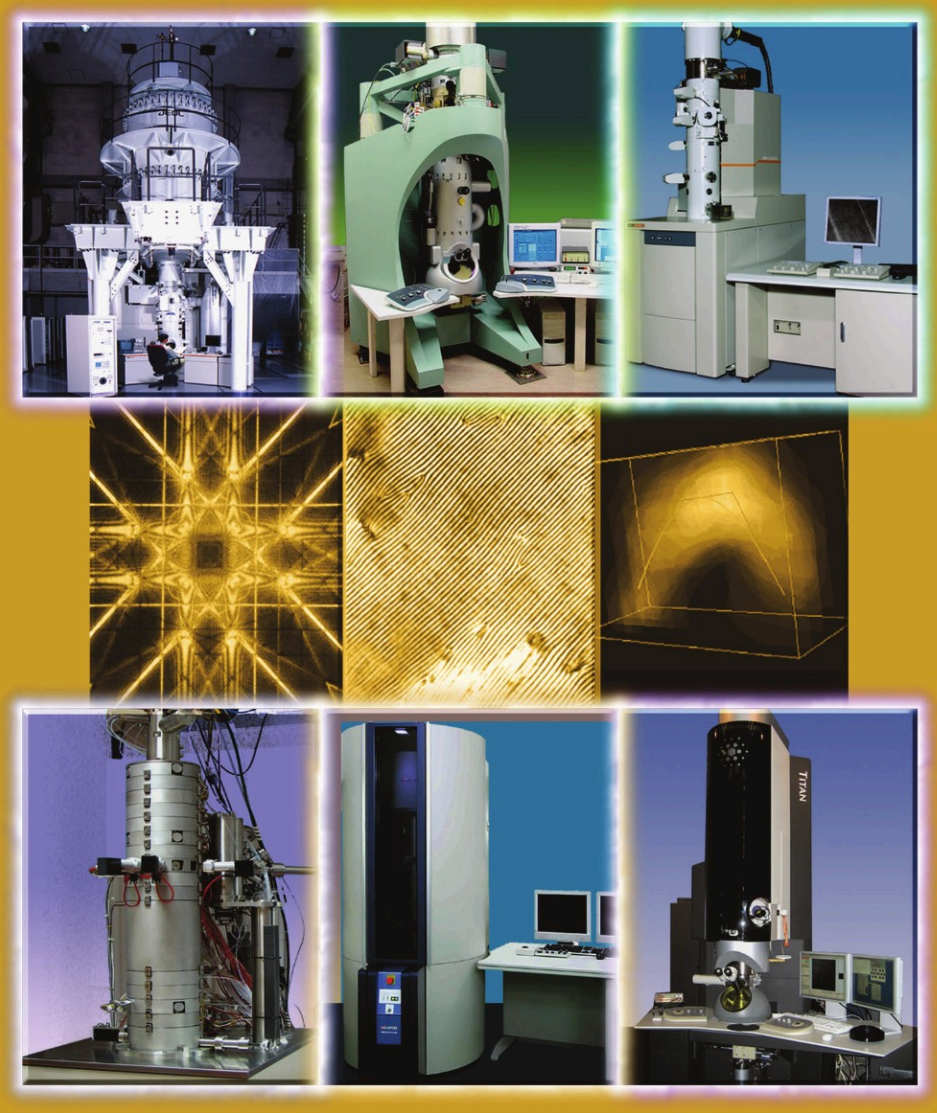
David B. Williams • C. Barry Carter

Transmission Electron Microscopy

Part 1: *Basics*



Second Edition

Springer

Transmission Electron Microscopy

A Textbook for Materials Science

Transmission Electron Microscopy

A Textbook for Materials Science

David B. Williams
C. Barry Carter

Springer

David B. Williams
The University of Alabama in Huntsville
Huntsville AL, USA
david.williams@uah.edu

C. Barry Carter
University of Connecticut
Storrs, CT, USA
cbcarter@engr.uconn.edu

ISBN 978-0-387-76500-6 hardcover
ISBN 978-0-387-76502-0 softcover (This is a four-volume set. The volumes are not sold individually.)
e-ISBN 978-0-387-76501-3

Library of Congress Control Number: 2008941103

© Springer Science+Business Media, LLC 1996, 2009
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.

Printed on acid-free paper

springer.com

To our parents

Walter Dennis and Mary Isabel Carter

and

Joseph Edward and Catherine Williams,
who made everything possible.

About the Authors

David B. Williams

David B. Williams became the fifth President of the University of Alabama in Huntsville in July 2007. Before that he spent more than 30 years at Lehigh University where he was the Harold Chambers Senior Professor Emeritus of Materials Science and Engineering (MS&E). He obtained his BA (1970), MA (1974), PhD (1974) and ScD (2001) from Cambridge University, where he also earned four Blues in rugby and athletics. In 1976 he moved to Lehigh as Assistant Professor, becoming Associate Professor (1979) and Professor (1983). He directed the Electron Optical Laboratory (1980–1998) and led Lehigh's Microscopy School for over 20 years. He was Chair of the MS&E Department from 1992 to 2000 and Vice Provost for Research from 2000 to 2006, and has held visiting-scientist positions at the University of New South Wales, the University of Sydney, Chalmers University (Gothenburg), Los Alamos National

Laboratory, the Max Planck Institut für Metallforschung (Stuttgart), the Office National d'Etudes et Recherches Aérospatiales (Paris) and Harbin Institute of Technology.

He has co-authored and edited 11 textbooks and conference proceedings, published more than 220 refereed journal papers and 200 abstracts/conference proceedings, and given 275 invited presentations at universities, conferences and research laboratories in 28 countries.

Among numerous awards, he has received the Burton Medal of the Electron Microscopy Society of America (1984), the Heinrich Medal of the US Microbeam Analysis Society (MAS) (1988), the MAS Presidential Science Award (1997) and was the first recipient of the Duncumb award for excellence in microanalysis (2007). From Lehigh, he received the Robinson Award (1979), the Libsch Award (1993) and was the Founders Day commencement speaker (1995). He has organized many national and international microscopy and analysis meetings including the 2nd International MAS conference (2000), and was co-chair of the scientific program for the 12th International Conference on Electron Microscopy (1990). He was an Editor of *Acta Materialia* (2001–2007) and the *Journal of Microscopy* (1989–1995) and was President of MAS (1991–1992) and the International Union of Microbeam Analysis Societies (1994–2000). He is a Fellow of The Minerals Metals and Materials Society (TMS), the American Society for Materials (ASM) International, The Institute of Materials (UK) (1985–1996) and the Royal Microscopical Society (UK).

C. Barry Carter

C. Barry Carter became the Head of the Department of Chemical, Materials & Biomolecular Engineering at the University of Connecticut in Storrs in July 2007. Before that he spent 12 years (1979–1991) on the Faculty at Cornell University in the Department of Materials Science and Engineering (MS&E) and 16 years as the 3 M

Heltzer Multidisciplinary Chair in the Department of Chemical Engineering and Materials Science (CEMS) at the University of Minnesota. He obtained his BA (1970), MA (1974) and ScD (2001) from Cambridge University, his MSc (1971) and DIC from Imperial College, London and his DPhil (1976) from Oxford University. After a postdoc in Oxford with his thesis advisor, Peter Hirsch, in 1977 he moved to Cornell initially as a postdoctoral fellow, becoming an Assistant Professor (1979), Associate Professor (1983) and Professor (1988) and directing the Electron Microscopy Facility (1987–1991). At Minnesota, he was the Founding Director of the High-Resolution Microscopy Center and then the Associate Director of the Center for Interfacial Engineering; he created the Characterization Facility as a unified facility including many forms of microscopy and diffraction in one physical location. He has held numerous visiting scientist positions: in the United States at the Sandia National Laboratories, Los Alamos National Laboratory and Xerox PARC; in Sweden at Chalmers University (Gothenburg); in Germany at the Max Planck Institut für Metallforschung (Stuttgart), the Forschungszentrum Jülich, Hannover University and IFW (Dresden); in France at ONERA (Chatillon); in the UK at Bristol University and at Cambridge University (Peterhouse); and in Japan at the ICYS at NIMS (Tsukuba).

He is the co-author of two textbooks (the other is *Ceramic Materials; Science & Engineering* with Grant Norton) and co-editor of six conference proceedings, and has published more than 275 refereed journal papers and more than 400 extended abstracts/conference proceedings. Since 1990 he has given more than 120 invited presentations at universities, conferences and research laboratories. Among numerous awards, he has received the Simon Guggenheim Award (1985–1986), the Berndt Matthias Scholar Award (1997/1998) and the Alexander von Humboldt Senior Award (1997). He organized the 16th International Symposium on the Reactivity of Solids (ISRS-16 in 2007). He was an Editor of the *Journal of Microscopy* (1995–1999) and of *Microscopy and Microanalysis* (2000–2004), and became (co-)Editor-in-Chief of the *Journal of Materials Science* in 2004. He was the 1997 President of MSA, and served on the Executive Board of the International Federation of Societies for Electron Microscopy (IFSEM; 1999–2002). He is now the General Secretary of the International Federation of Societies for Microscopy (IFSM; 2003–2010). He is a Fellow of the American Ceramics Society (1996) the Royal Microscopical Society (UK), the Materials Research Society (2009) and the Microscopy Society of America (2009).

Preface

How is this book different from the many other TEM books? It has several unique features but what we think distinguishes it from all other such books is that it is truly a *textbook*. We wrote it to be read by, and taught to, senior undergraduates and starting graduate students, rather than studied in a research laboratory. We wrote it using the same style and sentence construction that we have used in countless classroom lectures, rather than how we have written our countless (and much-less read) formal scientific papers. In this respect particularly, we have been deliberate in *not* referencing the sources of every experimental fact or theoretical concept (although we do include some hints and clues in the chapters). However, at the end of each chapter we have included groups of references that should lead you to the best sources in the literature and help you go into more depth as you become more confident about what you are looking for. We are great believers in the value of history as the basis for understanding the present and so the history of the techniques and key historical references are threaded throughout the book. Just because a reference is dated in the previous century (or even the antepenultimate century) doesn't mean it isn't useful! Likewise, with the numerous figures drawn from across the fields of materials science and engineering and nanotechnology, we do not reference the source in each caption. But at the very end of the book each of our many generous colleagues whose work we have used is clearly acknowledged.

The book consists of 40 relatively small chapters (with a few notable Carter exceptions!). The contents of most of the chapters can be covered in a typical lecture of 50-75 minutes (especially if you talk as fast as Williams). Furthermore, each of the four softbound volumes is flexible enough to be usable at the TEM console so you can check what you are seeing against what you should be seeing. Most importantly perhaps, the softbound version is cheap enough for all serious students to buy. So we hope you won't have to try and work out the meaning of the many complex color diagrams from secondhand B&W copies that you acquired from a former student. We have deliberately used color where it is useful rather than simply for its own sake (since all electron signals are colorless anyhow). There are numerous boxes throughout the text, drawing your attention to key information (green), warnings about mistakes you might easily make (amber), and dangerous practices or common errors (red).

Our approach throughout this text is to answer two fundamental questions:

Why should we use a particular TEM technique?
How do we put the technique into practice?

In answering the first question we attempt to establish a sound theoretical basis where necessary although not always giving all the details. We use this knowledge to answer the second question by explaining operational details in a generic sense and showing many illustrative figures. In contrast, other TEM books tend to be either strongly theoretical or predominantly descriptive (often covering more than just TEM). We view our approach as a compromise between the two extremes, covering enough theory to be reasonably rigorous without incurring the wrath of electron physicists yet containing sufficient hands-on instructions and practical examples to be useful to the materials engineer/nanotechnologist who wants an answer to a

materials problem rather than just a set of glorious images, spectra, and diffraction patterns. We acknowledge that, in attempting to seek this compromise, we often gloss over the details of much of the physics and math behind the many techniques but contend that the content is usually approximately right (even if on occasions, it might be precisely incorrect!).

Since this text covers the whole field of TEM we incorporate, to varying degrees, *all* the capabilities of the various kinds of current TEMs and we attempt to create a coherent view of the many aspects of these instruments. For instance, rather than separating out the broad-beam techniques of a traditional TEM from the focused-beam techniques of an analytical TEM, we treat these two approaches as different sides of the same coin. There is no reason to regard 'conventional' bright-field imaging in a parallel-beam TEM as being more fundamental (although it is certainly a more-established technique) than annular dark-field imaging in a focused-beam STEM. Convergent beam, scanning beam, and selected-area diffraction are likewise integral parts of the whole of TEM diffraction.

However, in the decade and more since the first edition was published, there has been a significant increase in the number of TEM and related techniques, greater sophistication in the microscope's experimental capabilities, astonishing improvements in computer control of the instrument, and new hardware designs and amazing developments in software to model the gigabytes of data generated by these almost-completely digital instruments. Much of this explosion of information has coincided with the worldwide drive to explore the nanoworld, and the still-ongoing effects of Moore's law. It is not possible to include all of this new knowledge in the second edition without transforming the already doorstop sized text into something capable of halting a large projectile in its tracks. It is still essential that this second edition teaches you to understand the essence of the TEM before you attempt to master the latest advances. But we personally cannot hope to understand fully all the new techniques, especially as we both descend into more administrative positions in our professional lives. Therefore, we have prevailed on almost 20 of our close friends and colleagues to put together with us a companion text (TEM; a companion text, Williams and Carter (Eds.) Springer 2010) to which we will refer throughout this second edition. The companion text is just as it says—it's a friend whose advice you should seek when the main text isn't enough. The companion is not necessarily more advanced but is certainly more detailed in dealing with key recent developments as well as some more traditional aspects of TEM that have seen a resurgence of interest. We have taken our colleagues' contributions and rewritten them in a similar conversational vein to this main text and we hope that this approach, combined with the in-depth cross-referencing between the two texts will guide you as you start down the rewarding path to becoming a transmission microscopist.

We each bring more than 35 years of teaching and research in all aspects of TEM. Our research into different materials includes metals, alloys, ceramics, semiconductors, glasses, composites, nano and other particles, atomic-level planar interfaces, and other crystal defects. (The lack of polymeric and biological materials in our own research is evident in their relative absence in this book.) We have contributed to the training of a generation of (we hope) skilled microscopists, several of whom have followed us as professors and researchers in the EM field. These students represent our legacy to our beloved research field and we are overtly proud of their accomplishments. But we also expect some combination of these (still relatively young) men and women to write the third edition. We know that they, like us, will find that writing such a text broadens their knowledge considerably and will also be the source of much joy, frustration, and enduring friendship. We hope you have as much fun reading this book as we had writing it, but we hope also that it takes you much less time. Lastly, we encourage you to send us any comments, both positive and negative. We can both be reached by e-mail: david.williams@uah.edu and cbcarter@engr.uconn.edu.

Foreword to First Edition

Electron microscopy has revolutionized our understanding of materials by completing the *processing-structure-properties* links down to atomistic levels. It is now even possible to tailor the microstructure (and mesostructure) of materials to achieve specific sets of properties; the extraordinary abilities of modern transmission electron microscopy—TEM—instruments to provide almost all the structural, phase, and crystallographic data allow us to accomplish this feat. Therefore, it is obvious that any curriculum in modern materials education must include suitable courses in electron microscopy. It is also essential that suitable texts be available for the preparation of the students and researchers who must carry out electron microscopy properly and quantitatively.

The 40 chapters of this new text by Barry Carter and David Williams (like many of us, well schooled in microscopy at Cambridge and Oxford) do just that. If you want to learn about electron microscopy from specimen preparation (the ultimate limitation); or via the instrument; or how to use the TEM correctly to perform imaging, diffraction, and spectroscopy—it's all there! This, to my knowledge, is the only complete text now available that includes all the remarkable advances made in the field of TEM in the past 30 to 40 years. The timing for this book is just right and, personally, it is exciting to have been part of the development it covers—developments that have impacted so heavily on materials science.

In case there are people out there who still think TEM is just taking pretty pictures to fill up one's bibliography, please stop, pause, take a look at this book, and digest the extraordinary intellectual demands required of the microscopist in order to do the job properly: crystallography, diffraction, image contrast, inelastic scattering events, and spectroscopy. Remember, these used to be fields in themselves. Today, one has to understand the fundamentals of *all* these areas before one can hope to tackle significant problems in materials science. TEM is a technique of characterizing materials down to the atomic limits. It must be used with care and attention, in many cases involving teams of experts from different venues. The fundamentals are, of course, based in physics, so aspiring materials scientists would be well advised to have prior exposure to, for example, solid-state physics, crystallography, and crystal defects, as well as a basic understanding of materials science, for without the latter, how can a person see where TEM can (or may) be put to best use?

So much for the philosophy. This fine new book definitely fills a gap. It provides a sound basis for research workers and graduate students interested in exploring those aspects of structure, especially defects, that control properties. Even undergraduates are now expected (and rightly) to know the basis for electron microscopy, and this book, or appropriate parts of it, can also be utilized for undergraduate curricula in science and engineering.

The authors can be proud of an enormous task, very well done.

G. Thomas
Berkeley, California

Foreword to Second Edition

This book is an exciting entry into the world of atomic structure and characterization in materials science, with very practical instruction on how you can see it and measure it, using an electron microscope. You will learn an immense amount from it, and probably want to keep it for the rest of your life (particularly if the problems cost you some effort!).

Is nanoscience "the next industrial revolution"? Perhaps that will be some combination of energy, environmental and nanoscience. Whatever it is, the new methods which now allow control of materials synthesis at the atomic level will be a large part of it, from the manufacture of jet engine turbine-blades to that of catalysts, polymers, ceramics and semiconductors. As an exercise, work out how much reduction would result in the transatlantic airfare if aircraft turbine blade temperatures could be increased by 200°C. Now calculate the reduction in CO_2 emission, and increased efficiency (reduced coal use for the same amount of electricity) resulting from this temperature increase for a coal-fired electrical generating turbine. Perhaps you will be the person to invent these urgently needed things! The US Department of Energy's Grand Challenge report on the web lists the remarkable advances in exotic nanomaterials useful for energy research, from separation media in fuel cells, to photovoltaics and nano-catalysts which might someday electrolyze water under sunlight alone. Beyond these functional and structural materials, we are now also starting to see for the first time the intentional fabrication of atomic structures in which atoms can be addressed individually, for example, as quantum computers based perhaps on quantum dots. 'Quantum control' has been demonstrated, and we have seen fluorescent nanodots which can be used to label proteins.

Increasingly, in order to find out exactly what new material we have made, and how perfect it is (and so to improve the synthesis), these new synthesis methods must be accompanied by atomic scale compositional and structural analysis. The transmission electron microscope (TEM) has emerged as the perfect tool for this purpose. It can now give us atomic-resolution images of materials and their defects, together with spectroscopic data and diffraction patterns from sub-nanometer regions. The field-emission electron gun it uses is still the brightest particle source in all of physics, so that electron microdiffraction produces the most intense signal from the smallest volume of matter in all of science. For the TEM electron beam probe, we have magnetic lenses (now aberration corrected) which are extremely difficult for our X-ray and neutron competitors to produce (even with much more limited performance) and, perhaps most important of all, our energy-loss spectroscopy provides unrivalled spatial resolution combined with parallel detection (not possible with X-ray absorption spectroscopy, where absorbed X-rays disappear, rather than losing some energy and continuing to the detector).

Much of the advance in synthesis is the legacy of half a century of research in the semiconductor industry, as we attempt to synthesize and fabricate with other materials what is now so easily done with silicon. Exotic oxides, for example, can now be laid down layer by layer to form artificial crystal structures with new, useful properties. But it is also a result of the spectacular advances in materials characterization, and our ability to see structures at the atomic level. Perhaps the best example of this is the discovery of the carbon nanotube, which was first identified by using an electron

microscope. Any curious and observant electron microscopist can now discover new nanostructures just because they look interesting at the atomic scale. The important point is that if this is done in an environmental microscope, he or she will know how to make them, since the thermodynamic conditions will be recorded when using such a 'lab in a microscope'. There are efforts at materials discovery by just such combinatorial trial-and-error methods, which could perhaps be incorporated into our electron microscopes. This is needed because there are often just 'too many possibilities' in nature to explore in the computer — the number of possible structures rises very rapidly with the number of distinct types of atoms.

It was Richard Feynman who said that, "if, in some catastrophe, all scientific knowledge was lost, and only one sentence could be preserved, then the statement to be passed on, which contained the most information in the fewest words, would be that matter consists of atoms." But confidence that matter consists of atoms developed surprisingly recently and as late as 1900 many (including Kelvin) were unconvinced, despite Avagadro's work and Faraday's on electrodeposition. Einstein's Brownian motion paper of 1905 finally persuaded most, as did Rutherford's experiments. Muller was first to see atoms (in his field-ion microscope in the early 1950s), and Albert Crewe two decades later in Chicago, with his invention of the field-emission gun for his scanning transmission electron microscope (STEM). The Greek Atomists first suggested that a stone, cut repeatedly, would eventually lead to an indivisible smallest fragment, and indeed Democritus believed that "nothing exists except vacuum and atoms. All else is opinion." Marco Polo remarks on the use of spectacles by the Chinese, but it was van Leeuwenhoek (1632-1723) whose series of papers in *Phil. Trans.* brought the microworld to the general scientific community for the first time using his much improved optical microscope. Robert Hooke's 1665 Micrographica sketches what he saw through his new compound microscope, including fascinating images of facetted crystallites, whose facet angles he explained with drawings of piles of cannon balls. Perhaps this was the first resurrection of the atomistic theory of matter since the Greeks. Zernike's phase-plate in the 1930s brought phase contrast to previously invisible ultra-thin biological 'phase objects', and so is the forerunner for the corresponding theory in high-resolution electron microscopy.

The past fifty years has been a wonderfully exciting time for electron microscopists in materials science, with continuous rapid advances in all of its many modes and detectors. From the development of the theory of Bragg diffraction contrast and the column approximation, which enables us to understand TEM images of crystals and their defects, to the theory of high-resolution microscopy useful for atomic-scale imaging, and on into the theory of all the powerful analytic modes and associated detectors, such as X-rays, cathodoluminescence and energy-loss spectroscopy, we have seen steady advances. And we have always known that defect structure in most cases controls properties — the most common (first-order) phase transitions are initiated at special sites, and in the electronic oxides a whole zoo of charge-density excitations and defects waits to be fully understood by electron microscopy. The theory of phase-transformation toughening of ceramics, for example, is a wonderful story which combines TEM observations with theory, as does that of precipitate hardening in alloys, or the early stages of semiconductor-crystal growth. The study of diffuse scattering from defects as a function of temperature at phase transitions is in its infancy, yet we have a far stronger signal there than in competing X-ray methods. The mapping of strain-fields at the nanoscale in devices, by quantitative convergent-beam electron diffraction, was developed just in time to solve a problem listed on the Semiconductor Roadmap (the speed of your laptop depends on strain-induced mobility enhancement). In biology, where the quantification of TEM data is taken more seriously, we have seen three-dimensional image reconstructions of many large proteins, including the ribosome (the factory which makes proteins according to DNA instructions). Their work should be a model to the materials science community in the constant effort toward better quantification of data.

Like all the best textbooks, this one was distilled from lecture notes, debugged over many years and generations of students. The authors have extracted the heart from

many difficult theory papers and a huge literature, to explain to you in the simplest, clearest manner (with many examples) the most important concepts and practices of modern transmission electron microscopy. This is a great service to the field and to its teaching worldwide. Your love affair with atoms begins!

J.C.H. Spence
Regent's Professor of Physics
Arizona State University and Lawrence
Berkeley National Laboratory

Acknowledgments

We have spent over 20 years conceiving and writing this text and the preceding first edition and such an endeavor can't be accomplished in isolation. Our first acknowledgment must be to our respective wives and children: Margie, Matthew, Bryn, and Stephen and Bryony, Ben, Adam, and Emily. Our families have borne the brunt of our absences from home (and occasionally the brunt of our presence). Neither edition would have been possible without the encouragement, advice, and persistence of (and the fine wines served by) Amelia McNamara, our first editor at Plenum Press, then Kluwer, and Springer.

We have both been fortunate to work in our respective universities with many more talented colleagues, post-doctoral associates, and graduate students, all of whom have taught us much and contributed significantly to the examples in both editions. We would like to thank a few of these colleagues directly: Dave Ackland, Faisal Alamgir, Arzu Altay, Ian Anderson, Ilke Arslan, Joysurya Basu, Steve Baumann, Charlie Betz, John Bruley, Derrick Carpenter, Helen Chan, Steve Claves, Dov Cohen, Ray Coles, Vinayak Dravid, Alwyn Eades, Shelley Gillis, Jeff Farrer, Joe Goldstein, Pradyumna Gupta, Brian Hebert, Jason Hefflefinger, John Hunt, Yasuo Ito, Matt Johnson, Vicki Keast, Chris Kiely, Paul Kotula, Chunfei Li, Ron Liu, Charlie Lyman, Mike Mallamaci, Stuart McKernan, Joe Michael, Julia Nowak, Grant Norton, Adam Papworth, Chris Perrey, Sundar Ramamurthy, René Rasmussen, Ravi Ravishankar, Kathy Repa, Kathy Reuter, Al Romig, Jag Sankar, David A. Smith, Kamal Soni, Changmo Sung, Caroline Swanson, Ken Vecchio, Masashi Watanabe, Jonathan Winterstein, Janet Wood, and Mike Zemyan.

In addition, many other colleagues and friends in the field of microscopy and analysis have helped with the book (even if they weren't aware of it). These include Ron Anderson, Raghavan Ayer, Jim Bentley, Gracie Burke, Jeff Campbell, Graham Cliff, David Cockayne, Peter Doig, the late Chuck Fiori, Peter Goodhew, Brendan Griffin, Ron Gronsky, Peter Hawkes, Tom Huber, Gilles Hug, David Joy, Mike Kersker, Roar Kilaas, Sasha Krajnikov, the late Riccardo Levi-Setti, Gordon Lorimer, Harald Müllejans, Dale Newbury, Mike O'Keefe, Peter Rez, Manfred Rühle, John-Henry Scott, John Steeds, Peter Swann, Gareth Thomas, Patrick Veyssière, Peter Williams, Nestor Zaluzec, and Elmar Zeitler. Many of these (and other) colleagues provided the figures that we acknowledge individually at the end of the book.

We have received financial support for our microscopy studies through several different federal agencies; without this support none of the research that underpins the contents of this book would have been accomplished. In particular, DBW wishes to thank the National Science Foundation, Division of Materials Research for over 30 years of continuous funding, NASA, Division of Planetary Science (with Joe Goldstein) and The Department of Energy, Basic Energy Sciences (with Mike Notis and Himanshu Jain), Bettis Laboratories, Pittsburgh, and Sandia National Laboratories, Albuquerque. While this edition was finalized at the University of Alabama in Huntsville, both editions were written while DBW was in the Center for Advanced Materials and Nanotechnology at Lehigh University, which supports that outstanding electron microscopy laboratory. Portions of both editions were written while DBW was on sabbatical or during extended visits to various microscopy labs: Chalmers University, Göteborg, with Gordon Dunlop and Hans Nordén; The Max Planck

Institut für Metallforschung, Stuttgart, with Manfred Rühle; Los Alamos National Laboratory with Terry Mitchell; Dartmouth College, Thayer School of Engineering, with Erland Schulson; and the Electron Microscope Unit at Sydney University with Simon Ringer. CBC wishes to acknowledge the Department of Energy, Basic Energy Sciences, the National Science Foundation, Division of Materials Research, the Center for Interfacial Engineering at the University of Minnesota, The Materials Science Center at Cornell University, and the SHaRE program at Oak Ridge National Laboratories. The first edition was started while CBC was with the Department of Materials Science and Engineering at Cornell University. This edition was started at the Department of Chemical Engineering and Materials Science at the University of Minnesota where the first edition was finished and was finalized while CBC was at the University of Connecticut. The second edition was partly written while CBC was on Sabbatical Leave at Chalmers University with Eva Olssen (thanks also to Anders Tholen at Chalmers), at NIMS in Tsukuba with Yoshio Bando (thanks also to Dmitri Golberg and Kazuo Furuya at NIMS at Yuichi Ikuhara at the University of Tokyo) and at Cambridge University with Paul Midgley. CBC also thanks the Master and Fellows of Peterhouse for their hospitality during the latter period.

CBC would also like to thank the team at the Ernst Ruska Center for their repeated generous hospitality (special thanks to Knut Urban, Markus Lenzen, Andreas Thust, Martina Luysberg, Karsten Tillmann, Chunlin Jia and Lothar Houben)

Despite our common scientific beginnings as undergraduates in Christ's College Cambridge, we learned our trade under different microscopists: DBW with Jeff Edington in Cambridge and CBC with Sir Peter Hirsch and Mike Whelan in Oxford. Not surprisingly, the classic texts by these renowned microscopists are referred to throughout this book. They influenced our own views of TEM tremendously, contributing to the undoubted bias in our opinions, notation, and approach to the whole subject.

List of Initials and Acronyms

The field of TEM is a rich source of initials and acronyms (these are words formed by the initials), behind which we hide both simple and esoteric concepts. While the generation of new initials and acronyms can be a source of original thinking (e.g., see ALCHEMI), it undoubtedly makes for easier communication in many cases and certainly reduces the length of voluminous textbooks. You have to master this strange language before being accepted into the community of microscopists, so we present a comprehensive listing that you should memorize.

ACF absorption-correction factor
ACT automated crystallography for TEM
A/D analog to digital (converter)
ADF annular dark field
AEM analytical electron microscope/microscopy
AES Auger electron spectrometer/spectroscopy
AFF aberration-free focus
AFM atomic force microscope/microscopy
ALCHEMI atom location by channeling-enhanced micro-analysis
ANL Argonne National Laboratory
APB anti-phase domain boundary
APFIM atom-probe field ion microscope/microscopy
APW augmented plane wave
ASW augmented spherical wave
ATW atmospheric thin window

BF bright field
BFP back-focal plane
BSE backscattered electron
BZB Brillouin-zone boundary

C1,2 condenser 1, 2, etc. lens
CASTEP electronic-potential calculation software
CAT computerized axial tomography
CB coherent bremsstrahlung
CBED convergent-beam electron diffraction
CBIM convergent beam imaging
CCD charge-coupled device
CCF cross-correlation function
CCM charge-collection microscopy
CDF centered dark field
CF coherent Fresnel/Foucault
CFE cold field emission
CL cathodoluminescence
cps counts per second
CRT cathode-ray tube
CS crystallographic shear
CSL coincident-site lattice
CVD chemical vapor deposition

DADF displaced-aperture dark field
DDF diffuse dark field
DF dark field
DFT density-functional theory
DOS density of states
DP diffraction pattern
DQE detection quantum efficiency
DSTEM dedicated scanning transmission electron microscope/microscopy
DTSA desktop spectrum analyzer

EBIC electron beam-induced current/conductivity
EBSD electron-backscatter diffraction
EELS electron energy-loss spectrometer/spectrometry
EFI energy-filtered imaging
EFTEM energy-filtered transmission electron microscope
ELNES energy-loss near-edge structure
ELP™ energy-loss program (Gatan)
EMMA electron microscope microanalyzer
EMS electron microscopy image simulation
(E)MSA (Electron) Microscopy Society of America
EPMA electron-probe microanalyzer
ESCA electron spectroscopy for chemical analysis
ESI electron-spectroscopic imaging
EXAFS extended X-ray-absorption fine structure
EXELFS extended energy-loss fine structure

FEFF ab-initio multiple-scattering software
FEG field-emission gun
FET field-effect transistor
FFP front-focal plane
FFT fast Fourier transform
FIB focused ion beam
FLAPW full-potential linearized augmented plane wave
FOLZ first-order Laue zone
FTP file-transfer protocol
FWHM full width at half maximum
FWTM full width at tenth maximum

GB grain boundary
GIF Gatan image filter™
GIGO garbage in garbage out
GOS generalized oscillator strength

HAADF high-angle annular dark field
HOLZ higher-order Laue zone
HPGe high-purity germanium
HREELS high-resolution electron energy-loss spectrometer/spectrometry
HRTEM high-resolution transmission electron microscope/microscopy
HV high vacuum
HVEM high-voltage electron microscope/microscopy

ICC incomplete charge collection
ICDD International Center for Diffraction Data
ID identification (of peaks in spectrum)
IDB inversion domain boundary
IEEE International Electronics and Electrical Engineering
IG intrinsic Ge
IVEM intermediate-voltage electron microscope/microscopy

K-M Kossel-Möllenstedt

LACBED large-angle convergent-beam electron diffraction
LCAO linear combination of atomic orbitals
LCD liquid-crystal display
LDA local-density approximation
LEED low-energy electron diffraction
LKKR layered Korringa-Kohn-Rostoker

MAS Microbeam Analysis Society
MBE molecular-beam epitaxy
MC minimum contrast
MCA multichannel analyzer
MDM minimum detectable mass
MLS multiple least-squares
MMF minimum mass fraction
MO molecular orbital
MRS Materials Research Society
MS multiple scattering
MSA multivariate statistical analysis
MSDS material safety data sheets
MT muffin tin
MV megavolt

NCEMSS National Center for Electron Microscopy simulation system
NIH National Institutes of Health
NIST National Institute of Standards and Technology
NPL National Physical Laboratory

OIM orientation-imaging microscopy
OR orientation relationship

PARODI parallel recording of dark-field images
PB phase boundary
P/B peak-to-background ratio
PEELS parallel electron energy-loss spectrometer/spectrometry
PIPS Precision Ion-Polishing System™
PIXE proton-induced X-ray emission
PM photomultiplier
POA phase-object approximation
ppb/m parts per billion/million
PDA photo-diode array
PSF point-spread function
PTS position-tagged spectrometry

QHRTEM quantitative high-resolution transmission electron microscopy

RB translation boundary (yes, it does!)
RDF radial distribution function
REM reflection electron microscope/microscopy
RHEED reflection high-energy electron diffraction

SACT small-angle cleaving technique
SAD(P) selected-area diffraction (pattern)
SCF self-consistent field
SDD silicon-drift detector
SE secondary electron
SEELS serial electron energy-loss spectrometer/spectrometry
SEM scanning electron microscope/microscopy
SESAMe sub-eV sub-Å microscope
SF stacking fault
SHRLI simulated high-resolution lattice images
SI spectrum imaging
SI Système Internationale
SIGMAK K-edge quantification software
SIGMAL L-edge quantification software
SIMS secondary-ion mass spectrometry
S/N signal-to-noise ratio
SOLZ second-order Laue zone
SRM standard reference material
STEM scanning transmission electron microscope/microscopy
STM scanning tunneling microscope/microscopy

TB twin boundary
TEM transmission electron microscope/microscopy
TFE thermal field emission
TMBA too many bloody acronyms

UHV ultrahigh vacuum
URL uniform resource locator
UTW ultra-thin window

V/F voltage to frequency (converter)
VLM visible-light microscope/microscopy
VUV vacuum ultra violet

WB weak beam
WBDF weak-beam dark field
WDS wavelength-dispersive spectrometer/spectrometry
WP whole pattern
WPOA weak-phase object approximation
WWW World Wide Web

XANES X-ray absorption near-edge structure
XEDS X-ray energy-dispersive spectrometer/spectrometry
XPS X-ray photoelectron spectrometer/spectrometry
XRD/F X-ray diffraction/fluorescence

YAG yttrium-aluminum garnet
YBCO yttrium-barium-copper oxide
YSZ yttria-stabilized zirconia

ZAF atomic number/absorption/fluorescence correction
ZAP zone-axis pattern
ZLP zero-loss peak
ZOLZ zero-order Laue zone

List of Symbols

We use a large number of symbols. Because we are constrained by the limits of our own and the Greek alphabets, we often use the same symbol for different terms, which can confuse the unwary. We have tried to be consistent where possible but undoubtedly we have not always succeeded. The following (not totally inclusive) list may help if you remain confused after reading the text.

a	interatomic spacing
a	relative transition probability
a	width of diffraction disk
a_0	Bohr radius
a_0	lattice parameter
a, b, c	lattice vectors
a*, b*, c*	reciprocal-lattice vectors
A	absorption-correction factor
A	active area of X-ray detector
A_0	amplitude
A	amplitude of scattered beam
A	amperes
A	atomic weight
A	Richardson's constant
Å	Ångstrom
$\mathcal{A}$	Bloch wave amplitude
$A(\mathbf{u})$	aperture function
A, B	fitting parameters for energy-loss background subtraction
b	beam-broadening parameter
b	separation of diffraction disks
$\mathbf{b}_e$	edge component of the Burgers vector
$\mathbf{b}_p$	Burgers vector of partial dislocation
$\mathbf{b}_T$	Burgers vector of total dislocation
B	beam direction
B	magnetic field strength
B	background intensity
$B(\mathbf{u})$	aberration function
c	centi
c	velocity of light
C	composition
C	contrast
C	coulomb
C_a	astigmatism-aberration coefficient
C_c	chromatic-aberration coefficient
$C_\mathbf{g}$	**g** component of Bloch wave
C_s	spherical-aberration coefficient
C_X	fraction of X atoms on specific sites
C_0	amplitude of direct beam
C_ε	combination of the elastic constants
$(C_s\lambda)^{1/2}$	scherzer
$(C_s\lambda^3)^{1/4}$	glaser
c/o	condenser /objective
d	beam (probe) diameter
d	diameter of spectrometer entrance aperture
d	interplanar spacing

d	spacing of moire fringes
d_c	effective source size
d_d	diffraction-limited beam diameter
d_{eff}	effective entrance-aperture diameter at recording plane
d_g	Gaussian beam diameter
d_{hkl}	*hkl* interplanar spacing
d_i	image distance
d_{im}	smallest resolvable image distance
d_o	object distance
d_{ob}	smallest resolvable object distance
d_s	spherical-aberration limited beam diameter
d_t	total beam diameter
dz	thickness of a diffracting slice
$d\sigma/d\Omega$	differential cross section of one atom
D	aperture diameter
D	change in focus
D	dimension (as in 1D, 2D...)
D	distance from projector crossover to recording plane
D	electron dose
D_A	distance from beam crossover to spectrometer entrance aperture
D_{im}	depth of focus
D_{ob}	depth of field
D_1, D_2	tie-line points on dispersion surfaces in presence of defect
e	charge on the electron
E	energy
E	electric-field strength
E	Young's modulus
E	total energy
$\mathcal{E}$	energy loss
E_a	spatial-coherence envelope
E_c	chromatic-coherence envelope
E_c	critical ionization energy
E_d	displacement energy
E_F	Fermi energy/level
$E_{h/l}$	high/low energy for background-subtraction window
$E_{K/L/M}$	ionization energy for K/L/M-shell electron
$E_{K/L/M}$	energy of K/L/M X-ray
$\mathcal{E}_m$	average energy loss
E_P	plasmon energy
$\mathcal{E}_P$	plasmon energy loss
E_s	sputtering-threshold energy
E_t	threshold energy
E_0	beam energy
$E(\mathbf{u})$	envelope function
$E_c(\mathbf{u})$	envelope function for chromatic aberration
$E_d(\mathbf{u})$	envelope function for specimen drift
$E_D(\mathbf{u})$	envelope function for the detector
$E_s(\mathbf{u})$	envelope function for the source
$E_v(\mathbf{u})$	envelope function for specimen vibration
f	focal length
$f(\mathbf{r})$	strength of object at point (x,y)
$f(\theta)$	atomic-scattering factor
$f(\mathbf{k})$	atomic-scattering amplitude
f_x	scattering factor for X-rays
$f_i(x)$	residual of least-squares fit
F	Fano factor
F	fluorescence-correction factor
F	Lorentz force
F	relativistic-correction factor
F	Fourier transform
F'	Fourier transform of edge intensity
F_B	fraction of alloying element B
$F_\mathbf{g}$	special value of $F(\theta)$ when θ is the Bragg angle
$F(P)$	Fourier transform of plasmon intensity
$F(\mathbf{u})$	Fourier transform of $f(\mathbf{r})$
$F(0)$	Fourier transform of elastic intensity
$F(1)$	Fourier transform of single-scattering intensity
$F(\theta)$	structure factor
$\mathbf{g}/\bar{\mathbf{g}}$	diffraction vector (magnitude of +/- **K** at the Bragg angle)
$\mathbf{g}_{hkl}$	diffraction vector for *hkl* plane
g	gram
$g(\mathbf{r})$	intensity of image at point (x,y)
G	Bragg reflection
G	radius of a HOLZ ring
G	giga
$G(\mathbf{u})$	Fourier transform of $g(\mathbf{r})$
Gy	gray (radiation unit)
h	Planck's constant
h	distance from specimen to the aperture
$h(\mathbf{r})$	contrast-transfer function
(hkl)	Miller indices of a crystal plane
hkl	indices of diffraction spots from *hkl* plane
H	spacing of the reciprocal-lattice planes parallel to beam
$H(\mathbf{u})$	Fourier transform of $h(\mathbf{r})$
i	beam current
i	imaginary number
i	number of atoms in unit cell
I	intensity
I	intrinsic line width of the XEDS detector
i_e	emission current

i_f	filament-heating current
I_g	intensity in the diffracted beam
$I_{K/L/M}$	K/L/M-shell intensity above background
$I(\mathbf{k})$	kinematical intensity
$I(1)$	single-scattering intensity
I_P	intensity in the first plasmon peak
I_T	total transmitted intensity
I_0	intensity in the zero-loss peak
$\boldsymbol{I_0}$	intensity in the direct beam
$I(t)$	low-loss spectrum intensity
J	current density
J	joule
J	sum of spin and angular quantum numbers
k	magnitude of the wave vector
k	Boltzmann's constant
k	kilo
$\mathbf{k}_I$	**k**-vector of the incident wave
$\mathbf{k}_D$	**k**-vector of the diffracted wave
k_{AB}	Cliff-Lorimer factor/sensitivity factor
K	bulk modulus
K	Kelvin
K	Kramers' constant
K	sensitivity factor
$K/L/M$	inner-shell/characteristic X-ray/ ionization edge
$\mathbf{K}$	change in **k** due to diffraction
$\mathbf{K}_B$	magnitude of **K** at the Bragg angle
K_o	kernel
l	angular quantum number
L	camera length
L	lattice spacing in beam direction
L	length of magnetic field
L_0	length of magnetic field along optic axis
L	path difference
L	width of composition line-profile
m	meters
m	milli
m	mirror plane
m	number of focal increments
m_0	rest mass of the electron
M	magnification
M	mega
M_A	angular magnification
M_T	transverse magnification
M_1, M_2	tie-line points on dispersion surfaces
n	integer
n	free-electron density
n	number of counts
n	number of scattered electrons
n_0	number of incident electrons
n	nano
n	principal quantum number
$\mathbf{n}$	vector normal to the surface
n_s	number of electrons in the ionized sub-shell
N	$h + k + l$
N	newton
N	noise
N	number of counts in ionization edge
N	number of atoms/unit area
N_V	number of atoms/unit volume
$N(E)$	number of bremsstrahlung photons of energy E
N_0	Avogadro's number
O	direct beam
p	integer
$\mathbf{p}$	momentum
p	pico
P	probability of scattering
P	peak intensity
P	FWHM of a randomized electronic-pulse generator
Pa	pascal
$P_{K/L/M}$	probability of K/L/M-shell ionization
$P(z)$	scattering matrix for a slice of thickness z
q	charge
Q	cross section
r	radius
r	distance a wave propagates
r	distance between contamination spots
r	minimum resolvable distance/resolution
r	power term to fit background in EEL spectrum
r_M	image-translation distance
$\mathbf{r}_n$	lattice vector
$\mathbf{r}^*$	reciprocal-lattice vector
r_{ast}	radius of astigmatism disk
r_{chr}	radius of chromatic-aberration disk
r_{sph}	radius of spherical-aberration disk
r_{min}	minimum disk radius
r_{th}	theoretical disk radius
$\mathbf{r}'_n$	lattice vector in strained crystal
r_0	maximum radius of DP in focal plane of spectrometer
R	ALCHEMI intensity ratio
R	count rate
$\mathbf{R}$	crystal-lattice vector

R	distance on screen between diffraction spots
R	radius of curvature of EEL spectrometer
R	resolution of XEDS detector
R	spatial resolution
R	reduction in partial cross section with increasing α
R_{MAX}	diameter of beam emerging from specimen
$\mathbf{R}_n$	lattice-displacement vector
$\mathbf{R}(\mathbf{r})$	displacement
$\mathbf{s}$	excitation error/deviation parameter
s	second
s	spin quantum number
$\mathbf{s_R}$	excitation error due to defect
$\mathbf{s}_z(\mathbf{s_g})$	excitation error
s_{eff}	effective excitation error
S	distance from specimen to detector
S	signal
S	standard deviation for n measurements
sr	steradians
$\mathbf{t}$	shift vector between the ZOLZ and HOLZ
t	student (t) distribution
t	thickness
t'	absorption path length
t_0	thickness at zero tilt
T	absolute temperature
T	tesla
T_c	period of rotation
$T(\mathbf{u})$	objective-lens transfer function
$T_{eff}(\mathbf{u})$	effective transfer function
$\mathbf{u}$	reciprocal lattice vector
$\mathbf{u}$	unit vector along the dislocation line
$\mathbf{u}^*$	vector normal to the ZOLZ
U	overvoltage
U_g	Fourier component of the perfect-crystal potential
$[UVW]$	indices of a crystal direction
UVW	indices of beam direction
v	velocity
V	accelerating voltage
$\mathcal{V}$	potential energy
V_c/V	volume of the unit cell
V_c	inner potential of cavity
V_t	projected potential through specimen thickness
$V(\mathbf{r})$	crystal inner potential
w	$s\xi_g$ (excitation error multiplied by extinction distance)
w	projected width of planar defect
w	width
x	distance
$\times$	times (magnification)
x, y, z	atom coordinates
X	FWHM due to XEDS detector
X	rotation axis
y	displacement at the specimen
y	number of counts in channel
y	parallax shift in the image
z	distance within a specimen
z	distance along optic axis
z	specimen height
Z	atomic number/atomic-number correction factor

Greek symbols

α	phase shift due to defect
α	semi-angle of incidence/convergence
α	X-ray take-off angle
α_0	beam divergence semi-angle at gun crossover
α_{opt}	optimum convergence semi-angle
β	brightness
β	ratio of electron velocity to light velocity
β	semi-angle of collection
β_{opt}	optimum collection semi-angle
γ	degree of spatial coherence
γ	phase of direct beam
γ	relativistic-correction factor
γ	specimen tilt angle
Δ	change/difference
Δ	width of energy window
Δd	change in lattice parameter
$\Delta\phi$	phase difference
$\Delta\theta_i$	angle between Kossel–Möllenstedt fringes
Δ_{AB}	difference in mass-absorption coefficients
ΔE	energy width /spread
ΔE_P	plasmon-line width/change in plasmon energy
Δf	maximum difference in focus
Δf	defocus error due to chromatic aberration
Δf_{AFF}	aberration-free (de)focus
Δf_{MC}	minimum contrast defocus
Δf_{opt}	optimum defocus
Δ	change (in height)

Δh	relative depth in specimen
ΔI	change in intensity
Δp	parallax shift
ΔV	change in the inner potential
Δx	path difference/image shift
Δx	half-width of image of undissociated screw dislocation
Δx_{res}	resolution at Scherzer defocus
Δf_{sch}	Scherzer defocus
δ	angle between XEDS detector normal and line from detector to specimen
δ	angle between beam and plane of defect
δ	diameter of disk image
δ	diffuseness of interface
δ	precipitate/matrix misfit
δ	small increment
δ	smallest resolvable distance (resolution)
ε	deflection angle
ε	detector efficiency
ε	energy to create an electron-hole pair
ε	specimen-tilt angle
ε	strain
ε_0	permittivity of free space (dielectric constant)
η	phase change
η	angle between excess Kikuchi lines at $\mathbf{s} = 0$ and $\mathbf{s} > 0$
$\eta(\theta)$	phase of the atomic-scattering factor
Φ	phase shift accompanying scattering
Φ	work function
ϕ	rotation angle between image and diffraction pattern
ϕ	angle between Kikuchi line and diffraction spot
ϕ	angle between two Kikuchi-line pairs
ϕ	angle between two planes
ϕ	angle between two plane normals
ϕ	angle of tilt between stereo images
ϕ	phase of a wave
ϕ^*	complex conjugate of ϕ
$\phi_{\mathbf{g}}$	amplitude of the diffracted beam
ϕ_0	amplitude of the direct beam
ϕ_{x}	angle of deflection of the beam
$\phi(\rho t)$	depth distribution of X-ray production
$\boldsymbol{\chi}$	wave vector outside the specimen
$\boldsymbol{\chi}_{\mathrm{G}}$	wave vector terminating on the point G in reciprocal space
$\boldsymbol{\chi}_{\mathrm{O}}$	wave vector terminating on the point O in reciprocal space
χ^2	goodness of fit (between standard and experimental spectra)
$\chi(\mathbf{u})$	phase-distortion function
$\chi(\mathbf{k})$	momentum transfer
κ	thermal conductivity
$\xi_{\mathbf{g}}$	extinction distance for the diffracted beam
$\xi_{\mathbf{g}'}$	absorption parameter
ξ_0	extinction distance for the direct beam
ξ_{eff}	effective extinction distance ($s \neq 0$)
$\xi_{\mathbf{g}}^{\mathrm{abs}}$	absorption-modified ξ_{g}
λ_{c}	coherence length
λ	mean-free path
λ	wavelength
$\lambda_{\mathrm{K/L/M}}$	mean-free path for K/L/M-shell ionization
λ_{P}	plasmon mean-free path
λ_{R}	relativistic wavelength
λ^{-1}	radius of Ewald sphere
μ	micro
μ	refractive index
μ/ρ	mass-absorption coefficient
$\mu^{(j)}(\mathbf{r})$	Bloch function
ν	frequency
ν	Poisson's ratio
ψ	amplitude of a wave
ψ	the wave function
ψ_{sph}	amplitude of spherical wave
ψ_{tot}	total wave function
ψ_0	amplitude
ρ	angle between two directions
ρ	density
$\rho_{\mathrm{c/s}}$	information limit due to chromatic/spherical aberration
$\rho(\boldsymbol{r})$	radial distribution function
ρt	mass thickness
ρ_i^2	area of a pixel
σ	scattering cross section of one atom
σ	standard deviation
σ	stress
$\sigma_{\mathrm{K/L/M}}$	ionization cross section for K/L/M-shell electron
σ_{T}	total ionization cross section
$\sigma_{\mathrm{K/L/M}}(\beta\Delta)$	partial ionization cross section for K/L/M-shell electron
θ	scattering semi-angle

θ_B	Bragg angle
θ_C	cut-off semi-angle
θ_E	characteristic scattering semi-angle
θ_0	screening parameter
τ	XEDS detector time constant
τ	dwell time
τ	analysis time
ω	fluorescence yield
ω_c	cyclotron frequency
ω_p	plasmon frequency
Ω	filter for energy loss
Ω	solid angle of collection of XEDS
Ω	volume of unit cell
ζ	zeta factor
$\otimes$	convolution (multiply and integrate)

About the Companion Volume

As described in our Preface, the many years since the publication of the first edition have seen a significant increase in the number of TEM (and related) techniques and the sophistication of the microscope's experimental capabilities, as well as new hardware designs, astonishing improvements in computer control of the instrument and amazing developments in software to handle and model the gigabytes of data generated by these (now almost completely digital) instruments. Much of this explosion of information has coincided with the world-wide drive to explore the nanoworld, and the still-ongoing effects of Moore's law. It is not possible to include all of this new knowledge in a textbook, and the primary objective of the second edition is still to teach you to understand the essence of the TEM before you attempt to master the latest advances. We also personally cannot hope to comprehend fully all of the new techniques, especially as we both descend into more administrative positions in our professional lives.

Therefore, we have prevailed on almost 20 of our close friends and colleagues to put together with us a companion applications text (Carter and Williams, eds., Springer, 2010) to which we will refer throughout this second edition. The companion text is just as it says; it's a friend whose advice you should seek when the main text is not enough. The companion is not intended to be more advanced but it certainly provides much more detail on key recent developments and some more traditional aspects of TEM that have seen a resurgence of interest. We have taken our colleagues' contributions and worked with them to produce chapters that are in a similar conversational vein to this main text. While *Transmission Electron Microscopy, Second Edition*, is a completely stand-alone textbook, we think that you will find the cross-referencing between the two texts to be of great value as you continue along the rewarding path of becoming a transmission microscopist.

Contents

About the Authors vii

Preface xi

Foreword to First Edition xiii

Foreword to Second Edition xv

Acknowledgments xix

List of Initials and Acronyms xxi

List of Symbols xxv

About the Companion Volume xxxi

Figure Credits xlix

PART 1 BASICS 1

1 The Transmission Electron Microscope 3

Chapter Preview 3
1.1 What Materials Should We Study in the TEM? 3
1.2 Why Use Electrons? 4
1.2.A An Extremely Brief History 4
1.2.B Microscopy and the Concept of Resolution 5
1.2.C Interaction of Electrons with Matter 7
1.2.D Depth of Field and Depth of focus 8
1.2.E Diffraction 8
1.3 Limitations of the TEM 9
1.3.A Sampling 9
1.3.B Interpreting Transmission Images 9
1.3.C Electron Beam Damage and Safety 10
1.3.D Specimen Preparation 11
1.4 Different Kinds of TEMs 11
1.5 Some Fundamental Properties of Electrons 11
1.6 Microscopy on the Internet/World Wide Web 15
1.6.A Microscopy and Analysis-Related Web Sites 15
1.6.B Microscopy and Analysis Software 15
Chapter Summary 17

2 Scattering and Diffraction **23**

Chapter Preview 23
2.1 Why Are We Interested in Electron Scattering? 23
2.2 Terminology of Scattering and Diffraction 25
2.3 The Angle of Scattering 26
2.4 The Interaction Cross Section and Its Differential 27
2.4.A Scattering from an Isolated Atom 27
2.4.B Scattering from the Specimen 28
2.4.C Some Numbers 28
2.5 The Mean Free Path 28
2.6 How We Use Scattering in the TEM 29
2.7 Comparison to X-ray Diffraction 30
2.8 Fraunhofer and Fresnel Diffraction 30
2.9 Diffraction of Light from Slits and Holes 31
2.10 Constructive Interference 33
2.11 A Word About Angles 34
2.12 Electron-Diffraction Patterns 34
Chapter Summary 36

3 Elastic Scattering **39**

Chapter Preview 39
3.1 Particles and Waves 39
3.2 Mechanisms of Elastic Scattering 40
3.3 Elastic Scattering from Isolated Atoms 41
3.4 The Rutherford Cross Section 41
3.5 Modifications to the Rutherford Cross Section 42
3.6 Coherency of the Rutherford-Scattered Electrons 43
3.7 The Atomic-Scattering Factor 44
3.8 The Origin of $f(\theta)$ 45
3.9 The Structure Factor $F(\theta)$ 46
3.10 Simple Diffraction Concepts 47
3.10.A Interference of Electron Waves; Creation of the Direct and Diffracted Beams 47
3.10.B Diffraction Equations 48
Chapter Summary 49

4 Inelastic Scattering and Beam Damage **53**

Chapter Preview 53
4.1 Which Inelastic Processes Occur in the TEM? 53
4.2 X-ray Emission 55
4.2.A Characteristic X-rays 55
4.2.B Bremsstrahlung X-rays 60
4.3 Secondary-Electron Emission 60
4.3.A Secondary Electrons 60
4.3.B Auger Electrons 61
4.4 Electron-Hole Pairs and Cathodoluminescence (CL) 62
4.5 Plasmons and Phonons 63
4.6 Beam Damage 64
4.6.A Electron Dose 65
4.6.B Specimen Heating 65
4.6.C Beam Damage in Polymers 66
4.6.D Beam Damage in Covalent and Ionic Crystals 66
4.6.E Beam Damage in Metals 66
4.6.F Sputtering 68
Chapter Summary 68

5 Electron Sources **73**

Chapter Preview 73
5.1 The Physics of Different Electron Sources 73
5.1.A Thermionic Emission 74
5.1.B Field Emission 74
5.2 The Characteristics of the Electron Beam 75
5.2.A Brightness 75
5.2.B Temporal Coherency and Energy Spread 76
5.2.C Spatial Coherency and Source Size 77
5.2.D Stability 77
5.3 Electron Guns 77
5.3.A Thermionic Guns 77
5.3.B Field-Emission Guns (FEGs) 80
5.4 Comparison of Guns 81
5.5 Measuring Your Gun Characteristics 82
5.5.A Beam Current 82
5.5.B Convergence Angle 83
5.5.C Calculating the Beam Diameter 83
5.5.D Measuring the Beam Diameter 85
5.5.E Energy Spread 85
5.5.F Spatial Coherency 86
5.6 What kV should You Use? 86
Chapter Summary 87

6 Lenses, Apertures, and Resolution **91**

Chapter Preview 91
6.1 Why Learn About Lenses? 91
6.2 Light Optics and Electron Optics 92
6.2.A How to Draw a Ray Diagram 92
6.2.B The Principal Optical Elements 94
6.2.C The Lens Equation 94
6.2.D Magnification, Demagnification, and Focus 95
6.3 Electron Lenses 96
6.3.A Polepieces and Coils 96
6.3.B Different Kinds of Lenses 97
6.3.C Electron Ray Paths Through Magnetic Fields 99
6.3.D Image Rotation and the Eucentric Plane 100
6.3.E Deflecting the Beam 101
6.4 Apertures and Diaphragms 101
6.5 Real Lenses and their Problems 102
6.5.A Spherical Aberration 103
6.5.B Chromatic Aberration 104
6.5.C Astigmatism 106
6.6 The Resolution of the Electron Lens (and Ultimately of the TEM) 106
6.6.A Theoretical Resolution (Diffraction-Limited Resolution) 107
6.6.B The Practical Resolution Due to Spherical Aberration 108
6.6.C Specimen-Limited Resolution Due to Chromatic Aberration 109
6.6.D Confusion in the Definitions of Resolution 109
6.7 Depth of Focus and Depth of Field 110
Chapter Summary 111

7 How to 'See' Electrons 115

Chapter Preview 115
7.1 Electron Detection and Display 115
7.2 Viewing Screens 116
7.3 Electron Detectors 117
7.3.A Semiconductor Detectors 117
7.3.B Scintillator-Photomultiplier Detectors/TV Cameras 118
7.3.C Charge-Coupled Device (CCD) Detectors 120
7.3.D Faraday Cup 121
7.4 Which Detector Do We Use for which Signal? 122
7.5 Image Recording 122
7.5.A Photographic Emulsions 122
7.5.B Other Image-Recording Methods 124
7.6 Comparison of Scanning Images and Static Images 124
Chapter Summary 125

8 Pumps and Holders 127

Chapter Preview 127
8.1 The Vacuum 127
8.2 Roughing Pumps 128
8.3 High/Ultra High Vacuum Pumps 129
8.3.A Diffusion Pumps 129
8.3.B Turbomolecular Pumps 129
8.3.C Ion Pumps 130
8.3.D Cryogenic (Adsorption) Pumps 130
8.4 The Whole System 130
8.5 Leak Detection 131
8.6 Contamination: Hydrocarbons and Water Vapor 132
8.7 Specimen Holders and Stages 132
8.8 Side-Entry Holders 133
8.9 Top-entry Holders 134
8.10 Tilt and Rotate Holders 134
8.11 In-Situ Holders 135
8.12 Plasma Cleaners 138
Chapter Summary 138

9 The Instrument 141

Chapter Preview 141
9.1 The Illumination System 142
9.1.A TEM Operation Using a Parallel Beam 142
9.1.B Convergent-Beam (S)TEM Mode 143
9.1.C The Condenser-Objective Lens 145
9.1.D Translating and Tilting the Beam 147
9.1.E Alignment of the C2 Aperture 147
9.1.F Condenser-Lens Defects 148
9.1.G Calibration 149
9.2 The Objective Lens and Stage 150
9.3 Forming DPs and Images: The TEM Imaging System 152
9.3.A Selected-Area Diffraction 152
9.3.B Bright-Field and Dark-Field Imaging 155
9.3.C Centered Dark-Field Operation 155
9.3.D Hollow-Cone Diffraction and Dark-Field Imaging 157
9.4 Forming DPs and Images: The STEM Imaging System . . 158

9.4.A Bright-Field STEM Images . . . 159
9.4.B Dark-Field STEM Images . . . 161
9.4.C Annular Dark-Field Images . . . 161
9.4.D Magnification in STEM . . . 161
9.5 Alignment and Stigmation . . . 161
9.5.A Lens Rotation Centers . . . 161
9.5.B Correction of Astigmatism in the Imaging Lenses . 162
9.6 Calibrating the Imaging System . . . 164
9.6.A Magnification Calibration . . . 164
9.6.B Camera-Length Calibration . . . 165
9.6.C Rotation of the Image Relative to the DP . . . 167
9.6.D Spatial Relationship Between Images and DPs . . . 168
9.7 Other Calibrations . . . 168
Chapter Summary . . . 169

10 Specimen Preparation . . . 173

Chapter Preview . . . 173
10.1 Safety . . . 173
10.2 Self-Supporting Disk or Use a Grid? . . . 174
10.3 Preparing a Self-Supporting Disk for Final Thinning . . . 175
10.3.A Forming a Thin Slice from the Bulk Sample . . . 176
10.3.B Cutting the Disk . . . 176
10.3.C Prethinning the Disk . . . 177
10.4 Final Thinning of the Disks . . . 178
10.4.A Electropolishing . . . 178
10.4.B Ion Milling . . . 178
10.5 Cross-Section Specimens . . . 182
10.6 Specimens on Grids/Washers . . . 183
10.6.A Electropolishing—The Window Method for Metals and Alloys . . . 183
10.6.B Ultramicrotomy . . . 183
10.6.C Grinding and Crushing . . . 184
10.6.D Replication and Extraction . . . 184
10.6.E Cleaving and the SACT . . . 186
10.6.F The 90° Wedge . . . 186
10.6.G Lithography . . . 187
10.6.H Preferential Chemical Etching . . . 187
10.7 FIB . . . 188
10.8 Storing Specimens . . . 189
10.9 Some Rules . . . 189
Chapter Summary . . . 191

PART 2 DIFFRACTION . . . 195

11 Diffraction in TEM . . . 197

Chapter Preview . . . 197
11.1 Why Use Diffraction in the TEM? . . . 197
11.2 The TEM, Diffraction Cameras, and the TV . . . 198
11.3 Scattering from a Plane of Atoms . . . 199
11.4 Scattering from a Crystal . . . 200
11.5 Meaning of *n* in Bragg's Law . . . 202
11.6 A Pictorial Introduction to Dynamical Effects . . . 203
11.7 Use of Indices in Diffraction Patterns . . . 204
11.8 Practical Aspects of Diffraction-Pattern Formation . . . 204
11.9 More on Selected-Area Diffraction Patterns . . . 204
Chapter Summary . . . 208

12 Thinking in Reciprocal Space 211

Chapter Preview 211
12.1 Why Introduce Another Lattice? 211
12.2 Mathematical Definition of the Reciprocal Lattice 212
12.3 The Vector **g** 212
12.4 The Laue Equations and their Relation to Bragg's Law . 213
12.5 The Ewald Sphere of Reflection 214
12.6 The Excitation Error 216
12.7 Thin-Foil Effect and the Effect of Accelerating Voltage . 217
Chapter Summary 218

13 Diffracted Beams 221

Chapter Preview 221
13.1 Why Calculate Intensities? 221
13.2 The Approach 222
13.3 The Amplitude of a Diffracted Beam 223
13.4 The Characteristic Length ξ_g 223
13.5 The Howie-Whelan Equations 224
13.6 Reformulating the Howie-Whelan Equations 225
13.7 Solving the Howie-Whelan Equations 226
13.8 The Importance of $\gamma^{(1)}$ and $\gamma^{(2)}$ 226
13.9 The Total Wave Amplitude 227
13.10 The Effective Excitation Error 228
13.11 The Column Approximation 229
13.12 The Approximations and Simplifications 230
13.13 The Coupled Harmonic Oscillator Analog 231
Chapter Summary 231

14 Bloch Waves 235

Chapter Preview 235
14.1 Wave Equation in TEM 235
14.2 The Crystal 236
14.3 Bloch Functions 237
14.4 Schrödinger's Equation for Bloch Waves 238
14.5 The Plane-Wave Amplitudes 239
14.6 Absorption of Bloch Waves 241
Chapter Summary 242

15 Dispersion Surfaces 245

Chapter Preview 245
15.1 Introduction 245
15.2 The Dispersion Diagram When $U_g = 0$ 246
15.3 The Dispersion Diagram When $U_g \neq 0$ 247
15.4 Relating Dispersion Surfaces and Diffraction Patterns . . 247
15.5 The Relation Between U_g, $\boldsymbol{\xi}_g$, and s_g 250
15.6 The Amplitudes of Bloch Waves 252
15.7 Extending to More Beams 253
15.8 Dispersion Surfaces and Defects 254
Chapter Summary 254

16 Diffraction from Crystals 257

Chapter Preview 257
16.1 Review of Diffraction from a Primitive Lattice 257
16.2 Structure Factors: The Idea 258

16.3 Some Important Structures: BCC, FCC and HCP 259
16.4 Extending fcc and hcp to Include a Basis 261
16.5 Applying the bcc and fcc Analysis to Simple Cubic 262
16.6 Extending hcp to TiAl 262
16.7 Superlattice Reflections and Imaging 262
16.8 Diffraction from Long-Period Superlattices 264
16.9 Forbidden Reflections 265
16.10 Using the International Tables 265
Chapter Summary 267

17 Diffraction from Small Volumes 271

Chapter Preview 271
17.1 Introduction 271
17.1.A The Summation Approach 272
17.1.B The Integration Approach 273
17.2 The Thin-Foil Effect 273
17.3 Diffraction from Wedge-Shaped Specimens 274
17.4 Diffraction from Planar Defects 275
17.5 Diffraction from Particles 277
17.6 Diffraction from Dislocations, Individually and Collectively 278
17.7 Diffraction and the Dispersion Surface 279
Chapter Summary 281

18 Obtaining and Indexing Parallel-Beam Diffraction Patterns 283

Chapter Preview 283
18.1 Choosing Your Technique 284
18.2 Experimental SAD Techniques 284
18.3 The Stereographic Projection 286
18.4 Indexing Single-Crystal DPs 287
18.5 Ring Patterns from Polycrystalline Materials 290
18.6 Ring Patterns from Hollow-Cone Diffraction 291
18.7 Ring Patterns from Amorphous Materials 293
18.8 Precession Diffraction 295
18.9 Double Diffraction 296
18.10 Orientation of the Specimen 298
18.11 Orientation Relationships 302
18.12 Computer Analysis 303
18.13 Automated Orientation Determination and Orientation Mapping 305
Chapter Summary 305

19 Kikuchi Diffraction 311

Chapter Preview 311
19.1 The Origin of Kikuchi Lines 311
19.2 Kikuchi Lines and Bragg Scattering 312
19.3 Constructing Kikuchi Maps 313
19.4 Crystal Orientation and Kikuchi Maps 317
19.5 Setting the Value of S_g 318
19.6 Intensities 319
Chapter Summary 320

20 Obtaining CBED Patterns 323

Chapter Preview 323
20.1 Why Use a Convergent Beam? 323

20.2 Obtaining CBED Patterns 324
20.2.A Comparing SAD and CBED 325
20.2.B CBED in TEM Mode 326
20.2.C CBED in STEM Mode 326
20.3 Experimental Variables 327
20.3.A Choosing the C2 Aperture 327
20.3.B Selecting the Camera Length 328
20.3.C Choice of Beam Size 329
20.3.D Effect of Specimen Thickness 329
20.4 Focused and Defocused CBED Patterns 329
20.4.A Focusing a CBED Pattern 330
20.4.B Large-Angle (Defocused) CBED Patterns 330
20.4.C Final Adjustment 332
20.5 Energy Filtering 334
20.6 Zero-Order and High-Order Laue-Zone Diffraction 335
20.6.A ZOLZ Patterns 335
20.6.B HOLZ Patterns 336
20.7 Kikuchi and Bragg Lines in CBED Patterns 338
20.8 HOLZ Lines 339
20.8.A The Relationship Between HOLZ Lines and Kikuchi Lines 339
20.8.B Acquiring HOLZ Lines 341
20.9 Hollow-Cone/Precession CBED 342
Chapter Summary 343

21 Using Convergent-Beam Techniques 347

Chapter Preview 347
21.1 Indexing CBED Patterns 348
21.1.A Indexing ZOLZ and HOLZ Patterns 348
21.1.B Indexing HOLZ Lines 351
21.2 Thickness Determination 352
21.3 Unit-Cell Determination 354
21.3.A Experimental Considerations 354
21.3.B The Importance of the HOLZ-Ring Radius 355
21.3.C Determining the Lattice Centering 356
21.4 Basics of Symmetry Determination 357
21.4.A Reminder of Symmetry Concepts 357
21.4.B Friedel's Law 358
21.4.C Looking for Symmetry in Your Patterns 358
21.5 Lattice-Strain Measurement 361
21.6 Determination of Enantiomorphism 363
21.7 Structure Factor and Charge-Density Determination 364
21.8 Other Methods 365
21.8.A Scanning Methods 365
21.8.B Nanodiffraction 366
Chapter Summary 366

PART 3 IMAGING 369

22 Amplitude Contrast 371

Chapter Preview 371
22.1 What Is Contrast? 371
22.2 Amplitude contrast 372
22.2.A Images and Diffraction Patterns 372
22.2.B Use of the Objective Aperture or the STEM Detector: BF and DF Images 372

22.3 Mass-Thickness Contrast 373
22.3.A Mechanism of Mass-Thickness Contrast 373
22.3.B TEM Images 374
22.3.C STEM Images 376
22.3.D Specimens Showing Mass-Thickness Contrast . . 377
22.3.E Quantitative Mass-Thickness Contrast 378
22.4 *Z*-Contrast 379
22.5 TEM Diffraction Contrast 381
22.5.A Two-Beam Conditions 381
22.5.B Setting the Deviation Parameter, s 382
22.5.C Setting Up a Two-Beam CDF Image 382
22.5.D Relationship Between the Image and the Diffraction Pattern 384
22.6 STEM Diffraction Contrast 384
Chapter Summary 386

23 Phase-Contrast Images 389

Chapter Preview 389
23.1 Introduction 389
23.2 The Origin of Lattice Fringes 389
23.3 Some Practical Aspects of Lattice Fringes 390
23.3.A If $s = 0$ 390
23.3.B If $s \neq 0$ 390
23.4 On-Axis Lattice-Fringe Imaging 391
23.5 Moiré Patterns 392
23.5.A Translational Moiré Fringes 393
23.5.B Rotational Moiré Fringes 393
23.5.C General Moiré Fringes 393
23.6 Experimental Observations of Moiré Fringes 393
23.6.A Translational Moiré Patterns 394
23.6.B Rotational Moiré Patterns 394
23.6.C Dislocations and Moiré Fringes 394
23.6.D Complex Moiré Fringes 396
23.7 Fresnel Contrast 397
23.7.A The Fresnel Biprism 397
23.7.B Magnetic-Domain Walls 398
23.8 Fresnel Contrast from Voids or Gas Bubbles 399
23.9 Fresnel Contrast from Lattice Defects 400
23.9.A Grain Boundaries 402
23.9.B End-On Dislocations 402
Chapter Summary 402

24 Thickness and Bending Effects 407

Chapter Preview 407
24.1 The Fundamental Ideas 407
24.2 Thickness Fringes 408
24.3 Thickness Fringes and the DP 410
24.4 Bend Contours (Annoying Artifact, Useful Tool, Invaluable Insight) 411
24.5 ZAPs and Real-Space Crystallography 412
24.6 Hillocks, Dents, or Saddles 413
24.7 Absorption Effects 413
24.8 Computer Simulation of Thickness Fringes 414
24.9 Thickness-Fringe/Bend-Contour Interactions 414
24.10 Other Effects of Bending 415
Chapter Summary 416

25 Planar Defects **419**

Chapter Preview 419
25.1 Translations and Rotations 419
25.2 Why Do Translations Produce Contrast? 421
25.3 The Scattering Matrix 422
25.4 Using the Scattering Matrix 423
25.5 Stacking Faults in fcc Materials 424
25.5.A Why fcc Materials? 424
25.5.B Some Rules 425
25.5.C Intensity Calculations 426
25.5.D Overlapping Faults 426
25.6 Other Translations: π and δ Fringes 427
25.7 Phase Boundaries 429
25.8 Rotation Boundaries 430
25.9 Diffraction Patterns and Dispersion Surfaces 430
25.10 Bloch Waves and BF/DF Image Pairs 431
25.11 Computer Modeling 432
25.12 The Generalized Cross Section 433
25.13 Quantitative Imaging 434
25.13.A Theoretical Basis and Parameters 434
25.13.B Apparent Extinction Distance 435
25.13.C Avoiding the Column Approximation 435
25.13.D The User Interface 436
Chapter Summary 436

26 Imaging Strain Fields **441**

Chapter Preview 441
26.1 Why Image Strain Fields? 441
26.2 Howie-Whelan Equations 442
26.3 Contrast from a Single Dislocation 444
26.4 Displacement Fields and Ewald's Sphere 447
26.5 Dislocation Nodes and Networks 448
26.6 Dislocation Loops and Dipoles 448
26.7 Dislocation Pairs, Arrays, and Tangles 450
26.8 Surface Effects 451
26.9 Dislocations and Interfaces 452
26.10 Volume Defects and Particles 456
26.11 Simulating Images 457
26.11.A The Defect Geometry 457
26.11.B Crystal Defects and Calculating the Displacement Field 458
26.11.C The Parameters 458
Chapter Summary 459

27 Weak-Beam Dark-Field Microscopy **463**

Chapter Preview 463
27.1 Intensity in WBDF Images 463
27.2 Setting S_g Using the Kikuchi Pattern 464
27.3 How to Do WBDF 466
27.4 Thickness Fringes in Weak-Beam Images 467
27.5 Imaging Strain Fields 468
27.6 Predicting Dislocation Peak Positions 469
27.7 Phasor Diagrams 470
27.8 Weak-Beam Images of Dissociated Dislocations 473
27.9 Other Thoughts 477

27.9.A Thinking of Weak-Beam Diffraction as a Coupled Pendulum . . . 477
27.9.B Bloch Waves . . . 478
27.9.C If Other Reflections are Present . . . 478
27.9.D The Future Is Now . . . 478
Chapter Summary . . . 479

28 High-Resolution TEM . . . 483

Chapter Preview . . . 483
28.1 The Role of an Optical System . . . 483
28.2 The Radio Analogy . . . 484
28.3 The Specimen . . . 485
28.4 Applying the WPOA to the TEM . . . 487
28.5 The Transfer Function . . . 487
28.6 More on $\chi(u)$, $\sin \chi(u)$, and $\cos \chi(u)$ . . . 488
28.7 Scherzer Defocus . . . 490
28.8 Envelope Damping Functions . . . 491
28.9 Imaging Using Passbands . . . 492
28.10 Experimental Considerations . . . 493
28.11 The Future for HRTEM . . . 494
28.12 The TEM as a Linear System . . . 494
28.13 FEG TEMs and the Information Limit . . . 495
28.14 Some Difficulties in Using an FEG . . . 498
28.15 Selectively Imaging Sublattices . . . 500
28.16 Interfaces and Surfaces . . . 502
28.17 Incommensurate Structures . . . 503
28.18 Quasicrystals . . . 504
28.19 Single Atoms . . . 505
Chapter Summary . . . 506

29 Other Imaging Techniques . . . 511

Chapter Preview . . . 511
29.1 Stereo Microscopy and Tomography . . . 511
29.2 $2\frac{1}{2}$D Microscopy . . . 512
29.3 Magnetic Specimens . . . 514
29.3.A The Magnetic Correction . . . 514
29.3.B Lorentz Microscopy . . . 515
29.4 Chemically Sensitive Images . . . 517
29.5 Imaging with Diffusely Scattered Electrons . . . 517
29.6 Surface Imaging . . . 519
29.6.A Reflection Electron Microscopy . . . 519
29.6.B Topographic Contrast . . . 521
29.7 High-Order BF Imaging . . . 521
29.8 Secondary-Electron Imaging . . . 522
29.9 Backscattered-Electron Imaging . . . 523
29.10 Charge-Collection Microscopy and Cathodoluminescence 523
29.11 Electron Holography . . . 524
29.12 In Situ TEM: Dynamic Experiments . . . 526
29.13 Fluctuation Microscopy . . . 528
29.14 Other Variations Possible in a STEM . . . 528
Chapter Summary . . . 529

30 Image Simulation . . . 533

Chapter Preview . . . 533
30.1 Simulating images . . . 533
30.2 The Multislice Method . . . 533

30.3 The Reciprocal-Space Approach 534
30.4 The FFT Approach 536
30.5 The Real-Space approach 536
30.6 Bloch Waves and HRTEM Simulation 536
30.7 The Ewald Sphere Is Curved 537
30.8 Choosing the Thickness of the Slice 537
30.9 Beam Convergence 538
30.10 Modeling the Structure 540
30.11 Surface Grooves and Simulating Fresnel Contrast 540
30.12 Calculating Images of Defects 542
30.13 Simulating Quasicrystals 543
30.14 Bonding in Crystals 544
30.15 Simulating *Z*-Contrast 545
30.16 Software for Phase-Contrast HRTEM 545
Chapter Summary 545

31 Processing and Quantifying Images 549

Chapter Preview 549
31.1 What Is Image Processing? 549
31.2 Processing and Quantifying Images 550
31.3 A Cautionary Note 550
31.4 Image Input 550
31.5 Processing Techniques 551
31.5.A Fourier Filtering and Reconstruction 551
31.5.B Analyzing Diffractograms 552
31.5.C Averaging Images and Other Techniques 554
31.5.D Kernels 556
31.6 Applications 556
31.6.A Beam-Sensitive Materials 556
31.6.B Periodic Images 557
31.6.C Correcting Drift 557
31.6.D Reconstructing the Phase 557
31.6.E Diffraction Patterns 558
31.6.F Tilted-Beam Series 559
31.7 Automated Alignment 560
31.8 Quantitative Methods of Image Analysis 561
31.9 Pattern Recognition in HRTEM 562
31.10 Parameterizing the Image Using QUANTITEM 563
31.10.A The Example of a Specimen with Uniform Composition 563
31.10.B Calibrating the Path of R 565
31.10.C Noise Analysis 565
31.11 Quantitative Chemical Lattice Imaging 567
31.12 Methods of Measuring Fit 568
31.13 Quantitative Comparison of Simulated and Experimental HRTEM Images 570
31.14 A Fourier Technique for Quantitative Analysis 571
31.15 Real or Reciprocal Space? 572
31.16 Software 573
31.17 The Optical Bench—A Little History 573
Chapter Summary 575

PART 4 SPECTROMETRY 579

32 X-ray Spectrometry 581

Chapter Preview 581

32.1 X-ray Analysis: Why Bother? . . . 581
32.2 Basic Operational Mode . . . 584
32.3 The Energy-Dispersive Spectrometer . . . 584
32.4 Semiconductor Detectors . . . 585
32.4.A How Does an XEDS Work? . . . 585
32.4.B Cool Detectors . . . 586
32.4.C Different Kinds of Windows . . . 586
32.4.D Intrinsic-Germanium Detectors . . . 587
32.4.E Silicon-Drift Detectors . . . 588
32.5 Detectors with High-Energy Resolution . . . 589
32.6 Wavelength-Dispersive Spectrometers . . . 589
32.6.A Crystal WDS . . . 589
32.6.B CCD-Based WDS . . . 590
32.6.C Bolometers/Microcalorimeters . . . 590
32.7 Turning X-rays into Spectra . . . 591
32.8 Energy Resolution . . . 593
32.9 What You Should Know about Your XEDS . . . 594
32.9.A Detector Characteristics . . . 594
32.9.B Processing Variables . . . 596
32.10 The XEDS-AEM Interface . . . 598
32.10.A Collection Angle . . . 598
32.10.B Take-Off Angle . . . 599
32.10.C Orientation of the Detector to the Specimen . . . 599
32.11 Protecting the Detector from Intense Radiation . . . 600
Chapter Summary . . . 601

33 X-ray Spectra and Images . . . 605

Chapter Preview . . . 605
33.1 The Ideal Spectrum . . . 605
33.1.A The Characteristic Peaks . . . 605
33.1.B The Continuum Bremsstrahlung Background . . . 606
33.2 Artifacts Common to Si(Li) XEDS Systems . . . 606
33.3 The Real Spectrum . . . 608
33.3.A Pre-Specimen Effects . . . 608
33.3.B Post-Specimen Scatter . . . 611
33.3.C Coherent Bremsstrahlung . . . 613
33.4 Measuring the Quality of the XEDS-AEM Interface . . . 614
33.4.A Peak-to-Background Ratio . . . 614
33.4.B Efficiency of the XEDS System . . . 614
33.5 Acquiring X-ray Spectra . . . 615
33.5.A Spot Mode . . . 615
33.5.B Spectrum-Line Profiles . . . 616
33.6 Acquiring X-ray Images . . . 616
33.6.A Analog Dot Mapping . . . 617
33.6.B Digital Mapping . . . 618
33.6.C Spectrum Imaging (SI) . . . 619
33.6.D Position-Tagged Spectrometry (PTS) . . . 620
Chapter Summary . . . 620

34 Qualitative X-ray Analysis and Imaging . . . 625

Chapter Preview . . . 625
34.1 Microscope and Specimen Variables . . . 625
34.2 Basic Acquisition Requirements: Counts, Counts, and More Caffeine . . . 626

34.3 Peak Identification 627
34.4 Peak Deconvolution 630
34.5 Peak Visibility 632
34.6 Common Errors 634
34.7 Qualitative X-ray Imaging: Principles and Practice 634
Chapter Summary 636

35 Quantitative X-ray Analysis 639

Chapter Preview 639
35.1 Historical Perspective 639
35.2 The Cliff-Lorimer Ratio Technique 640
35.3 Practical Steps for Quantification 641
35.3.A Background Subtraction 641
35.3.B Peak Integration 644
35.4 Determining k-Factors 646
35.4.A Experimental Determination of k_{AB} 646
35.4.B Errors in Quantification: The Statistics 647
35.4.C Calculating k_{AB} 648
35.5 The Zeta-Factor Method 652
35.6 Absorption Correction 654
35.7 The Zeta-Factor Absorption Correction 656
35.8 The Fluorescence Correction 656
35.9 ALCHEMI 657
35.10 Quantitative X-ray Mapping 658
Chapter Summary 660

36 Spatial Resolution and Minimum Detection 663

Chapter Preview 663
36.1 Why Is Spatial Resolution Important? 663
36.2 Definition and Measurement of Spatial Resolution 664
36.2.A Beam Spreading 665
36.2.B The Spatial-Resolution Equation 666
36.2.C Measurement of Spatial Resolution 667
36.3 Thickness Measurement 668
36.3.A TEM Methods 669
36.3.B Contamination-Spot Separation Method 670
36.3.C Convergent-Beam Diffraction Method 671
36.3.D Electron Energy-Loss Spectrometry Methods . . 671
36.3.E X-ray Spectrometry Method 671
36.4 Minimum Detection 672
36.4.A Experimental Factors Affecting the MMF 673
36.4.B Statistical Criterion for the MMF 673
36.4.C Comparison with Other Definitions 674
36.4.D Minimum-Detectable Mass 674
Chapter Summary 675

37 Electron Energy-Loss Spectrometers and Filters 679

Chapter Preview 679
37.1 Why Do EELS? 679
37.1.A Pros and Cons of Inelastic Scattering 679
37.1.B The Energy-Loss Spectrum 680
37.2 EELS Instrumentation 681
37.3 The Magnetic Prism: A Spectrometer and a Lens 681
37.3.A Focusing the Spectrometer 682
37.3.B Spectrometer Dispersion 683

37.3.C Spectrometer Resolution 683
37.3.D Calibrating the Spectrometer 684
37.4 Acquiring a Spectrum 684
37.4.A Image and Diffraction Modes 685
37.4.B Spectrometer-Collection Angle 685
37.4.C Spatial Selection 688
37.5 Problems with PEELS 688
37.5.A Point-Spread Function 688
37.5.B PEELS Artifacts 689
37.6 Imaging Filters 690
37.6.A The Omega Filter 691
37.6.B The GIF 692
37.7 Monochromators 693
37.8 Using Your Spectrometer and Filter 694
Chapter Summary 696

38 Low-Loss and No-Loss Spectra and Images 699

Chapter Preview 699
38.1 A Few Basic Concepts 699
38.2 The Zero-Loss Peak (ZLP) 701
38.2.A Why the ZLP Really Isn't 701
38.2.B Removing the Tail of the ZLP 701
38.2.C Zero-Loss Images and Diffraction Patterns 702
38.3 The Low-Loss Spectrum 703
38.3.A Chemical Fingerprinting 704
38.3.B Dielectric-Constant Determination 705
38.3.C Plasmons 705
38.3.D Plasmon-Loss Analysis 707
38.3.E Single-Electron Excitations 709
38.3.F The Band Gap 709
38.4 Modeling The Low-Loss Spectrum 710
Chapter Summary 711

39 High Energy-Loss Spectra and Images 715

Chapter Preview 715
39.1 The High-Loss Spectrum 715
39.1.A Inner-Shell Ionization 715
39.1.B Ionization-Edge Characteristics 717
39.2 Acquiring a High-Loss Spectrum 721
39.3 Qualitative Analysis 723
39.4 Quantitative Analysis 723
39.4.A Derivation of the Equations for Quantification 724
39.4.B Background Subtraction 726
39.4.C Edge Integration 728
39.4.D The Partial Ionization Cross Section 728
39.5 Measuring Thickness from the Core-Loss Spectrum 730
39.6 Deconvolution 731
39.7 Correction for Convergence of the Incident Beam 733
39.8 The Effect of the Specimen Orientation 733
39.9 EFTEM Imaging with Ionization Edges 733
39.9.A Qualitative Imaging 734
39.9.B Quantitative Imaging 734
39.10 Spatial Resolution: Atomic-Column EELS 735

39.11 Detection Limits 736
Chapter Summary 737

40 Fine Structure and Finer Details 741

Chapter Preview 741
40.1 Why Does Fine Structure Occur? 741
40.2 ELNES Physics 742
40.2.A Principles 742
40.2.B White Lines 744
40.2.C Quantum Aspects 744
40.3 Applications of ELNES 745
40.4 ELNES Fingerprinting 746
40.5 ELNES Calculations 747
40.5.A The Potential Choice 748
40.5.B Core-Holes and Excitons 749
40.5.C Comparison of ELNES Calculations and Experiments 750
40.6 Chemical Shifts in the Edge Onset 750
40.7 EXELFS 751
40.7.A RDF via EXELFS 752
40.7.B RDF via Energy-Filtered Diffraction 753
40.7.C A Final Thought Experiment 753
40.8 Angle-Resolved EELS 755
40.9 EELS Tomography 755
Chapter Summary 757

Index I-1

Figure Credits

TEM is a visual science, and any TEM text is heavily dependent on figures, halftones, and (more recently) full color images to transmit its message. We have been fortunate to work with many colleagues over the years who have generously given us fine examples of the art and science of TEM; we would like to acknowledge them here. We have also used our own work, and the work of others, whose permission has been sought as listed below.

Chapter 1

Figure 1.1: From Ruska, E (1980) *The Early History of the Electron Microscope*, Fig. 6 reproduced by permission of S. Herzel Verlag GmbH & Co.
Figure 1.2B,C: Specimen courtesy of Y Ikuhara and T Yamamoto, University of Tokyo, reproduced by permission of JEOL Ltd.
Figure 1.4: Courtesy of M Watanabe.
Figure 1.6: Courtesy of KS Vecchio.
Figure 1.7: Courtesy of T Hayes, from Hayes, T (1980) in O Johari Ed. SEM-1980 1 1, Fig. 8 reproduced by permission of Scanning Microscopy International.
Figure 1.9A: Courtesy of M Kersker, reproduced by permission of JEOL USA Inc.
Figure 1.9B: Courtesy of E Essers, reproduced by permission of Carl Zeiss SMT.
Figure 1.9C: Courtesy of K Jarausch, reproduced by permission of Hitachi High Technologies.
Figure 1.9D: Courtesy of M Kersker, reproduced by permission of JEOL USA Inc.
Figure 1.9E: Courtesy of OL Krivanek, reproduced by permission of NION Inc.
Figure 1.9F: Courtesy of JS Fahy, reproduced by permission of FEI Co.

Chapter 2

Figure 2.4: Courtesy of J Bruley and VJ Keast.
Figure 2.11: Modified from Hecht, E (1988) Optics, Fig. 10.21 Addison-Wesley.
Figure 2.13A,D: Courtesy of KS Vecchio.
Figure 2.13C: Courtesy of DW Ackland.

Chapter 3

Figure 3.3: Courtesy of DE Newbury, modified from Newbury, DE (1986) in DC Joy *et al.* Eds. *Principles of Analytical Electron Microscopy* p 6, Fig. 2 original reproduced by permission of Plenum Press.
Figure 3.4: Courtesy of DE Newbury, modified from data in Newbury, DE (1986) in DC Joy *et al.* Eds. *Principles of Analytical Electron Microscopy* p 8, Table II reproduced by permission of Plenum Press.

Chapter 4

Figure 4.1: Courtesy of DE Newbury, modified from Newbury, DE (1986) in DC Joy *et al.* Eds. *Principles of Analytical Electron Microscopy* p 20, Fig. 4 original reproduced by permission of Plenum Press.
Figure 4.3: Modified from Woldseth, R (1973) *X-ray Energy Spectrometry*, Fig. 3 original reproduced by permission of Kevex Instruments.
Figure 4.4: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 4.3 reproduced by permission of Philips Electron Optics.
Figure 4.11: Courtesy of LW Hobbs, modified from Hobbs, LW (1979) in JJ Hren *et al.* Eds. *Introduction to Analytical Electron Microscopy*, Fig. 17.2 original reproduced by permission of Plenum Press.
Figure 4.12: Courtesy of LW Hobbs, modified from Hobbs, LW (1979) in JJ Hren *et al.* Eds. *Introduction to Analytical Electron Microscopy*, Fig. 17.4 original reproduced by permission of Plenum Press.
Table 4.1: Courtesy of JI Goldstein, from Goldstein, JI *et al.* (1992) *Scanning Electron Microscopy and X-ray Microanalysis,* 2nd Edition, Table 3.11 reproduced by permission of Plenum Press.
Table 4.2: Data obtained from National Physical Laboratory, Teddington, UK, web site. http://www.kayelaby.npl.co.uk/atomic_and_nuclear_physics/4_2/4_2_1.html
Table 4.3: Courtesy of NJ Zaluzec and JF Mansfield, from Zaluzec, NJ and Mansfield, JF (1987) in K Rajan Ed. *Intermediate Voltage Electron Microscopy and Its Application to Materials Science* p 29, Table 1 reproduced by permission of Philips Electron Optics.

Chapter 5

Figure 5.1: Modified from Hall, CE (1966) *Introduction to Electron Microscopy*, Fig. 7.8 McGraw-Hill.
Figure 5.4B: Courtesy of JI Goldstein, modified from Goldstein, JI *et al*. (1992) *Scanning Electron Microscopy and X-ray Microanalysis,* 2nd Edition, Fig. 2.7 original reproduced by permission of Plenum Press.
Figure 5.5: Courtesy of DW Ackland.
Figure 5.6A: Modified from Crewe, AV *et al*. (1969) Rev. Sci. Instrum. **40** 241, Fig. 2.
Figure 5.6B: Courtesy of DW Ackland.
Figure 5.7: Courtesy of M Watanabe, modified from Watanabe, M and Williams, DB (2006) J. Microsc. **221** 89, Fig. 14.
Figure 5.10: Courtesy of JR Michael, modified from Michael, JR and Williams, DB (1987) J. Microsc. **147** 289, Fig. 3 original reproduced by permission of the Royal Microscopical Society.
Figure 5.11: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 2.12B Philips Electron Optics.
Figure 5.12: Courtesy of JR Michael, from Michael, JR and Williams, DB (1987) J. Microsc. **147** 289, Fig. 2 original reproduced by permission of the Royal Microscopical Society.
Figure 5.13A: Courtesy of DW Ackland.
Figure 5.13B: Reproduced by permission of NSA Hitachi Scientific Instruments Ltd.

Chapter 6

Figure 6.7: Courtesy of DW Ackland.
Figure 6.8A: Reproduced by permission of Philips Electronic Instruments Inc.
Figure 6.8B: Reproduced by permission of Kratos Ltd.
Figure 6.8C: From Mulvey, T (1974) Electron Microscopy-1974 17, Fig. 1 reproduced by permission of the Australian Academy of Science.
Figure 6.8D: From Reimer, L (1993) *Transmission Electron Microscopy,* 3rd Edition, Fig. 2.12 reproduced by permission of Springer Verlag.
Figure 6.9: Modified from Reimer, L (1993) Transmission Electron Microscopy, 3rd Edition, Fig. 2.3 Springer Verlag.
Figure 6.10B: Courtesy of AO Benscoter.
Figure 6.11: Modified from Reimer, L (1993) *Transmission Electron Microscopy,* 3rd Edition, Fig. 2.13 Springer Verlag.
Figure 6.12A: Courtesy of OL Krivanek, reproduced by permission of NION Inc.
Figure 6.12A: Courtesy of M Haider, reproduced by permission of CEOS GmbH.
Figure 6.15: Modified from Reimer, L (1993) *Transmission Electron Microscopy,* 3rd Edition, Fig. 4.23 Springer Verlag.

Chapter 7

Figure 7.1: Modified from Stephen, J et al. (1975) J. Phys. E 8 607, Fig. 2.
Figure 7.5: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 1.2 Philips Electron Optics.
Figure 7.6: Modified from Berger, SD et al. (1985) *Electron Microscopy and Analysis* p 137, Fig. 1 original by permission of The Institute of Physics Publishing.

Chapter 8

Figure 8.1: Courtesy of WC Bigelow, modified from Bigelow, WC (1994) *Vacuum Methods in Electron Microscopy*, Fig. 4.1 original by permission of Portland Press Ltd.
Figure 8.2: Courtesy of WC Bigelow, modified from Bigelow, WC (1994) *Vacuum Methods in Electron Microscopy*, Fig. 5.1 original by permission of Portland Press Ltd.
Figure 8.3: Reproduced by permission of Leybold Vacuum Products Inc.
Figure 8.4: Courtesy of WC Bigelow, modified from Bigelow, WC (1994) *Vacuum Methods in Electron Microscopy*, Fig. 7.1 original by permission of Portland Press Ltd.
Figure 8.6: Reproduced by permission of Gatan Inc.
Figure 8.7: Modified from Valdrè, U and Goringe, MJ (1971) in U Valdrè Ed. *Electron Microscopy in Materials Science* p 217, Fig. 6 original by permission of Academic Press Inc.
Figure 8.8: Courtesy of NSA Hitachi Scientific Instruments Ltd.
Figure 8.9A,B: Reproduced by permission of Gatan Inc.
Figure 8.10A: Reproduced by permission of Gatan Inc.
Figure 8.11: Reproduced by permission of Gatan Inc.
Figure 8.12: Modified from Komatsu, M *et al*. (1994) J. Amer. Ceram. Soc. **77** 839, Fig. 1 original by permission of The American Ceramic Society.
Figure 8.13: Original by permission of NSA Hitachi Scientific Instruments Ltd.
Figure 8.14A: Courtesy PE Fischione, reproduced by permission of EA Fichione Instruments Inc.
Figure 8.14B: Courtesy PE Fischione, reproduced by permission of EA Fichione Instruments Inc.
Figure 8.15A,B: Courtesy NJ Zaluzec.

Chapter 9

Figure 9.5: Courtesy of J Rodenburg, modified from original diagram on web site.
Figure 9.6: Modified from Reimer, L (1993) *Transmission Electron Microscopy,* 3rd Edition, Fig. 4.14 Springer Verlag.
Figure 9.10B: Courtesy of M Watanabe, modified from Watanabe, M *et al*. (2006) Microsc. Microanal. **12** 515, Fig. 6
Figure 9.15: Courtesy of R Ristau.
Figure 9.17: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 1.7 original reproduced by permission of Philips Electron Optics.
Figure 9.19B-D: Courtesy of DW Ackland.
Figure 9.20: Modified from Edington, JW (1976) Practical Electron Microscopy in Materials Science, Fig. 1.5 original reproduced by permission of Philips Electron Optics.
Figure 9.21: Courtesy of S Ramamurthy.
Figure 9.22: Courtesy of DW Ackland.
Figure 9.24: Courtesy of DW Ackland.
Figure 9.25: Courtesy of DW Ackland.
Figure 9.26: Courtesy of S Ramamurthy.
Table 9.1: From Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Table 2.4 reproduced by permission of Philips Electron Optics.
Table 9.2: From Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Table 2.2 reproduced by permission of Philips Electron Optics.

Chapter 10

Figure 10.1: Modified from Médard, L *et al*. (1949) Rev. Met. **46** 549, Fig. 5.
Figure 10.3: Reproduced by permission of SPI Inc.
Figure 10.4: Reproduced by permission of South Bay Technology.
Figure 10.5A: Reproduced by permission of Electron Microscopy Sciences.
Figure 10.5B: Reproduced by permission of VCR Inc.
Figure 10.8A: Modified from Thompson-Russell, KC and Edington, JW (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 9 original reproduced by permission of Philips Electron Optics.
Fig. 10.8B: Modified from Thompson-Russell, KC and Edington, JW (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 7 original reproduced by permission of Philips Electron Optics.
Figure 10.9A: Modified from Thompson-Russell, KC and Edington, JW (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 12 original reproduced by permission of Philips Electron Optics.
Figure 10.9B: Courtesy PE Fischione, reproduced by permission of EA Fichione Instruments Inc.
Figure 10.10: Modified from Thompson-Russell, KC and Edington, JW (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 11 Philips Electron Optics.
Figure 10.11: Courtesy of R Alani, reproduced by permission of Gatan Inc.
Figure 10.12: Courtesy of AG Cullis, from Cullis, AG *et al*. (1985) Ultramicrosc. **17** 203, Figs. lA, 3 reproduced by permission of Elsevier Science BV.
Figure 10.13: Modified from van Hellemont, J *et al*. (1988) in J Bravman *et al*. Eds. *Specimen Preparation for Transmission Electron Microscopy of Materials* Mater. Res. Soc. Symp. Proc. **115** 247, Fig. 1 original by permission of MRS.
Figure 10.16A,B: Modified from Thompson-Russell, KC and Edington, JW (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Figs. 20, 21 original reproduced by permission of Philips Electron Optics.
Figure 10.17A: Modified from Thompson-Russell, KC and Edington, JW (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 25 original reproduced by permission of Philips Electron Optics.
Figure 10.17B: Courtesy of M Aindow.
Figure 10.19A-F: Courtesy of SD Walck.
Figure 10.20: Modified from Hetherington, CJD (1988) in J Bravman *et al*. Eds. *Specimen Preparation for Transmission Electron Microscopy of Materials* Mater. Res. Soc. Symp. Proc. **115** 143, Fig. 1 original reproduced by permission of MRS.
Figure 10.21: Modified from Dobisz, EA et al. (1986) J. Vac. Sci. Technol. B 4 850, Fig. 1 original reproduced by permission of MRS.
Figure 10.22A,B: After Fernandez, A (1988) in J Bravman *et al*. Eds. *Specimen Preparation for Transmission Electron Microscopy of Materials*. Mater. Res. Soc. Symp. Proc. **115** 119, Fig. 1.
Figure 10.22C,D: Courtesy of J Basu.
Figure 10.23: Reproduced by permission of FEI Inc.
Figure 10.24A-F: Courtesy of L Giannuzzi.
Figure 10.25A,B: Thanks to JR Michael.
Figure 10.26: Modified from Goodhew, PJ (1988) in J Bravman *et al*. Eds. *Specimen Preparation for Transmission Electron Microscopy of Materials,* Mater. Res. Soc. Symp. Proc. **115** 52.
Table 10.1: Courtesy of T Malis.

Chapter 11

Table 11.1: Modified from Hirsch, PB *et al.* (1977) *Electron Microscopy of Thin Crystals,* 2nd Edition p 19, Krieger, NY.

Chapter 13

Table 13.2: Modified from Reimer, L (1993) *Transmission Electron Microscopy,* 3rd Edition Table 7.2 p 296 Springer Verlag.

Chapter 14

Figure 14.2: Modified from Hashimoto, H *et al.* (1962) Proc. Roy. Soc. (London) **A269** 80, Fig. 2.
Table 14.2: Modified from Reimer, L (1993) *Transmission Electron Microscopy,* 3rd Edition Table 3.2 p 58, Springer Verlag.

Chapter 16

Figure 16.5: Courtesy of ML Jenkins, from Jenkins, ML *et al.* (1976) Philos. Mag. **34** 1141, Fig. 2 reproduced by permission of Taylor and Francis.
Figure 16.6: Courtesy of BC De Cooman.
Figure 16.7: From Dodsworth, J *et al.* (1983) Adv. Ceram. **6** 102, Fig. 3 reproduced by permission of the American Ceramic Society.
Figure 16.8: Courtesy of BC De Cooman.
Figure 16.9: Courtesy of M Gajdardziska-Josifovska, from Gajdardziska-Josifovska M *et al.* (1995) Ultramicrosc. **58** 65, Fig. 1 reproduced by permission of Elsevier Science BV.
Figure 16.10: Courtesy of S McKernan.
Figure 16.11: Modified from Hahn, T (Ed.) *International Tables for Crystallography A* pp 538–539, No. 164 original by permission of The International Union of Crystallography.
Table 16.1: Modified from Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Appendix 8 Van Nostrand Reinhold.

Chapter 17

Figure 17.2: Modified from Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Fig. 2.16 original reproduced by permission of Philips Electron Optics.
Figure 17.7: From Carter, CB *et al.* (1981) Philos. Mag. **A43** 441, Fig. 5C reproduced by permission of Taylor and Francis.
Figure 17.9: Modified from Hirsch, PB *et al.* (1977) *Electron Microscopy of Thin Crystals,* 2nd Edition, Fig. 4.11, Krieger.
Figure 17.10: From Driver, JH *et al.* (1972) Phil Mag. ***26*** 1227, Fig. 3 reproduced by permission of Taylor and Francis.
Figure 17.11A-C: From Lewis, MH and Billingham, J (1972) JEOL News 10e(l) 8, Fig. 3 reproduced by permission of JEOL USA Inc.
Figure 17.11D: Modified from Sauvage, M and Parthè, E (1972) Acta Cryst. **A28** 607, Fig. 2.
Figure 17.12: Modified from Carter, CB *et al.* (1981) Philos. Mag. **A43** 441, Fig. 5A,B.
Figure 17.13: Modified from Carter, CB *et al.* (1980) J. Electron Microsc. **63** 623, Fig. 8.
Figure 17.14: Modified from Carter, CB (1984) Philos. Mag. **A50** 133, Figs. 1–3.

Chapter 18

Figure 18.2: Modified from Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Fig. A1.7 original reproduced by permission of Philips Electron Optics.
Figure 18.7: Courtesy of S Ramamurthy.
Figure 18.9: Courtesy of S McKernan.
Figure 18.10A,C: Courtesy of S McKernan.
Figure 18.10B,D: Modified from Vainshtein, BK *et al.* (1992) in JM Cowley Ed. *Electron Diffraction Techniques* **1**, Fig. 6.13 original reproduced by permission of Oxford University Press.
Figure 18.10E: From Vainshtein, BK *et al.* (1992) in JM Cowley Ed. *Electron Diffraction Techniques* **1**, Fig. 6.13 reproduced by permission of Oxford University Press.
Figure 18.11: Modified from James, RW (1965) in L Bragg Ed. *The Optical Principles of the Diffraction of X-rays* The Crystalline State **II**, Figs. 170, 184 Cornell University Press.
Figure 18.12: Courtesy of DJH Cockayne, modified from Sproul, A *et al.* (1986) Philos. Mag. **B54** 113, Fig. 1 original by permission of Taylor and Francis.
Figure 18.13: From Graczyk, JF and Chaudhari, P (1973) Phys. stat. sol. b **58** 163, Fig. l0A reproduced by permission of Akademie Verlag GmbH.
Figure 18.14: Courtesy of A Howie, from Howie, A (1988) in PR Buseck *et al.* Eds. *High-Resolution Transmission Microscopy and Associated Techniques* p 60, Fig. 14.12 reproduced by permission of Oxford University Press.
Figure 18.15: Courtesy of LD Marks and CS Own, modified from Own, CS and Marks, LD (2005) Rev. Sci. Instrum. **76** 033703, Fig. 1.
Figure 18.16: Courtesy of J-P Morniroli.

Figure 18.17: From Tietz, LA *et al*. (1995) Ultramicrosc. **60** 241, Figs. 2–4 reproduced by permission of Elsevier Science BV.
Figure 18.18: Modified from Tietz, LA *et al*. (1995) Ultramicrosc. **60** 241, Fig. 5 original by permission of Elsevier Science BV.
Figure 18.19: Modified from Andrews, KW *et al*. (1971) *Interpretation of Electron Diffraction Patterns,* 2nd Edition, Fig. 41 original reproduced by permission of Plenum Press.
Figure 18.20: Modified from Andrews, KW *et al*. (1971) *Interpretation of Electron Diffraction Patterns,* 2nd Edition, Fig. 41 original reproduced by permission of Plenum Press.
Figure 18.21: Modified from Andrews, KW *et al*. (1971) *Interpretation of Electron Diffraction Patterns,* 2nd Edition, Fig. 41 original reproduced by permission of Plenum Press.
Figure 18.22: Modified from Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Fig. 2.20 original reproduced by permission of Philips Electron Optics.
Figure 18.24: Modified from Li, C and Williams, DB (2003) Interface Science **11** 461–472, Figs. 2A, 4. Courtesy of C Li.

Chapter 19

Figure 19.6A: Courtesy of G. Thomas, modified from Levine, E *et al*. (1966) J. Appl. Phys. **37** 2141, Fig. lA original reproduced by permission of the American Institute of Physics.
Figure 19.7: Modified from Okamoto, PR *et al*. (1967) J. Appl. Phys. **38** 289, Fig. 5.
Figure 19.8: Courtesy of S Ramamurthy.
Figure 19.9A: Courtesy of G. Thomas, modified from Thomas, G and Goringe, MJ (1979) *Transmission Electron Microscopy of Metals*, Fig. 2.30 John Wiley & Sons Inc.
Figure 19.9B: Modified from Edington. JW (1976) *Practical Electron Microscopy in Materials Science*, Fig. 2.27 Van Nostrand Reinhold.
Figure 19.11: Modified from Thomas, G and Goringe, MJ (1979) *Transmission Electron Microscopy of Metals*, Fig. 2.29 John Wiley & Sons Inc. Thanks to G Thomas.

Chapter 20

Figure 20.2A: Courtesy of KS Vecchio, from Williams, DB *et al*. Eds. (1992) *Images of Materials*, Fig. 6.5 reproduced by permission of Oxford University Press.
Figure 20.2B: Courtesy of KS Vecchio, from Williams, DB *et al*. Eds. (1992) *Images of Materials*, Fig. 6.17 reproduced by permission of Oxford University Press,
Figure 20.3: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 6.6 Philips Electron Optics.
Figure 20.5: Courtesy of JF Mansfield, from Mansfield, JF (1984) *Convergent Beam Diffraction of Alloy Phases*, Fig. 5.3 reproduced by permission of Institute of Physics Publishing.
Figure 20.6: From Lyman, CE *et al*. Eds. (1990) *Scanning Electron Microscopy, X-ray Microanalysis and Analytical Electron Microscopy—A Laboratory Workbook*, Fig. A 27.2 reproduced by permission of Plenum Press.
Figure 20.7A,B: Courtesy of J-P Morniroli, from Morniroli, J-P (2002) *Large-Angle Convergent-Beam Electron Diffraction*, Figs. V.8, V.12B (Thanks to SF Paris).
Figure 20.8A: Courtesy of J-P Morniroli, modified from Morniroli, J-P (2002) *Large-Angle Convergent-Beam Electron Diffraction*, Fig. VI.I.
Figure 20.8B-D: Courtesy of J-P Morniroli, from Morniroli, J-P (2002) *Large-Angle Convergent-Beam Electron Diffraction*, Fig. VI.2A-C (Thanks to SF Paris).
Figure 20.9: Courtesy of KS Vecchio, from Williams, DB *et al*. Eds. (1992) *Images of Materials*, Fig. 6.21 reproduced by permission of Oxford University Press.
Figure 20.10: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.
Figure 20.11D: Courtesy of R Ayer.
Figure 20.12: Modified from Ayer, R (1989) J. Electron Microscopy Tech. **13** 3, Fig. 3.
Figure 20.13: Courtesy of WAT Clark from Heilman, P *et al*. (1983) Acta Metall. **31** 1293, Fig. 4 reproduced by permission of Elsevier Science BV.
Figure 20.14: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 6.9 original by permission of Philips Electron Optics.
Figure 20.15: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 6.16 original by permission of Philips Electron Optics.
Figure 20.16: Courtesy of CM Sung.
Figure 20.17: Courtesy of M Terauchi.
Figure 20.18: Courtesy of CS Own and LD Marks.

Chapter 21

Figure 21.1: Courtesy of ZL Wang, after Wang, ZL *et al*. (2003) Phys. Rev. Lett. **91** 185502, Fig. 2.
Figure 21.2: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials*

Science, 2nd Edition, Fig. 6.13 original by permission of Philips Electron Optics.

Figure 21.3: Modified from Williams, DB 1987 *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 6.14 original by permission of Philips Electron Optics.

Figure 21.4: Courtesy of JM Zuo, simulation from WebEMAPS.

Figure 21.5: Courtesy of B Ralph, modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 6.18 original by permission of Philips Electron Optics.

Figure 21.6: From Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 4.29A reproduced by permission of Philips Electron Optics.

Figure 21.8: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 4.29B,C original by permission of Philips Electron Optics.

Figure 21.9: Courtesy of R Ayer, from Raghavan, M *et al.* (1984) Metall. Trans. **I5A** 783, Fig. 6 reproduced by permission of ASM International.

Figure 21.10A: Courtesy of KS Vecchio, modified from Williams, DB *et al.* Eds. (1992) *Images of Materials*, Fig. 6.23 original by permission of Oxford University Press.

Figure 21.10B: Courtesy of R Ayer, modified from Ayer, R (1989) J. Electron Microsc. Tech. **13** 3, Fig. 7 original by permission of John Wiley & Sons Inc.

Figure 21.12: Courtesy of KS Vecchio, modified from Williams. DB *et al.* Eds. (1992) *Images of Materials*, Fig. 6.19 original by permission of Oxford University Press.

Figure 21.13: Courtesy of JM Zuo, modified from Kim, M *et al.* (2004) Appl. Phys. Lett. 84 2181, Fig. 1.

Figure 21.14: Modified from Johnson, A (2007) Acta Cryst. **B63** 511, Fig.7 reproduced by permission of The International Union of Crystallography. Courtesy A Johnson.

Figure 21.15: Modified from Zuo, JM *et al.* (1999) Nature **401** 49, Fig. 3A reproduced by permission of Macmillan Magazines Ltd. Courtesy JCH Spence.

Figure 21.16: Courtesy of R McConville, from Williams, DB *et al.* Eds. (1992) *Images of Materials*, Fig. 6.33 reproduced by permission of Oxford University Press.

Figure 21.17: Courtesy of JM Cowley, from Liu, M and Cowley, JM (1994) Ultramicrosc. **53** 333, Figs. 1, 2 reproduced by permission of Elsevier Science BV.

Table 21.1: Data from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition p 79, reproduced by permission of Philips Electron Optics.

Table 21.2: Data from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition p 79, reproduced by permission of Philips Electron Optics.

Chapter 22

Figure 22.5: Courtesy of KA Repa.

Figure 22.6: From Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 3.7D reproduced by permission of Philips Electron Optics.

Figure 22.7: Courtesy of KB Reuter.

Figure 22.8: From Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 3.7C reproduced by permission of Philips Electron Optics.

Figure 22.9A,B: Courtesy of HL Tsai, from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 1.19A, B reproduced by permission of Philips Electron Optics.

Figure 22.9C: Courtesy of K-R Peters.

Figure 22.10: Modified from Williams, DB (1983) in Krakow, W *et al.* (Eds.) *Electron Microscopy of Materials* Mater. Res. Soc. Symp. Proc. **31** 11, Fig. 3A,B.

Figure 22.11A,B: Courtesy of IM Watt, from Watt, I (1996) *The Principles and Practice of Electron Microscopy,* 2nd Edition, Fig. 5.5A,B reproduced by permission of Cambridge University Press.

Figure 22.12: Courtesy of MMJ Treacy, from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 5.26B reproduced by permission of Philips Electron Optics.

Figure 22.14: Courtesy of SJ Pennycook, from Pennycook, SJ *et al.* (1986) J. Microsc. **144** 229, Fig. 8 reproduced by permission of the Royal Microscopical Society.

Figure 22.15A,B: Courtesy of SJ Pennycook, from Lyman, CE (1992) *Microscopy: The Key Research Tool*, special publication of the EMSA Bulletin **22** 7, Fig. 7 reproduced by permission of MSA.

Figure 22.15C: Courtesy of SJ Pennycook, from Browning, ND *et al.* (1995) Interface Science **2** 397, Fig. 4D reproduced by permission of Kluwer.

Figure 22.16A: From Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Fig. 2.34 reproduced by permission of Philips Electron Optics.

Figure 22.17: Courtesy of D Cohen.

Chapter 23

Figure 23.3A: From Izui, KJ *et al.* (1977) J. Electron Microsc. **26** 129, Fig. 1 reproduced by permission of the Japanese Society of Electron Microscopy.

Figure 23.3C: Courtesy of JCH Spence, from Spence, JCH *Experimental High-Resolution Electron Microscopy*, Fig. 5.15 reproduced by permission of Oxford University Press.

Figure 23.4B: Courtesy of JL Hutchison, from Hutchison, JL *et al.* (1991) in J Heydenreich and W Neumann Eds. *High-Resolution Electron Microscopy—Fundamentals and Applications* p 205, Fig. 3 reproduced by permission of Halle/Saale.
Figure 23.4C: Courtesy of S McKernan.
Figure 23.4D: From Carter, CB *et al.* (1989) Philos. Mag. **A63** 279, Fig. 3 reproduced by permission of Taylor and Francis.
Figure 23.8: From Tietz, LA *et al.* (1992) Philos. Mag. **A65** 439, Figs. 3A, 12A,C reproduced by permission of Taylor and Francis.
Figure 23.10: Courtesy of J Zhu.
Figure 23.12: Modified from Vincent, R (1969) Philos. Mag. **19** 1127, Fig. 4.
Figure 23.13: Modified from Norton, MG and Carter, CB (1995) J. Mater. Sci. 30, Fig. 6.
Figure 23.14: Courtesy of U Dahmen, from Hetherington, CJD and Dahmen, U (1992) in PW Hawkes Ed. Signal and Image Processing in Microscopy and Micro-analysis *Scanning Microscopy* Supplement **6** 405, Fig. 9 reproduced by permission of Scanning Microscopy International.
Figure 23.15: From Heidenreich, RD (1964) *Fundamentals of Transmission Electron Microscopy*, Figs. 5.4, 5.6 reproduced by permission of John Wiley & Sons Inc.
Figure 23.16A: Modified from Heidenreich, RD (1964) *Fundamentals of Transmission Electron Microscopy*, Fig. 11.2 original by permission of John Wiley & Sons Inc.
Figure 23.16B: From Boersch, H *et al.* (1962) Z. Phys. **167** 72, Fig. 4 reproduced by permission of Springer Verlag.
Figure 23.17: Courtesy of M Rühle.
Figure 23.18: Modified from Kouh, YM *et al.* (1986) J. Mater. Sci. **21** 2689, Fig, 9.
Figure 23.19: Courtesy of M Rühle, from Rühle, M and Sass, SL (1984) Philos. Mag. **A49** 759, Fig. 2 reproduced by permission of Taylor and Francis.
Figure 23.20B,E: From Carter, CB *et al.* (1986) Philos. Mag. **A55** 21, Fig. 11 reproduced by permission of Taylor and Francis.

Chapter 24

Figure 24.1: Courtesy of S Ramamurthy.
Figure 24.2: Redrawn after Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Fig. 3.2A.
Figure 24.3B: Courtesy of D Cohen.
Figure 24.3C: Courtesy of S King when not busy founding Cricinfo.
Figure 24.5: Courtesy of D Susnitzky.
Figure 24.7: Redrawn after Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Fig. 3.3.
Figure 24.8: Redrawn after Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Figs. 3.4B,D. Images reproduced by permission of Philips Electron Optics.
Figure 24.9: Courtesy of S Ramamurthy.
Figure 24.10: From Hashimoto, H *et al.* (1962) Proc. Roy. Soc. (London) **A269** 80, Fig. 11 reproduced by permission of The Royal Society.
Figure 24.11A: Courtesy of NSA Hitachi Scientific Instruments Ltd.
Figure 24.11B,C: Courtesy of D Cohen.
Figure 24.12: From Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Fig. 3.3D reproduced by permission of Philips Electron Optics.
Figure 24.13B,C: From De Cooman, BC *et al.* (1987) in JD Dow and IK Schuller Eds. *Interfaces, Superlattices, and Thin Films,* Mater. Res. Soc. Symp. Proc. **77** 187, Fig. 1 reproduced by permission of MRS.

Chapter 25

Figure 25.4A-D: Courtesy of D Cohen.
Figure 25.4E,F: Modified from Gevers, R *et al.* (1963) Phys. stat. sol. **3** 1563, Table 3.
Figure 25.5: From Föll, H *et al.* (1980) Phys. stat. sol. (a) **58** 393, Fig. 6A,C reproduced by permission of Akademie Verlag GmbH.
Figure 25.7A,B: From Lewis, MH (1966) Philos. Mag. **14** 1003, Fig. 9 reproduced by permission of Taylor and Francis.
Figure 25.7C,D: Courtesy of S Amelinckx, from Amelinckx, S and Van Landuyt, J (1978) in S Amelinckx *et al.* Eds. *Diffraction and Imaging Techniques in Material Science* **I** 107, Figs. 3, 18 North-Holland.
Figure 25.8: From Rasmussen, DR *et al.* (1991) Phys. Rev. Lett. **66** (20) 262, Fig. 2 reproduced by permission of The American Physical Society.
Figure 25.9: Courtesy of S Summerfelt.
Figure 25.13: Modified from Metherell, AJ (1975) in U Valdrè and E Ruedl Eds. *Electron Microscopy in Materials Science* **II** 397, Fig. 13 Commission of the European Communities.
Figure 25.14: From Hashimoto, H *et al.* (1962) Proc. Roy. Soc. (London) **A269** 80, Fig. 15 original by permission of The Royal Society.
Figure 25.16: Modified from Rasmussen, R *et al.* (1991) Philos. Mag. **63** 1299, Fig. 4.

Chapter 26

Figure 26.2B: Modified from Amelinckx, S (1964) Solid State Physics Suppl. **6**, Fig. 76.
Figure 26.6A-C: Modified from Carter, CB (1980) Phys. stat. sol. (a) **62** 139, Fig. 4.

Figure 26.6F: From Van Landuyt, J *et al.* (1970) Phys. stat. sol. **41** 271, Fig. 1 reproduced by permission of Akademie Verlag GmbH.
Figure 26.6G-H: Courtesy of BC De Cooman.
Figure 26.7: Modified from Hirsch, PB *et al.* (1977) *Electron Microscopy of Thin Crystals,* 2nd Edition, Fig. 7.8 Krieger.
Figure 26.8: From Delavignette, P and Amelinckx, S (1962) J. Nucl. Mater. **5** 17, Fig. 7 reproduced by permission of Elsevier Science BV.
Figure 26.10: From Urban, K (1971) in S Koda Ed. *The World Through the Electron Microscope* Metallurgy **V** 26, reproduced by permission of JEOL USA Inc.
Figure 26.11: Courtesy of A Howie, from Howie, A and Whelan, MJ (1962) Proc. Roy. Soc. (London) **A267** 206, Fig. 14 reproduced by permission of The Royal Society.
Figure 26.12: Modified from M Wilkens (1978) in S Amelinckx *et al.* Eds. *Diffraction and Imaging Techniques in Material Science* **I** 185, Fig. 4 North-Holland.
Figure 26.14: From Dupouy, G and Perrier, F (1971) in S Koda Ed. *The World Through the Electron Microscope* Metallurgy **V** 100, reproduced by permission of JEOL USA Inc.
Figure 26.15A: From Modeer, B and Lagneborg, R (1971) in S Koda Ed. *The World Through the Electron Microscope* Metallurgy **V** 44, reproduced by permission of JEOL USA Inc.
Figure 26.15B: Courtesy of DA Hughes, from Hansen, N and Hughes, DA (1995) Phys. stat. sol. (a) **149** 155, Fig. 5 reproduced by permission of Akademie Verlag GmbH.
Figure 26.16A: From Siems, F *et al.* (1962) Phys. stat. sol. **2** 421, Fig. 5A reproduced by permission of Akademie Verlag GmbH.
Figure 26.16C: From Siems, F *et al.* (1962) Phys. stat. sol. **2** 421, Fig. 15A reproduced by permission of Akademie Verlag GmbH.
Figure 26.17A: Modified from Whelan, MJ (1958–1959) J. Inst. Met. **87** 392, Fig. 25A.
Figure 26.17B: Courtesy of K Ostyn.
Figure 26.18: From Takayanagi, L (1988) Surface Science **205** 637, Fig. 5 reproduced by permission of Elsevier Science BV.
Figure 26.19A: From Tunstall, WJ *et al.* (1964) Philos. Mag. **9** 99, Fig. 9 reproduced by permission of Taylor and Francis.
Figure 26.19B: From Amelinckx, S in PG Merli and VM Anti-sari Eds. *Electron Microscopy in Materials Science* p 128, Fig. 45 reproduced by permission of World Scientific.
Figure 26.20: Courtesy of W Skrotski.
Figure 26.21: Courtesy of W Skrotski.
Figure 26.22: From Carter, CB *et al.* (1986) Philos. Mag. **A55** 21, Fig. 2 reproduced by permission of Taylor and Francis.
Figure 26.23: From Carter, CB *et al.* (1981) Philos. Mag. **A43** 441, Fig. 3 reproduced by permission of Taylor and Francis.
Figure 26.24: Courtesy of K Ostyn.
Figure 26.25: Courtesy of L Tietz.
Figure 26.26A: Courtesy of LM Brown, from Ashby, MF and Brown, LM (1963) Philos. Mag. **8** 1083, Fig. 10 reproduced by permission of Taylor and Francis.
Figure 26.26B: Modified from Whelan, MJ (1978) in S Amelinckx *et al.* Eds. *Diffraction and Imaging Techniques in Material Science* **I** 43, Fig. 36 North-Holland.
Figure 26.26C: Courtesy of LM Brown, from Ashby, MF and Brown, LM (1963) Philos. Mag. **8** 1083, Fig. 12 reproduced by permission of Taylor and Francis.
Figure 26.27: From Rasmussen, DR and Carter, CB (1991) J. Electron Microsc. Technique **18** 429, Fig. 2 reproduced by permission of John Wiley & Sons Inc.

Chapter 27

Figure 27.7: Courtesy of S King.
Figure 27.10: Courtesy of DJH Cockayne, from Cockayne, DJH (1972) Z. Naturforschung **27a** 452, Fig. 6 original by permission of Verlag der Zeitschrift für Naturforschung, Tübingen.
Figure 27.13: Modified from Carter, CB *et al.* (1986) Philos. Mag. **A55** 1, Fig. 9.
Figure 27.15: Modified from Föll, H *et al.* (1980) Phys. stat. sol. (a) **58** 393, Fig. 6B,C.
Figure 27.17: From Heidenreich. RD (1964) *Fundamentals of Transmission Electron Microscopy*, Fig. 9.20 reproduced by permission of John Wiley & Sons Inc.
Figure 27.18: Courtesy of DJH Cockayne, from Ray, ILF and Cockayne, DJH (1971) Proc. Roy. Soc. (London) **A325** 543, Fig. 10 reproduced by permission of The Royal Society.
Figure 27.23: Modified from Carter, CB (1979) J. Phys. (A) **54** (1) 395, Fig. 8A.

Chapter 28

Figure 28.4: Courtesy of R Gronsky, from Gronsky, R (1992) in DB Williams *et al.* Eds. *Images of Materials*, Fig. 7.6 original by permission of Oxford University Press.
Figure 28.5: Courtesy of S McKernan.
Figure 28.6: Courtesy of S McKernan.
Figure 28.7: Modified from Cowley, JM (1988) in PR Buseck *et al.* Eds. *High-Resolution Electron Microscopy and Associated Techniques*, Fig. 1.9 Oxford University Press.
Figure 28.8: Courtesy of JCH Spence, from Spence, JCH (1988) *Experimental High-Resolution Electron*

Microscopy, 2nd Edition, Fig. 4.3 original by permission of Oxford University Press.

Figure 28.10: From de Jong, AF and Van Dyck, D (1993) Ultramicrosc. **49** 66, Fig. 1 original by permission of Elsevier Science BV.

Figure 28.11: Courtesy of MT Otten, from Otten, MT and Coene, WMJ (1993) Ultramicrosc. **48** 77, Fig. 8 reproduced by permission of Elsevier Science BV.

Figure 28.12: Courtesy of MT Otten, from Otten, MT and Coene, WMJ (1993) Ultramicrosc. **48** 77, Fig. 11 reproduced by permission of Elsevier Science BV.

