$, with α denoting the angle between the X-ray helicity vector E (parallel to the X-ray propagation direction) and the magnetization M . Figure 9–19 shows X-ray absorption spectra of a Co film, which was magnetized parallel (solid line) and antiparallel (dotted line) to the X-ray polarization. The spectra were obtained by measuring the drain current on the sample using a sensitive picoammeter. The drain current is also called total electron yield (TEY) because it is equal to the total number of electrons that leave the sample in response to the photoexcitation. In the lower part of Figure 9–19, the XMCD difference spectrum is shown.

Linearly polarized X-rays probe the anisotropy of the electronic and magnetic structure and are particularly useful for the study of antiferromagnets, which do not possess a macroscopic ferromagnetic moment. In an antiferromagnet with a collinear magnetic structure, linearly polarized X-rays distinguish between a linear polarization vector

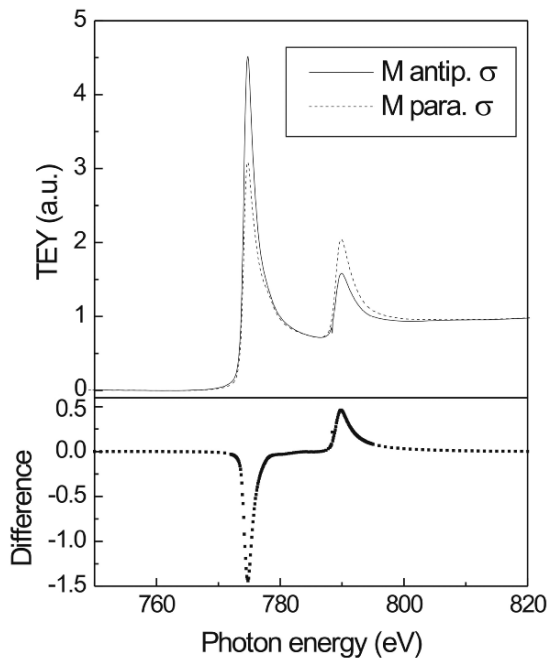


Figure 9–19. Co X-ray magnetic circular dichroism (XMCD) spectra acquired for a magnetization M parallel (dotted line) and antiparallel (solid line) to the X-ray polarization σ . At the bottom the difference spectrum is shown.

aligned with the magnetic spin axis and a linear polarization vector perpendicular to the magnetic spin axis (Ruiper et al., 1993; Alders et al., 1998; Lüning et al., 2003). In transition metal oxides, such as NiO, CoO, Fe₂O₃, Fe₃O₄, LaFeO₃, and others, an X-ray magnetic linear dichroism (XMLD) of up to 10% has been observed. Linearly polarized X-rays probe the angle α between the linear X-ray polarization vector $\mathbf{E}$ and the orientation of the magnetization axis $\hat{\mathbf{A}}$. The angle and magnetization dependence of the XMLD intensity is approximately given by $I_{\text{XMLD}} \sim (1 - 3 \cos^2 \alpha) \langle M^2 \rangle_T$. Here $\langle M^2 \rangle_T$ is the statistical average of the squared moment at the temperature T .

5.2 Imaging of Ferromagnetic Domains

A subtle balance of energies in a magnetic material causes the formation of magnetic domains. Contributions to the total energy are the exchange energy, the magnetic anisotropy, and the dipolar energy. Dipolar fields are the driving force behind domain creation. Domains minimize the magnetic stray field leaving the sample and therefore minimize the dipolar energy. Exchange and anisotropy fields resist domain formation because they usually favor aligned spins. In magnetic thin films interface effects such as interface exchange coupling, interlayer coupling, and interface anisotropy alter this delicate balance and lead to new magnetic phases. Transitions between different magnetic phases occur as a function of film thickness, interlayer thickness, magnetic field, and temperature, and are extensively studied because of the interesting physics underlying the often complex phase diagrams. High-resolution magnetic imaging is an important tool to study magnetic domains and phase transitions. X-ray techniques are especially powerful because they allow us to separate the magnetic state of individual layers in a multilayer, using element-specific imaging. For a review of domain imaging using X-PEEM see Schneider and Schönhense (2002).

As an example, the in- versus out-of-plane transition of a thin face-centered cubic (fcc) Fe film on a five-monolayer Ni/Cu(001) substrate is visualized in Figure 9–20 (Wu et al., 2004). The Ni layer produced an in-plane exchange field that lowered the critical thickness at which the Fe spin reorientation occurred to about 2.6 monolayers. The experiments were conducted on an Fe wedge, which was grown under ultra-high-vacuum (UHV) conditions. The shown images were acquired in the transition region of in- and out-of-plane magnetization. XMCD images were obtained by dividing a PEEM image taken at the Fe L₃ edge by an image taken at the L₂ edge. Since the magnetic dichroism has an opposite sign at these edges, the procedure enhanced the magnetic contrast and suppressed nonmagnet effects. Below the transition thickness, stripe domains formed, which had an out-of-plane magnetization. Approaching the transition, these domains shrank in size. Large in-plane domains appeared above the transition thickness. The experimentally observed thickness dependence of the stripe domain width could be explained within a theoretical model that took into account the exchange interaction, the anisotropy, the dipolar field, and the virtual field created by the Ni layer (Wu et al., 2004).

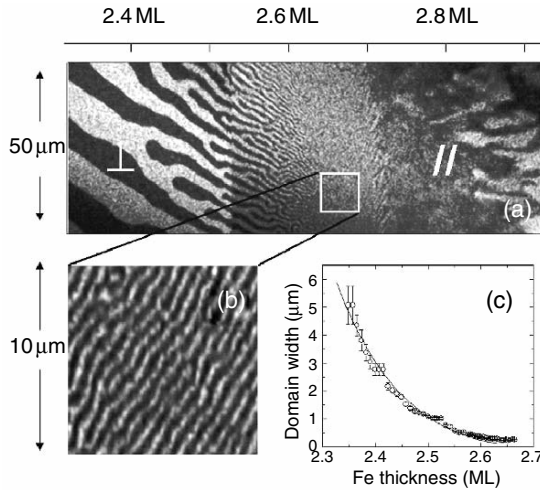


Figure 9–20. (a) Fe stripe domain structure in Fe/Ni/Cu(001) at the spin reorientation transition from an out-of-plane magnetization for small Fe thickness to an in-plane magnetization for larger thickness. (b) Magnification of the transition region. (c) Domain size as a function of Fe thickness. [Reprinted with permission from Wu et al. (2004). Copyright 2004 by the American Physical Society.]

5.3 Imaging of Antiferromagnetic Domains

Antiferromagnetism is a material property that occurs in important material classes: in transition metal oxides, for example, Ni oxide, Co oxide, and Mn oxides, as well as in metallic alloys, for example, Mn alloys. The unidirectional exchange coupling at the interface between an antiferromagnet and a ferromagnet has found particular attention, called exchange bias (Meiklejohn and Bean, 1956; Kiwi, 2001; Berkowitz and Takano, 1999; Nogues and Schuller, 1999). Exchange bias leads to a pinning of the magnetization of the ferromagnet after annealing of the antiferromagnet/ferromagnet structure in a magnetic field. It has found application in technology where it is used to pin the magnetization of a ferromagnetic reference layer in devices that are used as magnetic field sensors, e.g., in hard disk read heads. X-ray spectroscopy and X-ray microscopy have been instrumental in studying and understanding the microscopic magnetic structure of such ferromagnet/antiferromagnet interfaces, because of the sensitivity of X-ray spectroscopy and microscopy to antiferromagnetic order (Stöhr et al., 1999; Scholl et al., 2000, 2001, 2004a, 2004b; Ohldag et al., 2001a, 2001b, 2003; Kuch et al., 2002; Offi et al., 2003a, 2003b, 2004).

LaFeO₃ is an antiferromagnet with a bulk Néel temperature of 740 K. Figure 9–21 shows local X-ray absorption spectra and images acquired using X-PEEM of an epitaxially grown 40-nm LaFeO₃ film on an SrTiO₃(001) substrate (Scholl et al., 2000). The spectra were acquired using linearly polarized X-rays. XMLD effect is visible at the L_{3/2} edges and appears as a variation in the L edge fine structure. Images acquired at energy positions labeled A and B in the spectrum show a dark–bright contrast that due to antiferromagnetic domains

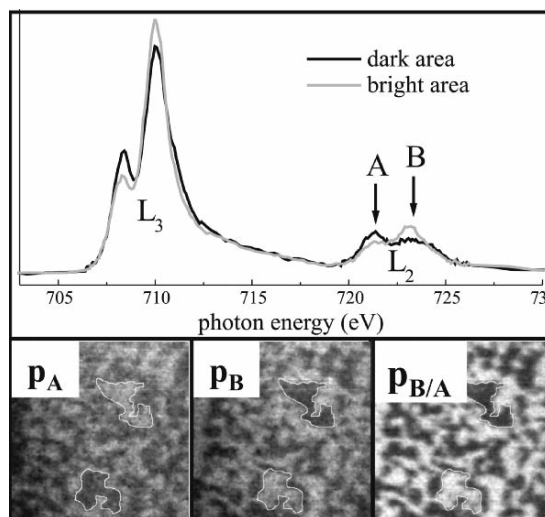


Figure 9–21. X-ray absorption spectra and images of micron-sized antiferromagnetic domains in an LaFeO_3 thin film. The spectra were measured in single domains and show X-ray magnetic linear dichroism at the L resonances. The images were acquired at energies labeled A and B. Image $p_{B/A}$ is the ratio image.

that are less than $1\mu\text{m}$ in size. The image contrast is the result of the different angle between the X-ray polarization and the antiferromagnetic axis in each domain. The ratio image or XMLD image $p_{B/A}$ shows the domains with improved contrast. The spectra were determined from a sequence of images, an image stack, acquired as a function of the energy of the incident X-rays. Local spectra were then calculated within single domains by plotting the local image intensity as a function of the energy at which the image was taken.

To prove that the perceived image contrast had a magnetic origin, the antiferromagnetic ordering transition was studied by acquiring temperature-dependent domain images (Scholl et al., 2000). A reduced ordering, or Néel, temperature of 670 K was observed in the film, compared to a bulk Néel temperature of 740 K (Figure 9–22). The domain contrast was compared with a fit function $\sim \langle M^2 \rangle_T$, which was estimated using mean field theory. The good match between fit and measurement demonstrated that the antiferromagnetic ordering transition in a thin film could be described in the mean field approximation.

The exchange coupling and exchange bias at the interface between a ferromagnet and an antiferromagnet has been attributed to uncompensated spins at the surface of the antiferromagnet (Takano et al., 1997). This conjecture was not directly proven until surface- and interface-sensitive X-ray microscopy techniques such as X-PEEM became available. Figure 9–23 shows magnetic domain images of a $\text{Co}/\text{NiO}(001)$ sample, measured using circular polarization (Ohldag et al., 2001; 2003). The Co film was grown by e-beam evaporation in UHV on a freshly cleaved and *in situ*-annealed $\text{NiO}(001)$ single crystal. Because of the surface sensitivity of PEEM, most of the signal in the images originated from the thin ferromagnet film and the surface region of the antiferromagnet. The Ni moment could be separated from the much larger Co moment, because of the element specificity of the method. The top

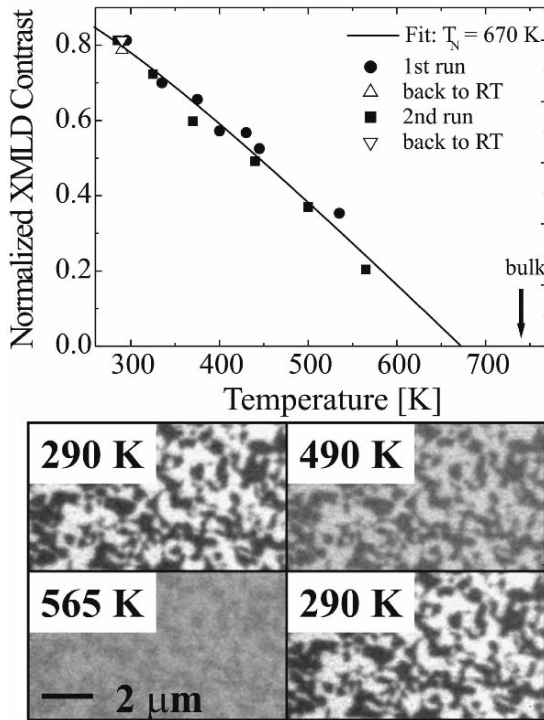


Figure 9–22. Magnetic contrast in LaFeO_3 as a function of temperature. The data of the phase transition are compared with a mean field calculation. [Reprinted with permission from Scholl et al. (2000). Copyright 2000 by AAAS/Science.]

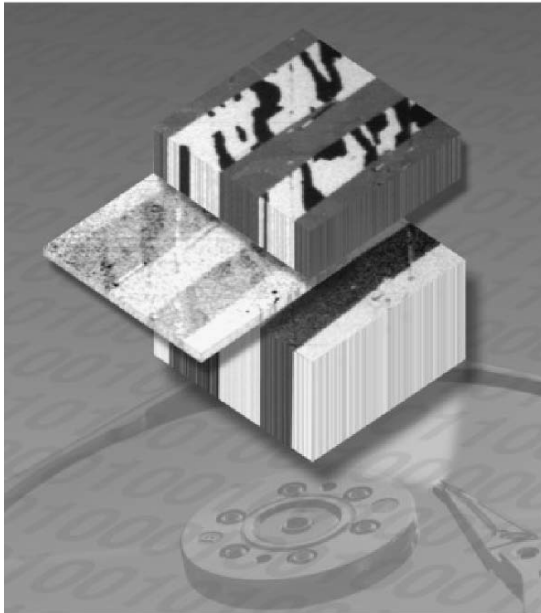


Figure 9–23. Magnetic domain images of Co (top layer, ferromagnetic), NiO (bottom layer, antiferromagnetic), and NiO interface layer (center layer, ferromagnetic).

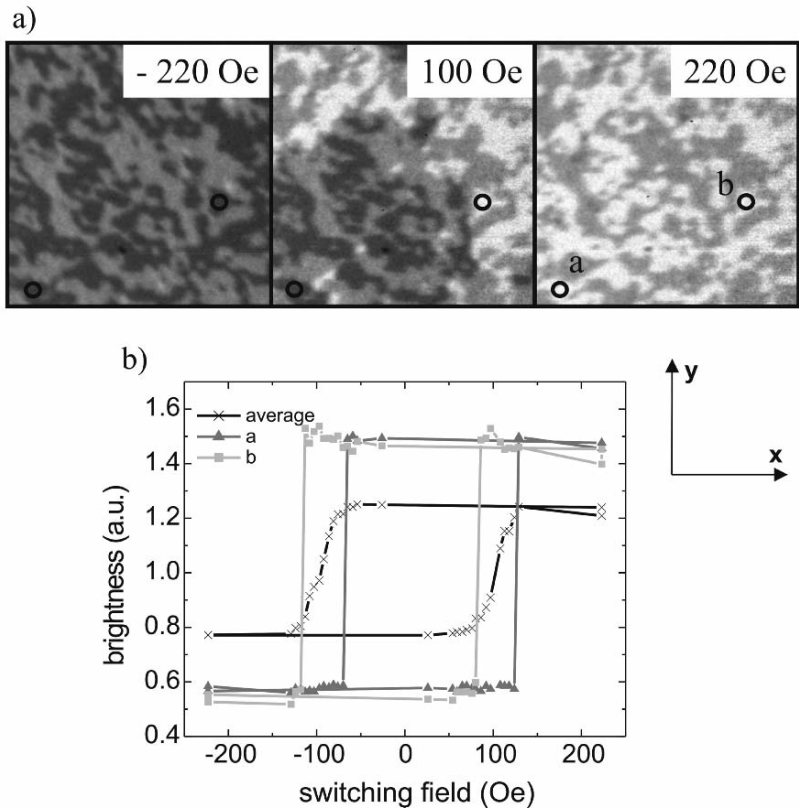


Figure 9–24. Magnetization reversal of Co/LaFeO₃ imaged using XMCD. A local bias exists in single domains (hysteresis loops a and b). Integrated over a larger sample area the bias vanishes. [Reprinted with permission from Nolting et al. (2000).¹⁰⁰ Copyright 2000 by Nature Publishing Group.]

image was acquired using circular polarization at the Co edge and shows ferromagnetic domains, which were aligned domain-by-domain with NiO antiferromagnetic domains (bottom image). The center image was acquired at the Ni edge using circular polarization and shows the small ferromagnetic moment of the Ni interface layer, which mediated the coupling between the NiO antiferromagnet and the Co ferromagnet. The moment was estimated to be close to the moment of one monolayer. Spectroscopy studies of the chemistry of the interface layer showed that the uncompensated spins were a result of a chemical reaction at the interface, which led to a partial oxidization of the Co and a reduction of the surface of the NiO.¹¹¹

Similar results were obtained on Co/LaFeO₃/SrTiO₃(001) (Nolting et al., 2000). The interface coupling led to a uniaxial interface anisotropy and a correlation of the antiferromagnetic and ferromagnetic domains. Furthermore, a unidirectional anisotropy or bias was observed locally in single Co domains. This local bias averaged out to zero over a large area but reached significant values in single, small domains. Remanent magnetization loops of single domains showed a different switching field for opposite directions of the applied field pulse (Figure 9–24). The

remanent loops were acquired by detecting the XMCD intensity in a single domain as a function of the strength of the pulsed field. Selected images of a full loop (about 30–40 images) are shown in Figure 9–24. After applying a field pulse in the $+y$ direction, dark regions that were magnetized into the $-y$ direction switched into the $+y$ direction and became bright. Gray domains were magnetized in the $\pm x$ direction and did not switch, held by the uniaxial anisotropy induced by the anti-ferromagnet. Below, remanent loops of two selected domains, marked in the domain pictures, are plotted. The local bias appears as a horizontal shift of the magnetization loop. The local bias reached values of up to 30 Oe. Also shown is a loop averaged over a large area that shows no macroscopic bias. The intensity of the loop is only half the intensity of the local loops because it averaged over $\pm y$ domains, which switched, and $\pm x$ domains, which did not switch. A statistical analysis of the local switching field demonstrated that the average value of the local bias increased linearly with decreasing inverse domain diameter (Scholl et al., 2004b).

5.4 Linear Dichroism Imaging of Ferromagnets

XMLD was also observed in experiments on ferromagnetic metals (Schwickert et al., 1998). Since linear dichroism is a quadratic effect with a 180° periodicity it allows us to distinguish between domains, which have a 90° rotated magnetization, in contrast to XMCD, which also distinguishes 180° domains. As an example PEEM images acquired with X-PEEM of an Fe/NiO(001) sample are shown (Figure 9–25). The

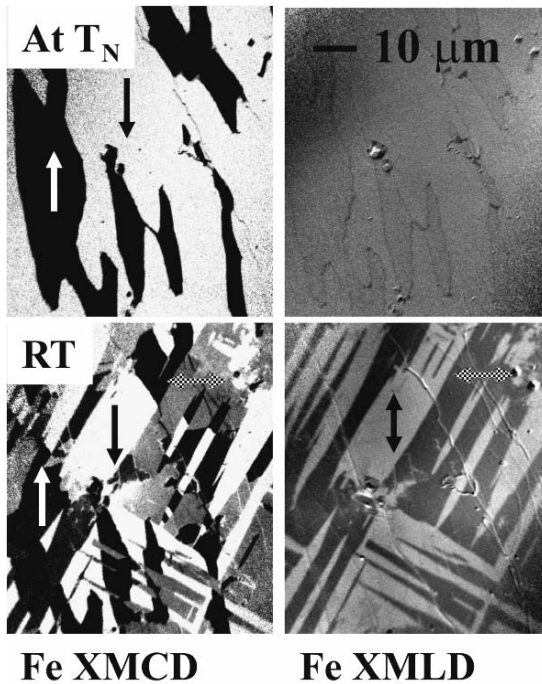


Figure 9–25. X-Ray magnetic circular dichroism (left) and linear dichroism (right) imaging of ferromagnetic Fe domains. Images were acquired at room temperature (bottom) and at the Néel temperature of the NiO substrate (top).

images in the left column were acquired using conventional XMCD at the Fe $L_{3/2}$ edges; the right column shows the corresponding XMLD images, acquired at the left and right side of the Fe L_2 edge using linearly polarized X-rays. In the XMCD image domains appear in three colors, white, black, and gray, corresponding to domains with a magnetization pointing up, down, or left-right; the last are not distinguished. In the XMLD image only two colors appear and domains with a magnetization pointing up or down are not distinguished. The images in the bottom row were acquired at room temperature, and the domain structure is a result of exchange coupling to the antiferromagnetic NiO(001) substrate, which forces the magnetization in the Fe layer along the four in-plane $\langle 110 \rangle$ directions. When the sample is heated above the Néel temperature of NiO (523 K) the coupling vanishes and only two Fe domains remain (up and down). Arrows indicate the local magnetization direction. The XMLD image acquired at elevated temperature shows contrast from Néel-type domain walls (dark lines at the boundary between 180° domains).

6 Time-Resolved Microscopy

6.1 Introduction

Time-resolved X-ray microscopy of magnetization dynamics, although still in its infancy, has made great progress over the past years. Storage rings produce X-ray pulses, which are typically between 30 and 100 ps long. Already with picosecond pulses interesting magnetic processes can be studied: magnetostatic modes, magnetization precession, and magnetic switching (Vogel et al., 2003; Stoll et al., 2004; Choe et al., 2004; Schneider et al., 2004). On a fundamental level, magnetization and spin dynamics have become increasingly important in the field of spin electronics, because the generation, transport, and decay of a spin polarization are important steps in the function of a spin-electronic device. The flow of energy and angular momentum in a magnetic material in response to an external excitation are therefore in the focus of current research (Wolf et al., 2001).

X-Ray dichroism techniques have qualities that make them extremely valuable for the study of magnetization dynamics. Besides elemental and chemical specificity, the XMCD sum rules allow us to disentangle and quantify spin and orbital momentum in each component of a multielement system. The sensitivity to both ferro- and antiferromagnetic order is particularly useful in magnetic oxides, which can be in either magnetic state, depending on composition or temperature. High spatial resolution is important for the investigation of nanoscale devices. In a stroboscopic, or a pump-probe experiment using X-rays, the response of the sample to an excitation (pump) pulse occurring at a time t_0 is measured after a variable delay time Δt (Figure 9–26). X-Ray pulses of current synchrotron radiation sources do not contain a sufficient number of photons for a single shot experiment. Therefore, the experiment has to be repeated at high frequency, the repetition frequency of the X-ray and the excitation pulse, typically many kilohertz

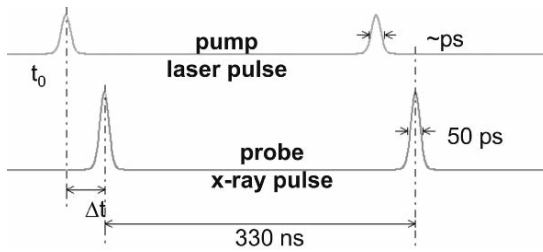


Figure 9-26. Pulse sequence in a laser-synchrotron pump-probe experiment. A fast optical laser pulse excites the sample (pump) and an X-ray pulse probes the dynamic after a variable delay Δt .

or megahertz. The signal is accumulated on the CCD detector over millions of cycles and then read out. The time evolution of the event is measured by varying the delay Δt between pump and probe pulse. Because of the signal averaging, only the repetitive or periodic component of the dynamics of the system is measured. Stochastic events cannot be measured by stroboscopic techniques. The limitation to repeatable dynamics is a lesser constraint in mesoscopic magnetic systems, in which the trajectory of the magnetization is simple and often unique. In the following, experiments of the vortex dynamics in small CoFe patterns will be summarized.

6.2 Experimental Setup

We will describe a technique developed at the Advanced Light Source PEEM-2 (Choe et al., 2004a, 2004b). Similar techniques were developed in other groups using electronic pulsers to produce fast electrical pulses in a waveguide, a high bandwidth circuit (Voget et al., 2003; Stoll et al., 2004; Schneider et al., 2004). At the Advanced Light Source, optically triggered photoconductive switches were used for the pulse generation. The photoconductive switches were triggered by a femto-second laser, synchronized to the X-ray source at a frequency of 125 MHz, producing current pulses with very short rise time. An interdigitated, finger-like switch structure maximized the current. Lithography techniques were used to integrate switch, waveguide, and magnetic structure on the GaAs substrate. The magnetic structures were placed on top or next to the waveguide in the magnetic field of the current pulse. The dynamics of the structures were then monitored by X-PEEM. A schematic drawing of sample and experiment is shown in Figure 9-27 (Choe et al., 2004c).

6.3 *In Situ* Measurement of the Field Pulse

The electrical potential difference between the waveguide and the surrounding ground plane during the field pulse caused a deflection of the X-ray emitted electrons. The deflection of the electrons led to an expansion or breathing of the PEEM image during the pulse. Figure 9-28 (left) shows narrow slices cut from PEEM images of the waveguide and assembled in a new image and ordered by delay time. The measured expansion is proportional to the time-dependent potential

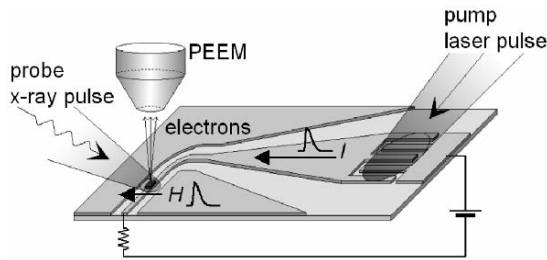


Figure 9–27. Layout of the time-resolved PEEM-2 at the Advanced Light Source experiment. The laser triggers a biased photoconductive GaAs switch and generates a field pulse in a waveguide. The PEEM images the local dynamics of the sample using an X-ray probe pulse. [Reprinted with permission from Choe et al. (2004a). Copyright 2004 by AAAS/Science.]

of the waveguide, and, assuming a resistive load, the current is proportional to that potential. It can be determined quantitatively since the average current drawn by the switch is known. To the right of Figure 9–28, the thus estimated current and the strength of the field pulse are plotted, showing a fast, rising edge, and a slower exponential decay, modulated by a small-amplitude ringing, due to an imperfect impedance match and reflections at the switch and the ground connection. The maximum current at 15 V bias was about 200 mA, which corresponded to a peak field of 13 mT (130 Oe).

6.4 Vortex Dynamics

Figure 9–29 shows a time sequence of PEEM XMCD images of a 2.5-μm-wide and 20-nm-thick triangular, soft-magnetic FeCo structure. The original images and images that were processed using a gradient filter are shown. The filter enhances intensity gradients, e.g., from domain walls. It was observed that the center of the magnetic vortex, a curling magnetization, which is the ground state of the triangular pattern, rotated on a spiraling or gyrotropic trajectory around

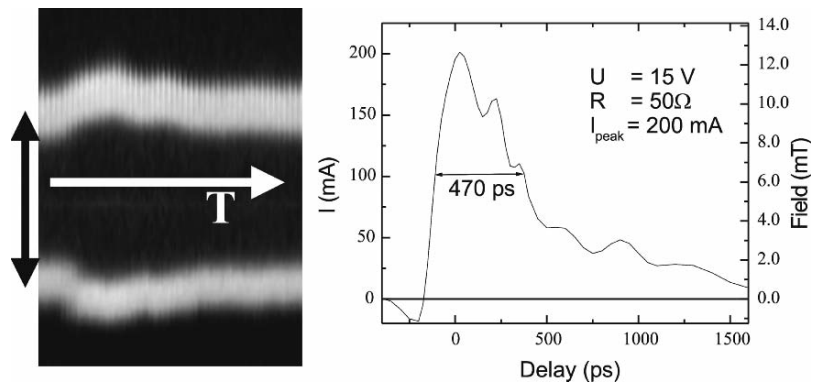


Figure 9–28. Expansion of the waveguide image caused by the pulsed current flowing through the circuit. A pulse with a fast rising edge and less than 500 ps length was recorded. [Reprinted with permission from Choe et al. (2004c). Copyright 2004 by the American Institute of Physics.]

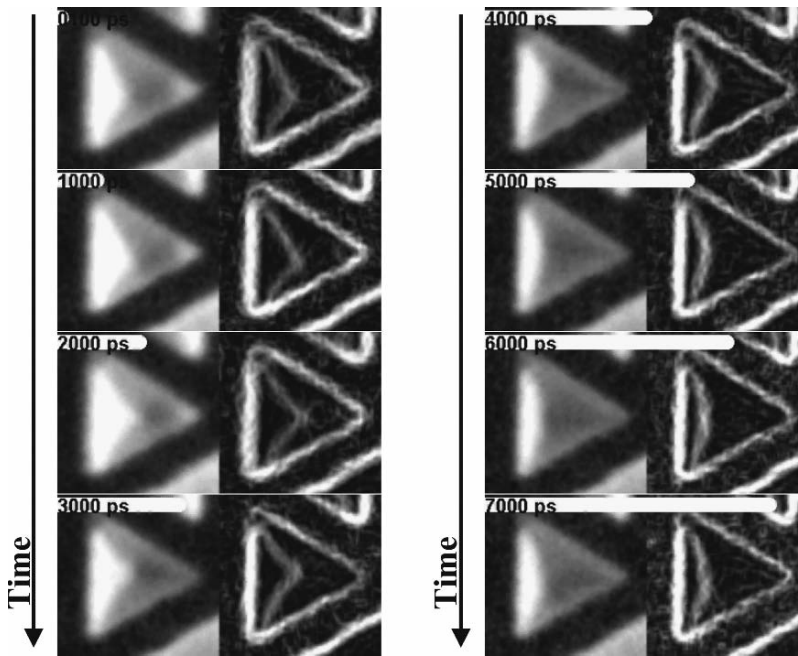


Figure 9–29. Vortex dynamics in a triangular FeCo structure after an in-plane field pulse at $t = 0$. Gradient images are shown next to the XMCD images, showing domain walls with enhanced contrast.

the pattern center. This rotation of the vortex center occurred at a frequency of 125 MHz, commensurate with the repetition rate of the pump laser and the field pulse. The vortex dynamics were quantitatively studied on a variety of patterns, using a particular advantage of full-field imaging techniques: the ability to study the dynamics of a large area in parallel. The vortex dynamics were analyzed in detail on four rectangular structures, which contained a single vortex (Choe et al., 2004a). Domain images and trajectories of the vortex center are shown in Figure 9–30. In all patterns a gyrotropic rotation was observed within the observation range of 8 ns, limited by the time gap between consecutive field pulses. The trajectories showed an initial acceleration parallel to the field pulse (blue arrow) and, subsequently, a gyrotropic rotation of the vortex center around the center of the magnetic structure. Interestingly, when the trajectories of patterns II and III were compared, different acceleration directions and rotation senses were observed in two patterns, which had identical in-plane domain structures. This was the result of a hidden parameter in the system: the magnetization in the center of the vortex, called the vortex core. Earlier microscopic studies observed that the vortex core contained a nanometer-sized region, in which the magnetization rotated out of the plane.¹¹⁸ The time-resolved PEEM study found that the vortex core magnetization is responsible for a handedness or chirality in the vortex pattern, which governed the field response and picosecond dynamics of the system. It was observed that a left-handed vortex accelerated parallel to an applied field, whereas a right-handed vortex accelerated opposite

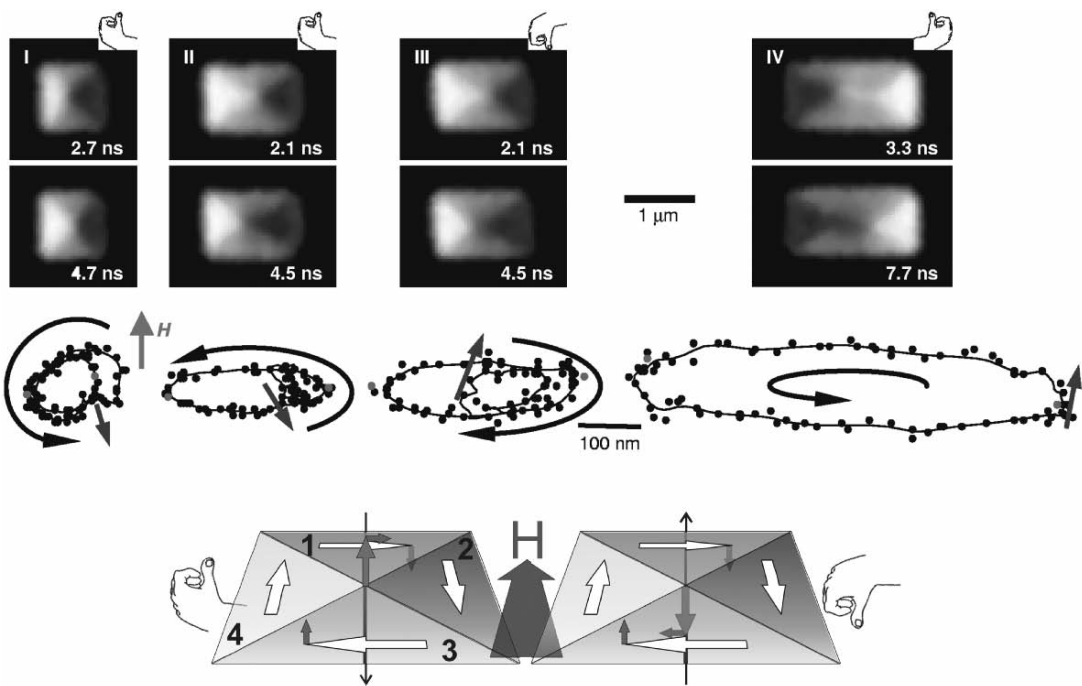


Figure 9–30. Top: Domain images of the in-plane magnetization of Pattern I ($1 \times 1 \mu\text{m}^2$), Patterns II and III ($1.5 \times 1 \mu\text{m}^2$), and Pattern IV ($2 \times 1 \mu\text{m}^2$), taken at the specified delay times after the field pulse. The images are part of a time series extending over 8 ns and were chosen so that the horizontal displacement of the vortex has maximum amplitude. Hands illustrate the vortex handedness and the out-of-plane core magnetization as determined from the vortex dynamics. Middle: Trajectories of the vortex core. The dots represent sequential vortex positions (100 ps step). Bottom: Spin structure of a left-handed (left side) and a right-handed (right side) square vortex. Red arrows represent the precessional torque generated by the external magnetic field (purple arrow). [Reprinted with permission from Choe et al. (2004a). Copyright 2004 by AAAS/Science.]

to the field, explaining the observed, opposite rotation sense in two patterns with identical in-plane domain structures but opposite core magnetization. The availability of high-resolution microscopy was essential for the real-space study of the vortex rotation, because of the small amplitude of the vortex motion ($\sim 100 \text{ nm}$).

7 Conclusion

X-PEEM as a hybrid technique stands between the worlds of electron microscopy, from which it inherits its high spatial resolution, and of X-ray spectroscopy with its sensitivity to the electronic, chemical, and magnetic structure of solids. It excels at chemical and magnetic imaging of thin films and surfaces, because of its surface sensitivity and its sensitivity to the chemical environment and magnetic structure. Fast dynamics on a picosecond time scale can be measured in pump-probe style using the short and intense X-ray pulses of modern synchrotron sources. New, faster X-ray sources and detection schemes are at the brink of being realized and promise even better temporal resolution.

New schemes of aberration correction, employing electron mirrors, will soon push down the resolution from a current 20–50 nm to only a few nanometers, opening the door to many exciting applications in the field of nanoscience.

Acknowledgments. The authors would like to thank H. Ohldag, S.-B. Choe, and F. Nolting for their contributions to the experimental work and a rewarding collaboration and in particular W. Wan, A.A. MacDowell, N. Kelez, and R.D. Schlueter for their contributions to the PEEM-3 design and development. This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

References

- Alders, D., Tjeng, L.H., Voogt, F.C., Hibma, T., Sawatzky, G.A., Chen, C.T., Vogel, J., Sacchi, M. and Iacobucci, S. (1998). *Phys. Rev. B (Cond. Matt.)* **57**, 11623–11631.
- Anders, S., Padmore, H.A., Duarte, R.M., Renner, T., Stammer, T., Scholl, A., Scheinfein, M.R., Stöhr, J., Seve, L. and Sinkovic, B. (1999). *Rev. Sci. Instr.* **70**, 3973–3981.
- Attwood, D.T. (2000). *Soft X-Rays and Extreme Ultraviolet Radiation: Principles and Applications* (Cambridge University Press, Cambridge, U.K.).
- Batson, P.E., Dellby, N. and Krivanek, O.L. (2002). *Nature* **418**, 617–620.
- Bauer, E. (1991). *Ultramicroscopy* **36**, 52–62.
- Bauer, E. (1994). *Rep. Prog. Phys.* **57**, 895.
- Bauer, E. (1994). *Surface Sci.* **299**, 102.
- Bauer, E. (1997). *Chemical, Structural, and Electronic Analysis of Heterogeneous Surfaces on Nanometer Scale* (Kluwer Academic Publ., Dordrecht).
- Bauer, E. (2001a). *J. Electron Spectrosc. Rel. Phenomena* **114–116**, 975–987.
- Bauer, E. (2001b). *J. Phys.—Cond. Matt.* **13**, 11391–11404.
- Berkowitz, A.E. and Takano, K. (1999). *J. Magn. Magn. Mat.* **200**, 552–570.
- Brüche, E. (1933). *Z. Phys.* **86**, 448.
- Brüche, E. and Johansson, H. (1932). *Ann. Phys.* **15**, 145.
- Carra, P., Thole, B.T., Altarelli, M. and Xindong, W. (1993). *Phys. Rev. Lett.* **70**, 694–697.
- Chmelik, J., Veneklasen, L. and Marx, G. (1989). *Optik* **83**, 155–160.
- Choe, S.B., Acremann, Y., Bauer, A., Scholl, A., Doran, A., Stöhr, J. and Padmore, H.A. (2004c). *AIP Conf. Proc.* **705**, 1391–1394.
- Choe, S.B., Acremann, Y., Scholl, A., Bauer, A., Doran, A., Stöhr, J. and Padmore, H.A. (2004a). *Science* **304**, 420–422.
- Dahl, D.A. (2000). *Int. J. Mass Spectrom.* **200**, 3.
- Davissón, C.J. and Calbick, C.J. (1932). *Phys. Rev.* **42**, 580.
- De Stasio, G., Perfetti, L., Gilbert, B., Fauchoux, O., Capozzi, M., Perfetti, P., Margaritondo, G. and Tonner, B.P. (1999). *Rev. Sci. Instr.* **70**, 1740–1742.
- Dellby, N., Krivanek, O.L., Nellist, P.D., Batson, P.E. and Lupini, A.R. (2001). *J. Electron Microsc.* **50**, 177–185.
- Dodin, A.L. and Nesvizhskii, M.B. (1981). *Soviet Phys.—Tech. Phys.* **26**, 539–541.
- Doran, A., Scholl, A., Feng, J., Padmore, H.A., Chao, W. and Anderson, E. in preparation.

- Einstein, A. (1905). *Ann. Phys.* **17**, 132.
- Engel, W. (1966). *Proc. 6th Int. Cong. Electron Microsc., Kyoto, Japan*, **217**.
- Feng, J., MacDowell, A.A., Duarte, R., Doran, A., Forest, E., Kelez, N., Marcus, M., Munson, D., Padmore, H., Petermann, K., Raoux, S., Robin, D., Scholl, A., Schlueter, R., Schmid, P., Stohr, J., Wan, W., Wei, D.H. and Wu, Y. (2004). *AIP Conf. Proc.* **705**, 1070–1073.
- Feng, J., Padmore, H., Wei, D.H., Anders, S., Wu, Y., Scholl, A. and Robin, D. (2002). *Rev. Sci. Instr.* **73**, 1514–1517.
- Fink, R., Weiss, M.R., Umbach, E., Preikszas, D., Rose, H., Spehr, R., Hartel, P., Engel, W., Degenhardt, R., Wichtendahl, R., Kühlenbeck, H., Erlebach, W., Ihmann, K., Schlogl, R., Freund, H.J., Bradshaw, A.M., Lilienkamp, G., Schmidt, T., Bauer, E. and Benner, G. (1997). *J. Electron Spectrosc. Rel. Phenomena* **84**, 231–250.
- Griffith, O.H. and Engel, W. (1991). *Ultramicroscopy* **36**, 1.
- Griffith, O.H. and Rempfer, G.E. (1987). *Adv. Opt. Electron Microsc.* **10**, 269.
- Haider, M., Rose, H., Uhlemann, S., Schwan, E., Kabius, B. and Urban, K. (1998). *Ultramicroscopy* **75**, 53–60.
- Haider, M., Uhlemann, S., Schwan, E., Rose, H., Kabius, B. and Urban, K. (1998). *Nature* **392**, 768–769.
- Hartel, P., Preiksaz, D., Spehr, R., Müller, E. and Rose, H. (2002). *Advances in Imaging and Electron Physics*, Vol. 120 (Elsevier Science, New York).
- Hartel, P., Preiksaz, D., Spehr, R., Müller, H. and Rose, H. (2002). *Advances in Imaging and Electron Physics*, Vol. 120 (Elsevier Science, New York).
- Harting, E. and Read, F. (1976). *Electrostatic Lenses* (Elsevier Scientific, Amsterdam).
- Hawkes, P., Ed. (1973). *Image Processing and Computer-Aided Design in Electron Optics* (Academic Press Inc, New York).
- Hawkes, P.W. and Kasper, E. (1996). *Principles of Electron Optics*, Vol. 2 (Academic Press, New York).
- Heyderman, L.J., Nolting, F. and Quitmann, C. (2003). *App. Phys. Lett.* **83**, 1797–1799.
- Jacobsen, C., Kirz, J. and Williams, S. (1992). *Ultramicroscopy* **47**, 55–79.
- Kel'man, V.M., Sekunova, L.M. and Yakushev, E.M. (1973). *Soviet Phys.—Tech. Phys.* **17**, 2279.
- Kel'man, V.M., Sekunova, L.M. and Yakushev, E.M. (1974). *Soviet Phys.—Tech. Phys.* **18**, 1142.
- Kiwi, M. (2001). *J. Magn. Magn. Mat.* **234**, 584–595.
- Knoll, M., Hontermans, F.G. and Schulzer, W. (1932). *Z. Phys.* **78**, 340.
- Kortright, J.B., Awschalom, D.D., Stöhr, J., Bader, S.D., Idzerda, Y.U., Parkin, S.S.P., Schuller, I.K. and Siegmann, H.C. (1999). *J. Magn. Magn. Mat.* **207**, 7–44.
- Kuch, W., Chelaru, L.I., Offi, F., Jing, W., Kotsugi, M. and Kirschner, J. (2004). *Phys. Rev. Lett.* **92**, 017201/1–4.
- Kuch, W., Gilles, J., Offi, F., Kang, S.S., Imada, S., Suga, S. and Kirschner, J. (2000). *J. Electron Spectrosc. Rel. Phenomena* **109**, 249–265.
- Kuch, W., Offi, F., Chelaru, L.I., Kotsugi, M., Fukumoto, K. and Kirschner, J. (2002). *Phys. Rev. B (Cond. Matt. Mat. Phys.)* **65**, 140408/1–4.
- Kuiper, P., Searle, B.G., Rudolf, P., Tjeng, L.H. and Chen, C.T. (1993). *Phys. Rev. Lett.* **70**, 1549–1552.
- Langmuir, D.B. (1937). *Proc. Inst. Rad. Eng.* **25**, 977.
- Lüning, J., Nolting, F., Scholl, A., Ohldag, H., Seo, J.W., Fompeyrine, J., Locquet, J.P. and Stöhr, J. (2003). *Phys. Rev. B* **6721**, 4433.
- Marcus, M. (2001). *Surface Sci.* **480**, 203–207.

- Meiklejohn, W.H. and Bean, C.P. (1956). *Phys. Rev.* **102**, 1413–1414.
- Müller, H., Preikszas, D. and Rose, H. (1999). *J. Electron Microsc.* **48**, 191–204.
- Nakajima, R., Stohr, J. and Idzerda, Y.U. (1999). *Phys. Rev. B (Cond. Matt.)* **59**, 6421–6429.
- Nepijko, S.A., Sedov, N.N., Schönhense, G., Escher, M., Bao, X. and Huang, W. (2000). *Ann. Phys.* **9**, 441–451.
- Nogues, J. and Schuller, I.K. (1999). *J. Magn. Magn. Mat.* **192**, 203–232.
- Nolting, F., Scholl, A., Stöhr, J., Seo, J.W., Fompeyrine, J., Siegwart, H., Locquet, J.P., Anders, S., Lüning, J., Fullerton, E.E., Toney, M.F., Scheinfein, M.R. and Padmore, H.A. (2000). *Nature* **405**, 767–769.
- Offi, F., Kuch, W., Chelaru, L.I., Fukumoto, K., Kotsugi, M. and Kirschner, J. (2003a). *Phys. Rev. B—Cond. Matt.* **67**, 94419–1–9.
- Offi, F., Kuch, W., Chelaru, L.I., Kotsugi, M. and Kirschner, J. (2003b). *J. Magn. & Magn. Mat.* **261**, L1–6.
- Ohara, H., Yamamoto, Y., Kajikawa, K., Ishii, H., Seki, K. and Ouchi, Y. (1999). *J. Synchrotron Radiat.* **6**, 803–804.
- Ohldag, H., Regan, T.J., Stöhr, J., Scholl, A., Nolting, F., Luning, J., Stamm, C., Anders, S. and White, R.L. (2001a). *Phys. Rev. Lett.* **87**, 247201/1–4.
- Ohldag, H., Scholl, A., Nolting, F., Anders, S., Hillebrecht, F.U. and Stöhr, J. (2001). *Phys. Rev. Lett.* **86**, 2878–2881.
- Ohldag, H., Scholl, A., Nolting, F., Arenholz, E., Maat, S., Young, A.T., Carey, M. and Stöhr, J. (2003). *Phys. Rev. Lett.* **9101**, 7203.
- Orloff, J., Ed. (1997). *Handbook of Charged Particle Optics* (CRC Press, Boca Raton, FL).
- Preikszas, D. and Rose, H. (1997). *J. Electron Microsc.* **46**, 1–9.
- Ramberg, E.G. (1949). *J. App. Phys.* **20**, 183–186.
- Reckangel, A. (1941). *Z. Phy.* **117**, 689.
- Regan, T.J., Ohldag, H., Stamm, C., Nolting, F., Lüning, J., Stöhr, J. and White, R.L. (2001). *Phys. Rev. B—Cond. Matt.* **64**, 214422/1–11.
- Rempfer, G.F. (1985). *J. App. Phys.* **57**, 2385.
- Rempfer, G.F. (1990). *J. App. Phys.* **67**, 6027–6040.
- Rempfer, G.F. and Griffith, O.H. (1989). *Ultramicroscopy* **27**, 273.
- Rempfer, G.F. and Mauck, M.S. (1992). *Optik* **92**, 3–8.
- Rempfer, G.F., Desloge, D.M., Skoczylas, W.P. and Hayes Griffith, O. (1997). *Microsc. Microanal.* **3**, 14–27.
- Rempfer, G.F., Skoczylas, W.P. and Griffith, O.H. (1991). *Ultramicroscopy* **36**, 196.
- Rockenberger, J., Nolting, F., Lüning, J., Hu, J. and Alivisatos, A.P. (2002). *J. Chem. Phys.* **116**, 6322–6328.
- Rose, H. (1994). *Ultramicroscopy* **56**, 11–25.
- Rose, H. and Preikszas, D. (1992). *Optik* **92**, 31–44.
- Rose, H. and Preikszas, D. (1995). *Nucl. Inst. Methods Phys. Res. Section A* **363**, 301–315.
- Scherzer, O. (1936). *Z. Phy.* **101**, 593.
- Schneider, C.M. and Schönhense, G. (2002). *Rep. Prog. Phys.* **65**, R1785–1839.
- Schneider, C.M., Kuksov, A., Krasnyuk, A., Oelsner, A., Neeb, D., Nepijko, S.A., Schönhense, G., Monch, I., Kaltoven, R., Morais, J., de Nadai, C. and Brookes, N.B. (2004b). *Appl. Phys. Lett.* **85**, 2562–2564.
- Schneider, G. (1998). *Ultramicroscopy* **75**, 85–104.
- Schönhense, G. and Spiecker, H. (2002). *J. Vacuum Sci. Tech. B* **20**, 2526–2534.
- Scholl, A., Liberati, M., Arenholz, E., Ohldag, H. and Stöhr, J. (2004a). *Phys. Rev. Lett.* **92**, 247201/1–4.
- Scholl, A., Nolting, F., Seo, J.W., Ohldag, H., Stöhr, J., Raoux, S., Locquet, J.-P. and Fompeyrine, J. (2004b). *App. Phys. Lett.* **85**, 4085–4087.

- Scholl, A., Stöhr, J., Lüning, J., Seo, J.W., Fompeyrine, J., Siegwart, H., Locquet, J.P., Nolting, F., Anders, S., Fullerton, E.E., Scheinfein, M.R. and Padmore, H.A. (2000). *Science* **287**, 1014–1016.
- Schütz, G., Wagner, W., Wilhelm, W., Kienle, P., Zeller, R., Frahm, R. and Materlik, G. (1987). *Phys. Rev. Lett.* **58**, 737–740.
- Schwickert, M.M., Guo, G.Y., Tomaz, M.A., O'Brien, W.L. and Harp, G.R. (1998). *Phys. Rev. B—Cond. Matt.* **58**, R4289–4292.
- Shao, Z. and Wu, X.D. (1990a). *Rev. Sci. Inst.* **61**, 1230–1235.
- Shao, Z. and Wu, X.D. (1990b). *Optik* **84**, 51–54.
- Shinjo, T., Okuno, T., Hassdorf, R., Shigeto, K. and Ono, T. (2000). *Science* **289**, 930–932.
- Smith, A.D., Schofield, P.F., Scholl, A., Pattrick, R.A.D. and Bridges, J.C. (2003). *J. Physique IV* **104**, 373–376.
- Stöhr, J. (1992). *NEXAFS Spectroscopy*, Vol. 25 (Springer-Verlag, Berlin).
- Stöhr, J., Scholl, A., Regan, T.J., Anders, S., Lüning, J., Scheinfein, M.R., Padmore, H.A. and White, R.L. (1999). *Phys. Rev. Lett.* **83**, 1862–1865.
- Stöhr, J., Wu, Y., Hermsmeier, B.D., Samant, M.G., Harp, G.R., Koranda, S., Dunham, D. and Tonner, B.P. (1993). *Science* **259**, 658–661.
- Stoll, H., Puzic, A., van Waeyenberge, B., Fischer, P., Raabe, J., Buess, M., Haug, T., Hollinger, R., Back, C., Weiss, D. and Denbeaux, G. (2004). *App. Phys. Lett.* **84**, 3328–3330.
- Takano, K., Kodama, R.H., Berkowitz, A.E., Cao, W. and Thomas, G. (1997). *Phys. Rev. Lett.* **79**, 1130–1133.
- Telieps, W. and Bauer, E. (1987). *Ultramicroscopy* **17**, 57.
- Thole, B.T., Paolo, C., Sette, F. and van der Laan, G. (1992). *Phys. Rev. Lett.* **68**, 1943–1946.
- Tonner, B.P. and Dunham, D. (1994). *Nucl. Inst. Methods Phys. Res. A* **347**, 436–440.
- Tonner, B.P. and Harp, G.R. (1988). *Rev. Sci. Inst.* **59**, 853–858.
- Urquhart, S.G., Hitchcock, A.P., Smith, A.P., Ade, H.W., Lidy, W., Rightor, E. G. and Mitchell, G.E. (1999). *J. Electron Spectrosc. Rel. Phenomena* **100**, 119–135.
- Vogel, J., Kuch, W., Bonfim, M., Camarero, J., Pennec, Y., Offi, F., Fukumoto, K., Kirschner, J., Fontaine, A. and Pizzini, S. (2003). *App. Phys. Lett.* **82**, 2299–2301.
- Wan, W., Feng, J. and Padmore, H.A. in preparation.
- Wan, W., Feng, J., Padmore, H.A. and Robin, D.S. (2004). *Nucl. Inst. Methods Phys. Res. Section A* **519**, 222.
- Watts, R.N., Liang, S., Levine, Z.H., Lucatorto, T.B., Polack, F. and Scheinfein, M.R. (1997). *Rev. Sci. Inst.* **68**, 3464–3476.
- Wegmann, L. (1972). *J. Microsc.* **96**, 1.
- Wei, D.H., Hsu, Y.J., Klauser, R., Hong, I.H., Yin, G.C. and Chuang, T.J. (2003). *Surface Rev. Lett.* **10**, 617–624.
- Wolf, S.A., Awschalom, D.D., Buhrman, R.A., Daughton, J.M., von Molnar, S., Roukes, M.L., Chtchelkanova, A.Y. and Treger, D.M. (2001). *Science* **294**, 1488–1495.
- Wu, Y., Stöhr, J., Hermsmeier, B.D., Samant, M.G. and Weller, D. (1992). *Phys. Rev. Lett.* **69**, 2307–2310.
- Wu, Y.K., Robin, D.S., Forest, E., Schlueter, R., Anders, S., Feng, J., Padmore, H. and Wei, D.H. (2004). *Nucl. Inst. Methods Phys. Res. Section A* **519**, 230–241.
- Wu, Y.Z., Won, C., Scholl, A., Doran, A., Zhao, H.W., Jin, X.F. and Qiu, Z.Q. (2004). *Phys. Rev. Lett.* **93**, 117205/1–4.

Zach, J. and Haider, M. (1995). *Optik* **98**, 112–118.

Zharnikov, M., Frey, S., Heister, K. and Grunze, M. (2002). *J. Electron Spectrosc. Rel. Phenomena* **124**, 15–24.

Zworykin, V.K. and Kosma, V. (1945). *Electron Optics and the Electron Microscope* (Wiley, New York).

10

Aberration Correction

Peter W. Hawkes

1 Introduction

It has been known since the early days of electron optics that the rotationally symmetric lenses employed in electron microscopes and similar instruments suffer from severe aberrations that cannot be eliminated by skillful lens design (Scherzer, 1936). Immense effort has been devoted to finding lenses with small aberrations and devising aberration correctors. The original demonstration that the two most important aberrations, spherical and chromatic, cannot be eliminated required that several conditions be satisfied and, by relaxing one or the other of these conditions, correctors can be designed. A near-exhaustive list was published by Scherzer (1947) and reviews charting trends in thinking about aberration correction and progress in implementing correctors are to be found in Septier (1966), Hawkes (1980), and Hawkes and Kasper (1989). These contain very full accounts of earlier attempts to correct aberrations with extensive reference lists and the material presented there is not always reproduced here. In particular, a survey of attempts to build apochromats and aplanatic lenses by H. Rose and colleagues in Darmstadt is to be found in the article by Hawkes (1980).

The types of corrector that seem most promising today are examined below but first, we describe the various kinds of aberration and explain why they are important. We then look more closely at the aberration coefficients themselves, which leads naturally to a study of the correctors.

Some familiarity with basic electron optics is assumed here. In particular, the reader is expected to be acquainted with the paraxial properties of lenses and the cardinal elements that characterize them. The *Handbook of Charged Particle Optics* edited by J. Orloff (1997) is recommended for readers who wish to brush up their knowledge of electron lenses, notably the chapters by Munro, Tsuno, Lencová, and Hawkes. For very full accounts, see Hawkes and Kasper (1989) and Rose (2008).

2 Types of Aberration

Lens aberrations are of three kinds: geometric aberrations, chromatic aberrations, and parasitic aberrations

2.1 Geometric Aberrations

An ideal lens would provide point-to-point mapping of the structure of an object into an image and such imaging is indeed predicted by the simplest approximate description of the effect of a lens on an electron beam, the *paraxial approximation*. This is valid provided that the electrons remain close to the optic axis (the axis of symmetry in the case of a round lens) and that the electron trajectory remains inclined at a small angle to this axis. When the electrons depart too far from the axis or the trajectories are inclined at a steeper angle, the paraxial approximation is perturbed by geometric aberrations.

In the paraxial approximation, electrons (of charge $-e$ and rest mass m_0) satisfy the linear, homogeneous, second-order differential equation

$$\frac{d}{dz}(\hat{\phi}^{1/2}x') + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}^{1/2}}x = 0 \quad (1)$$

with an identical equation for $y(z)$, in which $\phi(z)$ is the axial electrostatic potential and $B(z)$ is the magnetic induction on the optic axis, which coincides with the coordinate axis z . Primes denote differentiation with respect to z . The field and potential expansions are given in Appendix I; for derivations of these, see Hawkes and Kasper (1989, Chapter 7). Since the chapter by Rose in *High-Resolution Imaging and Spectrometry of Materials* edited by F. Ernst and M. Rühle is a very relevant reference (Rose, 2002c) to much of the material presented here, the relation between his notation and that adopted here is also given. The relativistically corrected potential $\hat{\phi}$ is given by

$$\hat{\phi} = \phi(1 + \varepsilon\phi) \quad (2)$$

where $\varepsilon = e/2m_0c^2 \approx 0.1 \text{ MV}^{-1}$; $\gamma = 1 + 2\varepsilon\phi$ and $\eta = (e/2m_0)^{1/2} \approx 3 \times 10^5 \text{ C}^{1/2} \text{ kg}^{-1/2}$. The distances $x(z)$ and $y(z)$ are the rotating coordinates routinely employed in the study of round magnetic lenses.

Such differential equations have the general solution

$$x(z) = Ax_1(z) + Bx_2(z)$$

where $x_1(z)$ and $x_2(z)$ are linearly independent solutions of Eq. (1) and A, B are constants. It is frequently convenient to choose for $x_1(z)$ and $x_2(z)$ the solutions $g(z)$ and $h(z)$ that satisfy the boundary conditions

$$\begin{aligned} g(z_0) &= h'(z_0) = 1 \\ g'(z_0) &= h(z_0) = 0 \end{aligned} \quad (3)$$

where z_0 is the object plane, whereupon we have

$$\begin{aligned} x(z) &= x_0g(z) + x'_0h(z) \\ y(z) &= y_0g(z) + y'_0h(z) \end{aligned} \quad (4)$$

The cardinal elements of the lens, its focal lengths and the positions of its focal planes and principal planes, are defined with the aid of the rays $g(z)$ and $h(z)$. It is convenient to write $u = x + iy$ and it can readily be shown that

$$\begin{pmatrix} u \\ u' \end{pmatrix} = \begin{pmatrix} \frac{z_{Fi} - z}{f_i} & T_{12} \\ -1/f_i & \frac{z_o - z_{Fo}}{f_i} \end{pmatrix} \begin{pmatrix} u_o \\ u'_o \end{pmatrix} \quad (5)$$

If u is proportional to u_o for all values of the gradient u'_o , the matrix element T_{12} vanishes. The plane z is then said to be conjugate to z_o and we write $z = z_i$. The matrix equation then collapses to

$$\begin{pmatrix} u_i \\ u'_i \end{pmatrix} = \begin{pmatrix} \frac{z_{Fi} - z_i}{f_i} & 0 \\ -1/f_i & \frac{z_o - z_{Fo}}{f_i} \end{pmatrix} \begin{pmatrix} u_o \\ u'_o \end{pmatrix} = \begin{pmatrix} M & 0 \\ -1/f_i & \frac{f_o}{f_i M} \end{pmatrix} \begin{pmatrix} u_o \\ u'_o \end{pmatrix} \quad (6)$$

Here, M denotes the magnification, $g(z_i) = M, f_o$ and f_i are the object and image focal lengths, and z_{Fo} and z_{Fi} are the object and image foci. (For more details of paraxial optics, see Chapters 16 and 17 of Hawkes and Kasper, 1989.) In the case of purely magnetic round lenses, $f_o = f_i$ and we denote the focal length by f .

The next higher approximation includes terms of third order in u_o and u'_o . Systems with a straight optic axis have no second-order terms and the rotational symmetry about the optic axis restricts the permitted third-order terms to the following:

$$\begin{aligned} \frac{u_i - M u_o}{M} = & C u'_o (x_o'^2 + y_o'^2) && \text{(spherical aberration)} \\ & + 2(K + ik) u_o (x_o'^2 + y_o'^2) + (K - ik) u_o u_o'^2 && \text{(coma)} \\ & + (A + ia) u_o^2 u_o'^* && \text{(astigmatism)} \\ & + F(x_o^2 + y_o^2) u'_o && \text{(field curvature)} \\ & + (D + id) u_o (x_o^2 + y_o^2) && \text{(distortion)} \end{aligned} \quad (7)$$

The spherical aberration term is of particular concern since it does not vanish or even dwindle for object points close to or on the axis. This is the region that is imaged in high-resolution operation. We examine this aberration closely in a later section. The other terms are of decreasing interest for objective or probe-forming lenses and, moreover, the coma (next in importance after the spherical aberration) can be avoided. Lenses have a “coma-free plane,” the exact position of which is determined by the relative magnitudes of the spherical aberration and coma coefficients. In practice, for magnetic lenses, it falls within the lens field, upstream from the image focus (the “diffraction plane”). Specifically, let us suppose that the aberrations are expressed in terms of ray position in the object plane, $z = z_o$ and some aperture plane, $z = z_a$. The paraxial solutions appearing in the aberration integrals will then be $s(z)$ and $t(z)$, which satisfy the conditions

$$\begin{aligned} s(z_o) &= t(z_a) = 1 \\ s(z_a) &= t(z_o) = 0 \end{aligned} \quad (8)$$

If another aperture position is selected, $z = \bar{z}_a$, the aberration coefficients will have exactly the same structure but the appropriate paraxial solutions will now be $\bar{s}(z)$ and $\bar{t}(z)$,

$$\begin{aligned} \bar{s}(z_o) &= \bar{t}(\bar{z}_a) = 1 \\ \bar{s}(\bar{z}_a) &= \bar{t}(z_o) = 0 \end{aligned} \quad (9)$$

Since there can be only two linearly independent paraxial solutions, we must be able to write

$$\begin{aligned} \bar{s}(z) &= \alpha s(z) + \beta t(z) \\ \bar{t}(z) &= \gamma s(z) + \delta t(z) \end{aligned} \quad (10)$$

and obviously $\alpha = 1$ and $\gamma = 0$; for β and δ , we have

$$\begin{aligned} \beta &= -s(\bar{z}_a)/t(\bar{z}_a) \\ \delta &= 1/t(\bar{z}_a) \end{aligned} \quad (11)$$

It is then easy to show that the coma coefficient for $z = \bar{z}_a$, $K(\bar{z}_a)$, is related to that for $z = z_a$, $K(z_a)$, as follows:

$$K(\bar{z}_a) \propto K(z_a) + \beta C_s(z_a) \quad (12)$$

The coma-free point is thus situated at the point for which

$$\frac{s(\bar{z}_a)}{t(\bar{z}_a)} = \frac{K(z_a)}{C_s(z_a)} \quad (13)$$

The distortion can be important in projector lenses, in which the trajectories may be relatively far from the axis while the gradient will be very small at high magnification. Astigmatism and field curvature are mainly of importance in devices in which the size of the field is a major preoccupation, such as lithography.

For quadrupoles, the situation is slightly more complicated. A quadrupole usually consists of four electrodes or four magnetic poles, though many other configurations that create quadrupole fields exist (see, for example, Hawkes, 1970 or Baranova and Yavor, 1989). For convenience, we assume throughout that the quadrupole is disposed in such a way that the converging and diverging planes coincide with the x - z and y - z planes (Figure 10-1). The paraxial properties of the quadrupole are then characterized by equations of motion for $x(z)$ and $y(z)$ that are uncoupled but no longer identical. They are again linear, homogeneous, second-order differential equations and hence have solutions analogous to those for round lenses. Now, however, we have *two* sets of cardinal elements, one set for the x - z plane and a second set for the y - z plane. These equations take the form

$$\begin{aligned} \frac{d}{dz}(\hat{\phi}^{1/2}x') + \frac{\gamma\phi'' - 2\mathcal{P}_2 + 4\eta Q_2\hat{\phi}^{1/2}}{4\hat{\phi}^{1/2}}x &= 0 \\ \frac{d}{dz}(\hat{\phi}^{1/2}y') + \frac{\gamma\phi'' + 2\mathcal{P}_2 - 4\eta Q_2\hat{\phi}^{1/2}}{4\hat{\phi}^{1/2}}y &= 0 \end{aligned} \quad (14)$$

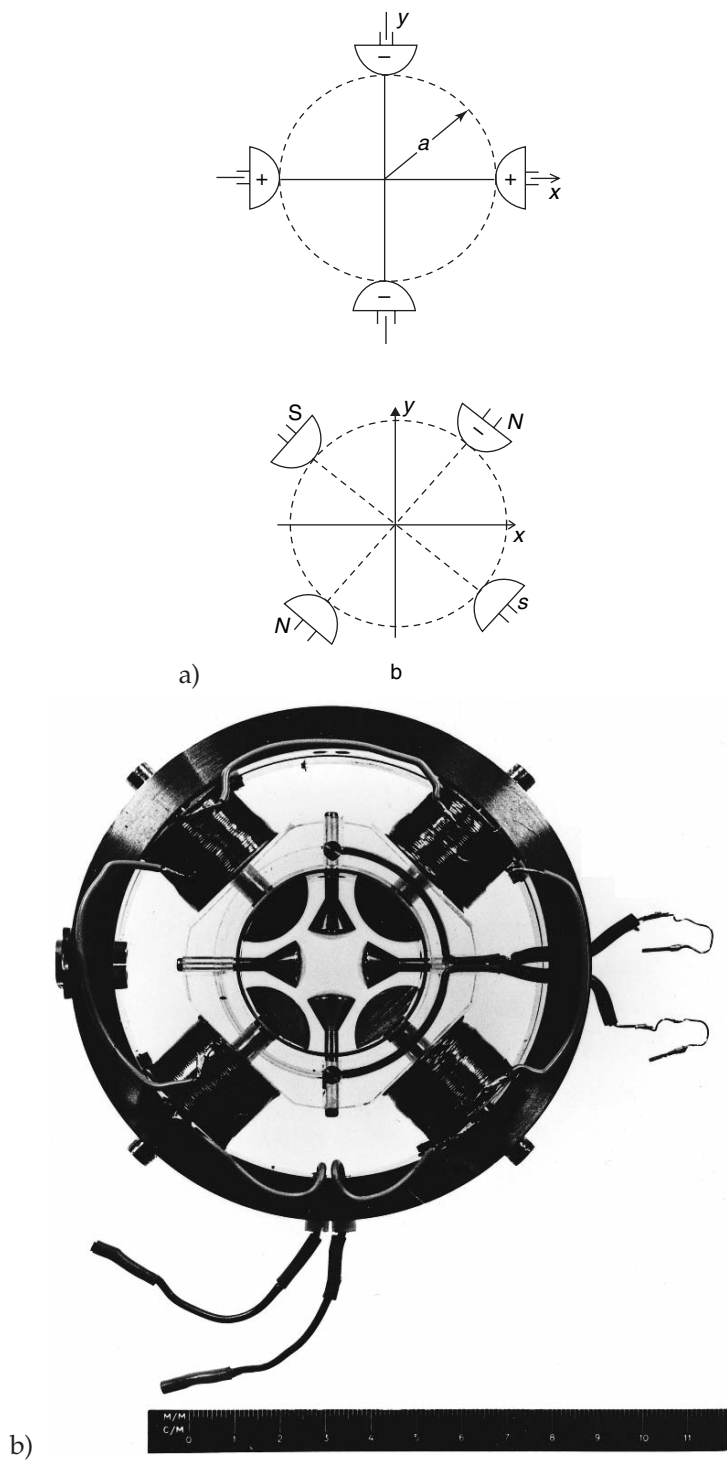
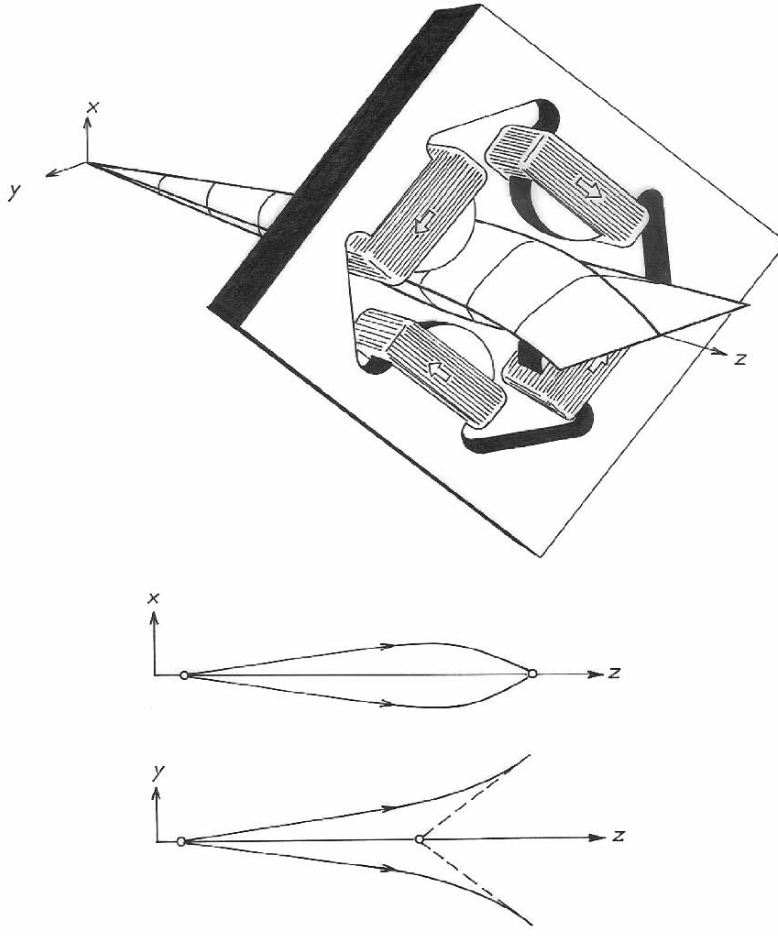


Figure 10-1. Quadrupoles. (a) Quadrupole orientation. The paraxial equations of motion are uncoupled when the x - z and y - z planes coincide with the planes of symmetry of electrostatic quadrupoles and fall midway between the poles of magnetic quadrupoles. (b) Appearance of a mixed magnetic-electrostatic quadrupole.



c)

Figure 10-1. *Continued* (c) Formation of a line image in a magnetic quadrupole; the arrows show the directions of the currents in the windings. (After Hawkes and Kasper, 1989, courtesy of Elsevier/Academic Press.)

in which the functions $p_2(z)$ and $Q_2(z)$ characterize the potential and field distributions (see Appendix I). We can now write

$$\begin{pmatrix} x(z_{ix}) \\ x'(z_{ix}) \end{pmatrix} = \begin{pmatrix} \frac{z_{Fi}^{(x)} - z_{ix}}{f_i^{(x)}} & 0 \\ -1/f_i^{(x)} & \frac{z_o - z_{Fo}^{(x)}}{f_i^{(x)}} \end{pmatrix} \begin{pmatrix} x_o \\ x'_o \end{pmatrix} = \begin{pmatrix} M_x & 0 \\ -1/f_i^{(x)} & \frac{f_o^{(x)}}{f_i^{(x)}} M_x \end{pmatrix} \begin{pmatrix} x_o \\ x'_o \end{pmatrix} \quad (15a)$$

and

$$\begin{pmatrix} y(z_{iy}) \\ y'(z_{iy}) \end{pmatrix} = \begin{pmatrix} \frac{z_{Fi}^{(y)} - z_{iy}}{f_i^{(y)}} & 0 \\ -1/f_i^{(y)} & \frac{z_o - z_{Fo}^{(y)}}{f_i^{(y)}} \end{pmatrix} \begin{pmatrix} y_o \\ y'_o \end{pmatrix} = \begin{pmatrix} M_y & 0 \\ -1/f_i^{(y)} & \frac{f_o^{(y)}}{f_i^{(y)}} M_y \end{pmatrix} \begin{pmatrix} y_o \\ y'_o \end{pmatrix} \quad (15b)$$

Since the cardinal elements are different in the two planes, the planes conjugate to a given object plane will not, in general, coincide. The system is astigmatic and if we operate two quadrupoles in tandem, the “object” of the second member will be astigmatic. Clearly, if we require quadrupoles to produce a stigmatic image of an object, we must use quadrupole multiplets and must somehow arrange that the cardinal elements of the multiplet are the same in the two planes. One multiplet configuration is of particular importance. It can be shown by symmetry arguments that if the multiplet is geometrically symmetric and electrically antisymmetric about its center plane, the focal lengths in the x - z and y - z planes will be equal. It is then only necessary to satisfy one condition (coincidence of the focal planes) to render the multiplet stigmatic. Quadruplets with this property have been extensively studied and are known as Russian Quadruplets, since their properties were first investigated by Dymnikov and Yavor (1963) in the Ioffe Institute in Saint Petersburg (Figure 10-2).

The aberrations of quadrupoles are more numerous than in the case of round lenses but they fall into the same categories: aperture aberrations (the analogues of the spherical aberration), comas, astigmatisms, and field curvatures and distortions. Here we consider only the aperture aberrations, since they will be exploited in aberration correctors. At this point, we note that octopoles, which have eight electrodes or magnetic poles, also have quadrupole symmetry and should hence be included in the formalism for the aberrations. They have no linear effect and hence have no effect on the paraxial behavior.

Our interest here is exclusively with aberration correctors, and we therefore assume that any multiplets of quadrupoles and octopoles are stigmatic and that the magnifications in the two planes are likewise equal ($M_x = M_y$). In these circumstances, the additional terms arising from the aperture aberrations take the following form:

$$\begin{aligned}\frac{\Delta x}{M} &= x'_o (C_x x_o'^2 + C_{xy} y_o'^2) \\ \frac{\Delta y}{M} &= y'_o (C_{xy} x_o'^2 + C_y y_o'^2)\end{aligned}\tag{16}$$

The other multipoles used for aberration correction are sextupoles (also known as hexapoles). Like octopoles, they have no linear

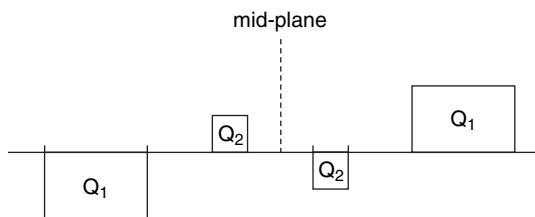


Figure 10-2. The “Russian” quadruplet. The excitations of the first and last quadrupoles are equal and opposite, as are those of the second and third quadrupoles. The quadrupoles are disposed symmetrically about the center plane.

effect on charged particles but unlike octopoles (and of course quadrupoles), their dominant effect is *quadratic*. Their primary aberrations are third order, like round lenses and quadrupoles, and their aperture aberration, the aberration that depends only on gradient and is independent of off-axis distance, has the same nature as the spherical aberration of round lenses. It is for this reason that sextupoles are potential correctors of C_s but some way of eliminating the quadratic effects must be devised. We shall see how this is achieved in a later section. For extensive discussion of their optical properties, see Rose (2002c).

Since the sextupoles will be used in conjunction with one or more round lenses, the paraxial solutions satisfy Eq. (1). We adopt a slightly different pair of solutions from those of Eq. (2), namely $h(x)$, which, as before, satisfies

$$h(z_o) = 0, h'(z_o) = 1 \quad (17a)$$

and $k(z)$,

$$k(z_o) = 1, k(z_k) = 0 \quad (17b)$$

where $z = z_k$ is the coma-free point discussed earlier. The expression for a general ray becomes

$$u(z) = \Omega_h h(z) + \Omega_k k(z) \quad (18)$$

in which Ω_h and Ω_k are simply related to the usual object coordinates (for ample details, see Rose, 2002, which is closely followed here). When this general ray passes through a sextupole lens, it will be deviated through a distance Δu , which has the form

$$\Delta u = \bar{\Omega}_h^2 u_{11} + \bar{\Omega}_h \bar{\Omega}_k u_{12} + \bar{\Omega}_k^2 \quad (19)$$

with

$$\begin{aligned} u_{11}(z) &= \frac{1}{2} \left(h(z) \int_{z_o}^z H h^2 k dz - k(z) \int_{z_o}^z H h^3 dz \right) \\ u_{12}(z) &= \left(h(z) \int_{z_o}^z H h k^2 dz - k(z) \int_{z_o}^z H h^2 k dz \right) \\ u_{22}(z) &= \frac{1}{2} \left(h(z) \int_{z_o}^z H k^3 dz - k(z) \int_{z_o}^z H h k^2 dz \right) \end{aligned} \quad (20)$$

The function $H(z)$ characterizes the field distribution in the sextupoles; in the most general case (in which both electrostatic and magnetic sextupoles may be present), H is given by

$$H(z) = -\frac{\exp(-3i\chi)}{\hat{\phi}_o^{1/2}} \left(\frac{(1 + \epsilon\phi)(p_3 + iq_3)}{\hat{\phi}_o^{1/2}} + i\eta(P_3 + iQ_3) \right) \quad (21)$$

where $\chi(z)$ characterizes the usual rotation in magnetic lenses.

All three contributions must vanish if the unwanted second-order effects of the sextupoles are to be eliminated; this can be achieved if the four integrals

$$\int_{z_0}^{z_e} H(z) h^{3-n} k^n dz, n = 0, 1, 2, \text{ and } 3 \quad (22)$$

all vanish. The form of these conditions, in two of which $h(z)$ is an even function while $k(z)$ may be odd, and in the other two the situation is reversed, indicates that symmetry can be used to eliminate all four quantities. The simplest arrangement is that shown in Figure 10–3. Here, one ray is symmetrical within each sextupole but antisymmetrical about the center plane of the combination; the other ray is antisymmetrical about the mid-plane of each sextupole but symmetrical about the center plane of the whole combination. The members of the round-lens doublet have equal and opposite excitations, chosen in such a way that the centers of the sextupoles are conjugates, with magnification -1 . Since the sextupole strength is a free parameter, it can, as we shall see in Section 3.2.2, be used to cancel the spherical aberration of an adjoining round lens.

2.2 Chromatic Aberrations

Chromatic aberrations are a consequence of the rapid variation of electron lens strength with electron energy and lens excitation. The electron beam from the gun will inevitably have some energy spread and there will be some variation in the lens excitations, however carefully they have been stabilized. The result is a “chromatic” effect, characterized by chromatic aberration coefficients, that blurs the image. We can include this in the paraxial formalism by writing

$$\frac{u_i - Mu_o}{M} = -\{C_c u'_o + (C_D + iC_\theta) u_o\} \left(\frac{\Delta \hat{\phi}_0}{\hat{\phi}_0} - 2 \frac{\Delta B_0}{B_0} \right) \quad (23)$$

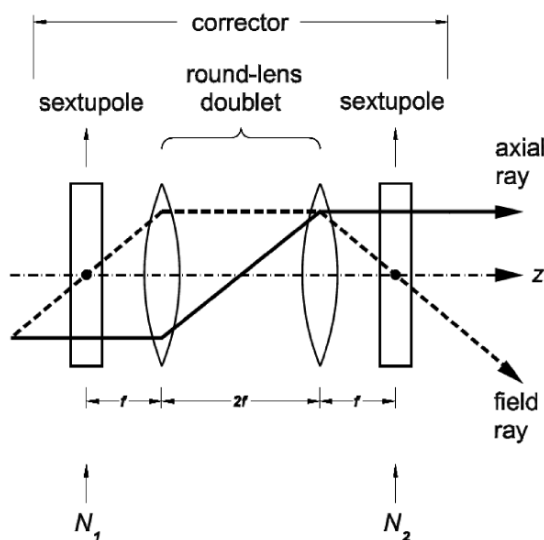


Figure 10–3. The simplest sextupole arrangement. No second-order aberrations are introduced outside the corrector. (After Rose, 2003c, courtesy of the author and Springer-Verlag.)

in which C_c is the (axial) chromatic aberration coefficient, C_D is a measure of the chromatic aberration of magnification, and C_θ is the anisotropic distortion. The quantity $\Delta\phi_0$ is a measure of the potential variation corresponding to the energy spread of the beam and ΔB_0 represents any fluctuations in the field strength of the lens caused by variations of the current. (Here we are considering only magnetic lenses; a similar reasoning applies to electrostatic lenses.)

We thus have two types of aberration: the chromatic aberration C_c , which is linear in gradient and hence does not vanish for object points close to the axis, just like the spherical aberration; and the complex chromatic aberration of magnification ($C_D + iC_\theta$), analogous to the distortion in that it is independent of gradient. We can therefore expect that the chromatic aberration, like the spherical aberration, will impose a limit on the resolution attainable in very high resolution work. In practice, it defines an “information limit.” which, in instruments that have not been corrected for spherical aberration, is usually less severe than the limit imposed by spherical aberration. With the arrival of spherical aberration correctors, however, the situation is reversed and it becomes imperative to correct the chromatic aberration as well, or find some way of rendering it innocuous. Incidentally, this is a return to the situation in the early days of electron microscopy, when chromatic effects dominated as a result of the relatively poor stabilization of circuitry at that time.

For quadrupoles, the chromatic aberrations are again different in the x - z and y - z planes:

$$\begin{aligned}\Delta x_i &= (C_{cx}x'_o + C_{Mx}x_o)\Delta_c \\ \Delta y_i &= (C_{cy}y'_o + C_{My}y_o)\Delta_c\end{aligned}\tag{24}$$

in which we have represented the potential and field variation by Δ_c [see Eq. (23)].

2.3 Parasitic Aberrations

Parasitic aberrations are, as their name suggests, not intrinsic defects of electron lenses. They result from imperfections of the construction or alignment of the lenses or of the entire instrument in question. The most serious is a (first-order) astigmatism, traditionally associated with imperfect circularity of the bore of the lens: in fact, Fourier analysis shows that this astigmatism will usually be the dominant parasitic effect, whatever the origin of the problem. It is as though a very weak quadrupole had been superimposed on the lens and the effect can be cancelled by adding a weak quadrupole with the opposite strength. Such a device, which in practice has a more complex structure, is known as a stigmator and several such correctors are routinely incorporated in commercial instruments. In high-resolution operation, other parasitic effects become noticeable and ways of correcting or compensating for these are now known.

A very detailed study of these parasitic aberrations in the context of electron microscope alignment has been made by Krivanek (Krivanek, 1994; see also Krivanek and Fan, 1992a,b; Krivanek and Leber, 1993,

1994; Krivanek and Stadelmann, 1995). For measurement techniques, see Saxton (1995a,b, 2000), Saxton et al. (1994), and Chand et al. (1995) as well as Ishizuka (1994). The work of Zemlin on alignment is also relevant here (Zemlin et al., 1978; Zemlin, 1979). Among the earlier literature on parasitic aberrations, we draw attention to the work of Glaser (1942), Sturrock (1951), Archard (1953), Glaser and Schiske (1953), Der-Shvarts (1954), and Stoyanov (1955). A long survey has been written by Yavor (1993) and other references are listed in Hawkes (1997). The many publications on the stigmator are also relevant; see Chapter 31 of Hawkes and Kasper (1989) for numerous references to these.

3 Aberration Correction

3.1 Introduction

The picture of a corrected electron microscope as a standard instrument to which a quadrupole–octopole or sextupole corrector has been added is an oversimplification. Before correction of the spherical aberration of the transmission electron microscope (TEM) and scanning transmission electron microscope (STEM) was achieved in the late 1990s, only sporadic attention was paid to any other aberrations that might become troublesome after C_s had been reduced. Now, however, largely due to the theoretical work of H. Rose and the practical designs of O. Krivanek and his group, strategies are being developed to correct all the potentially harmful aberrations in a systematic way. In this section, we first examine the basic correctors of spherical and chromatic aberration and then describe the much more complex correctors that should be capable of a broader range of correction. The latter are still at the planning stage or, at best, exist as prototypes.

Spherical aberration is the dominant resolution-limiting aberration in electron microscopes. It is characterized by the spherical aberration coefficient C_s , which can be expressed as an integral of the form

$$C_s = \int_{z_0}^{z_i} f[B(z), h(z)] dz \quad (25)$$

for magnetic lenses, in which as usual $B(z)$ denotes the magnetic flux on the axis and $h(z)$ is the particular solution of the paraxial ray equation (1) that vanishes at the object plane. A similar formula, in which the axial potential distribution $\phi(z)$ replaces $B(z)$, gives the spherical aberration coefficient of electrostatic lenses. Here we concentrate on magnetic lenses.

The integrand $f[B(z), h(z)]$ can be written in different ways [see Hawkes and Kasper, 1989 for many of these and for a general formula from which all the others can be generated; forms particularly useful for programming are given by Lencová and Lenc, 1990 (magnetic lenses) and 1994, 1977 (electrostatic lenses)]. In 1936, Otto Scherzer derived a nonnegative definite form of the integrand, a sum of squared terms, from which it is clear that the sign of the coefficient cannot

change. Scherzer's formula was nonrelativistic but Rose (1967/8, see corrigendum in Preikszas and Rose, 1995) has established a relativistic version, which essentially confirms Scherzer's conclusion. Efforts to find field or potential distributions for which the integrand vanishes (Glaser, 1940; Recknagel, 1941) failed, as they were sure to do given the form of the integrand found by Scherzer. (Attempts to find a loophole nevertheless continue; see Nomura, 2004 for a recent claim.) Tretner (1959) later established bounds on the coefficient.

However, all was not lost for the derivation of Scherzer's formula requires a certain number of conditions to be satisfied: the lens must possess rotational symmetry, the excitation must be static, only dioptric operation is permissible (excluding an electron mirror mode), and, in the case of electrostatic lenses, the potential distribution and its derivatives must be continuous. Object and image must both be real (not virtual). Scherzer showed (1947) that by relaxing one or the other of these conditions, a corrector could be devised.

Throughout the second half of the twentieth century, efforts were made to build and test the various types of corrector. These are described in detail in Septier (1966), Hawkes (1980), Hawkes and Kasper (1989, Chapter 41), and Hawkes (1997); some more recent information is included in Hawkes (2002). Departure from rotational symmetry has always seemed a promising approach and as early as the 1950s, Seeliger (1949, 1951) attempted to put Scherzer's suggestion for exploiting the idea into practice, as did Möllenstedt (1956). Archard (1954) showed that relatively simple configurations would create the desired field distribution and Burfoot (1953) found a three-electrode geometry that would in principle at least be aberration-free. Numerous further experiments were made over the years, including a reassuring proof-of-principle experiment by Deltrap (1964), but all attempts to correct a high-resolution microscope objective foundered essentially as a result of the complexity of the necessary system and of the intrinsic instability of the technique. A realistic corrector consists of four (or more) quadrupoles, probably with the "Russian" symmetry, and three octopoles. The asymmetry between the focusing in the x - z and y - z planes is not in itself beneficial, quite the contrary in fact. However, it creates a situation in which octopoles (which have no paraxial effect, we recall) can correct the combined spherical and aperture aberrations of a round lens and the quadrupoles. Even if the quadrupoles and octopoles are combined into single elements, the task of aligning all these individual components is clearly formidable. In addition, the aberrations introduced by the quadrupoles are themselves large, so we have a situation in which a well-designed round lens with a small spherical aberration coefficient is combined with a set of quadrupoles with large aperture aberrations and octopoles, whose role is to cancel the overall aperture aberrations of the combination. It is for this reason that we describe the situation as unstable.

For several decades, quadrupoles and octopoles were the only non-rotationally symmetric devices envisaged, although it was known that sextupoles suffer from a spherical aberration that would be suitable, in principle at least, for use as a corrector (Hawkes, 1965). However, the

three-fold symmetry and associated optical properties of such multipoles seemed to render them unsuitable for the purpose. But in 1979, Beck suggested that sextupoles could be combined in such a way that they would become attractive for correction, and this proposal was soon thoroughly studied (Crewe, 1980, 1982; Crewe and Kopf, 1980a,b; Rose, 1981; Ximen, 1983; Ximen and Crewe, 1985). In an interesting alternative approach, Dragt and Forest (1986) again showed the suitability of sextupoles as correctors.

As we shall see in the following sections, each type of multipole corrector is now in use. The quadrupole–octupole corrector is used in the STEM, a probe-forming instrument in which the very considerable difficulties of correcting a useful area or field of view do not arise. The sextupole arrangement is incorporated in TEMs. A valuable study of the performance of these correctors has been made by Haider et al. (2000).

Of the other ways of achieving correction, Scherzer had high hopes of the use of high frequency (Scherzer, 1946, 1947), hopes that were shared by Gabor (1950). The basic idea is easily understood. Electrons are pictured as reaching the lens from a region close to the optic axis at the object; those rays that are more steeply inclined will reach the outer region of the lens (here, an electric lens) slightly later than those leaving the object at the same time but remaining close to the axis. By illuminating the specimen with short pulses of electrons and reducing the excitation of the lens in such a way that the outer electrons encounter a weaker field, it should be possible to focus them all at the same place, thus eliminating the effect of spherical aberration, which would have focused the outer electrons too strongly if the lens strength had not been weakened. For electrons with energies in the tens of kilovolts range, the frequency required is found to be of the order of gigahertz and some experimental studies with a microwave cavity inserted between the polepieces of a magnetic lens were made by Oldfield (1973, 1974). These were not pursued, however, and the problems of generating short enough pulses and, above all, of dealing with the increased energy spread of the beam were clearly formidable. Recently, Calvo has reconsidered the technique along very different lines, but his findings are at present (2005) inconclusive (Calvo, 2002, 2004; Calvo and Lazcano, 2002; Calvo and Laroze, 2002). Yet another way of using dynamic fields for spherical and chromatic aberration correction has been proposed by Schönhense and Spiecker (2002, 2003), with photoemission and low-energy electron microscopes in mind. For chromatic aberration, an ingenious way of inverting the energy distribution of the electron beam has been found. The electrons are generated at the sample by a pulsed beam and, after acceleration and collimation, they enter a drift space, in which the faster electrons draw away from the slower ones, like horses galloping down the straight in a race. Beyond the drift space is an accelerator, initially switched off. When the fastest electrons emerge from it, the electric accelerating field is rapidly switched on and by suitable choice of the accelerator field, the slower electrons can be accelerated to a higher energy than the fast electrons, thus effectively inverting the original energy distribution (Figure 10–4). The chromatic

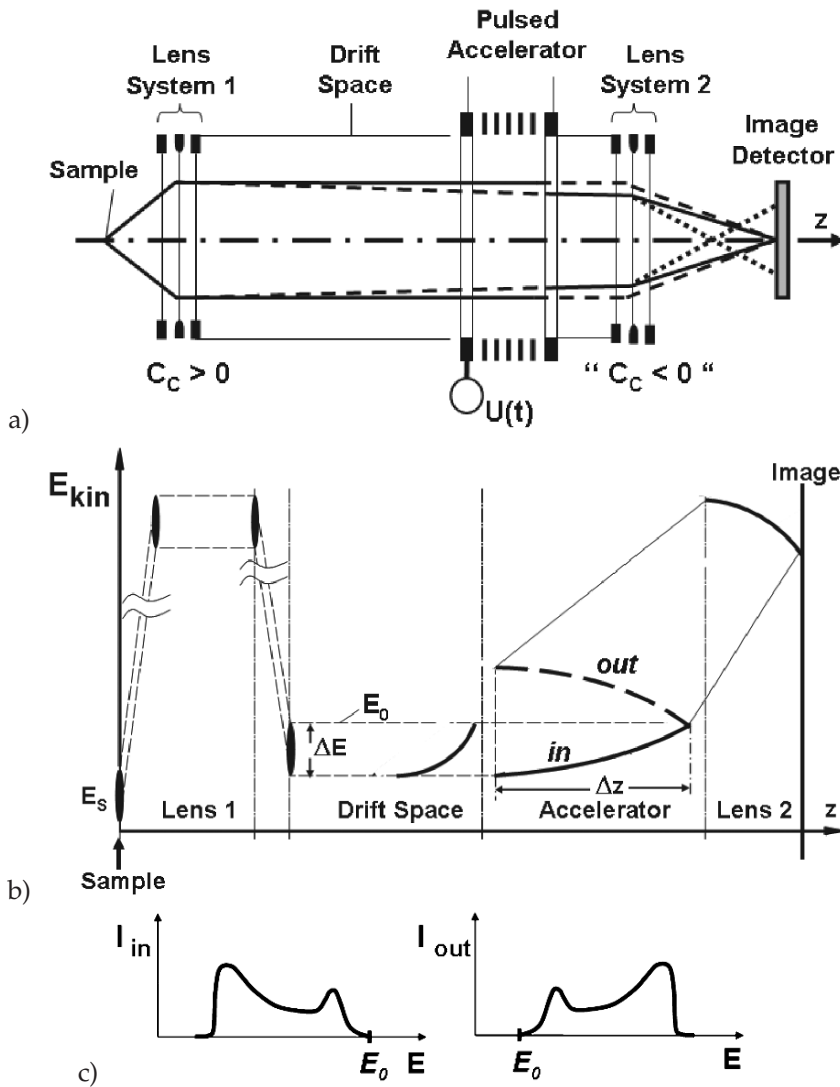


Figure 10-4. Chromatic correction based on inversion of the electron energy distribution. (a) Schematic cross-section of the system (greatly exaggerated in the radial direction). (b) Electron energy distribution as a function of the optical path; E_{kin} denotes the actual kinetic energy of the electron ensemble. The distribution E_{kin} as a function of z before and after passing the pulsed accelerating field is denoted by "in" and "out," respectively. (c) Schematic representation of the electron energy distribution before and after passing the accelerator, I_{in} and I_{out} . (After Schönhense et al., 2002, courtesy of the authors and the American Institute of Physics.)

aberration of the final lens will then bring the whole beam to a smaller focus than in the absence of energy inversion. For spherical aberration correction, Schönhense and Spiecker consider an "electrostatic" lens in which the lens strength is abruptly altered when the (pulsed) beam reaches the center. Schönhense and Spiecker explain the correcting effect by regarding the half-lens as a diverging lens (the converging

effect of the second half of the lens having been suppressed by switching off the lens). The image of the object is then virtual and Scherzer's result is no longer applicable. Another way of understanding this is to note that the abrupt change in lens strength is equivalent to a discontinuity in the field and, as for the foil correctors mentioned below, this is known to offer a way of achieving correction. See also Khursheed (2002, 2005).

The idea of combining an electron mirror (an electrostatic lens in which there is a potential barrier too high for the electrons to surmount) with a round lens in such a way that spherical and chromatic aberrations of the mirror cancel those of the round lens has been explored and projects based on this are currently being actively pursued. A major problem is to separate the beam incident on the mirror from the beam emerging from it and numerous ways of achieving this have been proposed (see Septier, 1966 and Hawkes, 1980 for earlier suggestions). We return to this in Section 3.3.4.

Finally, we draw attention to the method in which a field discontinuity is introduced into an electrostatic lens, by covering the bore with a very thin electron-transparent foil. This was suggested as a means of achieving chromatic correction by Scherzer in 1948 and has subsequently been investigated in great detail and with considerable ingenuity, notably by Typke (1968/9, 1972, 1976) and Scherzer (1980) in Darmstadt, Hoch et al. (1976) in Tübingen, Munro and Wittels (1977), and in Japan (Hanai et al., 1986, 1994, 1995, 1998); for recent work, see Matsuda et al. (2005). The idea has been revived but in a rather different form by van Aken et al. in Delft (van Aken et al., 2002a,b, 2004, 2005; van Aken, 2005). The use of a cloud of space charge also falls into this category (see, for example, Haufe, 1958, where earlier attempts are listed) and has recently been revived by Orloff (Wang et al., 1995; Tang et al., 1996a,b; Chao et al., 1997) for ion-optical systems.

Chromatic aberration correction also has a long history, though less effort has been devoted to it than to spherical aberration correction because, in the high-resolution imaging mode, developments in microscope design soon rendered the adverse effect of spherical aberration greater than that of chromatic aberration. This is obvious from an examination of the phase-contrast transfer function, which is a sinusoidal curve in the absence of any energy spread (chromatic effects or temporal partial coherence) and neglecting the nonvanishing source size. The sinusoidal curve is damped by an envelope function, representing the effect of energy spread, but the first zero of the sinusoidal curve (a measure of the limit of resolution determined by the spherical aberration) occurs well before the damping curve reduces it to an unacceptably small value (the so-called information limit). With the arrival of C_s -correctors, however, the situation has changed dramatically and it is now of interest, indeed essential to improve the information limit as well by reducing the undesirable effects of energy spread.

The (axial) chromatic aberration of electron lenses is characterized by a coefficient, C_c , which can be written as an integral of the form

$$C_c = \int_{z_0}^{z_i} \frac{\eta^2 B^2}{4 \hat{\phi}_0} h^2 dz \quad (26a)$$

for magnetic lenses and

$$C_c = \hat{\phi}_0^{1/2} \int \frac{\gamma \phi'^2 (3 + 2\epsilon\phi)}{8 \hat{\phi}^{5/2}} h^2 dz \quad (26b)$$

for electrostatic lenses. As in the case of C_s (and, indeed, of all the aberration coefficients), the integrands in Eqs. (26a) and (26b) can be written in different ways. The ones given here show immediately that C_c is positive definite and the best that can be hoped for in a round lens is a design for which C_c is small.

Over the years, two approaches to the problem of avoiding the limitations imposed by the chromatic aberration of round lenses have emerged. One is a natural continuation of the efforts to reduce the energy spread of the beam emitted by the electron gun; by introducing a monochromator into the column, electrons with energies outside the chosen range can be eliminated. There is of course some loss of beam current but since the energy spread of the filtered beam can be made appreciably narrower than that of the original beam, this reduction in current may be acceptable. This solution is attractive not only in the imaging mode but also for electron energy-loss spectroscopy (EELS) and energy-filtered transmission electron microscopy (EFTEM). We say no more about this here; extensive discussion is to be found in the following books, review articles and other publications: Plies (1978), Tang (1986), Tsuno et al. (1988–9, 1990, 2003, 2005), Tsuno (1991, 1992, 1993, 1999), Reimer (1995), Tsuno and Rouse (1996), Kahl and Rose (1998, 2000), Huber and Plies (1999, 2000), Mook and Kruit (1999a,b, 2000), Mook et al. (1999, 2000), Oshima (1999), Batson et al. (2000), Kahl and Voelkl (2001), Martínez and Tsuno (2002, 2004), Benner et al. (2003c), Mukai et al. (2003a, b), Plies and Bärtele (2003), Ioanoviciu et al. (2004), Huber et al. (2004), Freitag et al. (2005), and Browning et al. (2006).

The alternative to using a monochromator is to devise a corrector of chromatic aberration. There are several ways in which such correctors can be conceived, involving the use of superimposed round lens and quadrupole fields, mixed electrostatic–magnetic quadrupoles, or electron mirrors.

An early suggestion by Scherzer (1947) involved combining an electrostatic round lens and an electrostatic quadrupole in such a way that the overall chromatic aberration coefficient of the combination is negative. Such a device could then be used to correct the chromatic aberration of a round lens acting as an objective. This suggestion was taken up by Archard (1955) and has subsequently been investigated carefully.

In 1961, Kel'man and Yavor showed that the chromatic aberration coefficient of a combined electrostatic–magnetic quadrupole can have either sign, depending on the relative strengths of the component quadrupoles, and hence that such combined lenses could be used to correct chromatic aberration. The result was rediscovered by Septier (1963) and generalized by Hawkes (1964, 1965b). Such combined

quadrupoles appear in the latest designs of correctors, intended for the correction of both geometric and chromatic aberrations.

Another possibility, pointed out by Rose (1990), involves the use of a long Wien filter. This too has been studied carefully.

Finally, an electron mirror can be coupled to a round lens in such a way that the chromatic aberration is cancelled. A practical configuration was examined by Rempfer and Mauck (1985, 1986, 1992) and Skoczylas et al. (1991, 1994). For other proposals, see Shao and Wu (1989, 1990a,b), Rempfer (1990a,b), Crewe et al. (1995a,b, 2000), Tsai (2000), Crewe (2002), and Bimurzaev (2003, 2004). The simple beam-splitter employed by Rempfer and Mauck and later by Rempfer et al. (1997) to separate the incident and emergent beams was dispersive. An improved version, in which the separator is nondispersive, has been studied by Preikszas and Rose (1997) and is incorporated in the optical system of the European SMART (Fink et al., 1997; Wichtendahl et al., 1998; Müller et al., 1999; Hartel et al., 2000, 2002; Preikszas et al., 2000) and of the American PEEM3 (Wan et al., 2004; Wu et al., 2004; Feng et al., 2005; Schmid et al., 2005) projects.

3.2 Spherical Aberration Correctors

3.2.1 *Quadrupole–Octopole Correctors*

For the correction of spherical aberration, Scherzer proposed a sequence of cylindrical lenses and octopoles in his seminal paper of 1947 on ways of avoiding the consequences of his 1936 proof. Cylindrical lenses are the electron optical counterparts of glass lenses with cylindrical (as opposed to spherical) faces and are characterized by a round lens and a quadrupole potential distribution. It was soon realized (Archard, 1954) that quadrupole lenses could be used to advantage instead of cylindrical lenses and the basic corrector configuration, which has remained essentially unaltered, soon emerged (Figure 10–5): a sequence of three or four quadrupole lenses, with an octopole situated at each of the line foci to cancel or overcorrect the aperture aberrations in the x - z and y - z planes, together with a third octopole to complete the task of correction. (Turnbull, 2004, has recently revived the idea of using cylindrical lenses in a combined chromatic and spherical aberration

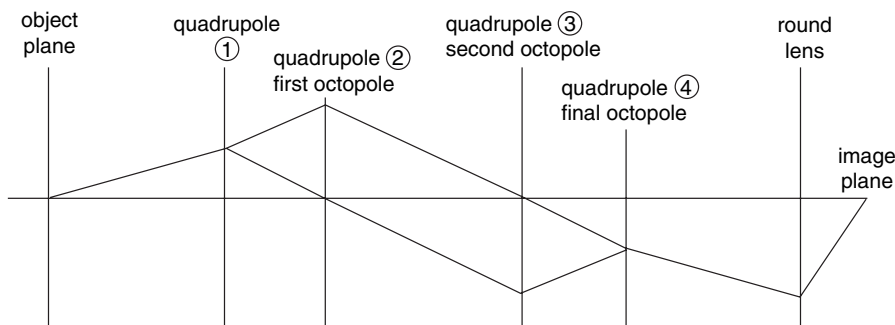


Figure 10–5. The basic quadrupole corrector arrangement, showing four quadrupoles with octopoles situated at the two line foci and at a third site.

corrector.) An important step forward was the introduction of the Russian quadruplet, which has geometric symmetry and electrical antisymmetry about its mid-plane, as we have already mentioned. In common with all multiplets possessing these symmetry properties, such quadruplets have the same focal length in the x - z and y - z planes. For a given geometry, the positions of the foci in these planes can then be made to coincide by suitable choice of the two excitations, whereupon the quadruplet has the same overall paraxial behavior as a round lens. Sets of load curves, showing the appropriate excitations as a function of geometry, are available (see Hawkes, 1970, for many such curves and Baranova and Yavor, 1989). Another interesting early contribution was made by Burfoot (1953), who sought the (electrostatic) configuration with the smallest number of electrodes that would be free of spherical aberration. He established suitably shaped apertures in a three-electrode lens (a remarkable achievement in precomputer times) but concluded that the necessary tolerances could not be achieved in practice; a simpler way of attaining the same objective was proposed by Archard (1958).

In 1964, Deltrap showed that the spherical aberration of a test lens could be reduced by means of a quadrupole–octopole corrector and thus confirmed that the principle of correction was sound. However, for the next three decades, all attempts to make a corrector capable of improving the performance of a well-designed objective failed; with hindsight, we can see that these repeated and disappointing failures were due to the natural complexity of the system and to the unstable character of the correction principle mentioned above. Considerable progress was made, notably in the Darmstadt project (Reichenbach and Rose, 1968/9; Rose, 1970, 1971a,b; Bastian et al., 1971; Pöhner 1976, 1977; Bernhard and Koops, 1977; Koops et al., 1977; Koops, 1978, 1978/9; Koops and Bernhard, 1978; Pejas, 1978; Kuck, 1979; Bernhard, 1980; and Fey, 1980; see Scherzer, 1978 for a summing up) and in the Chicago microscope (Crewe et al., 1968; Thomson, 1972; Beck and Crewe, 1974, 1976; Beck, 1977; Crewe, 1978), but neither succeeded in showing any real gain in electron microscope resolution. The tools necessary for the adjustment of such devices were not yet available. In the early 1990s, Zach showed that such correctors could improve the performance of scanning electron microscopes and this finding continues to be exploited in commercial instruments (Zach, 1989, 2000; Zach and Haider, 1995a,b; Honda et al., 2004a,b; Kazumori et al., 2004a,b; Uno et al., 2004a,b). Success came at last in 1997, when Krivanek and colleagues, working in the Cavendish Laboratory in Cambridge, built a corrector equipped with computer control, capable of making the many necessary adjustments rapidly and systematically. This corrector was fitted to a STEM and was hence required to reduce the size of an electron probe (or to allow the angular aperture and hence the probe current to be increased). The Krivanek corrector consists of the basic quadrupoles and octopoles, all under computer control, together with other multipole fields designed to compensate for misalignments and parasitic aberrations in general. In the second-generation Nion corrector, 16 quadrupoles are used together with three combined

quadrupole–octopole elements. An additional quadrupole triplet is situated between the corrector and the probe-forming lens. The corrector itself consists of an alternating sequence of quadrupole quadruplets and quadrupole–octopoles (Figure 10–6a). With this arrangement, the center planes of the quadrupole–octopole elements are all

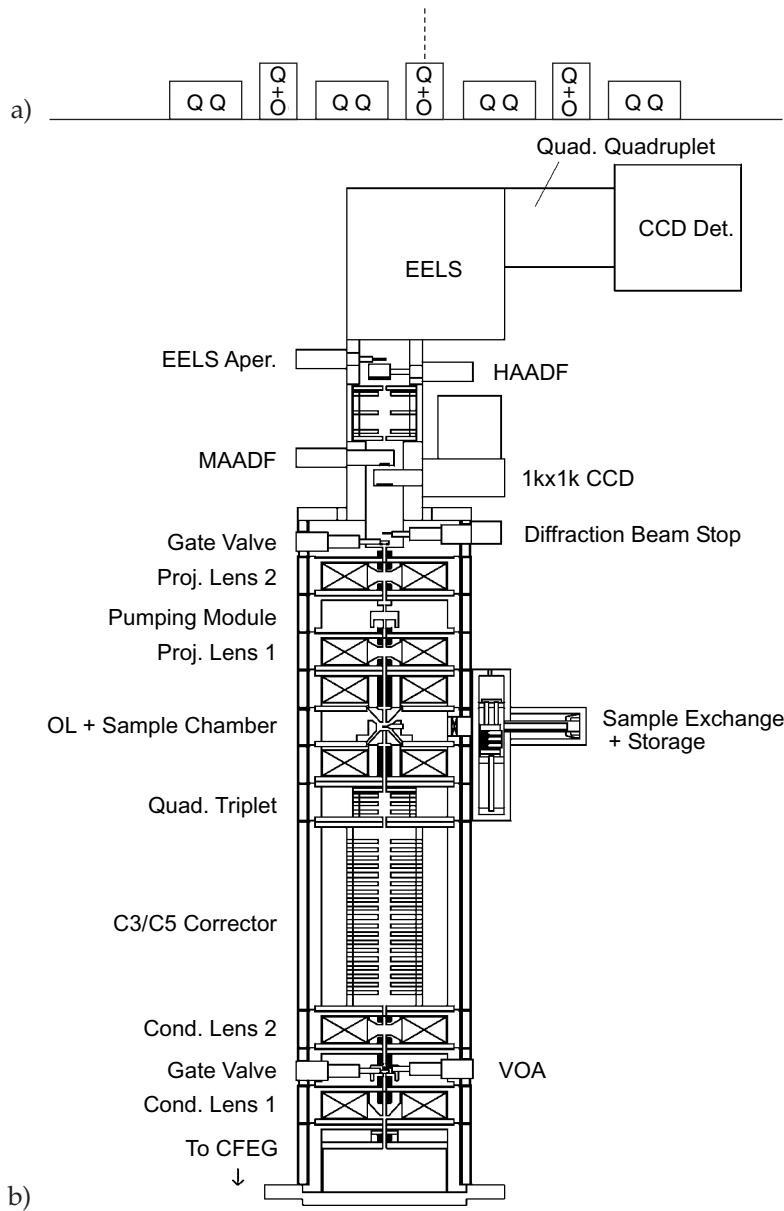


Figure 10–6. The Nion quadrupole–octopole corrector incorporated in scanning transmission electron microscopes. (a) The sequence of quadrupole quadruplets and quadrupole–octopoles of which the corrector is composed. Q-Q, quadrupole quadruplet; Q-O, quadrupole–octopole element. (b) Schematic view of the corrector incorporated in an STEM. (After Krivanek et al., 2004, courtesy of the authors and the Belgian Microscopy Society.)

conjugates and are also conjugate to a plane close to the coma-free plane of the probe-forming lens. In this way, the fifth-order geometric aberrations of the combination of corrector and probe-forming lens can be eliminated. Software will adjust the various components systematically. The evolution of the corrector, which has been fitted to many VG STEMs, to the two SuperSTEMs at Daresbury, and to the Nion STEM, can be studied in the following publications: Krivanek et al. (1997a,b, 1998, 1999a,b, 2000, 2001, 2002, 2003, 2004), Dellby et al. (2000, 2001), Batson et al. (2002), Batson (2003), Lupini et al. (2003), Pennycook et al. (2003), Nellist et al. (2004a,b), Dellby et al. (2005), Bacon et al. (2005), Krivanek et al. (2005), and Nellist et al. (2006). Figure 10–6b shows schematically the Nion corrector incorporated in a STEM. The quadrupole–octopole corrector designed for an FEI STEM/TEM is described by Mentink et al. (2004).

3.2.2 Sextupole Correctors

Sextupoles were not among the correctors envisaged by Scherzer in his 1947 paper. In 1965, it was pointed out that the third-order aberrations, including of course the spherical aberration, of sextupoles have the same dependence on gradient in the object plane as that of a round lens (Hawkes, 1965a). However, the fact that the lowest order optical effect of sextupoles is not linear, as it is in round lenses and quadrupoles, but quadratic seemed to rule out any hope of using them for aberration correction. It was not until 1979 that combinations of sextupoles and round lenses from which the quadratic effects had been eliminated by compensation were proposed and subsequent developments have confirmed that such correctors are suitable for incorporation into transmission electron microscopes. As we have seen, the second-order effect of a sextupole is characterized by four terms of the form $\int H(z)h^{3-n}k^n dz$, in which $H(z)$ represents the field distribution in the (electrostatic or magnetic) sextupole and $h(z)$, $k(z)$ are two linearly independent solutions of the familiar paraxial equation for round lenses (these solutions collapse to straight lines in the absence of any round lens component). The integer n takes the four values 0, 1, 2, and 3. All four terms can be made to vanish by suitable choice of the symmetry of the configuration; the simplest is shown in Figure 10–7. Before coupling such a device to a microscope objective, we must, however, ensure that the coma-free condition is satisfied. The (isotropic) coma-free plane of an objective is situated within the lens field and must hence be imaged onto the front focal plane of the round-lens doublet in the corrector by means of another doublet (Figure 10–8). If it should be necessary to eliminate the anisotropic coma as well as the isotropic coma, an objective design in which two coils are used in tandem would have to be adopted (Rose, 1971b). Sextupole correction may be traced in the following articles (in addition to the early publications already cited): Haider et al. (1982, 1995, 1998a–c), Rose (1990b, 2002a,b), Haider and Uhlemann (1997), Haider (1998, 2000, 2003), Foschepoth and Kohl (1998), Uhlemann et al. (1998), Urban et al. (1999), Müller et al. (2002), Kabius et al. (2002), Lentzen et al. (2002), Liu et al. (2002), Benner et al. (2003a,b, 2004a,b), Chang et al. (2003), Hosokawa and Sawada (2003), Hosokawa et al. (2003), Jia et al.

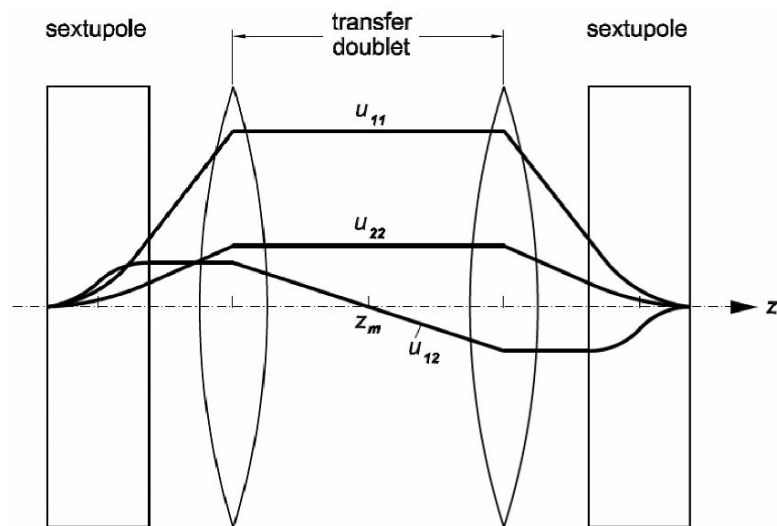


Figure 10–7. Sextupole correctors. Course of the second-order fundamental rays in the corrector shown in Figure 10–3. (After Rose, 2003c, courtesy of the author and Springer-Verlag.)

(2003), Sawada et al. (2004a–c), Haider et al. (2004), Hartel et al. (2004), Titchmarsh et al. (2004), Hutchison et al. (2005), Haider and Müller (2005), Müller et al. (2005), and Chang et al. (2006).

3.2.3 Foil Correctors

The suggestion that spherical aberration can be corrected by incorporating a foil in a rotationally symmetric lens, thus introducing a

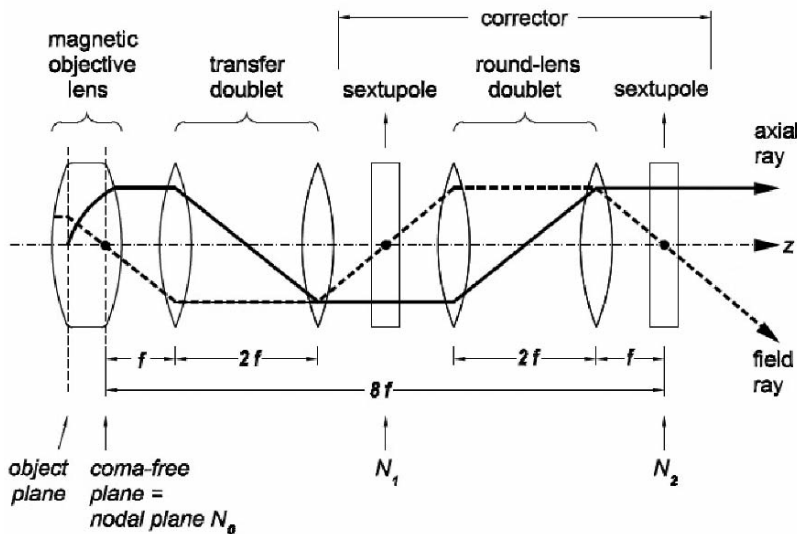


Figure 10–8. Sextupole correctors. Correction of coma as well as spherical aberration requires a more complex system. The transfer doublet between the objective lens and the corrector allows for the fact that the coma-free plane lies within the magnetic field of the objective. (After Rose, 2003c, courtesy of the author and Springer-Verlag.)

discontinuity in the electric field, has been very thoroughly explored by a Japanese group (see Hanai et al., 1986, 1994, 1995, 1998). In the earlier studies, the lens shown in Figure 10–9 was fitted in a TEM and was shown to reduce the spherical aberration. Subsequent work has extended this to probe-forming instruments, notably the STEM, where again the principle was shown to be valid. The lifetime of the foil was, however, rather short as a result of the build-up of contamination. For a recent work using a curved mesh, see Matsuda et al. (2005).

A very different type of foil corrector has been investigated by van Aken (van Aken et al., 2002a,b, 2004, 2005; van Aken, 2005), who attempted to exploit the marked increase in the mean free path of electrons in metal foils at very low voltages. If the latter is atomically flat, a mean free path of the order of 5 nm is expected for several metals

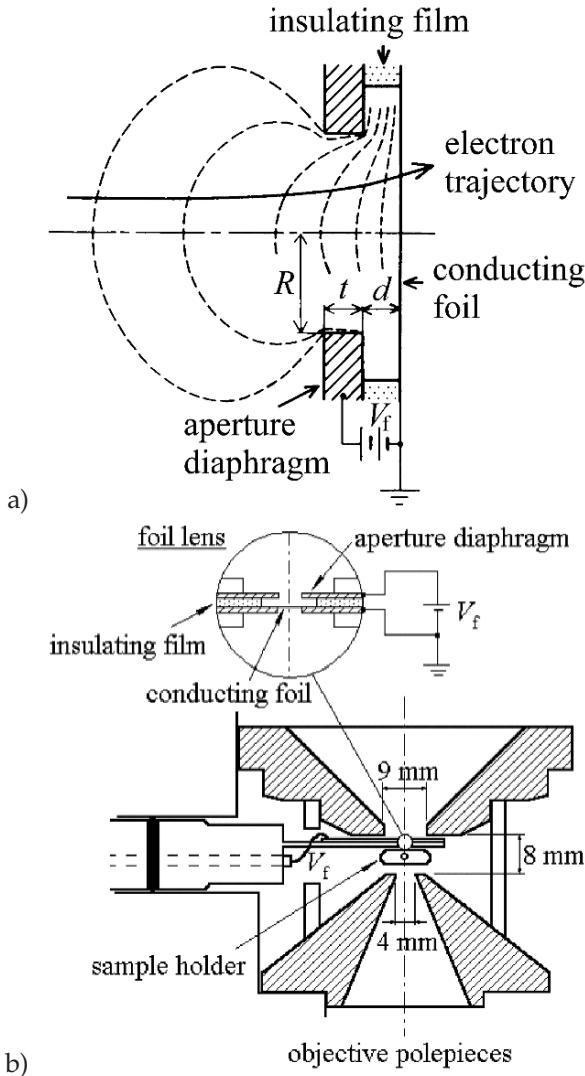


Figure 10–9. Foil lens. (a) Schematic cross section of the foil lens. (b) A practical design, showing the foil lens incorporated in a magnetic objective. (After Hanai et al., 1998, courtesy of the authors and Oxford University Press.)

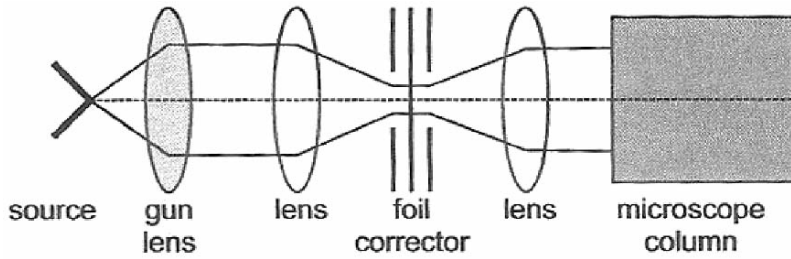


Figure 10–10. The van Aken low-energy foil corrector. Electrons are slowed down to a very low energy before passing through the correcting foil and then accelerated again before entering the microscope column. (After van Aken, 2005, courtesy of the author.)

at 5 eV above the Fermi level and a transmission of about 10% is predicted for thin foils (~ 5 nm). By sandwiching the foil between a pair of electrodes, which first retard the incident electrons and then accelerate them again, correction of both chromatic and spherical aberration is in principle possible. Figure 10–10 shows schematically a microscope column incorporating such a corrector.

3.3 Chromatic Aberration Correctors

3.3.1 All-Electrostatic Correctors

One form of the chromatic aberration coefficients of a system containing electrostatic round lens fields and electrostatic quadrupole fields is

$$\begin{aligned} C_{cx} &= \sqrt{\phi_0} \int \frac{1}{\phi^{1/2}} \left(h_x'^2 - \frac{\phi'}{2\phi} h_x h_x' \right) dz \\ C_{cy} &= \sqrt{\phi_0} \int \frac{1}{\phi^{1/2}} \left(h_y'^2 - \frac{\phi'}{2\phi} h_y h_y' \right) dz \end{aligned} \quad (27)$$

By introducing the Picht transformation,

$$H_x = \left(\frac{\phi}{\phi_0} \right)^{1/4} h_x, \quad H_y = \left(\frac{\phi}{\phi_0} \right)^{1/4} h_y \quad (28)$$

these take the form

$$\begin{aligned} C_{cx} &= \phi_0 \int \frac{1}{\phi} \left[\left(H_x' - \frac{\phi'}{2\phi} H_x \right)^2 - \frac{1}{16} \frac{\phi'^2}{\phi^2} H_x^2 \right] dz \\ C_{cy} &= \phi_0 \int \frac{1}{\phi} \left[\left(H_y' - \frac{\phi'}{2\phi} H_y \right)^2 - \frac{1}{16} \frac{\phi'^2}{\phi^2} H_y^2 \right] dz \end{aligned} \quad (29)$$

If the quadratic term could be reduced to zero (or at least made smaller than the absolute value of the negative term), the overall chromatic aberration coefficient would be negative and the device would hence be suitable for use as a corrector. The quadratic term vanishes if the field distribution is such that the paraxial ray $H_x(z) = C\phi^{1/2}$ and similarly for $H_y(z)$, where C is an arbitrary constant. In general, this is not an acceptable form for $H_x(z)$ and $H_y(z)$, which must vanish in the object (and

image) planes. If, however, the corrector is telescopic, such a form is permissible. By substituting this expression for $H_x(z)$ and $H_y(z)$, and hence $h_x(z)$ and $h_y(z)$, back into the paraxial equation, it is trivial to show that the field functions $\phi(z)$ and $p_2(z)$ must be related by the formula

$$p_2 = \phi'' - \frac{\phi'^2}{8\phi} \quad (30)$$

Configurations in which this condition (Scherzer, 1947) is closely satisfied have been found by Weißbäcker and Rose (2000, 2001, 2002) and by Maas and co-workers (Maas et al., 2000, 2001, 2003; Henstra and Krijn, 2000). In the studies of Weißbäcker and Rose, several configurations are examined, in which the complexity increases with the practical usefulness of the corrector. In the simplest design, four distinct elements are employed and the corrector is capable of correcting both the chromatic and the spherical aberration in a scanning instrument. The first element is a quadrupole, the purpose of which is to render the incoming beam astigmatic. This is followed by a three-element corrector consisting of quadrupole fields superimposed on a (round) einzel lens field. A second corrector unit cancels the chromatic aberration in the other plane. Octopoles are incorporated so that the spherical aberration can be corrected simultaneously.

Unfortunately, such a corrector introduces large off-axis aberrations and is hence not suitable for use with transmission (fixed-beam) electron microscopes. Weißbäcker and Rose therefore consider first an extended version in which a third such element is added (Figure 10–11) and then propose a doubly-symmetric electrostatic corrector (DECO), in which each correcting element is enclosed within two quadrupole doublets (Figure 10–12). The symmetry conditions can now be arranged in such a way that the chromatic aberration and the coma vanish, while the spherical aberration is corrected by means of octopoles as usual.

In the complementary investigations of Henstra, Maas, Mentink, and co-workers, a configuration consisting of nine elements (Figure 10–13) is explored. At the outer extremities are quadrupoles to create astigmatism and subsequently annihilate it. In the center are five combined round lens and quadrupole units. Two other quadrupoles are included to match the slowly decaying round lens field of the first and last combined elements. For further work on this approach, see Baranova et al. (2004).

3.3.2 Mixed Quadrupole Correctors

Quadrupole lenses consisting of four electrodes and four magnetic poles situated midway between the electrodes have the power of correcting the chromatic aberration of a round lens. They must of course be part of a suitable configuration and are currently incorporated in the complex superaplanator and ultracorrector described in Section 3.4. Such correctors have not been seriously considered for correction of C_c in the transmission electron microscope in recent years, for the simpler configurations increased the spherical (and other) aberrations unreasonably. However, they have been reconsidered recently (Haider and

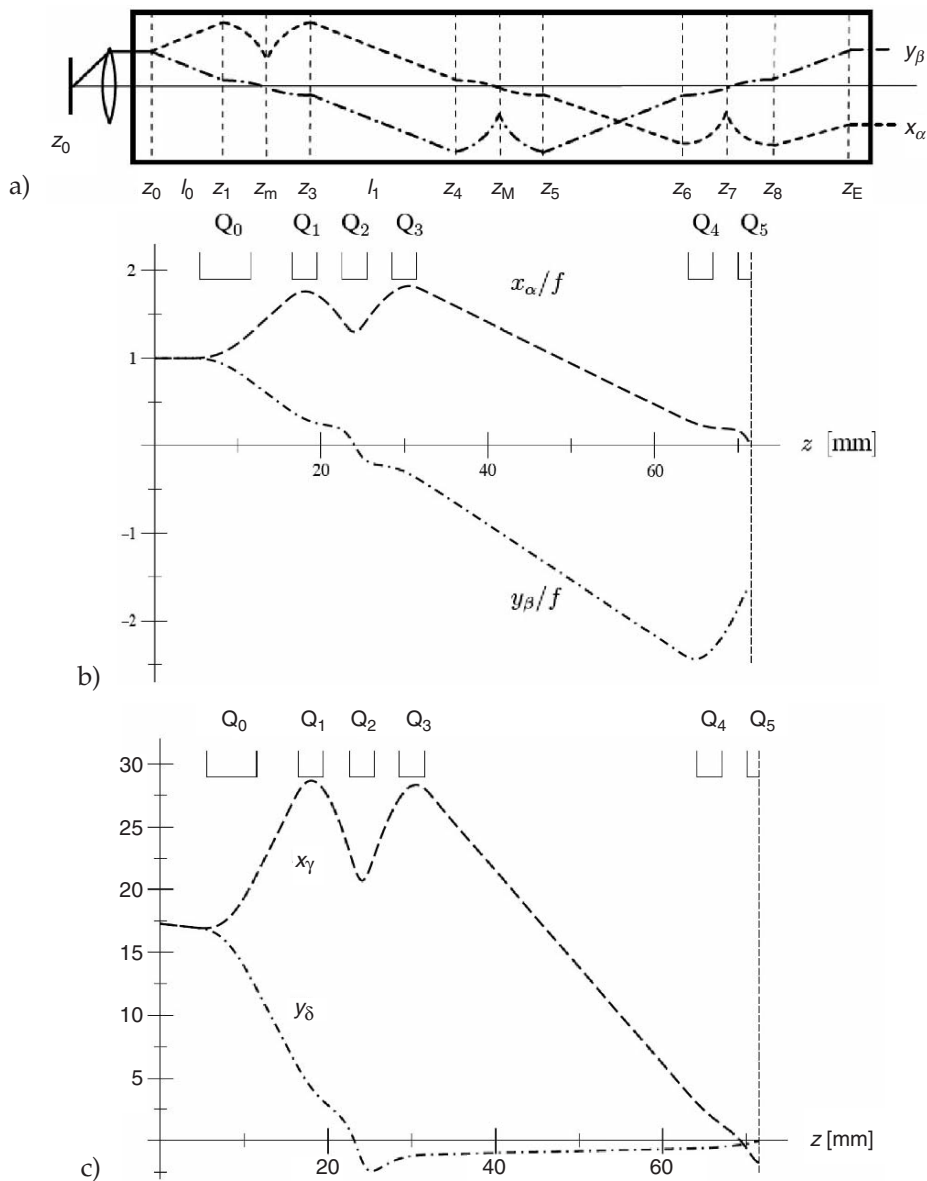
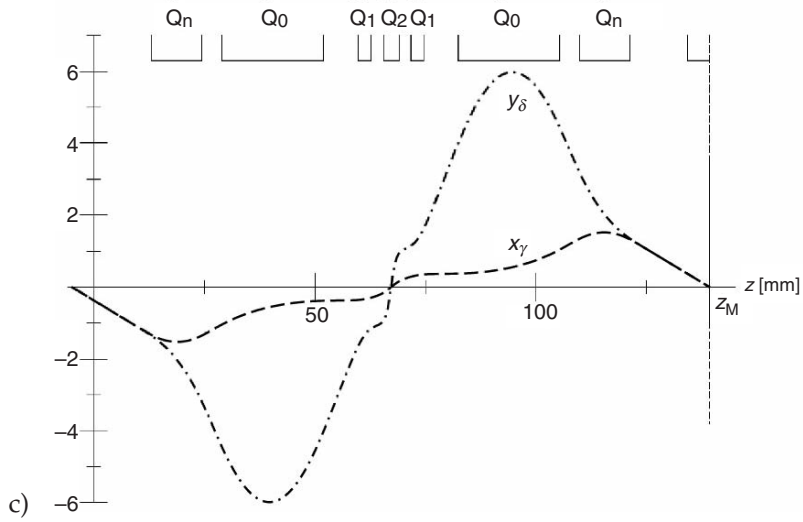
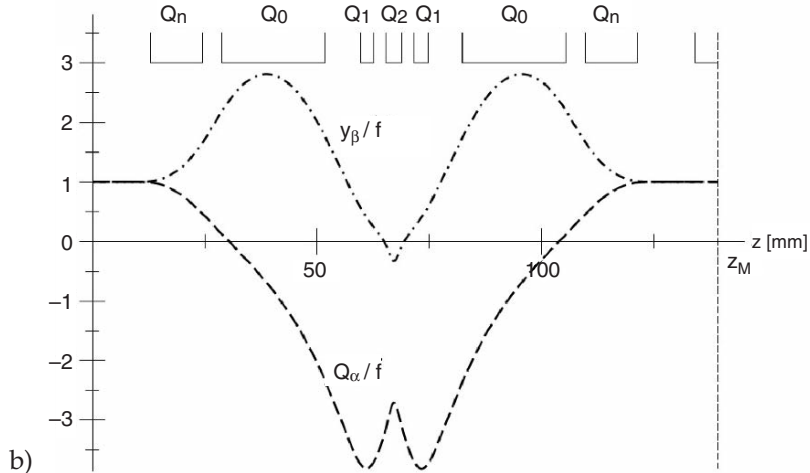
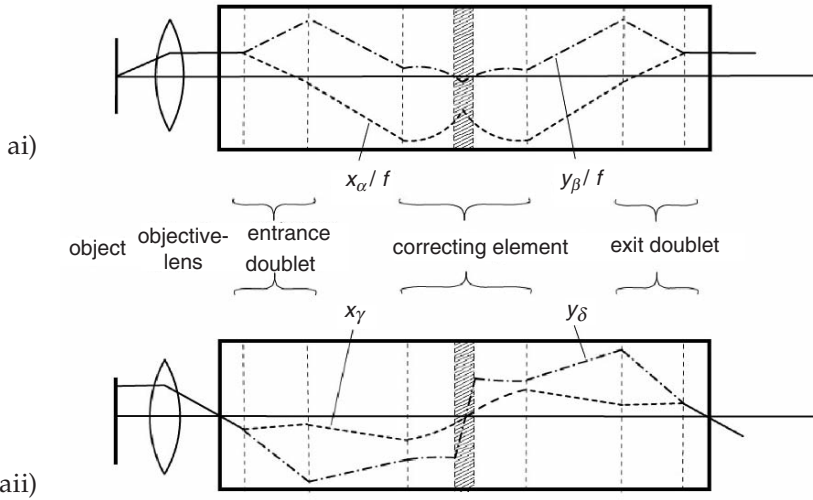


Figure 10-11. Aplanatic corrector of spherical and chromatic aberration. (a) A three-element corrector, in which the first and third elements cancel the chromatic aberration in the x - z plane while the central element cancels it in the y - z plane. (b) Axial rays in the first half of the corrector. (c) Field rays in the first half of the corrector. (After Weißbäcker and Rose, 2002, courtesy of the authors and Oxford University Press.)

Figure 10-12. Doubly symmetric electrostatic corrector (DECO) of spherical and chromatic aberration. (a) Axial and field rays in one-half of the DECO; each half consists of a corrector enclosed between quadrupole doublets. (b) Details of the axial rays in one unit of the DECO. (c) Details of the field rays in the first half of the DECO. (After Weißbäcker and Rose, 2002, courtesy of the authors and Oxford University Press.)



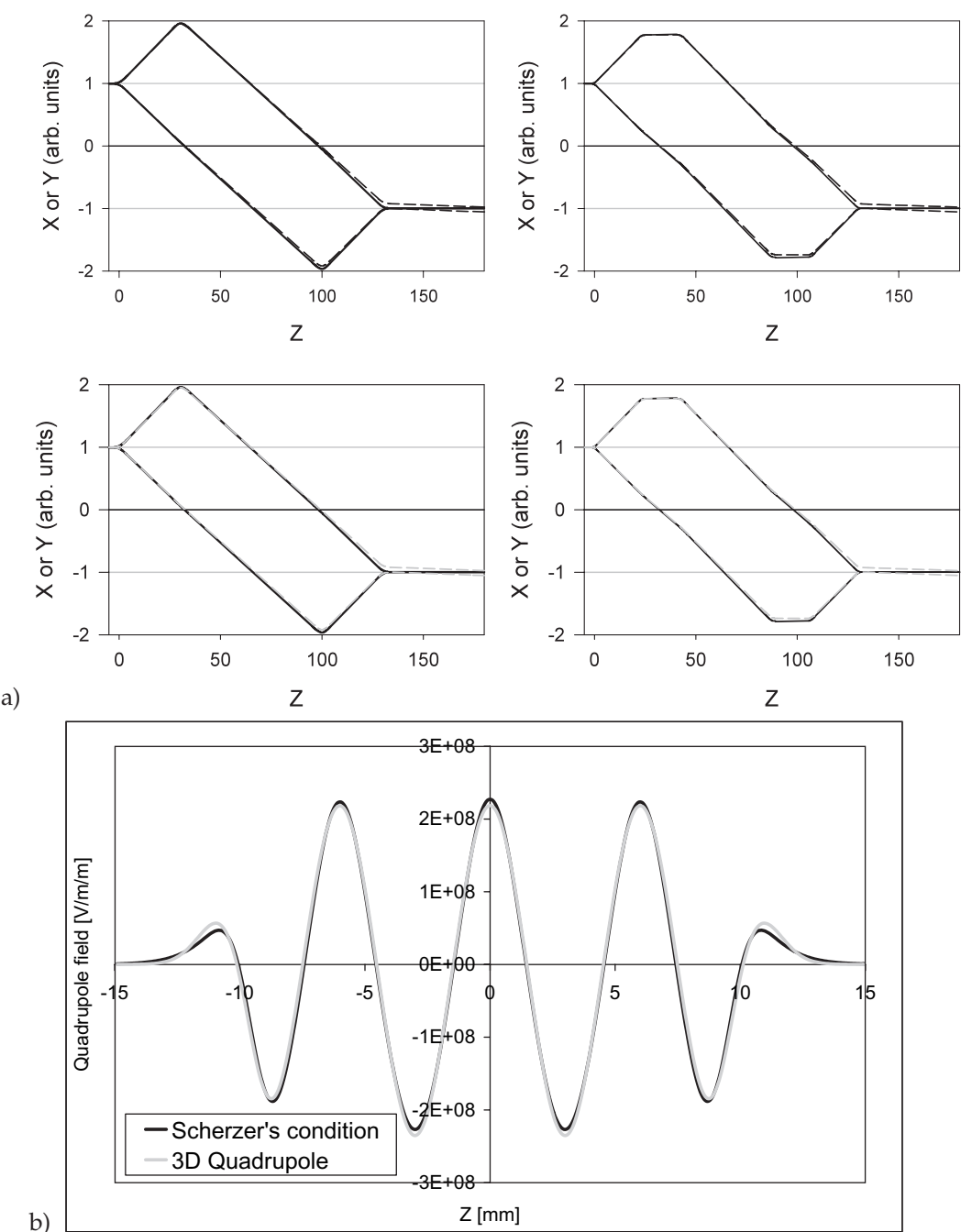


Figure 10–13. Electrostatic correction of chromatic aberration. (a) Rays in a quadrupole quadruplet or sextuplet. (b) Match between the potentials needed to satisfy Scherzer’s condition and those in the corrector. (See color plate.)

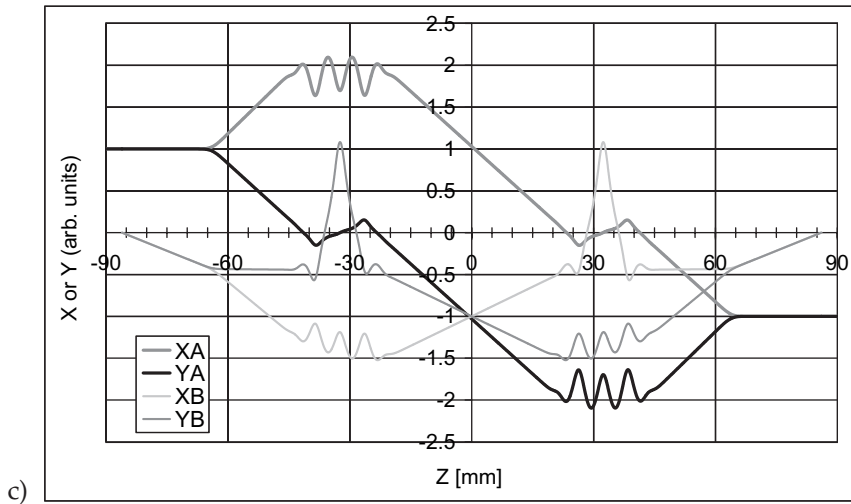


Figure 10-13. *Continued* (c) Gaussian rays in the sextuplet after correcting elements have been placed at the line foci. The chromatic aberration has been corrected and the residual chromatic aberration of magnification is small. (After Maas et al., 2001, 2003, courtesy of the authors, SPIE, and the Microscopy Society of America.)

Müller, 2004), for there exist more elaborate arrangements that do not have this handicap. Details are not available as yet.

3.3.3 Wien Filters and Correction

In an attempt to design a corrector that is reasonably easy to align and consists of as few separate elements as possible, Rose (1990) has examined the properties of an inhomogeneous Wien filter. If such a filter is to correct chromatic aberration instead of acting as a highly dispersive device, it must be nondispersive, double focusing, and free of all second rank aberrations (apart from the desired rotationally symmetric chromatic aberration). We have seen in other electron optical configurations how useful symmetry and antisymmetry are to eliminate certain aberrations. Here, all geometric second-order aberrations can be excluded by ensuring that the fundamental rays are antisymmetric with respect to the mid-plane of the filter or to the mid-plane of the first or second half of the filter. Rose's extremely complete study shows that a configuration can be found that will correct chromatic aberration. Furthermore, by adding sextupoles before and after the corrector (Figure 10-14), axial chromatic aberration and spherical aberration can be corrected independently.

This idea has been revived by Mentink et al. (1999) and Steffen et al. (2000), who were seeking further simplification and reduction of the number of optical elements and excitations. Rose's device consisted of a relatively long combined multipole (octopole or dodecapole) with both magnetic and electrostatic components (as in all Wien filters) for chromatic correction, with external sextupoles. In this new design, chromatic correction is achieved by means of the usual magnetic and electrostatic fields of the Wien filter onto which an electrostatic quad-

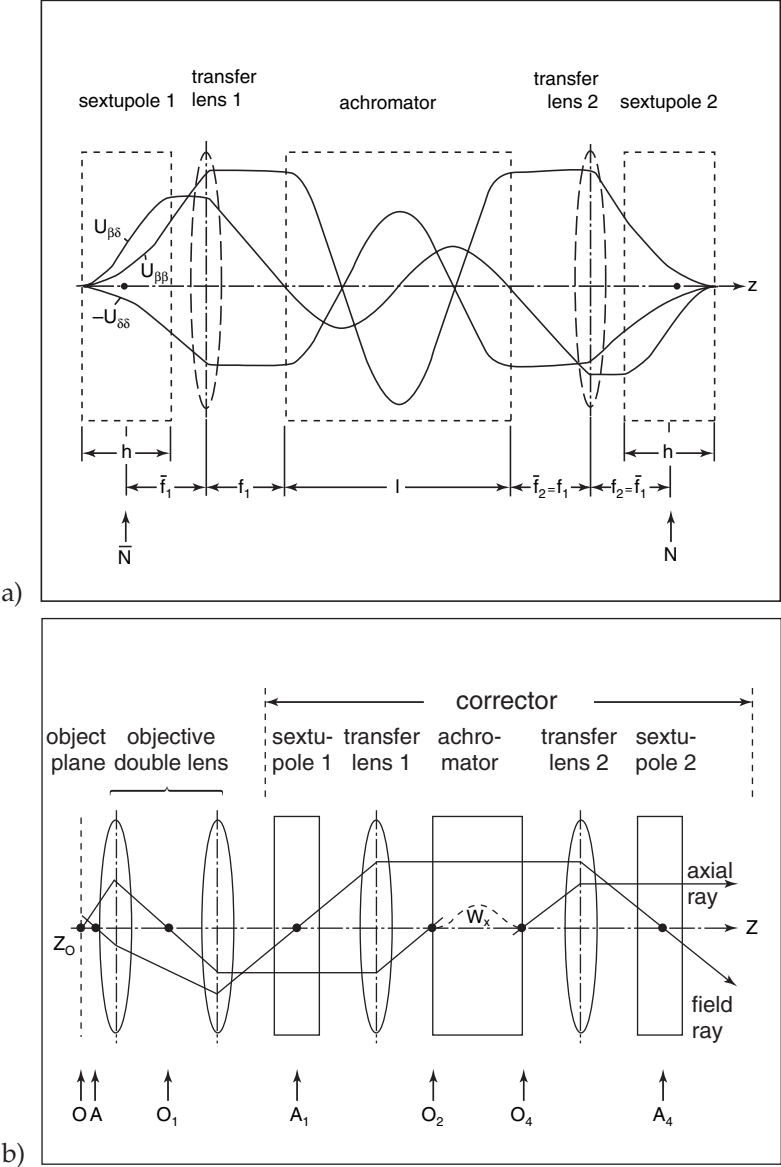


Figure 10–14. (a) Wien filter flanked by two sextupoles and the necessary transfer lenses, capable of correcting both chromatic and spherical aberration. The nodal points N and $\bar{N}$ coincide with the coma-free points. (b) The corrector in conjunction with a coma-free objective. (After Rose, 1990, courtesy of the author and the Wissenschaftliche Verlagsgesellschaft.)

rupolet is superimposed. No other magnetic fields are required. Spherical aberration is corrected by superimposing sextupole fields, which are of opposite sign in the first and second halves of the Wien filter zone. Unwanted aberrations are introduced by the fringing fields at the extremities of the corrector but it should be possible to control them by shaping the ends of the electrodes and magnetic poles.

3.3.4 Mirror Correctors

Several schemes for compensating the aberrations of round lenses by introducing an electron mirror (see Recknagel, 1937, for early work on electron mirrors) into the optical system have been proposed (Zworykin et al., 1945; Ramberg, 1949; Kasper, 1968/9). Aberration correction has now reached the stage at which simultaneous correction of all the aberrations that are liable to impair the performance of the instrument in question must be envisaged: correction of individual aberrations without considering the effect of the remainder is no longer sufficient. For this reason, we describe here only the scheme devised by Preikszas and Rose (1997) and surveyed by Hartel et al. (2002), intended for the spectromicroscope for all relevant techniques (SMART) project at BESSY II and also adopted for PEEM3 at the Lawrence Berkeley National Laboratory (Rose et al., 2004; Feng et al., 2005; Schmid et al., 2005). This differs from earlier schemes, notably that of Rempfer and Mauck (1992) and Rempfer et al. (1997), in that the beam splitter, the role of which is to separate the beam incident on the mirror from the beam emerging from it, is nondispersive (Figure 10–15); however, for the application for which it was designed, a low-energy emission microscope, the corrector shown in Figure 10–15 is effective. A four-electrode mirror, such as that shown in Figure 10–16, offers enough degrees of freedom to adjust the focal length and spherical and chromatic aberration coefficients satisfactorily. As an example, Preikszas and Rose (1997) show that the spherical and chromatic aberration coefficients can be chosen anywhere inside the shaded region in Figure 10–17 for a fixed position of the Gaussian image plane. Such a mirror could be combined with a dispersion-free magnetic beam splitter as shown in Figure 10–18a. Figure 10–18b shows the device incorporated in the SMART.

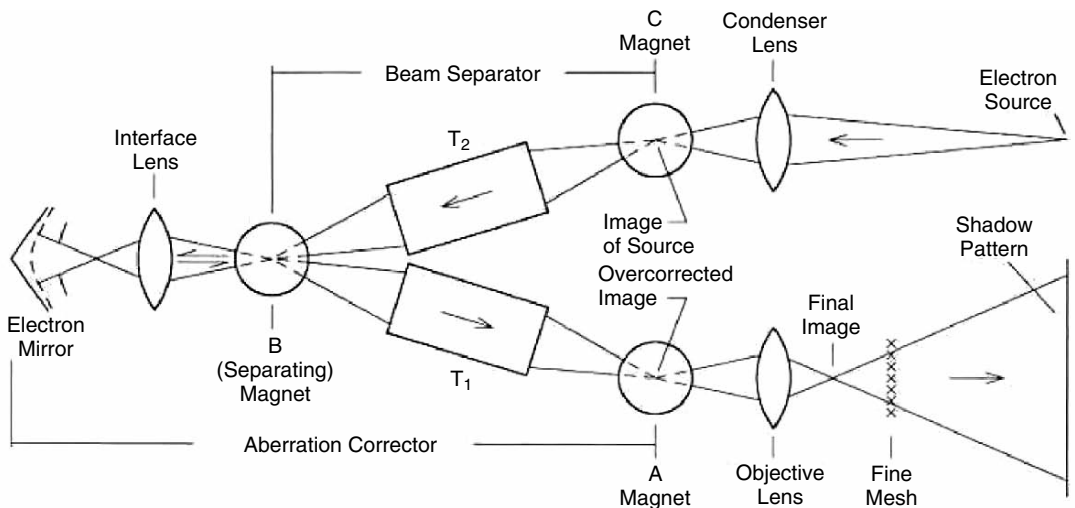


Figure 10–15. The optical system described by Rempfer et al. (1997) for correcting chromatic and spherical aberration with the aid of an electron mirror. T_1 and T_2 are lens triplets, which act as transfer lenses. The shadow thrown by the fine mesh gives a measure of the aberration coefficients. (After Rempfer et al., 1997, courtesy of the authors and the Microscopy Society of America.)

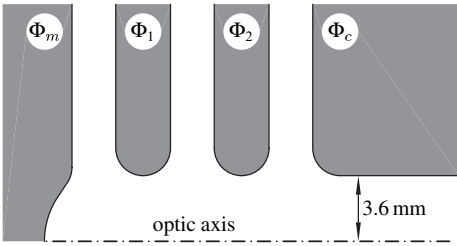


Figure 10-16. Section of a tetrode mirror. The potentials applied to the electrons determine the focal length and the chromatic and spherical aberration coefficients. (After Preikszas and Rose, 1997, courtesy of the authors and Oxford University Press.)

3.4 All-Purpose Correctors

In his keynote lecture on “High-Performance Electron Microscopes of the Future,” delivered at the twelfth European Congress on Electron Microscopy in Brno, 2000, Rose described a “hexapole planator” capable of correcting the field curvature and (third-order) astigmatism. Such correction is essential for projection lithography systems. This proposal was explored in detail by Munro et al. (2001) who, in an important paper, first examined the present limitations of projection systems for charged particles and then analyzed an electrostatic and a magnetic configuration free of all third-order geometric aberrations, inspired by Rose’s hexapole planator.

Subsequently, two very complex correctors have been proposed by Rose, which are capable of correcting the spherical and chromatic aberrations as well as other primary aberrations that could be harmful once the axial aberrations have been brought under control (Rose, 2003, 2004, 2005). These are the *superaplanator* and the *ultracorrector*. The first of these is suitable for a transmission electron microscope: it consists of two symmetrical quadrupole quintuplets and three (or more) octopoles and corrects spherical and chromatic aberration. Very high stability of the excitations is essential (Rose, 2005), of the order of a few parts in 10⁷, but this is attainable today. Symmetry is exploited to make the device as easy to use as possible. Thus the quadrupole fields are symmetric with respect to the center plane of each quintuplet; conversely, the whole

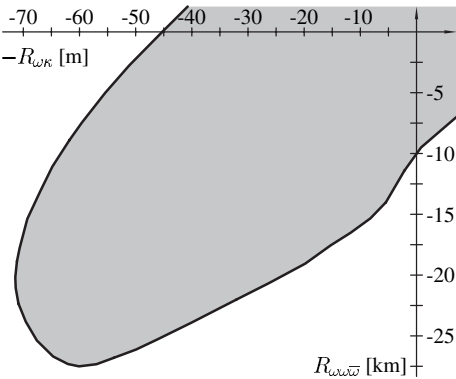


Figure 10-17. Behavior of the aberration coefficients of the tetrode mirror of Figure 10-16. For a fixed image plane, the spherical and chromatic aberration coefficients fall within the shaded zone. (After Preikszas and Rose, 1997, courtesy of the authors and Oxford University Press.)

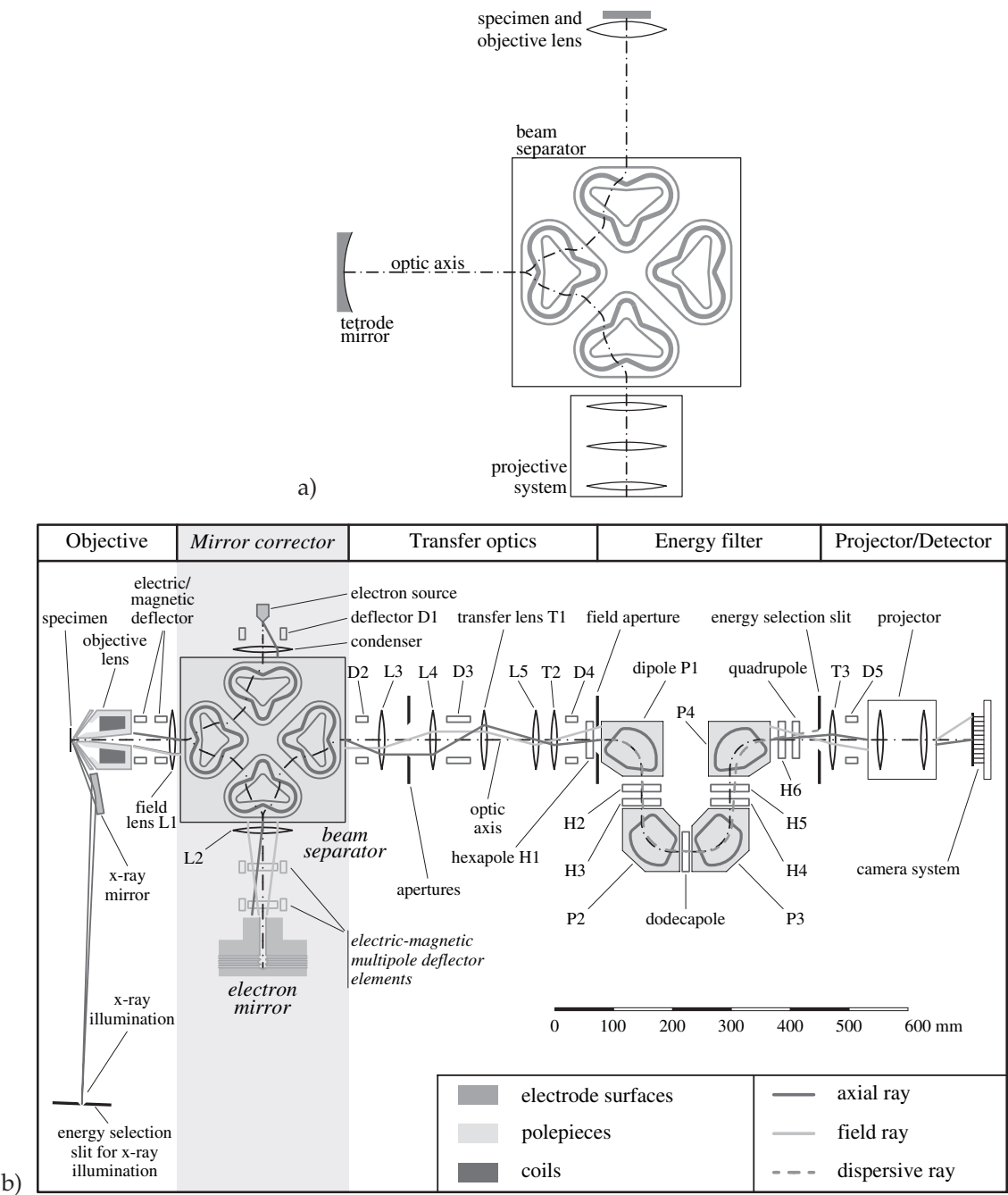


Figure 10–18. Incorporation of the tetrode mirror of Figure 10–16 into a complete system. The beam separator is free of dispersion. (a) Basic configuration. (b) The entire layout of the SMART (spectro-microscope for all relevant techniques). [(a) After Preikszas and Rose, 1997, courtesy of the authors and Oxford University Press. (b) After Hartel et al., 2002, courtesy of the authors and Elsevier.]

(double-quintuplet) unit exhibits antisymmetry about its mid-plane (the plane midway between the two quintuplets). An octopole is placed in this mid-plane and at the center of each quintuplet (Figure 10–19).

The superaplanator is adequate for the transmission electron microscope, where no particular attention need be paid to third-order astigmatism and field curvature, the zone to be imaged being so small. In electron projection lithography, however, it is desirable to correct all the primary chromatic and geometric aberrations. This can in principle be achieved with the configuration known as the ultra-corrector. Here, two identical multipole multiplets are used. Each of these consists of seven quadrupoles and seven octopoles, themselves disposed symmetrically about the center plane of the multiplet. Once again, the multiplet fields are antisymmetric with respect to the plane midway between two multiplets in which an additional octopole is situated (Figure 10–20). Detailed description of the modes of action of these complex structures is to be found in the work of Rose already cited. The suitability of these correctors for the Transmission Electron Aberration-corrected Microscope (TEAM) project is currently (2005) being examined (O’Keefe, 2003, 2004; Kabius et al., 2004).

4 Concluding Remarks

If any lesson is to be learned from past attempts to correct spherical and chromatic aberration, it is that complex systems require a rapid and efficient means of adjustment and that beyond a certain degree of complexity, only computer control offers the necessary speed and flexibility. Nevertheless, it remains important to simplify the task as far as possible and this is achieved by exploiting the symmetry properties. The intrinsic simplification of the optics provided by the Russian quadruplet (or multiplet) has long been known and, more recently, very

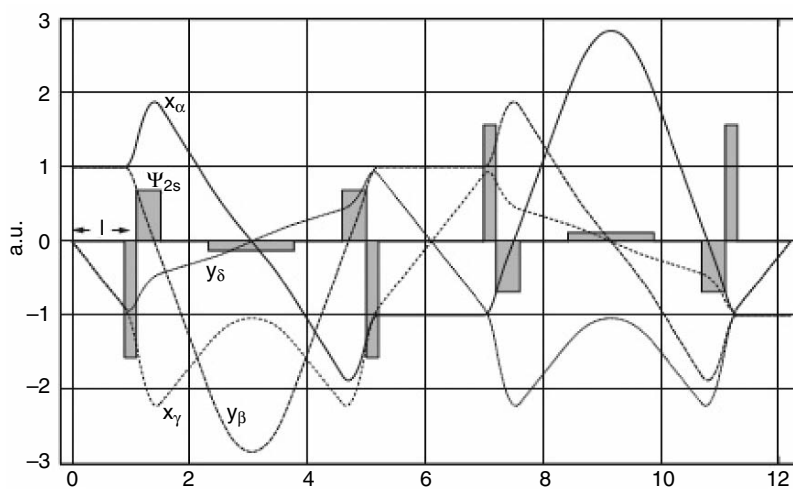


Figure 10–19. The superaplanator. (After Rose, 2005, courtesy of the author and Elsevier.)

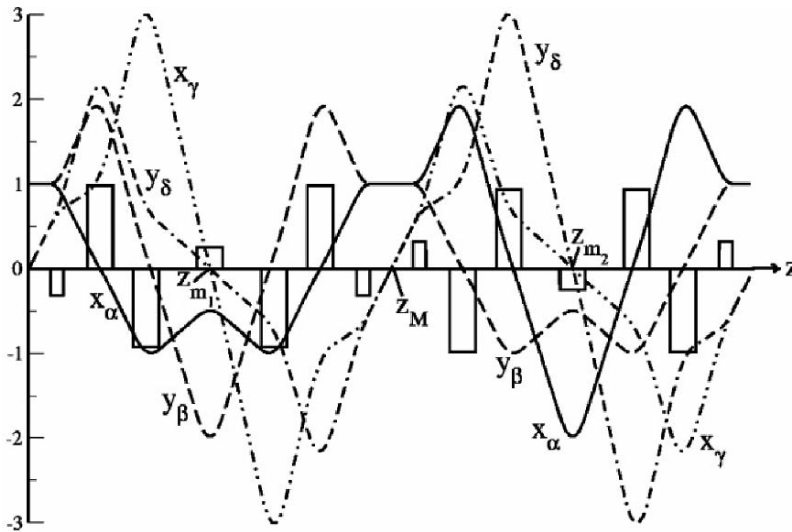


Figure 10–20. The ultracorrector. (After Rose, 2005, courtesy of the author and Elsevier.)

ingenious blends of symmetry and antisymmetry have been beneficial, or indeed indispensable, in the design of Ω and related types of filters, such as the mandoline filter (see Rose and Krahl, 1995, for a thorough account and Tsuno, 2001, for a schematic appraisal). The symmetry and antisymmetry on which the multicomponent structures of the superaplanator and the untracorrector are based not only help us to understand how they work but also render the adjustment more systematic and less liable to instability.

The quest for aberration correctors has always been fuelled by the desire for ever higher resolution but, now that correction has been achieved, microscopists are beginning to realise that such devices will be useful in many other areas of microscopy. *In-situ* microscopy requires lenses with ample space between the polepieces and the coefficients of spherical and chromatic aberration of such lenses are large. By correcting these aberrations, *in-situ* microscopy should be capable of providing much finer information. In the life sciences, chromatic correction will make it possible to work with thicker specimens. And no doubt this is only the beginning. The new century is witnessing a new era in electron microscopy.

5 Appendix I

5.1 Power Series Expansions of Electrostatic Potential and Vector Potential

The power series expansions used in the foregoing text are identical with those adopted in Hawkes and Kasper (1989) and are reproduced here for the reader's convenience. However, since the work of Rose is very frequently referred to, the relation between the notation used in his chapter in Ernst and Rühle (2003) and that employed here is also given.

5.1.1 Expansions for the Electrostatic Potential

In Hawkes and Kasper (1989), the electrostatic potential is written as

$$\begin{aligned}\Phi(x, y, z) = & \phi(z) - \frac{1}{4}(x^2 + y^2)\phi''(z) + \frac{1}{64}(x^2 + y^2)^2 - xF_1(z) - yF_2(z) \\ & + \frac{1}{8}(x^2 + y^2)(xF_1'' + yF_2'') + \frac{1}{2}(x^2 - y^2)p_2(z) + xyq_2(z) \\ & - \frac{1}{24}(x^2 + y^2)(x^2 - y^2)p_2'' - \frac{1}{12}(x^2 + y^2)xyq_2'' \\ & - \frac{1}{6}(x^3 - 3xy^2)p_3(z) + \frac{1}{6}(y^3 - 3x^2y)q_3(z) \\ & + \frac{1}{24}(x^4 - 6x^2y^2 + y^4)p_4(z) + \frac{1}{6}(x^2 - y^2)xyq_4(z)\end{aligned}$$

Rose (2003c) uses the expansion

$$\varphi(w, z) = \operatorname{Re} \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} (-)^l \frac{m!}{l!(m+l)!} \left(\frac{w\bar{w}}{4} \right)^l \bar{w}^m \Phi_m^{[2l]}(z)$$

in which

$$w = x + iy$$

and $[2l]$ signifies $2l$ -fold differentiation with respect to z . An overbar indicates the complex conjugate.

This leads to the following correspondences:

$$\begin{aligned}\phi(z) &= \Phi_0(z) \\ F_1(z) &= -\Phi_1^{(r)}, F_2(z) = -\Phi_1^{(i)}, \Phi_1 = -(F_1 + iF_2) \\ p_2(z) &= 2\Phi_2^{(r)}, q_2(z) = 2\Phi_2^{(i)}, \Phi_2 = \frac{1}{2}(p_2 + iq_2) \\ p_3(z) &= -6\Phi_3^{(r)}, q_3(z) = -6\Phi_3^{(i)}, \Phi_3 = -\frac{1}{6}(p_3 + iq_3) \\ p_4(z) &= 24\Phi_4^{(r)}, q_4(z) = 24\Phi_4^{(i)}, \Phi_4 = \frac{1}{24}(p_4 + iq_4)\end{aligned}$$

and we have written

$$\Phi_m = \Phi_m^{(r)} + i\Phi_m^{(i)}$$

5.1.2 Expansions for the Vector Potential

In Hawkes and Kasper, the components of the vector potential are written as

$$\begin{aligned}A_x = & -\frac{y}{2} \left(B - \frac{1}{8}(x^2 + y^2)B'' \right) + \frac{1}{4}(x^2 - y^2)B_2' \\ & - \frac{1}{48}(x^2 - y^2)(x^2 + y^2)B_2'' - \frac{1}{2}xyB_1' + \frac{1}{24}xy(x^2 + y^2)B_1''' \\ & - \frac{1}{12}(x^3 - 3xy^2)Q_2' - \frac{1}{12}(y^3 - 3x^2y)P_2' \\ & + \frac{1}{48}(x^4 - 6x^2y^2 + y^4)Q_3' - \frac{1}{12}(x^2 - y^2)xyP_3'\end{aligned}$$

$$\begin{aligned}
A_y = & \frac{x}{2} \left(B - \frac{1}{8}(x^2 + y^2)B'' \right) + \frac{1}{4}(x^2 - y^2)B'_1 \\
& - \frac{1}{48}(x^2 - y^2)(x^2 + y^2)B'''_1 + \frac{1}{2}xyB'_2 - \frac{1}{24}xy(x^2 + y^2)B'''_2 \\
& - \frac{1}{12}(x^3 - 3xy^2)P'_2 + \frac{1}{12}(y^3 - 3x^2y)Q'_2 \\
& + \frac{1}{48}(x^4 - 6x^2y^2 + y^4)P'_3 + \frac{1}{12}(x^2 - y^2)xyQ'_3
\end{aligned}$$

$$\begin{aligned}
A_z = & -xB_2(z)_1 - yB_1(z) + \frac{1}{8}(x^2 + y^2)(xB''_2 - yB''_1) \\
& + \frac{1}{2}(x^2 - y^2)Q_2(z) - xyP_2(z) - \frac{1}{24}(x^2 + y^2)(x^2 - y^2)Q''_2 \\
& + \frac{1}{12}(x^2 + y^2)xyP''_2 - \frac{1}{6}(x^3 - 3xy^2)Q_3(z) \\
& - \frac{1}{6}(y^3 - 3x^2y)P_3(z) + \frac{1}{24}(x^4 - 6x^2y^2 + y^4)Q_4(z) \\
& - \frac{1}{6}(x^2 - y^2)xyP_4(z)
\end{aligned}$$

In Rose, the same gauge is adopted and the components of the vector potential are given by

$$\begin{aligned}
A = A_x + iA_y &= \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \frac{(-)^l}{2i} \frac{m!}{l!(m+1+l)!} \left(\frac{w\bar{w}}{4} \right)^l w^{m+1} \bar{\Psi}_m^{[2l+1]}(z) \\
A_z = \text{Im} \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} &(-)^l \frac{m!}{l!(m+l)!} \left(\frac{w\bar{w}}{4} \right)^l \bar{w}^m \Psi_m^{[2l]}(z)
\end{aligned}$$

The correspondence is now

$$\begin{aligned}
\Psi_1 &= -(B_1 + iB_2) \\
\Psi_2 &= \frac{1}{2}(P_2 + iQ_2) \\
\Psi_3 &= -\frac{1}{6}(P_3 + iQ_3) \\
\Psi_4 &= \frac{1}{24}(P_4 + iQ_4)
\end{aligned}$$

References

- Aken, R.H. van (2005). Low-energy electron beams through ultra-thin foils, applications for electron microscopy. Proefschrift, Delft.
- Aken, R.H. van, Hagen, C.W., Barth, J.E. and Kruit, P. (2002a). Low-energy foil aberration corrector. *Ultramicroscopy* **93**, 321–330.
- Aken, R.H. van, Hagen, C.W., Barth, J.E. and Kruit, P. (2002b). Concept for a low energy foil corrector. In *Proceedings of the 15th International Congress on Electron Microscopy*, Durban (R. Cross, J. Engelbrecht, M. Witcomb, and P. Richards, Eds.), Vol. 1, 1101–1102.

- Aken, R.H. van, Lenc, M. and Barth, J.E. (2004). Aberration integrals for the low-voltage foil corrector. *Nucl. Instrum. Methods Phys. Res. A* **519**, 205–215.
- Archard, G.D. (1953). Magnetic electron lens aberrations due to mechanical defects. *J. Sci. Instrum.* **30**, 352–358.
- Archard, G.D. (1954). Requirements contributing to the design of devices used in correcting electron lenses. *Br. J. Appl. Phys.* **5**, 294–299.
- Archard, G.D. (1955). A possible chromatic aberration system for electron lenses. *Proc. Phys. Soc. (London)* **68**, 817–829.
- Archard, G.D. (1958). An unconventional electron lens. *Proc. Phys. Soc. (London)* **72**, 135–137.
- Bacon, N.J., Corbin, G.J., Dellby, N., Hrnčirik, P., Krivanek, O.L., McManama-Smith, A., Murfitt, M.F. and Szilagy, Z.S. (2005). Nion UltraSTEM: an aberration-corrected STEM for imaging and analysis. *Microsc. Microanal.* **11** (Supplement 2), 1422–1423.
- Baranova, L.A. and Yavor, S. Ya. (1989). The optics of round and multipole electrostatic lenses. *Adv. Electron. Electron Phys.* **76**, 1–207.
- Baranova, L.A., Read, F.H. and Cubric, D. (2004). Computational simulation of an electrostatic aberration corrector for a low-voltage scanning electron microscope. *Nucl. Instrum. Methods Phys. Res. A* **519**, 42–47.
- Bastian, B., Spengler, K. and Typke, D. (1971). An electric-magnetic octupole element to correct spherical and chromatic aberrations of electron lenses. *Optik* **33**, 591–596.
- Batson, P.E. (2003). Aberration correction results in the IBM STEM instrument. *Ultramicroscopy* **96**, 239–249.
- Batson, P.E., Mook, H.W. and Kruit, P. (2000). High brightness monochromator for STEM. In *Microbeam Analysis 2000* (D.B. Williams and R. Shimizu, Eds.), 213–214 (Institute of Physics, Bristol).
- Batson, P.E., Dellby, N. and Krivanek, O.L. (2002). Sub-ångström resolution using aberration corrected optics. *Nature* **418**, 617–620.
- Beck, V. (1977). Experiments with a quadrupole-octupole corrector in a STEM. *Proc. Annu. Meeting EMSA* **35**, 90–91.
- Beck, V.D. (1979). A hexapole spherical aberration corrector. *Optik* **53**, 241–255.
- Beck, V. and Crewe, A.V. (1974). A quadrupole octupole corrector for a 100 kV STEM. *Proc. Annu. Meeting EMSA* **32**, 426–427.
- Beck, V. and Crewe, A.V. (1976). Progress in aberration correction in a STEM. *Proc. Annu. Meeting EMSA* **34**, 578–579.
- Benner, G., Orchowski, A., Haider, M. and Hartel, P. (2003a). State of the first aberration-corrected, monochromatized 200 kV FEG-TEM. *Microsc. Microanal.* **9** (Supplement 2), 938–939.
- Benner, G., Matijevich, M., Orchowski, A., Schindler, B., Haider, M. and Hartel, P. (2003b). State of the first aberration-corrected, monochromatized 200 kV FEG-TEM. *Microsc. Microanal.* **9** (Supplement 3), 38–39.
- Benner, G., Essers, E., Huber, B. and Orchowski, A. (2003c). Design and first results of SESAM. *Microsc. Microanal.* **9** (Supplement 3), 66–67.
- Benner, G., Essers, E., Matijevic, M., Orchowski, A., Schlossmacher, P. and Thesen, A. (2004a). Performance of monochromized and aberration-corrected TEMs. *Microsc. Microanal.* **10** (Supplement 2), 108–109.
- Benner, G., Matijevic, M., Orchowski, A., Schlossmacher, P., Thesen, A., Haider, M. and Hartel, P. (2004b). Sub-Ångström and sub-eV resolution with the analytical SATEM. In *Materials Research in an Aberration-Free Environment*. *Microsc. Microanal.* **10** (Supplement 3), 6–7.
- Bernhard, W. (1980). Erprobung eines sphärisch und chromatisch korrigierten Elektronenmikroskops. *Optik* **57**, 73–94.

- Bernhard, W. and Koops, H. (1977). Kompensation der Farbabhängigkeit der Vergrößerung und der Farbabhängigkeit der Bilddrehung eines Elektronenmikroskops. *Optik* **47**, 55–64.
- Bimurzaev, S.B., Serikbaeva, G.S. and Yakushev, E.M. (2003). Electrostatic mirror objective with eliminated spherical and axial chromatic aberrations. *J. Electron Microsc.* **52**, 365–368.
- Bimurzaev, S.B., Serikbaeva, G.S. and Yakushev, E.M. (2004). Calculation of the focussing quality of the electrostatic mirror objective free of the third-order spherical aberration. *Nucl. Instrum. Methods Phys. Res. A* **519**, 70–75.
- Browning, N.D., Arslan, I., Erni, R., Idrobo, J.C., Ziegler, A., Bradley, J., Dai, Z., Stach, E.A. and Bleloch, A. (2006). Monochromators and aberration correctors: taking EELS to new levels of energy and spatial resolution. *J. Phys. Conf.* **26**, 59–64 [EMAG-NANO 2005]
- Burfoot, J.C. (1953). Correction of electrostatic lenses by departure from rotational symmetry. *Proc. Phys. Soc. (London)* **B66**, 775–792.
- Calvo, M. (2002). Transfer theory and image formation in a time dependent aberrationless magnetic field lens. *Optik* **113**, 233–240.
- Calvo, M. (2004). Optimal resolution of a time-dependent aberrationless magnetic lens. *Ultramicroscopy* **99**, 179–187.
- Calvo, M. and Lazcano, P. (2002). Theory of a time dependent magnetic lens. *Optik* **113**, 31–34.
- Calvo, M. and Laroze, D. (2002). Aberrations theory of a time dependent magnetic field lens. *Optik* **113**, 429–434.
- Chand, G., Saxton, W.O. and Kirkland, A.I. (1995). Aberration measurement and automated alignment of the TEM. In *Electron Microscopy and Analysis 1995* (D. Cherns, Ed.), 297–300 (Institute of Physics, Bristol).
- Chang, L.Y., Chen, F.R., Kirkland, A.I. and Kai, J.J. (2003). Calculations of spherical aberration-corrected imaging behaviour. *J. Electron Microsc.* **52**, 359–364.
- Chang, L.Y., Kirkland, A.I. and Titchmarsh, J.M. (2006). On the importance of fifth-order spherical aberration for a fully corrected electron microscope. *Ultramicroscopy* **106**, 301–306.
- Chao, L.C., Orloff, J. and Wang, L. (1997). Spherical aberration corrector using space charge. *J. Vac. Sci. Technol.* **B15**, 2732–2736.
- Crewe, A.V. (1978). Is there a future for the STEM? In *Proceedings of the Ninth International Congress on Electron Microscopy*, Toronto (J.M. Sturgess, Ed.), Vol. 3, 197–204 (Microscopical Society of Canada, Toronto).
- Crewe, A.V. (1980). Studies on sextupole correctors. *Optik* **57**, 313–327.
- Crewe, A.V. (1982). A system for the correction of axial aperture aberrations in electron lenses. *Optik* **60**, 271–281.
- Crewe, A.V. (2002). Some Chicago aberrations. *Microsc. Microanal.* **8** (Supplement 2), 4–5.
- Crewe, A.V. and Kopf, D. (1980a). A sextupole system for the correction of spherical aberration. *Optik* **55**, 1–10.
- Crewe, A.V. and Kopf, D. (1980b). Limitations of sextupole correctors. *Optik* **56**, 391–399.
- Crewe, A.V., Cohen, D. and Meads, P. (1968). A multipole element for the correcting of spherical aberration. In *Proceedings of the Fourth Regional Conference on Electron Microscopy*, Rome (D.S. Bocciarelli, Ed.), Vol. I, 183 (Tipographia Poliglotta Vaticana, Rome).
- Crewe, A.V., Ruan, S., Tsai, F. and Korda, P. (1995a). The first test on a magnetically focused mirror corrector. In *Electron Microscopy and Analysis 1995* (D. Cherns, Ed.), 301–304 (Institute of Physics, Bristol).
- Crewe, A.V., Tsai, F., Korda, F. and Ruan, S. (1995b). The first test on a magnetically focused mirror corrector. In *Microscopy and Analysis 1995* (G.W. Bailey,

- M.H. Ellisman, R.A. Hennigar and N.J. Zaluzec, Eds.), 562–563 (Jones and Begell, New York).
- Crewe, A.V., Ruan, S., Korda, P. and Tsai, F.C. (2000). Studies of a magnetically focused electrostatic mirror. I. Experimental test of the first order properties. *J. Microsc.* **197**, 110–117.
- Dellby, N., Krivanek, O.L. and Lupini, A.R. (2000). Progress in aberration-corrected STEM. *Microsc. Microanal.* **6** (Supplement 2), 100–101.
- Dellby, N., Krivanek, O.L., Nellist, P.D., Batson, P.E. and Lupini, A.R. (2001). Progress in aberration-corrected scanning transmission electron microscopy. *J. Electron Microsc.* **50**, 177–185.
- Dellby, N., Krivanek, O.L., Murfitt, M.F. and Nellist, P.D. (2005). Design and testing of a quadrupole/octupole C_3/C_5 aberration corrector. *Microsc. Microanal.* **11** (Supplement 2), 2130–2131.
- Deltrap, J.M.H. (1964). Correction of spherical aberration with combined quadrupole–octopole units. In *Proceedings of the Third European Regional Conference on Electron Microscopy*, Prague (M. Titlbach, Ed.), Vol. A, 45–46 (Czechoslovak Academy of Sciences, Prague).
- Der-Shvarts, G.V. (1954). [Influence of imperfections of the rotational symmetry of the focusing fields on the resolution of the magnetic objectives of electron microscopes]. *Zh. Tekh. Fiz.* **24**, 859–870.
- Dragt, A.J. and Forest, E. (1986). Lie algebraic theory of charged-particle optics and electron microscopes. *Adv. Electron. Electron Phys.* **67**, 65–120.
- Dunin-Borkowski, R.E., Kasama, T., Cervera, L., Twitchett, A.C., Midgeley, P. A., Robins, A.C., Smith, D.W., Gronsky, J.J., Thomas, C.M., Fischione, P.E., Hetherington, C.J.D. and Kirkland, A.I. (2004). Aberration correction: Some advantages and alternatives. In *Materials Research in an Aberration-Free Environment*. *Microsc. Microanal.* **10** (Supplement 3), 22–23.
- Dymnikov, A.D. and Yavor, S. Ya. (1963). [Four quadrupole lenses as an analogue of an axially symmetric system]. *Zh. Tekh. Fiz.* **33**, 851–858; *Sov. Phys. Tech. Phys.* **8**, 639–643.
- Feng, J., Forest, E., Macdowell, A.A., Marcus, M., Padmore, H., Raoux, S., Robin, D., Scholl, A., Schlueter, R., Schmid, P., Stöhr, J., Wan, W., Wei, D.H. and Wu, Y. (2005). An x-ray photoemission electron microscope using an electron mirror aberration corrector for the study of complex materials. *J. Phys. Condensed Matter* **17**, S1339–S1350.
- Fey, G. (1980). Elektrische Versorgung eines elektronenoptischen Korrektivs. *Optik* **55**, 55–65.
- Fink, R., Weiss, M.R., Umbach, E., Preikszas, D., Rose, H., Spehr, R., Hartel, P., Engel, W., Degenhardt, R., Wichtendahl, R., Kuhlenbeck, H., Erlebach, W., Ihmann, K., Schlögl, R., Freund, H.-J., Bradshaw, A.M., Lilienkamp, G., Schmidt, T., Bauer, E. and Benner, G. (1997). SMART: A planned ultrahigh-resolution spectromicroscope for BESSY II. *J. Electron Spectrosc. Relat. Phenom.* **84**, 231–250.
- Foschepoth, M. and Kohl, H. (1998). Amplitude contrast—a way to obtain directly interpretable high-resolution images in a spherical-aberration-corrected transmission electron microscope. *Phys. Status Solidi A* **166**, 357–366.
- Freitag, B., Kujawa, S., Mul, P.M., Tiemeijer, P.C. and Snoeck, E. (2004a). First experimental proof of spatial resolution improvement in a monochromized and Cs-corrected TEM. In *Proceedings of the Eighth Asia-Pacific Conference on Electron Microscopy*, Kanazawa (N. Tanaka, Y. Takano, H. Mori, H. Seguchi, S. Iseki, H. Shimada and E. Simamura, Eds.), 18–19 (8APEM Publication Committee, Uchinada, Ishikawa).
- Freitag, B., Kujawa, S., Mul, P.M. and Tiemeijer, P.C. (2004b). First experimental proof of spatial resolution improvement in a monochromized and Cs-

- corrected TEM. In *Materials Research in an Aberration-Free Environment. Microsc. Microanal.* **10** (Supplement 3), 4–5.
- Freitag, B., Kujawa, S., Mul, P.M., Ringnalda, J. and Tiemeijer, P.C. (2005). Breaking the spherical and chromatic aberration barrier in transmission electron microscopy. *Ultramicroscopy* **102**, 209–214.
- Gabor, D. (1950). Electron optics at high frequencies and at relativistic velocities. *Rev. Opt.* **29**, 209–231.
- Glaser, W. (1940). Über ein von sphärische Aberration freies Magnetfeld. *Z. Phys.* **116**, 19–33 and 734–735.
- Glaser, W. (1942). Über elektronenoptische Abbildung bei gestörter Rotations-symmetrie. *Z. Phys.* **120**, 1–15.
- Glaser, W. and Schiske, P. (1953). Bildstörungen durch Polschuhasymmetrien bei Elektronenlinsen. *Z. Angew. Phys.* **5**, 329–339.
- Haider, M. (1998). Correctors for electron microscopes: Tools or toys for scientists? In *Proceedings of EUREM-11, the 11th European Conference on Electron Microscopy*, Dublin 1996 (CESEM, Ed.), Vol. I, 363–364 (CESEM, Brussels).
- Haider, M. (2000). Towards sub-Ångström point resolution by correction of spherical aberration. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.145–I.148 (Czechoslovak Society for Electron Microscopy, Brno).
- Haider, M. (2003). Current and future developments in order to approach a point resolution of 0.5 Å. *Microsc. Microanal.* **9** (Supplement 2), 930–931.
- Haider, M. and Müller, M. (2004). Design of an electron optical system for the correction of the chromatic aberration C_c of a TEM objective lens. In *Materials Research in an Aberration-Free Environment. Microsc. Microanal.* **10** (Supplement 3), 2–3.
- Haider, M. and Müller, H. (2005). Is there a road map of aberration correction towards ultra-high resolution in TEM and STEM? *Microsc. Microanal.* **11** (Supplement 2), 546–547.
- Haider, M. and Uhlemann, S. (1997). Seeing is not believing: Reduction of artefacts by an improved point resolution with a spherical aberration corrected 200 kV transmission electron microscope. *Microsc. Microanal.* **3** (Supplement 2), 1179–1180.
- Haider, M., Bernhardt, W. and Rose, H. (1982). Design and test of an electric and magnetic dodecapole lens. *Optik* **63**, 9–23.
- Haider, M., Braunshausen, G. and Schwan, E. (1995). Correction of the spherical aberration of a 200 kV TEM by means of a hexapole-corrector. *Optik* **99**, 167–179.
- Haider, M., Uhlemann, S., Schwan, E., Rose, H., Kabius, B. and Urban, K. (1998a). Electron microscopy image enhanced. *Nature* **392**, 768–769.
- Haider, M., Rose, H., Uhlemann, S., Kabius, B. and Urban, K. (1998b). Towards 0.1 nm resolution with the first spherically corrected transmission electron microscope. *J. Electron Microsc.* **47**, 395–405.
- Haider, M., Rose, H., Uhlemann, S., Schwan, E., Kabius, B., and Urban, K. (1998c). A spherical-aberration-corrected 200 kV transmission electron microscope. *Ultramicroscopy* **75**, 53–60.
- Haider, M., Uhlemann, S. and Zach, J. (2000). Upper limits for the residual aberrations of a high-resolution aberration-corrected STEM. *Ultramicroscopy* **81**, 163–175.
- Haider, M., Müller, M. and Hartel, P. (2004). Present state and future trends of aberration correction. In *Proceedings of the Eighth Asia-Pacific Conference on Electron Microscopy*, Kanazawa (N. Tanaka, Y. Takano, H. Mori, H. Seguchi, S. Iseki, H. Shimada and E. Simamura, Eds.), 16–17 (8APEM Publication Committee, Uchinada, Ishikawa).

- Hanai, T., Hibino, M. and Maruse, S. (1986). Measurement of axial geometrical aberrations of the probe-forming lens by means of the shadow image of fine particles. *Ultramicroscopy* **20**, 329–336.
- Hanai, T., Hibino, M. and Maruse, S. (1994). Effect of the three-fold astigmatism on the performance of the foil lens used for correction of the spherical aberration of a probe-forming lens. *Optik* **97**, 86–90.
- Hanai, T., Arai, S. and Hibino, M. (1995). Increase of current density of the electron probe by correction of the spherical aberration with a side-entry type foil lens. *J. Electron Microsc.* **44**, 301–306.
- Hanai, T., Yoshida, H. and Hibino, M. (1998). Characteristics and effectiveness of a foil lens for correction of spherical aberration in scanning transmission electron microscopy. *J. Electron Microsc.* **47**, 185–192.
- Hartel, P., Preikszas, D., Spehr, R. and Rose, H. (2000). Performance of the mirror corrector for an ultrahigh-resolution spectromicroscope. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.153–I.154 (Czechoslovak Society for Electron Microscopy, Brno).
- Hartel, P., Preikszas, D., Spehr, R., Müller, H. and Rose, H. (2002). Mirror corrector for low-voltage electron microscopes. *Adv. Imaging Electron Phys.* **120**, 41–133.
- Hartel, P., Müller, H., Uhlemann, S. and Haider, M. (2004). Residual aberrations of hexapole-type C_s -correctors. In *Proceedings of the 13th European Microscopy Congress*, Antwerp (D. van Dyck and P. van Oostveldt, Eds.), Vol. I, 41–42 (Belgian Society for Microscopy, Liège).
- *Hartel, P., Müller, H., Uhlemann, S. and Haider, M. (2007). First experimental results with advanced hexapole C_s -correctors. *Microsc. Microanal.* **13**, Supplement 3, 4–5.
- Haufe, G. (1958). Korrektur des Öffnungsfehlers bei elektrostatischen Linsen durch Raumladung. *Optik* **15**, 521–537.
- Hawkes, P.W. (1964). The paraxial chromatic aberrations of quadrupole systems. In *Proceedings of the Third European Regional Conference on Electron Microscopy*, Prague (M. Titlbach, Ed.), Vol. A, 5–6 (Czechoslovak Academy of Sciences, Prague).
- Hawkes, P.W. (1965a). The geometrical aberrations of general electron optical systems, I and II, *Philos. Trans. Roy. Soc. A* **257**, 479–552.
- Hawkes, P.W. (1965b). The paraxial chromatic aberrations of rectilinear orthogonal systems, *Optik* **22**, 543–551.
- Hawkes, P.W. (1970). Quadrupoles in electron lens design. *Adv. Electron. Electron Phys.* (Supplement 7).
- Hawkes, P.W. (1980). Methods of computing optical properties and combating aberrations for low-intensity beams. *Adv. Electron. Electron Phys.* (Supplement 13A), 45–157.
- Hawkes, P.W. (1997). Aberrations. In *Handbook of Charged Particle Optics* (J. Orloff, Ed.), 223–274 (CRC Press, Boca Raton, FL).
- Hawkes, P.W. (2002). Signposts in electron optics. *Adv. Imaging Electron Phys.* **123**, 1–28.
- *Hawkes, P.W. (2008). Aberrations. In *Handbook of Charged Particle Optics*, 2nd ed. (J. Orloff, Ed.; CRC Press, Boca Raton)
- *Hawkes, P.W., Ed. (2008). Aberration-corrected electron microscopy. *Adv. Imaging & Electron Phys.* **155**.
- Hawkes, P.W. and Kasper, E. (1989, 1994). *Principles of Electron Optics* (Academic Press, London).
- Henstra, A. and Krijn, M.P.C.M. (2000). An electrostatic achromat. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.155–I.156 (Czechoslovak Society for Electron Microscopy, Brno).

- Herrmann, K.-H., Schiske, P. and Zemlin, F. (1976). Einfluß und Nachweis des dreizähligen Astigmatismus in der Hochauflösungsmikroskopie. *Mikroskopie* **32**, 235.
- Hoch, H., Kasper, E. and Kern, D. (1976). Der Öffnungsfehler 3. Ordnung und die axiale Farbfehler von rotationssymmetrischen Elektronenlinsen mit gekrümmter geladener transparenter Folie. *Optik* **46**, 463–473.
- Honda, K., Uno, S., Nakamura, N., Matsuya, M., Achard, B. and Zach, J. (2004a). An automatic geometrical aberration correction system of scanning electron microscopes. In *Proceedings of the 13th European Microscopy Congress*, Antwerp (D. van Dyck and P. van Oostveldt, Eds.), Vol. I, 43–44 (Belgian Society for Microscopy, Liège).
- Honda, K., Uno, S., Nakamura, N., Matsuya, M. and Zach, J. (2004b). An automatic geometrical aberration correction system of scanning electron microscopes. In *Proceedings of the Eighth Asia-Pacific Conference on Electron Microscopy*, Kanazawa (N. Tanaka, Y. Takano, H. Mori, H. Seguchi, S. Iseki, H. Shimada and E. Simamura, Eds.), 44–45 (8APEM Publication Committee, Uchinada, Ishikawa).
- Hosokawa, F. and Sawada, H. (2003). A C_s -corrected TEM. *Denshi Kenbikyō* **38** (Supplement 2), 55–58.
- Hosokawa, F., Tomita, T., Naruse, M., Honda, T., Hartel, P. and Haider, M. (2003). A spherical aberration-corrected 200 kV TEM. *J. Electron Microsc.* **52**, 3–10.
- Huber, A. and Plies, E. (1999). Numerical ray tracing of electrons in different 3D fringing fields of spherical deflectors. *Proc. SPIE* **3777**, 264–274.
- Huber, A. and Plies, E. (2000). Simulation of electron trajectories in the 3D field of two coupled hemispherical electrostatic deflectors. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, 1473–1477 (Czechoslovak Society for Electron Microscopy, Brno).
- Huber, A., Bärtle, J. and Plies, E. (2004). Initial experiences with an electrostatic Ω -monochromator for electrons. *Nucl. Instrum. Methods Phys. Res.* **A519**, 320–324.
- Hutchison, J.L., Titchmarsh, J.M., Cockayne, D.J.H., Doole, R.C., Hetherington, C.J.D., Kirkland, A.I. and Sawada, H. (2005). A versatile double aberration-corrected energy filtered HREM/STEM for materials science. *Ultramicroscopy* **103**, 7–15.
- Ioanoviciu, D., Tsuno, K. and Martínez, G. (2004). Third order aberration theory of double Wien filters. *Rev. Sci. Instrum.* **75**, 4434–4441.
- Ishizuka, K. (1994). Coma-free alignment of a high-resolution electron microscope with three-fold astigmatism. *Ultramicroscopy* **55**, 407–418.
- Jia, C.L., Lentzen, M. and Urban, K. (2003). Atomic-resolution imaging of oxygen in perovskite ceramics. *Science* **299**, 870–873.
- Kabius, B., Haider, M., Uhlemann, S., Schwan, E., Urban, K. and Rose, H. (2002). First application of a spherical-aberration corrected transmission electron microscope in materials science. *J. Electron Microsc.* **51**, S51–S58.
- Kabius, B., Miller, D.J. and Zaluzec, N.J. (2004). Aberration correction for the TEAM project: Design and applications. In *Proceedings of the 13th European Microscopy Congress*, Antwerp (D. van Dyck and P. van Oostveldt, Eds.), Vol. I, 25–26 (Belgian Society for Microscopy, Liège).
- Kahl, F. and Rose, H. (1998). Outline of an electron monochromator with small Boersch effect. In *Proceedings of the 14th International Conference on Electron Microscopy*, Cancún (H.A. Calderón and M.J. Yacamán, Eds.), Vol. 1, 71–72 (Institute of Physics, Bristol).
- Kahl, F. and Rose, H. (2000). Design of a monochromator for electron sources. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno

- (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.459–I.460 (Czechoslovak Society for Electron Microscopy, Brno).
- Kahl, F. and Voelkl, E. (2001). Present concepts and designs for gun monochromators. *Microsc. Microanal.* **7** (Supplement 2), 922–923.
- Kasper, E. (1968/9). Die Korrektur des Öffnungs- und Farbfehlers im Elektronenmikroskop durch Verwendung eines Elektronenspiegels mit überlagertem Magnetfeld. *Optik* **28**, 54–64.
- Kazumori, H., Honda, K., Matsuya, M., Date, M. and Nielsen, C. (2004a). Field emission SEM with a spherical and chromatic aberration corrector. *Microsc. Microanal.* **10** (Supplement 2), 1370–1371.
- Kazumori, H., Honda, K., Matuya, M. and Date, M. (2004b). Field emission SEM with a spherical and chromatic aberration corrector. In *Proceedings of the Eighth Asia-Pacific Conference on Electron Microscopy*, Kanazawa (N. Tanaka, Y. Takano, H. Mori, H. Seguchi, S. Iseki, H. Shimada and E. Simamura, Eds.), 52–53 (8APEM Publication Committee, Uchinada, Ishikawa).
- Kel'man, V.M. and Yavor, S. Ya. (1961). Achromatic quadrupole electron lenses. *Zh. Tekh. Fiz.* **31**, 1439–1442; *Sov. Phys. Tech. Phys.* **6**, 1052–1054.
- Khursheed, A. (2002). A low voltage time of flight electron emission microscope. *Optik* **113**, 505–509.
- Khursheed, A. (2005). A method of dynamic chromatic correction in low-voltage scanning electron microscopes. *Ultramicroscopy* **103**, 255–260.
- Koops, H. (1978). Aberration correction in electron microscopy. In *Proceedings of the Ninth International Congress on Electron Microscopy*, Toronto (J.M. Sturgess, Ed.), Vol. 3, 185–196 (Microscopical Society of Canada, Toronto).
- Koops, H. (1978/9). Erprobung eines chromatisch korrigierten elektronenmikroskopischen Objektives. *Optik* **52**, 1–18.
- Koops, H. and Bernhard, W. (1978). An objective lens for an electron microscope with compensated axial chromatic aberration. In *Proceedings of the Ninth International Congress on Electron Microscopy*, Toronto (J.M. Sturgess, Ed.), Vol. 1, 36–37 (Microscopical Society of Canada, Toronto).
- Koops, H., Kuck, G. and Scherzer, O. (1977). Erprobung eines elektronenoptischen Achromators. *Optik* **48**, 225–236.
- Krivanek, O. (1994). Three-fold astigmatism in high-resolution transmission electron microscopy *Ultramicroscopy* **55**, 419–433.
- Krivanek, O.L. and Fan, G.Y. (1992a). Complete HREM autotuning using automated diffractogram analysis. *Proc. EMSA*, **50**, Part 1, 96–97.
- Krivanek, O.L. and Fan, G.Y. (1992b). Application of slow-scan charge-coupled device (CCD) cameras to on-line microscope control. *Scanning Microsc.* (Supplement 6), 105–114.
- Krivanek, O.L. and Leber, M.L. (1993). Three-fold astigmatism: An important TEM aberration. In *Proceedings of the 51st Annual Meeting Microscopy Society of America* (G.W. Bailey and C.L. Rieder, Eds.), 972–973 (San Francisco Press, San Francisco).
- Krivanek, O.L. and Leber, M.L. (1994). Autotuning for 1 Å resolution. In *Proceedings of the 13th International Conference on Electron Microscopy*, Paris (B. Joffrey, C. Colliex, J.-P. Chevalier, F. Glas, P.W. Hawkes, D. Hernandez-Verdun, J. Schrevel and D. Thomas, Eds.), Vol. 1, 157–158 (Editions de Physique, Les Ulis).
- Krivanek, O.L. and Stadelmann, P.A. (1995). Effect of three-fold astigmatism on high-resolution electron micrographs. *Ultramicroscopy* **60**, 103–113.
- Krivanek, O.L., Dellby, N., Spence, A.J., Camps, A. and Brown, L.M. (1997a). Aberration correction in the STEM. In *Proceedings of EMAG 1997* (J.M.

- Rodenburg, Ed.), 35–39 (Institute of Physics, Bristol). Krivanek, O.L., Dellby, N., Spence, A.J., Camps, R.A. and Brown, L.M. (1997b). On-line aberration measurement and correction in STEM. *Microsc. Microanal.* **3** (Supplement 2), 1171–1172.
- Krivanek, O.L., Dellby, N. and Brown, L.M. (1998). Spherical aberration corrector for a dedicated STEM. In *Proceedings of EUREM-11, the 11th European Conference on Electron Microscopy*, Dublin 1996 (CESEM, Ed.), Vol. I, 352–353 (CESEM, Brussels).
- Krivanek, O.L., Dellby, N. and Lupini, A.R. (1999a). Towards sub-Å electron beams. *Ultramicroscopy* **78**, 1–11.
- Krivanek, O.L., Dellby, N. and Lupini, A.R. (1999b). STEM without spherical aberration. *Microsc. Microanal.* **5** (Supplement 2), 670–671.
- Krivanek, O.L., Dellby, N. and Lupini, A.R. (2000). Advances in C_s -corrected STEM. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.149–I.150 (Czechoslovak Society for Electron Microscopy, Brno).
- Krivanek, O.L., Dellby, N., Nellist, P.D., Batson, P.E. and Lupini, A.R. (2001). Aberration-corrected STEM: The present and the future. *Microsc. Microanal.* **7** (Supplement 2), 896–897.
- Krivanek, O.L., Dellby, N., Murfitt, M., Nellist, P.D. and Szilagyi, Z. (2002). STEM aberration correction: Where next? *Microsc. Microanal.* **8** (Supplement 2), 20–21.
- Krivanek, O.L., Nellist, P.D., Dellby, N., Murfitt, M.F. and Szilagyi, Z. (2003). Towards sub-0.5 Å beams. *Ultramicroscopy* **96**, 229–237.
- Krivanek, O.L., Corbin, G.J., Dellby, N., Murfitt, M., Nagesha, K., Nellist, P.D. and Szilagyi, Z. (2004). Nion UltraSTEM: a new STEM for sub-0.5 Å imaging and sub-0.5 eV analysis. In *Proceedings of the 13th European Microscopy Congress*, Antwerp (D. van Dyck and P. van Oostveldt, Eds.), Vol. I, 35–36 (Belgian Society for Microscopy, Liège).
- Krivanek, O.L., Dellby, N., McManama-Smith, A., Murfitt, M., Nellist, P.D. and Own, C.S. (2005). An aberration-corrected STEM for diffraction studies. *Microsc. Microanal.* **11** (Supplement 2), 544–545.
- *Krivanek, O.L., Corbin, G.J., Dellby, N., Elston, B.F., Keyse, R.J., Murfitt, M.F., Own, C.S., Szilagyi, Z.S. and Woodruff, J.W. (2007). An electron microscope for the aberration-corrected era. *Ultramicroscopy*.
- Kuck, (1979). *Erprobung eines elektronenoptischen Korrektors für Farb- und Öffnungsfehler*. Dissertation, Darmstadt.
- Lencová, B. (1997). Electrostatic lenses. In *Handbook of Charged Particle Optics* (J. Orloff, Ed.), 177–221 (CRC Press, Boca Raton, FL).
- Lencová, B. and Lenc, M. (1990). Computation of multi-lens focusing systems. *Nucl. Instrum. Methods Phys. Res.* **A298**, 45–55.
- Lencová, B. and Lenc, M. (1994). Computation of properties of electrostatic lenses. *Optik* **97**, 121–126.
- Lencová, B. and Lenc, M. (1997). Third order geometrical and first order chromatic aberrations of electrostatic lenses. *Optik* **105**, 121–128.
- Lentzen, M., Jahnen, B., Jia, C.L., Thust, A., Tillmann, K. and Urban, K. (2002). High-resolution imaging with an aberration-corrected transmission electron microscope. *Ultramicroscopy* **92**, 233–242.
- Liu, H., Munro, E., Rouse, J. and Zhu, X. (2002). Simulation methods for multipole imaging systems and aberration correctors. *Ultramicroscopy* **93**, 271–291.
- Lupini, A.R., Varela, M., Borisevich, A.Y., Travaglini, S.M. and Pennycook, S.J. (2003). Advances in aberration corrected STEM at ORNL. In *Proceedings of EMAG 2003* (S. McVitie and D. McComb, Eds.), 211–214 (Institute of Physics, Bristol).

- Maas, D.J., Henstra, A., Krijn, M.P.C.M. and Mentink, S.A.M. (2000). Electrostatic correction in LV-SEM. *Microsc. Microanal.* **6** (Supplement) 2, 746–747.
- Maas, D.J., Henstra, S., Krijn, M. and Mentink, S. (2001). Electrostatic aberration correction in low-voltage SEM. *Proc. SPIE* **4510**, 205–217.
- Maas, D., Mentink, S. and Henstra, S. (2003). Electrostatic aberration correction in low-voltage SEM. *Microsc. Microanal.* **9** (Supplement 3), 24–25.
- Martínez, G. and Tsuno, K. (2002). BEM simulation of Wien filters. *Ultramicroscopy* **93**, 253–261.
- Martínez, G. and Tsuno, K. (2004). 2nd order aberration corrected Wien filter. In *Proceedings of the 13th European Microscopy Congress*, Antwerp (D. van Dyck and P. van Oostveldt, Eds.), Vol. I, 63–64 (Belgian Society for Microscopy, Liège).
- Matsuda, H., Daimon, H., Kato, M. and Kudo, M. (2005). Approach for simultaneous measurement of two-dimensional angular distribution of charged particles: spherical aberration correction using an ellipsoidal mesh. *Phys. Rev. E* **71**, 066503:1–8.
- Mentink, S.A.M., Steffen, T., Tiemeijer, P.C. and Krijn, M.P.C.M. (1999). Simplified aberration corrector for low-voltage SEM. In *Proceedings of EMAG 1999* (C.J. Kiely, Ed.), 83–86 (Institute of Physics, Bristol).
- Mentink, S.A.M., Zande, M.J. van der, Kok, C. and Rooy, T.L. van (2004). Development of a C_s corrector for a Tecnai 20 FEG STEM/TEM. In *Proceedings of EMAG 2003* (S. McVitie and D. McComb, Eds.), 165–168 (Institute of Physics, Bristol).
- Möllenstedt, G. (1956). Elektronenmikroskopische Bilder mit einem nach O. Scherzer sphärisch korrigierten Objektiv. *Optik* **13**, 209–215.
- Mook, H.W. and Kruit, P. (1998). Fringe field monochromator for high brightness electron sources. In *Proceedings of the 14th International Conference on Electron Microscopy*, Cancún (H.A. Calderón and M.J. Yacamán, Eds.), Vol. 1, 73–74 (Institute of Physics, Bristol).
- Mook, H.W. and Kruit, P. (1999a). On the monochromatisation of high brightness sources for electron microscopy. *Ultramicroscopy* **78**, 43–51.
- Mook, H.W. and Kruit, P. (1999b). Optics and design of the fringe-field monochromator for a Schottky field-emission gun. *Nucl. Instrum. Methods Phys. Res. A* **427**, 109–120.
- Mook, H.W. and Kruit, P. (2000). Construction and characterization of the fringe field monochromator for a field emission gun. *Ultramicroscopy* **81**, 129–139.
- Mook, H.W., Batson, P.E. and Kruit, P. (1999). Operation of the fringe field monochromator in field emission STEM. In *Proceedings of EMAG 1999* (C.J. Kiely, Ed.), 223–226 (Institute of Physics, Bristol).
- Mook, H.W., Batson, P.E. and Kruit, P. (2000). Monochromator for high brightness electron guns. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.315–I.316 (Czechoslovak Society for Electron Microscopy, Brno).
- Mukai, M., Kaneyama, T., Tomita, T., Tsuno, K., Terauchi, M., Tsuda, K., Saitoh, K., Naruse, M., Honda, T. and Tanaka, M. (2003a). ‘MIRAI–21’ electron microscope—performance of the monochromator. *Microsc. Microanal.* **9** (Supplement 2), 954–955.
- Mukai, M., Kaneyama, T., Tomita, T., Tsuno, K., Terauchi, M., Tsuda, K., Saitoh, K., Kersker, M., Naruse, M., Honda, T. and Tanaka, M. (2003b). ‘MIRAI–21’ electron microscope—performance of the monochromator. In *Proceedings of the 15th International Congress on Electron Microscopy*, Durban (R. Cross, J. Engelbrecht and M. Witcomb, Eds.), Vol. 3, 321–322 (Microscopy Society of Southern Africa, Onderstepoort).
- Müller, H., Preikszas, D. and Rose, H. (1999). A beam separator with small aberrations. *J. Electron Microsc.* **48**, 191–204.

- Müller, H., Uhlemann, S. and Haider, M. (2002). Benefits and possibilities for C_c -correction in TEM/STEM. *Microsc. Microanal.* **8** (Supplement 2), 12–13.
- Müller, H., Uhlemann, S., Hartel, P. and Haider, M. (2005). Optical design, simulations and alignment of present-day and future aberration correctors for the transmission electron microscope. *Paul-Scherrer-Institut Proc.* **05–01**, 6.
- *Müller, H., Uhlemann, S., Hartel, P. and Haider, M. (2006). Advancing the hexapole C_s -corrector for the scanning transmission electron microscope. *Microsc. Microanal.* **12**, 442–455.
- Munro, E. (1997). Computational techniques for design of charged particle optical systems. In *Handbook of Charged Particle Optics* (J. Orloff, Ed.), 1–76 (CRC Press, Boca Raton, FL).
- Munro, E. and Wittels, N. (1977). Calculation of the optical properties of electron lenses with electron-transparent foils. *Optik* **47**, 25–38.
- Munro, E., Zhu, X., Rouse, J.A. and Liu, H. (2001). Aberration correction for charged particle lithography. *Proc. SPIE* **4510**, 218–224.
- Nellist, P.D., Dellby, N., Krivanek, O.L., Murfitt, M.F., Szilagy, Z., Lupini, A. R. and Pennycook, S.J. (2004a). Towards sub 0.5 angstrom beams through aberration corrected STEM. In *Proceedings of EMAG 2003* (S. McVitie and D. McComb, Eds.), 159–164 (Institute of Physics, Bristol).
- Nellist, P.D., Chisholm, M.F., Dellby, N., Krivanek, O.L., Murfitt, M.F., Szilagy, Z.S., Lupini, A.R., Borisevich, A., Sides, W.H. and Pennycook, S.J. (2004b). Direct sub-Angstrom imaging of a crystal lattice. *Science* **305**, 1741.
- Nellist, P.D., Chisholm, M.F., Lupini, A.R., Borisevich, A., Sides, W.H., Pennycook, S.J., Dellby, N., Keyse, R., Krivanek, O.L., Murfitt, M.F. and Szilagy, Z.S. (2006). Aberration corrected STEM: current performance and future directions. *J. Phys. Conf.* **26**, 7–12.
- Nomura, S. (2004). Aberration-free electron microscope composed of round lenses. In *Proceedings of the Eighth Asia-Pacific Conference on Electron Microscopy*, Kanazawa (N. Tanaka, Y. Takano, H. Mori, H. Seguchi, S. Iseki, H. Shimada and E. Simamura, Eds.), 34–35 (8APEM Publication Committee, Uchinada, Ishikawa).
- O’Keefe, M.A. (2003). HRTEM at half-Ångström resolution: From 0.5 Å to TEAM. *Microsc. Microanal.* **9** (Supplement 2), 936–937.
- O’Keefe, M.A. (2004). Sub-Ångstrom resolution with aberration-corrected TEM: Present and future. In *Materials Research in an Aberration-free Environment*. *Microsc. Microanal.* **10** (Supplement 3), 68–69.
- Oldfield, L.C. (1973). Computer design of high frequency electron-optical systems. In *Image Processing and Computer-Aided Design in Electron Optics* (P.W. Hawkes, Ed.), 370–399 (Academic Press, London).
- Oldfield, L.C. (1974). The use of microwave cavities as electron lenses. In *Proceedings of the 8th International Congress on Electron Microscopy*, Canberra (J.V. Sanders and D.J. Goodchild, Eds.), Vol. I, 152–153 (Australian Academy of Science, Canberra).
- Orloff, J., Ed. (1997). *Handbook of Charged Particle Optics* (CRC Press, Boca Raton, FL).
- Oshima, C. (1999). Monochromatic field electron emission from a Nb superconductor. *Ultramicroscopy* **78**, 27–32.
- Pejas, W. (1978). Magnetische Abschirmung eines korrigierten Elektronenmikroskops. *Optik* **50**, 61–72.
- Pennycook, S.J., Lupini, A.R., Kadavanich, A., McBride, J.R., Rosenthal, S.J., Puetter, R.C., Yahil, A., Krivanek, O.L., Dellby, N., Nellist, P.D., Duscher, G. and Pantelides, S.T. (2003). Aberration-corrected scanning transmission electron microscopy: The potential for nano- and interface science. *Z. Metall.* **94**, 350–357.

- Plies, E. (1978). Proposal for an electrostatic energy filter and a monochromator. In *Proceedings of the Ninth International Congress on Electron Microscopy*, Toronto (J.M. Sturgess, Ed.), Vol. 1, 50–51 (Microscopical Society of Canada, Toronto).
- Plies, E. and Bärtle, J. (2003). Proposal of a novel highly symmetric Wien filter monochromator. *Microsc. Microanal.* **9** (Supplement 3), 28–29.
- Pöhner, W. (1976). Ein in dritter und fünfter Ordnung sphärisch korrigierter elektronenoptischer Aplanat. *Optik* **45**, 443–454.
- Pöhner, W. (1977). Ein sphärisch korrigierter elektronenoptischer Apochromat. *Optik* **47**, 283–297.
- Prekšzas, D. and Rose H., (1995). Procedures for minimizing the aberrations of electromagnetic compound lenses. *Optik* **100**, 179–187.
- Prekšzas, D. and Rose, H. (1997). Correction properties of electron mirrors. *J. Electron Microsc.* **46**, 1–9.
- Prekšzas, D., Hartel, P., Spehr, R. and Rose, H. (2000). SMART electron optics. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.181–I.184 (Czechoslovak Society for Electron Microscopy, Brno).
- Ramberg, E.G. (1949). Aberration correction with an electron mirror. *J. Appl. Phys.* **20**, 183–186.
- Recknagel, A. (1937). Zur Theorie des Elektronenspiegels. *Z. Phys.* **104**, 381–394.
- Recknagel, A. (1941). Über die sphärische Aberration bei elektronenoptischer Abbildung. *Z. Phys.* **117**, 67–73.
- Reichenbach, M. and Rose, H. (1968/9). Entwurf eines korrigierten magnetischen Objektivs. *Optik* **28**, 475–487.
- Reimer, L., Ed. (1995). *Energy-Filtering Transmission Electron Microscopy* (Springer, Berlin).
- Rempfer, G.F. (1990a). Comment on “A study on hyperbolic mirrors as correctors.” *Optik* **86**, 73–74.
- Rempfer, G.F. (1990b). A theoretical study of the hyperbolic electron mirror as a correcting element for spherical and chromatic aberration in electron optics. *J. Appl. Phys.* **67**, 6027–6040.
- Rempfer, G.F. and Mauck, M.S. (1985). Aberration-correcting properties of the hyperbolic electron mirror. *Proc. Annu. Meeting EMSA* **43**, 132–133.
- Rempfer, G.F. and Mauck, M.S. (1986). An experimental study of the hyperbolic electron mirror. *Proc. Annu. Meeting EMSA* **44**, 886–887.
- Rempfer, G.F. and Mauck, M.S. (1992). Correction of chromatic aberration with an electron mirror. *Optik* **92**, 3–8.
- Rempfer, G.F., Desloge, D.M., Skoczylas, W.P. and Griffith, O.H. (1997). Simultaneous correction of spherical and chromatic aberrations with an electron mirror. *Microsc. Microanal.* **3**, 14–27.
- Rose, H. (1967/8). Über die Korrigierbarkeit von Linsen für schnelle Elektronen. *Optik* **26**, 289–298.
- Rose, H. (1970). Berechnung eines elektronenoptischen Apochromaten. *Optik* **32**, 144–164.
- Rose, H. (1971a). Abbildungseigenschaften sphärisch korrigierter elektronenoptischer Achromate. *Optik* **33**, 1–24.
- Rose, H. (1971b). Elektronenoptische Aplanate. *Optik* **34**, 285–311.
- Rose, H. (1981). Correction of aperture aberrations in magnetic systems with threefold symmetry. *Nucl. Instrum. Methods* **187**, 187–199.
- Rose, H. (1987). The retarding Wien filter as a high-performance imaging filter. *Optik* **77**, 26–34.
- Rose, H. (1990a). Inhomogeneous Wien filter as a corrector for the chromatic and spherical aberration of low-voltage electron microscopes. *Optik* **84**, 91–107.

- Rose, H. (1990b). Outline of a spherically corrected semiaplanatic medium-voltage TEM. *Optik* **85**, 19–24.
- Rose, H. (2000). *High-Performance Electron Microscopes of the Future* (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.21–I.24 (Czechoslovak Society for Electron Microscopy, Brno).
- Rose, H. (2002a). Theory of electron-optical achromats and apochromats. *Ultramicroscopy* **93**, 293–303.
- Rose, H. (2002b). Correction of aberrations—past, present and future. *Microsc. Microanal.* **8** (Supplement 2), 6–7.
- Rose, H. (2002c). Advances in electron optics. In *High-Resolution Imaging and Spectrometry of Materials* (F. Ernst and M. Rühle, Eds.) (Springer, Berlin).
- Rose, H. (2003). Outline of an ultracorrector compensating for all primary chromatic and geometrical aberrations of charged-particle lenses. *Microsc. Microanal.* **9** (Supplement 3), 32–33.
- Rose, H. (2004). Outline of an ultracorrector compensating for all primary chromatic and geometrical aberrations of charged-particle lenses. *Nucl. Instrum. Methods Phys. Res. A* **519**, 12–27.
- Rose, H. (2005). Prospects for aberration-free microscopy. *Ultramicroscopy* **103**, 1–6.
- Rose, H. (2008). *Geometrical Electron Optics*. In preparation.
- Rose, H. and Krahel, D. (1995). Electron optics of imaging energy filters. In *Energy-Filtering Transmission Electron Microscopy* (L. Reimer, Ed.), 43–149 (Springer, Berlin).
- Rose, H. and Pejas, W. (1979). Optimization of imaging magnetic filters free of second-order aberrations. *Optik* **54**, 235–250.
- Rose, H. and Preikszas, D. (1992). Outline of a versatile corrected LEEM. *Optik* **92**, 31–44.
- Rose, H., Hartel, P. and Preikszas, D. (2004). Outline of the mirror corrector for SMART and PEEM3. In *Materials Research in an Aberration-Free Environment*. *Microsc. Microanal.* **10** (Supplement 3), 28–29.
- Sawada, H., Tomita, T., Kaneyama, T., Hosokawa, F., Naruse, M., Honda, T., Hartel, P., Haider, M., Tanaka, N., Hetherington, C.J.D., Doole, R.C., Kirkland, A.I., Hutchison, J.L., Titchmarsh, J.M. and Cockayne, D.J.H. (2004a). C_s corrector for imaging. *Microsc. Microanal.* **10** (Supplement 2), 978–979.
- Sawada, H., Tomita, T., Naruse, M., Honda, T., Hartel, P., Haider, M., Hetherington, C.J.D., Doole, R.C., Kirkland, A.I., Hutchison, J.L., Titchmarsh, J.M. and Cockayne, D.J.H. (2004b). C_s corrector for illumination. *Microsc. Microanal.* **10** (Supplement 2), 1004–1005.
- Sawada, H., Tomita, T., Naruse, M., Honda, T., Hartel, P., Haider, M., Hetherington, C.J.D., Doole, R.C., Kirkland, A.I., Hutchison, J.L., Titchmarsh, J.M. and Cockayne, D.J.H. (2004c). 200 kV TEM with C_s correctors for illumination and imaging. In *Proceedings of the Eighth Asia-Pacific Conference on Electron Microscopy*, Kanazawa (N. Tanaka, Y. Takano, H. Mori, H. Seguchi, S. Iseki, H. Shimada and E. Simamura, Eds.), 20–21 (8APEM Publication Committee, Uchinada, Ishikawa).
- Saxton, W.O. (1995a). Observation of lens aberrations for very high resolution electron microscopy. I: Theory. *J. Microsc.* **179**, 201–213.
- Saxton, W.O. (1995b). Simple prescriptions for measuring three-fold astigmatism. *Ultramicroscopy* **58**, 239–243.
- Saxton, W.O. (2000). A new way of measuring aberrations. *Ultramicroscopy* **81**, 41–45.

- Saxton, W.O., Chand, G. and Kirkland, A.I. (1994). Accurate determination and compensation of lens aberrations in high resolution EM. In *Proceedings of the 13th International Conference on Electron Microscopy*, Paris (B. Jouffrey, C. Colliex, J.-P. Chevalier, F. Glas, P.W. Hawkes, D. Hernandez-Verdun, J. Schrevel and D. Thomas, Eds.), Vol. 1, 203–204 (Editions de Physique, Les Ulis).
- Scherzer, O. (1936). Über einige Fehler von Elektronenlinsen. *Z. Phys.* **101**, 593–603.
- Scherzer, O. (1946). Zur Korrigierbarkeit von Elektronenlinsen. *Phys. Bl.* **2**, 110.
- Scherzer (1947). Sphärische und chromatische Korrektur von Elektronenlinsen. *Optik* **2**, 114–132.
- Scherzer, O. (1948). Ein elektronenoptisches Apochromat. *Z. Naturforsch.* **3a**, 544–545.
- Scherzer, O. (1978). Limitations for the resolving power of electron microscopes. In *Proceedings of the Ninth International Congress on Electron Microscopy*, Toronto (J.M. Sturgess, Ed.), Vol. 3, 123–129 (Microscopical Society of Canada, Toronto).
- Scherzer, O. (1980). Eine sphärisch korrigierte Folien-Linse für die Phasenmikroskopie mit Elektronen. *Optik* **56**, 133–147.
- Schmid, P., Feng, J., Padmore, H., Robin, D., Rose, H., Schlueter, R., Wan, W., Forest, E. and Wu, Y. (2005). Correction and alignment strategies for the beam separator of the photoemission electron microscope 3 (PEEM3). *Rev. Sci. Instrum.* **76**, 023302:1–14.
- Schönhense, G. and Spiecker, H. (2002). Correction of chromatic and spherical aberration in electron microscopy utilizing the time structure of pulsed electron sources. *J. Vac. Sci. Technol.* **B20**, 2526–2534.
- Schönhense, G. and Spiecker, H. (2003). Chromatic and spherical correction using time-dependent acceleration and lens fields. *Microsc. Microanal.* **9** (Supplement 3), 34–35.
- Seeliger, R. (1949). Versuche zur sphärischen Korrektur von Elektronenlinsen mittels nicht rotationssymmetrischer Abbildungselemente. *Optik* **5**, 490–496.
- Seeliger, R. (1951). Die sphärische Korrektur von Elektronenlinsen mittels nicht-rotationssymmetrischer Abbildungselemente. *Optik* **8**, 311–317.
- Septier, A. (1963). Lentille quadrupolaire magneto-électrique corrigée de l'aberration chromatique. Aberration d'ouverture de ce type de lentilles. *C. R. Acad. Sci. Paris* **256**, 2325–2328.
- Septier, A. (1966). The struggle to overcome spherical aberration in electron optics. *Adv. Opt. Electron Microsc.* **1**, 204–274.
- Shao, Z. and Wu, X.D. (1989). Adjustable four-electrode electron mirror as an aberration corrector. *Appl. Phys. Lett.* **55**, 2696–2697.
- Shao, Z. and Wu, X.D. (1990a). Properties of a four-electrode adjustable electron mirror as an aberration corrector. *Rev. Sci. Instrum.* **61**, 1230–1235.
- Shao, Z. and Wu, X.D. (1990b). A study on hyperbolic mirrors as correctors. *Optik* **84**, 51–54.
- Skoczymas, W.P., Rempfer, G.F. and Griffith, O.H. (1991). A proposed modular imaging system for photoelectron and electron probe microscopy with aberration correction, and for mirror microscopy and low energy electron microscopy. *Ultramicroscopy* **36**, 252–261.
- Skoczymas, W.P., Rempfer, G.F. and Griffith, O.H. (1994). Electron optical benches for in-line and branched systems. A new bench designed for mirror-based aberration correction and low energy electron microscopy. *Rev. Sci. Instrum.* **65**, 3183–3193.
- Steffen, T., Tiemeijer, P.C., Krijn, M.P.C.M. and Mentink, S.A.M. (2000). Correction of chromatic and spherical aberration using a Wien filter. In *Proceedings*

- of the 12th European Congress on Electron Microscopy (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.151–I.152 (Czechoslovak Society for Electron Microscopy, Brno).
- Stoyanov, P.A. (1955). [The effect of departures of the geometrical shape of the polepieces of an objective from circular symmetry on the resolving power of the electron microscope]. *Zh. Tekh. Fiz.* **25**, 625–635.
- Sturrock, P.A. (1950). The aberrations of magnetic electron lenses due to asymmetries. *Proceedings of the [First] Conference on Electron Microscopy*, Delft, 1949 (A.L. Houwink, J.B. Le Poole and W.A. Le Rütte, Eds.), 89–93 (Hoogland, Delft).
- Sturrock, P.A. (1951). The aberrations of magnetic electron lenses due to asymmetries. *Phil. Trans. Roy. Soc. London A* **243**, 387–429.
- Tanaka, M., Terauchi, M., Tsuda, K., Saitoh, K., Honda, T., Tsuno, K., Naruse, M., Tomita, T. and Kaneyama, T. (2000). An 0.2 eV energy resolution analytical electron microscope. In *Microbeam Analysis 2000* (D.B. Williams and R. Shimizu, Eds.), 217–218 (Institute of Physics, Bristol).
- Tanaka, M., Terauchi, M., Tsuda, K., Saitoh, K., Mukai, M., Kaneyama, T., Tomita, T., Tsuno, K., Kersker, M., Naruse, M. and Honda, T. (2002). Development of an 0.2 eV energy resolution analytical electron microscope. *Microsc. Microanal.* **8** (Supplement 2), 68–69.
- Tang, T.-t. (1986). Aberration analysis of a crossed-field analyzer. *Optik* **74**, 51–56.
- Tang, T.-t., Orloff, J. and Wang, L. (1996a). Modeling and design of space charge lenses/aberration correctors for focused ion beam systems. *J. Vac. Sci. Technol.* **B14**, 80–84.
- Tang, T.-t., Orloff, J. and Wang, L. (1996b). Monte-Carlo simulation on discrete space charge effects of space charge lenses. *Optik* **102**, 141–146.
- Terauchi, M. and Tanaka, M. (2000). High energy-resolution electron energy-loss spectroscopy based on electron microscopy. In *Microbeam Analysis 2000* (D.B. Williams and R. Shimizu, Eds.), 211–212 (Institute of Physics, Bristol).
- Thomson, M.G.R. (1972). The primary aberrations of a quadrupole corrector system. *Optik* **34**, 528–534.
- Tiemeijer, P.C. (1999a). Operation modes of a TEM monochromator. In *Proceedings of EMAG 1999* (C.J. Kiely, Ed.), 191–194 (Institute of Physics, Bristol).
- Tiemeijer, P.C. (1999b). Measurement of Coulomb interactions in an electron beam monochromator. *Ultramicroscopy* **78**, 53–62.
- Tiemeijer, P.C., Lin, J.H.A. van, Freitag, B.H. and Jong, A.F. de (2002). Monochromized 200 kV (S)TEM. *Microsc. Microanal.* **8** (Supplement 2), 70–71.
- Titchmarsh, J.M., Cockayne, D.J.H., Doole, R.C., Hetherington, C.J.D., Hutchison, J.L., Kirkland, A.I. and Sawada, H. (2004). A versatile double aberration-corrected, energy-filtered HREM/STEM for materials science. In *Proceedings of the 13th European Microscopy Congress*, Antwerp (D. van Dyck and P. van Oostveldt, Eds.), Vol. I, 27–28 (Belgian Society for Microscopy, Liège).
- Tretner, W. (1959). Existenzbereiche rotationssymmetrischer Elektronenlinse. *Optik* **16**, 155–184.
- Tsai, F.C. (2000). Studies of a magnetically focused electrostatic mirror. II. Aberration corrections. *J. Microsc.* **197**, 118–135.
- Tsuno, K. (1991). Aberration analysis of a Wien filter for electrons. *Optik* **89**, 31–40.
- Tsuno, K. (1992). Electron optical analysis of a high-resolution electron energy loss spectrometer with a retarding Wien filter. *Rev. Sci. Instrum.* **63**, 4112–4121.

- Tsuno, K. (1993). Electron optical analysis of a retarding Wien filter for electron spectroscopic imaging. *Rev. Sci. Instrum.* **64**, 659–666.
- Tsuno, K. (1997). Magnetic lenses for electron microscopy. In *Handbook of Charged Particle Optics* (J. Orloff, Ed.), 143–175 (CRC Press, Boca Raton, FL).
- Tsuno, K. (1999). Optical design of electron microscope lenses and energy filters. *J. Electron Microsc.* **48**, 801–820.
- Tsuno, K. (2001). “S filter,” a compact, high dispersion in-column energy filter. *Proc. SPIE* **4510**, 71–79.
- Tsuno, K. and Rouse, J. (1996). Simulation of electron trajectories of Wien filter for high-resolution EELS installed in TEM. *J. Electron Microsc.* **45**, 417–427.
- Tsuno, K., Terauchi, M. and Tanaka, M. (1988–9). Electron trajectory calculation of a stigmatic-focus Wien filter for electron energy loss spectroscopy I, II, III. *Optik* **78**, 71–77; **80**, 149–154; **83**, 77–84.
- Tsuno, K., Terauchi, M. and Tanaka, M. (1990). Design and performance of a retarding Wien filter installed in a TEM for EELS. In *Proceedings of EMAG–MICRO 89* (P.J. Goodhew and H.Y. Elder, Eds.), Vol. 1, 71–74 (Institute of Physics, Bristol).
- Tsuno, K., Ioanoviciu, D. and Martínez, G. (2003). Aberration corrected Wien filter as a monochromator of high spatial and high energy resolution electron microscopes. *Microsc. Microanal.* **9** (Supplement 2), 944–945.
- Tsuno, K., Ioanoviciu, D. and Martínez, G. (2005). Third-order aberration theory of Wien filters for monochromators and aberration correctors. *J. Microsc.* **217**, 205–215.
- Turnbull, W. (2004). Direct spherical and chromatic aberration correction for charged particle optical systems. *J. Vac. Sci. Technol.* **B22**, 3560–3564.
- Typke, D. (1968/9). Die Öffnungsfehler von rotationssymmetrischen Elektronenlinsen mit Raumladung. *Optik* **28**, 488–494.
- Typke, D. (1972). Sphärische und chromatische Korrektur von Elektronenlinsen mit Hilfe von Influenzladen, I, II. *Optik* **34**, 573–594 and **36**, 124–138.
- Typke, D. (1976). Sphärische und chromatische Korrektur von Elektronenlinsen mit Hilfe von Influenzladen, erratum. *Optik* **44**, 509.
- Uhlemann, S., Haider, M., Schwan, E. and Rose, H. (1998). Towards a resolution enhancement in the corrected TEM. In *Proceedings of EUREM-11, the 11th European Conference on Electron Microscopy*, Dublin 1996 (CESEM, Ed.), Vol. I, 365–366 (CESEM, Brussels).
- Uno, S., Honda, K., Nakamura, N., Matsuya, M., Achard, B. and Zach, J. (2004a). An automated aberration correction method in scanning electron microscopes. In *Proceedings of the 13th European Microscopy Congress*, Antwerp (D. van Dyck and P. van Oostveldt, Eds.), Vol. I, 37–38 (Belgian Society for Microscopy, Liège).
- Uno, S., Honda, K., Nakamura, N., Matsuya, M. and Zach, J. (2004b). Probe shape extraction and automatic aberration correction in scanning electron microscopes. In *Proceedings of the Eighth Asia–Pacific Conference on Electron Microscopy*, Kanazawa (N. Tanaka, Y. Takano, H. Mori, H. Seguchi, S. Iseki, H. Shimada and E. Simamura, Eds.), 46–47 (8APEM Publication Committee, Uchinada, Ishikawa).
- Urban, K., Kabius, B., Haider, M. and Rose, H. (1999). A way to higher resolution: Spherical-aberration correction in a 200 kV transmission electron microscope. *J. Electron Microsc.* **48**, 821–826.
- Wan, W., Feng, J., Padmore, H.A. and Robin, D.S. (2004). Simulation of a mirror corrector for PEEM. *Nucl. Instrum. Methods Phys. Res.* **A519**, 222–229.
- Wang, L., Orloff, J. and Tang, T.-t. (1995). Study of space-charge devices for focused ion beam systems. *J. Vac. Sci. Technol.* **B13**, 2414–2418.

- Weißbäcker, C. and Rose, H. (2000). Electrostatic correction of the chromatic and spherical aberration of charged particle lenses. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.157–I.158 (Czechoslovak Society for Electron Microscopy, Brno).
- Weißbäcker, C. and Rose, H. (2001). Electrostatic correction of the chromatic and the spherical aberration of charged-particle lenses, I. *J. Electron Microsc.* **50**, 383–390.
- Weißbäcker, C. and Rose, H. (2002). Electrostatic correction of the chromatic and the spherical aberration of charged-particle lenses, II. *J. Electron Microsc.* **51**, 45–51.
- Wichtendahl, R., Fink, R., Kühlenbeck, H., Preikszas, D., Rose, H., Spehr, R., Hartel, P., Engel, W., Schlögl, R., Freund, H.-J., Bradshaw, A.M., Lilienkamp, G., Schmidt, T., Bauer, E., Benner, G. and Umbach, E. (1998). SMART: An aberration-corrected XPEEM/LEEM with energy filter. *Surface Rev. Lett.* **5**, 1249–1256.
- Wu, Y.K., Robin, D.S., Forest, E., Schleuter, R., Anders, S., Feng, J., Padmore, H. and Wei, D.H. (2004). Design and analysis of beam separator magnets for third generation aberration compensated PEEMs. *Nucl. Instrum. Methods Phys. Res. A* **519**, 230–241.
- Ximen, J.-y. (1983). The aberration theory of a combined magnetic round lens and sextupoles system. *Optik* **65**, 295–309.
- Ximen, J.-y. and Crewe, A.V. (1985). Correction of spherical and coma aberrations with a sextupole–round lens–sextupole system. *Optik* **69**, 141–146.
- Yavor, M.I. (1993). Methods for calculation of parasitic aberrations and machining tolerances in electron optical systems. *Adv. Electron. Electron Phys.* **86**, 225–281.
- Zach, J. (1989). Design of a high-resolution low-voltage scanning electron microscope. *Optik* **83**, 30–40.
- Zach, J. (2000). Aspects of aberration correction in LVSEM. In *Proceedings of the 12th European Congress on Electron Microscopy*, Brno (P. Tomanek and R. Kolarik, Eds.), Vol. III, I.169–I.172 (Czechoslovak Society for Electron Microscopy, Brno).
- Zach, J. and Haider, M. (1995a). Correction of spherical and chromatic aberration in a low-voltage SEM. *Optik* **98**, 112–118.
- Zach, J. and Haider, M. (1995b). Aberration correction in a low voltage SEM by a multipole corrector. *Nucl. Instrum. Methods Phys. Res. A* **363**, 316–325.
- Zemlin, F. (1979). A practical procedure for alignment of a high resolution electron microscope. *Ultramicroscopy* **4**, 241.
- Zemlin, F., Weiss, K., Schiske, P., Kunath, W. and Herrmann, K.-H. (1978). Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms. *Ultramicroscopy* **3**, 49–60.
- Zworykin, V.K., Morton, G.A., Ramberg, E.G., Hillier, J. and Vance, A.W. (1945). *Electron Optics and the Electron Microscope* (Wiley, New York and Chapman & Hall, London).

*References added since the first printing.

Part II

IMAGING WITH PHOTONS

Two-Photon Excitation Fluorescence Microscopy

Alberto Diaspro, Marc Schneider, Paolo Bianchini,
Valentina Caorsi, Davide Mazza, Mattia Pesce,
Ilaria Testa, Giuseppe Vicidomini, and Cesare Usai

1 Introduction

Two-photon excitation (TPE) fluorescence microscopy (Denk et al., 1990; Pennisi, 1997; Esposito et al., 2004) can be considered an important example of the continuing growth of interest in optical microscopy (Diaspro, 1996; Koster and Klumperman, 2003). In spite of its low spatial resolution compared to other modern imaging techniques, such as scanning near-field microscopy (Dürig and Pohl, 1986), scanning probe microscopy (Binnig et al., 1986), or electron microscopy (Ruska and Knoll, 1931), light microscopy techniques, including TPE microscopy, have unique capabilities in the investigation of biological structures in a hydrated state, in living specimens, or at least under conditions that are close to physiological states (Pawley, 1995; Periasamy, 2001; Diaspro, 2002). This fact, coupled with advances in fluorescence labeling, permits the study of the complex and delicate relationships existing between structure and function in the four-dimension (x - y - z - t) biological systems domain (Arndt-Jovin et al., 1985; Beltrame et al., 1985; Wang and Herman, 1996; Herman and Tanke, 1998). As well, the advances achieved in the field of biological markers, especially the design of application-suited chromophores, the development of the so-called quantum dots (Jaiswal et al., 2004), visible fluorescent proteins (VFPs) from the green fluorescent protein (GFP) and its natural homologues to specifically engineered variants of these molecules (Patterson and Lippincott-Schwarz, 2002; Wiedenmann et al., 2004), and the improvements in resolution by means of special optical schemes (Egner et al., 2004; Gugel et al., 2004), are enabling TPE to move from microscopy to nanoscopy (Hell, 2003; Bastiaens and Hell, 2004).

There is also ongoing research to use TPE in new fields where its special features can be advantageously applied to improve and to optimize existing schemes (McConnell and Riis, 2004). This covers new online detection systems such as endoscopic imaging based on gradient refractive index fibers (Jung et al., 2004), the development of new substrates with higher fluorescence output (Kappel et al., 2004), as well as the use of TPE to systematically cross-link protein matrices and

control the diffusion (Basu and Campagnola, 2004). Furthermore, the combination of TPE applications with other techniques has great potential use (Bird et al., 2004; Nemet et al., 2004; Periasamy and Diaspro, 2004).

TPE microscopy belongs to the category of three-dimensional (3D) optical microscopy methods, which have been widespread since the 1970s (Weinstein and Castleman, 1971; Agard et al., 1989; Bianco and Diaspro, 1989; Diaspro et al., 1990; Brakenhoff et al., 1979; Sheppard and Wilson, 1980; Wilson and Sheppard, 1984; Carlsson et al., 1985; Shotton, 1993). During the past 10 years, confocal microscopes have proved to be extremely useful research tools, notably in the life sciences. The evolution has also brought optical microscopy from 3D ($x-y-z$) to 5D ($x-y-z-t-\lambda$) analysis allowing researchers to probe even deeper into the intricate mechanisms of living systems (Cheng, 1994; Pawley, 1995; Masters, 1996; Sheppard and Shotton, 1997; Periasamy, 2001; Diaspro, 2002). Here, TPE microscopy (Denk et al., 1990; Diaspro, 1999a-c), or more generally multiphoton excitation (MPE) microscopy (König, 2000; Gratton et al., 2001), can probably be considered the most relevant advance in fluorescence optical microscopy since the introduction of confocal imaging in the 1980s (Wilson and Sheppard, 1984; White et al., 1987; Pawley, 1995; Webb, 1996; Sheppard and Shotton, 1997; Diaspro, 2002; Amos, 2000).

TPE microscopy couples a 3D intrinsic ability, shared with confocal microscopy, with almost five other interesting properties. First, TPE greatly reduces photo interactions and allows imaging of living specimens over long time periods. Second, it allows operation in a high-sensitivity background-free acquisition scheme. Third, TPE microscopy can penetrate turbid and thick specimens down to a depth of a few hundred micrometers. Fourth, due to the distinct character of the multiphoton absorption spectra of many of the fluorophores, TPE allows simultaneous excitation of different fluorescent molecules reducing colocalization errors. Fifth, TPE can prime photochemical reactions within a subfemtoliter volume inside solutions, cells, and tissues.

Furthermore, this form of nonlinear microscopy also favored the development and application of several investigative techniques starting from TPE microscopy (Denk et al., 1990), namely, three-photon excited fluorescence (Hell et al., 1996; Maiti et al., 1997), second harmonic generation (Gannaway and Sheppard, 1978; Gauderon et al., 1999; Campagnola et al., 1999; Zoumi et al., 2002), third-harmonic generation (Mueller et al., 1998; Squier et al., 1998), fluorescence correlation spectroscopy (Berland et al., 1995; Schwille et al., 1999, 2000; Schwille, 2001; Heinze et al., 2004; Ruan et al., 2004), image correlation spectroscopy (Wiseman et al., 2000, 2002), lifetime imaging (König et al., 1996; French et al., 1997; Sytsma et al., 1998; Straub and Hell, 1998), single molecule detection schemes (Mertz et al., 1995; Xie and Lu, 1999; Sonleitner et al., 1999, 2000; Cannone et al., 2003b; Chirico et al., 2001, 2003b), photodynamic therapies (Bhawalkar et al., 1997), two-photon photoactivation and photoswitching of visible fluorescent proteins (Chirico et al., 2004, 2005; Post et al., 2005; Schneider et al., 2005), and others (White and Errington, 2000; Masters, 2002; Periasamy,

2001; Periasamy and Diaspro, 2003; Cahalan et al., 2003; Miller et al., 2003; Diaspro, 1998, 2004; Piston, 2005, Cruz and Luscher, 2005).

2 Brief Chronological Notes

The important work done by Abbe indicated how to optimize the optical microscope. In 2005 the “Focus on Microscopy” conference, held in Jena, was dedicated to Abbe’s work 100 hundred years after his death (www.focusonmicroscopy.org). Abbe’s approach in defining factors influencing the microscope’s resolution was fundamental. Confocal and TPE microscopy show how to extend optical parameters to obtain a better resolution. In 2000 the Optical Society of America honored Paul Davidovits, M. David Egger, and Marvin Minsky with the R.W. Wood Prize for “seminal contributions to confocal microscopy.” The fact is that in 1975 Minsky invented a confocal microscope identical in the concept to the one developed by Egger and Davidovits at Yale (Davidovits and Egger, 1969, 1971), by Sheppard and Wilson at Oxford (Sheppard and Choudhury, 1977; Sheppard and Wilson, 1980; Wilson and Sheppard, 1984), and by Brakenhoff and colleagues in Amsterdam (Brakenhoff et al., 1979, 1989). As reported by Minsky (1988), the circumstances are also remarkable in that Minsky only published his invention as a patent (Figure 11–1). In addition, the idea for a confocal microscope was previously presented by Naora, who built an optical setup based upon a concept of Koana, as recently indicated by Guy Cox. It was not until the end of the 1970s, with the advent of affordable computers and lasers, and the development of digital image processing software, that the first single-beam confocal laser scanning microscopes became available in a number of laboratories and were applied to biological and materials specimens. A new revolution was developing (Sheppard and Kompfner, 1978): TPE second harmonic and fluorescence microscopy. The TPE story dates back to 1931, having its roots in the theory originally developed by Maria Göppert-Mayer (1931). The keystone of TPE theory lies in the prediction that one atom or molecule can simultaneously absorb two photons in the same quantum event.

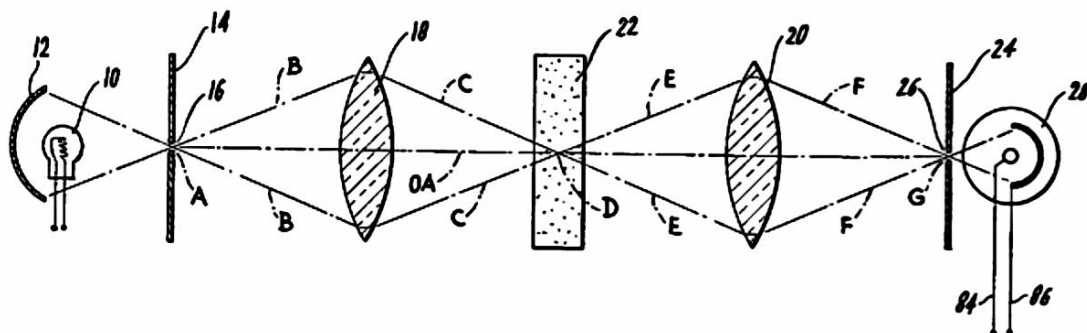


Figure 11–1. Historical sketch of the confocal setup as reported by Marvin Minsky in his patent (U.S. patent 3013467: Microscopy Apparatus, filed 7 November 1957).

Now, to indicate the rarity of the event, consider that “simultaneously” here implies “within a temporal window of 10^{-16} – 10^{-15} s”: in bright daylight a good one- or two-photon excitable fluorescent molecule absorbs a photon through a one-photon interaction about once a second and a photon pair by two-photon simultaneous interaction every 10 million years (Denk and Svoboda, 1997). To increase the probability of the event, a very high density of photons is needed, i.e., a laser source. As in confocal microscopy, the laser is the key to the development and dissemination of the technique. In fact, it was only in the 1960s, after the development of the first laser sources (Svelto, 1998; Wise, 1999), that it was possible to find experimental evidence for Maria Göppert-Mayer’s prediction. Kaiser and Garret (1961) reported TPE of fluorescence in $\text{CaF}_2:\text{Eu}^{2+}$ and Singh and Bradley (1964) were able to estimate the three-photon absorption cross-section for naphthalene crystals. These two results consolidated other related experimental achievements obtained by Franken et al. (1961) of second harmonic generation in a quartz crystal using a ruby laser. Later, Rentzepis and colleagues (1970) observed three-photon excited fluorescence from organic dyes, and Hellwarth and Christensen (1974) collected second-harmonic generation signals from ZnSe polycrystals at a microscopic level. In 1976, Berns reported a probable two-photon effect as a result of focusing an intense pulsed laser beam onto chromosomes of living cells, and such interactions form the basis of modern nanosurgery (König et al., 1999). However, the original idea of generating 3D microscopic images by means of such nonlinear optical effects was first suggested and attempted in the 1970s by Sheppard, Kompfner, Gannaway, and Choudhury of the Oxford group (Sheppard et al., 1977; Gannaway and Sheppard, 1978; Sheppard and Kompfner, 1978). It should also be emphasized that for many years the application of two-photon absorption was mainly related to spectroscopic studies (Friedrich and McClain, 1980; Friedrich, 1982; Birge, 1986; Callis, 1997). The real “TPE boom” took place at the beginning of the 1990s at the W.W. Webb laboratories (Cornell University, Ithaca, NY). However, it was the excellent and effective work done by Winfried Denk and colleagues (Denk et al., 1990) that was responsible for spreading the technique and that revolutionized fluorescence microscopy imaging.

3 Basic Principles on Confocal and Two-Photon Excitation of Fluorescent Molecules

3.1 Fluorescence

Fluorescence optical microscopy is very popular for imaging in biology since fluorescence is highly specific either as exogenous labeling or endogenous autofluorescence (Beltrame et al., 1985; Arndt-Jovin et al., 1985; Periasamy, 2001). Fluorescent molecules allow both spatial and functional information to be obtained through specific absorption, emission, lifetime, anisotropy, photodecay, diffusion, and other contrast mechanisms (Diaspro, 2002; Zoumi et al., 2002). This means that it is possible to efficiently study, for example, the distribution and

dynamics of proteins, DNA, and chromatin as well as ion concentration, voltage, and temperature within living cells (Chance, 1989; Tsien, 1995; Robinson, 2001). TPE of fluorescent molecules is a nonlinear process related to the simultaneous absorption of two photons whose total energy equals the energy required for conventional, one-photon excitation (Birks, 1970; Denk et al., 1995; Callis, 1997). In any case the energy required to prime fluorescence is the energy sufficient to produce a molecular transition to an excited electronic state. The excited fluorescent molecules then decay to an intermediate state giving off a photon of light having an energy lower than needed to prime excitation. This means that the energy E provided by photons should equal the molecule energy gap ΔE_g , and, considering the relationship between photon energy E and radiation wavelength λ , it follows that

$$\Delta E_g = E = \frac{hc}{\lambda} \quad (1)$$

where $h = 6.6 \times 10^{-34} \text{ J} \cdot \text{s}$ is Planck's constant and $c = 3 \times 10^8 \text{ ms}^{-1}$ is the value of the speed of light (considered in vacuum and to a reasonable approximation). Conventional techniques for fluorescence excitation use ultraviolet (UV) or visible radiation and excitation occurs when the absorbed photons are able to match the energy gap to the ground from the excited state. Due to energetic aspects, the fluorescence emission is shifted toward a wavelength longer than the one used for excitation. This shift typically ranges from 50 to 200 nm (Birks, 1970; Cantor and Schimmel, 1980). For example, a fluorescent molecule that absorbs one photon at 340 nm, in the ultraviolet region, exhibits fluorescence at 420 nm in the blue region.

A 3D reconstruction of the distribution of fluorescence within a three-dimensional object such as a living cell starting with the acquisition of the two-dimensional distribution of specific intensive properties is one of the most powerful properties of the optical microscope. In fact, this allows complete morphological analysis through volume rendering procedures (Kriete, 1992; Robinson, 2001) of living biological specimens, where the opportunity of optical slicing allows information to be obtained from different planes of the specimen without being invasive, thus preserving the structures and functionality of the different parts (Weinstein and Castleman, 1971; Agard, 1984; Agard et al., 1989; Diaspro et al., 1990).

3.2 Confocal Principle and Laser Scanning Microscopy

Conventional wide-field microscopes involve a specimen entirely bathed in the radiation from the light source, viewed directly by eye or through any capture device [charge coupled device (CCD) camera, for instance or photosensitive film]. As reported in the paper by Minsky (1961), an ideal microscope would examine each point of the specimen and measure the amount of light scattered, absorbed, or emitted by that point, excluding contributions from another part of the sample from the actual or from adjacent planes (Figure 11-2). Unfortunately, in trying to obtain images by making many such measurements at the

same time, every focal image point will be clouded by aberrant rays of scattered light deflected from points of the specimen that are not the point of interest. This means that samples undergo continuous full excitation, leading to in- and out-of-focus light points and contributing to overlapping and worsening axial resolution and producing that typical haziness in the collected images that, together with light-diffraction effects, limits the performance of the instrument. Most of those extra rays would be absent if we could illuminate only one specimen point at a time. There is no way of eliminating every undesired ray, because of multiple scattering, but it is comparatively straightforward to remove all rays not initially aimed at the focal point by using a sort of second microscope (instead of a condenser lens) to image a pinhole aperture (a small aperture in an opaque screen) on a single point of the specimen. This reduces the amount of light in the specimen by orders of magnitude without reducing the focal brightness. Even then, some of the initially focused light will be scattered by out-of-focus specimen points onto other points in the image plane affecting the clarity of the final acquisition, i.e., of the observed image *o*. But it is possible to reject undesired rays as well, by placing a second pinhole aperture in the image plane that lies beyond the exit side of the objective lens. We end up with an elegant, symmetrical geometry: a pinhole and an objective lens on each side of the specimen. This leads to the use of two lenses on both the excitation and detection sides of the microscope, combining the two lenses for a unique effect (Figure 11–3).

To acquire an image the excitation light has to be fully delivered to each point of the sample and the emission signal collected and displayed. This is usually accomplished by either of two possible but different strategies.

The first one is based upon scanning the sample in a raster pattern such that over every fixed period of time, the necessary amount of information from the focal plane is collected and the emitted light signal, usually detected through a photomultiplier tube (PMT), is displayed by a mapping of each single point light emission. Sometimes, the use of a slit moving in one direction (rather than a single point) is

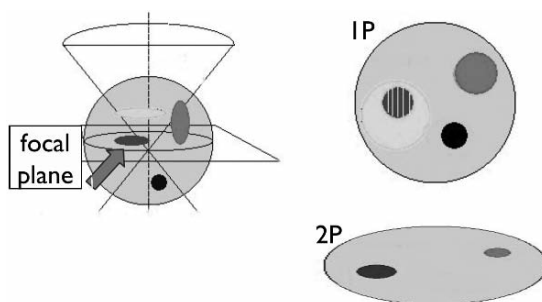


Figure 11–2. Comparison between conventional (1P) and two-photon (2P) excitation with respect to image formation. When focusing on the actual focal plane under 1P, a contribution from adjacent planes that are physically excluded in the 2P process is obtained, as happens in a confocal setup. (From Giuseppe Vicidomini, LAMBS, MicroScoBio, University of Genoa.) (See color plate.)

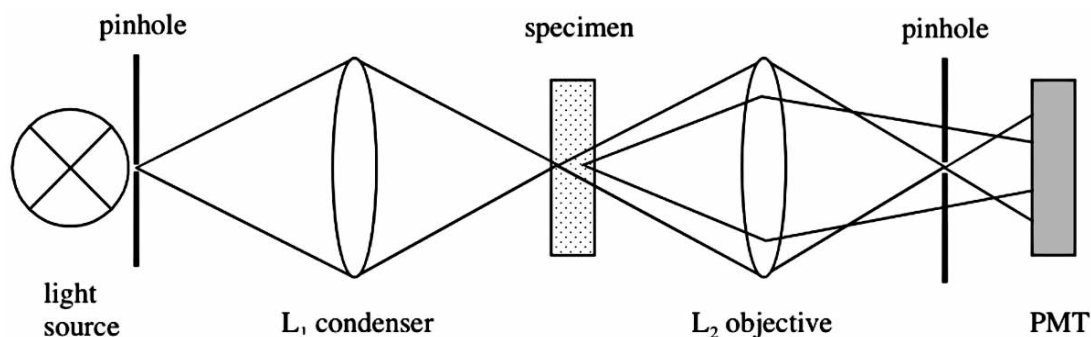


Figure 11-3. Equivalent optical configuration for a confocal setup (see text).

preferred for speeding up the scanning rate, although this leads to an evident worsening of the spatial resolution and of the three-dimensional imaging capability.

A second possible approach to form confocal images consists of employing a multipinhole Nipkow spinning disk (Petran et al., 1968; Kino and Corle, 1989). This is a disk containing multiple sets of spirally arranged holes placed in the image plane of the objective lens. A large parallel beam of light is then pointed at a particular region of the disk and the light passing through the illuminated pinholes is focused by the objective lens straight onto the specimen. When spinning the disk at a rapid rate, the sample may undergo excitation several hundred times per second: emitted light is collected and imaged typically by a high-resolution and high-quantum-efficiency CCD camera. Concerning optical sectioning, every architecture is built such that the sample is placed along the light path at a conjugate focal plane and the movements along the optical axis keep the focus at a fixed distance from the objective, making it possible to effectively scan different fields of view through the specimen (due to a step-by-step motor device attached to the fine focus) and collect a series of in-focus optical slices for 3D reconstruction.

The degree of confocality is a function of the pinhole size: the use of smaller pinholes improves the discrimination of focused light from stray light, thus involving a thinner plane in the image formation process and improving resolution, at the cost of lower light throughput, which makes things complicated when dealing with particularly dim samples.

In these architectures, z -resolution and optical sectioning thickness (which are basically the parameters involved in every optical sectioning process) depend on a number of factors such as the numerical aperture (NA) of the objective lens, the wavelength of the excitation/emission light, the pinhole size, the refractive index of components along the light path, and finally the overall alignment of the instrument.

3.3 Theoretical Analysis

The development of an effective theoretical model for describing the properties of an optical system needs some preliminary, *realistic* assumptions to be made to simplify calculations.

From this point of view, a linear space invariant (LSI) model is a good choice, pliable enough to obtain important insights and develop suitable mathematical tools for the analysis of most concrete situations. Let us consider the situation sketched in Figure 11–3, showing a typical confocal setup.

A monochromatic point light source is focused onto a sample in the focal plane through a lens L_1 (condenser) and the emitted radiation from the sample (which is also supposed to be monochromatic) is collected through a second lens L_2 (objective) by a point detector. Let h_{ex} and h_{em} be, respectively, the impulse response of L_1 and L_2 , i.e., the lens response to a point-like light source. Under this hypothesis it can be written that $U_{\text{ex}}(x) = (h_{\text{ex}} \otimes \delta_s)(x) = h_{\text{ex}}(x)$, where the excitation light source is modeled by a Dirac impulse.

It can be shown that with U_{ex} being the signal reaching the sample, the emitted signal scales with the fluorescent dye density D according to $U_{\text{em}} = D \cdot U_{\text{ex}}$. The emitted radiation is then focused on the point detector through L_2 .

This leads to $U_{\text{det}}(x) = [(h_{\text{em}} \otimes U_{\text{em}}) \cdot \delta_d](x)$ where the point detector function is assumed to be a Dirac impulse. The overall signal collected by the detector is

$$\begin{aligned} I_{\text{tot}} &= \int U_{\text{det}}(x) dx = \int dx \delta_d(x) (h_{\text{em}} \otimes U_{\text{em}})(x) \\ &= \int dx \delta_d(x) \int dy h_{\text{em}}(x - y) D(y) h_{\text{ex}}(y) \\ &= \int dy D(y) h_{\text{ex}}(y) \int dx \delta_d(x) h_{\text{em}}(x - y) \\ &= \int dy D(y) h_{\text{ex}}(y) h_{\text{em}}(-y) \end{aligned}$$

If we now limit ourselves to a point-like sample:

$$\int U_{\text{det}}(x) dx = \int dy \delta(y) h_{\text{ex}}(y) h_{\text{em}}(-y) = h_{\text{ex}}(0) h_{\text{em}}(0)$$

where $h_{\text{ex}} = h_{\text{em}}$ under the hypothesis of $L_1 = L_2$ and $\lambda_{\text{ex}} = \lambda_{\text{em}}$.¹

Since an x - y - z scanning process is generally coupled to the imaging one, it is natural to write, for a general point $P(x, y, z)$: $I_{\text{tot}} = h^2(x, y, z)$, which is the general expression for the point spread function (PSF), that is, the system impulse response. A mathematical expression for $h(x, y, z)$ can be obtained through the scalar electromagnetic waves theory.

The formulation, based on Fraunhofer diffraction, leads to

$$h(u, v) \propto \left| \int_0^1 J_0(v\rho) e^{-0.5i u \rho^2} \rho d\rho \right|^2 \quad (2)$$

Where u and v are suitable dimensionless variables defined according to the following:

$$\begin{aligned} u &\propto z \\ v &\propto \sqrt{x^2 + y^2} \end{aligned}$$

¹ The equivalence of $h_{\text{ex}} = h_{\text{em}}$ is valid only for pinhole sizes ≤ 0.25 AU. AU is the so-called Airy unit, which represents the diameter of the Airy disk.

By limiting the discussion to the points along the optical axis and in the focal plane, we have:

$$h(0, v) \propto \left[\frac{2J_1(v)}{v} \right]^2 \quad h(u, 0) \propto \left[\frac{\sin(u/4)}{u/4} \right]^2$$

that is

$$I_{\text{tot}}(0, v) \propto \left[\frac{2J_1(v)}{v} \right]^4 \quad I_{\text{tot}}(u, 0) = \left[\frac{\sin(u/4)}{u/4} \right]^4$$

Compared to a conventional microscope, where $I_{\text{tot}} \approx h(u, v)$, the calculation of the fall width at half-maximum (FWHM), representing the system resolution, leads to an improvement in resolution by a factor 1.4 (Brakenhoff et al., 1979; Wilson and Sheppard, 1984; Diaspro et al., 1999a; Jonkman and Stelzer, 2002; Torok and Sheppard, 2002).

3.4 Remarks and Comments

Comparisons between the ideal PSF in the case of strict confocality with that of conventional microscopes account for improvements in resolution. However, despite the use of this mathematical formalism for concrete situations, further drawbacks need to be highlighted. First, there is a natural relation between the pinhole size and the PSF: the more the pinhole size is increased, the more the confocal microscope response tends to fit conventional responses.

This means that in the case of dim or highly photosensitive specimens some compromise has to be found between the resolution and the amount of the collected signals, according to the kind of analysis to be performed (whether a morphometric one or intensity one).

Second, the PSF is obviously dependent on many physical parameters, inter alia the refractive index of the sample, immersion medium, its turbidity, the degree of homogeneity of the sample, and the photochemical properties of the dyes used. For these reasons the development of complicated computations often leads to poorly applicable results in practice, since conditions often change dramatically for the different measurements.

One of the most meaningful factors on which PSF depends is the refractive index mismatch between the objective immersion medium and that of the sample solution. This results in a loss of axial resolution and a corruption of the shape (Diaspro et al., 2002a).

Table 11–1 reports the value of theoretical and experimental FWHM of confocal PSFs using different pinhole sizes. A sample of subresolu-

Table 11–1. FWHM of confocal PSFs for different pinhole sizes.

	Oil ($n = 1.5$)			
	Lateral (nm)		Axial (nm)	
	Pinhole 20 μm	Pinhole 50 μm	Pinhole 20 μm	Pinhole 50 μm
Experimental	186 $\pm$ 6	215 $\pm$ 5	489 $\pm$ 6	596 $\pm$ 4
Theoretical	180	210	480	560

Table 11–2. Variation of lateral FWHM with focusing depth.

Depth (μm)	Air		Glycerol		Oil	
	Lateral (nm)	Axial (nm)	Lateral (nm)	Axial (nm)	Lateral (nm)	Axial (nm)
0	187 ± 8	484 ± 24	183 ± 14	495 ± 29	186 ± 6	489 ± 6
30	244 ± 10	623 ± 9	221 ± 5	545 ± 12	197 ± 10	497 ± 21
60	269 ± 11	798 ± 10	252 ± 7	628 ± 9	186 ± 12	496 ± 19
90	277 ± 5	1063 ± 24	268 ± 8	797 ± 26	191 ± 9	484 ± 12

tion beads [Polyscience, diameter = $(0.064 \pm 0.009) \mu\text{m}$] has been imaged by means of a 100× NIKON oil-immersion objective (NA = 1.3; WD = 0.20 mm) under argon laser excitation ($\lambda = 488 \text{ nm}$). Theoretical values are those expected (in the absence of mismatch) and are calculated by means of web-based deconvolution software (<http://www.powermicroscope.com>).

As can be seen from the reported values, the system resolution is worse along the optical axis and is strictly dependent on the pinhole size: these results are in accordance with what is expected from the above theory. Asymmetry of the plots in the real case is typical and becomes even more evident when focusing through different stratified media (Diaspro et al., 2002a).

The theory, developed within the context of electromagnetic waves focusing across stratified media, suggests a progressive broadening of the PSF with respect to the focusing depth, becoming even more noticeable under refractive-index mismatch conditions. Furthermore, on a higher level of complexity, the coexistence of different refractive indices within the sample and the resulting artifacts can be considered.

As a consequence of this, the largest percentage of variation of the lateral FWHM, with respect to the focusing depth, goes from 6% (oil-immersed PSF), to 48% (air-immersed PSF), whereas the axial FWHM varies up to 130% (air-immersed PSF) (see Table 11–2).