Figure 28.13: Courtesy of MT Otten, from Otten, MT and Coene, WMJ (1993) Ultramicrosc. **8** 77, Fig. 10 reproduced by permission of Elsevier Science BV.

Figure 28.14A,B: From Amelinckx, S *et al.* (1993) Ultramicrosc. **51** 90, Fig. 2 original by permission of Elsevier Science BV.

Figure 28.15: From Amelinckx, S *et al.* (1993) Ultramicrosc. **51** 90, Fig. 3 reproduced by permission of Elsevier Science BV.

Figure 28.16: From Rasmussen, DR *et al.* (1995) J. Microsc. **179** 77, Fig. 2C,D reproduced by permission of the Royal Microscopical Society.

Figure 28.18A: Courtesy of S McKernan.

Figure 28.18B: From Berger, A *et al.* (1994) Ultramicrosc. **55** 101, Fig. 4B reproduced by permission of Elsevier Science BV.

Figure 28.18C: Courtesy of S Summerfelt.

Figure 28.18D: Courtesy of S McKernan.

Figure 28.19: Courtesy of DJ Smith.

Figure 28.21A: From Van Landuyt, J *et al.* (1991) in J Heydenreich and W Neumann Eds. *High-Resolution Electron Microscopy—Fundamentals and Applications,* p 254, Fig. 6 reproduced by permission of Halle/Saale.

Figure 28.21D: From Van Landuyt, J *et al.* (1991) in J Heydenreich and W Neumann Eds. *High-Resolution Electron Microscopy—Fundamentals and Applications,* p 254, Fig. 8 reproduced by permission of Halle/Saale.

Figure 28.22: From Nissen, H-U and Beeli, C (1991) in J Heydenreich and W Neumann Eds. *High-Resolution Electron Microscopy—Fundamentals and Applications,* p 272, Fig. 4 reproduced by permission of Halle/Saale.

Figure 28.23: From Nissen, H-U and Beeli, C (1991) in J Heydenreich and W Neumann Eds. *High-Resolution Electron Microscopy—Fundamentals and Applications,* p 272, Fig. 2 reproduced by permission of Halle/Saale.

Figure 28.24: From Parsons, JR *et al.* (1973) Philos. Mag. **29** 1359, Fig. 2 reproduced by permission of Taylor and Francis.

Table 28.1: Modified from de Jong, AF and Van Dyck, D (1993) Ultramicrosc. **49** 66, Table 1.

Chapter 29

Figure 29.2: Courtesy of R Sinclair, from Sinclair, R *et al.* (1981) Met. Trans. 12A, 1503, Figs. 13, 14 reproduced by permission of ASM International.

Figure 29.4A,B: From Marcinkowksi, MJ and Poliak, RM (1963) Philos. Mag. 8, 1023, Fig. 15a,b reproduced by permission of Taylor and Francis.

Figure 29.4C,D: Courtesy of J Silcox, from Silcox, J (1963) Philos. Mag. 8, 7, Fig. 7 reproduced by permission of Taylor and Francis.

Figure 29.5: Courtesy of AJ Craven, from Buggy, TW *et al.* (1981) *Analytical Electron Microscopy-1981,* p 231, Fig. 5 reproduced by permission of San Francisco Press.

Figure 29.6D,E: Courtesy of NSA Hitachi Scientific Instruments Ltd. and S McKernan.

Figure 29.7: Courtesy of R Sinclair.

Figure 29.8: From Kuesters, K-H *et al.* (1985) J. Cryst. Growth **71**, 514, Fig. 4 reproduced by permission of Elsevier Science BV.

Figure 29.9: Courtesy of M. Mallamaci.

Figure 29.10A: Modified from De Cooman, BC *et al.* (1985) J. Electron Microsc. Tech. **2**, 533, Fig. 1.

Figure 29.10B: Courtesy of SM Zemyan.

Figure 29.10C-E: Courtesy of BC De Cooman.

Figure 29.12: Courtesy of G Thomas, from Bell, WL and Thomas, G (1972) in G Thomas *et al.* Eds. *Electron Microscopy and Structure of Materials,* p 53, Fig. 28 reproduced by permission of University of California Press.

Figure 29.13: Courtesy of K-R Peters, from Peters, K-R (1984) in DF Kyser *et al.* Eds. *Electron Beam Interactions with Solids for Microscopy, Microanalysis and Lithography,* p 363, Fig. 1 original by permission of Scanning Microscopy International.

Figure 29.14: Courtesy of R McConville, from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 3.11 reproduced by permission of Philips Electron Optics.

Figure 29.15: Courtesy of Philips Electronic Instruments, from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 3.10 reproduced by permission of Philips Electron Optics.

Figure 29.16: Courtesy of H Lichte, from Lichte, H (1992) Scanning Microscopy, p 433, Fig. 1 reproduced by permission of Scanning Microscopy International.

Figure 29.17: Modified from Lichte, H (1992) Ultramicrosc. 47, 223, Fig. 1.

Figure 29.18: Modified from Tonomura, A. Courtesy of NSA Hitachi Scientific Instruments Ltd.

Figure 29.19A-C: From Tonomura, A (1992) Adv. Phys. 41, 59, Fig. 29 reproduced by permission of Taylor and Francis.
Figure 29.19D: From Tonomura, A (1987) Rev. Mod. Phys. 59, 639, Fig. 41 reproduced by permission of The American Physical Society.
Figure 29.20A: From Tonomura, A (1992) Adv. Phys. 41, 59, Fig. 38, reproduced by permission of Taylor and Francis.
Figure 29.20B: From Tonomura, A (1992) Adv. Phys. 41, 59, Fig. 42 reproduced by permission of Taylor and Francis.
Figure 29.20C: From Tonomura, A (1992) Adv. Phys. 41, 59, Fig. 44 reproduced by permission of Taylor and Francis.
Figure 29.21: Courtesy of R Sinclair, from Sinclair, R *et al*. (1994) Ultramicrosc. 56, 225, Fig. 5 reproduced by permission of Elsevier Science BV.
Figure 29.22: Courtesy of M Treacy.

Chapter 30

Figure 30.2A,B: Courtesy of MA O'Keefe, from O'Keefe, MA and Kilaas, R (1988) in PW Hawkes *et al*. Eds. Image and Signal Processing in Electron Microscopy, Scanning Microscopy Supplement 2 p 225, Fig. 1 original by permission of Scanning Microscopy International.
Figure 30.3: From Kambe, K (1982) Ultramicrosc **10** 223, Fig. lA-D reproduced by permission of Elsevier Science BV.
Figure 30.4: Courtesy of MA O' Keefe, from O'Keefe, MA and Kilaas, R (1988) in PW Hawkes *et al*. Eds. Image and Signal Processing in Electron Microscopy Scanning Microscopy Supplement 2 p 225, Fig. 4 reproduced by permission of Scanning Microscopy International.
Figure 30.5: Modified from Rasmussen, DR and Carter, CB (1990) Ultramicrosc. **32** 337, Figs. 1, 2.
Figure 30.8: From Beeli, C and Horiuchi, S (1994) Philos. Mag. **B70** 215, Fig. 6A-D reproduced by permission of Taylor and Francis.
Figure 30.9: From Beeli, C and Horiuchi, S (1994) Philos. Mag. **B70** 215, Fig. 7A-D reproduced by permission of Taylor and Francis.
Figure 30.10: From Beeli, C and Horiuchi, S (1994) Philos. Mag. **B70** 215, Fig. 8 reproduced by permission of Taylor and Francis.
Figure 30.11: From Jiang, J *et al*. (1995) Phil. Mag. Lett. **71** 123, Fig. 4 reproduced by permission of Taylor and Francis.

Chapter 31

Figure 31.1: Courtesy of J. Heffelfinger.
Figure 31.2: From Rasmussen, DR *et al*. (1995) J. Microsc. 179, 77, Fig. 1b original by permission of The Royal Microscopical Society.
Figure 31.3: From Rasmussen, DR *et al*. (1995) J. Microsc. 179, 77, Fig. 5 reproduced by permission of The Royal Microscopical Society.
Figure 31.4: Courtesy of OL Krivanek, from Krivanek, OL (1988) in PR Buseck *et al*. Eds. *High-Resolution Electron Microscopy and Associated Techniques,* Fig. 12.6 reproduced by permission of Oxford University Press.
Figure 31.5: Courtesy of OL Krivanek, Krivanek, OL (1988) in PR Buseck *et al*. Eds. *High-Resolution Electron Microscopy and Associated Techniques,* Fig. 12.7 reproduced by permission of Oxford University Press.
Figure 31.6A: Courtesy of JCH Spence, from Spence, JCH and Zuo, JM (1992) *Electron Microdiffraction,* Fig. A1.3 reproduced by permission of Plenum Press.
Figure 31.6B: Courtesy of OL Krivanek, from Krivanek, OL (1988) in PR Buseck *et al*. Eds. *High-Resolution Electron Microscopy and Associated Techniques,* Fig. 12.8 reproduced by permission of Oxford University Press.
Figure 31.7: Courtesy of S McKernan.
Figure 31.8: Courtesy of ZL Wang. Wwang, ZL *et al*. (2007) MRS Bulletin, 109–116. Reproduced by permission of MRS.
Figure 31.9: Courtesy of ZC Lin, from Lin, ZC (1993) Ph.D. dissertation, Fig. 4.15, University of Minnesota.
Figure 31.10: Courtesy of O Saxton, from Kirkland, AI (1992) in PW Hawkes Ed. Signal and Image Processing in Microscopy and Microanalysis, *Scanning Microscopy* Supplement 6, 139, Figs. 1–3 reproduced by permission of Scanning Microscopy International.
Figure 31.11: From Zou, XD and Hovmöller, S (1993) Ultramicrosc. 49, 147, Fig. 1 reproduced by permission of Elsevier Science BV.
Figure 31.12A: From Kirkland, AI *et al*. (1995) Ultramicrosc. 57, 355, Fig. 1 reproduced by permission of Elsevier Science BV.
Figure 31.12B; From Kirkland. AI *et al*. (1995) Ultramicrosc. 57, 355, Fig. 3 reproduced by permission of Elsevier Science BV.
Figure 31.13A-C: From Kirkland, AI *et al*. (1995) Ultramicrosc. 57, 355, Fig. 8 reproduced by permission of Elsevier Science BV.
Figure 31.14: Courtesy of OL Krivanek, from Krivanek, OL and Fan, GY (1992) in PW Hawkes (Ed.) Signal and Image Processing in Microscopy and Microanalysis, Scanning Microscopy Supplement 6, p 105, Fig. 4 reproduced by permission of Scanning Microscopy International.
Figure 31.15: Courtesy of OL Krivanek, Krivanek, OL and Fan, GY (1992) in PW Hawkes (Ed.) Signal and Image Processing in Microscopy and Microanalysis,

Scanning Microscopy Supplement 6, p 105, Fig. 5 reproduced by permission of Scanning Microscopy International.

Figure 31.16: After U Dahmen, from Paciornik, S *et al*. (1996) Ultramicrosc. 62, 15, Fig. 1.

Figure 31.17: Courtesy of U Dahmen, from Paciornik, S *et al*. (1996) Ultramicrosc. 62, 15, Fig. 5 reproduced by permission of Elsevier Science BV.

Figure 31.18: Courtesy of A Ourmazd, from Kisielowski, C *et al*. (1995) Ultramicrosc. 58, 131, Figs. 2–4 reproduced by permission of Elsevier Science BV.

Figure 31.19: Courtesy of A Ourmazd, from Kisielowski, C *et al*. (1995) Ultramicrosc. 58, 131, Figs. 8, 10, 12 reproduced by permission of Elsevier Science BV.

Figure 31.20A,B: Data from Ourmazd, A *et al*. (1990) Ultramicrosc. 34, 237, Fig. 1.

Figure 31.20C,D: From Ourmazd, A *et al*. (1990) Ultramicrosc. 34, 237, Figs. 2, 5 reproduced by permission of Elsevier Science BV. Courtesy of A Ourmazd.

Figure 31.21A-F: From Kisielowski, C *et al*. (1995) Ultramicrosc. 34, 237, Fig. I5 reproduced by permission of Elsevier Science BV. Courtesy of A Ourmazd.

Figure 31.22: Courtesy of U Dahmen, from Paciornik, S *et al*. (1996) *Ultramicroscopy,* in press, Fig. 2 reproduced by permission of Elsevier Science BV.

Figure 31.23: From King, WE and Campbell, GH (1994) Ultramicrosc. 56, 46, Fig. 1 reproduced by permission of Elsevier Science BV.

Figure 31.24: From King, WE and Campbell, GH (1994) Ultramicrosc. 56, 46, Fig. 6 reproduced by permission of Elsevier Science BV.

Figure 31.25: Courtesy of M Rühle, from Möbus, G *et al*. (1993) Ultramicrosc. 49, 46, Fig. 6 reproduced by permission of Elsevier Science BV.

Figure 31.26: From Thon, F (1970) in U Valdrè Ed. *Electron Microscopy in Materials Science,* p 571, Fig. 36 reproduced by permission of Academic Press.

Figure 31.27: Courtesy of J Heffelfinger.

Chapter 32

Figure 32.1: Courtesy of JE Yehoda, from Messier, R and Yehoda, JE (1985) J. Appl. Phys. **58** 3739, Fig. 1 reproduced by permission of the American Institute of Physics.

Figure 32.2: Courtesy of M Watanabe.

Figure 32.3B,C: Courtesy of JH Scott.

Figure 32.4A: Courtesy of JH Scott.

Figure 32.4B: Courtesy of P Statham, reproduced by permission of Oxford Instruments.

Figure 32.5: Courtesy of N Rowlands, reproduced by permission of Oxford Instruments.

Figures 32.6,7: Courtesy of SM Zemyan.

Figure 32.8A-C: Courtesy of JH Scott, modified from figure originally supplied by Photon Detector Technologies.

Figure 32.8D: Courtesy of DE Newbury.

Figure 32.9A-C: Courtesy of M Terauchi.

Figure 32.9D: Courtesy of D Wollman, SW Nam and DE Newbury.

Figure 32.11: Courtesy of SM Zemyan, modified from Zemyan, S and Williams, DB (1995) in DB Williams *et al*. Eds. *X-ray Spectrometry in Electron Beam Instruments*, Fig. 12.9 original by permission of Plenum Press.

Figure 32.12A: Courtesy of SM Zemyan, modified from Zemyan, SM and Williams, DB (1995) in DB Williams *et al*. Eds. *X-ray Spectrometry in Electron Beam Instruments,* Fig. 12.10 original by permission of Plenum Press.

Figure 32.12B: Courtesy of JH Scott, modified from diagram originally supplied by Photon Detector Technologies .

Figure 32.13: Courtesy of JJ Friel, modified from Mott, RB and Friel, JJ (1995) in DB Williams *et al*. Eds. *X-ray Spectrometry in Electron Beam Instruments,* Fig. 9.8 original reproduced by permission of Plenum Press.

Figure 32.14: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 4.5A original reproduced by permission of Philips Electron Optics.

Figure 32.15: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 4.30 original reproduced by permission of Philips Electron Optics.

Chapter 33

Figure 33.1: Courtesy of SM Zemyan.

Figure 33.2: Courtesy of SM Zemyan.

Figure 33.3: Courtesy of DE Newbury, from Newbury, DE (1995) in DB Williams *et al*. Eds. *X-ray Spectrometry in Electron Beam Instruments*, Fig. 11.18 reproduced by permission of Plenum Press.

Figure 33.4: Courtesy of SM Zemyan.

Figure 33.5A,B: Courtesy of G Cliff, modified from Cliff, G and Kenway, PB (1982) Microbeam Analysis-1982 p 107, Figs. 5, 4 original reproduced by permission of San Francisco Press.

Figure 33.6: Modified from Williams, DB and Goldstein, JI (1981) in KFJ Heinrich *et al*. Eds. *Energy-Dispersive X-ray Spectrometry* p 346, Fig. 7A NBS.

Figure 33.7: Courtesy of SM Zemyan.

Figure 33.8: Courtesy of SM Zemyan.

Figure 33.9A: Courtesy of SM Zemyan.

Figure 33.9B: Courtesy of KS Vecchio, modified from Vecchio, KS and Williams, DB (1987) J. Microsc. **147** 15, Fig. 1 original by permission of the Royal Microscopical Society.

Figure 33.10A: Courtesy of SM Zemyan.

Figure 33.10B: Courtesy of SM Zemyan, modified from Zemyan, SM and Williams, DB (1994) J. Microsc. **174** 1, Fig. 6.

Figure 33.10C: Courtesy of SM Zemyan, modified from Zemyan, SM and Williams, DB (1995) in DB Williams *et al*. Eds. *X-ray Spectrometry in Electron Beam Instruments*, Fig. 12.7.
Figure 33.11: Courtesy of M Watanabe.
Figure 33.12: Courtesy of M Watanabe.
Figure 33.13A-C: Courtesy of CE Lyman, from Lyman, CE (1992) in CE Lyman *et al*. Eds. *Compositional Imaging in the Electron Microscope: An Overview,* Microscopy: The Key Research Tool p 1, Fig. 2.
Figure 33.14A: Modified frontispiece image by Hunneyball, PD *et al*. (1981) in GW Lorimer *et al*. Eds. *Quantitative Microanalysis with High Spatial Resolution* The Institute of Metals, London.
Figure 33.14B,C: Courtesy of DT Carpenter.
Figure 33.15: Courtesy of M Watanabe.

Chapter 34

Figure 34.1: Courtesy of M Watanabe.
Figure 34.2: Courtesy of SM Zemyan and M Watanabe.
Figure 34.3: Courtesy of SM Zemyan and M Watanabe.
Figure 34.4A-D: Courtesy M Watanabe, from Watanabe, M and Williams, DB (2003) *Microsc. Microanal.* **9** Suppl. 2 p 124, Figs. 1–4 reproduced by permission of the Microscopy Society of America.
Figures 34.5–7: Courtesy of SM Zemyan and M Watanabe.
Figure 34.8: Courtesy of CH Kiely. Full report given by Enache, DI *et al*. (2006) Science **311** 362.

Chapter 35

Figures 35.1–4: Courtesy of SM Zemyan and M Watanabe.
Figure 35.5A: Modified from Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 4.20 original by permission of Philips Electron Optics.
Figure 35.5B,C: Courtesy of SM Zemyan.
Figure 35.6: Courtesy of SM Zemyan.
Figure 35.7: Modified from Wood, JE *et al*. (1984) J. Microsc. **133** 255, Fig. 2.8 original by permission of the Royal Microscopical Society.
Figure 35.8: Modified from Bender, BA *et al*. (1980) J. Amer. Ceram. Soc. **63** 149, Fig. 1 original by permission of the American Ceramic Society.
Figure 35.10B: Courtesy of JA Eades, from Christenson, KK and Eades, JA (1986) Proc. 44th EMSA Meeting p 622, Fig. 2 original by permission of the Electron Microscopy Society of America.
Figure 35.11A-C: Courtesy of M Watanabe and MG Burke.
Figure 35.11D: Courtesy of M Watanabe, modified from Watanabe, M *et al*. (2006) Microsc. Microanal. **12** 515, Figs. 10, 11 reproduced by permission of the Microscopy Society of America.
Tables 35.1, 2: From Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Table 4.2A,B reproduced by permission of Philips Electron Optics.
Table 35.3A,B: From Wood, JE *et al*. (1984) J. Microsc. **133** 255, Tables 9, 11 reproduced by permission of the Royal Microscopical Society.
Table 35.4: Courtesy of M Watanabe.

Chapter 36

Figure 36.1: Courtesy of M Watanabe.
Figure 36.2: Courtesy of M Watanabe.
Figure 36.3: Courtesy of JR Michael, modified from Williams, DB *et al*. (1992) Ultramicrosc. **47** 121, Fig. 1.
Figure 36.4: Courtesy of JR Michael, modified from Williams, DB *et al*. (1992) Ultramicrosc. **47** 121, Fig. 2 original by permission of Elsevier Science BV.
Figure 36.5A-C: Courtesy of M Watanabe, modified from Watanabe, M *et al*. (2006) Microsc. Microanal. **12** Suppl. 2, 1568, Figs. 2–4 reproduced by permission of the Microscopy Society of America.
Figure 36.8A,B: From Williams, DB (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition, Fig. 4.27 reproduced by permission of Philips Electron Optics.
Figure 36.9: After Williams, DB *et al*. (2002) J. Electron Microscopy **51** (Suppl), S113, Figure courtesy of M Watanabe.
Figure 36.10: Courtesy of CE Lyman, modified from Lyman, CE (1987) in J Kirschner *et al*. Eds. *Physical Aspects of Microscopic Characterization of Materials* p 123, Fig. 1, original reproduced by permission of Scanning Microscopy International.
Figure 36.11: Courtesy of M Watanabe, modified from Watanabe, M *et al*. (2006) Microsc. Microanal. **12** 515, Fig. 13; in turn modified from Lyman, CE (1987) in J Kirschner *et al*. Eds. *Physical Aspects of Microscopic Characterization of Materials* p 123, Fig. 7.
Figure 36.12: Courtesy of M Watanabe, modified from Watanabe, M *et al*. (2006) Microsc. Microanal. **12** 515, Fig. 5 reproduced by permission of the Microscopy Society of America.
Figure 36.13: Courtesy of M Watanabe.

Chapter 37

Figure 37.1: Courtesy of J Bruley.
Figure 37.2A Courtesy of JA Hunt, original by permission of Gatan Inc.
Figure 37.2B,C: Courtesy of RF Egerton, modified from Egerton, RF (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope,* 2nd edition,

Fig. 2.2 original reproduced by permission of Plenum Press.
Figure 37.3: Courtesy of K Scudder, reproduced by permission of Gatan Inc.
Figure 37.4A: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.
Figure 37.4B: Courtesy of M Watanabe.
Figure 37.5: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.
Figure 37.9A: Courtesy of JA Hunt, modified from Hunt, JA and Williams, DB (1994) Acta Microsc. **3** 1, Fig. 7 original by permission of the Venezuelan Society for Electron Microscopy.
Figure 37.9B,C: Courtesy of C Colliex.
Figure 37.10: Courtesy of JA Hunt, from Hunt, JA and Williams, DB (1994) Acta Microsc. **3** 1, Fig. 5 reproduced by permission of the Venezuelan Society for Electron Microscopy.
Figure 37.11: Courtesy of JA Hunt, from Hunt, JA and Williams, DB (1994) Acta Microsc. **3** 1, Fig. 4 reproduced by permission of the Venezuelan Society for Electron Microscopy.
Figure 37.12: Courtesy of J Bruley, from Hunt, JA and Williams, DB (1994) Acta Microsc. **3** 1, Fig. 6 reproduced by permission of the Venezuelan Society for Electron Microscopy.
Figure 37.13A: Courtesy of OL Krivanek, modified from Krivanek, OL *et al*. (1991) Microsc. Microanal. Microstruct. **2** 315, Fig. 8.
Figure 37.13B: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.
Figure 37.14: Courtesy of M Watanabe.
Figure 37.15A,B: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.
Figure 37.16A,B: Courtesy of F Hofer.
Figure 37.17A: Courtesy of V Dravid.
Figure 37.17B: Courtesy of C Colliex.
Figure 37.17C,D: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.

Chapter 38

Figure 38.1: Courtesy of J Bruley.
Figures 38.2–4: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.
Figure 38.5A: Courtesy of J Bruley.
Figure 38.5B: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.
Figure 38.6A,B: Courtesy of RH French, modified from van Benthem, K, Elsässer, C and French, RH (2001) J. Appl. Phys. **90** 6156, Figs.1, 7.
Figure 38.7A: Courtesy of J Bruley.
Figure 38.7B: Courtesy of JA Hunt, reproduced by permission of Gatan Inc.
Figure 38.8A,B: From Williams, DB and Edington, JW (1976) Acta Metall. **24** 323, Fig. 7 reproduced by permission of Elsevier Science BV.
Figure 38.8C: Courtesy of AJ Strutt.
Figure 38.9A: Courtesy of JA Hunt.
Figure 38.9B: Courtesy of M Libera, modified from Kim *et al*. (2006) J. Am. Chem. Soc**.** 128 6570, Figs.1, 2.
Figure 38.10A,B: Courtesy of JA Hunt.
Figure 38.11: Courtesy of VJ Keast.
Table 38.1: Courtesy of RF Egerton, from Egerton, RF (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscopy,* 2nd edition p 157, Table 3.2.

Chapter 39

Figure 39.1: Courtesy of OL Krivanek, modified from Ahn, CC and Krivanek, OL (1983) *EELS Atlas* p iv, original by permission of Gatan Inc.
Figure 39.2: Courtesy of J Bruley, modified from Joy, DC (1986) in DC Joy *et al*. Eds. *Principles of Analytical Electron Microscopy* p 249, Fig. 8 Plenum Press.
Figure 39.3: Courtesy of J Bruley.
Figure 39.4: Courtesy of CE Lyman, modified from Lyman, CE (1987) in J Kirschner *et al*. Eds. *Physical Aspects of Microscopic Characterization of Materials* p 123, Fig. 2 original by permission of Scanning Microscopy International.
Figure 39.5: Courtesy of DC Joy, modified from Joy, DC (1979) in JJ Hren *et al*. Eds. *Introduction to Analytical Electron Microscopy* p 235, Fig. 7.6 original by permission of Plenum Press.
Figure 39.6: Courtesy of M Kundmann.
Figure 39.7: Courtesy of J Bruley.
Figure 39.8: Courtesy of K Sato and Y Ishiguro, modified from Sato, K and Ishiguro, Y (1996) Materials Transactions, **37** 643, Figs. 1, 7 Japan Institute of Metals.
Figure 39.9: Courtesy of J Bruley.
Figure 39.10: Courtesy of JA Hunt, from Hunt, JA and Williams, DB (1994) Acta Microsc. **3** 1, Fig. 14 original by permission of the Venezuelan Society for Electron Microscopy.
Figure 39.11: Courtesy of JA Hunt, from Williams, DB and Goldstein, JI (1992) Microbeam Analysis **1** 29, Fig. 11 reproduced by permission of VCH.
Figures 39.12,13: Courtesy of JA Hunt, from Hunt, JA and Williams, DB (1994) Acta Microsc. **3** 1, Fig. 17A,B reproduced by permission of the Venezuelan Society for Electron Microscopy.
Figure 39.14: Courtesy of RF Egerton, modified from Egerton, RF (1993) Ultramicrosc. **50** 13, Fig. 6.
Figure 39.15–18: Courtesy of J Bruley.
Figure 39.19A,B: Courtesy of JA Hunt, from Hunt, JA and Williams, DB (1994) Acta Microsc. **3** 1, Fig. 16

reproduced by permission of the Venezuelan Society for Electron Microscopy.
Figure 39.20: Courtesy of F Hofer.
Figure 39.21: Modified from Egerton, RF (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope,* 2nd edition, Fig. 1.11 original by permission of Plenum Press. Courtesy of RF Egerton.
Figure 39.22: Courtesy of M Varela.
Figure 39.23: Courtesy of M Varela.

Chapter 40

Figure 40.4: Modified from Zaluzec, NJ (1982) Ultramicrosc. **9** 319, Fig. 3. Courtesy of NJ Zaluzec.
Figure 40.5B: Courtesy of J Bruley.
Figure 40.6: Courtesy of PE Batson, from Batson, PE (1993) Nature **366** 727, Fig. 1 reproduced by permission of Macmillan Journals Ltd.
Figures 40.7A,B: Courtesy of VJ Keast, modified from Keast, VJ *et al*. (1998) Acta Mater. **46** 481.
Figure 40.8. Courtesy of RF Brydson, modified from Garvie, LAJ *et al*. (1994) Amer. Mineralogist **79** 411, Fig. 4.
Figure 40.11: Courtesy of J Bruley and J Mayer.
Figure 40.12: Courtesy of J Bruley.
Figure 40.13 Modified from Alamgir, FM *et al*. (2000) Microscopy and Microanalysis Suppl. **2** 194. Courtesy FM Almagir.
Figure 40.14 Modified from Botton, GA (2005) J. Electr. Spect. Rel. Phen. **143** 129, Fig. 5. Courtesy of GA Botton.
Figure 40.15: Courtesy of M Aronova and RD Leapman, modified from Leapman, RD and Aronova, MA (2006) in *Cellular Electron Microscopy* Ed. JR McIntosh.

Part 1

Basics

1

The Transmission Electron Microscope

CHAPTER PREVIEW

A typical commercial transmission electron microscope (TEM) costs about $5 for each electron volt (eV) of energy in the beam and, if you add on all available options, it can easily cost up to $10 per eV. As you'll see, we use beam energies in the range from 100,000 to 400,000 eV, so a TEM is an extremely expensive piece of equipment. Consequently, there have to be very sound scientific reasons for investing such a large amount of money in one microscope. In this chapter (which is just a brief overview of many of the concepts that we'll talk about in detail throughout the book) we start by introducing you to some of the historical development of the TEM because the history is intertwined with some of the reasons why you need to use a TEM to characterize materials. Other reasons for using a TEM have appeared as the instrument continues to develop, to the point where it can seriously be claimed that no other scientific instrument exists which can offer such a broad range of characterization techniques with such high spatial and analytical resolution, coupled with a *completely* quantitative understanding of the various techniques. Indeed as nanotechnology and related areas seize both the public and the technological community's imaginations, it is increasingly obvious that the TEM is the central tool for complete characterization of nanoscale materials and devices. Unfortunately, coupled with the TEM's advantages are some serious drawbacks and you must be just as aware of the instrument's limitations as you are of its advantages, so we summarize these also.

A TEM can appear in several different forms, all of which are described by different acronyms such as HRTEM, STEM, and AEM, and we'll introduce you to these different instruments. We'll also use the same acronyms or initials (go back and read p. xxi) to denote both the technique (microscopy) and the instrument (microscope). We regard all of the different types of TEM as simply variations on a basic theme and that is why only 'TEM' is in the book title. We will also describe some of the basic physical characteristics of the electron. Throughout the book you'll have to confront some physics and mathematics every now and again because understanding what we can do with a TEM and why we operate it in certain ways is governed by the fundamental physics of electrons, how electrons are controlled by magnetic fields in the microscope, how electrons interact with materials, and how we detect the many signals emitted from a specimen in the TEM.

Finally we will summarize some of the most popular computer software packages for TEM. We will refer to many of these throughout the text. We are including them in the first chapter to emphasize the central role of the computer in today's TEM operation and analysis. A basic lesson to take away from this chapter is not just the versatility of the TEM but the fact that it is fundamentally a signal-generating and detecting instrument rather than simply a microscope for high-resolution images and diffraction patterns (we'll call them DPs), which is how it operated for many decades.

1.1 WHAT MATERIALS SHOULD WE STUDY IN THE TEM?

The materials scientist has traditionally examined metals, alloys, ceramics, glasses, polymers, semiconductors, and composite mixtures of these materials, with sporadic adventures into wood, textiles, and concrete. In addition to thinning them from the bulk state, particles and fibers of some of these materials are also commonly studied and, in such shapes, they are sometimes thin enough for direct TEM examination. Nanotechnology, which will feature as a common theme throughout

this book, is defined as "the ability to understand and control matter at dimensions of roughly 1 to 100 nanometers, where unique phenomena enable novel applications. Encompassing nanoscale science, engineering and technology, nanotechnology involves imaging, measuring, modeling, and manipulating matter at this length scale" (URL #1).

THE CRUCIAL WORDS

"Imaging, measuring, modeling, and manipulating matter" can be accomplished with the help of the TEM and are often thrown together as part of the emerging field of 'nanocharacterization,' a term which we will try not to use too often.

When we create nanoscale materials, they come with specific dimensional limits in 1D, 2D, or 3D and the TEM is well suited to observing them, precisely because of these limits. We will include examples of archetypal dimensionally limited structures throughout the book. For example, single layers (such as graphene sheets or quantum wells), nanotubes and nanowires, quantum dots, nanoparticles, and most catalyst particles can be viewed as 1D structures. We can put all of these types of specimen into the TEM without modification, since 1D is always thin enough for direct observation; 2D nanomaterials include interfaces, and complex 3D nanomaterials are typified by multilayer, semiconductor devices, functional materials, or nanoporous structures such as substrates for catalyst-particle dispersions. Lastly, we should note the rapidly growing interface between the nano- and the bio-worlds. While much of biological electron microscopy has been superceded in the last decade or more by less-damaging techniques such as confocal, two-photon, multi-photon, and near-field light microscopies, there is still a major role for TEM in biomaterials, bio/inorganic interfaces, and nano-bio/biomaterials.

1.2 WHY USE ELECTRONS?

Why should we use an electron microscope? Historically TEMs were developed because of the limited image resolution in light microscopes, which is imposed by the wavelength of visible light. Only after electron microscopes were developed was it realized that there are many other equally sound reasons for using electrons, most of which are utilized to some extent in a modern TEM. By way of introduction to the topic, let's look at how the TEM developed and the pros and cons of using such an instrument.

1.2.A An Extremely Brief History

Louis de Broglie (1925) first theorized that the electron had wave-like characteristics, with a wavelength substantially less than visible light. Then in 1927 two

FIGURE 1.1. The electron microscope built by Ruska (in the lab coat) and Knoll, in Berlin in the early 1930s.

research groups, Davisson and Germer and Thomson and Reid, independently carried out their classic electron-diffraction experiments, which demonstrated the wave nature of electrons. It didn't take long for the idea of an electron microscope to be proposed, and the term was first used in the paper of Knoll and Ruska (1932). In this paper they developed the idea of electron lenses into a practical reality and demonstrated electron images taken on the instrument shown in Figure 1.1. This was a most crucial step, for which Ruska received the Nobel Prize ("somewhat late" he was quoted as saying), in 1986, shortly before his death in 1988. Within a year of Knoll and Ruska's publication, the resolution limit of the light microscope was surpassed. Ruska, surprisingly, revealed that he hadn't heard of de Broglie's ideas about electron waves and thought that the wavelength limit didn't apply to electrons. Some idea of the power of Ruska's breakthrough is the fact that commercial TEMs were first developed only 4 years later. The Metropolitan-Vickers EM1 was the first such instrument and was built in the UK in 1936. Apparently it didn't work very well and regular production of commercial TEMs was really started by Siemens and Halske in Germany in 1939. TEMs became widely available from several other sources (Hitachi, JEOL, Philips, and RCA, inter alia) after the conclusion of World War II.

For materials scientists a most important development took place in the mid-1950s when Bollman in Switzerland and Hirsch and co-workers in Cambridge, in the UK, perfected techniques to thin metal foils to electron transparency. (In fact, because so much of the early TEM work examined metal specimens, the word 'foil' came to be synonymous with 'specimen' and we'll often use it this way.) In addition, the Cambridge group also developed the theory of electron-diffraction contrast with which we can now identify, often in a quantitative manner, *all* known line and planar crystal defects

in TEM images. This work is summarized in a formidable but essential text often referred to as the 'Bible' of TEM (Hirsch et al. 1977). For the materials scientist, practical applications of the TEM for the solution of materials problems were pioneered in the United States by Thomas and first clearly expounded in his text. Other materials-oriented texts followed, notably the first student-friendly 'hands-on' text by Edington.

Today TEMs constitute arguably the most efficient and versatile tools for the characterization of materials over spatial ranges from the atomic scale, through the ever-growing 'nano' regime (from < 1 nm to ~ 100 nm) up to the micrometer level and beyond. If you want to read a history of the TEM, the book by Marton (1968) is a compact, personal monograph and the text edited by Hawkes in 1985 contains a series of individual reminiscences. Fujita's (1986) paper emphasizes the substantial contribution of Japanese scientists to the development of the instrument. The field is now at the point where many of the pioneers have put their memoirs down on paper, or Festschrifts have been organized in their honor (e.g., Cosslett 1979, Ruska 1980, Hashimoto 1986, Howie 2000, Thomas 2002, Zeitler 2003) which detail their contributions over the decades, and compile some useful overview papers of the field. If you enjoy reading about the history of science, we strongly recommend the review of *Fifty Years of Electron Diffraction* edited by Goodman (1981) and *Fifty Years of X-ray Diffraction* edited by Ewald (1962) (the spelling of X-ray is discussed in the *CBE Manual*, 1994). More recently, Haguenau et al. (2003) compiled an extensive list of references describing key events in the history of electron microscopy. As always, there is a wealth of information, some of it accurate, available on the Web.

1.2.B Microscopy and the Concept of Resolution

When asked "what is a microscope?," most people would answer that it is an instrument for magnifying things too small to see with the naked eye, and most likely they would be referring to the visible-light microscope (VLM). Because of the general familiarity with the concept of the VLM, we will draw analogies between electron and light microscopes wherever it's instructive.

The smallest distance between two points that we can resolve with our eyes is about 0.1–0.2 mm, depending on how good our eyes are, and assuming that there's sufficient illumination by which to see. This distance is the *resolution* or (more accurately) the *resolving power* of our eyes. So any instrument that can show us pictures (or *images* as we'll often refer to them) revealing detail finer than 0.1 mm could be described as a microscope, and its highest useful magnification is governed by its resolution. A major attraction to the early developers of the TEM was that, since electrons are smaller than atoms, it should be possible, at least theoretically, to build a microscope that could 'see' detail well below the atomic level. The idea of being able to 'see' with electrons may be confusing to you. Our eyes are not sensitive to electrons. If a beam of high-energy electrons was aimed into your eye, you would most likely be blinded as the electrons killed your retinal cells, but you wouldn't see anything (ever again!). So an integral part of any electron microscope is a viewing screen of some form (now usually a flat-panel computer display), which displays electron intensity as light intensity, which we first observe and then record photographically or store digitally. (We'll discuss these screens and other ways of recording electron images in Chapter 7.)

VLM

We'll try to avoid the phrases 'optical microscope' (they all are) and 'light microscope' (some are very heavy).
'Visible-light microscope/y' is simple and appropriate use of the hyphen.

The resolution of a TEM means different things for different functions of the instrument, and we'll discuss them in the appropriate chapters. It's easiest to think of the image resolution in TEM in terms of the classic Rayleigh criterion for VLM, which states that the smallest distance that can be resolved, δ, is given approximately by

$$\delta = \frac{0.61\lambda}{\mu \sin \beta} \tag{1.1}$$

In equation 1.1, λ is the wavelength of the radiation, μ the refractive index of the viewing medium, and β the semi-angle of collection of the magnifying lens. For the sake of simplicity we can approximate $\mu \sin \beta$ (which is sometimes called the numerical aperture) to unity and so the resolution is equal to about half the wavelength of light. For green light in the middle of the visible spectrum, λ is about 550 nm, and so the resolution of a good VLM is about 300 nm. In TEMs we can approximate the best resolution using an expression similar to equation 1.1 (actually $\sim 1.22\lambda/\beta$) which, as we'll see later, is very small.

Now although 300 nm is a small dimension to us, it corresponds to about 1000 atom diameters, and, therefore, many of the features that control the properties of materials are on a scale well below the resolution of the VLM. Also, 300 nm is well above the upper limit of the nano regime which we defined earlier. So there's a real need in nano/materials science and engineering to image details, all the way down to the atomic level, if we want to understand and ultimately control the properties of materials, and that's a major reason why TEMs are so useful.

This limit of light microscopy was well understood at the turn of the last century and prompted Ernst Abbe, one of the giants in the field, to complain that "it is poor comfort to hope that human ingenuity will find ways

and means of overcoming this limit." (He was right to be so depressed because he died in 1905, some 20 years before de Broglie's ingenuity solved the problem.) Louis de Broglie's famous equation shows that the wavelength of electrons is related to their energy, E, and, if we ignore relativistic effects, we can show approximately (and exactly in Section 1.4 below) that (ignoring the inconsistency in units)

$$\lambda = \frac{1.22}{E^{1/2}} \qquad (1.2)$$

In this equation E is in electron volts (eV) and λ in nm.

So from equation 1.2 you can work out that for a 100 keV electron, $\lambda \sim 4$ pm (0.004 nm), which is much smaller than the diameter of an atom.

V AND eV

Remember that we should be precise in our use of these units: V represents the accelerating voltage of the microscope while eV refers to the energy of the electrons in the microscope (look ahead to equation 1.4 to see the relation between the two).

We'll see later that we cannot yet build a 'perfect' TEM that approaches this wavelength-limited limit of resolution, because we can't make perfect electron lenses (see Chapter 6). Until recently, a top of the line lens could rightly be compared to using the bottom of a Coca-Cola™ bottle as a lens for light microscopy. Progress was rapid after Ruska's early work on lenses and since the mid-1970s many commercial TEMs have been capable of resolving individual columns of atoms in crystals, creating the field of high-resolution transmission electron microscopy or HRTEM, which we'll discuss in Chapter 28. A typical HRTEM image is shown in Figure 1.2A.

The advantages of shorter wavelengths led in the 1960s to the development of high-voltage electron microscopes (HVEMs), with accelerating potentials between 1 and 3 MV. In fact, rather than push the resolution limits, most of these instruments were used to introduce controlled amounts of radiation damage into specimens, in an attempt to simulate nuclear-reactor environments. Three-Mile Island and Chernobyl contributed to changes in the emphasis of energy research; recently there has not been much call for HVEMs. Today, climate change is forcing a reconsideration of nuclear power. Only one HVEM (1 MV) for HRTEM imaging was constructed in the 1980s and three 1.25 MV machines in the 1990s. Intermediate voltage electron microscopes (IVEMs) were introduced in the 1980s. These TEMs operate at 200–400 kV, but still offer very high resolution, close to that achieved at 1 MV. In fact, progress is such that most IVEMs purchased these days are, effectively, HRTEMs with atomic resolution.

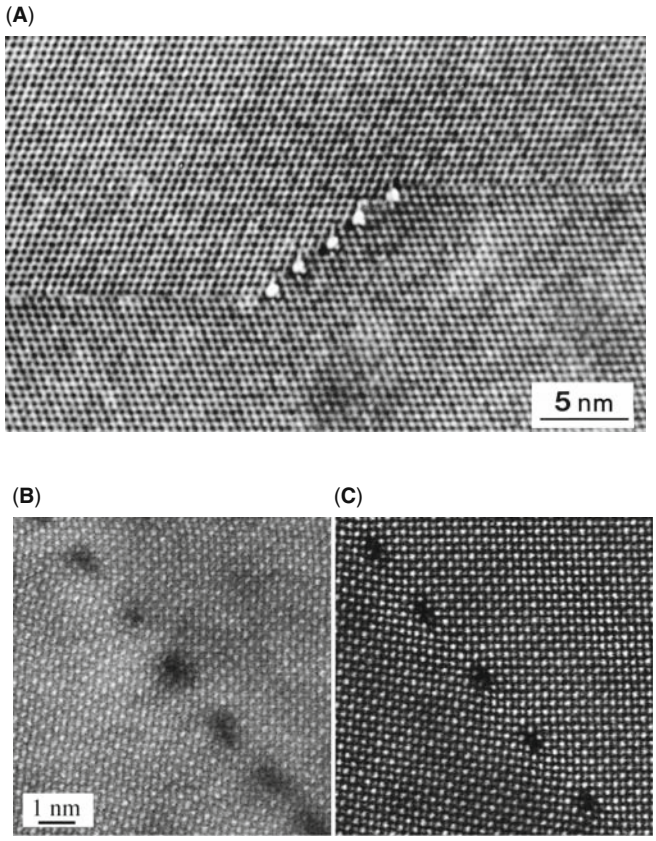


FIGURE 1.2. (A) A twin boundary in spinel stepping from one {111} plane to another parallel plane. The white dots are columns of atoms. The change in atomic orientation across the twin boundary can be readily seen even if we do not know what causes the white dots or why indeed they are white. (B) A grain boundary in $SrTiO_3$ imaged without C_s correction and (C) with C_s correction. As you can see, the effect is just as dramatic as putting on your reading glasses (if you need them).

We are still improving the resolution, and recent breakthroughs in spherical- and chromatic-aberration corrections (see Chapters 6 and 37, respectively) are revolutionizing the TEM field. Among many advantages, corrections of spherical aberration (which, for reasons we'll explain in Chapter 6, we abbreviate to C_s) and chromatic aberration (C_c) allow us to produce sharper atomic-resolution images. By filtering out electrons of different wavelengths we can also better image thicker specimens.

The combination of IVEMs and C_s correction has pushed TEM image resolution to well below the 0.1 nm (1 Å) barrier. Today the point has perhaps been reached where the drive for much better resolution is now no longer paramount and the TEM will develop more constructively in other ways. As we'll illustrate many times throughout the book and elaborate in the companion text, C_s correction is perhaps the most exciting advance in TEM technology in several decades and Figure 1.2B and C shows beautifully the difference in a typical atomic-resolution image with and without C_s correction. The advantages of C_s and C_c aberration correction in

TEM are explored in depth in chapters on C_s correction and energy-filtered TEM (EFTEM) in the companion text.

C_S,C_C AND MAGNIFICATION

Having extolled the virtues of C_s correction it is worth pointing out that most TEM images are recorded at magnifications where such correction makes *no* discernible difference. Most TEM specimens are not thin enough to produce images with resolution that benefits from C_s correction. For thicker specimens C_c correction via energy filtering is much more useful.

1.2.C Interaction of Electrons with Matter

Electrons are one type of ionizing radiation, which is the general term given to radiation that is capable of removing the tightly bound, inner-shell electrons from the attractive field of the nucleus by transferring some of its energy to individual atoms in the specimen.

One of the advantages of using ionizing radiation is that it produces a wide range of secondary signals from the specimen and some of these are summarized in Figure 1.3. Many of these signals are used in analytical electron microscopy (AEM), giving us chemical information and a lot of other details about our specimens. AEM uses X-ray energy-dispersive spectrometry (XEDS) and electron energy-loss spectrometry (EELS). For example, Figure 1.4A shows X-ray spectra from very small regions of the TEM specimen imaged in Figure 1.4B. The spectra exhibit characteristic peaks, which identify the different elements present in different regions. We can transform such spectra into quantitative

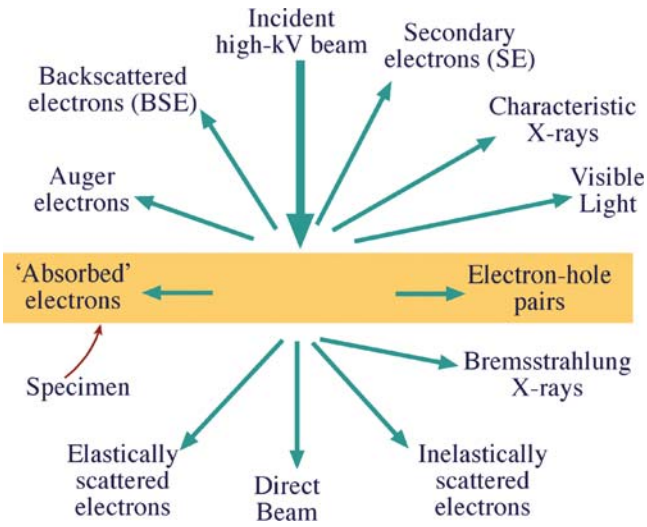


FIGURE 1.3. Signals generated when a high-energy beam of electrons interacts with a thin specimen. Most of these signals can be detected in different types of TEM. The directions shown for each signal do not always represent the physical direction of the signal, but indicate, in a relative manner, where the signal is strongest or where it is detected.

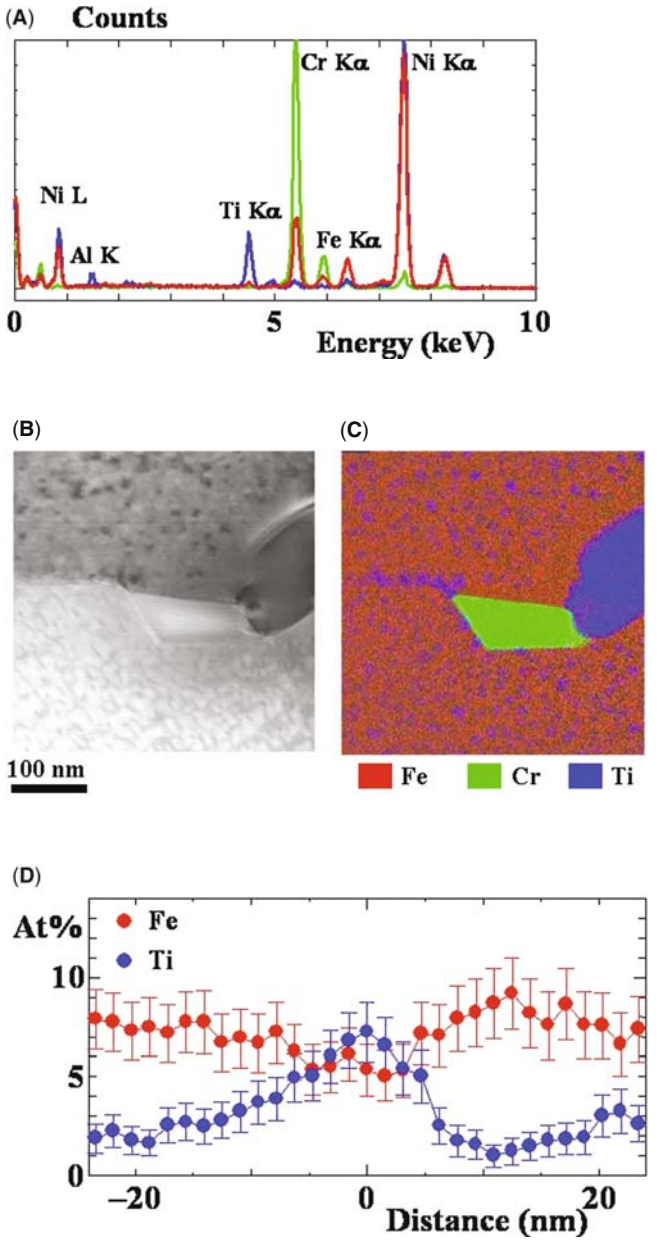


FIGURE 1.4. (A) X-ray spectra from three different regions of a specimen of Ni-base superalloy imaged in (B). The spectra are color-coded to match the different regions of the specimen highlighted in (C) which is a quantitative map showing the distribution of the elements detected in the spectra in (A) (e.g., green areas are rich in Cr, blue areas contain predominantly Ti, etc.). Quantitative composition profiles showing the localized changes in composition across one of the small matrix precipitates in (C) are shown in (D).

images of the distributions of all the elements present in the specimen (Figure 1.4C) and from such images extract quantitative data describing elemental changes associated with inhomogeneous microstructures as shown in Figure 1.4D. This and similar analyses with EELS comprise Part 4 of the book. In contrast, microscopes using non-ionizing radiation, such as visible light, usually only generate light (but not much heat, which is good).

In order to get the best signal out of our specimens we have to put the best signal in, and so the electron source is critical. We are now very accomplished in this respect, as you'll see in Chapter 5; modern TEMs are very good signal-generating instruments. To localize these signals we need our TEM to produce a very small electron beam (or *probe* as it is often called), typically <5 nm and at best < 0.1 nm in diameter. We combine TEM and scanning electron microscope (SEM) technology to create a scanning transmission electron microscope (STEM). The STEM is both the basis for AEMs and a unique scanning-imaging (or scanned-probe) microscope in its own right. In fact there are instruments that are only capable of operating in scanning mode and these are sometimes referred to as dedicated STEMs or DSTEMs. AEMs offer improved analytical performance at intermediate voltages, similar to the improvement in image resolution gained in standard TEMs.

Most importantly, C_s correction permits the generation of smaller electron probes with higher currents, thus significantly improving both analytical spatial resolution and sensitivity. Chromatic-aberration correction (i.e., energy filtering) also offers the opportunity to form images of electrons with a whole range of specific energies, permitting such breakthroughs as band-gap imaging and chemical-bond imaging.

1.2.D Depth of Field and Depth of Focus

The depth of field of a microscope is a measure of how much of the *object* that we are looking at remains in focus at the same time; the term depth of focus refers to the distance over which the *image* can move relative to the object and still remain in focus. If you are confused, it may help to recall that *depth of field* and *field of view* both refer to the *object* in everyday photography. The lenses in the TEM govern these properties just as they determine the resolution. Electron lenses are not very good, as we've already mentioned, and one way to improve their performance is to insert very small limiting apertures, narrowing the beam down to a thin 'pencil' of electrons which at most is a few micrometers across. These apertures obviously cut down the intensity of the electron beam, but they also act to increase the depth of field of the specimen and depth of focus of the images that we produce, as we explain in detail in Chapter 6.

While this large depth of field is chiefly used in the SEM to produce 3D-like images of the surfaces of specimens with large changes in topography, it is also critical in the TEM. It turns out that in the TEM, your specimen is usually in focus from the top to the bottom surfaces at the same time, independent of its topography, so long as it's electron transparent! Figure 1.5 shows a TEM image of some dislocations in a crystal.

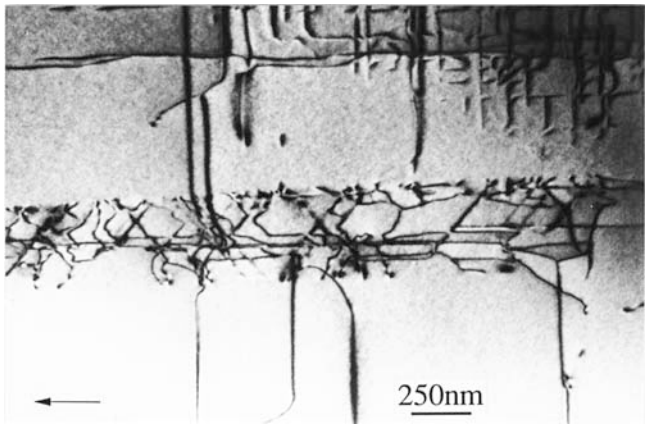


FIGURE 1.5. TEM image of dislocations (dark lines) in GaAs. The dislocations in the band across the middle of the image are on slip planes close to 90° to one another and thread through the thin specimen from the top to the bottom but remain in focus through the foil thickness.

The dislocations appear to start and finish in the specimen, but in fact they are threading their way through the specimen from the top to the bottom surfaces, and they remain in sharp focus at all times. (By the time you finish reading this book, you should be able to work out which is the top and which is the bottom surface of the specimen.) Furthermore, you can record the final image at different positions below the final lens of the instrument and it will still be in focus (although the magnification will change). Compare these properties with the VLM where, as you probably know, unless the surface of your specimen is flat within the wavelength of light, it is not all in focus at the same time. This aspect of TEM gives us both advantages and disadvantages in comparison to the VLM. You should note that, in this rare situation, C_s correction is not an advantage since it permits the use of larger apertures without degrading the resolution of the lens. But *smaller* apertures are the ones that give better depth of focus and depth of field (see Section 6.7). However, if you are using a C_s corrector, your specimen has to be so thin that it will still remain in focus except under extreme conditions. We'll see more on this topic in the companion text and also mention using TEM in a 'confocal' mode.

1.2.E Diffraction

As we've noted, Thompson, Reid, Davisson, and Germer independently showed that electrons could be diffracted when passing through thin crystals of nickel. Performing electron diffraction in TEMs was first realized by Kossel and Möllenstedt (1939). Today, electron diffraction is an indispensable part of TEM and is arguably the most useful aspect for materials scientists and nanotechnologists for whom crystal structure (and particularly crystal defects) is an essential characteristic when it comes to controlling properties. Figure 1.6

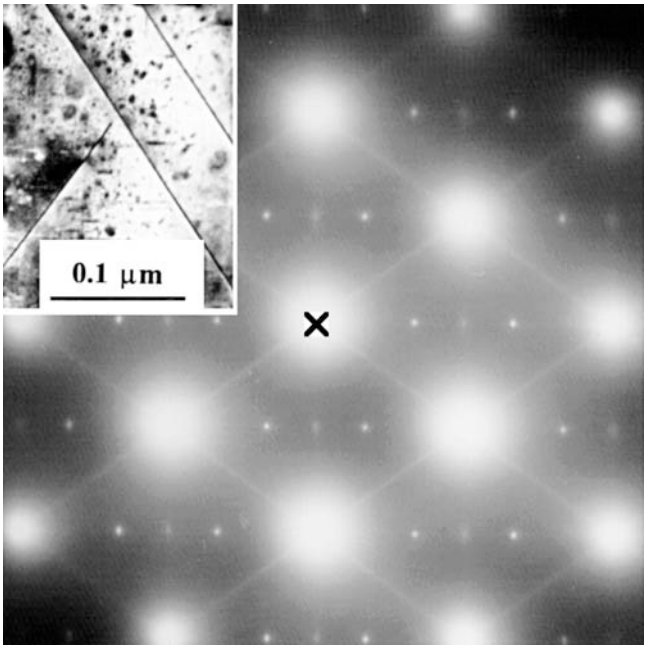


FIGURE 1.6. TEM DP from a thin foil of Al-Li-Cu containing various precipitate phases, shown in the inset image. The central spot (X) contains electrons that come directly through the foil and the other spots and lines are diffracted electrons which are scattered from the different crystal planes.

shows a TEM DP that contains information on the crystal structure, lattice repeat distance, and specimen shape (as well as being a most striking pattern). We'll see that the pattern can always be related to the image of the area of the specimen from which it came, in this case shown in the inset. You will also see in Part 2 that, in addition to the things we just listed, if you converge the usually parallel TEM beam to a focused probe, then you can produce even more striking convergent-beam patterns (see Figure 2.13D) from which you can conduct a complete crystal-symmetry analysis of minuscule crystals, including such esoteric aspects as point-group and space-group determination. You shouldn't be surprised by now if we tell you that aberration correction can produce even better DPs, which are both sharper (by reducing chromatic aberration) and come from smaller regions of the specimen (by reducing C_s). The crystal structure produces no diffraction information in a VLM because of the relatively large wavelength of visible light.

KEY POINT TO REMEMBER

At all times the crystallographic information in the DP (and all the analytical information) can be related to the image of your specimen.

So a TEM can produce atomic-resolution images, it can generate a variety of signals telling you about your specimen chemistry and crystallography, and you can always produce images that are in focus. There are many other good reasons why you should use electron microscopes. We hope they will become evident as you read through this book. At the same time there are many reasons why you should *not* always seek to solve your problems with the TEM, and it is most important that you realize what the instrument *cannot* do, as well as knowing its capabilities.

1.3 LIMITATIONS OF THE TEM

1.3.A Sampling

All the above advantages of the TEM bring accompanying drawbacks. First of all, the price to pay for any high-resolution imaging technique is that you only look at a small part of your specimen at any one time. The higher the resolution therefore, the worse the sampling abilities of the instrument. Von Heimendahl (1980) reported a calculation by Swann around 1970 estimating that all TEMs, since they first became available commercially (~15 years), had only examined 0.3 mm^3 of material! Extending that calculation to the present time probably increases this volume to no more than 10^3 mm^3. So we have an instrument that is not a good sampling tool! This sampling problem only serves to emphasize that, if you're just starting your research, before you put your specimen in the TEM you must have examined it with techniques that offer poorer resolution but better sampling, such as your eyes, the VLM, and the SEM. In other words, know the forest before you start looking at the veins in the leaves on the trees.

1.3.B Interpreting Transmission Images

Another problem is that the TEM presents us with 2D images of 3D specimens, viewed in transmission. Our eyes and brain routinely understand reflected light images but are ill-equipped to interpret TEM images and so we must be cautious. Hayes illustrates this problem well by showing a picture of two rhinoceros side by side such that the head of one appears attached to the rear of the other (see Figure 1.7). As Hayes puts it "when we see this image we laugh" (because we understand its true nature in 3D) "but when we see equivalent (but more misleading) images in the TEM, we publish!" So beware of artifacts which abound in TEM images.

One aspect of this particular drawback (sometimes called the projection-limitation) is that generally all the TEM information that we talk about in this book (images, DPs, spectra) is averaged through the thickness of the specimen. In other words, a single TEM image has no depth sensitivity. As we noted in Figure 1.5 there often is information about the top and bottom surfaces of the thin foil, but this is not immediately apparent. So

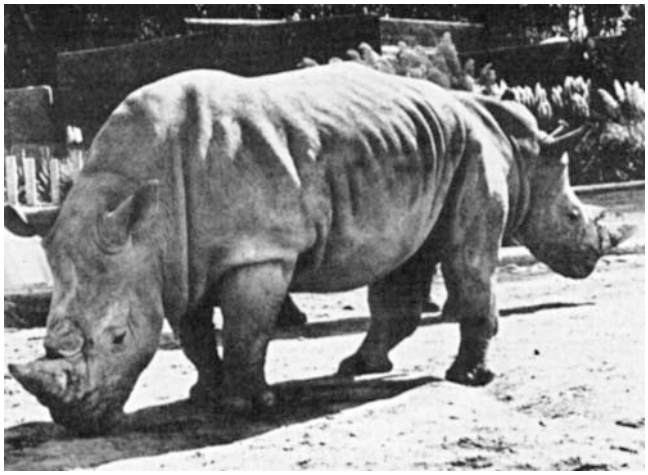


FIGURE 1.7. Photograph of two rhinos taken so that, in projection, they appear as one two-headed beast. Such projection artifacts in reflected-light images are easily discernible to the human eye but similar artifacts in TEM images are easily mistaken for 'real' features.

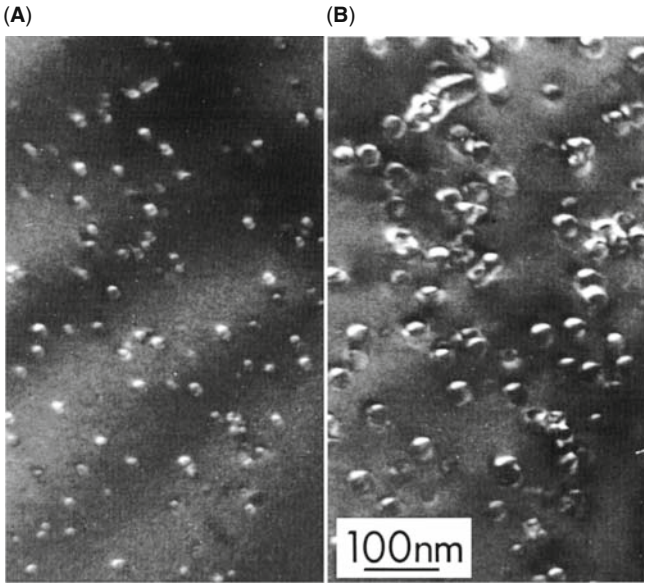


FIGURE 1.8. Beam damage (bright bubble-like regions) in quartz after bombardment with 125 keV electrons. With increasing time from (A) to (B) the damaged regions increase in size.

other techniques which are more clearly surface sensitive or depth sensitive, such as field-ion microscopy, scanning-probe microscopy, Auger spectroscopy, and Rutherford backscattering, are necessary complementary techniques if you want a full characterization of your specimen.

However, there has been progress in overcoming this limitation, which was much more of a problem for biologists interested in the shape of complex molecules, cells, and other natural structures. So they invented the technique of electron tomography, which uses a sequence of images taken at different tilts to create a 3D image, identical in principle to the more familiar medical CAT (computerized-axial tomography) scans using X-rays. Recently, there has been rapid improvement in specimen-holder design to permit full 360° rotation and, in combination with easy data storage and manipulation, nanotechnologists have begun to use this technique to look at complex 3D inorganic structures such as porous materials containing catalyst particles. This relatively new aspect of TEM for materials scientists is explored in depth in the companion text.

1.3.C Electron Beam Damage and Safety

A detrimental effect of ionizing radiation is that it can damage your specimen, particularly polymers (and most organics) or certain minerals and ceramics. Some aspects of beam damage are exacerbated at higher voltages, and with commercial instruments offering up to 400 kV, beam damage can now limit much of what we do in the TEM, even with refractory metals. The situation is even worse with more intense beams made possible because of advances in C_s correction. Figure 1.8 shows an area of a specimen damaged by high-energy electrons.

However, all is not lost and we can combine more intense electron sources with more sensitive electron detectors and use computer enhancement of noisy images to minimize the total dose received by the specimen to levels below the damage threshold. Minimum-dose microscopy techniques, often combined with specimen cooling (cryo-microscopy) and low-noise, charge-coupled device (CCD) cameras (see Chapters 7 and 31, respectively), are standard approaches in biological TEM and permit images to be obtained even when only a few hundred electrons/nm^2 are hitting the specimen. These approaches are finding increasing usage in TEM of materials where digital control of the beam in STEMs is another way to minimize radiation damage.

The combination of high kV beams with the intense electron sources that are available means that you can destroy almost any specimen, if you are not careful. At the same time comes the danger that should *never* be forgotten, that of exposing yourself to ionizing radiation. Modern TEMs are remarkably well engineered and designed with safety as a primary concern, but *never* forget that you are dealing with a potentially dangerous instrument that generates radiation levels that will kill tissue (and managed to damage some operators in the early days of the technique). So *never* modify your microscope in any way without consulting the manufacturer and without carrying out routine radiation-leak tests. If in doubt, don't do it!

1.3.D Specimen Preparation

Your specimens have to be thin if you're going to get any information using transmitted electrons in the TEM. 'Thin' is a relative term, but in this context it means electron transparent. For a specimen to be transparent to electrons, it must be thin enough to transmit sufficient electrons such that enough intensity falls on the screen, CCD, or photographic plate to give an interpretable image in a reasonable time. Generally this requirement is a function of the electron energy and the average atomic number (Z) of your specimen. Typically for 100-keV electrons, specimens of aluminum alloys up to ~1 μm would be thin, while steel would be thin up to about several hundred nanometers. However, it is an axiom in TEM that, almost invariably, *thinner is better* and specimens <100 nm should be used wherever possible. In extreme cases such as doing HRTEM or electron spectrometry, specimen thicknesses <50 nm (even <10 nm) are essential. These demands become less strict as the beam voltage increases, but this is offset by the production of beam damage.

Again these words of caution are balanced by the development of a specialized specimen-preparation tool called a focused ion beam (FIB) which is essential to the semiconductor-device fabricators who use them (by the dozen) to prepare, in a matter of a few tens of minutes, thin foils of specific, individual gates or junctions from one of the many millions of such on a 12-in. VLSI wafer. The only drawback is that, to buy a FIB, you have to pay as much as it costs to buy a TEM. We'll talk about this and other methods of specimen preparation in Chapter 10 and in the companion text.

THIN SPECIMENS

A *major* limitation of the TEM is we need thin specimens. Methods to prepare thin specimens exist for almost all materials, and we talk about them in Chapter 10. But as a general rule, the thinning processes that we use do affect the specimens, changing both their structure and chemistry. So you need to be aware of the drawbacks of specimen preparation and learn to recognize the artifacts introduced by standard preparation methods.

A terminological distinction is worth noting here. The words sample and specimen are often used interchangeably but, in this text, we'll distinguish the specimen specifically as the thin piece of material that you insert into the TEM in a specimen holder and we'll assume that the specimen was thinned from a much larger, bulk sample of the material you are interested in studying. Sometimes we'll mix the terms to test your understanding.

So it should be obvious to you by now that while TEM and associated techniques are tremendously powerful characterization tools when used properly, they should *never* be used in isolation to solve a materials problem. You must understand your material at low magnification with your eyes and with VLM and SEM before venturing into TEM studies. Otherwise you may fall-foul of some of the limitations we have just listed. But you should also have got the message that we are constantly improving the technique and overcoming these limitations while, at the same time, making the positive aspects even better.

1.4 DIFFERENT KINDS OF TEMs

As you read through the previous sections you will have seen that TEMs come in a wide variety of types: HRTEMs, HVEMs, IVEMs, STEMs, and AEMs. Complete books have been written on each of these instruments, but it is our philosophy that all these are simply different forms of the basic TEM. So in this book we intend to treat them as such. Indeed a current 200 or 300 keV TEM can combine aspects of *all* the above microscope types. Figure 1.9 shows several of the different kinds of TEMs we have mentioned. It is instructive to consider some of the features of the instruments shown here. An HVEM usually requires a two- or three-story room; the operator shows the scale of this instrument. A modern machine essentially is an electron-optical column in which we can maintain a good vacuum but the lenses and most other functions can be controlled by one or more computers. Note that the DSTEM only has a flat-panel display: there is no (internal) viewing screen. This aspect is becoming a more popular design feature in TEMs because, if the screen is not in the microscope, then the operator doesn't have to be in the room or in the building or even in the same country. Removing the operator from proximity to the TEM overcomes many of the factors that limit the performance of the best instruments. Remote operation (or telepresence) is an increasingly attractive feature, which will give many more researchers access to the most sophisticated TEMs, as already happens in the world of astronomy with major telescopes.

1.5 SOME FUNDAMENTAL PROPERTIES OF ELECTRONS

Many times in the book we'll have to refer to some of the basic properties of electrons. You know that electrons show both particle and wave characteristics, illustrating one of the great puzzles of quantum physics that we all seem to accept without too much trouble. In fact the TEM routinely demonstrates both the particle and wave characteristics of the electron, repeating the electron analog of GI Taylor's famous experiment in which he

FIGURE 1.9. (Continued)

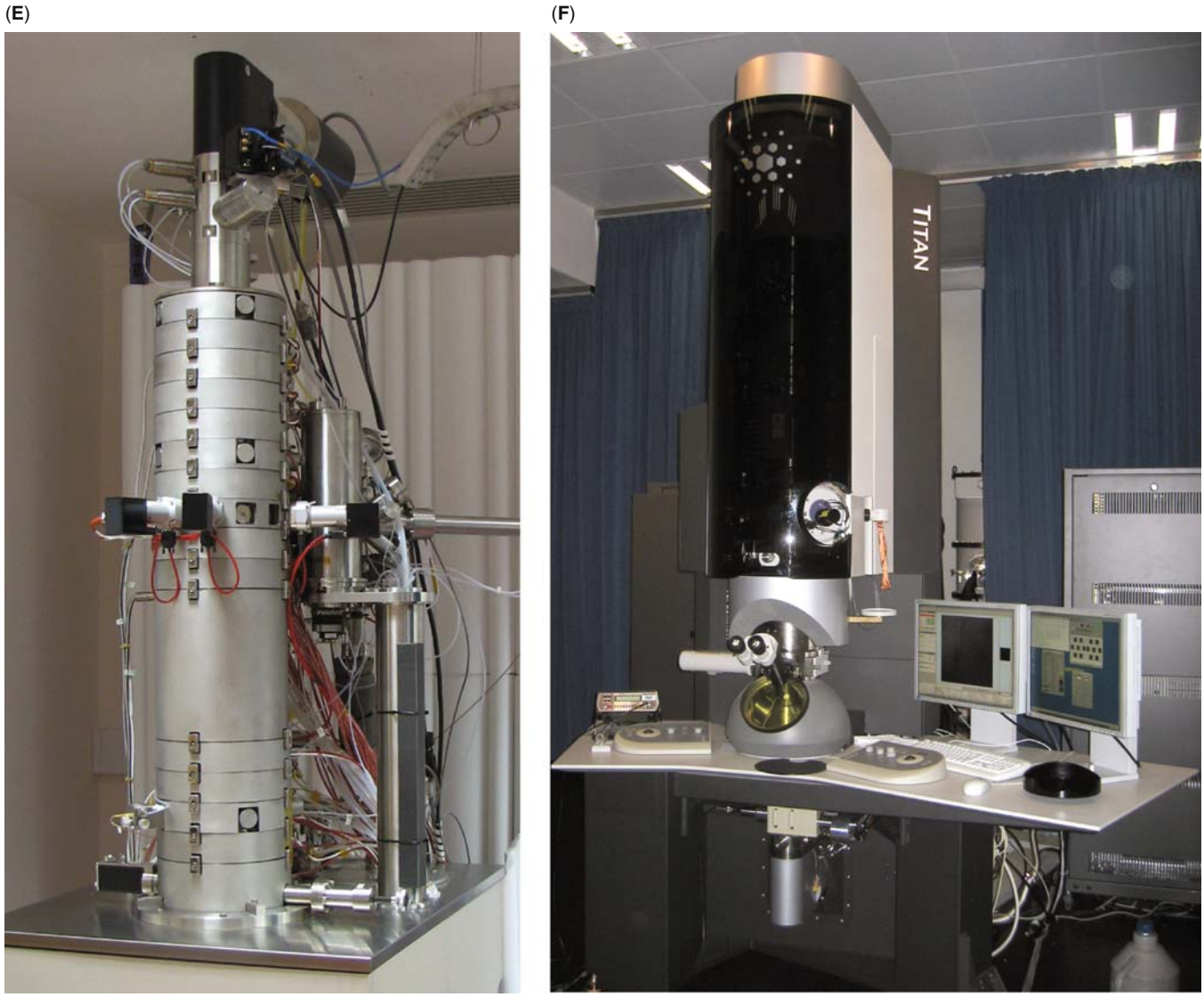


FIGURE 1.9. A selection of different commercial TEMs: (A) JEM 1.25 MeV HVEM. Note the size of the instrument; often the high-voltage tank is in another room above the column. (B) Zeiss HRTEM with a C_s corrector and an in-column energy filter. Note the large frame to provide high mechanical stability for the highest-resolution performance. (C) Hitachi 200 keV dedicated STEM; note the absence of a viewing chamber. Such instruments are often designed to aid failure analysis for the semiconductor device manufacturers. Specimens thinned from wafers on the production line can be easily transferred and examined. (D) JEOL 200 keV TEM/STEM; note also the absence of a viewing chamber. (E) Nion 200 keV ultrahigh vacuum SuperSTEM; the only US-manufactured (S)TEM and current holder of the world record image resolution (F) FEI Titan. Comparison with Ruska's instrument (Figure 1.1), which is 70–80 years older than these instruments, is instructive.

demonstrated Young's slits interference patterns despite using such a weak light source that only one photon passed through either slit at any one time. The electron beam current in a TEM can be as high as ~ 0.1–$1\ \mu A$, which corresponds to about 10^{12} electrons passing through the specimen plane. But, as we'll see below, with 100 keV energy, these electrons travel at about 0.5c (actually $\sim 1.6 \times 10^8$ m/s), so they are separated by ~ 1.6 mm and this means that there is never more than one electron in the specimen at any one time. Nevertheless, electron diffraction and interference occur, both of which are wave phenomena, and require interaction between different electron waves. Despite this dilemma, we know a lot about the electron and its behavior, and some of the basic characteristics are summarized in Table 1.1, along with some relevant physical constants.

There are a few important equations that you should know. First of all, based on de Broglie's ideas of the wave-particle duality, we can relate the particle momentum p to its wavelength λ through Planck's constant, thus

$$\lambda = \frac{h}{p} \tag{1.3}$$

TABLE 1.1 Fundamental Constants and Definitions

Charge (e)	(–) 1.602×10^{-19} C
1 eV	1.602×10^{-19} J
Rest mass (m_0)	9.109×10^{-31} kg
Rest energy (m_0c^2)	511 keV
Kinetic energy (charge × voltage)	1.602×10^{-19} N m (for 1 volt potential) = J
Planck's constant (h)	6.626×10^{-34} N m s
1 A	1 C/s
Speed of light in vacuum (c)	2.998×10^{8} m/s

In the TEM we impart momentum to the electron by accelerating it through a potential drop, V, giving it a kinetic energy eV. This potential energy must equal the kinetic energy, thus

$$eV = \frac{m_0 v^2}{2} \tag{1.4}$$

Now we can equate the momentum p to the electron mass, m_0, times the velocity, v, and substituting for v from equation 1.4

$$p = m_0 v = (2m_0 eV)^{1/2} \tag{1.5}$$

These three simple equations define the relationship between the electron wavelength λ and the accelerating voltage of the electron microscope, V

$$\lambda = \frac{h}{(2m_0 eV)^{1/2}} \tag{1.6}$$

If you look back, this equation is equivalent to equation 1.2. The inverse relationship between λ and V introduces a very important concept: by increasing the accelerating voltage we decrease the wavelength of the electrons. So you, the operator, can do this whenever you wish!

Equations 1.2 and 1.6 are useful expressions for deducing ballpark estimates, but be careful to note the differences. We can use equation 1.6 to calculate the non-relativistic electron wavelength for typical commercial TEM operating voltages as listed in Table 1.2.

The simple treatment we just went through neglects relativistic effects and, unfortunately for electron microscopists, relativistic effects cannot be ignored at energies $> \sim$100 keV because the velocity of the electrons (as particles) becomes greater than half the speed of light! So to be exact we must modify equation 1.6 to give

$$\lambda = \frac{h}{\left[2m_0 eV\left(1 + \frac{eV}{2m_0 c^2}\right)\right]^{1/2}} \tag{1.7}$$

A full listing for many more voltages can easily be generated by putting equations 1.6 and 1.7 into a spreadsheet. The effect of relativity is greater for higher accelerating voltages as shown in Table 1.2 which comprises all the commercial TEM accelerating voltages.

There will be many times when it's useful to refer back to these numbers, especially when we consider the resolution of the microscope and when we need to make calculations about the way electrons interact with matter.

A word about units: as we noted above, we should all be using SI units. We don't for two reasons: first, some special units are ideal for the purpose at hand; second we forget to include special conversion factors in some formulas. The difference between, e.g., the Gaussian system of units and SI units is summarized in the invaluable reference by Fischbeck and Fischbeck (1987) or in the electronic version of the almost 100-year-old standard source, Kaye and Laby (1986) (URL #2), or on the NIST database from which you can quickly find any number or constant that you need (URL #3).

TABLE 1.2 Electron Properties as a Function of Accelerating Voltage

Accelerating voltage (kV)	Non-relativistic wavelength (nm)	Relativistic wavelength (nm)	Mass ($\times m_0$)	Velocity ($\times 10^8$ m/s)
100	0.00386	0.00370	1.196	1.644
120	0.00352	0.00335	1.235	1.759
200	0.00273	0.00251	1.391	2.086
300	0.00223	0.00197	1.587	2.330
400	0.00193	0.00164	1.783	2.484
1000	0.00122	0.00087	2.957	2.823

1.6 MICROSCOPY ON THE INTERNET/ WORLD WIDE WEB

TEM users are well integrated into the Internet and the World Wide Web (WWW) and this is a source of useful information (and also some useful knowledge!) about what's going on in the field. You can view research TEMs in real time over the Internet and, as we've seen, you may not only see other instruments but you may be able to operate them remotely (a quick search on telemicroscopy or telepresence microscopy will get you some useful hits). So rather than spend a week in sunny Storrs operating an advanced TEM you can (regretfully) do it from the comfort of your own office or lab (or even from the beach in Aruba if there's a good broadband connection). As Internet2 and the National Lambda Rail further penetrate research labs in the United States and equivalent systems spread throughout the world, the rapid exchange of experimental information and the parallel simulation of data via access to high-end computational resources will continue to expand the options available to microscopists and analysts.

In addition, specialized software packages are also available on the Web which allow you to carry out many of the advanced analyses that we will introduce in this text (e.g., DP analysis and image/diffraction/spectral simulation). In many cases access to this software is limited (i.e., you have to pay for it) but any serious microscopy operator should have access to such software on site. Sometimes it is useful to explore the possibilities before you purchase.

We have already referenced a few important URLs in this chapter and we will continue to do so throughout the book and give lists at the end of each chapter. We have tried to keep the referenced URLs to sites associated with longstanding, reputable organizations such as national labs, professional societies, and major publishers. A list of broadly useful sites is included below and, not surprisingly, much of this list is very different from the one that we gave more than a decade ago when the first edition of this text was published. Nothing much has changed in this respect and so we cannot guarantee that these same URLs will still be active when you get round to searching for them. Such is life in the age of the Web.

1.6.A Microscopy and Analysis-Related Web Sites

http://www.amc.anl.gov This is the best source for TEM information on the Web in the United States and it is run by NJ Zaluzec at Argonne National Laboratory (ANL). Through it you can get access to the Microscopy ListServer and a Software Library. There is a connection to the Microscopy & Microanalysis FTP Site and to access Software/Image Libraries. Other useful connections through this site include:

http://microscopy.com/MicroscopyListserver/ This is an email-based discussion forum giving members of the EM community a centralized Internet address to which questions/comments/answers in the various fields of microscopy or analysis can be rapidly distributed to a list of (subscribed) individuals by electronic mail.

http://www.microscopy.com/MMMeetingCalendar.pl Listing of currently planned meetings, symposia, and courses.

http://zaluzec.com/cgi-bin/ANLWWWListingSQL.pl?SearchOrg=society This is a listing of the Web sites of various national and international microscopy and analysis societies. There are also connections to university, government, and individual microscopy-related sites.

http://cimewww.epfl.ch/EMYP/emyp.html This is the home of the Electron Microscopy Yellow Pages which is a similar operation to the ANL site, but based at the Ecole Polytechnique Fédérale de Lausanne in Switzerland, run by P Stadelmann. The Yellow Pages contain electron microscopy laboratories, software for electron microscopy, learned societies, instruments, equipment and consulting education in electron microscopy, data and databases, news and publications-related sources of information, conferences, workshops and schools, and getting somewhere else on the Web.

http://cimesg1.epfl.ch/CIOL/ and **http://cimesg1.epfl.ch/CIOL/summary.html** Stadelmann also offers access to very sophisticated EMS software for high-resolution image analysis and diffraction simulation and much more detail about this is covered in Chapter 30 and in the companion text.

http://iucr.org/resources/commissions A great resource operated for the International Union of Crystallography by JCH Spence at Arizona State University giving listings of electron diffraction and microscopy-related software, etc.

http://www.numis.northwestern.edu/IUCR_CED/ L Marks resource site for electron microscopy and diffraction

http://tem.msae.wisc.edu/emdb/index.html NSF supported TEM data base maintained by P. Voyles.

1.6.B Microscopy and Analysis Software

There is a lot of software available on the WWW and this is an aspect of TEM which is changing on a rapid basis, but you can now buy excellent software packages for all the fundamental aspects of microscopy: diffraction, imaging, and analysis. Many of these programs will be referenced throughout the text, but here is a brief summary of

the best that are currently used (with an indication of the source of the software) some of which are still free! There are many more packages than we have listed here but these are the ones with which we are familiar.

- **Cross sections for electron scattering:** The NIST version 3.0 of this database provides values of differential elastic-scattering cross sections, total elastic-scattering cross sections, phase shifts, and transport cross sections for elements with atomic numbers from 1 to 96 and for electron energies between 50 eV and 300 keV (in steps of 1 eV). Free downloads at http://www.nist.gov/srd/nist64.htm.
- **CRISP**: It is a commercial package running under Windows on a PC. It is designed for imaging process of HRTEM images. It can be combined with ELD (see below) and is available from Calidris, Manhemsvägen 4, S-191 46 Solltuna, Sweden (46 8 625 00 41). www.calidris-em.com/crisp.htm—the site does not change very often.
- **DigitalMicrograph™ (DM)**: It is a complete system for the acquisition, control, and processing of digital images from any electron microscope and is the central software for the Gatan Microscopy Suite™ (GMS) (see below). Gatan is to TEM as Microsoft is to PC.
- **DTSA** (Desk-Top Spectrum Analyzer): The NIST/NIH Desktop Spectrum Analyzer generates, interprets, and analyzes X-ray spectra from specimens under electron bombardment. This remarkable software/database package simulates the experimental environment and emulates specimen properties to generate spectra reflecting the relevant physics, chemistry, and statistics of a real-world application. DTSA incorporates many widely accepted X-ray data analysis procedures. Technical Contact: johnhenry.scott@nist, (301) 975-4981; http://www.cstl.nist.gov/div837/Division/outputs/DTSA/DTSA.htm.
  Really essential for the X-ray analyst but currently only available from NIST for Macintosh users. A PC version has been developed by Masashi Watanabe at Lehigh University.
- **ELD**: It is a commercial package from the producers of CRISP running under Windows on a PC. It is intended for quantitative analysis of DPs and is available from Calidris, Manhemsvägen 4, S-191 46 Solltuna, Sweden (46 8 625 00 41). www.calidris-em.com/eld.htm—still not changing often.
- **ELP**: The original Gatan energy-loss acquisition, processing, and analysis program now incorporated as part of the Gatan Microscopy Suite™ (GMS) (see below).
- **EMS** and **jEMS**: Image simulation program; see the listing of its capabilities in Section 1.6.A.
- **Gatan Microscopy Suite™ (GMS)**: It permits data acquisition, processing, and analysis for a wide range of TEM applications including EELS analysis, and energy-filtered compositional mapping, DP analysis, and 3D tomography acquisition. Open to individual scripting for user-specific needs. Widely used by many microscopists and analysts for acquiring, analyzing, and processing any kind of image data. Use with a CCD camera that provides digital images from the TEM or interface to any STEM system. From Gatan Inc., 5933 Coronado Lane, Pleasanton, CA 94588, (925) 463-0200; info@gatan.com.
- **ImageJ**: It is a versatile and extremely powerful image-processing and analysis open-source freeware available from NIH (has largely replaced the popular NIH Image). Developed by Wayne Rasband and aided by input from more than 1400 users round the world, ImageJ runs on Linux, Mac OS 9, Mac OS X, Windows. It is the world's fastest pure Java image-processing program. It can filter a 2048 × 2048 image in 0.1 s (40 million pixels per second!). Open and save all supported data types as TIFF (uncompressed) or as raw data. Open and save GIF, JPEG, BMP, PNG, PGM, FITS, and ASCII. Open DICOM. Open TIFFs, GIFs, JPEGs, DICOMs, and raw data using a URL. Open and save many other formats using plugins. Supports smoothing, sharpening, edge detection, median filtering, and thresholding on both 8-bit grayscale and RGB color images. Measure area, mean, standard deviation, min, and max of selection or entire image. Measure lengths and angles. Use real-world measurement units such as millimeters. Calibrate using density standards. Generate histograms and profile plots and much more.
  Available from http://rsb.info.nih.gov/ij/.
  Also look for Image SXM, a version of NIH Image extended to handle the loading, display, and analysis of scanning images: http://www.liv.ac.uk/~sdb/ImageSXM/.
- **Maclispix**: A Macintosh-based image-processing program, which works in conjunction with NIH Image (see below) or ImageJ (see above); permits analysis of
  Stacks: Movies, Depth profiles, Cropping and saving of large data sets.
  Groups: Coordinated measurements, Color overlays, Scatter diagrams.
  Pixels: bit, byte, integer, RGB, real, complex.
  Statistical measurements. Signal/noise determinations.
  Also diffraction analysis, segmentation (blobbing and measurement), registration, principal component analysis.
  Developed by David Bright at NIST (micro@nist.gov) and can be downloaded (free) from http://www.nist.gov/lispix/MLxDoc/mlx.html.
- **MacTempas and CrystalKit**: MacTempas is a Macintosh-based image analysis program for the simulation of high-resolution images, DPs, and crystal structures. Features include

Full multislice calculation of HRTEM Images and dynamical DPs.
Automatic calculation of the correct unit cell for any given beam orientation.
Automatic selection of correct array size and shape for the multislice calculation.
Display of atomistic models from any direction in both color and shades of gray.
Display of projected potentials, Fourier coefficients, diffractogram of images.
Plotting of amplitudes/phases of diffracted beams vs. thickness, contrast transfer functions.
Database on the 230 space groups; calculates the associated symmetry operators and atomic coordinates.
http://www.totalresolution.com/MacTempas.html.

CrystalKit: It works within MacTempas or EMS and builds models of crystalline defects, from point defects to grain boundaries and precipitates. A geometric grain boundary involving several thousand atoms can be generated in a matter of minutes by specifying the orientation relation between the grains, the interface plane, and the zone axis. CrystalKit allows the user to freely rotate the crystal, identify planes, measure angles and distances between atoms, visually move atoms, delete atoms, and add new atoms or create an arbitrary path interface.
http://www.totalresolution.com/CrystalKit.html.

- **Monte Carlo Simulations**: Software to simulate electron-beam trajectories through materials for estimating the spatial resolution of X-ray analysis or the backscattered electron yield. The best source of information is NIST's Web site http://www.cstl.nist.gov/div837/837.02/epq/index.html.
  For thin specimens you should use David Joy's Monte Carlo program described in his book (Joy 1995) and available at the University of Tennessee Web site http://web.utk.edu/~srcutk/htm/simulati.htm.
- **NIH-Image** (and **ImageJ**): It is a public domain software from NIH, developed by Wayne Rasband for general image manipulation with a limited set of image-processing tools. Details can be found at http://rsb.info.nih.gov/nih-image/about.html but it has been largely replaced by ImageJ (see above).
- **Adobe Photoshop**: Professional page layout programs for presentations and labeling your figures. Still the standard and available for a nominal fee at http://www.adobe.com/products/photoshop/.
- **Diffraction-Pattern Indexing**: Start with http://emaps.mrl.uiuc.edu JM Zuo's excellent free site at the University of Illinois. After this you should check the EM Yellow Pages and the EMS software and also the Web site for the International Union of Crystallography (both already referenced in Section 1.6.A). Also try SingleCrystal™; part of the extensive CrystalMaker® suite of DP and crystal structure software. Free demo download at: http://www.crystalmaker.co.uk/singlecrystal/index.html.

CHAPTER SUMMARY

TEMs comprise a range of different instruments that make use of the properties of electrons, both as particles and as waves. The TEM generates a tremendous range of signals so we can obtain images, DPs, and several different kinds of spectra from the same small region of the specimen. In the rest of this book we'll take you through the fundamental aspects of TEM, trying to explain at all times *why* we do certain things in certain ways. We'll also explain to some degree *how* we carry out certain operations. Since many different commercial TEMs exist, there's no point in being specific in how to operate a particular TEM, but we can explain in a generic sense, in many cases, what you have to do to get your particular one to deliver the enormous amounts of information that it can generate. Not least of course, we also describe what you need to know to *interpret* the images, DPs, and spectra that you obtain.

If you count up the different imaging, diffraction, and spectroscopic operations that are available in a TEM there are almost 40 different modes of forming an image, DP, or spectrum, each of which produces different information about your specimen. (Your last homework, when you've finished the book, is to validate this claim!) Each of these information planes can be understood in a *quantitative* manner, to the extent that we can simulate all TEM images, DPs, and spectra in a computer. *No* other characterization technique comes close to the combination of versatility and quantification that is produced by this remarkable instrument, particularly when you consider the enormous range of magnifications over which the information is obtainable.

There is a wealth of other sources of information about TEM and in the general reference list below we give a selection of appropriate books that emphasize materials science and

nanotechnology (most of which remain in print) as well as some standard journals and regular conference proceedings. We also encourage you to get on the WWW and see what's out there, but be careful about the validity of the content on any particular site. It is probably reasonable to trust the Web sites from reputable institutions such as government labs, commercial TEM and related equipment manufacturers organizations where, once you get past any sales pitch, there's often great educational material, and of course the Web pages of professional electron microscopists such as that associated with this text (URL #4).

REFERENCES PHILOSOPHY

In the reference sections throughout the book, we will annotate the references but we won't necessarily spell them out in the text (we'll give clues); we don't want the chapters to look like journal papers themselves and we want to encourage you to look at these reference sections. (This chapter is the exception.) In addition, don't forget the references for the figures right at the end of the textbook where you'll find much more than is in the figure captions. We do encourage you to explore the literature—you really have no choice if you want to understand TEM.

GENERAL TEM BOOKS

Amelinckx, S, van Dyck, D, van Landuyt, J and van Tendeloo, G (Eds.) 1997 *Electron Microscopy: Principles and Fundamentals* VCH Weinheim Germany. An expensive text containing review articles by leading microscopists covering TEM and much more. A good overview to put TEM in context, but make sure your library buys this along with the even more expensive *Handbook of Microscopy: Applications in Materials Science, Solid State Physics and Chemistry* edited by the same authors and published by VCH in 1997.

DeGraef, M 2003 *Introduction to Conventional Transmission Microscopy* Cambridge University Press New York. A complementary text to this one, in many ways. It uses different materials specimens as a thread to introduce different techniques. Supported by an excellent Web site (http://ctem.web.cmu.edu/) but not really an introduction!

Edington, JW 1976 *Practical Electron Microscopy in Materials Science* Van Nostrand-Reinhold New York. The original out-of-print 1976 edition has been reprinted by TechBooks, 2600 Seskey Glen Court, Herndon, VA 22071. A very helpful, if outdated, text full of examples and hands-on operations; no AEM or HREM, just diffraction-based imaging.

Egerton, RF 2006 *Physical Principles of Electron Microscopy; An Introduction to TEM, SEM, and AEM* Springer New York. If you need a general introduction to EM, this is a good choice.

Ernst, F and Rühle, M (Eds.) 2003 *High-Resolution Imaging and Spectrometry of Materials* Springer Series in Materials Science 50 Springer Berlin. A collection of review articles covering some aspects of TEM and other high-resolution techniques.

Fultz, B and Howe, JM 2002 *Transmission Electron Microscopy and Diffractometry of Materials* 2nd Ed. Springer New York. A broad-based text emphasizing diffraction-based imaging and crystallography via studies with X-ray and electrons.

Goodhew, PJ, Humphreys, FJ and Beanland, R 2001 *Electron Microscopy and Analysis* 3rd Ed. Taylor & Francis New York. A succinct summary of SEM, TEM, and AEM.

Hall, CE 1953 *Introduction to Electron Microscopy* McGraw-Hill New York. A wonderful but nowadays neglected book. The level is very close to this text. Historically minded students will enjoy the preface.

Hawkes, PW and Spence, JCH (Eds.) 2007 *Science of Microscopy* Springer New York. Comprehensive, up-to-date, multi-author review of many forms of microscopy. Get your library to buy it.

Heidenreich, RD 1964 *Fundamentals of Transmission Electron Microscopy* Interscience Publisher New York NY. Another wonderful, but sometimes forgotten, classic.

Hirsch, PB, Howie, A, Nicholson, RB, Pashley, DW and Whelan, MJ 1977 *Electron Microscopy of Thin Crystals* 2nd Ed. Krieger Huntington NY. For many years, the 'Bible' for TEM users and still required reading for true TEM converts!

Marton, L 1968 *Early History of the Electron Microscope* San Francisco Press San Francisco.

McLaren, AC 1991 *Transmission Electron Microscopy of Minerals and Rocks* Cambridge University Press New York. Invaluable for the geologist or ceramist.

Reimer, L 1997 *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis* 4th Ed. Springer New York. Essential reference text. Strong physics background required; never uses a few words where a triple integral will do.

Ruska, E 1980 *The Early Development of Electron Lenses and Electron Microscopy* (translated by T Mulvey) S Hirzel Verlag Stuttgart.

Sawyer, LC, Grubb, DT and Meyers, GF 2008 *Polymer Microscopy* 3rd Ed. Springer New York. An expensive but useful qualitative introduction to TEM and SEM of polymers.
Thomas, G and Goringe, MJ 1979 *Transmission Electron Microscopy of Metals* Wiley New York. Invaluable for classical imaging and diffraction topics. The original out-of-print 1979 edition has been reprinted by TechBooks, 2600 Seskey Glen Court, Herndon, VA 22071.
Watt, IM 1997 *The Principles and Practice of Electron Microscopy* 2nd Ed. CUP New York NY. A basic, practical introduction to SEM and TEM.
Wenk, H-R 1976 *Electron Microscopy in Mineralogy* Springer New York NY. Required reading for microscopy of geological or ceramic materials. From the library.
Yao, N and Wang, ZL 2005 *Microscopy for Nanotechnology* Kluwer New York. In-depth review articles on techniques for nano-characterization. Half of the 22 chapters are devoted to EM methods and more than 80% of those are TEM.

SPECIALIZED TEM BOOKS

Ahn, CC (Ed.) 2004 *Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS Atlas* 2nd Ed. Wiley-VCH Berlin. An excellent, in-depth review of EELS and the best database for ionization-edge energies.
Brydson, R 2001 *Electron Energy Loss Spectroscopy* Bios (Royal Microsc. Soc.) Oxford UK. You must read this before going on to Egerton's classic.
Champness, PE 2001 *Electron Diffraction in the TEM* Bios (Royal Microsc. Soc.) Oxford UK. Outstanding, essential introductory text by a crystallographer who also knows TEM.
Cowley, JM (Ed.) 1992 *Electron Diffraction Techniques* Vols. 1 and 2 Oxford University Press New York. Another collection of excellent individual review articles.
Egerton, RF 1996 *Electron Energy Loss Spectroscopy in the Electron Microscope* 2nd Ed. Plenum Press New York. The quintessential text for this aspect of AEM.
Frank, J 1992 *Electron Tomography* Plenum Press New York. A biological EM text but the basis for much of what's going to happen in TEM of materials.
Garratt-Reed, AJ and Bell, DC 2002 *Energy-Dispersive X-ray Analysis in the Electron Microscope* Bios (Royal Microsc. Soc.) Oxford, UK. Basic introduction pitched at a similar level to this text.
Hawkes, PW and Kasper, E 1989, 1994 *Principles of Electron Optics* Vols. 1–3 Academic Press New York, 1900 pp. Comprehensive but advanced. The third volume deals with many aspects of imaging in the TEM, simulation and processing with ~118 pages of TEM references; an exceptional resource.
Head, AK, Humble, P, Clarebrough, LM, Morton, AJ and Forwood, CT 1973 *Computed Electron Micrographs and Defect Identification* North-Holland New York NY. Long out of print but often referenced.
Horiuchi, S 1994 *Fundamentals of High-Resolution Transmission Electron Microscopy* North-Holland Amsterdam. As it says; a one topic book.
Kirkland, EJ 1998 *Advanced Computing in Electron Microscopy* Plenum Press New York. Required reading for the mathematically inclined microscopist.
Jones, IP 1992 *Chemical Microanalysis Using Electron Beams* Institute of Materials London. Great introduction to AEM with lots of calculations to introduce the principles of quantitative analysis.
Loretto, MH 1994 *Electron Beam Analysis of Materials* 2nd Ed. Chapman and Hall New York. A concise overview of the subject.
Royal Microscopical Society Handbook series: a broad range of introductory texts covering many aspects of TEM but also SEM and visible-light microscopy. Easy reading and not expensive (URL #5).
Shindo, D and Oikawa, T 2002 *Analytical Electron Microscopy for Materials Science* Springer New York. Brief summary of X-ray and electron spectrometry.
Spence, JCH 2003 *High Resolution Electron Microscopy* 3rd Ed. Oxford University Press New York. Practical HREM combined with lots of sound theory.
Spence, JCH and Zuo, JM 1992 *Electron Microdiffraction* Plenum Press New York. Quantitative convergent-beam diffraction in great depth.
Tonomura, A 1999 *Electron Holography* Springer New York. A good introduction.
von Heimendahl, M 1980 *Electron Microscopy of Materials* Academic Press New York NY. An introductory-level text, no AEM or HRTEM component. From the library.
Wang, ZL 1995 *Elastic and Inelastic Scattering in Electron Diffraction and Imaging* Plenum Press New York. Everything you ever need to know about scattering of electrons in the TEM.

THE COMPANION TEXT

Throughout this chapter and the rest of the text we will refer to the companion text. This new textbook is not a required reading for all students but does contain chapters on special topics that are only covered briefly in the present text.

JOURNALS

Advances in Imaging and Electron Physics Ed. PW Hawkes. Peter Hawkes is an editor, author, and historian without peer in the field of electron microscopy. Any text he edits or book he writes is worth reading. This journal merges two long-running serial journals *Advances in Electronics and Electron Physics* and *Advances in Optical and Electron Microscopy*. Reviews on particle optics at high and low energies, image science and digital image processing, electromagnetic-wave propagation, electron microscopy, and computing methods associated with all these topics. http//www.elsevier.com/wps/find/bookdescription.cws_home/711044/description#description.

Journal of Electron Microscopy Official Journal of the Japanese Society for Microscopy, becoming a more widely appreciated resource now that the papers are all in English. Oxford University Press Oxford. http//jmicro.oxfordjournals.org/archive/.

Journal of Microscopy Official Journal of the Royal Microscopical Society, the International Society for Stereology, the Microscopical Society of Ireland, the Polish Society for Microscopy, and the Austrian Society for Electron Microscopy. Many groundbreaking papers in the TEM of materials have appeared in this journal despite the strong biological theme of many of the papers. Blackwell Publishing Ltd Oxford, UK. http//www.rms.org.uk/journal.shtml.

Micron The International Research and Review Journal for Microscopy. Elsevier Amsterdam The Netherlands. http//www.elsevier.com/wps/find/journaldescription.cws_home/475/description#description.

Microscopy and Microanalysis Official Journal of the Microscopy Society of America, Microbeam Analysis Society (USA), Microscopical Society of Canada/Société de Microscopie du Canada, Mexican Microscopy Society, Brazilian Society for Microscopy and Microanalysis, Venezuelan Society for Electron Microscopy, European Microbeam Analysis Society, Australian Microscopy and Microanalysis Society. Has the largest circulation of any EM journal. Cambridge University Press New York NY. http//www.msa.microscopy.org/MSAUnits/Journal/MscopyManalysis.html.

Microscopy, Microanalysis, Microstructure Official Journal of the Société Française des Microscopies. Publication continued by The European Physical Journal (Applied Physics) in 1998 since when it has lost much visibility, but there are key papers in here before 1998.

Microscopy Research and Technique A place to publish your new techniques and methods for microscopy, specimen preparation, or any related aspect of TEM. John Wiley & Sons Hoboken NJ. http//www3.interscience.wiley.com/cgi-bin/jhome/38527.

Ultramicroscopy An international journal affiliated with multiple national societies, committed to the advancement of new methods, tools, and theories in microscopy, where much cutting-edge TEM research is published. Look out for Peter Hawkes' occasional reviews of the state of the microscopy literature. Elsevier Amsterdam The Netherlands http//www.elsevier.com/wps/find/journaldescription.cws_home/505679/description#description.

SELECTED CONFERENCE PROCEEDINGS

Asia-Pacific Electron Microscopy Conference organized by the Committee of Asia-Pacific Societies for Electron Microscopy (CAPSEM) every 4 years (2012).

Australian Microscopy and Microanalysis Society every 2 years (2010).

Electron Microscopy and Analysis Group (EMAG), Institute of Physics, London, UK every 2 years (2011).

European Microbeam Analysis Society every 2 years (2011).

European Microscopy Congress organized by the European Microscopy Society, every 4 years (2012).

Inter-American Congress for Electron Microscopy organized by Comité Interamericano De Sociedades De Microscopía Electrónica (CIASEM), every 2 years (2011).

International Congress for Microscopy organized by the International Federation of Societies for Microscopy (IFSM), every 4 years (2010 in Rio de Janeiro). The world gathers here.

International Union of Microbeam Analysis Societies every 4 years (2012).

Japanese Society for Microscopy annually.

Microscopy & Microanalysis combined meeting of the Microscopy Society of America, the Microbeam Analysis Society and occasionally others (including the Canadians); annually.

USEFUL SOURCES OF NUMERICAL DATA AND CONSTANTS

Fischbeck, HJ and Fischbeck, KH 1987 *Formulas, Facts and Constants* 2nd Ed. Springer New York. An invaluable reference. SI units are described in Chapter 2. Relevant equations in Gaussian units are related to SI units on page 127.

Jackson, AG 1991 *Handbook for Crystallography for Electron Microscopists and Others* Springer New York. Ideal for the microscopist but see the review by JA Eades Microsc. Res. Tech. **21** 368.

Kaye, GWC and Laby, TH 1986 *Tables of Physical and Chemical Constants* 15th Ed. This invaluable resource was first published in 1911 and is now online courtesy of the National Physical Laboratory (UK) (URL #2).

SPECIFIC REFERENCES IN CHAPTER 1

Usually we will group these into topics but this time the topic is 'The Introduction.'

CBE (Council of Biology Editors) 1994 *Scientific Style and Format* 6th Ed. Cambridge University Press New York.

Cosslett, VE 1979 *The Cosslett Festschrift* J. Microsc. **117** 1–184.

Davisson, CJ and Germer, LH 1927 *Diffraction of Electrons by a Crystal of Nickel* Phys. Rev. **30** 705–740. Early work by Clinton J. Germer and Lester H. German; read the introduction for encouragement.

de Broglie, L 1925 *Recherches sur la Theorie des Quanta (Researches on the Quantum Theory)* Ann. Phys. **3** 22–128.

Ewald, PP 1962 *Fifty Years of X-ray Diffraction* International Union of Crystallography D. Reidel Dordrecht.

Fujita, H 1986 *History of Electron Microscopes* Business Center for Academic Societies Japan.

Goodman, P 1981 *Fifty Years of Electron Diffraction* International Union of Crystallography Utrecht.

Haguenau, F, Hawkes, PW, Hutchison JL, Satiat-Jeunemaitre, B, Simon, G and Williams, DB 2003 *Key Events in the History of Electron Microscopy* Microsc. Microanal. **9** 96–138.

Hashimoto, H 1986 *High Resolution and High Voltage Electron Microscopy* J. Elec. Microsc. Tech. **3** 1.

Hawkes, PW (Ed.) 1985 *The Beginnings of Electron Microscopy, Advances in Electronics and Electron Physics* Academic Press New York NY.

Hayes, TL 1980 *Biophysical Aspects of Scanning Electron Microscopy* SEM-1980 **1** 1–10 Ed. O Johari SEM Inc. AMF O'Hare IL.

Howie, A 2000 *A Symposium in Honor of Professor Archie Howie's 65th Birthday* Eds. PL Gai, ED Boyes, CB Carter, DJH Cockayne, LD Marks and SJ Pennycook. Microsc. Microanal. **6** 281–284.

Joy, DC 1995 *Monte Carlo Modeling for Electron Microscopy and Microanalysis* Oxford University Press New York.

Knoll, M and Ruska, E 1932 *Das Elektronenmikroskop (Electron Microscope)* Z. Phys. **78** 318–339.

Kossel, W and Möllenstedt, G 1939 *Electroneninterferenzen im Konvergenten* Ann. Phys. **36** 113–140.

Thomas, G 2002 *A Symposium in Honor of Gareth Thomas' 70th Birthday* Eds. DG Howitt, CB Carter, U Dahmen, R Gronsky, DB Williams and R Sinclair Microsc. Microanal. **8** 237–364.

Thomson, GP 1928 *Experiments on the Diffraction of Cathode Rays* Proc. Roy. Soc. Lond. **117** 600–609. George Paget Thomson was the son of J. J. Thomson; he shared the 1937 Nobel Prize for physics with Davisson. A good read.

Zeitler, E 2003 *Zeitler Festschrift* Eds. GA Botton, K Moore and D Su Micron **34** 119–260.

SPECIFIC URLs

1. http//www.nano.gov/html/facts/whatIsNano.html
2. http//www.kayelaby.npl.co.uk/
3. http//www.physics.nist.gov/cuu/Constants/archive1998.html
4. http//www.TEMbook.com
5. http//www.rms.org.uk/other-publications.shtml

SELF-ASSESSMENT QUESTIONS

Q1.1 Define ionizing radiation and explain why it's useful.
Q1.2 List four signals generated by electrons interacting with the atoms in your specimen.
Q1.3 What makes electrons interact strongly with matter?
Q1.4 What is the resolution or the resolving power of a microscope?
Q1.5 What limits the resolution of VLMs?
Q1.6 What limits the resolution of TEMs?
Q1.7 How can we get high resolution in a TEM?
Q1.8 Define depth of field for a specimen in the TEM.
Q1.9 Explain why sampling is a problem with TEMs and how to combat this limitation.

Q1.10 List three benefits of using a TEM rather than any other instrument to characterize nanoparticles.
Q1.11 What is the major difficulty with interpreting all TEM images?
Q1.12 Why is TEM so powerful a characterization tool?
Q1.13 What signals are most commonly used in the conventional TEM?
Q1.14 Define the acronyms TEM, AEM, STEM, HREM, HVEM. Are there any other kinds of electron microscopes?
Q1.15 Which two signals produced when electrons interact with matter are used in AEM?
Q1.16 What happens to the wavelength of the electron beam when the accelerating voltage of the TEM is increased?
Q1.17 What does 'thin' mean when referring to a TEM specimen?
Q1.18 Give a ballpark figure for a suitably 'thin' TEM specimen.
Q1.19 Why is TEM such an important experimental tool for materials scientists and nanotechnologists?
Q1.20 What is a typical electron beam current in a TEM?
Q1.21 What are the typical causes of specimen damage in the TEM?
Q1.22 What effect should we consider in any calculations when the energy of the electrons is 100 keV or higher and why?

TEXT-SPECIFIC QUESTIONS

T1.1 If you were a physicist in the 1920s why would you think that the fact that electrons can be diffracted by crystals is a good reason to develop an electron microscope?
T1.2 Calculate an approximate wavelength for 50 kV electrons via equations 1.1 and 1.2 and then estimate the Rayleigh-defined resolving power of a microscope using such electrons. We tell you to assume $\mu \sin \beta \sim \beta$ so what does this tell you about what numbers might be reasonable for the refractive index and numerical aperture in a TEM?
T1.3 Examine Figure 1.2A; draw a schematic diagram to summarize all that you can deduce about the crystallography of the specimen with the information given in the text.
T1.4 Look at the arrows in Figure 1.3. From what little you know at this stage about how electrons interact with matter, can you work out which signals are most likely to be used in a TEM and which are less likely to be used? (Hint: relate the diagram to the actual construction of the TEM.)
T1.5 Examine Figure 1.4; as a materials scientist, do you expect there to be more of the Ti (blue) alloying element present in the smaller precipitates in the matrix, the larger precipitates on the boundary or the large particle on the boundary, the big dark lumps up the middle of the image or in the matrix? Are there reasons to be concerned about trying to estimate compositions from the color in the image only?
T1.6 In Figure 1.5 some dislocations appear to begin or end in the material but you will recall that dislocation lines cannot begin or end in a crystal. (Refer, e.g., to Hull, D and Bacon, DJ Introduction to Dislocations, Butterworth-Heinemann, 2001 ISBN 0750646810 or Weertman, J and Weertman, J, Elementary Dislocation Theory, Oxford University Press; 1992 ISBN 0195069005.) So why do the lines appear to do just this?
T1.7 Why do you think it is useful to be able to relate the directions in the DP in Figure 1.6 to the directions of the plate-like precipitates in the inset in the same figure?
T1.8 Can you think of an analogy to Figure 1.7 involving images viewed in transmission using visible light? (Hint: think about looking at images of objects at varying distance apart viewed through a transparent medium, e.g., air or water).
T1.9 Examine Figure 1.8; what information in the images would lead you to infer that electron-beam damage was occurring and what could you do to try and avoid further damage? (Hint: what is the difference between the two images and what is causing it?)
T1.10 Would you expect a higher-voltage TEM to be capable of more damage to your specimen than a lower-voltage instrument?
T1.11 Use equations 1.6 and 1.7 in a spreadsheet to reproduce Table 1.2 and add in lines for electrons accelerated to 50 and 250 kV.
T1.12 Compare the usefulness of Google searches for 'TEM' and 'transmission electron microscopy,' 'AEM', and 'analytical electron microscopy.' What does this tell you about acronyms (TMBA?)
T1.13 See how many of the URLs listed in Section 1.6 are still active. Make your own list of useful URLs related to your own interests in electron microscopy.
T1.14 Confirm for yourself the calculation referred to in Section 1.4 that there is never more than one electron inside a typical thin specimen in a 100 keV TEM.
T1.15 Looking at Figure 1.9, you'll notice that high-voltage microscope is bigger than other ones. Can you think of any reasons why this is so?
T1.16 Why do some TEMs have viewing screens and some only have computer displays?
T1.17 Why might you want to build an ultrahigh vacuum TEM and why are there so few available?

2

Scattering and Diffraction

CHAPTER PREVIEW

The electron is a low-mass, negatively charged particle. As such, it can easily be deflected by passing close to other electrons or the positive nucleus of an atom. These Coulomb (electrostatic) interactions cause electron scattering, which is the process that makes TEM feasible. We will also discuss how the wave nature of the electron gives rise to diffraction effects. What we can already say is that if the electrons weren't scattered, there would be no mechanism to create TEM images or DPs and no source of spectroscopic data. So it is essential to understand both the particle approach and the wave approach to electron scattering in order to be able to interpret all the information that comes from a TEM. Electron scattering from materials is a reasonably complex area of physics, but it isn't necessary to develop a detailed comprehension of scattering theory to be a competent microscopist.

We start by defining some terminology that recurs throughout the book and then we introduce a few fundamental ideas that have to be grasped. These fundamental ideas can be summarized in the answers to four questions.

- What is the probability that an electron will be scattered when it passes near an atom?
- If the electron is scattered, what is the angle through which it is deviated?
- What is the average distance an electron travels between scattering events?
- Does the scattering event cause the electron to lose energy or not?

The answer to the first question concerning the probability of scattering is embodied in the idea of a cross section. The angle of scattering (usually determined through the differential cross section) is also important because it allows you as the TEM operator to control which electrons form the image and therefore what information is contained in the image. We will develop this point much further when we talk about image contrast in Part 3 of the book. The third question requires defining the mean-free path, an important concept given that we use thin specimens. The answer to the fourth question requires distinguishing elastic and inelastic scattering. The former constitutes most of the useful information in DPs obtained in the TEM, discussed in Part 2, while the latter is the source of X-rays and other spectroscopic signals discussed in Part 4. The distinction between electrons that lose energy and those that don't is important enough that we devote the next two chapters to each kind of electron and expand on the basic ideas introduced here.

2.1 WHY ARE WE INTERESTED IN ELECTRON SCATTERING?

We need to know about electron scattering because it is fundamental to all electron microscopy (not just TEM). You know well that your eyes cannot see any object unless it interacts with visible light in some way, for example through reflection or refraction, which are two forms of scattering (e.g., we can't see a light beam unless it is scattered by dust within it or it hits a surface). Similarly, we cannot see anything in EM images unless the specimen interacts with and scatters the electrons in some way. Thus, any non-scattering object is invisible and we will come across situations where 'invisibility' is an important criterion in TEM images. In the TEM we are usually most interested in those electrons that do not deviate far from the incident-electron direction. This is because the TEM is constructed to gather these electrons primarily and they also give us the information we seek about the internal structure and chemistry of the specimen. Other forms of scattering, such as electrons which are scattered through large angles (e.g., backscattered

electrons) and electrons ejected from the specimen (such as secondary electrons) are also of interest and we will not totally neglect them (although they are of much greater interest in the SEM where they provide atomic number contrast and surface-sensitive, topographical images, respectively).

WAVE AND PARTICLE

The electron is treated in two different ways throughout this book: in electron scattering it is a succession of particles, while in electron diffraction it is treated by wave theory. The analogy to X-rays or visible light would be to compare a beam of photons and an electromagnetic wave. However, you must always remember that electrons are charged particles and that Coulomb forces are very strong.

In this chapter we introduce the fundamental ideas of electron scattering; then, in the next two chapters, we discuss the two principal forms of scattering, namely, elastic and inelastic. Both forms are useful to us, but you'll see that the latter has the unfortunate side effect of being responsible for specimen damage and ultimately limits what we can do with a TEM.

To give you some feel for the importance of electron scattering, it is worth illustrating at this stage the basic principles of the TEM. You will see in due course that in a TEM we illuminate a thin specimen with a broad beam of electrons in which the intensity is uniform over the illuminated area.

We will often refer to incident and scattered electrons as beams of electrons, because we are dealing with many electrons, not an individual electron; these electrons are usually confined to well-defined paths in the microscope. So the electrons that hit the specimen are often called the incident beam and those scattered by the specimen are called scattered (or sometimes specifically, diffracted) beams. Electrons coming through a thin specimen are separated into those that suffer no angular deviation and those scattered though measurable angles. We call the undeviated electrons the 'direct beam' (in contrast to most texts that describe this as the 'transmitted beam' despite the fact that all electrons coming through the specimen have been 'transmitted'). As the electrons travel through the specimen they are either scattered by a variety of processes or they may remain unaffected. The end result, however, is that a non-uniform distribution of electrons emerges from the exit surface of the specimen, as shown schematically in Figure 2.1. It is this non-uniform distribution that contains all the structural, chemical, and other information about our specimen. So everything we learn about our specimen using TEM can be attributed to some form of electron scattering.

DIRECT BEAM

The beam that comes through the specimen, but remains parallel to the direction of the incident electrons is a very important beam, which we will term the **direct beam**.

We'll see in Chapter 9 that the TEM is constructed to display this non-uniform distribution of electrons in two different ways. First the *spatial distribution* (Figure 2.1A) of scattering can be observed as contrast in *images* of the specimen, and the *angular* distribution of scattering (Figure 2.1B) can be viewed in the form of scattering patterns, usually called *diffraction* patterns. A simple (and fundamental) operational step in the TEM is to use a restricting aperture, or an electron detector, of a size such that it only selects electrons that have suffered more or less than a certain angular deviation. So you as the operator have the ability to choose which electrons you want to use and thus you control what information

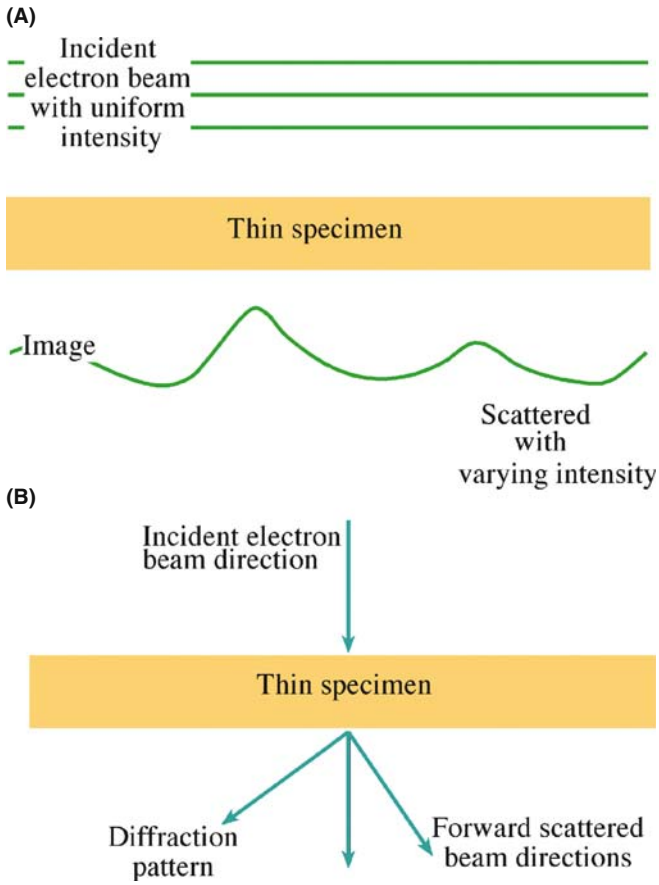


FIGURE 2.1. (A) A uniform intensity of electrons, represented by the horizontal lines, falls on a thin specimen. Scattering within the specimen changes both the spatial and angular distributions of the emerging electrons. The spatial distribution (intensity) is indicated by the wavy line. (B) The change in angular distribution is shown by an incident beam of electrons being transformed into several forward-scattered beams.

will be present in the image. Therefore, to comprehend these images, you have to understand what causes electrons to scatter in the first place. The same is true for DPs since you can also control (to a lesser extent) the angular-scattering distribution, e.g., by tilting your specimen.

We devote the whole of Part 2 to diffraction phenomena and Part 3 to images. Lastly, Part 4 deals with ways in which we use inelastic scattering for analytical electron microscopy (AEM) to study, e.g., the chemistry and the bonding of the atoms in our specimen.

2.2 TERMINOLOGY OF SCATTERING AND DIFFRACTION

Electron-scattering phenomena can be grouped in different ways. We've already used the most important terms: *elastic* and *inelastic* scattering. These terms, respectively, describe scattering that results in no loss of energy or in some measurable loss of energy (usually very small with respect to the beam energy). In either case, we can consider the beam electrons and specimen atoms as particles, and scattering of the incident electrons by the atoms in the specimen can often be approximated to something like billiard balls colliding. The billiard-ball analogy will be good through Section 2.7 after which we'll be talking about waves.

ELECTRON SCATTERING
This theme permeates the whole text and connects ALL aspects of TEM.

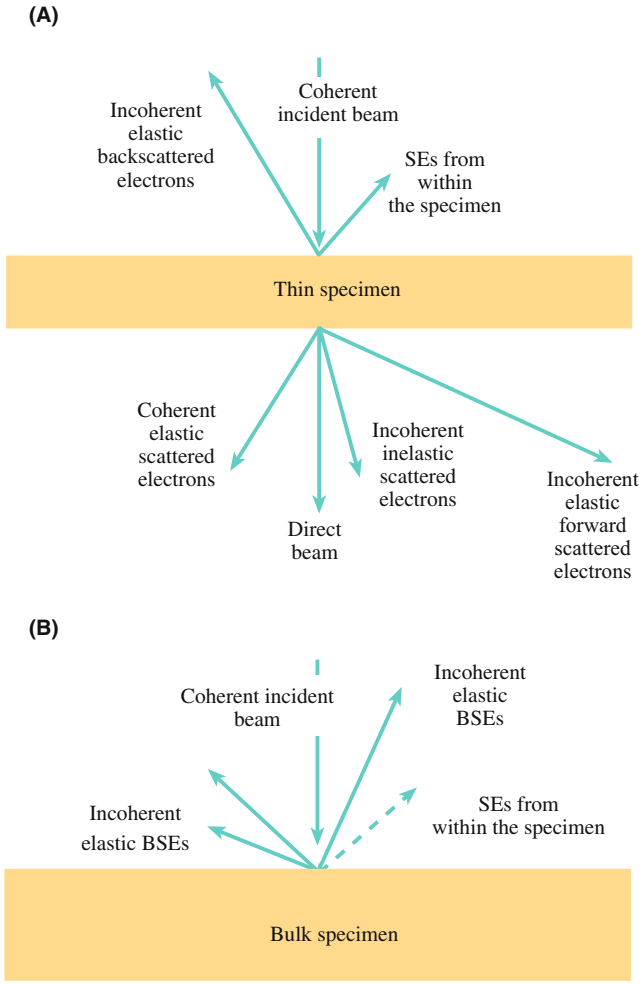


FIGURE 2.2. Different kinds of electron scattering from (A) a thin specimen and (B) a bulk specimen: a thin specimen permits electrons to be scattered in both the forward and back directions while a bulk specimen only backscatters the incident-beam electrons.

However, we can also separate scattered electrons into *coherent* and *incoherent*, which refers, of course, to their wave nature. These distinctions are related since elastically scattered electrons are usually coherent and inelastic electrons are usually incoherent (note the modifier 'usually'). Let's assume that the incident electron waves are coherent, that is, they are essentially in step (in phase) with one another and of a fixed wavelength, governed by the accelerating voltage. You'll see that this isn't a bad assumption in most circumstances. Then, coherently scattered electrons are those that remain in step and incoherently scattered electrons have no phase relationship, after interacting with the specimen.

The nature of the scattering can result in different angular distributions. Scattering can be either *forward scattering* or *back scattering* (usually written as one word) wherein the terms refer to the angle of scattering with respect to the incident beam and a specimen that is normal to that beam. (Note: you will sometimes see the term 'forward scattering' used in another sense.) If an electron is scattered through $< 90^\circ$, then it is forward scattered and $> 90^\circ$ it is backscattered. These various terms are related by the following general principles, summarized in Figure 2.2.

- Elastic scattering is usually coherent, if the specimen is thin and crystalline (think in terms of waves).
- Elastic scattering usually occurs at relatively low angles (1–10°), i.e., it is strongly peaked in the forward direction (waves).
- At higher angles ($>\sim 10^\circ$) elastic scattering becomes more incoherent (now think of particles).
- Inelastic scattering is almost always incoherent and is very low angle ($<1^\circ$) scattering (think particles).
- As the specimen gets thicker, fewer electrons are forward scattered and more are backscattered. Incoherent, backscattered electrons are the only remnants of the incident beam that emerge from bulk, non-transparent specimens (think particles).

The notion that electrons can be scattered through different angles is related to the fact that an electron can also be scattered more than once. Generally, the more scattering events, the greater the angle of scattering (although sometimes a second scattering event can redirect the electron back into the direct beam, so it appears to have undergone no scattering).

The simplest scattering process is *single scattering* and we often approximate all scattering within a TEM specimen to this process (i.e., an electron either undergoes a single-scattering event or it suffers no scattering). We'll see that this can be a very reasonable assumption if the specimen is very thin (something you can control). If the electron is scattered more than once, we use the term *plural scattering* and if it is scattered >20 times, we say *multiple scattering*. It is generally safe to assume that, unless you have a particularly thick specimen (through which you probably can't see anything anyhow), multiple scattering will not occur in the TEM. The greater the number of scattering events, the more difficult it is to predict what will happen to the electron and the more difficult it is to interpret the images, DPs, and spectra that we gather. So, once again, we emphasize the importance of the 'thinner is better' criterion, i.e., if you create thin enough specimens so that the single-scattering assumption is plausible, your TEM research will be easier.

Diffraction is a very special form of elastic scattering and the terminology used can be confusing. Collins' Dictionary defines *diffraction* as 'a deviation in the direction of a wave at the edge of an obstacle in its path' while *scattering* is defined as 'the process in which particles, atoms, etc., are deflected as a result of collision.' The word scatter can also be a noun denoting the act of scattering. So scattering might best apply to particles and diffraction to waves; both terms thus apply to electrons! You should also note that the term diffraction is not limited to Bragg diffraction which we'll emphasize in TEM; it refers to *any* interaction involving a wave, but many texts are not consistent in this respect.

DEFINE DIFFRACTION
An interaction between a wave of any kind and an object of any kind (Taylor 1987).

In the TEM we utilize the electrons that go through a specimen; it is important to note that such electrons are not simply 'transmitted' in the sense of visible light through window glass. Electrons are scattered mainly in the forward direction, i.e., parallel to the incident beam direction (and we've already noted the confusion between 'direct' and 'transmitted'). We'll tell you in a short while what fraction of the electrons are forward scattered and how this varies with the thickness of the specimen and atomic number of the 'target' atom. This scattering is a direct consequence of the fact that there is such a strong interaction between electrons and matter.

Forward scattering includes the direct beam, most elastic scattering, diffraction, particularly Bragg diffraction (see Chapter 3), refraction, and inelastic scattering (see Chapter 4). Because of forward scattering through our thin specimen, we see a DP or an image on the viewing screen, and detect an X-ray spectrum or an electron energy-loss spectrum outside the TEM column. But don't neglect backscattering; it is an important imaging mode in the SEM.

FORWARD SCATTERING
The cause of most of the signals used in the TEM.

When physicists consider the theory of electron interactions within a solid, they usually consider scattering of electrons by a single, isolated atom, then progress to agglomerations of atoms, first in amorphous solids and then in crystalline solids and we'll follow a similar path.

2.3 THE ANGLE OF SCATTERING

When an electron encounters a single, isolated atom it can be scattered in several ways which we will cover in the next two chapters. For the time being, let's imagine simply that, as shown in Figure 2.3, the electron is scattered through an angle θ (radians) into some solid angle ω, measured in steradians (sr). We have to define this angle first because you'll see that it plays an important role in the subsequent discussion of cross sections.

SEMI-ANGLE
Note that the scattering angle θ is in fact a semi-angle, not a total angle of scattering. Henceforth, whenever we say "scattering angle" we mean "scattering semi-angle."

Often we assume that θ is small enough such that $\sin\theta \approx \tan\theta \approx \theta$. When θ is this small, it is often convenient to use milliradians or mrads; 1 mrad is 0.0573°, 10 mrads is ~0.5°.

SMALL ANGLE
A convenient upper limit is <10 mrads.

The characteristics of the scattering event are controlled by factors such as the incident-electron energy and the atomic number/weight of the scattering atom. When we consider a specimen rather than a single

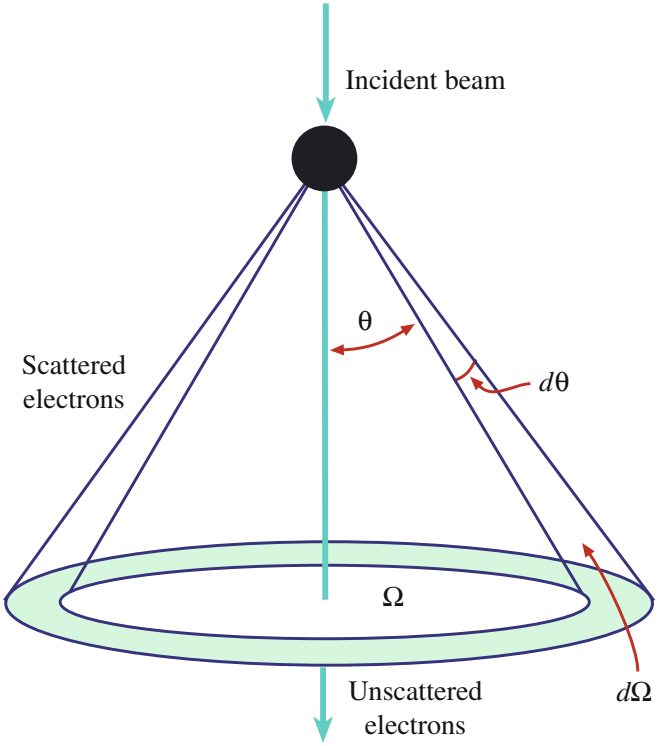


FIGURE 2.3. Electron scattering by a single isolated atom. The electrons are scattered through an angle θ and the total solid angle of scattering is Ω. An incremental increase in scattering angle $d\theta$ gives an incremental increase in a solid angle $d\Omega$, which is the basis for determining the differential scattering cross section.

atom, factors such as the thickness, density, crystallinity, and angle of the specimen to the incident beam also become important. To understand these variables, we need to examine the physics of scattering in more detail. Of necessity, we'll be rather brief and often imprecise since we're trying to condense much of Mott and Massey's substantial and classic textbook into just a few pages.

2.4 THE INTERACTION CROSS SECTION AND ITS DIFFERENTIAL

The chance of a particular electron undergoing any kind of interaction with an atom is determined by an interaction *cross section*. The concept of a cross section is well described by the following analogy given by Rudolf Peierls (Rhodes 1986)

> "If I throw a ball at a glass window one square foot in area, there may be one chance in ten that the window will break and nine chances in ten that the ball will just bounce. In the physicist's language this particular window, for a ball thrown in this particular way, has a disintegration (inelastic!) cross section of 0.1 square feet and an elastic cross section of 0.9 square feet."

So each possible interaction has a different cross section which depends on the energy of the particle, in our case the beam energy. The cross section (for which we'll use the Greek letter σ) has units of area (not square feet as used in Peierls' analogy, but a tiny fraction of the area of an atom termed a 'barn'). One barn is 10^{-28} m^2 (that's $(10^{-5}$ nm$)^2$) and the name arises because of the perverse sense of humor of some of the early atomic physicists who considered that this unimaginably small area is 'as big as a barn door.' The cross section does *not* represent a physical area but, when divided by the actual area of the atom, it represents a *probability* that a scattering event will occur.

2.4.A Scattering from an Isolated Atom

First of all we'll consider the scattering cross section for a single isolated atom, then extend the concept to a specimen with many atoms. We'll use a generalized form to start with in this chapter and then break down the concept of a total cross section into cross sections for individual processes such as elastic scattering and the various inelastic processes in the next two chapters.

SCATTERING PROBABILITY
The larger the cross section, the better the chances of scattering.

Following Heidenreich (1964), we can define the cross section (an area) in terms of the *effective radius* of a single, isolated atom, r

$$\sigma_{\text{atom}} = \pi r^2 \tag{2.1}$$

where r has a different value for each scattering process as we'll see in the next chapter. What interests us in the TEM is whether or not the scattering process deviates the incident-beam electrons outside a particular scattering angle θ such that, e.g., they do not go through the aperture in the lens or they miss the electron detector. So we have to know the *differential cross section* $(d\sigma/d\Omega)$ which describes the angular distribution of scattering from an atom. As shown in Figure 2.3 electrons are scattered through an angle θ into a solid angle Ω and there is a simple geometrical relationship between θ and Ω

$$\Omega = 2\pi(1 - \cos\theta) \tag{2.2}$$

and therefore

$$d\Omega - 2\pi \sin\theta d\theta \tag{2.3}$$

So the differential scattering cross section for a single, isolated atom can be written as

$$\frac{d\sigma}{d\Omega} = \frac{1}{2\pi \sin\theta}\frac{d\sigma}{d\theta} \tag{2.4}$$

Now, we can calculate σ_{atom} for scattering into all angles *greater than* θ by integrating equation 2.4 from θ to π

$$\sigma_{atom} = \int_{\theta}^{\pi} d\sigma = 2\pi \int_{\theta}^{\pi} \frac{d\sigma}{d\Omega} \sin\theta \, d\theta \tag{2.5}$$

The limits of the integration are governed by the fact that the values of θ can vary from 0 to π, depending on the specific type of scattering. If we work out the integral we find that σ decreases as θ increases (which makes physical sense). Since $d\sigma/d\Omega$ is often what is measured experimentally (but not in the TEM), equation 2.5 gives us an easy way to determine the cross section for a given atom (σ_{atom}) for all values of θ, by working out the integral of $(d\sigma/d\Omega)\sin\theta$ from 0 to π.

INTEGRATE

If we integrate from 0 to θ then we determine the cross section for scattering into all angles less than θ which is in fact more relevant to the TEM situation.

2.4.B Scattering from the Specimen

So let's move on from the cross section for a single isolated atom (with units of area) and consider that the specimen contains N atoms/unit volume. Therefore, we can define the total cross section for scattering from the specimen (in units of m^{-1}) as

$$\sigma_{total} = N\sigma_{atom} \tag{2.6}$$

Since $N = N_0\rho/A$ where N_0 is Avogadro's number (in units of atoms mol^{-1}), A is the atomic weight of the scattering atoms in the specimen (kg mol^{-1}) which has density ρ (kg m^{-3}), we can write

$$\sigma_{total} = N\sigma_{atom} = \frac{N_0\sigma_{atom}\rho}{A} \tag{2.7}$$

Thus, σ_{total} is the number of scattering events per unit distance that the electron travels through the specimen. If the specimen has thickness t, then the probability of scattering from the specimen is given by

$$\sigma_{total}t = \frac{N_0\sigma_{atom}(\rho t)}{A} \tag{2.8}$$

Here we've gathered together the product of ρ and t because this is called the mass thickness of the specimen (e.g., doubling ρ produces the same effect as doubling t) and we'll come across this term again when we discuss image contrast and also X-ray absorption. Equation 2.8 is an important expression, since it contains all the variables that affect the scattering probability from a real specimen. We'll use it again when we consider how certain kinds of image contrast arise in the TEM.

So we can now appreciate, through a few (rather simplified) equations, the relationship between the physics of electron scattering and the information we collect in the TEM.

Expressions for the cross section become more complicated as they are modified to give better approximations for the scattering in a real specimen, as we'll see in the next couple of chapters. However, the more complex equations don't alter the basic scattering behavior predicted by the simple equations we've just given.

2.4.C Some Numbers

Because of all the variables that affect σ_{atom} and σ_{total}, it is only possible to give a ballpark (barnyard?) value for the cross section. For TEM electron energies (100–400 keV), the elastic cross section is almost always the dominant component of the total scattering. If you look ahead to Figure 3.3, typical small-angle, elastic cross sections for transition metals bombarded by 100-keV electrons are $\sim 10^{-22}$ m^2. This is a good number to remember for typical elastic scattering. Inelastic cross sections are generally smaller and range from $\sim 10^{-22}$ m^2 down to 10^{-26} m^2 depending on the specific type of scattering and the material. Going back to equation 2.1, a typical scattering radius (r) is $\sim 10^{-11}$ m or ~ 0.01 nm, which might seem a bit small (about a tenth of an atomic radius) but, since the scattering is localized to the inner or core shells which are closer to the nucleus or to a particular electron-electron interaction, perhaps this isn't such a bad estimate, given all the caveats.

2.5 THE MEAN FREE PATH

Instead of using an area to describe the interaction we can use a length since the distance an electron travels between interactions with atoms is clearly going to be important when we are using thin specimens. This new parameter is then the average distance that the electron travels between scattering events. This distance is important because, if we know what it is, we can work out how thin we have to make our specimen, so plural scattering is not significant, thus making it easier to interpret our images and spectroscopic data in terms of single-scattering theory. The term σ_{total} can be expressed as the inverse

of the mean free path, λ. Because the dimensions of σ_{total} are m^{-1} (you can check this using equations 2.1 and 2.7) there is a simple expression for the mean free path λ which has units of length

$$\lambda = \frac{1}{\sigma_{total}} = \frac{A}{N_0 \sigma_{atom} \rho} \qquad (2.9)$$

Typical values of λ for scattering at TEM voltages are of the order of tens of nm, so single-scattering approximations imply specimen thicknesses of this order. It is, unfortunately, conventional to use λ to denote the mean free path; it is *not* the wavelength of the electron. From equation 2.9 we can define a probability of scattering p as the electron travels through a specimen thickness t

$$p = \frac{1}{\lambda} = \frac{N_0 \sigma_{atom} (\rho t)}{A} \qquad (2.10)$$

which is $(\sigma_{total} t)$ from equation 2.8.

Although computational resources are constantly improving, our knowledge of the values of σ, λ, and θ is imprecise at best, particularly at the 100–400 keV beam energies used in TEMs. Cross sections and mean free paths for particular scattering events may only be known within a factor of two, but we can often measure θ very precisely in the TEM. We can combine all our knowledge of scattering in Monte Carlo simulations to predict the electron paths as a beam is scattered through a thin foil.

MONTE CARLO SIMULATION

So called because random numbers are used in the computer programs; the outcome is always predicted by statistics!

The first Monte Carlo calculation was developed by two of the United States' foremost mathematicians, John von Neumann and Stanley Ulam, at Los Alamos in the late 1940s. Ulam actually rolled dice and made hand (!) calculations to determine the paths of neutrons through deuterium and tritium which proved that Teller's design for the 'Super' (H-bomb) was not feasible (Rhodes 1995). Monte Carlo methods are used more often in SEM imaging and X-ray calculations (see, e.g., NIST's Web site (URL #1), Joy 1995, Goldstein et al. 2003) but they have a role in TEM in determining the expected spatial resolution of X-ray analysis as we'll discuss in Chapter 36. Figure 2.4 shows Monte Carlo simulations of electron paths through thin foils of Cu and Au.

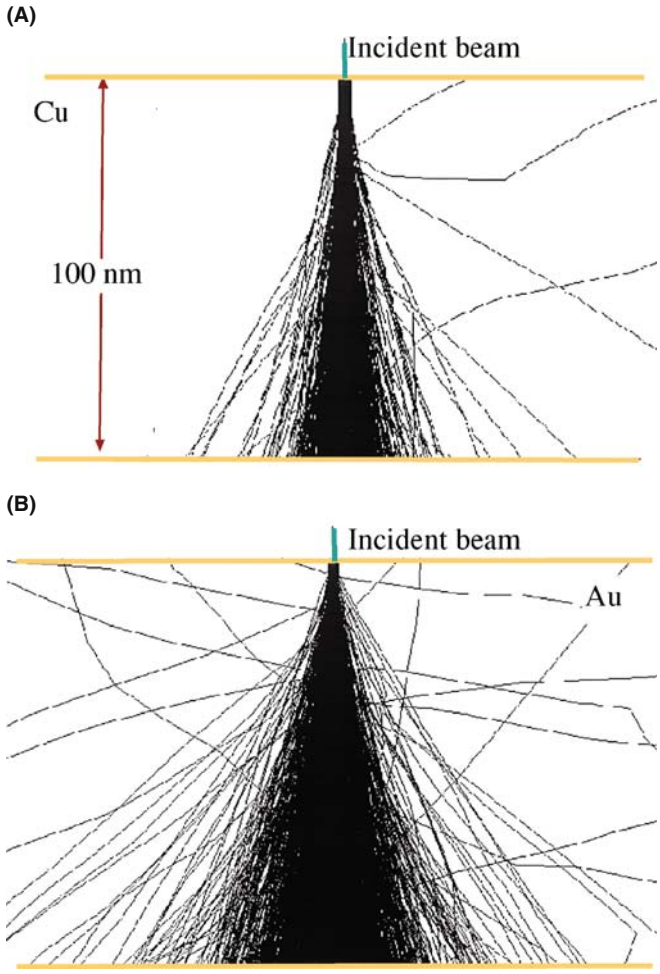


FIGURE 2.4. Monte Carlo simulation of the paths followed by 10^3 100-keV electrons as they pass through thin foils of (A) Cu and (B) Au. Notice the increase in scattering angle with atomic number and the small number of electrons that are scattered through >90°.

2.6 HOW WE USE SCATTERING IN THE TEM

So why have we made you go through all this math? Because when you select electrons that have been scattered through a certain angle (choosing a θ), you are changing the effective scattering cross section (σ_θ), because the scattering strength generally decreases as the angle of scattering increases. Therefore, there will generally be less scattering at higher angles, which explains why we said at the start of the chapter that we are mainly interested in forward scattering in the TEM. Most scattered electrons are contained within ±5° of the direct beam.

300 VERSUS 100 kV

Total σ decreases as E_0 increases; electron scattering at 300 kV will be less than at 100 kV. Higher-density regions of your specimen scatter more than lower-density regions. The target becomes smaller as the bullets become faster!

You also have control of the scattering cross section in other ways. First, the accelerating voltage, which determines the electron energy E_0 (eV), will affect the cross section as implied in equation 2.3 (specifically for elastic scattering). In fact, for all forms of scattering, the total cross section decreases as E_0 increases. Therefore, intermediate- and higher-voltage TEMs will result in *less* electron scattering than typical 100-kV instruments and, as we'll see in Chapter 4, this has important implications for electron-beam damage in delicate specimens, such as polymers. Second, and more intuitively, you can choose specimens with different densities. Denser materials scatter more strongly, so you have to make them thinner to keep the single-scattering approximation valid.

We shall see in the next two chapters that the effect of the atomic number of the atom is more important in elastic than inelastic scattering and, as Z increases, elastic scattering dominates. This behavior helps when we consider ways to enhance scattering (and therefore contrast) in low-Z materials such as polymers and biological tissue.

2.7 COMPARISON TO X-RAY DIFFRACTION

There is a very good reason why electrons are used in microscopy: they have a 'suitable interaction' with matter. Most descriptions of the interaction of electrons with matter are based on scattering. You will come across such topics as kinematical scattering and dynamical scattering in addition to elastic and inelastic scattering, and we will use the formalism of a scattering factor to describe the process mathematically. It is this scattering process that varies with the structure or composition of the specimen, permitting us ultimately to image a microstructure, record a DP, or collect a spectrum. We'll see in the next chapter that scattering factors are used when we consider electrons as waves and their diffraction as a specific form of scattering.

So now it's time to move from billiard balls to waves. Historically, it was diffraction that provided most of the crystallographic information we have about materials, and the majority of those studies used X-rays. This is partly why X-ray diffraction is so well documented in the scientific literature. A good understanding of X-ray diffraction helps considerably in understanding electron diffraction; however, the primary processes by which electrons are scattered are very different to the processes by which X-rays are scattered. Electron scattering is much more complex.

X-rays are scattered by the *electrons* in a material through an interaction between the negatively charged electrons and the electromagnetic field of the incoming X-rays. The electrons in the specimen respond to the applied field of the X-ray flux, oscillating with the period of the X-ray beam. These accelerated charged particles then emit their own electromagnetic field, identical in wavelength and phase to the incident X-rays. The resultant field, which propagates radially from every scattering source, is called the scattered wave.

ELECTRONS VERSUS X-RAYS

Electrons are scattered much more strongly than X-rays.

Electrons are scattered by *both* the electrons and the nuclei in a material; the incoming negatively charged electrons interact with the local electromagnetic fields of the specimen. The incoming electrons are therefore directly scattered by the specimen; it is not a field-to-field exchange as occurs for the case of X-rays.

2.8 FRAUNHOFER AND FRESNEL DIFFRACTION

Diffraction of visible light is well understood, so we should carry over as much of the analysis as possible. Optics is a venerable discipline with a history of several hundred years and what we're trying to do here is condense the principal messages from classic texts such as Hecht (2003) into a few pages. So, as with electron scattering, we'll be making a few simplifications. If you have any experience with diffraction of visible light you will have encountered Fraunhofer and Fresnel diffraction.

- Fraunhofer diffraction occurs when a flat wavefront interacts with an object. Since a wave emitted by a point becomes planar at large distances, this is known as far-field diffraction.
- Fresnel diffraction occurs when it's not Fraunhofer. This case is also known as near-field diffraction.

We will see later that electron-diffraction patterns correspond closely to the Fraunhofer case while we 'see' the effects of Fresnel diffraction in our images.

In TEM we will find both forms of diffraction. We will briefly go through the Huygens' explanation of how a wave propagates, then consider Fraunhofer diffraction from two slits (Young's slits) and then extend this process to many slits. So why discuss these topics now? There are two reasons for reviewing this analysis

- It reminds us that coherent interference is purely a matter of physical optics.
- We can introduce the concept of phasor diagrams which we'll use in later chapters.

Huygens explained the propagation of any wavefront by imagining that each point on the wavefront itself acts as a new source for a spherical wavelet. The wavelets interfere with one another to give the new wavefront and the process is repeated.

2.9 DIFFRACTION OF LIGHT FROM SLITS AND HOLES

In this section we will very briefly review the topic known as physical or geometric optics as it relates to diffraction. Much of what we know about diffraction of electron waves has been carried over from the understanding of the diffraction of visible light and X-rays. There are textbooks on this topic if you don't vaguely remember it from high school.

Two slits (the Young's slits experiment)

We start with diffraction caused by a wavefront incident on a pair of very narrow slits. We then select just two of the Huygens wavelets; these wavelets then must have the same phase at the slits. As they propagate past the slits, their phases differ depending on the position of the detector. The important term is the path difference $L = d \sin\theta$ as shown in Figure 2.5. The two wavelets propagating in direction $\boldsymbol{r}$ have a path difference of L and a phase difference of $2\pi L/\lambda$, i.e., $2\pi d \sin\theta/\lambda$. If d and λ are such that this phase difference is actually a multiple of 2π (so $d \sin\theta/\lambda =$ an integer, n) then the rays are again in phase and their amplitudes add. The condition for this additive interference is thus that $d \sin\theta = n\lambda$. Therefore, there is an inverse relationship between d and θ for a given d; as d decreases, $\sin\theta$ increases. If we think of each wavelet as having an amplitude and a phase we can represent each by a vector—a phasor. When the phasors are parallel to one another (in phase) they add; when they are antiparallel, they cancel (since they have equal lengths). A phasor diagram is a way to plot the amplitude and phase of the total scattered wave; in other words, when we add the amplitudes of beams we must take into account their phase.

THE INVERSE RELATIONSHIP

$\theta \propto d^{-1}$ solely due to the positions of the slits. We'll come across an identical relationship when we talk about electron diffraction.

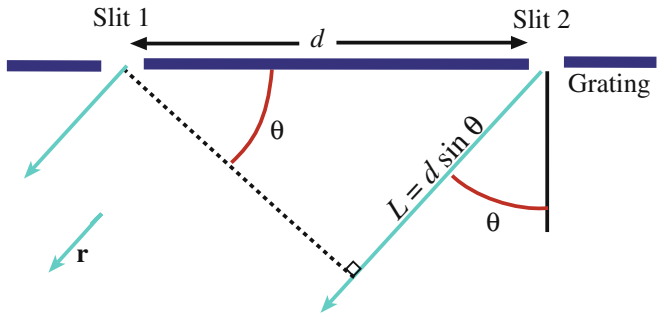


FIGURE 2.5. An incident plane wave is scattered by two slits, distance d apart. The scattered waves are in phase when the path difference $d \sin\theta$ is $n\lambda$.

Many slits (the diffraction grating)

When we extend this analysis to more than two slits we see a similar result but with added subsidiary peaks. The origin of the subsidiary peaks can best be illustrated by considering a series of phasor diagrams. (We'll find similar diagrams useful when we discuss TEM images in Chapter 27.) We'll examine the case of five slits. Each of the polyhedra in Figure 2.6 represents a different value of θ. When θ is zero, the five rays are all in phase and we simply add all of the amplitudes (the phasor vectors are aligned); as θ increases the rays become out of phase but the phasors can still add to give a large resultant vector but can also add to give zero. For example, when θ is exactly 72° (360°/5 for 5 slits) the phasor diagram is a closed pentagon (shown in the figure) and the resultant amplitude is zero. This process repeats at 144° (2×720°/5) and 216° (3×360°/5). In between these values at 108°

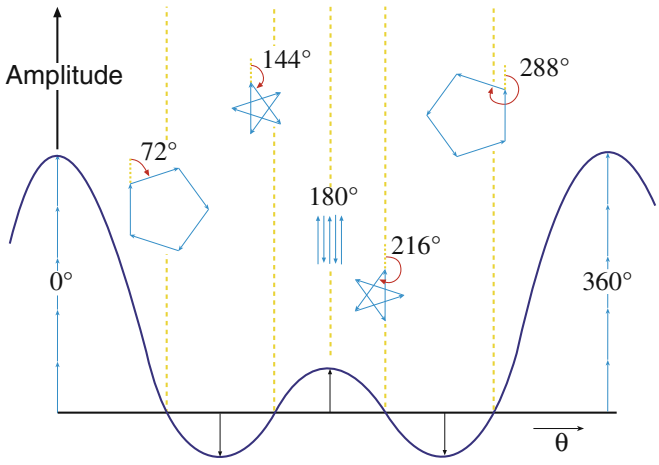


FIGURE 2.6. A phasor diagram showing how the total amplitude produced by summing five waves produced by five slits varies with the phase angle θ between the different waves. The individual phasors from each of the five slits sum to create a total amplitude of zero at $\theta = 72°, 144°, 216°$, and 288°, large positive amplitudes at $\theta = 0°$ and 360°, a single-phasor negative amplitude at $\theta = 108°$ and 252°, and a single-phasor positive amplitude at $\theta = 180°$. Remember the intensity is governed by the square of the amplitude, so positive and negative values both contribute to diffracted intensity.

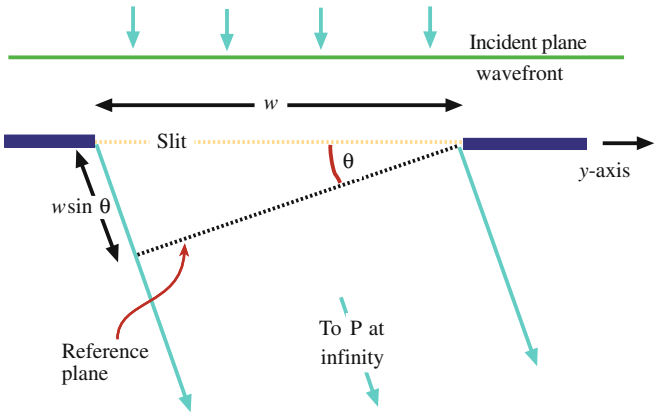


FIGURE 2.7. Geometry for the scattering from an individual slit.

$(1.5\times360^\circ/5)$ we produce a local maximum in amplitude which is repeated at 180° $(2.5\times360^\circ/5)$. If we plot the amplitude as a function of θ, we produce the curve with a series of subsidiary maxima shown in Figure 2.6. From this figure you see that the amplitude is a strong function of θ and you'll learn in the next chapter that the electron intensity (which is what we see in images and DPs) is proportional to the square of the amplitude (so negative amplitudes don't mean anything) and the scattered electron intensity is, therefore, a similarly strong function of θ.

A single wide slit

What happens if we allow the slit to have some width as shown in Figure 2.7? Now the rays from *within* a single slit will interfere with each other. We can think of the single slit as being many adjacent slits of width δw. Imagine dividing the one slit into 11 slits of width $\delta w/11$. This one slit would then produce a phasor diagram as shown in Figure 2.8; if we make δw increasingly small, the phasor diagram becomes a curve: instead of having Figure 2.8, we have Figure 2.9 (for several different values of θ). If you do the full analysis you'd find that the amplitude from a single slit varies as $A = A_0\phi^{-1}\sin\phi$ where ϕ is the phase $\pi w \sin\theta/\lambda$ for a slit of width w (which reminds you of the analysis from Figure 2.5). For one slit, we would see a zero in the phasor diagram when $\phi = \pm n\pi$. If we plot the intensity (rather than A) we obtain the Airy plot shown in Figure 2.10.

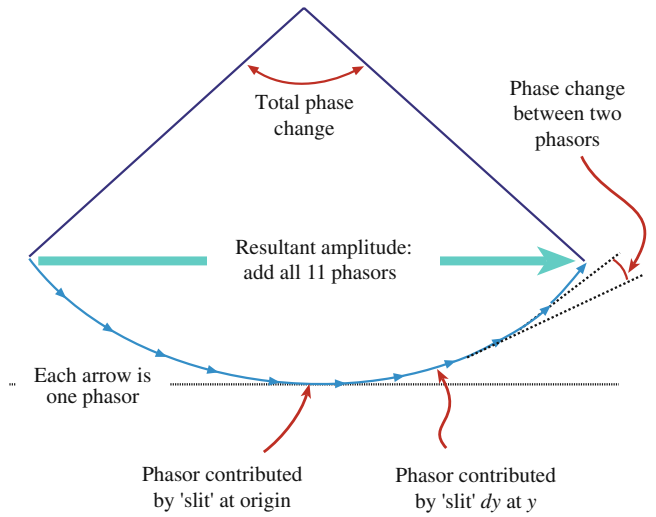


FIGURE 2.8. How the phasors from within an individual slit can be added to give the total phasor for the slit shown in Figure 2.7.

THE AIRY DISK

The disk of radius $r = 1.22\lambda/D$ is named after Airy and is one of the fundamental limits on the achievable resolution in TEM, as we will discuss in Chapter 6. If we introduce *any* aperture into *any* microscope we will limit the ultimate resolution of the instrument.

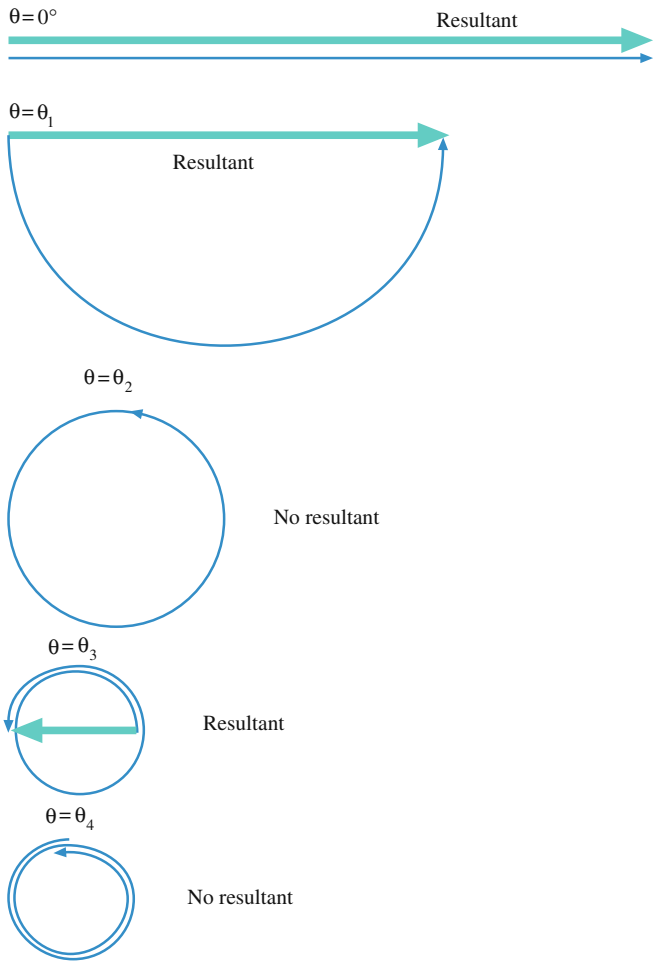


FIGURE 2.9. How a single slit can produce a beam which has zero amplitude for certain values of θ in Figure 2.7. The circles are directly comparable to the polyhedra in Figure 2.6. The total length of the phasor increments (from each dy) is the same in each figure.

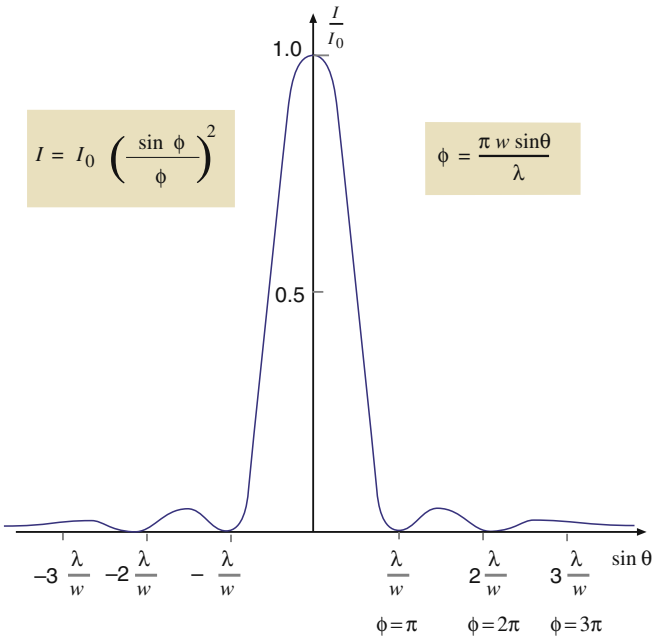


FIGURE 2.10. The plot of the resulting intensity for scattering from the slit shown in Figure 2.7; this is known as the Fraunhofer DP from a single slit; w is the slit width defined in Figure 2.7.

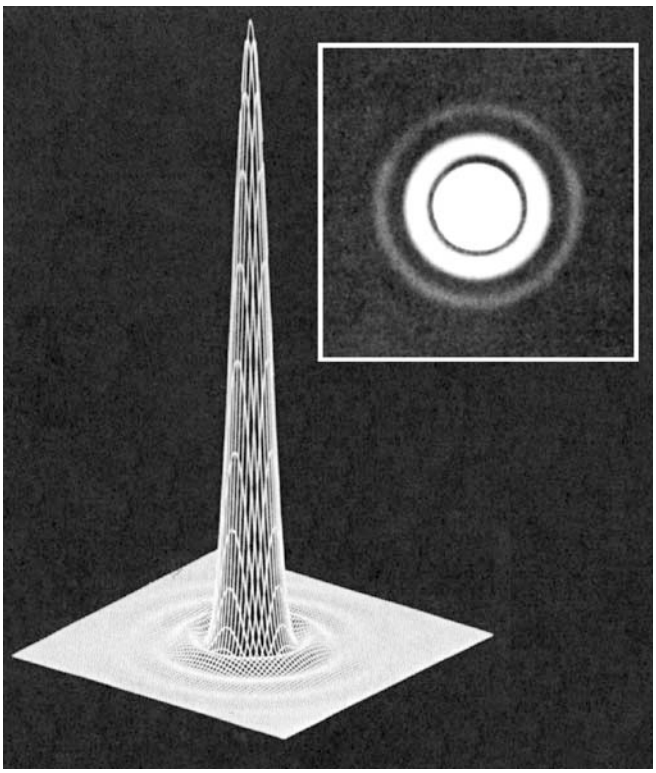


FIGURE 2.11. The visible-light intensity produced by a 0.5-mm-diameter circular aperture and the observed Airy rings (inset). The width of the central intense region is $1.22\lambda/D$.

Scattering from a circular hole

Now the real purpose of the exercise: without going into the detailed math we can replace the slit of width w, by a circular hole or aperture of diameter D. The resulting peak width in the plot of amplitude versus θ then has a maximum at $1.22\lambda/D$ as shown in Figure 2.11, which is a 3D representation of Figure 2.10 (but the third dimension is I, not I/I_0).

Because of the circular symmetry of the aperture, the calculation needed to obtain the number 1.22 involves the use of Bessel functions which you can find in texts on physical optics, a few of which we reference at the end of the chapter.

As the diameter of the aperture, D, decreases, the minimum resolvable spacing, r, increases (i.e., the resolving power gets worse). This expression for the Airy disk diameter also shows that as λ decreases, r decreases (so decreasing λ by increasing the accelerating voltage of the TEM will improve resolution).

Why is this relevant to TEM?

The important point about this analysis for TEM is that we'll see the same relationship in several later chapters. In those chapters, we will replace the slits by an aperture or we'll replace the hole by an atom or by your specimen. In other words, this analysis of diffraction from slits and holes is just geometry applied to optics—it's geometric optics.

2.10 CONSTRUCTIVE INTERFERENCE

To expand on this point, consider an infinite plane wave described by the usual characteristics of amplitude and phase. We can describe the wave function ψ by the standard expression

$$\psi = \psi_0 \exp[i\phi] \tag{2.11}$$

where ψ_0 is the amplitude and ϕ the phase of the wave. The phase depends on position x, such that if the path length changes by one wavelength λ, the phase difference is 2π. Stated another way, the phase difference Δϕ between any two monochromatic (same wavelength) waves is related to the path difference Δx they must travel in going from the source to the detector. The relationship is

$$\Delta\phi = \frac{2\pi}{\lambda}\Delta x \tag{2.12}$$

This phenomenon of constructive interference is precisely what we discussed in Figure 2.6. Constructive interference between waves relies on the fact that the amplitudes of the waves add when you take account of the phase. If all waves scattered by all of the atoms in the specimen are to interfere constructively, they must all differ in phase by integral multiples of 2π. Clearly this condition requires that the path differences

traveled by all of the waves be integer multiples of the wavelength of the incident wave. We can ensure this by requiring that the scattering centers be periodically spaced; fortunately this can happen for all crystals. So the mathematical description of constructive interference is simplified (as we'll see in Part 2 of this text). The point here is that this analysis was carried out for X-rays and was not changed for electrons since it does not depend on the scattering mechanism, only on the geometry.

2.11 A WORD ABOUT ANGLES

Since angles (remember we mean semi-angles) are so important in the TEM (you can control some of them and the specimen controls others) we want to try to be consistent in our terminology.

- We can control the angle of incidence of electrons on the specimen and we will define the angle of incidence as α, as shown in Figure 2.12.
- In the TEM we use apertures or detectors to collect a certain fraction of the scattered electrons and we will define any angle of collection as β.
- We will define all scattering angles controlled by the specimen as θ. This may be a specific angle, such as twice the Bragg angle (where $\theta = 2\theta_B$) (see Section 11.4) or a general scattering angle θ. So θ is the scattering semi-angle for diffraction even though it is $2\theta_B$!

In fact the only angle of interest in TEM which is not given as a semi-angle is the solid angle of collection of X-rays by the XEDS detector (see Chapter 32) which is such a miserably small fraction of the total solid angle of X-ray generation (4π sr) that it is traditionally given in terms of the full collection angle!

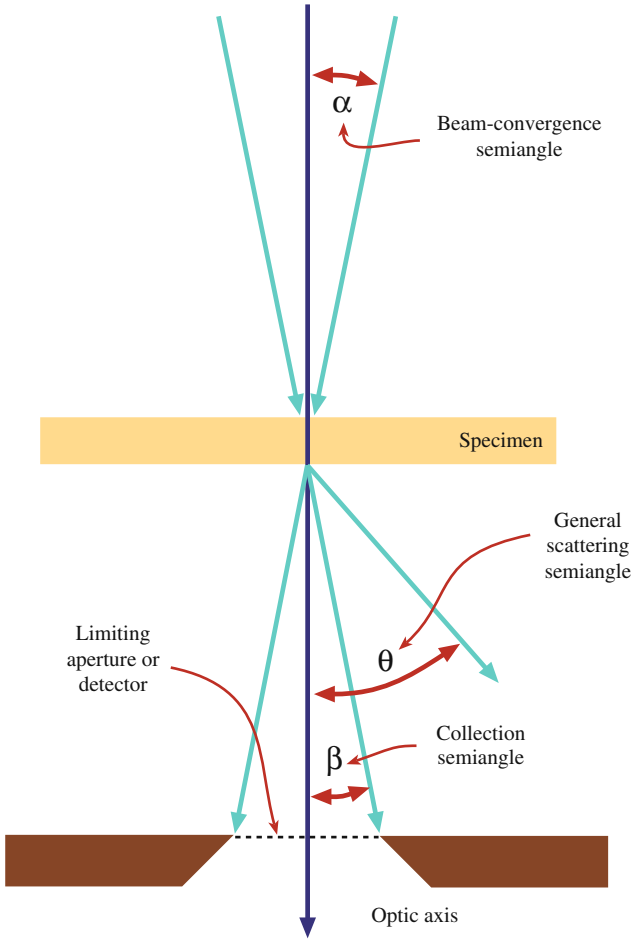


FIGURE 2.12. Definition of the major angles (i.e., semi-angles) in TEM. Any incidence/convergence angle of the beam is α; any collection angle is β and general scattering angles are θ. All the angles are measured from the optic axis, an imaginary line along the length of the TEM column.

2.12 ELECTRON-DIFFRACTION PATTERNS

We've mentioned a couple of times that the TEM is uniquely suited to take advantage of electron scattering because it can form a picture (DP) of the distribution of scattered electrons, which we'll discuss in Part 2 in much more detail. To understand fully how a DP is formed in the TEM, you need to go to Chapter 6 to see how electron lenses work and then to Chapter 9 to find how we combine several lenses to create the TEM imaging system. But before we take you through these concepts it is worth just showing a few of the many kinds of DPs that can be formed in the TEM. At this stage, all you have to do is imagine that a photographic film is placed directly after the thin specimen and that electrons scattered by the specimen as in Figure 2.1B impinge directly on the film. Under these circumstances, the greater the angle of scatter, the further off center the electron hits the film.

ON THE 'FILM'
Thus in a DP, distances on the film correspond to angles of scatter at the specimen.

Even using this simple description, however, you can comprehend some of the basic features of DPs. Figure 2.13 is a montage of several kinds of DPs, all of which are routinely obtainable in a TEM. You can see that several points we've already made about scattering are intuitively obvious in the patterns. First, most of the intensity is in the direct beam, in the center of the pattern, which means that most electrons appear to travel straight through the specimen. Second, the scattered intensity decreases with increasing θ (increasing distance from the direct beam), which reflects the

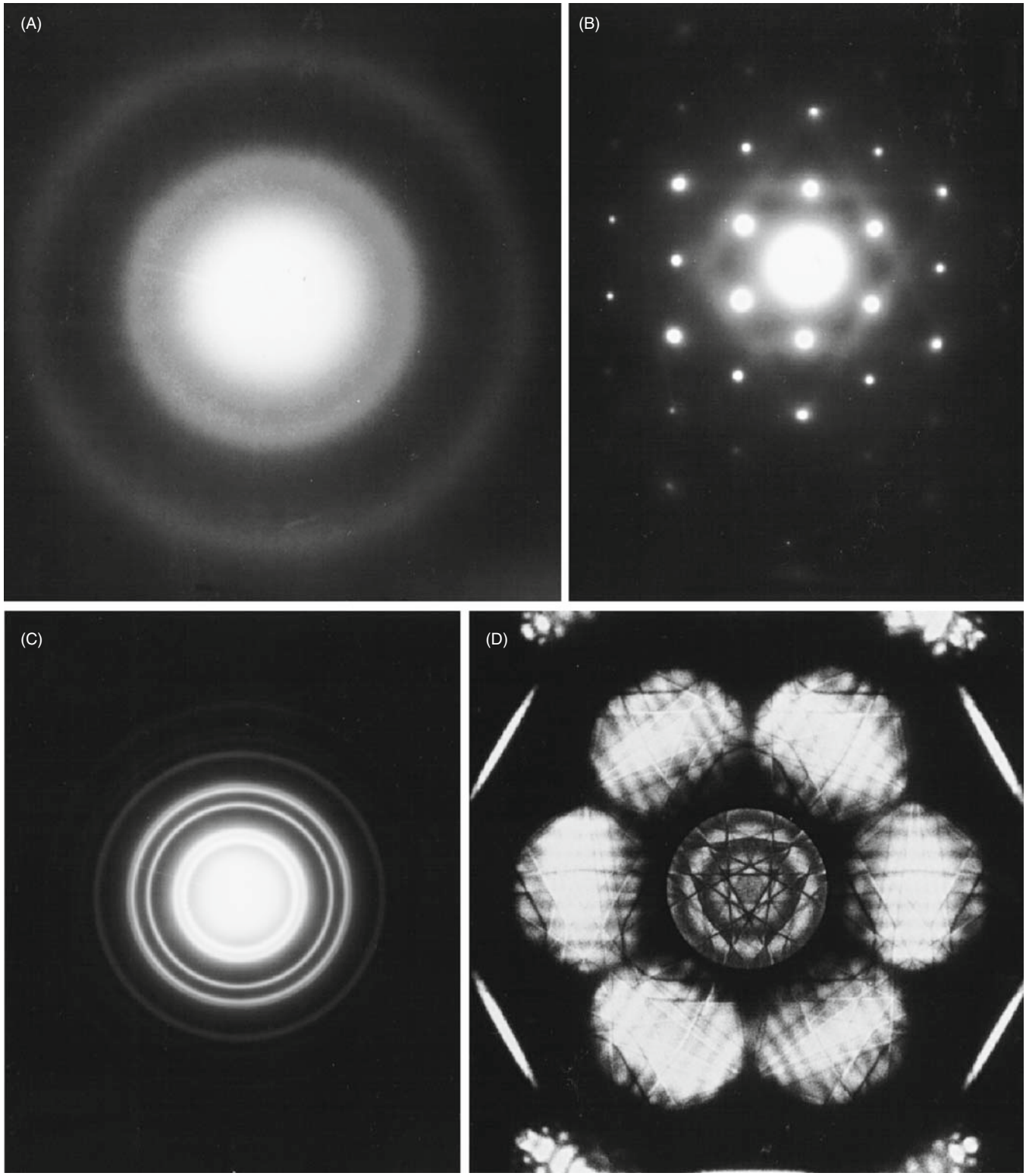


FIGURE 2.13. Several kinds of DPs obtained from a range of materials in a conventional 100-kV TEM: (A) amorphous carbon, (B) an Al single crystal, (C) polycrystalline Au, (D) Si illuminated with a convergent beam of electrons. In all cases the direct beam of electrons is responsible for the bright intensity at the center of the pattern and the scattered beams account for the spots or rings that appear around the direct beam.

decrease in the scattering cross section with increasing θ. Third, the scattering intensity varies strongly with the structure of the specimen. You'll see much more of this in Part II.

ANGLE OF SCATTER AND DISTANCES IN DPs
This relationship is different to the usual interpretation of images in which distances correspond to distances in the specimen, but it is critical to our understanding of diffraction patterns.

So far, we've only considered the amplitude/intensity of the electron wave and neglected the phase. When a wave is scattered, it will change its phase with respect to the incident wave. This is because a wave cannot change direction and remain in step with a wave that is not scattered. The phase of the scattered wave is most important in the specific topic of phase-contrast images, which have until recently been the principal form of high-resolution, atomic-level images such as shown back in Figure 1.2. We'll also come across the importance of the phase of the scattered wave when we consider the intensity of diffracted electron beams and the intensity in diffraction-contrast images. But at this stage all you need to know is that the electrons in the beam are in phase when they hit the specimen and the process of scattering, in any form, results in a loss of phase between the scattered and direct beams.

CHAPTER SUMMARY

Remember that electrons are strongly scattered because they are charged particles. This is the big difference compared to X-rays. Electrons are scattered by the electron cloud and by the nucleus of an atom. Remember X-rays are only scattered by the electron cloud. (In case you are physics oriented, a quantum-mechanical calculation does give the same distribution as the classic calculation for the Coulomb force.)

We have defined four important parameters in this chapter:

σ_{atom}	the scattering cross section of one atom
σ_{total}	the number of scattering events per unit distance traveled in the specimen
$d\sigma/d\Omega$	the differential scattering cross section of one atom
λ	the mean free path of (average distance traveled by) an electron between scattering events

Finally, a note on grammar! Should we discuss electron *scatter* or electron *scattering*? Electrons are scattered and we observe the results of this scattering (a gerund) but in fact we see the scatter (noun) of the electrons, which can be measured. However, if you've been observant you'll have noticed that we have always used *scattering* to denote the effect. Our practice is also consistent with the popular usage, which goes back to the early work of Bragg and others.

SCATTERING AND CROSS SECTIONS

Born, M and Wolf, E 1999 *Principle of Optics* 7th (yes, 7th!) Ed. Cambridge University Press New York. Perhaps *the* optics textbook in terms of classical treatments and number of editions.

Heidenreich, RD 1964 *Fundamentals of Transmission Electron Microscope* Interscience Publisher New York NY.

Jones 1992 gives a succinct introduction to scattering and Newbury (1986) gives a clear exposition on the units of cross sections. If you want to see a fuller description, read Wang (1995). If you're a glutton for punishment, the classic text is by Mott and Massey (1965) as we've already mentioned. You should realize that we've introduced you to some of the giants of electron optics, e.g., Airy, Fresnel, and Fraunhofer, who never knew about electron waves.

Jones, IP 1992 *Chemical Microanalysis Using Electron Beams* The Institute of Materials London.

Mott, NF and Massey, HSW 1965 *The Theory of Atomic Collisions* Oxford University Press Oxford.

Newbury, DE 1986 in *Principles of Analytical Electron Microscopy* p 1 Eds. DC Joy, AD Romig Jr and JI Goldstein Plenum Press New York.

Wang, ZL 1995 *Elastic and Inelastic Scattering in Electron Diffraction and Imaging* Plenum Press New York. An in-depth treatment of scattering using a much more rigorous mathematical approach than in this chapter.

OPTICS

We should have references to some of the founders of optics here, especially Abbe, Airy, Fraunhofer, and Fresnel, but we'll leave you to chase those up in the optics texts.

Fishbane, PM, Gasiorowicz, S and Thornton, ST 2004 *Physics for Scientists and Engineers* 3rd Ed. Prentice Hall Englewood Cliffs NJ.

Goodman, JW 2004 *Introduction to Fourier Optics* 3rd Ed. Roberts & Company Greenwood Village CO. An excellent source for the advanced student.

Hecht, E 2003 *Optics* 4th Ed. Addison-Wesley Reading MA. A favorite.

Klein, MV and Furtak, TE 1985 *Optics* 2nd Ed. Wiley & Sons New York NY. Not for the faint-hearted.

Smith, FG and Thomson, JH 1988 *Optics* 2nd Ed. Wiley & Sons New York.

Taylor, C 1987 *Diffraction* Adam Hilger Bristol UK.

SOME MICROANALYSIS AND MORE

Goldstein, JI, Newbury, DE, Joy, DC, Lyman, CE, Echlin, P, Lifshin, E, Sawyer, LC and Michael, JR 2003 *Scanning Electron Microscopy and X-ray Microanalysis* 3rd Ed. Kluwer New York.

Joy, DC 1995 *Monte Carlo Modeling for Electron Microscopy and Microanalysis* Oxford University Press New York.

Rhodes, R 1986 *The Making of the Atomic Bomb* Simon and Schuster New York. See p 282.

Rhodes, R 1995 *Dark Sun: The Making of the Hydrogen Bomb* Simon and Schuster New York. See p 423.

These books are well worth reading because of both the historical and the scientific content.

URLs

1) http://www.cstl.nist.gov/div837/837.02/epq/index.html

SELF-ASSESSMENT QUESTIONS

Q2.1 What is a cross section and in what units is it measured?
Q2.2 Distinguish between total, atomic, and differential cross sections.
Q2.3 Why are we interested in variations in the scattering intensity and the angular distribution of electron scattering?
Q2.4 What is the mean free path of an electron?
Q2.5 What do we mean by the term *electron beams* and why do we ask this question?
Q2.6 How is the direct beam different from or similar to the scattered beams?
Q2.7 Distinguish scatter and scattering.
Q2.8 What's the difference between forward scattering and backscattering?
Q2.9 What distinguishes elastic and inelastic scatterings?
Q2.10 Distinguish between coherent and incoherent scattering.
Q2.11 Describe what distinguishes diffraction from other kinds of scattering.
Q2.12 Distinguish between Fraunhofer and Fresnel diffractions.
Q2.13 Distinguish the angles α, β, θ, and Ω.
Q2.14 List the different ways a specimen can scatter electrons.
Q2.15 How many different ways can you control the scattering processes in the TEM?
Q2.16 How can you select electrons that have suffered a specific kind of scattering?
Q2.17 What's the fundamental difference between electron scattering and X-ray scattering?
Q2.18 What is a phasor diagram?
Q2.19 Why would you want to draw a phasor diagram in TEM?
Q2.20 How small is a small angle in the TEM and why are scattering angles in the TEM usually this small?

TEXT-SPECIFIC QUESTIONS

T2.1 Write down concise definitions of coherent, incoherent, elastic, and inelastic as we use them and link these definitions to the information in Figure 2.2.
T2.2 Explain in a paragraph the relationship between scattering cross section and atomic scattering factor, mentioning the important factors that influence them.
T2.3 Explain the link between the information in Figures 1.3 and 2.1.
T2.4 Distinguish the scattering angles θ and Ω in Figure 2.3 and the information that can be gathered within them. Relate these angles to the relevant angles in a TEM described in Figure 2.12.
T2.5 Sketch the intensity projected onto a photographic plate or viewing screen from the scattering produced by the Cu and Au specimens in Figure 2.4. The result does not look like the intensity in either a typical TEM image or DP shown in many figures throughout the book. Explain why this is so.

T2.6 Why are the backscattered electrons so few in number in Figure 2.4A and B and why do they all scatter to the one side in Figure 2.4A?

T2.7 Draw equivalent diagrams to Figure 2.5 for (a) 2 slits $d/2$ apart; (b) 2 slits $2d$ apart; (c) 5 slits d apart. What does this tell you about the effect on the scattering distribution of both the number and the spacing of the scattering centers?

T2.8 Draw a phasor diagram like Figure 2.6, but for three slits only.

T2.9 What is the relationship between Figure 2.10 and Figure 2.11?

T2.10 Make a copy of Figure 2.13. Cut out two circular holes with diameters ~ 5 and ~ 40 mm in another sheet of paper corresponding to different collection angles (β) in Figure 2.12. Superimpose the smaller circular hole on the different patterns in different positions to simulate the selection of electrons for forming images in a TEM. Note how easy it is to select electrons scattered in specific directions, but also note how many electrons are excluded when you do this. (a) What does this tell you about the advantages and disadvantages of a small selection aperture (or small detector)? Now superimpose the larger hole and note how many more electrons can be selected. (b) What does this tell you about the advantages and disadvantages of a large selection aperture (or large detector)?

3

Elastic Scattering

CHAPTER PREVIEW

Elastically scattered electrons are the major source of contrast in TEM images. They also create much of the intensity in DPs, so we need to understand what controls this process. First we'll consider elastic scattering from single, isolated atoms and then from many atoms together in the specimen. To comprehend elastic scattering we need to invoke both particle and wave characteristics of electrons.

Scattering from isolated atoms can occur either as a result of electrons interacting with the negatively charged electron cloud, which results in angular deviations of only a few degrees, or by attraction toward the positive nucleus which scatters the electrons through much larger angles, up to 180°. Such scattering can often be interpreted in terms of billiard-ball type, particle-particle collisions, cross sections, and mean free paths that we introduced in the previous chapter. We'll introduce the Rutherford differential cross section, which explains the strong dependence of high-angle elastic scattering on the atomic number (Z) of the atom. Later in the book, we'll show how to use this Z dependence to form images that reflect the chemistry of the specimen. When we treat the electrons as waves, their *coherency* becomes important. The coherency of the scattered electrons is related to their *angle* of scattering (θ). As this angle becomes larger, the degree of coherency becomes less and electrons that are Rutherford-scattered out to high angles are incoherent.

In contrast to Rutherford scattering, electrons that are scattered elastically through small angles (which we'll define as $<3°$) are coherent. The intensity of this low-angle scattering is strongly affected by the arrangement of atoms within the specimen. As we introduced in the previous chapter, such collective scattering by the atoms is referred to as *diffraction* and can only be understood if we treat the electron as a wave. Diffraction is controlled mainly by the angle of incidence of the electron beam to the atomic planes in the specimen, the spacing between these planes, and interatomic distances within the planes. So this small-angle, coherent scattering is invaluable for characterizing the crystallography of the specimen and is undoubtedly *the* most significant scattering phenomenon in the TEM.

So as we discuss elastic scattering, you'll see that we often use the wave-particle duality simultaneously, because both lines of thought are necessary for a full understanding.

3.1 PARTICLES AND WAVES

We have two different ways of looking at how an electron beam interacts with our TEM specimens. We can consider the beam as a succession of particles or as a number of waves. What we want to do is understand the relationship between the two approaches and we can summarize the two viewpoints thus:

Electrons are *particles* so they have the following properties, which we introduced in Chapter 2.

- They have a scattering cross section and a differential scattering cross section.
- They can be scattered through particular angles (remember our angles are semi-angles).
- The electrons interact with the nucleus and the electron cloud through Coulomb forces.
- We can relate this process to scattering of other particles such as α particles, so lots of analysis can carry over from other systems.

PARTICLES AND WAVES

When we discuss X-ray and electron spectrometry you'll see that we have to use a particle description. When we discuss imaging, HRTEM, and DPs you'll see that we use a wave description.

Electrons have a *wave* nature and the electron beam is almost a *plane wave*, hence:

- Waves are diffracted by atoms or 'scattering centers.'
- How strongly a wave is scattered by an atom is determined by the atomic-scattering amplitude.
- When we gather atoms together into a solid, the diffraction process gets much more complicated but it is central to TEM.
- We can relate the process to the diffraction of X-rays, so lots of analysis already exists.

3.2 MECHANISMS OF ELASTIC SCATTERING

In the previous chapter we simply stated that electrons going through a thin specimen are either scattered or not scattered and either lose energy or don't lose energy. It's now time to describe the ways in which this scattering occurs and in this chapter we'll confine our attention to elastic events, saving inelastic scattering for Chapter 4.

It's convenient to divide elastic-scattering mechanisms into two principal forms: electron scattering from single, isolated atoms and collective scattering from many atoms together within the specimen. We'll start in the same way we did in the previous chapter by looking first at the interaction of a single electron with an isolated atom. In this situation, elastic scattering can occur in one of two ways, both of which involve Coulomb forces. As shown in Figure 3.1, the electron may interact with the electron cloud, resulting in a small angular deviation. Alternatively, if an electron penetrates the electron cloud and approaches the nucleus, it will be strongly attracted and may be scattered through a larger angle that, in rare cases in the TEM, can approach 180° (i.e., complete backscattering).

ELASTIC?

You should be aware that either of these two interactions may not be truly elastic, so our separation of scattering into elastic and inelastic is a bit of a simplification.

In fact many electron-electron interactions are inelastic, as we'll see in the next chapter. We'll also see, for example, that the nuclear interaction may result in the generation of a bremsstrahlung X-ray or may even result in the displacement of the atom from its site in the crystal, both of which involve some energy loss for the electron. Indeed, the higher the angle of scattering of an electron emerging from the specimen, the greater the chance that it will have undergone an inelastic event at some time during its passage through the specimen. Despite all this, we'll continue to ignore any inelastic effects in this chapter.

The second principal form of elastic scattering occurs when the electron wave interacts with the specimen as a whole. We've already mentioned the best-known form of this interaction, namely, diffraction, which is particularly important at low-scattering angles. Understanding diffraction involves treating the electron beam as a wave, rather than as a particle as we did in Figure 3.1. Following the original approach of Huygens for the diffraction of visible light, we imagine each atom in the specimen that interacts with the incident plane wave acts as a source of secondary spherical wavelets, as illustrated in Figure 3.2.

INTERFERENCE

These wavelets reinforce one another in certain angular directions and cancel in other directions: both reinforcement and cancellation are extremely useful phenomena.

Thus, the low-angle, elastic scattering distribution is modified by the crystal structure of the specimen, and strong diffracted beams emerge at specific angles. The

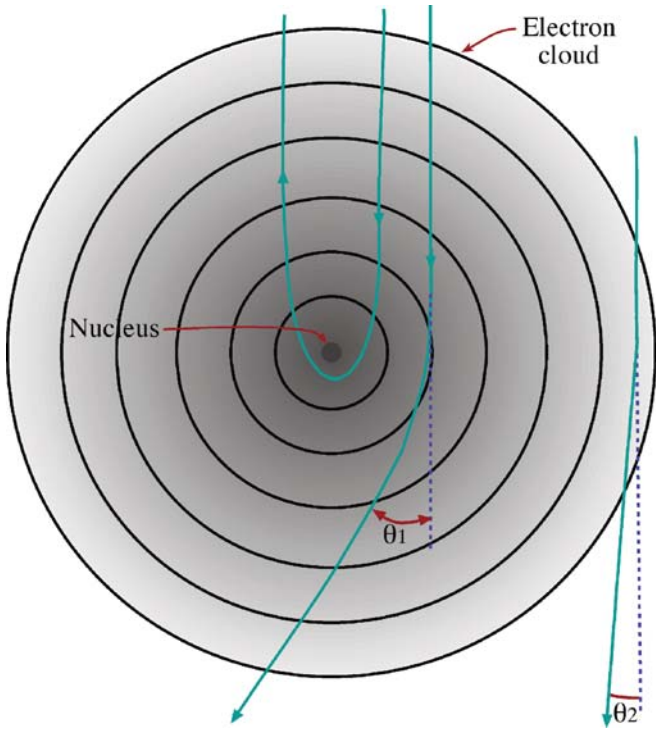


FIGURE 3.1. An isolated atom can scatter a high-energy electron by two mechanisms. Coulombic interaction within the electron cloud results in low-angle scattering; Coulombic attraction by the nucleus causes higher-angle scattering (and perhaps complete backscatter when $\theta > 90°$). The potential within the electron cloud is always positive.

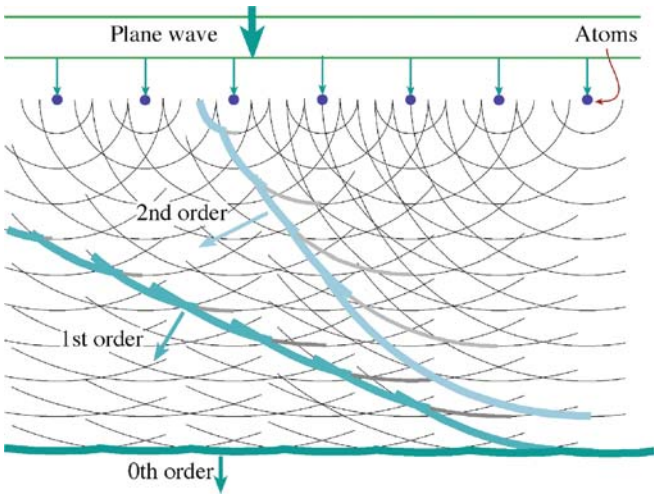


FIGURE 3.2. A plane, coherent electron wave generates secondary wavelets from a row of scattering centers (e.g., atoms in the specimen). The secondary wavelets interfere, resulting in a strong direct (zero-order) beam and several (higher order) coherent beams scattered (diffracted) at specific angles.

diffracted beam scattered through the smallest angle is called the first-order beam and we'll discuss these and higher-order effects in depth in Chapters 11 and 12. We'll now go on to examine these two forms of elastic scattering in more detail, starting with the billiard-ball approach. Then we will briefly describe the scattering of a wave to show how it relates to this particle-based treatment and later we'll use the wave approach as the basis for a full analysis of diffraction.

3.3 ELASTIC SCATTERING FROM ISOLATED ATOMS

So let's look at two possible paths for a beam electron interacting with an isolated atom as shown in Figure 3.1. Whether it interacts more strongly with the nucleus or the electrons, the electron is scattered through an angle θ.

SCATTERING ANGLE

Elastic electron-electron interactions usually result in a relatively low scattering angle, while electron-nucleus interactions cause higher-angle scattering.

If we just consider an electron, charge e, accelerated through a voltage V before being scattered from an isolated atom, the electron-electron and electron-nucleus scattering paths are hyperbolic and can be given by two simple equations (Hall 1953) which are useful because they summarize the principal factors that control elastic scattering

$$r_e = e/V\theta \tag{3.1}$$

$$r_n = Ze/V\theta \tag{3.2}$$

where r is the radius of the scattering field of the nucleus and the electron. The different cross sections for scattering through angles $> \theta$ are given by πr_n^2 for the nucleus and $Z\pi r_e^2$ for the scattering by Z electrons in the cloud. If we sum the two components and (just as back in equation 2.8) multiply by $N_0\rho t/A$ we'll get a sense of the total elastic scattering through a film of thickness t.

This approach is "defective in many respects" as Hall says but gives you a good qualitative sense of the various parameters that affect elastic scattering. You can see that the atomic number Z of the atom controls the elastic interaction with the nucleus, but the electron-electron scattering is more a function of the incident-beam energy V (which has to be in esu if the dimensions of these equations are to be correct). We'll see later in Chapter 22 that the strong effect of Z becomes important when we need to enhance scattering in low-Z materials, such as polymers and biological tissue, in order to get better TEM image contrast. Notice that when the electron passes close to the nucleus (r_n is small) the angle θ will be large. We'll see in Chapter 22 that this dependence on θ directly relates to TEM-image contrast. The electron-beam energy can also control the image contrast to some extent. So Z, V, and θ all affect image contrast and are the three major reasons why you cannot avoid having to study the physics of electron scattering.

3.4 THE RUTHERFORD CROSS SECTION

For the next three sections, we'll ignore the low-angle, electron-electron scattering and concentrate only on scattering by the nucleus. The high-angle, electron-nucleus interaction is analogous to the backscattering of α particles from a thin metal foil. The first observation of such backscattering in 1911 by H. Geiger (of *counter* fame) and a Manchester University *undergraduate*, E. Marsden, enabled their professor, Ernest Rutherford, to deduce the existence of the nucleus (never overlook undergraduate research results!). Rutherford (1911) described backscattering as "the most incredible event that has ever happened to me" (even though he'd already won the Nobel Prize for Chemistry in 1908), and he derived the following expression for the differential cross section for this high-angle scattering by the nucleus alone

$$\sigma_R(\theta) = \frac{e^4 Z^2}{16(4\pi\varepsilon_0 E_0)^2} \frac{d\Omega}{\sin^4 \frac{\theta}{2}} \tag{3.3}$$

All the terms in this equation were defined back in Chapter 2. The expression ignores relativistic effects and assumes that the incident electron does not lose

significant energy through inelastic processes, so that the energy of the electrons, E_0 (in keV), is fixed. As we've already noted, although strictly inaccurate, these assumptions can be tolerated in the TEM (at least at 100 keV or below).

3.5 MODIFICATIONS TO THE RUTHERFORD CROSS SECTION

You'll often see the Rutherford differential cross section in different, but mathematically similar, forms. For example, equation 3.3 neglects the so-called screening effect of the electron cloud. Screening can be thought of as making the nucleus appear somewhat less positive to the incident electron (although the overall charge within the electron cloud is *always* positive). So the differential cross section is effectively reduced and the amount of scattering is lowered. Screening is only important when the beam electron passes far from the nucleus and under these circumstances the scattering angle will be small (say $<\sim 3^\circ$). If we wish to account for screening, we replace the $\sin^2(\theta/2)$ term with $[\sin^2(\theta/2) + (\theta_0/2)^2]$ where θ_0 is called the screening parameter given by

$$\theta_0 = \frac{0.117Z^{1/3}}{E_0{}^{1/2}} \tag{3.4}$$

(Here E_0 is in keV.) What we are saying is that the screening parameter can be described by a particular scattering angle, θ_0. When the scattering angle is greater than θ_0 we can neglect electron-electron interactions and the electron-nucleus interaction is dominant. The value of θ_0 at 100 keV is only $\sim 2^\circ$ for Cu and less for lighter elements, so above a few degrees, all scattering can be approximated to Rutherford high-angle scattering.

As we've noted, so far all our equations are non-relativistic, which is unfortunate since relativistic effects are significant for electrons with energies $>\sim 100$ keV (which is the case for most materials investigations in the TEM). Fortunately, we can easily correct for relativity to give a more accurate cross section by using λ_R, the relativistically corrected electron wavelength (see equation 1.7), and a_0, the Bohr radius of the scattering atom

$$a_0 = \frac{h^2\varepsilon_0}{\pi m_0 e^2} \tag{3.5}$$

where ε_0 is the dielectric constant. Using the other constants listed in Table 1.1 we find a_0 is 0.0529 nm (if you're old enough you can easily remember this as 0.5 Å). The net result of adding screening and relativity corrections is that

$$\sigma_R(\theta) = \frac{Z^2\lambda_R^4}{64\pi^4 a_0^2}\frac{d\Omega}{\left[\sin^2\left(\frac{\theta}{2}\right) + \frac{\theta_0^2}{4}\right]^2} \tag{3.6}$$

This expression describes the screened, relativistic, differential Rutherford cross section. One very important effect of incorporating screening into these equations is that the cross section does not go to infinity as the scattering angle goes to zero which is an important limitation of all the simpler equations that we used initially.

The screened Rutherford cross section is the one most widely used for TEM calculations, although it has particular limitations at the highest operating voltages (300–400 kV) and for heavier elements ($Z>30$) which scatter electrons through large angles. Under these circumstances, you should use another cross section, such as that of Mott, for which you should consult the text by Mott and Massey (which we referred to in Chapter 2) or Newbury (1986).

So, as we did for the basic cross section back in Chapter 2, we can integrate this expression to obtain the total cross section over specific angular ranges. We can substitute appropriate values for the various constants and integrate the differential cross section from θ to π to obtain the total nuclear cross section (in scattering events/electron/atom/m^2) for electrons elastically scattered into angles $> \theta$

$$\sigma_{\text{nucleus}} = 1.62 \times 10^{-24}\left(\frac{Z}{E_0}\right)^2 \cot^2\frac{\theta}{2} \tag{3.7}$$

(From what we told you in Chapter 2 you ought to be able to work out the integration necessary to determine the probability of electrons being scattered into angles $<\theta$.) Again we see that the beam energy (E_0), the angle of scattering (θ), and the atomic number (Z) all affect the probability that an electron will be scattered by the nuclei of atoms in the specimen. If you simplify this last expression by assuming that θ is small you should be able to see some parallels with Hall's less accurate equation 3.2 for nuclear scattering. However, there's much more to this whole scattering process than we have covered here and you should read Newbury (1986) and Jones (1992) for further discussion of these calculations once you really appreciate their significance.

The best way to summarize the characteristics of cross sections is to present some data. Figure 3.3 shows the variation of the screened Rutherford cross section in equation 3.7 with scattering angle for (a) three different elements and (b) two different beam energies. As you can see for Cu, the cross section decreases by several orders of magnitude from $\sim 10^{-22}$ to $\sim 10^{-28}$ m^2 as the scattering angle increases from 0 to 180°; so, as we've already told you, scattering is most likely to occur in the forward ($\theta \sim 0^\circ$) direction and drops off rapidly

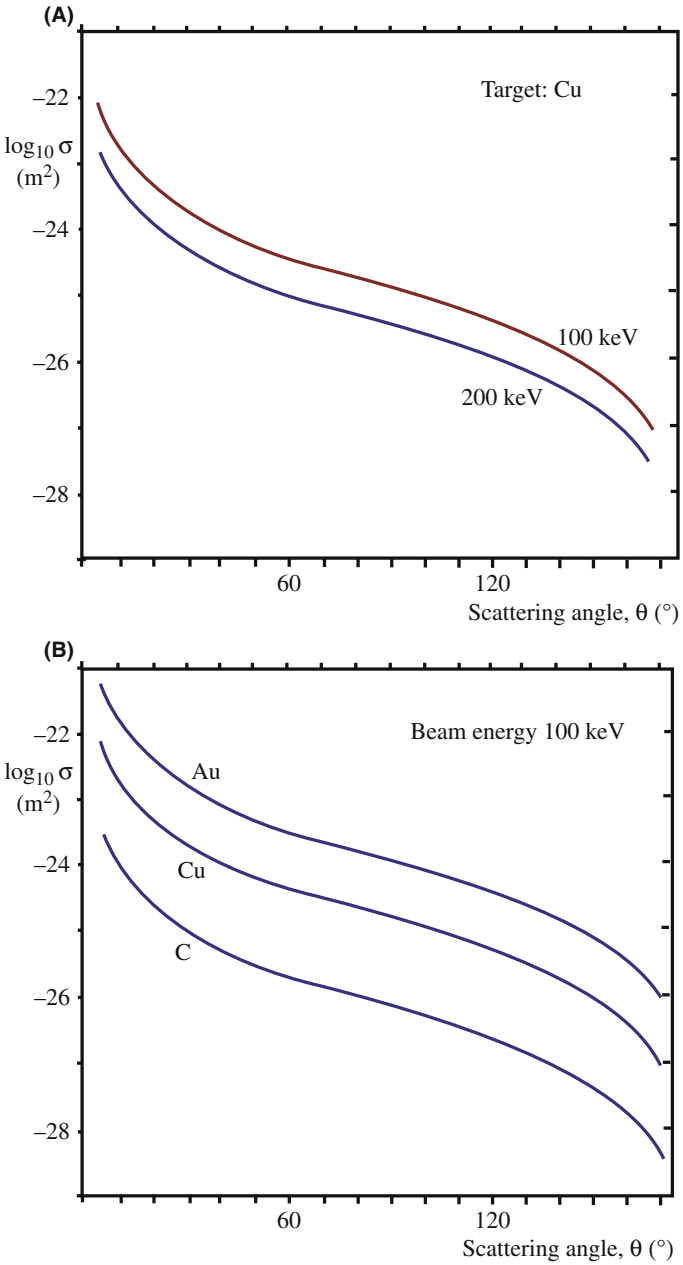


FIGURE 3.3. The variation of the logarithm of the screened relativistic Rutherford cross section with scattering angle from equation 3.7, describing the change in cross section for electrons scattered at angles $> \theta$ (A) for different elements at 100 keV and (B) for scattering from Cu at different accelerating voltages.

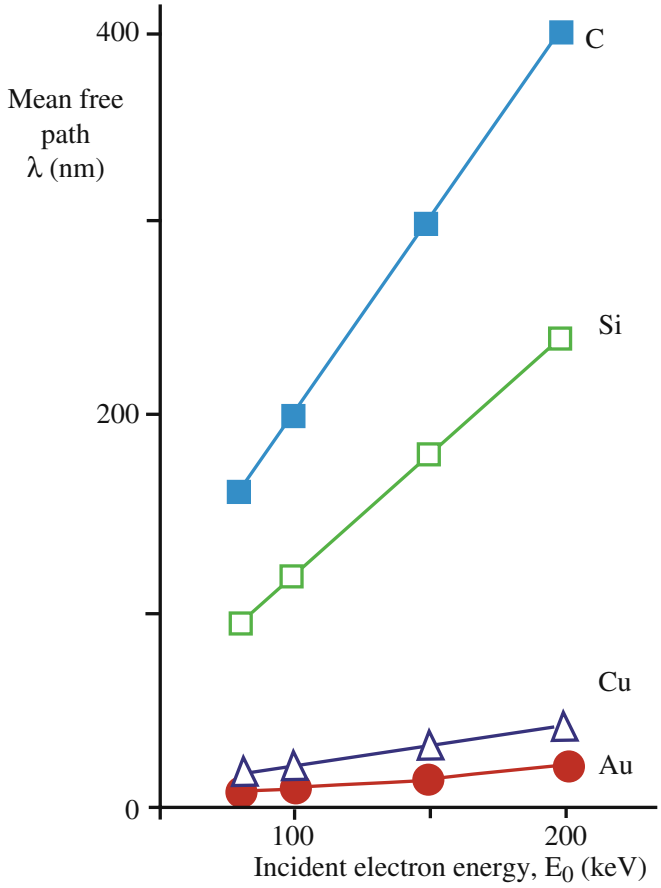


FIGURE 3.4. The variation of the mean free paths of elastic scattering for four different elements as a function of the beam energy, calculated assuming a screened, relativistic Rutherford cross section.

above a few degrees. Increasing Z from carbon to gold can increase the cross section by a factor of ~100, which is why you need proportionately thinner TEM specimens if you want to 'see' through higher-Z materials. Doubling the electron-beam energy can lower the cross section by a factor of two or three, which is why higher-energy electrons are less likely to be scattered by your specimen than lower-energy electrons, all else being equal. Figure 3.4 plots the related mean free paths for elastic scattering. From this graph you can see that very few high-angle elastic scattering events will occur if you can make your specimen < 100 nm in thickness. Within such specimens, most electrons either undergo a single-scattering event or are not scattered and we'll assume that this simplification is a viable approximation to what's actually going on in the microscope many times throughout this text. This approximation is the main reason why, as we've already noted that in almost all TEM studies, the 'thinner is better' criterion applies.

3.6 COHERENCY OF THE RUTHERFORD-SCATTERED ELECTRONS

Up to now, in this chapter, we've treated the electron as a particle, but there is useful insight to be gained if we examine the wave nature of the scattered electron. High-angle Rutherford-scattered electrons are *incoherent*: i.e., there is no phase relationship between them. This is a tricky concept because we are scattering particles. Such incoherent scattering is important in two respects. First, the high-angle, forward scattering can be used to form exceptionally high-resolution

images of a crystalline specimen in which the image contrast is due solely to the value of Z, not the orientation of the specimen (as is the case for low-angle coherent diffraction). Such *Z-contrast images,* as we'll see in Chapter 22, provide qualitative atomic-resolution elemental analysis, in addition to showing atomic-resolution detail at interfaces between regions of different Z. Compared to other image-contrast mechanisms, Z-contrast imaging is a relatively new technique for most microscopists but, particularly since the availability of C_s-correctors, it has consistently held the record for the highest-resolution images and analysis in the TEM (e.g., Varela et al. 2005) and is already beginning to revolutionize our understanding of the atomic-level structure and chemistry of crystal defects.

COHERENCY

Coherency of the scattered electron is a wave property. If the scattered electron waves have a phase relationship they must be coherent.

Second (but much less important), the high-angle backscattered electrons (BSEs) can be used to form images of the beam-entrance surface of the specimen, in which the contrast is not only due to differences in Z, but also to changes in surface topography of the specimen. BSE images are rarely used in the TEM because the BSE signal is small. If you go back and look at the Monte Carlo simulation in Figure 2.4 you'll see that out of 10^3 incident electrons in Cu only about three (0.3%) were backscattered. Therefore, the quality of this signal is very poor, the images are noisy, and the contrast is low. The contrast is much better for bulk specimens in an SEM in which many more electrons are backscattered (e.g., about 30% in Cu) and BSEs provide a stable, high-contrast imaging technique in SEMs in which you can discriminate between the signals from adjacent elements in the periodic table. In principle, Z-contrast should be able to do the same in the TEM.

3.7 THE ATOMIC-SCATTERING FACTOR

The classic Rutherford differential cross section cannot be used to calculate the cross section exactly, because it ignores the wave nature of the electron beam. A full treatment involves wave mechanics and is well beyond the scope of this text. Perhaps the most familiar aspect of the wave approach to cross sections is the concept of the atomic-scattering factor $f(\theta)$, which is related to the differential elastic cross section by a simple equation (more on this in Section 3.8)

$$|f(\theta)|^2 = \frac{d\sigma(\theta)}{d\Omega} \tag{3.8}$$

What we will now do is to highlight some of the important features that lead to this result by outlining the basic arguments.

- $f(\theta)$ is a measure of the amplitude of an electron wave scattered from an isolated atom.
- $|f(\theta)|^2$ is proportional to the scattered intensity.

From these two statements and given the importance of scattered-electron intensity in images and DPs, you can appreciate why $f(\theta)$ is such an important parameter in TEM.

The scattering-factor approach is complementary to the Rutherford differential cross section analysis, because it is most useful for describing the low-angle (i.e., $<\sim 3^\circ$) elastic scattering where the Rutherford model is inappropriate. Usually, $f(\theta)$ is defined in the following manner

$$f(\theta) = \frac{\left(1 + \frac{E_0}{m_0 c^2}\right)}{8\pi^2 a_0} \left(\frac{\lambda}{\sin\frac{\theta}{2}}\right)^2 (Z - f_x) \tag{3.9}$$

All the terms have been previously defined (note that we've dropped the screening term, so remember what this implies). If you need a more detailed approach you could consult the physics-based text by Reimer. Because we're now thinking in terms of waves, we need the wavelength λ (controlled of course by the beam energy E_0), and f_x is the scattering factor for X-rays, which is well known. The most widely referenced source of electron-scattering factors for TEM was usually the classic work of Doyle and Turner (1968), but you can now find values in software packages (see Section 1.6) and you can even do your own calculations using the free software on the NIST database (also in Section 1.6 and URL #1). The appearance of f_x in equation 3.9 is a reminder that $f(\theta)$ is a fundamental result of the wave nature of the electron.

***f*(θ)**

The atomic-scattering factor $f(\theta)$ depends on λ, θ, and Z.

We can plot this angular variation for a single isolated atom. Figure 3.5 summarizes graphically what we already know about the magnitude of elastic scattering (see equations 3.1 and 3.2)

- It decreases as θ increases ($\theta = 0^\circ$ for the incident-beam direction).
- It decreases as λ decreases (i.e., as the accelerating voltage (V) increases).
- It increases with Z for any value of θ.

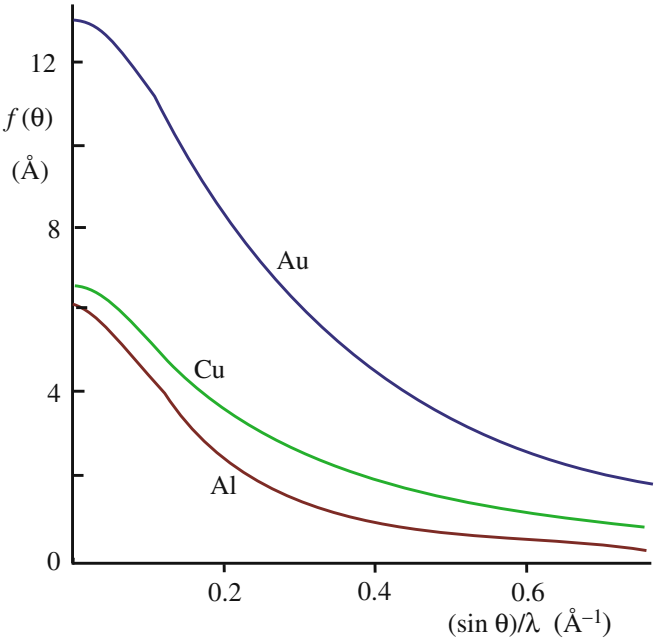


FIGURE 3.5. Change in the atomic scattering factor $f(\theta)$ with scattering angle θ (calculated from equation 3.9) showing that elastic scattering decreases with angle away from the incident beam direction ($\theta = 0°$) and increases with Z.

This expression (equation 3.9) for $f(\theta)$ contains components of both elastic nuclear scattering (the Z term) and elastic electron-cloud scattering (the f_x term). We'll see later in the chapters on diffraction in Part 2 that the $f(\theta)$ approach is used exclusively and, if we neglect the f_x term, then it can be shown that $|f(\theta)|^2$ is mathematically equivalent to the high-angle Rutherford differential cross section, as we defined it in equation 3.6. So now we've tied together the particle and wave approaches to elastic scattering.

ANGLE VARIATION

The important point to remember is that both the differential cross section and the scattering factor are simply measures of how the electron-scattering intensity varies with θ.

3.8 THE ORIGIN OF $f(\theta)$

Since $f(\theta)$ relates to the amplitude of a scattered wave, we'll consider briefly how it arises. The following analysis is not intended to be completely rigorous, but only to give the fundamental ideas behind the meaning of $f(\theta)$ and its relation to the differential scattering cross section. You can safely delay studying this topic until curiosity wins, then you can go and read the really thick physics textbooks.

To find the total elastic-scattering cross section, we have to integrate $d\sigma/d\Omega$. Note that this is a particle model, but you should also be aware of how the wave nature of the electrons is brought in. We can consider the wave nature by looking at Figure 3.6 (which you should realize is closely related to Figures 2.3 and 2.12).

We can describe the incident beam as a wave of amplitude ψ_0 and phase $2\pi kr$

$$\psi = \psi_0 e^{2\pi i k r} \tag{3.10}$$

In this definition of phase, k is the magnitude of the wave vector and r is the distance that the wave has propagated, as we'll discuss in detail later in Chapter 11. When the incident plane wave is scattered by a point charge, a spherical scattered wave is created which has a different amplitude ψ_{sc} but keeps the same phase apart from a $\pi/2$ addition which we return to in a moment

$$\psi_{sc} = \psi_0 f(\theta) \frac{e^{2\pi i k r}}{r} \tag{3.11}$$

In this equation, $f(\theta)$ is the amplitude we would have if $\psi_0 = 1$, i.e., it is the *atomic-scattering amplitude*.

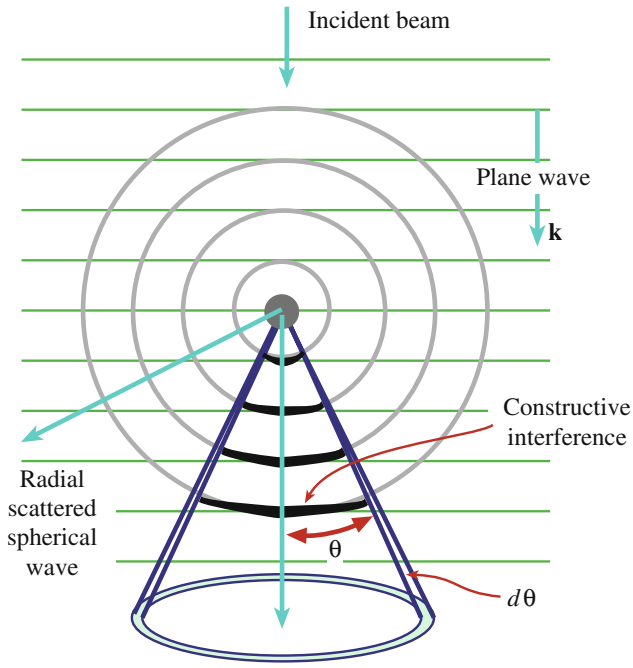


FIGURE 3.6. Generating a scattered wave by the interaction of a plane wave (horizontal line, wavelength λ) with a point charge. The circles represent the scattered spherical wavefronts which are in phase and retain the original λ. The in-phase, constructive interference between the plane and spherical waves is shown by the dark areas. The angles θ and $d\theta$ are the same as in Figure 2.3.

So obviously we need to know $f(\theta)$ and an acceptable model is essential to make the problem manageable. Up to this point, our treatment has been quite rigorous and ideally, the model would distinguish between a neutral atom in a metal, a covalently bonded atom, and an ion. If you're desperate, the quantity $f(\theta)$ can always, in principle, be calculated from the Schrödinger equation. In practice, however, we usually use a simple approximation which we'll now describe.

If we write down the expression for the scattering process shown in Figure 3.6, then we have

$$\psi_{sc} = \psi_0 \left[e^{2\pi i \mathbf{k}_I \cdot \mathbf{r}} + i f(\theta) \frac{e^{2\pi i k r}}{r} \right] \qquad (3.12)$$

You should note first of all that, as usual for Huygens wavelets, there is a 90° phase shift (shown by the inclusion of 'i' in the second term) between the incident and scattered beams and secondly, that $f(\theta)$ can be expressed as

$$f(\theta) = |f(\theta)| e^{i\eta(\theta)} = |f(\theta)|(\cos\eta(\theta) + i \sin\eta(\theta)) \qquad (3.13)$$

which means that the phase, $\eta(\theta)$, of $f(\theta)$ also depends on the angle of scatter, θ.

First aside: In writing equation 3.12, we have introduced two wave-propagation parameters: the vector $\mathbf{k}_I$ for the incident plane wave and the scalar k for the spherical scattered wavelet. By writing the 2π factor separately as part of the phase term, we have implicitly defined k to be $1/\lambda$. Many physics textbooks include the 2π in k so they have k given by $2\pi/\lambda$. Just be careful when you compare similar formulas in two textbooks.

1/λ AND 2π/λ

Sometimes $k = 1/\lambda$ and sometimes $k = 2\pi/\lambda$ and it's sometimes difficult to find out which definition is being used.

Second aside: The 90° phase change for the scattered-wave component in equation 3.13 can be easily understood if you consider the following. If the amplitude of the wave is initially $\psi_0 \sin(2\pi kr)$ then, after it has passed through the specimen, it will be ψ_{tot}. After scattering, the phase is increased by ϕ, so we can express the new ψ_{tot} as

$$\psi_{tot} = \psi_0 \sin(2\pi kz + \phi) = \psi_0 \sin(2\pi kz)\cos\phi + \psi_0 \cos(2\pi kz)\sin\phi \qquad (3.14)$$

Now if ϕ is small, then $\cos\phi \approx 1$ and $\sin\phi \approx \phi$; $\cos\theta$ is always the same as $\sin(\theta + \pi/2)$, hence

$$\psi_{tot} = \psi_0 \sin(2\pi kz) + \psi_0 \phi \sin\left(2\pi kz + \frac{\pi}{2}\right) \qquad (3.15)$$

The $\pi/2$ term would arise if we used the exponential rather than the sine to denote phase, so we can now write equation 3.15 as

$$\psi_{tot} = \psi + i\psi_{sc} \qquad (3.16)$$

This equation has the same form as that given in equation 3.12.

3.9 THE STRUCTURE FACTOR *F*(θ)

The next introductory step in discussing electron scattering is to take the idea of individual atoms scattering electrons (i.e., $f(\theta)$), which we've just discussed in some detail, and consider what happens when the atoms are stacked together regularly in a crystal structure. (We can, in principle, also do this for an amorphous solid but we'll stick to crystals for simplicity.) We will deal with this approach in great detail in Chapter 13, but for now we can introduce the structure factor $F(\theta)$, which is a measure of the amplitude scattered by a unit cell of a crystal structure. Because $F(\theta)$ is an amplitude like $f(\theta)$, it also has dimensions of length. We can define $F(\theta)$ as the sum of the $f(\theta)$ terms from all the i atoms in the unit cell (with atomic coordinates $x_i\, y_i\, z_i$) multiplied by a phase factor. The phase factor takes account of the difference in phase between waves scattered from atoms on different but parallel atomic planes with the same Miller indices (hkl). The scattering angle θ is the angle between the incident and scattered electron beams. So we can write

$$F(\theta) = \sum_i^{\infty} f_i e^{2\pi i(hx_i + ky_i + lz_i)} \qquad (3.17)$$

The amplitude (and hence its square, the intensity) of scattering is influenced by the *type* of atom ($f(\theta)$), the *position* of the atom in the cell (x,y,z), and the specific *atomic planes* (hkl) that make up the crystal structure. None of this is very surprising, but it turns out that this equation predicts that in certain circumstances the amplitude of scattering is zero. This behavior is intrinsic to the scattering process, is implicit back in Figure 3.2, and is often a very useful diagnostic test when determining crystal structures in the TEM.

ZERO SCATTERING

Under specific conditions, electrons scattering in a crystal may result in ZERO scattered intensity. Why might this occur?

We'll return to this point in Chapter 13 in much more detail.

3.10 SIMPLE DIFFRACTION CONCEPTS

As we mentioned earlier, electron diffraction is by far the most important scattering phenomenon in the TEM. The reason for this importance, as we'll show you in Chapters 11 and 12, is that we can use diffraction to determine the spacing of planes in crystals and, as you'll see later in Chapters 20, 21, and in the companion text, there is a whole field termed *electron crystallography* which gives an unprecedented amount of crystallographic information from space-group symmetry data right down to the dimensions of single unit cells. At the most basic level, the interplanar spacings in different crystal structures are characteristic of that structure.

We'll see that the *positions* of the diffracted beams of electrons are determined by the size and shape of the unit cell and the *intensities* of the diffracted beams are governed by the distribution, number, and types of atoms in the specimen. We'll also show you in Part 3 how diffraction leads to contrast in TEM images which is controlled by the orientation of a crystal with respect to the electron beam and which you can control simply by tilting your specimen.

DP PLUS IMAGE

We can distinguish different crystal structures by observing and measuring DPs. The combination of the DP and the electron image(s) is a most powerful tool for characterizing crystals and particularly their defects.

It's easy to see, in a qualitative manner, how diffraction modifies the distribution of the low-angle scattering, described by $f(\theta)$, and shown for a single atom in Figure 3.5. When we consider the effect of the arrangement of atoms in the specimen, then Figure 3.5 has to be modified. For an amorphous specimen, the atoms are almost (but not quite) randomly arranged. A random arrangement would result in a plot similar to Figure 3.5, but there are certain interatomic spacings that tend to occur in an amorphous structure (e.g., first- and second-nearest neighbor spacings are usually relatively well defined). As a result, the amplitude (and hence the intensity) of diffraction is stronger at some angles than at others, so we see diffuse, bright rings on the TEM screen. If the specimen is crystalline, then the intensity of the diffracted beams is a maximum at specific angles because the interplanar spacings are very well defined. The variation of $f(\theta)$ with θ plotted in Figure 3.7A and B is equivalent to the radial intensity variation across the DPs in Figure 2.13A and C, respectively, and thus emphasizes the strong relationship between $f(\theta)$ and diffracted intensity. We'll describe this important relationship mathematically in Section 3.10.B below.

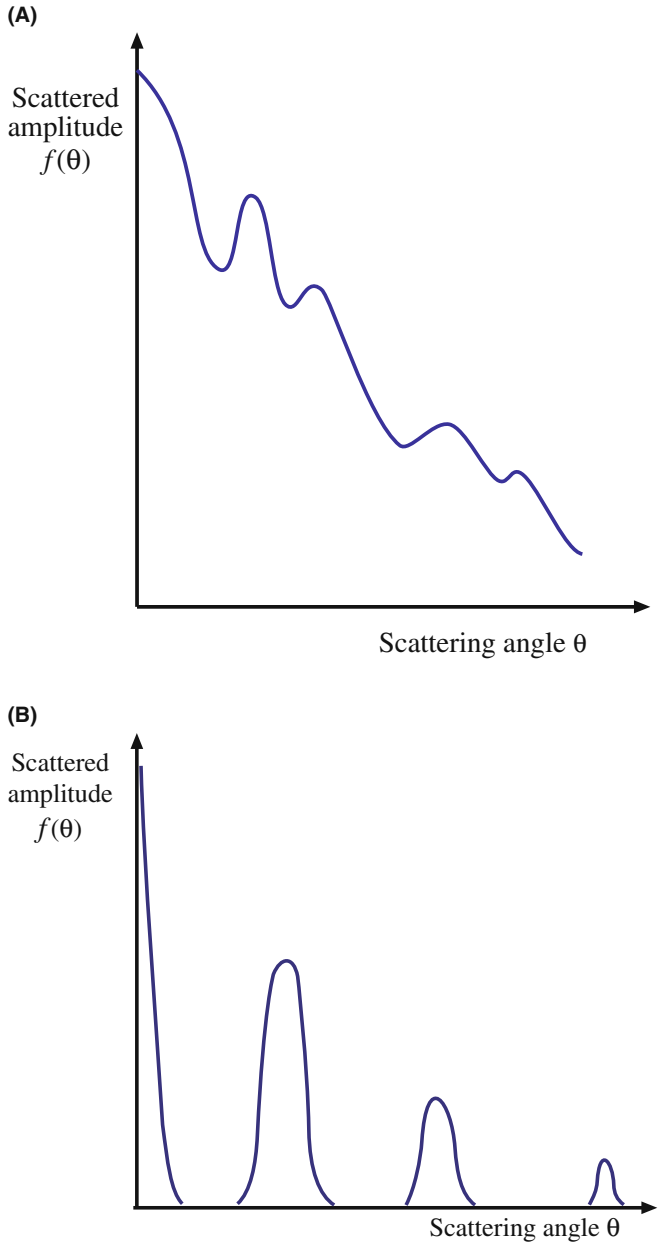


FIGURE 3.7. Change in $f(\theta)$ with θ for (A) an amorphous specimen and (B) a crystalline specimen. The amplitude (and therefore the intensity) of scattering generally decreases with increasing θ but the smooth decrease is modified at certain scattering angles (compare these curves with the intensity variation along a radius of the DPs in Figure 2.13A and C, respectively).

3.10.A Interference of Electron Waves; Creation of the Direct and Diffracted Beams

To interpret low-angle elastic scattering (which is primarily from the electron cloud) it is best to think in

terms of electron waves and not in terms of particle-particle interactions that characterize high-angle Rutherford scattering. If you go back and look at Figure 3.2 you see a periodic one-dimensional array of scattering centers (slits), and a monochromatic wave (i.e., fixed λ) is advancing toward these centers. Each center acts as a new source of a wave of the same λ. Thus many new waves are created and, when more than one wave is present, the waves can interfere with one another. This process happens from even the thinnest specimens and is entirely a wave phenomenon that doesn't need concepts such as cross section, which we apply when we think of the electron as a particle.

A rule of wave theory is that waves reinforce one another (this is constructive interference) when they are in phase. Waves also cancel one another (destructive interference) when they are out of phase. What you see in Figure 3.2 is that the diffracted waves are in phase with one another only in certain directions. There is invariably a *zero-order wave* that proceeds in the same direction as the incident wave, which in the TEM we'll refer to as the direct beam of electrons, as we defined at the start of Chapter 2. There are also *higher-order waves* that propagate in forward directions that are at some fixed (but very small) angle to the incident wave and we'll call these the diffracted beams.

So diffraction creates many electron beams traveling at specific angles relative to a single monochromatic incident beam. In the chapters on diffraction in Part 2, we'll find ways to measure these angles and relate them to the spacing of the scattering planes.

DIRECT AND DIFFRACTED

The direct beam consists of electron, scattered in the same direction as the incident beam. Often in TEM terminology these electrons are called the transmitted beam but this term is ambiguous since, in fact, all forward-scattered beams are 'transmitted' through the specimen.

3.10.B Diffraction Equations

Here we'll introduce the mathematical relationships that describe the diffraction process. The idea of using diffraction to probe the atomic structure of materials was credited to von Laue (1913) in Germany, although others such as Ewald were working on similar ideas at the same time. Von Laue's crucial idea was that much shorter electromagnetic rays than light would cause diffraction or interference phenomena in a crystal. Although his colleague Sommerfeld, with whom he discussed the idea while skiing, disagreed, Friedrich, one of Sommerfeld's assistants, and Knipping tested the idea experimentally by irradiating a copper sulfate crystal and became the first to observe diffraction from crystal planes. In fact it was a remarkable stroke of luck that the $CuSO_4$ diffracted the X-rays at all because of the strict equations that govern diffraction.

Von Laue used the well-known light-optics approach to argue that the diffracted waves are in phase if the path difference between waves scattered by adjacent scattering centers is a whole number of wavelengths, $h\lambda$ (h is an integer). Thus, as shown in Figure 3.8, if the scattering centers (B and C) are spaced some distance a apart and the incident beam (wavelength λ) makes an angle θ_1 with the line connecting the scattering centers and is diffracted at an angle θ_2, then the path difference AB – CD is

$$a(\cos\theta_1 - \cos\theta_2) = h\lambda \tag{3.18}$$

Now in three dimensions, two more Laue equations can be written for two more distances b and c and appropriate angles θ_n

$$b(\cos\theta_3 - \cos\theta_4) = k\lambda \tag{3.19}$$

$$c(\cos\theta_5 - \cos\theta_6) = l\lambda \tag{3.20}$$

These three simultaneous equations bear von Laue's name and for his original suggestion and the analysis of the experiments of Sommerfeld's students, he received the Nobel Prize in Physics in 1914 (nice work: three equations). We'll show in Chapter 11 that

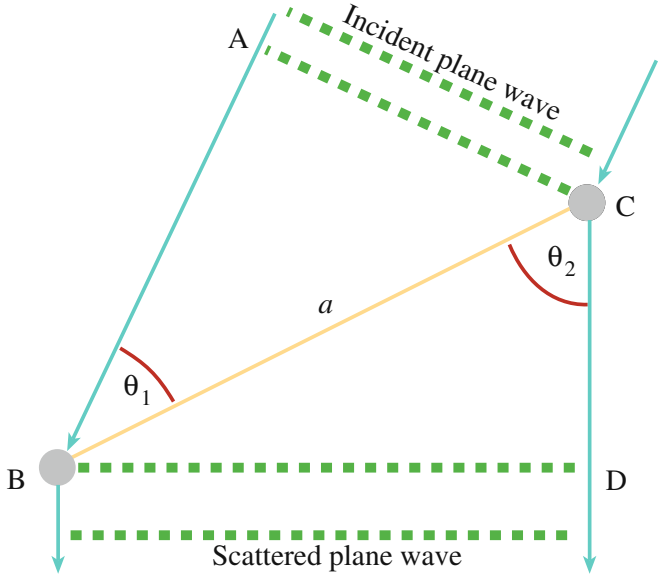


FIGURE 3.8. The approach used by von Laue to calculate the path difference for a wave (wavelength λ). In this one-dimensional figure the wave is incident at an angle θ_1 and scattered at an angle θ_2 from two atoms (B and C) spaced distance a apart. The path difference between scattered waves is AB – CD.

in a TEM specimen, when all three Laue equations are satisfied simultaneously a diffracted beam is produced. We'll also show you in Chapters 11 and 12 that the letters *hkl* are the indices of the diffracted beam and are equivalent to the Miller indices (*hkl*) of the diffracting crystal plane (or some multiple thereof).

Usually in TEM, we use a simpler approach to describe diffraction. Von Laue's approach was simplified by the family team of Sir William H. (obviously the dad) and Mr. W. Lawrence Bragg (the son) in England who proposed (Bragg and Bragg 1913) that the waves behaved as if they were reflected off atomic planes as shown in Figure 3.9.

In parallel with von Laue's optical approach, the Braggs argued that waves reflected off adjacent scattering centers must have a path difference equal to an integral number of wavelengths, if they are to remain in phase. So, in the TEM the path difference between electron

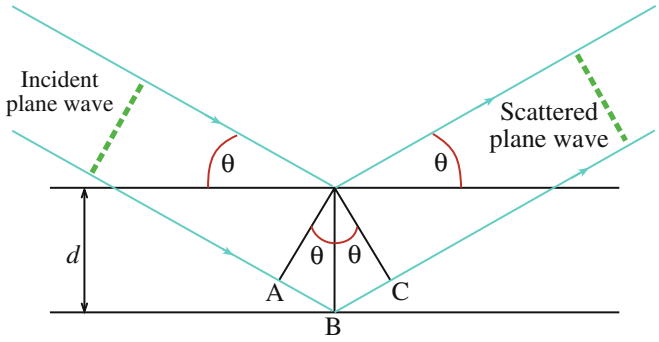


FIGURE 3.9. The Bragg description of diffraction in terms of the reflection of a plane wave (wavelength λ) incident at an angle θ to atomic planes of spacing *d*. The path difference between reflected waves is AB + BC.

waves reflected from the upper and lower planes in Figure 3.9 is (AB + BC). Thus, if the 'reflecting' *hkl* planes are spaced a distance *d* apart and the wave is incident and reflected at an angle θ_B, both AB and BC are equal to $d \sin \theta_B$ and the total path difference is $2d \sin \theta_B$. So we can write what is known as Bragg's (although grammatically and historically it should be Braggs') law

$$n\lambda = 2d \sin \theta_B \tag{3.21}$$

We'll reserve θ_B for the Bragg angle, which is *the* most important scattering angle (remember we really mean *semi*-angle) in TEM and you'll come across it many more times in this text. The Braggs also received a Nobel Prize in Physics a year after von Laue but this time for only one equation (even nicer work!) and despite the fact that the idea of reflected electrons, while mathematically correct, is physically wrong. We'll continue to use the term Bragg reflection to describe diffraction in the TEM because everyone does so, even though it's inaccurate, and because it is extremely useful. However, we'll demonstrate to you, in a rigorous fashion, the mathematical equivalence of the Bragg and von Laue approaches in Chapter 12.

It is simple to see from the Bragg equation that atomic planes which are closer together give rise to larger angles of scatter. This reciprocal relationship (*d* is proportional to $1/\theta$; see Chapter 12) is very important in diffraction-pattern interpretation. So, if you know λ for the incident electron (which you control by choosing the accelerating voltage) and you can measure θ experimentally, you can work out the interplanar spacings in your specimen. It is this crystallographic information that makes diffraction such an important aspect of the TEM.

CHAPTER SUMMARY

What should you remember from this chapter? Until you have time to study this material very carefully you may find it difficult, so here are a few suggestions:

- Know the words! In particular, we can describe the scattering process by three parameters

  $\sigma(\theta)$ the scattering cross section

  $\dfrac{d\sigma(\theta)}{d\Omega}$ the differential scattering cross section

  $f(\theta)$ the atomic-scattering amplitude

In particular, don't be put off because 'differential scattering cross section' sounds difficult. All three terms are *very* important in different parts of TEM.

- The relationships between $f(\theta)$ and $\sigma(\theta)$ are very important (as a principle, but not much used in practice).

- The relationship between $f(\theta)$ and the intensity in a DP is very important.

Remember that, although we often write $\sigma(\theta)$ as σ, there is an angle involved in any σ

- The fact that the electron is a charged particle is critical to the whole scattering process.
- The strength of the scattering, $f(\theta)$, depends inversely on the scattering angle, θ.

Yes, a really rigorous treatment of scattering would take into account the wave nature of the electron (wave mechanics), relativity, and the electron charge, all at the same time. Because we're good guys we won't inflict this on you or ourselves. Fortunately, if required, we can do very well using compiled tables of cross sections and scattering data, which are available on the web (e.g., URL #1).

We can describe the effect of the crystal structure on the electron scattering by one more parameter, the structure factor $F(\theta)$

- $F(\theta)$ is a measure of the amplitude scattered by a unit cell and $|F(\theta)|^2$ is proportional to the scattered intensity.

The diffraction process from a TEM specimen is usually described by the Bragg equation which tells us the important reciprocal relationship between atomic-plane spacings and scattering angles.

A final point to think about: remember that $f(\theta)$ is the property of a 'scattering center.' We usually think of this center as being an atom. What happens if the scattering center is an ion (i.e., if it is charged)? Is the scattering process affected by how this atom is bonded to its neighbors? What changes if the atom has a covalent rather than a metallic bond? These are important questions (otherwise we wouldn't ask them) and we'll teach you the answers as we go on.

DIFFRACTION AND SCATTERING

Andrews, KW, Dyson, DJ and Keown, SR 1967 *Electron Diffraction Patterns* The Institute of Physics Bristol UK. The original text focused on diffraction in the TEM.

Bragg, WH and Bragg, WL 1913 *The Reflection of X-rays by Crystals* Proc. Roy. Soc. Lond. **A88** 428–438. The paper that led to the Nobel Prize.

Doyle, PA and Turner, PS 1968 *Relativistic Hartree-Fock X-ray and Electron Scattering Factors* Acta Crystallogr. **A24** 390–397.

Mott, NF and Massey, HSW 1965 *The Theory of Atomic Collisions* Oxford University Press New York.

Reimer, L 1997 *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis* 4th Ed. Springer New York. Rigorous thorough treatment of the scattering process, especially as used in Section 3.8.

Rutherford, E 1911 *The Scattering of* α *and* β *Particles by Matter and the Structure of the Atom* Phil. Mag. **21** 669–688. His scattering.

von Laue, M 1913 *Kritische Bemerkungen zu den Deutungen der Photoframme von Friedrich und Knipping* Phys. Z. **14** 421–423. Paper that led to the Nobel Prize.

Wang, ZL 1995 *Elastic and Inelastic Scattering in Electron Diffraction and Imaging* Plenum Press New York. Much more detailed than the approach used here.

SCATTERING APPLIED TO EM

Hall, CE 1953 *Introduction to Electron Microscopy* p 229 McGraw-Hill New York.

Jones, IP 1992. *Chemical Microanalysis Using Electron Beams* Institute of Materials London.

Newbury, DE 1986 *Electron Beam-Specimen Interactions in the Analytical Electron Microscope* in *Principles of Analytical Electron Microscopy* p 1 Eds. DC Joy, AD Romig Jr and JI Goldstein Plenum Press New York.

Varela, M, Lupini, AR, van Benthem, K, Borisevich, AY, Chisholm, MF, Shibata, N, Abe, E and Pennycook, SJ 2005 *Materials Characterization in the Aberration-Corrected Scanning Transmission Electron Microscope* Annu. Rev. Mat. Sci. **35** 539–569. Review of Z-contrast imaging.

URLs

(1) www.nist.gov/srd/nist64.htm NIST Standard reference database #64 provides values of the differential elastic-scattering cross sections, total elastic-scattering cross sections, phase shifts, and transport cross sections for elements with $Z = 1$ to 96 and for beam energies from 50 eV to 300 keV (in steps of 1 eV).

SELF-ASSESSMENT QUESTIONS

Q3.1 What are the primary causes of elastic scattering?
Q3.2 What do we mean by the term 'wave-particle duality'?
Q3.3 What forces act on an electron as it interacts with atoms?
Q3.4 What term describes the strength of the scattering process?
Q3.5 What factors control the interference between waves?
Q3.6 How is the scattering amplitude related to the intensity of the scattered beams that we see in the microscope?
Q3.7 What are the two principal forms of elastic scatter?
Q3.8 Relate the general form of the Rutherford differential cross section (equation 3.3) to the equation describing the cross section for nucleus scattering proposed by Hall (equation 3.2).
Q3.9 What is a screening parameter and why do we need to incorporate it in the equations that describe scattering?
Q3.10 Why is it important to include a screening parameter in the Rutherford cross section?
Q3.11 Why do elastic electron-electron interactions usually result in a relatively low scattering angle, while elastic electron-nucleus interactions cause higher-angle scattering?
Q3.12 From your answer to the previous question describe the different information that might be contained in low-angle and high-angle scattered electrons and how you might obtain that information
Q3.13 How thin should your specimen be so that scattering within it approaches the ideal of a single event per electron?
Q3.14 What is the relationship between the atomic scattering factor $f(\theta)$ and the structure factor $F(\theta)$?
Q3.15 Why do crystalline and amorphous specimens give rise to different scattering distributions?
Q3.16 What are the fundamental differences between the von Laue and Bragg approaches to diffraction and what are the similarities?
Q3.17 Put some reasonable values for d and λ into equation 3.21 and calculate a typical Bragg angle in a TEM.
Q3.18 Why is the Bragg approach fundamentally incorrect?
Q3.19 What do we mean by the term 'scattering center'?
Q3.20 What is the relationship between the spacing of the lattice planes and the angle of scatter?

TEXT-SPECIFIC QUESTIONS

T3.1 In Figure 3.1, why are the electrons interacting with both the nucleus and the electron cloud shown to deviate in the same directions (i.e., both are bent through an angle θ) when the nucleus and the electron clouds in fact have opposite electrical charges?
T3.2 Why do we show the electron close to the nucleus in Figure 3.1 as being turned around rather than being pulled directly into the (highly positively charged) nucleus?
T3.3 Look again at Figure 3.1 and explain why elastic electron-electron interactions usually result in a relatively low scattering angle, while elastic electron-nucleus interactions cause higher-angle scattering.
T3.4 In Figure 3.2 why don't we see a third-order scattered beam?
T3.5 Relate Figure 3.3 to equations 3.1 and 3.2.
T3.6 Can you show that the data in Figures 3.3 and 3.4 are consistent? (Hint: assume that θ is small (i.e., $\sim 0^\circ$) for elastic scatter.)
T3.7 Relate Figure 3.5 to equations 3.1 and 3.2.
T3.8 Is Figure 3.5 plotted for a screened or unscreened atomic potential? Explain your answer.
T3.9 In Figure 3.6 why don't we show constructive interference of waves going back in the direction of the incident beam?
T3.10 Figures 3.3, 3.5, and 3.7 all have the same general form. Why is this?
T3.11 Discuss the advantages and disadvantages of the Rutherford cross section for elastic scattering. Put in some values into equation 3.7 and determine the value of the cross section.
T3.12 Write down concise definitions of coherent, incoherent, elastic, and inelastic as we use them. (Hint: first take a look at Webster's.)
T3.13 Explain in a paragraph the relationship between scattering cross section and atomic scattering factor mentioning the important factors that influence them.
T3.14 In Figure 3.6, if this process were Bragg diffraction, how would the Bragg angle relate to θ?
T3.15 Use equation 3.21 to determine the value of θ if $n = 1$ and $d = 2$ for each of the wavelengths in Table 1.2. Thus, discuss whether or not relativistic corrections are important.
T3.16 Copy Figure 3.8 and draw on it where other atoms in the diffracting planes might be positioned. (Hint: look at Figure 3.9.)
T3.17 Explain why we talk about the Bragg angle in Figure 3.9 as being a semi-angle of scattering.

4

Inelastic Scattering and Beam Damage

CHAPTER PREVIEW

In the previous chapter, we discussed elastic scattering of the electron beam in which the incident electrons lost no energy as they traversed the specimen. Inelastically scattered (often termed energy-loss) electrons are equally important and we'll discuss many of these processes here, but leave the applications till later. Why are we interested in inelastic scattering? Well, such scattering generates a whole range of signals, each of which can tell us more about the chemistry of the specimen than we can find out from the elastic electrons. In addition to the energy-loss electrons themselves, the most important signals are the characteristic X-rays, secondary electrons, and, occasionally, visible light (cathodoluminescence (CL)) and so we'll emphasize how these arise. We will also tell you why all these signals are useful, to varying degrees, to materials scientists, engineers, and nanotechnologists.

So how do we use these inelastic signals? First we have to detect them and we'll describe electron detection in general in Chapter 7 and the spectrometers that disperse electrons according to their energy in Chapter 37. Then in Chapters 38–40 we'll talk about analyzing these energy-loss electrons. We will discuss how to detect X-rays in Chapter 32 and how we get quantitative, elemental information from the spectra in Chapters 33–36. In all cases we get complementary information to that gained in TEM images and DPs. We'll briefly discuss CL images and spectra in Chapter 29. Obviously there's a lot of useful information in these signals and this is a major advantage to using ionizing radiation. However, the other side of the coin is that all the inelastic processes deposit energy in your specimen, which can be damaged if it is beam sensitive. So we must also look at the downside of the inelastic processes and we end the chapter by discussing this problem under the general topic of beam damage or radiation damage.

A warning: This chapter is based on some quite difficult theoretical, physical concepts. However, these concepts form the basis of AEM, which constitutes Part 4 of the book, so we have to address the material. You can safely delay studying much of this material in detail until you reach Chapter 32 and beyond.

- What you need to get out of this chapter is an appreciation that a modern TEM is primarily a signal-generating and detecting instrument, not just a tool for producing high-magnification images.

One redeeming feature is that in all these processes we can treat the electron as a particle. So this is (almost) a wave-free chapter, which generally means that it is more easily understandable unless you have an in-depth physics background.

4.1 WHICH INELASTIC PROCESSES OCCUR IN THE TEM?

Historically, the conventional TEM used only two *elastic* signals, namely, the direct beam and the diffracted beam(s). As we've seen, these signals constitute the DP and we'll see in due course how they can be used to produce images. In operating a TEM in this classical manner we are being extraordinarily inefficient; we throw away a vast amount of information about our specimen which is contained in the signals that result from *inelastic* scatter. Some of these signals are shown back in Figure 1.3 and are often used predominantly in related instruments such as the SEM and the Auger electron spectrometer (AES), but we can also use TEMs to detect many of these signals,

thus allowing for a more complete characterization of the specimen.

Because some of the incident-beam electrons lose energy, all these signals are related to the general topic of electron energy-loss spectrometry (EELS). The EELS signals and the accompanying X-ray signals constitute analytical electron microscopy (AEM), which we cover in Part 4. In seeking to detect more signals from the specimen, we find that practically we cannot do everything at once, nor can we do it all with equal efficiency. Nevertheless, various kinds of analytical TEMs exist which, in one form or another, can detect all the signals shown in Figure 1.3. With the advent of aberration correction, the spatial resolution and the detection limits of the various techniques have reached or closely approach the single-atom level and so are very well suited to the characterization of nanostructured materials.

In this chapter we'll cover all the signals that are detectable and what use (if any) they are to the nanotechnologist. We need to know

- What are the inelastic-scattering interactions?
- What is the range of energy losses associated with each process?
- What is the likelihood that each energy-loss process will occur?
- What is the scattering angle for the various energy-loss electrons?

When a high-energy electron encounters an atom, it first penetrates the outer, loosely bound electron cloud, then it passes the more tightly bound inner (or core) shell electrons, and finally it may encounter the nucleus.

RULE OF THUMB

The deeper the electron penetrates into the atom, the greater the energy that may be lost. It is possible (but very rare) for the electron to lose all its energy in a single nuclear interaction.

This range of inelastic scattering produces a range of scattering angles, but there is no simple relationship between the energy lost and the scattering angle. We'll separate the inelastic processes into three components

- Processes that generate X-rays
- Processes that generate other (secondary) electrons
- Processes that result from collective interactions with many atoms or electrons

We know the first two rather well, but the third is usually poorly defined. Figure 4.1 shows the cross sections for the most important inelastic processes

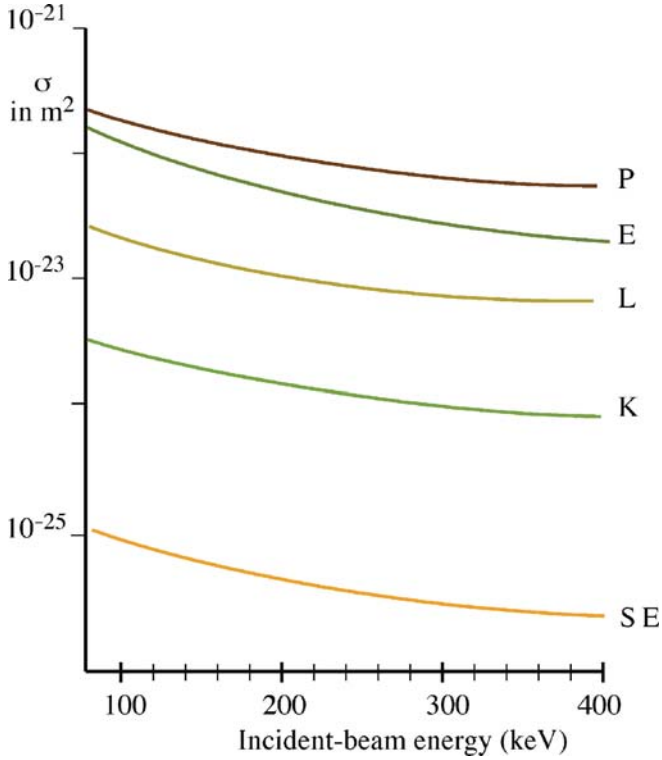


FIGURE 4.1. Cross sections for the various inelastic scattering processes in Al as a function of the incident electron energy, assuming a small angle of scattering ($\theta \sim 0°$); plasmon (P), K- and L-shell ionization (K,L), secondary-electron generation (SE). For comparison purposes the elastic cross section (E) is also included. The values are relatively insensitive to the beam energy.

that we'll talk about. As you can see, the elastic and one of the inelastic (plasmon excitation) processes are by far the most likely events and together account for almost all of the *total scattering cross section* that we discussed back in Chapter 2 (with the caveat that the data in this figure are for small-angle scattering only). These cross sections vary over several orders of magnitude and this fact alone should give you some feel for the relative generation probability of each signal. We'll discuss the specific cross sections for inelastic scattering in more detail as we describe each individual scattering event.

We'll see throughout the book that energy-loss processes are both useful and damaging. For example, in Chapter 19 we describe how energy-loss electrons cause Kikuchi lines to arise in DPs and these are extraordinarily useful. In contrast, some of those same energy-loss electrons cause diffuse scatter that lowers the signal to background information in all DPs and images. If your specimen is thick enough, the energy-loss electrons hide all the useful contrast information, but we'll see in Part 4 how to use EELS to filter out those electrons from images and DPs. This filtering improves the quality of both image and DP and allows the study of much thicker specimens.

4.2 X-RAY EMISSION

We'll consider X-ray emission first because it's the most important secondary signal generated in the specimen. From X-rays we can quickly find out what elements constitute the part of the specimen interacting with the electron beam and we can also quantify the amount of each element in quite a straightforward manner. (The way to do all of this is described in Part 4.) Two kinds of X-rays are produced

- Characteristic X-rays; we'll see these are very useful for local elemental analysis of nano-structured materials and crystal defects.
- Bremsstrahlung X-rays, which are useful to the biologist, but generally regarded as a nuisance by most materials scientists (nano or otherwise).

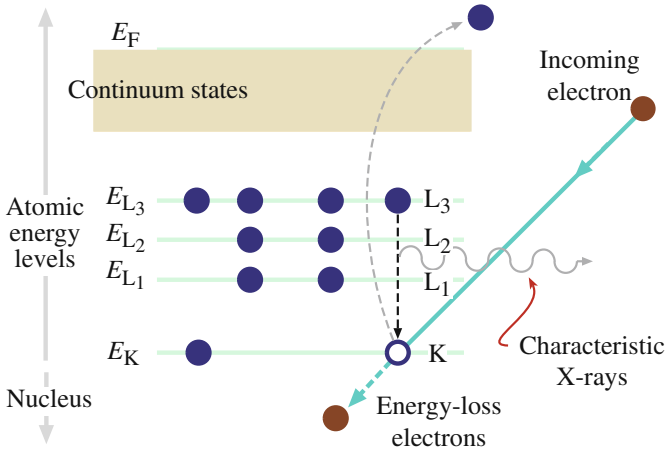


FIGURE 4.2. The ionization process. An inner (K) shell electron is ejected from the atom by a high-energy electron. When the hole in the K shell is filled by an electron from the L shell, characteristic (K_α) X-ray emission occurs. The beam electron loses energy but continues on through the specimen.

4.2.A Characteristic X-rays

How do we produce characteristic X-rays and of what are they 'characteristic'? First of all, a high-energy beam electron must penetrate through the outer conduction/valence bands and interact with the inner-shell (or core) electrons. If more than a critical amount of energy is transferred to an inner-shell electron, that electron is ejected; i.e., it escapes the attractive field of the nucleus, leaving a hole in the inner shell. In an isolated atom, the electron is ejected into the vacuum while in a solid it escapes above the Fermi level into the unfilled states. The atom is then left in an excited state because it has more energy than it would like, and we describe it as ionized.

The ionized atom can return almost to its lowest energy (ground state) by filling in the hole with an electron from an outer shell. This transition is accompanied by the emission of either an X-ray or an Auger electron. This latter process was first described by Auger (1925) and won him the Nobel Prize for Physics. (Since the discoverer was French, we pronounce his name "Ozhay" with a soft *g* as in beige.) In both the X-ray and Auger cases, the energy of the emission is *characteristic* of the difference in energy between the two electron shells involved and this energy difference is unique to the atom. The process of X-ray emission is shown schematically in Figure 4.2. We'll cover Auger emission in Section 4.3.B

Note that characteristic X-rays can also be produced if an atom is ionized by a process other than electron irradiation. For example, ionization can occur as a result of X-ray bombardment also, in which case we use the term fluorescence. It is customary *not* to refer to electron-induced X-ray emission as fluorescence, although you may occasionally come across such usage in the literature.

We've been able to detect X-rays in electron microscopes for many years, but Auger electron detection is rather specialized and usually carried out in a dedicated AES. More recently, however, we've found ways to detect the Auger signal in ultrahigh vacuum (UHV) TEMs and so we'll discuss this in Section 4.3.B below.

You need to know several aspects of the ionization process to understand why the characteristic X-rays are so useful and what it takes to generate them

- What are *electron shells*?
- Why do we use the terms X-ray *lines, families, and weights*?
- What is the *critical ionization energy* and the *ionization cross section*?
- What controls the *X-ray energy* and *wavelength*?
- What is the *fluorescence yield*?

Electron shells: We use a specific terminology to identify the different characteristic X-rays. To understand the terminology you must be familiar with the simple Bohr theory of atomic structure in which the electrons are circling the nucleus in specific shells. (The electrons stay in their shells rather than spiral into the nucleus because of the constraints imposed by quantum theory.)

Aside: For historical reasons, the innermost electron shell is called the K shell and the next innermost is the L shell, the next the M, and so on, as used in Figure 4.2. All the shells (except the K shell) may themselves have subshells (e.g., L_1, L_2, etc.). We name the characteristic X-rays in terms of the shell being filled and the shell from which the electron comes. (The K, L, M, etc., terminology was first introduced by Charles Barkla, an early X-ray spectroscopist and has nothing to do with Royal

Dutch Airlines. The reason Barkla chose K as the first shell may have been because he wasn't sure if he'd need a J shell but knew he'd need an L shell!)

Remember that the difference between the two shell energies equals the energy of the characteristic X-ray. Thus, if we fill a K-shell hole from the L shell we get a K_α X-ray, but if we fill it from the M shell we get a K_β X-ray. If the hole is in the L shell and we fill it from the M shell we get an L_α X-ray, and if we fill it from the N shell we get an L_β X-ray. The notation is in fact much more complex because we differentiate the α X-rays in terms of α_1 and α_2 depending on which subshell of the outer shell the electron falls from to fill the hole. The α_1 X-ray is from the outermost subshell (e.g., the L_{III} or M_V), the α_2 from the next innermost (the L_{II} or M_{IV}). To make this a bit clearer you can look at the diagram in Figure 4.3, although it may in fact confuse you more since it's not at all obvious why, with all these possible electron transitions, only a small fraction generates sufficient X-rays for us to use. Suffice it to say that X-ray physics is an arcane discipline. Fortunately, for X-ray detection in the TEM

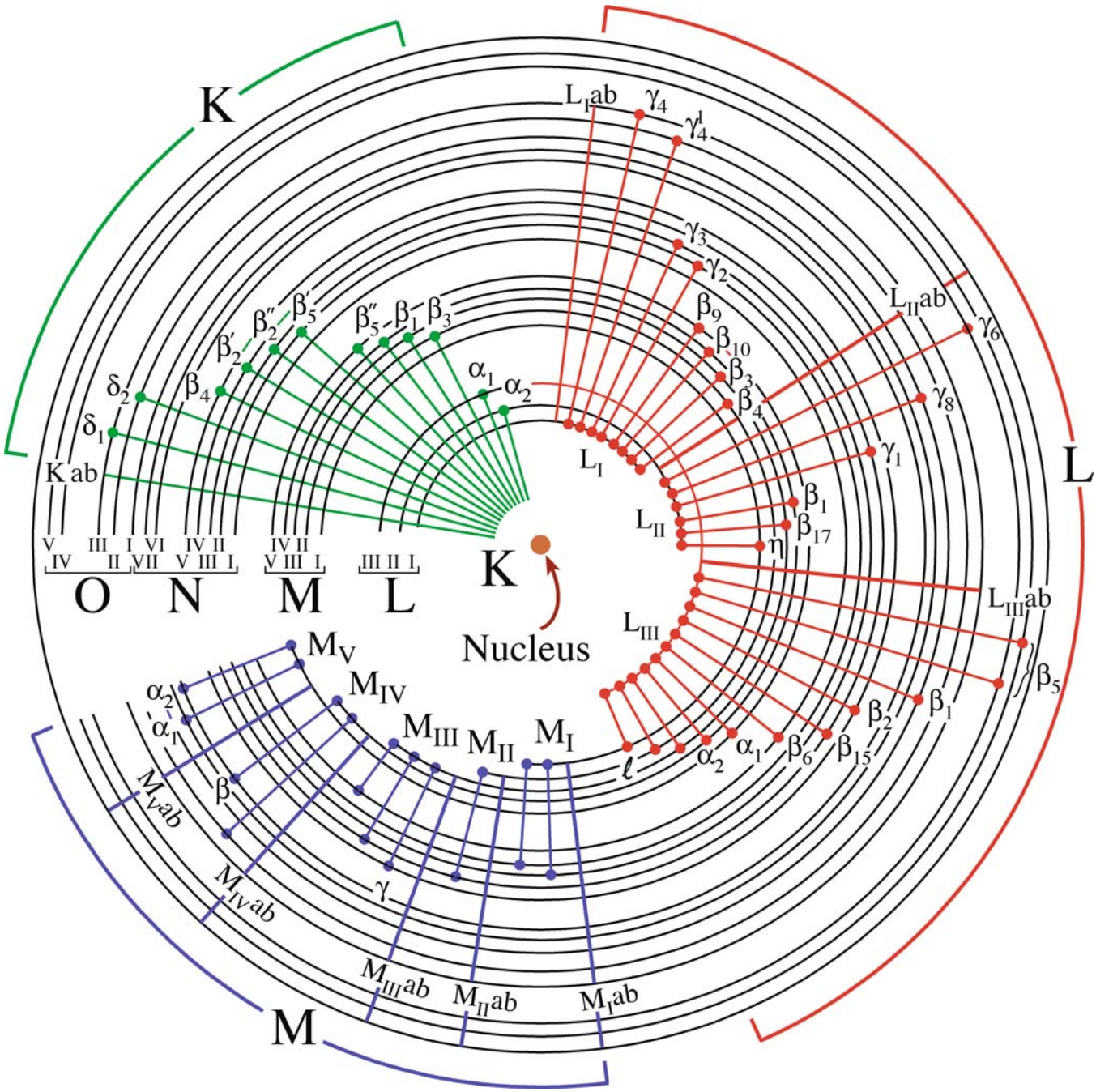


FIGURE 4.3. The complete range of possible electron transitions that give rise to K, L, and M characteristic X-rays. Not all these X-rays are detectable by the XEDS in the TEM.

you don't need to worry about such details because, as you'll see later, the detectors we use can't usually discriminate between the X-rays from different subshells, except at the highest X-ray energies, so K, L, and M and α and β are about all you'll need to remember.