This phenomenon is related to a subsequent weakening of the signal with respect to the focusing depth, which turns out to be more evident in the case of refractive index mismatch. Table 11–3 gives typical observed values of the percentage of variation of the PSF intensity peak under different mismatch conditions and at different focusing depths (referred to the coverslip).

Table 11–3. Variations in PSF intensity with mismatch conditions and focusing depths.

Medium	% at 30 μm depth	% at 60 μm depth	% at 90 μm depth
Oil	3	6	7
Glycerol	17	27	34
Air	44	51	60

3.5 Resolution and Three-Dimensional Optical Sectioning

Three-dimensional reconstruction of an object starting from the acquisition of 2D confocal slices is one of the most powerful procedures for morphological analysis and volume rendering, especially within biological sciences, where optical slicing allows information to be obtained from different planes of the specimen without being invasive, thus preserving the structure and functionality of the different parts.

Three-dimensional optical sectioning is intrinsically coupled with the axial resolution of the confocal microscope. For pinhole diameters smaller than 1 AU the approximation of a point-like pinhole is used. For this case the FWHM of the total PSF in the z direction can be expressed as

$$r_z = \frac{0.64\bar{\lambda}}{n - \sqrt{n^2 - NA^2}} \quad (3)$$

This expression describes the axial resolution and the effective optical slice thickness for the sectioning of the specimen. Here $\bar{\lambda}$ is $\bar{\lambda} \approx \sqrt{2}(\lambda_{em}\lambda_{ex}/\sqrt{\lambda_{ex}^2 + \lambda_{em}^2})$: a mean wavelength. This technique is essentially based on an automatic fine z stepping either of the objective or of the sample stage, coupled with the usual x - y point-to-point scanning of the focal plane and image capturing.

The synchronous x - y - z scanning allows the collection of a set of in-focus 2D images, which are less affected by signal cross-talk from other planes of the sample as more strictly confocal conditions are respected.

This means that when a set of 2D images is acquired at various focus positions and under certain conditions, in principle the 3D shape of the object can be recovered. However, the observed image $o(x,y,z)$, produced by the true intensity distribution $i(x,y,z)$, is corrupted by the characteristic PSF of the image formation system $s(x,y,z)$, by noise stemming from different sources $n(x,y,z)$, and by cross information coming from different planes rather than from the focus one.

At a certain plane of focus z_0 within the sample or, at discrete planes along the z axis, the simplest way to describe such a process for the j th plane can be regarded as the following:

$$o_j = i_j \otimes s_0 + i_{j-1} \otimes s_{j-1} + i_{j+1} \otimes s_{j+1} + (\text{other plane contributions if relevant}) + n \quad (4)$$

where the subscripts on i and o refer to the z plane numbers, while the subscripts on s refer to the number of interplane spacings z away from the “in-focus” position at the actual j th plane.

This relationship is usually transferred to the Fourier frequency domain, where the convolution operator becomes an algebraic multiplication (Diaspro et al., 1990). Image restoration algorithms (deconvolution) aim to invert such equations in order to extract the true measured quantity $i(x,y,z)$. Thus a 3D sample reconstruction is possible directly by piling up 2D images. Therefore further scale corrections are performed accounting for axial distortion phenomena linked to the refractive index mismatch.

The generic considerations for this method evidently demonstrate the crucial importance of getting to know the system performance under different working conditions, by means of its PSF. This knowledge will extend the possibilities of applying the confocal technique to a wide range of studies.

In practice, one wants to find the best estimate, accordingly to some criterion, of $i(x,y,z)$ through the knowledge of the observed images, the distortion or PSF of the image formation system, and the additive noise within a restoration scheme classical for space invariant linear systems (Diaspro et al., 1990; Bertero and Boccacci, 1998; Boccacci and Bertero, 2002). Figure 11–4 shows an example of digital restoration of microscopic data obtained after solving the appropriate set of equations. So far, this can be computationally done starting from any data set of optical slices. Recently a WWW service, named Power-Up-Your-Microscope (Diaspro et al., 2002c; Bonetto et al., 2004), has become available that produces the best estimate of $i(x,y,z)$ accordingly to the acquired data set of optical slices. Interested readers can find information and check the service through the webpage <http://www.power-microscope.com> for free (Figure 11–4).

Image restoration is needed only to correct PSF distortions that are less than in the conventional case. Unfortunately, a drawback occurs. In fact, during the excitation process of the fluorescent molecules the whole thickness of the specimen is harmed by every scan, within an hourglass-shaped region (Bianco and Diaspro, 1989). This means that even though out-of-focus fluorescence is not detected, it is generated, with the negative effect of potential induction of those photobleaching and phototoxicity phenomena previously mentioned. The situation becomes particularly serious when there is the need for 3D and temporal imaging coupled to the use of fluorochromes that require excita-

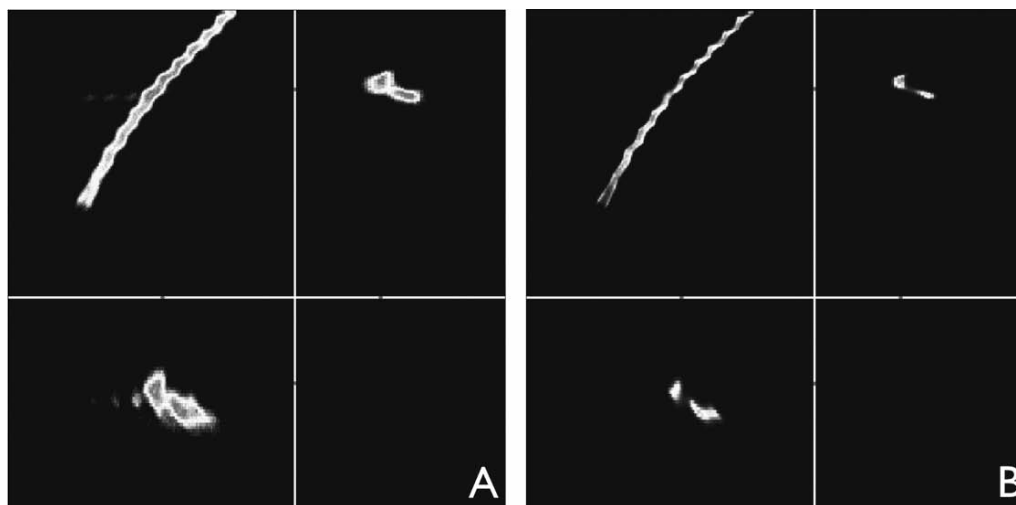


Figure 11–4. Comparison between 3D views of a helicoidal biological sample before (left) and after (right) image processing utilizing 3D deconvolution strategies as implemented at <http://www.power-microscope.com> as the web-based computational facility (Difato et al., 2004).

tion in the ultraviolet regime (Stelzer et al., 1994; Denk, 1996). As earlier reported by König and colleagues (1996a,b), even using UVA (320–400 nm) photons may modify the activity of the biological system. DNA breaks, giant cell production, and cell death can be induced at radiant exposures of the order of magnitude of a few J/cm², accumulated during 10 scans with a 5-μW laser scanning beam at approximately 340 nm and a 50-μs pixel dwell time. In this context, TPE of fluorescent molecules provides an immediate practical advantage over confocal microscopy (Denk et al., 1990; Potter, 1996; Centonze and White, 1998; Gu and Sheppard, 1995; Diaspro, 1998; Piston, 1999; Squirrel et al., 1999; Diaspro and Robello, 2000; So et al., 2000; Wilson, 2002). In fact, reduced overall photobleaching and photodamage are generally acknowledged as major advantages of TPE in laser scanning microscopy of biological specimens (Brakenhoff et al., 1996; Denk and Svoboda, 1997; Patterson and Piston, 2000) even though photobleaching in the focal plane can be accelerated (Patterson and Piston, 2000). However, the excitation intensity has to be kept low considering a regime under 10 mW of average power as a normal operation mode. When laser power is increased above 10 mW some nonlinear effects might arise, evidenced through abrupt rising of the signals (Hopt and Neher, 2001). Moreover photothermal effects should be induced, especially when focusing on single molecule detection schemes (Chirico et al., 2003a).

4 Two-Photon Excitation

Let us now move from conventional excitation of fluorescence as used in computational optical sectioning and confocal microscopy to a special case of multiphoton excitation, i.e., TPE. All considerations can be easily extended to the TPE. The physical suppression of contributions from adjacent planes is realized in a completely different way, thus moving again to 3D optical sectioning ability.

In TPE, two low-energy photons are involved in the interaction with absorbing molecules. The excitation process of a fluorescent molecule can take place only if two low-energy photons are able to interact simultaneously with the very same fluorophore. As mentioned in the introduction, the time scale for simultaneity is the time scale of molecular energy fluctuations at photon energy scales, as determined by the Heisenberg uncertainty principle, i.e., 10⁻¹⁶–10⁻¹⁵ s (Louisell, 1973). These two photons do not necessarily have to be identical, but their wavelengths, λ_1 and λ_2 , have to be such that

$$\lambda_{1P} \equiv \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \right)^{-1} \quad (5)$$

where λ_{1P} is the wavelength needed to prime fluorescence emission in a conventional one-photon absorption process according to the energy relation given in Eq. (1). This situation, compared to the conventional one-photon excitation process shown in Figure 11–5, is illustrated using

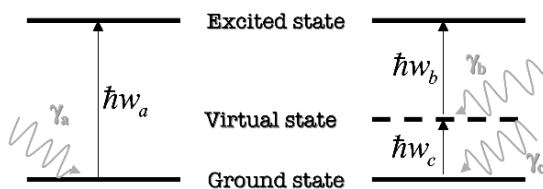


Figure 11-5. Perrin-Jablonsky-like diagram illustrating the difference between conventional (left) and two-photon (right) excitation (Esposito et al., 2004). In the second case, photons delivering one-half the energy conventionally needed for bringing the fluorescent molecule to an excited state are used (see text).

a Perrin-Jablonski-like diagram. It is worth noting that for practical reasons the experimental choice is usually such that (Denk et al., 1990; Diaspro, 2001; Girkin and Wokosin, 2002)

$$\lambda_{2P} = 2\lambda_{1P} \quad \lambda_1 = \lambda_2 \approx 2\lambda_{1P} \quad (6)$$

and

$$\Delta E_g = \frac{2\hbar c}{\lambda_{1P}} \quad (7)$$

Considering this as a nonresonant process and assuming the existence of a virtual intermediate state, the resident time, τ_{virt} in this intermediate state should be calculated using the time-energy uncertainty consideration for TPE:

$$\Delta E_g \cdot \tau_{virt} \cong \hbar/2 \quad (8)$$

where $\hbar = h/2\pi$. It follows that

$$\tau_{virt} \cong 10^{-15} - 10^{-16} \text{ s} \quad (9)$$

This is the temporal window available to two photons to coincide in the virtual state.

So far, in a TPE process it is hence crucial to combine sharp spatial focusing with temporal confinement of the excitation beam.

The process can be extended to n -photons requiring higher photon densities temporally and spatially confined. Thus, near infrared (circa 680–1100 nm) photons can be used to excite UV and visible electronic transitions producing fluorescence. The typical photon flux densities are of the order of more than 10^{24} photons $\text{cm}^{-2} \text{s}^{-1}$, which implies intensities around MW–TW cm^{-2} (Goppert Mayer, 1931). An elegant treatment in terms of quantum theory for two-photon transition has been proposed by Nakamura (1999) using perturbation. He clearly described the process by a time-dependent Schrödinger equation, where the Hamiltonian contains electric dipole interaction terms. Using a perturbation expansion, the first-order solution is found to be related to one-photon excitation while higher order solutions are related to n -photon ones (Esposito et al., 2004).

The dependence of TPE on I^2 should be evident and is demonstrated by using simple arguments (Diaspro and Sheppard, 2002).

The fluorescence intensity per molecule, $I_f(t)$, can be considered to be proportional to the molecular cross-section $\delta_2(\lambda)$ and to the square of $I(t)$ as follows:

$$I_f(t) \propto \delta_2 \cdot I(t)^2 \propto \delta_2 \cdot P(t)^2 \left[\pi \frac{(NA)^2}{hc\lambda} \right]^2 \quad (10)$$

where $P(t)$ is the laser power and (NA) is the numerical aperture of the focusing objective lens. The last term of Eq. (10) simply takes care of the distribution in time and space of the photons by using the paraxial approximation in an ideal optical system (Born and Wolf, 1980).

It follows that the time-averaged two-photon fluorescence intensity per molecule within an arbitrary time interval T , $\langle I_f(t) \rangle$, can be written as

$$\langle I_f(t) \rangle = \frac{1}{T} \int_0^T I_f(t) dt \propto \delta_2 \left[\pi \frac{(NA)^2}{hc\lambda} \right]^2 \frac{1}{T} \int_0^T P(t)^2 dt \quad (11)$$

in the case of continuous wave (CW) laser excitation.

Now, because the present experimental situation for TPE is related to the use of ultrafast lasers, we consider that for a pulsed laser $T = 1/f_p$, where f_p is the pulse repetition rate. This implies that a CW laser beam, where $P(t) = P_{ave}$, allows transformation of Eq. (11) into

$$\langle I_{f,cw}(t) \rangle \propto \delta_2 \cdot P_{ave}^2 \left[\pi \frac{(NA)^2}{hc\lambda} \right]^2 \quad (12)$$

For a pulsed laser beam with pulse width, τ_p , repetition rate, f_p , and average power

$$P_{ave} = D \cdot P_{peak}(t)$$

where $D = \tau_p \cdot f_p$, the approximated $P(t)$ profile can be described as

$$\begin{aligned} P(t) &= P_{ave}/D & \text{for } 0 < t < \tau_p \\ P(t) &= 0 & \text{for } \tau_p < t < (1/f_p) \end{aligned} \quad (13)$$

We can write Eq. (12) as:

$$\langle I_{f,p}(t) \rangle \propto \delta_2 \frac{P_{ave}^2}{\tau_p^2 f_p^2} \left[\pi \frac{(NA)^2}{hc\lambda} \right]^2 \frac{1}{T} \int_0^{\tau_p} dt = \delta_2 \frac{P_{ave}^2}{\tau_p f_p} \left[\pi \frac{(NA)^2}{hc\lambda} \right]^2 \quad (14)$$

The conclusion here is that CW and pulsed lasers operate at the very same excitation efficiency, i.e., fluorescence intensity per molecule, if the average power of the CW laser is kept higher by a factor of $1/\sqrt{\tau \cdot f_p}$. This means that 10 W delivered by a CW laser, allowing the same efficiency of conventional excitation performed at approximately 10^{-1} mW, is nearly equivalent to 30 mW for a pulsed laser (Diaspro and Chirico, 2002).

5 Fluorescent Molecules under TPE Regime

The above steps lead to the most popular relationship reported below, which is related to the practical situation of a train of beam pulses focused through a high numerical aperture objective, with a duration τ_p and repetition rate f_p . In this case, the probability, n_a , that a certain

fluorophore simultaneously absorbs two photons during a single pulse, in the paraxial approximation, is (Denk et al., 1990)

$$n_a \propto \frac{\delta_2 \cdot P_{\text{ave}}^2}{\tau_p f_p^2} \left(\frac{NA^2}{2hc\lambda} \right)^2 \quad (15)$$

where P_{ave} is the time-averaged power of the beam and λ is the excitation wavelength. Introducing 1 GM (Goppert-Mayer) = 10^{-58} ($\text{m}^4 \cdot \text{s}$), for a δ_2 of approximately 10 GM per photon (Denk et al., 1990; Xu, 2002), focusing through an objective of $NA > 1$, an average incident laser power of ≈ 1 –50 mW, and operating at a wavelength ranging from 680 to 1100 nm with 80–150 fs pulsewidth and 80–100 MHz repetition rate, would saturate the fluorescence output as for one-photon excitation. This suggests that for optimal fluorescence generation, the desirable repetition time of pulses should be on the order of a typical excited-state lifetime, which is a few nanoseconds for commonly used fluorescent molecules. For this reason the typical repetition rate is around 100 MHz. A further condition that makes Eq. (15) valid is that the probability that each fluorophore will be excited during a single pulse has to be smaller than one. The reason lies in the observation that during the pulse time (10^{-13} s of duration and a typical excited-state lifetime in the 10^{-9} s range) the molecule has insufficient time to relax to the ground state. This can be considered a prerequisite for absorption of another photon pair. Therefore, whenever n_a approaches unity saturation effects start to occur. The use of Eq. (15) makes it possible to choose optical and laser parameters that maximize excitation efficiency without saturation. In case of saturation the resolution declines and the image becomes worse (Cianci et al., 2004). It is also evident that the optical parameter for enhancing the process in the focal plane is the lens numerical aperture, NA , even if the total fluorescence emitted is independent of this parameter as shown by Xu (2002). This is usually confined around 1.3–1.4 as the maximum value. Now, it is possible to estimate n_a for a common fluorescent molecule like fluorescein that possesses a two-photon cross-section of 38 GM at 780 nm (Diaspro and Chirico, 2003).

To this end, we can use $NA = 1.4$, a repetition rate at 100 MHz, and a pulse width of 100 fs within a range of P_{ave} values of 1, 10, 20, and 50 mW, and substituting the proper values in Eq. (15) we get $n_a \cong 5930 \cdot P_{\text{ave}}^2$. This result for $P_{\text{ave}} = 1, 20$, as a function of 1, 10, 20, and 50 mW, gives values of 5.93×10^{-3} , 5.93×10^{-1} , 1.86, and 2.965, respectively. It is evident that saturation begins to occur at 10 mW (Diaspro and Sheppard, 2002).

The related rate of photon emission per molecule, at a nonsaturation excitation level, in the absence of photobleaching (Patterson and Piston, 2000; So et al., 2001), is given by n_a multiplied by the repetition rate of the pulses. This means approximately 5×10^7 photons s^{-1} in both cases. It is worth noting that, when considering the effective fluorescence emission, a further factor given by the so-called quantum efficiency of the fluorescent molecules should also be considered. At present, the quantum efficiency value is usually known from conventional one-photon excitation data (Diaspro, 2002).

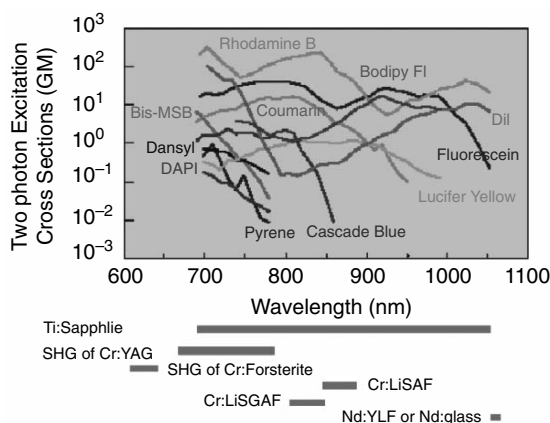


Figure 11-6. Two-photon cross-sections for popular fluorescent molecules as a function of the excitation wavelength. Red bars indicate the emission range of some common laser sources utilized in TPE microscopy and spectroscopy. (See color plate.)

Now, even if the quantum-mechanical selection rules for TPE differ from those for one-photon excitation, several common fluorescent molecules can be used. Unfortunately, knowing the one-photon cross-section for a specific fluorescent molecule does not allow any quantitative prediction of the two-photon trend, except for a sort of “rule of thumb.” This simple rule states that, in general, a TPE cross-section may be expected to peak at double the wavelength needed for one-photon excitation. However, the cross-section parameter has been measured for a wide range of dyes (Xu et al., 1995; Albota et al., 1998b; Diaspro and Chirico, 2003). It is worth noting that due to the increasing dissemination of TPE microscopy, new “ad hoc” organic molecules, endowed with large engineered two-photon absorption cross-sections, have recently been developed (Albota et al., 1998; Abbotto et al., 2005). Figure 11-6 summarizes the properties of some commonly used fluorescent molecules under two-photon excitation (Xu et al., 1995; So et al., 2000). TPE fluorescence from NAD(P)H, flavoproteins (Piston et al., 1995; So et al., 2000), tryptophan, and tyrosine in proteins (Lakowicz and Gryczynski, 1992) has been measured. In addition, the autofluorescent biological proteins such as the GFP and its molecular variants are important molecular markers (Chalfie et al., 1994; Chalfie and Kain, 1998; Potter, 1996; Zimmer, 2002). Their TPE cross-sections are between 6 and 40 GM (Blab et al., 2001). As a comparison consider that the cross-section for NADH, at the excitation maximum of 700 nm, is approximately 0.02 GM (So et al., 2000). Moving to quantum dots there is an increase of cross-section up to 2000 GM.

6 Optical Consequences of TPE

In terms of optical consequences the two-photon effect limits the excitation region to within a subfemtoliter volume. The 3D confinement of the TPE volume can be understood with the aid of optical diffraction theory

(Born and Wolf, 1980). Using excitation light with wavelength λ , the intensity distribution at the focal region of an objective with numerical aperture $NA = \sin(\alpha)$ is described [see also Eq. (2)] in the paraxial regime (Born and Wolf, 1980; Sheppard and Gu, 1990) by

$$I(u, v) = \left| 2 \int_0^1 J_0(v\rho) e^{-(i/2)u\rho^2} \rho d\rho \right|^2 \quad (16)$$

where ρ is a dimensionless radial, J_0 is the zeroth-order Bessel function, ρ is a radial coordinate in the pupil plane, and $u = 8\pi \sin^2(\alpha/2)z/\lambda$ and $v = 2\pi \sin(\alpha)r/\lambda$ are dimensionless axial and radial coordinates, respectively, normalized to the wavelength (Wilson and Sheppard, 1984). Now, the intensity of fluorescence distribution within the focal region has an $I(u, v)$ behavior for the one-photon case and $I^2(u/2, v/2)$ for the TPE case as demonstrated above. The arguments of $I^2(u/2, v/2)$ take into proper account the fact that in the latter case wavelengths are utilized that are approximatively twice those used for one-photon excitation. As compared with the one-photon case, the TPE intensity distribution is axially confined (Nakamura, 1993; Gu and Sheppard, 1995; Jonkman and Stelzer, 2002). In fact, considering the integral over v , keeping u constant, its behavior is constant along z for one-photon and has a half-bell shape for TPE. This behavior, better discussed in Wilson (2002), Torok and Sheppard (2002), and Jonkman and Stelzer (2002), explains the three-dimensional discrimination property in TPE.

Now, the most interesting aspect is that the excitation power falls off as the square of the distance from the lens focal point, within the approximation of a conical illumination geometry. In practice this means that the quadratic relationship between the excitation power and the fluorescence intensity results in the fact that TPE falls off as the fourth power of distance from the focal point of the objective. This fact implies that those regions away from the focal volume of the objective lens, directly related to the numerical aperture of the objective itself, therefore do not suffer photobleaching or phototoxicity effects and do not contribute to the signal detected when a TPE scheme is used. Because they are simply not involved in the excitation process, a confocal-like effect is obtained without the necessity of a confocal pinhole. It is also immediately evident that in this case an optical sectioning effect is obtained. In fact, the observed image $o(x, y, z)$ at a plane j , produced by the true fluorescence distribution $i(x, y, z)$ at plane j , distorted by the microscope through s , plus noise n , again corresponds to the confocal ideal situation where contributions from adjacent k planes can be set to zero as in the confocal situation: $o_j = i_j * s_j + n$.

This means that TPE microscopy is intrinsically three dimensional. It is worth noting that the optical sectioning effect is obtained in a very different way with respect to the confocal solution. No fluorescence has to be removed from the detection pathway. In this case it should be possible to collect as much fluorescence is possible. In fact fluorescence can come only and exclusively from the small focal volume traced in Figure 11-7, which also shows a comparison with the confocal mode, that is of the order of a fraction of a femtoliter.

In TPE over 80% of the total intensity of fluorescence comes from a 700- to 1000-nm-thick region about the focal point for objectives with numerical apertures in the range of 1.2–1.4 (Brakenhoff et al., 1979;

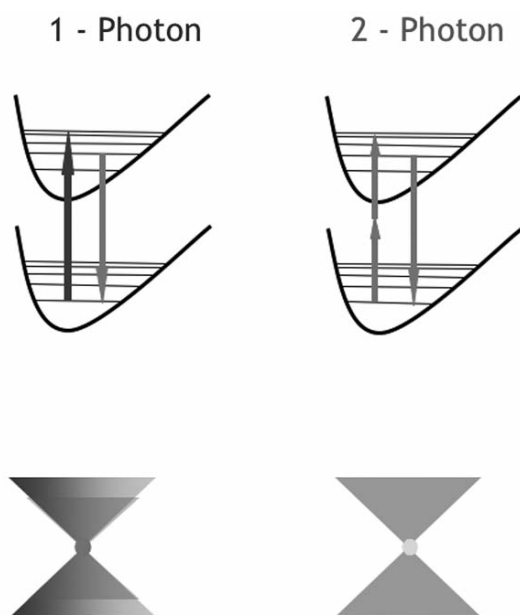


Figure 11-7. Illustration of the two different modalities for selecting 3D information under a confocal (left) and TPE regime (right). In the confocal case the selection is realized during the emission process. The different case of two-photon excitation shows how the 3D selection can be realized during the fluorescence excitation process.

Wilson and Sheppard, 1984; Wilson, 2002; Jonkman and Stelzer, 2002; Torok and Sheppard, 2002). This fact implies a reduction in background that allows compensation of the poorer spatial resolution compared to the single-photon confocal mode due to the longer wavelength utilized. However, the utilization of an infrared wavelength instead of UV-visible ones also allows deeper penetration than in the conventional case (So et al., 2001; Periasamy et al., 2002; König and Tirlapur, 2002). The long wavelengths used in TPE, or in general in multiphoton excitation, will be scattered less than the ultraviolet-visible wavelengths used for conventional excitation (de Grauw and Gerritsen, 2001). Hence deeper targets within a thick sample can be reached. Of course, for fluorescence light, scattering on the way back can be overcome by acquiring the emitted fluorescence using a large area detector and collecting not only ballistic photons (Soeller and Cannel, 1999; Bauhler et al., 1999; Girkin and Wokosin, 2002).

7 The Optical Setup

A TPE architecture including confocal modality includes the following: a high peak-power laser delivering moderate average power (femtosecond or picosecond pulsed at a relatively high repetition rate) emitting infrared or near-infrared wavelengths (650–1100 nm), CW laser sources for confocal modes, a laser beam scanning system or a confocal laser scanning head, high numerical aperture objectives (>1), a high-throughput

microscope pathway, and a high-sensitivity detection system (Denk et al., 1995; So et al., 1996; Soeller and Cannell, 1996; Wokosin and White, 1997; Centonze and White, 1998; Potter et al., 1996; Wolleschensky et al., 1998; Diaspro et al., 1999a,b; Wier et al., 2000; Soeller and Cannell, 1999; Tan et al., 1999; Mainen et al., 1999; Majewska et al., 2000; Diaspro, 2002; Girkin and Wokosin, 2002; Iyer et al., 2002). Figure 11–8 shows a general scheme for a TPE microscope incorporating a confocal mode.

In typical TPE or confocal microscopes, images are built by raster scanning the x – y mirrors of a galvanometrically driven mechanical scanner (Webb, 1996). This fact implies that image formation speed is mainly determined by the mechanical properties of the scanner, i.e., for single line scanning it is of the order of milliseconds. Faster beam-scanning schemes can be realized, even if the “eternal triangle of compromise” should be considered for sensitivity, spatial resolution, and temporal resolution. While the x – y scanners provide lateral focal-point scanning, axial scanning can be achieved by means of different positioning devices, the most popular being a belt-driven system using a DC motor and a single objective piezo nanopositioner, such as the PIFO (Physik Instrumente, Germany). Usually, it is possible to switch between confocal and TPE modes retaining x – y – z positioning on the sample being imaged (Diaspro, 2001; Diaspro and Chirico, 2003). Acquisition and visualization are generally completely computer controlled by dedicated software. Figure 11–9 shows a TPE microscope.

Let us now consider two popular approaches that can be used to perform TPE microscopy, namely, the descanned and nondescanned

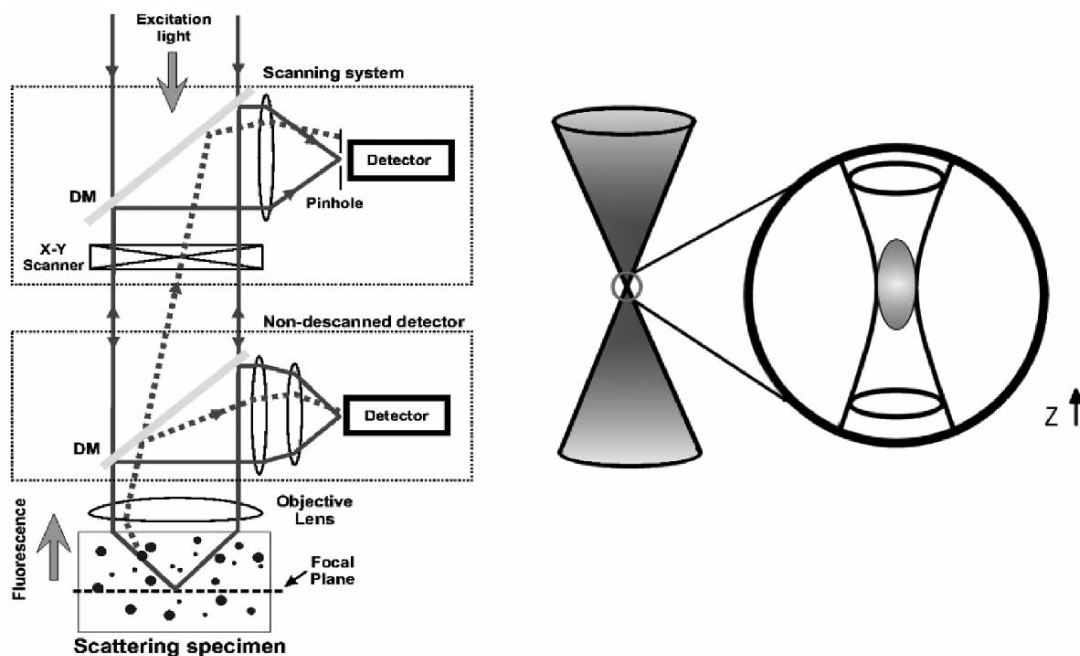


Figure 11–8. Optical configuration for a TPE microscope operating in a descanned (upper inset box) and nondescanned (lower inset box) mode; see text. (Courtesy of M. Cannel and C. Soeller.)



Figure 11–9. The TPE setup at LAMBS, MicroScoBio Research Center of the University of Genoa (from left to right: Ilaria Testa, Paolo Bianchini, and Davide Mazza).

modes. They are sketched in Figure 11–8. The former uses the very same optical pathway and mechanism employed in confocal laser scanning microscopy. The latter mainly optimizes the optical pathway by minimizing the number of optical elements encountered on the way from the sample to detectors, and increases the detector area. The TPE nondescanned mode provides very good performances giving a superior signal-to-noise ratio inside strongly scattering samples (Masters et al., 1997; Daria et al., 1998; Centonze and White, 1998; So et al., 2000).

In the descanned approach pinholes are removed or set to their maximum aperture and the emission signal is captured using an excitation scanning device on the back pathway. For this reason it is called the descanned mode. In the latter, the confocal architecture has to be modified in order to increase the collection efficiency: pinholes are removed and the emitted radiation is collected using dichroic mirrors on the emission path or external detectors without passing through the galvanometric scanning mirrors. A high-sensitivity detection system is another critical issue (Wokosin et al., 1998; So et al., 2000; Girkin and Wokosin, 2002).

The fluorescence emitted is collected by the objective and transferred to the detection system through a dichroic mirror along the emission path (Figure 11–8). Due to the high excitation intensity, an additional barrier filter is needed to avoid mixing the excitation and emission light at the detection system that is differently placed depending on the acquisition scheme being used. Photodetectors that can be used include photomultiplier tubes, avalanche photodiodes, and CCD cameras (Denk et al., 1995; Murphy, 2001). Photomultiplier tubes are the most commonly used. This is due to their low cost, good sensitivity in the blue-green spectral region, high dynamic range, large size of the sensitive area, and single-photon counting mode availability (Hamamatsu Photonics, 1999). They have a quantum efficiency around 20–40% in the blue-green spectral region that drops down to <1% moving to the red region. This is a good condition, especially in MPE mode, because it is desirable to reject as much as possible wavelengths above 680 nm

that are mainly used for excitation. Another advantage is that the large size of the sensitive area of photomultiplier tubes allows efficient collection of signal in the nondescanned mode within a dynamic range of the order of 10^8 . Avalanche photodiodes are excellent in terms of sensitivity exhibiting quantum efficiency close to 70–80% in the visible spectral range. Unfortunately they are high in cost and the small active photosensitive area, <1 mm size, could introduce drawbacks in the detection scheme and requires special descanning optics (Farrer et al., 1999). CCD cameras are used in video rate multifocal imaging (Fuijta and Takamatsu, 2002; Girkin and Wokosin, 2002). However, once the best quality image possible has been obtained then image restoration algorithms can be applied to enhance the features of interest to the biological researcher and to improve the quality of data to be used for three-dimensional modeling, such as those used for single-photon optical sectioning microscopy, available at <http://www.powermicroscope.com> (van der Voort et al., 1995; Shotton, 1995; Diaspro et al., 1990, 2000; Boccacci and Bertero, 2002; Carrington, 2002; Difato et al., 2004; Bonetto et al., 2004).

Laser sources, as often happened in optical microscopy, represent an important resource, especially in fluorescence microscopy (Gratton and van de Ven, 1995; Svelto, 1998). For nonresonant TPE, owing to the comparatively low TPE cross-sections of fluorophores, high photon flux densities are required, $>10^{24}$ photons $\text{cm}^{-2}\text{s}^{-1}$ (König, 2000). Using radiation in the spectral range of 600–1100 nm for TPE, excitation intensities in the MW–GW cm^{-2} range are required. This high energy can be obtained by the combined use of focusing lens objectives and CW (Hanninen and Hell, 1994; König et al., 1995) or pulsed (Denk et al., 1990) laser radiation of 50 mW mean power or less (Girkin and Wokosin, 2002; Diaspro and Sheppard, 2002). TPE microscopes have been realized using CW, femtosecond, and picosecond laser sources (Periasamy, 2001; Diaspro, 2001, 2002; Masters, 2002). Since the original successful experiments in TPE microscopy, advances have been made in the technological field of ultrashort pulsed lasers. Today laser sources suitable for TPE can be described as “turnkey” compact systems (Fisher et al., 1997; Wokosin et al., 1996; Diaspro, 2001).

Figure 11–10 shows a new generation ultrafast Ti:sapphire laser source. The emission range between 700 and 1050 nm of the Ti:sapphire laser allows a large number of commonly used fluorescent molecules to be excited. Other laser sources used for TPE are Cr:LiSAF, pulse-compressed Nd:YLF in the femtosecond regime, and mode-locked Nd:YAG and picosecond Ti-sapphire lasers in the picosecond regime (Gratton and Van de Ven, 1995; Wokosin et al., 1996). Most of the laser sources used for TPE operate in a mode-locking mode. This endows the laser with the ability to generate a train of very short pulses by modulating the gain or excitation of a laser at a frequency with a period equal to the roundtrip time of a photon within the laser cavity (Fisher et al., 1997; Svelto, 1998) (Figure 11–11). The resulting pulsewidth is in the 50–150 fs regime. The parameters that are more relevant in the selection of the laser source are average power, pulsewidth and repetition rate, and wavelength also according to Eq. (15). The most

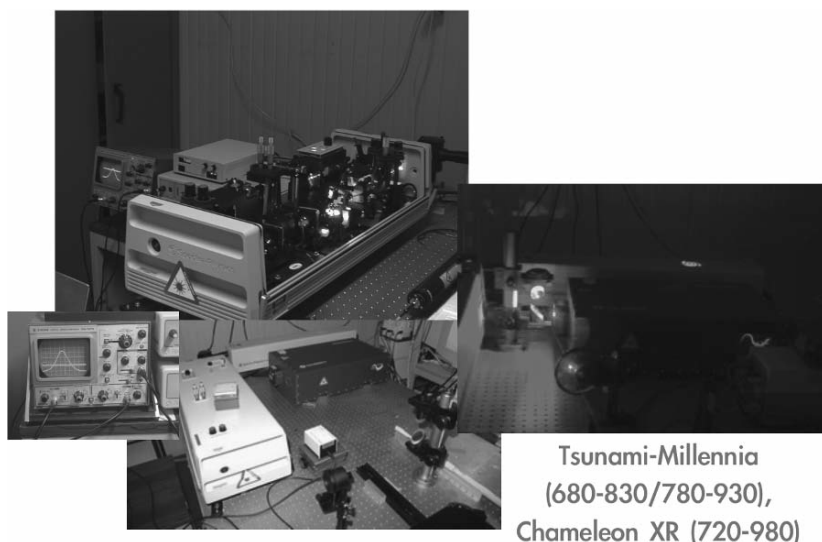


Figure 11-10. Typical laser sources in use for TPE microscopy.

popular features for an infrared pulsed laser are 700 mW–1 W average power, 80–100 MHz repetition rate, and 100–150 fs pulse width.

At present, the use of short pulses and small duty cycles are mandatory to allow image acquisition in a reasonable time while using power levels that are biologically tolerable (Denk et al., 1994; Denk, 1996; Koester et al., 1999; König et al., 1996, 1998; König, 2000; König and Tirlapur, 2002). To minimize pulse width dispersion problems König (2000) suggested working with pulses around 150–200 nm, and this constitutes a very good compromise both for pulse stretching and sample viability. It should always be remembered that a shorter pulse broadens more than a longer one. Pulse width measurement is a very delicate issue. In fact, because it is not very easy to measure it at the focal volume within the sample, little can be definitely said about it (Hanninen and Hell, 1994; Guild et al., 1997; Wolleschensky et al., 2002). Although users do not perform measurement of the pulse width at the

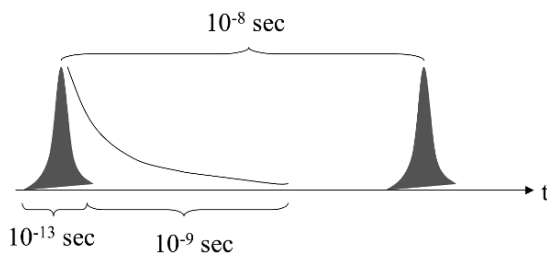


Figure 11-11. Laser emission time scale for TPE excitation: a short pulse at high photon density is released for approximately 100 fs; this laser shot is able to prime fluorescence without damaging the sample so fluorescence occurs in the next few nanoseconds. The laser is silent for 10 ns and then delivers a new high-density photon pulse. This modality allows TPE to be experienced at tolerable time-averaged power (see text).

sample when they use two-photon microscopy, which would require a specific procedure that, even if not too complex for a researcher in the field, could be irksome for the majority of users, it is a reasonable approximation to assume that at the focal volume, 1.5–2. times temporal pulse broadening occurs using high-quality optics (Wolleschensky, 2002; Girkin and Wokosin, 2002). As an example, for a measured laser pulse width of about 100 fs, an estimate at the sample is about 150–180 fs under favorable experimental conditions, sample characteristics included. Sample properties are mentioned because for thick samples the role played by thickness, also in terms of pulse width broadening, is not so obvious (de Grauw and Gerritsen, 2002; So et al., 2001; Gu et al., 2000; Saloma et al., 1998).

8 Conclusion

Confocal microscopy, in the authors' opinion, constitutes one of the most significant advances in optical microscopy within the past decades, and has become a powerful investigative tool for the molecular, cellular, and developmental biologist, the materials scientist, the biophysicist, and the electronic engineer. It is entirely compatible with the range of "classical" light microscopic techniques, and, at least in scanned beam instruments, can be applied to the same specimens on the same optical microscope stage. Its peculiar advantages result in its ability to generate multidimensional (x – y – z – t) images by noninvasive optical sectioning with a virtual absence of out-of-focus blur, its capacity for multiparametric imaging of multiply labeled samples, and its property of investigating at microscopic resolution large objects as a result of the rejection of scattered light. So far, the advent of confocal microscopy in the mid-1980s favored the rapid spreading of two- and multiphoton excitation microscopy, since Denk's report at the beginning of the 1990s, bringing dramatic changes in designing experiments that utilize fluorescent molecules and, more specifically, in fluorescence 3D optical microscopy.

While confocal microscopy is moving to spectral and fast-scanning architectures in terms of acquisition, it is mainly two-photon microscopy that occupies the scene of advances in fluorescence optical microscopy. TPE microscopy, with its intrinsic three-dimensional resolution, the absence of background fluorescence, and the attractive possibility of exciting UV excitable fluorescent molecules, thus increasing sample penetration, constitutes significant progress in science. In fact, in a TPE scheme two 720-nm photons combine to produce the very same fluorescence conventionally primed at ~360 nm, and to be utilized in a classical confocal microscope using conventional excitation of fluorescent molecules. The excitation of the fluorescent molecules bound to the specific components of the biological systems being studied mainly takes place (80%) in an excitation volume of the order of magnitude of 0.1 fl. This results in an intrinsic 3D optical sectioning effect. What is invaluable for cell imaging and, in particular, for live-cell imaging is the fact that weak endogenous one-photon absorption and

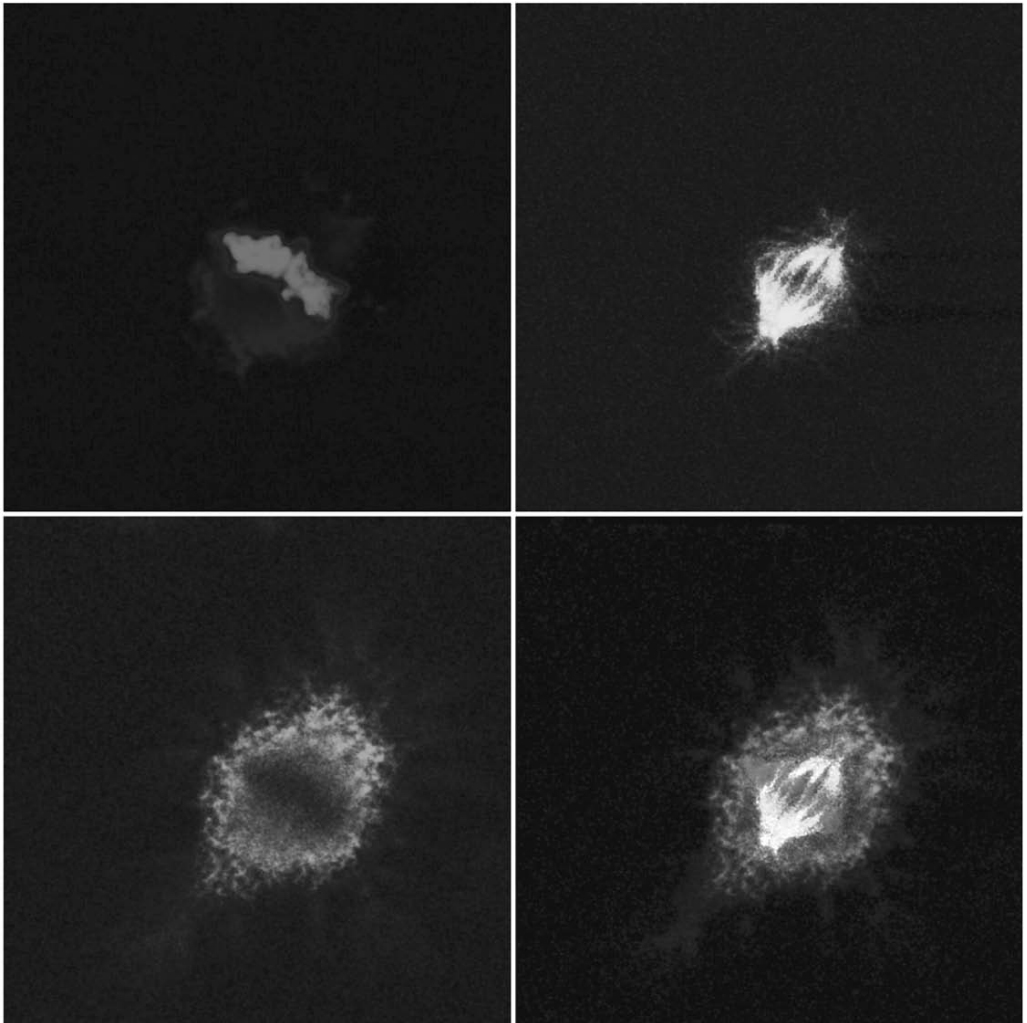


Figure 11-12. Multiple excitation of three fluorescent dyes using 740 nm under a TPE regime. The conventional excitation would have required the utilization of 360 or 405 nm, 488 nm, and 543 nm laser lines. The final image (lower right quadrant) is realized by merging the three subsets. (This image has been acquired by students of the Biotechnology School during the course of Advanced Microscopy Techniques activated at the University of Genoa, academic year 2005. Advisors: Grazia Tagliafierro and Alberto Diaspro.) (See color plate.)

highly localized spatial confinement of the TPE process dramatically reduce phototoxicity stress. To summarize the unique characteristics and advantages of TPE we recall the following properties:

1. Spatially confined fluorescence excitation in the focal plane of the specimen can be considered the key feature of TPE microscopy. It is one of the advantages over confocal microscopy, where fluorescence emission occurs across the entire thickness of the sample being excited by the scanning laser beam. A strong implication is that there is no photon signal from sources out of the geometric position of the optical focus within the sample. Therefore, the signal-to-noise ratio increases, photo-

degradation effects decrease, and optical sectioning is immediately available without the need for pinhole or deconvolution algorithms. In addition, very efficient acquisition schemes can be implemented such as the nondescanned one operating at an excellent signal-to-noise ratio.

2. The use of near-infrared (IR)/IR wavelengths permits examination of thick specimens in depth. This is due to the fact that, apart from some cases such as pigmented samples and portions of the absorption spectral window of water, cells and tissues absorb poorly in the near-IR/IR region. Cellular damage is globally minimized, thus allowing cell viability to be prolonged with long-term 3D sessions. Moreover, scattering is reduced and deeper targets can be reached with fewer problems than in one-photon excitation. The depth of penetration can be up to 0.5 mm. In addition, whereas in one-photon excitation, the emission wavelength is comparatively close to the excitation one (about 50–200 nm longer), in TPE the fluorescence emission occurs at a wavelength substantially shorter and at a larger spectral distance than in one-photon excitation. Thus separation of the excitation light and the emitted light can be easily performed.

Continuing research in this field is focused on very intriguing problems (www.focusonmicroscopy.org offers a complete scenario of the evolution of three-dimensional microscopy in the past 5 years) such as local heating from absorption of IR light by water at high laser power (Schonle and Hell, 1998) and photothermal effects on fluorescent molecules (Chirico et al., 2003a), phototoxicity from long wavelength IR excitation and short wavelength fluorescence emission (König et al., 1996c; Tyrrel and Keyse, 1990; König, 2000; Hopt and Neher, 2001; König and Tirlapur, 2002), photoactivation and photocycling of visible fluorescent proteins (Post et al., 2004; Chirico et al., 2004; Schnedier et al., 2005), development of new fluorochromes better suited for TPE and multiphoton excitation (Albota et al., 1998a; Abbotto et al., 2005), and the investigation of the cross-sections of uncharacterized molecules (Gostkowski et al., 2004; Wokosin et al., 2004).

One of the major benefits in setting up an MPE microscope is the flexibility in choosing the measurement modality favored by the simplification of the optical design. In fact, a TPE microscope offers a greater variety of measurement options without changing any optics or hardware. This means that during the very same experiments real multimodal information can be obtained from the specimen being studied (Zoumi et al., 2002; Wang et al., 2004). Moreover, the usefulness of the TPE scheme for spectroscopic and lifetime studies (So et al., 1996; Sytsma et al., 1998; Schwille et al., 2000; Diaspro et al., 2001; Wiseman et al., 2002), for optical data storage and microfabrication (Cumpston et al., 1999; Kawata et al., 2001), and for single molecule detection (Mertz et al., 1995; Farrer et al., 1999; So et al., 2000; Chirico et al., 2001; Cannone et al., 2003b) has been well documented. Other very interesting applications involve the study of impurities affecting the growth of protein crystals (Caylor et al., 1999), TPE imaging in the field of plant biology (Tirlapur and König, 2002), and measurements in living systems (Squirrel et al., 1999; Yoder and Kleinfeld, 2002; Diaspro et al., 2002b,d; Post et al., 2004). Here the combination of MPE and

second-harmonic generation offers the opportunity to investigate the morphometric properties on the basis of the microstructure of blood cells (Zoumi et al., 2004). Another promising field is the investigation of complex formation where the TPE properties will improve the information accessible (Heinze et al., 2004). Another, more indirect usage that provides a look at the sample with nanometer resolution is the excitation of an evanescent wave at a metal surface (Novotny et al., 1998). For microscopic purposes the evanescent wave needs to be localized at a nanoparticle or a fine metal tip (Sánchez et al., 1999; Gerton et al., 2004). The MPE microscope can also be used as an active device, with increasing applications related to nanosurgery (König, 2000), selective uncaging of caged compounds (Diaspro et al., 2003), and photodynamic therapy (Bhalwalkar et al., 1997; So et al., 2000). Recently TPE microscopy, even if in an evanescent-field-induced configuration, has been extended to large area structures of the order of square centimeters (Duveneck et al., 2001). This has application in the realization of biosensing platforms such as genomic and proteomic microarrays based upon large planar waveguides. It is easy to perceive that the range of applicability of MPE microscopes is rapidly increasing in the biomedical, biotechnological, and biophysical sciences and is expanding to clinical applications (Diaspro, 2002; Masters, 2002; Periasamy and Diaspro, 2003).

Acknowledgments. The first Italian TPE architecture realized at LAMBS has been supported by INFN grants. LAMBS-MicroScoBio is currently funded by IFOM (Istituto FIRC di Oncologia Molecolare, FIRC Institute of Molecular Oncology, Milano). This chapter is dedicated to the memory of Osamu Nakamura, who passed away January 23, 2005 at Handai Hospital.

References

- Abbe, E. (1910). In: *Die Lehre von der Bildentstehung in Mikroskop* (O. Lummer, Ed.). F. Reiche, Braunschweig.
- Abbotto, A., Baldini, G., Beverina, L., Chirico, G., Collini, M., D'alfonso, L., Diaspro, A., Magrassi, R., Nardo, L. and Pagani, G.A. (2005). Dimethyl-Peep: A Dna probe in two-photon excitation cellular imaging. *Biophys. Chem.* **114**(1), 35–41.
- Agard, D.A. (1984). Optical sectioning microscopy: Cellular architecture in three dimensions. *Annu. Rev. Biophys.* **13**, 191–219.
- Agard, D.A., Hiraoka, Y., Shaw, P.J. and Sedat, J.W. (1989). Fluorescence microscopy in three-dimensions. *Methods Cell. Biol.* **30**, 353–378.
- Albota, M., Beljonne, D., Bredas, J.L., Ehrlich, J.E., Fu, J.Y., Heikal, A.A., Hess, S.E., Kogej, T., Levin, M.D., Marder, S.R. and others. (1998a). Design of organic molecules with large two-photon absorption cross sections. *Science* **281**(5383), 1653–1656.
- Albota, M.A., Xu, C. and Webb, W.W. (1998b). Two-photon fluorescence excitation cross sections of biomolecular probes from 690 to 960 nm. *Appl. Opt.* **37**, 7352–7356.
- Amos, B. (2000). Lessons from the history of light microscopy. *Nature Cell Biol.* **2**, E151–E152.
- Andrews, D.L. (1985). A simple statistical treatment of multiphoton absorption. *Am. J. Phys.* **53**, 1001–1002.

- Arndt-Jovin, D.J., Nicoud, R.M., Kaufmann, J. and Jovin, T.M. (1985). Fluorescence digital-imaging microscopy in cell biology. *Science* **230**, 1333–1335.
- Bastiaens, P.I. and Hell, S.W. (Eds.) (2004). Recent advances in light microscopy. *J. Struct. Biol.* **147**, 1–89.
- Basu, S. and Campagnola, P.J. (2004). Properties of crosslinked protein matrices for tissue engineering applications synthesized by multiphoton excitation. *J. Biomed. Mat. Res. Part A* **71A**(2), 359–368.
- Beltrame, F., Bianco, B., Castellaro, G. and Diaspro, A. (1985). Fluorescence, absorption, phase-contrast, holographic and acoustical cytometries of living cells. In: *Interactions between Electromagnetic Fields and Cells* (A. Chiabrera and H.P. Schwan, Eds.). NATO ASI Series, Vol. 97, pp. 483–498. (Plenum Press, New York).
- Benedetti, P. (1998). From the histophotometer to the confocal microscope: The evolution of analytical microscopy. *Eur. J. Histochem.* **42**, 11–17.
- Benham, G.S. and Schwartz, S. (2002). Suitable microscope objectives for multiphoton digital imaging. In: *Multiphoton Microscopy in the Biomedical Sciences II* (A. Periasamy and P.T.C. So, Eds.). Proc. SPIE 4620, 36–47.
- Berland, K. (2002). Basics of fluorescence. In: *Methods in Cellular Imaging* (A. Periasamy, Ed.), pp. 5–19. (Oxford University Press, New York).
- Berland, K.M., So, P.T.C. and Gratton, E. (1995). Two-photon fluorescence correlation spectroscopy: Method and application to the intracellular environment. *Biophys. J.* **68**, 694–701.
- Berns, M.W. (1976). A possible two-photon effect in vitro using a focused laser beam. *Biophys. J.* **16**, 973–977.
- Bertero, M. and Boccacci, P. (1998). Introduction to Inverse Problems in Imaging. (IOP Publishing, Bristol).
- Bhawalkar, J.D., Kumar, N.D., Zhao, C.F. and Prasad, P.N. (1997). Two-photon photodynamic therapy. *J. Clin. Laser Med. Surg.* **15**, 201–204.
- Bianco, B. and Diaspro, A. (1989). Analysis of the three dimensional cell imaging obtained with optical microscopy techniques based on defocusing. *Cell Biophys.* **15**(3), 189–200.
- Binnig, G., Quate, C.F. and Gerber, C. (1986). Atomic force microscope. *Phys. Rev. Lett.* **56**(9), 930–933.
- Bird, D.K., Eliceiri, K.W., Fan, C.H. and White, J.G. (2004). Simultaneous two-photon spectral and lifetime fluorescence microscopy. *Appl. Opt.* **43**(27), 5173–5182.
- Birge, R.R. (1986). Two-photon spectroscopy of protein-bound fluorophores. *Acc. Chem. Res.* **19**, 138–146.
- Birks, J.B. (1970). Photophysics of Aromatic Molecules. (Wiley Interscience, London).
- Blab, G.A., Lommerse, P.H.M., Cognet, L., Harms, G.S. and Schmidt, T. (2001). Two-photon excitation action cross-sections of the autofluorescent proteins. *Chem. Phys. Lett.* **350**(1–2), 71–77.
- Boccacci, P. and Bertero, M. (2002). Image restoration methods: Basics and algorithms. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), pp. 253–270. (Wiley-Liss, New York).
- Bonetto, P., Boccacci, P., Scarito, M., Davolio, M., Epifani, M., Vicidomini, G., Tacchetti, C., Ramoino, P., Usai, C. and Diaspro, A. (2004). Three-dimensional microscopy migrates to the Web with “PowerUp Your Microscope.” *Microsc. Res. Tech.* **64**(2), 196–203.
- Born, M. and Wolf, E. (1980). *Principles of Optics*, 6th ed. (Cambridge University Press, Cambridge, UK).
- Brakenhoff, G.J., Blom, P. and Barends, P. (1979). Confocal scanning light microscopy with high aperture immersion lenses. *J. Microsc.* **117**, 219–232.

- Brakenhoff, G.J., van Spronsen, E.A., van der Voort, H.T. and Nanninga, N. (1989). Three-dimensional confocal fluorescence microscopy. *Method. Cell. Biol.* **30**, 379–398.
- Brakenhoff, G.J., Muller, M. and Ghauharali, R.I. (1996). Analysis of efficiency of two-photon versus single-photon absorption for fluorescence generation in biological objects. *J. Microsc.* **183**, 140–144.
- Cahalan, M.D., Parker, I., Wei, S.H. and Miller, M.J. (2002). Two-photon tissue imaging: Seeing the immune system in a fresh light. *Nat. Rev. Immunol.* **2**(11), 872–880.
- Callis, P.R. (1997). Two-photon-induced fluorescence. *Annu. Rev. Phys. Chem.* **48**, 271–297.
- Campagnola, P., Mei-de Wei, Lewis, A. and Loew, L. (1999). High-resolution nonlinear optical imaging of live cells by second harmonic generation. *Biophys. J.* **77**, 3341–3351.
- Cannell, M.B. and Soeller, C. (1997). High resolution imaging using confocal and two-photon molecular excitation microscopy. *Proc. R. Microsc. Soc.* **32**, 3–8.
- Cannone, F., Chirico, G., Baldini, G. and Diaspro, A. (2003a). Measurement of the laser pulse width on the microscope objective plane by modulated autocorrelation method. *J. Microsc. (O_{XF})*, **210**(2), 149–157.
- Cannone, F., Chirico, G. and Diaspro, A. (2003b). Two-photon interactions at single fluorescent molecule level. *J. Biomed. Opt.* **8**(3), 391–395.
- Cantor, C.R. and Schimmel, P.R. (1980). *Biophysical Chemistry. Part II: Techniques for the Study of Biological Structure and Function*. (Freeman and Co., New York).
- *Caorsi, V., Ronzitti, E., Vicidomini, G., Krol, S., McConnell, G. and Diaspro, A. (2007). FRET measurements on fuzzy fluorescent nano-structures. *Microsc. Res. Tech.* **70**, 452–458.
- Carlsson, K., Danielsson, P.E., Lenz, R., Liljeborg, A., Majlof, L. and Aslund, N. (1985). Three-dimensional microscopy using a confocal laser scanning microscope. *Opt. Lett.* **10**, 53–55.
- Carrington, W. (2002). Imaging live cells in 3-d using wide field microscopy with image restoration. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 33–346. (Wiley-Liss, New York).
- Caylor, C.L., Dobrianov, I., Kimmer, C., Thorne, R.E., Zipfel, W. and Webb, W.W. (1999). Two-photon fluorescence imaging of impurity distributions in protein crystals. *Phys. Rev. E* **59**, 3831–3834.
- Centonze, V.E. and White, J.G. (1998). Multiphoton excitation provides optical sections from deeper within scattering specimens than confocal imaging. *Biophys. J.* **75**, 2015–2024.
- Chalfie, M. and Kain, S. (Eds.) (1998). *Green Fluorescent Protein. Properties, Applications and Protocols*. (Wiley-Liss, New York).
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–805.
- Chance, B. (1989). *Cell Structure and Function by Microspectrofluorometry*. (Academic Press, New York).
- Cheng, P.C. (Ed.) (1994). *Computer Assisted Multidimensional Microscopies*. (Springer-Verlag, New York).
- Chirico, G., Cannone, F., Beretta, S., Baldini, G. and Diaspro, A. (2001). Single molecule studies by means of the two-photon fluorescence distribution. *Microsc. Res. Tech.* **55**, 359–364.
- Chirico, G., Cannone, F., Baldini, G. and Diaspro, A. (2003a). Two-photon thermal bleaching of single fluorescent molecules. *Biophys. J.* **84**, 588–598.
- Chirico, G., Cannone, F. and Diaspro, A. (2003b). Single molecule photodynamics by means of one- and two-photon approach. *J. Phys. D: Appl. Phys.* **36**, 1–7.

- Chirico, G., Cannone, F., Diaspro, A., Bologna, S., Pellegrini, V., Nifosì, R. and Beltram, F. (2004). Multiphoton switching dynamics of single green fluorescent proteins. *Phys. Rev. E* **70**, 030901.
- Chirico, G., Diaspro, A., Cannone, F., Collini, M., Bologna, S., Pellegrini, V. and Beltram, F. (2005). Selective fluorescence recovery after bleaching of single E2gfp proteins induced by two-photons excitation. *Chemphyschem* **6**(2), 328–335.
- *Chirico, G., Cannone, F. and Diaspro, A. (2006). Unfolding time distribution of GFP by single molecule fluorescence spectroscopy. *Eur. Biophys. J.* **35**, 663–674.
- Cianci, G.C., Wu, J. and Berland, K. (2004). Saturation modified point spread functions in two-photon microscopy. *Microsc. Res. Tech.* **64**(2), 135–141.
- Cox, G. (2002). Biological confocal microscopy. *Materials Today* **5**, No. 3, 34–41.
- Cruz, H.G. and Luscher, C. (2005). Applications of two-photon microscopy in the neurosciences. *Front Biosci.* **10**, 2263–2278.
- Cumpston, B.H., Ananthavel, S.P., Barlow, S., Dyer, D.L., Ehrlich, J.E., Erskine, L.L., Heikal, A.A., Kuebler, S.M., Lee, I.Y.S., McCord-Maughon, D. and others. (1999). Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication. *Nature* **398**(6722), 51–54.
- Daria, V., Blanca, C.M., Nakamura, O., Kawata, S. and Saloma, C. (1998). Image contrast enhancement for two-photon fluorescence microscopy in a turbid medium. *Appl. Opt.* **37**, 7960–7967.
- Davidovits, P.D. and Egger, M.D. (1969). Scanning laser microscope. *Nature* **223**, 831.
- Davidovits, P.D. and Egger, M.D. (1971). Scanning laser microscope for biological investigations. *Appl. Opt.* **10**, 1615–1619.
- de Grauw, K. and Gerritsen, H. (2002). Aberrations and penetration depth in confocal and two-photon microscopy. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 153–170. (Wiley-Liss, New York).
- Denk, W. (1996). Two-photon excitation in functional biological imaging. *J. Biomed. Opt.* **1**, 296–304.
- Denk, W. and Svoboda, K. (1997). Photon upmanship: Why multiphoton imaging is more than a gimmick. *Neuron* **18**, 351–357.
- Denk, W., Strickler, J.H. and Webb, W.W. (1990). Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76.
- Denk, W., Delaney, K.R., Gelperin, A., Kleinfeld, D., Strowbridge, B.W., Tank, D.W. and Yuste, R. (1994). Anatomical and functional imaging of neurons using two-photon laser scanning microscopy. *J. Neurosci. Methods* **54**, 151–162.
- Denk, W., Piston, D. and Webb, W.W. (1995). Two-photon molecular excitation in laser scanning microscopy. In: *Handbook of Confocal Microscopy* (J.B. Pawley, Ed.), 445–457. (Plenum, New York).
- Diaspro, A. (guest editor) (1996). New world microscopy. *IEEE Eng. Med. Biol. Mag.* **15**(1), 29–100.
- Diaspro, A. (1998). Two-photon fluorescence excitation. A new potential perspective in flow cytometry. *Minerva Biotechnol.* **11**(2), 87–92.
- Diaspro, A. (guest editor) (1999a). Two-photon microscopy. *Microsc. Res. Tech.* **47**, 163–212.
- Diaspro, A. (guest editor) (1999b). Two-photon excitation microscopy. *IEEE Eng. Med. Biol. Mag.* **18**(5), 16–99.
- Diaspro, A. (1999c). Two-photon excitation of fluorescence in three-dimensional microscopy. *Eur. J. Histochem.* **43**, 169–178.
- Diaspro, A. (2001). Building a two-photon microscope using a laser scanning confocal architecture. In: *Methods in Cellular Imaging* (A. Periasamy, Ed.), 162–179. (Oxford University Press, New York).

- Diaspro, A. (Ed.) (2002). *Confocal and Two-Photon Microscopy: Foundations, Applications, and Advances*. (Wiley-Liss, New York).
- Diaspro, A. (2004). Rapid dissemination of two-photon excitation microscopy prompts new applications. *Microsc. Res. Tech.* **63**(1), 1–2.
- *Diaspro, A. (2007). Introduction: advanced multiphoton and fluorescence lifetime imaging techniques. *Microsc. Res. Tech.* **70**, 397. Special issue.
- Diaspro, A. and Chirico, G. (2003). Two-photon excitation microscopy. *Adv. Imaging Elect. Phys.* **126**, 195–286.
- Diaspro, A. and Robello, M. (2000). Two-photon excitation of fluorescence for three-dimensional optical imaging of biological structures. *J. Photochem. Photobiol. B* **55**, 1–8.
- Diaspro, A. and Sheppard, C.J.R. (2002). Two-photon excitation microscopy: Basic principles and architectures. In: *Confocal and Two-Photon Microscopy: Foundations, Applications, and Advances* (A. Diaspro, Ed.), 39–74. (Wiley-Liss, New York).
- Diaspro, A., Sartore, M. and Nicolini, C. (1990). Three-dimensional representation of biostructures imaged with an optical microscope: I. Digital optical sectioning. *Image Vision Comp.* **8**, 130–141.
- *Diaspro, A., Chirico, G. and Collini, M. (2005). Two-photon fluorescence excitation and related techniques in biological microscopy. *Q. Rev. Biophys.* **38**, 97–166.
- Diaspro, A., Beltrame, F., Fato, M., Palmeri, A. and Ramoino, P. (1997). Studies on the structure of sperm heads of *eledone cirrhosa* by means of CLSM linked to bioimage-oriented devices. *Microsc. Res. Tech.* **36**, 159–164.
- Diaspro, A., Annunziata, S., Raimondo, M. and Robello, M. (1999a). Three-dimensional optical behaviour of a confocal microscope with single illumination and detection pinhole through imaging of subresolution beads. *Microsc. Res. Tech.* **45**(2), 130–131.
- Diaspro, A., Corosu, M., Ramoino, P. and Robello, M. (1999b). Adapting a compact confocal microscope system to a two-photon excitation fluorescence imaging architecture. *Microsc. Res. Tech.* **47**, 196–205.
- Diaspro, A., Annunziata, S. and Robello, M. (2000). Single-pinhole confocal imaging of sub-resolution sparse objects using experimental point spread function and image restoration. *Microsc. Res. Tech.* **51**, 464–468.
- Diaspro, A., Chirico, G., Federici, F., Cannone, F., Beretta, S. and Robello, M. (2001). Two-photon microscopy and spectroscopy based on a compact confocal scanning head. *J. Biomed. Opt.* **6**, 300–310.
- Diaspro, A., Federici, F. and Robello, M. (2002a). Influence of refractive-index mismatch in high-resolution three-dimensional confocal microscopy. *Appl. Opt.-OT* **41**, 685–690.
- Diaspro, A., Silvano, D., Krol, S., Cavalleri, O. and Gliozzi, A. (2002b). Single living cell encapsulation in nano-organized polyelectrolyte shells. *Langmuir* **18**, 5047–5050.
- Diaspro, A., Boccacci, P., Bonetto, P., Scarito, M., Davolio, M. and Epifani, M. (2002c). “Power-up your Microscope.” www.powermicroscope.com.
- Diaspro, A., Fronte, P., Raimondo, M., Fato, M., De Leo, G., Beltrame, F., Cannone, F., Chirico, G. and Ramoino, P. (2002d). Functional imaging of living paramecium by means of confocal and two-photon excitation fluorescence microscopy. In: *Functional Imaging* (D. Farkas, Ed.). Proc. SPIE 4622, 47–53.
- Diaspro, A., Federici, F., Viappiani, C., Krol, S., Pisciotta, M., Chirico, G., Cannone, F. and Gliozzi, A. (2003). Two-photon photolysis of 2-nitrobenzaldehyde monitored by fluorescent-labeled nanocapsules. *J. Phy. Chem. B* **107**(40), 11008–11012.

- *Diaspro, A., Bianchini, P., Vicidomini, G., Faretta, M., Ramoino, P. and Usai, C. (2006). Multi-photon excitation microscopy. *Biomed. Eng. Online*. **5**, 36.
- *Diaspro, A., Krol, S., Campanini, B., Cannone, F. and Chirico, G. (2006). Enhanced Green Fluorescent Protein (GFP) fluorescence after polyelectrolyte caging. *Opt. Express* **14**, 9815–9824.
- Difato, F., Mazzone, F., Scaglione, S., Fato, M., Beltrame, F., Kubinova, L., Janacek, J., Ramoino, P., Vicidomini, G. and Diaspro, A. (2004). Improvement in volume estimation from confocal sections after image deconvolution. *Microsc. Res. Tech.* **64**(2), 151–155.
- Dürig, U. and Pohl, D.W. (1986). Near-field optical-scanning microscopy. *J. Appl. Phys.* **59**(10), 3318–3327.
- Duveneck, G.L., Bopp, M.A., Ehrat, M., Haiml, M., Keller, U., Bader, M.A., Marowsky, G. and Soria, S. (2001). Evanescent-field-induced two-photon fluorescence: Excitation of macroscopic areas of planar waveguides. *Appl. Phys. B* **73**, 869–871.
- Esposito, A., Federici, F., Usai, C., Cannone, F., Chirico, G., Collini, M. and Diaspro, A. (2004). Notes on theory and experimental conditions behind two-photon excitation microscopy. *Microsc. Res. Tech.* **63**, 12–17.
- Faisal, F.H.M. (1987). *Theory of Multiphoton Processes*. (Plenum Press, New York).
- Farrer, R.A., Previte, M.J.R., Olson, C.E., Peyser, L.A., Fourkas, J.T. and So, P. T.C. (1999). Single molecule detection with a two-photon fluorescence microscope with fast scanning capabilities and polarization sensitivity. *Opt. Lett.* **24**, 1832–1834.
- Fay, F.S., Carrington, W. and Fogarty, K.E. (1989). Three-dimensional molecular distribution in single cells analyzed using the digital imaging microscope. *J. Microsc.* **153**, 133–149.
- Feynman, R.P. (1985). *QED: The Strange Theory of Light and Matter*. (Princeton University Press, Princeton, NJ).
- Fisher, W.G., Watcher, E.A., Armas, M. and Seaton, C. (1997). Titanium: sapphire laser as an excitation source in two-photon spectroscopy. *Appl. Spectrosc.* **51**, 218–226.
- Ford, B.J. (1991). *The Leeuwenhoek Legacy*. (Biopress and Parrand, Bristol).
- Franken, P.A., Hill, A.E., Peters, C.W. and Weinreich, G. (1961). Generation of optical harmonics. *Phys. Rev. Lett.* **7**, 118–119.
- French, T., So, P.T.C., Weaver, D.J., Coelho-Sampaio, T. and Gratton, E. (1997). Two-photon fluorescence lifetime imaging microscopy of macrophage-mediated antigen processing. *J. Microsc.* **185**, 339–353.
- Friedrich, D.M. (1982). Two-photon molecular spectroscopy. *J. Chem. Educ.* **59**, 472–483.
- Friedrich, D.M. and McClain, W.M. (1980). Two-photon molecular electronic spectroscopy. *Annu. Rev. Phys. Chem.* **31**, 559–577.
- Gannaway, J.N. and Sheppard, C.J.R. (1978). Second harmonic imaging in the scanning optical microscope. *Opt. Quant. Electron.* **10**, 435–439.
- Gauderon, R., Lukins, R.B. and Sheppard, C.J.R. (1999). Effects of a confocal pinhole in two-photon microscopy. *Microsc. Res. Tech.* **47**, 210–214.
- Gerton, J.M., Wade, L.A., Lessard, G.A., Ma, Z. and Quake, S.R. (2004). Tip-enhanced fluorescence microscopy at 10 nanometer resolution. *Phy. Rev. Lett.* **93**(18), 180801.
- Girkin, J. and Wokosin, D. (2002). Practical multiphoton microscopy. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 207–236. (Wiley-Liss, New York).
- Göppert-Mayer, M. (1931). Über Elementarakte mit zwei Quantensprüngen. *Ann. Phys.* **9**, 273–295.
- Gosnell, T.R. and Taylor, A.J. (Eds.) (1991). *Selected Papers on Ultrafast Laser Technology*. SPIE Milestone Series. (SPIE Press, Bellingham, WA).

- Gostkowski, M.L., Allen, R., Plenert, M.L., Okerberg, E., Gordon, M.J. and Shear, J.B. (2004). Multiphoton-excited serotonin photochemistry. *Biophys. J.* **86**(5), 3223–3229.
- Gratton, E. and van de Ven, M.J. (1995). Laser sources for confocal microscopy. In: *Handbook of Confocal Microscopy* (J.B. Pawley, Ed.), 69–97. (Plenum, New York).
- Gratton, E., Barry, N.P., Beretta, S. and Celli, A. (2001). Multiphoton fluorescence microscopy. *Methods* **25**, 103–110.
- Gu, M. and Sheppard, C.J.R. (1995). Comparison of three-dimensional imaging properties between two-photon and single-photon fluorescence microscopy. *J. Microsc.* **177**, 128–137.
- Gu, M., Gan, X., Kisteman, A. and Xu, M.G. (2000). Comparison of penetration depth between two-photon excitation and single-photon excitation in imaging thorough turbid tissue media. *Appl. Phys. Lett.* **77**(10), 1551–1553.
- Guild, J.B., Xu, C. and Webb, W.W. (1997). Measurement of group delay dispersion of high numerical aperture objective lenses using two-photon excited fluorescence. *Appl. Opt.* **36**, 397–401.
- Hamamatsu Photonics K.K. (1999). *Photomultiplier Tubes: Basics and Applications*, 2nd ed. (Hamamatsu Photonics K.K., Japan).
- Hanninen, P.E. and Hell, S.W. (1994). Femtosecond pulse broadening in the focal region of a two-photon fluorescence microscope. *Bioimaging* **2**, 117–121.
- Harper, I.S. (2002). Fluorophores and their labeling procedures for monitoring various biological signals. In: *Methods in Cellular Imaging* (A. Periasamy, Ed.), 20–39. (Oxford University Press, New York).
- Haughland, P.R. (Ed.) (2002). *Handbook of Fluorescent Probes and Research Chemicals*. (Molecular Probes, Eugene, OR).
- Heinze, K.G., Jahnz, M. and Schwill, P. (2004). Triple-color coincidence analysis: One step further in following higher order molecular complex formation. *Biophys. J.* **86**(1), 506–516.
- Hell, S.W. (guest editor) (1996). Nonlinear optical microscopy. *Bioimaging* **4**, 121–172.
- Hell, S.W. (2003). Toward fluorescence nanoscopy. *Nat. Biotechnol.* **21**, 1347–1355.
- Hell, S.W., Bahlmann, K., Schrader, M., Soini, A., Malak, H., Gryczynski, I. and Lakowicz, J.R. (1996). Three-photon excitation in fluorescence microscopy. *J. Biomed. Opt.* **1**, 71–74.
- Hellwarth, R. and Chistensen, P. (1974). Nonlinear optical microscopic examination of structures in polycrystalline ZnSe. *Opt. Commun.* **12**, 318–322.
- Herman, B. and Tanke, H.J. (1998). *Fluorescence Microscopy*. (Springer-Verlag, New York).
- Hooke, R. (1961). *Micrographia* (facsimile). (Dover, New York).
- Hopt, A. and Neher, E. (2001). Highly nonlinear photodamage in two-photon fluorescence microscopy. *Biophys. J.* **80**, 2029–2036.
- Iyer, V., Hoogland, T.M., Losavio, B.E., McQuiston, A.R. and Saggau, P. (2002). Compact two-photon laser scanning microscope made from minimally modified commercial components. In: *Multiphoton Microscopy in the Biomedical Sciences II* (A. Periasamy, P.T.C. So, Eds.). Proc. SPIE 4620, 274–280.
- Jaiswal, J.K., Goldman, E.R., Mattoussi, H. and Simon, S.M. (2004). Use of quantum dots for live cell imaging. *Nat. Methods* **1**(1), 73–78.
- Jonkman, J. and Stelzer, E. (2002). Resolution and contrast in confocal and two-photon microscopy. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 101–126. (Wiley-Liss, New York).
- Jung, J.C., Mehta, A.D., Aksay, E., Stepnoski, R. and Schnitzer, M.J. (2004). In vivo mammalian brain imaging using one- and two-photon fluorescence microendoscopy. *J. Neurophysiol.* **92**(5), 3121–3133.

- Kaiser, W. and Garrett, C.G.B. (1961). Two-photon excitation in $\text{CaF}_2\text{:Eu}^{2+}$. *Phys. Rev. Lett.* **7**, 229–231.
- Kappel, C., Selle, A., Fricke-Begemann, T., Bader, M.A. and Marowsky, G. (2004). Giant enhancement of two-photon fluorescence induced by resonant double grating waveguide structures. *Appl. Phys. B-Lasers Opt.* **79**(5), 531–534.
- Kawata, S., Sun, H.-B., Tanaka, T. and Takada, K. (2001). Finer features for functional microdevices. *Nature* **412**, 697–698.
- Kino, G.S. and Corle, T.R. (1989). Confocal scanning optical microscopy. *Phys. Today* **42**, 55–62.
- Koana, Z. (1943). *J. Illumination Eng. Inst.* **26**, 371.
- Koester, H.J., Baur, D., Uhl, R. and Hell, S.W. (1999). Ca^{2+} fluorescence imaging with pico- and femtosecond two-photon excitation: Signal and photodamage. *Biophys. J.* **77**, 2226–2236.
- König, K. (2000). Multiphoton microscopy in life sciences. *J. Microsc.* **200**, 83–104.
- König, K. and Tirlapur, U.K. (2002). Cellular and subcellular perturbations during multiphoton microscopy. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 191–206. (Wiley-Liss, New York).
- König, K., Liang, H., Berns, M.W. and Tromberg, B.J. (1995). Cell damage by near-IR microbeams. *Nature* **377**, 20–21.
- König, K., Krasieva, T., Bauer, E., Fiedler, U., Berns, M.W., Tromberg, B.J. and Greulich, K.O. (1996a). Cell damage by UVA radiation of a mercury microscopy lamp probed by autofluorescence modifications, cloning assay and comet assay. *J. Biomed. Opt.* **1**, 217–222.
- König, K., So, P.T.C., Mantulin, W.W., Tromberg, B.J. and Gratton, E. (1996b). Two-photon excited lifetime imaging of autofluorescence in cells during UVA and NIR photostress. *J. Microsc.* **183**, 197–204.
- König, K., Liang, H., Berns, M.W. and Tromberg, B.J. (1996c). Cell damage in near infrared multimode optical traps as a result of multiphoton absorption. *Opt. Lett.* **21**, 1090–1092.
- König, K., Boehme, S., Leclerc, N. and Ahuja, R. (1998). Time-gated autofluorescence microscopy of motile green microalgae in an optical trap. *Cell. Mol. Biol.* **44**, 763–770.
- König, K., Riemann, I., Fischer, P. and Halbhauer, K.J. (1999). Intracellular nanosurgery with near infrared femtosecond laser pulses. *Cell. Mol. Biol.* **45**, 195–201.
- Kriete, A. (Ed.) (1992). *Visualization in Biomedical Microscopies*. (VCH, Weinheim).
- Lakowicz, J.R. and Gryczynski, I. (1992). Tryptophan fluorescence intensity and anisotropy decays of human serum albumin resulting from one-photon and two-photon excitation. *Biophys. Chem.* **45**, 1–6.
- Loudon, R. (1983). *The Quantum Theory of Light*. (Oxford University Press, London).
- Louisell, W.H. (1973). *Quantum Statistical Properties of Radiation*. (Wiley, New York).
- Mainen, Z.F., Malectic-Savic, M., Shi, S.H., Hayashi, Y., Malinow, R. and Svoboda, K. (1999). Two-photon imaging in living brain slices. *Methods* **18**, 231–239.
- Maiti, S., Shear, J.B., Williams, R.M., Zipfel, W.R. and Webb, W.W. (1997). Measuring serotonin distribution in live cells with three-photon excitation. *Science* **275**, 530–532.
- Majewska, A., Yiu, G. and Yuste, R. (2000). A custom-made two-photon microscope and deconvolution system. *Pflugers Arch. Eur. J. Physiol.* **441**(2/3), 398–408.

- Masters, B.R. (1996). *Selected Papers on Confocal Microscopy*. SPIE Milestone Series. (SPIE Press, Bellingham, WA).
- Masters, B.R. (2002). *Selected Papers on Multi-photon Excitation Microscopy*. SPIE Milestone Series. (SPIE Press, Bellingham, WA).
- Masters, B.R., So, P.T. and Gratton, E. (1997). Multiphoton excitation fluorescence microscopy and spectroscopy of in vivo human skin. *Biophys J.* **72**, 2405–2412.
- *Mazza, D., Cella, F., Vicidomini, G., Krol, S. and Diaspro, A. (2007). Role of three-dimensional bleach distribution in confocal and two-photon fluorescence recovery after photobleaching experiments. *Appl. Opt.* **46**, in press.
- McConnell, G. and Riis, E. (2004). Two-photon laser scanning fluorescence microscopy using photonic crystal fiber. *J. Biomed. Opt.* **9**(5), 922–927.
- Mertz, J., Xu, C. and Webb, W.W. (1995). Single molecule detection by two-photon excited fluorescence. *Opt. Lett.* **20**, 2532–2534.
- Miller, M.J., Wei, S.H., Cahalan, M.D. and Parker, I. (2003). Autonomous T cell trafficking examined in vivo with intravital two-photon microscopy. *Proc. Natl. Acad. Sci. USA* **100**(5), 2604–2609.
- Minsky, M. (1988). Memoir of inventing the confocal scanning microscope. *Scanning* **10**, 128–138.
- *Mondal, P.P., Diaspro, A. (2007). Reduction of higher-order photobleaching in two-photon excitation microscopy. *Phys Rev E Stat Nonlin Soft Matter Phys.* **75**, 061904.
- Moscatelli, F.A. (1986). A simple conceptual model for two-photon absorption. *Am. J. Phys.* **54**, 52–54.
- Mueller, M., Squier, J., Wilson, K.R. and Brakenhoff, G.J. (1998). 3D microscopy of transparent objects using third-harmonic generation. *J. Microsc.* **191**, 266–274.
- Murphy, D.B. (2001). *Fundamentals of Light Microscopy and Electronic Imaging*. (Wiley-Liss, New York).
- Nakamura, O. (1993). Three-dimensional imaging characteristics of laser scan fluorescence microscopy: Two-photon excitation vs. single-photon excitation. *Optik* **93**, 39–42.
- Nakamura, O. (1999). Fundamentals of two-photon microscopy. *Microsc. Res. Tech.* **47**, 165–171.
- Naora, H. (1951). Microspectrophotometry and cytochemical analysis of nucleic acids. *Science* **114**, 279–280.
- Nemet, B.A., Nikolenko, V. and Yuste, R. (2004). Second harmonic imaging of membrane potential of neurons with retinal. *J. Biomed. Opt.* **9**(5), 873–881.
- Novotny, L., Sanchez, E.J. and Xie, X.S. (1998). Near-field optical imaging using metal tips illuminated by higher-order Hermite-Gaussian beams. *Ultramicroscopy* **71**(1–4), 21–29.
- Patterson, G.H. and Lippincott-Schwarz, J.A. (2002). Photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **297**, 1873.
- Patterson, G.H. and Piston, D.W. (2000). Photobleaching in two-photon excitation microscopy. *Biophys. J.* **78**, 2159–2162.
- Pawley, J.B. (Ed.) (1995). *Handbook of Biological Confocal Microscopy*. (Plenum Press, New York).
- Pennisi, E. (1997). Biochemistry: Photons add up to better microscopy. *Science* **275**, 480–481.
- Periasamy, A. (Ed.) (2001). *Methods in Cellular Imaging*. (Oxford University Press, New York).
- Periasamy, A. and Diaspro, A. (2003). Multiphoton microscopy. *J. Biomed. Opt.* **8**(3), 327–328.
- Periasamy, A., Skoglund, P., Noakes, C. and Keller, R. (1999). An evaluation of two-photon excitation versus confocal and digital deconvolution fluores-

- cence microscopy imaging in *Xenopus* morphogenesis. *Microsc. Res. Tech.* **47**, 172–181.
- Periasamy, A., Noakes, C., Skoglund, P., Keller, R. and Sutherland, A.E. (2002). Two-photon excitation fluorescence microscopy imaging in *Xenopus* and transgenic mouse embryos. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 271–284. (Wiley-Liss, New York).
- Petran, M., Hadravsky, M., Egger, M.D. and Galambos, R. (1968). Tandem-scanning reflected-light microscope. *J. Opt. Soc. Am.* **58**, 661–664.
- Piston, D.W. (1999). Imaging living cells and tissues by two-photon excitation microscopy. *Trends Cell Biol.* **9**, 66–69.
- Piston, D.W. (2005). When two is better than one: Elements of intravital microscopy. *PLoS Biol.* **3**(6), e207. Epub June 14, 2005.
- Piston, D.W., Masters, B.R. and Webb, W.W. (1995). Three-dimensionally resolved NAD(P)H cellular metabolic redox imaging of the *in situ* cornea with two-photon excitation laser scanning microscopy. *J. Microsc.* **178**, 20–27.
- Post, J.N., Lidke, K.A., Rieger, B. and Arndt-Jovin, D.J. (2005). One- and two-photon photoactivation of a paGFP-fusion protein in live *Drosophila* embryos. *FEBS Lett.* **579**(2), 325–330.
- Potter, S.M. (1996). Vital imaging: Two-photons are better than one. *Curr. Biol.* **6**, 1596–1598.
- Potter, S.M., Wang, C.M., Garrity, P.A. and Fraser, S.E. (1996). Intravital imaging of green fluorescent protein using two-photon laser-scanning microscopy. *Gene* **173**, 25–31.
- Rentzepis, P.M., Mitschele, C.J. and Saxman, A.C. (1970). Measurement of ultrashort laser pulses by three-photon fluorescence. *Appl. Phys. Lett.* **17**, 122–124.
- Robinson, J.P. (2001). *Current Protocols in Cytometry*. (John Wiley, New York).
- Ruan, Q., Cheng, M.A., Levi, M., Gratton, E. and Mantulin, W.W. (2004). Spatial-temporal studies of membrane dynamics: Scanning fluorescence correlation spectroscopy (SFCS). *Biophys. J.* **87**(2), 1260–1267.
- Rudolf, R., Mongillo, M., Magalhaes, P.J. and Pozzan, T. (2004). In vivo monitoring of Ca²⁺ uptake into mitochondria of mouse skeletal muscle during contraction. *J. Cell Biol.* **166**(4), 527–536.
- Ruska, E. and Knoll, M. (1931). Die magnetische Sammelspule fuer schnelle Elektronenstrahlen. *Z. Tech. Phys.* **12**(389–400 and 488).
- Saloma, C., Saloma-Palmes, C. and Kondoh, H. (1998). Site-specific confocal fluorescence imaging of biological microstructures in a turbid medium. *Phys. Med. Biol.* **43**, 1741.
- Sánchez, E.J., Novotny, L. and Xie, X.S. (1999). Near-field fluorescence microscopy based on two-photon excitation with metal tips. *Phys. Rev. Lett.* **82**(20), 4014–4017.
- Schneider, M., Barozzi, S., Testa, I., Faretta, M. and Diaspro, A. (2005). Two-photon activation and excitation properties of Pa-Gfp in the 720–920 nm region. *Biophys. J.* **89**(2), 1346–1352.
- Schonle, A. and Hell, S.W. (1998). Heating by absorption in the focus of an objective lens. *Opt. Lett.* **23**, 325–327.
- Schwille, P. (2001). Fluorescence correlation spectroscopy and its potential for intracellular applications. *Cell Biochem. Biophys.* **34**, 383–405.
- Schwille, P., Haupts, U., Maiti, S. and Webb, W.W. (1999). Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophys. J.* **77**, 2251–2265.
- Schwille, P., Kummer, S., Heikal, A.A., Moerner, W.E. and Webb, W.W. (2000). Fluorescence correlation spectroscopy reveals fast optical excitation-driven intramolecular dynamics of yellow fluorescent proteins. *Proc. Natl. Acad. Sci. USA* **97**, 151–156.

- Sheppard, C.J.R. (2002). The generalized microscope. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 1–18. (Wiley-Liss, New York).
- Sheppard, C.J.R. and Choudhury, A. (1977). Image formation in the scanning microscope. *Opt. Acta* **24**, 1051–1073.
- Sheppard, C.J.R. and Gu, M. (1990). Image formation in two-photon fluorescence microscopy. *Optik* **86**, 104–106.
- Sheppard, C.J.R. and Kompfner, R. (1978). Resonant scanning optical microscope. *Appl. Opt.* **17**, 2879–2882.
- Sheppard, C.J.R. and Shotton, D.M. (1997). *Confocal Laser Scanning Microscopy*. (BIOS, Oxford, UK).
- Sheppard, C.J.R. and Wilson, T. (1980). Image formation in confocal scanning microscopes. *Optik* **55**, 331–342.
- Sheppard, C.J.R., Kompfner, R., Gannaway, J. and Walsh, D. (1977). The scanning harmonic optical microscope. *IEEE/OSA Conf. Laser Eng. Appl.*, (Washington, DC).
- Shih, Y.H., Strekalov, D.V., Pittman, T.D. and Rubin, M.H. (1998). Why two-photon but not two photons? *Fortschr. Phys.* **46**, 627–641.
- Shotton, D.M. (Ed.) (1993). *Electronic Light Microscopy: Techniques in Modern Biomedical Microscopy, Vol. 1*. (Wiley-Liss, New York).
- Shotton, D.M. (1995). Electronic light microscopy—present capabilities and future prospects. *Histochem. Cell Biol.* **104**, 97–137.
- Singh, S. and Bradley, L.T. (1964). Three-photon absorption in naphthalene crystals by laser excitation. *Phys. Rev. Lett.* **12**, 162–164.
- So, P.T.C., Berland, K.M., French, T., Dong, C.Y. and Gratton, E. (1966). Two-photon fluorescence microscopy: Time resolved and intensity imaging. In: *Fluorescence Imaging Spectroscopy and Microscopy* (X.F. Wang, B. Herman, Eds.). Chemical Analysis Series, Vol. 137, 351–373. (John Wiley, New York).
- So, P.T.C., Dong, C.Y., Masters, B.R. and Berland, K.M. (2000). Two-photon excitation fluorescence microscopy. *Annu. Rev. Biomed. Eng.* **2**, 399–429.
- So, P.T.C., Kim, K.H., Buehler, C., Masters, B.R., Hsu, L. and Dong, C.Y. (2001). Basic principles of multi-photon excitation microscopy. In: *Methods in Cellular Imaging* (A. Periasamy, Ed.), 152–161. (Oxford University Press, New York).
- Soeller, C. and Cannell, M.B. (1996). Construction of a two-photon microscope and optimisation of illumination pulse duration. *Pfluegers Arch. Eur. J. Physiol.* **432**, 555–561.
- Soeller, C. and Cannell, M.B. (1999). Two-photon microscopy: Imaging in scattering samples and three-dimensionally resolved flash photolysis. *Microsc. Res. Tech.* **47**, 182–195.
- Sonnleitner, M., Schutz, G.J. and Schmidt, T. (1999). Imaging individual molecules by two-photon excitation. *Chem. Phys. Lett.* **300**, 221–226.
- Sonnleitner, M., Schutz, G., Kada, G. and Schindler, H. (2000). Imaging single lipid molecules in living cells using two-photon excitation. *Single Mol.* **1**, 182–183.
- Spence, D.E., Kean, P.N. and Sibbett, W. (1991). 60-fsec pulse generation from a self-mode-locked Ti:sapphire laser. *Opt. Lett.* **16**, 42–45.
- Squier, J.A., Muller, M., Brakenhoff, G.J. and Wilson, K.R. (1998). Third harmonic generation microscopy. *Opt. Express* **3**, 315–324.
- Stelzer, E.H.K., Hell, S., Lindek, S., Pick, R., Storz, C., Stricker, R., Ritter, G. and Salmon, N. (1994). Non-linear absorption extends confocal fluorescence microscopy into the ultra-violet regime and confines the illumination volume. *Opt. Commun.* **104**, 223–228.

- Straub, M. and Hell, S.W. (1998). Fluorescence lifetime three-dimensional microscopy with picosecond precision using a multifocal multiphoton microscope. *App. Phys. Lett.* **73**, 1769–1771.
- Svelto, O. (1998). *Principles of Lasers*, 4th ed. (Plenum, New York).
- Sytsma, J., Vroom, J.M., De Grauw, C.J. and Gerritsen, H.C. (1998). Time-gated fluorescence lifetime imaging and microvolume spectroscopy using two-photon excitation. *J. Microsc.* **191**, 39–51.
- Tan, Y.P., Llano, I., Hopt, A., Wurriehausen, F. and Neher, E. (1999). Fast scanning and efficient photodetection in a simple two-photon microscope. *J. Neurosci. Methods* **92**, 123–135.
- Tirlapur, U.K. and König, K. (2002). Two-photon near infrared femtosecond laser scanning microscopy in plant biology. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 449–468. (Wiley-Liss, New York).
- Torok, P. and Sheppard, C.J.R. (2002). The role of pinhole size in high aperture two and three-photon microscopy. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 127–152. (Wiley-Liss, New York).
- Tsien, R.Y. (1995). Design of Molecules to Probe Living Cells. Abstracts of Papers of the American Chemical Society 209:4-SOCED.
- Tyrrell, R.M. and Keyse, S.M. (1990). The interaction of UVA radiation with cultured cells. *J. Photochem. Photobiol. B* **4**, 349–361.
- *Vidomini, G., Schneider, M., Bianchini, P., Krol, S., Szellas, T. and Diaspro, A. (2007). Characterization of uniform ultrathin layer for z-response measurements in three-dimensional section fluorescence microscopy. *J. Microsc.* **225**, 88–95.
- Volkman, J. (1996). Ernst Abbe and his work. *Appl. Opt.* **5**, 1720–1731.
- Wang, X.F. and Herman, B. (1996). *Fluorescence Imaging Spectroscopy and Microscopy*. (Wiley-Liss, New York).
- Wang, Y., Wang, X.F., Wang, C. and Ma, H. (2004). Simultaneously multi-parameter determination of hematocrit cell apoptosis by two-photon and confocal laser scanning microscopy. *J. Clin. Lab. Anal.* **18**(5), 271–275.
- Webb, R.H. (1996). Confocal optical microscopy. *Rep. Progr. Phys.* **59**, 427–471.
- Weinstein, M. and Castleman, K.R. (1971). Reconstructing 3-D specimens from 2-D section images. *Proc. SPIE* **26**, 131–138.
- White, J.G., Amos, W.B. and Fordham, M. (1987). An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy. *J. Cell Biol.* **105**, 41–48.
- White, N.S. and Errington, R.J. (2000). Improved laser scanning fluorescence microscopy by multiphoton excitation. *Adv. Imag. Elect. Phys.* **113**, 249–277.
- Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Rocker, C., Salih, A., Spindler, K.-D. and Nienhaus, G.U. (2004). EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proc. Natl. Acad. Sci. USA* **101**(45), 15905–15910.
- Wier, W.G., Balke, C.W., Michael, J.A. and Mauban, J.R. (2000). A custom confocal and two-photon digital laser scanning microscope. *Am. J. Physiol.* **278**, H2150–H2156.
- Wilson, T. (2002). Confocal microscopy: Basic principles and architectures. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 19–38. (Wiley-Liss, New York).
- Wilson, T. and Sheppard, C.J.R. (1984). *Theory and Practice of Scanning Optical Microscopy*. (Academic Press, London).
- Wise, F. (1999). Lasers for two-photon microscopy. In: *Imaging: A Laboratory Manual* (R., Yuste, F., Lanni, A. Konnerth, Eds.), 18.1–9. (Cold Spring Harbor Press, Cold Spring Harbor, NY).

- Wiseman, P.W., Squier, J.A., Ellisman, M.H. and Wilson, K.R. (2000). Two-photon image correlation spectroscopy and image cross-correlation spectroscopy. *J. Microsc.* **200**, 14–25.
- Wiseman, P.W., Capani, F., Squier J.A. and Martone, M.E. (2002). Counting dendritic spines in brain tissue slices by image correlation spectroscopy analysis. *J. Microsc. (Oxf)* **205**, 177–186.
- Wokosin, D.L. and White, J.G. (1997). Optimization of the design of a multiphoton excitation laser scanning fluorescence imaging system. In: *Three-Dimensional Microscopy: Image, Acquisition and Processing IV*. Proc. SPIE 2984, 25–29.
- Wokosin, D.L., Centonze, V.E., White, J., Armstrong, D., Robertson, G. and Ferguson, A.I. (1996). All-solid-state ultrafast lasers facilitate multiphoton excitation fluorescence imaging. *IEEE J. Selected Top. Quant. Electron.* **2**, 1051–1065.
- Wokosin, D.L., Amos, W.B. and White, J.G. (1998). Detection sensitivity enhancements for fluorescence imaging with multiphoton excitation microscopy. *Proc. IEEE Eng. Med. Biol. Soc.* **20**, 1707–1714.
- Wokosin, D.L., Loughrey, C.M. and Smith, G.L. (2004). Characterization of a range of Fura dyes with two-photon excitation. *Biophys. J.* **86**(3), 1726–1738.
- Wolleschensky, R., Feurer, T., Sauerbrey, R. and Simon, U. (1998). Characterization and optimization of a laser scanning microscope in the femtosecond regime. *Appl. Phys. B* **67**, 87–94.
- Wolleschensky, R., Dickinson, M. and Fraser, S.E. (2002). Group velocity dispersion and fiber delivery in multiphoton laser scanning microscopy. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 171–190. (Wiley-Liss, New York).
- Xie, X.S. and Lu, H.P. (1999). Single molecule enzymology. *J. Biol. Chem.* **274**, 15967–15970.
- Xu, C. (2002). Cross-sections of fluorescence molecules used in multiphoton microscopy. In: *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances* (A. Diaspro, Ed.), 75–100. (Wiley-Liss, New York).
- Xu, C., Guild J., Webb, W.W. and Denk, W. (1995). Determination of absolute two-photon excitation cross sections by *in situ* second-order autocorrelation. *Opt. Lett.* **20**, 2372–2374.
- Yoder, E.J. and Kleinfeld, D. (2002). Cortical imaging through the intact mouse skull using two-photon excitation laser scanning microscopy. *Microsc. Res. Tech.* **56**(4), 304–305.
- Zimmer, M. (2002). Green fluorescence protein (GFP): Applications, structure, and related photophysical behavior. *Chem. Rev.* **102**(3), 759–781.
- Zoumi, A., Yeh, A. and Tromberg, B.J. (2002). Imaging cells and extracellular matrix *in vivo* by using second-harmonic generation and two-photon excited fluorescence. *Proc. Natl. Acad. Sci. USA* **99**(17), 11014–11019.
- Zoumi, A., Lu, X., Kassab, G.S. and Tromberg, B.J. (2004). Imaging coronary artery microstructure using second-harmonic and two-photon fluorescence microscopy. *Biophys. J.* **87**(4), 2778–2786.

*References added since the first printing.

Nanoscale Resolution in Far-Field Fluorescence Microscopy

Stefan W. Hell and Andreas Schönle

1 Introduction

The discovery of the diffraction barrier in 1873 by Ernst Abbe (1873) has shown that, relying on propagating light waves and regular lenses, the traditional light microscope cannot resolve spatial structures that are smaller than about half the wavelength of the focused light. This physical insight has been accepted as an unalterable limitation of focusing light microscopy and has consequently triggered the invention of non-optical imaging techniques, such as electron and scanning probe microscopy. In spite of the tremendous improvement in resolution brought about by these methods, light microscopy has maintained its importance in many fields of science. The reasons are mainly a number of exclusive advantages, the most prominent of which is the ability to noninvasively image (living) specimens. Light microscopy also entails the possibility of using fluorescence as a highly specific signature of the specimen features of interest. Fluorescence is particularly attractive when provided by endogenous fluorescence markers, i.e., proteins in physiologically intact cells. Mapped with a confocal or multiphoton excitation microscope (Sheppard and Kompfner, 1978; Wilson and Sheppard, 1984; Denk et al., 1990), fluorescence emission readily yields protein three-dimensional (3D) distributions, or that of other fluorescently labeled molecules from the strongly convoluted inside of biological specimens.

Abandoning the concept of focusing light altogether, near-field optical microscopy is the earliest practically relevant attempt to overcome the diffraction barrier with visible light (Pohl and Courjon, 1993). To this end, scanning near-field optical microscopes employ ultrasharp tips or tiny apertures to confine the interaction of the light field with the object to subdiffraction dimensions. However, applying this technique to soft biological matter has not been very successful to date and it has to be applied carefully to avoid imaging artifacts (Hecht et al., 1997). In any case, a near-field optical microscope is confined to imaging surfaces, again underscoring the importance of improving the resolution in an optical microscope that still preserves focusing. For

convenience we refer to the traditional light microscope as a “far-field” instrument.

This formidable problem has been approached by many scientists (Toraldo di Francia, 1952; Lukosz, 1966), but only in recent years have methods been characterized (Hell and Kroug, 1995; Hell, 1997, 2003) that effectively broke Abbe’s diffraction barrier, implying the potential to attain molecular resolution with regular lenses and visible light (Hell et al., 2003; Hell, 2003; Westphal and Hell, 2005).

2 The Resolution Limit

The limited resolution of a focusing light microscope is readily described by the form and extent of the effective focal spot, commonly referred to as the (effective) point spread function (PSF). The PSF is the image that an infinitely small object would create. For an incoherent imaging mode such as fluorescence, the image is therefore given by the convolution of the object with the PSF:

$$I = h \otimes G \quad (1)$$

Here I , h , and G denote the image, the PSF, and the object, respectively, with the object consisting of, e.g., a distribution of fluorophores in space. The convolution actually means that the object is “smeared out” by the focal spot. While a careful analysis of the imaging properties requires a detailed analysis of the PSF, for PSFs with a single peak, assessing the full width half maximum (FWHM) usually gives a good estimate of the microscope’s resolution. If identical molecules are within the FWHM distance, the molecules cannot be separated in the image. Therefore, it becomes evident that improving the resolution is largely equivalent to narrowing the PSF of the microscope.

In a conventional fluorescence microscope, the FWHM of the PSF is about $\lambda/(2n\sin\alpha) = \lambda/(2 \text{ NA})$, with λ denoting the wavelength, n the refractive index, α the semiaperture angle, and NA the numerical aperture of the lens. Without changes to the principal method, the spot size can only be decreased by using shorter wavelengths or larger aperture angles (Abbe, 1873; Born and Wolf, 1993). However, the lens half-aperture is technically limited to $\sim 70^\circ$ and the wavelength λ cannot be reduced below 350 nm because shorter wavelengths are not compatible with live cell imaging. In the best case, established far-field microscopes resolve 180 nm in the focal plane (x, y) and merely 500–800 nm along the optic axis (z) (See Figure 2–1a) (Pawley, 1995). All attempts to improve the resolution by a mere improvement of the optical components remain limited by diffraction. This is arguably better understood in the frequency domain. Here, the resolving power is described by the optical transfer function (OTF) giving the strength with which these spatial frequencies are transferred from the object to the image. The OTF is readily computed as the Fourier transform of the PSF (Wilson and Sheppard, 1984; Goodman, 1968). Due to the convolution theorem, Eq. (1) becomes

$$\hat{I} = \hat{h}\hat{G} = o\hat{G} \quad (2)$$

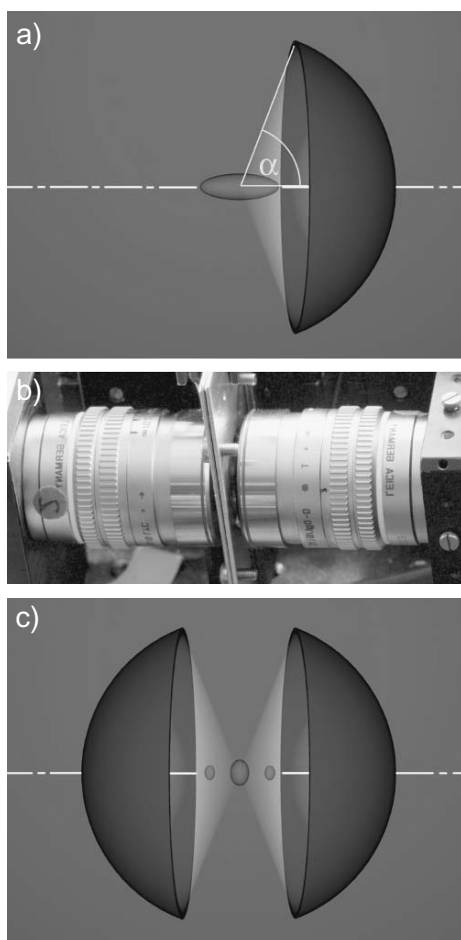


Figure 12-1. (a) The wavefront created by a single lens is a spherical cap. For the highest available semiaperture angle ($\sim 70^\circ$), diffraction theory dictates an elongated spot along the optic axis (z) that is approximately three times longer than it is wide (xy). The elongation can also be explained by the fact that the wavefront is only a cap rather than a sphere, bringing about an asymmetry of the focusing process. (b) Coherent addition of two spherical wavefronts created by opposing lenses completes a major part of the missing wavefront toward a more spherical one, which is the tenet of the “4Pi concept.” (c) The central focal spot, i.e., the main diffraction maximum, becomes more spherical as a result. However, because the wavefront is not approaching a sphere side-maxima appear along the optic axis.

where the hat signifies a 3D Fourier transform and o is the OTF. Due to the multiplication on the right hand side it is evident that even under optimal imaging conditions, object frequencies are lost where the OTF is zero. This loss is physically irrecoverable. Hence, the ultimate resolution limit is given by the highest frequency where the OTF is nonzero, i.e., the extent of the support of the OTF.

The highest spatial frequency produced by a (quasi)monochromatic wave passing through the setup is $k = 2\pi n/\lambda$. For focused light, the Fourier transform of its electric field, $C(k)$, is therefore given by a spherical cap with radius k , as seen in Figure 12-2. The excitation probability is well approximated by the absolute square of the electric field, corresponding to an autocorrelation in frequency space. This signifies that frequencies of up to $2k$ are contained in the excitation OTF. However, resolving optics can be employed both in the illumination and the detection path. This introduces a spatially varying detection probability for the emitted fluorescence. Since the phase of the illuminating light is lost during excitation, fluorescence is emitted incoherently. Therefore the effective PSF is given by the multiplication of the excitation PSF with its normalized detection counterpart, i.e., the detection PSF (Wilson and Sheppard, 1984), which is equivalent to a

convolution in frequency space. Thus the achievable cutoff is at $2k_{\text{ex}} + 2k_{\text{det}}$ where k_{ex} and k_{det} are the frequencies corresponding to the excitation and detection wavelengths in the medium, respectively.

These considerations are illustrated in Figure 12–2. As an example, if fluorescence is excited at 488nm and detected at around 530nm, the cutoff is given by $(2n/488\text{nm} + 2n/530\text{nm})^{-1} \equiv 127\text{nm}/n$,

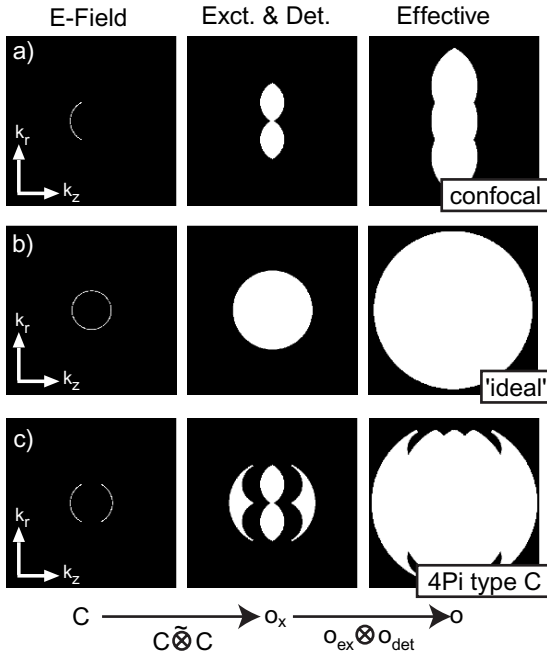


Figure 12–2. OTF supports in frequency space. (a) For a single lens, the Fourier transform of the electric field, $C(k)$, is a spherical cap with a cutoff angle determined by the aperture angle of the objective lens, which has been technically limited to $<70^\circ$. Excitation and detection (intensity) PSFs depend on the absolute square of the electric field; hence, the respective OTFs are given by the autocorrelation of the cap. Finally, in a confocal fluorescence microscope, the effective PSF is given by the product of the excitation and detection (intensity) PSF. Its OTF is therefore the convolution of the corresponding excitation and detection OTF, resulting in a squeezed OTF along the z direction accounting for the degraded axial resolution. (b) With a scalar field a truly spherical wavefront could ideally be generated. The autocorrelation of such a complete spherical shell is nonzero within a sphere of radius $2k$. Convolution of the excitation and detection OTFs again results in a spherical OTF support of radius $2k_{\text{ex}} + 2k_{\text{det}}$, which is tantamount to isotropic resolution. This radius defines the theoretical resolution limit of all microscopes in which the resolution is solely created by the optical arrangement and in which the phase of light is lost at the sample, as in the excitation–emission process. (c) The 4Pi microscope of type C with coherent excitation and detection through both lenses with large aperture angles generates the largest wavefront currently achievable in practice. For the highest available cutoff angles of approximately 70° this results in an almost spherical effective OTF. The resulting resolution is very close to the possible limit. While a semiaperture angle somewhat greater than 70° would still give a better coverage, the study also shows that the true solid angle of 4π is actually not required to gain the desired improvement in axial resolution. The putative use of aperture angles approaching 80° and higher is also complicated by the transverse nature of electromagnetic waves.

corresponding to approximately 83 nm for oil and 96 nm for water immersion. Therefore, all far-field light microscopes relying solely on improvements of the instrument have an ultimate resolution limit of just below 100 nm in all directions, even when all conceivable spatial frequencies are transmitted and restored in the image. This theoretical benchmark is indeed almost achieved by the 4Pi microscope described in more detail later (Egner et al., 2002; Hell et al., 1997).

To overcome this fundamental limit, higher frequencies need to be generated. Obviously this can be achieved by introducing nonlinearities in the interaction of the excitation light with a dye: For example, when using multiphoton (m -photon) excitation, the probability of producing a fluorescence photon depends on the m th order of the illumination intensity (Sheppard and Kampfner, 1978; Göpper-Mayer, 1931; Bloembergen, 1965). The original excitation OTF will be convolved m times with itself, thus extending the support region to m times higher frequencies. Therefore, it has sometimes been argued that superresolution can be attained by multiphoton excitation. It is, however readily understood why the resolution limit cannot really be pushed (Denk et al., 1990). While it is true that the excitation volume is narrower than the focal spot, which is due to the nonlinear dependence, using m -photon excitation to excite the same dye also means that the excitation energy has to be split between the photons (Denk et al., 1990). Consequently the photons have an m times lower amount of energy, thus an m times longer wavelength $m\lambda_{\text{ex}}$, which results in m times larger focal spots to begin with. Consequently, the theoretical cutoff remains equal at $2m(k_{\text{ex}}/m) + 2k_{\text{det}} = 2k_{\text{ex}} + 2k_{\text{det}}$. The multiplication by m stems from the m correlations or convolutions in frequency space and the division from the longer excitation wavelength. Obviously, this equally holds for excitation with several photons of different energies such as two-color two-photon excitation (Lakowicz et al., 1996). However, in the case of multiphoton excitation, higher frequencies within the support are usually damped, making the resolution poorer as compared to the 1-photon case. Multiphoton excitation is therefore not a viable option for significantly increasing the resolution. In addition, multiphoton excitation, especially for $m > 3$, requires very high intensities (Xu et al., 1996) leading to damage to the sample. Bleaching is largely restricted to the focal plane, but there it can be much stronger than for one-photon excitation. For some dyes, however, the excitation wavelength can be chosen to be shorter than $m\lambda_{\text{ex}}$ and when imaging deep into scattering samples such as brain slices, multiphoton imaging can feature a superior signal-to-noise ratio (SNR). In these cases, multiphoton excitation is a valuable method for reasons unrelated to the classical resolution issue.

Recognizing that the energy subdivision prevents any resolution increase, “multiphoton” concepts have been proposed, where the detection of a photon occurs only after the consecutive absorption of multiple excitation photons. Since the photons induce a linear optical transition in the dye, high intensities are not required and photon-energy subdivision does not occur (Hänninen et al., 1996; Schönle et al., 1999; Schönle and Hell, 1999). However, these concepts require specific conditions or complicated dye systems, also complicating their practical realization.

A radically different concept for improving the far-field fluorescence microscopy resolution appeared in the mid-1990s in the form of stimulated emission depletion (STED) and ground state depletion (GSD) microscopy (Hell and Wichmann, 1994; Hell and Kroug, 1995). Both concepts were based on one common principle: The required nonlinearities in the relationship between the irradiating light and the signal detected from the dye are provided by a spatially modulated reversible saturable optically linear (fluorescence) transition between two molecular states (RESOLFT). Since there is no physical limit to the degree of saturation, there is no longer a theoretical limit to the resolution. The limit is largely the size of the marker molecule itself. Therefore, concepts based on the RESOLFT principle, such as STED or GSD microscopy, truly break the diffraction barrier. Due to their principles of operation they can achieve nanoscale resolution in *all* directions (Hell, 1997; Hell et al., 2003; Heintzmann et al., 2002).

Fundamentally departing from earlier approaches to resolution increase, an introduction to the general idea will follow in the second part of this chapter, and we shall outline different possibilities for realizing the RESOLFT concept. We will see that initial applications of RESOLFT-based superresolution microscopy have recently yielded encouraging results, such as the first demonstration of a spatial resolution of $\lambda/25$ with focused light using regular lenses (Westphal et al., 2003). A potential road map toward imaging with nanometer resolution in live cells will be discussed, opening the prospect of bridging the gap between electron and current light microscopy, and providing a “nanoscope” working with focused light.

3 Axial Resolution Improvement by Aperture Enlargement: 4Pi Microscopy and Related Approaches

Figure 12–2 illustrates that the suboptimal axial resolution of a far-field light microscope can readily be motivated by the fact that the focusing angle of the objective lens is far from covering a full solid angle of 4π . The axially elongated PSF evidently also implies an axially compressed OTF. If the focused wavefronts were almost spherical, so would the focal spot, as well as the OTF of the system. In this case, the resolution would be largely isotropic. Therefore, an obvious way to achieve optimal optical resolution is to synthesize a focusing angle that comes closer to the solid angle of 4π . Since the focusing angle of a lens is limited by technical constraints, a larger wavefront can be obtained only by pasting together two or more wavefronts. A wavefront synthesis sharpening the focus along the optic axis is attained by employing two opposing lenses (Figure 12–1) that either coherently illuminate the sample from both sides or that detect the light through both lenses in a coherent manner (Hell, 1990; Hell and Stelzer, 1992a; Gustafsson et al., 1995). The resultant support of the OTF is displayed in Figure 12–2c. In the spatial domain the effect can be pictured as two counter-propagating focused beams interfering either at the focal spot or at a common point at the detector, for the illumination and the detection,

respectively. For the common high-angle lenses with a semiaperture $64^\circ < \alpha < 72^\circ$, constructive interference produces a maximum with an approximately three to four times narrower axial FWHM. The basic idea of the 4Pi microscope is to produce and take advantage of this axially narrowed PSF. Due to the fact that the semiaperture angle α is considerably smaller than 90° , the main maximum is also accompanied by one or two axial satellite lobes originating from the “missing lateral part” of a potentially full solid angle (Figure 12–1c).

Importantly, the narrowed main focal maximum is not $\lambda/(4n)$, as would be anticipated for a flat standing wave, but somewhat larger. By the same token, the side maxima are not located at $m\lambda/(2n)$, ($m = 0, 1, 2, 3 \dots$), but slightly farther away from the focal point, constantly decreasing in height with increasing order. As we shall see later, this difference is essential to the axial resolution improvement brought about by the coherent use of two opposing lenses and the actual reason why it is not possible to improve the axial resolution just by the interference of counterpropagating plane waves (Lanni, 1986; Bavley et al., 1993). The latter is the tenet of the standing wave microscope (SWM), which utilizes a *flat* standing wave of laser light and thus a set of excitation nodal planes in the sample along the optic axis. The SWM uses wide-field detection on a camera, like any conventional fluorescence microscope; the flat standing wave of excitation light is produced either with a mirror beneath the sample or by adding counterpropagating waves from two lenses. A substantial increase in axial resolution has initially been claimed by SWM. However, the resulting multiple thin interference layers in a sample (Lanni, 1986) do not unambiguously resolve features extending over a minimal axial range that is larger than half the wavelength ($0.2\text{--}0.4\mu\text{m}$) (Krishnamurthi et al., 1996; Freimann et al., 1997). While SWM might prove useful for specialized applications, it fails in delivering axial images of arbitrary objects (Nagorni and Hell, 2001a, 2001b). An explanation for this is that the interference excitation maxima have almost equal intensity, leading to strongly blurred “ghost images.” The deeper-rooted physical reason is that SWM does not increase the aperture of the system. This can be readily observed when analyzing the SWM in the frequency space (Figure 12–3). The excitation OTF has only three delta peaks located at $-k$, 0 , and k on the inverse optical axis. When convolved with the detection OTF, large gaps remain in the support of the effective OTF representing potential object frequencies that are not transferred from the object to the image.

It is a major physical insight (into the problem of axial resolution improvement with coherently used opposing lenses) that these gaps can be closed only when the light is *focused*, that is with spherical wavefronts. Focusing to the same point requires the accurate alignment of the two lenses, but the real physical challenge is to identify feasible mechanisms helping to further reduce the fluorescence contributed by the side-maxima still present when focusing at $64^\circ < \alpha < 72^\circ$. Reduction of the contribution from the side-maxima is equivalent to completing the support of the OTF and eventually to avoiding imaging artifacts. When spot-scanning confocal and nonconfocal 4Pi microscopy (Hell, 1990;

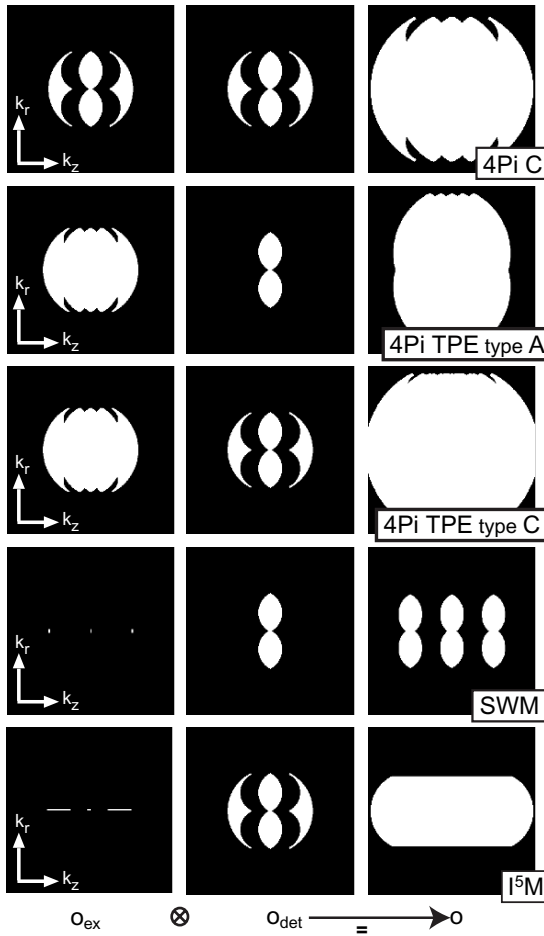


Figure 12–3. Comparison of the supports of excitation (left column), detection (middle column), and effective OTFs for techniques using two opposing lenses coherently. The effective OTF is the convolution of the excitation and detection OTFs. For two-photon excitation in the 4Pi mode, the Fourier transform of the excitation intensity is given by the excitation OTF of the 4Pi type C microscope but scaled by $488\text{ nm}/800\text{ nm} = 0.61$. The excitation OTF is therefore very similar to the effective OTF of 4Pi type C. Note the gaps in the support of the SWM microscope’s effective OTF.

Hell and Stelzer, 1992b) (Figure 12–3) and wide-field I^5M microscopy (Gustafsson et al., 1995) finally demonstrated the improvement of axial resolution (Hell et al., 1997; Schrader and Hell, 1996; Gustafsson, 1999; Gustafsson et al., 1999), each method used lenses with high numerical aperture and relied at least on one of the following three lobe-reducing mechanisms: (1) Using both focused excitation *and* confocal detection (Hell, 1990; Hell and Stelzer, 1992a) suppresses fluorescence from higher order side lobes of the excitation PSF; (2) two-photon excitation (TPE) (Hell and Stelzer, 1992b) emphasizes differences in the intensities of the main and the side-maxima due to its quadratic dependence on the excitation intensity; and (3) finally, spatial disparities between the maxima of excitation and fluorescence wavelengths can be used (Hell

and Stelzer, 1992b; Gustafsson et al., 1995) if light is detected coherently through both lenses. That is, the intensity maxima for excitation and detection are located at different points in space.

Three major types of 4Pi microscopy have been reported (Hell and Stelzer, 1992a). They differ on whether the spherical wavefronts are coherently added for illumination, for detection, or for both simultaneously; they are referred to as type A, B, and C, respectively. Usually the detection has been confocalized, but in conjunction with TPE successful axial separation with nonconfocal detection has also been reported. Here we will concentrate on the TPE 4Pi (type A), the 4Pi type C, and the TPE 4Pi type C confocal microscopes. Of these three, the TPE 4Pi confocal microscope has been applied to the largest number of imaging problems. It uses the very effective lobe-reducing measure of TPE combined with “point-like” detection. In reality the size of the “point-like” detector amounts to about the size of the main maximum of the diffraction-limited fluorescence spot (Airydisk), when imaged into the focal plane of the objective lens.

Clearly, nonconfocal wide-field detection and regular illumination would make 4Pi microscopy more versatile. Therefore, the related approach of I⁵M (Gustafsson et al., 1995, 1996, 1999; Gustafsson, 1999) confines itself to using the simultaneous interference of both the excitation and the (Stokes-shifted) fluorescence wavefront pairs; the latter are spherical as in a 4Pi microscope. The potential benefits of I⁵M are readily stated: single-photon excitation with arguably less photobleaching, an additional 20–50% gain in fluorescence signal, and lower cost. This method has so far yielded 3D images of actin filaments with an axial resolution slightly better than 100 nm in fixed cells (Gustafsson et al., 1999). To remove the side-lobe artifacts, I⁵M-recorded data are deconvolved offline. While the consideration of the OTF support in Figure 12–3 suggests that this single mechanism is indeed sufficient, it turns out that the relaxation of the side-lobe suppression comes at the expense of an increased vulnerability to sample-induced aberrations, especially with nonsparse objects (Nagorni and Hell, 2001a, 2001b). Thus I⁵M imaging, which has so far relied on oil immersion lenses, has required mounting the cell in a medium with $n = 1.5$ (Gustafsson et al., 1999). Live cells inevitably necessitate aqueous media ($n = 1.34$). Moreover, water immersion lenses have an inferior focusing angle and therefore larger lobes to begin with (Bahlmann et al., 2001). Potential strategies for improving the tolerance of I⁵M are the implementation of a nonlinear excitation mode and its combination with pseudoconfocal or patterned illumination (Gustafsson, 2000). While these measures again add physical complexity, they may have the potential to render I⁵M more suitable for live cells.

However, at this stage, the implementation of at least two of the mechanisms above proved more reliable: After initial demonstration of TPE 4Pi confocal microscopy (Schrader and Hell, 1996), superresolved axial separation was applied to fixed cells (Hell et al., 1997). The image quality could be improved further by applying nonlinear restoration (Holmes, 1988; Carrington et al., 1995; Holmes et al., 1995). Under biological imaging conditions, this typically improves the resolution up to a factor of two in both the transverse and the axial direction.

Therefore, in combination with image restoration, TPE 4Pi confocal microscopy has resulted in a resolution of $\sim 100\text{ nm}$ in all directions, as first witnessed by the imaging of filamentous actin¹⁹ and immunofluorescently labeled microtubules (Nagorni and Hell, 1998; Hell and Nagorni, 1998) in mouse fibroblasts.

While a very useful and explanative comparison of the OTF *supports* has been published (Gustafsson, 1999) it is obvious from the above that consideration of the supports alone is not sufficient to understand the respective benefits and limitations of SWM, I⁵M, and 4Pi microscopy. Rather than comparing their technical implementation (Gustafsson et al., 1999), a quantitative analysis of their PSFs and OTFs is needed to clarify under which conditions these microscopes will be able to deliver 3D-resolved images with superior resolution. The success of increasing the axial resolution with coherently used opposing lenses not only depends on the achievable bandwidth, but also on the strength with which the respective systems transfer the spatial frequencies *within* this bandwidth. Gaps and weak parts that occur in some systems (Gustafsson et al., 1995; Krishnamurthi et al., 1996) must be quantified for a particular optical setting because they may render the removal of artifacts impossible, thus precluding an increase in axial resolution. Below we demonstrate that these gaps are intimately connected with the optical arrangement and therefore are inherent to some of the methods described.

3.1 The Optical Transfer Function of 4Pi-Microscopy and Related Systems

Before expanding on the analysis of PSFs and OTFs, we quickly review their theoretical derivation. The excitation PSF of the confocal microscope, as well as the *detection* PSF of the wide-field, confocal, SWM, and 4Pi-type microscopes, are regular intensity PSFs. Assuming that excitation and emission involve a dipole transition of the dye, the excitation and emission PSF of a single lens is well approximated by the absolute square of the electric field in the focal region:

$$h = |\mathbf{E}_1(z, r, \phi)|^2 \quad (3)$$

It depends on the axial and radial distance to the geometric focus (z and r , respectively) and the polar angle ϕ . The field is usually calculated using the vectorial theory of Richards and Wolf (1959). In our case, we assumed circular polarization of the light. In frequency space the Fourier transform of the electrical field is then simply given by a spherical cap of radius k_{ex} (Figure 12–2). This is equivalent to approximating the spherical wavefronts emerging at the exit pupil as plane waves close to the focal spot. The absolute square in Eq. (3) corresponds to an autocorrelation in frequency space. When calculating the OTF, this convolution can be carried out directly (Schönle and Hell, 2002; Sheppard et al., 1993) or the OTF can be determined by Fourier transformation of Eq. (3).

For the excitation PSF of the 4Pi confocal microscopes (type A and C) and for the detection PSF of the I⁵M and the 4Pi confocal (type B and C), a calculation of constructive interference between the two spherical wavefronts is required. The PSF is therefore given by the coherent addition of two beams

$$h = |\mathbf{E}_1(z, r, \phi) + \mathbf{E}_2(z, r, \phi)|^2 \quad (4)$$

The field of the opposing lens is given by

$$\mathbf{E}_2(\mathbf{r}) = \mathbf{M}\mathbf{E}_1(\mathbf{M}^{-1}\mathbf{r}) \quad (5)$$

$\mathbf{M}$ is the coordinate transform from the system of lens number 2 to lens number 1 and is a diagonal matrix inverting the z -components for a triangular cavity and the y - and z -components for a rectangular cavity (Bahmann and Hell, 2000). In frequency space, the Fourier transform of the electric field is now given by two caps corresponding to the two focused wavefronts. Consequently, the OTF consists of an autocorrelation part equivalent to that for single lens excitation or detection and a cross-correlation of two opposite spherical caps represented by the outer “brackets” in Figures 12–2, 12–3, and 12–4. For TPE the excitation PSF is simply given by squaring the one-photon PSF scaled to the

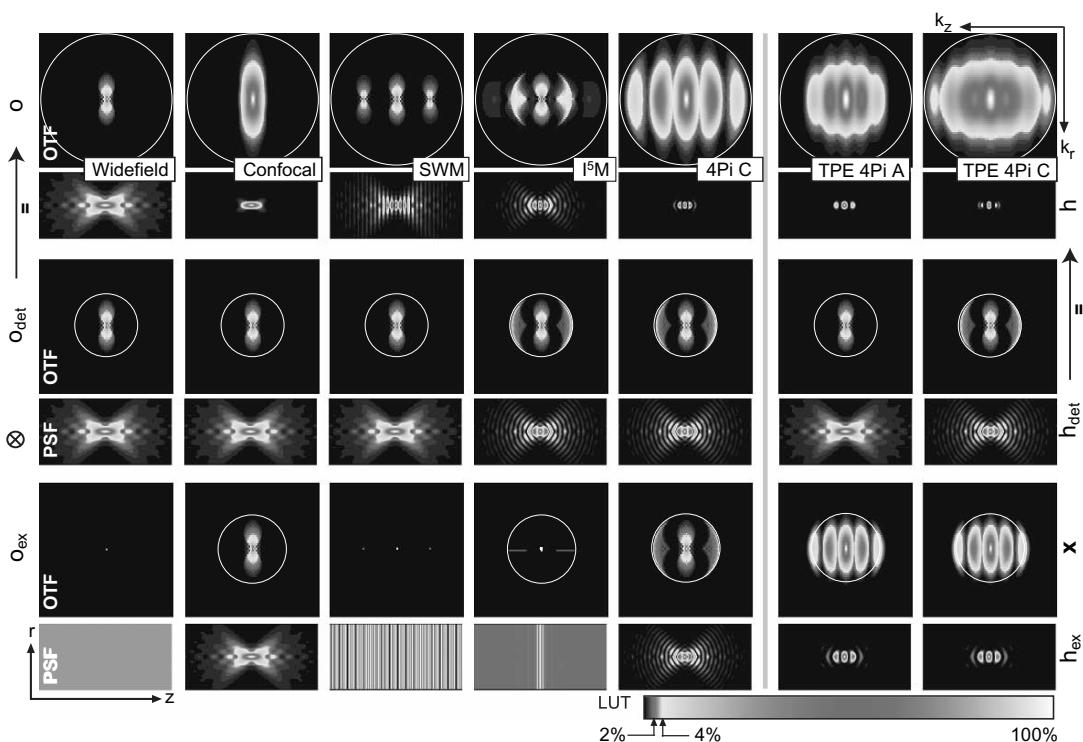


Figure 12–4. Overview of the excitation (bottom), detection (center), and effective (top) OTFs’ modulus and the corresponding PSFs of the wide-field, confocal, standing wave (SWM), I³M, 4Pi confocal type C, TPE 4Pi confocal type A, and TPE 4Pi confocal type C microscopes. The color look-up table (LUT) has been designed to emphasize the important weak OTF regions. The OTFs are shown in the squares above the corresponding PSFs; the zero frequency point is in the center and the largest frequency displayed is $2\pi/80\text{ nm}^{-1}$. The circles represent the maximum possible carrier as explained in Figure 12–3. For TPE, the excitation OTFs slightly extend over these circles because the excitation wavelength of 800 nm is less than double the one-photon excitation wavelength of 488 nm. While all these methods extend the OTF along the axial direction, they fundamentally differ in contiguity and absolute strength within the support region. For example, there are pronounced frequency gaps for the SWM and depressions for the I³M. The rectangular images of the PSFs represent a region of $5 \times 2.5\text{ }\mu\text{m}$ with the geometric focus in the center. (See color plate.)

appropriate TPE wavelength and similarly the OTF is the autoconvolution of the scaled one-photon OTF (Denk et al., 1990).

The excitation intensity of the SWM is given by a plane standing wave along the optic axis,

$$h = I_0 \cos^2(k_{\text{ex}}z) \quad (6)$$

where I_0 denotes a constant, k_{ex} is the wavenumber; and we assumed constructive interference to occur at the common geometric focus. The excitation OTF is its Fourier transform and given by

$$o = I_0 [\delta(k) + \delta(k - 2k_{\text{ex}})/2 + \delta(k + 2k_{\text{ex}})/2] / 2 \quad (7)$$

Again, this is the result of auto-correlating the electric field's Fourier transform that consists of delta peaks at $\pm 2k_{\text{ex}}$.

While the 4Pi microscope uses a spatially coherent point-like laser illumination, in the I^5M microscope wide-field illumination is used, normally in the Köhler mode, either with a lamp or a laser. The physical consequences of this difference are best explained as follows. If the two spherical wavefronts of the 4Pi illumination are decomposed into plane waves incident from different angles and corresponding to different points of the illumination apertures, *all* plane waves of the aperture interfere with each other in the focal region. In the I^5M the illumination light is not coherent throughout the aperture. Therefore only *pairs* of plane waves originating from corresponding points (mirror images about the focal plane) of the illumination apertures are mutually coherent, forming a standing wave in the focal region. The period of these standing waves scales with the cosine of the azimuth angle θ . The excitation PSF of the I^5M can then be calculated by adding the intensity of these plane standing waves. Assuming uniform intensity throughout the exit pupil of the lens, the PSF is given by

$$\begin{aligned} h(z) &= I_0 \int d\phi \int_0^\alpha d\theta \sin\theta \cos^2(k_{\text{ex}}z \cos\theta) \\ &= 2\pi I_0 \int_0^\alpha d\theta \sin\theta \cos^2(k_{\text{ex}}z \cos\theta) \end{aligned} \quad (8)$$

The support of the excitation OTF is readily inferred: For each θ the integrand in Eq. (8) is the excitation PSF of the SWM and thus the total OTF consists of a superposition of expressions of Eq. (6) for wave vectors ranging from $k_{\text{ex}} \cos(\alpha)$ to k_{ex} . Loosely speaking, this incoherent superposition smears out the delta peaks at the sides, forming the lines in Figures 12–3 and 12–4. This excitation mode contributes to avoiding the gaps that remain in the SWM's support after convolution with the detection OTF. If, on the other hand, the incoherent light source is imaged into the focal plane of the lens, i.e., critical illumination, mutually coherent points form wavefronts focused onto and interfering at the conjugate point in the image of the light source. The individual 4Pi PSFs produced by each point of the light source as a result are incoherently summed up, giving an integral of the 4Pi excitation PSF over the field of view in the focal plane. The OTF becomes nonzero exclusively on the inverse optical axis where it is given by the values of the 4Pi OTF, altogether leading to a result not very much different from that predicted by Eq. (8). However, critical illumination is problematic due

to potential nonuniformities in the light source and will be omitted in our analysis.

In any case, once the fluorescence light is generated in the sample, the I⁵M collects the spherical fluorescent wavefronts just as in a 4Pi microscope of type B or C. The two counterpropagating spherical wavefronts of fluorescence collected by each lens interfere constructively in a common point on the camera. Disturbance of the interference pattern of neighboring points of fluorescence emission is damped by the spatial incoherence in the focal plane: the radius of spatial coherence is largely given by the Airy disk associated with the fluorescence light at the aperture in use. Thus the I⁵M implements the highest possible degree of parallelization of 4Pi detection.

Figure 12–4 shows the numerically calculated excitation, detection, and effective PSFs/OTFs of the 4Pi, I⁵M, and SWM setups, along with those of the conventional epifluorescence and confocal microscope. The epifluorescence microscope features uniform illumination intensity throughout the sample volume; its OTF is a single deltapeak at the origin. To obtain a practically relevant comparison, we assumed a numerical aperture of $NA = 1.35$ and oil immersion with a refractive index $n = 1.51$. In the case of single-photon excitation of the dye, an excitation and detection (i.e., central fluorescence) wavelength of 488 nm and 530 nm, respectively, is assumed. For TPE, an excitation wavelength of 800 nm was chosen. Finite-sized pinholes can be taken into account by convolving the detection PSF with the image of the pinholes in the focal plane: a disk of radius r_{PH} . In frequency space this corresponds to a multiplication with the disk's Fourier transform given by

$$\hat{h}_{\text{PH}} = J_1(kr_{\text{PH}})/k \quad (9)$$

The first root of the Bessel function is at ~ 3.83 and thus the detection OTF becomes zero at $k = 3.83/r_{\text{PH}}$. However, the largest frequency present in the detection OTF is given by half the wavelength and therefore its support is unaltered if the pinhole radius is smaller than $3.83\lambda/(4\pi NA) \cong 0.3\lambda/NA$, which corresponds to half the Airy disk radius. Pinholes smaller than this can virtually be neglected in the computation, while for sizes around this value and larger, the PSF will widen laterally, suppressing the OTF at higher lateral frequencies. The effect of the pinhole size on axial resolution remains small as long as it does not exceed that of the Airy disk. We will therefore neglect the pinhole in our further analysis. All PSFs were numerically computed in a volume of $128 \times 128 \times 512$ pixels in x -, y -, and z -directions, respectively, for cubic pixels with 20 nm length. The OTFs were calculated by Fourier transformation; of the 512 pixels in the z -direction only data based on the central 256 pixels are shown in Figure 12–4. The color look-up-table (LUT) has been chosen so that the regions of weak signal are emphasized for both PSF and OTF. This reveals important differences between the systems. Areas of low but nonnegligible intensity are important since they cover a large volume and substantially contribute to the image formation.

Let us first consider the PSFs (Figure 12–4, narrow columns). Immediately, some differences between the various approaches become

apparent. While the excitation modes of the SWM and I^5M are similar, the local minima are not zero for the latter due to the incoherent addition of the standing wave spectrum. The most important difference is observed when comparing the 4Pi microscopes. As a result of focusing, their PSFs are confined in the lateral direction so that contributions from the outer lateral parts of the focal region are reduced. The confinement has important consequences. Whereas the 4Pi confocal microscope, especially its two-photon version, exhibits only two pronounced but rather low lobes, the I^5M and even more so the SWM feature a multitude of lobes and fringes on either side of the focal plane, despite the fact that all of them rely on the same aperture. The second consequence is that due to its quadratic or cubic dependence on the excitation distributions, the 4Pi confocal PSFs can be separated into an axial and a radial function in good approximation (Hell et al., 1995; Schrader et al., 1998):

$$h(r, z) \cong c(r)h_i(z) \quad (10)$$

Separability is a particular feature of the 4Pi confocal and multiphoton arrangements and we shall see later that it is the prerequisite for simple online removal of sidelobe effects in the image. TPE leads to a further suppression of the outer parts of the excitation focus and thus of the side lobes of the 4Pi illumination mode. Figures 12–3 and 12–4 also reveal that in conjunction with coherent detection (type C), TPE 4Pi confocal microscopy features an almost lobe-free PSF. Its OTF almost fills the maximum support region. In the SWM and the I^5M , the number and relative heights of the lobes increase dramatically when moving away from the focal point because an effective suppression mechanism is missing. In the SWM the lobes become even higher than the central peak itself. In the I^5M the secondary maxima are as high as the first maxima of the single-photon 4Pi confocal microscope of type C.

3.2 Removing Periodic Artifacts through Deconvolution

Even if they are small, side lobes and resulting ghost images in the raw data remain a common feature of all methods employing two lenses coherently. Next, we turn to the OTF to understand the circumstances under which image processing can effectively be used to remove the artifacts induced by the lobes. It is obvious from Eq. (2) that if the OTF were nonzero everywhere, we could divide the Fourier transform of the image by it. Subsequent Fourier back-transformation of the data would render the object. In practice, the OTF is limited in bandwidth and has weak regions. As division outside the OTF support is impossible, these frequencies are lost. But even in regions where the OTF is small, division strongly amplifies noise-producing artifacts.

Image restoration techniques, whether linear or nonlinear, aim at restoring as many frequencies as possible while trying to avoid this effect. Linear deconvolution is based on the division approach but introduces a special treatment for frequencies not transmitted by the OTF. It is capable of restoring frequencies only where the OTF is above the noise level. If frequencies are missing, a correct representation of

the object in the image can be given only if these frequencies are extracted from a priori knowledge of the object. This extraction is mathematically more complex and often not viable. Linear deconvolution, on the other hand, is computationally facile and fast. Speed is of particular importance because the interference artifacts are ideally removed online making the final image readily accessible. Therefore, one of the most prominent advantages of a uniformly strong OTF is the possibility of applying a linear deconvolution.

The comparison of the effective OTFs in Figure 12–4 highlights the severe gaps in the SWM, making deconvolution impossible. Linear deconvolution is reportedly possible in the I⁵M (Gustafsson et al., 1999). However, as the gaps in the I⁵M are filled with rather low amplitudes, this method will create image artifacts for objects that are not sparse, not very bright, or objects containing spatial frequencies coinciding with the gap. Owing to the contiguity of its support and the strong amplitudes of the OTF, 4Pi confocal microscopy fulfills the preconditions for linear deconvolution. In fact, linear deconvolution along the axial direction based on the separability of the PSF has been applied for the removal of interference artifacts arising in the recording of complex objects. Thus superresolved axial imaging has been shown in the dense filamentous actin⁵⁵ and in the microtubular network (Nagorni and Hell, 1998) of a mouse fibroblast cell. We will have a closer look at this important procedure in the subsequent section.

Approximating the 4Pi confocal PSF as in Eq. (10) allows us to perform a computationally inexpensive *one-dimensional* linear deconvolution that simply eliminates the effect of the lobes in the image. The axial factor of the PSF can basically be decomposed into the convolution of a function $h_p(z)$ describing the shape of a single peak, and a lobe function $l(z)$ containing the position and relative heights of the lobes:

$$h_l(z) \cong h_p(z) \otimes l(z) \quad (11)$$

h_p quantifies the blur and l describes the replication that is responsible for the “ghost images” in unprocessed 4Pi images. The effect of the lobe function can be eliminated using algebraic inversion. We use a discrete notation, where each lobe is represented by a component l_i of the vector $\mathbf{l}$ with l_0 denoting the strength of the central lobe and the index running from $-n$ to n . Negative indices denote lobes to the left of the central peak. If the lobe distance in pixels is denoted by d , the values of the object along the line are given by O_j and those of the image by I_j . Thus, the convolution is given by

$$I_j = \sum_k l_k O_{j-dk} \quad (12)$$

Looking for a filter l^{-1} inverting this convolution we need

$$O_j = \sum_s l_s^{-1} I_{j-ds} = \sum_s l_s^{-1} l_k O_{j-ds-dk} \quad (13)$$

for all possible objects. The inverse filter needs to fulfill the condition

$$\sum_s l_s^{-1} l_{j-s} = \delta_{j0} \quad (14)$$

At this point we can arbitrarily choose the length of the inverse filter, assuming an index running from $-m$ to m . The index j in Eq. (13) can take values from $-m - n$ to $m + n$. Therefore we have a system of $2(m + n) + 1$ equations with $2m + 1$ unknowns, which is usually not solvable. An approximation is found by considering the equations for $j = -m \dots m$ only. Now, the problem is equivalent to solving a linear Toeplitz problem with the Toeplitz matrix given by the vector $\mathbf{l}$ (Nagorni and Hell, 2001; Press et al., 1993). The approximation is good if the edges of the inverse filters are small, since the remaining equations are nearly satisfied. This holds if the first-order lobes are $<45\%$; in this case, the error is practically not observable. A typical length of the inverse filter is 11. In conjunction with a lobe height of 45% , the edges of the inverse filter feature a modulus below 1% of the filter maximum. For lobes of 35% relative height, this value drops to a value of only about 0.08% . Thus, using this technique, the separability of the 4Pi confocal PSF provides a quick way to obtain a final image that is equivalent to imaging with an “ideal” optical microscope that has a single narrow main maximum at the focal point. The inverse filter is discrete and nonzero only at a few points. This very effective side lobe removal method is therefore referred to as point deconvolution. Exploiting the characteristics of the PSF of the 4Pi microscope, it is applicable only in conjunction with this method.

Axial lobes in the PSF entail suppressed OTF regions along the optic axis. Thus, point deconvolution restores suppressed frequencies by linear deconvolution even though the actual calculation takes place in real space. For PSFs that cannot be written as in Eq. (10) point deconvolution is insufficient, because the transverse directions (x, y) have to be involved as well. In this case, it is mathematically clearly preferable to pass the data through the frequency domain. Established linear deconvolution algorithms rely on inverting Eq. (2) but also accommodate for the vanishing regions of the OTF. An estimate of the frequency spectrum of the object can be obtained by using

$$\hat{E}(k) = \hat{I}(k) o^*(k) / (|o(k)|^2 + \mu) \quad (15)$$

The regularization parameter μ (Bertero et al., 1990) sets a lower threshold on the denominator to avoid amplification of frequencies where the modulus of the OTF is so small that the frequency spectrum of the image is dominated by noise. If the OTF is a convex function that is in the absence of “gaps,” the effect of regularization is similar to smoothing. In most cases considered in this chapter, however, the OTF is not convex and the situation is more complicated. Small OTF values are found not only at the OTF boundaries, but also within its region of support, e.g., in the vicinity of the minima (Figure 12–4) or at the frequency gaps (Figure 12–3). If the lobes are too high (typically $>50\%$), the level of the minima of the OTF is comparable or smaller than the noise level. This is definitely the case in the SWM, but also in the presence of slight aberrations in the I⁵M, as well as in an aberrated 4Pi microscope. Therefore, the necessity of implementing a certain value of μ is closely connected with the lobe height and with the potential artifacts induced by the lobes.

If the PSF is separable in a peak function and an axial lobe function, the lobe removal can be elegantly targeted:

$$h(\mathbf{r}) = h_p(\mathbf{r}) \otimes l(z) \quad (16)$$

Contrary to the decomposition in Eq. (11), the peak function $h_p(\mathbf{r})$ also contains the dependence of the peak on the lateral coordinates. We no longer require the PSF to separate in a radial and an axial part, yet the lobe function still gives the relative height of the lobes as well as their location:

$$l(z) = \sum_s l_s \delta(z - ds) \quad (17)$$

with d now being the lobe distance in units of length. The frequency spectrum of the image is then given by

$$\hat{I}(\mathbf{k}) = \hat{G}(\mathbf{k}) \cdot o(\mathbf{k}) = \hat{G}(\mathbf{k}) \cdot \hat{h}_p(\mathbf{k}) \cdot \hat{l}(k_z) \quad (18)$$

If the lobe function is symmetric with respect to the focal plane (i.e., for constructive interference), we can write

$$\hat{l}(k_z) = 1 + \sum_{s>0} 2l_s \cos(sdk_z) \quad (19)$$

This decomposition immediately discloses why the critical lobe height is 50%: If the PSF consists of a main maximum and two primary lobes of 50%, the right-hand side of (19) vanishes for the axial frequencies associated with the distance d . So, if the lobes are >50% the frequency represented by the lobes is not contained in the OTF, and hence not transferred to the image. In reality, the critical lobe height is slightly above 50%. The reason is the influence of the secondary lobes that have been neglected in our reasoning. Nonetheless, the 50% threshold is an excellent rule for the critical lobe height, which applies equally to SWM, I⁵M, and 4Pi microscopy, for fundamental reasons.

Equation (18) implies that the effect of the lobes can be removed by direct Fourier inversion. This is the case if the Fourier transform of the lobe function remains above the noise level throughout the relevant frequency spectrum. The spectrum of the lobe-free image is then obtained by dividing the image spectrum by the Fourier transform of the lobe function. Figure 12–5 shows a typical data set acquired with a TPE type A 4Pi confocal microscope and illustrates lobe removal both in the spatial and the frequency domain. The avalanche photodiode used as a detector had a typical dark count rate <1 count/pixel, which is negligible (Hell et al., 1997). Hence, the only significant source of noise was the Poisson noise of the photon-counting process, manifesting itself as white noise that is independent of the spatial frequency. Since the primary lobes are well below 50%, the first minima of the OTF are at 19%, which is well above the noise level of typically 0.5–1%. Direct lobe removal is straightforward in this example, underscoring the importance of nonvanishing amplitudes in the OTF. If the OTF exhibited regions close to zero, as is predicted for the I⁵M, the multiplication with a high number in this region would result in a strong amplification of noise, and compromise the obtained image. In the

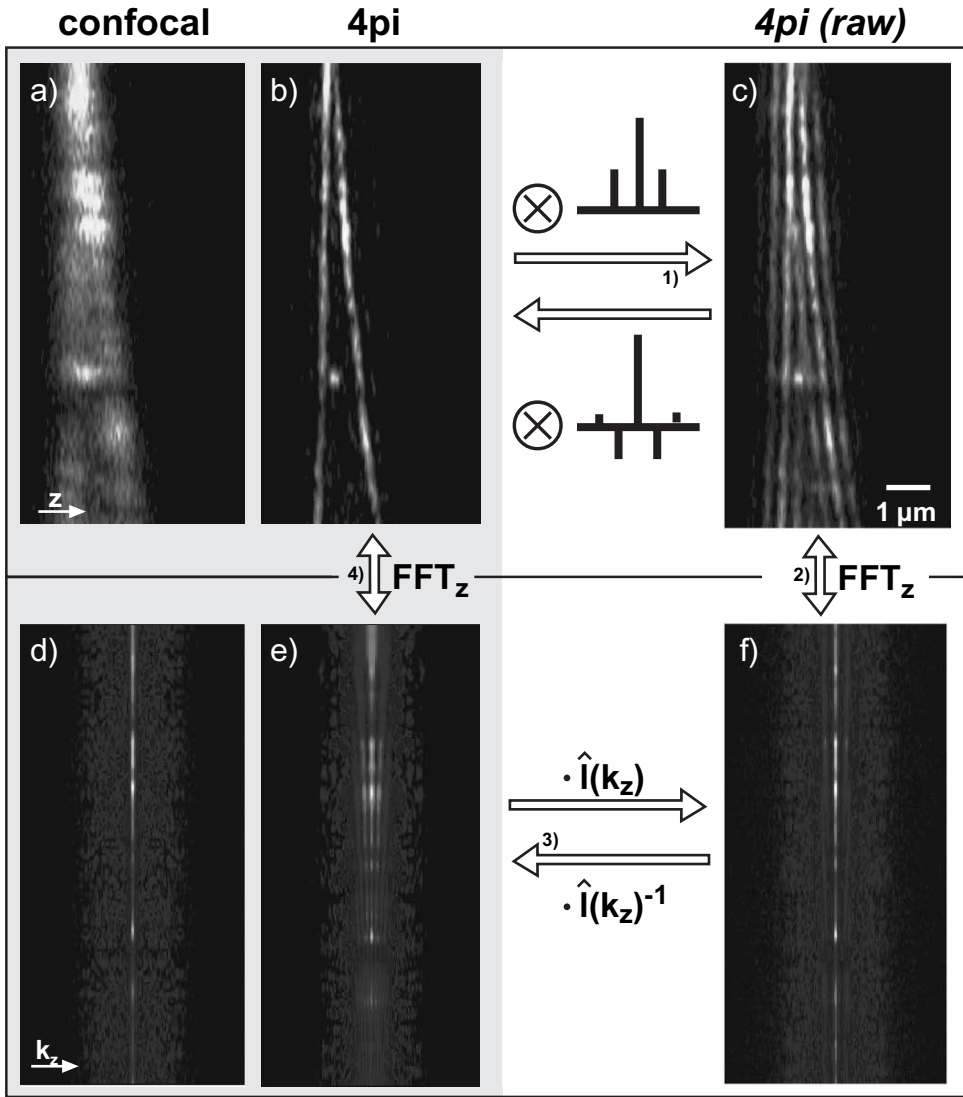


Figure 12-5. Lobe removal and deconvolution in 4Pi microscopy. The figure shows images of the same pair of actin fibers in a fixed mouse fibroblast cell recorded in the TPE confocal (a) and TPE 4Pi type A (b and c) mode. The corresponding Fourier transform along the optic axis is also shown (d, e, and f). The five-fold axial resolution increase (a vs. b) and the correspondingly extended OTF (d vs. e) are immediately visible. The side lobes are well below 50% and the factorization of the PSF's axial and lateral dependence is possible in 4Pi microscopy. Therefore, an inverse discrete filter can be found and its application to the raw data (c) yields a valid and almost artifact-free image (b). Alternatively, lobe removal can be performed in the frequency domain. Equation (16) indicates that the Fourier transforms of the raw data (f) is given by the product of the Fourier transform of the lobe-free image (e) and the lobe function $\hat{l}(k_z)$. Thus, (2) Fourier transforming the raw data, (3) multiplying with the inverse of the lobe function's Fourier transform, $\hat{l}^{-1}(k_z)$, and (4) Fourier backtransforming lead to almost the same lobe-free image. This method can be applied even if the separation of axial and lateral dependence is impossible for the PSF. The Fourier transforms along the axial direction (3 and 4) merely have to be replaced by their 3D counterparts. (See color plate.)

SWM, it is impossible to remove the interference artifacts on a general basis.

While an in-depth analysis of *nonlinear* iterative deconvolution (i.e., image restoration) algorithms is far beyond the scope of this chapter, it can be safely stated that these suffer from limitations similar to the linear deconvolution approaches. They cannot really restore information that has been lost in the imaging process. A connected and largely convex OTF that does not imply regions of weak transmission is almost equally important to these algorithms. Nonlinear iterative algorithms take advantage of a priori information about the object that linear deconvolution cannot uncover unambiguously. The simplest a priori information is the positivity of the object and of the image. Carefully applied to 4Pi and I⁵M data, nonlinear restoration can yield impressive results, but extreme care must be taken not to compromise the reliability of the outcome by false or even biased assumptions.

3.3 Discussion: Improved Axial Resolution in Practice

While the use of coherent beams from two opposing lenses extends the OTF in the axial direction, a detailed analysis shows that conditions have to be met to exploit this advantage in an effective manner. Reliable and artifact-free restoration is possible only if the region of the OTF is convex, that is, if it does not contain nonnegligible “weak” regions or gaps. It was shown that the OTF of an SWM exhibits marked gaps along the optic axis that make the removal of the interference artifacts almost impossible. It is therefore not surprising that previous experimental studies confirmed that in the SWM it is impossible to unambiguously distinguish two axially separated objects, unless the object is thinner than 50% of the wavelength.³⁴ Unambiguous resolution of axially extended objects is impossible with SWM for fundamental physical reasons. Nevertheless, both SWM and 4Pi microscopy have been successfully applied to the ultraprecise measurement of axial distances (Albrecht et al., 2002; Schneider et al., 2000; Schmidt et al., 2000) and object sizes (Egner et al., 2000; Failla et al., 2002), which is conceptually and practically less demanding.

The OTF of the I⁵M is superior to that of the SWM, because it remains nonzero throughout its support. Still, it is weak over a considerable region when compared to the giant zero-frequency peak, which actually is a singularity. Hence, to benefit from this contiguity, I⁵M data must be recorded with a very large signal-to-noise ratio. Besides, the method may be applicable to sparse nonextended objects only, such as points or sparse fine lines. The OTF of the reported 4Pi microscopes are truly contiguous. In the critical regions, the 4Pi confocal OTF (Figure 12–4) exhibits significant values in the 19–23% range. This feature is of key relevance to the removal of the interference artifacts and hence for unambiguous, object-independent 3D microscopy with improved axial resolution.

Another major difference between the effective PSF of the SWM and I⁵M on the one hand, and the 4Pi confocal on the other, is the fact that the PSF is spatially much more confined in the latter case. This allows

us to reduce the process of lobe removal from a 3D deconvolution to a one-dimensional linear problem with the highly useful side effect that single layers can be acquired and deconvolved while all other methods require acquisition of a full 3D data stack.

The 4Pi concept is the result of rigorously maximizing the aperture angle employed in the imaging process. This results in a superior excitation OTF due to its lateral filling, which is rooted in the fact that the (exciting and detected) light is *focused*. Hence, while scanning with a focused beam inevitably reduces imaging speed, this procedure also results in fundamentally improved imaging properties of the microscope. Flat field standing wave excitation inherent to the I⁵M and SWM trades off collected spatial frequencies. The loss of optical frequencies is so significant (Figures 12–3 and 12–4) that the ability to provide unambiguous axial resolution is either put at risk or not viable.

Nevertheless, axial resolution improvement and imaging similar to the 4Pi microscope are reportedly possible when combining I⁵M with offline image restoration (Gustafsson et al., 1999). The combination of I⁵M with fringe pattern illumination and subsequent image restoration has also been suggested to (1) alleviate the problem of the zero frequency singularity (Gustafsson, 2000; Heintzmann and Cremer, 1998) and (2) to increase the lateral resolution to that of (restored) confocal microscopy. Leaving aside the technical complexity of controlling the interference patterns of typically four pairs of beams, this suggestion confirms that an unambiguous axial resolution requires the employment of a wider angular spectrum. In fact, the spherical beams in a 4Pi microscope can be regarded as a complete spectrum of interfering plane waves coming from all angles available. Hence, from the standpoint of imaging theory, the combination of I⁵M with fringe pattern illumination is a modification of I⁵M toward a scheme that is more similar to the 4Pi arrangement; the improvements of the OTF are gained by conditions that converge to the “focusing” conditions found in the 4Pi microscope.

In real samples the OTF of the microscopes is compromised by aberrations that were not included in our comparison of the concepts. Residual misalignments of the foci induced by variations of the refractive index in the sample play a role. However, successful 4Pi imaging of the mouse fibroblast cytoskeleton (Nagorni and Hell, 1998) and I⁵M imaging of similar structures (Gustafsson, 2000) have revealed that aberration effects are surmountable. Image deconvolution requires prior knowledge of the PSF and hence its explicit determination with a point-like object. This is particularly important since the PSF depends on the relative phase of the two counterpropagating wavefronts. It has been shown that in 4Pi microscopy the type of interference in the focal plane (constructive, destructive, or anything in between) is of lesser importance (Hell and Nagorni, 1998). Depending on its influence on the OTF, the same will also apply to I⁵M. So far, the structure of the PSF has been determined by measuring the response of test objects, such as fibers or fluorescent beads. However, it has been shown that the relative phase can be extracted directly from the image data (Blanca et al., 2002). This is of great importance in cases in which the imaged object itself alters the relative phase of the interfering beams.

A filled OTF entails robustness in operation and lower amenability to potential misalignment. Thus 4Pi microscopy allowed superresolved imaging of specimens in aqueous media using water immersion lenses (Bahlmann et al., 2001). This is noteworthy since water lenses feature semiaperture angles of not more than 64° as compared to 68° available for oil or glycerol immersion, therefore increasing side lobes and rendering the problem of missing frequencies in the OTF more severe. Imaging a watery sample is one of the prerequisites of live-cell imaging.

For realistic aperture angles, the inherent lack of contiguity of the OTF in the SWM renders the removal of interference ambiguities impossible. The I⁵M becomes more viable through filling the frequency gaps present in the SWM, albeit with values that are weak with respect to the zero-frequency components. Both systems have yet to prove their applicability to live cell imaging. In fact, recent investigations have cast doubt on whether I⁵M is able to provide reliable data from live cells with the available water immersion lenses. If water immersion lenses of 70° and higher were available, the robustness and applicability of I⁵M would significantly increase. Thus, this method would become viable for a much larger variety of objects, at least for not too dense or not too 3D-convoluted ones. The fact that the enlargement of the semiaperture angle is the key to the performance of the I⁵M highlights the exploitation of the entire *spherical* wavefronts as the central physical element. In a sense, the I⁵M can be viewed as a 4Pi system that maximizes the degree of parallelization in the focal plane.

Parallelization is readily accomplished in 4Pi and 4Pi confocal microscopy, as well. A multifocal variant of TPE 4Pi confocal microscopy, termed multifocal multiphoton 4Pi microscopy (MMM-4Pi) (Egnor et al., 2002), has indeed translated the typical 140 nm axial resolution of a TPE 4Pi microscope into live cell imaging. Importantly, although present and noticeable, phase changes induced by the cell proved more benign than anticipated. Nevertheless, phase alterations and wavefront aberrations due to refractive index changes within the specimen are likely to confine 4Pi microscopy and related techniques to the imaging of individual cells or thin cell layers. In conjunction with nonlinear image restoration this novel imaging method has displayed ~ 100 -nm 3D resolution under live cell conditions. For example, MMM-4Pi microscopy has provided superior 3D images of the reticular network of GFP-labeled mitochondria in live budding yeast cells.

More recently, 4Pi microscopy has been realized in a compact optical unit that was interlaced with a state-of-the-art single beam scanning confocal fluorescence microscope (Leica TCS-SP2, Mannheim, Germany). Operating in type C mode, i.e., coherent spherical wavefronts both for excitation and for confocal detection, this compact and rugged system displayed a seven-fold improved axial resolution over confocal microscopy in live cells (Gugel et al., 2004). This system is now commercially available as the first far-field optical microscope with axial superresolution.

The fact that *single-photon* excitation 4Pi confocal microscopy has not been realized to date has been due to the superiority of the OTF in the TPE version. Nevertheless, it is clear from imaging theory that a single-photon 4Pi confocal microscope of type C features a more contiguous

and better filled OTF than an I³M, which inevitably operates in the single-photon mode. Given further improvements in aberration compensation and lens manufacturing, it is feasible that single-photon excitation 4Pi microscopy (of type C) can also be realized in a reliable manner. An inherent advantage of single-photon excitation over its two-photon counterpart is that the much shorter excitation wavelengths lead to axially narrower focal spots. For example, using a wavelength of ~400 nm, a single-photon 4Pi microscope of type C would deliver an axial resolution of ~73 nm. This number further underscores the potential of 4Pi microscopy to deliver 3D images with axial sections well in the tens of nanometer regime.

In summary, the resolution of a far-field fluorescence light microscope can be improved down to the range of 60–100 nm along the optic axis by coherently adding the focal light fields of two opposing lenses. The addition of the fields leads to improved resolution for both excitation at the focal point and detection of fluorescence at a common point. The axial resolution is improved only if the main maximum of the resulting PSF is at least twice as large as the primary side maxima arising from the coherent addition. The latter condition can be fulfilled only if, at least for one of the processes, the added light field is a *spherical* wavefront covering the aperture angle of the lens. This in turn shows that the key physical element to improving the axial resolution by the coherent use of two opposing lenses is not the production of an interference pattern, but the enlargement of the aperture of the system.

4 Breaking the Diffraction Barrier

Increasing the total aperture of the focusing system with two opposing lenses improves the 3D resolution of a far-field microscope, but does not break the diffraction barrier. On the contrary, 4Pi microscopy exploits the full potential of diffraction-limited imaging. This particularly applies to (multiphoton) type C 4Pi confocal microscopy, which can be regarded as the far-field optical microscope with the largest possible aperture. For a microscope, being subject to the diffraction limit implies that the system features the maximum resolution that cannot be surmounted. On the other hand, breaking the diffraction barrier implies the potential for featuring an infinitely sharp focal spot or an OTF with an infinitely large bandwidth.

The first concept to break the diffraction barrier was STED fluorescence microscopy.⁹ The physical concept of STED microscopy was described as early as 1994 and was soon followed by GSD microscopy, a related concept that also entailed diffraction-unlimited resolution (Hell and Kroug, 1995). Whereas STED microscopy utilized stimulated emission to deplete the excited state of the fluorophore, GSD aimed at depleting its ground state by transiently pumping the dye into a long-lived dark state, i.e., its triplet state. Importantly, both concepts share the same principle for breaking Abbe's barrier: a focal intensity distribution featuring a zero point (or at least a strong gradient in space) effects a *saturated depletion* of one of the molecular states that are essential to the fluorescence.^{10,11} Following depletion, the state is populated

again, that is, the saturated transition is reversible. A saturated transition, such as a depletion of a state, introduces vast nonlinearities that eventually prove essential for breaking the diffraction barrier.

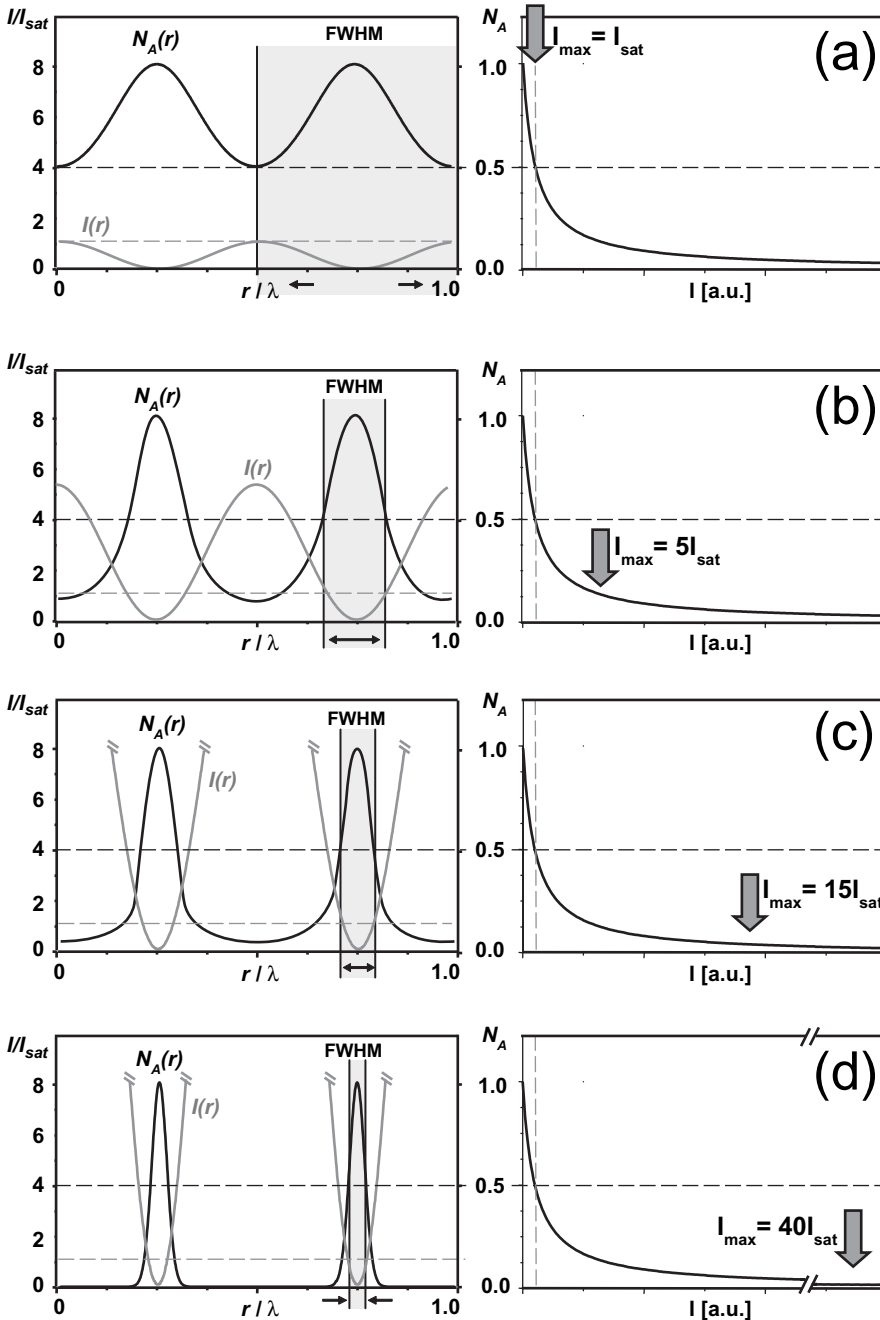
An even closer examination of the underlying concept shows that in fact any saturable transition between two states where the molecule can be returned to its initial state is a potential candidate for breaking the diffraction barrier (Hell, 1997; Hell et al., 2003; Dyba and Hell, 2002). The general concept has therefore been termed RESOLFT. The transition utilized can be selected to match the practical conditions of the imaging problem at hand, such as the required intensities, the available light sources, and the avoidance of photobleaching. An important aspect with respect to biological applications is the compatibility with live cells.

The basic idea of the RESOLFT concept can be understood by considering a molecule with two arbitrary states A and B between which the molecule can be transferred. In fluorophores, typical examples for these states are the ground and first excited electronic states and conformational and isomeric states. The transition $A \rightarrow B$ is induced by light, but no restriction is made about the transition $B \rightarrow A$. It may be spontaneous, but may also be induced by light, heat, or any other mechanism. The only further assumption is that at least one of the two states is critical to the generation of the signal. In fluorescent microscopy this means that the dye can fluoresce only (or much more intensively) in state A. Such a system may be exploited to generate diffraction-unlimited resolution in fluorescence imaging (or any kind of manipulation, probing, etc. that depends on one of the states), which is illustrated in Figure 12–6.

We begin with all molecules or entities in the sample being in state A. Our goal is to generate a diffraction-*un*limited distribution of molecules in state A. To this end, the sample is illuminated with light that drives the transition from A to B. The intensity distribution $I(r)$ of the illuminat-

Figure 12–6. The RESOLFT principle. Diffraction-unlimited spatial resolution is achieved by saturating a linear but reversible optical transition from state A to state B. The simple explanation: A uniform dye distribution of state A is illuminated by a strongly modulated intensity light distribution that transfers the dye molecules to state B; ideally the modulation is perfect, so that the local intensity minima actually are zeros. Here, we choose the narrowest possible modulation $I(r) = \cos^2(2\pi r/\lambda)$. The left-hand panel shows $I(r)$ and the resultant probability $N_A(r)$ of the dye molecules being in state A. The right-hand panel depicts the dependence of $N_A(r)$ on the local intensity $I(r)$, exhibiting the typical saturation behavior. I_s is defined as the intensity at which half the molecules are transferred to state B. (a) At $I(r) < I_s$, the modulation of the light is replicated in $N_A(r)$. The narrowest distribution of $N_A(r)$ will also be limited by diffraction, because $I(r)$ is diffraction limited and the relationship between I and N_A is basically linear. (b) Increasing the maximum intensity moves the points at which $I(r)$ reaches I_s closer to the local intensity minimum, e.g., the zero. Because these are also the points at which $N_A(r)$ drops to 0.5, the FWHM of $N_A(r)$ is correspondingly reduced. (c) Further increasing the maximum intensity beyond I_s leads to a further reduction of the FWHM. If A is the fluorescent state, $N_A(r)$ is read out as the desired signal stemming from a narrow region. To obtain an image, the local minima are scanned across the sample. If the signal stems from state A, the intensity values $N_A(r)$ are read out subsequently and the image is assembled in a computer. If the signal is generated by the “majority population in state B” (e.g., state B is the fluorescent state) the function $1 - N_A(r)$ is read out and the image must be “inverted” later. This approach is challenged by signal-to-noise issues. (d) There is no theoretical limit to this method and its resolution is ultimately determined by the available laser power and the potentially limiting photodamage.

ing light is of course diffraction limited but features one or several positions with zero intensity. After illumination, the probability, $N_A(r)$, of finding molecules in state A depends on $I(r)$ and displays peaks where $I(r)$ is zero and no transitions to B are induced. If we choose the maximum intensity $I_{\max}$ in such a way that it is many times (ζ times, $\zeta = I_{\max}/I_s$ is called the saturation factor) higher than the saturation intensity, I_s , of the transition $A \rightarrow B$ (the threshold at which 50% of the molecules are



transferred into state B), then molecules will be almost exclusively in state B even at positions where $I(r)$ is only a small fraction ($\sim 5/\zeta$) of $I_{\max}$. This means that while molecules remain in state A in the intensity zeros, they are transferred to state B even at locations in the immediate vicinity of the zero. Thus, the peaks of $N_A(r)$ become very narrow, featuring steep edges at the same time. Figure 12–6 is presented in one dimension for clarity, but this idea is readily extended to all directions in space and hence to 3D imaging.

Their Fourier transform (the effective ex-citation OTF) of an increasingly sharp peak correspondingly becomes wider with an increasing $I_{\max}$. A complementary mathematical explanation for this fact is that higher-order contributions that are described by several autoconvolutions of the depletion pattern become more and more important for higher intensities. This is in contrast to very low intensities where the dependence of N_A on the depletion intensity is approximately linear: $N_A(r) = 1 - I(r)/I_s$. Here, a single additional convolution is introduced in frequency space and the theoretical cutoff of the support can only be pushed from the $4k$ we derived for conventional imaging to $6k$. Thus, the diffraction barrier is not truly broken for low intensities, but only shifted to a slightly higher value.

The generated probability distribution can be used for high-resolution imaging by applying scanning. Since the fluorescence stems exclusively from the immediate vicinity of the positions where $I(r)$ is zero, scanning the zero(s) through the sample with simultaneous recording of the fluorescence signal allows an image with basically unlimited spatial resolution in the far field to be assembled. For particular conditions, the resolution of this process is determined by the FWHM of the peaks in $N_A(r)$, which in turn is solely determined by the saturation factor. The FWHM can become infinitely narrow, that is down to the size of a single molecule. The reason why Abbe's spatial frequency cutoff does not apply in this case is simply the fact that in the RESOLFT concept the lens merely acts as a condensor collecting the fluorescence. For example, when scanning with a single zero, the detector can be placed right next to the sample to measure $N_A(r)$. Sequential readout is absolutely mandatory in a far-field optical system with a broken diffraction barrier, because—and Abbe was perfectly correct in this regard—the objective lens cannot transmit the higher spatial frequencies through the lens (Hell, 2007).

However, this does not imply that a RESOLFT microscope needs to be based on *single*-beam scanning. Parallelization is readily possible and hence detection with a conventional camera is feasible if the nodes are further apart than the classical resolution limit of the microscope. When the sharply localized fluorescence from the nodes is imaged onto the camera, the resulting spot will certainly be blurred as a result of diffraction, and possibly extend over several pixels on the camera. However, if the nodes are further apart than Abbe's diffraction barrier, each of the blurred spots can be assigned to the respective nodes. Sequential read-out of the camera makes it possible to integrate the measured signal of each blurred spot or line separately and to associate it with the corresponding sharply localized region in the sample from which the signal emerged.

Importantly, implementing such a “wide-field” detection in a microscope does not imply that subdiffraction resolution is possible in conventional microscopy, since scanning is still an essential element in the process of image formation. In a sense, “wide-field” is not the most appropriate term in any event, since RESOLFT-based concepts always imply that a part of the sample (at the very least, one single point) is omitted from the saturable transition. Consequently, camera-based RESOLFT approaches are essentially parallelized scanning concepts.

This scanning “wide-field” detection is readily explained in frequency space as well. The PSF describing the generated probability of the molecules of being in state A, $N_A(r)$, consists of periodic sharp peaks, and can be described as a convolution of a single peak with a comb function or with its multidimensional equivalent. In frequency space, this is equivalent to a multiplication of the broad frequency band of a single peak with a comb function. Thus the Fourier transform of $N_A(r)$ which dominates the excitation OTF if given by delta-peaks separated by the inverse distance of the nodes in $I(r)$. The speed with which they drop off towards large frequencies depends on the width of a single peak in $N_A(r)$. The effective OTF is given by the convolution of this series of delta-peaks with the detection OTF. Thus it will be continuous whenever the usable support of the latter is larger than the distance between the delta peaks in the excitation OTF. Not surprisingly, this just means that the intensity zeros of $I(r)$ are far enough apart so that the detection system can separate their fluorescence.

Note that complete depletion of A (or complete darkness of B) is *not* required. It is sufficient that the nonnodal region features a constant, notably lower probability to emit fluorescence, so that it can be distinguished from its sharp counterpart. Even if not A, but B is the brighter state, it is possible to read out B and obtain the same superresolved image after mathematical postprocessing that entails subtraction of signals from each other. While reading out B, in principle, also delivers unlimited resolution, this version is heavily challenged by signal-to-noise issues. This stems from the fact that the “bright” light from the nonnodal regions contributes with a substantial amount of photon shot noise.

It is also important to keep in mind that the nonlinearities introduced in these concepts are not analogous to the nonlinear interactions connected with m -photon excitation, m th harmonics generation, coherent anti-Stokes–Raman scattering (Sheppard and Kompfner, 1978; Shen, 1984), etc. In the latter cases, the nonlinear signal stems from the simultaneous action of *more than one photon* at the sample, which would work only at high focal intensities. In contrast, the nonlinearity brought about by saturation and depletion stems from a change in the *population* of the involved states, which is effected by a *single-photon* process, namely stimulated emission. Therefore, unlike m -photon processes, strong nonlinearities are achieved at comparatively low intensities.

Next, let us derive an estimate for the resolution achievable in such a system at finite depletion intensities (Hell, 2003, 2004). We denote the rates of $A \rightarrow B$ and $B \rightarrow A$ with k_{AB} and k_{BA} , respectively. The time evolution of the normalized populations of the two states n_A and n_B is then given by

$$dn_A/dt = -k_{AB}n_A + k_{BA}n_B = -dn_B/dt \quad (20)$$

Independently from its initial state, after a time

$$t \geq (k_{BA} + k_{AB})^{-1} \quad (21)$$

the equilibrium is approximately reached and the population of state A is given by

$$N_A = k_{BA}/(k_{BA} + k_{AB}) \quad (22)$$

During the process described in Figure 12-6, state A is depleted at a rate $k_{AB} = \sigma I$, where σ denotes the molecular cross section and the intensity I is written as photon flux per unit area. Hence, the equilibrium population is given by

$$N_A(r) = k_{BA}/[\sigma I(r) + k_{BA}] \quad (23)$$

And $N_A = 1/2$ for the saturation intensity

$$I_s = k_{BA}/\sigma \quad (24)$$

From Eq. (23) we see that where $I(r) \gg I_s$ all molecules end up in B. Thus, if we choose

$$I(r) = I_{\max} f(r) \quad (25)$$

with $I_{\max} \gg I_s$, molecules in state A are found only in the nodes of the diffraction-limited distribution function $f(r)$. As an example, we choose a sine-square intensity distribution, such as produced by a standing wave

$$f(x) = \sin^2(2\pi n x/\lambda) \quad (26)$$

for illumination, where n denotes the index of refraction of the medium. A simple calculation shows that the FWHM of the peaks of N_A and hence the resolution of the microscope is then given by

$$\Delta x = \frac{\lambda}{\pi n} \arcsin(1/\zeta) \cong \frac{\lambda}{\pi n \sqrt{\zeta}} \quad (27)$$

A saturation factor of $\zeta = 1000$ yields $\Delta x \approx \lambda/100$, but in principle the spot of "A molecules" can be continuously squeezed by increasing ζ . Scanning with such a spot and simultaneous recording of the signal from it deliver diffraction-unlimited resolution. Equation (27) quantitatively describes this possibility in a microscope using diffraction-limited beams.

If the intensity distribution $I(r)$ is produced by a lens, the largest frequency will be determined by the numerical aperture of the lens $n \sin \alpha$, due to diffraction. In this case, Eq. (27) changes into

$$\Delta x \cong \frac{\lambda}{\pi n \sin \alpha \sqrt{\zeta}} \quad (27')$$

If at the same time the molecules are being driven back from B to A by a light intensity distribution following a cosine-square form, we have $k_{BA} = \sigma_{BA} I_{\max}^{\text{BA}} \cos^2(2\pi n \sin \alpha x/\lambda)$, with σ_{BA} denoting the cross section of the transition. By defining the saturation intensity as $I_s = \sigma_{BA} I_{\max}^{\text{BA}}/\sigma$, we obtain

$$\Delta x \equiv \frac{\lambda}{\pi n \sin \alpha \sqrt{1 + \zeta}} \quad (27'')$$

This equation is particularly appealing, since for a vanishing saturation factor $\zeta = 0$ it almost assumes Ernst Abbe's diffraction limited form, whereas for $\zeta \rightarrow \infty$ the spot becomes infinitely small.

There are also adverse effects that need to be taken into account. The first is that state A cannot be completely emptied by even very intense illumination (e.g., because there is an excitation by the same beam) and the second is that state B may also contribute to the signal. Both can be considered by including a constant offset in Eq. (23):

$$N_A(r) = (1 - \delta)k_{BA}/[\sigma I(r) + k_{BA}] + \delta \quad (28)$$

This would result in the image consisting of a superresolved image plus a (weak) conventional image. The latter does not alter the frequency content of the image. Therefore, given sufficient SNR, the resolution will not deteriorate. In other words, if δ is sufficiently small so as not to swamp the image with noise, the conventional contribution can be subtracted (Hell and Kroug, 1995; Hell, 1997).

A further experimental problem is caused by imperfections of the intensity zeros. Imagine the standing wave is aberrated and approximately described by

$$f(x) = (1 - \gamma) \sin^2(2\pi n \sin \alpha x / \lambda) + \gamma \quad (29)$$

Such zeros with insufficient "depth" turn out to have a more serious impact on performance. The maximum signal in the intensity minima drops by a factor $(1 + \zeta \gamma)$, as a result. Following the same calculation as above we obtain

$$\Delta x = \frac{\lambda}{\pi n \sin \alpha} \sqrt{\gamma + \frac{1}{\zeta}} \quad (30)$$

and therefore the maximum achievable resolution is given by $\lambda \sqrt{\gamma} / \pi n \sin \alpha$. At $\gamma \sim 1\%$ resolutions of $\lambda/20$ at saturation factors of 100 can be achieved without losing more than half the signal in the intensity minima.

While RESOLFT is far more intuitively explained in the way presented above, it is also helpful to take a look at the frequency space to relate these findings to the concepts and results presented in the first part of this chapter. The dependence of the effective excitation PSF, i.e., the distribution of the probability that a molecule actually emits a fluorescence photon, is governed by the saturation level-dependent value of $N_A(r)$ expressed in Eq. (23). If for $I(r) = 0$ the microscope begins, for example, with a conventional PSF $h_c(r)$ used for imaging the distribution $N_A(r)$ onto a camera, the effective PSF of the system is given by

$$h(r) = h_c(r) \frac{1}{[1 + \sigma I(r) / k_{BA}]} \quad (31)$$

Now let $\sigma I_{\max} / k_{BA} = \zeta$ and $I(r) = I_{\max} f(r)$ as above, then we can expand (31) in a Taylor series

$$h(r) = \frac{h_c(r)}{1 + \zeta} \sum_v \left[\frac{\zeta}{\zeta + 1} \right]^v (1 - f(r))^v \quad (32)$$

With $g(r) = 1 - f(r)$ and $\xi = \zeta/(1 + \zeta)$ we obtain the OTF after Fourier transformation

$$o(k) = \frac{\hat{h}_c(k)}{1 + \zeta} \otimes (\delta(k) + \xi \hat{g} + \xi^2 \hat{g} \otimes \hat{g} + \xi^3 \hat{g} \otimes \hat{g} \otimes \hat{g} + \dots) \quad (33)$$

At low intensities, ζ and therefore ξ is so small that only the linear term is relevant and the convolution extends the support to $6k$ as discussed above. The larger the maximum intensity, the more important higher orders of the Taylor series will become. These involve multiple auto-convolutions of the function g extending the support further and further.

While a quantitative treatment in frequency space is more complicated and less intuitive than the one introduced in the previous section, the following analysis gives a feel for the effect of the saturation factor and also illustrates the possible vast expansion of the OTF support. For the sake of simplicity we assume a Gaussian form of the light distribution function

$$f(x) = 1 - \exp(-x^2/2a^2) \quad (34)$$

The properly normalized m -fold auto-convolution of g is then given by

$$\otimes_m \hat{g}(k) = a \sqrt{2\pi/m} \exp(-a^2 k^2/(2m)) \quad (35)$$

Now let us assume that the useful support ends at a frequency where the OTF is attenuated to a small fraction ε of its value at small frequencies. For large saturation factors the influence of the convolution with the confocal OTF on the cut-off frequency can be neglected and we have to calculate the sum in brackets in (33). Substituting (35) into equation (33) and approximating the sum by an integral we get for the term in brackets

$$o(k, \zeta) \cong -i\sqrt{\pi/\ln \xi} \exp(-iak\sqrt{2\ln \xi}) \quad (36)$$

For large saturation factors we can write $\ln \xi = \ln \zeta - \ln(\zeta + 1) \cong -1/\zeta$ and obtain

$$o(k, \zeta) \cong \sqrt{\pi\zeta} \exp(-ak\sqrt{2/\zeta}) \quad (37)$$

This means that the attenuation of the modulus of the OTF at large frequencies is anti-proportional to the square-root of the saturation factor. This is equivalent to saying that the resolution increases with the square root of the saturation factor just as we expected from our previous analysis.

4.1 STED Microscopy

STED microscopy produces subdiffraction resolution and subdiffraction-sized fluorescence volumes in exactly the manner described above by the depletion of the fluorescent state of the dye. Depletion inherently implies saturation of the depleting transition. At present, it is realized in a (partially confocalized) spot-scanning system due to a number of technical advantages, but it has been conceptually clear from the outset

that nonconfocalized detection is viable as well (Hell and Wichmann, 1994). The principal idea, a schematic setup and an exemplary measurement of the resolution increase, is shown in Figure 12–7. The fluorophore in the fluorescent state S_1 (state A) is stimulated to the ground

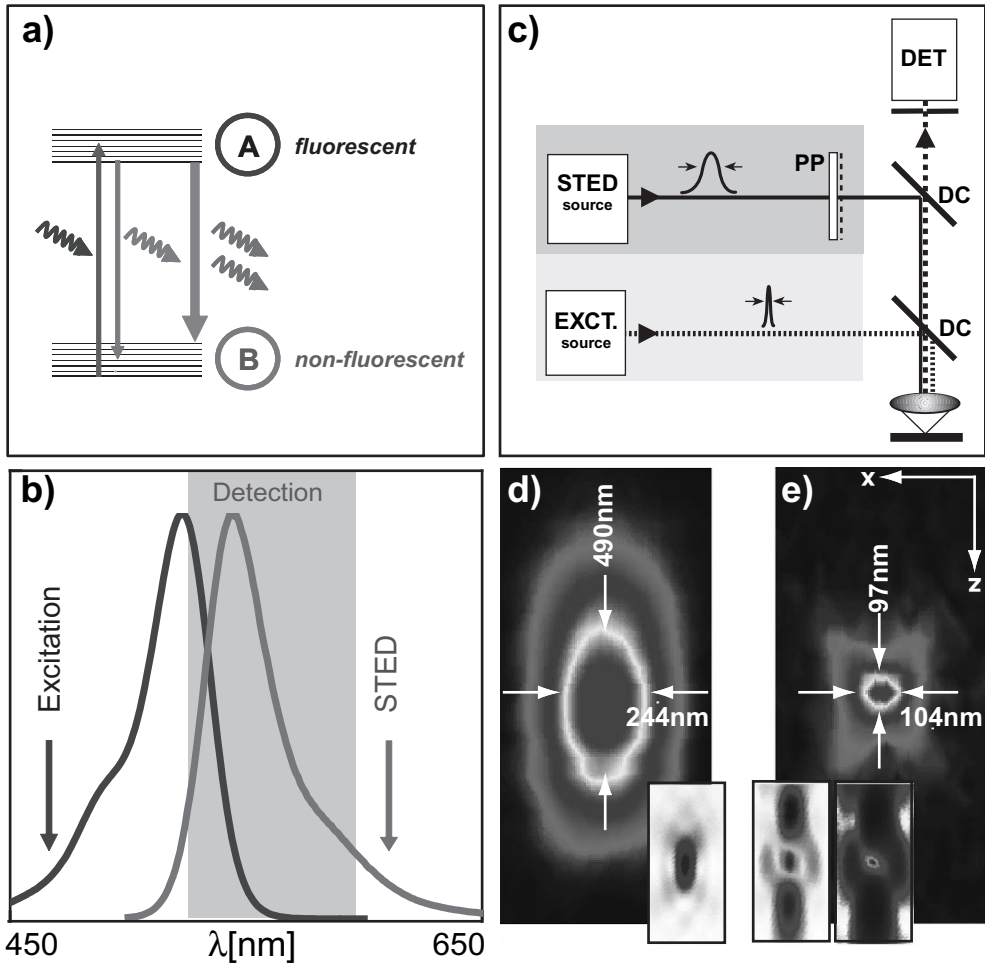


Figure 12–7. Stimulated emission depletion (STED) was the first implementation of the RESOLFT principle. (a) Dye molecules are excited into the S_1 (state A) by an excitation laser pulse. (b) Fluorescence is detected over most of the emission spectrum. However, molecules can be quenched back into the ground state S_0 (state B) using stimulated emission before they fluoresce by irradiating them with a light pulse at the edge of the emission spectrum shortly after the excitation pulse and before they are able to emit a fluorescence photon. Saturation is realized by increasing the intensity of the depletion pulse and consequently inhibiting fluorescence everywhere except at the “zero points” of the focal distribution of the depletion light. (c) Schematic of a point-scanning STED microscope. Excitation and depletion beams are combined using appropriate dichroic mirrors (DC). The excitation beam forms a diffraction-limited excitation spot in the sample (inset in d) while the depletion beam is manipulated using a phase-plate (PP) or any other device to tailor the wavefront in such a way that it forms an intensity distribution with a nodal point in the excitation maximum (left inset in e). The third inlay shows the resulting quenching probability when saturating the depletion process. (d) and (e) show an experimental comparison between the confocal PSF and the effective PSF after switching on the depleting beam. Note the doubled lateral and five-fold improved axial resolution. The reduction in dimensions (x, y, z) yields ultrasmall volumes of subdiffraction size, here 0.67 al (Klar et al., 2000), corresponding to an 18-fold reduction compared to its confocal counterpart. The spot size is not limited on principle grounds but by practical circumstances such as the quality of the zero and the saturation factor of depletion. (See color plate.)

state S_0 (state B) with a doughnut-shaped beam. The saturated depletion of S_1 confines fluorescence to the central intensity zero. With typical saturation intensities ranging from 1 to 100 MW/cm², saturation factors of up to 120 have been reported (Klar et al., 2000, 2001). This should yield a 10-fold resolution improvement over the diffraction barrier, but imperfections in the doughnut have limited the improvement to 5 to 7-fold in experiments (Klar et al., 2001).

As already stated, light microscopy resolution can be described either in real space or in terms of spatial frequencies. In real space, the resolution is assessed by the FWHM of the focal spot. The measurements depicted in Figure 12–8 were carried out with an excitation wavelength of $\lambda = 635$ nm, an oil immersion lens with a numerical aperture of 1.4, and with the smallest possible probe: a single fluorescent molecule (Westphal and Hell, 2005; Westphal et al., 2003). Figure 12–8a shows the measured profile of the PSF in the focal plane (x) for a conventional

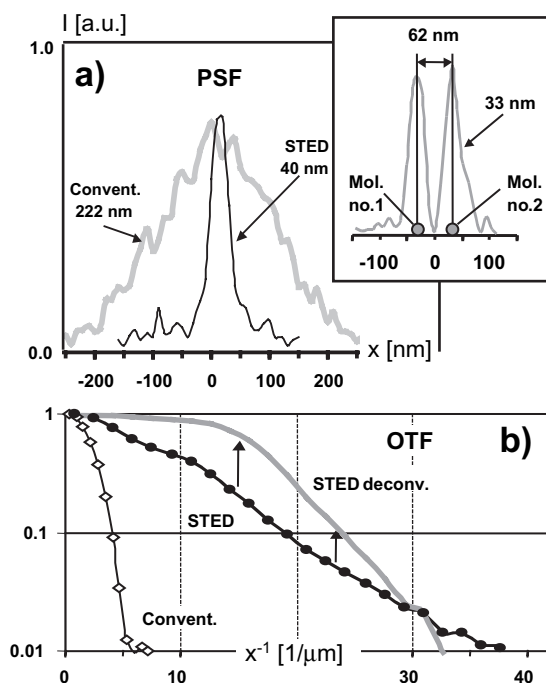


Figure 12–8. (a) Comparison of the effective PSF's lateral intensity profile for confocal and STED microscopy indicating an ~5.5-fold resolution increase in the latter. (b) Lateral cuts through the effective OTFs giving the bandwidth of the lateral spatial frequencies passed to the image. The data plotted in (a) and (b) are gained by probing the fluorescent spot of a scanning microscope with a single molecule of the fluorophore JA 26 using a numerical aperture of 1.4 (oil) objective lens and at wavelengths of 635 nm (excitation), 650–720 nm (fluorescence collection), and 790 nm (STED). The inset demonstrates subdiffraction resolution with STED microscopy. Two identical molecules located in the focal plane that are only 62 nm apart can be entirely separated by their intensity profile in the image. A similar clear separation by conventional microscopy would require the molecules to be at least 300 nm apart. (Date adapted from Westphal et al.)

fluorescence microscope along with its sharper subdiffraction STED fluorescence counterpart. STED leads to improvement in resolution by a factor of approximately 5.5.

Figure 12–8b shows the OTF of a conventional microscope along with the enlarged OTF of the STED fluorescence microscope. As expected, the effective OTF's support in the confocal case ends at approximately $(2/635\text{ nm} + 2/720\text{ nm}) = 5.91/\mu\text{m}$. For STED, we included the OTF after successful linear deconvolution, which restores higher spatial frequencies that are not swamped by noise. The region of usable OTF support is approximately marked by the region where frequencies are enhanced by the deconvolution process without producing artifacts and is ~ 5.5 times larger than for the confocal case. This marks a fundamental breaking of Abbe's diffraction barrier in the focal plane. The inlay demonstrates the resulting subdiffraction resolution exemplified by the linearly deconvolved STED image of two molecules at a distance of 62 nm. They are distinguished in full by two narrow peaks (Westphal et al., 2003). As a result of deconvolution, the individual peaks are sharper (33 nm FWHM) than the initial peak of 40 nm FWHM.

Very recently, utilizing STED wavelengths of $\lambda = 750\text{--}800\text{ nm}$, a lateral FWHM of down to 16 nm has been achieved in experiments with single JA 26 molecules spin-coated on a glass slide.¹⁴ Measuring the resolution as a function of the applied STED intensity confirmed the predicted increase of the resolving power with the square root of the saturation factor (see Figure 12–9). Of course the cutoffs presented are based on a somewhat arbitrary definition of what can be considered “usable frequencies” at a certain signal-to-noise ratio. However low this threshold is set, the confocal support cannot extend beyond $\sim 6/\mu\text{m}$ while the STED-OTF's support is theoretically unlimited.

The one-dimensional (1 D) phase-plate yielding 16 nm FWHM is optimized for maximum resolution improvement in the lateral direction perpendicular to the polarization of the depleting light and leads to an intensity distribution with two strong peaks at either side of the excitation maximum (Keller et al., in preparation). Because the depleting light is polarized, the resolution gain depends on the orientation of the molecules. However, a considerable increase in resolution is still possible for the second phase-plate, which yields a doughnut-shaped intensity distribution and thus an almost isotropic resolution increase in the lateral directions when using circularly polarized light.

To “squeeze” the fluorescence spot in both lateral directions two STED beams aberrated with 1 D phase-plates oriented at 90° to each other can be combined. Together with circularly polarized excitation, almost uniform resolution in the focal plane is achieved as shown in Figure 12–10. A series of xy images acquired with different STED beam powers demonstrate the resolution increase and concomitant widening of the OTF when the applied saturation factor increases (Schönle et al., in preparation). This combination of two incoherent beams causes the resolution to depend on the orientation of the transition dipole and results in spikes along the x and y direction of the OTF when imaging randomly oriented fluorophores (see Figure 12–10). New phase-plates have been proposed to avoid such effects and to improve the effective

saturation factors at a given total STED power. Incoherent combination can then be used to improve the resolution in all three spatial dimensions. The resulting PSFs exhibit very weak dependence on dipole orientation (Keller et al., in preparation) and allow application of STED to the imaging of biological specimen and reliable subsequent linear deconvolution (Willing et al., in preparation).

STED microscopy has also been successfully applied to the imaging of biological samples. Subdiffraction images with three-fold enhanced axial and doubled lateral resolution have been obtained with membrane-labeled bacteria and live budding yeast cells (Klar et al., 2000). While there is some evidence for increased nonlinear photobleaching of some dyes when increasing the depletion intensity (Dyba and Hell, 2003), there is no reason to believe that the intensities currently applied would be detrimental to live cells. This is not surprising since the intensities are two to three orders of magnitude lower than those used in multiphoton microscopy (Denk et al., 1990). Moreover, STED has proven to be single molecule sensitive, despite the proximity of the

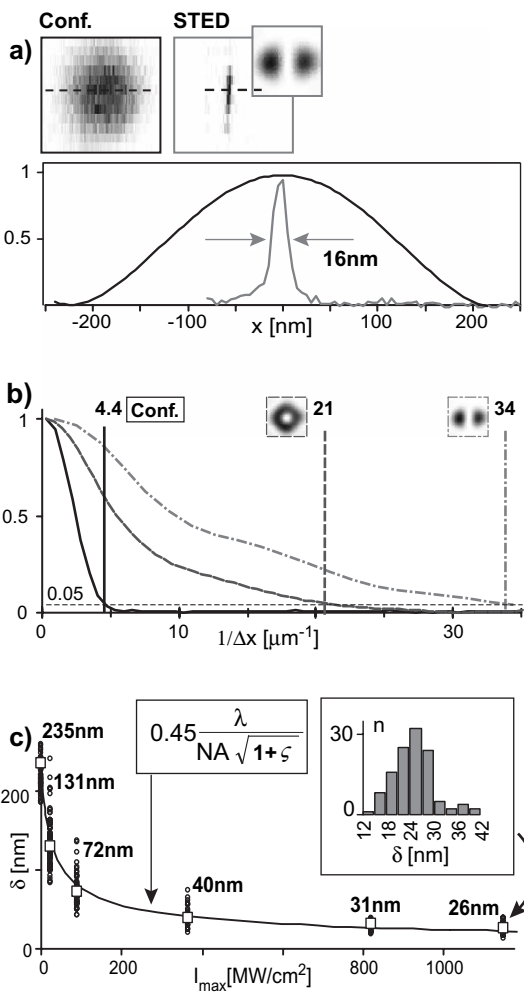


Figure 12-9. STED microscopy reduces the fluorescence focal spot size to a degree far below the diffraction limit: (a) spot of a confocal microscope (left) compared with that in an STED microscope (right) utilizing a y -oriented intensity valley for STED (upper right inset, not to scale) squeezing the spot in the x direction to 16 nm width. (b) As also observed in Figure 12-8, the bandwidth in STED is fundamentally increased over confocal microscopy. The graph shows the normalized magnitude of optical transfer function (OTF) as a function of inverse distance. For the “1D” depletion scheme, the usable support of the OTF is increased by almost a factor of 8. When using a doughnut-shaped depletion beam with a “wider” intensity zero, the OTF support is still extended almost five-fold. (c) The average focal spot size decreases with the STED intensity following a square-root law, in agreement with Eq. (27). Because the resolution depends on molecule orientation, the spot sizes were measured for several tens of single molecules. The curves follow the mean values (squares) and the inset discloses the histogram of the measured spot sizes at 1100 (MW/cm^2) with the minimum FWHM at 16 nm and a 26 nm average.

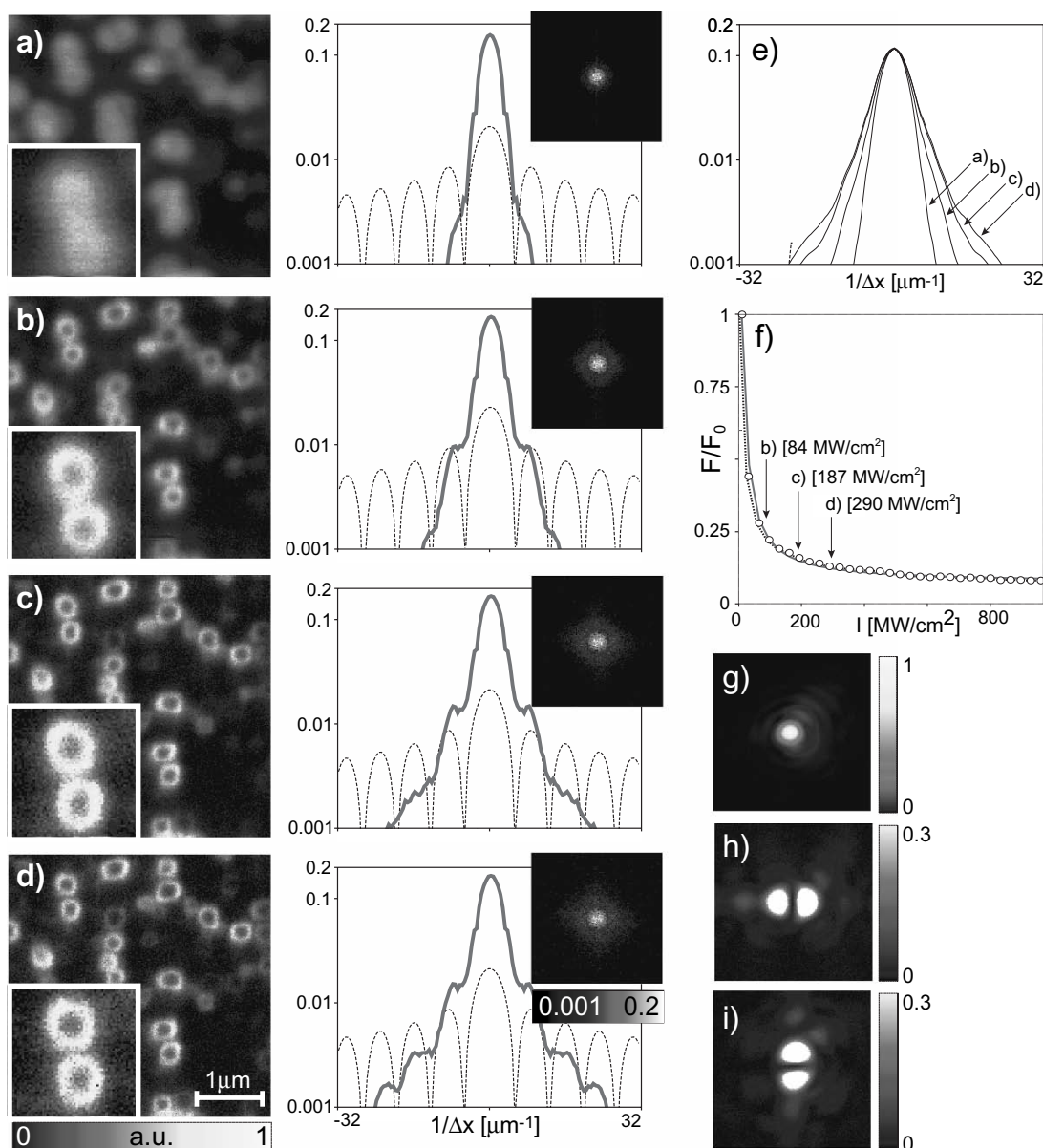


Figure 12-10. Images of a wetted Al_2O_3 matrix featuring z-oriented holes (Whatman plc, Brentford, UK) with a spin cast of a dyed (JA 26) polymethyl methacrylate solution. The rings formed in this way are ~ 250 nm in diameter and are barely resolved in confocal mode. (a–d) The confocal image (a) and STED images with two depleting beams perpendicularly polarized and aberrated by “1D” phase-plates (b–d). The excitation PSF (g) and the STED PSF for y polarization (h) and x polarization (i) are shown on the right. The STED intensity was chosen at the spots marked in the saturation curve (f). The smaller effective spot size also results in an extended OTF as seen in the second column. Here, the insets show the 2D Fourier transformation of the images in the left and the graphs show a profile along the x direction. Note the logarithmic scales. The Fourier transform of the image is given by the product of OTF and the Fourier transform of the object [Eq. (2)]. For such regular structures, an estimate for the modulus of the OTF can therefore be gained by estimating the latter and solving for the OTF. The dashed line shows the Fourier transform of a ring with a diameter of 275 nm and a width of 50 nm and the estimated OTF is presented in (e). (f) The suppression of fluorescence resulting from stimulated emission. The phase-plates were removed and the ratio of fluorescence without STED light (F_0) and with the STED beams switched on (F) was recorded. The intensities are pulse intensities per beam at the global maximum. (See color plate.)

STED wavelength to the emission peak. In fact, individual molecules have been switched on and off by STED upon command (Kastrup and Hell, 2004; Westphal et al., 2003).

The power of STED and 4Pi microscopy has been synergistically combined to demonstrate for the first time an axial resolution of 30–40 nm in focusing light microscopy (Dyba and Hell, 2002). The intensity distribution of the depleting light is formed by a 4Pi setting with destructive interference at the geometric focus leading to a zero intensity there and two neighboring maxima at a distance of approximately $\lambda/4$. This results in superior xz images, and the technique has initially been successfully applied to membrane-labeled bacteria (Dyba and Hell, 2002). More recently, STED-4Pi microscopy has been extended to immunofluorescence imaging (Figure 12–11). A spatial resolution of ~50 nm has been demonstrated in the imaging of the microtubular meshwork of a mammalian cell.⁷⁶ These results indicate that the basic physical obstacles to attaining a 3D resolution of the order of a few tens of nanometers have been overcome. Since the samples were mounted in an aqueous buffer (Dyba and Hell, 2002; Dyba et al., 2003), the results indicate that the optical conditions for obtaining subdiffraction resolution are met under the physical conditions encountered in live cell imaging.

It is to be expected that ultrasmall detection volumes created by STED will also be useful in a number of sensitive bioanalytical techniques. Fluorescence correlation spectroscopy (FCS) (Magde et al., 1972) relies on small focal volumes to detect rare molecular species or interactions in concentrated solutions (Eigen and Rigler, 2001; Elson and Rigler, 2001). While volume reduction can be obtained by nanofabricated structures (Levene et al., 2003), STED may prove instrumental in attaining ultrasmall spherical volumes at the nanoscale inside samples that do not allow for mechanical confinement. The latter fact is particularly important to avoid an alteration of the measured fluctuations by the nanofabricated surface walls.

In fact, the viability of STED FCS has recently been shown in an experiment (Kastrup et al., 2005). In a particular implementation STED FCS has witnessed a reduction of the focal volume by a factor of five along the optic axis and a concomitant reduction of the axial diffusion time. The initial experiments showed that for particular dyewavelength combinations the evaluation of the STED FCS data might be complicated by a seemingly uncorrelated background at the outer wings of the fluorescence spot where STED may not completely suppress the signal. Further investigations will show whether this challenge is easily overcome in the near future. In any case, published results suggest a further decrease of the volume by another order of magnitude (Westphal et al., 2003; Irie et al., 2002). An inherent disadvantage of STED is the necessity of an additional pulsed light train that is tuned to the red edge of the emission spectrum of the dye. Nevertheless STED is to date the only known method for “squeezing” a fluorescence volume to the zeptoliter scale without making mechanical contact. Thus, the creation of ultrasmall volumes, tens of nanometers in diameter, by STED may be a pathway to improving the sensitivity of fluorescence-based bioanalytical techniques (Weiss, 2000; Laurence and Weiss, 2003).

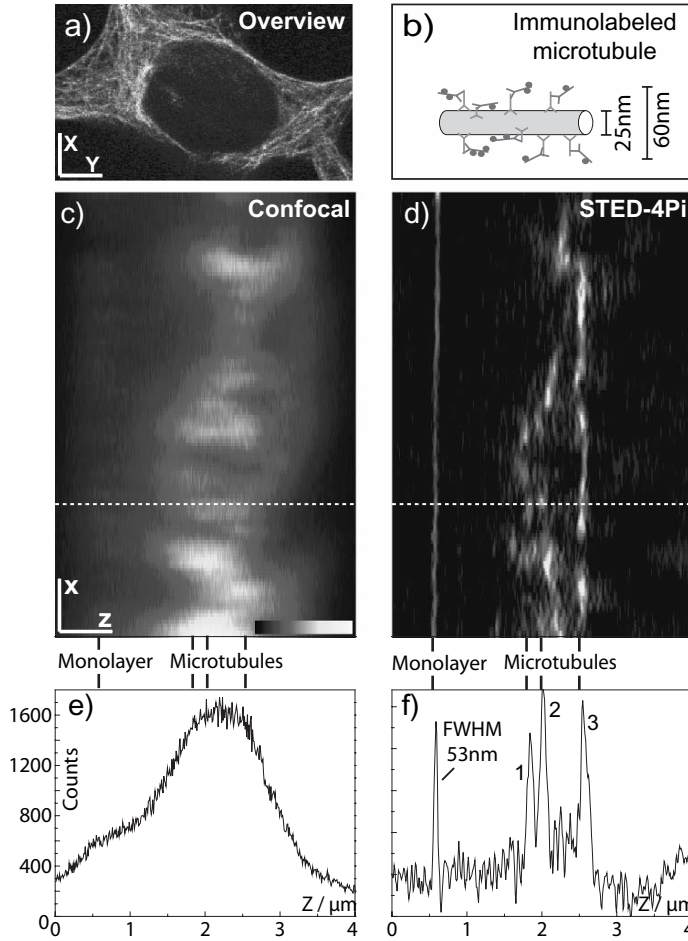


Figure 12–11. Subdiffraction immunofluorescence imaging with STED-4Pi microscopy. (a) Overview image (xy) of the microtubular network of an HEK cell. (b) Sketch of typical dimensions of a labeled microtubule fluorescently decorated via a secondary antibody. (c) and (d) Standard confocal and STED-4Pi xz image recorded at the same site of the cell; the straight line close to the cell stems from a monomolecular fluorescent layer attached to the adjacent coverslip. In both images, the pixel size was 95×9.8 nm in the x and z direction, respectively; the dwell time per pixel was 2 ms. Note the fundamentally improved clarity in (d). The STED-4Pi microscope's PSF features two low side lobes caused by the secondary minima STED intensity distribution. These lobes are $<25\%$ and were removed in the STED-4Pi image using linear filtering as outlined in the text [see Eq. (16ff)]. (e) and (f) Corresponding profiles of the image data along the dashed lines in (b) and (c) quantify the improved axial resolution of the STED-4Pi microscopy mode (f) over the confocal benchmark. Peaks 1, 2, and 3 due to microtubules are broader than the response to the monolayer. Note the ability of the STED-4Pi microscope to distinguish adjacent features. (See color plate.)

An important step toward far-reaching applicability of STED microscopy was the demonstration of the suitability of laser diodes both for excitation and for depletion (Westphal et al., 2003). However, several issues remain to be addressed. Due to the considerably smaller detection volumes, the signal per pixel is reduced and the amount of pixels to be recorded increases. Therefore, it will be important to incorporate STED into fast, ideally parallelized scanning systems.

While the transition to shorter wavelengths will further increase resolution by a factor of ~ 1.5 , it is most likely that the ultimate resolution limit in STED will be set by the stability of the marker used. The photostability of current markers was considerably improved by stretching the depleting pulse to >300 ps (Dyba and Hell, 2003), but it might not be easily possible to attain saturation factors $\zeta > 200$ in the near future. Nevertheless, according to Eq. (27) $\zeta = 200$ should already yield an improvement by one order of magnitude, provided that the actual intensity value at the “intensity zero” is indeed negligible at this saturation level.

As explained above (Eq. 30), the actual “depth” of the zero codetermines the attainable resolution, because for relatively high saturation factors the saturable transition also becomes effective at the zero point or points. So far, typical depths were in the range of $\gamma = 1\text{--}2.5\%$ of the global maximum of the depleting intensity $I(r)$. The zero could be a single point, as in a single beam scanning system, but in the case of a parallelized system, it may also be a line or an array of points or lines. The actual depth of the zeros will certainly depend on the particular setup and the quality of optical components and proper alignment. Independently of implementation details, active optical elements such as wavefront phase modulators will be a valuable tool to further “deepen” the zeros, which in turn will allow the full potential of the attained saturation level to be exploited for improvement in resolution.

4.2 Variations of RESOLFT Microscopy and Producing Large Saturation Factors at Low Power

At this point, we reiterate that RESOLFT is not restricted to the process of stimulated emission, but can exploit any reversible (linear) transition driven by light; the attainable resolution is determined by the ratio of driving intensity and the competing transition rate k_{BA} . If the applicable intensity is limited by the onset of photodamage to the marker or even to the sample, marker constructs must be found where high saturation levels are attained at lower intensities. This is certainly the case if the rate competing with the transition to be saturated is lower.

One such example is the GSD mentioned earlier. In this version of the RESOLFT concept the ground state (now state A) is depleted by targeting an excited state (B) with a comparatively long lifetime (Hell and Kroug, 1995; Hell, 1997), such as the meta-stable triplet state T_1 . In many fluorophores T_1 can be reached through the S_1 with a quantum efficiency of $1\text{--}10\%$ (Lakowicz, 1983). Being a forbidden transition, the relaxation of the T_1 is $10^3\text{--}10^5$ times slower than that of the S_1 , thus yielding $I_s = 0.1\text{--}100$ kW/cm². The signal to be measured (from the intensity zero) is the fluorescence of the molecules that remained in the singlet system; this measurement can be accomplished through a synchronized further excitation (Hell and Kroug, 1995). For many fluorophores, this approach is not straightforward, because T_1 is involved in the process of photobleaching, but there are potential alternatives such as the meta-stable states of rare earth metal ions that are fed through chelates.

Also proposed has been depleting the ground state S_0 by populating the S_1 (now B) (Heintzmann et al., 2002). This is the technically simplest

realization of saturated depletion, since it requires only excitation wavelength matching. However, as the fluorescence emission maps the spatially extended “majority population” in state B, the superresolved images (represented by state A) are negative images hidden under a bright signal from B. Hence photon noise from the large signal might swamp the fluorescence minima that occur when intensity zeros, where no fluorescence is excited, are colocalized with fluorophores. The subsequent computational extraction of the positive image is therefore very dependent on an excellent signal-to-noise ratio. The saturation intensity is of the same order as in STED microscopy, because the saturation of fluorescence also competes with the spontaneous decay of S_1 . This probably results in photostability issues similar to the case of STED. In fact, the photobleaching should be exacerbated, since the saturated transition is effected with higher energy photons that are generally more prone to facilitating photochemical reactions. Pumping the dye to a higher state rather than into the ground state also favors photolability. Moreover, the fact that a large number of dye molecules constantly undergo excitation–emission cycles to image a comparatively small spot adds to the problem. Finally, saturation of the S_1 will be possible only if the long-lived triplet state is not allowed to build up during repeated excitation. As most dyes feature a triplet relaxation rate of $>1\mu\text{s}$ (that strongly depends on the environment), effective triplet relaxation requires a pulse repetition rate $<500\text{kHz}$. Nevertheless, due to the simplicity of raw data acquisition it may remain an attractive method for the imaging of very bright and photostable samples.

One possible solution to the quest for large saturation factors at low intensities should be compounds with two (semi)stable states (Hell et al., 2003; Dyba and Hell, 2002). If the rate k_{BA} (and the spontaneous rate k_{AB}) almost vanish, large saturation factors are attained at very low intensities. The lowest useful intensity is then determined by the slowest acceptable imaging speed, which is ultimately determined by the switching *rate*. A favorable aspect is that in most bistable compounds the speed of the actual switching mechanism, i.e., of the conformational change, is less than a few nanoseconds, which is much faster than the typical pixel dwell time in scanning. In the ideal case, the marker indeed is a bistable fluorescent compound that can be photoswitched at separate wavelengths, from a fluorescent state A to a dark state B, and vice versa, where spontaneous rates will not influence this compromise.

Recently, a photoswitchable coupled molecular system, based on a photochromic diarylethene derivative and a fluorophore, has been reported (Irie et al., 2002). Using the kinetic parameters reported, Eq. (27) predicts that focusing of less than $100\mu\text{W}$ of deep-blue “switchoff light” to an area of 10^{-8}cm^2 for $50\mu\text{s}$ should yield better than 5nm spatial resolution. Future targeted optimization of photochromic or other compounds to fatigue-free switching and visible light operation could therefore open up radically new avenues in microscopy and data storage (Hell et al., 2003).

For live cell imaging, fluorescent proteins have many advantages over synthetic dyes. Many of them feature dark states with light-driven transitions (Hell, 1997; Hell et al., 2003). If the spontaneous lifetimes of

these states are longer than 10 ns such proteins may permit much larger saturation factors than those used in STED microscopy today. The most attractive solution, however, involves fluorescent proteins that can be “switched on and off” at different wavelengths (Hell et al., 2003). An example is as FP595 (Lukyanov et al., 2000), insertion of the published data into Eq. (27) predicts saturated depletion of the fluorescence state with intensities of less than a few W/cm^2 and, under favorable switching conditions, spatial resolutions of better than 10 nm.¹² The involved intensities should also enable parallelization of saturation through an array of minima or dark lines.

Initial realization of very low-intensity depletion microscopes may be challenged by switching fatigue (Irie et al., 2002) and overlapping action spectra (Lukyanov et al., 2000). Nevertheless, the prospect of attaining nanoscale resolution with regular lenses and focused light is an incentive to surmount these challenges by strategic fluorophore modification (Hell et al., 2003) and this or similar types of fluorescent proteins are a good starting point for these efforts.

5 Conclusion

The coherent use of opposing lenses enables the axial resolution of a far-field microscope to be improved by a factor of 3–7. The improvement in resolution occurs if the *spherical* wavefronts of illumination are coherently added at the focus, or the emerging spherical wavefronts of fluorescence are coherently added at the detector, because in both cases the total aperture of the system is enlarged. The latter fact is the basic tenet of the concept of 4Pi microscopy. The mere implementation of interference of low aperture (or flat) wavefronts is insufficient, because the “4Pi concept” of aperture increase is the underlying physical element of the improvement in resolution. Consequently, the success of increasing the axial resolution by two opposing lenses fully relies on large aperture angles and on the degree to which the aperture angle is utilized in the particular implementation.

Adding the spherical wavefronts both for illumination and detection, (multiphoton) 4Pi confocal microscopy of type C utilizes the lenses’ aperture as much as possible. As a result, this imaging mode features a contiguous and weakly modulated effective OTF that is robust enough for live cell imaging in an aqueous environment. On the other hand, I⁵M uses a mutually incoherent set of flat standing waves with varying spatial frequency for illumination. To be viable, I⁵M adds the spherical wavefronts of fluorescence emission at the detector; in other words, it uses a 4Pi scheme for detection. The compromise in the illumination path with regard to the “4Pi concept” bestows the OTF of the I⁵M with internal regions of very weak frequency transfer. The I⁵M is therefore more prone to artifacts and probably not reliably applicable in live cells.

With the 4Pi idea as the key physical element in the process, it is not surprising that both 4Pi microscopy and I⁵M would benefit from lenses with higher semiaperture angles than those currently available. An

increase of the semiaperture angle even by only a few degrees would make a large difference in I^5M . This difference would be even more decisive than in 4Pi microscopy, which already exploits the available aperture to the highest possible degree.

Although 4Pi microscopy has demonstrated a 3D resolution in the 100 nm range following deconvolution, the method is still limited by diffraction. The latter is no longer the case with the emerging approaches exploiting a reversible saturable transition between two states of a marker, which we termed RESOLFT. In a microscope using the RESOLFT principle, the resolution is no longer limited by diffraction but by the attainable level of saturation of a (linear) optical transition in the marker molecule. Therefore the “hard” theoretical resolution barrier is replaced by a “soft” barrier determined solely by practical conditions such as available laser power, cross sections, and the stability of the dye and the sample. The enhanced resolution has already been demonstrated (Hänninen, 2002) in a number of imaging experiments.

It is important to realize that while an effectively nonlinear interaction between light and dye is the basis for this resolution increase, these methods do not require transitions involving more than one photon at a time, such as m -photon excitation, m th harmonics generation, or coherent anti-Stokes–Raman scattering (Sheppard and Kompfner, 1978; Shen, 1984). This means that the required intensities are not determined by the very small cross sections of these processes and the requirement for ultrahigh (peak) intensities. By contrast, the saturation of a linear optical transition depends on the basic kinetics of the population of the involved states. Consequently, pulse-length requirements are less strict and, most importantly, the required intensities can be significantly reduced by choosing appropriate spectroscopic systems.

To date STED microscopy is the most advanced implementation of the RESOLFT principle. It has been well characterized by its application to single fluorescent molecules and has been applied to imaging fixed and live, albeit simple biological specimens. Improvements in resolution by a factor of up to eight have already been demonstrated and the resulting fluorescent volumes are the smallest that have ever been created with focused light. The results hitherto obtained with STED must not be considered as a new limit but as proof of the viability of the concept of RESOLFT and STED microscopy in particular. Because this method is very young, future research on spectroscopy conditions and on practical aspects (Stephens and Allen, 2003) should lead to further improvements.

Nevertheless, due to the rather fast relaxation of the excited state STED requires intensities of the order of at least several tens of MW/cm². Hence, there will be a practical limit to the applicable intensity and to the saturation level attainable under practical conditions. Intriguingly, the intensities required for obtaining a high saturation level can be fundamentally lowered in other implementations of the RESOLFT concept. This is particularly true when utilizing optically bistable markers, such as photoswitchable dyes and photochromic fluorescent proteins. Both are very promising candidates for providing high levels of saturation at ultralow intensities of light (Hell et al., 2003). In fact, we

expect them to play an important role in translating nanoscale resolution into the noninvasive, all-optical imaging of live cells. Dedicated synthesis or protein engineering might eventually uncover a whole new range of suitable markers.

Light microscopy is still commonly portrayed as fundamentally resolution limited. However, about a decade ago, concepts emerged that broke the diffraction barrier postulated by Abbe in 1873. These developments are poised to radically extend the field of application for far-field light microscopy and eventually lead to far-field “nanoscopes” operating with regular lenses and visible light.

Acknowledgments. The authors thank all members of the Department of NanoBiophotonics for contributions to this work and valuable discussions. Much of the results described in this chapter have been adopted from original work with significant contributions by A. Egner, M. Nagorni (4Pi), M. Dyba, V. Westphal, B. Harke, and J. Keller (STED). We thank J. Jethwa, J. Bewersdorf, and G. Donnert for valuable discussions and critical reading of the manuscript.

References

- Abbe, E. (1873). Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Arch. Mikr. Anat.* **9**, 413–420.
- Albrecht, B., Failla, A.V., Schweizer, A. and Cremer, C. (2002). Spatially modulated illumination microscopy allows axial distance resolution in the nanometer range. *Appl. Opt.* **41**(1), 80–87.
- Bahlmann, K. and Hell, S.W. (2000). Polarization effects in 4Pi confocal microscopy studied with water-immersion lenses. *Appl. Opt.* **39**(10), 1653–1658.
- Bahlmann, K., Jakobs, S. and Hell, S.W. (2001). 4Pi-confocal microscopy of liver cells. *Ultramicroscopy* **87**, 155–164.
- Bailey, B., Farkas, D.L., Taylor, D.L. and Lanni, F. (1993). Enhancement of axial resolution in fluorescence microscopy by standing-wave excitation. *Nature* **366**, 44–48.
- Bertero, M., Boccacci, P., Brakenhoff, G.J., Malfanti, F. and Van der Voort, H. T.M. (1990). Three-dimensional image restoration and super-resolution in fluorescence confocal microscopy. *J. Microsc.* **157**, 3–20.
- *Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J. and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometre resolution. *Science* **313**, 1642–1645.
- Blanca, C.M., Bewersdorf, J. and Hell, S.W. (2002). Determination of the unknown phase difference in 4Pi-confocal microscopy through the image intensity. *Opt. Commun.* **206**, 281–285.
- Bloembergen, N. (1965). *Nonlinear Optics*. (Benjamin, New York).
- Born, M. and Wolf, E. (1993). *Principles of Optics*, 6th ed. (Pergamon Press, Oxford).
- *Bretschnneider, S., Eggeling, C. and Hell, S.W. (2007). Breaking the diffraction barrier in fluorescence microscopy by optical shelving. *Phys. Rev. Lett.* **98**, 218103.
- Carrington, W.A., Lynch, R.M., Moore, E.D.W., Isenberg, G., Fogarty, K.E. and Fay, F.S. (1995). Superresolution in three-dimensional images of fluorescence in cells with minimal light exposure. *Science* **268**, 1483–1487.
- Denk, W., Strickler, J.H. and Webb, W.W. (1990). Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76.

- *Donnert, G., Keller, J., Medda, R., Andrei, M.A., Rizzoli, S.O., Luhrmann, R., Jahn, R., Eggeling, C. and Hell, S.W. (2006). Macromolecular-scale resolution in biological fluorescence microscopy. *Proc. Natl. Acad. Sci. USA* **103**, 11440–11445.
- *Donnert, G., Keller, J., Wurm, C.A., Rizzoli, S.O., Westphal, V., Schönle, A., Jahn, R., Jakobs, S., Eggeling, C. and Hell, S.W. (2007). Two-colour far-field fluorescence nanoscopy. *Biophys. J.* **92**, L67–L69.
- Dyba, M. and Hell, S.W. (2002). Focal spots of size $1/23$ open up far-field fluorescence microscopy at 33 nm axial resolution. *Phys. Rev. Lett.* **88**, 163901.
- Dyba, M. and Hell, S.W. (2003). Photostability of a fluorescent marker under pulsed excited-state depletion through stimulated emission. *Appl. Opt.* **42**(25), 5123–5129.
- Dyba, M., Jakobs, S. and Hell, S.W. (2003). Immunofluorescence stimulated emission depletion microscopy. *Nature Biotechnol.* **21**(11), 1303–1304.
- Egner, A., Jakobs, S. and Hell, S.W. (2002). Fast 100-nm resolution 3D-microscope reveals structural plasticity of mitochondria in live yeast. *Proc. Natl. Acad. Sci. USA* **99**, 3370–3375.
- *Egner, A., Geisler, C., von Middendorff, C., Bock, H., Wenzel, D., Medda, R., Andresen, M., Stiel, A.C., Jakobs, S., Eggeling, C., Schönle, A. and Hell, S.W. (2007). Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophys. J.* **93**, 3285–3290.
- Eigen, M. and Rigler, R. (1994). Sorting single molecules: Applications to diagnostics and evolutionary biotechnology. *Proc. Natl. Acad. Sci. USA* **91**, 5740–5747.
- Elson, E.L. and Rigler, R. Eds. (2001). *Fluorescence Correlation Spectroscopy. Theory and Applications*. (Springer, Berlin).
- Failla, A.V., Spoeri, U., Albrecht, B., Kroll, A. and Cremer, C. (2002). Nanosizing of fluorescent objects by spatially modulated illumination microscopy. *Appl. Opt.* **41**(34), 7275–7283.
- Freimann, R., Pentz, S. and Hörler, H. (1997). Development of a standing-wave fluorescence microscope with high nodal plane flatness. *J. Microsc.* **187**(3), 193–200.
- Göpper-Mayer, M. (1931). Über Elementarakte mit zwei Quantensprüngen. *Ann. Phys. (Leipzig)* **9**, 273–295.
- Goodman, J.W. (1968). *Introduction to Fourier Optics*. (McGraw-Hill, New York).
- Gugel, H., Bewersdorf, J., Jakobs, S., Engelhardt, J., Storz, R. and Hell, S.W. (2004). Cooperative 4Pi excitation and detection yields 7-fold sharper optical sections in live cell microscopy. *Biophys. J.* **87**, 4146–4152.
- Gustafsson, M.G., Agard, D.A. and Sedat, J.W. (1996). 3D widefield microscopy with two objective lenses: Experimental verification of improved axial resolution. In: *Three-Dimensional Microscopy: Image Acquisition and Processing III*. Proc. SPIE.
- Gustafsson, M.G.L. (1999). Extended resolution fluorescence microscopy. *Curr. Opin. Struct. Biol.* **9**, 627–634.
- Gustafsson, M.G.L. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198**(2), 82–87.
- Gustafsson, M.G.L., Agard, D.A. and Sedat, J.W. (1995). Sevenfold improvement of axial resolution in 3D widefield microscopy using two objective lenses. *Proc. SPIE* **2412**, 147–156.
- Gustafsson, M.G.L., Agard, D.A. and Sedat, J.W. (1999). I5M: 3 widefield light microscopy with better than 100 nm axial resolution. *J. Microsc.* **195**, 10–16.
- Hänninen, P. (2002). Beyond the diffraction limit. *Nature* **419**, 802.

- Hänninen, P.E., Lehtelä, L. and Hell, S.W. (1996). Two- and multiphoton excitation of conjugate dyes with continuous wave lasers. *Opt. Commun.* **130**, 29–33.
- Hecht, B., Bielefeldt, H., Inouyne, Y., Pohl, D.W. and Novotny, L. (1997). Facts and artifacts in near-field optical microscopy. *J. Appl. Phys.* **81**, 1492–2498.
- Heintzmann, R. and Cremer, C. (1998). Laterally modulated excitation microscopy: Improvement of resolution by using a diffraction grating. *SPIE Proc.* **3568**, 185–195.
- Heintzmann, R., Jovin, T.M. and Cremer, C. (2002). Saturated patterned excitation microscopy—a concept for optical resolution improvement. *J. Opt. Soc. Am. A: Opt. Image Sci. Vision* **19**(8), 1599–1609.
- Hell, S. and Stelzer, E.H.K. (1992a). Properties of a 4Pi-confocal fluorescence microscope. *J. Opt. Soc. Am. A* **9**, 2159–2166.
- Hell, S.W. (1990). Double-scanning confocal microscope. European Patent.
- Hell, S.W. (1997). Increasing the resolution of far-field fluorescence light microscopy by point-spread-function engineering. In *Topics in Fluorescence Spectroscopy* (J.R. Lakowicz, Ed.), 361–422 (Plenum Press, New York).
- Hell, S.W. (2003). Toward fluorescence nanoscopy. *Nature Biotechnol.* **21**(11), 1347–1355.
- Hell, S.W. (2004). Strategy for far-field optical imaging and writing without diffraction limit. *Phys. Lett. A* **326**(1–2), 140–145.
- *Hell, S.W. (2007). Far-field optical nanoscopy. *Science* **316**, 1153–1158.
- Hell, S.W. and Kroug, M. (1995). Ground-state depletion fluorescence microscopy, a concept for breaking the diffraction resolution limit. *Appl. Phys. B* **60**, 495–497.
- Hell, S.W. and Nagorni, M. (1998). 4Pi-confocal microscopy with alternate interference. *Opt. Lett.* **23**(20), 1567–1569.
- Hell, S.W. and Stelzer, E.H.K. (1992b). Fundamental improvement of resolution with a 4Pi-confocal fluorescence microscope using two-photon excitation. *Opt. Commun.* **93**, 277–282.
- Hell, S.W. and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: Stimulated emission depletion microscopy. *Opt. Lett.* **19**(11), 780–782.
- Hell, S.W., Schrader, M., Hänninen, P.E. and Soini, E. (1995). Resolving fluorescence beads at 100–200 distance with a two-photon 4Pi-microscope working in the near infrared. *Opt. Commun.* **117**, 20–24.
- Hell, S.W., Jakobs, S. and Kastrop, L. (2003). Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Appl. Phys. A* **77**, 859–860.
- Hell, S.W., Schrader, M. and van der Voort, H.T.M. (1997). Far-field fluorescence microscopy with three-dimensional resolution in the 100 nm range. *J. Microsc.* **185**(1), 1–5.
- *Hess, S.T., Girirajan, T.P.K. and Mason, M.D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272.
- *Hofmann, M., Eggeling, C., Jacobs, S. and Hell, S.W. (2005). Breaking the diffraction barrier at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl. Acad. Sci. USA* **102**, 17565–17569.
- Holmes, T.J. (1988). Maximum-likelihood image restoration adapted for non-coherent optical imaging. *JOSA A* **5**(5), 666–673.
- Holmes, T.J., Bhattacharyya, S., Cooper, J.A., Hanzel, D., Krishnamurthi, V., Lin, W., Roysam, B., Szarowski, D.H. and Turner, J.N. (1995). Light microscopic images reconstruction by maximum likelihood deconvolution. In *Handbook of Biological Confocal Microscopy* (J. Pawley, Ed.) 389–400 (Plenum Press, New York).

- Irie, M., Fukaminato, T., Sasaki, T., Tamai, N. and Kawai, T. (2002). A digital fluorescent molecular photoswitch. *Nature* **420**(6917), 759–760.
- Kastrup, L. and Hell, S.W. (2004). Absolute optical cross section of individual fluorescent molecules. *Angew. Chem. Int. Ed.* **43**, 6646–6649.
- Kastrup, L., Blom, H., Eggeling, C. and Hell, S.W. (2005). Fluorescence fluctuation spectroscopy in subdiffraction focal volumes. *Phys. Rev. Lett.* **94**(17), 178104.
- Klar, T.A., Engel, E. and Hell, S.W. (2001). Breaking Abbe's diffraction resolution limit in fluorescence microscopy with stimulated emission depletion beams of various shapes. *Phys. Rev. E* **64**, 1–9.
- Klar, T.A., Jakobs, S., Dyba, M., Egner, A. and Hell, S.W. (2000). Fluorescence microscopy with diffraction resolution limit broken by stimulated emission. *Proc. Natl. Acad. Sci. USA* **97**, 8206–8210.
- Krishnamurthi, V., Bailey, B. and Lanni, F. (1996). Image processing in 3-D standing wave fluorescence microscopy. *Proc. SPIE* **2655**, 18–25.
- Lakowicz, J.R. (1983). *Principles of Fluorescence Spectroscopy*. (Plenum Press, New York).
- Lakowicz, J.R., Gryczynski, I., Malak, H. and Gryczynski, Z. (1996). Two-color two-photon excitation of fluorescence. *Photochem. Photobiol.* **64**, 632–635.
- *Lang, M., Müller, T., Engelhardt, J. and Hell, S.W. (2007). 4Pi microscopy of type A with 1-photon excitation in biological fluorescence imaging. *Opt. Express* **15**, 2459–2467.
- Lanni, F. (1986). *Applications of Fluorescence in the Biomedical Sciences*, 1st ed., D.L. Taylor, Ed. 520–521. (Liss, New York).
- Laurence, T.A. and Weiss, S. (2003). How to detect weak pairs. *Science* **299**(5607), 667–668.
- Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G. and Webb, W.W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686.
- Lukosz, W. (1966). Optical systems with resolving powers exceeding the classical limit. *J. Opt. Soc. Am.* **56**, 1463–1472.
- Lukyanov, K.A., Fradkov, A.F., Gurskaya, N.G., Matz, M.V., Labas, Y.A., Savitsky, A.P., Markelov, M.L., Zaraisky, A.G., Zhao, X., Fang, Y., Tan, W. and Lukyanov, S.A. (2000). Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *J. Biol. Chem.* **275**(34), 25879–25882.
- Magde, D., Elson, E.L. and Webb, W.W. (1972). Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* **29**(11), 705–708.
- Nagorni, M. and Hell, S.W. (1998). 4Pi-confocal microscopy provides three-dimensional images of the microtubule network with 100- to 150-nm resolution. *J. Struct. Biol.* **123**, 236–247.
- Nagorni, M. and Hell, S.W. (2001a). Coherent use of opposing lenses for axial resolution increase in fluorescence microscopy. I. Comparative study of concepts. *J. Opt. Soc. Am. A* **18**(1), 36–48.
- Nagorni, M. and Hell, S.W. (2001a). Coherent use of opposing lenses for axial resolution increase in fluorescence microscopy. II. Power and limitation of nonlinear image restoration. *J. Opt. Soc. Am. A* **18**(1), 48–54.
- Pawley, J., Ed. (1995). *Handbook of Biological Confocal Microscopy*. (Plenum Press, New York).
- Pohl, D.W. and Courjon, D. (1993). *Near Field Optics*. (Kluwer, Dordrecht).
- Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1993). *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge).

- Richards, B. and Wolf, E. (1959). Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system. *Proc. R. Soc. Lond. A* **253**, 358–379.
- *Rust, M.J., Bates, M. and Zhuang, X.-w. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–796.
- Schmidt, M., Nagorni, M. and Hell, S.W. (2000). Subresolution axial distance measurements in far-field fluorescence microscopy with precision of 1 nanometer. *Rev. Sci. Instrum.* **71**, 2742–2745.
- Schneider, B., Albrecht, B., Jaeckle, P., Neofotistos, D., Söding, S., Jäger, T. and Cremer, C. (2000). Nanolocalization measurements in spatially modulated illumination microscopy using two coherent illumination beams. *Proc. SPIE* **3921**, 321–330.
- Schönle, A. and Hell, S.W. (1999). Far-field fluorescence microscopy with repetitive excitation. *Eur. Phys. J. D* **6**, 283–290.
- Schönle, A. and Hell, S.W. (2002). Calculation of vectorial three-dimensional transfer functions in large-angle focusing systems. *J. Opt. Soc. Am. A* **19**(10), 2121–2126.
- Schönle, A., Hänninen, P.E. and Hell, S.W. (1999). Nonlinear fluorescence through intermolecular energy transfer and resolution increase in fluorescence microscopy. *Ann. Phys. (Leipzig)* **8**(2), 115–133.
- Schrader, M. and Hell, S.W. (1996). 4Pi-confocal images with axial superresolution. *J. Microsc.* **183**, 189–193.
- Schrader, M., Bahlmann, K., Giese, G. and Hell, S.W. (1998). 4Pi-confocal imaging in fixed biological specimens. *Biophys. J.* **75**, 1659–1668.
- Shen, Y.R. (1984). *The Principles of Nonlinear Optics*, 1st ed. (John Wiley, New York).
- Sheppard, C.J.R. and Kompfner, R. (1978). Resonant scanning optical microscope. *Appl. Optics* **17**, 2879–2882.
- Sheppard, C.J.R., Gu, M., Kawata, Y. and Kawata, S. (1993). Three-dimensional transfer functions for high-aperture systems. *J. Opt. Soc. Am. A* **11**(2), 593–596.
- Stephens, D.J. and Allen, V.J. (2003). Light microscopy techniques for live cell imaging. *Science* **300**, 82–91.
- Torraldo di Francia, G. (1952). Supergain antennas and optical resolving power. *Nuovo Cimento Suppl.* **9**, 426–435.
- Weiss, S. (2000). Shattering the diffraction limit of light: A revolution in fluorescence microscopy? *Proc. Natl. Acad. Sci. USA* **97**(16), 8747–8749.
- Westphal, V., Blanca, C.M., Dyba, M., Kastrup, L. and Hell, S.W. (2003). Laser-diode-stimulated emission depletion microscopy. *Appl. Phys. Lett.* **82**(18), 3125–3127.
- Westphal, V., Kastrup, L. and Hell, S.W. (2003). Lateral resolution of 28 nm ($\lambda/25$) in far-field fluorescence microscopy. *Appl. Phys. B* **77**(4), 377–380.
- Westphal, V.H. and Hell, S.W. (2005). Nanoscale resolution in the focal plane of an optical microscope. *Phys. Rev. Lett.* **94**(14), 143903.
- *Willig, K.I., Rizzoli, S.O., Westphal, V., Jahn, R. and Hell, S.W. (2006). STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis. *Nature* **440**, 953–959.
- Wilson, T. and Sheppard, C.J.R. (1984). *Theory and Practice of Scanning Optical Microscopy*. (Academic Press, New York).
- Xu, C., Zipfel, W., Shear, J.B., Williams, R.M. and Webb, W.W. (1996). Multiphoton fluorescence excitation: New spectral windows for biological nonlinear microscopy. *Proc. Natl. Acad. Sci. USA* **93**, 10763–10768.

*References added since the first printing.

Principles and Applications of Zone Plate X-Ray Microscopes

Malcolm Howells, Chris Jacobsen, and Tony Warwick

1 Introduction

1.1 Background

In the 1949 issue of *Scientific American*, an article by Stanford physicist Paul Kirkpatrick on “The X-ray Microscope” (Kirkpatrick, 1949) was described by the editors as follows:

“It would be a big improvement on microscopes using light or electrons, for X-rays combine short wavelengths, giving fine resolution, and penetration. The main problems standing in the way have now been solved.”

With the perspective of a half century, we might change “improvement on” to “complement to” and say that further problems were solved after 1949, but here in essence is the character of X-ray microscopes.

In this chapter, we outline some of the properties of X-ray microscope systems in operation today, and highlight some of their present applications. We will not discuss the history of X-ray microscopes prior to about 1975 but instead refer the reader to a series of conference proceedings known as “X-ray Optics and X-ray Microanalysis,” which began in 1956. Originally these had valuable material on X-ray microscopy but this diminished after about 1970. The first five were at Cambridge (1956) (Cosslett et al., 1957), Stockholm (1959) (Engström et al., 1960), Stanford (1962) (Pattee et al., 1963), Orsay (1965) (Castaing et al., 1966) and Tübingen (1968) (Molenstedt et al., 1969). We also recommend the historical perspectives by A. Baez (Baez, 1989, 1997) and the book by Cosslett and Nixon (1960). There is a recognisable thread of continuity between today’s status of the field and efforts that began slowly around 1975 (Niemann et al., 1976; Parsons, 1978; Kirz and Sayre, 1980c; Parsons, 1980) and blossomed with the availability of synchrotron light sources and nanofabrication technologies; this thread can be traced in part via the proceedings of another conference series that began in 1984 (Schmahl and Rudolph, 1984a) and has continued until today (Sayre et al., 1988; Michette et al., 1992; Aristov and Erko, 1994; Thieme et al., 1998b; Meyer-Ilse et al., 2000b; Susini et al., 2003). Zone-plate X-ray

microscopes now exist at roughly two dozen international synchrotron radiation research centers (see Table 13–3), and commercial lab-based instruments are also available. Three types are in especially widespread use. Transmission X-ray microscopes (TXMs) specialize in the rapid acquisition of 2D images using high flux sources, and in the collection of tilt sequences of projection images for 3D imaging by tomography. Scanning transmission X-ray microscopes (STXMs) specialize in the acquisition of reduced dose images and point spectra with high energy resolution for elemental and chemical state mapping, and require high source brightness. Scanning fluorescence X-ray microprobes (SFXMs) are similar to STXMs except that fluorescence X-rays are collected by energy-resolving detectors for trace element mapping. All three approaches are now working below 100 nm resolution, to the point of reaching 15 nm resolution in some demonstrations (Chao et al., 2005). While many of the new technical developments continue to be pursued by specialists in X-ray optics and microscopy, much of present-day activity comes from scientists in other fields of research who are using X-ray microscopes to address their particular questions. This chapter is mainly aimed at scientists from the latter group as well as those from the other communities represented in the content of this series of books.

1.2 X-Ray Interactions

A microscope requires illumination, magnification, and contrast. The characteristics of X-ray interactions with matter affect all three. In Figure 13–1, we show the cross-section (Hubbell et al., 1980) for photoelectric absorption, coherent (elastic or Thomson) scattering, and inco-

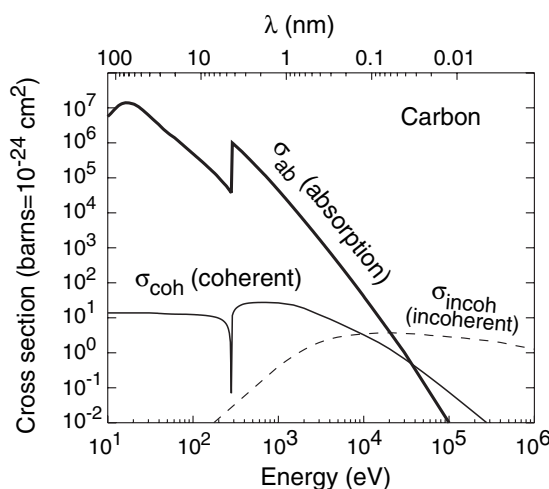


Figure 13–1. X-ray interaction cross sections in carbon. At energies below about 10 keV, absorption dominates so that images are free from the complications of multiple scattering. [Data from Henke et al. (1993) and the NIST XCOM database (Saloman and Hubbell, 1987).]

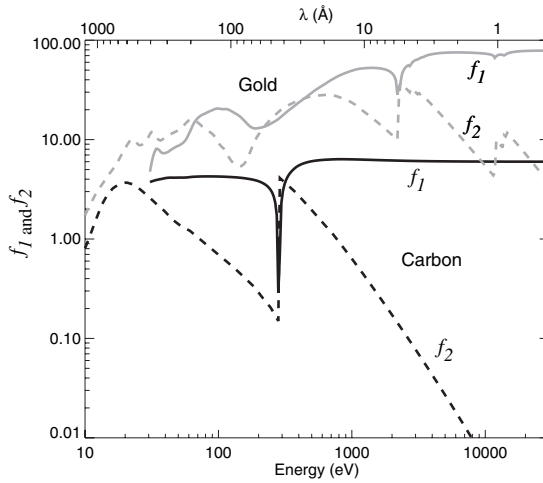


Figure 13–2. The frequency-dependent oscillator strength ($f_1 + if_2$) for carbon and gold. At X-ray absorption edges (such as 290 eV for carbon), f_2 has step increments, while f_1 undergoes anomalous dispersion resonances. (Reprinted from Henke et al., © 1993, with permission from Elsevier.)

herent (inelastic or Compton) scattering for carbon. Below 10 keV, absorption dominates, so multiple X-ray scattering is usually not of concern (an X-ray is much more likely to be absorbed following any scattering event than scattered again) nor is inelastic scattering. However, what is not evident in Figure 13–1 is the fact that the propagation of X-rays in materials can also include refractive effects, and in fact it was Einstein (1918) who first pointed out that the refractive index is slightly less than unity. The X-ray refractive index for a wave forward propagated as $\exp[-i(k\tilde{n}x - \omega\tau)]$ is often written as $\tilde{n} = 1 - \delta - i\beta$ where δ represents the phase-shifting part of the refractive index and β represents absorption according to a linear coefficient $\mu = 4\pi\beta/\lambda$ in the Lambert-Beer law $I = I_0 \exp[-\mu t]$. In an anomalous dispersion model, the refractive index terms can furthermore be written as $(\delta + i\beta) = \alpha\lambda^2(f_1 + if_2)$ with $\alpha = n_a r_e / 2\pi$, where $n_a = \rho N_A / A$ gives the number density of atoms, $r_e = 2.82 \times 10^{-15}$ m is the classical radius of the electron, and $(f_1 + if_2)$ represents the frequency-dependent oscillator strength of an atom. This oscillator strength ($f_1 + if_2$) has been tabulated with very good absolute accuracy for all elements by Henke, Gullikson et al. (Henke, 1993) over the energy range 10–30,000 eV (see Figure 13–2). In examining Figure 13–2, two features immediately jump out: f_1 is somewhat constant except near absorption edges, so the thickness $t_\pi = \lambda / 2\delta = 1 / (2\alpha\lambda f_1)$ needed to provide a phase advance $\exp[ik\delta t_\pi]$ equal to π increases as λ^{-1} , while, because f_2 scales as E^{-2} or λ^2 , the thickness $1/\mu = 1 / (4\pi\alpha\lambda f_2)$ that produces an attenuation of $1/e$ increases as E^3 or λ^{-3} . As a result, phase contrast becomes the dominant contrast mechanism as one goes to shorter wavelengths (Schmahl and Rudolph, 1987).

Much of modern X-ray microscopy centers on the exploitation of X-ray absorption edges. X-ray absorption edges arise when the X-ray photon reaches the threshold energy needed to completely remove an electron from an inner-shell orbital. The energy at which this occurs is approximately given by the Bohr model as $E_n = (13.6\text{eV})(Z - z_{\text{shield}})^2/n^2$, where Z is the atomic number, z_{shield} approximates the partial screening of the nucleus' charge by other inner-shell electrons ($z_{\text{shield}} \approx 1$ for K edges), and n is the principal quantum number ($n = 1$ for K edges, 2 for L edges, and so on). This produces the step-like rise in the cross-section for photoelectric absorption that can be seen in the plots of f_2 in Figure 13–2. If one takes one image I_1 at an energy E_1 just below an element's absorption edge where the incident flux is I_{01} , and a similar image I_2 at an energy just above an absorption edge, one can recover the mass per area m_x/A of the element x from (Engström, 1946)

$$\frac{m_x}{A} = \rho \frac{(E_1/E_2)^3 \ln(I_1/I_{01}) - \ln(I_2/I_{02})}{\mu_2 - \mu_1 (E_1/E_2)^3} \tag{1}$$

This approach works well for mass concentrations greater than about 1%. Another way in which X-ray absorption edges are exploited is by means of the “water window.” At X-ray energies between the carbon and oxygen absorption edges at 290 and 540 eV, respectively, organic materials show strong absorption contrast while water layers up to several μm thick are reasonably transmissive (Wolter, 1952); this is particularly valuable for imaging hydrated biological and environmental science specimens.

For those elements which have absorption edges below the energy of incident X-rays so that inner-shell ionization occurs, the aftermath of absorption involves the emission of either a fluorescent photon or an Auger electron of characteristic energy. The energy of these fluorescent photons, and the fluorescence yield (Krause, 1979) (the fraction of events which result in fluorescence rather than Auger electron emission), are both shown in Figure 13–3. At X-ray energies below 1 keV,

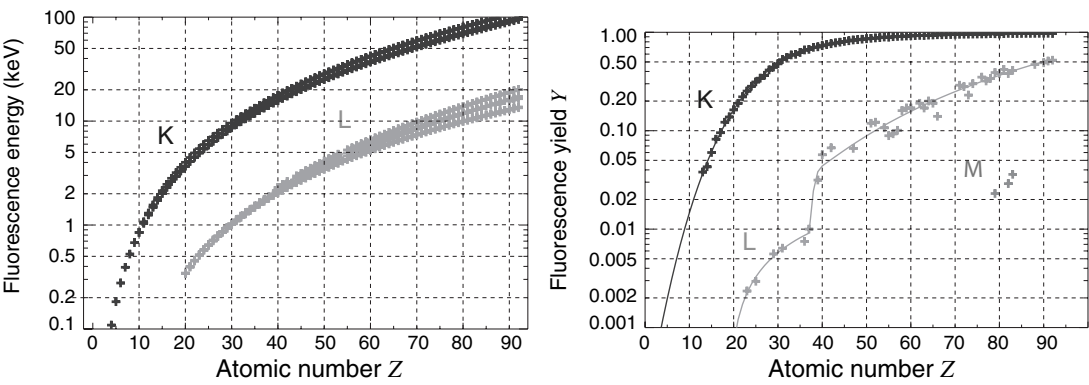


Figure 13–3. Energies (left) and fluorescence yields (right) for K and L edge emission. (Krause, 1979.)

Auger emission dominates, and scanning photoemission microscopes (SPEM) use electron spectrometers to exploit these electrons for surface studies (Ade et al., 1990a; Günther et al., 1997; Warwick et al., 1997; Ko et al., 1998). At higher energies, the fluorescence signal dominates and detection of these characteristic X-rays provides information on the concentration of various elements in the specimen. Most scanning fluorescence X-ray microprobes (SFXM) (Horowitz and Howell, 1972; Sparks, 1980) use energy dispersive detectors where the number of electron-hole pairs created by each fluorescent photon is used to measure its energy, though crystal-based wavelength dispersive spectrometers can also be used. Exact quantitation of the elemental concentration requires accurate knowledge of a number of factors, including the solid angle acceptance of the detector and its quantum efficiency, the degree to which fluorescent photons are reabsorbed in the specimen, and other factors, so that in most cases comparison is made with standards with known elemental concentration and matrix concentration similar to that of the specimen under study. When compared with electron microprobes, X-ray microprobes do not suffer from expansion of the probe beam due to electron scattering, or a large continuum background, so that the sensitivity to trace elements is often in the 100 parts per billion range.

Because X-ray interactions are well understood and do not involve significant complications due to multiple scattering at energies below about 10 keV, reliable predictions of image contrast can be made. If we have a normalized signal I_f from a feature-containing pixel and I_b from a background region, the signal to noise ratio obtained with N illuminating photons is (Glaeser, 1971; Sayre et al., 1977a)

$$\text{SNR} = \frac{\text{Signal}}{\text{Noise}} = \sqrt{N} \frac{I_f - I_b}{\sqrt{I_f + I_b}} = \sqrt{N} \Theta \quad (2)$$

where we have used the Gaussian approximation to Poisson statistics (which is quite good for NI greater than about 10) and the assumption that there are no other noise sources with significant fluctuations. The contrast parameter Θ is different from the usual definition of contrast due to the square root in the denominator. With this definition, the number of photons required to see a feature with a desired signal to noise ratio SNR is given by $N = (\text{SNR})^2 / \Theta^2$, and a common choice for the minimum detectable signal to noise ratio is the Rose criterion of $\text{SNR} = 5$ (Rose, 1946). Using this approach, Sayre *et al.* showed that “water window” X-ray microscopes are able to image organic specimens in micrometer-thick water layers with greatly reduced radiation dose compared to electron microscopy (Sayre, 1977b; Sayre, 1977a). This conclusion remains true even when modern energy-filtered electron microscopes are considered (Grimm et al., 1998; Jacobsen et al., 1998) (see Figure 13–4). Other investigators have extended the same approach to include the effects of phase contrast (Rudolph et al., 1990; Götz, 1992) (see Figure 13–5) and the reduction of modulation transfer at high spatial frequencies (Schneider, 1998), while Kirz *et al.* have used this approach to compare elemental mapping using both differential absorption and X-ray fluorescence (Kirz et al., 1978, 1980a, 1980b).

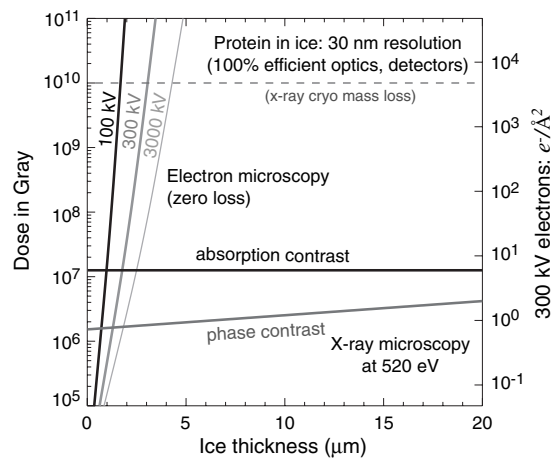


Figure 13–4. Estimated radiation dose required for imaging 30-nm-protein features as a function of ice thickness for 200 keV electron microscopes and 520 eV soft X-ray microscopes. (Reprinted from Jacobsen et al., 1998, with permission from Springer Science+Business Media.)

1.3 Focusing Optics

Microscopes require focusing optics, or some other means to provide a magnified view of the object. X-rays reflect well from single refractive interfaces only at grazing angles of incidence less than a critical angle of $\theta_c \approx \sqrt{2\delta}$ which is typically in the range of 1–5° for soft X-rays. (Once a particular angle has been selected, this same relationship gives a critical energy above which the reflectivity becomes low; this can be used to low-pass-filter the energy spectrum from a radiation source). While a number of labs have explored the use of axially symmetric paraboloid or hyperboloid optics (Wolter, 1952; Aoki, 1994), most present efforts center on the use of two orthogonal cylindrical grazing mirrors in the Kirkpatrick-Baez geometry (Kirkpatrick and Baez, 1948). Advantageous characteristics of these optics include their relatively long focal

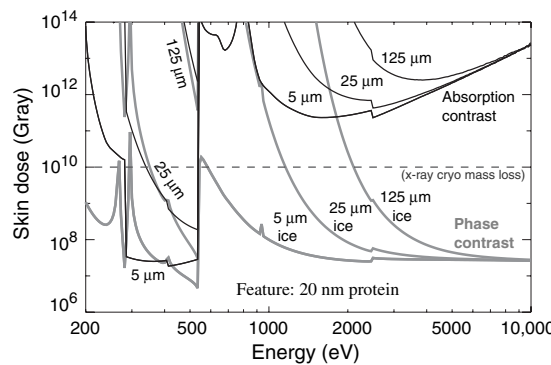


Figure 13–5. Estimated radiation dose required for imaging 20-nm-thick protein features in various ice thicknesses as a function of X-ray energy; 100% efficient optics and detectors are assumed.

length (several centimeters is typical) and their low chromaticity, so that the incident beam energy can be tuned for spectroscopy without any need to adjust the focus on the specimen. Optics of this sort have recently achieved better than 100 nm resolution probe sizes using 12 keV X-rays (Hignette et al., 2003; Yamamura et al., 2003; Mimura et al., 2004), although the profile of the focus always has some degree of “tail” outside of the geometrical image of the source due to scattering from the residual surface roughness of even the best available mirrors. Synthetic multilayer X-ray mirrors (Spiller, 1972; Barbee et al., 1981) can increase the incidence angle well beyond θ_c for narrow-bandwidth radiation, and can achieve good reflection efficiencies for normal incidence reflection at photon energies below about 200 eV. This approach has seen rapid improvements due to the development of EUV projection lithography at 95 eV. However, notwithstanding recent progress in mirror manufacture, it is important to recognize that even a perfectly made Kirkpatrick-Baez mirror system still suffers from aberrations, especially obliquity of field, which severely restrict its field of view and therefore its performance as a microscope. On the other hand, it is still well-able to focus points on or near the optical axis, which has led to a resurgence in its popularity for microprobes and relay mirrors that are imaging small sources such as synchrotrons.

When Röntgen discovered X-rays, he immediately tried to focus them using refractive lenses but without success. The reason for this is now well known: the focal length for a planoconvex lens with radius of curvature R_c is given by $f_R = -R_c/\delta$, so that at 10 keV a glass lens with $R_c = 1$ cm would have a focal length of about 2 km. This does not preclude the usefulness of refractive optics, however; a series of lenses with small R_c can be placed together to produce a significant net focusing effect. One simple way to achieve this result in 1D is to drill a series of holes in a solid block (Snigirev et al., 1996), and more recent work using parabolic optics has demonstrated a resolution of about 100 nm for hard X-ray imaging (Lengeler et al., 2002) with theoretical promise for sub-10 nm resolution imaging (Schroer and Lengeler, 2005). Because the ratio of phase shift to absorption increases with increasing X-ray energy, these optics work primarily at energies above about 5 keV, and at higher energies one will ultimately need to consider the contributions of inelastic scattering to the image due to the overall thickness of the optic. Still, this approach is of interest especially since these optics can be easily water cooled for high power applications.