X-ray lines, families, and weights: Often we refer to specific characteristic X-rays as 'lines' because they originally appeared as lines on photographic plates from early spectrometers. Each characteristic X-ray line has a specific wavelength or energy. Groups of lines that arise from transitions to a specific K, L, or M shell are often called 'families.' Much more detail can be found in books on X-rays and X-ray spectrometry, and we'll tell you more in Chapter 34.

Not all electron transitions are equally probable (think of those cross sections again) and this is taken into account by the different 'weights' (i.e., relative intensity compared with the most intense line) of the X-ray lines that are given in Table 4.1. These weights are only important within a given K, L, or M family and do not relate families (e.g., K to L), because experimental conditions affect X-ray generation in each family differently. In X-ray analysis in the TEM we only use the most intense lines, usually the α lines (or, if the spectrometer can't resolve them, both α and β lines). This will become more obvious when you've learned about X-ray qualitative analysis in Chapter 34.

K,L,M,...
A family is a group of X-ray lines and each member of the family has a different relative intensity, which unfortunately is called its weight.

Critical ionization energy: The electron beam has to transfer an amount of energy greater than a certain value to the inner-shell electron in order to ionize the atom. This energy is called the critical ionization energy (E_c). If we want to generate a useful number of X-rays, then the beam energy E_0 must be significantly greater than E_c. The value of E_c increases as the electrons are more tightly bound to the nucleus so, for a given element, the innermost shell (K) has a higher E_c than the next (L) shell, and so on. Atoms with higher Z have more protons binding the core electrons to the nucleus and, therefore, have a higher E_c. You can see this effect if you go and look at an X-ray spectrum, e.g., in Figure 1.4A (or one of many others throughout Part 4) in which the energy of the X-ray peaks increases with increasing atomic number. Since there's a lot of shells and a lot of atoms, the list of critical ionization energies is long, as you'll find out if you consult any X-ray textbook. A similar list is also invaluable in EELS since the E_c obviously corresponds to a critical amount of energy lost by a beam electron and thus gives rise to peaks (usually called edges) in the energy-loss spectrum. As we'll see in Chapter 39, EELS edges, like characteristic X-ray peaks, can also be used to identify uniquely the presence of a particular element in the specimen.

TABLE 4.1 Relative Weights of X-ray Lines

K_α (1)	K_β (0.1)		
$L_{\alpha 1,2}$ (1)	$L_{\beta 1}$ (0.7)	$L_{\beta 2}$ (0.2)	$L_{\gamma 1}$ (0.08)
	$L_{\gamma 3}$ (0.03)	L_l (0.04)	L_η (0.01)
M_α (1)	M_β (0.6)	M_ζ (0.06)	M_γ (0.05)
	$M_{II}N_{IV}$ (0.01)		

The cross section for ionization (σ): It is shown in Figure 4.1 for K- and L-shell electrons. It is not a strong function of energy and has a relatively large value, and so we expect to see X-rays generated in all TEMs. In low-voltage SEMs we have to worry about another parameter called the overvoltage, U, which is the ratio of the beam energy E_0 to the ionization energy E_c. The cross section varies with U as shown in Figure 4.4. What this figure tells you is that if E_0 is close to E_c then there isn't much chance of ionization occurring. However, in the TEM E_0 is $\geq$ 100 keV and E_c is generally $<$ 20 keV so U is usually $>$ 5. Therefore, X-ray generation is expected and the cross section is pretty constant with energy. Despite this relatively simple behavior, there is considerable uncertainty about the absolute value of the ionization cross sections because few reliable experimental measurements have been made at TEM voltages. Most models are variations on the original expression given by Bethe (1930) which describes the total (not the differential) ionization cross section as

$$\sigma_T = \left(\frac{\pi e^4 b_s n_s}{E_0 E_c}\right) \log\left(\frac{c_s E_0}{E_c}\right) \qquad (4.1)$$

where the only new terms are n_s, which is the number of electrons in the ionized subshell, and b_s and c_s, which are constants for that shell. We are not particularly concerned with any angular variation in the ionization process. The differential form of the Bethe expression shows two features

- The electron that ionized the atom is deviated through a really small angle (<∼10 mrads).

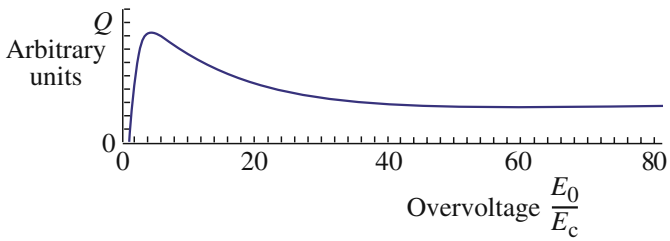


FIGURE 4.4. The variation of the ionization cross section with overvoltage. Ionization is most probable if the beam energy is ∼5× the critical ionization energy. The cross section decreases, but not substantially, at higher overvoltages, typical of a TEM.

- The resultant characteristic X-ray is a spherical wave emitted uniformly over 4π sr.

Remember the Bethe expression is an *inelastic* cross section but, like the Rutherford (elastic) cross section, it also needs to be corrected for the effect of relativity at TEM beam energies. So we substitute the term $m_0v^2/2$ for the beam energy and introduce a standard relativistic factor, β ($= v/c$) (Williams 1933)

$$\sigma = \left(\frac{\pi e^4 b_s n_s}{\left(\frac{m_0 v^2}{2}\right) E_c}\right)\left[\log\left[c_s\left(\frac{m_0 v^2}{2E_c}\right)\right] - \log(1-\beta^2) - \beta^2\right] \tag{4.2}$$

This modified Bethe cross section can be manipulated to fit almost any X-ray data just by altering b_s and c_s, although such parameterization is not always justified. Several cross section models have been developed, all of which are modifications to Bethe's approach (e.g., Newbury 1986, Goldstein et al. 1986). A good source of cross section data can be found on the NIST Web site (URL #1).

The X-ray energy/wavelength: X-rays are electromagnetic radiation and so we usually think of them of as waves with a specific wavelength λ. But, just like electrons, X-rays can show particle-like characteristics and then we describe them as photons with a specific energy such as E_K or E_L, where the subscript refers to the shell from which the core electron was ejected.

There is a similar inverse relationship between the X-ray wavelength and its energy, as we saw for electrons back in Chapter 1. However, there are a couple of important differences which you *must* remember.

- An X-ray is a photon of electromagnetic energy, so the concepts of rest mass and momentum embodied in the electron energy are irrelevant; an X-ray has *no mass*.
- X-rays, like all electromagnetic radiation, travel at the speed of light (c) in vacuum and consequently we don't have to make increasing relativistic corrections as their energy increases. So the quantized X-ray energy is just $h\nu$ where h is Planck's constant and ν is the frequency and in order to express this energy in eV we equate it to E, where E is the X-ray energy.

Thus

$$E = h\nu = \frac{hc}{\lambda} \tag{4.3}$$

Now since h and c are constants we can substitute values with appropriate units into the equation and find that the X-ray wavelength is given by

$$\lambda = \frac{1.24}{E} \tag{4.4}$$

where λ is in nm and E in keV. This expression is *very* similar to the expression for the uncorrected *electron* wavelength ($1.22/E^{1/2}$) where E is the electron energy in eV (not keV) that we gave back in Chapter 1. You can easily confuse the two, so beware!

SPEED

Don't confuse electromagnetic waves (e.g., X-rays), which always travel at the speed of light, with electron waves whose speed depends on their energy.

Because the X-ray energy depends on the difference in the inner-shell energies and these differences increase monotonically with Z, we can use the detection of a characteristic X-ray with a specific energy as an unambiguous sign of the presence of an element in the specimen (although it doesn't necessarily mean that that element is intrinsic to your specimen, as we'll see in Chapter 33). The concept of the atomic number (Z) of the specimen and its relationship to the X-ray energy/wavelength was reported by the brilliant young physicist, HGJ Moseley. Soon after his discovery, Moseley volunteered for the British army and, despite his talents, was dispatched to the trenches of Gallipoli in 1915 where he was promptly killed before he could be nominated for the Nobel Prize, which would undoubtedly have been his. He is remembered by Moseley's law which states

$$\lambda = \frac{B}{(Z-C)^2} \tag{4.5}$$

where B and C are constants. So we can also generate a list of X-ray energies which are associated with each atomic transition. As with E_c the complete list is enormous and given in Bearden's tables. More compact lists are given in the software attached to the XEDS system on your TEM or in handy 'slide rules' by the manufacturers of X-ray spectrometers, in textbooks, review articles, or on trusted Web sites such as NIST (URL #2) or NPL (URL #3).

If you compare the value of E_c and the relevant characteristic X-ray energy you'll see that they are not quite identical. The X-ray energy, E_K or E_L, is invariably less than E_c. Table 4.2 lists a comparison of critical ionization energies and corresponding X-ray energies for a range of elements. Note how the differences in energy increase with increasing Z. The differences arise because the atom doesn't return completely to ground state when the X-ray is emitted. If the electron that fills the hole in the ionized inner shell comes from an outer shell then this process will leave a hole in that outer shell.

TABLE 4.2 Difference Between E_c and E_K

Element	Critical ionization energy E_c (keV)	X-ray energy E_K (keV)
C	0.282	0.277
Al	1.562	1.487
Ca	4.034	3.692
Cu	8.993	8.048
Ag	25.531	22.163

Note that the energies may be affected by bonding states but shifts will only be a few eV.

This hole must also be filled by another electron with perhaps the emission of another X-ray (much more likely as Z increases) and so on until eventually a free electron from the conduction or valence band fills the last hole in the outermost core shell.

An example: A Cu K-shell electron requires 8.993 keV of energy for ionization (E_c = 8.993 keV). One possible sequence by which this extra energy within the atom is lost is first by the creation of a Cu K_α X-ray (8.048 keV), then an L_α X-ray (0.930 keV). These X-ray energies total 8.978 keV and the remaining few eV could come from the hole in the M shell being filled from the conduction band with the emission of a photon or the generation of phonons (see below).

CASCADE TO GROUND

So the ionized atom returns to ground state not via a single event but by a cascade of transitions, depending on the complexity of the electronic structure of the atom.

The possible variations are enormous and affected by such events as Coster-Kronig transitions, in which the atomic shells rearrange their energies after the electron transition. The situation is further complicated if the ionized atom is bound to a different atom, in which case the energy of the X-ray can be shifted slightly (<~5 eV). Such detail is well beyond what you need to know now but a textbook on X-ray spectrometry will give you more, if you so wish.

Fluorescence yield: Remember that an ionized atom does not have to lose energy by giving off a characteristic X-ray but can emit an Auger electron instead. The probability of X-ray versus Auger emission is described by the fluorescence yield, ω, which is the ratio of X-ray emissions to inner-shell ionizations. The fluorescence yield is a strong function of atomic number as shown in Figure 4.5, decreasing at a rate proportional to Z^4 as Z decreases. One common expression for ω gives

$$\omega = \frac{Z^4}{a + Z^4} \tag{4.6}$$

where $a \sim 10^6$ for the K shell. While an approximation, this equation still describes a formidable dependence on Z. For carbon (Z = 6), ω is $\sim 10^{-3}$ and for Ge (Z = 32), ω is ~ 0.5. So you have to ionize 1000 carbon atoms before you get a single C K_α X-ray but only two ionizations are needed to produce a Ge X-ray. So if you ionize low-Z atoms, the chances are you won't see an X-ray and therefore XEDS is *not* the best way to analyze light elements; you should use EELS (see Part 4) because we can always detect the energy-loss electron whether or not it has generated an X-ray. If you want to know more about the chances of generating X-rays in your microscope then the database by Hubbell et al. is a good source.

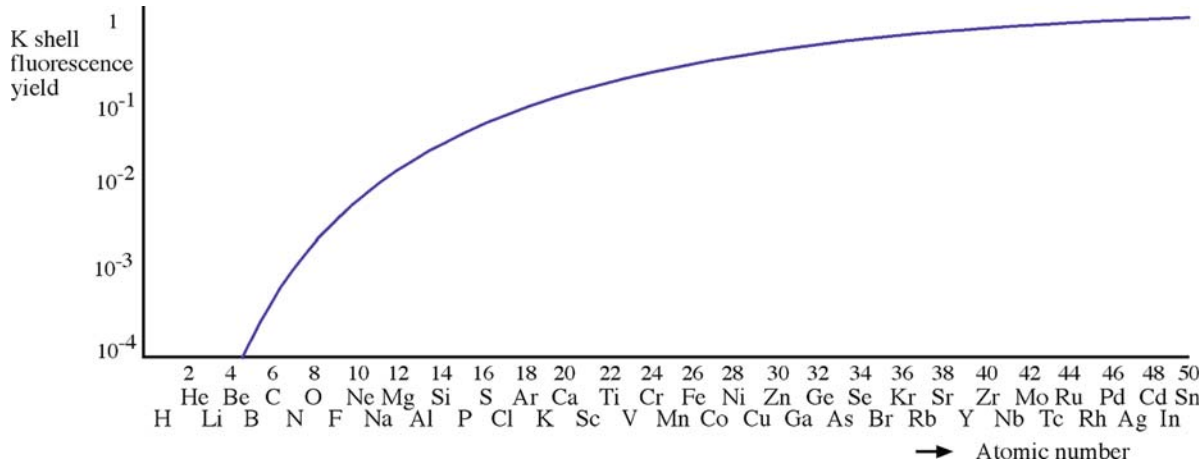


FIGURE 4.5. Fluorescence yield for K-shell X-rays as a function of atomic number. Note the rapid decrease at low atomic numbers. X-rays from elements below Be are not detectable.

4.2.B Bremsstrahlung X-rays

If the electrons in the beam penetrate completely through the electron shells they can interact inelastically with the nucleus. If the electron interacts with the Coulomb (charge) field of the nucleus, it can suffer a substantial change in momentum and during this process it may emit an X-ray. Since the electron can suffer any amount of energy loss, depending on the strength of its interaction, then these X-rays can have any energy up to the beam energy. Such X-rays are known by their original German name of *bremsstrahlung* which can be translated as 'braking radiation.'

The likelihood of bremsstrahlung creation is usually described by the cross section derived by Kramers. This expression is often used for thin TEM specimens, although it was originally derived for bulk samples. It is common to use the Kramers cross section to predict the bremsstrahlung production rather than the probability of interaction. The approximate expression used is

$$N(E) = \frac{KZ(E_0 - E)}{E} \tag{4.7}$$

where $N(E)$ is the number of bremsstrahlung photons of energy E (i.e., the intensity) produced by electrons of energy E_0, K is Kramers' constant, and Z is the atomic number of the ionized atom. This relationship predicts that it is far more likely that the interaction causes a small loss of energy and exceedingly rare that the electron loses all its energy in one deceleration at the nucleus. So the bremsstrahlung intensity as a function of energy is shown in Figure 4.6. In contrast to the isotropic emission of the characteristic X-rays, the bremsstrahlung is highly anisotropic, showing strong forward scattering which increases as E_0 increases. This anisotropy is very useful since it allows us to design spectrometers that collect many more useful characteristic X-rays than relatively useless bremsstrahlung X-rays.

The bremsstrahlung has a continuous energy spectrum on which the characteristic X-rays that we just talked about are superimposed, as also shown schematically in Figure 4.6 and experimentally in the spectrum back in Figure 1.4. Since the characteristic X-rays have a narrow energy range, they appear as sharp peaks in the spectrum centered at specific energies, indicated by computer-generated lines on the display (now another reason to call them 'lines'). The bremsstrahlung intensity depends on the average Z of the specimen and this is useful to biologists or polymer scientists who are interested in this aspect of their specimens. But materials scientists generally dismiss the bremsstrahlung as a signal, which only succeeds in obscuring characteristic lines. We'll come back to the X-ray spectrum in more detail in Chapters 32–36.

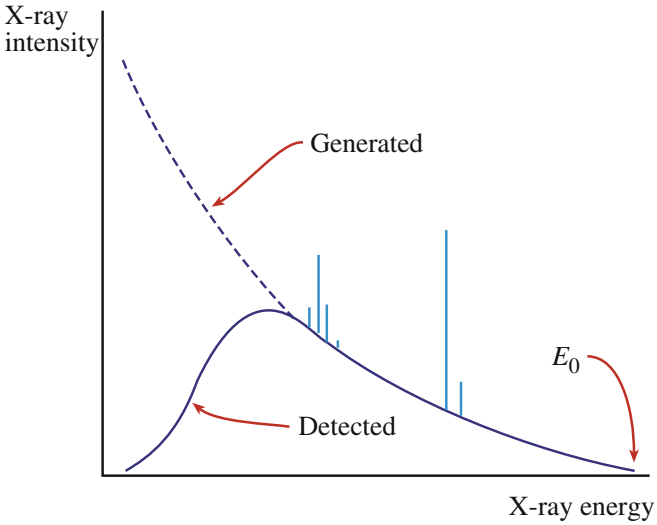


FIGURE 4.6. The bremsstrahlung X-ray intensity as a function of energy. The generated intensity increases rapidly with decreasing X-ray energy but at energies $<\sim$ 2 keV the bremsstrahlung is absorbed in the specimen and in any detector being used so the observed intensity in the detected spectrum drops rapidly to zero. E_0 is the energy of the electrons that cause the X-ray emission. Two families of characteristic lines at specific energies are also shown superimposed on the bremsstrahlung.

4.3 SECONDARY-ELECTRON EMISSION

Secondary electrons (SEs) are electrons within the specimen that are ejected by the beam electron.

- If the electrons are in the conduction or valence bands then it doesn't take much energy to eject them and they typically have energies $<\sim$ 50 eV.
- If the electrons are ejected from an inner shell by the energy released when an ionized atom returns to the ground state, then these SEs are called Auger electrons. The process is often termed a non-radiative transition (since no X-ray emerges from the atom) and the energy undergoes an 'internal conversion' (which is not quite a religious experience).

Historically, SEs were usually considered only in relation to the SEM where they are used to form (often stunning) images which are sensitive to surface topology. We'll now discuss each of these SE signals and their relative importance in the TEM.

4.3.A Secondary Electrons

SEs are ejected from the conduction or valence bands of the atoms in the specimen. The actual emission process can be quite complex and no simple cross section model covers all production mechanisms. The data in Figure 4.1 indicate that SE emission is a far less likely process than all the other inelastic processes we've discussed, but enough are generated for them to be useful in the TEM.

Usually, SEs are assumed to be free electrons, i.e., they are not associated with a specific atom and so they contain no specific elemental information. Because SEs are weak they can only escape if they are near the specimen surface. So we use them in SEMs for forming images of the specimen surface. While SEs are the standard signal used in SEMs, they are also used in STEMs where they can provide very high resolution, topographic images of the specimen surface. We'll discuss ways to detect SEs in Chapter 7 and we'll talk about the images themselves in Chapter 29.

We'll discuss several reasons for the improved (SE) resolution in STEM in Chapter 29. However, recent developments in high-resolution field-emission gun (FEG) SEMs have produced SE image resolution <0.5 nm (close to surface atom resolution) at 30 kV. (We discuss FEGs in the next chapter.) A STEM at 100 kV can offer similar or better resolution even without an FEG, so the SEs are very useful. Aberration correction in STEM naturally brings about even higher-resolution SE images, close to the atomic level.

SE RESOLUTION

SE images in a STEM have much better resolution than SE images in the (relatively) low-kV SEMs.

The number of SEs does depend on energy; it rises to a maximum at about 5 eV and drops close to zero with energies >~50 eV. (You should know that, on rare occasions, strongly bound inner-shell electrons can be ejected with energies up to about 50% of the beam energy. Such fast SEs are generally ignored because they do not seem to limit the resolution of XEDS in the TEM.) The SE yield (number of SEs/incident-beam electron) is generally regarded as being independent of E_0; if there is any Z dependence (which is still a matter of some debate) then it is very small. The angular distribution of emitted SEs is not important since the detector uses a strong field to gather SEs emerging from the surface at any angle. But the *number* of SEs increases with specimen tilt because SEs escape more easily as the surface is tilted parallel to the beam. This behavior is a critical aspect of SE emission because it mimics Lambert's cosine law of visible-light reflection, accounting for the great similarity between SE images of rough specimens and the everyday, reflected-light images we are accustomed to seeing with our eyes.

4.3.B Auger Electrons

Remember we said at the start of this chapter that the emission of Auger electrons is an alternative to X-ray emission as an ionized atom returns to its ground state.

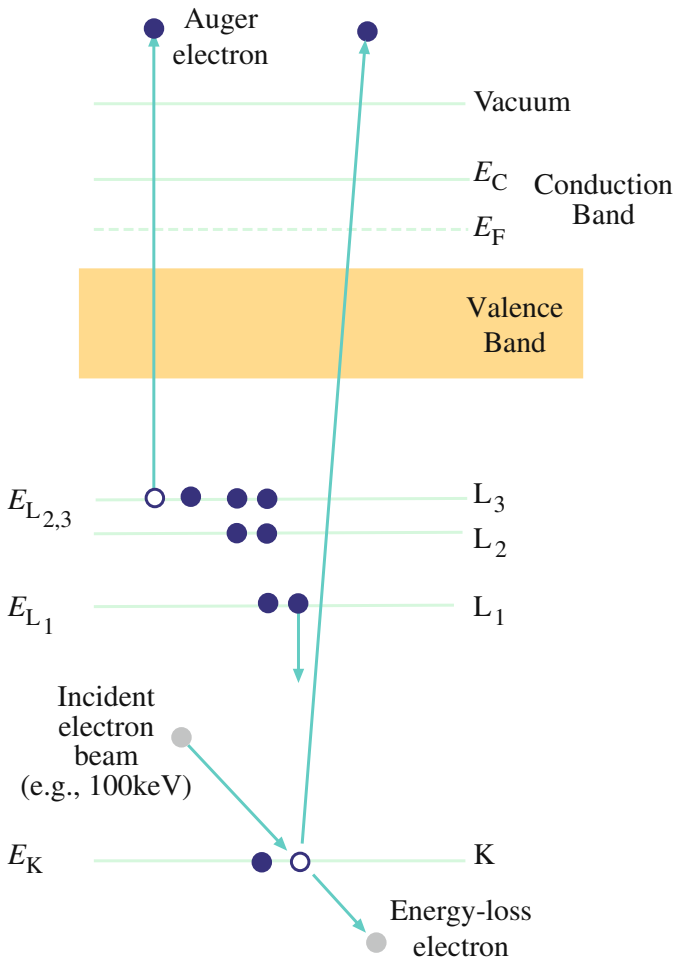


FIGURE 4.7. The process of inner (K) shell ionization and subsequent Auger-electron emission. The energy released when the L_1 electron fills the hole in the K shell is transferred to an electron in the $L_{2,3}$ shell which is ejected as a $KL_1L_{2,3}$ Auger electron.

Figure 4.7 shows how such an atom ejects an outer-shell (Auger) electron; it's instructive to compare with Figure 4.2 for X-ray emission. The ejected electron has an energy given by the difference between the original excitation energy (E_c) and the binding energy of the outer shell from which the electron was ejected. This explains the rather complex nomenclature used to describe each Auger electron (see the caption to Figure 4.7). So the Auger electron has a characteristic energy that is dependent on the electronic structure of the ionized atom and is almost identical to the energy of the alternative, characteristic X-ray.

Because they have such low energies, the Auger electrons that do escape come from very close to the specimen surface. They contain chemical information and consequently AES is a recognized surface-chemistry technique. Because of the similarity in energy between Auger electrons and characteristic X-rays, you might ask, why is light-element X-ray analysis in the TEM not just a surface technique? What you have to

remember is that characteristic X-rays are much less strongly absorbed in the specimen than electrons of similar energy. So most X-rays generated in a thin TEM specimen can escape and be detected. (So it's all to do with the cross sections for interaction in the first place.)

AUGER

The Auger process is favored in atoms having small binding energies, i.e., the lighter elements. Typical Auger electron energies are in the range of a few hundred eV to a few keV and are strongly absorbed within the specimen.

Because Auger emission is a surface phenomenon, the state of the specimen surface is paramount. Oxidation or contamination will prevent interpretable Auger analysis of the true surface chemistry and so we only carry out AES in a UHV system. As a result, the Auger signal has traditionally been ignored by electron microscopists and confined to the realm of surface chemistry, along with such techniques as ESCA and SIMS. However, as TEMs are being built with better vacuums and UHV STEMs become more common, the Auger signal may receive more interest. Unfortunately, it is not simple to attach an Auger system to a STEM unless you build a dedicated instrument in which routine AEM is difficult, so such studies are still very rare.

4.4 ELECTRON-HOLE PAIRS AND CATHODOLUMINESCENCE (CL)

These two signals are closely related. We'll see in Chapter 7 that one way to detect electrons is to use a semiconductor that creates electron-hole pairs when hit by high-energy electrons. So if your specimen happens to be a direct-gap semiconductor then electron-hole pairs will be generated inside it.

Cathodoluminescence is explained schematically in Figure 4.8. The emitted photon has a frequency (i.e., color) equal to the energy of the gap (E_G) divided by Planck's constant (h). If the band gap varies for some reason, there will be a spectrum of light given off or the color of the light will vary depending on what part of the specimen is being observed. So CL spectroscopy has applications in the study of semiconductors and impurity effects therein. While the spatial resolution of CL is not down to the nanometer level like X-rays or secondary electrons, it is still well within the nano-scale range, typically defined as <100 nm.

CATHODOLUMINESCENCE

Electrons and holes will recombine and in doing so give off light; this process is referred to as CL.

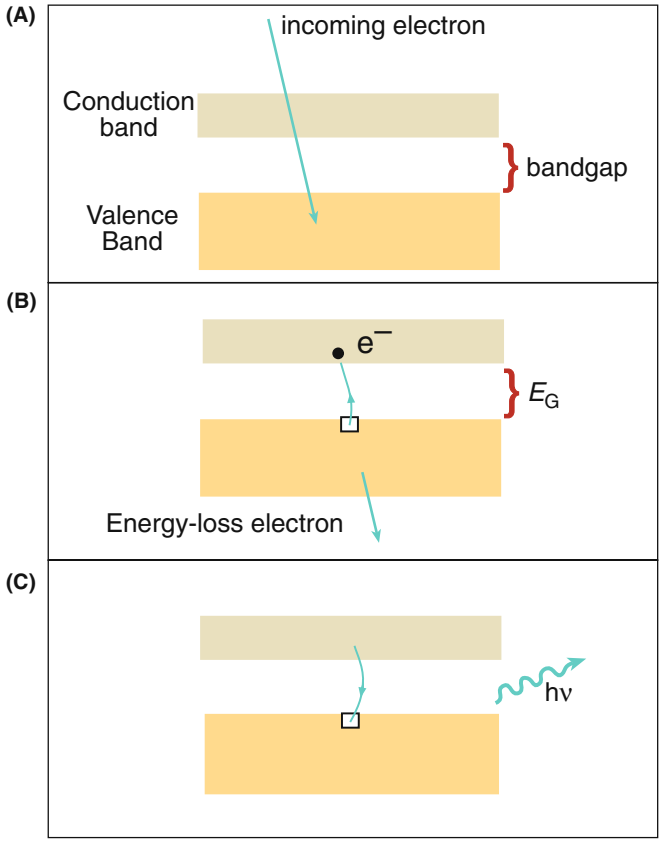


FIGURE 4.8. Schematic illustration of CL. (A) Initial state before a beam electron interacts with valence-band electrons. (B) A valence-band electron is excited across the gap into the conduction band, leaving a hole in the valence band. (C) The hole is filled by a conduction-band electron falling back into the valence-band hole. Upon recombination a photon of light is emitted with a frequency determined by the band gap.

Now if you apply a bias to your specimen or if it happens to be a p-n junction or a Schottky-barrier diode, then the electrons and holes can be separated by the internal bias. You can pick up this charge if you ground the specimen through a picoammeter. In this situation, your specimen is acting as its own detector! The current you then detect is sometimes called the electron-beam-induced current (EBIC) signal. If you detect this signal and use it to form an image then you are doing charge-collection microscopy (CCM).

The CL and CCM modes of operation are standard methods of characterizing bulk samples in the SEM. In principle, there is nothing to prevent us doing the same in a STEM, and a few people have built dedicated instruments, but generally the space available in the TEM stage limits the efficiency of signal collection. This may improve with C_s correction, but, in general, these two techniques are rare and mainly limited to studies of semiconductors (e.g., Boyall et al.) although some minerals also exhibit CL. We'll describe

CL detectors in Chapter 7 and show you an image in Chapter 29. Just remember that CL and CCM are potentially powerful, but highly specialized, techniques.

4.5 PLASMONS AND PHONONS

We can link these two phenomena because they are both examples of what we call collective oscillations.

We can consider plasmons as analogous to sound waves, since they are longitudinal oscillations of the free-electron gas, which create regions of varying electron density as shown schematically in Figure 4.9. These oscillations are damped out in less than a femtosecond and the wave is localized to <10 nm. If you go back to Figure 4.1, you'll see that the plasmon process has the largest cross section so it's by far the most common inelastic interaction occurring in materials and, as we'll see in Chapter 38, plasmon peaks are strong features of EEL spectra. Plasmons can occur in any material with weakly bound or free electrons, but they occur predominantly in metals, particularly ones like aluminum which have a large Fermi surface and, therefore, a high free-electron density. The plasmon oscillation is quantized and the mean free path for plasmon excitation is ~100 nm. As we'll also see in Section 38.3.C this quantization makes the number of plasmon excitations a useful way to measure your specimen thickness. Also, the plasmon energy is a function of the free-electron density and this changes with composition (see Chapter 38 also), so the plasmon excitation is chemically dependent, although we rarely use it for elemental analysis.

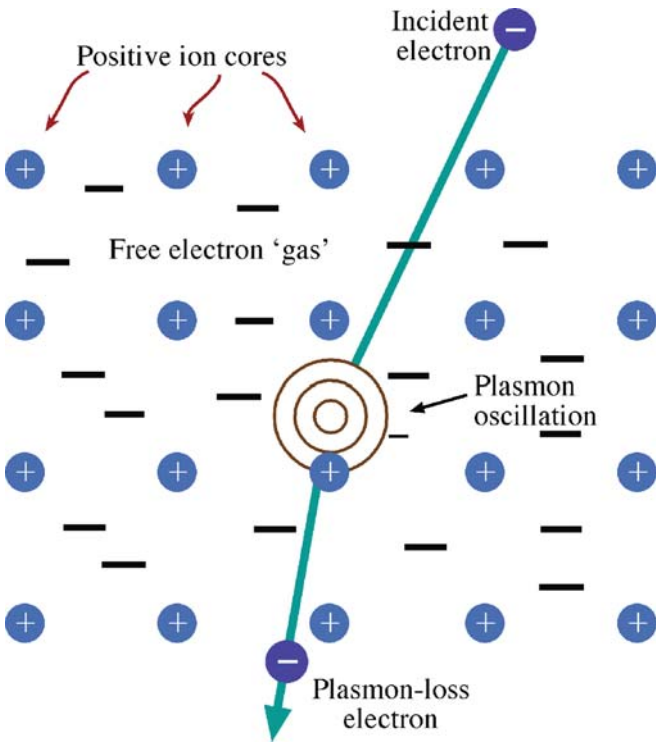


FIGURE 4.9. Schematic diagram of a high-energy beam electron exciting a plasmon oscillation in a free-electron gas that permeates the ion cores in a metal.

PLASMONS AND PHONONS

Plasmons are collective oscillations of free electrons that occur when a beam electron interacts with the free electron 'gas.'

Phonons are collective oscillations of atoms in a solid that arise when the atomic lattice is struck by a beam electron.

The differential cross section for plasmon excitation has a general Lorenztian form

$$\frac{d\sigma_\theta}{d\Omega} = \frac{1}{2\pi a_0}\left(\frac{\theta_E}{\theta^2+\theta_E^2}\right) \tag{4.8}$$

where a_0 is the Bohr radius, θ is the scattering angle, and θ_E is the so-called characteristic scattering angle given by $E_P/2E_0$ (which is always going to be small given the large value of E_0 in TEMs). Since E_P, the plasmon energy, is almost fixed (~15–25 eV), the cross section is a strong function of θ, dropping rapidly to zero at values much above 10 mrads, indicating once again the strong forward scattering of such energy-loss electrons.

When a high-energy electron strikes an atom in the specimen, the lattice shakes, just like hitting a chain-link fence with a stick. This process occurs because, as shown in Figure 4.10, all the atoms are linked together elastically. Phonons can also be generated by other inelastic processes occurring within the atom; for example, the energy of Auger or X-ray emission or an interband transition is sometimes converted internally to lattice vibrations. Any shaking of the atoms is equivalent to heating up the specimen and the net result of all phonons is that the specimen gets warmer. As we will see, this is particularly damaging to some specimens.

The incident electron can generate phonons in any solid specimen, even amorphous ones in which there is no periodic crystal structure. Typically, a phonon vibration causes a very small energy loss of < 0.1 eV but the phonon-loss electrons are scattered out to quite large angles (5–15 mrads), and these electrons account for the diffuse background intensity present between the Bragg intensity maxima in DPs. Phonon-scattered electrons carry no useful chemical information nor do they carry contrast useful to the microscopist.

It is not important to know the phonon-scattering cross section exactly, but it is useful to remember that

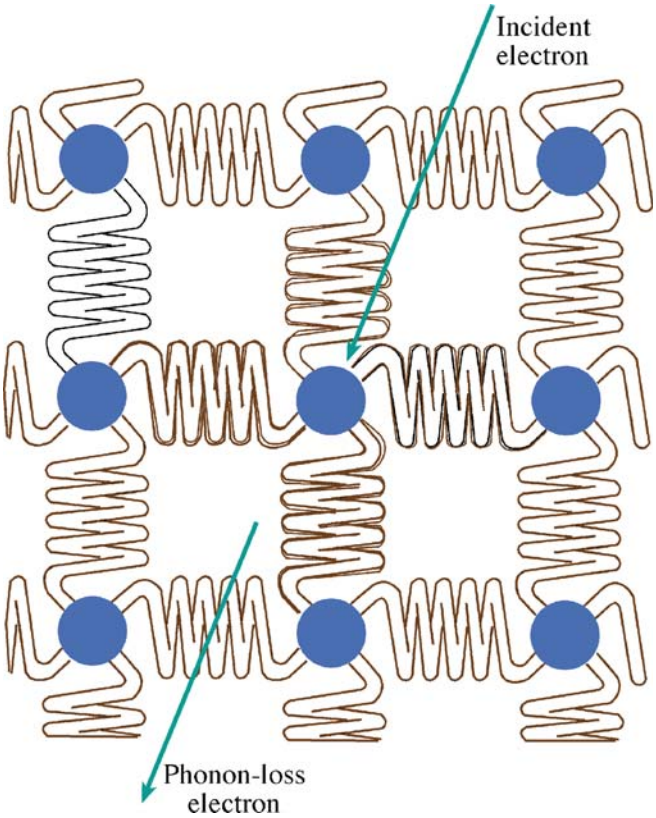


FIGURE 4.10. An illustration of the crystal lattice as a group of atoms linked elastically by springs. The bonds vibrate when struck by a high-energy electron creating lattice oscillations or phonons and these vibrations are equivalent to heating the specimen.

phonon scattering increases with Z with a dependence $\sim Z^{3/2}$, which is a somewhat weaker dependence than for true elastic scattering. Also, because of the effect of temperature on atomic vibration, phonon scattering increases as the temperature rises. This accounts for the increase in thermal-diffuse scattering with temperature and is the major reason why we cool specimens if we want to obtain good, sharp DPs. The mean free path for phonon scattering at room temperature varies from a couple of nm for Au up to about 350 nm for Al, and at liquid-He temperatures these values increase $\sim$ 2–3X.

We don't use either plasmons or phonons directly to form images (although in principle this is possible), but we do detect the electrons that caused them, and we'll discuss the (rather limited) uses of plasmon energy-loss electrons in Chapter 38.

PHONONS

These oscillations involve all the atoms in the crystal lattice vibrating collectively. Such vibrations are equivalent to specimen heating. You can reduce the number of phonons by cooling your specimen.

There are other inelastic processes that can occur as the beam electron traverses the specimen, such as inter-band and intra-band transitions within the valence or conduction bands. The bonding between the atoms and local atomic arrangements within the specimen can also affect many of the inelastic events that we have described in ways that we can both predict and discern. Other electronic features such as the band gap can affect the possibility of certain inelastic interactions occurring. In principle, any atomic or electronic characteristic of your particular specimen that reduces the energy of a beam electron during its passage through the specimen can be detected, measured, and perhaps quantified and simulated. These signals can all be used to form images containing information which complements that contained in the elastic-electron (mainly diffraction-contrast) images that constitute traditional TEM studies of materials. We'll give you plenty of examples of such inelastic images in Part 4.

4.6 BEAM DAMAGE

The inelastic collisions that give us all the useful signals we've just discussed bring with them an unfortunate side effect, electron-beam damage. We are often less precise and call this phenomenon radiation damage. The damage, which affects the structure and/or the chemistry of the specimen, depends, in some form or other, on the incident-beam energy. Certain materials are more susceptible than others but, in the end, you can damage virtually anything that you put into the TEM, particularly now that aberration correction permits even more electron current to be focused into even smaller beams. Therefore, damage represents a real physical limit on what the TEM can do and may be regarded as the microscopists' analog of the Heisenberg uncertainty principle in that the very act of observing your specimen can change it.

DAMAGE

Once its structure or chemistry is changed, your thin specimen is not representative of its parent material and interpreting any of your TEM images, DPs or spectra becomes more difficult.

On the other hand, we can sometimes use beam damage to aid certain in-situ transformations that are speeded up by the damage process or we can use electron damage to emulate other forms of radiation damage. Generally, however, beam damage must be considered undesirable.

Damage takes one of three principal forms

- *Radiolysis:* Inelastic scattering (mainly electron-electron interactions such as ionization) breaks the chemical bonds of certain materials such as polymers and alkali halides.

- *Knock-on damage or sputtering:* Knock-on damage is the displacement of atoms from the crystal lattice and creates point defects. If atoms are ejected from the specimen surface we call it sputtering. These processes are ubiquitous if the beam energy (E_0) is high enough.
- *Heating:* Phonons heat your specimen and heat is a major source of damage to polymers and biological tissue.

We will see that, paradoxically, radiolysis is reduced at higher E_0 while knock-on damage is increased; so there is sometimes no way around the damage problem; you can get it at all energies. Cooling your specimen can obviously help if it is likely to be damaged by heating.

All these processes occur in the voltage range available in commercial TEMs and so you must be aware of the dangers. The actual processes can be very complicated and are also specimen-specific, so we could get bogged down in an enormous amount of detail. What we'll do, however, is describe the fundamental processes in different materials, explain how you can determine if your specimen is being damaged and describe how you can minimize or eliminate the problem.

If you find you need to know more about radiation damage, the text by Jenkins and Kirk is the place to start. But we should also note upfront that whole areas of TEM such as in-situ studies and environmental TEM actually take advantage of radiation damage to enhance particular reactions.

We'll start our brief overview of damage in different kinds of TEM specimens by explaining the terms we use to measure damage.

4.6.A Electron Dose

In the TEM we define the electron dose as the charge density (C/m^2) hitting the specimen. It is easy to convert this to the number of electrons/unit area (usually e/nm^2) knowing that e = 1.6×10^{-19} C. This term is *not* the same as for radiation effects on the human body, for which we define dose as the energy absorbed per unit volume. This human dose is defined by the Gray (Gy) which is the absorption of 1 J of ionizing radiation/kg of material and 1 Gy = 100 rads (in pre-SI units). If we convert the incident electron dose to an absorbed dose it can easily be shown that typical electron exposures inside the TEM are well above the lethal limit for human tissue and early microscopists occasionally found this out to their cost (although we hasten to add that, to our knowledge, while no deaths occurred, digits were apparently lost). While this fact is another warning about the dangers inherent in TEM it is more pertinent as a reminder to you that we put an enormous amount of energy into our specimens. This latter point is well illustrated if you calculate the total power input into the specimen, as we do in the next chapter. Fortunately, such a small fraction of the beam energy is transferred to a thin specimen that most specimens survive this, otherwise hostile, environment.

4.6.B Specimen Heating

Specimen heating is difficult to measure experimentally because of the many variables that can affect the result, such as the thermal conductivity, thickness, and surface condition of the specimen as well as the beam size, energy, and current. Hobbs has calculated the effects of beam current and thermal conductivity on the specimen temperature, as shown in Figure 4.11. From these results we can say that, as a rule for metals and other good conductors, beam heating is negligible under standard TEM conditions but, for insulators, it can be quite substantial. To minimize heating, follow the instructions given at the end of the next section.

You will often hear questions about beam heating from people who don't use TEM.

- If thermal conduction is very high, heating is negligible.
- If thermal conduction is poor, heating can be quite substantial.

So, beam heating for metals is usually minimal but small ceramic particles may be heated by the beam to temperatures of ~1700°C. If a good thermal conductor is thermally insulated from its surroundings, considerable

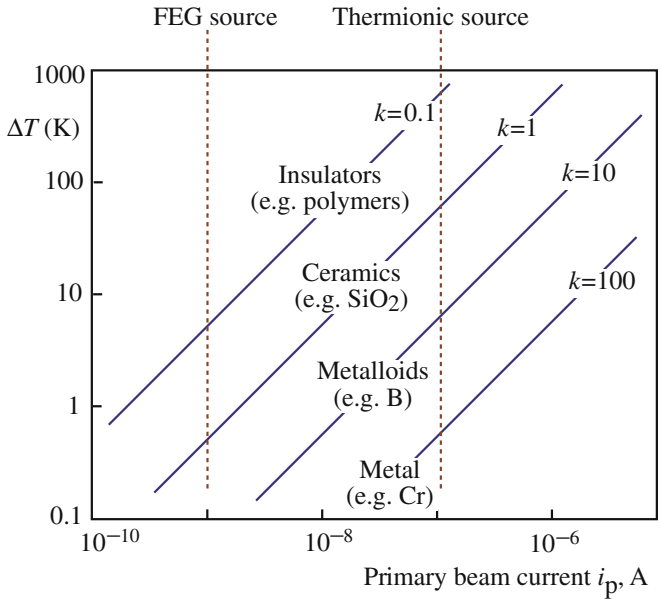


FIGURE 4.11. The increase in specimen temperature as a function of the beam current and the thermal conductivity of the specimen (k, in W/m K). Typical materials are noted, but should not be considered representative, since k varies substantially in any class of materials.

beam heating can occur. Most electrons go through a thin specimen and lose very little energy, as we'll see in our discussion of EELS.

4.6.C Beam Damage in Polymers

Polymers are particularly sensitive to the electron-electron interactions since these can break chemical bonds creating new structures; we call this process radiolysis.

- Electrons can cause the main polymer chain to break, thus changing its basic structure.
- Electrons can cause side groups to break off, leaving reactive free radicals which may crosslink to form a new structure.

Breaking a polymer chain this way is known as scission. Generally, polymers show a tendency either to break down or to crosslink under electron irradiation. In the former case, the polymer will continue to lose mass while, in the latter, the polymer eventually becomes mainly carbon. Mass loss can sometimes be measured directly by EELS in the TEM and it can also manifest itself as a major dimensional change in your specimen because, ultimately, a hole appears in the damaged area; if you're watching carefully you'll see that the image contrast will usually change before the hole appears!

If your polymer specimen was originally crystalline, then radiation damage results in a loss of crystallinity, which you can measure quantitatively from the loss of diffraction contrast in the image or the loss of sharp peaks in the DP (which are gradually replaced with diffuse scattered intensity characteristic of amorphous structures (go back and look at Figure 2.13A). Sometimes you can preserve the crystal structure by staining with a heavy metal such as Pb or U. However, whenever you stain your specimen you affect its structure and change the chemistry, so this isn't ideal.

There are several methods you can use to minimize beam damage in polymers

- Use low-dose imaging techniques (see Chapter 31).
- Cool the specimen to liquid N_2 temperatures or lower T, if possible.
- Coat the specimen with a conducting metal film.
- Use STEM imaging (Section 22.3).
- Do all of the above, if necessary.

In addition to these practical steps, any contribution to damage from beam heating is generally minimized by reducing the cross section for inelastic scatter, i.e., by using the highest available voltage. So HVEMs are better for the study of heat-sensitive materials. If the specimen is thinner than the mean free path for inelastic interaction then less energy is transferred to the specimen and the result is less damage due to heating effects.

4.6.D Beam Damage in Covalent and Ionic Crystals

In covalent and ionic materials such as ceramics and minerals, radiolysis can change the specimen chemistry and possibly its structure through a series of reactions driven by the electron beam. The inelastic interaction primarily responsible for radiolysis is the interband transition, similar to that which causes CL. The transition of a mobile, valence-band electron to the conduction band leaves a hole in the original energy level. Rather than emitting a photon, the electrons and holes may partially recombine via an intermediate, metastable state called an exciton which, through a rather complicated sequence of events, can create an anion vacancy and a cation interstitial. Crystalline quartz (although a very hard material) can be amorphized by a similar process. Often radiolysis can result in the formation of new compounds, which can be identified *in situ* as they form, by electron diffraction and AEM. The formation of Ag from Ag halides in a photographic emulsion is a (rare) example of useful radiolysis. So, somewhat paradoxically, if we use photographic film we depend on radiation damage to record the information we generate in our TEMs (more about this in Chapter 7).

We can't stop radiolysis simply by cooling or coating our specimen, since it isn't affected by heat-transfer considerations. The best way is to lower the cross section for the initiating electron-electron interactions and we can do this by using higher voltages and thinner specimens (thinner is better again). Nevertheless, radiolysis remains a major limitation in the TEM when looking at certain ceramics and minerals and most polymers.

4.6.E Beam Damage in Metals

The primary way that metals are damaged is by knock-on or displacement/sputtering damage. This process occurs by the direct transfer of the beam energy to atoms in the solid, knocking them out of their atomic site and creating a combination vacancy and interstitial (or Frenkel pair). For an atom to be kicked out of its very comfortable and stable crystal lattice site, the beam electron has to penetrate close to the nucleus and be effectively stopped in its tracks by the Coulombic attraction, thus transferring most, if not all, of its energy to the atom.

KNOCK-ON DAMAGE
It is directly related to the incident beam energy.

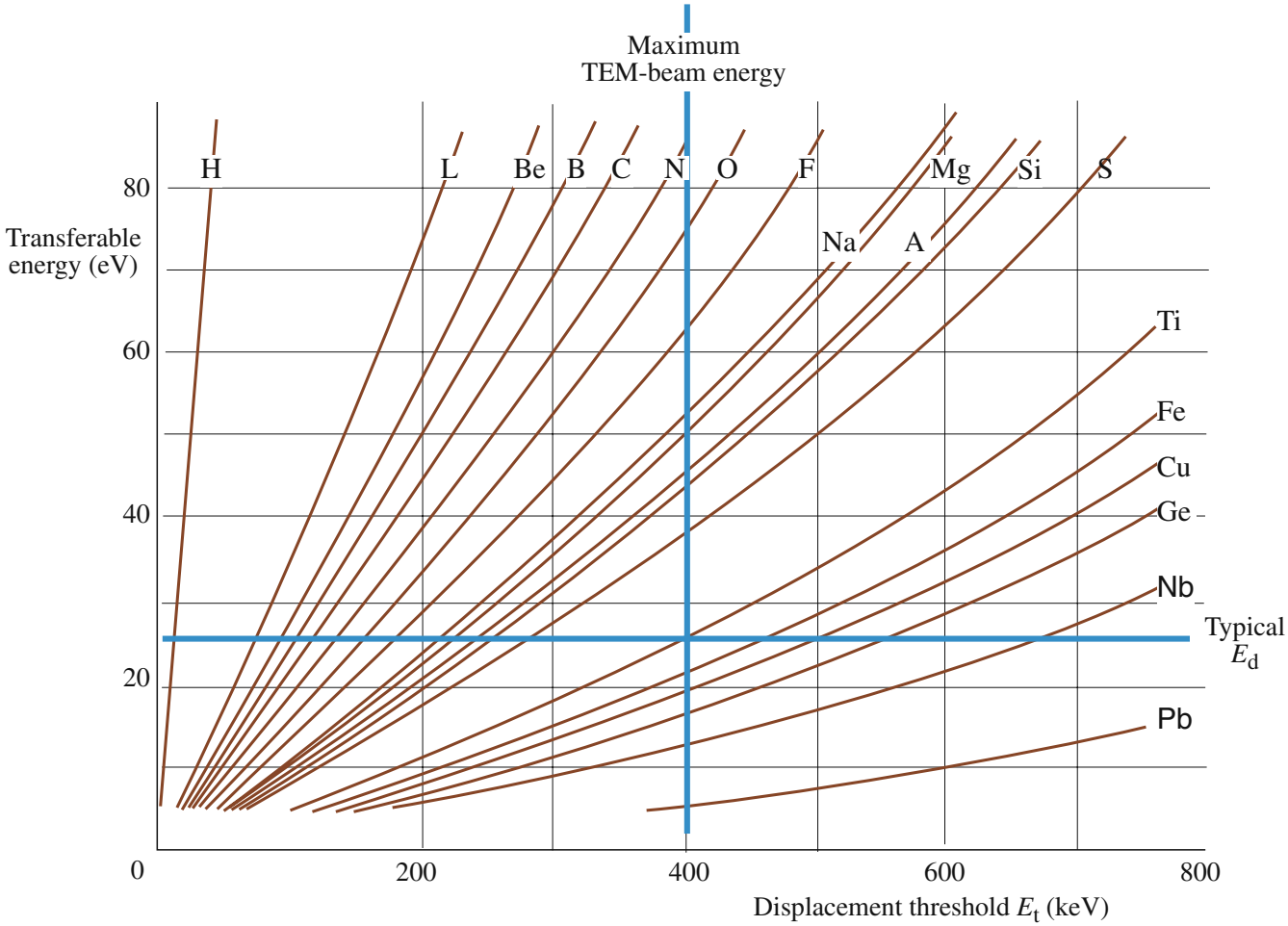


FIGURE 4.12. The maximum transferable energy for a range of atoms as a function of the displacement-threshold energy. A maximum IVEM beam energy is indicated (400 keV) and a typical E_d is shown as ~25 eV, but it can vary substantially with bond strength in different materials. In regions of the graph above 400 keV and below E_d, damage will not occur.

How strongly the atoms are bonded to their neighbors will also be a factor. A simple expression given by Hobbs for the displacement energy E_d allows us to determine the threshold energy (E_t) for displacement of atoms of atomic weight A

$$E_t = \frac{\left(\frac{100 + AE_d}{5}\right)^{1/2} - 10}{20} \qquad (4.9)$$

where E_t is in MeV and E_d is in eV. E_d is typically in the range from 5 to 50 eV, but varies with bonding type. If you can transfer more than the threshold energy to an atom then you will displace it from its site. This concept is summarized in Figure 4.12. So, for example, at 400 keV, $>$ 80 eV can be transferred to a carbon atom, about 45 eV can be transferred to an Al atom, and about 25 eV to a Ti atom. If we assume an average displacement energy of $\sim$ 25 eV then it is quite evident that if you have a 400-kV intermediate voltage TEM, you can displace any atom with an atomic weight below about Ti (unless it happens to be bonded so that E_d is much greater than the typical value in Figure 4.12). If you're using an HVEM with beam energies of 1 MeV or more you will *invariably* cause displacement damage, except perhaps in the heaviest elements. The only way to avoid displacement damage is to operate below threshold and you should determine this energy experimentally given the very approximate nature of Figure 4.12.

How can you identify displacement damage? The only sure way is to record images of the same area before and after radiation and compare the contrast under the same imaging conditions. Knock-on damage often manifests itself as small vacancy clusters which appear as black-white lobe contrast or dot contrast, as we showed back in Figure 1.8. Sometimes damage is discernible as dislocation loops and stacking fault tetrahedra caused by the agglomeration of vacancies. Such crystal defects could of course be easily confused with defects that are intrinsic to the material rather than introduced by the act of observing the materials in the TEM. Displacement damage can also occur in polymers

and minerals, of course. The problem here is that we just suggested going to higher voltages as one way of minimizing thermal effects and radiolysis and in doing so, we may well induce knock-on damage. So depending on your material, there may in fact be no way to avoid damaging your specimen in one way or another except, perhaps, by becoming a metallurgist.

Perhaps the only bright side to displacement damage is that we can study it for its own sake. It can be argued, though by no means conclusively, that electron-beam damage in materials can be equivalent to neutron damage, such as that occurring in nuclear reactors. A general rule of thumb is that a few minutes' exposure in an HVEM is equivalent in terms of damage to many years in a nuclear reactor and so accelerated studies of materials degradation in reactor environments were possible. With this justification, an enormous amount of work was carried out in the 1960s when nuclear power was in vogue. Three Mile Island and Chernobyl seriously reduced the number of such studies but there are many reviews extant in the older literature and an occasional more modern reference. Given that the current political climate (!) is more favorable toward nuclear energy than for some decades, it is not unreasonable to suspect that beam damage might see a resurgence in importance. So, if you haven't settled on a career yet, you may wish to spend a little time studying this field because all known radiation-damage microscopists have retired or are close, and the hard-earned knowledge from the early halcyon years is in danger of being forgotten!

Vacancies caused by displacement damage can enhance diffusion processes which, in turn, can speed diffusional transformations when they're being studied in situ in the IVEM/HVEM. There are many other problems that can arise during in-situ observations, so interpretation isn't always straightforward. The book by Butler and Hale is recommended for more facts, practical hints, and many beautiful images of reactions occurring, in real time, in thin foils. More recently, Gai has edited a volume of contributed chapters on in-situ TEM and there is much excitement in the field about the possibilities for growth of this important area due (not surprisingly) to the advent of C_s correction bringing about TEM stages with larger pole-piece gaps, thus permitting easier insertion of gas-reaction stages.

4.6.F Sputtering

The displacement of surface atoms, or sputtering, occurs in the TEM at voltages which are $<0.5E_t$. If your specimen is quite thick then this problem is minor since the average, through-thickness, characteristics of the specimen are not changed significantly by any changes on the surface. But, as we've already noted many times, your specimen should really be very thin (thinner is better yet again) if you want to get the best images and the best analytical information. In these specimens, modifications of the surface structure and chemistry may be sufficient to affect the image interpretation and/or change the average through-thickness composition enough to affect the accuracy of any quantitative analysis. Table 4.3 lists typical sputtering threshold energies (E_s) compared with displacement thresholds (E_d) and their relative values compared to how much energy can be transferred by electrons of different energies. As you can see, there is cause for concern even with 100-keV beams. A STEM can easily drill a hole through MgO.

TABLE 4.3 Comparison of Maximum Transferable Kinetic Energy (T) at 100, 200, 300 and 400 keV with Displacement (E_d) and Sputtering (E_s) Energies

Element	T (eV)				E_d (eV)	E_s (eV)
	100 keV	200 keV	300 keV	400 keV		
Al	8.93	19.5	31.6	45.3	16	4–8
Ti	5.00	11.0	17.8	25.5	15	4–8
V	4.73	10.3	16.72	24.0	29	7–14
Cr	4.63	10.1	16.38	23.5	22	5–11
Fe	4.31	9.40	15.25	21.8	16	4–8
Co	4.08	8.91	14.45	20.7	23	5–12
Ni	4.10	8.94	14.5	20.8	22	6–11
Cu	3.79	8.26	13.4	19.2	18	4–9
Zn	3.69	8.03	13.03	18.7	16	4–8
Nb	2.59	5.65	9.17	13.2	24	6–12
Mo	2.51	5.47	8.88	12.7	27	7–14
Ag	2.23	4.87	7.90	11.3	28	7–14
Cd	2.14	4.67	7.58	10.9	20	5–10
Ta	1.33	2.90	4.71	6.75	33	8–16
Pl	1.23	2.69	4.37	6.26	33	8–16
Au	1.22	2.67	4.32	6.2	36	9–18

CHAPTER SUMMARY

Inelastic scattering transfers energy to your specimen, generating a lot of useful signals with which we can form different images or get spectroscopic information about the chemistry and electronic structure of the specimen. Much of this information is localized at the nanometer level or below.

- The generation of many different characteristic signals all localized with nanometer-scale (or below) spatial resolution is a most powerful aspect of TEM.

- Unfortunately, these same inelastic processes create beam damage and heat which can be disastrous under certain conditions for all kinds of TEM specimens. To minimize heat transfer, cool your specimen and use higher voltages and thinner specimens but beware:
- If the accelerating voltage is high enough, knock-on and sputtering damage will occur in all materials, creating non-characteristic crystal defects and changing the surface chemistry.

On the brighter side, beam damage can be a positive help in the simulation of nuclear-radiation effects and can also enhance in-situ transformation studies and environmental TEM.

In addition to the following references, many more will be found in the companion text.

IN-SITU

Butler, EP and Hale, KF 1981 *Dynamic Experiments in the Electron Microscope* Practical Methods in Electron Microscopy **9** Ed. AM Glauert Elsevier Amsterdam. Radiation damage plays a major role in many of the topics discussed.

Gai, PL Ed. 1997 *In-Situ Microscopy in Materials Research* Springer-Verlag New York.

Inokuti, M 1971 *Inelastic Collisions of Fast Charged Particles with Atoms and Molecules—The Bethe Theory Revisited* Rev. Mod. Phys. **43** 297–347. A classic and most comprehensive review of progress over the 40 years following Bethe's first cross section description.

Wang, ZL 1995 *Elastic and Inelastic Scattering in Electron Diffraction and Imaging* Plenum Press New York.

BEARDEN'S TABLES AND MORE

Bearden, JA 1964 NYO-10586 US Atomic Energy Commission Oak Ridge TN.

Deslattes, RD, Kessler, RD Jr, Indelicato, P, de Billy, L, Lindroth, E and Anton, J 2003 *X-ray Transition Energies: New Approach to a Comprehensive Evaluation* Rev. Mod. Phys. **75** 35–99.

SOME HISTORY

Auger, MP 1925 *Sur L'Effet Photoélectrique Composé* J. Phys Radium **6** 205–208.

Bethe, HA 1930 *Zur Theorie des Durchgangs Schneller Korpuskularstrahlen Durch Materie* Ann. der Phys. Leipzig **5** 325–400.

Kramers, HA 1923 *On the Theory of X-ray Absorption and Continuous X-ray Spectrum* Phil. Mag. **46** 836.

Moseley, HGJ 1914 *High Frequency Spectra of the Elements* Phil. Mag. **26** 1024–1032.

Williams, EJ 1933 *Applications of the Method of Impact Parameter in Collisions* Proc. Roy. Soc. London **A139** 163–86.

SPECIMEN DAMAGE

Egerton, RF, Li, P and Malac, M 2004 *Radiation Damage in the TEM and SEM* Micron **35** 399–409. More recent paper on radiation damage.

Hobbs, LW 1979 in *Introduction to Analytical Electron Microscopy* Eds. JJ Hren, JI Goldstein and DC Joy Plenum Press New York p437 gives calculations of beam heating.

Jenkins, ML and Kirk, MA 2000 *Characterization of Radiation Damage by Transmission Electron Microscopy* Institute of Physics Bristol and Philadelphia. The only text on this topic; helpful reading if specimen damage affects your work. Includes discussion of knock-on damage.

Sawyer, LC, Grubb, DT and Meyers, GF 2008 *Polymer Microscopy* 3rd Ed. Springer New York. For polymers, of course.

SPECIAL TECHNIQUES

Boyall, NM, Durose, K and Watson IM 2003 *A Method of Normalizing Cathodoluminescence Images of Electron Transparent Foils for Thickness Contrast Applied to InGaN Quantum Wells* J. Microsc. **209** 41–46.

Goldstein, JI, Williams DB and Cliff, G 1986 *Quantitative X-ray Analysis* in *Introduction to Analytical Electron Microscopy* Eds. JJ Hren, JI Goldstein and DC Joy p 155 Plenum Press New York.

Hubbell, JH, Trehan, PN, Singh, N, Chand, B, Mehta, D, Garg, ML, Garg, RR, Singh, S and Puri, S 1994 *A Review, Bibliography and Tabulation of K, L and Higher Atomic Shell X-ray Fluorescence Yields.* J. Phys. Chem. Ref. Data **23** 339–364.

Markowicz, AM and van Grieken, RE 2002 *Handbook of X-ray Spectrometry* Marcel Dekker New York. Excellent text for when you really need more detail on topics such as families of X-rays.

Newbury, DE 1986 in *Introduction to Analytical Electron Microscopy* Eds JJ Hren, JI Goldstein and DC Joy p 6 Plenum Press New York.

Venables, JA, Hembree, GG, Drucker, J, Crozier PA and Scheinfein MR 2005 *The MIDAS Project at ASU: John Cowley's Vision and Practical Results* J. Electr. Microsc. **54** 151–162. Rare example of attaching an Auger system to a STEM.

URLs

1. www.physics.nist.gov/PhysRefData/Ionization/Xsection.html
2. www.physics.nist.gov/PhysRefData/XrayTrans/index.html

SELF-ASSESSMENT QUESTIONS

Q4.1 Distinguish background, continuum and bremsstrahlung X-rays.
Q4.2 'Characteristic' X-rays are characteristic of what?
Q4.3 Why shouldn't we refer to an 'ionized' electron?
Q4.4 Approximately how large is an electron shell relative to an electron?
Q4.5 What is the critical ionization energy?
Q4.6 What is the ionization cross section?
Q4.7 What controls the energy and wavelength of the characteristic X-ray?
Q4.8 What is the fluorescence yield?
Q4.9 What do we mean by the term 'weight of an X-ray line'?
Q4.10 What is overvoltage?
Q4.11 What is the difference in the angular distribution of the ionizing electrons and the emitted X-rays?
Q4.12 Does an X-ray have kinetic energy like an electron?
Q4.13 Why is all the energy transferred to the atom during ionization (E_c) not recovered by the emission of the characteristic X-ray(s)?
Q4.14 How is energy conserved in the overall electron-atom interaction?
Q4.15 Why should you be concerned about fast secondary electrons in TEM?
Q4.16 What is cathodoluminescence?
Q4.17 How can you minimize electron-beam damage to your specimen?
Q4.18 Why does sputtering of atoms from the surface of the specimen take less energy than displacing atoms in the interior?
Q4.19 What is radiolysis?
Q4.20 What's the best reason for using an HVEM to intentionally displace atoms in a specimen?

TEXT-SPECIFIC QUESTIONS

T4.1 Look at Figure 4.1 and explain why
   A. We need to be concerned about plasmon energy losses whenever inelastic scatter compromises the operation of the TEM.
   B. All scattering processes show approximately the same dependence on beam energy.
T4.2 What does Figure 4.1 tell us about the advantages and disadvantages of operating at 400 kV rather than 100 kV?
T4.3 From Figure 4.2
   A. Explain what kind of characteristic X-ray has actually been emitted. (Hint: go to Figure 4.3.)
   B. What happens to the hole left by the electron falling out of the L_3 energy level?
   C. Under what conditions is the presence of the vacuum energy level expected?
T4.4 If, as stated in the text, K-shell X-rays have higher energy than L-shell X-rays and so on, why does the Ag L line (2.3keV) have a higher energy than the Al K line (1.5keV)?
T4.5 Use equation 4.6 to determine the relative fluorescence yields of (a) Be and N and (b) Si and Ag. Cross-check your answer with Figure 4.5. Use your answers to explain why X-ray analysis is challenging at lower atomic numbers and quantification is relatively straightforward for higher atomic numbers. Do Auger electrons have an equivalent 'yield'? If so, derive the approximate dependence on Z.
T4.6 Can you explain why we talk about 'braking' radiation when the nucleus and the electron have opposite charges that one might expect to attract one another very strongly?
T4.7 Looking at Figure 4.6, the generated bremsstrahlung intensity increases for lower X-ray energies, but we have just learned that the characteristic X-ray intensity decreases at lower energies (from lower atomic number elements). Explain this difference.
T4.8 Why are the conduction and valence bands shown separately in Figure 4.8 but not in Figure 4.2?
T4.9 Calculate the electron dose in a 1-nm electron probe containing 1 nA of current.
T4.10 Using Equation 4.9, compare the displacement threshold energies for Li, Al, Cu and Au and estimate the sputtering energies. Compare your answers where possible with the data in Table 4.3 and Figure 4.12.

T4.11 Inelastic scattering ultimately limits the information we can obtain from the TEM because the transfer of energy to the specimen can change its structure or chemistry (the two characteristics that the TEM is best at discerning) either temporarily or permanently. So, why do we take great pains to both maximize the generation and detection of inelastic scattering?

T4.12 Relate the amount of energy lost by the electron to the typical potential-well model of the atom.

T4.13 If you look at Table 4.1 you'll see that the cross sections are smaller for interactions that result in larger energy transfer. Why is this a good thing for TEM operators?

T4.14 Why don't we routinely try to use Auger electrons for surface analysis in the TEM?

T4.15 Based on Figures 4.1, 4.4 and 4.5 explain the most likely signal(s) generated when (a) a 100-kV electron beam strikes specimens of Be and Si and (b) when a 300-kV beam does the same.

T4.16 Compare and contrast plasmons and phonons. (Hint: consider their generation processes, energies, scattering angles, cross sections, range, effects on the specimen.) From your answer discuss their relative importance to the TEM user in terms of the information they contain or destroy.

T4.17 (Extra challenge) If the electron beam sputters Au from the surfaces of a thin film, which surface is affected most?

5

Electron Sources

CHAPTER PREVIEW

A reliable source of electrons to 'illuminate' the specimen is one of the most important parts of a TEM. Fortunately, electron sources are plentiful, but to get the best images and other signals out of our expensive microscope, we need to use the best available source. There are stringent requirements to produce the beam of electrons with the necessary properties and these are best met by only two types of source: thermionic and field-emission sources (or 'guns' as they are often called). Thermionic sources are (now rarely) tungsten filaments or (now commonly) lanthanum hexaboride (LaB_6) crystals, and field emitters are fine tungsten needles. In this chapter we'll first explain briefly the physics of the different electron-emission processes because then you'll understand why we operate the sources in certain ways. Next we'll tell you the characteristics we need from our electron beam. Then we'll compare the sources and show you that no one source is best for all aspects of TEM, but all have their roles. Finally, we'll explain ways to check that a particular source meets your needs.

Because the source is so critical to the performance of the microscope, the technology is advancing rapidly with the aim being to have complete computer control, which would leave you, the operator, with very little to do except push the 'on' button. This state of affairs is most advanced for the field-emission source, and since these are both delicate and expensive, it is just as well. But the majority of TEMs still use thermionic sources, and these may need a fair bit of operator control. In these circumstances, you should know how these sources work and why you do certain things to them. So we'll spend much of this chapter talking about thermionic sources, although field emission is essential for the best performing TEMs of any kind (imaging, analytical, etc.), so there's a good chance that field emission will expand to be the source of choice in the future.

5.1 THE PHYSICS OF DIFFERENT ELECTRON SOURCES

We use two kinds of electron sources in TEMs: the first kind is called a thermionic source, which, as the name suggests, produces electrons when heated, and the second type is a field-emission source, which produces electrons when a large electric potential is applied between it and an anode. Schottky sources combine both heat and field emissions. These sources are part of an assembly which we refer to as the electron 'gun.' Now, from a physics standpoint, it is really quite interesting to know the details of how electron sources work and there's a great deal of active research into new and improved sources, and carbon nanotubes show some real promise. (Like future TV screens but using higher voltages.)

From a practical standpoint, you don't have to know too much about the physics, and we can summarize the essential points very briefly, using two simple equations. Keep in mind two points as you read about sources

- Your TEM will use a thermionic source (W or LaB_6) or a field-emission (W) source and the two cannot be interchanged.
- Field-emission sources give more monochromatic electrons; thermionic sources are less monochromatic and give 'whiter' electrons.

The analogy here is to X-rays or visible light. The 'color' of electrons depends on their energy spread (which translates into a frequency or a wavelength range using equation 1.6); we'll discuss this in Section 5.2. You don't always need to use monochromatic electrons, even if your field-emission TEM did cost twice as much as a conventional microscope would with a thermionic source.

5.1.A Thermionic Emission

If we heat any material to a high-enough temperature, we can give the electrons sufficient energy to overcome the natural barrier that prevents them from *leaking out* from the surface. This barrier is termed the work function (Φ) and has a value of a few eV.

The physics of thermionic emission can be summarized in Richardson's law which relates the current density from the source, J, to the operating temperature, T in Kelvin

$$J = AT^2 e^{-\frac{\Phi}{kT}} \tag{5.1}$$

Here k is Boltzmann's constant (8.6×10^{-5} eV/K) and A is Richardson's constant ($A/m^2 K^2$), which is only constant for a given source material. From this equation you can see that we need to heat the source to a temperature T such that energy $> \Phi$ is given to the electrons; then they will escape from the source and be available to form an electron beam. Unfortunately, when we put a few eV of thermal energy into most materials they either melt or vaporize. So the only viable thermionic sources are either refractory (high melting point) materials or those with an exceptionally low Φ. The source used for the first several decades of TEMs (and still used in some SEMs) was tungsten which melts at 3660 K and the only thermionic source used by modern TEMs is lanthanum hexaboride (LaB_6) which has a low Φ. If you look at Table 5.1, you'll see the relative values of J_c, T, and Φ for tungsten and LaB_6.

We use several different words to describe the sources. We called tungsten sources *filaments*, because tungsten was drawn into fine wire similar to the filament used in an incandescent light bulb. LaB_6 crystals (which should not be called filaments) are usually grown with a <110> orientation to enhance emission. Sometimes we call both tungsten and LaB_6 sources *cathodes* because, as we'll see, the complete gun assembly acts as a triode system in which the source is the cathode.

So all you need to know from the physics is that heating up a thermionic source gives you a higher J. But there is a limit, because higher temperatures shorten the source life through evaporation and/or oxidation. So we seek a compromise operating temperature. We thus operate under a condition called 'saturation' which we'll discuss in 'Thermionic Guns.'

5.1.B Field Emission

Field-emission sources, usually called FEGs (for field-emission guns: pronounced either as 'F-E-Gs' or as 'fegs') operate in a fundamentally different way to thermionic sources. The principle behind FE is that the strength of an electric field E is considerably increased at sharp points because, if we have a voltage V applied to a (spherical) point of radius r, then

$$E = \frac{V}{r} \tag{5.2}$$

We call the fine needles 'tips.' The technique of atom-probe field-ion microscopy (APFIM) is another well-established experimental tool for materials characterization. APFIM uses *specimens* with a very fine needle shape, and so there's a lot of expertise available to help produce FE tips. One of the easiest materials to produce with a fine needle point is tungsten wire which can readily be given a tip radius of < 0.1 μm. If we apply a 1-kV potential to this tip then E is 10^{10} V/m and this lowers the work-function barrier sufficiently for electrons to tunnel out of the tungsten. The tunneling process is the same as you've met in semiconductor devices. Applying such high fields imposes a severe stress on the tip and the material must be mechanically strong to remain intact. FE, like thermionic emission from LaB_6, varies with the orientation of the tungsten crystalline tip; the <310> orientation is found to be best.

For FE to occur, the surface has to be pristine, i.e., it must be free of contaminants and oxide. We can achieve this by operating in ultra-high vacuum (UHV) conditions ($<10^{-9}$ Pa). In this case the tungsten is operated at ambient temperatures and the process is called 'cold' FE. Alternatively, we can keep the surface in a pristine condition at a poorer vacuum by heating the tip. The thermal energy assists in electron emission

TABLE 5.1 Characteristics of the Principal Electron Sources

	Units	Tungsten	LaB_6	Schottky FEG	Cold FEG
Work function, Φ	eV	4.5	2.4	3.0	4.5
Richardson's constant	A/m^2K^2	6×10^9	4×10^9		
Operating temperature	K	2700	1700	1700	300
Current density (at 100 kV)	A/m^2	5	10^2	10^5	10^6
Crossover size	nm	$> 10^5$	10^4	15	3
Brightness (at 100 kV)	A/m^2sr	10^{10}	5×10^{11}	5×10^{12}	10^{13}
Energy spread (at 100 kV)	eV	3	1.5	0.7	0.3
Emission current stability	%/hr	<1	<1	<1	5
Vacuum	Pa	10^{-2}	10^{-4}	10^{-6}	10^{-9}
Lifetime	hr	100	1000	>5000	>5000

so much that, in fact, the electrons don't tunnel through the barrier. For such 'thermal' FE, surface treatments with ZrO_2 improve the emission characteristics, particularly the stability of the source, and such Schottky emitters are the most popular. There are pros and cons for both CFE and TFE, which we'll talk about later in the chapter.

5.2 THE CHARACTERISTICS OF THE ELECTRON BEAM

The electron beam in a TEM requires certain characteristics which are controlled by the source itself and how we integrate the source into a gun assembly. We describe the performance of an electron source by such words as brightness, coherency, and stability. While these words mean something to you already, they have very precise meanings in TEM terminology, so we'll go through the various characteristics, tell you what they mean and why they are important in the TEM. We'll then compare the properties of the various sources that you may have in your microscopes. You'll see that there's no best source for all applications, but for specific applications one source or other is usually better.

Before we define the electron-beam characteristics needed in a TEM, it is worth summarizing a few of the properties of electron beams in general and how these vary with kV.

5.2.A Brightness

The word *brightness* is often confused with *intensity* and indeed the two terms are related. For instance, when we look at the viewing screen of a TEM, we may say how 'bright' it is, when we are really referring to the intensity of light coming from the screen. When we think of the intensity of any radiation source, it is in terms of the flux emanating from it. For a light bulb, it would be the number of photons per unit area per unit time. For electron sources we talk about the current density, which is the number of electrons (or charge) per unit area per unit time.

BRIGHTNESS
While current density can be a useful term, it is more important to define the brightness. Brightness is the current density per unit solid angle of the source.

Electron sources differ considerably in their size and, as a result, the electrons leave the source with a range of divergent angles, so we can't ignore the angular distribution of the electrons. Brightness is particularly important when we are using very small convergent probes, as we do in AEM and STEM. The concept of brightness is less important in conventional TEM where we use a relatively large, defocused beam, but it is still relevant to the intensity we see on the screen, and so it affects how easy it is to operate the microscope and see our images and DPs.

So we can consider an electron source as having the following characteristics

- Diameter d_0
- Cathode emission current i_e
- Divergence angle α_0 (remember when we say angle we mean semi-angle)

We'll describe the actual way in which these characteristics are achieved in Section 5.3, where we discuss the complete gun assembly, but if you look at Figure 5.1 you'll see that i_e, d_0, and α_0 are actually defined at the gun crossover, that is, the point at which the electrons are focused *after* leaving the source. The current density

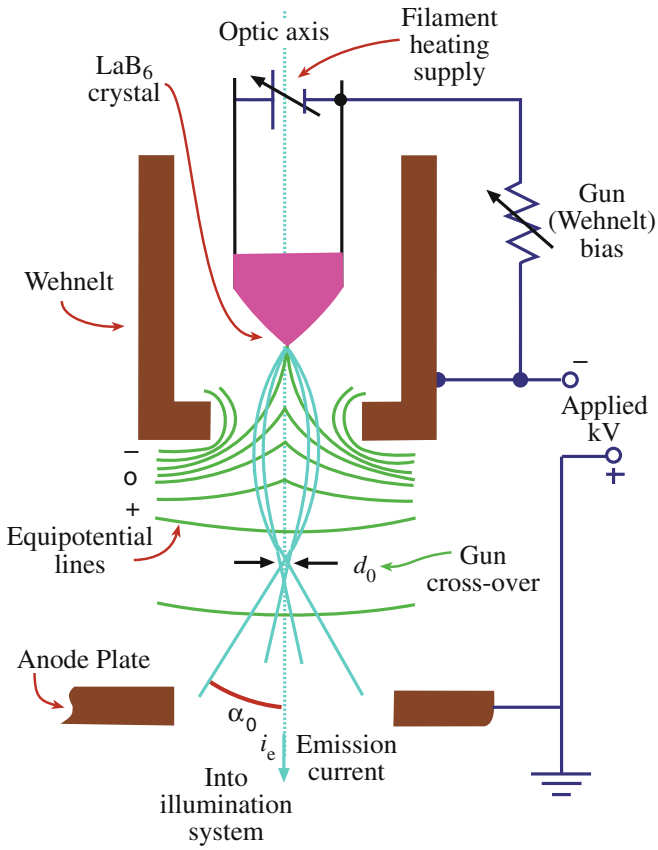


FIGURE 5.1. Schematic diagram of a thermionic electron gun. A high voltage is placed between the cathode and the anode, modified by a potential on the Wehnelt which acts as the grid in a triode system. The electric field from the Wehnelt focuses the electrons into a crossover, diameter d_0 and convergence/divergence angle α_0 which is the true source (object) for the lenses in the TEM illumination system.

(current per unit area) is $i_e/\pi(d_0/2)^2$ and the solid angle of the source is $\pi\alpha^2$, so we define the brightness β as

$$\beta = \frac{i_e}{\pi\left(\frac{d_0}{2}\right)^2 \pi(\alpha_0)^2} = \frac{4i_e}{(\pi d_0 \alpha_0)^2} \tag{5.3}$$

This brightness equation is an important one which you should remember. What is not shown in this equation is the important fact that β increases linearly with increasing accelerating voltage for thermionic sources. This was one reason for the development of intermediate-voltage (300–400 kV) instruments.

Obviously, the higher the value of β, the more electrons we can put into a beam of a given size, and so the more information we can generate from the specimen and also the more we can damage sensitive specimens. The beam current is an important part of the brightness equation. Measuring the beam current in situ can be a very good diagnostic tool. We'll talk about this later in the chapter when we discuss measuring β, but for the time being you can again look at Table 5.1 to see how the various sources compare in brightness.

UNITS FOR BRIGHTNESS

β is in units of $A/m^2\,sr$.

Now we can consider some real numbers. With a cold FEG at 100 keV, we can put 1 nA into an area of diameter 1 nm. If you convert this current density to units of power (1 watt = 1 J/s), you'll find that the energy the electron beam puts into this small area of the specimen is nearly $150\,MW/mm^2$. By comparison, the output of a typical electric power-generating turbine is anywhere from 350 to 1000 MW. If all this energy were in fact absorbed by the TEM specimen the technique would be useless since the specimen would vaporize. We'll find out later why this doesn't happen but clearly we can alter our specimen when we look at it in the TEM, as we discussed in relation to beam damage in the previous chapter. The energy density we just calculated is such that a TEM electron source is the brightest, continuously radiating source known; it is considerably brighter than a supernova.

The brightness is particularly important in AEM, which is the quantitative analysis of the many signals that come from a specimen irradiated by an electron beam, shown back in Figure 1.3. As you'll see in Part 4, we need to put the most beam current into the smallest probe to optimize both spatial resolution and analytical sensitivity. Similarly, as we go to higher magnifications in HRTEM, the intensity of light coming from the viewing screen (see Chapter 7) becomes less because we are viewing only a small fraction of the illuminated area of the specimen. The electron density can be increased by using the brightest source. Then images can be recorded with reasonably short exposure times minimizing image drift and other instabilities. So brighter is better for AEM and HREM.

5.2.B Temporal Coherency and Energy Spread

The coherency of a beam of electrons is a way of defining how well the electron waves are 'in step' with one another. You know that white light is incoherent, because it consists of photons with a range of frequencies (colors), and so to get a coherent beam of electrons we must create one in which all the electrons have the same frequency (i.e., wavelength) just like monochromatic light. We refer to this aspect of coherency as temporal coherency, which is a measure of how similar the wave packets are. If the packets are all identical they have the same coherence length. A definition of the coherence length λ_c is

$$\lambda_c = \frac{v\mathrm{h}}{\Delta E} \tag{5.4}$$

where v is the electron velocity, ΔE is the energy spread of the beam, and h is Planck's constant. This means we must have stable power supplies to the source and a stable high-voltage supply (or high tension, as it is sometimes called for historical reasons) so that all the electrons have a small ΔE, thus giving a well-defined wavelength. We'll show in detail in Section 37.7 how we can add an energy-selecting spectrometer to the gun to choose electrons with an energy-spread of as little as 10 meV. Such monochromators are expensive and cut down the total beam current tremendously but, for certain very specialized applications, they are invaluable. This loss of current can be offset somewhat by C_s correction (even more expense) and instruments with combinations of monochromators and C_s corrector are the most expensive (and rare) TEMs in the world. If you look at Table 5.1 you'll see that, without a monochromator, typical ΔE values for the three sources are in the range 0.3–3 eV (which is still remarkably small compared with a total energy of 100–400 keV). So it isn't really correct to imply as we did at the start of the chapter that thermionic sources give 'white' electrons since ΔE is so very small. From these values of ΔE, if you take care to get the units consistent, you can calculate typical coherence lengths, which turn out to be a few hundred nanometers.

Temporal coherency is important when the energy spread of the electrons that are *incident* on the specimen affects the microscopy. Because we can make such good high-voltage power supplies, the incident electron-energy spread rarely limits any aspect of TEM except the highest energy-resolution EELS (see Chapters 37–40). In other words, for most practical purposes our electron sources are stable enough. However, we'll see that it's a very different matter when we have to consider the electrons that have come *through* the specimen because they may have lost substantial amounts of

energy and that's where energy-filtered TEM technology really comes into its own (see Section 37.6).

5.2.C Spatial Coherency and Source Size

Spatial coherency is related to the size of the source. Perfect spatial coherence would imply that the electrons were all emanating from the same point at the source. So source size governs spatial coherence and smaller sources give better coherency (just as they give higher brightness). The spatial coherence is strictly defined by looking at electron-interference fringes in the equivalent of a Fresnel biprism experiment in light optics, with which you may be familiar. We can define the distance d_c, the effective source size for coherent illumination, to be

$$d_c = \frac{\lambda}{2\alpha} \tag{5.5}$$

where λ is the electron wavelength and α is the angle subtended by the source at the specimen. We can control α by inserting an aperture in the illumination system, as we'll see when we describe the construction of a TEM in Chapter 9. But if this aperture is not limiting then it is the smallest source which subtends the smallest angle, and thus has the highest spatial coherence. Putting reasonable values for 100-keV electrons into equation 5.5 we find that the spatial coherence is at best only about a nanometer. To maximize the coherency, you can choose several approaches

- Make the source size, d_c, smaller, e.g., by using a FE source. This explains why research into using nanotubes as electron sources is ongoing.
- Use a smaller illumination aperture, thus reducing α.
- If your source size is large (e.g., a W hairpin) decrease the accelerating voltage and thus increase λ.

A small electron source subtends a small angle at the specimen, and we can help by using small limiting apertures. Small beams are more spatially coherent than large beams and give better spatial resolution of analysis (see Part 4). The more coherent and parallel the beam, the better the quality of the phase-contrast images (Part 3), the sharper the DPs (Part 2), and the better the diffraction contrast in images of crystalline specimens (Part 3). So that's why spatial coherence is important. The whole concept of coherence is rather more complex than we have described here and an in-depth and rather mathematical description of electron coherency in the TEM is given in the review by Hawkes and in the companion text.

COHERENCY

Spatial coherency is more important practically than temporal coherency; smaller source –> higher β, better spatial coherency, but lower stability.

5.2.D Stability

In addition to the stability of the high-voltage supply to the source, it is also important that the electron current coming from the source is stable. Otherwise the screen intensity would vary, making it difficult for you to take correctly exposed images and also making quantitative analytical measurements impossible. Thermionic sources are generally very stable except when they are first installed or when they are about to fail. Typically, you can expect variations of $< \pm 1\%$/hr in the current and TFE sources are similarly stable. For CFE sources, however, the emission current is not very stable, and electrical feedback circuits are required to maintain stability to $<\pm 5\%$. Stability improves with better UHV conditions in the gun.

To summarize: the important properties of electron sources are their brightness, temporal coherency, energy spread, spatial coherency and stability. A smaller source size gives higher β and better spatial coherency, but less stability.

Now that we know the critical characteristics required of electron sources, let's examine those used in commercial TEMs.

5.3 ELECTRON GUNS

It's no good just having a source. We need to be able to control the electron beam and direct it into the illumination system of the TEM. We do this by incorporating the source into a gun assembly which, in effect, acts as a lens to focus the electrons coming from the source. The design of the gun is different for thermionic sources and FE sources.

5.3.A Thermionic Guns

LaB_6 is the only thermionic source used in modern TEMs so we'll just describe these. The LaB_6 crystal is used as the cathode in a triode gun shown in Figure 5.1. In addition to the cathode, there is a grid called a Wehnelt cylinder and an anode at earth potential with a hole in its center. What these three components look like in practice is shown in Figure 5.2, where they are all separated. The cathode is attached to the high-voltage cable that, in turn, connects to the high-voltage power supply. The LaB_6 crystal is bonded to a metal wire such as rhenium, which is resistively heated to cause thermionic emission.

When the electrons leave the cathode they have a negative potential of whatever accelerating voltage you have chosen (say 100 kV) with respect to the earthed anode. So they accelerate through this potential difference, acquiring an energy of 100 keV and a velocity greater than half the speed of light.

Now to get a controllable beam of electrons through the hole in the anode and into the microscope itself, we apply a small negative bias to the Wehnelt cylinder. The

FIGURE 5.2. The three major parts of a thermionic gun, from top to bottom: the cathode, the Wehnelt cylinder and the anode shown separated. The Wehnelt screws into the cathode support and both are attached to the high-voltage cable which also contains power supplies for heating the cathode and biasing the Wehnelt. The anode sits just below the Wehnelt and the whole assembly sits on the top of the column of lenses that make up the rest of the TEM.

BEAM CURRENT

As the cathode heating current (i_f) increases, T increases until thermionic emission occurs; then an emission current from the cathode i_e can be measured. Sometimes you'll find this current referred to as the beam current. This is misleading; the true beam current is that which enters the specimen after the electrons have left the gun and gone through the illumination system of the TEM.

electrons coming off the cathode see the negative field and are converged to a point called a crossover between the Wehnelt and the anode as shown in Figure 5.1. We could operate the controls for the cathode heating and the Wehnelt bias independently, but the electronic circuitry of the gun is designed so that as the emission current increases, the Wehnelt bias increases; this arrangement is called a self-biasing gun. The result is shown in Figure 5.3, which plots the emission current (i_e) against the current used to heat the cathode (i_f). As you can see, i_e reaches a maximum such that a further increase in i_f doesn't increase the current going into the TEM column. This is the *saturation condition* and all thermionic sources should be operated at or just below saturation. Operating above saturation reduces the source life without any compensating advantage; operating significantly below saturation reduces the current into your specimen, thus reducing the intensity of all the signals coming out of your specimen although on occasions, as we'll see later, undersaturation can be useful.

WEHNELT

The Wehnelt acts as a simple electrostatic lens: the first lens the electrons go through in the TEM.

In addition to optimizing the source life, operating at saturation also optimizes brightness. If you look at Figure 5.1, the crossover is the source size d_0 that we used back in the brightness equation (equation 5.3) and the divergence angle at the crossover is α_0 in that same equation. The current in the crossover is the emission current i_e. Now, as shown in Figure 5.4A, if the Wehnelt bias were too low (diagram i) d_0 would not be very small and if the bias were too high (diagram iii) the cathode-emission current would be suppressed. In either case β would be low. The optimum β is at an intermediate-bias setting (diagram ii), as summarized in Figure 5.4B. You would be right if you thought that the small bias on the

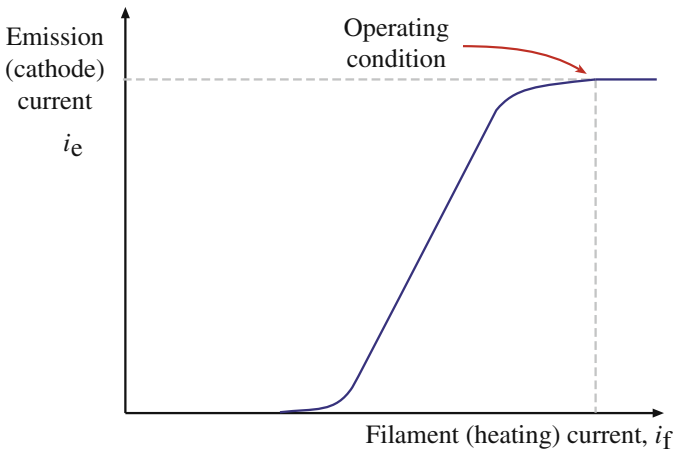


FIGURE 5.3. The relationship between the current emitted by the electron source and the source heating current for a self-biasing gun. Increasing the source current results in a maximum emission current termed saturation.

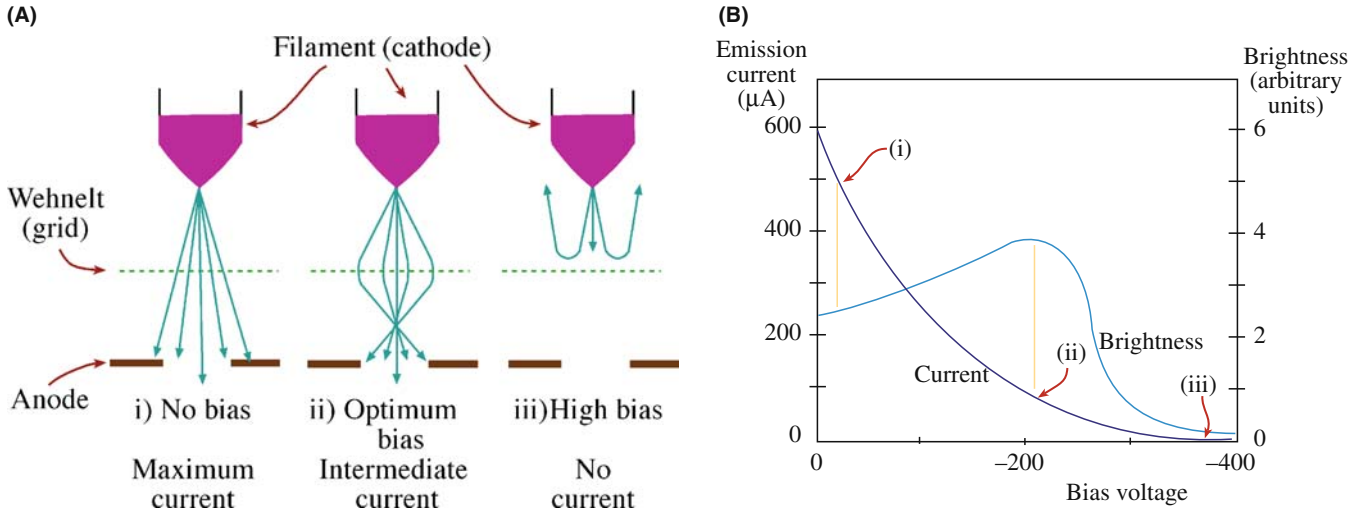


FIGURE 5.4. (A) The effect of increasing Wehnelt bias (i–iii) on the distribution of electrons coming through the anode. (B) The relationship between the bias and the emission current/gun brightness. Maximum brightness (ii) is achieved at an intermediate Wehnelt bias, and an intermediate emission current (ii).

Wehnelt acts against the accelerating voltage, so the true beam voltage should be the applied kV minus the Wehnelt bias (which may be up to 2 kV), but this is compensated for in the design of the gun, so don't worry.

So how do we achieve saturation? One way is to look at the meter which displays i_e and watch it rise to a maximum as i_f is continuously increased. This method may not be easy because the appropriate readouts may not be available or, if they are, they may not be very sensitive. So the standard way is to look at the image of the source crossover on the TEM screen. This image shows you the distribution of electrons coming off the source. As thermionic emission starts, the electrons may come from both the central tip and/or a region surrounding the tip of the crystal. Since LaB_6 sources have well-defined crystal facets (Figure 5.5A) the undersaturated image is as shown in Figure 5.5B. With increasing emission the halo collapses in on the central bright disk, although some structure may still be visible. The cathode is truly saturated when no structure is visible (Figure 5.5C).

It is best to operate an LaB_6 source just below saturation, since this will extend the source life without undue loss of signal. The electrons in the halo are more coherent than those in the central bright region and have reduced energy spread. LaB_6 crystals are susceptible to thermal shock which can cause them to break, and so you should take care when *heating* and *cooling* an LaB_6 source (often the TEM computer takes control of this). If you have to switch the source on and off manually then increase/decrease the heating current slowly, with 10–20 seconds pause between each setting. This is particularly critical after you've installed a new LaB_6 source.

The appearance of the image of the source, as in Figure 5.5, can also be used to align the gun assembly so that the beam is aligned along the optic axis of the TEM. This is the only other adjustment you have to do to the gun, apart from saturating it. The source is usually pre-aligned by the manufacturer, so alignment should be simple when it is put inside the Wehnelt. Typically, a misaligned, undersaturated source image is asymmetrical as in Figure 5.5B and, in those circumstances, all you have to do is adjust the gun to make it symmetrical prior to final saturation. Detailed instructions will be in the manufacturer's handbook. Most modern TEM guns are so well constructed that slight electronic corrections are all that you need to ensure alignment.

SMALL PROBES - FINE BEAMS

Achieving optimum β is critical in any operations that require a fine beam (<0.01 μm).

In an SEM, which always requires a small probe rather than a broad beam, the gun is carefully adjusted by the manufacturer to produce optimum β at saturation; you may not have any external control of the Wehnelt. In a TEM, particularly when you are operating in a broad-beam mode, there is no need to optimize β, but you may need to increase the current density and make the image appear brighter. You can do this by decreasing the Wehnelt bias, using the gun-emission control. When you decrease the bias, you should go back and adjust i_f to ensure you're at saturation, since the saturation condition will change with changing bias. So now you will have a greater current density falling on the screen, but the crossover size will have increased, thus decreasing β. This

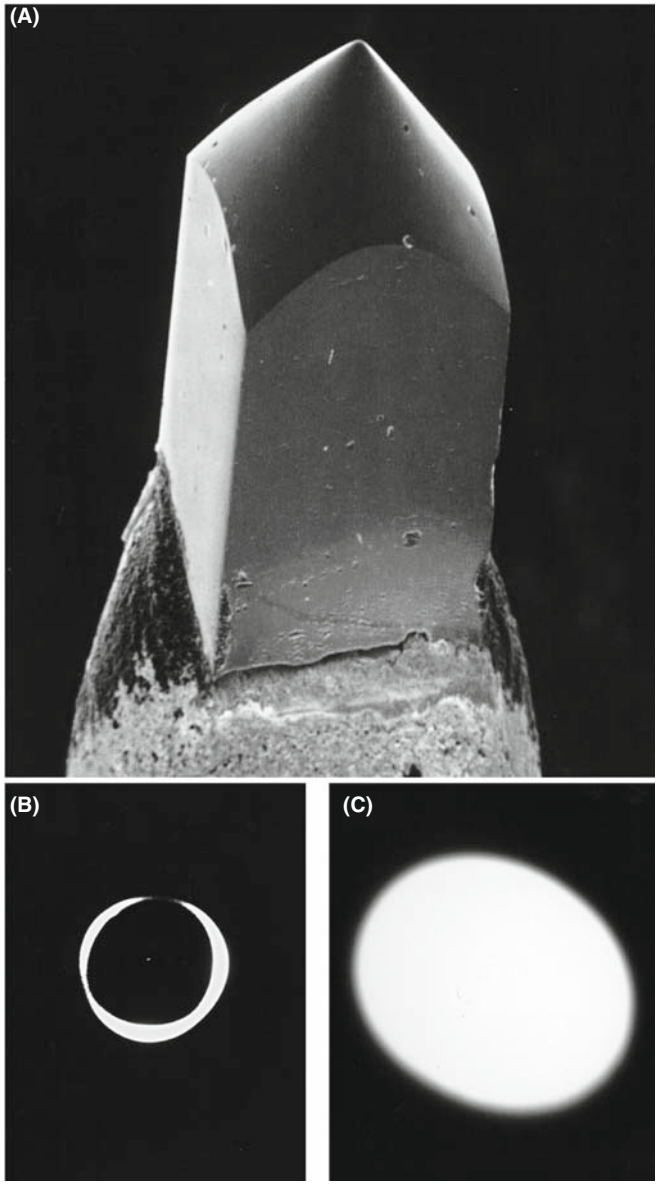


FIGURE 5.5. (A) An LaB_6 crystal and the electron distribution when the source is (B) undersaturated and (C) saturated.

is not important when you're operating with a broad beam. However, if you want to operate at maximum β with a focused beam, as is the case for AEM, then you need to be able to measure β; we'll show you how to do this in Section 5.5.

5.3.B Field-Emission Guns (FEGs)

In many ways, FEGs are much simpler than thermionic guns. In order to get a FEG to work we make it the cathode with respect to *two* anodes. The first anode is positively charged by several kV with respect to the tip. This charge produces the extraction voltage since it generates the intense electric field which enables electrons to tunnel out of the tip. Increasing the extraction voltage when you first switch on the gun has to be done slowly, so the thermo-mechanical shock doesn't fracture the tip. This is the only practical step you have to carry out to run a FEG; it is invariably computer controlled.

- Anode 1 provides the extraction voltage to pull electrons out of the tip.
- Anode 2 accelerates the electrons to 100 kV or more.

The electrons are accelerated through the appropriate voltage by the second anode. The combined fields of the anodes act like a refined electrostatic lens to produce a crossover, as shown in Figure 5.6A. This

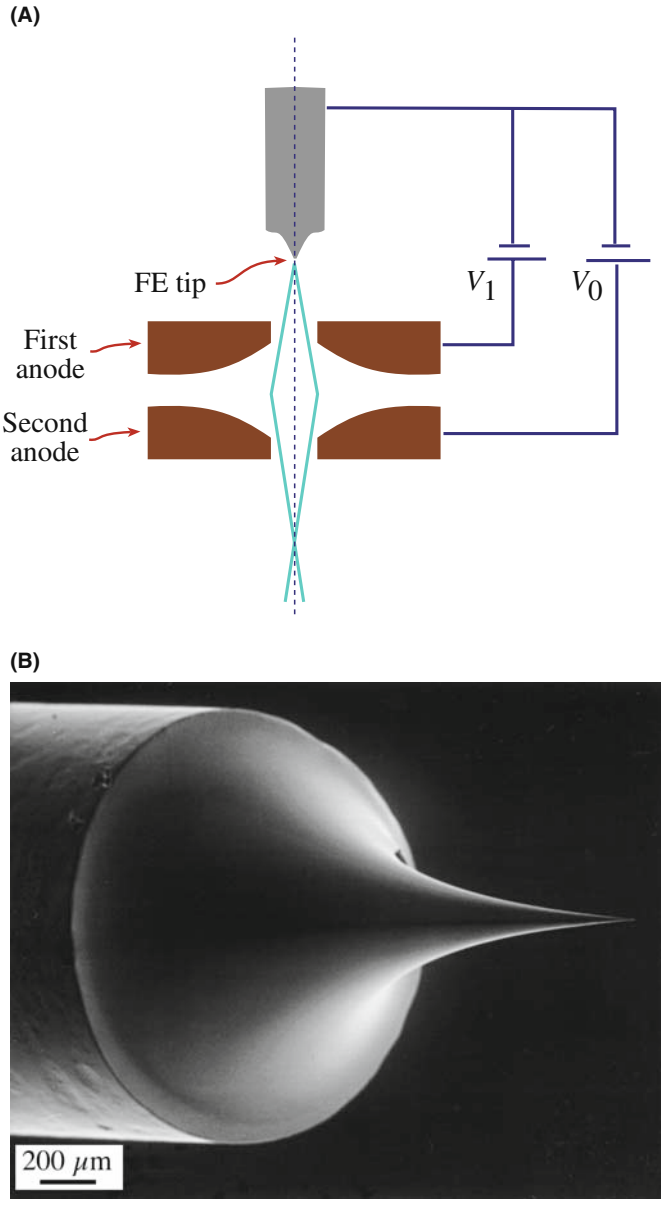


FIGURE 5.6. (A) Electron paths from a field-emission source showing how a fine crossover is formed by two anodes acting as an electrostatic lens. Sometimes an extra (gun) lens is added below the second anode. (B) A FEG tip, showing the extraordinarily fine W needle.

lens controls the effective source size and position, but it isn't very flexible. Incorporating a magnetic lens into the gun gives a more controllable beam and larger β. The faults (known as lens aberrations) in the gun lens are very important in determining the source size; we'll talk extensively about lens aberrations in Chapter 6.

VACUUM IS IMPORTANT

In a vacuum of 10^{-5} Pa, one monolayer of contaminants will form on a substrate in less than a minute. At 10^{-8} Pa, it will take 7 hours to form a monolayer.

We have already noted that CFE requires a pristine surface and, even in UHV conditions, surface contaminants build up on the tip. With time, the emission current falls and the extraction voltage must be increased to compensate. Eventually it becomes necessary to remove the contamination by 'flashing' the tip. This just means reversing the potential to the tip and 'blowing off' the surface layer of atoms, and/or heating the tip quickly to ~5000 K to evaporate the contaminants. In most CFE guns flashing occurs automatically, when the extraction voltage increases to a certain predetermined level. Because they are continuously heated, Schottky FEGs do not form the same surface contamination layer and so don't need flashing. A typical FEG tip is shown in Figure 5.6B.

A future generation of ultra-bright field-emission electron sources might well be based on carbon nanotubes which have already demonstrated a brightness in excess of 10^{14} A/m^2 sr although such a laboratory demonstration is a long way from reliable service in a commercial IVEM in a student-oriented lab. It is perhaps more likely that this technology will be developed for tips in scanning-probe microscopes or electron emitters for flat-panel display technology well before it becomes an electron source in a TEM.

5.4 COMPARISON OF GUNS

The important characteristics of the three guns we've talked about are summarized in Table 5.1. For historical reasons, we've included tungsten sources which are the worst in most respects (except price).

LaB_6 is a much more useful source for several reasons. While it is not as refractory as tungsten, LaB_6 has a much lower value of Φ and, since Φ appears in the exponential in the Richardson equation, its effect on the current density is dominant. LaB_6 crystals can be produced with a fine tip about 1 μm in radius, which accounts for the smaller crossover size. As a result, LaB_6 current densities are considerably higher than for tungsten and the brightness is typically 10 times greater, even though LaB_6 is usually operated at a much lower T to increase operating life. The decreased source size also results in improved coherency and the energy spread approaches 1 eV.

Because LaB_6 is highly reactive, the gun vacuum has to be good enough to minimize oxidation during operation, thus ensuring a reasonable life expectancy. Anything that requires improving the vacuum is good since better vacuums improve most aspects of TEM performance, but this improvement comes at a price.

The increased brightness, higher coherency, and longer life are tremendous advantages and explain why the only thermionic sources we should now use in TEM are LaB_6. You, as the operator, may have considerable control over its performance and unless the computer control overrides you, careless heating, cooling, and oversaturation can easily destroy a LaB_6 crystal. So treat LaB_6 sources gently and you will be well rewarded. If users are not careful, your TEM supervisor may extend the life of the LaB_6 to the point where it behaves no better than a W filament. LaB_6 sources don't die, they fade away.

In FEGs, the current density is enormous and β is correspondingly high. The values in Table 5.1 are all for 100 kV-accelerating voltage and you should remember that for thermionic sources, β increases linearly with kV, so there are advantages to using 300- and 400-kV instruments. However, the LaB_6 source brightness at 400 kV still does not approach β of a FEG at 100 kV. So for all applications that require a bright, coherent source, the FEG is best. This is the case for AEM, HRTEM, and such special applications as electron holography and Lorentz microscopy (for looking at magnetic domains). However, as we'll see later, the coherence of the source may produce a new complication: we must interpret the image!

There are significant differences between a CFE and a Schottky (TFE) sources and, depending on what you need from your TEM, one or the other may be significantly better. First of all, the extremely small source size of a cold FEG means that the beam is highly spatially coherent and the resulting energy spread is the smallest available without monochromation. Thermally assisted Schottky FEGs have a somewhat larger source size and larger energy spread but they provide greater stability of beam current and lower noise. A CFE requires UHV; such technology is expensive and generally requires a higher level of operator competence. However, a UHV brings other advantages such as a cleaner specimen stage and reduced contamination from the microscope system. Cleaning the tip by heating as we do for a Schottky rather than flashing as in a CFE lowers the stress on the tip, ensuring a longer life. But the cleaner UHV system for a CFE gun also

ensures a long life and so both FE sources offer similar lifetimes measured in the thousands of hours (if you're careful!). In summary, if you are doing EELS and need the lowest possible energy spread or you are doing the highest resolution STEM imaging and X-ray analysis in which the highest brightness and smallest probe size are required, then a cold FEG has advantages. For routine FEG work a Schottky gun is better, more reliable, and easier to operate.

Lastly, we should note that for routine, relatively low magnification (<50–100,000×) TEM imaging, a FEG is far from ideal because the source size is too small. It is thus not possible to illuminate large areas of your specimen at low image magnifications without losing current density, and therefore intensity, on the screen. So you can't see your image clearly at low magnifications. Under these circumstances, a LaB_6 source is better. However, increased computer control and the need for optimal performance means that FEG TEMs are increasingly popular and for the best high-resolution imaging and analytical performance, as we've made clear, there is no alternative. If you still don't get it, there's a good interactive summary of both thermionic and FE electron guns and how they work on URL #1.

5.5 MEASURING YOUR GUN CHARACTERISTICS

This section requires that you know how to operate a TEM. If you're a novice, you should skip this part of the chapter for now because we are going to refer ahead in the book for much of what you need to know.

For conventional TEM imaging and diffraction and many other routine uses, all you need to do is saturate and align the (thermionic) gun or just switch on the FEG and then ignore it. In many instruments the computer takes care of this. There are, however, times when we need to be able to measure the brightness and coherency. The source brightness is a most important parameter to measure in an AEM since, if the gun is not operating at its maximum β, then the quality of the analytical information that is generated will be poorer. Similarly, knowing the energy spread of your source is important for EELS and having a measure of the beam coherency can be important for some more advanced techniques that we've just mentioned. So let's see how we can measure the various parameters that we've just discussed. We'll start with β, then ΔE, and finally the coherency.

By measuring the three variables in equation 5.3, i.e., the beam current, the beam diameter, and the angle of convergence, we can determine β. However, while we can easily get a measure of the emission current at the gun, it is impossible to measure d_0 and α_0 there (think about why this is so). So we make the approximation that, if we neglect lens aberrations, β is constant throughout the electron-optical system so it doesn't matter where it is measured. It is easiest, practically, to determine β at the plane of the specimen and we'll now show you how to do this. (Neglecting lens aberrations is reasonable but you should be aware that C_s correction in TEMs can effectively increase the brightness of the electron beam *at the specimen* since we can use larger apertures to permit higher currents in the probe without broadening the probe dimensions.)

5.5.A Beam Current

You can measure the beam current at the specimen i_b directly using a Faraday cup in a specimen holder. A Faraday cup consists of a small aperture above a relatively deep hole in an earthed metal block. If the aperture is small enough (e.g., ~50 μm) and the metal block deep enough (~2 mm), and made of something light like Al to minimize backscatter, then it is a reasonable assumption that no electrons escape back out of the entrance aperture. All the electrons going into the aperture therefore go to earth, and you can measure the electron current using a picoammeter in the earth line. (Ideally a Faraday cup would be available permanently in the column of a TEM, and this would permit constant monitoring of the beam current but no TEM manufacturer offers such.) You should calibrate your Faraday-cup measurement against the TEM screen exposure meter or the electron energy-loss spectrometer shield current. Carrying out this procedure allows you to make a rapid estimate of i_b at any time you need it.

In modern TEMs equipped with a Schottky FEG, the beam-current fluctuation should be less than a few percent over many hours of operation. The stable beam current in Schottky-FEG TEMs does not need to be monitored frequently and can be calibrated easily through the readout from the viewing screen. Conversely, the beam current decreases with time in cold-FEG TEMs. Figure 5.7 shows the time dependence of the emission and the beam current measured after tip-flashing in the cold-FEG STEM. While the emission current decreases almost linearly, the beam current drops parabolically up to 3 hours after flashing. In thermionic-source TEMs, the beam current also decreases after stabilization. This variation would not be nearly so large for a Schottky FEG which, of course, never needs flashing.

MEASURING THE BEAM

A Faraday cup is a black hole for electrons and a very useful diagnostic tool for your TEM performance.

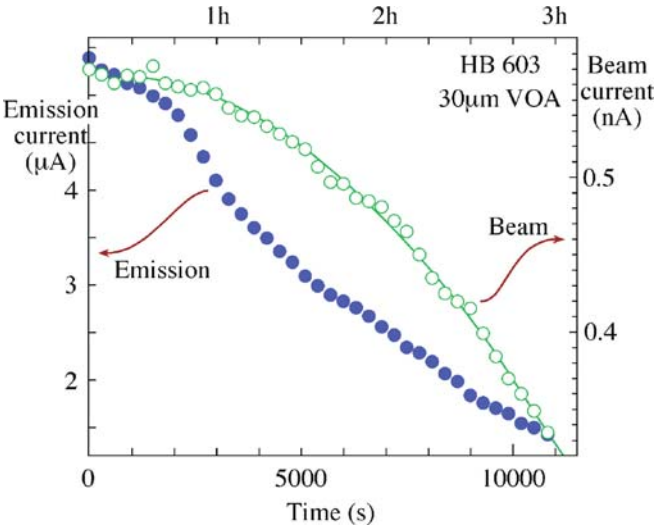


FIGURE 5.7. Time dependence of the emission and the beam currents of a cold FEG in a 300-keV VG HB603 STEM. Both the currents were measured after flashing the tip.

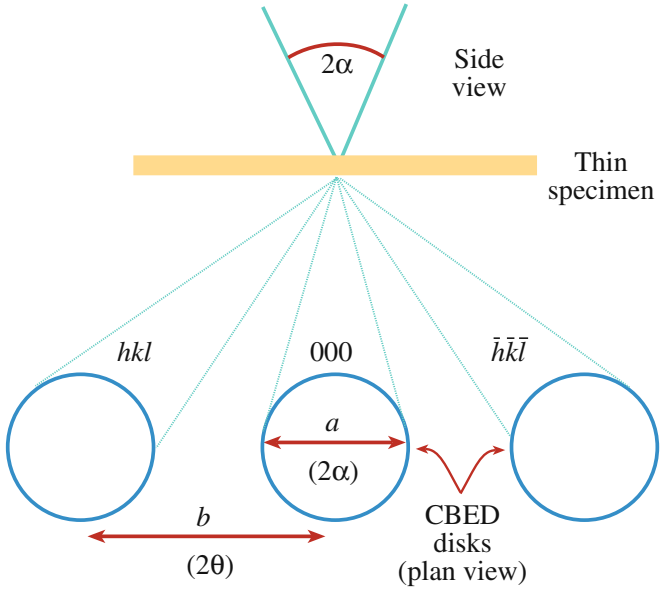


FIGURE 5.8. The distances on a convergent beam DP from which you can measure the beam convergence angle, α, which is proportional to the width of each diffraction disk.

As we'll show in Chapter 6, i_b is a strong function of the beam size. Therefore, the current is controlled by the first condenser (C1) lens strength and the size of the final beam-limiting aperture in the second condenser (C2) lens. If you look ahead to Figures 9.10 and 9.11 you will see the variation of i_b as a function of C1 lens strength and the effect of C2 aperture size on α_0.

- The beam current is usually in the range from nanoamps to picoamps.
- The emission current is typically several microamps.

So the current decreases by three to six orders of magnitude between the gun and the specimen: most is lost in the illumination system, as we'll see in Chapter 9.

5.5.B Convergence Angle

You can easily measure the convergence angle α from the convergent-beam electron diffraction (CBED) pattern, which you can see directly on the TEM screen. (You will need to read Chapter 21 in order to find out how to generate CBED patterns.) In the schematic diagram in Figure 5.8, the total convergence angle 2α is proportional to the width of the diffraction discs, *a*. This width can easily be calibrated if the specimen has a known Bragg angle $2\theta_B$ (see Chapter 11), since $2\theta_B$ is proportional to the distance, *b*, from the 000 disc to the *hkl* disc. Thus

$$2\alpha = 2\theta_B \frac{a}{b} \tag{5.6}$$

The convergence angle, α, at the specimen is not only important in the brightness equation, but we'll see that it also plays a major role in CBED, STEM imaging, XEDS, and EELS. So you must know how to measure and control α because it is essential in many aspects of TEM. The value of α is controlled by the size of the final limiting aperture in the illumination system and we'll see how this works in Chapter 6.

5.5.C Calculating the Beam Diameter

While it is a relatively simple matter to measure i_b and determine α, the measurement of *d*, the beam diameter, is not so straightforward. However, *d* is a major factor in all aspects of TEM such as AEM and STEM imaging where we use a fine focused beam. We can either calculate *d* or measure it experimentally. The former is easy but imprecise, the latter is difficult and can be equally imprecise. The first problem with determining *d* is that there is no universally accepted definition of the beam diameter. The manufacturer will give you a list of nominal beam sizes for each setting of the C1 lens. These values are *calculated* and may differ from the actual beam size by large amounts. The calculation assumes that the electron-intensity distribution in the beam is Gaussian, and the beam diameter is defined as the full-width at half-maximum (FWHM) of the Gaussian distribution, defined in Figure 5.9. To approach a Gaussian intensity distribution, the beam must be well aligned, any astigmatism in the condenser lenses corrected (see Chapter 9), and all apertures in the illumination system accurately centered. Even under these conditions you cannot obtain Gaussian conditions for

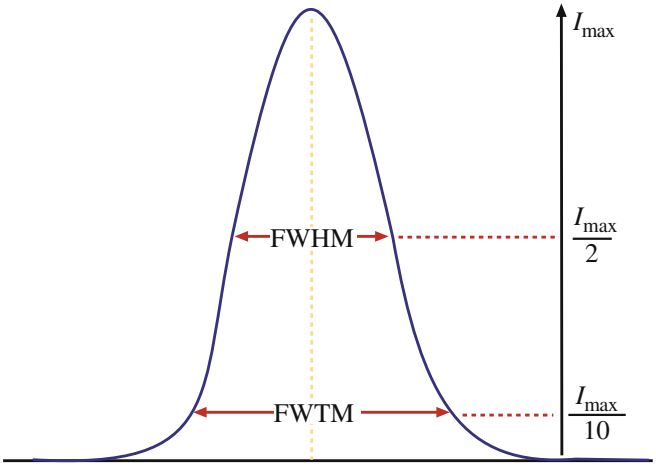


FIGURE 5.9. The definition of the full width at half maximum (FWHM) and the full width at tenth maximum (FWTM) of a Gaussian intensity distribution which is typical of a well-aligned electron beam. Ideally the beam hitting your specimen should always approximate to this kind of intensity distribution.

every possible beam size. For example, there may be six different C1 lens excitations, each of which gives a different calculated beam size, but there are invariably fewer than six C2 apertures available, so each beam size cannot be correctly apertured; spherical-aberration (C_s) effects will broaden the beam size beyond a true Gaussian (see Chapter 6). If you select too small an aperture, then the intensity distribution will be truncated at a fraction of the full Gaussian curve. If you use too large an aperture, the actual beam will extend out well beyond the calculated size and this has very important implications for XEDS in the TEM as we describe in Section 33.3.A.

To make a complete calculation of the beam size, we assume that it is determined by an initial Gaussian diameter at the gun (d_g). This diameter is broadened by the effects of spherical aberration in the beam-forming lens (d_s) and diffraction at the final aperture (d_d). All these terms can be added in quadrature (although for no better reason than that it seems reasonable) to give a total, calculated beam size, d_t

$$d_t = (d_g^2 + d_s^2 + d_d^2)^{1/2} \tag{5.7}$$

This equation gives us only a first-order estimate, since the contributions are not all Gaussian. We'll now briefly discuss the origin of each of these terms.

The value of d_g is a function of β, and a value of β has to be assumed for the purposes of calculation. The expression for d_g is

$$d_g = \frac{2}{\pi}\left(\frac{i}{\beta}\right)^2 \frac{1}{\alpha} \tag{5.8}$$

We have already defined i, β, and α.

The disc of minimum confusion caused by spherical aberration has a diameter given by

$$d_s = 0.5 C_s \alpha^3 \tag{5.9}$$

where C_s is the spherical-aberration coefficient, which we discuss in detail in Chapter 6. This is the full diameter containing 100% of the beam current. Clearly, this term is not Gaussian unless the beam is correctly apertured, which, as we just discussed, is not always possible. However, C_s correction is now available in probe-forming TEMs and then, of course, this contribution to the beam broadening goes away. The calculated diameter due to diffraction is

$$d_d = 1.22 \frac{\lambda}{\alpha} \tag{5.10}$$

which is the Rayleigh criterion that we discussed in Section 1.2.B and refers to a spacing between two overlapping images of the probe. Although all these definitions clearly do not define the same diameter of the electron distribution, the summation in quadrature is still assumed to give a first approximation of the FWHM of the beam. Figure 5.10 shows the result of calculations of the three contributions to the beam diameter in a VG HB501 STEM and the comparison with two experimental measurements carried out, as we'll now describe.

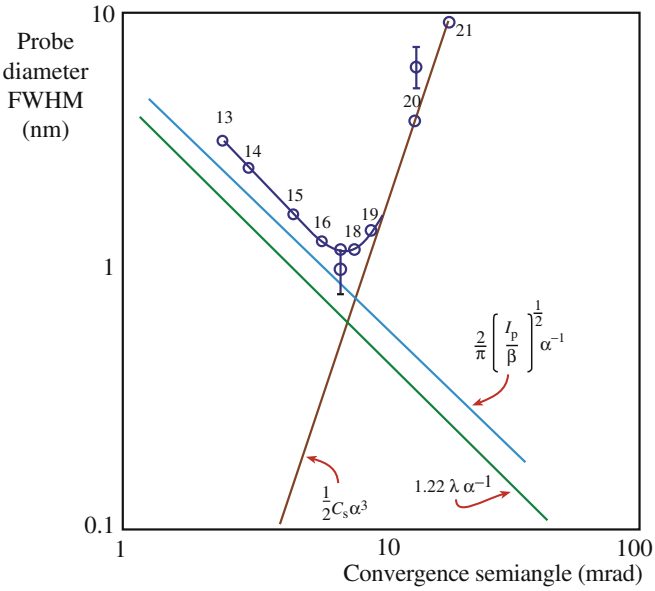


FIGURE 5.10. Calculations of the various contributions to the beam size as a function of the convergence angle α, in a FEG STEM with a probe current I_p of 0.85×10^{-8} A. Two experimental points are shown (with error bars) at first-condenser lens settings 17 and 20 (corresponding close to the minimum and the maximum probe sizes, respectively). The minimum diameter is ~ 1 nm with $\alpha < 10$ mrads.

Given all the (somewhat inaccurate) assumptions we made, the agreement with experiment is quite reasonable.

5.5.D Measuring the Beam Diameter

Given the uncertainties in the calculation of the probe size which we just described, it would seem much more reliable to measure d experimentally. To measure the beam size in a TEM you must form an image of the beam on the viewing screen or computer display under conditions where you know, or can calibrate the magnification. This is a non-trivial exercise and you may need to consult the manufacturer's handbook to be sure that you are doing it correctly. You can then *photograph* the beam and determine the intensity distribution from a microdensitometer trace across the film, or an electronic scan across the readout from the CCD detector (see Chapter 7) as shown in Figure 5.11. What you'll learn later on in Chapter 9 is that the first-condenser lens (C1) is responsible for controlling the beam size (and hence the current) and that is why different settings of C1 are mentioned in Figures 5.10–5.12. From Figure 5.11 a couple of important points can be drawn

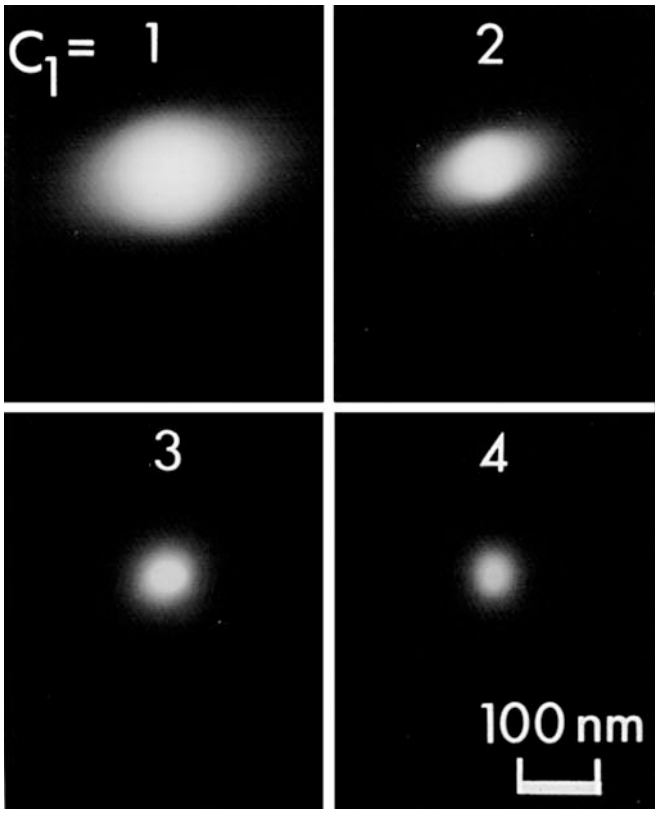


FIGURE 5.11. Four images of the beam formed on the TEM screen at different C1 lens settings. Spot #3 most closely corresponds to the Gaussian intensity distribution shown in Figure 5.9.

- The FWHM contains 50% of the integrated intensity. It is the value used by the manufacturers when they calculate beam sizes. It is also the important dimension when considering the effect of d on the (S)TEM image resolution.
- The full width at tenth maximum (FWTM) contains 90% of the integrated intensity. It is a more relevant dimension generally because the Faraday cup (or any probe-current measuring method) measures the current in the total beam, which is much closer in size to the FWTM. This dimension is also more relevant to measurement of the XEDS spatial resolution. (Think about it and see Chapter 36 later.)

When you insert the beam diameter in the brightness equation, either the FWHM or the FWTM can be used. The FWTM is equal to 1.82 × FWHM and this is also shown in Figure 5.9. You should note, therefore, that you overestimate β if you use the smaller FWHM.

In a dedicated STEM you can't image the beam directly, since there are no post-specimen lenses to magnify its image, no screen on which to project it, and no photographic film to record it. The value of d must be determined indirectly, as in other scanning instruments. The worst method (apart from all the rest) involves scanning the beam across a knife-edge specimen and monitoring the intensity change that occurs, for example, by recording the output from the annular dark-field detector (see Chapter 9). The specimen should be atomically sharp and not transparent to electrons until right at the edge. Such specimens don't exist. This approach yields an integrated intensity profile, as shown in Figure 5.12. In order to extract a value of the FWHM or FWTM from the profile, you must make measurements between various points determined by integrating the intensity from one side of a two-dimensional Gaussian to the other. Nevertheless, the two experimental beam-size measurements in Figure 5.10 show reasonable agreement with the values calculated from the brightness equation. The measurement of d is clearly not a simple procedure.

5.5.E Energy Spread

Remember that the energy spread (ΔE) of the electron beam is a measure of the temporal coherency. This spread is important in EELS and, in fact, the only way to measure the energy spread is to use an electron spectrometer. Under conditions where the spectrometer itself is not limiting the resolution of the spectrum, the value of ΔE can be simply measured by collecting a spectrum of electrons without a specimen in the way of the beam as we describe in Section 37.3.C. The spectrum then consists of a single Gaussian peak and the resolution of the spectrum is defined as the FWHM of this peak. Typical values of ΔE for the various electron

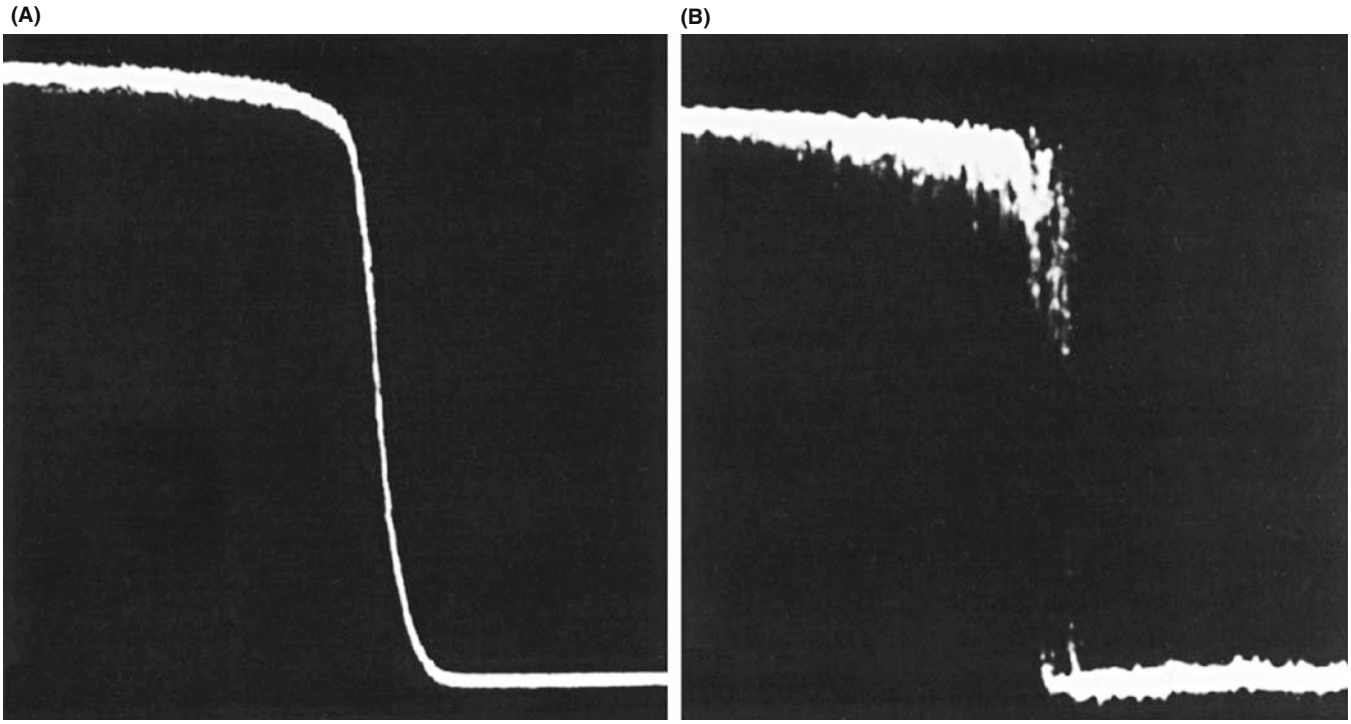


FIGURE 5.12. Intensity profiles obtained by scanning a fine beam across a sharp edge of a cube of MgO. The measured probe size (FWTM) in (A) is 7.4 nm (magnification 1×10^6) and in (B) 1.8 nm (magnification 11×10^6). The smaller probe contains a much smaller current and is thus a much noisier trace.

sources are given in Table 5.1; with a monochromator (see Section 37.7) ΔE for any source can be reduced to $<\sim$100 meV.

5.5.F Spatial Coherency

It's difficult to measure the coherency of the beam experimentally although, as we've discussed, small sources ensure spatial coherency. One practical way of measuring the coherency is to form an image of the edge of a hole in a specimen such as a thin holey carbon film. When you operate slightly out of focus you see alternating dark and bright fringes called Fresnel fringes as shown in Figure 5.13A. Typically, for a thermionic source, only one or two fringes are visible. These fringes are a phase-contrast effect (which we cover in great detail in Part 3). We can also use the fringes to correct astigmatism in the objective lens, as we'll see in Chapter 6. The number of visible fringes is a measure of the beam coherency and Figure 5.13B shows the enormous number that can be generated by a FEG.

5.6 WHAT kV SHOULD YOU USE?

For the materials scientist and nanotechnologist, this is usually an easy question to answer: we'll call it the kV axiom.

THE kV AXIOM
You should always operate at the maximum available kV (unless you shouldn't).

However, there are exceptions to this axiom and the most obvious is avoiding knock-on beam damage, but we'll see others later in the book. So don't forget that you can always operate a modern 300-kV TEM at 100 kV. It's like being able to change the wavelength of a monochromatic light source in a visible-light microscope. As we saw in Chapter 4, the threshold for displacement damage for most metals is less than 400 kV, which is the highest available voltage on 'off-the-shelf' TEMs. For lighter and more beam-sensitive materials such as some ceramics and polymers, lower kV may be better. For most materials specimens, there is not much use going below 100 kV since the images will be rather dim and you'll have to make *very* thin specimens to see anything useful. However, when studying a crystalline specimen by diffraction contrast, 100 kV is better than 200 kV is better than 300 kV, providing you can still see through the specimen! For biological specimens, there is significant advantage to the increased contrast available in lower-kV images and STEM in (30-kV) SEMs is an increasingly useful imaging tool.

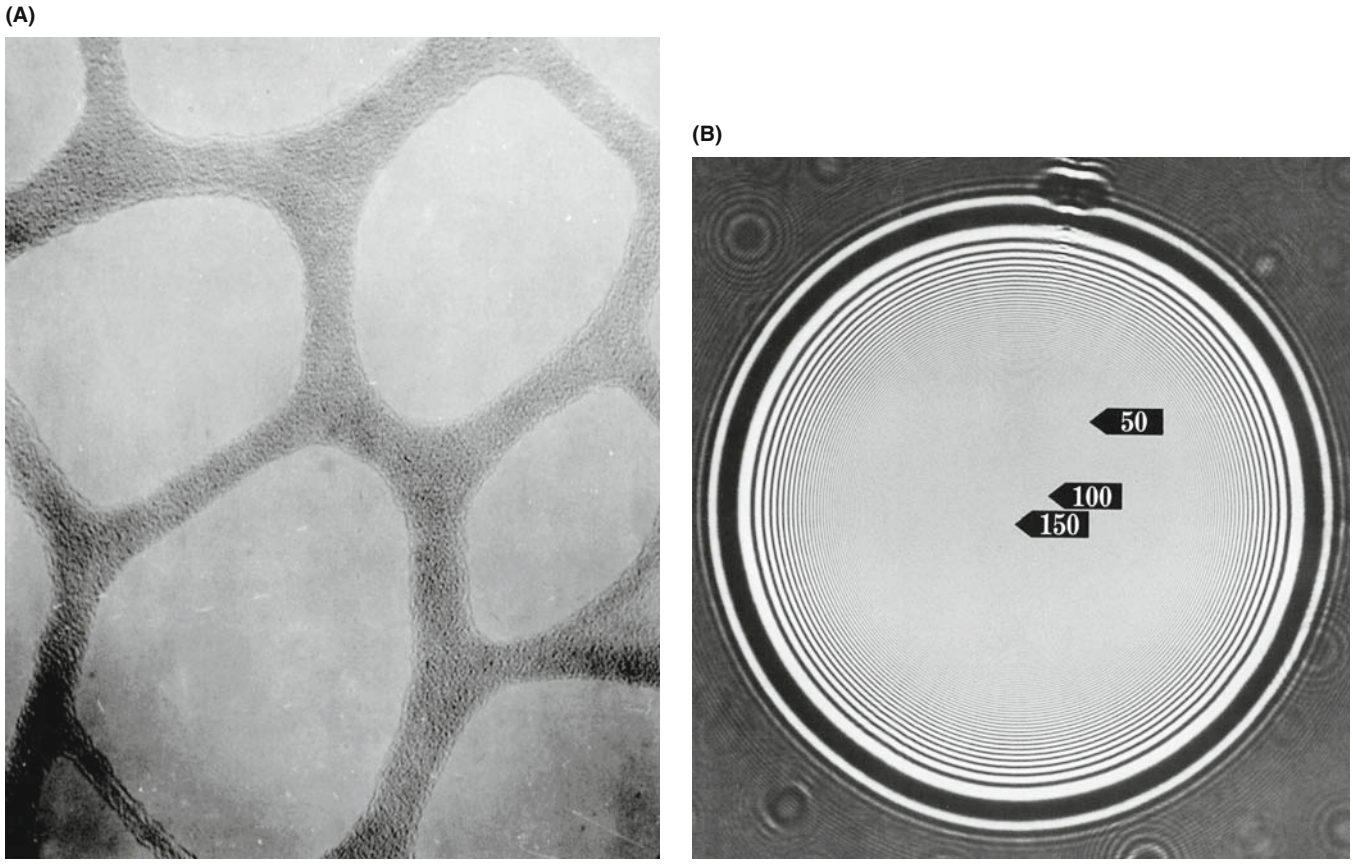


FIGURE 5.13. Fresnel fringes from (A) a thermionic source with poor coherency and (B) a FEG with high coherency.

Apart from these exceptions, the reasons for choosing the highest kV are

- The gun is brightest so you get the most signal to put into your specimen.
- The wavelength is shortest; the image resolution is potentially better.
- The cross section for elastic scatter is also reduced so beam broadening is reduced and analytical spatial resolution is enhanced.
- The cross section for inelastic scatter is smaller, so heating effects may be smaller.
- You can 'see' through thicker specimens.
- The peak to background ratio in X-ray spectra is improved (see Chapter 36).

When you've learned about EFTEM later on, return to this chapter and ask: why not 80 kV?

CHAPTER SUMMARY

Most TEMs use LaB_6 thermionic sources. Take care when heating and cooling your LaB_6 crystal; always operate just below saturation to maximize the lifetime of the source and (almost) always operate at the highest kV. If you're doing high-resolution work of any kind, find a FEG TEM and for high-resolution imaging, then the degree of coherency is important too. If you're doing XEDS, get some idea of the beam current that you can get from your FEG under typical operating conditions. Also measure the beam size and convergence angle to give a measure of β. If you're doing EELS then the energy spread is essential information and you may need to find one of the rare, monochromated TEMs. Treat your source particularly carefully if your TEM is of such a vintage that you have to change it, align it, saturate it, or switch it off. There's nothing more annoying than losing your source since it usually happens at some critical point during your work. Fortunately, computer control is making such events much rarer.

SOME HISTORY

Why include a note on history? Because you should/will study old papers. The sources used in the 1950s and 1960s were so-called W hairpins. The simple hairpins were replaced by pointed W filaments (hairpins with an atom-probe-like tip attached). In the 1970s, LaB_6 filaments replaced W as more efficient emitters. Other materials might be more efficient, but... In the late 1980s, FEGs began to be used in several labs. By the 2000s, these are the sources of choice, especially for those who can afford them. So, we are back to the pointed filament. Now, consider how the TEM techniques had to be different in those old papers.

SOURCES

Broers, AN 1974 in *Recent Advances in Scanning Electron Microscopy with Lanthanum Hexaboride Cathodes* SEM 1974 9–18, Ed. O Johari IITRI Chicago IL.

de Jonge, N and van Druten, NJ 2003 *Field Emission from Individual Multiwalled Carbon Nanotubes Prepared in an Electron Microscope* Ultramicroscopy **95** 85–91. Demonstrated a brightness in excess of $10^{14}A/m^2sr$.

Hawkes, PW 1978 *Coherence in Electron Optics* Adv. Opt. Electr. Microsc. **7** 101–184.

Orloff, J 1989 *Survey of Electron Sources for High-Resolution Microscopy* Ultramicroscopy **58** 88–97.

Veneklasen, LH 1972 *Some General Considerations Concerning the Optics of the Field Emission Illumination System* Optik **36** 410–433.

THE PROBE

Michael, JR and Williams, DB 1987 *A Consistent Definition of Probe Size and Spatial Resolution in the Analytical Electron Microscope* J. Microsc. **147** 289–303. Details on how to measure the diameter of the probe.

Mook, HW and Kruit, P 1999 *On the Monochromatisation of High Brightness Electron Sources for Electron Microscopy* Ultramicroscopy **78** 43–51. Application to EFTEM.

THE COMPANION TEXT

There is much more in-depth treatment of electron guns in the companion text. Electron coherence is a very tricky topic that the most experienced of us can still confuse. It is examined in the discussion of lenses and of holography in the companion text.

URLs

1) http://www.matter.org.uk/tem/electron_gun/electron_sources.htm

SELF-ASSESSMENT QUESTIONS

Q5.1 State the two types of electron sources currently used in TEMs and explain how they work.
Q5.2 Name two thermionic sources and the properties that make them useful.
Q5.3 What is the difference between a field-emission and thermionic source TEMs?
Q5.4 What is brightness and how does it change with kV?
Q5.5 When is high brightness most useful? When is low brightness useful?
Q5.6 Name the five most important properties or characteristics of an electron beam?
Q5.7 What are some reasons to choosing the highest kV when operating a TEM?
Q5.8 What is the purpose of the Wehnelt in a thermionic source? Why don't we need one in a FEG?
Q5.9 Describe the purpose of the two anodes in a FEG.
Q5.10 Why is a FEG operated under high-vacuum conditions?
Q5.11 If faced with the urge to 'crank up the temperature' on the LaB_6, how long between each heat setting should a wise microscopist wait?
Q5.12 What are the limitations of a field-emission TEM?
Q5.13 How would you maximize the coherency of the source?
Q5.14 What is the 'saturation condition' for a thermionic source?
Q5.15 Why would you operate your thermionic filament just under the saturation condition?
Q5.16 How do you know you have achieved gun saturation while using the TEM?
Q5.17 What is spatial coherency and why is it important?
Q5.17 What is temporal coherency and how is it measured?
Q5.19 Name three ways to increase the coherency of the beam.

TEXT-SPECIFIC QUESTIONS

T5.1 Carefully redraw Figure 5.1 to scale.
T5.2 Sketch Figures 5.5A and 5.5C with scale bars.
T5.3 By considering equation 5.3 explain
A. Why a LaB_6 source brightness varies as a function of the orientation of the LaB_6 crystal?

B. Why adjusting the voltage on the Wehnelt can change the apparent brightness of a thermionic source? (Hint: look at Figure 5.4.)
C. Why the brightness of a cold FEG is generally higher than that of a (thermally assisted) Schottky FEG?
D. Why a tungsten source gets brighter if you sharpen the tip of the hairpin?

T5.4 Why is the concept of gun brightness generally not relevant when observing a specimen in TEM mode? Under what operating conditions does the brightness become crucial and why?

T5.5 Estimate the approximate power density in the probe when a FEG puts 1nm of current into a spot of diameter 1nm and a LaB_6 gun puts 10 pA into 1nm. Show all steps and justify any approximations.

T5.6 If the emission current from a thermionic source is several hundred μA (see Figure 5.4B), why does the beam at the specimen contain only a few hundreds or thousands of picoamps?

T5.7 Calculate the beam size for a 100-keV FEG source from the data in Chapter 1, Table 5.1, and equations 5.7, 5.8, 5.9, and 5.10. State any assumptions. Compare your data with Figure 5.10.

T5.8 From Figure 5.10, explain why it would be useful to be able to use larger apertures in the probe-forming system and what prevents us from doing so in most TEMs until recently? (Hint: go back to the questions for Chapter 1.)

T5.9 From the data in Figure 5.10 calculate the beam brightness.

T5.10 Why do you need a different C2 aperture for each possible C1 lens setting? Do you have such a range of apertures on your microscope? (Hint: look at Figure 5.11.)

T5.11 Why is the probe size measured in different ways in a TEM and a dedicated STEM (compare Figures 5.11 and 5.12).

T5.12 Can you think of any use for the much larger number of Fresnel fringes in FEG-TEM images compared with thermionic-source TEM images, such as in Figure 5.13?

T5.13 How would the data in Figure 5.7 change if you had not flashed the tip?

T5.14 Can you think of any other suitable specimens for measuring the experimental probe size by the method shown in Figure 5.12?

T5.15 Can you think of what effect a non-Gaussian probe shape might have on (A) your STEM images and (B) your analyses in AEM?

T5.16 Which line in Figure 5.10 would change for a TEM with a spherical-aberration corrector? Which way would it move and what would be the consequence for the minimum probe dimension?

T5.17 Why is a Schottky field emitter so called?

6

Lenses, Apertures, and Resolution

CHAPTER PREVIEW

Electron lenses are the TEM's equivalent of the glass lenses in a visible light microscope (VLM) and, to a large extent, we can draw comparisons between the two. For example, the behavior of all the lenses in a standard TEM can be approximated to the action of a convex (converging) glass lens on monochromatic light. The lens is basically used to do two things

- Take all the rays emanating from a point in an object and recreate a point in an image
- Focus parallel rays to a point in the focal plane of the lens

The lens can't collect *all* the rays from the object and we often deliberately limit the collection angle with an aperture. We can draw ray diagrams showing how electron lenses control beams of electrons. These diagrams correspond directly to the ray diagrams used in physical optics. Of course the analogy with light fails for certain aspects, but basically it will pervade this chapter. So we'll start by reminding you of the principles of light optics insofar as they relate to electron optics. Then we'll discuss the electron lens in more detail, showing how an electron behaves as it passes through such a lens. We'll describe some actual lenses and tell you how we use different kinds of electron lenses to do different things in the microscope.

A major limit to the use of electron lenses is the fact that we aren't very good at making them. They suffer from rather severe spherical and chromatic aberrations which we usually control by inserting limiting apertures to select electrons nearest to the optic axis since these are least affected by the lens aberrations. Recent technical developments have permitted these aberrations to be largely overcome, but aberration-corrected TEMs are both rare and expensive; most microscopists still have to live with these limitations. So you need to understand lens aberrations, since they play a major role in deciding what we can and cannot do with the microscope. In particular, lens aberrations (rather than the wavelength of the electrons) limit the resolution of the TEM (unlike in the VLM). Since resolution is often the single most important reason for buying a TEM, you need a firm understanding of this concept. Unfortunately, we electron microscopists aren't always very precise in our definitions of resolution. Finally, we describe how the apertures we use aid both the depth of field and the depth of focus of the instrument.

6.1 WHY LEARN ABOUT LENSES?

Why should we learn about electron lenses? As in a VLM, the lenses in a TEM control all the basic operational functions of the instrument. As you are well aware, we have to physically move glass lenses up and down in a VLM to control the intensity of the illumination and to focus the image. The focal length of a glass lens is fixed so we have to change lenses to change the magnification. We choose stronger lenses for higher magnification. By contrast, in a TEM, the positions of the lenses are fixed but we can change the strength of the lens at will.

CHANGING THE LENS

We change focus, change the intensity of illumination or change magnification by changing the strength of the lenses.

As you'll see, in most cases the lenses we use are electromagnetic, so we change their strength by

changing the current through a coil around a soft-iron core which changes the strength of the resultant magnetic field. Almost any operation we carry out on the TEM involves changing magnification or focus; we use electron lenses to magnify and focus the electron beam, the images, and the DPs.

These factors are critical in the principal functions of a TEM: imaging, diffraction, and analysis which, respectively, comprise the next three parts of this book. An aperture is used to select different electron beams to form different images, thus manipulating the image contrast. Another aperture is used to select different regions of the specimen to contribute to the DP as we'll see in Chapter 9.

So knowing how these aperture/lens combinations work allows you to understand how we control the TEM and why we do certain operations on the microscope.

APERTURES

We use apertures in the lenses to control the divergence or convergence of electron paths through the lenses which, in turn, affects the lens aberrations and controls the current in the beam hitting the specimen.

An understanding of electron lenses will help us to answer such questions as

- Why can we see finer detail with an electron microscope than with a light microscope?
- Why can't we see as much detail as we might expect from physics?
- Why does the TEM have a better depth of field and depth of focus than the VLM?

We'll see that the answers to these questions lie in the quality of the lenses, and how we use them. In this chapter we'll discuss the basics of how a lens/aperture combination works. Throughout the book you'll come across different uses and combinations of lenses and apertures. So this is a central chapter for the serious microscope operator but it is only an introduction to the important aspects of electron optics, which is a field in itself. For this you need to explore Chapter 2 in the companion text and the electron-optics texts in the references. Apart from the tremendous advances brought about by aberration correction, electron optics is relatively static these days but traditional light optics is undergoing a renaissance from which electron optics can only benefit. If you are interested you should consult any of the optics textbooks that we referenced back in Chapter 2 and check out URL #1.

6.2 LIGHT OPTICS AND ELECTRON OPTICS

You are already familiar with the action of a magnifying glass lens on light rays. The magnifying glass is a convex lens. It can be used in two ways to control the light rays coming through it. First, it can produce a magnified image of the object you're looking at. Second, it can focus a parallel beam of light to a point, in the focal plane of the lens. (When younger, we've all used this latter property to set something or someone on fire by focusing the (parallel) sun's rays.) These two actions, forming an image of an object and focusing parallel rays to a point, are all we need in order to understand how the lenses in a TEM work. The reason that we can get away with this simple approach is because the electron lenses act, to a reasonable approximation, like convex glass lenses; in detail they're often equivalent to more complex combinations of convex lenses and aberration correction involves the equivalent of a divergent or concave lens. We introduce the practical use of lenses in the TEM in Chapter 9.

6.2.A How to Draw a Ray Diagram

In traditional light optics it's customary to draw diagrams of the paths of light rays through the lens and these ray diagrams are usually drawn horizontally because the traditional optical bench on which light-optical experiments are carried out is a horizontal setup. Likewise, we draw diagrams of the electron trajectories through electron lenses but, since the TEM is a vertical instrument, we will draw all our ray diagrams vertically assuming the gun is at the top of the column of lenses (although this isn't invariably so, as we describe in Chapter 9).

Let's start by drawing ray diagrams to illustrate the two fundamental lens actions of image formation and the focusing of parallel rays. In these and all subsequent diagrams we'll draw all the lenses in the TEM as convex lenses. We will draw all electron ray paths as straight lines outside the lens, and we'll start by assuming that the lenses are perfect. We'll also draw the lenses as so-called 'thin' lenses, which means their thickness is small compared to their radii of curvature. Actually, we'll make the lenses *very* thin. We'll see that these assumptions are all precisely wrong, yet sufficiently reasonable that traditional ray diagrams are nonetheless very useful.

The first thing we need to do is to have a base line on which to draw our diagrams; this line is called the optic axis (also called the rotation axis in the TEM because, as you'll learn, the electrons actually rotate through the lens even though we draw the ray paths as straight lines).

THE OPTIC AXIS

An imaginary line down the column of the TEM passing through the center of each lens.

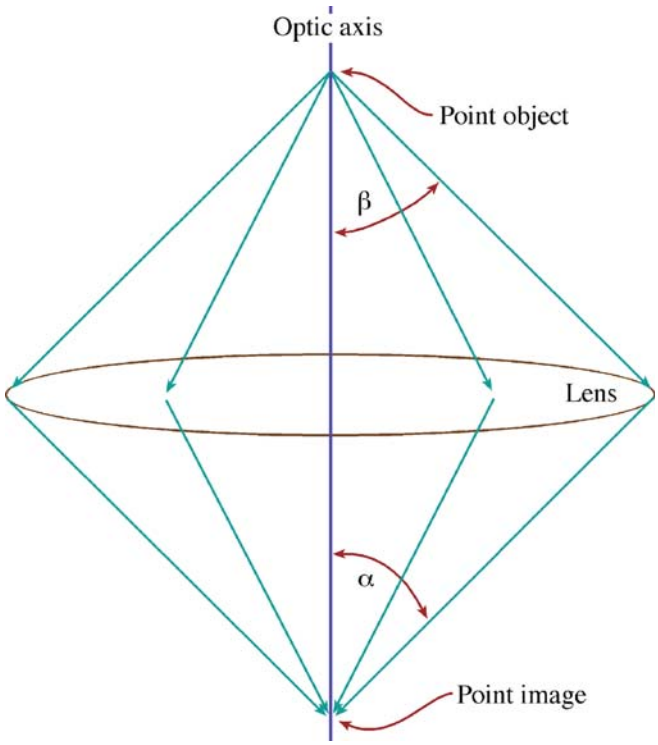


FIGURE 6.1. Image formation by a convex lens. A point object is imaged as a point and the collection angle of the lens is defined relative to the object (β) or the image (α).

Now the first action of a lens that we want to show is how it produces an image of an object. In a TEM the object will usually be the specimen itself or an image of it, but it may also be the electron source, which is an object for the illumination system. If we assume the object is a point and the radiation is emanating from that point (a so-called 'self-luminous object'), then a perfect lens will gather a fraction of that radiation and form a point image. This action is shown in Figure 6.1 in which the point is on the optic axis. The fraction of the rays from the object gathered by the lens is an important variable, defined by the angle β in Figure 6.1. Ultimately, as you can see, β is governed by the size of the lens, but we often choose to limit β by inserting an aperture, as we'll discuss later in this chapter. You'll often see the angle of collection defined as α, but we will reserve α for convergence angles (see Section 2.7). From now on, as we did in earlier chapters, we'll talk about angles when we actually mean semi-angles.

LENSES ARE FINITE

All lenses are imperfect insofar as they cannot gather all the radiation emitted by an object and so can never create a perfect image.

However, as you know from Chapters 2 to 4, most electrons are strongly forward scattered, so we can in practice gather a high fraction of the scattered electrons.

The angles in Figure 6.1 and in the other ray diagrams we'll draw are all greatly exaggerated.

EXAGGERATE ANGLES

In practice, a typical value of β is maybe a few tens of milliradians (10 mrad ~ 0.57°) so if the diagrams were drawn to scale they would be many times longer than they were wide and all the ray paths would be exceedingly narrow. Since drawing to scale is impractical, we always exaggerate the angles considerably in all electron ray diagrams.

If the object has a finite size, we can illustrate this by an arrow, asymmetrically positioned with respect to the optic axis, as in Figure 6.2. Then the lens creates an image of the arrow, rotated by 180°. To draw this figure, the first step is to draw line 1 from the arrowhead through the center of the lens, because rays crossing the optic axis in the lens (or on-axis rays which travel down the axis) are *not* affected by the lens at all and remain as a straight line.

The second step is to draw line 2 which is a ray from the arrowhead that is parallel to the optic axis. We could draw such a ray from any point along the arrow and the further away rays are from the optic axis, the more strongly they are bent by a convex lens. So we take line

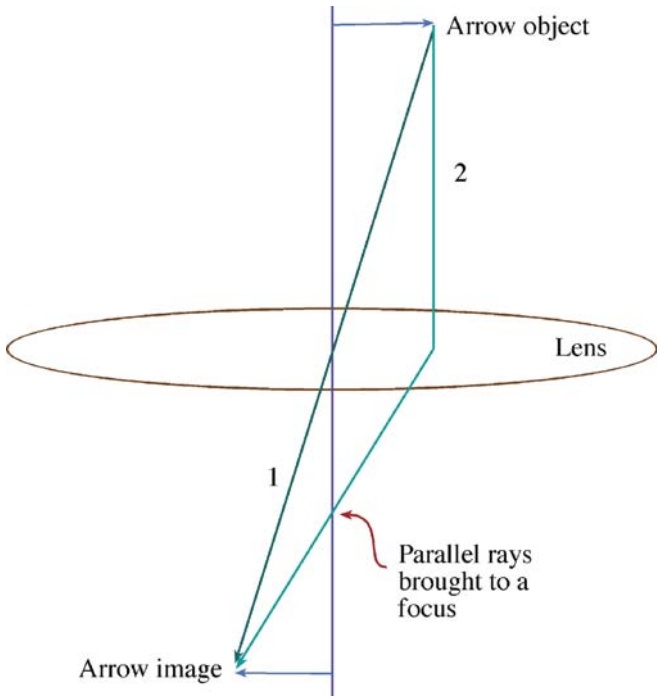


FIGURE 6.2. How to draw a ray diagram: first construct ray path #1 through the middle of the lens, then draw ray path #2 (initially parallel to the optic axis) to determine the lens strength. Where path #2 intersects the optic axis defines the focal plane.

2 and bend it toward the optic axis as it passes through the lens. We can choose to make the lens as strong as we wish, and the strength determines how much the ray is bent and where lines 1 and 2 meet to recreate an image of the arrowhead. Where ray path 2 intersects the optic axis (and thus intersects a ray path along the axis) defines the focal plane of the lens and thus illustrates the second fundamental action of a convex lens, i.e., the lens brings rays that are initially parallel to a focus.

FOR A THIN LENS

A fundamental principle of how a lens works is that an electron passing through the middle of the lens is unaffected so we can draw its path as a straight line. All other electron paths are bent when they pass through the lens.

Some important points on electromagnetic lenses

- The strength of the lens determines where the parallel electrons are focused: stronger lenses having shorter focal lengths.
- The focal plane is where initially parallel rays intersect after passing through the lens.
- The image formed by the lens is rotated by 180° with respect to the object.

Now a full ray diagram for an object of finite size, symmetrically positioned about the axis, combines aspects of Figures 6.1 and 6.2, as shown in Figure 6.3. In Figure 6.3, all rays from a point in the object are brought back to a point in the image and all parallel rays (whether parallel to the optic axis or not) are brought to a focus in a plane at a position depending on their angle to the axis.

Note that on-axis parallel rays are focused on axis and off-axis parallel rays are focused off axis.

This is a most important property, since it allows the lens to create DPs in the focal plane. We'll use this diagram to introduce you to the principal terms used in lens optics.

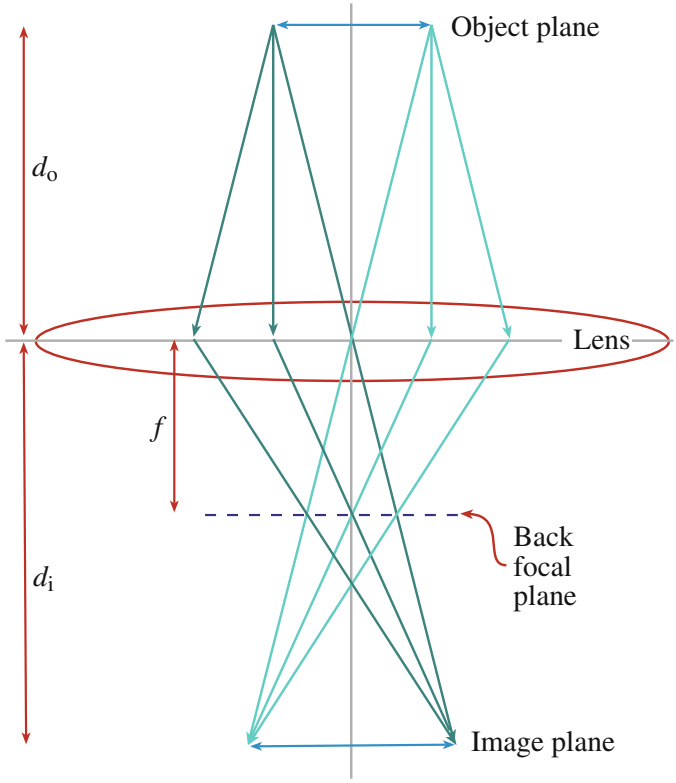


FIGURE 6.3. A complete ray diagram for a finite object, symmetrically positioned around the optic axis. All rays emerging from a point in the object (distance d_o from the lens) that are gathered by the lens converge to a point in the image (distance d_i from the lens) and all parallel rays coming from the object are focused in the focal plane (distance f from the lens).

6.2.B The Principal Optical Elements

From the above diagrams, we can define several *principal planes* to which we will often refer. The first plane is the plane of the lens. In a thin lens this plane can be imagined as a line through the middle of the lens. The object plane is the plane containing the object point in Figure 6.1 or the object arrow in Figures 6.2 and 6.3. The object plane always lies above the lens in question in the diagrams in this text. The image plane (sometimes called the Gaussian-image plane) is the plane containing the image point or arrow and it always lies below the lens. These two planes are said to be conjugate, which means optically equivalent. Rays leaving a point in one plane are brought to a point (if the lens is perfect) in a conjugate plane and vice versa. In other words, the electron doesn't care which way it goes through the lens and this is the basis for the theorem of reciprocity which we'll discuss when we compare TEM and STEM imaging in Chapter 9. The focal plane of the lens is the plane in which the parallel rays are brought to a focus as shown in Figures 6.2 and 6.3. In the image-forming process in a TEM, the focal plane lies after or 'behind' the lens and so the plane is sometimes called the back-focal plane (BFP). There is also an equivalent front-focal plane (FFP) and a convex lens would take all the rays coming from a point in the front-focal plane and create a parallel beam of radiation, in exactly the reverse manner to Figures 6.2 and 6.3.

6.2.C The Lens Equation

From the above diagrams we can define three important distances, labeled in Figure 6.3: the distance from the object plane to the lens (the object distance d_o), the distance from the lens to the image plane (the image distance d_i), and the distance from the lens to the back-focal plane (the focal

length f). Now if the lens is symmetric in strength either side of the lens plane (i.e., the front and back-focal planes are the same distance from the lens) then we can write the following basic equation

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} \tag{6.1}$$

THE PLANES
The principal planes of a lens comprise the object, image and focal planes.

which is known as Newton's lens equation. You'll find a proof in any standard optics text (several were referenced back in Chapter 2). The distances d_o and d_i are measured from the two different principal planes in a thick lens, but from the same plane in the middle of a thin lens, which we are assuming here. In all cases that we'll consider, the object distance (and therefore the image distance) is greater than the focal length. Thus a real image is produced on the other side of the lens beyond the back-focal plane. If the object were within the (front) focal length, then a virtual image would be produced on the same side of the lens as the object, and this is often the case in light optics. Since we don't deal with virtual images in the TEM we'll ignore this aspect.

6.2.D Magnification, Demagnification, and Focus

We can use Newton's lens equation to define the magnification of a convex lens as

$$M = \frac{d_i}{d_o} \tag{6.2}$$

M is also approximately equal to the ratio of the collection angles of the lens subtended at the object (β) and at the image (α) as shown in Figure 6.1, assuming that these angles are small, as they invariably are in a TEM. In this example the magnification is unity.

STRENGTH VERSUS MAGNIFICATION
Under conditions normally found in the TEM, strong lenses *magnify less* and *demagnify more*. In VLMs stronger lenses produce greater magnifications.

Now we may sometimes want to *demagnify* an object (for example, when we want to form a small image of the electron source, to create the smallest possible probe at the specimen). If that is the case, we define the demagnification as $1/M$. In a VLM we could change the magnification by moving the object relative to the lens or vice versa, and adjusting our eyes accordingly, but generally we rotate in another objective lens of different strength (curvature). In a TEM we change magnification in this latter way by changing the strength of the lens, but you'll see that we can do this without changing the lens itself. So electron lenses differ fundamentally from glass lenses in that one lens can be adjusted to a range of strengths.

If we make the lens stronger, then the focal length is shortened as shown in Figure 6.4. If f is shortened but d_o is unchanged, then d_i must be correspondingly shorter and the image magnification is smaller, or the demagnification is larger.

How do we get the high magnifications that we need to form images of atomic columns such as Figure 1.2? Since, as you'll see in Chapter 9, we tend to operate the objective lens of the TEM at a fixed strength, we move the object plane close to the lens thus making d_o small and M correspondingly large (see equation 2). We then make the image plane of the first lens, the object plane for the next lens and repeat this for several lenses in tandem one after the other. So we end up with a multi-lens system like a compound VLM. We'll discuss many more details of lens combinations in the illumination and imaging systems of the TEM in Chapter 9.

Now, in principle, there's nothing to stop us magnifying as much as we wish. However, above a certain magnification, we will see no more information because

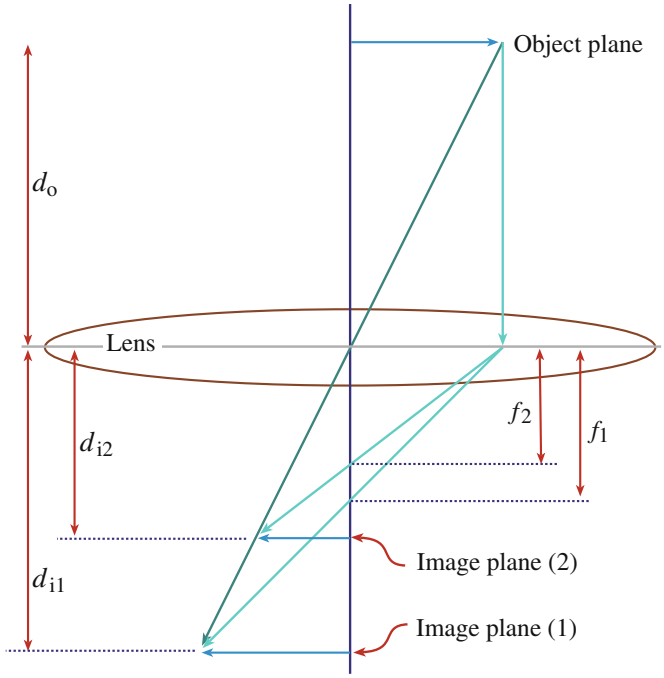


FIGURE 6.4. Strengthening the lens shortens the focal length f. So a weaker lens ($f1$) produces a higher magnification of the object than a stronger lens ($f2$) since the image distance d_i increases, but the object distance, d_o, is unchanged.

other factors limit the image detail and therefore the resolution of the microscope. We'll discuss this point later in Section 6.6. We'll also see that there are times when we want to look at an image of the focal plane (because this contains the DP). To do this, the back-focal plane of the upper lens must become the object plane for the subsequent lenses in the imaging system.

MAGNIFICATION VERSUS RESOLUTION
Don't confuse the two.

When discussing the focus of images we need another convention because we'll find that there is much useful information to be gained and certain technical advantages to operating out of focus. This situation is somewhat different to almost any other form of microscopy wherein out-of-focus images are generally less useful or, more likely, completely useless. However, in TEM we need to define the following two conditions relative to the plane in which a focused image is formed

- If the lens strength is increased such that the image forms above (i.e., before the rays get to) the image plane, then the image will be out of focus and we say the lens is *overfocused*.
- If the lens is weakened and the image forms below (i.e., after) the image plane, the image will be out of focus and the lens is said to be *underfocused*.

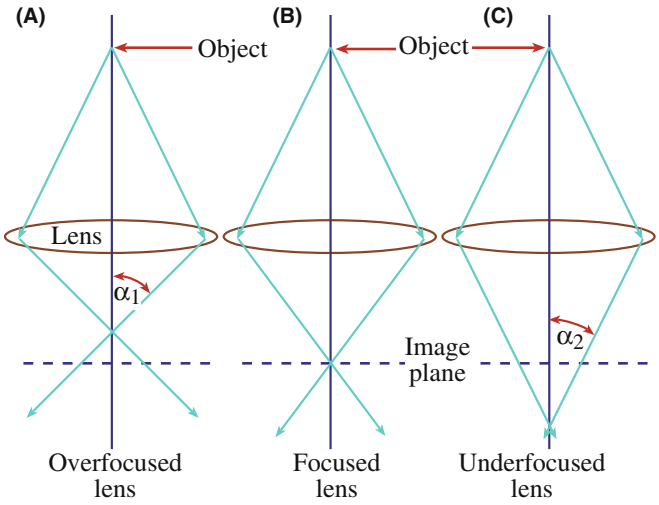


FIGURE 6.5. (A) The concept of overfocus in which a strong lens focuses the rays from a point in the object above the normal image plane where a focused image (B) of the object is usually formed. At underfocus (C) the lens is weakened and focuses the rays below the image plane. It is clear from (C) that at a given underfocus the convergent rays are more parallel than the equivalent divergent rays at overfocus ($\alpha_2 < \alpha_1$).

It's very easy to confuse these two terms unless you think in terms of the vertical frame of the microscope as shown in Figure 6.5. One point to note from Figure 6.5, which we'll find useful, is that the electrons are closer to being parallel to the optic axis when the lens is underfocused than when it is overfocused.

WEAK LENS
A weak, underfocused lens gives a more parallel electron beam. Remember α_1 and α_2 are very small.

We'll exploit underfocused imaging conditions on many occasions in the future. We'll also find there are times when we should operate with our DPs out of focus and also get different information to when it is in focus. So even dexterously challenged TEM operators or those with aging eyes can still do well!

6.3 ELECTRON LENSES

Electrons were first successfully focused by Busch in 1927; he used an electromagnet of the sort that Ruska later incorporated into the first TEM shown in Figure 1.1. Busch also showed that it was possible to focus electrons using electrostatic fields and we've already seen how this works in thermionic electron guns in Chapter 5. In practice, magnetic lenses are superior in many respects, particularly because they are not susceptible to high-voltage breakdown. The TEMs that we're discussing in this text all use magnetic lenses, so we won't discuss electrostatic lenses further here but they are examined in the companion text.

6.3.A Polepieces and Coils

To make a magnetic electron lens we need two parts. Both are drawn schematically in cross section in Figure 6.6. First there is a cylindrically symmetrical core of soft magnetic material such as soft iron, with a hole drilled through it. We call this soft iron a *polepiece* and the hole is called the *bore* of the polepiece. (Soft refers to the magnetic not the mechanical behavior.) In most lenses there are two polepieces (upper and lower), which can be part of the same piece of soft iron as in Figure 6.6 or they may be two separate pieces. The distance between the polepiece faces is called the *gap* and the bore-to-gap ratio is another important characteristic of such lenses, controlling the focusing action of the lens. Some polepieces are machined to a cone shape; the cone angle is then an important variable in the lens performance.

The second part of the lens is a coil of copper wire which surrounds each polepiece. When we pass a current through the coil, a magnetic field is created in the bore.

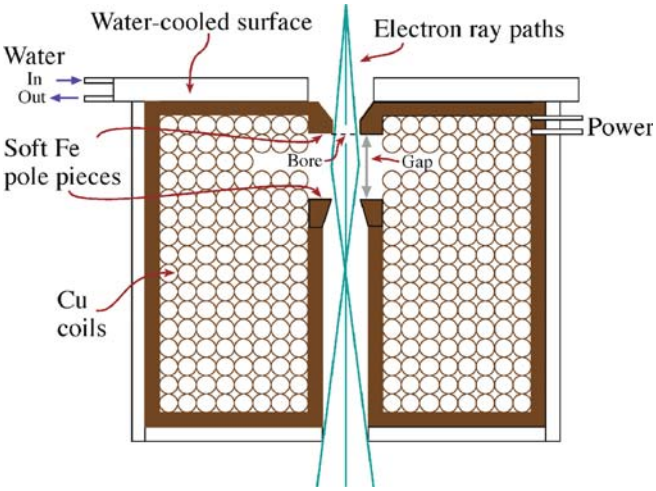


FIGURE 6.6. Schematic diagram of a magnetic lens. The soft-iron polepieces sit in the hole down the middle of the lens and are surrounded by the copper coils through which the current runs to magnetize the polepieces. When viewed in cross section, the bore and the gap between the polepieces are visible. The magnetic field is weakest on axis and increases in strength toward the sides of the polepiece, so the more the electrons travel off axis the more strongly they are deflected.

FIGURE 6.7. A real lens: the cylindrical shape conceals the copper wire coils. The two conical polepieces beside the lens sit inside the central hole in the lens. The three-pin electrical connections provide current to the coil to magnetize the polepieces, and cooling water is circulated in and out of the two holes in the top plate of the lens to dissipate the resistive heat generated in the coils. Compare this picture with the schematic in Figure 6.6.

This field is inhomogeneous along the length of the lens, but axially symmetric. It is the strength of the field in a magnetic lens that controls the electron trajectories or ray paths. As you can see, the electron path through the lens is a reasonable approximation to the schematic diagram back in Figure 6.1.

The resistive heating of the coil means that the lenses have to be cooled and a water recirculating system is an essential part of TEM lenses. A real lens removed from the column of a TEM is shown in Figure 6.7.

6.3.B Different Kinds of Lenses

The principles that we've just described are incorporated into different kinds of lenses used in the TEM. Most lenses in the microscope are weak lenses with large gaps. Either they act to demagnify the source image onto the specimen or they magnify the image or DP from the specimen and project it onto the viewing screen or CCD in ways that we'll see in Chapter 9. Typically these lenses are of the sort shown schematically in Figure 6.6. An aperture can be introduced into the bore of the lens, as we'll discuss later.

PRACTICAL HINT

You should be able to get a readout (on the TEM computer display) of the current through any lens coil. It is a useful thing to know the standard lens currents for your common operating modes such as imaging and diffraction and for creating various beam sizes.

Compared to the other lenses in a TEM, the objective lens is a very strong lens. Several types exist, depending on the needs of the particular TEM. The most flexible objective lens is that in which the upper and lower polepieces are separated and have their own coils as shown in Figure 6.8A. This geometry gives the space needed to allow us to insert both the specimen and the aperture between the polepieces. With this type of polepiece, other instruments such as X-ray spectrometers can have relatively easy access to the specimen. For the same reason, it is straightforward to design specimen holders that do a variety of tasks such as tilting, rotating, heating, cooling, straining, etc. This versatility accounts for the popularity of the split-polepiece lens in TEMs.

With split polepieces it is possible from to make the upper polepiece behave differently from the lower polepiece. The most common application of this is to excite the upper-objective polepiece very strongly. This kind of (asymmetrical) lens is ideal for an AEM/STEM because it can produce both the necessary broad beam of electrons for TEM and a fine beam of electrons for AEM and STEM. We'll see how this is accomplished in more detail in Chapter 9.

If high resolution is a major requirement, then we'll see that it is essential to keep the focal length of the objective lens short and this means a very strong lens is

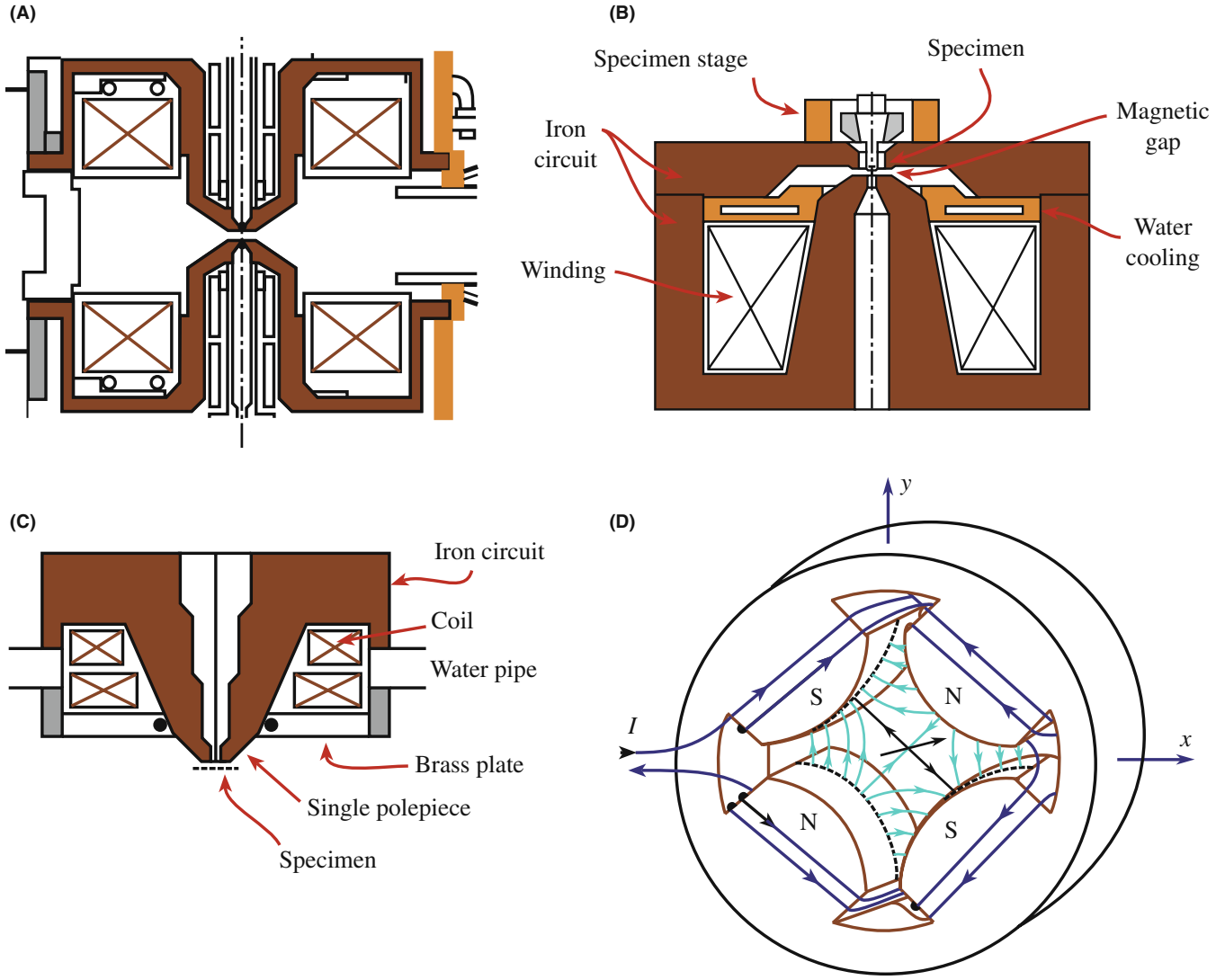


FIGURE 6.8. A selection of different lenses; (A) a split polepiece objective lens, (B) a top-entry immersion lens, (C) a snorkel lens, and (D) a quadrupole lens.

needed. This is traditionally accomplished by using an immersion lens. The specimen is dropped into (i.e., immersed in) the center of the lens field as shown in Figure 6.8B. In such a top-entry stage the specimen is surrounded by the objective lens and so it is a more difficult engineering feat to manipulate, heat or cool the specimen and it is not possible to get X-ray detectors near the specimen, so analytical microscopy is very inefficient. If the focal length is kept really short to give the highest resolution, then it becomes difficult to tilt the specimen more then a few degrees. So in the highest-resolution TEMs you can't do much apart from imaging and diffraction over a restricted range of tilt (see Chapter 8 on stages). This limitation can be overcome by designs such as the snorkel lens as shown in Figure 6.8C, which is a single polepiece lens with a small bore to give a strong lens. Spherical-aberration correction also reduces the need to have the strong lenses for high resolution, so larger gaps are feasible in aberration-corrected TEMs without compromising resolution.

THE OBJECTIVE LENS

The most important lens in the TEM. It forms the images and DPs that are magnified by the other lenses. It is also the most difficult to construct since the specimen must be located close to the plane of this lens.

The limitations of ferromagnetic polepieces can be overcome using superconducting lenses. We cannot make soft-iron polepieces stronger than their saturation magnetization and this limits the focal length and the probe-forming capability of the lens. Superconducting lenses can overcome these limitations but

since a superconductor generates a fixed field, it cannot be varied in the same way as a conventional ferromagnetic lens and so it is not very flexible. Periodically there are increased fluxes of papers describing superconducting lenses because they are small, they don't need water cooling, and they cool the area around the specimen which improves the vacuum, helps minimize contamination, and preserves biological or polymeric specimens. Such lenses also saw a brief flurry of activity after the discovery of high-Tc superconductors. These lenses can generate intense fields (>100 T compared to the maximum of ~2 T in electromagnetic lenses) which are very promising for forming fine probes with high-energy electrons (useful in AEM). Superconducting lenses are so strong that their aberrations (which we'll get to in Section 6.5) are inherently small and they could feasibly be used to construct very compact TEMs.

In addition to these variations on the theme of a single or double polepiece, it is also possible to design a quadrupole, sextupole, or octupole lens in which the focusing action is achieved by four, six, or eight polepieces, respectively. Adjacent polepieces are of opposite polarity as shown in Figure 6.8D. These lenses are not used in TEMs as magnifying lenses but are used to correct lens defects such as astigmatism (see Chapter 9), are used as lenses in aberration correctors (see Section 6.5.A) and also in electron energy-loss spectrometers (Chapter 37). These lenses require less power, and they don't introduce any rotation into the image, which as we'll now show, is a characteristic of standard, electromagnetic lenses.

6.3.C Electron Ray Paths Through Magnetic Fields

We need a bit of mathematics to explain how magnetic lenses actually work. When an electron with charge q (= –e) enters a magnetic field with a strength **B** (Tesla) and an electric field of strength **E**, it experiences a force **F**, known as the Lorentz force, which depends on the velocity of the electron, **v**. All these factors are related through the equation

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = -\mathrm{e}(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \qquad (6.3)$$

where the term in parentheses is a vector cross-product. Since we are not applying an electric field within the lens, the resulting (Lorentz) force **F** is a vector normal to **v** and **B,** which are inclined to one another at an angle θ. You can easily work out the relative directions of **E**, **v**, **B**, and **F** using the right-hand rule in which your thumb represents the direction of the force acting on a *positive* charge moving in the direction of the middle finger through a field in the direction of the index finger. So the force on the electron acts in the *opposite* direction to your thumb.

RIGHT-HAND RULE
Field: Forefinger
Velocity (Speed): Second finger
Thrust: Thumb

The force on an electron entering a uniform magnetic field, nearly 90° to **B** is

$$F = \mathrm{e}vB\sin\theta = \mathrm{e}vB = \frac{mv^2}{r} \qquad (6.4)$$

where r is the radial distance of the electron from the optic axis (sometimes called the cyclotron radius for historical reasons which you should be able to recognize) and m is the mass of the electron. We can rearrange equation 6.4 to give an expression for

$$r = \frac{mv}{\mathrm{e}B} \qquad (6.5)$$

Since v is a relativistic velocity, we should write this equation as

$$r = \frac{\left[2\mathrm{m}_0 E\left(1 + \frac{E}{2E_0}\right)\right]^{1/2}}{\mathrm{e}B} \qquad (6.6\,\mathrm{A})$$

where m_0 and E_0 are the rest mass and energy of the electron, respectively. This form of the equation allows us to substitute known constants to estimate r (in meters)

$$r = \frac{3.37 \times 10^{-6}\left[V\left(1 + 0.9788 \times 10^{-6}V\right)\right]^{1/2}}{B} \qquad (6.6\,\mathrm{B})$$

In deriving equation 6.4, we made a rather gross oversimplification. If θ equals 90°, the electron is traveling straight down the optic axis and is not focused; in fact it doesn't even notice that a lens is there! It is the *deviation* from θ = 90° that gives the lens effect. The next step, therefore, is to separate the electron velocity **v** in a magnetic field into two components, $\mathbf{v}_1$ perpendicular to, and $\mathbf{v}_2$ parallel to the magnetic-field direction **B**, as shown in Figure 6.9, where $v_1 = v\sin\theta$ and $v_2 = v\cos\theta$. The parallel component, $\mathbf{v}_2$, results in motion parallel to the optic axis in the z direction, with $z = v_2 t$, while the perpendicular component produces circular motion with a radius given by equation 6.5.

THE FIELD
For V = 100 kV and B = 1 Tesla, from equation 6.5 the radius, r, is < 1 mm.

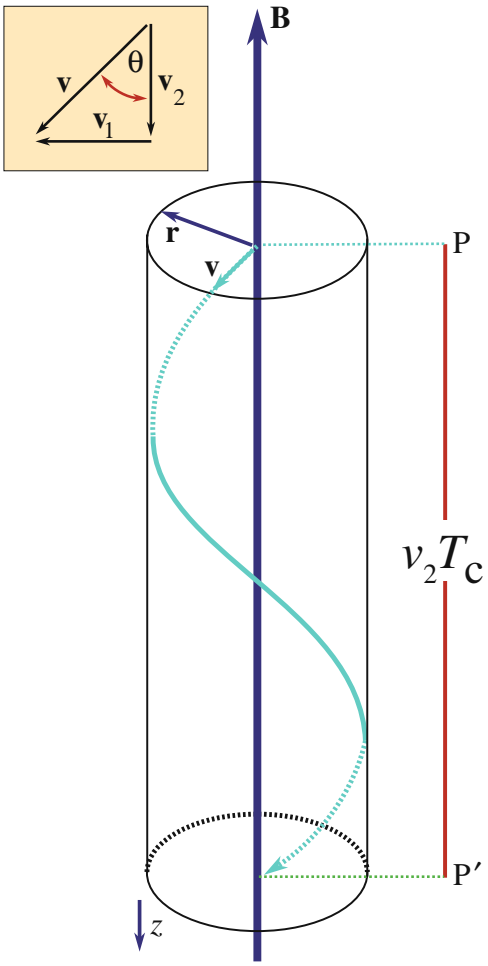


FIGURE 6.9. Electron trajectories in a homogeneous magnetic field, strength **B**. The electrons have velocity components parallel and perpendicular to the field, so long as they are not traveling at 90° to the direction of **B**. The Lorentz force causes electrons passing through point P on the optic axis to spiral through the field and intersect the axis again at P′. The electron's helical path defines the cyclotron radius, *r*.

So all the ray diagrams that we draw ignore this complicating factor which also explains why the optic axis is sometime referred to as the rotation axis. The period of rotation (T_c) through the field gives rise to the (cyclotron) frequency ω_c

$$\omega_c = \frac{2\pi}{T_c} = \frac{eB}{m} \tag{6.7}$$

From these various relationships, we can calculate the complete ray paths through the lens. The most important equations are called the *paraxial* (i.e., near-axis) ray equations. These equations determine both *r* and the angle of rotation (θ) about the axis as the electron moves around the axis in the direction *z*: it rotates under the influence of the rotationally symmetrical field, *B*. These equations, which neglect electron trajectories far off axis, are derived in texts on electron optics. As Hawkes succinctly states "a straightforward, but quite lengthy calculation yields"

$$\frac{d^2r}{dz^2} + \frac{\eta^2 B^2 r}{2\,V^{1/2}} = 0 \tag{6.8}$$

$$\frac{d\theta}{dz} = \frac{\eta B}{2\,V^{1/2}} \tag{6.9}$$

where V is the accelerating voltage of the microscope and η is $(e/2m_0c^2)^{1/2}$. You can see from equation 6.8 that the rate of change of *r* along the optic axis is smaller for more energetic electrons (larger V) and larger for more intense field strengths (larger **B**). Likewise, from equation 6.9, the angular rotation rate increases with increasing field strength and decreases for more energetic electrons.

SPIRAL

The electron spirals through the lens field: a helical trajectory. For electrons with higher keV, we must use stronger lenses (larger **B**) to get similar ray paths.

While these conclusions might be intuitively obvious, the implication is often missed. When we change the TEM accelerating voltage, we change the lenses in the microscope! (Think what this would mean in a VLM.) Therefore, the calibration of the TEM and the lens 'constants' change as we change the kV. Remember the initial paraxial assumption; we'll use non-paraxial rays to explain the effect of spherical aberration on resolution a little later in Section 6.5.A.

While all these ray equations are approximations, they form the basis for more detailed mathematical models of electron motion through lenses (see Chapter 2 in the companion text and URL #2). The more complete models are used in advanced software which simulates the effects of new lens shapes, bore/gap ratios, etc., and has permitted significant advances in the design of lenses to meet the more stringent demands of the latest TEMs.

PITCH OF THE HELIX

When we increase **B**, the pitch of the helical path becomes less if we do not change the energy, because the electrons rotate round the axis more often per unit path length along the axis (*z*).

6.3.D Image Rotation and the Eucentric Plane

So the electrons follow a helical path as they traverse the field along the axis of the lens. This rotation is rarely

shown on standard ray diagrams. You'll see this effect as you operate your TEM because the image or DP rotates on the display screen as you try and focus or if you change magnification. This rotation may require calibration as we'll see in Chapter 9, unless the manufacturer has compensated for it by including an extra lens.

We've already seen in Figure 6.4 that if we change the strength of the lens while keeping d_o fixed, the position of the focal plane and the image plane will also change. Because of this, we have to define a standard object plane for the main imaging lens of the microscope and we call this the *eucentric* plane. Your specimen height should always be adjusted to the eucentric plane because an image of an object in this plane will not move as you tilt the specimen around the primary tilt axis of the holder. (The image will still move if you tilt orthogonally, unless the TEM stage is completely computer controlled to compensate for this.) All other planes in the imaging system are defined with reference to the eucentric plane.

We'll tell you much more about this very important reference plane in Chapter 9.

EUCENTRIC PLANE

If your specimen is in the eucentric plane, then the objective lens strength is always the same when the image on the screen is in focus.

6.3.E Deflecting the Beam

There are many occasions during the operation of the TEM when we want to deflect the beam entering the lens. We may wish to deflect the beam laterally off axis or tilt it to a certain angle with respect to the optic axis. In STEM, these operations are essential to the whole process of forming a scanning image. It is also useful in AEM to be able to blank the beam, i.e., deflect it off axis so it goes into a Faraday cup to measure the current, or to prevent the beam from hitting the specimen when no useful spectroscopic data are being gathered. The way we do this is to apply an electromagnetic field to tilt or traverse the beam or an electrostatic field to blank it. Electromagnetic scan times are of the order of milliseconds while electrostatic blanking can occur in fractions of a microsecond.

Although we are assuming that the lens is thin and has effectively zero thickness along the optic axis, the magnetic field actually acts over a length L. The angle of deflection ε is (for small ε)

$$\varepsilon = \frac{eLB}{mv} \tag{6.10}$$

From this equation we can show that to tilt the beam by 5° we need a coil carrying about 0.2 A and ~100 turns applied along a length of 10 mm, giving a field of 0.01 T. For electrostatic blanking we need about 2 kV/mm.

6.4 APERTURES AND DIAPHRAGMS

We mentioned earlier that an aperture is often inserted into a lens. The aperture limits the collection angle (β) of the lens as shown schematically in Figure 6.10 and such an aperture in the objective lens allows us to control the resolution of the image formed by the lens, the depth of field and the depth of focus, the image contrast, the collection angle of the electron energy-loss spectrometer, the angular resolution of the DP, and so on. In other words, this aperture is important! Physically, the aperture may reside above, in, or below the plane of the lens as we draw it in ray diagrams (but it doesn't really matter since the actual effect will be the same and we've already seen that the electron doesn't care which way it is going). Apertures can also perform other functions, which we'll come across later, such as protecting the specimen from stray radiation in the illumination system, measuring the current in the beam or changing that current.

Usually the apertures are circular holes in metal disks and the disks are made of either Pt or Mo, which are both refractory metals.

A quick word on terminology: While the aperture is the hole in the disk, the metal surrounding the aperture is called the diaphragm (like the variable iris diaphragm in a VLM or your camera). We use the aperture to allow certain electrons to pass through the lens and exclude others by causing them to hit the surrounding diaphragm. This 'aperture/diaphragm' wording, while strictly correct English, is a bit cumbersome, and microscopists tend to be lazy and use 'aperture' in both the correct sense of a hole, but also incorrectly to describe the action of the diaphragm. So we might say that "the objective aperture was used to exclude high-angle scattered electrons from the image" or, as we indicated above, "the aperture protects the specimen from stray radiation" while, strictly speaking, the diaphragm did the excluding and protecting. We'll try to be both consistent and correct in our usage of both terms, but sometimes the precise terminology is awkward.

Diaphragms come in several forms, depending on their function and the particular microscope. They can be either individual disks, each with a particular aperture diameter, or they can be a series of different apertures in a single metal strip (as shown in Figure 6.10). The diameter can be as small as 10 μm, which is about

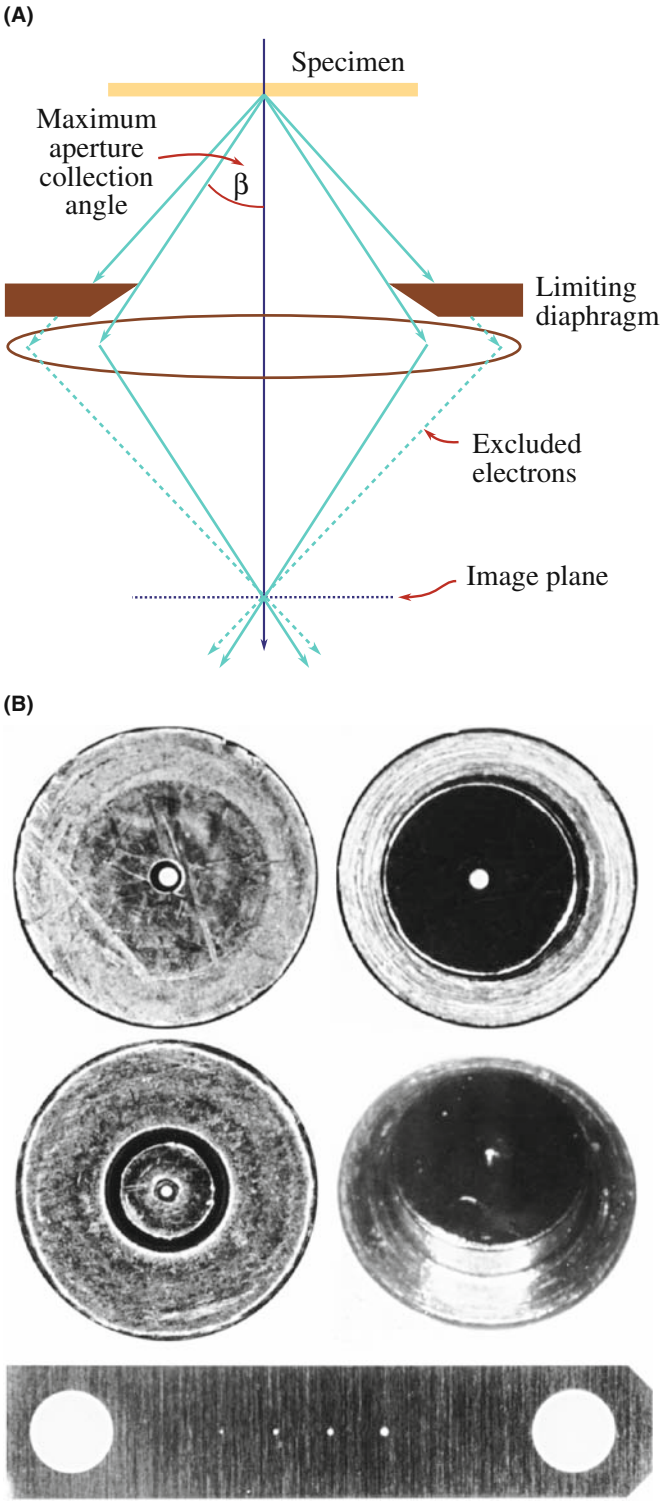


FIGURE 6.10. (A) Ray diagram illustrating how a diaphragm restricts the angular spread of electrons entering the lens. Only electrons emerging from the specimen scattered through angles less than that subtended by the aperture at the object (β) are included in the image-forming process (full lines). The excluded electrons are scattered at angles >β and are stopped by the diaphragm (dotted lines). (B) A selection of diaphragms: the top and middle left are upper and lower views, respectively, of a conventional objective diaphragm; the top/middle right are views of a 'top-hat' (thick) C2 diaphragm; below is a metal strip containing several apertures. Each diaphragm is ~3 mm across.

the smallest circular aperture we can make consistently, or up to ~ 0.3 mm (300 μm). The individual diaphragms or the strips are usually a heavy metal such as Mo or Pt and ~ 25–50 μm thick, but if their job is also to prevent X-rays from hitting the specimen they may be several millimeters thick (see Chapter 33), which means they can be quite expensive if they're made of Pt.

Often the diaphragm collects contamination caused by the electron beam cracking residual hydrocarbons in the vacuum (as we describe in Chapter 8). The contamination tends to accumulate on the edges of the aperture, destroying their circular shape and causing astigmatism. So the diaphragms need occasional cleaning, which can be done by heating them to red heat in the central blue part of a butane flame. In some TEMs, this problem is eliminated by making the diaphragm from very thin metal foil (e.g., Au or Mo). The foil gets hot in the electron beam and any contamination boils off. But such 'self-cleaning' diaphragms are delicate and often crack, thus allowing electrons through other gaps, which defeats the object of the exercise of producing a well-defined aperture.

A SAFETY NOTE ON X-RAYS

X-rays with energies up to the beam energy are generated within the lens wherever the electron beam hits a surface (particularly a limiting diaphragm). So substantial, carefully designed lead shielding is incorporated into the column of the TEM to prevent irradiation of the operator. Obviously, it could be very dangerous to tamper with the lenses or diaphragms of the microscope in any way and only qualified engineers should dismantle, take apart, or repair the lenses or remove the diaphragms.

6.5 REAL LENSES AND THEIR PROBLEMS

It might appear from what we've discussed so far that the analogy between electromagnetic lenses and convex glass lenses is complete, but that is not so. Over the 300 years since van Leeuwenhoek first constructed a light microscope, glass lenses have developed to a point where perfect lenses can be fabricated. In the 80 years since Busch's first magnetic lens, we haven't progressed so far and our lenses are still very imperfect. We've already compared the best electromagnetic lens to the equivalent of using the bottom of a well-known soft-drink bottle as a magnifying glass. Another common description is that if the lenses in your own eyes were as good as our best electromagnetic lens, then you'd be legally blind! So we have to modify all the ideal ray diagrams we've drawn to take into account the imperfections of the lenses. These imperfections all limit the resolution of the microscope but, paradoxically, help us

to get better depth of focus and depth of field from the microscope.

There are many kinds of lens defects (see Chapter 2 in the companion text) and, at one time or another, the effects of all the various defects can be seen in an image or DP. In practice, however, most of us don't need to know about all of them and we'll emphasize the ones that limit the microscope performance in substantial ways. These comprise spherical aberration, chromatic aberration, and astigmatism.

6.5.A Spherical Aberration

The term 'spherical aberration' has almost entered the popular vocabulary since its presence was discerned in the main optics of the Hubble Space Telescope (unfortunately after launch). This defect occurs when the lens field behaves differently for off-axis rays. For our electromagnetic lenses, the further off axis the electron is, the more strongly it is bent back toward the axis. As a result, a point object is imaged as a disk of finite size, which limits our ability to magnify detail because the detail is degraded by the imaging process. As we've told you many times, we can now correct for this aberration, but it still limits the resolution of most TEMs so we need to examine it carefully.

The effects of spherical aberration are shown in Figure 6.11. A point object P is imaged at P′ in the Gaussian image plane. The image is not a point but is instead a central high-intensity region surrounded by a halo of decreasing intensity (similar to Figure 2.11). Spherical aberration is most important in the objective lens because it degrades the detail that we can resolve in TEM images: all the other lenses magnify any error it introduces. It is equally deleterious in the condenser lenses in an AEM or STEM which we use to form the smallest probe with the most current. What we can accomplish at the resolution limits of all forms of TEM is almost always limited by spherical aberration; which is why we're so excited that we can now correct for it.

From Figure 6.11 you can see why we use the term 'spherical' to describe the aberration. The effect of this aberration is to take the curved (spherical) wavefront from the source and increase the curvature. Now if you go back and look at Figure 6.9, you'll see that electrons traveling through a point P on axis intersect the axis again at point P′ where the distance PP′ is given by

$$\mathrm{PP}' = v_2 T_c = v T_c \cos\theta = 2\pi \frac{mv}{eB}$$
$$\left(1 - \frac{\theta^2}{2} + \ldots\right) = L_0\left(1 - \frac{\theta^2}{2} + \ldots\right) \qquad (6.11)$$

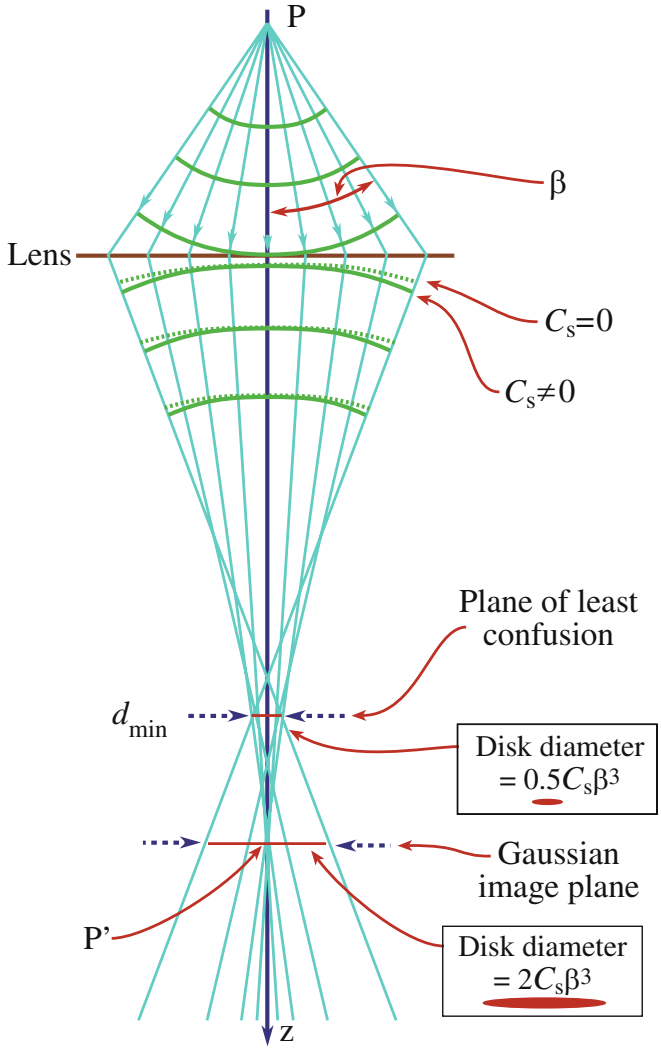


FIGURE 6.11. Spherical aberration in the lens causes wavefronts from a point object P to be spherically distorted by bending the rays at the outside of the lens more than those close to the axis. The point is thus imaged as a disk with a minimum radius in the plane of least confusion and a larger disk at P′ in the Gaussian-image plane. The plane of least confusion is where the smallest image of the object is formed. Schematic intensity distributions at these two important planes are shown beside the ray diagram.

In this relationship, $L_0 = \mathrm{PP}'_0$, where P'_0 is the Gaussian image of the point P for very small θ (i.e., paraxial (near-axis) conditions). As θ increases, the distance PP′ decreases because of spherical aberration and we can write

$$PP' = PP'_0 = -\Delta z \qquad (6.12)$$

where $\Delta z = 0.5 L_0 \theta^2$. Thus we get an expression describing the error, δ, in the Gaussian image position due to spherical aberration

$$\delta = \Delta z \tan\theta \sim \Delta z \theta = 0.5 L_0 \theta^3 \qquad (6.13)$$

So the diameter of the Gaussian image of a point *formed by paraxial rays* is given by this expression, which we will write as

$$\delta = C_s\theta^3 \qquad (6.14)$$

where C_s is a constant (a length) for a particular lens called the spherical aberration coefficient. As you will have noticed, we often use 'C_s' to mean spherical aberration as in 'C_s correction,' 'C_s corrector' and 'C_s corrected.' We'll see in a while that this equation is *very* important because of its effect on the resolution of the TEM and so we need to make a few points of clarification here.

- Equation 6.14 was for paraxial rays only. In a real TEM, the apertures are usually large enough that paraxial conditions do not apply and the sharp image is made more diffuse. As a result the spherically aberrated Gaussian image under non-paraxial conditions is broadened to a diameter of $\delta = 2C_s\theta^3$ (see Figure 6.11).
- You'll often see equation 6.14 written as $\delta = C_s\theta^3 M$ when referring to the image plane but because most discussion of TEM resolution refers back to the minimum distance that we can resolve in the object plane (i.e., the specimen) the magnification term is sometimes left out and we will use this approach.
- When referring to resolution in TEM images it is the *radius* of the point that is more important than the diameter.
- In a real lens the value of θ in equation 6.14 which describes the angle of the electron to the optic axis is replaced by the maximum angle of collection of the (objective-lens) aperture, β.

So, in the forthcoming discussion of resolution, we'll use the radius, we'll refer to the object plane, and we'll use β to define the objective-lens angle of collection. This latter approach is compared to our use of α when discussing beam size in Chapter 5, since α defined the angle of beam convergence. Be careful: many other TEM texts use α somewhat more indiscriminately for both collection and convergence angles.

So we end up with an expression for the radius of the spherically aberrated disk of intensity r_{sph} in the Gaussian image plane, referred back to the specimen plane, under non-paraxial (i.e., realistic) conditions, given by

$$r_{sph} = C_s\beta^3 \qquad (6.15)$$

Because β (in radians) is small, then β^3 is a very strong dependence. The units of r and C_s have to be the same and since C_s is typically a few mm, then we can measure r in (very small fractions of) mm. From this derivation (and linking equations 6.13 and 6.14) you can see that C_s has the dimensions of length and typically it is approximately equal to the focal length of the lens, which for objective lenses in most TEMs is 1–3 mm, but in high-resolution instruments may be well below 1 mm (or not, if they have a C_s corrector).

If you look at Figure 6.11, you will see that the smallest dimension of the cone of rays formed by the lens does not occur at the Gaussian image plane. As we note in the figure, the smallest dimension is formed at a plane closer to the lens which goes by the delightful term 'plane of least confusion' or sometimes 'plane of minimum confusion'; this disk has a radius of $0.25C_s\beta^3$ and a diameter of $0.5C_s\beta^3$. As we'll discuss in Section 6.6.C, some texts use this smaller dimension to define the resolution limit imposed by spherical aberration and it is popular with the TEM manufacturers since it is smaller than the disk in the Gaussian image plane and thus the resolution of the microscope appears better!

CONFUSION

Beware when reading about TEM image resolution because of the confusion between the definition that refers to the Gaussian image plane and that referring to the plane of least confusion.

The way that a corrector compensates for C_s in a magnetic lens is in effect to create a diverging (i.e., concave) lens which spreads out the off-axis beams such that they re-converge to a point rather than a disk in the Gaussian-image plane. In practice this correction is achieved by a highly complex, computer-controlled set of quadrupoles and hexapoles or octupoles. There are two main approaches to the solution of C_s correction. The first is due to the work of Rose and colleagues in Germany embodied in the CEOS commercial system and is used for both probe correction in STEMs and image correction in TEMs. The second is due to Krivanek et al. and is used in Nion dedicated STEMs and has also been retrofitted to several VG STEMs. Figure 6.12 shows schematic ray diagrams for the Nion corrector and the CEOS system. We'll leave a more in-depth discussion of C_s correction to the companion text.

6.5.B Chromatic Aberration

This term is related to the 'color' (i.e., frequency, wavelength, or energy) of the electrons. We've assumed so far that the electrons are monochromatic, but they aren't really. However, we can make very good high-voltage

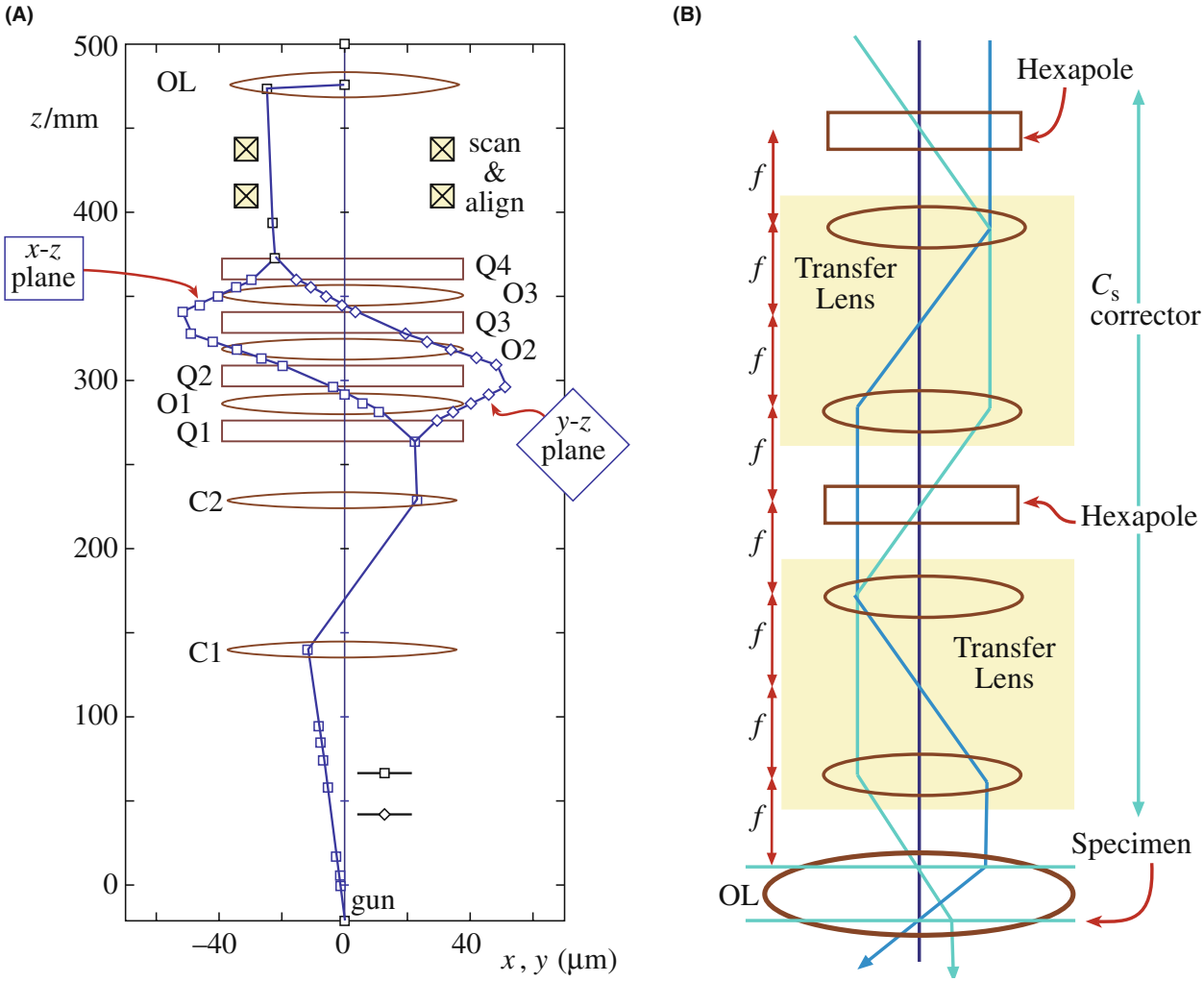


FIGURE 6.12. Ray diagrams showing how the two different commercial systems use (A) multiple quadrupole (Q) and octupole (O) lenses (Nion) or (B) hexapole and other transfer lenses (CEOS) to correct for C_s.

supplies and the variation of the electron energy due to the power supplies is usually smaller than one part in 10^6, which is 0.1 eV for a 100-keV beam. As we discussed in Chapter 5, depending on the electron source the actual energy spread in the beam may vary from ~0.3 eV (cold FEG) to ~1 eV (LaB_6). This range is still so small that we generally don't have to worry about chromatic aberration affecting the image resolution. The exception is if you happen to have a C_s corrector, in which case, after compensating for C_s, C_c is the next most-persistent aberration. Lens elements that can correct for C_c are being developed.

USING A MONOCHROMATOR

Correcting for C_c effects only makes sense if you are dealing with specimens that are thin enough such that specimen-induced chromatic effects do not dominate the resolution. (Correcting C_s is similar.)

Chromatic aberration could be almost completely ignored if we didn't put a specimen into the beam. Unfortunately, this rather essential action creates electrons of a whole range of energies emerging from the thin foil (for reasons we described in Chapter 4). The objective lens bends electrons of lower energy more strongly and thus electrons from a point in the object once again are blurred to form a disk in the Gaussian-image plane (Figure 6.13) (and a smaller disk in the plane of least confusion). The radius r_{chr} of this disk (referring to the object plane) is given by

$$r_{chr} = C_c \frac{\Delta E}{E_0} \beta \tag{6.16}$$

where C_c is the chromatic-aberration coefficient of the lens, ΔE is the energy loss of the electrons, E_0 is the initial beam energy, and β is the angle of collection of the lens. C_c, like C_s, is a length, approximately equal to the focal length. While ΔE in the incident electron beam

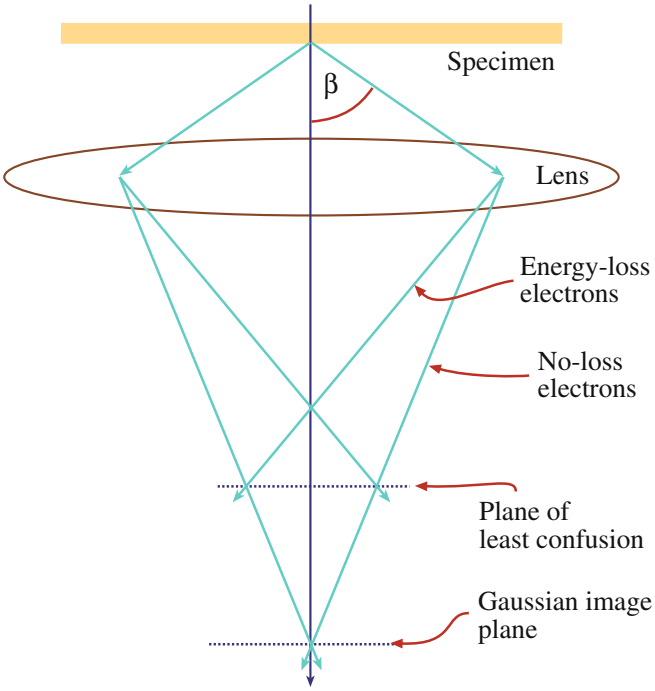


FIGURE 6.13. Chromatic aberration results in electrons with a range of energies being focused in different planes. Electrons emerging from the specimen with no loss of energy are less strongly focused than those that suffered energy loss within the specimen. So, as in Figure 6.11, a point in the object is imaged as a disk in the Gaussian image plane and there is a plane of least confusion.

is $<\sim$1 eV as we just noted, it is typically 15–25 eV for a good fraction of the electrons coming through a typical 50–100 nm thick foil so, as you can easily calculate, r_{chr} is quite a large number (compared to atomic dimensions). Chromatic aberration gets worse for thicker specimens and is better for thinner ones (remember the almost ubiquitous 'thinner-is-better' criterion). So you can do something cheaply to minimize the effects of this aberration; make thin specimens!

The mechanics of C_c correction depend on whether we're trying to compensate for beam effects or specimen-induced effects. We just reminded you that the energy range of electrons coming from the gun is governed by the type of electron source so there are limits depending on which source you have in your TEM. Monochromating the source is a (very expensive) solution. We discuss monochromators at various times in Chapters 37–40 on EELS because that is where such correction pays the greatest dividends.

Unfortunately, for the vast majority of TEM studies, our specimens are not thin enough and, when we have to live with C_c limitations due to having a thick specimen, energy-filtering (EF) is the best solution. EFTEM can correct for the poor resolution that arises when we form images or DPs with electrons that have lost substantial amounts of energy in the specimen, as we'll also discuss in detail in the EELS chapters.

6.5.C Astigmatism

Astigmatism occurs when the electrons sense a non-uniform magnetic field as they spiral round the optic axis. This defect arises because we can't machine the soft-iron polepieces to be perfectly cylindrically symmetrical down the bore. The soft iron may also have microstructural inhomogeneities which cause local variations in the magnetic field strength. Even if these difficulties were overcome, the apertures we introduce into the lens may disturb the field if they are not precisely centered around the axis. Furthermore, if the apertures are not clean, the contamination charges up and deflects the beam. So there are a variety of contributions to astigmatism, which distorts the image by an amount r_{ast} where

$$r_{ast} = \beta \Delta f \tag{6.17}$$

and Δf is the maximum difference in focus induced by the astigmatism. Fortunately, astigmatism is easily corrected using stigmators, which are small octupoles that introduce a compensating field to balance the inhomogeneities causing the astigmatism. There are stigmators in both the illumination (condenser lenses) system and the imaging system (objective lens) and we'll describe how to use them in Chapter 9.

In summary, spherical and chromatic aberration and astigmatism are the three major defects in electromagnetic lenses. There are several minor defects, such as barrel and pincushion distortion, which are self-explanatory in terms of how they distort the image. They are occasionally seen at very low magnification where electrons traveling well off axis and close to the bore of the polepiece appear in the image. Other defects such as coma, and field curvature we will ignore for now.

Again, if you want to learn more about any of these defects they are covered in Chapter 2 in the companion text.

6.6 THE RESOLUTION OF THE ELECTRON LENS (AND ULTIMATELY OF THE TEM)

Another note on terminology: We electron microscopists tend to be rather imprecise in our definition and use of the words 'resolution' and 'resolving power' and related expressions. We've borrowed these terms from classical VLM, which is concerned with the imaging of incoherent light waves through amplitude contrast. High-resolution performance in the TEM is a different matter and traditionally involves phase-contrast imaging of reasonably coherent electron waves, so perhaps we shouldn't be surprised if a different usage has developed. But we should at least define the terms we use. Now in VLM the word *resolution* strictly applies to the act of displaying fine detail in an *image*. The

resolving power of the microscope is the ability to distinguish in the *image* two points, which are closely adjacent in the object. The minimum distance apart of these points in the *object* is the *minimum-resolvable distance*. Since electron microscopists customarily talk about the resolution of the TEM in terms of distances in the *object* (usually a fraction of a nanometer), we should then use the term minimum-resolvable distance but instead everyone says resolution.

Because the lens defects that we've just discussed cause a point object to degrade into a Gaussian image with a finite radius (some combination of r_{sph}, r_{chr}, r_{ast}) they limit the resolution of the electron lens, and hence that of the microscope. The image resolution in the TEM is governed by the ability of the objective lens to image the object, while in the STEM the image resolution is governed by how much beam current we can put into a small probe which is an image of the electron source demagnified onto the specimen. In either case, aberrations limit the resolution.

RESOLUTION AGAIN

We will use the word resolution, but we define it to mean (usually) the minimum-resolvable distance in the object!

6.6.A Theoretical Resolution (Diffraction-Limited Resolution)

If there are *no* aberrations at all, the resolution of *any* lens (glass, electromagnetic, electrostatic. . .) is customarily defined in terms of the Rayleigh criterion, which we introduced back in Equation 1.1 for light optics. Rayleigh's criterion for resolution is arbitrary in the sense that it is not a fundamental physical rule but more a practical definition. This criterion gives us a figure of merit in terms of our eyes' ability to distinguish separate images of two self-luminous, incoherent point sources.

POINTS BECOME DISKS

A single point source will not be imaged as a point, even if no aberrations or astigmatism are present. The finite size of the lens results in diffraction of the rays at the outermost collection angle of the lens, usually defined by the limiting aperture.

This diffraction results in a point being imaged as a disk (called the Airy disk) which has a schematic cross section intensity profile as shown in Figure 6.14A (and also in Figure 2.11). This effect should be familiar to anyone who has encountered basic physical optics. If the two disks overlap so much that they cannot be resolved as in Figure 6.14B, then the points in the object cannot be resolved. Rayleigh assumed that if the maximum from one source lies over the first minimum from the other source, as shown in Figure 6.14C, then your eye can discern this dip as two overlapping images, thus indicating the presence of two separate objects. Under Rayleigh conditions, when the overall intensity profile exhibits a dip in the middle that is above 80% of I_{max}, the two points cannot be resolved. The separation of the two incoherent point sources is then defined as the theoretical resolution of the lens r_{th} and is given by the radius of the Airy disk, which is similar in form to equation 1.1

$$r_{th} = 1.22\frac{\lambda}{\beta} \tag{6.18}$$

BEWARE!

Sometimes in EM texts you'll find the diameter rather than the radius is used. Reason: the beam diameter defines image resolution in STEM and SEM; in TEM, the radius controls the image resolution. (Hence the factor 1.22 in equation 5.10.)

Any standard text on physical optics (which we've already referenced) will show you how to derive this criterion.

Strictly speaking we should not use this equation for electron sources because they are not incoherent. When dealing with high-resolution images, a different approach is used (see Chapter 28). But for our introductory purposes here, we will be content with this approximation.

From equation 6.18 we see that we can get higher resolution if we lower λ or increase β. This terminology can initially be confusing because, as we just did,

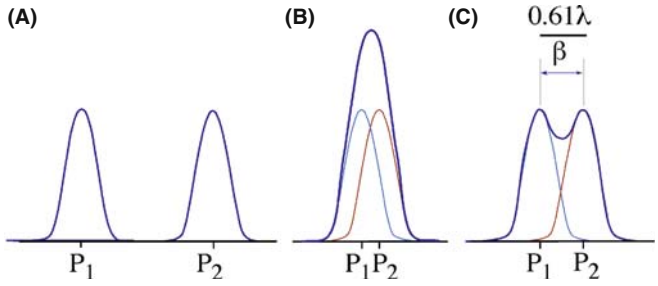


FIGURE 6.14. (A) The Airy-disk intensity profiles from two clearly separated point sources P_1 and P_2. In (B) the two Airy disks are so close that they cannot be distinguished, but in (C) the two are separated such that the maximum in the image of P_1 overlaps the minimum in P_2. This latter situation is the definition of resolution defined by the Rayleigh criterion and is the best (diffraction-limited) resolution.

microscopists often use the expression 'higher resolution' when in fact they mean 'better resolution.' The word higher is then associated with a lower number! It's a smaller r in equation 6.18. It's not just microscopists; a vacuum is also 'higher' if its magnitude is smaller. The improvement in resolution with lower λ is a major reason why there are intermediate and high-voltage TEMs since λ decreases with keV, as we saw back in equation 1.6. The obvious question is, why don't we just increase β (i.e., use a bigger lens aperture or remove it altogether). Well, we could do this if we had perfect lenses, but that isn't the case. All the lens aberrations increase as we increase β (see equations 6.15–6.17); which is why C_s correction is so interesting.

6.6.B The Practical Resolution Due to Spherical Aberration

Let's assume first of all that we have corrected for any astigmatism and our specimen is thin enough that chromatic aberration is negligible. Under these circumstances, the spherical aberration error r_{sph} limits the resolution. Now if you go back and look at equation 6.15 you'll see that r_{sph} increases with the cube of β, a *very* strong dependence. The resolution in the object, then, is given by some combination of the Rayleigh criterion and the aberration error. Hawkes gives a particularly clear description of how this combination leads to a value for the resolution of the microscope. Since this is very often *the* principal figure of merit used when investing hundreds of thousands or even millions of dollars in a TEM, it is essential that you understand that the definition is not exact.

We'll start by summing the radii of the Rayleigh disk and spherical-aberration disk (in the Gaussian image plane) in quadrature (remember it's radii for image resolution, diameters for probe-limited resolutions)

$$r = (r_{th}^2 + r_{sph}^2)^{1/2} \tag{6.19}$$

Therefore, since both these terms are approximate

$$r(\beta) \approx \left[\left(\frac{\lambda}{\beta}\right)^2 + \left(C_s\beta^3\right)^2\right]^{1/2} \tag{6.20}$$

Since the two terms vary differently with the aperture collection angle β, a compromise value exists when the differential of $r(\beta)$ with respect to β is set to zero and we find that

$$\frac{\lambda^2}{\beta^3} \approx C_s^2\beta^5 \tag{6.21}$$

So we come up with an optimum expression for β which Hawkes (1972) gives as

$$\beta_{opt} = 0.77\frac{\lambda^{1/4}}{C_s^{1/4}} \tag{6.22}$$

The exact value of the numerical factor depends on the assumptions made about the various terms included in the definition of resolution and so is often written simply as ***A***. Sometimes, this compromise value is determined by simply equating the equations for r_{th} and r_{sph} rather than going through the summation in quadrature. A quick calculation for 100-keV electrons (λ = 0.0037 nm) for an instrument with C_s = 3 mm gives a β_{opt} value of ~4.5 mrads.

If this expression for β_{opt} in equation 6.22 is substituted into equation 6.20 we get a minimum value of $r(\beta)$

$$r_{min} \approx 0.91\left(C_s\lambda^3\right)^{1/4} \tag{6.23}$$

This is the expression we want; it gives the *practical* resolution of the TEM.

The numerical factor in equation 6.23 is often written as ***B***. Typically, the value for r_{min} is ~0.25–0.3 nm and the best high-resolution instruments have r_{min} ~0.1–15 nm; 1-Å TEMs are about the best available without C_s correction and about 0.07 nm is (at the time of writing) the best reported resolution with C_s correction. So, as we showed back in Figure 1.2, we can resolve rows of atoms, which in most crystalline materials have a separation close to r_{min} (although low-index planes in some metals are still below this resolution). It's worth noting that since your eyes can resolve a distance of ~0.2 mm, then the maximum useful magnification of the best high-resolution TEM is $\sim 3\times10^6$. Above this magnification, no more detail will be revealed.

Hawkes (1972) reminds us that the decision to add in quadrature back in equation 6.19 was arbitrary, and simply summing r_{th} and r_{sph} is another possible way to determine r_{min} (as we'll see in Section 28.7). But any way you combine the two terms for r (or diameter if you're discussing a probe-limited resolution) leads to expressions that have the same general form as equation 6.22 for β_{opt} and equation 6.23 for r_{min}. In some cases, ***A*** and ***B*** are put equal to unity and not even included, and if you're not pushing any limits in your calculations or experiments this latter approach is a very reasonable approximation.

As we indicated right at the beginning of our discussion, electron microscopists are rather imprecise in our definition of the resolution. However, the resolution is often given as a very precise number!

6.6.C Specimen-Limited Resolution Due to Chromatic Aberration

Remember that we assumed in the previous section that there was no contribution from chromatic aberration. However, if you have a thick specimen then a significant number of electrons will lose 15–25 eV of energy (a typical value of the most probable (plasmon) energy loss; see Figure 4.1). If you put 20 eV into the chromatic-aberration resolution limit given by equation 6.16 you'll find that, at 100 keV with β_{opt} of 4.5 mrads from equation 6.16, the value of r_{chr} is ~2.5 nm.

C_C-LIMITED RESOLUTION

This is typically 10× larger than the C_s-limited resolution. When you're looking through a thick specimen the performance of your TEM is 10× worse than its specified resolution.

If you have a thick specimen, it doesn't matter what voltage you use or how low your C_s is; it doesn't matter if you've got a 1-MeV TEM or access to a C_s corrector; you'll have an image resolution in the 1–3 nm range and you can see all the available information in your specimen at a magnification as little as $\sim 10^5\times$. In fact, the vast majority of all recorded TEM images have C_c-limited resolution: your images will too!

So how thick is thick? Well, it depends on TEM voltage and the mean free path for elastic and inelastic scatter in your specimen, which increases with Z (see Chapter 4). For good high resolution at 100 keV your specimen should be $<\sim$ 30 nm, while at 300 keV you can probably get away with ~50 nm before C_c effects begin to control resolution, assuming $Z <\sim 30$. So for higher-Z specimens the 'thinner is better' axiom is even more important. A more restrictive rule of thumb given by Sawyer and Grubb is that, for biological and polymeric specimens, the resolution limit is about one tenth the specimen thickness. As we noted when we first talked about chromatic aberration, the solution to this problem is in your hands (and in Chapter 10).

6.6.D Confusion in the Definitions of Resolution

If you're new to the subject, you don't have to read this section because it may confuse you still further, but if you've read other TEM texts you may have noticed discrepancies in the definitions of resolution.

We used the expression for r_{sph}, measured at the Gaussian-image plane. Strictly speaking, it is only under ideal conditions (i.e., if $C_s = 0$) that we should use the Gaussian image as a measure of the resolution limited by the lens and it is only really correct to use the Gaussian image under *paraxial* conditions, that is with a *very* small objective aperture. As we've already noted, in the TEM β is usually large enough that paraxial conditions do not apply. So the disk in the plane of least confusion is the relevant feature from which to define the best image resolution, as shown back in Figure 6.11.

If this is so, why did we choose the definition of r_{sph} as the radius of the disk in the Gaussian image plane?

The answer to this question is discussed by Hawkes. Defocusing the image slightly, to bring the plane of least confusion to the Gaussian image plane, will indeed lead to a decrease in the value of the numerical factor in equation 6.23 from 0.91 to 0.43. Hawkes also comments that since this latter value is smaller, manufacturers tend to use it to define the resolution of their instrument! However, this whole treatment of resolution assumes incoherent illumination, which is *not* the case in the TEM. Also, the resolution depends on the contrast in the image and how the lens transfers information from the object to the image. As a result, Hawkes concludes that (see equation 6.23) $\boldsymbol{B} \sim 1$ (from the Gaussian image) is "a more prudent choice" (i.e., closer to reality) than $\boldsymbol{B} = 0.43$ (from the disk in the plane of least confusion) even though, strictly speaking, the plane of least confusion refers to the conditions operating in the TEM.

So it is basically a matter of opinion whether to use the diameter or the radius of the disk in the Gaussian image plane or that in the plane of least confusion. Fortunately, it doesn't really matter too much since, in the end, the choice only alters the value of the numerical terms $\boldsymbol{A}$ and $\boldsymbol{B}$, which we've already mentioned are often approximated to unity anyhow. For example, the value of $\boldsymbol{A}$ will depend on exactly which of the several quoted expressions was used for r_{sph}, e.g., if there was 0.25, 0.5, or 1 in front of $C_s\beta^3$. After these various terms are fed into the equations and the value of β_{opt} is extracted, $\boldsymbol{A}$ only varies by about ±15%. A small variation in $\boldsymbol{B}$ will occur also, for the same reason.

BEWARE!

1. There are inconsistencies in the definition of the terms used to describe the effects of C_s on TEM resolution
2. We use the Gaussian image radius referred back at the object plane, i.e., we use $r_{sph} = C_s\beta^3$.

We have tried to be consistent in our use of the radius of the Airy disk and the radius of the aberration/astigmatism error. Obviously, it doesn't really

matter whether you use the radius or the diameter, so long as you are consistent. Occasionally, however, you may find the Airy disk *radius* is used in combination with the *diameter* of the disk in the plane of least confusion or the Gaussian image plane, so this also contributes much to the discrepancy between various TEM texts.

The question to any student learning HRTEM is: do you know what your resolution really is?

6.7 DEPTH OF FOCUS AND DEPTH OF FIELD

You should have got the message that, because of the poor lens quality we have to use small apertures to minimize their aberrations. This generally means that we cut out many of the electrons that would otherwise be gathered by the lens. Fortunately, our electron sources are so bright that we can live with substantially reduced beam currents hitting our specimen. In fact there are advantages to using small apertures, despite the price we pay in image intensity, probe current, and diffraction-limited resolution. These advantages come in the form of better depth of focus and better depth of field. These terms can be confusing, and once again, the TEM literature is variable. So we need to go back to physical optics to find the correct definition of these terms.

Basically, we are trying to find out how much of the object (the specimen) is in focus at the same time and over what range the image is in focus. (This latter question is irrelevant in SEM and dedicated STEMs without post-specimen lenses where we don't use conventional lenses to form the image, so both terms are equivalent.) In TEM both terms are important.

The depth of field, D_{ob}, is measured at, and refers to, the *object*. It's the distance along the axis on both sides of the object plane within which we can move the object without detectable loss of focus in the image. The depth of focus, D_{im}, is measured in, and refers to, the *image plane*. It is the distance along the axis on both sides of the image plane within which the image appears focused (assuming the object plane and objective lens are fixed). Note that we say "appears" in both cases and this of course also depends on how good your eyes are.

We can derive expressions for these definitions using Figure 6.15. Imagine that ray 1 originates at the highest point up the column where the object can appear to be in focus within the resolution and that this ray arrives at the farthest point down the column where the image can appear to be in focus. Ray 2 represents the other extreme but travels at the same inclination to the optic axis. If these two rays appear to come from the same point (to within the resolution of the lens) the distances d_{ob} and d_{im} correspond to the smallest distances which we can resolve in the object or image, respectively. Note immediately that d_{im} is greater than d_{ob}. Now we can show that angles α_{im} and β_{ob}, which are both small, are given by

$$\alpha_{im} \approx \tan \alpha_{im} = \frac{d_{im}/2}{D_{im}/2} \tag{6.24}$$

and

$$\beta_{ob} \approx \tan \beta_{ob} = \frac{d_{ob}/2}{D_{ob}/2} \tag{6.25}$$

The angular magnification is thus

$$M_A = \frac{\alpha_{im}}{\beta_{ob}} \tag{6.26}$$

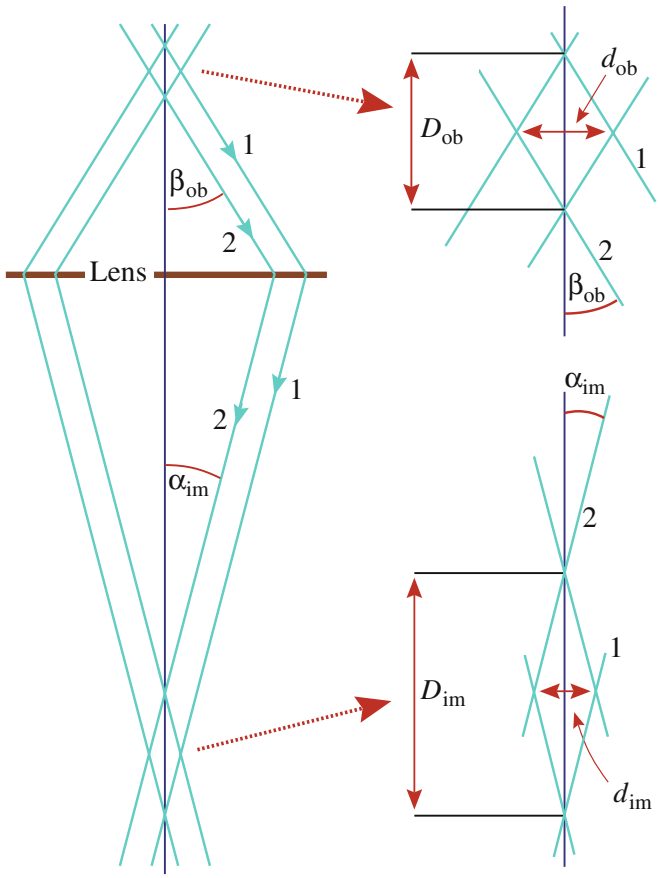


FIGURE 6.15. The definition of the depth of field and the depth of focus. Rays 1 and 2 represent the extremes of the ray paths that remain in focus when emerging $\pm D_{ob}/2$ either side of a plane in the specimen. Typically D_{ob} is greater than the specimen thickness. The same rays define the depth of field over which the image is in focus $\pm D_{im}/2$ either side of the image plane. The resolution in the object is d_{ob} and that in the image is d_{im}.

and the transverse magnification (which we simply call the magnification) is

$$M_T = \frac{d_{im}}{d_{ob}} \tag{6.27}$$

If these two magnifications are related in the usual way by

$$M_T = \frac{1}{M_A} \tag{6.28}$$

Then we can say that the depth of focus is given by

$$D_{im} = \frac{d_{ob}}{\beta_{ob}} M_T^2 \tag{6.29]}$$

and the depth of field is

$$D_{ob} = \frac{d_{ob}}{\beta_{ob}} \tag{6.30}$$

Notice that for a correct calculation of either D_{ob} or D_{im} you must be careful to select the right value of β. Under different circumstances, the limiting angle is defined by the illumination aperture α (in the C2 lens) or the objective aperture β_o (in the objective lens). In thin specimens, which scatter weakly, most electrons emerge from the specimen in a cone closer to that defined by α_{im}, which is often very small ($\sim 10^{-4}$ rad). In a thicker, more strongly scattering specimen, the objective aperture defines the angle and β_o is usually about 10^{-2} rad.

FOCUS AND FIELD

So we get a much greater depth of focus and field by using small apertures (small β).

For a collection angle, β_{ob}, of 10 mrad and a d_{ob} of 0.2 nm, equation 6.30 tells us that the depth of field will be 20 nm, i.e., a specimen of this thickness can all be in focus at the same time. If you only need 2-nm detail in your image, then you can use a specimen which is 200 nm thick and it will still all be in focus.

If we want to see detail at the 0.2 nm level we need to use a magnification of about 500,000×. Equation 6.29 tells us that, under these conditions, the depth of focus will then be 5 km! If we only need to see 2 nm, we can use a magnification of 50,000× and the depth of focus is 5 m. In either case, we have tremendous latitude in where we put the photographic negative or CCD camera because it would still record a focused image many meters either side of the screen. This explains why we can use a CCD camera which can be inserted just below the final projector lens, and still get a focused image with a TV camera well below the standard film camera. In fact, the TEM image would be in focus on the floor under the microscope (or maybe even the floor below if your TEM lab has the misfortune not to be on the lowest floor) if you projected it there but M_T would be different!

Now things get a little more complicated if you're using a C_s-corrected TEM because, for example, in STEM we can use much larger condenser apertures which give much larger convergence angles and thus define the probe a lot more strictly, over distances less than the specimen thickness. In this case, rather than worrying about the reduced depth of field it becomes feasible to think about focusing the probe at different levels within the specimen to explore any structural or chemical variations through the foil thickness. So now we can think about overcoming the projection limitation of TEM images. (Remember the rhino?) The first attempts at imaging single atoms at specific depths within a thin specimen have already been reported.

CHAPTER SUMMARY

We've introduced you to the principles of how an electromagnetic lens works, and how we describe its functions in simple ray diagrams. There are two principal operations: either we use the lens to create an image of an object or we use it bring parallel rays to a focus. We'll see in later chapters that the former operation is used to create magnified images of the specimen on the screen of the TEM. It is also used to create small electron probes (demagnified images of the electron source) at the plane of the specimen in a STEM or AEM. The latter operation is used to create DPs in the back-focal plane of the objective lens.

Our lenses are rather abysmal in their performance, resulting in the need for small limiting apertures. The lens aberrations limit the resolution of the microscope and we usually need an optimum aperture to get the minimum resolution. The small apertures cut down the electron beam intensity, but also give us remarkable depth of focus and depth of

field in our images and specimen, respectively. The recent development of aberration correctors for TEM will revolutionize much of what we've written in this chapter. However, very few TEMs are equipped with correctors and so, for the vast majority of users, it is important to understand the limits imposed on the resolution performance of TEMs by their lenses and by your specimens.

You don't need much skill to use a magnifying glass. The manufacturers may say the same for today's TEM. We say, the more you know about the TEM, the more you'll get out of it and the less likelihood you'll have of making embarrassing errors.

SOME HISTORY

Busch, H 1927 *Über die Wirkungsweise der Konzentrierungsspule beider braunschen Röhre* Arch. Elektrotech. **18** 583–594. The original paper on focusing electron beams.

Hawkes, PW (Ed.) 1997 *Advances in Imaging & Electron Physics Vol. 100: Partial Cumulative Index* Academic Press New York (now published by Elsevier). Essential reference for the historically minded.

Hawkes, PW 2004 *Recent Advances in Electron Optics and Electron Microscopy* Ann. Fond. Louis de Broglie **29** 837–855. An overarching yet concise review of recent advances in electron optics and microscopy, with a great collection of references, both historical and recent.

LENSES AND ELECTRON TRAJECTORIES

Reimer gives a summary of lens defects and more on the derivation of equation 6.11.

Grivet, P 1972 *Electron Optics* Pergamon Press New York.

Hawkes, PW 1972 *Electron Optics and Electron Microscopy* Taylor & Francis Ltd. London. This account is particularly clear if you have an interest in the physics of electron lenses. An important discussion of how to take account of many aberrations when giving a figure of merit. In 'Confusion in the Definitions of Resolution,' we follow Hawkes' clear reasoning regarding the plane of least confusion.

Hawkes, PW (Ed.) 1982 *Magnetic Electron Lenses* Springer New York. A collection of review articles in true Peter Hawkes style; thorough, sound, erudite, and informative.

Hawkes, PW and Kasper, E 1989, 1994 *Principles of Electron Optics* **1**–**3** Academic Press New York. Comprehensive but advanced. Volume **3** includes imaging in the TEM. If by now you're getting the idea that Hawkes is *the* source of electron optical information, then you are right.

Klemperer, O and Barnett, ME 1971 *Electron Optics* Cambridge University Press New York.

Munro, E 1997 *Electron and Ion Optical Design Software for Integrated Circuit Manufacturing Equipment* J. Vac. Sci. Technol. B **15** 2692–2701. More on electrons moving through the lens.

Rempfer, GF 1993 *Electrostatic Electron Optics in the 1940s and Today* MSA Bull. **23** 153–158. By an expert in the use of electrostatic lenses.

ABERRATION CORRECTION

The companion text goes into this in much more detail. In particular, you'll find there that C_s is actually better written as C_3. There are many more 'C_s' terms. These references give an introduction.

Chang, LY, Kirkland, AI and Titchmarsh, JM 2006 *On the Importance of Fifth-Order Spherical Aberration for a Fully Corrected Electron Microscope* Ultramicroscopy **106** 301–306.

Krivanek, OL, Delby, N and Lupini, AR 1999 *Towards Sub-Å Electron Beams* Ultramicroscopy **78** 1–11. Used in the Nioen STEM.

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. Electr. Microsc. **48** 821–826.

RESOLUTION

All texts on TEM will include a discussion of resolution. Particularly useful are those in Reimer 1997, Edington 1976, Fultz and Howe 2002 and Hirsch et al. 1977.

❑ Points to be wary of when reading about definitions of C_s-limited resolution: (see references in Chapter 1)

Sawyer and Grubb (2008) and Egerton 2005 use the Gaussian image radius referred back at the object plane, just as we do; i.e., $r_{sph} = C_s\beta^3$. Reimer 1997 and Fultz and Howe 2001 use the diameter of the disk in the plane of least confusion; i.e., $d_{sph} = 0.5C_s\beta^3$ although both also describe the radius at the Gaussian image plane as we do. Beware: Edington 1976 implies, and Hirsch et al. 1977 state, that $C_s\beta^3$ is the radius *of the disk in the plane of least confusion, which it is not, since by definition it must be less than the Gaussian-image radius (see Figure* 6.11).

Sawyer, LC, Grubb, DT and Meyers, DT 2008 *Polymer Microscopy* 3rd Ed. Springer New York. Rule of thumb for polymers.

❑ Points to be wary of when reading about depth of field and depth of focus

Bradbury et al. 1989 give a particularly clear discussion of the topic. Reimer 1997 uses the term depth of focus for the depth of field and uses depth of image for depth of focus; a rare inconsistency! The terms are used interchangeably in SEM because there is no lens between the object and the image.

Bradbury, S, Evennett, PJ, Haselmann, H and Piller, H 1989 *Dictionary of Light Microscopy* Royal Microscopical Society Handbook #15 Oxford University Press New York. For the VLM comparison.

SPECIAL TECHNIQUES

Borisevich, AY, Lupini, AR, Travaglini, S and Pennycook, SJ 2006 *Depth Sectioning of Aligned Crystals with the Aberration-Corrected Scanning Transmission Electron Microscope* J. Electr. Microsc. **55** 7–12. Moving to confocal imaging in the TEM.

THE COMPANION TEXT

The concepts of depth of field and depth of focus are explored in greater depth in Chapter 2 of the companion text. There, you'll also find more on the principles of lens optics and on Newton's lens equation in particular. Lenses may also be used in other components of the TEM such as electron spectrometers or, in effect, the electron gun; you'll find more on these topics in the same chapter. Quadrupoles, sextupoles, and octupoles are critical components in the correction of aberrations and in some spectrometers; we decided to leave even an introduction to the details until the specialized chapters in the companion text.

URLs

1) http://www.opticsinfobase.org/default.cfm. Optics information base courtesy of the Optical Society of America; lists of papers and journals.
2) http://www.mebs.co.uk/about_us.htm. Munro's company site; provides commercial software for electron optics; essential for the serious designer.

SELF-ASSESSMENT QUESTIONS

Q6.1 How do you focus an image in a TEM?
Q6.2 What kind of visible-light lens does the behavior of a magnetic lens resemble?
Q6.3 Name the main components of a magnetic lens and state their functions.
Q6.4 What are the back and front-focal planes of a magnetic lens?
Q6.5 What do we mean by the term 'optic axis'?
Q6.6 What force acts on an electron in a magnetic field and how can we control this force?
Q6.7 What effect does the magnetic lens have on the trajectory of the electron with respect to the optic axis?
Q6.8 To achieve the highest magnification, where should the specimen be located relative to the objective lens?
Q6.9 Define 'underfocused' and 'overfocused.'
Q6.10 Why is the objective lens the most important lens in a TEM?
Q6.11 Define the eucentric plane.
Q6.12 Explain the difference between a diaphragm and an aperture.
Q6.13 Why do we use apertures in the TEM?
Q6.14 What causes spherical aberration and how can we minimize it?
Q6.15 Define chromatic aberration and describe how to minimize it.
Q6.16 What causes astigmatism and how do we correct it?
Q6.17 Define resolution (strictly speaking, the resolving power) of the TEM?
Q6.18 What ultimately limits the TEM resolution?
Q6.19 In practice, what often limits the practical TEM resolution?
Q6.20 In the TEM, what is depth of field, what controls it, and why is it important?
Q6.21 In the TEM, what is depth of focus, what controls it, and why is it important?

TEXT-SPECIFIC QUESTIONS

T6.1 Estimate the limit of resolution in a 100-kV TEM if the specimen is very thin. Assume $C_s = 1$ mm and $\beta = 10$ mrads.
T6.2 Under the same conditions, estimate the limit of resolution if the specimen is thick enough so that each electron on average undergoes a plasmon loss of ~ 15 eV.
T6.3 If your specimen is pure Al how thick does it have to be such that each electron typically suffers a single plasmon loss. (Hint: go back to Chapter 4.)
T6.4 Go on the Web to find the image resolution offered by commercial manufacturers for a typical 200-kV TEM. (Find Web pages for FEI, Hitachi, JEOL, Zeiss.) Compare the resolution with your answers to questions 1

and 2. What does this exercise tell you about the assumptions that are being made when a TEM resolution is specified?

T6.5 Use suitable values of β_{ob} and α_{im} to deduce values for D_{ob} and D_{im} under the following two conditions: (a) 100-keV electrons, 20 k× magnification, looking for 1 nm detail in the image; (b) 200-keV electrons, 800 k× magnification, looking for 0.2 nm detail in the image.

T6.6 Examine Figure 6.9. Why is B exactly parallel to the optic axis? Is the electron traveling exactly parallel to B? Show, with diagrams, that you understand the reason for these questions.

T6.7 In deriving equation 6.14 we say this is used only for paraxial rays. Why?

T6.8 Explain why the electrons in Figure 6.13 are 'overfocused' if they have lost energy.

T6.9 Discuss the accuracy of using a Faraday cup to measuring beam current.

T6.10 Why do we always draw electron lenses as convex and why haven't we been able to build a concave lens for electrons using cylindrically symmetric lenses? (Hint: recently this problem has, in effect, been solved and has resulted in a solution to the long-standing problem of reducing spherical aberration.)

T6.11 Just for the heck of it try to draw Figure 6.1 to scale assuming that the focal length and object distance are about 3 and 1.5 mm, respectively, and the aperture in the bore of the lens is 60 μm diameter. Estimate values of α and β and explain which of the various lens properties wouldn't be such a good choice for a real lens in TEM.

T6.12 Calculate the radius of the spiral trajectory of 100- and 300-kV electrons in a magnetic field of 1 T.

T6.13 Use ray diagrams to distinguish the terms underfocus and overfocus. Usually it is good to operate any kind of microscope with the lenses in focus. Can you think of any occasion when underfocus (or overfocus) conditions might be useful? (Hint: refer to Figures 6.4 and 6.5.)

T6.14 Why do we use soft magnets and not permanent magnets for electron lenses? If we were in fact to use a permanent magnet what advantages might this bring to the design of a TEM?

T6.15 Why do we have to cool the electron lenses? List as many drawbacks as you can to having to cool the lenses. Can you think of any lenses that might not require cooling? Could we design a TEM in such a way that lens cooling is not required?

T6.16 How could you compensate for the rotation introduced into an electron beam by the action of the lens field? Why would you want to do this?

T6.17 Distinguish the plane of least (or minimum) confusion, the Gaussian image plane, the back-focal plane and the front-focal plane of a lens, using diagrams where necessary. (Hint: Figure 6.11 is a good place to start.)

T6.18 Explain (using diagrams where necessary) why paraxial rays from a point in an object are not subject to significant spherical aberration, yet are still brought to a focus in the Gaussian image plane as a disk rather than as a point.

T6.19 Calculate the radius of the image disk in the plane of least confusion and in the Gaussian image plane under spherical aberration conditions. Assume reasonable values for all terms and justify your assumptions.

T6.20 Calculate the optimum semi-angle of collection for the objective aperture to minimize the contributions of spherical aberration in a 200-keV microscope. State any assumptions. What is the practical resolution of the TEM under these conditions?

T6.21 How is the practical resolution of a TEM further compromised if thick specimens are used? Calculate the expected resolution (i.e., the radius of the disk in the plane of least confusion in the Gaussian image plane) assuming all the electrons suffer an energy loss of $\sim$ 15 eV. Assume reasonable values for all terms and justify your assumptions.

T6.22 Why are the terms *depth of field* and *depth of focus* distinctly differently in TEM, but used interchangeably in SEM?

7

How to 'See' Electrons

CHAPTER PREVIEW

If we are studying the structure of a material, when all is said and done, all we have to show for learning how to operate our expensive TEM, the many hours spent in specimen preparation, etc., is an image or a DP. These images and DPs, which are just different distributions of electron intensity, have first to be viewed in some manner. After viewing, we have to decide if we want to save the results for future reference, perhaps so we can print out the data for a presentation, technical report, or scientific publication. Since, as we noted in the opening chapter, our eyes are not sensitive to electrons, we have to find ways to translate the electron-intensity distributions generated by the specimen into visible-light distributions, which we can see. This chapter will explain how we 'see' electrons.

We'll break the process down into two parts: first, detection (and display) of the image, and second, recording of the image. Both these areas are undergoing rapid change because of ongoing advances in electronic imaging and storage technology, and so this chapter will undoubtedly contain anachronisms by the time you read it. In particular, numbers are favored over photographic data; how can we quantitatively compare two photographs? Comparing two sets of numbers is routine.

7.1 ELECTRON DETECTION AND DISPLAY

As we saw back in Figure 2.1, images and DPs are different kinds of two-dimensional, electron-density distributions which are produced when a thin specimen scatters electrons. We detect and display them in different ways depending on whether we are using a TEM or STEM, as we'll explain in Chapter 9. In a conventional TEM, the images and DPs are static, because the incident beam is fixed, and so we can easily project them onto a viewing screen within the microscope column. TEM images, for example, are *analog* images of electron-density variations in the image plane of the objective lens. We cannot manipulate the image or its contrast in any way between the electrons leaving the image plane and being projected onto the viewing screen. So we will briefly discuss the properties of the viewing screen. The manufacturer controls the choice of screen materials so you might think there's not much need to understand this aspect in any depth. You might be surprised by the limitations you don't need to accept or the improvements which could be made.

When we operate our TEM as a STEM, or we use a dedicated STEM, the image is not static; it is built up over time as the small probe is scanned across the area of interest. Under these circumstances, we can detect the electron signals in several ways. If we are seeking secondary electron (SE) or backscattered electron (BSE) signals, then these detectors sit in the specimen stage area. If we are seeking the same forward-scattered electrons that we view on the TEM screen, the detectors are in the viewing chamber of the TEM. After we've detected any one of these signals, it is usually digitized and the *digital* scanning image is presented on a fluorescent screen as an analog image. You may hear this fluorescent screen referred to as the CRT, which are the initials for cathode-ray tube and a relic from the early days of electron physics. It is becoming much more common for the image or DP to be displayed on a flat-panel screen beside the main TEM column (or even on a plasma or LCD screen on the wall of the EM lab) controlled by the TEM's computer.

We should point out that the sequential or serial nature of the scanning image makes it ideal for on-line image enhancement, image processing, and subsequent image analysis. The signal from any electronic detector can be digitized and electronically manipulated prior to display on the CRT or computer screen, in a way that is impossible with analog images. We can adjust the digital signal to enhance the contrast or to reduce the noise. Alternatively, we can store the digital information and process it mathematically. The availability of cheap memory and

fast computers permits on-line image processing and the rapid extraction of quantitative data from the scanning image; we discuss all this and more in Chapter 31. Because of developments in computer technology, there is great interest in recording analog TEM images via a TV camera in order to digitize them; charge-coupled device (CCD) cameras are readily available for on-line viewing and processing, particularly of HRTEM images. CCD technology is advancing rapidly, driven largely by the digital-camera market and microscopists will continue to benefit from the availability of ever-larger CCD detectors. So we'll spend part of this chapter on CCDs which you'll have now worked out are equally sensitive to visible light and high-energy electrons.

In attempting to compare the properties of detection and recording devices we often use the concept of the 'detection quantum efficiency' or DQE. If a detector is linear in its response then the DQE is defined simply as

$$\mathrm{DQE} = \frac{\left(\frac{S_{\mathrm{out}}}{N_{\mathrm{out}}}\right)^2}{\left(\frac{S_{\mathrm{in}}}{N_{\mathrm{in}}}\right)^2} \qquad (7.1)$$

where S/N is the signal-to-noise ratio of the output or input signal. So a perfect detector has a DQE of 1 and all practical detectors have a DQE <1.

Note on terminology: We use several different terms, often imprecisely, to describe how we 'see' electrons. Since our eyes can't in fact see electrons, we have to resort to the phenomenon of cathodoluminescence (CL) (which we introduced back in Section 4.4) in order to provide an interface between electrons and our eyes. Any electron display system that we look at relies on CL at some point. The CL process converts the energy of the electrons (cathode rays) to produce light (luminescence). As a result, any electron display screen emits light in proportion to the intensity of electrons falling on it. A few definitions are in order

- *Light emission* caused by ionizing radiation is *scintillation*.
- The process of *fluorescence* implies *rapid emission*.
- *Phosphorescence* implies that the wavelength and the *delay time* are longer than for fluorescence.