The third way to focus X-rays is to use diffraction. While bent crystals can provide focused beams of Bragg or Laue diffracted X-rays, most work in X-ray microscopy centers on the use of microfabricated diffractive optics in the form of Fresnel zone plates. Efforts in X-ray microscopy using zone plate optics date back nearly a half century (Baez, 1960, 1961), and X-ray Fresnel zone plates are now benefiting from a high degree of development. Apart from detailed literature that we will cite in Section 2, general reviews can be found in the books by (Michette, 1986; Attwood, 1999). Due to their popularity as high resolution optics for X-ray microscopy, the properties of Fresnel zone plates are described in some detail below.

1.4 Overview and Recent Trends

We now review some of the general trends of X-ray microscopy technology and how they might affect the planning of a new X-ray microscopy program today. In this section, the number of relevant references is essentially unlimited so we cite instead the subsections of this article where many of the references can be found.

Modern X-ray microscopy was based on the twin ideas of the water window (Section 1.1) and microfabricated zone-plate lenses (Section 2). These led initially to life-science experiments consisting of 2D imaging of wet samples in room-temperature air with “naturalness” as the unique capability of the technique. Early versions of both the TXM (Section 3.1.1) and STXM (Section 3.1.3) were capable of such imaging and this style of X-ray microscopy has proved to be good fit to many of the needs of the polymer (Section 4.3) and environmental research communities (Section 4.2) among others. However, the room-temperature approach could not be easily adapted to 3D imaging of radiation-sensitive samples and this was something for which there was, and still is, considerable demand. On the other hand the STXM, even in its earliest realizations, was ready to begin the development of spectromicroscopy (Sections 3.5 and 4). This development has continued over the last two decades or so and spectromicroscopy is now quite a mature field that spans a wide range of application areas.

The step which has brought 3D imaging of biological samples within reach has been the recent move toward X-ray microscopy of hydrated biological samples frozen to cryogenic temperatures. For this type of sample, X-ray microscopy fills an important gap in the range of resolution values and sample thicknesses which can be covered by other methods. The use of cryogenic temperatures is the key step in providing enough radiation-damage protection (see Figure 13–5) to enable tomographic (three-dimensional) imaging (see Section 3.4). However, 3D cryomicroscopy requires significant technical additions to the X-ray microscope including either the use of a vacuum sample chamber or a gas-stream approach to keeping the sample cold plus the mechanical devices required for recording a tilt series.

The interest in 3D imaging is also high in materials science and engineering where there is a similar gap in the coverage by other methods and where generally the samples have higher atomic number and have nonaqueous background materials (microcircuits for example). For these samples the water window has no advantage and X-ray microscopy at higher X-ray energies allows examination of bigger samples and has in fact been going on for some time. Even in biology there are factors pushing in the same direction. Ice is equally transparent at 1.5 keV as in the water window. Moreover, the 1.5–3.0 keV region gives less absorption and more phase contrast so there is only a moderate increase of the required radiation dose compared to the water window (see Figure 13–5). In this energy range zone plates also have longer focal lengths (important for sample tilting) and greater depth of focus (important for 3D reconstruction (Section 3.4.2)). Overall, multi-keV 3D X-ray microscopy has great promise but it poses some-

what different challenges with respect to the resolution/efficiency capabilities of the zone plates and the provision of a suitable condenser.

When possible we have cited publications in widely available journals rather than conference proceedings. However, good snapshots of the field are provided at three-year intervals by the series of conference proceedings dating back to 1984 (Section 1.1). As of the time of writing of this chapter the most recent one (of which the proceedings (Kagoshima et al., 2006) have not yet been published) was in Himeji, Japan, in 2005. We have in some cases cited papers in the proceedings of this conference because they have not appeared yet elsewhere. This is especially true for papers on multi-keV X-ray microscopy which was a growth subject at Himeji. Reports on the high-aspect-ratio zone plates needed for multi-keV X-ray microscopy (Section 2.4.6) were particularly promising. Zone plates of 50 nm outer zone width for 3–10 keV and 30 nm for 1–3 keV were reported with efficiencies of at least 10% (Section 2.4.6). Several of the former are operational in synchrotrons. Equally significantly, images were shown using a 50 nm zone plate in 3rd order to get sub-30-nm resolution (Section 2.4.6).

The resolution values of the best water-window and multi-keV three-dimensional images are both currently around 60 nm. The underlying reasons for this value probably include the depth of focus effect (3.4b) for water-window images and the zone plates for multi-keV ones. If this is true, then the above reports suggest that the tools are now in place for improvements to the 60 nm limit.

On condensers the news was also good. It was announced publicly for the first time that single-reflection monocapillary mirrors have been in use for some time as condensers for multi-keV TXMs and that extensive 3D imaging has already been done using them. The condenser system (Section 2.4.3) has traditionally been the Achilles heel of the TXM and this new generation of condensers is delivering an elegant solution which is described in more detail in Section 2.4.4. Apart from removing various limitations of the condenser-zone-plate, the most important innovation is energy-tunability. This combined with its multiplexed data collection could make the TXM competitive in the spectromicroscopy arena for specimens that are tolerant of its higher dose.

The TXM and STXM have always been seen as complementary devices; roughly equally popular (Table 13–3) and with important advantages on both sides. We do not believe that that general perception is likely to change in the near future. The STXM will always have the advantage in trace element mapping, the ability to instantly switch from high to low magnification, the ability to hold constant magnification even when the X-ray energy is changing, the ability to image at close to the theoretical minimum dose and so forth. On the other hand, the TXM is going through a period of technical enhancement. It has been moving into the multi-keV region and into cryomicroscopy which together with its traditional main asset, multiplexed data collection, is strengthening its capability in tomography which appears to us to be its natural home.

2 Fresnel Zone Plates

2.1 Introduction

A Fresnel zone plate is a circular diffraction grating that can be made to focus light waves in the manner of a lens. It consists of a series of concentric, usually metal, rings alternating with circular slots. Typically the rings are about equal in width to the slots and are fabricated on a thin membrane. The design is based on the idea, that by blocking, say, the even-numbered Fresnel half-period zones (Born and Wolf, 1999), the wavelets from the remaining (odd-numbered) zones will add constructively. To see this quantitatively, consider plane-wave illumination of a zone plate with n zones with radius r_n , outer zone width Δr_n , and a focal length f at wavelength λ (Figure 13–6). To get a first order *diffracted* beam in which the signals from all the open zones reinforce at the focus, we need a path difference λ between neighboring open zones. In other words the optical path $\sqrt{r_n^2 + f^2} - n\lambda/2$ should equal f . Expanding the square root and neglecting terms above fourth order, we obtain

$$\frac{n\lambda}{2} = \frac{r_n^2}{2f} - \frac{r_n^4}{8f^3} + \dots \quad (3)$$

Evidently the focusing condition is

$$r_n^2 = n\lambda f \quad (4)$$

to second order and

$$r_n^2 = n\lambda f + \frac{n^2\lambda^2}{4} \quad (5)$$

to fourth order. In view of Eq. (7), we can neglect the fourth order (spherical-aberration) term of Eq. (3) if the numerical aperture (NA) $\ll 1$ which is often the case for X-ray zone plates. If the fourth-order term is significant, then the zone plate can be made according to (5) and will be corrected for spherical aberration but the correction will only apply near the chosen wavelength and conjugate distances (∞ and f). If it can be neglected, we have $r_n = \sqrt{n\lambda f}$, and this defines a zone plate that will

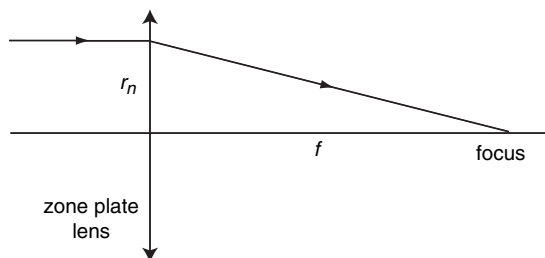
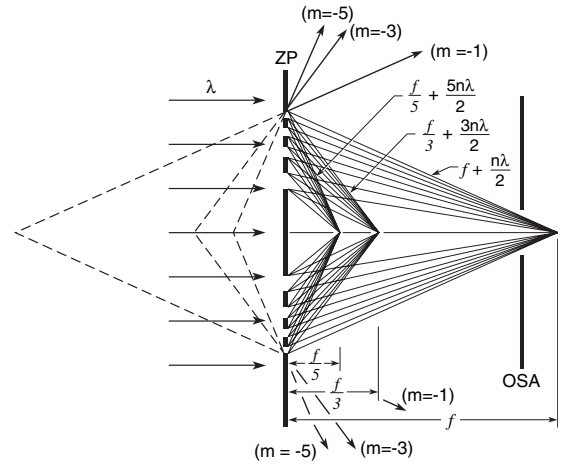


Figure 13–6. Geometry to calculate the radius of the n th half-period zone of a zone plate illuminated by parallel light. The path from the n th zone to the focus must be equal to f plus $n/2$ waves.

Figure 13–7. Diagram showing orders number $-5, -3, -1, 1, 3, 5$ of a zone plate. Compared to the incident beam, the negative orders diverge, the positive orders converge and the zero order (omitted for clarity) has the same shape. For plane-wave illumination, the virtual foci of the negative order beams and the real foci of the positive order beams are at distances $|f|/m$ from the zone plate where $|f|$ is the focal distance of first order. (From Attwood, © 1999, with permission of Cambridge University Press.)



focus well for a range of wavelengths, although the focal length will vary inversely with wavelength. Thus the chromatic aberration of a zone-plate lens is much larger than that of a refractive lens and, to get a good focus, the zone plate needs to be illuminated by monochromatic light. The required degree of monochromaticity for achievement of the diffraction-limited resolution is roughly given by $\Delta\lambda/\lambda \leq 1/n$ (Thieme, 1988).

Some useful quantities follow from the fundamental zone-plate Eq. (4). First we can take the difference between the n th and $(n - 1)$ th equations to get the outer zone width Δr_n

$$\Delta r_n = \frac{\lambda f}{2r_n}. \quad (6)$$

This allows the conclusion that all of the zones have equal area and also gives us the numerical aperture

$$NA \equiv \frac{r_n}{f} = \frac{\lambda}{2\Delta r_n}, \quad (7)$$

and thence the Rayleigh resolution

$$\delta_{\text{Rayleigh}} = \frac{0.61\lambda}{NA} = 1.22\Delta r_n. \quad (8)$$

Thus we see that a given zone plate can be specified by its r_n and Δr_n from which the resolution (which is independent of wavelength) and the focal length and numerical aperture at any given wavelength, follow from Eq. (8), Eq. (6) and Eq. (7) respectively. So far we have been discussing the first-order focus but in general, beams of all integral orders may be produced. Thus there is a zero-order (unfocused) beam, a series of positive-order converging beams with focal distance f/m and a series of negative-order diverging beams with focal distance $-f/m$ (Figure 13–7). In m th order the numerical aperture is m times larger

and hence the resolution is m times smaller (better) than in first order. As we will see, if the open and opaque zones are of equal width, then the even orders are missing.

2.2 Zone Plate Image Quality

2.2.1 Optical Path Function Analysis

We consider the imaging of a general point A (x, y, z) by a planar zone plate lying in the $y - z$ plane (Figure 13–8) (Kamiya, 1963). We will use the method of the optical path function so we start by calculating the optical path from A to a general point B (x', y', z') that we will later identify as the Gaussian image point. Without loss of generality we set $y = y' = 0$. We calculate the path APB where P ($0, w, l$) is a general point in the zone plate. The expression for the optical path will be a power series in the aperture coordinates, w and l , and the field angle, z/x and each term in the series will represent a specific aberration. Evidently

$$AP = \sqrt{x^2 + w^2 + (z - l)^2} \quad (9)$$

so expanding the square root and keeping terms up to fourth order, we have

$$AP = x \left\{ 1 + \frac{1}{2} \frac{w^2 + l^2}{x^2} + \frac{1}{2} \frac{z^2}{x^2} - \frac{2zl}{2x^2} - \frac{1}{8} \frac{1}{x^4} [(w^2 + l^2)^2 + z^4 + 4z^2l^2 - 4z^3l + 2(w^2 + l^2)z^2 - 2(w^2 + l^2)2zl] + \dots \right\} \quad (10)$$

There is an identical series for PB except that, for PB, the x , y and z are replaced by x' , y' , and z' (Figure 13–8). We are now in a position to write down the optical path function, F . Before doing so we drop terms

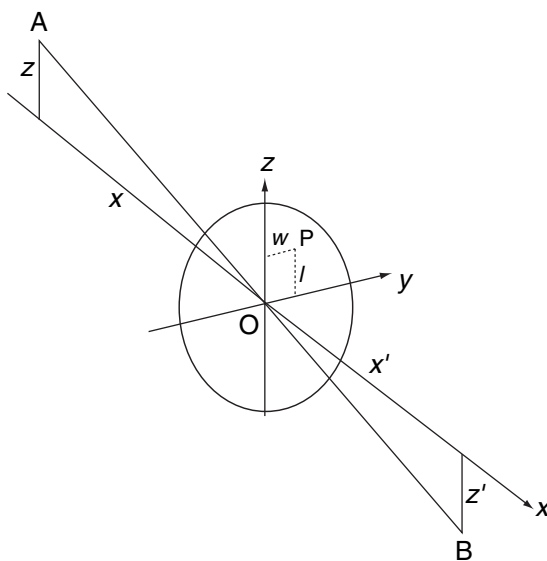


Figure 13–8. Notation for the optical-path-function analysis of a zone plate.

which do not depend on w , l or z/x because they do not represent aberrations, and introduce the term $-nm\lambda/2$ as we did to get Eq. (1). We choose to analyze the case of a parabolic zone plate (that is one built according to Eq. (4)) so initially, we use Eq. (4) to write the optical path function up to third order only, as follows.

$$\begin{aligned} F &= AP + PB - \frac{nm\lambda}{2} = AP + PB - \left| \frac{w^2 + l^2}{2} \frac{1}{f/m} \right| \\ &= \frac{w^2 + l^2}{2} \left[\frac{1}{x} + \frac{1}{x'} - \frac{1}{f/m} \right] - l \left[\frac{z}{x} + \frac{z'}{x'} \right] + \dots \end{aligned} \quad (11)$$

Specializing to the case when B is the Gaussian image point, the first (defocus) term vanishes and by considering the ray AOB we obtain

$$z/x = -z'/x' \quad (12)$$

so that the second term also vanishes. This leaves only the five fourth-order terms

$$\begin{aligned} F &= -\frac{(w^2 + l^2)^2}{8} \left(\frac{1}{x^3} + \frac{1}{x'^3} \right) - \frac{l^2 z^2}{2x^2} \frac{1}{f/m} + \frac{l}{2} \left(\frac{z^3}{x^3} + \frac{z'^3}{x'^3} \right) - \\ &\quad \frac{(w^2 + l^2)}{4} \frac{z^2}{x^2} \frac{1}{f/m} + \frac{(w^2 + l^2)}{2} \frac{lz}{x} \left(\frac{1}{x^2} + \frac{1}{x'^2} \right) + \dots \end{aligned} \quad (13)$$

These are the five Seidel aberrations; spherical aberration, astigmatism, distortion, field curvature and coma respectively. Because of Eq. (12), the distortion term vanishes identically which is a useful property of zone-plate lenses and we can therefore turn our attention to the remaining aberrations.

2.2.2 Ray Aberrations

We need to know the ray pattern delivered by the zone plate for a given point object. That is, we want the ray aberrations $\Delta y'$ and $\Delta x'$ relative to the Gaussian image point (coordinates identified by subscript zero). For a normal-incidence optic these are given by the following expression (Born and Wolf, 1999)

$$\Delta y' = x'_0 \frac{\partial F}{\partial w}, \quad \Delta z' = x'_0 \frac{\partial F}{\partial l} \quad (14)$$

We now apply this to the Seidel aberrations individually.

2.2.3 Spherical Aberration

We can rearrange the spherical aberration term using the magnification $M = x'_0/x$

$$F_{\text{sp ab}} = -\frac{(w^2 + l^2)^2}{8f^3} \frac{M^3 + 1}{(M + 1)^3} \quad (15)$$

The last term of this expression, which we will denote by Θ , approaches unity in the cases of interest to us namely M large (a microscope) or M small (a microprobe). If we consider the case that Θ does approach unity, then Eq. (15) reduces to the fourth-order term of Eq. (3). This is expected because we are reverting to the conjugates used to derive (3)

(∞ and f) and the parabolic zone plate we are analyzing is not corrected for spherical aberration. Treating the case of a microscope ($x \approx f$) and using the notation $r = \sqrt{w^2 + l^2}$, we now substitute from (15) into (14) to obtain the ray aberrations

$$\begin{aligned}\Delta y' &= x'_0 \frac{1}{8f^3} 2(w^2 + l^2) 2w = \frac{M}{2} \left(\frac{r}{f} \right)^2 w \\ \Delta z' &= x'_0 \frac{1}{8f^3} 2(w^2 + l^2) 2l = \frac{M}{2} \left(\frac{r}{f} \right)^2 l\end{aligned}\quad (16)$$

Thus we see that the image-plane figure produced by a point object via the rays passing through the rim of the lens ($r = r_n$), is a circle of diameter D_{SA} where

$$D_{SA} = M(NA)^2 r_n \quad (17)$$

This is produced irrespective of the position of the object point. The presence of uncorrected spherical aberration means there will be an optimum value of the NA , that is the largest NA for which the resolution is still diffraction-limited. This can be estimated (Michette, 1986) by requiring that the path error be less than the Rayleigh quarter-wave limit

$$NA_{opt} = \sqrt[4]{\frac{2\lambda}{\Theta f}} \quad (18)$$

In practice the spherical aberration of a parabolic zone plate is often not negligible in the soft X-ray region (see Figure 13–9 for example) and therefore soft X-ray zone plates are usually made according to Eq. (5), which means that the spherical aberration is corrected for the chosen wavelength. In these cases we may ask what happens when such a zone plate is used with a wavelength other than the correction wavelength. We therefore consider a zone plate corrected for a given wavelength in first positive order with M very small as shown in Figure 13–6 and we assume that the correction is also small ($NA \ll 1$). Now using a subscript zero to represent the properties of the corrected zone plate and subscript one to represent the properties of the same zone plate operating at another wavelength, we can apply Eq. (5) for the n th and $(n - 1)$ th zone to show that $\Delta r_n \equiv (\lambda_0 f_0 + n\lambda_0^2/2)/2r_n$. From that we can use the grating equation to get the ray deviation angle (ϵ_0) at the zone plate and thus the ray displacement ($\epsilon_0 f_0$) at the detection plane when the zone plate images a distant axial point (M small):

$$\epsilon_0 f_0 = r_n \left\{ 1 - \frac{n\lambda_0}{2f_0} \right\} = r_n - \frac{r_n}{2} (NA_0)^2, \quad (19)$$

where Eq. (4) and Eq. (7) have been used. For imaging an axial point with M large the ray displacement would be approximately $M\epsilon_0 f_0$. The first term of these expressions, r_n or Mr_n , is the ray displacement needed for correct imaging of an axial point and the second term is an aberration equal to minus the radius of the spherical aberration disk of a parabolic zone plate (compare Eq. (19) with Eq. (17)). This results in perfect cancellation of the spherical aberration as intended.

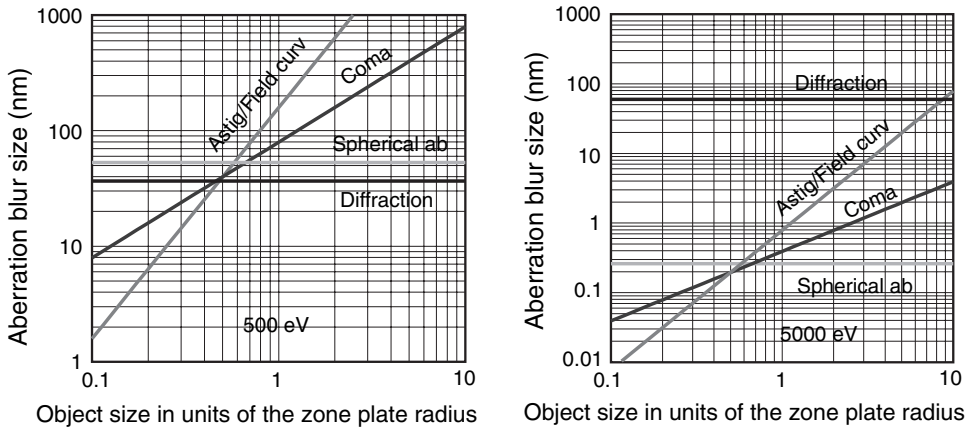


Figure 13–9. Size of the aberrations of a soft X-ray (500 eV; left) and a hard X-ray (5000 eV; right) zone plate as a function of object size. The soft X-ray zone plate has an outer zone width of 30 nm, diameter of 62 μm , and a focal length of 0.75 mm at 500 eV. The hard X-ray one has an outer zone width of 60 nm, diameter 124 μm and focal length 30 mm at 5000 eV. The graphs illustrate the general trend that hard X-ray zone plates have lower aberrations on account of their lower NA. In the example the hard X-ray zone plate has only negligible aberrations up to a sample diameter of about ten radii, which is far beyond the sample size allowed by penetration requirements. On the other hand the resolution of the soft X-ray zone plate is degraded by the field-angle-dependent aberrations for objects of diameter more than about half a radius. This is still good for many experiments but, as the diagram shows, spherical aberration is not negligible for the parabolic soft-X-ray zone plate. Therefore, as explained in Section 2.2.3, zone plates designed for soft-X-ray applications are normally made with built-in spherical-aberration correction according to Eq. (5).

When the same zone plate is used at another wavelength λ_1 with allowance for the change of focal length, Δr_n remains the same so Eq. (19) becomes

$$\epsilon_1 f_1 = \frac{f_1 \lambda_1}{f_0 \lambda_0} r_n \left\{ 1 - \frac{n \lambda_0}{2 f_0} \right\} = r_n - \frac{r_n}{2} (NA_0)^2 \quad (20)$$

In other words the amount of correction has not changed, whereas for exact correction it should now be $r_n (NA_1)^2 / 2$. Thus the residual error is an aberration disk of diameter $|r_n [(NA_1)^2 - (NA_0)^2]|$ or a new disk of diameter $|(\lambda_1^2 - \lambda_0^2) / \lambda_0^2|$ times the diameter of the original uncorrected disk. As an example, if the new wavelength differs by 10% from the wavelength of correction, then the aberration disk will be reduced to about 20% of its uncorrected size. This would be enough to make the aberration negligible in the soft X-ray example shown in Figure 13–9. Evidently this general conclusion applies equally to M very small or very large.

2.2.4 Astigmatism and Field Curvature

Taking the astigmatism and field curvature terms together and again calculating the ray aberrations by substituting the path function terms into Eq.(14), we obtain

$$\begin{aligned}\Delta y' &= -\frac{1}{2}x_0'w\frac{z^2}{x^2}\frac{1}{f} \\ \Delta z' &= -\frac{3}{2}x_0'l\frac{z^2}{x^2}\frac{1}{f}\end{aligned}\quad (21)$$

Squaring and adding we find that the marginal rays trace out an ellipse of major axis $3k$ and minor axis k

$$\frac{\Delta y'^2}{(k/2)^2} + \frac{\Delta z'^2}{(3k/2)^2} = 1 \quad \text{where} \quad k = M(NA)^2 r_n \bar{z}^2. \quad (22)$$

The parameter $\bar{z} = z/r_n$ expresses the position of the object point in units of the zone plate radius. Therefore, unlike spherical aberration, the size of this aberration does depend on the position of the object point, and therefore will limit the field of view of the microscope.

2.2.5 Coma

In this case, the above procedure leads to

$$\Delta y' = -2wlQ, \quad \Delta z' = -(w^2 + 3l^2)Q \quad \text{where} \quad Q = \frac{1}{2} \frac{z}{x} \frac{M-1}{f} \quad (23)$$

By expressing w and l in polar coordinates one can show that each circle of radius r in the lens produces an aberration figure

$$(\Delta z' - 2K)^2 + (\Delta y')^2 = K^2 \quad \text{where} \quad K = r^2 Q. \quad (24)$$

This is a series of circles of radius K , each shifted by $2K$ from the origin which is the usual “comet-shaped” figure associated with coma. The outer boundary of the figure has a length $3K_{\max}$ and the largest circle has a diameter $2K_{\max}$ where (assuming M is large so that $M-1 \approx M$)

$$2K_{\max} = M(NA)^2 r_n \bar{z}. \quad (25)$$

2.2.6 Relative Size of the Aberrations

The relative size of these aberration figures can be obtained from Eq. (17), Eq. (22) and Eq. (25) respectively. The diameter of the spherical aberration circle, the diameter of the largest coma circle and the major axis of the astigmatism ellipse are in the ratio $1:\bar{z}:3\bar{z}^2$ (Mchette, 1986). For points close to the axis ($\bar{z} \ll 1$) the field-angle-dependent aberrations coma and astigmatism/field curvature will be negligible and the resolution will be determined by diffraction if the numerical aperture is below NA_{opt} or by spherical aberration if it is above. On the other hand for points further from the axis, coma and astigmatism/field curvature become more significant and for points more than about a zone plate radius away from the axis ($\bar{z} \geq 1$) astigmatism/field curvature dominates. The general behavior of these aberrations is shown in Table 13–1, while Figure 13–9 shows examples of parabolic zone plates working at 0.5 keV and 5 keV. However, note that the plots in the figure show the size of the outer boundary of the aberration figures which is a conservative estimate of their contribution to the resolution because the light is somewhat concentrated near the center of the figure. For example, we can deduce from Eq. (17) that half of the light is concentrated in a circle of diameter about one third of D_{SA} .

Table 13–1. Seidel aberrations of a planar zone plate.

Name	Spherical aberration	Astigmatism/field curvature	Distortion	Coma
$w, l, z/x$ term	$(w^2 + l^2)^2$	$(w^2 + 3l^2)(z/x)^2$	$l\{(z/x)^2 + (z'/x')^2\}$	$(w^2 + l^2)l(z/x)$
Aberration figure boundary	circle	ellipse	vanishes identically	two lines through O at $\pm 30^\circ$ to Oz touching a family of circles
y parameter	circle diameter	major axis	na	diameter of the largest circle
Value y parameter	$M(NA)^2 r_n$	$3M(NA)^2 r_n \bar{z}^2$	na	$M(NA)^2 r_n \bar{z}$
z parameter	circle diameter	minor axis	na	distance from the origin to the far side of the largest circle
Value z parameter	same	$M(NA)^2 r_n \bar{z}^2$	na	$\frac{3}{2}M(NA)^2 r_n \bar{z}$

2.3 Zone Plate Efficiency

2.3.1 Idealized Structures

One can see from Eq. (4) that an ideal Fresnel zone plate is periodic in r^2 space with period λf , which means that its amplitude transparency function T_{zp} (shown in Figure 13–10) can be written as the Fourier series

$$T_{zp} = \sum_{-\infty}^{+\infty} a_m \exp\left(\frac{im\pi r^2}{\lambda f}\right) \quad \text{where} \quad a_m = q(-1)^m \text{sinc}(mq) \quad (26)$$

The sinc function is defined by $\text{sinc}(x) = \sin(\pi x)/(\pi x)$ and q is the fraction of each *period* that is opaque. For a classical zone plate, $q = 0.5$, so in that case (26) gives the power in the $\pm m$ th harmonic as $|a_m|^2 = 1/(m\pi)^2$ for m odd, zero for m even and 0.25 for $m = 0$. These are therefore the intensity efficiencies of the classical zone plate in those orders. The

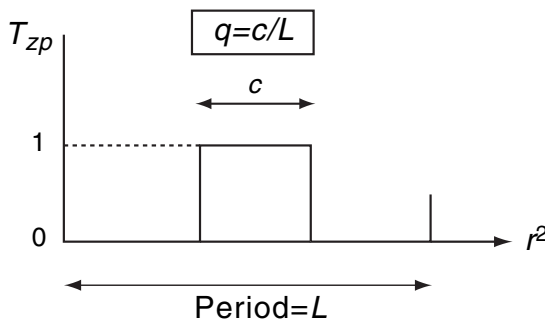


Figure 13–10. The transparency function of one period of a zone plate plotted in r^2 space. The fraction that has transparency unity is $q = c/L$.

main point here is that this type of zone plate has a maximum efficiency in the first order focus of about 10%. By taking the derivative of $|a_m|^2$ we can show that the classical zone plate ($q = 0.5$) is, in fact, the optimum choice of q for first-order efficiency. Similarly, the optimum choice for second order efficiency is $q = 0.25$ or 0.75 which may have some practical significance, as discussed by Simpson and Michette (1984). By applying Parseval's theorem to the series in (26) with $q = 0.5$, one can determine the disposition of the energy for the classical zone plate as shown in Table 13-2.

Another idealized type of zone plate, first proposed by Rayleigh (1888) and implemented by Wood (1898), is the phase plate. Here, the opaque rings are replaced by transparent rings, that impart a phase change of $\phi = \pi$. The phase plate transparency function is $T_{pp} = 1 + (e^{i\phi} - 1)T_{zp}$, so the Fourier series becomes

$$T_{pp} = \sum_{-\infty}^{+\infty} b_m \exp\left(\frac{im\pi r^2}{\lambda f}\right)$$

(27)

where $b_m = \delta_{m,0} + (e^{i\phi} - 1)q(-1)^m \text{sinc}(mq)$.

This shows that when $\phi = \pi$ and $q = 0.5$, which are the optimum values, the power in the $\pm m$ th harmonic is $|b_m|^2 = 4/(m\pi)^2$ for m odd, zero for m even and $|b_0|^2 = |1 + q(e^{i\phi} - 1)|^2 = 0$ for $m = 0$. The efficiency in the first-order focus is now about 40% and the disposition of energy is again shown in Table 13-2.

A third idealized type of zone plate is the Gabor plate (the hologram of a point) in which the rectangular profile of the last two devices is replaced by a sinusoid in r^2 space or a chirp function in r space. An absorption Gabor plate has only three orders, $m = \pm 1$ and 0 of which the $+1$ order has efficiency $1/16$. A phase Gabor plate, with a maximum phase change of 1.84 radians, has all the odd orders and the $+1$ order receives 34% of the light (Table 13-2).

Table 13-2. Efficiency of various types of zone plate.

Type of zone plate	Fresnel	Rayleigh-Wood	Gabor amplitude	Gabor phase
Type of zones	Opaque, transparent	Phase-change π , transparent	Sine-wave transparency	Sine-wave phase change max = 1.84 rad
Efficiency ± 1 order	1/4	$4/\pi^2$	1/16	0.34
Efficiency in general	$1/(m^2\pi^2)$ m odd 0 m even	$4/(m^2\pi^2)$ m odd 0 m even	0	$\neq 0$
Total positive orders ($m \neq 0$)	1/8	1/2	1/16	0.45
Total negative orders ($m \neq 0$)	1/8	1/2	1/16	0.45
Efficiency zero order	1/4	0	1/4	0.10
Absorbed	1/2	0	5/8	0
Overall total of last four rows	1	1	1	1

2.3.2 Real Structures

To make practically useful X-ray lenses, we must extend the treatment given so far to include the optical properties of real materials, suited to manufacturing zone plates. Such an extension was first provided in 1974 in an important paper by Kirz (1974) in which it was demonstrated that

1. Phase zone plates with primary efficiencies of 20–40% can be made from realistic materials.
2. Such zone plates can be designed to reduce or eliminate the zero-order beam and to reduce the absorbed fraction compared to a classical Fresnel zone plate.
3. These improvements can be effected essentially throughout the wavelength range 0.1–80 nm.
4. Realistic fabrication errors lead to only moderate deterioration in the optical performance.

For a phase-reversal zone plate made of a material with complex refractive index (Henke et al., 1993) $\tilde{n} = 1 - \delta - i\beta$, the efficiency $|b_m|^2$ can be found by making the replacement $\phi = \phi_1 + i\phi_2$ in (27) where $\phi_1 = kt\delta$, $\phi_2 = kt\beta$, $k = 2\pi/\lambda$, t is the thickness and we use the shorthand $r_a = e^{-\phi_2}$ for the amplitude attenuation factor. For $m \neq 0$, this gives

$$|b_m|^2 = \frac{1}{(m\pi)^2} (1 + r_a^2 - 2r_a \cos \phi_1) \quad (28)$$

and for $m = 0$

$$|b_0|^2 = \frac{1}{4} (1 + r_a^2 + 2r_a \cos \phi_1) \quad (29)$$

These equations (Kirz's (7) and (10) (Kirz, 1974)) give the efficiency of a planar zone plate of known thickness and refractive index. As discussed by Kirz and later by Michette (Michette, 1986), the optimum phase change is no longer π but is about 10–20% less. However, one can certainly choose a thickness to optimize the efficiency of a planar zone plate based on the use of Eq. (28). Plots of the theoretical efficiency, calculated using Eq. (28), for zone plates made of nickel and gold are shown in Figure 13–11.

2.4 Zone Plates: Fabrication and Examples

2.4.1 Fabrication Technique

The fabrication of X-ray zone plates involves several challenges. For high resolution imaging, one wishes to obtain the smallest possible value for the outermost zone width Δr_n or about 15–80 nm in high resolution modern examples. At the same time, the thickness t of the zone plate should ideally be that required to deliver a phase shift near π so as to maximize efficiency; this generally implies a thickness of greater than 100 nm for soft X-rays of 100–1000 eV energy, and a thickness near 1 μm for zone plates designed to operate at multi keV energies (see Figure 13–11). These two requirements are difficult to meet at the same time, for they lead to a demand for the fabrication of nanostructures

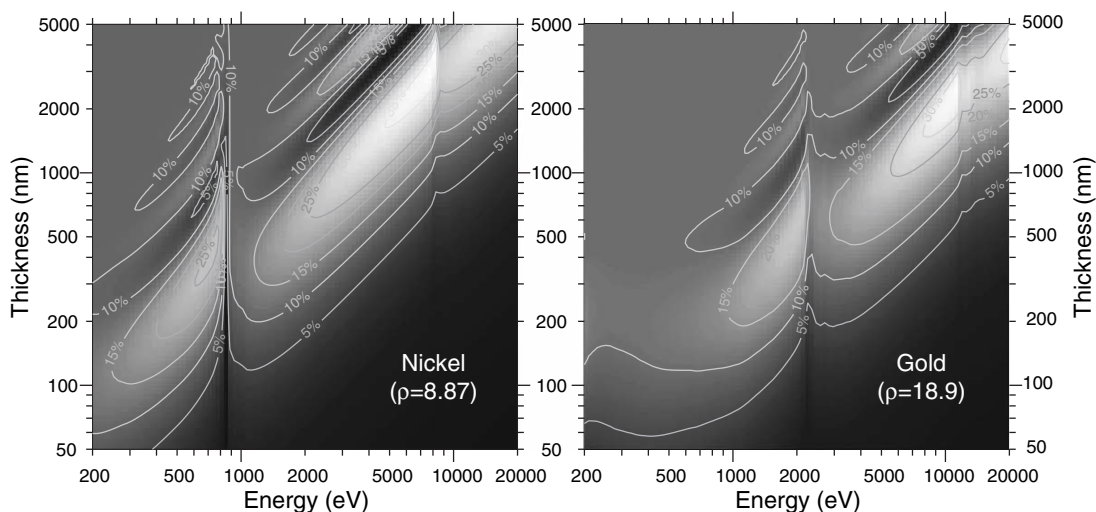


Figure 13-11. Contour plot showing the efficiency of planar zone plates made of Ni (left) and Au (right) according to Eq. (24) as a function of thickness and X-ray energy.

with very high aspect ratios (structure height over width). In addition, to obtain usable focal lengths of order 1 mm or larger, most zone plates used as high resolution objectives have diameters of 50–200 μm yet to minimize loss of efficiency all zones must be placed accurately to roughly one-third of their width (Simpson and Michette, 1983). This implies an absolute accuracy of zone placement of about 0.01% which is quite challenging but which can be achieved in modern 100 keV electron beam lithography systems that incorporate laser interferometer positioning control (Anderson et al., 2000; Tennant et al., 2000). Such zone plates typically have 300–1000 zones, and require corresponding quasimonochromatic illumination with $E/\Delta E > 300$ –1000 or better (Thieme, 1988).

Early demonstrations of X-ray zone plate fabrication used optical lithography to create free-standing zone plates with $\Delta r_n = 20 \mu\text{m}$ (Baez, 1960). An important early advance was the use of holographic methods to create zone plates with submicron zone widths (Schmahl and Rudolph, 1969; Niemann et al., 1974), eventually leading to outermost zone widths of $\Delta r_n = 56 \text{ nm}$ (Schmahl et al., 1984b). The method used by nearly all laboratories today involves electron beam lithography, which was first suggested by Sayre (Sayre, 1972) and subsequently demonstrated in several laboratories (Shaver et al., 1980; Kern et al., 1984; Buckley et al., 1985). In order to obtain high aspect ratio nanostructures, most laboratories now use some variation of tri-layer resist schemes (Tennant et al., 1981) with electroplating (Schneider et al., 1995) as illustrated in Figure 13-12. This approach allows for the writing of fine, dense features in an electron beam resist which is sufficiently thin that electron side scattering is minimized. A series of reactive ion etches are used to first transfer the e-beam pattern into a hard mask, and to use that hard mask to transfer the pattern into a second polymer

used either to define an etch mask for the underlying zone material (Tennant et al., 1991; David et al., 1995, 2000) or, more commonly, as a mold for electroplating of the zone structures (Schneider et al., 1995). Using these approaches, several groups have fabricated zone plates with outermost zone widths of 20 nm or below (Bögli et al., 1988; Schneider et al., 1995; Spector et al., 1997; Anderson et al., 2000; Peuker, 2001; Chao et al., 2003).

A recent promising approach (Chao et al., 2005) has been to write every other zone in one pass, and then to write the other half of the zones in a subsequent processing step (of course extremely high overlay accuracy is required). By increasing the distance from the next zone written in one pass, the proximity effect of electron beam lithography is reduced, leading to higher contrast and thus higher aspect ratio in the developed resist. In addition, the width of the resist used as a plating mold is tripled, thus reducing collapse during processing. This approach has been used to fabricate zone plates with an outermost zone width of 15 nm (Chao et al., 2005) and a theoretical zone efficiency of 6% (see next section). A key challenge lies in maintaining high efficiency as the outermost zone width is decreased. For applications requiring higher efficiency for greater flux in STXM, or reduced radiation dose in TXM, efficiencies of 10–18% have been obtained at 30 nm outermost zone width (Spector et al., 1997; Peuker, 2001). These efficiencies are for zone plates operating at 400–550 eV; zone plates for use at higher X-ray energies are discussed in Section 2.4.5.

2.4.2 Resolution-Determining Zone Plates

While a variety of groups are fabricating high resolution zone plates as described above, we consider here one example which is the effort of the Center for X-ray Optics at Lawrence Berkeley National Laboratory. By using a 100 keV electron beam lithography system with interferometric positioning control and customized circular pattern generation (Anderson et al., 1995), this group has made a series of zone plates with steadily improving resolution (Anderson et al., 2000; Chao et al., 2003) and have developed a sophisticated technique for determining the resolution (Jochum and Meyer-Ilse, 1995; Heck, 1998; Chao

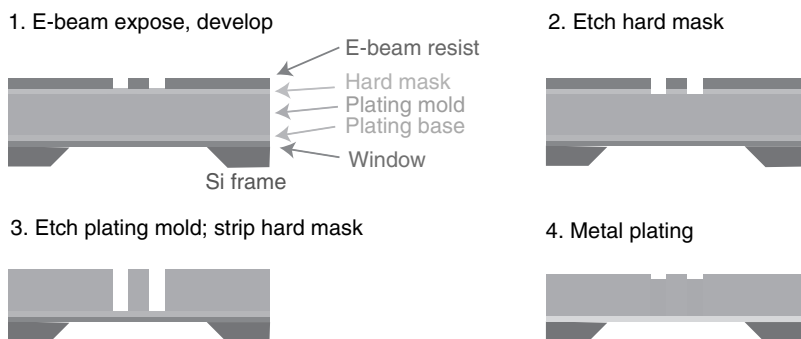


Figure 13-12. Schematic of zone plate fabrication technique as discussed in the text.

et al., 2003). The tests have been done using the XM-1 microscope at beam-line 6.1 at the Advanced Light Source at Berkeley USA (see Microscopes Layouts and Illumination Schemes (Section 3.1)). The latest results (Chao et al., 2005), which represent the best zone-plate-microscope resolution that we know of, show a resolution of $<15\text{ nm}$ at a photon energy of 815 eV (see Figure 13–13). This measured value depends on the degree of partial coherence of the illumination and we discuss it in that context in a later section. The zone plate had an outer zone width of 15 nm , which, with the XM-1 illumination system, gave a theoretically expected resolution of 12 nm . The zone plate was made by the above-mentioned new process in which the $n = 2, 6, 10, \dots$ and the $n = 4, 8, 12, \dots$ opaque zones are made in two separate groups. Chao and colleagues suggest that the new process is not yet at its limit and that 10 nm zone plates should be within reach.

2.4.3 Condenser Zone Plates

Condenser zone plates serve the dual function of imaging the source on to the sample (in critical illumination) and, in combination with a pinhole close to the sample, of acting as a moderate-resolution monochromator. Ideally they should deliver a beam which (1) has the same NA as the objective zone plate, (2) exactly fills the sample with light and (3) has a spectral bandwidth equal to the reciprocal of the number of zones of the objective. In practice, due to the fact that bending-magnet synchrotron-radiation sources usually have smaller phase space area than the microscope and due to the fabrication difficulties described below, conditions (1) and (2) cannot be fully met. Moreover, condition (3) implies that the condenser zone plate must be about $5\text{--}10\text{ mm}$ in diameter. We discuss the trade-offs involved here in more detail in Section 3.1.1.

The first condenser zone plates were fabricated by the Göttingen group (Schmahl and Rudolph, 1984b; Hettwer, 1998) using roughly the same holographic process then used to make objective zone plates. As a result, the finest line widths (and thus the NA s) of the condenser and

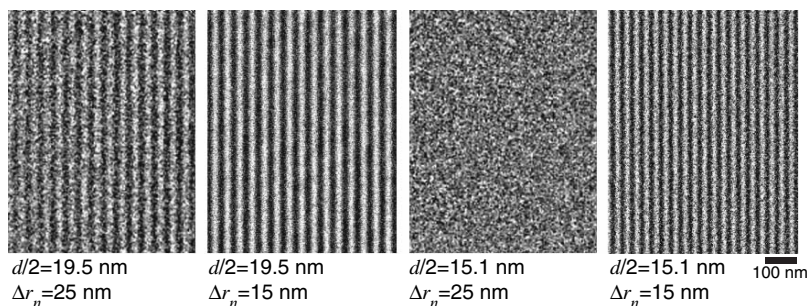


Figure 13–13. The soft X-ray images of square-wave test objects used by Chao and coworkers (2005) to demonstrate a microscope resolution of $<15\text{ nm}$ using a 15 nm - and a 25 nm -outer-zone-width zone plate as explained in the text. The half-period of the test object and the outer zone width of the zone plate for the four images are shown. (Reprinted from Chao et al., © 2005, with permission from Nature.)

objective tended to match, which is broadly what is required for best resolution (see for example Fig.10.13 of (Born and Wolf, 1999)). With electron beam lithography, new challenges have arisen. To achieve the same finest zone width and efficiency in the condenser as in the objective, would require the use of the same high-resolution but necessarily slow electron beam resists. Since the area of a 5mm diameter condenser is 10^4 times larger than a $50\mu\text{m}$ objective zone plate, it would take 10^4 times longer to fabricate the condenser with the same process. (Because aberrations on condensers do not degrade image quality, the requirements for zone placement accuracy do not scale up in the same fashion). For these reasons, as the resolution of the objectives has been pushed down below 20nm, the condenser zone plates have not kept up. Instead, typical condenser zone plates fabricated by electron beam lithography have outer zone widths of 50–60nm. As one example, the fabrication of a TXM condenser zone plate with 9mm diameter and 55nm outer-zone width required a 48-hour writing time (Anderson et al., 2000). Although process improvements have since reduced that time by about a factor of two, such large zone plates are still not widely available. The mismatch of numerical aperture between condenser and objective zone plates limits the modulation transfer at high spatial frequency (Born and Wolf, 1999) and thus limits the ability to detect small structures such as immunogold labels in bright field or dark field modes (Vogt et al., 2001a).

Besides the challenges of matching the NA of the best objectives and their limited availability, condenser zone plates are usually required to operate in an unfriendly environment, relatively near the source in a synchrotron beam line. Even with the protection of an energy-filtering mirror, they still often take significant heat load and the heat removal pathways are poor. This is an undesirable circumstance for optics that take so much effort to build. The problem of power deposition in condenser zone plates can be understood by solving the boundary value problem of a uniformly heated membrane (Howells et al., 2002). Assuming a square membrane of side a , thickness t and conductivity k , the solution for the temperature is

$$T(x, y) = \frac{16Qa^2}{\pi^4 kt} \sum_{\substack{m=1 \\ m, n \text{ odd}}}^{\infty} \sum_{n=1}^{\infty} \frac{1}{mn} \frac{\sin \frac{m\pi x}{a} \sin \frac{n\pi y}{a}}{m^2 + n^2} \quad (30)$$

where Q is the absorbed power density. The double sum is a simple function that has a peak of height 0.448 in the center and is zero at the edges. The center temperature is then given by the useful relation

$$T_{\text{max}} = 7.17 \frac{Qa^2}{\pi^4 kt}. \quad (31)$$

Since the maximum temperature depends on Q/t and, for a simple thin membrane, Q is proportional to t , the temperature is not reduced by making the membrane thicker. However, if the absorption is principally in the zone plate rings then a thicker membrane may help. A smaller or better-conducting membrane always helps.

The list of issues posed by condenser zone plates does not end here. Another serious limitation is that to change the X-ray energy one has to change the condenser zone plate or at least change its object and image distances. This is so difficult that systems fed by a condenser zone plate are effectively not energy tunable. Evidently there are plenty of reasons to seek alternatives to condenser zone plates and to this we now turn.

2.4.4 Alternatives to Condenser Zone Plates

Zone plates deflect the outermost ray by an angle equal to their NA (Eq. (7)). As explained in the previous section, manufacturable NAs of condenser zone plates are too small to match those of the best objectives by a factor of two or more. Now we know that a mirror can deflect an X-ray roughly by $2\sqrt{2\delta}$ or twice its critical angle. Thus we can characterize a mirror by an effective outer zone width $\Delta r_n = \lambda / (4\sqrt{2\delta})$, that a zone plate would need to have to produce a deflection equal to twice the critical angle of the mirror. Since δ is proportional to λ^2 (Section 1.1), one can see that the effective outer zone width will vary rather slowly with both energy and electron density. For example, for a platinum, mirror it varies from 6 nm at 0.5 keV to 4 nm at 5 keV. While for a silicon dioxide mirror it varies from 12 nm to 10 nm over the same energy range. This shows that if a suitable geometry can be arranged a single-reflection mirror system could be a very effective condenser. However it is important to note that a reflective condenser will not provide a monochromatic beam and that either a line source or a monochromator will be required. As we shall see, a separate monochromator coupled with a reflective condenser will have major advantages.

To produce the desired hollow-cone illumination, a grazing incidence ellipsoid of revolution in the form of a hollow tube would be suitable. The input angle should match the angle of the beam from the source or monochromator and the output angle should match or exceed the objective NA. For a synchrotron beam this design will normally lead to under filling of the sample which can be overcome by wobbling. The fabrication accuracy should be such that the point spread function of the mirror is small compared to the size of the object field. Mirrors of this general type have been made for some time (often starting from glass capillaries) and are quite widely used in the hard X-ray research community as reviewed for example by Bilderback (2003).

Single-reflection monocapillary X-ray mirrors have now been in use for microscopes at both laboratory and synchrotron sources (Section 1.4) for the last year or two and have enjoyed considerable success. For example, documentation of the performance of the X-ray microscope at the Taiwan synchrotron, including details of its condenser mirror, is provided by Tang et al. (2006) and Yin et al. (2006). These condenser mirrors have many advantages compared to condenser zone plates. Specifically, capillary-mirror condensers

1. Are more readily available
2. When fed by a separate monochromator allow the TXM to be truly energy-tunable
3. Are able to match the NA of any currently available objective zone plate

4. Are a factor 3–15 times more efficient with no unwanted orders
5. Are more robust, longer lived and easier to clean
6. Are less fragile with respect to thermal or mechanical damage
7. Do not require a band-width-selecting pinhole close to the focus and thus do not limit the size of the sample holder nor its ability to rotate.

As indicated earlier (Section 1.4) we believe that the advent of an energy-tunable TXM is important and may enable the TXM to become competitive for spectromicroscopy.

2.4.5 Zone Plates with Shaped Grooves

Until now we have talked about square-wave zone plates with a gap to period ratio of 0.5 that behaved according to the theory of a thin zone plate, even if the thickness was greater than Δr_n . Just as a blazed reflection grating with a saw-tooth profile has much better efficiency than a square-wave grating (even if the latter is a perfect phase grating with π phase shifts), so one can get higher efficiency from zone plates with shaped groove profiles. Considering that a zone plate is intended to synthesize a smooth spherical wave front from a succession of ring-shaped parts, we might expect that the optimum groove shape will be a parabola that increases the phase shift smoothly across the zone-plate period. In fact the mathematical treatment (Tatchyn et al., 1982; Michette, 1986) shows that the thickness function is

$$t_i(r) = \begin{cases} \sqrt{f^2 + r^2} - \sqrt{f^2 + r_{i-1}^2} & r_{i-1} \leq r < (r_i - d_i) \\ 0 & (r_i - d_i) \leq r \leq r_i \end{cases} / \delta \quad (32)$$

where d_i/r_i can be calculated (Tatchyn et al., 1984) and is the outer fraction of the i th period which is to be left open. The first-order efficiency of a nickel zone plate made according to this specification would be about 80% at 7 keV.

It is hard to microfabricate a smooth curve but one can still get much of the advantage of this scheme by approximating the parabolic profile by a stepped structure (Di Fabrizio et al., 1994; Yun et al., 1999). For example Di Fabrizio et al. (1999) have made a nickel zone plate with four equal width steps of optical delay 0, 0.25, 0.5 and 0.75 wavelengths. The measured first order efficiency of this zone plate was 55% at 7 keV, which represents a substantial improvement in efficiency and suppression of unwanted orders compared to traditional soft X-ray performance and shows the benefits of both groove-shaping and phase-plates. Evidently the use of several thickness steps implies that the outermost zone must be several times wider than the finest line width that can be achieved with the particular fabrication process, so that this approach involves a tradeoff between spatial resolution and efficiency.

2.4.6 Hard X-Ray Zone Plates

While much of the effort of the last three decades has gone into developing zone plate microscopy in the 290–540 eV “water window” region for studies of 0.1–10 μm thick specimens, there is increasing activity in hard-X-ray zone-plate imaging at energies of roughly 2–15 keV. Scanning fluorescence X-ray microprobes (SFXM) using zone plate optics are providing new capabilities for trace element mapping, and hard

X-ray transmission X-ray microscopes (TXMs) using absorption or especially phase contrast are able to image much thicker objects than their soft-X-ray counterparts. Zone plates for these energies must be much thicker to achieve good efficiency (see Figure 13–11) which places increasing demands on zone aspect ratio in lithographically patterned zone plates and means that the minimum zone width (and thus first order spatial resolution) were initially in the 50–100 nm range. However, as noted in Section 1.4, several 45 and 50 nm hard X-ray zone plates are now operational at synchrotrons. At the same time, because the ratio of phase shifting to absorption f_1/f_2 improves as the energy is increased, the achievable efficiency becomes much higher and the depth of focus increases considerably (Jacobsen, 1992), which is helpful for applications such as tomography. Quantitatively, the transverse resolution of a zone plate is given by $0.61\lambda/\text{NA} = 1.22\Delta r_n$ (Eq. (8)) and the depth of focus by $2\lambda/(\text{NA})^2 = 8(\Delta r_n)^2/\lambda$. Therefore, a zone plate with $\Delta r_n = 50$ nm has a depth of focus of about 160 μm at 10 keV as opposed to about 8 μm at 500 eV. In addition, the focal length $f = 2r_n(\Delta r_n)/\lambda$ (Eq. (6)) for such a zone plate with 100 μm diameter increases from 2 to 40 mm, which considerably eases some of the challenges of mechanical design for specimen temperature control, insertion of fluorescence detectors, and so on.

In spite of the challenges of fabricating thicker zone plates using lithographic techniques, much success has been achieved. In some cases the initial electron beam lithography write has been transferred into a thicker plating mold using reactive ion etching as described above; this has led to the commercial availability (Xradia, Inc.) of a variety of high-aspect-ratio zone plates including one with outer-zone width 50 nm and thickness 700 nm, or an aspect ratio of 14:1 (an example of an earlier Xradia zone plate is shown in Figure 13–14). Additional parameters of zone plates given at the latest X-ray microscopy conference (Himeji, Japan, June 2005) are given in Section 1.4. Other approaches have involved using an electron-beam-written zone

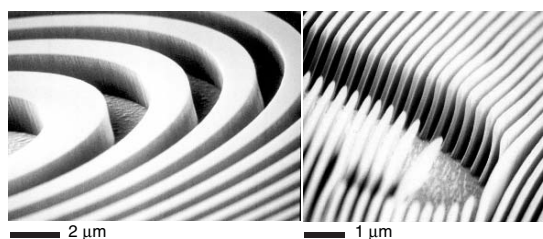


Figure 13–14. A hard X-ray zone plate with 100 nm outermost zone width and 1.6 μm thickness of gold for use at 5.4 keV in a commercial X-ray microscope system (Xradia, Inc.). The simultaneous achievement of narrow zone widths for high spatial resolution, and significant zone thickness so as to achieve a π phase shift, means that the achievement of high aspect ratio nanostructures is important. This zone plate has an aspect ratio of 16:1 and a theoretical focusing efficiency of 31%. (Figure Courtesy W. Yun, Xradia.)

plate as a mask for the subsequent processing of a thicker zone plate using X-ray lithography (Shaver et al., 1980; Lai et al., 1992), including the fabrication of $2.5\mu\text{m}$ thick zone plates with a finest zone width of $0.25\mu\text{m}$ (Krasnoperova et al., 1993). Another approach is to stack two or more zone plates together (Shastri et al., 2001).

Sputter-sliced or “jelly roll” zone plates (Schmahl et al., 1980) represent a completely different approach in fabrication. The goal of this approach is to start with a rotating wire and then build up alternating layers of weakly and strongly refractive material by sputtering or evaporation. The resulting structure is then sliced to yield zone plates of the appropriate thickness. In this case the achievement of high aspect ratios is not at all challenging; instead, the challenges include avoiding error and roughness accumulation in realizing the proper zone radii, the difficulties of maintaining perfect cylindrical symmetry, and the challenges involved in slicing the structure.

Recent results from the groups involved (Bionta et al., 1994; Tamura et al., 2002; Duvel et al., 2003) show that the technique is making steady progress to the point where a zone plate consisting of 70 Cu/Al layer pairs with outer zone width of $0.16\mu\text{m}$ and aspect ratio of more than a thousand has been used to focus a 100 keV beam from Spring 8 to $0.5\mu\text{m}$ FWHM (Kamijo et al., 2003). Similarly a sputter-slice soft X-ray zone plate with an aspect ratio of 200, made at Göttingen Germany, had 188 layer pairs of the alloy Ni₈₀-Cr₂₀ and silica. It showed a measured efficiency of 3.8% at 4.1 keV but had a focal spot size considerably greater than the 17-nm outer zone width. These are intriguing results, though for the moment, the sputter-sliced approach has not yet produced optics with an optical performance consistent with the intended geometrical parameters.

2.4.7 Thick Zone Plates

Up until now we have used kinematical diffraction theory to understand the properties of zone plates. In this theory, the incident wave and the diffracted signals from each volume element are all treated as independent. However, in reality, the “incident” wave and the “diffracted” waves in the solid structure are coherently coupled and if the zone plate is thick enough, the effect of this coupling will become evident at the output. Such coupling is known in perfect-crystal diffraction where it leads to anomalous transmission in Laue-geometry experiments (the Borrmann effect), while on a larger size scale, “Bragg-effect” holograms show similar behavior. Zone plates can be designed to exploit coupled-wave effects and these devices offer the possibility of very high efficiency and resolution in high diffraction orders, thus exceeding the resolution limit of the outermost zone width which applies when operating in first order. Theoretical treatments of diffraction by thick periodic structures have been developed in the hard-X-ray community (dynamical diffraction by crystals) (Batterman and Cole, 1964) and the optical-holography community (coupled-wave theory) (Kogelnik, 1969; Solymar and Cooke, 1981). The coupled-wave method has been applied to X-ray gratings and zone plates by Maser (1994) and more recently by Schneider (1997) who has given a solution that includes

the case of high orders and gap-to-period ratios other than 0.5. Schneider's solution predicts that, in the soft X-ray region, absolute efficiencies of 30–50% in a single high order are indeed possible with line-to-space ratios of 0.1–0.5 and aspect ratios greater than 30:1. Hambach et al. (2001) have followed up these calculations with a series of experiments, mostly with copolymer gratings with aspect ratio 10:1. The predictions of the theory were broadly confirmed and a maximum efficiency of 15.3% was achieved at 13 nm wavelength. This value was 75% of the prediction of the coupled wave theory and 25 times greater than the prediction of thin-grating (kinematic) theory. The authors suggested that zone plates based on this principle may find application as condensers for table-top microscopes using isotropically emitting sources. A similar verification of dynamical diffraction theory for the case of sectioned multilayers illuminated with 19.5 keV X-rays in Laue geometry has been published recently (Kang et al., 2005). A reflecting efficiency of 70% was observed. Both of these experiments used gratings as being representative of the diffraction-efficiency behavior of a conventional zone plate (Hambach) and a sputter-sliced zone plate (Kang) respectively. However, as pointed out by Maser (1994), thick zone plates, like volume holograms, have a directional selectivity based on Bragg's law. That is, the zones must be oriented so that the incoming wave is locally Bragg-reflected by the zones and there will be a rocking curve outside of which the high efficiency is lost or moved to another order. Eq. (6) can be regarded as showing that Bragg's law is obeyed at every zone of a conventional zone plate operating at magnification unity. On the other hand, for high magnification or demagnification applications, the zones must be tilted by an angle that varies with radius. The difficulty of doing this in practice is currently delaying the application of these ideas to practical zone plates.

A promising recent alternative to tilt angles that vary continuously with angle has been to apply the concepts of the "sputter-sliced" zone plate (see previous section) to produce linear half-zone-plates called "multilayer Laue lenses." One starts with an atomically smooth flat and deposits alternating zone materials, starting with the highest-order zones, so that error accumulation mainly affects the coarser, low-order zones. Two multilayer Laue lenses together can achieve two dimensional focusing in the manner of a Kirkparick-Baez mirror pair. Although tilting by a continuously variable angle is not exactly achieved by this approach it can be approximated by tilting the whole lens to a compromise Bragg angle. One dimensional line foci as narrow as 19 nm have already been reported by this method (Maser et al., 2004; Kang et al., 2006) and these authors believe that 5 nm may be possible in future work.

3 X-Ray Microscopes

In the previous sections we have described some of the characteristics of X-ray interactions and focusing optics. We now turn our attention to a discussion of X-ray microscopes currently in operation. They fall

into two classes: full-field imaging and scanning, which are both illustrated in Figure 13–15. A large number of microscopes are listed in Table 13–3. We also describe three specific microscopes as examples: a transmission X-ray microscope (TXM) operated at Lawrence Berkeley National Laboratory, a scanning transmission X-ray microscope (STXM) operated at Brookhaven National Laboratory, and a scanning fluorescence X-ray microprobe (SFXM) operated at Argonne National Laboratory.

A key difference between TXM, STXM, and SFXM concerns the illumination phase space that can be accepted. In STXM and in SFXM, the size of the spot delivered by the zone plate objective is a convolution of the geometric image of the source and the point spread function

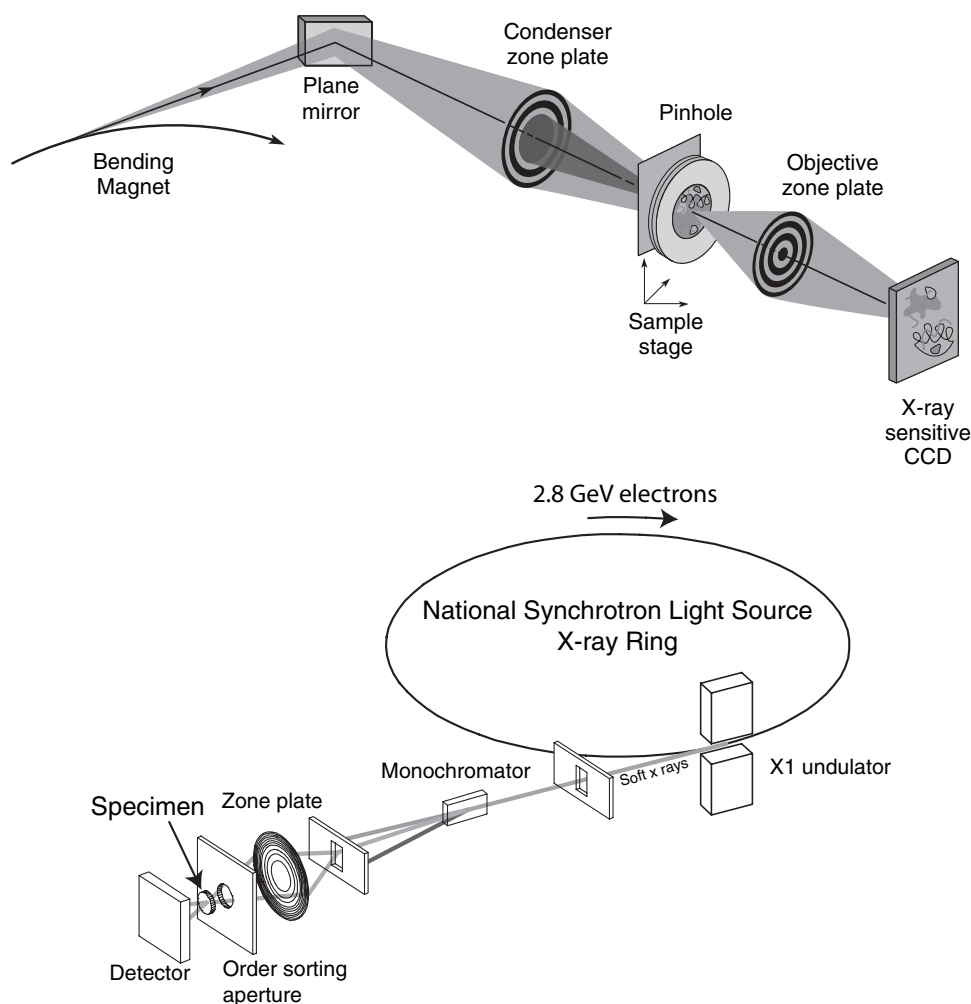


Figure 13–15. Schematic of the main components of a transmission X-ray microscope or TXM (top: courtesy of D. Attwood, Lawrence Berkeley National Laboratory) and a scanning transmission X-ray microscope or STXM (bottom: courtesy of Y. Wang, then of Stony Brook.) (See color plate.)

Table 13–3. Zone plate microscopes.

Microscope/ location	Light source	Illumination/ monochromator	Focusing, imaging	Contrast mechanisms	Techniques, X-ray energy	Citation
MES STXM	ALS undulator	grating	STXM	absorption, magnetization	NEXAFS, MCD 100 to 2000 eV	(Tyliszczak, 2004; Warwick, 2004)
Polymer STXM	ALS bend magnet	grating	STXM	absorption	NEXAFS 250 to 750 eV	(Warwick, 2002; Kilcoyne, 2003)
XM-1 TXM	ALS bend magnet	zone plate condenser/mono	TXM	absorption, magn- etization, phase	Tomog, MCD 200 to 1800 eV	(Meyer-Ilse, 2000a)
XM-2 TXM	ALS bend magnet	zone plate condenser/mono	TXM	absorption, phase	Tomog 200 to 7000 eV	
2-ID-B	APS undulator	multilayer-coated grating	STXM	abs, fluor, phase XANES, tomog	tomography 600 to 4000 eV	(McNulty, 2003a; McNulty, 2003b)
2-ID-D	APS undulator	Crystal/multilayer	STXM	diffraction	strain mapping 6 to 20 keV	(Cai, 2003, McNulty, 2003a)
2-ID-E	APS undulator	Crystal/multilayer	STXM	fluor, XANES diff, microdiff	5 to 35 keV	(McNulty, 2003a)
26-ID	APS undulator	Crystal/multilayer	STXM TXM	abs, fluor, diff XANES	3–30 keV	(McNulty, 2003a)
BL20B2	Spring8 bend magnet	crystal	STXM	absorption	4 to 113 keV opt testing, tomog	(Suzuki, 2003, Takano, 2003)
BL47XU	Spring8 undulator	crystal	TXM	absorption	5–37.7 keV tomography	(Suzuki, 2003, Uesugi, 2003)
BL20XU	Spring8 undulator	crystal—250 m beam line	STXM	absorption	8–37.3, 24–113 keV, μbeams	(Suzuki, 2003)
BL24XU	Spring8 undulator	crystal	TXM	phase contrast	8.77–12.85, 12.4–18.17 keV	(Tsusaka, 2001, Kagoshima, 2003)
BL12	Ritsumeikan bend magnet	zone plate condenser/mono	TXM	absorption	water window,	(Takemoto, 2003)

8A1 U7 SPEM	Pohang undulator	grating	STXM	photoemission	nanoXPS 100 to 1000 eV	(Shin, 2003, Yi, 2005)
1B2 hard xray	Pohang b. magnet	crystal	TXM	absorption	6.95 keV	(Youn, 2005)
U41TXM	BESSYII undulator	zone plate condenser/mono	TXM	absorption, phase	water window, 2D, 3D imaging	(Guttman, 2003, Wiesemann, 2003)
UE46TXM	BESSYII undulator	zone plate condenser	TXM	absorption, magnetization	MCD 0.2 to 2 keV	(Eimüller, 2003)
TWINMIC	ELETTRA Undulator	grating	STXM & TXM	absorption phase contrast	NEXAFS 250 to 2000 eV	(Kaulich, 2003)
BL2.2 ESCA	ELETTRA undulator	grating	STXM	absorption photoemission	nanoXPS 200 to 1400 eV	(Casalis, 1995; Kiskinova, 2003)
KINGS STXM	laser plasma	gas filtered spectrum	STXM	absorption	water window	(Michette, 2000)
X1A STXMs	NSLS undulator	grating	STXM	absorption diffraction phase contrast	NEXAFS, cryomicroscopy 250 to 1000 eV	(Jacobsen, 2000a)
ID21 microscopes	ESRF undulator	grating, crystal	STXM & TXM	absorption fluorescence diffraction phase contrast	NEXAFS 200 to 7000 eV	(Susini, 2000)
ID22 imaging	ESRF undulator	crystal	STXM	absorption fluorescence phase contrast	5 to 70 keV	(Weitkamp, 2000)
Aarhus TXM	ASTRID bend magnet	zone plate condenser/mono	TXM	absorption	typically 517 eV	(Uggerhøj, 2000)
XRADIA	Chromium anode	reflective condenser	TXM	absorption, phase	tomography 5.4 keV	(Scott, 2004)

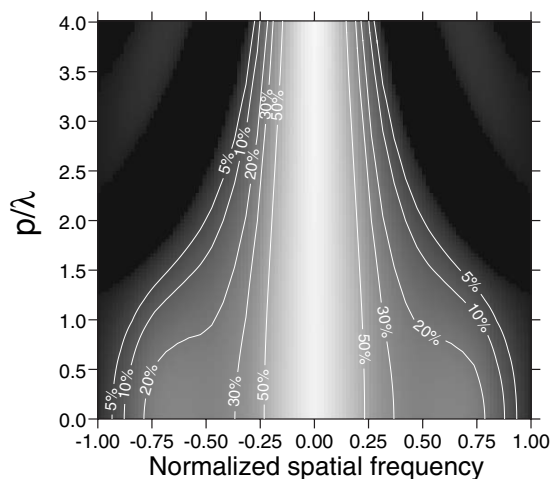


Figure 13–16. The aim in operating a scanning microscope or microprobe is to have a diffraction-limited focus. Therefore the source must be sufficiently demagnified so that it contributes negligibly to the focal width. This contour plot shows how the modulation transfer function (MTF) of an optic with a half-diameter central stop is affected by increasing the phase space parameter p of the source. This parameter $p = w\theta$ (source full width w times the full angle θ accepted by the optic) should be less than about the wavelength λ in both the x and y directions to achieve maximum spatial resolution. The normalized spatial frequency is defined to be unity at the MTF cutoff of $1/\Delta r_n$. (From Winn et al., 2000.)

of the optic. As Figure 13–16 shows, for an objective with diffraction-limited (as opposed to aberration-limited) resolution, the effect of the geometric source size becomes negligible if the product $p = w\theta$ of source width w times the full angle θ accepted by the optic is less than the wavelength λ in each dimension. This is commonly summarized by saying that scanning microscopes require single-mode illumination, although it is understood that a spatially filtered, incoherent source is not the exact equivalent of a single-mode optical cavity. The situation in TXM is much different; for incoherent bright field imaging, each pixel in the object can be imaged independently of its neighbors (within good approximation), so one can illuminate all object pixels simultaneously and with nominally incoherent light. If object resolution elements are imaged in 1:1 correspondence to detector pixels in a TXM, the number of “modes” of phase space p/λ that can be accepted in the x direction is approximately equal to the number of detector pixels in that direction and the same holds for y . As a result, TXMs are often operated with bending magnet synchrotron radiation sources or laboratory sources which deliver high flux (photons per solid angle), while STXMs and SFXMs are often operated with undulator sources which deliver high brightness (photons per solid angle per source area). The issues of microscope illumination and its effects on image formation will be discussed in more detail in Section 3.1.

3.1 Microscope Layouts and Illumination Schemes

3.1.1 Transmission X-Ray Microscope (TXM) Layout

Full-field transmission X-ray microscopes (TXMs) typically use a zone plate to produce a magnified image of the specimen on a 2D detector. This approach was pioneered by the group of G. Schmahl at the Universität Göttingen, who, after initial experiments including reflection-grating monochromators (Niemann et al., 1976) switched to using a condenser zone plate as the sole monochromator (Rudolph et al., 1984). This latter approach is now used by a number of TXMs, including the XM-1 at Lawrence Berkeley Lab (Meyer-Ilse et al., 1994, 2001) for which we provide some example numbers. As shown in Figure 13–15a, the beam from the synchrotron bending magnet source is deflected by a grazing-incidence mirror which filters out the power due to high-energy X-rays, passes through a thin metal filter to remove visible and ultraviolet radiation, and is then imaged by the condenser zone plate onto a pinhole located just upstream of the specimen. As noted in Section 2.4.3 on condenser zone plates, the condenser zone plate (of diameter $D = 9$ mm) and the pinhole (of diameter $d \approx 10$ – $20 \mu\text{m}$), together are equivalent to a monochromator of resolving power equal to $D/(2d)$ (Niemann, 1974). Because the light transmitted by the objective zone plate includes a significant undiffracted (zero order) component which must not reach the detector, the illumination of the sample needs to be hollow-cone and this is achieved by means of a stop built into the condenser, blocking a central circle of radius about one third to one half of the condenser radius. The objective zone plate used by XM-1 in the resolution test described above had the following characteristics: outer zone width $\Delta r_n = 15$ nm, diameter $d = 30 \mu\text{m}$, 500 zones of 80 nm thick gold (giving a maximum aspect ratio of 5:1), and focal length $f = 0.3$ mm at 815 eV. This is the highest resolution zone plate used to date and slightly larger outer zone widths (25–30) are used for routine user operations. The vertical phase space area of the synchrotron source is generally smaller than its horizontal phase-space area and smaller than that of the microscope (which equals object full-width d times twice the objective NA). Since the condenser zone plate cannot expand the phase space, both the object width and the numerical aperture of the objective of a TXM will generally be underfilled. To counter the under filling of the object field, the condenser is usually “wobbled” up and down during the course of an exposure. This type of microscope layout, in which the source is imaged on to the sample, is known as “critical illumination” (Born and Wolf, 1999) and is widely used for amplitude contrast.

Traditionally, the specimen has been placed in an atmospheric pressure environment and to accomplish this, thin vacuum windows (100 nm Si_3N_4 or Si are common) can be used between the condenser and the specimen, and also between the specimen and the objective zone plate. Because the focal length of the objective zone plate is quite small (for example, in the case of a 25-nm-outermost-zone-width, 60- μm -diameter zone plate operating at 530 eV it would be 1.3 mm.), the specimen region lying between these two windows is quite constrained.

The beam then re-enters a vacuum environment where the objective zone plate and the image detector are located. At energies below a few keV, the most common detector is a backside-thinned CCD which is directly illuminated by the X-ray beam; at higher energies, phosphor screens imaged by a visible light lens onto a CCD detector are commonly used. Because of the desire to deliver 10–50 nm resolution using detectors with 1–20 μm pixel size, the distance from the zone plate objective to the detector is often in the range 1–2 m to give acceptably high optical magnification.

The approach described above is commonly used with bending magnet and laboratory sources. Particular challenges arise when the source phase space area is dramatically smaller than desired, which is the case for undulator sources on low emittance storage rings. As an example, Niemann has studied a variety of solutions for the condenser of the current Berlin TXM which is illuminated by a BESSY II undulator (Niemann, 1998). The adopted solution (Niemann et al., 2000) involves a zone plate segment, and three flat mirrors. The magnification of the zone plate is chosen to fully illuminate the object field. The first mirror is fixed but must be tilted for a change of wavelength. The other two mirrors are mounted in a structure that rotates and delivers an incoherent hollow-cone beam of which the inner-to-outer angular difference ($\Delta\theta$ say) is determined by phase-space matching and the outer angle (the NA) is determined by the last-mirror reflection angle. Thus the illuminated area and the NA can both be chosen, no wobbling is required and Liouville's theorem (which states that the phase space area of an optical beam is a conserved quantity) is respected by allowing $\Delta\theta$ to float. This system represents an elegant optical solution, though it is mechanically quite complex. Other strategies to expand the phase space of an XUV beam have been explored by the microfabrication community who are concerned about "fringing" in XUV lithography (Murphy et al., 1993; White et al., 1995). One such approach is to design a pseudorandom diffractive optic specifically to "spoil" the phase space of a beam and match the object size and the NA (David et al., 2003); such optics must meet the challenge of evenly filling both the object plane and the back focal plane with light.

3.1.2 TXM Phase Contrast Layout

As noted above, phase contrast plays an important role in X-ray microscopy, particularly at higher photon energies. In order for phase variations at the specimen plane to produce intensity variations at the detector, some method of mixing the wave diffracted by the specimen with an undiffracted phase-reference wave must be employed. The most common approach in X-ray microscopes is that of Zernike. In light microscopes, Köhler illumination is provided by using a relay lens to image the source on to the front focal plane of the condenser. Points at this front focal plane deliver parallel beams to the object plane which (if undeviated by the object) are focused on to a phase ring at the back focal plane of the objective where they are phase shifted usually by $\pm\pi/2$. Thus the ring aperture at the front focal plane of the condenser provides a narrow, hollow cone of illumination of the speci-

men, and is conjugate to the phase ring. At the same time, light originating from a point scatterer in the object is focused to the detector, where it interferes with the phase-shifted unscattered light (see Figure 13–17).

In X-ray microscopes, it is more difficult to use a relay optic to work in the Köhler illumination condition; instead, the front focal plane ring aperture is illuminated by nearly parallel light from the source (i.e., critical illumination) so that a much smaller area of the condenser is illuminated. Because the light from the front aperture is much more collimated than would have been the case with Köhler illumination, the longitudinal location of the aperture and its corresponding phase ring is much less critical than it is with visible light microscopes so the

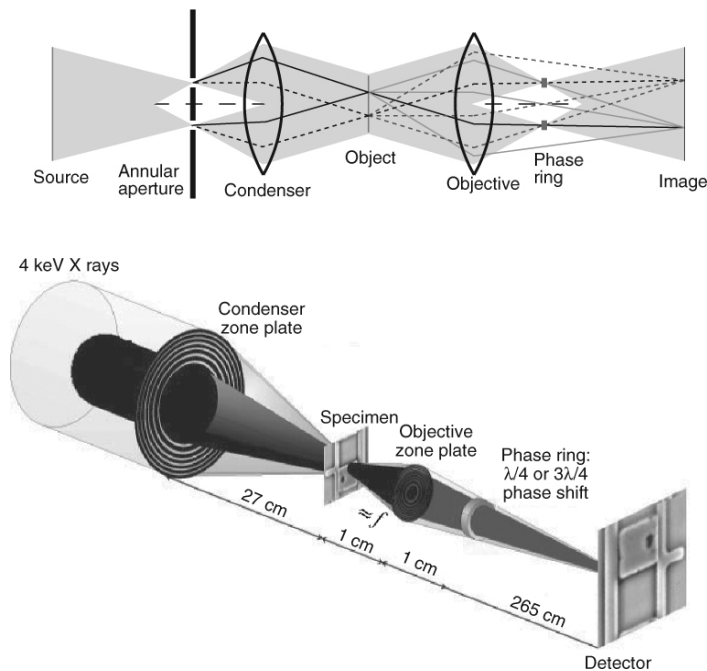


Figure 13–17. Two illustrations of a Zernike phase contrast optical system. The upper one shows the classical scheme used in light microscopes based on Köhler's illumination (Born and Wolf, 1999). Light from each point of the annular aperture, placed in the front focal plane of the condenser, is delivered to the object as a parallel beam. Two object points are shown receiving example rays from the source. The rays may be undeviated by the object, in which case they are seen to pass through the phase ring on their way to the detector. On the other hand the rays may be deviated (diffracted) by the object in which case they reach the detector without passing through the phase ring. Interferences between these two types of optical signal result in a mapping of the object phase variations into an intensity pattern on the detector. The lower diagram shows a practical synchrotron radiation implementation of a Zernike-phase-contrast TXM at the 4.1 keV beam line on ID21 at the European Synchrotron Radiation Facility in Grenoble. An undulator X-ray source is followed by a crystal monochromator illuminating a condenser zone plate (which can be small since it does not have to act as a linear monochromator). The condenser provides critical illumination to the sample rather than Köhler and, due to the good collimation of the incoming beam, the illuminated area of the condenser is projected on to a phase ring in the back-focal plane of the objective zone plate so as to provide phase contrast. The tradeoffs involved in choosing the illuminated area of the condenser pupil are discussed in Section 3.3.6 (Courtesy J. Susini, ESRF.)

primary alignment requirement is transverse to the X-ray beam direction. For computations it is convenient to consider the two rings as built into the lens pupil functions. Evidently when the source is well-collimated, as in the case of a synchrotron, there is no need for the two rings to be exactly conjugate. The use of the ring aperture obviously makes the illumination more coherent. The question of the best choice of width and radius for the two rings or, equivalently, how much coherence to have, we defer until later (Section 3.3.6). The phase ring itself is constructed out of a material with a large phase shift per absorption length $f_1/(2f_2)$, such as any material used at energies just below an absorption edge.

Zernike phase contrast X-ray microscopy was pioneered by Schmahl and Rudolph (1987) and Rudolph et al. (1990), and the Göttingen group has shown impressive results in phase contrast for water-window imaging of biological specimens. The ring aperture and phase ring used in these experiments could be rapidly inserted or retracted (Schmahl et al., 1994, 1995; Schneider, 1998). Phase contrast is arguably even more important at higher X-ray energies where it is the dominant contrast mechanism. For example a hard X-ray (4 keV) phase-contrast microscope, illuminated by a system using a crystal monochromator followed by a condenser zone plate, is operating at the ID21 beam line at the European Synchrotron Radiation Facility in Grenoble, France (see Figure 13–17). Imaging of functioning integrated circuits at 60 nm resolution has been demonstrated (Neuhäusler, 2003); see Section 4 for more information on this. In another example of this configuration, the National Synchrotron Radiation Research Center in Hsinchu, Taiwan has installed a TXM built by Xradia. This instrument (alluded to in Sections 1.4 and 2.4.4) has been used to demonstrate the long-existing idea of using zone plate higher focal orders for imaging. In water-window instruments which typically have focal lengths on the order of 1–2 mm in first order (and therefore 1/3–2/3 mm in third order), the idea has not been readily adopted due to practical considerations of working distance. However, in the Taiwan experiment (Tang et al., 2006) a phase contrast image of a fabricated test object was made at 8 keV using a 50 nm outer-zone-width zone plate in third order. Lines of minimum width 30 nm were imaged clearly and the authors estimate a resolution below 25 nm. This is evidently a most important development (see Section 5).

3.1.3 *Scanning Transmission X-Ray Microscope (STXM) Layout*

Scanning transmission X-ray microscopes (STXMs) typically use a zone plate to demagnify a pinhole source to a small focus spot through which the specimen is scanned. While initial demonstrations using synchrotron radiation used pinhole optics (Horowitz and Howell, 1972; Rarback et al., 1980), the use of zone plate optics in scanning microscopes was pioneered by Rarback et al. (1984) and later by (Niemann, 1987; Niemann et al., 1988). (Normal incidence optics with synthetic multilayer reflective coatings have also been used in the 50–120 eV range (Haelbich, 1980a; Haelbich et al., 1980b; Ng et al., 1990)). Since scanning microscopes require coherent illumination to reach their

maximum resolution, they have often used undulators as high brightness sources (Rarback et al., 1988; Kenney et al., 1989; Morrison et al., 1989b) though excellent performance has also been obtained using bending magnet sources on low emittance storage rings (Kilcoyne et al., 2003). While a large number of STXMs are now in operation, we describe here the characteristics of the most recent in a series (Rarback et al., 1988; Jacobsen et al., 1991; Feser et al., 1998, 2000) of undulator-based scanning microscopes built at Stony Brook University for operation at the National Synchrotron Light Source at Brookhaven National Laboratory in New York. A soft X-ray undulator plus spherical grating monochromator with an energy resolution that can be as good as 0.06 eV at 290 eV (Winn et al., 2000) is used to deliver soft X-rays to a 2D exit slit which can limit the beam size in the range 25–120 μm in both x and y . This slit then serves as a secondary radiation source for zone plates of either 80 or 160 μm diameter and zone widths of 30–45 nm (Spector et al., 1997; Tennant et al., 2000), producing a focal spot of 36–54 nm Rayleigh resolution. The beam emerges from the ultra high vacuum synchrotron beam line into an atmospheric pressure environment by passing through a 100 nm thick Si_3N_4 window. The zone plate includes a central stop of about half the zone plate diameter; this stop must be made quite thick (0.3 μm gold is common for soft X-ray applications) so that the undiffracted light transmitted through the large central stop is kept to a very small level compared to the flux in the focused X-ray beam. The zone plate is then followed by an order sorting or selecting aperture (OSA) so that a pure first-order focal spot is obtained.

While steering mirrors are used to scan the beam in visible light scanning microscopes, it is easier to maintain signal uniformity by keeping the beam and zone plate fixed and scanning the specimen through the focal spot. This is accomplished using an X-Y-Z stack of stepping motor stages for large motion with 1 μm precision, and a piezo scanning stage for 50–100 μm range and nanometer precision. Because piezos have nonlinearities and hysteresis in their response to scan voltages, some form of closed-loop feedback is generally used, based on position signals such as those provided by linear voltage differential transformers (Kenney et al., 1985), capacitance micrometers (Jacobsen et al., 1991), or laser interferometers (Shu et al., 1988; Kilcoyne et al., 2003); the latest Stony Brook STXM allows the user to choose between capacitive or laser interferometer feedback. The specimen is then followed by a high efficiency X-ray detector; common choices include the use of gas-based proportional counters which offer extremely high efficiency of detection for those X-rays that make it through a thin entrance window (Rarback et al., 1980; Kenney et al., 1985; Feser et al., 2000) but which suffer from a count-rate limit of about 1 MHz. Alternatives are phosphor-coated screens followed by photomultipliers to detect the resulting visible light (Maser et al., 2000), and solid state detectors which are capable of significantly higher signal rates (Barrett et al., 1998; Wiesemann et al., 2000; Feser et al., 2001, 2003; Guttmann et al., 2001). In the Stony Brook STXM, the user can choose between proportional counter and segmented silicon detectors, and a visible

light microscope is also placed on the detector stage with X-Y-Z motorized motion so as to pre-locate desired regions of the specimen. Another approach, which works for either a TXM or a STXM, is co-indexing of off-line light microscopes with the X-ray microscope (Meyer-Ilse et al., 1994, 2001; Kilcoyne et al., 2003).

Scanning microscopes offer different characteristics than full-field imaging systems do. These include the ability to quickly change from scanning very large areas at low resolution to taking high resolution, small field scans, and reduced radiation dose because the 5–20% efficient zone plate is located upstream of the specimen rather than downstream. Because of the need to mechanically scan the specimen in most present microscopes, and the need for coherent illumination, imaging times are generally longer (in the range of one or a few minutes, rather than seconds in the case of many TXMs). At the same time, the requirement for coherent illumination means that the etendue or phase space that the monochromator must accept is greatly reduced, so that aberrations are reduced and it is relatively easy to obtain very high spectral resolution. These characteristics make scanning transmission X-ray microscopes especially well suited to low-dose spectromicroscopy applications, as will be described below.

Phase contrast has historically seen less use in scanning transmission X-ray microscopes. However, refractive and diffractive effects by the specimen lead to a redistribution of signal on the detector which can be interpreted to give phase contrast images (as will be discussed below). The ultimate approach is to use a 2D detector (such as a CCD camera) to detect the entire intensity distribution at each pixel of a scanned image; Chapman has used this in an impressive demonstration of Wigner deconvolution microscopy to recover the phase and magnitude distribution of the specimen as well as the zone plate objective (Chapman, 1996a), while Morrison et al. have used this to obtain first moment images which reveal the dominant phase gradient at each pixel location (Eaton et al., 2000; Morrison et al., 2002). Coupled with the potential power of these approaches are significant challenges: the readout time of large pixel detectors is often not in the few or sub millisecond pixel timescale required for fast scanning, and the resulting 4D data files are quite large. More fundamental is the question of statistical significance in each pixel of a large array detector when radiation dose to the specimen must be considered; in some cases it may be preferable to divide a weaker signal into fewer detector segments. This approach has been used by Feser et al. (Figure 13–18) who have used a detector with only 8 segments to obtain quantitative phase contrast images while operating at per-pixel acquisition times of a few milliseconds and producing data files of manageable size (Feser et al., 2003) (see Figure 13–18). Additional approaches to obtaining phase contrast in scanning microscopes include the use of zone plate doublers (Kaulich et al., 2002) or phase modifiers (Polack et al., 2000) to produce differential interference contrast. Undoubtedly different experiments will involve different choices in the tradeoff of the fineness of segmentation of scanning microscope detectors, but in any case

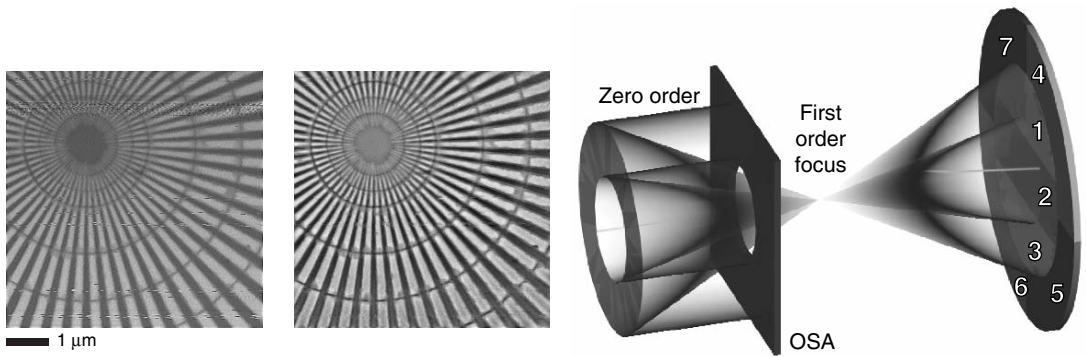


Figure 13-18. Amplitude (left) and phase (right) contrast images of a germanium test pattern imaged using a scanning transmission X-ray microscope with a segmented detector. Also shown is a schematic view of the detector. The undeflected beam from the first order focus is directed into bright field segments 1, 2 and 3 (these segments also allow differential interference contrast). The deflected beam is detected in the angular segments 4, 5, 6, and 7 for dark field imaging and differential phase contrast. (Reprinted from Feser et al., © 2003, with permission of EDP Sciences.)

it is clear that phase contrast plays an interesting role in scanning transmission X-ray microscopy as well as in microprobes as will be noted below.

3.1.4 Scanning Fluorescence X-Ray Microprobe (SFXM) Layout

Scanning fluorescence X-ray microprobes (SFXM) use a focused X-ray beam to stimulate the emission of characteristic fluorescence X-rays from specific elements in the specimen. When linearly polarized radiation (such as is usually obtained from synchrotron sources) is used, a fluorescence detector placed 90° to the beam in the polarization plane will detect a minimum of coherent scattering signal; this detector must then have some means of discriminating between different X-ray emission energies. Energy-dispersive detectors accomplish this by measuring the number of electron-hole pairs created by each X-ray in a semiconductor material, while wavelength-dispersive detectors use a crystal optic or a grating to separate the X-ray energies. Energy dispersive detectors generally have large solid angle collection, and multi-element detectors can be used to overcome the ~ 50 kHz count rate limit determined by charge readout time, while wavelength dispersive detectors offer better separation between nearby spectral lines and larger dynamic range for detecting low concentration elements amongst other fluorescing elements of higher concentration. There is a long and rich history of synchrotron-based microprobes (Horowitz and Howell, 1972; Sparks, 1980; Rivers et al., 1988; Thompson et al., 1988; Hayakawa et al., 1989), and a variety of optical approaches including the use of compound refractive lenses and Kirkpatrick-Baez mirror optics are now achieving submicron resolution. We outline here some of the characteristics of microprobes using zone plate optics (Barrett et al., 1998; Yun et al., 1998a; Suzuki et al., 2001; Kamijo et al., 2003) by

considering the example of the 2-ID-E microprobe at the Advanced Photon Source at Argonne National Laboratory near Chicago.

This microprobe operates using a side-deflecting crystal monochromator to transfer an off-axis part of the central cone produced by a hard X-ray undulator. In the vertical direction, the objective zone plate images the source directly onto the specimen, while in the horizontal direction the variable width monochromator exit slit is imaged. Astigmatism effects are avoided in the resulting focused beam by the fact that the depth of focus is much larger than the difference between the positions of the horizontal and vertical foci of the zone plate (in addition, the zone plate can be tilted to compensate for more severe source astigmatism). Zone plates of diameter 160–320 μm and outermost zone width of 100 nm are typically used, giving focal lengths of 12–25 cm at 10 keV. While the probe size can be as small as 150 nm, a larger horizontal source size is often chosen to give more flux at the cost of resolution. The specimen is mounted at 15° to the incident beam to provide access to both the incident X-ray beam and the fluorescence detector, and it is scanned by motor-driven stages with 0.1 μm step size. A multi-element germanium fluorescence detector is used to collect the fluorescent signal; one can either record the signal in a limited number of pre-defined energy windows for rapid analysis with modest data file size, or record the full fluorescence spectrum per pixel for improved quantitation of elements with closely spaced fluorescence energies. The region consisting of the specimen and detector is located inside a glovebox which can be purged with helium to eliminate fluorescence from argon in air which would otherwise obscure a number of low-Z elements, and to reduce the absorption of low-Z fluorescence signals by air. Per-pixel dwell times are on the order of one second, so that the experimenter must be judicious in the choice of scan area (the use of common position indexing between a visible light microscope and the microprobe aids in rapid specimen location).

Zone plates of diameter 50–100 μm and outermost zone width of 100–300 nm are typically used, giving focal lengths of 5–30 mm. While the probe size can be as small as the Rayleigh resolution of 1.22 times the outermost zone width, a larger virtual source size is often chosen to give more flux at the cost of resolution. The specimen is mounted at 45° to the incident beam to provide access to both the incident X-ray beam and the fluorescence detector, and it is scanned by motor-driven stages with 0.1 μm step size. A multi-element germanium fluorescence detector is used to collect the fluorescent signal; one can either record the signal in a limited number of pre-defined energy windows for rapid analysis with modest data file size, or record the full fluorescence spectrum per pixel for improved quantitation of elements with closely spaced fluorescence energies. The region consisting of the zone plate, specimen, and detector is all located inside a glovebox which can be purged with helium to eliminate fluorescence from argon in air which would otherwise obscure a number of low-Z elements. Per-pixel dwell times are on the order of one second, so that the experimenter must be judicious in the choice of scan area (the use of common position index-

ing between a visible light microscope and the microprobe aids in rapid specimen location).

Trace element mapping by fluorescence detection with sensitivities down to about 100 parts per billion, or about 10^{-17} grams of iron within a $(200\text{ nm})^2$ spot, represents the majority of microprobe applications. However, X-ray microprobes can be used in a number of other ways as well, including measurements of crystal strain in small regions (Rebonato et al., 1989; Cai et al., 1999; Soh et al., 2002) and differential-aperture measurements of microstructure and strain (Larson et al., 2002). The phase contrast methods described above for STXM are equally applicable in SFXM, and offer a much-needed way to image the overall mass and ultrastructure of specimens while simultaneously forming trace element or strain maps.

3.2 Fundamentals of Contrast in the TXM

It is useful to have an analytical treatment that provides insight into the way a microscope produces contrast and at the same time allows simple calculations to assess experimental plans. This was provided by Rudolph and coworkers (1990) in a form that allows amplitude-contrast, Zernike-phase-contrast and dark-field imaging, to be included in a unified description, that is largely independent of the microscope design. Assuming only that we have an *imaging* microscope, we consider first the Zernike phase-contrast TXM.

We are interested in the contrast C or the contrast parameter Θ (see Section 1) between an interesting feature F and a background feature B generated via the phase shifter S . F and B are defined to have the same thickness but in reality the background material (water for example) may be thicker than the feature so we allow for that by adding a layer L . If we define the complex transmission factors of F , B , S and L as $a_F \cdot p_F \equiv [\exp\{-2\pi\beta_F t_F/\lambda\}] \cdot [\exp\{2\pi i\delta_F t_F/\lambda\}]$ etc where $1-\delta-i\beta$ is the refractive index and t is the thickness, then following Rudolph et al. (1990) we can obtain the image intensities (I_F and I_B) and thence C and Θ

$$C = \frac{(I_F - I_B)}{(I_F + I_B)}, \quad \Theta = \frac{(I_F - I_B)}{\sqrt{I_F + I_B}}$$

$$I_F = \left\{ a_B^2 a_S^2 + 2a_F a_B a_S \operatorname{Re}[p_F p_B^* p_S^*] - 2a_B^2 a_S \operatorname{Re}[p_S^*] \right. \\ \left. + a_F^2 - 2a_F a_B \operatorname{Re}[p_F p_B^*] + a_B^2 \right\} a_L^2.$$

$$I_B = a_B^2 a_S^2 a_L^2$$

The dose D (the energy deposited per unit mass of sample) needed to detect a feature of area d^2 , thickness $t_F = d$ and density ρ with signal-to-noise ratio S/N can now be calculated (Rudolph et al., 1990) as follows

$$D = \left(\frac{S}{N} \right)^2 \frac{hc}{\lambda \rho d^3} \frac{1 - a_F^2 a_L^2}{\Theta^2}$$

where $hc = 1240\text{ eV}\cdot\text{nm}$ represents the product of Planck's constant and the velocity of light. The above relations are convenient because, in

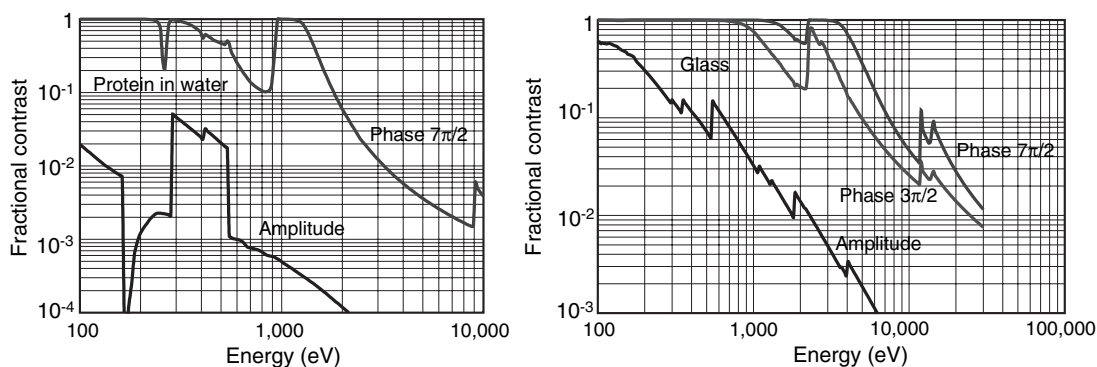


Figure 13–19. Intrinsic amplitude and Zernike phase contrast for two types of sample of thickness 30 nm relative to a background material of the same thickness: protein in water (left) and vacuum in glass, representing a crack (right). The protein is modeled assuming a density of 1.35 g/cm³, composition of H₅₀C₃₀N₉O₁₀S and a phase ring made of copper. Glass is modeled assuming a density of 2.5 g/cm³, composition Si₁₆Na₁₂K₁Ca₇Mg₆P₁O₅₇ and a phase ring made of gold. It is noteworthy that the phase contrast can be much greater than the amplitude contrast even in the 290–540 eV water window.

addition to phase contrast, they also describe amplitude-contrast ($t_s = 0$) and dark-field ($t_s = \text{large}$) experiments. The formula for D yields dose plots like Figure 13–5 and also tells us that the number of X-rays (of energy E) per unit area required to make the measurement with the given resolution and signal-to-noise ratio is $D\rho/(\mu E)$, where μ is the X-ray absorption coefficient.

In the multi-keV X-ray energy range, the phase contrast is substantially larger than the absorption contrast for suitable choices of the thickness of the phase shifter. The best result is typically achieved by attenuating the direct beam by the phase plate so that its amplitude is comparable to that of the scattered signal, resulting in an interference of two beams of similar amplitude. The available choices of phase plate thickness to optimize this are positive phase contrast (phase shift = $\pi/2$, $5\pi/2$, ...) or negative phase contrast (phase shift = $3\pi/2$, $7\pi/2$, ...). Figure 13–19 shows contrast plots of some of these possibilities. Although these plots are useful for providing comparative information, they represent a considerable idealization; the phase shift is assumed to be applied to 100% of the undiffracted light and 0% of the diffracted light, the thickness of the phase shifter is chosen, at each energy, to give the stated phase shift and the optical system is assumed 100% efficient. Under these assumptions, the dark-field contrast is identically equal to one. This suggests that dark-field has a dose advantage that will be dependent on the practical value of the nominally zero signal due to the undiffracted light and on the strength of the dark-field signal (Chapman et al., 1996c; Vogt et al., 2001a).

3.3 Partial Coherence

3.3.1 History

The resolution of microscopes, including X-ray microscopes, depends on the angular widths of the light beams delivered to, and collected

from, the sample. The analysis of this effect was pioneered in the 1950s by Hopkins, Wolf and others and was part of a movement to apply the linear-systems ideas, widely used by the engineering community, in the optical arena. This work has been reviewed by Hopkins (1957), Thompson (1969), and in various texts (Wilson and Sheppard, 1984; Goodman, 1985; Born and Wolf, 1999). The main point is that the finest features (highest spatial frequencies) in the sample diffract the illuminating beam by the largest angles θ . The best geometry to include such large deflection angles is therefore one that has a wide-angle beam both inward to, and outward from, the sample. This implies broadly that in a TXM, a large-area source providing *spatially incoherent* illumination gives better resolution than a point source giving *coherent* illumination (Figure 13–20), although such a comparison is not as simple as it sounds (Goodman, 1968). It would be wrong to conclude from this that STXMs which use coherent illumination have intrinsically worse resolution than TXMs. In fact, scanning microscopes with large area detectors and transmission microscopes with large illumination angles both deliver incoherent bright-field images with the same resolution and transfer function provided only that they use objective lenses of equal resolution.

The first application of linear-systems concepts in X-ray microscopy was in the analysis of STXM images (Jacobsen et al., 1991; Zhang et al.,

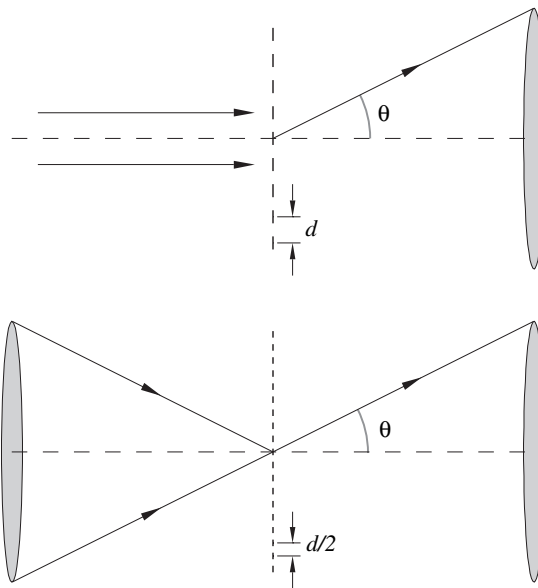


Figure 13–20. Schematic showing why the transfer function for incoherent imaging extends to twice the spatial frequency of coherent imaging for a given optic numerical aperture. For coherent imaging (top) the marginal ray is deviated by an angle θ due to diffraction by the sample periodicity d . For incoherent imaging (bottom) some light is deviated by 2θ due to the sample periodicity $d/2$. Both TXM and STXM with large area detector deliver incoherent bright-field images. (After Jacobsen et al., 1992b.)

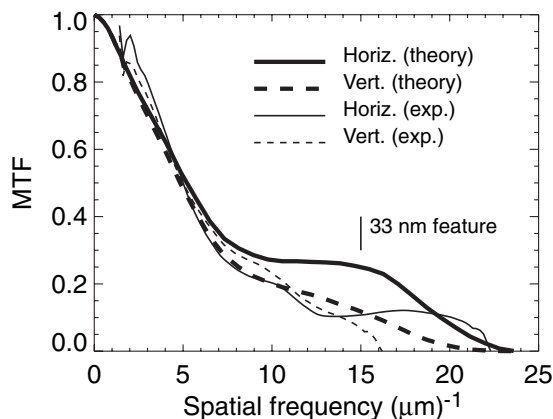


Figure 13–21. Measured and calculated modulation-transfer function for STXM imaging with a nickel zone plate of outer zone width 45 nm and diameter 100 nm. The calculated curve was derived from the known zone plate aperture function and the size and distance of the source pinhole. (Reprinted from Jacobsen, © 1991, with permission from Elsevier.)

1992) in which the intensity point spread function and its Fourier transform, the optical transfer function (OTF) were calculated. In fact, the magnitude of the OTF, known as the modulation transfer function (MTF), was both calculated and measured for the Stony Brook STXM and good agreement was obtained (Figure 13–21). Similar analysis has been provided for TXMs (Jochum and Meyer-Ilse, 1995; Niemann et al., 2000). Jochum and Meyer-Ilse provided a fairly general treatment of the application of coherence theory to X-ray microscopy including imaging of two-point and step objects by a realistic TXM in bright-field amplitude contrast. Other discussions of coherence issues have been provided by (Heck et al., 1998; Chao et al., 2003).

3.3.2 Fourier Optics Treatment

Partially coherent imaging by a microscope can be described generally by the methods of Fourier optics (Wilson et al., 1984; Goodman 1985; Born and Wolf, 1999). This method uses a real-space and a frequency-space description of waves in which frequencies (u) are closely related to directions (θ) according to the general (1D) relation $u = \sin \theta / \lambda$. Following Chapman et al. (1996c), we consider first a STXM with amplitude point spread function $h(\mathbf{x})$ imaging a sample of amplitude transparency $t(\mathbf{x})$. Using capital letters to represent Fourier transforms, the pupil function of the lens is $H(\mathbf{u})$ where $\mathbf{u}$ is the general frequency coordinate, that is conjugate to the object-plane spatial coordinate $\mathbf{x}$ and has a maximum value of NA/λ where NA refers to the beam-limiting lens. Any point in the lens pupil or the detection plane can be represented by a $\mathbf{u}$ value. Since the detector in a STXM is placed in the far field of the X-ray focal spot, the diffraction pattern formed in the detection plane in the absence of a sample will be $H^*(-\mathbf{u})$. When

the sample is present and the spot is at $\mathbf{x}_s$, the wave field immediately behind the sample is $h(\mathbf{x})t(\mathbf{x} - \mathbf{x}_s)$ and the field in the far-field detection plane is given by the Fourier transform of that. The detected intensity is therefore

$$F(\mathbf{u}, \mathbf{x}_s) = |H(\mathbf{u}) \otimes_{\mathbf{u}} T(\mathbf{u}) e^{2\pi i \mathbf{x}_s \cdot \mathbf{u}}|^2 \quad (33)$$

where $\otimes$ represents convolution and the convolution and shift theorems have been used. The same quantity $F(\mathbf{u}, \mathbf{x})$ can also be represented in another useful way. By inserting the representations of $H(\mathbf{u})$ and $T(\mathbf{u})$ as Fourier integrals into the convolution integral Eq. (33) and using the Fourier-integral definition of the delta function (Born and Wolf, 1999) and then its sifting property, we obtain (Chapman et al., 1996c)

$$F(\mathbf{u}, \mathbf{x}) = |h(\mathbf{x}) \otimes_{\mathbf{x}} t(\mathbf{x}) e^{-2\pi i \mathbf{x} \cdot \mathbf{u}}|^2, \quad (34)$$

The first of the above two equations represents the diffraction pattern formed in the detection plane by a STXM at each scan position as $F(\mathbf{u}, \mathbf{x}_s)$, regarded as a function of $\mathbf{u}$ for a given $\mathbf{x}_s$. The second equation represents a coherent image in a TXM, for illumination direction $\mathbf{u}$, as $F(\mathbf{u}, \mathbf{x})$, regarded as a function of $\mathbf{x}$ for a given $\mathbf{u}$. In the first case the exponential represents the scan shift $\mathbf{x}_s$ and in the second case it represents the incoming plane wave at direction $\mathbf{u}$. This optical equivalence of the STXM and TXM is known as “reciprocity” and is discussed further in Section 3.3.4. To get the delivered intensity image $I(\mathbf{x})$ in either case one has to integrate the signal in the detection plane over the particular distribution of $\mathbf{u}$ values that are used. That is in STXM we integrate over the intensity response function of the detector $|D(\mathbf{u})|^2$ while in TXM we similarly integrate over the intensity distribution in $\mathbf{u}$ delivered by the source $|S(\mathbf{u})|^2$.

$$\begin{aligned} I_{\text{STXM}}(\mathbf{x}) &= \int_{\text{DET}} F(\mathbf{u}, \mathbf{x}) |D(\mathbf{u})|^2 d\mathbf{u} \\ \text{and} \\ I_{\text{TXM}}(\mathbf{x}) &= \int_{\text{SOURCE}} F(\mathbf{u}, \mathbf{x}) |S(\mathbf{u})|^2 d\mathbf{u}. \end{aligned} \quad (35)$$

If the condenser is approximately incoherently illuminated (as specified in §10.5.1 Eq. 13 of (Born and Wolf, 1999) for example), which is often the case for TXMs, (Schneider, 1998; Vogt et al., 2001a), then the effective source (Hopkins, 1957), $|S(\mathbf{u})|^2$, will be the condenser lens aperture function. The expression for the fully incoherent bright field image ($|D(\mathbf{u})|^2 = 1$ or $|S(\mathbf{u})|^2 = 1$) is the same in both TXM and STXM and is obtained by inserting Eq. (33) into Eq. (35) and applying Parseval’s theorem (Chapman et al., 1996c)

$$I_{\text{BF}}(\mathbf{x}) = |h(\mathbf{x})|^2 \otimes_{\mathbf{x}} |t(\mathbf{x})|^2 \quad (36)$$

The coherent bright field image is available from a STXM by using an axial point detector and from a TXM by using an axial point source. Both are given by $F(0, \mathbf{x})$ although neither is widely used in X-ray microscopy. The process of integrating over S or D , which is carried out automatically by the hardware of the TXM or STXM, is generally convenient but it destroys potentially useful information about the

sample. A procedure for capturing this information, by storing the full detection-plane pattern at every pixel position of the STXM image, has been described and implemented to obtain phase- and amplitude-contrast images by Chapman et al. (1996a). The speed of the procedure was limited by the speed of early 1990s computers but, given the improvement of computers since then, it may well be time to re-examine this approach (see Section 3.1.3). Eq. (34) and Eq. (36) show respectively that coherent imaging is linear in the amplitude and incoherent imaging is linear in the intensity. On the other hand, as we see below, *partially coherent* imaging is not linear in either.

3.3.3 Contrast Transfer

In the case that we do not have $|D(\mathbf{u})|^2 = 1$ or $|S(\mathbf{u})|^2 = 1$, the above procedure used to obtain Eq. (36) does not lead to such a simple result but rather to the following expression representing partially coherent imaging in a STXM (Kintner et al., 1978; Wilson and Sheppard, 1984; Born and Wolf, 1999).

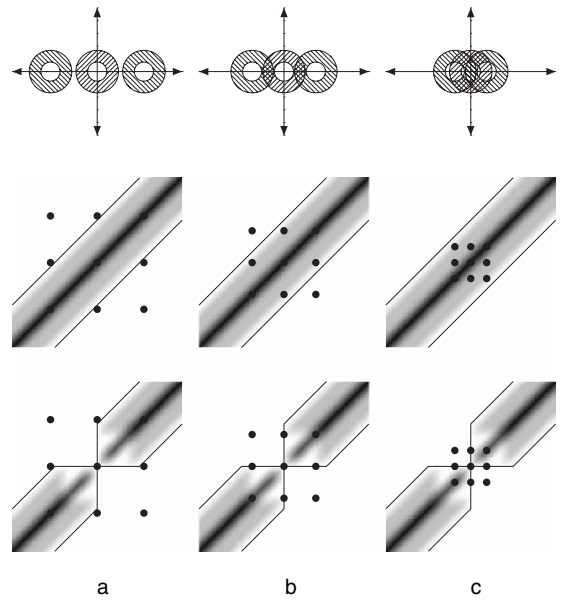
$$I(\mathbf{x}) = \int \int \int_{-\infty}^{+\infty} C(\mathbf{m}; \mathbf{p}) T(\mathbf{m}) T^*(\mathbf{p}) e^{-2\pi i[(\mathbf{m}-\mathbf{p}) \cdot \mathbf{x}]} d\mathbf{m} d\mathbf{p} \quad (37)$$

$$C(\mathbf{m}; \mathbf{p}) = \int \int_{-\infty}^{+\infty} |D(\mathbf{u})|^2 H(\mathbf{u} - \mathbf{m}) H^*(\mathbf{u} - \mathbf{p}) d\mathbf{u} \quad (38)$$

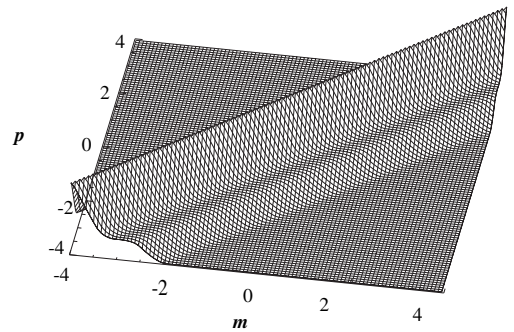
For a TXM S replaces D in the last equation. The integration variables $\mathbf{m}$ and $\mathbf{p}$ in Eq. (37) are frequencies similar to $\mathbf{u}$ but $\mathbf{m}$ represents a ray incident on the sample while $\mathbf{p}$ represents a ray emerging from it. The ranges of frequencies included in these beams by the form of S or D determine the range of periodicities $(\mathbf{m}-\mathbf{p})$ in the sample that contribute to the image and thus determine the extent of the MTF in frequency space. The function $C(\mathbf{m}; \mathbf{p})$ is known in optics as the transmission cross coefficient (Born and Wolf, 1999) or the partially coherent transfer function (Wilson and Sheppard, 1984) and provides a *sample-independent* description of the effect of both the illumination and the optical system on the transfer of information from object to image. It is not a true transfer function, since the transfer is not linear, but is a member of a wider class of “bilinear transfer functions.” Such functions are described, for example, by (Saleh, 1979) and have been applied to partially coherent X-ray imaging by Vogt et al. (2001a).

$C(\mathbf{m}; \mathbf{p})$ is widely used in the optical and electron microscopy communities and its properties have been worked out in detail; see for example (Sheppard and Wilson, 1980; Wilson and Sheppard, 1984). It is normally a 4D function but in the case of a 1D object it becomes the 2D function $C(m;p)$. The value of $C(m;p)$ is then equal to the overlap integral of the three appropriately shifted aperture functions in the integrand of (38) (Kintner et al., 1978; Wilson and Sheppard, 1984; Born and Wolf, 1999). For many cases of interest in both TXM and STXM, all three are circular disks or annuli (Figures 13–22 and 13–23). For the ideally incoherent bright-field image the value of D or S is taken to be unity for all frequencies and the overlap then depends only on the

Figure 13–22. Imaging of a one-dimensional cosine amplitude grating object using a STXM. The top row shows the signal in the detection plane. The center row shows gray-level plots of $C(m, p)$ in m - p space overlaid with the support boundary of $C(m, p)$ (solid lines) and the spectra $T(m)T^*(p)$ as in Eq. (37) (spots). The middle row shows bright field and the bottom row dark field. The three columns correspond to grating frequencies that are (a) >2 , (b) between 1 and 2 and (c) <1 expressed in units of the maximum values of m and p which are both equal to NA/λ . (Reprinted from Chapman et al., © 1996c, with permission from Elsevier.)



a



b

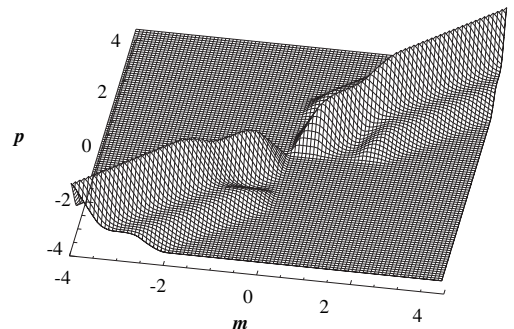


Figure 13–23. Bright-field (a) and dark-field (b) partially coherent transfer functions for a 1D object and an annular lens with inner radius equal to 0.44 times the outer (after Chapman et al., 1996c). The function $C(m, p)$ is plotted against m and p , expressed as multiples of their maximum value NA/λ . (Reprinted from Chapman et al., © 1996c, with permission from Elsevier.)

difference $m-p$ of the shifts of H and H^* . That is, there is only *one* response to the sample frequency $s = m - p$ irrespective of m , which indicates a linear system with MTF equal to $C(m-p;0)/C(0;0)$. For forms of D or S corresponding to partial coherence, the system is not linear. On the other hand a dark-field configuration must have detector (or source) and lens aperture functions which have zero overlap at $m = p = 0$. For example D or S could be the Babinet inverse of H . For circular functions of the latter type $C(m;p) = 0$ if $\text{sign}(m) \neq \text{sign}(p)$. Examples of both bright- and dark-field transfer functions for aperture geometries that are representative of a STXM and that show the above characteristics are given by Chapman et al. (1996c) (Figure 13–23). The response of the same systems to a grating-like object are also given (Figure 13–22). Dark-field STXM is particularly well suited to imaging samples with small features such as gold labels that scatter by large angles (Chapman et al., 1996b). The procedures outlined above allow the calculation of the MTF and the resolution behavior of both types of X-ray microscope based on a knowledge of the resolution-determining lens and the geometry of the source or detector. It is noteworthy that, as in other types of microscope, the resolution does not depend on aberrations of the condenser if there is one. As noted in Section 3.1.2 and illustrated in Figure 13–17, the placement of a ring aperture and phase ring to get Zernike phase contrast in a TXM may be modeled as modifications of the source and lens aperture functions. By this means the above method of analysis may be applied to this case as well (Mondal and Slansky, 1970; Sheppard and Wilson, 1980; Morrison, 1989a).

3.3.4 Reciprocity

The general conclusion of the above analysis is that the optical systems of the TXM and STXM are the same with the position of the lens, before or after the sample, interchanged and the role of the source and detector interchanged. This is the “reciprocity” relationship (Zeitler and Thomson, 1970) that has long been recognized in the visible-light and electron imaging communities and has been explained in the context of X-ray imaging by Morrison (1989a; Morrison et al., 2002). Thus we might expect that, given identical resolution-determining lenses, a TXM and a STXM (both operating in incoherent bright-field mode) could equally well utilize wide-angle beams and get good resolution. For TXM the requirement would be that the condenser should deliver a wide angle to the sample and for STXM that the detector should collect a wide angle from the sample. However, in the past, the practical realization of a wide-angle condenser for a TXM has been much harder than a wide-angle detector for a STXM as we discussed in the condenser zone plate section above.

In practice the TXM/STXM relationship is not quite as symmetrical as the above account suggests because of the general use of objective zone plates with a central stop for STXM (Section 3.1.3) but not for TXM. The stop produces a point-spread function, which has a narrower central peak but larger side lobes. As a consequence the frequency response (the MTF, see Figure 13–21) is increased in the high-frequency and decreased in the low-frequency region.

3.3.5 The Influence of Coherence on Resolution

Calculations of the transfer function as an overlap area of three aperture functions in the integrand of Eq. (38) were discussed in Section 3.3.3. For standard TXM, these functions are circular and two of them are the same. One can therefore follow (Hopkins, 1957) and characterize the illumination by a coherence parameter σ defined by the ratio of the condenser and objective numerical apertures or $\sigma = NA_c/NA_o$. Full coherence is represented by $\sigma = 0$ and full incoherence by $\sigma = \infty$, although $\sigma = 1$ is usually sufficient to get close to fully incoherent behavior. We start by considering the modulation (the percent dip in the valley between the peaks) for a two-point object with separation $0.61\lambda/NA_o$. The modulation is 26.5% for incoherent illumination (Born and Wolf, 1999) and according to the Rayleigh criterion, the two points are just resolved. It is common practice (Born and Wolf, 1999) to extend the Rayleigh criterion to other pairs of objects and define them as resolved if the modulation is at least 26.5%. An example is the two-point object with in-phase coherent illumination for which the just-resolvable separation is $0.82\lambda/NA_o$. Further detail of this can be seen from the plots in Figure 13–24 (Jochum and Meyer-Ilse, 1995). An important example for X-ray imaging is the modulation due to a periodic object, in particular a square-wave transmission object. Such an object can either be prepared by standard lithography methods (Jacobsen et al., 1991) or, for finer line widths, by preparing thin cross-sections of synthetic multilayers (Chao et al., 2003), to yield resolution test patterns for an X-ray microscope. If the resolution is defined as the half period of the finest square wave that can be imaged with 26.5% modulation and is expressed as $k_1\lambda/NA_o$, then according to (Chao et al., 2005), the diffraction-limited value of k_1 is 0.5 for a coherent system ($\sigma = 0$) and 0.4 for $\sigma = 0.38$ (the actual value for XM-1 illuminating the 15 nm zone plate). Thus the diffraction-limited resolution of their experiment was $0.8\Delta r_n$ while the

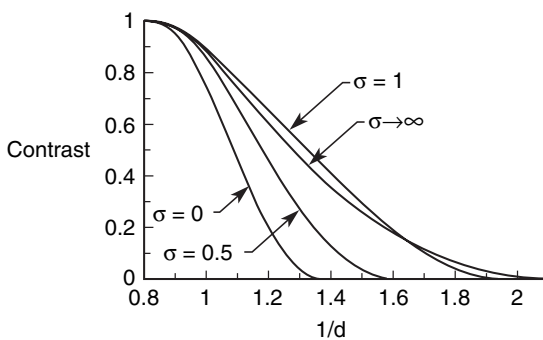


Figure 13–24. Image contrast as a function of $1/d$ where d is the point separation of a two-point object imaged by a lens with a circular pupil and coherence parameter σ equal to 0 (coherent case), 0.5, 1 and infinity (incoherent case). d is expressed in units of λ/NA . Note that the Rayleigh resolution corresponds to 15.3% intensity contrast (defined as $(I_{\max} - I_{\min})/(I_{\max} + I_{\min})$), which is the same as 26.5% modulation. It occurs at $d = 0.61\lambda/NA$ for both $\sigma = 1$ and $\sigma \rightarrow \infty$. (Reprinted from Jochum and Meyer-Ilse, © 1995, with permission from Optical Society of America.)

achieved value was about $1.0\Delta r_n$ (<15 nm). The data demonstrating the achieved resolution are shown in Figure 13–13. By reciprocity, the same arguments about the value of the diffraction-limited resolution would apply for a STXM in incoherent bright field using the same zone plate. With a large enough detector one would achieve $\sigma = 1$ and the same diffraction-limited resolution as a TXM with $\sigma = 1$.

3.3.6 Coherence in Zernike Phase Contrast

We return now to the question of the choice of width and radius for the two rings in a Zernike phase-contrast configuration of the TXM, or, equivalently, how much coherence is desirable in this case. This choice has been discussed by (Mondal and Slansky, 1970) and is essentially a trade off between light collection and the distorting effects of the fact that the phase ring must have finite area, which applies an unintended phase change to a certain portion of the diffracted light causing the so-called halo effect (Wilson and Sheppard, 1981). It is generally thought that one needs very little coherence, that is, the ring aperture can leave a large fraction of the condenser area open. This is true if the requirement is merely to make otherwise invisible phase features, especially phase jumps, become visible. However, with low coherence, a phase step is rendered as a double-peaked zero-crossing function and a phase rect function is rendered as two such double peaks. How much coherence do we need to get anything resembling a faithful rendition of the object? We have not found much attention to this point in the literature but the treatment by Martin (1966) provides an answer, which is confirmed by our own computer modeling. To get a rendition of a rect function that looks like the original function one needs to have the coherence width $w_c = \lambda f_{\text{cond}} / \Delta r_{\text{ring}}$ of light arriving at the sample at least equal to the width of the rect function.

3.3.7 Propagation-Based Phase Contrast

Another way to achieve phase contrast is to exploit the $\exp[i\pi(x^2 + y^2)/\lambda z]$ phase shifts that occur in the (x, y) plane as a result of the propagation of a coherent wavefield through a distance z . This is exploited in X-ray holographic microscopy which has had many successes (Aoki and Kikuta, 1974; Joyeux, et al. 1988; Jacobsen et al., 1990; McNulty et al., 1992; Snigirev et al., 1995; Lindaas et al., 1996; Eisebitt et al., 2004) but which is so far not used for routine X-ray imaging. The exception is in the use of holography for phase contrast tomography at higher X-ray energies, where Cloetens and coworkers have achieved considerable success in routine micrometer-resolution tomography using a phosphor/lens/CCD detector system (Cloetens et al., 1999). While it lies beyond the scope of the present article's emphasis on zone plate X-ray microscopy, this unique approach is providing impressive 3D reconstructions of difficult specimens including foams.

3.4 Tomography in X-Ray Microscopes

3.4.1 Principle of Operation

As was noted in Section 2.4.5 on hard X-ray zone plates, the transverse resolution of a zone plate operated in first diffraction order is given by

$0.61\lambda/\text{NA} = 1.22\Delta r_n$, and the depth of focus is $2\lambda/(\text{NA})^2 = 8(\Delta r_n)^2/\lambda$. Since present zone plates have outermost zone widths Δr_n that are much larger than the wavelength λ , this means that the depth of focus is necessarily large compared to the resolution as can be seen by the illustration of the 3D modulation transfer function in Figure 13–25. This provides an opportunity for 3D imaging if the object is smaller than the depth of focus because a 2D image can then be interpreted as providing a simple projection through the specimen, which is precisely what conventional tomography requires at each viewing angle. Tomography with electron microscopes is long established, and following earlier demonstrations by Haddad et al. (1994) using the Stony Brook STXM and Lehr (1997) using the Göttingen TXM, a number of groups are now using X-ray microscopes for tomographic studies of frozen hydrated biological specimens and integrated circuits, among other applications that will be described below in Section 4. However, although the technique used in these studies is improving, they have still not reached the resolution achieved by the same microscopes in 2D.

3.4.2 The Depth-of-Focus Limit

Given the successes achieved and the amount of current interest, what are the issues to be faced in improving the resolution of tomography

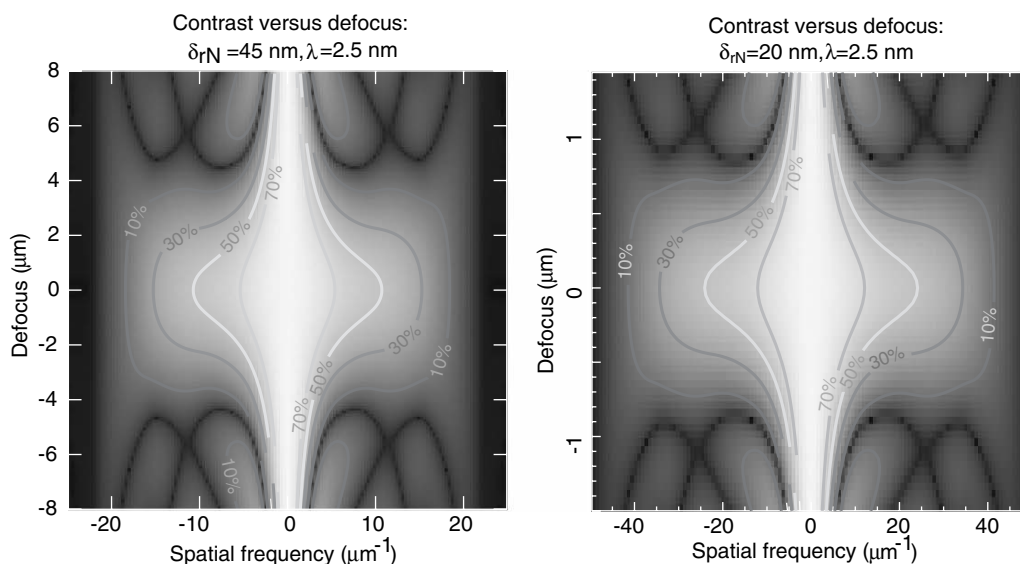


Figure 13–25. Properties of soft X-ray tomography using zone plate optics. At the left is shown the 3D modulation transfer function for monochromatic, spatially incoherent bright-field imaging with a 45-nm-outer-zone-width zone plate with a half-diameter central stop, as a function of depth. One can see that the good in-focus frequency response is preserved over a total depth of about $8\mu\text{m}$. This is useful for many tomography experiments that rely on the fact that the delivered image is a *projection* of the object. At the right is the same information for a 20-nm-zone plate; as can be seen, the figure scales quite well from the figure at the left according to the square of the ratio of the finest zone widths. With a 20-nm-zone plate and monochromatic illumination, good frequency response is only preserved over a depth of about $0.5\text{--}1\mu\text{m}$ which is much more restrictive, illustrating the challenges of improved resolution in TXM zone-plate tomography. (Reprinted from Wang et al., © 2000, with permission from Blackwell Publishing.)

in X-ray microscopes? One of them concerns the same depth of focus that makes such tomography straightforward. To our knowledge all demonstrations of soft X-ray tomography reconstructions have used zone plates with outermost zone width Δr_n no finer than 35 nm, so that the depth of focus in the water window region is at least 4 μm which has been comparable to the specimen size. As higher resolution zone plates are employed, the depth of focus will decrease as the square of improvements in transverse resolution, so that a 15 nm outermost-zone-width zone plate would have a depth of focus of about 0.8 μm . This approaches the $\sim 0.5 \mu\text{m}$ thickness accessible to cryo electron tomography of frozen hydrated cell regions at 6–8 nm resolution (Grimm et al., 1998; Medalia et al., 2002) and thus would seem to leave a much-reduced “window of opportunity” for soft X-ray tomographers and apparently deny them the important goal of high-resolution, 3D imaging of intact eucaryotic cells.

The depth-of-focus limit evidently poses a considerable challenge to those who would exploit zone plate X-ray microscopes for tomography. Accordingly we discuss in the sections that follow several different approaches to dealing with the challenge.

3.4.3 Avoiding the Depth-of-Focus Limit by Reconstructing a Partial Volume

Zone-plate-microscope X-ray tomography as described above is based on geometrical optics and is mathematically identical to the technique; Computed Axial Tomography (CAT). Therefore, following an established CAT procedure (see for example Ritman et al., 1997), it is possible to reconstruct partial volumes that are embedded in a larger overall object. This has been done in the X-ray microscope context by Xradia Inc. In this case the depth-of-focus limit applies to the partial volume, rather than the whole volume, and the signals generated by the out-of-focus parts of the sample reduce, after combining all members of the tilt series, to a smooth background. The size of the region that is always in focus will be roughly a sphere of diameter equal to the depth of focus.

3.4.4 Avoiding the Depth-of-Focus Limit by Wide-Band Illumination

In fact, the depth of focus calculation given above is for monochromatic illumination, which is applicable to demonstrations of tomography of frozen hydrated cells in a STXM with a high-energy-resolution monochromator (Wang et al., 2000). In some existing soft X-ray TXMs using the zone plate condenser as a linear monochromator, an energy resolving power of $E/\Delta E \approx 200$ has been used for objective zone plates with $N = 375$ zones (the recommended bandwidth for such an objective would usually have been $E/\Delta E > N$, (Thieme, 1988). However, Weiss et al. (2002) have shown that use of $E/\Delta E \approx 200$ leads to significant changes of the modulation transfer function (MTF) as a function of defocus (similar results for different cases have been given by Schneider et al. (2002a)). In their calculations for a zone plate with 40 nm outermost zone width, the MTF at a spatial frequency of $15 \mu\text{m}^{-1}$ (corresponding to a spatial half-period of 33 nm) declines from about 0.28 in focus to

about 0.015 at a defocus of $4\mu\text{m}$ in the monochromatic case. When the bandwidth is increased to $E/\Delta E \approx 200$, the MTF at $15\mu\text{m}^{-1}$ is reduced in focus to only about 0.11 but at the same $4\mu\text{m}$ defocus it degrades much less to 0.08. Therefore these calculations show that, although the transverse resolution and efficiency for the collection of structural information, are both made worse by the use of high-bandwidth radiation, the depth of field is improved. (To our knowledge no calculations have yet addressed the possibly interesting question of how this tradeoff with nonmonochromatic illumination compares with the tradeoff resulting from using a lower or higher resolution zone plate). In summary, it will require further investigation to determine whether manipulation of the bandwidth can enable significant improvements to the tomographic resolution without reduction of the reconstructible sample volume.

3.4.5 Avoiding the Depth-of-Focus Limit by Through-Focus Deconvolution

One possible approach to beat the depth of focus limit is to use through-focus deconvolution as is done in light and electron microscopy. In electron microscopy, the recording of defocus image sequences is routine; each defocus provides positive and negative phase contrast at various bands of spatial frequencies along with zeroes in the transfer function, and the combination of several images can provide a complete image of the specimen (Reimer, 1984). In fluorescence light microscopy, through-focus image sequences can yield a high quality 3D image through the use of deconvolution of the 3D point spread function (Agard and Sedat, 1983; Carrington et al., 1995). However, there are important differences between these examples and the situation present in X-ray microscopy. In electron microscopy, this approach is usually applied to thin samples for which phase contrast dominates (indeed the specimen focus can be quickly estimated by looking for a minimum in image contrast). In light microscopy, the use of fluorescence means that the object is a sparse, pure-real function (incoherent emission from independent fluorescence emitters with no sensitivity to the relative phase of the illumination) so that the deconvolution can be done based on the intensity point spread function. While some form of generalized through-focus deconvolution may provide the needed breakthrough, quantitatively reliable results such as are needed for assembly into a tomographic reconstruction will have to account for the fact that biological specimens imaged at water-window energies produce both absorption and phase contrast so one will require exact knowledge of the complex bilinear transfer function of the zone plate optic and illumination system. In other words, the problem of 3D deconvolution of a strongly absorbing, optically thick, complex object with partial coherence is much more difficult than the cases in which optical sectioning is typically used at present.

3.4.6 Avoiding the Depth-of-Focus Limit by Use of Higher Energy X-Rays

The use of shorter wavelengths (higher X-ray energies) to increase the monochromatic depth of field $8(\Delta r_n)^2/\lambda$ is a guaranteed way to extend

the depth of focus. The associated questions of what will be the cost in resolution, contrast and efficiency are also now beginning to be answered favorably. This approach has been used to obtain sub-100 nm resolution tomographic reconstructions of metallic layers within thinned integrated circuits using a laboratory X-ray microscope operating at 5.4 keV (Wang, 2002) as will be described in Section 4.3. For lower density specimens, the use of hard X-rays naturally leads to the use of phase contrast which is far more dose-efficient than absorption contrast in this energy region. As shown in Figure 13–5, this enables multi-keV imaging at similar dose levels to the water window. As noted in Section 3.3.7 there have already been quite successful demonstrations of phase contrast tomography using hard X-rays (Cloetens et al., 1999) in the past though not (to our knowledge) at the sub-100 nm resolution level accessible to zone plate microscopes. Now the group at NSRRC, Taiwan have recently used their 8 KeV phase-contrast TXM with a 50 nm zone plate to produce 3D images of a microcircuit with defects at 60 nm resolution (Yin et al., 2006) (see Sections 1.4 and 2.4.4).

3.4.7 Avoiding the Depth-of-Focus Limit by Lens-Free Imaging

The challenges of achieving the highest possible resolution in 3D imaging have led to the consideration of lens-free imaging. Of course crystallography is able to obtain exquisite 3D maps of the electron density of a unit cell in a crystal by interpretation of a tilt series of Bragg diffraction patterns. In the case of a noncrystalline specimen, one obtains a continuous rather than Bragg-sampled diffraction pattern but there has been considerable recent progress in obtaining X-ray images through the application of iterative phasing algorithms to diffraction data from objects known to be limited in size (Miao et al., 2002; Marchesini et al., 2003; Williams et al., 2003; Shapiro et al., 2005; Chapman et al., 2006). At the moment the data collection time in 3D experiments of this type is 10–20 hours and 10 nm resolution images of materials-science samples in 3D and 30-nm-resolution images of biological specimens in 2D have been obtained. Moreover, a modern beam line designed specifically for this type of experiment would easily bring image acquisition times down to a convenient level. It is noteworthy that the phasing algorithms depend on use of the Born approximation which sets an upper limit to the sample size which may become significant at low X-ray energies such as those in the water window.

3.5 X-Ray Spectromicroscopy

As noted in Section 1, X-ray absorption edges arise when the X-ray photon reaches the threshold energy needed to completely remove an electron from an inner-shell orbital. At photon energies within about 10 eV of the edge, electrons can also be promoted to unoccupied or partially occupied molecular orbitals (see Figure 13–26); photons over a narrow energy range are sometimes able to excite inner-shell electrons into such orbitals, giving rise to absorption resonances. This so-called X-ray absorption near-edge structure (XANES) or near-edge X-ray absorption fine structure (NEXAFS) is highly sensitive to the

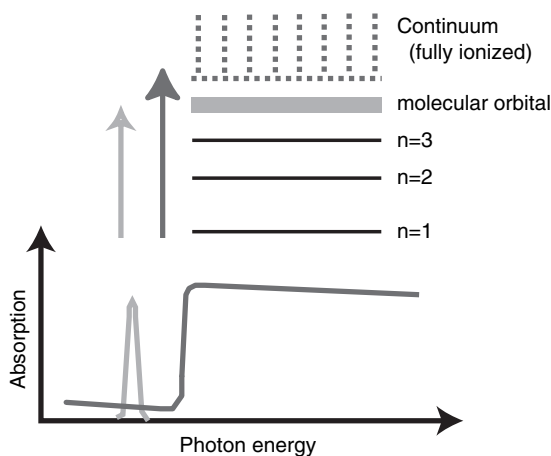


Figure 13–26. Schematic of an X-ray absorption edge, which involves the removal of an inner-shell electron, and a near-edge absorption resonance in which the electron is promoted to a partially occupied or vacant molecular orbital. These resonances are referred to as X-ray absorption near-edge structure (XANES) or near-edge X-ray absorption fine structure (NEXAFS).

local chemical bonding state of the atom in question (Stöhr, 1992) (see Figure 13–26).

One can exploit these resonances as an additional contrast mechanism in soft X-ray imaging. In electron energy loss spectroscopy (EELS), the equivalent contrast mechanism is known as ELNES for energy-loss near-edge structure and its use in energy-loss spectrum imaging (Jeanguillaume and Colliex, 1989; Hunt and Williams, 1991) is described elsewhere in this volume. Early efforts in X-ray imaging included the use of XANES resonances to enhance the sensitivity of differential absorption measurements of calcium in bone (Kenney et al., 1985), spectral imaging (King et al., 1989) and microspectroscopy in photoelectron microscopes (Harp et al., 1990), and photoelectron and transmission imaging at selected photon energies (Ade et al., 1990b, 1992). It is now common to take image sequences across X-ray absorption edges (Jacobsen et al., 2000b) yielding data sets with a full near-edge spectrum per pixel. When comparing spectrum imaging in electron versus X-ray microscopes, a few comments are in order:

- ELNES is typically done using a fixed electron energy in the range 80–200 keV. The ideal specimen thickness is under 100 nm in most cases.
- In ELNES, one gets spectroscopic information over a wide range of energies, including plasmon energies of ~ 10 eV, in a single measurement. However, plural inelastic scattering dominates the signal at higher energies (for example, electrons can lose 300 eV once, or 50 eV six times, etc.) resulting in poorer signal-to-background.
- In X-ray absorption spectroscopy, one must tune the incident X-ray energy across each absorption edge of interest. The optimum

specimen thickness of about $1/\mu(E)$ changes accordingly, so that in the ideal case one would require samples of several different thicknesses to study chemical speciation of several elements. However, X-rays suffer almost no plural inelastic scattering, which leads to improved signal-to-background.

- It is common to find scanning X-ray microscopes operating with monochromators with an energy resolution of 0.1 eV or better. Most electron microscopes have an energy resolution of 0.5–0.7 eV which leads to “blurring” of near-edge spectral features, although a limited number of higher energy resolution systems are starting to become available.
- Using XANES, one can exploit the favorable characteristics of X-ray microscopes including the ability to study hydrated specimens and/or specimens in an ambient atmosphere environment.

In X-ray microscopes, we obtain images (maps of transmitted flux I) according to the Lambert-Beer law for absorption: $I = I_0 \exp(-\mu t)$ where I_0 is the incident X-ray flux, μ is an absorption coefficient for a specific material, as discussed in Section 1.1, and t is thickness of that material. The value of $\mu(E)$ for near-edge absorption resonances can be calculated based on the electronic structure of specific molecules, and this has been employed in detailed studies via microscopy of the absorption spectra of polymers (Urquhart and Ade, 2002; Dhez et al., 2003) and amino acids (Kaznacheyev et al., 2002) (see Figure 13–27), to name two recent examples.

For a thickness t of a single material, a measurement of the transmitted flux $I(E)$ relative to the incident flux $I_0(E)$ provides a means to calculate the energy-dependent optical density $D(E) = -\ln(I(E)/I_0(E)) = \mu(E)t$. If, however, we measure the optical density not over a continuous energy range E but at some set of $n = 1 \dots N$ discrete energies E_n , we then measure

$$D_n = \mu_n t$$

for each of the $n = 1 \dots N$ photon energies. Let us next consider a mixture of $s = 1 \dots S$ different materials with partial thicknesses t_s ; our

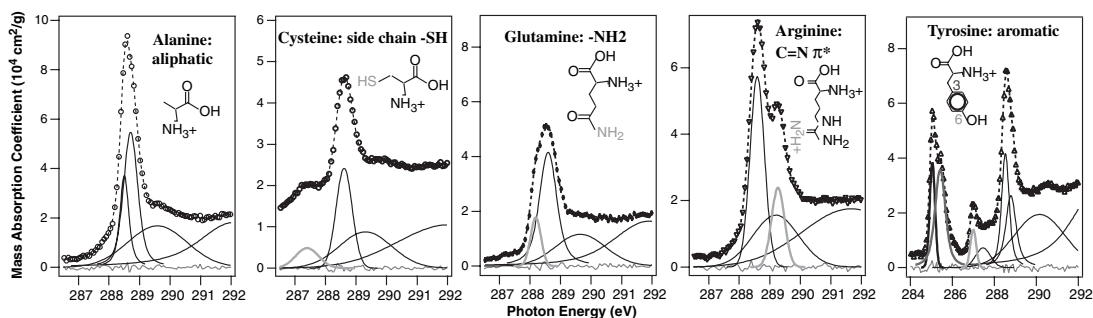


Figure 13–27. Near-carbon-edge absorption spectra of several amino acids, showing the effects of various molecular bonds in the absorption spectrum. These resonances can be used for chemical contrast in X-ray microscopy. (Reprinted from Kaznacheyev et al., © 2002, with permission from American Chemical Society.) (See color plate.)

total measurement of optical density D_n at one photon energy is given by the combined absorption of all the materials, or

$$D_n = \mu_{n1}t_1 + \mu_{n2}t_2 + \cdots + \mu_{nS}t_S.$$

Finally, if we carry out this measurement not from a single homogeneous uniform film, but from heterogeneous pixels $p = 1 \dots P$ indexed by $p = i_{\text{column}} + (i_{\text{row}} - 1) \cdot (\# \text{ columns})$ in an image, the optical density measured at one pixel p is given by

$$D_{np} = \mu_{n1}t_{1p} + \mu_{n2}t_{2p} + \cdots + \mu_{nS}t_{Sp}.$$

When all N photon energies are considered, we see that we have a data matrix D_{NP} of

$$\begin{bmatrix} D_{11} & \cdots & D_{1P} \\ \vdots & \ddots & \vdots \\ D_{N1} & \cdots & D_{NP} \end{bmatrix} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1S} \\ \vdots & \ddots & \vdots \\ \mu_{N1} & \cdots & \mu_{NS} \end{bmatrix} \cdot \begin{bmatrix} t_{11} & \cdots & t_{1P} \\ \vdots & \ddots & \vdots \\ t_{S1} & \cdots & t_{SP} \end{bmatrix}$$

or $D_{N \times P} = \mu_{N \times S} \cdot t_{S \times P}$. In other words the data represent a series of spectral signatures $\mu_{N \times S}$ and thickness maps $t_{S \times P}$.

When we acquire a series of images at different photon energies N , we are in fact measuring the data matrix $D_{N \times P}$. If we know the exact absorption spectrum μ_{Ns} for each of the $s = 1 \dots S$ components in the sample, then we can find the spatially resolved thicknesses $t_{S \times P}$ of the components by matrix inversion:

$$t_{S \times P} = \mu_{S \times N}^{-1} \cdot D_{N \times P}$$

The inversion of the matrix of spectra from all known components $\mu_{N \times S}$ can be accomplished in a robust fashion using singular value decomposition (Zhang et al., 1996; Koprinarov et al., 2002). This approach, as well as approaches which involve pixel-by-pixel least squares fits of all reference spectra, work well with specimens that involve mixtures of components that can all be measured separately. Examples using this approach are shown in the chemical imaging section of this chapter.

In many areas of research, such as biology or environmental science, the complexity of the specimen and the possibility of reactions between components means that one cannot know in advance the set of all absorption spectra $\mu_{N \times S}$ present in the specimen. In this case, one approach that has yielded recent success is to first use principal component analysis (King et al., 1989; Osanna and Jacobsen, 2000) to orthogonalize and noise-reduce the data matrix $D_{N \times P}$, and then use cluster analysis (a method of unsupervised pattern recognition) to group pixels together based on similarity of spectral signatures (Lerotic et al., 2004, 2005) (see Figure 13–34 below). This method yields a set of absorption spectra $\mu_{N \times S}$ where S now indexes the set of characteristic spectra found from the data. The power of this approach lies in its ability to improve the signal-to-noise of spectra of heterogeneous specimens by averaging noncontiguous pixels, to find even quite small regions with distinct spectroscopic signatures, and to deliver continuous “thickness” maps based on the distribution of the discovered signature spectra.

For studies at the carbon edge, one can characterize the observed set of near-edge resonances in terms of a limited number of functional group types (see e.g., Scheinost et al., 2001). While there are a number of open questions regarding this approach (for example, how many resonances should be used, with what range of allowed center photon energies, and what range of energy widths?), confidence in it can be enhanced by correlation with other spectroscopies such as solid-state nuclear magnetic resonance (Scheinost et al., 2001; Schumacher et al., 2005) and Fourier transform infrared (Solomon et al., 2005).

4 Applications

Two decades ago, nearly all research using X-ray microscopes was done by the groups that had developed the instruments. Today, most X-ray microscopes are operated as user facilities at synchrotron radiation research centers, and are used both by their developers but also by a wider community of scientists. As a result, while it was originally possible to see the major applications of X-ray microscopes in conference proceedings (Schmahl and Rudolph, 1984a; Sayre et al., 1988; Michette et al., 1992), papers in which X-ray microscopes were used to address the problem of interest now appear across a very wide array of scientific journals. In what follows, we do not presume to be exhaustive in coverage of all research using X-ray microscopes; instead, we will briefly highlight a few examples from some of the areas of present activity.

4.1 Biology

X-ray microscopes using zone plates and synchrotron radiation have been used for studies of biological specimens from the start (Niemann et al., 1976; Rarback et al., 1980), and a number of reviews have concentrated on biological applications of X-ray microscopes (see for example Kirz et al., 1995) for background information and older results, or Abraham-Peskir, (2000). One emphasis has been on high resolution imaging of whole cells at “water window” wavelengths (see Figure 13–28), including studies of human sperm (Chantler and Abraham-Peskir, 2004), malaria in red blood cells (Magowan et al., 1997), Kupffer cells (Scharf and Schneider, 1999) and COS cells (Yamamoto et al., 1998) from liver, protists (Abraham-Peskir, 1998), and chromosomes (Guttmann et al., 1992; Williams et al., 1993; Kinjo et al., 1994) among other examples. As soft X-ray microscopes push to higher spatial resolution, views through whole cells will involve a great deal of overlap of structure, but several developments offer information beyond two-dimensional images with natural contrast. One of these is to use molecular labeling methods to tag specific proteins (such as is done with great success in visible light microscopy). Several groups have demonstrated the use of gold labeling in X-ray microscopes, including detection by dark field (Chapman et al., 1996c) (see Figure 13–29) and bright field (Meyer-Ilse, 2001 et al.; Vogt et al., 2001b) approaches. One of the challenges faced thus far is that the label must

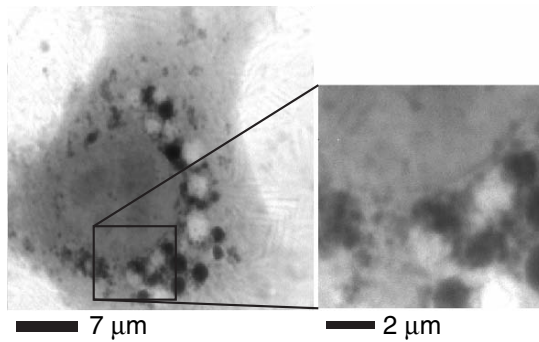


Figure 13-28. Whole fibroblast imaged in the frozen-hydrated state. The cell was cultured on a formvar-coated gold electron microscope grid, and rapidly frozen by plunging into liquid ethane. It was then imaged using a cryo STXM operated at 516 eV. In addition to this 2D image, 3D reconstructions were also obtained using tomography (Wang et al., 2000). (Reprinted from Maser et al., © 2000, with permission from Blackwell Publishing.)

be comparable in size to the resolution of the microscope for efficient detection (Chapman et al., 1996c; Vogt et al., 2001a), which means that in all studies carried out thus far the cell membrane has been permeabilized by agents such as methanol to allow relatively large labels to reach the cell's interior and this step must be preceded by chemical fixation. As a result, future improvements in X-ray microscope resolution will not only lead to improved visualization of unlabeled

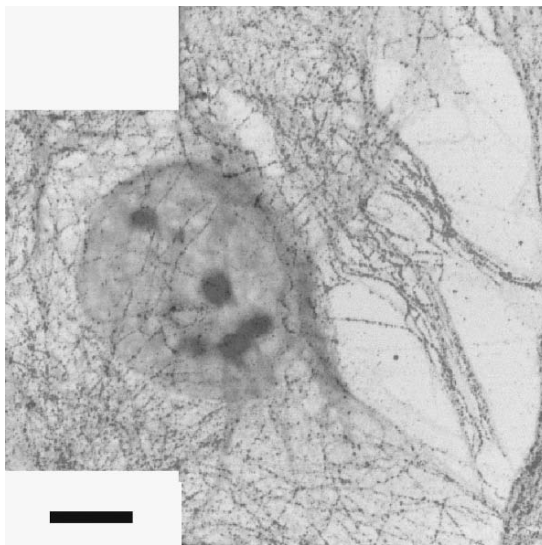


Figure 13-29. Human fibroblast with immunogold labeling for tubulin. This is a composite of two images: a bright field image (gray tones) to image overall mass, and a dark field image (red tones) to selectively image the silver-enhanced gold labels. This whole-mount cell was fixed and then permeabilized to allow for introduction of the immunogold labels, after which it was air dried. (From Chapman et al., © 1996b,c, courtesy of the Microscopy Society of America.) (See color plate.)

ultrastructure but will also make it possible to use smaller immunolabels with more “natural” preparation protocols.

Another approach to exploit the characteristics of X-ray microscopes is to go beyond two-dimensional imaging. One approach is to use XANES spectromicroscopy for mapping chemical speciation in bacteria and cells (Ito et al., 1996; Zhang et al., 1996; Lerotic et al., 2005) and biomaterials (Hitchcock et al., 2002) using the approaches outlined in Section 3.6 above. Another involves the use of tomography as has been discussed in Section 3.5. This was first used to study algae in a thin capillary by Weiss et al. (2000) (Figure 13–30) and to study whole-mount eukaryotic cells by Wang et al. (2000), followed by studies of yeast in capillaries (Larabell et al., 2004) (see Figure 13–31). In all of these cases, cells were studied in the frozen hydrated state for reasons that will be discussed in the following paragraph. A third approach beyond two-dimensional imaging is to use X-ray microscopes (Kenney et al., 1985; De Stasio et al., 1996; Buckley et al., 1997) or X-ray microprobes (Kawai et al., 2001; Ortega et al., 2003; Paunesku et al., 2003; Kemner et al., 2004; Behets et al., 2005; Wagner et al., 2005) to study elemental content and distribution in bacteria and cells, particularly in the case of calcium in bone and metals that regulate various biological activities.

When studying biological specimens, attention must be paid to the limitations set by radiation damage. Basic considerations of signal-to-noise and absorption indicate that the radiation dose that is necessarily imparted for X-ray imaging at 50 nm or better resolution is in excess of 10^6 Gray (Sayre et al., 1977a; Schneider, 1998). This is well in excess of

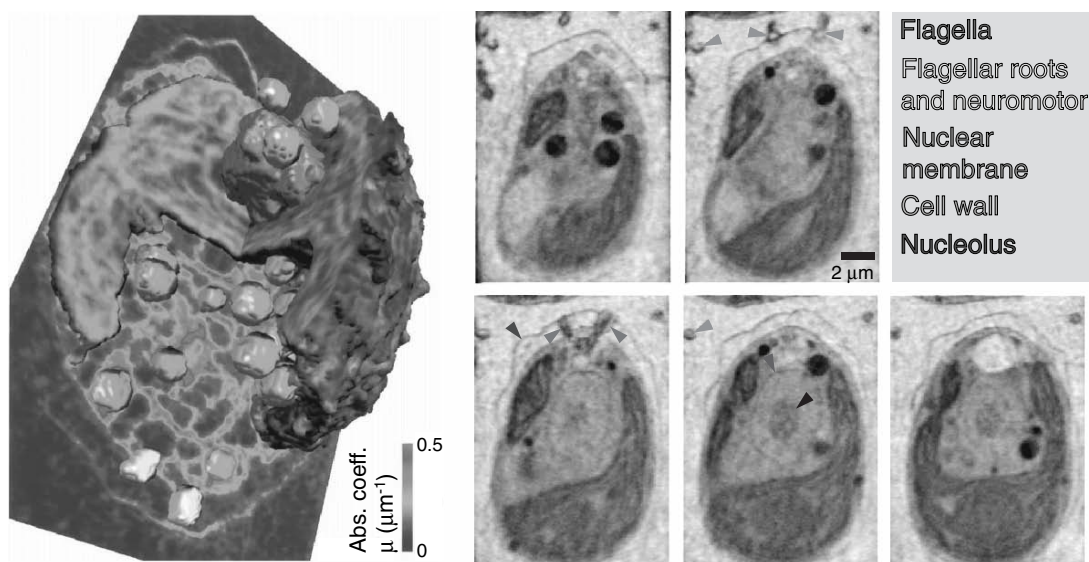


Figure 13–30. 3D rendering (left) and reconstruction slices (right) of the algae *Chlamydomonas reinhardtii* viewed by soft X-ray tomography at the BESSY I synchrotron. This alga was plunge-frozen in liquid ethane, and imaged over 180° rotation sequence. The reconstruction is given in terms of the quantitative linear absorption coefficient for 517 eV X-rays. (Reprinted from Wei et al., © 2000, with permission from Elsevier.) (See color plate.)

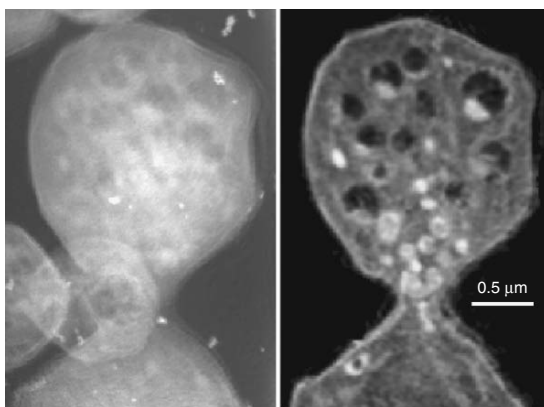


Figure 13-31. Single projection image (left) and slice from a tomographic reconstruction (right) of a frozen hydrated yeast *Saccharomyces cerevisiae*. A number of cells were loaded into a thin-walled, 10 μm diameter glass capillary and rapidly frozen using a jet of helium gas cooled by liquid nitrogen. A series of 45 images through a 180° tilt range was then acquired using the XM-1 TXM at the Advanced Light Source. This illustrates the ability of soft X-ray tomography to image the interior detail of cells rapidly frozen from a living state. (From Larabell et al., 2004. Reprinted from *Molecular Biology of the Cell*, with permission of the American Society for Cell Biology.)

the <10 Gray (1 Gray = 100 rad) dose that is lethal to humans when received over a short time interval. Studies of initially living cells have shown that doses of 10^6 Gray are at the approximate threshold for producing immediate changes in bacteria and are well above the dose needed to affect more complex cells in X-ray microscopy investigations (Gilbert et al., 1992; Pine and Gilbert, 1992; Bennett et al., 1993; Kirz et al., 1995). One of the main damage mechanisms is the creation of radiolytical free radicals in water. Some but not all chemically fixed, hydrated biological specimens will show effects such as mass loss, shrinkage, and the loss of ultrastructural information at these radiation doses as well (Ford et al., 1992; Williams et al., 1993) (of course, chemical fixation produces its own changes on many specimens (Coetzee and van der Merwe, 1984, 1989; Stead et al., 1992; O'Toole et al., 1993; Jearanaikoon and Abraham-Peskir, 2005). Fortunately, a ready solution was developed some years ago by electron microscopists: the use of rapidly frozen specimens in cryomicroscopy (Taylor and Glaeser, 1976; Steinbrecht and Zierold, 1987; Echlin, 1992). In X-ray microscopes, frozen hydrated biological specimens have been shown (Schneider, 1998; Maser et al., 2000) to be well preserved and free of easily visible structural changes and mass loss at radiation doses up to about 10^{10} Gray thus providing the required conditions for a variety of biological studies. The situation for spectroscopy is not yet so clear; cryo methods have been shown to be less effective in preserving XANES resonances in dry polymers (Beetz and Jacobsen, 2003) but they may be more advantageous in studies of frozen hydrated organic specimens due to the inactivation of the diffusion of free radicals ("cage" effect) (Schneider, 1998).

4.2 Environmental Science

Environmental science using synchrotron radiation is a broad topic, as discussed in a recent review (Brown, 2002); we point out here just a few examples using X-ray microscopes.

By placing microliter drops between two silicon nitride windows which are then drawn together by surface tension and some sort of seal, it is straightforward to make a specimen chamber with micrometer-thick water layers and study samples wet and at room temperature (Neuhäusler et al., 2000) (see Figure 13–32). Using this approach, one can use soft X-ray spectromicroscopy to study the role of bacteria and their biofilms in changing the reduction/oxidation state or sequestration of various elements in the environment (Lawrence et al., 2003; Yoon et al., 2004; Hitchcock et al., 2005) (see Figure 13–33), the growth of crystalline materials (Chan et al., 2004), and other geochemical reactions (Myneni et al., 1997; Tonner et al., 1999; Pecher et al., 2003). Spectromicroscopy at the carbon edge can be used to study a variety of organic processes, ranging from the diagenetic breakdown of organic material over geological timescales and its presence and preservation in fossilized plants and wood (Cody, 2000; Boyce et al., 2002, 2004) and coals (Botto et al., 1994; Cody et al., 1995), and the role of natural organic matter in the properties of soils (Thieme et al., 1994; Thieme and Niemeyer, 1998a; Scheinost et al., 2001; Schäfer et al., 2003; Solomon et al., 2005) including its role in the groundwater transport of radionuclides (Schäfer et al., 2005) (see Figure 13–34). Tomography has also been used to study bacterial “microhabitats” (Thieme et al., 2003). Other studies have considered the functional groups present in the soot produced by combustion in diesel engines (Braun et al., 2004).

The trace element mapping capabilities of X-ray microprobes are also very useful for studies in environmental science. Low concentration of iron sets a biotic limit to carbon uptake in the southern Pacific; Twining et al. (2003) have used microprobe studies to investigate this on a cell-by-cell basis (see Figure 13–35) since bulk chemistry measurements do

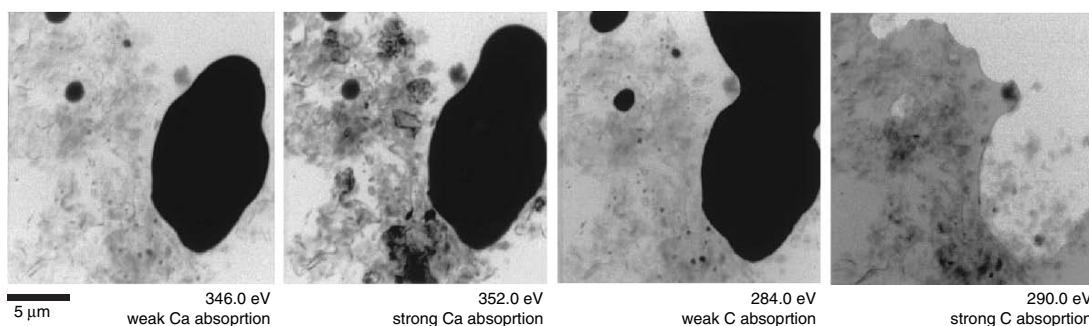


Figure 13–32. Images of a colloidal chemistry sample consisting of oil in water with clays and calcium-rich layered double hydroxides used to “cage” the oil droplet where present (left and bottom edges of the droplet). This illustrates the ability to highlight various elemental components in a room temperature wet specimen. (Courtesy of Neuhäusler, 1999.)

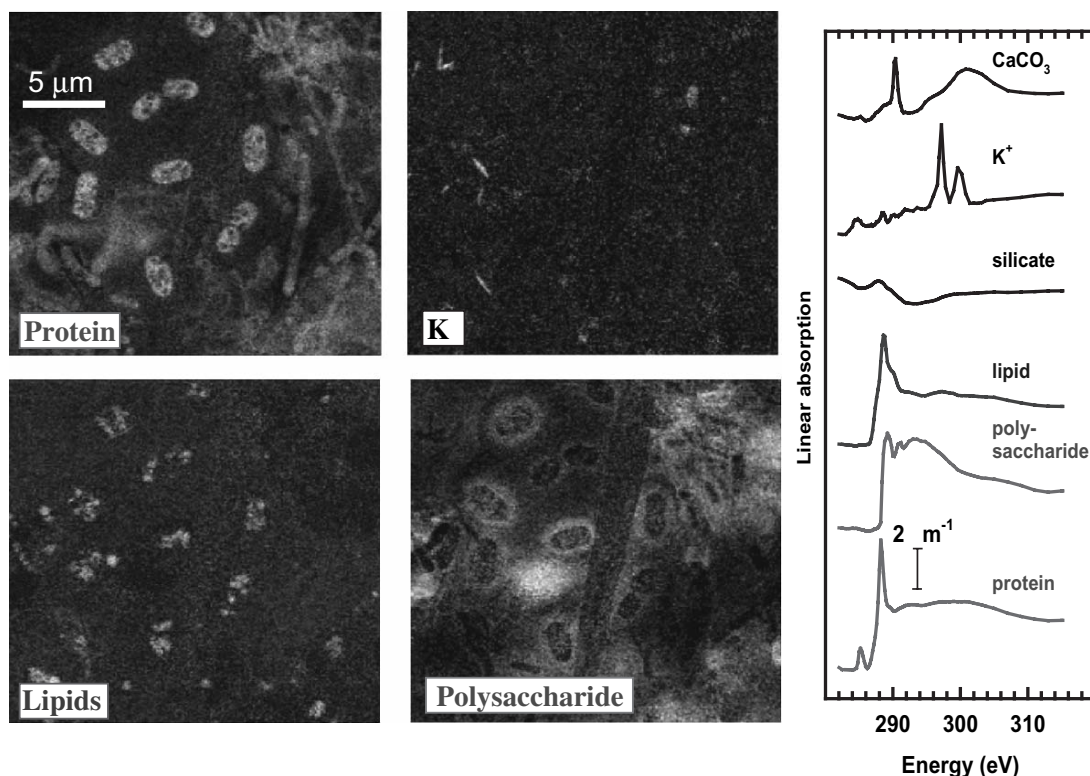


Figure 13-33. Quantitative chemical maps of protein, K⁺, lipids, and polysaccharides from a wet microbial colony from the South Saskatchewan river, derived from STXM images (880 × 880 pixels) and image sequences (52 energies, 230 × 230 pixels). The spatial distributions of the various chemical species are determined by fitting the spectra from each pixel with a linear combination of the absorption spectra of the constituents. X-ray absorption spectra in the C 1s region are shown for CaCO₃, K⁺, silicate, lipid, polysaccharides, and protein. The spectrum of CaCO₃ is from pure material. Those of the other five species are derived from the C 1s image sequence recorded from this biofilm using pixel identification and (for lipid, polysaccharides) spectral subtractions based on fits of the image sequence to the spectra of pure reference materials. (Courtesy A.P. Hitchcock, McMaster University.)

not allow one to differentiate between protist types and particulate matter at the same size scale. Other studies using zone plate microprobes have concentrated on the speciation of metals near the roots of healthy and diseased plants (Yun et al., 1998b), the presence of metals in soil bacteria (Kemner et al., 2004), sulfur speciation in bacteria (Labrenz et al., 2000), in natural silicate glasses (Bonnin-Mosbah et al., 2002), and in microbial filaments (Foriel et al., 2004), and elemental concentrations in atmospheric particles (Ma et al., 2004). These represent only early examples, as the number of projects being carried out using zone plate microprobes is increasing rapidly.

4.3 Materials Science

Applications of X-ray microscopes to material science include four broad categories of study: chemical state mapping in polymer systems

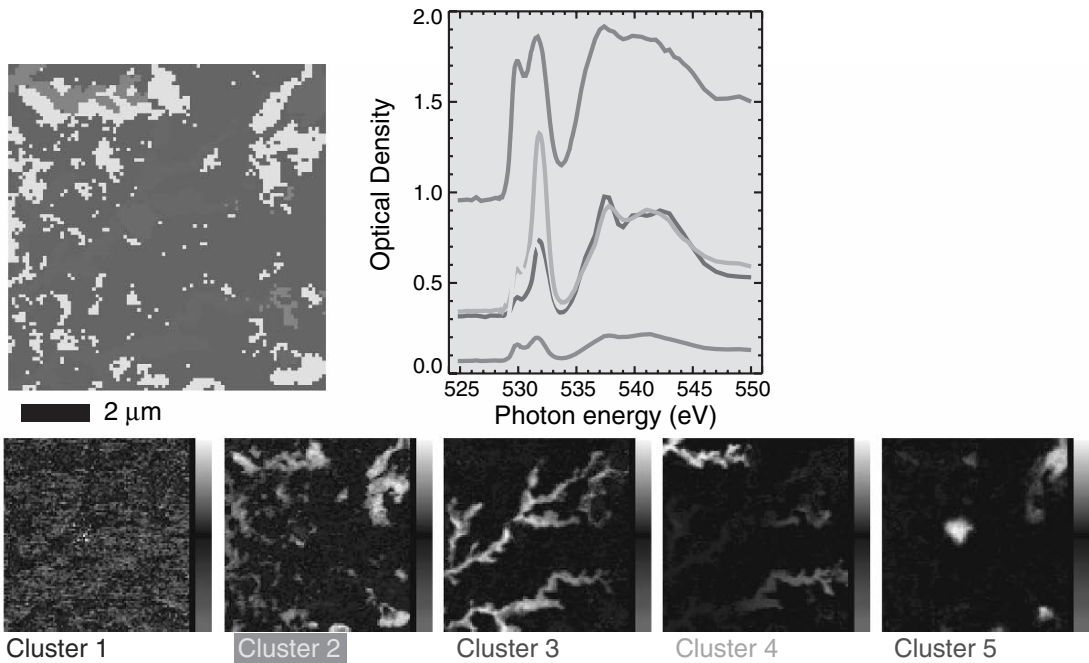


Figure 13–34. Cluster analysis in a spectromicroscopy study of lutetium in hematite. Lutetium is serving as a homologue to americium in an investigation of the uptake and transport of nuclear waste products in groundwater colloids. By using a pattern recognition algorithm to search for pixels with spectroscopic similarities, a set of signature spectra is automatically recovered from the data (shown here in a color-coded classification map) and thickness maps can be formed based on these signature spectra. Analysis at the oxygen edge reveals two different phases of reactivity for lutetium with hematite. Analysis by Lerotic (2004), from a study by T. Schäfer, INE Karlsruhe. (See color plate.)

using spectromicroscopy, imaging of the structure and electromigration failure of integrated circuits, measurements of strain in crystalline materials using microdiffraction, and studies of surface properties using photoelectrons. Studies of polymer systems represent one of the

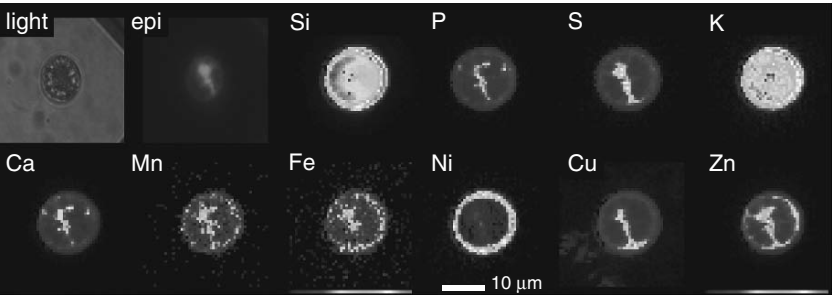


Figure 13–35. Visible light and epifluorescence micrographs, and false color X-ray fluorescence element maps of a centric diatom collected from the southern Pacific. In this region of the ocean, iron availability is a biolimiter with an impact on oceanic uptake of carbon dioxide from the atmosphere. X-ray microprobes allow one to study iron content on a protist-specific basis. (Reprinted from Twining et al., © 2003, with permission from American Chemical Society.) (See color plate.)

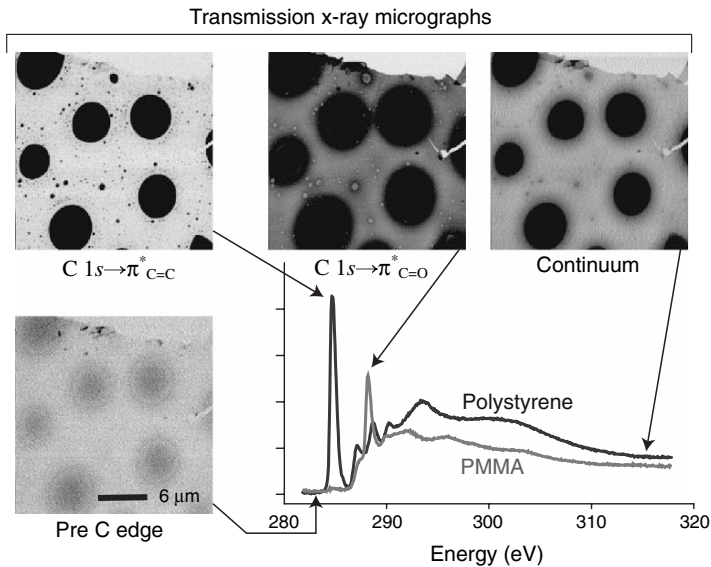


Figure 13-36. One of the first applications of X-ray transmission spectromicroscopy was to the study of polymers, where the chemical selectivity of near-edge absorption resonances allows one to make maps based on XANES spectral signatures. In this example, polymethylmethacrylate (PMMA) was spun cast with polystyrene (PS) before annealing, giving rise to phase segregation. Images acquired at specific absorption resonances show very different contrast and can be used to form compositional maps of the polymers. (Courtesy of D.A. Winest, NCSU.)

first uses of zone plate spectromicroscopy (Ade, 1992) (see Figure 13-36), and subsequent work has ranged from exploring fundamental questions such as confinement-induced miscibility (Zhu et al., 1999) to studies of specific industrially useful materials using both absorption contrast (Smith et al., 2001; Rightor et al., 2002; Hitchcock et al., 2003; Croll et al., 2005) and linear dichroism (Ade and Hsiao, 1993). Other studies have measured the degree to which polymers can seep into wood at the cellular level in particleboard (Buckley et al., 2002). These represent only a few examples; a much wider survey is given in recent reviews (Ade, 1998; Urquhart et al., 1999).

Modern integrated circuits are incredibly intricate, with oxidation layers sometimes only a few molecular layers thick, and metallization planes and vias which connect them, having dimensions in the 100 nm range. The ability of X-ray microscopes to image thick specimens (especially using phase contrast at higher energies; see Figure 13-37) is well suited to studies of the properties and failure modes of such circuits. As one example, Schneider et al. (2002b, 2002c, 2003) have studied electromigration failures as they take place, leading to observations of the propagation of voids from the point of their original formation (see Figures 13-38 and 13-39), while Levine et al. have done tomographic imaging of electromigration voids using a STXM (Levine et al., 2000). For industrial applications of chip inspection, a very significant development has been the recent commercial availability (Xradia, Inc.) of

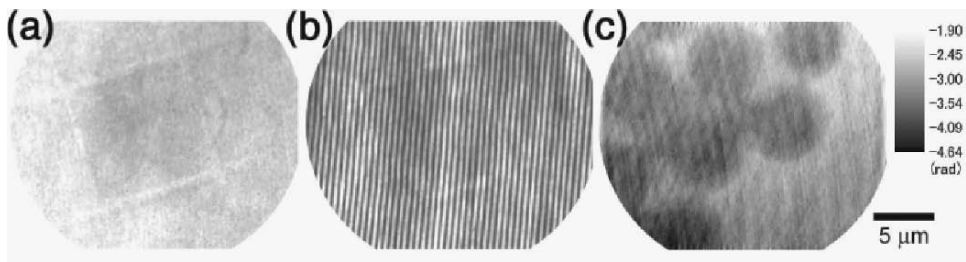


Figure 13–37. Interferometric TXM imaging of polystyrenes at 9 KeV. In these experiments, a hard X-ray micro-interferometer has been constructed by using two overlying objective zone plates with a slight transverse offset to produce an interferometric fringe pattern as shown in (b). Compared to the single-objective image (a), interference fringes with a visibility of as high as 60% can clearly be seen. Analysis of the interferometric image (b) is then used to obtain the quantitative contrast image of the polystyrene spheres shown in (c). This example shows how low absorption contrast objects can be imaged in hard X-ray microscopes. (Courtesy of T. Koyama, Himeji Institute of Technology.)

laboratory-based tomography systems using zone plate optics and operating at a sufficiently high energy (5.4 keV) so as to allow tomographic data sets to be acquired and reconstructed; this allows one to study various metallization layers in intact, working chips (Wang et al., 2002) (see Figure 13–40).

Another way in which zone plate X-ray microscopes are used to study material properties is through microdiffraction, where one

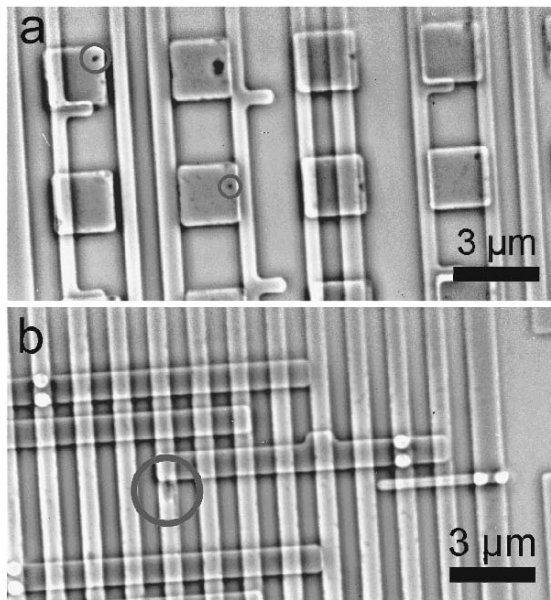


Figure 13–38. Zernike phase contrast provides one means to image the metallic layers of integrated circuits in regions where the underlying silicon wafer has been thinned. A common failure mode in integrated circuits is electromigration in which voids in a conducting layer or via are formed. These images obtained using a TXM operating at 4 keV at the European Synchrotron Radiation Research Facility or ESRF, show what appear to be such voids (circles) within test structures for advanced microprocessors. (Reprinted from Schneider et al., © 2003, with permission from Elsevier.)

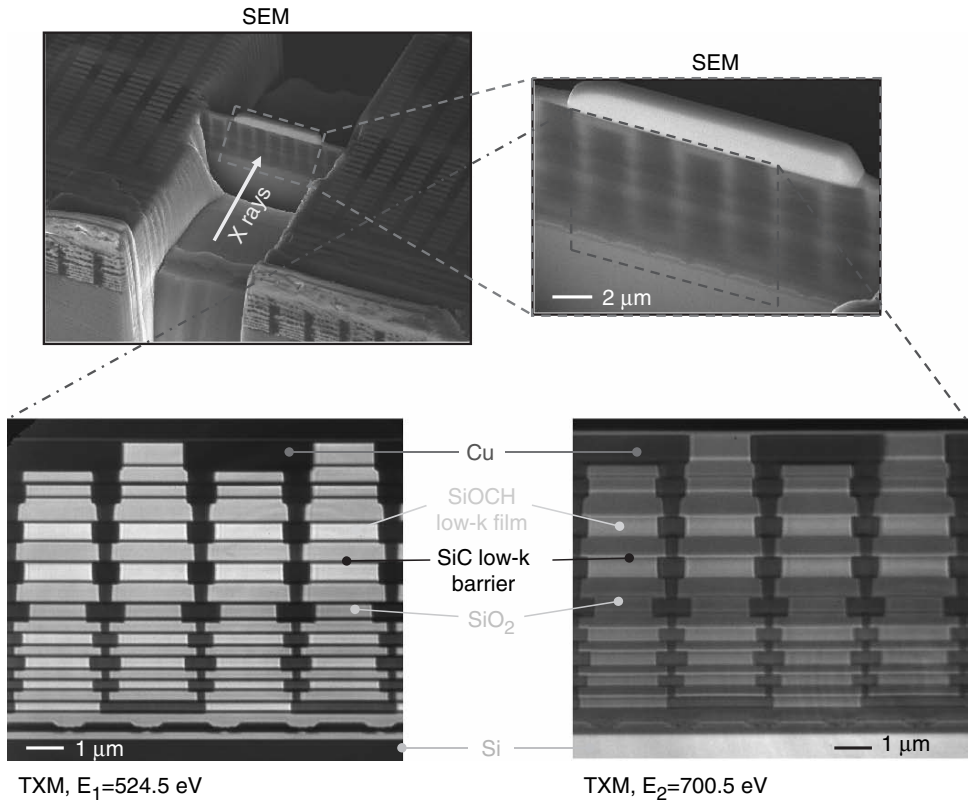


Figure 13–39. In studies of integrated circuits, it is often important to study changes in metallization from layer-to-layer by imaging in cross section. In this example, a focused ion beam (FIB) system was used to prepare a thinned, fully passivated cross section of copper interconnect structures within an electrically functional test structure as shown in the two scanning electron micrographs at top. A soft X-ray TXM at the BESSY II synchrotron facility in Berlin was then used to image this cross section at 525.5 eV (left) and 700.5 eV (right) to highlight different metal and dielectric layers in the chip, with features as fine as 20 nm visible. (Courtesy of G. Schneider, BESSY.)

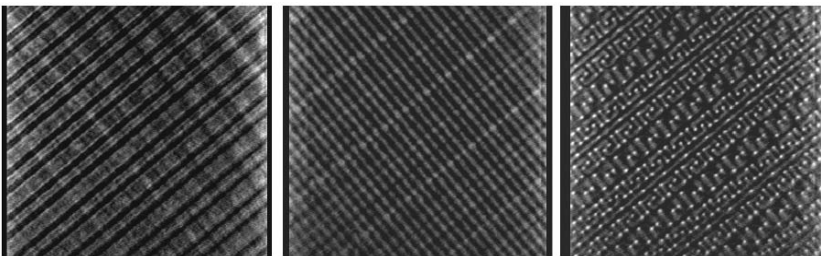


Figure 13–40. Tomographic imaging of an integrated circuit done with a commercial laboratory X-ray microscope (Xradia). An integrated circuit had the silicon wafer underneath a region of interest thinned to about 15 μm , after which a tilt series of TXM images was acquired over 8 hours using a rotating anode source operation at 5.4 keV. The figure shows slices extracted at depths corresponding to the center of three Cu interconnect layers in the tomographic reconstruction with an estimated resolution of 60 nm in the transverse dimension and 90 nm in depth. This system can be used for chip inspection at a chip fab plant, among other applications. (Courtesy W. Yun, Xradia.)

examines not the undeviated transmission image through the specimen, but the signal that is Bragg diffracted (usually in the Laue geometry) by specific crystalline regions within the specimen (see for example (Engström et al., 1995)). Measurement of the position of the Bragg peaks can give values of the local lattice constants so that repetition of the measurement over a grid of points provides a strain map of the sample. This has been applied to optoelectronic devices (Cai et al., 1999), magnetic domain evolution (Evans et al., 2002) as well as for examination of the strain at the midpoint and edges of mesoscopic structures (Murray et al., 2005) (see Figure 13–41).

Since photoelectrons emerge only from within the top 100 nm or so of a bulk specimen, methods that use photoelectron detection are ideal for studies of surface phenomena. Photoelectron emission microscopes using X-ray illumination of a broad area and sub-30 nm resolution electron optics are beyond the scope of our concentration on zone plate microscopes, though we note that they are used with great success and at very high spatial resolution (see Figure 13–42). Another type of photoelectron microscope is a Scanning PhotoEmission Microscope or SPEM using a zone plate to produce a fine focus and an electron spectrometer for signal detection (Ade et al., 1990b; Ko et al., 1998; Yi et al., 2005) (see Fig. 43); activities in this area were recently reviewed by Günther et al. (2002).

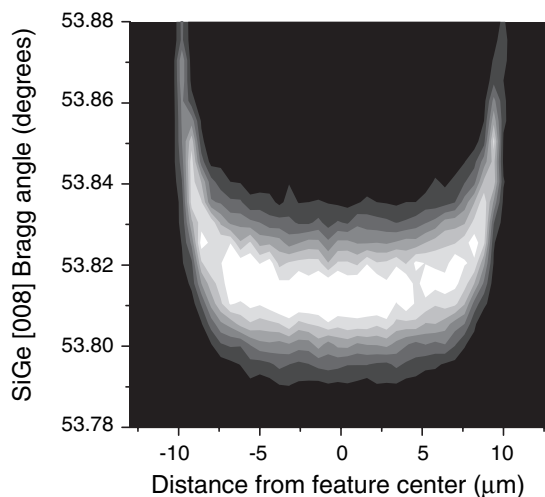


Figure 13–41. In X-ray microdiffraction, a detector is set to collect diffraction from small, crystalline features of the specimen that can be selectively illuminated by the microfocus beam. Local variations from perfect crystal order are seen as changes in the width or angle of the diffraction peaks. In this example, a 20 μm wide, 0.24 μm thick $\text{Si}_{0.86}\text{Ge}_{0.14}$ pseudomorphically strained film is located on a Si [001] surface. A determination of the angle of the SiGe [008] diffraction peak as a function of position on the sample reveals elastic relaxation at the free edges of the SiGe feature, and demonstrates the ability of a zone-plate STXM to study the strain distribution of patterned microstructures. (Reprinted from Murray et al., © 2005, with permission from American Institute of Physics.) (See color plate.)

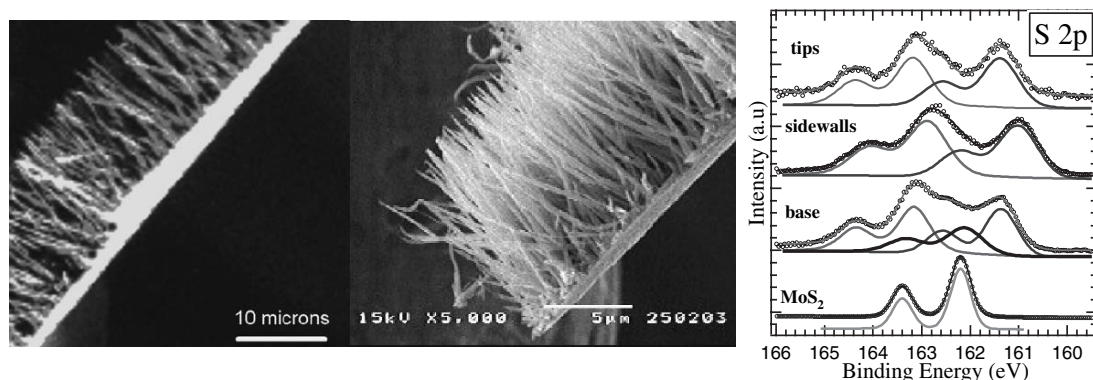


Figure 13-42. In Electron Spectroscopy for Chemical Analysis or ESCA microscopy, a monochromatic beam is used to illuminate a region several micrometers across; electron optics are then used to image a tunable electron ejection energy to reveal surface chemistry. Though this does not involve zone plate imaging, we include it here due to its widespread use with tunable X rays. In this case a 90-nm resolution ESCA microscope was used to locate aligned MoS_2 nanotube bundles and select certain areas along the axes of the tubes for detailed examination. The image at left was acquired using $\text{Mo } 3d$ electrons, while $\text{S } 2p$, $\text{Mo } 3d$, and valence band spectra taken at the tips and sidewalls and the growth base from the Si wafer appear strongly affected by the low dimensionality of the nanotubes and differ significantly from the corresponding spectra taken on a reference MoS_2 crystal. (Reprinted from Kiskinova et al., © 2003, with permission of EDP Sciences.)

4.4 Magnetic Materials

X-ray magnetic circular dichroism (XMCD) exploits changes in absorption due to the relative orientation of magnetic domains and incident circularly polarized radiation. It draws upon the fact that in magnetic

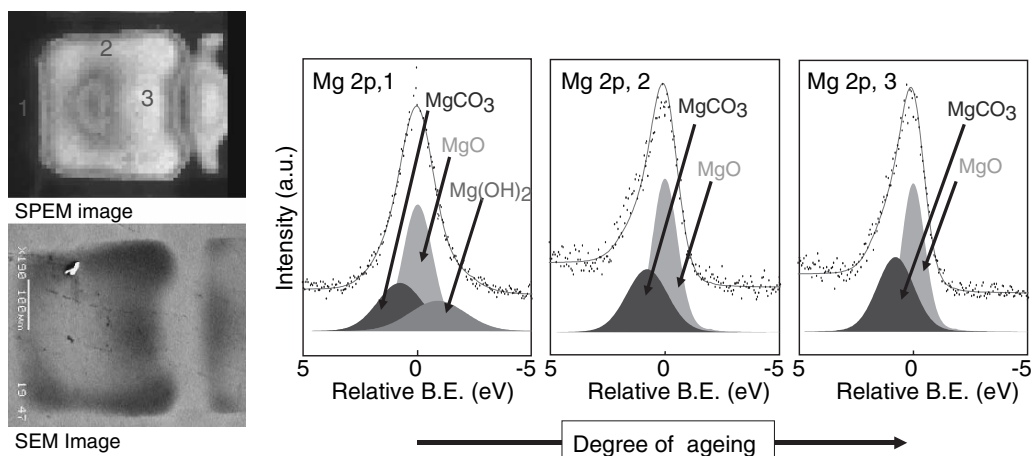


Figure 13-43. Scanning photoemission microscope (SPEM) study of a plasma display cell. In this microscope the specimen is scanned through the zone plate focus while photoelectrons are collected by an electron spectrometer. This figure shows a SPEM image, a scanning electron micrograph, and photoelectron spectra from several regions of the sample. In a plasma display cell, light of the appropriate color emerges through a front glass window which is protected from plasma damage by a composite insulating layer including MgO . The photoelectron spectra show aging in the Mg(OH)_2 component of the layer over the life of the display cell. (Reprinted from Yi et al., © 2005, with permission from the Institute of Pure and Applied Physics.) (See color plate.)

materials the density of certain electronic states is different for electrons of spin parallel to the magnetization, compared to electrons of spin anti-parallel. The absorption of circularly polarized photons selects between electron spins, and depends on the component of spin parallel to the helicity of the photon (the direction of the photon beam). Images taken with a particular polarization of the illumination beam, at saturated magnetization states, or at L_2 versus L_3 absorption edges can by themselves show magnetic contrast effects, while difference images between two polarization states at an absorption edge can be used to obtain element-specific images of magnetic contrast only. While much work has been done using photoemission microscopes (Stöhr et al., 1993) and there are recent exciting results using X-ray holography (Eisebitt et al., 2004), zone plate microscopy provides two primary approaches. One method is to use a TXM with a large-angle-collection condenser zone plate and exploit the fact that the radiation from synchrotron bending magnet sources is circularly polarized above and below the synchrotron plane; this was the first method demonstrated (Fischer et al., 1996) and it has led to considerable success for the study of out-of-specimen-plane magnetism (Fischer et al., 2001a) (see Figure 13–44) and has recently been extended to the study of in-specimen-

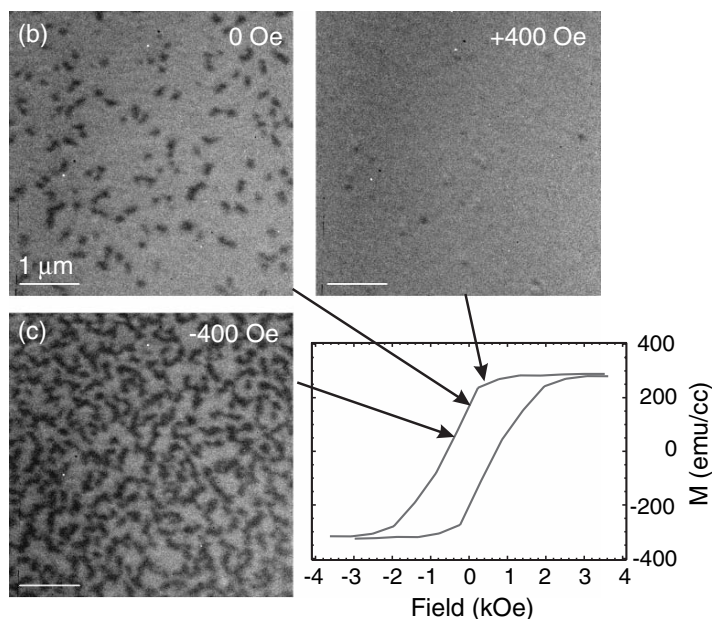


Figure 13–44. X-ray magnetic circular dichroism (XMCD) images of the magnetic domain structure of a 50-nm thick $(\text{Co}_{83}\text{Cr}_{17})_{87}\text{Pt}_{13}$ alloy film recorded at the Co L_3 absorption edge (777 eV) and in an external field of (a) +400 Oe, (b) 0 Oe, and (c) –400 Oe. (d) M vs. H hysteresis loop obtained via VSM measurement. The arrows indicate the point in the reversal cycle at which each image is recorded. Domain structure is apparent as the magnetization of the film is driven around the hysteresis loop and the net magnetization reversal can be seen to be the average of the reversal of individual domains, with the number of reversed domains increasing as the strength of the applied field is increased. (Reprinted from Im et al., © 2003, with permission from American Institute of Physics.)

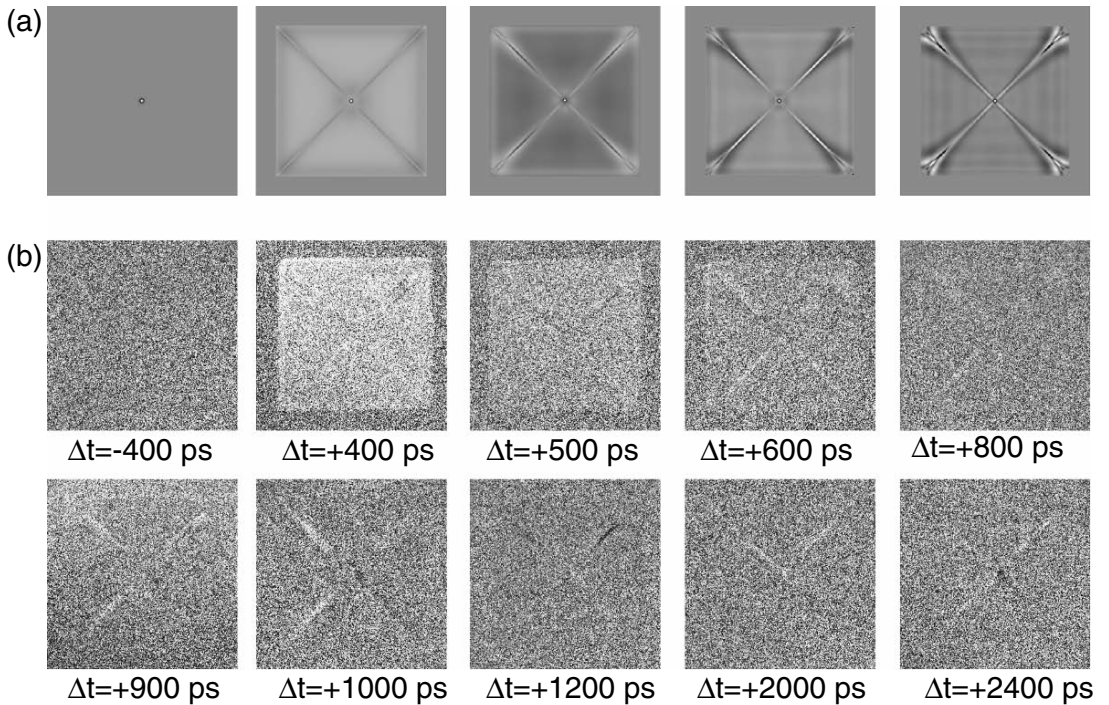


Figure 13-45. Time-resolved XMCD imaging of a magnetized Ni-Fe film patch as the magnetization is reversed in an applied magnetic field. a) The z-component of the dynamic magnetization at selected time delays obtained from micro-magnetic simulations (OOMMF). b) XMCD images from the XM-1 TXM taken with various time delays between the application of the pulsed magnetic field and the arrival of radiation from electron bunches in the storage ring. By integrating over many bunches with a particular time delay, one can study the temporal evolution of the z-component of the magnetization at delay times varying from probe pulse 400ps before the pump, up to 2400ps after the pump. (Reprinted from Stoll et al., © 2004, with permission from American Institute of Physics.)

plane magnetic structure as well (Fischer et al., 2001b). Another more recent approach is to use a STXM with a variable polarization undulator source. In either case, the pulsed nature of synchrotron radiation from electron bunches means that one can cycle an applied magnetic field in synchrony with the arrival of short (~ 100 psec) pulses of X-rays, and thereby accumulate images corresponding to controlled time delays before and after application of the pulsed field (Stoll et al., 2004) (see Figure 13-45). A more extended discussion of magnetic contrast X-ray microscopy is provided in a recent review by Fischer (Fischer, 2003).

5 Conclusion

In this chapter, we have outlined some of the principles and characteristics of X-ray microscopes using zone plate optics, and have attempted to convey an incomplete but representative survey of their applications in scientific studies. We have seen that the resolution and efficiency of

zone plates has improved considerably over the lifetime of the field, although, in spite of constant efforts and the application of the best technology, the rate of improvement has been slow. For some time the "Moore's Law" graph for zone plate resolution has had a slope of about a factor of two per decade. However, as we have seen, this area of development has been especially active in recent times. There is now some optimism that the 10nm barrier may be broken and the present art is nowhere close to hitting fundamental limits. Resolution is not the whole story, however; many applications are combining imaging with tilt of the specimen for tomography, with energy tunability for spectromicroscopy, and with fluorescence detection for elemental identification. These represent the application of zone plate optics to extend the boundaries of previously existing techniques with active communities, so these areas are likely to expand. Another general trend of the last few years has been the growth in hard X-ray applications of zone plate imaging. This has been especially beneficial for tomography and microanalysis and, as recent experiments have shown, the use of hard X-ray zone plates in high order may soon approach the best resolution of soft X-ray zone plates in first order. At the time of this writing it seems that technical developments in X-ray microscopy and its marriage with promising application areas is happening at an ever-increasing pace and we can now forecast that these activities have a bright future with more confidence than ever before.

Acknowledgments. Naturally an enterprise like writing this review depends greatly on the willingness of our colleagues around the X-ray microscopy community to provide us with advice information and images and we thank the many people who have done that. We especially thank Janos Kirz and Henry Chapman for reading the manuscript and our immediate colleagues at Stony Brook, Brookhaven and Berkeley for many helpful discussions. Work by MH and TW was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences Division of the U. S. Department of Energy, under Contract No. DE-AC03-76SF00098. Work by CJ was supported by the National Institutes of Health under grant R01 EB00479-01A1, and the National Science Foundation under grants DBI-9986819, ECS-0099893, and CHE-0221934.

References

- Abraham-Peskir, J. (1998). Structural changes in fully hydrated *Chilomonas paramecium* exposed to copper. *Eur. J. Protisto.* **34**, 51–57.
- Abraham-Peskir, J. (2000). X-ray microscopy with synchrotron radiation: Applications to cellular biology. *Cel. Molec. Biol.* **46**(6), 1045–1052.
- Ade, H. (1998). X-ray spectromicroscopy. *Experimental Methods in the Physical Sciences.* **32**, 225–262 (R. Celotta and T. Lucatorto, Eds.). (Academic Press New York).
- Ade, H. and Hsiao, B. (1993). X-ray linear dichroism microscopy. *Science* **262**, 1427–1429.

- Ade, H., Kirz, J., Hulbert, S., Johnson, E., Anderson, E. and Kern, D. (1990a). Scanning photoelectron microscope with a zone plate generated microprobe. *Nucl. Inst. Methods Phys. Res. A* **291**, 126–131.
- Ade, H., Kirz, J., Hulbert, S.L., Johnson, E., Anderson, E. and Kern, D. (1990b). X-ray spectromicroscopy with a zone plate generated microprobe. *Appl. Phys. Lett.* **56**, 1841–1843.
- Ade, H., Zhang, X., Cameron, S., Costello, C., Kirz, J. and Williams, S. (1992). Chemical contrast in X-ray microscopy and spatially resolved XANES spectroscopy of organic specimens. *Science* **258**, 972–975.
- Agard, D. and Sedat, J. (1983). Three-dimensional architecture of a polytene chromosome. *Nature* **302**, 676–681.
- Anderson, E.H., Bögli, V. and Muray, L.P. (1995). Electron beam lithography digital pattern generator and electronics for generalized curvilinear structures. *J. Vacuum Sci. Tech. B* **13**, 2529–2534.
- Anderson, E.H., Olynick, D.L., Harteneck, B., Veklerov, E., Denbeaux, G., Chao, W., Lucero, A., Johnson, L. and Attwood, D. (2000). Nanofabrication and diffractive optics for high-resolution X-ray applications. *J. Vacuum Sci. Tech. B* **18**(6), 2970–2975.
- Aoki, S. (1994). Grazing incidence X-ray microscope with a Wolter type mirror. In *Radiation in the Biosciences* (B. Chance, D. Deisenhofer, S. Ebashi, D.T. Goodhead, J.R. Helliwell, H.E. Huxley, T. Iizuka, J. Kirz, T. Mitsui, E. Rubenstein, N. Sakabe, T. Sasaki, G. Schmahl, H. Sturhmann, K. Wüthrich and G. Zaccai, Eds.) (Oxford University Press, New York).
- Aoki, S. and Kikuta, S. (1974). X-ray holographic microscopy. *Japan. J. Appl. Phys.* **13**, 1385–1392.
- Aristov, V.V. and Erko, A.I., Eds. (1994). *X-ray Microscopy IV*. (Bogorodskii Pechatnik, Chernogolovka, Russia).
- Attwood, D. (1999). *Soft X-rays and Extreme Ultraviolet Radiation*. (Cambridge Univ. Press, Cambridge).
- Baez, A.V. (1960). A self-supporting metal Fresnel zone-plate to focus extreme ultra-violet and soft X-rays. *Nature* **186**, 958.
- Baez, A.V. (1961). Fresnel zone plate for optical image formation using extreme ultraviolet and soft X radiation. *J. Opt. Soc. Amer.* **51**, 405–412.
- Baez, A.V. (1989). The early days of X-ray optics: A personal memoir. *J. X-ray Sci. Tech.* **1**, 3–6.
- Baez, A.V. (1997). Anecdotes about the early days of X-ray optics. *J. X-ray Sci. Tech.* **7**(2), 90–97.
- Barbee, T.W., Jr. (1981). Sputtered layered synthetic microstructure (LSM) dispersion elements. In *Proceedings of the Conference on Low Energy X-ray Diagnostics* (American Institute of Physics, Monterey, CA).
- Barrett, R., Kaulich, B., Oestreich, S. and Susini, J. (1998). Scanning microscopy end station at the ESRF X-ray microscopy beamline. In *X-ray Microfocusing: Applications and Techniques* (I. McNulty, Ed.). *Proceedings of the SPIE*, Vol. 3449, 80–90 (SPIE, Bellingham, WA).
- Batterman, B. and Cole, H. (1964). Dynamical diffraction of X-rays by perfect crystals. *Rev. Mod. Phys.* **36**, 681–717.
- Beetz, T. and Jacobsen, C. (2003). Soft X-ray radiation-damage studies in PMMA using a cryo-STXM. *J. Synchrotron Rad.* **10**(3), 280–283.
- Behets, G.J., Verberckmoes, S.C., Oste, L., Bervoets, A.R., Salome, M., Cox, A.G., Denton, J., De Broe, M.E. and D'Haese, P.C. (2005). Localization of lanthanum in bone of chronic renal failure rats after oral dosing with lanthanum carbonate. *Kidney Int.* **67**(5), 1830–1836.
- Bennett, P.M., Foster, G.F., Buckley, C.J. and Burge, R.E. (1993). The effect of soft X-radiation on myofibrils. *J. Microsc.* **172**, 109–119.

- Bilderback, D.H. (2003). Review of capillary X-ray optics from the second international capillary optics meeting. *X-ray Spectrom.* **32**(3), 195–207.
- Bionta, R.M., Skulina, K.M. and Weinberg, J. (1994). Hard X-ray sputtered-sliced phase zone plates. *Appl. Phys. Lett.* **64**(8), 945–947.
- Bögli, V., Unger, P., Beneking, H., Greinke, B., Guttman, P., Neimann, B., Rudolf, D. and Schmahl G. (1988). Microzoneplate fabrication by 100 keV electron beam lithography. In *X-ray Microscopy II* (Springer-Verlag, New York).
- Bonnin-Mosbah, M., Metrich, N., Susini, J., Salome, M., Massare, D. and Menez, B. (2002). Micro X-ray absorption near edge structure at the sulfur and iron K-edges in natural silicate glasses. *Spectrochimica Acta Part B* **57**(4), 711–725.
- Born, M. and Wolf, E. (1999). *Principles of Optics*. (Cambridge University Press, Cambridge).
- Botto, R.E., Cody, G.D., Kirz, J., Ade, H., Behal, S. and Disko, M. (1994). Selective Chemical Mapping of Coal Microheterogeneity by Scanning Transmission X-ray Microscopy. *Energy and Fuels* **8**, 151–154.
- Boyce, C.K., Cody, G.D., Feser, M., Jacobsen, C., Knoll, A.H. and Wirick, S. (2002). Organic chemical differentiation within fossil plant cell walls detected with X-ray spectromicroscopy. *Geology* **30**(11), 1039–1042.
- Boyce, C.K., Zwieniecki, M.A., Cody, G.D., Jacobsen, C., Wirick, S., Knoll, A.H. and Holbrook, N.M. (2004). Evolution of xylem lignification and hydrogel transport regulation. *Proc. Nat. Acad. Sci.* **101**(50), 17555–17558.
- Braun, A., Shah, N., Huggins, F.E., Huffman, G.P., Wirick, S., Jacobsen, C., Kelly, K. and Sarofim, A.F. (2004). A study of diesel PM with X-ray micro-spectroscopy. *Fuel* **83**, 997–1000.
- Brown, G.E. and Sturchio, N.C. (2002). An overview of synchrotron radiation applications to low temperature geochemistry and environmental science. In *Applications of Synchrotron Radiation in Low-Temperature Geochemistry and Environmental Sciences. Rev. Mineral Geochem.*, **49**, 1–115.
- Buckley, C.J., Browne, M.T. and Charalambous, P.S. (1985). Contamination lithography for the fabrication of zone plate X-ray lenses. In *Electron Beam, X-ray and Ion-Beam Techniques for Submicrometer Lithographies* (P.D. Blair, Ed.). *Proceedings of the SPIE*, Vol. 537, 213–217 (SPIE, Bellingham, WA).
- Buckley, C.J., Khaleque, N., Bellamy, S.J., Robins, M. and Zhang, X. (1997). Mapping the organic and inorganic components of tissue using NEXAFS. *J. Phys. IV* **7**, 83–90.
- Buckley, C.J., Phanopoulos, C., Khaleque, N., Engelen, A., Holwill, M.E.J. and Michette, A.G. (2002). Examination of the penetration of polymeric methylene di-phenyl-di-isocyanate (pMDI) into wood structure using chemical-state x-ray microscopy. *Holzforschung* **56**(2), 215–222.
- Cai, Z., Lai, B., Xiao, Y. and Xu, S. (2003). An X-ray diffraction microscope at the Advanced Photon Source. *J. Phys. IV* **104**, 17–20.
- Cai, Z.H., Rodrigues, W., Ilinski, P., Legnini, D., Lai, B., Yun, W., Isaacs, E.D., Lutterodt, K.E., Grenko, J., Glew, R., Spitz, S., Vandenberg, J., People, R., Alam, M.E., Hybertsen, M. and Ketelsen, L.J.P. (1999). Synchrotron x-ray microdiffraction diagnostics of multilayer optoelectronic devices. *App. Phys. Lett.* **75**(1), 100–102.
- Carrington, W.A., Lynch, R.M., Moore, E.D.M., Isenberg, G., Fogarty, K.E. and Fay, F.S. (1995). Superresolution three-dimensional images of fluorescence in cells with minimal light exposure. *Science* **268**, 1483–1487.
- Casalis, L., Jark, W., Kiskinova, M., Lonza, D., Melpignano, P., Morris, D., Rosei, R., Savoia, A., Abrami, A., Fava, C., Furlan, P., Pugliese, R., Vivoda, D., Sandrin, G., Wei, F.Q., Contarini, S., DeAngelis, L., Gariazzo, C., Nataletti,

- P. and Morrison, G. (1995). ESCA Microscopy beamline at ELETTRA. *Rev. Sci. Instr.* **66**, 4870–4875.
- Castaing, R., Deshamps, P. and Philibert, J., Eds. (1966). *X-ray Optics and X-ray Microanalysis* (Herman, Paris).
- Chan, C.S., De Stasio, G., Welch, S.A., Girasole, M., Frazer, B.H., Nesterova, M.V., Fakra, S. and Banfield, J.F. (2004). Microbial polysaccharides template assembly of nanocrystal fibers. *Science* **303**, 1656–1658.
- Chantler, E. and Abraham-Peskir, J.V. (2004). Significance of midpiece vesicles and functional integrity of the membranes of human spermatozoa after osmotic stress. *Andrologia* **36**(2), 87–93.
- Chao, W., Anderson, E., Denbeaux, G.P., Harteneck, B., Liddle, J.A., Olynick, D.L., Pearson, A.L., Salmassi, F., Song, C.Y. and Attwood, D.T. (2003) 20-nm-resolution soft x-ray microscopy demonstrated by use of multilayer test structures, *Opt. Lett.* **28**(21), 2019–2021.
- Chao, W., Harteneck, B., Liddle, A., Anderson, E. and Attwood, D. (2005). Soft X-ray microscopy at a spatial resolution better than 15 nm. *Nature* **345**, 1210–1213.
- Chapman, H.N. (1996a). Phase retrieval X-ray microscopy by Wigner-distribution deconvolution. *Ultramicroscopy* **66**, 153–172.
- Chapman, H.N., Fu, J., Jacobsen, C. and Williams, S. (1996b). Dark-field X-ray microscopy of immunogold-labeled cells. *J. Microsc. Soc. Amer.* **2**(2), 53–62.
- Chapman, H.N., Jacobsen, C. and Williams, S. (1996c). A characterisation of dark-field imaging of colloidal gold labels in a scanning transmission x-ray microscope. *Ultramicroscopy* **62**(3), 191–213.
- Chapman, H.N., Barty, A., Beetz, T., Cui, C., He, H., Howells, M.R., Marchesini, S., Noy, A., Spence, J.C.H. and Weierstall, U. (2006). High-resolution three-dimensional X-ray diffraction microscopy. *J. Opt. Soc. Amer.* **23**, 1179–1201.
- Cloetens, P., Ludwig, W., Baruchel, J., Van Dyck, D., Van Landuyt, J., Guigay, J.P. and Schlenker, M. (1999). Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays. *App. Phys. Lett.* **75**(19), 2912–2914.
- Cody, G., Botto, R.E., Ade, H. and Wirick, S. (1995). Soft x-ray microscopy and microanalysis: Applications in organic geochemistry. In *X-ray Microbeam Technology and Applications* (W. Yun, Ed.). *Proceedings of the SPIE*, Vol. 2516, 185–196 (SPIE, Bellingham, WA).
- Cody, G.D. (2000). Probing chemistry within the membrane structure of wood with soft X-ray spectral microscopy. In *X-ray Microscopy: Proceedings of the Sixth International Conference* (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.), (American Institute of Physics, Melville, NY).
- Coetzee, J. and van der Merwe, C.F. (1984). Extraction of substances during glutaraldehyde fixation of plant cells. *J. Microsc.* **135**, 147–158.
- Coetzee, J. and Merwe, C.F.V.D. (1989). Extraction of carbon 14 labeled compounds from tissue during processing for electron microscopy. *J. Electron Microsc. Tech.* **11**, 155–160.
- Cosslett, V.E. and Nixon, W.C. (1960). *X-ray Microscopy* (Cambridge University Press, London).
- Cosslett, V.E., Engström, A. and Pattee, H.H. Eds. (1957). *X-ray Optics and X-ray Microanalysis* (Academic Press, New York).
- Croll, L.M., Stover, H.D.H. and Hitchcock, A.P. (2005). Composite tectocapsules containing porous polymer microspheres as release gates. *Macromolecules* **38**(7), 2903–2910.
- David, C., Kaulich, B., Medenwaldt, R., Hettwer, M., Fay, N., Diehl, M., Thieme, J. and Schmahl, G. (1995). Low-distortion electron-beam lithography for

- fabrication of high-resolution germanium and tantalum phase zone plates. *J. Vacuum Sci. Tech. B* **13**(6), 2762–2766.
- David, C., Kaulich, B., Barrett, R., Salome, M. and Susini, J. (2000). High-resolution lenses for sub-100 nm x-ray fluorescence microscopy. *App. Phys. Lett.* **77**(23), 3851–3853.
- David, C., Nöhammer, B., Solak, H.H., Glaus, F., Haas, B., Grubelnik, A., Dolocan, A., Ziegler, E., Hignette, O., Burghammer, M., Kaulich, B., Susini, J., Bongaerts, J. and Veen, J.V.D. (2003). Diffractive soft and hard x-ray optics. *J. Phys. IV* **104**, 171–176.
- De Stasio, G., Mercanti, D., Ciotti, M.T., Droubay, T.C., Perfetti, P., Margaritonda, G. and Tonner, B.P. (1996). Synchrotron spectromicroscopy of cobalt accumulation in granule cells, glial cells and GABAergic neurons. *J. Phys. D: App. Phys.* **29**, 259–262.
- Dhez, O., Ade, H. and Urquhart, S.G. (2003). Calibrated NEXAFS spectra of some common polymers. *J. Electron Spectrosc. Rel. Phenomena* **128**(1), 85–96.
- Di Fabrizio, E., Gentili, M., Grella, L., Baciocchi, M., Krasnoperova, A., Cerrina, F., Yun, W., Lai, B. and Gluskin, E. (1994). High-performance multilevel blazed x-ray microscopy Fresnel zone plates: Fabricated using x-ray lithography. *J. Vacuum Sci. Tech. B* **12**(6), 3979–3985.
- Di Fabrizio, E., Romanato, F., Gentili, M., Cabrini, S., Kaulich, B., Susini, J. and Barrett, R. (1999). High-efficiency multilevel zone plate for keV x-rays. *Nature* **401**(6756), 895–898.
- Duvel, A., Rudolph, D. and Schmahl, G. (2003). Fabrication of thick zone plates for multi-kilovolt x-rays. *J. Phys. IV* **104**, 607–614.
- Eaton, W.J., Morrison, G.R. and Waltham, N.R. (2000). Configured detector for STXM imaging. In *X-ray Microscopy: Proceedings of the Sixth International Conference* (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- Echlin, P. (1992). *Low Temperature Microscopy and Analysis*. (Plenum Publishing, New York).
- Eimüller, T., Niemann, B., Guttman, P., Fischer, P., Englisch, U., Vatter, R., Wolter, C., Seiffert, S., Schmahl, G. and Schütz, G. (2003). The magnetic transmission X-ray microscope project at BESSY II. *J. Phys. IV* **104**, 91–94.
- Einstein, A. (1918). Lassen sich Brechungsexponenten der Körper für Röntgenstrahlen experimentell ermitteln? *Verhandlungen Deutsche Physikalische Gesellschaft* **20**, 86–87.
- Eisebitt, S., Lüning, J., Schlotter, W.F., Lörger, M., Hellwig, O., Eberhardt, W. and Stöhr, J. (2004). Lensless imaging of magnetic nanostructures by X-ray spectro-holography. *Nature* **432**, 885–888.
- Engström, A. (1946). Quantitative micro- and histochemical elementary analysis by Roentgen absorption spectrography. *Acta Radiologica (Supplementum LXIII)* **63**, 1–106.
- Engström, A., Cosslett, V.E. and Pattee, J.H.H., Eds. (1960). *X-ray Microscopy and X-ray Microanalysis*. (Elsevier, Amsterdam).
- Engström, P., Fiedler, S. and Riekel, C. (1995). Microdiffraction instrumentation and experiments on the microfocus beamline at the ESRF. *Rev. Sci. Instr.* **66**(2), 1348–1350.
- Evans, P.G., Isaacs, E.D., Aeppli, G., Cai, Z. and Lai, B. (2002). X-ray microdiffraction images of antiferromagnetic domain evolution in chromium. *Science* **295**, 1042–1045.
- Feser, M., Carlucci-Dayton, M., Jacobsen, C.J., Kirz, J., Neuhausler, U., Smith, G. and Yu, B. (1998). *Applications and Instrumentation Advances with the Stony Brook Scanning Transmission X-ray Microscope*. In *X-ray Microfocusing: Appli-*

- cations and Techniques* (I. McNulty, Ed.). *Proceedings of the SPIE*, Vol. 3449, 19–29. (SPIE, Bellingham, WA).
- Feser, M., Beetz, T., Carlucci-Dayton, M. and Jacobsen, C. (2000). Instrumentation advances and detector development with the Stony Brook scanning transmission X-ray microscope. In *X-ray Microscopy: Proceedings of the Sixth International Conference* (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.), (American Institute of Physics, Melville, NY).
- Feser, M., Jacobsen, C., Rehak, P., DeGeronimo, G., Holl, P. and Strüder, L. (2001). Novel integrating solid state detector with segmentation for scanning transmission soft x-ray microscopy. *X-ray Micro- and Nano-focusing: Applications and Techniques II* (I. McNulty, Ed.). *Proceedings of the SPIE*, Vol. 4499, 117–125. (SPIE, Bellingham, WA).
- Feser, M., Jacobsen, C., Rehak, P. and De Geronimo, G. (2003). Scanning transmission x-ray microscopy with a segmented detector. *J. Phys. IV* **104**, 529–534.
- Fischer, P., Schutz, G., Schmahl, G., Guttmann, P. and Raasch, D. (1996). Imaging of magnetic domains with the x-ray microscope at BESSY using x-ray magnetic circular dichroism. *Z. Phys. B* **101**, 313–316.
- Fischer, P., Eimüller, T., Schütz, G., Denbeaux, G., Pearson, A., Johnson, L., Attwood, D., Tsunashima, S., Kumazawa, M., Takagi, N., Köhler, M., Bayreuther, G. (2001a). Element-specific imaging of magnetic domains at 25 nm spatial resolution using soft x-ray microscopy. *Rev. Sci. Instrum.*, **72**(5), 2322–2324.
- Fischer, P., Eimüller, T., Schütz, G., Köhler, M., Bayreuther, G., Denbeaux, G. and Attwood, D. (2001b). Study of in-plane magnetic domains with magnetic transmission x-ray microscopy, *J. App. Phys.* **89**, 7159–7161.
- Fischer, P. (2003). Magnetic soft X-ray transmission microscopy, *Curr. Opin. Solid State Mater. Sci.* **7**, 173–179.
- Ford, T.W., Page, A.M., Foster, G.F. and Stead, A.D. (1992). Effects of soft x-ray irradiation on cell ultrastructure. In *Soft X-ray Microscopy* (C. Jacobsen and J. Trebes, Eds.). *Proceedings of the SPIE*, Vol. 1741, 325–332. (SPIE, Bellingham, WA).
- Foriel, J., Philippot, P., Susini, J., Dumas, P., Somogyi, A., Salome, M., Khodja, H., Menez, M., Fouquet, Y., Moreira, D. and Lopez-Garcia, P. (2004). High-resolution imaging of sulfur oxidation states, trace elements, and organic molecules distribution in individual microfossils and contemporary microbial filaments. *Geochimica et Cosmochimica Acta* **68**(7), 1561–1569.
- Gilbert, J.R., Pine, J., Kirz, J., Jacobsen, C., Williams, S., Buckley, C.J. and Rarback, H. (1992). Soft x-ray absorption imaging of whole wet tissue culture cells. In *X-ray Microscopy III* (A.G. Michette, G.R. Morrison and C.J. Buckley, Eds.). (Springer-Verlag, Berlin).
- Glaeser, R.M. (1971). Limitations to significant information in biological electron microscopy as a result of radiation damage. *J. Ultrastruct. Res.* **36**, 466–482.
- Gölz, P. (1992). Calculations on radiation dosages of biological materials in phase contrast and amplitude contrast x-ray microscopy. In *X-ray Microscopy III*. (A.G. Michette, G.R. Morrison and C.J. Buckley, Eds.). (Springer-Verlag, Berlin).
- Goodman, J.W. (1968). *Introduction to Fourier Optics*. (McGraw-Hill, San Francisco).
- Goodman, J.W. (1985). *Statistical Optics*. (John Wiley & Son, New York).
- Grimm, R., Singh, H., Rachel, R., Typke, D., Zillig, W. and Baumeister, W. (1998). Electron tomography of ice-embedded prokaryotic cells. *Biophys. J.* **74**, 1031–1042.

- Günther, S., Kolmakov, A., Kovac, J., Casalis, L., Gregoratti, L., Marsi, M. and Kiskinova, M. (1997). Scanning photoelectron microscopy of a bimetal/Si interface: Au coadsorbed on Ag/Si(111). *Surf. Sci.* **377**, 145–149.
- Günther, S., Kaulich, B., Gregoratti, L. and Kiskinova, M. (2002). Photoelectron microscopy and applications in surface and materials science. *Prog. Surf. Sci.* **70**(4–8), 187–260.
- Guttmann, P., Schneider, G., Robert-Nicoud, M., Niemann, B., Rudolph, D., Thieme, J., Jovin, T.M. and Schmahl, G. (1992). X-ray microscopic investigations on giant chromosomes isolated from salivary glands of *Chironomus thummi* larvae. In *X-ray Microscopy III*. (A.G. Michette, G.R. Morrison and C.J. Buckley, Eds.). (Springer-Verlag, Berlin).
- Guttmann, P., Niemann, B., Thieme, J., Hambach, D., Schneider, G., Wiesemann, U., Rudolph, D. and Schmahl, G. (2001). Instrumentation advances with the new X-ray microscopes at BESSY II. *Nucl. Inst. Methods Phys. Res. Sect. a* **467**, 849–852.
- Guttmann, P., Niemann, B., Rehbein, S., Knoechel, C., Rudolph, D. and Schmahl, G. (2003). The transmission X-ray microscope at BESSY II. *J. Phys. IV* **104**, 85–90.
- Haddad, W.S., McNulty, I., Trebes, J.E., Anderson, E.H., Levesque, R.A. and Yang, L. (1994). Ultra high resolution x-ray tomography. *Science* **266**, 1213–1215.
- Haelbich, R.-P. (1980). A scanning ultrasoft X-ray microscope with multilayer coated reflection optics: First test with synchrotron radiation around 60 eV photon energy. *Scanned Image Microscopy*. (E.A. Ash, Ed.). (Academic Press, London).
- Haelbich, R.-P., Staehr, W. and Kunz, C. (1980). A scanning ultrasoft x-ray microscope with large aperture reflection optics for use with synchrotron radiation. *Ann. NY Acad. Sci.* **342**, 148–157.
- Hambach, D., Schneider, G. and Gullikson, E. (2001). Efficient high-order diffraction of extreme-ultraviolet light and soft x-rays by nanostructured volume gratings. *Opt. Lett.* **26**(15), 1200–1202.
- Harp, G.R., Han, Z.-L. and Tonner, B.P. (1990). Spatially-resolved x-ray absorption near-edge spectroscopy of silicon in thin silicon-oxide films. *Phys. Scripta* **T31**, 23–27.
- Hayakawa, S., Iida, A., Aoki, S. and Gohshi, Y. (1989). Development of a scanning x-ray microprobe with synchrotron radiation. *Rev. Sci. Inst.* **60**, 2452–2455.
- Heck, J., Attwood, D., Ilse, W.M. and Anderson, E. (1998). Resolution determination in x-ray microscopy: An analysis of the effects of partial coherence and illumination spectrum. *J. X-ray Sci. Tech.* **8**, 95–104.
- Henke, B.L., Gullikson, E.M. and Davis, J.C. (1993). X-ray interactions: Photoabsorption, scattering, transmission and reflection at $E = 50\text{--}30,000\text{ eV}$, $Z = 1\text{--}92$. *Atom. Dat. Nucl. Dat. Tables* **54**, 181–342.
- Hettwer, M. and Rudolph, D. (1998). Fabrication of the x-ray condenser zone plate KZP7. In *X-ray Microscopy and Spectromicroscopy* (J. Thieme, G. Schmahl, D. Rudolph, E. Umbach, Eds.). (Springer, Berlin.)
- Hignette, O., Cloetens, P., Lee, W.-K., Ludwig, W. and Rostaing, G. (2003). Hard x-ray microscopy with reflecting mirrors status and perspectives of the ESRF technology. *J. Phys. IV* **104**, 231–234.
- Hitchcock, A.P., Morin, C., Heng, Y.M., Cornelius, R.M. and Brash, J.L. (2002). Towards practical soft X-ray spectromicroscopy of biomaterials. *J. Biomat. Sci.-Poly.* **13**(8), 919–937.
- Hitchcock, A.P., Araki, T., Ikeura-Sekiguchi, H., Iwata, N. and Tani, K. (2003). 3d chemical mapping of toners by serial section scanning transmission X-ray microscopy. *J. Phys. IV* **104**, 509–512.

- Hopkins, H.H. (1957). Applications of coherence theory in microscopy and interferometry. *J. Opt. Soc. Am.* **47**(6), 508–526.
- Horowitz, P. and Howell, J.A. (1972). A scanning x-ray microscope using synchrotron radiation. *Science* **178**, 608–611.
- Howells, M.R., Charalambous, P., He, H., Marchesini, S. and Spence, J.C.H. (2002). An off-axis zone-plate monochromator for high-power undulator radiation. In *Design and Microfabrication of Novel X-ray Optics* (D. Mancini, Ed.), *Proceedings of the SPIE*, Vol. 4783, 65–73 (SPIE, Bellingham, WA).
- Hubbell, J.H., Gimm, H.A. and Øverbø, I. (1980). Pair, triplet, and total atomic cross sections (and mass attenuation coefficients) for 1 MeV–100 GeV photons in elements $Z = 1$ –100. *J. Phys. Chem. Ref. Data* **9**, 1023–1147.
- Hunt, J.A. and Williams, D.B. (1991). Electron energy-loss spectrum-imaging. *Ultramicroscopy* **38**(1), 47–73.
- Im, M., Fischer, P., Eimuller, T., Denbeaux, G. and Shin, S. (2003). Magnetization reversal study of CoCrP alloy thin films on a nanogranular length scale using magnetic transmission soft X-ray microscopy. *Appl. Phys. Lett.*, **83**, 4589–4591.
- Ito, A., Shinohara, K., Nakano, H., Matsumura, T. and Kinoshita, K. (1996). Measurement of soft x-ray absorption spectra and elemental analysis in local regions of mammalian cells using an electronic zooming tube. *J. Microsc.* **181**, 54–60.
- Jacobsen, C. (1992). Making soft x-ray microscopy harder: Considerations for sub-0.1 micron resolution imaging at ~ 4 Å wavelengths. In *X-ray Microscopy III* (A.G. Michette, G.R. Morrison and C.J. Buckley, Eds.). (Springer-Verlag, Berlin).
- Jacobsen, C., Howells, M., Kirz, J. and Rothman, S. (1990). X-ray holographic microscopy using photoresists. *J. Opt. Soc. Amer. A* **7**, 1847–1861.
- Jacobsen, C., Williams, S., Anderson, E., Browne, M.T., Buckley, C.J., Kern, D., Kirz, J., Rivers, M. and Zhang, X. (1991). Diffraction-limited imaging in a scanning transmission x-ray microscope. *Opt. Comm.* **86**, 351–364.
- Jacobsen, C., Kirz, J. and Williams, S. (1992). Resolution in soft x-ray microscopes. *Ultramicroscopy* **47**, 55–79.
- Jacobsen, C., Medenwaldt, R. and Williams, S. (1998). A perspective on biological x-ray and electron microscopy. *X-ray Microscopy and Spectromicroscopy* (J. Thieme, G. Schmahl, E. Umbach and D. Rudolph, Eds.). (Springer-Verlag, Berlin).
- Jacobsen, C., Abend, S., Beetz, T., Carlucci-Dayton, M., Feser, M., Kaznacheyev, K., Kirz, J., Maser, J., Neuhäusler, U., Osanna, A., Stein, A., Vaa, C., Wang, Y., Winn, B. and Wirick, S. (2000a). Recent developments in scanning microscopy at Stony Brook. In *X-ray Microscopy: Proceedings of the Sixth International Conference* (American Institute of Physics, Berkeley, CA).
- Jacobsen, C., Flynn, G., Wirick, S. and Zimba, C. (2000b). Soft x-ray spectroscopy from image sequences with sub-100 nm spatial resolution. *J. Microsc.* **197**, 173–184.
- Jeanguillaume, C. and Colliex, C. (1989). Spectrum-image: The next step in EELS digital acquisition and processing. *Ultramicroscopy* **28**, 252–257.
- Jearanaikoon, S. and Abraham-Peskir, J. (2005). An x-ray microscopy perspective on the effect of glutaraldehyde fixation on cells. *J. Microsc.* **218**(2), 185–192.
- Jochum, L. and Meyer-Ilse, W. (1995). Partially coherent image formation with x-ray microscopes. *App. Opt.* **34**(22), 4944–4950.
- Joyeux, D., Lowenthal, S., Polack, F. and Bernstein, A. (1988). X-ray microscopy by holography at LURE. In *X-ray Microscopy II*. (D. Sayre, M.R. Howells, J. Kirz and H. Rarback, Eds.). (Springer-Verlag, Berlin).

- Kagoshima, Y., Yokoyama, Y., Nimi, T., Koyama, T., Tsusaka, Y., Matsui, J. and Takai, K. (2003). Hard X-ray phase contrast microscope for observing transparent specimens. *J. Phys. IV* **104**, 49–52.
- Kagoshima, Y., Ed. (2006). Proceedings of the 8th International Conference on X-ray Microscopy, Vol. 7, IPAP Conference Series, Tokyo.
- Kamijo, N., Suzuki, Y., Takano, H., Tamura, S., Yasumoto, M., Takeuchi, A. and Awaji, M. (2003). Microbeam of 100 keV x ray with a sputtered-sliced Fresnel zone plate. *Rev. Sci. Instr.* **74**(12), 5101–5104.
- Kamiya, K. (1963). Theory of Fresnel zone plate. *Sci. Light* **12**(2), 35–49.
- Kang, H., Stephenson, G., Liu, C., Conley, R., Macrander, A., Maser, J., Bajt, S. and Chapman, H. (2005). High-efficiency diffractive x-ray optics from sectioned multilayers. *App. Phys. Lett.* **86**, 151109.
- Kang, H.C., Maser, J., Stephenson, G.B., Liu, C., Conley, R., Macrander, A.T. and Vogt, S. (2006). Nanometer linear focusing of hard x-rays by a multilayer Laue lens. *Phys. Rev. Lett.* **96**, 127401.
- Kaulich, B., Wilhein, T., Di Fabrizio, E., Romanato, F., Altissimo, M., Cabrini, S., Fayard, B. and Susini, J. (2002). Differential interference contrast x-ray microscopy with twin zone plates. *J. Opt. Soc. Amer. A* **19**(4), 797–806.
- Kaulich, B., Bacescu, D., Cocco, D., Susini, J., Salomé, M., Dhex, O., David, C., Weitkamp, T., Fabrizio, E.D., Cabrini, S., Morrison, G., Charalambous, P., Thieme, J., Wilhein, T., Kovac, J., Podnar, M. and Kiscinova, M. (2003). Twinmic: A European twin microscope station combining full-field imaging and scanning microscopy. *J. Phys. IV* **104**, 103–107.
- Kawai, J., Takagawa, K., Fujisawa, S., Ektessabi, A. and Hayakawa, S. (2001). Microbeam XANES and X-ray fluorescence analysis of cadmium in kidney. *J. Trace Microprobe Tech.* **19**(4), 541–546.
- Kaznacheyev, K., Osanna, A., Jacobsen, C., Plashkevych, O., Vahtras, O., Ågren, H., Carravetta, V. and Hitchcock, A.P. (2002). Inner-shell absorption spectroscopy of amino acids. *J. Phys. Chem. A* **106**(13), 3153–3168.
- Kemner, K.M., Kelly, S.D., Lai, B., Maser, J., O'Loughlin, E.J., Sholto-Douglas, D., Cai, Z.H., Schneegurt, M.A., Kulpa, C.F. and Neilson, K.H. (2004). Elemental and redox analysis of single bacterial cells by X-ray microbeam analysis. *Science* **306**, 686–687.
- Kenney, J.M., Jacobsen, C., Kirz, J., Rarback, H., Cinotti, F., Thomlinson, W., Rosser, R. and Schidlovsky, G. (1985). Absorption microanalysis with a scanning soft x-ray microscope: Mapping the distribution of calcium in bone. *J. Microsc.* **138**, 321–328.
- Kenney, J.M., Morrison, G.R., Browne, M.T., Buckley, C.J., Burge, R.E., Cave, R.C., Charalambous, P.S., Duke, P.D., Hare, A.R., Hills, C.P.B., Michette, A.G., Ogawa, K. and Rogoyski, A.M. (1989). Evaluation of a scanning transmission x-ray microscope using undulator radiation at the SERC Daresbury Laboratory. *J. Phys. E* **22**, 234–238.
- Kern, D., Coane, P., Acosta, R., Chang, T.H.P., Feder, R., Houzgo, P., Molzen, W., Powers, J., Speth, A., Viswanathan, R., Kirz, J., Rarback, H. and Kenney, J. (1984). Electron beam fabrication and characterization of Fresnel zone plates for soft x-ray microscopy. In *Science with Soft X-Rays* (F.J. Himpsel and R.W. Klaffky, Eds.). *Proceedings of the SPIE*, Vol. 447, 204–213 (SPIE, Bellingham, WA).
- Kilcoyne, A.L.D., Tyliszczak, T., Steele, W.F., Fakra, S., Hitchcock, P., Franck, K., Anderson, E., Harteneck, B., Rightor, E.G., Mitchell, G.E., Hitchcock, A.P., Yang, L., Warwick, T. and Ade, H. (2003). Interferometer-controlled scanning transmission X-ray microscopes at the Advanced Light Source. *J. Synchrotron Radiat.* **10**(2), 125–136.

- King, P.L., Browning, R., Pianetta, P., Lindau, I., Keenlyside, M. and Knapp, G. (1989). Image-processing of multispectral x-ray photoelectron-spectroscopy images. *J. Vac. Sci. Tech. A* **7**(6), 3301–3304.
- Kinjo, Y., Shinohara, K., Ito, A., Nakano, H., Watanabe, M., Horiike, Y., Kikuchi, Y., Richardson, M.C. and Tanaka, K.A. (1994). Direct imaging in a water layer of human chromosome fibres composed of nucleosomes and their higher order structures by laser plasma X-ray contact microscopy. *J. Microsc.* **176**, 63–74.
- Kintner, E.C. (1978). Method for the calculation of partially coherent imagery. *App. Opt.* **17**(17), 2747.
- Kirkpatrick, P. (1949). The x-ray microscope. *Sci. Amer.* **180**, 44–47.
- Kirkpatrick, P. and Baez, A.V. (1948). Formation of optical images by x-rays. *J. Opt. Soc. Amer.* **38**, 766–774.
- Kirz, J. (1974). Phase zone plates for X rays and the extreme UV. *J. Opt. Soc. Amer.* **64**, 301–309.
- Kirz, J. (1980a). Mapping the distribution of particular atomic species. *Ann. NY Acad. Sci.* **342**, 273–287.
- Kirz, J. (1980b). Specimen damage considerations in biological microprobe analysis. *Scanning Electron Microscopy 1980*, **2**: 239–249 (Scanning Electron Microscopy International: Chicago).
- Kirz, J. and Sayre, D. (1980). Soft x-ray microscopy of biological specimens. In *Synchrotron Radiation Research* (H. Winick and S. Doniach, Eds.). (Plenum, New York).
- Kirz, J., Sayre, D. and Dilger, J. (1978). Comparative analysis of x-ray emission microscopies for biological specimens. *Ann. NY Acad. Sci.* **306**, 291–305.
- Kirz, J., Jacobsen, C. and Howells, M. (1995). Soft x-ray microscopes and their biological applications. *Q. Rev. Biophys.* **28**(1), 33–130.
- Kiskinova, M. (2003). Chemical, electronic and magnetic properties of surfaces and interfaces probed with X-ray microscopes at ELETTRA. *J. Phys. IV* **104**, 453–458.
- Ko, C.H., Klausner, R., Wei, D.-H., Chan, H.-H. and Chuang, T.J. (1998). The soft x-ray scanning photoemission microscopy project at SRRC. *J. Synchrotron Radiat.* **5**, 299–304.
- Kogelnik, H. (1969). Coupled wave theory for thick hologram gratings. *Bell Syst. Tech. J.* **48**, 2909–2930.
- Koprinarov, I.N., Hitchcock, A.P., McCrory, C.T. and Childs, R.F. (2002). Quantitative mapping of structured polymeric systems using singular value decomposition analysis of soft x-ray images. *J. Phys. Chem. B* **106**, 5358–5364.
- Krasnoperova, A.A., Xiao, J., Cerrina, F., Di Fabrizio, E., Luciani, L., Figliomeni, M., Gentili, M., Yun, W., Lai, B. and Gluskin, E. (1993). Fabrication of hard x-ray phase zone plate by x-ray lithography. *J. Vac. Sci. Tech. B* **11**, 2588–2591.
- Krause, M.O. (1979). Atomic radiative and radiationless yields for K and L shells. *J. Phys. Chem. Ref. Data* **8**, 307–327.
- Labrenz, M., Druschel, G.K., Thomsen-Ebert, T., Gilbert, B., Welch, S.A., Kemner, K.M., Logan, G.A., Summons, R.E., De Stasio, G., Bond, P.L., Lai, B., Kelly, S.D. and Banfield, J.F. (2000). Formation of sphalerite (ZnS) deposits in natural biofilms of sulfate-reducing bacteria. *Science* **290**, 1744–1747.
- Lai, B., Yun, W.-B., Legnini, D., Xiao, Y., Chrzas, J., Viccaro, P.J., White, V., Bajikar, S., Denton, D., Cerrino, F., Di Fabrizio, E., Gentili, M., Grella, L. and Baciocchi, M. (1992). Hard x-ray phase zone plate fabricated by lithographic techniques. *App. Phys. Lett.* **61**, 1877–1879.
- Larabell, C.A. and Le Gros, M.A. (2004). X-ray tomography generates 3-D reconstructions of the yeast, *Saccharomyces cerevisiae*, at 60-nm resolution. *Molec. Biol. Cell* **115**, 957–962.

- Larson, B.C., Yang, W., Ice, G.E., Budai, J.D. and Tischler, J.Z. (2002). Three-dimensional X-ray structural microscopy with submicrometre resolution. *Nature* **415**, 887–890.
- Lawrence, J.R., Swerhone, G.D.W., Leppard, G.G., Araki, T., Zhang, X., West, M.M. and Hitchcock, A.P. (2003). Scanning transmission X-ray, laser scanning, and transmission electron microscopy mapping of the exopolymeric matrix of microbial biofilms. *App. Env. Microbiol.* **69**(9), 5543–5554.
- Lehr, J. (1997). 3D x-ray microscopy: Tomographic imaging of mineral sheaths of bacteria *Leptothrix ochracea* with the Göttingen x-ray microscope at BESSY. *Optik* **104**(4), 166–170.
- Lengeler, B., Schroer, C.G., Benner, B., Gerhardus, A., Günzler, T.F., Kuhlmann, M., Meyer, J. and Zimprich, C. (2002). Parabolic refractive X-ray lenses. *J. Synchrotron Radiat.* **9**, 119–124.
- Lerotic, M., Jacobsen, C., Schäfer, T. and Vogt, S. (2004). Cluster analysis of soft x-ray spectromicroscopy data. *Ultramicroscopy* **100**, 35–57.
- Lerotic, M., Jacobsen, C., Gillow, J.B., Francis, A.J., Wirick, S., Vogt, S. and Maser, J. (2005). Cluster analysis in soft X-ray spectromicroscopy: Finding the patterns in complex specimens. *J. Electron Spectrosc. Rel. Phenomena* **144–147C**, 1137–1143.
- Levine, Z.H., Kalukin, A.R., Kuhn, M., Frigo, S.P., McNulty, I., Retsch, C.C., Wang, Y., Arp, U., Lucatorto, T.B., Ravel, B.D. and Tarrio, C. (2000). Tomography of an integrated circuit interconnect with an electromigration void. *J. App. Phys.* **87**(9), 4483–4488.
- Lindaas, S., Howells, M., Jacobsen, C. and Kalinovsky, A. (1996). X-ray holographic microscopy by means of photoresist recording and atomic-force microscope readout. *J. Opt. Soc. Amer. A* **13**(9), 1788–1800.
- Ma, C.J., Tohno, S., Kasahara, M. and Hayakawa, S. (2004). Determination of the chemical properties of residues retained in individual cloud droplets by XRF microprobe at SPring-8. *Nucl. Inst. Methods Phys. Res. Section B* **217**(4), 657–665.
- Magowan, C., Brown, J.T., Liang, J., Heck, J., Coppel, R.L., Mohandas, N. and Meyer-Ilse, W. (1997). Intracellular structures of normal and aberrant *Plasmodium falciparum* malaria parasites imaged by soft x-ray microscopy. *Proc. Nat. Acad. Sci. USA* **94**, 6222–6227.
- Marchesini, S., He, H., Chapman, H.N., Hau-Riege, S.P., Noy, A., Howells, M.R., Weierstall, U. and Spence, J.C.H. (2003). X-ray image reconstruction from a diffraction pattern alone. *Phys. Rev. B* **68**, 140101.
- Martin, L.C. (1966). *The Theory of the Microscope*. (American Elsevier, New York).
- Maser, J. (1994). Theoretical description of the diffraction properties of zone plates with small outermost zone width. *X-ray Microscopy IV*. (V. Aristov and A. Erko, Eds.). (Bogorodski Pechatnik, Chernogolovka, Russia).
- Maser, J., Osanna, A., Wang, Y., Jacobsen, C., Kirz, J., Spector, S., Winn, B. and Tennant, D. (2000). Soft x-ray microscopy with a cryo STXM: I. Instrumentation, imaging, and spectroscopy. *J. Microsc.* **197**, 68–79.
- Maser, J., Stephenson, G.B., Vogt, S., Yun, W., Macrander, A., Kang, H.C., Liu, C. and Conley, R. (2004). Multilayer Laue lenses as high-resolution x-ray optics. In *Design and Microfabrication of Novel X-ray Optics II* (A.A. Snigirev, D.C. Mancini, Eds). *Proceedings of the SPIE*, Vol. 5539, 185–194 (SPIE, Bellingham, WA).
- McNulty, I., Kirz, J., Jacobsen, C., Anderson, E., Kern, D. and Howells, M. (1992). High-resolution imaging by Fourier transform x-ray holography. *Science* **256**, 1009–1012.

- McNulty, I., Lai, B., Maser, J., Paterson, D., Evans, P., Heald, S., Ice, G., Isaacs, E., Rivers, M. and Sutton, S. (2003a). X-ray microscopy at the advanced photon source. *Synchrotron Radiat. News* **16**, 34–39.
- McNulty, I., Paterson, D., Arko, J., Erdmann, M., Frigo, S.P., Goetze, K., Ilinski, P., Krapf, N., Mooney, T., Retsch, C.C., Stampfl, A.P.J., Vogt, S., Wang, Y. and Xu, S. (2003b). The 2-ID-B intermediate-energy scanning X-ray microscope at the APS. *J. Phys. IV* **104**, 11–15.
- Medalia, O., Weber, I., Frangakis, A., Nicastro, D., Gerisch, G. and Baumeister, W. (2002). Macromolecular architecture in eukariotic cells visualized by cryoelectron tomography. *Science* **298**, 1209–1213.
- Meyer-Ilse, W., Attwood, D. and Koike, M. (1994). The X-ray Microscopy Resource Center at the Advanced Light Source. *Synchrotron Radiation in the Biociences*. (B. Chance, D. Deisenhofer, S. Ebashi, D.T. Goodhead, J.R. Helliwell, H.E. Huxley, T. Iizuka, J. Kirz, T. Mitsui, E. Rubenstein, N. Sakabe, T. Sasaki, G. Schmahl, H. Sturhmann, K. Wüthrich and G. Zaccai, Eds.). (Oxford University Press, New York).
- Meyer-Ilse, W., Denbeaux, G., Johnson, L.E., Bates, W., Lucero, A. and Anderson, E.H. (2000a). The high resolution x-ray microscope XM-1. *X-ray Microscopy: Proceedings of the Sixth International Conference*. (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- Meyer-Ilse, W., Warwick, T. and Attwood, D., Eds. (2000). *X-ray Microscopy: Proceedings of the Sixth International Conference*. (American Institute of Physics, Melville, NY).
- Meyer-Ilse, W., Hamamoto, D., Nair, A., Lelievre, S.A., Denbeaux, G., Johnson, L., Pearson, A.L., Yager, D., Legros, M.A. and Larabell, C.A. (2001). High resolution protein localization using soft X-ray microscopy. *J. Microsc.* **201**, 395–403.
- Miao, J.W., Ishikawa, T., Johnson, B., Anderson, E.H., Lai, B. and Hodgson, K.O. (2002). High resolution 3D x-ray diffraction microscopy. *Phys. Rev. Lett.* **89**(8), 088303.
- Michette, A.G. (1986). *Optical Systems for Soft X Rays*. (Plenum, New York).
- Michette, A.G., Morrison, G.R. and Buckley, C.J., Eds. (1992). *X-ray Microscopy III*. Springer Series in Optical Sciences. (Springer-Verlag, Berlin).
- Michette, A., Buckley, C., Pfauntsch, S., Arnot, N., Wilkinson, J., Wang, Z., Khaleque, N. and Dermody, G. (2000). The Kings College Laboratory Scanning X-ray Microscope, In *Proceedings of the Sixth International Conference on Synchrotron Radiation Instrumentation* (American Institute of Physics Melville, NY).
- Mimura, H., Yamauchi, K., Yamamura, K., Kubota, A., Matsuyama, S., Sano, Y., Ueno, K., Endo, K., Nishino, Y., Tamasaku, K., Yabashi, M., Ishikawa, T. and Mori, Y. (2004). Image quality improvement in a hard X-ray projection microscope using total reflection mirror optics. *J. Synchrotron Radiat.* **11**, 343–346.
- Mollenstedt, G. and Gaukler, K.H., Eds. (1969). *X-ray Optics and X-ray Microanalysis*. (Springer-Verlag, Berlin).
- Mondal, P.K. and Slansky, S. (1970). Influence de la position de l’anneau de phase sur le contraste de l’image dans un objectif à contraste de phase. *App. Opt.* **9**(8), 1879–1882.
- Morrison, G.R. (1989). Some aspects of quantitative x-ray microscopy. In *X-ray Instrumentation in Medicine and Biology, Plasma Physics, Astrophysics, and Synchrotron Radiation*. (R. Benattar, Ed.), *Proceedings of the SPIE*, Vol. 1140, 41–49 (SPIE, Bellingham, WA).
- Morrison, G.R., Bridgwater, S., Browne, M.T., Burge, R.E., Cave, R.C., Charalambous, P.S., Foster, G.F., Hare, A.R., Michette, A.G., Morris, D.,

- Taguchi, T. and Duke, P. (1989). Development of x-ray imaging at the Daresbury SRS. *Rev. Sci. Instr.* **60**, 2464–2467.
- Morrison, G.R., Eaton, W.J. and Charalambous, P. (2002). STXM imaging with a configured detector. *J. Phys. IV* **104**, 547–550.
- Murphy, J., White, D., MacDowell, A. and Wood, O. (1993). Synchrotron radiation sources and condensers for projection lithography. *App. Opt.* **32**(34), 6920–6928.
- Murray, C.E., Yan, H.-F., Noyan, I.C., Cai, Z. and Lai, B. (2005). High-resolution strain mapping in heteroepitaxial thin film features. *J. App. Phys.* **98**, 013504.
- Myneni, S.C.B., Tokunaga, T.K. and Brown, G.E. Jr. (1997). Abiotic selenium redox transformations in the presence of Fe(II,III) oxides. *Science* **278**, 1106–1109.
- Neuhäusler, U. Soft X-ray spectromicroscopy on hydrated colloidal and environmental science specimens, PhD thesis, Georg-August-Universität zu Göttingen, Germany, 1999.
- Neuhäusler, U., Jacobsen, C., Schulze, D., Stott, D. and Abend, S. (2000). A specimen chamber for soft x-ray spectromicroscopy on aqueous and liquid samples. *J. Synchrotron Radiat.* **7**, 110–112.
- Neuhäusler, U., Schneider, G., Ludwig, W., Meyer, M., Zschech, E. and Hambach, D. (2003). X-ray microscopy in Zernike phase contrast mode at 4 keV photon energy with 60 nm resolution. *J. Phys. D: App. Phys.* **36**, A79–A82.
- Ng, W., Ray-Chaudhuri, A.K., Crossley, S., Crossley, D., Gong, C., Guo, J., Hansen, R.W.C., Margaritondo, G., Cerrina, F., Underwood, J.H., Perera, R.C.C. and Kortright, J. (1990). The photoemission spectromicroscope multiple-application x-ray imaging undulator microscope (MAXIMUM). *J. Vacuum Sci. Tech.* **A8**, 2563–2565.
- Niemann, B. (1987). The Göttingen scanning x-ray microscope. In *Soft x-ray optics and technology*. (E.-E. Koch and G. Schmahl, Eds.), *Proceedings of the SPIE*, Vol. 733, 422–427 (SPIE, Bellingham, WA).
- Niemann, B. (1998). High numerical-aperture x-ray condensers for transmission x-ray microscopes. In *X-ray Microscopy and Spectromicroscopy*. (J. Thieme, G. Schmahl, E. Umbach and D. Rudolph, Eds.). (Springer-Verlag, Berlin).
- Niemann, B., Rudolph, D. and Schmahl, G. (1974). Soft x-ray imaging zone plates with large zone numbers for microscopic and spectroscopic applications. *Opt. Comm.* **12**, 160–163.
- Niemann, B., Rudolph, D. and Schmahl, G. (1976). X-ray microscopy with synchrotron radiation. *App. Opt.* **15**, 1883–1884.
- Niemann, B., Guttman, P., Hilkenbach, R., Thieme, J. and Meyer-Ilse, W. (1988). The Göttingen scanning x-ray microscope. In *X-ray Microscopy II*. (D. Sayre, M.R. Howells, J. Kirz and H. Rarback, Eds.). (Springer-Verlag, Berlin).
- Niemann, B., Guttman, P., Hambach, D., Schneider, G., Weiss, D. and Schmahl, G. (1999). The condenser-monochromator with dynamical aperture synthesis for the TXM at an undulator beam line at BESSY II. In *X-ray Microscopy: Sixth International Conference*. (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- Ortega, R., Deves, G., Fayard, B., Salome, M. and Susini, J. (2003). Combination of synchrotron radiation X-ray microprobe and nuclear microprobe for chromium and chromium oxidation states quantitative mapping in single cells. *Nucl. Inst. Methods Phys. Res. B* **210**, 325–329.
- Osanna, A. and Jacobsen, C. (2000). Principal component analysis for soft x-ray spectromicroscopy. In *X-ray Microscopy: Proceedings of the Sixth International*

- Conference. (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- O'Toole, E., Wray, G., Kremer, J. and McIntosh, J.R. (1993). High voltage cryo-microscopy of human blood platelets. *J. Struct. Biol.* **110**, 55–66.
- Parsons, D., Ed. (1978). *Short Wavelength Microscopy*. (Academy of Sciences, New York, NY).
- Parsons, D., Ed. (1980). *Ultrasoft X-ray Microscopy: Its Application to Biological and Physical Sciences*. (Academy of Sciences, New York).
- Pattee, H.H., Cosslett, V.E. and Engström, A., Eds. (1963). *X-ray Optics and X-ray Microanalysis*. (Academic Press, New York).
- Paunesku, T., Rajh, T., Wiederrecht, G., Maser, J., Vogt, S., Stojićević, N., Protić, M., Lai, B., Oryhon, J., Thurnauer, M. and Woloschak, G. (2003). Biology of TiO₂ oligonucleotide nanocomposites. *Nature Mater* **2**, 343–346.
- Pecher, K., McCubbery, D., Kneedler, E., Rothe, J., Bargar, J., Meigs, G., Cox, L., Nealson, K. and Tonner, B. (2003). Quantitative charge state analysis of manganese biominerals in aqueous suspension using scanning transmission X-ray microscopy (STXM). *Geochimica et Cosmochimica Acta* **67**(6), 1089–1098.
- Peuker, M. (2001). High-efficiency nickel phase zone plates with 20 nm minimum outermost zone width. *App. Phys. Lett.* **78**(15), 2208–2210.
- Pine, J. and Gilbert, J. (1992). Live cell specimens for x-ray microscopy. In *X-ray Microscopy III*. (A.G. Michette, G.R. Morrison and C.J. Buckley, Eds.). (Springer-Verlag, Berlin).
- Polack, F., Joyeux, D., Feser, M., Phalippou, D., Carlucci-Dayton, M., Kaznacheyev, K. and Jacobsen, C. (2000). Demonstration of phase contrast in scanning transmission x-ray microscopy: Comparison of images obtained at NSLS X-1A with numerical simulations. In *X-ray Microscopy: Proceedings of the Sixth International Conference*. (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- Rarback, H., Kenney, J., Kirz, J. and Xie, X.S. (1980). Scanning X-ray Microscopy—First Tests with Synchrotron Radiation. In *Scanned Image Microscopy* (E.A. Ash, Ed.). (Academic Press: London).
- Rarback, H., Kenney, J.M., Kirz, J., Howells, M.R., Chang, P., Coane, P.J., Feder, R., Houzago, P.J., Kern, D.P. and Sayre, D. (1984). Recent results from the Stony Brook scanning microscope. In *X-ray Microscopy* (G. Schmahl and D. Rudolph, Eds.). (Springer-Verlag, Berlin).
- Rarback, H., Shu, D., Feng, S.C., Ade, H., Kirz, J., McNulty, I., Kern, D.P., Chang, T.H.P., Vladimirsky, Y., Iskander, N., Attwood, D., McQuaid, K. and Rothman, S. (1988). Scanning x-ray microscope with 75-nm resolution. *Rev. Sci. Instr.* **59**, 52–59.
- Rayleigh, L. (1888). Wave theory of light. In *Scientific Papers by Lord Rayleigh*, Vol. III, 47–187 (Dover, New York).
- Rebonato, R., Ice, G.E., Habenschuss, A. and Bilello, J.C. (1989). High-resolution microdiffraction study of notch-tip deformation in Mo single crystals using X-ray synchrotron radiation. *Philos. Mag. A* **60**(5), 571–583.
- Reimer, L. (1984). *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*. (Springer-Verlag, Berlin).
- Rightor, E.G., Urquhart, S.G., Hitchcock, A.P., Ade, H., Smith, A.P., Mitchell, G.E., Priester, R.D., Aneja, A., Appel, G., Wilkes, G. and Lidy, W.E. (2002). Identification and quantitation of urea precipitates in flexible polyurethane foam formulations by X-ray spectromicroscopy. *Macromolecules* **35**(15), 5873–5882.
- Ritman, E.L., Jorgensen, S.M., Lundo, P.E., Thomas, P.J., Dunsmuir, J.H., Romero, J.C., Tumerc, R.T. and Bolander, M.E. (1997). Synchrotron-based

- micro-CT of in situ biological basic functional units and their integration. In *Developments in X-ray Tomography* (U. Bonse, Ed.). *Proceedings of the SPIE*, Vol. 3149, 13–24 (SPIE, Bellingham, WA).
- Rivers, M.L., Sutton, S.R. and Smith, J.V. (1988). A synchrotron X-ray-fluorescence microprobe. *Chem. Geol.* **70**(1–2), 179.
- Rose, A. (1946). Unified approach to performance of photographic film, television pickup tubes, and human eye. *J. Soc. Motion Pict. Eng.* **47**, 273–294.
- Rudolph, D., Niemann, B., Schmahl, G. and Christ, O. (1984). The Göttingen x-ray microscope and x-ray microscopy experiments at the BESSY storage ring. In *X-ray Microscopy*. (G. Schmahl and D. Rudolph, Eds.). (Springer-Verlag, Berlin).
- Rudolph, D., Schmahl, G. and Niemann, B. (1990). Amplitude and phase contrast in x-ray microscopy. In *Modern Microscopies*. (P.J. Duke and A.G. Michette, Eds.). (Plenum, New York).
- Saleh, B.E.A. (1979). Optical bilinear transformations: General properties. *Opt Acta* **26**(6), 777–799.
- Saloman, E.B. and Hubbell, J.H. (1987). Critical analysis of soft X-ray cross section data. *Nucl. Inst. Methods Phys. Res. A* **255**, 38–42.
- Sayre, D. (1972). Proposal for the utilization of electron beam technology in the fabrication of an image forming device for the soft x-ray region. Report RC 3974 (#17965) (IBM Research, Yorktown Heights, NY).
- Sayre, D., Kirz, J., Feder, R., Kim, D.M. and Spiller, E. (1977a). Transmission microscopy of unmodified biological materials: Comparative radiation dosages with electrons and ultrasoft x-ray photons. *Ultramicroscopy* **2**, 337–341.
- Sayre, D., Kirz, J., Feder, R., Klim, D.M., Spiller, E. (1977b) Potential operating region for ultrasoft x-ray microscopy of biological specimens. *Science*, **196**, 1339–1340.
- Sayre, D., Howells, M.R., Kirz, J. and Rarback, H. (1988). *X-ray Microscopy II*. (Springer-Verlag, Berlin).
- Schäfer, T., Claret, F., Bauer, A., Griffault, L., Ferrage, E. and Lanson, B. (2003). Natural organic matter (NOM)-clay association and impact on Callovo-Oxfordian clay stability in high alkaline solution: Spectromicroscopic evidence. *J. Phys. IV* **104**, 413–416.
- Schäfer, T., Buckau, G., Artinger, R., Kim, J.I., Geyer, S., Wolf, M., Bleam, W.F., Wirick, S. and Jacobsen, C. (2005). Origin and mobility of fulvic acids in the Gorleben aquifer system: Implications from isotopic data and carbon/sulfur XANES. *Org. Geochem.* **36**, 567–582.
- Scharf, J.-G. and Schneider, G. (1999). Ultrastructural characterization of isolated rat Kupffer cells by transmission x-ray microscopy. *J. Microsc.* **193**(3), 250–256.
- Scheinost, A.C., Kretzschmar, R., Christl, I. and Jacobsen C. (2001). Carbon Group Chemistry of Humic and Fulvic Acid: A Comparison of C-1s NEXAFS and ¹³C-NMR Spectroscopies. In *Humic Substances: Structures, Models, and Functions*. (E.A. Ghabbor and G. Davies, Eds.) (Royal Society of Chemistry, London).
- Schmahl, G. and Rudolph, D. (1969). Lichtstarke Zonenplatten als abbildende Systeme für weiche Röntgenstrahlung. *Optik* **29**, 577–585.
- Schmahl, G. and Rudolph, D., Eds. (1984). *X-ray Microscopy*. Springer Series in Optical Sciences. (Springer-Verlag, Berlin).
- Schmahl, G. and Rudolph, D. (1987). Proposal for a phase contrast x-ray microscope. *X-ray Microscopy: Instrumentation and Biological Applications* (P.C. Cheng and Jan, G.J. Eds.). (Springer-Verlag, Berlin).