All these terms are used in electron microscopy (interchangeably and often inaccurately) because the 'fluorescent' screen is coated with a long-delay phosphor (see Chapter 9).

7.2 VIEWING SCREENS

The viewing screen in a TEM is coated with a material such as ZnS, which emits light with a wavelength of ~450 nm. The ZnS is usually modified (doped) to give off green light at closer to 550 nm; hence you'll see screens of different shades of green which, being in the middle of the visible spectrum, is most relaxing for your eyes. As long as sufficient light is emitted, the main requirement of the viewing screen is that the ZnS particle (grain) size be small enough so that your eye cannot resolve individual grains. This means that grain sizes < 100 μm are acceptable (although you can see the grain size if you look at the screen through the auxiliary focusing binoculars). Typical screen coatings are made with a ZnS grain size of ~50 μm, although they may be as small as 10 μm for the highest-resolution screens.

As we've seen in Chapter 4, the cross section for inelastic interactions (and hence the emission intensity of most signals, including CL) decreases with increasing beam voltage. You would thus expect the light intensity to degrade at higher voltages, but this is offset by the increase in gun brightness. In some HVEMs the support for the small focusing screen is made of a heavy metal such as Pt to enhance backscatter and increase screen intensity, Of course, this backscattering will broaden the volume where light is generated and blur the image, so we don't gain very much. In fact most TEMs have very similar screens. Other signals are also given off by the viewing screen, such as X-rays, and whenever you look at the screen you are protected from this lethal radiation flux by lead glass, which is carefully selected to reduce transmitted radiation to levels at or below ambient background. In HVEMs this can amount to several tens of millimeters of glass and, invariably, the optical transmission capabilities are degraded as the glass gets thicker, but obviously we have no alternative if we want to view the screen directly.

A FEW WORDS OF CAUTION ABOUT YOUR SCREEN

There isn't much you can do about choosing the best material for the viewing screen since the manufacturer selects it for you, but you can extend its life substantially by taking care to minimize overexposure. The greatest source of screen damage is the intense direct beam that comes through thin specimens and constitutes the central spot in DPs. Using what you'll learn about operations of the TEM in Chapter 9, you can minimize burning of the screen by (a) only going to diffraction mode with the selected-area aperture inserted, (b) only going to diffraction mode with the C2 lens overfocused, and (c) if the spot appears exceptionally intense despite these precautions, then insert the beam stop while you're observing the pattern on the screen (but not when recording it).

While it is surprising that a modern TEM still relies on an analog screen, the end is perhaps already in sight. One of the latest TEM models (go back and look at Figure 1.9) is built without an operator's viewing screen; all the information is shown on flat-panel computer displays on a console that is separate from the column. Such a design breaks away from more than 70 years of TEM design but has the distinct advantages that

- Anyone in the room (or indeed anyone connected via the Internet) can see the images and DPs, which creates a much better teaching environment.
- The lights don't have to be out to view and record the information.
- The TEM column can be placed in a room that is separated from the operator, whose presence invariably reduces the resolution capabilities of the highest-performance microscopes.

Moving to digital display and recording brings with it the possibility of processing the image or DP to enhance or suppress information prior to publication or presentation. There are obvious ethical considerations here since the scientific community expects that published data be presented with sufficient background information that others would be able to reproduce and cross-check the experiment. So if you process digital images it is wise to publish the unprocessed data at the same time so others can see what data processing has been used. We'll talk a lot more about such ethical issues and related topics when we discuss image processing in Chapter 31.

7.3 ELECTRON DETECTORS

We have several alternatives to the fluorescent screen for detecting electrons. These other electron detectors play a major role in STEMs and AEMs (as well as in SEMs). They are actually essential to the STEM image-forming process that we'll describe in Chapter 9. Such detectors are usually one of two kinds: semiconductor (Si p-n junction) detectors or scintillator-photomultiplier systems. We'll examine the pros and cons of each of these two types and end with a section on CCDs.

7.3.A Semiconductor Detectors

A full understanding of how semiconductor detectors work requires a fair knowledge of solid-state physics. We'll just give a brief outline of the principles as they affect the use of the TEM but if you want to dig deeper the place to start is the excellent text by Pierret.

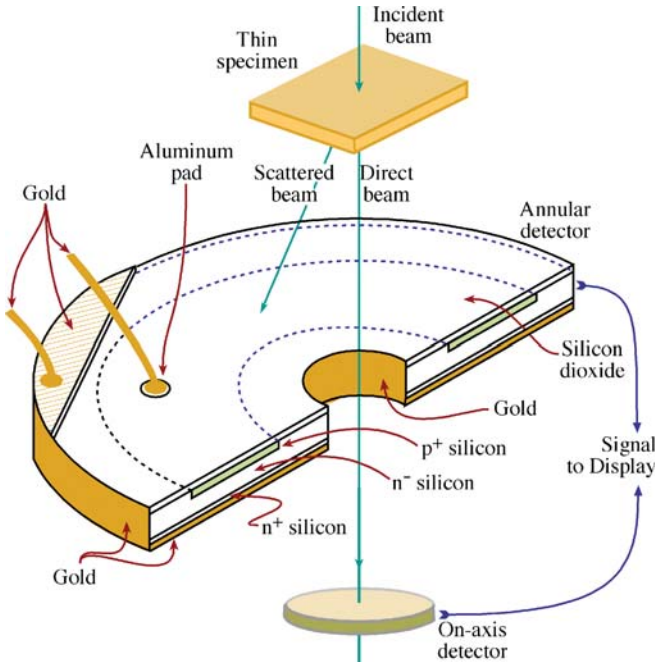


FIGURE 7.1. Semiconductor detector of the surface-barrier type, shown in a configuration where it would be used to detect high-energy, forward-scattered electrons. The direct beam is detected by a small circular detector on the optic axis of the microscope surrounded by a concentric wide-angle annular detector, which detects any scattered electrons.

The semiconductor detector, shown schematically in Figure 7.1, is a doped single-crystal sheet of Si (often inaccurately described as a solid-state detector). We make the Si into an electron-sensitive detector by creating a p-n junction beneath the Si surface in one of two ways. In one type of detector, we create the junction by doping the Si (e.g., by ion implantation of n-type impurity atoms into p-type Si or vice versa). This doping disturbs the equilibrium charge carrier concentration and creates a region across the p-n junction that is free of majority carriers which we call a 'depletion region.' A conducting metal layer is evaporated onto both surfaces to provide ohmic contacts. The alternative type of detector is called a surface-barrier detector (or sometimes a Schottky diode) and we fabricate this by evaporating a thin layer of Au on the surface of high-resistivity n-type Si, or evaporating Al onto p-type Si. This surface layer acts as an electrical contact and also creates a depletion layer and a p-n junction just inside the Si.

When we put either of these detectors into a beam of high-energy electrons, most of the beam energy is transferred to valence-band electrons in the Si which are excited across the band gap into the conduction band thus creating electron-hole pairs (see Figure 4.8). We can separate the electrons and holes most efficiently by applying an external reverse bias to the detector; that is, we put a negative bias on the p side of the junction and a positive bias on the n side. In practice, however, so many electron-hole pairs are created at TEM beam energies

that an external bias is not usually necessary, and the internal bias of the p-n junction acts to separate the electrons and holes. Because the electrons and holes move quite quickly in Si, it takes only a few nanoseconds to gather most of the carriers over an area of $\sim$1 μm^2. So the semiconductor detector is remarkably responsive to electrons. The net result is that the incoming electron signal is converted to a current in the external circuit between the surface contacts, as shown in the surface-barrier detector in Figure 7.1.

Since it takes approximately 3.6 eV to produce an electron-hole pair in Si at room temperature, a 100-keV electron can theoretically produce $\sim$28,000 electrons. This represents a maximum detector gain of close to 3×10^4 but in practice there are losses due to electron absorption in the metal contact layer and recombination of the electrons and the holes close to the Si surface (in a region called the dead layer), and we actually get a gain of closer to 2×10^4.

These semiconductor detectors are very efficient at picking up and amplifying electron signals. Unfortunately, they have an inherently large capacitance, so they are not very responsive to rapid changes in signal intensity. Such changes are quite likely to occur during the rapid scanning process of STEM imaging. In other words, the detector has a narrow bandwidth (typically 100 kHz); this is not a good property for a detector which is subject to widely varying signal intensities. We could lower the capacitance by decreasing the detector area, but if we do this, the signal-to-noise ratio will be lowered. It is the S/N ratio that ultimately limits the quality of all scanning images.

Semiconductor detectors have several advantages

- We can easily fabricate them.
- They are cheap to replace.
- They can be cut into any shape, as long as it is flat.

This latter advantage makes them ideal for squeezing into the confines of TEM stages and columns. For example, we can make the semiconductor detector in annular form so that the main electron beam goes through the hole in it, but the scattered electrons are very efficiently detected. This produces a dark-field (scattered electron) detector. We can also make detectors that are divided into halves or quadrants and each segment is insulated from the other(s). These detectors are very useful for discriminating directional signals such as those coming from magnetic specimens.

There are also some drawbacks to semiconductor detectors

- They have a large dark current (the current registered when no signal is incident on the detector). This dark current arises from thermal activation of electron-hole pairs, or from light falling on an uncoated detector. Since the detectors in a TEM invariably have a metal ohmic contact, the light problem is minimal because light can't penetrate the metal film. Now we could minimize thermal activation by cooling the detector to liquid-nitrogen temperatures but that step is impractical and introduces a cold surface into the vacuum which would simply collect contamination, so we live with noise due to the thermal activation.
- Because noise is inherent in the semiconductor detector, its DQE is poor for low-intensity signals, but rises almost to unity for high-intensity signals.
- The electron beam can damage the detector, particularly in intermediate voltage microscopes. In these circumstances, a doped p-n detector is less sensitive than a surface-barrier detector, because the depletion region is deeper in the Si.
- They are insensitive to low-energy electrons such as secondary electrons.

Despite these drawbacks, both types of Si detector are far more robust than the alternative scintillator, which we will now describe.

7.3.B Scintillator-Photomultiplier Detectors/TV Cameras

A scintillator emits visible light when struck by electrons because of the same CL process that occurs in fluorescent screens. While we are viewing a static TEM image, we want the fluorescent screen to continue emitting light for some time after the electrons hit it, so we use a long-delay scintillator. Of course, when we are using a scintillator to detect rapid changes in signal as in scanning beam imaging, we want the light emission to decay rapidly. So we don't use ZnS in scintillator detectors but rather materials such as Ce-doped yttrium-aluminum garnet (YAG) and various doped plastics and glasses. These materials have decay times on the order of nanoseconds rather than the microseconds needed for ZnS. Once we've converted the incoming electron signal to visible light, the light from the scintillator is amplified by a photomultiplier (PM) system, attached to the scintillator via a light pipe. Figure 7.2 shows a schematic diagram of a scintillator-PM detector setup to detect secondary electrons in a TEM, but the design used to detect primary scattered electrons in the STEM is essentially identical.

The scintillators that we use in STEMs or SEMs are often coated with a 100-nm-thick layer of Al to reflect any light generated in the microscope and stop it from entering the PM tube where it would add noise to the signal. If the detector is in the stage of the microscope, this light could come from the specimen itself if it is cathodoluminescent, or it could be light coming down

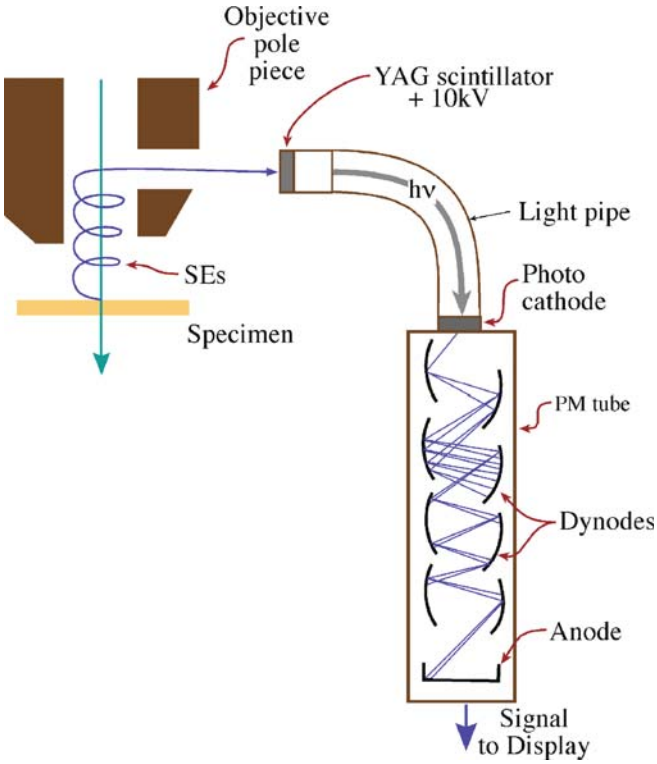


FIGURE 7.2. Scintillator-photomultiplier detector system for SE detection in a TEM. SEs from the specimen spiral back up through the objective lens polepiece and are accelerated by the high voltage onto the scintillator, generating visible light which travels via fiber optics to a photocathode. There the light is re-converted to electrons. The electron signal is then multiplied by several electrodes (dynodes) in the PM tube before being used to modulate the display screen.

the column from a thermionic source and reflected from the polished surface of the specimen. If you have an uncoated scintillator detector in the viewing chamber, then room light may also hit the detector, so you should cover the windows of the viewing chamber.

The advantages of the scintillator-PM system are

- The gain of the system is very high. The gain for the total detector system is of the order of 10^n, depending on the number (n) of electrodes (often called dynodes) in the PM. A value of 10^8 is not unusual (compare with $\sim 10^4$ for the semiconductor detector). This performance is reflected in a typical DQE of close to 0.9 for several commercial scintillators.
- The noise level in a scintillator is low compared with semiconductor detectors, and the bandwidth of the scintillator is in the MHz range. As a result, both low-intensity images and TV-rate images are easily displayed. There is a tremendous practical advantage to TV-rate imaging of digital signals, because such images, when suitably processed and displayed can be viewed, stored, and recorded under normal conditions of room illumination. So you don't have to work in the dark while operating your (S)TEM.

The disadvantages of the scintillator-PM system are

- The scintillator is not as robust as the semiconductor detector, being even more susceptible to radiation damage, particularly after longtime exposure to the beam.
- The scintillator-PM combination is also substantially more expensive and bulky compared to semiconductor detectors and therefore it does not fit well within the TEM stage, nor is it easily manufactured into multi-detector configurations; it is also more expensive. However, plastic scintillators can be shaped to give large collection angle, such as the Robinson BSE detector used in many SEMs.
- The energy-conversion efficiency of a scintillator is also rather low ($\sim$ 2%–20%) compared to a semiconductor detector and, typically, we only get about 4000 photons per incident 100-keV electron, $\sim 7\times$ less than for the semiconductor detector. This low efficiency is offset by the gain in the PM tube.

On balance, the scintillator-PM detector is preferred over the semiconductor detector for most general electron detection in TEM/STEM systems. However, you must take care to minimize any high-intensity beams that may damage the detector and lower its efficiency. Therefore, you need to take more care when operating scintillator detectors.

We've already mentioned that you can view the TEM image directly through a TV camera, rather than looking at the fluorescent screen. There are real advantages to TV cameras, e.g., for on-line viewing of faint HRTEM images (see Chapter 28) or for recording of dynamic in-situ events (see Chapter 29). Also, from a teaching standpoint, anything that pulls the TEM image, in real time, out of the viewing chamber (which only the operator can see into clearly) and onto a classroom or laboratory computer screen or plasma display makes life so much easier. With the increased interest in telemicroscopy, TV cameras and webcams are becoming much more common within the TEM room (e.g., URLs #1 and #2). TV cameras attached to TEMs come in both analog and digital forms. Typically the cameras are placed below the viewing screen so you have to lift up the screen to detect the TV image. The camera may have to be offset if there is another post-column attachment such as an EELS system. Sometimes the camera is placed within the column and is then moved on axis when needed.

Analog YAG-based scintillator TV cameras may be used for applications requiring image intensification if you are dealing with faint images (e.g., if you are using low-dose techniques because your specimen would otherwise be damaged by the beam). Also, wide-angle

cameras are available that collect a much greater area of the image or DP than the standard TEM photographic plate. However, the most widely used TV cameras use only digital-detection technology, which we'll now discuss.

7.3.C Charge-Coupled Device (CCD) Detectors

Electronic technology for recording images and spectra is rapidly closing in on the more traditional analog methods. CCD cameras are becoming the norm for real-time TV recording of images and DPs. They are also being used for two–dimensional arrays for parallel-collection EELS and energy-filtered images, as we describe in Chapter 37.

CCDs are metal-insulator-silicon devices that store charge generated by light or electron beams. CCD arrays consist of several mega- (millions of) pixels which are individual capacitors electrically isolated from each other through the creation of potential wells under each CCD cell, so they can accumulate charge in proportion to the incident radiation intensity, as shown in Figure 7.3A. The largest CCD arrays, as of writing this text are giga-pixels (10^9). Because such systems are so expensive, they are typically developed for use in major astronomical telescopes for detecting very faint light sources. In fact there isn't a good reference text for CCD use in electron microscopy but there is a great one by Howell for astronomy that gives excellent background material if you are interested in digital-image recording.

The maximum size of CCD currently available for TEM use is 4k×4k but the size will only increase with time. (It is also possible to stitch together multiple CCD images using software so, if you have the time, the size of the CCD itself is not a serious issue.) The individual cells currently can be as small as 6 μm although a more typical size range is ~10–15 μm. To create a picture, we have to read out the array. We do this by changing the applied potentials to transfer the charge serially from each potential well along a line in the array into an output amplifier, as shown in Figure 7.3B. With good design of the electrodes, charge-transfer efficiencies of 99.999% can be achieved. Once all the cells are empty the array can be re-exposed. This so-called 'full-frame' design is simple and robust and offers the highest resolution and highest pixel density.

Rather than serial readout of full-frame CCDs, it is also possible to have frame transfer CCDs in which the whole frame is transferred to an adjacent storage array leaving the main array free to collect a new signal flux. This method allows for shorter frame times, and thus a faster image acquisition, but is much more complex and, since more of the device is taken up with the storage, the CCD has lower resolution and much higher cost. The typical frame times needed to record TEM images are long enough that high-speed frame transfer is not usually needed.

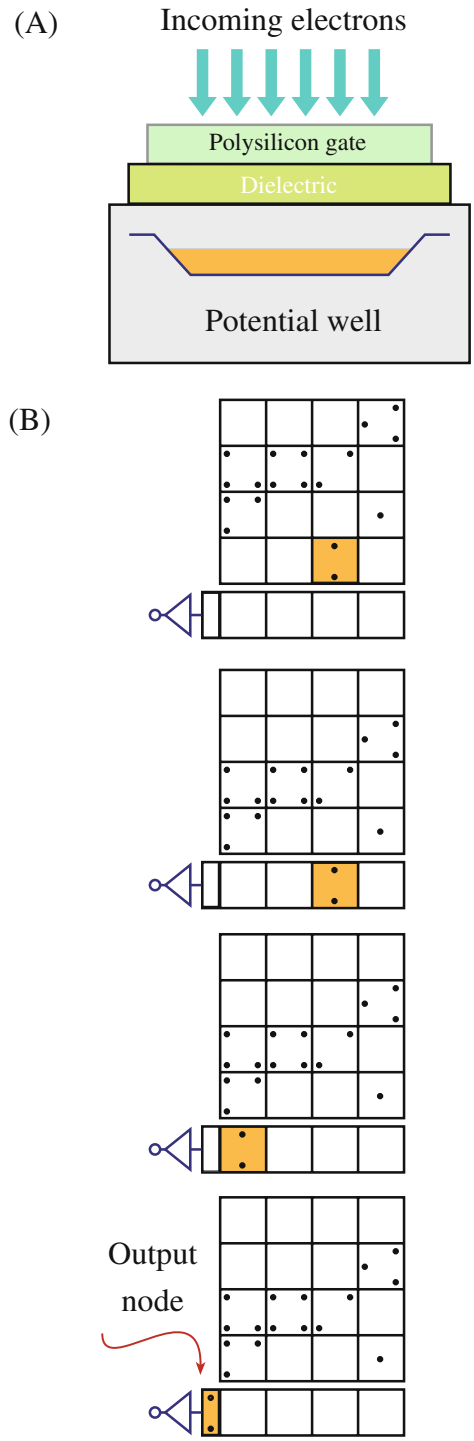


FIGURE 7.3. (A) A single cell in a CCD array showing the storage of charge in the potential well under one pixel. If we vary the applied potential to rows of pixels in sequence as in (B), one pixel row is shifted to the parallel register and is read out pixel by pixel, after which the next row is moved to the parallel register, and so on. The stored charge in each pixel is thus fed into an amplifier and digitized.

The frame time for reading the CCD depends on the size of the image and the specific technology used to readout the detected signal. Ultrahigh-speed CCD cameras are available with $>10^5$ frames/second, but such high speeds are not essential in standard TEM. It is worth noting, however, that time-resolved TEM is an area of growing importance and, in such dedicated instruments, ultra-fast recording is required. Routinely, frame times of <0.001 seconds, well below standard TV rates of 0.033 seconds, are available for in-situ recording of fast events. But the frame time can equally well be several minutes (e.g., for picking up diffuse scattering in faint DPs). Obviously, the longer the exposure, the more the image is susceptible to external vibration, drift, etc.; so long exposures, e.g., for HRTEM images, are not good.

CCD detectors have several advantages

- When cooled, they have very low noise and a good DQE (>0.5) even at low input signal levels.
- The dynamic range is high, making them ideal for recording DPs which can span an enormous intensity range.
- They respond linearly to changes in input signal and show a fairly uniform response across many pixels.

There are some drawbacks to CCDs, not least of which is their cost, but that is always decreasing, as with any Moore's-law-based technology. However, 'blooming,' which occurs when too much signal fills up the pixel and the signal overflows into surrounding pixels, can be a problem. This problem can be overcome substantially by building anti-blooming or overflow drain structures within the device. Apart from these minor factors it is clear that, in the end, CCDs, or other electronic technology, will eventually record and store all TEM images, DPs, and spectra.

7.3.D Faraday Cup

In conventional TEM there isn't much need to know the beam current, but for X-ray analysis in the AEM, it is essential, since there is often a need to compare analytical results obtained under identical beam current conditions. A Faraday cup is a detector that simply measures the total electron current in the beam. We don't use it for any imaging process, but rather as a way of characterizing the performance of the electron source as we saw in Chapter 5. Once the electrons enter the Faraday cup, they cannot leave except by flowing to ground through an attached picoammeter that measures the electron current.

You can easily construct a Faraday cup to go in an SEM, but it is more difficult to design one that fits in the stage of a TEM. A dedicated Faraday cup holder is shown in Figure 7.4A. The entrance aperture is small and the chamber is relatively deep and lined with a low-Z material to minimize backscatter. If you tilt it slightly, the electrons have little chance of being scattered directly back. With such a holder you can only find the hole if you can image the upper surface with SE or BSE detectors, and if these are not available then you must have a cup with a hole in the lower surface too. When the cup is not tilted, the electrons go straight through; if you tilt the cup, then all the electrons are trapped as shown in Figure 7.4A. The way to ensure that you are measuring the maximum current is to look at the picoammeter reading as you tilt the cup. Some manufacturers now incorporate a Faraday cup in the specimen holder. You can measure the current by deflecting the beam into the cup or partially extracting the holder so the beam falls into the cup (Figure 7.4B).

FARADAY CUP

Remember: A Faraday Cup is a Black Hole for Electrons.

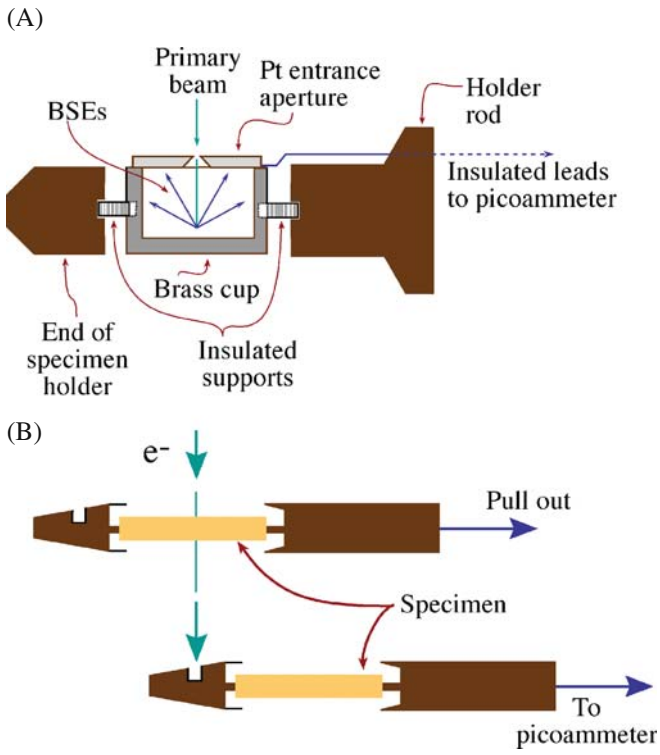


FIGURE 7.4. (A) Schematic diagram of a dedicated Faraday cup in the end of a side-entry specimen holder (more details about these holders in Chapter 8). The entrance aperture has to be found by imaging the top surface using SEs or BSEs. In (B) the specimen holder is retracted so the electrons fall into a cup on the tip of the holder (in which position of course the TEM image of your specimen cannot be seen). In either case, the electron current is measured as it goes to ground through a picoammeter attached to the outside of the holder.

If you don't have a Faraday cup, it is possible to get an approximate reading of the current by just measuring the current through an insulated line from a bulk region of your specimen and correcting for electron backscatter. Backscattering is independent of the accelerating voltage and approximately linear with atomic number up to about $Z = 30$. For example, the backscatter coefficient for Cu is about 0.3 and for Al is about 0.15. It is also possible to deflect the beam onto the last beam-defining diaphragm (see Chapters 6 and 9) and measure the current via an insulated feedthrough (also correcting for backscatter).

7.4 WHICH DETECTOR DO WE USE FOR WHICH SIGNAL?

As we mentioned at the start of the chapter, the principal electron signals that we can detect are the forward-scattered electrons (which as we'll see in Chapter 9 form the most common TEM images) and the BSE and SE signals from the beam-entry surface of the specimen.

Semiconductor detectors are only sensitive to electrons with sufficient energy (>5 keV) to penetrate the metal contact layer. So we use these detectors mainly for *high-energy* forward-scattered imaging and *high-energy* BSE imaging. Because of the surface contact layer we don't use semiconductor detectors for *low-energy* SEs but instead use a scintillator-PM system. Remember that the scintillator may also be coated with Al to prevent visible light from generating noise. This coating would also prevent low-energy SEs from being detected. So for SE detection, either there must be no coating or the electrons must be accelerated to an energy high enough to penetrate the coating; we achieve the latter by applying a high kV (>10 kV) positive bias to the scintillator.

The capacitance is relatively high for semiconductor detectors so they are not the detector of choice in dedicated STEMs where high scan-rate TV images are the normal viewing mode, i.e., you need a quick response. The scintillator-PM system is again preferred under these circumstances. As most microscopes move toward TV-rate display of scanning images it is likely that the scintillator-PM will be used increasingly for forward-scattered TEM imaging. Semiconductor detectors may only be used for BSEs, which is not a major imaging mode in TEMs. A summary of all the various electron detectors in a TEM/STEM is given in Figure 7.5. We'll talk more about the methods of imaging in Chapter 9 and later in Part 3 but the primary detectors in (S)TEM pick up the forward-scattered electrons on-axis (called the bright-field (BF) detector), forward-scattered through small angles $<\sim 3^\circ$ (called the annular dark-field (ADF) detector), or scattered out to higher angles (called the high-angle annular dark-field (HAADF) detector).

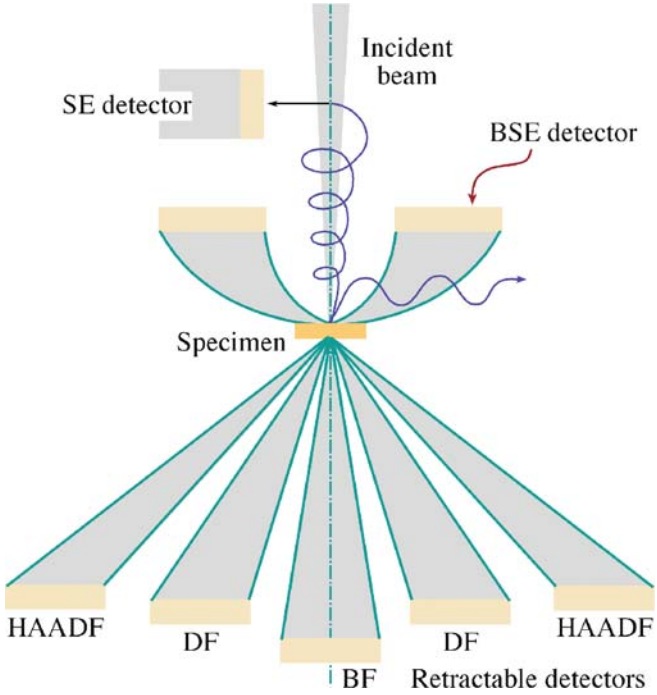


FIGURE 7.5. The various electron detectors in a STEM. Scintillator-PM detectors are invariably used for SE detection and semiconductor detectors for the BSE. The on-axis and annular forward-scattered and high-angle dark field detectors may be either type, depending on the microscope. The SE detector is rare, the BSE detector is a waste of time: only the forward-scattered electron detectors are standard.

Sometimes we examine specimens which themselves exhibit cathodoluminescence under electron bombardment. We discussed CL back in Chapter 4 and we'll give an example of why CL imaging might be useful in Chapter 29. A mirror is used to focus the light from the specimen into a scintillator-PM system; one design is shown in Figure 7.6. In this setup, the specimen must be moved and tilted until, in combination with the collimating lens, the maximum signal is detected by the PMT. This setup effectively prevents detection of most other signals, including X-rays, because the mirror occupies all the free space in the TEM stage. So you have to dedicate the TEM to CL detection alone and ignore other signals. There are only a few such CL-TEMs in the world.

7.5 IMAGE RECORDING

7.5.A Photographic Emulsions

Although photographic film is the oldest recording medium, it still retains sufficient advantages that we continue to use it in some TEMs (and most that are >10 years old). Photographic emulsions are suspensions of silver halide grains in a gel. Electrons strike the

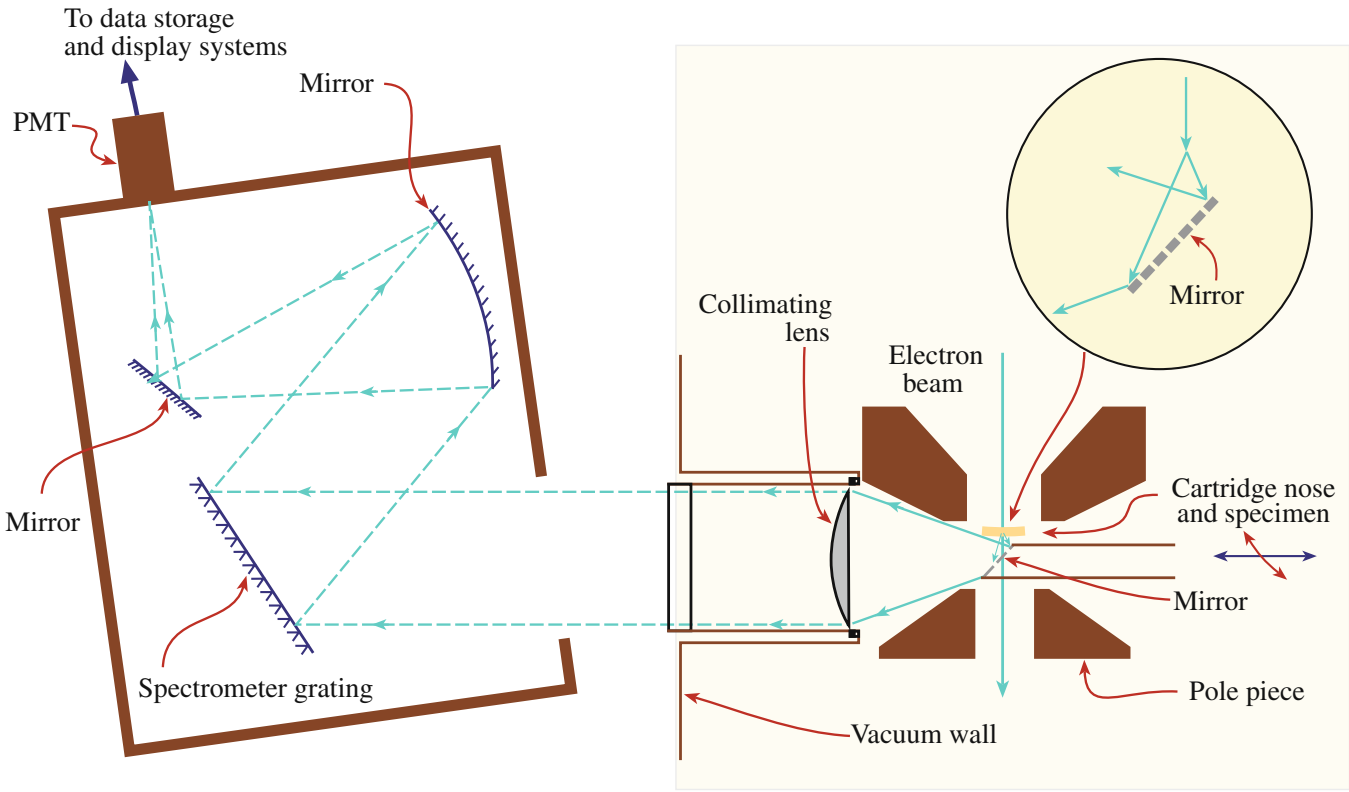


FIGURE 7.6. Cross-sectional diagram of a mirror detector below a thin cathodoluminescent specimen that collects a small fraction of the emitted light and focuses it via a collimating lens into a spectrometer-PM system. The CL signal is usually very weak to start with and so the detector has to be as large as possible, and it takes up much of the free volume in the TEM stage making it impossible to have any other detectors in the stage.

halide, ionize it, and transform it to silver. The emulsions are supported on a polymer film which, unlike the earlier glass plates, outgas and shrink during pre-pumping, or processing. However, glass plates were heavy, occupied an enormous volume compared to film, and Murphy's law meant that your best plates invariably broke because you spent more time handling them. Many aging microscopists still call their EM films 'plates.'

If you are still using film, you have a choice of photographic emulsions, just as you do for your camera (since you'll probably still have a film camera if you're using TEM film). Different speed emulsions are available, with the usual compromise that faster film means a larger grain size and therefore less resolution.

- In principle, for the highest-resolution images, the slowest (finest grain) film is best.
- In practice, we usually minimize the exposure time and go for the fastest film.

We usually want to minimize beam damage and blurring due to movement (drift) of the specimen/stage, so we keep the exposure short. In fact grain sizes for the faster film are about 5 μm compared to about 4 μm for the slowest film, so we don't lose much resolution. The loss of resolution is more than offset by the shorter exposure times which minimize the overall dose to the specimen. The only time you may need to use slower film is if you have a problem with poor image contrast. This problem is more common when imaging amorphous, biological, or polymer specimens.

Although the grain size of the emulsion may be as small as a few micrometers, the actual resolution of the recorded image is worse than this because of electron spreading in the emulsion. The practical resolution may only be about 20–50 μm. Despite this degradation we still have more than 10^7 picture elements or pixels available to store information in a standard 100 mm × 100 mm image (Kodachrome™ film has the silver halide equivalent of 1.8×10^7 pixels). Film has a high DQE, although its dynamic range is rather limited. What this means is that you can easily saturate the film (change all the halide to silver) and you lose any linear relationship between electron intensity and gray scale in the film. As we've already noted, CCDs have a high dynamic range, are linear in response, uniform in output, have anti-blooming technology and the latest CCDs boast $>10^9$ pixels. Perhaps the end is indeed in sight!

Removing photographic film from the TEM will be a major operational improvement, because the absorbed water degrades the vacuum.

PHOTOGRAPHIC FILM

A photographic emulsion on a polymer support is one of the worst things you can put into a high-vacuum instrument.

Both the emulsion and the support outgas, which is a major contribution to the residual high pressure of hydrocarbons and water vapor in the instrument which, in turn, cause contamination of your specimen.

We used to use Polaroid instant film for recording scanning images from the CRT display in STEM mode and there may be a few older instruments around that still require such. But it will generally be good to see the end of any kind of film because they were analog, expensive, used nasty chemicals, and created a lot of mess around the TEM lab. Also there was always the chance that someone before you had loaded the film incorrectly, thus rendering your precious time on the microscope useless and subsequently provoking occasional melees in graduate-student hangouts.

7.5.B Other Image-Recording Methods

Digital images can be stored and retrieved, for example, magnetically on hard drives, or optically on compact discs, DVDs, or, most recently, holographic storage disks. These devices are cheaper and easier to use than photographic recording and images on an optical disc will not degrade with time even after years of storage. However, the usual drawback to digital-storage technology is the question of how long will the technology be around to read any particular stored image. How many images do you still have on a 3.5" floppy? Probably not many! Or a zip drive? Perhaps lots more but you can't read them anymore.

An alternative is the image plate, manufactured by Fuji, which is a kind of re-usable, digital photographic plate. The plates can be used, read, and re-used and this capability permits on-line image processing of the data. While this technology is seeing a lot of use in medical X-ray laboratories, it has not caught on in the TEM field, mainly because it is so expensive.

To present a stored image for publication you still have to print it in some way; photographic methods are still occasionally used. However, laser printers now have the quality required for publishing the highest resolution images (1200 dpi or 48/mm) means that the dot size is well below the resolving power of human eye. Likewise the advent of artificial coloring of electron images, DPs, and spectral images over the last decade or so means that photographic methods of printing are becoming completely obsolete because color photography was never a satisfactory individual lab process.

7.6 COMPARISON OF SCANNING IMAGES AND STATIC IMAGES

We have a choice of creating analog static images in conventional TEM mode or digital scanning images via electronic detection and display. Which is best? While we can only form BSE and SE images in a scanning mode, the answer is not clear for conventional BF and DF images, and the answer depends somewhat on the contrast mechanism that is operating in the specimen, as we'll see in Chapter 22. Regardless of which detector you use, scanning images are always displayed on a computer screen, and this limits the amount of information in the image. Typically the viewing screen will have up to 10^3 lines with a maximum of 10^3 pixels per line, giving a total of 10^6 pixels in each frame. Currently, high-definition TV displays offer at least 1920 pixels per line and 1080 lines per frame giving a total of $2\times\ 10^6$ pixels. In contrast, as we just noted, a TEM image recorded directly onto photographic emulsion will have a higher information density, with $>10^7$ pixels of information available in a 100 mm $\times$ 100 mm image. Furthermore, if a scanning image is to be recorded in a reasonable time, the electron beam can only stay on each point in the image (i.e., each pixel on the display) for a very short time. Typical dwell times per pixel are $<<1$ ms and this means that the signal-to-noise ratio in a scanning image is liable to be quite low. The combination of the lower pixel density compared to a photographic emulsion and the short dwell times means that, almost invariably, STEM images are poorer in quality than static TEM images. However, with the increasing availability of FEG (S)TEMs, increased probe current through C_s correction and improved detection and display technology, STEM digital picture quality compares quite well with analog TEM images. The former technology will continue to improve with the main challenge being our ability to read/access data; the latter will continue to fade away, slowly turning to shades of sepia with increasing time.

CHAPTER SUMMARY

Although it might seem surprising for such a high-end scientific instrument, the TEM is still in the age of analog images. We look at fluorescent screens and computer displays and we still record some of our pictures on photographic film. But darkrooms have disappeared from many labs and the whole area of electron detection is in a state of rapid flux as new electronic technology develops. Semiconductor detectors, scintillators, and CCDs all bring with them the advantage of digital signal collection and therefore the images can be processed and subsequently stored either magnetically or optically. As we said back in 1996, "anything we say about this technology will probably be obsolete before it is published." It is probably safe to speculate that most analog detection, recording, and storage of images and DPs will eventually be replaced by digital methods and the CCD manufacturers are already pronouncing 'the end of film.' So, TEM will produce numbers, but remember that we can all interpret images on film from the 1880s. Can you read data from computer punchcards from the 1970s or from zip disks of the 1990s? A final thought: we encourage you to read the original papers on TEM. All those images were recorded using a photographic emulsion.

REFERENCES

The general references for SEM are the standard book by Goldstein et al. (3rd Ed.) and Reimer's SEM text. The other references here are interesting (often challenging) reading.

Chapman, JN, Craven, AJ and Scott, CP 1989 *Electron Detection in the Analytical Electron Microscope* Ultramicroscopy **28** 108–117.

Howell, SB 2006 *Handbook of CCD Astronomy* 2nd Ed. Cambridge University Press NY.

Knoll, GF 2000 *Radiation Detection and Measurement* 3rd Ed. John Wiley & Sons NY.

Pierret, RF 1996 *Semiconductor Device Fundamentals* Addison-Wesley Boston MA

Reimer, L 1985 *Scanning Electron Microscopy* Springer Verlag New York.

URLs

1) http://tpm.amc.anl.gov/
2) http://telescience.ucsd.edu/gts.shtml

SELF-ASSESSMENT QUESTIONS

Q7.1 How do we 'see' electrons?
Q7.2 What are viewing screens usually coated with?
Q7.3 How do you prolong the lifetime of your screen?
Q7.4 Why is green chosen as the color of the light emitted by the screen?
Q7.5 What is DQE?
Q7.6 What is CL?
Q7.7 How do semiconductor detectors work?
Q7.8 What are the advantages and disadvantages of semiconductor detectors?
Q7.9 How does a scintillator-photomultiplier detector work?
Q7.10 What are the advantages and disadvantages of scintillator-photomultiplier detectors?
Q7.11 What are the advantages of CCD cameras?
Q7.12 What is a Faraday cup?
Q7.13 Which detector should be used for which signal in the TEM?
Q7.14 What is a photographic emulsion, and how does the chemistry work?
Q7.15 What is the difference between scintillation, fluorescence, and phosphorescence?
Q7.16 Why are ZnS-based viewing screens doped?
Q7.17 Why aren't semiconductor devices cooled with LN_2 to reduce the large dark current as they are when used as X-ray detectors (see Chapter 32)?
Q7.18 Give a major disadvantage of photographic film when used in a TEM.
Q7.19 How do you deal with the problem of recording images of specimens with poor image contrast?
Q7.20 What does a 'static' TEM image mean, as opposed to a 'scanning' image?

TEXT-SPECIFIC QUESTIONS

T7.1 Why do we still use analog viewing screens in nearly all TEMs rather than viewing all images and DPs digitally on a computer screen? What would be the biggest advantages of removing the viewing screen?

T7.2 Why do we still sometimes record TEM images and DPs on photographic plates rather than always capturing and storing them digitally?

T7.3 Examine Figure 7.1 and explain why this kind of detector is good for imaging high-angle elastically scattered electrons?

T7.4 Why is this kind of detector (Figure 7.1) not very good at imaging DPs such as those in Figure 2.13B and D?

T7.5 Why would you not use the scintillator-PM detector for the on-axis detector to pick up the direct beam signal in Figure 7.1?

T7.6 Why is the scintillator-PM detector even better than the semiconductor detector for imaging very high-angle, elastically scattered electrons?

T7.7 Go on the Web and find the smallest pixel size available in a commercial CCD camera. How does this dimension compare with the emulsion size in a typical high-resolution photographic film?

T7.8 Go on the Web and find the largest number of pixels available in a commercial CCD camera. How does this number compare with the effective number of pixels in a typical high-resolution photographic film used in a TEM?

T7.9 A problem with recording DPs on CCD cameras is that the intense direct beam can cause 'blooming.' What is blooming and how can it be corrected? (Hint: go on the Web and look for non/anti-blooming cameras.)

T7.10 Why are CL detectors much less common in TEMs than in SEMs? (Hint: take a look at Figure 7.6 and consider the relative sizes of the CL source and the CL detector.)

T7.11 Even if you can, why should you not buy (or bother to use) a BSE detector in your TEM?

T7.12 Go online and see how many TEMs are available for you to watch in someone else's lab? Can you think of specific advantages that could be gained by this accessibility? Can you think of experiments that you could do on remotely accessible TEMs that you could not do in your own laboratory?

T7.13 Why is CL imaging a relatively common imaging method in the SEM but rarely used in the TEM?

8

Pumps and Holders

CHAPTER PREVIEW

In the past three chapters we've described the sources, lenses, and detectors that make up a TEM. The only other parts of the instrument you need to know about in detail are those that, if you are not careful, can seriously degrade the quality of the information you generate even if the rest is perfect. These two parts are the holder in which you put your specimen and the vacuum that surrounds it. While there isn't much you can do to improve the vacuum, beyond buying a better microscope, there is a lot you can do that will degrade the quality of the vacuum in the column and, in doing so, contaminate your specimen. So we'll tell you a few basics about how the vacuum pumps work, and how the vacuum system is put together. Although the vacuum system is under computer control in most TEMs, you still affect the vacuum by what you put in the microscope. Consequently, you need to know what not to do on those occasions when you might otherwise degrade the vacuum.

The vacuum in the stage of a typical TEM is $\sim 10^{-5}$ Pa, compared with atmospheric pressure of $\sim 10^{5}$ Pa. It is quite remarkable that we can transfer a specimen into the TEM, reducing the ambient pressure at its surface by 10 orders of magnitude in a matter of a few seconds. This rapid transfer is a testament to the skills of TEM designers, and particularly the construction of the specimen holder and the airlock system. Specimen holders are the physical contact between you and your specimen across this extraordinary vacuum range. You must transmit all the experimental variables that you want to inflict on your specimen by way of the holder. The most basic requirement is that you should be able to move the specimen laterally to look at different areas; to optimize the imaging you should also be able to move the specimen vertically. In addition we'll describe how you can tilt, rotate, heat, cool, strain, and bias the materials that you are studying. Unfortunately, the holder also transmits vibrations, drift, and contamination to the specimen and may be a source of X-rays that can degrade any analysis that you want to perform. Care of your specimen holders is extremely important since damaged or worn holders reduce the quality of the data generated by the microscope. If you are not careful, a $10,000 holder can easily limit the information generated by a million-dollar TEM.

8.1 THE VACUUM

You know already that electrons are strongly scattered by atoms, which accounts for the versatility of TEM, and the need for thin specimens. Strong scattering also occurs in gases and we can't transmit coherent, controlled electron beams very far through air, so all EMs operate under vacuum. This means that your specimen has to go through an airlock into the TEM. Therefore, you can only control your specimen remotely, not directly, and this makes TEMs more expensive to build. In addition to permitting the electron beam to travel through the instrument undisturbed, the vacuum also plays a role in keeping the specimen clean (or making it dirty). Contamination of the specimen by vacuum-borne contaminants such as hydrocarbons and water vapor can be a problem in many aspects of TEM. Generally, the better the vacuum, the less contamination, but it is the partial pressure of contaminants, not the absolute pressure, which is important. Fortunately, the vacuum systems in most TEMs today are reasonably clean, fully automated and their operation is transparent to the user. Despite this, you should have some understanding of vacuums and how to control them, so this chapter will cover, very superficially, the principles of vacuum systems and pumps.

First of all, a word on units which, as usual, are in disarray. The SI unit of pressure is the pascal; other non-SI units are the torr and the bar. You'll come across all three units in TEM texts and in manufacturers' handbooks, so you need to know the conversions.

We'll mainly use the pascal, but since the torr is still very common terminology, we'll occasionally put approximate torr values in parentheses to remind you of the conversion. Since we deal with very low pressures, the numbers are small, although we perversely use the expression 'high vacuum' for these low pressures. We think of vacuum in terms of rough, low, high, and ultrahigh. A roughing pump gives a pressure between 100 and 0.1 Pa; 0.1–10^{-4} Pa is low vacuum and 10^{-4}–10^{-7} Pa is high vacuum (HV). If the pressure is $<10^{-7}$ Pa you have an ultrahigh vacuum (UHV). These values are approximate, not standardized definitions. A typical modern TEM has a pressure inside the column of $\sim 1.3 \times 10^{-5}$ Pa (10^{-7} Torr), which is in the HV range. UHV TEMs operate below 10^{-7} Pa and the gun region of an FEG TEM operates at $\sim 10^{-9}$ Pa (10^{-11} Torr). To have an electron beam inside the TEM that is not scattered significantly by the air molecules in the column, the pressure must be $<\sim$0.1 Pa. This was achievable with simple mechanical pumps in the early days of the instrument, but there are good reasons to operate at much lower pressures (higher vacuums), for which you need more sophisticated and more expensive apparatus.

PRESSURE

1 Torr is $\sim$ 130 Pa
1 Pa is 7.5×10^{-3} Torr
One bar is atmospheric pressure ($\sim$760 Torr) and is equivalent to $\sim 10^5$ Pa.
The name is the torr; the unit is the Torr, but either way the torr is not an accepted SI unit.
100–0.1 Pa ($\sim$1–10^{-3} Torr) is a rough vacuum.
0.1–10^{-4} Pa ($\sim 10^{-3}$–10^{-6} Torr) is low vacuum.
10^{-4}–10^{-7} Pa ($\sim 10^{-6}$–10^{-9} Torr) is high vacuum (HV).
$<10^{-7}$ Pa ($\sim 10^{-9}$ Torr) is ultrahigh vacuum (UHV).
Be careful when you hear a phrase like "the vacuum in the gun is 10^{-8}" and remember the pascal unit is Pa and the torr unit is Torr.

Generally we use one type of pump to create a rough vacuum and another type to create the higher vacuum. The TEM is kept permanently under vacuum, unless it's being repaired or serviced. If you need access to the inside of the column to change specimens, electron sources, or photographic plates, you do this via an airlock system, which can be pumped separately, as we'll explain later. There are many different kinds of pumps used in TEMs, and you often have a choice when purchasing an instrument. As with most things, you get what you pay for; a clean UHV system is very expensive. We can divide pumps into roughing pumps and HV/UHV pumps, as we'll now discuss.

8.2 ROUGHING PUMPS

The most common roughing pump is a mechanical (rotary) pump in which a belt-driven, eccentrically mounted reciprocating mechanism sucks air through an inlet valve into a chamber and expels it through an exit valve, as shown in Figure 8.1. Such pumps are very reliable, relatively inexpensive, noisy, and dirty and only lower the pressure to $\sim 10^{-1}$ Pa ($\sim 10^{-3}$ Torr). Mechanical pumps should be housed outside your TEM room, and connected to the column through a line that doesn't transmit their vibration. These pumps use a hydrocarbon oil as a medium. If you have such a pump, the line from the pump to the vacuum should contain a foreline trap to condense out oil vapor before it is deposited in the column. Also, the exhaust line from the pump must also be trapped to prevent (possibly carcinogenic) oil vapor escaping into the room where you are working. There are alternative 'dry' roughing pumps which do not use oil. These are more expensive and somewhat less reliable; they do not pump to such low pressure.

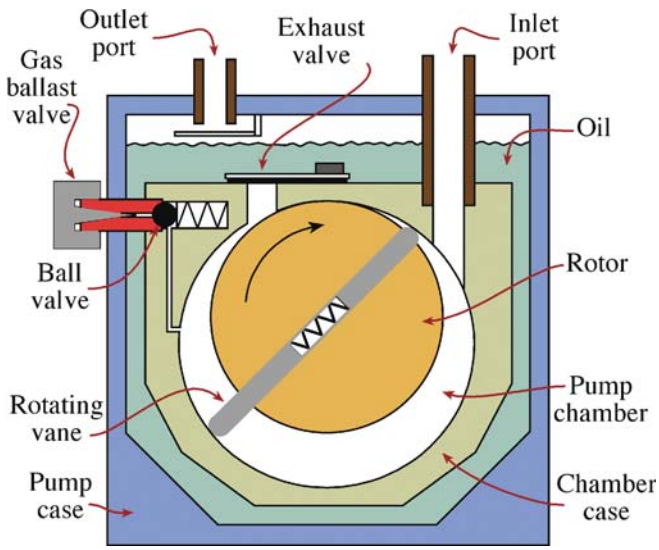


FIGURE 8.1. A mechanical pump for roughing vacuums. The eccentric motion of the pump creates a vacuum in the RH side when it rotates and the vacuum sucks air into the inlet valve. As the cylinder rotates further, it cuts off the inlet and forces the air through the outlet on the LH side, creating a vacuum again on the inlet side as it does so. Because of the constant contact between the rotating cylinder and the inside of the pump, oil is needed to reduce frictional heating.

THE ROOM

The microscope room should be acoustically quiet. You should never feel a draft. All computers with fans should be replaced or moved out of the room—the fan just moves heat around the room. All monitors should be flat panels.

8.3 HIGH/ULTRAHIGH VACUUM PUMPS

8.3.A Diffusion Pumps

These pumps use a hot plate to boil oil, which then forms a series of concentric vapor jets. The jets drag air molecules out of the microscope as shown in Figure 8.2, then condense onto a cold surface, freeing the air molecules which are extracted by the mechanical pump 'backing' the diffusion pump. While this may seem an inefficient way to move air, diffusion pumps can in fact transport more than a hundred liters of air per second, which is quite sufficient to pump out a TEM column. With no moving parts, diffusion pumps are inexpensive and very reliable, but they need external water cooling to aid condensation of the vapor. Failure of the cooling water supply and burnout of the hot plate are about the only possible causes of failure. The absence of moving parts ensures vibration-free operation. As with the mechanical pump, the oil diffusion pump would contaminate the vacuum in the TEM if oil vapor were to escape into the column. To minimize this you must use synthetic nonhydrocarbon oils with low vapor pressures, such as Fomblin™ or Santovac™. (Never use a silicone-based oil, of course.) A liquid-N_2 cold trap sits on top of the pump and condenses out any residual oil molecules. If you have diffusion pumps you must keep the cold traps full of liquid N_2 to maintain a clean system.

Diffusion pumps are capable of very efficient pumping from $\sim 10^{-1}$ to $\sim 10^{-9}$ Pa (10^{-11} Torr) and, if properly trapped, will provide a clean UHV system that is very reliable. The VG series UHV DSTEMs use only oil diffusion pumps to attain UHV conditions.

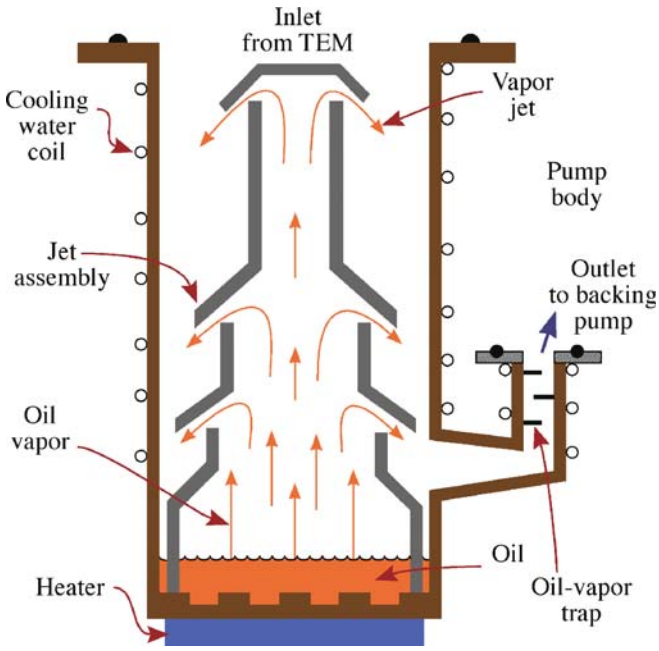


FIGURE 8.2. Principles of diffusion pump operation. A heater plate at the base of the pump boils synthetic oil. The expansion of the oil vapor on boiling creates a pressure which forces the vapor up the central column and out of several holes . The stream of oil vapor pulls gas molecules out of the top of the pump down to the base where the oil condenses and the air is pumped out of the base by a mechanical backing pump.

8.3.B Turbomolecular Pumps

Turbomolecular pumps, or turbopumps, as the name implies, use a turbine to force gases from the microscope. They have many parts moving at high speeds (in excess of 20,000–50,000 rpm is common), so they are more liable to fail than diffusion pumps. The mechanics of the pump are very simple as you can appreciate from Figure 8.3. They do not use oil so they don't introduce hydrocarbons to contaminate the microscope, and the best models (unlike earlier versions) are very quiet and almost vibration-free. In fact, modern turbopumps are being used to prepump the specimen chamber when this is critical, as in the cryotransfer technique (see Section 8.10). If you buy a turbopump, make sure to specify that its use will not transmit vibrations to the TEM column, where it would destroy the image resolution. The turbopump can start (slowly) at

FIGURE 8.3. A turbopump (with and without its casing), which is nothing more than a small turbine that rotates at high speed. Like a jet turbine it pulls air in at the front end and forces it out of the back. The blades are designed like airfoils to enhance the flow of gas through the system.

ambient pressures, increasing speed as the pressure is lowered, ultimately providing UHV conditions at high enough speeds. It is usual, however, to back the turbopump with a dry mechanical pump.

EXHAUST PUMPS

Mechanical, diffusion, and turbopumps are all exhaust pumps; they pull in air from one end and expel it from the other.

8.3.C Ion Pumps

Ion pumps do not contain oil, so they cannot contaminate the TEM column. They also have no moving parts, relying solely on the ionization process to remove air. The ion pump emits electrons from a cathode. These ions spiral in a magnetic field (see Section 6.3) and ionize air molecules, which are then attracted to the cathode. The energetic gas ions sputter Ti atoms from the cathode and they condense throughout the pump chamber, mainly on the cylindrical anode, trapping gas atoms. Thus ion pumps remove gas atoms in two ways; by chemisorption on the anode surfaces and by electrical attraction to the cathodes. The smaller the ion current between the electrodes, the lower the vacuum, so the pump acts as its own vacuum gauge. Ion pumps are only efficient at high vacuums, so they are usually switched on after a diffusion pump has lowered the pressure to $<\sim 10^{-3}$ Pa (10^{-5} Torr). It is common to add ion pumps directly to the stage or gun chambers of TEMs to focus their pumping action on these important regions. Since these pumps are very common on TEMs, we include a diagram (Figure 8.4) showing how they operate.

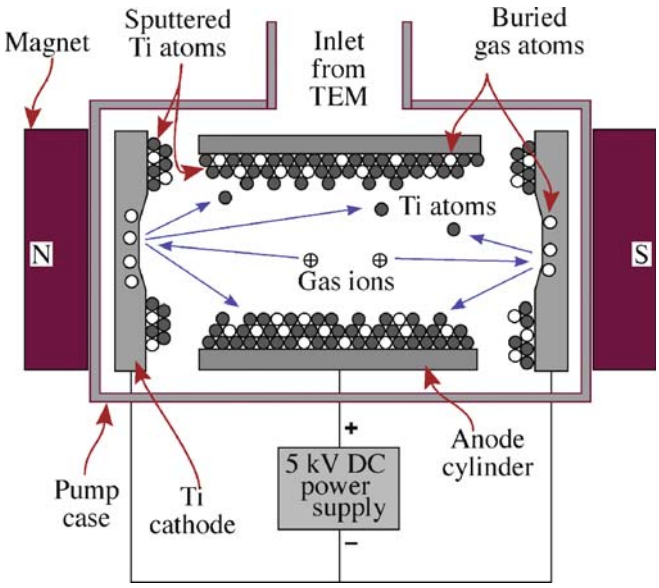


FIGURE 8.4. Schematic diagram showing how ion pumps trap ionized gas atoms by layers of Ti atoms at electrodes. Once trapped, the ions cannot escape until the pump is turned off.

8.3.D Cryogenic (Adsorption) Pumps

As the name implies, cryogenic pumps (cryopumps) rely on liquid N_2 to cool molecular sieves with large surface areas. The cold surface efficiently removes air molecules from ambient pressure down to $\sim 10^{-4}$ Pa (10^{-6} Torr). Because they are oil-free, cryopumps are also used to back ion pumps and prevent their accidental contamination through backstreaming from oil-bearing pumps.

We also use cold surfaces to enhance vacuums in the stage of most non-UHV TEMs. Such 'cold fingers' or 'anticontaminators' provide an alternative site (rather than your specimen) for condensation of residual components in the vacuum.

The same is true if the anticontaminator in your stage is allowed to warm up; then it will degrade the vacuum around your specimen. So you must use another pump such as a diffusion or mechanical pump to remove the air molecules as they are released from captivity. Otherwise, this outgassing will degrade the quality of the vacuum around your specimen, increasing contamination.

TRAPPING PUMPS

Ion pumps and cryopumps are trapping pumps. They keep the air molecules within them and release them when turned off or warmed up, respectively.

8.4 THE WHOLE SYSTEM

As shown schematically in Figure 8.5 the modern TEM has at least two separate pumping systems: one that evacuates the column and one that pumps the camera and screen chamber. We pump the camera separately because the film is one of the primary causes of vacuum degradation since outgassing occurs from the emulsion that contains the AgI grains. So this part of the TEM is usually pumped by a combination mechanical/diffusion pump. The stage is often pumped by a separate ion pump, turbopump, or cryopump, or some combination of these. If the instrument has an FEG, then there is a separate UHV pumping system for the gun region, which often consists of several ion pumps. Each part of the vacuum system consists of roughing pumps (mechanical or turbo) that pump out the appropriate part of the microscope to give the vacuum where the HV/UHV pumps can start to operate.

Looking at Figure 8.5, there are three valves, which are now all computer controlled.

- #1 connects the mechanical pump to the column (the roughing valve).

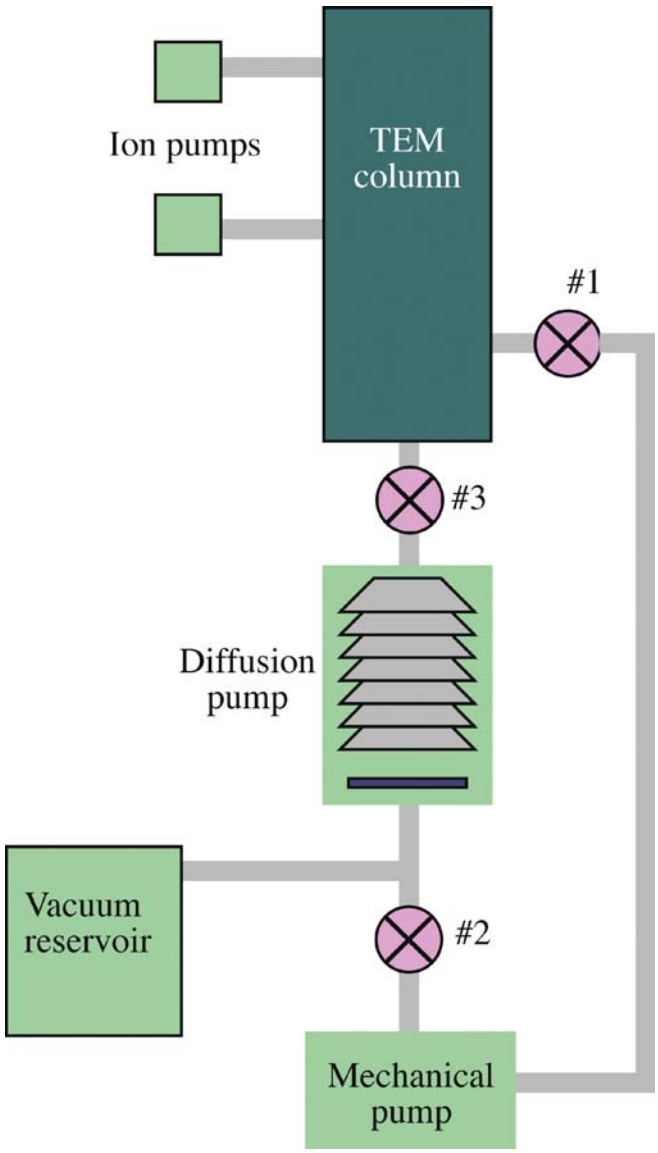


FIGURE 8.5. The principles of the TEM vacuum system. Often, the console display on the TEM will show a similar diagram. The mechanical pump can pump the column directly or back out the diffusion pump, which is connected directly to the base of the microscope. Ion pumps are often interfaced directly to the stage and gun areas. Computer-controlled valves separate the pumps from the column and from each other.

- #2 connects the mechanical pump to the bottom of the diffusion pump (the backing valve).
- #3 connects the diffusion pump directly to the TEM column (the butterfly valve).

If you're pumping down from atmospheric pressure, you first use the mechanical pump to back out the diffusion pump, till it gets to a low enough pressure so its heater can be safely switched on without oxidizing. So close #1, open #2, and close #3.

When the diffusion pump is warmed up, you rough out the column: open #1, close #2, and #3, until the column is at a low enough pressure that the diffusion pump can be used.

At this point, close #1, open #2, and then #3, so the diffusion pump is open to the TEM and may be continuously backed by the mechanical pump. The better approach incorporates a vacuum reservoir between the mechanical and diffusion pumps. When the reservoir is pumped to < 0.1 Pa, the mechanical pump is closed off and the diffusion pump exhausts into the reservoir. When the pressure builds in the reservoir, the mechanical pump will automatically switch on and lower the pressure.

Similar arrangements work for other pumps; e.g., a diffusion pump may be used to lower the pressure in the stage and gun sufficiently for the ion pumps to be switched on, and so on. In most TEMs the stage and gun have significantly better vacuums than the camera region, so the camera/screen is isolated from the rest of the column by a differential pumping aperture (not shown in Figure 8.5). This aperture often coincides with the BFP of the projector lens, since all the electrons have to pass through it and the DP in the BFP localizes all the electron trajectories close to the optic axis. A similar arrangement exists between the stage and gun in FEG systems to preserve the tip in case of a vacuum leak in the stage.

The advent of high-quality digital recording which will remove the need for film in the camera will do more to improve the quality of vacuums in TEMs than any advances in pumping technology.

8.5 LEAK DETECTION

"Nature abhors a vacuum," as Franois Rabelais said in 1534. That's the reason why the pumps must keep pumping: the TEM leaks. But some leaks are too large for the pumps to handle, and then the instrument performance degrades. If you can't run the electron gun, your TEM is useless. Under these circumstances, you have to find the leak, cure it, and repump the instrument, (this is usually a job for your service engineer) but some labs or older TEMs don't have service contracts. Leak detection involves using a mass spectrometer, which can be put into the pumping lines of the microscope. You then release helium gas close to the various parts of the TEM where you suspect a leak (the stage airlock, which sees a lot of use, is a common point of failure). The small He atoms are relatively easily sucked into the column through any leak and register on the mass spectrometer. When a leak is isolated, the TEM may have to be opened to the atmosphere to permit replacement of the defective part, such as the O-ring seals.

The most common cause of a leak is your specimen holder. The O-ring seal on the shaft of a side-entry holder (see the second half of this chapter) is easily contaminated with dust or a hair since it is continually inserted and extracted from the column, and left on the bench while you pore over it. Never touch the O-ring, make sure it doesn't dry out, but if it does (because you don't have someone else looking after the TEM), lubricate it with a very thin film of vacuum grease.

After repairing a leak, when you've pumped down again, it is often useful to 'bake' the column. Baking means heating the internal surfaces to >100°C (or >150–200°C in UHV TEMs) to boil off residual water vapor and hydrocarbons that may have entered the system when it was down to air. Usually, you can achieve the bakeout by leaving the lenses running without their cooling water (check this *very* carefully with the manufacturer before proceeding). In some cases, special heating panels are constructed around the column. Baking can also introduce other leaks as the whole system expands and then contracts, so sometimes leak detection and cure is an iterative process. For UHV systems, you *must* bake to reach the ultimate vacuum, and the higher the temperature the better.

Be wary, however, since sometimes the TEM accessories, such as XEDS and EELS systems, are not designed to be baked to the same high temperature as the column.

8.6 CONTAMINATION: HYDROCARBONS AND WATER VAPOR

As we said right at the start of the chapter, the vacuum (or rather what is still present in the column) can be a source of contamination. Residual hydrocarbons from the pump oil crack under the electron beam. Carbonaceous material then deposits on your carefully thinned specimen, making it difficult to do sensible high-resolution imaging or microanalysis. So a clean vacuum (one in which the hydrocarbon partial pressure is $< 10^{-9}$ Pa) is essential. Fortunately, most modern TEMs are relatively contamination-free, particularly if you use synthetic oils and appropriate traps on the pumps. (See also Section 8.12.)

However, even if you've paid dearly for a clean vacuum system, contamination often occurs; it comes primarily through the airlock with your specimen. You can minimize this by heating the specimen to >100°C in a heating holder or with a halogen lamp in the prepump chamber, or cooling the specimen to liquid-N_2 temperatures in a cooling holder. It may help if the prepump chamber is pumped with an oil-free pump. More recently, plasma cleaning of the specimen holder and specimen prior to insertion in the TEM has proven a very successful way to ensure a clean specimen (more on this in Section 8.12).

Polymers and biological specimens can easily introduce hydrocarbon contaminants, as they outgas in the vacuum, so it is sensible to cool such specimens (since heating or plasma cleaning destroys them). However, when you cool your specimen, it attracts water vapor which condenses as ice on the surface; so load your specimen first, then cool it down in the TEM before you switch on the beam. A low partial pressure of H_2O in the vacuum is obviously essential. Also, warm up any cooled specimens in the TEM before bringing them out to ambient atmosphere, otherwise they will immediately ice up (unless it's a *very* dry winter's day). There will be more about this in the sections on specimen holders.

In addition to the specimen, you personally can be a major source of contamination. Take care never to touch anything that will enter the vacuum, i.e., the specimen itself, the grids, specimen holder (beyond the O-ring seal on the rod), clamping rings, replacement diaphragms, new filaments, replacement Wehnelt, components of XEDS and EELS systems, etc. Use latex gloves whenever you load a specimen, and don't breathe on it. Store specimen holders and specimens in a dry box containing a desiccant such as silica gel, which should be replaced regularly. *Always* prepump fresh film in a vacuum desiccator (which is sometimes integrated into the TEM itself but probably best done elsewhere). Better still, never use film if you can avoid it. Simple precautions like this will minimize contamination of your specimens and the microscope in general and bring a much greater return in terms of good data per TEM session.

8.7 SPECIMEN HOLDERS AND STAGES

To look at your specimen, place it in a specimen holder and insert this assembly into the TEM stage. Therefore, there are two key components which are often not separated, namely, the holder and the stage. In this part of the chapter, we will emphasize the holder but the stage is also critical. Suitable design of the stage is the essential precursor to computer-controlled, remotely accessed TEM, which is already happening.

The cold trap, cold finger, or cryo-blades are a critical part of the stage. Ideally, this cold finger will completely surround the specimen: it cryo-pumps the region around the specimen. However, the cold surfaces, usually brass, provide a source of stray electrons and X-rays which is undesirable for AEM (see Chapter 33), so these blades should be removable for AEM.

X-ray diffractometers use goniometers to hold and tilt the specimen; so do TEMs. Conventional SEMs use a stub on which you mount the specimen so that you can bring the specimen close to the objective lens. However, some high-resolution SEMs use a specimen holder which

is very similar to those used in the TEM, because the specimen is inserted inside the lens, rather than underneath and outside it.

The reason the specimen holder is so important in TEM is that your specimen must invariably be located within the objective lens and the aberrations associated with the objective lens determine the resolution of the TEM.

Historically, microscopists have used two different designs and a lot of what you'll read here or elsewhere has a strong historical background.

- The traditional side-entry holder is a rod with a motor attached to tilt and/or rotate the specimen and a lead connecting it to a power supply and control box, or liquid-N_2 dewar.
- The traditional top-entry holder is a cartridge which you load into the TEM but is detached from the outside world when you use the microscope.

The actual cup that holds your specimen is either 2.3 or 3.05 mm in diameter, so the specimen disk or support grid has to be the same dimension, as we'll see in Chapter 10. The reasons for these dimensions are again partly historical. In the top-entry holder the specimen and part of the holder fit through the bore of the upper polepiece (see Figures 6.7 and 6.8). Clearly, the specimen must be smaller than the bore diameter. So the original top-entry holders used small specimens.

THE WORD 'TRADITIONAL'

Holders are changing; new manufacturers and new capabilities are emerging. Ideally, the holder should not be moved after it is inserted in the TEM.

Side-entry holders are more versatile and larger specimen dimensions first appeared when they were introduced. However, side-entry holders connect the specimen directly to the outside world via a long lever arm, which is undesirable, unstable, and also not necessary in many cases! Ideally, the side-entry holder should leave the specimen in the stage, not connected to the outside world, and all manipulations should be conducted through the stage itself, not the holder. This ideal is being approached as stages become more computer-controlled.

8.8 SIDE-ENTRY HOLDERS

Side-entry holders are now the standard, although their design has changed quite radically in recent years. The traditional design is shown in Figure 8.6. The key parts of this holder are:

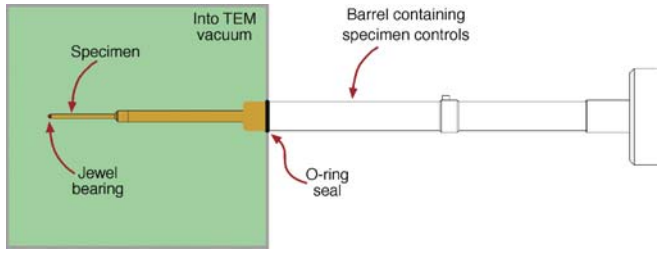


FIGURE 8.6. Principal parts of a side-entry holder that is held in the goniometer stage. The specimen is clamped into the cup at the end of the rod. A small jewel at the end of the rod (usually sapphire) fits into another jewel bearing in the stage to provide a stable base for manipulating the specimen. The O-ring seals the end of the holder inside the vacuum. Manipulating the specimen is accomplished from outside the column via controls within the rod. (See Figures 8.8–8.11 to see just where the specimen goes.)

- *The O-ring,* which is one mechanical link to the microscope column. Some holders have two O-rings and the gap between the O-rings is pumped separately to improve the vacuum.
- *The jewel bearing,* which is the other mechanical link to the microscope column. You push on this bearing to move your specimen back and forth and from side to side. Like the O-ring, you must keep the bearing clean otherwise the specimen will not be stable.
- *The cup,* which actually holds your specimen and thus provides the immediate environment which is seen by stray electrons and any X-rays coming down the column. So cups in holders for AEM are made of Be to minimize the generation of X-rays that would interfere with microanalysis.
- *The clamping ring or screw,* which holds the specimen in the cup. This ring (not shown in the figure), which may also be Be, must be carefully designed. It must hold your specimen firmly (so, e.g., magnetic disks cannot be pulled out of the cup by the lens field). However, the ring must not be so difficult to tighten that you put undue pressure on your specimen. Brittle disks may break as you are loading them. There are two kinds of retaining rings: screw-thread rings, which are easier to control and do not damage metals, but you'll find they may break ceramics because they transfer shear stresses to the disk; spring clips are difficult for the novice to master, but with practice you'll find they offer more control over the load that you put on the specimen, so we recommend them for the experienced ceramist. Unfortunately, no one we know makes Be spring clips!

In a more modern design, the jewel bearing is omitted so that the holder is supported at just one pivot point.

8.9 TOP-ENTRY HOLDERS

Top-entry holders are becoming less common because they essentially preclude XEDS analysis in the TEM. Also, it is more difficult to design such holders so that the specimen can be manipulated (e.g., rotated or strained). Their great advantage was that they were much less susceptible to drift since they were not connected directly to the outside, so early HRTEM required top-entry holders. Today, however, nearly all TEMs up to 400 kV use side-entry holders; only the DSTEMs retain the top-entry (bottom entry?) design.

Another drawback of such holders is that the bore of the objective lens must be asymmetric (using an upper and a lower polepiece), which actually limits the ultimate resolution by constraining the lens designer. Figure 8.7 shows a schematic diagram of such a holder.

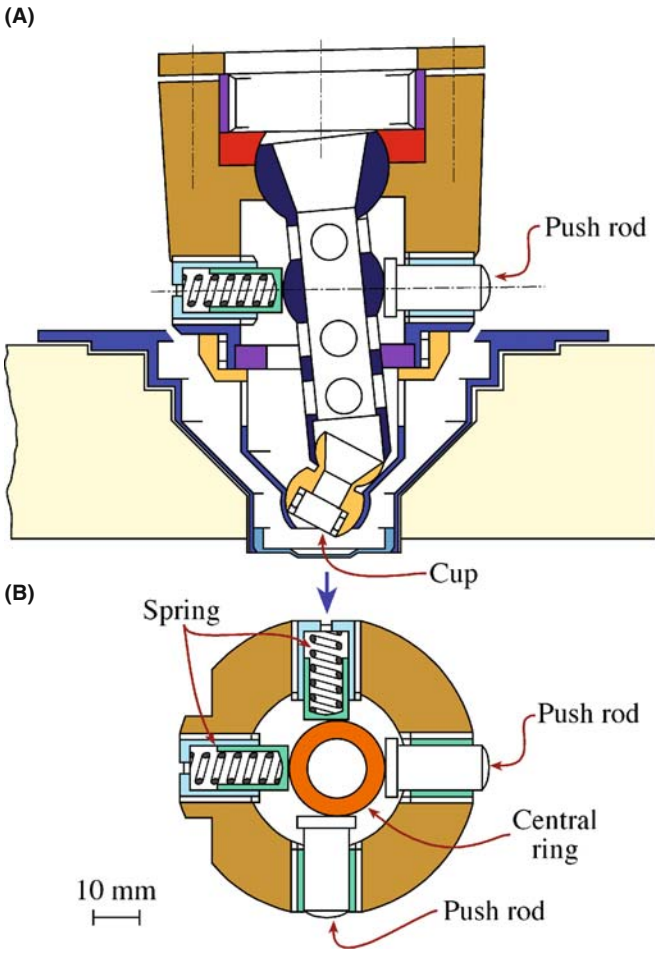


FIGURE 8.7. Top-entry holder: (A) cross section; (B) top view. The cartridge has a cone shape which fits into the tapered bore of the objective lens polepiece. The specimen sits in a cup at the base of a column through the cone down which the incident beam travels. Simple manipulations such as tilting or rotating require complex micromechanical design, since the specimen is at the base of the cartridge and completely surrounded by the polepiece. To tilt, e.g., as shown in (A), push rods are pressed against springs in two orthogonal directions, displacing a central ring around the column (see B), thus tilting the specimen cup.

8.10 TILT AND ROTATE HOLDERS

One feature of TEMs which may surprise you if you are a new user is that a wide variety of holders is available for the TEM. Figure 8.8 shows illustrations of these different designs for the side-entry holder:

- *Single-tilt holder:* This is the basic holder with which any novice should start practicing. You can only tilt around the axis of the rod. It is relatively cheap, robust, and can at least give you some idea of the usefulness of tilting a specimen for diffraction-contrast studies.
- *Quick change holder:* This is also a single-tilt holder that clamps the specimen with a lever arm which you raise and lower onto your disk or grid. It doesn't put a high stress on the specimen, but it doesn't hold it very strongly either. Don't use it for magnetic specimens, but it can be great for ceramics. Different retainers can be substituted for the clamp as shown in Figure 8.8 (bottom), creating a more versatile multipurpose holder
- *Multiple-specimen holder:* This is usually a single-tilt holder, but you can load up to five specimens into the column at one time as shown in Figure 8.9A. A two-specimen, double-tilt version is also available (Figure 8.9B). Such holders can be useful if you are not very good at specimen preparation, or you want to compare different specimens under identical conditions without turning off the beam. However, in

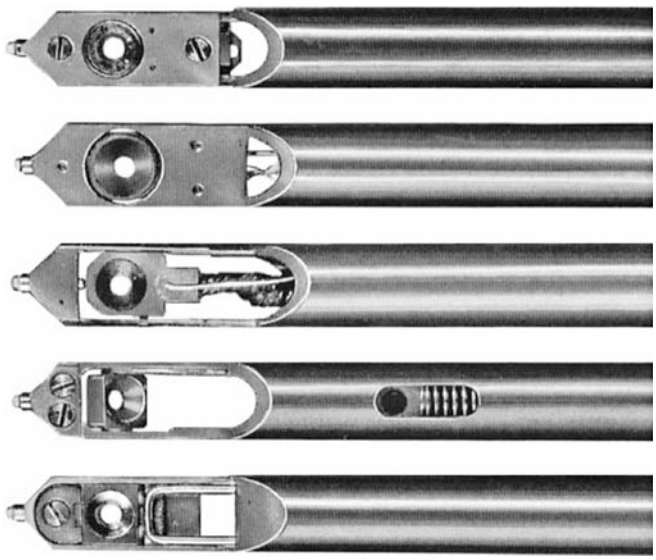


FIGURE 8.8. Examples of different designs for the side-entry holder. From the top, they are: a rotation holder, a heating holder, a cooling holder, a double-tilt holder, and a single-tilt holder.

FIGURE 8.9. Multiple-specimen holders: (A) two-specimen double-tilt and (B) five-specimen single-tilt.

modern TEMs, specimen exchange is relatively quick, except in UHV instruments where the multi-holder would probably be more useful although it is less common.

- *Bulk specimen holder:* This holder is used for surface imaging and diffraction, e.g., using SE or BSE in a STEM or for reflection diffraction and imaging in a TEM (see Chapter 29 for more about these techniques). The bulk specimen is larger than the traditional 3-mm disk (usually ~10 mm × 5 mm) so if you can create a thin specimen of these dimensions, the bulk holder will allow you to sample more of your material at one time (Figure 8.10).

  So don't always think that you are limited to 3-mm specimens!

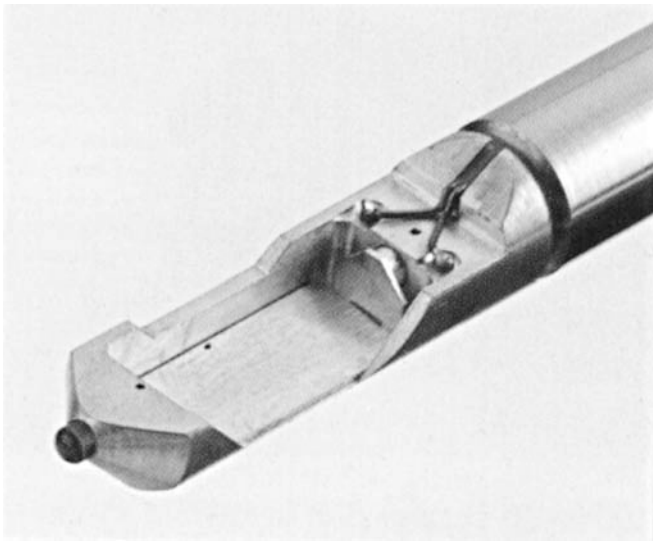


FIGURE 8.10. A bulk holder for large specimens.

- *Double-tilt holder:* This is the most popular holder since it gives you the most flexibility in orienting the specimen. It is absolutely essential for imaging and diffraction studies of crystalline specimens. The tilt axes are fixed as two orthogonal directions. In some designs, you can remove the cup while the specimen is in place which means that you can reinsert your specimen in the same orientation. This feature is extremely useful if your specimen is robust.
- *Tilt-rotate holder:* You would often like to be able to orient your specimen parallel to the tilt axis (along the rod). This holder lets you do just that. It's a strength for the side-entry holder: one tilt axis is always parallel to the rod of the holder which also gives the largest tilt angle.
- *Low-background holder:* The cup and clamping ring are made of Be to minimize the generation of bremsstrahlung X-rays and characteristic X-rays. So they are required for XEDS studies. They can be double or single tilt and may be cooled also.
- *Tomography holder:* This is a new design that allows you to tilt the sample through a full 360°. It's ideal for looking at needles (like an AFM or atom-probe tip).

8.11 IN-SITU HOLDERS

Special holders have been developed that allow you to change your specimen while you observe it in the TEM; in other words you can do experiments (heat, cool, strain, twist, compress, etc.) on a specimen in the TEM.

- *Heating holder:* Such holders in a conventional TEM can go to ~1300°C which is measured by a thermocouple attached to the cup. In HVEMs, the temperature can go higher because of the larger gap between the polepieces. You have to be careful to calibrate the temperature and remember that the temperature may be different for different specimens. You should also be sure that the material you are studying does not form a eutectic alloy with the material forming the holder! If the eutectic does form it will have a lower melting point, so you may deposit part of your specimen and the holder on the objective lens, or down onto the screen, if the microscope is well aligned.
- *Cooling holder:* This is available for either liquid-N_2 or liquid-He temperatures. These holders, which can be single or double tilt, are a great asset for XEDS, EELS, and CBED studies since they minimize surface-borne contamination. They are also essential for in-situ studies of superconducting materials and ideal for polymers or biological tissue. However, you should remember that the cold holder can also act as a small cryo-pump so that it actually attracts contamination.

Since you are necessarily changing the temperature at the specimen relative to its surroundings, be prepared for specimen drift. It takes time for the whole system to stabilize. The Polara from FEI is a modification of the cooling holder which fits in a special dedicated stage; the tip of the holder detaches from the rod with obvious advantages.

- *Cryo-transfer holder:* Certain specimens are prepared at cryogenic temperatures such as liquids, latex emulsions, and tissue in general. This holder permits you to transfer such cold specimens into the TEM without water vapor from the atmosphere condensing as ice on the surface.
- *Straining holder:* This holder clamps the specimen at both ends then applies a load to one end, via a load cell or screw-thread mechanism, as shown in Figure 8.11. The sample can be in the shape of a small tensile specimen and it is thinned in the middle of the gauge length (see inset). The motion of dislocations, cracks, etc., are then easily monitored, so a video camera is an essential accessory. You can vary the load, to study cyclic as well as tensile loading, and the strain rate is another variable that is easily controlled. In Figure 8.11 a furnace is present, so the specimen can be heated while under load. The use of piezoelectric drives is leading to great improvements in this type of holder.
- *Probing holders:* These are like AFM holders for the TEM. You can use them to 'poke' your specimens—just like an AFM, STM, or an indenter—while you observe the effect in the TEM.
- *EBIC and CL holders:* The essential feature is the electrical feed-through that allows you to control the charge recombination in a semiconductor or certain mineral specimens by applying a bias across the specimen surface.

Beware: heating and straining holders, in particular, can produce effects in thin foils that are totally uncharacteristic of your bulk sample. So you must use these holders carefully and interpret your results cautiously. Often, surface reactions will dominate internal reactions when you are trying to induce a phase transformation by heating. The surface may also stop grain boundaries from migrating at temperatures where they would do so in the bulk material. Obviously, defect motion under applied stress may also be strongly affected since the 3D stress field will be very different in bulk specimens compared to thin foils.

These problems can be overcome to some extent if you use thicker specimens and examine them in an HVEM, or at least an IVEM, and the whole field of in situ studies, particularly, heating and/or straining, is best performed in such microscopes (Butler and Hale 1981 and Section 29.12). However, the high-energy electrons in these microscopes may introduce lattice defects that affect the very phenomenon that you want to study, e.g., beam-induced vacancies can change diffusional phase-transformation kinetics very easily.

It is also possible, but much more difficult and expensive, to manipulate specimens in top-entry stages. The top-entry holder shown in Figure 8.12 is a heating-

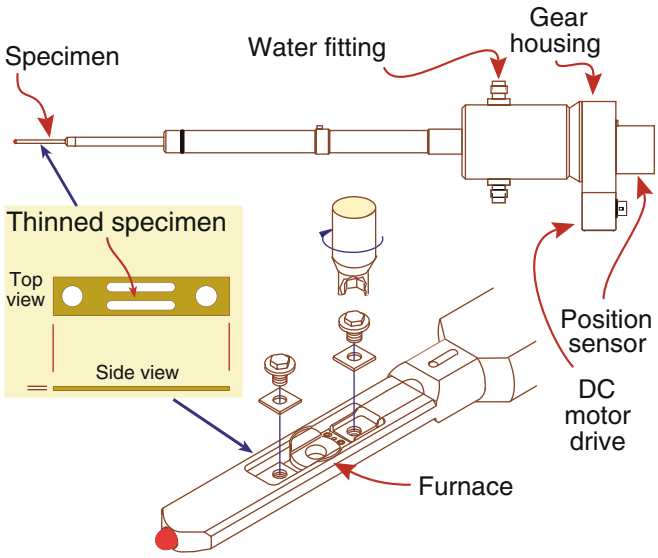


FIGURE 8.11. A side-entry combined straining and heating holder. The specimen looks like a miniature tensile specimen (inset) and is clamped at either end by hex screws. There is a screw-thread arrangement for pulling the specimen contained within the rod. The furnace surrounds the central thin portion of the specimen.

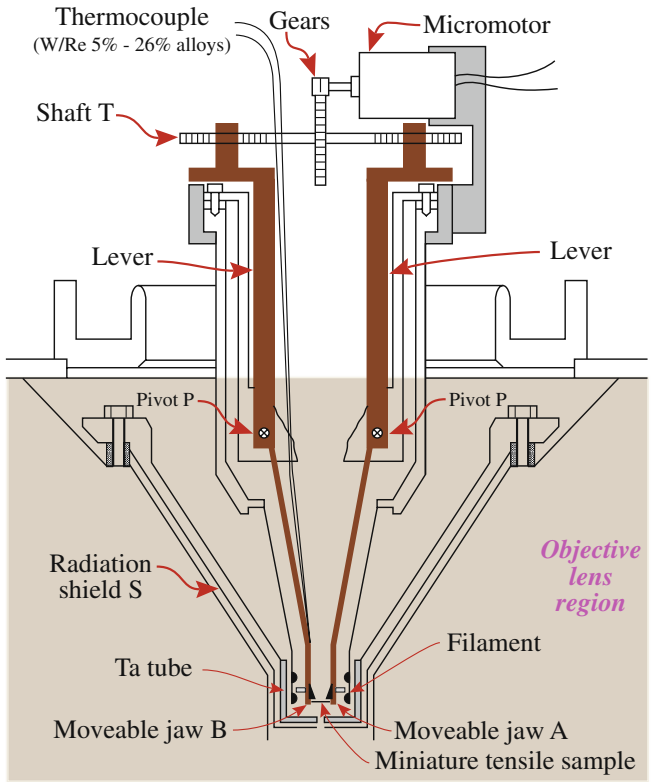


FIGURE 8.12. A top-entry, heating-straining holder which can be used at temperatures up to 2300 K in a 3-MV HVEM.

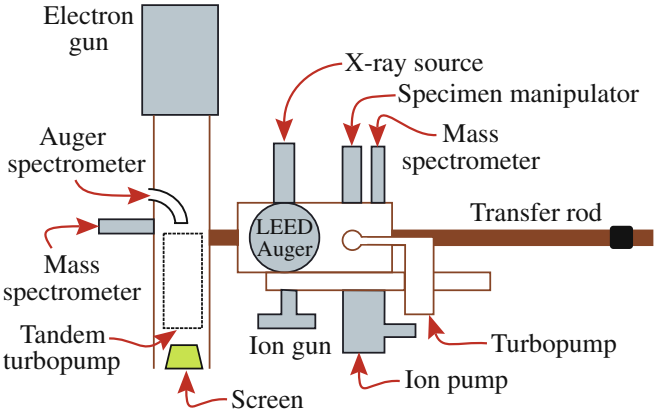


FIGURE 8.13. Schematic diagram showing the Hitachi H900 UHV TEM. This instrument is equipped with a prechamber with LEED, Auger, and an ion gun which can be used to clean the specimen, allowing UHV surface analysis to be carried out on the TEM specimen. The holder has to transfer the specimen through a prepump chamber where it is ion-cleaned before going into the column.

straining holder, which is capable of operating at temperatures up to 2300 K. The heat is provided by a coaxial Ta tube that supports the W heater filament as shown in the figure. The holder is used in a 3-MV microscope where the specimen diameter is 5 mm. The larger specimen diameter means that the disk can be shaped as a small tensile specimen and still be quite robust.

There are also special combinations of holders and stages which have been optimized for particular applications. The example shown in Figure 8.13 has been optimized to combine surface studies using low-energy electron diffraction (LEED) and Auger analysis with TEM. The prechamber is fitted with an ion gun to clean the sample before the surface is analyzed. The specimen can then be moved into the TEM column for transmission studies. A similar prechamber has been

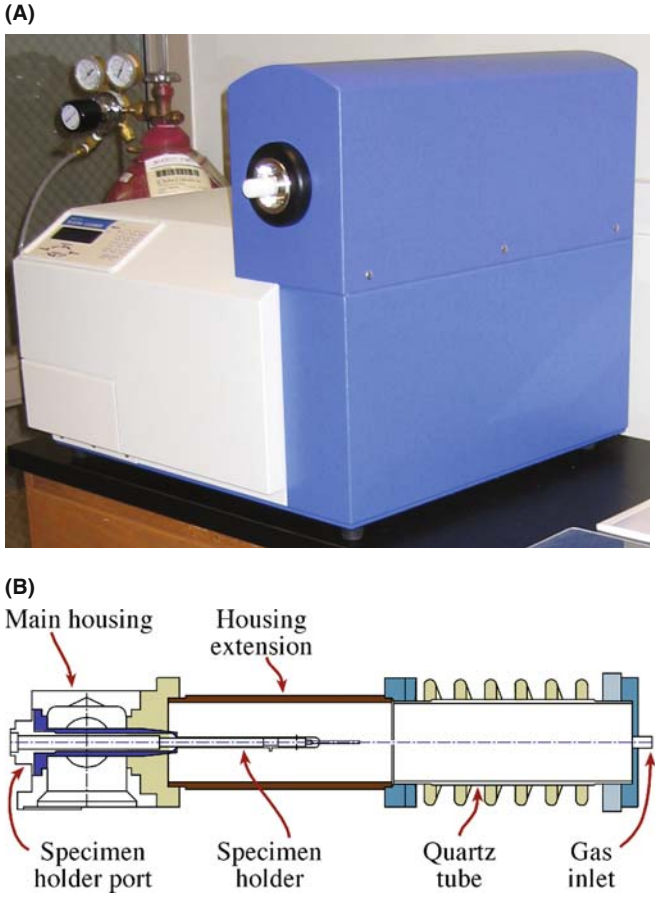


FIGURE 8.14. Plasma cleaner: (A) example and (B) schematic.

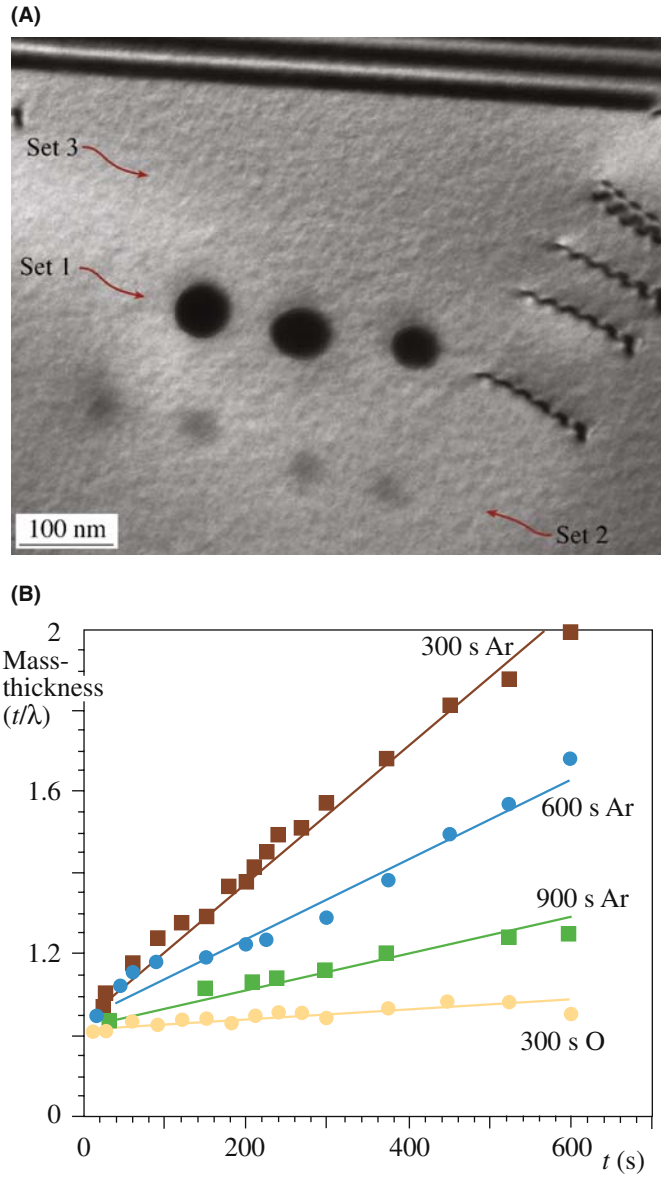


FIGURE 8.15. Cleaning the surface of the specimen using a plasma cleaner reduces the contamination produced by a focused electron beam. (A) Set 1 was produced before the specimen was cleaned. Sets 2 and 3 (which you can't see) were produced after 5 minutes of Ar cleaning and then an additional 5 minutes of oxygen cleaning, respectively. (B) The rate at which the specimen contamination builds up; the additional cleaning in pure oxygen always reduced the contamination rate.

used elsewhere to provide a method to clean the sample before growing thin films on the sample by molecular-beam epitaxy (MBE) or thermal evaporation.

One of the reasons for using higher accelerating voltages is that this gives more room in the specimen-stage region. Thus, even 400-kV microscopes can be fitted with a small, differentially pumped environmental chamber. Such a chamber allows in situ studies of corrosion, degradation of catalysts, etc., especially when combined with a heating holder.

Development of new holders and their use for in-situ studies is going to be one of the most exciting topics in TEM. We'll expand on this topic in the companion text.

8.12 PLASMA CLEANERS

Plasma cleaners have been available for removing surface contamination and modifying surfaces (for example, changing wettability) for over 30 years. The unit is now marketed as a small box (e.g., as shown in Figure 8.14A) that can usually accommodate one specimen holder; the rod of the holder is placed inside a plasma chamber, as shown in Figure 8.14B, just before it goes into the TEM. Plasma cleaners have long been used to modify/clean surfaces of glass, semiconductors and other ceramics, metals, and even polymers and biomaterials. The development of plasma cleaners for cleaning both the TEM holder (the part that goes into the vacuum of the TEM) and more specifically the specimen that has already been placed in the TEM holder, is more recent, but the possibilities are clear. This cleaning is generally regarded as essential for small-probe AEM. Most of the contamination that occurs in the electron beam originates on your specimen, not from a 'dirty' vacuum.

Although the processes occurring in the plasma are complex (like the plasma itself) the basic idea is illustrated by the removal of hydrocarbon contamination on the specimen shown in Figure 8.15A. The plasma consists of a mixture of energetic electrons and ions that bombard the surface and break the C–H bonds. With short duration exposure, the surface of the specimen itself is essentially unaffected. The hydrocarbon is thus gradually reduced in molecular weight and pumped away in the vacuum of the cleaner. The user has some flexibility in the choice of the plasma gas although the manufacturer usually limits this. Although many gasses can be used in principle, O_2, N_2, and Ar are the most common. The graph in Figure 8.15B shows the effect of using an oxygen plasma—it reduces the rate at which the contamination grows under the electron probe; the effect of the oxygen-reactive gas is similar to that found when using iodine in the ion mill.

CHAPTER SUMMARY

The vacuum and the holder are the two parts of the TEM that most closely affect your specimen. You have to treat both carefully if you want to be sure of getting the most out of your TEM. The vacuum is usually automated, so you don't have too much control over it. However, you can degrade the vacuum easily if you are a careless operator; e.g., if you don't bother to prepump your film. You don't want to touch any part (including the specimen holder) that will go into the vacuum. In fact, you should treat the specimen holder as if it were a rare jewel; it may actually contain a couple of synthetic ones and it certainly costs as much as a diamond of several carats! (It's also probably worth more than its weight in gold.)

With the range of holders available today, you can conduct many materials science experiments on your thin specimen while observing it in the TEM. However, if you're looking at crystalline material, the most common manipulation is still tilting the specimen in two orthogonal directions to orient different crystal planes parallel to the electron beam. You'll understand more fully why this is important after you've finished reading Parts 2 and 3 (and even in Part 4).

REFERENCES

There is a great deal of history for you to explore when you have time. Why are specimens 2.3 or 3.05 mm diameter? Why do we talk about loading the plates? In many cases, this history holds us back but we often don't recognize why!

VACUUM

A more detailed list of references is given in the chapter on in situ TEM in the companion text. For a full exposition of vacuum technology for TEMs, read Bigelow or the equally informative user's guide by O'Hanlon.

Bigelow, WC 1995 *Vacuum Methods in Electron Microscopy* in *Practical Methods in Electron Microscopy* **15** Ed. AM Glauert Portland Press London. An essential reference.
O'Hanlon, JF 1981 *A User's Guide to Vacuum Technology* John Wiley and Sons New York.

HOLDERS AND IN SITU

Butler, EP and Hale, KF 1981 *Dynamic Experiments in the Electron Microscope* in *Practical Methods in Electron Microscopy* **9** Ed AM Glauert Elsevier Amsterdam.
Watt, IM 1985 *The Principles and Practice of Electron Microscopy* Cambridge University Press New York. See appendix 1.
Komatsu, M, Mori, H and Iwasaki, K 1994 *Design of a Hot Tensile Stage for an Ultrahigh-voltage Electron Microscope and Its Application to In Situ Deformation of Sapphire at 1620 and 1720* K J. Am. Ceram. Soc. **77** 839–842. Illustrates the use of temperatures as high as 2300 K.
Valdrè, U and Goringe, MJ 1971 in *Electron Microscopy in Material Science*, 208–254 Ed. U Valdrè Academic Press New York. This article gives a detailed description of several TEM holders.

THE COMPANION TEXT

A full chapter on holders is included in the companion text with particular attention being paid to holders for in-situ experiments.

SELF-ASSESSMENT QUESTIONS

Q8.1 How does a diffusion pump work?
Q8.2 How does a mechanical pump work?
Q8.3 How does a turbomolecular pump work?
Q8.4 How does an ion pump work?
Q8.5 How does a cryogenic pump work?
Q8.6 If you are concerned about contaminants in the TEM from pumps, which pump(s) should you use (consider where and why)?
Q8.7 Where does the most common vacuum leak occur in a TEM and why at this point?
Q8.8 Name one way to localize the source of a leak.
Q8.9 Which types of specimens introduce contaminants and how can such introduction be prevented?
Q8.10 How can you personally introduce contamination into the microscope and what should you do to stop doing it?
Q8.11 Summarize the advantages and disadvantages of side-entry holders.
Q8.12 Why are TEM top-entry holders out of fashion?
Q8.13 What is a high vacuum (give a number with SI units)?
Q8.14 What is an ultrahigh vacuum and why is it worth paying a lot of money to buy a TEM with such a vacuum?
Q8.15 Which type of holder is the most flexible for TEM?
Q8.16 Why is the column baked? Why should this be done carefully?
Q8.17 What size is the typical specimen cup in the holder?
Q8.18 Why don't we make larger-diameter thin samples?
Q8.19 What are the drawbacks when using a straining holder?
Q8.20 Why would you cool a specimen?
Q8.21 What do we sometimes deliberately insert into a TEM that creates even more contamination and vacuum degradation than the specimen?

TEXT-SPECIFIC QUESTIONS

T8.1 How would you notice that there is a vacuum leak in the TEM (without watching the gauges)?
T8.2 Why can baking both improve the vacuum and cause more leaks?
T8.3 Explain with a schematic diagram (like Figure 8.5) how the combination of low and high vacuum pumps is used in tandem to pump down the microscope after: (a) it has been let down to air (e.g., for repair); (b) after it has been degraded by the insertion of a specimen; (c) after it has been degraded by changing the film.
T8.4 Copy Figure 8.1 and cut out the circle attached to the rotating vane. By placing that circle and the vane in different positions within the pump chamber show how air is alternately pulled out of the microscope (into the inlet port) then expelled from the pump (through the outlet port).
T8.5 Estimate the number of atoms in a cubic millimeter of air at 10^{-3} and 10^{-6} bar and thus estimate the number of Ti atoms that need to be ionized to reduce the vacuum in that cubic millimeter by that amount in each case. State any assumptions and be brief.
T8.6 Compare and contrast top-entry and side-entry holders from the points of view of stability (mechanical, thermal, etc.) versatility, cost, etc.

T8.7 Under what circumstances might you prefer to use a holder with just one tilt axis but the ability to rotate the specimen?

T8.8 Why do we choose to use Be for analytical holders rather than a lighter metal, or a heavier one such as Mg? What are the potential problems with using Be?

T8.9 What are the problems with manufacturing a stage and holders for UHV operations rather than for the generally poorer vacuum in most TEMs?

T8.10 Why is it difficult to manufacture holders that operate at extremely high or very low temperatures that do not compromise the routine operation of the microscope?

9

The Instrument

CHAPTER PREVIEW

Over the preceding four chapters we've now introduced all the essential components of the TEM and it's time to see how the guns (Chapter 5), lenses (Chapter 6), detectors/screens (Chapter 7), and specimen holders (Chapter 8) are combined to form the instrument. Just as we do for the VLM, it's convenient to divide up the TEM into three components: the illumination system, the objective lens/stage, and the imaging system. The illumination system comprises the gun and the condenser lenses and its role is to take the electrons from the source and transfer them to your specimen. You can operate the illumination system in two principal modes: parallel beam and convergent beam. The first mode is used primarily for TEM imaging and selected-area diffraction (SAD), while the second is used mainly for scanning (STEM) imaging, analysis via X-ray and electron spectrometry, and convergent-beam electron diffraction (CBED).

The objective lens and the specimen holder/stage system is the heart of the TEM. This critical region usually extends over a distance of ~10 mm at the center of the TEM. Here is where all of the beam-specimen interactions take place and the two fundamental TEM operations occur, namely, the creation of the various images and DPs that are subsequently magnified for viewing and recording. Clearly the objective lens is the most important lens in a TEM because its quality determines the quality of all the information about the specimen that we seek. Because of the various defects inherent in magnetic lenses, our specimens must be placed close to the center of the objective lens and this restricts our ability to manipulate the specimen and gain access to the various signals generated. However, as we've mentioned many times already, we can now overcome many of the defects that limit lens performance by inserting complex aberration-correction systems into the column, but only the most expensive TEMs are so equipped.

The imaging system uses several lenses to magnify the image or the DP produced by the objective lens and to focus these on the viewing screen or computer display via a detector, CCD, or TV camera. We'll refer to the magnifying lenses as the intermediate and diffraction lenses and the final lens as the projector lens (since it projects the final image or DP onto the viewing screen or detector). All TEM operations involve observing the electrons on a screen of some form, with or without a specimen in place. In many current TEMs you will have a button for focus, another for magnification, and another for diffraction (or a slide on the computer screen). These three components, the illumination system, stage, and imaging system, are often called the 'column', for obvious reasons. Throughout this book, the electrons will invariably be shown as traveling 'down' the column because this is the construction of most (but not all) TEMs.

The purpose of this chapter is to go through the principal functions of the many lenses in the column and give you some feel for what is happening in the microscope when you press a button, rotate a knob or trackball, or drag your mouse. The better you understand the operation of a TEM, the more you can be sure that you are getting the most out of it.

We do NOT intend to teach you how to operate a specific TEM since individual manufacturers choose different ways to manipulate the many lenses, apertures, deflectors, etc., in the TEM column to achieve the same ends. What we will do is introduce specific *operational procedures* that we hope are generic enough such that you should, in concert with the operating manual of your own TEM, be able to carry out all the basic procedures.

9.1 THE ILLUMINATION SYSTEM

The illumination system takes the electrons from the gun and transfers them to the specimen giving either a broad beam or a focused beam (often called a 'probe' or 'spot') entering the specimen. We can think of these two cases as analogous to a floodlight or a spotlight. In Chapter 5 we described how the gun produces an image of the source (called a crossover). This crossover acts as the object for the first lens in the illumination system that consists of several condenser lenses (C1, etc.). We will discuss the two different ways to use the illumination system: formation of a parallel electron beam (although it is almost never truly parallel) or a convergent beam (which may be divergent).

9.1.A TEM Operation Using a Parallel Beam

In the *traditional* TEM mode, the first two condenser lenses (C1, C2) are adjusted to illuminate the specimen with a parallel beam of electrons, typically several micrometers across at reasonable magnifications (20,000×–100,000×). As shown in Figure 9.1, the C1 lens first forms an image of the gun crossover. In the case of a thermionic source, the original crossover may be several tens of micrometers across, and this crossover is demagnified by an order of magnitude or more. In the case of a FEG, the source size may be less than the desired illumination area on the specimen so it may be necessary to magnify the crossover (so the condenser lenses don't always condense!). The simplest way to produce a parallel beam is shown in Figure 9.1A in

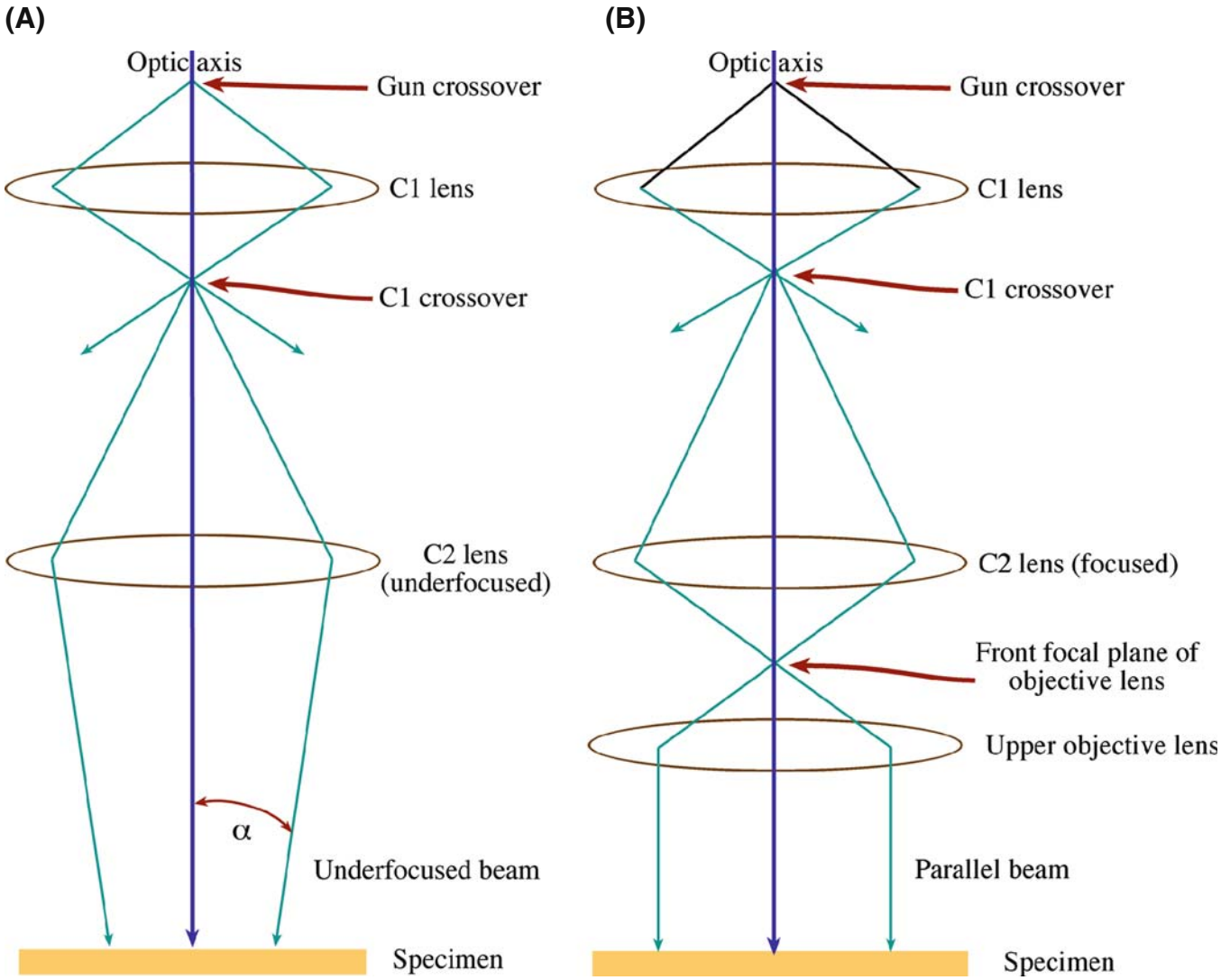


FIGURE 9.1. Parallel-beam operation in the TEM: (A) the basic principle illustrated, using just the C1 and an underfocused C2 lens. (B) The practical situation in most TEMs; using the C1 and C2 lenses to image the source at the FFP of the condenser-objective lens, thus creating a parallel beam at the specimen plane. Therefore, the upper objective is sometimes termed the C3 lens.

which the C2 lens is weakened to produce an underfocused image of the C1 crossover (although no TEM today has only two condenser lenses). Remember from Section 6.2.D and Figure 6.5 that any lens operates either in focus or out of focus and there are times when each mode is useful. In its out-of-focus condition, a lens is said to be overfocused if it is strong and the crossover occurs before the image plane and underfocused if the lens is weak and the crossover occurs after the image plane. Remember also that underfocus gives a more parallel beam at the specimen than overfocus.

CONVERGENCE ANGLES

α is small. All ray diagrams in this book are drawn with exaggerated (large) angles.

While the beam in Figure 9.1A is not exactly parallel to the optic axis, α under these conditions is $< 10^{-4}$ rads (0.0057°), which is effectively a parallel beam.

When generating the small probes we need in STEM and AEM, the upper polepiece of the objective lens acts as the C3 lens (see the next two sections) and controls the beam hitting the specimen as shown in Figure 9.1B. If the C2 lens is focused to produce an image (of the crossover) at the front-focal plane (FFP) of the upper-objective polepiece then a parallel beam of electrons is formed by the lens. The so-called condenser-objective (c/o) lens system is standard on almost all TEMs used for materials characterization but there is no c/o lens in dedicated STEMS, TEMs made before ~1980 and those designed primarily for imaging of biological specimens. The c/o lens system complicates the electron optics, as we shall see in Section 9.1.C below.

We'll see that parallel illumination is essential to get the sharpest selected-area DPs (SADPs) (Chapter 18) and, in principle, the best *classical* image contrast (Part 3). In fact, we usually assume that the incident beam is parallel when interpreting our *classical* images. Usually you should underfocus C2 until the illuminated area on your specimen fills the viewing screen. The higher the magnification the more you have to strengthen C2 (making the beam *less* parallel) in order to keep the illuminated area just filling the screen and thus illuminating less of the specimen.

CLASSICAL? PARALLEL?

Classical (or traditional or conventional) means... thinking in terms of waves and beams. Parallel illumination isn't really parallel, it's just not very convergent.

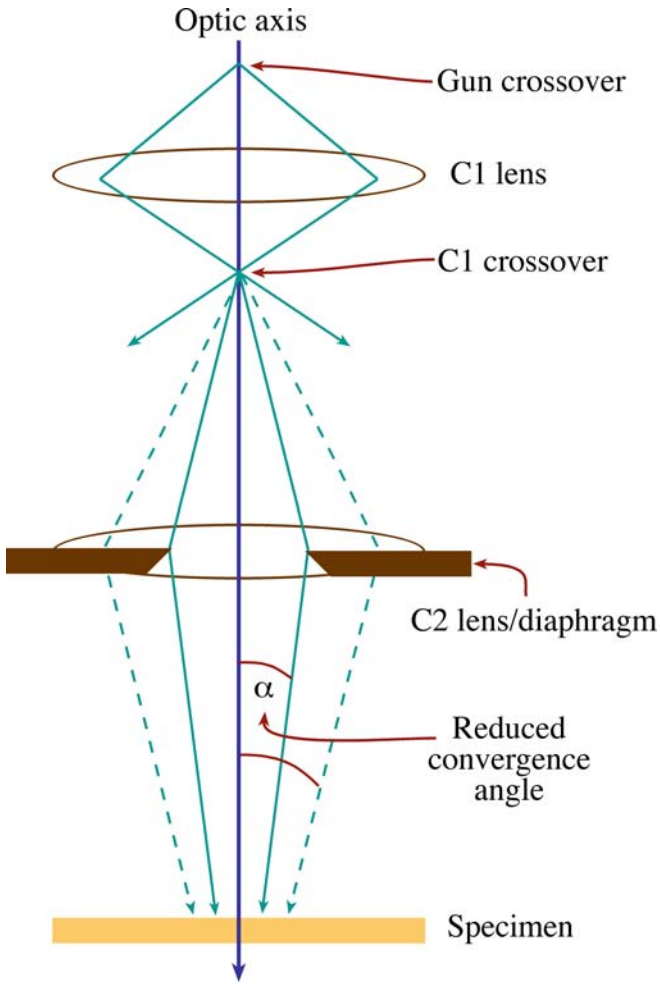


FIGURE 9.2. Effect of the C2 aperture on the parallel nature of the beam: a smaller aperture creates a more parallel beam at the expense of total number of electrons (i.e., reduced probe current) reaching the specimen.

In parallel-illumination TEM mode, such as you would use for diffraction-contrast imaging and SADP formation, there is usually no need to change C1, which is therefore kept at some intermediate setting, recommended by the manufacturer. The only other variable is the C2 aperture. A small aperture reduces the electron current reaching your specimen. However, if you use a smaller aperture, you decrease the angle of beam convergence and therefore make the beam more parallel, as is evident from Figure 9.2.

9.1.B Convergent-Beam (S)TEM Mode

Now, there are times when you may wish to focus the beam more, so that the intensity of the beam on a specific area of the specimen is increased. Let's look at various ways to do this.

If you want to *minimize* the area of the specimen that you are illuminating, you simply change the C2 lens so it is focused rather than defocused and you form an

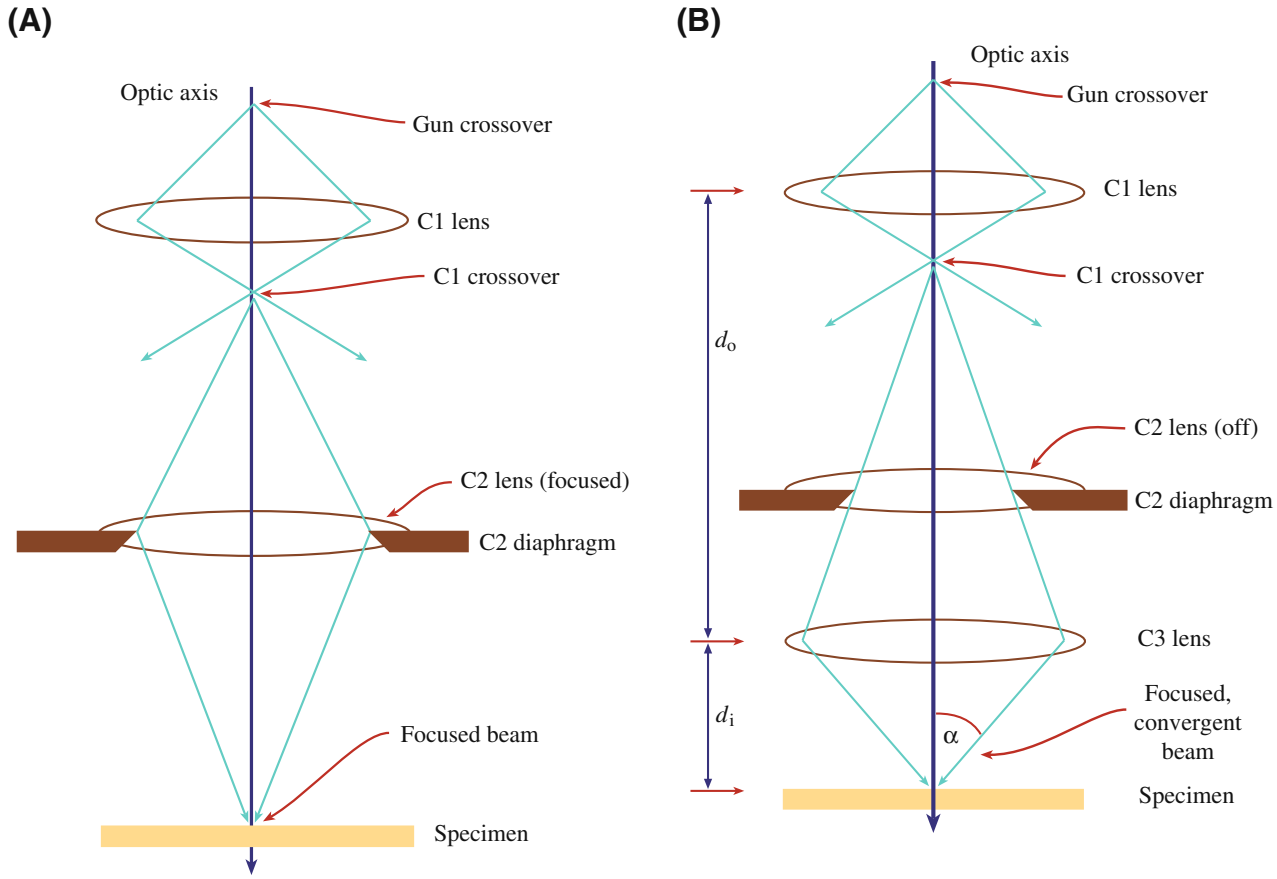


FIGURE 9.3. Convergent-beam/probe mode in the TEM. (A) The basic principle: a focused C2 lens illuminates a small area of the specimen with a non-parallel, convergent beam. (B) The practical situation in most TEMs: use of the upper-objective polepiece as the C3 lens gives the smallest possible probe and large convergence angles. Note that C2 is shown here as having no effect on the ray paths (i.e., it is effectively switched off). The large d_o/d_i ratio gives the maximum demagnification of the image of the gun crossover.

image of the C1 crossover at the specimen, as shown in Figure 9.3A. This is the condition under which you can view the source image to adjust its saturation (go back and look at Figure 5.5B and C) or to measure the dimensions of the beam (see Section 5.5.D). When C2 is focused like this, the beam is at its least parallel and most convergent. While the intensity of illumination on the viewing screen will be greatest, your image contrast will be reduced and any SADP will be distorted. Ideally, for routine TEM work, your specimen should always be thin enough so that you never have to operate with C2 focused but, in practice, you'll often find yourself focusing C2 to compensate somewhat for poor transmission through a thick portion of your specimen.

However, there are times when we do need to create a focused, convergent beam at the specimen. We then use the other principal way to operate the illumination system: the convergent-beam (or probe/spot) mode. When you use this mode you won't immediately see a useful image; the convergence destroys the parallelism and reduces the image contrast. So to see an image we have to scan the beam; this mode of operation of the illumination system is standard for STEM and AEM.

PROBE

The convergent beam is a probe. We use a probe to localize the signals coming from the specimen, as in XEDS, EELS, CBED, or HAADF.

Now, if you have a FEG it's possible to use the C1 and C2 lenses to produce Å-level probes but, with a thermionic source, it isn't possible to use just those two lenses (as in Figure 9.3A) to demagnify the relatively large thermionic-gun crossover to less than several nanometers. So to get the desired probe sizes of $\ll$ 1 nm for analysis, etc., the usual solution, as we just noted, is to introduce a C3 or c/o lens. We can only do this if the objective lens is split into two polepieces with separate coils and then we can make the upper polepiece of the objective lens much stronger than usual and weaken C2 or even turn it off, as shown in Figure 9.3B. In addition, C1 must be strongly excited so the image of the gun crossover is a long way from C3. Thus the C3 image distance (d_i) is much less than the object distance (d_o), which gives a large demagnification of the C1 crossover (see equation 6.2).

From Figure 9.3B you can see that, although C2 is switched off, the C2 aperture still controls the convergence angle (α) of the beam on the specimen. As was the case for parallel-beam mode, a smaller C2 aperture gives a smaller α. You'll see later in Chapter 20 that the correct choice of C2 aperture is important in CBED and also in defining the exact dimensions of the probe for XEDS or EELS analysis (see Part 4). Some TEMs such as the FEI and older JEOLs use a double-gap C1 lens (check the situation for your own TEM in the manufacturer's handbook). In these TEMs there is a single C1 lens winding but effectively four polepieces and hence two crossovers occur in the bore of the lens. In some other TEMs (e.g., the Akashi 002B, the JEOL 2010, and Hitachi 8100 and later instruments) there are actually three independent lenses. Without this additional crossover, it's not possible to get a probe $< \sim$10 nm. An additional condenser lens will yield a probe size about 10× smaller. Even though you only get to use two controls (C1 and C2) there are at least three crossovers. So, as you can see, condenser-lens configurations can get much more complicated than the basic principle shown in Figure 9.3B. The more lenses you have, the more flexible the optics and therefore the more operations that a single TEM can perform.

So, in probe-forming TEMs, the role of C1 is fundamentally different from its role in traditional, parallel-beam TEM. The C1 lens is used to control the probe size (and thus the current) at the specimen. As shown in Figure 9.4, a strong C1 gives you a small probe while a weak C1 creates a large probe. This difference occurs because increasing the strength of C1 shortens its d_i, thus lengthening d_o for the probe-forming C2 (or C3) lens and, therefore, increasing the demagnification (but remember that this is a very simplified diagram).