- Schmahl, G., Rudolph, D. and Niemann, B. (1980). Imaging and scanning soft x-ray microscopy with zone plates. In *Scanned Image Microscopy* (E.A. Ash, Ed.). (Academic Press: London).
- Schmahl, G., Rudolph, D., Guttman, P. and Christ, O. (1984). Zone plate lenses for x-ray microscopy. In *X-ray Microscopy*. (G. Schmahl and D. Rudolph, Eds.). (Springer-Verlag, Berlin).
- Schmahl, G., Rudolph, D., Schneider, G., Guttman, P. and Niemann, B. (1994). Phase contrast x-ray microscopy studies. *Optik* **97**, 181–182.
- Schmahl, G., Rudolph, D., Guttman, P., Schneider, G., Thieme, J. and Niemann, B. (1995). Phase contrast studies of biological studies with the x-ray microscope at BESSY. *Rev. Sci. Instr.* **66**(2), 1282–1286.
- Schneider, G. (1997). Zone plates with high efficiency in high orders of diffraction described by dynamical theory. *App. Phys. Lett.* **71**(16), 2242–2244.
- Schneider, G. (1998). Cryo x-ray microscopy with high spatial resolution in amplitude and phase contrast. *Ultramicroscopy* **75**, 85–104.
- Schneider, G., Schliebe, T. and Aschoff, H. (1995). Cross-linked polymers for nanofabrication of high-resolution zone plates in nickel and germanium. *J. Vac. Sci. Tech. B* **13**(6), 2809–2812.
- Schneider, G., Anderson, E., Vogt, S., Knöchel, C., Wliss, D., Leberos, M. and Larabell, C. (2002a). Computed tomography of cryogenic cells. *Surface Rev Lett.* **9**, 177–183.
- Schneider, G., Denbeaux, G., Anderson, E.H., Bates, B., Pearson, A., Meyer, M.A., Zschech, E., Hambach, D. and Stach, E.A. (2002b). Dynamical x-ray microscopy investigation of electromigration in passivated inlaid Cu interconnect structures. *App. Phys. Lett.* **81**(14), 2535–2537.
- Schneider, G., Meyer, M.A., Denbeaux, G., Anderson, E., Bates, B., Pearson, A., Knochel, C., Hambach, D., Stach, E.A. and Zschech, E. (2002c). Electromigration in passivated Cu interconnects studied by transmission x-ray microscopy. *J. Vac. Sci. Tech. B* **20**(6), 3089–3094.
- Schneider, G., Denbeaux, G., Anderson, E., Bates, W., Salmassi, F., Nachimuthu, P., Pearson, A., Richardson, D., Hambach, D., Hoffmann, N., Hasse, W. and Hoffmann, K. (2003). Electromigration in integrated circuit interconnects studied by X-ray microscopy. *Nucl. Instr. Methods Phys. Res. B* **199**, 469–474.
- Schroer, C.G. and Lengeler, B. (2005). Focusing hard x rays to nanometer dimensions by adiabatically focusing lenses. *Phys. Rev. Lett.* **94**, 054802.
- Schumacher, M., Christl, I., Scheinost, A.C., Jacobsen, C. and Kretzschmar, R. (2005). C-1s XANES spectroscopy of humic substances: Comparison with ¹³C-NMR results. *Eur. J. Soil Sci.* in press.
- Scott, D., Duewer, F., Kamath, S., Lyon, A., Trapp, D., Wang, S. and Yun, W. (2004). *A Novel X-Ray Microtomography System with High Resolution and Throughput for Non-Destructive 3D Imaging of Advanced Packages*. Thirtieth International Symposium on Testing and Failure Analysis. (ASM International).
- Shapiro, D., Thibault, P., Beetz, T., Elser, V., Howells, M., Jacobsen, C., Kirz, J., Lima, E., Miao, H., Neiman, A. and Sayre, D. (2004). Biological Imaging by Soft X-ray Diffraction Microscopy. *Proc. Nat. Acad. Sci. USA* **102**, 15343–15346.
- Shastri, S.P., Maser, J.M., Lai, B. and Tys, J. (2001). Microfocusing of 50-keV undulator radiation with two stacked zone plates. *Optics comm.* **197**, 9–14.
- Shaver, D.C., Flanders, D.C., Ceglie, N.M. and Smith, H.I. (1980). X-ray zone plates fabricated using electron-beam and x-ray lithography. *J. Vac. Sci. Tech.* **16**, 1626–1630.

- Sheppard, C. and Wilson, T. (1980). Fourier imaging of phase information in scanning and conventional optical microscopes. *Philos. Trans. R. Soc.* **295**(1415), 513–536.
- Shin, H.J., Lee, M.K., Kim, G., Hong, C., Lee, J., Kim, J., Park, S., Roh, Y. and Leong, K. (2003). A scanning photoelectron microscope for materials science spectromicroscopy at the Pohang light source. *J. Phys. IV* **104**, 67–70.
- Shu, D., Siddons, D.P., Rarback, H. and Kirz, J. (1988). Two-dimensional laser interferometric encoder for the soft x-ray scanning microscope at the NSLS. *Nucl. Inst. Methods Phys. Res. A* **266**, 313–317.
- Simpson, M.J. and Michette, A.G. (1983). The effects of manufacturing inaccuracies on the imaging properties of Fresnel zone plates. *Opt. Acta* **30**, 1455–1462.
- Smith, A.P., Ade, H., Koch, C.C. and Spontak, R.J. (2001). Cryogenic mechanical alloying as an alternative strategy for the recycling of tires. *Polymer* **42**(9), 4453–4457.
- Snigirev, A., Snigireva, I., Kohn, V., Kuznetsov, S. and Schelokov, I. (1995). On the possibility of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Rev. Sci. Instr.* **66**(12), 5486–5492.
- Snigirev, A., Kohn, V., Snigireva, I. and Lengeler, B. (1996). A compound refractive lens for focusing high energy x-rays. *Nature* **384**, 49–51.
- Soh, Y.A., Evans, P.G., Cai, Z., Lai, B., Kim, C.Y., Aeppli, G., Mathur, N.D., Blamire, M.G. and Isaacs, E.D. (2002). Local mapping of strain at grain boundaries in colossal magnetoresistive films using x-ray microdiffraction. *J. App. Phys.* **91**(10), 7742–7744.
- Solomon, D., Lehmann, J., Kinyangi, J., Liang, B. and Schäfer, T. (2005). Carbon K-edge NEXAFS and FTIR-ATR spectroscopic investigation of organic carbon speciation in soils. *Soil Sci. Soc. Amer. J.* **67**, 1721–1731.
- Solymar, L. and Cooke, D.J. (1981). *Volume Holography and Volume Gratings*. (Academic Press, London).
- Sparks, J., C.J. (1980). X-ray fluorescence microprobe for chemical analysis. *Synchrotron Radiation Research*. (H. Winick and S. Doniach, Eds.). (Plenum Press, New York).
- Spector, S., Jacobsen, C. and Tennant, D. (1997). Process optimization for production of sub-20 nm soft x-ray zone plates. *J. Vac. Sci. Tech. B* **15**(6), 2872–2876.
- Spiller, E. (1972). Low-loss reflection coatings using absorbing materials. *App. Phys. Lett.* **20**, 365–367.
- Stead, A.D., Cotton, R.A., Page, A.M., Dooley, M.D. and Ford, T.W. (1992). Visualization of the effects of electron microscopy fixatives on the structure of hydrated epidermal hairs of tomato (*Lycopersicum peruvianum*) as revealed by soft x-ray microscopy. In *Soft X-ray Microscopy* (C. Jacobsen and J. Trebes, Eds.). *Proceedings of the SPIE*, Vol. 174, 351–362 (SPIE, Bellingham, WA).
- Steinbrecht, R.A. and Zierold, K. (1987). *Cryotechniques in Biological Electron Microscopy*. (Springer-Verlag, Berlin).
- Stöhr, J. (1992). *NEXAFS Spectroscopy*. (Springer-Verlag, Berlin).
- Stöhr, J., Wu, Y., Hermsmeier, B.D., Samant, M.G., Harp, G.R., Koranda, S., Dunham, D., and Tonner, B.P. (1993). Element-specific magnetic microscopy with circularly polarized x-rays. *Science* **259**, 658–661.
- Stoll, H., Puzic, A., vonWayenberge, B., Fischer, P., Raabe, J., Buess, M., Huang, T., Hollinger, R., Back, C.H., Weiss, D. and Denbeaux, G. (2004). High resolution imaging of fast magnetization dynamics in magnetic nanostructures. *App. Phys. Lett.* **84**, 3328–3330.
- Susini, J., Barrett, R., Kaulich, B., Oestreich, S. and Salome, M. (2000). The X-ray microscopy facility at the ESRF: A status report. In *X-ray Microscopy: Proceed-*

- ings of the Sixth International Conference* (W. Meyer-Ilse, T. Warwick, D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- Susini, J., Joyeux, D. and Polak, F., Eds. (2003). X-ray Microscopy 2002. *J. Phys. IV* (EDP Sciences, Les Ulis).
- Suzuki, Y., Takeuchi, A., Takano, H., Ohigashi, T. and Takenaka. (2001). Diffraction-limited microbeam with Fresnel zone plate optics in hard x-ray regions. *J. App. Phys.* **40**, 1508–1510.
- Suzuki, Y., Awaji, M., Takeuchi, A., Takano, H., Uesugi, K., Kohmura, Y., Kamijo, N., Yasumoto, M. and Tamura, S. (2003). Hard X-ray microscopy activities at SPring-8. *J. Phys. IV* **104**, 35–40.
- Takano, H., Uesugi, K., Takeuchi, A., Takai, K. and Suzuki, Y. (2003). High-Sensitive Imaging with Scanning Transmission Hard X-ray Microscope. *J. Phys. IV* **104**, 41–44.
- Takemoto, K., Mizuno, T., Yoshikawa, T., Mishibata, H., Ueki, T., Uyama, T., Miyoshi, T., Sawa, D., Matsumoto, T., Wada, N., Onoda, H., Kojima, K., Niemann, B., Hettwer, M., Rudolph, D., Anderson, E., Attwood, D., Kern, D.P., Iwasaki, H. and Kihara, H. (2003). Xray Microscopy in Ritsumeikan Synchrotron Radiation Center. *J. Phys. IV* **104**, 57–61.
- Tamura, S., Yasumoto, M., Kamijo, N., Suzuki, Y., Awaji, M., Takeuchi, A., Takano, H. and Handa, K. (2002). Development of a multilayer Fresnel zone plate for high-energy synchrotron radiation X-rays by DC sputtering deposition. *J. Synchrotron Radiat.* **9**, 154–159.
- Tang, M.-T., Yin, G.-C., Song, Y.-F., Chen, J.-H., Tsang, K.-L., Liang, K.S., Chen, F.-R., Diewer, F. and Yun, W. (2006). Hard X-ray microscopy with sub-30 nm spatial resolution at NSRRC. In *Proceedings of the 8th International Conference on X-ray Microscopy* (Y. Kagoshima, Ed.). (IPAP, Tokyo).
- Tatchyn, R., Csonka, P.L. and Lindau, I. (1982). Outline of a variational formulation of zone-plate theory. *J. Opt. Soci. Amer.* **72**, 1630–1638.
- Tatchyn, R., Csonka, P.L. and Lindau, I. (1984). The constant-thickness zone plate as a variational problem. *Opt. Acta* **31**, 729–733.
- Taylor, K.A. and Glaeser, R.M. (1976). Electron microscopy of frozen hydrated biological specimens. *J. Ultrastruct. Res.* **55**, 448–456.
- Tennant, D., Jackel, L.D., Howard, R.E., Hu, E.L., Grabbe, P., Capik, R.J. and Schneider, B.S. (1981). 25 nm features patterned with trilevel e-beam resist. *J. Vac. Sci. Tech.* **19**, 1304–1307.
- Tennant, D.M., Gregus, J.E., Jacobsen, C. and Raab, E.L. (1991). Construction and test of phase zone plates for x-ray microscopy. *Opt. Lett.* **16**, 621–623.
- Tennant, D., Spector, S., Stein, A. and Jacobsen, C. (2000). Electron beam lithography of Fresnel zone plates using a rectilinear machine and trilayer resists. In *X-ray Microscopy: Proceedings of the Sixth International Conference* (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- Thieme, J. (1988). Theoretical investigations of imaging properties of zone plates and zone plate systems using diffraction theory. In *X-ray Microscopy II* (D. Sayre, M.R. Howells, J. Kirz and H. Raback, Eds.), Springer Series in Optical Sciences. (Springer-Verlag, Berlin).
- Thieme, J. and Niemeyer, J. (1998). Interaction of colloidal soil particles, humic substances and cationic detergents studied by x-ray microscopy. *Prog. Colloid Polymer Sci.* **111**, 193–201.
- Thieme, J., Wilhein, T., Guttman, P., Niemeyer, J., Jacob, K.-H. and Dietrich, S. (1994). Direct visualization of iron and manganese accumulating micro-organisms by x-ray microscopy. In *X-ray Microscopy IV*. (V. Aristov and A. Erko, Eds.). (Bogorodskii Pechatnik, Chernogolovka, Russia).

- Thieme, J., Schmahl, G., Umbach, E. and Rudolph, D. Eds. (1998). *X-ray Microscopy and Spectromicroscopy*. (Springer-Verlag, Berlin).
- Thieme, J., Schneider, G. and Knochel, C. (2003). X-ray tomography of a microhabitat of bacteria and other soil colloids with sub-100 nm resolution. *Micron* **34**(6–7), 339–344.
- Thompson, A.C., Underwood, J.H., Wu, Y., Giaque, R.D., Jones, K.W. and Rivers, M.L. (1988). Elemental measurements with an X-ray microprobe of biological and geological samples with femtogram sensitivity. *Nucl. Inst. Methods Phys. Res. A* **266**, 318–323.
- Thompson, B.J. (1969). Image formation with partially coherent light. In *Progress in Optics* (E. Wolf, Ed.). (North Holland, Amsterdam).
- Tonner, B.P., Droubay, T., Denlinger, J., Meyer-Ilse, W., Warwick, T., Rothe, J., Kneedler, E., Pecher, K., Nealson, K. and Grundl, T. (1999). Soft x-ray spectroscopy and imaging of interfacial chemistry in environmental specimens. *Surface Interface Anal.* **27**, 247–258.
- Tsutsuka, Y., Yokoyama, K., Takeda, S., Takai, K., Kagoshima, Y. and Matsui, J. (2001). Hyogo beamline at SPring-8: multiple station beamline with the TROIKA concept. *Nucl. Instruments Meth. A*, 467–468, 670–673.
- Twining, B.S., Baines, S.B., Fisher, N.S., Maser, J., Vogt, S., Jacobsen, C., Tovar-Sanchez, A. and Sañudo-Wilhelmy, S.A. (2003). Quantifying trace elements in individual aquatic protist cells with a synchrotron x-ray fluorescence microprobe. *Anal. Chem.* **75**, 3806–3816.
- Tyliszczak, T., Warwick, T., Kilcoyne, A., Fakra, S., Yoon, D.S., Brown, G., Andrews, S., Chembrolu, V., Strachan, J. and Acremann, Y. (2004). Soft X-ray scanning transmission microscope working in an extended energy range at the Advanced light source. In *Proceedings of the Eighth International Conference on Synchrotron Radiation Instrumentation*. (American Institute of Physics, Melville, NY).
- Uesugi, K., Tsuchiyama, A., Yasuda, H., Nakamura, M., Nakano, T., Suzuki, Y. and Yagi, N. (2003). Micro-tomographic imaging for material sciences at BL47XU in Spring-8. *J. Phys. IV* **104**, 45–48.
- Uggerhoj, E. and Abraham-Peskir, J. (2000). X-ray microscopy in Aarhus. In *X-ray Microscopy: Proceedings of the Sixth International Conference* (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- Urquhart, S.G. and Ade, H. (2002). Trends in the carbonyl core (C 1s, O 1s) $\rightarrow$ $\pi^*c = o$ transition in the near-edge X-ray absorption fine structure spectra of organic molecules. *J. Phys. Chem. B* **106**(34), 8531–8538.
- Urquhart, S.G., Hitchcock, A.P., Smith, A.P., Ade, H.W., Lidy, W., Rightor, E.G. and Mitchell, G.E. (1999). NEXAFS spectromicroscopy of polymers: Overview and quantitative analysis of polyurethane polymers. *J. Electron Spectrosc. Rel. Phenomena* **100**, 119–135.
- Vogt, S., Chapman, H.N., Jacobsen, C. and Medenwaldt, R. (2001a). Dark field X-ray microscopy: The effects of condenser/detector aperture. *Ultramicroscopy* **87**(1–2), 25–44.
- Vogt, S., Jager, M., Schneider, G., Schulze, E., Saumweber, H., Rudolph, D. and Schmahl, G. (2001b). Visualizing specific nuclear proteins in eukaryotic cells using soft X-ray microscopy. *Nucl. Inst. Methods Phys. Res. A* **467**, 1312–1314.
- Wagner, D., Maser, J., Lai, B., Cai, Z.H., Barry, C.E., Bentrup, K.H.Z., Russell, D.G. and Bermudez, L.E. (2005). Elemental analysis of *Mycobacterium avium*-, *Mycobacterium tuberculosis*-, and *Mycobacterium smegmatis*-containing phagosomes indicates pathogen-induced microenvironments within the host cell's endosomal system. *J. Immunol.* **174**(3), 1491–1500.

- Wang, S., Duewer, F., Kamath, S., Kelly, C., Lyon, A., Nill, K., Pombo, P., Scott, D., Trapp, D., Yun, W., Neogi, S., Kuhn, M., Bennet, C., Coon, P. and Yan, S. (2002). A transmission x-ray microscope (TXM) for non-destructive 3D imaging of ICs at sub-100nm resolution. In *Proceedings of the 28th International Symposium for Testing and Failure Analysis*. (AMS International).
- Wang, Y., Jacobsen, C., Maser, J. and Osanna, A. (2000). Soft x-ray microscopy with a cryo STXM: II. Tomography. *J. Microsc.* **197**, 80–93.
- Warwick, T., Ade, H., Hitchcock, A.P., Padmore, H., Rightor, E.G. and Tonner, B.P. (1997). Soft x-ray spectromicroscopy development for materials science at the Advanced Light Source. *J. Electron Spectrosc. Rel. Phenomena* **84**, 85–98.
- Warwick, T., Ade, H., Kilcoyne, D., Kraitscher, M., Tyliszczak, T., Fakra, S., Hitchcock, A., Hitchcock, P. and Padmore, H. (2002). A new bend-magnet beamline for scanning transmission X-ray microscopy at the Advanced Light Source. *J. Synchrotron Radiat.* **9**, 254–257.
- Warwick, T., Andresen, N., Comins, J., Kaznacheyev, K., Kortright, J., McKean, J., Padmore, H., Shuh, D., Stevens, T. and Tyliszczak, T. (2004). New implementation of an SX700 undulator beamline at the advanced light source. In *Proceedings of the Eighth International Conference on Synchrotron Radiation Instrumentation*. (American Institute of Physics, Melville, NY).
- Weisemann, U., Thieme, J., Guttman, P., Frueke, R., Rehbein, S., Neimann, B., Rudolph, D. and Schmahl, G. (2003). First results of the new scanning transmission X-ray microscope at BESSY II. *J. Phys. IV* **104**, 95–98.
- Weiss, D., Schneider, G., Niemann, B., Guttman, P., Rudolph, D. and Schmahl, G. (2000). Computed tomography of cryogenic biological specimens based on x-ray microscopic images. *Ultramicroscopy* **84**, 185–197.
- Weitkamp, T., Drakopoulos, M., Leitenberger, W., Raven, C., Schroer, C., Simionivici, A., Snigireva, I. and Snigirev, A. (2000). High resolution X-ray imaging and tomography at the ESRF beamline ID22. In *X-ray Microscopy: Proceeding of the Sixth International Conference on X-ray Microscopy* (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- White, D.L., O.R.W. II, Bjorkholm, J.E., Spector, S., MacDowell, A.A. and LaFontaine, B. (1995). Modification of the coherence of undulator radiation. *Rev. Sci. Instr.* **66**(2), 1930–1933.
- Wiesemann, U., Thieme, J., Guttman, P., Niemann, B., Rudolph, D. and Schmahl, G. (2000). The new scanning transmission x-ray microscope at BESSY II. In *X-ray Microscopy: Proceedings of the Sixth International Conference* (W. Meyer-Ilse, T. Warwick and D. Attwood, Eds.). (American Institute of Physics, Melville, NY).
- Williams, G.J., Pfeifer, M.A., Vartanyants, I.A. and Robinson, I.K. (2003). Three-dimensional imaging of microstructure in Au nanocrystals. *Phys. Rev. Lett.* **90**, 175501.
- Williams, S., Zhang, X., Jacobsen, C., Kirz, J., Lindaas, S., Hof, J.V.T. and Lamm, S.S. (1993). Measurements of wet metaphase chromosomes in the scanning transmission x-ray microscope. *J. Microsc.* **170**, 155–165.
- Wilson, T. and Sheppard, C. (1981). The halo effect of image processing by spatial frequency filtering. *Optik* **59**, 119–123.
- Wilson, T. and Sheppard, C. (1984). *Theory and Practice of Scanning Optical Microscopy*. (Academic Press, London).
- Winn, B., Ade, H., Buckley, C., Feser, M., Howells, M., Hulbert, S., Jacobsen, C., Kaznacheyev, K., Kirz, J., Osanna, A., Maser, J., McNulty, I., Miao, J., Oversluizen, T., Spector, S., Sullivan, B., Wang, Y., Wirick, S. and Zhang, H. (2000). Illumination for coherent soft X-ray applications: The new X1A beamline at the NSLS. *J. Synchrotron Radiat.* **7**, 395–404.

- Wolter, H. (1952). Spiegelsysteme streifenden Einfalls als abbildende Optiken für Röntgenstrahlen. *Ann. Phys.* **10**, 94–114, 286.
- Wood, R.W. (1898). Phase-reversal zone-plates, and diffraction telescopes. *Philos. Mag.* **45**, 511–522.
- Yamamoto, A., Masaki, R., Guttman, P., Schmahl, G. and Kihara, H. (1998). Studies on intracellular structures of COS cells by x-ray microscopy. *J. Synchrotron Radiat.* **5**, 1105–1107.
- Yamamura, K., Yamauchi, K., Mimura, H., Sano, Y., Saito, A., Endo, K., Souvorov, A., Yabashi, M., Tamasaku, K., Ishikawa, T. and Mori, Y. (2003). Fabrication of elliptical mirror at nanometer-level accuracy for hard x-ray focusing by numerically controlled plasma chemical vaporization machining. *Rev. Sci. Instr.* **74**(10), 4549–4553.
- Yi, Y., Cho, S., Noh, M., Wang, C., Jeong, K. and Shin, H. (2005). Characterization of surface chemical states of a thick insulator: Chemical state imaging on MgO surface. *J. J. App. Phys.* **44**, 861–864.
- Yin, G.-C., Tang, M.T., Song, Y.-F., Chen, F.-R., Liang, K.S., Duewer, F.W., Yun, W., Ko, C.-H., Shieh, H.-P.D. (2006). An energy-tunable transmission x-ray microscope for differential contrast imaging with near 60-nm resolution tomography. *Appl. Phys. Lett.* (accepted).
- Yoon, T.H., Johnson, S.B., Benzerara, K., Doyle, C.S., Tyliczszak, T., Shuh, D.K. and Brown, G.E. (2004). *In situ* characterization of aluminum-containing mineral-microorganism aqueous suspensions using scanning transmission X-ray microscopy. *Langmuir* **20**(24), 10361–10366.
- Youn, H., Baik, S. and Chang, C. (2005). Hard X-ray microscopy with a 130 nm spatial resolution. *Rev. Sci. Instr.* **76**, 23702.
- Yun, W. (2005). Private communication.
- Yun, W., Pratt, S.T., Miller, R.M., Cai, Z., Hunter, D.B., Jarstfer, A.G., Kemner, K.M., Lai, B., Lee, H.R., Legnini, D.G., Rodrigues, W. and Smith, C.I. (1998). X-ray imaging and microspectroscopy of plants and fungi. *J. Synchrotron Radiat.* **5**, 1390–1395.
- Yun, W., Lai, B., Krasnoperova, A.A., Di Fabrizio, E., Cai, Z., Cerrina, F., Chen, Z., Gentili, M. and Gluskin, E. (1999). Development of zone plates with a blazed zone profile for hard x-ray applications. *Rev. Sci. Instr.* **70**, 3537–3541.
- Zeitler, E. and Thomson, M.G.R. (1970). Scanning transmission electron microscopy. *Optik* **31**, 258–280, 359–366.
- Zhang, X., Jacobsen, C. and Williams, S. (1992). Image enhancement through deconvolution. In *Soft X-ray Microscopy* (C. Jacobsen and J. Trebes, Eds.). *Proceedings of the SPIE*, Vol. 1741, 251–259 (SPIE, Bellingham, WA).
- Zhang, X., Balhorn, R., Mazrimas, J. and Kirz, J. (1996). Measuring DNA to protein ratios in mammalian sperm head by XANES imaging. *J. Struc. Biol.* **116**, 335–344.
- Zhu, S., Liu, Y., Rafailovich, M.H., Sokolov, J., Gersappe, D., Winesett, A. and Ade, H. (1999). Confinement-induced miscibility in polymer blends. *Nature* **400**, 49–51.

Part III

NEAR-FIELD SCANNING PROBES

Scanning Probe Microscopy in Materials Science

Maxim P. Nikiforov and Dawn A. Bonnell

1 Introduction

The quest toward understanding the behavior of condensed matter has relied on measuring structure, bonding, and properties at increasingly local levels. This has driven advances in techniques that probe both soft and hard materials directly as well as indirectly. Many of these advances are described in other chapters of this volume. While structure and bonding-based probes have accessed molecular and atomic scales for decades, local determination of properties has been elusive. The emergence of scanning probes filled this gap to some extent.

There are three major classes of scanning probe techniques that access electronic, magnetic, optical, and mechanical properties. *Scanning tunneling microscopy* (STM) was the first and is based on electrons tunneling between a metal tip and a sample. The distance sensitivity of tunneling imparts intrinsically high spatial resolution and the voltage dependence yields local density of states. It is, however, applicable only to conducting materials. Another class of techniques is based on local optical responses induced and/or detected with a very fine optical fiber. These techniques, e.g., *near field optical microscopy*, find extensive application in organic and biological systems.

The focus of this chapter will be those probes that exploit the interactions of a small tip with a surface that are detected via the properties of a cantilever to which the tip is attached. The original cantilever probe, *atomic force microscopy* (AFM), is based on van der Waals interactions at the tip/surface junction. As a cantilever is mechanically oscillated near its resonant frequency (usually 10–500 kHz) and, at relatively small sample-tip separations (usually 0.5–100 nm), the van der Waals interaction causes a force that alters the oscillation. Cantilever motion is detected with laser reflection into a photodiode. If this measurement is made at every point as the tip is scanned across a surface, and a signal is used to maintain a profile at constant force, the topographic structure of the surface is mapped. It was almost immediately understood that other interactions can be detected with this scheme and, furthermore, that these interactions might be distinguished by their

distance dependence. For example, at a 10-nm sample/tip separation the van der Waals interactions can be used to determine the topographic structure, while at 200 nm an electrostatic force (EFM) or magnetic force (MFM) would dominate the measurement.

The utility of local probes is illustrated by the fact that even though the field is relatively young, upward of 2700 papers per year are published that cite “AFM or STM” as a key word.¹ Several monographs have summarized the state of this field and various books are available that provide an introduction to the field and general overviews.^{2–5} Most applications utilize SPM as a straightforward qualitative mapping tool. Some researchers interested in complex behavior of solids have examined fundamental tip–surface interactions and extended SPM to probe local electronic transport, dielectric, ferroelectric, and magnetic properties. Rather than address conventional STM, AFM, MFM, or EFM, this chapter will describe recent advances in nanometer probes of complex properties, highlighting potential insight, as well as remaining challenges. First, factors leading to atomic resolution imaging based on forces will be described along with some example applications. This is followed by summaries of approaches to probe spatially localized electrical and dielectric properties. Finally we will present a view of where this segment of the field is headed.

2 Imaging at Atomic Resolution with Force Interactions

There are a number of so-called “imaging modes” in AFM based on various combinations of force detection and feedback mechanism. Confusion can arise since there is no universal naming convention among microscope suppliers. Generally the cantilever is oscillating. From a fundamental perspective the force regimes categorize three imaging modes: **contact AFM** [the tip applies a small force (1–10 nN) normal to the surface of the sample], **intermittent contact AFM** (the oscillating cantilever tip is brought close to the sample so that it barely hits, or “taps” the sample at the bottom of the excursion), and **noncontact AFM** (the tip never contacts the surface). Superimposed upon this scheme are the tip responses that can be monitored to detect the tip–surface interaction: oscillation amplitude, frequency, or phase. The instrument configurations are illustrated in Figure 14–1. Atomic resolution imaging can be achieved in conventional AFM in contact mode for some materials and in noncontact mode under certain conditions. The latter is a relatively recent advance and offers the potential to investigate materials properties at the atomic scale, which is of importance not only from a fundamental, but also from a technological, point of view.

2.1 Operational Principles of Noncontact Atomic Force Microscopy

Noncontact atomic force microscopy (NC-AFM) is the general name for the group of techniques in which the tip of the cantilever oscillates in close proximity to the surface but never makes contact. Several review articles present details of imaging mechanisms, which we summarize

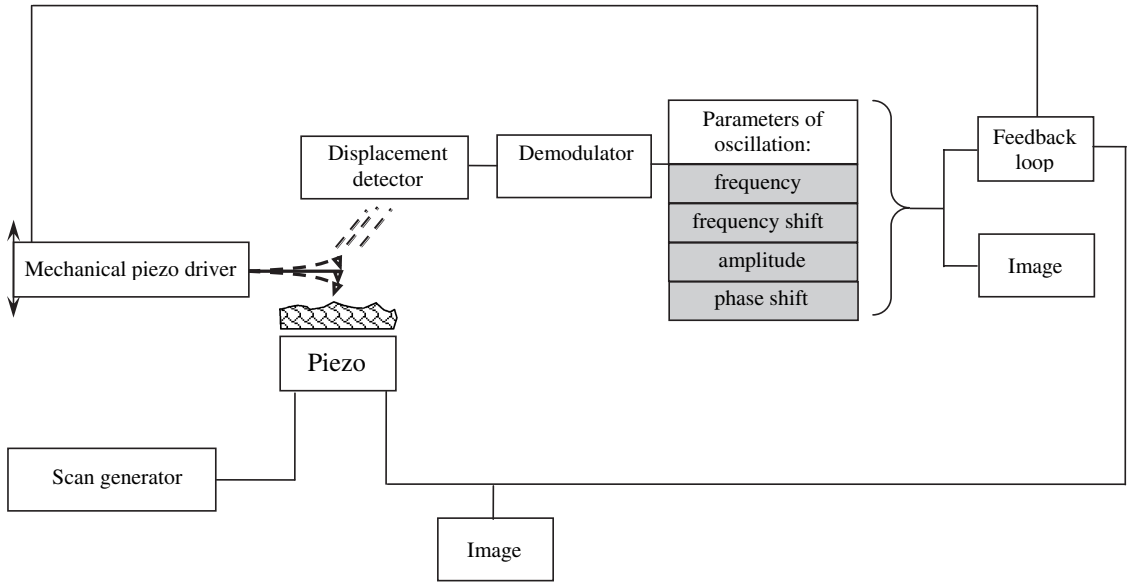


Figure 14–1. Basic block diagram for surface imaging with the NC-AFM. Varying one of the oscillation parameters and maintaining the others constant leads to the different NC-AFM techniques described in the text.

here.^{6–8} It should be noted that only specific types of NC-AFM (frequency modulated NC-AFM) achieve atomic resolution for the simple reason that the interactions between the sample and tip, i.e., long-range van der Waals and electrostatic forces, are power law functions of order 2. This is not sufficiently sensitive to track distances of fractions of angstroms. If, however, the geometry is configured such that the tip experiences short-range van der Waals and/or bonding interactions during a substantial part of the cantilever oscillation, sensitivity is enhanced. In this situation the local force between the sample and tip is the sum of the electrostatic, F_{elec} , van der Waals, F_{vdW} , and bonding F_{bonding} :

$$F_{\text{tot}} = F_{\text{elec}} + F_{\text{vdW}} + F_{\text{bonding}}$$

Figure 14–2 illustrates the range over which these forces operate based on the following estimations/assumptions.

$$F_{\text{tot}} = -\frac{q_{\text{surf}}q_{\text{tip}}}{4\pi\epsilon_0 r^2} + \frac{1}{2}\frac{\partial C}{\partial r}V^2 + -\frac{AR}{6r^2} + F_{\text{shortrange}}$$

where q_{surf} is the electrostatic charge accumulated on the surface, q_{tip} is the electrostatic charge accumulated on the tip apex, ϵ_0 is the dielectric constant of the vacuum, r is the distance from the tip apex to the surface, C is the capacitance between the tip and surface, V is the electrostatic potential difference between the tip and surface, A is the Hamaker constant, and R is the tip radius.

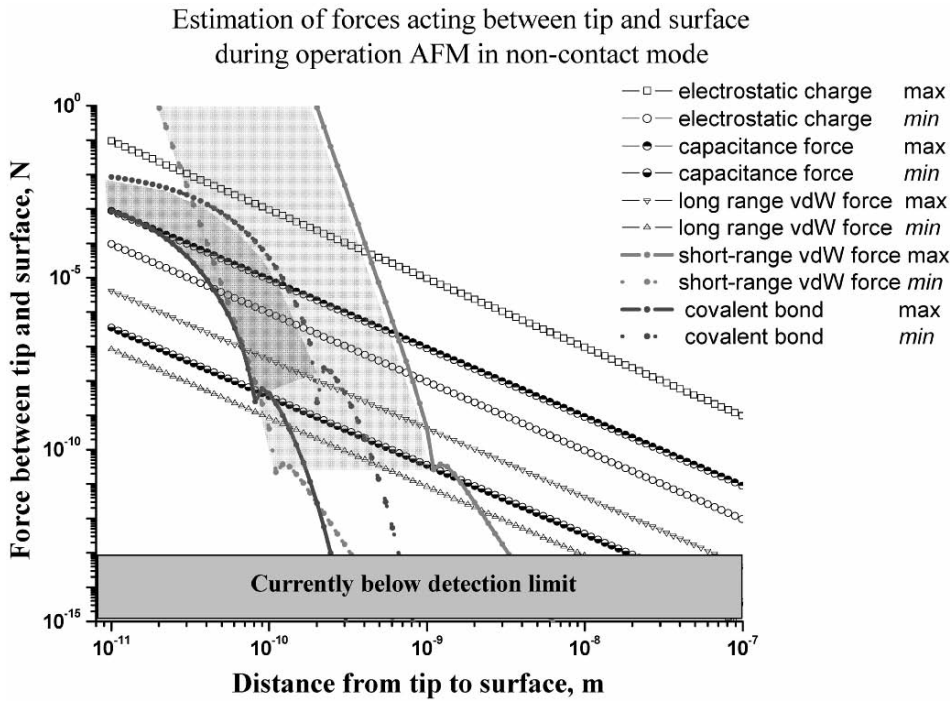


Figure 14–2. Distance dependencies of short- and long-range forces. Forces acting between the tip and surface during AFM operation in the noncontact mode are estimated.

Short-range forces play a significant role for separations less than 10 nm. For an illustration of the magnitude of the various forces we consider two types of short-range interactions: van der Waals and covalent bonding. The nature of short-range van der Waals forces is dipole–dipole interactions, which can be approximated by a Lenard–Jones potential.

$$F_{\text{srvdW}}(r) = 24 \frac{E_{\text{bond}}}{r} \left[2 \left(\frac{r_{\text{eq}}}{r} \right)^{12} - \left(\frac{r_{\text{eq}}}{r} \right)^6 \right]$$

where E_{bond} is the equilibrium bond energy (usual values 0.1–0.005 eV) and r_{eq} is the equilibrium distance between atoms (usual values 1–3 Å). A fair approximation for covalent bonding is the Morse potential. Lim showed that the Morse potential could be written as a function of two parameters (E_{bond} , r_{eq}):⁹

$$F_{\text{Morse}}(r) = \frac{12D}{r_{\text{eq}}} \left[\exp \left(-12 \frac{r - r_{\text{eq}}}{r_{\text{eq}}} \right) - \exp \left(-6 \frac{r - r_{\text{eq}}}{r_{\text{eq}}} \right) \right]$$

The typical parameters used in these calculations are $r_{\text{eq}} = 0.8 \text{ Å}$ and $E_{\text{bond}} = 1 \text{ eV}$ for the Morse *min* and $r_{\text{eq}} = 2 \text{ Å}$ and $E_{\text{bond}} = 10 \text{ eV}$ for the Morse *max*.

Figure 14–2 shows that $F_{\text{electrostatic}}$, $F_{\text{capacitance}}$, and F_{vdW} (large-range forces) are attractive for any tip–surface distances. At the same time short-range van der Waals forces and covalent forces (Morse potential) are attractive at large distances and repulsive at small distances. The repulsive part of the interactions is shaded with light gray for van der Waals forces and with dark gray for Morse forces in Figure 14–2. Although these considerations are based on rather simple approximations, Figure 14–2 illustrates that for short-range interactions to be the same order of magnitude as long-range interactions, a requirement for high resolution, oscillations should be within tens of nanometers of the surface.

Despite the fact that relatively simple theories qualitatively explain experimental results, a complete self-consistent theory that relates contrast to atomic structure has not been developed. Hofer et al.⁸ concluded that although consistent methods for simulating the basic aspects of scanning probe microscopy (SPM) exist, no model can reproduce all features of the experimental tip–surface interaction. Each model depends crucially on a set of assumptions, which is for scanning probe microscopy in general:

1. The ground-state properties of a system are equal to the properties at finite temperature (e.g., in an ambient environment).
2. There is a hierarchy of interactions that allows the separation of different effects in the theoretical models (no inclusion of, for example, cumulative or time-dependent interactions).

For *scanning tunneling microscopy* specifically:

1. Macroscopic interactions do not affect the tunneling current.
2. The resistance in the STM circuit is due only to the tunnel barrier.
3. The current cross section is centered at the apex atom of the tip.
4. Current flow does not change the properties of a system.

The last point has recently been analyzed by Todorov et al.,¹⁰ who found that the current flow slightly changes the position of the surface atoms.

For *scanning force microscopy* (SFM) specifically:

1. Charge transfer between the tip and surface is not a significant component of the interactions for an insulating surface.
2. The effects of the system electronics, such as apparent dissipation,^{6,11,12} do not affect the physics of the tip–surface interaction.

There has been one study, the adsorption of formate ions on TiO_2 (110),¹³ in which a chemical force model was used to establish chemical identification fairly conclusively. Theoretical interpretation of STM images helped to interpret noncontact AFM images of the same surface. Recent STM and noncontact (NC)-SFM theoretical modeling¹⁴ has confirmed the original experimental interpretation. Thus, the combined use of STM and AFM can resolve issues of contrast interpretation; however, by the nature of STM it is limited to surfaces that can be made to conduct. Further development of the theory of scanning force microscopy is needed.

2.2 NC-AFM Techniques

Mechanical oscillations of the cantilever lie at the heart of NC-AFM. Forces acting between the tip and surface change the oscillation parameters and these parameters are sensitive to atomic scale resolution. During operation three sets of parameters play the most important roles: cantilever-related (spring constant of the cantilever, the eigenfrequency of the cantilever, the quality factor of the cantilever), oscillation-related (amplitude, frequency shift), and image-related (either constant separation or constant height operation mode) parameters. Cantilever-related parameters are predetermined by the tip and are not variable during scanning. Changes in the other parameters result in different AFM techniques, as shown in Table 14–1. It should be noted that NC-AFM experiments are usually carried out in high vacuum or in liquid to avoid water condensation between the tip and surface.

2.2.1 Amplitude-Modulated AFM (AM-AFM)

As the name implies, in **AM-AFM** the amplitude of the tip oscillation is monitored while frequency shift and separation are kept constant by feedback. Usually AM-AFM is done at the resonant frequency of the cantilever; since this is the maximum oscillation amplitude. This mode is analogous to constant current STM, despite the difference in the physics of image formation. To date, atomic resolution has not been reported using this technique even on atomically smooth surfaces. A possible explanation is that the sensitivity is low because the amplitude depends on the tip–surface force averaged over the oscillation cycle [Eq. (1)].¹⁵

$$A \cong A_0 \left[1 - 4 \left(\frac{\langle F_{ts} \rangle}{F_0} \right)^2 \right]^{1/2} \tag{1}$$

where $\langle F_{ts} \rangle$ is the average force over the oscillation cycle and F_0 is the driving force. According to Eq. (1) AM-AFM senses the change of average force over the oscillation cycle at constant separation. The typical driving force might be estimated as $F_0 \sim kA_0 \sim 10 \text{ N/m } 50 \text{ nm} = 5 \times 10^{-7} \text{ N}$. The change in average force that produces atomic resolution is usually on the order of 10^{-8} – 10^{-10} N . Thus, the smaller the oscillation amplitude, the better the tip “senses” the surface because of the decrease in F_0 . To achieve atomic resolution (vertical $\sim 0.01 \text{ nm}$ and lateral $\sim 0.1 \text{ nm}$) the oscillation amplitude must be kept on the order of 1–10 nm. Small oscillation amplitude can be achieved either by decreasing the driving force or increasing the spring constant. A decrease of driving force

Table 14–1. Operational modes in noncontact AFM.

	Amplitude of tip oscillations	Frequency shift of tip oscillations	Tip–surface separation
AM-AFM	<i>Monitored (varied)</i>	Constant	Constant
FM-AFM (1)	Constant	<i>Monitored (varied)</i>	Constant
FM-AFM (2)	Constant	Constant	<i>Monitored (varied)</i>

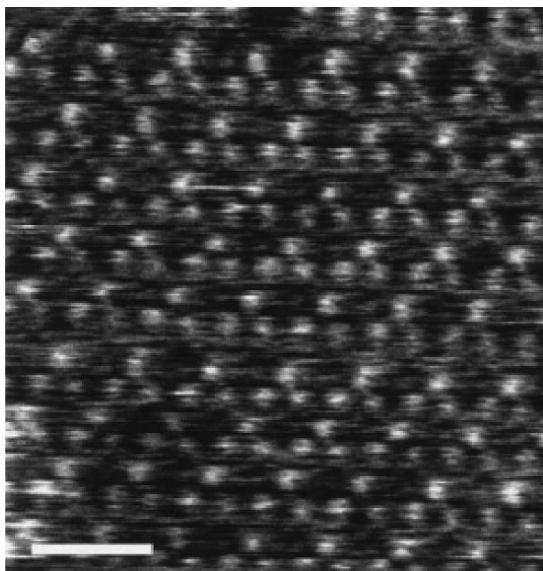


Figure 14–3. AM-AFM topographic image of the extracellular purple membrane surface in buffer solution. The scale mark is 10 nm. [Courtesy of Moller et al. Reprinted with permission from *Biophysical Journal*, 77(2), 1150–1158, 1999.]

leads to a decrease in the oscillation stability. A substantial increase of the spring constant is not easy due to manufacturing issues. At this time silicon cantilevers with spring constants from 0.1 to 100 N/m are on the market.

With spatial resolution on the order of ~ 2 nm, AM-AFM is ideal for many biological applications. In addition, the relatively small force minimizes destructive imaging of soft samples.¹⁶ Figure 14–3 illustrates AM-AFM on a purple membrane of *Halobacterium salinarum* stain ET1001, isolated as described in Oesterhelt and Stoekenius¹⁷ and imaged in 300 mM KCl, pH 7.8, 10 mM Tris-HCl. The authors observed the topography of the extracellular purple membrane surface, which exhibits a trimeric structure protruding 0.4 ± 0.1 nm above a lipid bilayer. The trimers are arranged in a trigonal lattice of 6.2 ± 0.2 nm length. In this experiment high lateral resolution (1.1–1.2 nm) was achieved. More examples of biological application AM-AFM are found in a recent review,⁶ which also describes the detailed theory of AM-AFM.

Similar resolution has been demonstrated on inorganic surfaces. Resolution close to atomic level was demonstrated on calcite with a modified AM-AFM technique. F.M. Ohnesorge^{18,19} operated the microscope in the separation region close to that for snap-into-contact instability. In this region feedback with an inverted sign provides reasonably stable images of high resolution.

2.2.2 Frequency-Modulated AFM (FM-AFM)

There are two FM-AFM techniques (FM-AFM(1) and FM-AFM(2) in Table 14–1). For these two techniques the oscillation amplitude is constant and either the frequency shift of oscillation (in constant height

mode) or the change in piezo movement (in constant force gradient mode) is varied. Atomic resolution was obtained almost simultaneously by Kitamura and Iwatsuki and Giessibl in 1995 using FM-AFM. Both groups operated the microscope in constant force gradient mode. The difference in the approaches was in the cantilever excitation: Kitamura and Iwatsuki used constant excitation mode and Giessibl used an automatic gain control of the excitation signal to maintain constant oscillation amplitude. Figure 14–4 shows one of the first images of the (7×7) reconstruction on silicon (111) obtained by Kitamura and Iwatsuki. The authors noted that in this early attempt the image stability is poor, but subsequent improvement in the frequency stability of the demodulator resulted in routine high-quality imaging of this surface. At this point the (7×7) reconstruction on silicon (111) has become a standard for NC-AFM calibration. From a practical perspective the quality of an NC-AFM image depends not only on the surface quality, but also on the quality of tip. Tips sputtered with an ion gun often provide the best results.

Note that an FM-AFM image is not only an aesthetically pleasing image, it is a quantitative description of tip–surface interactions, hence, mathematical modeling of these interactions is often required to interpret the image contrast. The basic equation for image formation is: $\Delta f(z_c) = (f_0/2k)k_{ts}(z_c)$, where Δf is the frequency shift, z_c is the lift height, f_0 is the base frequency, k is the cantilever spring constant, and k_{ts} is the force gradient.²⁰ Therefore the image in **FM-AFM(1)** mode is a map of force gradient over the surface at constant separation, and the image in **FM-AFM(2)** mode is a map of the heights of constant tip–surface force gradient. For stable imaging several requirements must be fulfilled:

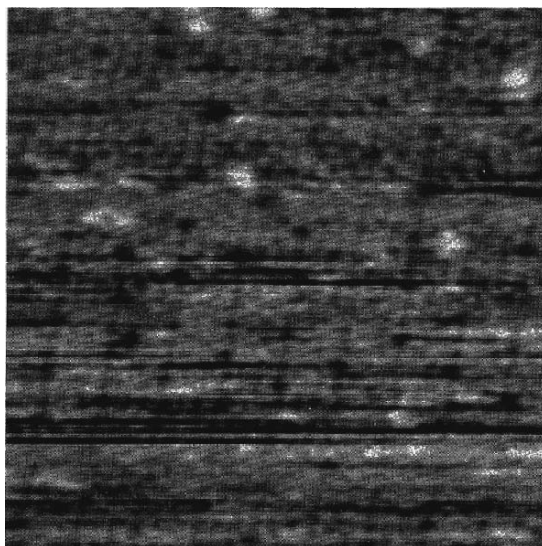


Figure 14–4. First NC-AFM image of the Si (111) 7×7 reconstruction. [Courtesy of Kitamura and Iwatsuki. Reprinted with permission from Japanese Journal of Applied Physics Part 2—Letters, 34(1B), L145–L148, 1995.]

$$\max \left| \frac{d^2 V_{ts}}{dz^2} \right| = k_{ts}^{\max} < k \text{ and}$$

$$\max \left| -\frac{dV_{ts}}{dz} \right| = |F_{ts}^{\max}| < kA_0$$

where V_{ts} is the tip–surface potential, k_{ts} is the tip–surface force gradient, k is the cantilever spring constant, $F_{ts}^{\max}$ is the tip–surface force, and A_0 is the oscillation amplitude. For atomic resolution stiff, high-frequency cantilevers are required to provide stable oscillation with a good quality factor. The basis for atomic resolution in FM-AFM is that the first derivative of the tip–surface force is monitored, in contrast to AM-AFM, in which the force itself is monitored. Experimental data obtained in FM-AFM(1) can be converted into FM-AFM(2) data only when the distance dependence of the tip–surface potential is known. Experimental parameters such as cantilever resonant frequency, spring constant, and oscillation amplitude vary from one experiment to another. To compare experimental results obtained under different conditions the concept of *normalized frequency shift* was developed by Giessibl.²⁰ Normalized frequency shift, $\gamma = (kA^{3/2}/f_0)\Delta f$, depends on the tip–surface potential and not on oscillation amplitude or oscillation frequency. Analytical solutions for basic tip–surface interactions (power law dependence and exponential decay)²¹ have also been developed. This allows the magnitude of the tip–surface force to be related to the image formation mechanism. The concept of normalized frequency shift will be used in subsequent discussions of various applications.

2.3 Role of the Tip in Image Interpretation

2.3.1 General Approach for Tip Modeling

Classical physics fails to quantitatively describe interactions between the tip and surface, while analytical expressions for a quantum mechanical treatment have not yet been developed. In other words at this point numerical calculations are required to calculate the quantum mechanical part of the problem. The currently used approach is to develop “nanotip” models based on a cluster at the end of the tip, which interacts with the surface obeying quantum mechanical laws. Using this model, “chemical” interactions between the “nanotip” and surface can be calculated based on the pair potentials between tip and surface atoms. Since numerical calculations are time consuming the “nanotip” is kept as small as possible. Obviously, the “nanotip” does not represent the entire tip–surface interaction. Among the models developed to incorporate long-range forces the most popular is the representation of the tip as a cone with a hemispherical end. A complete description of the tip then includes a cone with a hemispherical cap (analytical description) with a “nanotip” at the end of the hemisphere (numerical modeling). Some information regarding the tip shape, conductivity, and charging is inherent in the experimental dependence of the cantilever frequency change on tip–surface separation measured before and just after imaging.^{22–25}

2.3.2 “Nanotip” Models

In SPM simulations, the most common conception for Si and other semiconductor surfaces has been that the main component of the tip–surface interaction is the interaction of a dangling Si bond at the end of the tip with the surface atoms. This dangling bond can be well described using relatively small 4- or 10-atom Si clusters saturated by H atoms.²⁶ Another approach is to assume from the outset that the tip is ionic (MgO).^{27,28} Calculations showed that if the bottom of the tip was flat, i.e., no nanotip (only a macroscopic tip), then the interaction with the surface was averaged over several tip ions, and no contrast was produced. When a nanotip was included, it had to extend significantly beyond the main part of the tip to achieve atomic resolution.

Silicon cluster tip models have been successful in developing a qualitative understanding of the origins of image contrast in SPM on metals, semiconductors, and insulators; however, they fail in most cases to qualitatively reproduce image contrast. The solution is to compare images to simulations with a number of different tip models. Silicon tips under experimental conditions are likely to be contaminated by residual oxide, adsorbed hydrogen, and water,²⁹ or even materials transferred from the surface. To compare the properties of clean and contaminated silicon tips, the electronic structures of Si10 clusters with adsorbed contaminant species were calculated using the density functional theory by Sushko et al.³⁰ The results clearly showed that adsorbed hydrogen has no effect on the potential gradient from the uncontaminated silicon cluster; however, adsorbed oxygen and hydroxyl groups cause a significant change in the potential gradient. Both potentials decayed over a much longer distance than did that of the uncontaminated cluster, and the stronger gradients suggested a much stronger interaction with the surface. Interestingly, the potential gradient from an MgO cube corner with an O atom at the end was very similar to that of the oxygen-contaminated silicon cluster, a strong negative potential. An MgO cluster is a good model of a hard oxide tip, and has the important advantage that reliable interatomic potentials exist for MgO, alkali halides, and other oxides.

In an interesting experiment Bennewitz and co-workers atomically resolved a copper (111) substrate, as well as a unit cell thick NaCl grown on that surface.³¹ An MgO “nanotip” of only a few atoms would not atomically resolve the features observed in the experimental images. Different “nanotip” models for SFM were compared to determine which most closely matched experimental results.³¹ It was found that a 64-atom MgO “nanotip” with an oxygen atom at the very end of the tip imbedded in a macroscopic tip gives quantitative agreement with image contrast. The MgO cube can also be oriented with the Mg ion down, providing a strong positive potential. For many SFM experiments on insulators, the ionic MgO tip model provides excellent quantitative agreement with the image. Recent achievements in instrumental control reduce the possibility of tip contamination by sample material. It has also become important to use conducting tip models when studying insulating surfaces.

A good illustration of the importance of the “nanotip” structure is given by Hembacher and co-workers.³² The molecular orbitals of a W atom at the end of a tip were “imaged” using an sp^2 orbital of flat graphite as a probe. This experiment demonstrates that current instruments are capable of resolving not only atoms but atomic orbitals as well. Although orbitals are not immediately obvious in the image contrast, Hembacher et al.³² filtered the first harmonic of the signal and then summed all other harmonics with weighted coefficients to eliminate topographic contributions to the image, and the contrast due to the atomic orbitals became evident.

2.4 Applications of NC-AFM

In spite of the relatively short time since the demonstration of the atomic imaging with NC-AFM, its potential for characterizing nonconducting materials has motivated numerous studies. Illustrative examples of several classes of materials are summarized below.

2.4.1 Oxides: $SrTiO_3$ and TiO_2

Atomic resolution imaging of many oxide surfaces including Al_2O_3 ,³³ TiO_2 ,³⁴ $SrTiO_3$,³⁵ NiO ,³⁶ CeO_2 ,³⁷ mica,³⁸ and MoO_3 ³⁹ has been demonstrated using frequency modulation (FM)-AFM. Two representative examples, $SrTiO_3$ and TiO_2 , are illustrated here. The (100) surface of $SrTiO_3$ attracts much attention not least because it is the best substrate for the deposition of epitaxial films of superconductive oxides. Another attractive feature of $SrTiO_3$ is the ability to dope the surface with oxygen to very high concentrations ($\sim 10^{18} \text{ cm}^{-3}$)⁴⁰ suggesting the potential of applications as channel layers of field-effect transistor.

Unit cell resolution was obtained by Kubo and Nozoye³⁵ on an $SrTiO_3$ (100) single crystal, on which they observed the $\sqrt{5} \times \sqrt{5}$ reconstruction, using the FM-AFM(2) mode. The details of frequency shift were not provided so an estimation of the nature of the forces is not possible. W_2C -coated conductive tips were used to eliminate electrostatic charge on the tip and chemical interaction between the tip and surface, and to nullify the capacitance forces. As a result, unit cell resolution was achieved. The authors describe the $\sqrt{5} \times \sqrt{5}$ surface re-construction as an ordered array of Sr adatoms. Figure 14–5 compares STM and NC-AFM images of the surface. The STM resolves only the Sr adatoms, while the NC-AFM resolves both the adatom and the underlying lattice. STM contrast on oxides contains both a geometric and electronic structure contribution, the relative magnitude of which cannot be determined a priori. In this case the geometric height of the Sr adatom appears to preclude resolution of the Ti lattice in unfilled state images. The AFM contrast, in principle, contains integrated charge density of all atoms so the oxygen sublattice can be imaged simultaneously. Other orientations of $SrTiO_3$ single crystals were examined by NC-AFM; atomic rows were observed on $SrTiO_3$ (110)⁴¹ and atomic steps on $SrTiO_3$ (111).⁴² It should be noted that unit cell resolution is routinely achieved with STM as well, when the sample is either doped or significantly reduced to provide sufficient conductivity ($SrTiO_3$ is an insulator).

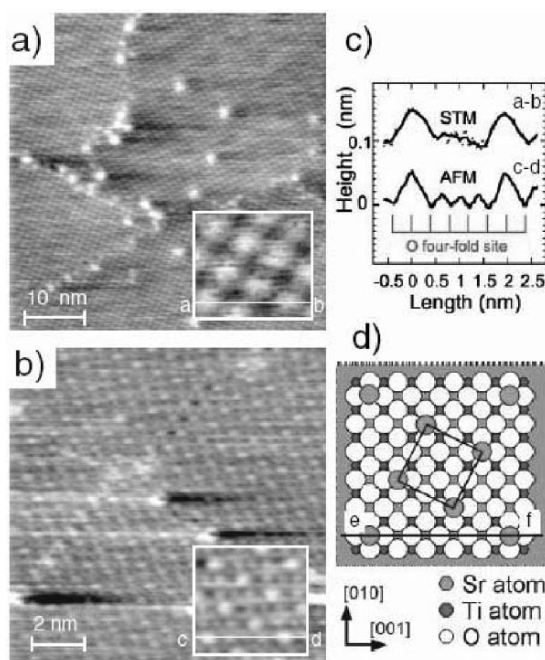


Figure 14–5. SPM on (100) SrTiO_3 . (a) STM image ($V = +0.65 \text{ V}$, $I = 0.11 \text{ nA}$); (b) NC-AFM image ($V = -1.49 \text{ V}$); (c) line profile comparison for STM and NC-AFM; note that STM shows the Sr adatoms on the surface while NC-AFM resolves unit cells underneath; (d) the model for $\sqrt{5} \times \sqrt{5}$ Sr adatoms on SrTiO_3 . [Courtesy of Kubo and Nozoye. Reprinted with permission from Physical Review Letters, 86(9), 1801–1804, 2001.]

As discussed earlier, the details of formate adsorption on TiO_2 (110) were demonstrated with atomic resolution STM and FM-AFM(2) with a conductive silicon tip. An untreated (and probably oxidized) Si tip was used to image the (110) surface.^{13,43} The (100) surface of TiO_2 is rather more complex than is the (110) surface, exhibiting a (1×3) microfaceted structure at high temperatures and a combination (1×1) and (1×3) structure at intermediate temperatures. Raza and co-workers compared the intermediate structure of the TiO_2 (100) surface with STM and AFM.⁴⁴ Figure 14–6 illustrates the level of detail evident in the AFM image contrast. The normalized frequency shift, γ , in this case is $0.147 \text{ fN/m}^{1/2}$, implying that short-range van der Waals or covalent (Morse potential) forces were accessed. It was noted that the lateral dimensions of the features in the AFM and STM images agree but the apparent corrugation amplitude is much larger in AFM (0.5 nm compared to 0.1 nm). This illustrates the importance of height calibration. In this case the (1×3) microfacet was used as an internal standard.

2.4.2 FM-AFM of Semiconductors: Si and GaAs

The Si (111) (7×7) reconstruction has long been a standard in surface science. The structure based on minimization of the number of dangling bonds took 20 years to solve. The “Dimer–Adatom–Staking Fault”

model proposed by Takayanagi et al.⁴⁵ is the currently accepted structure for the silicon (111) 7×7 surface. Immediately after the invention of the STM, Binnig and co-workers imaged the 7×7 reconstruction on Si (111) in real space by STM.⁴⁶ It took more than 10 years before the first NC-AFM images of the 7×7 reconstruction were obtained simultaneously by Kitamura and Iwatsuki⁴⁷ and Giessibl.⁴⁸ Both groups used FM-AFM(2) to obtain atomic resolution as described earlier (Figure 14-4).

A number of semiconductor surfaces have been imaged since then, including Si (111), (100), (110); GaAs (100); and InAs (110).⁴⁹ Several representative examples are presented here. The most technologically relevant Si (100) undergoes a 2×1 reconstruction. Figure 14-7 shows an atomic resolution FM-AFM(1) image by Yokoyama and co-workers.⁵⁰ The normalized frequency shift calculated for these images is $\sim 4.35 \times 10^{-14} \text{ N/m}^{1/2}$, a magnitude that suggests that the forces are either short-range van der Waals or Morse forces, hence atomic resolution. Although the surface is known to adopt the (2×1) reconstruction, the distance between the “dimer-like” features in Figure 14-7 is larger (0.35 nm) than

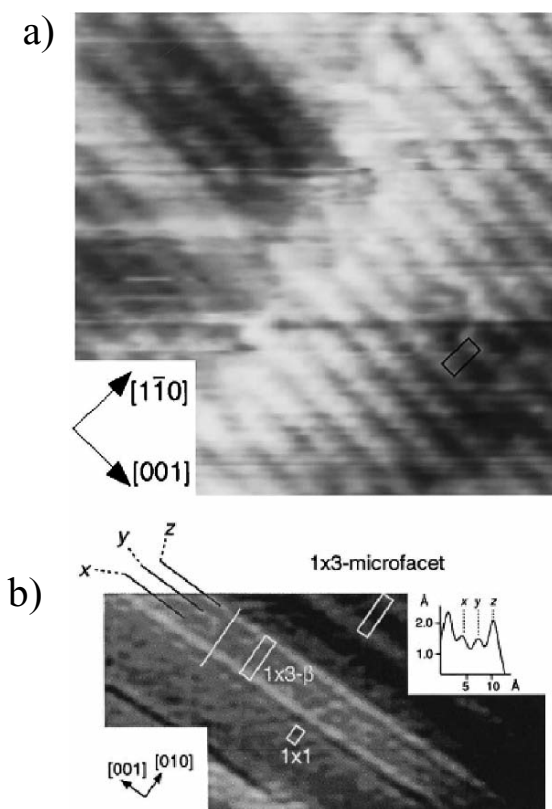


Figure 14-6. NC-AFM image of 1×1 reconstruction on TiO_2 (100) (a); faceted structure with 1×3 and 1×1 reconstruction on TiO_2 (100) (b). [Courtesy of Ashino et al. Reprinted with permission from Physical Review Letters, 86(19), 4334–4337, 2001.]

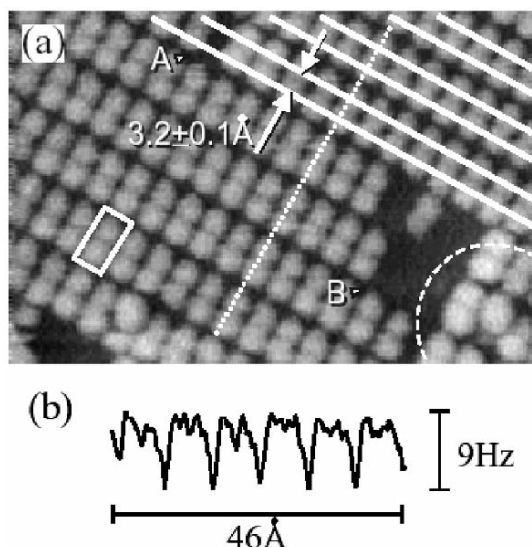


Figure 14-7. (a) NC-AFM image of Si (100) 2×1 reconstructed surface. One 2×1 unit cell is outlined by a box. (b) Line profile of the NC-AFM image across the dotted line. [Courtesy of Yokoyama et al. Reproduced with permission from Japanese Journal of Applied Physics Part 2—Letters, 39(2A), L113–L115, 2000.]

of the expected structure (0.23–0.20 nm). This result was compared with an NC-AFM image of the hydrogen passivated surface on which the lateral dimensions agree with the expected positions of hydrogen atoms. This dilemma is resolved if the image mechanism is that of a Si dangling bond of the tip interacting with the surface. In the case of the clean (2×1) surface the contrast then relates to surface dangling bonds rather than dimer bonds and in the case of the passivated surface the contrast relates to the dangling bond–hydrogen interaction. These results further emphasize the role of chemical/bonding forces in the image mechanism.

The surface of GaAs affords an opportunity to examine differences in tip–surface interactions on the same surface. Uehara et al.⁵¹ compare the corrugations of As atoms and Ga atoms on a GaAs (110) surface at different tip–surface distances. Again FM-AFM(2) was used to obtain atomic resolution. Unfortunately the oscillation amplitude is not specified so a quantitative analysis of the forces is not possible. In this case, though, the apparent distance between As atoms did not change with tip–sample distance, while the distance between Ga atoms did. This is explained by considering that the force on the surface atom increases with decreasing tip–sample distance. Therefore, the As feature is associated with valence charge distribution and is not sensitive to this range of forces, while the Ga feature is associated with a dangling bond, which can be displaced by the interaction with the tip. This is another case of using atomic bonding interactions for chemical specificity in AFM image contrast.

2.4.3 Layered Materials: Graphite

Numerous layered compounds have been studied by NC-AFM, including mica,⁵² MoO_3 ,³⁹ etc. Highly oriented pyrolytic graphite is perhaps the most general in that it is used as a calibration standard for SPM and as an atomically flat substrate in numerous studies. As with most layered compounds the graphite is easily prepared as a clean atomically flat surface by cleavage. Despite the proverbial ease of imaging graphite by STM with atomic resolution, every second atom in the hexagonal surface unit cell is usually unresolved. Giessibl and co-workers⁵³ demonstrated that a low-temperature FM-AFM(1) with pico-Newton force sensitivity reveals the hidden surface atom (Figure 14–8). This is another clear example that despite similarity in implementation and that combined AFM/STM microscopes are manufactured, differences between the image formation mechanisms of STM and NC-AFM are very important. In STM the tip images the density of states at the tip bias, while in NC-AFM the tip images an equiforce surface. With an oscillation amplitude of 300 pm, a cantilever spring constant of 1800 N/m, and an eigenfrequency of 18,076.5 Hz, the normalized frequency shift, γ , is $-2.7 \text{ fN/m}^{0.5}$. The magnitude of the normalized frequency shift indicates that the contrast in FM-AFM(1) is due either to short-range van der Waals or to Morse forces for the image in Figure 14–8.

2.4.4 Dielectrics: KBr and CaF_2

The ability to image insulating surfaces of high resistance has been noted as an advantage of NC-AFM. Of course the oxides presented in Section 2.4.1 are insulating compounds when fully oxidized and undoped. Here we illustrate compounds that are not made conducting. Benewitz and co-workers present an illustrative example⁵⁴ by imaging the potassium chloride–bromide surface. True atomic resolution was

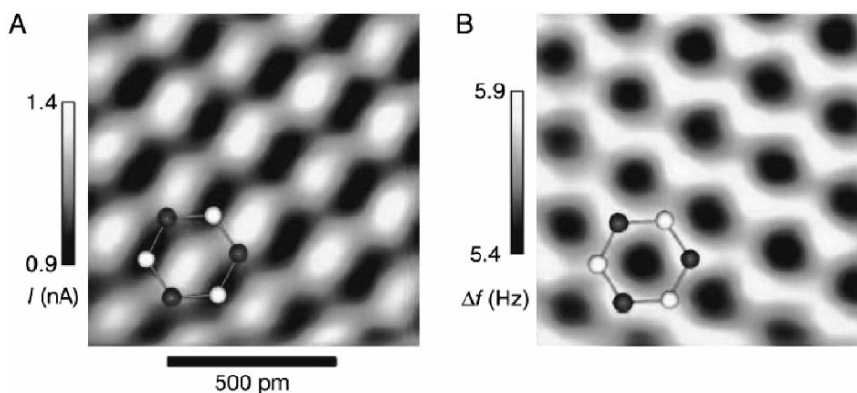


Figure 14–8. Low temperature STM (A) and NC-AFM (B) images on graphite. One hexagon of the graphite structure is superimposed on STM and NC-AFM images. One atom per unit cell for STM and two for NC-AFM. [Courtesy of Hembacher et al. Reproduced with permission from Proceedings of the National Academy of Sciences USA, 100(22), 12539–12542, 2003.]

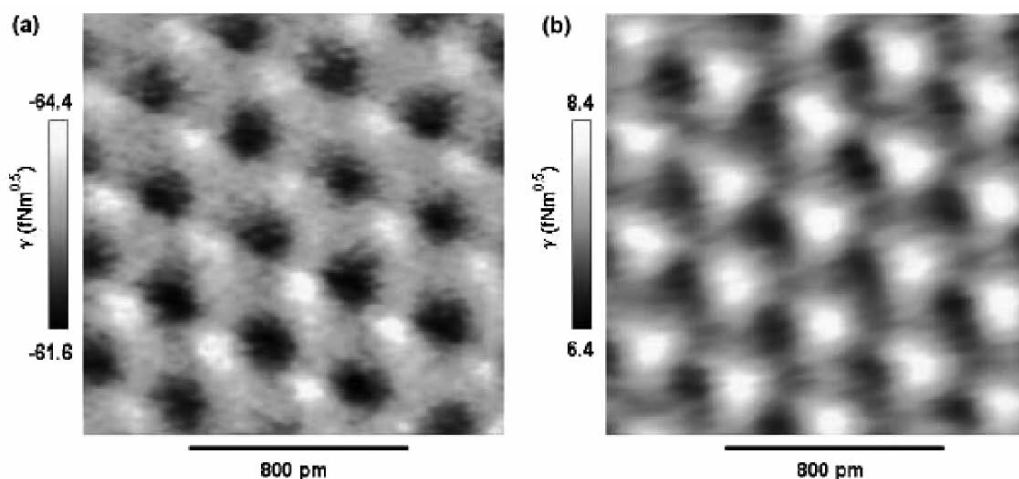


Figure 14–9. High-resolution constant height images of CaF_2 (111) in (a) the attractive and (b) the repulsive mode of FM-AFM(1) using qPlus sensor as cantilever. (Courtesy of Giessibl and Reichling. Reproduced with permission from *Nanotechnology*, 16, (2005) S118–S124, 2005.

not achieved in their analysis of $\text{K}(\text{Cl}, \text{Br})$ single crystals, but atomic corrugations with spacing equal to the size of unit cell were resolved. This difference in corrugation amplitude was attributed to anion positions on the surface, providing an additional example of chemical specificity in NC-AFM.

Giessibl and Reichling⁵⁵ imaged the CaF_2 (111) surface with atomic resolution in both attractive and repulsive modes (Figure 14–9). The level of spatial resolution was facilitated by the use of a quartz tuning fork to oscillate the tip. The normalized frequency shift was $-62 \text{ fN/m}^{0.5}$ for attractive and $7.4 \text{ fN/m}^{0.5}$ for repulsive interactions, suggesting that van der Waals forces are responsible for image formation. The quartz tuning fork has a high spring constant ($\sim 1000 \text{ N/m}$ compared to $\sim 10 \text{ N/m}$ for silicon cantilever), resulting in high stability of the small-amplitude oscillations. The quality of these images emphasizes the importance of small oscillation amplitudes and the proximity of these oscillations to the surface. Also, imaging in repulsive mode is challenging but yields high spatial resolution.

3 Imaging Properties: Advanced SPM Techniques

In addition to mapping topographic and atomic structure, variants of SPM probe local properties. The most common of these are EFM and magnetic force microscopy (MFM). These are straightforward measurements in which the tip is lifted above the surface to a distance at which only long-range forces are detected. By expanding the mechanisms of force detection and utilizing all possible image modes, the list of properties that can be probed is extensive (Table 14–2). The cantilever can be driven mechanically or electrically; this can be done in

contact or noncontact regimes; and not only amplitude, but also first, second, and third harmonics of amplitude (or phase) can be detected. The relation between these operational regimes is shown in Figure 14–10. For example, conventional AFM is a noncontact, mechanically driven oscillation with amplitude or phase-based detection. The newer piezoelectric force microscopy (PFM) is in contact mode, electrically driven, and with phase-based detection. Static or periodic electric or magnetic fields can be applied to the sample, independent of the tip signal. An underlying theme of the newest developments is the use of multiple signal modulations or high order harmonics of modulated signals. Within this framework several techniques that address transport and dielectric properties will be reviewed.

3.1 Advanced SPM Techniques for Transport Properties

Perhaps the most common of the so-called “advanced” SPM techniques is scanning surface potential microscopy (SSPM), sometimes referred to as Kelvin probe force microscopy (KPFM), which maps the work function of surfaces.^{56,57} In SSPM, the cantilever oscillation is driven electrically. This is a noncontact (generally 50–200 nm separation) probe, with feedback on the first harmonic (Figure 14–11). An ac voltage

Table 14–2. Properties accessible with scanning probe microscopy.

Technique	Mode	Property	References
AFM	nc/ic, mech, phase/amp	vdW interaction, topography	[2–5]
EFM	nc, mech, phase/amp	Electrostatic force	[2–5]
MFM	nc, mech, phase/amp	Magnetic force, current flow	[2–5]
SSPM (KPM)	nc, elec, 1st harmonic	Potential, work function, adsorbate enthalpy/entropy	[2–5, 56–75]
SCM	c, F, cap sensor	Capacitance, relative dopant density	[2–5, 82, 84–87,
SCFM	c, elec, 3rd harmonic	dC/dV , dopant profile	89–92]
SSRM	c, F, dc current	Resistivity, relative dopant density	[2–5, 81–83, 87]
SGM	nc, elec, amp	Current flow, local band energy, contact potential variation	[2–5, 97]
SIM	nc, elec, phase/amp	Interface potential, capacitance, time constant, local band energy, potential, current flow (in combination with SSPM)	[93, 97, 99]
NIM	c, F, freq spectrum	Interface potential, capacitance, time constant, dopant profiling	[102–104]
PFM	c, elec, phase/amp	d33	[63, 105–111, 125, 126, 128, 129]
NPFM	c, elec, 2nd harmonic	Switching dynamics, relaxation time and domain nucleation	[130]
SNDM	c, F, 1st or 3rd harmonic	dC/dV , dielectric constant	[133, 134]
NFMM	c, F, phase	Microwave losses, d33	[135–138]

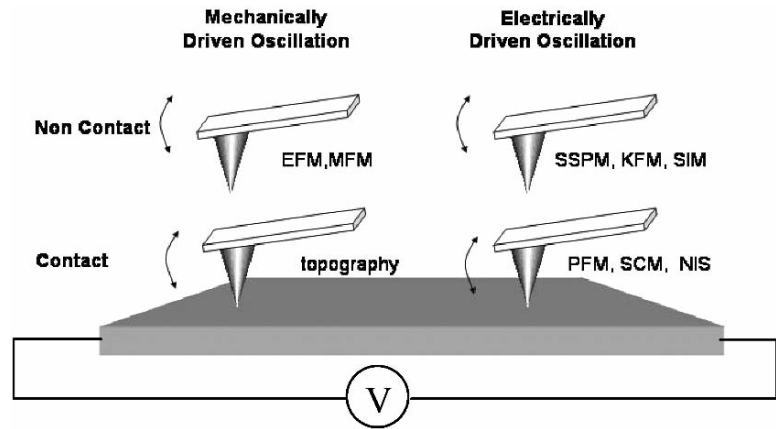


Figure 14–10. Variations of SPM are separated into four operational regimes distinguished by cantilever driving force and condition of contact.

is applied to the tip. V_{dc} is varied until it matches the surface potential, V_{surf} , at which point the first harmonic of the force F (with frequency ω) is nullified. Thus, the applied V_{dc} is equal to the surface potential. The tip potential is expressed as $V_{tip} = V_{dc} + V_{ac} \sin \omega t$, and the force acting between the tip and surface $F(z) = \frac{1}{2}(\partial C / \partial z)[V_{tip} - V_{surf}]^2$, so the first harmonic of the force is $F_{1w}(z) = (\partial C / \partial z)[V_{dc} - V_{surf}]V_{ac}$.

The apparent ease with which potential variations can be mapped belies the challenges in interpreting images on electrically^{58,59} and topographically inhomogeneous surfaces. In the cases of semiconductor and dielectric surfaces the electrostatic properties of a surface are not characterized solely by intrinsic potential and topography. SSPM images must be interpreted in terms of *effective* surface potential that includes capacitive interactions, surface and volume bound charges, double layers, and remnant polarization.^{60–63} For semiconductor surfaces tip-induced band bending⁶⁴ can offset local surface potential.⁶⁵ Despite these challenges the obvious need to examine variations in

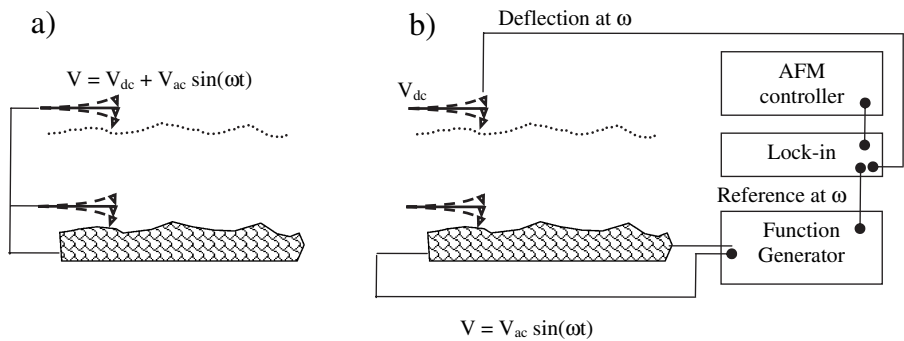


Figure 14–11. Comparison of working principles for SSPM (KFM) and SIM. In SSPM (a) the oscillating potential is applied between the sample and tip; in SIM (b) the oscillating potential is applied across the sample.

local potential in electronic nanodevices spurred efforts to overcome some of the obstacles with careful analytical treatments that determined limits in quantification. In the late 1990s SSPM was applied to semiconductor,^{66,67} organic,⁶⁸ and ferroelectric^{69,70} surfaces, as well as to defects,^{71,72} and photoinduced^{73,74} and thermal phenomena.⁷⁵ It is fair to say that at this point that absolute values of potential cannot be quantified, but variations in potential can be determined with energy resolution of 2–4 meV and spatial resolution of the order of 50–100 nm.

One interesting application is the measurements of surface potential of self-assembled monolayers (SAM). Figure 14–12 shows a typical potential map of two SAMs patterned by microcontact printing. In this case one SAM is an alkanethiol, while the other is a conjugated conductive molecule. The difference in surface potential is obvious but the basis of the difference is not entirely clear. There are, in principle, four contributions to the potential of SAM: the substrate work function, the dipole in the surface bond (usually S—M or SiO_x—Si), the dipole in the molecule, and the terminating end group. Eng and co-workers showed that the potential of alkanethiols on Au increases with the increase of chain length.⁷⁶ Sugimura et al. calculated dipole moments in a sequence of siloxane-coupled molecules and the trends agreed with the experiment.⁷⁷ These results provide evidence for the contribution of the molecular dipole to the measured surface potential. Alvarez and Bonnell⁷⁸ showed that for the SAM of alkanes on metal, the S—M bond dipole also affects the measured surface potential. While work remains to quantify the relative contribution of SAMs to surface potential, SSPM easily quantifies variations and is routinely used to characterize complex films.

On samples with morphological variations, such as grains, SSPM in the presence of a lateral field can elucidate the behavior of individual microstructural features. This is illustrated in Figure 14–13, which shows the behavior of a polycrystalline ZnO surface under different applied lateral biases. The topography contains features due to contaminants and inter- and intragranular pores. On application of 5 V

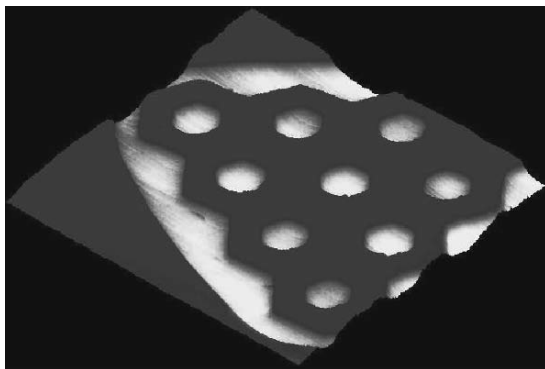


Figure 14–12. SSPM image of SAM of alkanethiol and conjugated molecules self-assembled on Au, patterned with a microcontact printing technique. (Courtesy of Rodolfo Alvarez.)

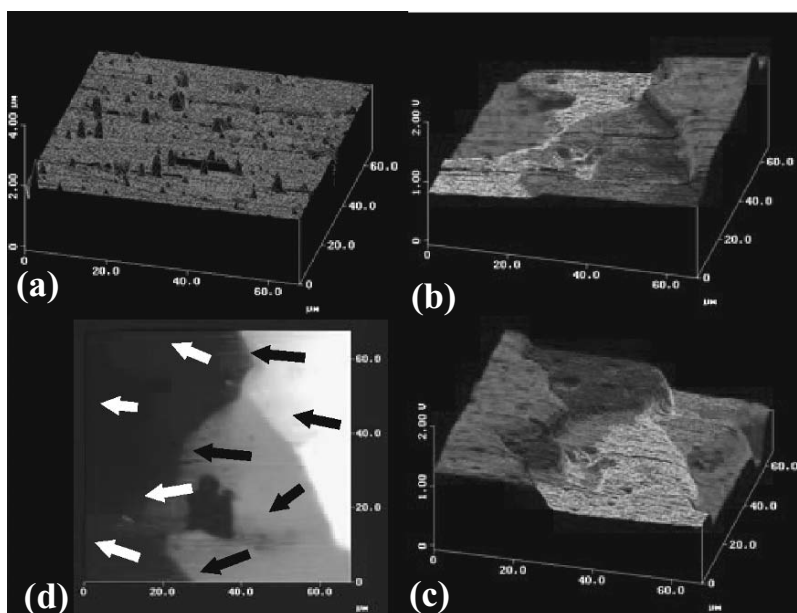


Figure 14-13. Transport imaging in polycrystalline ZnO. (a) Surface topography. (b and c) SSPM images under lateral bias exhibit potential drops at grain boundaries, indicative of grain boundary resistive behavior. Note that the direction of potential drops inverts with bias. (d) Current maps for positive and negative bias polarity. (Partially reproduced from Kalinin and Bonnell, *Zeitschrift für Metallkunde*, 90, 983–989, 1999.) (See color plate.)

lateral bias the potential steps at grain boundaries become evident (Figure 14-13b), and the contrast inverts when the direction of the applied bias is reversed (Figure 14-13c). Clearly the behavior of individual grains and grain boundaries is distinguished and the voltage dependence of all microstructural constituents can be examined separately. Furthermore, taking directional derivatives allows the current flow to be determined at all points on the surface, as shown in Figure 14-13d.

Although not explicitly stated, much SSPM is done in ambient conditions. Indeed the ease of measurement is a strong advantage for its use as a qualitative characterization tool. As with all ambient measurements, the possible interaction of the environment with the surface must be considered. There are, in fact, situations in which these effects dominate. Some of the early studies involved imaging ferroelectric domains in air⁷⁹ and it was eventually understood that the strong field due to domain polarization at a surface is almost completely screened, presumably by adsorption. Domains are easily imaged but quantifying the magnitude of the surface potential is impossible. In fact the sign of the measured potential is usually opposite that of the domain potential. Ferroelectric domains represent an extreme case of local electric fields, but even for grain boundaries intersecting a surface the effect is observed. Figure 14-14 compares the potential of an atomically abrupt boundary in SrTiO₃ measured in air and in ultrahigh vacuum (UHV).⁸⁰

The boundary contains an intrinsic charge, which results in a potential variation at the intersection with the surface. In ambient conditions the charge attracts compensating adsorbates and the measured difference in surface potential is under 10 mV. In UHV, the uncompensated difference in potential is on the order of 90 mV and opposite in sign.

Two contact techniques, scanning spreading resistance microscopy (SSRM)^{81–83} and scanning capacitance microscopy (SCM),^{84–86} have been developed to characterize semiconductors. In SSRM, a conducting tip is biased with respect to the sample and the dc current through the tip/surface contact is detected under force feedback control. The amount of current is determined by the local spreading resistance of the surface, which is related to the local conductivity. Because the native oxide layer on Si hinders tip/surface contact, carrier profiling in Si-based devices usually requires a tip coated with a hard material and a high spring constant cantilever to provide strong indentation forces ($\sim 20 \mu\text{N}$).⁸⁷ In SCM, a high-frequency capacitance sensor detects the tip/sample capacitance as the tip is scanned across the sample, which allows a smaller indentation force than is used in SSRM. Diamond-coated tips are used in spreading resistance measurements, but lower indentation force makes metal-coated probes suitable for SCM.⁸⁸ An ac voltage applied to the tip induces the depletion and accumulation of carriers, resulting in a change in capacitance, ΔC . In a semiconductor, the depletion/accumulation width is inversely related to the carrier concentration, so mapping $\Delta C/\Delta V$ yields a carrier concentration profile. Difficulties in quantification arise if the dopant concentration is non-uniform, and when spatial resolution decreases due to low dopant concentrations. A recent modification incorporates an additional feedback that adjusts ΔV to maintain a constant ΔC during the scan, maintaining a constant depletion width.⁸⁸ Analysis of SCM results are mathematically

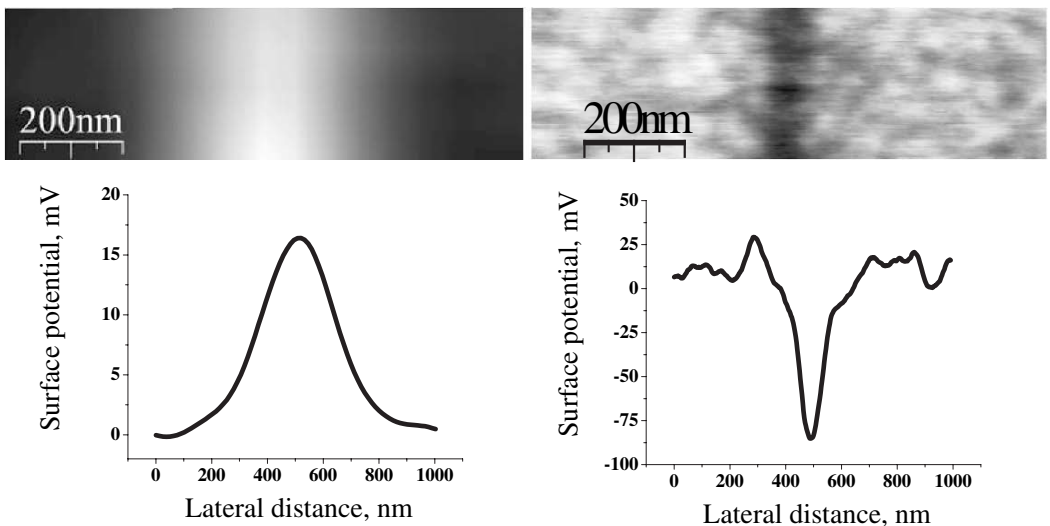


Figure 14–14. KFM of the $\Sigma 5$ grain boundary in SrTiO_3 (100). Left, in air; right, in UHV. (Courtesy of Rui Shao.)

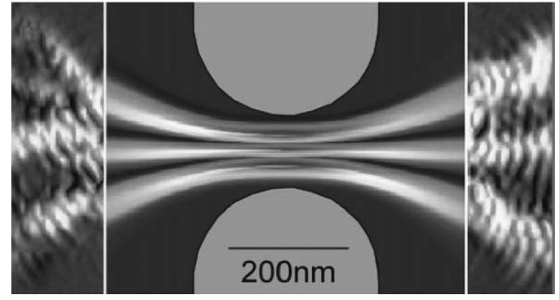
challenging, usually requiring 2D or 3D numerical approaches.^{89–92} Note that SCM distinguishes the difference between n- and p-dopants in the sample, whereas SSRM does not. On the other hand metallic and insulating samples give no contrast in an SCM image, while SSRM images defects in insulations and in metallic films.⁸⁸

Another technique for electronic property characterization on the nanoscale is scanning gate microscopy (SGM). SGM is a two-scan technique. During the first scan topography of the sample is monitored in intermittent contact mode. During the second scan an ac voltage is applied across the sample; a biased tip is brought within proximity of the sample, and the magnitude of the current *across the sample* is measured as a function of tip position. The first implementation of SGM was demonstrated on a single wall carbon nanotube in 2000.⁹³ The tip acted as a gate electrode in this nanotube circuit. It was found that the Fermi surface of the nanotube was not uniform along its length. Researchers from Stanford University,⁹⁴ University of California at Berkeley⁹⁵ and University of Pennsylvania^{96,97} used SGM to good effect in nanotube-based electronic circuits illustrating the ability to quantify local properties in individual nanotubes and to demonstrate three-terminal device behavior. Recently researchers from the University of California at Irvine⁹⁸ measured the gating effect of a biased AFM tip on ZnO nanowires and found it to be 200 times smaller than that of single-wall carbon nanotubes.⁹⁴ The change of resistance in these nanowires differs because of the difference in the widths of the single-wall carbon nanotube (1–2 nm) and ZnO nanowires (~400 nm).

Research in the group of R.M. Westervelt extended the concept of using an SPM tip as a perturbing probe to visualize the flow of electron waves in a two-dimensional electron gas. They produced a planar electron gas 57 nm below the surface in a GaAs/AlGaAs heterostructure with a quantum point contact formed with a pair of gates on the surface. A biased tip near the surface will capacitively couple to the underlying electron gas. As the tip is scanned across the surface, it decreases the conductance through the contacts when it is over a region of high electron flow and has no effect over a region with low electron flow. Similar to SGM, the image is the current or conductance across the gap as a function of tip position on the surface. Figure 14–15 illustrates the geometry of the quantum point contact and the pattern of electron flow that results from a system with a density of $4.5 \times 10^{11} \text{ cm}^{-2}$, a mobility of $1 \times 10^6 \text{ cm}^2/\text{Vs}$, a mean free path of 11 μm , and a Fermi wavelength $\lambda_F = 37 \text{ nm}$. Images were recorded with the sample and the SPM at $T = 1.7 \text{ K}$.

Many phenomena underlying materials behavior involve time-dependent processes that could be accessed if signals applied to the sample and/or tip spanned a wide frequency space. The first introduction of frequency dependence in scanning probes is referred to as scanning impedance microscopy (SIM).⁹⁹ This is a noncontact, first harmonic detection in which the oscillating electrical signal is applied to the sample instead of the tip. The implementation of SIM is compared to SSPM in Figure 14–11. The tip can act as a nonperturbing probe or as a local gate in a configuration that allows both the amplitude, which is related to potential, and the phase, which is related to

Figure 14–15. Experimental images (outside) and theoretical simulations (inside) of the flow of electron waves through a quantum point contact. Fringes spaced by half the Fermi wavelength demonstrate coherence in the flow. [Courtesy of Westervelt et al. Reproduced with permission from *Physica E*, 24, 63–69, 2004.] (See color plate.)



loss, to be quantified. The frequency dependence can be used to isolate relaxation associated with electron traps at interfaces and defects.

Figure 14–16a and b illustrates SIM of an interface and grain boundary. For a prototypical ideal sample composed of a single electroactive

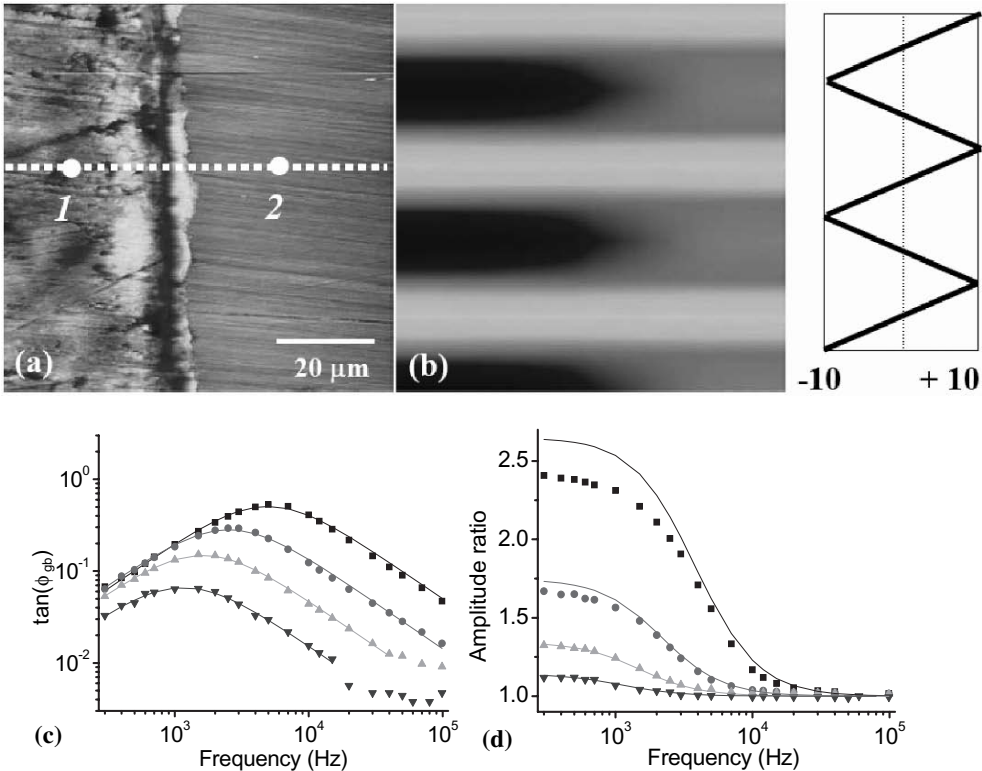


Figure 14–16. Surface topography (a) of the cross-sectioned diode. The potential profiles were acquired along the dotted line. The changes of potential, phase, and amplitude were determined from positions 1 and 2. Surface potential (b) during a 0.002-Hz triangular voltage ramp to the sample for $R = 500\ \Omega$. The scale is 300 nm (a) and 10 V (b). (c) Frequency dependence of SIM phase shift and (d) amplitude ratio of grain boundary in an Nb-doped $\Sigma 5$ SrTiO₃ bicrystal. Solid lines on (c) are fits for frequency independent grain boundary resistance and capacitance and on (d) the calculations are from using interface resistance and capacitance from phase data. Data are shown for circuit terminations 148 Ω (■), 520 Ω (●), 1.48 k Ω (▲), and 4.8 k Ω (▼). (Courtesy of Kalinin and Bonnell. Reprinted with permission from *Physical Review B* 70, 235304, 2004).

interface with resistance, R_d , and capacitance, C_d , in series with two current-limiting resistors, R , the phase shift across the interface as a function of frequency is $\tan(\varphi_d) = (\omega C_d R_d^2)/(R + R_d) + R\omega^2 C_d^2 R_d^2$, where $\omega = 2\pi f$ and f is the oscillation frequency. See Figure 14–11 for the exact configuration. For high frequencies the interface phase shift has the simple form $\tan(\varphi_d) = 1/\omega C_d R$ and interface capacitance can be calculated directly from the frequency shift. At low frequencies the amplitude ratio across the interface, $A_1/A_2 = (R + R_{gb})/R$. In addition, a dc potential can be applied across the surface establishing a potential step at the interface, which can then be measured by SSPM. As the dc bias across the interface of a metal/silicon diode is switched from positive to negative, the forward/reverse bias behavior is evident in the potential image (Figure 14–16a). The phase image of an atomically abrupt SrTiO₃ grain boundary (Figure 14–16b) shows the large effect of a 2D defect. The frequency dependences of both the amplitude (which is related to potential) and phase shift (which is related to interface processes) of the SrTiO₃ boundary are shown in Figure 14–16c and d. The combination of SIM and SSPM allows local transport spectroscopy of individual interfaces, e.g., direct measurement of the spatially resolved interface C – V characteristics. The quantitative nature of this approach has been confirmed on studies of Si/metal diodes.¹⁰⁰

SIM also contains an intrinsic improvement in spatial resolution relative to scanning surface potential or SGM. In combination with other SPM techniques, SIM has been used to map the rectifying behavior of a Schottky junction, and characterize defect-mediated transport in a single nanotube,⁹⁷ as well as provide transport measurements of idealized grain boundaries. Recently SIM was used to image current transport in carbon nanotube networks imbedded in a polymer.¹⁰¹ Similar to SSPM, the spatial resolution in SIM is ultimately limited by capacitive tip–surface interactions.

A second approach to accessing frequency-dependent transport expands the frequency range to eight orders of magnitude and provides higher spatial resolution. Nanoimpedance microscopy and spectroscopy (NIM)^{102–104} is a contact probe with force feedback, in which the oscillating bias signal is applied to the tip; current phase and amplitude are detected at the sample. Figure 14–17a illustrates the local impedance between a tip and lateral electrode, across a ZnO grain boundary. By accessing this range of frequencies, the data can be presented in a conventional Cole–Cole plot that describes both the real and imaginary components of impedance. In these plots each of the semicircles corresponds to the properties of a microstructural element. The interpretation is somewhat model dependent, but for sample configurations such as thin films with bottom electrodes or single boundaries between electrodes, quantitative characterization of the local properties is possible. The local boundary potential, capacitance, charge, and depletion lengths can be extracted with spatial resolution on the order of tens of nanometers. In a configuration with an electrode under the sample, the impedance of individual grains can be imaged, as shown in Figure 14–17b.

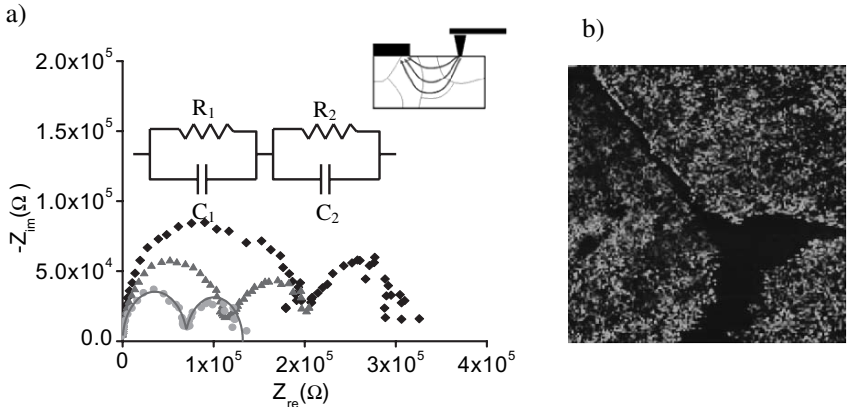


Figure 14-17. (a) Cole–Cole plot of local impedance spectra between the SPM tip and the top electrode (nanoimpedance spectroscopy) at tip/sample biases of +5 V (●), +3 V (Δ), and +2 V (◆) for fixed tip location. Solid line is the fitting of impedance data to the equivalent circuit of two RC elements in series (—). Note that interface resistance decreases with dc bias, indicative of varistor behavior. (b) Representative NIM phase image; the dark region corresponds to nonconductive inclusion in the sample (phase $\theta = -90^\circ$).

3.2 Advanced SPM Techniques for Dielectric Properties

In principle, utilizing higher order harmonic signals and clever detector design allows dielectric constant, electrostriction, and piezoelectric properties to be detected (Table 14-2). In practice a number of probes have been developed to quantify linear and nonlinear dielectric properties locally. These have focused on electromechanical coupling coefficients, hysteretic ferroelectric domain switching, etc.

PFM is a scanning probe that is used increasingly to determine electromechanical coupling coefficients at local scales. It is a contact, electrically driven probe technique with feedback based on phase lag. Like NIM, an oscillatory electric field is applied to the tip, which is in contact with the sample (Figure 14-18). If the material is piezoelectric,

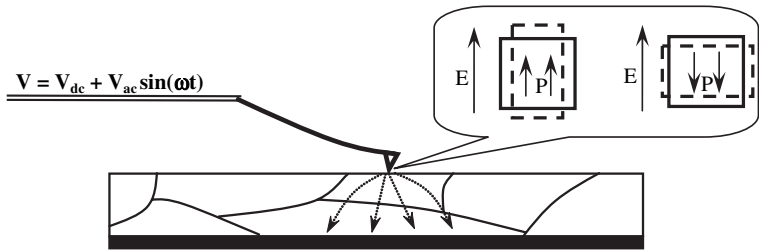


Figure 14-18. Working principle of PFM. An ac bias is applied to the tip, which is in contact with the sample, and an electric field is generated. If the material is ferroelectric the size of the region around the tip will change. A schematic of the responses of polarized domains is presented in the inset.

the field results in a local deformation of the surface that oscillates the tip, i.e., piezoresponse (PR).^{105,106} In a ferroelectric material domains with downward polarization vector contract with a positive voltage, producing a phase shift of $\delta = 180^\circ$. For upward oriented domains, the situation is reversed, and $\delta = 0^\circ$, because the deformation is in phase with the field. The phase, therefore, indicates the orientation of polarization. The piezoresponse amplitude, $A = A_{1\omega}/V_{ac}$, defines the magnitude of the interaction. For the ideal case of a (100) surface of a tetragonal compound, $A = \alpha d_{33}$, where α is proportionality coefficient close to unity,¹⁰⁷ the piezoelectric constant, d_{33} , is related to the polarization, P , as $d_{33} = \epsilon\epsilon_0 Q_{33}P$, where Q is the second-order electromechanical coefficient. For the general case there are in-plane components of polarization that can be accessed by measuring the lateral response of the tip to a field variation.^{108,109} Furthermore, the piezoelectric response is a tensorial function, the complexity of which depends on the symmetry of the compound and the orientation of the grain or crystal. Harnagea et al.¹¹⁰ have shown that even for BaTiO₃ with relatively high symmetry either the grain orientation or the in-plane component must also be known to determine domain orientation.

This is illustrated nicely by Gruverman and co-workers¹¹¹ who undertook the three-dimensional high-resolution reconstruction of the polarization vectors in a (111)-oriented Pb(Zr,Ti)O₃ ferroelectric capacitor by detecting the in-plane and out-of-plane polarization components using PFM. Figure 14–19a and b shows the vertical and lateral contributions to the phase images of a region that is nominally poled in the vertical direction. In spite of exhibiting uniform vertical contrast, the lateral component exhibits significant variation. Knowing that this material is oriented in the (111) direction, the individual domain orientation can be determined, as shown on Figure 14–19c. In this case the vertical components of domains polarization have the same magnitude, while the lateral component varies between domains.

In PFM voltage spectroscopy,¹¹² piezoresponse, $A_{1\omega}$, and phase, f , are measured as a function of dc potential offset, V_{dc} , on the tip.^{109,113} PFM spectroscopy yields local electromechanical hysteresis loops, quantifying remnant response and coercive bias, on the 20–50 nm level. Piezoresponse force spectroscopy of PbTiO₃ is illustrated in Figure 14–20. These hysteresis curves are acquired on a PbTiO₃ film with 100 nm grain size. The phase signal is related to polarization and therefore has the shape of a conventional P–E hysteresis curve. The amplitude signal shows that strain, in addition to switching, occurs with voltage and again traces the conventional butterfly shape of a curve. These curves illustrate the critical tip bias required to achieve polarization switching. A number of attempts to relate local hysteresis to crystallographic orientation and piezoresponse amplitude have been reported.¹¹⁴

One critical issue with regard to quantification and spatial resolution limit is the question of what volume is accessed by PFM. The answer is found by determining the decay of the electric field below the probe tip, which in turn depends on the dielectric constant and conductivity of the material. For oxide ferroelectric compounds the volume in on the order of 20–200 nm.^{115,116} Consequently, for films thinner than this,

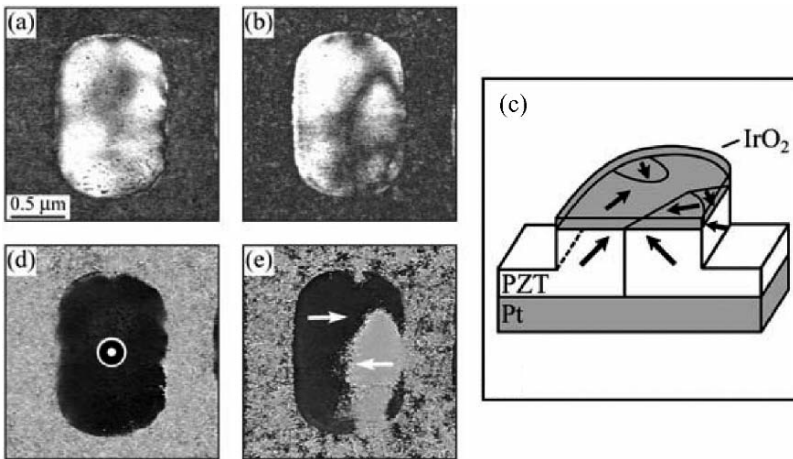


Figure 14-19. Investigation of a PZT capacitor with vertical and lateral force PFM. Images of amplitude (a) and phase (d) in vertical PFM and amplitude (b) and phase (e) in lateral PFM are presented. (c) Schematics of the domain orientation in the sample followed from PFM measurements. Note that only a combination of three PFM modes gives complete information about the direction of the polarization vector in the sample (only two are presented on the picture). [Courtesy of Rodriques et al. Reproduced with permission from Journal of Applied Physics, 95(4), 1958–1962, 2004.]

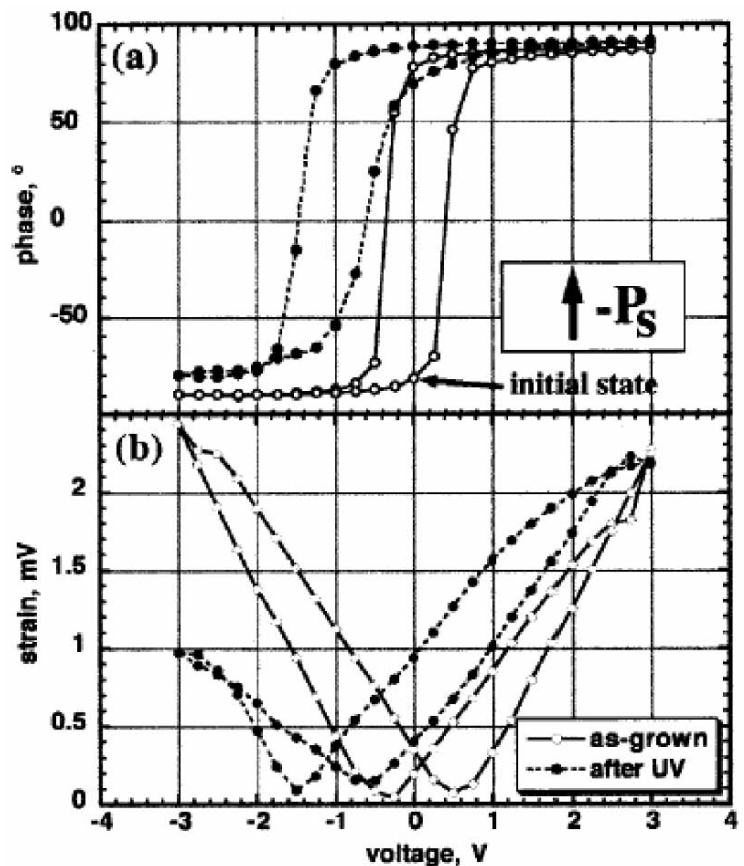


Figure 14-20. (a) Vertical phase and (b) amplitude hysteresis loops of PbTiO_3 thin films. (Courtesy of Gruveman and co-workers. Reprinted with permission from Journal of Applied Physics, 92, 2734, 2002).

or inclined domain structures, quantification requires accounting for this volume.

The question arises as to whether the PFM influences the properties it is measuring. Recently, it was shown that the mechanical strain produced by the tip can suppress local polarization¹¹⁷ or induce local ferroelectroelastic polarization switching.^{118,119} A quantitative analysis of the tip-induced potential and stress distribution is required to characterize local ferroelectric properties by SPM.^{120–124} A rigorous treatment of the image contrast includes simultaneous electrostatic and electro-mechanical interactions. One complication in PFM is that both long-range electrostatic forces and the electroelastic response of the surface contribute to the PFM signal.^{63,125} Even under optimal conditions, the basis of the electroelastic contribution, A_{piezo} is not straightforward because of the complex geometry of the tip–surface junction. Some progress in the quantitative understanding of PFM has been achieved.^{126–129} Depending on the tip radius of curvature and the indentation force, PFM may correspond to the electroelastic response of the surface induced by the contact area (strong indentation limit) or be dominated by the electroelastic response of the surface due to the field produced by the spherical part of the tip (weak indentation limit). In these cases, the magnitude of surface and tip displacements is determined by the electromechanical coupling in the material. Alternatively, the signal can be dominated by the electrostatic tip–surface interactions (electrostatic limit) and have little relation to the properties of the material. Taking an approach familiar to materials scientists, the analytical solutions of these interactions can be presented as contrast mechanism maps that relate experimental conditions to properties of the material and delineate the conditions under which quantitative measurements can be obtained (Figure 14–21).¹²⁶

The relationship between higher order harmonics of the PR function and time dependence of domain switching has been developed into a probe of switching dynamics called second harmonic piezo force microscopy (SH-PFM).¹³⁰ When the field and the measured electrostrictive strain are in the z direction, electrostriction is expressed in terms of the field-induced polarization P as $x = Q_{33}P^2$. For a ferroelectric with spontaneous polarization P^S and field-induced polarization P^E , the strain becomes $x = Q_{33}(P^S)^2 + 2Q_{33}P^SP^E + Q_{33}(P^E)^2$, where the second and third terms are the piezoelectric response (because $d_{33} = 2Q_{33}P^S/\epsilon$) and the electrostrictive response, respectively. Under external field, $E_3(\omega) = E_3 \cos \omega t$, induced

$$x = Q_{33} \left[(P^S)^2 + \frac{(P^E)^2}{2} \right] + 2Q_{33}P^SP^E \cos \omega t + \frac{1}{2} Q_{33} (P^E)^2 \cos 2\omega t$$

In macroscopic measurements electrostriction is quantified with interferometry.^{131,132} In SH-PFM the cantilever oscillation in contact mode determines the local electrostrictive properties.

A complementary strategy to accessing linear and nonlinear dielectric properties is referred to as scanning nonlinear dielectric microscopy^{133,134} and near-field microwave micro-scopy (NFMM).^{135,136} The

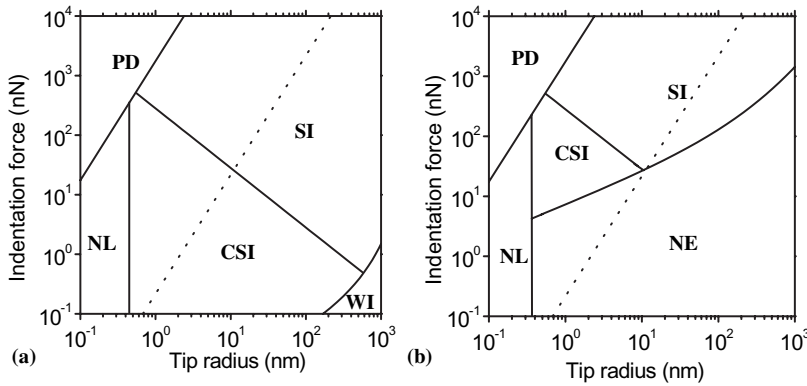


Figure 14-21. Contrast mechanism maps of piezoresponse force microscopy. SI, strong indentation regime; CSI, contact-limited strong indentation; WI, weak indentation regime; NE, nonlinear electrostatic regime; NL, nonlocal interactions; PD, plastic deformation. The dotted line delineates the region where stress-induced switching is possible. (a) $w = 0.1$ nm, $\Delta V = V_{\text{tip}} - V_s = 0$ V; (b) $w = 0.1$ nm, $\Delta V = 5$ V. (Courtesy of Kalinin and Bonnell. Reprinted with permission from Physical Review B, 65, 125408, 2002.)

approach utilizes a coaxial probe in which a sharp, center conductor “tip” protrudes. The probe is actually the end of a transmission line resonator, which is coupled to a microwave source. The concentration of the microwave fields at the tip changes the boundary condition of the resonator, and, hence, the resonant frequency and quality factor. The magnitude of the perturbation depends on the dielectric properties of the sample. Specifically, the change in resonant frequency, $\Delta f_0/f_0 = g\Delta\epsilon'$, where g is a constant and $\Delta\epsilon'$ is the real part of the complex dielectric constant; $\Delta(1/Q) = q\Delta\epsilon''$, where q is a constant and $\Delta\epsilon''$ is the imaginary part of the complex dielectric constant. The spatial resolution of the microscope in this mode of operation is about $1\text{ }\mu\text{m}$. NFMM with coaxial resonators has been used successfully to quantitatively image sheet resistance, dielectric constant, dielectric polarization, topography, magnetic permeability, and Hall effect. Lu et al.¹³⁷ have used NFMM on ferroelectrics (001) LiNbO_3 single crystal to distinguish variations in dielectric constant and ferroelectric domains (Figure 14-22). The growth process of this crystal results in a periodic composition change that should alter the local dielectric constant. Different ferroelectric domains in this crystal have the same dielectric constant. The two periodic variations in composition and ferroelectric domain orientation are not coincident. Since $\Delta f_0/f_0$ relates to dielectric constant and ΔQ relates to loss associated with ferroelectric domain boundaries, the images should demonstrate different periodicities. This is illustrated in Figure 14-22a and b. Lu et al.¹³⁷ claimed that their NFMM has submicrometer lateral resolution, so the prognosis for further improvement of spatial resolution is good. Lee and Anlage¹³⁸ demonstrated that high-order harmonic powers acquired by NFMM can be used to spatially resolve the local nonlinearity. In their work, the grain boundary

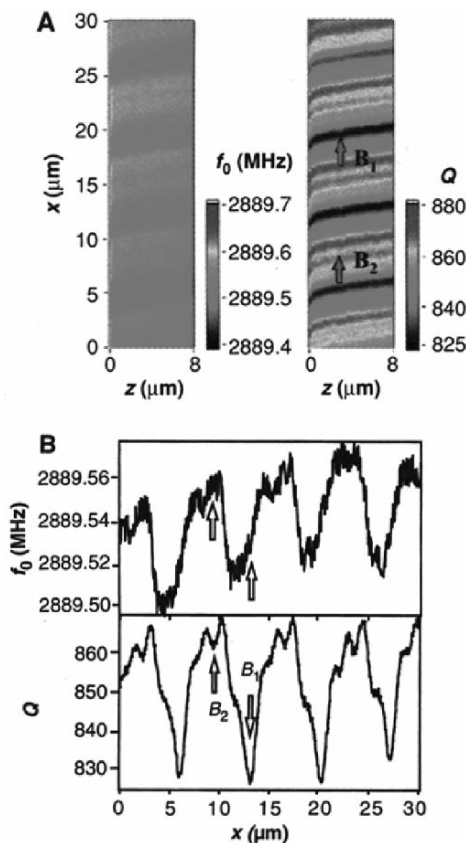


Figure 14-22. (A) Images of the profiles of periodic dielectric constant (on the left) and ferroelectric domain boundaries (on the right). (B) Line profiles of frequency (top) and quality factor (bottom) images. (See color plate.)

area of superconducting YBCO thin film deposited on an SrTiO_3 bicrystal was spatially resolved from the ratio of the second and the third power.

4 Future Trends

The rapid pace of advances in SPM shows no sign of slowing. Atomic resolution imaging is becoming ever more common and the range of local properties that can be quantified is expanding. In the next few years, several areas of research will likely push the limits of SPM even further.

Combinations of multiple modulations and high order harmonics will continue to be used in new ways to access complex properties. Table 14-2 illustrates the potential of such combinations but is not a comprehensive list of all possibilities. Neither the advanced magnetic probes nor the near-field optical probes are included. Frequency mixing has not yet been exploited.

There is recent evidence that some of the property probes may achieve atomic resolution. Eguchi et al. obtained atomic scale imaging in an SPPM (KFM).¹³⁹ Figure 14-23 compares NC-AFM and SSPM of a Ge/Si (105) surface with a model of the atomic structure. An in another

community, Rugar et al. demonstrated the measurement of a single spin of an electron.¹⁴⁰ While further work is needed to confirm the absence of artifacts in these measurements, these observations are exciting in that they portend a future of atomic resolution property imaging. Advances in high precision cantilevers and tips could unite NC-AFM imaging of structure with similar resolution imaging of potential, work function, dielectric function, etc. Concomitant advances in the physics of tip–surface interactions will be necessary when we need to interpret, for example, atomic resolution “dielectric constant.”

Finally, in the ideal world it would be useful to combine multiple probes with a variety of options for sample stimulation for comprehensive analysis. This vision is schematically illustrated in Figure 14–24. Scanning probes with multiple transport-related tips are commercially

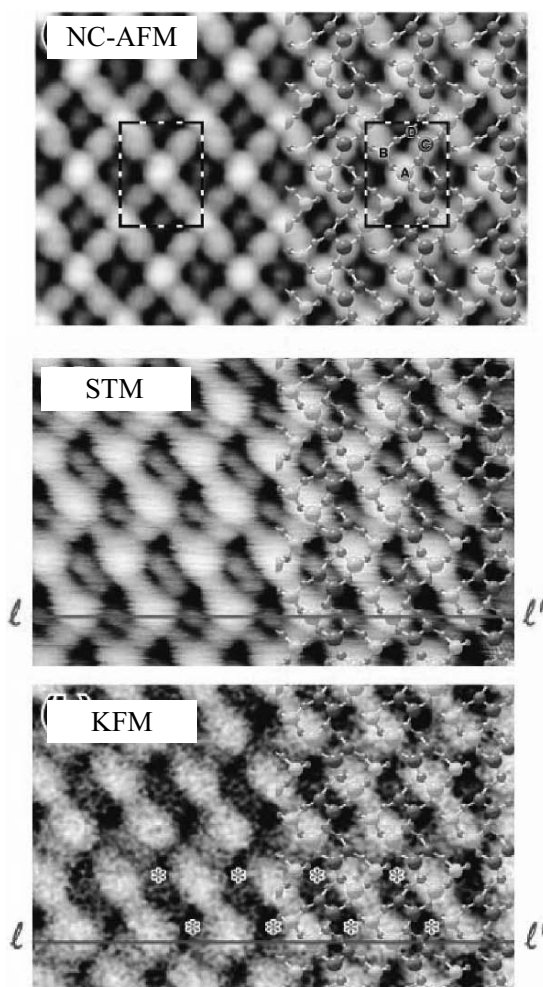


Figure 14–23. Atomic resolution NC-AFM, STM, and KFM on Ge/Si (105). [Courtesy of Eguchi et al. Reproduced with permission from Physical Review Letters, 93(26), 2004.] (See color plate.)

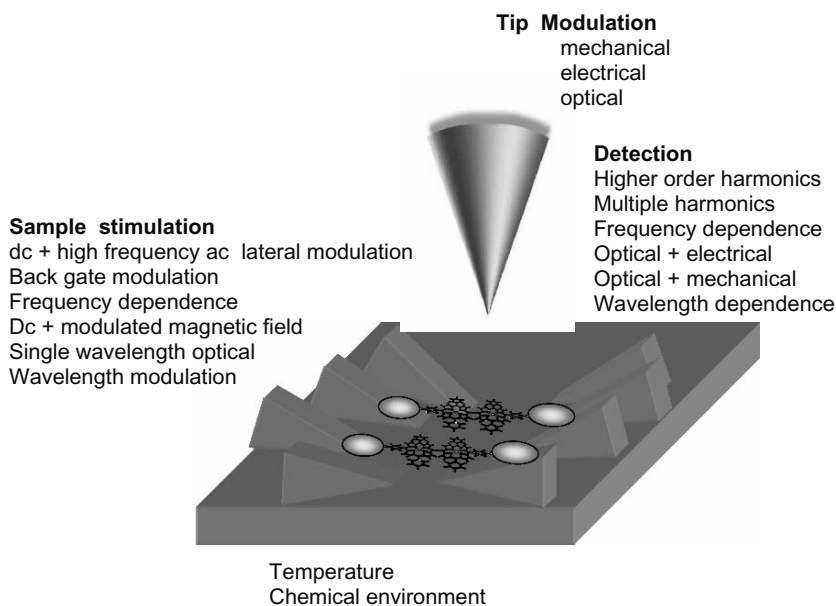


Figure 14–24. Generalized approach to SPM. An ideal probe system. Tip modulation, detection, sample stimulation, and environment might be changed to characterize the system for each specific case.

available, but there is a long way to go before combined electrical, optical, and magnetic probes become available. The gap between invention/discovery and the availability of these tools through commercial vendors is on the order of 8–10 years. Minimizing this gap would advance numerous fields that rely on high-resolution structure/property characterization.

Acknowledgments. The authors would like to acknowledge financial support from NSF (DMR05-20020, DMR-0425780), DoE (DE-FG02-00ER45813-A000). Dr. Sergei Kalinin is gratefully acknowledged for extremely helpful and informative discussion.

References

1. Compindex, as determined from COMPENDEX for 2004, 2004.
2. D. A. Bonnell, *Scanning Probe Microscopy and Spectroscopy: Theory, Techniques and Applications*, 2nd ed. New York: Wiley VCH, 2000.
3. R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy—Methods and Applications*. Cambridge, UK: Cambridge University Press, 1994.
4. G. Friedbacher and H. Fuchs, "Classification of scanning probe microscopies (Technical Report)," *Pure and Applied Chemistry*, 71, 1337–1357, 1999.
5. L. A. Bottomley, "Scanning probe microscopy," *Analytical Chemistry*, 70, 425R–475R, 1998.

6. R. Garcia and R. Perez, "Dynamic atomic force microscopy methods," *Surface Science Reports*, 47, 197–301, 2002.
7. F. J. Giessibl, "Advances in atomic force microscopy," *Reviews of Modern Physics*, 75, 949–983, 2003.
8. W. A. Hofer, A. S. Foster, and A. L. Shluger, "Theories of scanning probe microscopes at the atomic scale," *Reviews of Modern Physics*, 75, 1287–1331, 2003.
9. T. C. Lim, "The relationship between Lennard-Jones (12-6) and Morse potential functions," *Zeitschrift fur Naturforschung Section a-a Journal of Physical Sciences*, 58, 615–617, 2003.
10. T. N. Todorov, J. Hoekstra, and A. P. Sutton, "Current-induced forces in atomic-scale conductors," *Philosophical Magazine B—Physics of Condensed Matter Statistical Mechanics Electronic Optical and Magnetic Properties*, 80, 421–455, 2000.
11. M. Gauthier and M. Tsukada, "Theory of non-contact dissipation force microscopy," *Physical Review B*, 60, 11716–11722, 1999.
12. S. Morita, R. Wiesendanger, and E. Meyer, *Non-Contact Atomic Force Microscopy*. Berlin: Springer, 2002.
13. K. Fukui, H. Onishi, and Y. Iwasawa, "Imaging of individual formate ions adsorbed on $\text{TiO}_2(110)$ surface by non-contact atomic force microscopy," *Chemical Physics Letters*, 280, 296–301, 1997.
14. S. H. Ke, T. Uda, and K. Terakura, "First-principles studies of tip-sample interaction and STM-AFM image formation on $\text{TiO}_2(110)-1 \times 1$ and $\text{TiO}_2(110)-1 \times 2$ surfaces," *Physical Review B*, 65, 125417, 2002.
15. A. San Paulo and R. Garcia, "Tip-surface forces, amplitude, and energy dissipation in amplitude-modulation (tapping mode) force microscopy," *Physical Review B*, 64, 193411, 2001.
16. C. Moller, M. Allen, V. Elings, A. Engel, and D. J. Muller, "Tapping-mode atomic force microscopy produces faithful high-resolution images of protein surfaces," *Biophysical Journal*, 77, 1150–1158, 1999.
17. D. Oesterhelt and W. Stoeckenius, "Isolation of the cell membrane of *Halobacterium halobium* and its fraction into red and purple membrane," *Methods in Enzymology*, 31, 667–678, 1974.
18. F. M. Ohnesorge, "Towards atomic resolution non-contact dynamic force microscopy in a liquid," *Surface and Interface Analysis*, 27, 797–797, 1999.
19. F. M. Ohnesorge, "Towards atomic resolution non-contact dynamic force microscopy in a liquid," *Surface and Interface Analysis*, 27, 379–385, 1999.
20. F. J. Giessibl, "Forces and frequency shifts in atomic-resolution dynamic-force microscopy," *Physical Review B*, 56, 16010–16015, 1997.
21. F. J. Giessibl and H. Bielefeldt, "Physical interpretation of frequency-modulation atomic force microscopy," *Physical Review B*, 61, 9968–9971, 2000.
22. L. N. Kantorovich, A. S. Foster, A. L. Shluger, and A. M. Stoneham, "Role of image forces in non-contact scanning force microscope images of ionic surfaces," *Surface Science*, 445, 283–299, 2000.
23. M. Guggisberg, M. Bammerlin, C. Loppacher, O. Pfeiffer, A. Abdurixit, V. Barwich, R. Bennewitz, A. Baratoff, E. Meyer, and H. J. Guntherodt, "Separation of interactions by noncontact force microscopy," *Physical Review B*, 61, 11151–11155, 2000.
24. S. Sounilhac, E. Barthel, and F. Creuzet, "The electrostatic contribution to the long-range interactions between tungsten and oxide surfaces under ultrahigh vacuum," *Applied Surface Science*, 140, 411–414, 1999.
25. C. Loppacher, M. Bammerlin, M. Guggisberg, S. Schar, R. Bennewitz, A. Baratoff, E. Meyer, and H. J. Guntherodt, "Dynamic force microscopy of

- copper surfaces: Atomic resolution and distance dependence of tip-sample interaction and tunneling current," *Physical Review B*, 62, 16944–16949, 2000.
26. R. Perez, I. Stich, M. C. Payne, and K. Terakura, "Surface-tip interactions in noncontact atomic-force microscopy on reactive surfaces: Si(111)," *Physical Review B*, 58, 10835–10849, 1998.
 27. A. L. Shluger, A. I. Livshits, A. S. Foster, and C. R. A. Catlow, "Models of image contrast in scanning force microscopy on insulators," *Journal of Physics-Condensed Matter*, 11, R295–R322, 1999.
 28. A. L. Shluger and A. L. Rohl, "A model of the interaction of ionic tips with ionic surfaces for interpretation of scanning force microscope images," *Topics in Catalysis*, 3, 221–247, 1996.
 29. M. A. Zaibi, J. P. Lacharme, and C. A. Sebenne, "Water vapour adsorption on the Si(111)-(7 × 7) surface," *Surface Science*, 377, 639–643, 1997.
 30. P. V. Sushko, A. S. Foster, L. N. Kantorovich, and A. L. Shluger, "Investigating the effects of silicon tip contamination in noncontact scanning force microscopy (SFM)," *Applied Surface Science*, 145, 608–612, 1999.
 31. R. Bennewitz, A. S. Foster, L. N. Kantorovich, M. Bamberlin, C. Loppacher, S. Schar, M. Guggisberg, E. Meyer, and A. L. Shluger, "Atomically resolved edges and kinks of NaCl islands on Cu(111): Experiment and theory," *Physical Review B*, 62, 2074–2084, 2000.
 32. S. Hembacher, F. J. Giessibl, and J. Mannhart, "Force microscopy with light-atom probes," *Science*, 305, 380–383, 2004.
 33. A. Sasahara, H. Uetsuka, and K. Onishi, "Non-contact-mode atomic force microscopy observation of α - Al_2O_3 (0001) surface," *Japanese Journal of Applied Physics Part 1—Regular Papers Short Notes & Review Papers*, 39, 3773–3776, 2000.
 34. K. Fukui, H. Onishi, and Y. Iwasawa, "Atom-resolved image of the TiO_2 (110) surface by non-contact atomic force microscopy," *Physical Review Letters*, 79, 4202–4205, 1997.
 35. T. Kubo and H. Nozoye, "Surface structure of SrTiO_3 (100)-(root 5 × root 5)-R26.6 degrees," *Physical Review Letters*, 86, 1801–1804, 2001.
 36. H. Hosoi, K. Sueoka, K. Hayakawa, and K. Mukasa, "Atomic resolved imaging of cleaved NiO (100) surfaces by NC-AFM," *Applied Surface Science*, 157, 218–221, 2000.
 37. K. Fukui, Y. Namai, and Y. Iwasawa, "Imaging of surface oxygen atoms and their defect structures on CeO_2 (111) by noncontact atomic force microscopy," *Applied Surface Science*, 188, 252–256, 2002.
 38. R. V. Coleman, Q. Xue, Y. Gong, and P. B. Price, "Atomic-force microscope study of etched tracks of low-energy heavy-ions in mica," *Surface Science*, 297, 359–370, 1993.
 39. S. Suzuki, Y. Ohminami, T. Tsutsumi, M. M. Shoaib, M. Ichikawa, and K. Asakura, "The first observation of an atomic scale noncontact AFM image of MoO_3 (010)," *Chemistry Letters*, 32, 1098–1099, 2003.
 40. K. Ueno, I. H. Inoue, H. Akoh, M. Kawasaki, Y. Tokura, and H. Takagi, "Field-effect transistor on SrTiO_3 with sputtered Al_2O_3 gate insulator," *Applied Physics Letters*, 83, 1755–1757, 2003.
 41. A. Gunhold, L. Beuermann, K. Gomann, G. Borchardt, V. Kempter, W. Maus-Friedrichs, S. Piskunov, E. A. Kotomin, and S. Dorfman, "Study of the electronic and atomic structure of thermally treated SrTiO_3 (110) surfaces," *Surface and Interface Analysis*, 35, 998–1003, 2003.
 42. S. Sekiguchi, M. Fujimoto, M. G. Kang, S. Koizumi, S. B. Cho, and J. Tanaka, "Structure analysis of SrTiO_3 (111) polar surfaces," *Japanese Journal of Applied Physics Part 1—Regular Papers Short Notes & Review Papers*, 37, 4140–4143, 1998.

43. M. Ashino, Y. Sugawara, S. Morita, and M. Ishikawa, "Atomic resolution noncontact atomic force and scanning tunneling microscopy of $\text{TiO}_2(110)-(1 \times 1)$ and $-(1 \times 2)$: Simultaneous imaging of surface structures and electronic states," *Physical Review Letters*, 86, 4334–4337, 2001.
44. C. L. Pang, H. Raza, S. A. Haycock, and G. Thornton, "Imaging reconstructed TiO_2 surfaces with non-contact atomic force microscopy," *Applied Surface Science*, 157, 233–238, 2000.
45. K. Takayanagi, Y. Tanishiro, M. Takahashi, and S. Takahashi, "Structural-analysis of $\text{Si}(111)-7 \times 7$ by UHV-transmission electron-diffraction and microscopy," *Journal of Vacuum Science & Technology a—Vacuum Surfaces and Films*, 3, 1502–1506, 1985.
46. G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, " 7×7 reconstruction on $\text{Si}(111)$ resolved in real space," *Physical Review Letters*, 50, 120–123, 1983.
47. S. Kitamura and M. Iwatsuki, "Observation of 7×7 Reconstructed structure on the silicon (111) surface using ultrahigh-vacuum noncontact atomic-force microscopy," *Japanese Journal of Applied Physics Part 2—Letters*, 34, L145–L148, 1995.
48. F. J. Giessibl, "Atomic-Resolution of the Silicon (111)-(7×7) Surface by atomic-force microscopy," *Science*, 267, 68–71, 1995.
49. A. Schwarz, W. Allers, U. D. Schwarz, and R. Wiesendanger, "Simultaneous imaging of the In and As sublattice on $\text{InAs}(110)-(1 \times 1)$ with dynamic scanning force microscopy," *Applied Surface Science*, 140, 293–297, 1999.
50. K. Yokoyama, T. Ochi, A. Yoshimoto, Y. Sugawara, and S. Morita, "Atomic resolution imaging on $\text{Si}(100)2 \times 1$ and $\text{Si}(100)2 \times 1:\text{H}$ surfaces with noncontact atomic force microscopy," *Japanese Journal of Applied Physics Part 2—Letters*, 39, L113–L115, 2000.
51. N. Uehara, H. Hosoi, K. Sueoka, and K. Mukasa, "Non-contact atomic force microscopy observation on $\text{GaAs}(110)$ surface with tip-induced relaxation," *Japanese Journal of Applied Physics Part 1—Regular Papers Short Notes & Review Papers*, 43, 4676–4678, 2004.
52. Y. Seo, H. Choe, and W. Jhe, "Atomic-resolution noncontact atomic force microscopy in air," *Applied Physics Letters*, 83, 1860–1862, 2003.
53. S. Hembacher, F. J. Giessibl, J. Mannhart, and C. F. Quate, "From the cover: Revealing the hidden atom in graphite by low-temperature atomic force microscopy," *Proceedings of the National Academy of Sciences USA*, 100, 12539–12542, 2003.
54. R. Bennewitz, O. Pfeiffer, S. Schar, V. Barwich, E. Meyer, and L. N. Kantorovich, "Atomic corrugation in nc-AFM of alkali halides," *Applied Surface Science*, 188, 232–237, 2002.
55. F. J. Giessibl and M. Reichling, "Investigating atomic details of the $\text{CaF}_2(111)$ surface with a qPlus sensor," *Nanotechnology*, 16, S118–S124, 2005.
56. J. M. R. Weaver and D. W. Abraham, "High-resolution atomic force microscopy potentiometry," *Journal of Vacuum Science & Technology B*, 9, 1559–1561, 1991.
57. M. Nonnenmacher, M. P. Oboyle, and H. K. Wickramasinghe, "Kelvin probe force microscopy," *Applied Physics Letters*, 58, 2921–2923, 1991.
58. A. K. Henning and T. Hochwitz, "Scanning probe microscopy for 2-D semiconductor dopant profiling and device failure analysis," *Materials Science and Engineering B—Solid State Materials for Advanced Technology*, 42, 88–98, 1996.

59. H. O. Jacobs, P. Leuchtmann, O. J. Homan, and A. Stemmer, "Resolution and contrast in Kelvin probe force microscopy," *Journal of Applied Physics*, 84, 1168–1173, 1998.
60. S. V. Kalinin and D. A. Bonnell, "Local potential and polarization screening on ferroelectric surfaces," *Physical Review B*, 63, 125411, 2001.
61. S. Cunningham, I. A. Larkin, and J. H. Davis, "Noncontact scanning probe microscope potentiometry of surface charge patches: Origin and interpretation of time-dependent signals," *Applied Physics Letters*, 73, 123–125, 1998.
62. S. V. Kalinin, C. Y. Johnson, and D. A. Bonnell, "Domain polarity and temperature induced potential inversion on the BaTiO₃(100) surface," *Journal of Applied Physics*, 91, 3816–3823, 2002.
63. K. Franke, H. Huelz, and M. Wehnacht, "How to extract spontaneous polarization information from experimental data in electric force microscopy," *Surface Science*, 415, 178–182, 1998.
64. C. Donolato, "Electrostatic tip-sample interaction in immersion force microscopy of semiconductors," *Physical Review B*, 54, 1478–1481, 1996.
65. Y. Leng, C. C. Williams, L. C. Su, and G. B. Stringfellow, "Atomic ordering of GaInP studied by Kelvin probe force microscopy," *Applied Physics Letters*, 66, 1264–1266, 1995.
66. M. Tanimoto and O. Vatel, "Kelvin probe force microscopy for characterization of semiconductor devices and processes," *Journal of Vacuum Science & Technology B*, 14, 1547–1551, 1996.
67. T. Hochwitz, A. K. Henning, C. Levey, C. Daghljan, J. Slinkman, J. Never, P. Kaszuba, R. Gluck, R. Wells, J. Pekarik, and R. Finch, "Imaging integrated circuit dopant profiles with the force-based scanning Kelvin probe microscope," *Journal of Vacuum Science & Technology B*, 14, 440–446, 1996.
68. M. Fujihira, "Kelvin probe force microscopy of molecular surfaces," *Annual Review of Materials Science*, 29, 353–380, 1999.
69. X. Q. Chen, H. Yamada, T. Horiuchi, K. Matsushige, S. Watanabe, M. Kawai, and P. S. Weiss, "Surface potential of ferroelectric thin films investigated by scanning probe microscopy," *Journal of Vacuum Science & Technology B*, 17, 1930–1934, 1999.
70. T. Tybell, C. H. Ahn, and J. M. Triscone, "Ferro-electricity in thin perovskite films," *Applied Physics Letters*, 75, 856–858, 1999.
71. P. M. Bridger, Z. Z. Bandic, E. C. Piquette, and T. C. McGill, "Measurement of induced surface charges, contact potentials, and surface states in GaN by electric force microscopy," *Applied Physics Letters*, 74, 3522–3524, 1999.
72. Q. Xu and J. W. P. Hsu, "Electrostatic force microscopy studies of surface defects on GaAs/Ge films," *Journal of Applied Physics*, 85, 2465–2472, 1999.
73. A. Chavezpiron, O. Vatel, M. Tanimoto, H. Ando, H. Iwamura, and H. Kanbe, "Nanometerscale imaging of potential profiles in optically-excited N-I-P-I heterostructure using Kelvin probe force microscopy," *Applied Physics Letters*, 67, 3069–3071, 1995.
74. T. Meoded, R. Shikler, N. Fried, and Y. Rosenwaks, "Direct measurement of minority carriers diffusion length using Kelvin probe force microscopy," *Applied Physics Letters*, 75, 2435–2437, 1999.
75. S. V. Kalinin and D. A. Bonnell, "Dynamic behavior of domain-related topography and surface potential on the BaTiO₃ (100) surface by variable temperature scanning surface potential microscopy," *Zeitschrift fur Metallkunde*, 90, 983–989, 1999.
76. J. Lu, E. Delamarche, L. Eng, R. Bennewitz, E. Meyer, and H. J. Guntherodt, "Kelvin probe force microscopy on surfaces: Investigation

- of the surface potential of self-assembled monolayers on gold," *Langmuir*, 15, 8184–8188, 1999.
77. H. Sugimura, K. Hayashi, N. Saito, N. Nakagiri, and O. Takai, "Surface potential microscopy for organized molecular systems," *Applied Surface Science*, 188, 403–410, 2002.
 78. R. A. Alvarez and D. A. Bonnell, Unpublished (presented on Fall MRS 2004), 2005.
 79. M. Abplanalp, L. M. Eng, and P. Gunter, "Mapping the domain distribution at ferroelectric surfaces by scanning force microscopy," *Applied Physics a—Materials Science & Processing*, 66, S231–S234, 1998.
 80. R. Shao and D. A. Bonnell, Unpublished, 2005.
 81. P. De Wolf, T. Clarysse, W. Vandervorst, and L. Hellemans, "Low weight spreading resistance profiling of ultrashallow dopant profiles," *Journal of Vacuum Science & Technology B*, 16, 401–405, 1998.
 82. P. De Wolf, R. Stephenson, T. Trenkler, T. Clarysse, T. Hantschel, and W. Vandervorst, "Status and review of two-dimensional carrier and dopant profiling using scanning probe microscopy," *Journal of Vacuum Science & Technology B*, 18, 361–368, 2000.
 83. P. De Wolf, J. Snauwaert, L. Hellemans, T. Clarysse, W. Vandervorst, M. D'Olieslaeger, and D. Quaeys, "Lateral and vertical dopant profiling in semiconductors by atomic force microscopy using conducting tips," *Journal of Vacuum Science & Technology a—Vacuum Surfaces and Films*, 13, 1699–1704, 1995.
 84. J. R. Matey and J. Blanc, "Scanning capacitance microscopy," *Journal of Applied Physics*, 57, 1437–1444, 1985.
 85. R. C. Barrett and C. F. Quate, "Charge storage in a nitride-oxide-silicon medium by scanning capacitance microscopy," *Journal of Applied Physics*, 70, 2725–2733, 1991.
 86. Y. Huang, C. C. Williams, and M. A. Wendman, "Quantitative two-dimensional dopant profiling of abrupt dopant profiles by cross-sectional scanning capacitance microscopy," *Journal of Vacuum Science & Technology a—Vacuum Surfaces and Films*, 14, 1168–1171, 1996.
 87. T. Hantschel, P. Niedermann, T. Trenkler, and W. Vandervorst, "Highly conductive diamond probes for scanning spreading resistance microscopy," *Applied Physics Letters*, 76, 1603–1605, 2000.
 88. P. De Wolf, E. Brazel, and A. Erickson, "Electrical characterization of semiconductor materials and devices using scanning probe microscopy," *Materials Science in Semiconductor Processing*, 4, 71–76, 2001.
 89. J. F. Marchiando and J. J. Kopanski, "Regression procedure for determining the dopant profile in semiconductors from scanning capacitance microscopy data," *Journal of Applied Physics*, 92, 5798–5809, 2002.
 90. J. Yang and F. C. J. Kong, "Simulation of interface states effect on the scanning capacitance microscopy measurement of p-n junctions," *Applied Physics Letters*, 81, 4973–4975, 2002.
 91. S. Lanyi, J. Torok, and P. Reherek, "Imaging conducting surfaces and dielectric films by a scanning capacitance microscope," *Journal of Vacuum Science & Technology B*, 14, 892–896, 1996.
 92. S. Belaidi, P. Girard, and G. Leveque, "Electrostatic forces acting on the tip in atomic force microscopy: Modelization and comparison with analytic expressions," *Journal of Applied Physics*, 81, 1023–1030, 1997.
 93. S. J. Tans and C. Dekker, "Molecular transistors—Potential modulations along carbon nanotubes," *Nature*, 404, 834–835, 2000.
 94. T. W. Tombler, C. W. Zhou, J. Kong, and H. J. Dai, "Gating individual nanotubes and crosses with scanning probes," *Applied Physics Letters*, 76, 2412–2414, 2000.

95. A. Bachtold, M. S. Fuhrer, S. Plyasunov, M. Forero, E. H. Anderson, A. Zettl, and P. L. McEuen, "Scanned probe microscopy of electronic transport in carbon nanotubes," *Physical Review Letters*, 84, 6082–6085, 2000.
96. M. Freitag, A. T. Johnson, S. V. Kalinin, and D. A. Bonnell, "Role of single defects in electronic transport through carbon nanotube field-effect transistors," *Physical Review Letters*, 89, art. No. 216801, 2002.
97. S. V. Kalinin, D. A. Bonnell, M. Freitag, and A. T. Johnson, "Tip-gating effect in scanning impedance microscopy of nanoelectronic devices," *Applied Physics Letters*, 81, 5219–5221, 2002.
98. Z. Fan and J. G. Lu, "Electrical properties of ZnO nanowire field effect transistors characterized with scanning probes," *Applied Physics Letters*, 86, 032111, 2005.
99. S. V. Kalinin and D. A. Bonnell, "Scanning impedance microscopy of electroactive interfaces," *Applied Physics Letters*, 78, 1306–1308, 2001.
100. S. V. Kalinin and D. A. Bonnell, "Scanning impedance microscopy of an active Schottky barrier diode," *Journal of Applied Physics*, 91, 832–839, 2002.
101. S. V. Kalinin, S. Jesse, J. Shin, A. P. Baddorf, M. A. Guillorn, and D. B. Geohegan, "Scanning probe microscopy imaging of frequency dependent electrical transport through carbon nanotube networks in polymers," *Nanotechnology*, 15, 907–912, 2004.
102. R. Shao, S. V. Kalinin, and D. A. Bonnell, "Local impedance imaging and spectroscopy of polycrystalline ZnO using contact atomic force microscopy," *Applied Physics Letters*, 82, 1869–1871, 2003.
103. R. O'Hayre, M. Lee, and F. B. Prinz, "Ionic and electronic impedance imaging using atomic force microscopy," *Journal of Applied Physics*, 95, 8382–8392, 2004.
104. R. O'Hayre, G. Feng, W. D. Nix, and F. B. Prinz, "Quantitative impedance measurement using atomic force microscopy," *Journal of Applied Physics*, 96, 3540–3549, 2004.
105. C. Durkan and M. E. Welland, "Investigations into local ferroelectric properties by atomic force microscopy," *Ultramicroscopy*, 82, 141–148, 2000.
106. A. Gruverman, O. Kolosov, J. Hatano, K. Takahashi, and H. Tokumoto, "Domain-Structure and Polarization Reversal in Ferroelectrics Studied by Atomic-Force Microscopy," *Journal of Vacuum Science & Technology B*, 13, 1095–1099, 1995.
107. S. V. Kalinin, E. Karapetian, and M. Kachanov, "Nanoelectromechanics of piezoresponse force microscopy," *Physical Review B*, 70, art. No. 184101, 2004.
108. L. M. Eng, H. J. Guntherodt, G. A. Schneider, U. Kopke, and J. M. Saldana, "Nanoscale reconstruction of surface crystallography from three-dimensional polarization distribution in ferroelectric barium-titanate ceramics," *Applied Physics Letters*, 74, 233–235, 1999.
109. A. Roelofs, U. Bottger, R. Waser, F. Schlaphof, S. Trogisch, and L. M. Eng, "Differentiating 180 degrees and 90 degrees switching of ferroelectric domains with three-dimensional piezoresponse force microscopy," *Applied Physics Letters*, 77, 3444–3446, 2000.
110. C. Harnagea, A. Pignolet, M. Alexe, and D. Hesse, "Piezoresponse scanning force microscopy: What quantitative information can we really get out of piezoresponse measurements on ferroelectric thin films," *Integrated Ferroelectrics*, 38, 667–673, 2001.
111. B. J. Rodriguez, A. Gruverman, A. I. Kingon, R. J. Nemanich, and J. S. Cross, "Three-dimensional high-resolution reconstruction of polarization in ferroelectric capacitors by piezoresponse force microscopy," *Journal of Applied Physics*, 95, 1958–1962, 2004.

112. A. Gruverman, B. J. Rodriguez, R. J. Nemanich, and A. I. Kingon, "Nanoscale observation of photoinduced domain pinning and investigation of imprint behavior in ferroelectric thin films," *Journal of Applied Physics*, 92, 2734–2739, 2002.
113. M. Alexe, A. Gruverman, C. Harnagea, N. D. Zakharov, A. Pignolet, D. Hesse, and J. F. Scott, "Switching properties of self-assembled ferroelectric memory cells," *Applied Physics Letters*, 75, 1158–1160, 1999.
114. S. V. Kalinin, A. Gruverman, and D. A. Bonnell, "Quantitative analysis of nanoscale switching in $\text{SrBi}_2\text{Ta}_2\text{O}_9$ thin films by piezoresponse force microscopy," *Applied Physics Letters*, 85, 795–797, 2004.
115. L. M. Eng, M. Bammerlin, C. Loppacher, M. Guggisberg, R. Bennewitz, R. Luthi, E. Meyer, T. Huser, H. Heinzelmann, and H. J. Guntherodt, "Ferroelectric domain characterisation and manipulation: A challenge for scanning probe microscopy," *Ferroelectrics*, 222, 411–420, 1999.
116. X. M. Lu, F. Schlaphof, S. Grafstrom, C. Loppacher, L. M. Eng, G. Suchaneck, and G. Gerlach, "Scanning force microscopy investigation of the $\text{Pb}(\text{Zr}_{0.25}\text{Ti}_{0.75})\text{O}_3/\text{Pt}$ interface," *Applied Physics Letters*, 81, 3215–3217, 2002.
117. A. L. Kholkin, V. V. Shvartsman, A. Y. Emelyanov, R. Poyato, M. L. Calzada, and L. Pardo, "Stress-induced suppression of piezo-electric properties in $\text{PbTiO}_3\text{:La}$ thin films via scanning force microscopy," *Applied Physics Letters*, 82, 2127–2129, 2003.
118. M. Abplanalp, J. Fousek, and P. Gunter, "Higher order ferroic switching induced by scanning force microscopy," *Physical Review Letters*, 86, 5799–5802, 2001.
119. M. Labardi, C. Polop, V. Likodimos, L. Pardi, M. Allegrini, E. Vasco, and C. Zaldo, "Surface deformation and ferroelectric domain switching induced by a force microscope tip on a La-modified PbTiO_3 thin film," *Applied Physics Letters*, 83, 2028–2030, 2003.
120. A. L. Roytburd, S. P. Alpay, V. Nagarajan, C. S. Ganpule, S. Aggarwal, E. D. Williams, and R. Ramesh, "Measurement of internal stresses via the polarization in epitaxial ferroelectric films," *Physical Review Letters*, 85, 190–193, 2000.
121. C. S. Ganpule, A. Stanishevsky, S. Aggarwal, J. Melngailis, E. Williams, R. Ramesh, V. Joshi, and C. P. de Araujo, "Scaling of ferroelectric and piezoelectric properties in $\text{Pt}/\text{SrBi}_2\text{Ta}_2\text{O}_9/\text{Pt}$ thin films," *Applied Physics Letters*, 75, 3874–3876, 1999.
122. M. Alexe, C. Harnagea, D. Hesse, and U. Gosele, "Patterning and switching of nanosize ferroelectric memory cells," *Applied Physics Letters*, 75, 1793–1795, 1999.
123. J. J. Urban, J. E. Spanier, O. Y. Lian, W. S. Yun, and H. Park, "Single-crystalline barium titanate nanowires," *Advanced Materials*, 15, 423–426, 2003.
124. W. S. Yun, J. J. Urban, Q. Gu, and H. Park, "Ferroelectric properties of individual barium titanate nanowires investigated by scanned probe microscopy," *Nano Letters*, 2, 447–450, 2002.
125. S. Hong, J. Woo, H. Shin, J. U. Jeon, Y. E. Pak, E. L. Colla, N. Setter, E. Kim, and K. No, "Principle of ferroelectric domain imaging using atomic force microscope," *Journal of Applied Physics*, 89, 1377–1386, 2001.
126. S. V. Kalinin and D. A. Bonnell, "Imaging mechanism of piezoresponse force microscopy of ferroelectric surfaces," *Physical Review B*, 65, 125408, 2002.
127. S. V. Kalinin, "Ph.D. thesis," in *Materials Science and Engineering*. Philadelphia: University of Pennsylvania, 2002.

128. C. S. Ganpule, "Nanoscale phenomena in ferroelectric thin films." Ph.D. thesis. College Park: University of Maryland, 2001.
129. C. Harnagea, "Local piezoelectric response and domain structures in ferroelectric thin films investigated by voltage modulated force microscopy." Dr. Rer. Nat. thesis. Halle: Martin-Luther-Universitaet Halle Wittenberg, 2001.
130. R. Shao and D. A. Bonnell, "Scanning probes of nonlinear properties in complex materials," *Japanese Journal of Applied Physics Part 1—Regular Papers Short Notes & Review Papers*, 43, 4471–4476, 2004.
131. Q. M. Zhang, W. Y. Pan, and L. E. Cross, "Laser interferometer for the study of piezoelectric and electrostrictive strains," *Journal of Applied Physics*, 63, 2492–2496, 1988.
132. I. L. Guy and Z. Zheng, "Piezoelectricity and electrostriction in ferroelectric polymers," *Ferroelectrics*, 264, 1691–1696, 2001.
133. C. Gao and X. D. Xiang, "Quantitative microwave near-field microscopy of dielectric properties," *Review of Scientific Instruments*, 69, 3846–3851, 1998.
134. Y. S. Cho, A. Kirihaara, and T. Saeki, "Scanning nonlinear dielectric microscope," *Review of Scientific Instruments*, 67, 2297–2303, 1996.
135. D. E. Steinhauer, C. P. Vlahacos, F. C. Wellstood, S. M. Anlage, C. Canedy, R. Ramesh, A. Stanishevsky, and J. Melngailis, "Imaging of microwave permittivity, tunability, and damage recovery in (Ba, Sr)TiO₃ thin films," *Applied Physics Letters*, 75, 3180–3182, 1999.
136. D. E. Steinhauer, C. P. Vlahacos, S. K. Dutta, F. C. Wellstood, and S. M. Anlage, "Surface resistance imaging with a scanning near-field microwave microscope," *Applied Physics Letters*, 71, 1736–1738, 1997.
137. Y. L. Lu, T. Wei, F. Duewer, Y. Q. Lu, N. B. Ming, P. G. Schultz, and X. D. Xiang, "Nondestructive imaging of dielectric-constant profiles and ferroelectric domains with a scanning-tip microwave near-field microscope," *Science*, 276, 2004–2006, 1997.
138. S. C. Lee and S. M. Anlage, "Spatially-resolved nonlinearity measurements of YBa₂Cu₃O_{7-δ} bicrystal grain boundaries," *Applied Physics Letters*, 82, 1893–1895, 2003.
139. T. Eguchi, Y. Fujikawa, K. Akiyama, T. An, M. Ono, T. Hashimoto, Y. Morikawa, K. Terakura, T. Sakurai, M. G. Lagally, and Y. Hasegawa, "Imaging of all dangling bonds and their potential on the Ge/Si(105) surface by noncontact atomic force microscopy," *Physical Review Letters*, 93, art. No. 266102, 2004.
140. D. Rugar, R. Budakian, H. J. Mamin, and B. W. Chui, "Single spin detection by magnetic resonance force microscopy," *Nature*, 430, 329–332, 2004.

Scanning Tunneling Microscopy in Surface Science

Peter Sutter

1 Introduction

After its invention in 1982 by Gerd Binnig and Heinrich Rohrer (Binnig et al., 1982, 1983), who were awarded the 1986 Nobel Prize for this discovery, scanning tunneling microscopy (STM) has rapidly become a standard technique for high-resolution imaging of conducting surfaces. As such, it has revolutionized the way surface science is conducted. Having relied almost entirely on diffraction methods for determining surface structures, and electron spectroscopy for measuring surface chemistry and electronic structure, surface scientists immediately embraced the powerful imaging and spectroscopy capabilities of STM. Earlier techniques for high-resolution surface microscopy include field ion microscopy (Muller and Tsong, 1969), low-energy electron microscopy (Bauer, 1998), ultrahigh-vacuum scanning electron microscopy (Venables, 2000), and reflection electron microscopy (Yagi, 1982). Field ion microscopy, the first surface imaging technique to routinely provide atomic resolution on metals over small sample areas, has recently seen a revival in the form of modern atom probes that provide tomographic images of three-dimensional sample volumes (Cerezo et al., 2001), nicely complementing the surface imaging capabilities of STM. The other techniques, none of which achieves atomic resolution, have been developed further for specific applications. As an example, low-energy electron microscopy has become a powerful technique for real-time microscopy of fast surface processes, such as epitaxial growth or surface reactions (Bauer, 1998). STM itself has recently been at the center of yet another scientific revolution, enabling systematic studies on individual structures composed of a small number of atoms or molecules, and on the order of a few nanometers in size.

This chapter attempts to provide an introduction of the basic concepts of STM, together with illustrations of applications of the technique. Over two decades after its invention, the applications of STM for imaging, spectroscopy, and manipulation at the atomic level have clearly become too numerous to allow for a comprehensive review. In view of the wide range of applications and overall maturity of the

technique, a historic overview may also be of limited value. Hence, I have tried to identify representative examples in the current literature, which are discussed briefly to illustrate the different uses of STM. The connection to historic developments can in most cases be made easily via literature references. For a more in-depth discussion of basic concepts and application examples, several dedicated monographs on STM and other scanning probe techniques (Chen, 1993; Wiesendanger, 1994; Güntherodt and Wiesendanger, 1992; Bonnell, 2001) represent an invaluable resource.

This chapter is organized as follows. In Section 2 basic principles of STM imaging are introduced, and the imaging methodology as well as practical and instrumentation requirements are discussed. The approach taken in surface imaging by STM is illustrated by the example of silicon surfaces. Section 3 highlights an application of STM that has gained ever-increasing importance: atomic scale spectroscopy. It provides a survey of the spectroscopy capabilities of STM, and introduces a variety of techniques for local spectroscopy and spectroscopic imaging. In addition, pathways toward obtaining chemical and element specificity at the atomic scale—traditionally a weakness of STM—are discussed. The extension of the operating conditions of STM to high and low temperatures has opened up new avenues of investigation. Variable temperature STM of dynamic surface processes as well as atom and molecule manipulation at cryogenic temperatures are the topics of Section 4. Section 5 discusses STM imaging and spectroscopy on subsurface structures, using ballistic electrons to probe buried interfaces or cross-sectional STM on cleavage faces of III–V semiconductors to image embedded nanostructures. The chapter concludes with a brief discussion of STM image simulation techniques in Section 6.

2 Basic Principles of STM Imaging

2.1 Elastic Vacuum Tunneling and Scanning Tunneling Microscopy

STM is based on the vacuum tunneling of electrons between two solids, one of which is a sharp tip and the other a sample. To remove an electron from a solid and bring it into vacuum with zero kinetic energy requires energy equal to the work function $\phi = E_{\text{vac}} - E_{\text{F}}$, i.e., the difference between the vacuum level, E_{vac} , and the Fermi energy of the material, E_{F} . To move an electron from one solid to another, it has to cross the same vacuum barrier. If the two solids are separated by a microscopic distance of the order of the decay length of electronic states into the vacuum (typically about 1–5 nm), electrons can cross the barrier by quantum mechanical tunneling (Figure 15–1).

If, say, two metals are brought to within tunneling distance, a rapid charge transfer takes place between them until an equilibrium state is established in which their Fermi levels are aligned. Once in equilibrium, no net charge is transferred on average. Consider now a slightly modified situation, in which metal A is biased relative to B by applying

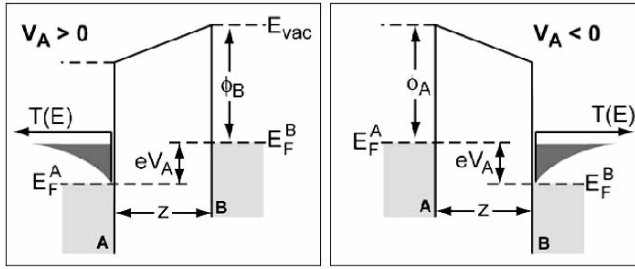


Figure 15-1. Band diagrams (energy vs. distance along the tunneling direction) for vacuum tunneling between two metal electrodes, A and B.

a voltage, V_A . The bias offsets the Fermi energies on either side of the vacuum barrier by eV_A . Under steady-state conditions, this potential difference establishes a net tunneling current, I , between A and B.

In a simple planar tunneling model using the Wentzel–Kramers–Brillouin (WKB) approximation, the tunneling current is given by an integral over the energy range between the Fermi energies on either side of the vacuum gap,

$$I = \int_0^{eV_A} \rho_A(E) \rho_B(E - eV_A) T(E, eV_A) dE \quad (1)$$

where ρ_A and ρ_B denote the local density of states at energy E at the surface of A and B, respectively, and $T(E, eV_A)$ is the tunneling transmission probability for electrons with energy E and applied bias V_A :

$$T(E, eV_A) = \exp\left(-\frac{2z\sqrt{2m}}{\hbar}\right) \sqrt{\frac{\phi_A + \phi_B}{2} + \frac{eV_A}{2} - E} \quad (2)$$

Here, $\phi_{A,B}$ are the work functions of the two solids, z is their separation, and m is the mass of the electron. Evaluation of T for positive ($V_A > 0$) and negative bias ($V_A < 0$) shows that the transmission probability is highest for electrons at the Fermi energy of the material that is negatively biased, and falls off exponentially for lower energies down to a lower cutoff at the Fermi energy of the positively biased electrode. This general observation has important consequences in tunneling spectroscopy and spectroscopic imaging, as discussed in Section 3.

The tunneling current depends strongly on the separation, z , between the two solids. To illustrate this fact, we simplify Eq. (1) by assuming constant densities of states, independent of energy. In the limit of low bias voltage, $V_A/\phi_{A,B} \ll 1$, the tunneling current is then given by

$$I = \rho_A \rho_B V_A \exp\left(-2\sqrt{m(\phi_A + \phi_B)/\hbar^2 z}\right) \quad (3)$$

i.e., it depends exponentially on the separation. It is from this strong z dependence of the tunneling current that STM derives its exquisite height resolution, typically of the order of 0.1 pm, or below 1/100 monolayer for most solids.

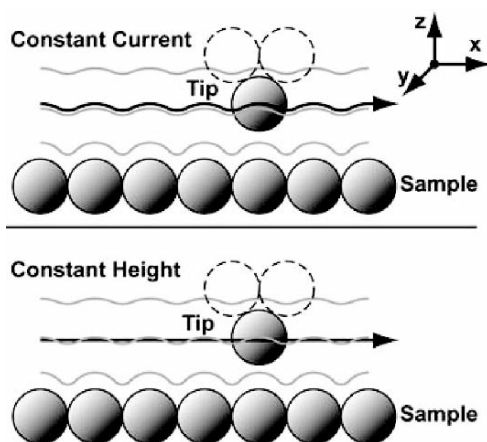


Figure 15-2. Schematic tip trajectory for STM imaging in constant current and constant height mode.

In actual STM imaging, an atomically sharp tip is used to obtain laterally confined tunneling at a well-defined position of a sample. A piezoelectric positioning element controls the in-plane (x , y) position and height (z) of the tip with picometer resolution. While a constant bias voltage is applied between sample and tip, the tip is scanned in a line-by-line fashion to continuously change the tunneling position and build up an image of a chosen area of interest on the sample. The strong dependence of the tunneling current on the tip-sample separation can be used for two basic modes of STM imaging: constant-current and constant-height imaging (Figure 15-2).

In constant-current imaging, the tunneling current is measured at each pixel of the scan and is compared with a chosen current set point. Deviations between the measured tunneling current and the set point are corrected by applying an appropriate voltage to the z -piezo, thus adjusting the tip-sample separation. This feedback mechanism maintains a constant tunneling current during the scan, while the trajectory $z(x, y)$ followed by the tip is used to generate a map of the sample surface or, more accurately, a map of a particular charge density contour above the surface.

In constant-height mode, the tip height is not modified during the scan. The tunneling current is again measured at each image pixel, but is now used directly as a representation of the sample surface via a current map, $I(x, y)$. The current maps represent a cut through charge density contours in the plane in which the tip is scanned above the sample. Obviously, since the tip-sample separation is uncontrolled during the scan, constant-height imaging is limited to samples with low corrugation and/or small scan sizes. A more practical implementation of constant-height STM uses a slow feedback system to adjust the tip-sample separation on time scales that are long compared to the residence time at each pixel, thus compensating for sample surface topography or sample tilt. As the current comparison and tip z correction are eliminated and only the tunneling current is measured, data acquisition can be faster than in constant-current imaging. The constant-height mode is thus preferred for fast STM image acquisition.

2.2 Inelastic Vacuum Tunneling

The above considerations assumed elastic tunneling between two materials, e.g., a sharp tip and the sample. In this case, the energy of a tunneling electron is conserved in the transit through the vacuum barrier. Only inside the opposite electrode the electron thermalizes to the Fermi energy by phonon emission. Tunneling electrons can, however, undergo inelastic scattering during the tunneling process. Vibrational modes of molecules adsorbed on the sample surface, for example, can be excited by inelastic scattering of a small fraction of the tunneling electrons (Figure 15–3).

The onset of inelastic tunneling at characteristic energies can then be detected and used as a fingerprint to identify individual molecular bonds. Such STM-based single molecule vibrational spectroscopy will be discussed in Section 3. In addition, controlled amounts of energy can be deposited locally into individual adsorbates, important for the manipulation of adsorbed atoms or molecules, as discussed in Section 4.

2.3 Practical Requirements: Tips, Samples, and Operating Environment

Obtaining a highly localized tunneling contact requires very sharp probe tips, ideally terminated by a single atom or a small cluster. D-band metals (e.g., W) or alloys (PtIr) are generally thought to be superior for obtaining high spatial resolution due to the strongly directional nature of the d-wave function. However, atomic resolution has also been obtained with Au tips, i.e., tip materials with primarily s-electrons at the Fermi level.

Electrochemically etched tungsten wires are often used for high-resolution STM in vacuum, while less reactive PtIr or Au tips are preferred for imaging under ambient conditions. Sharp W tips are produced by electrochemical etching in ~2M KOH or NaOH solution, using a stainless steel or Pt foil as a counter electrode. dc or ac etching at typical bias voltages of 5–10 V can be used. Preferential etching at the meniscus of the solution leads to a tapering of the immersed wire, causing one end to eventually break off. To prevent further etching that would blunt the tip apex, an electronic circuit detects the break-off

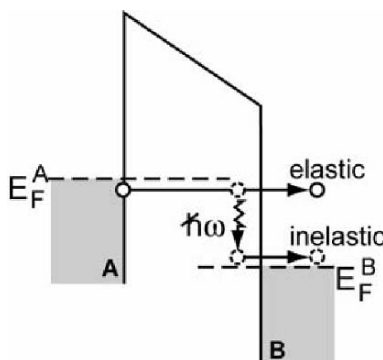


Figure 15–3. Band diagrams for elastic and inelastic electron tunneling.

current and switches the bias to zero within a few microseconds. Etched tips are rinsed in deionized water to remove traces of the etch solution, and loaded into ultrahigh vacuum (UHV). Contaminants and surface oxide are desorbed from the tip by a heat treatment in UHV prior to mounting the tip into the STM. Electron beam bombardment of the tip apex provides an efficient local cleaning of the thin part of the tip, and the applied electric field in this procedure may lead to an additional tip sharpening (Kuk and Silverman, 1986). Alternative tip cleaning and modification techniques while in tunneling contact involve voltage pulses of the order of 4–10 V while scanning the sample, or tip forming in the field emission regime (i.e., at tip–sample large separation) with applied bias voltages between 10 and 100 V and currents of several nA. Ideally, the atomic configuration of the tip would be characterized by field ion microscopy (FIM) prior to its use in the STM. Attempts have been made to combine FIM with STM. However, in practice such efforts are of limited benefit since tip changes occur frequently during scanning, rendering a complete characterization of the tip impractical.

To allow the controlled biasing of tip and sample, and to prevent charging, conventional direct current STM is limited to metal or semiconductor samples. Alternating current STM (Kochanski, 1989) has been demonstrated on insulators and organic layers, but is not widely used. Clean and ordered crystalline surfaces with low surface roughness are preferred for most studies, in particular for atomic-resolution imaging. Reactive semiconductor or metal surfaces are best prepared and imaged in UHV, but metal surfaces are also prepared and imaged routinely in electrolytes (Gewirth and Niece, 1997). In vacuum, sample preparation often involves a thermal cleaning step to remove contaminants. Flash cleaning to 1200°C for a brief period of time (few seconds) while keeping the vacuum in the low 10^{-9} torr range removes the native oxide and provides atomically clean Si samples with a low density of SiC contaminants. Metal surfaces (e.g., Cu, Al, Pt, Au), and some semiconductors (e.g., Ge) are prepared in a two-step procedure involving repeated cycles of Ar^+ ion sputtering and annealing. In refractory metals such as Ru, C contamination is removed by many (several hundred) cycles of oxygen adsorption near room temperature and flashing to 1500°C. *In situ* cleavage can be a very effective method for preparing sample surfaces that cannot be sputtered or heated to high temperatures. Samples best prepared by cleavage include III–V compound semiconductors (GaAs, InAs), as well as high- T_C superconductors ($\text{Yb}_2\text{Cu}_3\text{O}_{7-x}$, $\text{BiSr}_3\text{Cu}_2\text{O}_{8+x}$). *In situ* cleavage is also the preparation method of choice for cross-sectional STM (see Section 5), commonly performed on (110) cleavage planes of III–V semiconductors. Bulk insulating metal oxides, most prominent among them TiO_2 , can be imaged successfully by STM following Ar^+ ion sputtering and annealing, which generates oxygen vacancies and renders the surface region sufficiently conductive for stable STM imaging. Au and highly oriented pyrolytic graphite (HOPG) are among the few materials on which atomic resolution is obtained under ambient conditions. Well-ordered, (111) oriented surfaces of Au single crystals and thin films on mica can be prepared

by flame annealing in air (Robinson et al., 1992). Preparation of layered materials, notably HOPG, is particularly simple. It involves the removal of bundles of graphene layers by adhesive tape from the crystal. Since the exposed graphene sheet has no dangling bonds and is very inert, such samples can be imaged with atomic resolution in air for extended time periods.

For many materials, in particular metals and alloys, electrochemical STM is a powerful alternative to imaging in UHV. In this environment, almost the entire portion of the tip that is immersed in the solution must be coated by an insulating layer (e.g., wax or glass) to minimize the faradaic (i.e., ionic) current, which is typically much larger than the tunneling current and would otherwise dominate the measured signal. Only a small fraction near the apex of the tip remains uncoated to allow for tunneling between the tip and sample.

2.4 STM Instrumentation

In tunneling contact the probe tip is only a distance of the order of 1 nm away from the sample. As the tunneling current is exponentially sensitive on the tip-sample separation, any mechanical vibrations of the tip relative to the sample have to be minimized. To this end, a two-fold strategy is commonly pursued. The microscope head itself is built as compact and stiff as possible by closely integrating sample and tip on a rigid platform. This results in a high resonance frequency for oscillations of the tip relative to the sample. In addition, the entire microscope is isolated mechanically from outside vibration sources, e.g., by suspending it on soft extension springs or long bungy cords. The overall goal is to achieve a maximal mismatch between the mechanical modes of a soft suspension system and the high resonance frequency of the stiff microscope head. In addition, efficient vibration damping is required, which can be achieved by magnetically induced eddy currents. The springs used in many suspension systems have their own mechanical modes that can be damped by polymer strips woven into their coils. Direct coupling of acoustic noise into the microscope, finally, is a problem that is best solved by operating the STM in vacuum (Figure 15-4).

A few additional components are part of any practical tunneling microscope. A coarse approach mechanism has to be in place, which moves the tip from an initial distance of several millimeters into tunneling range without making physical contact with the sample. Although this has been achieved by mechanical approach mechanisms (Smith and Binnig, 1986), piezoelectric driven stick-slip motors are now preferred since they allow the entire coarse approach to be performed automatically under computer control. For the positioning and scanning of the tip during STM imaging, a segmented piezoelectric tube scanner is used (Binnig and Smith, 1986). By applying high voltages to the individual segments, the scanner is independently actuated in the (x , y) plane parallel to the sample surface and along the perpendicular z axis with little crosstalk. The tunneling current in the pA to nA range is converted into a voltage by a sensitive, low-noise amplifier

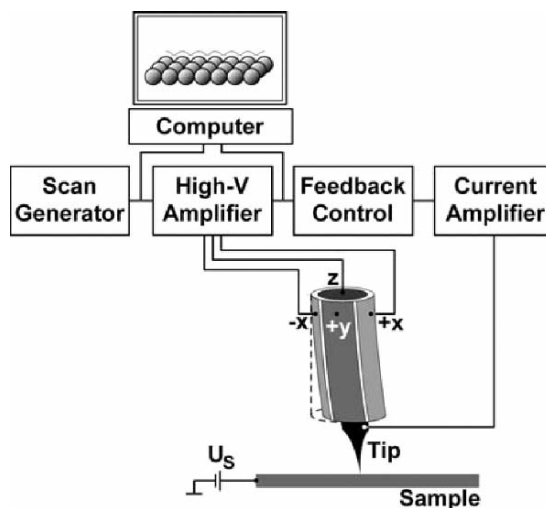


Figure 15–4. Schematic illustration of the various components of an STM system, including piezoelectric tube scanner for tip positioning, and control electronics (current amplifier, feedback, scan generator, high-voltage amplifier).

with adjustable transconductance of 10^7 – 10^{10} V/A, which provides the input signal for the tunneling current feedback electronics. Modern STM controllers typically use digital feedback and scan modules implemented by specific software programs running on fast digital signal processor (DSP) chips. The measured analog tunneling current signal is digitized using an analog-to-digital (A/D) converter, and is compared with the current set point by the feedback program on the DSP. Necessary corrections in the tip height z are computed, and are converted into analog voltages by digital-to-analog (D/A) converters. These voltages may be amplified further by a high-voltage amplifier stage to levels suitable to drive the tube scanner. The scan signal, i.e., independent voltage ramps driving the piezo scanner along the in-plane x and y directions, is also generated digitally on a DSP, followed by conversion and amplification steps analogous to those of the z signal. A program running on a host computer and communicating with the DSP, finally, acquires, displays, and stores the measured tip trajectory $z(x, y)$ (in constant-current imaging) or position-dependent tunneling current $I(x, y)$ (in constant-height mode).

2.5 STM Imaging: Application Examples

High-resolution imaging of surface structure is probably the most widespread application of STM. As a direct imaging technique, STM has added significantly to the diffraction techniques used traditionally to determine the structure of clean and adsorbate covered surfaces. Examples of typical applications include the determination of the structure of reconstructed surfaces of clean semiconductors, for instance vicinal Si surfaces (Erwin et al., 1996) and oxides, such as TiO_2 (Diebold, 2003), and of adsorbate-induced reconstructions on metals

(e.g., O/Ru(0001); Meinel et al., 1997); the mapping of the evolution of surface morphology during epitaxial growth (e.g., metal epitaxy; Chambliss et al, 1995), etching (Boland and Weaver, 1998), and energetic particle bombardment (e.g., electrons on Si; Nakayama and Weaver, 1999); and the imaging and structural identification of non-periodic structures such as surface defects, steps, as well as surface-supported solid (e.g., silicide nanowires on Si; Chen et al., 2000) and molecular nanostructures [e.g., molecular rotors on Cu(100); Gimzewski et al., 1998]. Although electrically conducting samples are a prerequisite for stable tunneling, high-resolution imaging is feasible on ultrathin insulating films (for a recent review, see Schintke and Schneider, 2004) and self-assembled organic monolayers [e.g., alkanethiol monolayers on Au(111); Cygan et al., 1998) supported by metal substrates.

In view of the large number of materials systems to which STM has been applied, a comprehensive survey would be beyond the scope of this chapter. Instead, we discuss in some detail recent contributions of STM imaging to one specific material: silicon.

2.5.1 STM Imaging—Silicon Surfaces

Due to the important role of Si in electronics, extensive STM studies on the structure of clean Si surfaces with different orientation have contributed to making Si one of the best-characterized materials system. Important early milestones in the characterization of clean Si surfaces include the determination of surface bonding and reconstructions on Si(111) (Binnig et al., 1983) and on Si(001) (Tromp et al., 1985) used as a substrate in microelectronics. STM imaging served to establish the thermodynamics of terraces and steps (Swartzentruber et al., 1990; Men et al., 1988) on these surfaces, to study electromigration and step-bunching (Yang et al., 1996), and to survey the stable facets involved in the equilibrium shape of Si crystals (Gai et al., 1998).

Figure 15–5 shows an example of constant current STM images on Si(111), a complex reconstructed surface whose geometric (Binnig and

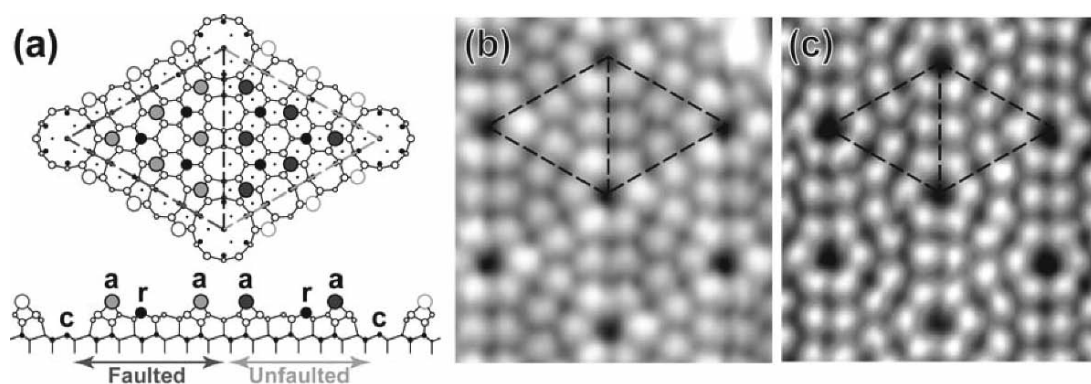


Figure 15–5. (a) Ball-and-stick model of the Si(111)–(7 × 7) surface. Top: top view; bottom: side view. a, adatom; r, rest atom; c, corner hole. (b) Filled-state constant-current STM on Si(111)–(7 × 7); $V = -1.0$ V; $I = 0.4$ nA. (c) Empty-state constant-current STM; $V = +1.2$ V; $I = 0.4$ nA.

Rohrer, 1983) and electronic (Hamers et al., 1986) structure was solved in pioneering STM experiments, and which continues to play an important role in STM technique development. Figure 15–5a shows a structure model of the (7×7) reconstruction (Takayanagi et al., 1985). Figure 15–5b, obtained at negative sample bias ($V = -1.0$ V), i.e., with electrons tunneling from occupied sample states to unoccupied tip states, shows the corrugation associated with filled states of the sample. Conversely, Figure 15–5c, obtained at positive sample bias ($V = +1.2$ V), maps the empty states of the sample. Both images show atomic resolution, clearly resolving the twelve adatoms (“a”) per unit cell. In addition characteristic deep “corner holes” (“c”) bounding the diagonals of the $4.6 \text{ nm} \times 2.9 \text{ nm}$ rhombohedral unit cell are imaged.

While all adatoms are mapped uniformly in the empty state image, the filled state scan shows one-half of the unit cell somewhat higher than the other, an effect on the charge density due to the different stacking sequence of atomic layers in the two halves of the unit cell. A comparison of the images obtained at opposite bias polarity suggests that in addition to surface topography, the electronic structure of the sample surface adds substantially to the contrast observed in STM imaging. This is implicit also in Eq. (1) via the dependence of the tunneling current on the local densities of states of both tip and sample at the tunneling contact. Note also that the rest atoms (“r”), a second near-surface species with dangling bonds protruding into the vacuum, are imaged neither at positive nor at negative sample bias. A detailed discussion of bias-dependent imaging and other tunneling spectroscopy methods used to assess the local electronic structure of a sample surface is given in Section 3.

Apart from the structure of clean Si surfaces, Si-based surface chemistry has been studied widely by STM. Initial studies provided a knowledge base for technological processes, such as reactive ion etching, doping, and chemical vapor deposition (CVD). STM was used to probe interactions of Si with halogens (Boland, 1993) and with small molecules involved in surface passivation (H_2 , H; Laracuenta and Whitman, 2001), doping (PH_3 ; Wang et al., 1994a), and CVD growth (Si_2H_6 ; Wang et al., 1994b). More recent research directions include the integration of molecular electronic elements with Si. Under this perspective, the covalent bonding of a wide variety of organic molecules on Si has been studied (for a review, see Wolkow, 1999), including small molecules such as ethylene (Mayne et al., 1993) and acetylene (Li et al., 1997), simple alkenes (e.g., propylene; Lopinski et al., 1998) and polyenes (e.g., 1,3-cyclohexadiene; Hovis and Hamers, 1997), pentacene (Kasaya et al., 1998), and benzene (Borovski et al., 1998). On clean Si(001), adsorption is from the gas phase in UHV. H-passivated Si can also be modified by *ex situ* wet-chemical techniques, and reintroduced into UHV for STM imaging.

Traditionally, an important aspect of the surface science of semiconductors has been epitaxial growth. STM observations at initial growth stages, i.e., at coverages of fractions of a monolayer (ML) up to several ML, can provide direct insight into fundamental processes such as adatom diffusion, incorporation into steps, and nucleation of mono-

layer islands (for a review, see Zhang and Lagally, 1998). Observations of growth and equilibrium structures allow identifying kinetic and thermodynamic factors affecting the growth process, as well as the role of defects, of elastic strain, etc.

STM studies of Si epitaxy are commonly performed in either of two modes. Conceptually preferred is the dynamic observation, via time-lapse STM “movies,” of the growth surface with the sample held in the microscope at high temperature and in the presence of a deposition flux. This difficult technique will be discussed in more detail in Section 4. As an alternative, STM imaging can be performed at room temperature on samples that have been quenched at key stages in the growth process. If measures are taken to verify that the quenched-in morphology is indeed representative of that at high temperature during growth, this “quench-and-look” technique offers the advantage of higher spatial resolution and larger scan sizes over high-temperature STM.

Early STM experiments focused on the initial stages of Si homoepitaxy and Ge/Si heteroepitaxy, mostly on the technologically important (111) or (001) surfaces. Key results include the identification of regimes of step-flow growth and nucleation, determination of the activation energy and atomistic pathway of surface diffusion (Mo et al., 1991), explanation of growth and equilibrium shapes of two-dimensional (2D) islands (Mo et al., 1989), and exploration of the modification of the growth process by surfactants (Horn-von Hoegen, 1994). Figure 15–6 illustrates the identification of the initial stage of island formation in

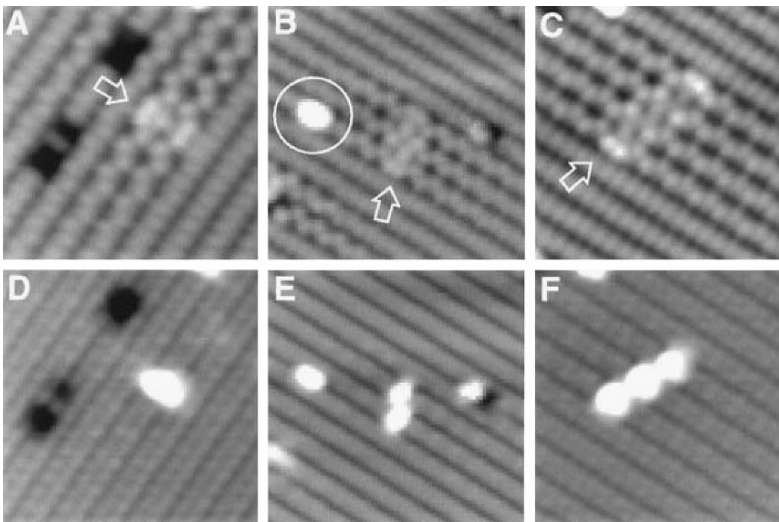


Figure 15–6. STM images of 0.01 monolayers Ge on Si(001). The long diagonal bands are substrate dimer rows. Filled-state images (A–C; $V = -2$ V; $I = 0.2$ nA) and corresponding empty-state images (D–F; $V = +2$ V; $I = 0.2$ nA), showing rows of symmetric (bean-shaped) and asymmetric (buckled) substrate dimers, and Ge adatom-induced chain-like structures (marked by arrows). The chain-like paired adatom structures are seen as metastable precursors to the formation of monolayer islands in Ge/Si homoepitaxy (Reprinted with permission from Qin and Lagally, © 1997 AAAS).

Ge/Si(001) epitaxy (Qin and Lagally, 1997). Classic theories of nucleation in thin film growth are based on the notion of a “critical nucleus,” defined as the structural entity for which the addition of one more atom will for the first time reduce the free energy. Employing very low Ge coverages and high-resolution dual bias STM imaging to uncover the link between monomer adsorption and initial 2D growth islands, this study shows that island formation may not involve a “critical nucleus,” but instead may proceed via a family of metastable structures of varying size, with significant consequences on many growth models that are based on nucleation and require the size of a critical nucleus as input.

More recently, interest has concentrated on strained layer heteroepitaxial growth, e.g., of $\text{Si}_{1-x}\text{Ge}_x$ alloys on Si(001), as an elegant way of producing large arrays of nanostructures. Si and Ge are miscible over the entire composition range, and the lattice mismatch in the $\text{Si}_{1-x}\text{Ge}_x$ /Si(001) system can be tuned between 0 and 4% by alloying. Over a wide range of compositions, $\text{Si}_{1-x}\text{Ge}_x$ alloys initially wet the Si substrate, and at higher coverage develop coherent (i.e., dislocation free) faceted three-dimensional (3D) islands that lower the free energy by relaxing part of the lattice mismatch strain. Potentially useful in electronics or optoelectronics, these faceted nanostructures have shown a strikingly complex array of growth phenomena, which make them interesting for fundamental growth studies.

Constant current STM images with atomic resolution typically encompass fields of view of few tens of nanometers. Combining a drift-stable microscope with state-of-the-art control electronics, substantially larger atom-resolved images have become feasible. The ability to obtain image sizes in excess of $1\mu\text{m}^2$ with high resolution is potentially powerful for imaging growth processes, since it provides access to all relevant length scales as well as image statistics far superior to that of conventional small scans. Figure 15–7 illustrates this capability with an STM image of 1.5 monolayers Ge on Si(111) with a field of view of $0.75\mu\text{m}$ and 0.05nm pixel size. The large scan provides an excellent overview of the step and island structures resulting from monolayer Ge deposition. At progressive zoom into the image, individual terraces, and, ultimately, single surface defects and reconstruction domains (here coexisting 7×7 and 5×5 domains) can be examined.

Figure 15–8 gives an example in which the superior statistics resulting from large atom-resolved images was used successfully to pinpoint the mechanism of periodic surface roughening in Ge/Si(001) growth (Sutter et al., 2003a). Via a sequence of “quench-and-look” experiments at different Ge coverages, short-range interactions between surface steps and vacancy lines, periodic arrays of linear chains of dimer vacancies, are shown to cause a highly correlated surface roughness in the form of anisotropic 2D islands with progressively higher aspect ratio (a–c). Finally, images obtained at a critical Ge coverage of four atomic layers (d) show the transition from 2D to 3D growth by formation of the first faceted islands on the rough wetting layer, and identify the atomic-scale pathway of the 2D–3D transition.

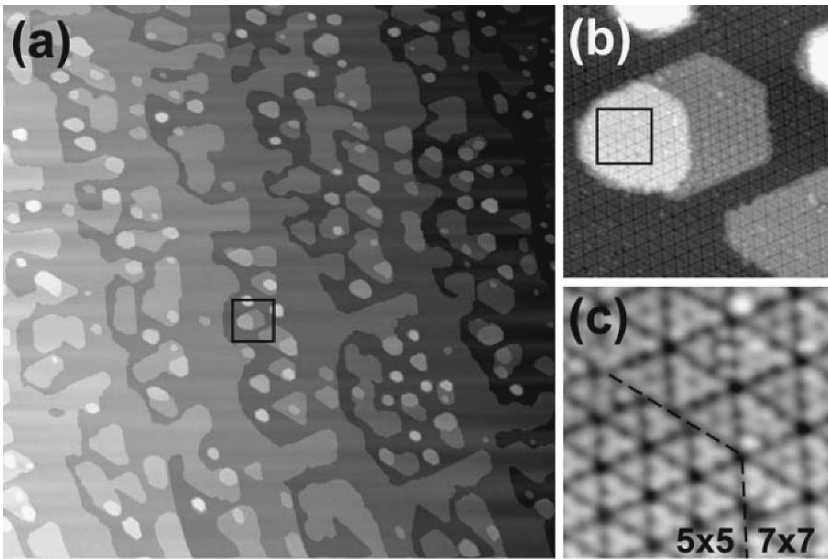


Figure 15-7. Large-area atom-resolved STM. (a) STM image with $0.75\text{-}\mu\text{m}$ field of view and 0.05-nm pixel size, showing the surface morphology of Si(111) after deposition of 1.5 ML Ge at 550°C . An array of parallel steps, running diagonally from the upper left to lower right, coexists with islands nucleated during Ge deposition. (b) Zoomed-in view (50 nm) of the area marked by a square in (a), showing one of the islands. (c) Further zoom-in (10 nm), showing the coexistence of (7×7) and (5×5) reconstructed domains in the area marked in (b).

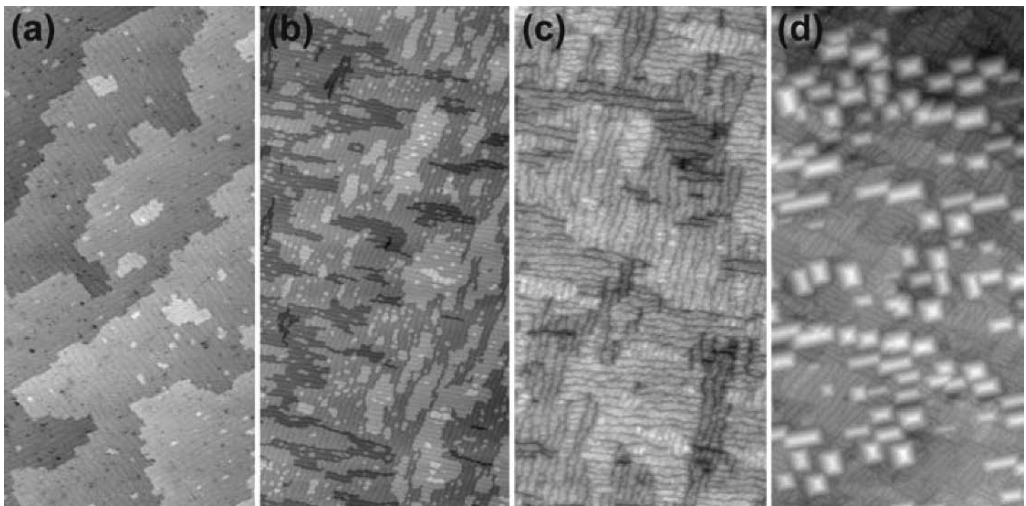


Figure 15-8. Transition from 2D to 3D morphology in the growth of Ge on Si(001). Large-area atom-resolved STM images ($250\text{ nm} \times 125\text{ nm}$ sections are shown here) illustrate the surface morphology at (a) 1.5 ML , (b) 2.3 ML , (c) 3.5 ML , and (d) 4.0 ML Ge coverage. The transition to 3D growth is preceded by the formation of highly correlated, anisotropic surface roughness. With the superior statistics of large-area high-resolution images, the repulsive interaction between surface steps and defects (dimer vacancy lines) was identified as the origin of the ordering of this surface roughness. (Reprinted with permission from Sutter et al., © 2003a by the American Physical Society.)

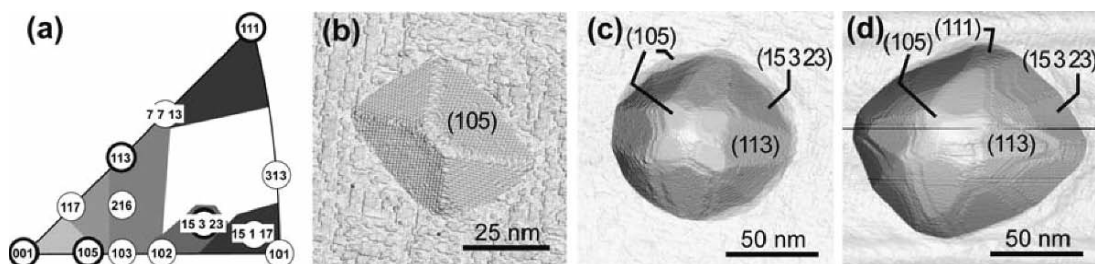


Figure 15-9. Shapes of coherent (i.e., dislocation-free) 3D islands induced by lattice mismatch strain during $\text{Si}_{1-x}\text{Ge}_x/\text{Si}(001)$ heteroepitaxy. (a) Unit stereographic triangle for Si, showing the major and minor stable facets (Gai et al., 1998). (b) STM top view, showing a shallow (105) faceted “hut.” (c) Multifaceted “dome” shape, terminated by (105), (113), and (15 3 23) facets. (d) “Barn”-shaped 3D island, observed for low-misfit $\text{Si}_{1-x}\text{Ge}_x$ alloys on Si(001), transformed from a “dome” by introduction of additional (111) facets (Reprinted with permission from Sutter et al., © 2004 American Institute of Physics.).

Initial 3D islands with shallow facets, such as those shown in Figure 15-8d, evolve into more complex shapes involving steeper facets, which give rise to more efficient strain relaxation (Figure 15-9). For $\text{Si}_{1-x}\text{Ge}_x$ epitaxy on Si(001), the facets bounding the islands identified by STM are to a large extent major stable facets of Si and Ge (Figure 15-9a). At early stages of 3D growth, shallow (105) faceted “huts” are invariably observed (Figure 15-9b; Mo et al., 1990). These islands then undergo a shape transformation to become multifaceted “domes” (Figure 15-9c; Medeiros-Ribeiro et al., 1998). At later stages in the shape evolution, elastic strain relaxation increasingly competes with plastic relaxation via dislocations. For Ge/Si(001) dislocations are introduced before the “dome” shape can transform further by introducing additional steeper facets. By reducing the Ge concentration of $\text{Si}_{1-x}\text{Ge}_x$ alloys, and thus lowering the lattice mismatch strain, the coherent shape evolution has been extended beyond the “dome” shape to the end-point in the stereographic triangle, where steep (111) facets are introduced (Sutter et al., 2004).

3 Tunneling Spectroscopy

Equation (1) gives an expression for the tunneling current I in the WKB approximation for planar tunneling. In this approximation the tunneling current is an integral, over the energy range of width $|eV|$ in which tunneling can occur, of a product of the densities of states (DOS) of tip $[\rho_T(E)]$ and sample $[\rho_S(E)]$ multiplied by the transmission probability, $T(E, eV)$. While this simple approximation cannot address the high spatial resolution of STM—a more elaborate theory capable of addressing this aspect will be discussed in Section 6—it demonstrates that the tunneling current is affected by the electronic structure of both the tip and the sample. In a large number of applications it would be desirable to merely map the atomic scale “topography” of a sample. The sensitivity of the tunneling current to the DOS can then be a disadvantage, since it often prevents a simple interpretation of a tunneling image as

a “topographic” map. For many metals with small variations in the DOS near the Fermi energy such a simple interpretation is often possible, i.e., height maxima in a constant current STM image are correlated with atomic positions. However, for semiconductors with strongly varying DOS near E_F , the simple picture frequently breaks down.

An example is the (001) surface of Si. Geometrically, this surface consists of Si dimers, in which each atom is bonded to two atoms in the second layer and to its dimer partner. Constant-current STM images obtained on Si(001) at positive sample bias indeed show two maxima in each dimer (Figure 15–10B), which could be interpreted as the “atomic positions.” At negative sample bias, however, only a single broader bean-shaped structure is observed (Figure 15–10A), which extends across each dimer. In fact, the tunneling current in empty-state STM (positive sample bias) reflects the charge distribution in a π^* anti-bonding state just above the Fermi energy, whose wave function has a node at the center of the dimer. Even in this case in which the STM image shows two maxima per dimer, these maxima do not reflect the atomic positions in the dimer. Worse yet, in filled-state STM (negative sample bias) a π bonding state is imaged whose charge density peaks at the center of the dimer bond, thus giving rise to a single “topographic” maximum at that position.

The sensitivity to the electronic structure of tip and sample, however, is not merely a complication in STM imaging. It is also one of the major strengths of the technique, since it allows measuring and mapping surface electronic structure with very high spatial resolution well below 1 nm. While most other techniques, such as ultraviolet or x-ray photoelectron spectroscopy or Auger electron spectroscopy, average over macroscopic sample areas, recently developed spectromicroscopy techniques, such as synchrotron photoelectron emission microscopy, provide spectroscopic imaging with 10 nm spatial resolution (see Chapter 8, this volume). However, measurements of surface electronic

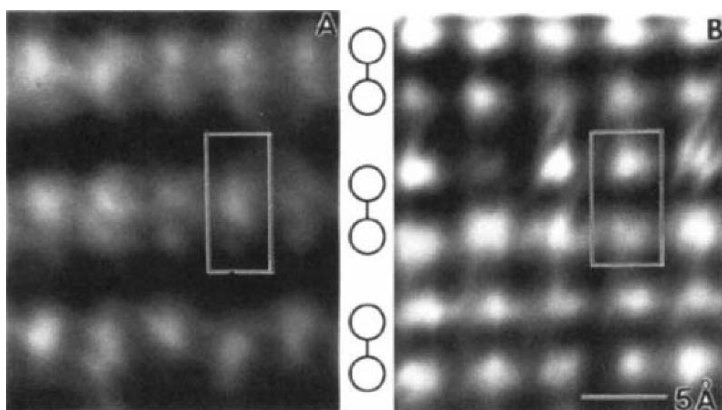


Figure 15–10. Constant-current STM images of Si(001) at negative [$V = -1.6$ V (A)] and positive [$V = +1.6$ V, (B)]. The (2×1) unit cell is outlined, and the locations of the dimers are shown in the center. (Reprinted with permission from the Annual Review of Physical Chemistry, Volume 40 © 1989 by Annual Reviews www.annualreviews.org.)

structure “atom-by-atom,” summarized under the term scanning tunneling spectroscopy, are feasible only by STM.

3.1 Bias-Dependent Imaging

The simplest scenario for gathering electronic structure information involves the successive measurement of constant-current STM images at different sample bias. As discussed in Section 1, the tunneling transmission probability is generally highest at the Fermi energy of the negatively biased electrode, and falls off exponentially at lower energies. For positive sample bias, i.e., empty-state STM, elastic tunneling occurs mainly from states near the tip Fermi energy into unoccupied sample states. In this polarity, varying the bias voltage probes different surface states of the sample, particularly if the electronic structure of the tip varies slowly in the energy range defined by the bias voltages used. A series of constant-current STM images at different bias can thus be used to map the spatial distribution of unoccupied surface states of the sample.

For negative voltages applied to the sample, i.e., filled-state STM, the situation is not as fortunate. The tunneling current is now dominated by states near the Fermi level of the sample. As a result, the occupied sample state with highest energy is invariably imaged preferentially (Becker et al., 1989). Even if lower-lying surface states exist, as indicated schematically in Figure 15–11, constant-current STM imaging is rather insensitive to those states.

As a possible solution to this problem in bias-dependent filled-state imaging it has been suggested that metal probe tips with smoothly varying density of states near the Fermi energy could be replaced by a semiconductor tip with a strongly modulated DOS. Band structure effects have long been recognized in the interaction of low-energy electrons with metal surfaces (Bauer, 1994). As an example, the high normal incidence reflectivity on W(110) at low electron energies is due to a (110) projected band gap extending over 5 eV, i.e., a lack of extended bulk states in the energy range of this gap. For W(100), such a gap does

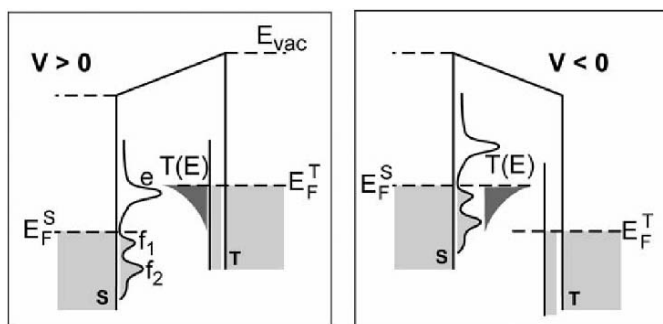


Figure 15–11. Band diagrams illustrating the bias-dependent tunneling between tip (T) and sample (S). Sample bias $V > 0$: empty-state STM. $V < 0$: filled-state STM. $T(E)$ shows schematically the transmission probability as a function of energy for states participating in the tunneling process.

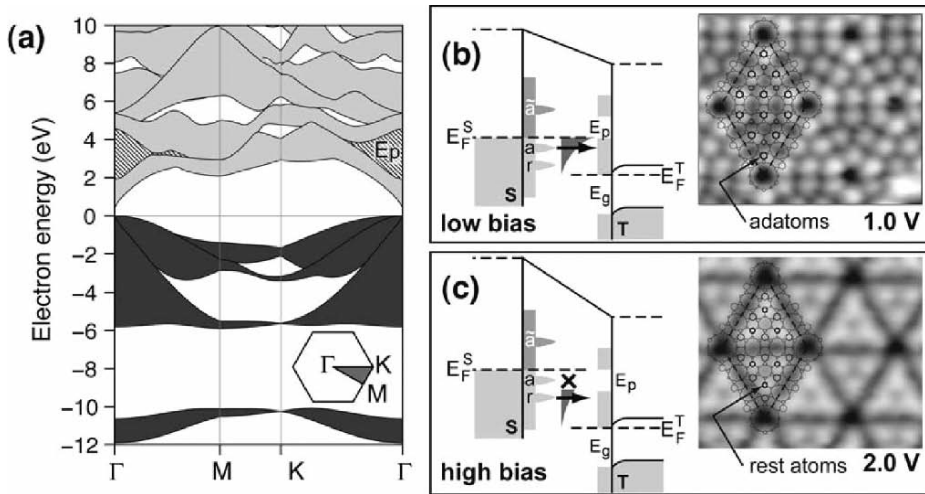


Figure 15-12. Energy-filtered STM. (a) Bulk band structure of monocrystalline InAs probe tips, projected along the (111) zone axis parallel to the tunneling direction. E_p denotes a projected gap in the bulk conduction band. (b) Band diagram and STM image for low-bias filled state STM with InAs tips, demonstrating preferential imaging of the adatom state on Si(111)-(7 × 7). (c) High-bias filled state InAs-tip STM on Si(111)-(7 × 7): alignment of the adatom state with the projected tip gap reduces its contribution to the tunneling current, giving rise to preferential imaging of the energetically lower rest-atom state (Reprinted with permission from Sutter et al., © 2003b by the American Physical Society).

not exist. Low-energy electrons can propagate into extended states, and the reflectivity in the same energy range is low. Similarly, it can be expected that a lack of final states due to projected band gaps in the STM tip material can significantly affect the tunneling probability in the energy range spanned by such gaps. This modulation of the transmission probability could be used for efficient energy filtering of the tunneling current in constant-current STM, specifically in filled-state imaging.

Calculations by density functional theory (DFT) show a number of gaps in the [111]-projected band structure of III-V compound semiconductors such as InAs. III-V compound semiconductors cleave easily at (110) planes, and cleavage corners bounded by {110} and (100) planes are sufficiently sharp for atomic-resolution STM (Nunes and Amer, 1993). In addition, InAs(110) has the advantage of an unpinned surface with minimal band bending and slight electron accumulation. The feasibility of energy-filtered STM has been demonstrated by using a cleaved InAs tip to selectively image different occupied surface states on Si(111)-(7 × 7) (Sutter et al., 2003b). In these experiments, summarized in Figure 15-12, a [111]-projected gap in the InAs conduction band, centered at Γ and about 2.6 eV wide, is used to modulate the tunneling current in filled-state constant current STM. The “adatom” state (“a”) on Si(111)-(7 × 7), which lies only 0.35 eV below the Fermi energy and is imaged by STM using a metal tip, dominates the tunneling current at low tip-sample bias. At higher bias, this state aligns with a projected gap in the InAs conduction bands and its tunneling probability is reduced sharply. Under these conditions the “rest atom”

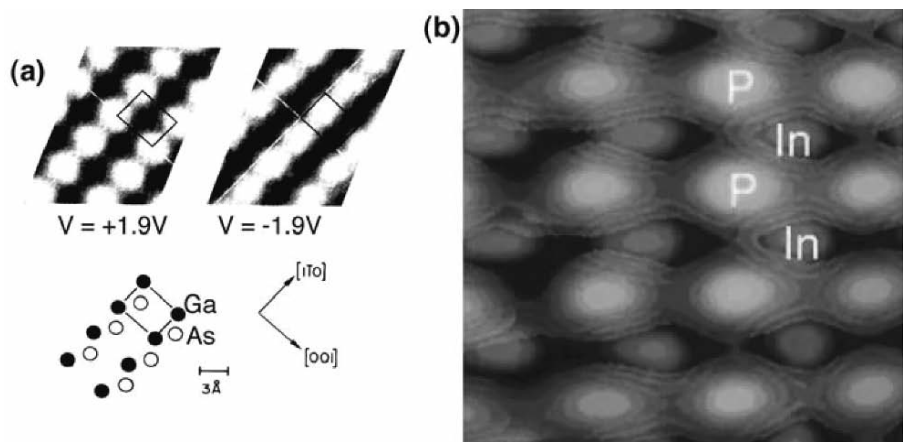


Figure 15-13. (a) Bias-dependent STM on GaAs(110): selective imaging of Ga and As sublattices at positive and negative sample bias, respectively (Reprinted with Permission from Feenstra et al., ©1987 by the American Physical Society). (b) Compound STM image of the InP(110) surface, assembled from separate positive and negative bias scans (Reprinted from Ebert et al., ©1992 with permission from Elsevier). (See color plate.)

state (“r”), a surface state 0.8eV below E_F that is not accessible in conventional STM imaging, contributes a majority of the tunneling current and can be imaged selectively. Importantly, these results suggest that robust bulk band structure effects, which are independent of the atomic structure of the tip, can be used to modulate the transmission probability in vacuum tunneling. Once developed into a routine imaging technique, energy-filtered STM has the potential to provide rich spectroscopic contrast within the relatively simple framework of bias-dependent constant current STM.

A classic example of a scenario in which conventional bias-dependent imaging provides directly interpretable spectroscopic information is imaging of cation and anion sites on (110) cleavage planes of compound semiconductors such as GaAs (Feenstra et al., 1987) or InP (Ebert et al., 1992) (Figure 15-13). For (110) surfaces of compound semiconductors, such as GaAs, the occupied state density (imaged at negative sample voltage) is localized on the anions (As, P), while the unoccupied state density (positive sample voltage) is localized on the cations (Ga, In). As a result, bias-dependent STM imaging can be used to selectively image the cation and anion sublattices on these surfaces.

3.2 Local I-V Measurements and Current Imaging Tunneling Spectroscopy

In many instances, a complete electronic structure map is not required, but one would like to determine, at one or several positions on a sample, the surface density of states in an interval of a few eV around the Fermi energy. Such local measurements can, for example, serve to determine if a specific surface reconstruction or a small cluster is semiconducting or metallic, to measure the band gap at the surface of semiconductor samples, and to determine for a given location the energies of surface states.

Local current–voltage [$I(V)$] spectra serve this purpose. The basic information content of a tunneling $I(V)$ spectrum can be demonstrated by differentiating Eq. (1) [$V > 0$]:

$$\frac{dI}{dV} = e\rho_s(eV)\rho_T(0)T(eV, eV) + \int_0^{eV} [e\rho_s(E)\rho_T(E - eV)\partial T(E, eV) / d(eV) - e\rho_s(E)\rho_T'(E - eV)\partial T(E, eV)] dE \quad (4)$$

For a given energy eV , the tunneling conductance dI/dV reflects the density of states of the sample at that energy, $\rho_s(eV)$.

Feenstra et al. (1987) noted difficulties with conductance spectroscopy, resulting from the fact that the expression for dI/dV diverges exponentially both in voltage (V) and separation (z). These divergences can be eliminated by normalizing dI/dV by the conductance of the tunneling junction, $[dI/dV]/[I/V] \approx d(\ln I)/d(\ln V)$. In this form, spectra obtained at different tip–sample separation z can be compared directly.

Local spectroscopy is quite simple to perform. The STM tip is positioned and stabilized over a chosen point on the sample. The feedback loop is switched off while the sample voltage is varied over the desired values, and a local measurement of tunneling current as a function of bias voltage, $I(V)$, is performed. As an alternative, the tunneling conductance, $dI(V)/dV$, can be measured by ramping the bias voltage with a small ac voltage added, and measuring the ac component of the tunneling current at the modulation frequency of the ac part by lock-in techniques. After the $I(V)$ or $dI(V)/dV$ spectrum is recorded, the tip feedback is reactivated, and STM imaging can resume.

Figure 15–14a shows tunneling conductance spectra obtained on p- and n-type GaAs(110) (Feenstra et al., 1987), which provide a measure

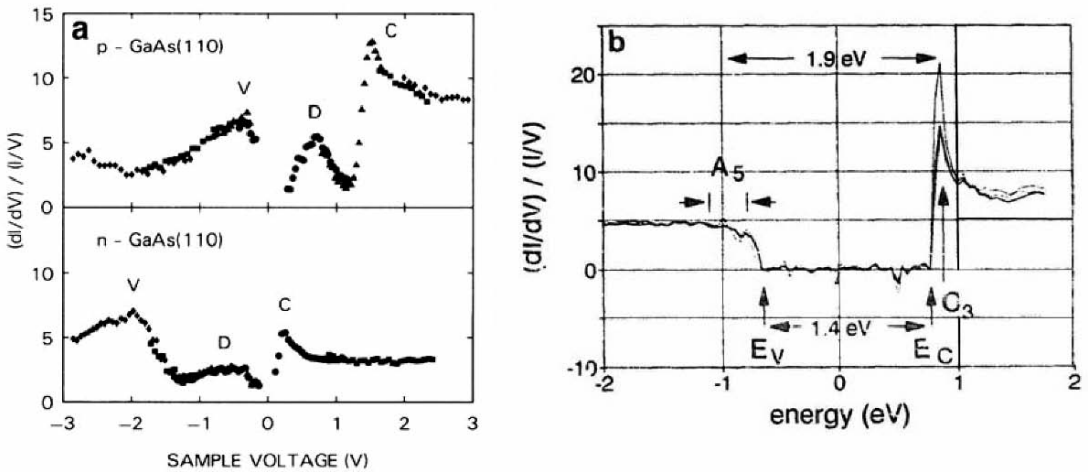


Figure 15–14. (a) Local normalized conductance spectra obtained on n- and p-doped GaAs(110), showing three peaks (V, D, C) originating from surface states (Feenstra et al., 1987). (b) Normalized conductance spectra obtained on InP(110), illustrating the valence and conduction band edges delimiting the bulk bandgap (1.4 eV), and the A_5 and C_3 surface states defining the surface gap (1.9 eV) (Reprinted from Ebert et al., © 1992 with permission from Elsevier).

of the surface state density. The spectra show three distinct peaks associated with surface states. Peak C is related to a state in the GaAs conduction band, localized on the Ga atoms, while peak V stems from a valence state localized on the As anions. Peak D is associated with tunneling of dopant-induced carriers within the bulk band gap. The separation between the leading edges of the C and V peaks is close to the bulk band gap of GaAs ($E_g = 1.43\text{ eV}$), i.e., in this system tunneling spectroscopy can be used to determine the band gap. Figure 15–14b) shows tunneling conductance spectra on the (110) cleavage surface of another III–V compound, InP. Also for InP, the band gap ($\sim 1.4\text{ eV}$) is evident from sharp onsets of significant state density at the valence and conduction band edges. However, for this system it was argued that a wider surface band gap ($\sim 1.9\text{ eV}$) is indirect, delimited by a surface state C_3 at the edge of the surface Brillouin zone and a broadened state $A5$ at the zone center.

A natural, albeit experimentally much more complex extension of local $I(V)$ spectroscopy is the measurement of tunneling spectra, as discussed above, at each image pixel of a constant-current STM scan. This measuring scheme is called current-imaging tunneling spectroscopy (CITS). Since complete $I(V)$ spectra are obtained at each image point, the corresponding data sets can provide a full range of spectroscopic information. As an example, current maps $I(x, y)$ at fixed tip–sample bias can be produced. Instead, the voltage dependent tunneling conductance $dI(V)/dV$ can be calculated numerically and mapped as a function of sample position. If the chosen voltage corresponds to the energy of a surface state of the sample, conductance maps will provide a direct image of the spatial distribution of that state. Due to the complexity of the data acquisition, the experimental requirements for CITS are quite stringent. Since complete $I(V)$ curves are measured at each image pixel, requiring the stopping of a scan and deactivation of the feedback loop for a fraction of a second per spectrum, a very high stability of the tunneling gap and low lateral drift of the tip relative to the sample are of key importance. These conditions are more easily fulfilled at cryogenic temperatures, where a z -stability of the order of 1 pm and lateral drift velocities of the order of few $\text{\AA}/\text{hour}$ are possible. Low temperatures also reduce the thermal broadening of the tunneling transmission coefficient $T(E)$, and thus narrow the linewidth of features in the tunneling spectra.

A classic example of the application of CITS, the measurement of the electronic structure of Si(111)-(7 $\times$ 7), is shown in Figure 15–15 (Hamers et al., 1986). Shown are a constant current image of the surface at $+2\text{ V}$ sample bias (a), as well as current images acquired during the same scan at bias voltages of $+1.45\text{ V}$ and -1.45 V , showing spatial maps of unoccupied (+)/occupied (–) sample states at energies up to 1.45 eV above/below the Fermi energy. CITS difference images, a precursor to modern dI/dV maps, discussed below, calculated by numerically subtracting current images at energies bracketing those of specific surface states, allowed the mapping of the adatom state, dangling bond, and backbond state on this complex reconstructed surface.

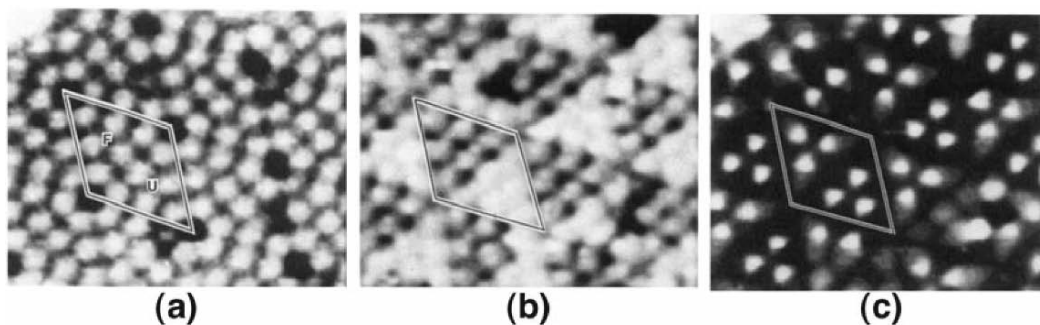


Figure 15–15. Current-imaging tunneling spectroscopy on Si(111)-(7 × 7) (Hames et al., 1986). (a) Constant-current STM image ($V = +2$ V), with one (7 × 7) unit cell outlined (F, faulted; U, unfaulted half). (b and c) Current images obtained at +1.45 V and –1.45 V, respectively. Note the strong similarity between the rest-atom contrast in the current image (c), and the contrast obtained in rest-atom imaging by energy-filtered constant-current STM (Figure 15–12).

More recently, this complex data acquisition mode has been used, for example, to map modifications in the electronic structure of individual carbon nanotubes arising with the supramolecular assembly of C_{60} molecules inside the hollow tube to form nanotube “peapods” (Hornbaker et al., 2002). dI/dV spectra obtained along the symmetry axis of C nanotubes show distinct peaks in the density of unoccupied states with a spatial periodicity corresponding to the average spacing of embedded C_{60} molecules observed by transmission electron microscopy, suggesting that the insertion of C_{60} causes a significant modulation of the electronic structure. A control measurement was constructed by using the STM tip to push the C_{60} away, and remeasuring dI/dV spectra along the same nanotube section without embedded C_{60} . The empty tube obtained in this way shows small and smooth variations in the density of states along the nanotube axis. Qualitatively, the electronic structure of the peapods was explained by the single wall C-nanotube acting as a conduit that enhances the coupling between C_{60} molecules nested inside it, and the Bragg scattering of nanotube states due to the periodic arrangement of the C_{60} . Calculations of the peapod electronic structure confirm the formation of a hybrid electronic band, which derives its character from both the nanotube states and the C_{60} molecular orbitals (Figure 15–16).

3.3 Differential Conductance (dI/dV) Mapping

For a given energy eV , the tunneling conductance dI/dV reflects the density of states of the sample at that energy, $\rho_s(eV)$. As a result, it can be expected that high-resolution maps of tunneling conductance can provide detailed information on spatial variations in sample electronic structure. For example, by plotting dI/dV at an energy corresponding to a localized state at the sample surface, the spatial distribution of that sample state can be mapped.

In a typical experiment a constant current image, at a specific bias voltage and tunneling current used to stabilize the tunneling gap, is

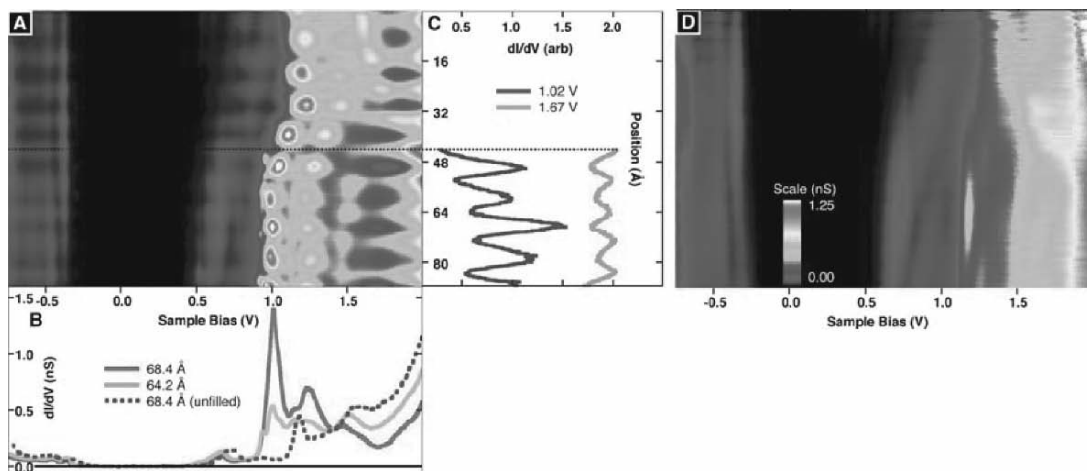


Figure 15–16. $I(V)$ tunneling spectroscopy on C_{60} /C-nanotube “peapods” (Reprinted with permission from Hornbaker et al., © 2002 AAAS). (A) Map of an array of full dI/dV spectra along the axis of a C-nanotube “peapod.” Sample bias voltage is plotted on the horizontal axis and displacement along the tube on the vertical axis. (B) Representative dI/dV spectra at selected positions along the tube. Large conductance peaks are found at positions of embedded C_{60} molecules. (C) Variation of tunneling conductance along the tube axis. (D) Reference spectroscopic map on an empty C-nanotube section without embedded C_{60} , in which no strong modulation of the tunneling conductance is observed. (See color plate.)

measured simultaneously with dI/dV maps at one or several voltages/energies. To obtain the spectroscopic data, the feedback loop is turned off at each point of the STM scan, and the selected bias voltages are applied. The tunneling conductance is measured directly using a modulation technique. At the end of the measurement, the feedback is reactivated, and the tip is moved to the next image pixel.

If a small modulation voltage $dV(t) = a \cos \omega t$ with amplitude a and frequency ω is added to the dc sample bias V_0 , the resulting tunneling current

$$\begin{aligned} I(t) &= I_0 + dI(t) = I_0 + \left. \frac{dI}{dV} \right|_{V_0} dV(t) + \dots \\ &= I_0 + \left. \frac{dI}{dV} \right|_{V_0} a \cos \omega t + \dots \end{aligned} \quad (5)$$

has a small ac component. Its amplitude at the modulation frequency ω , which can be measured using a lock-in amplifier, is proportional to $dI/dV|_{V_0}$. Via measurements at different dc bias V_0 , a set of tunneling conductance values at different energies can be determined at each pixel during an STM scan, and can later be displayed as maps of the sample density of states at those energies. Although quite simple conceptually, the long dwell times at each image pixel in dI/dV mapping pose stringent demands on microscope stability, which are best met at cryogenic temperatures. Under these conditions, thermal broadening of features in the density of states is also minimized.

dI/dV mapping has been applied in a wide variety of measurements of electronic states at surfaces and in nanostructures, such as atomic

chains (Crain and Pierce, 2005) or single molecules (Repp et al., 2005). A particularly widespread application is the imaging of quasiparticle interference. In an ideal metal, the Landau-quasiparticle eigenstates are Bloch wavefunctions with specific wavevectors $\mathbf{k}$ and energies ϵ . Their dispersion cannot be measured directly by STM. However, in the presence of disorder, e.g., impurities or crystal defects, elastic scattering mixes eigenstates with different $\mathbf{k}$ that are located on the same quasiparticle contour of constant energy in k -space. When states $\mathbf{k}_1$ and $\mathbf{k}_2$ are mixed by scattering, an interference pattern with wavevector $\mathbf{q} = \mathbf{k}_2 - \mathbf{k}_1$ appears in the norm of the quasiparticle wavefunction. The interference leads to modulations in the local density of states with wavelength $\lambda = 2\pi/|\mathbf{q}|$, which can be observed as modulations in the differential tunneling conductance. Via differential conductance mapping, interference of electronic eigenstates has been probed in metals (e.g., Cu(111), Crommie et al., 1993; Au(111), Hasegawa and Avouris, 1993) and semiconductors (e.g., InAs(110), Wittneven et al., 1998). When density of states modulations are detected by STM, certain contours of constant energy in k -space can be reconstructed by analyzing the Fourier transform of the real-space density of states map. Such Fourier transform spectroscopy simultaneously yields real-space and momentum-space information on wavefunctions, scattering processes, and quasiparticle dispersion.

Elastic scattering of Bogoliubov quasiparticles in superconductors can also give rise to conductance modulations (Hoffman et al., 2002a), suggesting that STM could be used as a local probe to study electron correlation in superconductors. Quasiparticle interference has indeed been demonstrated for cuprate high-temperature superconductors, such as $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (Hoffman et al., 2002a; Vershinin et al., 2004) and $\text{Ca}_{2-x}\text{Na}_x\text{CuO}_2\text{Cl}_2$ (Hanaguri et al., 2004). On $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ crystals cleaved at the BiO plane at cryogenic temperature in UHV, a constant-current image (Figure 15–17A) and differential conductance maps (Figure 15–17B) are obtained simultaneously in a low-temperature

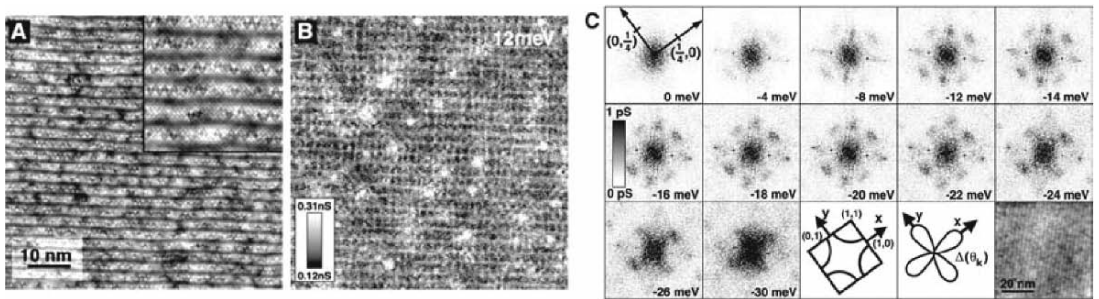


Figure 15–17. Fourier transform spectroscopy on a cleaved $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ high-temperature superconductor. (A) Constant-current STM image (65 nm, 0.13 nm resolution). (B) Simultaneously acquired differential conductance map ($\sim$ local DOS) at an energy of 12 meV below E_F , showing a pronounced modulation of the electronic structure. dI/dV is measured with 2 mV modulation. (C) Fourier transforms of DOS maps similar to (B), for energies up to 30 meV, representing maps of quasiparticle scattering wave vectors in this energy range. (Reprinted with permission from Hoffman et al., © 2002a AAAS.)

STM. The conductance map shows a checkerboard-like modulation of the local density of states. Fourier transforms of dI/dV maps obtained over a range of energies show dominant $\mathbf{q}$ -vectors associated with quasiparticle scattering. Different spatial patterns and wavelengths are observed at different energies. Applying a magnetic field perpendicular to the sample, the quasiparticle scattering can be modified by the controlled introduction of vortices (Hoffman et al., 2002b). Although the interpretation of these complex data sets is still somewhat controversial (Vershinin et al., 2004), these experiments clearly demonstrate the power of an STM-based approach for studying electronic structure and ordering phenomena in high-temperature superconductors.

A second example in which dI/dV mapping plays a key role is spin-polarized STM. A magnetic probe tip, e.g., a conventional etched W tip coated with a thin Fe film and polarized in a magnetic film perpendicular to the tip axis, is used for tunneling on a magnetic sample. Figure 15–18 illustrates the principle of spin-polarized STM for the example of a ferromagnetic Gd(0001) sample and a hard magnetic Fe tip (Bode et al., 1998). In a weak external magnetic field applied parallel to the sample surface, Gd(0001) has an exchange-split surface state with majority and minority parts at binding energies of -220 meV and $+500\text{ meV}$, respectively (Figure 15–18a). In the experiment, the magnetization direction of the Fe tip remains fixed independent of the direction of the applied magnetic field, while the magnetization of the Gd sample can be switched by reversing the direction of the applied field. Due to a spin valve effect, the tunneling current of the surface state spin component parallel to the tip magnetization is enhanced. If the majority spin is aligned with the tip magnetization (blue curve in Figure 15–18b) the tunneling conductance of the occupied majority

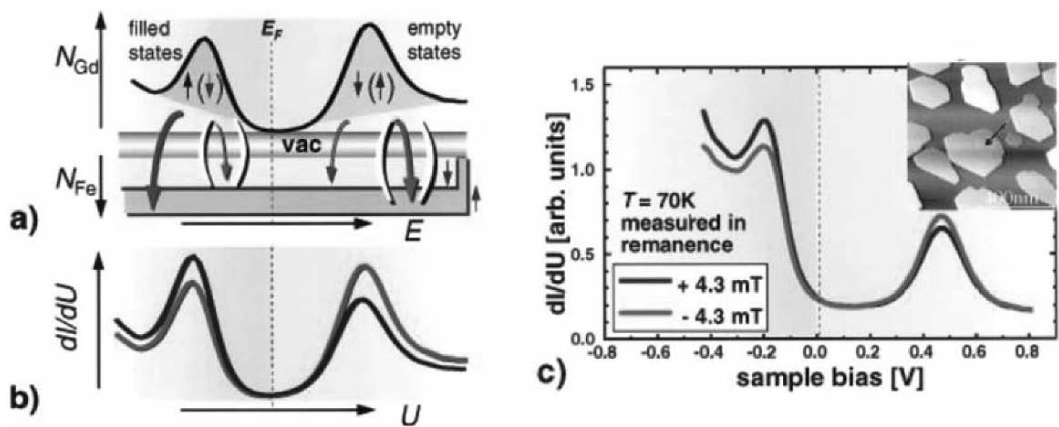


Figure 15–18. Spin-polarized STM on a Gd(0001) sample with an exchange-split surface state and a magnetic Fe tip with constant spin polarization close to E_F . (a) Due to the spin-valve effect the tunneling current of the surface state spin component parallel to the tip magnetization is enhanced. (b) Illustration of the reversal in the dI/dV signal at the surface state peak position upon switching the sample magnetically. (c) Experimental observation of this reversal in tunneling into an isolated Gd island (Reprinted with permission from Bode et al., ©1998 by the American Physical Society). (See color plate.)

state will be high, while that of the minority state is low. After reversal of the applied magnetic field, the minority spin will align with the tip magnetization. The conductance of the minority state thus increases, while that of the majority state is reduced. Difference spectra obtained by subtracting the traces, measured at two polarities of the external field, provide a direct measure of the sample magnetization, and can be used to build maps of magnetic order. Spin polarized STM has been shown to be a very powerful technique for mapping magnetic and antiferromagnetic order with atomic resolution, as demonstrated by studies on 2D antiferromagnetic ordering in Mn/W(110) (Heinze et al., 2000) and on magnetic hysteresis in Fe/W(110) (Pietzsch et al., 2001).

3.4 Inelastic Tunneling: Vibrational Spectroscopy

A major drawback of conventional STM imaging is its lack of chemical and element specificity, which stems from the fact that only states within few eV of the Fermi energy can be probed. As a result, characteristic core levels, which could provide element specificity, are not accessible. For single molecules, an alternative pathway to chemically specific imaging employs vibrational signatures, which are characteristic of specific molecular bonds.

The use of inelastic electron tunneling spectroscopy (IETS) for vibrational fingerprinting on molecules embedded in planar tunnel junctions dates back to the 1960s (Jaklevic and Lambe, 1966). The technique is based on the detection of a slight increase in tunneling conductance at an electron energy $eV = \hbar\omega$, corresponding to a vibrational mode with frequency ω of an embedded molecule. When the applied bias reaches this threshold, an inelastic channel for tunneling opens up in addition to the elastic tunneling at lower bias. Inelastically scattered electrons make up only a small fraction of the total tunneling current (typically a few percent). Nevertheless, the second derivative of the tunneling current, (d^2I/dV^2) , usually shows clear signatures of inelastic tunneling in the form of peaks at the characteristic bias voltages corresponding to different vibrational modes. As in dI/dV maps, d^2I/dV^2 can be measured by lock-in techniques, by adding a small modulation voltage to the tunneling bias V_0 . The signal amplitude measured at twice the modulation frequency is proportional to $d^2I/dV^2|_{V_0}$.

Conventional IETS in planar oxide tunnel junctions samples large ensembles of embedded molecules. Soon after the invention of STM, it was suggested that IETS should be feasible in local tunneling between a tip and a sample in STM (Binnig et al., 1985). A first convincing experimental demonstration of inelastic tunneling through few molecules was obtained in a setup of two crossed Au wires with finely adjustable normal force, allowing the trapping and IET spectroscopy at cryogenic temperatures on a small number of hydrocarbons (Gregory, 1990).

IETS on single molecules in a low-temperature STM was finally demonstrated by Ho and co-workers in 1998 (Stipe et al., 1998a). A beautiful experiment on single acetylene molecules adsorbed on Cu(100) not only provides clear evidence for inelastic tunneling at

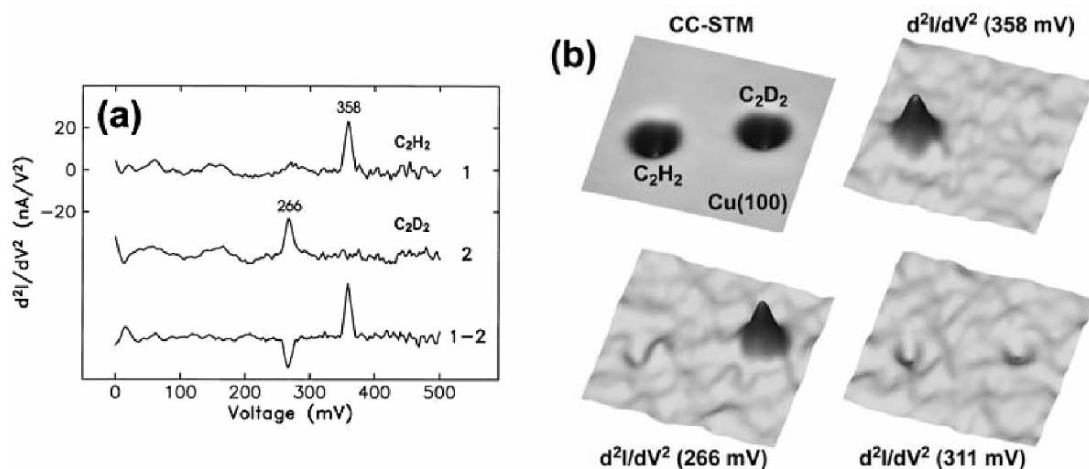


Figure 15-19. Single molecule vibrational spectroscopy and microscopy. (a) d^2I/dV^2 recorded over individual C_2H_2 and C_2D_2 molecules on $Cu(100)$ at 8 K. (1) Signature of inelastic electron tunneling by exciting the C—H stretch mode with an energy of 358 meV. (2) Isotope shifted (266 meV) inelastic tunneling peak of the C—D stretch mode. (b) Constant current (CC, 4.8 nA) STM of adjacent C_2H_2 and C_2D_2 molecules on $Cu(100)$, along with simultaneously acquired d^2I/dV^2 maps obtained at 358 mV and 266 mV, imaging the C—H and C—D bonds selectively. The symmetric round appearance is attributed to rotations of the molecules during the experiment. A d^2I/dV^2 map obtained at intermediate energy (311 mV) shows no vibrational contrast. (Reprinted with permission from Stipe et al., © 1998 AAAS.)

energies close to vibrational frequencies determined in the gas phase, but it also shows the expected isotope shift between measurements on normal and deuterated acetylene. Robust increases in differential conductance of the order of 3–12% are observed for this system.

In addition to inelastic tunneling spectra on individual molecular bonds, as shown in Figure 15-19a, inelastic electron tunneling microscopy with molecular resolution was demonstrated on small molecules, such as acetylene (Figure 15-19b). In constant-current STM imaging, both C_2H_2 and C_2D_2 molecules are imaged as identical depressions, i.e., are indistinguishable. On the other hand, C_2H_2 molecules are selectively imaged in (d^2I/dV^2) maps obtained at the characteristic energy of the C—H stretch mode (358 meV), while C_2D_2 molecules are imaged at the isotope shifted energy (266 meV). In a control experiment at an energy between these two characteristic values, none of the two isotopes gives rise to contrast.

The development of single-molecule vibrational spectroscopy and microscopy provides a powerful tool for performing and viewing chemistry at the ultimate limit of concentration (individual molecules) and spatial resolution (atomic). Local energy input by inelastic tunneling allows the precise manipulation of individual molecules (rotation, translation, see Section 4) and the cleavage and formation of individual molecular bonds, as well as the characterization of the individual and final configurations on the surface.

3.5 Element-Specific Imaging

In contrast to chemical specificity at the level of individual small molecules, which IETS can provide independent of the particular sample structure, no universal approach has been identified to obtain element specific contrast in STM. Various techniques have been suggested to achieve generic element specificity, among them a scanning tunneling atom probe operated by controlled atom transfer to the STM tip and desorption into a time-of-flight chamber (Weierstall and Spence, 1998), and a scanning probe energy loss spectrometer using an STM at high tip-sample bias in combination with an electron spectrometer for local probe electron energy loss spectrometry (Festy and Palmer, 2004). None of the techniques demonstrated to date provided a universal and practical pathway toward element specific imaging at the atomic scale. However, a number of sample-tip combinations have been identified, in which element contrast is indeed achieved on a case-by-case basis. Prime examples are bimetallic metal surfaces, on which different atomic species can be discriminated in constant current STM images.

In a few simple cases, a mere difference in atomic size can give rise to contrast, e.g., for submonolayer coverages of Pb/Cu(111) with a difference in metallic radii of nearly 50 pm (Nagl et al., 1994). In a broader class of metal alloys, electronic effects, i.e., differences in local density of states above different atoms, give rise to element contrast. Examples of systems with substantial density of states differences near the Fermi energy include alloys between transition metals with a partially filled d-shell and noble metals (Au, Ag, Cu). STM images of the (111) surface of AgPd alloys (Wouda et al., 1998) indeed show the Ag atoms with lower apparent height than the transition metal atoms, as expected based on a simple density of states argument. For alloys of transition metals (e.g., PtRh; Wouda et al., 1996), the prediction of the STM contrast is less straightforward, and generally requires *ab initio* calculations of the local density of states. For the PtRh(100) surface shown in Figure 15-20a, a local density of states difference at low tunneling resistance produces a height difference of about 20 pm between Pt and Rh, with Rh imaged at larger apparent height. In some systems, where sufficient differences in local density of states to provide STM contrast do not exist, such differences can be induced by suitable modification of the sample surface. In an important semiconductor system, $\text{Si}_{1-x}\text{Ge}_x$ alloys, Si and Ge atoms are indistinguishable in conventional constant current STM, which severely hampers the understanding of epitaxial growth since crucial local composition information is lacking. Covering a heterogeneous (111) surface composed of Si- and Ge-rich areas with 1 atomic layer of Bi induces robust differences in apparent height of nearly 0.1 nm between the chemically different regions, thus providing a means for imaging and characterizing lateral heterostructures of monolayer-thick alternating epitaxial Si and Ge stripes (Kawamura et al., 2003).

A third mechanism that can give rise to atomic-scale contrast between different elements is thought to be based on a direct interaction between tip and sample, in what could be described as a “precursor” chemical

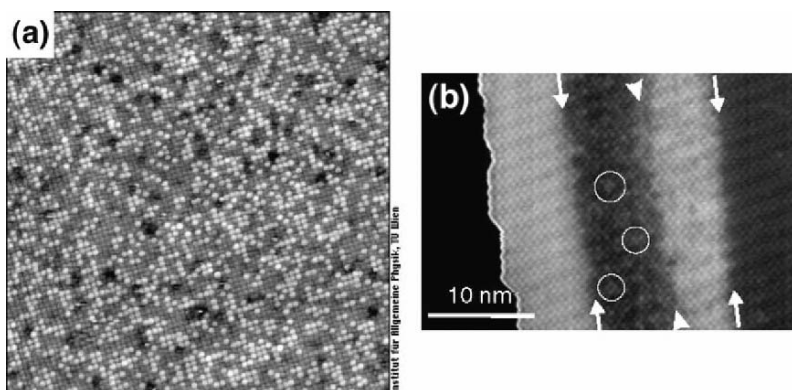


Figure 15–20. Element-specific imaging with atomic resolution by STM. (a) PtRh(100): for a bulk composition of 50% Pt and Rh, there are approximately 31% Rh (bright) and 69% Pt atoms (darker) at the surface. Black spots are caused by residual carbon. Pt and Rh atoms tend to cluster in small groups of the same species (Reprinted with permission from Wouda et al., © 1996 with permission from Elsevier). (b) Atomically and chemically resolved image of Ge (bright)/Si (dark) nanowires formed by sequential deposition of submonolayer amounts of Ge and Si in step flow mode on Bi-terminated Si(111). The initial Si/Ge boundary (right arrows) is nearly atomically sharp, while interdiffusion is observed at the Ge/Si boundary (arrowheads) (Reprinted with permission from Kawamura et al., © 2003 by the American Physical Society).

bond. With the tip–sample distance too large to allow true chemical bonding, either atomic relaxation or the local charge density can change as the tip is moved across different atomic species, which in turn affects the tunneling current and gives rise to apparent height differences in constant current STM. Qualitatively, surface atoms with higher chemical affinity to the tip atom should have higher tunneling conductance and thus will be imaged at larger apparent height. Several metal alloy systems, such as PtNi(111) and PtRh(111), are believed to show element contrast that is based on tip–sample interaction. Clearly, the nature of the tip apex would be of key importance for this type of contrast. This may explain why for some systems element contrast is obtained only after deliberate chemical modification of the STM tip.

4 STM at High and Low Temperatures

STM combines unique capabilities in both high-resolution imaging and spectroscopy at surfaces. Some form of controlled environment—such as UHV, an inert gas or fluid—may be required, mainly to ensure reproducible sample surface conditions. But many basic microscopy tasks, including a broad range of imaging and spectroscopy experiments, are performed conveniently at room temperature. Soon after the invention of STM, however, it was recognized that control over the sample temperature enables a wide variety of new experiments that are impossible to perform at room temperature. As an example, the structure or chemical reactivity of a given surface, e.g., of a catalyst, may depend on temperature. To characterize its properties in a particular temperature regime, the sample has to be heated or cooled during

STM imaging. Further, though imaging of thin film growth in a “quench and look” mode is possible, as discussed previously, studying growth processes directly at relevant sample temperatures would often be preferred. And finally, investigation of surface diffusion, chemical reactions, or molecular dissociation at surfaces often requires cooling of the sample to cryogenic temperatures to slow the rates of these thermally activated processes sufficiently to map them by relatively slow STM imaging. Probably the best known example of the benefits of cryogenic STM is atom and molecule immobilization for atomic or molecular manipulation, which allows artificial structures to be built from individual atoms and can provide highly idealized structures for probing chemistry and condensed matter physics at the ultimate spatial limit. Below, the technical challenges involved in temperature-dependent STM are discussed briefly, and selected examples of the main classes of STM experiments at high and low temperature—the imaging of dynamic surface processes via STM movies and the assembly of nanostructures “one atom at a time” by atomic/molecular manipulation with the STM tip—are presented.

4.1 STM at High and Low Temperatures: Motivation and Issues

Compared to room temperature operation, STM experiments at high, low, or even variable temperatures involve a number of additional issues. A primary experimental difficulty arises from thermal drift due to incomplete thermalization and the resulting variations in thermal expansion across the microscope, which cause relative motion between the STM tip and sample. Thermal expansion coefficients are temperature dependent, and generally approach zero as $T \rightarrow 0$. As a result, thermal drift rates tend to be low and the overall instrument stability is improved in low-temperature STM. Stable low-temperature microscopes reach lateral drift rates of 0.1 nm/h (Meyer and Rieder, 1997), and values of vertical stability of the tunneling gap approaching 1 pm have been reported (Stipe et al., 1998a). Beyond instrument stability, there are strong additional incentives for operation at cryogenic temperatures. The rates of thermally activated processes, such as surface diffusion of adsorbed atoms or molecules, can be slowed considerably at low temperatures. The freeze-out of thermal motion paves the way for time-lapse movies documenting diffusion processes, and for controlled atom or molecule manipulation using the STM tip. On the other hand, the thermal broadening of the tunneling distribution and of the electronic structure can be minimized at cryogenic temperatures, which results in higher energy resolution in spectroscopy.

STM imaging at high temperatures is generally affected by high drift rates, which may require active drift compensation between consecutive scans to allow imaging a given region on the sample for extended periods of time. Intrinsic drift rates of 3 nm/min without and 0.2 nm/min with drift correction have been reported for dedicated high-temperature microscopes (Voigtländer, 2001). Additional practical difficulties involve sample heating, temperature measurement, as well as the need for high scan speeds to capture dynamic phenomena at high temperatures.

4.2 Capturing Dynamic Surface Processes: STM Movies and Atom Tracking

The most popular approach for imaging dynamic processes involves repeated scanning of the same sample area to generate a time-lapse STM “movie.” This capability has been used to study a wide variety of phenomena, including step dynamics (Kuipers et al., 1993), epitaxial growth (Voigtländer and Zinner, 1993), self-diffusion on semiconductor (Borovski et al., 1997) and metal surfaces (Horch et al., 1999), as well as diffusion of adsorbates (Wintterlin et al., 1997) and large molecules (Schunack et al., 2002). If rates of dynamic processes, such as surface diffusion, are measured at several temperatures, activation energies of these processes can be determined. Complex reaction pathways, including adsorption, diffusion, dissociation, etc., can be elucidated at the atomic scale, and active sites can be identified.

As a near-field microscopy technique STM is based on raster scanning a probe tip. The scan speed is limited by the resonance frequency of the scanning element, typically below about 10 kHz, and by the electronic bandwidths of the tunneling current amplifier and feedback circuit. Frame rates in STM movies can be expected to be inherently lower than those in parallel imaging techniques, such as transmission electron microscopy or low-energy electron microscopy (Bauer, 1998; Tromp, 2000). To achieve adequate time resolution, the scan sizes in STM movies tend to be small, of the order of 10^4 pixels. Additional measures are typically necessary to deal with slow imaging rates and capture the detailed time evolution of a given process. Whenever a phenomenon under consideration permits, cooling of the substrate can be employed to slow the rates of thermally activated processes. In thin film growth, slow evaporation rates are employed. In studying surface reactions involving adsorption from the gas phase, e.g., on catalysts, low partial pressures of the reactive species in the gas phase are maintained. As an alternative to these often rather restrictive provisions, the development of high-speed microscopes and control electronics has been a focus of active research recently. As an example, scanners with mechanical resonances in the range between 50 and 100 kHz have been developed. Combined with a fast current amplifier (600 kHz bandwidth) and feedback loop (1 MHz), atomically resolved images have been obtained at rates approaching 200 frames/s (Rost et al., 2005). Such exciting instrument developments may in the future allow the imaging of new classes of dynamic surface phenomena at relaxed ambient conditions (higher temperatures, pressure, and growth rates) by STM.

To illustrate the use of STM movies at cryogenic temperatures to identify and quantify dynamic surface processes, we discuss the example of rutile $\text{TiO}_2(110)$, which has emerged as the prototypical system for fundamental surface science studies of transition metal oxides (Figure 15–21). TiO_2 has numerous applications in areas as diverse as heterogeneous catalysis, solar cells, photocatalysis, and organic waste remediation, and STM plays a key role in elucidating its fundamental surface processes. The $\text{TiO}_2(110)-(1 \times 1)$ surface consists

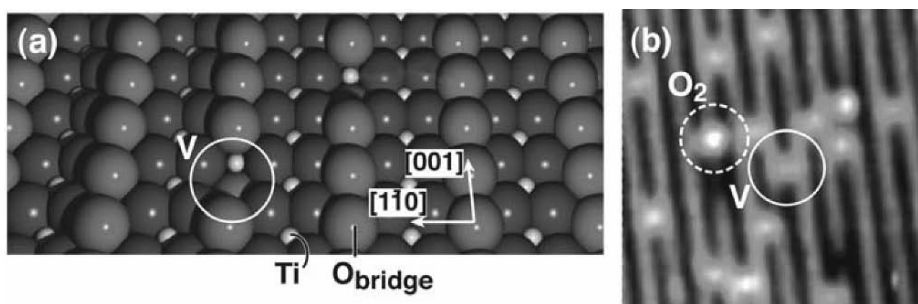


Figure 15–21. STM on Rutile $\text{TiO}_2(110)$. (a) Structure of the $\text{TiO}_2(110)$ surface. V, a single oxygen vacancy. (b) Constant-current STM on $\text{TiO}_2(110)$. Bright rows are assigned to five-fold coordinated Ti atoms and dark rows to bridging oxygen atoms. Oxygen vacancies are imaged as protrusions between the Ti rows. Adsorbed O_2 molecules are associated with bright protrusions on the Ti rows. (Reprinted with permission from Schaub et al., © 2003 AAAS.)

of alternating rows of Ti and O atoms aligned along the $[001]$ direction. Due to electronic effects, STM images Ti as protruding rows and bridging oxygen rows, which are geometrically highest on the surface by about 0.12 nm, as troughs.

The reactivity of $\text{TiO}_2(110)$ is affected to a great extent by the presence of oxygen vacancies, which are generated in the process of reducing the surface by annealing in vacuum. Using STM movies, the reaction mechanisms at these defect sites were studied for model reactions such as water dissociation (Brookes et al., 2001; Schaub et al., 2001). The diffusion of oxygen vacancies follows an intriguing mechanism mediated by O_2 molecules (Figure 15–22) (Schaub et al., 2003). Oxygen vacancies

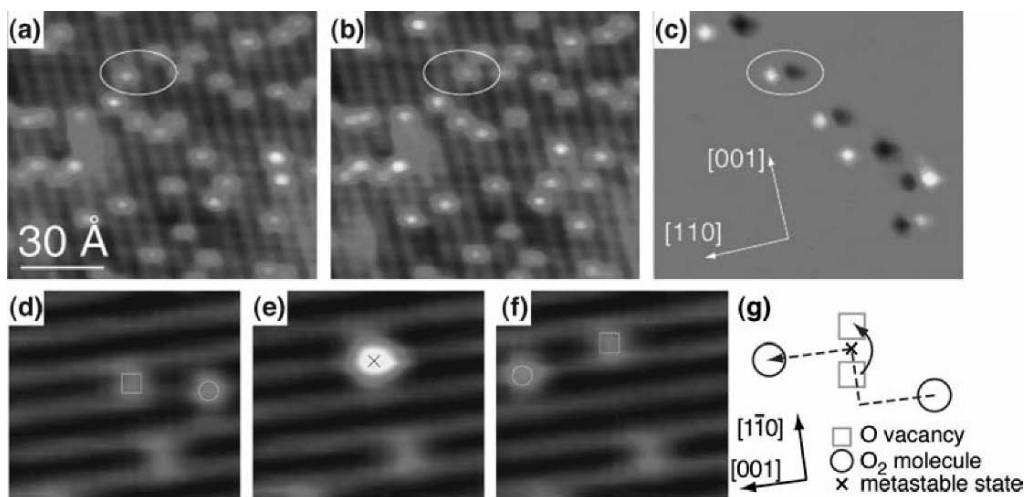


Figure 15–22. O_2 -mediated diffusion of oxygen vacancies on $\text{TiO}_2(110)$. (a and b) Consecutive frames of an STM movie on the motion of oxygen vacancies on $\text{TiO}_2(110)$ ($T = 300$ K, 8.5 s/frame). (c) Difference image of (a) and (b), highlighting changes due to the diffusion of single oxygen vacancies. Vacancies diffuse perpendicular to the bridging oxygen rows. (d–f) Time-lapse STM of the O_2 -assisted diffusion of an individual oxygen vacancy ($T = 230$ K, 1.1 s/frame). (g) Schematic illustration of the O_2 -mediated vacancy diffusion mechanism. (Reprinted with permission from Schaub et al., © 2003 AAAS.)

are immobile on the adsorbate-free surface. Vacancy diffusion is greatly enhanced by adsorbed O_2 . At temperatures sufficiently low that the diffusion of adsorbed O_2 molecules can be captured by STM, the subtraction of consecutive frames in time-lapse STM shows that single oxygen vacancies diffuse along $[\bar{1}10]$ from one bridging O row to the next, always in the presence of neighboring O_2 molecules.

The role of O_2 molecules in the vacancy diffusion process is established from detailed investigation of single vacancy hops, again based on time-lapse STM movies at low temperature. As an O_2 molecule diffusing along a Ti row approaches an oxygen vacancy, it dissociates and contributes one oxygen atom toward healing the vacancy, thus creating a metastable intermediate consisting of a single O atom. The O adatom is highly reactive, as corroborated in separate experiments involving dosing of atomic oxygen. It rapidly recombines with a bridging O atom and emerges as an O_2 molecule. If in this process the bridging O atom is removed from one of the adjacent rows, the net result is a diffusion jump of an oxygen vacancy by one bridging oxygen row. Given this O_2 -mediated mechanism of oxygen vacancy diffusion, the rate of diffusion events is expected to scale linearly with O_2 coverage. STM movies obtained at different O_2 exposure show that this is indeed the case.

While early imaging of dynamic surface processes was performed almost invariably in UHV, several applications require STM imaging in what is seen as more “realistic” environments for those applications. A prominent example is heterogeneous catalysis. It has been recognized that actual reactions under technologically relevant conditions, often involving elevated temperatures and pressures at or above atmospheric pressure, can involve surface structures and compositions, and entire reaction mechanisms that differ substantially, even qualitatively, from those of “simulated” reactions running in UHV, a situation commonly termed the “pressure gap” problem of heterogeneous catalysis. To address the need for imaging with high spatial and temporal resolution at elevated pressure, a family of dedicated STM instruments was developed (Rasmussen et al., 1998; Jensen et al., 1999; Lægsgaard et al., 2001; Rößler et al., 2005). These instruments allow sample preparation and surface analysis in UHV, followed by exposure to reactants at high pressure and simultaneous STM imaging. A particularly elegant implementation of this concept is the “reactor STM,” allowing dynamic STM imaging of surfaces exposed to reactants in a compact catalytic flow reactor in combination with the simultaneous analysis of the reaction products by mass spectrometry.

Figure 15–23 shows an example of a complex dataset obtained during high-pressure CO oxidation on Pt(110) (Hendriksen and Frenken, 2002). The upper panel traces mass spectrometer signals for O_2 , CO, and CO_2 , showing initial exposure to CO, followed by the introduction of molecular oxygen into the reactor. The panel below shows representative STM images obtained at specific stages of the reaction, during which the sample is kept at a constant temperature of 425 K. Images A, B, E, F, and H show flat terraces separated by steps, representing the metallic, CO-covered Pt(110) surface. Image C shows the change in

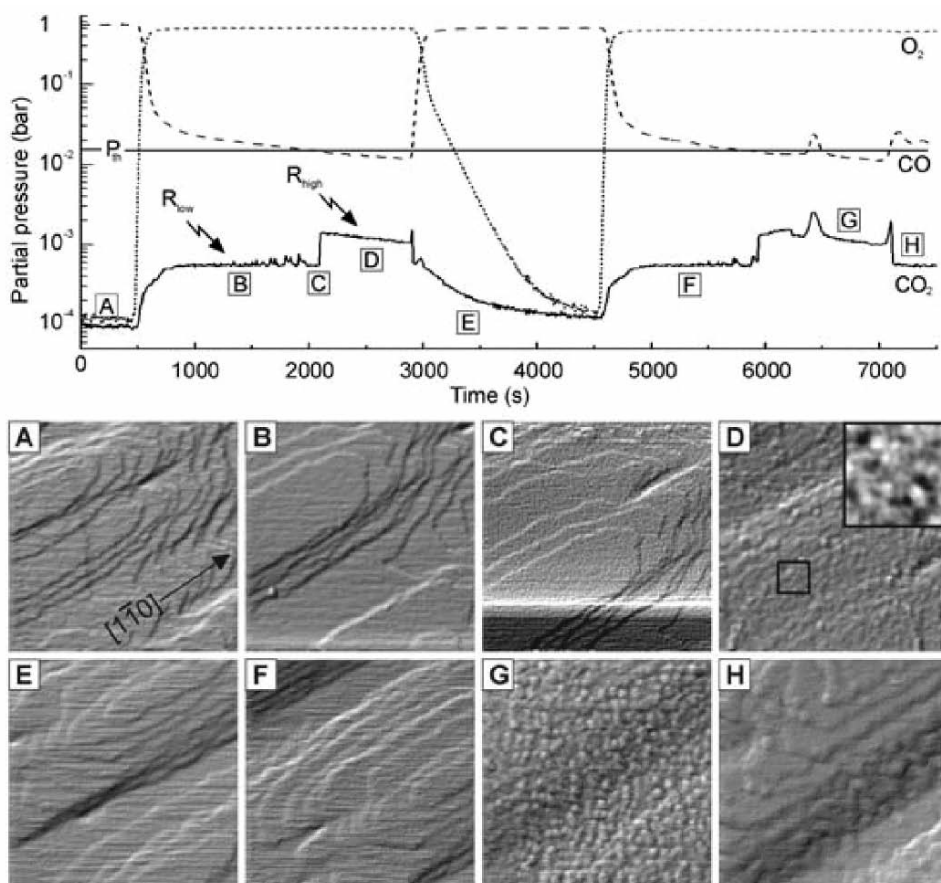


Figure 15-23. Simultaneous mass spectrometry and STM imaging in a catalytic flow reactor: CO oxidation on Pt(110). (Top) Mass spectrometer signals of O_2 , CO , and CO_2 measured at the output of the flow reactor cell. (Bottom) STM images on the Pt(110) catalyst surface acquired at selected stages of the process: CO adsorption (A), O_2 flow (B–D), and repetition of the sequence (E–H). Note the strong surface roughening in (D), associated with increased CO_2 evolution. (Reprinted with permission from Hendriksen and Frenken, © 2002 by the American Physical Society.)

surface morphology, a pronounced surface roughening, during a step in activity giving rise to a sudden increase in CO_2 evolution, demonstrating a direct link between surface roughness and activity for this surface involving a mechanism that is not observed at low pressure.

Although capable of mapping dynamic surface phenomena, STM movies have obvious limitations in imaging fast dynamic processes, such as surface diffusion at room temperature or above. A possible solution is the development of novel approaches and instruments for very high-speed STM imaging. As an alternative, frame-by-frame imaging can be abandoned altogether if only a map of the trajectory of the diffusing species is desired. The recognition of this fact led to the development of atom tracking STM (Swartzentruber, 1996). In atom tracking, the STM tip is locked onto a diffusing surface species using a 2D lateral feedback mechanism. Once locked, the feedback maintains

the tip over that species and tracks its coordinates as it diffuses over the substrate. In the atom-tracking mode, the STM spends all of its time measuring the diffusion trajectory, which results in substantially improved time-resolution compared to time-lapse STM movies.

Atom tracking STM has been used to measure the diffusion kinetics of Si (Swartzentruber, 1996) and SiGe (Qin et al., 2000) dimers on Si(001) above room temperatures, of water molecules on Pd(111) (Mitsui et al., 2002), and of Pd atoms in a Pd/Cu(001) surface alloy (Grant et al., 2001). The latter example is illustrated in Figure 15–24. Pd atoms in the surface alloy are imaged as protrusions in STM. Locking the STM tip onto individual Pd atoms, their diffusion pathway can be tracked. Analyzing the residence time (the time between hops) and jump length leads to the conclusion that there is no time correlation between individual Pd diffusion events, and that the diffusion is mediated by surface vacancies rather than Cu adatoms, i.e., involves rapidly diffusing vacancies visiting Pd atoms in the surface layer. From the temperature dependence of the Pd hop rate—determined from temperature-dependent tracking experiments—the activation energy of the overall process, in this case equal to the sum of the vacancy formation energy and the energy barrier for lateral Pd-vacancy exchange, can be measured.

In contrast to time-lapse STM, in which the tip is scanned rapidly across a larger field of view, the tip maintains close contact with the diffusing entity in atom tracking. Hence, the observed diffusion process could be affected by tip–sample interactions. While this question has to be studied on a case-by-case basis, at least one system [SiGe dimer

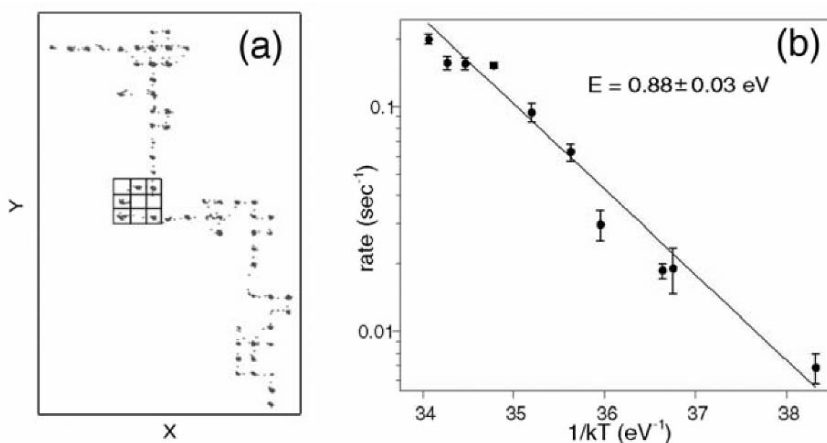


Figure 15–24. Diffusion kinetics of Pd atoms in the Pd/Cu(001) surface alloy. (a) Site visitation map of an individual Pd atom, obtained by atom-tracking STM at a temperature of 62°C. The square array marks the position of the Cu(001) unit mesh. In this dataset the atom hopped 853 times in a time interval of 5557 s. (b) Temperature dependence between 31 and 69°C of the average hop rate of an incorporated Pd atom. The average residence time of Pd atoms decreases from 145.3 to 5.0 s in this temperature range. The data follow an Arrhenius form with an activation energy of 0.88 eV and measured prefactor of $10^{12.4 \pm 0.4}$ Hz. (Reprinted with Permission from Grant et al., © 2001 by the American Physical Society.)

diffusion on Si(001)] has been identified in which the diffusion mechanism depends on the sign of the electric field between tip and sample (Sanders et al., 2003), indicating that the presence of the probe tip can indeed affect the measurement.

4.3 Atom and Molecule Manipulation

Atom and molecule manipulation (Hla and Rieder, 2003) experiments utilize the sharp STM probe tip and the ability to control it laterally and vertically with picometer resolution to build and modify nanometer-scale structures, typically in combination with their analysis by STM imaging and tunneling spectroscopy. While conceptually related to atom tracking, manipulation experiments are performed in a different regime: the sample is cooled to temperatures low enough that all thermal diffusion is frozen out, and the tip is typically brought into close contact to induce strong tip–adatom/molecule interactions.