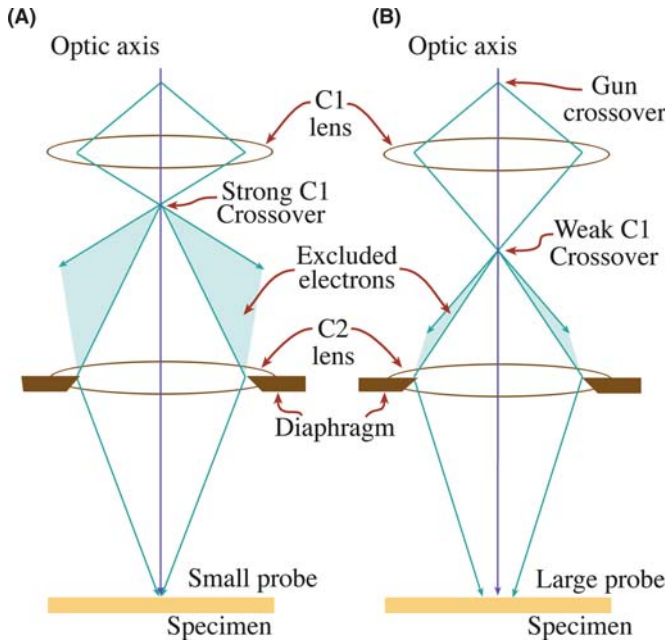


FIGURE 9.4. Effect of the C1 lens strength on probe size: a stronger C1 lens (A) results in greater demagnification by any subsequent lens (C2 and/or C3), giving a smaller electron beam at the specimen. A weaker lens (B) gives a broader probe. Note how changing C1 also changes the number of electrons that hit the C2 diaphragm and therefore, do not contribute to the current in the probe hitting the specimen.

9.1.C The Condenser-Objective Lens

When convergent-beam TEMs were first constructed, it was not possible to design an objective lens that would give both a parallel or a convergent beam at the specimen. So you had to change polepieces in order to change operating modes, which was highly inconvenient. To solve this problem, the condenser/objective (c/o) lens was developed in the mid-1970s. In a c/o lens, the magnetic field between the polepieces of the objective lens is very strong ($\sim$ 2 Tesla). This strong field has a dual function. First, it acts as a condenser lens and effectively causes the beam to converge on the specimen. Thus, as we mentioned above, we can refer to the condenser part of the lens as the C3. Second, once the beam has penetrated the specimen, it is focused just as it would be in a conventional lens. The beam is almost never parallel in a c/o lens and so, as shown in Figure 9.5, yet another lens is introduced into the column between the C2 lens and the objective lens. This extra lens is a condenser mini-lens (called a twin-lens in Philips/FEI instruments) and it is standard on TEMs that also operate as STEMs or analytical TEMs (i.e., it's standard on all modern TEMs used in materials characterization). Without the condenser mini-lens, the action of the c/o is always similar to the objective lens in a dedicated STEM because, in such instruments, parallel-beam imaging conditions are never required. All imaging and analyses are accomplished with a convergent probe. However, in TEM where both parallel and convergent beams are needed, when the mini-lens is activated, the convergent effect of the c/o lens can be offset by forming a crossover at the FFP of the objective. Here the FFP is actually acting as the condenser part of the c/o lens. Thus, unlike in conventional objective lenses (see Figure 6.5), the parallel-beam condition is met in the *overfocus* condition in the c/o lens. If you're feeling confused, take another look at Figure 9.5.

THE c/o LENS

In older TEMs without a c/o lens, underfocus gives a more parallel beam at the plane of the specimen than overfocus. This is not the case if the objective lens is a c/o lens and the beam is more parallel at the specimen when the c/o lens is overfocused.

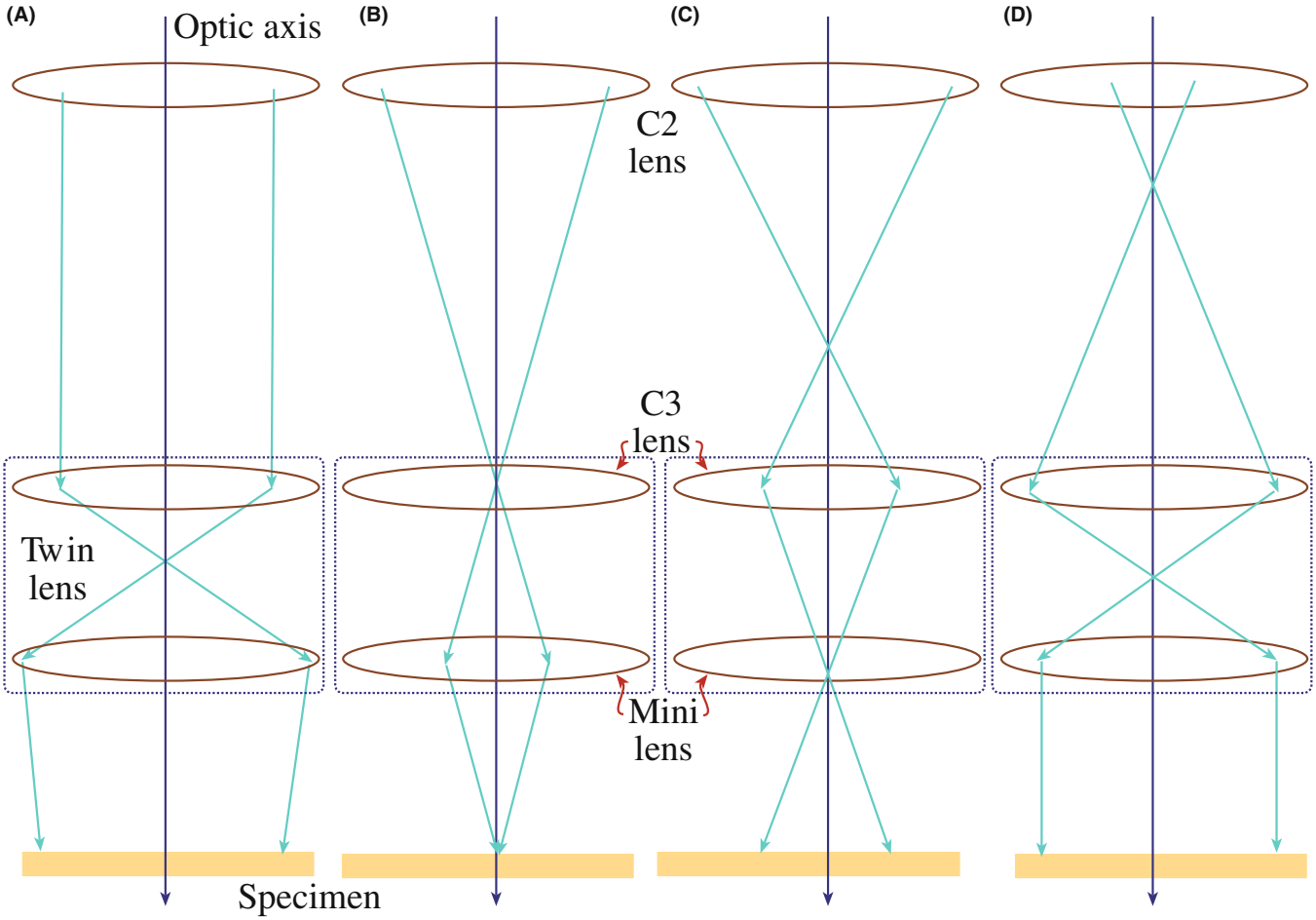


FIGURE 9.5. Condenser/objective lens operation as a function of the strength of the C2 lens. As the C2 strength is increased the beam on the specimen varies from (A) almost parallel, through (B) convergent, and (C) divergent, then (D) back to parallel. You can see that the crossover moves up the column as the C2 strength increases and the C3 lens introduces its own crossover prior to the minilens depending on the convergence of the ray paths entering it.

Figure 9.5 shows the various ray paths through a c/o lens, with a mini/twin lens, as the strength of the C2 lens is changed. As you can see, the beam at the specimen varies from parallel through convergent, divergent, and then back to parallel, as C2 is strengthened. Therefore, because of this interaction between C2 and the c/o lens, changing C2 changes the focus of the probe at the specimen and thus changes the focus of the STEM image and the CBED pattern and also changes the spatial resolution of any XEDS or EELS analyses, as we'll describe in detail in Part 4. So you have to be aware that, when you are operating in probe mode, two lenses control the focus, which can really mess things up unless you are fully aware of the situation. If you are operating in conventional parallel-beam TEM mode then you can simply use the objective lens (lower polepiece) to focus the image (see Section 9.3) and you don't need to adjust the illumination-system optics.

All recent probe-forming TEMs operate at a very specific c/o lens excitation where the STEM probe can be imaged in TEM mode on the viewing screen/computer display. This c/o condition for simultaneous focus and image is a very limited one. (Look in the manufacturer's instructions to find out how to set this condition exactly on your own TEM.) If you don't do this, you're not at the correct objective-lens excitation and the beam can act in contrary ways; the beam convergence can decrease when you expect it to increase and the electron distribution in the probe can be spread into a broad halo rather than a fine focused point. So find out how to do this whenever you're operating in probe-forming mode.

In a c/o lens, while the convergence angle can be effectively reduced to zero, the helical path of the electrons through the strongly excited upper-objective polepiece never permits truly parallel illumination. In contrast, in the older conventional C1/C2 optics, the objective pre-field is not strong and the spiraling is greatly reduced. Therefore, image contrast and other effects arise as a result of c/o lens operation. If you read Christenson, and Eades all of this is very well explained.

9.1.D Translating and Tilting the Beam

There are certain operations where we need to translate the beam laterally on the specimen (e.g., to position a fine probe on a feature of interest for analysis). Similarly, there are times when we need to tilt the beam off the optic axis so it impinges on the specimen at a specific angle (e.g., for centered-DF imaging using a specific diffraction spot which we describe in Section 9.3.C or for hollow-cone imaging/diffraction (Section 9.3.D) and precession diffraction (Section 18.8)). Both translating and tilting are also essential for aligning the beam down the column and are accomplished by varying the current through potentiometers (which we call 'scan coils') to generate a local magnetic field which deflects (rather than focuses) the beam. There are several sets of scan coils in the column, some of which tilt and others which translate the beam. Ray diagrams to explain translating and tilting the beam are shown in Figure 9.6A and B.

When we create a scanning beam for STEM imaging, the beam must always move parallel to the optic axis to mimic the parallel-beam illumination conditions of standard TEM. Such scanning is accomplished by tilting the beam twice with two sets of scan coils (one above the other) to ensure that the beam crosses the optic axis at the FFP of the upper-objective-lens polepiece. Then, wherever the beam enters the upper-objective-lens field, it is tilted to follow a path parallel to the optic axis (so sometimes you have to tilt in order to translate!). You need computer control to do this rather complex adjustment (look ahead to Figure 9.17). Like many other procedures on a modern TEM, this adjustment is made automatically when you select a particular (in this case STEM) operating mode.

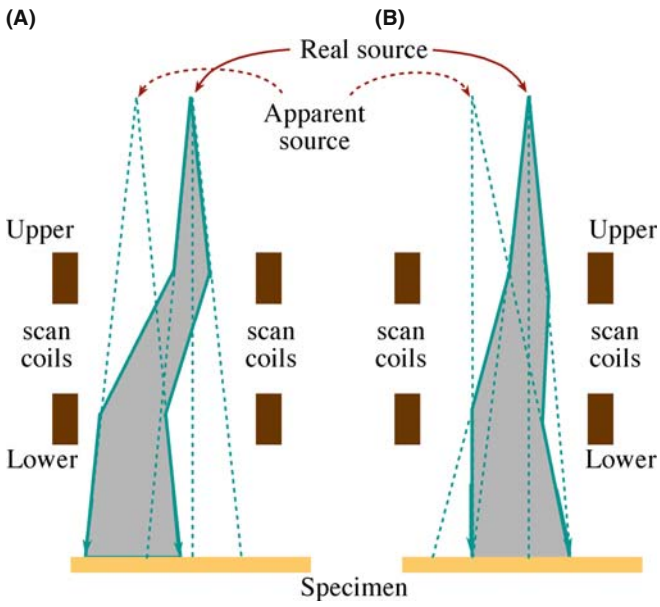


FIGURE 9.6. The use of pre-specimen scan coils for (A) traversing the beam and (B) tilting the beam. Traversing moves the beam to a different area of the specimen but it stays parallel to the optic axis. Conversely, tilting the beam illuminates the same area of the specimen, but from a different angle of incidence.

9.1.E Alignment of the C2 Aperture

Instructions for alignment vary for different TEMs so we'll simply describe the principles. Even if you won't be doing the alignment you will want to check that the microscope is correctly aligned; you can recognize if the wheels are not aligned on your car and you know it is important to balance them for best performance of the vehicle, even if you have to ask someone else to perform the task. If you want the best out of your machine, you'll want to be able to fine-tune this alignment.

If your illumination system is correctly aligned, the gun crossover is on the optic axis and the electrons can then travel down the optic axis in a straight line through the lenses and apertures until they hit the specimen. We'll assume that your gun is aligned so that the source crossover is on the optic axis of the column, as we described back in Chapter 5. Alignment of the illumination system used to be a tedious, manual affair involving tilting and translating each of the condenser lenses and centering the apertures on axis. Now most of the lens components are machined accurately enough so that only minor electronic tilts and translations are needed and usually these are computer controlled. Under these circumstances, centering the C2 aperture remains the most critical step in obtaining the best performance out of your TEM, particularly if you intend to operate in scanning mode for STEM imaging and analysis. (Other apertures in the illumination system are usually left untouched.) The C2 alignment remains a manual operation on most TEMs, although undoubtedly this will not always be the case.

Operational Procedure #1 You must always have the C2 aperture accurately centered on the optic axis of the TEM. If the aperture is misaligned, the image of the beam on the screen moves off axis and distorts as you underfocus or overfocus C2, as shown in Figure 9.7. To adjust the diaphragm so the aperture is on axis we have to alternately focus and defocus the C2 lens (we'll arbitrarily start by overfocusing but it actually doesn't matter if you under- or overfocus first). It's best to start at a low magnification so you can find the aperture image then repeat at the magnification range where you wish to operate.

- First, overfocus C2 (strengthen the lens) so the beam is spread on the screen and the image of the C2 aperture is visible (make sure any other apertures in the *imaging* system are out of the column).
- Then use the external manual drives (or computer control if you are lucky enough) to move the diaphragm so the aperture image is centered on the screen.
- Next, weaken C2 so the image of the beam is focused to its minimum diameter.

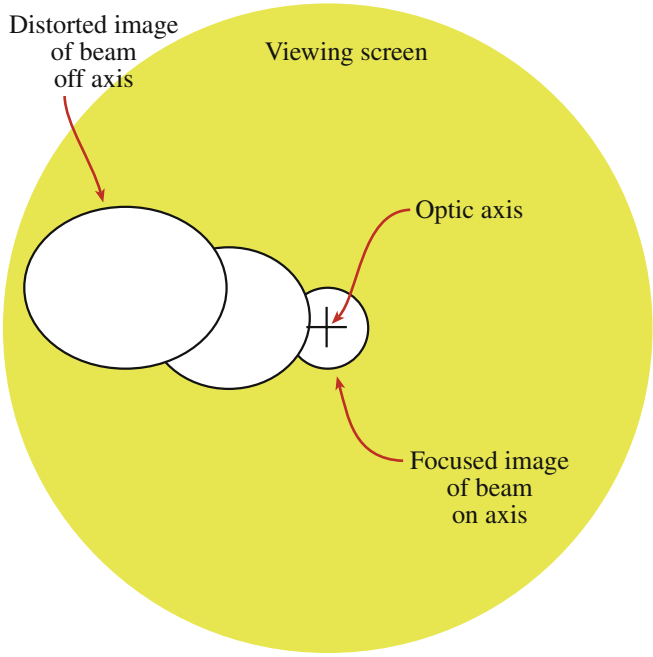


FIGURE 9.7. If the C2 aperture is misaligned, wobbling (alternately underfocusing and overfocusing the C2 lens) causes the image of the beam to sweep off axis (i.e., move across the viewing screen) and to become distorted.

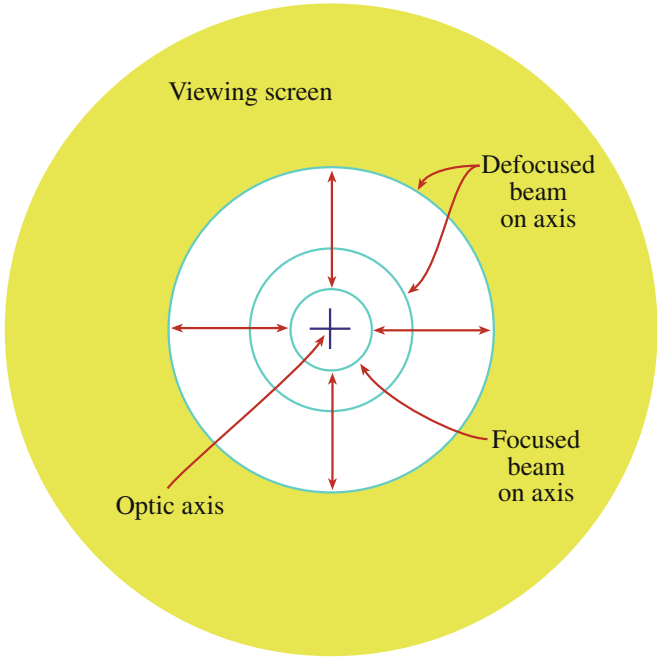


FIGURE 9.8. If the C2 aperture is aligned on axis, the image of the beam remains circular and expands or contracts about the optic axis as C2 is wobbled.

- Then, center the beam with the condenser deflector controls.
- Now underfocus the C2 lens (i.e., continue to weaken it) until you can again see the aperture image and center the image again with the external drives.
- Now strengthen the lens until the spot is focused and center it with the deflector coils.
- You have to repeat this whole operation iteratively until the image of the beam expands and contracts around the center of the screen as shown in Figure 9.8.

WOBBLING

Usually there's a control that will introduce an AC current into the lens coil, in effect 'wobbling' the lens setting either side of focus. This 'wobbler' saves you from manually underfocusing and overfocusing the lens, but think what you are doing.

In TEM mode this operation should only be necessary on the rare occasions when you change the diaphragm or the C1-lens setting but, in STEM, any change of probe size requires re-alignment of the C2 aperture.

9.1.F Condenser-Lens Defects

The illumination-system lenses suffer from all the standard lens defects we described in Chapter 6, such as aberrations and astigmatism. These defects don't really limit the operation of the TEM in parallel-beam mode, but they are crucial if you're intent on forming the finest probe possible for STEM and analytical work. Let's look at the role of each of the major defects.

Spherical Aberration This defect plays no role in limiting parallel-beam formation. However, as we discussed in Chapter 5, in adjusting the illumination system to form the finest possible probe with the maximum available current, spherical aberration in the probe-forming lens (C3) controls the minimum possible probe size. In exactly the same manner as we control the image resolution (see Chapter 6), spherical aberration limits the probe dimensions to a minimum radius (equation 6.22) of $r_{min} \sim 0.91(C_s\lambda^3)^{1/4}$. So the C3 probe-forming lens has a short focal length to minimize C_s. The final probe-limiting aperture in C2 needs to be carefully chosen to be the optimum value (equation 6.23) for the selected probe size $\alpha_{opt} = 0.77\,\lambda^{1/4}/C_s^{1/4}$. In practice, however, there are always more C1 settings than available C2 apertures, so it is not possible to choose the optimum aperture for each probe. This can cause problems if you need a specific probe size for a certain spatial resolution, as we discuss in Chapter 36. Of course, if you are lucky enough to be using a STEM with a C_s-corrected probe then this problem is minimized.

Chromatic Aberration Remember this aberration depends on the energy spread of the electrons. Since the electrons in the beam (particularly from a FEG)

have a very small energy spread we generally regard them as monochromatic and there is no detectable degradation of the probe dimensions. However, for the very best image resolution or the very best EELS, C_c-correction via monochromation of the beam can help, as we'll describe later in Section 37.7.

Astigmatism This is the most common defect in the TEM illumination system and arises either because the final limiting aperture in C2 is misaligned or it is contaminated and charging up, thus deflecting the beam. We'll assume you've centered the C2 aperture as we just described and we'll now tell you how to detect and correct any residual astigmatism due to contaminated apertures. The condenser stigmators introduce a compensating field (like a scan coil) which corrects this distortion.

Operational Procedure #2A — Detecting Astigmatism You can discern astigmatism in the illumination system by looking at an image of the electron source on the display/screen as follows

- Operate in image mode with no specimen inserted (or look through a hole in your specimen).
- Focus C2 so the beam is a minimum diameter (i.e., in spot mode with the size governed by C1).
- Adjust the C2 aperture and the beam traverses so the spot is in the middle of the screen and the image of the beam is circular, as you did when aligning the aperture.
- Wobble the C2 lens either side of the focal setting so the image of the beam expands and contracts about its minimum dimension.
- If there is astigmatism, the image is not circular, but distorts elliptically and rotates through 90° either side of focus, as shown in Figure 9.9.

Operational Procedure #2B — Correcting Astigmatism

- First overfocus the beam so you can see the effect of the astigmatism (i.e., the beam appears elliptical) in one direction.
- Then adjust the stigmators so the image appears circular.
- Now underfocus the beam and repeat the correction.
- Repeat the whole over/underfocus procedure iteratively until the image of the beam remains circular as you expand and contract it on the screen with the C2 lens (see Figure 9.8). Again, a wobbler can help here once you get the hang of the interplay between defocusing and the stigmators.
- If you can't make the image circular, you'll have to increase the range of strength of the stigmators. If you are on maximum strength, then there is too much contamination on the aperture and you either need to replace or flame-clean the diaphragm, as we described in Chapter 6.

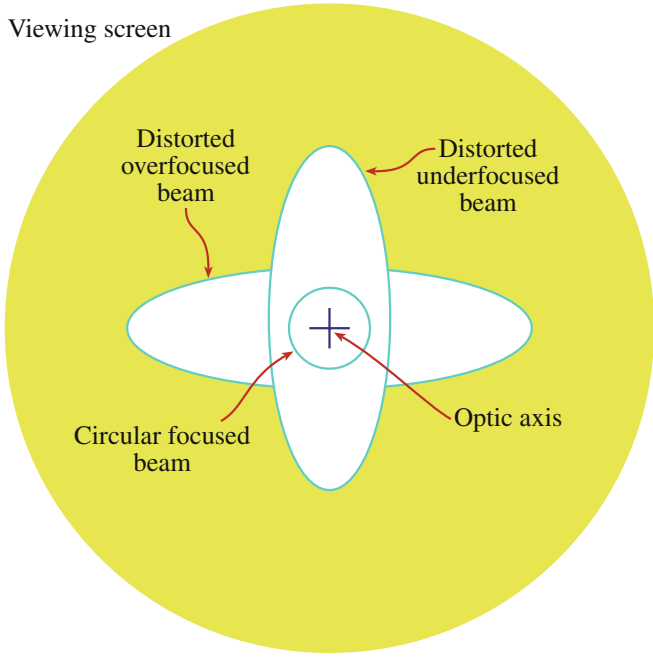


FIGURE 9.9. The effect of astigmatism in the illumination system is to distort the image of the beam elliptically as the C2 lens is wobbled. Correction of this astigmatism results in an image that remains circular as the C2 lens is defocused (as in Figure 9.8).

9.1.G Calibration

We've already described in Section 5.5 what it takes to measure the performance of the electron gun and to optimize the brightness in STEM operation so that the maximum beam current goes into the minimum probe size. We also have to calibrate the illumination system. The major variables are the probe size (and current) for various C1 settings and the convergence angle for various C2 aperture sizes.

The C1 lens strength controls the probe size at the plane of the specimen. We've described in some detail how to measure the beam dimensions at the specimen back in Section 5.5.C. in Chapter 5. Figure 9.10A shows the variation of the calculated (not measured) probe size as a function of the C1 lens setting for a typical TEM. These calculations are approximate, since they define the probe width as the FWHM and assume the C2 aperture correctly limits the Gaussian distribution. Despite these approximations you can clearly see the expected trend of decreasing probe size with increasing C1 strength. The probe current is a strong function of the probe size. If you look in the text by Goldstein et al. you'll find that the maximum probe current is proportional to (almost) the cube of the probe diameter: if you increase the probe size by a factor of 10, the current will increase by almost 1000×. This is an extraordinary dependence and explains why, particularly for probe-related TEM work (XEDS, EELS, STEM imaging, and CBED), we are very

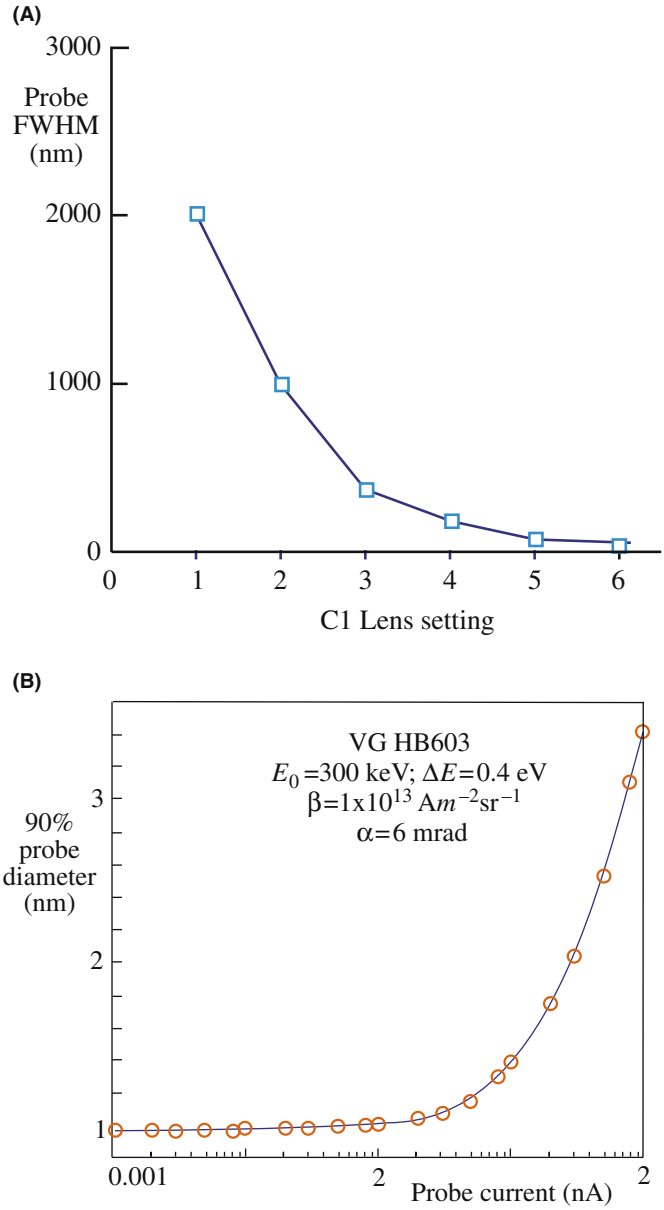


FIGURE 9.10. (A) Calibration of the illumination system requires determining the variation of the probe dimension with C1 lens strength. (B) The probe current is strongly related to the probe size. These data are from a 300-keV FEG-STEM; the dependence would be even greater in a thermionic-source TEM.

concerned about knowing and measuring the probe size and current. (There's much more about this in Part 4.) Figure 9.10B shows the variation of probe current with probe size for a FEG AEM.

The C2 aperture size governs the convergence (semi) angle α, as we also discussed in Chapter 5 when we were determining the gun brightness.

- We measure the total convergence angle 2α from a CBED pattern (Figure 5.8).
- We increase α by increasing the C2 aperture size (Figure 9.11).

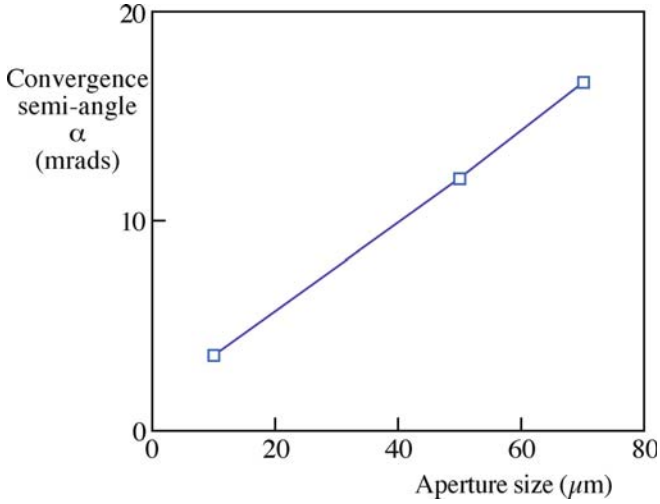


FIGURE 9.11. Increase of the beam-convergence angle, α, with increasing C2 aperture size. Increasing α will also increase the probe current.

Obviously the probe current will also be affected by the aperture size (e.g., if you double the aperture diameter you would expect four times the number of electrons to come through it). But, as you learned back in Chapter 6 (equations 6.22 and 6.23), because of the increasing lens aberrations with increasing α, there is an optimum aperture that gives the maximum probe current (and hence the best image resolution) that we just mentioned. You might well expect that, if there is a C_s corrector in the TEM illumination system, this limit can be overcome and you would be correct.

On some instruments the C2 aperture is virtual (so you have an effective aperture size), which makes it rather difficult to measure 2α. (See the Goldstein et al. text for a detailed description of this problem, which is common in SEMs.) Furthermore, if the C2 lens is excited, it can also change α and then you have to calibrate α both as a function of the aperture size and the C2 lens setting, which is an extremely tedious exercise.

9.2 THE OBJECTIVE LENS AND STAGE

This combination is the heart of the TEM. We use the stage to clamp the specimen holder in the correct position so the objective lens can form images and DPs in a reproducible manner. As we discussed in Chapter 8, there are two different types of holder, top-entry and side-entry and these determine the geometry of the polepiece and the flexibility which you have to make adjustments. We'll only consider side-entry holders here since they are standard but top-entry holders require the same adjustment of the *z*-control or specimen height.

As a practical consideration, you'll find that you would like to be able to tilt the specimen without changing its height on the optic axis. Otherwise you will be continuously using the *z*-control whenever you tilt. Clearly this means that you should ensure that the region of the specimen you want to work on is located close to the tilt axis of the specimen rod.

FOCUSING AND HEIGHT

We cannot emphasize too strongly that we need to fix the height of our specimen on the optic axis so we can always work at the same objective-lens current and thus at a fixed objective-lens magnification.

This is **really** important! This is *usually* ignored!

The central requirement here is the need to define a reference plane so that our calibrations will be reproducible. The reference plane (see Chapter 6) for a side-entry holder is called the eucentric plane. This plane is normal to the optic axis and contains the axis of the specimen-holder rod; clearly there could be many such planes. What is special about the eucentric plane is that when your specimen is located at this plane and the image is in focus, the objective-lens current is a standard reference value. The position of this plane within the objective lens is known as the eucentric height. If you put your specimen in the eucentric plane, then a point on the optic axis does not move laterally when you tilt it around the holder axis, which makes many essential imaging and diffraction operations very easy. Unfortunately, if you tilt your specimen normal to the holder axis, or rotate/traverse it off axis, then the area you're examining almost invariably moves out of the eucentric plane upon tilting and the image can easily disappear from the screen. This limitation can only be overcome by complete computer control of all five axes (three translations (x, y, z) and two tilts (parallel and orthogonal to the axis of the holder)) and such stages are often called double-eucentric. They are becoming more commonly available and are extremely useful. Get one if you can!

Operational Procedure #3A The first thing you must always do when inserting your specimen into the TEM is to ensure that it is in the eucentric plane. To do this, first find the *z*-control, which is usually a knob on the outside of the goniometer stage near the hole where you insert the holder. Then

- Select a recognizable feature in the image and position it at the center of the screen at a reasonable magnification (20–50,000×).
- Tilt the goniometer stage a few degrees clockwise or anticlockwise until the feature moves close the limit of the viewing screen. (If the slightest tilt has this effect then go to a lower magnification until you can tilt a few degrees and still see the feature on the screen.)
- Adjust the height of the specimen holder until the feature returns to the center of the screen.
- Tilt the goniometer the other way and repeat the process.
- Continue to increase the tilt both ways and compensate with the *z*-control until the image of the specimen remains stationary when you tilt the specimen through about ±30° either side of zero tilt.
- Repeat at a higher magnification as needed.

Operational Procedure #3B: Another Way to Set the Eucentricity The eucentric position of the stage is a mechanical point located in the column. As we mentioned earlier, the optimal c/o lens setting in a STEM/TEM is a very well-defined and reproducible excitation of the objective lens. This lens condition also defines a mechanical point somewhere between the polepieces. We need to make the two points (eucentric position and the optimal c/o lens excitation position) the same. To find the eucentric point when these conditions are met, you need to

- Set the lens condition to the optimal c/o operating condition.
- Focus the image using the *z*-control.
- The specimen height will now be set to the eucentric position.

With practice you can set these two points to differ by <10 μm, and < 3 μm is not uncommon.

With computer control and auto-focusing techniques becoming common, this operation can be automated. As a result, in a completely eucentric stage, your specimen doesn't move off the optic axis and remains in focus no matter around what axis it is tilted or rotated. If you don't have such a computer-controlled stage, be cautious and develop the skill to tilt and change the *z*-control simultaneously.

The eucentric plane should also be symmetrically positioned with respect to the upper and lower objective polepiece fields. At this condition, the eucentric plane coincides with the plane at which the electron beam is imaged, in both TEM and STEM modes. If the symmetric plane and the eucentric plane are not coincident, then the images and DPs will appear at different magnifications and different focus settings in TEM and STEM. Obviously this requirement has no meaning in a DSTEM where there is no TEM mode.

Ensuring coincidence of the eucentric and symmetric planes is usually carried out by the manufacturer. You can check it by comparing the focus of a DP or an image in TEM and STEM modes. You should not have to refocus the image or DP with the objective lens when you change from one mode to the other. If you find this is a problem, talk to your technical staff.

9.3 FORMING DPs AND IMAGES: THE TEM IMAGING SYSTEM

You know already that the objective lens takes the electrons emerging from the exit surface of the specimen, disperses them to create a DP in the back-focal plane (BFP), and recombines them to form an image in the image plane (see Figure 6.3). We can use this ray diagram to introduce the basic operations for forming static-beam images and DPs in the TEM. We'll then describe how to do the same thing with a scanning beam in STEM mode.

In this discussion, we will skip many of the details and concentrate on the role of the instrument. In Chapter 11 we'll discuss the details of the diffraction process and then expand these ideas in Chapters 16 through 21. We'll then discuss the images formed in the TEM in Chapters 22 through 31.

PUSH-BUTTON TEM

There are buttons to push on the console or icons to click on the computer screen by which you choose either **image mode** or **diffraction mode**. These are the primary modes, central to TEM operation. Now TEM is so easy it has never been more difficult!

The first operation that you need to master when using the TEM is viewing the DP (diffraction mode). In all the subsequent imaging modes, we'll form our images by using the objective aperture and the DP to select electrons that have been scattered through particular angles.

- *Diffraction mode*: to see the DP you have to adjust the imaging-system lenses so that the *BFP* of the objective lens acts as the object plane for the intermediate lens. Then the DP is projected onto the viewing screen/CCD as shown in Figure 9.12A.
- *Image mode*: if you want to look at an image instead, you readjust the intermediate lens so that its object plane is the *image plane* of the objective lens. Then an image is projected onto the viewing screen/CCD, as shown in Figure 9.12B.

Let's look now at the details of these two fundamental operations from the point of view of the instrument. In subsequent chapters, we will discuss how to understand the images/DPs and why we form them in the ways we do.

9.3.A Selected-Area Diffraction

As you can see from Figure 9.12A, the DP contains electrons from the whole area of the specimen that we illuminate with the beam. Such a pattern is not very useful because the specimen will often be buckled. Furthermore, the direct beam is often so intense that it will damage the viewing screen or saturate the CCD camera. So we perform a basic TEM operation which both selects a specific area of the specimen to contribute to the DP and reduces the intensity of the direct beam in the DP on the screen.

If you look at Figure 9.12A, there are two ways we could reduce the illuminated area of the specimen contributing to the DP.

- We could make the beam smaller
- We could insert an aperture above the specimen which would only permit electrons that pass through it to hit the specimen

The first option involves using C2 and/or C3 to converge the beam at the specimen. We use this approach to form CBED patterns, which we'll discuss in great detail in Chapters 20 and 21. Converging the beam destroys any parallelism, and spots in the pattern are not sharply defined but spread into disks. If we wish to obtain a DP with a parallel beam of electrons, the standard way is to use the selected-area aperture. Now, we can't insert an aperture at the specimen plane, because the specimen is already there! But if we insert an aperture in a plane *conjugate with* the specimen, i.e., in one of the image planes of an imaging lens, then it creates a virtual aperture at the plane of the specimen. This is exactly what we do. This operation is called selected-area diffraction or SAD, and is shown in Figure 9.12A.

The conjugate plane that we choose is the image plane of the objective lens. As shown in Figure 9.13, we create SADPs by inserting what we call the SAD aperture into the image plane of the objective lens and we center this aperture on the optic axis in the middle of the viewing screen. You can see this aperture if you want to change it or center it, by projecting the image plane onto the viewing screen/CCD, as we'll discuss in Section 9.3.B below.

Operational Procedure #4 The specific steps to form an SADP are as follows (while it looks complicated you'll soon find that it only takes a few seconds to do all these steps)

- Choose image mode so you can see an image of your thin specimen on the viewing screen.

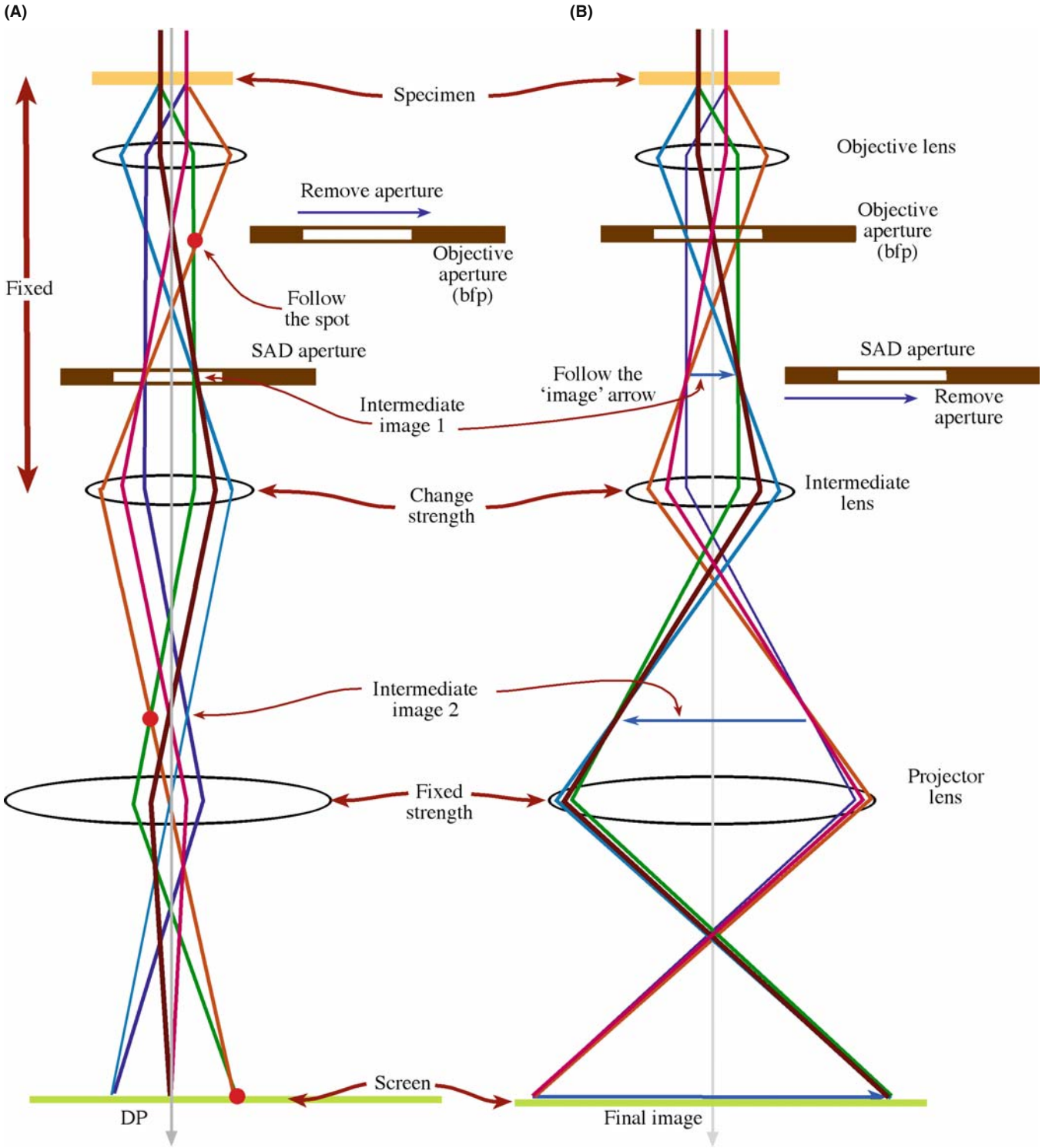


FIGURE 9.12. The two basic operations of the TEM imaging system involve (A) diffraction mode: projecting the DP onto the viewing screen and (B) image mode: projecting the image onto the screen. In each case the intermediate lens selects either the BFP (A) or the image plane (B) of the objective lens as its object. The imaging systems shown here are highly simplified. Most TEMs have many more imaging lenses, which give greater flexibility in terms of magnification and focusing range for both images and DPs. The SAD and objective diaphragms are also shown appropriately inserted or retracted. NOTE: This is a highly simplified diagram showing only three lenses. Modern TEM columns have many more lenses in their imaging systems.

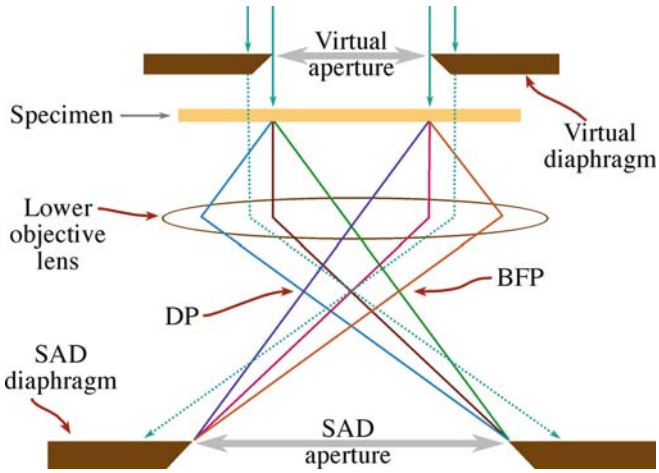


FIGURE 9.13. Ray diagram showing SADP formation: the insertion of an aperture in the image plane results in the creation of a virtual aperture in the plane of the specimen (shown here slightly above the specimen plane). Only electrons falling inside the dimensions of the virtual aperture at the entrance surface of the specimen will be allowed through into the imaging system to contribute to the SAD pattern. All other electrons (dotted lines) will hit the SAD diaphragm.

Under these conditions, with an SAD aperture inserted and the objective apertures removed, any electron that hits the specimen outside the area defined by the virtual aperture will hit the diaphragm when it travels onto the image plane. Such electrons will thus be excluded from contributing to the DP that is projected onto the viewing screen. In practice, we can't make apertures smaller than ~10 μm, and the demagnification back to the plane of the specimen is only ~25× which gives a minimum selected area of ~0.4 μm—which isn't as small as we'd like, particularly in today's nano-world. We'll discuss in Chapter 11 whether or not smaller values would be useful. As always, you really need to know these values for the TEM you are using, so do the calibration.

SAD

It is a basic principle of TEM operation that when you want to look at the DP (i.e., the **BFP** of the objective lens), you put an SAD aperture into the **image plane** of the objective lens.

- Spread the beam by underfocusing C2.
- Make sure that the objective diaphragm is retracted. The image contrast will be poor and it will be difficult to focus but that's all right at this stage.
- Insert the SAD diaphragm and start by choosing the largest available aperture to make finding the aperture easy (there may be three to five different apertures depending on the specific TEM you are using).
- If the screen then goes blank and you can't see anything, the aperture you chose is not on axis and you should lower the image magnification until you can see the aperture and center it on the axis.
- If the illuminated area of the image isn't restricted when you insert the aperture then you are at too high a magnification; so lower it until you can see the image of this aperture on the viewing screen.
- Traverse the specimen if necessary to make sure the specific region you are interested in is visible through the aperture.
- You must focus the SAD aperture by adjusting the intermediate lens so it is conjugate with (i.e., exactly in the plane of) the image of the specimen that we focused with the objective lens.
- Switch to diffraction mode and an SADP will appear on the screen.
- Focus the SADP with the diffraction-focus controls so the spots are sharp. Further underfocusing C2 may also help at this stage.
- If you need a smaller aperture to select a smaller region of the specimen, simply click the diaphragm holder in or out to choose a smaller aperture and repeat the above procedure.

The SADP is often displayed on the viewing screen at a fixed magnification so that you can recognize differences in the magnitude of **g**.

By analogy with an X-ray diffractometer or pinhole camera, we define a distance called the 'camera length' (L). We think of this length as being the distance of the recording plane from the DP (look ahead to Figure 9.23); it's analogous to a real distance in a real camera. We choose the value of L such that inter-spot or ring spacings in the DP are easily discernible on the viewing screen and on the recording medium. This magnification can be changed by adjusting the intermediate lenses. We'll describe how we calibrate this magnification in Section 9.6.B.

At this stage, however, it's worth noting that, with a c/o lens, calibration of L can be a challenge because changing the condenser-lens focus as we just described changes the diffraction crossover and, if you re-focus the SADP with the intermediate lens, then you'll measure a different L.

Beware: in all the early TEM books, SAD is the only standard diffraction technique. As a result, some microscopists only use SAD to obtain diffraction information. However, CBED, which we discuss in Chapters 20 and 21, provides a great deal of complementary diffraction information and *must* be used by all TEM operators in the materials and nanotechnology fields. But there are still certain times when your best option is to form an SADP

- When you need to select a spot from which to form a BF or DF image (see next section)
- When diffraction spots are very close to one another and would overlap in CBED patterns (see examples in Chapters 23 and 24)

- When you are looking for fine structure in the DP such as streaks (see Chapter 17)
- When your specimen is beam sensitive

On all other occasions, when the diffraction maxima provide the most important information in the pattern, you should use CBED.

9.3.B Bright-Field and Dark-Field Imaging

When the SADP is projected onto the viewing screen/CCD, we can use the pattern to perform the two most basic imaging operations in the TEM. No matter what kind of specimen you're looking at, the SADP will contain a bright central spot that contains the direct-beam electrons and some scattered electrons (as shown in Figure 2.13A–C), the distribution of which will depend on the nature of your specimen.

IMAGING
It is another basic principle of TEM operation that if you want to view an image (i.e., the **image plane** of the objective lens), you insert an aperture called the objective aperture into the **BFP** of the objective lens.

Operational Procedure #5 When we form images in the TEM, we either form an image using the central spot or we use some or all of the scattered electrons. The way we choose which electrons form the image is

- Starting with both SAD and objective diaphragm holders retracted, select image mode to observe an image of the desired region of your specimen on the screen. It will not be a good image because there is no objective aperture inserted, but that is OK at this stage.
- Form an SADP of a selected region as we described in operational procedure #4.
- Insert the objective diaphragm so an aperture appears in the BFP of the objective lens, thus blocking out most of the DP, except that small area which is visible through the aperture. As with the SAD operation above, we use the external drives to move the aperture.
- Adjust the external diaphragm drives so the aperture selects the direct beam on the optic axis.
  This centering procedure is absolutely critical to forming the best image. If you have binoculars, view the TEM screen at a much higher magnification to ensure the best possible centering; otherwise do the equivalent using the digital image.
- Return to image mode, remove the SAD aperture, and focus the image with the objective lens. We call the resultant image, formed by the direct-beam electrons, a bright-field (BF) image (Figure 9.14A).
- If instead we choose scattered electrons (e.g., a specific diffracted beam (usually) or a portion of a diffraction ring), then we call the image a dark-field (DF) image (Figure 9.14B).

You can view the BF and DF images at any magnification simply by adjusting the intermediate lenses of the microscope and typical magnification ranges will be ~25,000× to 100,000× for a wide field of view, but up to 10^6× for high-resolution images. Usually you'll have to calibrate the actual magnification and also be able to relate directions in the image at any magnification to directions in the DP at a fixed camera length. These are the two basic calibrations required for any TEM.

BF AND DF
Select the direct beam to form a BF image.
Select only electrons that are not in the direct beam to form a DF image.

The insertion and removal of the SAD and objective apertures can be confusing to the beginner and often the wrong aperture is inserted or not removed when it should be. You have to practice obtaining SADPs and BF/DF images to get used to what aperture should be inserted and when. Both apertures are inserted below the objective lens. The objective aperture goes into the BFP, so it is closer to the lens (i.e., higher up the column) than the SAD aperture, which is in the image plane. Remember that if you're looking at a DP the (lower) SAD aperture should be inserted and the (upper) objective aperture removed. If you want to look at an image, the objective aperture should be inserted and the SAD aperture removed.

OBJECTIVE APERTURE
This is the most important aperture in the TEM. When inserted, its size controls the collection angle (β); hence, it determines the effect of the aberrations of the most important lens in the instrument and thus directly influences the resolution.

9.3.C Centered Dark-Field Operation

If you look at Figure 9.14B, the electrons that are selected by the aperture travel off the optic axis, since we displace the aperture to select the scattered electrons. The more off-axis the electrons are the greater the aberrations and astigmatism they suffer. Therefore, such a displaced-aperture DF (DADF) image is difficult to focus on an older TEM. You'll find that the image will move on the screen as you adjust the objective-lens strength. While there may be situations where you want to use such a DADF image, almost invariably you have to get any

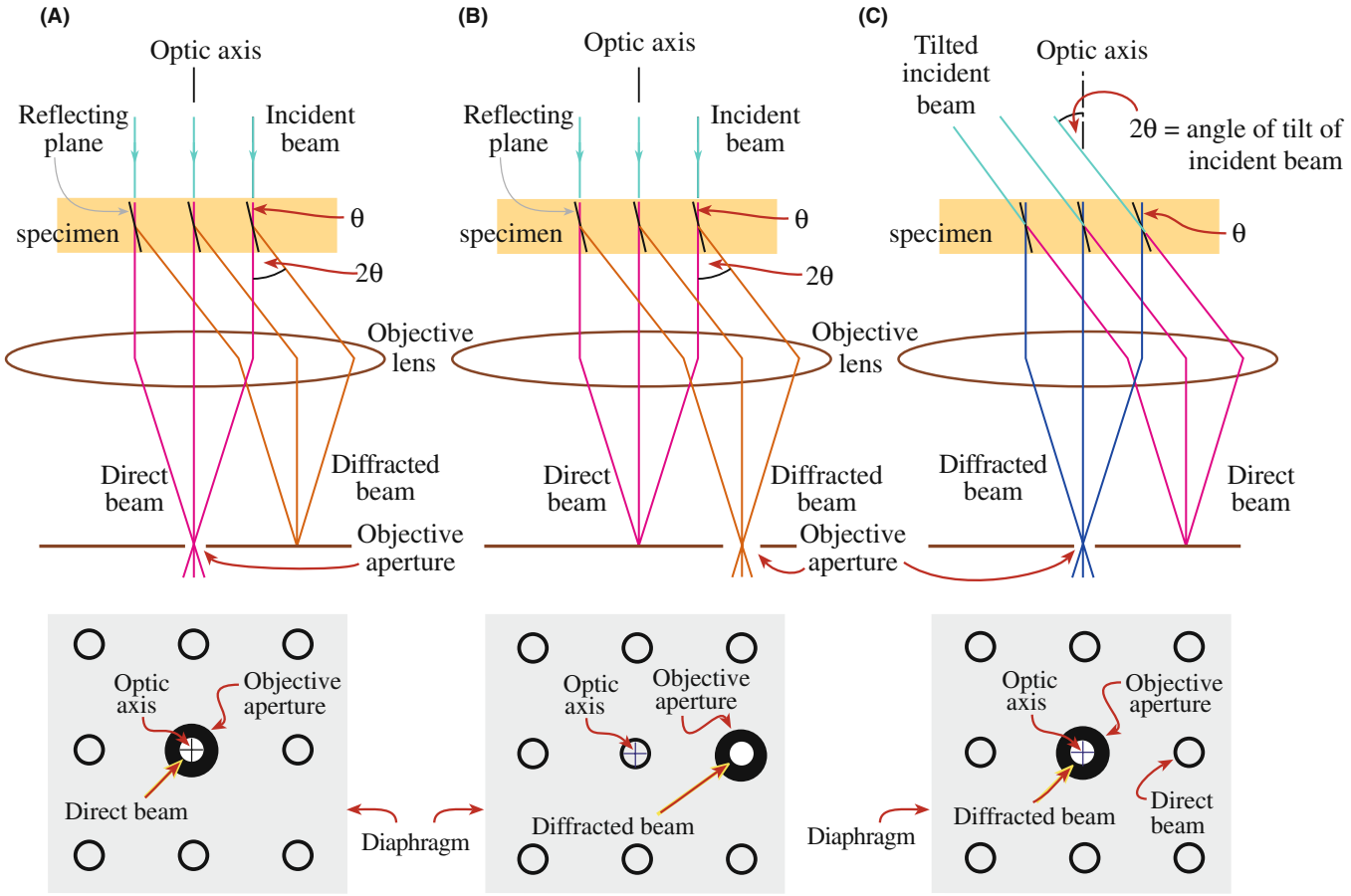


FIGURE 9.14. Ray diagrams showing how the objective lens and objective aperture are used in combination to produce (A) a BF image formed from the direct electron beam, (B) a displaced-aperture DF image formed with a specific off-axis scattered beam, and (C) a CDF image where the incident beam is tilted so that the scattered beam emerges on the optic axis. The area of the DP selected by the objective aperture, as seen on the viewing screen, is shown below each ray diagram. (Images comparing BF and DF are given in Part 3.)

scattered electrons back on the optic axis and then form a DF image under this condition. This operation is called centered dark-field (CDF) imaging. It is the conventional way to do DF imaging so we'll tell you how it's done. If you have a new TEM (with a small C_s) you should consider using only DADF for SAD because it's easier and you've eliminated the reason for forming CDF images!

Operational Procedure #6

- Start with an SADP on the screen and instead of moving the objective diaphragm so the aperture selects scattered electrons, adjust the aperture so it is on axis, as for BF imaging.
- Retract the objective diaphragm drive so you can see the SADP.
- Switch on and adjust the beam-tilt (DF) potentiometers above the objective lens so that the scattered-electron beam that you wish to use to form the CDF image moves toward the central, on-axis position.
- Looking through the binoculars, switch the potentiometers on and off to ensure that the scattered beam exactly superimposes where the direct beam was when the potentiometers are on and the direct beam is back on axis when the potentiometers are off.
- Re-introduce the objective diaphragm and check that the aperture is still centered around the on-axis scattered beam.
  As with BF imaging, this aperture-centering procedure is absolutely critical to forming the best CDF image.
- Select image mode again and focus the CDF image.

What you are doing here is making the incident beam hit the specimen at an angle equal and opposite to the scattering angle. In this way the scattered electrons will now travel down the optic axis, as shown in Figure 9.14C.

We'll return to BF, CDF, DADF, and SAD operations when we discuss specific contrast mechanisms that occur in TEM images in Chapter 22.

9.3.D Hollow-Cone Diffraction and Dark-Field Imaging

One obvious limitation to either DADF or centered CDF imaging is that we are only using a small fraction of the scattered electrons. When that fraction is a specific diffraction spot from a crystalline specimen that can indeed be very useful, as we'll see in Part 3 of the book. However, if, for example, we want to see *all* the portions of the specimen or *all* the precipitate phases that are diffracting into a *set* of diffraction spots (e.g., {111} rather than (111)) then we can either take separate CDF images from *all* the individual (111) reflections in the SADP (which is obviously a pain) or we can carry out hollow-cone (also called conical) diffraction and DF imaging. Likewise, if your specimen is micro/nanocrystalline so it generates a ring pattern, or if it is amorphous so diffuse-intensity rings are formed, then we can use the same technique to maximize the information in the images. There are two ways to do hollow-cone DF imaging: hardware and software. In essence, in either method, a specific set of diffracted beams is collected by the objective aperture when the conical-scanning beam satisfies a particular Bragg reflection condition (see Section 3.10 and Chapter 11).

- *Hardware*: Use an annular condenser aperture such that a cone of electrons illuminates the specimen at a fixed angle to the optic axis. Annular apertures fabricated by FIB techniques (see Chapter 10) are becoming more commonly available. (Think about how to make an annular aperture so that the middle doesn't just fall out of the doughnut shape leaving just a large circular hole?) So then we basically have multiple DPs formed; the direct (000) beam travels through the specimen and emerges off axis and a set of diffracted beams (depending on the cone angle) is always scattered on axis.
- *Software*: Use computer control of the scan coils to spin the incident (and thus the direct) beam around the optic axis. This method is obviously more flexible because, while an annular C2 aperture gives a fixed cone (semi) angle, the scan coils can be adjusted to give a variable cone angle. Thus, with the software approach, all the electrons scattered at a specific angle can be integrated into a single DF image which you can envisage as spinning (Figure 9.14C) around so that a set of diffraction spots always rotates through the optic axis. The principle of hollow-cone diffraction and imaging is shown in Figure 9.15.

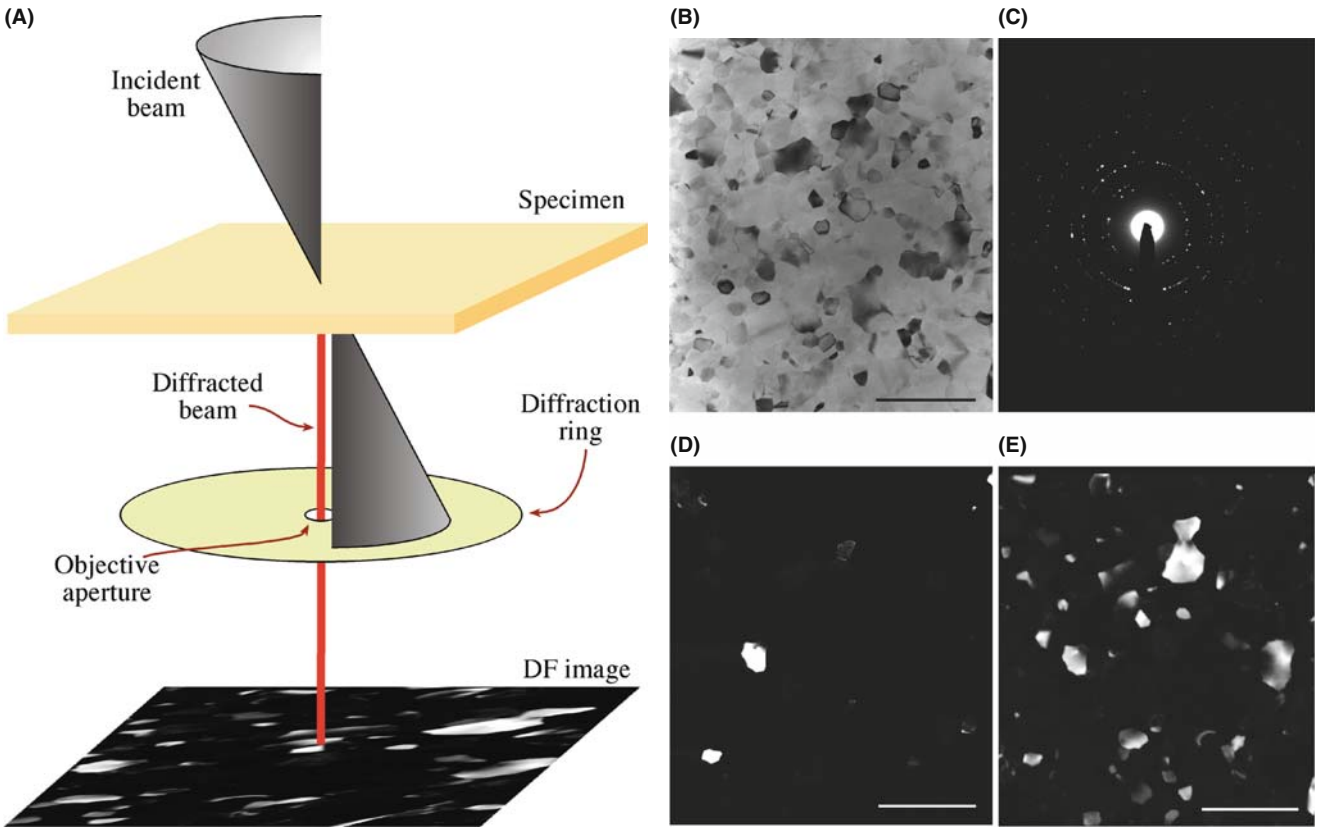


FIGURE 9.15. (A) Ray diagram showing hollow-cone illumination conditions. The direct beam is always off axis but electrons diffracted at the cone angle are always scattered on axis. (B) A BF image of a nanocrystalline Al film. (C) A hollow-cone SAD pattern from the film. Thus, while a single (220) CDF image (D) reveals only a couple of strongly diffracting crystals, a hollow-cone DF image (E) from all the {220} reflections shows diffracted intensity from dozens of grains. The scale bar is 500 nm.

Computer control of the scan coils allows you to select the angle, i.e., the radius of the circle, thus selecting which (*hkl*) ring is to be on the optic axis. You can also control the speed of the rotation so you can synchronize a single rotation with sufficient exposure time to record the DF image. This process is analogous to the normal DF tilt controls that shift the DP in the *x*–*y* plane (by tilting the incident beam as in Figure 9.14C) being replaced with *r*–θ axis controls. If you observe the rotating DP with the objective aperture removed, the direct (000) beam describes a circle, radius *r*, around the optic axis. Any point in the DP that is distance *r* from the (000) spot will at some time pass through the optic axis during the rotation of the DP. An example of a hollow-cone DP (not spinning) is shown in Figure 9.15C. The SAD aperture was about the size of the image (~ 50 μm across) and the objective aperture was small enough to permit diffracted intensity from only one ring to be collected. Figure 9.15B, D and E shows, respectively, a BF image of a polycrystalline metal film from which the DP in Figure 9.15C was obtained, a conventional CDF from a few spots in the DP ring and the hollow-cone DF, revealing many more diffracting grains. You usually have to take a series of time exposures to get a good hollow-cone image. Figure 9.15E was a 20-second exposure during which time the DP rotated about 1000 times through the objective aperture!

Hollow-cone DPs are also discussed in Section 18.6. A closely related technique, precession diffraction, is covered in Section 18.8. The STEM equivalent of hollow-cone DF imaging is annular-DF imaging, which we'll talk about in the next section.

9.4 FORMING DPs AND IMAGES: THE STEM IMAGING SYSTEM

If you want to use a fine probe to form STEM images then the objective-lens optics are a little more complex than in TEM. The key feature to remember is that the scanning beam must not change direction as the beam is scanned (unlike in an SEM where the scanning beam simply pivots about a point above the specimen). If the incident direction varies then the electron scattering (particularly the diffraction) processes would change as the beam intercepts the specimen at different angles. So interpreting the image contrast would be rather difficult to say the least.

STEM

The beam must scan parallel to the optic axis at all times so that it mimics the parallel beam in a TEM even though it's scanning.

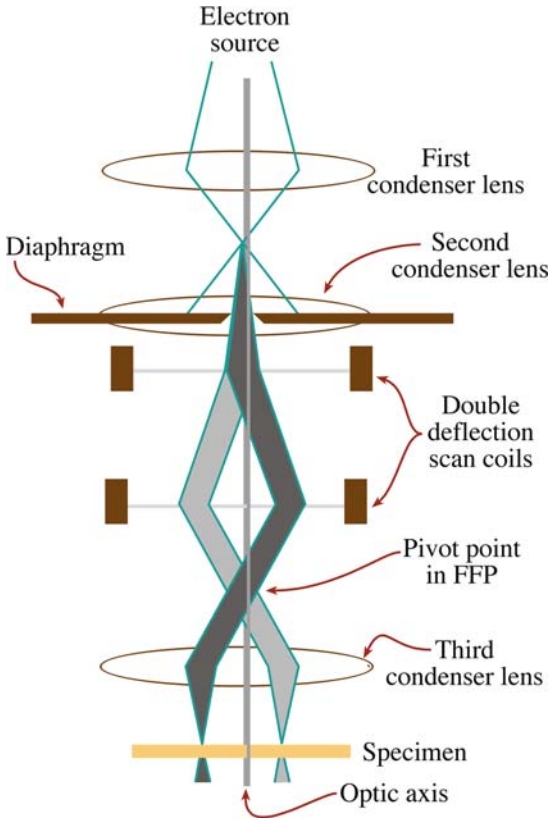


FIGURE 9.16. Scanning the convergent probe for STEM image formation using two pairs of scan coils between the C2 lens (usually switched off) and the upper-objective polepiece. The double-deflection process ensures that the probe remains parallel to the optic axis as it scans across the specimen surface.

As we show in Figure 9.16, the way we achieve parallel incidence is to use two pairs of scan coils to pivot the beam about the FFP of the upper objective (C3) polepiece. The C3 lens then ensures that all electrons emerging from the pivot point are brought parallel to the optic axis and an image of the C1 lens crossover is formed in the specimen plane. Now, if the objective lens is symmetrical, and the lower objective polepiece is similarly strong, then a stationary DP is formed in the BFP (this pattern does not move, even though the beam is scanning, since it is conjugate with the FFP, as shown in Figure 9.17). If we stop the beam from scanning, then we have a CBED pattern in the BFP and we can project that onto the TEM computer screen if we wish. So let's first discuss how to form STEM images.

THE STEM IMAGING LENS?

The STEM image quality depends on the probe. The probe has aberrations because we use a lens to form it. So the STEM image quality does depend on a lens (just not on an imaging lens).

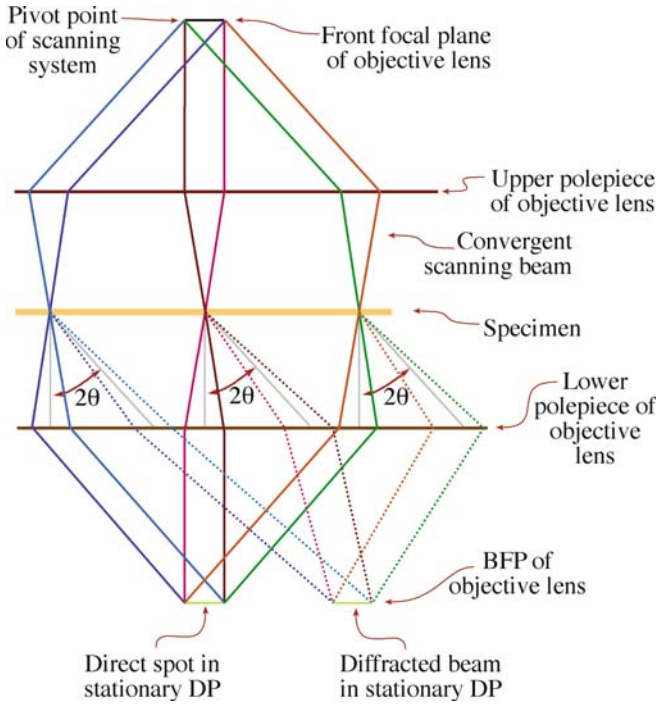


FIGURE 9.17. The creation of a stationary (convergent-beam) DP in the BFP of the objective lens is a necessary prerequisite for STEM imaging. Note that electrons scattered through the same angle (2θ) at different points in the specimen are focused at the same point in the BFP.

One potentially very big advantage for STEM is that, just like in an SEM, we don't use lenses to form the image. So defects in the *imaging lenses* do not affect our image resolution, which is limited by the beam dimensions. Hence chromatic aberration which (as we saw back in Section 6.5.B) can seriously limit TEM image resolution, is absent in STEM images. This is a great advantage if you're dealing with a thick specimen. However, there are drawbacks also, as we'll discuss below and in Part 3, and STEM images aren't widely used, particularly for crystalline specimens.

9.4.A Bright-Field STEM Images

Image formation in the scanning mode is fundamentally different from static-beam TEM image formation. As you've just seen in the TEM, we select a portion of the electrons emerging from an area of the specimen and project that distribution onto a screen. The principle of scanning-image formation is shown in Figure 9.18. Simply stated, we scan the beam on the specimen by adjusting the scan coils; these same coils are used to scan the computer display synchronously. The electron detector acts as the interface between the electrons coming from the specimen and the image viewed on the display screen. Since it takes up to 2048 scan lines to build up an image on the recording screen, the whole process of creating a STEM image is much slower than TEM imaging: it's serial recording instead of parallel recording.

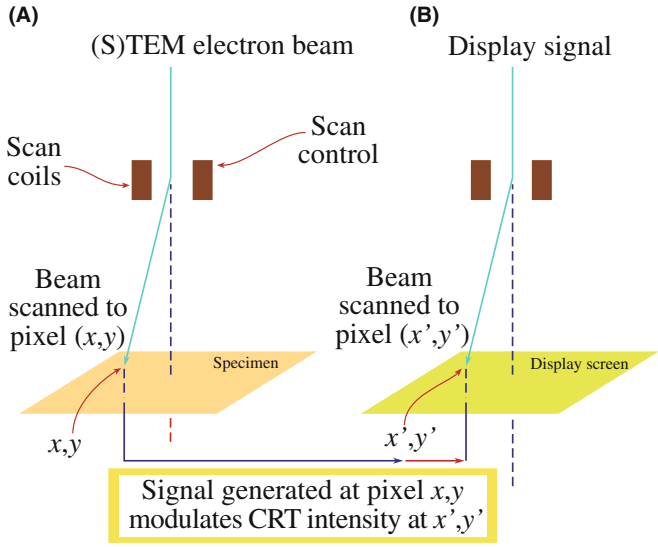


FIGURE 9.18. The principle of forming a scanning image, showing how the same scan coils in the microscope control (A) the beam-scan on the specimen and (B) the beam-scan on the computer display screen of the STEM. Thus no lenses are required to form the image.

THE STEM SIGNAL

The signal is generated at a point on the specimen, detected, amplified, and a corresponding signal displayed at an equivalent point on the computer display. The image builds up over several seconds or even minutes.

This process is exactly the same principle as used in any scanning-beam instrument such as an SEM or an STM (scanning-tunneling microscope). Remember that to form a TEM-BF image, we inserted an aperture into the plane of the TEM DP and only allowed the direct electrons through it into the imaging system. In STEM mode we use an electron detector, in exactly the same way as we use the aperture: we only allow the electrons that we want to contribute to the image to hit the detector. So we insert a BF (either a semiconductor or scintillator-PM) detector onto the axis of the microscope and it intercepts the direct-beam electrons no matter where the beam is scanning on the specimen, as shown in Figure 9.19A. So a variable, direct-beam signal travels from the detector via an amplification system to modulate the signal on the computer display, thus building up a BF image as also shown in Figure 9.19D.

BF DETECTOR

The BF detector picks up the direct beam which varies in intensity depending on the specific point on the specimen illuminated by the probe at that specific time.

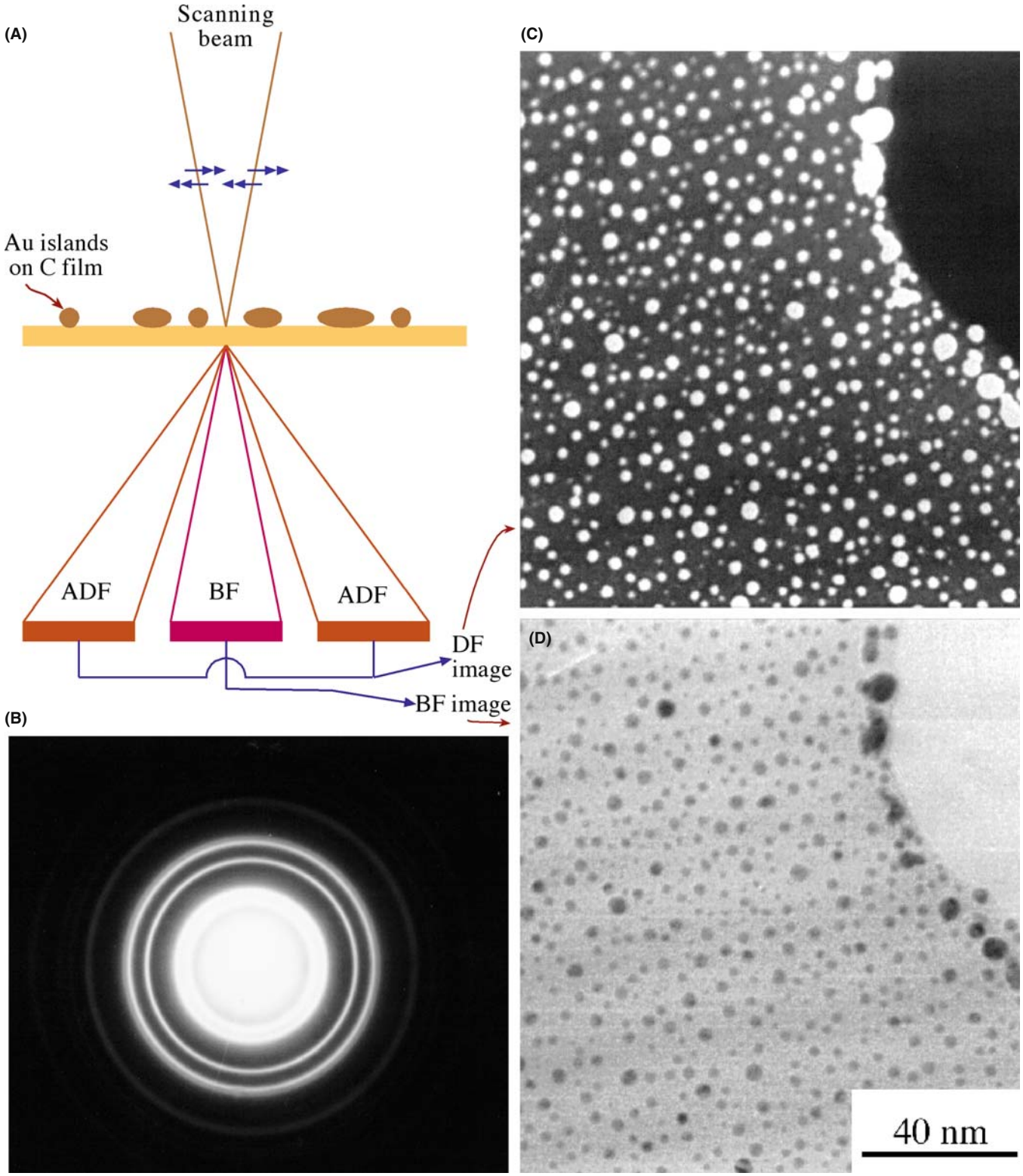


FIGURE 9.19. STEM image formation: A BF detector is placed in a conjugate plane to the BFP to intercept the direct beam (A) and a concentric annular DF detector intercepts the diffracted electrons whose distribution is shown in the SAD pattern in (B). The signals from either detector are amplified and modulate the STEM computer display. The specimen (Au islands on a C film) gives complementary ADF (C) and BF (D) images.

Now in a TEM we can't physically put the detector in the BFP of the objective lens to form a STEM image, because it would interfere with the objective aperture. Therefore, we usually insert the detector into a conjugate plane to the stationary DP (Figure 9.19B). So when you form a STEM image in a TEM, you operate the TEM in diffraction mode and insert a detector into the viewing chamber of the TEM, either above or below (in which case you raise) the screen. The stationary DP falls on the detector and the signal goes to the display. In a DSTEM, there may not be any imaging-system (or post-specimen) lenses, in which case the detector is indeed positioned immediately after the objective lens. Much of what we've just said is automatically done when you 'hit the STEM button'. The message is the same: understand what is happening in your microscope and why.

9.4.B Dark-Field STEM Images

The approach is analogous to that of TEM. We form a DF image by selecting any or all of the scattered electrons, rather than the direct-beam electrons. Remember, in a TEM we tilt the incident beam so the scattered electrons we want to form the image travel down the optic axis and are selected by the objective aperture. In a STEM, we do things rather differently.

DF STEM

If we want a specific beam of scattered electrons to fall on the BF detector, we can simply shift the stationary DP so that the scattered beam is on the optic axis and hits the BF detector.

It's simple to do this with the DP centering controls or you could also displace the C2 aperture. The former is to be preferred since doing the latter misaligns the illumination system.

9.4.C Annular Dark-Field Images

Rather than using the BF detector for DF imaging, we usually use an annular detector, which surrounds the BF detector, and then all the scattered electrons fall onto that detector. We call this process annular dark-field (ADF) imaging and it has certain advantages, depending on the contrast mechanism operating in the specimen, as we'll see in Chapter 22. As we show in Figure 9.19A, the ADF detector is centered on the optic axis and has a hole in the middle, within which the BF detector sits. The resultant ADF image in this simple example (Figure 9.19C) is complementary to the BF image (Figure 9.19D). As you'll see in Chapter 22, we can also use another annular detector that sits around the ADF and picks up the electrons scattered out to even higher angles forming so-called high-angle (HA) ADF (or Z-contrast) images in which Rutherford-scattering effects are maximized and diffraction-contrast effects are smoothed out.

We can take this idea further and make detectors of any size or shape we wish. For example, we could design a detector in which the annulus is split into two halves or four quadrants and electrically isolate each part of the detector. Then we can form different images from electrons that fall on different parts of the detector. It's impossible to do this in a TEM, because the objective aperture that does the selecting is a hole and can't be cut up like a semiconductor detector. We'll talk more about these kinds of detectors when we discuss specific contrast mechanisms in TEM and STEM images in Chapter 22.

9.4.D Magnification in STEM

All the STEM images that we have just described appear on the computer screen at a magnification that is controlled by the scan dimensions on the specimen, *not* the lenses of the TEM. This is a fundamental difference between scanning and static image formation.

STEM IMAGES

STEM images are *NOT* magnified by lenses.

Because scanning images are not magnified by lenses, they are not affected by aberrations in the imaging lenses. They are, however, affected by aberrations of the probe itself and, therefore, can be improved by (illumination-system) aberration correctors.

If the scanned area on the specimen is 10 mm × 10 mm, and the resultant image is displayed on a computer screen with an area 100 mm by 100 mm, then the image magnification is 10×. If the scan dimension is reduced to 1 mm, the magnification on that same screen is 100×, and so on, up to magnifications in excess of 10^7×, which are common in dedicated STEMs. As with the TEM, we have to calibrate the STEM magnification and the camera length of the DP we use to create the images.

9.5 ALIGNMENT AND STIGMATION

9.5.A Lens Rotation Centers

You only need to perform two alignments to ensure that the imaging system is operating correctly. By far the most important is the alignment of the objective lens center of rotation and the second is the alignment of the DP on the optic axis. To get the best out of your TEM, you *must* master these two fundamental alignments.

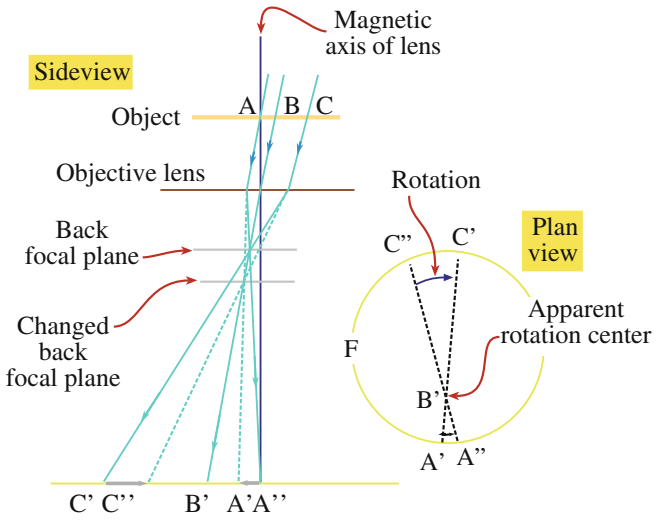


FIGURE 9.20. When the objective-lens center of rotation is misaligned, the image appears to rotate about a point away from the center of the viewing screen when the lens is wobbled about focus. When the rotation center is correctly aligned, the image will rotate about the center of the screen.

Basically, the idea of the objective-lens rotation alignment is to ensure that the objective lens field is centered around the optic axis, so that the direct-beam electrons emerging from the specimen see a symmetric field as they pass through the lens. If the field is off-center, then the electrons will move off axis, suffer more aberrations, and your image will rotate about a position off-axis as you change the objective lens (focus), as shown schematically in Figure 9.20.

Operational Procedure #7

- To center the objective rotation, start at a relatively low magnification (say 10,000×), select an obvious reference point in the image and move it to the middle of the screen, and observe the way the point rotates as you wobble the objective lens from overfocus to underfocus and back again. If the point rotates without moving off center, the lens is aligned, but check the accuracy of the alignment by going to much higher magnifications ($>10^5\times$).
- If the point moves off center at 10,000× then use the beam tilts to move the point in the image that *is* the center of rotation to the middle of the screen while continuously wobbling the objective. Repeat the process at higher magnifications.
- Above $\sim 10^5\times$ the wobbler may introduce too large a rotation, so you may have to defocus the objective lens manually. The actual steps to do this are instrument-dependent, so consult the manufacturer's handbook. This process is also called 'current centering.'
- When the image wobbles (rotates) about the center of the screen at magnifications $>10^5\times$, the objective lens rotation center is well aligned. The higher the magnification at which you can achieve this, the better the alignment and the better quality all your pictures will be.

This operation may be computer controlled in modern TEMs. In some instruments you can also perform 'voltage centering' in which a varying voltage is applied to the gun and the objective lens is aligned to ensure that the electrons remain on axis through the lens as their energy varies. Not all instruments are capable of this alignment.

If you change the DP magnification (i.e., the camera length, L) the whole pattern will move off axis if the diffraction center is misaligned. To align the center you have to adjust the projector lens until the central spot in the DP is on axis and it rotates around the axis as L is changed. Check the manufacturer's handbook on exactly how to do this alignment.

Centering the DP is useful for STEM-image formation, since you have to center the DP such that the direct beam hits the BF detector and the scattered beams hit the ADF detector. Apart from this simple operation, the STEM imaging system needs no lens alignment.

9.5.B Correction of Astigmatism in the Imaging Lenses

After you've centered the image and DP, the main cause of problems in the imaging system is astigmatism in the objective and intermediate lenses.

ASTIGMATISM

Objective-lens astigmatism occurs if the objective aperture is misaligned, so you must always carefully center the aperture on the optic axis, symmetrically around the electron beam used to form the BF or DF image.

Despite careful centering of the objective aperture, residual contamination may also cause astigmatism and then you have to use the objective stigmators to introduce a compensatory field. You'll find that the effects of objective astigmatism are harder to see than condenser astigmatism, which is easily visible on the screen as we just described. Correct objective astigmatism as follows:

Operational Procedure #8 Either look for a small hole in your specimen or look at a rough corner where the specimen edge curves through $> 90^\circ$. Ideally you might use a holey carbon film to correct residual astigmatism, *before* you insert your specimen, especially while learning this procedure, which we illustrate in Figure 9.21.

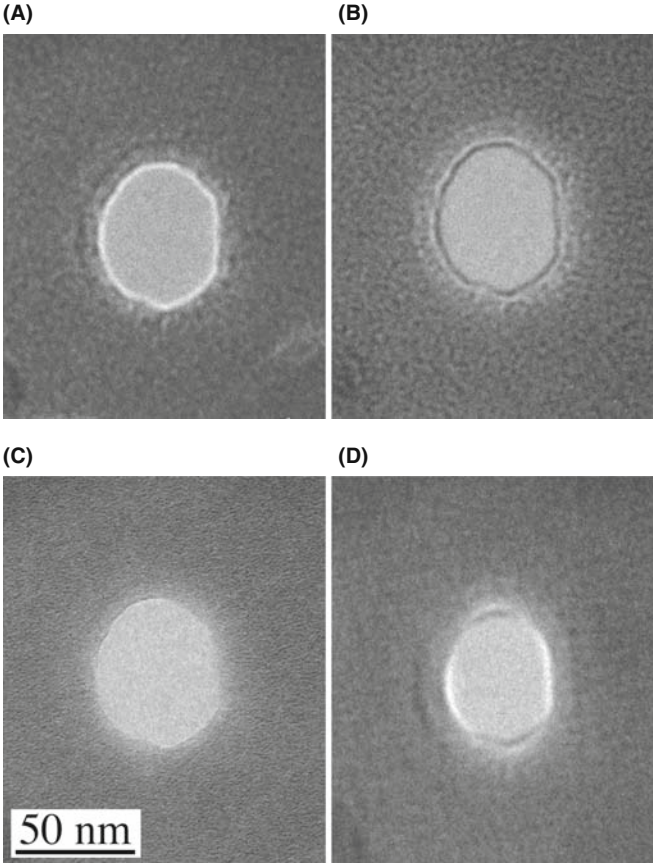


FIGURE 9.21. The image of a hole in an amorphous carbon film illuminated with a parallel beam showing that (A) with the beam underfocused, a bright Fresnel fringe is visible; (B) with the beam overfocused, a dark fringe is visible; (C) at exact focus there is no fringe; and (D) residual astigmatism distorts the fringe. Correcting the astigmatism means changing any image similar to (D) to one similar to (A) or (B).

- Often you can only see objective astigmatism at the highest magnifications so form a BF image of the hole or the specimen edge at high magnification ($>10^5\times$). First, adjust the illumination system to ensure a reasonably parallel beam. (Remember: how to do this depends on whether you have a c/o lens or not.)
- Now defocus the objective lens (either over- or underfocus; it doesn't matter). A Fresnel fringe (which is a phase-contrast effect, as we'll see in Chapter 27) should be visible at the thin edge of your specimen.
- Alternately underfocus and overfocus the objective lens (i.e., use the wobbler again).
- If there is astigmatism, you should see streaking in the image and this streaking effect will rotate through 90° as you wobble either side of exact focus. The streaking is most easily seen by watching the Fresnel fringe at the edge, as we describe below for Figure 9.21.
- Adjust the objective stigmators to compensate for the streaking at overfocus and then again at underfocus (until you're very skilled, you'll probably have to switch off the wobbler to do this and manually defocus the objective).
- Repeat these steps until there is no obvious image streaking as you defocus and the image merely blurs when going out of focus.
- Go to a higher magnification and repeat the procedure. As with the rotation center, the higher the magnification at which you can correct the astigmatism, the better and $>250{,}000\times$ should be possible. But remember that at higher magnifications the image intensity will be reduced. Don't try and compensate for this reduced intensity by condensing the beam because doing so will destroy the parallelism, thus reducing the Fresnel-fringe (phase) contrast.

As shown in Figure 9.21A, when you underfocus the objective lens, there is a bright Fresnel fringe round the edge of the hole. If this fringe is uniform around the hole, then there is no astigmatism. If the fringe varies in intensity, as in Figure 9.21D, then the focus of the lens is changing around the hole because of astigmatism. Then you have to adjust the objective stigmators to make the fringe uniform. The same operation must be repeated at overfocus, when there is a dark fringe around the edge of the hole (Figure 9.21B). At exact focus, you should see no fringe and the image contrast is minimized (Figure 9.21C).

FRESNEL FRINGES
Underfocus = bright fringe
Overfocus = dark fringe

This method of correcting the astigmatism is reasonable at magnifications up to several hundred thousand times.

In practice, if you're working at such high magnifications, you'll probably have to check the astigmatism throughout your TEM session, so you should get used to looking at the Fresnel fringes on a thin, curved edge of your specimen rather than the ideal holey carbon film.

For high-resolution imaging at magnifications of $>300{,}000\times$, we have to use the streaking in the image to correct for astigmatism. We'll talk about this when we discuss HRTEM in Chapter 28.

Intermediate-lens astigmatism is of secondary importance and only affects the DP. Because the DP is at zero magnification in the objective lens, the intermediate lenses are responsible for magnifying it. So if there is residual astigmatism in these lenses, then the DP will show orthogonal distortions as you take it through focus. This effect is small and can only be seen in the binoculars as you focus the DP with the diffraction

focus (intermediate lens) control. Make sure that the incident beam is strongly underfocused to give the sharpest spots. As with objective astigmatism in the image, simply adjust the intermediate stigmators to compensate for any spot distortion at underfocus, and overfocus until the spots expand and contract uniformly in all directions through focus. You should be aware that not all instruments have the requisite intermediate stigmators to carry out this correction.

9.6 CALIBRATING THE IMAGING SYSTEM

Your TEM should be calibrated when it is first installed and then periodically throughout its life, especially if you wish to carry out accurate measurements from images or DPs. If the instrument is modified substantially (e.g., a burned-out lens coil is replaced), then it must be recalibrated. In all cases you must specify a set of standard conditions under which the calibrations are carried out (e.g., objective lens current and other lens settings, eucentric height, etc.). You can find a full description of all the detailed concerns with all the TEM calibrations in Edington (1976). Remember, if you're serious about TEM don't rely on someone else's calibrations.

THE LAST USER

Since you usually will not be the first user, you should take the time to check the existing calibration. Don't assume it is correct.

9.6.A Magnification Calibration

We use standard specimens to calibrate the magnification. The most common specimen we use is a thin carbon-film replica of an optical-diffraction grating of known spacing, as shown in Figure 9.22A. The typical linear density of lines in the replica is 2160/mm (giving a line spacing of 0.463 μm with an error that will depend on how many grating spacings you can measure). Grating replicas enable calibration up to magnifications of ~200,000×. Above this magnification, individual grating spacings are wider than the recording film. So we can then use small latex spheres (50–100 nm diameter) although they are susceptible to beam damage and shrinkage under electron bombardment. At the highest magnifications, phase-contrast images of known crystal spacings, such as the 0002 spacing in the graphite structure (0.344 nm) can be used or the 111 fringes of Si. What we are doing is just using a known periodicity in the crystal; careful consideration of the objective-lens defocus and specimen thickness is required before the phase-contrast lattice image can be directly interpreted, as we discuss in detail in Chapters 27 and 28.

MAGNIFICATION

Magnification calibration is so sensitive to so many variables that some users deposit a standard material on the material they are studying so that the calibration will be done under exactly the same conditions and will appear in the same image as your area of interest.

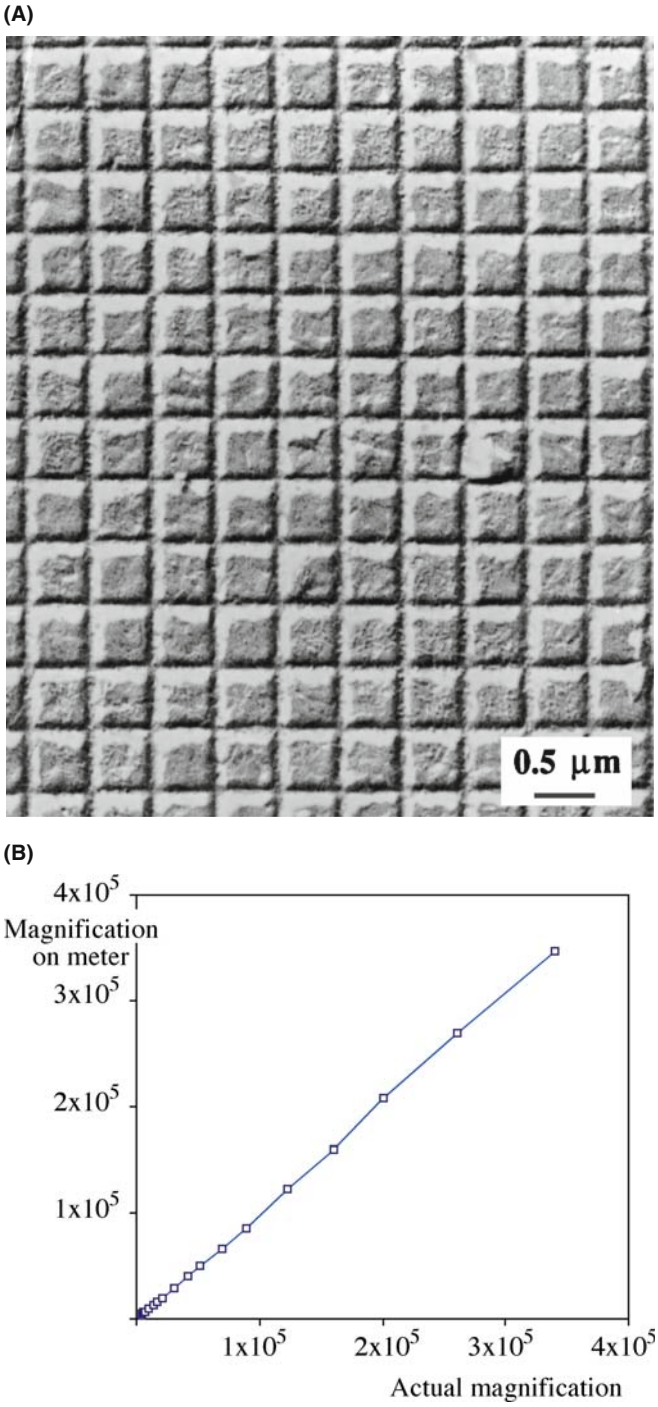


FIGURE 9.22. (A) An image of a diffraction-grating replica for which the actual spacing of the grating is known. (B) The TEM magnification can thus be calibrated, permitting specific magnification settings (usually with approximate/calculated magnifications) to be assigned more accurate, experimental magnifications. If the plot for your TEM is not linear then there are serious problems either with your measurements or with the microscope.

Operational Procedure #9

- Insert the calibration specimen and ensure the holder is at the eucentric height.
- Ensure the illumination is parallel and the objective aperture is well centered on the 000 spot in the SADP.
- Switch back to image mode and then focus the BF image (see below for the best way to be consistent in focusing).
- Record images of the diffraction grating at all magnification settings, taking care to re-check the focus after each change of magnification.
- Calculate the magnifications experimentally from the images knowing the true spacing.

Figure 9.22B shows the classical magnification calibration for a Philips CM30 TEM. Today, you'll use a crystal lattice at ~300 k×, relate it to the same area at say 50 k×, repeat at 10 k×, and then at all magnifications in between.

HYSTERESIS

Electromagnetic lenses have hysteresis. Everything you do must be reproducible if you want your results to be.

You have to calibrate the magnification because the TEM imaging system does not give stable and reproducible lens strengths. The lens strengths will change with ambient temperature (e.g., how many people are in the room), with the efficiency of the cooling system of the lenses and with lens hysteresis. Therefore, if you want to make accurate measurements from your TEM images, you must carry out the magnification calibration *at the time you make the measurements*. In particular, you have to minimize the lens hysteresis by always approaching image focus consistently from overfocus or underfocus and/or reversing the lens polarity several times before finally coming to focus. Also, you must remember that there may be barrel or pincushion distortions in the image particularly at very low magnification (< 5000×) where the lens is not designed to work well. (As students you ought to be able to work out what the first distortion does to your image, but you might not be familiar with the latter distortion at this early stage of your life!)

THE MAGNIFICATION

The image size will change during recording, printing and publishing. *Never* state an image magnification in the caption. Indicate a distance on the recorded image that corresponds to a distance on the actual specimen at the time the image was recorded.

Remember that the area of the specimen you are working on must be at exactly the right 'height' in the column. Because of the magnification error, the TEM is not the best instrument for absolute measurement of particle sizes, etc. However, relative measurement is easily done with reasonable accuracy (±5%), so long as you note the precautions we've just described. Without a calibration, the digital readout is probably no better than ±10% accurate, and so it is unwise to state magnifications to better than ±10%. You should be suspicious of any micrographs that you see in the literature with a magnification that is more precise than three (or even two) significant figures such as 52,550×. It may indicate that the microscopist does not understand the limitations of their instrument and the work should be interpreted with due caution.

We used to call these lines 'micron bars' for obvious reasons but TEM magnifications have become larger and 'nano bars' would often be more appropriate (even though neither one is related to barrel distortion). All modern TEMs automatically put such a marker on the negative. However, its accuracy is only as good as your operational skills.

You can use an identical procedure to calibrate the STEM-display magnification despite the fact that the digital STEM-image magnification is, in principle, easily calculated from the scan-coil strengths. The image magnification differs from the digital readout because of variations in the objective lens. Table 9.1 shows the difference between a typical digital STEM magnification and the experimentally determined magnification using a diffraction-grating replica.

9.6.B Camera-Length Calibration

As mentioned already, we describe the magnification of the DP by the camera length (L), a term that arises from X-ray projection-diffraction cameras which operate without lenses (because focusing X-rays is very difficult). In such cameras, magnification is increased by moving the recording film further away from the specimen, in exactly the same way that you magnify your computer-slide presentations onto a screen by moving the projector further away. (You can always tell non-microscopists who try to *focus* their slides in the same way, or those

TABLE 9.1. Magnification Calibration for a Philips EM400T in STEM Mode at 120 kV

Digital readout	Calculated magnification
3,200	3,420
6,400	6,850
12,500	12,960
25,000	27,000
50,000	54,000
100,000	108,000

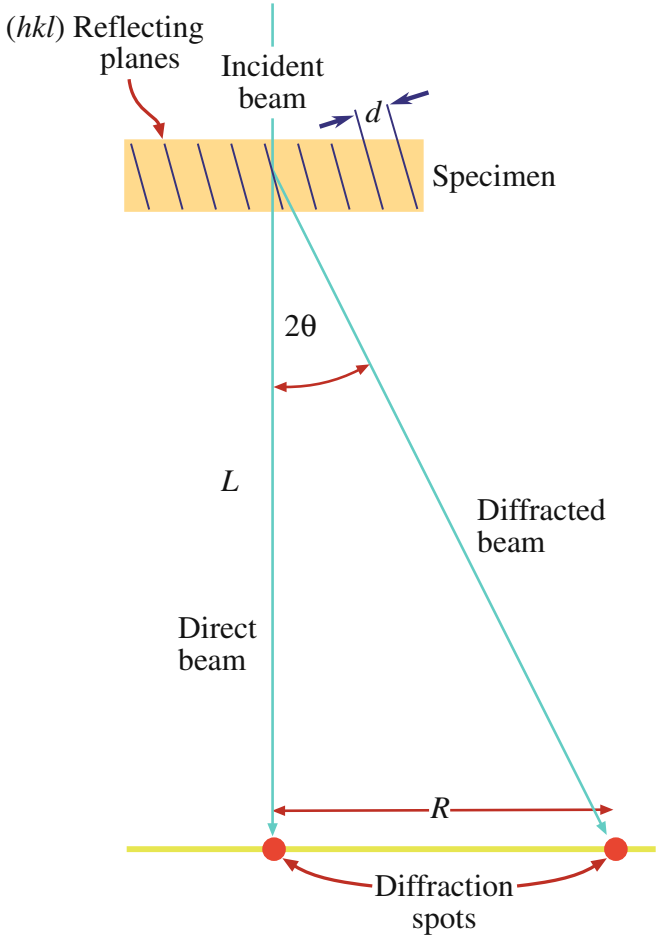


FIGURE 9.23. The spacing R between the direct beam and a scattering maximum (such as a diffracted beam or the radius of a ring of diffracted intensity) is related to the camera length, L. Increased magnification corresponds to effectively increasing L, although in practice, this is accomplished with lenses.

who try to change the magnification by adjusting the projector lens focus!)

The camera-length concept can be applied in the TEM as shown in Figure 9.23. This figure represents the imaging system, but without the lenses drawn in.

CAMERA LENGTH

If we increase the magnification of the lenses between the specimen and the viewing screen, we increase the effective distance L between the specimen and the screen.

The camera length is a calculated value rather than a physical distance. If electrons are scattered through an angle 2θ at the specimen (as in a typical diffraction event described in Chapters 2 and 11), then the distance between the direct and diffracted beams as measured on the screen (R) is determined by L, since

$$\frac{R}{L} = \tan 2\theta \sim \theta \qquad (9.1)$$

From the Bragg equation (equation 3.21) we know that $\lambda/d = 2\sin\theta \sim 2\theta$ and so we can write

$$Rd = \lambda L \qquad (9.2)$$

Operational Procedure #10 To calibrate the magnification of the DP we need to record DPs from a specimen with a known crystal spacing (d) such as a thin film of polycrystalline Au or Al which gives a ring pattern (see Figure 2.11).

- Make sure your specimen is at the eucentric height and the BF image is in focus.
- Insert the SAD aperture and switch to diffraction mode.
- Ensure the beam is parallel so that the DP spots or rings can be focused to the smallest size. If the pattern is not focused then focus with the first intermediate lens (diffraction focus) control. Be careful if you have a c/o lens because changing the illumination conditions (e.g., trying to make the beam more parallel) will change the DP focus. If you then sharpen the DP by changing the diffraction focus, the camera length will change. If in doubt, consult the manufacturer's handbook.
- Since we know the lattice parameter of the metal, we can measure the ring radius R on the photographic film or computer display for any plane that is diffracting (see Chapter 18 to find out exactly how we do this).
- Since we know λ we can easily determine L from equation 9.2.

A typical TEM camera-length calibration is shown in Table 9.2. The STEM camera-length calibration may be different to the TEM if the objective lens setting is not exactly the same in TEM and STEM modes, and this depends on the vintage and make of your instrument. So you should check with the manufacturer before taking the time to perform the calibration.

TABLE 9.2. Comparison of Experimentally Measured Camera Length (and Camera Constant) with the Digital Readout for a Philips EM400T Operating at 120 kV (λ = 0.0335 Å)

Camera length setting	Digital readout (mm)	Measured camera length, L (mm)	Camera constant λL (mm Å)
1	150	270	9.04
2	210	283	9.47
3	290	365	12.22
4	400	482	16.14
5	575	546	18.28
6	800	779	26.08
7	1150	1084	36.29
8	1600	1530	51.22
9	2300	2180	72.99
10	3200	3411	114.20

It's worth explaining in more detail the effect of a c/o lens on the SAD calibration. In a c/o lens, as shown back in Figure 9.5, the beam can go from divergent to convergent conditions at the specimen plane. However, unless the beam is parallel, your DP focus will change with illumination angle. In this case, the diffraction-lens setting at the point of focus of your DP will differ for different illumination conditions. So L will change (by as much as 15%) depending on the angle of convergence/divergence. In order to ensure the same L, you must pick one condition and use it for every calibrated DP. There are several ways to do this.

- Set the diffraction lens to the same excitation for each pattern and focus the DP with the C2 lens.
- Focus the pattern at the objective aperture (especially if the BFP of the objective and the objective aperture share the same position).
- Go to the maximum value of the C2 lens, focus the DP, and use the C1 lens to adjust the brightness of the DP.

There are other possibilities which you'll be able to work out as you get more skilled but, above all, be consistent and always *use the conditions you chose when you did the calibration*.

9.6.C Rotation of the Image Relative to the DP

Anyone studying crystalline materials must determine the angle between directions in the image and directions in the DP. At a fixed camera length, the DP always appears on the screen in a fixed orientation. But if you record images at different magnifications, the images will rotate by an angle ϕ with respect to the fixed DP. (In some TEMs this rotation has been removed by the addition of a compensating projector lens and, in this case, there is always a fixed rotation (ideally 0°) between directions common to both image and DP.)

THE GOLDEN RULE

Always do the calibrations yourself. Do not rely on factory calibrations. The conditions you use in your laboratory may differ from those chosen by the manufacturer.

Operational Procedure #11 To determine this rotation, we often use a specimen of α-MoO_3, because it forms thin, asymmetric crystals with a long edge known to be parallel to the 001 direction in the crystal.

- Take care to ensure that, as usual, the image is focused with the specimen at the eucentric plane.
- Insert the SAD aperture and ensure that it is focused (using the intermediate lenses) to coincide with the image plane.
- Switch to diffraction mode with the beam parallel and adjust the diffraction focus to give sharp diffraction maxima.
- Take a double exposure of the DP and the image as shown in Figure 9.24A.
- Repeat the whole exercise for different magnifications and plot out the variation of the angle ϕ as shown in Figure 9.24B.
- You can do the same, if necessary, for different values of L, which introduce a systematic change in ϕ. For this reason, we recommend that you carry out all your SAD work at a standard value of L; 500–1000 mm is usually optimum.

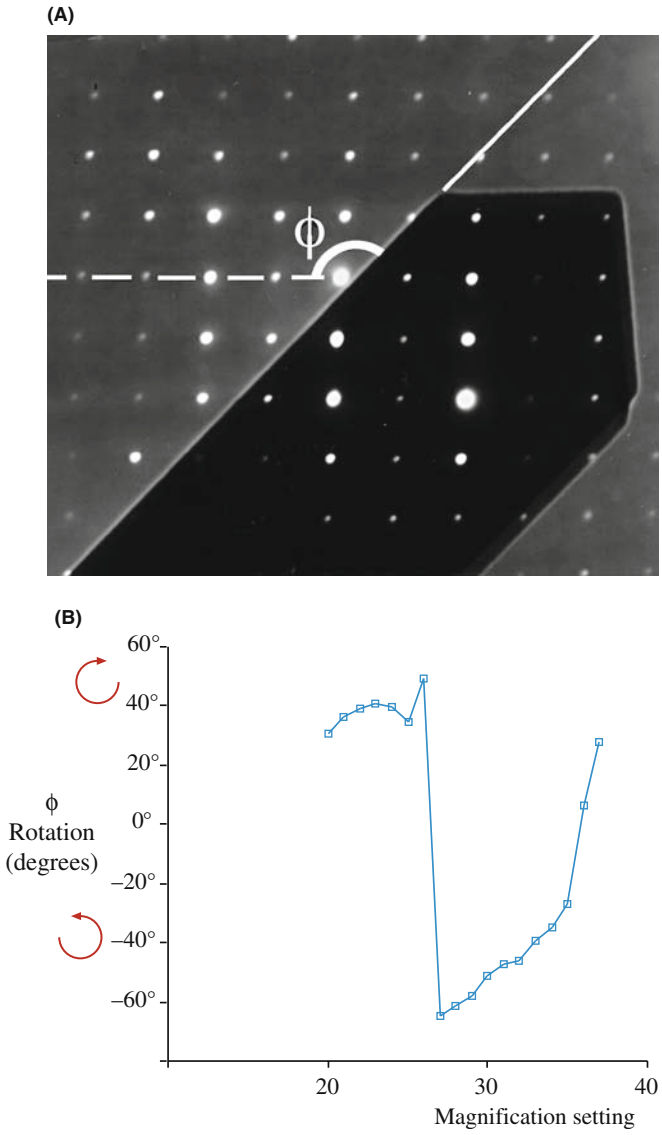


FIGURE 9.24. (A) A double exposure showing the superposition of an image of an α-MoO_3 crystal on a DP from the same crystal, defining the rotation angle ϕ. (B) The rotation calibration gives the angle ϕ between equivalent directions in the image and the DP as the magnification is varied. The calibration assumes a constant camera length. The rotation angle will change significantly if the imaging system switches lenses on or off as the magnification is changed (e.g., between settings 26 and 27).

A further complicating factor is that, as the image magnification is increased, the TEM lens control logic may switch off, or switch on, one of the imaging-system lenses. When this happens, a 180° inversion is introduced into the image. You can see this happen if you watch the image carefully as you change the magnification. This inversion has to be included in the rotation calibration, otherwise a 180° error will be made in the assignment of directions in the image. One way to see if the image has a 180° inversion is to look at the DP and defocus it slightly so the BF image in the direct beam can be seen directly at very low magnification. The 180° inversion is immediately obvious, as shown in Figure 9.25. As we already noted, some manufacturers adjust the imaging-system lenses to compensate for the rotation, thus keeping a constant image-DP rotation angle at all magnifications. Likewise, in STEM mode, since the imaging lenses are not used for magnification, there is a fixed angle between the images and DPs.

McCaffrey and Baribeau used molecular-beam epitaxy to create a cross section TEM specimen made from a single-crystal Si/Ge wafer specimen that can perform these three major calibrations (magnification, camera-length, and image/DP rotation). Similar specimens are available from suppliers of TEM consumables (see URLs 1–3, for example). These suppliers also offer a range of standard specimens for other TEM needs (e.g., XEDS and EELS analysis). If you are serious about your microscopy, buy your own standards, keep them safe and clean, and use them often.

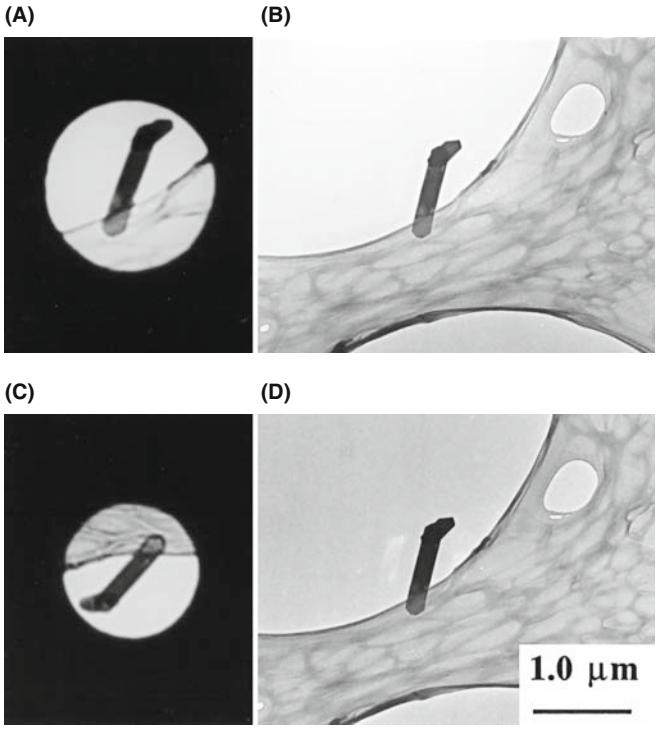


FIGURE 9.25. Defocused direct beam in a DP from α-MoO_3 compared with a BF image, showing how to determine if a 180° inversion exists or not. If the image of the specimen inside the expanded image of the beam is rotated with respect to the image on the screen, as in (C) and (D), then a 180° inversion is required to determine the correct angle between directions in the DP and directions in the image. In (A) and (B), no rotation occurs between the DP and BF image.

9.6.D Spatial Relationship Between Images and DPs

If you don't use a double exposure when comparing images and DPs (or indeed when comparing directions in any two films) you need a fixed reference line. This line must be independent of slight variations that may arise depending on the film size, how you loaded it, etc. The best reference line is the edge of the plate-numbering system that is superimposed on each film.

Yesterday's advice:
Whenever you're comparing images and DPs, it is essential to compare the photographic negatives with the *emulsion side up*. This is contrary to usual photographic practice, but it's necessary to preserve the relationship between manipulations of your specimen and what you see happening on the screen. *If you don't do this,* it is easy to introduce a 180° error into the relationships between images and DPs.
Today's advice:
Don't use film. If you use a CCD, you'll never think of inverting the image!

9.7 OTHER CALIBRATIONS

The accelerating voltage: The selected voltage may differ from the absolute voltage by detectable amounts. There are several ways to determine the actual voltage: First, you can measure the electron wavelength, λ, by measuring the angle, ϕ, between two Kikuchi line pairs (see Chapter 19) that intersect a distance R from the direct beam

$$\tan\phi = \frac{R}{L} = \frac{\lambda}{d} \tag{9.3}$$

Alternatively, you can match simulations of HOLZ lines to experimental lines in CBED patterns (see Chapter 21) and determine which λ gives the best match. Finally, in principle, you could get your XEDS computer system to display the X-ray spectrum (Chapter 34) out to E_0, the beam energy, and the bremsstrahlung intensity vanishes to zero at the exact beam energy (this is called the Duane-Hunt limit). However, XEDS computer displays rarely show energies > 30–40 keV.

The specimen tilt axis and the sense of tilt: In a side-entry stage, the principal tilt axis is parallel to the specimen-holder rod. Since the image is usually rotated on the screen relative to the specimen, how can you locate this tilt-axis direction? Well it's easy if you can move the specimen in a known direction. From this movement, you can determine the direction of the tilt axis for a specimen of known geometry.

Yesterday's advice:
If you gently push on the end of your side-entry specimen holder, the image moves parallel to the principal tilt axis. But don't pull the holder in the other direction since you may inadvertently cause air to leak into the airlock.
Today's advice:
We're not letting you remove the draft-excluder (which covers the end of your specimen holder) to touch the holder after it has equilibrated. Monitor how you load the specimen instead which takes care and practice—but so do most TEM operations.

If you are looking at the DP, defocus the pattern so you can see the BF image in the direct-beam spot, as in Figure 9.26, then carry out the same exercise. If you are using a top-entry holder you will need to calibrate this tilt using a known specimen geometry, which is more challenging.

Focal increments of the objective lens: If you're going to do high-resolution, phase-contrast imaging and you don't have a C_s corrector, then you need to know the value of each defocus step of the objective lens so you can correctly calculate and interpret the image contrast. There is a simple method for determining this step value. Superimpose a focused image and an image defocused by a known number of objective-lens focal increments (Δf). The two images will be separated by a distance Δx which is related to Δf by

$$\Delta f = \frac{\Delta x}{2\,Mm\theta} \tag{9.4}$$

where M is the magnification, m is the number of focal increments, and θ is the Bragg angle for the reflection used to form the image. If you insert some typical values, you'll find that it is difficult to be very accurate with this method. We'll return to this topic in Chapter 28.

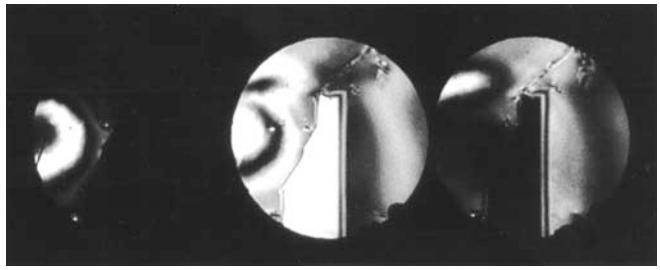


FIGURE 9.26. Defocused multiple-DF image showing how it is possible to determine simultaneously the direction of features in the image (e.g., the vertical twin boundary) and directions in the DP (e.g., the horizontal vector between the diffraction disks). If the specimen holder is moved in the direction of the principal tilt axis, the image will move and identify the relationship between that tilt axis and the crystallographic direction in the DP.

CHAPTER SUMMARY

We've now shown you how a TEM is put together. While the manufacturer does the best job possible, there are still some essential steps for you, the operator, to carry out. You must understand how to align the illumination system so the beam is on axis. You can then create a parallel beam for TEM and a convergent one for STEM. The C2 aperture is a crucial part of the whole illumination system and the most easily misaligned. Astigmatism is not too much of a problem if the instrument in general and the apertures in particular are kept clean. The objective lens/stage combination controls all the useful information that is created as the beam is scattered by your specimen. *Always* start a microscope session by fixing the eucentric height and, before you do any worthwhile imaging, align the objective center of rotation, and minimize any objective astigmatism at high magnification. Diffraction and STEM operations require a centered, focused DP.

If you want to make any quantitative measurements from your images and DPs (and you really ought to do this if you have any aspirations to be a real microscopist), then calibration cannot be avoided. Your images and DPs are relatively useless unless you know their magnification and camera length, respectively, and the angular relationship between the two. So take the time to do this early in your studies. In doing so you will not only ensure that you produce quality data, but you will also learn an enormous amount about how these complex instruments work.

You'll notice that some of the data (e.g., Figure 9.12B and Table 9.1) are for older TEMs. It would be better to use data from even older machines ("Why?" you should ask) but you are unlikely to see such machines in operation.

SOME HISTORY

SAD was invented by LePoole in 1947. Some early machines used a pair of blades with a 90° bend that slid past one another to give a square aperture! See, e.g., LePoole, JB 1947 *A New Electron Microscope with Continuously Variable Magnification* Philips Tech. Rept. **9**(2) 33–45. You can find further discussion of TEM calibration in Edington's text.

OPERATING THE TEM

Chapman, SK 1980 *Understanding and Optimizing Electron Microscope Performance 1. Transmission Microscopy* Science Reviews Ltd. London. An old book but the principles are so clearly outlined that, if you are a serious user, this must be in your laboratory.

Chapman, SK 1986 *Maintaining and Monitoring the TEM* Royal Microscopical Society Handbook No. 8 Oxford University Press New York. If you can't find this on the shelf in the office of the EM technical staff, be sure to buy it for them for Christmas. Your small investment will be repaid many times over.

Chescoe, D and Goodhew, PJ 1990 *The Operation of Transmission and Scanning Microscopes* Royal Microscopical Society Handbook No. 20 Oxford University Press New York. A great in-depth description of the operational principles described briefly in this chapter.

Christenson, KK and Eades, JA 1988 *Skew Thoughts on Parallelism*, Ultramicroscopy **26** 113–132.

Edington, JW 1976 *Practical Electron Microscopy in Materials Science* Van Nostrand Reinhold New York. Now very dated, but still a valuable text for the details of hands-on, diffraction-contrast imaging and SAD; it's especially good on calibrations. However, this text was written before c/o lenses were introduced which, as we've noted, introduce complications into hitherto simple operations like forming parallel beams.

Goldstein, JI, Newbury, DE, Joy, DC, Lyman, CE, Echlin, P, Lifshin, E, Sawyer, LC and Michael, JR 2003 *Scanning Electron Microscopy and X-ray Microanalysis* 3rd Ed, Kluwer New York. Dealing with a virtual C2 aperture, scanning imaging and much more.

Keyse, RJ, Garratt-Reed, AJ, Goodhew, PJ and Lorimer, GW 1997 *Introduction to Scanning Transmission Electron Microscopy* Royal Microscopical Society Handbook No. 39 Bios Scientific Publishers Oxford. Basic introductory level but the only book available that emphasizes the operation of dedicated STEMs which historically were made by one company (VG) but are now available through at least three manufacturers (Hitachi, JEOL, and Nion). The DSTEM may well come to dominate the semiconductor industry in the way the SEMs did in the 1980s. If so, these instruments will become fixtures in any self-respecting EM Center. If, by now, you are getting the idea that the RMS handbook series is useful then you are right (see http://www.rms.org.uk/other-publications.shtml).

McCaffrey, JP and Baribeau, JM 1995 *A Transmission Electron Microscope (TEM) Calibration Standard Sample for All Magnification, Camera Constant, and Image/Diffraction Pattern Rotation Calibrations* Microsc. Res. Tech. **32** 449–454.

Watt, IM, 1997 *The Principles and Practice of Electron Microscopy* 2nd Ed. Cambridge University Press New York. Basic, hands-on, easy-to-read book, full of instructive images and diagrams, and easily understood explanations.

URLs

1) http://www.emsdiasum.com/microscopy/products/catalog.aspx
2) http://www.tedpella.com/calibrat_html/TEM6.htm
3) http://www.2spi.com/catalog/stand.html

COMPANION TEXT

C_c correction will be discussed in the chapters on HRTEM and EELS. The C_c monochromator takes on several forms but each decreases the total beam current.

SELF-ASSESSMENT QUESTIONS

Q9.1 Name the three principal components of the TEM.
Q9.2 How is the beam tilted or translated?
Q9.3 How do you form a parallel beam; why would you want to do this and why is it not exactly parallel?
Q9.4 How do you form a convergent beam; why would you want to do this and why is it sometimes divergent?
Q9.5 How do you align the gun and why should you bother?
Q9.6 How can you tell when your C2 aperture is misaligned and how would you correct it?
Q9.7 How does the probe size change with the C1 setting and why might you want to change the size?

Q9.8 What is SAD and how is it different from CBED?
Q9.9 What is the difference between a BF image and a DF image?
Q9.10 What is the advantage of forming images using STEM rather then TEM?
Q9.11 What causes lens astigmatism? How do you correct for it (a) in the illumination and (b) in the imaging system?
Q9.12 Why does the objective lens rotation need to be calibrated?
Q9.13 Why would you want to form hollow-cone DPs and images?
Q9.14 Explain CDF, DADF, HCDF, SADP, SDP, HCDP, FFP, BFP? If you can't find all these acronyms/initials in the text, suggest what they mean.
Q9.15 What is the most important aperture in the TEM and why?
Q9.16 How do you change the image magnification in a TEM?
Q9.17 How do you change the image magnification in STEM? Explain why this method confers advantages over the answer to the previous question.
Q9.18 Why do we need to do the magnification calibration?
Q9.19 What are the reasons you might need to translate the beam?
Q9.20 How should you always start a microscope session (apart from quiet prayer)?
Q9.21 How should you always finish a microscope session (apart from a loud prayer)?

TEXT-SPECIFIC QUESTIONS

T9.1 To converge the beam, which lenses do you use and how does this differ depending on the electron source? (Hint: go back to Chapter 5 and also look at Figures 9.3–9.5.)
T9.2 Explain the differences between the various methods used to create a parallel beam at the specimen plane in Figures 9.1 and 9.2.
T9.3 A smaller probe has a smaller current in it. Use Figure 9.4 to explain why this is so and state some consequences of reducing the probe current.
T9.4 Why use SAD, instead of just forming DPs without a selecting aperture?
T9.5 What illumination-system operation is essential to form Fresnel fringes for astigmatism correction?
T9.6 What is the most common problem with TEM illumination systems (apart from the previous user of the microscope)?
T9.7 How can you tell when your specimen is in the eucentric plane and why do you want it to be there?
T9.8 Typically, how accurate is the magnification of images of particles in TEM? What do you have to do to get better accuracy?
T9.9 List the operational steps necessary to see a SADP if you have a BF image on the TEM screen.
T9.10 List the operational steps necessary to see a CDF image if you have a DP on the TEM screen.
T9.11 Explain the principle shown in Figure 9.13 of using a virtual rather than a real aperture in the specimen plane for selected-area diffraction.
T9.12 Use Figures 9.16 and 9.19 to explain why we can't change the direction of the incident beam in STEM imaging.
T9.13 Why might we say that STEM images don't suffer from the defects in the imaging lenses and would we be correct?
T9.14 Is Figure 9.21 taken with a thermionic or FEG source? Explain your answer.
T9.15 Discuss the pros and cons of a c/o lens system using Figure 9.5 as a basis for your discussion. Justify why every materials-research TEM must have a c/o system.
T9.16 Distinguish traversing and tilting a beam and, using specific figures in the text, explain how these operations can be used for DADF operation, hollow-cone imaging and diffraction, stationary-DP formation, and scanning-image formation.
T9.17 When might you want to use a wobbler? Sketch a ray diagram explaining how a wobbler works. Sketch an image of how an astigmatic beam changes as it is wobbled and how a misaligned beam changes as it is wobbled.
T9.18 Use the references to find an equation governing the relationship between probe current and probe size.
T9.19 Estimate the relationship between the final aperture radius in the illumination system and the convergence angle of the beam and explain how this is relevant to Figure 9.11.
T9.20 Why does converging the beam expand the spots in the DP and why is there different contrast information (images) in different expanded spots, as in Figure 9.26?
T9.21 Why is it important to calibrate the rotation of your DP with respect to your image?
T9.22 Estimate the accuracy of the magnification and rotation calibrations.
T9.23 Look up the lattice parameters of α-MoO_3, make some reasonable assumptions about the TEM, and calculate an approximate camera length for the DP in Figure 9.24.

10

Specimen Preparation

CHAPTER PREVIEW

Specimen preparation is a very broad subject; there are books devoted to this topic alone. The intention here is to summarize the techniques, suggest routes that you might follow, and above all to emphasize that there are many ways to produce a TEM specimen; the one you choose will depend on the information you need, time constraints, availability of equipment, your skill, and the material. So we'll concentrate on the 'principles of cooking,' but won't try to list all the possible 'recipes.' One important point to bear in mind is that your technique must not affect what you see or measure, or if it does, then you must know how. Specimen preparation artifacts may be interesting but they are not usually what you want to study. Incidentally, we'll make 'specimens' from the 'sample' we're investigating so we'll look at 'TEM specimens,' but sometimes we, and everyone else, will interchange the two words.

The TEM specimen, when you've made it, must be electron transparent (usually) and representative of the material you want to study. In most cases (but not all) you would like your specimen to be uniformly thin, stable under the electron beam and in the laboratory environment, conducting, and non-magnetic (we'll discuss some exceptions as we proceed). Few specimens approach the ideal and usually you have to compromise. In general we can divide specimens into two groups: self-supporting specimens and specimens resting on a support grid or thin washer; the grid or washer is usually Cu but could be Au, Ni, Be, C, Pt, etc. Before discussing these two groups we will briefly review the most important part of specimen preparation, namely, safety. You may damage the microscope later, but this is the stage where you could do much worse to yourself and your colleagues.

It is often assumed that preparation of the TEM specimen will take several hours. Actually this time could be as short as 5 minutes or as long as 2 days even for the same material. For example, as you'll see, if you want to examine a piece of $YBa_2Cu_3O_{6+x}$, the high-temperature superconductor, you could crush the sample in a pestle and mortar using a nonaqueous solvent, catch the small particles on a carbon film, and put the specimen in the TEM; time required is about 10 minutes. Alternatively, you might cut the sample into thin slices using a diamond saw, cut 3-mm-diameter disks from the slice, thin the disk on a grinding wheel, dimple the thinned disk, then ion mill to electron transparency at liquid-nitrogen temperatures, carefully warm the specimen to room temperature in a dry environment, and put it in the TEM; time required is 1 or 2 days. Which method you choose would depend on what you want to learn about your material.