Since the first demonstration of atomic manipulation of Xe/Ni(110) by Eigler and Schweizer (1990), manipulation experiments have branched out considerably, and now encompass scenarios as diverse as controlled chemical reactions of individual molecules (Lee and Ho, 1999), contacting single molecules with atomic metal wires (Nazin et al., 2003), construction of mechanical logic gates from adsorbed molecules (Heinrich et al., 2002), and control of excess charge on individual atoms at insulator surfaces (Repp et al., 2004). While generally based on some form of tip–sample interaction, atom or molecule manipulation can involve a number of different physical mechanisms: close proximity and short-range chemical interaction between a tip atom and adatom to affect the potential landscape seen by the adatom on the surface, vibrational excitation within adsorbed molecules or between substrate and adatom, electric field, or direct charging by tunneling electrons. Different patterns of adsorbate motion can be induced, including lateral hopping between adsorption sites on the substrate, vertical transfer between sample and tip, rotation, as well as the controlled making and breaking of individual bonds. Substrate surfaces with relatively low symmetry, e.g., (110) surface orientations or stepped surfaces, are often used to establish one-dimensional diffusion pathways between substrate adsorption sites that help guide the manipulation process.

Lateral manipulation is based on establishing close proximity between an adatom and the STM tip to increase the interaction between them. The approach process is controlled via the tunneling resistance, which is lowered from typically several hundred M Ω during imaging to values of the order of 100 k Ω for manipulation. Depending on the particular system, the lateral force needed to move an adsorbate between adjacent adsorption sites can be either repulsive or attractive, giving rise to manipulation by pushing and pulling, respectively. The manipulation process itself is monitored by measuring the tip height (or *z*-piezo voltage) during the lateral motion of the tip, as illustrated in Figure 15–25 for Pb/Cu(211) (pulling) and CO/Cu(211) (pushing) (Meyer and Rieder, 1998).

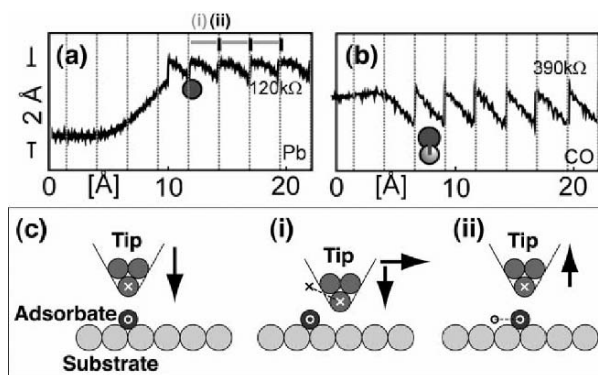


Figure 15–25. Mechanisms of lateral atom manipulation. (a) “Pulling” manipulation via attractive interaction between a Pb atom on Cu(211) and the W tip. (b) “Pushing” manipulation via repulsive interaction between a CO adsorbate and the W tip. In both cases a sawtooth-like tip height profile with the periodicity of the substrate adsorption sites is observed (Reprinted with permission from Meyer and Rieder, © 1998 The Materials Research Society). (c) Illustration of the tip and adsorbate motion during manipulation via attractive interaction. (i) Lateral motion of the tip, accompanied by an approach toward the substrate. (ii) Hopping of the adsorbate, followed by a sharp retraction of the tip.

The adsorbate motion in both pushing and pulling modes is strongly influenced by the preferred adsorption sites defined by the substrate. A sawtooth-like vertical motion of the tip accompanies each jump of the adsorbate between neighboring adsorption sites (Figure 15–25a and b). The origin of this tip motion is illustrated in Figure 15–25c for the example of an attractive tip–adsorbate interaction. Following the approach to the adsorbate, the tip is moved laterally. Initially, the adsorbate, held in the potential well of a substrate adsorption site, does not follow. The tunneling gap thus increases, and the tip moves forward to maintain a constant tunneling current (i). Ultimately, the interaction with the tip induces a lateral jump of the adsorbate into a neighboring potential minimum. The tunneling gap closes abruptly and the feedback loop causes the tip to retract.

The precision and complexity of structures achievable by lateral manipulation is illustrated in Figure 15–26 (Heinrich et al., 2002). CO molecules adsorbed in monomer, dimer, or trimer configurations on Cu(111) give rise to distinct contrast in constant current STM. While monomers and dimers are stable at low temperature (5K in this example), there are three distinct configurations for trimers: a stable three-fold symmetric “close-packed” arrangement, and metastable “straight-line” and “bent-line” (“chevron”) configurations, which relax with time constants of the order of seconds into the stable state. Complex structures consisting of up to 545 CO molecules were built with atomic precision such that any neighboring molecules were initially in stable dimer configurations, but could be changed to an unstable “chevron” configuration with a place exchange of a single molecule in the vicinity. In this way, molecular cascades, similar to toppling rows of standing

dominoes, could be constructed. Mechanical computation was demonstrated by molecular cascades set up as logic AND gates, two-input, and three-input sorters.

Apart from lateral manipulation, vertical manipulation, i.e., transfer of adsorbates between sample and tip, is possible. Vertical manipulation of CO molecules, for example, has been used for controlled modification of the STM tip and to induce single molecule reactions (Lee and Ho, 1999). Vertical transfer is also important for moving adatoms or molecules across obstacles such as step edges, if they cannot be surmounted by controlled lateral hops. Manipulation mechanisms other than direct chemical interaction between the STM tip and adsorbates have received increased interest recently. Notably the different roles of vibrational excitations are actively investigated (Komeda et al., 2002; Stroscio and Celotta, 2004). Inelastic tunneling can, for instance, drive controlled rotations of adsorbed molecules (Stipe et al., 1998b). An intriguing example of atomic manipulation based on inelastic tunneling, the controlled manipulation of excess charge on single Au atoms, is illustrated in Figure 15–27 (Repp et al., 2004). Individual Au

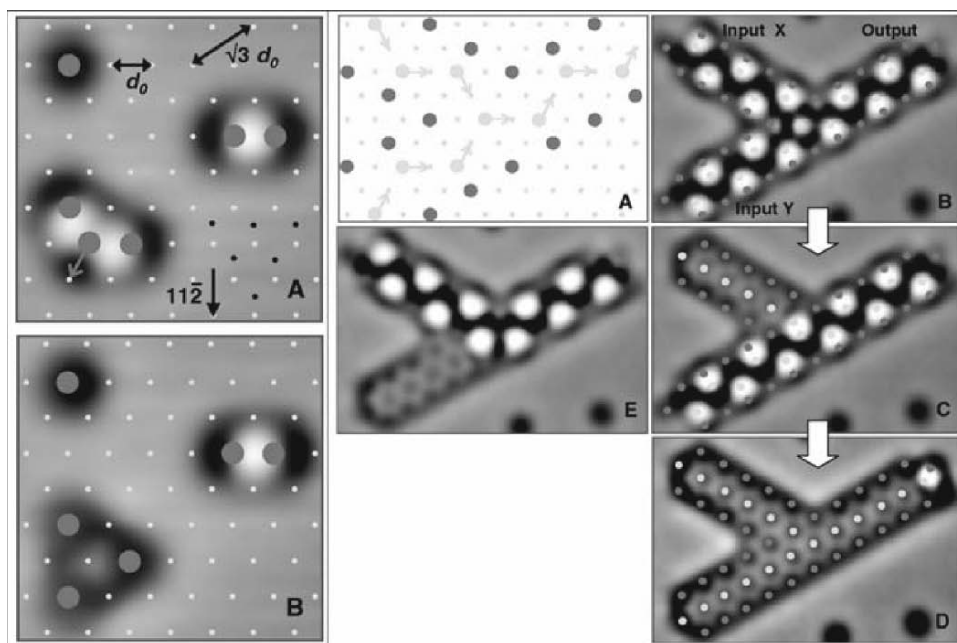


Figure 15–26. Molecule cascades. (Left) (A) Configurations of neighboring CO molecules on Cu(111): isolated CO molecule, dimer, and trimer in a “chevron” configuration. Large circles mark the positions of the molecules and small dots indicate the surface layer Cu atoms (1.9 nm scans; $T = 5$ K). (B) The same area after one CO molecule in the trimer has hopped to generate a stable close-packed trimer. (Right) Demonstration of a molecular mechanical AND gate, implemented via cascades of “chevron”-type CO trimers. (A) Model of the AND gate. (B–D) Sequence of STM images (5.1 nm by 3.4 nm) showing the operation of the AND gate: (B) Initialization. (C) Result after input X was triggered with the STM tip. (D) When input Y was triggered, the cascade propagated all the way to the output. (E) Result of triggering only input Y. (Reprinted with permission from Heinrich et al., © 2002 AAAS.)

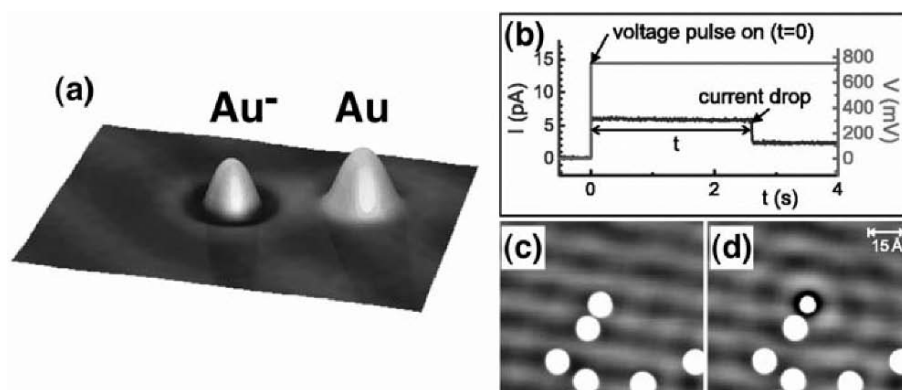


Figure 15–27. Charging of individual Au atoms on ultrathin NaCl/Cu(111). (a) Constant-current STM on a pair of Au atoms, one of which was charged negatively with the other remaining neutral. (b) Signature of the charging event. Successful charging, achieved by a voltage pulse while maintaining the tip above a single Au atom, is indicated by a sudden drop in tunneling current. (c and d) Quasi-particle scattering: neutral Au adatoms do not scatter NaCl/Cu(111) interface-state electrons (c), whereas the negatively charged adatom acts as a scatterer (d). (Reprinted with permission from Repp et al., © 2004 AAAS.)

atoms are deposited on ultrathin, insulating NaCl supported by Cu(111). Electrons can tunnel between the metal substrate and the STM tip through the ultrathin insulator, i.e., STM imaging, spectroscopy, and manipulation are possible for this system. However, the coupling of the Au adsorbate to the substrate is affected profoundly by the insulating support. Voltage pulses can be used as a means to reversibly deposit an excess electron on individual Au atoms. The charging occurs via an inelastic electron tunneling mechanism enabled by a weak coupling of the adatom and Cu substrate electronic states. The data suggest that the coupling is so weak that the lifetime of a negative ion resonance state of the adatom is in the range of ionic vibrational periods for the NaCl interlayer, allowing the relaxation of the NaCl lattice, shift of the negative ion resonance state below the Fermi energy, and capture of the electron. The excess charge is maintained due to the substantial relaxation of the underlying NaCl lattice. This stabilization prevents discharging by tunneling into the metal, and the electron resides on the Au atoms until removed by a voltage pulse of opposite sign. The charged Au atom has a distinct signature in constant-current STM imaging, and the long-range electric field due to the charged Au atom strongly scatters electrons in interface states at the Cu/NaCl interface. Charging individual adsorbed atoms is a potentially powerful approach to tuning their physical properties. For Au/NaCl, significant differences in surface diffusivity were identified, with the onset temperature of significant surface diffusion lowered from 60 K (Au) to 40 K (Au^-). Other possible modifications due to controlled charging at the atomic level include changes in catalytic activity and the controlled manipulation of the net spin magnetic moment of individual atoms.

5 Heterostructures and Buried Interfaces: BEEM, Quantum Size Effects, and Cross-Sectional STM

The exponential dependence of the tunneling current on the tip–sample separation makes STM a versatile tool for atomic-resolution microscopy and spectroscopy at surfaces. As a result of its unique resolution and surface sensitivity STM has brought important advances in fields such as catalysis or epitaxial growth, in which surface processes play a key role. For many technological applications of semiconductor materials, e.g., in device structures such as transistors, detectors, or lasers, the active regions encompass complex heterostructures, and device performance is affected much less by the free surface than by buried interfaces. Therefore, a need arises for a technique capable of probing subsurface structures, electronic properties, and carrier transport. The mapping of interfacial and transport properties should occur with nanometer spatial resolution to provide data that are relevant for semiconductor devices with progressively reduced dimensions. Since the early 1990s, several STM-based techniques have been developed to provide high-resolution imaging and spectroscopy of buried interfaces and heterostructures. Ballistic electron emission microscopy (BEEM) operates in a three-terminal configuration, in which the STM tip, whose height is controlled by a constant-current feedback loop, injects hot carriers into a thin film or heterostructure while an additional collector contact measures the current of carriers that are transmitted through buried interfaces. Electron interference in a thin metal film, again with carriers injected from an STM tip and traveling ballistically in the metal, provides detailed thickness maps of metallic overlayers, and can—under favorable circumstances—even map the atomic structure of a buried metal–semiconductor interface. Cross-sectional STM, finally, is used to image complex heterostructures such as superlattices, embedded self-assembled quantum dots, or substitutional magnetic impurities, on nonpolar (110) cleavage planes of III–V compound semiconductors.

5.1 Probing Buried Interfaces (I)—BEEM

BEEM was invented by Kaiser and Bell (1988) to probe Schottky contacts, i.e., metal–semiconductor interfaces with high spatial resolution (for a recent review, see Narayanamurti and Kozhevnikov, 2001). Beyond Schottky barriers, the technique can determine the height of other subsurface potential barriers and it has also been applied to measure band offsets between different semiconductors, the energies of transmission resonances in semiconductor quantum wells and superlattices, and bound states in buried quantum wires and dots. In addition, electron scattering at subsurface linear and point defects has been characterized.

We illustrate the operating principle of BEEM using the example of a metal–semiconductor junction. The technique operates in a three-terminal configuration, i.e., adds an additional collector electrode to the STM tip and sample contact, as shown schematically in Figure 15–28.

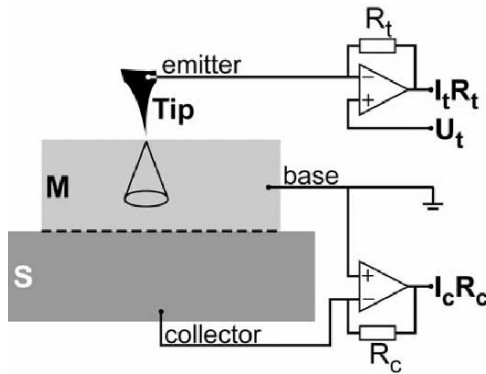


Figure 15–28. Schematic setup of a typical BEEM experiment on a metal (M)–semiconductor (S) heterostructure.

The STM tip is scanned at constant current I_t , measured between the tip and base electrode, over the surface of the base layer. The tip serves as a point source of hot carriers, electrons or holes, injected into the metal base. If the metal thickness is small compared to the mean free path for electron–phonon, electron–electron, or impurity scattering, most carriers reach the metal–semiconductor interface *ballistically*, i.e., with their original energy and momentum distribution. In metals, the mean free path for electron–phonon scattering of electrons with energies of a few eV is typically of the order of 10 nm. In significantly thicker films, electron–phonon scattering would broaden the momentum distribution of the injected carriers, i.e., mostly affect the spatial resolution of BEEM, while having little effect on the energy distribution. Hence, the injected ballistic carriers generally have a well-defined energy distribution and can be used to perform spectroscopic measurements on buried potential barriers.

At the metal–semiconductor interface, ballistic carriers with energies below a threshold equal to the Schottky barrier height $e\phi_B$ are reflected back into the metal base, while carriers with energy above the threshold are transmitted into the semiconductor collector (Figure 15–29). If

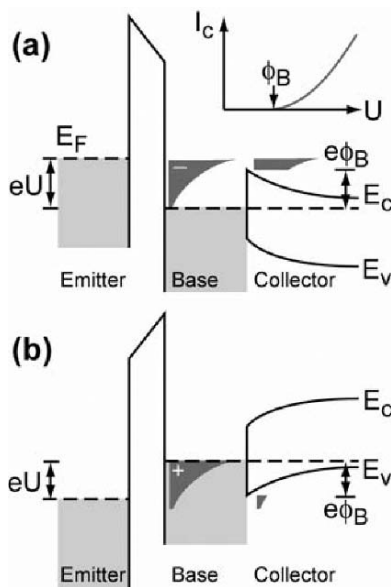


Figure 15–29. Band diagrams for (a) ballistic electron and (b) ballistic hole injection into the metal base on an n- and p-type semiconductor, respectively. The dark shaded areas indicate the energy distribution of the injected carriers, as well as of those transmitted into the collector. The inset in (a) illustrates the onset of the collector current at a threshold bias corresponding to the Schottky barrier height, $e\phi_B$. (After Bell et al., 1991.)

other barriers exist in the collector layer, the transport through that material can involve additional interfacial reflections. Carriers that are transmitted through the entire collector layer contribute to the collector current (or BEEM current), which is measured between base and collector contacts.

The energy of the injected carriers is varied by changing the voltage applied between the STM tip and metal base. For low bias, none of the injected carriers is transferred across the metal–semiconductor interface. At voltages above the threshold, some carriers have energy above the Schottky barrier and can cross the interface, causing an increase in collector current with increasing tip–sample bias above threshold. From the onset of the BEEM current, the barrier height $e\phi_B$ can be determined, e.g., by fitting the measured $I_c(V)$ to a theoretical expression. In a simple one-dimensional theory,

$$I_c(V) = RI_t \int dE [f(E) - f(E - eV)] \times \theta[E - (E_F - eV + e\phi_B)] \quad (6)$$

where R is a bias independent parameter and $f(E)$ is the Fermi function.

Figure 15–30 shows experimental BEEM spectra on an Au/*n*-Si(001) junction, measured at room temperature in N₂ atmosphere, fitted using Eq. (6) to determine a Schottky barrier $e\phi_B = 0.92$ eV. Due to its low reactivity and oxidation resistance, Au was used in many of the early BEEM experiments as a nonepitaxial metal base on a variety of other semiconductors. Schottky barrier heights were determined, for example, for Au/*n*-GaAs(001) ($e\phi_B = 0.70$ eV at 77 K) (Bell et al., 1990) and Au/*n*-GaP(110) ($e\phi_B = 1.41$ eV) (Prietsch and Ludeke, 1991). For Au/*n*-CdTe(001) ($e\phi_B = 0.69$ – 1.07 eV) (Fowell et al., 1990), strong lateral variations in the

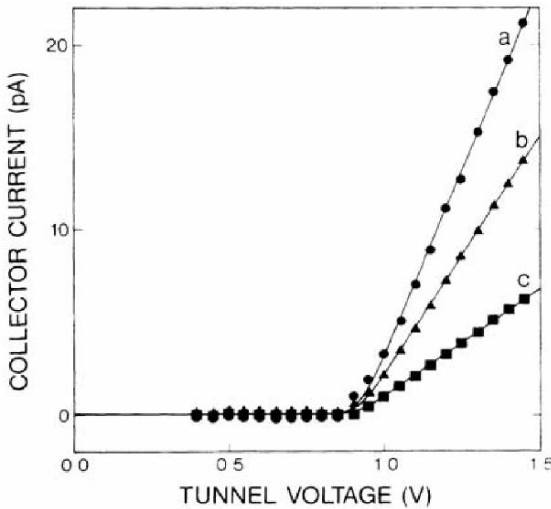


Figure 15–30. BEEM spectra obtained on a polycrystalline Au/*n*-Si(001) junction. Spectra a–c (symbols) are measured at different tunneling currents. The calculated spectra (solid lines) correspond to a common Schottky barrier height value $\phi_B = 0.92$ eV and R value of 0.045 eV^{−1}. (Reprinted with permission from Kaiser and Bell, © 1988 by the American Physical Society.)

measured Schottky barrier height were detected, emphasizing the need for high-resolution maps of interfacial transport.

Carrier transport in BEEM is affected by all stages of the injection pathway: (1) tunneling from the tip into the metal base layer, (2) hot carrier transport through the base, (3) transmission across the metal–semiconductor interface, and (4) transport in the semiconductor, including transmission through additional heterostructure interfaces. In addition to local Schottky barrier heights, the technique is therefore sensitive to a number of factors, e.g., surface topography, elastic or inelastic scattering in the different layers or at interfaces, and the band structures of the junction partners, which affect not only interface transmission probabilities but also the spatial resolution obtainable in BEEM via metal band structure-induced focusing or defocusing.

BEEM current maps can be acquired simultaneously with constant-current STM images by measuring I_c spatially resolved during a constant-current scan. Figure 15–31 shows STM and BEEM current images obtained on 2.5 nm epitaxial $\text{CoSi}_2/\text{Si}(111)$ (Sirringhaus et al., 1994). The lattice mismatch of 1.2% between the metallic silicide and the Si substrate is accommodated by an interfacial dislocation network. In STM topography, the location of these line defects is detected via their elastic strain field at the CoSi_2 surface. Strikingly, the defects are mapped in BEEM as sharply localized regions with increased collector current with a width of only 0.8 nm, as expected for ballistic transport of carriers with a strongly forward focused tunneling momentum distribution, i.e., $k_{\parallel} \sim 0$. While the Schottky barrier was found to be uniform, as expected for an epitaxial interface, the sharp local increase in transmissivity of the CoSi_2/Si interface was explained by a significant increase in in-plane momentum due to scattering by the dislocation core, facilitating the transmission into the Si conduction band minima.

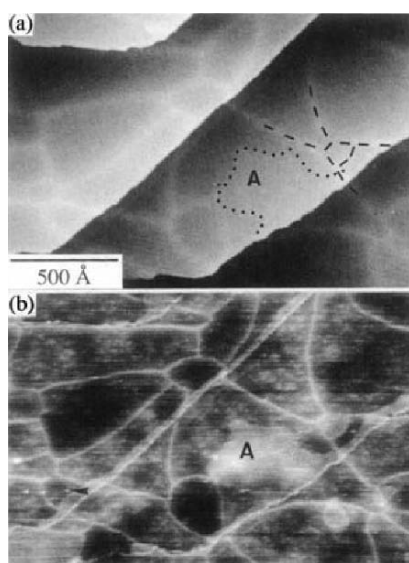


Figure 15–31. (a) Constant-current STM image and (b) corresponding BEEM image obtained on epitaxial 2.5 nm $\text{CoSi}_2/\text{Si}(111)$. A dislocation network at the silicide/silicon interface is indicated by dashed lines. In the BEEM image brighter areas indicate regions of higher collector current. Hot electron scattering at the interfacial dislocations causes a sharply localized increase in collector current.

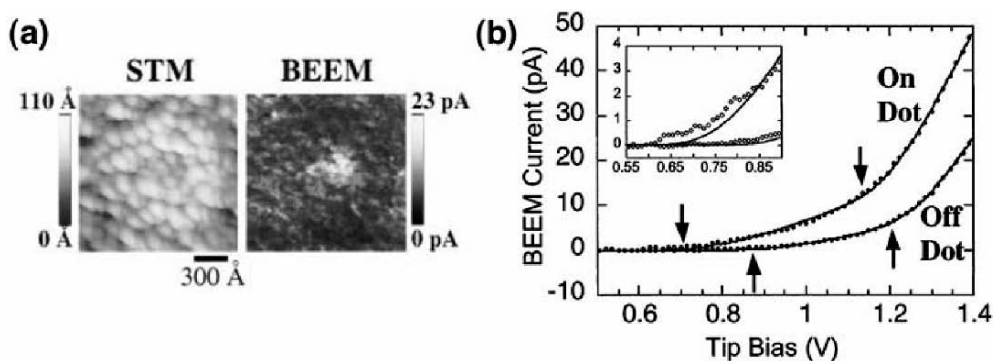


Figure 15–32. BEEM on self-assembled quantum dots. (a) STM and BEEM images of an individual InAs/GaAs quantum dot under a polycrystalline Au base. (b) Comparison of BEEM spectra obtained by ballistic electron injection into an InAs dot and into the wetting layer between dots. (Reprinted with permission from Rubin et al., © 1996 by the American Physical Society.)

In experiments such as those on $\text{CoSi}_2/\text{Si}(111)$, growth and probing of the entire heterostructure in the same UHV environment is critical to achieve ordered interfaces and reliable measurements on reactive metal surfaces. For the microscopy itself, low temperatures have several practical advantages, such as improved energy resolution due to reduced thermal broadening of the tunneling distribution, and reduced thermal drift of the STM. An additional benefit is a lower thermal noise due to a reduction of thermionic emission across interfacial barriers, which allows shallower potential barriers to be probed. As a result, semiconductor band offsets, resonant transport in semiconductor heterostructures, and interfacial barriers between an organic layer and a metal base (Troade et al., 2005) can be measured successfully by low-temperature BEEM. Examples are embedded self-assembled quantum dots (Rubin et al., 1996), lithographically patterned quantum wires (Eder et al., 1996), double-barrier resonant tunneling structures (Sajoto et al., 1995), and superlattices (Heer et al., 1998). Figure 15–32 shows the band profile for carrier injection into an individual InAs quantum dots in GaAs (Rubin et al., 1996). Also shown are STM and BEEM images, as well as BEEM spectra obtained with the tip positioned on top and next to the dot. The semiconductor heterostructure was coated with a polycrystalline Au base layer, whose morphology largely dominates the contrast in STM. A strong enhancement in BEEM current in a circular region with 30 nm diameter is associated with carrier transport through a single InAs quantum dot. A comparison of BEEM spectra obtained on and between dots shows signatures of transport through two zero-dimensional states of the dot.

Ballistic carrier transport and scattering can, in principle, be used to probe a wide variety of other systems, beyond Schottky barriers and semiconductor heterostructures. Examples are metal–insulator–semiconductor structures (Cuberes et al., 1994) or magnetic multilayers (Rippard and Buhrmann, 2000). In the latter, an Au/Si(111) Schottky barrier is used as an analyzer for spin-dependent scattering of carriers

in a Co/Cu/Co trilayer structure, used in spin-valve devices, integrated on top of the Au base layer. Unpolarized carriers injected from an STM tip into this magnetic multilayer base will undergo spin-dependent scattering in the ferromagnetic Co layers. In areas in which the Co layers couple ferromagnetically, only one spin component will be scattered heavily and a large fraction of unscattered carriers is transmitted across the Au/Si interface. If the magnetization direction of the two Co layers is misaligned, both spin components are strongly attenuated by scattering, causing a sharp drop in collector current.

5.2 Probing Buried Interfaces (II)—Quantum Size Effects

In BEEM hot carriers, locally injected into a metal/semiconductor heterostructure, are used to measure the transmitted current across a buried interface. Instead, carriers that impinge with energies below threshold and are reflected back into the metal film can also be considered. Due to the long mean free path, these carriers can set up standing waves in the metal film. An STM tip, probing the evanescent tails into the vacuum, can then be used to image fringes due to interference of these electrons. The concept of electron interference in metal films dates back to Jaklevic et al. (1971), and interference effects have been observed by angle-resolved photoemission, low-energy electron diffraction and low-energy electron microscopy.

An implementation of an STM experiment to probe interference effects in the system Pb/Si(111) is shown in Figure 15–33. Pb evaporated onto a stepped Si(111) substrate forms (111)-oriented islands whose surface is atomically flat. Due to the substrate vicinality, the Pb

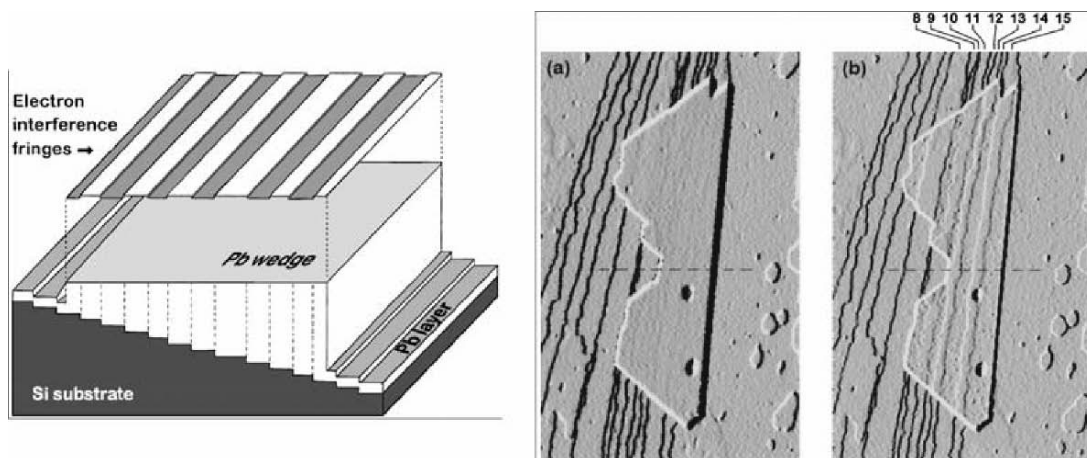


Figure 15–33. Electron interference in a Pb “quantum wedge” on Si(111). (Left) Geometry of the Pb wedge on a stepped Si(111) substrate. The Pb thickness varies in increments of roughly one Si(111) step height, giving rise to electron interference fringes in the direction of the substrate steps. (Right) Constant-current STM images ($730\text{ nm} \times 1100\text{ nm}$) at -5 V (a) and $+5\text{ V}$ (b) sample bias. The image obtained at positive bias shows apparent height changes at the surface of the wedge due to electron interference in the Pb film. (Reprinted with permission from Altfeder et al., © 1997 by the American Physical Society.)

island is wedge shaped, with a thickness that changes in integer multiples of the Si(111) step height from one substrate terrace to the next. While STM at positive sample bias, i.e., injection of electrons from the tip, images the atomically smooth Pb surface, STM images at opposite bias show bands of apparent terraces and steps, aligned with steps of the Si substrate, at the surface of the Pb wedge. These images, and associated tunneling spectra, are interpreted as signatures of quantum well states in the Pb wedge (Altfeder et al., 1997), and can be used to determine the position of subsurface steps as well as the position-dependent absolute thickness of the Pb film. Bands with constructive and destructive interference alternate with a thickness change d_0 of one Pb(111) monolayer if $d_0 \approx \lambda_F/4$, where λ_F denotes the Fermi wavelength. Similar interference effects were also observed by STM for epitaxial silicide layers, such as $\text{CoSi}_2/\text{Si}(111)$ (Lee et al., 1994) and $\text{NiSi}_2/\text{Si}(111)$ (Kubby and Greene, 1992).

Strikingly, electron interference can even be used to image interfacial atomic structures buried under as much as 10 nm of metal. In the Pb/Si(111) system, the (7×7) reconstruction of the Si(111) surface remains essentially intact upon low temperature evaporation of Pb, except for some intermixing by replacing Si adatoms by Pb (Altfeder et al., 1998). Due to the topology of the Fermi surface of Pb, in particular a large mismatch between the electron effective mass in in-plane and normal directions, the quantized electron states in the Pb film can be used to map interfacial structure with a resolution of 0.6 nm for overlayer thicknesses exceeding 10 nm, or roughly 10 times the Fermi wavelength in the metal.

5.3 Imaging Buried Heterostructures—Cross-Sectional STM

The strong interest in low-dimensional semiconductor structures—quantum wells, wires, and dots—has stimulated widespread activity in nanoscale imaging of electronic materials with reduced dimensionality. Recent efforts have focused on self-assembled quantum dots, generated by lattice mismatched heteroepitaxial growth. Semiconductor quantum dots with lateral size in the 10–100 nm range are readily imaged by STM if they are exposed as islands on a free surface. Consequently, a large number of studies have been devoted to studying epitaxial growth and quantum dot self-assembly by conventional STM imaging. However, almost any technological applications of self-assembled quantum dots require embedding in a matrix, often consisting of the substrate material. The embedding process causes significant modifications to the dots that include segregation and intermixing, shape changes, dopant redistribution, and adjustments to the local strain field in the dot and in the surrounding material. All these factors make it desirable to image embedded rather than exposed nanostructures.

Cross-sectional STM (X-STM), originally demonstrated by Feenstra et al. (1987) for imaging and spectroscopy on (110) surfaces of III–V compound semiconductors, offers an elegant solution to this challenge. Semiconductors that cleave easily, as most III–V compounds do,

are used as a substrate for the growth of embedded self-assembled quantum dots. A sample is then cleaved in UHV, and STM imaging is performed on the cleavage face, typically a nonpolar (110) plane, which provides large step-free areas for atomic resolution imaging. In this geometry, X-STM gives access to subsurface structures over the entire thickness of the epitaxial layer and the underlying substrate.

X-STM has been used to image a variety of buried semiconductor heterostructures, such as superlattice structures in GaAs/AlGaAs (Salemink and Albrektsen, 1991), InAs/GaSb (Feenstra et al., 1994), and GaN/GaAs (Goldman et al., 1996), showing atomic-scale composition fluctuations, interface roughness, lateral stain variations, and phase separation in these systems. The technique has seen a strongly revived interest with the advent of self-assembled quantum dots and quantum dot superlattices (Legrand et al., 1998). Specifically for the imaging of quantum dots, X-STM relies on the fact that the dots are small and form rather dense populations. Hence, a random cleavage will cut through a large number of these nanostructures and provide a cross-sectional view of their atomic structure on the cleavage plane (Figure 15–34).

Much of the power of X-STM imaging derives from the fact that bias-dependent imaging provides chemical contrast on (110) cleavage faces of III–V compounds (Feenstra et al., 1987). Empty-state imaging gives atomically resolved maps of the cation sublattice and allows a direct identification of atomic species, e.g., indium atoms in a GaAs matrix (Pfister et al., 1995). This electronic structure effect not only provides strong contrast to determine the shape of individual buried InAs quantum dots and their stacking in multilayer structures, but can also be used to quantify interface roughness, intermixing, and surface segregation, near the dots and the wetting layer, with atomic precision.

To access this information in high-resolution images, a background subtraction has to be performed to remove topographic contrast of the

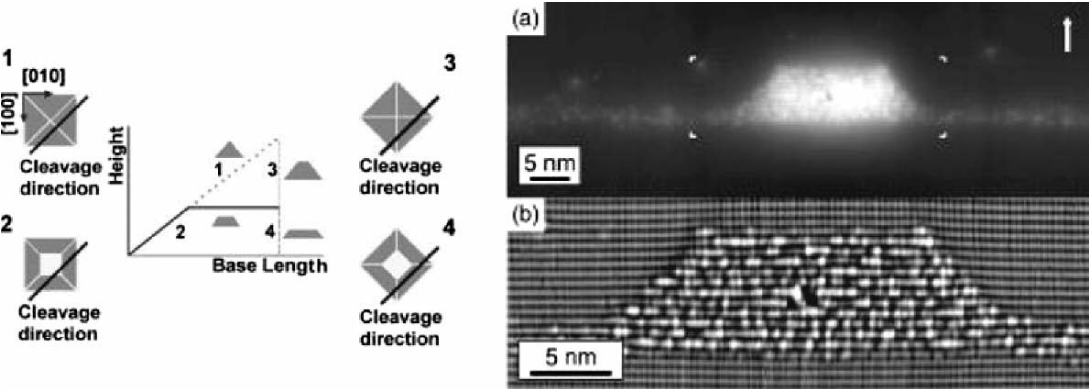


Figure 15–34. Cross-sectional STM on self-assembled quantum dots. (Left) Illustration of the possible apparent geometries observed due to cleavage at random positions through quantum dots with pyramidal and truncated pyramidal shape. (Right) Filled-state constant-current STM of an InAs quantum dot embedded in GaAs. Part of the image in (a) is treated by a local mean equalization filter to accentuate the atomic corrugations in the dot and the surrounding GaAs matrix, as shown in (b). (Reprinted with permission from Bruls et al., © 2002 American Institute of Physics; reprinted with permission from Gong et al., © 2004 American Institute of Physics.)

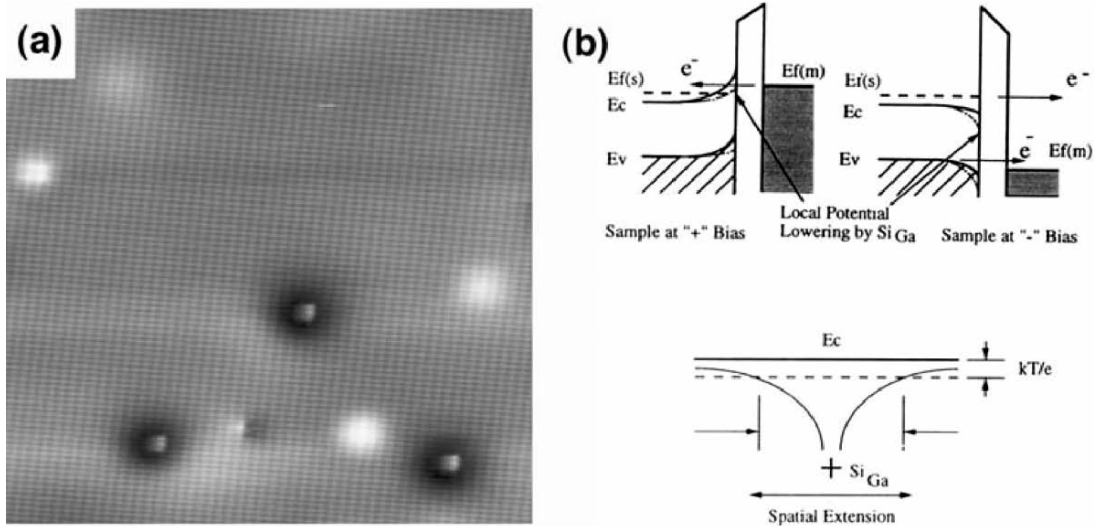


Figure 15–35. Imaging of individual dopants by X-STM. (a) Filled-state constant-current STM image ($V = -1.6$ V) of cleaved GaAs(110). Individual Si_{Ga} substitutional donors appear as protrusions of different apparent height. (b) Band diagram illustrating the electron tunneling process between a metal tip and the GaAs(110) surface in the presence of tip-induced band bending. The Coulomb potential of a donor ion locally alters the band bending, causing an increase in tunneling current above the donor. (Reprinted with permission from Zheng et al., © 1994 by the American Physical Society.)

pronounced local deformation of the cleavage plane due to the relaxation of the strained dots and matrix (Davies et al., 2002). Conversely, when combined with modeling, the elastic relaxation at the free cleavage surface itself can be used to quantitatively determine the strain field in and around individual dots.

Apart from the imaging of buried semiconductor quantum structures, an application of X-STM that has been receiving increasing attention is the mapping of electronic dopants for electronics (Figure 15–35) and, more recently, of magnetic dopants for spintronics applications. Buried substitutional Zn and Be acceptors in *p*-GaAs have been imaged as protrusions in filled-state constant-current STM via a contrast mechanism assigned to an increased local state density directly above a dopant (Johnson et al., 1993). Si donors at Ga sites (Si_{Ga}) in GaAs have been imaged on GaAs(110) as protrusions, a few nanometers in size and with discrete values of apparent height, superimposed on the background lattice. The observed contrast in filled-state constant current STM was attributed to a local perturbation of the near-surface band bending by the Coulomb potential of the Si_{Ga}. The discrete apparent heights were interpreted as a consequence of a distribution of the donor atoms over five subsurface layers beneath the cleavage surface (Zheng et al., 1994).

Deep acceptor states due to magnetic impurities in III–V semiconductors, such as substitutional Mn_{Ga} in GaAs, are expected to play an important role in the hole-mediated coupling between magnetic impu-

rities, and thus determine the magnetic properties of hole-mediated ferromagnetic semiconductors such as $\text{Ga}_{1-x}\text{Mn}_x\text{As}$, important for emerging spintronics applications. X-STM at room temperature was used to map the wave function of the hole bound to an individual Mn acceptor in GaAs (Yakunin et al., 2004). Via bias-induced changes to the local band bending, the acceptor could be imaged in both the neutral ($U_s = +0.6\text{ V}$) and ionized state ($U_s = -0.7\text{ V}$). The acceptor ground state has a highly anisotropic structure due to a significant contribution of d-wave envelope functions, which is well-reproduced by simulated images based on a tight-binding model of the Mn acceptor structure (Tang and Flatté, 2004).

6 STM Image Simulation

Although STM is a very powerful experimental technique in its own right, geometric and spectroscopic contrast tend to be difficult to separate, as discussed previously. In many cases, the unequivocal identification of surface structures or of adsorption geometries requires a comparison of experimental STM images with contrast simulations. The conventional approach to STM image simulations follows a two-step process. Relaxed atomic positions of candidate surface structures are calculated by *ab initio* theoretical methods, such as density functional theory. The actual STM contrast is then computed using the Tersoff–Hamann theory of STM (Tersoff and Hamann, 1983, 1985). By combining bias-dependent atomic-resolution STM images with simulated images for various candidate structures, even completely unknown surface structures can in principle be “solved” on the basis of STM imaging alone. The difficulty lies less with the microscopy than with the identification of plausible candidate surface structures. In the past, this key step has typically been approached intuitively, i.e., by guessing structures based on minimizing the density of dangling bonds, starting from a truncated bulk structure. Recently developed systematic techniques, based on stochastic optimization, for generating comprehensive sets of candidate structures promise to become a powerful tool for solving a wide range of surface structures by STM imaging and image simulation (Ciobanu and Predescu, 2004).

The Tersoff–Hamann theory, the basis of most STM image simulations to date, provides a transfer matrix formalism to calculate the tunneling matrix element and tunneling current for realistic configurations of tip and sample. For values of tip–sample separation typical in STM imaging ($\sim 1\text{ nm}$), the coupling between tip and sample wave functions is weak, and the tunneling process can be treated by first-order perturbation theory. Within this framework the tunneling current is given by

$$\begin{aligned}
 I &= \frac{2\pi e}{\hbar} \sum_{\mu, \nu} [f(E_\mu) - f(E_\nu)] |M_{\mu\nu}| \delta(E_\nu + V - E_\mu) \\
 &\approx \frac{2\pi}{\hbar} e^2 V \sum_{\mu, \nu} |M_{\mu\nu}|^2 \delta(E_\mu - E_F) \delta(E_\nu - E_F)
 \end{aligned} \tag{7}$$

where $f(E)$ is the Fermi function, V denotes the tunneling bias, $M_{\mu\nu}$ the tunneling matrix element between a tip state ψ_μ and a sample state ψ_ν , and $E_{\mu(\nu)}$ the energy of state $\psi_{\mu(\nu)}$. In the second line, the Fermi function was replaced by a step function (zero- T approximation), and the limit of low tunneling bias has been assumed. If, after Bardeen (1961), the tunneling matrix element is written as a surface integral, and the sample and tip wavefunctions are expanded in plane waves, one obtains

$$M_{\mu\nu} = -\frac{4\pi^2\hbar^2}{m} \int d\mathbf{q} a_{\mathbf{q}} b_{\mathbf{q}}^* e^{-\kappa_{\mathbf{q}} z_t} e^{i\mathbf{q} \cdot \mathbf{x}_t} \quad (8)$$

Here $\mathbf{q}$ is the Fourier wavevector and $a_{\mathbf{q}}$, $b_{\mathbf{q}}$ are expansion coefficients in the plane wave expansion of the sample and tip wavefunctions, x_t and z_t denote the lateral and vertical tip position, and $\kappa_{\mathbf{q}}$ is the decay constant. Given the wavefunctions of sample and tip, this expression provides the tunneling matrix element, from which the tunneling current can be calculated via Eq. (7).

In practice, a difficulty arises from the fact that the atomic-scale geometry and chemical composition of the tip are unknown. Thus, assumptions have to be made as to the tip wavefunctions (i.e., $b_{\mathbf{q}}$). For an s-wavefunction of the tip (as for an ideal “point”-tip), the tunneling current is

$$I \propto \sum |\psi_\nu(\mathbf{r}_t)|^2 \delta(E_\nu - E_F) \equiv \rho(\mathbf{r}_t, E_F) \quad (9)$$

i.e., is proportional to the local density of states of the sample at the Fermi energy, a property of the sample surface alone! For finite tunneling bias, V , the tunneling current can be written as an integral over the energy range between the Fermi energy E_F and $E_F + V$:

$$I = \int_0^{eV} \rho_t(E) \rho_s(E - eV) T(E, eV) dE \quad (10)$$

and the tunneling conductance at bias V is proportional to the position-dependent density of states of the sample at energy eV from the Fermi energy

$$\left(\frac{dI}{dV} \right) \bigg|_V \approx \rho_s(eV) \rho_t(0) T(eV, V) \quad (11)$$

Contrast calculations thus involve the computation of the position- and energy-dependent sample DOS to obtain simulated maps of tunneling conductance $dI/dV(x, y)$ or tunneling current $I(x, y)$.

Although originally derived for small voltages, as used for STM imaging on metals, with this extension the Tersoff–Hamann formalism can be applied to simulate images at higher bias as well, and has proven a powerful simulation tools for a wide range of imaging scenarios. Tromp et al. (1986) were the first to show that it could be applied to semiconductor surfaces, as demonstrated by their successful contrast simulation for the Si(111)-(7 × 7) reconstruction. The systems simulated since then include surface structures of semiconductors (Fujikawa et al., 2002; Klijn et al., 2003), ultrathin insulators (Olsson et al., 2005),

as well as adsorbed atoms (Repp et al., 2004) and molecules (Olsson et al., 2003; Kühnle et al., 2002).

Figures 15–36 and 15–37 illustrate two of the numerous examples of this approach in the literature.

Shallow Ge quantum dots self-assembled during heteroepitaxy on Si(001) invariably have surface termination by (105) facets, with surface normal just 11° away from [001]. These (105) facets appear extremely stable, likely due to a low density of dangling bonds and additional strain stabilization by surface strain compensating for some of the 4% lattice mismatch strain between the Ge overlayer and the Si substrate. To calculate the surface energy, and thus explain the stability of the facet, a detailed knowledge of the surface structure is necessary. Dual bias constant-current STM combined with STM contrast calculations was used to identify a best match with one of two proposed structures of the (105) surface: “paired dimer” (Structure A) and “rebonded step” (Structure B) (Fujikawa et al., 2002). The STM contrast is clearly identified as that of a “rebonded step” structure, the structure that not only has the lowest dangling bond density but also causes tensile surface strain, key to the strain stabilization of the (105) facet.

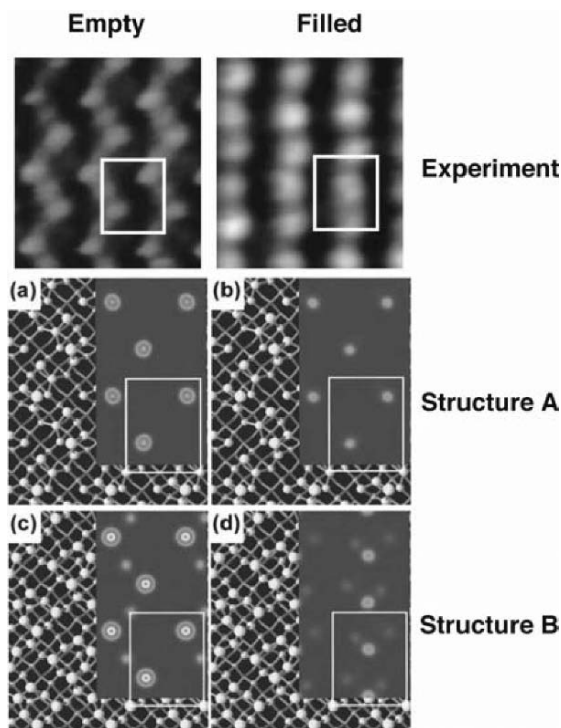


Figure 15–36. Identification of the surface structure of Ge(105). Empty- and filled-state constant-current STM images are compared with two candidate structures, the paired dimer Structure A and rebonded step Structure B (c and d). A Tersoff–Hamann calculation of the STM contrast for these two structures clearly shows that the experimental STM images arise from the rebonded step structure. (Reprinted with permission from Fujikawa et al., © 2002 by the American Physical Society.)

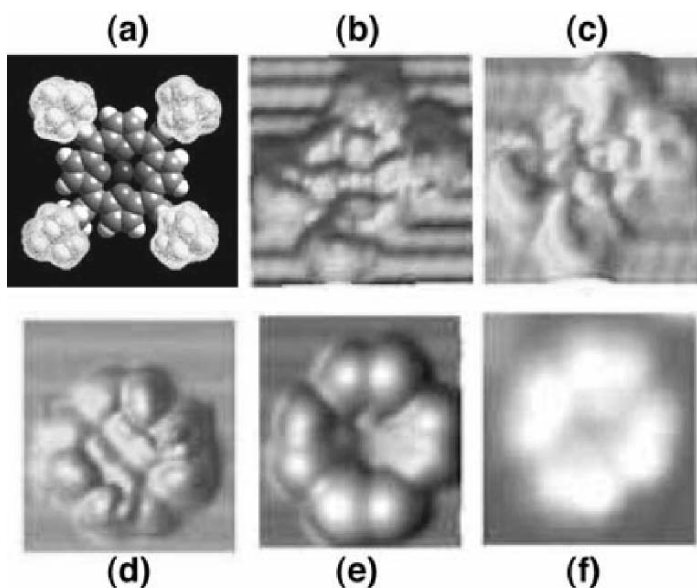


Figure 15–37. Identification of the molecular conformation of Cu-DTBPP on Cu(211). (a) Molecular model of Cu-DTBPP, with the four-lobed pattern observed by STM marked in yellow. (b–e) STM contrast calculation for different angles between the four legs and the substrate: (b) 60°, (c) 45°, (d) 30°, and (e) 10°. (f) Experimental STM image of the molecule. (Reprinted from Moresco et al., ©2002 with permission from Elsevier.) (See color plate.)

Simulations of STM contrast of adsorbed molecules pose a somewhat different problem than simulations of surface structures. For solid surfaces, the combination of STM imaging with image simulation can discriminate between several candidate structures, as discussed in the previous example. For molecules, the structure is typically well known. However, the adsorption geometry (site, orientation) as well as the conformation for larger, more flexible molecules enter as unknowns into the simulation. A relatively simple example is the stiff molecular cage structure of fullerene C_{60} (Hou et al., 1999; Pascual et al., 2000; Lu et al., 2003). In this case, the structure and conformation are well known, and the only parameter of the contrast calculation is the orientation. A more complex system, Cu-tetra(3,5-di-*tert* butyl phenyl) porphyrin (Cu-DTBPP) on Cu(211), is illustrated in Figure 15–37.

Cu-DTBPP consists of a central porphyrin ring with four symmetrically attached di-*tert*-butyl phenyl (DTBP) groups (Figure 15–37a). These four bulky groups determine the shape of the molecule, which is imaged in STM as a four-leaf clover structure, and define the interaction with the metal substrate (Jung et al., 1996). The orientation of the side groups is the main parameter that needs to be optimized to determine the conformation of the adsorbed molecule, and to simulate the experimental STM images. Figure 15–37 shows several candidate conformations, with leg angles of 60° (b), 45° (c), 30° (d), and the optimized angle of 10° relative to the substrate (e), which was found to best reproduce the experimental STM image of the molecule.

Acknowledgments. The compilation of this chapter was supported by the U.S. Department of Energy under Contract No. DE-AC02-98CH10886.

References

- Altfeder, I.B., Matveev, K.A. and Chen, D.M. (1997). *Phys. Rev. Lett.* **78**, 2815.
- Altfeder, I.B., Chen, D.M. and Matveev, K.A. (1998). *Phys. Rev. Lett.* **80**, 4895.
- Bardeen, J. (1961). *Phys. Rev. Lett.* **6**, 57.
- Bauer, E. (1994). *Rep. Prog. Phys.* **57**, 895.
- Bauer, E. (1998). *Surf. Rev. Lett.* **5**, 1275.
- Becker, R.S., Swartzentruber, B.S., Vickers, J.S. and Klitsner, T. (1989). *Phys. Rev. B* **39**, 1633.
- Bell, L.D., Hecht, M.H., Kaiser, W.J. and Davis, L.C. (1990). *Phys. Rev. Lett.* **64**, 2679.
- Bell, L.D., Kaiser, W.J., Hecht, M.H. and Davis, L.C. (1991). *J. Vac. Sci. Technol. B* **9**, 594.
- Binnig, G. and Rohrer, H. (1983). *Surf. Sci.* **126**, 236.
- Binnig, G., Rohrer, H., Gerber, Ch. and Weibel, E. (1982). *Appl. Phys. Lett.* **40**, 178.
- Binnig, G., Rohrer, H., Gerber, Ch. and Weibel, E. (1983). *Phys. Rev. Lett.* **50**, 120.
- Binnig, G., Garcia, N. and Rohrer, H. (1985). *Phys. Rev. B* **32**, 1336.
- Binnig, G. and Smith, D.P.E. (1986). *Rev. Sci. Instrum.* **57**, 1688.
- Bode, M., Getzlaff, M. and Wiesendanger, R. (1998). *Phys. Rev. Lett.* **81**, 4256.
- Boland, J.J. (1993). *Science* **262**, 1703.
- Boland, J.J. and Weaver, J.H. (1998). *Phys. Today* **51**, 34.
- Bonnell, D.A., Ed. (2001). *Scanning Probe Microscopy and Spectroscopy—Theory, Techniques, and Applications*, 2nd ed., Wiley-VCH, New York.
- Borovski, B., Krueger, M. and Ganz, E. (1997). *Phys. Rev. Lett.* **78**, 4229.
- Borovski, B., Krueger, M. and Ganz, E. (1998). *Phys. Rev. B* **57**, 4269.
- Brookes, I.M., Muryn, C.A. and Thornton, G. (2001). *Phys. Rev. Lett.* **87**, 266103.
- Bruls, D.M., Vugs, J.W.A.M., Roenraad, P.M., Salemink, H.W.M., Wolter, J.H., Hopkinson, M., Skolnick, M.S., Long, F. and Gill, S.P.A. (2002). *Appl. Phys.* **81**, 1708.
- Cerezo, A., Larson, D.J. and Smith, G.D.W. (2001). *MRS Bull.* **26**, 102.
- Chambliss, D.D., Wilson, R.J. and Chiang, S. (1995). *IBM J. Res. Dev.* **39**, 639.
- Chen, C.J. (1993). *Introduction to Scanning Tunneling Microscopy* (Oxford University Press, Oxford).
- Chen, Y., Ohlberg, D.A.A., Medeiros-Ribeiro, G., Chang, Y.A. and Williams, R.S. (2000). *Appl. Phys. Lett.* **76**, 4004.
- Ciobanu, C.V. and Predescu, C. (2004). *Phys. Rev. B* **70**, 085321.
- Crain, J.N. and Pierce, D.T. (2005). *Science* **307**, 703.
- Crommie, M.F., Lutz, C.P. and Eigler, D.M. (1993). *Nature* **363**, 524.
- Cuberes, M.T., Bauer, A., Wen, H.J., Prietsch, M. and Kaindl, G. (1994). *J. Vac. Sci. Technol. B* **12**, 2646.
- Cygan, M.T., Dunbar, T.D., Arnold, J.J., Bumm, L.A., Shedlock, N.F., Burgin, T.P., Jones, L., Allara, D.L., Tour, J.M. and Weiss, P.S. (1998). *J. Am. Chem. Soc.* **120**, 2721.
- Davies, J.H., Bruls, D.M., Vugs, J.W.A.M. and Koenraad, P.M. (2002). *J. Appl. Phys.* **91**, 4171.
- Diebold, U. (2003). *Surf. Sci. Rep.* **48**, 53.
- Ebert, Ph., Cox, G., Poppe, U. and Urban, K. (1992). *Surf. Sci.* **271**, 587.

- Eder, C., Smoliner, J. and Strasser, G. (1996). *Appl. Phys. Lett.* **68**, 2876.
- Eigler, D.M. and Schweizer, E.K. (1990). *Nature* **344**, 524.
- Erwin, S.C., Baski, A.A. and Whitman, L.J. (1996). *Phys. Rev. Lett.* **77**, 687.
- Feenstra, R.M., Stroscio, J.A., Tersoff, J. and Fein, A.P. (1987). *Phys. Rev. Lett.* **58**, 1192.
- Feenstra, R.M., Collins, D.A., Ting, D.Z.-Y., Wang, M.W. and McGill, T.C. (1994). *Phys. Rev. Lett.* **72**, 2749.
- Festy, F. and Palmer, R.E. (2004). *Appl. Phys. Lett.* **85**, 5034.
- Fowell, A.E., Williams, R.H., Richardson, B.E. and Shen, T.-H. (1990). *Semicond. Sci. Technol.* **5**, 348.
- Fujikawa, Y., Akiyama, K., Nagao, T., Sakurai, T., Lagally, M.G., Hashimoto, T., Morikawa, Y. and Terakura, K. (2002). *Phys. Rev. Lett.* **88**, 176101.
- Gai, Z., Li, X., Zhao, R.G. and Yang, W.S. (1998). *Phys. Rev. B* **57**, 15060.
- Gewirth, A.A. and Niece, B.K. (1997). *Chem. Rev.* **97**, 1129.
- Gimzewski, J.K., Joachim, C., Schlittler, R.R., Langlais, V., Tang, H. and Johannsen, I. (1998). *Science* **281**, 531.
- Goldman, R.S., Feenstra, R.M., Briner, B.G., O'Steen, M.L. and Hauenstein, R. J. (1996). *Appl. Phys. Lett.* **69**, 3698.
- Gong, Q., Offermans, P., Nötzel, R., Koenraad, P.M. and Loolter, J.H. (2004). *Appl. Phys.* **85**, 5697.
- Grant, M.L., Swartzentruber, B.S., Bartelt, N.C. and Hannon, J.B. (2001). *Phys. Rev. Lett.* **86**, 4588.
- Gregory, S. (1990). *Phys. Rev. Lett.* **64**, 689.
- Güntherodt, H.-J. and Wiesendanger, R., Eds. (1992). *Scanning Tunneling Microscopy Vol. I & II* (Springer, Berlin).
- Hamers, R.J. (1989). *Annu. Rev. Phys. Chem.* 531.
- Hamers, R.J., Tromp, R.M. and Demuth, J.E. (1986). *Phys. Rev. Lett.* **56**, 1972.
- Hanaguri, T., Lupien, C., Kohsaka, Y., Lee, D.-H., Azuma, M., Takano, M., Takagi, H. and Davis, J.C. (2004). *Nature* **430**, 1001.
- Hasegawa, Y. and Avouris, Ph. (1993). *Phys. Rev. Lett.* **71**, 1071.
- Heer, R., Smoliner, J., Strasser, G. and Gornik, E. (1998). *Appl. Phys. Lett.* **73**, 3138.
- Heinrich, A.J., Lutz, C.P., Gupta, J.A. and Eigler, D.M. (2002). *Science* **298**, 1381.
- Heinze, S., Bode, M., Kubetzka, A., Pietzsch, O., Nie, X., Blügel, S. and Wiesendanger, R. (2000). *Science* **288**, 1805.
- Hendriksen, B.L.M. and Frenken, J.W.M. (2002). *Phys. Rev. Lett.* **89**, 046101.
- Hla, S.-W. and Rieder, K.-H. (2003). *Annu. Rev. Phys. Chem.* **54**, 307.
- Hoffman, J.E., McElroy, K., Lee, D.-H., Lang, K.M., Eisaki, H., Uchida, S. and Davis, J.C. (2002a). *Science* **297**, 1148.
- Hoffman, J.E., Hudson, E.W., Lang, K.M., Madhavan, V., Eisaki, H., Uchida, S. and Davis, J.C. (2002b). *Science* **295**, 466.
- Horch, S., Lorensen, H.T., Helveg, S., Lægsgaard, E., Stensgaard, I., Jacobsen, K.W., Nørskov, J.K. and Besenbacher, F. (1999). *Nature* **398**, 134.
- Hornbaker, D.J., Kahng, S.-J., Misra, S., Smith, B.W., Johnson, A.T., Mele, E.J., Luzzi, D.E. and Yazdani, A. (2002). *Science* **295**, 828.
- Horn-von Hoegen, M. (1994). *Appl. Phys. A* **59**, 503.
- Hou, J.G., Jinlong, Y., Haiqian, W., Qunxiang, L., Changgan, Z., Hai, L., Wang, B., Chen, D.M. and Qinshi, Z. (1999). *Phys. Rev. Lett.* **83**, 3001.
- Hovis, J.S. and Hamers, R.J. (1997). *Surf. Sci.* **402**, 1.
- Jaklevic, R.C. and Lambe, J. (1966). *Phys. Rev. Lett.* **17**, 1139.
- Jaklevic, R.C., Lambe, J., Mikkor, M. and Vassell, W.C. (1971). *Phys. Rev. Lett.* **26**, 88.
- Jensen, J.A., Rider, K.B., Chen, Y., Salmeron, M. and Somorjai, G.A. (1999). *J. Vac. Sci. Technol. B* **17**, 1080.

- Johnson, M.B., Albrechtsen, O., Feenstra, R.M. and Salemink, H.W.M. (1993). *Appl. Phys. Lett.* **63**, 2923.
- Jung, T.A., Schlittler, R.R., Gimzewski, J.K., Tang, H. and Joachim, C. (1996). *Science* **271**, 181.
- Kaiser, W.J. and Bell, L.D. (1988). *Phys. Rev. Lett.* **60**, 1406.
- Kasaya, M., Tabata, H. and Kawai, T. (1998). *Surf. Sci.* **400**, 367.
- Kawamura, M., Paul, N., Cherepanov, V. and Voigtländer, B. (2003). *Phys. Rev. Lett.* **91**, 096102.
- Klijn, J., Sacharov, L., Meyer, C., Blügel, S., Morgenstern, M. and Wiesendanger, R. (2003). *Phys. Rev. B* **68**, 205327.
- Kochanski, G.P. (1989). *Phys. Rev. Lett.* **62**, 2285.
- Komeda, T., Kim, Y., Kawai, M., Persson, B.N.J. and Ueba, H. (2002). *Science* **295**, 2055.
- Kubby, J.A. and Greene, W.J. (1992). *Phys. Rev. Lett.* **68**, 329.
- Kühnle, A., Linderöth, T.R., Hammer, B. and Besenbacher, F. (2002). *Nature* **415**, 891.
- Kuipers, L., Hoogeman, M.S. and Frenken, J.W.M. (1993). *Phys. Rev. Lett.* **71**, 3517.
- Kuk, Y. and Silverman, P.J. (1986). *Appl. Phys. Lett.* **48**, 1597.
- Lægsgaard, E., Österlund, L., Rasmussen, P.B., Stensgaard, I. and Besenbacher, F. (2001). *Rev. Sci. Instrum.* **72**, 3537.
- Laracuenta, A. and Whitman, L.J. (2001). *Surf. Sci.* **476**, L247.
- Lee, E.Y., Sirringhaus, H. and von Känel, H. (1994). *Phys. Rev. B* **50**, 5807.
- Lee, H.J. and Ho, W. (1999). *Science* **286**, 1719.
- Legrand, B., Grandidier, B., Nys, J.P., Stiévenard, D., Gérard, J.M. and Thierry-Mieg, V. (1998). *Appl. Phys. Lett.* **73**, 96.
- Li, L., Tindall, C., Takaoka, O., Hasegawa, Y. and Sakurai, T. (1997). *Phys. Rev. B* **56**, 4648.
- Lopinski, G.P., Moffatt, D.J., Wayner, D.D.M. and Wolkow, R.A. (1998). *Nature* **392**, 909.
- Lu, X., Grobis, M., Khoo, K.H., Louie, S.G. and Crommie, M.F. (2003). *Phys. Rev. Lett.* **90**, 096802.
- Mayne, A.J., Avery, A.R., Knall, J., Jones, T.S., Briggs, G.A.D. and Weinberg, W.H. (1993). *Surf. Sci.* **284**, 247.
- Medeiros-Ribeiro, G., Bratkovski, A.M., Kamins, T.I., Ohlberg, D.A.A. and Williams, R.S. (1998). *Science* **279**, 353.
- Meinel, K., Wolter, H., Ammer, Ch., Beckmann, A. and Neddermeyer, H. (1997). *J. Phys. Condens. Matter* **9**, 4611.
- Men, F.K., Packard, W.E. and Webb, M.B. (1988). *Phys. Rev. Lett.* **61**, 2469.
- Meyer, G. and Rieder, K.-H. (1997). *Surf. Sci.* **377–379**, 1087.
- Meyer, G. and Rieder, K.-H. (1998). *MRS Bull.* **23**, 28.
- Mitsui, T., Rose, M.K., Fomin, E., Ogletree, D.F. and Salmeron, M. (2002). *Science* **297**, 1850.
- Mo, Y.-W., Swartzentruber, B.S., Kariotis, R., Webb, M.B. and Lagally, M.G. (1989). *Phys. Rev. Lett.* **63**, 2393.
- Mo, Y.-W., Savage, D.E., Swartzentruber, B.S. and Lagally, M.G. (1990). *Phys. Rev. Lett.* **65**, 1020.
- Mo, Y.-W., Kleiner, J., Webb, M.B. and Lagally, M.G. (1991). *Phys. Rev. Lett.* **66**, 1998.
- Moresco, F., Meyer, G., Rieder, K.-H., Ping, J., Tang, H. and Joachim, C. (2002). *Surf. Sci.* **499**, 94.
- Muller, E.W. and Tsong, T.T. (1969). *Field Ion Microscopy* (Elsevier, New York).
- Nagl, C., Haller, O., Platzgummer, E., Schmid, M. and Varga, P. (1994). *Surf. Sci.* **321**, 237.

- Nakayama, K. and Weaver, J.H. (1999). *Phys. Rev. Lett.* **82**, 980.
- Narayanamurti, V. and Kozhevnikov, M. (2001). *Phys. Rep.* **349**, 447.
- Nazin, G.V., Qiu, X.H. and Ho, W. (2003). *Science* **302**, 77.
- Nunes, G., Jr. and Amer, N.M. (1993). *Appl. Phys. Lett.* **63**, 1851.
- Olsson, F.E., Lorente, N. and Persson, M. (2003). *Surf. Sci.* **522**, L27.
- Olsson, F.E., Persson, M., Repp, J. and Meyer, G. (2005). *Phys. Rev. B* **71**, 075419.
- Pascual, J.I., Gomez-Herrero, J., Rogero, C., Baro, A.M., Sanchez-Portal, D., Artacho, E., Ordejon, P. and Soler, J.M. (2000). *Chem. Phys. Lett.* **321**, 78.
- Pfister, M., Johnson, M.B., Alvarado, S.F., Salemink, H.W.M., Marti, U., Martin, D., Morier-Genoud, F. and Reinhard, F.K. (1995). *Appl. Phys. Lett.* **67**, 1459.
- Pietzsch, O., Kubetzka, A., Bode, M. and Wiesendanger, R. (2001). *Science* **292**, 2053.
- Prietsch, M. and Ludeke, R. (1991). *Phys. Rev. Lett.* **66**, 2511.
- Qin, X.R. and Lagally, M.G. (1997). *Science* **278**, 1444.
- Qin, X.R., Swartzentruber, B.S. and Lagally, M.G. (2000). *Phys. Rev. Lett.* **85**, 3660.
- Rasmussen, P.B., Hendriksen, B.L.M., Zeijlemaker, H., Ficke, H.G. and Frenken, J.W.M. (1998). *Rev. Sci. Instrum.* **69**, 3879.
- Repp, J., Meyer, G., Olsson, F.E. and Persson, M. (2004). *Science* **305**, 493.
- Repp, J., Meyer, G., Stojkovic, S.M., Gourdon, A. and Joachim, C. (2005). *Phys. Rev. Lett.* **94**, 026803.
- Rippard, W.H. and Buhrmann, R.A. (2000). *Phys. Rev. Lett.* **84**, 971.
- Robinson, K.M., Robinson, I.K. and O'Grady, W.E. (1992). *Surf. Sci.* **262**, 387.
- Rößler, M., Geng, P. and Wintterlin, J. (2005). *Rev. Sci. Instrum.* **76**, 023705.
- Rost, M.J., Crama, L., Schakel, P., van Tol, E., van Velzen-Williams, G.B.E.M., Overgaw, C.F., ter Horst, H., Dekker, H., Okhuijsen, B., Seynen, M., Vijftigschild, A., Han, P., Katan, A.J., Schoots, K., Schumm, R., van Loo, W., Oosterkamp, T.H. and Frenken, J.W.M. (2005). *Rev. Sci. Instrum.* **76**, 053710.
- Rubin, M.E., Medeiros-Ribeiro, G., O'Shea, J.J., Chin, M.A., Lee, E.Y., Petroff, P.M. and Narayanamurti, V. (1996). *Phys. Rev. Lett.* **77**, 5268.
- Sajoto, T., O'Shea, J.J., Bhargava, S., Leonard, D., Chin, M.A. and Narayanamurti, V. (1995). *Phys. Rev. Lett.* **74**, 3427.
- Salemink, H. and Albrechtsen, O. (1991). *J. Vac. Sci. Technol. B* **9**, 779.
- Sanders, L.M., Stumpf, R., Mattson, T.R. and Swartzentruber, B.S. (2003). *Phys. Rev. Lett.* **91**, 206104.
- Schaub, R., Thostrup, P., Lopez, N., Stensgaard, I., Nørskov, J.K. and Besenbacher, F. (2001). *Phys. Rev. Lett.* **87**, 266104.
- Schaub, R., Wahlström, E., Rønnau, A., Lægsgaard, E., Stensgaard, I. and Besenbacher, F. (2003). *Science* **299**, 377.
- Schintke, S. and Schneider, W.-D. (2004). *J. Phys: Cond. Matter* **16**, R49.
- Schunack, M., Linderroth, T.R., Rosei, F., Lægsgaard, E., Stensgaard, I. and Besenbacher, F. (2002). *Phys. Rev. Lett.* **88**, 156102.
- Sirringhaus, H., Lee, E.Y. and von Känel, H. (1994). *Phys. Rev. Lett.* **73**, 577.
- Smith, D.P.E. and Binnig, G. (1986). *Rev. Sci. Instrum.* **57**, 2630.
- Stipe, B.C., Rezaei, M.A. and Ho, W. (1998a). *Science* **280**, 1732.
- Stipe, B.C., Rezaei, M.A. and Ho, W. (1998b). *Science* **279**, 1907.
- Stroscio, J.A. and Celotta, R.J. (2004). *Science* **306**, 242.
- Sutter, P., Schick, I., Ernst, W. and Sutter, E. (2003a). *Phys. Rev. Lett.* **91**, 176102.
- Sutter, P., Zahl, P., Sutter, E. and Bernard, J.E. (2003b). *Phys. Rev. Lett.* **90**, 166101.
- Sutter, E., Sutter, P. and Bernard, J.E. (2004). *Appl. Phys. Lett.* **84**, 2262.

- Swartzentruber, B.S. (1996). *Phys. Rev. Lett.* **76**, 459.
- Swartzentruber, B.S., Mo, Y.-W., Kariotis, R., Lagally, M.G. and Webb, M.B. (1990). *Phys. Rev. Lett.* **65**, 1913.
- Takayanagi, K., Tanishiro, Y., Takahashi, S. and Takahashi, M. (1985). *Surf. Sci.* **164**, 367.
- Tang, J.-M. and Flatté, M.E. (2004). *Phys. Rev. Lett.* **92**, 047201.
- Tersoff, J. and Hamann, D.R. (1983). *Phys. Rev. Lett.* **50**, 1998.
- Tersoff, J. and Hamann, D.R. (1985). *Phys. Rev. B* **31**, 805.
- Troadec, C., Kunardi, L. and Chandrasekhar, N. (2005). *Appl. Phys. Lett.* **86**, 72101.
- Tromp, R.M. (2000). *IBM J. Res. Dev.* **44**, 503.
- Tromp, R.M., Hamers, R.J. and Demuth, J.E. (1985). *Phys. Rev. Lett.* **55**, 1303.
- Tromp, R.M., Hamers, R.J. and Demuth, J.E. (1986). *Phys. Rev. B* **34**, 1388.
- Venables, J.A. (2000). *Introduction to Surface and Thin Film Processes* (Cambridge University Press, Cambridge).
- Vershinin, M., Misra, S., Ono, S., Abe, Y., Ando, Y. and Yazdani, A. (2004). *Science* **303**, 1995.
- Voigtländer, B. (2001). *Surf. Sci. Rep.* **43**, 127.
- Voigtländer, B. and Zinner, A. (1993). *Appl. Phys. Lett.* **63**, 3055.
- Wang, Y., Chen, X. and Hamers, R.J. (1994a). *Phys. Rev. B* **50**, 4534.
- Wang, Y., Bronikowski, M.J. and Hamers, R.J. (1994b). *Surf. Sci.* **311**, 64.
- Weierstall, U. and Spence, J. (1998). *Surf. Sci.* **398**, 267.
- Wiesendanger, R. (1994). *Scanning Probe Microscopy and Spectroscopy—Methods and Applications* (Cambridge University Press, Cambridge).
- Wintterlin, J., Trost, J., Renisch, S., Schuster, R., Zambelli, T. and Ertl, G. (1997). *Surf. Sci.* **394**, 159.
- Wittneven, Chr., Dombrowski, R., Morgenstern, M. and Wiesendanger, R. (1998). *Phys. Rev. Lett.* **81**, 5616.
- Wolkow, R.A. (1999). *Annu. Rev. Phys. Chem.* **50**, 413.
- Wouda, P.T., Nieuwenhuys, B.E., Schmid, M. and Varga, P. (1996). *Surf. Sci.* **359**, 17.
- Wouda, P.T., Schmid, M., Nieuwenhuys, B.E. and Varga, P. (1998). *Surf. Sci.* **417**, 292.
- Yagi, K. (1982). *Scan. Electron Microsc.* **4**, 1421.
- Yakunin, A.M., Silov, A.Yu., Koenraad, P.M., Wolter, J.H., Van Roy, W., De Boeck, J., Tang, J.-M. and Flatté, M.E. (2004). *Phys. Rev. Lett.* **92**, 216806.
- Yang, Y.-N., Fu, E.S. and Williams, E.D. (1996). *Surf. Sci.* **356**, 101.
- Zhang, Z.Y. and Lagally, M.G. (1998). *Science* **276**, 377.
- Zheng, J.F., Liu, X., Newman, N., Weber, E.R., Ogletree, D.F. and Salmeron, M. (1994). *Phys. Rev. Lett.* **72**, 1490.

Atomic Force Microscopy in the Life Sciences

Matthias Amrein

He sees the face and the moving hands, even hears it ticking. If he is ingenious he may form some picture of a mechanism, which could be responsible for all the things he observes, but he may never be quite sure his picture is the only one that could explain his observations. He will never be able to compare his picture with the real mechanism and he cannot even imagine the possibility of the meaning of such a comparison.

Albert Einstein, describing a man trying to understand a closed watch.

1 Introduction

Cells are highly organized in compartments and functional units down to the macromolecular level. The frameworks of these functional and structural units are either protein complexes or complexes of proteins with nucleic acids or with lipids. The recently completed map of the human genome and the systematic mapping of the proteins expressed in tissues and cells (*proteomics*) are part of a concerted effort to rapidly advance the understanding of the functions of macromolecular units and the cell. But proteomics reveals only the basic inventory of a cell and the inventory is insufficient to explain the function of each element and the orchestration of the components. As with Einstein's image of the closed watch, understanding life is inconceivable without observing the structures behind function. Microscopy plays an important role by placing the molecular elements into a structural context. Because the pace of discovery of these elements has been increased substantially by proteomics, the need for more sophisticated microscopy has also substantially increased in recent years. Atomic force microscopy (AFM) plays a specific role in life science microscopy by allowing imaging to be combined with locally probing functions of macromolecular elements.

Most microscopes depend on radiation, emitted and recorded at a distance from the sample, using lenses. The resolution power of these microscopes in the life sciences is limited by diffraction and/or damage to the sample by the illuminating beam. In contrast, the scanning tunneling microscope (STM) has been the first microscope to rely on an effect only present in the immediate vicinity of a physical probe and the sample. In an STM, an electrically conductive needle approaches the sample until

current flows. The current is based on a quantum-mechanical tunneling effect and flows before the tip and sample physically touch. During scanning, the tip–sample distance remains constant by keeping the tunneling current constant. The movement of the tip relative to the sample reflects the sample topography and is recorded. Not only has the atomic surface lattice structure been revealed for many crystalline samples, but also single atom defects imaged directly. The STM immediately created great interest among biologists because of its outstanding resolution power and the absence of radiation damage (Amrein et al., 1989, 1993; Baró et al., 1985; Voelker et al., 1988; Guckenberger et al., 1989; Stemmer et al., 1989; Welland et al., 1989; Miles et al., 1990; Arscott and Bloomfield, 1993; Lindsay and Tao, 1993; Miles and McMaster, 1993). But the application of STM in the life sciences suffered from the poor electrical conductivity of most biological matter, and reproducible images were obtained only after metal coating the specimens (Travaglini et al., 1987; Zasadzinski et al., 1988; Blackford and Jericho, 1991; Amrein et al., 1988, 1989b, 1993). Nevertheless, early applications demonstrated the resolution power and high signal-to-noise ratio at the macromolecular level offered by the new microscopy. Although unintentional at first, it also became clear that the probe could be used to manipulate macromolecular structures individually (Figure 16–1).

The basic principles of the STM were soon extended to form a suite of new devices, the scanning probe microscopes (SPM). For all SPM, a physical probe is scanned over the sample in nanometer distance at most. Sample properties are mapped from a very small interaction volume in the near field of the sample and the probe. The meaning of the term of *near field* is different for the various types of SPM. In an AFM, short-range forces between a tip and the sample take the role of the tunneling current in the STM, thus making the new paradigm applicable to electrically nonconductive samples. This is how AFM gained ground in the life sciences.

The direct measurement of the position of the surface means that the achievable resolution is not limited by the environment—the AFM lends itself well to imaging molecular and cellular specimens under physiological conditions, in buffer. In many cases, the sample prepara-

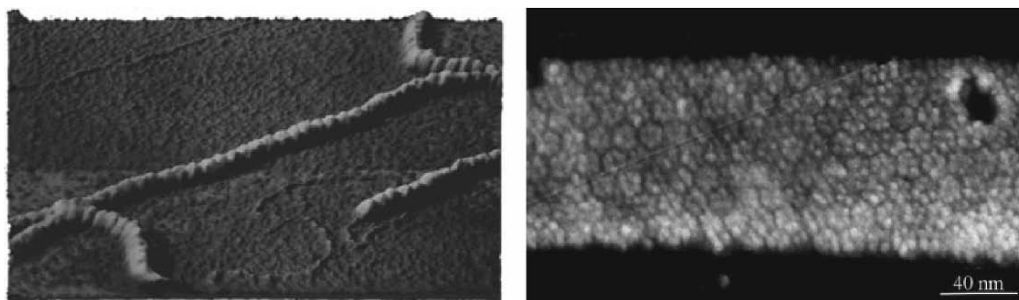


Figure 16–1. STM topographies of Rec-DNA complex (left) and an aberrant bacteriophage capsid (right). In the upper right corner, a hole has been created applying a voltage pulse between the tip and the sample. Both images immediately reveal the molecular arrangement of the respective complex structure. (Left, from Amrein et al., 1988, reproduced with permission. Right, from Amrein et al., 1998b, reproduced with permission.) (See color plate.)

tion is very straightforward, and at a basic level requires only the immobilization of the sample to be imaged. In the case of adherent cells, for instance, the sample can be imaged directly without any special preparation. The local measurement at the sample surface also allows direct quantitative measurement of dimensions and forces. In addition to the topographic image that is built up by scanning the tip over the surface, AFM can simultaneously probe the mechanical properties of the surface, by dynamically oscillating the cantilever that supports the tip. Tip-sample adhesion can also be probed, and maps of surface (visco)elasticity or adhesion built up. One application is to coat the tip with appropriate molecules, such as antibodies, other proteins, or sugars to generate maps of specific ligand-receptor binding over the surface. The AFM is also well suited to combination with other techniques—a combination with optical microscopy is often particularly worthwhile for life science research. Other techniques, such as electrochemistry or electrophysiology, can also be used simultaneously, and *in situ* probing of cell response to electrical, chemical, or mechanical stimulation is possible.

The foundation of AFM in the life sciences has now been laid by a profound understanding of the interactions between the tip and the sample and the possibility to subtly control the interactions by choosing an appropriate sensor and by adjusting the imaging environment (Weisenhorn et al., 1992; Butt et al., 1995; Heuberger et al., 1996).

To appreciate the contribution of AFM to the life sciences, it is necessary to consider it in relation to other microscopy techniques currently in use. In the realm of single molecules, the motivation for using AFM comes mainly from the high resolution achievable and the direct (real-space) measurement of samples without coating. Most of the knowledge so far gained by AFM in the life sciences comes from studying single macromolecules or macromolecular complex structures. Biological membranes, for example, can be imaged in their native state at a lateral resolution of 0.5–1 nm and a vertical resolution of 0.1–0.2 nm. Conformational changes that are related to functions can be resolved to a similar resolution. In the study of macromolecular structure, AFM competes with transmission electron microscopy. In many cases, the outcome is comparable and either microscope can be used with equal success. For example, measuring the contour length of a plasmid DNA or determining the binding site of a protein to a DNA may be performed by either microscope. However, solving the three-dimensional structure of macromolecules or macromolecular complexes that cannot be solved by X-ray crystallography or by nuclear magnetic resonance (NMR) is better accomplished by (cryo)electron microscopy, because AFM reveals only a topographical view of the structures. On the other hand, when a molecular assembly is not highly defined structurally and individual units differ from each other substantially, AFM may allow a specific question to be answered in a straightforward manner, because of its uniquely high signal-to-noise ratio at molecular dimensions.

The strength of AFM does not lie in imaging alone, but in the possibility of combining microscopy with an experiment at the molecular level. In a number of cases, molecular images have been obtained with

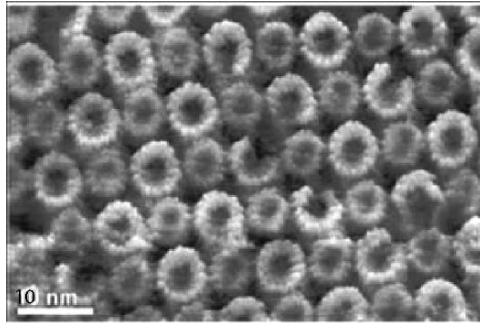


Figure 16–2. The ATP synthase is a proton-driven molecular motor with a stator (seen here) and a rotor. Simply counting the number of individual proteins inside the ring structure has not been possible from electron micrographs. (From Seelert et al., 2000, reprinted with permission.)

sufficient resolution to individually recognize single macromolecules and then address the molecules individually with the stylus of the AFM (Figure 16–2).

Most prominently, single molecule force spectroscopy combined with single molecule imaging has provided unprecedented possibilities for analyzing intramolecular and intermolecular forces including the study of the folding pathway of proteins. Probing the self-assembly of macromolecular complexes and measuring the mechanical properties of macromolecular “springs” are other examples in which AFM has made substantial contributions in the life sciences.

Imaging living cells with AFM is also performed in conjunction with local probing of the sample. The images by themselves usually are used only to obtain proper experimental control. The extreme flexibility of the cell membrane means that the images often show a combination of the cell topography with the mechanical stiffness of cytoskeletal fibers and vesicles below the cell surface. An important example for a meaningful application of AFM in living cells is the measurement of the constrained diffusion of elements of the plasma membrane. These measurements have substantially contributed to establishing the nature of lipid rafts. In other applications, the cellular response on defined mechanical stress has been studied in conjunction with hearing or with the myogenic effect (Figure 16–3).

Future applications and future instrument developments in life sciences AFM are certain to capitalize even more strongly on performing local spectroscopy or manipulation of samples at the nanoscopic scale in addition to imaging. Such experiments will certainly extend the possibilities of probing function directly in many ways. But future AFM technology might also contribute to revealing the spatial relationships of the macromolecules of cells and tissues. Today, immunohistochemistry and immunocytochemistry are mostly in use to establish these structural relationships in cells and tissues. This implies that macromolecular identity is revealed only after labeling. However, the number of different proteins and other macromolecules at work at any time in a cell makes it inconceivable to extend this approach to determine even a substantial subset of all of its macromolecular relation-

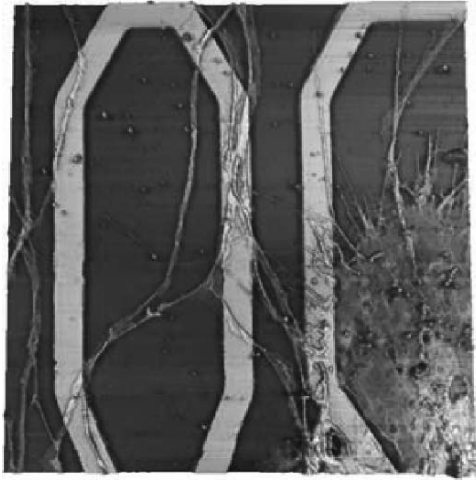


Figure 16-3. Neuronal cell, cultured on an electronic chip. The chip is designed such as to pick up an action potential of the cell. The image demonstrates proper tracing of the cell surface. In a future application, an appropriately designed stylus might be used as an additional electrode to excite or record an action potential at any location of the cell body or a process of the neuronal cell. (See color plate.)

ships by current techniques. AFM might offer a means to overcome the need for labeling. It has recently been demonstrated that thin sections of fixed and embedded tissue and cell samples make the cell's interior accessible to AFM. High-resolution cell images give a clear picture of the cell's ultrastructure, sometimes down to the molecular level. Although AFM will not likely allow identifying macromolecules by their shape, it could be combined with powerful local spectroscopy. It is tempting to speculate that spectroscopy might then enable microscopy to identify macromolecular specimens under the tip directly (Figure 16-4).

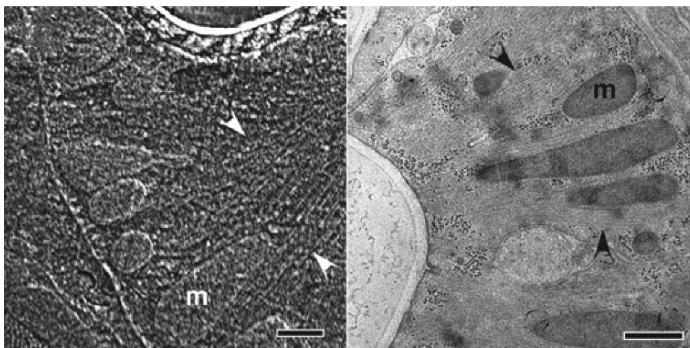


Figure 16-4. AFM image of the block face surface (left) and TEM image of an ultrathin section (right) from the nematode *Caenorhabditis elegans* embedded in epoxy resin. Scale bars equal 500nm. Arrows point to actin filaments, m, mitochondria; G, gut. (From Matsko and Mueller, 2004, reprinted with permission.)

In the following section, AFM and its elements are described. The basics for preparing macromolecular and cellular samples are then described. Finally, a few select examples highlight AFM experiments, in which a combination of imaging with sample manipulation has been used to understand macromolecular or cellular function. A comprehensive review all AFM applications in the life sciences is beyond the scope of this chapter.

2 Instrumentation and Imaging

2.1 Introduction

In AFM, the topology of the sample is traced by a sharp stylus that is scanned line by line over the sample. For most setups, the stylus is sitting at the free end of a cantilever spring. Every elevation on the sample causes the stylus to move up and bend the cantilever upward and every depression makes the lever move down. Stylus and cantilever are usually microfabricated from silicon or silicon nitride. The cantilever is typically a fraction of a millimeter long and a few micrometers thick. The softer the specimen, the softer the cantilever spring should be for it to trace the sample surface rather than deform it. The shape of the stylus is crucial too. It may be tetrahedral or extended, with a high aspect ratio, depending on need. The radius of curvature at the apex of the stylus may be as small as 2 nm (the apex of the stylus is also referred to as the *tip*) (Figure 16–5).

Accurate measurement of the deflection of the cantilever is the basis for accurate measurement of the sample topology. It also allows proper control of the loading force of the stylus onto the sample. It turns out that this latter aspect is particularly important in life science applications of AFM (see below). The first AFMs used an STM behind the cantilever to measure the deflection. Deflection has also been measured by the change in electrical capacitance between the cantilever and a reference electrode, or by means of a piezoresistor integrated with the cantilever (Tortonese et al., 1993). Most AFMs now use an

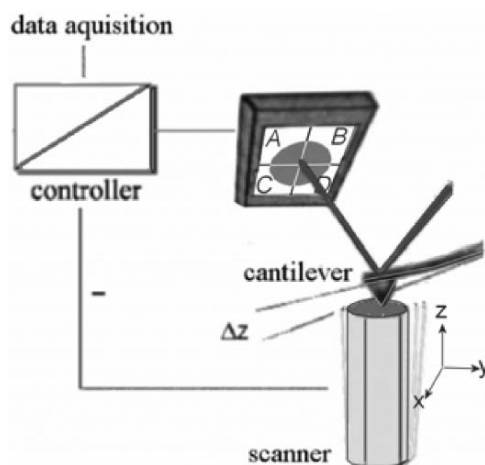


Figure 16–5. The elements of an AFM.

optical pointer to measure cantilever deflection. This detection system is fully adequate as it poses no limitation to AFM resolution.

In the optical pointer detection, a laser beam is focused onto the back of the free end of the cantilever (Figure 16–5). The laser beam is then reflected off the cantilever onto a four-segment photodiode. Prior to imaging, the four-segmented laser diode is moved until all four segments are equally illuminated. For imaging, the stylus is then loaded onto the sample. This causes the free end of the cantilever to bend upward and the laser beam now illuminates the two upper segments more strongly. The signals from the two upper segments of the diode are compared to the two lower segments $[(A + B) - (C + D)]$ to derive the amount of deflection of the lever in the z direction. The load is preset by the user, depending on the application and is related to the deflection of the cantilever:

$$F = c \cdot \Delta S$$

ΔS is the deflection in the z direction and c is the spring constant of the cantilever. The force is typically selected within the range of less than 100 pN to a few nanonewtons, depending on the application.

In operation, the cantilever is deflected from the preset value by the sample topology and the reflected laser beam is moved up or down. The original deflection is then restored via a feedback loop by a motion of the scanner perpendicular to the sample plane (referred to as the z direction). The position of the scanner with respect to the tip is recorded and used as the AFM topographical image.

Torsion of the cantilever may also occur during scanning, when the tip is experiencing friction with the sample. When the cantilever is becoming twisted, the laser beam is moved sideways. The amount of torsion and, hence, friction is then derived from comparison of the signals from the two right and two left segments of the photodiode $[(A + C) - (B + D)]$. Maps of local friction are used to reveal materials contrast in addition to the topographical image.

In dynamic AFM modes, the cantilever is oscillated and the amplitude and phase of the oscillation are monitored using the laser signal on the photodiode rather than a static deflection.

An AFM does not necessarily need to be based on a cantilever at all. In an instrument combining AFM topographical imaging with near-field optical imaging (the scanning near-field optical microscope, SNOM or NSOM), a tapered optical fiber is used as the stylus in most current instruments. It is oscillated parallel to the sample. Dampening of this oscillation is used as the feedback signal. In another alternative setup to the cantilever-based AFM, the sample is mounted on the membrane of an electret microphone and the output of this microphone is used for feedback. This setup performs equally well as the more traditional cantilever setup (Figure 16–6).

In addition to a highly sensitive probe, AFM depends on a precise scanner. The scanner is attached either to the probe or the sample. It allows the sample to be scanned with respect to the stylus in the plane of the sample (referred to as the x,y plane) and adjusting the relative height of the sample and the probe (referred to as the z direction) with

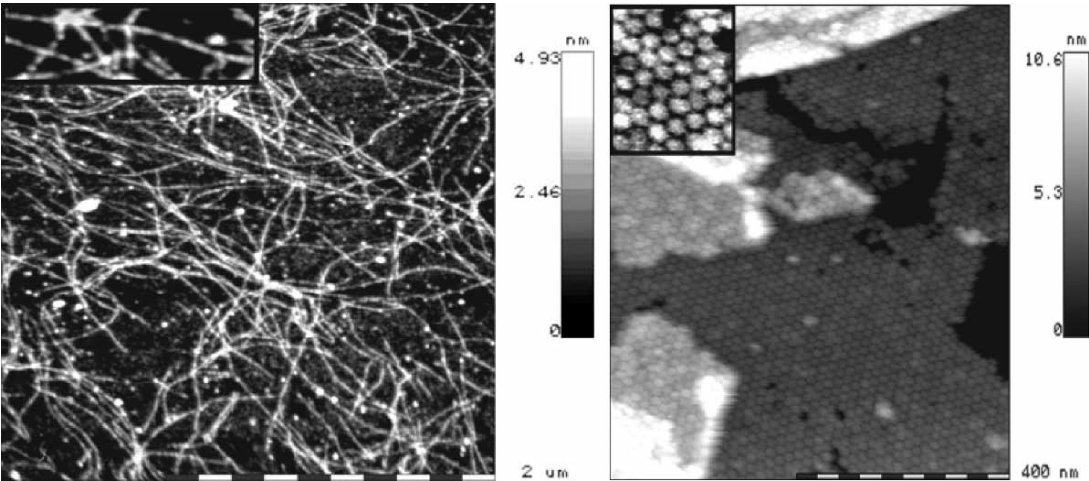


Figure 16–6. Actin filaments (left) and an archeobacterial S-layer (HPI-layer, right) imaged with an AFM based on an electret sensor, rather than a cantilever sensor (Schenk et al., 1994, 1996).

subatomic precision (Amrein et al., 1997). AFM scanners are made of voltage-driven piezoceramic elements.

AFM imaging is a mechanical process, usually operated in a closed feedback loop. Even an apparently crisp, high-resolution topographical image need not necessarily reflect the true sample topography. The faithfulness of the topographical images depends on the properties of the feedback loop, the tip and sample geometries, and on how much the sample deforms upon imaging. Eventually, the accuracy of the topographical images is limited by noise (Figure 16–7). Each of these aspects is discussed below. Also discussed is how development of AFM technology might lead to improved instrument performance.

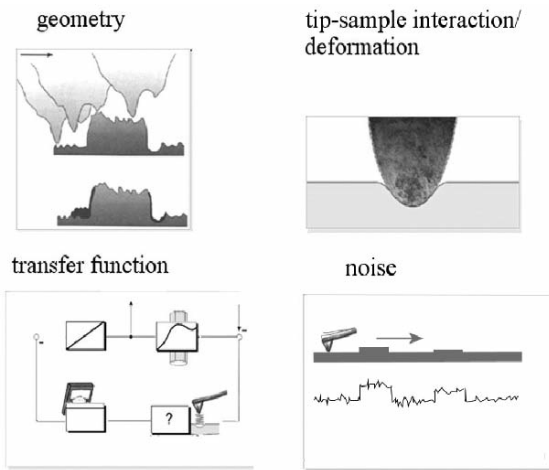


Figure 16–7. The faithfulness of an AFM topographical image depends on the sample and tip geometries, the sample deformation, the quality of the mechanical feedback loop of the instrument, and noise.

2.2 Geometry of the Stylus

When the stylus moves over a sample, the effective point of contact of the tip with the sample also changes. The topographical image is, colloquially speaking, convoluted with the tip geometry. The early days of AFM (and STM) were plagued by ill-characterized tips. Multiple whiskers created images that contained the same object multiple times (Figure 16–8). For a blunt stylus, each prominent object of the sample resulted in a local image of the tip itself rather than the local sample topology, and the overall appearance of such images has been cloudy. Because both the tip geometry and the sample topology matter, high-resolution images were sometimes obtained with an apparent blunt tip for very flat samples. This is because even the bluntest of tips has a rough surface, being covered with fine asperities.

Commercially available cantilevers now come with a well-characterized stylus and the apex may have a very small radius of curvature. The shape of the stylus needs to be selected with the sample in mind. Biological membranes or two-dimensional arrays of proteins with little overall height variations, for example, are well imaged by a pyramid-shaped stylus that ends in an apex of small radius of curvature, whereas a sample with prominent topology with steep flanks needs to be scanned by an elongated, needle-like stylus of high aspect ratio. During imaging, even a well-characterized, sharp stylus may become mechanically damaged or may pick up contaminate that renders it blunt. In these cases, the probe needs to be cleaned (e.g., by washing in ultrapure water, containing a detergent). If not successful, it has to be replaced.

When selecting an appropriate stylus, not only the expected sample topology must be considered. The sample compliance is an equally important aspect as a sharp tip may strongly deform a soft sample. This aspect is described below.

2.3 Tip–Sample Interaction

The tip–sample interactions need to be considered carefully, because they influence critically the success of the experiments. In AFM, most of the time the stylus is loaded onto the sample either intermittently (referred to as intermittent contact mode or tapping mode) or con-

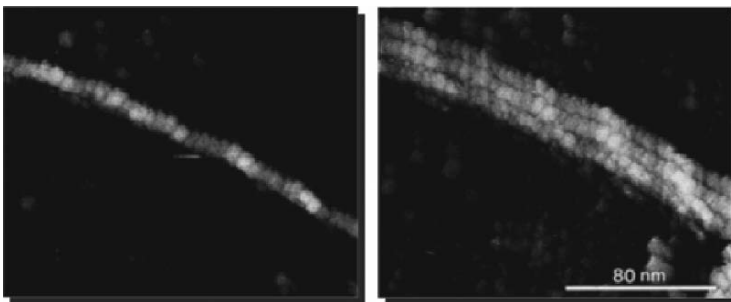


Figure 16–8. A single active filament, imaged with a single tip (left) and after the tip has become a triple tip (right).

stantly (contact mode; see below). This is achieved by approaching the probe and sample and bending the cantilever until the desired loading force is achieved.

In addition to the loading force, exerted by the cantilever spring, there are additional forces F_{StSa} acting between the sample and the stylus (see below). They may be repulsive or attractive. An attractive force makes the effective load of the tip onto the sample greater than what would be assumed from the bending of the cantilever. A repulsive force that acts prior to physical contact reduces the effective load of the tip onto the sample. The total loading force of the tip onto the sample becomes

$$F = c \cdot \Delta S + F_{\text{StSa}}$$

where ΔS is the deflection of the cantilever and c is the spring constant of the cantilever.

Evaluating these interactions may be pursued by acquiring force-versus-distance curves (referred to also as “force spectroscopy”). Thereby, the tip is approached to the sample and the deflection of the cantilever recorded as a function of the vertical position of the scanner with respect to the sample. The deflection of the cantilever can then be converted into a force using the spring constant of the lever (Figure 16–9).

2.3.1 Sample Deformation

When physical contact of the tip and the sample is established, the sample will deform until the contact area has sufficiently increased such that the load is accommodated (Figure 16–10).

The deformation strongly determines the resolution and trustworthiness of AFM imaging and must therefore be considered carefully.

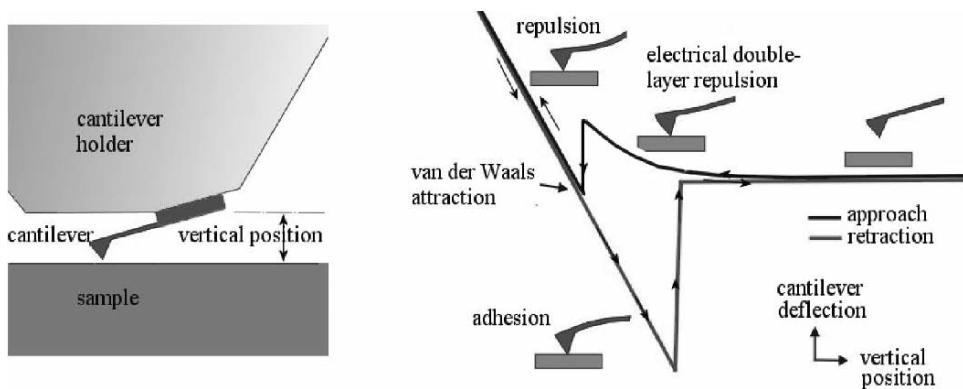


Figure 16–9. Force versus distance curve. For the example shown here, the tip first experiences a long-range repulsive force upon approaching the sample, even before the tip and sample are in physical contact. Close to the sample, the tip becomes strongly attracted by the van der Waals force. In this instance, the attractive force gradient becomes greater than the force gradient by the cantilever spring. This causes the tip to snap into physical contact with the sample (the perpendicular part of the approach curve). Once physical contact has been made, the cantilever is deflected linearly by the approaching scanner. On the way back, the tip may stick to the sample by adhesion until the pull by the cantilever forces it out of contact.

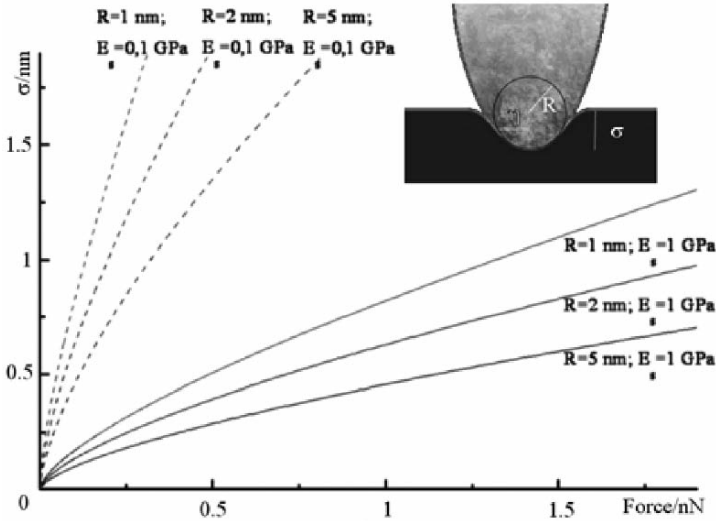


Figure 16–10. Parameter plot of the loading force F versus the penetration depth σ for three different radii of curvature of the apex R and for two types of differently stiff samples. A Young's modulus of 0.1 GPa might reflect a living cell; 1 GPa could be ascribed to a protein structure. The plot shows that even at a very low force, a sharp stylus dives right into the softer sample. It is therefore crucial to select the right kind of an apex radius for each application.

There are a number of models that relate the deformation of a solid body to the loading force of a stylus. According to Sneddon (Heuberger et al., 1996), for example, the repulsive force F for a stylus being loaded onto a solid, homogeneous body is

$$F = \frac{E_s}{2(1-\nu_s^2)} \left[(\eta^2 - R^2) \ln \left(\frac{R+\eta}{R-\eta} \right) - 2\eta R \right]$$

where E_s is the Young's modulus, ν_s is the Poisson's ratio of the sample, and R is the apex radius of the stylus (the deformation of the stylus is neglected). With increasing loading force, the radius η of the contact area between the tip and the sample increases. The penetration depth σ of the tip and the radius of the contact area η are related as (Heuberger et al., 1996):

$$\sigma = \frac{1}{2} \eta \ln \left(\frac{R+\eta}{R-\eta} \right)$$

Hence, for highest resolution, macromolecular samples need to be imaged at minimal load. Under optimal conditions, subnanometer scale resolution has been obtained on protein samples (Schabert and Engel, 1994); for a review, see Engel and Müller (2000). Müller et al. found on a two-dimensional regular array of the protein bacteriorhodopsin that at a load exceeding about 100 pN, the resolution dropped and the molecules became deformed vertically and laterally (Müller and Büldt, 1995). Because the atomic structure of bacteriorhodopsin is known, the change in topography could be assigned to a distinct

conformational change of the protein upon the higher load (Figure 16–11).

A convenient way to detect sample deformation upon too high a load comes from comparison of images from trace and retrace (Hoh et al., 1993; Müller et al., 1996; Schabert and Engel, 1994). Trace and retrace refer to the line-by-line motion of the scanner. All commercial AFMs allow two individual images to be acquired, one using the line traces when the stylus moves from left to right and the other one using the traces on the way back.

The actual load leading to high-resolution images might be even smaller than expected from the force setting of the microscope. Yang and co-workers (1996) proposed that short-range interactions with a local asperity give rise to high-resolution contrast while longer-range interactions with blunter parts of the tip help support the load of the tip in contact with the sample. Muller et al. (1997) have argued that in solution, the long-range force required for repulsion of the body of the tip is electrostatic. They adjust the supporting electrolyte so that the asperity just touches the sample lightly.

To obtain details of living cells and tissues by AFM, a meaningful image is often obtained only after a loading force of a few nanonewtons has been applied. Such high loads lead to deformation of the cell up to several hundred nanometers (Hoh and Schoenenberger, 1994). The cell membrane is then pressed onto intracellular structures such

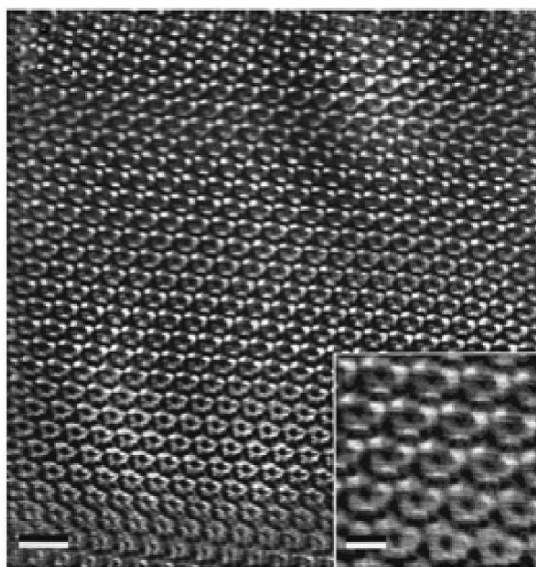


Figure 16–11. Force-dependent surface topography of bacteriorhodopsin (scale bar represents 10 nm) demonstrating the effect of force variations on the topography of the cytoplasmic purple membrane surface. The initial force of 300 pN (bottom of image) was decreased during the scan to 100 pN (top of image). A conformational change is distinct: donut-shaped bacteriorhodopsin trimers transform into units with three pronounced protrusions at their periphery. Inset: noise reduced image at higher magnification (scale bar represents 4 nm). (Courtesy of Müller and Büldt, 1995).

as the nucleus, cytoskeletal elements, and vesicles, which in consequence become visible (Chang et al., 1993; Fritz et al., 1994; Hoh and Schoenenberger, 1994). In these cases, the AFM images reflect the local plasticity of the living cells more than their true surface topography.

2.3.2 Forces between the Apex of the Stylus and the Sample

A van der Waals force F_{vdW} is always present between the tip and the sample. The main contribution to F_{vdW} is the dispersion force, caused by the dipole-induced dipole interaction and is present between all kinds of materials. The Lifshitz theory, a combination of quantum electrodynamics theory and spectroscopic data allows the forces between two geometrically shaped surfaces to be calculated. In the case of a flat surface (representing the sample) and a sphere (being used as an approximation for the apex of the stylus) F_{vdW} is (Israelachvili, 1991)

$$F_{\text{vdW}} = \frac{-H_a R}{6d^2}$$

where R is the radius of the tip (radius of the tip apex) and d is the distance between apex and sample. H_a represents the Hamaker constant, which characterizes the interaction of the two surfaces (media) across a third medium. For example, for two mica surfaces in water H_a is 2.2×10^{-20} J and for two silicon oxide surfaces in water H_a is 8.3×10^{-21} J (Israelachvili, 1991). For hydrocarbons in water, H_a lies between $(0.2 - 1) \times 10^{-20}$ J (Butt, 1992). The van der Waals force between particles is always attractive in air and attractive for most situations in an aqueous solution.

When imaging in aqueous solution, additional interactions between the apex of the stylus and the sample need to be taken into account. Many of the commonly used supports and probes as well as most biological samples are charged in an aqueous environment. This is because they usually carry weak acidic and basic functional groups. They dissociate in an aqueous solution, according to their equilibrium constants. The net charge density of a surface in water depends on the density of the functional groups, their pK values, and the pH of the buffer solution.

The DLVO (Derjaguin, Landau, Verwey, Overbeek) theory quantitatively describes the total force between charged interfaces in aqueous solution. It considers the electrostatic double-layer interaction caused by surface charges and the van der Waals forces and neglects entropic or steric contributions. Unlike the van der Waals interaction, the electrical double-layer repulsion depends on the sign and magnitude of the surface charge density, the ion concentration, and the pH. A charged surface attracts counterions in the water. At the solid-liquid interface, a charge cloud on the order of molecular dimensions is created as a transition region. In this so-called electrical double layer (EDL) the counterions balance the charge of the surface. The density of the charges surrounding the surface falls off exponentially with distance z from the surface (Debye-Hückel approximation):

$$\psi = \psi_0 e^{-z/\lambda_D}$$

ψ_0 represents the potential at the surface. The Debye length λ_D is the thickness of the EDL:

$$\lambda_D = \sqrt{\frac{\epsilon_0 \epsilon_e k T}{e^2 \sum_i c_i q_i^2}}$$

where ϵ_0 represents the vacuum permittivity, ϵ_e the dielectric permittivity of the electrolyte, k the Boltzmann constant, T the absolute temperature, e the unit charge, and c_i the concentration and q_i the ionic charge of the i th component of the liquid. Note the strong dependence of the double layer thickness on the valence of the ions.

If the tip and the sample are approaching each other, the electrical double layers of the two interfaces become perturbed when they begin to overlap (Figure 16–12). This results in a force that is known as the double layer force F_{el} . F_{el} decreases exponentially with distance d between the two surfaces. For a stylus with an apex radius of curvature of R and a planar sample (at a surface potential <50 mV) (Butt, 1992):

$$F_{el} = \frac{4\pi R \sigma_1 \sigma_2 \lambda_D}{\epsilon \epsilon_0} e^{-d/\lambda_D}$$

where σ_1 and σ_2 represent the surface charge density of the stylus and the specimen, respectively, ϵ_0 is the vacuum permittivity, ϵ_e the dielectric permittivity of the electrolyte, and d the distance between the two surfaces. This equation is a simplification of any real situation. At a distance much below λ_D , it is necessary to resort to numerical solutions for which there are no simple expressions.

In addition, when the two charged surfaces approach each other, the local ion concentration also changes. This means a shift in the equilibrium conditions of the charged groups with the ions in solution. Hence, the ionizable functional groups of the surface may become neutralized for finite dissociation constants, according to the new equilibrium conditions, and the surface charge density is decreased. This phenomenon is called charge regulation and causes a less strong repulsive force for surfaces charged with similar sign than would occur without charge regulation. F_{el} can even become attractive at very small distances (Israelachvili, 1991).

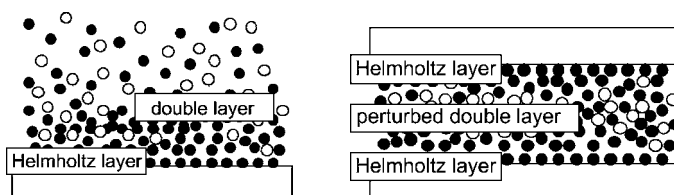


Figure 16–12. Specific interaction of ions at solid–liquid interfaces. The surfaces are negatively charged. (Left) The ions produce an interfacial region of excess solute concentration, the electrical double layer, which consists of the Debye layer and the counterions bound at the surface (Helmholtz layer). (Right) Two negatively charged surfaces at very small separation. The ion clouds of the electrical double layer overlap and cause a repulsive force (Amrein and Müller, 1999). (From Mueller et al., 1997a, reprinted with permission.)

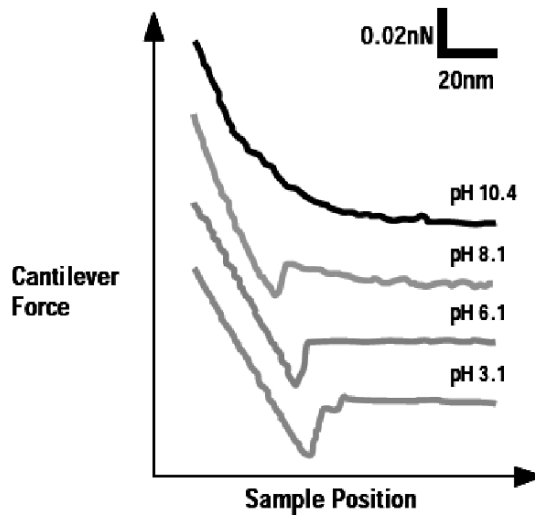


Figure 16-13. Force versus distance curves showing the DVLO force between a silicon nitride cantilever and a mica surface under varying pH conditions and at fixed ion concentration. Mica is naturally negatively charged at neutral pH, less charged at low pH, and increasingly charged at higher pH. At low surface charge, the van der Waals attraction dominates and causes the tip to snap onto the sample. At higher charge, the van der Waals attraction becomes increasingly screened by the repulsive electrical double layer repulsion. (From Butt, 1992, reprinted with permission).

To evaluate the overall force that is relevant for the interaction between the stylus and the sample, the electrical double-layer force and the van der Waals force are summed up to the total DLVO force (F_{DLVO}) (Figure 16-13).

$$F_{\text{DLVO}} = \frac{4\pi R\sigma_1\sigma_2\lambda_D}{\epsilon\epsilon_0} e^{-d/\lambda_D} + \frac{-H_a R}{6d^2}$$

For high-resolution imaging, the electrical double-layer repulsion may have to be reduced such that the stylus can effectively come in physical contact with the sample rather than “riding” on the electrical double layer. This can be achieved by increasing the (bivalent) ion concentration (Butt 1991, 1992; Butt et al., 1995; Ducker et al., 1991).

There are a number of interactions that need to be considered in addition to the above-mentioned DLVO force, including steric, hydrophobic, and hydrophilic interactions.

When the microscopy is performed in air, there is often a water bridge occurring between stylus and sample. This is because under ambient conditions, most surfaces are covered by a thin water layer. The resulting meniscus force may be quite strong (e.g., in the order of 10^{-7} N) and may pull the tip effectively onto the sample. This will result in poor resolution and may even cause damage to the sample and stylus. It is notable that the van der Waals interaction in air is usually about an order of magnitude stronger than in an aqueous environment. The detrimental effects occurring in air through these forces may par-

tially be overcome by choosing an appropriate imaging mode (intermittent contact mode, see below). The properties of the stylus may also be changed to avoid formation of a water bridge between sample and tip. Knapp et al. (1995) have established a method for rendering the Si_3N_4 probes hydrophobic. The probes became coated with a Teflon-like polymer through glow discharge in a hexafluoropropene (HFP, Hoechst AG, Frankfurt, Germany) atmosphere (Guckenberger et al., 1994; Knapp et al., 1995).

2.4 The Feedback Loop of the AFM

During scanning, the sample topology causes the stylus to move up or down and the cantilever is deflected accordingly. The deflection is measured by means of the optical pointer and compared to the set point in the regulator. The set point represents the loading force chosen by the user. Deviation from the set point is relayed to the driver of the scanner and the height of the sample with respect to the stylus is readjusted, etc. Hence, the scanning process translates the sample topology into a time-dependent signal. The signal may be decomposed (by the Fourier transform) into a spectrum of sine waves. Each sine wave of the spectrum is distinct based on its amplitude and frequency. The smaller the features are in the plane of the sample and the faster the probe is scanned over the sample, the higher the frequencies. The taller the features of the sample, the higher the amplitude of the respective frequencies. Each frequency is also distinct by its phase. The phase describes how much one sine wave is shifted with respect to all other sine waves. The quality of the feedback loop refers to how accurate phase and amplitude is relayed through feedback loop as a function of the frequency.

The feedback loop is only able to respond adequately up to a certain speed (frequency). Above that speed, the stylus does no longer trace properly the sample topology or the feedback becomes unstable. When the probe is scanned fast over the sample, for example, the speed of the feed-back loop may be too low to adjust the tip to sample distance in time. This results in variable and at times high interaction forces (Weisenhorn et al., 1993).

The quality of the feedback loop with respect to speed and accuracy depends on the quality of its elements. The elements of the feedback are the regulator, the cantilever spring, the scanner element that moves the sample up and down, and also the interaction between the sample and the apex of the stylus and the sample topology itself (Figure 16–14).

2.4.1 Stability of the Feedback Loop

Each element of the feedback loop has its own transfer function or frequency response. It amplifies the signal in a frequency-dependent manner. Moreover, it relays the signal with a delay. A given delay relates to a larger and larger phase shift with increasing frequency. The phase shifts from all elements add up. At a certain frequency the total phase shift will have become too large and causes the feedback loop to react in the wrong direction, i.e., instead of moving up, the scanner

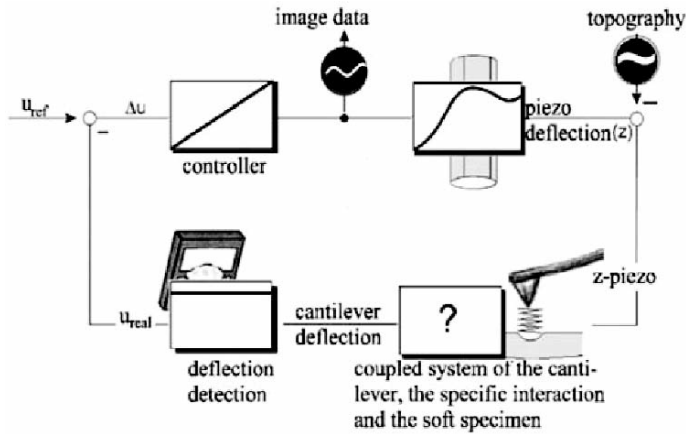


Figure 16–14. The feedback loop of the AFM. The boxes symbolize the frequency response of the elements of the loop.

would move down, etc. Only the frequency range below this limiting frequency must be used for imaging. The higher frequencies must be suppressed by an appropriate setting of the regulator. Otherwise, the feedback loop will oscillate. Because the frequency response of the feedback loop may be different for each new sample and probe and may even change during imaging, the regulator needs to be regularly adjusted by the user. Most of the time, the regulator allows the amplitude for all frequencies to be adjusted proportionally (P-regulator) and in a frequency-dependent way (integral-regulator).

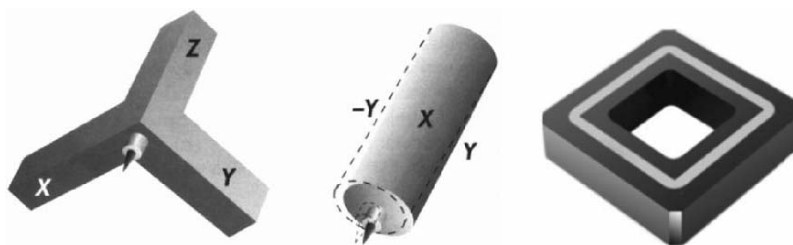
The bandwidth (the usable frequency range) of an AFM is usually limited either by the scanner or the probe in contact with the sample. The optical pointer and the regulator should not pose a limitation to speed or accuracy of the feedback loop.

2.4.2 Elements of the Feedback Loop

2.4.2.1 The Scanner

Subatomic-precision scanners are the key to the success of STM, AFM, and nanotechnology in general. Such scanners are based on piezoceramic elements. They deflect in an electrical field. Within a certain range, the deflection is approximately linear to the applied voltage. A scanner is characterized by its scan range (x,y), the z range, and the frequency response. The scan range of the actuator limits the size an image can have. The z range puts a limit on the maximum specimen corrugation the scanner can trace. The first scanners used were tripods, with each leg representing one axis of space (Binnig and Rohrer, 1982). Actuators used today in AFM are often made of single piezoceramic tubes (Binnig and Smith, 1986). The bending of the tube allows the probe or the sample to be moved in two dimensions (on approximately a spherical surface). The length adjusts the sample-to-probe distance. Piezoceramic tube scanners are superior in many respects to piezoceramic tripods that were used in the beginning of SPMs. The tube scanners are symmetrical with respect to the z axis and therefore are

less susceptible to thermal drift. They can be made much smaller to obtain a scan range similar to the tripods. Most recent microscopes also have scanners with an open frame for access from above and below the sample. In these cases, the x,y scanner is separate from the elements moving the scanner up and down.



The frequency response determines how fast the actuator can respond to an applied voltage. Today's scan units are a compromise with respect to these properties. On the one hand, a high range and sensitivity are accomplished by increasing the size of the actuators; on the other hand, larger scanners are slow to react. The bandwidth (i.e., the usable frequency range) of the scanner is limited by its first mechanical resonant frequency (also *natural frequency*) in the z direction and this depends on the size and shape of the element as well as material properties (Table 16–1). An almost linear relation between an applied voltage and the deflection is achieved only when the actuator is used well below its resonant frequency. Close to the resonance, the sensitivity rises sharply and, even more important, the deflection becomes increasingly delayed with respect to the driving voltage (i.e., there is a negative phase shift between the driving voltage and the deflection). As discussed above, too much of a delay is not acceptable, because the probe could then no longer trace the relief of a specimen in real time.

To date the scanner is usually the slowest element in the feedback loop of the AFM. It thus determines the bandwidth of the whole instrument. From this condition a slow scan speed results and, hence, a long image acquisition time that not only can make the SPM a tedious task, but also puts a corresponding limit on the time domain for the observation of dynamic events. This is surprising, since there are ways to strongly increase the bandwidth of AFM by a faster scanner without limiting the scan range in any direction (Figure 16–15).

2.4.2.2 The Probe and the Sample

The dynamic response of the probe is more difficult to understand than the frequency response of the scanner. Prior to establishing contact with the sample, the probe (cantilever) may be treated as a beam with one end fixed and one end free. Establishing the frequency response is straightforward, as described in Table 16–1. The smaller and the stiffer the cantilever, the higher its first resonant frequency lies. Note that the frequency response of a free cantilever is always established prior to imaging in the intermittent contact mode (tapping mode). However, after establishing contact, the beam is in an intermediate state between a beam with both ends fixed and one end fixed. Moreover, contact to the

Table 16-1. Natural frequencies for the elements of a scanner.^a

	Natural frequency (Hz)		
	Longitudinal	Lateral	Torsional
Beam with quadratic cross section: one end fixed (E = Young's modulus, ρ = density, L = length, and a = edge length)	$f_z = \frac{1}{4L} \sqrt{\frac{E}{\rho}}$	$f_{xy} = 0.1615 \frac{a}{L^2} \sqrt{\frac{E}{\rho}}$	
Tube with attached mass: one end fixed A = cross-sectional area, m = mass, M = attached mass, I = area-moment of inertia, d = outer diameter, w = wall thickness, I_z = moment of inertia, I_M = moment of inertia of attached mass, and $k = \frac{\pi E}{32} \frac{d^4 - (d - 2w)^4}{2.6L}$	$f_z = \frac{1}{2\pi} \sqrt{\frac{EA}{L(M + \frac{m}{3})}}$	$f_w = \frac{1}{2\pi} \sqrt{\frac{3EI}{(M + 0.23m)L^3}}$	$f_t = \frac{1}{2\pi} \sqrt{\frac{k}{I_M + \frac{I_z}{3}}}$ $I_z = \frac{m}{4} \left[\left(\frac{d}{2} \right)^2 - \left(\frac{d}{2} - w \right)^2 \right]$

^a From Harris (1987).

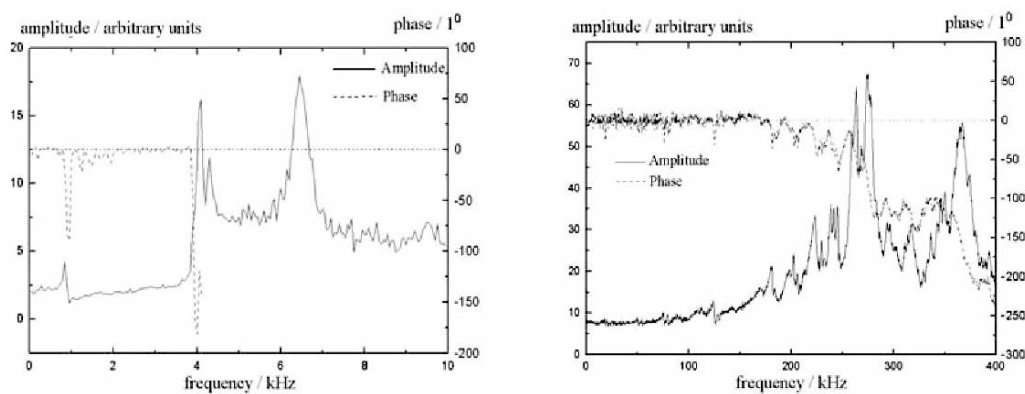


Figure 16–15. (Left) The frequency response of a typical commercial scanner (the J scanner from NanoScope III, digital instruments). The scanner may be used up to approximately 4kHz. Above, the phase shift is too large for stable imaging. (Right) A scanner developed in the group of the author. The scan range and range in the z direction are comparable to the J-scanner. However, the usable frequency range is one to two orders of magnitude wider (Knebel et al., 1997).

sample is complicated. The cantilever and the sample are coupled via the force interaction between the tip and the sample and the compliant sample itself. Mechanically, the situation resembles a series of three dampened springs, the cantilever, the tip sample interaction, and the viscoelastic sample. There is no analytical solution for this situation (Budó, 1976) and the frequency response is best determined experimentally (Figure 16–16). Smaller and stiffer cantilevers will still lead to a broader bandwidth of this element of the feedback loop.

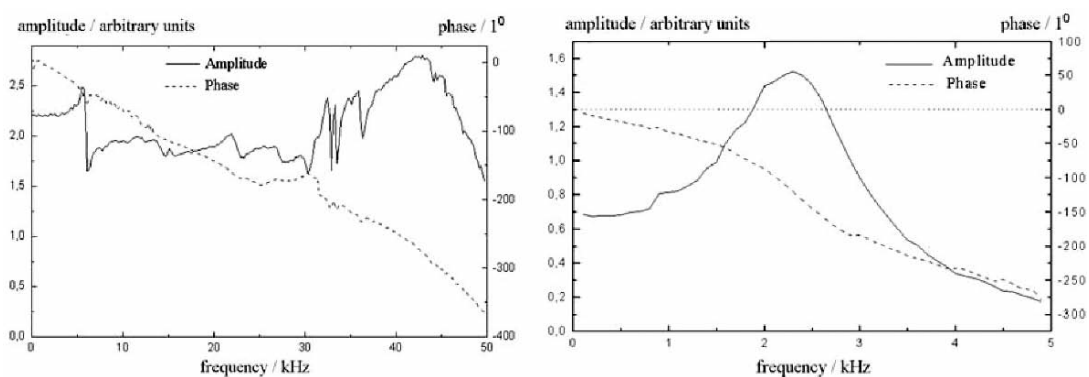


Figure 16–16. (Left) Frequency response of a silicon nitride cantilever (100μm long; $c = 0.1\text{ N/m}$; Olympus Ltd., Tokyo, Japan) in contact with a mica sample. The loading force was 1 nN. The frequency limit posed by this element of the feedback loop lies at about 10kHz, above which the phase shift has become too great for a stable feedback loop. (Right) Frequency response of the closed feedback loop of a NanoScope III (digital imaging, Santa Barbara, CA) equipped with a J-scanner. The same cantilever has been used. The measurement was carried out in water. Instead of scanning across the mica sample, the sample was oscillated, mimicking a corrugated sample. The oscillation was continuously changed from low frequency (mimicking broad objects) to a higher frequency (representing smaller objects). The amplitude of the oscillation was constant over the total measuring range. The regulator of the microscope was adjusted as for imaging to optimally track the oscillation of the sample. The amplitude shown corresponds to the output of the feedback for image acquisition.

2.4.3 Frequency Response of the Closed Feedback Loop

For imaging, all elements work together in the closed feedback loop (Figure 16–16A). The frequency response of the closed feedback loop is therefore a direct measure of the quality and trustworthiness of AFM imaging. In the example shown in Figure 16–16B, lower frequencies (larger object details) up to about 1.5 kHz are all represented with about the same amplitude. Higher frequencies up to about 2.5 kHz (smaller object details) are amplified increasingly more strongly by the system. Hence, in an image these details would be more prominent than they should be. Very small details would no longer be traced. For proper interpretation of the sample structure at high resolution, images should be taken at various scan speeds.

2.5 Noise

The instrument's noise degrades the accuracy with which the topology is traced. For a well-designed instrument, from all the elements of the feedback, the thermal motion of the cantilever is the major source of noise. The cantilever “produces” one $k_B T$ per resonance interval (k_B is Boltzmann's constant and T is the absolute temperature). This will cause the cantilever typically to oscillate by a few Angstroms. When the stylus rests in firm contact on the sample, the thermal motion of the cantilever is strongly reduced and will not affect resolution directly. However, the thermal energy now creates a fluctuation of the loading force. In the case of a soft, susceptible sample, this in turn will create a marked fluctuation of the sample deformation. The thermal force noise within the context of the simple harmonic oscillator model is given by (Sarid, 1991)

$$\Delta F = 23 \sqrt{\frac{4k_B B k}{\omega_0 Q}}$$

where ΔF is the rms force noise, B the bandwidth at which the microscope operates, T the absolute temperature, k the spring constant of the cantilever, ω_0 the radial resonant frequency, and Q the mechanical quality factor. Q parameterizes the sharpness of the resonance, being the ratio of the resonant frequency to the full-width at half-height of the peak. In air or vacuum, very high Q (on the order of thousands) can be achieved. In water, Q is much lower. As an example, for a typical cantilever used to scan a molecular sample in an aqueous solution ($k = 0.1$ N/m, $\omega_0 = 20$ kHz, $Q = 1$, $B = 10$ kHz) the thermal force noise is 5–20 pN. From the equation above it follows that an ideal cantilever for low-noise operation should have a low spring constant, but still a high resonant frequency. For a given size of a cantilever, however, the stiffer the lever, the higher the resonant frequency. This dilemma can be overcome by employing smaller (shorter) cantilevers (Hansma, 1996). The effect of the thermal noise on the image quality will also be reduced by operating the microscope at a lower bandwidth. But this comes at the cost of reduced scan speed.

It is notable that noise is not evenly spread over the entire frequency spectrum. For large Q , there is a large reduction of the thermal noise

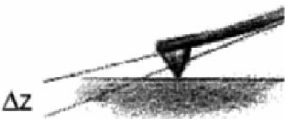
away from the resonant peak. In water, the thermal noise is quite evenly distributed over the whole frequency range. At low frequencies, other sources of mechanical motion become progressively more important. This is referred to as $1/f$ noise and is notable foremost in contact mode imaging (see below).

2.6 Imaging Modes

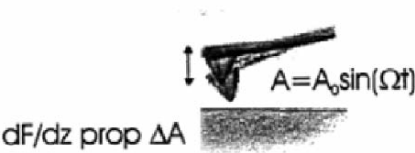
2.6.1 Contact Mode

In contact mode, the tip is permanently touching the sample. On a relatively solid sample, the position of the cantilever in the z direction is well defined and is affected only by force noise. It has therefore provided the best resolution to date on molecular samples. Contact mode imaging works best with densely packed samples, to provide stability against the sideways push by the tip. On soft samples, small cantilevers with a low spring constant are desirable. For operation at minimal force, the force during scanning is reduced until the cantilever takes off from the sample and then increases it just enough to touch back down and trace the topology properly.

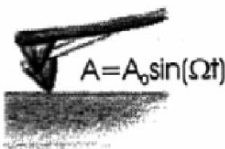
contact mode AFM: high spatial resolution



non-contact AFM: low forces



tapping-mode AFM: no lateral forces



2.6.2 Noncontact Mode

In noncontact mode, the probe tip is excited to oscillate near the resonant frequency. To chose the excitation frequency, it is necessary to perform a sweep from low to high frequencies and observe the amplitude and phase of the cantilever oscillation. The fundamental bending frequency of the cantilever will stand out as a prominent peak and the

excitation frequency chosen to be close by (on the lower-frequency flank of the peak). In close vicinity to the sample but prior to contact, the interaction between the tip and the sample will change the resonance conditions for the lever (as the cantilever now enters an intermediate state between a lever free on one side and a lever fixed on both sides). This in turn will cause the amplitude to decrease, because the driving frequency is now further away from the resonant frequency. The shift in resonant frequency is a function of the force gradient (dF/dz) between the tip and the sample. The microscope is operated to keep the amplitude reduction at a defined level and, hence, the tip at a defined distance to the sample. Although noncontact imaging promises to be the most subtle mode for highly deformable, weakly immobilized objects, it has not become popular so far in life science applications because it does not usually allow for stable high-resolution tracing of an interface of a biological sample. This might be because the tip is kept at a distance to the sample where it still can oscillate considerably by thermal noise and instrument instabilities. Very stable microscopes, very small cantilevers, and very sharp tips with a high aspect ratio may, in the future, bring about a strong improvement for this potentially attractive imaging mode.

2.6.3 Intermittent Contact Mode (Tapping Mode, Dynamic Force Microscopy)

Intermittent contact mode AFM in a certain sense combines the advantages of noncontact and contact AFM. The probe oscillates as in the noncontact operation. But the set point for the amplitude is now chosen so that the tip can make brief contact with the sample in each cycle of its oscillation. As in the noncontact mode, there are no lateral forces exerted on the sample. As in contact mode AFM, the local topological height is determined accurately by the well-defined contact point. This mode has therefore had success in imaging well-separated molecules bound to an underlying support rather weakly that did not withstand the lateral forces of contact mode imaging. It is notable, however, that the loading force during contact is not smaller than with contact mode AFM. Intermittent contact mode has resulted in resolution of molecular samples on the order of 1 nm (Figure 16–17).

As in noncontact mode, intermittent contact mode AFM in air is sensitive to a shift in resonance frequency that then brings about a change in amplitude of the oscillation. In buffer, resonance effects are small and the microscope is largely sensitive to the restriction of the oscillation brought about by the physical barrier posed by the sample. Its sensitivity is thus limited by thermal fluctuations and is not significantly enhanced by operating at resonance. In water, it is desirable to choose as small an amplitude as possible to minimize disturbance of the sample. In practice, the overall oscillation amplitude is set to the smallest value that provides stable imaging. Then the set point amplitude change is increased (increasing the resolution) to the largest value that does not damage the sample. However, stable operation requires amplitude sufficient to pull the tip out of attractive interactions.

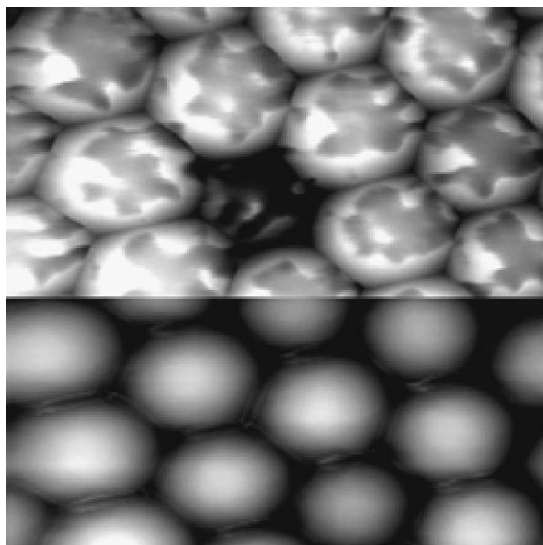


Figure 16–17. Latex-beads imaged in tapping mode. In the top portion of the image, the set change in amplitude was chosen large and the beads were deformed. In the lower portion, the set amplitude change was reduced, resulting in proper topographical imaging ($1\mu\text{m} \times 1\mu\text{m}$).

Two methods for exciting the cantilever oscillator are common, acoustic drive (Hansma et al., 1994; Putman et al., 1994) and magnetic drive. In acoustic drive an oscillator supplies a voltage drive to a piezoelectric actuator that generates sound waves in the cantilever holder. The frequencies are typically from tens of kilohertz to megahertz. When the driving frequency is near a bending-mode resonance of the cantilever, the cantilever is driven into an oscillation. A frequency scan with acoustic drive results in many peaks, only some of which yield a usable signal (Putman et al., 1994; Schäffer et al., 1996). This is because the transmission of the acoustic wave is not uniform over the frequency range. In magnetic drive, the cantilever is coated with a magnetic film, or a magnetic particle is glued onto the end of the cantilever (Lantz et al., 1994; Han et al., 1996). The oscillating magnetic field of a solenoid in the vicinity of the cantilever causes the cantilever to oscillate.

In contrast to contact mode, which senses a force, intermittent contact mode AFM is a technique that senses an interfacial stiffness. As frequency is increased, intermittent AFM also becomes sensitive to increased interfacial damping.

2.7 Optimizing AFM Design for Life Science Applications

The range of applications for AFM is very large, from atomic resolution imaging on some crystal surfaces to imaging or manipulation of whole cells. An AFM designed for working under high vacuum to look at defects in atomic lattices has different basic requirements from one that is designed predominantly for samples in liquid, with full optical microscope integration as a priority. On biological samples, atomic

resolution is generally not possible; for samples of macromolecules and larger structures, the resolution is mainly defined by issues such as the tip-sample interaction and sample deformation. Then technical issues such as integrated optics, flexibility of handling, and sample environment carry more weight.

2.7.1 Combining AFM and Optical Microscopy

Optical microscopy still dominates life science research, and techniques for enhancing contrast, labeling, or staining to recognize and study biological samples have been developed. It makes no sense to throw away this vast reserve of information about samples that are to be studied using the AFM, and indeed it is not necessary to compromise the quality of optical images for light microscopy to be combined with AFM. For a well-designed AFM it is possible to simultaneously obtain high-contrast and high-resolution optical images, including techniques such as laser scanning microscopy (LSM), total internal reflection fluorescent (TIRF) microscopy, and fluorescence resonance energy transfer (FRET) microscopy (Figure 16–18).

In one sense it is relatively straightforward to build a combined optical and atomic force microscope, since the AFM can be built around a standard inverted optical microscope. This allows the use of the optics through the objective lens under the sample (if the sample is transparent, which is often the case), while at the same time the AFM can approach from above. In practice, care must be taken that the entire system has the required mechanical stability, that sources of drift or vibration are minimized, and that the “open” nature required of such an AFM does not compromise its stability and reduce the resolution achievable. There are also two potential problems that must be considered in the instrument design that relate directly to the optical imaging: the optical quality of the support for the sample, through which the optical image must be formed, and the light path to illuminate the sample from above.

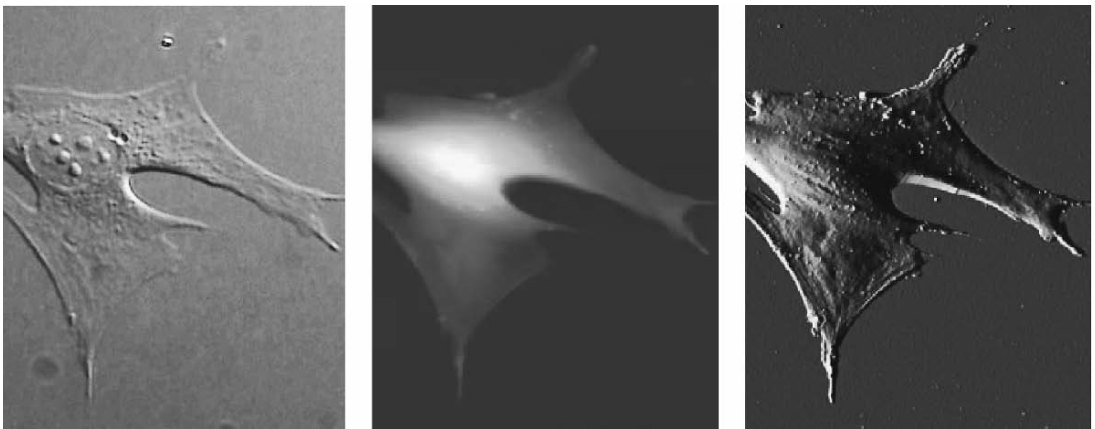


Figure 16–18. Images of a fixed 3T3 cell in buffer, on a coverslip. Left to right: Optical (differential interference contrast, DIC) and simultaneous AFM (JPK Instruments, contact mode) topography and feedback signal. (See color plate.)

For high-quality optical images, the objective lenses are optimized for thin glass coverslips, which are mechanically unstable sample supports for AFM imaging. The working distance of high-magnification objectives is very small, so the objective lens must be positioned close beneath the sample, often with an oil or water droplet between the lens and the coverslip. These considerations must be addressed by appropriate design of sample holders (which must generally also function as a liquid cell) that are able to provide enough mechanical support for the sample without interfering with access for the microscope objective beneath.

For common biological samples such as cells or vesicles, there is very low optical contrast in brightfield illumination, so some enhanced optical contrast mechanism is required. The most common choices are phase contrast (which provides contrast based on the sample refractive index) or differential interference contrast (DIC, which is sensitive to local differences in the refractive index from point to point). In both cases, the path length and optical quality of any components in the illumination path are critical. To use the optical microscope in its standard configuration, and hence have the optimal optical system, the microscope requires illumination from above, using the condensor optics. This means that the illumination light must pass “through” the AFM in some way. The ideal optical solution is to be able to use the condensor optics provided by the optical microscope manufacturer, and build the AFM on an open frame around the optical path that does not interfere with the sample illumination.

Another issue for experiments involving simultaneous AFM and optical microscopy is the laser illumination used to measure the cantilever deflection. Although the majority of this is reflected by the cantilever, a significant proportion spills over the cantilever edges and can pass through the sample into the optical microscope image. For brightfield, phase contrast, or DIC this can be removed using a simple filter in the optical path to cut the required wavelength, but for fluorescence this can severely limit the wavelength regions available and hence dyes that can be used for labeling. This is particularly a problem when the sample should be labeled with more than one dye to identify different components, or in more advanced optical techniques such as FRET, where two fluorescent dyes are used simultaneously. Typically most AFMs used a detection laser in the red region of the optical spectrum, because of the availability of small, low-power red laser diodes. More recently, AFMs are becoming available with infrared laser illumination, which allows the full visible spectrum to be used for optical imaging. Using an infrared detection laser also results in better performance of the AFM itself, in that it removes artifacts caused by optical interference between the light reflected from the cantilever and from the sample in the path to the photodiode detector.

2.7.2 *Scanners*

Over time, the scan range available from commercial AFMs has increased, particularly the z-range or distance the tip is able to move up and down over the surface. For cell imaging, a z-range of at least

10–15 μm is required, otherwise the tip will be unable to move high enough to scan over the cell nucleus. Larger x and y scan ranges of 100 μm or more are also generally required for cell imaging, and sometimes also for imaging much smaller samples, which may be distributed inhomogeneously over the surface. The piezoceramic material used for all AFM scanners suffers from various problems of nonlinearity, hysteresis, and creep. Although it is possible to move the tip very precisely, this is against a large background of position changes due to longer term effects in the piezo material, as it continues moving slowly for a long time after a voltage jump is applied (creep) or moves variable distances depending on its history (hysteresis). These problems have been addressed by adding position sensors (such as capacitive or strain-based sensors) along the movement axes, so that the tip movement is no longer set merely by converting the desired position into a simple voltage. The current position of the tip is constantly read by the position sensors and the nonlinearity of the piezo material and its changes over time can be constantly corrected, using another feedback loop. These “linearized” piezo systems are also becoming available in commercial AFM systems and are likely to become standard.

There are two possibilities generally used for the lateral scanning for an AFM—tip scanning or sample scanning. For general AFM, these are equivalent, and both have their advantages and disadvantages, but in terms of optimizing an AFM for biological samples, tip scanning generally has clear advantages. On a basic practical level, much of AFM imaging for life science research has to take place in aqueous solutions, usually containing a reasonably high concentration of salt, and there is always a greater risk of expensive damage when the high-voltage piezoelectronics sit at a lower level than the sample. Even using a tip-scanner, the AFM must be carefully designed so that the electronics and piezo elements above the sample are protected from accidental spillage or water vapor. This is particularly important if the sample is to be heated to 37°C, when the evaporation is significant, and condensed water vapor could easily collect in an unsealed AFM head. There is also a more fundamental problem with sample scanning, however, as simultaneous AFM and optical imaging cannot generally be performed. When the sample is moved in the z direction, which usually lies along the optical axis, the objective lens can be tracked with the sample using a piezoactuated lens holder, but this is not possible for lateral scanning movements across the optical axis.

2.7.3 Sample Environment

In addition to the considerations already discussed about the need for using coverslips as a sample support for many optical imaging applications, there are other practical issues raised by life science samples. Temperature control of the sample is often important for studying molecular reactions or whole cells under physiological conditions, and, for example, many lipid bilayers used as model cell membranes undergo phase transitions over the range from around 10 to 37°C and above. Perfusion is also important, particularly for *in situ* experiments and to introduce molecules for reactions, blocking, or to change the properties

of the solution. For live cell work, it may also be necessary to allow a gas exchange to equilibrate the cell medium with 5% CO₂.

The priorities for the design of the cantilever and sample holders should also include the easy disassembly and cleaning of all the components that are in contact with the sample, so that, for example, ultrasound or autoclaving is possible. When working in liquid, all parts of the instrument that come in contact with the fluid are sources of possible sample contamination. The issues are somewhat different depending on the applications. For instance, if the aim is to image or stimulate living cells, then molecular-sized contamination is not so important for AFM imaging itself, but the sample must remain sterile and the cells must not be exposed to any chemical contaminants that will produce a biological response. For single molecule imaging, the molecules may not be affected by traces of certain ions leaching from metal surfaces, but every macromolecular contaminant that could adhere to the surface is a problem for AFM imaging.

3 Sample Preparation

To study their native structures and probe their functions, most cellular or macromolecular samples need to be kept in an aqueous environment. Hydrophilic and hydrophobic interactions promote correct folding of the polypeptide chains into a protein and are responsible for the formation of micelles, bilayers, and membranes from lipids and proteins. The conformation of membrane proteins is determined by hydrophobic interactions with lipidic tails and hydrophilic interactions with their heads and the surrounding water (Haltia and Freire, 1995; Engel et al., 1992; Jap et al., 1992). The pH, electrolyte type, and its concentration and temperature also influence the structure and function. The function of macromolecular structures depends not only on their native conformation but often requires even more exacting environmental conditions with respect to pH and temperature and may depend on the presence of coenzymes or ATP, for example. Living cells are sensitive not only to pH, ionic strength, temperature, and CO₂ levels; they usually need a specific and often highly complex medium in which to grow. When biological structures are allowed to air dry, they are subjected to a high force caused by the change in surface tension as the water evaporates. The energy involved is considerable. As a result, even macromolecules become severely flattened and collapsed (Baumeister et al., 1986; Kellenberger et al., 1982; Kellenberger and Kistler, 1979; Wildhaber et al., 1985). For these reasons, most sample preparation techniques described below are for samples in buffer. Most of the time, immobilizing cells and macromolecular structures is all that is needed for AFM analysis. AFM analysis of macromolecular samples is demanding with respect to the cleanliness of the support and the purity of the buffer. All surfaces become immediately covered with hydrocarbons when exposed to ambient air. Even double-distilled water can be a source of organic contaminants. A layer of these hydrocarbons on the sample or the probe can be most disturbing for AFM.

As a result, the sample supports should be prepared or activated immediately before use. Ultrapure water (fresh milli-Q water; $\leq 18 \text{ M}\Omega \text{ cm}^{-1}$) should be used to prepare all buffer and rinsing solutions because it contains fewer hydrocarbons and macroscopic contaminants than conventional bidistilled water.

In the following sections, we will describe suitable supports and immobilization techniques for both macromolecular and cellular samples.

3.1 Macromolecular Samples

Immobilizing macromolecular structures aims for a homogeneous distribution of the specimens in a close-to-native conformation. Tight binding of the biological specimens to the support surface will prevent them from clustering. They may also better withstand the forces that arise between the probe and the sample for most AFM. On the other hand, the structure can be substantially distorted and proteins may even denature by strong binding. This is well known from transmission electron microscopy (TEM) as well as from STM of biological specimens (Baumeister et al., 1986; Wang et al., 1990). The specimens may adsorb with preferential orientations depending on the binding conditions (Fisher et al., 1978; Hayward et al., 1978; Karrasch et al., 1993; Müller et al., 1996). In addition to suitable binding properties, the support should be as smooth as possible so that it does not interfere with the structure of the biological specimen in the final image. Furthermore, it should be relatively chemically inert to prevent contamination due to the solution or nonspecific reactions with the biological system.

3.1.1 Specimen Supports

Glass coverslips are widely used as an amorphous specimen support and can be used either unmodified or altered to change their physisorption or chemisorption properties. The surface can be almost featureless on the scale of macromolecular specimens. They are best suited for all experiments in which visible light is transmitted across the sample, as in SNOM or in the combined light and atomic force microscopy. Before use, organic contaminants, dust, and other particles are removed by washing one time with concentrated HCl/HNO₃ (3:1) and five times for 1 min with Millipore water in an ultrasonic bath (50 kHz). This process makes the coverslips clean and smooth (rms roughness $\sim 0.5 \text{ nm}$).

The most commonly used support for imaging biological specimens in the AFM is mica. Mica minerals are characterized by their layered crystal structure. Mica can be readily cleaved for a clean, atomically flat surface. Muscovite mica (Mica New York Corporation, New York, NY) is the most commonly used form. The average surface charge density of muscovite mica in water is $\sigma_m \approx -0.0025 \text{ C/m}^2$ (0.015 electron per surface unit cell).

Gold surfaces can be easily prepared by vapor deposition. They are chemically inert against O₂ and stable against radicals. They bind organic thiols or bifunctional disulfides with high affinity, which can

be used to covalently attach biological macromolecules (see below). Hegner et al. (1993) developed a relatively simple and reliable method for preparing ultraflat gold (Knebel et al., 1997) surfaces. They consist of atomically flat terraces many microns in diameter. Thin carbon films commonly used in TEM are smooth on molecular dimensions and adsorb macromolecules well when freshly prepared. High-vacuum carbon evaporators are common in electron microscopy laboratories.

3.1.2 Physisorption, DLVO Force

The most common technique for immobilizing biomolecules onto a support is by physisorption. Usually, the objects are immobilized out of an aqueous solution. They become attached to a support when there is an overall attractive force that pulls the surfaces into contact. The relevant force for adsorption is the DLVO force. Hydrophobic and hydrophilic interactions may also play a role. The DLVO force for two planar surfaces is (Israelachvili, 1991)

$$F_{\text{DLVO}} = F_{\text{el}}(z) + F_{\text{vdW}}(z) = \frac{2\sigma_{\text{su}}\sigma_{\text{sp}}}{\epsilon_0\epsilon_e} e^{-z/\lambda_D} + \frac{-H_a}{6\pi z^3}; \quad \text{per unit area}$$

The DLVO force between charged surfaces is highly susceptible to ion concentration and conditions can thus be adjusted to achieve good adsorption (Müller et al., 1997) (Figure 16–19).

When the electrical double layer repulsion between the two surfaces has “vanished,” they rapidly coalesce to minimize the interaction energy (Israelachvili, 1991). For the example shown in Figure 16–19, at electrolyte concentrations above 50 mM KCl, the amount of adsorbed membranes rapidly increased and reached its maximum at about 150 mM KCl. Note that adsorption occurred even though both mica and the sample carried a net negative charge.

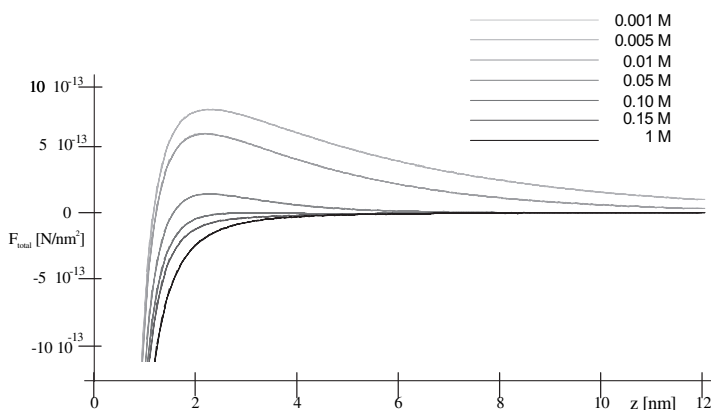


Figure 16–19. Dependence of the DLVO force on ion concentration (1:1 monovalent electrolyte) and distance between a macromolecular sample (purple membrane) and a mica support. (From Mueller et al., 1997a, reprinted with permission.) Whereas the attractive van der Waals force is mainly unaffected by the electrolyte, the double layer repulsion decreases with increasing salt concentration. The surface charge densities were -0.0025 C/m^2 for mica (Israelachvili, 1991) and -0.05 C/m^2 for purple membrane (Butt, 1992), respectively. The Hamaker constant was $3 \times 10^{-19} \text{ J}$.

3.1.3 *Physisorption, Hydrophobic and Hydrophilic Interaction*

There is an attractive interaction between hydrophobic surfaces in water. The attractive interaction potential is larger than the van der Waals potential and can be very long range. The nature of these long-range forces is not yet fully elucidated. Hydrophilic molecules, on the other hand, tend to disorder the surrounding water molecules and prefer contact with water molecules. Hence, the molecules repel each other. These repulsive, hydrophilic forces are also referred to as hydration, structural, or solvation forces. They may cause the DLVO theory to fail at small distances between two hydrophilic surfaces. With respect to adsorption, hydrophobic molecules do not attach to a hydrophilic surface and vice versa. For example, hydrophilic purple membranes did not adsorb to highly hydrophobic supports such as derivatized glass. The hydrophilic and hydrophobic interaction can cause an oriented adsorption of molecular structures.

3.1.4 *Physisorption, Preparation of the Support*

With mica, an active surface is conveniently obtained by cleaving the layered mica crystals prior to specimen adsorption. For most other supports, the active surface cannot be produced so easily. These supports are usually covered by hydrocarbon contaminants and behave more or less hydrophobically. Glass, silicon wafers, and many thin films can be rendered hydrophilic by exposure to glow discharge (for example, in a Harrick Plasma cleaner, 1 min, $p = 0.1$ m bar, with air as the residual gas) right before use. Thin carbon films become negatively charged. For those specimens that adsorb better to hydrophobic surfaces, glow discharge must be omitted.

Coating is another way to improve physisorption on many specimen supports and it has been used for a long time by electron microscopists (Jacobson and Branton, 1977; Mazia et al., 1975). For example, poly-L-lysine can be used for coating glass and mica and render the coated surfaces positively charged. This allows cells, tissues, and plasma membranes that are usually negatively charged to be readily adsorbed. Objects that carry charge in an uneven distribution can be adsorbed in a defined orientation on a poly-L-lysine-coated surface. For example, purple membrane mainly consists of a light-driven proton pump that builds up an electrochemical potential across the membrane. Illuminated by light, purple membranes show an asymmetric charge distribution and adsorb to polylysine-coated surfaces in an oriented fashion (Fisher et al., 1977, 1978; Hayward et al., 1978). More than 90% of the membranes attach with their cytoplasmic surface toward the poly-L-lysine under specific conditions (pH 9). At a pH below 4, the majority of the membranes (>94%) were directed with their extracellular surface toward the coated surface.

3.1.5 *Chemical Bonding*

Covalent bonding can be a very reliable technique to allow firm binding of biological specimens to a support. Some of the first high-resolution AFM topographies of protein structures in buffer solution have been obtained using this technique (Karrasch et al., 1993). It appears that covalent binding does not interfere with the macromolecular structure

anymore than physisorption. Bonding of the macromolecular specimens can be accomplished using chemically modified supports. Karrasch et al. (1993) developed a protocol to cross-link biological systems to a silanized glass coverslip. The silane (APTES, Fluka Chemie AG, Buchs, Switzerland) contained a free amino group that allowed it to react with the succinimide ester group of the photocrosslinker ANB-NOS (Fluka Chemie; $\lambda = 312\text{nm}$). Proteins were then bound to the interface by activating the photocrosslinker with UV radiation. This method resulted in the first high-resolution images of protein structures by AFM in buffer (Figure 16–20).

Epitaxial gold surfaces can effectively be functionalized by alkane-thiols. They form ordered, self-assembled monolayers that are tightly bound to the gold surface via chemisorption of the sulfur atoms. The monolayers are further stabilized by the lateral hydrophobic interactions of the alkyl chains (Hegner et al., 1993; Sellers et al., 1993; Wagner et al., 1994; Wolf et al., 1995). The latter can carry head groups at the free end that allow oriented covalent anchoring of macromolecular structures (Allison and Thundat, 1993; Hegner et al., 1993, 1996; Wagner et al., 1994). Wagner et al. (1995, 1996) bound protein structures via their amino groups with an *N*-hydroxysuccinimide-terminated monolayer on gold.

3.1.6 Langmuir–Blodgett Films

There are amphiphilic substances that naturally form insoluble monomolecular films on an air–water interface. They exhibit a water-soluble polar or charged head group and a highly apolar tail. This causes them to attach to an air–water interface with the head group immersed in the water and the tail toward the air. The most prominent example is the pulmonary surfactant that forms at the interface of the respiratory

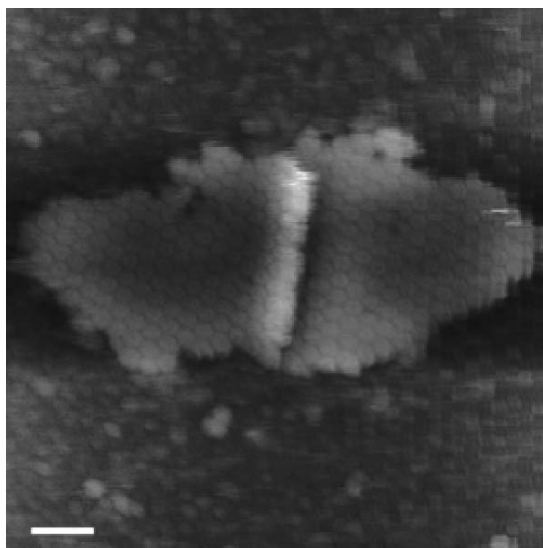


Figure 16–20. Hexagonally packed intermediate layer. Scale bar = 70nm. (From Karrasch et al., 1993, reprinted with permission.)

gas lumen and the solvation layer that covers the alveolar epithelium of lungs. Surfactant layers can be formed *ex vivo* in a Langmuir trough to study their biophysical properties under defined conditions or for the purpose of microscopic examination. Langmuir films of lipids have also been used to mimic biological membranes (for references see Bader et al., 1984), or they served as a substrate to bind and crystallize proteins in two dimensions for TEM and AFM investigations (e.g., Brisson et al., 1994). AFM proved to be outstandingly well suited to study the structure and mechanical properties of these thin layers.

To prepare films for microscopy, the amphiphilic substances are spread at the air–water interface of a Langmuir trough. They are then compressed by a movable barrier by a desired amount. To perform AFM on the air side of the film, the monolayers may be transferred from the air–water interface onto a solid support by slowly pulling a hydrophilic support out of the aqueous phase across the interface (Langmuir–Blodgett transfer; Blodgett and Langmuir, 1937). The film is deposited as the support is moved vertically across the air–water interface. It is then inspected by AFM in air. To do microscopy on the aqueous side of the film, the monolayer may also be deposited by dipping a hydrophobic substrate from the air side across the interface into the water. If a first lipid layer is deposited on the upstroke onto a hydrophilic substrate and then another layer added on the down stroke of the sample, a complete lipid bilayer has formed. This bilayer may contain membrane proteins. It is interesting to note that deposition of a bilayer onto a mica substrate arrests the lipid of the first lipid layer. These lipids are no longer free to diffuse in the plane of the membrane. If the support is glass, both the lipids bound to the support and those within the second layer facing the aqueous phase are free to diffuse.

Finally, Langmuir–Blodgett transfer may not be necessary and films of pulmonary surfactant have been studied directly at the air–water interface (Knebel et al., 2002).

3.2 Cells

Successful immobilization of living cells largely depends on the cell type. There are cells with adherent growth (for example, epithelial cells, fibroblasts, or glial cells), and cells that grow in suspension without contact to a substrate (for example, bacterial cells or erythrocytes). Adhesive cells are more readily imaged with the AFM, whereas cells that grow in suspension have to be immobilized to be imaged. It is notable that cells may change their shape, physiology, and even their life cycle once bound to a substrate. A variety of techniques have been developed to immobilize living cells. Cells are best imaged with an AFM that is combined with a light microscope.

3.2.1 Adsorbing Cells on Glass Coverslips

Cells that naturally adhere to a substrate can either be cultured on an appropriate support and subsequently imaged, or plated on the support and monitored shortly after they have established cell–substrate contact. For both procedures, the glass coverslip must be thoroughly cleaned. If cleaned with water, the glass support has to be dried in air or a stream

of N_2 to prevent plated cells from possible osmotic shock. Coverslips have been coated with poly-L-lysine, collagen (Henderson et al., 1992), proteoglycans, laminin, or fibronectin to improve adhesion.

For imaging individual, adherent cells with the SFM, the density of the cell suspension has to be chosen such that enough space remains for the cells to spread out. The time required for the cells to attach and spread depends on the cell type. Before imaging, the samples have to be rinsed with buffer solution to remove cells that are not firmly attached and, if feasible, examined by conventional light microscopy. Specific cells that were cultured on a solid support spread out to a thickness of less than 100 nm over large areas in the periphery (Fritz et al., 1994; Henderson et al., 1992; Hoh and Schoenenberger, 1994; Kasas and Ikai, 1995). In these thin regions it is possible to monitor the organization of the intracellular cytoskeleton (Figure 16–13).

3.2.2 *Immobilizing Nonadhering Cells*

A stable immobilization of cells that grow in suspension and do not establish substrate interactions in their natural environment is difficult to obtain. Hörber et al. (1992) have a method to trap single cells by a micropipette and image the exposed part with the AFM. The setup makes it possible to use the advantages of the micropipette technique and to enhance the inner pressure of the cell. This is an advantage, because the “spring constant” of a cell surface may be very low [for example, ~ 0.002 N/m (Hoh and Schoenenberger, 1994)]. Hence, the cell is extensively deformed by any reasonable interaction force with an AFM probe.

Permeable supports provide the possibility of measuring additional properties of cells (for example, permeability, diffusion, and voltage characteristics of the plasma membrane) while they are imaged by AFM. The cells may attach onto substrates with a much smaller pore size than the average diameter of the cell, or individual cells may be trapped in pores that are only slightly smaller than the average cell diameter. Kasas and Ikai (1995) have used Millipore filters (Millipore PCF, Millipore Corp., Bedford, MA) with pore sizes similar to that of the cell diameter for trapping yeast cells. Hoh and Schoenenberger (1994) have cultured MDCK epithelial cells (average lateral diameter $\sim 10 \mu\text{m}$) on polycarbonate filter supports (Millipore PCF, 12 mm diameter) with a much smaller pore size ($0.4 \mu\text{m}$) than the average cell diameter.

4 Imaging and Locally Probing Macromolecular and Cellular Samples: Examples

4.1 Imaging

By controlling the ionic strength and composition of the deposition buffer, individual collagen molecules can be assembled and adsorbed onto a mica surface in various different conformations (Jiang et al., 2004). Around five collagen molecules associate form microfibrils, which have a lateral size of around 3–5 nm. These microfibrils are also likely to be an

intermediate stage in the formation of the larger collagen fibers seen in natural tissue. A monolayer of these microfibrils can then be adsorbed to the surface to form a nanostructured, biologically active surface. The fibrils can be aligned through adsorbing under conditions of hydrodynamic flow (Figure 16–21); Figure 16–21 shows a case in which the fibrils are adsorbed at very low coverage, and the 67-nm banding repeat can be seen along the filament length. For samples in which a complete monolayer is formed, the composition of the adsorption buffer can be used to control whether the filaments organize themselves such that the bands on adjacent filaments are aligned or not (Jiang et al., 2004).

The nanometer-scale topography of these surfaces can be measured using AFM, and then used for cultivating cells *in situ*. The orientation and growth direction of fibroblast cells grown on the aligned collagen supports depend critically on the alignment of the D-banding between adjacent collagen fibrils. The overall alignment of the direction of the collagen fibers is not sufficient to produce a response in the cells, but when the banding of the fibers is aligned, the cells show a strong response (Poole et al., 2005). This is one case in which the ability to combine AFM and optical microscopy allows the study of a biological structure/function question over the size scale from the molecular structure and organization to the response of whole cells.

4.2 Beyond Imaging

One strength of the AFM is that it is able to combine imaging modes sensitive to different properties of the sample with direct measurements of forces and interactions. These measurements can be carried out at particular points (selected, for instance, from a sample that has just been imaged), or built up over a grid to “map” the surface properties.

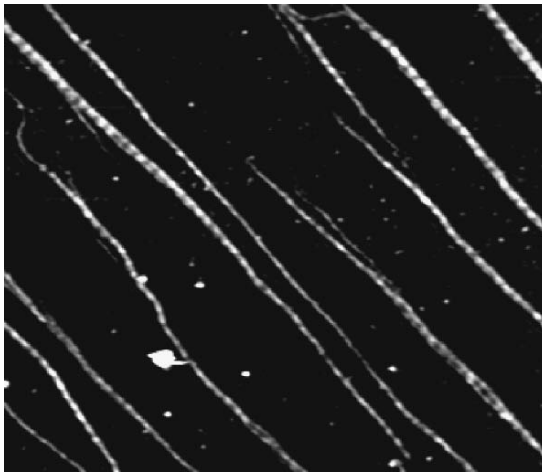


Figure 16–21. Collagen fibrils adsorbed on mica; sample courtesy of Müller; imaging JPK Instruments, intermittent contact mode in buffer. Scan area = $2.6 \times 2.3 \mu\text{m}$, z range = 2.5 nm. The 67-nm banding along the axis of the individual collagen fibrils can be seen.

Force spectroscopy in AFM refers to the measurement of tip–sample interaction forces at a point as the height of the cantilever base is varied. This allows measurement of forces either pushing into the sample (elasticity, rheology, etc.), or pulling away from the sample (adhesion, including recognition of specific molecular binding, unfolding, or stretching of molecules bound between the tip and the sample surface). Alternatively, manipulation of the surface, such as dissection or alignment, is possible using the tip to apply forces and modify the sample.

The ability to locally probe and manipulate a sample in addition to imaging is a unique strength of AFM. The local experiments are as diverse as the research questions and address both isolated macromolecular structures and cells. This is demonstrated for a few examples.

4.2.1 Pulmonary Surfactant

A mixed lipid–protein film of pulmonary surfactant covers the hydrated lung epithelia to the air. This highly cohesive and mechanically stable film reduces the otherwise high surface tension of the air–aqueous interface to almost zero as is required for the structural stability of the alveolar lung and ease of breathing. The mechanical stability of the film depends on a pattern of monolayer domains of disaturated phospholipids intercepted by multilayer areas, containing the unsaturated lipids also present in surfactant. This molecular architecture is conveyed to the film by the surfactant-specific proteins SP-B and/or SP-C. The proteins cross-link the multilayer areas to the monolayer.

This molecular arrangement has been discovered by AFM (von Nahmen et al., 1997). First, a topographical image was obtained (Figure 16–22). Then, the loading force was increased such as to physically remove the lipid stacks on top of the molecular monolayer. After removal of the multilayers, the monolayer showed crevices on the circumference of the earlier multilayer patch and holes, evenly

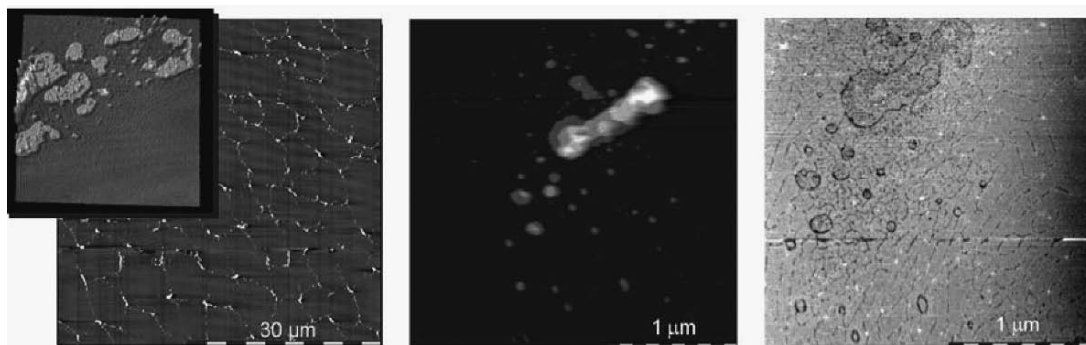


Figure 16–22. Functional pulmonary surfactant forms molecular films of molecular monolayer regions and areas of monolayers with stacks of lipid bilayers cross-linked to it (left). Cross-linking is revealed by physically removing the bilayers by an intentionally high load of the AFM probe (middle and right).

distributed over the area earlier covered. This was indicative of a cross-linking structure that had also been removed by the stylus. Mere lipid layers, formed on top of a lipid monolayer in the absence of surfactant proteins, left no traces behind after being removed by the stylus.

Multilayers, cross-linked to the monolayer as opposed to multilayers, merely adsorbed to the monolayer appear to be crucial for surfactant function. At low surface tension, the interfacial film is subjected to high lateral pressure. Cross-linked multilayered areas take up the lateral load together with the monolayer. Otherwise, the load resides on the monolayer alone. But lipid monolayer films containing unsaturated lipids are not able to withstand a high film pressure without collapsing. They therefore cannot reduce tension by a degree required for proper function of the lung. This condition is likely responsible for lung failure after acute lung injury where the surfactant is distinct by an elevated level of cholesterol (Karagiorga et al., 2006). Excess cholesterol prevents formation of cross-linked multilayer stacks and surface tension is high (Gunasekara et al., 2005).

4.2.2 Collagen Matrices

Müller and his associates (Jang et al., 2003) used the atomic force microscope to mechanically direct collagen into two-dimensional, 3-nm-thick layers of defined structure. During a critical, initial period of time this structure was found to be highly malleable and the collagen molecules could be reoriented by the stylus of the AFM. After this time, the collagen coating hardened and remained stable for several months without loss of fiber orientation or mechanical strength. It was proposed that collagen fiber formation *in vivo* may also have a critical period during which cellular forces remodel a collagen network that acts as a matrix for later tissue growth. In tissues, cells are indeed thought to organize the collagen fibrils by exerting tension on the matrix, and the realignment forces required in the experiment (300 pN) were within the range of forces that fibroblasts can generate. Besides explaining an important phenomenon on how tissue might organize *in vivo*, this discovery might also find technical applications. Engineered collagen matrices might find use as a platform for cell biological and tissue engineering applications (Figure 16–23).

4.2.3 Imaging and Unfolding a Bacterial Cell Wall Protein

Unfolding proteins and DNA and separating molecular binding partners have become important and active offspring of AFM, referred to as force spectroscopy. It provides insight into protein and DNA folding and conformation about the nature of binding pockets. Müller and his associates have used this technique in conjunction with high-resolution imaging to gain insight into the interaction forces between the individual protomers of a regular bacterial surface layer, the hexagonally packed intermediate (HPI) layer of *Deinococcus radiodurans*. After imaging the HPI layer, the AFM stylus was attached to individual protomers by enforced stylus–sample contact to allow force spectroscopy experiments. Imaging of the HPI layer after recording force–extension curves allowed unfolding forces to be

correlated with the structures being unfolded. By using this approach, individual protomers of the HPI layer were found to be removed at pulling forces of ≈ 300 pN. Furthermore, it was possible to sequentially unzip entire bacterial pores formed by six HPI protomers (Figure 16–24).

4.2.4 *Sphingolipid–Cholesterol Rafts*

In the plasma membrane, nanoscopic, condensed sphingolipid–cholesterol rafts “float” in a fluid matrix and likely act as prefabricated subunits of functional complexes. The functions these platforms perform involve essential cellular processes such as the initiation and down-regulation of signaling cascades that regulate cell growth, survival, and death, or endo- and exocytosis. A rapidly growing list of diseases has been identified as being related to rafts, including Alzheimer’s

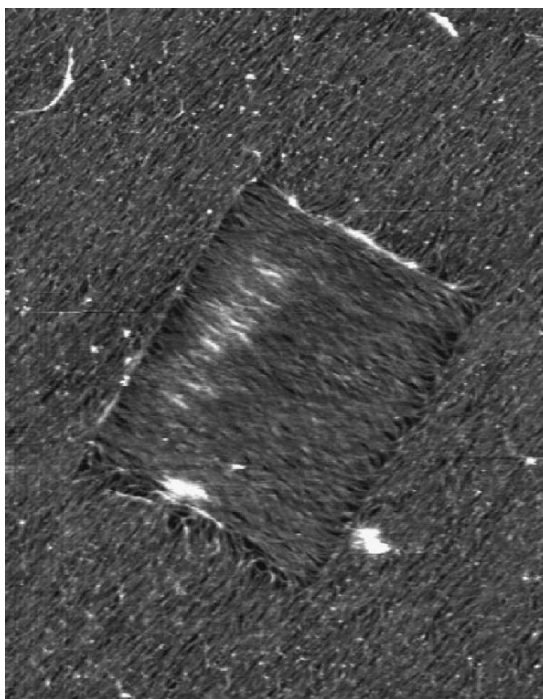


Figure 16–23. After collagen microfibrils were assembled out of solution on the supporting surface, their orientation has been changed in a controlled manner using the stylus of an AFM in the central region of the image. For this, the force was increased to 300 pN while the stylus was scanned over this operator-defined region. The collagen microfibrils in the region were now aligned with the fast-scanning direction of the AFM stylus and were approximately perpendicular to the original orientation. The surface was then reimaged at low force to create the image. (From Jiang et al., 2004b, reproduced with permission.)

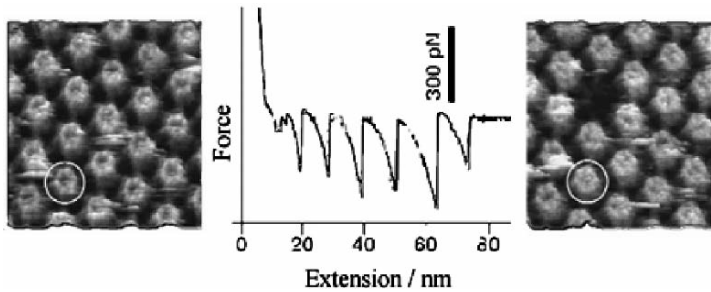


Figure 16–24. The six protomers of an individual pore can be sequentially pulled out of the HPI layer. Left: AFM topograph of the inner surface of the HPI layer prior to (left) and after (right) the pulling experiment. Note that one entire pore is missing. (Middle) The force-extension curve shows a sawtooth pattern with six force peaks of about 300 pN each corresponding to the extraction of one protomer. The height of the force peaks corresponds to the binding force for a protomer to its neighbors; the stretching distances between protomer disruption events (7.3 ± 1.6 nm) corresponds to the length of the molecular linker connecting a protomer to its neighbors. (From Engel et al., 2000, reproduced with permission.)

and Parkinson's diseases, muscular dystrophy, asthma and allergic responses, hypertension, arteriosclerosis, and prion diseases. Bacteria (e.g., *Vibrio cholerae*), viruses such as HIV-1, measles, and Ebola virus, as well as parasites (e.g., malaria) depend on lipid rafts for their purposes. Despite the outstanding importance of lipid rafts, most questions about how rafts are formed, composed, and behave in the plasma membrane of living cells are not well understood. Although several chemical groups that anchor proteins to rafts are known (e.g., the GPI, dipalmitoyl group), it is not known why, and to what extent, certain proteins become associated with one type and others to another kind of raft. But even the most basic questions regarding the size of rafts and whether they are transient structures are controversially discussed (Figure 16–25).

Strong experimental evidence of the existence of rafts as entities of defined size and with an extended lifetime has come from measuring the diffusion and stability over time of individual rafts in the plasma membrane of BHK-21 cells by photonic force microscopy, a variety of AFM (Hörber et al., 1992; Pralle et al., 2000). A bead in a laser trap was scanned over a cell sample. Like with the tip of a cantilever-based AFM, the height of the laser spot was readjusted by feedback when the bead was forced out of the focal spot of the laser trap by the sample. A topographical image of the sample was thus recorded. In a set of experiments, beads were coated with antibodies against a raft-associated protein. After binding to the cell, the laser trap was kept still and the diffusion of the bead attached to a raft was observed within the narrow confinement of the laser focal spot. The diffusion reflected the diffusion of the raft within the membrane and allowed the size of

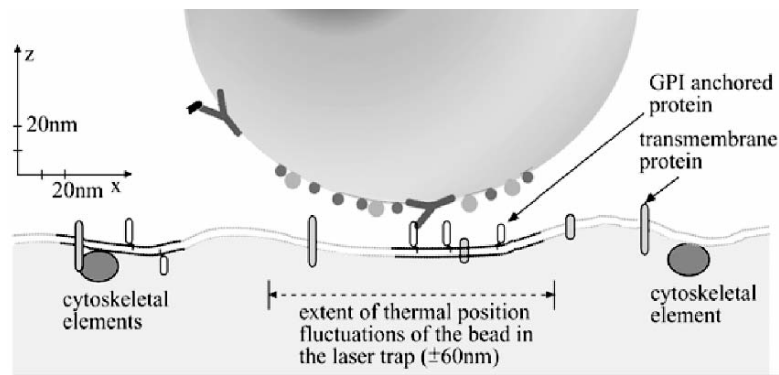


Figure 16–25. The size and temporal stability of a lipid raft may be derived by observing its diffusion within the plasma membrane. Observing the free diffusion of a raft is problematic, because rafts collide frequently with the elements of the cytoskeleton, anchored to the membrane. A more accurate picture is derived from observation of the constraint diffusion of a raft offered by photonic force microscopy. (From Pnalle et al., 2000, reproduced with permission.)

the raft to be derived. For those rafts addressed by the selected antibodies, these entities turned out to be about 50 nm in diameter. The experiment also showed that these rafts were not transient structures within the timeframe of the measurement (Figure 16–26).

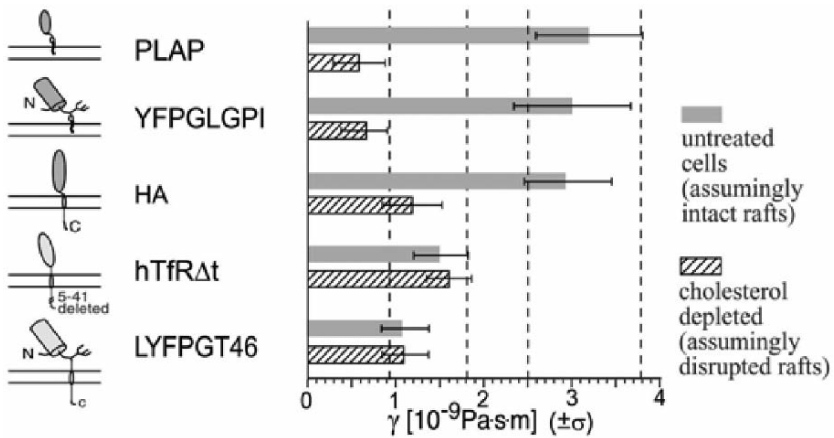


Figure 16–26. Diffusion coefficients for raft proteins (PLAP, YFPGLGPI, HA) and nonraft membrane proteins (hTfR Δ t, LYFPGT46). The raft proteins diffuse together with the lipid raft; the nonraft proteins diffuse freely in the fluid phase of the plasma membrane. Depriving cells in culture of one of the mandatory raft elements, cholesterol, abolishes rafts. As a consequence, the former raft proteins diffuse individually in the membrane. (From Pnalle et al., 2000, reproduced with permission.)

References

- Allison, D. and Thundat, T. (1993). Imaging entire genetically functional DNA molecules with the scanning tunneling microscope. *J. Vac. Sci. Technol. A* **11**, 816.
- Amrein, M., Stasiak, A., Gross, H. and Travaglini, G. (1988). Scanning tunneling microscopy of recA-DNA complexes coated with a conducting film. *Science* **240**, 514–516.
- Amrein, M., Dürr, R., Stasiak, S., Gross, H. and Travaglini, G. (1989b). Scanning tunneling microscopy of uncoated recA-DNA complexes. *Science* **243**, 1708–1711.
- Amrein, M., Dürr, R., Winkler, H., Travaglini, G., Wepf, R. and Gross, H. (1989b). STM of freeze-dried and Pt-Ir-C-coated bacteriophage T4 polyheads. *J. Ultrastruct. Mol. Struct. Res.* **102**, 170–177.
- Amrein, M., Gross, H. and Guckenberger, R. (1993). STM of proteins and membranes. In *STM and AFM in biology* (O. Marti and M. Amrein, Eds.), 127–175 (Academic Press, London).
- Amrein, M., von Nahmen, A. and Sieber, M. (1997). A scanning force- and fluorescence light microscopy study of the structure and function of a model pulmonary surfactant. *Eur. Biophys. J.* **26**(5), 349–357.
- Arscott, P.G. and Bloomfield, V.A. (1993). *STM of DNA and RNA* (Academic Press, San Diego).
- Bader, H., Dorn, K., Hupfer, B. and Ringsdorf, H. (1985). Polymeric monolayers and liposomes as models for biomembranes—how to bridge the gap between polymer science and membrane biology. *Adv. Polymer Sci.* **64**, 1–62.
- Baró, A.M., Miranda, R., Alaman, J., Garcia, N., Binnig G., Rohrer, H., Gerber, C. and Carracosa, J.L. (1985). Determination of surface topography of biological specimens at high resolution by scanning tunnelling microscopy. *Nature* **315**, 253–254.
- Baumeister, W., Barth, M., Hegerl, R., Guckenberger, R., Hahn, M. and Saxton, W.O. (1986). Three-dimensional structure of the regular surface layer (HPI layer) of *Deinococcus radiodurans*. *J. Mol. Biol.* **187**, 241–253.
- Binnig, G. and Rohrer, H. (1982). Scanning tunneling microscopy. *Helv. Phys. Acta* **55**, 726–735.
- Binnig, G. and Smith, D.P.E. (1986). Single-tube three-dimensional scanner for scanning tunneling microscopy. *Rev. Sci. Instrum.* **57**(8), 1688–1689.
- Blackford, B.L. and Jericho, M.H. (1991). A metallic replica/anchoring technique for scanning tunneling microscope or atomic force microscope imaging of large biological structures. *J. Vac. Sci. Technol. B* **9**(2), 1253–1258.
- Blodgett, K.A. and Langmuir, I. (1937). *Phys. Rev.* **51**, 2492.
- Brisson, A., Olofsson, A. and Ringler, P. (1994). Two-dimensional crystallization of proteins on planar lipid films and structure determination by electron crystallography. *Biol. Cell* **80**, 221–228.
- Budó, A. (1976). *Theoretische Mechanik* (VEB Deutscher Verlag der Wissenschaften, Berlin).
- Butt, H.-J. (1991). Measuring electrostatic, van der Waals, and hydration forces in electrolyte solutions with an atomic force microscope. *Biophys. J.* **60**, 1438–1444.
- Butt, H.-J. (1992). Measuring local surface charge densities in electrolyte solutions with a scanning force microscope. *Biophys. J.* **63**, 578–582.
- Butt, H.-J., Jaschke, M. et al. (1995). Measuring surface forces in aqueous solution with the atomic force microscope. *Bioelect. Bioenerg.* **38**, 191–201.
- Chang, L., Kious, T., Yorgancioglu, M., Keller, D. and Pfeiffer, J. (1993). Cytoskeleton of living, unstained cells imaged by scanning force microscopy. *Biophys. J.* **64**, 1282–1286.

- Ducker, W.A., Senden, T.J. and Pashley, R.M. (1991). Direct measurements of colloidal forces using an atomic force microscope. *Nature* **353**, 239–241.
- Engel, A. and Müller, D.J. (2000). Observing single biomolecules at work with the atomic force microscope. *Nature Struct. Biol.* **7**(9), 715–718.
- Engel, A., Hoenger, A., Hefti, A., Henn, C., Ford, R.C., Kistler, F. and Zulauf, M. (1992). Assembly of 2-D membrane protein crystals: dynamics, crystal order, and fidelity of structure analysis by electron microscopy. *J. Struct. Biol.* **109**, 219–234.
- Fisher, K.A., Yanagimoto, K. and Stoeckenius, W. (1977). Purple membrane bound to polylysine glass: effects of pH and light. *J. Cell Biol.* **75**, 200a.
- Fisher, K.A., Yanagimoto, K. and Stoeckenius, W. (1978). Oriented adsorption of purple membrane to cationic surfaces. *J. Cell Biol.* **77**, 611–621.
- Fritz, M., Radmacher, M. and Gaub, H.E. (1994). Granula motion and membrane spreading during activation of human platelets imaged by atomic force microscopy. *Biophys. J.* **66**(5), 1328–1334.
- Guckenberger, R., Wiegräbe, W., Hillebrand, A., Hartmann, T., Wang, Z. and Baumeister, W. (1989). Scanning tunneling microscopy of a hydrated bacterial surface protein. *Ultramicroscopy* **31**, 327–332.
- Guckenberger, R., Heim, M., Cevc, G., Knass, H.F., Weigräbke, W. and Hillebrand, A. (1994). Scanning tunneling microscopy of insulators and biological specimens based on lateral conductivity of ultrathin water films. *Science* **266**, 1538–1540.
- Gunasekara, L., Schürch, S., Schoel, W.M., Nag, K., Leonenko, Z., Haufs, M. and Amrein, M. (2005). Pulmonary surfactant function is abolished by an elevated proportion of cholesterol. *Biochim. Biophys. Acta* **1737**, S27–35.
- Haltia, T. and Freire, E. (1995). Forces and factors that contribute to the structural stability of membrane proteins. *Bba-Bioenergetics* **1228**(1), 1–27.
- Han, W., Lindsay, S.M. and Jing, T.W. (1996). A magnetically driven oscillating probe microscope for operation in liquids. *Appl. Phys. Lett.* **69**(26), 4111–4113.
- Hansma, P. (1996). New AFM technology for better biological imaging. *Am. Phys. Soc.* **41**, 515.
- Hansma, P.K., Cleveland, J.P., Radmacher, M., Walters, D.A., Hillner, P.E., Bezantilla, M., Fritz, M., Vie, D., Hansma, H.G. and Prater, C.B. (1994). Tapping mode atomic force microscopy in liquids. *Appl. Phys. Lett.* **64**(13), 1738–1740.
- Harris, C.M. (1987). *Shock and Vibration Handbook* (McGraw-Hill, New York).
- Hayward, S.B., Grano, D.A., Glaeser, R.M. and Fisher, K.A. (1978). Molecular orientation of bacteriorhodopsin within the purple membrane of *Halobacterium halobium*. *Proc. Natl. Acad. Sci. USA* **75**(9), 4320–4324.
- Hegner, M., Wagner, P. and Semenza, G. (1993). Immobilizing DNA on gold via thiol modification for atomic force microscopy imaging in buffer solutions. *FEBS Lett* **336**(3), 452–456.
- Hegner, M., Dreier, M., Wagner, P., Semenza, G. and Güntherodt, H.J. (1996). Modified DNA immobilized on bioreactive self-assembled monolayer on gold for dynamic force microscopy imaging in aqueous buffer solution. *J. Vac. Sci. Technol.* **14**, 1418–1421.
- Henderson, E., Haydon, P.G. and Sakaguchi, D.S. (1992). Actin filament dynamics in living glial cells imaged by atomic force microscopy. *Science* **257**, 1944–1945.
- Heuberger, M., Dietler, G. and Schlapbach, L. (1996). Elastic deformations of tip and sample during atomic force microscope measurements. *J. Vac. Sci. Technol. B* **14**(2), 1250–1254.
- Hoh, J.H. and Schoenenberger, C.-A. (1994). Surface morphology and mechanical properties of MDCK monolayers by atomic force microscopy. *J. Cell Sci.* **107**(5), 1105–1114.

- Hoh, J.H., Sosinski, G.E., Revel, J.-P. and Hansma, P.K. (1993). Structure of the extracellular surface of the junction by atomic force microscopy. *Biophys. J.* **65**, 149–163.
- Hörber, J.K.H., Häberle, W., Ohnesorge, F., Binnig, G., Liebich, H.G., Czerny, C.P., Bähnel, H. and Mayr, A. (1992). Investigation of living cells in the nanometer regime with the scanning force microscope. *Scan. Microsc.* **6**(4), 919–930.
- Israelachvili, J. (1991). *Intermolecular & Surface Forces* (Academic Press, London).
- Jacobson, B.S. and Branton, D. (1977). Plasma membrane: rapid isolation and exposure of the cytoplasmic surface by use of positively charged beads. *Science* **195**, 302–304.
- Jap, B.K., Zulauf, M., Scheybani, T., Hefti, A., Baumeister, W., Aepli, U. and Engel, A. (1992). 2D crystallization: from art to science. *Ultramicroscopy* **46**, 45–84.
- Jiang, F., Hörber, H., Howard, J. and Muller, D.J. (2004a). Assembly of collagen into microribbons: effects of pH and electrolytes. *J. Struct. Biol.* **148**(3), 268–278.
- Jiang, F.Z., Khairy, K., Poole, K., Howard, J. and Muller, D.J. (2004b). Creating nanoscopic collagen matrices using atomic force microscopy. *Microsc. Res. Techni.* **64**, 435–440.
- Karagiorga, G., Nakos, G., Galiatson, E. and Lekka, M.E. (2006). Biochemical parameters of bronchoalveolar lavage fluid in fat embolism. *Intensive Care med.* **32**, 116–123.
- Karrasch, S., Dolder, M., Schabert, F., Ramsden, J. and Engel, A. (1993). Covalent binding of biological samples to solid supports for scanning probe microscopy in buffer solution. *Biophys. J.* **65**, 2437–2446.
- Kasas, S. and Ikai, A. (1995). A method for anchoring round shaped cells for atomic force microscope imaging. *Biophys. J.* **68**(5), 1678–1680.
- Kellenberger, E. and Kistler, J. (1979). The physics of specimen preparation. In *Advances in Structure Research by Diffraction Methods* (W.H.R. Mason, Ed.), III, 49–79 (Friedrich Vieweg, Wiesbaden).
- Kellenberger, E., Häner, M. and Wurtz, M. (1982). The wrapping phenomenon in air-dried and negatively stained preparations. *Ultramicroscopy* **9**, 139–150.
- Knapp, H., Wiegräbe, W., Heim, M., Eschrich, R. and Guckenberger, R. (1995). Atomic force microscope measurements and manipulation of Langmuir-Blodgett films with modified tips. *Biophys. J.* **69**, 708–715.
- Knebel, D., Amrein, M. and Reichelt, R. (1997). A fast and versatile scan-unit for scanning probe microscopy. *Scanning* **19**(4), 264–268.
- Knebel, D., Sieber, M., Reichelt, R., Galla, H.J. and Amrein, M. (2002). Scanning force microscopy at the air-water interface of an air bubble coated with pulmonary surfactant. *Biophys. J.* **82**(1 Pt 1), 474–480.
- Lantz, M.A., O'Shea, S.J. and Welland, M.E. (1994). Force microscopy imaging in liquids using ac techniques. *Appl. Phys. Lett.* **65**(4), 409–411.
- Lindsay, S.M. and Tao, N.J. (1993). Potentiostatic deposition of molecules for STM. In *STM and AFM in Biology* (O. Marti and M. Amrein, Eds.), 229–258 (Academic Press, London).
- Matsko, N. and Mueller, M. (2004). AFM of biological material embedded in epoxy resin. *J. Struct. Biol.* **146**(3), 334–343.
- Mazia, D., Schatten, G. and Sale, W. (1975). Adhesion of cells to surfaces coated with polylysine: applications to electron microscopy. *J. Cell. Biol.* **66**, 198–200.
- Miles, M.J. and McMaster, T.J. (1993). *Protein Assemblies and Single Molecules Imaged by STM* (Academic Press, San Diego).

- Miles, M.J., McMaster, T., Carr, H.J., Tatham, A.S., Shewry, P.R., Field, J.M., Belton, B.S., Jeenes, D., Hanley, B. and Whittam, M. (1990). Scanning tunneling microscopy of biomacromolecules. *J. Vac. Sci. Technol. A* **8**(1), 698–702.
- Müller, D.J. and Büldt, G. (1995). Force-induced conformational change of bacteriorhodopsin. *J. Mol. Biol.* **249**, 239–243.
- Müller, D.J., Baumeister, W. and Engel, A. (1996). Conformational change of the hexagonally packed intermediate layer imaged by atomic force microscopy. *J. Bacteriol.* **78**, 3025–3030.
- Müller, D.J., Amrein, M. and Engel, A. (1997a). Adsorption of biological molecules to a solid support for scanning probe microscopy. *J. Struct. Biol.* **119**(2), 172–188.
- Müller, D.J., Engel, A. and Amrein, M. (1997b). Preparation techniques for the observation of native biological systems with the atomic force microscope. *Biosensors Bioelectron.* **12**(8), 867–877.
- Nahmen von, A., Schenk, M., Sieber, M. and Amrein, M. (1997). The structure of a model pulmonary surfactant as revealed by scanning force microscopy. *Biophys. J.* **72**, 463–469.
- Poole, K., Khairy, K., Friedrichs, J., Franz, C., Cisneros, D.A. Howard, J. and Mueller, D. (2005). Molecular-scale topographic cues induce the orientation and directional movement of fibroblasts on two-dimensional collagen surfaces. *J. Mol. Biol.* **349**, 380–386.
- Pralle, A., Keller, P., Florin, E.-L., Simons, K. and Horber, J.K.H. (2000) Sphingolipid-cholesterol rafts diffuse as small entities in the plasma membrane of mammalian cells. *J. Cell Biol.* **148**, 997–1008.
- Putman, C.A.J., Van der Werf, K.O., Degrooth, B.G., Vanhulst, N.F. and Greve, J. (1994). Tapping mode atomic force microscopy in liquid. *Appl. Phys. Lett.* **64**(18), 2454–2456.
- Sarid, D., Ed. (1991). *Scanning Force Microscopy with Application to Electric, Magnetic and Atomic Forces* (Oxford University Press, New York).
- Schabert, F.A. and Engel, A. (1994). Reproducible acquisition of *Escherichia coli* porin surface topographs by atomic force microscopy. *Biophys. J.* **67**, 2394–2403.
- Schäffer, T.E., Cleveland, J.P., Ohnesorge, F., Walters, D.A. and Hansma, P.K. (1996). Studies of vibrating atomic force microscope cantilevers in liquid. *J. Appl. Phys.* **80**(7), 3622–3627.
- Schenk, M., Amrein, M., Dietz, P. and Reichelt, R. (1994). *Simultaneous STM/AFM Imaging of Biological Macromolecular Assemblies: Experimental Set-up and First Results*. 13th International Congress on Electron Microscopy (Les Editions de Physique Les Ulis, Paris).
- Schenk, M., Amrein, M. and Reichelt, R. (1996). An electret microphone as a force sensor for combined scanning probe microscopies. *Ultramicroscopy* **65**, 109–118.
- Seelert, H., Poetsch, A., Dencher, N.A., Engel, A., Stahlberg, H. and Muller, D.J. (2000). Structural biology: proton-powered turbine of a plant motor. *Nature* **405**, 418–419.
- Sellers, H., Ulman, A., Schnidman, Y. and Eilers, J.E. (1993). Structure and binding of alkanethiolates on gold and silver surfaces—implications for self-assembled monolayers. *J. Am. Chem. Soc.* **115**(21), 9389–9401.
- Stemmer, A., Hefti, A., Aepli, U. and Engel, A. (1989). Scanning tunneling and transmission electron microscopy on identical areas of biological specimens. *Ultramicroscopy* **30**, 263–280.
- Tortorese, M., Barret, R.C. and Quate, C.F. (1993). Atomic resolution with an atomic force microscope using piezoresistive detection. *Appl. Phys. Lett.* **62**, 834–836.

- Travaglini, G., Rohrer, H., Amrein, M. and Gross, H. (1987). Scanning tunneling microscopy on biological matter. *Surf. Sci.* **181**, 380–390.
- Voelker, M.A., Hameroff, S.R., He, J.D., Dereniak, E.L., McCuskey, R.S., Schneiker, C.W., Chvapil, T.A., Bell, T.S. and Weiss, L.B. (1988). STM imaging of molecular collagen and phospholipid membranes. *J. Microsc.* **152**, 557–573.
- Wagner, D., Moy, V.T., Hofman, U.G., Benoit, M., Ludwig, M. and Gaub, H.E. (1994). *AFM—The Art of Touching Molecules (Invited)*. 13th International Congress on Electron Microscopy (Les Editions de Physique Les Ulis, Paris).
- Wagner, P., Hegner, M., Guntherodt, H.J. and Semenza, G. (1995a). Formation and in-situ modification of monolayers chemisorbed on ultraflat template-stripped gold surfaces. *Langmuir* **11**(10), 3867–3875.
- Wagner, P., Hegner, M., Kernen, P., Zaug, F. and Semenza, G. (1995b). Covalent immobilization of native biomolecules onto Au (111) via N-hydroxysuccinimide ester functionalized self-assembled monolayers for scanning probe microscopy. *Biophys. J.* **70**, 2052–2066.
- Wang, Z., Hartmann, T., Baumeister, W. and Guckenberger, R. (1990). Thickness determination of biological samples with a z-calibrated scanning tunneling microscope. *Proc. Natl. Acad. Sci. USA* **87**, 9343–9347.
- Weisenhorn, A.L., Maivald, P., Butt, H.-J. and Hansma, P.K. (1992). Measuring adhesion, attraction, and repulsion between surfaces in liquids with an atomic-force microscope. *Phys. Rev. B* **45**(19), 11226–11232.
- Weisenhorn, A.L., Khorsandi, M., Kasas, S., Gotzos, V. and Butt, H.-J. (1993). Deformation and height anomaly of soft surfaces studied with an AFM. *Nanotechnology* **4**, 106–113.
- Welland, M.E., Miles, M.J., Lambert, N., Morris, V.J., Loombs, J.H. and Pethia, J.B. (1989). Structure of the globular protein vicilin revealed by scanning tunnelling microscopy. *Int. J. Biol. Macromol.* **11**(1), 29–32.
- Wildhaber, I., Gross, H., Engel, A. and Baumeister, W. (1985). The effects of air-drying and freeze-drying of the structure of a regular protein layer. *Ultramicroscopy* **16**, 411–422.
- Wolf, H., Ringsdorf, H., Delamarche, E., Takami, T., Kang, H., Michel, B., Gerber, C., Jaschke, M., Butt, H.J. and Bamberg, E. (1995). End-group-dominated molecular order in self-assembled monolayers. *J. Phys. Chem.* **99**(18), 7102–7107.
- Yang, J., Mou, J. and Shao, Z. (1996). The effect of deformation on the lateral resolution of force microscopy. *J. Microsc.* **182**(2), 106–113.
- Zasadzinski, J.A.N., Schneir, J., Gurley, J., Elings, V. and Hansma, P.K. (1988). Scanning tunneling microscopy of freeze-fracture replicas of biomembranes. *Science* **239**, 1013–1015.

Low-Temperature Scanning Tunneling Microscopy

Uwe Weierstall

1 Introduction

The scanning tunneling microscope (STM) has revolutionized surface science since its invention in 1982 (Binnig and Rohrer, 1982) by providing a means to directly image atomic scale spatial and electronic structure. Using the combination of a coarse approach and piezoelectric transducers, a sharp, metallic tip is brought into close proximity with the sample. The distance between tip and sample is less than 1 nm, which means that the electron wave functions of tip and sample start to overlap. A bias voltage is applied between tip and sample that causes electrons to tunnel through the barrier. The tunneling current is a quantum mechanical effect: tunneling of electrons can occur between two electrodes separated by a thin insulator or a vacuum gap and the tunneling current decays on the length scale of one atomic radius. The tunneling current is in the range of picoamperes to nanoamperes and is measured with a preamplifier. In an STM, the tip is scanned over the surface and electrons tunnel from the very last atom of the tip apex to single atoms on the surface, providing atomic resolution. The exponential dependence of the tunneling current on the tip-sample distance can be exploited to control the tip-sample distance with high precision. There are four basic operation modes for any STM: constant current imaging, constant height imaging, spectroscopic imaging, and local spectroscopy. Their interpretation and realization will be briefly discussed below. For details about other modes and a comprehensive introduction to electron tunneling and STM see Wiesendanger (1994).

To acquire constant current images, a feedback loop adjusts the height of the tip during scanning so that the tunneling current flowing between tip and sample is kept constant. The height z is adjusted by applying an appropriate voltage V_z to the z -piezoelectric drive while the lateral tip position (x,y) is determined by the corresponding voltages applied to the x and y piezoelectric drives. The recorded signal V_z can be translated into a topography $z(x,y)$ if the sensitivity of the piezoelectric drives is known. The word

topography should be used with caution: since the local density of states at the Fermi level is measured, a molecule adsorbed on a metal surface that reduces the local density of states and may actually be imaged as a depression.

To acquire constant height images, the feedback loop is switched off, i.e., the tip is scanned at constant height above the surface, and variations in the current are measured. This mode has the advantage that the finite response time of the feedback loop does not limit the scan speed. It can be used to collect images at video rates, offering the opportunity to observe dynamic processes at surfaces. However, thermal drift limits the time of the experiment and there is an increased risk of crashing the tip.

To measure differential conductance (dI/dV) maps with the STM, a high-frequency sinusoidal modulation voltage is superimposed on the constant dc bias voltage V_{bias} between tip and sample. The modulation frequency is chosen higher than the cutoff frequency of the feedback loop, which keeps the tunneling current constant. By recording the tunneling current modulation, which is in phase with the applied bias voltage modulation, with a lock in amplifier, a spatially resolved spectroscopic signal $dI/dV|_{V_{\text{bias}}}$ can be obtained simultaneously with the constant current image (Binnig et al., 1985a,b).

By measuring the differential conductance dI/dV at a fixed tip position with open feedback loop (constant tip-sample distance z) while sweeping the applied bias voltage, an energy-resolved spectrum can be obtained. This is useful for probing, e.g., band-gap states in semiconductors or the onset of surface states on metals.

The tunneling current I at a given tip position is approximately equal to the integrated local density of states (ILDOS), integrated over the energy range between the Fermi energy E_F of the sample and eV , where V is the applied bias voltage. Therefore the differential conductance dI/dV is approximately proportional to the local density of states (LDOS) of the sample at the energy eV , and a constant current image should represent a contour of constant ILDOS. For measurements close to E_F , i.e., at low bias voltages, the LDOS and ILDOS are essentially the same and a constant current image at low bias (a few millivolts) is therefore approximately proportional to the sample LDOS at the Fermi energy E_F (assuming the tip has a uniform density of states and the temperature is low). To illustrate how to arrive at the picture presented above, the theoretical treatment of electron tunneling is briefly outlined.

A one-dimensional WKB approximation predicts that the tunneling current at low temperatures (where the Fermi distribution is a step function) is given by

$$I = \int_0^{eV} \rho_s(E, x) \rho_t(-eV + E, x) T(E, eV, x) dE \quad (1)$$

where $\rho_s(E)$ and $\rho_t(E)$ are the density of states of the sample and the tip at the location x and energy E , measured with respect to their individual Fermi levels, and V is the applied bias voltage (Hamers, 1989). The

tunneling transmission probability $T(E, eV, x)$ for electrons with energy E and applied voltage V is given by

$$T(E, eV, x) = \exp\left(-z(x)\sqrt{\frac{4m}{\hbar^2}(\phi_t + \phi_s + eV - 2E)}\right) \quad (2)$$

where ϕ_s and ϕ_t are the work functions of sample and tip and z is the tip-sample distance. Therefore, assuming that the tip electronic structure is featureless, Eq. (1) shows that the tunneling current at position x is approximately equal to the ILDOS of the sample integrated between E_F and eV , weighted by the transmission probability T . Examination of Eq. (2) shows that if $eV < 0$ (i.e., negative sample bias), the transmission probability is largest for $E = 0$ (corresponding to electrons at the Fermi level of the sample). If $eV > 0$ (positive sample bias) the probability is largest for $E = eV$ (corresponding to electrons at the Fermi level of the tip). Therefore the tunneling probability is always largest for electrons at the Fermi level of whichever electrode is negatively biased. This is shown schematically in Figure 17-1.

Differentiating Eq. (1) gives the differential conductance

$$\begin{aligned} \frac{dI}{dV} &\propto \rho_s(eV, x)\rho_t(0, x)T(eV, eV, x) \\ &+ \int_0^{eV} \rho_s(E, x)\rho_t(E - eV, x)\frac{dT(E, eV, x)}{dV}dE \end{aligned} \quad (3)$$

The first term is the product of the density of states of the sample, the density of states of the tip, and the tunneling transmission probability. The second term contains the voltage dependence of the tunneling transmission probability. Since T is a smooth monotonically

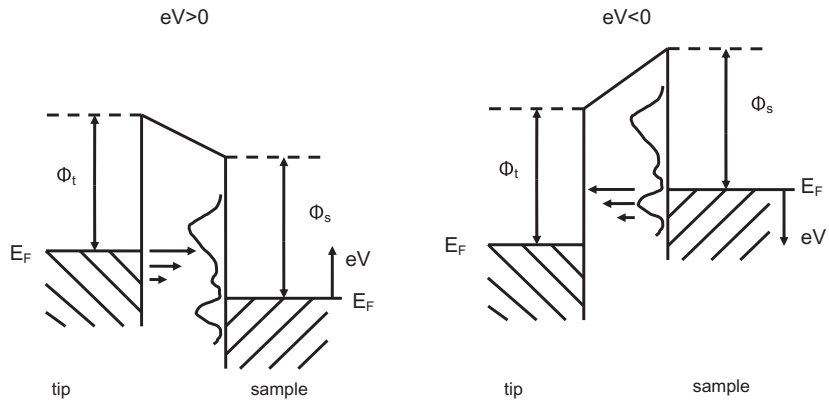


Figure 17-1. Energy level diagram of sample and tip with V being the sample voltage relative to the tip. Left: electrons tunneling from tip to sample with positive sample bias voltage. Right: electrons tunneling from sample to tip with negative sample bias voltage. The curve represents the density of states of the sample; the tip density of states is assumed featureless. The different lengths of the arrows illustrate the fact that the tunneling probability is largest for electrons at the Fermi level of the negatively biased electrode. Φ_t and Φ_s are the tip and sample work functions.

increasing function of V , structure in dI/dV can be assigned to changes in the sample LDOS at eV , assuming a tip with featureless density of states.

In general, however, the tip electronic structure is unknown and may not be featureless. The small size of the STM tip is expected to significantly modify its electronic structure from that of a bulk material. Therefore, it is usually necessary to compare tunneling spectra acquired at different locations on the surface to distinguish the spatially invariant contribution of the tip and the spatially varying contribution from the sample. Comparing spectra taken with different tips also helps to eliminate tip contributions.

At finite temperatures Eq. (1) contains an additional Fermi distribution factor, which imposes a limit on spectroscopic resolution. At room temperature, with $k_B T \approx 0.026$ eV, the spread of the tip and sample energy distribution are each $2k_B T \approx 0.052$ eV. Therefore the total spread is $\Delta E \approx 4k_B T \approx 0.1$ eV. To make high-resolution spectroscopic measurements with ΔE in the millielectron volt range, experiments must be conducted at cryogenic temperatures.

To work with clean metal and semiconductor surfaces, STM measurements have to be done in ultrahigh vacuum (UHV). Operating a UHV-STM at cryogenic temperature has several major advantages:

1. Close coupling of the microscope to a large temperature bath that keeps constant temperature over hours or days ensures a reduction of thermal drift and allows long-term measurements. The reduction of thermal drift during low-temperature operation is further improved by the fact that the thermal expansion coefficients at liquid helium temperature are two or more orders of magnitude smaller than at room temperature.

2. Superior vacuum conditions: If the microscope is incorporated in a cryostat that acts as an effective cryopump, surfaces are kept free from contamination over days.

3. Thermal diffusion of adsorbates and defects is suppressed—stable imaging becomes possible even for weakly bound species. Moreover, low temperatures might also stabilize the atomic configuration at the tip end by preventing sudden jumps of the most loosely bound foremost atoms due to thermal activation.

4. Small thermal broadening at the Fermi energy is a necessary condition for spectroscopic investigations with high-energy resolution.

5. Individual adsorbates can be manipulated with the STM to qualitatively probe their interaction with the substrate.

6. Physical properties can be studied as a function of temperature or physical effects can be examined that occur only at low temperatures (e.g., superconductivity, Kondo effect, nanoscale magnetism).

7. Piezo nonlinearities and hysteresis (creep) affecting the piezoelectric scanners of the STM decrease substantially at low temperatures.

In an effort to improve the stability of the microscope and the atoms or molecules under investigation, a variety of low-temperature UHV STM designs have been developed.

2 Design Principals

All low-temperature STMs (LT-STMs) operate in UHV, which is a prerequisite for obtaining a clean surface. To reach a low final temperature and a short cooling time, the thermal conductivity from the microscope to the cryogen has to be maximized, and the thermal load from room temperature has to be minimized. Heat transfer through the electrical connections and contacts has to be considered as well as thermal radiation. The discussion here is restricted to liquid helium (LHe)-cooled instruments. In designing an LT-STMs, there is a choice between a flow cryostat and a bath cryostat to cool the STM. Then there are two basic designs: one in which only the sample is cooled and one in which the whole STM is cooled and surrounded by a radiation shield.

2.1 STM with LHe Continuous-Flow Cryostat

If the ability to change temperature on a relatively short time scale is of importance, the heat reservoir should be small. Therefore in variable-temperature STMs that can work from room temperature down to liquid helium temperature, a flow cryostat is usually used and only the sample is cooled to ensure a small thermal mass. The sample is thermally connected to the cryostat via a flexible Cu or Au braid. Cryostat vibrations and instabilities caused by boiling of the coolant as well as different thermal expansion coefficients of various materials during temperature cycling require special attention. To reduce the amplitude of vibrations introduced by the cryostat, it is important to mechanically decouple the braid to a heavy mass (Bott et al., 1995) (see Figure 17-2). Since only the sample is cooled, the scanner has to be thermally isolated from the sample to avoid heat transfer to the sample. One advantage of this approach for variable-temperature operation is that the repeated recalibrations of the STM made necessary by temperature-dependent changes in the piezo coefficients are avoided. But large thermal gradients can create image drift problems and possible disturbances due to heat transfer from the “hot” tip scanning the cold sample (Xu et al., 1994).

An example of a variable-temperature STM with flow cryostat cooling the sample only is shown in Figure 17-3 (Behler et al., 1997). Other examples are given in Bott et al. (1995), Horch et al. (1994), and Petersen et al. (2001).

LHe flow cryostats can also be used to cool the entire STM. This results in improved thermal stability due to thermal equilibrium between all STM parts. If fast cooldown times are required, the STM can be connected rigidly to the cold end of the flow cryostat (Mugele et al., 1998; Zhang et al., 2001). This mandates a high mechanical stability (high resonance frequency) of the STM so that the vibration frequencies caused by the He flow are well below the eigenfrequency of the STM. The microscope is surrounded by one or two radiation shields. Higher mechanical stability can be obtained if the STM is suspended with springs inside a radiation shield to provide vibration isolation

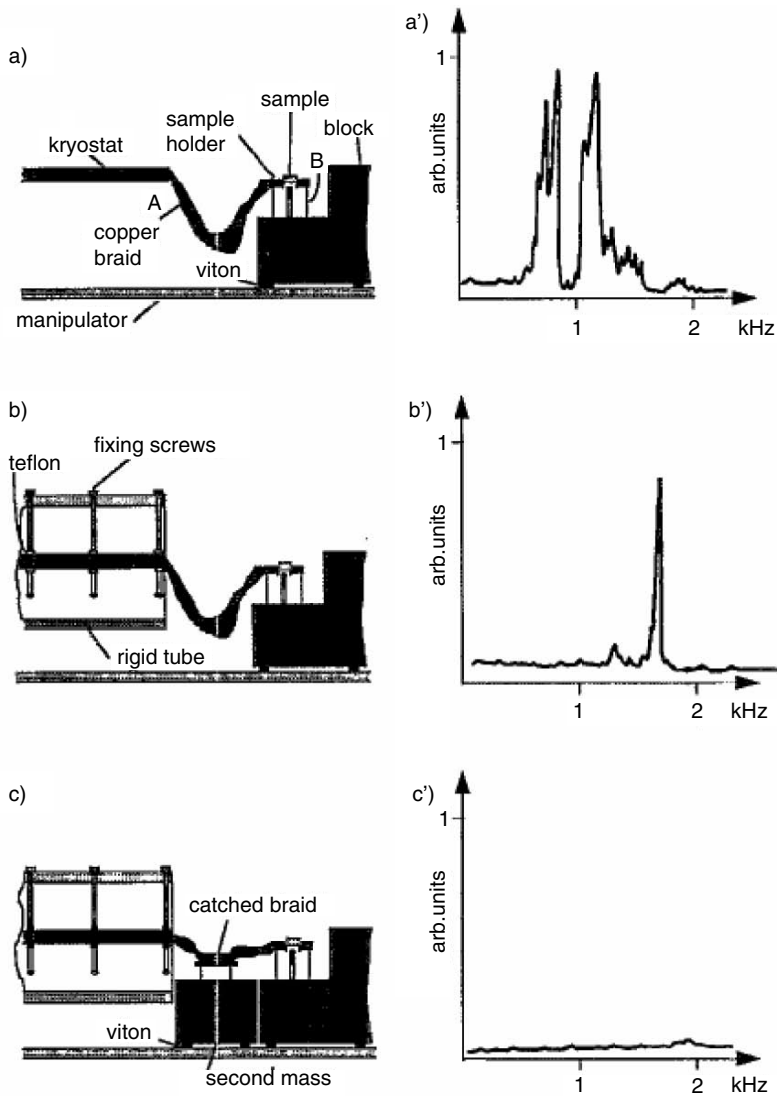


Figure 17–2. Mechanical decoupling of a copper braid connecting a flow cryostat and a sample holder. Shown are modifications and their effect on the noise spectrum. (a) Initial arrangement of the sample holder connected to the cryostat by a copper braid. (a') Fourier spectrum of the tunneling current with noise showing up as peaks. (b) Resonance frequencies of cryostat are increased by defining nodes with a rigid tube mounted on the cryostat with three screws separated by 120° at each node. (b') Fourier spectrum after modification; noise frequencies are shifted to higher frequencies. (c) Noise from the copper braid coupling is damped out by fixing the copper braid thermally isolated to a massive block. (c') Noise is greatly reduced. (From Bott et al., 1995.)

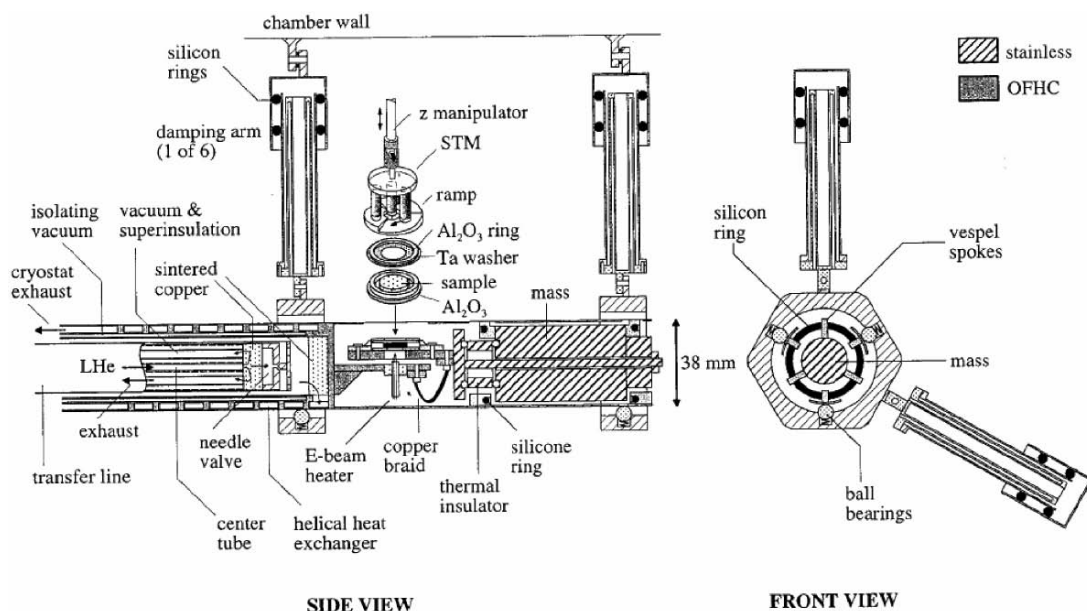


Figure 17-3. Example of a variable-temperature STM for operation in the temperature range of 20–300 K. Only the sample is cooled. Continuous flow cryostat on the left; arrows show helium flow. The sample is thermally connected to the cryostat by a copper braid and mechanically decoupled with a heavy stainless steel mass supported by O-rings. (From Behler et al., 1997.)

from the cryostat (Wolkow, 1995; Stipe et al., 1999b) (see Figures 17–4 and 17–5).

2.2 STM with LHe Bath Cryostat

A bath cryostat is usually used if the whole STM is being cooled. The bath cryostat serves as a large thermal reservoir that is being cooled down prior to the measurements. This approach has been successfully applied on a number of 4 K microscopes (Eigler and Schweizer, 1990; Gaisch et al., 1992; Stranick et al., 1994b; Rust et al., 1997; Meyer, 1996; Becker et al., 1998; Ferris et al., 1998; Harrell and First, 1999; Stroscio, 2000). If all parts of the system are allowed to reach thermal equilibrium, thermal drift may be virtually eliminated. Bath cryostat low-temperature STMs operating in a rotatable magnetic field have also been built (Wittneven et al., 1997); some of them are ^3He refrigerated and operate at about 250 mK (Pan et al., 1999; Kugler et al., 2000; Matsui et al., 2003). Sample turnaround times for the very-low-temperature variants are quite long since it can take 36 h to reach thermal equilibrium (Pan et al., 1999). LT-STMs with bath cryostats are difficult to use for variable temperature measurements, e.g., study of diffusion and phase transitions, since they use a large cold reservoir to cool the STM and the temperature cannot be changed easily. Temperature control by altering the exchange gas pressure in combination with a PID-controlled heater has been achieved (Rust et al., 2001). Another way to achieve measurements at different temperatures is to simply remove

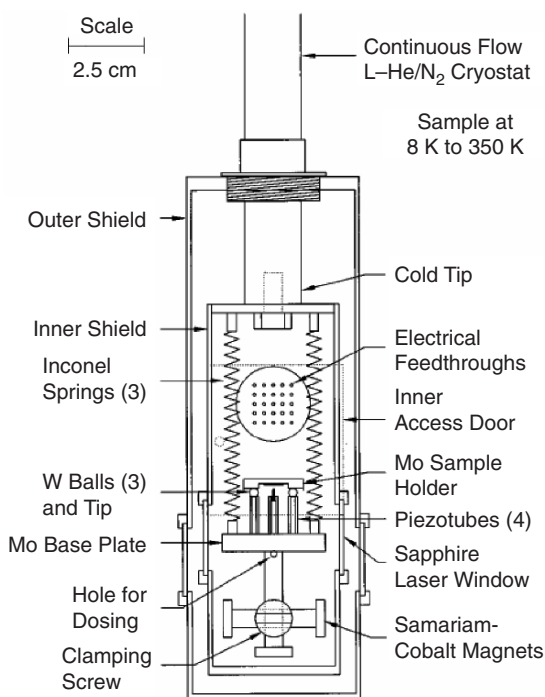


Figure 17-4. A variable-temperature scanning tunneling microscope cooled with a continuous-flow cryostat. The STM and the sample are suspended from three springs that provide the second stage of vibrational isolation. The design allows for *in situ* dosing and irradiation of the sample as well as the exchange of samples and tips. (From Stipe et al., 1999b.)

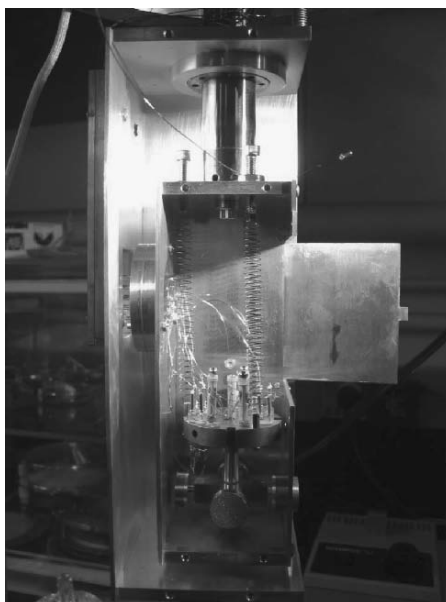


Figure 17-5. Low and variable temperature STM in the author's laboratory, which is very similar in design to the one by Stipe et al. (1999b). Refer to Figure 17-4 for identification of different parts. The front of the outer radiation shield and the sides of both radiation shields are missing to enable the inside to be viewed.

liquid helium from the cryostat, giving rise to a temperature increase of the microscope over days, thus enabling drift-free STM measurements at well-defined temperatures up to 300 K. (Jeandupeux et al., 1999)

Since the tunneling current has an exponential dependence on the distance between tip and sample, vibrations can cause strong noise. Therefore the construction of the STM scanner, consisting of tip holder, sample holder, and actuators, should be as rigid as possible to increase the mechanical eigenfrequency of the scanner. For typical tube scanner setups this eigenfrequency lies in the range of 1–10 kHz. A two-stage damping system consisting of an external damper (e.g., air damped feet suspension of the whole UHV chamber) and an internal damper (e.g., scanner suspended on springs and damped by eddy current dampers) effectively isolates the tunneling gap from building vibrations and acoustic noise. The two damping stages should have resonance frequencies well below building and acoustic frequencies and should be strongly damped (low Q-factor).

With a low-temperature STM the boiling cryogenic liquid produces additional vibrations after the first damping stage. Even worse, the boiling liquid has to be in close proximity to the vibration-sensitive scanner to enable good thermal contact to the scanner. Therefore there are two contradictory demands for the design of a low-temperature STM: mechanical decoupling of the STM head and thermal coupling to the cryostat.

This problem has been solved by different groups in different ways and in the following we will discuss three design examples. They all have in common a helium bath cryostat surrounded by a liquid nitrogen cryostat, which acts as a radiation shield. They all have a two-stage damping system; in the first stage the whole UHV chamber is mechanically decoupled from the floor by pneumatic dampers. The designs differ in the way the second damping stage is realized (see Figure 17–6):

The IBM-Rueschlikon/University Lausanne LT-STM designed by R. Gaisch (Gaisch et al., 1992) uses a bellows to mechanically decouple the liquid helium cryostat from the liquid nitrogen cryostat and the UHV chamber. The STM, however, is rigidly connected to the liquid helium cryostat, which results in very effective cooling. However, vibrations from the cryostat can reach the STM unattenuated.

The LT-STM designed by G. Meyer and K.H. Rieder (Meyer, 1996) has no damping between the liquid helium and the liquid nitrogen cryostat; instead the STM is suspended from the He cryostat by small extension springs and thereby is mechanically decoupled during measurements. During cooldown the STM is lowered onto a copper plate connected to the He cryostat to achieve good thermal contact. Since the thin springs are not very good thermal conductors, the STM has to be effectively shielded against incoming thermal radiation. This is achieved by a two-stage radiation shield with the inner shield connected to the He bath and the outer shield connected to the liquid nitrogen bath. The STM is then almost thermally isolated during measurements (see Figure 17–7).

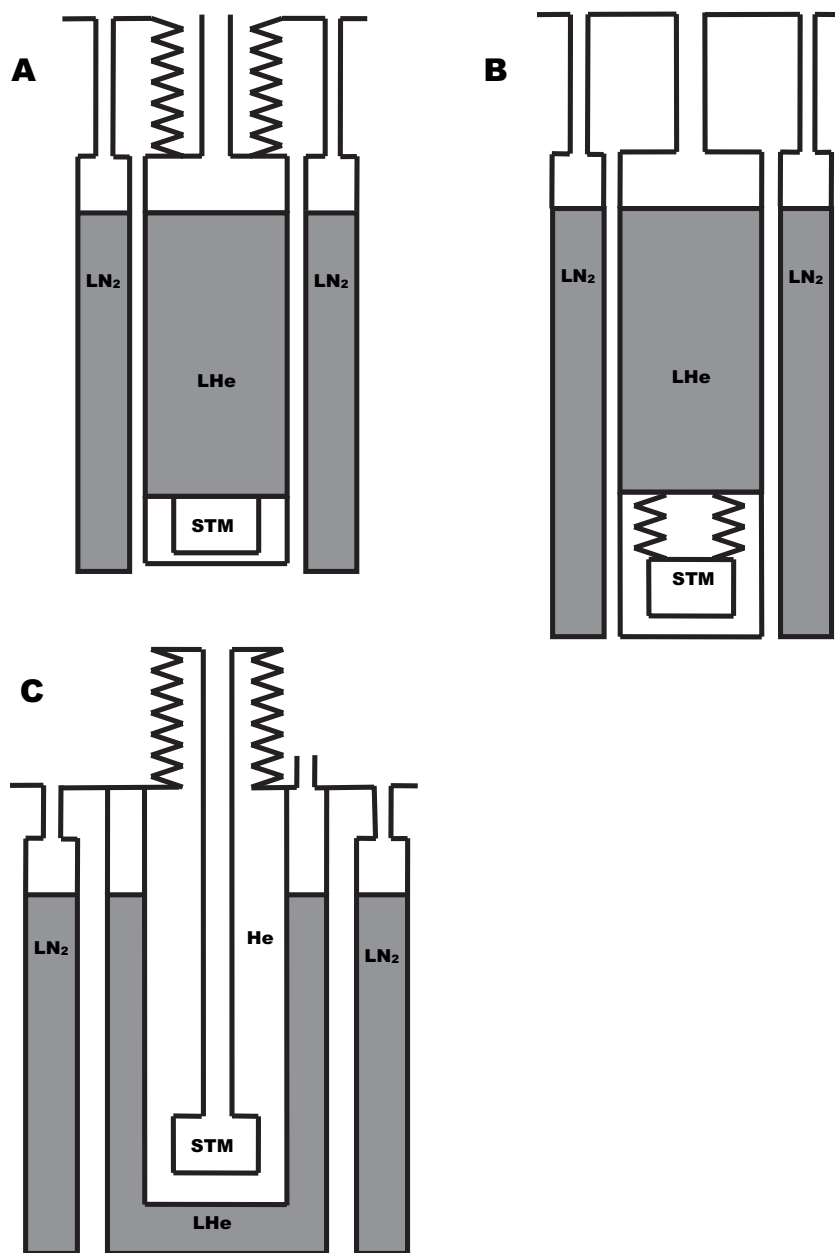


Figure 17-6. Schematic comparison of different LT-STMs with bath cryostat. (A) R. Gaisch/IBM Rueschlikon, (B) G. Meyer/FU Berlin, (C) D. Eigler/IBM Almaden. The connection to the LHe cryostat and the second damping stage of the STM is shown. (A) and (B) have a cryoshield surrounding the STM. In (A) the LHe cryostat is decoupled from the LN₂ cryostat by bellows. In (B) the STM is decoupled from the LHe cryostat by springs. In (C) the STM is decoupled from the LHe cryostat by a bellows-supported pendulum. Thermal contact is made by He exchange gas surrounding the evacuated pendulum.

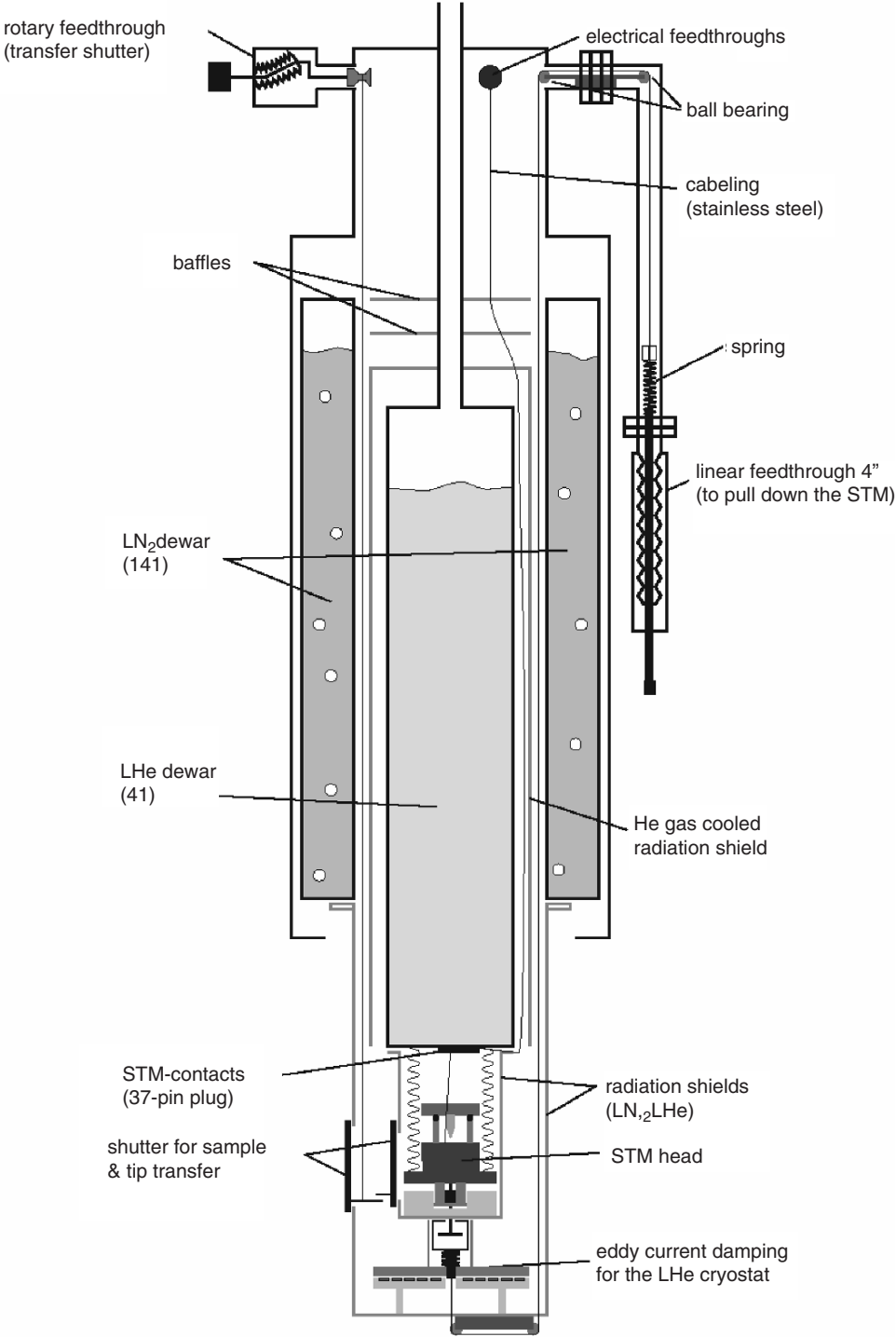
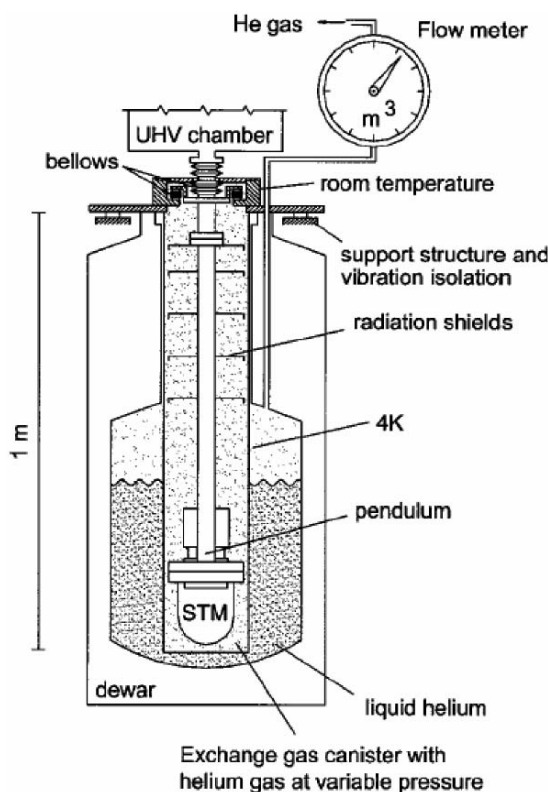


Figure 17-7. Design of the LT-STM by Prof. Rieder and Dr. Meyer (Free University Berlin), courtesy of CreaTec GmbH, distributed by SPECS GmbH.

Figure 17-8. Illustration of the basic design principle for the Eigler-style low-temperature STM. The STM is mounted at the end of a bellows-supported pendulum, which is encapsulated in an exchange gas canister inside of a liquid helium dewar. Limited variable temperature operation is possible by controlling the exchange gas pressure. (From Rust et al., 2001.)



The LT-STM build by D. Eigler (Eigler et al., 1990) at IBM-Almaden and a similar instrument at Berlin are different from the previous two. In both previous instruments, the STM vacuum system and the isolation vacuum of the He and nitrogen reservoirs are connected. In the Eigler design the cryogenic vacuum system is not connected to the STM vacuum system. Instead the STM is mounted at the end of a bellows-supported evacuated pendulum with a resonance frequency of about 0.6 Hz. Helium exchange gas provides thermal coupling as well as acoustic isolation between the surrounding liquid helium dewar and the STM flange. The temperature of the STM can be set by controlling the helium gas pressure in the exchange gas chamber (Rust et al., 2001) (see Figure 17-8). In all designs, mechanical disturbances from bubbling liquid nitrogen in the surrounding liquid nitrogen cryostat can be prevented by solidifying the nitrogen by pumping it with a rotary pump (which is located far from the microscope) (Jeandupeux et al., 1999).

3 Applications

There is a large volume of literature on experiments using a LT-STM covering electronic structure and lifetime measurements, measurements on superconductors, atomic and molecular manipulations, molecular vibrational spectroscopy, photon-emission spectroscopy,

and measurements of magnetic properties with spin polarized electrons. We can discuss only a small subset of the work done and refer to the literature for further reading. For the applications shown here, low-temperature operation of the STM has proven to be essential.

3.1 Electronic Structure

Understanding the distribution and response of electrons in solids, the *electronic structure*, is key to explaining and exploiting fundamental and technologically important properties of materials. The method of choice to obtain information about the electronic structure of solids and surfaces in the past decades has been angle-resolved photoemission spectroscopy (ARPES) and inverse photoemission spectroscopy (IPES). The development of low-temperature scanning tunneling microscopes operating under ultrahigh vacuum conditions has provided new opportunities for investigating electronic states at metal surfaces. At low temperatures, due to reduced broadening of the Fermi level of the STM tip and sample, rather high-energy resolution is achievable. Moreover, the absence of diffusion together with the spatial resolution of the STM enables detailed studies of the interaction of electronic states with single atoms and other nanoscale structures.

3.1.1 LDOS Oscillations

Electrons occupying surface states on the close-packed surfaces of noble metals are bound in the direction perpendicular to the surface, but have free-electron-like characteristics parallel to the surface (Gartland and Slagsvold, 1975; Heimann et al., 1977; Zangwill, 1988). On the (111) face of noble metals these surface states arise as a result of the gap that exists along the Γ -L line in their bulk Brillouin zone (Shockley, 1939). Surface state electrons play an important role in a variety of physical processes, including epitaxial growth (Mommel and Bertel, 1995), in determining equilibrium crystal shapes (Garcia and Serena, 1995), molecular ordering (Stranick et al., 1994a), surface catalysis (Bertel et al., 1995), and physisorption (Bertel, 1997).

Surface state electrons scattered from defects and step edges give rise to quantum interference patterns in the electron density, which can be probed using the scanning tunneling microscope (Davis et al., 1991). Standing wave patterns in the LDOS on the Cu(111) surface have been observed with an LT-STM for the first time by Crommie et al. (1993b). LDOS oscillations at surfaces are the analog to the Friedel oscillations of the total charge density (Friedel, 1958). The experiments were done at 4 K in an LHe bath cryostat STM (Eigler and Schweizer, 1990). Figure 17-9 shows a constant current image of the Cu(111) surface with static spatial oscillations in the LDOS. The oscillations decay away from step edges and point defects and have a characteristic period of $\sim 15 \text{ \AA}$. The amplitude of the corrugations is $\sim 0.02 \text{ \AA}$ and is greater near the top of step edges than near the bottom. Spatial variations in the LDOS of the surface at the energy $E = E_F + eV$ can be approximately mapped out by measuring dI/dV at bias voltage V (Crommie et al., 1993b). Figure 17-10 shows dI/dV linescans as a function of distance to a monoatomic step on Cu(111). The wavelength of

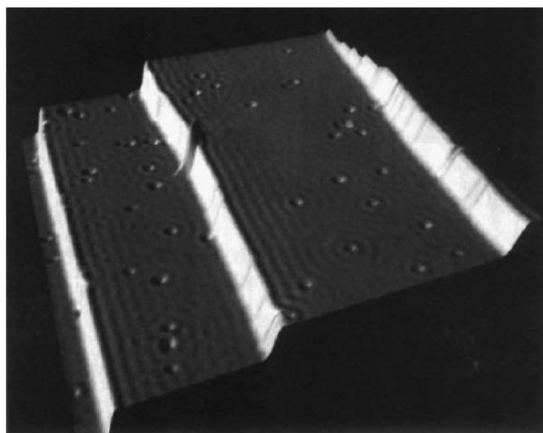


Figure 17-9. Standing wave patterns on copper: constant current image of the Cu(111) surface ($V = 0.1$ V, $I = 1.0$ nA). Three monoatomic steps and several point defects are visible. Spatial oscillations of the tunneling current originating from step edges and defects are evident. Image size: $500 \text{ \AA} \times 500 \text{ \AA}$. (From Crommie et al., 1993b.)

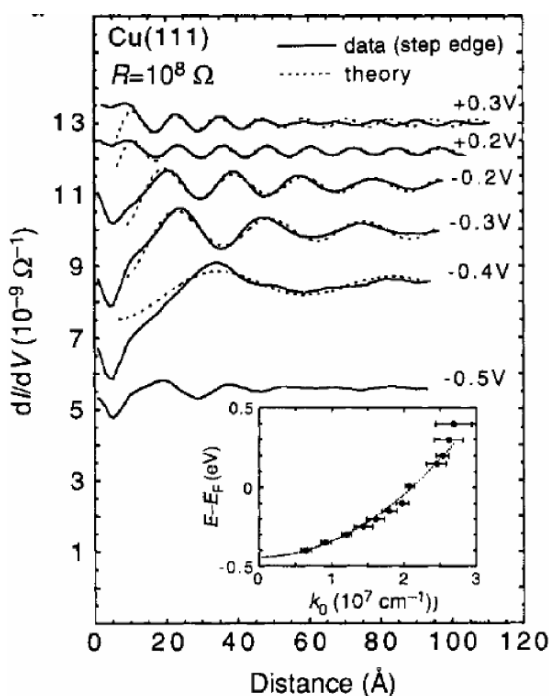


Figure 17-10. Spatial dependence of dI/dV , measured as a function of distance (along upper terrace) from step edge on Cu(111) at different bias voltages. Zero distance corresponds to the lower edge of the step. The measured wavelength of the surface LDOS oscillations changes as a function of energy. Solid lines, experiment; dashed lines, theoretical fit modeling the surface as a two-dimensional electron gas in the presence of a single hard wall barrier. Inset shows the extracted values of k plotted against energy defining an experimental dispersion curve of the surface state. (From Crommie et al., 1993b.)

the oscillation in the surface LDOS increases with decreasing energy. At ~ 0.45 eV below E_F the LDOS sharply decreases in magnitude. Measuring a full dI/dV spectrum at a fixed point on the surface revealed the origin of this transition: Figure 17–11 shows dI/dV spectra recorded with a stationary tip away from steps and on a step while ramping the bias from 1.0 to -1.0 V. The spectra taken on a terrace show a sharp drop in dI/dV at $V \approx -0.45$ V. This corresponds to a sudden decrease in the surface LDOS at energies more than 0.45 eV below E_F and marks the bottom of the Shockley surface state band on Cu(111).

By fitting the oscillations of the dI/dV linescan data to a simple free particle model with hard-wall confinement at surface steps, the major features of the spectra could be explained and a value of the surface state electron wavevector $k_{\parallel}$ for each electron energy could be extracted. A plot of these electron wavevectors against electron energy resulted in an experimental dispersion curve, as shown in the inset of Figure 17–10. The $k_{\parallel}$ data for different energies $E - E_F$ were fitted with the dispersion relation for electrons confined to two dimensions (dotted curve in inset of Figure 17–10):

$$E(k_{\parallel}) = E_0 + \frac{\hbar^2 k_{\parallel}^2}{2m^*} \quad (4)$$

where E_0 is the surface state band edge, m^* is the effective mass of the surface state electron, and $k_{\parallel} = \sqrt{k_x^2 + k_y^2}$ is the electron wavevector parallel to the surface. This procedure yielded the surface state effective mass $m^* = 0.38 \pm 0.02 m_e$ and the surface state band edge $E_0 = -0.44 \pm 0.01$ eV below E_F . The extracted value of the surface state band edge matched the location of the peak in the dI/dV spectrum of Figure 17–11. The measured values were roughly in agreement with previous photoemission results.

In the previous example, scattering of surface state electrons has been modeled as scattering at infinite potential walls. However, absorp-

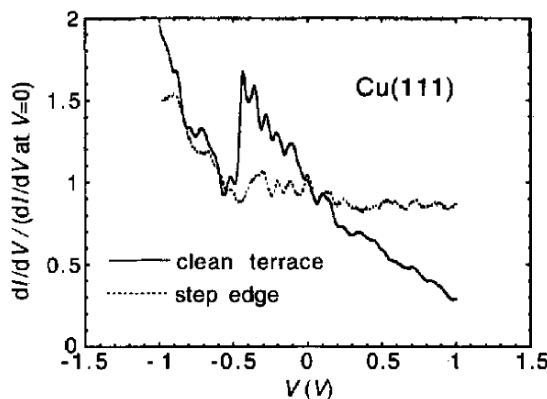


Figure 17–11. Differential conductance dI/dV spectra taken with tip over clean terrace (solid line) and over the center of a step edge (dashed line). The increase at ~ 0.45 V is due to the onset of the Cu(111) surface state. (From Crommie et al., 1993b.)

tion processes at step edges due to bulk coupling are disregarded with this model.

Therefore, a Fabry–Perot resonator model with a step reflection amplitude and scattering phase has been introduced to analyze the measured LDOS patterns of electrons confined between two parallel steps on Ag(111) at 4.9 K. (Burgi et al., 1998). The electron reflectivity has been found to be independent of crystallographic step structure, but depends on the step morphology (ascending, descending). Reflectivity and scattering phase could be quantified for any two parallel steps, providing insight into the scattering mechanism.

It has been assumed so far that for measurements close to E_F (small bias voltage), the LDOS and ILDOS are essentially the same, i.e., a constant-current image (topograph) and a dI/dV differential conductance image should be identical. However, with increasing bias voltage the constant-current image should systematically diverge from the LDOS (dI/dV) as states from a range of energies contribute to the integrated state density.

One-dimensional electron confinement has been used to measure the bias-dependent difference between the ILDOS and LDOS (Pivetta et al., 2003) with a low-temperature STM. The measurements were performed in a home-built STM, operating in UHV at 4.8 K (Gaisch et al., 1992). The system studied was a quantum box consisting of two-dimensional surface states on Ag(111) confined by two parallel surface steps (see Figure 17–12). Constant current cross sections (measuring the ILDOS) and dI/dV cross sections (measuring the LDOS) from the quantum box perpendicular to the steps were compared at different bias voltages. Oscillations in both, constant current and dI/dV curves were observed (Figure 17–13). These oscillations are due to the interference of the incident and scattered surface state electrons at the box boundaries. The measured data were compared with a model calculation of the LDOS and ILDOS assuming a free particle model with hard wall confinement at surface steps.

The surface state electrons are free in the direction parallel to the steps, but due to the boundary conditions given by the one-dimensional confinement in a box of width L , the component of the electron wavevector perpendicular to the step edges is quantized: $k_{\perp n} = \pi n/L$. Each corresponding eigenenergy $E_n(k_{\perp n})$ defines the onset of a one-dimensional subband $E_n(k_{\parallel})$, where $E(k_{\parallel})$ is the dispersion relation (4) for the Ag(111) surface state electrons. Since the tunneling current I is calculated by integrating over all energies between the bias eV and E_F , it contains contributions from several subbands $E_n(k_{\parallel})$. Calculating the differential conductance dI/dV also requires integration of the DOS over a small range of energies given by the modulation amplitude used for lock-in detection (here 1–10 mV). Comparison of the model calculations and experimental measurements could explain the observed oscillation patterns in I and dI/dV in terms of the way in which different subbands contribute to the signals. The measured data of Pivetta et al. (2003) indicate that the difference between differential conductance and constant-current measurements can already be pronounced for bias voltages as small as 19 mV. Thus, even close to E_F , the interpre-

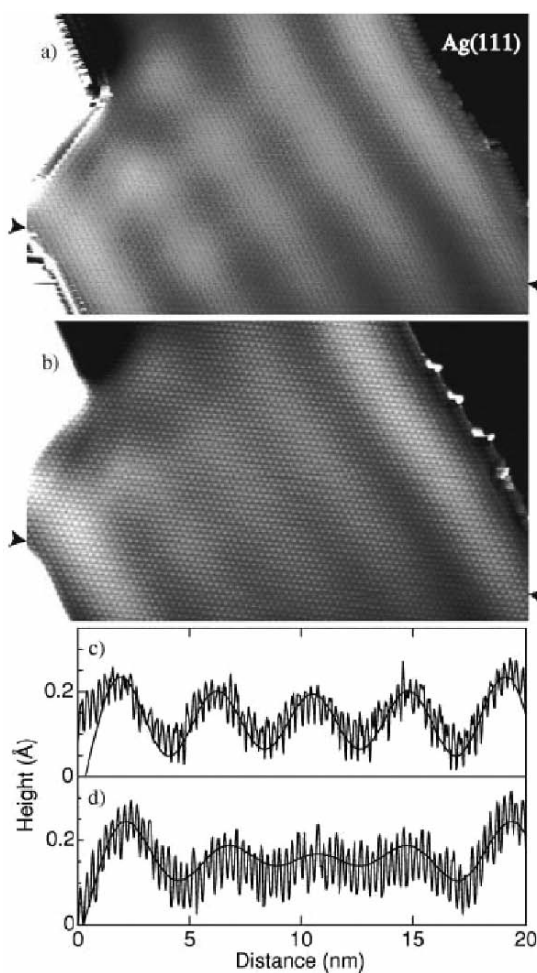


Figure 17-12. Surface states confined by two parallel steps on Ag(111). Atomically resolved STM constant current images at 4.8 K, $20\text{ nm} \times 12\text{ nm}$, $I = 1.8\text{ nA}$ at (a) +21 mV and (b) +6 mV showing standing-wave patterns. (c and d) Cross sections along the black arrows in (a) and (b), respectively, showing the atomic corrugation superimposed on the standing waves. The solid curves are the calculated ILDOS. (From Pivetta et al., 2003.)

tation of constant current images in terms of the LDOS may not be fully correct.

3.1.2 Energy Dispersion Measurements

Energy dispersion measurements as a function of the electron wave-vector $E(k)$ of surface states on metals is usually performed by techniques such as ARPES for occupied states or IPES for empty states, which are selective in both energy and wavevector. As shown in Section 3.1.1, the energy dispersion relation of the surface state can also be measured with scanning tunneling spectroscopy (STS) by measuring the LDOS on single line scans perpendicular to a straight substrate step. Measurement of the LDOS ρ_s at the surface by means of measuring the differential conductance dI/dV is affected by local variations of the tip-sample distance z and therefore does not exactly represent ρ_s (Crommie et al., 1993b; Heller et al., 1994; Hormandinger, 1994). As a consequence, the surface state dispersion $E(k)$ determined from dI/dV maps deviates somewhat from photoemission spectroscopy results (Hormandinger, 1994). Simultaneous measurements of z and dI/dV

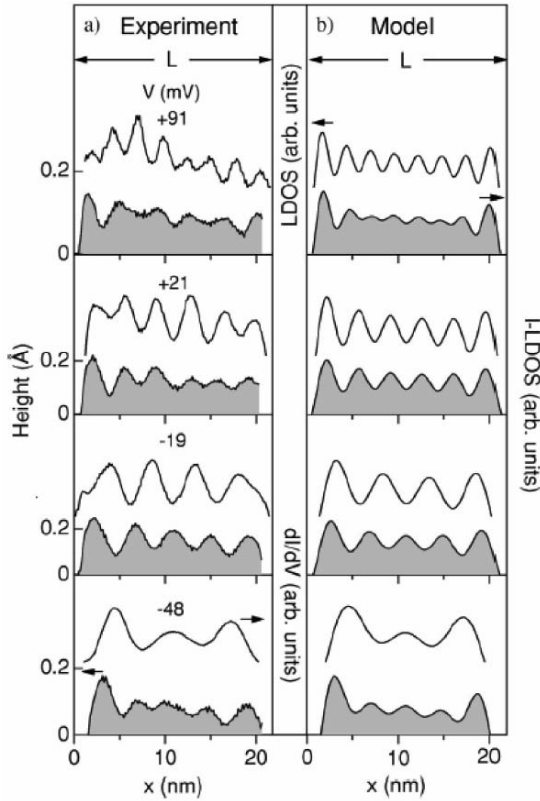


Figure 17-13. Bias-dependent evolution of the LDOS and ILDOS oscillations of surface states confined by parallel surface steps. (a) Simultaneously recorded constant current (gray shaded) and dI/dV (curves above) cross sections from a quantum box of width $L = 21.8\text{ nm}$ formed by two parallel steps on Ag(111). Bias voltages (mV) are indicated (bias applied to the sample). (b) Calculated electron density (gray shaded) and dI/dV . Note the difference between dI/dV (LDOS) (three maxima) and topography (ILDOS) (five maxima) at -48 mV , a bias voltage that might be considered to be close to E_F . (From Pivetta et al., 2003.)

have been used to recover the oscillating LDOS on a close-packed Ag(111) surface at $T = 50\text{ K}$ (Li et al., 1997).

A dependence of dI/dV on local z variations follows from Eq. (3). If the second term in Eq. (3) is neglected, the differential conductance is proportional to the product of ρ_s and T . The exponential z -dependence of T causes an error in the measurement of ρ_s in terms of dI/dV , if the tip-sample distance is not constant, but instead is controlled in constant current mode at the same bias voltage where dI/dV is measured.

At low bias, the tunneling current results from an integration over a narrow energy window and consequently the standing wave oscillations in the ILDOS get more pronounced (less k values contribute to the pattern). This causes an oscillation of z , since at every position x , the tip height $z(x)$ is adjusted by the feedback loop to follow the ILDOS oscillations at the preset constant-current value. Then the simultane-

ously measured dI/dV signal does not reflect ρ_s anymore, because the z adjustments influence the barrier transmission T , which depends exponentially on z . Since dI/dV is proportional to the product of ρ_s and T , ρ_s at $E = eV$ can be recovered from dI/dV by division with T , which is measured by measuring $z(x)$, i.e., a constant-current image. With these corrections, good agreement between LT-STM-derived dispersion curves $E(k)$ and PES-derived dispersion curves has been achieved (Li et al., 1997).

Another solution to avoid convolution between standing waves in the tip height z of the constant-current line scan and those in the simultaneously recorded dI/dV spectra is to control z at a large negative bias voltage (relative to the sample). Under these conditions, the current I contains contributions from electronic states with many different oscillation periods ($k_{\parallel}$ values), which minimizes the standing waves in the z signal (ILDOS). Thus the tip moves to a good approximation parallel to the surface plane, unaffected by interference patterns in the LDOS. The differential conductance is then roughly proportional to the LDOS of the sample (Hormandinger, 1994). Such a measurement has been published by Jeandupeux et al. (1999), and the result is shown in Figure 17–14. The upper graph displays the constant-current line scan on which the tip was moved while taking differential conductance maps. It can be seen that at a bias voltage of $V = 0.3$ V the tip–surface distance is almost unaffected by standing waves and follows the real topography. The differential conductance data are represented by gray levels as a function of the distance x from the step edge and the energy E . This plot already illustrates the dispersion of the Ag(111) surface state: from top to bottom the wavelength of the LDOS oscillations increases until it diverges at the band edge at $E_0 = -65$ meV. Analyzing constant energy cuts of the differential conductance plot in Figure 17–14 in quantitative terms by modeling the reflection of electrons at a potential barrier with a reflected amplitude and a phase shift leads to values for the wave number k for each energy, and thus the energy dispersion relation $E(k)$ of the Ag(111) surface state, which is shown in Figure 17–15. The data were in excellent agreement with other STS-derived data (Li et al., 1997) and PES data (Paniago et al., 1995).

The previously shown measurements of the dispersion relation on noble metal surfaces were limited to $k_{\parallel}$ vectors around the center of the surface Brillouin zone (SBZ) and it has been found that in this limit the surface state is free electron like, i.e., has an isotropic parabolic dispersion. LT-STs measurements on Ag(111) and Cu(111) over an extended energy range have shown a significant deviation from free electron behavior for large $k_{\parallel}$ vectors approaching the symmetry points at the SBZ boundary (Burgi et al., 2000b).

Direct visualization of the surface state dispersion on Ag(110) by means of Fourier transformation of differential conductance data taken with an LT-STM at 4 K at energies up to the vacuum level has also been shown (Pascual et al., 2001b). Low temperatures (4 K) enabled the necessary high (2 meV) energy resolution and high stability to allow long recording times for the measurement of the differential conductance (dI/dV) using lock-in amplification techniques.

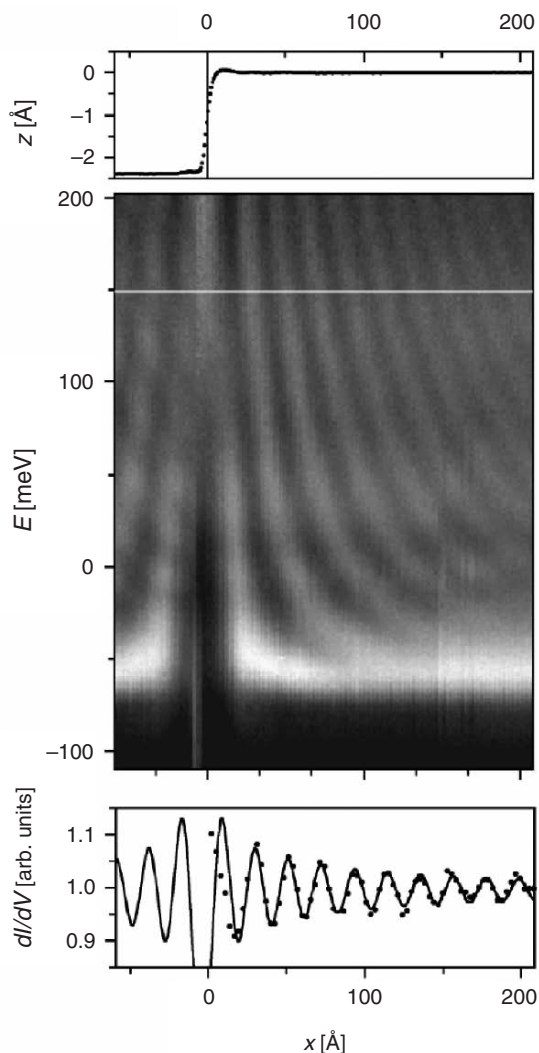


Figure 17-14. LDOS oscillations on Ag(111) at 5K represented as gray levels as a function of the lateral distance from the step x and of bias energy $E = eV$ with respect to E_F . The upper graph shows the constant-current line scan ($V = 0.3\text{ V}$, $I = 2.0\text{ nA}$) on which the STM tip was moved while taking the dI/dV spectra. The lower graph is a cut of the dI/dV plot taken along the white line at $E = 148\text{ meV}$ (dots), and the line is a fit to theory. (From Jeandupeux et al., 1999.)

Figure 17-16a shows the LDOS oscillations in front of a step on Ag(110) represented by a gray scale, plotted as a function of the distance from the step edge and the energy. Figure 17-16b shows the one-dimensional Fourier transform of the spatial one-dimensional conductance profiles in Figure 17-16a and visualizes the dispersion of the surface state in reciprocal space. A calculation shown in Figure 17-16c and d with a Bloch wave function limited to the first-order terms ($G = -1, 0, 1$) could explain the features of Figure 17-16a and b. It could

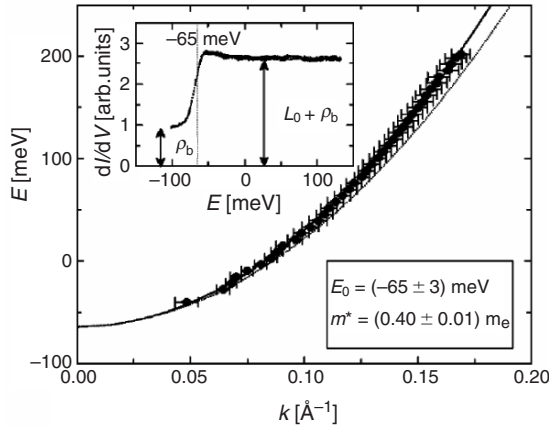


Figure 17-15. Energy dispersion relation of the Ag(111) surface state obtained from the measurement shown in Figure 17-14. The solid line is a quadratic fit to the STM data and the dotted line shows results from photoemission data at 65 K (Paniago et al., 1995). The inset shows a dI/dV spectrum taken on a perfect terrace ($V = 497 \text{ mV}$, $I = 5.0 \text{ nA}$) showing the onset of the surface state at 65 mV below the Fermi energy. (From Jeandupeux et al., 1999.)

be shown that the underlying atomic lattice gives rise to additional features in reciprocal space arising from coherent interference with surface state-derived Bloch waves.

3.1.3 Electron Confinement

When electrons are confined to length scales close to the de Broglie wavelength, their behavior is dominated by quantum mechanical

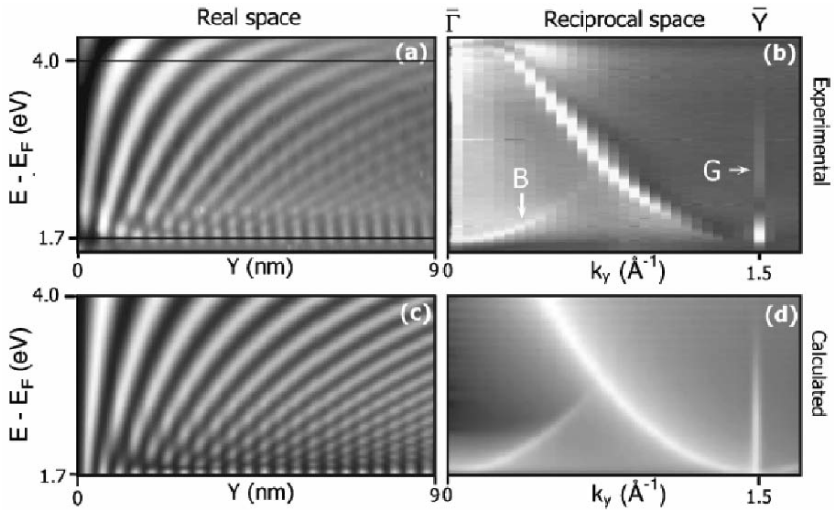


Figure 17-16. Direct visualization of the surface state dispersion $E(k)$. (a) Experimental dI/dV data from Ag(110) at 4 K as a function of energy and distance from a monoatomic step. (b) One-dimensional Fourier transform of the dI/dV vs. Y data, which is a direct visualization of the surface state dispersion. (c and d) Results from calculations. (From Pascual et al., 2001b.)

effects, i.e., discrete energy levels are formed. The locality of the probe in STM has an advantage over ARPES for the measurement of the effects of lateral confinement of surface-state electrons on the surface-state energies. Theoretical calculations have suggested that lateral confinement raises the energies of the surface-state electrons, resulting in a depopulation of the surface-state band and a modification of surface properties associated with the surface-state electrons (Bertel et al., 1995; Memmel and Bertel, 1995; Garcia and Serena, 1995). The energy levels of surface-state electrons confined within artificial nanostructures have been measured with an LT-STM for the first time by Crommie et al. (1993a). By manipulation with the STM tip, they positioned 48 Fe atoms on Cu(111) in the form of a circle, forming a so-called quantum corral, which confines surface-state electrons in the inner area (Figure 17-17). A standing wave pattern was observed within the corral. dI/dV spectra taken in the center showed that the energy levels of the confined electrons are discrete. Good agreement with quantum mechanical calculations of the energy levels for a particle in a box was found. For ideal confinement of electrons (i.e., infinitely high potential walls)

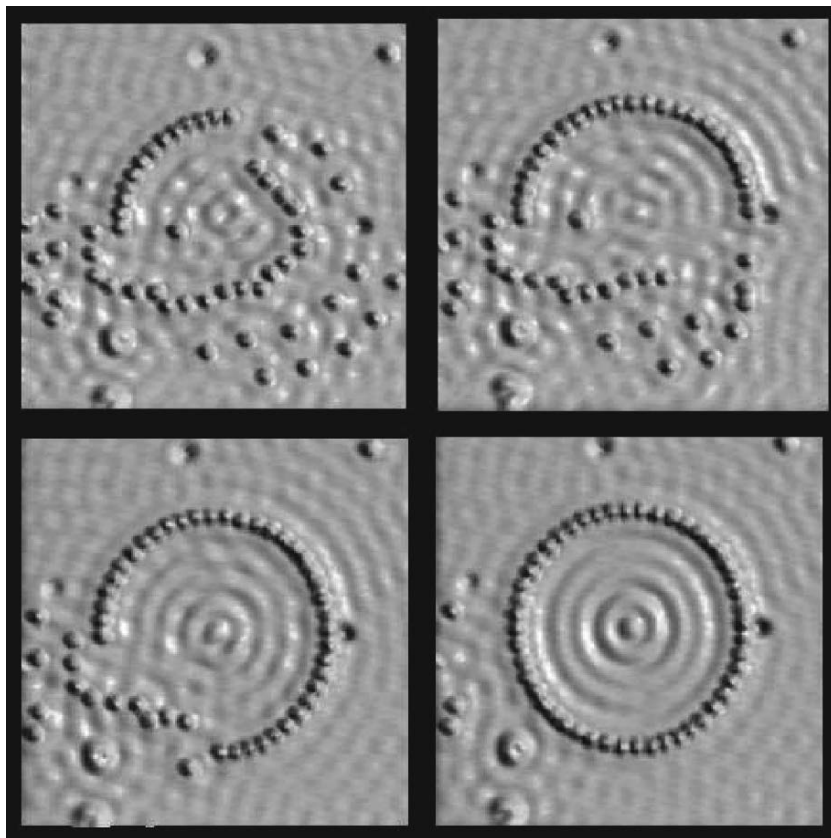


Figure 17-17. Various stages during the construction of a circular quantum corral consisting of 48 Fe atoms on Cu(111) (measured at 4 K). (From Crommie et al., 1993a. Reproduced by permission of IBM Research, Almaden Research Center.)

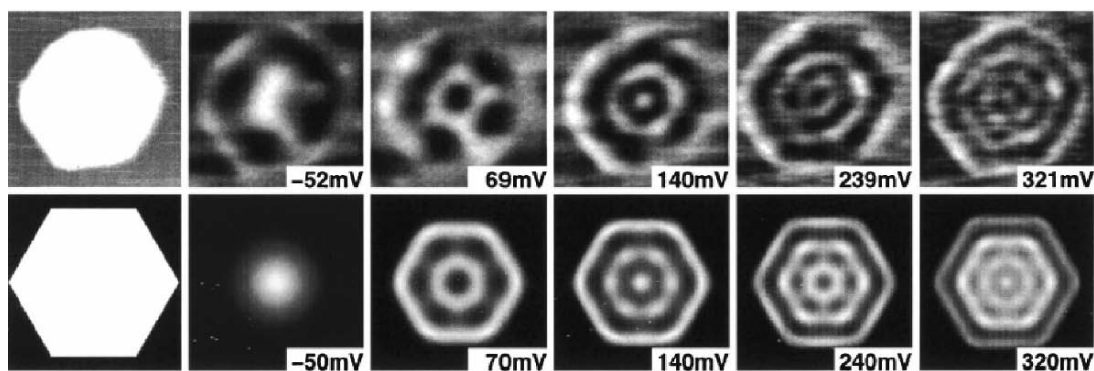


Figure 17–18. Electron confinement on nanoscale hexagonal Ag islands. Upper row: constant-current image of a hexagonal Ag island on Ag(111) (area $\sim 94 \text{ nm}^2$) and a series of dI/dV maps recorded at 50 K at different bias voltages as indicated. Lower row: calculated local density of states for the hexagonal box shown at the left. (From Li et al., 1998b.)

the energy levels E_n should be infinitely sharp apart from instrumental contributions. The measured peak width was interpreted as a lifetime effect via $\Delta t_n \approx \hbar/\Delta E_n$, dominated by scattering into bulk states (see Section 3.1.4).

Another experiment with a 76-atom “stadium” by the same group (Heller et al., 1994) gave further information about the scattering mechanism of surface-state electrons at the Fe adatoms. Comparison with s -wave multiple scattering theory showed that the Fe atoms reflect only about 25% of the incident electron wave, while 25% are transmitted and 50% are absorbed. Absorption is related to scattering into bulk states.

More recently a systematic study of electron confinement on nanoscale hexagonal Ag islands on Ag(111) was performed (Li et al., 1998b). In contrast to adatom corrals, these are stable structures even at elevated bias voltages in the STM. Figure 17–18 shows a series of dI/dV maps taken on a hexagonal Ag island on Ag(111). For voltages around -65 mV and lower (tunneling form of the sample) the image of the interior of the island is featureless, since this is below the onset of the surface state at -67 mV (Li et al., 1997). At higher voltages standing wave patterns are observed, which originate from oscillations in the LDOS of surface-state electrons, confined by the rapidly rising potential at the edges of the island. By recording local dI/dV spectra above the center of an island with an open feedback loop and varying tunnel voltage, peaks at different energy where recorded, which correspond to the energy levels of the confined surface-state electrons. The validity of the particle in a box model for these confined surface states was confirmed. Analyzing the measured energy levels in terms of a particle in a box model showed that the energies conform to the expected scaling behavior down to the smallest island sizes. The experiments confirmed that lateral confinement is a mechanism for surface-state depopulation since it causes a discretization of the formerly continuous SS band and an upward shift of the surface-state levels. The fact that ARPES on vicinal Cu(111) surfaces (Sanchez et al., 1995) failed to observe the shifts in the surface state

energies expected on the basis of ideal confinement is most likely due to the presence of a sizable distribution of terrace widths within the spot area of the incident radiation.

Electron confinement effects have also been observed in metallic single-walled carbon nanotubes (SWCNTs) (Maltezopoulos et al., 2003). With an LT-STM operating at 14 K (Pietzsch et al., 2000b), dI/dV spectra and maps on SWCNTs have been recorded (Figure 17–19). Peaks close to the Fermi energy are restricted to a certain area of the nanotube and have been assigned to confined states within the SWCNT created by defect scattering (see Figure 17–20). This shows that defects can lead to significant backscattering within individual metallic SWCNTs. These findings are important since SWCNTs are considered promising candidates for molecular electronics.

LT-STM measurements of the two-dimensional structure of individual electron wavefunctions in SWCNTs have been reported (Lemay et al., 2001). An SWCNT was deposited on an Au(111) surface. To increase the electronic energy-level spacing, the nanotube was cut to less than 40 nm length by applying short bias voltage pulses between tip and sample. Local spectroscopy with the tip above the nanotube showed an energy spectrum with a series of sharp peaks, representing confinement-induced energy levels. Differential conductance maps measured at the energies of the peaks revealed spatial patterns that could be understood from the electronic structure of a single graphite sheet. The patterns observed are a representation of individual molecular wavefunctions $|\psi(r)|^2$. The wavefunctions exhibited “beating” between electron waves with slightly different wavevectors, and this interference effect was exploited to directly probe the electronic dispersion relation of an individual nanotube.

3.1.4 Lifetime Measurements

Improved energy resolution (~ 3 meV) in photoemission spectroscopy has made it possible to study the lifetimes of photoholes in a surface

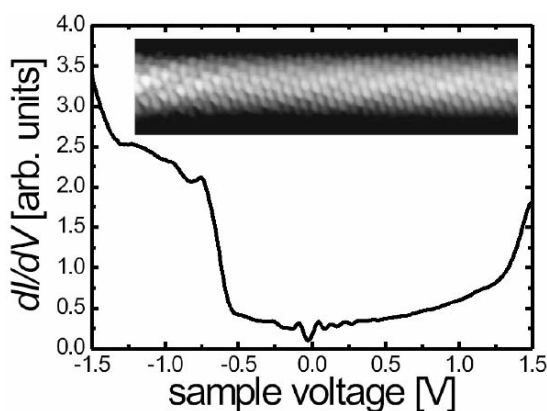


Figure 17–19. Electron confinement in single wall carbon nanotubes: dI/dV spectrum of a metallic SWCNT. The oscillations around the Fermi energy (0 V) are shown in more detail in Figure 17–20. Inset shows atomically resolved STM image of SWCNT at 14 K. (From Maltezopoulos et al., 2003.)

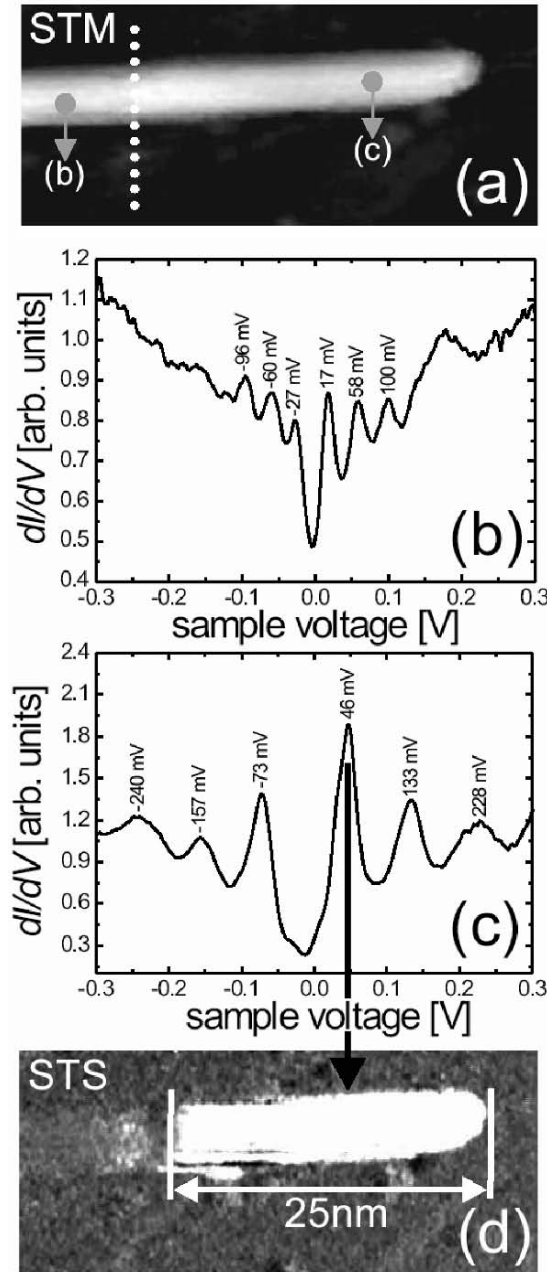


Figure 17-20. (a) STM image of an SWCNT end. (b) STS data ($-LDOS$) taken on the left-hand side of the dotted line in (a). (c) STS data taken on the right-hand side of the dotted line in (a). (d) Spatially resolved dI/dV map taken at the energy of the 46 mV peak shows confinement of these states to the area on the left. (From Maltezopoulos et al., 2003.)

state. Furthermore, femtosecond time-resolved two photon photoemission (2PPE) experiments opened up a new path to measure lifetimes of hot electrons in metals. A hole in an energy band or a hot electron are quasiparticles, elementary excitations of an interacting Fermi liquid. A quasiparticle has a lifetime that is the duration of the excitation. In combination with its velocity, the lifetime determines the mean free path of the quasiparticle and thereby the range of influence of the

excitation. Lifetimes are determined by inelastic scattering processes, i.e., electron–electron (e–e) scattering and electron–phonon (e–ph) scattering. Fermi liquid theory predicts an inverse second power law for the energy dependence of the e–e lifetime. It also predicts that e–e scattering dominates e–ph scattering at low temperatures and large excess energies (≥ 0.5 eV), whereas at energies very close to the Fermi level inelastic scattering is dominated by e–ph processes at all temperatures (Pines, 1966).

In metallic systems, quasiparticle states in two-dimensional systems are important in surface science (Memmel, 1998) and nanoscale technology (Himpsel et al., 1998). One example for a two-dimensional system is a surface state on a noble metal, and here the *L*-gap Shockley-type surface state on the (111) surface of noble metals has played a special role as model systems for the investigation of two-dimensional electron systems. The lifetime of a hole in such a surface state band has been extensively studied both theoretically and experimentally (Matzdorf, 1998), but the discrepancy between experiment and theory remained large. This lifetime is of interest as it determines the effective range of surface-mediated interactions and is important for epitaxial growth and chemical reactions at surfaces (Memmel and Bertel, 1995; Bertel et al., 1995). Theoretical models predicted significantly longer lifetimes than were observed with photoemission spectroscopy. Using

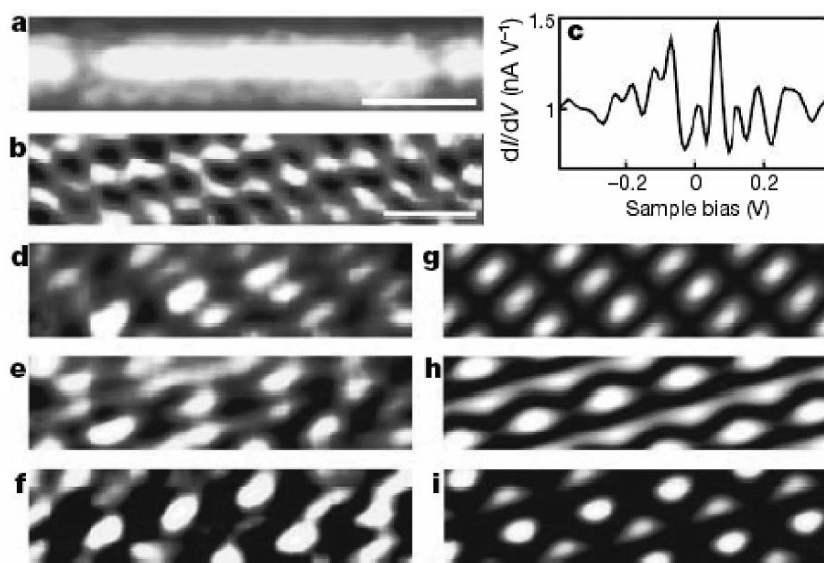


Figure 17-21. Imaging individual molecular wavefunctions in a carbon nanotube at 4.6 K. (a) Constant-current image of metallic SWNT cut to a length of 34 nm by bias voltage pulses (scale bar 10 nm). (b) High-resolution constant-current image of the shortened nanotube (scale bar 0.5 nm). (c) Differential conductance spectrum (LDOS) averaged over the area shown in (b) showing particle in a box states. (d–f) Differential conductance maps showing individual states at -95 , 30 , and 96 meV. (g–i) Calculated spatial maps of individual molecular wavefunctions. The characteristic features of each image are well reproduced. (From Lemay et al., 2001.)

ARPES from a surface state, the lifetime of a photohole state can be determined by analyzing the spectral linewidth Γ of a photoemission excitation. The spectral linewidth of a photoemission line gives access to the lifetime τ via $\Gamma = \hbar/\tau$. ARPES integrates over a large surface area (typically $\geq 1 \text{ mm}^2$) and contains contributions from surface imperfections over the whole investigated surface range (Theilmann et al., 1997). Therefore peak widths in ARPES are influenced by surface conditions (defects, roughness) as well as finite energy and angular resolution (so-called nonlifetime effects). Defects on the surface are known to strongly couple the surface-state electrons to bulk states and thus reduce lifetimes (Crampin et al., 1994). It would be desirable to apply the concept of measuring lifetimes from the linewidth of spectral features directly to STS. The nonlifetime effects could then be avoided by measuring the lifetime locally with an STM tip on a spot bare of impurities. The ability of the STM to detect surface topography and to identify minute amounts of contamination well below the limits of conventional surface analytical techniques ensures that an effectively defect-free surface is studied. Since STS does not offer momentum resolution, measuring the lifetime of a specific state defined by its momentum $k_{\parallel}$ and energy E may seem impossible. This problem can be circumvented to some extent as shown below.

The surface-state (excited hole) lifetime on Ag(111) has been determined at the dispersion minimum ($k_{\parallel} = 0$) by evaluating the linewidth of the surface state onset in dI/dV tunneling spectra (Li et al., 1998a; Pivetta et al., 2003; Kliewer et al., 2000). Calculated dI/dV spectra show (Li et al., 1998a) that this linewidth depends on the imaginary part Σ of the electron self-energy. As Σ increases, the onset of the surface-state contribution is seen to broaden. Therefore the measured linewidth can be used to estimate Σ or the lifetime $\tau = \hbar/2\Sigma$ (Li et al., 1998a) of a hole in the surface state at the band minimum. The first lifetime measurements (Li et al., 1998a) were performed at a temperature of 5 K using W tips on Ag(111). Differential conductance spectra were recorded under open feedback loop conditions, adding a modulation voltage (1–10 mV at $\omega = 230 \text{ Hz}$) and using a lock-in amplifier to record the signal at ω . Figure 17–22 shows a dI/dV spectrum measured on a large defect-free terrace showing the characteristic surface state band onset of width Δ , used to estimate the spectral linewidth (also called lifetime width or inverse lifetime) $\Gamma = \hbar/\tau$. The geometric width Δ is a combination of the spectral linewidth Γ of the surface state, thermal broadening, and additional broadening due to the modulation technique used to measure the spectrum. The electron self-energy Σ and the corresponding lifetime $\tau = 67 \pm 8 \text{ fs}$ were determined from a line-shape analysis. This measured spectral linewidth was considerably smaller than previous PES measurements. A more recent lifetime measurement (Pivetta et al., 2003) with the same method resulted in even smaller spectral linewidths, which were about 1 meV smaller than recent photoemission (Nicolay et al., 2000) results and theoretical predictions (Eiguren et al., 2002). Since STS is measured on locally defect-free surface regions, a smaller spectral linewidth (i.e., a larger lifetime of surface hole states) than in photoemission is expected. Lifetimes

resulting from STS lineshape analysis where $\tau = 120, 35$, and 27 fs for excited holes at the surface-state band edge on Ag, Au, and Cu, respectively (Kliwer et al., 2000). A comparison of linewidths obtained from STM measurements with previous PES data (Kliwer et al., 2000) showed that for Ag(111) and Au(111), STM measures linewidths smaller by a factor of 3 compared to PES (i.e., longer lifetimes). This illustrates how large the contribution from electron scattering at surface imperfections is. Cu(111) extrapolation of the PES linewidth to zero defect density (Theilmann et al., 1997) resulted in good agreement with the STM value, which provides evidence that STS measurements avoid defect-induced broadening of the linewidth.

In the absence of defect scattering, two processes limit the lifetime of the hole: inelastic e-e scattering and e-ph coupling. Inelastic e-e scattering results in the hole being filled by interband transitions involving bulk electrons near the surface or by intraband transitions involving surface-state electrons (Figure 17-23). These contributions have been estimated theoretically and good agreement of STS measurements of the lifetime width with theoretical values has been achieved (Kliwer et al., 2000).

With the method described, lifetimes can be studied only at a single energy, namely the energy E_F at the surface band minimum. Therefore, in a different approach, the electron lifetimes have been measured with an LT-STM (Burgi et al., 1999; Jeandupeux et al., 1999; Burgi et al., 2000a; Vitali et al., 2003) by studying the decay of quantum mechanical interference patterns from surface-state electrons scattering off step edges. The decay length is influenced by the loss of coherence and hence the phase relaxation length. The phase-relaxation length, i.e., the distance an electron can travel without losing its phase information, is directly related to the lifetime. If the lifetime is shorter than the time necessary

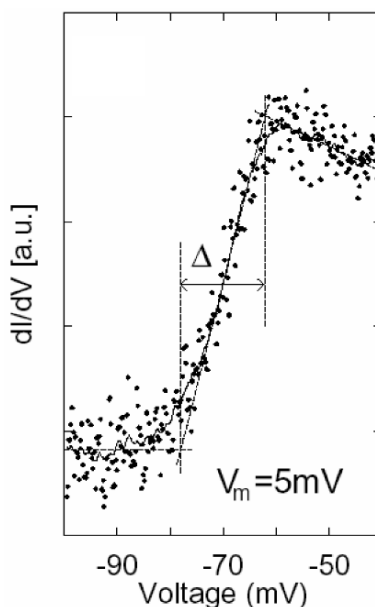


Figure 17-22. Surface state lifetime measurement. dI/dV spectrum measured on a large defect-free terrace on Ag(111), showing steplike onset of the surface state near -70 mV. The surface state lifetime can be determined from the width Δ by fitting it to calculated spectra. (From Li et al., 1998a.)

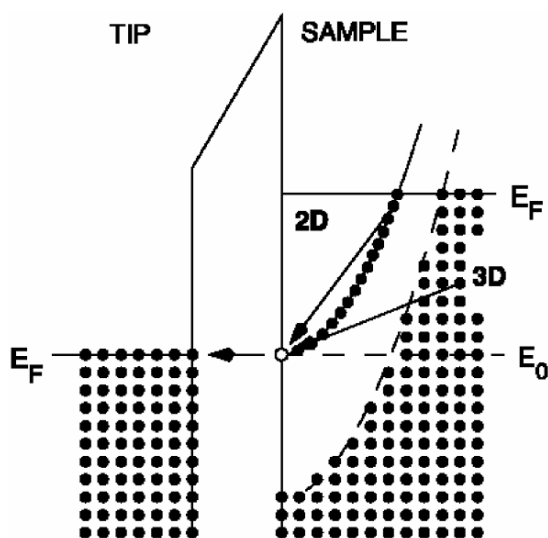


Figure 17-23. Schematic energy diagram of electron (solid circles) tunneling from the bottom of the occupied surface state band of the sample to a metallic tip (sample bias negative with respect to the tip). Sample states are shown as a projected band structure. The dashed parabola indicates the maximum of the bulk bands. The intrinsic lifetime of the hole left behind by an electron tunneling from sample to tip is determined by inelastic electron–electron scattering, which results in either an intraband transition (two-dimensional) or an interband transition (three-dimensional). (From Kliewer et al., 2000.)

to reflect the electron wave back from the step, the phase information is lost and the electron wave cannot interfere with itself anymore. The electron lifetime τ is related to the phase relaxation length by $\tau = L_\phi/v_g$, where v_g is the group velocity (Datta, 1995). The LDOS oscillation induced by scattering at a surface step is intrinsically damped and its intensity falls off like $1/\sqrt{kx}$, where k is the electron wave vector and x the distance from the step. The phase relaxation length leads to an additional damping e^{-x/L_ϕ} of the interference pattern and can be determined by fitting a model function to the data. $L_\phi(E)$ and thus $\tau(E)$ could be extracted from dI/dV scans across step edges for the Shockley-type surface states on Ag(111) and Cu(111). This method has the advantage that quasiparticle lifetimes can be measured as a function of excess energy and not only at the surface-state band edge $\bar{\Gamma}$ as with the STS peak width method described by Li et al. (1998a). Lifetimes of hot electrons injected by the STM tip have been measured for energies from the Fermi level to 3 eV excess energy (Vitali et al., 2003; Burgi et al., 1999). Figure 17-24 shows a schematic energy diagram of the tunneling process: an electron tunnels from the Fermi level of the tip into an unoccupied level of the metal, from which, after a short time, it will decay to the ground state. Experimental lifetimes of hot electrons as a function of excess energy are shown in Figure 17-25. These energy-dependent lifetime measurements of hot electrons have been interpreted in terms of e–e scattering, since e–ph lifetimes are independent of energy and exceed the measured lifetimes considerably.

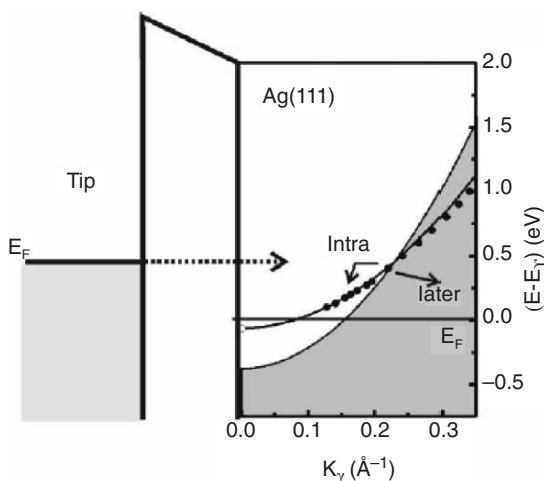


Figure 17-24. Schematic energy diagram of the tunneling process of hot electrons into the surface state band (sample bias positive with respect to the tip). Dots indicate experimental measurements in Vitali et al. (2003). Electrons injected at an energy E scatter inelastically with other electrons into unoccupied states with an energy smaller than E . (From Vitali et al., 2003.)

The experimental data are found to scale with $(E - E_F)^{-2}$ above 0.5 eV as predicted by Fermi-liquid theory, but deviate from the quadratic scaling law at low energies. Calculations showed that the main scattering process at low energies are inelastic intraband e-e scattering and e-ph scattering, whereas at higher energies e-e interband scattering with bulk electrons dominates.

To determine the e-ph scattering contribution to the lifetime, temperature-dependent lifetime measurements have been done with low-energy quasiparticles (Jeandupeux et al., 1999). Constant-current

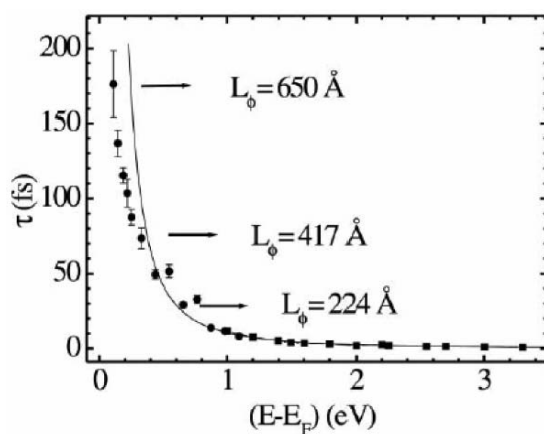


Figure 17-25. Experimental lifetime of hot surface state electrons at the Ag(111) surface as a function of energy. The solid line is a fit of the high energy data (solid squares) to the expected $(E - E_F)^{-2}$ behavior in the Fermi liquid model. (From Vitali et al., 2003.)

line scans across the oscillating LDOS at very low bias voltage perpendicular to step edges on Ag(111) have been measured in the temperature range of 3.5–178 K and compared with theoretical calculations. By measuring at the Fermi energy, the lifetimes are not reduced by e–e scattering (since the e–e lifetime is large at E_F). The measurements showed that the temperature-dependent damping of the standing waves at step edges could be very well described by Fermi–Dirac broadening alone. Therefore only a lower limit for the e–ph lifetime could be given, pointing to significantly longer lifetimes than all previous data.

Below 1 eV the phase relaxation length L_ϕ becomes longer and longer and the damping effect for LDOS oscillations at step edges becomes small. Measuring the additional damping of the oscillating LDOS caused by L_ϕ beyond the inherent $1/\sqrt{kx}$ decay at energies close to the Fermi level is very demanding. It has been proposed that by choosing a closed scattering geometry such as a quantum corral assembled with the low-temperature STM, reflections add up and contribute to a high total intensity, which allows the measurement of small effects. This idea was realized and the lifetime of surface-state electrons confined in an artificial atomic structure has been measured at various energies on an Ag(111) surface at 6 K (Braun and Rieder, 2002). Ag atoms were taken out of the surface by applying short voltage pulses to the tip. By means of lateral manipulation, single Ag atoms were then assembled along closed-packed row directions to form a triangle with a base length of 24 nm as shown in Figure 17–26. The surface-state electrons are scattered by the positioned Ag atoms, resulting in a standing wave pattern as can be seen in Figure 17–26. For the lifetime determination, dI/dV maps were recorded in constant height mode inside the triangle. Lifetimes have been determined by fitting theoretically calculated dI/dV maps to the experimental maps that have been recorded in the energy range of –55 to +796 meV with respect to the Fermi level. The model to calculate the dI/dV maps contained three fit parameters: the phase shift and absorption of the electron wave during scattering events and the phase-relaxation length. The extracted $\tau(E)$ data also reproduced the $\tau \propto E^{-2}$ law predicted by Fermi liquid theory for a 2DEG (Pines, 1966). Electrons injected at an energy E scatter inelastically with other electrons into unoccupied states with an energy smaller than E . This results in a singularity of the lifetime at the Fermi energy due to the reduction of the number of available states into which to scatter. In accordance with theory, all STM lifetime measurements showed a sharp maximum at the Fermi energy.

3.1.5 Stark Effect

The influence of the high electric field between the STM tip and the sample on STS spectra has been measured in an LT-STM study (Kroger et al., 2004). In metals, the electric field is efficiently screened by the conduction electrons. Nevertheless, the wavefunction of surface state electrons is evanescent into vacuum and can thus be affected by the electric field. By varying the electric field of the microscope through the resistance of the junction from 1 G Ω (low field) down to 60 k Ω (high field), it could be demonstrated that the STS spectra of the Au(111) and

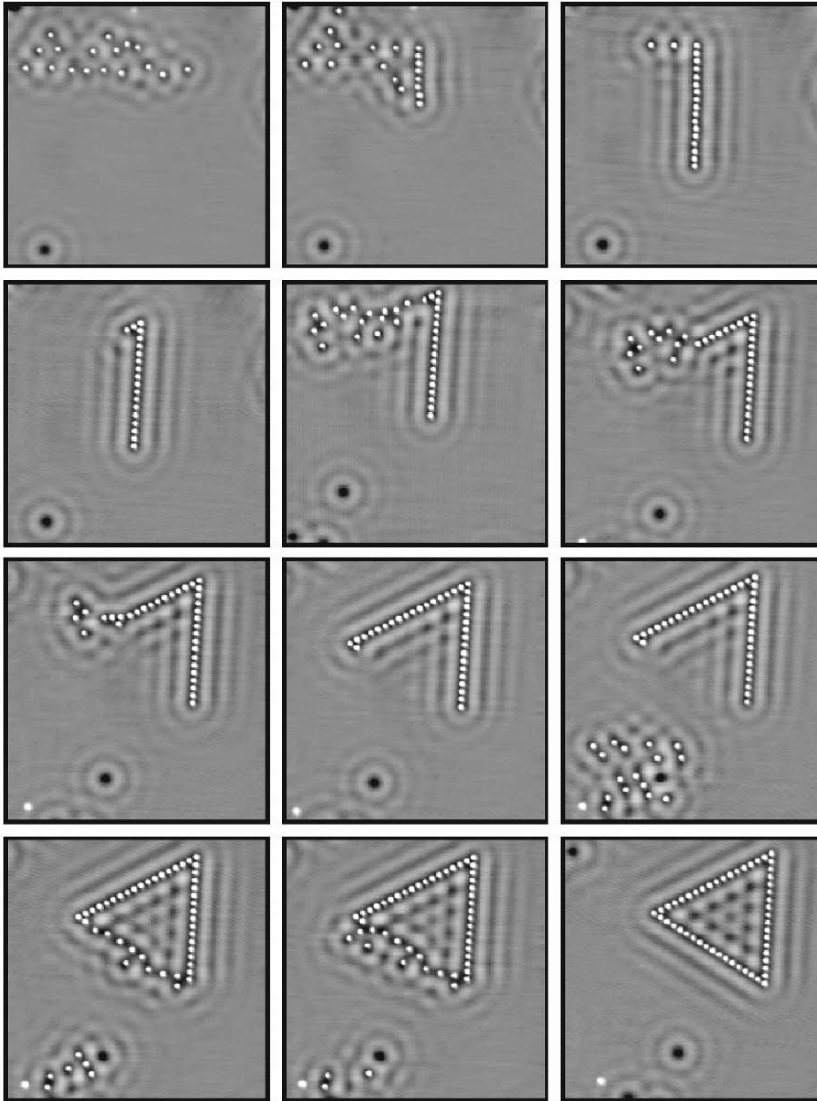


Figure 17–26. Series of images showing the construction of the triangle consisting of 51 Ag atoms on an Ag(111) surface (scan area $49.3\text{ nm} \times 49.3\text{ nm}$). (From Braun, 2001.)

Cu(111) surface states undergo a downward shift that was attributed to the Stark effect. The presence of a Stark effect in the noble metal surface states even at usual tunneling parameters suggests that this effect is quite common for STS.

3.1.6 Long-Range Interactions between Adatoms

The previous studies on standing wave patterns in the electron density of surface electron states concentrated on the effects caused by the adatom scatterers on the surface-state electrons. On the other hand, surface-state electrons give rise to an interaction between the scatter-

ers. This surface-state-mediated interaction is long ranged and oscillatory in nature. Quantitative measurements of the long-range interaction energy between single Cu adatoms on the Cu(111) surface have been performed with a low-temperature STM operated at 9–21 K (Repp et al., 2000). The main criteria of the theory of surface-state-mediated interaction energy (Lau and Kohn, 1978) could be reproduced in the experimental data: the interaction energy is oscillatory with a period of $\lambda_F/2$, where λ_F is the Fermi wavelength, and the envelope of the magnitude decays as $1/d^2$, where d is the adatom separation.

3.1.7 Magnetic Atoms at Surfaces: Kondo Effect

The smallest magnetic structure in condensed matter physics is a single magnetic atom in a nonmagnetic host. The localized spin of the magnetic atom interacts with the spin of the surrounding delocalized conduction electrons. For temperatures below a characteristic Kondo temperature T_K , this interaction causes the conduction electrons of the host metal to condense into a many-body ground state that collectively screens the local spin of the impurity. This screening cloud exhibits a set of low-energy excitations called the Kondo resonance. Due to the formation of the screening cloud, the LDOS near the Fermi energy is enhanced at the site of the impurity. The Kondo resonance disappears at temperatures above T_K , and it energetically splits by the Zeeman energy in an applied magnetic field (Hewson, 1993). The Kondo resonance shows up as a peak at the Fermi level in high-resolution photoemission spectroscopy.

This resonance can be probed locally by STS and manifests itself as a sharp (~ 10 meV wide) depression in the differential conductivity at the Fermi energy. The feature is localized to within ~ 1 nm of the Kondo impurity, and is observed only at low temperatures. Co impurities on Au(111) have been studied (Madhavan et al., 1998) at 4 K. The Kondo resonance line shape has been interpreted in terms of quantum interference resulting of two possible tunneling channels, one direct channel into the Co d -orbital and another channel into the surrounding continuum of conduction band states. A similar observation has been made for Ce adatoms on Ag(111) (Li et al., 1998c). A later experiment with Co adatoms on Cu(100) and Cu(111) (Knorr et al., 2002) showed that while at the Cu(111) surface both tunneling into the localized impurity state and into the substrate conduction band contribute to the Kondo resonance, tunneling into the conduction band dominates for Cu(111) (Figure 17–27).

Another landmark STM experiment showed that when a magnetic atom is placed at one focus of a properly sized empty elliptical corral built from magnetic atoms, a “mirage” of the Kondo signature in the LDOS is cast to the opposite empty focus (Manoharan et al., 2000) more than 7 nm away. The experiments were performed with Co atoms on Cu(111) with a low-temperature STM at 4 K well below the Kondo temperature of the system. Co atoms were evaporated onto the clean Cu surface at 4 K. The ellipses were constructed by use of the adatom sliding process (Eigler and Schweizer, 1990; Stroscio and Eigler, 1991). Figure 17–28 shows the results of these measurements, simultaneously

acquired constant current images, and dI/dV maps of the corrals. dI/dV maps measured on the corrals without an interior atom have been subtracted from the maps measured with an interior atom to remove the contribution of the LDOS oscillation and emphasize the Kondo signal. The constant current images (Figure 17-28a and b) show the location of the Co atoms and oscillations in the LDOS due to two-dimensional confinement of the surface-state electrons whereas the differential conductance spectra show the Kondo signature of Co inside the corral. The striking results reveal two interior positions that show a Kondo resonance: one signal is centered on the real Co atom at the left focus, while the other signal is centered on the empty right focus. The localized electronic structure has been projected from the occupied focus to the unoccupied focus. Tunneling spectra taken at both foci showed that the mirage at the right focus is a faithful spectroscopic replica of the real atom at the left focus attenuated by a factor of eight.

An energy-dependent phase shift of electrons scattering off magnetic impurities has been shown to explain the observed quantum

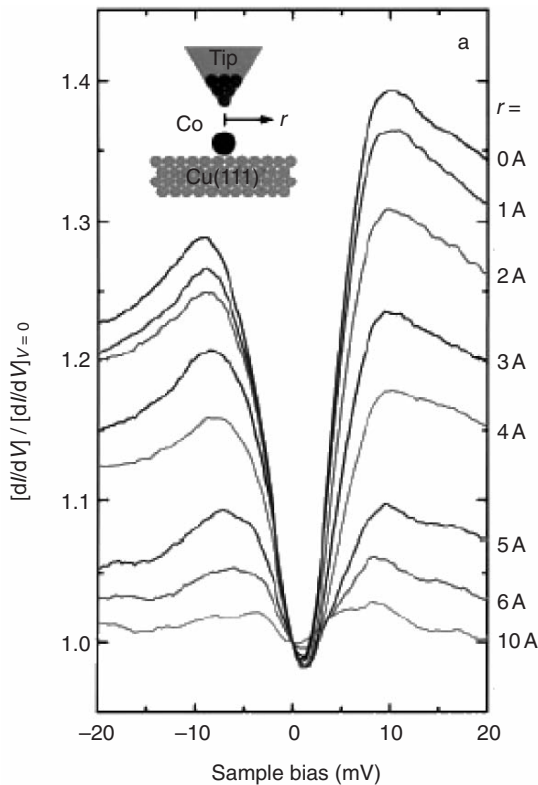


Figure 17-27. Spectroscopic signature of the Kondo resonance around a single Co atom on Cu(111). Tunneling spectra shown have been acquired over the Co atom with increasing lateral displacement r . The spectroscopic feature is narrow (9 mV FWHM) and dies off over a lateral length scale of 1 nm. Measurements with different tips showed nearly identical features. (From Manoharan et al., 2000. Reproduced by permission of IBM Research, Almaden Research Center.)

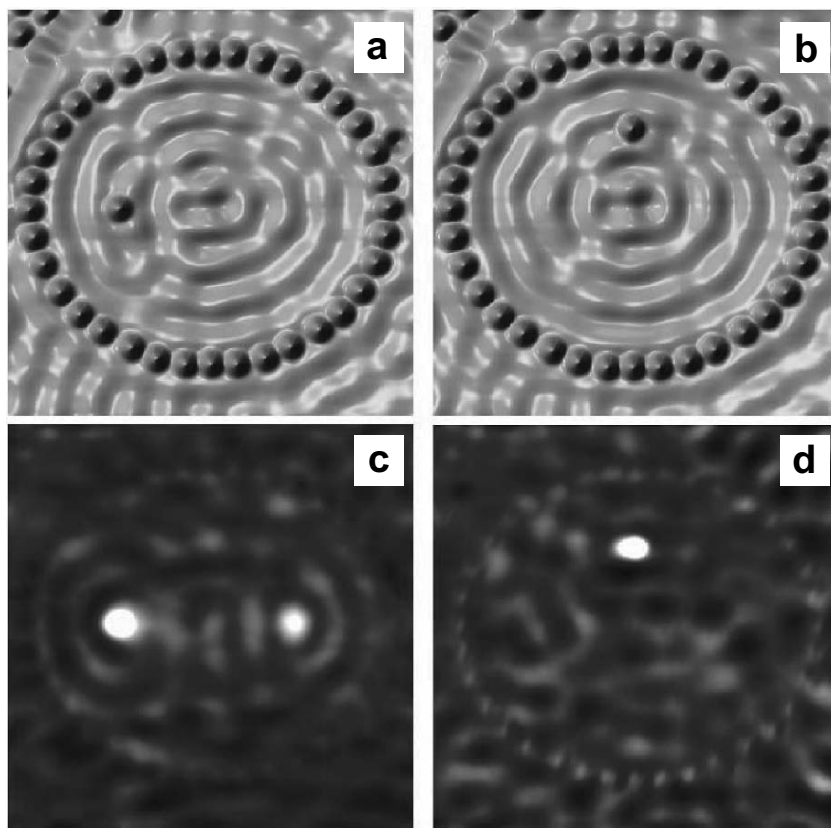


Figure 17–28. Quantum mirage. STM topographs of two elliptical corrals with (a) a Co atom at one focus and (b) a Co atom off focus. Associated dI/dV maps showing in (c) the Kondo effect projected to the empty right focus, resulting in a Co atom mirage; in (d) the mirage vanishes. (From Manoharan et al., 2000. Reproduced by permission of IBM Research, Almaden Research Center.)

mirage (Fiete et al., 2001) and Schneider et al. (2002) have shown experimentally that this phase shift can also be accurately determined with low-temperature STM when the Co atoms are simply adsorbed on Ag(111) and are not placed in artificial resonators.

3.1.8 Fermi Contour Imaging

It has been demonstrated that with a low-temperature STM an image of the surface Fermi contour is directly obtainable by Fourier transforming a low-bias constant-current image of the standing waves in the LDOS (Petersen et al., 1998, 2000ab; Sprunger et al., 1997). The wavevectors of electrons with an energy at the Fermi level are confined to the surface Fermi contour, and consequently standing wave patterns observed in low-bias STM images contain information about the entire Fermi contour. This information can be extracted by performing a simple two-dimensional Fourier transform of the STM image. Thus the power spectrum of a low-bias STM image contains an image of the surface Fermi contour. For an isotropic free electron like surface state a ring is observed in the Fourier transform of the oscillating LDOS image,

as shown in Figure 17–29. The radius of this ring is twice the Fermi wave vector $2k_F$. This is due to the fact that waves in the charge density are observed rather than in the actual wavefunction. Application of this method to anisotropic surfaces was undertaken by Hofmann et al. (1997), who studied the Be(10 $\bar{1}$ 0) surface. An image of the standing waves on Be(10 $\bar{1}$ 0) recorded at 4 K is shown in Figure 17–30a. In contrast to isotropic surfaces, standing wave patterns on Be(10 $\bar{1}$ 0) are visible only when originating from steps in the $\bar{\Gamma}\bar{M}$ direction, whereas no waves from the steps in the $\bar{\Gamma}\bar{A}$ direction are present. The two-dimensional Fourier transform of the standing wave pattern shows the Fermi contour at the SBZ boundary as half ellipses. Obviously the screening of defects at the surface is highly anisotropic due to the orientation of the Fermi contour. This is a beautiful example of the intimate coupling between the Fermi contour and the screening at the surface.

3.2 Single Atom and Molecule Manipulation

The tip of an STM always exerts a force on an adsorbate atom or molecule. This force contains both van der Waals and electrostatic contributions. By adjusting the position and the voltage of the tip, the magnitude and direction of this force may be tuned, thereby enabling atomic scale manipulation processes. Two basic modes of transport of atoms or molecules have been identified: lateral manipulation, i.e., moving particles along the surface by which they maintain contact with the substrate, and vertical manipulation, by which the particles are picked up by the tip and released back to the surface at a desired place. To realize a lateral manipulation experiment, the tip is positioned on the adsorbate and its height reduced to 0.2–0.4 nm by reducing the tunneling resistance from $\sim 1\text{ M}\Omega$ to $100\text{ k}\Omega$. The manipulation is then performed in constant-current mode and the particle is dragged

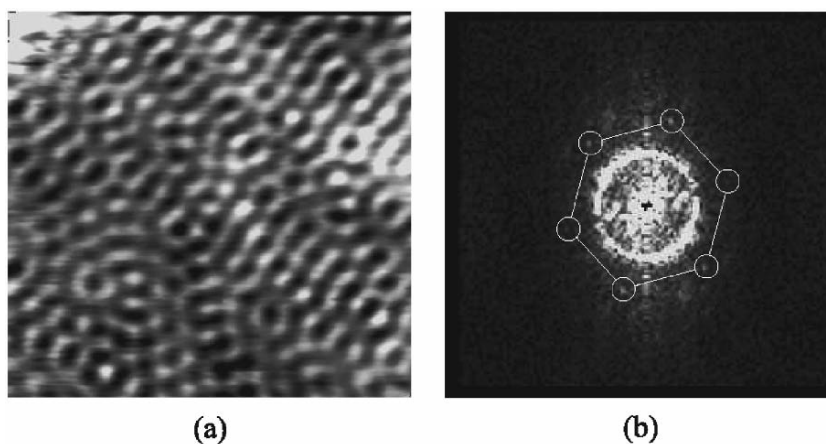


Figure 17–29. Imaging the Fermi contour. (a) Constant-current image of the Be(0001) surface at 150 K showing LDOS oscillations. (b) Two-dimensional Fourier transform of (a) showing a ring-shaped Fermi contour as expected for an isotropic free electron-like surface state. The spots are reciprocal lattice vectors and scale reciprocal space. (From Petersen et al., 2000a.)

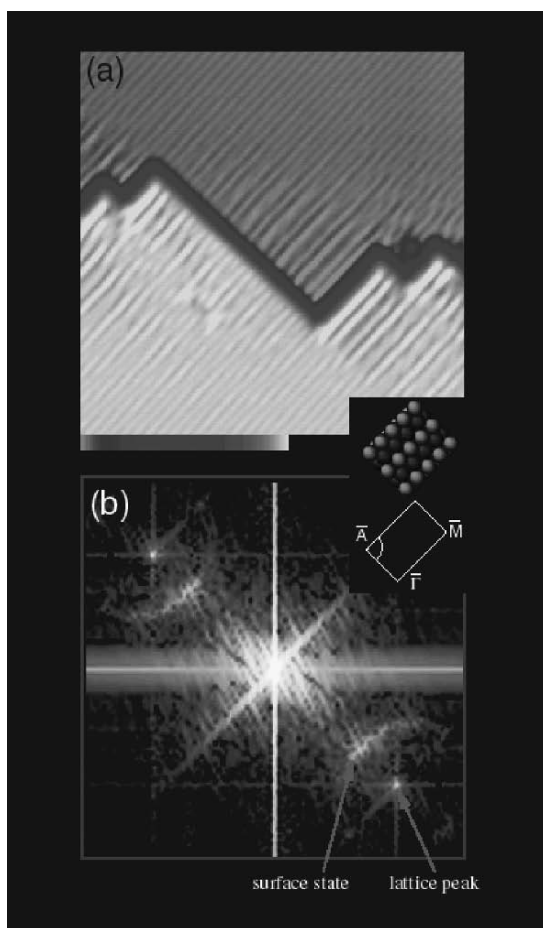


Figure 17-30. Imaging the Fermi contour. (a) Constant current image of the Be(1010) surface at 4 K with a surface step. Anisotropic scattering at step edges results in LDOS oscillations mainly in one direction. (b) Logarithm of the power spectrum of the image. The elliptically shaped features originate from the Fermi contour of the anisotropic surface states. Inset shows a real space top view of the surface and the surface Brillouin zone with the Fermi contour. (From Hofmann et al., 1997.)

by lateral movement of the tip to the final position. As single metal atoms or small molecules are mobile at ambient temperatures, experiments on small adsorbates located on low index metal surfaces have been conducted at low temperatures. Using an LT-STM reduces or eliminates thermally activated diffusion of adatoms on the surface and enables the experimenter to build stable nanostructures or synthesize molecules atom by atom. The electronic properties of these artificial structures can then be studied by STS.

Reliable lateral manipulation and build-up of extended nanostructures on an atomic basis with adsorbed atoms and small molecules at low temperatures have been demonstrated by several groups. Don Eigler's group at IBM-Almaden assembled arrays of Xe atoms on Ni(110) at 4 K (Eigler and Schweizer, 1990) (Figure 17-31). An atomic switch has

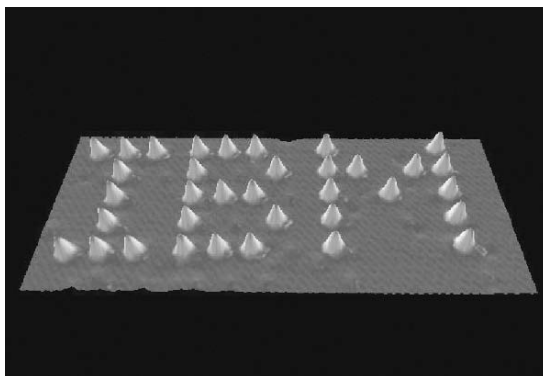


Figure 17–31. Patterned array of xenon atoms on an Ni(110) surface. The atomic structure of the nickel surface is not resolved. Each letter is 5 nm from top to bottom. (From Eigler and Schweizer, 1990. Reproduced by permission of IBM Research, Almaden Research Center.)

been realized where a xenon atom is moved reversibly between two stable positions on the STM tip and a nickel surface (Eigler et al., 1991). Heat-assisted electromigration (Ralls et al., 1989) has been identified as a possible explanation for the vertical manipulation of a xenon atom between a nickel surface and the tip (Eigler et al., 1991). Several other possible mechanisms responsible for lateral and vertical manipulation processes have been investigated (Stroscio and Eigler, 1991). Field-assisted diffusion, which relies on the static and induced dipole moment of the adatom in the strong electric field under the tip, has been identified as being responsible for lateral manipulation of Cs atoms on GaAs(110) at room temperature (Whitman et al., 1991). The lateral manipulation of Xe on Ni(110) has been attributed to the chemical bonding force between the adatom and the tip, since it is not sensitive to the sign or the magnitude of the electric field, the voltage, or the current. It depends only on the separation of the tip from the substrate (Stroscio and Eigler, 1991). A possible mechanism for vertical manipulation of adatoms is transfer on contact, whereby the tip is moved close to the adsorbate until the energy barrier separating them is lowered enough that thermal activation is sufficient for atom transfer. This process does not need an electric field, whereas field evaporation, another possible mechanism, relies on the strong field to create ions that are evaporated over the Schottky barrier. Inelastic tunneling, which leads to multiple vibrational excitations of the atom–surface bond until the bond is broken, has been used to explain the desorption of hydrogen atoms along single dimer rows of the Si(100)- 2×1 surface. Linewidths as small as 1 nm have been achieved (Foley et al., 1998) in these experiments. Inelastic tunneling was also used to selectively cut single molecular bonds in O₂ on Pt(111) (Stipe et al., 1997b).

Lateral manipulation of Cu atoms and Co and C₂H₄ molecules on Cu(211) in constant current mode has been demonstrated by Meyer et al. (2001) at 15 K. The measured tip height curves during pulling of a Cu atom along a close-packed row on Cu(111) showed characteristic

jumps due to the atom hopping from one fcc adsite to the next. By recording the tip height during adatom manipulations, it is possible to distinguish three different types of lateral manipulation: pushing, pulling, and sliding (Bartels et al., 1997a). During a pulling process the adsorbate is situated behind the tip apex with respect to the manipulation direction (attractive tip–adatom interaction). By applying larger forces than for pulling, the adsorbate remains under the tip apex without escaping sideways from the tip trajectory (sliding mode, attractive tip–adatom interaction). Small molecules can be manipulated by the pushing mode, where the adsorbate is in front of the tip (repulsive tip–molecule interaction).

Bias polarity-independent STM-induced desorption of individual NH_3 molecules from Cu(111) has been demonstrated at 15 K, which sometimes leads to transfer of the molecule to the tip (Bartels et al., 1999). Vertical manipulation of C_3H_6 molecules at low temperatures has also been shown (Meyer et al., 1996). Pb monomers and dimers on Cu(211) have been moved laterally at 20 K. Vertical manipulation of CO molecules on Cu(111) has been shown (Bartels et al., 1997b), whereby a CO molecule could be reliably transferred between the surface and the tip. This ability adds chemical sensitivity to STM: with a tip having a CO molecule at its apex, chemical contrast between otherwise similar appearing adsorbates has been achieved: CO molecules on Cu(111) always appear as depressions independent of the bias polarity when imaged with a clean tip. However, when imaged with a tip having a CO molecule at its apex, they appear independently of bias polarity as protrusions. This inversion of shape allows chemical-sensitive imaging, as shown in Figure 17–32. Here, O_2 and CO molecules were adsorbed on the surface. Both adsorbates were imaged as depressions in the STM images with a clean tip. Picking up the dark spot indicated with a white arrow yielded a contrast reversal for most dark spots, while some of them (black arrow) stayed the same. The conclusion was that such spots originate from oxygen, while the others are CO molecules.

Not only can adsorbates be manipulated, but single native substrate atoms can be removed in a controlled manner from differently coordinated sites of the substrate by using lateral manipulation techniques (Meyer et al., 1997). This ability may be of importance in gathering information about subsurface defects or to identify surface atoms with a time-of-flight analyzer (Weierstall and Spence, 1998) after transferring the atoms to the tip. Reversible lateral displacement of specific Si adatoms on the Si(111)- 7×7 surface has been reported at low temperatures (30–175 K) (Stipe et al., 1997a). A single adatom could be reversibly displaced as an atomic switch and its position monitored with the tunneling current.

The continued miniaturization of electronic devices is leading to an increasing interest in the application of single molecules in nanoelectronics (Joachim et al., 2000). In this context, low-temperature STM is a fundamental technique to study different molecular conformations and to manipulate single molecules, bringing them in electronic contact with atomically ordered nanoelectrodes. Extension of the lateral

manipulation mode from small to large molecules has been demonstrated at low temperatures (Moresco et al., 2001). Although big molecules have been positioned in a controlled way by room temperature STM (Jung et al., 1996, 1997; Gimzewski and Joachim, 1999), the stability and low noise level of the LT-STM are necessary to obtain detailed and quantitative information about the manipulation process. Complex organic molecules show a different translational movement under the influence of the STM tip than atoms or simple molecules. It has been shown that Cu-TBPP, a specially designed porphyrin-based molecule, cannot be moved at low temperatures by lateral manipulation in constant-current mode. Therefore a constant-height mode for lateral manipulation of these molecules was used, where the current signal can be used to extract information about the internal movement of the molecule under the action of the tip (Moresco et al., 2001). The central group of a $C_{90}H_{98}$ molecule (known as Lander) has been used as a model system for a molecular wire. The Lander molecule has the same molecular legs as Cu-TBTT, but it has a central polyaromatic molecular wire instead of a central porphyrin ring (Figure 17-33). The molecule was designed as a central molecular board lifted above the substrate by four legs and should act as a short (1.7-nm-long) molecular wire when contacted to an atomic step (Langlais et al., 1999). In an LT-STM study the conformational changes induced in the Lander upon adsorption on Cu(001) have been investigated (Kuntze et al., 2002). It has been found that the legs of adsorbed Lander molecules are rotated and deformed. This leg rotation results in a lowering of the central wire to

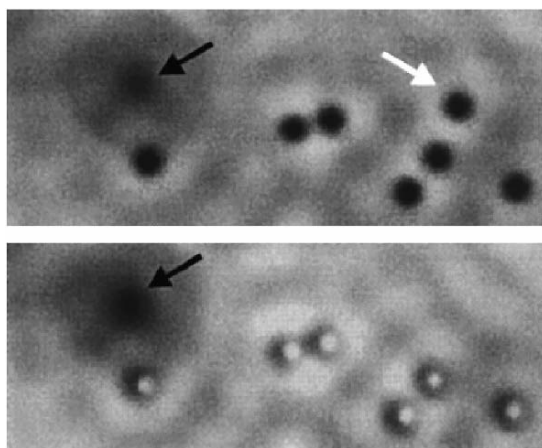


Figure 17-32. Chemical contrast with functionalized tip at 15K. (Top) STM image of oxygen and CO on Cu(111) obtained with a clean metal tip. All adsorbates appear as indentations (dark spots) in the image. (Bottom) Same part of the surface imaged after picking up the adsorbate marked with a white arrow. All adsorbates imaged as protrusions correspond now to CO molecules, whereas the adsorbate marked by the black arrow still appears as an indentation and is therefore identified as oxygen. The fact that the appearance of CO changes when imaged with a CO molecule at the tip has been verified on a surface where only CO molecules were adsorbed. (From Bartels et al., 1997b.)

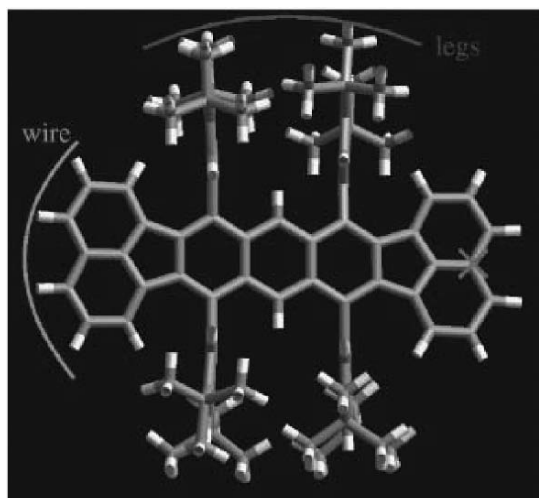


Figure 17–33. Top view of the chemical structure of the Lander molecule, which was designed to be a model system for a molecular wire. The molecular wire and the four legs that support the wire are shown. (From Moresco et al., 2003.)

0.37 nm above the surface compared to 0.7 nm for the gas phase molecule. Upon adsorption at room temperature the molecules diffuse to step edges and are stabilized with their wire parallel to step edges, which prevents good electronic contact (Kuntze et al., 2002).

The problem of controlling the electronic contact between the molecule and its electrodes has been addressed in an investigation with the Lander molecule. Reproducible contact formation was obtained by lateral manipulation of the Lander on Cu(111) with an LT-STM (Moresco et al., 2003). The molecules were adsorbed at 70 K instead of room temperature to avoid postdeposit thermal diffusion. Figure 17–34 (A3) shows an STM image of a single Lander on a terrace where the four bumps are attributed to the four legs of the molecule. When a molecule is pushed to the step with its central wire parallel to it, it reaches a final conformation imaged in Figure 17–34 (B3). Separated by the legs, the central wire is not interacting with the step edge and the standing wave pattern (LDOS oscillations) (B4) on the upper terrace has not changed. If the molecule is repositioned with its wire oriented perpendicular to the step edge, a notable modification of the standing wave pattern on the upper terrace is observed. The amplitude of the standing wave pattern is reduced at the contact point compared to the clean step edge case (4C). Simulations of the standing wave pattern indicated that the perturbation is caused by the terminal part of the molecular wire on the upper terrace. This contact area is visible in the STM image as an additional small bump (C3) and its location was confirmed by elastic-scattering quantum chemistry STM image calculations (Sautet and Joachim, 1991) (C2). The molecule could also be decontacted by reverse lateral manipulation, and the original step edge and molecule image were recovered.

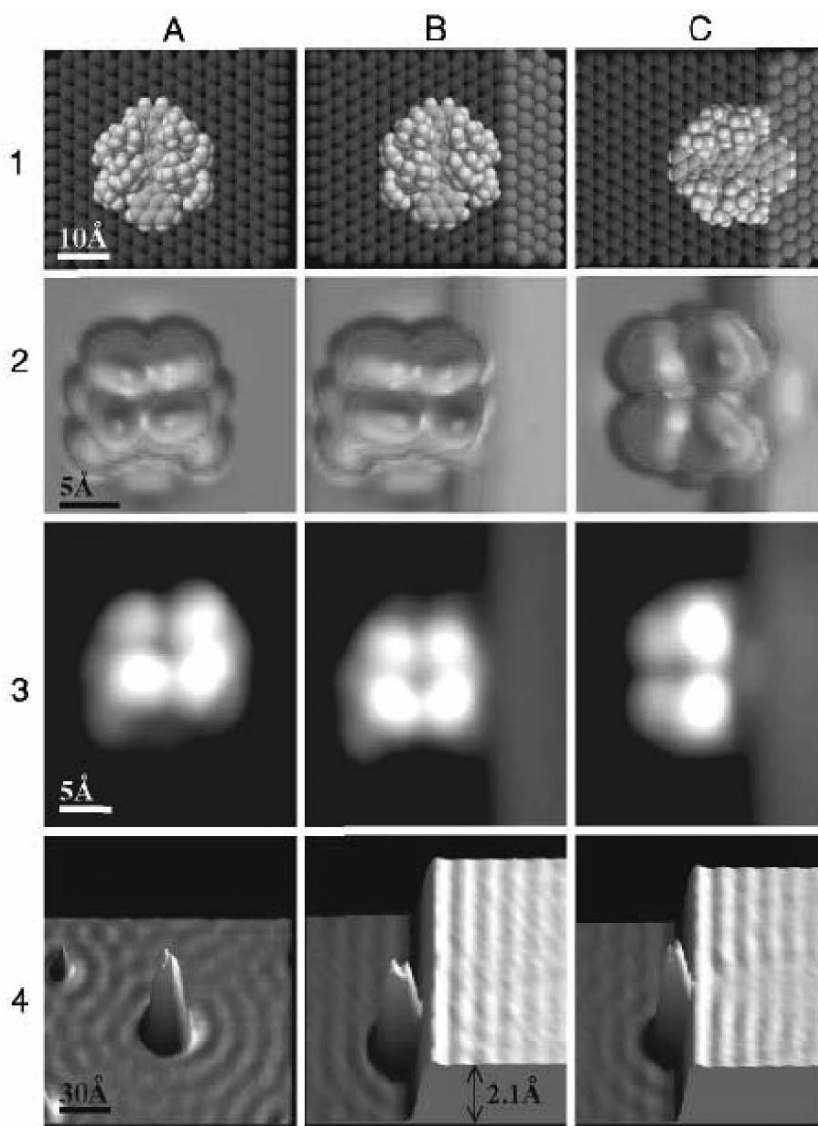


Figure 17-34. Contacting a molecular wire to a step: Lander molecules on a step-free Cu(111) surface (column A) and contacted to a (100) step. Molecular wire parallel (column B) and orthogonal (column C) to the step. Row 1: Sphere models of molecular structures. Row 2: Calculated STM images corresponding to the sphere models above. Row 3: STM images at 8 K. Row 4: STM measurements showing LDOS oscillations. In (C2) and (C3), an additional bump appears corresponding to the contact point of the wire to the step. Modification of the standing wave pattern on the upper terrace is observed only when the wire is orthogonal to the step (C4). (From Moresco et al., 2003.)

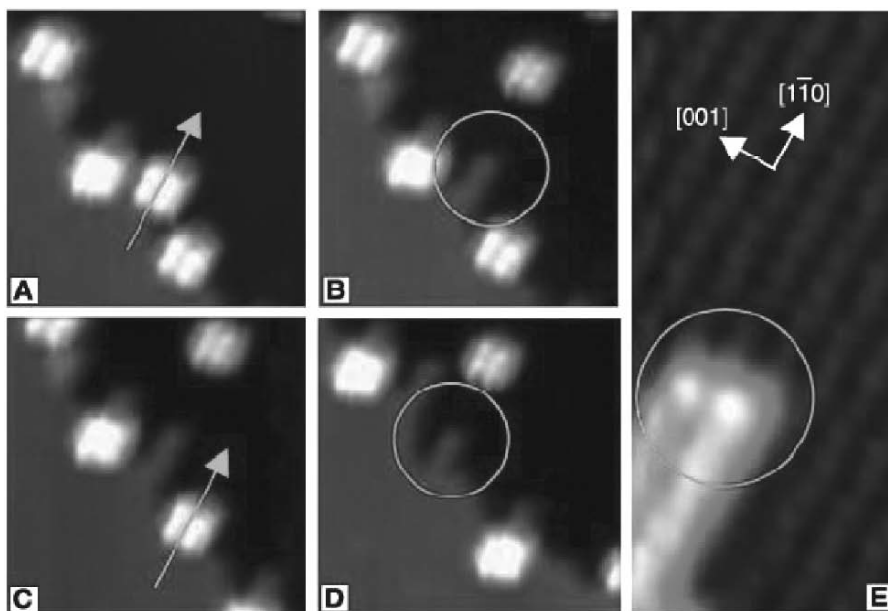


Figure 17-35. Self-assembly of nanowires at step edges initiated by Lander molecule adsorption. (A–D) Low-temperature STM images of a manipulation sequence of the Lander molecules from a step edge on Cu(110). The arrows show which molecule is being pushed aside; the circles mark the tooth-like structures that are visible on the step where the molecule was docked. (E) Zoom-in STM image showing the characteristic two-row width of the tooth-like structure after removal of a single Lander molecule. (From Rosei et al., 2002.)

In a variable-temperature STM experiment it has been found that the Lander molecule can act as a template, self-fabricating short metallic nanostructures at step edges. Lander molecules were adsorbed at room temperature on the Cu(110) surface and their conformation and anchoring at step edges were studied at 100 K. At room temperature the Cu kink atoms are highly mobile and the Lander molecule reshapes the fluctuating Cu step adatoms into tooth-like nanostructures perpendicular to the step edges. The dimension and shape of the Lander molecule form a perfect template for a double row of Cu atoms. Moving the molecule away from the step edge by lateral manipulation at low temperature revealed the underlying restructuring of the step edges (Figure 17-35). Upon adsorption of the Lander molecules at 150 K, no restructuring of the Cu step edges was observed, since the mobility of the Cu kink atoms at this temperature is not high enough for the molecular template to be effective.

At low temperatures, the short Cu nanowire acts as a sliding “rail” for the Lander molecule. By moving the molecule to the end of such a nanostructure, a model geometry can be obtained where one end of the central molecular wire is electronically connected to the metallic wire. A detailed study of the lateral manipulation of the Lander molecule along such an atomic wire has been performed with an LT-STM at 8 K (Grill et al., 2004). Lateral displacements of the molecule have been separated into monoatomic steps and it has been shown that single molecular legs can be rotated reversibly while keeping the

central wire fixed. Comparison with theory confirmed that the central wire is contacted to the metallic Cu atomic wire.

In a recent LT-STM study of the Lander molecule, it has been shown that despite the fact that the legs elevate the molecular wire away from the surface, there is still an electronic interaction between the central wire and the surface states of the substrate (Gross et al., 2004). This was shown by comparing the standing wave patterns of surface-state electrons scattered off the molecule with calculated patterns taking into account scattering from different areas of the molecule.

Chemical bond formation was studied with an LT-STM (Lee and Ho, 1999). Individual iron atoms were evaporated and coadsorbed with CO molecules on an Ag(110) surface at 13K. A CO molecule was transferred from the surface to the STM tip and bonded with an Fe atom on the surface to form Fe(CO). A second CO molecule could then be added to form Fe(CO)₂. This is shown in Figure 17–36. The adsorption sites of

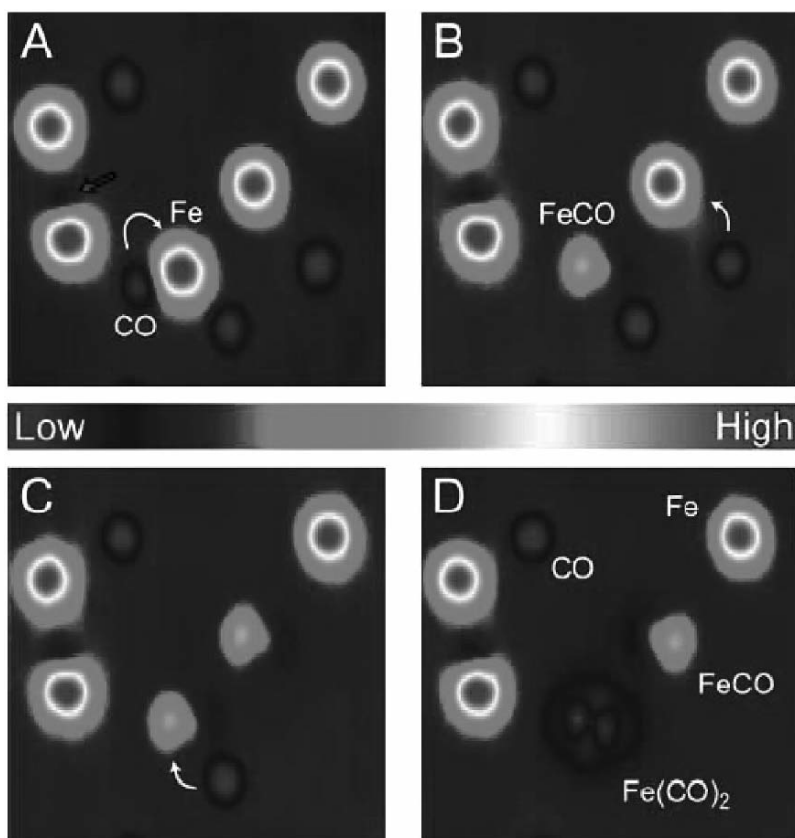


Figure 17–36. Bond formation induced with STM tip. A sequence of STM constant-current images at 13K showing the formation of Fe–CO bonds by vertical manipulation. Fe atoms are imaged as protrusions and CO molecules as depressions. The white arrows indicate the pair of adsorbates involved in each bond formation step. In (B) and (C) a CO molecule has been picked up and bonded to an Fe atom to form Fe(CO). In (D) a second CO molecule has been bonded to Fe(CO) to form Fe(CO)₂. (From Lee and Ho, 1999.) (See color plate.)

the reactants could be determined by resolving the underlying Ag lattice with a CO molecule attached to the tip, which leads to increased resolution in the constant-current image. This increase in spatial resolution can be attributed to the more localized wavefunction of the molecule-terminated tip. Analysis with inelastic tunneling spectroscopy provided spectroscopic support for the identification of the created single molecule products with $\text{Fe}(\text{CO})$ and $\text{Fe}(\text{CO})_2$.

Assembly of an artificial nanostructure composed of a copper(II) phthalocyanine (CuPc) molecule bonded to two gold atomic chains on NiAl(110) has been realized with an LT-STM (Nazin et al., 2003b). The electronic structure of this model metal–molecule–metal junction was studied by spatially resolved STS and systematically tuned by varying the number of gold atoms in the chains. Splitting and shifting of molecular orbital energies and modification of the local electronic structure of the electrodes were observed. These effects determine the alignment of the molecular orbital energies with respect to the Fermi energy of the metal and affect the conductivity of the junction.

3.3 Local Inelastic Electron Tunneling Spectroscopy

Besides the dominant elastic electron tunneling process, for which the electron energy is equal in the initial and final state, inelastic tunneling can occur if the tunneling electrons couple to some modes ω in the tunneling junction. Figure 17–37 shows an energy diagram for $T = 0$, illustrating elastic and inelastic tunneling processes. In the case of inelastic tunneling the electron loses energy $\hbar\omega$ to a mode in the tunneling barrier. According to the Pauli exclusion principle, tunneling is possible only if the final state after the inelastic tunneling event is initially unoccupied. The bias voltage dependence of the tunneling current with inelastic tunneling is shown schematically in Figure 17–38. The elastic tunneling current increases linearly, proportional to V . As long as the bias voltage is smaller than the lowest energy mode that can be excited in the gap, inelastic tunneling processes cannot occur. At the threshold bias $V = \hbar\omega/e$, the inelastic channel opens up, and the number

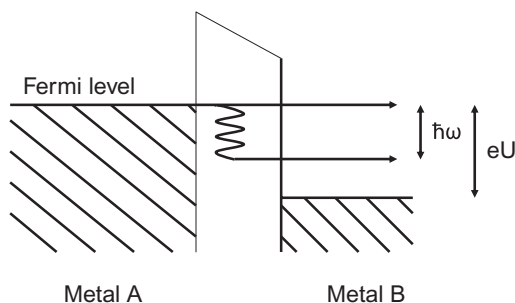
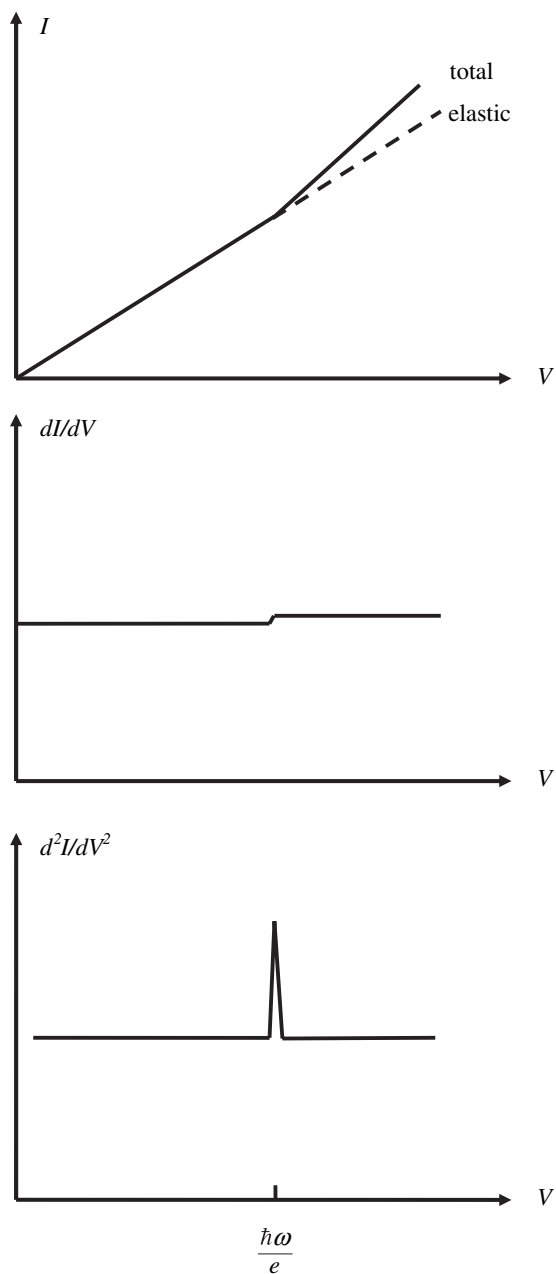


Figure 17–37. Elastic and inelastic tunneling channels. Tunneling electrons can excite a molecular vibration of energy $\hbar\omega$ only if $eU > \hbar\omega$. For smaller energies, there is no final state into which the electron can tunnel. Therefore the inelastic current has a threshold at $\hbar\omega/e$. The increase in conductance at the threshold is typically 1–10% in an STM experiment.

Figure 17–38. Schematic current versus voltage curves with elastic and inelastic tunneling. A kink is observed when the inelastic tunneling channel opens up. The kink becomes a step in the first derivative and a peak in the second derivative.



of electrons using this channel increases linearly with V . Therefore the total current has a kink at the threshold bias voltage. In the differential conductance curve dI/dV , the kink becomes a step and the second derivative d^2I/dV^2 exhibits a peak at the threshold. If several modes ω_i can be excited in the tunneling process, each mode leads to a peak in the differential conductance at the corresponding voltage $V_i = \hbar\omega_i/e$. Inelastic electron tunneling spectroscopy (IETS) can therefore be regarded as a special form of electron energy loss spectroscopy.

IETS has been shown to be a powerful technique for measuring the vibrational spectra of molecules that have been intentionally incorporated into a metal–oxide–metal tunneling junction (Jaklevic and Lambe, 1966). Vibrational spectroscopy can be performed with a variety of other techniques including electron energy loss spectroscopy, infrared absorption spectroscopy, Raman spectroscopy, inelastic neutron scattering, and helium atom scattering. All of these techniques have in common with IETS that they rely on macroscopic numbers of molecules to achieve detectable signal levels. The signal is therefore an average over molecules whose local environment can vary. The major drawback of traditional IETS with planar metal–oxide–metal junctions is that the molecules are buried within the junction, which is difficult to characterize microscopically.

Replacing the oxide layer by vacuum and the top planar electrode by a sharp STM metal tip has made it possible to extend IETS to single adsorbed molecules. One great advantage of performing vibrational spectroscopy with the STM is that the high spatial resolution of STM images permits changes in molecular spectra to be correlated with variations in the local environment on an atomic scale. STM-IETS was proposed as early as 1985 (Binnig et al., 1985b). Since the changes in tunneling conductance resulting from opening of additional inelastic tunneling channels are typically 0.1–1% for planar junctions and 1–10% for STM junctions, the relative stability of the tunneling current has to be better than 1% to obtain reasonable IET spectra with the STM. The physics of tunneling then dictates a tunneling gap stability of better than $\sim 0.005 \text{ \AA}$ over the time it takes to complete one scan of the spectrum (Lauhon and Ho, 2001). Because the vibrational features are very sharp, liquid helium temperatures are required to avoid thermal broadening of the Fermi levels. Hansma (1982) estimated an effective resolution of $5.4k_{\text{B}}T$ ($\sim 140 \text{ mV}$ at room temperature) for inelastic tunneling, while vibrational features are typically only a few millivolts wide. For those reasons, vibrational spectroscopy with the STM has proved difficult. First experiments probing a cluster of sorbic acid molecules adsorbed on graphite at 4 K reported large jumps in the *first* derivative spectrum instead of the expected second derivative spectrum (Smith et al., 1987). The peaks were attributed to characteristic vibrations of molecules. However, due to molecular diffusion events during the measurements, the spectra were not very reproducible and the energies of the peaks were different from those measured in bulk tunnel junctions. Reproducible single-molecule vibrational spectroscopy has been achieved only recently with an LT-STM (Stipe et al., 1998). In these landmark experiments, a Cu(100) surface was dosed with acetylene (C_2H_2) and deuterated acetylene (C_2D_2). Vibrational spectra were acquired at 8 K above single molecules with the use of a tracking scheme to position the tip at the center of the molecule, with lateral and vertical resolution of better than 0.1 and $0.01 \text{ \AA}$, respectively. Contributions from the electronic spectrum of the tip and the substrate could be minimized by subtracting spectra taken over a clean area of the surface from the molecular spectra. The I – V curves from a single molecule and the clean surface (Figure 17–39A) show the expected linear dependence

for metallic junctions. The differential conductance dI/dV shows an increase of 4.2% at 358 mV, resulting from the excitation of the C—H stretch mode (Figure 17–39B). The second derivative d^2I/dV^2 reveals a distinct peak at 358 mV (Figure 17–39C, compare with the idealized view of Figure 17–38). An isotopic shift to 266 mV was observed for deuterated acetylene and the C—D stretch mode. These values are in close agreement with those obtained by EELS for the same molecules on Cu(100). The ability to spectroscopically identify molecules with the STM makes it possible to implement chemical-sensitive STM imaging. This has been demonstrated by recording a d^2I/dV^2 map above both acetylene isotopes. When the dc voltage was fixed at 358 mV, only one of the two molecules (C_2H_2) was imaged, whereas at 266 mV, the other molecule (C_2D_2) was imaged (Figure 17–40). Thus, individual adsorbed molecules can be identified by their vibrational spectra and inelastic images. In contrast, identification and characterization of adatoms and

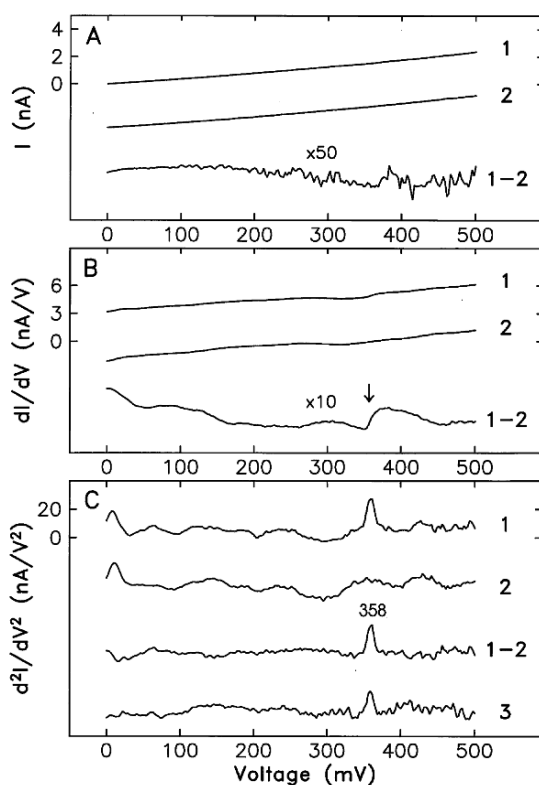


Figure 17–39. Molecular vibrational spectra observed with the STM at 8 K. (A) I – V curves recorded with the STM tip directly over the center of an acetylene molecule (1) and over the bare Cu(100) surface (2). (B) dI/dV on the molecule (1) and on the substrate (2). (C) d^2I/dV^2 on the molecule (1) and on the substrate (2). A peak at 358 mV is visible in the difference spectrum. (3) An average over 279 scans of 2 min each (10 h total data acquisition time directly above the molecule) with a different tip. The conductance change due to inelastic tunneling was 3–4% with different tips. (From Stipe et al., 1998.)

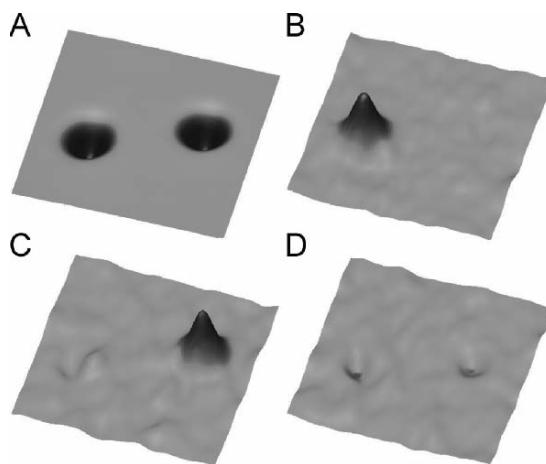


Figure 17-40. Chemical sensitive imaging. Spectroscopic spatial images of the inelastic channels for C_2H_2 and C_2D_2 . (A) Constant-current image of a C_2H_2 (left) and a C_2D_2 (right) molecule. This image is an average of STM images recorded simultaneously with the vibrational images. The molecules appear identical in this normal imaging mode. d^2I/dV^2 maps (vibrational images) of the same area were recorded at (B) 358 mV, (C) 266 mV, and (D) 311 mV. In (B) only C_2H_2 is visible, whereas in (C) C_2D_2 is visible. The symmetric, round appearance of the molecules is attributed to the rotation of the molecule between two equivalent orientations during the experiment. (From Stipe et al., 1998.)

molecules by electronic spectroscopy with the STM are problematic because the electronic energy levels are broadened and shifted upon adsorption and the adsorbed molecule's spectrum becomes convolved with the STM tip's electronic spectrum (Crommie et al., 1993c).

STM-IETS has been used to determine the orientation of individual C_2HD molecules (deuterated acetylene) adsorbed on the Cu(100) surface at 8 K (Stipe et al., 1999a). By setting the bias voltage to the vibrational energy of the C—D stretch mode in C_2HD , and simultaneous recording the constant-current image and the vibrational image (d^2I/dV^2 map), the spatial distribution of the C—D stretch signal within the molecule could be determined. Since the inelastic image has its maximum near the midpoint of the C—D bond, it locates the position of the bond in this case.

The previous two examples show the ability of the LT-STM to resolve internal vibrations of molecules adsorbed on surfaces. The internal modes can be used in surface chemical analysis for the identification of adsorbed species. External vibrations, i.e., vibrations of adsorbed molecules with respect to the substrate, are more sensitive to the interaction of the adsorbed molecule with the substrate. These external modes generally have lower energy than the internal ones, and are not easily accessed by some of the averaging vibrational spectroscopy methods mentioned above. External vibrational modes of benzene molecules on an Ag(110) surface have been detected with an LT-STM at 4 K (Pascual et al., 2001a). These measurements confirmed that the

external vibrations are strongly sensitive to the nature of the molecule–substrate bond. For internal vibration modes, the inelastic signal has been shown to be very localized at the position of the particular bond excited (Stipe et al., 1999a). In contrast the spatial distribution of the inelastic tunneling signal for external modes extends over the whole area of the molecule (Pascual et al., 2001a).

In a detailed study of temperature effects, electronic structure contributions, and tip effects on STM-IETS, Lauhon and Ho (2001) suggested that functionalization of the tip by transfer of a known molecule to the tip offers a means of accessing different vibrational modes. This has been confirmed in later experiments (Moresco et al., 1999; Hahn and Ho, 2001) where CO and C₂H₄ molecules have been transferred to the tip and single CO and O₂ molecules were probed with this tip.

There are a few caveats regarding the application and interpretation of STM-IETS spectra. In STM-IETS as in traditional IETS, there are no strict selection rules. Modes involving motion parallel and perpendicular to the surface can be excited. Modes are not observed for all molecules and not all modes are necessarily observed for any particular molecule. The symmetry of vibrational modes and electronic resonances of an adsorbate seem to give rise to selection rules for vibrational mode detection in STM-IETS (Lorente et al., 2001). A dependence of the STM-IETS signal on the molecular orientation on the surface has been shown for C₆₀ on Ag(110) (Pascual et al., 2002). The spectroscopic maps showed a correlation between the enhanced vibrational signal and orientational symmetry of the adsorbed molecule observed in constant-current mode.

To complicate matters more, it has been shown that vibrational excitation can lead to suppression of elastic tunneling and produce dips instead of peaks in the differential conductance spectrum (Hahn et al., 2000). This has been recently explained theoretically by Lorente (2004). Despite this complexity, there are many advantages of STM-IETS, i.e., that the adsorbate geometry is well defined and the effect of adsorbate orientation can be studied systematically. Recent progress in the theoretical analysis of STM-IETS may greatly enhance its ability to probe chemistry at the spatial limit (Mingo and Makoshi, 2000; Makoshi and Mingo, 2002; Lorente and Persson, 2000; Lorente, 2004).

3.4 STM-Induced Photon Emission

Injection of electrons or holes from the tip of an STM into the surface leads to the emission of light for many materials. The first observation of light emitted from the tunneling junction of an STM in the low-bias tunneling regime ($eV < \Phi$ where Φ is the work function) goes back to Coombs et al. (1988). The highly localized tunneling current allows high spatial resolution that enables experiments with single nanostructures and molecules. Photon emission from the tunneling junction of an STM can be used to measure the optical properties of the sample surface in the nanometer regime. Photon emission from metals involves surface plasmons, which are inelastically excited by the tunneling elec-

trons and then decay by the emission of photons (Persson and Baratoff, 1992). Besides these inelastic excitations, photons can also be emitted due to electron-hole recombination in semiconductors (Abraham et al., 1990). Given that the tunneling currents used for excitation are rather small—typically in the nanoampere range—the photon count rates tend to be small as well. The photon creation efficiency is typically of the order of 10^{-4} – 10^{-3} photons per tunneling electron (Persson and Baratoff, 1992). Increasing the tunneling current in many cases leads to spontaneous surface modifications, which prevents reproducible experiments. Given the low photon count rates, the photon detector has to be optimized for maximum collection efficiency. A number of designs have been reported, which make use of lenses (Berndt et al., 1991b; Hoffmann et al., 2002a), mirrors (Berndt et al., 1991b; Nilius et al., 2000), transparent tips (Smolyaninov et al., 1990; Murashita, 1999), and optical fibers (Arafune et al., 2001). The emitted light can be experimentally investigated using a variety of methods (Reihl et al., 1989):

1. Isochromat spectroscopy: The photon energy is kept fixed while the bias voltage is being scanned (at constant tunneling current). This allows study of the influence of the excitation energy on the intensity of particular emission features. During the acquisition of isochromat spectra the tip–sample distance is varied by the STM feedback loop to maintain constant current. The detectors used are avalanche photodiodes or photomultipliers, where the spectral response is limited by an optical bandpass filter.

2. Fluorescence spectroscopy: The electron energy (i.e., the bias voltage) is kept fixed while the wavelength-resolved distribution of the emitted photons is measured. This involves a grating spectrometer and a cooled CCD detector. Direct information on light-emitting transitions is obtained. This mode allows the observed photon emission to be assigned to elementary processes such as interface plasmons in the tip–sample cavity or interband transitions in the sample.

3. Luminescence spectroscopy: A special case of (2) for semiconductors, where the injected electron thermalizes to the bottom of the conduction band emitting a characteristic photon.

4. Spatial mapping: A selected feature in the spectra is used to generate a high-resolution picture of the intensity of the emitted photons. Spatial maps are usually acquired simultaneously with a constant-current image to relate the spectral map to topographic features. In the case of luminescence this allows the spatial identification of defects, grain boundaries, and dopant concentrations on semiconductor surfaces similar to cathodoluminescence in electron microscopy.

STM-induced photon emission from semiconductors has been compared to cathodoluminescence (CL) (Gustafsson et al., 1998) in scanning electron microscopy, photoluminescence (PL) (Montelius et al., 1992) and inverse photoemission spectroscopy (IPS) (Reihl et al., 1989). Measuring luminescence from semiconductors locally by inducing it with an STM has distinct advantages: Electron tunneling from the tip provides a bright and extremely localized source of electrons. In addition and unique to STM measurements, holes can be injected by biasing

the tip positive with respect to the sample. Comparative measurements of STM-induced photon emission in the field emission regime ($eV > \Phi$) on Si(111) with STS and normal incidence IPS demonstrated a correlation between those methods, but showed also differences due to the local nature and the strong electrostatic field of the STM measurement (Reihl et al., 1989).

STM-induced photon emission from metal surfaces involves excitation of a plasmon localized within the region near the end of the tip (Aizpurua et al., 2000). The plasmon then decays into photons, which are detected in the far field. These plasmons are excited either by inelastic tunneling, or by elastic tunneling into the substrate and subsequent thermalization via plasmon generation. Model calculations favor inelastic tunneling, which occurs in the tunneling gap as the excitation mechanism (Berndt et al., 1991a). This is in contrast to STM-induced luminescence on semiconductors, where hot electron decay within the semiconductor was found to dominate (Abraham et al., 1990). The resulting photon spectrum resulting from plasmon decay is quite broad and has a characteristic energy cutoff determined by the sample bias voltage at eV_{bias} . The spectrum is also sensitive to the geometry of the tunnel junction and the tip material (Berndt et al., 1993). Fluorescence due to hot electron thermalization, on the other hand, is expected to be insensitive to the tunneling gap conditions and produce a distinct peak (or a series of peaks) in the photon spectrum. Measurements on Cu(111) in the field emission regime found a close correlation between oscillations in the conductance, which arise from standing waves in the tip-sample gap, and oscillations in isochromat photon spectra (Berndt and Gimzewski, 1993). These findings supported the view that for metals inelastic tunneling and coupling to a tip-induced plasmon mode occurs in the tunneling gap.

Low temperature in CL (Murashita et al., 1993), PL, and IPS is known to increase the intensities of spectral features and/or reduce thermal broadening, which can obscure important details in the spectrum. This is one reason to perform STM photoemission measurements at low temperatures; the other main reason is to improve the long-term drift stability necessary to record spatial maps (Hoffmann et al., 2004) and photon spectra from single molecules (Qiu et al., 2003).

Low-temperature luminescence experiments on GaAs/AlAs using an LT-STM with a transparent tip as detector have been reported (Murashita, 1997). Isochromat spectra from individual several-atom silver chains assembled with an LT-STM on the NiAl(110) surface have been measured showing sensitivity to the number of atoms in the silver chain (Nazin et al., 2003a). The changes in photon emission with chain size were explained as a quantum size effect. Recently experiments with superconducting tips on a superconducting sample were reported that showed an energy cutoff at $2eV_{\text{bias}}$ in the isochromat photon spectrum. This could not be explained by single electron tunneling and was attributed to tunneling of Cooper pairs (Uehara et al., 2001).

The modes of measurement described above (1–4) either provide lateral or spectral resolution. Spectral and spatial resolution can be obtained simultaneously by recording a spatial map where at each

point in the image a complete fluorescence spectrum is recorded. The tip motion during constant-current imaging is delayed at each pixel for a photon detection time of 0.05–1 s (depending on detector sensitivity). This mode was named spectroscopic imaging (Hoffmann et al., 2002a). From the resulting three-dimensional data set isochromat spatial maps at various wavelengths can be extracted. Moreover, changes of the sample or tip are easily recognized. The first measurements of this kind were conducted by Nishitani et al. (1998) at room temperature. The recording time for a complete spectroscopic image was 1 h. Subsequent improvements in the photon detection system enabled faster recording times and the use of an LT-STM provided less thermal drift (Hoffmann et al., 2002a). Atomically resolved isochromat spatial maps have been recorded of the Au(110) surface at 4 K where the atomic rows appear darker than the troughs (see Figure 17–41). The finding that atomic scale structures cause clear variations of the photon emission characteristics was surprising when it was first observed (Berndt et al., 1995). Theoretically the local photon emission on metals was expected to extend laterally over approximately $50 \text{ \AA}$, which is the lateral extent of the tip-induced plasmons. The first finding of atomic scale features in STM-induced photoemission on Au(110) was explained in terms of the tip–sample distance dependence of the photon emission, which is given by the coupling to localized plasmon modes (Berndt et al., 1995). This model could not explain later measurements that showed that photon emission is reduced by a factor of 5 at Ag(111) steps, whereas it is

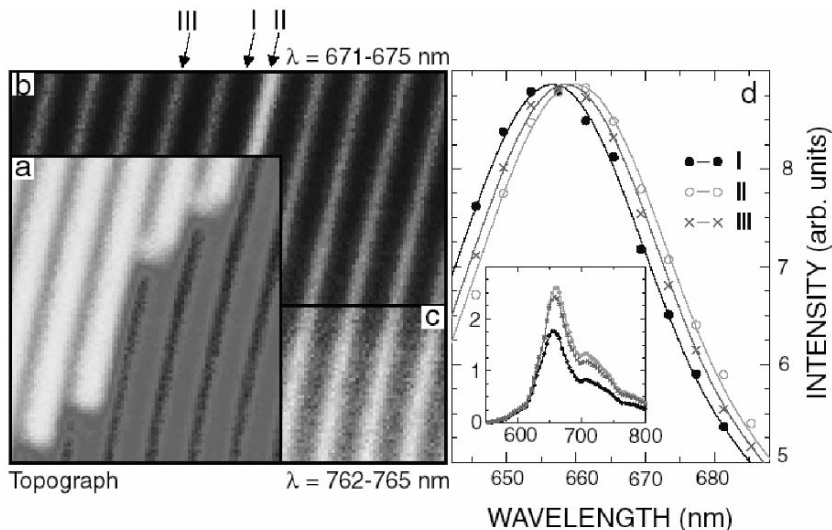


Figure 17–41. Atomic scale detail in photon maps: (a) Constant current image of a reconstructed Au(110) surface showing two terraces separated by a monoatomic step. Marker I: on top of an atomic row. Marker II: below a monoatomic step. Marker III: between atomic rows. (b and c) Simultaneously acquired photon map is in registry with the topographic data. (b) Photon wavelength $\lambda = 671\text{--}675 \text{ nm}$. (c) Photon wavelength $\lambda = 762\text{--}765 \text{ nm}$. (d) Representative spectra normalized to the same peak intensity. Inset: Unnormalized spectra. (From Hoffmann et al., 2004.)

increased by 50% at Au(110) steps. In the new study of atomic resolution spectroscopic imaging on Au(110) mentioned above (Hoffmann et al., 2004), a new model was presented that is consistent with the existing data. Atomic resolution on Au(110) and the contrast at steps was attributed to the local electronic structure and its effect on the elastic and inelastic tunneling channels (Hoffmann et al., 2004).

The sensitivity of STM-induced photon emission from metals to the conditions of the tunneling gap has been discussed above. New phenomena are observed when molecules are placed in this cavity. In spatial photon maps of close packed C₆₀ monolayers on Au(110), individual molecules have been resolved as distinct maxima (Berndt et al., 1994). This could not be explained by photon emission from tip-induced plasmon modes alone, since on metals the emission intensity would be reduced if the tip-metal distance is increased. Therefore if the emission from the C₆₀-covered metal surface was directly due to plasmon modes, the increase in tip-metal distance due to the molecules should weaken the plasmon modes. But actually the opposite is observed. The role of the molecules in the emission process is unclear, although molecular fluorescence has been suggested.

Spatial photon maps and fluorescence spectra from single hexa-*tert*-butyl-decacyclene (HBDC) molecules on Au(111), Ag(111), and Cu(111) have been recorded at 4K (Hoffmann et al., 2002b). Low temperature was essential for this measurement, since these molecules are mobile on the surface at room temperature. Results similar to C₆₀ are reported, i.e., increased emission on the molecule, but no new spectral features attributable to molecular fluorescence were observed. Only a 4-nm shift of the HBDC photon spectrum toward shorter wavelengths was observed compared to the substrate spectrum.

On a metal surface, the electronic levels of a molecule are considerably broadened whereas light emission is strongly quenched (Barnes, 1998), making it difficult to detect molecule-specific emission. In an LT-STM study it has been shown that molecular fluorescence could be observed when the molecule (ZnEtiol) was supported on a thin aluminum oxide (Al₂O₃) film grown on an NiAl(110) surface (Qiu et al., 2003). The oxide spacer reduces the interaction between the molecule and the metal. The photon emission spectra and intensities varied with different tips and increased photon emission efficiency was observed with Ag tips compared to W tips. This has been attributed to plasmon excitation in the tunneling gap. To reduce the plasmon signature in the spectra the tips were voltage pulsed. After that, photon emission spectra with sharp features could be seen when the tip was positioned directly above a molecule. Furthermore, the spectra were very sensitive to the tip position above the molecule. Light emission has been found to be almost identical for molecules of the same conformation. This is shown in photon spectra from three different molecules (in the same conformation) measured with three different STM tips (Figure 17-42A). The spectra could be explained as a superposition of tip-induced plasmon emission (which depend on tunneling gap conditions) and molecular fluorescence (which should be nearly independent of gap conditions). Molecular fluorescence is due to electrons tunneling elasti-

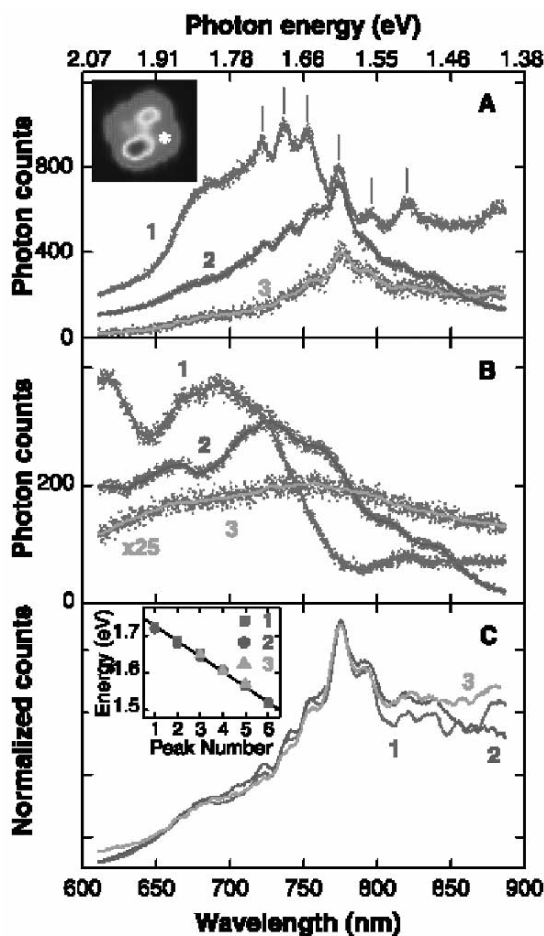


Figure 17-42. Molecular fluorescence from the excitation of vibrational modes in a single molecule. (A) Photon emission spectra for three different experiments with three different tips. The spectra were obtained over a single ZnEtIol molecule shown in the inset (constant current image) at the marked spot. The raw data are plotted together with the smoothed curves. Spectra 1 and 2 were taken with two different Ag tips and have been offset for clarity. Spectrum 3 was obtained with a W tip. The differences in the spectra are due to different tip plasmon properties. (B) NiAl light emission spectra measured with the same tips and voltages as in (A). (C) Molecular spectra from (A) divided by the corresponding NiAl spectra from (B) show remarkable similarity. The inset shows the photon energy of each peak determined for the three spectra. (From Qiu et al., 2003.)

cally into the molecule and creating excited vibrational states that make a radiative transition. To eliminate the influence of plasmons from the spectra, they have been divided by the spectra taken from the NiAl surface (Figure 17-42B) with the same tips. The resulting curves (Figure 17-42C) show remarkable similarity, indicating that the molecules have nearly identical light emission properties independent of the tip. The vibrational features were equidistant in energy with spacing of 40 ± 2 meV (Figure 17-42, inset).

A similar picture for the emission process has been used to explain the photon emission measured with an LT-STM from quantum well structures formed by Na layers on a Cu(111) surface (Hoffmann et al., 2001). The photon emission in this experiment was assigned to transitions between the electronic levels of the quantum structure and was observed together with tip-induced plasmon emission.

Recently another experiment with a room temperature STM has shown molecular fluorescence from organic molecules decoupled from an Au(100) surface by several adsorbed layers of the same molecule (Dong et al., 2004). Molecular fluorescence peaks in the spectra became sharper with increased coverage while the plasmon-related band was suppressed. Due to the higher thermal drift at room temperature, the spectra in this observation were averaged over several molecules.

Weakness of interaction between a molecule and the substrate turns out to be essential for the observation of STM-induced molecular fluorescence. STM-induced fluorescence combined with imaging and STS can be used to probe the interdependence between conformational structure, energy levels, and optical properties of single molecules. The inherent stability of the LT-STM together with the suppressed molecular diffusion at low temperatures should enable the study of electron dynamics in organic molecules with submolecular resolution.

3.5 Spin-Polarized Scanning Tunneling Microscopy

STM can yield information about magnetic properties by use of spin-polarized tunneling between a magnetic tip and substrate. The spin valve effect predicts that the tunneling current depends on the relative orientation of the magnetic moments of the tunneling electrodes (Julliere, 1975; Slonczewski, 1988). A magnetic tip acts as a source of spin-polarized electrons, probing the spin-split density of states of the magnetic sample. This technique allows imaging with atomic spatial resolution and, like conventional STM, is mostly sensitive to the topmost atomic layer. The ability to probe topography, crystallography, and magnetism at the same time renders spin-polarized scanning tunneling microscopy (SP-STM) a very powerful tool for the investigation of magnetic surfaces and monolayers. Since many magnetic phenomena exist only below a certain critical temperature, it is advantageous to use a low- or variable-temperature STM to observe magnetic materials. In addition, measurements on low-dimensional magnetic systems like ultrathin films or superparamagnetic particles demand low temperatures, since the Curie temperature T_c in general scales with dimension. In addition to those advantages of low temperatures specific to SP-STM, all other advantages mentioned in the introduction apply.

Wiesendanger et al. (1990) explored SP-STM using CrO_2 tips on an antiferromagnetic Cr(001) surface. These early experiments represented a mixture between topographic and spin-dependent contrast. Discriminating between magnetically caused contrast and contrast from other features of the electronic density of states near the Fermi level requires images taken with different tips (magnetic and nonmagnetic). An experimental setup for an LT-STM, which uses an external

magnetic field combined with the ability to rotate the sample without changing the tip position, has been described (Wittneven et al., 1997). This setup enables the determination of the relative magnetic orientation between the end of the tip and the sample.

Other magnetic imaging techniques and their respective lateral resolution are magneto-optical Kerr effect (MOKE) microscopy (~ 300 nm), MOKE with scanning near-field optical microscopy (SNOM) (~ 50 nm), scanning electron microscopy with polarization analysis (SEMPA) (~ 40 nm), magnetic force microscopy (MFM) (~ 20 nm), X-ray magnetic linear dichroism photoelectron emission microscopy (XMLD-PEEM) (~ 20 nm), off-axis electron holography, and Lorentz microscopy for thin films (~ 20 nm). SP-STM has proven to be capable of atomic resolution, which has been shown for the first time on a manganese monolayer on W(110) at 16 K (Heinze et al., 2000). A two-dimensional antiferromagnetic structure as predicted by theory with a periodicity of $4.5 \text{ \AA}$ has been observed as shown in Figure 17–43. This is a considerable advance, considering that previous characterizations of antiferromagnetic domains could not go beyond micrometer resolution. Spectroscopic studies with SP-STM allow the correlation of structural, local electronic, and local magnetic properties down to the atomic level. This has been demonstrated in low-temperature studies of ferromagnetic rare earth (Bode et al., 1998) and transition metal (Pietzsch et al., 2000a) systems as well as antiferromagnets (Kleiber et al., 2000). The tips used in these experiments are nonmagnetic tips coated with a thin layer of ferromagnetic (Kleiber et al., 2000) or antiferromagnetic (Kubetzka et al., 2002) material. Fe-coated tips are magnetized perpendicular to the tip axis at the apex and are therefore sensitive to the in-plane component of the sample's magnetization (Bode et al., 1998, Kleiber et al., 2000). FeGd-coated tips exhibit a perpendicular magnetic anisotropy and are therefore sensitive to the out-of-plane component of the sample's magnetization (Kubetzka et al., 2002).

If the LT-STM is built completely from nonmagnetic materials, SP-STM can be performed in the presence of high magnetic fields, which enables observation of hysteresis on a nanometer scale (Pietzsch et al., 2001). The internal spin structure of magnetic vortex states on three-dimensional Fe islands on W(110) has been resolved recently with an LT-STM, as shown in Figure 17–44. A Cr-coated tip was used to map out both the curling in plane magnetization around the vortex core as well as the perpendicular magnetization within the vortex core (see Figure 17–45). The width of the core was determined as 9 ± 1 nm in agreement with theory.

The extreme surface sensitivity of the SP-STM (as of all types of STM) may be its only weakness with regard to the investigation of coupling phenomena. Possible changes in the magnetic configuration of buried layers therefore are hidden from the measurement. These can be analyzed with XMLD-PEEM (Scholl, 2003), which, because of its elemental specificity and relatively long probing depth (3–5 nm), is able to investigate layered systems at more modest spatial resolution. However, SP-STM is unrivaled for the investigation of the magnetic structure of magnetic monolayers, surfaces, or surface alloys at atomic resolution.

Figure 17–43. Atomic resolution magnetic imaging with SP-STM. (A) Constant-current image of one monolayer of Mn on W(110) recorded with a non-magnetic W tip at 16 K. (B) Image recorded with a magnetic Fe tip showing an antiferromagnetic configuration as predicted by theory. The colored insets show calculated STM images. (C) Experimental and theoretical line sections from (A) and (B). The image size is 2.7 nm by 2.2 nm. (From Heinze et al., 2000.) (See color plate.)

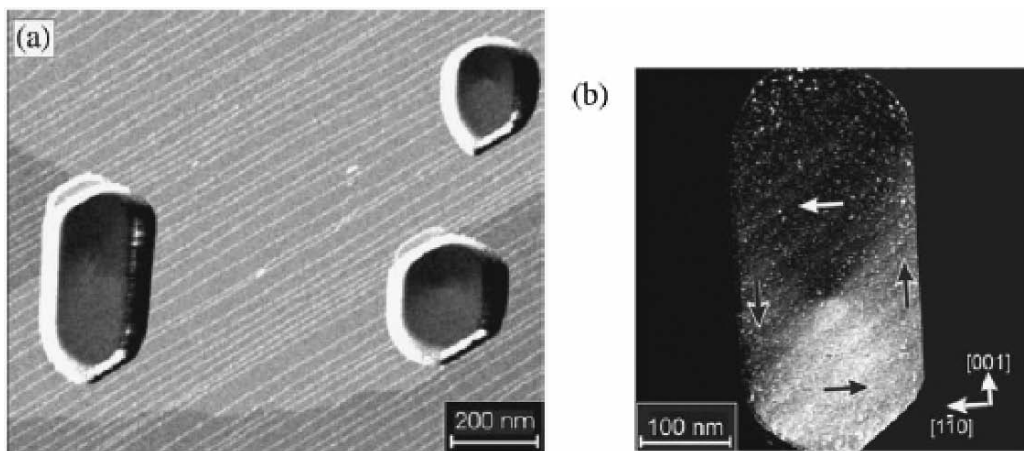
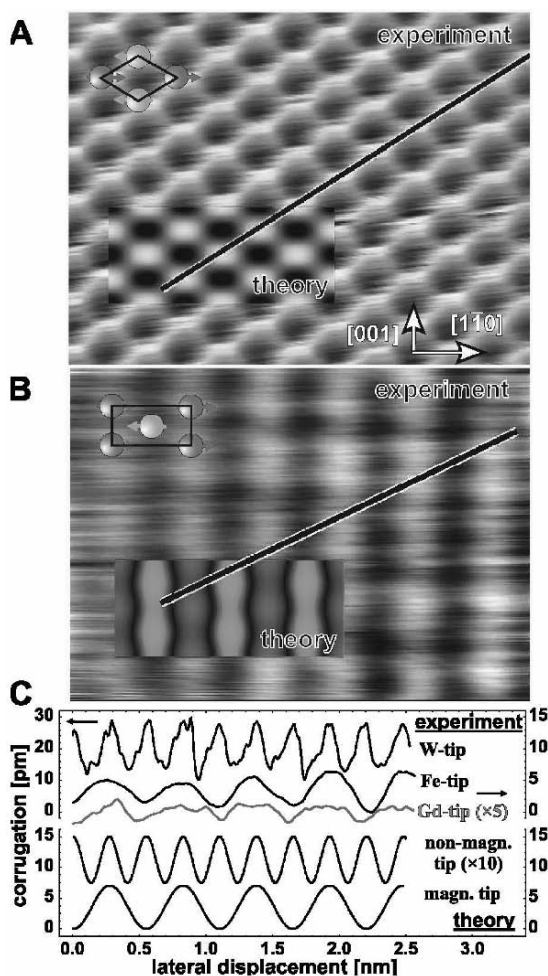


Figure 17–44. Magnetic vortex states on Fe islands imaged with SP-STM. (a) Spin-resolved dI/dV map of seven monolayers Fe on W(110) measured at 16 K with a Cr-coated tip being sensitive to the in-plane component of the sample magnetization. The islands exhibit a magnetic vortex state, as visible in (b). (From Wiesendanger et al., 2004.)

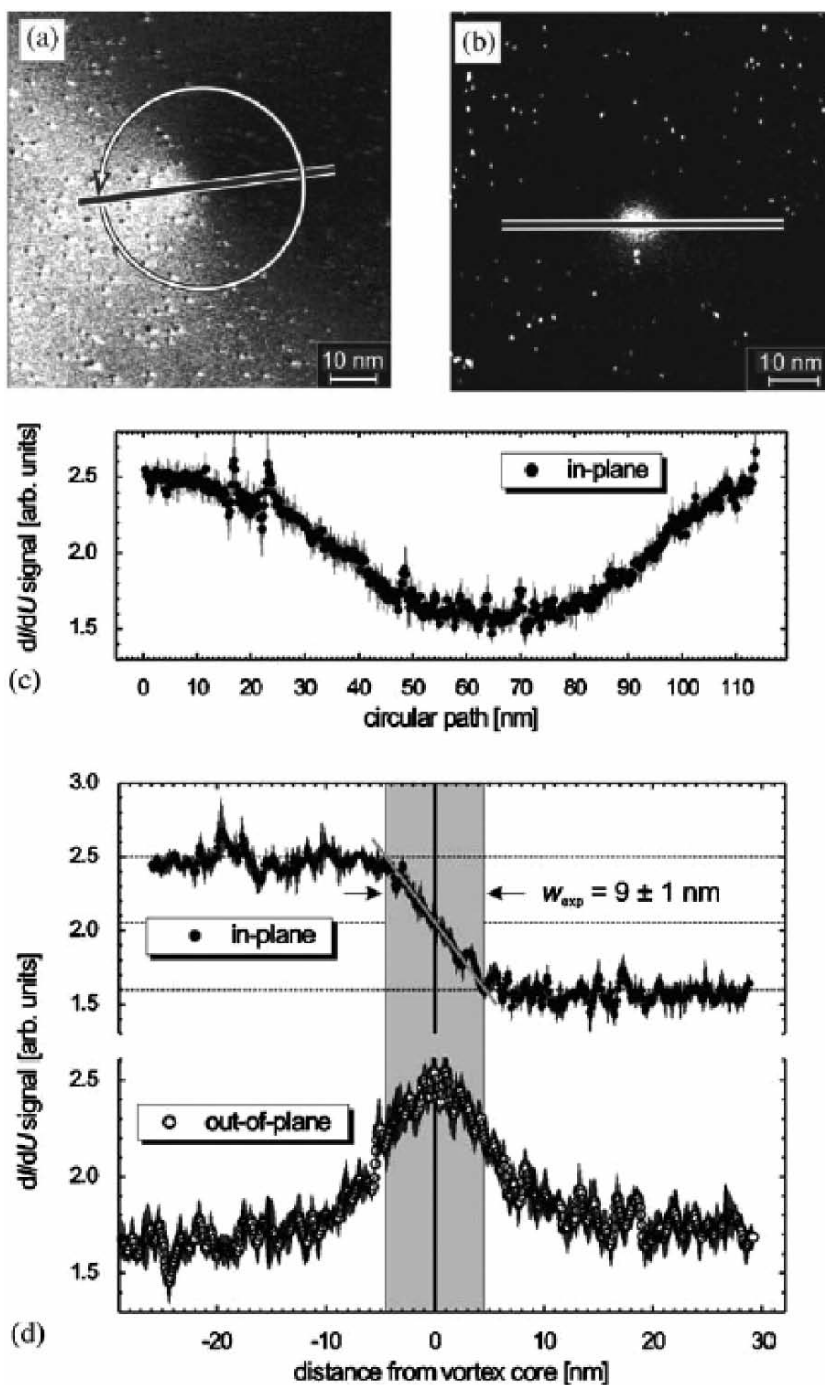


Figure 17-45. Magnetic vortex states on Fe islands at imaged higher magnification. Spin resolved dI/dV maps measured with an (a) in-plane and (b) out-of-plane sensitive Cr-coated tip. The curling in plane magnetization around the vortex core is visible in (a), whereas the magnetization component in the direction of the surface normal is seen in (b) as a bright area. (c) Spin-resolved dI/dV signal measured along the circle in (a) indicates that the in-plane component continuously curls around the core. (d) Spin-resolved dI/dV signal measured along the lines in (a) and (b) showing a vortex core width of 9 ± 1 nm. (From Wiesendanger et al., 2004.)

3.6 Superconductivity

The technique of tunneling was initially used as a tool to understand the bulk properties of superconductors. Electron tunneling through thin insulating films into superconductors is used to study how the density of states in a superconductor is modified. This technique averages over a macroscopic area and requires a high-quality insulating tunneling barrier. By using an STM for these measurements, a controllable vacuum barrier is used instead and superconductors can be studied, which do not support the formation of good, high-quality insulating films. Moreover, tunneling takes place over only a small area and variations in the density of states can be studied on an atomic scale.

According to BCS theory, the superconducting state is a new ground state where electrons of opposite momentum and spin form pairs. Any excitation of this ground state requires a Cooper pair to be broken into two single particle excitations. A minimum energy of Δ , the energy gap, is required to create one of these quasiparticles. If an electron tunnels from a metallic STM tip into the superconductor, it will enter into one of these quasiparticle states. Therefore the tunneling current provides a direct measure of the density of these states. Superconductors expel a magnetic field up to a certain critical field (Meissner effect). For larger fields, superconductivity is destroyed. In type II superconductors a mixed phase exists for certain magnetic field strengths, consisting of hexagonal lattices of normal conducting vortices (flux lines) in the superconductor, the so-called Abrikosov vortex lattice. Many experiments have confirmed the existence of these flux lines, e.g., electron microscopy (Bosch et al., 1985) and electron holography (Matsuda et al., 1989; Harada et al., 1992; Bonevich et al., 1993). These techniques are sensitive to the magnetic field variations. STM, in contrast, is sensitive to the electronic structure. The first LT-STM observation of flux lines and the flux lattice has been reported by Hess (Hess, 1991; Hess et al., 1989, 1990a,b, 1991). Surfaces of NbSe₂ have been studied in a ultralow temperature STM in a temperature range from 50 mK to 7 K. Differential conductance spectra in zero applied field showed the development of the superconducting energy gap below 7.2 K, which opens up to about 1.1 meV at the lowest temperatures. A second discontinuity was observed in the density of states at ± 34 mV. This feature resulted from the charge density wave (CDW) gap, since this material also supports a CDW state. Differential conductance spectra taken at 0 mV from a single vortex core at an applied field of 500 G showed the normal conducting vortex as a star pattern, which was explained as resulting from the 6-fold anisotropy of the atomic crystalline band structure. At 0 mV no state should exist in the superconductor, since this is the middle of the superconducting gap. Therefore the normal conducting vortex core was imaged bright on a dark background. At higher fields, the vortices organize into a hexagonal lattice. Figure 17-46 shows this lattice at a field of 10 kG. The dI/dV map at 1.3 mV was recorded under constant current feedback conditions with 1.3 mV bias. The question arises whether the localized current from the STM tip is sufficiently large to cause local perturbations such as heating, exceeding the local critical

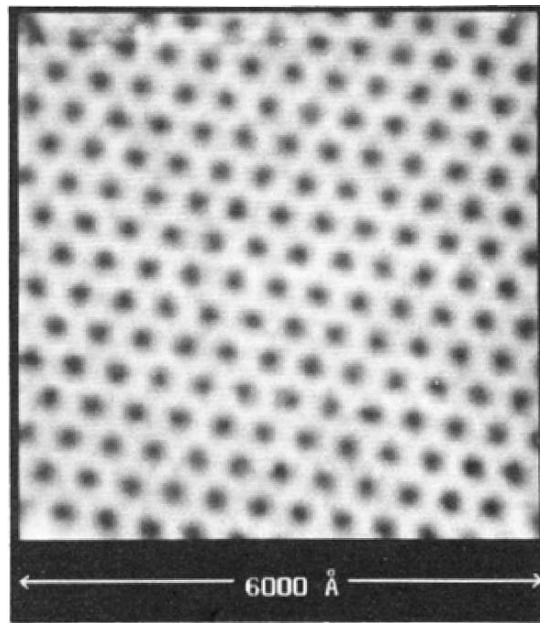


Figure 17-46. Abrikosov flux lattice imaged by LT-STM. Differential conductance map at 1.3 mV (above the superconducting gap) recorded on NbSe₂ at 1.8 K and 10 kG magnetic field. An enhanced density of states is observed away from the core and the density of states decreases at the vortex center. (From Hess et al., 1989. Reproduced by permission of Bell Labs, Lucent Technologies Inc.)

current or creating other nonequilibrium effects in the superconductor. No such current-dependent effects could be observed.

LT-STM provides a novel probe of inhomogeneous superconducting structures, of which the normal metal-to-superconductor planar interface is the most basic configuration. If the electrical contact between the two metals is good, superconductivity is weakened in the superconductor and induced in the normal metal. The phenomenon is known as the proximity effect. LT-STM has been used as a local probe of the superconducting proximity effect across a normal metal–superconductor interface of a short coherence length superconductor (Tessmer et al., 1996). Both the topography and the local electronic density of states were measured at 1.6 K on a superconducting NbSe₂ crystal decorated with nanometer-sized Au islands. A quasiparticle bound state was observed not only in the Au islands, but even when tunneling directly into the NbSe₂, which is evidence for a significant reduction of the superconductivity inside the NbSe₂ induced by the proximity of the Au overlayer. Local tunneling spectroscopy for an Nb/In/As/Nb superconducting proximity system was demonstrated with a low-temperature scanning tunneling microscope at 4.2 K (Inoue and Takayanagi, 1991). It was found that the local electron density of states in the InAs region is spatially modulated by the neighboring superconductor.

The local effects of individual Zn impurity atoms in the high-temperature superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (BSCCO) have been studied with an LT-STM at 4.2 K (Figure 17-47) (Pan et al., 2000). Recording a dI/dV map at zero bias, i.e., in the superconducting gap, revealed the Zn impurity scattering centers as bright dots. This is an image of the quasiparticle density of states in the vicinity of the impurity atoms. Imaging of the spatial dependence of the quasiparticle density of states in the vicinity of the impurity atoms reveals a four-fold symmetric quasiparticle “cloud” aligned with the nodes of the d -wave supercon-

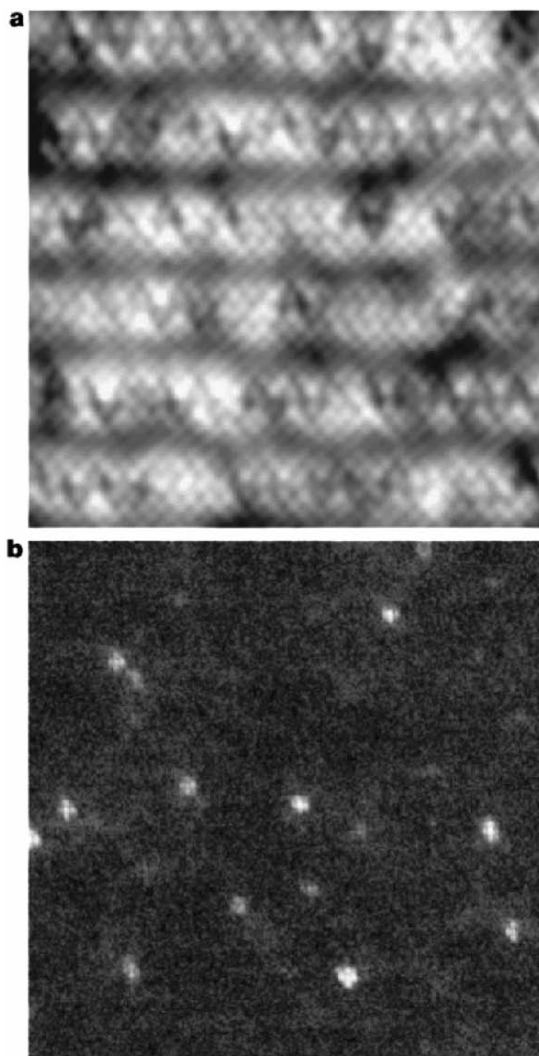


Figure 17-47. Constant current image (a) and zero-bias differential conductance map (b) of a cleaved single crystal of BSCCO at 4 K. (a) The surface BiO plane is exposed after cleavage. The atoms form a supermodulation along the crystal b -axis. (b) A map of the DOS at the Fermi energy ($V = 0$) Most of the image is dark, which is due to a low quasiparticle DOS at the Fermi level as expected for a superconductor far below T_c . The Zn impurities show up as bright features $\sim 15 \text{ \AA}$ in extent, each exhibiting a cross-shaped structure. (From Pan et al., 2000.)

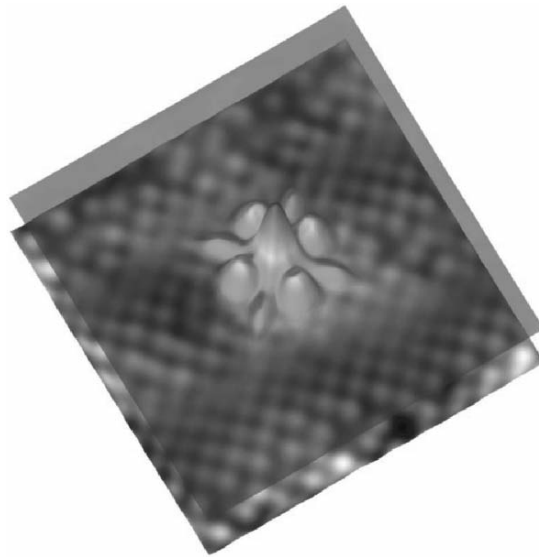


Figure 17–48. Quasiparticle density of states around a Zn impurity on the high T_c superconductor BSCCO: Relationship between the position of the Bi atoms on the crystal surface (lower layer), and the resonant DOS structure at the Zn atom (upper layer). Both layers were simultaneously acquired; the lower layer is a 6×6 -nm high-spatial-resolution image of the topography of the exposed BiO plane and the upper layer is the differential conductance at -1.5 mV bias voltage. The bright center of the scattering resonance coincides with the position of a Bi atom on the exposed surface BiO plane. The Zn scattering center is sitting on the CuO_2 plane two layers below. The inner bright cross in the quasiparticle DOS around the impurity (green bumps around the center) is aligned with the nodes of the d -wave superconducting gap. The weaker outer features, oriented at 45° to the inner cross, are aligned with the gap maxima. (From Pan et al., 2000.)

ducting gap, which is believed to characterize superconductivity in these materials (Figure 17–48).

References

- Abraham, D.L., Veider, A., Schonenberger, C., Meier, H.P., Arent, D.J. and Alvarado, S.F. (1990). *App. Phys. Lett.* **56**, 1564–1566.
- Aizpurua, J., Apell, S.P. and Berndt, R. (2000). *Phys. Rev. B*, **62**, 2065–2073.
- Arafune, R., Sakamoto, K., Meguro, K., Satoh, M., Arai, A. and Ushioda, S. (2001). *JPN J. App. Phys. Part 1* **40**, 5450–5453.
- Barnes, W.L. (1998). *J. Mod. Opt.* **45**, 661–699.
- Bartels, L., Meyer, G. and Rieder, K.H. (1997a). *Phys. Rev. Lett.* **79**, 697–700.
- Bartels, L., Meyer, G. and Rieder, K.H. (1997b). *App. Phys. Lett.* **71**, 213–215.
- Bartels, L., Wolf, M., Klamroth, T., Saalfrank, P., Kuhnle, A., Meyer, G. and Rieder, K.H. (1999). *Chem. Phys. Lett.* **313**, 544–552.
- Becker, T., Hovel, H., Tschudy, M. and Reihl, B. (1998). *App. Phys. a—Mat. Sci. Proc.* **66**, S27–S30.
- Behler, S., Rose, M.K., Dunphy, J.C., Ogletree, D.F., Salmeron, M. and Chape-lier, C. (1997). *Rev. Sci. Inst.* **68**, 2479–2485.
- Berndt, R. and Gimzewski, J.K. (1993). *Ann. Phys.* **2**, 133–140.

- Berndt, R., Gimzewski, J.K. and Johansson, P. (1991a). *Phys. Rev. Lett.* **67**, 3796–3799.
- Berndt, R., Schlittler, R.R. and Gimzewski, J.K. (1991b). *J. Va. Sci. Tech. B* **9**, 573–577.
- Berndt, R., Gimzewski, J.K. and Johansson, P. (1993). *Phys. Rev. Lett.* **71**, 3493–3496.
- Berndt, R., Gaisch, R., Schneider, W.D., Gimzewski, J.K., Reihl, B., Schlittler, R.R. and Tschudy, M. (1994). *Surface Sci.* **309**, 1033–1037.
- Berndt, R., Gaisch, R., Schneider, W.D., Gimzewski, J.K., Reihl, B., Schlittler, R.R. and Tschudy, M. (1995). *Phys. Rev. Lett.* **74**, 102–105.
- Bertel, E. (1997). *Phys. Status Sol. a—App. Res.* **159**, 235–242.
- Bertel, E., Roos, P. and Lehmann, J. (1995). *Phys. Rev. B* **52**, 14384–14387.
- Binnig, G. and Rohrer, H. (1982). *Helvetica Phys. Acta* **55**, 726–735.
- Binnig, G., Frank, K.H., Fuchs, H., Garcia, N., Reihl, B., Rohrer, H., Salvan, F. and Williams, A.R. (1985a). *Phys. Rev. Lett.* **55**, 991–994.
- Binnig, G., Garcia, N. and Rohrer, H. (1985b). *Phys. Rev. B* **32**, 1336–1338.
- Bode, M., Getzlaff, M. and Wiesendanger, R. (1998). *Phys. Rev. Lett.* **81**, 4256–4259.
- Bonevich, J.E., Harada, K., Matsuda, T., Kasai, H., Yoshida, T., Pozzi, G. and Tonomura, A. (1993). *Phys. Rev. Lett.* **70**, 2952–2955.
- Bosch, J., Gross, R., Koyanagi, M. and Huebener, R.P. (1985). *Phys. Rev. Lett.* **54**, 1448–1451.
- Bott, M., Michely, T. and Comsa, G. (1995). *Rev. Sci. Instr.* **66**, 4135–4139.
- Braun, K.-F. (2000). PhD Thesis. Freie Universitaet, Berlin.
- Braun, K.F. and Rieder, K.H. (2002). *Phys. Rev. Lett.* **88**, 096801.
- Burgi, L., Jeandupeux, O., Hirstein, A., Brune, H. and Kern, K. (1998). *Phys. Rev. Lett.* **81**, 5370–5373.
- Burgi, L., Jeandupeux, O., Brune, H. and Kern, K. (1999). *Phys. Rev. Lett.* **82**, 4516–4519.
- Burgi, L., Brune, H., Jeandupeux, O. and Kern, K. (2000a). *J. Electron Spectrosc. Rel. Phenomena* **109**, 33–49.
- Burgi, L., Petersen, L., Brune, H. and Kern, K. (2000b). *Surface Sci.* **447**, L157–L161.
- Coombs, J.H., Gimzewski, J.K., Reihl, B., Sass, J.K. and Schlittler, R.R. (1988). *J. Microsc.—Oxf.* **152**, 325–336.
- Crampin, S., Boon, M.H. and Inglesfield, J.E. (1994). *Phys. Rev. Lett.* **73**, 1015–1018.
- Crommie, M.F., Lutz, C.P. and Eigler, D.M. (1993a). *Science* **262**, 218–220.
- Crommie, M.F., Lutz, C.P. and Eigler, D.M. (1993b). *Nature* **363**, 524–527.
- Crommie, M.F., Lutz, C.P. and Eigler, D.M. (1993c). *Phys. Rev. B* **48**, 2851–2854.
- Datta, S. (1995). *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge).
- Davis, L.C., Everson, M.P., Jaklevic, R.C. and Shen, W.D. (1991). *Phys. Rev. B* **43**, 3821–3830.
- Dong, Z.C., Guo, X.L., Trifonov, A.S., Dorozhkin, P.S., Miki, K., Kimura, K., Yokoyama, S. and Mashiko, S. (2004). *Phys. Rev. Lett.* **92**, 086801.
- Eigler, D.M. and Schweizer, E.K. (1990). *Nature* **344**, 524–526.
- Eigler, D.M., Weiss, P.S., Schweizer, E.K. and Lang, N.D. (1990). *Abstr. Pap. Amer. Chem. Soc.* **200**, 85–PHYS.
- Eigler, D.M., Lutz, C.P. and Rudge, W.E. (1991). *Nature* **352**, 600–603.
- Eiguren, A., Hellsing, B., Reinert, F., Nicolay, G., Chulkov, E.V., Silkin, V.M., Hufner, S. and Echenique, P.M. (2002). *Phys. Rev. Lett.* **88**, 066805.
- Ferris, J.H., Kushmerick, J.G., Johnson, J.A., Youngquist, M.G.Y., Kessinger, R.B., Kingsbury, H.F. and Weiss, P.S. (1998). *Rev. Sci. Instr.* **69**, 2691–2695.

- Fiete, G.A., Hersch, J.S., Heller, E.J., Manoharan, H.C., Lutz, C.P. and Eigler, D.M. (2001). *Phys. Rev. Lett.* **86**, 2392–2395.
- Foley, E.T., Kam, A.F., Lyding, J.W. and Avouris, P. (1998). *Phys. Rev. Lett.* **80**, 1336–1339.
- Friedel, J. (1958). *Nuovo Cimento* **7**, 287.
- Gaisch, R., Gimzewski, J.K., Reihl, B., Schlittler, R.R., Tschudy, M. and Schneider, W.D. (1992). *Ultramicroscopy* **42**, 1621–1626.
- Garcia, N. and Serena, P.A. (1995). *Surface Sci.* **330**, L665–L667.
- Gartland, P.O. and Slagsvold, B.J. (1975). *Phys. Rev. B* **12**, 4047–4058.
- Gimzewski, J.K. and Joachim, C. (1999). *Science* **283**, 1683–1688.
- Grill, L., Moresco, F., Jiang, P., Joachim, C., Gourdon, A. and Rieder, K.H. (2004). *Phys. Rev. B* **69**, 035416.
- Gross, L., Moresco, F., Savio, L., Gourdon, A., Joachim, C. and Rieder, K.H. (2004). *Phys. Rev. Lett.* **93**.
- Gustafsson, A., Pistol, M.E., Montelius, L. and Samuelson, L. (1998). *J. App. Phys.* **84**, 1715–1775.
- Hahn, J.R. and Ho, W. (2001). *Phys. Rev. Lett.* **87**19, art. no. 196102.
- Hahn, J.R., Lee, H.J. and Ho, W. (2000). *Phys. Rev. Lett.* **85**, 1914–1917.
- Hamers, R.J. (1989). *Ann. Rev. Phys. Chem.* **40**, 531–559.
- Hansma, P. (1982). *Tunneling Spectroscopy: Capabilities, Applications and New Techniques* (Plenum Press, New York).
- Harada, K., Matsuda, T., Bonevich, J., Igarashi, M., Kondo, S., Pozzi, G., Kawabe, U. and Tonomura, A. (1992). *Nature* **360**, 51–53.
- Harrell, L.E. and First, P.N. (1999). *Rev. Sci. Instr.* **70**, 125–132.
- Heimann, P., Neddermeyer, H. and Roloff, H.F. (1977). *J. Phys. C-Sol. State Phys.* **10**, L17–L22.
- Heinze, S., Bode, M., Kubetzka, A., Pietzsch, O., Nie, X., Blugel, S. and Wiesendanger, R. (2000). *Science* **288**, 1805–1808.
- Heller, E.J., Crommie, M.F., Lutz, C.P. and Eigler, D.M. (1994). *Nature* **369**, 464–466.
- Hess, H.F. (1991). *Physica C* **185**, 259–263.
- Hess, H.F., Robinson, R.B., Dynes, R.C., Valles, J.M. and Waszczak, J.V. (1989). *Phys. Rev. Lett.* **62**, 214–216.
- Hess, H.F., Robinson, R.B., Dynes, R.C., Valles, J.M. and Waszczak, J.V. (1990a). *J. Vac. Sci. Tech. a—Vac. Surfaces Films* **8**, 450–454.
- Hess, H.F., Robinson, R.B. and Waszczak, J.V. (1990b). *Phys. Rev. Lett.* **64**, 2711–2714.
- Hess, H.F., Robinson, R.B. and Waszczak, J.V. (1991). *Physica B* **169**, 422–431.
- Hewson, A.C. (1993). *The Kondo Problem to Heavy Fermions* (Cambridge University Press, Cambridge).
- Himpsel, F.J., Ortega, J.E., Mankey, G.J. and Willis, R.F. (1998). *Adv. Phys.* **47**, 511–597.
- Hoffmann, G., Kliewer, J. and Berndt, R. (2007). *Phys. Rev. Lett.* **87**, 176803.
- Hoffmann, G., Kroger, J. and Berndt, R. (2002a). *Rev. Sci. Instr.* **73**, 305–309.
- Hoffmann, G., Libioulle, L. and Berndt, R. (2002b). *Phys. Rev. B* **65**, 212107.
- Hoffmann, G., Maroutian, T. and Berndt, R. (2004). *Phys. Rev. Lett.* **93**, 076102.
- Hofmann, P., Briner, B.G., Doering, M., Rust, H.P., Plummer, E.W. and Bradshaw, A.M. (1997). *Phys. Rev. Lett.* **79**, 265–268.
- Horch, S., Zeppenfeld, P., David, R. and Comsa, G. (1994). *Rev. Sci. Instr.* **65**, 3204–3210.
- Hormandinger, G. (1994). *Phys. Rev. Lett.* **73**, 910.
- Inoue, K. and Takayanagi, H. (1991). *Phys. Rev. B* **43**, 6214–6215.
- Jaklevic, R.C. and Lambe, J. (1966). *Phys. Rev. Lett.* **17**, 1139.

- Jeandupeux, O., Burgi, L., Hirstein, A., Brune, H. and Kern, K. (1999). *Phys. Rev. B* **59**, 15926–15934.
- Joachim, C., Gimzewski, J.K. and Aviram, A. (2000). *Nature* **408**, 541–548.
- Julliere, M. (1975). *Phys. Lett. A* **54**, 225–226.
- Jung, T.A., Schlittler, R.R., Gimzewski, J.K., Tang, H. and Joachim, C. (1996). *Science* **271**, 181–184.
- Jung, T.A., Schlittler, R.R. and Gimzewski, J.K. (1997). *Nature* **386**, 696–698.
- Kleiber, M., Bode, M., Ravlic, R. and Wiesendanger, R. (2000). *Phys. Rev. Lett.* **85**, 4606–4609.
- Kliwer, I., Berndt, R., Chulkov, E.V., Silkin, V.M., Echenique, P.M. and Crampin, S. (2000). *Science* **288**, 1399–1402.
- Knorr, N., Schneider, M.A., Diekhoner, L., Wahl, P. and Kern, K. (2002). *Phys. Rev. Lett.* **88**, 096804.
- Kroger, J., Limot, L., Jensen, H., Berndt, R. and Johansson, P. (2004). *Phys. Rev. B* **70**, 033401.
- Kubetzka, A., Bode, M., Pietzsch, O. and Wiesendanger, R. (2002). *Phys. Rev. Lett.* **88**, 057201.
- Kugler, M., Renner, C., Fischer, O., Mikheev, V. and Batey, G. (2000). *Rev. Sci. Inst.* **71**, 1475–1478.
- Kuntze, J., Berndt, R., Jiang, P., Tang, H., Gourdon, A. and Joachim, C. (2002). *Phys. Rev. B* **65**, 233405.
- Langlais, V.J., Schlittler, R.R., Tang, H., Gourdon, A., Joachim, C. and Gimzewski, J.K. (1999). *Phys. Rev. Lett.* **83**, 2809–2812.
- Lau, K.H. and Kohn, W. (1978). *Surface Sci.* **75**, 69–85.
- Lauhon, L.J. and Ho, W. (2001). *Rev. Sci. Inst.* **72**, 216–223.
- Lee, H.J. and Ho, W. (1999). *Science* **286**, 1719–1722.
- Lemay, S.G., Janssen, J.W., van den Hout, M., Mooij, M., Bronikowski, M.J., Willis, P.A., Smalley, R.E., Kouwenhoven, L.P. and Dekker, C. (2001). *Nature* **412**, 617–620.
- Li, J.T., Schneider, W.D. and Berndt, R. (1997). *Phys. Rev. B* **56**, 7656–7659.
- Li, J.T., Schneider, W.D., Berndt, R., Bryant, O.R. and Crampin, S. (1998a). *Phys. Rev. Lett.* **81**, 4464–4467.
- Li, J.T., Schneider, W.D., Berndt, R. and Crampin, S. (1998b). *Phys. Rev. Lett.* **80**, 3332–3335.
- Li, J.T., Schneider, W.D., Berndt, R. and Delley, B. (1998c). *Phys. Rev. Lett.* **80**, 2893–2896.
- Lorente, N. (2004). *App. Phys. a—Mat. Sci. Proc.* **78**, 799–806.
- Lorente, N. and Persson, M. (2000). *Phys. Rev. Lett.* **85**, 2997–3000.
- Lorente, N., Persson, M., Lauhon, L.J. and Ho, W. (2001). *Phys. Rev. Lett.* **86**, 2593–2596.
- Madhavan, V., Chen, W., Jamneala, T., Crommie, M.F. and Wingreen, N.S. (1998). *Science* **280**, 567–569.
- Makoshi, K. and Mingo, N. (2002). *Surface Sci.* **502**, 34–40.
- Maltezopoulos, T., Kubetzka, A., Morgenstern, M., Wiesendanger, R., Lemay, S.G. and Dekker, C. (2003). *App. Phys. Lett.* **83**, 1011–1013.
- Manoharan, H.C., Lutz, C.P. and Eigler, D.M. (2000). *Nature* **403**, 512–515.
- Matsuda, T., Hasegawa, S., Igarashi, M., Kobayashi, T., Naito, M., Kajiyama, H., Endo, J., Osakabe, N., Tonomura, A. and Aoki, R. (1989). *Phys. Rev. Lett.* **62**, 2519–2522.
- Matsui, T., Kambara, H., Ueda, I., Shishido, T., Miyatake, Y. and Fukuyama, H. (2003). *Phys. B—Cond. Matt.* **329**, 1653–1655.
- Matzdorf, R. (1998). *Surface Sci. Rep.* **30**, 153–206.
- Mommel, N. (1998). *Surface Sci. Rep.* **32**, 93–163.
- Mommel, N. and Bertel, E. (1995). *Phys. Rev. Lett.* **75**, 485–488.
- Meyer, G. (1996). *Rev. Sci. Inst.* **67**, 2960–2965.

- Meyer, G., Bartels, L., Zophel, S., Henze, E. and Rieder, K.H. (1997). *Phys. Rev. Lett.* **78**, 1512–1515.
- Meyer, G., Zophel, S. and Rieder, K.H. (1996). *App. Phys. Lett.* **69**, 3185–3187.
- Meyer, G., Moresco, F., Hla, S.W., Repp, J., Braun, K.F., Folsch, S. and Rieder, K.H. (2001). *J. J. App. Phys. Part 1* **40**, 4409–4413.
- Mingo, N. and Makoshi, K. (2000). *Phys. Rev. Lett.* **84**, 3694–3697.
- Montelius, L., Pistol, M.E. and Samuelson, L. (1992). *Ultramicroscopy* **42**, 210–214.
- Moresco, F., Meyer, G. and Rieder, K.H. (1999). *Mod. Phys. Lett. B* **13**, 709–715.
- Moresco, F., Meyer, G., Rieder, K.H., Tang, H., Gourdon, A. and Joachim, C. (2001). *App. Phys. Lett.* **78**, 306–308.
- Moresco, F., Gross, L., Alemani, M., Rieder, K.H., Tang, H., Gourdon, A. and Joachim, C. (2003). *Phys. Rev. Lett.* **91**, 036601.
- Mugele, F., Rettenberger, A., Boneberg, J. and Leiderer, P. (1998). *Rev. Sci. Inst.* **69**, 1765–1769.
- Murashita, T. (1997). *J. Vac. Sci. Tech. B* **15**, 32–37.
- Murashita, T. (1999). *J. Vac. Sci. Tech. B* **17**, 22–28.
- Murashita, T., Wada, K. and Inoue, N. (1993). *App. Surface Sci.* **68**, 223–226.
- Nazin, G.V., Qiu, X.H. and Ho, W. (2003a). *Phys. Rev. Lett.* **90**, 216110.
- Nazin, G.V., Qiu, X.H. and Ho, W. (2003b). *Science* **302**, 77–81.
- Nicolay, G., Reinert, F., Schmidt, S., Ehm, D., Steiner, P. and Hufner, S. (2000). *Phys. Rev. B* **62**, 1631–1634.
- Nilius, N., Ernst, N. and Freund, H.J. (2000). *Phys. Rev. Lett.* **84**, 3994–3997.
- Nishitani, R., Umeno, T. and Kasuya, A. (1998). *App. Phys. a—Mat. Sci. Proc.* **66**, S139–S143.
- Pan, S.H., Hudson, E.W. and Davis, J.C. (1999). *Rev. Sci. Inst.* **70**, 1459–1463.
- Pan, S.H., Hudson, E.W., Lang, K.M., Eisaki, H., Uchida, S. and Davis, J.C. (2000). *Nature* **403**, 746–750.
- Paniago, R., Matzdorf, R., Meister, G. and Goldmann, A. (1995). *Surface Sci.* **336**, 113–122.
- Pascual, J.I., Jackiw, J.J., Song, Z., Weiss, P.S., Conrad, H. and Rust, H.P. (2001a). *Phys. Rev. Lett.* **86**, 1050–1053.
- Pascual, J.I., Song, Z., Jackiw, J.J., Horn, K. and Rust, H.P. (2001b). *Phys. Rev. B*, **6324**, art. no. 241103.
- Pascual, J.I., Gomez-Herrero, J., Sanchez-Portal, D. and Rust, H.P. (2002). *J. Chem. Phys.* **117**, 9531–9534.
- Persson, B.N.J. and Baratoff, A. (1992). *Phys. Rev. Lett.* **68**, 3224–3227.
- Petersen, L., Sprunger, P.T., Hofmann, P., Laegsgaard, E., Briner, B.G., Doering, M., Rust, H.P., Bradshaw, A.M., Besenbacher, F. and Plummer, E.W. (1998). *Phys. Rev. B* **57**, R6858–R6861.
- Petersen, L., Hofmann, P., Plummer, E.W. and Besenbacher, F. (2000a). *J. Electron Spectrosc. Rel. Phenomena* **109**, 97–115.
- Petersen, L., Schaefer, B., Laegsgaard, E., Stensgaard, I. and Besenbacher, F. (2000b). *Surface Sci.* **457**, 319–325.
- Petersen, L., Schunack, M., Schaefer, B., Linderroth, T.R., Rasmussen, P.B., Sprunger, P.T., Laegsgaard, E., Stensgaard, I. and Besenbacher, F. (2001). *Rev. Sci. Inst.* **72**, 1438–1444.
- Pietzsch, O., Kubetzka, A., Bode, M. and Wiesendanger, R. (2000a). *Phys. Rev. Lett.* **84**, 5212–5215.
- Pietzsch, O., Kubetzka, A., Haude, D., Bode, M. and Wiesendanger, R. (2000b). *Rev. Sci. Inst.* **71**, 424–430.
- Pietzsch, O., Kubetzka, A., Bode, M. and Wiesendanger, R. (2001). *Science* **292**, 2053–2056.
- Pines, D. (1966). *The Theory of Quantum Liquids* (Benjamin, New York).

- Pivetta, M., Silly, F., Patthey, F., Pelz, J.P. and Schneider, W.D. (2003). *Phys. Rev. B* **67**, 193402.
- Qiu, X.H., Nazin, G.V. and Ho, W. (2003). *Science* **299**, 542–546.
- Ralls, K.S., Ralph, D.C. and Buhrman, R.A. (1989). *Phys. Rev. B* **40**, 11561–11570.
- Reihl, B., Coombs, J.H. and Gimzewski, J.K. (1989). *Surface Sci.* **211**, 156–164.
- Repp, J., Moresco, F., Meyer, G., Rieder, K.H., Hyldgaard, P. and Persson, M. (2000). *Phys. Rev. Lett.* **85**, 2981–2984.
- Rosei, F., Schunack, M., Jiang, P., Gourdon, A., Laegsgaard, E., Stensgaard, I., Joachim, C. and Besenbacher, F. (2002). *Science* **296**, 328–331.
- Rust, H.P., Buisset, J., Schweizer, E.K. and Cramer, L. (1997). *Rev. Sci. Instr.* **68**, 129–132.
- Rust, H.P., Doering, M., Pascual, J.I., Pearl, T.P. and Weiss, P.S. (2001). *Rev. Sci. Instr.* **72**, 4393–4397.
- Sanchez, O., Garcia, J.M., Segovia, P., Alvarez, J., Deparga, A.L.V., Ortega, J.E., Prietsch, M. and Miranda, R. (1995). *Phys. Rev. B* **52**, 7894–7897.
- Sautet, P. and Joachim, C. (1991). *Chem. Phys. Lett.* **185**, 23–30.
- Schneider, M.A., Vitali, L., Knorr, N. and Kern, K. (2002). *Phys. Rev. B* **65**, 121406.
- Scholl, A. (2003). *Curr. Opinion Sol. State Mat. Sci.* **7**, 59–66.
- Shockley, W. (1939). *Phys. Rev.* **56**, 317–323.
- Slonczewski, J.C. (1988). *J. Phys.* **49**, 1629–1630.
- Smith, D.P.E., Kirk, M.D. and Quate, C.F. (1987). *J. Chem. Phys.* **86**, 6034–6038.
- Smolyaninov, II, Khaikin, M.S. and Edelman, V.S. (1990). *Phys. Lett. A* **149**, 410–412.
- Sprunger, P.T., Petersen, L., Plummer, E.W., Laegsgaard, E. and Besenbacher, F. (1997). *Science* **275**, 1764–1767.
- Stipe, B.C., Rezaei, M.A. and Ho, W. (1997a). *Phys. Rev. Lett.* **79**, 4397–4400.
- Stipe, B.C., Rezaei, M.A., Ho, W., Gao, S., Persson, M. and Lundqvist, B.I. (1997b). *Phys. Rev. Lett.* **78**, 4410–4413.
- Stipe, B.C., Rezaei, M.A. and Ho, W. (1998). *Science* **280**, 1732–1735.
- Stipe, B.C., Rezaei, H.A. and Ho, W. (1999a). *Phys. Rev. Lett.* **82**, 1724–1727.
- Stipe, B.C., Rezaei, M.A. and Ho, W. (1999b). *Rev. Sci. Instr.* **70**, 137–143.
- Stranick, S.J., Kamna, M.M. and Weiss, P.S. (1994a). *Science* **266**, 99–102.
- Stranick, S.J., Kamna, M.M. and Weiss, P.S. (1994b). *Rev. Sci. Instr.* **65**, 3211–3215.
- Stroscio, J.A. and Eigler, D.M. (1991). *Science* **254**, 1319–1326.
- Stroscio, J.A., Celotta, R.J., Blankenship, S., Hudson, E. and Fein, A.P. (2000). Proceedings of the 4th International Workshop on Quantum Functional Devices, 145–146 (Kanazawa, Japan).
- Tersoff, J. and Hamann, D.R. (1985). *Phys. Rev. B* **31**, 805–813.
- Tessmer, S.H., Tarlie, M.B., Van Harlingen, D.J., Maslov, D.L. and Goldbart, P.M. (1996). *Phys. Rev. Lett.* **77**, 924–927.
- Theilmann, F., Matzdorf, R., Meister, G. and Goldmann, A. (1997). *Phys. Rev. B* **56**, 3632–3635.
- Uehara, Y., Fujita, T., Iwami, M. and Ushioda, S. (2001). *Rev. Sci. Instr.* **72**, 2097–2099.
- Vitali, L., Wahl, P., Schneider, M.A., Kern, K., Silkin, V.M., Chulkov, E.V. and Echenique, P.M. (2003). *Surface Sci.* **523**, L47–L52.
- Weierstall, U. and Spence, J.C.H. (1998). *Surface Sci.* **398**, 267–279.

- Whitman, L.J., Stroscio, J.A., Dragoset, R.A. and Celotta, R.J. (1991). *Science* **251**, 1206–1210.
- Wiesendanger, R. (1994). *Scanning Probe Microscopy and Spectroscopy*. (Cambridge University Press, Cambridge).
- Wiesendanger, R., Guntherodt, H.J., Guntherodt, G., Gambino, R.J. and Ruf, R. (1990). *Phys. Rev. Lett.* **65**, 247–250.
- Wiesendanger, R., Bode, M., Kubetzka, A., Pietzsch, O., Morgenstern, M., Wachowiak, A. and Wiebe, J. (2004). *J. Magn. Magn. Mat.* **272–76**, 2115–2120.
- Wittneven, C., Dombrowski, R., Pan, S.H. and Wiesendanger, R. (1997). *Rev. Sci. Instr.* **68**, 3806–3810.
- Wolkow, R.A. (1995). *Abstr. Papers Amer. Chem. Soc.* **209**, 381–PHYS.
- Xu, J.B., Lauger, K., Moller, R., Dransfeld, K. and Wilson, I.H. (1994). *J. App. Phys.* **76**, 7209–7216.
- Zangwill, A. (1988). *Physics at Surfaces* (Cambridge University Press, Cambridge).
- Zhang, H., Memmert, U., Houbertz, R. and Hartmann, U. (2001). *Rev. Sci. Instr.* **72**, 2613–2617.

Some recent additions to the literature (Chapter 19)

Additional developments prior to 2008 include the application of diffractive imaging to femtosecond soft-X-ray imaging (Chapman et al., 2006) and the application of a combined ptychography/iterative method in X-ray imaging, which, by removing the need for compact support, allows use of thin-film samples (Rodenburg et al., 2007). There has also been a renaissance of interest in Fourier Transform holography (for x-ray imaging), which also solves this problem to provide femtosecond, single-shot imaging (Marchesini et al., 2007). For divergent-beam in-line X-ray holography (the original Gabor arrangement) combined with iterative phasing to solve the twin-image problem and improve resolution, see Williams et al., (2006). A unified theoretical overview of all these new phasing methods can be found in Marchesini (2007); see also the review by Rodenburg (2008).

- Chapman, H. et al. (2006). Femtosecond diffractive imaging with a soft X-ray free electron laser, *Nature Physics* **2**, 839–843.
- Marchesini, S. (2007). A unified evaluation of iterative projection algorithms for phase retrieval, *Rev. Sci. Instrum.* **78**, 011301.
- Marchesini, S. et al. (2007). Ultrafast, ultrabright, high resolution Fourier–Hadamard transform x-ray holography. *Science*, in press.
- Rodenburg, J.M. (2008). Ptychography and related diffractive imaging methods. *Adv. Imaging & Electron Phys.* **150**.
- Rodenburg, J., Hurst, A.C., Collins, A.G., Dobson, B.K., Pfeiffer, F., Bunk, O., David, C., Jefimovs, K., and Johnson, I. (2007). Hard x-ray lensless imaging of extended objects, *Phys. Rev. Lett.* **98**, 034801.
- Williams, G.J., Quiney, H.M., Dhal, B.B., Nugent, K.A., Peele, A.G., Paterson, D. and de Jonge, M.D. (2006). Fresnel coherent diffractive imaging, *Phys. Rev. Lett.* **97**, 025506.

Part IV

HOLOGRAPHIC AND LENSLESS MODES

Electron Holography

Rafal E. Dunin-Borkowski, Takeshi Kasama,
Martha R. McCartney, and David J. Smith

1 Introduction

Electron holography, as originally described by Gabor (1949), is based on the formation of an interference pattern or “hologram” in the transmission electron microscope (TEM). In contrast to most conventional TEM techniques, which only record spatial distributions of image intensity, electron holography allows the phase shift of the high-energy electron wave that has passed through the specimen to be measured directly. The phase shift can then be used to provide information about local variations in magnetic induction and electrostatic potential. This chapter provides an overview of the technique of electron holography. It begins with an outline of the experimental procedures and theoretical background that are needed to obtain phase information from electron holograms. Medium-resolution applications of electron holography to the characterization of magnetic domain structures and electrostatic fields are then described, followed by a description of high-resolution electron holography and alternative modes of electron holography. The majority of the experimental results described below are obtained using the off-axis, or “sideband,” TEM mode, which is the most widely used mode of electron holography at present. For further details about electron holography, the interested reader is referred to several recent books (e.g., Tonomura et al., 1995; Tonomura, 1998; Völkl et al., 1998) and review papers (e.g., Tonomura, 1992; Midgley, 2001; Lichte, 2002; Matteucci et al., 2002).

1.1 Basis of Off-Axis Electron Holography

The off-axis mode of electron holography involves the examination of an electron-transparent specimen using defocused illumination from a highly coherent field-emission gun (FEG) electron source. To acquire an off-axis electron hologram, the region of interest on the specimen should be positioned so that it covers approximately half the field of view. An electron biprism, which takes the form of a fine ($<1\ \mu\text{m}$ diameter) wire (Möllenstedt and Düker, 1954), is located below the sample,

usually in place of one of the conventional selected-area apertures. The application of a voltage to the biprism results in overlap of a “reference” electron wave that has passed through vacuum (or through a thin region of support film) with the electron wave that has passed through the specimen, as shown schematically in Figure 18–1a. If the illumination is sufficiently coherent, then holographic interference fringes are

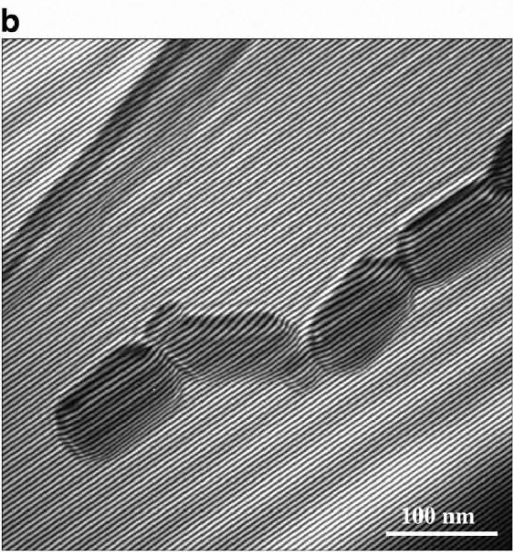
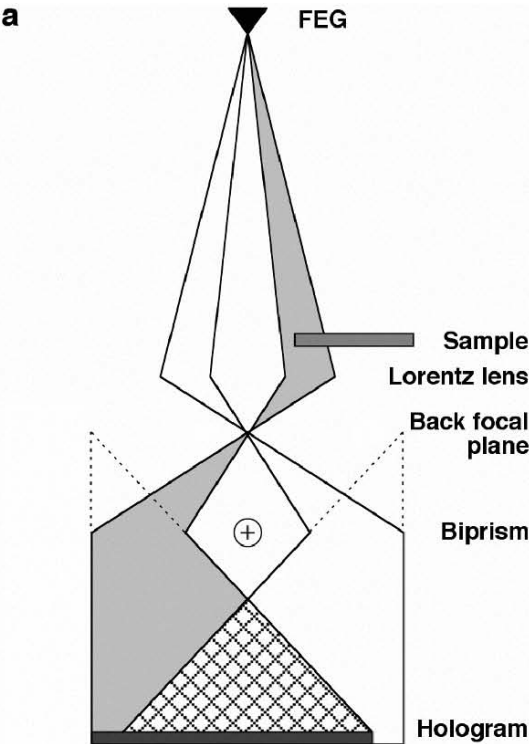


Figure 18–1. (a) Schematic illustration of the set-up used for generating off-axis electron holograms. The specimen occupies approximately half the field of view. Essential components are the field-emission gun (FEG) electron source, which provides coherent illumination, and the electron biprism, which causes overlap of the object and (vacuum) reference waves. The Lorentz lens allows imaging of magnetic materials in close to field-free conditions. (b) Off-axis electron hologram of a chain of magnetite (Fe_3O_4) crystals, recorded at 200 kV using a Philips CM200 FEGTEM. The crystals are supported on a holey carbon film. Phase changes can be seen in the form of bending of the holographic interference fringes as they pass through the crystals. Fresnel fringes from the edges of the biprism wire are also visible. (Reprinted from Dunin-Borkowski et al., 2004a.)

formed in the overlap region, with a spacing that is inversely proportional to the biprism voltage (Matteucci et al., 1998). The amplitude and the phase shift of the electron wave from the specimen are recorded in the intensity and the position of the holographic fringes, respectively. A representative off-axis electron hologram of a chain of magnetite nanocrystals is shown in Figure 18-1b.

For coherent TEM imaging, the electron wavefunction in the image plane can be written in the form

$$\psi_i(\mathbf{r}) = A_i(\mathbf{r}) \exp[i\phi_i(\mathbf{r})] \quad (1)$$

where $\mathbf{r}$ is a two-dimensional vector in the plane of the sample, A and ϕ refer to amplitude and phase, and the subscript i refers to the image plane. The recorded intensity distribution is then given by the expression

$$I_i(\mathbf{r}) = |A_i(\mathbf{r})|^2 \quad (2)$$

Thus, the image intensity can be described as the modulus squared of an electron wavefunction that has been modified by the specimen and the objective lens. The intensity distribution in an off-axis electron hologram can be represented by the addition of a tilted plane reference wave to the complex specimen wave, in the form

$$I_{\text{hol}}(\mathbf{r}) = |\psi_i(\mathbf{r}) + \exp[2\pi i \mathbf{q}_c \cdot \mathbf{r}]|^2 \quad (3)$$

$$= 1 + A_i^2(\mathbf{r}) + 2A_i(\mathbf{r}) \cos[2\pi i \mathbf{q}_c \cdot \mathbf{r} + \phi_i(\mathbf{r})] \quad (4)$$

where the tilt of the reference wave is specified by the two-dimensional reciprocal space vector $\mathbf{q} = \mathbf{q}_c$. It can be seen from Eq. (4) that there are three separate contributions to the intensity distribution in a hologram: the reference wave, the image wave, and an additional set of cosinusoidal fringes with local phase shifts and amplitudes that are exactly equivalent to the phase and amplitude of the electron wavefunction in the image plane, ϕ_i and A_i , respectively.

1.2 Hologram Reconstruction

To obtain amplitude and phase information, the off-axis electron hologram is first Fourier transformed. From Eq. (4), the complex Fourier transform of the hologram is given by

$$\begin{aligned} FT[I_{\text{hol}}(\mathbf{r})] &= \delta(\mathbf{q}) + FT[A_i^2(\mathbf{r})] \\ &\quad + \delta(\mathbf{q} + \mathbf{q}_c) \otimes FT\{A_i(\mathbf{r}) \exp[i\phi_i(\mathbf{r})]\} \\ &\quad + \delta(\mathbf{q} - \mathbf{q}_c) \otimes FT\{A_i(\mathbf{r}) \exp[-i\phi_i(\mathbf{r})]\} \end{aligned} \quad (5)$$

Equation (5) describes a peak at the reciprocal space origin corresponding to the Fourier transform of the reference image, a second peak centered on the origin corresponding to the Fourier transform of a bright-field TEM image of the sample, a peak centered at $\mathbf{q} = -\mathbf{q}_c$ corresponding to the Fourier transform of the desired image wavefunction, and a peak centered at $\mathbf{q} = +\mathbf{q}_c$ corresponding to the Fourier transform of the complex conjugate of the wavefunction.

The reconstruction of a hologram to obtain amplitude and phase information is illustrated in Figure 18-2. Figure 18-2a-c shows a

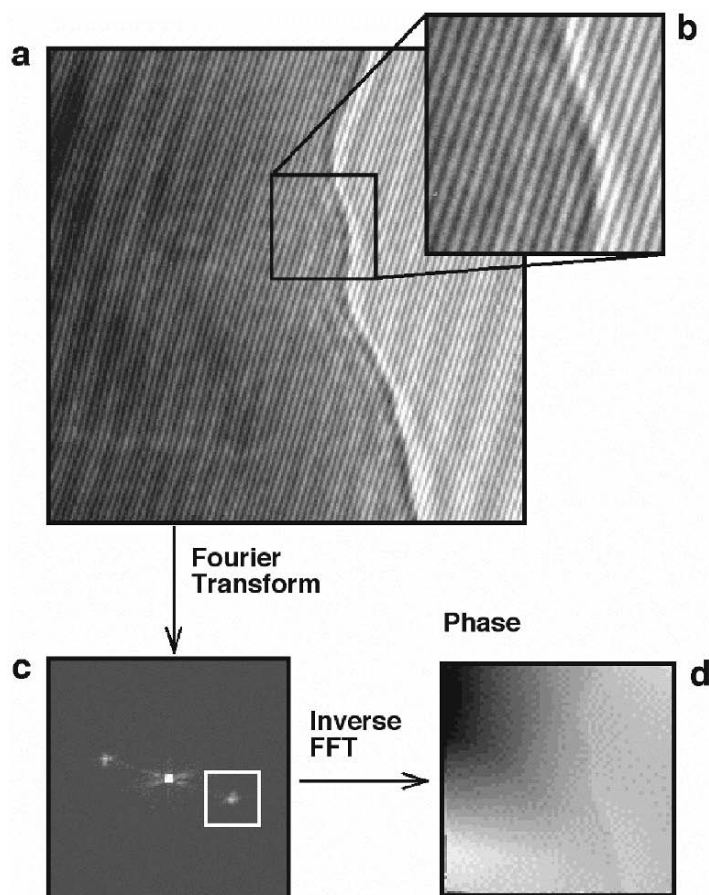


Figure 18–2. (a) Off-axis electron hologram recorded from a thin crystal. (b) Enlargement showing interference fringes within the specimen. (c) Fourier transform of the electron hologram. (d) Phase image obtained after inverse fast Fourier transformation (IFFT) of the sideband marked with a box in (c). (Reprinted from Dunin-Borkowski et al., 2000.)

hologram of a thin crystal, an enlargement of part of the hologram, and a Fourier transform of the entire hologram, respectively. To recover the amplitude and the relative phase shift of the electron wavefunction, one of the two sidebands is selected digitally and inverse Fourier transformed, as shown in Figure 18–2d. The amplitude and phase of this complex wavefunction are then easily calculated.

The impact of electron holography results from the dependence of the phase shift on the electrostatic potential and the in-plane component of the magnetic induction in the specimen. Neglecting dynamical diffraction (i.e., assuming that the specimen is thin and weakly diffracting), the phase shift can be expressed in the form

$$\phi(x) = C_E \int V(x, z) dz - \left(\frac{e}{\hbar} \right) \iint B_{\perp}(x, z) dx dz \quad (6)$$

where

$$C_E = \left(\frac{2\pi}{\lambda} \right) \left[\frac{E + E_0}{E(E + 2E_0)} \right] \quad (7)$$

z is in the incident electron beam direction, x is in the plane of the specimen, $B_{\perp}$ is the component of the magnetic induction within and outside the specimen perpendicular to both x and z , V is the electrostatic potential, λ is the (relativistic) electron wavelength, and E and E_0 are, respectively, the kinetic and rest mass energies of the incident electron (Reimer, 1991). C_E has values of 7.29×10^6 , 6.53×10^6 , and $5.39 \times 10^6 \text{ rad V}^{-1} \text{ m}^{-1}$ at 200 kV, 300 kV, and 1 MV, respectively. If neither V nor $B_{\perp}$ varies along the electron beam direction within a sample of thickness t , then Eq. (6) can be simplified to

$$\phi(x) = C_E V(x) t(x) - \left(\frac{e}{\hbar} \right) \int B_{\perp}(x) t(x) dx \quad (8)$$

By making use of Eqs. (6) and (8), information about V and $B_{\perp}$ can be recovered from a measurement of the phase shift ϕ , as described below.

1.3 Experimental Considerations

In practice, several issues must be addressed to record and analyze an electron hologram successfully. A key experimental requirement is the availability of a vacuum reference wave that can be overlapped onto the region of interest on the specimen, which usually implies that the hologram must be recorded from a region close to the specimen edge. This restriction can be relaxed if a thin, clean, and weakly diffracting region of electron-transparent support film, rather than vacuum, can be overlapped onto the region of interest.

As phase information is stored in the lateral displacement of the holographic interference fringes, long-range phase modulations arising from inhomogeneities in the charge and thickness of the biprism wire, as well as from lens distortions and charging effects (e.g., at apertures), must be taken into account by using a reference hologram obtained by removing the specimen from the field of view without changing the optical parameters of the microscope. Correction is then possible by performing a complex division of the sample and reference waves in real space to obtain the distortion-free phase of the image wave (de Ruijter and Weiss, 1993).

The need for this procedure is illustrated in Figure 18–3. Figure 18–3a shows a reconstructed phase image of a wedge-shaped crystal of indium phosphide (InP) obtained before distortion correction, with the vacuum region outside the sample edge on the left side of the image. Figure 18–3b shows the corresponding vacuum reference phase image, which was acquired with the sample removed from the field of view but with all other imaging parameters unchanged. The equiphase contour lines now correspond to distortions that must be removed from the phase image of the sample. Figure 18–3c shows the distortion-corrected phase image of the sample, which was obtained by dividing the two complex image waves. The vacuum region in Figure 18–3c is

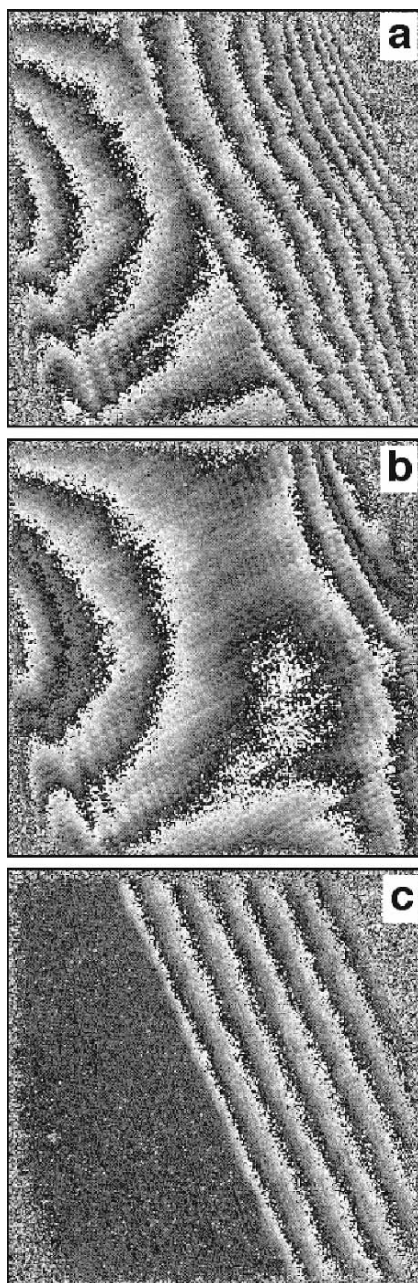


Figure 18-3. Illustration of distortion correction procedure. (a) Initial (six times amplified) phase image of a wedge-shaped InP crystal recorded at 100 kV using a Philips 400ST FEGTEM equipped with a Gatan 679 slow-scan CCD (charge coupled device) camera. (b) Phase image obtained from vacuum, with the specimen removed from the field of view. (c) Corrected phase image, obtained by subtracting image (b) from image (a). This procedure is carried out by dividing the complex image waves obtained by inverse Fourier transforming the sideband obtained from each hologram, and by calculating the phase of the resulting wavefunction. (Reprinted from Smith et al., 1998.)

flattened substantially by this procedure, which allows relative phase changes within the sample to be interpreted much more reliably. The acquisition of a reference hologram has the additional advantage that it allows the center of the sideband in Fourier space to be determined accurately. The use of the same location for the sideband in the Fourier transforms of the sample and reference waves removes any tilt of the recorded wave that might be introduced by an inability to locate the exact (subpixel) position of the sideband frequency in Fourier space.

Electron holograms have traditionally been recorded on photographic film, but digital acquisition using charge coupled device (CCD) cameras is now widely used due to their linear response, dynamic range, and high detection quantum efficiency, as well as the immediate accessibility to the recorded information (de Ruijter, 1995; Meyer and Kirkland, 2000). Whether a hologram is recorded on film or digitally, the field of view is typically limited to approximately $5\mu\text{m}$ by the dimensions of the recording medium and the sampling of the holographic fringes. A phase image that is calculated digitally is usually evaluated modulo 2π , meaning that 2π phase discontinuities that are unrelated to specimen features will appear at positions where the phase shift exceeds this amount. The phase image must often then be “unwrapped” using suitable algorithms (Ghiglia and Pritt, 1998).

The high electron beam coherence that is required for electron holography requires the use of an FEG electron source, a small spot size, a small condenser aperture, and a low gun extraction voltage. The coherence may be improved further by adjusting the condenser lens stigmators in the microscope to provide elliptical illumination that is wide in the direction perpendicular to the biprism when the condenser lens is overfocused (Smith and McCartney, 1998). The contrast of the holographic interference fringes is determined primarily by the lateral coherence of the electron wave at the specimen level, the mechanical stability of the biprism wire, and the point spread function of the recording medium. The fringe contrast

$$\mu = \left(\frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \right) \quad (9)$$

can be determined from a holographic interference fringe pattern that has been recorded in the absence of a sample, where $I_{\max}$ and $I_{\min}$ are the maximum and minimum intensities of the interference fringes, respectively (Völkl et al., 1995). Should the fringe contrast decrease too much, reliable reconstruction of the image wavefunction will no longer be possible.

The phase detection limit for electron holography (Harscher and Lichte, 1996) can be determined from the effect on the recorded hologram of Poisson-distributed shot noise, the detection quantum efficiency and point spread function of the CCD camera, and the fringe contrast. The minimum phase difference between two pixels that can be detected is given by the expression

$$\Delta\phi_{\min} = \frac{SNR}{\mu} \sqrt{\frac{2}{N_{\text{el}}}} \quad (10)$$

where SNR is the signal-to-noise ratio, μ is defined in Eq. (9), and N_{el} is the number of electrons collected per pixel (Lichte, 1995). In practice, some averaging of the measured phase is often implemented, particularly if the features of interest vary slowly across the image or only in one direction. A final artifact results from the presence of Fresnel diffraction at the biprism wire, which is visible in Figure 18–1b and causes phase and amplitude modulations of both the image and the reference

wave (Yamamoto et al., 2000). These effects can be removed to some extent by using a reference hologram, and by Fourier filtering the sideband before reconstruction of the image wave. More advanced approaches for removing Fresnel fringes from electron holograms based on image analysis (Fujita and McCartney, 2005) and double-biprism electron holography (Harada et al., 2004; Yamamoto et al., 2004) have recently been introduced. Great care should also be taken to assess the effect on the reference wave of long-range electromagnetic fields that may extend outside the sample and affect both the object wave and the reference wave (Matteucci et al., 1994).

2 Measurement of Mean Inner Potential and Sample Thickness

Before describing the application of electron holography to the characterization of magnetic and electrostatic fields, the use of the technique to measure local variations in specimen morphology and composition is considered. Such measurements are possible from a phase image that is associated solely with variations in mean inner potential and specimen thickness. When a specimen has uniform structure and composition in the electron beam direction, and in the absence of magnetic and long-range electrostatic fields (such as those at depletion regions in semiconductors), Equation (8) can be rewritten in the form

$$\phi(x) = C_E V_0(x)t(x) \quad (11)$$

where the mean inner potential of the specimen, V_0 , is the volume average of the electrostatic potential (Spence, 1993). Values of V_0 can in principle be calculated from the equation

$$V_0 = \left(\frac{h^2}{2\pi m e \Omega} \right) \sum_{\Omega} f_{el}(0) \quad (12)$$

by treating the specimen as an array of neutral atoms. In Eq. (12), $f_{el}(0)$ are electron scattering factors at zero scattering angle for each atom, which have been calculated by, for example, Doyle and Turner (1968) and Rez et al. (1994), and Ω is typically the volume of the unit cell in a crystalline material. However, values of V_0 that are calculated using Eq. (12) are invariably overestimated as a result of bonding in the specimen (Radi, 1970; Gajdardziska-Josifovska and Carim, 1998). It is therefore important to obtain experimental measurements of V_0 . According to Eq. (11), an independent measure of the specimen thickness profile is required to determine V_0 , for example, by examining a specimen in which the thickness changes in a well-defined manner, as shown in Figure 18-4a for a phase profile obtained from a 90° wedge of GaAs tilted to a weakly diffracting orientation. If the specimen thickness profile is known, then V_0 can be determined by measuring the gradient of the phase $d\phi/dx$, and making use of the relation

$$V_0 = \left(\frac{1}{C_E} \right) \left(\frac{d\phi/dx}{dt/dx} \right) \quad (13)$$

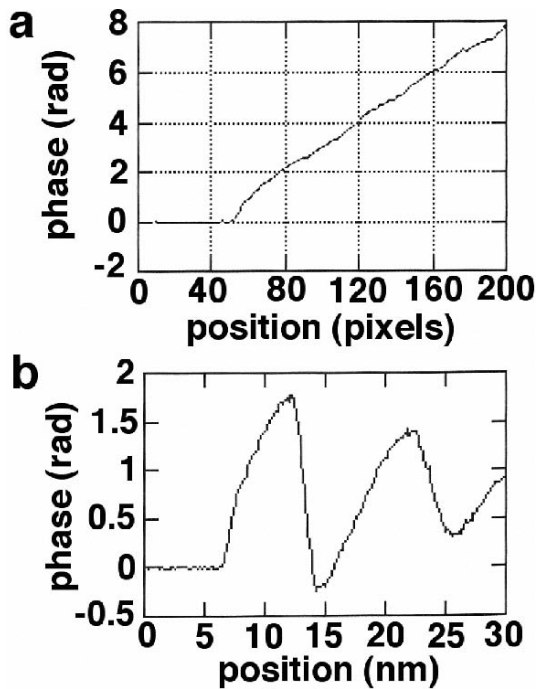


Figure 18–4. (a) Phase profile plotted as a function of distance into a 90° GaAs cleaved wedge specimen tilted to a weakly diffracting orientation. The phase change increases approximately linearly with specimen thickness. (b) Phase profile obtained from a GaAs wedge tilted close to a $[100]$ zone axis, showing strong dynamical effects. (Reprinted from Gajdardziska-Josifovska and Carim, 1998.)

This approach has been used successfully to measure the mean inner potential of cleaved wedges and cubes of Si, MgO, GaAs, PbS (Gajdardziska-Josifovska et al., 1993; de Ruijter et al., 1992), and Ge (Li et al., 1999). The resulting values of V_0 that were determined for MgO, GaAs, PbS, and Ge using this approach are 13.0 ± 0.1 , 14.5 ± 0.2 , 17.2 ± 0.1 , and 14.3 ± 0.2 V, respectively. In a similar study, wedge-shaped Si samples with stacked Si oxide layers on their surfaces were used to measure the mean inner potentials of the oxide layers (Rau et al., 1996). Experimental measurements of V_0 have been obtained from 20- to 40-nm-diameter Si nanospheres coated in layers of amorphous SiO_2 (Wang et al., 1997). The mean inner potential of crystalline Si was found to be 12.1 ± 1.3 V, that of amorphous Si 11.9 ± 0.9 V, and that of amorphous SiO_2 10.1 ± 0.6 V. Similar measurements obtained from spherical latex particles embedded in vitrified ice have provided values for V_0 of 8.5 ± 0.7 and 3.5 ± 1.2 V for the two materials, respectively (Harscher and Lichte, 1998).

Dynamical contributions to the phase shift complicate the determination of V_0 from crystalline samples. These corrections can be taken into account by using either multislice or Bloch wave algorithms. The

fact that Eq. (11) is no longer valid when the sample is tilted to a strongly diffracting orientation is demonstrated in Figure 18–4b for a 90° cleaved wedge sample of GaAs that has been tilted to a [100] zone axis. The phase shift varies nonlinearly with sample thickness, and is also very sensitive to small changes in sample orientation. Additional experimental factors that may affect measurements of V_0 include the chemical and physical state and the crystallographic orientation of the specimen surface (O’Keeffe and Spence, 1994), and specimen charging (Lloyd et al., 1997).

When V_0 is already known, then measurements of the phase shift can be used to determine the local specimen thickness t . Alternatively, the specimen thickness can be inferred from a holographic amplitude image in units of λ_{in} , the mean free path for inelastic scattering, by making use of the relation

$$\frac{t(x)}{\lambda_{\text{in}}} = -2 \ln \left[\frac{A_i(x)}{A_r(x)} \right] \quad (14)$$

where $A_i(x)$ and $A_r(x)$ are the measured amplitudes of the sample and reference holograms, respectively (McCartney and Gajdardziska-Josifovska, 1994). If desired, the thickness dependence of both the phase and the amplitude image can be removed by combining Eqs. (11) and (14) in the form

$$\frac{\phi(x)}{-2C_E \ln \left[\frac{A_i(x)}{A_r(x)} \right]} = V_0(x) \lambda_{\text{in}}(x) \quad (15)$$

Equation (15) can be used to generate an image where the contrast is the product of the local values of the mean inner potential and the inelastic mean free path. These parameters depend only on the local composition of the sample, and thus can be useful for interpreting images obtained from samples with varying composition and thickness (Weiss et al., 1991).

When the mean inner potential in a specimen is constant or if its variation across the specimen is known, then the morphologies of nanoscale particles (for which dynamical contributions to the phase shift are likely to be small) can be measured using electron holography by making use of Eq. (11). Examples of the measurement of particle shapes using this approach include the characterization of faceted ZrO_2 crystals (Allard et al., 1996), carbon nanotubes (Lin and Dravid, 1996), bacterial flagellae (Aoyama and Ru, 1996), and atomic-height steps on clean surfaces of MoS_2 (Tonomura et al., 1985). Such measurements can, in principle, be extended to three dimensions by combining electron holography with electron tomography, as demonstrated by the acquisition and analysis of tilt series of electron holograms of latex particles (Lai et al., 1994). The demanding nature of these latter measurements results from the fact that specimen tilt angles of at least $\pm 60^\circ$, as well as small tilt steps and accurate alignment of the resulting phase images, are all essential to avoid reconstruction artifacts.

3 Measurement of Magnetic Fields

The most successful and widespread applications of electron holography have involved the characterization of magnetic fields within and surrounding materials at medium spatial resolution. When examining magnetic materials, the normal objective lens is usually switched off, as its strong magnetic field will likely saturate the magnetization in the sample along the electron beam direction. A high-strength mini-lens located below the objective lens can instead be used to provide reasonably high magnification ($\sim 50\text{--}75\times$) with the sample still in a magnetic field-free environment.

3.1 Early Experiments

Early examples of the examination of magnetic materials using electron holography involved the reconstruction of holograms using a laser bench, and included the characterization of horseshoe magnets (Matsuda et al., 1982), magnetic recording media (Osakabe et al., 1983), and vortices in superconductors (Matsuda et al., 1989, 1991; Bonevich et al., 1993). The most elegant series of experiments involved the confirmation of the Aharonov–Bohm effect (Aharonov and Bohm, 1959), which states that when an electron wave from a point source passes on either side of an infinitely long solenoid then the relative phase shift that occurs between the two parts of the wave should result from the presence of a vector potential. In this way, the Aharonov–Bohm effect provides the only observable confirmation of the physical reality of gauge theory. Electron holography experiments were carried out on 20-nm-thick permalloy toroidal magnets that were covered with 300-nm-thick layers of superconducting Nb, which prevented electrons from penetrating the magnetic material and confined the magnetic flux by exploiting the Meissner effect (Tonomura et al., 1982, 1983). The observations showed that the phase difference between the center of the toroid and the region outside was quantized to a value of 0 or π when the temperature was below the Nb superconducting critical temperature (5 K), i.e., when a supercurrent was induced to circulate in the magnet. The observed quantization of magnetic flux, and the measured phase differences with the magnetic field entirely screened by the superconductor, provided unequivocal confirmation of the Aharonov–Bohm effect.

3.2 Digital Acquisition and Analysis

Recent applications of electron holography to the characterization of magnetic fields in nanostructured materials have been based on digital recording and processing. The examples chosen here highlight the different approaches that can be used to separate a weak magnetic signal from a recorded phase image, as well as illustrating the magnetic properties of the materials. The off-axis mode of electron holography is ideally suited to the characterization of such materials because unwanted contributions to the contrast from local variations in

composition and specimen thickness can usually be removed from a phase image more easily than from images recorded using other TEM phase contrast techniques. For example, the Fresnel and Foucault modes of Lorentz microscopy and differential phase contrast (DPC) imaging provide signals that are approximately proportional to either the first or the second differential of the phase shift. These techniques inherently enhance contributions to the contrast from rapid variations in specimen thickness and composition, as compared to the weak and slowly varying magnetic signal.

The *digital* acquisition, reconstruction, and analysis of electron holograms have allowed magnetic fields within samples with small feature sizes and rapid variations in thickness or composition to be examined. The key advantage of digital analysis is that the magnetic and mean inner potential contributions to the measured holographic phase shift can be separated, particularly at the edges of nanostructured particles, where rapid changes in specimen thickness often dominate both the measured phase and the phase gradient. The approaches that can be used to achieve this separation are described below. Digital analysis also facilitates the construction of line profiles from phase images, which can provide quantitative information such as the widths of domain walls.

Determination of the phase gradient is particularly useful for a magnetic material because of the following relationship, obtained by differentiating Eq. (8):

$$\frac{d\phi(x)}{dx} = C_E \frac{d}{dx} [V(x)t(x)] - \left(\frac{e}{\hbar}\right) B_{\perp}(x)t(x) \quad (16)$$

According to Eq. (16), for a specimen of uniform thickness and composition the phase gradient is proportional to the in-plane component of the magnetic induction in the specimen

$$\frac{d\phi(x)}{dx} = -\left(\frac{et}{\hbar}\right) B_{\perp}(x) \quad (17)$$

A direct graphic representation of the magnetic induction can therefore be obtained by adding contours to a magnetic phase image, where a phase difference of 2π corresponds to an enclosed magnetic flux of 4×10^{-15} Wb. Significantly, an experimental phase image does not need to be unwrapped to evaluate its first differential digitally. Instead, if the reconstructed image wave is designated ψ , then the phase differential can be determined directly from the expression

$$\frac{d\phi(x, y)}{dx} = \text{Im} \left(\frac{\frac{d\psi(x, y)}{dx}}{\psi(x, y)} \right) \quad (18)$$

Most of the results that are described below were acquired using Philips CM200ST and CM300ST FEGTEMs equipped with rotatable electron biprisms, and with Lorentz minilenses located in the bores of their objective lens pole-pieces. The Lorentz lenses allow electron holo-

grams to be recorded at magnifications of up to $\sim 75\times$ with the specimens located in a magnetic field-free environment.

3.2.1 NdFeB Hard Magnets

Figure 18–5a shows a Lorentz (Fresnel defocus) image of a $\text{Nd}_2\text{Fe}_{14}\text{B}$ specimen, in which magnetic domains can be seen (McCartney and Zhu, 1998). Such images provide little information about the direction of the local magnetic induction in the specimen. An electron holographic phase image acquired from the same area using an interference fringe spacing of 2.5 nm is shown in Figure 18–5b. Gradients of the phase image were calculated along the $+x$ and $-y$ directions, as shown in Figure 18–5c and d, respectively. These images were combined to form a vector map of the magnetic induction, as shown in Figure 18–5e. The map is divided into 20-nm squares, and has a low contrast phase gradient image superimposed on it for reference purposes. The minimum vector length is zero (corresponding to out-of-plane induction), while the maximum vector length is consistent with a measured induction $B = 1.0\text{ T}$. A vector map of the region marked in Figure 18–5e is shown at higher magnification in Figure 18–5f. In this map, magnetic vortices show Bloch-like character, with vanishingly small vector lengths. Care is needed when interpreting the fine details in such induction maps due to the undetermined effects of magnetic fringing fields immediately above and below the sample, as well as possible contributions from variations in specimen thickness. A single pixel line scan across a 90° domain wall, which appears as the bright ridge near the central part of Figure 18–5b, is shown in Figure 18–5g. This line profile places an upper limit of 10 nm on the domain wall width, which agrees well with theoretical estimates.

3.2.2 Co Nanoparticle Chains

The dominant nature of the mean inner potential contribution to the phase shift recorded from a nanoscale magnetic particle is illustrated in Figure 18–6. Figure 18–6a and b shows a hologram and a reconstructed phase image of a chain of Co particles suspended over a hole in a carbon support film (de Graef et al., 1999). Figure 18–6c and d shows corresponding line traces determined from the phase image across the centers of two particles. Each trace is obtained in a direction perpendicular to the chain axis. The magnetic induction and mean inner potential of each particle can be determined by fitting simulations to the experimental line traces. Analytical expressions for the expected phase shifts can be derived for a uniformly magnetized sphere of radius a , magnetic induction $B_\perp$ (along y), and mean inner potential V_0 in the form

$$\begin{aligned} \phi(x, y) \Big|_{(x^2+y^2) \leq a^2} = & 2C_E V_0 \sqrt{a^2 - (x^2 + y^2)} \\ & + \left(\frac{e}{\hbar} \right) B_\perp a^3 \left(\frac{x}{x^2 + y^2} \right) \\ & \left\{ 1 - \left[1 - \left(\frac{x^2 + y^2}{a^2} \right) \right]^{\frac{3}{2}} \right\} \end{aligned} \quad (19)$$

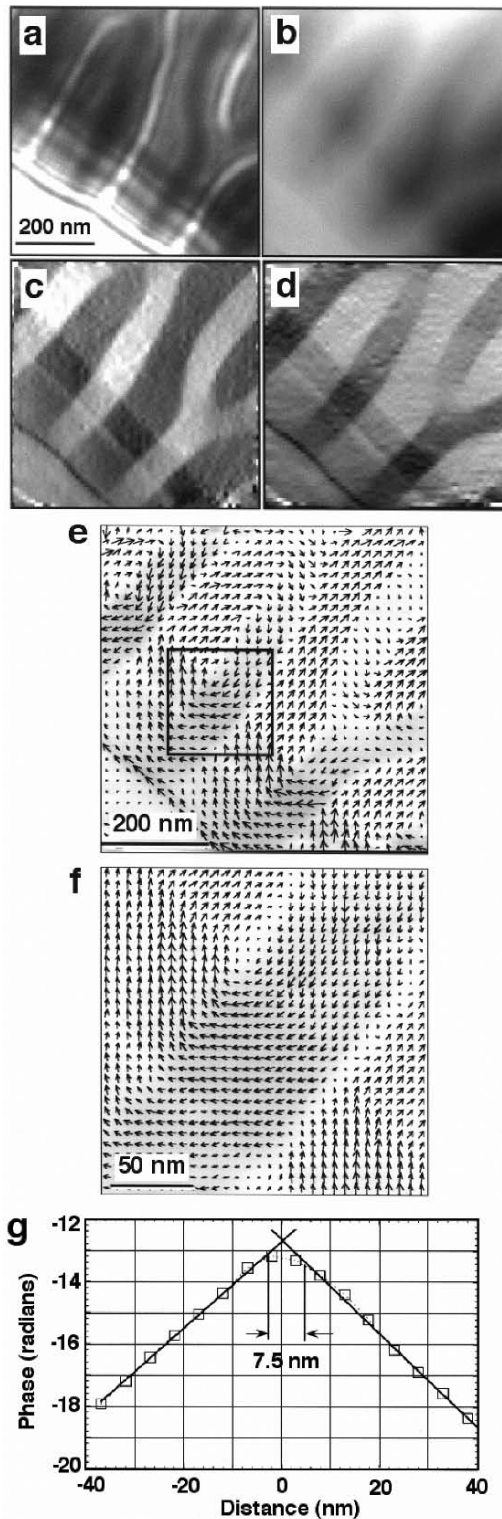


Figure 18-5. (a) Room temperature Lorentz (Fresnel underfocus) image of an $\text{Nd}_2\text{Fe}_{14}\text{B}$ hard magnet, recorded at 200 kV using a Philips CM200 FEGTEM operated in Lorentz mode. (b) Phase image of the same region of the specimen reconstructed from an electron hologram obtained using an interference fringe spacing of 2.5 nm. (c and d) Gradients of the phase image shown in (b), calculated parallel to the +x and -y directions, respectively. (e) Induction map derived from the phase gradients shown in (c) and (d). (f) Enlargement of the area indicated in (e). (g) Linescan obtained across a 90° domain wall that appears as a bright ridge near the center of image (b). The line profile provides an upper limit for the domain wall width of 10 nm. (Reprinted from McCartney and Zhu, 1998.)

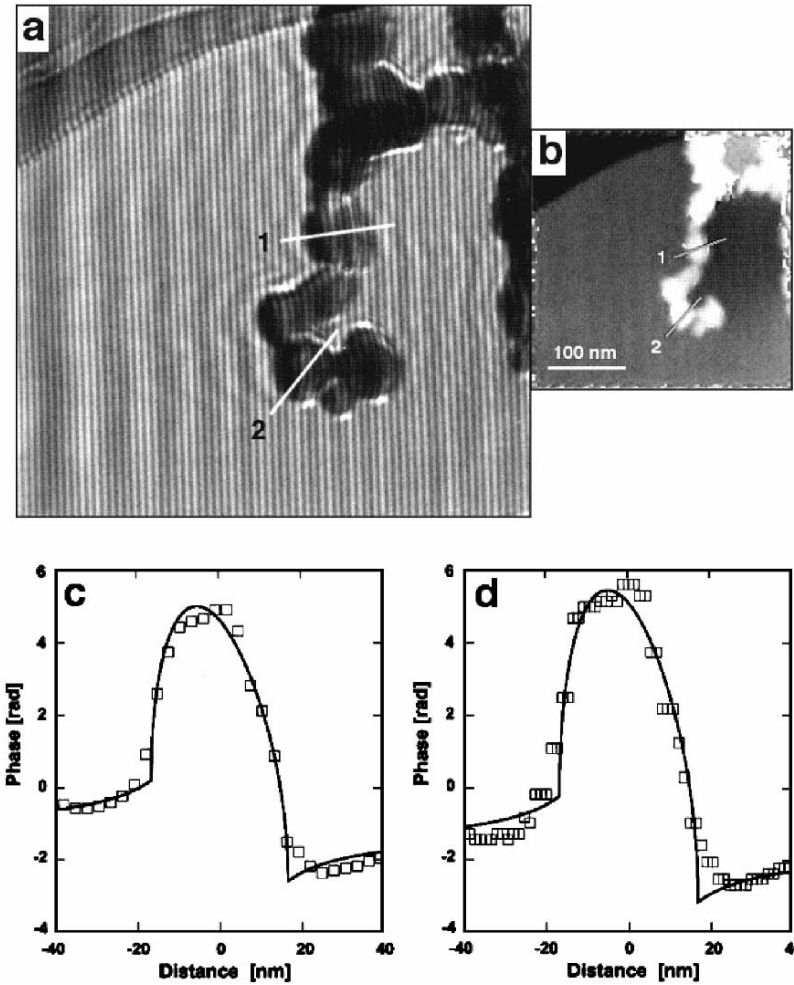


Figure 18–6. (a) Off-axis electron hologram of a chain of Co particles suspended over a hole in a carbon support film, acquired at 200 kV using a Philips CM200 FEGTEM and a biprism voltage of 90 V. (b) Corresponding unwrapped phase image. (c and d) Experimental line profiles formed from lines 1 and 2 in (b), and fitted phase profiles generated for spherical Co nanoparticles. (Reprinted from de Graef et al., 1999.)

$$\phi(x, y)|_{(x^2+y^2)>a^2} = \left(\frac{e}{\hbar}\right) B_{\perp} a^3 \left(\frac{x}{x^2 + y^2} \right) \quad (20)$$

For line profiles obtained through the centers of the particles in a direction perpendicular to that of $B_{\perp}$, these expressions reduce to

$$\begin{aligned} \phi(x)|_{x \leq a} = & 2C_E V_0 \sqrt{a^2 - x^2} \\ & + \left(\frac{e}{\hbar}\right) B_{\perp} \left[\frac{a^3 - (a^2 - x^2)^{\frac{3}{2}}}{x} \right] \end{aligned} \quad (21)$$

$$\phi(x)|_{x > a} = \left(\frac{e}{\hbar}\right) B_{\perp} \left(\frac{a^3}{x} \right) \quad (22)$$

Least squares fits of Eqs. (21) and (22) to the experimental data points, which are also shown in Figure 18–6c and d, provide best-fitting values for a , $B_{\perp}$, and V_0 of 17 nm, 1.7 T, and 26 V, respectively.

3.3 Separation of Magnetic and Mean Inner Potential Contributions

When characterizing magnetic fields inside nanostructured materials, the mean inner potential contribution to the measured phase shift must in general be removed to interpret the magnetic contribution of primary interest. Several approaches can be used. The sample may be inverted to change the sign of the magnetic contribution to the signal and a second hologram recorded. The sum and the difference of the two phase images can then be used to provide twice the magnetic contribution, and twice the mean inner potential contribution, respectively (Wohlleben, 1971; Tonomura et al., 1986). Alternatively, two holograms may be acquired from the same area of the specimen at two different microscope accelerating voltages. In this case, the magnetic signal is independent of accelerating voltage, and subtraction of the two phase images can be used to provide the mean inner potential contribution. A more practical method of removing the mean inner potential contribution involves performing magnetization reversal *in situ* in the electron microscope, and subsequently selecting pairs of holograms that differ only in the (opposite) directions of the magnetization. The magnetic and mean inner potential contributions to the phase can be calculated by taking half the difference, and half the sum, of the phases. The mean inner potential contribution can then be subtracted from all other phase images acquired from the same specimen region (Dunin-Borkowski et al., 1998). *In situ* magnetization reversal, which is required both for this purpose and for performing magnetization reversal experiments in the TEM, can be achieved by exciting the conventional microscope objective lens slightly and tilting the specimen to apply known magnetic fields, as shown schematically in Figure 18–7. Subsequently, electron holograms can be recorded with the conventional microscope objective lens switched off and the Lorentz lens

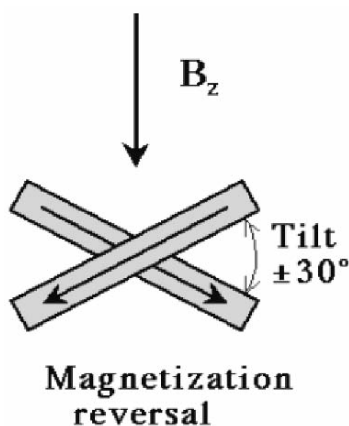


Figure 18–7. Schematic diagram illustrating the use of specimen tilt to provide an in-plane component of the external field for *in situ* magnetization reversal experiments. (Reprinted from Dunin-Borkowski et al., 2000.)

switched on, while the magnetic specimen is located in a magnetic field-free environment.

3.3.1 Magnetite Nanoparticle Chains

The chain of magnetite nanoparticles shown in Figure 18–8 illustrates the fact that both the mean inner potential and the magnetic contribution to the phase shift can provide useful information. In particular, the mean inner potential contribution can be used to interpret the morphologies and orientations of nanoparticles, as discussed above. Figure 18–8a and b shows phase contours generated from, respectively, the mean inner potential and magnetic contributions to the phase shift at the end of a chain of magnetite (Fe_3O_4) crystals from a magnetotactic bacterium collected from a brackish lagoon at Itaipu in Brazil. The magnetic moment that the crystals impart to the bacterial cell results in its alignment and subsequent migration along the Earth's magnetic field lines (Blakemore, 1975; Bazylinski and Moskowitz, 1997; Dunlop and Özdemir, 1997). Separation of the mean inner potential and magnetic contributions to the phase shift was achieved *in situ* by using the field of the conventional microscope objective lens to magnetize each chain parallel and then antiparallel to its length, as illustrated in Figure 18–7. The contours in Figure 18–8a and b have been overlaid onto the mean inner potential contribution to the phase. In Figure 18–8a, they are associated with variations in specimen thickness and are confined primarily to the crystals, while in Figure 18–8b they correspond to magnetic lines of force, which extend smoothly from within the crystals to the surrounding region. Figure 18–8c shows line profiles measured across the large and small magnetite crystals visible close to the centers of Figure 18–8a and b, in a direction perpendicular to the chain axis. Individual experimental data points are shown as open circles. Corresponding simulations based on Eq. (8) are shown on the same axes. The darker solid line shows the best-fitting simulation to the data for the larger crystal, on the assumption that the external shape is formed from a combination of {111}, {110}, and {100} faces. The simulation corresponds to a distorted hexagonal shape in cross section (shown as an inset above the figure). The lighter line shows a worse fit, provided by assuming a diamond shape in cross section.

3.3.2 Co Nanoparticle Rings

An illustration of the characterization of magnetostatic interactions between particles that each contains a single magnetic domain is provided by the examination of rings of 20-nm-diameter crystalline Co particles, as shown in Figure 18–9. Such rings are appealing candidates for high density information storage applications because they are expected to form chiral domain states that exhibit flux closure (FC). Nanoparticle rings are also of interest for the development of electron holography because their magnetization directions cannot be reversed by applying an in-plane external field. As a result, phase images were obtained both before and after inverting the specimen. The resulting pairs of phase images were aligned in position and angle, and their sum and difference calculated as described above. Figure 18–9a shows a low magnification bright-field image of the Co rings (Tripp et al.,

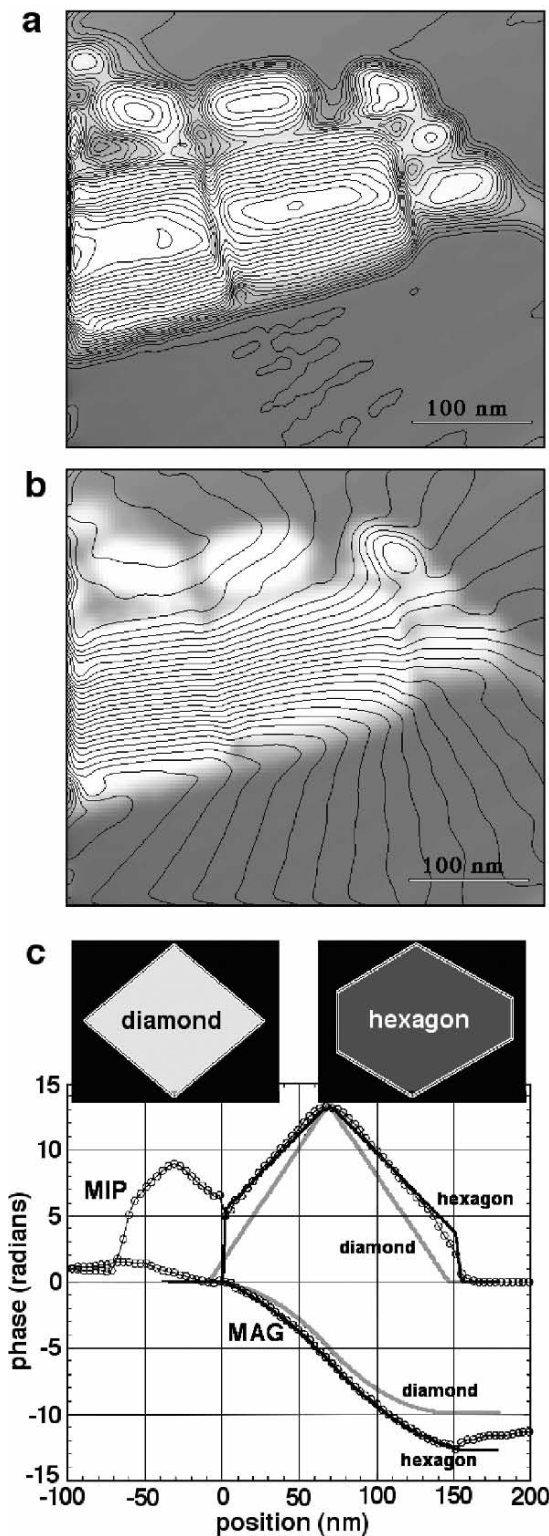
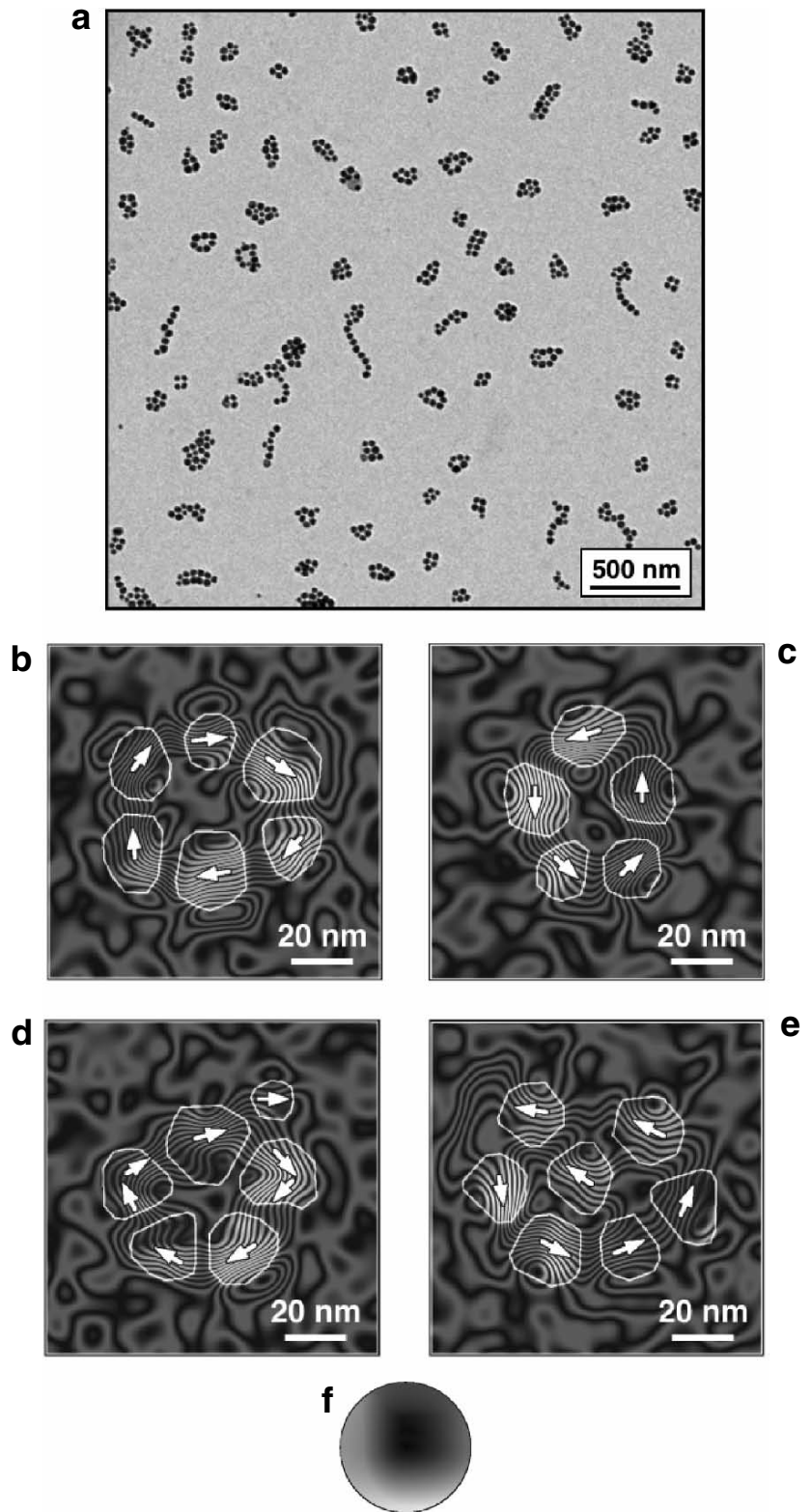


Figure 18-8. Phase contours showing (a) the mean inner potential and (b) the magnetic contribution to the phase shift at the end of a chain of magnetite crystals from magnetotactic bacteria collected from a brackish lagoon at Itaipu in Brazil. The contours have been overlaid onto the mean inner potential contribution to the phase. (c) Line profiles obtained from images (a) and (b) across the large and small magnetite crystals close to the center of each image. The experimental data are shown as open circles. The darker line shows the best-fitting simulation to the data for the larger crystal, corresponding to a distorted hexagonal cross section (shown above the figure). The lighter line shows the worse fit that results from assuming a diamond shape in cross section (also shown above the figure). (Reprinted from Dunin-Borkowski et al., 2004a.)

Figure 18–9. (a) Low magnification bright-field image of self-assembled Co nanoparticle rings and chains deposited onto an amorphous carbon support film. Each Co particle has a diameter of between 20 and 30 nm. (b–e) Magnetic phase contours (128× amplification; 0.049 radian spacing), formed from the magnetic contribution to the measured phase shift, in four different nanoparticle rings. The outlines of the nanoparticles are marked in white, while the direction of the measured magnetic induction is indicated both using arrows and according to the color wheel shown in (f) (red = right, yellow = down, green = left, blue = up). (Reprinted from Dunin-Borkowski et al., 2004b.) (See color plate.)



2002). A variety of self-assembled structures is visible, including five- and six-particle rings, chains, and closely packed aggregates. The particles are each encapsulated in a 3- to 4-nm oxide shell. Figure 18–9b–d shows magnetic FC states in four different Co particle rings, measured using electron holography at room temperature in zero-field conditions (Tripp et al., 2003). The magnetic flux lines, which are formed from the cosine of 128 times the magnetic contribution to the measured phase shift, reveal the in-plane induction within each ring ensemble. Further electron holography experiments show that the chirality of the FC states can be switched *in situ* in the TEM by using an out-of-plane magnetic field.

3.3.3 FeNi Nanoparticle Chains

The magnetic properties of nanoparticle chains have been studied for many years (e.g., Jacobs and Bean, 1955). However, there are few experimental measurements of the critical sizes at which individual particles that are arranged in chains are large enough to support magnetic vortices rather than single domains. Previous electron holography studies of magnetic nanoparticle chains (e.g., Seraphin et al., 1999; Signoretti et al., 2003) have never provided direct images of such vortex states. Here, we illustrate the use of electron holography to characterize chains of ferromagnetic FeNi crystals, whose average diameter of 50 nm is expected to be close to the critical size for vortex formation (Hýtch et al., 2003). Figure 18–10a shows a chemical map of a chain of $\text{Fe}_{0.56}\text{Ni}_{0.44}$ nanoparticles, acquired using a Gatan imaging filter. The particles are each coated in a 3-nm oxide shell. A defocused bright-field image and a corresponding electron hologram from part of a chain are shown in Figure 18–10b and c, respectively. The mean inner potential contribution to the phase shift was again determined by using the field of the microscope objective lens to magnetize each chain parallel and then antiparallel to its length. The external field was removed before finally recording holograms in field-free conditions. Figure 18–11a and b shows the remanent magnetic states of two chains of $\text{Fe}_{0.56}\text{Ni}_{0.44}$ particles, measured using electron holography. For a 75-nm $\text{Fe}_{0.56}\text{Ni}_{0.44}$ particle sandwiched between two smaller particles (Figure 18–11a), closely spaced contours run along the chain in a channel of width 22 ± 4 nm. A comparison of the result with micromagnetic simulations (Hýtch et al., 2003) indicates that the particle contains a vortex with its axis parallel to the chain axis, as shown schematically in Figure 18–11c. In Figure 18–11b, a vortex can be seen end-on in a 71-nm particle at the end of a chain. The positions of the particle's neighbors determine the handedness of the vortex, with the flux channel from the rest of the chain sweeping around the core to form concentric circles (Figure 18–11d). The vortex core, which is now perpendicular to the chain axis, is only 9 ± 2 nm in diameter. The larger value of 22 nm observed in Figure 18–11a results from magnetostatic interactions along the chain. Vortices were never observed in particles below 30 nm in size, while intermediate states were observed in 30- to 70-nm particles. Similar particles with an alloy concentration of $\text{Fe}_{0.10}\text{Ni}_{0.90}$ contain wider flux channels of diameter ~ 70 nm, and single domain states when the par-

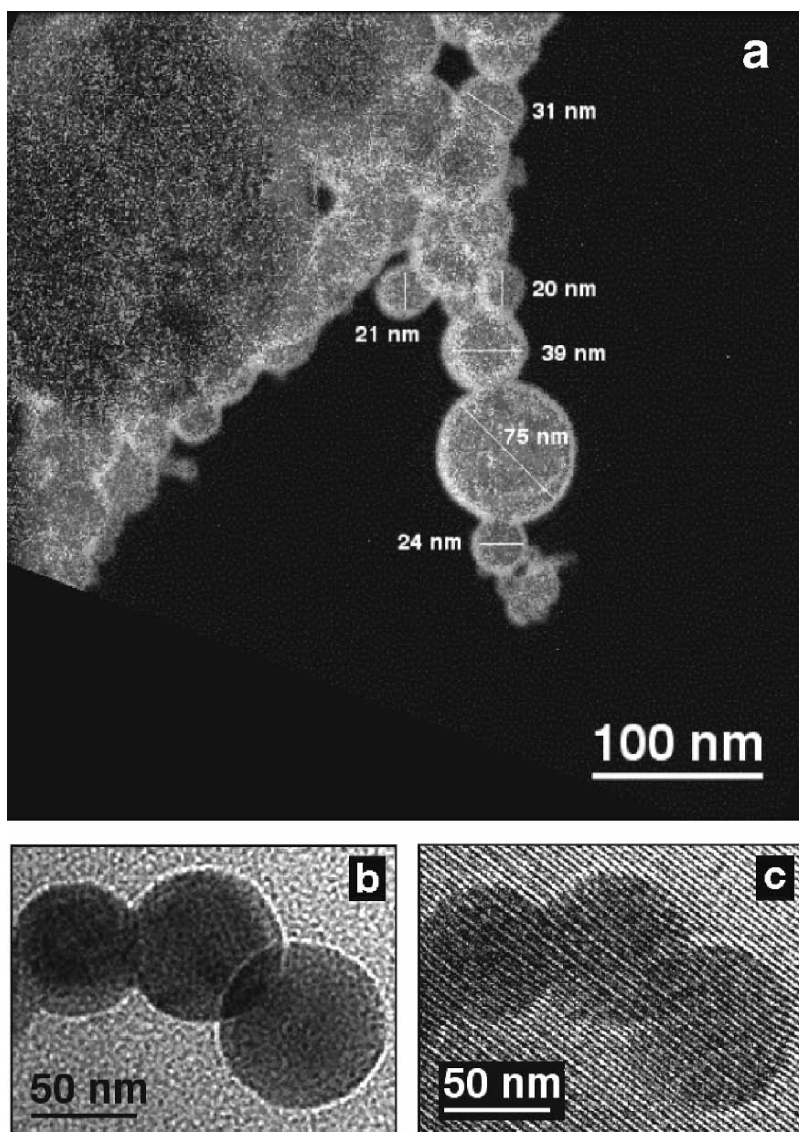


Figure 18–10. (a) Chemical map of $\text{Fe}_{0.56}\text{Ni}_{0.44}$ nanoparticles, obtained using three-window background-subtracted elemental mapping with a Gatan imaging filter, showing Fe (red), Ni (blue), and O (green). (b) Bright-field image and (c) electron hologram of the end of a chain of $\text{Fe}_{0.56}\text{Ni}_{0.44}$ particles. The hologram was recorded using an interference fringe spacing of 2.6 nm. (Reprinted from Dunin-Borkowski et al., 2004b.) (See color plate.)

ticles are above $\sim 100\text{ nm}$ in size. The complexity of such vortex states highlights the importance of controlling the shapes, sizes, and positions of closely spaced magnetic nanocrystals for applications in magnetic storage devices.

3.3.4 Planar Arrays of Magnetite Nanoparticles

The magnetic behavior of the chains and rings of nanomagnets described above contrasts with that of a regular two-dimensional array

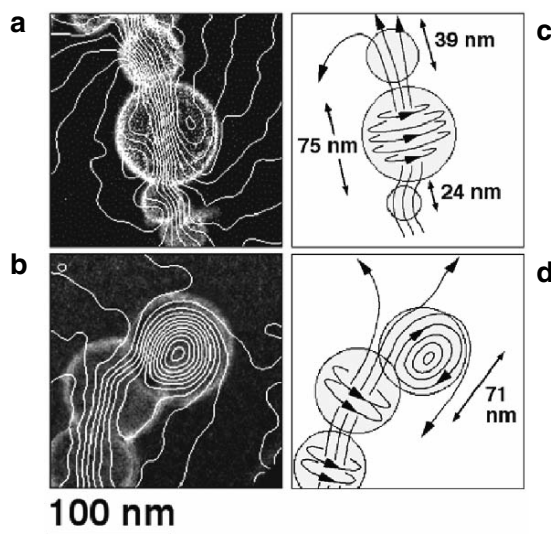


Figure 18-11. (a and b) Experimental phase contours showing the strength of the local magnetic induction (integrated in the electron beam direction) in two different chains of $\text{Fe}_{0.56}\text{Ni}_{0.44}$ particles, recorded with the electron microscope objective lens switched off. The particle diameters are (a) 75 nm between two smaller particles and (b) 71 nm at the end of a chain. Contours, whose spacings are 0.083 and 0.2 radians for images (a) and (b), respectively, have been overlaid onto oxygen maps of the particles recorded using a Gatan imaging filter. The mean inner potential contribution to the measured phase shift has been removed from each image. (c and d) Schematic representations of the magnetic microstructure in the chains. Magnetic vortices spinning about the chain axis are visible in (c) and (d). A vortex spinning perpendicular to the chain axis is also visible in (d). (Reprinted from Dunin-Borkowski et al., 2004b.)

of closely spaced crystals. Figure 18-12 shows chemical maps of a crystalline region of a naturally occurring magnetite-ulvöspinel ($\text{Fe}_3\text{O}_4\text{-Fe}_2\text{TiO}_4$) mineral specimen, which has exsolved during slow cooling to yield an intergrowth of magnetite-rich blocks separated by nonmagnetic ulvöspinel-rich lamellae (Price, 1981). The Fe and Ti chemical maps shown in Figure 18-12 were obtained using three-window background-subtracted elemental mapping with a Gatan imaging filter. Exsolution lamellae subdivide the grain into a fairly regular array of magnetite-rich blocks. The specimen thickness increases from 70 nm at the top of the region to 195 nm at the bottom. The magnetite blocks are, therefore, roughly equidimensional. Remanent magnetic states were recorded by tilting the specimen in zero field and then turning the objective lens on fully to saturate the sample, to provide a known starting point from which further fields could be applied. The objective lens was then turned off, the specimen tilted in zero field in the opposite direction, and the objective lens was excited partially to apply a known in-plane field component to the specimen in the opposite direction. The objective lens was switched off and the sample tilted back to 0° in zero field to record each hologram. This

procedure was repeated for a number of different applied fields (Harrison et al., 2002). Mean inner potential contributions to the measured phase shifts were removed using a procedure different from that used for the chains and rings of nanoparticles described above. Although both thickness and composition vary in the magnetite-ulvöspinel specimen, the different compositions of magnetite and ulvöspinel are compensated by their densities in such a way that their mean inner potentials are exactly equal. As a result, only a thickness correction is required. The local specimen thickness across the region of interest was determined in units of inelastic mean free path by using energy-filtered imaging. This thickness measurement was then used to determine the mean inner potential contribution to the phase shift, which was in turn used to establish the magnetic contribution to the phase. Figure 18–13 shows eight of the resulting remanent magnetic states recorded after applying the in-plane fields indicated. The black contour lines provide the direction and magnitude of the magnetic induction in the plane of the sample, which can be correlated with the positions of the magnetite blocks (outlined in white). The direction of the induction is also indicated using colors and arrows, according to the color wheel shown at the bottom of the figure. Figure 18–13 shows that the magnetic domain structure in this sample is extremely complex. In Figure 18–13, the smallest block observed to form a vortex is larger than the predicted minimum size of 70 nm for vortices to form in *isolated* cubes of magnetite. The abundance of single domain states implies that they have lower energy than vortex states in the presence of strong interactions. The demagnetizing energy, which normally destabilizes the single domain state with respect to the vortex state in isolated particles, is greatly reduced in an array of strongly interacting particles.

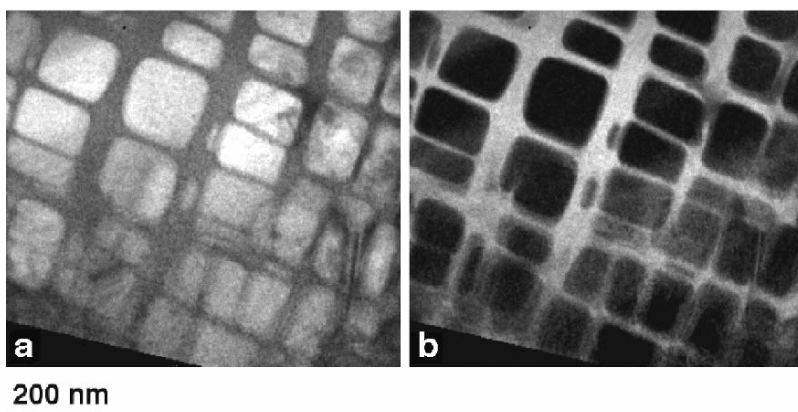
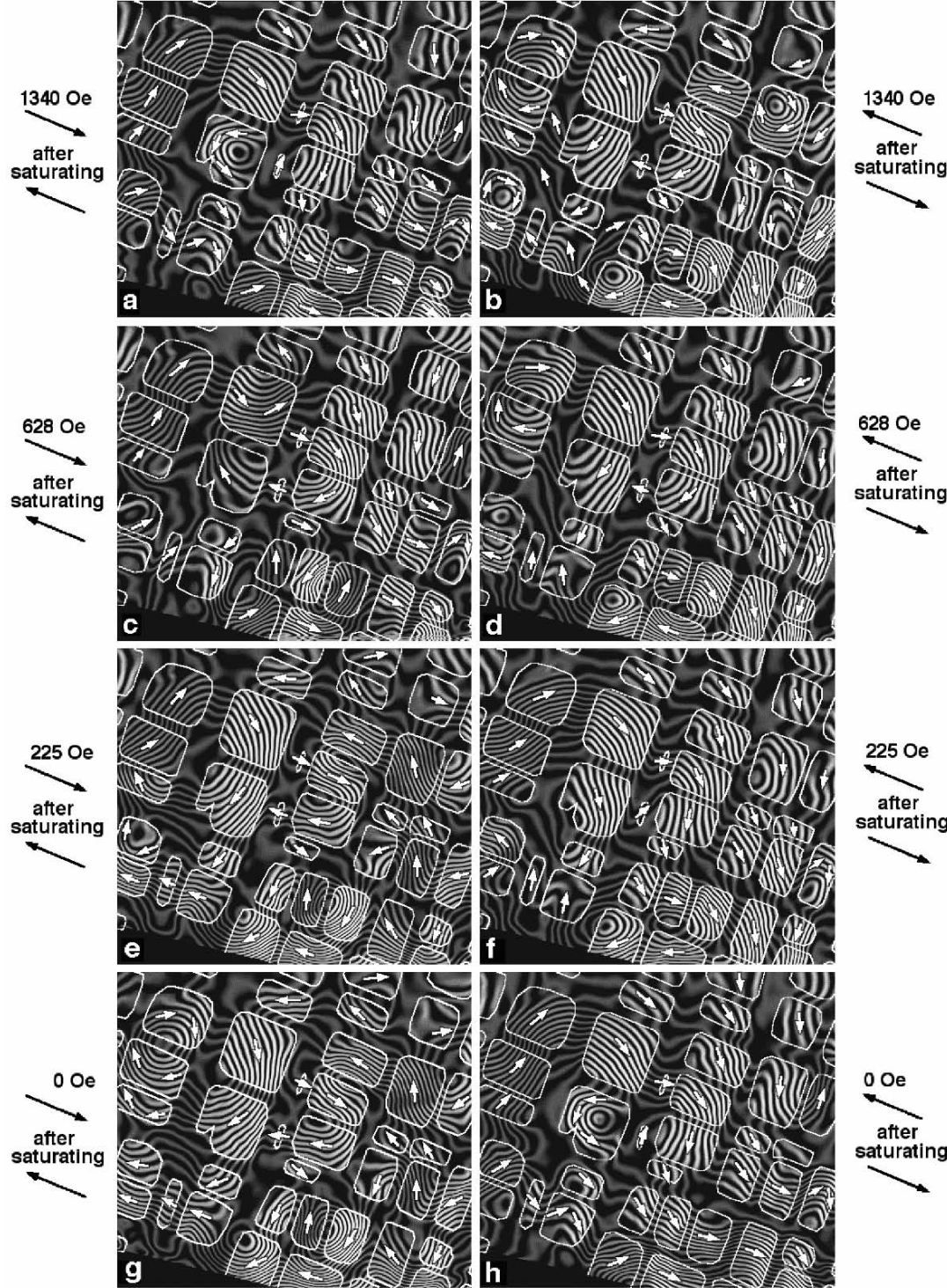


Figure 18–12. Three-window background-subtracted elemental maps acquired from a naturally occurring titanomagnetite sample with a Gatan imaging filter using (a) the Fe L edge and (b) the Ti L edge. Brighter contrast indicates a higher concentration of Fe and Ti in (a) and (b), respectively. (Reprinted from Dunin-Borkowski et al., 2004b.)



3.3.5 Lithographically Patterned Magnetic Nanostructures

Specimen preparation presents a challenge for many samples of interest that contain nanostructured magnetic materials. One example is provided by nanomagnet arrays that have been fabricated directly onto an Si substrate using interferometric lithography (Ross, 2001). Figure 18–14a shows a scanning electron microscope image of nominally 100-nm-diameter 20-nm-thick Co dots fabricated on Si in a square array of side 200 nm. The dots were prepared for TEM examination using focused ion beam milling in plan-view geometry, by micromachining a trench from the substrate side of the specimen to leave a free-standing $10 \times 12\text{-}\mu\text{m}$ membrane of crystalline Si, which was approximately 100 nm in thickness and contained over 3000 Co dots. Figure 18–14b shows an off-axis electron hologram recorded from part of the electron-transparent membrane containing the dots. The specimen was tilted slightly away from zone axis orientations of the underlying Si substrate to minimize diffraction contrast. The specimen edge is toward the bottom left of the figure (Dunin-Borkowski et al., 2001). Figure 18–14c and d shows contours of spacing $0.033 \approx \pi/94$ radians that have been added to the (slightly smoothed) magnetic contribution to the holographic phase, for two different remanent magnetic states of the Co dots. In Figure 18–14c, which was recorded after saturating the dots upward and then removing the external field, the dots are oriented magnetically in the direction of the applied field. In contrast, in Figure 18–14d, which was formed by saturating the dots upward, applying a 382 Oe downward field and then removing the external field, the dots are magnetized in a range of directions. The experiments show that the dots are sometimes magnetized out of the plane (e.g., at the bottom left of Figure 18–14d). The measured saturation magnetizations are smaller than expected for pure Co, possibly because of oxidation or damage sustained during specimen preparation. Similar electrodeposited 57-nm-diameter 200-nm-high Ni pillars arranged in square arrays of side 100 nm, which were prepared for TEM examination using focused ion beam milling in a cross-sectional geometry, have also been examined. Despite their shape, not all of the Ni pillars were magnetized parallel to their long axes. Instead, they interacted with each other strongly, with two, three, or more adjacent pillars combining to form vortices.



Figure 18–13. Magnetic phase contours from the region shown in Figure 8–12, measured using electron holography. Each image was acquired with the specimen in magnetic field-free conditions. The outlines of the magnetite-rich regions are marked in white, while the direction of the measured magnetic induction is indicated both using arrows and according to the color wheel shown at the bottom of the figure (red = right, yellow = down, green = left, blue = up). Images (a), (c), (e), and (g) were obtained after applying a large ($>10,000\text{ Oe}$) field toward the top left, then the indicated field toward the bottom right, after which the external magnetic field was removed for hologram acquisition. Images (b), (d), (f), and (h) were obtained after applying identical fields in the opposite directions. (Reprinted from Harrison et al., 2002.) (See color plate.)

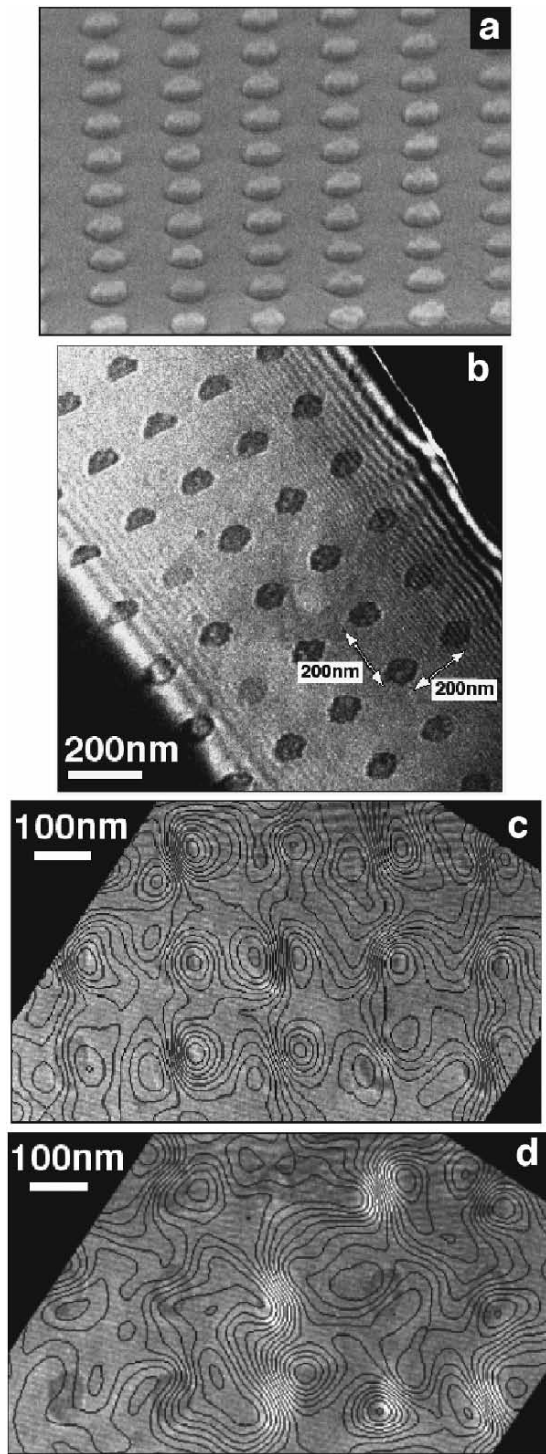


Figure 18–14. (a) Scanning electron microscope image of 100-nm-diameter 20-nm-thick Co dots fabricated on Si in a square array of side 200 nm using interferometric lithography. (b) Off-axis electron hologram of part of an electron-transparent membrane containing the dots, prepared using focused ion beam milling. The hologram was acquired at 200 kV using a Philips CM200 FEGTEM, a biprism voltage of 160 V, a holographic interference fringes spacing of 3.05 nm, and an overlap width of 1.04 μm . (c and d) Magnetic contributions to the measured electron holographic phase shift for two remanent magnetic states. The contour spacing is 0.033 radians: (c) was formed by saturating the dots upward and then removing the external field; (d) was formed by saturating the dots upward, applying a 382 Oe downward field, and then removing the external field. (Reprinted from Dunin-Borkowski et al., 2001.)

Results similar to those shown in Figure 18–14 have been obtained from a wide range of larger lithographically patterned structures, many of which show multidomain behavior (Dunin-Borkowski et al., 2000; Hu et al., 2005; Heumann et al., 2005). Few phase contours are visible outside such elements when they support magnetic flux closure states. Electron holography has also been used to provide information about magnetic interactions between closely separated ferromagnetic layers within individual Co/Au/Ni spin-valve elements (Smith et al., 2000). The presence of two different contour spacings at different applied fields in such elements is associated with the reversal of the magnetization direction of the Ni layer in each element before the external field is reduced to zero as a result of flux closure associated with the strong fringing field of the magnetically more massive and closely adjacent Co layer.

3.3.6 Co Nanowires

An important question relates to the minimum size of a nanostructure in which magnetic fields can be characterized successfully using electron holography. This point is now addressed through the characterization of 4-nm-diameter single crystalline Co nanowires (Snoeck et al., 2003). The difficulty of this measurement results from the fact that the mean inner potential contribution to the phase shift at the center of a 4-nm wire relative to that in vacuum is 0.57 radians (assuming a value for V_0 of 22 V), whereas the step in the magnetic contribution to the phase shift across the wire is only 0.032 radians (assuming a value for B of 1.6 T). Figure 18–15a shows a bright-field TEM image of a bundle of 4-nm-diameter Co wires, which are each between a few hundred nanometers and several hundred micrometers in length. Magnetic contributions to the phase shift were obtained by recording two holograms from each area of interest, where the wires were magnetized parallel and then antiparallel to their length by tilting the sample by $\pm 30^\circ$ about an axis perpendicular to the wire axis and using the conventional microscope objective lens to apply a large in-plane field to the specimen. The lens was then switched off and the sample returned to zero tilt to record each electron hologram. This procedure relies on the ability to reverse the magnetization in the sample exactly, which is a good assumption for these narrow and highly anisotropic wires. Figure 18–15b shows the magnetic contribution to the measured phase shift for an isolated wire, in the form of contours that are spaced 0.005 radians apart. The contours have been overlaid onto an image showing the mean inner potential contribution to the phase shift, so that they can be correlated with the position of the wire. The magnetic signal is weak and noisy, and was smoothed before forming the contours. The closely spaced contours along the length of the wire confirm that it is magnetized along its axis. The fact that they are not straight is intriguing. However, this effect may result from smoothing of the signal, which is noisy and weak. Figure 18–16a shows a montage of three holograms obtained close to the end of a bundle of Co wires, which was magnetized approximately parallel to its length. The magnetic contribution to the measured phase shift is shown in Figure

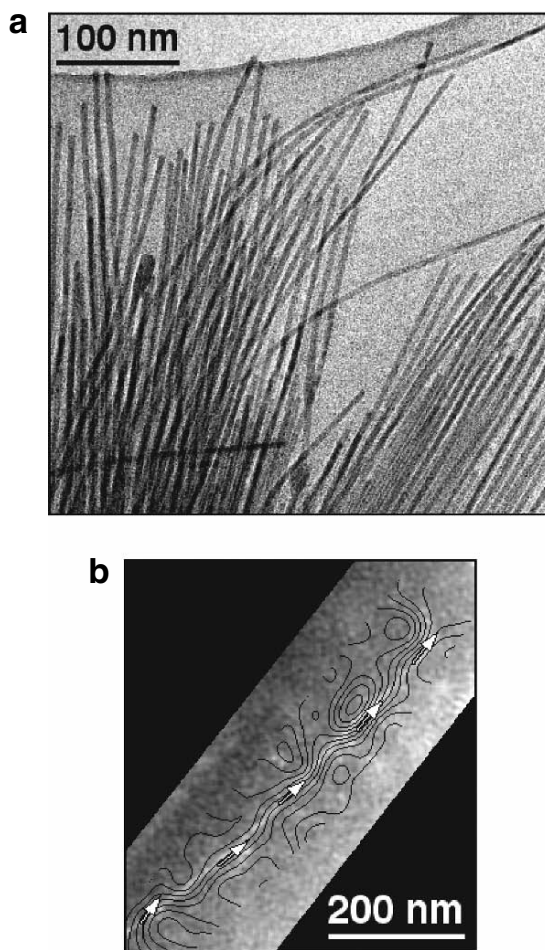


Figure 18-15. (a) Bright-field image of the end of a bundle of Co nanowires adjacent to a hole in a carbon support film. (b) Contours (0.005 radian spacing) generated from the magnetic contribution to the phase shift for a single isolated Co nanowire, superimposed onto the mean inner potential contribution to the measured phase shift. (Reprinted from Snoeck et al., 2003.)

18-16b in the form of contours, which are spaced 0.25 radians apart. The wires channel the magnetic flux efficiently along their length, and they fan out as the field decreases in strength at the end of the bundle. Although the signal from the bundle appears overall to obscure that from individual wires and junctions, these details can be recovered by increasing the density of the contours (Snoeck et al., 2003). The slight asymmetry between the contours on either side of the bundle in Figure 18-16b may result from the fact that the reference wave is affected by the magnetic leakage field of the bundle, which acts collectively as though it were a single wire of larger diameter. The step in phase across the bundle is (9.0 ± 0.2) radians, which is consistent with the presence of (280 ± 7) ferromagnetically coupled wires.

3.3.7 Cross-Sectional Specimens

One of the most challenging problems for electron holography of magnetic materials is the quantitative measurement of the magnetic properties of nanometer-scale magnetic layers when examined in cross section. The primary difficulty is the presence of rapid and unknown variations in both the composition and the thickness of the specimen, from which the weak magnetic signal must be separated. In a cross-sectional sample, the effects of variations in specimen thickness on the measurements cannot be eliminated by using the normalized amplitude of the hologram [Eq. (14)], both because the mean free path in each material in such a cross-sectional specimen is usually unknown

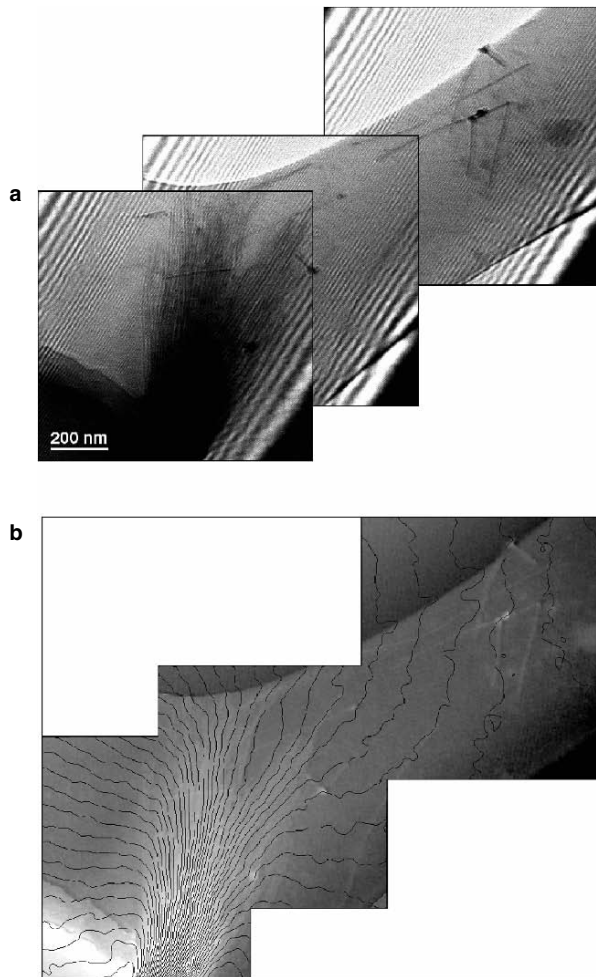


Figure 18–16. (a) Montage of three electron holograms acquired from the end of a bundle of Co nanowires. The biprism voltage is 210 V, the acquisition time for each hologram 16 s, the holographic interference fringe spacing is 3.9 nm, and the holographic overlap width is 1160 nm. No objective aperture was used. (b) Magnetic remanent state, displayed in the form of contours (0.25 radian spacing), generated from the measured magnetic contribution to the electron holographic phase shift after saturating the wires in the direction of the axis of the bundle. The contours are superimposed onto the mean inner potential contribution to the phase shift. (Reprinted from Snoeck et al., 2003.)

and because the amplitude image is in general noisy and may contain strong contributions from diffraction and Fresnel contrast. However, by rearranging Eqs. (8) and (16), it can be shown that specimen thickness effects may be removed by plotting the difference in the phase gradient between images in which the magnetization has reversed divided by the average of their phases, multiplied by a constant and by the value of the mean inner potential of each magnetic layer separately. Formally, this procedure can be written

$$\left(\frac{C_E \hbar V_0(x, y)}{e} \right) \left\{ \frac{\Delta[d\phi(x, y)/dx]}{\langle \phi(x, y) \rangle} \right\} \frac{\Delta B_{\perp}(x, y)}{\{1 - [e/C_E \hbar V_0(x, y)]\} [\langle \int B_{\perp}(x, y) t(x, y) dx \rangle / t(x, y)]} \quad (23)$$

According to Eq. (23), by combining phase profiles and their gradients (evaluated in a direction perpendicular to the layers) from successive holograms with the magnetization direction reversed, the specimen thickness profile can be eliminated and the magnetic induction in each layer determined quantitatively. Both the magnitude and the sign of $\Delta B_{\perp}(x, y) = 2B_{\perp}(x, y)$ are obtained exactly using Eq. (23) if the magnetization reverses exactly everywhere in the sample. (The denominator on the right-hand side of the equation is then unity.) Furthermore, nonzero values are returned only in regions where the magnetization has changed. Figure 18–17 illustrates the application of Eq. (23) to a cross-sectional magnetic tunnel junction that contains a layer sequence of 22 nm Co/4 nm HfO₂/36 nm CoFe on an Si substrate (McCartney and

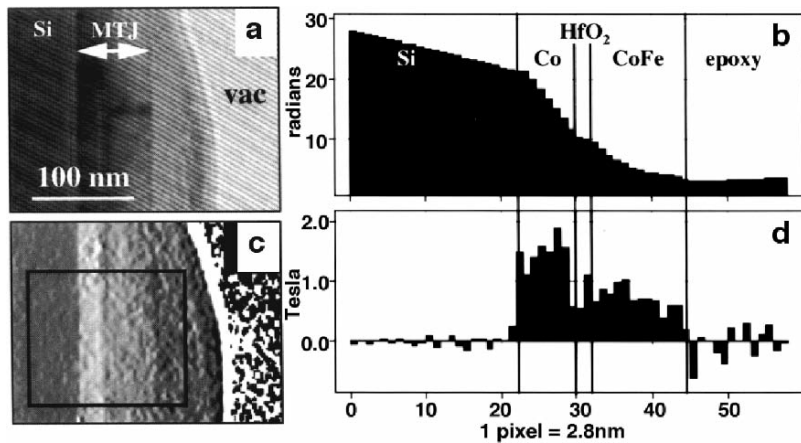


Figure 18–17. (a) Off-axis electron hologram obtained from a magnetic tunnel junction containing a 4-nm HfO₂ tunnel barrier. (b) Measured phase profile across the layers in the tunnel junction structure. (c) Image formed by recording two holograms with opposite directions of magnetization in the specimen, and subsequently taking the difference between the recorded phase gradients (calculated in a direction perpendicular to the layers) and dividing by the average of the two phases. (d) Measured magnetic induction in the tunnel junction sample, generated by multiplying a line profile obtained from image (c) by a constant (see text for details), with the vertical scale now plotted in units of Tesla. (Reprinted from McCartney and Dunin-Borkowski, 1998.)

Dunin-Borkowski, 1998). Two holograms were obtained, similar to that shown in Figure 18–17a, between which the magnetization directions of the Co and CoFe layers in the specimen were reversed *in situ* in the electron microscope. Figure 18–17b shows an unwrapped phase profile obtained from the hologram in Figure 18–17a by taking a line profile in the direction perpendicular to the layers. Phase profiles from the two holograms appeared almost identical irrespective of the direction of magnetization. The application of Eq. (23) to the two phase images results in the image shown in Figure 18–17c. The line profile in Figure 18–17d was obtained by averaging Figure 18–17c parallel to the direction of the layers. As predicted, Figure 18–17d, which should by now be independent of variations in composition and specimen thickness, is nonzero only in the magnetic layers and yields a value for the magnetic induction in the Co layer of 1.5 T (assuming a mean inner potential of 25 V). In a similar experiment, holograms of $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ have recently been acquired both above and below the Curie temperature of the material to remove specimen thickness and mean inner potential contributions from the measured phase (Loudon et al., 2003).

3.4 Quantitative Measurements, Micromagnetic Simulations, and Resolution

A particular strength of electron holography is its ability to provide quantitative information about magnetic properties. For example, the magnetic moment of a nanoparticle can be obtained from the relation

$$m_x = \left(\frac{\hbar}{e}\right) \int_{y=-\infty}^{y=+\infty} \int_{x=-\infty}^{x=+\infty} \frac{\partial}{\partial y} \phi_{\text{mag}}(x, y) dx dy \quad (24)$$

where ϕ_{mag} is the magnetic contribution to the phase shift and y is a direction perpendicular to x in the plane of the specimen. According to Eq. (24), the magnetic moment in a given direction can be obtained by measuring the area under the first differential of ϕ_{mag} evaluated in the perpendicular direction. The contribution of stray magnetic fields to the moment is included in this calculation if the integration is carried out over a large enough distance from the particle.

The need to compare electron holographic measurements with micromagnetic simulations results from the sensitivity of the magnetic domain structure in nanoscale materials and devices to their detailed magnetic history. Differences in the starting magnetic states on a scale that is too small to be distinguished visually, as well as interelement coupling and the presence of out-of-plane magnetic fields, are all important for the formation of subsequent domain states, and, in particular, to the sense (the handedness) with which magnetic vortices unroll (Dunin-Borkowski et al., 1999). The sensitivity of the domain structure to such effects emphasizes the need to correlate high quality experimental holographic measurements with micromagnetic simulations.

The spatial resolution that can be achieved in phase images is determined primarily by the spacing of the holographic interference fringes.

However, the contrast of these fringes decreases as their spacing is reduced, and the recording process is also dominated by Poisson-distributed shot noise (Lichte et al., 1987). These parameters are affected by the illumination diameter, exposure time, and biprism voltage. The final “phase resolution” (Harscher and Lichte, 1996) and “spatial resolution” are always inherently linked, in the sense that a small phase shift can be measured with high precision and poor spatial resolution, or with low precision but high spatial resolution. In each of the examples described above, the recorded phase images were always smoothed slightly to remove noise, and the spatial resolution of the magnetic information was estimated typically to be between 10 and 20 nm. This procedure is necessarily subjective, and great care is required to ensure that artifacts are not introduced. Higher spatial and phase resolution might possibly also be achieved by recording several holograms of each area of interest and subsequently averaging the resulting phase images.

4 Measurement of Electrostatic Fields

In this section, the application of electron holography to the characterization of electrostatic fields is reviewed. Initial examples are taken from the characterization of electrostatic fringing fields outside electrically biased nanowires. The challenges that are associated with imaging dopant contrast at depletion layers in semiconductors are then described, before discussing the characterization of interfaces at which both charge redistribution and changes in chemistry are possible.

4.1 Field-Emitting Carbon Nanotubes

Early experiments on tungsten microtips demonstrated that electron holography could be used to measure electrostatic fringing fields in biased samples (Matteucci et al., 1992). Further studies were made on pairs of parallel 1- μm -diameter Pt wires held at different potentials (Matteucci et al., 1988) and on single conducting wires (Kawasaki et al., 1993), and simulations were presented for electrostatic phase plates (Matsumoto and Tonomura, 1996). A more recent example involves the use of electron holography to map the electrostatic potential around the end of an electrically biased multiwalled carbon nanotube (Cumings et al., 2002). Nanotubes were mounted on a three-axis manipulation electrode using conducting epoxy and positioned approximately 6 μm from a gold electrode, as shown in Figure 18–18a. Depending on the applied bias, electrons were emitted from the nanotube. The left hand column of Figure 18–18b shows contoured phase images recorded before a bias V_b was applied to the specimen, and for a bias above the threshold for field emission (approximately 70 V). The upper phase shift map ($V_b = 0$) is featureless around the nanotube, whereas the lower map ($V_b = 120$ V) shows closely spaced 2π phase contours. The right hand column in Figure 18–18b shows the corresponding phase gradient for each image. When $V_b = 0$, the phase gradient is featureless around the nanotube, whereas it is concentrated around the nanotube

tip when $V_b = 120$ V. The images shown in Figure 18–18b were interpreted by comparison with simulations, calculated on the assumption that the nanotube could be approximated by a line charge, where the charge distribution was varied until a close fit to the data was found. The fit to the 120 V phase data in Figure 18–18b provided a value of 1.22 V/nm for the electric field at the nanotube tip. This field was stable over time, even when the emission current varied.

4.2 Dopant Potentials in Semiconductors

One of the most elusive yet tantalizing challenges for electron holography has been the quest for a reliable, quantitative approach to the characterization of electrostatic potentials associated with charge redistribution at depletion regions in doped semiconductors. Attempts to tackle this problem have been made since the 1960s using many forms of electron interferometry, both experimentally (e.g., Titchmarsh et al., 1969; Frabboni et al., 1987) and theoretically (e.g., Pozzi and Vanzi, 1982; Beleggia et al., 2000). It is now recognized that TEM specimen preparation can have a profound effect on the contrast visible in holographic phase images of doped semiconductors, either because of physical damage to the specimen surface or because of the implantation of dopant ions such as Ar or Ga during ion milling. An electrically inactive surface layer and/or a doped layer, with a thickness depending on the specimen preparation method, may form at the sample surface. In addition, the specimen may charge up during observation, to such an extent that all dopant contrast is lost. The effects of specimen

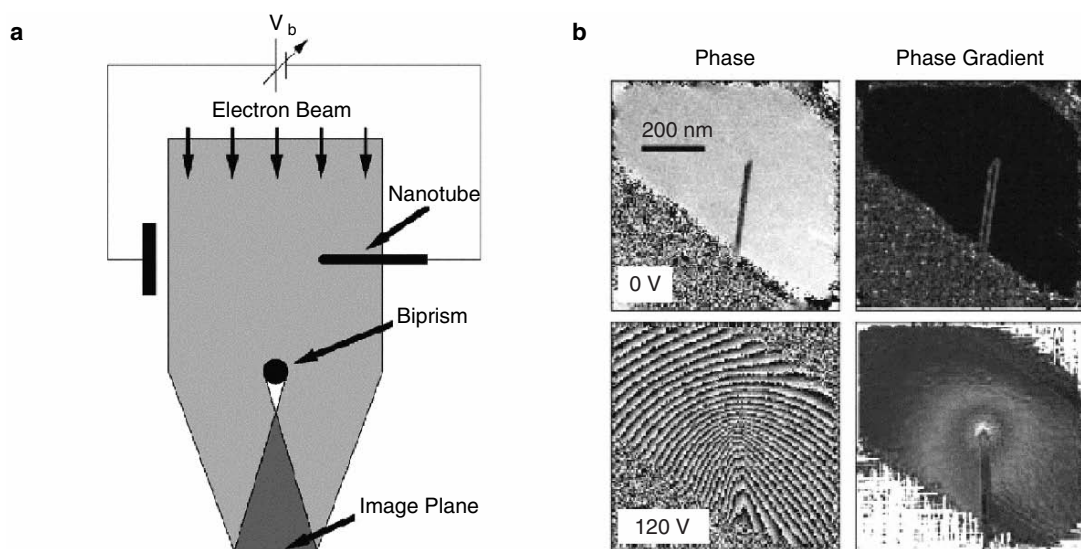


Figure 18–18. (a) Schematic diagram of the experimental set-up used to record electron holograms of field-emitting carbon nanotubes. (b) Phase shift and phase gradient maps determined from electron holograms of a single multiwalled carbon nanotube at bias voltages of 0 and 120 V. The phase gradient indicates where the electric field is strongest. Note the concentration of the electric field at the nanotube tip when the bias voltage is 120 V. (Reprinted from Cumings et al., 2002.)

preparation, and in particular the electrical state of near-surface regions, most likely account for many of the anomalous results in early experiments. Recent studies indicate that it may be possible to resolve these problems.

The first unequivocal demonstration of two-dimensional mapping of the electrostatic potential in an unbiased doped semiconductor using electron holography was achieved for metal–oxide–semiconductor (MOS) Si transistors (Rau et al., 1999). The source and drain regions were visible in phase images with a spatial resolution of 10 nm and an energy resolution of close to 0.10 eV. Differential thinning was discounted as a cause of the observed phase shifts, and an optimal specimen thickness of 200–400 nm was identified for such experiments. The transistors were prepared for TEM examination using conventional mechanical polishing and Ar ion milling. A 25-nm-thick electrically altered layer was identified on each surface of the specimen, which resulted in measured built-in voltages of 0.9 ± 0.1 V across each p – n junction, which was lower than the value of 1.0 V predicted for the specified dopant concentration.

More recently, electron holography studies of transistors have been compared with process simulations (Gribelyuk et al., 2002). Figure 18–19a shows a contoured image of the electrostatic potential associated with a 0.35- μ m Si device inferred from an electron hologram, where the contours correspond to potential steps of 0.1 V. The B-doped source and drain regions are delineated clearly. In this study, the specimen was prepared primarily using tripod wedge-polishing, followed by limited low-angle Ar ion milling at 3.5 kV. Significantly, no electrically dead surface layer had to be taken into account to quantify the results. Figure 18–19b and c shows a comparison between line profiles obtained from Figure 18–19a and simulations, both laterally across the junction and with depth from the Si surface. Simulations for “scaled loss” and “empirical loss” models, which account for B-implant segregation into the adjacent oxide and nitride layers, are shown. The scaled loss model, which leads to stronger B diffusion, assumes uniform B loss across the device structure, whereas the empirical loss model assumes segregation of the implanted B at the surfaces of the source and drain regions. In both Figure 18–19b and Figure 18–19c, the empirical loss model provides a closer match to the experimental results. Figure 18–19d shows a simulated electrostatic potential map for the same device based on the “empirical loss” model, which matches closely with the experimental image in Figure 18–19a. Overall, this study demonstrated successful mapping of the electrostatic potential in 0.13- μ m and 0.35- μ m device structures with a spatial resolution of 6 nm and a sensitivity of 0.17 eV.

In early applications of electron holography to dopant delineation, which were carried out on chemically thinned Si samples under conditions of reverse bias (e.g., Frabboni et al., 1985), differences between phase images recorded at different bias voltages were used to visualize external electrostatic fringing fields close to the positions of p – n junctions. Electrostatic potential profiles have recently been measured for reverse-biased Si p – n junctions that were prepared for TEM

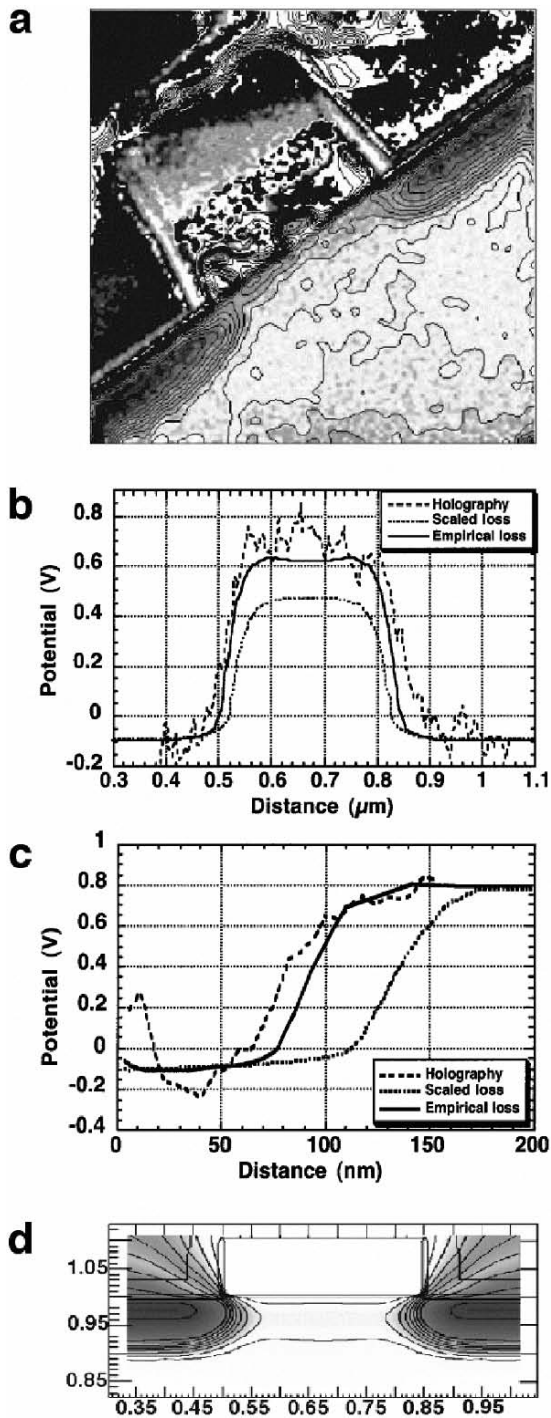


Figure 18–19. (a) Reconstructed maps of the electrostatic potential distribution in a 0.35- μm semiconductor device structure, with a contour step of 0.1 V, recorded at an accelerating voltage of 200 kV using a Philips CM200 FEGTEM. (b) Lateral and (c) depth profiles obtained from the image shown in (a). Predictions from process simulations for “scaled loss” and “empirical loss” models are also shown. (d) Two-dimensional simulated map of the potential based on the “empirical loss” model, with a contour step of 0.1 V. The dimensions are in micrometers. (Reprinted from Gribelyuk et al., 2002.)

examination using focused ion beam milling (Twitchett et al., 2002, 2004, 2005). It is significant to note here that focused ion beam milling is currently the technique of choice for preparing TEM specimens from site-specific regions of integrated circuits. It is therefore important to establish whether holography results obtained from unbiased specimens prepared by focused ion beam milling are reliable. It is also useful to develop a specimen geometry that allows electrical currents to be passed through TEM specimens prepared using this technique.

Specimens for *in situ* electrical biasing were prepared by using a 30-kV FEI 200 focused ion beam workstation to machine parallel-sided electron-transparent membranes at the corners of 1×1 -mm 90° cleaved squares of wafer, as shown schematically in Figure 18–20a. This geometry allowed electrical contacts to be made to the front and back surfaces of each specimen using a modified single-tilt holder, as shown in Figure 18–20b. Care was taken to expose the region of interest to the focused beam of Ga ions only at a glancing angle to its surface. Figure 18–20c shows a representative holographic phase image recorded from an unbiased Si p – n junction sample prepared by focused ion beam milling. The crystalline thickness was measured independently to be 550 nm using convergent beam electron diffraction. The p -type and n -type regions are delineated clearly as areas of darker and lighter contrast, respectively. The additional “gray” band at the specimen edge is likely to be associated with the presence of an electrically altered layer, which is visible in cross section but is thought to extend around the entire specimen surface. No electrostatic fringing field is visible outside the specimen, indicating that its surface must be an equipotential. Line profiles across the junction were obtained from phase images acquired with different reverse bias voltages applied to a specimen of 390 nm crystalline thickness (Figure 18–20d), as well as from several unbiased specimens. Each profile in Figure 18–20d is qualitatively consistent with the expected potential profile for a p – n junction in a specimen of uniform thickness. The height of the potential step across the junction, $\Delta\phi$, increases linearly with reverse bias voltage V_{appl} , as shown in Figure 18–20e. This behavior is described by the equation

$$\Delta\phi = C_E(V_{\text{bi}} + V_{\text{appl}})t_{\text{active}} \quad (25)$$

where C_E is defined in Eq. (7) and the p – n junction is contained in an electrically active layer of thickness t_{active} in a specimen of total thickness t . Measurement of the gradient of Figure 18–20e, which is equal to $C_E t_{\text{active}}$, provides a value for t_{active} of 340 ± 10 nm, indicating that 25 ± 5 nm of the crystalline thickness on each surface of the TEM specimen is electrically inactive. The intercept with the vertical axis is $C_E V_{\text{bi}} t_{\text{active}}$, which provides the expected value for the built-in voltage across the junction of 0.9 ± 0.1 V. Depletion widths across the junction measured from the line profiles are higher than expected, suggesting that the electrically active dopant concentration in the specimen is lower than the nominal value. These experiments also show that electrical biasing reactivates some of the dopant that has been passivated by specimen preparation (Dunin-Borkowski et al., 2002). Figure 18–20f shows a four-times-amplified phase image obtained from a 90° cleaved

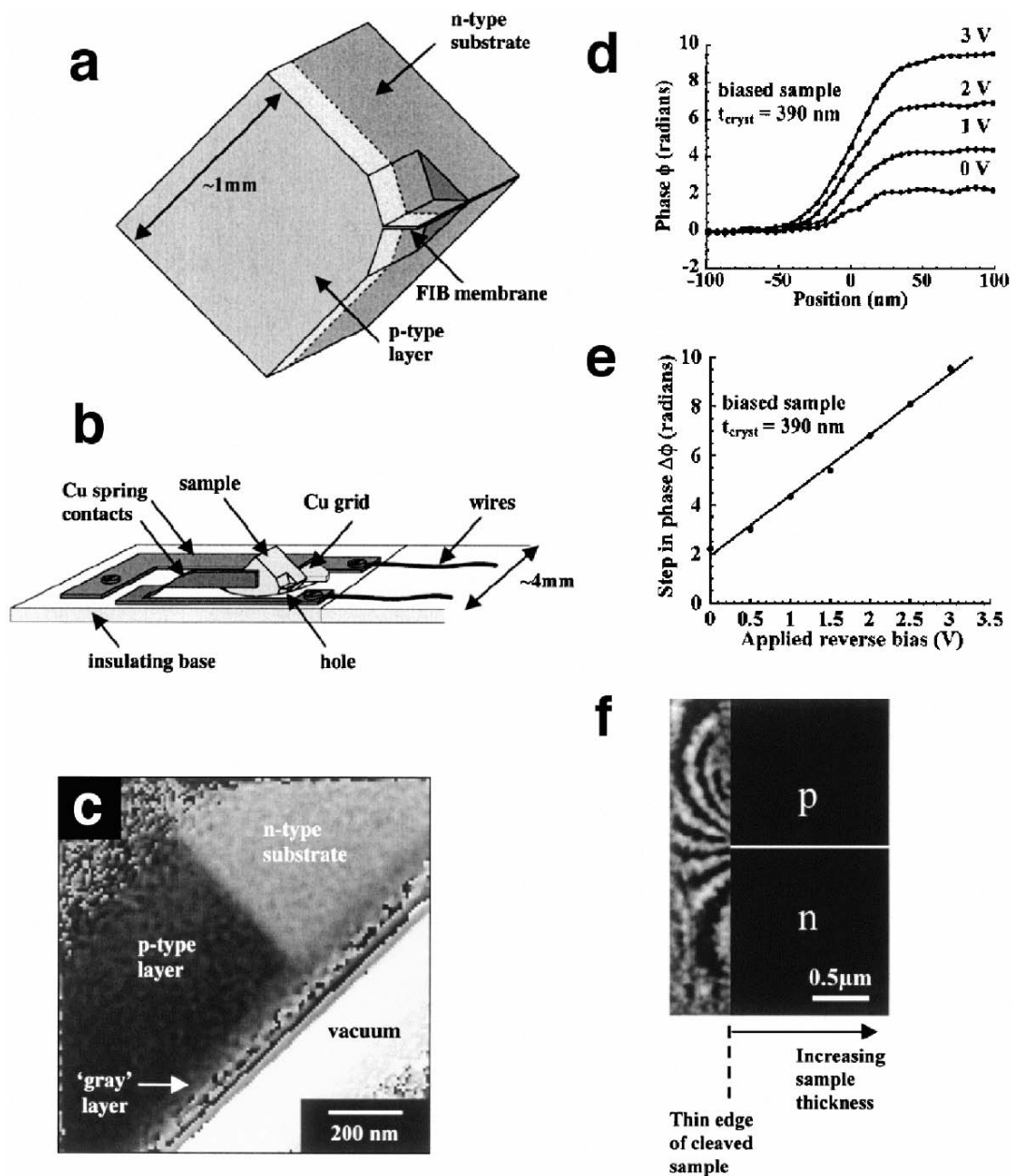


Figure 18–20. (a) Schematic diagram showing the specimen geometry used for applying external voltages to focused ion beam milled semiconductor device specimens containing p - n junctions *in situ* in the TEM. In the diagram, focused ion beam (FIB) milling has been used to machine a membrane of uniform thickness that contains a p - n junction at one corner of a 90° cleaved wedge. (b) Schematic diagram showing the specimen position in a single tilt electrical biasing holder. The specimen is glued to the edge of a Cu grid using conducting epoxy and then clamped between two spring contacts on an insulating base. (c) Reconstructed phase image acquired from an unbiased Si sample containing a p - n junction. Note the “gray” layer running along the edge of the specimen, which is discussed in the text. No attempt has been made to remove the 2π phase “wraps” at the edge of the specimen. (d) Phase shift measured across a p - n junction as a function of reverse bias for a single sample of 390 nm crystalline thickness (measured using convergent beam electron diffraction). (e) The height of the measured step in phase across the junction is shown as a function of reverse bias. (f) Four times-amplified reconstructed phase image, showing the vacuum region outside a p - n junction in a 2-V reverse-biased cleaved wedge sample that had not been focused ion beam milled. (Reprinted from Twitchett et al., 2002.)

wedge that had *not* been prepared by focused ion beam milling, for an applied reverse bias of 2 V, where an external electrostatic fringing field is visible. Such fringing fields were never observed outside *unbiased* cleaved wedges or any focused ion beam milled specimens, indicating that the surfaces of the present TEM specimens prepared by focused ion beam milling are equipotentials under applied bias.

The importance of minimizing and assessing damage, implantation, and specimen thickness variations when examining focused ion beam milled TEM specimens that contain p - n junctions has been highlighted by results from unbiased samples (Wang et al., 2002a–c, 2005). The most elegant of these experiments involved the use of focused ion beam milling to form a 45° specimen thickness profile, from which both the phase change across the junction and the absolute phase shift relative to vacuum on each side of the junction could be plotted as a function of specimen thickness. The slopes of the phase profiles were then used to determine the built-in voltage across the junction, the mean inner potentials on the p and n sides of the junction, and the electrically altered layer thickness. Using this approach, the built-in voltage across a junction with a dopant concentration of approximately 10^{15} cm^{-3} was measured to be $0.71 \pm 0.05 \text{ V}$, while the mean inner potentials of the p and n sides of the junction were measured to be 11.50 ± 0.27 and $12.1 \pm 0.40 \text{ V}$, respectively. The electrically altered layer thickness was measured to be approximately 25 nm on each surface of the specimen.

The electrical nature of the surface of a TEM specimen that contains a doped semiconductor can be assessed by comparing experimental holography results with simulations. Such a comparison, performed using commercial semiconductor process simulation software (Beleggia et al., 2001), suggests that electron beam-induced positive charging of the surface of a TEM specimen, at a level of 10^{13} – 10^{14} cm^{-2} , creates an inversion layer on the p -side of the junction. This layer may explain the absence of electrostatic fringing fields outside the specimen surface, which would otherwise dominate the observed phase contrast (Dunin-Borkowski and Saxton, 1996). Figure 18–21 shows the results of an alternative set of numerical simulations, in which semiclassical equations are used to determine the charge density and potential in a parallel-sided Si sample that contains a p - n junction. The Fermi level on the surface of the specimen is set to a single value to ensure that it is an equipotential (Somodi et al., 2005). The simulations in Figure 18–21 are for symmetrical junctions with dopant concentrations of 10^{18} , 10^{17} , and 10^{16} cm^{-3} . Contours of spacing 0.05 V are shown in each figure. As either the dopant concentration or the specimen thickness decreases, a correspondingly smaller fraction of the specimen retains electrical properties that are close to those of the bulk device. In the simulations, the average step in potential across the junction through the thickness of the specimen, which is insensitive to the surface state energy, is reduced from that in the bulk device. This reduction is greatest for low sample thicknesses and low dopant concentrations. In practice, as a result of additional complications from oxidation, physical damage, and implantation, the simulations shown in Figure 18–21 are likely to

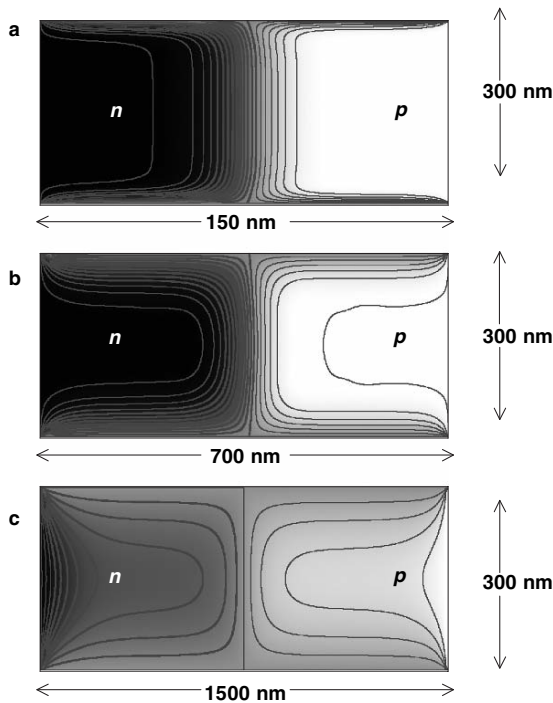


Figure 18–21. Simulations of electrostatic potential distributions in parallel-sided slabs of thickness 300 nm containing abrupt, symmetrical Si p - n junctions formed from (a) 10^{18} , (b) 10^{17} , and (c) 10^{16} cm $^{-3}$ of Sb (n -type) and B (p -type) dopants. The potential at the specimen surfaces is 0.7 eV above the Fermi level, and contours of spacing 0.05 V are shown. The horizontal scale is different in each figure to show the variation in potential close to the position of the junction. The simulations were generated using a two-dimensional rectangular grid. (Reprinted from Somodi et al., 2005.)

be an underestimate of the full modification of the potential from that in the original device.

The ways in which the sample preparation technique of “wedge-polishing” affects both the dead layer thickness and specimen charging have been explored experimentally for a one-dimensional p - n junction in Si by McCartney et al. (2002). A specimen was prepared from a p -type wafer that had been subjected to a shallow B implant and a deeper P implant, resulting in the formation of an n -type well and a p -doped surface region. Phase images were obtained before and after coating one side of the specimen with approximately 40 nm of carbon. Profiles obtained from the uncoated sample showed an initial increase in the measured phase going from vacuum into the specimen, then dropping steeply and becoming negative at large thicknesses. This behavior was not observed after carbon coating, suggesting that it is associated with sample charging that results from the electron beam-induced emission of secondary electrons.

Similar charging effects can be seen directly in two dimensions in Figure 18–22. Figure 18–22a shows a bright-field image of a linear array

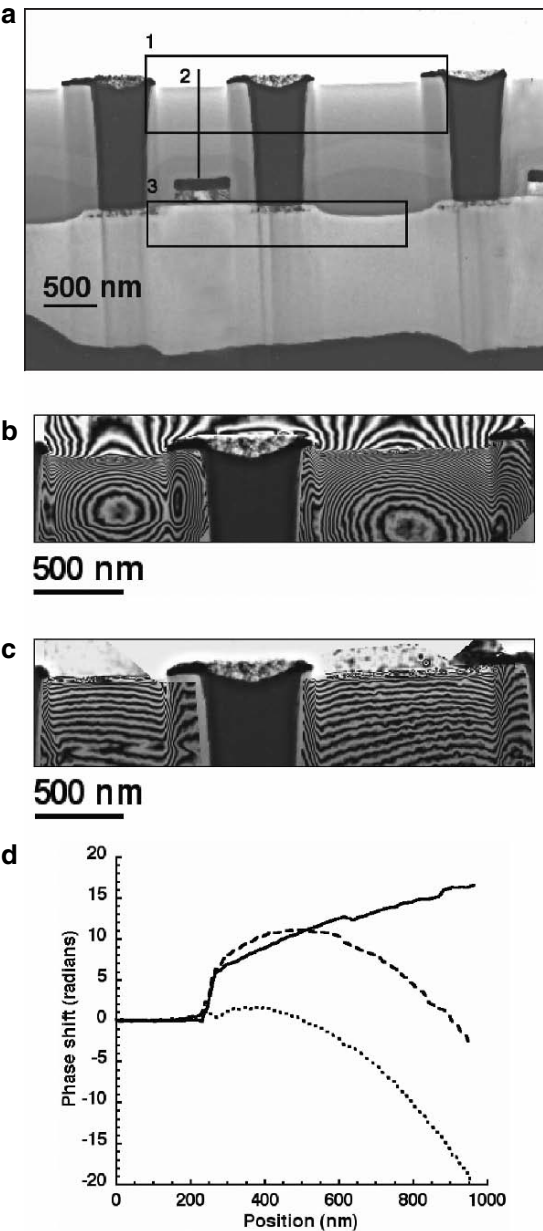


Figure 18–22. Results obtained from a cross-sectional semiconductor device specimen of nominal thickness 400nm prepared using conventional “trench” focused ion beam milling. (a) Bright-field TEM image of a PMOS (0.5- μm gate) transistor, which forms part of a linear array of similar transistors, indicating the locations of the regions analyzed in more detail in the subsequent figures. The dark bands of contrast above the transistors are W contacts. Thickness corrugations are visible in the Si substrate in each image. The gates are formed from W silicide, while the amorphous layers above the gates and between the W plugs are formed from Si oxides that have different densities. (b) Eight times-amplified phase contours were calculated by combining phase images from several holograms obtained across the region marked “1” in (a) using a microscope accelerating voltage of 200kV and a biprism voltage of 160V. Specimen charging results in the presence of electrostatic fringing fields in the vacuum region outside the specimen edge, as well as elliptical phase contours within the Si oxide layers between the W contacts. (c) An equivalent phase image obtained after coating the specimen on one side with approximately 20nm of carbon to remove the effects of charging. The phase contours now follow the expected mean inner potential contribution to the phase shift in the oxide layers, and there is no electrostatic fringing field outside the specimen edge. (d) One-dimensional line profiles obtained from the phase images in (b) and (c) along the line marked “2” in (a). The dashed and solid lines were obtained before and after coating the specimen with carbon, respectively. The dotted line shows the difference between the solid and dashed lines. (Reprinted from Dunin-Borkowski et al., 2005.)

of transistors, which were originally located $\sim 5\mu\text{m}$ below the surface of a wafer and separated from its surface by metallization layers. Such transistors present a significant but representative challenge for TEM specimen preparation for electron holography both because the metallization layers are substantial and can result in thickness corrugations in the doped regions of interest and because these overlayers must, at least in part, be removed to provide a vacuum reference wave for electron holography. An additional difficulty results from the possibility that the overlayers, which contain silicon oxides, may charge during

examination in the electron microscope. Conventional “trench” focused ion beam milling (Park, 1990; Szot et al., 1992) was used to prepare the specimen, which has a nominal thickness of 400 nm. Figure 18–22b shows eight-times-amplified phase contours obtained from the region marked “1” in Figure 18–22a. Instead of the expected phase distribution, which should be proportional to the mean inner potential multiplied by the specimen thickness, elliptical contours are visible in each oxide region, and an electrostatic fringing field is present outside the specimen (at the top of Figure 18–22b). Both the elliptical contours and the fringing field are associated with the build-up of positive charge in the oxide layer. The elliptical contours are centered several hundreds of nanometers from the specimen edge. Figure 18–22c shows a similar phase image obtained after coating the specimen on one side with approximately 20 nm of carbon. The effects of charging are now absent, there is no fringing field outside the specimen edge, and the phase contours follow the change in specimen thickness. One-dimensional phase profiles were generated from the phase images used to form Figure 18–22b and c along the line marked “2” in Figure 18–22a, and are shown in Figure 18–22d. The dashed and solid lines correspond to results obtained before and after coating the specimen with carbon, respectively, while the dotted line shows the difference between the solid and dashed lines. If the charge is assumed to be distributed through the thickness of the specimen, then the electric field in the oxide is approximately 2×10^7 V/m. This value is just below the breakdown electric field for thermal SiO₂ of 10^8 V/m (Sze, 2002). Equivalent results obtained from a specimen of 150 nm nominal thickness show that the elliptical contours are closer to the specimen edge. The effect of specimen charging on the dopant potential (in the source and drain regions of the transistors) is just as significant. The phase gradient continues into the substrate, and the dopant potential is undetectable before carbon coating, whether or not a phase ramp is subtracted from the images. If focused ion beam milling from the substrate side of the wafer (Schwarz et al., 2003) is used, then specimen charging no longer occurs, presumably as a result of Si redeposition onto the specimen surface. McCartney et al. (2003) provide an overview of this and other techniques for the preparation of semiconductor devices for electron holography.

Although questions still remain about phase contrast observed at simple *p–n* junctions, electron holographic data have been interpreted from more complicated semiconductor device structures, in which changes in composition as well as doping concentration are present. One example is a strained *n*-Al_{0.1}Ga_{0.9}N/In_{0.1}Ga_{0.9}N/*p*-Al_{0.1}Ga_{0.9}N heterojunction diode, in which strong piezoelectric and polarization fields are used to induce high two-dimensional electron gas concentrations (McCartney et al., 2000). To interpret experimental measurements of the potential profile across the heterojunction, after corrections for specimen thickness changes (assuming a linear thickness profile and neglecting contributions to the measured phase from variations in mean inner potential), additional charge had to be added to simulations. In particular, a sheet of negative charge was included at the

bottom of the InGaN well. The sheet charge density at this position was $2.1 \times 10^{13} \text{cm}^{-2}$. More recently, electron holography has been used to measure internal electrostatic potentials across InGaN quantum wells with thicknesses ranging from 2 to 10 nm (Stevens et al., 2004). In this study, the electric field strengths across the wells were observed to decrease in strength above a well thickness of 6 nm.

4.3 Space Charge Layers at Grain Boundaries

Electron holography has been used to characterize space charge layers at doped and undoped grain boundaries in electroceramics, although several contributions to the electron holographic phase shift can complicate interpretation. The space charge distribution that is predicted to form at such a grain boundary (Frenkel, 1946) is often described as a double (back-to-back) Schottky barrier. For Mn-doped and undoped grain boundaries in SrTiO_3 , a decrease in the measured phase shift at the boundary relative to that in the specimen was observed (Ravikumar et al., 1995). The changes in phase measured at the doped boundaries were larger in magnitude and spatial extent than at similar undoped boundaries. Possible contributions to the contrast from changes in density, composition, specimen thickness, dynamic diffraction, and electrostatic fringing fields (Pozzi, 1996; Dunin-Borkowski and Saxton, 1997) were considered, and the remaining contributions to the measured phase shifts at the doped boundaries were attributed to space charge. The sign of the space charge contribution to the specimen potential was consistent with the presence of Mn^{2+} and Mn^{3+} ions on Ti sites at the boundaries. The results were finally interpreted in terms of a narrow (1–2 nm) region of negative grain boundary charge and a wider (3–5 nm) distribution of positive space charge.

A similar approach has recently been applied to the characterization of grain boundaries in ZnO, at which a space charge layer width of approximately 150 nm has been measured (Elfving and Olsson, 2002). In an earlier study, defocus contrast recorded from delta-doped layers in Si and GaAs was also attributed to the presence of space charge (Dunin-Borkowski et al., 1994). Defocus contrast has been used to assess possible space charge contributions to electrostatic potential profiles across grain boundaries in doped and undoped SrTiO_3 (Mao et al., 1998). The contrast observed in these experiments was not consistent with a dominant contribution to the signal from space charge. Related experiments have been performed to measure polarization distributions across domain boundaries in ferroelectric materials such as BaTiO_3 and PbTiO_3 (Lichte, 2000; Lichte et al., 2003). There are many opportunities for further work on this topic.

5 High-Resolution Electron Holography

Aberrations of the objective lens, which result in modifications to the amplitude and phase shift of the electron wave, rarely need to be taken into account when characterizing magnetic and electrostatic fields at medium spatial resolution, as described in Sections 3 and 4. However,

these aberrations must be considered when interpreting electron holograms that have been acquired at atomic resolution, in which lattice fringes are visible.

The back focal plane of the objective lens contains the Fraunhofer diffraction pattern, i.e., the Fourier transform, of the specimen wave $\psi_s(r) = A_s(r)\exp[i\phi_s(r)]$, denoted $\psi(q) = FT[\psi_s(r)]$. Transfer from the back focal plane to the image plane is then represented by an inverse Fourier transform. For a perfect thin lens, neglecting magnification and rotation of the image, the complex image wave would be equivalent to the object wave $\psi_s(r)$. Modifications to the electron wave that result from objective lens aberrations can be represented by multiplication of the electron wavefunction in the back focal plane by a transfer function of the form

$$T(q) = B(q)\exp[i\chi(q)] \quad (26)$$

In Eq. (26), $B(q)$ is an aperture function that takes a value of unity for q within the objective aperture and zero beyond the edge of the aperture. The effects of two objective lens aberrations, defocus and spherical aberration, can be included in the phase factor in the form

$$\chi(q) = \pi\Delta z \lambda q^2 + \frac{\pi}{2} C_s \lambda^3 q^4 \quad (27)$$

where Δz is the defocus of the lens and C_s is the spherical aberration coefficient. The complexity of Eq. (27) increases rapidly as further aberrations are considered. The complex wave in the image plane can then be written in the form

$$\psi_i(r) = FT^{-1}[FT[\psi_s(r)] \times T(q)] \quad (28)$$

$$= \psi_s(r) \otimes t(r) \quad (29)$$

where $t(r)$ is the inverse Fourier transform of $T(q)$, and the convolution $\otimes$ of the specimen wave $\psi_s(r)$ with $t(r)$ represents the smearing of information that results from lens imperfections. Since both $\psi_s(r)$ and $t(r)$ are in general complex, the intensity of a conventional bright-field image, which can be expressed in the form

$$I(r) = |\psi_s(r) \otimes t(r)|^2 \quad (30)$$

is no longer related simply to the structure of the specimen.

The effects of lens aberrations can be removed by multiplying the complex image wave by a suitable phase plate corresponding to $T^*(q)$ to provide the amplitude and the phase shift of the *specimen* wave $\psi_s(r)$ rather than the *image* wave $\psi_i(r)$. Hence, the interpretable resolution of the image can be improved beyond the point resolution of the electron microscope. The optimal defocus that maximizes the resolution of the reconstructed specimen wave after correction of aberrations (Lichte, 1991, 1992; Lichte and Rau, 1994; Ishizuka et al., 1994) is given by the expression

$$\Delta z_{\text{opt}} = -\frac{3}{4} C_s (\lambda q_{\text{max}})^2 \quad (31)$$

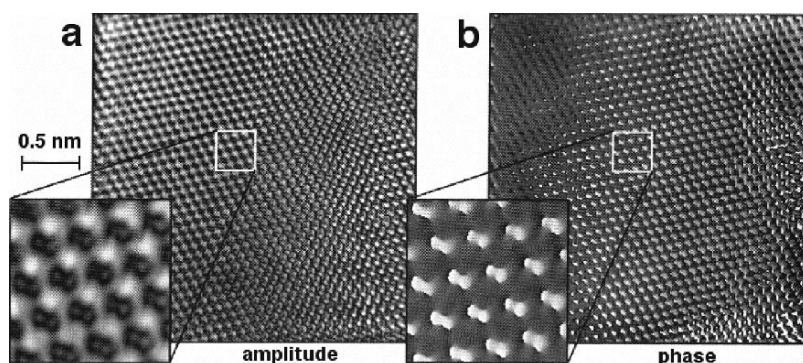


Figure 18–23. High-resolution (a) amplitude and (b) phase images of the aberration-corrected specimen wave reconstructed from an electron hologram of [110] Si, obtained at 300 kV on a CM30 FEGTEM. The spacing of the hologram fringes was 0.05 nm. The sideband contained {111}, {220}, {113}, and {004} reflections, corresponding to lateral information of 0.136 nm. The characteristic Si dumbbell structure is visible only after aberration correction. (Reprinted from Orchowski et al., 1995.)

where $q_{\max}$ is the maximum desired spatial frequency.

Figure 18–23 illustrates the application of aberration correction to a high-resolution electron hologram of crystalline Si imaged at the $\langle 110 \rangle$ zone axis, at which characteristic “dumbbell” contrast, of spacing 0.136 nm, is expected (Orchowski et al., 1995). The original hologram was acquired using an interference fringe spacing of 0.05 nm on a CM30 FEGTEM, which has a point resolution of 0.198 nm and an information limit of 0.1 nm at 300 kV. Figures 18–23a and b show, respectively, the reconstructed amplitude and phase shift of the hologram after aberration correction using a phase plate. The phase image reveals the expected white “dumbbell” contrast, at a spatial resolution that is considerably better than the point resolution of the microscope, after lens aberrations, including residual astigmatism and off-axis coma, have been measured and removed. Note also that the projected atom column positions are visible as black contrast in the amplitude image.

More recent developments in high-resolution electron holography, including the application of the technique to a wide range of materials problems, have been reviewed by Lehmann et al. (1999) and Lehmann and Lichte (2005).

6 Alternative Forms of Electron Holography

Many different forms of electron holography can be envisaged and implemented both in the TEM and in the scanning TEM (STEM) (Cowley, 1992). Equally, there are several ways in which the off-axis mode of TEM electron holography can be implemented. A full discussion of these various schemes, which include interferometry in the diffraction plane of the microscope (Herring et al., 1995) and reflection

electron holography (Banzhof and Herrmann, 1993), is beyond the scope of this chapter. Here, some of the more important developments are reviewed.

The need for a vacuum reference wave is a major drawback of the standard off-axis mode of TEM holography since this requirement restricts the region that can be examined to near the specimen edge. In many applications, the feature of interest is not so conveniently located. The implementation of a DPC mode of electron holography in the TEM enables this restriction to be overcome. DPC imaging is well established as a technique in the STEM, involving the use of various combinations of detectors to obtain magnetic contrast (Dekkers and de Lang, 1974; Rose, 1977; Chapman et al., 1978). It has also been shown (Mankos et al., 1994) that DPC contrast can be obtained using far-out-of-focus STEM electron holography (see below). An equivalent TEM configuration can be achieved by using an electron biprism located in the condenser aperture plane of the microscope (McCartney et al., 1996). Figure 18–24a shows a schematic ray diagram that illustrates the electron-optical configuration for this differential mode of off-axis TEM holography. The application of a positive voltage to the biprism results in the formation of two closely spaced, overlapping plane waves, which appear to originate from sources S_1 and S_2 to create an interference fringe pattern at the specimen level. When the observation plane is defocused by a distance Δz with respect to the specimen plane, the two coherent beams produced by the beam splitter, which are labeled k_1 and k_2 in Figure 18–24a, impinge upon different parts of the specimen. For a magnetic material, the difference in the component of the magnetic induction parallel to the biprism wire between these two points in the specimen plane determines the relative phase shift of the holographic fringes, thus giving differential phase contrast. Since the hologram is acquired under out-of-focus conditions, it is in effect the superposition of a pair of Fresnel images. The biprism voltage must be adjusted so that the feature of interest or the desired spatial resolution is sampled by at least three interference fringes. An appropriate postspecimen magnification should be chosen to ensure that the interference fringes are properly sampled by the recording medium. Figure 18–24b shows a composite phase image formed from a series of eight DPC holograms of a 30-nm-thick Co film. The fringe system was shifted progressively across the specimen plane between exposures. In addition to the holographic interference fringes, the image shows black and white lines that delineate walls between magnetic domains, with magnetization ripple visible within the domains. All of the image features are doubled due to the split incident beam. Figure 18–24c shows the final reconstructed DPC image obtained from Figure 18–24b, in which the contrast is proportional to the component of the magnetic induction parallel to the holographic fringes. The arrow below the image indicates the direction of the component of the induction analyzed in this experiment. Several magnetic vortices, at which the measured field direction circles an imperfection in the film, are visible. One such vortex is indicated by an arrow in the lower right corner of the image. For characterization of both components of the in-plane induction

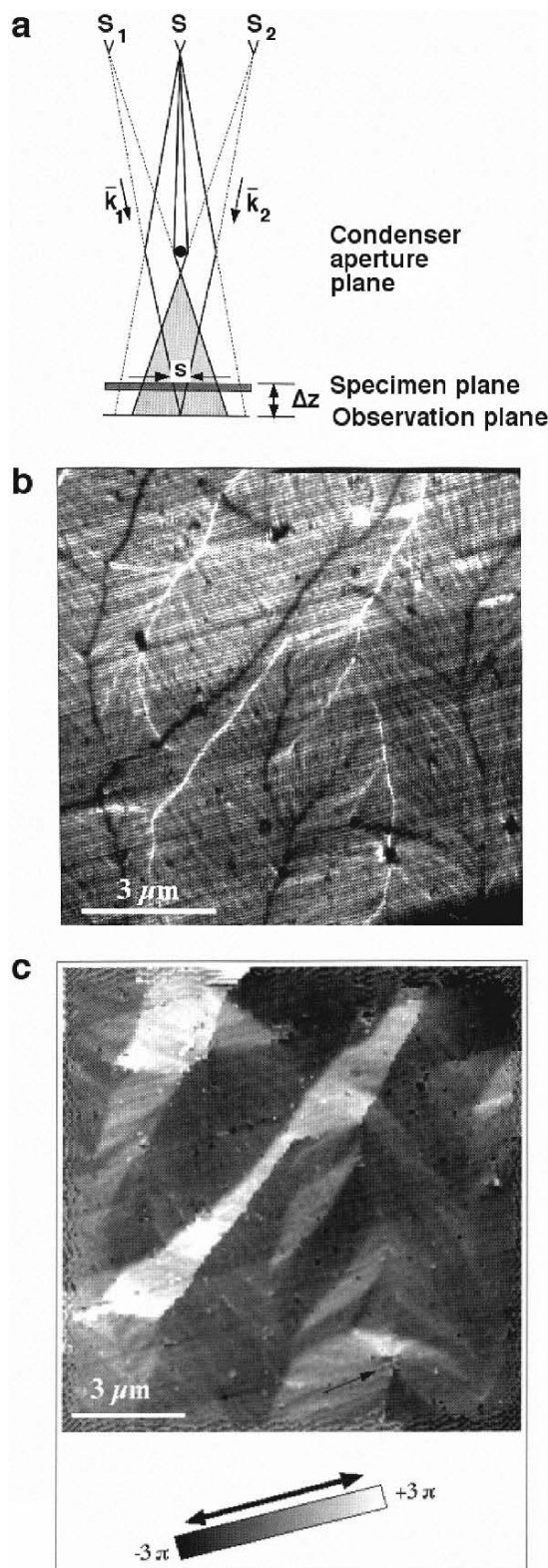


Figure 18-24. (a) Schematic ray diagram for the differential mode of off-axis TEM electron holography. The symbols are defined in the text. (b) Composite differential mode electron hologram formed from a series of eight holograms of a 30-nm-thick Co film. (c) Differential phase contrast image obtained from the hologram shown in (b). The arrow below the image indicates the direction of the magnetic signal analyzed. The arrow close to the bottom of the image indicates the position of a magnetic vortex. (Reprinted from McCartney et al., 1996.)

without removing the sample from the microscope, a rotating biprism or a rotating sample holder is required.

An alternative scheme that is conceptually similar to the differential mode of electron holography in the TEM, but which does not require the use of an electron biprism or a field emission electron gun, is termed amplitude division electron holography. Whereas conventional modes of electron holography involve splitting the wavefront of the incident illumination and thus require high spatial coherence to form interference fringes, this coherence requirement can be removed by dividing the amplitude of the electron beam instead of the wavefront. Division of the amplitude of the electron wave can be achieved by using a crystal film located before the specimen. The lattice fringes of the crystal film are then used as carrier fringes. The original configuration for this scheme involved placing the specimen in the selected-area-aperture plane of the microscope (Pozzi, 1983; Matteucci et al., 1982). The specimen can also be inserted into the normal object plane by placing a single-crystal thin film and the sample of interest on top of each other, in close proximity (Ru, 1995). The single crystal film is then tilted to a strong Bragg condition and used as an electron beam splitter. As a result of the separation of the crystal and the specimen, the hologram plane contains two defocused images of the specimen that are shifted laterally with respect to one another. One of these images is carried by the direct beam and the other by the Bragg-reflected beam. When the distance between the two images is greater than the size of the object, the images separate perfectly and interfere with adjacent plane waves to form an off-axis electron hologram. Because the single crystal is in focus and the object is out of focus, a Fresnel electron hologram of the object is obtained. The defocus of the object can be corrected at the reconstruction stage by using a phase plate, although high coherence of the incident illumination is then required. The coherence used when forming the image therefore determines the spatial resolution of the final reconstructed image. Although amplitude-division electron holography has several disadvantages over wavefront-division holography, the final phase image is not affected by effects such as Fresnel diffraction from the edges of the biprism.

An approach that can be used to increase the phase sensitivity of electron holography is termed phase-shifting electron holography. This approach is based on the acquisition of several off-axis holograms while the phase offset (the initial phase) of the image is changed, either by tilting the incident electron beam or by shifting the biprism (Ru et al., 1992). Electron holograms are recorded at successive values of the incident beam tilt, such that the phase is shifted by at least 2π over the image series. The fringe shift can be monitored in the complex Fourier spectra of the holograms. Although three holograms can in principle be used to reconstruct the object wave, in practice as many holograms as possible should be used to reduce noise. The advantages of the phase-shifting approach are greatly improved phase sensitivity and spatial resolution. Moreover, objects that are smaller than one fringe width can be reconstructed. Care is required if the object is out of focus, as tilting

the beam will also induce an image shift between successive images. Very small phase shifts have been observed from individual unstained ferritin molecules using this approach (Kawasaki et al., 1986).

Electron holograms can be acquired at video rate and subsequently digitized and processed individually to record dynamic events, but this procedure is time consuming. An alternative real-time approach for acquiring and processing holograms has been demonstrated by using a liquid-crystal panel to reconstruct holograms (Chen et al., 1994). The holograms are recorded at TV rate and transferred to a liquid-crystal spatial light modulator, which is located at the output of a Mach-Zender interferometer. The liquid crystal panel is illuminated using an He-Ne laser, and interference micrographs are observed at video rate on the monitor beside the microscope as the specimen is examined. In an alternative configuration, a liquid-crystal panel can also be used as a computer-controlled phase plate to correct for aberrations.

Whereas an off-axis electron hologram is formed by the interference of an object and a reference wave that propagate in different directions in the electron microscope, the simplest way of recording an electron hologram without using an electron biprism involves using the transmitted wave as the reference wave to form an in-line hologram. Gabor's original paper described the reconstruction of an image by illuminating an in-line hologram with a parallel beam of light and using a spherical aberration correcting plate and an astigmatism corrector. The reconstructed image is, however, disturbed by the presence of a "ghost" or "conjugate" twin image. If the hologram is recorded and subsequently illuminated by a plane wave, then the reconstructed image and a defocused conjugate image of the object are superimposed on each other. The most effective method of separating the twin images is to use Fraunhofer in-line holography. Here, in-line holograms are recorded in the Fraunhofer diffraction plane of the object (Thomson et al., 1966; Tonomura et al., 1968). Under this condition, the conjugate image is so blurred that its effect on the reconstructed image is negligible (Matsumoto et al., 1994).

The STEM holographic mode used for DPC imaging, which has similarities with the TEM differential mode of electron holography described above, is a point projection technique in which a stationary beam in an STEM is split by a biprism preceding the sample so that two mutually coherent electron point sources are formed just above the specimen. In this operating mode, the objective lens is excited weakly so that the hologram is formed in the diffraction plane rather than the image plane (Cowley, 1990). By greatly defocusing the objective lens, a shadow image of the object is formed, which has the appearance of a TEM hologram, although it is distorted by spherical aberration and defocus. The image magnification and the separation of the sources relative to the specimen are flexible in this configuration, and can be adjusted by changing the biprism voltage and/or the objective or post-specimen lens settings. The far-out-of-focus mode of STEM holography has been applied to the characterization of a range of magnetic materials (Mankos et al., 1995; Cowley et al., 1996).

A rapid approach that can be used to visualize equiphase contours involves superimposing a hologram of the specimen onto a reference hologram acquired under identical conditions, with the specimen removed from the field of view (Matteucci and Muccini, 1994). Interference effects between the holographic fringes in the two images then provide widely spaced, low-contrast bright and dark bands that reveal phase contours directly. By defocusing the combined image slightly, the unwanted finely spaced holographic interference fringes can be removed. The technique has been applied to image both electrostatic and magnetic fields.

A related approach involves the use of two parallel or perpendicular electron biprisms to generate an interference pattern between either three or four electron waves, respectively. Equiphase contours are then displayed in the recorded hologram. This method has been used to form images of electric fields outside charged latex and alumina particles and magnetic fields outside ferrite particles (Hirayama, 1999), and to expose a resist to fabricate a 100-nm-period two-dimensional grating lithographically (Ogai et al., 1995). Two parallel biprisms have also been used to form a “trapezoidal” biprism to record a double-exposure hologram with the biprism voltage changed (Tanji et al., 1996) and with the reference wave unaffected by the biprism voltage (Tanji et al., 1999).

7 Discussion and Conclusion

In this chapter, the technique of off-axis electron holography has been described, and its recent application to a wide variety of materials has been reviewed. Results have been presented from the characterization of magnetic fields in arrangements of closely spaced nanocrystals, patterned elements and nanowires, and electrostatic fields in field emitters and doped semiconductors. *In situ* experiments, which allow magnetization reversal processes to be followed and electrostatic fields in working semiconductor devices to be characterized, have been described, and the advantages of using digital approaches to record and analyze electron holograms have been highlighted. High-resolution electron holography and alternative modes of electron holography have also been described. Although the results that have been presented are specific to the dimensions and morphologies of the particular examples chosen, they illustrate the ways in which electron holography can be adapted to tackle different materials problems.

Future developments in electron holography are likely to include the development and application of new forms of electron holography and instrumentation, the introduction of new approaches for enhancing weak magnetic and electrostatic signals, the formulation of a better understanding of the effect of different TEM sample preparation techniques on phase images recorded from semiconductors and ferroelectrics, and the combination of electron holography with electron tomography to record both electrostatic and magnetic fields *inside* nanostructured materials in *three* dimensions rather than simply in

projection. If the application of electron holographic tomography to the characterization of magnetic vector fields inside materials in three dimensions, which requires two high-tilt series of holograms to be recorded about orthogonal axes, is ultimately successful, the primary difficulty may lie in the measurement and subtraction of the unwanted mean inner potential contribution to the measured phase shift at every one of the tilt angles required.

The unique capability of electron holography to provide quantitative information about magnetic and electrostatic fields in materials at a resolution approaching the nanometer scale, coupled with the increasing availability of field-emission-gun transmission electron microscopes and quantitative digital recording, ensures that the technique has a very promising future.

Acknowledgments. Thanks are due to the Royal Society, the EPSRC, FEI, and the Isaac Newton Trust for support, and to C. B. Boothroyd, P. R. Buseck, R. B. Frankel, K. Harada, R. J. Harrison, M. J. Hytch, B. Kardynal, J. Li, J. C. Loudon, P. A. Midgley, S. B. Newcomb, M. Pósfai, G. Pozzi, A. Putnis, C. A. Ross, M. R. Scheinfein, E. Snoeck, A. Tonomura, S. L. Tripp, A. C. Twitchett, A. Wei, and Y. Zhu for discussions and for ongoing collaborations. R.D.B. acknowledges the C.N.R.S. and the European Community for support through a European research network (GDR-E) entitled "Quantification and Measurement in Transmission Electron Microscopy," grouping laboratories in France, the United Kingdom, Germany, and Switzerland. We acknowledge the use of electron microscopy facilities at the John M. Cowley Center for High Resolution Electron Microscopy at Arizona State University.

References

- Aharonov, Y. and Bohm, D. (1959). *Phys. Rev.* **115**, 485.
- Allard, L.F., Völkl, E., Carim, A., Datye, A.K. and Ruoff, R. (1996). *Nano. Mater.* **7**, 137.
- Aoyama, K. and Ru, Q. (1996). *J. Microsc.* **182**, 177.
- Banzhof, H. and Herrmann, K.-H. (1993). *Ultramicroscopy* **48**, 475.
- Bazylinski, D.A. and Moskowitz, B.M. (1997). *Mineralog. Soc. Am. Rev. Mineral.* **35**, 181.
- Beleggia, M., Cristofori, D., Merli, P.G. and Pozzi, G. (2000). *Micron* **31**, 231.
- Beleggia, M., Cardinali, G.C., Fazzini, P.F., Merli, P.G. and Pozzi, G. (2001). *Inst. Phys. Conf. Ser.* **169**, 427.
- Beleggia, M., Fazzini, P.F., Merli, P.G. and Pozzi, G. (2003). *Phys. Rev. B* **67**, 045328.
- Blakemore, R.P. (1975). *Science* **190**, 377.
- Bonevich, J.E., Harada, K., Matsuda, T., Kasai, H., Yoshida, T., Pozzi, G. and Tonomura, A. (1993). *Phys. Rev. Lett.* **70**, 2952.
- Chapman, J.N., Batson, P.E., Waddell, E.M. and Ferrier, R.P. (1978). *Ultramicroscopy* **3**, 203.
- Chen, J., Hirayama, T., Tanji, T., Ishizuka, K. and Tonomura, A. (1994). *Opt. Commun.* **110**, 33.
- Cowley, J.M. (1990). *Ultramicroscopy* **34**, 293.
- Cowley, J.M. (1992). *Ultramicroscopy* **41**, 335.

- Cowley, J.M., Mankos, M. and Scheinfein, M.R. (1996). *Ultramicroscopy* **63**, 133.
- Cummings, J., Zettl, A., McCartney, M.R. and Spence, J.C.H. (2002). *Phys. Rev. Lett.* **88**, 056804.
- de Graef, M., Nuhfer, T. and McCartney, M.R. (1999). *J. Microsc.* **194**, 84.
- Dekkers, N.H. and de Lang, H. (1974). *Optik* **41**, 452.
- de Ruijter, W.J. (1995). *Micron* **26**, 247.
- de Ruijter, W.J. and Weiss, J.K. (1993). *Ultramicroscopy* **50**, 269.
- de Ruijter, W.J., Gajdardziska-Josifovska, M., McCartney, M.R., Sharma, R., Smith, D.J. and Weiss, J.K. (1992). *Scanning Microsc. Suppl.* **6**, 347.
- Doyle, P.A. and Turner, P.S. (1968). *Acta Crystallogr. A* **24**, 390.
- Dunin-Borkowski, R.E. and Saxton, W.O. (1996). In *Atomic Resolution Microscopy of Surfaces and Interfaces* (D.J. Smith and R.J. Hamers, Eds.), Materials Research Society Symposium Proceedings **466**, 73. (Materials Research Society, Warrendale, PA.)
- Dunin-Borkowski, R.E. and Saxton, W.O. (1997). *Acta Crystallogr. A* **53**, 242.
- Dunin-Borkowski, R.E., Stobbs, W.M., Perovic, D.D. and Wasilewski, Z.R. (1994). In *Proceedings of ICEM13* (B. Jouffrey and C. Colliex, Eds.), **1**, 411. Paper presented at the 13th International Conference on Electron Microscopy, Paris, France, July 17–22, 1994.
- Dunin-Borkowski, R.E., McCartney, M.R., Smith, D.J. and Parkin, S.S.P. (1998). *Ultramicroscopy* **74**, 61.
- Dunin-Borkowski, R.E., McCartney, M.R., Kardynal, B., Smith, D.J. and Scheinfein, M.R. (1999). *Appl. Phys. Lett.* **75**, 2641.
- Dunin-Borkowski, R.E., McCartney, M.R., Kardynal, B., Parkin, S.S.P., Scheinfein, M.R. and Smith, D.J. (2000). *J. Microsc.* **200**, 187.
- Dunin-Borkowski, R.E., Newcomb, S.B., McCartney, M.R., Ross, C.A. and Farhoud, M. (2001). *Inst. Phys. Conf. Ser.* **168**, 485.
- Dunin-Borkowski, R.E., Twitchett, A.C. and Midgley, P.A. (2002). *Microsc. Microanal.* **8**(Suppl. 2), 42.
- Dunin-Borkowski, R.E., McCartney, M.R. and Smith, D.J. (2004a). *Encyclopaedia of Nanoscience and Nanotechnology* (H.S. Nalwa, Ed.), **3**, 41. (American Scientific Publishers, Stevenson Ranch, CA.)
- Dunin-Borkowski, R.E., Kasama, T., Wei, A., Tripp, S.L., Hÿtch, M.J., Snoeck, E., Harrison, R.J. and Putnis, A. (2004b). *Microsc. Res. Techn.* **64**, 390.
- Dunin-Borkowski, R.E., Newcomb, S.B., Kasama, T., McCartney, M.R., Weyland, M. and Midgley, P.A. (2005). *Ultramicroscopy* **103**, 67.
- Dunlop, D.J. and Özdemir, Ö. (1997). *Rock Magnetism*. (Cambridge University Press, Cambridge, UK.)
- Elfving, M. and Olsson, E. (2002). *J. Appl. Phys.* **92**, 5272.
- Frabboni, S., Matteucci, G., Pozzi, G. and Vanzi, M. (1985). *Phys. Rev. Lett.* **55**, 2196.
- Frabboni, S., Matteucci, G. and Pozzi, G. (1987). *Ultramicroscopy* **23**, 29.
- Frenkel, J. (1946). *Kinetic Theory of Liquids* (Oxford University Press, Oxford).
- Fujita, T. and McCartney, M.R. (2005). *Ultramicroscopy* **102**, 279.
- Gabor, D. (1949). *Proc. R. Soc. London A* **197**, 454.
- Gajdardziska-Josifovska, M. and Carim, A. (1998). In *Introduction to Electron Holography* (E. Völkl, L.F. Allard and D.C. Joy, Eds.). (Kluwer Academic/Plenum Publishers, New York.)
- Gajdardziska-Josifovska, M., McCartney, M.R., de Ruijter, W.J., Smith, D.J., Weiss, J.K. and Zuo, J.M. (1993). *Ultramicroscopy* **50**, 285.
- *Geiger, D., Lichte, H., Linck, M. and Lehmann, M. (2007/8), Electron holography with a C_s -corrected electron microscope. *Microsc. Microanal.* (in press)

- Ghiglia, D.C. and Pritt, M.D. (1998). *Two-Dimensional Phase Unwrapping. Theory, Algorithms and Software* (Wiley, New York).
- Gribelyuk, M.A., McCartney, M.R., Li, J., Murthy, C.S., Ronsheim, P., Doris, B., McMurray, J.S., Hegde, S. and Smith, D.J. (2002). *Phys. Rev. Lett.* **89**, 025502.
- Harada, H., Tonomura, A., Togawa, Y., Akashi, T. and Matsuda, T. (2004). *Appl. Phys. Lett.* **84**, 3229.
- Harrison, R.J., Dunin-Borkowski, R.E., and Putnis, A. (2002). *Proc. Natl. Acad. Sci. USA* **99**, 16556.
- Harscher, A. and Lichte, H. (1996). *Ultramicroscopy* **64**, 57.
- Harscher, A. and Lichte, H. (1998). In *Electron Microscopy 98* (H.A. Calderón Benavides and M.J. Yacamán, Eds.), **1**, 553. Paper presented at the 14th International Conference on Electron Microscopy, Cancun, Mexico, August 31 to September 4, 1998 (Institute of Physics Publishing, Bristol, UK).
- Herring, R.A., Pozzi, G., Tanji, T. and Tonomura, A. (1995). *Ultramicroscopy* **60**, 153.
- Heumann, M., Uhlig, T. and Zweck, J. (2005). *Phys. Rev. Lett.* **94**, 077202.
- Hirayama, T. (1999). *Mater. Characterization* **42**, 193.
- Hu, H., Wang, H., McCartney, M.R. and Smith, D.J. (2005). *J. Appl. Phys.* **97**, 054305.
- Hýtch, M.J., Dunin-Borkowski, R.E., Scheinfein, M.R., Moulin, J., Duhamel, C., Mazelayrat, F. and Champion, Y. (2003). *Phys. Rev. Lett.* **91**, 257207.
- Ishizuka, K., Tanji, T., Tonomura, A., Ohno, T. and Murayama Y. (1994). *Ultramicroscopy* **55**, 197.
- Jacobs, I.S. and Bean, C.P. (1955). *Phys. Rev.* **100**, 1060.
- Kawasaki, T., Endo, J., Matsuda, T., Osakabe, N. and Tonomura, A. (1986). *J. Electron Microsc.* **35**, 211.
- Kawasaki, T., Missiroli, G.F., Pozzi, G. and Tonomura, A. (1993). *Optik* **92**, 168.
- Lai, G., Hirayama, K., Ishizuka, T., Tanji, T. and Tonomura, A. (1994). *Appl. Opt.* **33**, 829.
- Lehmann, M. and Lichte, H. (2005). *Cryst. Res. Technol.* **40**, 149.
- Lehmann, M., Lichte, H., Geiger, D., Lang, G. and Schweda, E. (1999). *Mater. Characterization* **42**, 249.
- Li, J., McCartney, M.R., Dunin-Borkowski, R.E. and Smith, D.J. (1999). *Acta Crystallogr. A* **55**, 652.
- Lichte, H. (1991). *Ultramicroscopy* **38**, 13.
- Lichte, H. (1992). *Ultramicroscopy* **47**, 223.
- Lichte, H. (1995). In *Electron Holography* (A. Tonomura, L.F. Allard, G. Pozzi, D.C. Joy and Y.A. Ono, Eds.), pp. 11–31. (Elsevier, Amsterdam.)
- Lichte, H. (2000). *Crystal Res. Technol.* **35**, 887.
- Lichte, H. (2002). *Phil. Trans. R. Soc. Lond. A* **360**, 897.
- Lichte, H. and Rau, W.D. (1994). *Ultramicroscopy* **54**, 310.
- Lichte, H., Herrmann, K.H. and Lenz, F. (1987). *Optik* **77**, 135.
- Lichte, H., Reibold, M., Brand, K. and Lehmann, M. (2003). *Ultramicroscopy* **93**, 199.
- *Lichte, H., Formanek, P., Lenk, A., Linck, M., Matzeck, C., Lehmann, M. and Simon, P. (2007). Electron holography: Applications to materials questions. *Annu. Rev. Mater. Res.* **37**, 539–588.
- Lin, X. and Dravid, V.P. (1996). *Appl. Phys. Lett.* **69**, 1014.
- Lloyd, S.J., Dunin-Borkowski, R.E. and Boothroyd, C.B. (1997). *Inst. Phys. Conf. Ser.* **153**, 113.
- Loudon, J.C., Mathur, N.D., and Midgley, P.A. (2002). *Nature* **420**, 797.
- Mankos, M., Scheinfein, M.R. and Cowley, J.M. (1994). *J. Appl. Phys.* **75**, 7418.
- Mankos, M., Cowley, J.M. and Scheinfein, M.R. (1995). *Mat. Res. Soc. Bull.* **20**, 45.

- Mao, Z., Dunin-Borkowski, R.E., Boothroyd, C.B. and Knowles, K.M. (1998). *J. Am. Ceram. Soc.* **81**, 2917.
- Matsuda, T., Tonomura, A., Suzuki, R., Endo, J., Osakabe, N., Umezaki, H., Tanabe, H., Sugita, Y. and Fujiwara, H. (1982). *J. Appl. Phys.* **53**, 5444.
- Matsuda, T., Hasegawa, S., Igarashi, M., Kobayashi, T., Naito, M., Kajiyama, H., Endo, J., Osakabe, N. and Tonomura, A. (1989). *Phys. Rev. Lett.* **62**, 2519.
- Matsuda, T., Fukuhara, A., Yoshida, T., Hasegawa, S., Tonomura, A. and Ru, Q. (1991). *Phys. Rev. Lett.* **66**, 457.
- Matsumoto, T. and Tonomura, A. (1996). *Ultramicroscopy* **63**, 5.
- Matsumoto, T., Tanji, T. and Tonomura, A. (1994). *Ultramicroscopy* **54**, 317.
- Matteucci, G. and Muccini, M. (1994). *Ultramicroscopy* **53**, 19.
- Matteucci, G., Missiroli, G.F. and Pozzi, G. (1982). *Ultramicroscopy* **8**, 403.
- Matteucci, G., Missiroli, G.F. and Pozzi, G. (1988). *Physica B* **151**, 223.
- Matteucci, G., Missiroli, G.F. Muccini, M. and Pozzi, G. (1992). *Ultramicroscopy* **45**, 77.
- Matteucci, G., Muccini, M. and Hartmann, U. (1994). *Phys. Rev. B* **50**, 6823.
- Matteucci, G., Missiroli, G.F. and Pozzi, G. (1998). *Adv. Imaging Electron Phys.* **99**, 171.
- Matteucci, G., Missiroli, G.F. and Pozzi, G. (2002). *Adv. Imaging Electron Phys.* **122**, 173.
- McCartney, M.R. and Gajdardziska-Josifovska, M. (1994). *Ultramicroscopy* **53**, 283.
- *McCartney, M.R. and Smith, D.J. (2007). Electron holography: Phase imaging with nanometer resolution. *Annu. Rev. Mater. Res.* **37**, 729–767.
- McCartney, M.R. and Zhu, Y. (1998). *Appl. Phys. Lett.* **72**, 1380.
- McCartney, M.R., Kruit, P., Buist, A.H. and Scheinfein, M.R. (1996). *Ultramicroscopy* **65**, 179.
- McCartney, M.R. and Dunin-Borkowski, R.E. (1998). In *Electron Microscopy 98* (H.A. Calderón Benavides and M. José Yacamán, Eds.), **2**, 497. Paper presented at the 14th International Conference on Electron Microscopy, Cancun, Mexico, August 31 to September 4, 1998. (Institute of Physics Publishing, Bristol, UK.)
- McCartney, M.R., Ponce, F.A., Cai, J. and Bour, D.P. (2000). *Appl. Phys. Lett.* **76**, 3055.
- McCartney, M.R., Gribelyuk, M.A., Li, J., Ronsheim, P., McMurray, J.S. and Smith, D.J. (2002). *Appl. Phys. Lett.* **80**, 3213.
- McCartney, M.R., Li, J., Chakraborty, P., Giannuzzi, L.A. and Schwarz, S.M. (2003). *Microsc. Microanal.* **9**(Suppl. 2), 776.
- Meyer, R.R. and Kirkland, A.I. (2000). *Micr. Res. Techn.* **49**, 269.
- Midgley, P.A. (2001). *Micron* **32**, 167.
- Möllenstedt, G. and Düker, H. (1954). *Naturwissenschaften* **42**, 41.
- Ogai, K., Kimura, Y., Shimizu, R., Fujita, J. and Matsui, S. (1995). *Appl. Phys. Lett.* **66**, 1560.
- O’Keeffe, M. and Spence, J.C.H. (1994). *Acta Crystallogr. A* **50**, 33.
- Orchowski, A., Rau, W.D. and Lichte, H. (1995). *Phys. Rev. Lett.* **74**, 399.
- Osakabe, N., Yoshida, K., Horiuchi, Y., Matsuda, T., Tanabe, H., Okuwaki, T., Endo, J., Fujiwara, H. and Tonomura, A. (1983). *Appl. Phys. Lett.* **42**, 746.
- Park, K.H. (1990). *Mater. Res. Soc. Symp. Proc.* **199**, 271.
- Pozzi, G. (1983). *Optik* **66**, 91.
- Pozzi, G. (1996). *J. Phys. D: Appl. Phys.* **29**, 1807.
- Pozzi, G. and Vanzi, M. (1982). *Optik* **60**, 175.
- Price, G.D. (1981). *Am. Mineral.* **66**, 751.
- Radi, G. (1970). *Acta Crystallogr. A* **26**, 41.
- Rau, W.D., Baumann, F.H., Rentschler, J.A., Roy, P.K. and Ourmazd, A. (1996). *Appl. Phys. Lett.* **68**, 3410.

- Rau, W.D., Schwander, P., Baumann, F.H., Höppner, W. and Ourmazd, A. (1999). *Phys. Rev. Lett.* **82**, 2614.
- Ravikumar, V., Rodrigues, R.P. and Dravid, V.P. (1995). *Phys. Rev. Lett.* **75**, 4063.
- Reimer, L. (1991). *Transmission Electron Microscopy* (Springer-Verlag, Berlin).
- Rez, D., Rez, P. and Grant, I. (1994). *Acta Crystallogr., Sect. A* **50**, 481.
- Rose, H. (1977). *Ultramicroscopy* **2**, 251.
- Ross, C.A. (2001). *Ann. Rev. Mater. Res.* **31**, 203.
- Ru, Q. (1995). *J. Appl. Phys.* **77**, 1421.
- Ru, Q., Endo, J., Tanji, T. and Tonomura, A. (1992). *Appl. Phys. Lett.* **59**, 2372.
- Schwarz, S.M., Kempshall, B.W., Giannuzzi, L.A. and McCartney, M.R. (2003). *Microsc. Microanal.* **9**(Suppl. 2), 116.
- Seraphin, S., Beeli, C., Bonard, J.M., Jiao, J., Stadelmann, P.A. and Chatelain, A. (1999). *J. Mat. Res.* **14**, 2861.
- Signoretti, S., Del Blanco, L., Pasquini, L., Matteucci, G., Beeli, C. and Bonetti, E. (2003). *J. Magn. Magn. Mat.* **262**, 142.
- Smith, D.J. and McCartney, M.R. (1998). In *Introduction to Electron Holography* (E. Völkl, L.F. Allard and D.C. Joy, Eds.), 87–106 (Kluwer Academic/Plenum Publishers, New York).
- Smith, D.J., de Ruijter, W.J., Weiss, J.K. and McCartney, M.R. (1998). In *Introduction to Electron Holography* (E. Völkl, L.F. Allard and D.C. Joy, Eds.), 116 (Kluwer Academic/Plenum Publishers, New York).
- Smith, D.J., Dunin-Borkowski, R.E., McCartney, M.R., Kardynal, B. and Scheinfein, M.R. (2000). *J. Appl. Phys.* **87**, 7400.
- Snoeck, E., Dunin-Borkowski, R.E., Dumestre, F., Renaud, P., Amiens, C., Chaudret, B. and Zurcher, P. (2003). *Appl. Phys. Lett.* **82**, 88.
- Somodi, P.K., Dunin-Borkowski, R.E., Twitchett, A.C., Barnes, C.H.W. and Midgley, P.A. (2005). In *Electron Microscopy of Molecular and Atom-Scale Mechanical Behavior, Chemistry and Structure* (D. Martin, D.A. Muller, E. Stach and P.A. Midgley, Eds.), *Mater. Res. Soc. Proc.* **839**, P3.2 (Materials Research Society, Pittsburgh, PA).
- Spence, J.C.H. (1993). *Acta Crystallogr. A* **49**, 231.
- Stevens, M., Bell, A., McCartney, M.R., Ponce, F.A., Marui, H. and Tanaka, S. (2004). *Appl. Phys. Lett.* **85**, 4651.
- Sze, S.M. (2002). *Physics of Semiconductor Devices* (Wiley, New York).
- Szot, J., Hornsey, R., Ohnishi, T. and Minagawa, J. (1992). *J. Vac. Sci. Technol. B* **10**, 575.
- Tanji, T., Ru, Q. and Tonomura, A. (1996). *Appl. Phys. Lett.* **69**, 2623.
- Tanji, T., Manabe, S., Yamamoto, K. and Hirayama, T. (1999). *Ultramicroscopy* **75**, 197.
- Thomson, B.J., Parrent, G.B., Ward, J.H. and Justh, B. (1966). *J. Appl. Meteorol.* **5**, 343.
- Titchmarsh, J.M., Lapworth, A.J. and Booker, G.R. (1969). *Phys. Stat. Sol.* **34**, K83.
- Tonomura, A. (1992). *Adv. Phys.* **41**, 59.
- Tonomura, A. (1998). *The Quantum World Unveiled by Electron Waves* (World Scientific, Singapore).
- Tonomura, A., Fukuhara, A., Watanabe, H. and Komoda, T. (1968). *Jpn. J. Appl. Phys.* **7**, 295.
- Tonomura, A., Matsuda, T., Suzuki, R., Fukuhara, A., Osakabe, N., Umezaki, H., Endo, J., Shinagawa, K., Sugita, Y. and Fujiwara, H. (1982). *Phys. Rev. Lett.* **48**, 1443.
- Tonomura, A., Umezaki, H., Matsuda, T., Osakabe, N., Endo, J. and Sugita, Y. (1983). *Phys. Rev. Lett.* **51**, 331.
- Tonomura, A., Matsuda, T., Kawasaki, T., Endo, J. and Osakabe, N. (1985). *Phys. Rev. Lett.* **54**, 60.

- Tonomura, A., Matsuda, T., Endo, J., Arii, T. and Mihama, K. (1986). *Phys. Rev. B: Solid State* **34**, 3397.
- Tonomura, A., Allard, L.F., Pozzi, G., Joy, D.C. and Ono, Y.A., Eds. (1995). *Electron Holography* (Elsevier, Amsterdam).
- Tripp, S.L., Pusztay, S.V., Ribbe, A.E. and Wei, A. (2002). *J. Am. Chem. Soc.* **124**, 7914.
- Tripp, S.L., Dunin-Borkowski, R.E. and Wei, A. (2003). *Angew. Chem.* **42**, 5591.
- Twitchett, A.C., Dunin-Borkowski, R.E. and Midgley, P.A. (2002). *Phys. Rev. Lett.* **88**, 238302.
- Twitchett, A.C., Dunin-Borkowski, R.E., Hallifax, R.J., Broom, R.F. and Midgley, P.A. (2004). *J. Microsc.* **214**, 287.
- Twitchett, A.C., Dunin-Borkowski, R.E., Hallifax, R.J., Broom, R.F. and Midgley, P.A. (2005). *Microsc. Microanal.* **11**, 66.
- *Twitchett-Harrison, A.C., Yates, T.J.V., Newcomb, S.B., Dunin-Borkowski, R.E. and Midgley, P.A. (2007). High-resolution three-dimensional mapping of semiconductor dopant potentials. *Nano Lett.* **7**, 2020–2023.
- Völkl, E., Allard, L.F., Datye, A. and Frost, B. (1995). *Ultramicroscopy* **58**, 97.
- Völkl, E., Allard, L.F. and Joy, D.C., Eds. (1998). *Introduction to Electron Holography* (Plenum, New York).
- Wang, Y.C., Chou, T.M., Libera, M. and Kelly, T.F. (1997). *Appl. Phys. Lett.* **70**, 1296.
- Wang, Z., Hirayama, T., Kato, T., Sasaki, K., Saka, H. and Kato, N. (2002a). *Appl. Phys. Lett.* **80**, 246.
- Wang, Z., Sasaki, K., Kato, N., Urata, K., Hirayama, T. and Saka, H. (2002b). *J. Electron Microsc.* **50**, 479.
- Wang, Z., Kato, T., Shibata, N., Hirayama, T., Kato, N., Sasaki, K. and Saka, H. (2002c). *Appl. Phys. Lett.* **81**, 478.
- Wang, Z., Kato, T., Hirayama, T., Kato, N., Sasaki, K. and Saka, H. (2005). *Surf. Int. Anal.* **37**, 221.
- Weiss, J.K., de Ruijter, W.J., Gajdardziska-Josifovska, M., Smith, D.J., Völkl, E. and Lichte, H. (1991). In *Proceedings of the 49th Annual EMSA Meeting* (G.W. Bailey, Ed.), 674 (San Francisco Press, San Francisco).
- Wohlleben, D.J. (1971). *Electron Microscopy in Materials Science*, Vol. 2, 712 (Academic Press, New York).
- Yamamoto, K., Tanji, T. and Hibino, M. (2000). *Ultramicroscopy* **85**, 35.
- Yamamoto, K., Hirayama, T. and Tanji, T. (2004). *Ultramicroscopy* **101**, 265.

*References added since the first printing.

19

Diffraction (Lensless) Imaging

John C.H. Spence

1 Introduction

Diffraction (or lensless) imaging refers to the use of theoretical methods and computer algorithms to solve the phase problem for scattering by a nonperiodic object. The name coherent X-ray diffraction imaging (CXDI) is used in the X-ray community, which we could generalize to CDI. Additional information about the object, such as the sign of the scattering potential and the approximate boundary of the object, may be combined with the measured scattered intensity to solve for the phases of the scattered amplitudes. In this way, under conditions of single scattering (and other approximations that often apply in optics and in electron, X-ray, and neutron diffraction) it may be possible to reconstruct a real-space image of an object by Fourier transform of the complex scattering distribution, or Fraunhofer diffraction pattern. (Applications to Fresnel near-field imaging are also possible. In this geometry, resolution is, however, limited by detector pixel size, since the magnification is unity if lenses are not used.) In this review we will not discuss the recently developed and powerful transport of intensity method, which is also applicable to the near field (Paganin and Nugent, 1998). By avoiding the need for a lens, the aberrations and resolution limits introduced by lenses are thus avoided. Within the past decade this process has been demonstrated experimentally for neutron, X-ray, and electron scattering, so that the field has reached an exciting point. The electron work has produced atomic-resolution images, while experiments with soft X-rays have finally produced three-dimensional (tomographic) reconstructions. It now offers the real possibility of diffraction-limited imaging with any radiation for which lenses do not exist. Since each radiation interacts differently with matter, the method can be expected to provide us with new information on matter in fields as diverse as biology, materials science, and astronomy. Much current interest focuses on tomographic imaging of whole cells, nanoparticles and mesoporous materials, and on the imaging of proteins at near-atomic resolution. At present, the inversion of even a two-dimensional diffraction pattern is a lengthy process, and while three-dimensional

inversion may involve weeks of work for a $1024 \times 1024 \times 1024$ voxel data set (with much time devoted to data merging and preparation issues), it is clear that automated processing will soon be possible. While it can be shown rigorously that some of the algorithms described here do not diverge, because of the nonconvex nature of the constraints involved no formal proof of convergence or of the uniqueness of solutions exists. Yet computational trials for simulated data with little noise practically always converge rapidly to the known solution, while the inevitable stagnation occurs in the presence of excessive noise. [The ability of these algorithms to “climb out” of local minimum and find a global minimum in a data set with perhaps a million unknown parameters (phases) is one of their most remarkable features.] The challenge now is to perform experiments under conditions in which the experimental parameters are sufficiently well known that inversion of experimental data performs as well as the simulations. In summary, over the past decade many algorithms for lensless imaging have been published that work well on simulated noisy data; the more difficult problem is experimental data collection under the conditions for which the algorithms converge. (This may require, for example, that the experimental scattered intensity distribution possesses inversion symmetry, consistent with the theoretical assumption of a two-dimensional hermitian potential or “real object.”)

2 History

The phase problem in optics has early origins: in a letter to Michelson, Rayleigh comments that “the phase problem in interferometry is insoluble without a-priori information on the symmetry of the data” (Strutt, 1892). Several fields have contributed to the development of our current working solution to this problem, including signal processing, wavefront sensing, astronomy, electron microscopy, image processing, and X-ray crystallography, in addition to applied mathematics. Thus a wonderfully rich set of ideas has contributed, from communications theory to diffraction physics and set theory. Since my background is in diffraction physics, I must apologize to those workers in other areas (such as acoustics) whose contributions to the vast literature on this subject I have overlooked. In this review we will not discuss the crystallographic phase problem, except to note parallels (such as solvent flattening) with our noncrystallographic or “single particle” problem (Carrozzini et al., 2004). [By the obvious method of periodic continuation, it has been shown that the highly successful direct methods of crystallography can be applied immediately to the continuous scattering from an isolated object, if the correct choice of sampling interval is made (Spence et al., 2003b), however this may not be the most efficient solution.] In retrospect, we can now see that the key to a successful solution was given in the crystallography literature as early as 1952. D. Sayre in “Some implications of a theorem due to Shannon” (Sayre, 1952) appears to have been the first to consider the relationship between Shannon’s sampling theorem and Bragg’s law. He pointed out that for

centrosymmetric crystals, where only the sign of structure factors F_g is unknown, these could be determined if intensities $|F|_g^2$ were observable at half-integral values of g , since a change of sign requires a zero crossing of the molecular scattering factor there. By Shannon's theorem, these fractional and integral order intensities are needed and sufficient to completely define the autocorrelation function of the molecule. Sayre's fractional orders could thus be generated only by a molecule that filled half the cell, leaving the remaining "half" empty. [The Patterson function or autocorrelation function of an inorganic, continuously bonded crystal will differ from that of the basis (one "molecule"), since the autocorrelation of the molecule is twice as large as the unit cell. These concepts will be clarified in the next section.] Sayre concludes that a solution to the phase problem for centric crystals could as well be obtained from measurements of the *intensity* of half-order reflections as from study of the Bragg intensities. [The paper was prompted by the observation of nonintegral reflections (following hydration) in experimental patterns from hemoglobin by Perutz and others, and a talk by Gay.] Since diffraction from an isolated nonperiodic object is a continuous function of scattering angle it provides immediate access to these "fractional orders." Thus a solution in principle to the phase problem for real, centric single particles has existed in the literature since 1952.

Sayre's paper had no impact in the image processing or signal processing communities, and we will not trace here its important influence on the development of density modification techniques in protein crystallography. (Here the molecule is often surrounded by a water jacket of comparable volume to the molecule, effectively doubling the cell size and possibly giving rise to "half-integral" reflections.) For imaging, the next important development was work by Gerchberg and Saxton (G-S) (Gerchberg and Saxton, 1971), which posed the following question: if the intensity of a general complex two-dimensional image and its corresponding diffraction pattern intensity is known, can the complex image (and diffraction pattern) be reconstructed? (By a complex image we refer here to the sample exit-face wavefunction, as discussed in the next section.) These data are available in a modern electron microscope. By solving the resulting quadratic equations (equal in number, as they comment, to the number of unknown phases) using an iterative algorithm, they showed successful inversion for one-dimensional data. It has often been noted that these Fourier equations, one for each pixel in the detector, may not be independent, are nonlinear if a sign constraint is used, and are corrupted by noise. Dependence in the equations may be introduced by choice of support shape. Other approaches have been explored for solving these nonlinear Fourier equations, but the G-S paper established the use of iterations between real and reciprocal space, with known information imposed repeatedly in each domain. G-S do not establish uniqueness or convergence in general.

The connection with Sayre's paper was not made; however, in retrospect we see that the sampling interval used by G-S for their analysis of a nonperiodic object necessarily corresponded to the Shannon half-integral orders of Sayre's paper. Modern work is focused on three-

dimensional data in which only the diffraction pattern intensity from a real or complex object is measured. However, for a pure phase object, the modulus of the exit face wavefunction is known a priori to be unity everywhere, greatly reducing the number of unknowns in the G–S analysis. (This unit modulus constraint is, unfortunately, nonconvex.) For a real object (or, equivalently, a weak phase object) the diffraction pattern has Hermitian symmetry (Friedel’s law), resulting in a further reduction in unknowns. Following the Gerchberg and Saxton paper (and a second paper, Gerchberg and Saxton, 1972), it was still necessary to determine the minimum information needed about a real object to solve the phase problem if the diffracted intensity was given. The role of the object boundary (defining a “support”) and the need to sample the continuous scattering finely enough to satisfy Shannon’s theorem (treating the autocorrelation function as the “bandlimit”) were all recognized at this time.

The autocorrelation of a two-dimensional object is twice as large as the object in any direction, so Shannon’s theorem requires sampling at half the Bragg angle for a periodically continued object. Bates referred to this as “oversampling,” a term still in use. (Since it is in fact *optimum* sampling of the diffracted *intensity*, we avoid that term here.) The problem was vigorously attacked by authors such as Bates, Fiddy, Fienup, Gonslaves, and Papoulis in the early nineteen eighties, and the use of lensless imaging based on these ideas using X-ray scattering was advocated (Sayre, 1980). The early series of papers by Bates and co-workers are especially significant, and can be traced through Bates and McDonnell (1989), while Fiddy provided a new approach based on analysis of zeros in the diffraction pattern (e.g., Liao et al., 1997, and earlier work). It was shown, for example, that the set of all bandlimited functions having a given set of real zero crossings is convex, and that, in the absence of noise, the data might be factorizable. By 1982 a useful working solution had been obtained (Fienup, 1982) for real two-dimensional images by the addition of a feedback feature to the G–S error-reduction (ER) algorithm. An important realization at about this time was that the “landscape” (a map of error metric in the N -dimensional search space) for the phase problem was not rugged, and usually consisted of a single shallow global minimum with a few small bumps; the essential difficulty is not caused by the number of local false minima, but rather by the large number of directions in which search is possible, in a space containing one dimension for each image pixel. This Fienup algorithm has come to be known as the hybrid input–output or HIO algorithm. Feedback greatly speeds up convergence and improves the ability of the algorithm to escape from local minima, however, most modern work is based on a combination of the ER and HIO algorithms, since only the ER algorithm supplies a reliable error metric. Fienup showed that the ER algorithm converges, in the sense that the error monotonically decreases (Fienup, 1997). Real images could then be reconstructed from Fraunhofer diffraction pattern intensity data, provided that the sign of the scattering potential was known together with an approximate estimate of the object boundary (support). Important work on the uniqueness problem for the HIO algorithm was

published by Bruck and Sodin (1979) and Baraket and Newsam (1984) who showed that ambiguous solutions are obtained with these constraints only on “pathologically rare” occasions. The aim of this work is to show that only one image function (or images equivalent to this) satisfies the sign, support, and Fourier modulus constraints.

The HIO algorithm, which we give below, is the basis for most successful modern work, and has led to the development of many variants and much detailed mathematical analysis. The book by Stark (1987) provided an excellent overview of the subject up to 1987, and can be strongly recommended to students, together with an important review article by Millane (1990), which unified the crystallographic and single-particle approaches and so brought the field to the attention of a much larger audience. [The striking resemblance of the HIO algorithm to the independently developed solvent flattening methods of X-ray crystallography can be traced through Wang (1985).] The water-jacket around a protein in a crystal plays the same role as the “zero-density” region outside the support of our single particles.

The HIO algorithm and its variants have been highly successful for real objects, however, rather less attention has been paid to the more difficult problem of complex image reconstruction. Except for special cases, it is found that a very precise knowledge of the support is needed, and that in one and two dimensions this support should be disjoint (separated into two parts) (Fienup, 1987). Iterative inversion schemes fail for one-dimensional real data that are not disjoint, and work better in higher dimensions.

The third more recent period of work since about 1990, during which the field has grown rapidly, is discussed in the next sections. During that time a major effort was made to apply these methods to soft X-ray transmission data at the Brookhaven synchrotron (Sayre et al., 1998; Miao et al., 1998). Our understanding of the success of the HIO algorithm has increased considerably in that period due to the contribution of mathematicians using powerful methods based on convex set theory; Bregman projections and constraint theory were first described as early as 1982 (Youla and Webb, 1982) but not taken up until recently (for an excellent review, see Bauschke et al., 2002). Questions of uniqueness in dimensions greater than two have been considered, where the problem is overdetermined (Millane, 1995), and it is generally found that the iterative algorithms work better. In addition, a number of developments of the HIO algorithm have been published (Bauschke et al., 2003; Elser, 2003); the standard symbols used by mathematicians for the theory of projections has been adopted to describe the problem, and a new algorithm, which dispenses with any need for knowledge of the object support [the “shrinkwrap” algorithm (Marchesini et al., 2003b)], has been published. The problem of lost information within the synchrotron beam-stop has been addressed by reconstructing the diffuse scattering around a Bragg reflection from a nanocrystal (Robinson et al., 2001). In 2001 the first of a biannual series of conferences on the noncrystallographic phase problem was held at Lawrence Berkeley Laboratory (Spence et al., 2001), with a second in Cairns, Australia. The latest, held at Porquelles in France in 2005, can be found

at <http://www.esrf.fr/NewsAndEvents/Events/Non-Crys20-06-05/>. An entirely different approach to the solution of the phase problem, based on the transport of intensity equations, has also been fully developed during this period and applied to optical, X-ray, and electron imaging (Paganin and Nugent, 1998).

These theoretical developments, often based on computer simulations, have been supported by far fewer demonstrations of inversions from experimental data. Only experimental results can truly give confidence in theory, and early results were somewhat disappointing. An early application of the HIO algorithm to experimental optical speckle data can be found in Cederquist et al. (1988), and applications to images formed with laser light can be found in Weierstall et al. (2001). In the X-ray imaging community the work at Brookhaven finally paid off in 1999, when images of lithographed characters were reconstructed from soft X-ray diffraction patterns (Miao et al., 1999). More recently, an atomic-resolution image of a single double-walled carbon nanotube has been reconstructed from its electron microdiffraction pattern (Zuo et al., 2003). These successes have led to rapid growth of the subject in recent years.

3 Objects, Images, and Diffraction Patterns: Validity Domains of Approximations

The terms *object*, *image*, *diffraction pattern*, *exit-face wavefunction*, *real and complex object*, and *transmission function* have sometimes been confused in the literature and are defined here for clarity. We define the *object* by its ground-state charge density $\rho(\mathbf{R})$ (which diffracts X-rays) and by the corresponding electrostatic potential $V(\mathbf{R})$ (which diffracts electrons). $V(\mathbf{R})$ is related to the density $\rho(\mathbf{R})$ by Poisson's equation. As a result of inelastic processes, both may be complex, and could be referred to generally as the “complex optical potential,” but in this chapter unless otherwise stated we take them to be real. Here $\mathbf{R}$ is a three-dimensional vector while $\mathbf{r}$ will be two dimensional. We define the *exit-face wavefunction* $\Psi(\mathbf{r})$ across the downstream face of a thin slab of sample in the transmission diffraction geometry with *transmission function* $T(\mathbf{r})$ by

$$\Psi(\mathbf{r}) = T(\mathbf{r}) \Psi_0(\mathbf{r})$$

where $\Psi_0(\mathbf{r})$ is the wavefield incident on the sample. [For perfectly coherent radiation from a point source, collimated by a lens, $\Psi_0(\mathbf{r})$ is approximately a plane-wave.]

Transmission functions for X-ray and electron diffraction are derived below in terms of the wanted object properties $\rho(\mathbf{R})$ or $V(\mathbf{R})$. We define the *image* as any magnified, resolution-limited or aberrated copy of $\Psi_0(\mathbf{r})$, formed by a lens. Unfortunately this term is now widely used in the CDI literature as a synonym for either *object* or *exit-face wavefunction*. This use of “image” to refer to an exit-face wavefunction is now firmly established and will be continued here; however, the more important distinction between object and exit-face wavefunction is likely to

become important in future work, and should be preserved. The terms “real object” and “complex object” have also been widely adopted in optics, but should strictly refer to the nature of the exit-face wavefunction for two-dimensional imaging. [For a strong phase object described by Eq. (1), the object Δn may be real, but the exit-face wavefunction complex. Most authors would refer to this as a “complex object,” however, the object property we wish to recover is $\Delta n(\mathbf{r})$.] It is also useful to reserve the word *image* for real-space functions, and *diffraction* for reciprocal space, so that the term “diffraction image,” when referring to a diffraction pattern, should be avoided. The measured diffraction pattern intensity is $I(\mathbf{u}) = |\Phi(\mathbf{u})|^2$, with $\Phi(\mathbf{u})$ the Fourier transform of $\Psi(\mathbf{r})$ and scattering vector $|\mathbf{u}| = \Theta/\lambda$ for small scattering angles Θ . (Here λ is the X-ray, or relativistically corrected de Broglie wavelength for electrons.) $I(\mathbf{u}) = I(-\mathbf{u})$ if $\Psi(\mathbf{r})$ is real, or, for complex objects, if $\Psi(\mathbf{r}) = \Psi(-\mathbf{r})$. $\Phi(\mathbf{u}) = \Phi^*(-\mathbf{u})$ for real $\Psi(\mathbf{r})$, while $\Phi(\mathbf{u})$ is a signed real quantity if $\Psi(\mathbf{r}) = \Psi(-\mathbf{r})$ and $\Psi(\mathbf{r})$ is real. The aim of diffractive imaging is to reconstruct the *object* from the scattered intensity $I(\mathbf{u})$; however, as a first step the exit-face wavefunction $\Psi(\mathbf{r})$, which is simply related to $\Phi(\mathbf{u})$ by a Fourier transform, is obtained. [This reduces to determination of the sign of $\Phi(\mathbf{u})$ if $\Psi(\mathbf{r}) = \Psi(-\mathbf{r})$ and $\Psi(\mathbf{r})$ is real.] The further recovery of object properties $\rho(\mathbf{R})$ or $V(\mathbf{R})$ from $\Psi(\mathbf{r})$ may be possible only in the absence of multiple scattering or inelastic scattering. For electron diffraction, the weak-phase approximation generates a “real object” in the language of optics and diffractive imaging, as described below. For X-ray diffraction, where single-scattering conditions are common, the absence of spatially dependent absorption (due to the photoelectric effect) provides such a “real object,” so that imaging should be performed at energies that avoid absorption edges for any elements present if phase contrast is expected. A single, spatially uniform absorptive process, however, may allow the phase contrast formulation to be used. The introduction of the transmission function allows a simple extension to the case of coherent convergent-beam illumination and related methods for phasing (Spence and Zuo, 1992).

We now relate $\Psi(\mathbf{r})$ to the wanted object properties $\rho(\mathbf{R})$ or $V(\mathbf{R})$ for the case of visible light, electron beams, and X-rays in the projection, iconal, or “flat Ewald sphere” approximation. Both refractive and dissipative (inelastic) processes may occur. In each case it is necessary to consider whether an iconal or projection approximation may be made, and the question of whether three-dimensional (tomographic) information (discussed later) may be extracted.

The simplest case for each radiation is that in which the projection approximation holds, in which case the image $\Psi(\mathbf{r})$ may be treated as a simple projection of some property of the sample, taken in the beam direction. Then, if a transmission sample in the form of a thin plate of thickness t is illuminated by a plane-wave,

$$\psi(\mathbf{r}) = T(\mathbf{r})\psi_0(\mathbf{r}) = \exp[-2\pi i\Delta n_p(\mathbf{r})/\lambda]\exp(-2\pi i\mathbf{u}_0 \cdot \mathbf{r}) \quad (1)$$

For normal-incidence plane-wave illumination $\mathbf{u}_0 = 0$, and we may set $\Psi_0(\mathbf{r}) = 1$, and $\Psi(\mathbf{r}) = \exp[-i\theta(\mathbf{r})]$. Here Δn_p is proportional to the

complex refractive index of the sample for the radiation concerned. We see that a real (“mask-like”) object can be obtained only if the real part of Δn_p is independent of $\mathbf{r}$, and all structural information is contained in the imaginary part. These experimental conditions (pure “absorption” contrast) can be obtained only at relatively low spatial resolution (under incoherent conditions) for both electrons and X-rays.

For X-rays,

$$\Delta n_p(\mathbf{r}) = \int_0^t [\delta(\mathbf{R}) - i\beta(\mathbf{R})] dz \quad (2)$$

where d is a positive quantity (Kirz et al., 1995). In terms of mean values, the complex index of refraction for X-rays is $n = 1 - \Delta n_p = (1 - \delta) + i\beta$, where δ describes refraction and β absorption (mainly the photo-electric effect, arising from absorption edges). The linear absorption coefficient is $\mu = 4\pi\beta/\lambda$. The dependence of $\Delta n(\mathbf{r})$ on the real and imaginary parts of the atomic scattering factors f and f' is given by $\delta = (r_e\lambda^2/2\pi)n_a f$, and $\beta = (r_e\lambda^2/2\pi)n_a f'$, with n_a atoms per unit volume and r_e the classical electron radius. Away from absorption edges, the electronic charge density (excluding the nuclear contribution) is

$$\rho(\mathbf{R}) = \left(\frac{2\pi}{r_e\lambda^2} \right) \delta(\mathbf{R}) \quad (3)$$

If small bonding effects are ignored, $\rho(\mathbf{R})$ is obtainable from tabulated X-ray scattering factors for neutral atoms.

Since δ is about 10^{-3} at 6 kV for light materials, a thickness of about $0.3\mu\text{m}$ of sapphire is needed to obtain a phase shift of $\pi/2$, allowing a first-order expansion of Eq.(1). Then the diffracted amplitudes are simply proportional to the Fourier transform of the projected charge density of the object.

For an electron beam of kinetic energy eV_0 ,

$$\Delta n_p(\mathbf{r}) = \int_0^t \frac{V_c(\mathbf{R})}{2V_0} dz \quad (4)$$

where $V_c(\mathbf{R})$ is the complex “optical” potential for high energy electrons (Radi, 1970; Howie and Stern, 1972), and the mean refractive index for electrons is $n = 1 + \Delta n$. The real part of this, for high energies, is the positive electrostatic or Coloumb potential, also obtainable from X-ray scattering factors using Poisson’s equation (Spence and Zuo, 1992). The imaginary part (typically about a tenth of the real part) accounts for depletion of the elastic wavefield by inelastic scattering events such as plasmon, inner-shell, and phonon scattering. The average value V_0 of $\text{Re}(V_c)$ is about 12 eV for light materials, and is positive, so this may be used as a constraint. Both real and imaginary parts of the potential $V_c(\mathbf{R})$ are positive if the mean inner potential is included, since electron beams are attracted predominantly to the positive nuclei. Electron diffraction in the transmission geometry is not, however, sensitive to the mean potential V_0 , which produces only a constant phase shift. The mean inner potential can be meaningfully defined only for a finite

object with zero total charge (O'Keeffe and Spence, 1993). Although the potential due to an (unphysical) isolated negative ion may have small negative excursions, the mean value depends on the volume assumed, and, in a real crystal, any long range ionic potential is also screened. In summary, in the absence of its mean value, the real part of the optical potential used to describe electron diffraction (when synthesized from diffraction data, with resolution limited by a temperature factor) is positive, except for possible very small negative excursions around negative ions. The positivity of the imaginary part is guaranteed by the requirement that energy gain is forbidden in inelastic processes if very small virtual processes are ignored. Fourier coefficients of the total optical potential may have either sign, and those of the real (elastic) and imaginary (inelastic) potential become mixed in objects without inversion symmetry. Hence a sign condition on the optical potential may be used as a convex constraint unless atomic resolution reconstructions are attempted of sufficiently high accuracy to detect the very weak bonding effects.

By estimating the maximum value of $[\text{Re}(V_c) - V_0]$, the maximum thickness can be estimated for which a first-order expansion of Eq.(1) may be made, the weak-phase approximation, in which $\theta(\mathbf{r}) \ll \pi/2$. This is also a limited case of the single scattering approximation. For X-rays, the maximum thickness allowable in the weak-phase object approximation can be estimated using the CXRO web calculator page supported by the Lawrence Berkeley Laboratory at <http://www-cxro.lbl.gov/>. For electrons the Fourier coefficients of electrostatic potential published for many materials by Radi may be useful (Radi, 1970), since these also provide an estimate of the imaginary part of the optical potential for electron diffraction due to inelastic scattering.

For visible light, $\Delta n_p(r)$ has a similar interpretation as for X-rays (with n given by the square root of the complex dielectric constant), however, $n > 1$. Thus visible light and electrons ($n > 1$) are bent toward the normal on entering a denser medium, while X-rays undergo total external reflection, with $n < 1$.

In summary, if inelastic processes are neglected (away from absorption edges), at high energies, transmission samples of thickness t are phase objects for electrons and X-rays, for which the refractive index is proportional to the electron density for X-rays and to the total electrostatic potential, including the nuclear contribution, for electrons. Poisson's equation relates these. The magnitudes of these quantities are such that a first-order expansion of Eq.(1) (weak phase object approximation) is justified $[2\pi i n_p(\mathbf{r})/\lambda < \pi/2]$ if $t < 20\text{ nm}$ for electrons (light elements, $V_0 = 200\text{ kV}$) but $t < 0.3\text{ }\mu\text{m}$ for X-rays (light inorganic material, 6 kV). Higher order terms in the expansion of Eq.(1) correspond to the multiple-scattering terms of the Born series in a "flat Ewald sphere" approximation.

This "flat sphere" or projection approximation, on which Eq.(1) is based, depends on the ratio of wavelength to smallest detail d of interest (with spatial frequency $u = 1/d$), and on the thickness of the sample. Scattering kinematics restrict elastic scattering to regions of reciprocal space near the energy and momentum-conserving Ewald sphere of

radius $1/\lambda$. The projection approximation holds if the “excitation error” distance $S_u \approx \lambda u^2/2$ from this sphere onto a plane in reciprocal space normal to the beam (passing through the origin) is less than either $1/t$ or $1/\xi_u$, whichever is the smallest. (ξ_u is a multiple-scattering extinction distance for spatial frequency u . Thus samples never look thicker than ξ_u to diffracting radiation.) Hence we require

$$\frac{\lambda u^2}{2} < \frac{1}{t} \text{ or } \frac{1}{\xi_u} \quad (5)$$

or

$$\left(\frac{\lambda t}{2}\right)^{1/2} < d \quad (6)$$

for the validity of Eq.(1). Since the width of the first Fresnel fringe due to propagation over distance t is approximately $w = (\lambda t/2)^{1/2}$, this condition requires that the spreading of the wavefield due to free-space propagation over a distance equal to the thickness of the sample be small compared to the resolution required.

The failure of Eq.(1) may require either single or multiple scattering treatments, depending on the strength of the interaction and sample thickness. We note that Eq.(1) is an exact solution that sums the Born series, including all multiple scattering effects, in the limit of vanishing wavelength. For X-ray tomography, use of Eq.(1) greatly simplifies the collection of data for three-dimensional reconstruction.

The preceding discussion has concerned two-dimensional imaging. For tomography there is a different analysis. Again, a single-scattering approximation must be used. But the introduction of a curved Ewald sphere does not prevent three-dimensional reconstruction (Miao et al., 2002), since data may be collected for various object orientations, and in each case assigned to points on the sphere until all of the reciprocal space volume is filled. (Experimentally this is done by rotating the sample about an axis normal to the beam and recording a diffraction pattern at each orientation.) Use of the Fienup algorithm with three-dimensional Fourier transform iterations can then take advantage of the improved convergence in three dimensions. A sign constraint may or may not be applied to both real and imaginary parts of the scattering potential, and a general spherical support, enclosing the object, has been found useful.

4 The HIO Algorithm and Its Variants

We now assume the weak-phase approximation, in which the exit-face wavefunction is computed along a single optical path along the beam direction, and no multiple scattering or inelastic processes are permitted other than an overall, spatially independent exponential attenuation with thickness. The recorded intensity of the diffraction pattern, excluding the central portion, is then given as

$$I(\mathbf{u}) = \Phi(\mathbf{u})\Phi(\mathbf{u})^* = |FT(\Psi(\mathbf{r}))|^2 \quad (7)$$

If the phase of the Fourier transform $FT[\Psi(\mathbf{r})]$ could be recovered, than the aberration-free complex exit wave $\Psi(\mathbf{r})$ could be reconstructed. Aberrations of any diffraction lenses in the electron case that magnify the diffraction pattern (astigmatism, distortion, etc.) could alter the intensity distribution $I(\mathbf{u})$ and so complicate recovery.

For nonperiodic objects the phase problem can now be solved if the object has a finite support, i.e., the image is nonzero only within a finite region of image space and if the image is real and nonnegative. For a weak phase object the exit-face wavefunction is $\Psi(\mathbf{r}) \sim 1 - i\theta(\mathbf{r})$ where $\theta(\mathbf{r})$ may be treated as real and positive. The diffracted amplitude is $\Psi(\mathbf{u}) = \delta(u) - i\Phi(u)$, for which the scattered part differs in phase by 90° from the direct beam given by the first term. This direct “unscattered” beam is absorbed by the beam-stop; difficulties arising from the finite size of this are discussed later, and we note that the average value of $\theta(\mathbf{r})$ is also lost within the beam stop. Including a larger area around the object known a priori to have zero or unit transmission function is equivalent to “oversampling” in Fourier space, and the spatial coherence of the illumination must span this larger area, at least equal to that of the autocorrelation function of the object. Hence diffractive imaging requires spatial coherence twice that of coherent imaging with a lens (Spence et al., 2004), increasing exposure time by a factor of four for a given source and resolution limit.

We assume that $|FT[\Psi(x, y)]|$ has been measured, and that the support $S(x, y)$ of $\theta(\mathbf{r})$ (which is the same as that of the object) is known. $S(x, y)$ is the region outside which the object is known to be zero. The iterations start with an initial estimate $\tilde{G}_1(u, v) = |FT[\Psi(x, y)]| \exp[i\theta_1(u, v)]$ of the spectrum. $\theta_1(u, v)$ is chosen to be an array of independent pseudorandom real numbers distributed between 0 and 2π . The iterative Fourier transform algorithm consists of the following steps (with subscript k labeling quantities at the k th iteration):

1. Inverse Fourier transform $\tilde{G}_k(u, v)$ to obtain the image $\tilde{g}_k(x, y)$
2. Define $g_{k+1}(x, y)$ as

$$g_{k+1}(x, y) = \begin{cases} \tilde{g}_k(x, y) & \text{if } (x, y) \in S(x, y) \\ g_k(x, y) - \beta \tilde{g}_k(x, y) & \text{if } (x, y) \notin S(x, y) \end{cases} \quad (8)$$

This constitutes the HIO version of the algorithm. β is a constant chosen between 0.5 and 1. In the ER version of the algorithm this step is replaced by

$$g_{k+1}(x, y) = \begin{cases} \tilde{g}_k(x, y) & \text{if } (x, y) \in S(x, y) \\ 0 & \text{if } (x, y) \notin S(x, y) \end{cases} \quad (9)$$

3. Fourier transform $g_{k+1}(x, y)$ to obtain $G_{k+1}(u, v)$
4. Define new Fourier domain function $\tilde{G}_{k+1}(u, v)$ using the known Fourier modulus $|FT[\Psi(x, y)]|$ with the computed phase:

$$\tilde{G}_{k+1}(u, v) = |FT[\Psi(x, y)]| \exp[i\theta_{k+1}(u, v)]$$

5. Go to step 1 with k replaced by $(k + 1)$.

To monitor the progress of the algorithm, the object space error metric ε_k is calculated during each iteration:

$$\epsilon_k = \frac{\sum_{(x,y) \notin S} |\tilde{g}_k(x,y)|^2}{\sum_{(x,y)} |\tilde{g}_k(x,y)|^2} \quad (10)$$

ϵ_k is the amount by which the reconstructed image violates the image-space constraints. Physically, in the X-ray case, it is the normalized amount of charge that remains outside the boundary of the object, which should be zero. In all our calculations we have used $\beta = 0.7$ and a combination of the ER and HIO algorithms, with 20 ER iterations followed by 50 HIO cycles, all repeated until the error ϵ_k drops below a certain level. Simulations based on the above procedure with small noise levels invariably converge to the correct solution (Weierstall et al., 2001; Spence et al., 2002). Since $g(x, y)$, $g^*(-x - a_1, -y - a_2)\exp(i\theta)$ and $g(x - a_1, y - a_2)\exp(i\theta)$ all have the same Fourier modulus they cannot be distinguished by the algorithm. Each run of the algorithm started with different random phases may thus produce images centered on different origins or related by inversion symmetry. We call these equivalent images. (Similarly, in three dimensions, enantiomorphs cannot be distinguished.)

Some understanding of the inversion can be obtained by considering the Fourier equations relating the object to its diffraction pattern. We reverse the domains originally considered by Shannon, and treat the object space as if it applies a “bandlimit” to the diffraction pattern (the object is assumed compact). Shannon’s theorem then specifies either the sampling interval on the diffraction pattern *intensity* needed to fully reconstruct the autocorrelation function of the object, or the sampling of the complex scattered amplitude needed to reconstruct the object. Consider a simple one-dimensional complex exit-face wavefunction $f(x)$, which is nonzero only for $0 < x < W$, so the support has width W . We first treat this function as the “bandlimit” on the diffraction pattern $F(u)$, which must therefore be sampled at intervals $u_n = n/W$ by the detector to satisfy Shannon’s theorem. We then have the N equations (one for each pixel in the detector)

$$|F(u_n)| = \left| \sum_{i=1}^N f(x_i) \exp(2\pi i n x_i / W) \right| \quad n = 1 \dots N$$

relating measured intensities $|F|$ to the $2N$ unknown complex values of $f(x_i)$ we seek. Since there are more unknowns than equations, the complex values of $f(x)$ cannot be found from these measurements.

However, now consider the same function placed within a domain of width $2W$, so that the values of $f(x)$ in $W < x < 2W$ are known a priori to be zero. The sampling interval on $F(u)$ is now $1/(2W)$ and we thus have $2N$ equations, but the number of unknown values of $f(x)$ remains at $2N$, so that the system of equations now becomes solvable in principle. Loosely speaking, we compensate for the missing half of the data in the diffraction domain (the phases) by requiring that half of the object values be known (they are zero outside the support of width W). Since the HIO algorithm assigns a random set of phases initially, the

equations are also linear within the algorithm unless a sign constraint is applied. Experimentally, real data from a CCD may not provide independent equations (symmetrical support shapes also lead to dependent equations) and the effects of noise must be considered. Large systems of nonlinear equations are normally extremely difficult or impossible to solve, so it is remarkable that the HIO algorithm does this so effectively, for reasons that are not fully understood, since few methods exist for analyzing nonconvex optimization. However, although the Fourier modulus constraint is nonconvex, it is known that the error surface in hyperspace is not a rugged landscape, but consists of a single rather smooth minimum with a few small bumps. Use of a boundary estimate (support) without symmetry also drives the algorithm toward one particular enantiomorph $\rho(\mathbf{r})$ rather than a spurious mixture involving $\rho(-\mathbf{r})$.

This example may be extended to other cases: for complex, two-dimensional objects there are $2N^2$ unknowns but N^2 equations (if the CCD has linear pixel dimension N). But by placing empty space around our compact object, which increases the diffracting volume by $2^{1/2}$ in both dimensions, we sample the diffraction pattern more finely and recover the $2N^2$ equations needed to solve the phase problem. It is the knowledge of the support (the object boundary) that ensures that known object pixel values may be inserted in the algorithm correctly outside the support. Experimentally, this creates the greatest difficulty of the diffractive imaging method; it is necessary to know a priori that the diffraction pattern comes from *an isolated object whose size is approximately known*. The existence of any material outside the assumed support that unwittingly contributes to the diffraction pattern will result in inconsistent constraints being applied by the algorithm, which will then not converge.

When experimental data are analyzed, the physical support in the object may not be known accurately. To avoid confusion we call the support actually present in the experiment the “physical support” and the support estimate used in the data analysis the “computational support.” The term “loose support” describes the situation in which the computational support is larger than the physical support. “Tight support” means the physical support is the same as the experimental support. For two-dimensional simulations with real objects, it is found that a “default” triangular-shaped support may invariably be used, if it is chosen to lack any symmetry and enclose the object. With experimental data, convergence may be slower (or nonexistent) due to high noise levels, the absence of data around the origin at the beam-stop, and excessively loose support. Then the simplest initial choice of support is the boundary of the autocorrelation function (obtained by Fourier transform of the diffracted intensity). This estimate is rapidly improved upon by the shrinkwrap algorithm, which uses a specified intensity threshold to find an improved smaller boundary at every iteration of the HIO algorithm (Marchesini et al., 2003b). This appears to be the most useful practical algorithm at present. Depending on noise levels, the support estimate has been found to be sufficiently accurate to deal with complex objects in many cases.

We conclude this section with some comments on additional constraints, the beam stop problem, support determination, and recent algorithm developments. In the general case of strong multiple scattering no simple closed-form expression relates the object to the exit-face wavefunction, and the two are related only by symmetry constraints. For a general (“strong”) phase object [Eq. (1)], a unit modulus constraint may be applied, so that reconstructed pixels are forced to lie on a unit circle on an Argand diagram; however, this constraint is non-convex and so has not been found very useful in practice. For such an object, to which a spatially independent absorption term is added [so that $\rho(\mathbf{R})$ or $V_C(\mathbf{R})$ has a constant known imaginary part], the reconstructed image pixels may be constrained to lie on a given spiral on an Argand diagram. A summary of the many constraints that have been tried and experimental results is given in Weierstall et al. (2001). Other constraints include atomicity, symmetry (most useful for reducing computing time), imposition of a known histogram of gray-levels for the object density, and low-resolution imaging by a different technique to provide a support estimate. In the following section the use of prepared objects is demonstrated to allow the method of Fourier transform holography to be used to provide a support estimate. The reference object need not be a simple point scatterer (He et al., 2004), and may have an extended complicated known shape (Szoke, 1997). The shrink-wrap algorithm described below rapidly improves iteratively on any initial support estimate and is found to converge for both real and complex objects under a wide range of conditions. For arrays of semiconductor devices the support will often be known, so that tomographic imaging of defects within an array element might be based on a support provided by the lithography pattern used to make the array.

A remarkable new “flipping” algorithm has recently appeared (Oszlanyi and Suto, 2004) for the crystallographic phase problem, which may also be adapted for nonperiodic objects, in which case it reduces to Fienup’s output–output algorithm with feedback parameter $\beta = 2$ and a dynamic support defined by an adjustable threshold (Wu et al., 2004b). This algorithm operates as follows: First, random phases are assigned to the structure factors, which must extend to atomic resolution. These are transformed to yield a real charge density. All density values below a certain threshold have their sign reversed. The result is transformed, and Fourier magnitudes are replaced with measured values. The process is continued to convergence. This algorithm (far simpler than the direct methods normally used in crystallography) has been used to solve new crystal structures from X-ray data (Wu et al., 2004a), and is found to perform very well. It does not require a support estimate or knowledge of atomic scattering factors, but does require atomic-resolution data. The “atomicity” constraint in crystallography assumes that the solution density consists of a set of smooth peaks, and requires that diffraction data extend to atomic resolution. The support then consists of spheres around each atom; most of the density within a crystal consists of empty space between atoms, akin to the zero-density band generated by oversampling around an object in the HIO

algorithm. Since all matter in the cold universe consists of atoms (whose scattering factors are known), this constraint provides an extremely powerful *ab initio* assumption if one has atomic-resolution data, and is the basis of the direct methods algorithms of X-ray crystallography. Other known “building blocks” (such as gold balls or lithographed dots) have been used to reduce the number of unknown parameters in image definition (Spence et al., 2003b). (For proteins, for example, both the sequence and the atomic structure of the 20 amino acids of which they consist are usually known a priori, together with a typical gray-level histogram for the density maps.) An algorithm that has used the atomicity idea for nonperiodic data is Speden (Hau-Riege et al., 2004), which has been applied to the data of Figure 19–3. Finally, we note the use of the HIO algorithm for two-dimensional protein crystals in electron microscopy. This technique uses Fourier transforms of conventional weak-phase-object images (formed with a magnetic lens) to provide the phases of structure factors, and diffraction patterns for their magnitudes. The subnanometer resolution images must be obtained over as large a range of tilts as possible, creating serious experimental difficulty. These two-dimensional monolayer crystals are nonperiodic in the direction normal to the plane of the crystal, so that lines of diffraction are generated in reciprocal space. By oversampling along these lines it is possible to phase these reciprocal lattice rods individually. The phase relation linking them can then be obtained from a few high-resolution images recorded at small tilts. In this way the number of electron microscope images needed for three-dimensional imaging of two-dimensional organic crystals can be greatly reduced, with most of the reconstruction based on easy-to-obtain diffraction data (Spence et al., 2003a).

We note in passing that the nonuniqueness of the crystallographic phase problem has been well studied; the so called “homometric” crystal structures studied by Pauling, Burger, and others have the same diffraction patterns, but different structures. (They are not enantiomorphs.) Fortunately these are very rare.

Several approaches have been made to the problem of data lost behind a synchrotron beam-stop, which is essential to protect a sensitive area detector. Any additional blooming can mean much loss of low-frequency data. In several HIO applications these missing values have simply been treated as free adjustable parameters, and the algorithm was found to converge. Calibrated absorption filters have been placed in front of the inner portion of the detector. Another solution is to use a sample consisting of an unknown object filling a small hole in an otherwise opaque mask (Eisebett et al., 2004; Weierstall et al., 2001). Stray scattering of X-rays from the edges of the holes in the mask can be a difficulty; however, surface roughness, which is smaller than the wavelength of the X-rays, produces little scattering. The use of very small silicon nitride windows (e.g., 2 μm width) greatly reduces the intensity of the direct beam and blooming effects; however, the detailed shape of the partially transparent silicon wedge around the window must then be modeled and used as a support for inversion. (When making samples of small particles deposited from solution, it is most

efficient to use a focused ion beam to remove all but a favorably located and isolated particle in the center of the window.) Finally, the “diffuse” X-ray scattering around Bragg peaks (which forms the “shape transform” of the object) from a crystallite has been inverted to an image, thus avoiding the direct-beam scattering (Robinson et al., 2001; Williams et al., 2003).

5 Experimental Results

Figure 19–1 shows the first X-ray images reconstructed by this lensless method in 1999 (Miao et al., 1999). The test object consists of letters formed from gold dots, 100 nm in diameter and 80 nm thick, on a transparent silicon nitride membrane. The transmission X-ray diffraction pattern formed with 1.7-nm monochromatic soft X-rays is shown in Figure 19–1A and the reconstructed image in Figure 19–1B. The object was illuminated through a coherently filled 10- μm -diameter pinhole, and a 25-cm camera length was used. Missing data from the central region within the beamstop were obtained from a lower-resolution optical image. The exposure time was 15 mins at the Brookhaven synchrotron. The Fienup algorithm was used for reconstruction, with a sign constraint applied to both real and imaginary parts of the scattering potential. A square support was used (irregular shapes usually work better) and 1000 iterations were needed for convergence. The resolution is about 75 nm. A detailed description of an improved version of the apparatus used to obtain this and other recent results at the Advanced Light Source is given in Beetz et al. (2005).

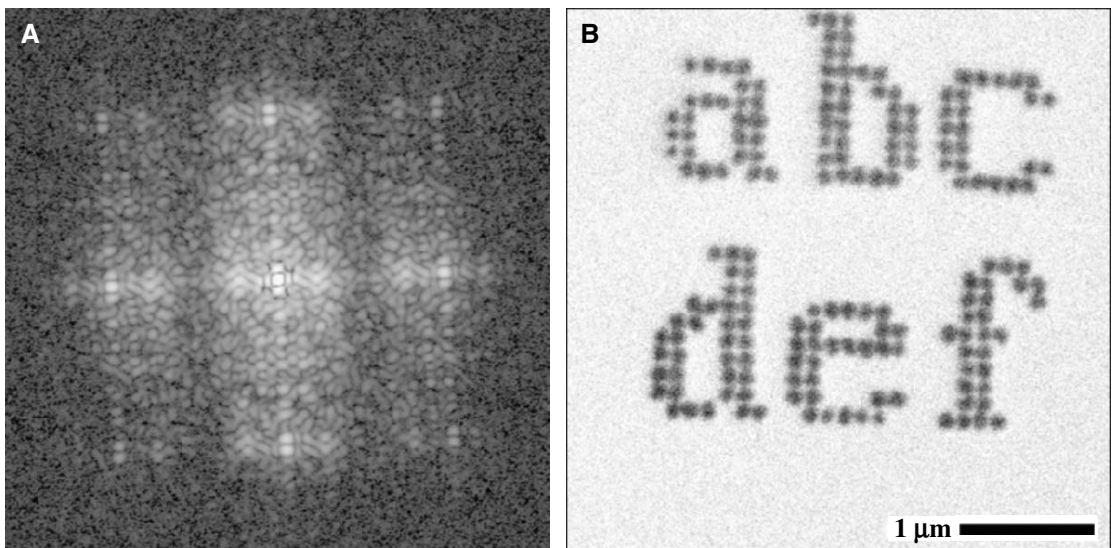


Figure 19–1. (A) Soft X-ray transmission diffraction pattern formed with 1.7-nm X-rays from the set of lithographed letters shown in (B). (B) Image recovered from (A) using a modified form of the HIO algorithm. (From Miao et al., 1999.)

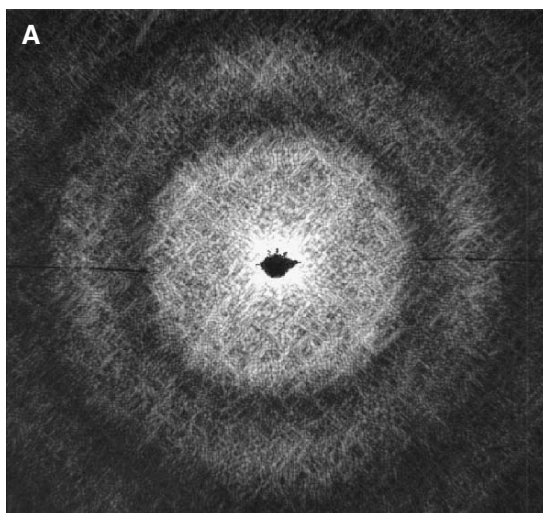
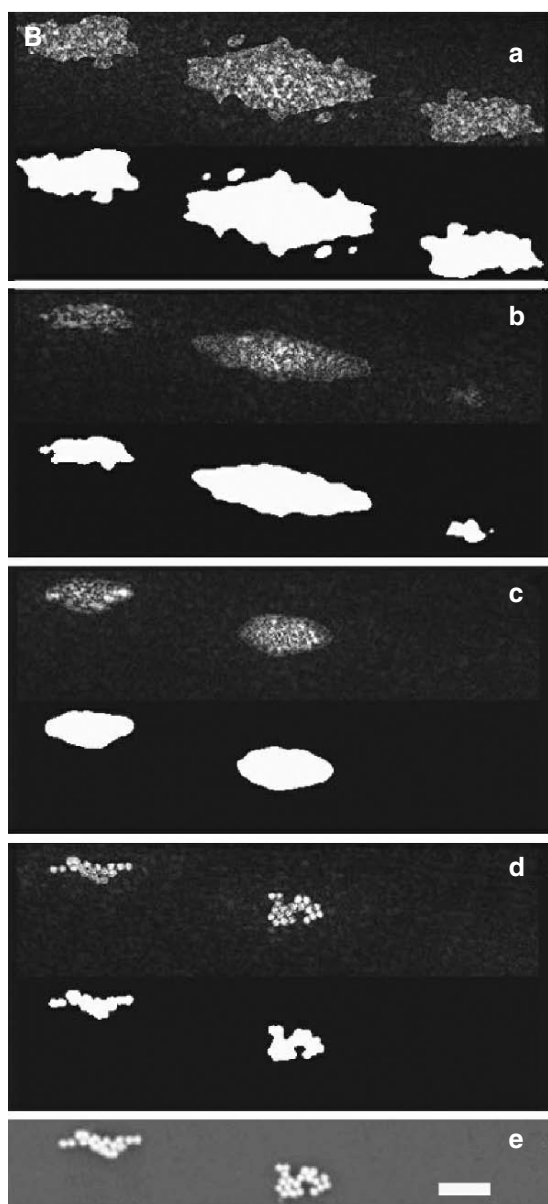


Figure 19-2. (A) Soft X-ray transmission diffraction pattern from clusters of 50-nm gold balls lying on a silicon nitride membrane. The X-ray wavelength is 2 nm (600 eV). The resolution at the midpoint of the sides corresponds to a spatial periodicity $u^{-1} = 17.4$ nm, or a Rayleigh resolution of 8.7 nm. (B) Image reconstruction series using the shrinkwrap algorithm, in which the HIO algorithm refines the support during iterations. The top image (a) is the centrosymmetric autocorrelation function, with the support estimate shown as a mask below. Intermediate iterations lead to a final converged image of gold ball clusters (e) shown at the bottom. The marker is 1 μ m in length. The inversion symmetry is lost at (c).

Figure 19-2A shows a transmission diffraction pattern obtained using 600-eV monochromatic soft X-rays from clusters of gold balls, 50 nm in diameter, lying on a silicon nitride membrane. The silicon nitride membrane is almost transparent to the X-rays, so the object provides a useful test object for reconstruction. The pattern resembles the Airy's disk-like pattern from one ball, crossed by "speckle" fringes due to interference between different balls. An image reconstruction series using the shrinkwrap algorithm, in which the HIO algorithm

refines the support during iterations, is shown in Figure 19–2B. The top image is the centrosymmetric autocorrelation function, with the support estimate shown as a mask below. This mask was obtained by Fourier transform of the diffraction pattern intensity (to produce the autocorrelation function shown), followed by the selection of a contour corresponding to a certain threshold of intensity. This thresholding operation is repeated after each HIO-ER iteration cycle, to generate a new improved estimate of the object support (see Marchesini et al., 2003b, for details). Intermediate iterations lead to the final converged image of the gold ball clusters (e) shown at the bottom. The marker is 1 μm in length. We note that the inversion symmetry necessarily possessed by the autocorrelation function at (a) is lost at (c) as it changes smoothly into the correctly phased image.

Figure 19–3 shows an instructive case, indicating the way in which “prepared objects” may be used to assist reconstruction. [Full experimental details for CXDI are given in He et al. (2003), from which Figure 19–3 is taken.] Figure 19–3A shows a scanning electron microscope (SEM) image of a set of gold balls lying on a silicon nitride membrane. One ball at A is isolated. The autocorrelation function obtained from an experimental soft X-ray transmission diffraction pattern (not shown) taken from this object is given in Figure 19–3B. This may be interpreted as the self-convolution of the object with its inverse, or, for a collection of point-like objects, as the set of all interpoint vectors. Some interball vectors are shown in Figure 19–3A and indicated again in Figure 19–3B. The convolution of the single isolated ball A with the three balls at B produces the autocorrelation function in Figure 19–3B a faithful image of the three balls, blurred by the image of one ball. This process is

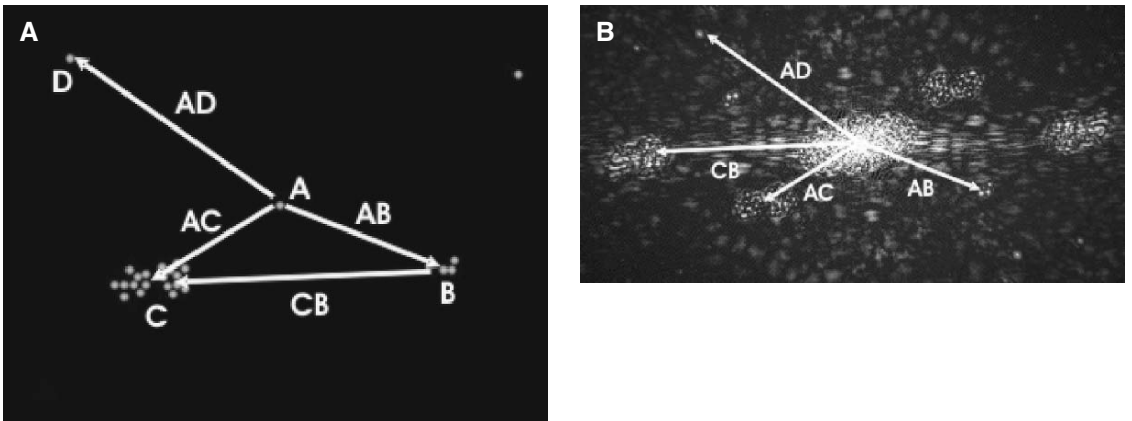


Figure 19–3. (A) SEM image of several clusters of gold balls, each 50nm in diameter. Some interball vectors are indicated. The balls lie on an X-ray transparent substrate. (B) The Fourier transform of the X-ray diffraction pattern taken from (A). This is the autocorrelation function of the density in (A) and is a map of all interball vectors or the self-convolution of the object with its inverse. Because the object in (A) includes a single isolated ball at A, the vector AB leads to a faithful image of the triple-ball cluster at B. (The convolution of one ball with three gives a blurred image of three.) (From He et al., 2003.)

similar to the heavy-atom method of X-ray crystallography, or the method of Fourier transform holography in optics (Collier et al., 1971). If such a gold ball or strong “point” scatterer can be placed near an unknown object, the autocorrelation function will contain a useful first estimate of the desired image of the unknown, which can also be used to provide a support for further HIO iterations aimed at improving resolution. This process is demonstrated in He et al. (2004), where, following the original suggestion of Stroke (1997), it is found that the resolution in the autocorrelation “image” may be considerably improved beyond the size of the reference ball by simple deconvolution. By using a larger reference object, or one consisting of a cluster of small balls, the intensity of scattering from the reference object can be increased. It has been noted that a randomly placed cluster of point scatterers can provide a high resolution image in Fourier transform holography when used as a reference object (see He et al., 2004; Collier et al., 1971; Eisebett et al., 2004, for more details and references).

The first successful application of CDI to the electron diffraction patterns provided by a transmission electron microscope (TEM) is described in Weierstall et al. (2001), where a complete description of the method can be found. An important asset of the TEM is its ability to provide an image of the same region that contributes to the micro-diffraction pattern, so that this image can be used to supply the support. However, electron scattering is so strong that any scattering contribution from a supporting film, however thin, is found to prevent successful CDI. The resolution of the best TEM instruments in direct phase-contrast imaging mode using lenses is now about 1 Å. Figure 19–4 shows a more recent application of CDI to an electron diffraction pattern using a TEM (Zuo et al., 2003). This remarkable image is the first atomic-resolution CDI image, and possibly the first atomic-resolution image of a nanotube. The image gives us the helicity of the tube, its dimensions, and the number of walls. The double-walled nanotube spans a hole in a thin amorphous carbon film, while the electron beam diameter (about 50 nm) is smaller than the hole, so that there is no background contribution from the carbon film. A conventional TEM image was used to provide the support function for HIO iterations along the edges of the tube, and it is suggested that the boundary of the support across the tube is provided by loss of coherence at the edge of the electron probe due to rapid phase variations arising from the aberrations of the probe-forming lens. (In general a tight support is desirable for CDI.) The image shows higher resolution detail than conventional TEM images of nanotubes. Resolution is limited perhaps only by the temperature factor, or by distortions in electron lenses used to magnify the diffraction pattern. It remains to be seen if tomographic imaging at atomic resolution is simplest by this method or by direct TEM imaging using lenses. If CDI is used, the difficult problem of supporting a nanoparticle for diffraction over a range of orientations will need to be solved. Radiation damage may be reduced in diffraction mode under some conditions.

Three-dimensional (tomographic) CDI of inorganic samples has now been demonstrated (Miao et al., 2002; Williams et al., 2003; Chapman

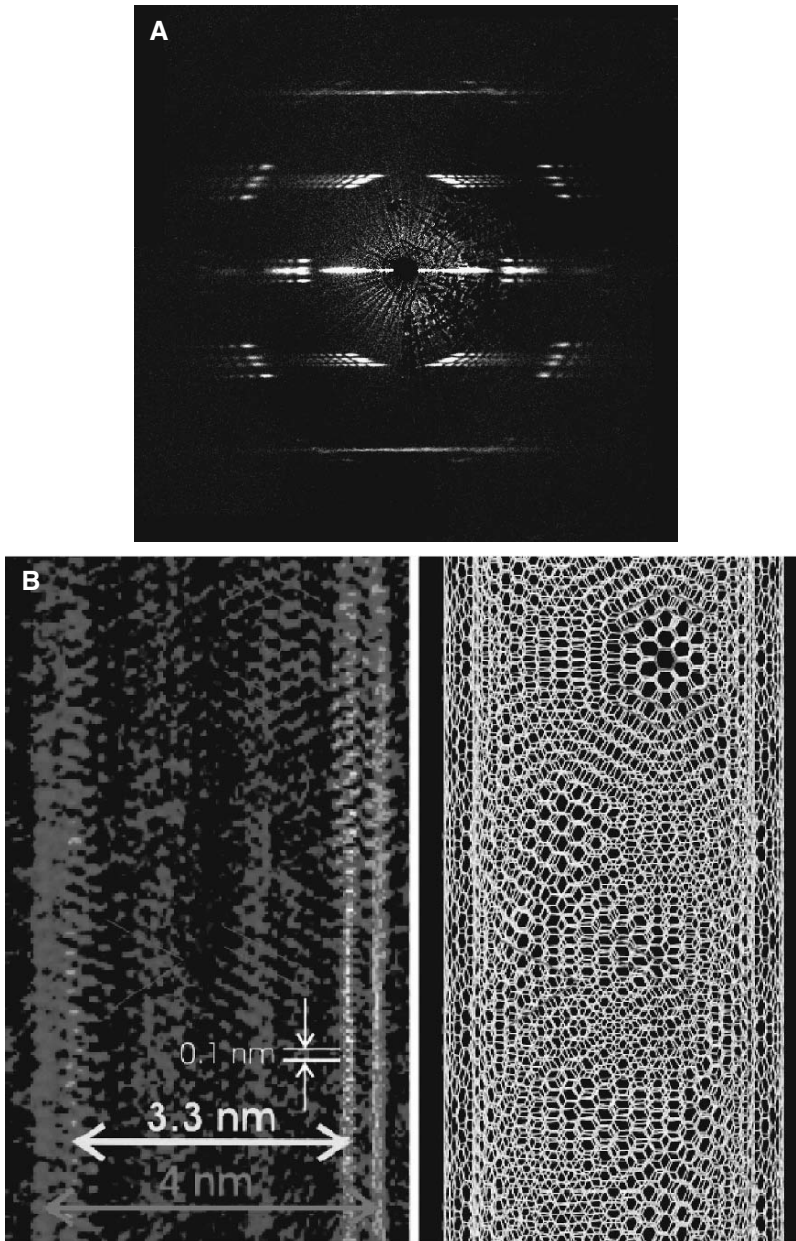


Figure 19-4. (A) Electron microdiffraction pattern from a single double-walled nanotube. This consists of a rolled-up sheet of graphite. Fine details arise from the helical structure (Zuo et al., 2003). (B) At left is the experimental image of the double-walled nanotube reconstructed from the electron diffraction pattern in (A). At right is shown a corresponding model of the structure (Zuo et al., 2003).

et al., 2006) using soft X-rays at a resolution of about 10 nm. This raises hopes of direct imaging, for example, of whole cells by this method if radiation damage considerations allow this at usefully high resolution. The X-ray source is a synchrotron and undulator, providing coherent

radiation at about 600 eV. In recent work (Chapman et al., 2006), a simple zone plate was used as a monochromator, following by a beam-defining aperture of about 10 μm in diameter, coherently filled. A nude soft X-ray CCD camera, employing 1024×1024 24- μm pixels was used. The sample is mounted in the center of a silicon nitride window fitted to a TEM single-tilt holder, which provides automated rotation about a single axis normal to the X-ray beam. The window is rectangular, with the long axis normal to both the beam and the holder axis. Diffraction patterns are recorded at 1° rotation increments, with a typical recording time of about 15 min per orientation. The maximum tilt angle is then limited by the thickness of the silicon frame around the window to perhaps 80° , resulting in a missing wedge of data. In addition, data may be missing around the axial beamstop. The development of software for automated tomographic diffraction data collection and merging is a large undertaking (Frank et al., 1996), and much can be learned from the prior experience of tomography in biological electron microscopy, where these techniques have been perfected (Frank, 2006). In that case, however, the registration of successive images at different tilts is greatly facilitated by direct observation of image features. The use of shadow images or X-ray zone-plate images for similar purposes has been suggested. With no direct imaging mode, much time is wasted in X-ray work locating the beam on the sample, which, with current CCD detectors, will typically be smaller than 2 μm in diameter. The final resolution (in one dimension), allowing for an “oversampling” factor of 2, will then be $4000/1024 = 3.9 \text{ nm}$. The camera length (sample-to-detector distance) of the diffraction camera must then be selected to allow half this spatial frequency to fall at the edge of the CCD camera at $u_{\text{max}} = \theta_{\text{max}}/\lambda = 0.5/3.9 \text{ nm}^{-1}$, so that the maximum scattering angle is $\theta_{\text{max}} = 0.25 \text{ rad}$ for $\lambda = 2 \text{ nm}$. Then the finest periodicity in the object ($3.9/0.5 \text{ nm}$) is sampled twice in every period, according to Shannon’s requirement (two points are required to define the period and amplitude of a sine wave if aliasing is excluded). For a CCD with linear pixel number N , the ratio of the finest detail to largest dimension is $N/2$, so that developments in detector technology limit CDI. The transverse spatial coherence of the beam must exceed 4 μm , as discussed together with monochromator requirements below.

Tomographic or three-dimensional imaging can provide the ability to “see inside” an object, but this requires that the intensity at a point in a projection be proportional to a line integral of some simple property of the object, such as the charge density. Then methods such as filtered back-projection can reassemble these two-dimensional projections into a volume density. Contours of equal density may then be isolated and presented to show the internal structure. For CXDI, a different approach is used, and some simplifications occur. It is no longer necessary to make the resolution-limiting “flat Ewald sphere” approximation, since diffraction data collected at one tilt can be assigned to points lying on the curved Ewald sphere in reciprocal space. (This is the momentum and energy-conserving sphere that describes elastic scattering in reciprocal space.) The sample is then rotated through this sphere around a single axis, until all of the reciprocal space is filled,

out to a given resolution. Three-dimensional interpolation of data points near the sphere is needed, and careful intensity scaling may be necessary if several exposures with different times are required to cover the full dynamic range of the data. It is often found that missing data points in the central region can be treated as adjustable parameters in the HIO iterations. Once a roughly spherical volume has been filled in reciprocal space (perhaps with missing wedge and beam-stop region), the three-dimensional iterations of the HIO algorithm may be applied [Eqs. (8), (9), etc., extended to three dimensions]. The computing demands are severe, as outlined below. The converged data will provide a three-dimensional density map, proportional to the local charge density, if the single-scattering approximation of X-ray diffraction theory applies and if the spatial variation in attenuation of the beam due to the photoelectric effect can be neglected. Figure 19–5 shows such a tomographic reconstruction, from which three-dimensional surfaces of constant density may be obtained. These surfaces allow us to “see inside” materials, and may eventually permit maps to be obtained that distinguish regions of different chemical composition.

The usefulness of tomographic CXDI in biology remains to be determined; at present the method appears to have the advantages over electron microscopy by allowing observation of thicker samples under a wider range of environments (for example, in the “water window” around 580 eV for soft X-rays). By comparison with X-ray zone-plate “full-field” imaging, the method allows a much larger numerical aperture to be used, and hence makes more efficient use of scattered

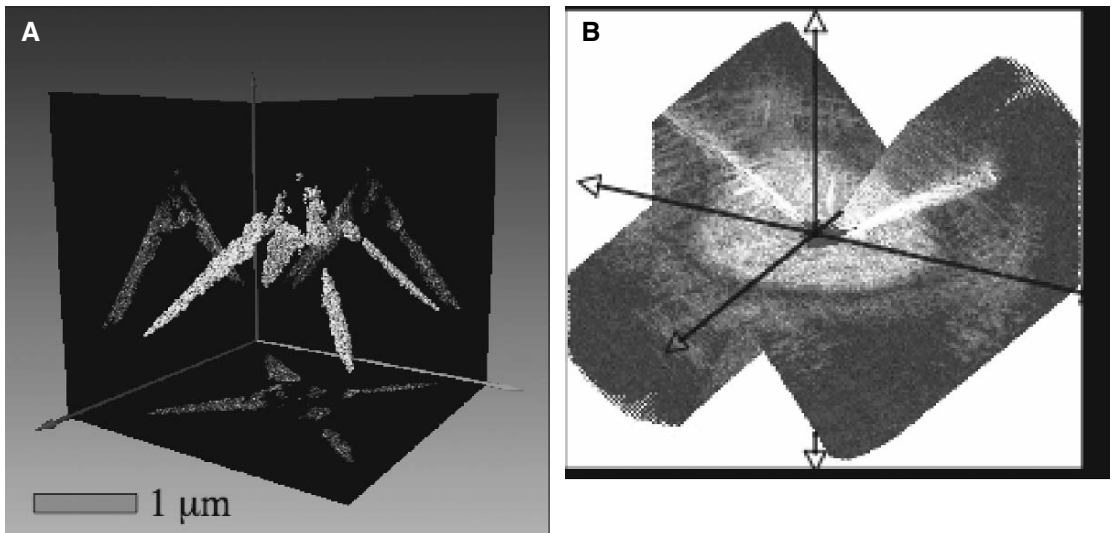


Figure 19–5. (A) Tomographic reconstruction from a soft X-ray diffraction pattern shown in (B). The object consists of gold balls (50 nm diameter) lying along the edges of a pyramidal-shaped silicon nitride structure. This is one image from a rotation series. From the complete series, three-dimensional surfaces of constant density can be constructed. (B) The volume of soft X-ray diffraction data collected to obtain the three-dimensional reconstruction in (A). (See color plate.)

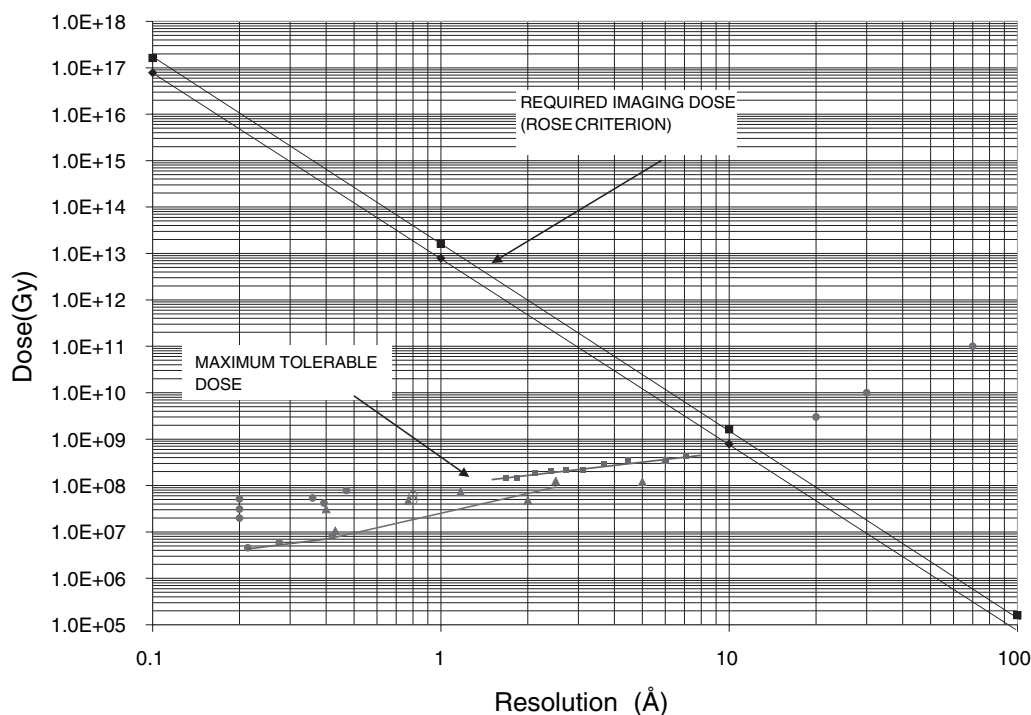


Figure 19–6. Summary of experimental measurement on organics for various microscopies on a plot of dose against resolution (Howells et al., 2005). Literature experimental values (e.g., Reimer, 1989) are shown as follows: Filled circles are from X-ray crystallography, filled triangles from electron crystallography, open circles from single-particle electron cryoelectron microscopy, open triangles from electron tomography, diamonds from soft X-ray microscopy, and filled squares from recent X-ray crystallography work on spot-fading experiments by Holton on the ribosome at the Advanced Light Source. Viable single-molecule microscopies must fall above the Rose equation line (to give a statistically significant image) and below the maximum tolerable damage line. Those below the Rose equation line succeeded by using crystallographic redundancy and form a periodically averaged image of perhaps 10^8 molecules in a crystal. The required imaging dose is calculated for a protein of empirical formula $\text{H}_{50}\text{C}_{30}\text{N}_9\text{O}_{10}\text{S}_1$ and density 1.35 g/cm^3 against a background of water, imaged with 10 keV X-rays (upper Rose line) and 1 keV (lower Rose line).

photons, while providing potentially higher resolution. At high resolution the depth of focus λ/θ^2 may become less than the sample thickness, which prevents tomographic reconstruction by back-projection methods based on simple projections. Then tomography may best be undertaken using optical sectioning rather than reconstruction from projections. CDI, based on three-dimensional diffraction data, provides a third alternative. The resolution limit imposed by radiation damage in CXDI, as expressed by the Rose equation (Spence, 2003), remains to be determined experimentally, but is likely to be significantly poorer than 1 nm, which has already been achieved in three-dimensional single-particle electron microscopy of proteins. Howells et al. (2005) and Marchesini et al. (2003a) provide a detailed discussion of this large subject, including a plot of dose against resolution for various microscopies in biology, as discussed in Section 9, shown in

Figure 19–6 (see also Henderson, 1995). The dose-fractionation theorem of Hegel and Hoppe is also relevant (Reimer, 1989). Recently, dramatic images of whole yeast cells have been imaged by CDI (Shapiro et al., 2005) using the apparatus described by Beetz et al. (2005).

6 Iterated Projections

A breakthrough in understanding the remarkable success of the HIO algorithm occurred in 1984, when Levi and Stark (1984) (based on earlier work by Youla and Webb, 1982) showed that the algorithm could be understood as successive Bregman projections between convex and nonconvex sets. Here an image is represented as a single vector $\mathbf{R}$ in an N -dimensional space, with one coordinate for each pixel. The addition of two such vectors adds together two images. Distance between images (vectors) in this space has the form of the familiar χ^2 goodness of fit index, so that similar images are near each other. The set of all images subject to a given constraint (e.g., known symmetry, known Fourier modulus, known sign of density, known support) is considered to occupy a volume in this space. The operation of taking a current estimate of the image, performing a Fourier transform, replacing the magnitudes of the diffracted amplitudes with the measured values, and inverse transforming was shown to be a projection onto the set of images subject to the Fourier modulus constraint. Vectors $\mathbf{R}$ between the boundaries of two constrained sets of images are considered. If it is shown that all the images within the only overlap between two constrained sets are equivalent solutions, then the phase problem reduces to finding this volume, where $\mathbf{R} = \chi^2$ and the ER error metric are a minimum. Constraints may be of two types—convex and nonconvex. For a convex set, all points on any line segment terminating within the set lie inside the set. A set $\mathbf{P}$ is convex if $\alpha\mathbf{R} + (1 - \alpha)\mathbf{R}'$ lies within $\mathbf{P}$ for all $\mathbf{R}$ and $\mathbf{R}'$. Here $0 < \alpha < 1$ is a scalar defining position along the line. In two dimensions, a kidney-shaped set is nonconvex and an ellipse is convex. Bregman has shown that iterative projections between convex sets must lead directly to a unique solution if it exists, without stagnation. In this manner the global optimization problem is solved without an exhaustive search for the case in which a unique solution and convex constraints are known to exist. For our problem the Fourier modulus constraint (the known diffraction intensities) is nonconvex, so that this approach has been of limited value. However, it provides a powerful geometric way of thinking about the algorithm as a trajectory in Hilbert space, which is usually drawn in two dimensions for simplicity. The effects of variations in feedback parameter β can be understood, convergence properties studied, and new algorithms proposed. For the ER algorithm, the path is a zig-zag between the boundaries of sets; for the HIO it is a spiral. Some desirable convex constraints include known support, a knowledge of phase rather than amplitude, the sign constraint, symmetry, a known histogram of density levels (Zhang and Main, 1990) (such as exists for proteins),

entropy minimization, and (for nonoverlapping atoms) atomicity. A list of constraints used in protein crystallography can be found in the relevant section of volume F of the international tables on crystallography. Application of these constraints thus avoids the common problem whereby optimization programs become trapped in local minima.

The application of constraints can be viewed as projections in the N -dimensional space. Recall that the support $\mathbf{S}$ is defined as the set of points for which the density is nonzero. For example, application of the support constraint $\mathbf{P}_s$ corresponds to setting many pixels to zero, that is, to projecting onto a space of lower dimension. For the Fourier modulus constraint, we note that Parseval's theorem ensures that distances in the N -dimensional real space are equal to those in a similar N -dimensional space of Fourier coefficients. Consider an Argand diagram for a particular Fourier component, for which the modulus constraint restricts solutions to a circle, whose radius is given by the measured value of the Fourier modulus. An estimate provided by the algorithm (e.g., outside this circle) must be projected (by operator $\mathbf{P}_m$) onto the nearest point on the circle, along a line that will pass through the origin. (This corresponds to the numerical process during one iteration of retaining the current phase estimate, but replacing the magnitude with the measured magnitude, in the HIO algorithm.) Since the linear addition of two vectors terminating on the circle does not produce a third that terminates on the circle, the modulus constraint is not convex. Note, however, that the addition of two vectors of arbitrary length but equal phase produces a new complex number with the same phase, so that a knowledge of phase is a convex constraint, and thus more powerful than a knowledge of amplitudes. The identity operation $\mathbf{I}$ is also useful, and a reflector operation $\mathbf{R}_s = 2\mathbf{P}_s - \mathbf{I}$ can be defined, which reverses the sign of the density outside the support $\mathbf{S}$. Using these operators, all the iterative algorithms can be represented simply and analyzed as alternating projections onto convex (and non-convex) sets (POCS). In this context, we may define these more limited projections as "projectors," which takes a given vector $\mathbf{R}$ to the nearest point of a nearby constrained set (usually on its boundary). Then

1. The error-reduction (ER) (Gerchberg–Saxton) algorithm may be written

$$\rho^{(n+1)} = \mathbf{P}_s \mathbf{P}_m \rho^{(n)}$$

2. The charge-flipping (CF) algorithm may be written

$$\rho^{(n+1)} = \mathbf{R}_s \mathbf{P}_m \rho^{(n)}$$

3. The hybrid input–output (HIO) algorithm may be written

$$\begin{aligned} \rho^{(n+1)}(r) &= \mathbf{P}_m \rho^{(n)}(r) \quad \text{if } r \in \mathbf{S} \\ \rho^{(n+1)}(r) &= (\mathbf{I} - \beta) \mathbf{P}_m \rho^{(n)}(r) \quad \text{if } r \notin \mathbf{S} \end{aligned}$$

4. The averaged successive reflections (ASR) algorithm may be written

$$\rho^{(n+1)} = 0.5(\mathbf{R}_s \mathbf{R}_m + \mathbf{I}) \rho^{(n)}$$

Similar descriptions of the difference map method (Elser, 2003), the hybrid projection reflection (HPR) (Bauschke et al., 2002), and the relaxed averaged alternating reflectors (RAAR) (Luke, 2005) have been given. For $\beta = 1$, the HIO, HPR, ASR, and RAAR algorithms are identical. A comparison of the performance of all of these, together with the powerful shrinkwrap algorithm (HIO with dynamic support) and simple geometric representations of the trajectory of the error metric for few-dimensional cases, can be found in Marchesini (2006).

Using this approach, it has also been shown that the HIO algorithm is equivalent to the Douglas–Rachford algorithm, and is related to classical convex optimization methods (Bauschke et al., 2002). The text by Stark (1987) is recommended as a tutorial introduction to this large subject.

7 Coherence Requirements for CDI: Resolution

It is readily shown (Spence et al., 2004) that the lateral or spatial coherence requirement for diffractive imaging is, in one dimension, that the coherence width $X_c \sim \lambda/\theta_c$ be at least equal to twice the largest lateral dimension W of the object. (This is similar to the requirement in crystallography that X_c exceed the dimensions of a primitive unit cell to avoid overlap of Bragg beams, with beam divergence θ_c . For phasing by the oversampling method, this cell must be about twice as large as the molecule.) This fixes the incident beam divergence and hence the exposure time for a given object size and source. Since at the unapertured diffraction limit ($\Theta = 90^\circ$) the resolution is approximately equal to the wavelength and about two pixels are required per resolution element, a total of about $(4X_c/\lambda)^2$ image pixels would be needed for a coherence width X_c and oversampling factor 2. Physically, this just means that the coherence patch must include the “known” region of vacuum (zero density) surrounding the object boundary (support). It is necessary to diffract coherently from an area twice as large as the isolated object of interest.

The temporal coherence length L_c is also important. For a field of view W at the object (so that the first oversampling point occurs at scattering angle λ/W) and finest (bandlimited) object spatial frequency d^{-1} , the optical path difference between points on opposite sides of the object and a distant detector point is $W \sin \theta = W\lambda/d$, which should not exceed the longitudinal coherence length for X-rays $L_c = \lambda E/\Delta E$. Hence the fractional energy spread allowable in the beam to record spatial frequency d^{-1} is $E/\Delta E > W/d = N$, where d is the sampling interval in the object and N the linear number of pixels needed to sample the object space in the HIO algorithm. A more detailed calculation, considering the shape of the temporal coherence function, gives the requirement on longitudinal coherence as about $E/\Delta E > N/3$, which improves on the estimate in Spence et al. (2004). This determines the quality of the monochromator needed. In practice values of $E/\Delta E = 500$ have yielded good results in soft X-ray work using CCD detectors with N^2 pixels, where $N = 1024$. Then the in-line arrangement of a simple

zone-plate monochromator can be used (Howells et al., 2002). For CDI using electron beams, the coherence requirements are easily met for the nanostructures of most interest. We note that the drive for higher resolution, for a fixed number of object pixels, reduces the demand on coherence. The number of pixels, especially in tomography, is likely to be limited by computer processing power to less than 1024^3 in the medium-term future, as discussed below.

It is important to devise a consistent definition of resolution in CDI. Each spatial frequency $u = d^{-1} = \theta/\lambda$ in the object diffracts energy at scattering angle θ into the far field. Consider a square area detector used for diffractive imaging for which the largest scattering angle into the midpoint of the side of the detector (not the corner) is $\theta_{\max}$. Then, in the small angle approximation, this angle defines a cutoff in the “transfer function” (see below) at $u_{\max} = \theta_{\max}/\lambda$, and the full period of the corresponding finest periodicity in the object that can be reconstructed is $d_{\min} = \lambda/\theta_{\max} = 1/u_{\max}$. This value of $d_{\min}$ has frequently been quoted as the resolution limit in CXDI. However, while it is a most important experimental parameter, it fails to consider the accuracy of the phasing process, and is not equal to the corresponding Rayleigh resolution limit. There are other considerations, which we now discuss.

The Rayleigh resolution limit was intended for the imaging of binary stars, which are incoherent point sources, unlike the phase contrast usually important for CXDI. If we do nevertheless wish to apply the Rayleigh condition to this situation, we may consider that the CXDI area detector in the far field plays the role of a square aperture in the back-focal plane of an ideal lens. The numerical aperture imposed on the reconstruction is then $\theta_{\max}$, and, if the reconstruction is perfect (no errors in phasing), the image will be given by the ideal object charge density convoluted with a sinc function amplitude (the impulse response for linear imaging), whose full width at half maximum is $d_{\min}/2 = 0.5/u_{\max}$. (The distance between first minima is $d_{\min}$. For a circular area detector the factor 0.5 becomes 0.61—the square detector does slightly better because of contributions from the corners.) Thus our adapted “Rayleigh resolution” is half the finest spatial periodicity, and is therefore equal to the sampling interval needed in real space for linear phase-contrast imaging, to avoid loss of information. Two samples are needed for every full periodicity in the object. (For an incoherent imaging model the impulse response becomes a sinc squared function, for which a sampling interval of $d_{\min}/4$ is needed, since the autocorrelation of the transform of the sinc squared function is a triangular function, doubling the bandwidth and resolution, as pointed out in Rayleigh’s original paper.) For a lens-based system with aberrations, the factor 0.5 depends on the aberrations of the lens, but takes its minimum value for the diffraction-limited CXDI case.

There are two further considerations. For phase contrast, the ability to distinguish adjacent small objects will depend on the phase shift each introduces, and thus the resolution becomes a property of the sample, not only of the instrument. It is then impossible to define resolution in a meaningful sample-independent manner. In coherent optics,

this problem is partly addressed by introducing the concept of a lens coherent transfer function (CTF). Such a function has been introduced in a way that also tests the reliability of the phasing process in an excellent recent proposal by V. Elser for a resolution definition for CXDI. The algorithm is repeatedly run to convergence, each time starting with a different set of random phases. The results for image contrast are plotted as a function of spatial frequency, showing, if noise is not too severe, a relatively smooth curve that falls to zero at some $u_{\max}$. (If the phasing process fails, the average of these many runs will be zero at each spatial frequency.) The resolution is then $0.5/u_{\max}$ if a square detector is used. This appears to be the best current definition of resolution for CXDI; however, it ignores the dependence of resolution on sample properties for phase contrast.

For a known object, the faithfulness of the reconstruction may be indicated by a cross-correlation function between the reconstructed estimate and the known object, or crystallographic R-factor, as discussed in detail elsewhere (Spence et al., 2003a). For an unknown object, the ER error metric ε defined in Eq. (10) has been shown by computational trials against known objects to vary monotonically with a cross-correlation function (Fienup, 1997). The best resolution achieved in CXDI is currently about 8 nm.

8 Computer Processing Demands

While noise-free data converge in a few iterations, several hundred iterations of the HIO algorithm are typically needed for convergence of good quality two-dimensional experimental data, or several thousand if many data are missing in three dimensions. Computer processing power can therefore impose serious limitations on CXI, especially for tomography. Recall that for an area detector of linear pixel dimension N , the ratio of largest to smallest feature size is $N/2$. In many cases we wish to obtain phase-contrast images of a real object, so that sample thickness is limited by the need to avoid spatial variations in absorption, which would lead to the possible complications of complex object restoration. (For CXDI, avoiding a beam energy near an inner-shell ionization edge may therefore be important.) A second limit on thickness may be set by the need to satisfy the weak-phase object condition, or avoid extinction effects, both of which introduce multiple scattering and a complex object. [The inversion of Eq. (1) for Δn_p is also referred to as the phase unwrapping problem.] Fixing this limiting thickness and N for a data cube then fixes the lowest resolution and largest field of view possible.

For $N = 1024$, a double-precision complex N^3 array occupies 16 gigabytes, and three or four of these must be accessible directly in RAM (plus a single-bit array for the support mask) to perform HIO iterations. For the Mac G5, a single fast Fourier transform (FFT) on this array takes about 1 min with the FFTW routine, although it is not possible to store sufficient data in RAM at present to work with a single processor, and hard disk transfers are prohibitively time consuming. Clusters of 16

dual processor G5 machines with TCP gigabit ethernet or Infiniband appear to be ideal, when times of approximately 15 and 7 s per iteration, respectively, are achieved for complex $1K \times 1K \times 1K$ arrays (Barty, 2005). Many procedures may be used to reduce the total data processing time. For example, working with some low-resolution 512^2 projections selected from the three-dimensional data set can provide useful support estimates using the two-dimensional shrinkwrap algorithm, before attempting full three-dimensional Fourier iterations. Thus overnight (or longer) computing times must be expected for tomographic CDI in the near future. Once the three-dimensional data set has been phased and transformed to real space, a variety of commercial tomographic viewing programs can be used to present the data, making it possible to “see inside” a nanostructure, such as a cell or metallic foam.

9 Summary

The past decade has been an exciting time for diffractive imaging. In it, we have seen the first successful applications of the phasing algorithms developed in the 1980s to experimental data, together with the development of many new algorithms, ideas, and stimulating intellectual discussion at workshops and conferences. The interdisciplinary nature of the subject has been striking and exciting. Experimental results have appeared for phase-contrast imaging by neutrons (Nugent, 2003), electrons (Weierstall et al., 2001), electrons at atomic resolution (Zuo et al., 2003), and X-rays in both two and three dimensions (Miao et al., 1999, 2002; Chapman et al., 2006). Applications of the method have been slower to develop, but already both in materials science and biology it is clear that the power of imaging *inside* nanostructures (going beyond the projection approximation) in three dimensions presents an exciting prospect for the future. Preliminary applications of phase-contrast X-ray imaging at lower resolution have included the imaging of nanostructures within composite materials, mesoporous silicates and foams, and the imaging of crack tips (Salvo et al., 2003). Bone is also of interest, since it consists of 20-nm nanocrystals of hydroxyapatite in a labyrinthine structure. There is every reason to suppose that the pursuit by tomographic CXDI imaging of these type of materials to higher resolution will be possible and of great interest to scientists. Using medium energy X-rays, the imaging of much thicker material should be possible than that studied by tomographic electron microscopy in materials science (Midgely et al., 2001). For CDI by electron diffraction in materials science the limiting problem is the method of sample support in the microscope. One obvious solution is to use a thin crystal of known structure (such as graphite) as the physical supporting membrane for the nanostructure. Then the known atom positions of the graphite may be used instead of the zero-density region in the HIO iterations. (This allows a connection with the phasing method of fragment completion in crystallography—the graphite atoms provide a kind of reference structure in projection.) Experiments along these lines are in progress. In biology, the situation will be clarified only when careful studies of radiation damage by spot-fading and other methods have been

completed. These are needed to determine the domain of applicability of the CXDI method in comparison with other microscopies. Figure 19–6 shows a plot dose against resolution. The domain of applicability of many microscopies is discussed on this figure (and on plots of resolution against thickness) by Howells et al. (2005), showing the niche for CXDI. Whole cell imaging has been pursued with both the zone-plate X-ray microscope and by cryoelectron microscopy, where, using the latter technique, a resolution of about 2 nm is possible in samples up to 100 nm thick or more. It seems likely that radiation damage will prevent competitive performance by CXDI; however, the method may provide useful images at perhaps 5 nm resolution in much thicker samples using medium energy X-rays in an environment of vitreous ice. The optimum choice of X-ray energy involves many issues, including the variation of synchrotron undulator brightness with beam energy and the variation in phase contrast with beam energy. While the coherent flux B available for a given synchrotron source brightness varies as λ^2 , the required fluence (from the X-ray cross section) A scales as λ^{-2} , so that the recording time A/B varies as λ^{-4} in a most unfavorable manner as X-ray beam energy increases. The dose in Grays needed to scatter a given number of photons into a voxel varies inversely as the fourth power of the resolution in tomography. An analysis of the variation of dose against resolution for several microscopies including CDI can be found in Marchesini et al. (2003a). Here the statistical demands of good imaging (based on the Rose equation) are compared with the maximum tolerable dose for a given resolution for single-particle imaging.

In summary, the demand for higher resolution three-dimensional noninvasive imaging with old and new radiation sources continues unabated in both materials science and biology, and it seems clear now that diffractive imaging will soon be able to make a decisive contribution.

Spectacular 25 femtosecond single-shot images have just been obtained at DESY by CXDI using 30 nm X-rays from a free-electron laser with 90 nm spatial resolution (H. Chapman et al., *Nature*, 2006, in press). It is clear that diffractive imaging will play a major role in future time-resolved imaging efforts.

Acknowledgments. This chapter has summarized the work of many groups and many of my collaborators, as indicated in the references. I am particularly grateful for the help of Malcolm Howells, Uwe Weierstall, and Anton Barty during the preparation of this review. The work was supported by NSF, CBST, and IDBR award.

References

- Barakat, R. and Newsam, G. (1984). *J. Math. Phys.* **25**, 3190.
- Barty, A. (2005). Personal communication.
- Bates, R. and McDonnell, M. (1989). *Image Restoration and Reconstruction*. (Oxford University Press, New York).
- Bauschke, H., Combettes, P.I. and Luke, D.R. (2002). *J. Opt. Soc. Am.* **19**, 1334.

- Bauschke, H., Combettes, P.I. and Luke, D.R. (2003). *J. Opt. Soc. Am.* **A20**, 1025.
- Beetz, T., Howells, M., Jacobsen, C., Koa, C., Kirz, J., Lima, E., Menten, T., Miao, H., Sanchez-Hanke, C., Sayre, D. and Shapiro, D. (2005). *Nucl. Instr. Methods* **545**, 459.
- Bruck, Y. and Sodin, L. (1979). *Opt. Commun.* **30**, 304.
- Carrozzini, B., Cascarnao, G., Giovavazzo, C., Chapman, H., Marchesini, S., Howells, M., He, H., Wu, J., Weierstall, H. and Spence, J. (2004). *Acta. Crystallogr.* **A60**, 331.
- Cederquist, J.N., Fienup, J.R. et al. (1988). *Opt. Lett.* **13**, 619.
- Chapman, H., Barty, A., Beetz, C., Cui, He, H., Howells, M., Marchesini, S., Noy, A., Rosen, R., Spence, J.C.H. and Weierstall, U. (2006). *J. Opt. Soc. Am.* **23**, 1179.
- Collier, R., Burkhadt, C.B. Lin. (1971). *Optical Holography*. (Academic Press, New York).
- Eisebett, S., Lorgen, M. and Eberhart, W. (2004). *Appl. Phys. Lett.* **84**, 3373.
- Elser, V. (2003). *J. Opt. Soc. Am.* **20**, 40.
- Fienup, J.R. (1982). *Appl. Opt.* **21**, 2758.
- Fienup, R. (1987). *J. Opt. Soc. Am.* **A4**, 118.
- Fienup, J.R. (1997). *Appl. Opt.* **36**, 8352.
- Frank, J. Ed. (2006). *Electron Tomography*. (Plenum, New York).
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladiadj, M. and Leith, A. (1996). *J. Struct. Biol.* **116**, 190.
- Gerchberg, R.W. and Saxton, W.O. (1971). *Optik*. **34**, 275.
- Gerchberg, R. and Saxton, W. (1972). *Optik*. **35**, 237.
- Goodman, J.W. (1968). *Introduction to Fourier Optics*. (McGraw-Hill, New York).
- Hau-Riege, S., Szoke, H., Chapman, H., Marchesini, S., Noy, A., He, H., Howells, M., Weierstall, U. and Spence, J. (2004). *Acta. Crystallogr.* **1760**, 294.
- He, H., Howells, M., Marchesini, S., Chapman, H., Weierstall, U., Padmore, H. and Spence, J. (2003). *Acta. Crystallogr.* **A59**, 143.
- He, H., Marchesini, S., Howells, M., Weierstall, U., Chapman, H., Hau-Riege, S., Noy, A. and Spence, J.C.H. (2003). *Phys. Rev.*, p. 174114.
- Henderson, R. (1995). *Q. Rev. Biophys.* **28**, 171.
- Howells, M., Charalambous, P., He, H., Marchesini, S. and Spence, J.C.H. (2002). A zone-plate monochromator for soft X-rays. *Design and Microfabrication of Novel X-Ray Optics*. (D. Mancini, Ed.). Bellingham, WA.
- Howells, M., Beetz, T., Chapman, H., Cui, C., Holton, J., Jacobsen, C., Kirz, J., Lima, E., Marchesini, S., Miao, H., Sayre, D., Shapiro, D. and Spence, J.C.H. (2006). *J. Spectrosc. Rel. Phenom.* (in press).
- Howie, A. and Stern, R.M. (1972). *Z. Naturforsch.* **A27**, 382.
- Kirz, J., Jacobsen, C. and Howells, M. (1995). *Q. Rev. Biophys.* **28**.
- Levi, A. and Stark, H. (1984). *J. Opt. Soc. Am.* **A1**, 932.
- Liao, C., Fiddy, M. and Byrne, C. (1997). *J. Opt. Soc. Am.* **A14**, 3155.
- Luke, R. (2005). *Inverse Prob.* **21**, 37.
- Marchesini, S., Chapman, H.N., Hau-Riege, S.P., London, R.A., Szoke, A., He, H., Howells, M.R., Padmore, H., Rosen, R., Spence, J.C.H. and Weierstall, U. (2003a). *Opt. Express* **11**, 2344.
- Marchesini, S., He, H., Chapman, H., Hau-Riege, S., Noy, M., Howells, M., Weierstall, U. and Spence, J.C.H. (2003b). *Phys. Rev.* **B68**, 140101(R).
- Marchesini, S. (2006). *Rev. Sci. Instr.* (arxiv:physics/0603201).
- McMahon, P., Allman, B., Jacobsen, D., Arif, M., Werner, S. and Nugent, K. (2003). *Phys. Rev. Lett.* **91**, 145502.
- Miao, J., Sayre, D. and Chapman, H. (1998). *J. Opt. Soc. Am.* **A15**, 1662.
- Miao, J., Charalambous, C., Kirz, J. and Sayre, D. (1999). *Nature* **400**, 342.

- Miao, J., Ishikawa, T., Johnson, E., Lai, B. and Hodgson, K. (2002). *Phys. Rev. Lett.* **89**, 088303.
- Midgely, P.A., Weyland, M., Thomas, J.M. and Johnson, F.G. (2001). *Chem. Commun.* **2001**, 907.
- Millane, R. (1990). *J. Opt. Soc. Am.* **7**, 394.
- Millane, R.P. (1995). *Opt. Soc. Am.* **13**, 725.
- O'Keefe, M.A. and Spence, J.C.H. (1993). *Acta. Crystallogr.* **A50**, 33.
- Oszlanyi, G. and Suto, A. (2004). *Acta. Crystallogr.* **A60**, 134.
- Paganin, D. and Nugent, K. (1998). *Phys. Rev. Lett.* **80**, 2586.
- Radi, G. (1970). *Acta. Crystallogr.* **A26**, 41.
- Reimer, L. (1989). *Transmission Electron Microscopy*. (Springer-Verlag, New York).
- Robinson, I.K., Vartanyants, I.A., Williams, G., Pfeifer, M. and Pitrey, J. (2001). *Phys. Rev. Lett.* **87**, 195505.
- Salvo, L., Cloetens, P., Maire, E., Zabler, S., Blandin, J., Buffiere, J., Ludwig, W., Boller, E., Bellet, D. and Josserond, C. (2003). *Nucl. Instr. Methods* **B200**, 273.
- Sayre, D. (1952). *Acta. Crystallogr.* **5**, 843.
- Sayre, D. (1980). *Image Processing and Coherence in Physics*. Springer Lecture Notes in Physics (M. Schlenker, Ed.), Vol. 112, 229. (Springer, New York).
- Sayre, D., Chapman, H. and Miao, J. (1998). *Acta. Crystallogr.* **A54**, 232.
- Shapiro, D., Thibault, P., Beetz, T., Elser, V., Howells, M., Jacobsen, C., Kirz, J., Lima, E., Miao, H., Neiman, A.M. and Sayre, D. (2005). *Proc. Natl. Acad. Sci. USA* **102**, 15343.
- Spence, J.C.H. (2003). *High Resolution Electron Microscopy*. (Oxford University Press, New York).
- Spence, J.C.H. and Zuo, J.M. (1992). *Electron Microdiffraction*. (Plenum, New York).
- Spence, J.C.H., Howells, M., Marks, L.D. and Maio, J. (2001). *Ultramicroscopy* **90**, 1.
- Spence, J., Weierstall, U., Fricke, J. Glaeser, R. and Downing, K. (2003a). *J. Struct. Biol.* **144**, 209.
- Spence, J.C.H., Wu, J., Giovacazzo, C., Carrozzini, B., Cascarano, G. and Padmore, H. (2003b). *Acta. Crystallogr.* **A59**, 255.
- Spence, J., Weierstall, U. and Howells, M. (2004). *Ultramicroscopy*. **101**, 149.
- Stark, H. (1987). *Image Recovery: Theory and Applications*. (Academic Press, New York).
- Strutt, J.W. (1892). *Phil. Mag.* **34**, 407.
- Szoke, A. (1997). *J. Imaging Sci. Technol.* **41**, 332.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90.
- Weierstall, U., Chen, Q., Spence, J.C.H., Howells, M., Isaacson, M. and Panepucci, R. (2001). *Ultramicroscopy* **90**, 171.
- Williams, G.J., Pfeifer, M.A. and Vartanyants, I.A. (2003). *Phys. Rev. Lett.* **90**, 175501.
- Wu, J., Spence, J., O'Keefe, M. and Groy, T. (2004a). *Acta. Crystallogr.* **A60**, 326.
- Wu, J., Weierstall, U., Spence, J.C.H. and Koch, C. (2004b). *Opt. Lett.* **29**, 1.
- Youla, D. and Webb, H. (1982). *IEEE-Trans. Med. Imaging* **MI-1**, 81.
- Zhang, K.Y.J. and Main, P. (1990). *Acta. Crystallogr.* **A46**, 41.
- Zuo, J.M., Vartanyants, I.A., Gao, M., Zhang, M. and Nagahara, L.A. (2003). *Science* **300**, 1419.

(See page 1138)

The Notion of Resolution

S. Van Aert, Arnold J. den Dekker, D. Van Dyck, and A. Van den Bos

1 Introduction

In microscopy, resolution has always been, and still is, an important issue. Since it is not an unambiguously defined physical quantity, it is interpreted in many ways (den Dekker and van den Bos, 1997). The purpose of this chapter is, on the one hand, to briefly review past and existing resolution definitions and methods, and, on the other hand, to present alternative quantitative definitions of resolution based on model fitting. Throughout this chapter, emphasis will be placed on electron microscopy.

Using model fitting, the resolution will principally be discussed in terms of the precision with which unknown quantities, atom positions in particular, can be measured. It will be shown that a precision of the order of 0.01 Å is in principle possible even with an electron microscope that is not corrected for spherical and chromatic aberration. Once atom positions can be measured with a precision of 0.01 Å they can be used as input data for theoretical *ab initio* calculations (Muller, 1998, 1999; Spence, 1999; Kisielowski et al., 2001b). Such calculations make it possible to calculate the properties of a material with a given structure. In this way, the combination of precise experimental structure determination and theoretical calculations contributes to the understanding of the properties–structure relation. A complete understanding of this relation, combined with recent progress in building materials atom by atom, will enable materials science to evolve toward materials design, that is, from describing and understanding toward predicting materials with interesting properties (Wada, 1996; Olson, 1997, 2000; Reed and Tour, 2000; Browning et al., 2001).

The detailed outline of the sections in this chapter is as follows. In Section 2, classical two-point resolution criteria are discussed. The most widely known classical resolution criterion is that of Lord Rayleigh. The Rayleigh criterion is derived from the assumption that the human visual system needs a minimal contrast to discriminate two points in its composite intensity distribution. Other classical criteria can be seen as modified versions of Rayleigh’s criterion. The classical criteria are expressed in terms of the width of the point spread function

of the imaging instrument. Section 3 discusses resolution in the spatial frequency domain, addressing literature relating linear systems theory to resolution. In this section, the diffraction limit to resolution and its relation to the classical resolution criteria are introduced. Furthermore, the notion of superresolution is discussed. It will be shown that to attain superresolution, prior knowledge is required. In Section 4, deterministic model-based resolution is considered. Here, prior knowledge is taken into account in the form of a parametric model. However, the images are supposed to be noise free. The unknown parameters, such as the positions of projected atoms, are measured by means of model fitting. Then, in Section 5, statistical model-based resolution is studied. It is taken into account that the observations fluctuate about their expectations due to the unavoidable presence of electron counting noise in the images. The parameters are measured by means of model fitting using parameter estimation methods. Hence, statistical model-based resolution is discussed in terms of the precision with which the parameters can be estimated. Finally, in Section 6, ultimate model-based resolution is contemplated. It is shown that depending on the observations, the solutions of the position estimates may be exactly coinciding. Section 7 includes a discussion and conclusions.

2 Classical Two-Point Resolution

Two-point resolution is a widely used criterion for the resolving capabilities of an imaging system. It is defined as the system's ability to resolve two point sources of equal brightness. Due to the finite size of the system's optical components, a point source is not imaged as a point but as the diffraction pattern of the system's effective aperture. This diffraction pattern therefore represents the system's point spread function. In astronomical problems, two-point resolution is more than just a resolution measure. It has direct practical significance, since in astronomical problems many objects are effectively point sources.

2.1 Rayleigh Resolution

The most widely used criterion for two-point resolution is that of Lord Rayleigh (Strutt, 1899). Rayleigh estimated the minimal resolvable distance between two points of equal brightness that are imaged by a diffraction-limited imaging system. According to the Rayleigh criterion, two point sources are just resolved if the central maximum of the diffraction pattern generated by one point source coincides with the first zero of the diffraction pattern generated by the second. This means that Rayleigh's resolution limit is given by the distance between the central maximum and the first zero of the point spread function of the imaging system. The criterion can be generalized to include point spread functions that have no zero in the neighborhood of their central maximum, by taking the resolution limit as the distance for which the intensity at the central dip in the composite image is 81% of that at the maxima on either side. This corresponds to the original Rayleigh limit for a rectangular aperture.

For example, let us consider the Rayleigh limit, or alternatively, the *point resolution*, of an imaging system with a (two-dimensional) Gaussian point spread function of the following form:

$$p(\mathbf{r}) = p(r) = \frac{1}{2\pi\rho^2} \exp\left(-\frac{x^2 + y^2}{2\rho^2}\right) = \frac{1}{2\pi\rho^2} \exp\left(-\frac{r^2}{2\rho^2}\right) \quad (1)$$

where r is the absolute value of the two-dimensional vector $\mathbf{r} = (x \ y)^T$ and ρ is the width of the Gaussian function. The superscript T denotes transposition. According to Rayleigh, the point resolution ρ_p , which is the smallest distance at which two points can be resolved, is given by the requirement that the value of the cross section of the composite intensity distribution halfway between these two points is about 0.8 times the value at the maxima. Thus

$$2 \exp\left(\frac{-\rho_p^2}{8\rho^2}\right) = 0.8 \quad (2)$$

from which it follows that

$$\rho_p \approx 2\sqrt{2}\rho \quad (3)$$

2.2 Sparrow Resolution

Rayleigh's choice of resolution limit is based on presumed resolving abilities of the human visual system. Since Rayleigh's days, several other resolution criteria have been proposed that are similar to Rayleigh's (for a review, refer to den Dekker and van den Bos, 1997). A notable example of such so-called classical criteria for two-point resolution is that of Sparrow (1916). Compared to Rayleigh resolution, which is based on presumed capabilities of the human visual system, the Sparrow resolution is based on a less subjective criterion that is valid for a hypothetical perfect imaging instrument. It states that the smallest resolvable distance between two points is the point at which the minimum in the composite image intensity distribution just disappears. When this definition of resolution is applied to an imaging system with a Gaussian point spread function, it follows from Eqs. (1) and (3) that,

$$\rho_s \approx \frac{\sqrt{2}\rho_p}{2} \quad (4)$$

From the examples given in this section, it can be noted that classical two-point resolution criteria are expressed in terms of the width of the point spread function of the imaging instrument. A narrower point spread function corresponds to an improved Rayleigh or Sparrow resolution.

3 Resolution in the Spatial Frequency Domain: Diffraction Limit and Superresolution

3.1 The Diffraction Limit

In the previous section, classical resolution criteria have been discussed. An alternative resolution criterion is based on linear systems

theory. It is assumed that the imaging system is linear and shift invariant. Coherent imaging systems are linear in complex amplitude and incoherent imaging systems are linear in intensity. The characteristics of a shift-invariant linear imaging system are defined by its point spread function, or, equivalently, by its transfer function, which is the Fourier transform of the point spread function. Point spread functions are more directly useful in the assessment of telescopes or spectroscopic instruments. Transfer functions, on the other hand, are more often used in the case of microscopes and cameras. The amplitude (coherent imaging) or intensity (incoherent imaging) distribution in the image produced by a linear and shift-invariant system is the convolution of the amplitude (or intensity) distribution of the object and the amplitude (or intensity) point spread function of the imaging system. For the spatial frequency domain, the imaging system acts as a filter for spatial frequencies. Each spatial frequency is transferred from the object to the image plane independently of all other frequencies present; the corresponding amplitude (coherent imaging) or intensity (incoherent imaging) is multiplied by the transfer function, or frequency response function, of the system. Due to the finite size of the system's aperture, transfer functions of (both coherent and incoherent) imaging systems are band limited, i.e., they are equal to zero for all frequencies above a certain cutoff frequency. Born and Wolf (1999) discuss the more general class of partially coherent imaging systems and show that these systems are also strictly band limited. Therefore, independent of the degree of coherence, spectral components beyond the cutoff frequency are not transferred by the imaging system. For this reason, the cutoff frequency is called the *diffraction limit* to resolution.

3.2 The Diffraction Limit and Its Relation to Rayleigh and Sparrow Resolution

The diffraction limit, or equivalently, the cutoff frequency, is related to the Rayleigh resolution. This can be seen as follows. It is well known that a narrow function has a broad Fourier transform and vice versa. Therefore a point spread function with a narrow main lobe corresponds to a transfer function with a high cutoff frequency. In fact, the product of the Rayleigh limit and the cutoff frequency is a constant that is close or even equal to one. For example, this constant is equal to 1 and 1.22 for incoherent imaging systems with a rectangular and a circular aperture, respectively.

As mentioned above, transfer functions of practical imaging systems are strictly band limited (due to the finite size of the system's aperture). In this work, it will often be assumed that the point spread function of the imaging systems under study can be described by a Gaussian function, such as (1). This assumption is made so as to keep calculations simple. The transfer function of the imaging system of which the point spread function is described by Eq. (1) is given by the two-dimensional Fourier transform of Eq. (1):

$$P(g) = \exp(-2\pi^2 \rho^2 g^2) \quad (5)$$

where g is the absolute value of the two-dimensional spatial frequency vector $\mathbf{g}$. Although this transfer function tends to zero for increasing frequency values, it is not *strictly* band limited. Nevertheless, the Gaussian approximation is sufficiently accurate for the purpose of this chapter. From the condition that the product of the Rayleigh resolution limit and the cutoff frequency should approximately be equal to one, it follows from Eq. (3) that the cutoff frequency corresponding to Rayleigh resolution is

$$g_p = \frac{1}{\rho_p} \approx \frac{1}{2\sqrt{2}\rho} \quad (6)$$

At that spatial frequency the modulus of the transfer function, which is given by Eq. (5), is reduced to only 8%. Thus, for Gaussian point spread functions as described by Eq. (1), the Rayleigh resolution limit can also be defined as the inverse of the spatial frequency for which the transfer function is reduced to 8% of its peak value.

The diffraction limit and its relation to Rayleigh and Sparrow resolution will now be discussed for conventional transmission electron microscopy (TEM). Thus far, only the diffraction limited point spread function of the imaging instrument has been taken into account. However, for electron microscopy, this should be extended to include the point spread function describing the effect of thermal vibrations of the atom, the effect of the environment, and the detector (de Jong and Dyck, 1993). Moreover, it has to be noted that the atoms are not point scatterers. Hence, an extension from points to objects of finite size has to be made. As shown in Figure 20-1, each effect contributing to the imaging process can be represented by a transfer function, which acts as a low pass filter. The transfer function of the electron microscope consists of a damping function, which is mainly due to chromatic aberration, and a phase shift, which causes the oscillations. Since there are many ways to get rid of the oscillations, such as focal series reconstruction (Schiske, 1973; Saxton, 1978; Van Dyck and Coene, 1987; Van Dyck et al., 1993; Coene et al., 1996; Thust et al., 1996) and correction of the spherical aberration (Rose, 1990), the Rayleigh resolution of the electron microscope can be assumed to be given by the so-called *information limit*, which is proportional to the inverse of the highest spatial frequency that is still transferred with appreciable intensity. For simplicity, it will first be assumed that the imaging process is linear. This requires that the interaction between the electron and the object also is linear, which means that there is a simple linear relation of the *electron exit wave* and the projected electrostatic potential. The electron exit wave is a complex wave function in the plane at the exit face of the object, resulting from the interaction of the electron beam with the object. For example, the imaging process of weak phase objects, for which the so-called weak phase object approximation holds (Buseck et al., 1988), may be considered to be linear. If the object is a crystal, viewed along a zone axis, the electrostatic potential of all the atoms along the atom column is superimposed, which makes the interaction very strong and highly nonlinear. In that particular case, due to the focusing effect of the successive atoms, the scattering is increased to

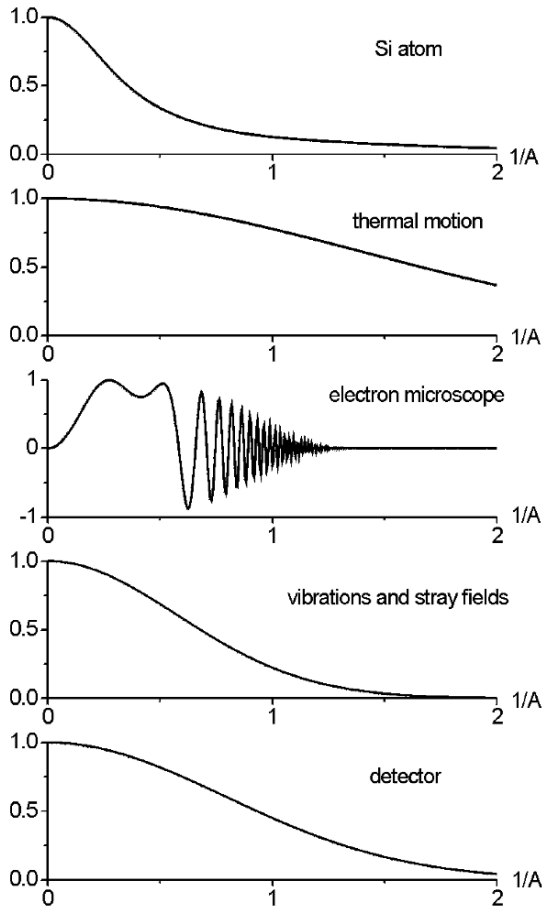


Figure 20-1. Transfer functions of the different subchannels of electron microscopic imaging.

higher angles. This effect is explained by the channeling theory (Howie, 1966; Van Dyck et al., 1989; Pennycook and Jesson, 1991; Van Dyck and Chen, 1999). However, for amorphous objects, the atoms are stacked in a disordered fashion, so that in projection their cores do not overlap, except by coincidence. As a result, the interaction remains linear for much larger object thicknesses and may be described by the weak phase object approximation. If the imaging is linear, all transfer functions have to be multiplied, or, equivalently, the point spread functions have to be convolved. If it is assumed that all constituent point spread functions are Gaussian, such as in Eq. (1), the resulting function is a Gaussian as well, with a Rayleigh resolution ρ_p determined by

$$\rho_p^2 = \rho_A^2 + \rho_T^2 + \rho_{EM}^2 + \rho_v^2 + \rho_D^2 \quad (7)$$

with ρ_A the “width” of the electrostatic potential of the atom, ρ_T the Rayleigh resolution limited by thermal vibrations of the atom, ρ_{EM} the Rayleigh resolution of the electron microscope, ρ_v the Rayleigh resolution limited by the environment (vibrations and stray fields), and ρ_D

the Rayleigh resolution limited by the detector. Today, for the best electron microscopes, ρ_{EM} is somewhat below 1 Å (O’Keefe et al., 2001; Kisielowski et al., 2001a; Batson et al., 2002a). In future instrumental developments the Rayleigh and Sparrow resolution can be improved by improving the resolutions of all different subchannels. However, a factor that cannot be improved is ρ_{A} , that is, the intrinsic “width” of the atom itself. It is important to note that beyond a certain point, it will be useless to further improve the Rayleigh resolution of the electron microscope since the transfer at high spatial frequencies is limited by the scattering factor. It is already difficult to find suitable objects that can be used to demonstrate the true Rayleigh resolution ρ_{EM} of an electron microscope. Consider, for example, amorphous silicon. From Figure 20–1, it follows that ρ_{A} is about 1 Å. Therefore, for the best electron microscopes, it follows from Eq. (7) that for amorphous silicon

$$\rho_{\text{p}} \approx 1 \text{ Å} \quad (8)$$

and from Eq. (4) that

$$\rho_{\text{s}} \approx 0.7 \text{ Å} \quad (9)$$

From this example, it can be concluded that for the best electron microscopes, the atoms themselves limit classical resolution criteria and hence the diffraction limit. However, it should be noted that the discussion of Rayleigh resolution and the diffraction limit is far more complicated in case of nonlinear electron–object interaction, for example, in the case of atom columns viewed along the column direction. It will then also depend on the assumptions regarding the scattering of the electrons on their way through the object. Furthermore, for coherent imaging, such as in TEM, Goodman (1968) has shown that the Rayleigh resolution will depend on the “phase distribution” associated with the object. Depending on the relative phase associated with two atoms or atom columns, the central dip in the composite image will be absent or present. For particular values of the relative phase shift, the dip will even be greater than the dip corresponding to an incoherent image of these two atoms or atom columns. From this example, it can be concluded that there is no simple generalization as to which type of imaging, coherent or incoherent, is preferred in the sense of Rayleigh resolution. So the assumption that incoherent imaging, for example scanning transmission electron microscopy (STEM), will yield a “better” resolution than coherent imaging, for example, TEM, is in general not valid.

In the remainder of this chapter, it will be shown that by using super-resolution algorithms, frequency components lying beyond the diffraction limit of the imaging system may be reconstructed. Then, other definitions of resolution are of interest.

3.3 Superresolution

Superresolution refers to reconstructing frequency components that lie beyond the cutoff frequency of the imaging system. At first sight, superresolution seems impossible. Knowledge of the system’s transfer

function makes it possible to reconstruct the object spectrum within the passband of the imaging system by means of inverse filtering of the image spectrum, but frequency components beyond the diffraction limit seem irrevocably lost. Indeed, in the absence of any restriction as to the nature of the object, there are an infinite number of objects that can produce the same image. Under certain conditions, however, super-resolution is possible. The key to superresolution is prior knowledge. For example, suppose that it is known that the object is of finite size, that is, it is nonzero only in a region of finite extent. This single condition guarantees that the object spectrum is analytic. A well-known property of an analytic function is that if it is known over a specified interval, it can always be reconstructed in its entirety (Castleman, 1979). This process of reconstruction is called analytic continuation. It can be shown that this method is perfect in theory. If the images are noise free, it leads to an exact and complete reconstruction of the object spectrum (Harris, 1964). However, noise limits its practical use (Frieden, 1967). Nevertheless, many effective superresolution algorithms (digital image processing methods) have been proposed in the literature [for a review, see Frieden (1975), Hunt (1994), and Meinel (1986)]. Both empirically and theoretically, it has been shown that there are certain necessary conditions to be satisfied by a superresolution algorithm to be successful (Hunt, 1994). First, the algorithm should explicitly utilize a mathematical description of the image formation process that relates object and image via the point spread function of the imaging system. Second, the images should be sufficiently oversampled to avoid aliasing after reconstruction of spatial frequencies beyond the diffraction limit. For Nyquist sampled images (Gonzalez and Woods, 2002), the algorithm should contain some suitable form of interpolation. Last, but not least, the algorithm should contain prior knowledge of the object. Examples of such prior knowledge used by superresolution algorithms include the following:

- Finite extent of the object (as discussed above) (e.g., Harris, 1964; Gerchberg, 1974, 1989).
- Positivity of the object (e.g., Schell, 1965; Biraud, 1969; Walsh and Nielsen-Delaney, 1994).
- Upper and lower bounds on the object intensity (e.g., Janson et al., 1970).
- Object statistics (e.g., Frieden, 1980; Hunt and Sementilli, 1992).
- Parametric model of the object.

Obviously, the performance of any superresolution algorithm will be limited by noise. In the remainder of this chapter, we will assume that the available prior knowledge of the object to be reconstructed consists of a parametric model. Then, superresolution can be achieved by computing the relatively small number of unknown parameters characterizing the object from the available observations. In electron microscopy these observations may be electron counting results made at the pixels of a CCD camera. The image reconstruction problem thus becomes a parameter estimation problem. For example, in the case of two-point resolution, the object can be described by a two-component model

parametric in the locations of the point sources. Hence, if the point spread function is known, a parametric model of the composite image of the two-point sources can be derived. By fitting this model to the image optimizing some criterion of goodness of fit, we obtain estimates of the parameters of the model. For a correct model, in the absence of noise, and apart from potential computational problems that will be discussed in Section 4, this would result in a perfect fit. That is, the object parameters can be estimated with unlimited precision so that the object can be reconstructed perfectly. This means that there would be no limit to resolution no matter how closely the point sources are spaced. However, in practice it is noise that limits the accuracy and precision with which parameters can be measured and therefore limits the resolution. This will be the subject of Sections 5 and 6.

4 Deterministic Model-Based Resolution

Classical resolution criteria disregard the possibility of using prior knowledge to extract analytic results from observations by means of model fitting (den Dekker and van den Bos, 1997). In this section, prior knowledge is taken into account in the form of a model describing the observations. Thus far, the observations are assumed to be noise free. Compared to Section 2, the model will be extended from two-peak models to one or more-peak models. Instead of classical resolution, we will speak of *deterministic* model-based resolution. It will be shown that the relevant limits to deterministic model-based resolution are in any case computational. However, if the model is inaccurate, which means that it systematically deviates from the exact noise-free observations, it will be shown that the relevant limits are both computational and fundamental.

Imagine that Lord Rayleigh would image stars today. First, the image of one star, which can be treated as a point object, will be considered. Now, there exists a model for the object, namely that it consists of a point. The point spread function of the telescope is also exactly known. So, it is known how an image of one star should look like. Thus, there is no interest in the detailed form of this image, but only in the position of the star. The only objective of the experiment is to determine this position. Obviously, in the absence of noise, numerically fitting the known one-peak model to the image with respect to the position parameter would result in a perfect fit. The resulting solution for this location would be exact, and despite the blurring effect of the point spread function, it imposes no limit to location resolution.

This line of reasoning can be extended to position measurements of atoms or atom columns from noise-free electron microscopic observations. Suppose that these observations λ_{kl} are made at the pixels (k, l) at the position $(x_k, y_l)^T$. The model that describes these observations is called $f_{kl}(\tau)$ with τ the vector of unknown parameters, among which are the locations of the atoms or atom columns. An example of such a model is the following:

$$\lambda_{kl} = f_{kl}(\tau) = \zeta + \sum_{n=1}^{n_c} \frac{\eta_n}{2\pi\rho^2} \exp \left[\frac{-(x_k - \beta_{xn})^2 - (y_l - \beta_{yn})^2}{2\rho^2} \right] \quad (10)$$

where ζ is the constant background, η_n is the column-dependent height of the Gaussian peak, ρ is the width of the Gaussian peak, n_c is the total number of atom columns, and β_{xn} and β_{yn} are the x - and y -coordinate of the n th atom or atom column, respectively. The width ρ is supposed to be identical for different atom columns. The parameter vector τ is equal to $(\beta_{x1} \dots \beta_{xnc} \beta_{y1} \dots \beta_{ync} \eta_1 \dots \eta_{nc} \rho \zeta)^T$ and contains $R = 3n_c + 2$ elements. The unknown parameters can be measured by fitting the model to the observations. In a sense, we are then looking for the optimum value of a criterion in a parameter space whose dimension is equal to R , that is, the number of parameters to be measured. Each possible combination of the R parameters can be represented by a point in an R -dimensional space. The search for the global optimum of the criterion of goodness of fit in this space is an iterative numerical optimization procedure.

The problem may be of a computational kind. The existing optimization methods fail if the dimension of the parameter space is so high that it is not possible to avoid ending up at a local optimum instead of at the global optimum of the criterion of goodness of fit, so that the wrong structure is derived. To solve this dimensionality problem, that is, to find a pathway to the global optimum, a good starting structure is required, that is, initial conditions should be available for the parameters. For example, neighboring atoms or atom columns should be discriminated in an image. In other words, the structure has to be resolved. This corresponds to X-ray crystallography, where it is first necessary to resolve a structure by using, for example, direct methods, and afterward to refine the structure. Moreover, the computing time needed to reach convergence of the iterative procedure increases with the dimension of the parameter space. In the following example, the problems related to the study of the amorphous object with atomic resolution TEM will be discussed. More details can be found in Van Dyck et al. (2003).

Example 1 (Amorphous Object) For an amorphous object, the number of parameters increases with thickness. Therefore, from a certain thickness on, it will be difficult to resolve the structure in projection. For example, consider Figures 20–2 and 20–3. In Figure 20–2, the amorphous foil is thin, whereas in Figure 20–3, it is thicker. Therefore, the number of projected atoms is larger in Figure 20–3 than in Figure 20–2. It is clear that it will be more likely to resolve the structure for the example given in Figure 20–2 than for that corresponding to Figure 20–3. To resolve the structure, it will be assumed that the distances between neighboring projected atom positions should be larger than or equal to the Sparrow resolution ρ_s . The reason for choosing this criterion is that the computer will then be able to distinguish the individual atoms, since the observations are assumed to be noise free. However, it should be noted that this criterion is not exact and, therefore, it will give only rough guidelines. Suppose that the mean concentration of atoms per cubic ångström is equal to

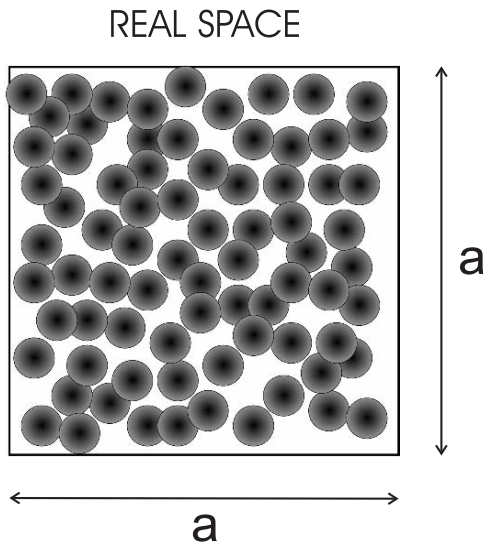


Figure 20-2. Amorphous structure containing clearly separable projected atoms.

V. Then, the mean concentration A of projected atoms per square ångstrom is given by

$$A = Vz \tag{11}$$

where z is the thickness of the amorphous foil. On the other hand, if it is assumed that each projected atom occupies a circle with a diameter equal to the average distance d , averaged over distances between nearest-neighbor projected atoms, then

$$A \approx \frac{1}{\pi(d/2)^2} \tag{12}$$

From Eqs. (11) and (12), it follows that the thickness of the amorphous foil is approximately given by

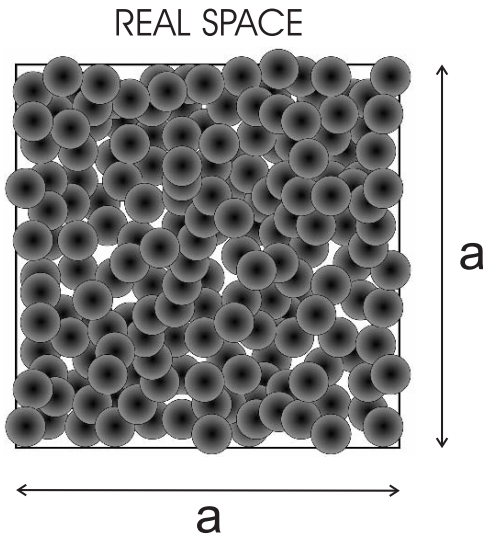


Figure 20-3. Amorphous structure containing severely overlapping projected atoms.

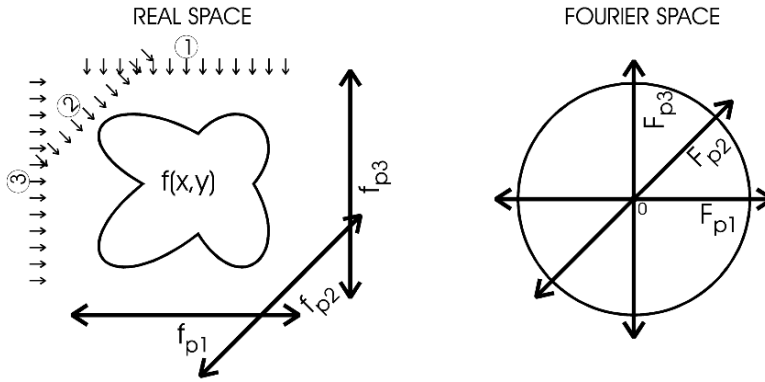


Figure 20-4. The principle of tomography.

$$z \approx \frac{4}{\pi d^2 V} \quad (13)$$

To resolve the structure and therefore to avoid dimensionality problems, it will be assumed that the following condition is met:

$$d \geq 2\rho_s \quad (14)$$

The factor 2 is arbitrarily chosen. Requiring that d is larger than or equal to ρ_s would not be sufficient. In that case, a substantial part of distances between neighboring atoms would be smaller than ρ_s and hence it is not possible to resolve the structure. In principle, this can still occur if inequality (14) is fulfilled, but the probability that it occurs is lower. Then, it follows from Eqs. (13) and (14), that

$$z \leq \frac{1}{\pi \rho_s^2 V} \quad (15)$$

For amorphous silicon, it follows from Eq. (9) that ρ_s is approximately equal to $0.7 \text{ \AA}$ and, furthermore, V is approximately equal to $0.05 \text{ atoms/\AA}^3$. Hence, it follows from Eq. (15) that the amorphous silicon foil should not exceed thicknesses of the order of $13 \text{ \AA}$ so as to avoid dimensionality problems. This thickness is rather small, which means that it is unrealistic to expect that atomic resolution TEM is able to resolve amorphous silicon samples with realistic foil thicknesses from only one projection (Cowley, 2001). It can thus be stated that the structure of a realistic amorphous object cannot be determined from one image alone. However, the situation can be improved drastically by using a tomographic technique in which the sample is tilted and many projections from different viewing directions are combined as shown in Figure 20-4. The Fourier transform of a projection yields a section through the origin of the three-dimensional Fourier space. By combining many different projections, it is possible to reconstruct the whole Fourier space (Frank, 1992). In this way an ideal microscope can resolve about one atom per cubic ångstrom, which is sufficient to resolve amorphous structures.

In the foregoing, it has been assumed that the model is accurate. However, if the model is inaccurate, the estimated position parameters

may deviate from the true positions, even if the observations would be noise free. This is called systematic error. Obviously, systematic errors would place fundamental limits on deterministic model-based resolution.

In this section, it has been shown that deterministic model-based resolution is computationally limited. Computational problems can be overcome only if the structure can be resolved. Moreover, if the model is inaccurate, the resolution is also fundamentally limited, since this results in a systematic error.

5 Statistical Model-Based Resolution

In Section 4, it was assumed that the observations are noise free. However, in any real-life experiment, the observations will “contain errors.” Then, the resolution depends fundamentally on the signal-to-noise ratio in the detected image. In this section, the resolution will be considered in the framework of statistical parameter estimation theory and will be called *statistical* model-based resolution. It will be shown that the relevant limits are both computational and fundamental.

Suppose there is a CCD camera that is able to count the individual photons forming the image of a single point object or of two point objects. The images as measured by this camera appear as in Figure 20–5 or as in Figure 20–6 for single or two-point objects, respectively. The noise on these images stems from the counting statistics. The position parameters can be estimated by numerically fitting the known parameterized mathematical model to the images with respect to the component positions, in the same way as expressed in Section 4. However, if one repeated this experiment several times, one would, due to the statistical nature of the observations, find different values

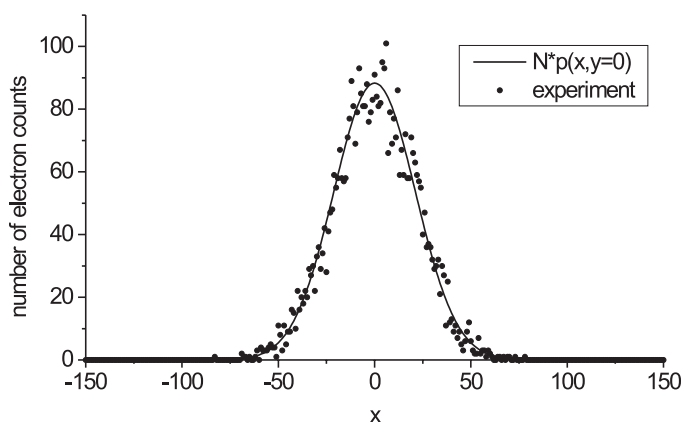


Figure 20–5. Simulation experiment of the image of a point as measured by a CCD camera (300×300 pixels), with pixel size $\Delta x = \Delta y = 1$. The point spread function used is a two-dimensional normalized Gaussian function $p(x, y)$ as in Eq. (1) with $\rho = 21$. The experimental as well as the expectation values $N \times p(x, y)$ are shown within the section $y = 0$. The number of imaging particles N is equal to 250,000.

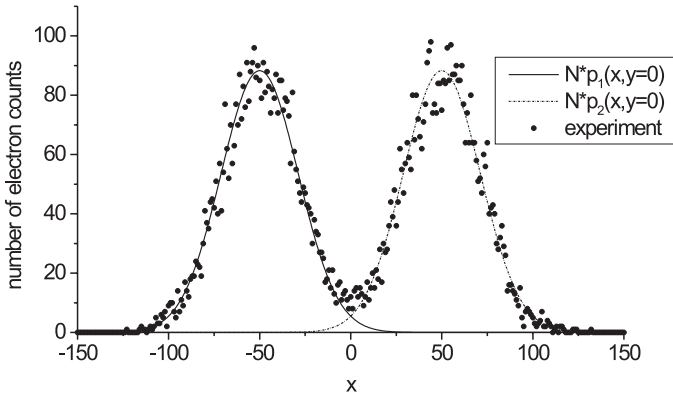


Figure 20–6. The results of a computer-simulated image of two neighboring points. The experimental as well as the expectation values of the individual peaks $[N \times p_{1,2}(x, y)]$ are plotted within the section $y = 0$, where $p_{1,2}(x, y)$ are two-dimensional normalized Gaussian point spread functions, as in Eq. (1). The number of imaging particles $2N$ is equal to 500,000. The peaks are clearly separable.

for the position or the distance estimates for single or two-point objects, respectively. The position and distance estimates are statistically distributed about their mean values. Then, the obvious criterion to quantify statistical model-based resolution is the precision of the estimate, which is given by the variance of this distribution, or, by its square root, the standard deviation. In a sense, the standard deviation is the “error bar” on the position or on the distance. Applying statistical parameter estimation theory, the *attainable* precision can be adequately quantified in the form of the so-called *Cramér–Rao lower bound* (CRLB) (van den Bos and den Dekker, 2001). This is a lower bound on the variance of any unbiased estimator of a parameter. It means that the variance of different estimators, such as, the least squares or the maximum likelihood (ML) estimator, can never be lower than the theoretical CRLB on the variance. Fortunately, the ML estimator attains the CRLB asymptotically, that is, if the number of observations is sufficiently large. Note that in accordance with the available literature, the CRLB on the variance of the position estimate of the image of a single point object and the CRLB on the variance of the distance estimate of the image of two point objects are measures of what is called *single-source* (Falconi, 1964) and *differential (two-source)* resolution (Falconi, 1967), respectively. Single-source resolution is defined as the instrument’s capacity to determine the position of a point object that is observed in a background of noise. Differential resolution is defined as the instrument’s ability to determine the separation of two point objects.

In this section the fluctuating behavior of the observations will be described in Section 5.1 using parametric statistical models of observations. Next, in Section 5.2, it will be shown how an adequate expression for the attainable statistical precision of the parameter estimates, that is, the CRLB, can be derived from such a parametric statistical model. Then, in Section 5.3, the ML estimator of the parameters will be derived

from the parametric model. This estimator is important since it achieves the CRLB asymptotically. In Section 5.4, an important purpose for which the expressions for the CRLB can be used will be introduced, namely, *statistical experimental design*.

5.1 Parametric Statistical Models of Observations

Generally, due to the inevitable presence of noise, sets of observations made under the same conditions nevertheless differ from experiment to experiment. The usual way to describe this behavior is to model the observations as *stochastic variables*. The reason is that there is no viable alternative and that it has been found to work (van den Bos, 1999; van den Bos and den Dekker, 2001). By definition, a stochastic variable is characterized by its probability density function, while a set of stochastic variables has a joint probability density function.

Consider a set of stochastic observations w_m , $m = 1, \dots, M$ made at the measurement points $x_1, \dots, x_M$. These measurement points are assumed to be exactly known. In electron microscopy, the observations are, for example, electron counting results made at the pixels of a CCD camera, where M represents the total number of pixels. The reader should not be misled by the fact that the observations and measurement points are here represented in a one-dimensional way. It is intended as a general representation. Also if the observations and measurement points would be two- or higher-dimensional, they can easily be transformed to a one-dimensional representation. For example, if the observations and measurement points are two-dimensional, say, w_{kl} , $k = 1, \dots, K$, $l = 1, \dots, L$ and x_{kl} , $k = 1, \dots, K$, $l = 1, \dots, L$, respectively, they can also be represented as w_m , $m = 1, \dots, M$ and x_m , $m = 1, \dots, M$, respectively, with $M = K \times L$. The $M \times 1$ vector w defined as

$$w = (w_1 \dots w_M)^T \quad (16)$$

is the column vector of these observations. It represents a point in the Euclidean M space having $w_1, \dots, w_M$ as coordinates. This will be called *space of observations* (van den Bos and den Dekker, 2001). The expectations of the observations, that is, the mean values of the observations, are defined by their probability density function. The vector of expectations

$$E[w] = (E[w_1] \dots E[w_M])^T \quad (17)$$

is also a point in the space of observations and the observations are distributed about this point. The symbol $E[\cdot]$ denotes the expectation operator. The expectations of the observations are described by the *expectation model*, that is, a physical model, that contains the unknown parameters to be estimated, such as the position coordinates of the projected atoms or atom columns. In a sense, this model has first been introduced in Section 4 on the understanding that it now describes the expectations of the observations whereas in Section 4, the model describes noise-free observations. The unknown parameters are represented by the $R \times 1$ parameter vector $\tau = (\tau_1 \dots \tau_R)^T$. Thus, it is supposed that the expectation of the m th observation is described by

$$E[w_m] = f_m(\tau) = f(x_m; \tau) \quad (18)$$

where $f_m(\tau)$ represents the expectation model, which is evaluated at the measurement point x_m and which depends on the parameter vector τ . Apart from the unknown parameters τ , the expectation model may contain known parameters and experimental settings as well. An example of an expectation model is given by Eq. (10).

Electron microscopic observations are electron counting results detected, for example, with a CCD camera. Under the assumption that the quantum efficiency of this detector is sufficiently large to detect single electrons, these observations may be assumed to be Poisson distributed. This means that the probability that the observation w_m is equal to ω_m is given by (Papoulis, 1965)

$$\frac{\lambda_m^{\omega_m}}{\omega_m!} \exp(-\lambda_m) \quad (19)$$

where the parameter λ_m is equal to the expectation of the observation w_m , which, in its turn, is described by the expectation model. Therefore,

$$E[w_m] = \lambda_m = f_m(\tau) \quad (20)$$

with $f_m(\tau)$ given by Eq. (18). A property of the Poisson distribution is that the variance of the observation w_m is equal to λ_m :

$$\text{var}(w_m) = \lambda_m \quad (21)$$

Moreover, electron microscopic observations may be assumed to be statistically independent. The probability $P(\omega; \tau)$ that a set of observations $w = (w_1 \dots w_M)^T$ is equal to $\omega = (\omega_1 \dots \omega_M)^T$ is thus equal to the product of all probabilities described by Eq. (19):

$$P(\omega; \tau) = \prod_{m=1}^M \frac{\lambda_m^{\omega_m}}{\omega_m!} \exp(-\lambda_m) \quad (22)$$

This function is called the *joint probability density function* of the observations. It represents the parametric statistical model of the observations. The parameters τ to be estimated enter $P(\omega; \tau)$ via λ_m .

If the expectation $E[w_m] = \lambda_m$ increases, the Poisson distribution tends to a normal distribution with both expectation $E[w_m]$ and variance $\text{var}(w_m)$ equal to λ_m of the Poisson distribution (Mood et al., 1974). This normal approximation is justified if the magnitude of the observations w_m is large with respect to the square root of this number (Koster et al., 1987). Moreover, if the contrast in the images is low, the deviations of the observations from their expectations may be supposed to be identically distributed, that is, $\text{var}(w_m) = \sigma^2$ (Miedema et al., 1994). Under these conditions the joint probability density function of the observations is given by

$$P(\omega; \tau) = \prod_{m=1}^M \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{\omega_m - \lambda_m}{\sigma}\right)^2\right] \quad (23)$$

In Section 5.2, the parameterized joint probability density function will be used to derive the CRLB, that is, an expression for the attainable precision with which the unknown parameters can be estimated unbiasedly from the observations. In Section 5.3, from the joint probability density function, the ML estimator of the parameters will be derived. This estimator actually achieves the CRLB asymptotically, that is, for the number of observations going to infinity.

5.2 The Cramér–Rao Lower Bound

In this section, the parameterized probability density function of the observations, which is derived in Section 5.1, will be used to define the *Fisher information matrix* and to compute the CRLB on the variance of unbiased estimators of the parameters of the expectation model. The CRLB will also be extended to include unbiased estimators of vectors of *functions of these parameters*. As discussed in the introduction of Section 5, the CRLB is used as a criterion for statistical model-based resolution. The reader is referred to van den Bos (1982), Frieden (1998), and van den Bos and den Dekker (2001) for the details of the CRLB.

First, the *Fisher information matrix* F with respect to the elements of the $R \times 1$ parameter vector $\tau = (\tau_1 \dots \tau_R)^T$ is introduced. It is defined as the $R \times R$ matrix

$$F = -E \left[\frac{\partial^2 \ln P(\mathbf{w}; \tau)}{\partial \tau \partial \tau^T} \right] \quad (24)$$

where $P(\omega; \tau)$ is the joint probability density function of the observations $w = (w_1 \dots w_M)^T$. The expression between square brackets represents the Hessian matrix of $\ln P$, for which the (r, s) th element is defined by $\partial^2 \ln P(\omega; \tau) / \partial \tau_r \partial \tau_s$. For electron microscopic observations, where $P(\omega; \tau)$ is given by Eq. (22), it follows from Eqs. (20), (22), and (24) that the (r, s) th element of F is equal to

$$F_{rs} = \sum_{m=1}^M \frac{1}{\lambda_m} \frac{\partial \lambda_m}{\partial \tau_r} \frac{\partial \lambda_m}{\partial \tau_s} \quad (25)$$

Furthermore, if the approximation of Eq. (22) by Eq. (23) is justified, the (r, s) th element of F is equal to

$$F_{rs} = \sum_{m=1}^M \frac{1}{\sigma^2} \frac{\partial \lambda_m}{\partial \tau_r} \frac{\partial \lambda_m}{\partial \tau_s} \quad (26)$$

Next, it can be shown that the covariance matrix $\text{cov}(\hat{\tau})$ of any unbiased estimator $\hat{\tau}$ of τ satisfies

$$\text{cov}(\hat{\tau}) \geq F^{-1} \quad (27)$$

This inequality indicates that the difference of the matrices $\text{cov}(\hat{\tau})$ and F^{-1} is positive semidefinite. Since the diagonal elements of $\text{cov}(\hat{\tau})$ represent the variances of $\hat{\tau}_1, \dots, \hat{\tau}_R$ and since the diagonal elements of a positive semidefinite matrix are nonnegative, these variances are larger than or equal to the corresponding diagonal elements of F^{-1} :

$$\text{var}(\hat{\tau}_r) \geq [F^{-1}]_{rr} \quad (28)$$

where $r = 1, \dots, R$ and $[F^{-1}]_{rr}$ is the (r, r) th element of the inverse of the Fisher information matrix. In this sense, F^{-1} represents a lower bound to the variances of all unbiased $\hat{\tau}$. The matrix F^{-1} is called the CRLB on the variance of $\hat{\tau}$.

The CRLB can be extended to include unbiased estimators of vectors or functions of the parameters instead of the parameters proper. Let $\gamma(\tau) = [\gamma_1(\tau) \dots \gamma_C(\tau)]^T$ be such a vector and let $\hat{\gamma}$ be an unbiased estimator of $\gamma(\tau)$. Then, it can be shown that

$$\text{cov}(\hat{\gamma}) \geq \frac{\partial \gamma}{\partial \tau^T} F^{-1} \frac{\partial \gamma^T}{\partial \tau} \quad (29)$$

where $\partial \gamma / \partial \tau^T$ is the $C \times R$ Jacobian matrix defined by its (r, s) th element $\partial \gamma_r / \partial \tau_s$ (van den Bos, 1982). The right-hand member of this inequality is the CRLB on the variance of $\hat{\gamma}$.

It should be noted that the CRLB may be computed only if the probability density function of the observations is known. At first sight, this seems to be a problem since the true parameters of the probability density function are unknown. Nevertheless, even if the CRLB is a function of the unknown parameters, it remains an extremely useful tool. For nominal values of the unknown parameters it enables one to quantify variances that might be achieved, to detect possibly strong covariances between parameter estimates, and, as will be shown in Section 5.4, to optimize the experimental design (van den Bos, 1982). Moreover, the estimates obtained using an estimator that achieves the CRLB may be substituted for the true parameters in the expression for the CRLB so as to obtain a level of confidence to be attached to these estimates (den Dekker and Van Aert, 2002; den Dekker et al., 2005; Van Aert et al., 2005). This will briefly be discussed in Section 5.3.

The *attainable* precision with which position and distance parameters of one or two point objects can be measured has been investigated. The observations consist of counting results in a one- or two-dimensional pixel array. The model describing the expectations of these observations has been assumed to consist of Gaussian peaks with unknown position. Under this assumption, the CRLB, which usually has to be calculated numerically, may be approximated by a simple rule in closed analytical form. Although the expectation model of images obtained in practice are generally of higher complexity than Gaussian peaks, the rules are suitable to provide insight in the attainable precision of position and distance parameters using quantitative atomic resolution TEM. In Bettens et al. (1999) and Van Aert et al. (2002a), the details of this study can be found. In the following examples, only the main results will be presented.

Example 2 (Position Measurement) *In this example, the attainable precision of the position measurement of one isolated point object will be considered. If the point spread function is assumed to be Gaussian and defined by Eq. (1) and if the total number of imaging particles is N , the lower bound on the*

standard deviation s_{LB} of the coordinates estimates of the position, that is, the square root of the CRLB, is given by

$$s_{\text{LB}} \approx \frac{\rho}{\sqrt{N}} \quad (30)$$

or, from Eq. (3)

$$s_{\text{LB}} \approx \frac{\rho_p}{2\sqrt{2N}} \quad (31)$$

Thus, the precision with which the position can be determined is a function of both the Rayleigh resolution ρ_p and the number of imaging photons N . If N is large, the precision can be orders of magnitude higher than the point resolution ρ_p .

Example 3 (Distance Measurement) Here, the attainable precision of the distance measurement of two neighboring point objects will be considered. If the imaging is supposed to be linear, the image consists of the superposition of the two corresponding point spread functions as simulated in Figures 20–6 and 20–7. The total number of imaging photons used in the simulation is equal to $2N$. This means that in agreement with Figure 20–5, the number of photons per peak is equal to N . The results are different for distances smaller than or larger than $\rho_p/2$ and may be summarized as follows.

- $d > \rho_p/2$

This is shown in Figure 20–6. Now, the distance d between the atoms is larger than the half of the Rayleigh resolution. Then, the lower bound on the standard deviation s_{MIN} on the distance d is minimal, independent of d , and given by

$$s_{\text{MIN}} \approx \frac{\rho_p}{2\sqrt{N}} \quad (32)$$

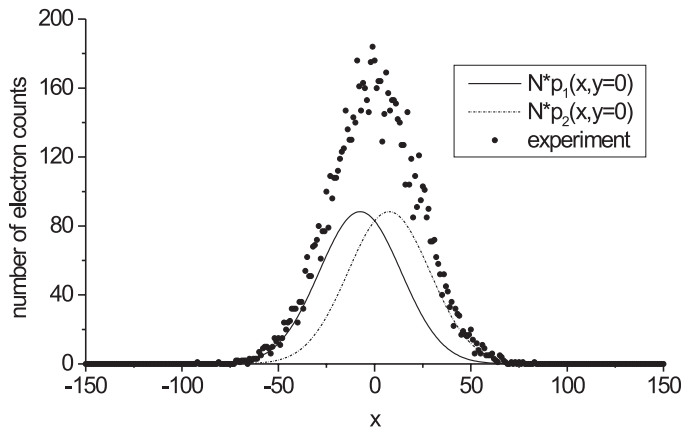


Figure 20–7. The results of a computer-simulated image of two neighboring points with a Gaussian point spread function, where $p_{1,2}(x, y)$ are two-dimensional normalized Gaussian point spread functions, as in Eq. (1). The experimental as well as the expectation values of the individual peaks $[N \times p_{1,2}(x, y)]$ are plotted within the section $y = 0$. The number of imaging particles $2N$ is equal to 500,000. The peaks overlap severely.

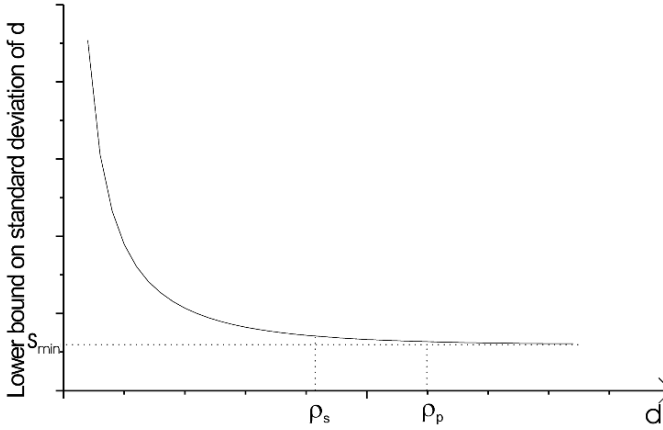


Figure 20–8. The lower bound on the standard deviation of the distance d as a function of the distance.

Thus, from Eq. (31) it follows that the CRLB on the variance of the distance is twice the CRLB on the variance of the position of an isolated point object:

$$s_{\text{MIN}}^2 = 2s_{\text{LB}}^2 \quad (33)$$

The reason for this is that the coordinate estimates of neighboring atoms are uncorrelated for distances larger than $\rho_p/2$.

- $d \leq \rho_p/2$

This is shown in Figure 20–7. Now, the distance d between the atoms is smaller than the half of the Rayleigh resolution. Then, the lower bound on the standard deviation s_{LB} increases inversely proportionally to the distance d , following approximately the relation

$$s_{\text{LB}}(d) \approx \frac{s_{\text{MIN}}\rho_p}{2d} \quad (34)$$

where s_{MIN} is given by Eq. (32).

In Figure 20–8, the lower bound on the standard deviation of the distance d is shown as a function of the distance.

In this section, statistical model-based resolution has been discussed in terms of the CRLB. In this discussion, the parametric statistical model of the observations, that is, the joint probability density function, has been assumed to be exact. Then, it is clear that statistical model-based resolution is limited by the lower bound on the variance or on the standard deviation. This limit is fundamental. Moreover, if the model is inaccurate, the introduced systematic error determines another fundamental limit to statistical model-based resolution. Apart from these fundamental limits, computational limits exist as well, just as for deterministic model-based resolution. The computing time needed to fit the model to the images with respect to the unknown parameters increases with the total number of parameters to be

measured. Furthermore, the optimization methods fail if the structure is not resolved.

5.3 Maximum Likelihood Estimation

In this section, the derivation of the ML estimator of the parameters from the parameterized probability density function, which is discussed in Section 5.1, will be discussed. This estimator is very important since it achieves the CRLB asymptotically, that is, for the number of observations going to infinity. Thus, it is asymptotically most precise and is therefore often used in practice. The ML estimator is clearly discussed in van den Bos and den Dekker (2001) and den Dekker et al. (2005). A summary is given here.

The ML method for estimation of the parameters consists of three steps:

1. The available observations $w = (w_1 \dots w_M)^T$ are substituted for the corresponding independent variables $\omega = (\omega_1 \dots \omega_M)^T$ in the probability density function, for example, in Eq. (22) or in Eq. (23). Since the observations are numbers, the resulting expression depends only on the elements of the parameter vector $\tau = (\tau_1 \dots \tau_R)^T$.
2. The elements of $\tau = (\tau_1 \dots \tau_R)^T$, which are the hypothetical true parameters, are considered to be variables. To express this, they are replaced by $t = (t_1 \dots t_R)^T$. The logarithm of the resulting function, $\ln P(w; t)$, is called the *log-likelihood function* of the parameters t for the observations w , which is denoted as $q(w; t)$.
3. The ML estimates $\hat{\tau}_{ML}$ of the parameters τ are defined by the values of the elements of t that maximize $q(w; t)$, or

$$\hat{\tau}_{ML} = \arg \max_t q(w; t) \quad (35)$$

For independent normally distributed observations with equal variance, for which the joint probability density function is given by Eq. (23), it can easily be shown that the ML estimator is equal to the well-known *uniformly weighted least-squares estimator* (van den Bos and den Dekker, 2001; den Dekker et al., 2005). The uniformly weighted least-squares estimates $\hat{\tau}_{LS}$ of the unknown parameters τ are given by the values of t that minimize the *uniformly weighted least-squares criterion*:

$$\hat{\tau}_{LS} = \arg \min_t \sum_{m=1}^M [w_m - f_m(t)]^2 \quad (36)$$

where the function f_m is defined by Eq. (18).

The most important properties of the ML estimator are the following:

- *Consistency.* Generally, an estimator is said to be consistent if the probability that an estimate deviates more than a specified amount from the true value of the parameter can be made arbitrarily small by increasing the number of observations used.
- *Asymptotic normality.* If the number of observations increases, the probability density function of an ML estimator tends to a normal distribution.

- *Asymptotic efficiency.* The asymptotic covariance matrix of an ML estimator is equal to the CRLB. In this sense, the ML estimator is most precise.
- *Invariance property.* The ML estimates $\hat{\gamma}_{\text{ML}}$ of a vector of functions of the parameters τ , that is, $\gamma(\tau) = [\gamma_1(\tau) \dots \gamma_C(\tau)]^T$, are equal to $\gamma(\hat{\tau}_{\text{ML}}) = [\gamma_1(\hat{\tau}_{\text{ML}}) \dots \gamma_C(\hat{\tau}_{\text{ML}})]^T$ (Mood et al., 1974).

To evaluate the level of confidence to be attached to the obtained ML estimates, confidence regions and intervals associated with these estimates are required. A summary of existing methods to compute such regions and intervals is presented in den Dekker et al. (2005). One of these methods is based on the asymptotic normality of the ML estimator. It is preferred, especially if the experiment cannot be replicated. For example, an approximate $100(1 - \alpha)\%$ confidence interval for the r th element of τ , τ_r , is given by

$$[\hat{\tau}_{\text{ML}}]_r \pm \lambda_{(1-\alpha/2)} \sqrt{[F^{-1}]_{rr}} \quad (37)$$

with $[\hat{\tau}_{\text{ML}}]_r$ the r th element of the parameter vector $\hat{\tau}_{\text{ML}}$, $[F^{-1}]_{rr}$ the (r, r) element of F^{-1} , and $\lambda_{(1-\alpha/2)}$ the $(1 - \alpha/2)$ quantile of the standard normal distribution. The meaning of a $100(1 - \alpha)\%$ confidence interval is that it covers the true element τ_r of τ with probability $1 - \alpha$. Usually, F^{-1} is a function of the parameters to be estimated. The confidence intervals (37) are then derived by using approximations of F^{-1} . A useful approximation $\hat{F}^{-1}$ of F^{-1} may be obtained by substituting $\hat{\tau}_{\text{ML}}$ for τ in the expression for the CRLB yielding

$$\hat{F}^{-1} = F^{-1}|_{\tau=\hat{\tau}_{\text{ML}}} \quad (38)$$

The effectiveness of the ML method will be shown in the following example, where it will be applied to experimental high-resolution TEM (HRTEM) images of an aluminum crystal. The details of this study can be found in Van Aert et al. (2005).

Example 4 (Maximum Likelihood Estimation of Structure Parameters from HRTEM Images of an Aluminum Crystal) Structure parameters, atom column distances in particular, have been estimated from HRTEM images of an aluminum crystal using the ML method. Therefore, 20 images have been recorded and afterward been corrected for specimen drift using cross-correlation. From these corrected images, a particular region consisting of 51 times 50 pixels has been selected. The individual images resulting from this procedure are shown in Figure 20–9. The observations corresponding to these individual images, that is, the 51 times 50 image pixel values, are supposed to be independent normally distributed with equal variance. For such observations, the joint probability density function is given by Eq. (23). Moreover, the expectation model has been supposed to be given by Eq. (10). The unknown parameters of this model have then been estimated using the ML estimator, which for these types of observations is equal to the uniformly weighted least-squares estimator given by Eq. (36). Furthermore, ML estimates of the atom column distances have been obtained using the invariance property of the ML estimator. Also confidence intervals for the parameters have been computed. For example, for the atom column distances, the

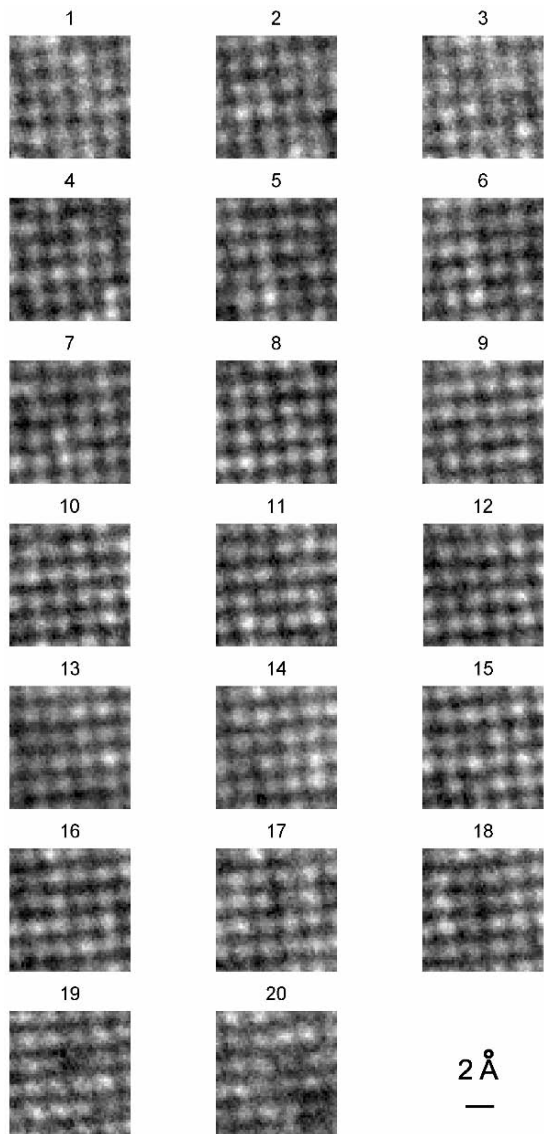


Figure 20-9. HRTEM observations of an aluminum crystal.

precision has been shown to be of the order of $0.03 \text{ \AA}$, with the precision represented by the square root of the approximated CRLB on the variance of the distance with the approximated CRLB given by Eq. (38). Obviously, the value to be attached to the obtained estimates and confidence intervals depends on the validity of the model. Therefore, it is important to test the proposed model before attaching confidence to these estimates. This has been done using the model assessment methods described in den Dekker et al. (2005) and Van Aert et al. (2005). From these methods, it could be concluded that the proposed model is a sufficiently adequate description of the experimental observations. The model evaluated at the obtained ML estimates is shown in Figure 20-10. Note that this figure may be regarded as an optimal image reconstruction.

Finally, the variation of the atom column distance estimates has been compared to the variance that would be expected if the statistical fluctuations of the observations would be the only source of variation of the distance estimates. The former has been measured by means of the well-known sample variance (Mood et al., 1974) and the latter by means of the CRLB. If statistical fluctuations would be the only source of contribution to the variation of the distance estimates, the sample variance and the estimated CRLB should be about equal to one another. However, from a comparison of both, it followed that the sample variance was about 10 times larger than the estimated CRLB. Hence, there must be a further, dominant contribution to the variation of the distance estimates. Because of this the atom columns of the crystal were not observed at a perfectly periodic crystal lattice in the experiment and, in

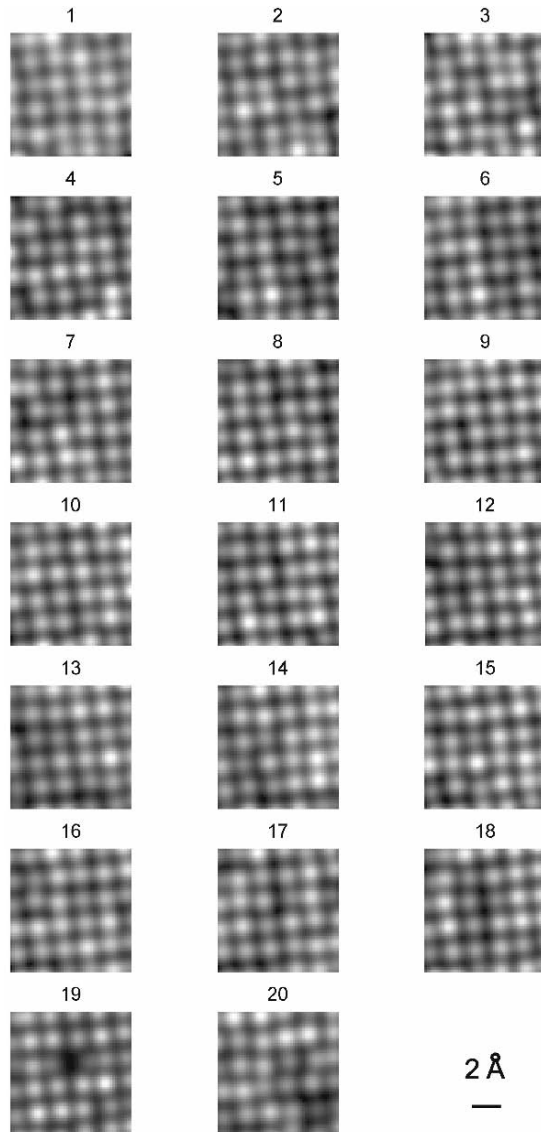


Figure 20-10. The expectation models as described by Eq.(10) evaluated at the estimated parameters obtained from the experimental images shown in Figure 20-9 using the ML method.

addition, the atom columns were spotted at different locations in the course of time.

The ML method is, of course, also applicable to other types of experiments for the estimation of unknown parameters. An interesting application to EELS spectra can be found in Verbeeck and Van Aert (2004).

5.4 Statistical Experimental Design

In this section, it will be shown how the statistical model-based resolution of a quantitative atomic resolution TEM experiment may be improved by optimizing its design.

Usually, instrumental developments in microscopy, telescopy, or even spectrometry aim at improving the Rayleigh resolution or Rayleigh-like resolution criteria. Examples of such developments in electron microscopy are the spherical aberration corrector (Rose, 1990; Haider et al., 1998) and the monochromator (Mook and Kruit, 1999). A resolution (in the sense of Rayleigh) beyond 1 Å allows microscopists to visually distinguish atom columns of solids oriented along a main zone axis. In addition, electron microscopes now have increased versatility of microscope settings, which will preferably be computer controllable in the future. Electron microscopists can choose between TEM or STEM, between imaging or diffraction techniques, focus, accelerating voltage, spherical aberration, energy spread of incoming electrons, beam tilt, and crystal tilt. The main constraints of the experiment will be the radiation sensitivity of the object or the specimen drift. The question then arises as to which microscope and which settings are optimal in keeping the incident electron dose per square ångström or the recording time subcritical. In this section this question will be addressed using statistical experimental design. Statistical experimental design can be defined as the selection of free experimental settings in an experiment to improve the precision with which unknown parameters can be estimated. For quantitative atomic resolution TEM experiments, these parameters are the atom or atom column positions in particular. The optimal design is then given by the combination of experimental settings for which the precision of the atom or atom column positions is highest. In the optimization of the experimental design, the CRLB is taken as the optimal criterion. This criterion thus aims at an improvement of statistical model-based resolution rather than an improvement of Rayleigh resolution or Rayleigh-like resolution criteria. Note that the optimal statistical experimental design may be different for different objects under study.

In Section 5.2, it was shown how from the joint probability density function, which was introduced in Section 5.1, the elements of the Fisher information matrix may be calculated explicitly. From the latter, the CRLB on the variance of the parameters of the expectation model and on the variance of functions of these parameters may be computed from the right-hand side of Eqs. (27) and (29), respectively. The diagonal elements of the CRLB give a lower bound on the variance of any

unbiased estimator of the parameters. Since the joint probability density function is a function of the experimental settings, the CRLB is a function of these settings as well. Therefore, the CRLB may be used to evaluate and to optimize the experimental design in terms of precision. However, simultaneous minimization of the diagonal elements of the CRLB, that is, the right-hand side of Eq. (28), is usually impossible. Therefore, statistical parameter estimation theory provides different optimality criteria, which are functions of the elements of the CRLB. These are scalar measures. Experimenters may choose one of these criteria or may produce their own criterion, reflecting their purpose. The interested reader may find a selection of such criteria in Fedorov (1972), Pázman (1986), and Van Aert et al. (2004).

Statistical experimental design can be illustrated as follows. Suppose the microscope is able to visualize individual atoms or atom columns so that the structure can be resolved. In other words, neighboring atoms or atom columns can be discriminated in the images. Furthermore, it will be assumed that the expectations of the Poisson distributed image pixel values can be modeled as a summation of Gaussian peaks centered at the atom or atom column positions. The position coordinates of these atoms or atom columns are the unknown parameters in the expectation model. It then follows from Eq. (31) that the lower bound on the standard deviation of these coordinate estimates, that is, the square root of the CRLB, is approximately given by $\rho_p / \sqrt{2N}$. In this expression, ρ_p represents the Rayleigh resolution, which depends on the Rayleigh resolution of all different subchannels contributing to the imaging process as follows from Eq. (7) and N represents the number of detected electrons per atom or atom column. It is now clear that it is not only the Rayleigh resolution that matters but the electron dose as well. If the Rayleigh resolution can be improved only at the expense of a decrease in the number of detected electrons, both effects have to be balanced under the existing physical constraints so as to produce the highest precision. Furthermore, it follows from Eq. (7) that below a certain value for ρ_p , it will be useless to further improve the Rayleigh resolution of the electron microscope since scattering is dominant or, equivalently, ρ_p is almost equal to the width of the electrostatic potential of an atom, that is, ρ_A . This result is meaningful in practice. For example, in STEM experiments, further narrowing of the probe, which represents the point spread function of the electron microscope, is not so beneficial in terms of precision since the width of the probe is currently almost equal to the width of an atom (Krivanek et al., 2002).

In the following examples, the main results obtained from the numerical optimization and evaluation of the design of quantitative atomic resolution TEM experiments will be presented. Also the possible benefits of new developments in instruments will be discussed. In these examples, expectation models with a solid physical base, instead of Gaussian peaks, have been assumed and the observations are assumed to be independent and Poisson distributed. The results obtained may intuitively be interpreted using the rules for the CRLB, which have been obtained from Gaussian peaked expectation models.

Comprehensive descriptions of these results can be found in den Dekker et al. (1999, 2001), Van Aert and Van Dyck (2001), and Van Aert et al. (2002a–c, 2004).

Example 5 (Statistical Experimental Design of Scanning Transmission Electron Microscopy) The expectation model that has been used in the evaluation and optimization of the experimental design of STEM experiments is based on the simplified channelling theory to describe the dynamic, elastic scattering of the electrons with atom columns (Broeckx et al., 1995; Geuens and Van Dyck, 2002; Pennycook and Jesson, 1992; Van Aert et al., 2002b; Van Dyck and Op de Beeck, 1996). Therefore, it has been assumed that the dynamic motion of an electron in a column may be primarily expressed in terms of so-called tightly bound 1s states, which are peaked at the atom column positions. Furthermore, in this model, temporal incoherence due to chromatic aberration has not been taken into account since STEM imaging seems to be robust to chromatic aberration (Batson, et al., 2002; Krivanek et al., 2002; Nellist and Pennycook, 1998, 2000). Thermal diffuse, inelastic scattering has not been taken into account for the following reasons. Thermal diffuse scattered electrons are predominantly collected in the detector at high angles (Treacy, 1982). Therefore, increasing the inner detector angle of an annular detector has the effect of increasing thermal diffuse, inelastic scattering relative to elastic scattering (Wang, 2001). Usually, the inner detector angle is chosen to be relatively large since the detected signal then strongly depends on the atomic number Z , hence the name, Z -contrast imaging. The disadvantage of increasing this detector angle, however, is the accompanied decrease of dose efficiency, which leads to a decrease in signal-to-noise ratio (SNR). It can be shown that as a result of this decrease in SNR, the optimal inner collection angle in terms of precision is small compared to the angles where thermal diffuse scattering is important. This justifies the fact that thermal diffuse scattering has not been taken into account. The details of the expectation model can be found in Van Aert et al. (2002b, 2004).

It has been shown in Van Aert and Van Dyck (2001) and Van Aert et al. (2002b, 2004) that the optimal aperture radius in terms of precision may be considerably smaller than the aperture radius at the Scherzer conditions for incoherent imaging, which is referred to as being optimal in terms of direct visual interpretability (Scherzer, 1949; Pennycook and Jesson, 1991). This indicates that the optimal width of diffraction-limited probes, for which the width is inversely proportional to the aperture radius, is not as small as possible. The optimal objective aperture radius, which determines the size of a diffraction-limited probe, has been found to be mainly determined by the object under study. More specifically, for isolated atom columns, it is proportional to the depth of the two-dimensional electrostatic potential of the atom column projected along the column direction, that is, the difference between the maximum and minimum potential energy. In terms of the column-specific 1s state, this means that the main lobe of the optimal probe is broader than the 1s state. This is shown in Figure 20–11, where both the 1s state and the amplitude of the optimal probe are shown for a silicon and a gold [100] atom column. However, in the presence of neighboring columns, it has been found that the optimal aperture radius may increase so as to avoid strong overlap of neighboring columns in the image.

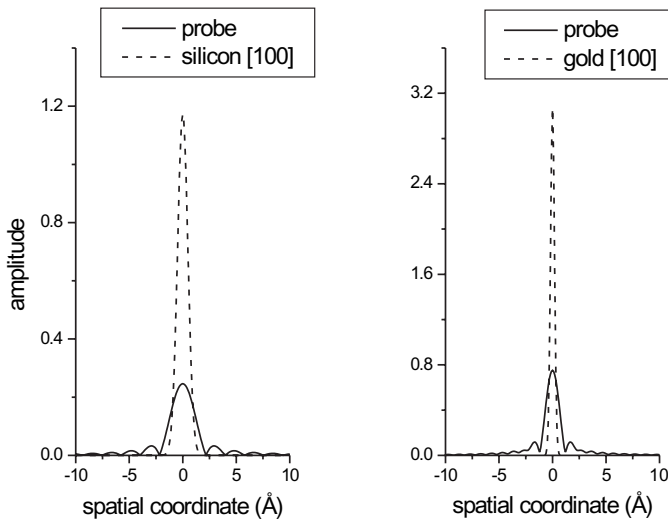


Figure 20-11. The dashed curves represent the 1s state for a silicon [100] (left) and a gold [100] (right) atom column, respectively. The solid curves represent the amplitude of their associated optimal STEM probes for C_s equal to 0.5mm.

In the evaluation of the experimental design, annular and axial detector types have been compared. Apart from some exceptions, an annular detector usually results in higher precision than an axial one. The optimal inner radius of an annular detector has been found to be equal to the optimal objective aperture radius, whereas in a conventional approach the radius of the hole is usually taken to be much larger than the aperture radius (Pennycook et al., 1995; Hartel et al., 1996). The optimal outer radius of an axial detector is usually slightly smaller than the optimal aperture radius. However, if this detector leads to very low contrast of the image, the optimal detector radius decreases.

Moreover, it has been found that a spherical aberration corrector improves the precision. The accompanied gain depends on the object under study. Correction of the spherical aberration is more useful in terms of precision for heavy than for light atom columns. This is shown in Figures 20-12 and 20-13, where the ratio of the lower bound on the standard deviation of the position coordinates for a given spherical aberration constant to the lower bound for a spherical aberration constant of 0mm is shown as a function of the spherical aberration constant. This is done for an isolated silicon and gold [100] atom column, respectively. Moreover, the evaluation is done for an annular as well as an axial detector. For each spherical aberration constant considered, the objective aperture radius is set to its optimal value. Furthermore, the detector radius is taken to be equal to this optimal objective aperture radius. For silicon, which is a light atom column, it follows from Figure 20-12 that the precision that is gained by reducing the spherical aberration constant from 0.5mm to 0mm is a factor of 1.0009 and 1.0011 for an annular and axial detector, respectively. These gains are negligible. For gold, which is a heavy atom column, it follows from Figure 20-13 that the precision that is gained by reducing the spherical aberration constant from 0.5mm to 0mm is a factor of 1.21 and 1.39 for an annular and axial detector, respectively, which is still

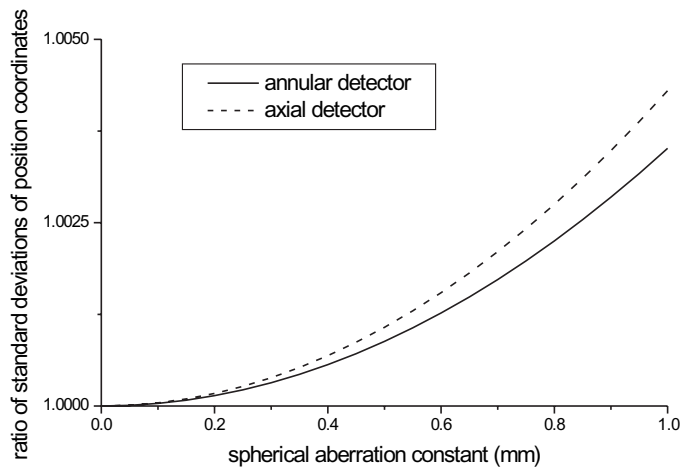


Figure 20-12. Ratio of the lower bound on the standard deviation of the position coordinates for a given spherical aberration constant to the lower bound for a spherical aberration constant of 0mm for an isolated silicon [100] atom column as a function of the spherical aberration constant for an annular as well as for an axial STEM detector.

relatively small. This makes it questionable whether such a corrector is needed to obtain a prespecified precision of the atom column positions.

Example 6 (Statistical Experimental Design of High-Resolution Transmission Electron Microscopy) *The optimal experimental design of HRTEM experiments has been reconsidered in terms of statistical model-based resolution (den Dekker et al., 1999, 2001; Van Aert et al., 2004). The expecta-*

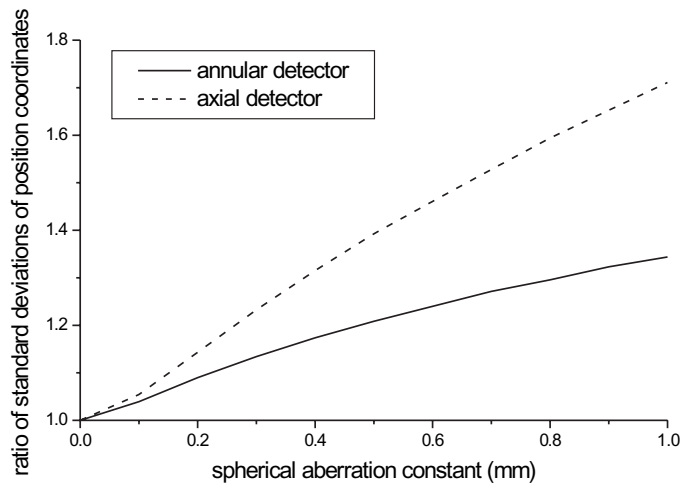


Figure 20-13. Ratio of the lower bound on the standard deviation of the position coordinates for a given spherical aberration constant to the lower bound for a spherical aberration constant of 0mm for an isolated gold [100] atom column as a function of the spherical aberration constant for an annular as well as for an axial STEM detector.

tion model that has been used in the evaluation and optimization of HRTEM experiments is based on the simplified channeling theory (Geuens and Van Dyck, 2002; Van Dyck and Op de Beeck, 1996). For microscopes operating at intermediate accelerating voltages of the order of 300kV, it has been shown that use of a spherical aberration corrector or a chromatic aberration corrector is only of limited value and that the use of a monochromator usually is not worthwhile in terms of precision, presuming that specimen drift places a practical constraint on the experiment. These results follow from Figure 20–14, where the lower bound on the standard deviation of the position coordinates of a gold atom column is evaluated as a function of the spherical aberration constant. The solid curve corresponds to a microscope without correction for chromatic aberration, that is, a microscope without a chromatic aberration corrector and monochromator. The dashed curve corresponds to a microscope with a chromatic aberration corrector, for which the chromatic aberration constant is equal to 0mm. The dotted curve corresponds to a microscope with a monochromator, for which the energy spread corresponds to a typical full width at half maximum height of 200 meV (Batson, 1999). In Figure 20–14 it is assumed that the specimen drift is the relevant physical constraint. Hence, the recording time is kept constant. Consequently, the total number of detected electrons is smaller in the presence of a monochromator. The values for the other microscope settings, that is, for the original, nonoptimized microscope settings, are in accordance with values that are typical for today's electron microscopes. Therefore, it also follows from Figure 20–14 that for today's electron microscopes it is in principle possible to obtain a precision of the order of 0.01 Å, even with a microscope that is not corrected for spherical and chromatic aberration.

However, for amorphous instead of crystalline structures, the conclusions are different since amorphous structures are very sensitive to radiation damage,

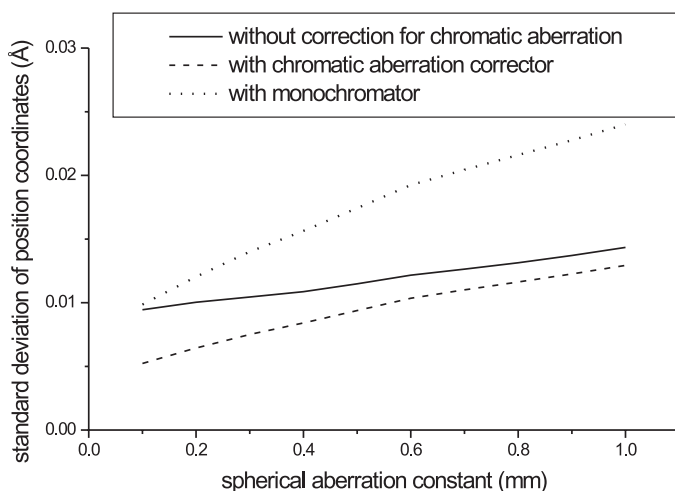


Figure 20–14. The lower bound on the standard deviation of the position coordinates of a gold [100] atom column as a function of the spherical aberration constant for a microscope operating at 300kV equipped with or without a chromatic aberration corrector or monochromator. In this evaluation, the recording time is kept constant.

so that radiation sensitivity rather than specimen drift will be the limiting factor in the experiment. Although the precision improves by increasing the incident dose, it has to be taken into account that every incident electron has a finite probability of damaging the structure. The structure can be damaged either by displacing an atom from its position or by changing chemical bonds due to ionization. Therefore, a compromise between precision and radiation damage has to be made, which turns out to be optimal when the accelerating voltage is reduced. However, at low accelerating voltages, the instrumental aberrations become important. For this reason, correction of both the spherical aberration and the chromatic aberration by either a chromatic aberration corrector or a monochromator will be essential. By means of statistical experimental design, it has been shown that a substantial improvement of the precision may be obtained if both the spherical and chromatic aberration are corrected (Van Dyck et al., 2003).

6 Ultimate Model-Based Resolution

In the preceding section, the CRLB on the precision of the estimated distance of two peaks was used as a resolution criterion. Such a criterion makes sense only if an estimator exists that actually achieves this bound. Such an estimator exists in the form of the so-called ML estimator. This estimator produces estimates that are normally distributed about the true values of the parameters with a variance equal to the CRLB. However, this is true only if the number of observations used by the ML estimator is sufficiently large. For small numbers of observations, the estimates have statistical properties that are very different. A detailed description is given in van den Bos (1992) and van den Bos and den Dekker (1995, 2001).

6.1 Two Types of Observations

Central in this section is the space of observations, which has been introduced in Section 5.1. This is the Euclidean space in which every coordinate direction corresponds to a particular observation. Therefore, the dimension of the space of observations is equal to the number of observations and a particular set of observations is represented by a single point in the space. For example, the observations shown in Figure 20–7 correspond to a single point in the space of observations. It has been shown in the references mentioned that two types of sets of two-peak observations may occur. If the two-peak model is fitted to the first type of observations, *distinct* values for the locations are obtained. Then the distance of the peaks is different from zero. If, on the other hand, the two-peak model is fitted to observations of the second type, the estimates for the locations *exactly coincide*. Then the distance of the peaks is equal to zero. This means that the peaks exactly coincide. It is clear that from the first type of observations both peaks are resolved in the sense that they occur at distinct locations. On the other hand, for the second type of observations, the peaks occur at the same location. Therefore, the peaks add up to form a single peak and cannot be resolved. From these considerations, it follows that the (hyper)surface

separating both types of observations in the space of observations may be used as a resolution criterion. If the point representing a particular set of observations is on the one side of the hypersurface, the peaks are resolved, but if it is on the other side, they are not.

This perhaps unexpected behavior of the solutions for the locations of the peaks has to do with the *structure* of criteria of goodness of fit for the estimation of parameters such as the locations. The structure is the pattern of the stationary points of the criterion, where a stationary point is a point at which the gradient is equal to zero, such as a minimum or maximum. The structure is important since the solution of the model fitting problem is represented by a stationary point: the absolute optimum of the criterion of goodness of fit. It has been found that for observations made on overlapping peaks, two different structures may occur and that these correspond to distinct or coinciding solutions for the locations, respectively. Which of these structures occurs depends on the particular realization of the observations.

6.2 Consequences for Resolution

If the observations are statistical, they are, in the space of the observations, distributed about the point representing their expectations. This *expectation point*, therefore, represents ideal exact observations. If the model fitted is correctly specified, it perfectly fits these exact observations. Then the solutions found for the locations are distinct and equal the hypothetical true locations, no matter how closely located the peaks. Therefore, the expectation point will be located on the “two-peak side” of the separating hypersurface and not on the “one-peak side,” since the solutions are distinct. Observations distributed about the expectation point will, with a certain probability, only be on the two-peak side of the separating hyperplane. It is concluded that resolution should be described in terms of *probability of resolution*. The distance of the expectation point and the separating hypersurface becomes smaller and smaller as the hypothetical true locations are closer. This means that the probability that a set of observations is located on the one-peak side of the separating hyperplane increases and the probability of resolution decreases correspondingly.

The separating hypersurface is computed from and, therefore, specific to the peak model fitted. In this chapter, this is often the Gaussian peak. The expectation point, however, is specific to the model actually present in the observations. If the model fitted and the model present differ, the expectation point may even be located on the one-peak side of the separating hyperplane. This implies that the peaks are not resolved even if the statistical errors in the observations are very small.

It is clear that, ultimately, resolution is limited by errors, both systematic (modeling) errors and nonsystematic (statistical) errors. This limit is fundamental: neither an increase in computer power nor improved software can remove it. It would be present even if initial conditions would be available for the locations of all peaks and sufficient computer power would be available for fitting them. Removing these limits requires more and preferably different observations of the same peaks.

7 Discussion and Conclusions

In this chapter, resolution has been investigated in detail in successive steps. Throughout this chapter, emphasis has been put on electron microscopy.

First, classical resolution criteria, such as Rayleigh's and Sparrow's, have been discussed. These criteria are expressed in terms of the width of the point spread function. The narrower the point spread function, the higher the resolution.

Next, the diffraction limit to resolution and its relation to Rayleigh and Sparrow resolution have been considered. The diffraction limit is given by the cutoff frequency, which is the highest spatial frequency that is transferred by the imaging system. It has been shown that the diffraction limit is inversely proportional to the Rayleigh resolution. For electron microscopy applications, it has been shown that the "width" of the electrostatic potential of the atoms as well as the effect of thermal vibrations of the atoms, the environment, and the detection have to be taken into account. As a result of these extensions, it should be concluded that the atoms themselves limit the diffraction limit and classical resolution criteria. Also the notion of superresolution has been introduced. Using superresolution algorithms, frequency components lying beyond the diffraction limit can be reconstructed. Such algorithms use prior knowledge of the object imaged by the imaging system.

Such prior knowledge has been incorporated in the form of a parametric model. Indeed, classical resolution criteria are in fact concerned with calculated images, that is, noise-free images exactly describable by a known parameterized mathematical model. Then, the unknown parameters, such as the positions of projected atoms, can be measured by means of model fitting. The practical limits to this so-called deterministic model-based resolution are of a computational kind. As an example, amorphous structures studied by atomic resolution TEM have been discussed. For these structures, computational problems can be overcome only if the thickness is very small. An upper bound to the thickness is derived, which for amorphous silicon is of the order of 10 Å. For realistic thicknesses, electron tomography is needed. Moreover, fundamental limits to deterministic model-based resolution exist if the model is inaccurate, since then, systematic errors are introduced.

Then, unavoidable noise in the observations has been incorporated by considering detected images instead of calculated images. The fundamental limit to this so-called statistical model-based resolution is determined by the CRLB, which is a lower bound on the variance with which the positions of, or the distance between, projected atoms or atom columns can be measured using parameter estimation methods. In this discussion, the ML estimator is very important since it achieves the CRLB asymptotically. The effectiveness of the ML method has been shown in an example, in which it has been applied to experimental HRTEM images of an aluminium crystal. Moreover, it has been shown that if the model is inaccurate, systematic errors determine fundamental limits to statistical model-based resolution as well. Apart from these fundamental limits, computational limits exist, just as for deterministic

model-based resolution. It has also been shown how to improve statistical model-based resolution using statistical experimental design.

Finally, ultimate model-based resolution is considered. Depending on the observations, the solutions of the position estimates may coincide exactly. It is shown that ultimately, resolution is limited by errors, both systematic and nonsystematic. This limit is fundamental: no increase in computer power or improved software can remove it. It would be present even if the structure is resolved and sufficient computer power would be available for fitting them. Removing these limits requires more and preferably different observations of the same structure.

Acknowledgments. Dr. S. Van Aert gratefully acknowledges the financial support of the Fund for Scientific Research—Flanders (FWO). The research of Dr. A.J. den Dekker has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

References

- Batson, P.E. (1999). Advanced spatially resolved EELS in the STEM. *Ultramicroscopy* **78**, 33–42.
- Batson, P.E., Dellby, N. and Krivanek, O.L. (2002). Sub-ångstrom resolution using aberration corrected electron optics. *Nature* **418**, 617–620.
- Bettens, E., Van Dyck, D., den Dekker, A.J., Sijbers, J. and van den Bos, A. (1999). Model-based two-object resolution from observations having counting statistics. *Ultramicroscopy* **77**, 37–48.
- Biraud, Y. (1969). A new approach for increasing the resolving power by data processing. *Astron. Astrophys.* **1**, 124–127.
- Born, M. and Wolf, E. (1999). *Principles of Optics*, 7th (expanded) ed. (Cambridge University Press, Cambridge).
- Broeckx, J., Op de Beeck, M. and Van Dyck, D. (1995). A useful approximation of the exit wave function in coherent STEM. *Ultramicroscopy* **60**, 71–80.
- Browning, N.D., Arslan, I., Moeck, P. and Topuria, T. (2001). Atomic resolution scanning transmission electron microscopy. *Phys. Stat. Sol. B* **227**, 229–245.
- Buseck, P.R., Cowley, J.M. and Eyring, L. (1988). *High-Resolution Transmission Electron Microscopy and Associated Techniques*. (Oxford University Press, Oxford).
- Castleman, K.R. (1979). *Digital Image Processing*. (Prentice-Hall International, London).
- Coene, W.M.J., Thust, A., Op de Beeck, M. and Van Dyck, D. (1996). Maximum-likelihood method for focus-variation image reconstruction in high resolution transmission electron microscopy. *Ultramicroscopy* **64**, 109–135.
- Cowley, J.M. (2001). The quest for ultra-high resolution. In *Progress in Transmission Electron Microscopy—Concepts and Techniques* (X.F. Zhang and Z. Zhang, Eds.), Vol. 1, 35–79, Berlin (Springer, 1).
- de Jong, A.F. and Van Dyck, D. (1993). Ultimate resolution and information in electron microscopy II. The information limit of transmission electron microscopes. *Ultramicroscopy* **49**, 66–80.
- den Dekker, A.J. and Van Aert, S. (2002). Quantitative high resolution electron microscopy and Fisher information. In *Proceedings of the 15th International Congress on Electron Microscopy, Interdisciplinary and Technical Forum Abstracts*

- 2002 in Durban, South Africa (R. Cross, Ed.), Vol. 3, 185–186 (Microscopy Society of Southern Africa, Onderstepoort).
- den Dekker, A.J. and van den Bos, A. (1997). Resolution: A survey. *J. Opt. Soc. Amer. A* **14**, 547–557.
- den Dekker, A.J., Sijbers, J. and Van Dyck, D. (1999). How to optimize the design of a quantitative HREM experiment so as to attain the highest precision. *J. Microsc.* **194**, 95–104.
- den Dekker, A.J., Van Aert, S., Van Dyck, D., van den Bos, A. and Geuens, P. (2001). Does a monochromator improve the precision in quantitative HR-TEM? *Ultramicroscopy* **89**, 275–290.
- den Dekker, A.J., Van Aert, S., van den Bos, A. and Van Dyck, D. (2005). Maximum likelihood estimation of structure parameters from high resolution electron microscopy images. Part I: a theoretical framework. *Ultramicroscopy*, **104**, 83–106.
- Falconi, O. (1964). Maximum sensitivities of optical direction and twist measuring instruments. *J. Opt. Soc. Amer.* **54**, 1315–1320.
- Falconi, O. (1967). Limits to which double lines, double stars, and disks can be resolved and measured. *J. Opt. Soc. Amer.* **57**, 987–993.
- Fedorov, V.V. (1972). *Theory of Optimal Experiments*. (Academic Press, New York).
- Frank, J. (1992). *Electron Tomography—Three-Dimensional Imaging with the Transmission Electron Microscope*. (Plenum Press, New York).
- Frieden, B.R. (1967). Band-unlimited reconstruction of optical objects and spectra. *J. Opt. Soc. Amer.* **57**, 1013–1019.
- Frieden, B.R. (1975). *Image Enhancement and Restoration* Vol. 6. (Springer-Verlag, New York).
- Frieden, B.R. (1980). Statistical models for the image restoration problem. *Comp. Graph. Image. Proc.* **12**, 40–59.
- Frieden, B.R. (1998). *Physics from Fisher Information—A Unification*. (Cambridge University Press, Cambridge).
- Gerchberg, R.W. (1974). Super-resolution through error energy reduction. *Opt. Acta.* **21**, 709–720.
- Gerchberg, R.W. (1989). Superresolution through error function extrapolation. *IEEE Transact. Acoustics, Speech, Sign Proc.* **37**, 1603–1606.
- Geuens, P. and Van Dyck, D. (2002). The S-state model: A work horse for HRTEM. *Ultramicroscopy* **93**, 179–198.
- Gonzalez, R.G. and Woods, R.E. (2002). *Digital Image Processing*, 2nd Ed. (Prentice Hall, Upper Saddle River, NJ).
- Goodman, J.W. (1968). *Introduction to Fourier Optics*. (McGraw-Hill, San Francisco).
- Haider, M., Uhlemann, S., Schwan, E., Rose, H., Kabius, B. and Urban, K. (1998). Electron microscopy image enhanced. *Nature* **392**, 768–769.
- Harris, J.L. (1964). Diffraction and resolving power. *J. Opt. Soc. Amer.* **54**, 931–936.
- Hartel, P., Rose, H. and Dinges, C. (1996). Conditions and reasons for incoherent imaging in STEM. *Ultramicroscopy* **63**, 93–114.
- Howie, A. (1966). Diffraction channelling of fast electrons and positrons in crystals. *Philos. Mag.* **14**, 223–237.
- Hunt, B.R. (1994). Prospects for image restoration. *Internat. J. Mod. Phys. C* **5**, 151–178.
- Hunt, B.R. and Sementilli, P. (1992). Description of a poisson imagery super resolution algorithm. In *Astronomical Data Analysis Software and Systems* (Worrall et al., Ed.), Vol. 25, 196–199 (Astronomical Society of the Pacific, San Francisco).

- Janson, P.A., Hunt, R.H. and Plyler, E.K. (1970). Resolution enhancement of spectra. *J. Opt. Soc. Amer. A* **60**, 596.
- Kisielowski, C., Hetherington, C.J.D., Wang, Y.C., Kilaas, R., O'Keefe, M.A. and Thust, A. (2001a). Imaging columns of the light elements carbon, nitrogen and oxygen with sub Ångstrom resolution. *Ultramicroscopy* **89**, 243–263.
- Kisielowski, C., Principe, E., Freitag, B. and Hubert, D. (2001b). Benefits of microscopy with super resolution. *Physica. B* **308–310**, 1090–1096.
- Koster, A.J., van den Bos, A. and van der Mast, K.D. (1987). An autofocus method for a TEM. *Ultramicroscopy* **21**, 209–222.
- Krivanek, O.L., Dellby, N. and Nellist, P.D. (2002). Aberration correction in the STEM. In *Proceedings of the 15th International Congress on Electron Microscopy, Interdisciplinary and Technical Forum Abstracts 2002 in Durban, South Africa* (R. Cross, Ed.), Vol. 3, 29–30 (Microscopy Society of Southern Africa, Onderstepoort).
- Meinel, E.S. (1986). Origins of linear and nonlinear recursive restoration algorithms. *J. Opt. Soc. Amer. A* **3**, 787–799.
- Miedema, M.A.O., van den Bos, A. and Buist, A.H. (1994). Experimental design of exit wave reconstruction from a transmission electron microscope defocus series. *IEEE Transact. Instr. Meas.* **43**, 181–186.
- Mood, A.M., Graybill, F.A. and Boes, D.C. (1974). *Introduction to the Theory of Statistics*, 3rd ed. (McGraw-Hill, Singapore).
- Mook, H.W. and Kruit, P. (1999). Optics and design of the fringe field monochromator for a Schottky field emission gun. *Nucl. Instr. Methods. Phys. Res. A* **427**, 109–120.
- Muller, D.A. (1998). Core level shifts and grain boundary cohesion. In *Proceedings Microscopy and Microanalysis 1998 in Atlanta, Georgia* (G.W. Bailey, Ed.), Vol. 4, 766–767. (Springer, New York).
- Muller, D.A. (1999). Why changes in bond lengths and cohesion lead to core-level shifts in metals, and consequences for the spatial difference method. *Ultramicroscopy* **78**, 163–174.
- Nellist, P.D. and Pennycook, S.J. (1998). Subångstrom resolution by underfocused incoherent transmission electron microscopy. *Phys. Rev. Lett.* **81**, 4156–4159.
- Nellist, P.D. and Pennycook, S.J. (2000). The principles and interpretation of annular dark-field Z-contrast imaging. In *Advances in Imaging and Electron Physics* (P.W. Hawkes, Ed.), Vol. 113, 147–199 (Academic Press, San Diego).
- O'Keefe, M.A., Hetherington, C.J.D., Wang, Y.C., Nelson, E.C., Turner, J.H., Kisielowski, C., Malm, J.-O., Mueller, R., Ringnald, J., Pan, M. and Thust, A. (2001). Sub-Ångstrom high-resolution transmission electron microscopy at 300 keV. *Ultramicroscopy* **89**, 215–241.
- Olson, G.B. (1997). Computational design of hierarchically structured materials. *Science* **277**, 1237–1242.
- Olson, G.B. (2000). Designing a new material world. *Science* **288**, 993–998.
- Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes*. (McGraw-Hill, New York).
- Pázman, A. (1986). *Foundations of Optimum Experimental Design*. (D. Reidel Publishing Company, Dordrecht).
- Pennycook, S.J. and Jesson, D.E. (1991). High-resolution Z-contrast imaging of crystals. *Ultramicroscopy* **37**, 14–38.
- Pennycook, S.J. and Jesson, D.E. (1992). Atomic resolution Z-contrast imaging of interfaces. *Acta. Met. Mat. Supplement* **40**, 149–159.
- Pennycook, S.J., Jesson, D.E., Chisholm, M.F., Browning, N.D., McGibbon, A.J. and McGibbon, M.M. (1995). Z-contrast imaging in the scanning transmission electron microscope. *J. Microsc. Soc. Amer.* **1**, 231–251.

- Reed, M.A. and Tour, J.M. (2000). Computing with molecules. *Sci. Amer.* **282**, 68–75.
- Rose, H. (1990). Outline of a spherically corrected semiaplanatic medium-voltage transmission electron microscope. *Optik*. **85**(1), 19–24.
- Saxton, W.O. (1978). *Computer Techniques for Image Processing in Electron Microscopy*, 236–248 (Academic Press, New York).
- Schell, A.C. (1965). Enhancing the angular resolution of incoherent sources. *The Radio and Electronic Engineer* **29**, 21–26.
- Scherzer, O. (1949). The theoretical resolution limit of the electron microscope. *J. App. Phy.* **20**, 20–28.
- Schiske, P. (1973). Image processing using additional statistical information about the object. In *Image Processing and Computer-aided Design in Electron Optics* (P.W. Hawkes, Ed.), 82–90 (Academic Press, London).
- Sparrow, C.M. (1916). On spectroscopic resolving power. *Astrophys. J.* **44**, 76–86.
- Spence, J.C.H. (1999). The future of atomic resolution electron microscopy for materials science. *Mat. Sci. Eng. R* **26**, 1–49.
- Strutt, J.W. (1899). Investigations in optics, with special reference to the spectroscope. In *Scientific papers by John William Strutt, Baron Rayleigh*, Vol. 1, 415–459 (Cambridge University Press, Cambridge).
- Thust, A., Coene, W.M.J., Op de Beeck, M. and Van Dyck, D. (1996). Focal-series reconstruction in HRTEM: Simulation studies of non-periodic objects. *Ultramicroscopy* **64**, 211–230.
- Treacy, M.M.J. (1982). Optimising atomic number contrast in annular dark field images of thin films in the scanning transmission electron microscope. *J. Microsc. Spectrosc. Elect.* **7**, 511–523.
- Van Aert, S. and Van Dyck, D. (2001). Do smaller probes in a scanning transmission electron microscope result in more precise measurement of the distances between atom columns? *Philos. Mag. B* **81**, 1833–1846.
- Van Aert, S., den Dekker, A.J., Van Dyck, D. and van den Bos, A. (2002a). High-resolution electron microscopy and electron tomography: Resolution versus precision. *J. Struct. Biol.* **138**, 21–33.
- Van Aert, S., den Dekker, A.J., Van Dyck, D. and van den Bos, A. (2002b). Optimal experimental design of STEM measurement of atom column positions. *Ultramicroscopy* **90**, 273–289.
- Van Aert, S., den Dekker, A.J., van den Bos, A. and Van Dyck, D. (2002c). High-resolution electron microscopy: From imaging toward measuring. *IEEE Transact. Instr. Meas.* **51**, 611–615.
- Van Aert, S., den Dekker, A.J., van den Bos, A. and Van Dyck, D. (2004). Statistical experimental design for quantitative atomic resolution transmission electron microscopy. In *Advances in Imaging and Electron Physics* (P.W. Hawkes, Ed.), Vol. 130, 1–164 (Academic Press, San Diego).
- Van Aert, S., den Dekker, A.J., van den Bos, A., Van Dyck, D. and Chen, J.H. (2005). Maximum likelihood estimation of structure parameters from high resolution electron microscopy images. Part II: A practical example. *Ultramicroscopy*, **104**, 107–125.
- van den Bos, A. (1982). Parameter estimation. In *Handbook of Measurement Science* (P.H. Sydenham, Ed.), Vol. 1, 331–337 (Wiley, Chichester).
- van den Bos, A. (1992). Ultimate resolution—A mathematical framework. *Ultramicroscopy* **47**, 298–306.
- van den Bos, A. (1999). Measurement errors. In *Encyclopedia of Electrical and Electronics Engineering* (J.G. Webster, Ed.), 448–459 (Wiley, New York).
- van den Bos, A. and den Dekker, A.J. (1995). Ultimate resolution in the presence of coherence. *Ultramicroscopy* **60**, 345–348.

- van den Bos, A. and den Dekker, A.J. (2001). Resolution reconsidered—Conventional approaches and an alternative. In *Advances in Imaging and Electron Physics*, (P.W. Hawkes, Ed.), Vol, **117**, 241–360 (Academic Press, San Diego).
- Van Dyck, D. and Chen, J.H. (1999). A simple theory for dynamical electron diffraction in crystals. *Sol. State Comm.* **109**, 501–505.
- Van Dyck, D. and Coene, W. (1987). A new procedure for wave function restoration in high resolution electron microscopy. *Optik* **77**(3), 125–128.
- Van Dyck, D. and Op de Beeck, M. (1996). A simple intuitive theory for electron diffraction. *Ultramicroscopy* **64**, 99–107.
- Van Dyck, D., Danckaert, J., Coene, W., Selderslaghs, E., Broddin, D., Van Landuyt, J. and Amelinckx, S. (1989). The atom column approximation in dynamical electron diffraction calculations. In *Computer Simulation of Electron Microscope Diffraction and Images*. (W. Krakow and M. O’Keefe, Eds.), 107–134. (The Minerals, Metals & Materials Society, Warrendale).
- Van Dyck, D., Op de Beeck, M. and Coene, W. (1993). A new approach to object wave function reconstruction in electron microscopy. *Optik* **93**(3), 103–107.
- Van Dyck, D., Van Aert, S., den Dekker, A.J. and van den Bos, A. (2003). Is atomic resolution transmission electron microscopy able to resolve and refine amorphous structures? *Ultramicroscopy* **98**, 27–42.
- Verbeeck, J. and Van Aert, S. (2004). Model based quantification of EELS spectra. *Ultramicroscopy* **101**, 207–224.
- Wada, Y. (1996). Atom electronics: A proposal of nano-scale devices based on atom/molecule switching. *Microelectr. Eng.* **30**, 375–382.
- Walsh, D.O. and Nielsen-Delaney, P.A. (1994). Direct method for superresolution. *J. Opt. Soc. Amer. A* **11**, 572–579.
- Wang, Z.L. (2001). Inelastic scattering in electron microscopy—Effects, spectrometry and imaging. In *Progress in Transmission Electron Microscopy 1—Concepts and Techniques* (X.-F. Zhang and Z. Zhang, Eds.), 113–159. (Springer-Verlag, Berlin).

Index

- Aberration
 - coefficients, of tetrode mirror, 726
 - corrected microscope, using electron mirror, 674–675
 - corrected PEEM microscope
 - beam separator in, 675–677
 - resolution/transmission of, 677–678
 - correction, 696–731
 - introduction to, 696, 706–712
 - of PEEM microscopes, 671–677
 - STEM and, 123–125
 - by tetrode electron mirror, 676
 - correctors
 - foil correctors and, 716–718
 - HRTEM and, 39–42
 - quadrupole–octopole correctors and, 708–711
 - sextupole correctors and, 715–716, 715–718
 - geometric, 697–706
 - phase shift, 71
 - resolution and, on uncorrected PEEM, 673
 - size, zone plate image quality and, 849
 - types of, 697–706
- Abrikosov flux lattice, LT-STM and, 1130
- Absorption, XPEEM and, 659–661
- Absorption length dependence, on elevation angle/foil thickness, 336
- Absorption limit, thickness criteria for, 337
- Absorption path, for two sample configurations, 338
- Acceleration voltage/emission, for Cryo-ET, 560
- Acceptance angle, resolution limit and, 621
- Acridine orange, 178
- Actin filaments, AFM image of, 1032
- ADC. *See* Analog-to-digital-converter
- ADF detector. *See* Annular dark-field detector
- AE. *See* Auger electrons
- AEEM. *See* Auger electron emission microscopy
- AEM. *See* Analytical electron microscopy
- AFM. *See* Atomic force microscope
- Ag(111)
 - LDOS oscillations on, 1089
 - surface states on, 1086
- Ag atom triangle, images of, 1101
- Ag islands, electron confinement, 1092
- Al alloy, energy-filtered image of, 354
- Al films, on Si(111), 636
- Al₂O₃ matrix, images of, 823
- Alignment, Cryo-ET and, 576–577
- All-electrostatic correctors, chromatic aberration correctors and, 718–719
- All-purpose correctors, 726–728
- Aluminum crystal, HRTEM of, 1249, 1250, 1251
- Aluminum oxide, high-resolution micrographs of, 221
- AM-AFM. *See* Amplitude-modulated atomic force microscope
- Amino acids, near-carbon-edge absorption spectra of, 890
- Amorphization, phase transformations and, 448–449
- Amorphous structure
 - with overlapping projected atoms, 1238
 - with separable projected atoms, 1238
- Amorphous to crystalline transformation, in phase change materials, 450
- Amplitude-modulated atomic force microscope (AM-AFM)
 - NC-AFM and, 934–935
 - topographic image with, 935
- Amplitude/phase contrast images, of germanium test pattern, 873
- Analog-to-digital-converter (ADC), 158
- Analytical electron microscopy (AEM), 273–399
 - EDXS and, 273
 - EELS and, 273
 - electron optical configuration of, 290–293
 - electron sources/probes for, 282–290
 - fundamentals of, 310–331
 - instrumentation for, 282–300
 - sample holders for, 378
 - sample preparation requirements for, 308
 - source comparisons for, 285
 - TEM and, 273
- A_nB_nO_{3n+2} compounds, structural models of, 50

I2 Index

- Angle/energy distributions, of SIMION simulations, of emitted electrons, 669
- Angle-resolved photoemission spectroscopy (ARPES), 1082
- Angular dependence, of partial cross sections, 324
- Angular distribution, of 50eV electrons, 607
- Angular/energy dependence, of the GOS, 321–323
- Annealed surface, cooled surface and, 632
- Annular dark-field (ADF) detector, 68
 - images of thicker samples, 90–94
 - interference detection, 86
- Annular dark-field (ADF) detector scanning
 - transmission microscope image, of Si<112>, 125
- Annular dark-field (ADF) imaging, 82–97
 - availability of, 83–84
 - early analyses of, 91
 - of GaAs<110>, 95
 - introduction of, 83
 - of NiS₂/Si(001)<110> interface, 95
 - of Pt atoms, 96
 - structure determination with, 94–95
- Annular dark-field (ADF) micrographs, of SE, 176
- Antiferromagnetic domains
 - imaging of, 681–685
 - X-ray absorption spectra of, 682
- Aperture diameter, 81
- Aperture function, 71
- Aplanatic corrector
 - of chromatic aberrations, 720
 - of spherical/chromatic aberration, 720
- Applications
 - of DTEM, 433–435
 - of HRSEM, 214–221
 - of LEEM, 630–642
 - of low-loss spectroscopy, 394–399
 - of LT-STM, 1081–1132
 - of LVSEM, 226–237
 - of NC-AFM, 939–944
 - of SEM, 203
 - in life science, 203
 - of STM, 975–976
 - of VLVSEM, 226–237
 - of X-ray microscopy, 892–906
 - in biology, 892–895
 - in environmental science, 896–897
 - in magnetic materials, 903–905
 - in materials science, 897–903
 - of zone plate X-ray microscopes, 836–906
- ARPES. *See* Angle-resolved photoemission spectroscopy
- As-cast niobium specimen, EBSD patterns from, 255
- ASEM. *See* Atmospheric scanning electron microscope
- Astigmatism/field curvature, zone plate image quality and, 849–850
- Atmosphere interaction, electron beam-induced phenomena and, 511
- Atmospheric scanning electron microscope (ASEM), 237
- Atmospheric thin windows (ATW), 295
- Atomic force microscope (AFM), 217, 929, 1048
 - actin filaments image with, 1032
 - bacterial cell wall protein image with, 1061–1062
 - beyond imaging with, 1059–1064
 - of block face surface image, 1029
 - cells and, 1057–1058
 - cellular samples and, 1058–1064
 - chemical bonding and, 1055–1056
 - collagen matrices and, 1061
 - contact mode of, 930, 1046
 - design optimization, for life sciences, 1048–1052
 - elements of, 1030
 - feedback loop of, 1040–1045
 - elements of, 1041–1042
 - probe/sample and, 1042
 - schematic drawing of, 1041
 - stability of, 1040–1041
 - fixed 3T3 cell image with, 1049
 - frequency response of, closed feedback loop, 1045
 - imaging modes of, 1046–1048
 - instrument noise in, 1045–1046
 - instrumentation for, 1030–1052
 - intermittent contact mode of, 930, 1047–1048
 - Langmuir-Blodgett films and, 1056–1057
 - in life sciences, 1025–1064
 - macromolecular samples for, 1053–1057
 - physisorption/DLVO force for, 1054
 - specimen supports for, 1053–1054
 - natural frequencies for, scanner elements, 1043
 - noncontact, 930
 - optical microscopy and, 1049–1050
 - physisorption/DLVO force and, 1054
 - physisorption/hydrophobic/hydrophilic interaction and, 1055
 - pulmonary surfactant and, 1060–1061
 - sample deformation and, 1034–1037
 - sample environment for, 1051–1052
 - sample preparation for, 1052–1058
 - scanners for, 1050–1051
 - sphingolipid-cholesterol rafts, 1062–1064
 - stylus geometry and, 1033
 - tip geometries and, 1032
 - tip-sample interaction and, 1033–1040
- Atomic number
 - total inelastic cross sections and, 318
 - X-ray energy and, 276
- Atomic resolution, 69
 - element-specific imaging with, 996
 - imaging, with force interactions, 930–944
 - magnetic imaging, with SP-STM, 1127
 - NC-AFM/STM/KFM, on Ge/Si (105), 959
- Atomic resolution transmission electron microscopy, 3–58
- Atomic scattering, interference features from, 87
- Atom/molecule manipulation, STM and, 1003–1006
- ATP synthase, 1028
- Attenuation coefficients μ, ν , energy dependence of, 610
- ATW. *See* Atmospheric thin windows

- Au, energy dependence of, inelastic mean free electron path of, 610
- Au atoms, charging of, 1006
- Au nanoparticles, ronchigrams of, 78
- Au on Si deposition, 465
- Au sample, electron trajectories simulation of, 349
- Auger electron emission microscopy (AEEM), 606
- Auger electrons (AE), 109, 149
 - SEM and, 181
 - yield, 179
- Auger microscopy, 181
- Austenite
 - orientation map of, 257
 - secondary electron micrograph of, 256
- Automated tomography, experimental setup for, 571
- Automation, Cryo-ET and, 571–576
- Axial diffractograms images, from HRTEM, 20
- Axial resolution improvement, by aperture enlargement, 795–811

- Ba M_{45} edges, Nb M_{45} edges and, 330
- Background
 - in EDXS, 330–331
 - in EELS, 330–331
- Backscattered electrons (BSE), 166, 172–175
 - detector, ETD and, 187
 - energy dependence of, 608
 - of fern fossil, 205
 - micrographs
 - of frozen yeast cells, 218
 - of steel ball, 155, 185, 190
 - of tartar, 186
 - yield
 - normalized, 173
 - SE yield and, 171
- Backscattering contrast, on Co squares, 624
- Bacterial cell wall protein image, with AFM, 1061–1062
- Bacteriorhodopsin, force dependent surface tomography of, 1036
- Ball-and-stick model, of Si(111), 977
- Ballistic electron, band diagrams for, 1008
- Ballistic electron emission microscopy (BEEM), 1007–1012
 - constant-current STM images and, 1010
 - on polycrystalline Au/*n*-Si(001), 1009
 - schematic setup of, 1008
 - on self-assembled quantum dots, 1011
- Band diagrams
 - for ballistic electron, 1008
 - for elastic/inelastic electron tunneling, 973
 - for vacuum tunneling, 971
- Be(0001), Fermi contour imaging of, 1105
- Be(1010), Fermi contour imaging of, 1105
- Beam current, v. incoherent imaging, 119
- Beam damage, Cryo-ET and, 554–557
- Beam interaction volumes, schematic diagram of, 108
- Beam scanner, 133
- Beam separator
 - in aberration corrected PEEM microscope, 675–677
 - for LEEM, 622
- Beam-induced phase transitions, 513–514
- Beam-vapor interaction, nanostructures and, 512
- BEEM. *See* Ballistic electron emission microscopy
- Bethe stopping power, mean ionization potential and, for carbon/protein, 206
- Bethe's theory, 320
- BF detector. *See* Bright-field detector
- $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$, Fourier transform spectroscopy on, 991
- Bias-dependent evolution, of LDOS/ILDOS oscillations, surface states, 1087
- Bias-dependent imaging, STM and, 984–986
- Bias-dependent STM, on GaAs(110), 986
- Bias-dependent tunneling, band diagrams of, 984
- Biological electron microscopy, 536
- Biological X-ray microscopy, 892–895
- Birch pollen, secondary electron micrograph of, 217
- Bloch wave approach, 56
- Bloch wave function, 94
- Block face surface, AFM image of, 1029
- Block oxide structure, oxidation of, 470
- Bond formation, with STM tip, 1113
- Bragg's law, 254
- Bright-field (BF) detector, 68
- Bright-field (BF) image, of Co nanowires, 1168
- Bright-field (BF) scanning transmission electron microscope (STEM)
 - crystalline sample schematic diagram of, 79
 - large detector incoherent, 82
 - lattice imaging in, 79–81
 - phase contrast imaging in, 81–82
- Brightness sufficiency, for STEM, 114–115
- BSE. *See* Backscattered electrons
- Bulk materials, deformation in, 488
- Buried interfaces, heterostructures and, 1007–1016

- $\text{C}_2\text{H}_2/\text{C}_2\text{D}_2$, chemical sensitive imaging of, 1118
- C_{60}/C -nanotube, $I(V)$ tunneling spectroscopy, 989–993
- Ca, EELS of, 374
- CaF_2
 - high-resolution constant height images of, 944
 - NC-AFM of, 943–944
- Carbon K-edge, specific peaks in, 106
- Carbon nanostructures, growth of, 469–472
 - in situ*, 471
- Carbon nanotubes (CNTs), 498
 - mechanical properties of, 498
 - molecular wavefunctions in, 1095
- Carbon/gold, frequency-dependent oscillator strength, 837
- Carbon/protein, Bethe stopping power and, mean ionization potential of, 206
- Catalysis, *in situ* transmission electron microscopy and, 466–467
- Catalysts, *in situ* imaging of, 468

14 Index

- Catalytic flow reactor, STM, simultaneous mass spectrometry and, 1001
- Cathode lenses configurations, for LEEM, 620
- Cathode ray tube (CRT), 139
- Cathodoluminescence (CL), 109, 149, 1120
detectors, 151–152
SEM and, 177
- CBED. *See* Convergent beam electron diffraction
- CC. *See* Charge collection
- CCC, histogram of, 586
- CCD. *See* Charge-coupled device
- CE. *See* Collection efficiency
- Cells
absorbing, on glass coverslips, 1057
AFM and, 1057–1058
immobilizing nonadhering, 1058
- Cellular cryoelectron tomography, of magnetotactic microorganism, 543
- Cellular samples
imaging of, 1058–1064
locally probing macromolecular and, 1058–1064
- Cerenkov effect, 312
- CFEG. *See* Cold field emission gun
- Charge collection (CC), 192
- Charge-coupled device (CCD), 36, 37, 99, 412, 535, 755
array, 73
cameras, MTFs and, 40
chip, 37
detector, schematic diagram of, 38
- Charged particle optics (CPO), 665
- Chemical bonding, AFM and, 1055–1056
- Chemical contrast
with functionalized tip, 1109
in XPEEM, 663
- Chemical imaging, of manganese nodule, 663
- Chemical maps, of water microbial colony, 897
- Chemical sensitive imaging, of C_2H_2/C_2D_2 , 1118
- Chemical vapor deposition (CVD), 978
- Chromatic aberrations, 146, 704–705
aplanatic corrector of, 720
correctors, 718–726
all-electrostatic correctors and, 718–719
mixed quadrupole correctors and, 719–723
of electrostatic correction, 722–723
STEM and, 118–119
- Chromatic correction, energy electron distribution and, 709
- Chromatic defocus spreads, probe profile plots with, 122
- Circular quantum corral, of Fe atoms, 1091
- CITS. *See* Current-imaging tunneling spectroscopy
- CL. *See* Cathodoluminescence
- Cluster analysis, of lutetium, 898
- CNTs. *See* Carbon nanotubes
- Co, XMCD of, 679
- Co dots, SEM of, 1166
- Co nanoparticle, self-assembled, low magnification
bright-field image, 1159
- Co nanoparticle chains, in electron holography, 1153–1156
- Co nanoparticle rings, electron holography and, 1157–1160
- Co nanowires, electron holography and, 1167–1168
- Co particles, off-axis electron holography of, 1155
- Co squares, backscattering contrast on, 624
($Co_{83}Cr_{17}$)₈₇Pt₁₃ alloy, XMCD of, 904
- Coherence, 74
HRTEM and, 26–29
requirements, for CXDI, 1221–1223
resolution and, 883–884
types of, 26–27
in Zernike phase contrast, 884
- Coherent convergent beam electron diffraction
ronchigrams and, 73–78
Si<110> pattern of, 74
- Coherent X-ray diffractive imaging (CXDI), 1196
coherence requirements for, 1221–1223
computer processing demands for, 1223–1224
resolution, 1221–1223
- Co/LaFeO₃, magnetization reversal of, 684
- Cold field emission, schematic diagram of, 116
- Cold field emission gun (CFEG), 68, 209
schematic diagram of, 116
for STEM, 11–115
- Cole-Cole plot, of local impedance spectra, 953
- Collagen fibrils, adsorbed on mice, 1059
- Collagen matrices, AFM and, 1061
- Collagen microfibrils, 1062
- Collection efficiency (CE), 149
- Colloidal chemistry sample, images of, 896
- Color film, FIB of, secondary electron micrograph of, 200
- Column design, electron optics and, 109–113
- Coma, zone plate image quality and, 850
- Commercial cross-grating replica, SE micrographs of, 163
- Complex oxides, exit-wave reconstruction of, 48–51
- Compositionally sensitive imaging, examples of, 96–97
- Condenser
OL and, 110
zone plates, 856–858
- Condenser aperture, 80, 113
- Condenser-objective lens, schematic drawing of, 113
- Cone formation, from diffraction, scheme of, 253
- Confocal principles
laser scanning microscopy and, 755–757
TPE of, fluorescent molecules and, 754–763
- Confocal PSFs, FWHM of, 759
- Confocal setup
equivalent optical configuration for, 757
historical sketch of, 753
- Confocal two-photon excitation, illustrated modalities of, 769
- Co/NiO, magnetic domain imaging of, 683
- Constant-current STM images
and BEEM, 1010
of crystal BSCCO, 1131
of Si(001), 983
- Contact atomic force microscope (AFM), 930

- Contact mode, of AFM, 1046
- Contamination
 - beam-induced, 202
 - SEM and, 201–203
- Continuous wave (CW), 407, 408
- Continuous-flow cryostat, variable-temperature STM with, 1077
- Contour plot, of planar zone plates, 853
- Contrast, in TXM, 885
- Contrast formation
 - FESEM and, 212–214
 - for LVSEM, 225–226
 - SEM and, 182–195
 - for VLVSEM, 225–226
- Contrast mechanisms, with LEEM, 624–630
- Contrast transfer, partial coherence and, 880–882
- Contrast transfer functions (CTFs), phase/amplitude, 425
- Conventional excitation, v. two-photon excitation, 756
- Conventional scanning electron microscope (CSEM), 139–208
 - schematic drawing of, 140
- Conventional transmission electron microscope (CTEM), 65, 535
 - incoherent imaging and, 84–90
- Conventional transmission electron microscope/Scanning transmission electron microscope instruments, 112–113
- Convergence correction, EELS and, 344–346
- Convergent beam electron diffraction (CBED), 73, 517
- Cooling system, for Cryo-ET, 563
- Copper, standing wave patterns of, 1083
- Copper braid, mechanical decoupling of, 1075
- Copper deposition, using electrochemical cell, 512
- Copper oxidation, 469
- Core electron, excitation on, 660
- Core-loss edges, 101
- Correction
 - factors, for spectra quantification, 345
 - techniques, 339–341
 - Wien filters and, 723–724
- Cosine amplitude grating imaging, with STXM, 881
- Co(0001) surface, SPLEEM image of, 644
- CPO. *See* Charged particle optics
- Cramér-Rao lower bound (CRLB)
 - distance measurement with, 1246–1248
 - position measurement with, 1245–1246
 - statistical model-based resolution and, 1244–1248
- Cr-Fe-alloy, X-ray microanalysis of, 250, 251
- CRLB. *See* Cramér-Rao lower bound
- Cross correlation function (XCF), PCF and, 23
- Cross sections
 - calculations of, 321
 - comparison of, of scattering processes, 325
 - of TDS, 92
- Cross-fractured semiconductor structure, secondary electron micrograph of, 227
- Cross-sectional scanning tunneling microscopy (XSTM), 1013, 1014
 - dopant imaging by, 1015
 - imaging buried heterostructures and, 1013–1016
 - on self-assembled quantum dots, 1014
- Cross-sectional semiconductor device specimen, experimental results from, 1180
- Cross-sectioned diode, surface topography of, 951
- CRT. *See* Cathode ray tube
- Cryoelectron tomography (Cryo-ET), 535–592, 543–547
 - acceleration voltage/emission for, 560
 - alignment and, 576–577
 - automation and, 571–576
 - beam damage and, 554–557
 - cooling system for, 563
 - cryopreparation, 549–552
 - cryosectioning, 552–554
 - detector for, 563–566
 - dual-axis tilting for, 559–560
 - electron optics for, 561
 - energy filter for, 566–570
 - goniometer tilt stage for, 562–563
 - instrumentation/automation for, 560–576
 - introduction to, 535–538
 - macromolecular complexes, identification strategies of, 584
 - major difficulties in, 547–589
 - plunge freezing instrumentation and, 550
 - plunge freezing process steps and, 551
 - radiation sensitivity and, 554–557
 - reconstruction and, 577–582
 - sample preparation with, 547–549
 - side entry cryo-holder for, 562
 - with single particle approach, 547
 - single-axis tilting for, 557
 - strategies/developments in, 589–592
 - tilting geometry with, 557–560
 - visualization/image analysis and, 582–584
- Cryo-ET. *See* Cryoelectron tomography
- Cryosectioning, Cryo-ET and, 552–554
- Crystal BSCCO, constant-current STM images of, 1131
- Crystal orientation contrast, SEM and, 193
- Crystal strain, lattice displacements in, 93
- Crystal structure analysis, 141
 - by EBSD, 252–257
- Crystalline materials, ronchigrams of, 74–76
- Crystallization
 - metal-induced, 449
 - on patterned substrates, 474, 475, 476
 - phase transformations and, 448–449
 - surface reactions and, 463–476
- Crystallographic characterization, of polycrystalline materials, 134
- CSEM. *See* Conventional scanning electron microscope
- CTEM. *See* Conventional transmission electron microscope
- CTFs. *See* Contrast transfer functions
- Cu Lu₂₃ edge, 329

I6 Index

- Cu-DTBPP, molecular conformation identification of, 1019
- Cu-Nb multifilamentary composite, deformation in, 491
- Current-imaging tunneling spectroscopy (CITS), 988
 - local I-V measurements and, 986–989
 - on Si(111), 989
- Cu(111) surface column, Lander molecule on, 1111
- CVD. *See* Chemical vapor deposition
- CXDI. *See* Coherent X-ray diffractive imaging
- Dark-field imaging, phase identification by, 626
- Data cube, 104
 - of spectrum image, 104
 - three-dimensional, 105
- DECO. *See* Doubly symmetric electrostatic corrector
- Deconvolution, periodic artifacts removal through, 803–807
- Defocus difference, PCF and, 24
- Deformation
 - in bulk materials, 488
 - controlled atmosphere effect on, 493
 - in Cu-Nb multifilamentary composite, 491
 - microscopic phenomenon during, 487–493
 - of multiphase/composite/layered materials, 490–493
 - in polycrystalline materials, 487–490
 - in single crystals, 487–490
- Density problem sampling, in Fourier space, illustration of, 581
- Depth-of-focus limit
 - high-energy X-rays and, in x-ray microscopes, 887–888
 - lens-free imaging and, in x-ray microscopes, 888
 - through-focus deconvolution and, in x-ray microscopes, 887
 - wide-band illumination and, in x-ray microscopes, 886–887
 - in X-ray microscope tomography, 885–886
- Detection geometries, FESEM and, 211
- Detection limits
 - for EDXS, 375–377
 - EDXS v. EELS, 381–384
 - for EELS, 379–381
 - in microanalysis, 374–384
- Detector(s)
 - characterization of, 38–39
 - for Cryo-ET, 563–566
 - EDXS, 293
 - FESEM and, 211
 - HRTEM and, 36–39
- Detector efficiency, detector windows and, 296
- Detector quantum efficiency (DQE), 38, 39, 412
- Detector windows
 - detector efficiency and, 296
 - for EDXS, 295–297
- Detectors/detection strategies
 - for LVSEM, 225
 - for VLVSEM, 225
- Detector/signal processing artifacts, EDXS and, 298–299
- Deterministic model-based resolution, 1236–1240
- Dielectric(s)
 - advanced SPM techniques for, 953–958
 - NC-AFM of, 943–944
 - real/imaginary part of, 395
- Differential conductance, 1071
 - spatial dependence of, 1083
 - with tip over clean terrace/steep edge, 1084
- Differential conductance mapping, 989–993
- Diffraction
 - barrier, 790
 - breaking of, 811–827
 - contrast, from Si surfaces, 625
 - limit
 - and Rayleigh resolution, 1231–1234
 - Sparrow resolution and, 1231–1234
 - superresolution and, 1230–1236
- Diffraction limited probe, 72
- Diffractive imaging, 1196–1225
 - applicability domains of, 1218
 - diffractive patterns and, 1201–1205
 - experimental results of, 1211–1219
 - HIO algorithm and, 1199–1200, 1205–1211
 - history of, 1197–1201
 - images and, 1201–1205
 - introduction to, 1196–1197
 - iterated projections of, 1219–1221
 - objects and, 1201–1205
 - validity domains of approximation and, 1201–1205
- Diffractive patterns, diffractive imaging and, 1201–1205
- Diffractogram measurements, HRTEM and, 21–22
- Diffusion coefficients, for raft proteins, 1064
- Diffusion kinetics, of Pd atoms, 1002
- Digital imaging recording, with SEM, 157–158
- Digital signal processor (DSP), 976
- 1,2-dipalmitoyl-*sn*-glycero-3-phosphothioethanol (DPPTE), secondary electron micrograph of, 231
- Direct visualization, of surface state dispersion, 1090
- Directly coupled CCD camera system, 412
- Disc of least confusion, 70
- Discrete angle sampling, illustration of, 579
- Dislocation
 - dynamics, in SiGe heterostructures, 494
 - in H atmosphere, 492
- Distance dependencies, of short/long range forces, 932
- Distance measurement
 - with CRLB, 1246–1248
 - lower bound of standard deviation of, 1247
- Distortion correction procedure, in electron holography, 1146
- DLVO force
 - force v. distance curve with, 1039
 - ion dependence of, 1054
- DLVO theory, 1037

- Domain images, of in-plane magnetization, 690
- Domain wall motion, in ferroelectrics, 485
- Dopant imaging, by X-STM, 1015
- Dopant potentials, in semiconductors, electrostatic field measurement and, 1173–1182
- Double reflection electron emission microscope (DREEM), 618
- Double-walled nanotube (DWNT)
 - electron microdiffraction pattern from, 1215
 - simultaneous structure analysis of, 504
- Doubly symmetric electrostatic corrector (DECO), of spherical/chromatic aberration, 721
- DPPE. *See* 1,2-dipalmitoyl-*sn*-glycero-3-phosphothioethanol
- DQE. *See* Detector quantum efficiency
- DREEM. *See* Double reflection electron emission microscope
- DSP. *See* Digital signal processor
- DTEM. *See* Dynamic transmission electron microscope
- Dual-axis tilting, for Cryo-ET, 559–560
- DWNT. *See* Double-walled nanotube
- Dynamic surface processes, STM and, 998–1003
- Dynamic transmission electron microscope (DTEM), 406
 - anatomy of, 407–414
 - applications for, 433–435
 - future developments for, 435–436
 - image quality simulation, 411
 - image resolution for, 423–428
 - lasers for, 419–421
 - limitations of, 421–433
 - photo of, 408
 - photoelectron guns for, 414–419
 - pulse compression and, 435–436
 - relativistic beams and, 436
 - sample damage with, 428–433
 - technologies of, 407–421
 - time resolution for, 421–423
 - TU Berlin, 409
 - ultrafast electron diffraction timeline development with, 434
- EBIC. *See* Electron beam-induced current
- EBIV. *See* Electron beam-induced voltage
- EBSD. *See* Electron backscattered diffraction
- ECP. *See* Electron channeling pattern
- Edge shapes
 - of EELS, 326–329
 - labels and, for elements, 382
 - schematic drawing of, 326
- Edge signals, extrapolation of, 344
- EDS. *See* Energy-dispersive spectroscopy
- EDXS. *See* Energy-dispersive X-ray spectroscopy
- EELM. *See* Electron energy loss microscopy
- EELS. *See* Electron energy loss spectroscopy
- EFTEM. *See* Energy-filtered transmission electron microscopy
- Einstein phonon dispersion, 93
- Elastic cross sections, 312–314
- Elastic electron scattering, inelastic electron scattering and, 310–312
- Elastic vacuum tunneling, STM and, 970–972
- Elastic/inelastic electron tunneling, band diagrams for, 973
- Elastic/inelastic tunneling channels, 1114
- Elastic/plastic deformation, *in situ* TEM, 486–500
- Elastic/total inelastic angular scattering distributions, 316
- Electric insulators, ESEM and, 135
- Electrical measurement
 - on nanostructures, 503–504
 - on TEM samples, 500–503
- Electrochemical cell, copper deposition using, 510
- electrochemical deposition, 507–509
- Electrode geometry, of symmetric unipotential lens, 667
- Electromagnetic lenses, 146
- Electron backscattered diffraction (EBSD), 134
 - crystal structure analysis by, 252–257
 - patterns, from as-cast niobium specimen, 255
- Electron beam, broadening of, 356
- Electron beam-induced current (EBIC), SEM and, 192–193
- Electron beam-induced phenomena, 510–517
 - atmosphere interaction and, 511
 - point defects formation/clustering and, 511–513
- Electron beam-induced voltage (EBIV), 192
- Electron beam-specimen interaction, with LEEM, 606–613
- Electron channeling pattern (ECP), 193, 253
- Electron confinement
 - on nanoscale hexagonal Ag islands, 1092
 - in SWCNTs, 1093
- Electron detectors, 148–152
 - strategies with, 152–154
 - types of, 150
- Electron diffraction, dynamic models of, 90
- Electron emission, schematic energy distribution of, 166
- Electron energy loss microscopy (EELM), 606
- Electron energy loss spectroscopy (EELS), 68, 97–107, 280
 - AEM and, 273
 - analysis/quantification limitations in, 346–347
 - background in, 330–331
 - of Ca, 374
 - convergence correction and, 344–346
 - detection limits, 379–381
 - of edge shapes, 326–329
 - edges, background extrapolation of, 342
 - full spectrum for, 279
 - inelastic electron scattering and, 99–101
 - mapping of, 362–374
 - microanalysis, energy-filtered microscopy and, 353–358
 - model-based quantification of, 343
 - near-edge structure and, 385
 - overview of, 278–280
 - parallel spectrometer, schematic drawing of, 305

- Electron energy loss spectroscopy (cont.)
 - quantification in, 341–347
 - quantification procedures in, 341–344
 - quantitative imaging with, 366–370
 - schematic drawing of, 279
 - signals, spatial localization of, 101–104
 - spectrometer designs of, 98–99
 - spectrometers and, 301–307
 - spectrum information from, 281
 - spectrum schematic of, 100
- Electron guns
 - characteristic parameters of, 145
 - HRTEM and, 209–221
 - SEM and, 143–145
- Electron hole pairs, 192
- Electron holography, 1141–1190
 - alternative forms of, 1184–1189
 - analysis and, 1151–1153
 - Co nanoparticle chains in, 1153–1156
 - Co nanoparticle rings and, 1157–1160
 - Co nanowires and, 1167–1168
 - cross-sectional specimens and, 1169–1171
 - digital acquisition and, 1151–1153
 - discussion on, 1189–1190
 - distortion correction procedure in, 1146
 - electrostatic field measurement in, 1172–1182
 - experimental considerations with, 1145–1148
 - FeNi nanoparticle chains and, 1160–1161
 - high-resolution, 1182–1184
 - introduction to, 1141–1148
 - lithographically patterned magnetic nanostructures and, 1164–1167
 - magnetic field measurement in, 1151–1172
 - magnetic/mean inner potential contributions in, separation of, 1156–1171
 - magnetite nanoparticle chains and, 1157
 - mean interior potential and, 1148–1150
 - micromagnetic simulations in, 1171–1172
 - montage of, 1169
 - NdFeB hard magnets in, 1153
 - off-axis, basis of, 1141–1148
 - phase profile plotting in, 1149
 - quantitative measurements in, 1171–1172
 - recording setup diagram for, 1173
 - resolution in, 1171–1172
 - sample thickness measurement and, 1148–1150
 - space charge layers, at grain boundaries, 1182
- Electron intensity dispersion, real space intensity plots of, 316
- Electron interference, in Pb, 1012
- Electron lens aberrations, in PEEM, 673
- Electron lenses
 - HRSEM and, FESEM and, 211
 - SEM and, 145–148
- Electron microdiffraction pattern, from double-walled nanotube, 1215
- Electron microscopy (EM), 535
 - applicability domains of, 1218
 - transfer functions of, 1233
- Electron mirror, aberration corrected microscope using, 674–675
- Electron optical configuration
 - of AEM, 290–293
 - of illumination system, 291
 - of PEEM, schematic drawing of, 664
 - for STEM, 292
- Electron optics
 - column design and, 109–113
 - for Cryo-ET, 561
 - LEEM and, 619–624
 - new developments in, 310
- Electron probe current
 - in SEM, 147
 - variation in, 288
- Electron probe size, variation in, 288
- Electron rays, in OL, 667
- Electron sources
 - for HRTEM, 30
 - for STEM, 114–117
- Electron sources/probes, for AEM, 282–290
- Electron spectroscopy for chemical analysis (ESCA) microscopy, 903
- Electron trajectories, of Monte Carlo simulation, 167
- Electron tunneling, schematic energy diagram of, 1098
- Electron tunneling, hot, schematic energy diagram of, 1099
- Electronic samples, broadening of, 350
- Electronic structure, LT-STM and, 1082–1105
- Electronic structure/bonding environment, ELNES and, 278
- Electron-specimen interaction, 182
 - signal generation and, 163–182
- Electroscan environmental scanning electron microscope, schematic cross section of, 239
- Electrostatic correction, of chromatic aberration, 722–723
- Electrostatic field measurement
 - in electron holography, 1172–1182
 - field-emitting carbon nanotubes and, 1172–1173
 - semiconductors and, dopant potentials in, 1173–1182
- Electrostatic potential, power series expansions of, 729–730
- Electrostatic potential distribution
 - in semiconductor device structure, 1175
 - simulations of, 1179
- Elemental mapping
 - for AEM, 358–374
 - color-coded, 359
 - of EDXS, 358–362
 - of NiAl multilayer, 361
 - of Ni-based alloy, 360
- Elements, edge shapes for, labels and, 382
- Element-specific imaging, 995–996
 - with atomic resolution, 996
- Elevated pressure, SEM at, 237–244
- Elevation angle/foil thickness, absorption length dependence on, 336
- Elliptically polarizing undulators (EPU), 659
- ELNES. *See* Energy loss near-edge structure

- EM. *See* Electron microscopy
- Embedded nanostructures, size-dependent transformations in, 458–460
- Emitted secondary electron, energy distribution of, 660
- Energy dependence
 of attenuation coefficients μ, ν , 610
 of backscattering, 608
 of elastic mean free electron path, of Xe, 611
 of inelastic mean free electron path, of Au, 610
 of inelastic mean free electron path of Fe³⁸, 612
 resolution and, 621
- Energy dispersion measurements, LT-STM and, 1086–1090
- Energy distribution, of emitted secondary electron, 660
- Energy electron distribution, chromatic correction and, 709
- Energy filter, for Cryo-ET, 566–570
- Energy filters, for HRTEM, 35–36
- Energy level diagrams, 275
 of sample/tip, 1072
- Energy level quanta, 275
- Energy loss fine structures, 384–399
- Energy loss near-edge structure (ELNES), 101, 385–392
 calculation examples of, 391, 392
 electronic structure/bonding environment and, 278
 regions/energy ranges for, 384
 shape, multiple inelastic losses and, 399
- Energy loss spectra, diagram of, 281
- Energy-dispersive spectroscopy (EDS), 107
 spectrum, 276
- Energy-dispersive X-ray spectroscopy (EDXS), 68, 134, 280
 AEM and, 273
 background in, 330–331
 detection limits for, 375–377
 detector, 293
 diagram of, 294
 schematic drawing of, 293
 detector front, schematic drawing of, 295
 detector interface, 300
 detector windows for, 295–297
 detector/signal processing artifacts, 298–299
 elemental mapping, 358–362
 features of, 249
 geometry of, 299–300
 instrument contributions in, 377–379
 microanalysis of, 331–347, 348–353
 overview of, 274–277
 peak shapes and, 297–298
 peak/background intensities in, 376
 quantification in, 331–335
 elemental maps of, 352
- Energy-filtered electron diffraction pattern, of Si(110), 307
- Energy-filtered electron microscopes, schematic drawing of, 302
- Energy-filtered image, of Al alloy, 354
- Energy-filtered microscopy, EELS microanalysis and, 353–358
- Energy-filtered scanning transmission microscope, 985
- Energy-filtered transmission electron microscopy (EFTEM), 105
 elemental mapping
 expected resolution of, 373
 spatial resolution of, 371–374
 imaging approaches, 363
 mapping, 362–365
 spectrum imaging technique for, 364
- Energy/fluorescence yields, for K/L edge emission, 838
- Environmental scanning electron microscope (ESEM), 237–238
 electric insulators and, 135
 secondary electron micrograph with, 242
- Environmental science applications, for X-ray microscopes, 896–897
- Environmental secondary electron detector (ESD), 240
- Epifluorescence micrographs, visible light and, 898
- Epitaxial Fe ribbon crystals, on W(110), 647
- Epitaxial/polycrystalline thin film growth, 472–474
- EPU. *See* Elliptically polarizing undulators
- Equipotential contours, of PEEM, objective lens, 666
- Equivalent optical configuration, for confocal setup, 757
- ER algorithm. *See* Error-reduction algorithm
- Error-reduction (ER) algorithm, 1199
- ESCA microscopy. *See* Electron spectroscopy for chemical analysis microscopy
- ESD. *See* Environmental secondary electron detector
- ESEM. *See* Environmental scanning electron microscope
- ETD. *See* Everhart-Thornley detector
- Eukaryotic cell, ET of, 545
- Everhart-Thornley detector (ETD), 149, 186, 187
 BSE and, 187
 schematic drawing of, 151
 topographic contrast and, 184
- EXAFS. *See* Extended X-ray absorption fine structure
- Excitation supports, comparison of, 797
- Exit-plane wavefunction, reconstructed modulus of, 49
- Exit-wave reconstruction
 of complex oxides, 48–51
 experimental geometries and, 47–48
 HRTEM and, 43–51
 theory of, 44–47
- Expected edges, summary of, 329
- Experimental geometries, exit-wave reconstruction and, 47–48
- Expitaxially strained materials, relaxation of, 493–495
- Extended X-ray absorption fine structure (EXAFS), 101
- External voltage application, schematic drawing of, 1177

- Fabrication techniques, for zone plate, 853–855
- Fabry-Perot resonator, 1085
- Far field fluorescence microscopy, nanoscale resolution in, 790–829
- Fast high-voltage switches, 414
- Fe atoms, in circular quantum corral, 1091
- Fe film, quantum size contrast of, 629
- Fe islands, magnetic vortex states on, 1127
- Fe stripe domain structure, in Fe/Ni/Cu(001), 681
- Fe-Co alloy layer, spin reorientation transition in, 646
- FEGs. *See* Field-emission guns
- FeNi nanoparticle chains
 - chemical map of, 1161
 - electron holography and, 1160–1161
- Fe/Ni/Cu(001), Fe stripe domain structure in, 681
- Fermi contour imaging
 - of Be(0001), 1105
 - of Be(1010), 1105
 - LT-STM and, 1104–1105
- Fern fossil
 - backscattered electron micrograph of, 205
 - secondary electron micrograph of, 205
- Ferroelectrics
 - domain wall motion in, 485
 - switching, 483–485
- Ferromagnetic domains, imaging of, 680–681
- Ferromagnetic materials, magnetic domain structure/motion in, 476–481
- Ferromagnets, linear dichroism imaging of, 685–686
- FESEM. *See* Field-emission scanning electron microscopy
- FET. *See* Field effect transistor
- FFP. *See* Front-focal plane
- FIB. *See* Focused ion beam
- Fibroblast, human, with immunogold labeling, 893
- Fibroblast imaged, in frozen-hydrated state, 893
- Field effect transistor (FET), 295
- Field ion microscopy (FIM), 974
- Field-emission electron gun, schematic drawing of, 286
- Field-emission guns (FEGs), 112, 117, 134, 1141
 - schematic drawing of, 210
- Field-emission scanning electron microscopy (FESEM), 134, 209
 - contrast formation and, 212–214
 - detection geometries and, 211
 - detectors and, 211
 - HRSEM and
 - electron guns and, 209–210
 - electron lenses and, 211
 - modern high-resolution, 135
 - resolution and, 212–214
 - specimen stages and, 212
- Field-emitting carbon nanotubes, electrostatic field measurement and, 1172–1173
- 50eV electrons, angular distribution of, 607
- FIM. *See* Field ion microscopy
- Finite element calculation (FEM), 33
- Finite source size, STEM and, 118–119
- FITC. *See* Fluoresceine isothiocyanate
- Fixed 3T3 cell, AFM image of, 1049
- FLM. *See* Fluorescence light microscopy
- Fluoresceine, 178
- Fluoresceine isothiocyanate (FITC), 178
- Fluorescence, 754–755
 - correction, 340
 - definition of, 339
 - focal spot, STED and, 822
 - yield, of X-rays, 180
- Fluorescence light microscopy (FLM), 149
- Fluorescence resonance energy transfer (FRET), 1049
- Fluorescent dyes, multiple excitation of, 775
- Fluorescent molecules
 - under TPE regime, 765–767
 - two-photon cross sections for, 767
- Fly head/eye, micrograph of, 142
- FM-AFM. *See* Frequency-modulated atomic force microscope
- Focus depth, of SEM, 142
- Focused ion beam (FIB)
 - of color film, secondary electron micrograph of, 200
 - SEM and, 196, 197, 200
 - of tyre, secondary electron micrograph of, 200
- Focusing optics, of zone plate X-ray microscopes, 840–843
- Foil correctors, aberration correctors and, 716–718
- Foil lens, schematic drawing of, 717
- Force dependent surface tomography, of bacteriorhodopsin, 1036
- Force v. distance curve, 1034
 - with DLVO force, 1039
- Force v. penetration depth, parameter plot of, 1035
- 4Pi-microscopy
 - lobe removal/deconvolution in, 807
 - OTF of, 799–803
- Fourier domain, schematic sectors of, 558
- Fourier optics treatment, partial incoherence and, 878–880
- Fourier space, density problem sampling in, illustration of, 581
- Fourier transform spectroscopy, on Bi₂Sr₂CaCu₂O_{8+δ}, 991
- Free-standing nanoparticles, phase transformations/sintering in, 461–463
- Frequency response
 - of closed feedback loop, in AFM, 1045
 - of commercial scanner, 1044
 - of silicon nitride cantilever, 1044
- Frequency space, OTF and, 793
- Frequency-dependent oscillator strength, for on/gold, 837
- Frequency-modulated atomic force microscope (FM-AFM)
 - of GaAs<110>, 940–942
 - NC-AFM and, 935–937
 - of semiconductors, 940–942
 - of Si, 940–942

- Fresnel zone plates
 introduction to, 844–846
 zone plate efficiency and, 851–853
 zone plate image quality and, 846–850
 zone plate X-ray microscopes and, 844–862
- FRET. *See* Fluorescence resonance energy transfer
- Front-focal plane (FFP), 70
- Frozen-hydrated state, whole fibroblast imaged in, 893
- Full width at half-maximum (FWHM), 72, 73
 of confocal PSFs, 759
- Functionalized tip, chemical contrast with, 1109
- FWHM. *See* Full width at half-maximum
- Ga thermometer, 451
- GaAs(110)
 ADF detector images, 95
 bias-dependent STM on, 986
 FM-AFM of, 940–942
 local normalized conductance spectra on, 987
- Gaseous secondary electron detector (GSED), 240
- Gaussian distribution functions, 16
- Gd, spectrum imaging for, 105
- Gd(0001), spin-polarized STM on, 992
- Ge
 2D to 3D morphology transition of, 981
 STM imaging of, 979
- Ge(105), surface structure identification of, 1019
- Ge island nucleation, 473
- General oscillator strength (GOS)
 angular/energy dependence of, 321–323
 two-dimensional angular/energy distribution, 322
- Germanium test pattern, amplitude/phase contrast images of, 873
- Ge/Si (105), atomic resolution, NC-AFM/STM/KFM on, 959
- Glass knife, secondary electron micrograph of, 234
- Glass micropipette, secondary electron micrograph of, 233
- Gold(100), lower bound of standard deviation for, 1257
- Gold balls
 SEM image of, 1213
 soft X-ray transmission diffraction pattern from, 1212
- Goniometer
 for HRTEM, 34–35
 tilt stage, for Cryo-ET, 562–563
- GOS. *See* General oscillator strength
- Grain boundaries, at space charge layers, in electron holography, 1182
- Grain boundary motion, grain growth and, 453
- Grain growth
 grain boundary motion and, 453
 phase transformations and, 448, 452
- Graphite, NC-AFM of, 943
- Growth process, 513–514
- GSED. *See* Gaseous secondary electron detector
- Gun anodes, 286
- H atmosphere, dislocation in, 492
- Hard X-ray zone plates, 859–861
 aberration, size of, 849
 image of, 860
- Helicoidal biological sample, 3D views of, 762
- Heterostructures, buried interfaces and, 1007–1016
- HFEG. *See* Hot field emission gun
- HfO₂, from magnetic tunnel junction, off-axis electron holography and, 1170
- High energy incident electrons, signals from, 274
- High speed electron microscopy, 406–437
 uses of, 406
- High vacuum (HV), 135
- High-energy X-rays, depth-of-focus limit and, in x-ray microscopes, 887–888
- Higher-order Laue zone (HOLZ), 93
- High-resolution constant height images, of CaF₂, 944
- High-resolution electron holography, 1182–1184
- High-resolution micrographs
 of aluminum oxide, 221
 of PVME, 221
- High-resolution scanning electron microscopy (HRSEM)
 with FESEM, 213
 FESEM and, 209–221
 electron lenses and, 211
 selected applications of, 214–221
- High-resolution transmission electron microscopy (HRTEM), 3
 aberration correctors and, 39–42
 advances in, 5
 of aluminum crystal, 1249, 1250
 axial diffractograms images from, 20
 coherence and, 26–29
 detectors and, 36–39
 diffractogram measurements and, 21–22
 electron guns and, 209–221
 electron sources for, 29–31
 physical properties of, 30
 energy filters for, 35–36
 essential theory of, 8–29
 exit-wave reconstruction and, 43–51
 future prospects on, 57–58
 goniometer for, 34–35
 historical summary of, 6–8
 image formation and, 8–13
 optical ray diagram schematic of, 9
 image simulation of, 51–57
 information limit of, 14
 instrumentation for, 29–43
 isoplanatic aberration coefficients in, 18
 of lattice matched heterojunction between InP and As, 5
 monochromators and, 40, 42–43
 of nanocrystalline gold particle, 6
 nanoscale materials and, 5
 optics for, 32–34
 PCD and, 12
 PCTF and, 14

- High-resolution transmission electron microscopy
(cont.)
photo of, 4
POA and, 9
prevalence of, 3
resolution limits of, 13–17
specimen stages for, 34–35
spherical aberration and, 32
statistical experimental design of, 1256–1258
studies on, 3–5
TCC and, 27–28
tilt-induced displacements and, 21
twins, stacking faults image from, 4
wave aberration function and, 17–26
WPOA and, 10
- HIO algorithm. *See* Hybrid input-output algorithm
- Histogram, of CCC, 586
- Hologram reconstruction, 1143–1145
- HOLZ. *See* Higher-order Laue zone
- Homoepitaxial growth, of Si(100), 633
- Hot field-emission gun (HFEG), 210
- HRSEM. *See* High-resolution scanning electron microscopy
- HRTEM. *See* High-resolution transmission electron microscopy
- HV. *See* High vacuum
- Hybrid input-output (HIO) algorithm
diffractive imaging and, 1199–1200, 1205–1211
variants of, 1205–1211
- Idealized structures, zone plate efficiency and, 851–853
- IETS. *See* Inelastic electron tunneling spectroscopy
- ILDOS. *See* Integrated local density of states
- Illumination system, electron optical configuration and, 291
- Image contrast, 883
in XPEEM, 661–663
- Image interpretation, tip and, 937–939
- Image resolution, for DTEM, 423–428
- Image simulation
of HRTEM, 51–57
multislice formalism and, 52–57
- Imaging buried heterostructures, cross-sectional STM and, 1013–1016
- Imaging modes, of AFM, 1046–1048
- Imaging radiation dose, for protein features, 840
- Imaging-to-diffraction mode, in STEM column, 110–112
- Immunogold labeling, human fibroblast, 893
- Implants, investigation of, 207
- “In lens” detection, 154
schematic drawing of, 155
- In situ* imaging, of catalysts, 468
- In situ* indentation/straining, of thin films, 495–496
- In situ* transmission electron microscopy, 445–516
catalysis and, 466–467
definition of, 445–447
elastic/plastic deformation in, 486–500
future outlook for, 517–518
grain boundary dynamics in, 452
magnetic/ferroelectric/superconducting materials in, 476–486
phase transformations and, 447–462
crystallization and, 448–449
grain growth and, 448, 452
melting and, 448–449
in nanostructured materials, 458–463
structural phase transitions in, 453–457
superconducting materials in, 481–483
surface reactions and, crystal growth, 463–476
- Incoherent illumination, with TXM, schematic drawing of, 877
- Incoherent imaging, 103
CTEM and, 84–90
v. beam current, 119
- Incoherent optical transfer function, plot of, 121
- In-column energy filtered microscope, schematic drawing of, 304
- In-column spectrometer configurations, 303
- Inelastic electron scattering
diagram of, 319
EELS and, 99–101
elastic electron scattering and, 310–312
localization of, 358
processes of, 312
- Inelastic electron tunneling spectroscopy (IETS), 993
local, 1114–1119
- Inelastic imaging, spatial localization of, 101–104
- Inelastic mean free electron path, of Au, energy dependence and, 610
- Inelastic mean free electron path of Fe³⁸, energy dependence of, 612
- Inelastic scattering cross sections, 314–320
- Inelastic tunneling, vibrational spectroscopy and, 993–994
- Inelastic vacuum tunneling, 973
- Information limit, of HRTEM, 14
- In-plane magnetization, domain images of, 690
- Instrument noise, in AFM, 1045–1046
- Instrumentation
for AEM, 282–300
for AFM, 1030–1052
for Cryo-ET, 550, 560–576
for HRTEM, 29–43
of LEEM, 614–619
for SPLEEM, 614–619
for STM, 975–976
- Integrated circuits
studies in, 901
tomographic imaging of, 901
with vertical cross sections, 198
Zernike phase contrast of, 900
- Integrated local density of states (ILDOS), 1071
oscillations, bias-dependent evolution of, 1087
- Interfaces/grain boundaries, geometry analysis for, 338
- Interference detection, with ADF detector, 86

- Interferometric TXM image, of polystyrene spheres, 900
- Intermediate layer, hexagonally packed, 1056
- Intermediate/projective lens, on PEEM, 666–668
- Intermittent contact atomic force microscope, 930
- Intermittent contact mode, of AFM, 1047–1048
- Intermixing, in small particles, 462
- Intrinsic amplitude, Zernike phase contrast and, in protein, 876
- Inverse photoemission spectroscopy (IPES), 1082, 1120
- Ion dependence, of DLVO force, 1054
- Ion implantation, 514–515
- Ion/electron beam-induced processes, 510–517
- Ions at solid-liquid interfaces, specific interaction of, 1038
- IPES. *See* Inverse photoemission spectroscopy
- Iron oxide clusters, XPEEM image of, 662
- Irradiation damage, 513
- Isoplanatic aberration coefficients
 - accuracy of, 19
 - in HRTEM, 18
- Isoplanatic approximation, 17
- K edges, trends in, 327
- K_{AFe} , experimental factors of, 334
- KBr, NC-AFM of, 943–944
- Kelvin force microscopy (KFM), 946
 - of $\Sigma 5$ grain boundary in SrTiO_3 (100), 949
- Kelvin probe force microscopy (KPFM), 945
- Keratinocyte, secondary electron micrograph of, 236
- K-factor approach, 346
- KFM. *See* Kelvin force microscopy
- K/L edge emission, energy/fluorescence yields for, 838
- Kondo effect, LT-STM and, 1102–1104
- Kondo resonance, spectroscopic signature of, 1103
- KPFM. *See* Kelvin probe force microscopy
- Kramer's constant, 179
- L_{23} edge, trends in, 328
- La B_6 . *See* Lanthanum hexaboride
- La FeO_3 , magnetic contrast in, 683
- Lander molecule
 - chemical structure of, 1110
 - on $\text{Cu}(111)$ surface column, 1111
- Langmuir-Blodgett films, AFM and, 1056–1057
- Lanthanum hexaboride (La B_6), 144
- Large detector incoherent, BF STEM, 82
- Large-area atom-resolved, STM, 981
- Laser(s)
 - for DTEM, 419–421
 - sources, for TPE, 773
- Laser emission time scale, for TPE, 773
- Laser scanning microscopy (LSM), 1049
 - confocal principles and, 755–757
- Laser-synchrotron, pulse sequence in, 687
- Lateral atom manipulation, mechanisms of, 1004
- Latex-beads, tapping mode imaging of, 1048
- Lattice displacements, in crystal strain, 93
- Lattice imaging, in BF STEM, 79–81
- Lattice matched heterojunction between InP/As , HRTEM of, 5
- Lattice mismatch strain, during $\text{Si}_{1-x}\text{Ge}_x/\text{Si}(001)$ heteroepitaxy, 982
- LCD. *See* Liquid crystal display
- LCF. *See* Local cross-correlation function
- LDOS. *See* Local density of states
- Leaf cuticular, secondary electron micrograph of, 237
- LEED. *See* Low-energy electron diffraction
- LEELM. *See* Low-electron energy loss microscopy
- LEEM. *See* Low-energy electron microscopy
- Lens aberrations, 33
- Lens-free imaging, depth-of-focus limit and, in x-ray microscopes, 888
- Lensless imaging. *See* Diffractive imaging
- LHe continuous-flow cryostat, STM with, 1074–1076
- Lifetime measurements, LT-STM and, 1093–1100
- Linear dichroism
 - imaging, of ferromagnets, 685–686
 - magnetic circular and, 678–680
- Lipid raft, size/temporal stability of, 1064
- Liquid crystal display (LCD), 139
- Liquid film, imaging under, 506–507
- Liquid phase processes, 505–510
 - reaction *in situ*, 508
- Lithographed letters, soft X-ray transmission diffraction pattern from, 1211
- Lithographically patterned magnetic nanostructures, electron holography and, 1164–1167
- Lobe removal/deconvolution, in 4pi microscopy, 807
- Local cross-correlation function (LCF), 585
- Local density of states (LDOS), 1071
 - oscillations, 1082–1086
 - on $\text{Ag}(111)$, 1089
 - bias-dependent evolution of, 1087
- Local element analysis, 134
- Local impedance spectra, Cole-Cole plot of, 953
- Local inelastic spectroscopy, 1114–1119
- Local I-V measurements, current imaging tunneling spectroscopy, 986–989
- Local magnetic induction, phase contours of, 1162
- Local normalized conductance spectra, on $\text{GaAs}(110)$, 987
- Locally probing macromolecular, cellular samples and, 1058–1064
- Lorentz image, of NdFeB hard magnets, 1154
- Low magnification bright-field image, of self-assembled Co nanoparticle, 1159
- Low vacuum (LV), 135
- Low voltage scanning electron microscope (LVSEM), 148, 221–237
 - contrast formation for, 225–226
 - detectors/detection strategies for, 225
 - selected applications for, 226–237

- Low-electron energy loss microscopy (LEELM), 615
- Low-energy electron diffraction (LEED), 605
 - spot movements, drawing of, 627
- Low-energy electron microscopy (LEEM), 605–647
 - applications of, 630–642
 - background on, 630–631
 - beam separator for, 622
 - cathode lenses configurations for, 620
 - commercial instrument of, 616
 - contrast mechanisms with, 624–630
 - cross section of, 616
 - electron beam-specimen interaction with, 606–613
 - electron optics and, 619–624
 - flange-on, schematic drawing of, 617
 - instrument schematic of, 614
 - instrumentation for, 614–619
 - lenses resolution for, 620
 - metal layers on metals and, 639–640
 - metal surface reactions and, 640–641
 - metal surfaces in, 638–639
 - nitrides and, 641
 - oxides and, 641
 - phase contrast imaging in, conditions for, 628
 - Si(100) and, 632–634
 - Si(111) surface with, 631–632
- Lower bound of standard deviation
 - distance measurement of, 1247
 - of gold(100), 1257
 - for spherical aberration, 1256
- Low-loss region, 100
- Low-loss spectroscopy, 393–399
 - applications of, 394–399
 - fundamentals of, 393–394
 - variation example of, 396
- Low-temperature scanning tunneling microscopy (LT-STM), 1070–1132
 - Abrikosov flux lattice and, 1130
 - applications of, 1081–1132
 - design drawing of, 1080
 - design principle illustration of, 1081
 - design principles of, 1073–1081
 - electronic structure and, 1082–1105
 - energy dispersion measurements and, 1086–1090
 - Fermi contour imaging and, 1104–1105
 - Kondo effect and, 1102–1104
 - lifetime measurements and, 1093–1100
 - magnetic atoms at surfaces and, 1102–1104
 - schematic drawing of, 1079
 - single atom and, molecule manipulation, 1105–1114
 - SP-STM and, 1125–1132
 - stark effect and, 1100–1101
 - STM-induced photo emission and, 1119–1125
 - superconductivity and, 1129–1132
- Low/variable-temperature STM, photo of, 1077
- LSM. *See* Laser scanning microscopy
- LT-STM. *See* Low-temperature scanning tunneling microscopy
- Lutetium, cluster analysis of, 898
- LV. *See* Low vacuum
- LVSEM. *See* Low voltage scanning electron microscope
- Macromolecular complexes, identification strategies of, 584
- Macromolecular samples, for AFM, 1053–1057
- Macromolecules, detection/identification of, 584
- Magnetic atoms at surfaces, LT-STM and, 1102–1104
- Magnetic circular, linear dichroism and, 678–680
- Magnetic contrast
 - in La FeO₃, 683
 - scheme of type-1, 195
 - SEM and, 194–195
 - types of, 194–195
- Magnetic detection of electrons, schematic drawing of, 154
- Magnetic domain imaging, 678–686
 - of Co/NiO, 683
- Magnetic domain structure/motion, in ferromagnetic materials, 476–481
- Magnetic fields
 - early experiments in, 1151
 - measurement, in electron holography, 1151–1172
- Magnetic force microscopy (MFM), 1126
- Magnetic materials
 - applications of, for X-ray microscopes, 903–905
 - phase transformations in, 482
 - phase transitions in, 480–481
- Magnetic random access memory (MRAM), 678
- Magnetic vortex states, on Fe islands, 1127, 1128
- Magnetic/ferroelectric/superconducting materials, for *in situ* transmission electron microscopy, 476–486
- Magnetic/mean inner potential contributions, separation of, 1156–1171
- Magnetite nanoparticle chains, electron holography and, 1157
- Magnetization reversal, of Co/LaFeO₃, 684
- Magneto-optical Kerr effect (MOKE) microscopy, 1126
- Magnetotactic microorganism, cellular cryo-ET of, 543
- Magnification calibration, for SEM, 162–163
- MAL. *See* Maximum likelihood
- Manganese nodule, chemical imaging of, 663
- Material contrast, SEM and, 189–191
- Materials science
 - applications, of X-ray microscopy, 897–903
 - scanning probe microscopy in, 929–960
- Maximum likelihood (MAL), 46
 - estimation/resolution and, 1248–1252
- MDFF. *See* Mixed dynamic form factor
- Mean interior potential
 - electron holography and, 1148–1150
 - phase contours of, 1158
 - of representative materials, in volts, 10
- Mean ionization potential, Bethe stopping power and, for carbon/protein, 206

- Mechanical decoupling, of copper braid, 1075
- Melting, phase transformations and, 448–449
- MEM. *See* Mirror electron microscopy
- MEMS. *See* Microelectromechanical systems
- Metal layers on metals, LEEM and, 639–640
- Metal surfaces
 - in LEEM, 638–639
 - reactions of, LEEM and, 640–641
- Metal-induced crystallization, 449
- Metallic emitters, photoelectron guns and, 417–419
- Metastable impact electron emission microscopy (MIEEM), 618, 642
- MFM. *See* Magnetic force microscopy
- Microanalysis
 - detection limits in, 374–384
 - resolution in, 348
 - in SEM, 246–252
- Microchannel plate detector, 151
- Micro-diffraction pattern, 73
- Microelectromechanical systems (MEMS), 207
- Micromagnetic simulations, in electron holography, 1171–1172
- Microscopic phenomenon, during deformation, 487–493
- Microtechnology, 207
- Microtubulis, in vitreous ice, imaging of, 568
- MIEEM. *See* Metastable impact electron emission microscopy
- MIMAPI. *See* Multiple input maximum a posteriori
- Mirror correctors, 725–726
- Mirror electron microscopy (MEM), 606
- Mixed dynamic form factor (MDFF), 102
- Mixed quadrupole correctors, chromatic aberration correctors and, 719–723
- Modulation transfer function (MTF), 37, 878
 - CCD cameras and, 40
 - formula for, 38
 - NTFs and, 37
 - of PEEM, 670
 - for STXM imaging, 878
- MOKE microscopy. *See* Magneto-optical Kerr effect microscopy
- Molecular conformation identification, of Cu-DTBPP, 1019
- Molecular fluorescence, 1122
- Molecular vibrational spectra, with STM, 1117
- Molecular wavefunctions, in carbon nanotube, 1095
- Molecular wire contact, to step, 1111
- Molecule cascades, 1005
- Molecule manipulation, single atom and, LT-STM and, 1105–1114
- Monatomic steps, on MO(101) surface, 627
- Monochromators, 307–308
 - HRTEM and, 42–43
 - implementations of, 308
 - visibility peak from, 309
 - ZL peak from, 309
- Monte Carlo simulation, of electron trajectories, 167
- MO(101) surface, monatomic steps on, 627
- MRAM. *See* Magnetic random access memory
- MTF. *See* Modulation transfer function
- Multilayered film *in situ*, magnetization of, 479
- Multiphase/composite/layered materials,
 - deformation of, 490–493
- Multiple inelastic losses, on ELNES shape, 399
- Multiple input maximum a posteriori (MIMAPI), 46
- Multislice algorithm, schematic diagram of, 52–57
- Multislice formalism
 - image simulation and, 52–57
 - phonon scattering and, 92
- Multispecimen chamber/cartridges, 564
- Multiwalled nanotube (MWNT), 504
- MWNT. *See* Multiwalled nanotube
- Nanobelt, mechanical properties of, 499
- Nanocrystalline gold particle, HRTEM of, 6
- Nanoelectromechanical systems (NEMS), 221
- Nanoimpedance microscopy (NIM), 952
- Nanoindentation, of steel, 496
- Nanomanipulation, tribology and, 497–498
- Nanoparticle melting, 459
- Nanoscale materials, HRTEM and, 5
- Nanoscale resolution, in far field fluorescence microscopy, 790–829
- Nanosecond multiframe operation, 407
- Nanostructured materials, phase transformations in, 458–463
- Nanostructures
 - beam-vapor interaction and, 512
 - electrical measurement on, 503–504
 - mechanical properties of, 495–500, 498–500
 - from surface mobility, 466
- “Nanotip” models, in scanning probe microscopy, 938–939
- Nanowires
 - electrical properties of, 505
 - self-assembly of, at step edges, 1112
- Nb M₄₅ edges, Ba M₄₅ edges and, 330
- NC-AFM. *See* Noncontact atomic force microscope
- NdFeB hard magnets
 - in electron holography, 1153
 - Lorentz image of, 1154
- Near field optical microscopy, 929
- Near-carbon-edge absorption spectra, of amino acids, 890
- Near-edge structure
 - EELS and, 385
 - examples of, 386, 387
- Near-field microwave microscopy (NFMM), 956
- NEMS. *See* Nanoelectromechanical systems
- Neuronal cell, 1029
- NFMM. *See* Near-field microwave microscopy
- Ni grain boundary, S segregation at, 351
- NiAl, phase stability in, 454
- NiAl multilayer, elemental mapping of, 361
- Ni-based alloy, elemental mapping of, 360
- Ni-Fe film, XMCD of, 905
- NIM. *See* Nanoimpedance microscopy

- Nion quadrupole-octopole corrector, schematic drawing of, 712
- NiS₂/Si(001)<110> interface, ADF imaging of, 95
- Nitrides, LEEM and, 641
- Noise transfer functions (NTFs), 37
- Noncontact atomic force microscope (NC-AFM), 930
 - AM-AFM and, 934–935
 - applications of, 939–944
 - block diagram of, 931
 - of CaF₂, 943–944
 - of dielectrics, 943–944
 - FM-AFM and, 935–937
 - of graphite, 943
 - of KBr, 943–944
 - operational principles of, 930–933
 - of Si(100), 942
 - of Si(111), 936
 - of SrTiO₃, 939–940
 - techniques of, 934–937
 - of TiO₂(100), 941
- Noncontact mode, of AFM, 1046–1047
- Noncrystalline materials, ronchigrams of, 77–78
- Nonpurified samples, imaging of, 589
- Normal incidence specular reflectivity, of W(110) surface, 608
- NTFs. *See* Noise transfer functions
- Nuclear magnetic resonance (NMR), 1027
- Nucleation, surface step structure growth and, 631
- Nyquist frequency, 56

- O₂-mediated diffusion, of oxygen vacancies, on TiO₂(110), 999
- Object function, 89
- Objective aperture, 80, 113
- Objective lens (OL)
 - condenser and, 110
 - electron rays in, 667
 - of PEEM, 665
- Objects, diffractive imaging and, 1201–1205
- Off-axis electron holography
 - basis of, 1141–1148
 - of Co particles, 1155
 - from magnetic tunnel junction, of HfO₂, 1170
 - ray diagram of, 1186
 - schematic drawing of, 1142
 - from thin crystal, 1144
- OIM. *See* Orientation imaging microscopy
- OL. *See* Objective lens
- 1s state for, Si(100)/gold(100), 1255
- Operating environment, for STM, 973–975
- Optical microscopy, AFM and, 1049–1050
- Optical path function analysis, zone plate image quality and, 846–847
- Optical system, schematic drawing of, 725
- Optical transfer function (OTF), 88, 791, 878
 - of 4Pi-microscopy, 799–803
 - excitation/detection/effective modulus overview of, 800
 - frequency space and, 793
 - Optics, for HRTEM, 32–34
 - Orders number, of zone plates, 845
 - Orientation imaging microscopy (OIM), 255
 - OTF. *See* Optical transfer function
 - Outer shell excitations, 320
 - Oxidation
 - of block oxide structure, 470
 - of surface, 467–469
 - Oxides
 - LEEM and, 641
 - SPM of, 939, 940
 - Oxygen vacancies, of O₂-mediated diffusion, on TiO₂(110), 999
 - Oxygen/CO, STM of, 1109

 - Parabolic mirrors, 152
 - Parallel electron energy-loss spectroscopy (PEELS), 99
 - Parallel pencil beam, 111
 - Parametric statistical models of observation, statistical model-based resolution with, 1242–1244
 - Parasitic aberrations, 705–706
 - Partial coherence
 - contrast transfer and, 880–882
 - reciprocity and, 882
 - transfer functions, for 1D object, 881
 - in X-ray microscopes, 876–884
 - Partial incoherence, Fourier optics treatment and, 878–880
 - Partial volume, depth-of-focus limit and, in x-ray microscopes, 886
 - Partial/total ionization cross sections, 323–325
 - Patterned substrates, crystal growth on, 474, 475, 476
 - Pb, electron interference in, 1012
 - Pb growth, on Si(111), 636
 - PbTiO₃, vertical phase/amplitude hysteresis loops of, 955
 - PCD. *See* Projected charge density
 - PCF. *See* Phase correlation function
 - PCF/PCI approach. *See* Phase correlation function/phase contrast index function approach
 - PCTF. *See* Phase contrast transfer function
 - Pd atoms, diffusion kinetics of, 1002
 - Peak shapes, EDXS and, 297–298
 - Peak/background intensities, in EDXS, 376
 - PEELS. *See* Parallel electron energy-loss spectroscopy
 - PEEM. *See* Photoemission electron microscopy
 - Periodic artifacts removal, through deconvolution, 803–807
 - Periodic dielectric constant, 958
 - PFM. *See* Piezoelectric force microscopy
 - Phase change materials, amorphous to crystalline transformation in, 450
 - Phase component maps, STEM images and, 368
 - Phase contours
 - of local magnetic induction, 1162
 - of mean inner potential, 1158

- Phase contrast imaging
 - in BF STEM, 81–82
 - in LEEM, conditions for, 628
- Phase contrast index function (PCI)
 - HRTEM and, 24
 - for mismatched values, 25
- Phase contrast layout, of TXM, 868–870
- Phase contrast transfer function (PCTF), 82
 - HRTEM, 14
 - plottings of, 15
- Phase correlation function (PCF)
 - defocus difference and, 24
 - relative defocus and, 23–24
 - XCF and, 23
- Phase correlation function/phase contrast index function (PCF/PCI) approach, 25
- Phase diagram, of water, 238
- Phase differences, step contrast for, 629
- Phase identification, by dark-field imaging, 626
- Phase object approximation (POA), HRTEM and, 9
- Phase problem, in optics, 1197
- Phase profile plotting, in electron holography, 1149
- Phase stability, in NiAl, 454
- Phase transformations
 - amorphization and, 448–449
 - crystallization and, 448–449
 - grain growth and, 448, 452
 - in magnetic material, 482
 - melting and, 448–449
 - in nanostructured materials, 458–463
 - of silicide, 457
 - sintering and, in free-standing nanoparticles, 461–463
 - in situ* transmission electron microscopy and, 447–462
 - solid-liquid interface and, 450–453
 - solid-to-liquid transformation and, 450–453
- Phase transitions, in magnetic materials, 480–481
- Phase-only filtering (POF), 585
- Phonon scattering, 91
 - multislice formalism and, 92
- Phospholipid/protein film, secondary electron micrograph of, 229, 230
- Photoelectron guns, for DTEM, 414–419
 - materials for, 416–419
 - metallic emitters and, 417–419
 - semiconducting emitters and, 416–417
- Photoemission, XPEEM and, 659–661
- Photoemission electron microscopy (PEEM), 606, 657–691
 - aberration correction of, 671–677
 - electron lens aberrations in, 673
 - electron optical configuration of, schematic drawing of, 664
 - equipotential contours, objective lens, 666
 - history of, 657–658
 - intermediate/projective lens on, 666–668
 - layout of, 659
 - MTFs of, 670
 - objective lens of, 665
 - schematic drawing of, 676
 - spatial resolution of, 668–675
 - time-resolved, layout of, 688
 - transmission of, 668–675
 - uncorrected, 664–671
 - X-ray sources for, 658–667
- Photoluminescence (PL), 1120
- Photomultiplier tube (PMT), 755
- Photon maps, atomic scale detail in, 1122
- Physisorption/DLVO force, for AFM
 - macromolecular samples, 1054
- Physisorption/hydrophobic/hydrophilic interaction, AFM and, 1055
- Piezoelectric force microscopy (PFM), 945
 - contrast mechanism maps of, 957
 - PZT capacitor and, 955
 - working principle of, 953
- Piezoresponse (PR), 954
- PL. *See* Photoluminescence
- Planar zone plate
 - contour plot of, 853
 - Seidel aberration of, 851
- Planar zone plates, contour efficiency plot of, 854
- Plasma display cell, SPEM of, 903
- Plasma energy, variation of, 395
- Plasmon, 100
- Plunge freezing instrumentation, Cryo-ET and, 550
- Plunge freezing process steps, Cryo-ET and, 551
- PMMA. *See* Polymethyl methacrylate
- PMT. *See* Photomultiplier tube
- POA. *See* Phase object approximation
- POF. *See* Phase-only filtering
- Point defects formation/clustering, electron beam-induced phenomena and, 511–513
- Point image, simulation experiment of, 1240
- Point spread function (PSF), 412, 759, 791
 - STED and, 820
 - variations in
 - with focusing depth, 760
 - with mismatch conditions, 760
- Polycrystalline Au/*n*-Si(001), BEEM on, 1009
- Polycrystalline materials
 - crystallographic characterization of, 134
 - deformation in, 487–490
- Polycrystalline sample, cross section of, 193
- Polycrystalline ZnO, transport imaging in, 948
- Polyfin-polycarbonate polymer composite, section of, 369
- Polymers, X-ray transmission spectromicroscopy of, 899
- Polymethyl methacrylate (PMMA), secondary electron micrograph of, 208
- Poly-N-isopropylacrylamide (PNIPA Am), secondary electron micrograph of, 235
- Polystyrene spheres, interferometric TXM image of, 899
- Polyvinyl methyl ether (PVME), high-resolution micrographs, 221
- Pore protomers, 1063
- Position measurement, with CRLB, 1245–1246

- Postcolumn imaging filter, 306
- Power series expansions
 - of electrostatic potential, 729–730
 - of vector potential, 730
- PR. *See* Piezoresponse
- Precipitate growth mechanism, 460
- Primary electrons, 141
- Principle of reciprocity, 80
- Prism spectrometer system, 301
- Probe electron density, real space intensity
 - distribution of, 315
- Probe function, 71
- Probe intensity, 103
- Probe profile plots, with chromatic defocus
 - spreads, 122
- Probing buried interfaces I, 1007–1012
- Probing buried interfaces II, 1012–1013
- Projected charge density (PCD), HRTEM and, 12
- Projection-slice theorem, 579
- Propagation-based phase contrast, 884
- Protein
 - features, imaging radiation dose for, 840
 - surface layer, secondary electron micrograph of, 216
 - Zernike phase contrast in, intrinsic amplitude and, 876
- PSF. *See* Point spread function
- Pulmonary surfactant, AFM and, 1060–1061
- Pulse compression, DTEM and, 435–436
- Pulse sequence, in laser-synchrotron, 687
- Pulsed circuit, waveguide image and, 688
- Punkt1, 248
- PVME. *See* Polyvinyl methyl ether
- PZT capacitor, PFM and, 955

- Quadrupole, 700–701
- Quadrupole corrector, schematic drawing of, 712
- Quadrupole-octopole correctors, spherical
 - aberration correctors and, 712–715
- Quantitative measurements, in electron
 - holography, 1171–1172
- Quantum efficiencies, of less sensitive metals, 420
- Quantum size contrast, of Fe film, 629
- Quantum size effects, 1012–1013

- Radial artery, secondary electron micrograph of, 207
- Radiation damage, SEM and, 201–203
- Radiation sensitivity, Cryo-ET and, 554–557
- Radiation-induced dislocation glide, 515
- Radiation-induced dislocation motion, 514
- Radiolarian, stereopair of, 161
- Raft proteins, diffusion coefficients for, 1064
- Ray aberrations, zone plate image quality and, 847
- Rayleigh definition, of resolution, 353
- Rayleigh resolution, 1229–1230
 - diffraction limit and, 1231–1234
- RC. *See* Ring cluster
- Real space intensity distribution, of probe electron
 - density, 315
- Real space intensity plots, of electron intensity
 - dispersion, 316
- Real structures, zone plate efficiency and, 853
- Rec-DNA complex, STM tomographies of, 1026
- Reciprocal inelastic mean free path, 612
- Reciprocity, partial coherence and, 882
- Reconstructed modulus, of exit-plane wavefunction, 49
- Reconstruction, Cryo-ET and, 577–582
- Reference image plane, 25
- Regions/energy ranges, for ELNES, 384
- Relative defocus, PCF and, 23–24
- Relativistic beams, DTEM and, 436
- Resistivity correlation, with structure, 502
- RESOLFT. *See* Reversible saturable optically linear
 - fluorescence transition
- Resolution, 1228–1261
 - calculations for, parameter definitions for, 350
 - coherence and, 883–884
 - for CXDI, 1221–1223
 - deterministic model-based, 1236–1240
 - discussion of, 1260–1262
 - electron holography in, 1171–1172
 - energy dependence and, 621
 - FESEM and, 212–214
 - introduction to, 1228–1229
 - limits, 791–795
 - acceptance angle and, 621
 - HRTEM of, 13–17
 - of STEM, 117–125
 - maximum likelihood estimation and, 1248–1252
 - partial coherence and, 881
 - Rayleigh, 353, 1229–1230
 - SEM and, 182–195
 - Sparrow, 1230
 - in spatial frequency domain, 1230–1236
 - statistical experimental design and, 1252–1258
 - statistical model-based, 1240–1258
 - three-dimensional optical sectioning and, 761–763
 - two-point, 1229–1230
 - ultimate model-based, 1258–1259
 - variation, electron beam thickness and, 352
- Resolution-determining zone plates, 855–856
- Resolution/transmission, of aberration corrected
 - PEEM microscope, 677–678
- Reversible saturable optically linear fluorescence
 - transition (RESOLFT), 795
 - large saturation factors, at low powers, 826–827
 - principle of, 813
- Ribosome(s), mapping of, 585
- 70S ribosome, averaged structure of, 588
- Ring cluster (RC), 637
- Road lines, secondary electron micrograph of, 205
- Ronchigrams
 - of Au nanoparticles, 78
 - coherent convergent beam electron diffraction and, 73–78
 - of crystalline materials, 74–76

- of noncrystalline materials, 77–78
- schematic of, 76
- Rotational ellipsoidal mirror, 152
- Rutile $\text{TiO}_2(110)$, STM on, 999
- S segregation, at Ni grain boundary, 351
- SAD. *See* Selected area diffraction
- Sample. *See also* Specimen
 - damage, with DTEM, 428–433
 - deformation, AFM and, 1034–1037
 - environment, for AFM, 1051–1052
 - preparation
 - for AFM, 1052–1058
 - with Cryo-ET, 547–549
 - for STM, 973–975
 - thickness measurement, electron holography and, 1148–1150
 - tip, energy level diagrams of, 1072
- Saturated environment, imaging in, 506–507
- SBZ. *See* Surface Brillouin zone
- Scanners
 - for AFM, 1050–1051
 - commercial, frequency response of, 1044
 - elements of, AFM natural frequencies for, 1043
- Scanning capacitance microscopy (SCM), 949
- Scanning electron acoustic microscopy (SEAM), 182
- Scanning electron microscope (SEM), 65, 133–257
 - abbreviations with, 135–137
 - AE and, 181
 - applications for, 203
 - attached equipment for, 155–157
 - CL and, 177
 - of Co dots, 1166
 - contamination and, 201–203
 - contrast formation and, 182–195
 - conventional, 139–208
 - crystal orientation contrast and, 193
 - digital imaging recording with, 157–158
 - EBIC and, 192–193
 - electron guns and, 143–145
 - electron lenses and, 145–148
 - electron probe current in, 147
 - at elevated pressure, 237–244
 - FIB and, 196, 197, 200
 - focus depth of, 142
 - of gold balls, 1213
 - life science applications of, 203
 - magnetic contrast, 194–195
 - magnification calibration for, 162–163
 - material contrast and, 189–191
 - microanalysis in, 246–252
 - radiation damage and, 201–203
 - resolution and, 182–195
 - SEs and, 169–172
 - in situ* treatments for, 203
 - special topics on, 157–163
 - specimen applications for, 201
 - specimen preparation for, 195–201
 - specimen stages for, 155–157
 - specimen tilting with, 158–162
 - stereo imaging with, 158–162
 - symbols with, 137–139
 - tilt compensation/dynamic focusing in, 159, 160
 - topographic contrast and, 183–189
 - voltage contrast and, 191–192
 - X-rays and, 178–180
- Scanning fluorescence X-ray microscope (SFXM), layout of, 873–875
- Scanning force microscopy (SFM), 933
- Scanning gate microscopy (SGM), 950
- Scanning impedance microscopy (SIM), 950
- Scanning low-energy electron microscopy (SLEEM), 221–222
- Scanning Microscope Analysis and Resolution Testing (SMART), 183–189
- Scanning near-field optical microscope (SNOM), 1031
- Scanning photoemission microscope (SPEM), of plasma display cell, 903
- Scanning probe microscopy (SPM)
 - advanced techniques, 944–958
 - for dielectric properties, 953–958
 - for transport properties, 945–953
 - future trends in, 958–960
 - generalized approach to, 960
 - imaging properties of, 944–958
 - in materials science, 929–960
 - of oxides, 939, 940
 - variations of, 946
- Scanning surface potential microscopy (SSPM), 945
 - of SAM, 947
 - working principles for, 946
- Scanning transmission electron microscope (STEM), 65–127
 - aberration correction and, 123–125
 - chromatic aberration and, 118–119
 - column, imaging-to-diffraction mode in, 110–112
 - as dedicated instrument, 110–112
 - electron optical configuration for, 292
 - electron sources for, 114–117
 - finite source size and, 118–119
 - images, phase component maps and, 368
 - introduction to, 65–69
 - operating principles of, 65–68
 - photograph of, 67
 - probe, 69–73
 - diffraction-limited, 72
 - resolution limits of, 117–125
 - schematic of, 66
 - spectrum, schematic description of, 366
 - spectrum imaging in, 104–107
 - spectrum imaging technique, application of, 367
 - spherical aberration of, geometric optics view of, 69, 70
 - statistical experimental design of, 1254–1256
 - TEM and, 175
 - X-ray analysis, 107–109
- Scanning transmission electron microscope/
 - Electron energy-loss spectroscopy STEM-EELS, mapping of, 365–366

- Scanning transmission X-ray microscope (STXM)
 cosine amplitude grating imaging with, 881
 layout of, 870–873
 MTF for, 878
 schematic drawing of, 863
- Scanning tunneling microscopy (STM), 494, 495, 915, 929
 atom/molecule manipulation and, 1003–1006
 bias-dependent, imaging and, 984–986
 capturing dynamic surface processes and, 998–1003
 elastic vacuum tunneling and, 970–972
 energy-filtered, 985
 Ge imaging with, 979
 at high/low temperatures, 996–1006
 motivation/issues with, 997
 image simulation, 1016–1019
 imaging applications for, 975–976
 imaging principles of, 969–970
 instrumentation for, 975–976
 introduction to, 969–970, 1070–1073
 large-area atom-resolved, 981
 with LHe continuous-flow cryostat, 1074–1076
 molecular vibrational spectra with, 1117
 movies/atom tracking, 998–1003
 operating environment for, 973–975
 of oxygen/CO, 1109
 of Rec-DNA complex, 1026
 on rutile $\text{TiO}_2(110)$, 999
 samples for, 973–975
 silicon surface imaging with, 977–982
 simultaneous mass spectrometry and, in catalytic flow reactor, 983
 surface manipulation, 497
 in surface science, 969–1019
 of SWCNT end, 1094
 tip, bond formation with, 1113
 tip trajectory for, 972
 tips for, 973–975
 variable-temperature, 1076
- Scanning tunneling microscopy (STM)-induced photo emission, LT-STM and, 1119–1125
- Scattered electrons, for vitreous ice, 567
- Schematic current v. voltage curves, 1115
- Schottky emission cathode (SEC), 209
- Schottky field-emission gun, 117
- Scintillation detector, 149
- SCM. *See* Scanning capacitance microscopy
- SDD. *See* Silicon drift detector
- SE. *See* Secondary electron(s)
- SEAM. *See* Scanning electron acoustic microscopy
- SEC. *See* Schottky emission cathode
- Secondary electron(s) (SE), 109
 ADF micrographs of, 176
 detector, 68
 generation, schematic drawing of, 168
 micrographs, of commercial cross-grating replica, 163
 SEM and, 169–172
 yield
 BSE yield and, 171
 schematic drawing of, 170
- Secondary electron micrograph
 of austenite, 256
 of birch pollen, 217
 of cross-fractured semiconductor structure, 227
 of DPPTe, 231
 with ESEM, 242
 with ESEM-E3, 243
 of fern fossil, 205
 of FIB color film, 200
 of FIB tyre, 200
 of glass knife, 234
 of glass micropipette, 233
 of keratinocyte, 236
 of leaf cuticula, 237
 of phospholipid/protein film, 229, 230
 of PMMA, 208
 of PNIPa Am, 235
 of protein surface layer, 216
 of radial artery, 207
 of road lines, 205
 of silicon calibration standard, 228
 of starch glycolys D, 244
 of steel ball, 184, 185
 SuperSharpSilicon AFM probe silicon cantilevers of, 229
 of tartar, 186
 of thiol monolayer, 230
 of zeolite FAU, 219
- Second-difference technique, 382
- Segregated B-doped Si(100), images of, 634
- Seidel aberration, of planar zone plate, 851
- Selected area diffraction (SAD), 110, 111
 aperture, 110
- Self-assembled Co nanoparticle, low magnification
 bright-field image, 1159
- Self-assembled monolayers (SAM), SSPM of, 947
- Self-assembled quantum dots
 BEEM on, 1011
 cross-sectional STM on, 1014
- SEM. *See* Scanning electron microscope
- Semiconductor(s)
 dopant potentials in, electrostatic field measurement and, 1173–1182
 FM-AFM of, 940–942
 thin films on, 635–638
- Semiconductor detector, 151
- Semiconductor device structure, electrostatic potential distribution in, 1175
- Semiconductor emitters, photoelectron guns and, 416–417
- Sextupoles
 arrangement, 704
 correctors
 aberration correctors and, 715–716
 schematic drawing of, 716
 spherical aberration correctors and, 715–716
 Wien filters with, 724

- SFM. *See* Scanning force microscopy
- SFXM. *See* Scanning fluorescence X-ray microscope
- SGM. *See* Scanning gate microscopy
- Shaped grooves, zone plates with, 859
- Short/long range forces, distance dependencies of, 932
- Si, surfaces, diffraction contrast from, 625
- Si(001)
 constant-current STM images of, 983
 surface, $\text{Si}_{0.86}\text{Ge}_{0.14}$ pseudomorphically strained film on, 902
- Si(100)
 homoepitaxial growth of, 633
 LEEM and, 632–634
 NC-AFM of, 942
- Si(110), energy-filtered electron diffraction pattern of, 307
- Si(111)
 Al films on, 636
 ball-and-stick model of, 977
 CITS on, 989
 FM-AFM of, 940–942
 NC-AFM image of, 936
 Pb growth on, 636
 surface, with LEEM, 631–632
- Si(112), annular dark-field detector scanning transmission microscope image of, 125
- $\text{Si}_{0.86}\text{Ge}_{0.14}$ pseudomorphically strained film, on
 Si(001) surface, 902
- $\text{Si}_{1-x}\text{Ge}_x/\text{Si}(001)$ heteropitaxy, lattice mismatch strain during, 982
- Side entry cryo-holder, for Cryo-ET, 562
- SiGe heterostructures, dislocation dynamics in, 494
- Signal generation, electron-specimen interaction and, 163–182
- Signal processing, 297
- Signals for thin sample, schematic drawing of, 141
- Signal-to-noise ratio (SNR), 411
- Si(100)/gold(100), 1s state for, 1255
- Silicide formation, solid state diffusion reaction and, 455–456
- Silicide phase transformations, 457
- Silicon calibration standard, secondary electron micrograph of, 228
- Silicon drift detector (SDD), 248
- Silicon nitride cantilever, frequency response of, 1044
- Silicon oxidation, 470
- Silicon surface, STM imaging of, 977–982
- Silicon-on-insulator (SOI), 634
- SIM. *See* Scanning impedance microscopy
- SIMION simulations, of angle/energy distributions, of emitted electrons, 669
- Simultaneous mass spectrometry, STM imaging and, in catalytic flow reactor, 1001
- Simultaneous structure analysis, for DWNT, 504
- Single active filament, 1033
- Single atom, molecule manipulation and, LT-STM and, 1105–1114
- Single axis tilt series data acquisition scheme, 540
- Single crystals, deformation in, 487–490
- Single lens, wavefront from, 792
- Single molecule vibrational spectroscopy/microscopy, 994
- Single particle approach, 540–543
 Cryo-ET with, 547
 of giant protein complex TPP II, 541
- Single projection image, of frozen hydrated yeast, 895
- Single-walled carbon nanotubes (SWCNTs), 1093
 electron confinement in, 1093
 end, STM image of, 1094
- Sintering, phase transformations and, in free-standing nanoparticles, 461–463
- Si(Li) X-ray diode, schematic drawing of, 247
- Size-dependent transformations, in embedded nanostructures, 458–460
- SLEEM. *See* Scanning low-energy electron microscopy
- Small elements, magnetization of, 480
- Small magnetic elements, 478–480
- SMART. *See* Scanning Microscope Analysis and Resolution Testing; Spectromicroscope for All Relevant Techniques
- Sn into Bi alloying, 462
- SNOM. *See* Scanning near-field optical microscope
- SNR. *See* Signal-to-noise ratio
- Soft X-ray(s) zone plates
 aberration size of, 849
 diffraction pattern, tomographic reconstruction from, 1217
 of square-wave test objects, 856
 tomography, 885
 transmission diffraction pattern
 from gold balls, 1212
 from lithographed letters, 1211
- SOI. *See* Silicon-on-insulator
- Solid state diffusion reaction, silicide formation and, 455–456
- Solid-liquid interface, phase transformations and, 450–453
- Solid-to-liquid transformation, phase transformations and, 450–453
- Source coherence, for sample illumination, 289
- Space charge layers, at grain boundaries, in electron holography, 1182
- Sparrow resolution, 1230
 diffraction limit and, 1231–1234
- Spatial dependence, of differential conductance, 1083
- Spatial frequency domain, resolution in, 1230–1236
- Spatial localization
 of EELS signals, 101–104
 of inelastic imaging, 101–104
- Spatial resolution
 in EFTEM elemental mapping, 371–374
 history of, 671
 of PEEM, 668–675
- Spatial/temporal resolution, classification by, 407

- Specimen. *See also* Sample
 preparation, for SEM, 195–201
 schematic drawing of, 196
 rotation, around tilt axis, geometric model of, 572
 stages
 FESEM and, 212
 for HRTEM, 34–35
 for SEM, 155–157
 surface profile, schematic drawing of, 186
 tilting
 schematic drawing of, 1156
 with SEM, 158–162
- Spectra quantification, for correction factors, 345
- Spectra visibility, variation of, with sample thickness, 347
- Spectrometers, EELS and, 301–307
- Spectromicroscope for All Relevant Techniques (SMART), schematic drawing of, 618
- Spectroscopic photoemission and low-energy electron microscopy (SPELEEM), 606
 operation modes of, 623
- Spectroscopy techniques, comparison of, 282
- Spectrum imaging
 of data cube, 104
 for Gd, 105
 in STEM, 104–107
- Spectrum imaging technique, for EFTEM, 364
- SPELEEM. *See* Spectroscopic photoemission and low-energy electron microscopy
- SPeM. *See* Scanning photoemission microscope
- Spherical aberrations, 146
 aplanatic corrector of, 720
 correctors
 quadrupole-octopole correctors and, 712–715
 sextupole correctors and, 715–716
 DECO of, 721
 and HRTEM, 32
 lower bound of standard deviation for, 1256
 zone plate image quality and, 847–849
- Sphingolipid-cholesterol rafts, AFM and, 1062–1064
- Spin manipulator, schematic drawing of, 643
- Spin reorientation transition, in Fe-Co alloy layer, 646
- Spin-polarized low-energy electron microscopy (SPLEEM), 605–647, 642–647
 Co(0001) surface image of, 644
 instrumentation for, 614–619
- Spin-polarized scanning tunneling microscopy (SP-STM)
 atomic resolution magnetic imaging with, 1127
 on Gd(0001), 992
 LT-STM and, 1125–1132
- SPLEEM. *See* Spin-polarized low-energy electron microscopy
- SPM. *See* Scanning probe microscopy
- SP-STM. *See* Spin-polarized scanning tunneling microscopy
- Square-wave test objects, soft x-ray images of, 856
- SrTiO₃, NC-AFM of, 939–940
- SrTiO₃ (100), KFM of Sigma5 grain boundary in, 931
- SSPM. *See* Scanning surface potential microscopy
- Standing wave patterns, on copper, 1083
- Starch glycolys D, secondary electron micrograph of, 244
- Stark effect, LT-STM and, 1100–1101
- Statistical experimental design
 of HRTEM, 1256–1258
 resolution and, 1252–1258
 of STEM, 1254–1256
- Statistical model-based resolution, 1240–1258
 CRLB and, 1244–1248
 parametric statistical models of observation with, 1242–1244
- STED. *See* Stimulated emission depletion
- Steel, nanoindentation of, 496
- Steel ball
 backscattered electron micrograph of, 155
 BSE micrograph of, 185
 BSE micrographs, 190
- STEM. *See* Scanning transmission electron microscope
- Step contrast, for phase differences, 629
- Stereo imaging, with SEM, 158–162
- Stereopair, of radiolarian, 161
- Stimulated emission depletion (STED), 818–826
 fluorescence focal spot and, 822
 images with, 819
 PSF and, 820
 subdiffraction immunofluorescence imaging with, 825
- STM. *See* Scanning tunneling microscopy
- Structural phase transitions, for *in situ* transmission electron microscopy, 453–457
- Structure correlation, with resistivity, 502
- Structure determination, using ADF images, 94–95
- STXM. *See* Scanning transmission X-ray microscope
- Stylus apex/sample, forces between, 1037–1040
- Stylus geometry, AFM and, 1033
- Subdiffraction immunofluorescence imaging, with STED, 825
- Superalloy NIMONIC, surface of, 220
- Superconducting materials, *in situ* TEM, 481–483
- Superconductivity, LT-STM and, 1129–1132
- Superconductors, vortex motion in, 483
- Superresolution, 1234–1236
 diffraction limit and, 1230–1236
- SuperSharpSilicon AFM probe silicon cantilevers, secondary electron micrograph of, 229
- Surface Brillouin zone (SBZ), 1088
- Surface imaging, 134
- Surface manipulation, STM and, 497
- Surface mobility, nanostructures from, 466
- Surface reactions, 513–514
 crystal growth and, 463–476
- Surface science, STM in, 969–1019
- Surface state dispersion, direct visualization of, 1090
- Surface state electrons, hot, experimental lifetime of, 1099

- Surface state lifetime measurement, 1097
- Surface states, on Ag(111), 1086
- Surface step structure growth, nucleation and, 631
- Surface structure
 - measurement/modification of, 464–466
 - mechanical properties of, 495–500
- Surface structure identification, of Ge(105), 1019
- Surface topography, of cross-sectioned diode, 951
- SWCNTs. *See* Single-walled carbon nanotubes
- Symmetric unipotential lens, electrode geometry of, 667
- Tapping mode, latex-beads imaged in, 1048
- Tartar, BSE micrograph of, 186
- TCC. *See* Transmission cross coefficient
- TDS. *See* Thermal diffuse scattering
- TEM. *See* Transmission electron microscope
- Tetrode electron mirror, aberration correction by, 676
- Tetrode mirror
 - aberration coefficients of, 726
 - section of, 726
 - within system, 727
- TEY. *See* Total electron yield
- TFE. *See* Thermal field emitter
- Thermal diffuse scattering (TDS), 84
 - cross sections of, 92
- Thermal field emitter (TFE), 209
- Thermionic emission triode gun, schematic drawing, 143
- Thermionic gun, 134
- Thick zone plates, 861–862
- Thin films
 - magnetic materials, 477–478
 - mechanical properties of, 495–500
 - phase transformations in, 458–463
 - semiconductors on, 635–638
 - in situ* indentation/straining of, 495–496
- Thiol monolayer, secondary electron micrograph of, 230
- Three-dimensional cryoelectron microscopy, 538–547
- Three-dimensional imaging, principles of, 538–540
- Three-dimensional optical sectioning, resolution and, 761–763
- Three-widow background-subtracted elemental maps, 1163
 - magnetic phase contours of, 1164
- Through-focus deconvolution, depth-of-focus limit and, in x-ray microscopes, 887
- “Through-the-lens” detection, 154
- Tilt axis, 160
 - specimen rotation of, geometric model of, 572
- Tilt compensation/dynamic focusing, in SEM, 159, 160
- Tilt/defocus dataset, schematic representation of, 26
- Tilt-induced displacements, HRTEM and, 21
- Tilting geometry, with Cryo-ET, 557–560
- Time resolution, for DTEM, 421–423
- Time-resolved microscopy, 686–690
 - experimental setup for, 687
 - introduction to, 686–687
 - in situ* measurement, of field pulse, 687–688
 - vortex dynamics and, 688–690
- TiO₂, NC-AFM of, 939–940
- TiO₂(100), NC-AFM of, 941
- TiO₂(110), oxygen vacancies on, of O₂-mediated diffusion, 999
- Tip(s)
 - geometries, AFM and, 1032
 - image interpretation and, 937–939
 - modeling of, 937
 - “nanotip” models and, 938–939
 - for STM, 973–975
 - trajectory, for STM, 972
- Tip-sample interaction, AFM and, 1033–1040
- TIRF. *See* Total internal reflection fluorescent
- Tomography
 - of integrated circuit, 901
 - principle of, 1239
 - in X-ray microscopes, 884–888
 - depth-of-focus limit, 885–886
 - operation principles for, 884–885
- Topography, contrast
 - ETD and, 184
 - SEM and, 183–189
- Total electron yield (TEY), 679
- Total inelastic cross sections, atomic number and, 318
- Total internal reflection fluorescent (TIRF), 1049
- TPE. *See* Two-photon excitation
- Transfer functions
 - comparison of, 48
 - of EM, 1233
- Transition metals, 636
- Transmission, of PEEM, 668–675
- Transmission cross coefficient (TCC)
 - HRTEM and, 27–28
- Transmission electron microscope (TEM), 65, 175
 - AEM and, 273
 - photo of, 274, 538
 - speed of, 406
 - STEM and, 175
- Transmission X-ray microscope (TXM), 863
 - contrast in, 873–875
 - incoherent illumination with, schematic drawing of, 877
 - layout of, 867–868
 - phase contrast layout of, 868–870
 - schematic drawing of, 863
 - Zernike phase contrast, schematic drawing of, 869
- Transmitted electrons, 175–177
 - scattered, 176
- Transparency function, of one period, of zone plate, 851
- Transparent support film, 175

- Transport imaging, in polycrystalline ZnO, 948
- Transport properties, advanced SPM techniques of, 945–953
- Triangular FeCo structure, vortex dynamics in, 689
- Tribology, nanomanipulation and, 497–498
- TU Berlin, DTEM, 409
- I(V) tunneling spectroscopy, on C₆₀/C-nanotube, 989–993
- Two neighboring points image, computer-simulated experiment of, 1241, 1246
- Two-dimensional angular/energy distribution, of GOS, 322
- Two-photon cross sections, for fluorescent molecules, 767
- Two-photon excitation (TPE), 751
 - analysis of, 763–765
 - of fluorescent molecules and, confocal principles, 754–763
 - fluorescent molecules under, 765–767
 - laser emission time scale for, 773
 - laser sources for, 773
 - optical consequences of, 767–769
 - optical setup for, 769–774
 - photo of, 771
 - schematic drawing of, 770
 - remarks/comments on, 759–761
 - theoretical analysis of, 757–759
 - v. conventional excitation, 756
- Two-photon excitation (TPE) fluorescence microscopy, 751–777
 - chronology of, 753–754
 - introduction to, 751–753
- Two-point resolution, 1229–1230
- TXM. *See* Transmission X-ray microscope
- Tyre, FIB of, secondary electron micrograph of, 200

- UED. *See* Ultrafast electron diffraction
- UHV. *See* Ultrahigh vacuum
- UHV SEM. *See* Ultrahigh vacuum scanning electron microscopy
- UHV-STM. *See* Ultrahigh vacuum scanning tunneling microscopy
- Ultimate model-based resolution, 1258–1259
 - observations on, 1258–1259
 - resolution consequences of, 1259
- Ultracorrector, 729
- Ultrafast electron diffraction (UED), 406
 - DTEM timeline development with, 434
- Ultrahigh vacuum (UHV), 31, 135
- Ultrahigh vacuum scanning electron microscopy (UHV SEM), 245–246
- Ultrahigh vacuum scanning tunneling microscopy (UHV-STM), 1073
- Ultrathin continuous metal layer-coated specimen, schematic cross section of, 214
- Ultrathin windows (UTW), 295
- Ultraviolet light-excited photoemission electron microscopy (UVPEEM), 642
- Uncorrected photoemission electron microscopy (PEEM), aberration/resolution on, 673

- UTW. *See* Ultrathin windows
- UVPEEM. *See* Ultraviolet light-excited photoemission electron microscopy

- Vacuum tunneling, band diagrams for, 971
- Validity domains of approximation, diffractive imaging and, 1201–1205
- Van Aken low-energy foil corrector, schematic drawing of, 718
- Van Cittert-Zirnicke theorem, 88
- Variable pressure secondary electron (VPSE) detector, 240
- Variable-temperature STM, 1077
 - with continuous-flow cryostat, 1077
- Vector potential, power series expansions of, 730
- Vertical phase/amplitude hysteresis loops, of PbTiO₃, 955
- Very low voltage scanning electron microscope (VLVSEM), 221–237
 - contrast formation for, 225–226
 - detectors/detection strategies for, 225
 - selected applications for, 226–237
- Vibrational spectroscopy, inelastic tunneling and, 993–994
- Visible light, epifluorescence micrographs and, 898
- Visualization/image analysis, Cryo-ET and, 582–584
- Vitreous ice
 - microtubules in, imaging of, 568
 - scattered electrons for, 567
- VLVSEM. *See* Very low voltage scanning electron microscope
- Voltage contrast, SEM and, 191–192
- Voltage curves, v. schematic current, 1115
- Voltage step generation, across image, 413
- Vortex dynamics
 - time-resolved microscopy and, 688–690
 - in triangular FeCo structure, 689
- Vortex motion, in supercomputers, 483
- VPSE detector. *See* Variable pressure secondary electron detector

- W(110)
 - epitaxial Fe ribbon crystals on, 647
 - surface, normal incidence specular reflectivity of, 608
- Water, phase diagram of, 238
- Water microbial colony, chemical maps of, 897
- Wave aberration function
 - HRTEM and, 17–26
 - schematic diagram of, 17
- Wavefront, from single lens, 792
- Wavelength dispersive spectrometer (WDX), 134
 - features of, 249
- Wavelength-dispersive X-ray (WDX), principle of, 247
- WD. *See* Working distance
- WDX. *See* Wavelength dispersive spectrometer; Wavelength-dispersive X-ray

- Weak phase object approximation (WPOA), HRTEM and, 10
- Wide-band illumination, depth-of-focus limit and, in x-ray microscopes, 886–887
- Wide-band semiconductors, 637
- Wien filters
 - correction and, 723–724
 - with sextupoles, 724
- Working distance (WD), 140–141
- WPOA. *See* Weak phase object approximation
- XANES. *See* X-ray absorption near-edge structure
- XCF. *See* Cross correlation function
- XEDS. *See* X-ray energy dispersive spectroscopy
- Xenon atoms
 - energy dependence of, elastic mean free electron path of, 611
 - patterned array of, 1107
 - precipitates, dynamics of, 516
- XMCD. *See* X-ray magnetic circular dichroism
- XMLD. *See* X-ray magnetic linear dichroism
- XMLD-PEEM. *See* X-ray magnetic linear dichroism photoelectron emission microscopy
- XPEEM. *See* X-ray photoemission electron microscopy
- X-ray(s)
 - analysis
 - energy dispersive, 107–109
 - in STEM, 107–109
 - energy, atomic number and, 276
 - fluorescence yield of, 180
 - generation of, schematic drawing of, 277
 - interactions
 - carbon cross sections and, 836
 - zone plate X-ray microscopes and, 836–839
 - SEM and, 178–180
 - soft/hard, aberration size of, 849
 - sources, for PEEM, 658–667
 - spectrum of, 178–179
 - schematic drawing of, 179
- X-ray absorption edge, schematic drawing of, 889
- X-ray absorption near-edge structure (XANES), 101
- X-ray absorption spectra, of antiferromagnetic domains, 682
- X-ray energy dispersive spectroscopy (XEDS), 107
- X-ray magnetic circular dichroism (XMCD), 678, 904
 - of $(\text{Co}_{83}\text{Cr}_{17})_{87}\text{Pt}_{13}$ alloy, 904
 - of Co, 679
 - of Ni-Fe film, 905
 - XMLD and, 685
- X-ray magnetic linear dichroism (XMLD), 680
- XMCD and, 685
- X-ray magnetic linear dichroism photoelectron emission microscopy (XMLD-PEEM), 1126
- X-ray microanalysis, of Cr-Fe-alloy, 250, 251
- X-ray microscopy, 862–892
 - applications of, 892–906
 - background on, 862–863
 - biological applications of, 892–895
 - depth-of-focus limit and, partial volume in, 886
 - environmental science applications for, 896–897
 - history of, 873–874
 - magnetic materials applications for, 903–905
 - materials science applications of, 897–903
 - partial coherence in, 876–884
 - tomography in, 884–888
- X-ray photoelectron spectroscopy (XPS), 281
- X-ray photoemission electron microscopy (XPEEM), 606, 658–667
 - absorption and, 659–661
 - chemical contrast in, 663
 - elemental contrast in, 661–662
 - image contrast in, 661–663
 - iron oxide clusters image with, 662
 - Photoemission and, 659–661
- X-ray spectromicroscopy, 888–892
- X-ray spectrum images (XSI), 248
- X-ray tomography, soft, 885
- X-ray transmission spectromicroscopy, of polymers, 899
- X-ray zone plates, hard, 859–861
- XSI. *See* X-ray spectrum images
- X-STM. *See* Cross-sectional scanning tunneling microscopy
- Yeast cells, frozen, BSE micrographs of, 218
- Yeast, frozen hydrated, single projection image of, 895
- Z-contrast imaging, 311
- Zeolite FAU, secondary electron micrograph of, 219
- Zernike phase contrast
 - coherence in, 884
 - of integrated circuits, 900
 - intrinsic amplitude and, in protein, 876
 - TXM, 869
 - schematic drawing of, 869
- Zero-loss filtering, influence of, 570
- Zero-loss (ZL) peak, 100
 - from monochromators, 309
- ZL peak. *See* Zero-loss peak
- Zn impurity, quasiparticle density of states around, 1132
- Zone plate(s)
 - condenser, 856–858
 - efficiency
 - Fresnel zone plates and, 851–853
 - idealized structures and, 851–853
 - real structures and, 852–853
 - fabrication techniques for, 853–855
 - schematic drawing of, 855
 - hard X-ray, 859–861
 - image quality of
 - aberration size and, 849
 - astigmatism/field curvature and, 849–850
 - coma and, 850
 - Fresnel zone plates and, 846–850
 - optical path function analysis and, 846–847

Zone plate(s) (cont.)

- ray aberrations and, 847
- spherical aberration, 847–849
- orders number of, 845
- resolution-determining zone plates and, 855–856
- with shaped grooves, 859
- thick, 861–862
- transparency function, of one period, 851
- types of, efficiency of, 852
- zone of, 844

Zone plate microscopes, table of, 864–865

Zone plate X-ray microscopes

- background on, 835–836
 - focusing optics of, 840–843
 - Fresnel zone plates and, 844–862
 - introduction to, 835–846
 - overview/recent trends in, 842–843
 - principles/applications of, 836–906
 - X-ray interactions and, 836–839
- Zr on Ta cathode, image of, 419

Figure 1–14. Illustration of certain lens aberrations. (a) A perfect lens focuses a point source to a single image point. (b) Spherical aberration causes rays at higher angles to be over-focused. (c) Chromatic aberration causes rays at different energies (indicated by color) to be focused differently.

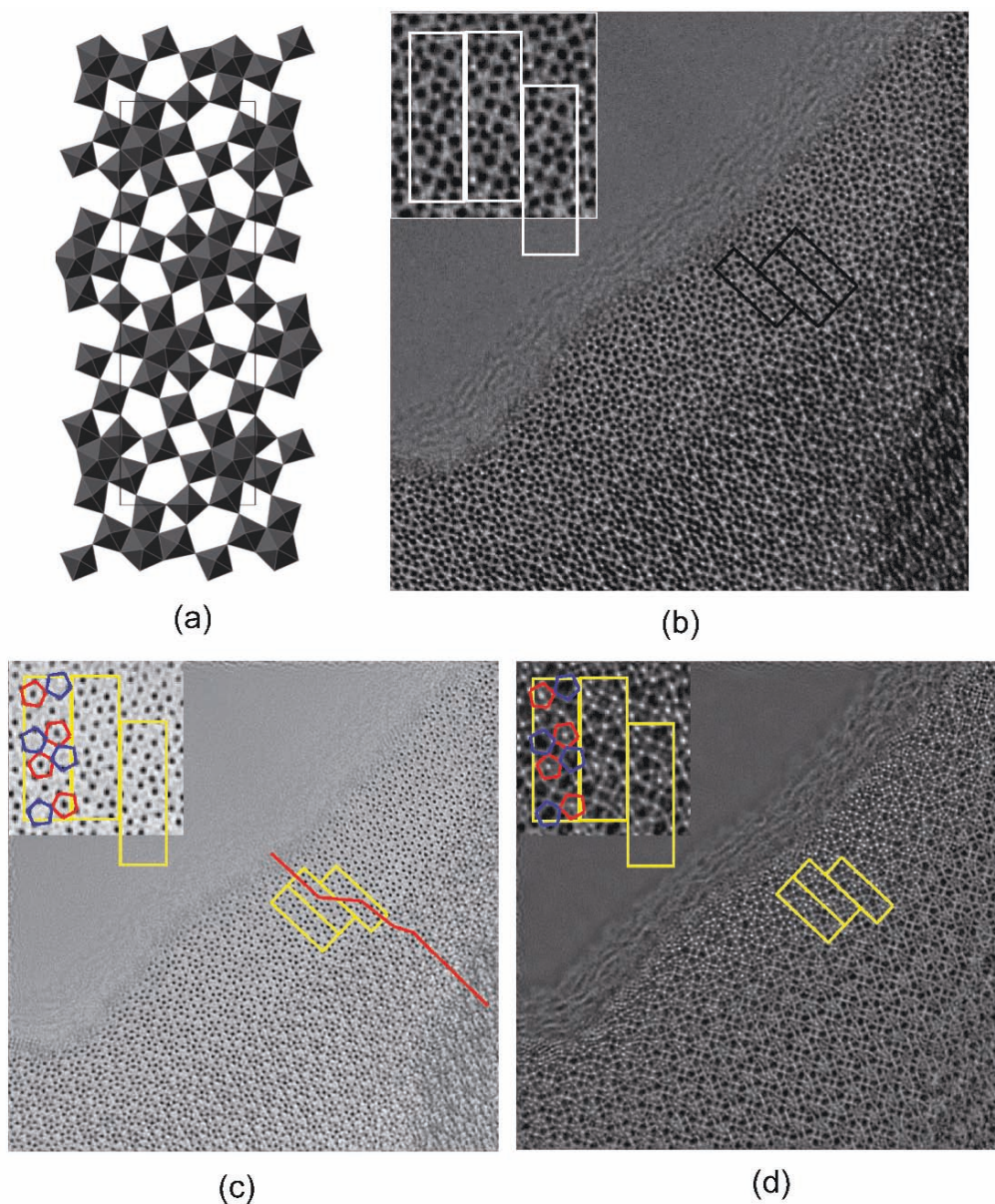
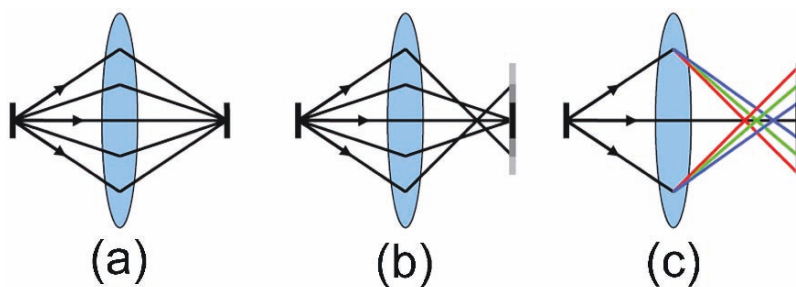


Figure 1–18. (a) Structural model of the complex oxide $\text{Nb}_{16}\text{W}_{18}\text{O}_{94}$ projected along [001]. (b) Conventional axial HRTEM image recorded at the Scherzer defocus of a thin crystal. (c) Reconstructed modulus of the exit-plane wavefunction of $\text{Nb}_{16}\text{W}_{18}\text{O}_{94}$ with the marked area enlarged (inset), which directly shows the cation positions (black) with improved resolution compared to the axial image. The line indicates a stacking fault with a shift of a third of a unit cell along [010]. (d) Reconstructed phase of the exit-plane wavefunction with the marked area enlarged (inset). The cation sites in the phase are recovered with positive (white) contrast and additional weak between the cation atomic columns which indicate the positions of the oxygen anions are also resolved. The reconstructed phase and modulus are shown at the same scale.

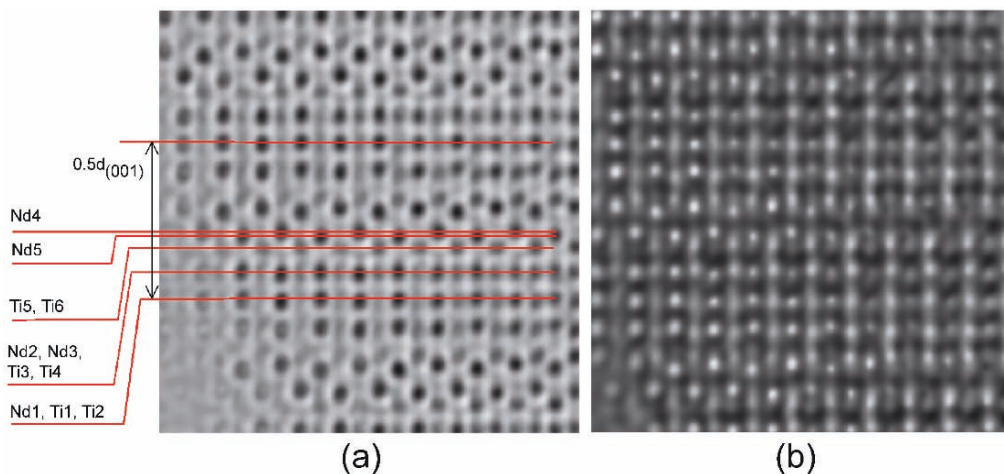


Figure 1-20. (a) Enlarged region taken from the reconstructed modulus calculated from a tilt azimuth dataset of an $\text{Nd}_4\text{SrTi}_5\text{O}_{17}$ crystal edge in the $[010]$ projection showing a small difference in positions of the Nd(4) and Nd(5) cations at the interface between adjacent perovskite slabs. (b) Reconstructed phase of the same region as (a) showing details of the oxygen anion sublattice between the Ti sites. The experimental data were recorded using the tilt series geometry described in the text with a JEOL JEM-3000F FEGTEM, 300kV, $C_3 = 0.57$ mm with an injected tilt of 1.9 mrad.

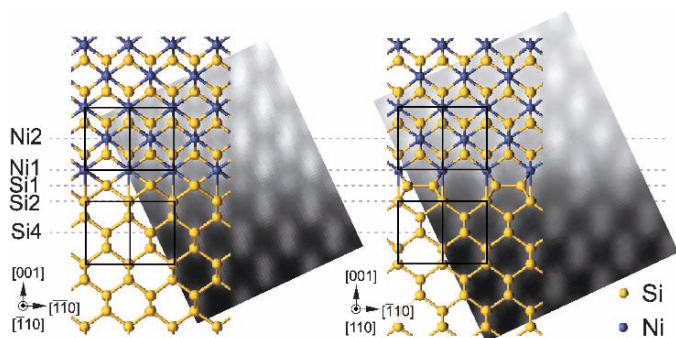


Figure 2-14. An ADF image of an $\text{NiS}_2/\text{Si}(001)$ interface with the structure determined from the image overlaid. [Reprinted with permission from Falke et al. (2004). Copyright (2004) by the American Physical Society.]

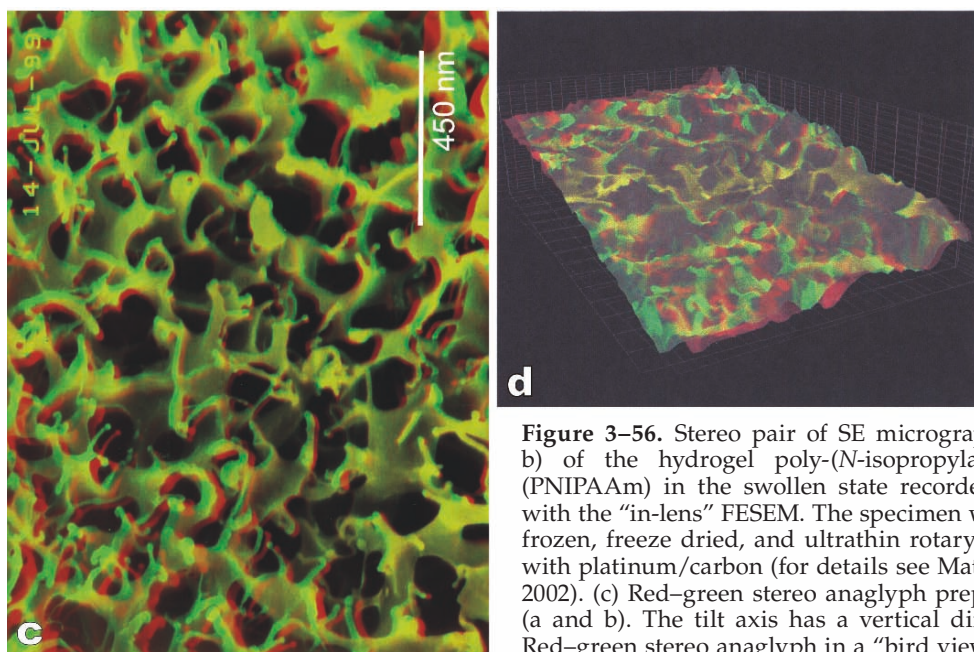


Figure 3-56. Stereo pair of SE micrographs (a and b) of the hydrogel poly-(*N*-isopropylacrylamide) (PNIPAAm) in the swollen state recorded at 2 keV with the "in-lens" FESEM. The specimen was rapidly frozen, freeze dried, and ultrathin rotary shadowed with platinum/carbon (for details see Matzelle et al., 2002). (c) Red-green stereo anaglyph prepared from (a and b). The tilt axis has a vertical direction. (d) Red-green stereo anaglyph in a "bird view."

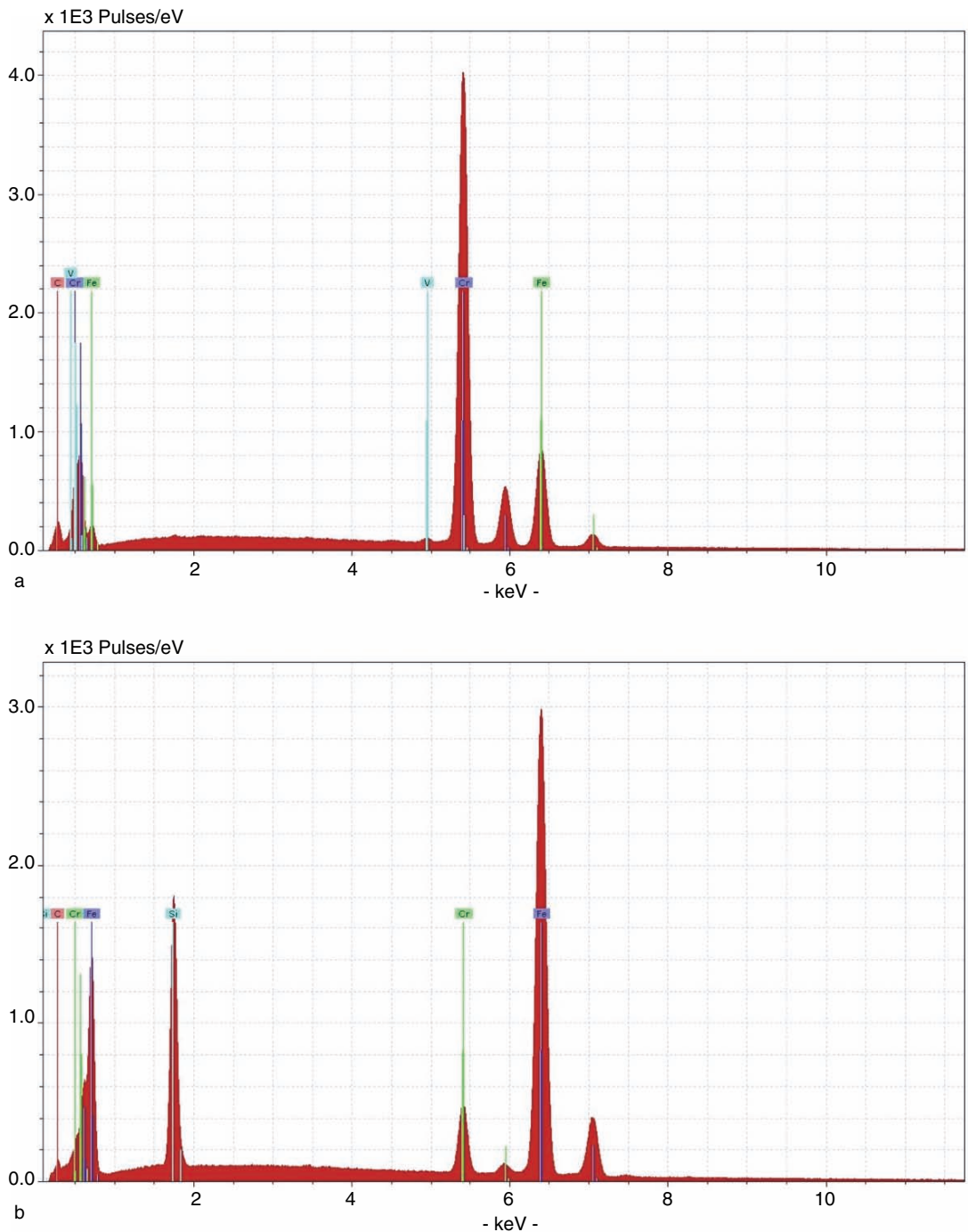


Figure 3–66. X-Ray microanalysis of a Cr–Fe-alloy with a Si phase. The EDX spectra (a and b) were recorded with the Röntec XFlah3001 from locations “Punkt1” and “Punkt2” marked in the SE micrograph of the specimen (c). The positions of the characteristic X-ray energies for the various elements emerging in the spectra are indicated by thin colored lines, which are labeled with the chemical symbol of the corresponding chemical element. The elements iron, chromium, and vanadium occur with one K_{α} peak each in the energy range from 4.95 to 6.40 keV and with one less intense L_{α} peak each in the energy range from 0.51 to 0.71 keV. Elemental distribution maps of Fe (d), Cr (e), Si (f), and Ti (g) were recorded using the K_{α} lines. (h) Mixed micrograph obtained by superimposition of the SE image and the maps of the distribution of Fe, Cr, Si, and Ti within the field of view. Experimental conditions: SEM, LEO 438VP. For recording spectra: $E_0 = 20$ keV; count rate, $\approx 3 \times 10^3$ cps; acquisition time, 300 s. For recording maps: $E_0 = 25$ keV; count rate $\approx 1.5 \times 10^5$ cps; acquisition time 600 s. (EDX spectra, SE micrograph, and elemental distribution maps were kindly provided by Röntec GmbH, Berlin, Germany.)

(continued)

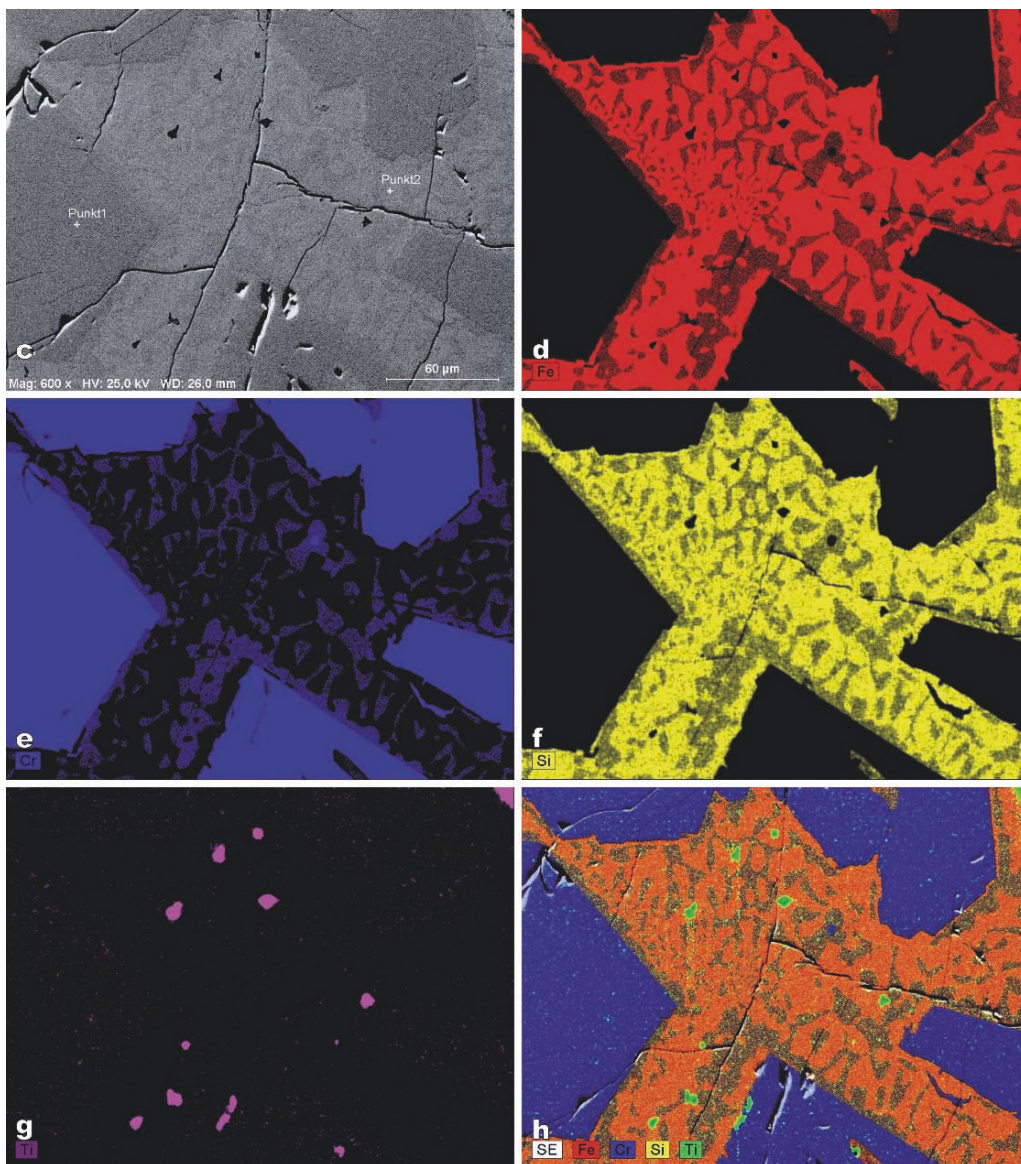


Figure 3-66. *Continued*

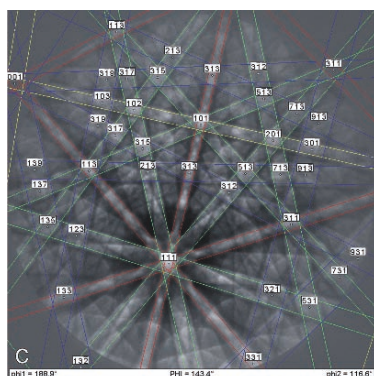


Figure 3-68. EBSD patterns from an as-cast niobium specimen. (c) EBSD pattern from (a) with colored pairs of Kikuchi lines generated by automatic indexing. [EBSD patterns were kindly provided by Dr. S. Zaefferer, Max-Planck-Institut für Eisenforschung, Düsseldorf, Germany. (a and b) From Zaefferer, 2004; with kind permission from JEOL (Germany) GmbH, München, Germany.

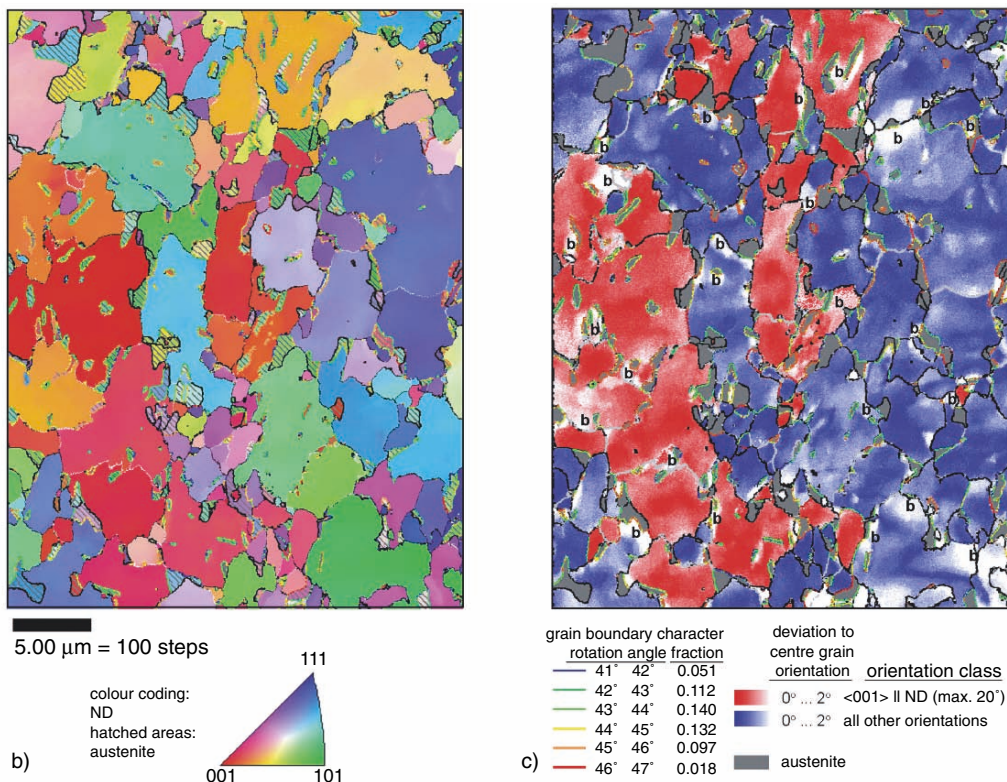


Figure 3–69. Secondary electron micrograph of austenite (b) Orientation map of (a) measured by automated crystal orientation mapping and color coded for the crystal direction parallel to the normal direction (ND) of the sheet. Hatched areas correspond to austenite grains. (c) The boundary character between γ - and α -grains, different orientation components [(001)||ND, red; all others blue] and the orientation variations inside of each grain (color shading; b, bainite). The micrograph and the maps are recorded with thermal FESEM JSM-6500F. Note: the extension toward the top and bottom of the measured are in (b) and (c) is larger than the area marked in (a). (Reprinted from Zaefferer et al., 2004; copyright 2004, with permission from Elsevier.)

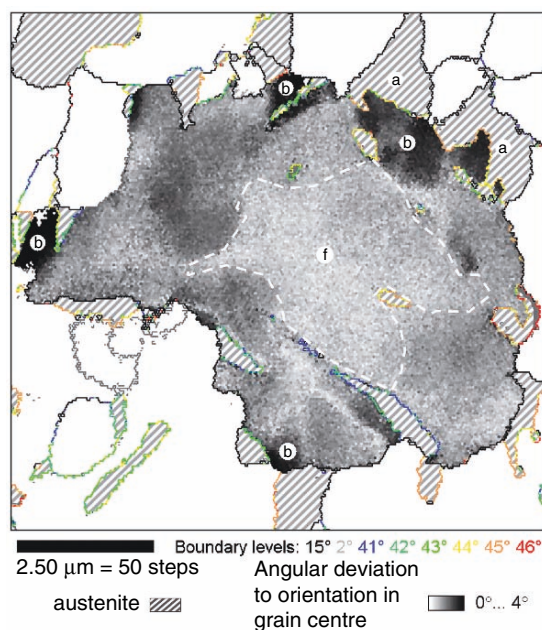


Figure 3–70. Orientation map of one grain from the microstructure in Figure 3–69a. Color code: angular deviation of every mapping point to one orientation in the center of the grain. Bainite appears in conjunction with a steep orientation gradient in ferrite. The white line marks the maximum extension of austenite at austenization temperature. f, ferrite; a, austenite (hatched); b, bainite. (Reprinted from Zaefferer et al., 2004; copyright, 2004, with permission from Elsevier.)

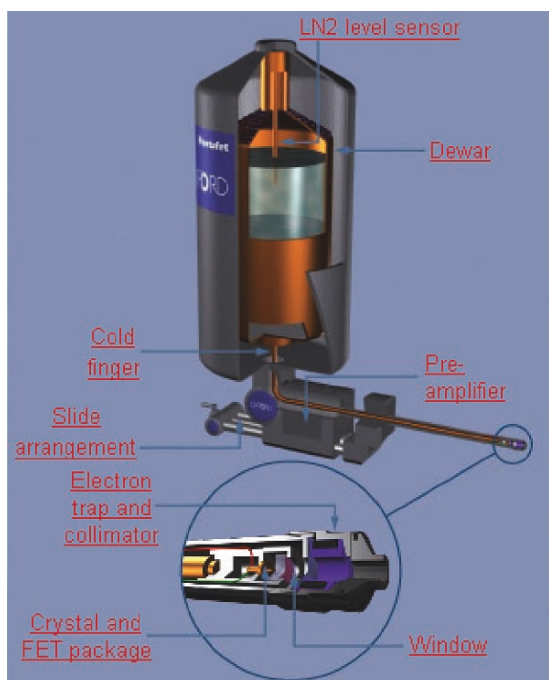


Figure 4-19. Schematics of a commercial EDXS detector showing the detector front, the Dewar system to cool the detector, and various components that are interfaced in the electron microscope. (Courtesy of N. Rowlands, Oxford Instruments.)

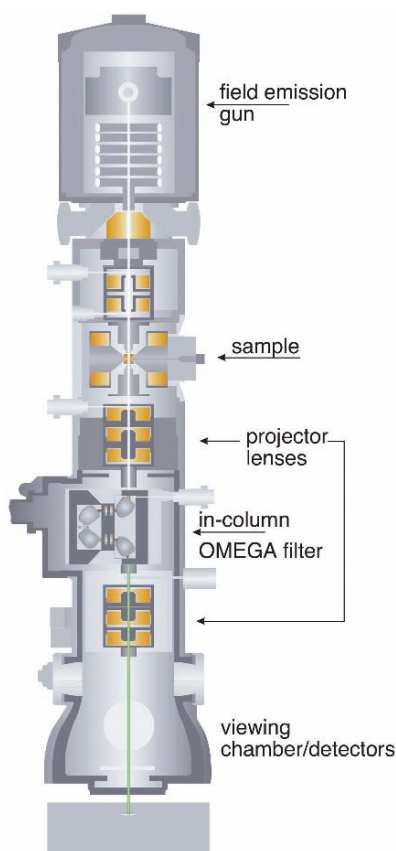


Figure 4-27. Schematic diagram of the in-column energy-filtered microscope (Zeiss-Libra 200). (Courtesy of P. Schlossmacher, Zeiss SMT.)

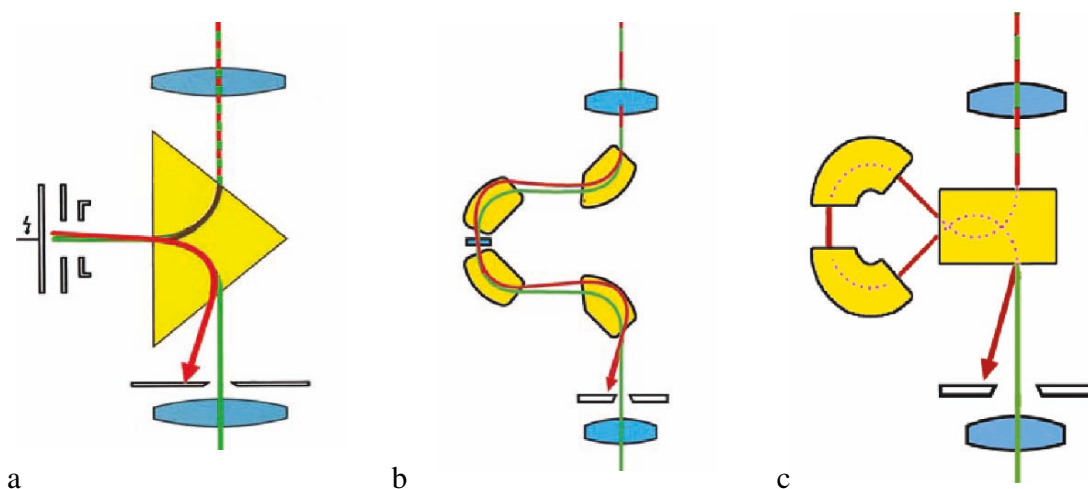
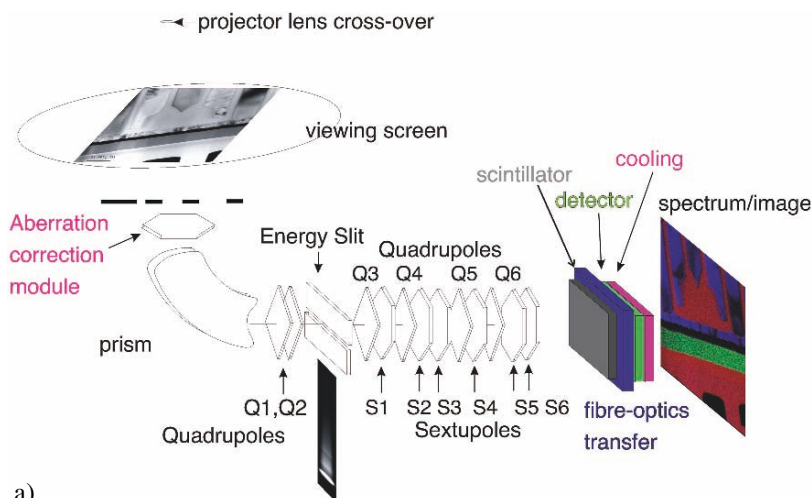
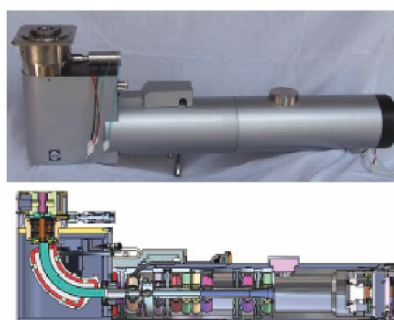


Figure 4-26. Various in-column spectrometer configurations. (a) Mirror-prism spectrometer, (b) OMEGA filter, and (c) Mandolin filter. (Courtesy of P. Schlossmacher, Zeiss SMT.)



a)



b)

Figure 4–29. (a) Postcolumn imaging filter (Gatan Imaging Filter GIF); schematic diagram of the electron optics components and detection system (top diagram). (Adapted from Krivanek, Gubbens et al., 1991a.) (b) Actual spectrometer [Gatan's GIF 2000 series spectrometer (Tridiem model)] and components (bottom diagram). (Courtesy of M. Kundman, Gatan.)

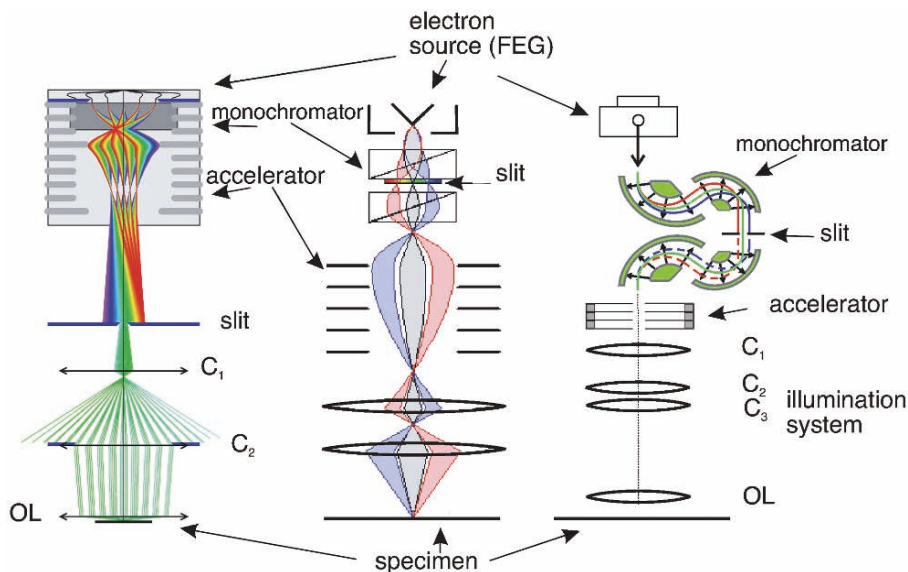


Figure 4–31. Various implementations of monochromators in commercially available instruments. Left: The FEI monochromator, single Wien filter. (Courtesy of P. Tiemeijer, FEI Company.) Center: The JEOL monochromator double-Wien filter. (Courtesy of JEOL Ltd.) Right: The electrostatic omega filter implemented in the Libra Zeiss microscope. (Courtesy of M. Haider, CEOS GmbH.)

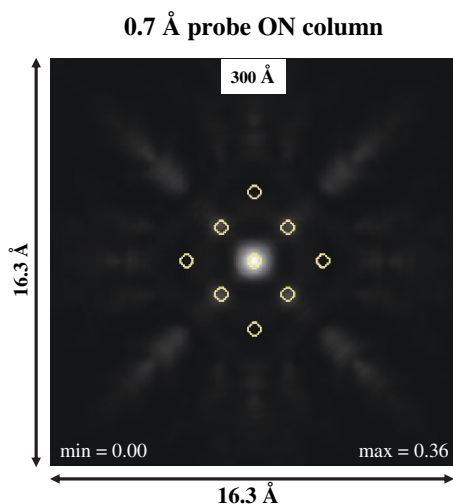


Figure 4–37. Real space intensity plots demonstrating the dispersion of the electron intensity when the electron beam is located on top of the atomic column and when it is located between two atomic columns. The bright empty circles indicate the position of the atoms in the cell closest to the point of impact of the electron beam. Channeling is observed when the beam is positioned on the atomic column (a) while much stronger dispersion is observed when the electron beam is not on the atomic column (b). (Data courtesy of C. Dywer and J. Etheridge.)

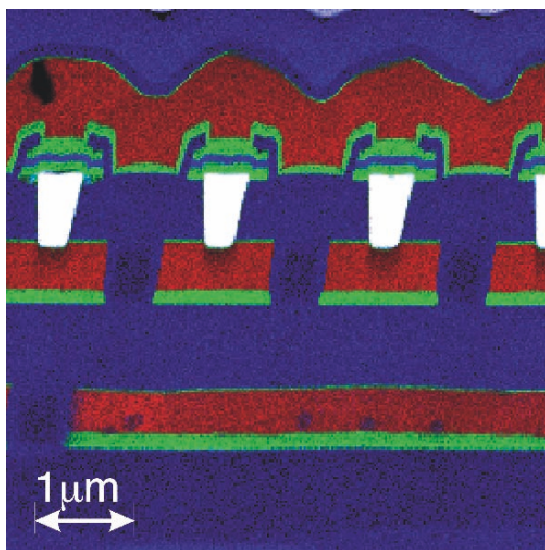


Figure 4–74. Color-coded elemental map of a device showing the distribution of elements in the rastered area: red, Al-rich area; blue, Si-rich area; green, Ti-rich area. White, W. Interdiffusion of Si into the Al is noted through the bottom barrier layer containing Ti.

Figure 4–77. Various approaches to EFTEM imaging. Zero-loss (ZL) filtered imaging (selecting only electrons that have lost no significant amounts of energy), plasmon imaging (selecting only electrons that have lost energy in the 10–30eV range), and core loss imaging with the three-windows technique (extrapolation of the background under the edge) and the jump-ratio technique.

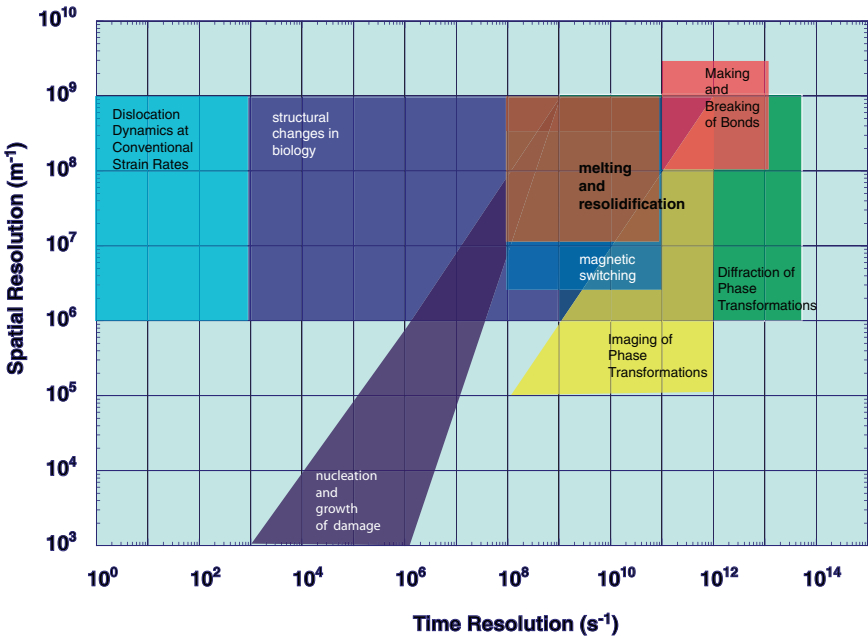
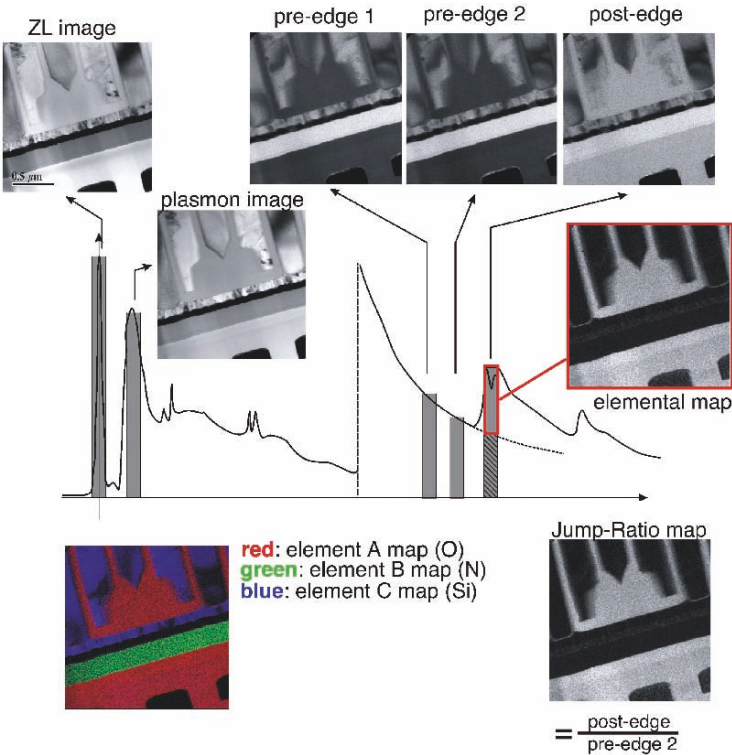


Figure 5–1. Phenomena classified by spatial and temporal resolution. Spatial resolution, is defined as follows: (1) if the technique is an imaging method, the low end of the bounding box would be defined by the smallest resolvable feature and the high end by the typical field of view or (2) for a nonimaging technique, the bounding box would be defined by the range of probe or spot sizes. Time resolution is defined as that for a single-shot investigation of irreversible processes. So time resolution is defined as the single-shot exposure time to obtain data that demonstrate a particular spatial resolution.

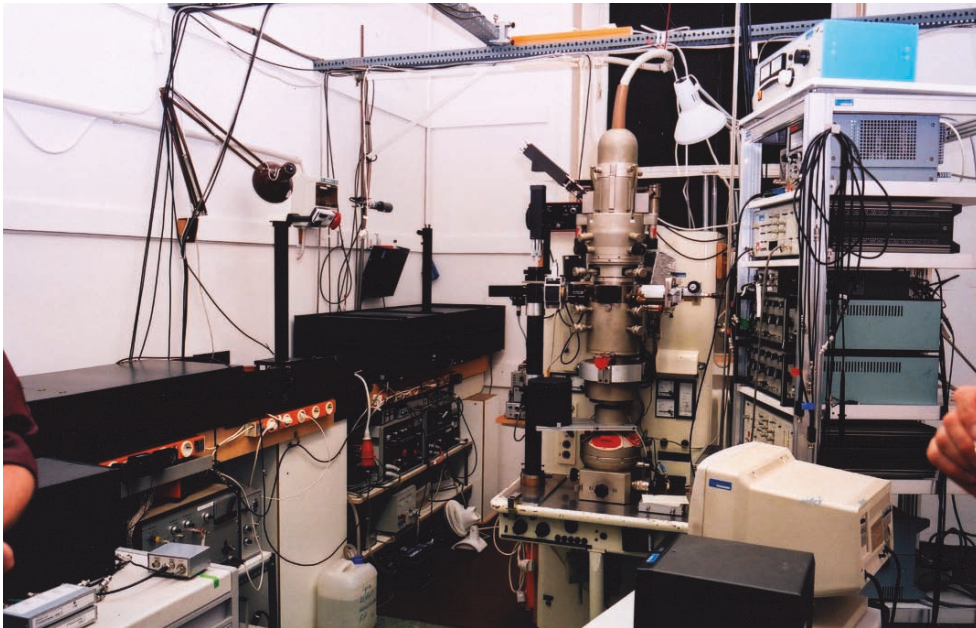


Figure 5–2. The DTEM at TU Berlin. Cathode and sample drive lasers are at left.

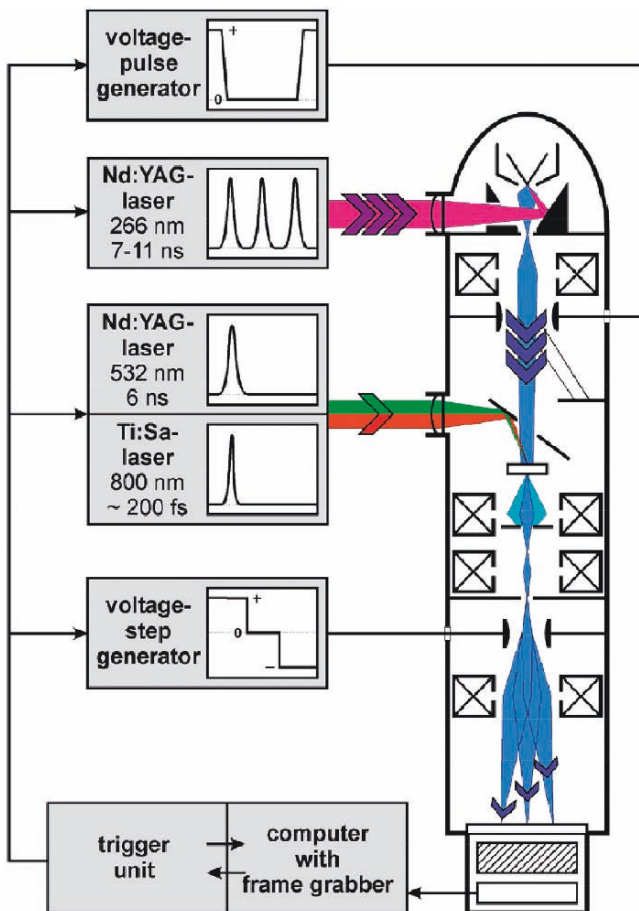


Figure 5–3. TU Berlin dynamic transmission electron microscope.

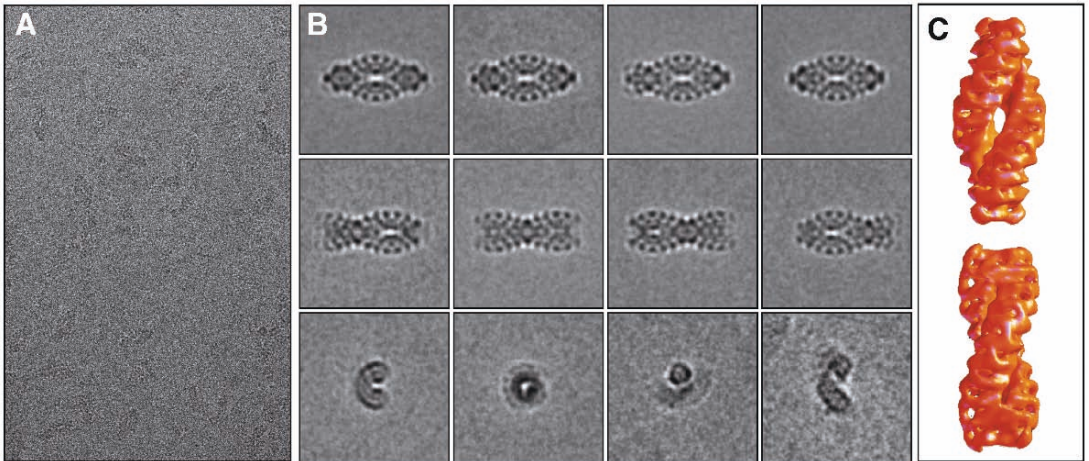


Figure 7-3. Single particle investigation of the giant protein complex TPP II from *Drosophila melanogaster* embedded in vitrified ice. In eukaryotes, tripeptidyl peptidase II (TPPII) is a crucial component of the protein degradation pathway. The 150-kDa subunits of *Drosophila* TPPII assemble into a giant proteolytic complex of 6 MDa with a remarkable architecture consisting of two segmented and twisted strands that form a spindle-shaped structure (length 56 nm, width 24 nm). a) Cryo-electron micrograph of isolated TPP II complexes illustrating the very weak image contrast and the high level of noise. b) Averaging and classification of a large number of equivalent projections of separate molecules. Once a large set of views is available, a preliminary 3D reconstruction can be computed and refined iteratively. c) The 3D model obtained by cryo-electron microscopy, reveals details of the molecular architecture and, in conjunction with biochemical data, provides insight into the assembly mechanism. The building blocks of this complex are apparently dimers, within which the 150 kDa monomers are oriented head to head. Stacking of these dimers leads to the formation of twisted single strands, two of which comprise the fully assembled spindle (Rockel et al., 2002 and 2005).

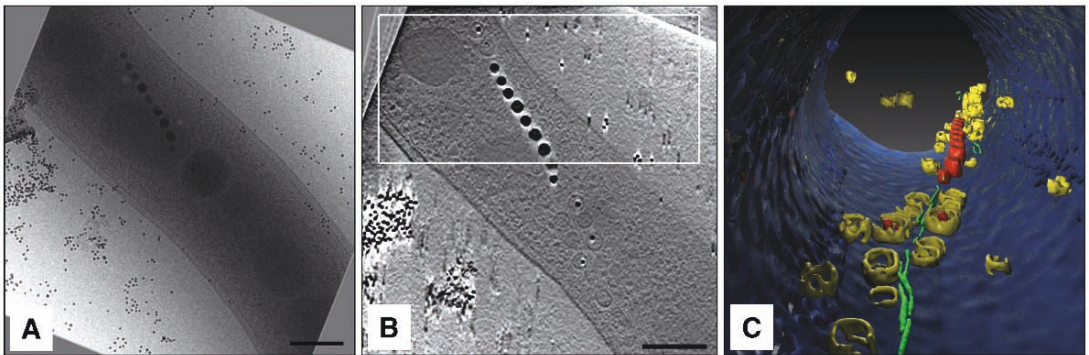


Figure 7-4. Cellular cryo-electron tomography of the magnetotactic microorganism *Magnetospirillum griffithswaldense*. The entire bacterium is oriented like a compass needle inside the magnetic field in its search for optimal living conditions. The miniature cellular compass is made by a chain of single nano-magnets, called magnetosomes (the scale bar represents 200 nm). a) The two-dimensional image represents one projection (at 0°) from an angular tilt-series. b) x - y slices along the z axis through a typical three-dimensional reconstruction (tomogram). c) Surface-rendered representation of the inside of the cell showing the membrane (blue), vesicles (yellow), magnetite crystals (red) and a filamentous structure (green). Until now, it was not clear how the cells organise magnetosomes into a stable chain, against their physical tendency to collapse by magnetic attraction. However, the biochemical analysis revealed a protein responsible for the chain formation and the 3D investigation a cytoskeletal structure, which aligns the magnetosomes like pearls on a string (Scheffel et al., 2005).

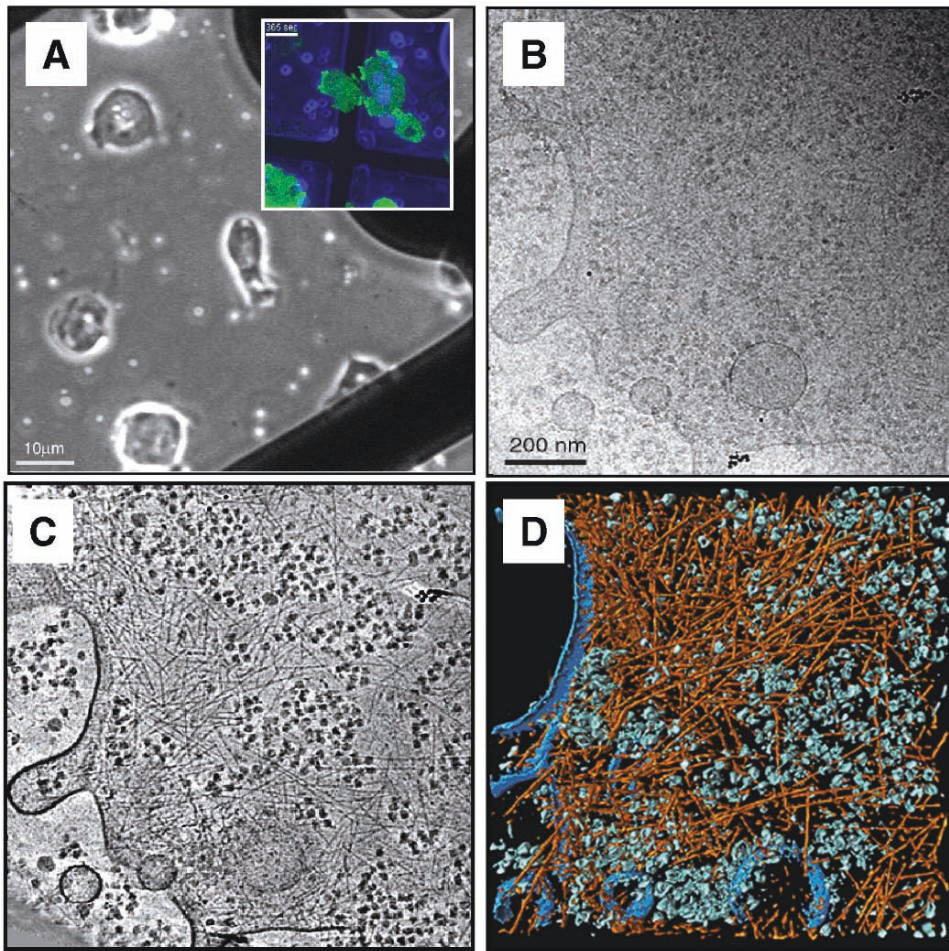


Figure 7-5. First electron tomographic investigation of a eukaryotic cell; the slime mould *Dictyostelium discoideum* embedded in vitrified ice. a) Phase contrast image and corresponding fluorescence image (inset) of cells on TEM grids. b) Cryo-electron micrograph at 0° tilt (conventional 2D projection) of a ~200nm thin peripheral region of the cell. c) Tomographic reconstruction from a complete tilt-series (120 images) and d) visualization by segmentation. Large macromolecular complexes, e.g. Ribosomes are shown in a green color, the actin filament network in orange-red and the cells' membrane in blue. Cryo-tomograms of *Dictyostelium discoideum* cells grown directly on carbon support films have provided unprecedented insights into the organization of actin filaments in an unperturbed cellular environment. The tomograms show, on the level of individual filaments, their modes of interaction (isotropic networks, bundles, etc.), they allow us to determine the branching angles precisely (in 3-D), and they reveal the structure of membrane attachment sites (Medalia et al., 2002).

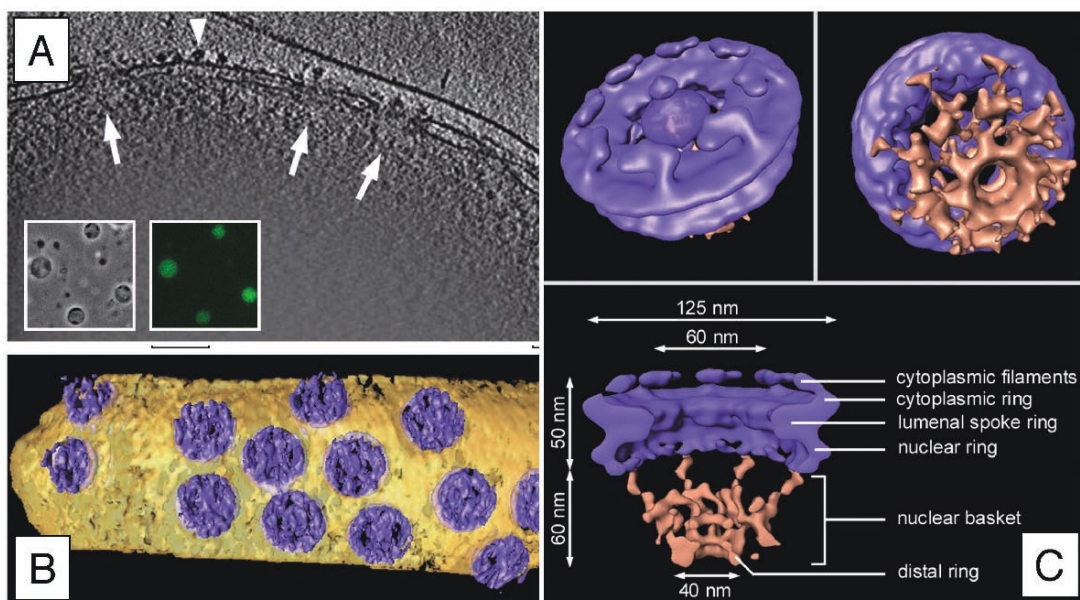


Figure 7-6. CET in combination with the single particle approach of transport-competent *Dictyostelium discoideum* nuclei. a) Three-dimensional reconstruction of the peripheral rim of an intact nucleus. X-y slice of 10nm thickness along the z axis through a typical tomogram. Side views of nuclear pore complexes (NPCs) are indicated by arrows. Ribosomes connected to the outer nuclear membrane are visible (arrowheads). Inset displays a phase-contrast image and the corresponding fluorescence image. b) Surface rendered representation of a segment of nuclear envelope (NPCs in blue, membranes in yellow). c) Structure of the Dictyostelium NPC after classification and averaging of subtomograms. Cytoplasmic face of the NPC (upper left); the cytoplasmic filaments are arranged around the central channel. Nuclear face of the NPC (upper right); the distal ring of the basket is connected to the nuclear ring by the nuclear filaments. Cross sectional view of the NPC (bottom). The dimensions of the main features are indicated. All views are surface-rendered (nuclear basket in brown; Beck et al. 2004).

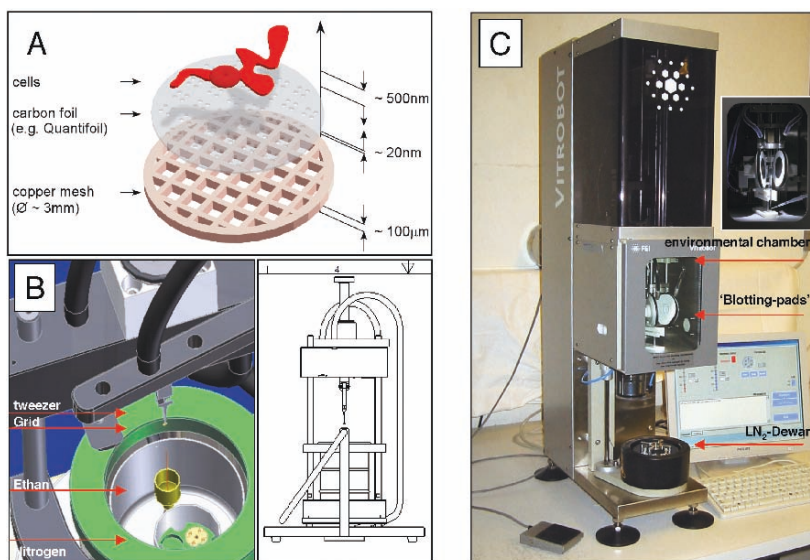


Figure 7-7. Plunge freezing instrumentation. a) Typical arrangement of cells cultured on TEM grid just prior to plunging. The schematic indicates the dimensions. b) CAD (computer aided design) image of a home-built plunge freezing apparatus (Images courtesy of R. Gatz, MPI of Biochemistry, Martinsried (near Munich), Germany). The small reservoir (yellow) in the middle of the dewar (green) contains the liquid ethane, which is chilled by a surrounding bath of liquid nitrogen. The forceps, which hold the grid are attached to a weighted arm. When the arm is released by means of a foot-trigger, the grid is gravity-plunged into the ethane. The cross-sectional scetch (left part of image B) illustrates the 'guillotine-like' arrangement. c) The Vitrobot (FEI company, Eindhoven, The Netherlands), a 'robot' for vitrification, allows the control of environmental and all necessary processing parameters.

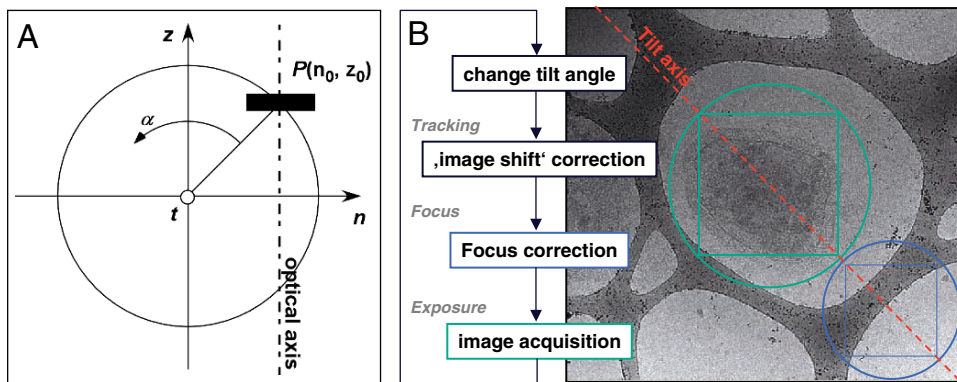


Figure 7-18. a) Geometric model of specimen rotation around a tilt axis. b) Automated data collection scheme for the 'full-tracking/ full-focusing' case. For tracking, focusing and the final image acquisition the same magnification is used and the beam is adjusted in a way that the areas where tracking, focusing and exposure is done do not overlap (blue and green circle). Using very short exposure times in combination with high binning values (4×8 or even 8×8) the exposure to the sample in auto-tracking and auto-focusing can be minimized.

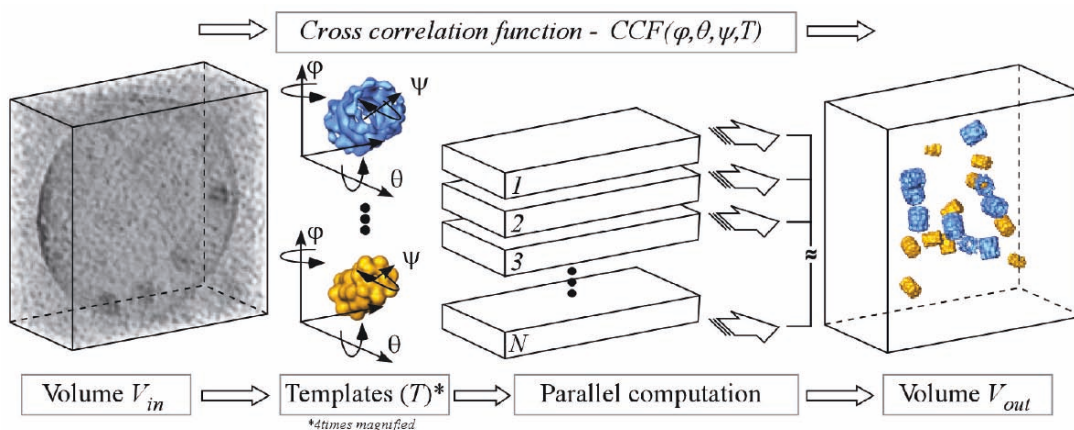


Figure 7-22. Detection and identification of individual macromolecules in cellular tomograms is based on their structural signature. Because of the crowded nature of the cytoplasm and 'contamination' with noise, an interactive segmentation and feature extraction is not feasible. It requires sophisticated pattern recognition techniques to exploit the information contained in the tomograms. A volume rendered presentation of a 3D image is presented on the left. Even though some high-density features may be visible, an unambiguous identification of individual structures would be difficult if not impossible given the residual noise. An approach, which has proven to work, is based on template matching. Templates of the macromolecules under scrutiny are obtained by a high- or medium resolution technique (X-ray crystallography, NMR, electron crystallography or single particle analysis). These templates (4 times magnified in this figure; 20S proteasome and thermosome) are then used to search the entire volume of the tomogram systematically for matching structures by three-dimensional cross-correlation and the result is refined by multivariate statistical analysis. In principle the 3D image has to be scanned for all possible Eulerian angles φ, ψ and θ around three different axes, with templates of all the different protein structures one is interested in e.g. the thermosome (blue) and proteasome (yellow). The search procedure is computationally very demanding but can be parallelized with respect to the different angular combinations in a highly efficient manner. Finally, the position and orientation of the different complexes can be mapped directly in the 3D image.

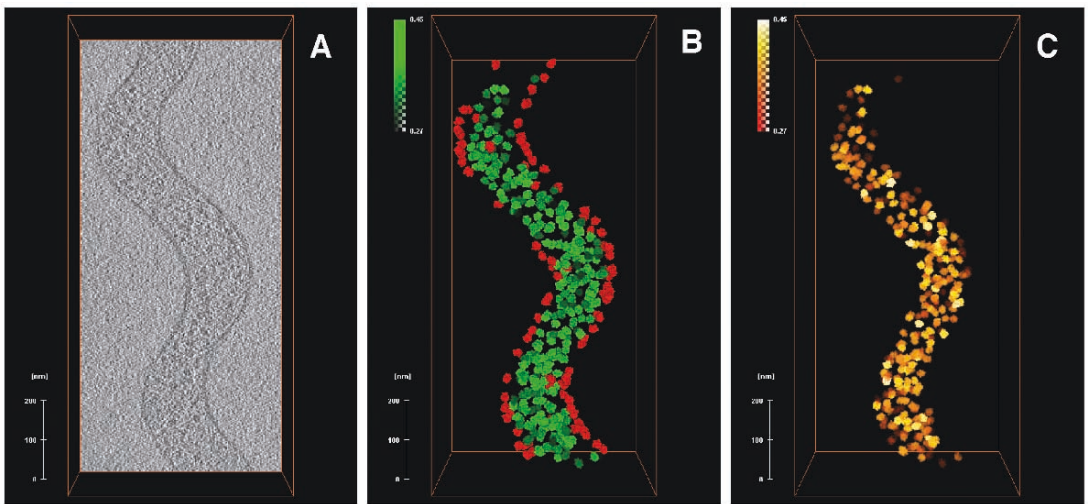


Figure 7-23. Mapping ribosomes in an intact *S. melliferum* cell. a) x - y slice from the corresponding tomogram. Image analysis of this portion of the cell is displayed in subsequent panels. b) Locations and orientation of all ribosomes detected by template matching. Each 70S ribosome is represented by the averaged density (see Fig. 26) derived from the tomogram. The colour coding indicates the detection fidelity; green is high, yellow is intermediate, red is low and probably represents false positives. The brightness of the colour corresponds to correlation peaks heights. c) Final ribosome atlas after removal of putative false positives (images courtesy of J. Ortiz, MPI of Biochemistry, Martinsried, Germany).

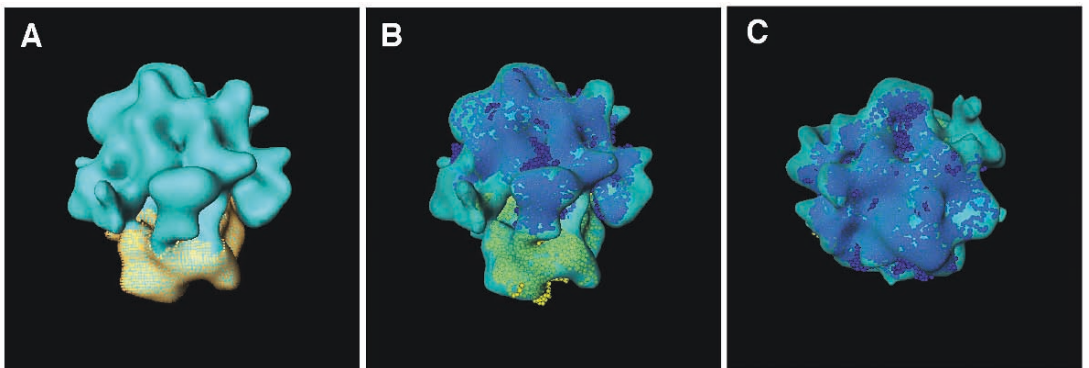


Figure 7-25. a) Averaged structure of the 70S ribosome derived from the dataset above (average from 300 individual particles to ~4nm resolution). The map highlights the 30S subunit in yellow. b) Docking of high resolution structures of the 70S ribosome into the map shown in A. c) "Crown view" of B (images courtesy of J. Ortiz, MPI of Biochemistry, Martinsried, Germany).

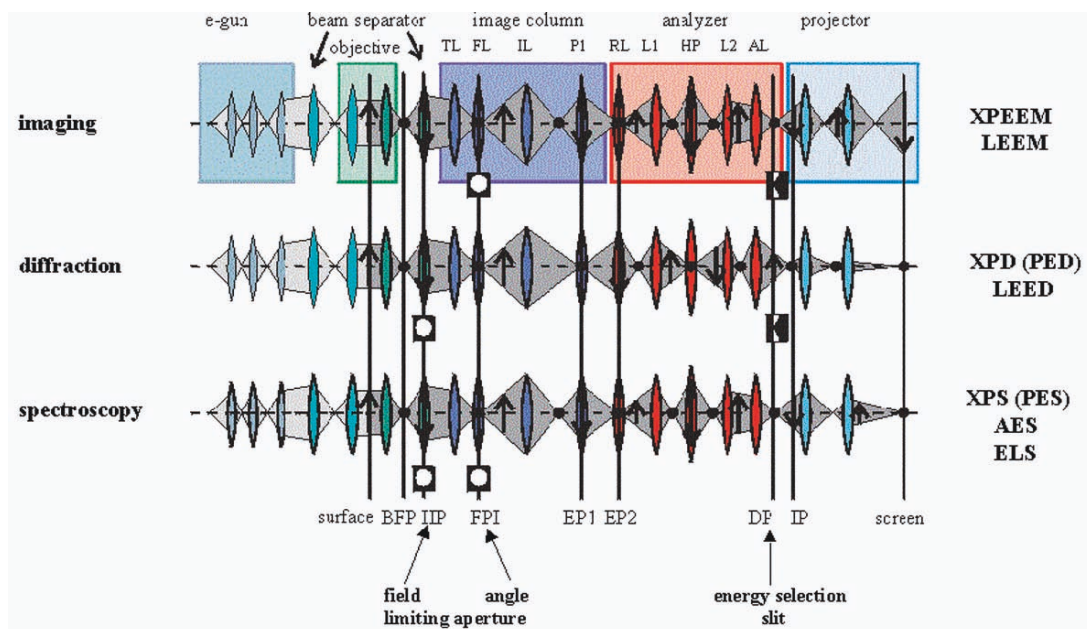


Figure 8-18. The three fundamental operation modes of a SPELEEM system. The various sections of the instrument are shown folded into one plane. In imaging and diffraction the energy selection slit is inserted in the dispersive plane DP and the image/diffraction pattern behind DP is imaged with the projector. The intermediate lens IL is used to switch between imaging and diffraction, simultaneously with the exchange of the contrast aperture in FPI and the field-limiting aperture in IIP. For fast spectroscopy both apertures are inserted, the energy selection slit is removed, and the dispersive plane is imaged by the projector.¹⁵

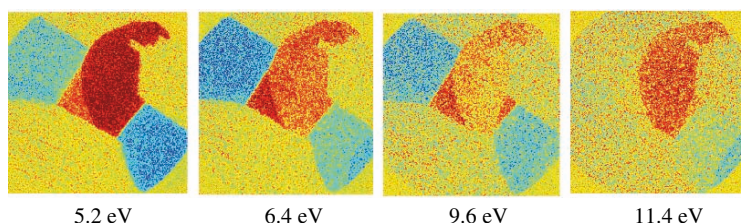


Figure 8-26. Quantum size contrast between regions with different thickness of an Fe film on W(110), taken with different electron energies, that is wavelengths. The images in the top row show the intensity and those at the bottom the magnetic signal (exchange asymmetry). Blue and red correspond to opposite magnetization directions.⁸⁸

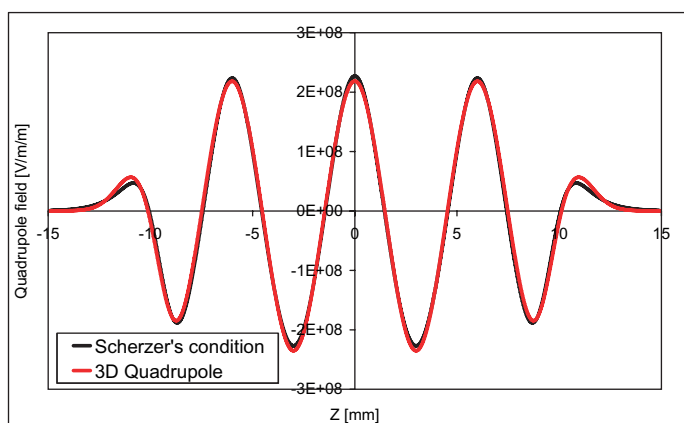


Figure 10-13. Electrostatic correction of chromatic aberration. (a) Rays in a quadrupole quadruplet or sextuplet. (b) Match between the potentials needed to satisfy Scherzer's condition and those in the corrector.

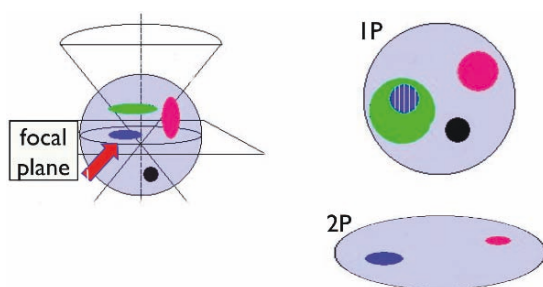


Figure 11-2. Comparison between conventional (1P) and two-photon (2P) excitation with respect to image formation. When focusing on the actual focal plane under 1P, a contribution from adjacent planes that are physically excluded in the 2P process is obtained, as happens in a confocal setup. (From Giuseppe Vicidomini, LAMBS, MicroScoBio, University of Genoa.)

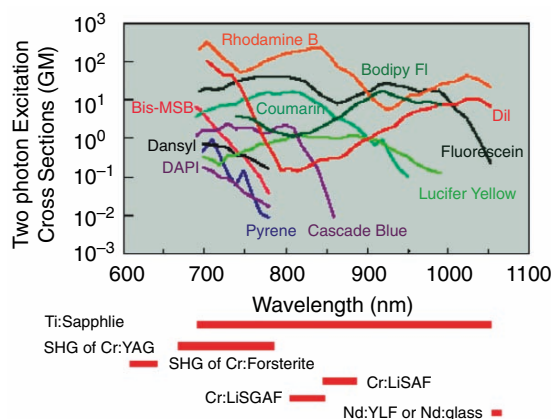


Figure 11-6. Two-photon cross-sections for popular fluorescent molecules as a function of the excitation wavelength. Red bars indicate the emission range of some common laser sources utilized in TPE microscopy and spectroscopy.

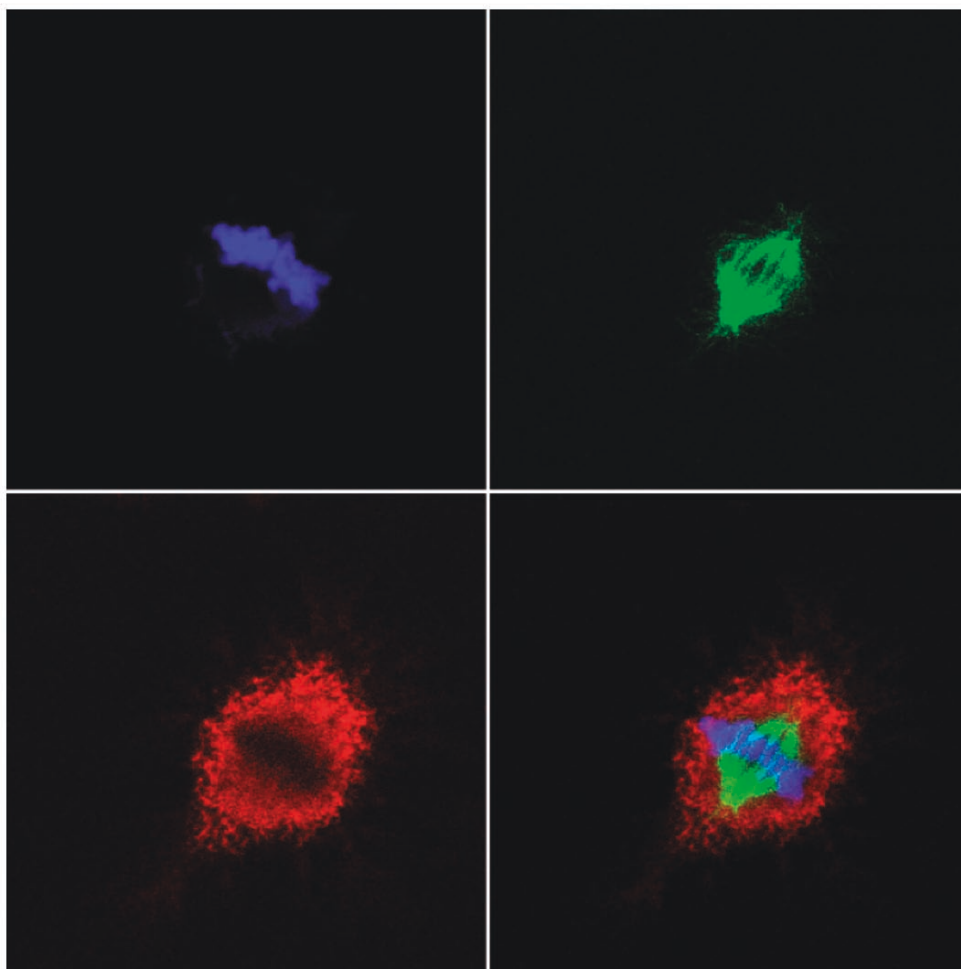


Figure 11-12. Multiple excitation of three fluorescent dyes using 740 nm under a TPE regime. The conventional excitation would have required the utilization of 360 or 405 nm, 488 nm, and 543 nm laser lines. The final image (lower right quadrant) is realized by merging the three subsets. (This image has been acquired by students of the Biotechnology School during the course of Advanced Microscopy Techniques activated at the University of Genoa, academic year 2005. Advisors: Grazia Tagliafierro and Alberto Diaspro.)

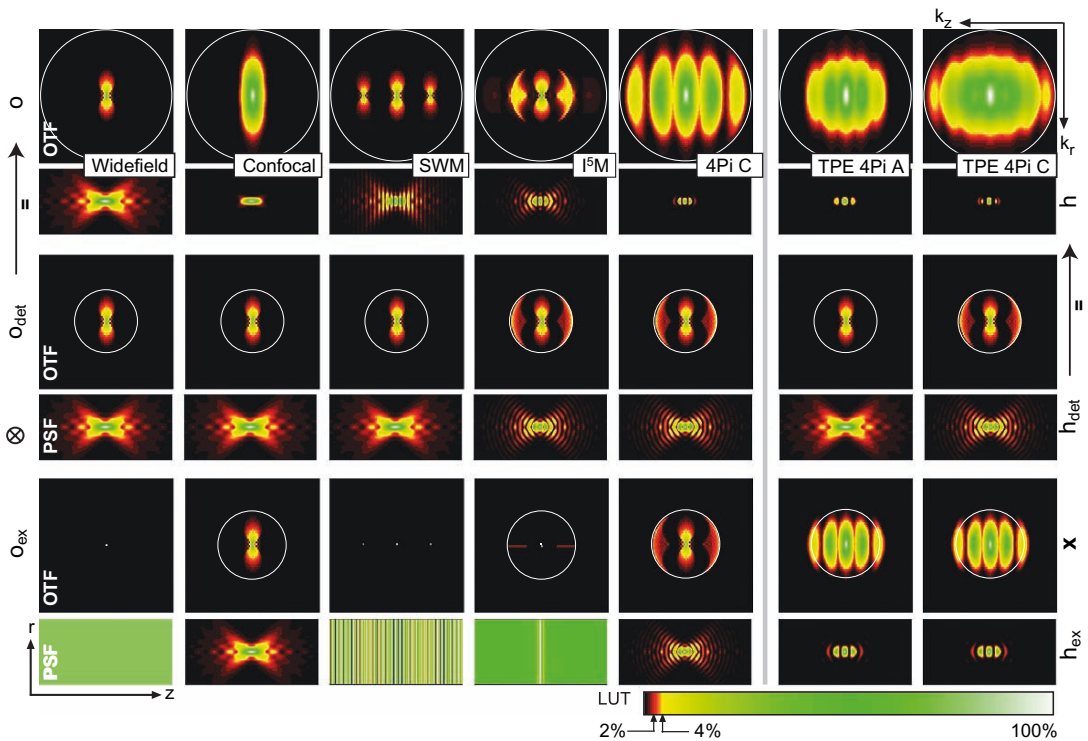


Figure 12-4. Overview of the excitation (bottom), detection (center), and effective (top) OTFs' modulus and the corresponding PSFs of the wide-field, confocal, standing wave (SWM), I⁵M, 4Pi confocal type C, TPE 4Pi confocal type A, and TPE 4Pi confocal type C microscopes. The color look-up table (LUT) has been designed to emphasize the important weak OTF regions. The OTFs are shown in the squares above the corresponding PSFs; the zero frequency point is in the center and the largest frequency displayed is $2\pi/80\text{nm}^{-1}$. The circles represent the maximum possible carrier as explained in Figure 12-3. For TPE, the excitation OTFs slightly extend over these circles because the excitation wavelength of 800nm is less than double the one-photon excitation wavelength of 488nm. While all these methods extend the OTF along the axial direction, they fundamentally differ in contiguity and absolute strength within the support region. For example, there are pronounced frequency gaps for the SWM and depressions for the I⁵M. The rectangular images of the PSFs represent a region of $5 \times 2.5\mu\text{m}$ with the geometric focus in the center.

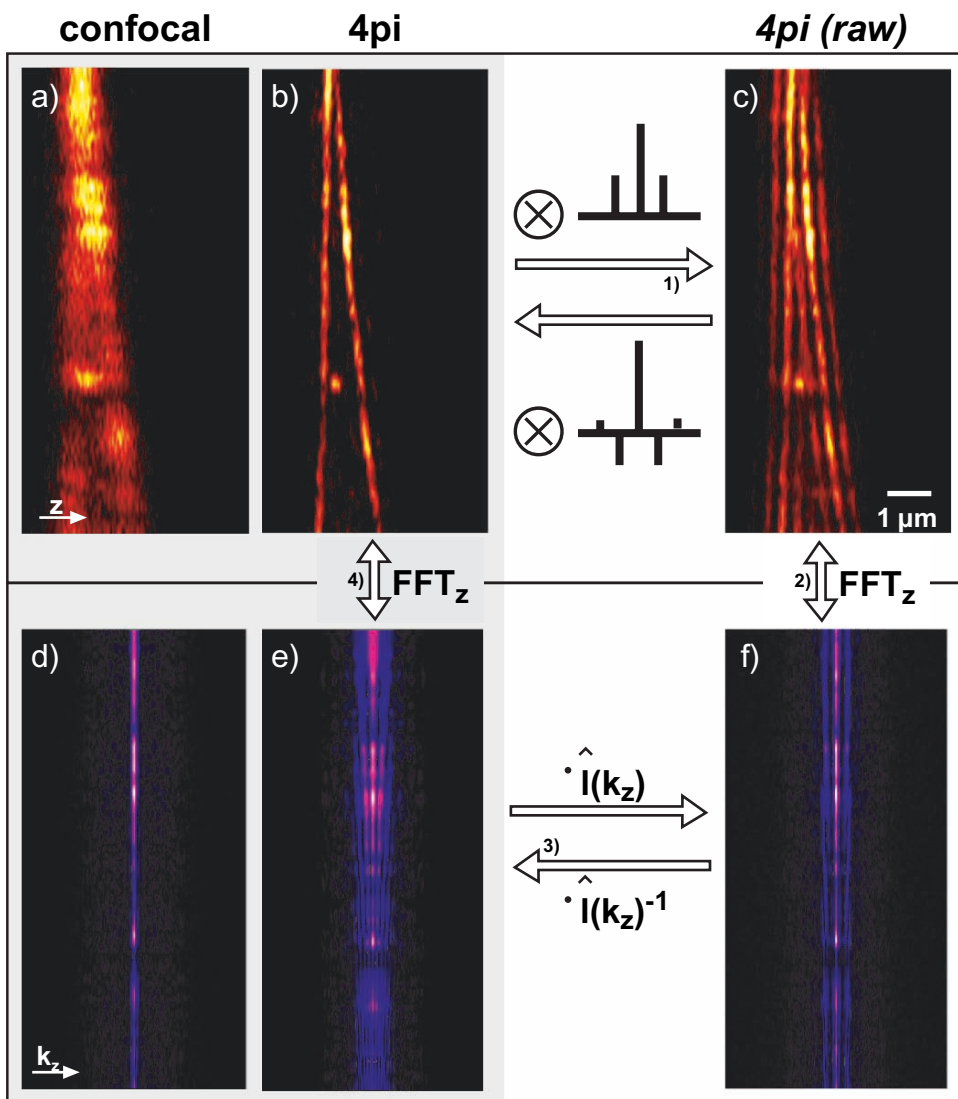


Figure 12–5. Lobe removal and deconvolution in 4Pi microscopy. The figure shows images of the same pair of actin fibers in a fixed mouse fibroblast cell recorded in the TPE confocal (a) and TPE 4Pi type A (b and c) mode. The corresponding Fourier transform along the optic axis is also shown (d, e, and f). The five-fold axial resolution increase (a vs. b) and the correspondingly extended OTF (d vs. e) are immediately visible. The side lobes are well below 50% and the factorization of the PSF's axial and lateral dependence is possible in 4Pi microscopy. Therefore, an inverse discrete filter can be found and its application to the raw data (c) yields a valid and almost artifact-free image (b). Alternatively, lobe removal can be performed in the frequency domain. Equation (16) indicates that the Fourier transforms of the raw data (f) is given by the product of the Fourier transform of the lobe-free image (e) and the lobe function $\hat{l}(k_z)$. Thus, (2) Fourier transforming the raw data, (3) multiplying with the inverse of the lobe function's Fourier transform, $\hat{l}^{-1}(k_z)$, and (4) Fourier backtransforming lead to almost the same lobe-free image. This method can be applied even if the separation of axial and lateral dependence is impossible for the PSF. The Fourier transforms along the axial direction (3 and 4) merely have to be replaced by their 3D counterparts.

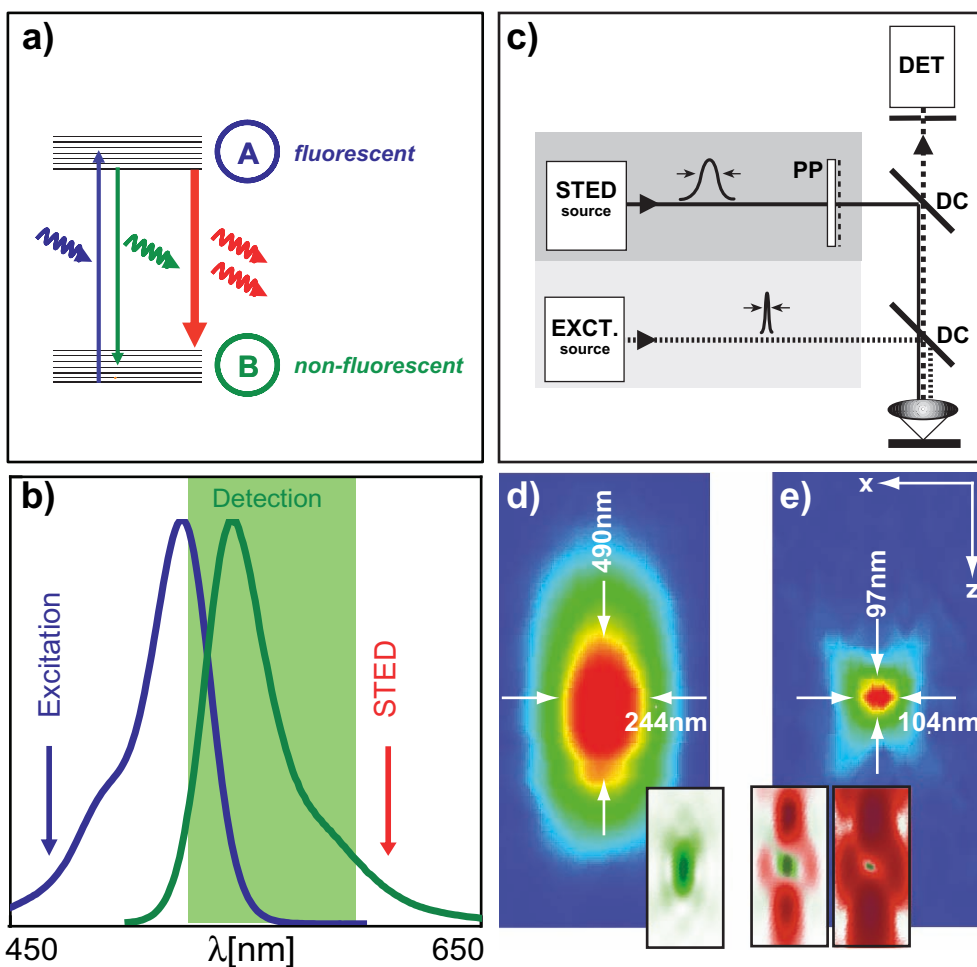


Figure 12-7. Stimulated emission depletion (STED) was the first implementation of the RESOLFT principle. (a) Dye molecules are excited into the S_1 (state A) by an excitation laser pulse. (b) Fluorescence is detected over most of the emission spectrum. However, molecules can be quenched back into the ground state S_0 (state B) using stimulated emission before they fluoresce by irradiating them with a light pulse at the edge of the emission spectrum shortly after the excitation pulse and before they are able to emit a fluorescence photon. Saturation is realized by increasing the intensity of the depletion pulse and consequently inhibiting fluorescence everywhere except at the “zero points” of the focal distribution of the depletion light. (c) Schematic of a point-scanning STED microscope. Excitation and depletion beams are combined using appropriate dichroic mirrors (DC). The excitation beam forms a diffraction-limited excitation spot in the sample (inset in d) while the depletion beam is manipulated using a phase-plate (PP) or any other device to tailor the wavefront in such a way that it forms an intensity distribution with a nodal point in the excitation maximum (left inset in e). The third inlay shows the resulting quenching probability when saturating the depletion process. (d) and (e) show an experimental comparison between the confocal PSF and the effective PSF after switching on the depleting beam. Note the doubled lateral and five-fold improved axial resolution. The reduction in dimensions (x, y, z) yields ultrasmall volumes of subdiffraction size, here 0.67 a.l (Klar et al., 2000), corresponding to an 18-fold reduction compared to its confocal counterpart. The spot size is not limited on principle grounds but by practical circumstances such as the quality of the zero and the saturation factor of depletion.

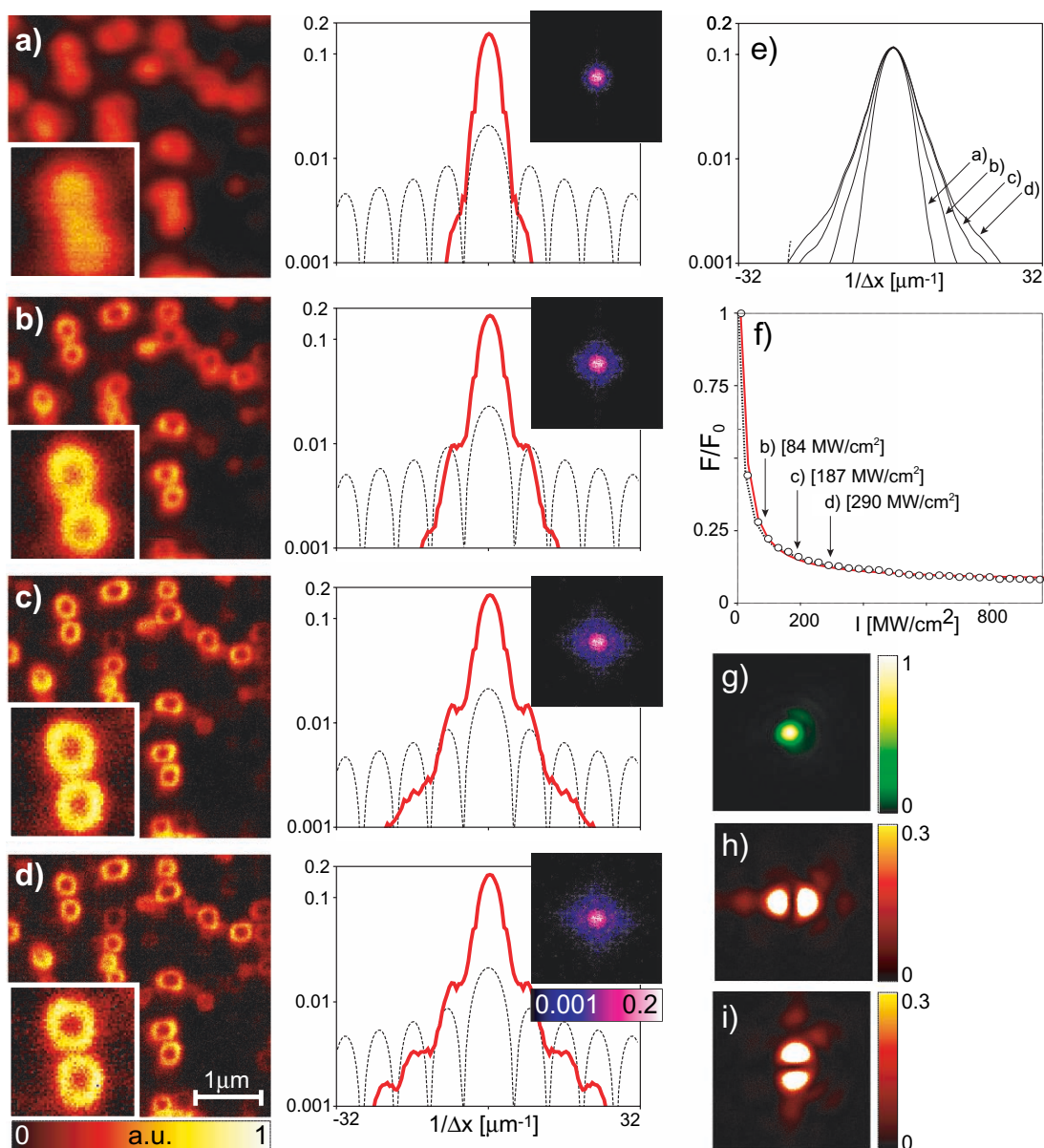


Figure 12-10. Images of a wetted Al_2O_3 matrix featuring z -oriented holes (Whatman plc, Brentford, UK) with a spin cast of a dyed (JA 26) polymethyl methacrylate solution. The rings formed in this way are ~ 250 nm in diameter and are barely resolved in confocal mode. (a–d) The confocal image (a) and STED images with two depleting beams perpendicularly polarized and aberrated by “1D” phase-plates (b–d). The excitation PSF (g) and the STED PSF for y polarization (h) and x polarization (i) are shown on the right. The STED intensity was chosen at the spots marked in the saturation curve (f). The smaller effective spot size also results in an extended OTF as seen in the second column. Here, the insets show the 2D Fourier transformation of the images in the left and the graphs show a profile along the x direction. Note the logarithmic scales. The Fourier transform of the image is given by the product of OTF and the Fourier transform of the object [Eq. (2)]. For such regular structures, an estimate for the modulus of the OTF can therefore be gained by estimating the latter and solving for the OTF. The dashed line shows the Fourier transform of a ring with a diameter of 275 nm and a width of 50 nm and the estimated OTF is presented in (e). (f) The suppression of fluorescence resulting from stimulated emission. The phase-plates were removed and the ratio of fluorescence without STED light (F_0) and with the STED beams switched on (F) was recorded. The intensities are pulse intensities per beam at the global maximum.

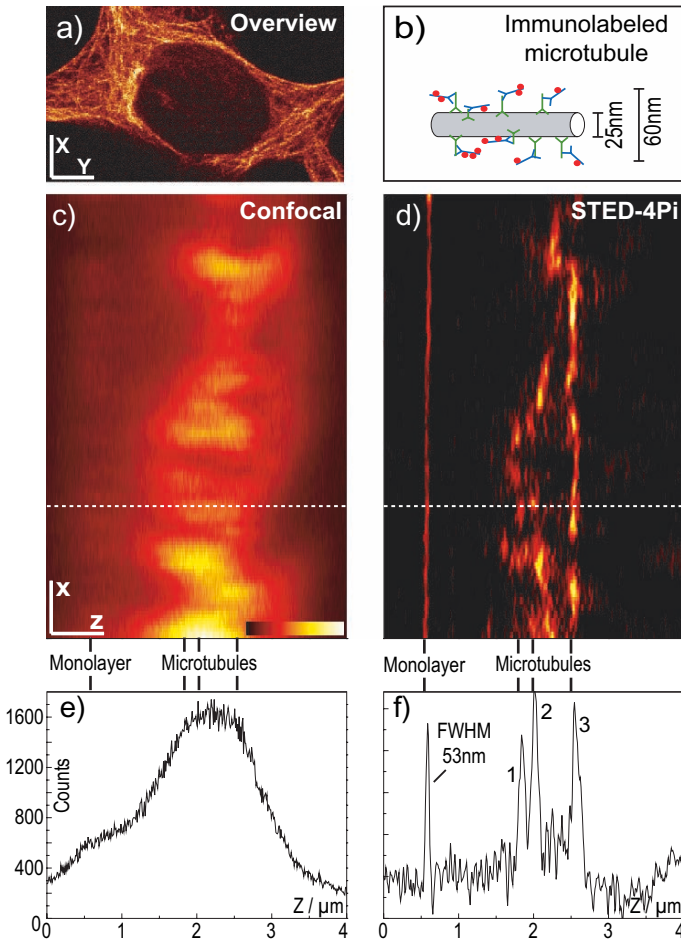


Figure 12–11. Subdiffraction immunofluorescence imaging with STED-4Pi microscopy. (a) Overview image (xy) of the microtubular network of an HEK cell. (b) Sketch of typical dimensions of a labeled microtubule fluorescently decorated via a secondary antibody. (c) and (d) Standard confocal and STED-4Pi xz image recorded at the same site of the cell; the straight line close to the cell stems from a monomolecular fluorescent layer attached to the adjacent coverslip. In both images, the pixel size was 95×9.8 nm in the x and z direction, respectively; the dwell time per pixel was 2 ms. Note the fundamentally improved clarity in (d). The STED-4Pi microscope's PSF features two low side lobes caused by the secondary minima STED intensity distribution. These lobes are $<25\%$ and were removed in the STED-4Pi image using linear filtering as outlined in the text [see Eq. (16ff)]. (e) and (f) Corresponding profiles of the image data along the dashed lines in (b) and (c) quantify the improved axial resolution of the STED-4Pi microscopy mode (f) over the confocal benchmark. Peaks 1, 2, and 3 due to microtubules are broader than the response to the monolayer. Note the ability of the STED-4Pi microscope to distinguish adjacent features.

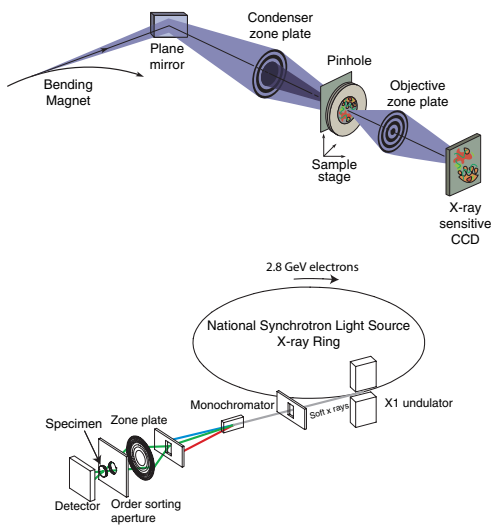


Figure 13–15. Schematic of the main components of a transmission X-ray microscope or TXM (top: courtesy of D. Attwood, Lawrence Berkeley National Laboratory) and a scanning transmission X-ray microscope or STXM (bottom: courtesy of Y. Wang, then of Stony Brook.)

Figure 13–27. Near-carbon-edge absorption spectra of several amino acids, showing the effects of various molecular bonds in the absorption spectrum. These resonances can be used for chemical contrast in X-ray microscopy. (Reprinted from Kaznacheyev et al., © 2002, with permission from American Chemical Society.)

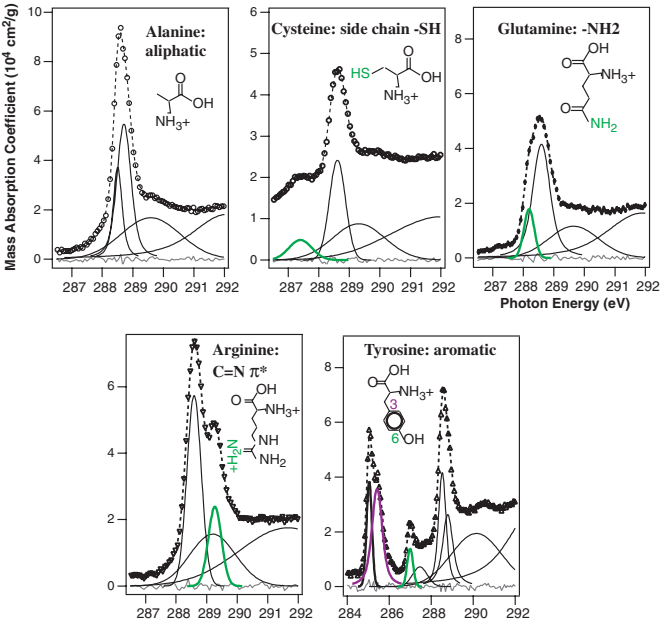


Figure 13–29. Human fibroblast with immunogold labeling for tubulin. This is a composite of two images: a bright field image (gray tones) to image overall mass, and a dark field image (red tones) to selectively imaging the silver-enhanced gold labels. This whole-mount cell was fixed and then permeabilized to allow for introduction of the immunogold labels, after which it was air dried. (From Chapman et al., © 1996b,c, courtesy of the Microscopy Society of America.)

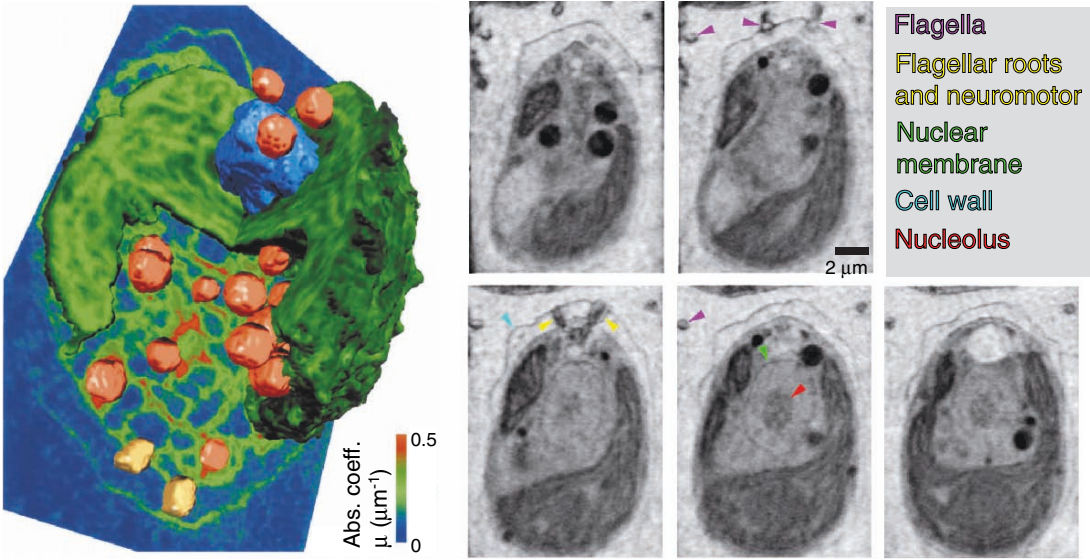
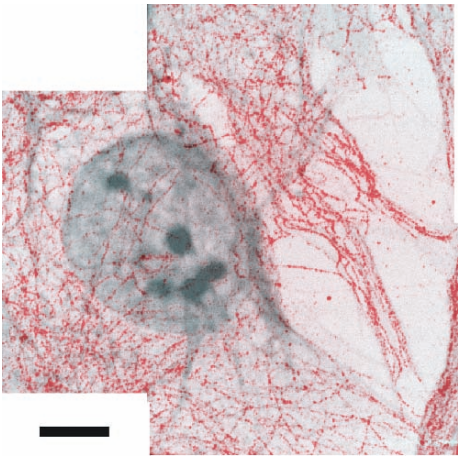


Figure 13–30. 3D rendering (left) and reconstruction slices (right) of the algae *Chlamydomonas reinhardtii* viewed by soft X-ray tomography at the BESSY I synchrotron. This alga was plunge-frozen in liquid ethane, and imaged over 180° rotation sequence. The reconstruction is given in terms of the quantitative linear absorption coefficient for 517 eV X-rays. (Reprinted from Wei et al., © 2000, with permission from Elsevier.)

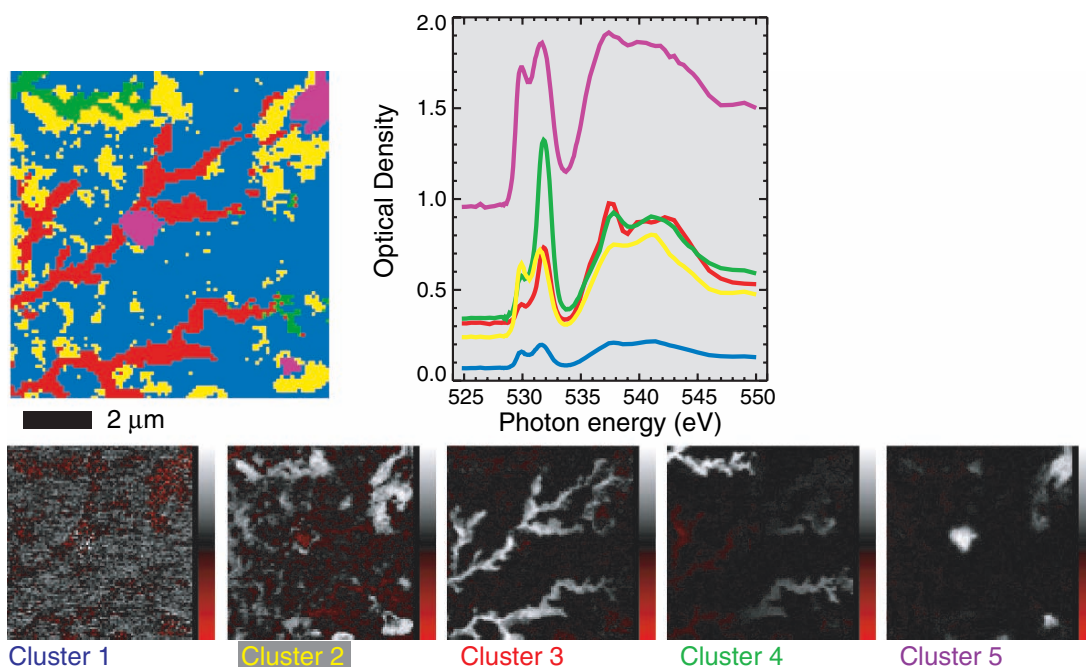


Figure 13–34. Cluster analysis in a spectromicroscopy study of lutetium in hematite. Lutetium is serving as a homologue to americium in an investigation of the uptake and transport of nuclear waste products in groundwater colloids. By using a pattern recognition algorithm to search for pixels with spectroscopic similarities, a set of signature spectra is automatically recovered from the data (shown here in a color-coded classification map) and thickness maps can be formed based on these signature spectra. Analysis at the oxygen edge reveals two different phases of reactivity for lutetium with hematite. Analysis by Lerotic (2004), from a study by T. Schäfer, INE Karlsruhe.

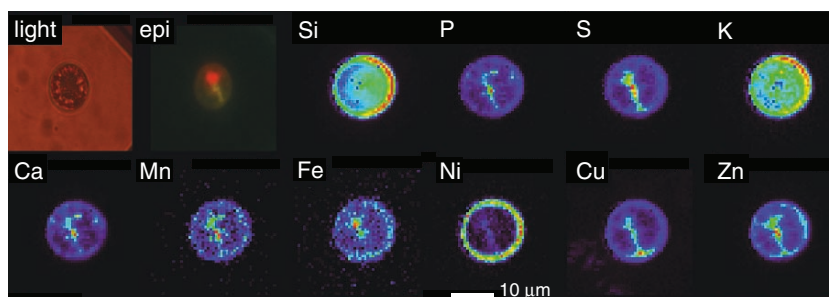


Figure 13–35. Visible light and epifluorescence micrographs, and false color X-ray fluorescence element maps of a centric diatom collected from the southern Pacific. In this region of the ocean, iron availability is a biolimiter with an impact on oceanic uptake of carbon dioxide from the atmosphere. X-ray microprobes allow one to study iron content on a protist-specific basis. (Reprinted from Twining et al., © 2003, with permission from American Chemical Society.)

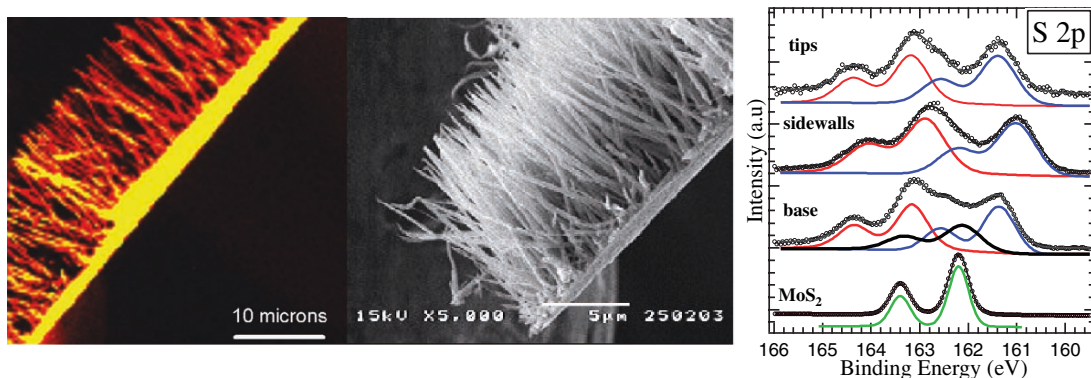


Figure 13–42. In Electron Spectroscopy for Chemical Analysis or ESCA microscopy, a monochromatic beam is used to illuminate a region several micrometers across; electron optics are then used to image a tunable electron ejection energy to reveal surface chemistry. Though this does not involve zone plate imaging, we include it here due to its widespread use with tunable X rays. In this case a 90-nm resolution ESCA microscope was used to locate aligned MoS₂ nanotube bundles and select certain areas along the axes of the tubes for detailed examination. The image at left was acquired using Mo 3*d* electrons, while S 2*p*, Mo 3*d*, and valence band spectra taken at the tips and sidewalls and the growth base from the Si wafer appear strongly affected by the low dimensionality of the nanotubes and differ significantly from the corresponding spectra taken on a reference MoS₂ crystal. (Reprinted from Kiskinova et al., © 2003, with permission of EDP Sciences.)

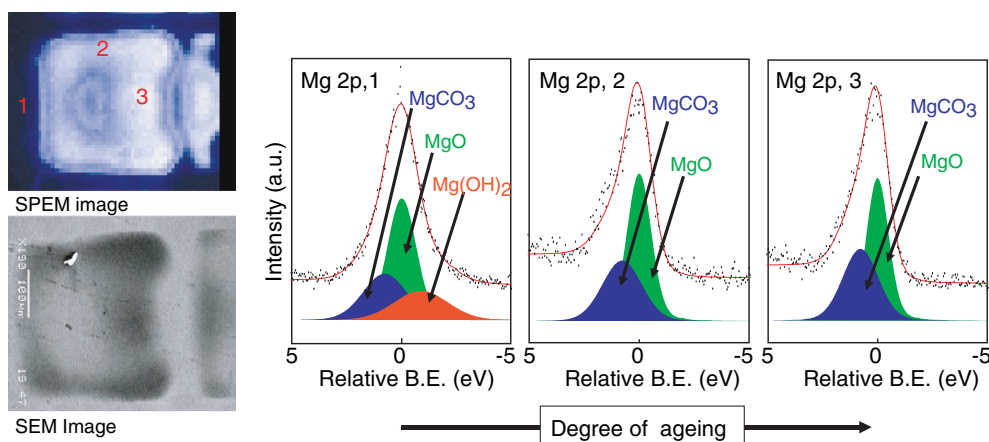
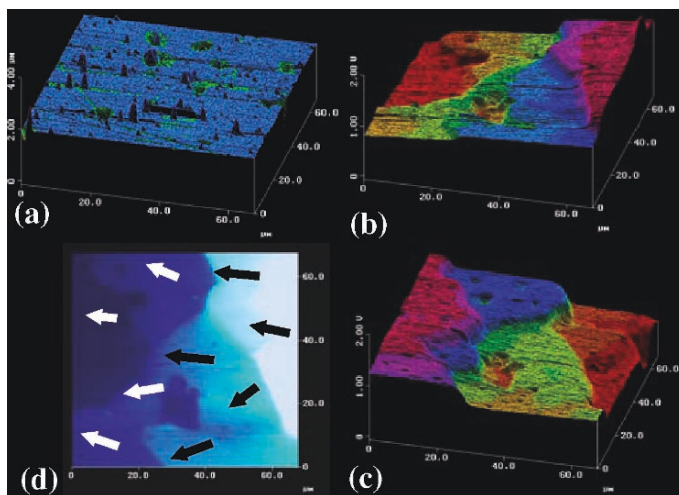


Figure 13–43. Scanning photoemission microscope (SPED) study of a plasma display cell. In this microscope the specimen is scanned through the zone plate focus while photoelectrons are collected by an electron spectrometer. This figure shows a SPED image, a scanning electron micrograph, and photoelectron spectra from several regions of the sample. In a plasma display cell, light of the appropriate color emerges through a front glass window which is protected from plasma damage by a composite insulating layer including MgO. The photoelectron spectra show aging in the Mg(OH)₂ component of the layer over the life of the display cell. (Reprinted from Yi et al., © 2005, with permission from the Institute of Pure and Applied Physics.)

Figure 14–13. Transport imaging in polycrystalline ZnO. (a) Surface topography. (b and c) SSPM images under lateral bias exhibit potential drops at grain boundaries, indicative of grain boundary resistive behavior. Note that the direction of potential drops inverts with bias. (d) Current maps for positive and negative bias polarity. (Partially reproduced from Kalinin and Bonnell, *Zeitschrift für Metallkunde*, 90, 983–989, 1999.)



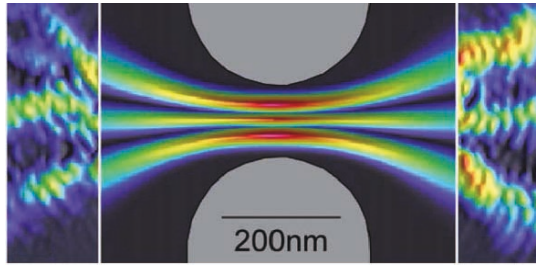


Figure 14–15. Experimental images (outside) and theoretical simulations (inside) of the flow of electron waves through a quantum point contact. Fringes spaced by half the Fermi wavelength demonstrate coherence in the flow. [Courtesy of Westervelt et al. Reproduced with permission from Physica E, 24, 63–69, 2004.]

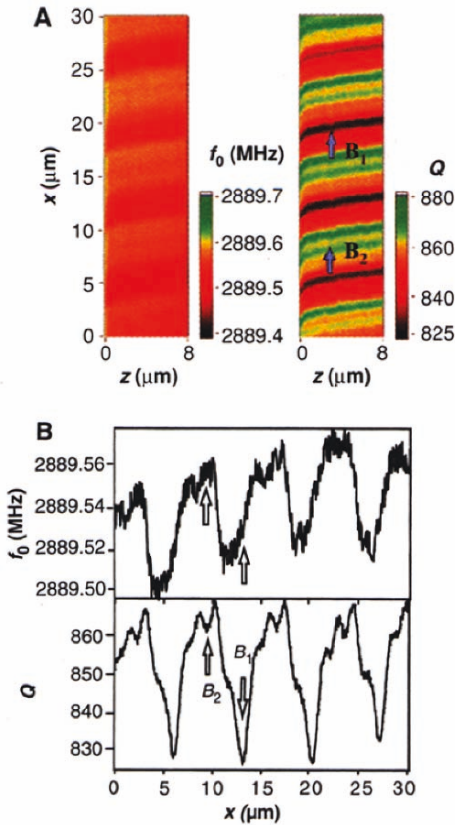


Figure 14–22. (A) Images of the profiles of periodic dielectric constant (on the left) and ferroelectric domain boundaries (on the right). (B) Line profiles of frequency (top) and quality factor (bottom) images.

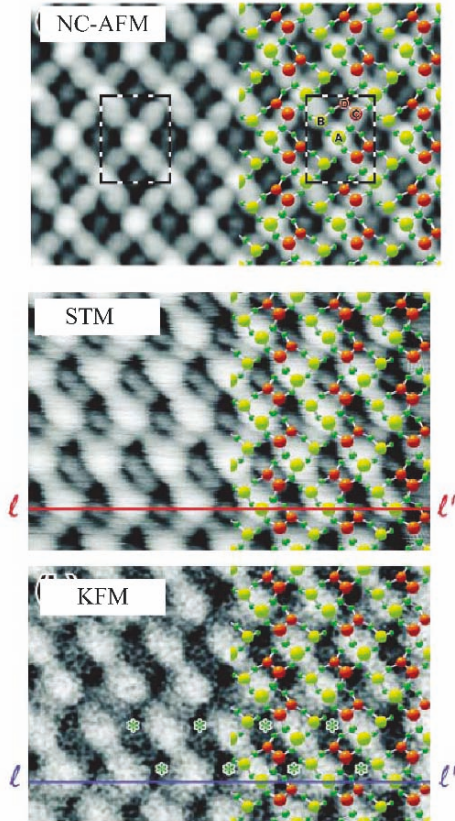


Figure 14–23. Atomic resolution NC-AFM, STM, and KFM on Ge/Si (105). [Courtesy of Eguchi et al. Reproduced with permission from Physical Review Letters, 93(26), 2004.]

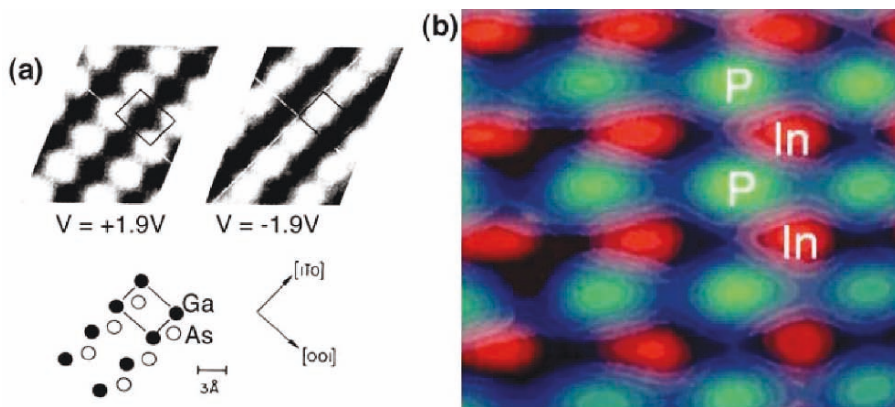


Figure 15-13. (a) Bias-dependent STM on GaAs(110): selective imaging of Ga and As sublattices at positive and negative sample bias, respectively (Reprinted with Permission from Feenstra et al., ©1987 by the American Physical Society). (b) Compound STM image of the InP(110) surface, assembled from separate positive and negative bias scans (Reprinted from Ebert et al., ©1992 with permission from Elsevier).

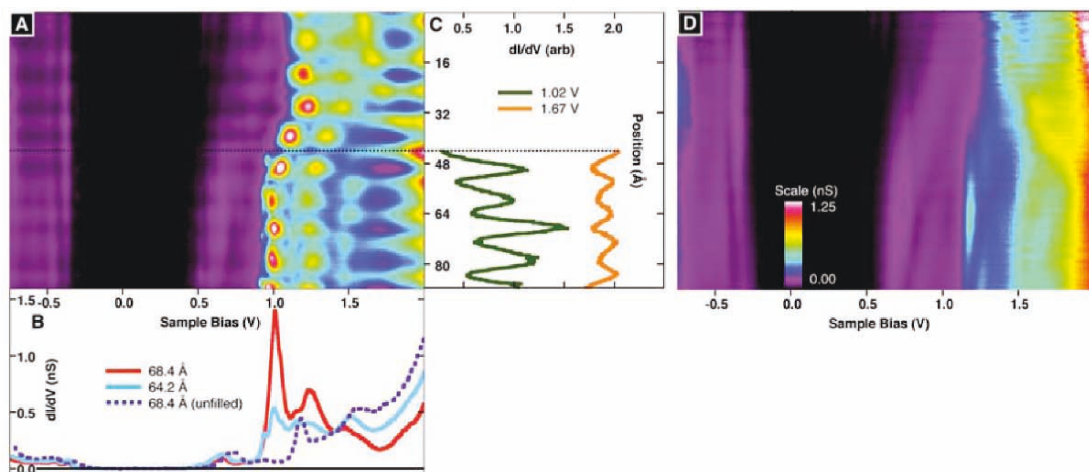


Figure 15-16. $I(V)$ tunneling spectroscopy on C_{60}/C -nanotube "peapods" (Reprinted with Permission from Hornbaker et al., ©2002 AAAS). (A) Map of an array of full dI/dV spectra along the axis of a C -nanotube "peapod." Sample bias voltage is plotted on the horizontal axis and displacement along the tube on the vertical axis. (B) Representative dI/dV spectra at selected positions along the tube. Large conductance peaks are found at positions of embedded C_{60} molecules. (C) Variation of tunneling conductance along the tube axis. (D) Reference spectroscopic map on an empty C -nanotube section without embedded C_{60} , in which no strong modulation of the tunneling conductance is observed.

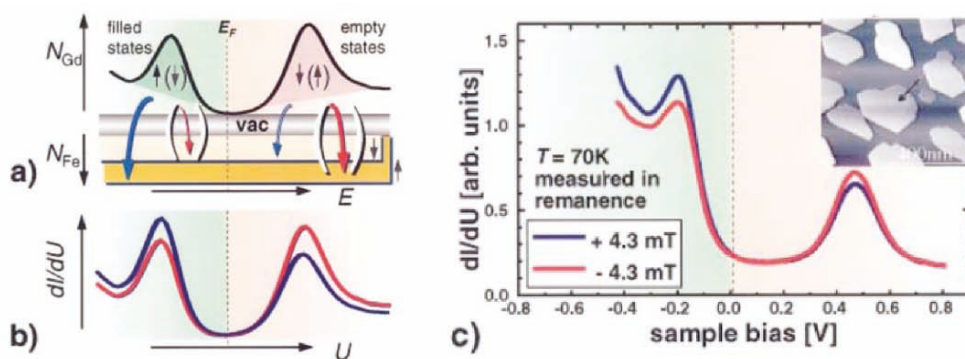


Figure 15-18. Spin-polarized STM on a Gd(0001) sample with an exchange-split surface state and a magnetic Fe tip with constant spin polarization close to E_F . (a) Due to the spin-valve effect the tunneling current of the surface state spin component parallel to the tip magnetization is enhanced. (b) Illustration of the reversal in the dI/dU signal at the surface state peak position upon switching the sample magnetically. (c) Experimental observation of this reversal in tunneling into an isolated Gd island (Reprinted with permission from Bode et al., ©1998 by the American Physical Society).

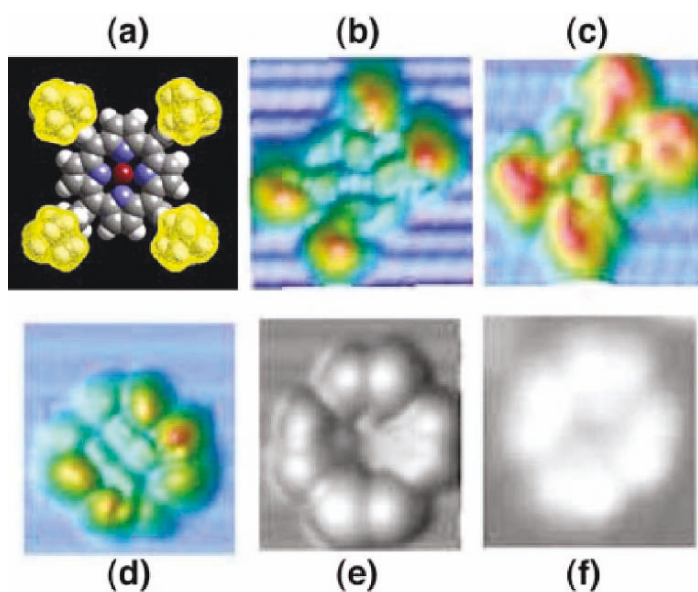


Figure 15-37. Identification of the molecular conformation of Cu-DTBPP on Cu(211). (a) Molecular model of Cu-DTBPP, with the four-lobed pattern observed by STM marked in yellow. (b–e) STM contrast calculation for different angles between the four legs and the substrate: (b) 60°, (c) 45°, (d) 30°, and (e) 10°. (f) Experimental STM image of the molecule. (Reprinted from Moresco et al., ©2002 with permission from Elsevier.)

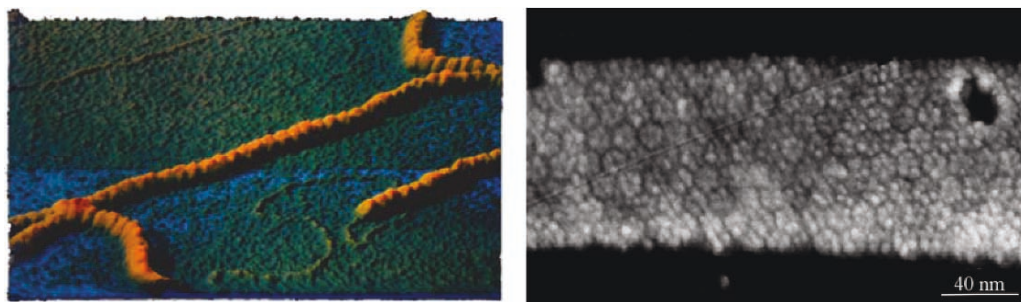


Figure 16-1. STM topographies of Rec-DNA complex (left) and an aberrant bacteriophage capsid (right). In the upper right corner, a hole has been created applying a voltage pulse between the tip and the sample. Both images immediately reveal the molecular arrangement of the respective complex structure. (Left, from Amrein et al., 1988, reproduced with permission. Right, from Amrein et al., 1998b, reproduced with permission.)



Figure 16-3. Neuronal cell, cultured on an electronic chip. The chip is designed such as to pick up an action potential of the cell. The image demonstrates proper tracing of the cell surface. In a future application, an appropriately designed stylus might be used as an additional electrode to excite or record an action potential at any location of the cell body or a process of the neuronal cell.

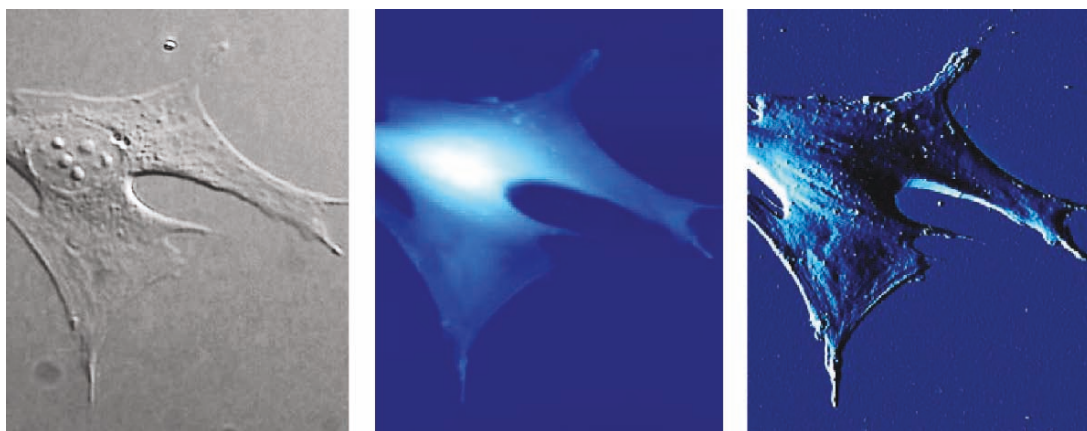
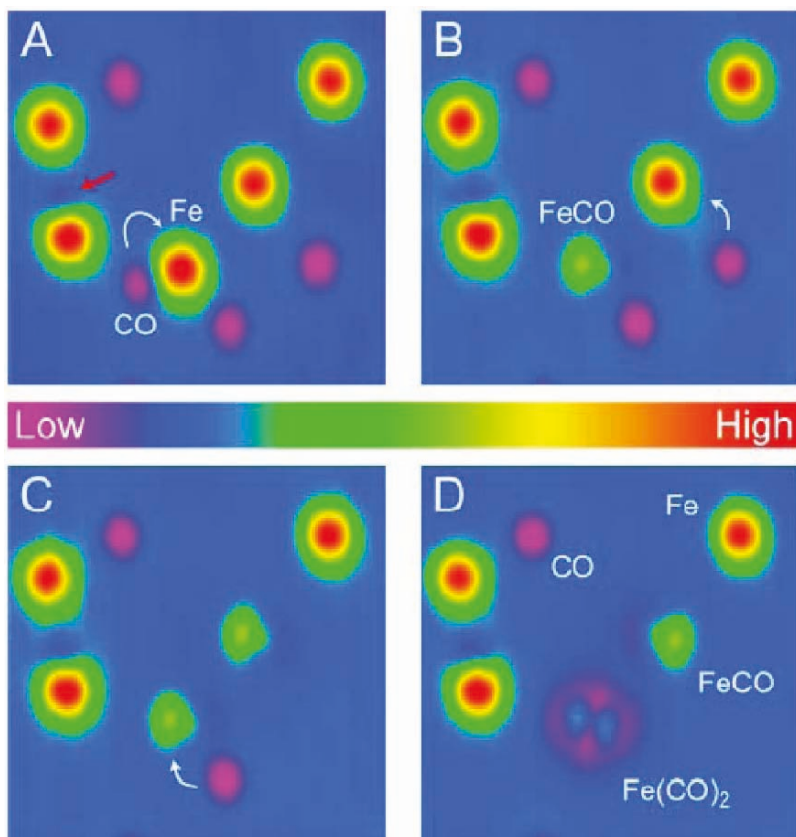


Figure 16-18. Images of a fixed 3T3 cell in buffer, on a coverslip. Left to right: Optical (differential interference contrast, DIC) and simultaneous AFM (JPK Instruments, contact mode) topography and feedback signal.

Figure 17-36. Bond formation induced with STM tip. A sequence of STM constant-current images at 13K showing the formation of Fe-CO bonds by vertical manipulation. Fe atoms are imaged as protrusions and CO molecules as depressions. The white arrows indicate the pair of adsorbates involved in each bond formation step. In (B) and (C) a CO molecule has been picked up and bonded to an Fe atom to form Fe(CO). In (D) a second CO molecule has been bonded to Fe(CO) to form Fe(CO)₂. (From Lee and Ho, 1999.)



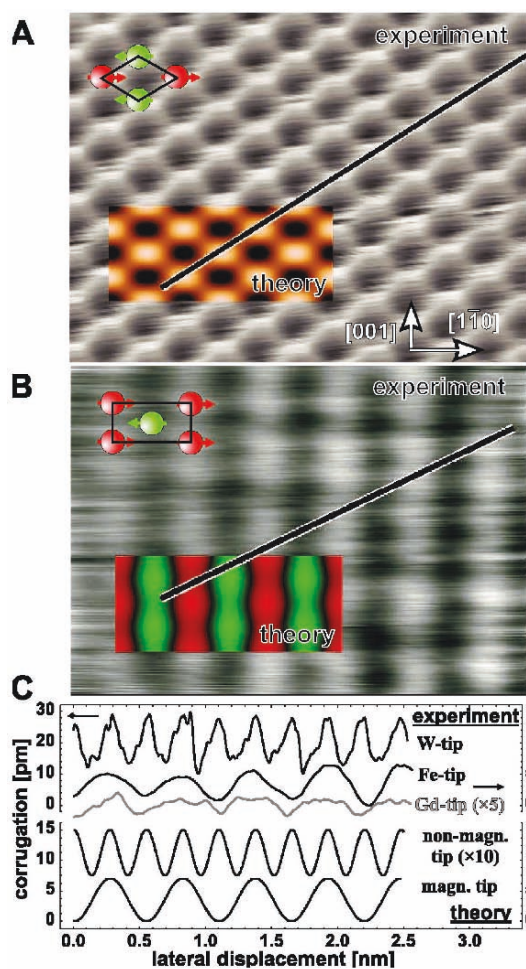


Figure 17-43. Atomic resolution magnetic imaging with SP-STM. (A) Constant-current image of one monolayer of Mn on W(110) recorded with a non-magnetic W tip at 16 K. (B) Image recorded with a magnetic Fe tip showing an antiferromagnetic configuration as predicted by theory. The colored insets show calculated STM images. (C) Experimental and theoretical line sections from (A) and (B). The image size is 2.7 nm by 2.2 nm. (From Heinze et al., 2000.)

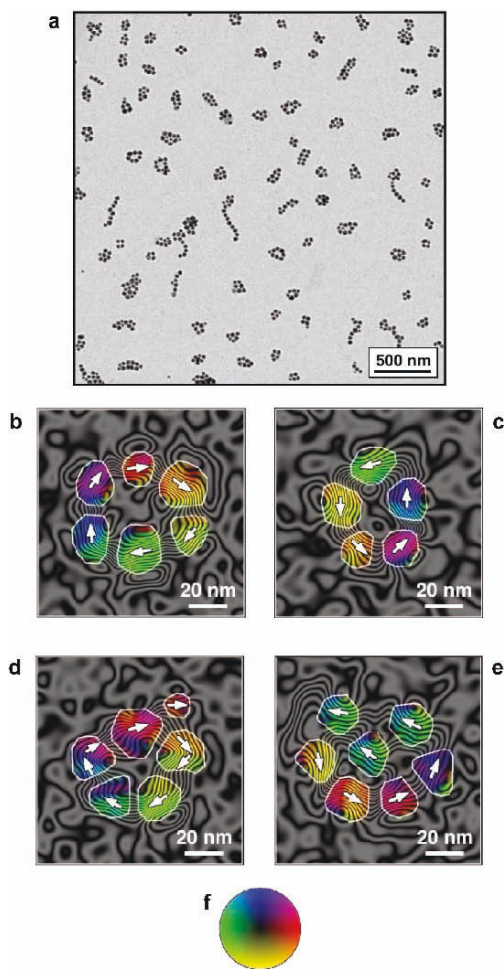


Figure 18-9. (a) Low magnification bright-field image of self-assembled Co nanoparticle rings and chains deposited onto an amorphous carbon support film. Each Co particle has a diameter of between 20 and 30 nm. (b-e) Magnetic phase contours (128 $\times$ amplification; 0.049 radian spacing), formed from the magnetic contribution to the measured phase shift, in four different nanoparticle rings. The outlines of the nanoparticles are marked in white, while the direction of the measured magnetic induction is indicated both using arrows and according to the color wheel shown in (f) (red = right, yellow = down, green = left, blue = up). (Reprinted from Dunin-Borkowski et al., 2004b.)

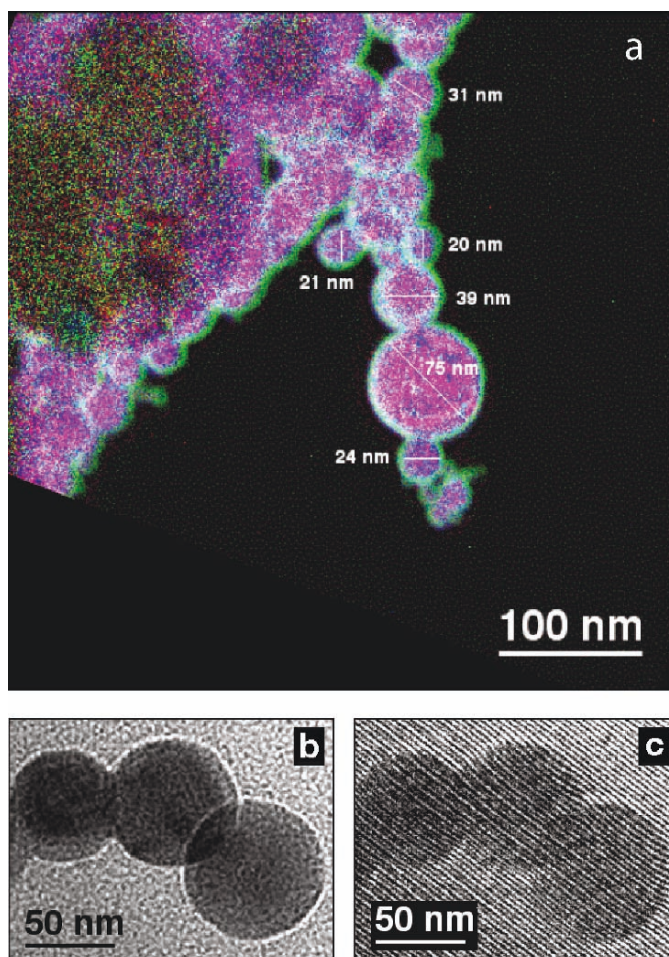


Figure 18–10. (a) Chemical map of $\text{Fe}_{0.56}\text{Ni}_{0.44}$ nanoparticles, obtained using three-window background-subtracted elemental mapping with a Gatan imaging filter, showing Fe (red), Ni (blue), and O (green). (b) Bright-field image and (c) electron hologram of the end of a chain of $\text{Fe}_{0.56}\text{Ni}_{0.44}$ particles. The hologram was recorded using an interference fringe spacing of 2.6 nm. (Reprinted from Dunin-Borkowski et al., 2004b.)

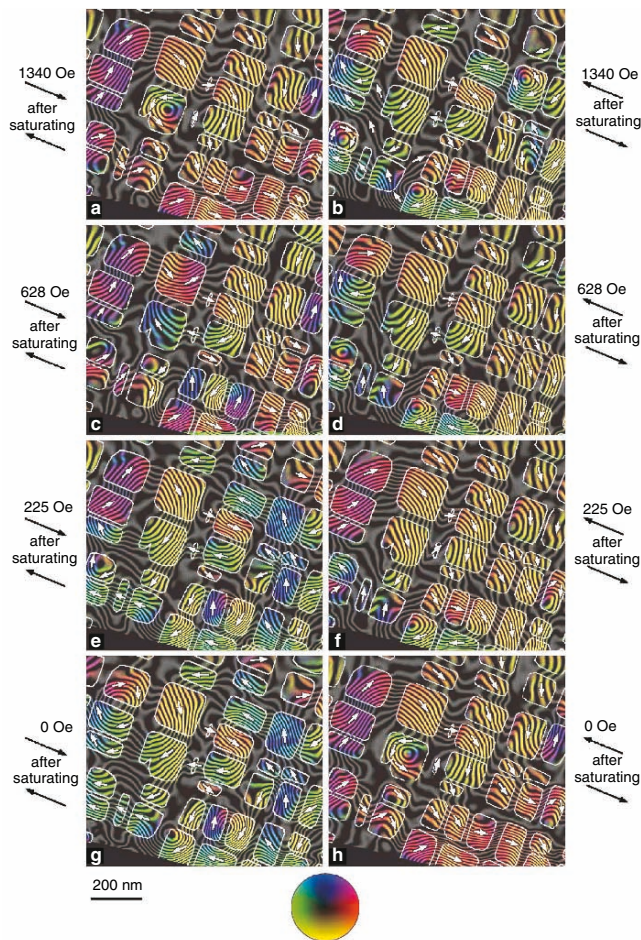


Figure 18-13. Magnetic phase contours from the region shown in Figure 8-12, measured using electron holography. Each image was acquired with the specimen in magnetic field-free conditions. The outlines of the magnetite-rich regions are marked in white, while the direction of the measured magnetic induction is indicated both using arrows and according to the color wheel shown at the bottom of the figure (red = right, yellow = down, green = left, blue = up). Images (a), (c), (e), and (g) were obtained after applying a large ($>10,000$ Oe) field toward the top left, then the indicated field toward the bottom right, after which the external magnetic field was removed for hologram acquisition. Images (b), (d), (f), and (h) were obtained after applying identical fields in the opposite directions. (Reprinted from Harrison et al., 2002.)

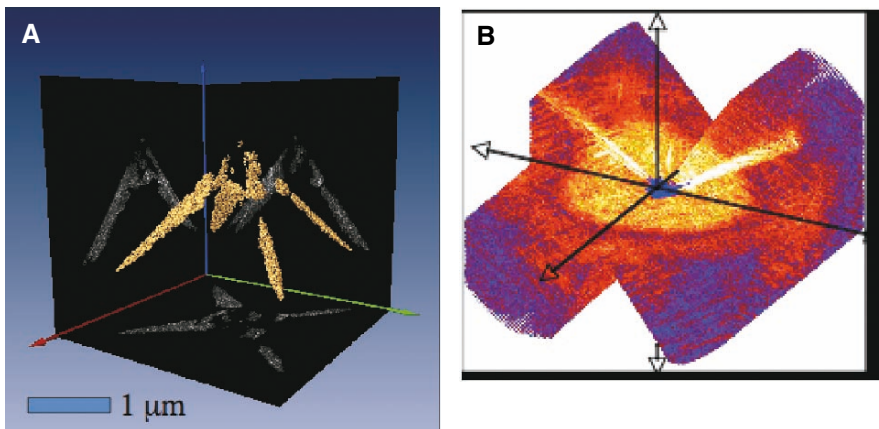


Figure 19-5. (A) Tomographic reconstruction from a soft X-ray diffraction pattern shown in (B). The object consists of gold balls (50 nm diameter) lying along the edges of a pyramidal-shaped silicon nitride structure. This is one image from a rotation series. From the complete series, three-dimensional surfaces of constant density can be constructed. (B) The volume of soft X-ray diffraction data collected to obtain the three-dimensional reconstruction in (A).