10.1 SAFETY

Either the specimen itself or the best method for preparing it for viewing in the TEM may require extreme care. Even materials which are safe and relatively inert in bulk form may be hazardous in powder form. Four favorite (because they work so well) liquids for polishing solutions are hydrogen cyanide, hydrofluoric acid, nitric acid, and perchloric acid. These liquids may be poisonous, corrosive (HF quickly penetrates the body and then dissolves the bone), or explosive (perchloric acid and nitric acid when mixed with certain organic solvents). It is clearly essential that you check with your laboratory manager, the reference texts, and the appropriate materials safety data sheets (MSDS) before you begin specimen preparation. This checking might also save a lot of time.

SAFETY FIRST

This whole section should, of course, be in a big red box. Some of the chemicals we use are really dangerous. We remember HF, perchloric acid, and HCN all being used concurrently in one small specimen-prep room.

In spite of these restrictions you may still need/want to use these acids and acid/solvent mixtures. The ion thinner may not be available or you may not be able to accept the damage that ions produce. In this event there are five brief points that you should bear in mind.

- Be sure that you can safely dispose of the waste product *before* you start.
- Be sure you have the 'antidote' at hand.
- Never work alone in the specimen-preparation laboratory. Always wear safety glasses when preparing specimens and/or full protective clothing, including face masks and gloves, if so advised by the safety manual.
- Only make up enough of the solution for the one polishing session. Never use a mouth pipette for measuring any component of the solution. Dispose of the solution after use.
- Always work in a fume hood when using chemicals. Check that the extraction rate of the hood is sufficient for the chemical used.

Since these four acids can be so dangerous, we'll mention them specifically, but remember—always seek advice before chemically preparing specimens.

Cyanide solutions: If possible, avoid this solution even though you may see it in the textbooks. The only metal where it really excels is gold and you can thin this by very careful ion milling.

Perchloric acid in ethanol or methanol: If you have to use this 'universal polish' you should be aware that many laboratories require that you use a special dedicated hood which can be completely washed down since crystallized perchloric acid is explosive. The phase diagram in Figure 10.1 for the perchloric-acetic (acid)-water system makes the message clear. If you have to use perchloric-acetic acid mixtures or indeed when using any perchloric-containing mixtures, keep the density below 1.48. If you are very careful, if you *always* add the acid to the solvent, and you make sure that the liquid *never* becomes warm, then perchloric acid solutions can be used to produce excellent TEM specimens of Al, stainless steel, and many other metals and alloys.

Nitric acid: In combination with ethanol, this acid can produce explosive mixtures, especially if left for long periods of time and exposed to sunlight. It is preferable to use methanol rather than ethanol, but in either case, keep the mixture cool and dispose of it properly.

HF: This acid is widely used in the semiconductor industry and in 'frosting' light bulbs; the reason in both cases is that it dissolves SiO_2 leaving no residue. Careful use of dilute solutions can produce specimens that have large thin areas. Remember: if you use HF, completely cover any exposed skin; HF rapidly penetrates the flesh and dissolves bone and you won't even feel it!

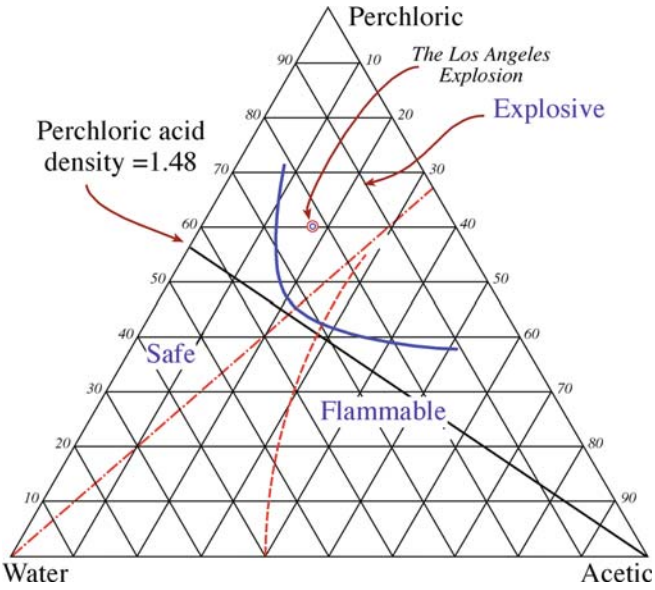


FIGURE 10.1. Perchloric-acetic-water phase diagram showing the hazardous regions and the recommended density line for safe use of all perchloric solutions. Always operate to the left of this line.

10.2 SELF-SUPPORTING DISK OR USE A GRID?

The type of TEM specimen you prepare depends on what you are looking for so you need to think about the experiment that you are going to do *before* you start thinning. For example, is mechanical damage to be avoided at all costs, or can it be tolerated so long as chemical changes don't occur—or vice versa? Is the specimen at all susceptible to heat or radiation? Depending on the answers to these questions, some of the following methods will be inappropriate. A flow diagram summarizing the different preparation philosophies is shown in Figure 10.2.

NANOMATERIALS

Think about your choice of material for the supporting grid.

A self-supporting specimen is one where the whole specimen consists of one material (which may be a composite). Other specimens are supported on a grid or on a Cu washer with a single slot. Several grids are

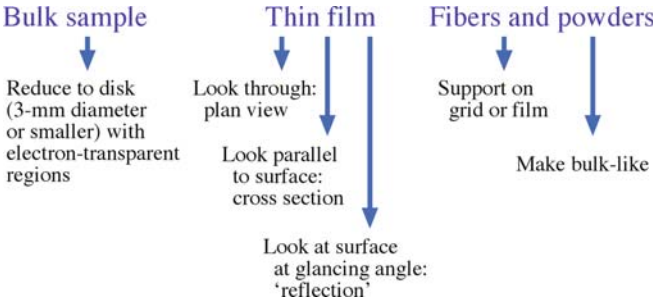


FIGURE 10.2. Flow chart summarizing the different sample geometries you may encounter.

shown in Figure 10.3. Usually the specimen or grid will be 3 mm in diameter.

Both approaches have advantages and disadvantages. Both offer you a convenient way of handling the thin specimen, since either the edge of the self-supporting disk or the grid will be thick enough to pick up with tweezers. If possible, never touch your specimen when it is thin. We recommend vacuum tweezers, but you'll need to practice using them; you can quite easily vibrate the specimen and break the thin area. You can get round this by using mouth-vacuum tweezers but see the section on safety first. Mechanical stability is always crucial. For example, single crystals of GaAs or NiO break very easily, so it is usually an advantage to have your specimen mounted on a grid since you then 'handle' the grid. However, if you are performing X-ray analysis on a specimen the grid may contribute to the signal, because the X-rays can also arise from the grid. Thus you see a Cu peak in the X-ray spectrum where no Cu is present in the specimen. We'll talk in Chapter 33 about how to minimize this artifact. Of course, the self-supporting specimen essentially has the same problem—it's just not as obvious! In fact, the preferred geometry for such analysis is usually the one where the specimen is thinnest.

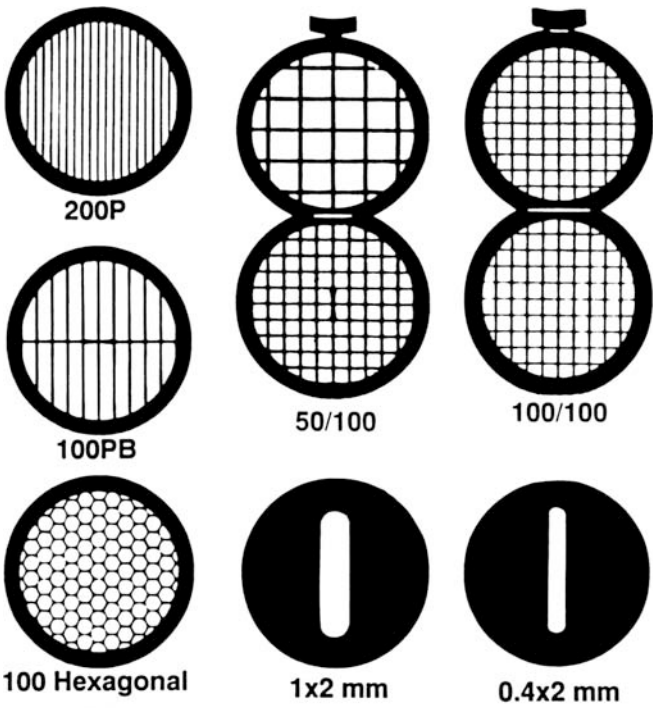


FIGURE 10.3. A variety of specimen support grids of different mesh size and shape. At top right is the oyster grid, useful for sandwiching small slivers of thin material.

THE DIAMETER OF TEM SPECIMENS

Why are the TEM specimens 3.05 mm in diameter? Because the manufacturers say so. Must this always be the case? Only if you need to double-tilt or tilt-rotate.

Why 3-mm disks? The disk diameter is usually a nominal 3.05 mm. We thus refer to the specimen as a 3-mm disk. Occasionally you will encounter a microscope which uses a 2.3-mm disk. The smaller diameter was used in earlier microscopes and has two important advantages, which are not fully exploited by modern machines. Ideally, the region of the specimen which you want to study will be located at the center of your disk, no matter how large the disk is. As we saw in Chapter 9, the reason is that, as you tilt the specimen in the microscope, the region of interest will then stay at the same position (height) above the objective lens and on the optic axis. Since, for a self-supporting disk, the rim of the specimen must be relatively thick and the total area of the material you'll study is small and confined to the center of the disk, you can make more 2.3-mm specimens from a given volume of material. This may be very important if the specimen is particularly special (expensive, rare) or if specimens break easily. A sample which is 5 mm × 5 mm will give one 3-mm disk or four 2.3-mm disks. The second advantage of such specimens relates to tilting; the smaller specimen holder can be manufactured to allow a greater tilt angle. Don't forget that if you need only one axis of tilt you may find the bulk holder useful. Then you can use a specimen which may be up to 10 mm long and 3 mm wide.

10.3 PREPARING A SELF-SUPPORTING DISK FOR FINAL THINNING

Preparation for final thinning involves three parts

- Initial thinning to make a slice of material between 100 and 200 μm thick.
- Cut the 3-mm disk from the slice.
- Prethin the central region from one or both faces of the disk to a few micrometers.

The method you use will depend on what you want to study and the physical characteristics of the material (whether it is soft or hard, ductile or brittle, delicate or robust, single phase or a composite, etc.).

10.3.A Forming a Thin Slice from the Bulk Sample

The materials you may need to thin can vary enormously. Clearly, we have to treat ductile and brittle materials differently.

(a) *Ductile materials such as metals*. Usually you don't want to introduce mechanical damage. For example, you may want to study the defect structure or the density of defects in processed materials. The ideal method is to use a chemical wire/string saw, a wafering saw (not diamond—the soft metal will dull the blade), or spark erosion (electro-discharge machining) to get a thin slice < 200 μm. (A string saw works by passing the string through an acid or solvent and then across the sample until the string 'cuts' through the sample; for example, you can use dilute acid to cut copper.) You could also roll the material to, very thin sheet, then anneal it to remove the defects introduced by rolling but that's a different material-processing route.

(b) *Brittle materials such as ceramics*. Here there are two cases: (i) where you must not introduce mechanical damage, (ii) where you don't mind introducing mechanical damage or the material won't damage. You have several options depending on the material. Some materials (Si, GaAs, NaCl, MgO) can be cleaved with a razor blade; these are materials with a well-defined cleavage plane and it is possible to carry out repeated cleavage to electron transparency (see Section 10.6.E). The ultramicrotome (see Section 10.6.B) allows you to cut very thin slices for immediate examination. If you don't want to cleave the specimen or you want to prepare a specimen parallel to a plane that doesn't cleave, you will need to use a diamond wafering saw. There are special techniques for some materials: you can, for example, use water as the solvent on a string saw to cut rock salt. One of the main limitations with sawing is that the process destroys some of your sample.

10.3.B Cutting the Disk

The same constraints hold for the coring process as for cutting slices: if the material is reasonably ductile and mechanical damage is not crucial, then the disks can be cut using a mechanical punch. A well-designed punch can cut disks with only minimal damage around the perimeter, but the shock can induce shear transformations in some materials. For more brittle materials the three principal methods are spark erosion (also essential when you need to avoid damage in a metal), ultrasonic drilling, and using a grinding drill. In each case the cutting tool is a hollow tube with an inner diameter of 3 mm. Again, you want the wall of the tube to be thin to minimize the amount of material that is wasted. Spark erosion is used for conducting samples and introduces the least amount of mechanical damage. The choice between an ultrasonic drill (vibrating in H_2O) and a grinding (or slurry) drill is often a matter of personal preference or availability. Both remove material mechanically and are widely used for ceramics and semiconductors. The drill may leave small particles in the specimen, and *all* mechanical thinning methods leave some surface damage. As a rule of thumb, abrasives produce damage to 3× their grit size. So a 1-μm abrasive will cause damage to 3 μm below the surface of each side of the specimen. Hence the final disk must be thicker than the 2× the damage depth or else mechanical damage will always be visible in the final specimen. Four different coring instruments from one manufacturer are shown in Figure 10.4.

CORING
Like extracting the core of an apple or a rock sample from the earth.

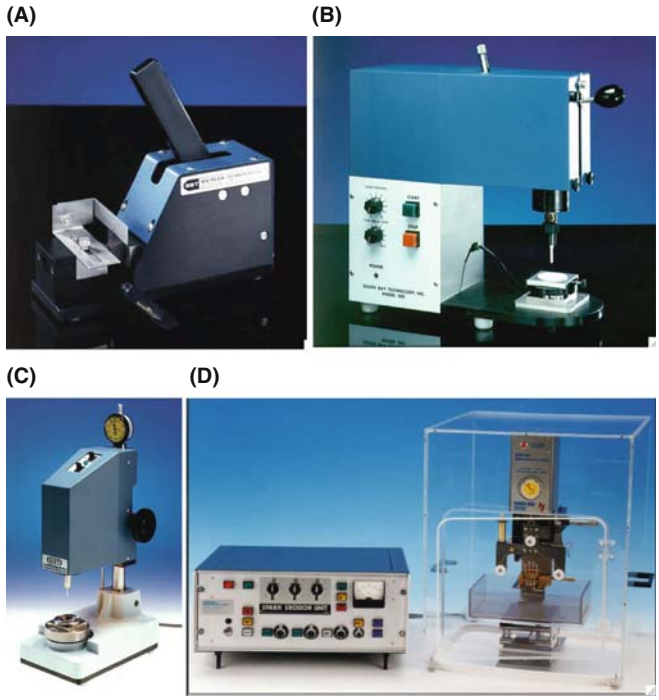


FIGURE 10.4. Four different coring tools from South Bay Technology. (A) A mechanical punch for stamping disks from thin sheets of ductile materials. A sheet sample is placed in the punch and the handle on the right is pushed down, ejecting a 3-mm-diameter disk suitable for thinning. (B) An abrasive-slurry disc cutter uses a rotary motion of the coring tube to drill round the disk. (C) An ultrasonic cutter. (D) A spark-erosion cutter; the erosion takes place under a solvent and behind a safety shield.

Note that there are variations for all these techniques: e.g., for Si, GaAs, and some other materials, you can glue the sample to a support, coat it with a protective layer, and cut circles through the film—then chemically etch the desired region. You need to experiment but the method should introduce no mechanical damage.

10.3.C Prethinning the Disk

The aim of this process is to thin the center of the disk while minimizing damage to the surface of the sample. In general we will refer to this stage as 'dimpling' no matter how the thinning is achieved. Any damage you create at this stage will have to be removed during the final thinning process (if you're interested in defects or if the damage changes your chemistry).

Most commercial mechanical dimplers use a small-radius tool to grind and polish the disk to a fixed radius of curvature in the center. Although the first instruments for dimpling were 'home built' the commercial models (see Figure 10.5) are now well developed. You can control the load, precisely determine the thickness of removed material (the depth of the dimple), quickly change the polishing tool, and interrupt the process to remove the sample for closer examination before continuing. The investment is well justified for materials laboratories. One alternative that has been used successfully is a (recycled) dentist's drill and some imagination. Typically dimpling can be carried out to produce regions ~10 μm thick although, in principle, precision dimpling with microprocessor control can sometimes produce electron-transparent specimens which are <1 μm thick.

For mechanical dimpling, as a general rule, the same guidelines apply as to all mechanical polishing; always gradually decrease the 'grit' size and conclude with the finest available, again ensuring that the final specimen thinness is >2× the damage depth of the smallest grit dimension. The better the polished surface, the better the final specimen. If both sides of a disk are dimpled the chances of final perforation occurring in the center are substantially increased, but in some cases you may wish to preserve one side of the specimen and thin from the other side only. One-sided dimpling is then essential prior to thinning to perforation.

Dimpling can also be performed chemically. Often in the case of Si this is achieved by allowing a jet of HF and HNO_3 to impinge (from below as shown in Figure 10.6) on the Si disk which has the edges lacquered to produce a supporting rim. The HNO_3 oxidizes the Si and the HF removes the SiO_2. Similar approaches use Br and methanol for thinning GaAs. This dimpling method uses dangerous chemicals, but it is very efficient. It can even be carried to final perforation with care.

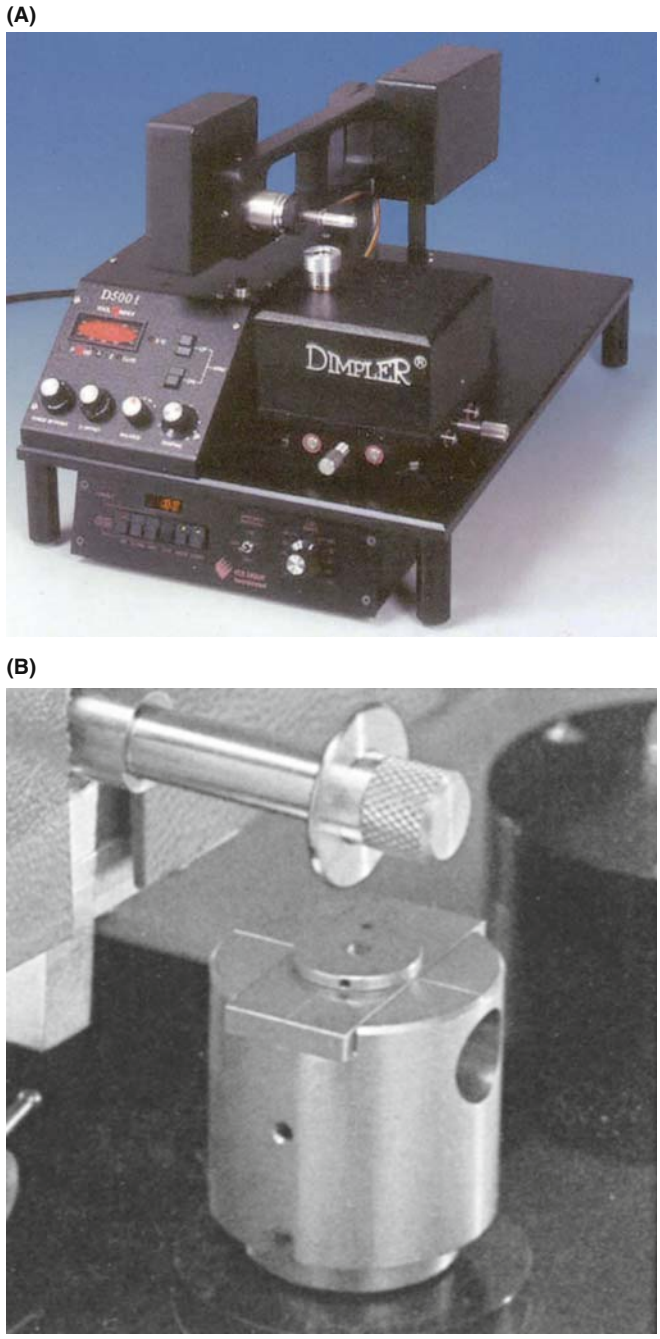


FIGURE 10.5. (A) Dimpling apparatus; (B) the grinding tool and specimen support block.

TEM specimen preparation has been revolutionized through the development of the tripod polisher; this tool can help you to thin your sample mechanically to less than 1 μm. You must consult the general references at the end of the chapter before using this tool. The tripod polisher, so called because it has three feet, is simply a device to hold your specimen while you mechanically thin it on a polishing wheel. You can purchase a tripod polisher commercially or build your own.

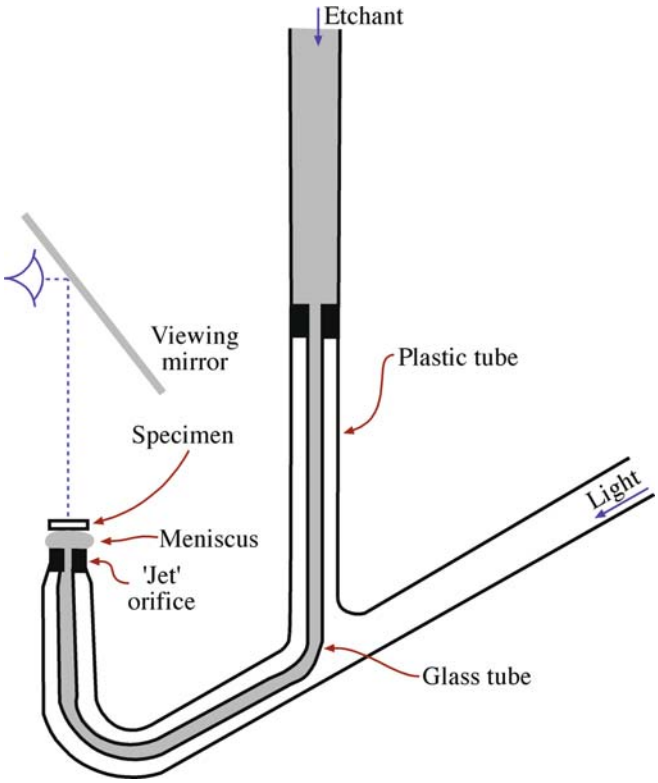


FIGURE 10.6. Surface dimpling using a chemical solution, e.g., to remove Si from one side of a disk. The light pipe permits visual detection of perforation using the mirror.

TRIPOD POLISHER

The tripod polisher has been used for many decades in the gem industry. It just looked different.

For some materials, such as Si, you can use this polisher to thin the specimen to electron transparency.

There are, however, several secrets in using the tripod polisher.

- You must use a very flat polishing wheel; the recommended approach is to use a glass platen. Take the greatest care in adjusting the micrometer to level the tripod.
- You need a supply of fine diamond lapping films; these are not inexpensive but it is a false economy to use them after they are worn. Always use a new sheet for polishing the second side of your sample since it is then particularly vulnerable.
- The diamond lapping films must not have an adhesive backing; you 'fix' them to the glass platen using the surface tension of the water and ensure that they are flat using a wiper blade. Bumps under the films will destroy your specimen.
- Any debris on the film will reduce its useful life; if the pad dries with polishing paste still present you should discard it.
- Minimize the effect of debris, which you produce on the polishing film as you thin your sample, by paying careful attention to where you place the specimen on the polishing wheel; orient interfaces in cross section samples normal to the radius and don't cross the debris trail.

With practice, tripod polishing can dramatically reduce the time required for the final thinning step. This tool has had a major impact on making TEM a quality-control instrument, particularly in the semiconductor industry.

10.4 FINAL THINNING OF THE DISKS

10.4.A Electropolishing

Electropolishing can only be used for electrically conducting samples such as metals and alloys. The method can be relatively quick (a few minutes to an hour or so) and it can produce foils with no mechanical damage. But it can change the surface chemistry of the specimen and it can be hazardous to your health as you can see from the safety section at the start of the chapter.

The basic premise is that there is a certain applied voltage at which the current due to anodic dissolution of the specimen creates a polished surface rather than etching or pitting, as shown in Figure 10.7. The classical jet polish is shown in Figure 10.8A. By keeping the volume of the reservoir constant, the jet falls under constant pressure. The voltage is applied between the tip of the pipette and the specimen. A twin-jet apparatus can be used to pump a jet of electrolyte onto both sides of the dimpled disk, as shown schematically in Figure 10.8B. A laser beam or light sensor detects transparency and a warning sound is given. At the warning, the electrolyte flow must be cut off immediately to prevent loss of thin area and the disk must be rapidly extracted from the electrolyte and washed in solvent to remove any residual film of electrolyte which may etch the surface.

Undoubtedly you get better at electropolishing with practice, but reproducing the correct conditions of temperature, electrolyte solution chemistry, stirring rate, applied voltage, polishing current, etc., can only be achieved through trial and error.

10.4.B Ion Milling

Ion milling involves bombarding your delicate thin TEM specimen with energetic ions or neutral atoms and sputtering material from your film until it is thin enough to be studied in the TEM. A schematic diagram and a commercial model are shown in Figure 10.9. The variables which you control include the voltage,

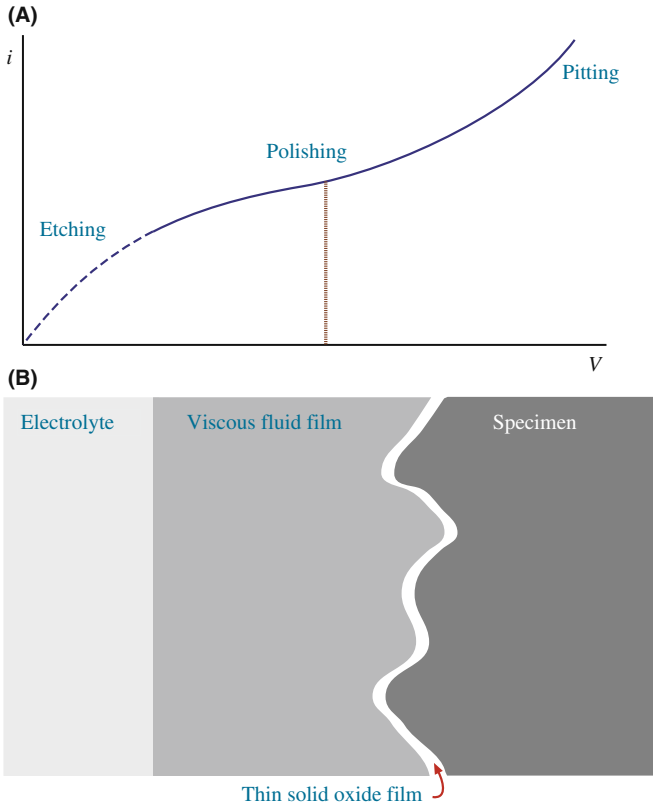


FIGURE 10.7. (A) Electropolishing curve showing the increase in current between the anode and the cathode as the applied voltage is increased. Polishing occurs on the plateau, etching at low voltages, and pitting at high voltages. (B) The ideal conditions for obtaining a polished surface require the formation of a viscous film between the electrolyte and the specimen surface.

temperature of the specimen (e.g., cold milling (liquid N2)), the nature of the ion (Ar, He, or a reactive ion (iodine)) and the geometry (the angle of incidence).

ION THINNING

Variables are ion energy, angle of incidence, vacuum, initial surface topology, initial chemistry, initial orientation, initial crystallography of the surface, beam energy, and beam profile. Note the word *initial*.

An accelerating voltage of 4–6 keV is usually used. The ion beam will always penetrate the specimen to some extent, so we minimize this by inclining the incident ion beam to the surface of the specimen. In the past, we often aligned the ion beam at an angle of 15–25° to the surface. However, Barna has shown that this angle of incidence should be avoided in many cases since it leads to compositional thinning; use an inclination of $\leq 5°$ to avoid preferential thinning and minimize ion implantation. Some implantation will occur so that the chemistry of the near-surface region is changed and the

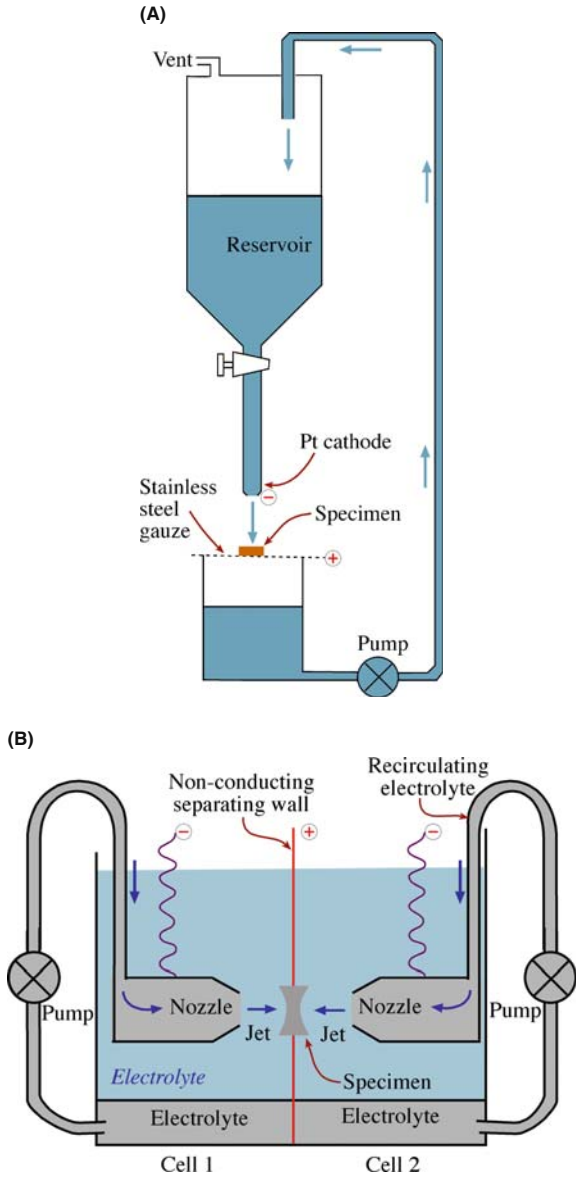


FIGURE 10.8. (A) Jet electropolishing by allowing a single jet of gravity-fed electrolyte to thin a disk supported on a positively charged gauze. The disk has to be rotated periodically. (B) Schematic of a twin-jet electropolishing apparatus. The positively charged specimen is held in a Teflon holder between the jets. A light pipe (not shown) detects perforation and terminates the polishing.

material is physically damaged (the top layer is often amorphized). If you use a low angle of incidence ($< 5°$), you'll deposit the energy of the ion beam in a region close to the surface of the specimen. A lower beam energy or a lower Z ion will also do less damage, but in both cases milling time will increase (Figure 10.10). In principle, you could also control the vacuum around the sample.

One thing you must remember is that ion thinning is closely related to ion-beam deposition. One manufacturer uses a similar arrangement to coat samples for SEM. The result is that material removed from one

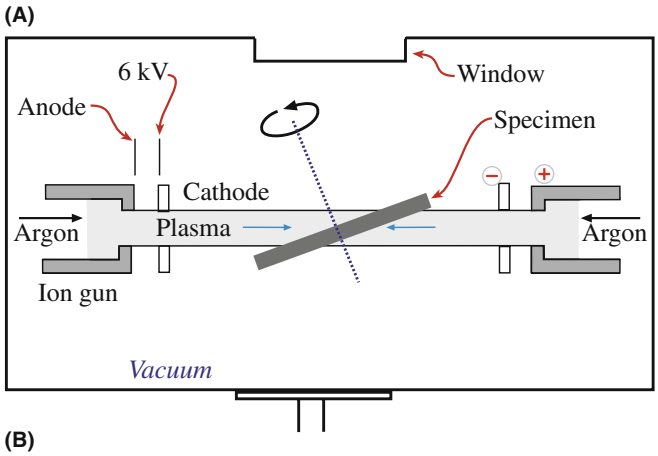


FIGURE 10.9. (A) Schematic diagram of an ion-beam thinning device: Ar gas bleeds into an ionization chamber where a potential up to 6 keV creates a beam of Ar ions that impinge on a rotating specimen. Although not shown, the whole apparatus is under vacuum. The specimen may be cooled to liquid-N_2 temperatures and perforation is detected by the penetration of ions through the specimen. (B) Typical ion mill.

part of the sample can easily be redeposited elsewhere on the sample.

The theory of ion milling is complex. We can define the sputtering yield to be the number of atoms ejected per incident ion; the yield depends on the mass of the incoming ion. The yield also depends on the ion used and the sample being milled. The principal variables are

- *The ion*: mass, energy, charge, and angle of incidence
- *The 'target'*: mass density, atomic mass, crystallinity, crystal structure, and orientation

Ar is used because it is inert, heavy, and not naturally present in most samples. Special applications may use reactive iodine, or add oxygen, etc.; this idea of reactive-ion etching is commonly used in semiconductor processing. The problem is that the reactive ion may contaminate or corrode your thinning device, the diffusion pumps, etc. Heavy ions give less penetration, but create more damage.

Most of the thinning parameters are generally fixed except the ion energy, the angle of incidence and any rotation, and the temperature of the specimen. A typical approach is to start with rapid thinning conditions (heavy ions, high incidence angle) and slow the thinning rate as perforation approaches. The effect of incidence angle on the thinning process is shown in Figure 10.10. Cooling the specimen is recommended for almost all materials; otherwise, it is possible that the ion beam might heat it to 200°C or higher. Even in metals which have good thermal conductivity, the creation of vacancies through ion damage can cause diffusional changes equivalent to heat treatment at such temperatures.

You may encounter discussions of whether to use ions or neutral atoms; one idea is that neutralized ions should not be affected by charging of a non-conducting specimen. It is not clear that neutral atoms remain neutral throughout the thinning process so this may be a moot point.

Ion milling is the most versatile thinning process, being used for ceramics, composites, polyphase semiconductors and alloys, and many cross section specimens. In addition, fibers and powders, which constitute a wide range of important materials, can also be thinned by ion milling. To do this, you have to first embed the particles or fibers in epoxy and transfer the mixture into a 3-mm brass tube for strength. The next step is to saw the tube/epoxy mixture into 3-mm disks and finally dimple and ion mill to electron transparency, as shown in Figure 10.11. A similar method (but without the brass tube) can be used prior to ultramicrotomy of powders and fibers (see Section 10.6.B).

Remember: Always beware of artifacts: some stories best illustrate this. Goodhew reports that Ar bubbles

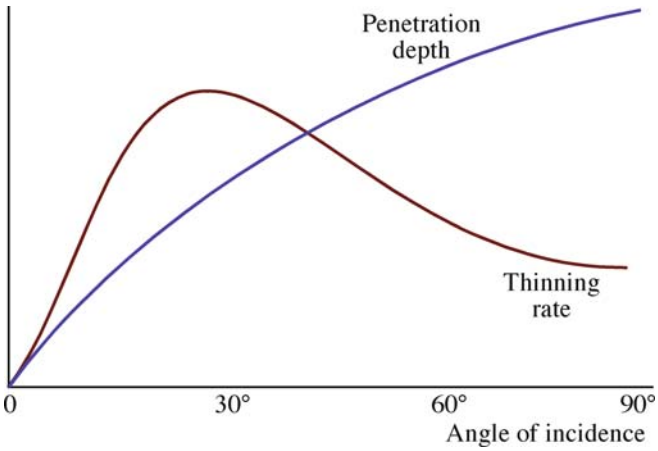


FIGURE 10.10. Variation in penetration depth and thinning rate with the angle of incidence. High-incidence angles promote implantation, which is undesirable. The rate of thinning reaches a maximum at ~20° incidence, after which the beam penetrates rather than sputters the sample surface. Initial thinning should start at 20–30° reducing to <10° as perforation approaches.

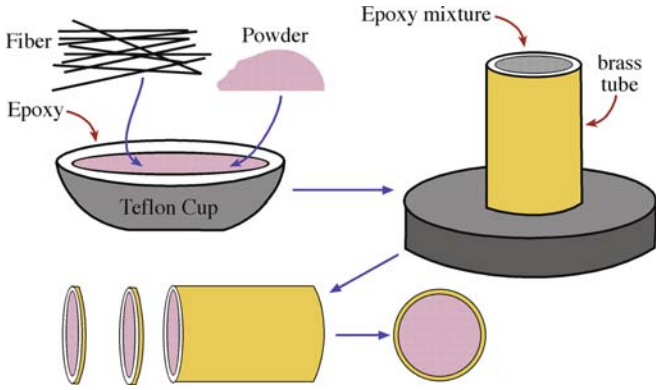


FIGURE 10.11. Sequence of steps for thinning particles and fibers by first embedding them in epoxy and forcing the epoxy into a 3-mm (outside) diameter brass tube prior to curing the epoxy. The tube and epoxy are then sectioned into disks with a diamond saw, dimpled, and ion milled to transparency.

form in silicon at a depth of ~10 nm after 5-keV thinning. Elemental analysis (XEDS) of some β-aluminas which had the correct structure by HRTEM (composition $K_2O \cdot 11Al_2O_3$) gave a composition with the K completely replaced by Ar. Glasses and zeolites can also accommodate large amounts of Ar. Cooling the specimen can often reduce contamination and surface damage. It is best to use two ions guns. If this is not acceptable, because you want to study the surface region, then you may want to coat one side with a polymer-protective lacquer and then dissolve this coating after thinning to remove sputtered material.

Why Rotate and Cool the Specimen? The specimen is usually rotated (at a few rpm) during thinning, otherwise you tend to get surface structure—grooves which run in certain directions; if you see these, check to see that the rotation has not stopped. In the preparation of cross-section specimens, you may use beam blockers and rotation control. In the first, you physically block the sample to shield it in certain directions from the ion beam so that it cannot thin, say an interface, preferentially. In the second, you vary the rate at which you rotate the sample to achieve the same effect. The latter is preferred if it is available since the time spent thinning the specimen is maximized.

Why cool the specimen? You can minimize atom migration in or on the specimen. We noted above that the specimen might be heated to >200°C otherwise. An additional advantage is that the cooling system also cools the surroundings to give a contribution of cryopumping and simple cryotrapping. However, you have to give the specimen time to warm up after milling which can increase preparation times.

Tilting the Specimen This depends on your ion miller but if you're choosing a new machine there may be an advantage in tilting the gun rather than the specimen. If the specimen is inclined, then you need a clamping ring and you may sputter this when you thin the specimen. This has led to the development of ion polishing instruments (see later) where the ion thinner has been optimized to provide a low angle without a retaining clamp. The specimen rests on a support and can be thinned at an angle of 4–5°.

Practical Design of the Ion Miller The schematic diagram in Figure 10.9 doesn't do justice to a modern ion miller which is a highly sophisticated piece of equipment. Two ion guns are available to thin from each side. The operating vacuum is $< 10^{-3}$ Pa without Ar and $10–10^{-1}$ Pa when Ar is bled into the gun. The ion guns are basically hollow chambers into which the Ar is introduced; then it is ionized and accelerated through a hole in the cathode. The hole gradually enlarges due to ion sputtering and cathodes need replacing after some time to maintain a high-intensity ion beam. More advanced gun designs incorporate saddle fields to focus the ion beam at the specimen and increase the thinning rate. The beam can be neutralized in some systems if the charged ions cause too much damage.

Some special phrases you'll encounter

- Reactive ion milling. The classic example is the use of iodine in the work described by Cullis and Chew. Iodine has a clear advantage for InP where the formation of In islands under Ar thinning is suppressed. In CdTe only growth defects were observed in iodine-thinned specimens, but many other defects were found in the same material thinned using Ar ions (Figure 10.12).
- Beam blockers and variable rotation speeds. Often the epoxy in a cross-section specimen thins faster than the specimen. Therefore we want to direct the ion beam at the different materials for different amounts of time. The two approaches used are blocking the beam geometrically using 'beam blockers' or varying the rotation velocity; e.g., you don't want the beam to thin along the interface. The latter approach can be extended further to oscillate the specimen, always keeping the ion beam at the same angle of incidence, so that it is never parallel to the interface.
- Low-angle, low-energy ion mills. Examples include the PIPS (Gatan's precision ion polishing system) and Baltec's Gentle Mill. These ion mills combine high-powered ion guns and a low angle of incidence (4°) to thin one side of a specimen with minimum surface damage and heating. The low incidence angle removes any surface roughness and differential thinning problems, while the high-power guns ensure reasonable thinning rates. The Gentle Mill can even be used to thin FIBbed specimens (when not on a C foil) (see Section 10.7).

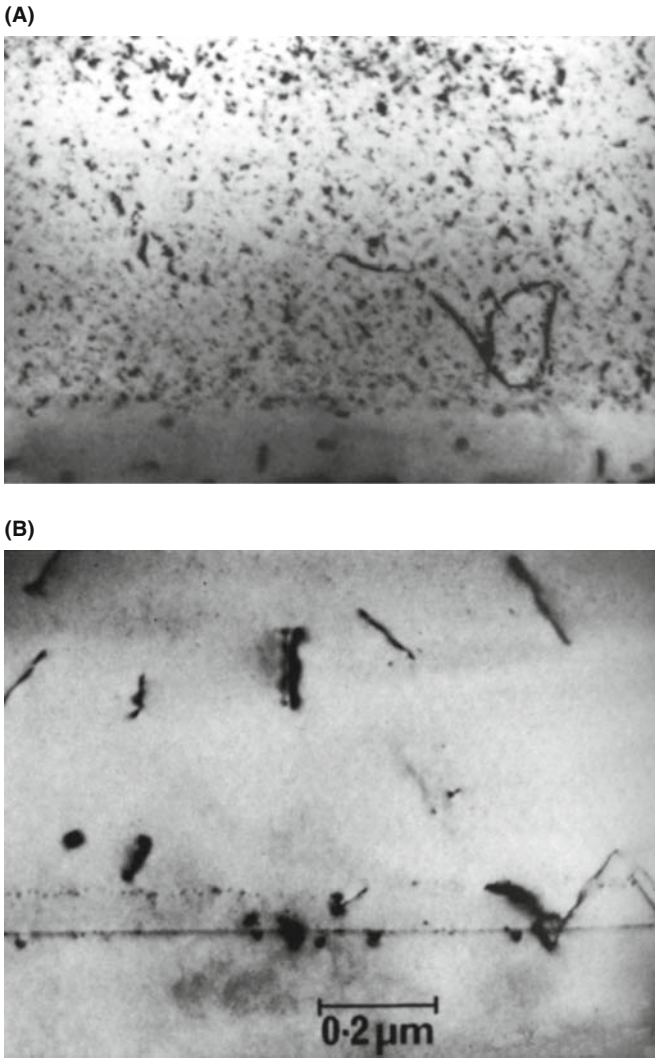


FIGURE 10.12. BF images of CdTe showing (A) defects (dark spots) in Ar-thinned specimen and (B) undamaged crystal thinned by reactive-iodine ion milling. The residual defects in (B) were formed during CdTe crystal growth.

Some final points to remember

- Materials thin at different rates. It's a good idea for the person responsible for the ion millers to run a test specimen periodically with nominally the same conditions, to be sure that the machine is still working optimally.
- Don't start with a thick sample. Always make the surface as smooth as possible before beginning to ion thin.
- Keep a record of what conditions you use: record the beam current, angle of incidence, rotation rate, and kV.
- Ion milling will form a layer on one surface or both which will probably be a combination of amorphous, highly damaged, and implanted material! The chemistry of the layer will differ from the rest of the specimen. The thickness of crystalline material will thus be less than the total thickness.

10.5 CROSS-SECTION SPECIMENS

The cross-section specimen is a special type of self-supporting disk. You must master this preparation technique if you are studying interfaces. We have often stressed that one of the principal limitations of the TEM is its insensitivity to variations in the structure and chemistry of the specimen in the direction of the electron beam. Therefore, if we are to look at structural and chemical variations close to an interface we have to prepare specimens in which the interface is parallel to the electron beam and this involves cross sectioning the sample. The most widely studied cross-section samples are semiconductor devices which often have multiple layers and therefore have multiple interfaces. But any composite materials, samples with surface layers (e.g., oxide-metal interfaces), MBE specimens, quantum-well heterostructures, etc., are candidates for this type of preparation.

There are numerous techniques for preparing cross-section specimens. Many details are reported in four MRS proceedings so we'll only describe a few basic principles. First, rather than trying to thin one interface only, the sample can be cut and glued together to produce several layers, rather like a club sandwich. Then the sandwich is sectioned such that we can see the layers, as shown schematically in Figure 10.13. In this process, a critical step is the gluing of the sections to form the sandwich. Several epoxies are available that cure at low temperatures, so that you won't heat treat the specimen inadvertently. The thickness of the epoxy layer must be such that it is thick enough for good adhesion, but not so thick that it is completely thinned away during final ion milling.

You can then cut the glued sections into 3-mm rods using an ultrasonic drill. Alternatively, you can cut the

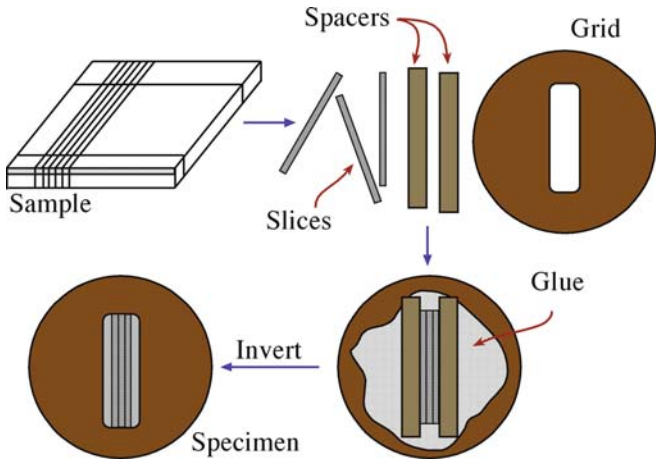


FIGURE 10.13. Schematic sequence for cross-section specimen preparation: the sample is cut into thin slices normal to the interfaces which are glued together between spacers which could be Si, glass, or some other inexpensive material so that they are wider than the slot in the grid. The 'club' sandwich is then itself glued to the grid (over the slot) and ion milled to perforation.

samples smaller and encase them in a 3-mm thin-walled tube. Section the filled tube into disks which you can then ion thin. The advantage of this method is that the final specimen has a thick ring of the tube metal around it, which gives it mechanical stability. With multiple interfaces the final thinning is almost always guaranteed to produce electron transparency at a useful region.

10.6 SPECIMENS ON GRIDS/WASHERS

The alternative to self-supporting disks is to make small electron-transparent portions of the specimen or create particles and support them on a thin film on a grid or washer. We can deposit these small particles on amorphous or crystalline films. The classic example is the amorphous carbon film (the holey carbon film), but this is not always the best choice. Some of the particles of the material of interest will be located partially over a hole so that they do not overlap anything else.

The thin supporting film should have a uniform thickness; the idea is that you are not actually interested in this material and therefore want to minimize its effect on the image of the material you are interested in.

The particles may stick to the film or may have to be clamped between two grids. Special hinged 'oyster' grids (see Figure 10.3) are available which make this very easy. Some of the processes we've already discussed can be used to make these specimens.

10.6.A Electropolishing—The Window Method for Metals and Alloys

Electropolishing is an application of electrochemistry and is regarded by many as a 'black art': a recipe which works one day but might not work the next. We can electropolish a thin sheet of metal. First cut the sheet into a square ~10 mm on the side, then seal the edges with a polymer lacquer to prevent preferential attack. The 'window' of exposed metal is immersed in electrolyte (usually cooled to slow the rate of dissolution), surrounded by a cathode and a voltage is applied, as in Figure 10.14A. The solution may or may not be stirred. The correct voltage will ensure that a viscous layer of electrolyte builds up at the surface of the specimen which results in uniform controlled thinning without pitting or corrosion. After some time, which you have to determine experimentally, the sheet is removed, cleaned, and turned through 180° and replaced in the bath as shown schematically in Figure 10.14B. If this procedure is done correctly (and this might require several rotations) the sheet will finally thin in the center. If final thinning occurs too near the top of the sheet, the edge of the perforation is smooth and relatively thick. After perforation, remove the sheet and cut off slivers of material from around the perforation using a scalpel under an inert solvent such as ethanol. Catch the floating slivers on oyster grids, dry them and they are ready for viewing.

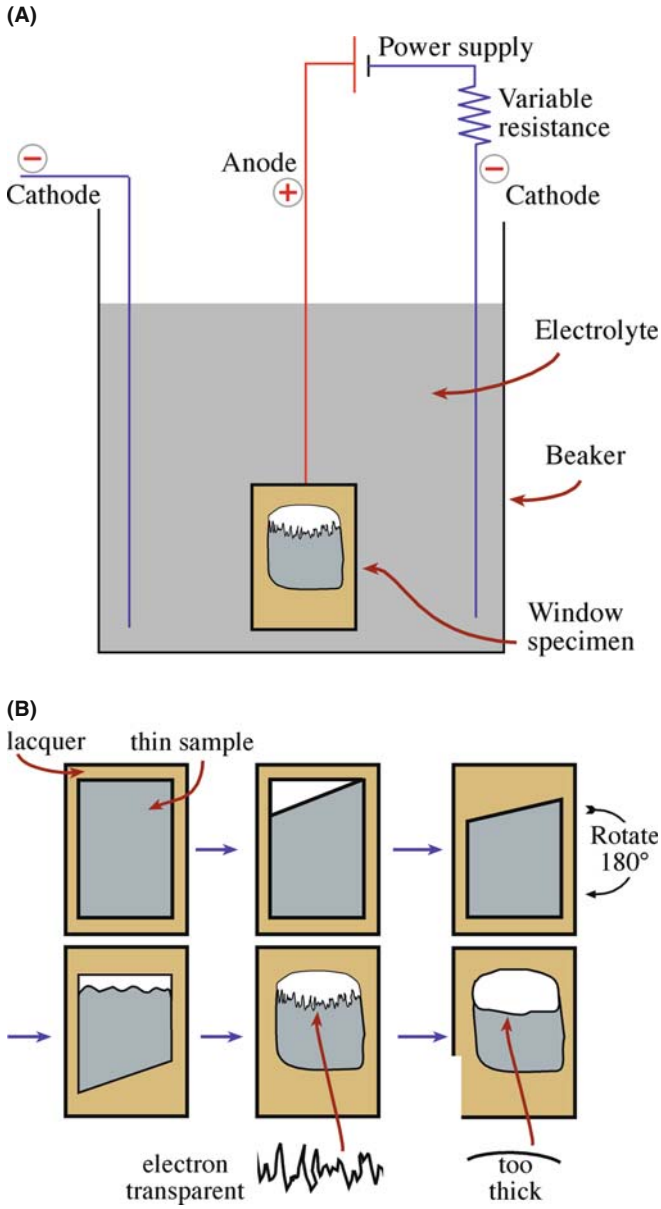


FIGURE 10.14. Window polishing. (A) A sheet of the metal ~100 mm^2 is lacquered around the edges and made the anode of an electrolytic cell. (B) Progress during thinning: the initial perforation usually occurs at the top of the sheet; lacquer is used to cover the initial perforation and the sheet is rotated 180° and thinning continues to ensure that final thinning occurs near the center of the sheet; if the final edge is smooth rather than jagged it is probably too thick.

10.6.B Ultramicrotomy

The microtome has long been used for sectioning biological materials. A tome is a 'piece cut off' so a microtome refers to the instrument used to cut a very thin tome (not like the one you're reading). With care and

much practice the biologist can reconstruct a 3D picture of the specimen. For visible-light microscopy the specimens are usually <0.1 mm thick; for the TEM the slices may be <100 nm thick and the instrument is known as an ultramicrotome. These instruments are routinely used for biological samples or for polymers where the samples tend to be quite soft. More recently, they have been used for many studies of crystalline materials. The principal advantages of the technique are that it leaves the chemistry unchanged and is thus ideal for AEM specimens, and you can use it to create uniform thin films of multiphase material. The main disadvantage, of course, is that it fractures and/or deforms the samples and therefore is most useful in cases where the defect structure is of secondary (i.e., zero) importance.

The ultramicrotome operates by moving the specimen past a knife blade. The blade can be glass (cheap) for soft materials but will be diamond for harder ones. Since there are so many possible applications, we will describe a few and refer to the references at the end of the chapter for more details. Two processes can occur in principle: the knife can cut a soft sample or it can cause a partly controlled fracture in a hard/brittle sample. In either case the limiting process is usually plastic deformation of the sample. The principles of this technique are shown in Figure 10.15.

You may also find ultramicrotomy useful if you want to study particles or fibers which are too small to thin individually but are too large to be electron transparent. You can embed the sample as we saw for the ion-thinned particles but without using the metal sheath (see Figure 10.11). We also use epoxy if the sample contains so many interconnected pores that it cannot be thinned mechanically. For porous materials, place the sample in a vacuum chamber, pump out the chamber, and coat the sample with epoxy using a dropper in the chamber. When the sample is fully encapsulated, admit air to the chamber so as to push the epoxy into the pores. After curing, you can ultramicrotome the sample in the usual way.

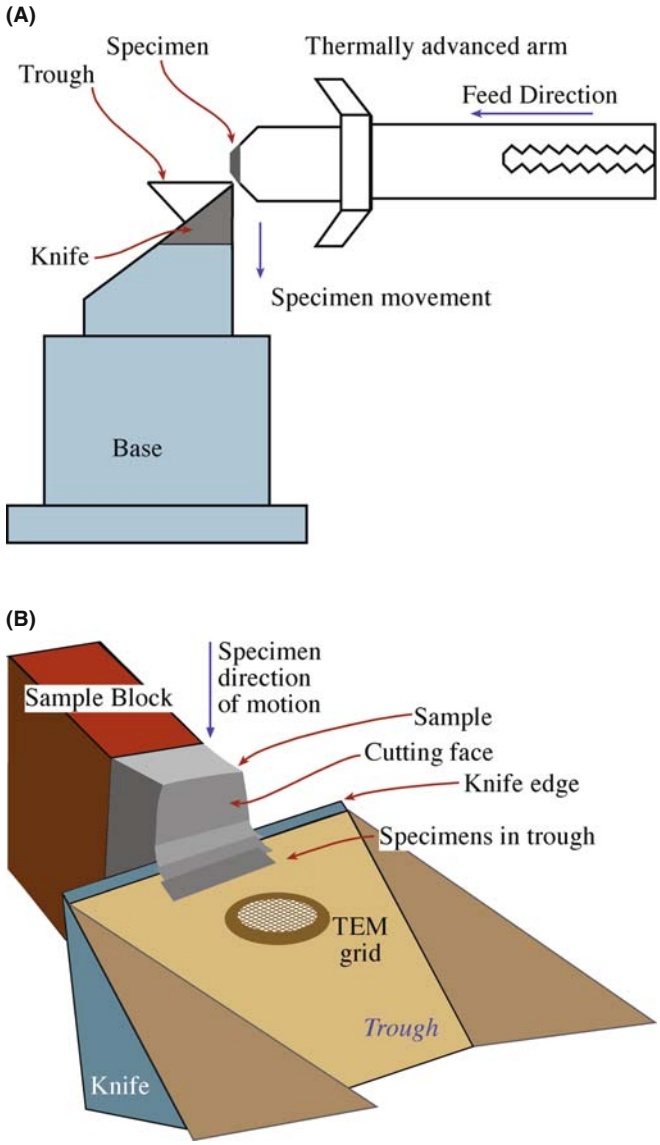


FIGURE 10.15. Ultramicrotomy. (A) The sample is first embedded in epoxy or some other medium or the whole sample is clamped and moved across a knife edge. (B) The thin flakes float off onto water or an appropriate inert medium, from where they are collected on grids.

10.6.C Grinding and Crushing

Many brittle materials such as ceramics and minerals are most easily prepared by crushing in a clean pestle and mortar (preferably in an inert liquid). The liquid containing the particles can then be ultrasonically stirred and allowed to settle. Particles suitable for TEM are too small to be seen by eye and the supernatant liquid in which they remain should appear clear. A drop of this liquid, if placed on a holey carbon film on a grid, will evaporate in a dry environment, leaving a distribution of the particles on the support film. If the particles have to be crushed dry, then agglomeration can be a problem. Electrostatic forces sometimes cause small particles to clump together and distributing them on a grid can be very difficult. In these cases, it sometimes pays to mix up the crushed material in an epoxy, then ultramicrotome the epoxy, as we just described in the previous section.

We can collect dust particles found in airborne pollution by simply exposing a support film (on a grid) to the atmosphere for a period of time. Interstellar dust can be sampled from a spacecraft or a high-flying plane.

10.6.D Replication and Extraction

These methods are among the oldest TEM specimen-preparation techniques. We use direct replication to

study fracture surfaces or surface topography in general. Evaporate a carbon film on the surface of interest, then etch away the underlying surface with an acid so that the carbon film floats off. If you coat this film with a heavy metal at an oblique angle, you will thus produce a sample that shows enhanced mass-thickness contrast (see Chapter 22); support the film on a grid for observation. As an alternative (Figure 10.16A) you can first replicate the surface by softening a plastic, pressing it on the surface, and allowing it to harden. Pull off the plastic replica, coat it with carbon, then dissolve the plastic with a suitable solvent, and pick up the carbon replica on a support grid. If the carbon replica is produced directly from a metal surface, it may be necessary to dissolve some of the metal with acid then float off the carbon onto distilled water before picking up on a grid, as shown in Figure 10.16B. After picking up on a grid it may be useful to coat the replica obliquely with a heavy metal to enhance any topographic (thickness) contrast.

Extraction replication has seen a resurgence of interest since AEM techniques appeared, because we can extract a particle from its surrounding matrix, thus allowing us to analyze that phase alone without interference from electron scattering into the matrix.

The various steps for extraction are shown in Figure 10.17A. The sample is polished metallographically to expose the particles on the surface. An appropriate etching process is used to remove the matrix such that the particles stand proud of the surface. A carbon film is evaporated onto the surface and scored into ~2 mm squares. Then the etching is continued. As the matrix is dissolved, the squares of carbon film float to the surface carrying the particles with them. Catch one of these squares on a grid and you have your specimen ready for the TEM as shown in Figure 10.17B. Again, oblique shadowing may be useful to enhance image contrast, but not if you plan to use AEM.

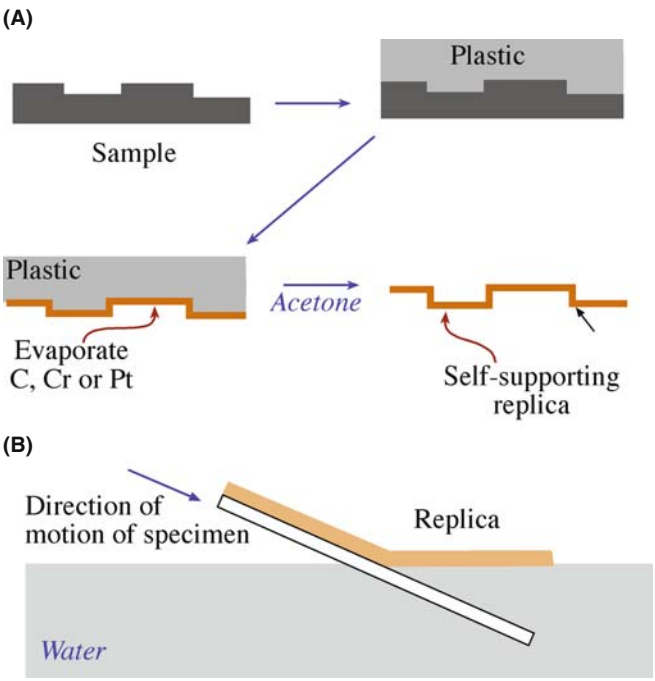


FIGURE 10.16. (A) Replication of a surface by the two-step method: spray acetone on the surface to be replicated before pressing a plastic (usually cellulose acetate) onto the surface which softens in contact with the acetone; the plastic is removed from the surface when it has hardened and a C, Cr, or Pt film is evaporated onto the replicated plastic surface; the plastic is then dissolved with acetone and the evaporated film retains the original topography. (B) Alternatively, the direct carbon replica of a metal surface may be floated off on distilled water after scratching the carbon and etching to free the film, which may subsequently be shadowed obliquely to enhance the topography.

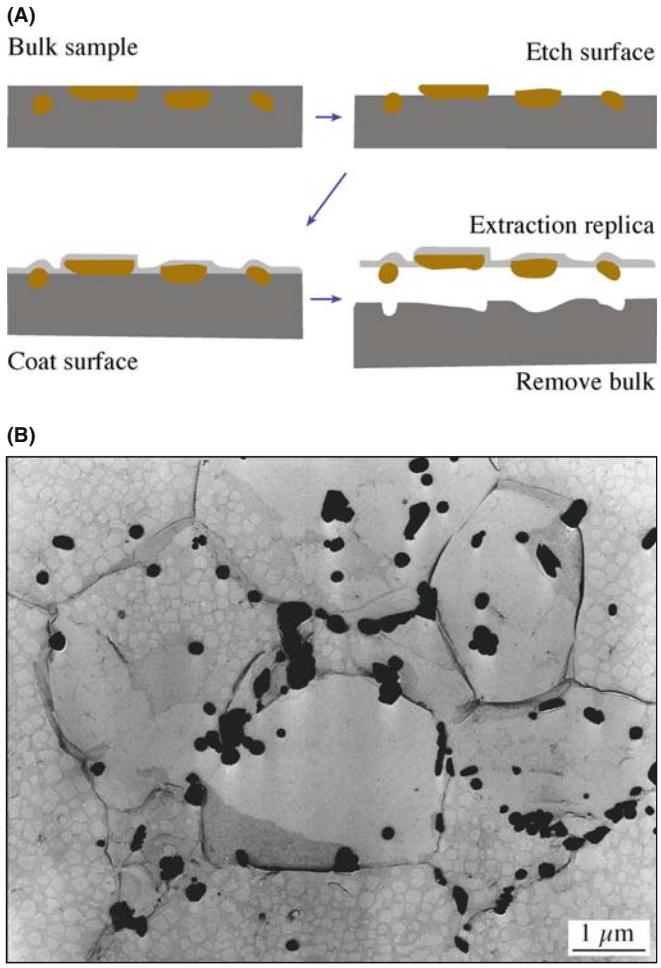


FIGURE 10.17. (A) Making the extraction replication: particles embedded in a matrix are revealed by etching the matrix, which leaves the particles standing proud of the surface; a thin amorphous carbon film is evaporated over the particles, then the rest of the matrix is etched away leaving the particles adhering to the carbon film. (B) Example from a γ/γ' alloy showing not only that the particles are mainly located at the grain boundaries but also the different contrast from γ' grains and two-phase γ/γ' grains.

10.6.E Cleaving and the SACT

Cleaving is one of the oldest techniques and has been used to make thin specimens of graphite, mica, and other layer materials that are weakly bonded along one plane. The classical idea is to attach adhesive tape to both sides of the sample and then pull the two pieces of tape apart. This process is repeated until the specimen is thin enough for TEM. You can really only tell this by experience: as it becomes thinner, graphite becomes a lighter shade of gray in transmitted visible light. Molybdenite (MoS_2) becomes a lighter shade of green as illustrated in Figure 10.18. Place the tape with the thin flake of material in a solvent to dissolve the glue (all traces of glue *must* be removed). This technique is not as easy as it once was. The glues used to be readily soluble in trichlorethylene which is now a known carcinogen.

A special variation on cleaving, known as the small-angle cleaving technique (SACT), is illustrated in Figure 10.19. The idea is to propagate a crack through the sample along a plane that is not a natural (crystallographic) fracture plane and then create another crack that is shallowly inclined to the first fracture surface. The technique can be applied to crystalline samples such as Si coated with thin films or to coated glass samples (as shown in Figure 10.19) that have no preferred fracture plane. In the case of glass, this is particularly attractive since the only alternative is ion milling which tends to implant argon into the open structure of the glass and is thus not suitable for AEM specimens.

Figure 10.19A shows a sample that has been scribed with several parallel lines. Each of the rectangular samples can then be fractured again by pressing on it with a stylus to produce a sharply wedged specimen. With luck, the resulting sample is electron transparent as seen in Figure 10.19D where a small particle of NaCl is imaged between a glass substrate and a coating layer. Although it is a hit-and-miss technique, you can make so many specimens in one sitting that a hit is assured (with lots of practice). You can then put this sample into the FIB and produce an even better sample without spending too much FIB time (money).

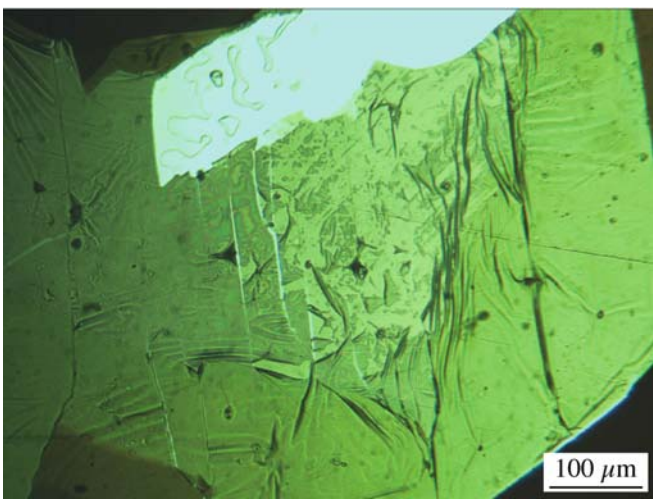


FIGURE 10.18. Image of cleaved MoS_2 showing regions of different shades of green, which correspond to different thicknesses.

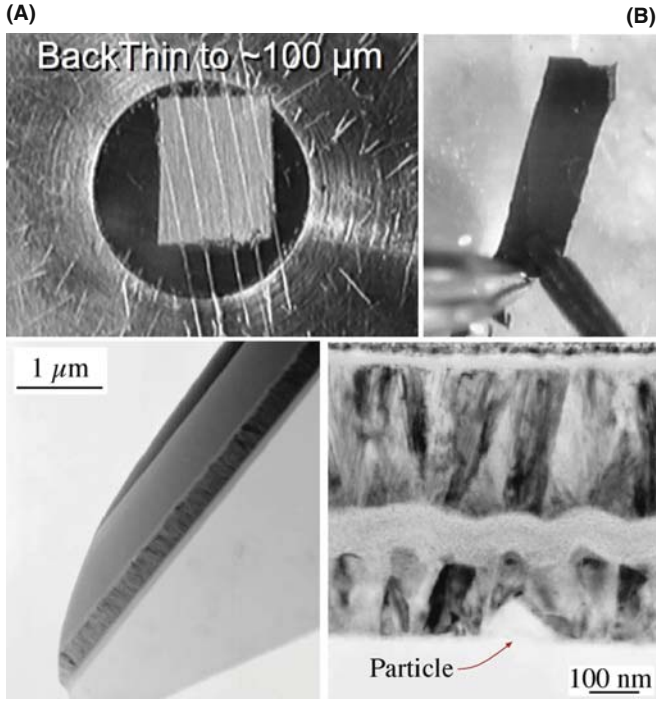


FIGURE 10.19. SACT of a coating on glass. (A) Scratch the sample; (B) cleaving along the scratch; (C, D) TEM images.

SACT

The small-angle cleaving technique is invaluable for films on Si or glass where there is no crystal structure; making the 90° wedge would be a LACT!

10.6.F The 90° Wedge

The 90°-wedge specimen was developed because many compound semiconductors such as GaAs are grown with a (001) surface and can be easily cleaved on the (110) and ($1\bar{1}0$) planes that are perpendicular to this growth surface. When you are practiced at cleaving the sample as shown in Figure 10.20, you can examine a specimen in the TEM within 30 minutes of completing the growth.

Mount the specimen as shown in the figure, preferably so that you won't need to tilt it in the microscope. Although the specimen is only transparent close to the edge of the 'hole,' you will have a long strip of material suitable for viewing. As always, beware of artifacts. If your specimen is perfect, you will know exactly how

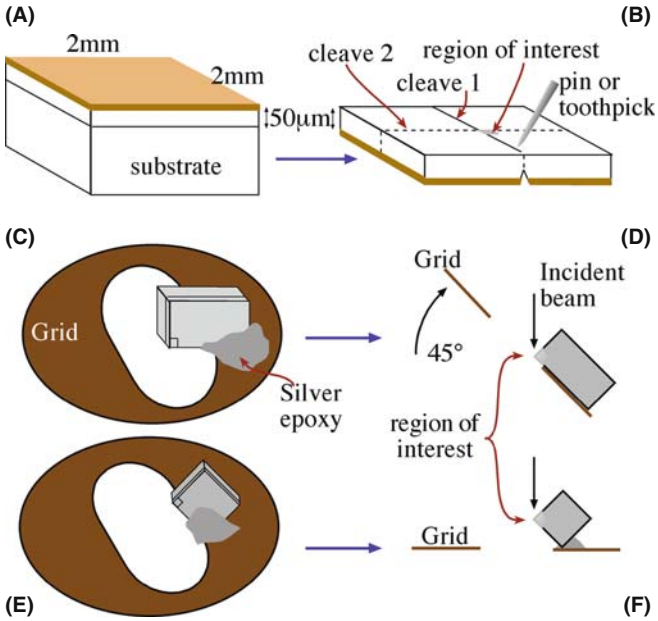


FIGURE 10.20. The 90°-wedge specimen: (A) prethin to create a 2-mm square of the multilayers on a Si substrate; (B) scribe the Si through the surface layers, turn over, and cleave; inspect to make sure the cleavage is clean, giving a sharp 90° edge; reject if not; (C, E) mount the 90° corner over the edge of a hole in a Cu grid; (D, F) then insert in the TEM; note that two different orientations are available from a single cleavage operation.

thick it is at the position you choose for study. We will find this wedge useful when we discuss image contrast in Part 3.

10.6.G Lithography

Here we use a technique developed for advanced engineering applications. Lithography is used in the microelectronics industry to define fine lines of width down to 100 nm. An illustration of how lithography can be used specifically to prepare TEM specimens (as opposed to generating a structure which might best be characterized by TEM) is shown in Figure 10.21. We can draw lines on the layered material using standard lithographic techniques. Material on either side of the lines is then removed by etching (chemical or ion) to give a plateau which is thin in one direction. We then remove most of the remaining substrate and attach the specimen to a support washer. We can then observe the specimen directly in the TEM. Although the width (formerly height) of the electron-transparent region is narrow it can extend across the entire hole in the 3-mm disk. The major disadvantages or limitations of the technique are (i) the dimension in the direction of the electron beam is fixed by the lithographic capabilities and (ii) tilting the specimen may quickly cause the thicker region to block the electron beam.

10.6.H Preferential Chemical Etching

The principle behind this technique is the same as for lithography: we remove part of the sample to leave an area which is electron transparent. The trick is to keep part of the final specimen thick enough for handling, or ideally for supporting, the specimen. Naturally, this approach only works with certain materials although the principle might be extended to other thin films. The technique has been used for III–V compounds where $Al_{1-x}Ga_xAs$ acts as an etch stop for GaAs and for Si where an etch stop can be produced by implanting with boron (Figure 10.22). In both cases, the resulting thin layers may be used as substrate materials for thin-film studies rather than as the subject of study in their own right. An extreme example of this approach is the use of thin films of Si_3N_4 as an 'etch stop' so that a uniform layer of amorphous Si_3N_4 remains across the window. Such specimen supports are commercially available; diamond films can be made in a similar way.

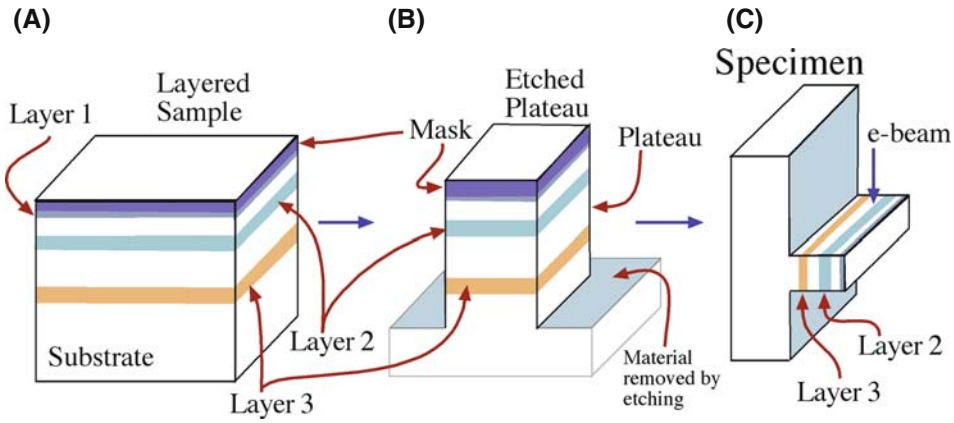


FIGURE 10.21. Etching of a multilayer sample (A). Etch away most of the sample, leaving a small etched plateau (B); mask a region <50 nm across and etch away the majority of the surrounding plateau. If this thin region is turned 90° and mounted in a specimen holder (C), the interfaces are now parallel to the electron beam.

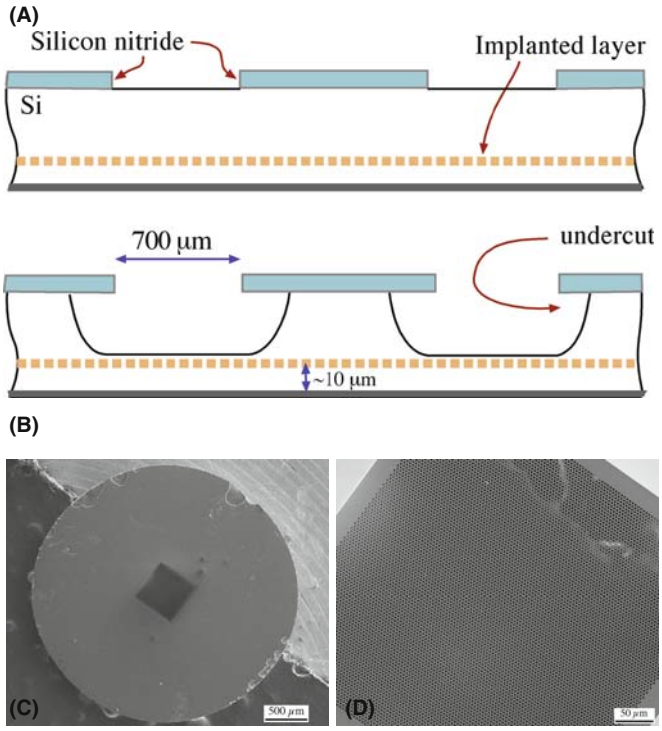


FIGURE 10.22. Lithographic techniques applied to thinning a multilayer specimen: (A) the unthinned sample is shown with a grid of Si_3N_4 barrier layers evident. Etching between the barrier layers, shown in (B), produces an undercutting down to the implanted layer which acts as an etch stop, producing a uniform layer ~10 μm thick. Further thinning with a different solution produces large areas of uniformly thin material (not shown) supported by the Si_3N_4 grid and the remaining unthinned regions. (C, D) A commercial Si_3N_4 thin films support disk; (D) is the enlarged view.

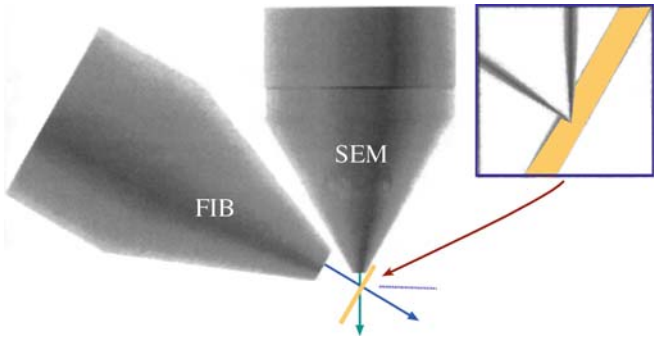


FIGURE 10.23. Schematic of a two-beam (electron and ion) FIB instrument.

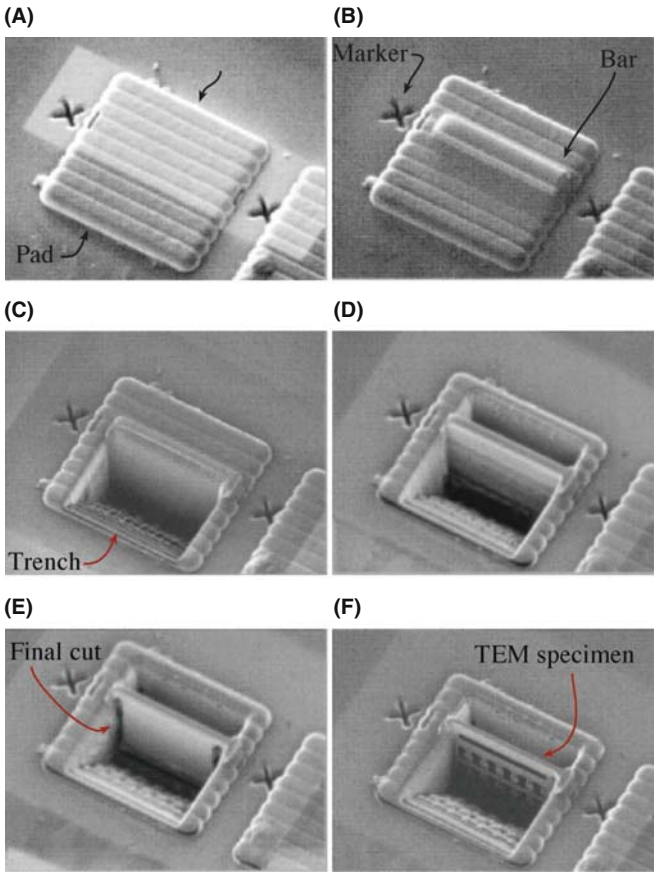


FIGURE 10.24. Stages in making TEM samples using a FIB instrument. (A) The area of interest has been marked. (B) A Pt bar is deposited to protect this area from the Ga beam. (C, D) The two trenches are cut. (E) The bottom and sides of the slice are (final) cut. (F) The TEM specimen is polished in place before extracting it.

10.7 FIB

The focused-ion beam (FIB) instrument is becoming much more readily available as prices moderate and their value is realized; however, a FEG-equipped FIB can still cost more than your TEM. We deal with FIB more extensively in the companion text but include a summary here because you must know about it even if you can't get access to one yet.

When preparing TEM specimens we can think of the FIB essentially as an SEM with a built-in ion mill. (Sometimes it's an ion gun with an SEM attachment.) The single ion gun produces a well-controlled beam of Ga ions (rather than Ar used in the ion mill). In the simplest (cheaper) design the ion beam also acts as the electron beam of the SEM with the secondary electrons being used to form the 'SEM' image of the sample. A schematic of the FIB is given in Figure 10.23. The various stages in the preparation process are shown in Figure 10.24. The pad in (A) is the coating of Pt. The two Xs, to mark the region of interest, have been drawn on the sample using the ion beam, and another Pt strap is deposited between them (B). Next, two staircases are cut out on either side to leave the thin wall shown in (C) and (D). In (E) the 'wall' has been trimmed away at the sides so that it is only supported at the top. The last step

(F) is to ion-polish the thin wall until it is really a thin TEM specimen and finally attach it to a probe for lift out or use static electricity to lift it out and place it on a supporting (usually C) film. Attaching the FIBbed specimen to a probe is becoming the norm since it allows the FIBbed specimen to be further cleaned to remove Ga contamination and/or reduce the thickness further. An FEI version of the instrument is shown in Figure 10.25. We'll discuss the details of FIB in the companion text because not everyone can afford to buy a FIB or can afford to use one even if it is 'available'!

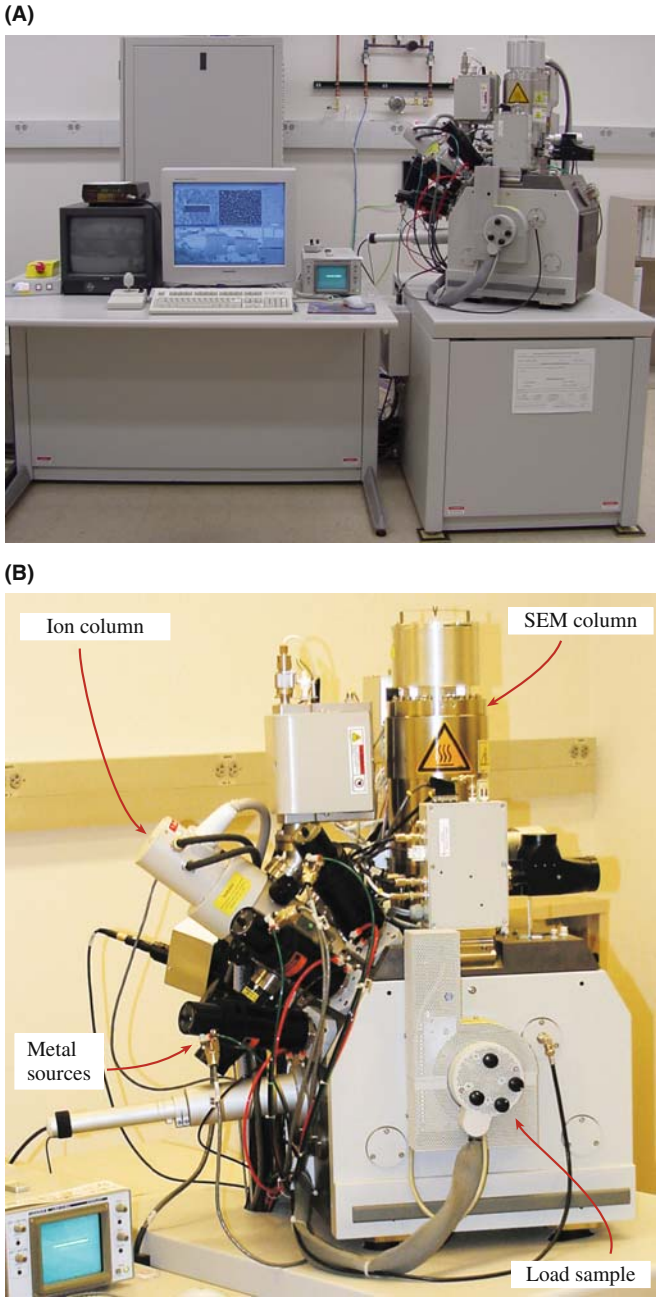


FIGURE 10.25. A dual-beam FIB instrument. (A) Overview; (B) enlarged view to compare with Figure 10.23.

10.8 STORING SPECIMENS

The best advice is to look at your specimen as soon as possible after preparation. If that is not possible, then keep your specimens under optimum conditions. Usually this means keeping them dry (water vapor affects the surface region of most materials), perhaps in an inert atmosphere (dry nitrogen works well, or a dry-pumped, oil-free vacuum desiccator) and in an inert container (a glass petri dish with filter paper).

The next problem is long-term storage; for periods up to 1 month, you can use the above procedure. If you want to keep the specimen longer, your choices can be more difficult. Don't use gelatin capsules for anything resembling 'delicate' material. Don't use slotted grid-holders for anything which might deform (break or bend) during handling; that rules out self-supporting ceramics, metals, and semiconductors. Always use vacuum tweezers to manipulate delicate specimens (remember safety when using mouth-vacuum tweezers). Remember, your most important specimen is the one most likely to break, bend, interact with sharp tweezers, or jump onto the floor.

Last, old specimens can be cleaned by ion polishing or chemical cleaning. This process does thin the specimen further so you may lose the area you originally studied. Ion polishing can also be useful for 'sectioning' specimens. The 'safer' re-cleaning (or refreshing) method is to use a plasma cleaner but this can change your sample.

If your specimen is a collection of nanoparticles, you will realize that such particles can be so reactive that they may have changed before they reach the TEM even for the first time. You may then have to use techniques that are more familiar to cryo-transfer users, such as transferring the specimen from the preparation chamber to the TEM in a controlled environment. Such a procedure is not routine using conventional holders.

10.9 SOME RULES

We stress once again that you must know what you want to study in your specimen before you begin specimen preparation. Figure 10.26 is a second flow chart to summarize the various possible options. Be aware of the limitations of the method you choose, particularly the artifacts introduced. Table 10.1 summarizes the artifacts introduced by various methods.

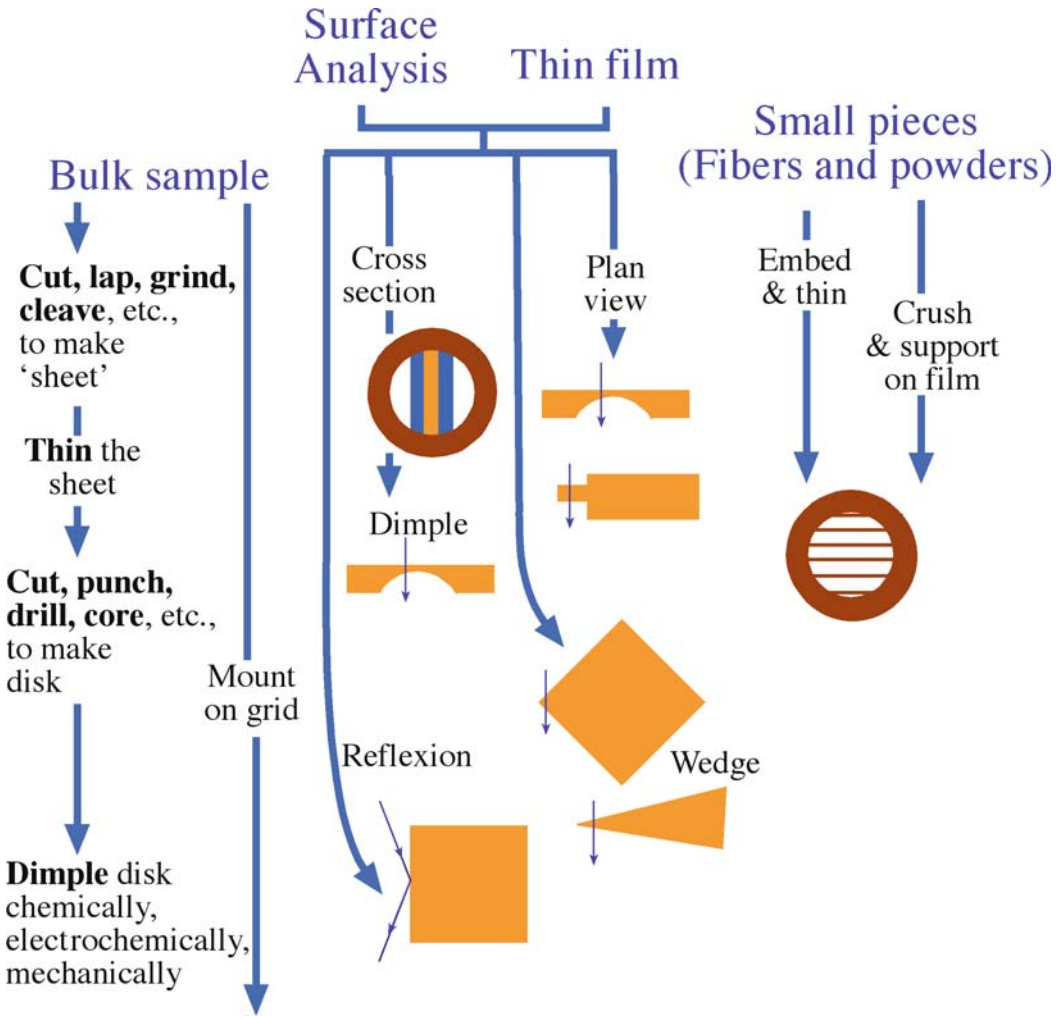


FIGURE 10.26. Summary flow chart for specimen preparation.

TABLE 10.1 Artifacts Produced During Specimen Preparation (after T. Malis)

Artifact/problem	Consequence
Variable thickness	
	▪ Limited local area for chemical mapping (EP, IT, C, CD)
	▪ Very limited area for EELS
	▪ Somewhat limited area for absorption-free XEDS
	▪ Omission of low-density defects
	▪ Distorted defect densities (EP, IT, TP)
Uniform thickness	
	▪ Limited diffraction information (UM)
	▪ Limited microstructure information (UM)
	▪ Handling difficulties (UM)
Surface films	
	▪ Bath residue, spec. dissolution and/or redeposition EP
	▪ Enhanced surface oxide (EP)
	▪ Extremely irregular topographies (IT)
	▪ Faster contamination buildup under beam (EP, R)
	▪ Retention of matrix on extracted particle
	▪ C-redeposition (UM—embedded, UM, C, R—support films)

Table 10.1(continued)

Artifact/problem	Consequence
	■ Cu_2O formation from Cu grids upon heating (R, UM, C) ■ Ion amorphization, diffusion-pump oil, redeposition (IT*)
Differential thinning	
	■ Different phases thin at different rates (EP, IT) ■ Different orientations thin at different rates (IT) ■ Grain/phase boundary grooving (EP, IT) ■ Anodic attack of matrix/particle (UM)
'Selectivity'	
	■ Perforation influenced by local defect structure (EP, IT) ■ Very limited or no microstructure information (C, R) ■ Weak local regions debond and fall out (all)
'False' defects	
	■ Microstructure obscured by high defect density (UM, CD) ■ Deformation-induced defects (EP, TP) ■ Ion-induced loops, voids (IT) ■ Heat-altered defects (EP, IT)

EP: electropolished; UM: ultramicrotomed; CD: controlled dimpling; R: extraction replication; IT: ion thinned; TP: tripod polish; C: cleavage (grinding, crushing).

CHAPTER SUMMARY

Specimen preparation is a craft and there is no substitute for hard work, careful, detailed experimentation and lots of practice as you seek to master it. This is the most tedious aspect of all of TEM work but, if you invest the time, your reward will be the best of times on the TEM itself. The quality of your data is at least directly proportional to the quality of your specimen (and this relationship is often far stronger than the linear nature just implied). You simply have to find the method that works best for your particular material. While there are many cookbooks available, the recipes are often too individualized and not to your specific taste.

There are few rules for specimen preparation except that thinner is usually better, although such specimens are more prone to artifacts. Think about each step and what it might do to change the microstructure or microchemistry of your material. Take care to avoid the physical dangers that are present whenever you use dangerous chemicals, ionizing radiation, or sharp knives. Be clean, use fresh materials, tidy up after yourself, and apply all the other lessons that you learned in kindergarten!

Although all the equipment mentioned here is available commercially, most were originally developed on a shoestring budget in someone's lab so you can always build your own electropolisher or even an ion mill. If you are working with brittle materials, buy or build a tripod polisher and learn how to use it.

A last reminder: The recipe books listed below are a great source of ideas. New recipes are appearing all the time. As is often the case in cooking it helps to see an expert chef in action to realize what is possible. In other words, when you have seen a really good TEM specimen, you'll know what yours should look like.

REFERENCES

These references only give a sampling. More extensive lists, especially for the more specialized techniques like FIB, are given in the companion text.

GENERAL TECHNIQUES

An extensive list of references is included in the chapter on specimen preparation in the companion text. The first four references below are essential: from the MRS Proceedings.

i. Bravman, JC, Anderson, RM and McDonald, ML (Eds.) 1988 *Specimen Preparation for Transmission Electron Microscopy of Materials* Mater. Res. Soc. Symp. Proc. **115** MRS Pittsburgh PA. (Number I in

the series.) We've updated the flow chart in the article on p51 by Goodhew, PJ, The tripod polisher is described by Klepeis, SJ, Benedict, JP and Anderson, RM on p179. (Pictures of the gem variety are shown in Figure 36.3 of *Ceramic Materials* by Carter and Norton.) Brown, JM and Sheng, TJ describe the use of lithography on p229.

ii. Anderson, RM (Ed.) 1990 *Specimen Preparation for Transmission Electron Microscopy of Materials, II* Mater. Res. Soc. Symp. Proc. **199** MRS Pittsburgh PA.

iii. Anderson, RM, Tracy, B and Bravman, JC (Eds.) 1992 *Specimen Preparation for Transmission Electron Microscopy of Materials, III* Mater. Res. Soc. Symp. Proc. **254** MRS Pittsburgh PA. Alani, R and Swann, PR (1992) discuss ion milling on p43.

iv. Anderson, RM and Walck, SD (Eds.) 1997 *Specimen Preparation for Transmission Electron Microscopy of Materials, IV* Mater. Res. Soc. Symp. Proc. **480** MRS Pittsburgh PA.

CHEMICAL POLISHING

Thompson-Russell, KC and Edington, JW (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science* Macmillan Philips Technical Library Eindhoven Netherlands. Many recipes.

ION MILLING AND FIB

Barber, DJ 1970 *Thin Foils of Non-metals Made for Electron Microscopy by Sputter-Etching* J. Mater. Sci. **5(1)** 1–8. First use of ion milling to make a ceramic TEM sample.

Barna, A 1992 *Topographic Kinetics and Practice of Low Angle Ion Beam Thinning* in MRS Proc. **254** 3–22. An early advocate of using low-energy ion beams.

Cullis, AG, Chew, NG and Hutchinson, JL 1985 *Formation and Elimination of Surface Ion Milling Defects in Cadmium Telluride, Zinc Sulphide and Zinc Selenide* Ultramicroscopy **17** 203–211. The paper on reactive-ion milling of TEM samples.

Giannuzzi, LA and Stevie, FA 2004 *Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice* Springer Verlag NY. The guide to FIB.

Harriott, LR 1991 *The Technology of Finely Focused Ion Beams* Nucl. Instr. Meth. Phys. Res. Section B: Beam Interactions with Materials and Atoms **55B** (1–4) 802–810. One place to start if you're interested in 'Why Ga?'

Medard, L, Jacquet, PA and Sartorius, R 1949 *Sur les dangers d'explosion des bains aceto-perchloriques de polissage electrolytique (Explosion Hazard of Acetic and Perchloric Acid Mixture Used as Solution in Electrolytic Polishing)* Rev. Metall. **46**(8) 549–560. Jacquet has many other papers on specimen preparation if your French is good.

OTHER MATERIALS

Carter, CB and Norton, MG 2007 give some illustrations of the tripod polisher used to facet diamond in *Ceramic Materials: Science and Engineering* Springer Verlag NY.

Malis, TF 1989 *AEM Specimens: Staying One Step Ahead* in Microbeam Analysis-1989 487–490 Ed. P.E. Russell, San Francisco Press San Francisco gives a discussion of using the microtome for hard samples.

Sawyer, LC, Grubb, DT and Meyers, GF 2008 *Polymer Microscopy* 3rd Ed. Springer Verlag NY gives the details for preparing polymer specimens.

THE COMPANION TEXT

A complete chapter in the companion text not only discusses more specialized techniques but also extends this chapter's discussion.

SELF-ASSESSMENT QUESTIONS

Q10.1 Name two ways you can damage yourself more than your specimen while preparing it.

Q10.2 What are the differences between self-supporting specimens and specimens resting on a grid or thin washer?

Q10.3 What is the main problem associated with using a grid?

Q10.4 What does dimpling mean?

Q10.5 What should we keep in mind when using the tripod polisher to pre-thin the specimen?

Q10.6 What is the difference in the use between electropolishing and ion milling?

Q10.7 Why do we need to use a low angle of incidence in ion milling?

Q10.8 Why do the specimens need to be rotated and cooled during the ion-thinning process?

Q10.9 Why are various chemical etching techniques used in TEM specimen preparation?

Q10.10 Why do you want the region of interest of your specimen to be in the center of the grid?

Q10.11 What methods can be used for preparing thin specimens (initial thinning) of ductile materials?

Q10.12 How is the thickness of your specimen affected by the damage done to the surface?

Q10.13 Why might you need to cool the electrolyte during electropolishing?
Q10.14 What are the advantages and disadvantages of ultramicrotomy?
Q10.15 When is it reasonable to use a mechanical punch?
Q10.16 As a 'rule of thumb,' how much damage does a typical abrasive cause to a specimen?
Q10.17 How would you go about creating a GaAs cross-sectional TEM specimen? If the material is heat sensitive (i.e., it reacts) what can be done to minimize the heat load?
Q10.18 If specimen preparation is so important, why do we spend relatively little time discussing this issue in the book, in class, and in our training?
Q10.19 List the four most widely used polishing solvents.
Q10.20 List at least three precautionary measures you should bear in mind while using hazardous chemicals.
Q10.21 What precautions are necessary for short-term storing of specimen?

TEXT-SPECIFIC QUESTIONS

T10.1 When and why would you take the time to use the window method (Figure 10.14) to prepare thin foils of a metal sheet rather than dimpling and jet polishing (Figure 10.8)?
T10.2 What TEM technique would benefit most from ultramicrotomed specimens and which technique would benefit least?
T10.3 List five different particle samples that might be easily prepared for viewing in a TEM. Give a literature (research journal) reference for each one.
T10.4 Nanotubes, nanowires, buckyballs, etc., are nano-scale specimens that have recently received significant attention. How would you prepare such specimens for examination in the TEM?
T10.5 If you had a large piece of a weld from which you wanted to prepare a thin specimen of a very specific region (e.g., the heat-affected zone or a single-pass region) for chemical analysis, how would you go about preparing such a specific thin specimen?
T10.6 List the advantages and disadvantages of making a replica of a specimen surface as in Figure 10.16 to view in the TEM rather than viewing the surface directly.
T10.7 List the advantages and disadvantages of making an extraction replica of particles from a specimen surface as in Figure 10.17 to view in the TEM rather than viewing the particles within a thin foil.
T10.8 What kinds of difficulties might arise when attempting to perform elemental analysis in the AEM when your specimen has been (a) electropolished or (b) ion milled? How might those difficulties be overcome?
T10.9 Why might ultramicrotomy be the ideal specimen preparation method for AEM but totally unsuitable for routine TEM imaging and diffraction?
T10.10 Why should your specimens remain isolated from you and other human beings after they are thinned?
T10.11 Use Figure 10.26 where appropriate to propose a method for getting a TEM specimen from
A. the join in a soldered copper alloy
B. catalyst particles on a substrate
C. a specific junction in a semiconductor device

Part 2

Diffraction

11

Diffraction in TEM

CHAPTER PREVIEW

This chapter will set the stage for our discussion of imaging using diffraction contrast. Put simply, diffraction contrast arises because the intensity of the diffr3acted beams is different in different regions of the specimen. These variations may arise because of changing diffracting conditions or because of differences in specimen thickness. In our study of diffraction in the TEM, we will see spots—lots of them. Sometimes the 'spots' will be small faint points and other times they will be large disks, which themselves contain 'structure' and more information. Other patterns will contain lines that we will examine in Chapters 19–21.

We need to know how to use the information that these spot patterns (diffraction patterns or DPs) contain. We will discuss the practical question of how we can best record the DPs, so that we can maximize the information they contain, but we will not try to give a rigorous proof of every equation used. These DPs give direct crystallographic information about small areas of the specimen. This capability is one of the most important features of the TEM, because we can relate the crystallography to the images we see.

In reading this chapter you should remember our discussion of the scattering of waves using an array of slits (Chapter 2). Much of the analysis is geometrically the same as we found for visible light—it's essentially the same physical optics. The big differences are that we have 'modulated' holes which are located in 3D space and both our wavelengths and the spacing of the 'holes' are very small.

A note on history: spot patterns in DPs have always been recorded using a photographic emulsion but many TEMs no longer have photographic plates. The CCD has a much greater dynamic range but you may need to record the DP using several different camera lengths. Plus—burning a hole in photographic film is one thing, doing the same to the CCD is another.

11.1 WHY USE DIFFRACTION IN THE TEM?

Let's begin by looking at an experimental DP. The pattern shown in Figure 11.1, like those we introduced in Chapter 2, was recorded from a thin specimen, in this case silicon. The main features to note are there are many spots and the spots vary in intensity and size (these are related effects).

We can list some of the questions you might ask on first seeing such a DP.

- What is it?
- What can we learn from it?
- Why do we see it?
- What determines the scale? What determines the distances between the spots or the positions of the lines?

What do we want to know about our specimen? To a materials scientist, perfect crystals are often pretty boring and can usually be better studied using such techniques as X-ray diffraction (for structural characterization), the electron microprobe (for chemical characterization), etc., although new EM techniques may change this situation. The TEM is the instrument of choice when the specimen is not perfect, particularly when the feature of interest is what makes the material imperfect or, paradoxically, useful!

The questions that we can address using DPs obtained in the TEM include the following

- Is the specimen crystalline? Crystalline and amorphous materials have very different properties.
- If it is crystalline, then what are the crystallographic characteristics (lattice parameter, symmetry, etc.) of the specimen?
- Is the specimen monocrystalline? If not, what is the grain morphology, how large are the grains, what is the grain-size distribution, etc.?

FIGURE 11.1. An experimentally observed DP showing the central, intense, direct beam and an array of diffraction spots from different atomic planes. Such a pattern, with sharply focused spots, is best obtained by underfocusing the beam.

- What is the orientation of the specimen or of individual grains with respect to the electron beam?
- Is more than one phase present in the specimen? If so, how are they oriented?

In general, if we see spots then the specimen is at least partly crystalline. (We'll discuss quasicrystals later.) The ability to determine crystallographic orientations locally (down to the nm level) gives TEM its great advantage over SEM and visible-light microscopes. Later on we can make this determination even more precise (to an accuracy of $\sim 0.001^{\circ}$) using convergent-beam patterns, as we'll see in Chapter 21.

In this chapter we will restrict the discussion to the geometry of the spot patterns. These are necessarily associated with crystalline materials. We'll see that spot patterns provide a great deal of information themselves; they also provide the basis for understanding other DPs. We will find that standard DPs which are common to a group of materials allow us quickly to recognize particular orientations and even certain grain boundaries and twin boundaries, etc., without having to index the pattern from scratch. For example, in a particular orientation, all cubic crystals give the same array of spots although some of the spots may have no intensity! We will consider the intensity of the spots in Chapter 12.

Remember, however, that SADPs patterns are not always the most useful DPs, since CBED (Chapters 20 and 21) can give you other useful information. Nevertheless, we are emphasizing SADPs here, since we use them to explain the contrast in TEM images, in Part III.

11.2 THE TEM, DIFFRACTION CAMERAS, AND THE TV

The use of electron diffraction for materials studies began around 1930 using diffraction cameras which very much resembled X-ray tubes in their physical appearance. Later on, if you pursue TEM in depth, you will find many of the earlier texts on electron diffraction useful for gaining a deeper understanding of TEM. It will be helpful to bear in mind some of the historical circumstances behind these developments when reading some of these texts. For example, many articles show ray diagrams with the optic axis horizontal. One reason for this is that much of the early theoretical analysis was developed as an extension of X-ray diffraction (XRD) or by researchers who were actively using either X-ray or electron-diffraction cameras. In each case, the optic axis of the instrument was horizontal as is still the case for visible-light optical benches. The optic axis of all electron microscopes is now vertical although the beam may originate at either the top or the bottom of the column. Actually, more than one of the early TEMs, e.g., the Philips EM100, was built with the optic axis horizontal and the electron beam directed at the observer. This arrangement is similar to that used for television, but remember that in TEM we are using very high energy electrons ($\geq$100 keV rather than 20 keV used in a TV). References to some of the early texts, and their historical significance, are given at the end of this chapter. When you are reading early texts on TEM remember that many were written at a time when most TEMs operated at 100 kV. This fact may easily be overlooked but it affects many features of diffraction including the camera length.

We will be talking about *positions* of spots and *not their intensities* for most of the time in this book. This type of analysis differs from many X-ray studies. The reason that beam intensities are not measured in TEM is that the electron beams are diffracted many times in a typical TEM specimen. A similar, but not identical, situation actually occurs when producing powder patterns by X-ray diffraction (XRD); diffraction then occurs in many different grains at the same time. We can compare the electron-diffraction pattern with that encountered in XRD. In the X-ray case, if you have a single crystal, then you either have to rotate the crystal to 'see' all the beams or use 'white' radiation (i.e., essentially use a range of wavelengths). Electron diffraction is very different. We can use a single wavelength and still see many diffracted beams. The techniques differ also with respect to the time

it takes to record a DP on a photographic plate; XRD takes minutes or hours unless you have a synchrotron or a position-sensitive detector to count every photon, while electron-diffraction patterns can be recorded in <1 second although longer is usually better because it means you will have 'spread the beam' (made the incident beam more parallel); for photographic film you'll use several seconds to a minute or more.

Much of the discussion of electron diffraction follows directly from the analysis of XRD. This has advantages and disadvantages, depending on whether or not you are familiar with XRD. Several references to XRD are given at the end of the chapter. When considering diffraction, remember that there are important differences between electrons and X-rays:

- Electrons have a much shorter wavelength than the X-rays commonly encountered in the research lab.
- Electrons are scattered more strongly because they interact with both the nucleus and the electrons of the scattering atoms through Coulomb forces.
- Electron beams are easily directed because electrons are charged particles.

It is particularly important that the electron beam can be deflected off the optic axis a short distance above the specimen and then pass through the specimen; this process of tilting the beam was described in Section 9.1.D. The most obvious effect of this deflection on the DP is that the whole DP is translated relative to the viewing screen. The more subtle effect results from the change in the direction of the incident beam with respect to the crystal lattice, as we will discuss in subsequent chapters.

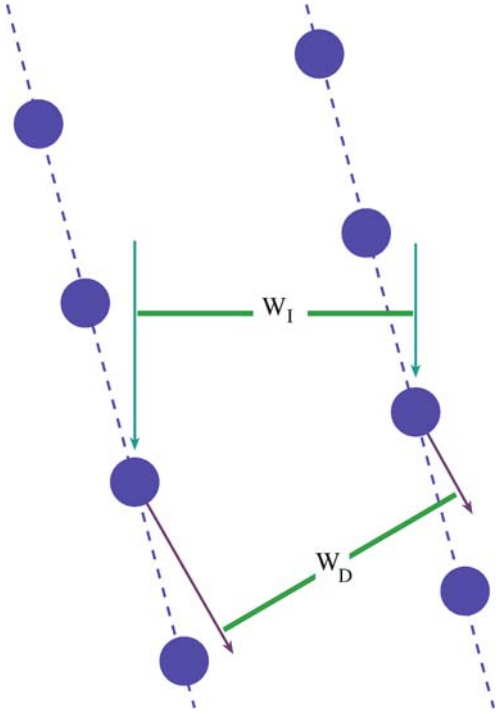


FIGURE 11.2. Scattering from two planes of atoms. W_I and W_D are the incident and diffracted wavefronts, respectively.

11.3 SCATTERING FROM A PLANE OF ATOMS

If you go back to Chapter 3 on elastic scattering you'll see that we introduced the two different ways of thinking about diffraction: the Laue conditions and Bragg's law. In this chapter we'll derive Bragg's law again, introducing a vector notation that we'll use throughout the rest of the book. In Chapter 12, we'll do the same with the Laue conditions.

The simple diagram in Figure 11.2 shows an *initial* wavefront, W_I, being scattered by a plane of atoms to produce a *diffracted* wavefront, W_D. Whether or not W_D corresponds to a diffracted beam will depend on whether the atoms are scattering in phase, which itself is determined by the angles between the incident beam, the diffracted beam, and the diffracting planes. The conditions for the individual waves being in phase are known as the Laue conditions, which we introduced in Section 3.10.B. To analyze the situation we first simplify the diagram as shown in Figures 11.3 and 11.4. These figures define the wave-propagation vectors, which we will refer to simply as the wave vectors or the **k**-vectors. We begin by considering scattering from only two atoms.

Notice that we are already mixing the concepts of waves and beams.

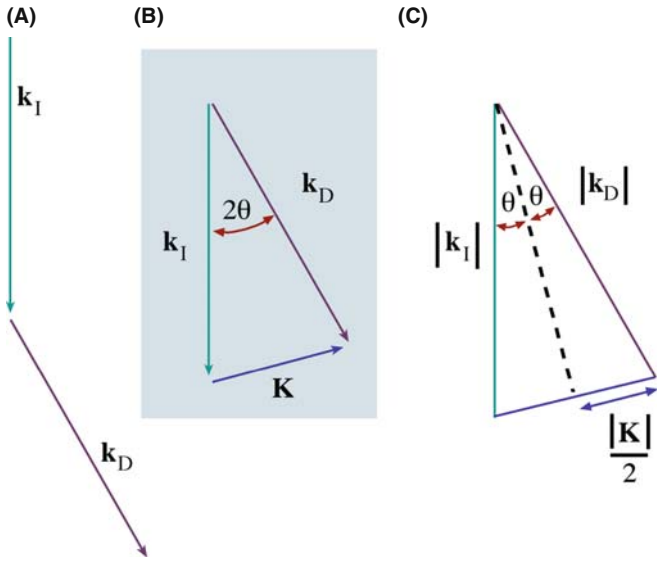


FIGURE 11.3. Definition of the scattering vectors: (A) the incident wavefront normal is $\mathbf{k}_I$, the diffracted wave normal is $\mathbf{k}_D$; (B) **K** is the difference vector ($= \mathbf{k}_D - \mathbf{k}_I$); (C) sin θ is defined as $\mathbf{K}/2\mathbf{K}_I$.

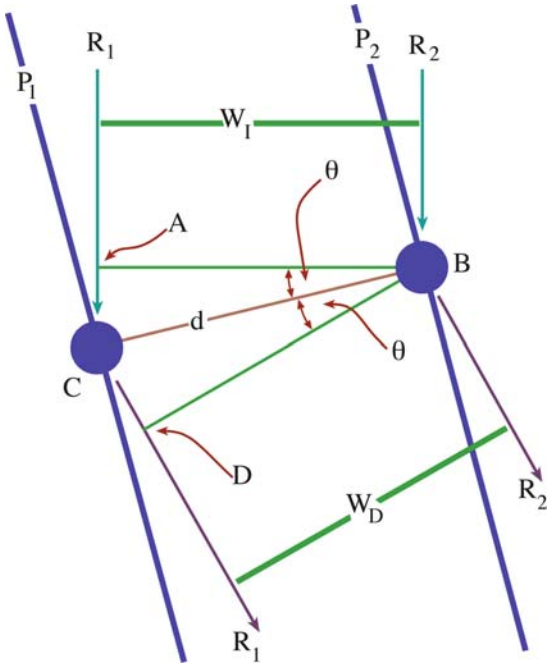


FIGURE 11.4. Two beams are scattered from two points, C and B, which lie on different planes, P_1 and P_2. The rays travel different distances giving a path difference of AC + CD.

We'll only consider plane wavefronts, i.e., the wavefront is flat and **k** is normal to this wavefront. The diagrams in Figure 11.3A and B define vectors $\mathbf{k}_I$, $\mathbf{k}_D$, and **K** and gives us the following important equation (which is just vector addition)

$$\mathbf{K} = \mathbf{k}_D - \mathbf{k}_I \tag{11.1}$$

where $\mathbf{k}_I$ and $\mathbf{k}_D$ are the **k**-vectors of the incident and diffracted waves, respectively. The vector **K** is thus the change in **k** due to diffraction. An important feature of this analysis is that this construction can be made for any $\mathbf{k}_D$ and thus for any value of **K**; the angle θ shown here need not be a Bragg angle.

Following our discussion in Section 3.10.B, we can always write that

$$|\mathbf{k}_I| = |\mathbf{k}_D| = \frac{1}{\lambda} = |\mathbf{k}| \tag{11.2}$$

providing the energy of the electron is unchanged during diffraction, i.e., the scattering process is elastic. From Figure 11.3C we can write an expression for θ using simple trigonometry

$$\sin\theta = \frac{|\mathbf{K}|/2}{|\mathbf{k}_I|} \tag{11.3}$$

or

$$|\mathbf{K}| = \frac{2\sin\theta}{\lambda} \tag{11.4}$$

$|\mathbf{K}|$, like $|\mathbf{k}_I|$, has units of nm^{-1} if λ is measured in nm. **K** and $\mathbf{k}_I$ are then referred to as *reciprocal-lattice vectors*. Note that this scattering process is taking place inside the crystal and therefore the **k**-vectors are all appropriate to the electrons inside the crystal (rather than in the vacuum).

EQUATION 11.4 IS VERY IMPORTANT
Whenever you see the term $\sin\theta/\lambda$ remember that it is just $\mathbf{K}/2$ and is thus related to a change in wave vector.

If we now extend this argument to consider the interference between waves scattered from two points (which you can visualize as being atom sites) then we have the situation sketched in Figure 11.4. This figure should remind you of the idea of constructive and destructive interference which we discussed back in Section 3.10. You will recognize that the geometry of Figure 11.4 is essentially a cross section of the two slits used by Young to demonstrate the wave nature of light (see also Section 2.10). We can then define two planes P_1 and P_2 to be normal to the vector **CB** which has length *d*. The distance traveled by ray R_1 is then larger than that traveled by ray R_2 by the path difference AC + CD. Simple geometry shows that

$$\mathrm{AC} + \mathrm{CD} = 2d\sin\theta \tag{11.5}$$

which is the basis for Bragg's law, as we'll now see.

11.4 SCATTERING FROM A CRYSTAL

We introduced the Bragg angle in Figure 3.9 as the most important scattering angle in TEM; at the Bragg angle the electron waves interfere constructively. If we now analyze Figure 11.4 further, we see that when θ equals the Bragg angle, θ_B, equation 11.4 becomes

$$|\mathbf{K}| = \frac{2\sin\theta_B}{\lambda} \tag{11.6}$$

When θ is θ_B, the path difference in equation 11.5 is $n\lambda$, where *n* is any integer, and the equation becomes

$$n\lambda = 2d\sin\theta_B \tag{11.7}$$

This is Bragg's law (equation 3.21). If *n* is 1

$$2\sin\theta_B = \frac{\lambda}{d} \tag{11.8}$$

but we already know from equation 11.6 that, at the Bragg angle,

$$2\sin\theta_B = \lambda|\mathbf{K}| \tag{11.9}$$

so when we are at the Bragg angle, the magnitude of the vector **K** has a special value, $\mathbf{K}_B$,

$$|\mathbf{K}_B| = \frac{1}{d} \tag{11.10}$$

and we define this vector, $\mathbf{K}_B$, to be **g** so that

$$\mathbf{K}_B = \mathbf{g} \tag{11.11}$$

This sequence of steps may seem rather pedantic but the conclusion is extremely important. Bragg's law and the geometry used to 'prove' it will be used so frequently in our discussions that it is worthwhile to delve a little into what it really tells us. Although it is not really a valid treatment of the phenomenon we are seeing, Bragg's law gives us a very useful physical picture of the diffraction process because the diffracting planes appear to behave as mirrors for the incident electron beam. Therefore, the diffracted beams, or the spots in the DP, are often called 'reflections' and we sometimes refer to the vector **g** as the diffraction vector. This derivation is simply geometry. In Section 12.3 we will derive the Laue equations and hence deduce Bragg's law from first principles.

DIFFRACTION, NOT REFLECTION

Don't forget we are really dealing with diffraction, not reflection, and we derived Bragg's law by considering just two atoms. We still examine reflections, but they're not.

The reason that this derivation of Bragg's law is not really valid (it just gives the right answer) is that it really applies to scattering at a glancing angle where the beam exits the same surface as it enters, not transmission.

We mentioned earlier that the angles shown in all of our figures are exaggerated for the case of diffraction in the TEM. For example, for 111 planes in Cu, d is 0.21 nm; λ is 3.35 pm (0.00335 nm or 0.0335 Å) for 120-kV electrons; equation 11.8 then gives $\theta = 7.97$ mrads (0.46°) for $n = 1$. As a rule of thumb, the Bragg angles of interest are usually no more than 1° when we are forming images (although important information may be present in DPs at much larger, 10–20°, angles). You will find it useful to remember the order of magnitude of these numbers.

DEGREES AND RADIANS

Remember 10 mrads is 0.573°, i.e., about 0.5°.

We can now generalize from single atoms to planes of atoms. Let's imagine that Figure 11.4 shows two 'planes of atoms,' P_1 and P_2, and that the points B and C are not necessarily atoms but are simply points on these planes and that d is the shortest distance between the two planes. How is the 'in-phase' nature changed if we move atom B but keep it on plane P_2?

Consider scattering from a single plane as shown in Figure 11.5. Geometry shows that while ray R_1 travels a distance EJ, ray R_2 travels a distance HF and that these two distances are equal. Thus there is no path difference for scattering from atoms located anywhere on a particular plane. This seemingly trivial result means that we can generalize our conclusions from Figure 11.4.

This result is summarized in Figure 11.6. Rays R_1, R_2, and R_3 all scatter in phase if $\theta = \theta_B$.

SCATTERING FROM A PLANE

It does not matter how the atoms (scattering centers) are distributed on these two planes; the scattering from any two points on planes P_1 and P_2 will produce the same path difference $2d \sin\theta$.

Next we extend this analysis to include many parallel planes each a distance d from its neighbors, as is shown in Figure 11.7.

In the special case of Bragg diffraction, the scattering semiangle between the incident beam and the Bragg-diffracted beam is twice the Bragg angle ($2\theta_B$). See Figure 11.7 and refer back to Section 2.12.

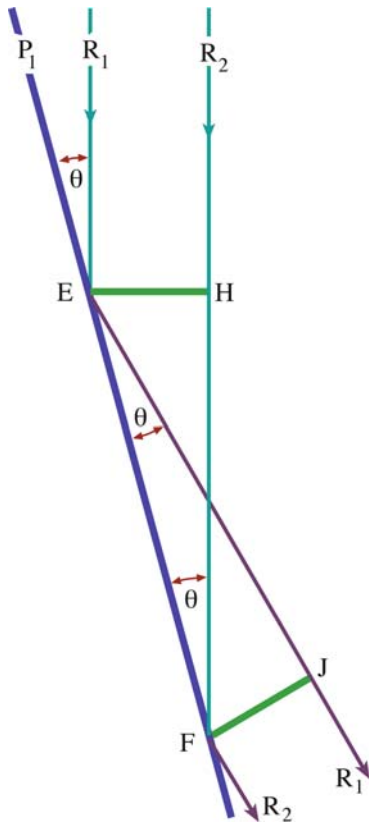


FIGURE 11.5. Two beams are scattered from two points, E and F, which lie on the same plane P_1. This simple diagram shows that the two beams travel the same distance since triangles EHF and FJE are congruent.

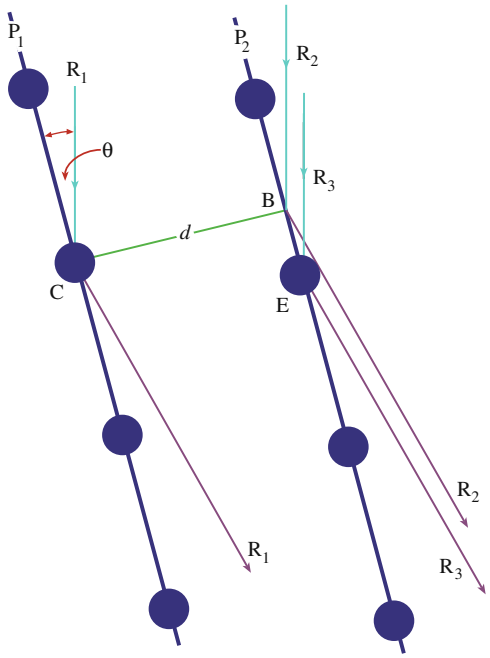


FIGURE 11.6. Scattering from three points on two planes. The path difference for scattering from points B and C is $2d\sin\theta$, so the path difference for scattering from points C and E is also $2d\sin\theta$. Hence, scattering in the direction of the diffracted beam from all points shown will be in phase if $2d\sin\theta = n\lambda$.

The Bragg reflection, **g**, is then perpendicular to the set of planes. Clearly this is just another way of expressing equation 11.11. Figures 11.2 and 11.7 remind us that Bragg diffraction occurs when **K** has the value **g**.

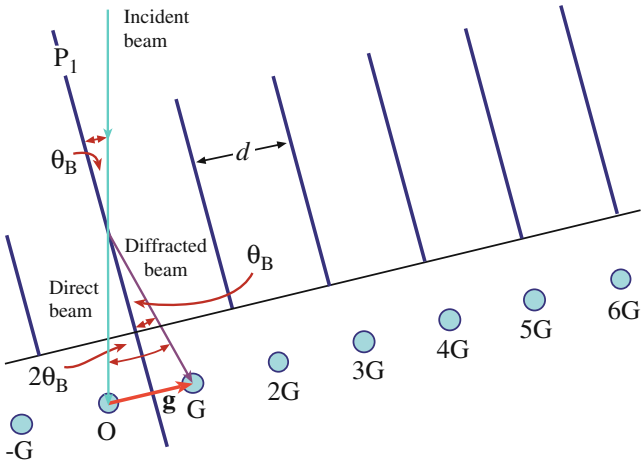


FIGURE 11.7. Diffraction from a set of planes a distance d apart. The planes have been oriented to be in the Bragg diffracting condition (θ_B is the incident angle). Note that the planes are not parallel to the incident beam. The resultant diffraction spots (reciprocal-lattice points) are labeled G, 2G, etc. The vector **g** from the origin (O) to the first diffraction spot G is normal to the diffracting plane.

THE BRAGG ANGLE

Be very careful with this definition; it is too easy to assume that the Bragg angle, θ_B, is the scattering semiangle: it is not! In all other cases, θ really is the semiangle.

11.5 MEANING OF *n* IN BRAGG'S LAW

As is shown in Figure 11.7, and in the DP in Figure 11.1, in practice there will not just be one Bragg reflection but a series of reflections which are periodically spaced along a line; these are known as a *systematic row* of reflections, –G, 0, G, 2G, 3G, etc., with corresponding diffraction vectors, $\bar{\mathbf{g}}$, **0**, **g**, 2**g**, 3**g**, etc.

Notation: When discussing beams in diffraction patterns, the letter O will refer to the 'direct' beam which is present even when there is no specimen; the letter G (not bold—it's not a vector) will refer to any single diffracted beam; the number **0** (bold) will refer to the diffraction vector for beam O (it is a vector of zero length) and the letter **g** (always bold to remind us that it is a vector) will denote the diffraction vector (in the DP) for beam G. Having said that, many microscopists use G and **g** interchangeably, so beware.

BAR g

The vector $\bar{\mathbf{g}}$ is pronounced "bar g" and connects O to −G, which is pronounced "minus g"! You will also hear $\bar{\mathbf{g}}$ pronounced "g bar."

These other reflections ($n\mathbf{g}$, where $n \neq 1$), called higher-order reflections, are particularly important in TEM. Pictorially, you can imagine them as arising from the interference of electrons scattered from planes which are a distance nd apart, where n is a rational fraction. To understand the physical meaning of this statement, put a plane P_3 half way between P_1, and P_2 as shown in Figure 11.8.

Now planes P_1, P_2, and P_3 will scatter in phase when

$$2\left(\frac{d}{2}\right)\sin\theta = \lambda \qquad (11.12)$$

because the new 'd' is $d/2$. Thus coherent scattering will occur when

$$|\mathbf{g}_2| = \frac{2}{d} \qquad (11.13)$$

i.e., when

$$|\mathbf{g}_2| = 2|\mathbf{g}| \qquad (11.14)$$

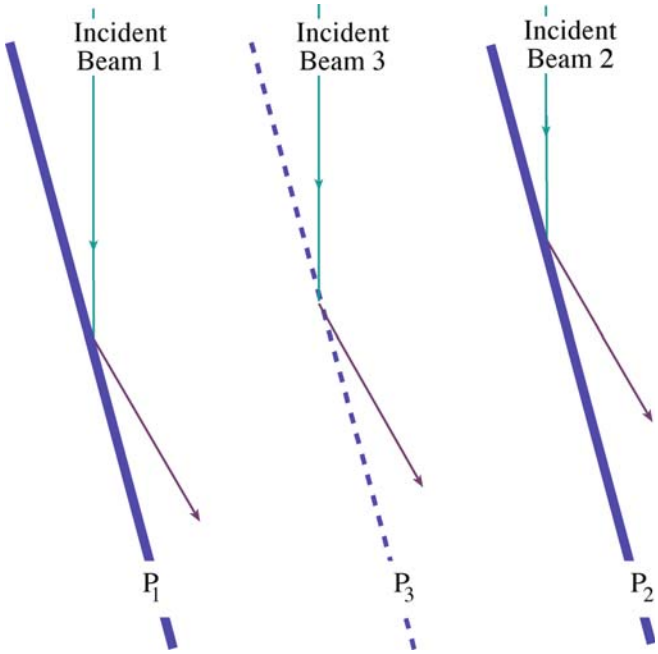


FIGURE 11.8. Scattering from three planes with plane P_3 positioned exactly half way between planes P_1 and P_2.

As we noted in the discussion of Figure 11.3, this scattering from plane P_3 will occur no matter how the atoms (scattering centers) are distributed on this plane—even if there are no atoms on the plane! Thus we will always see $\mathbf{g}_2 = 2\mathbf{g}$ and similarly $\mathbf{g}_3 = 3\mathbf{g}$, etc. So we can generalize equation 11.12 to be

$$2\left(\frac{d}{n}\right)\sin\theta = \lambda \tag{11.15}$$

or rewrite this as

$$2d\sin\theta = n\lambda \tag{11.16}$$

which gives a physical explanation for the n in equation 11.7.

To summarize: electrons are diffracting from a set of planes of spacing d such that we have both constructive and destructive interference. We can consider n in equation 11.15 as indicating that electrons are diffracting from a set of planes with spacing d/n rather than d. This equation can then be applied to planes that are occupied by different atoms. Although this treatment is not rigorous, it will prove to be very useful in practice. The alternative, but equivalent, view is obtained by considering the Laue equations which we will do in Chapter 12. You might consider why we don't have diffraction from planes which are nd apart instead of d/n.

11.6 A PICTORIAL INTRODUCTION TO DYNAMICAL EFFECTS

Dynamical diffraction traditionally strikes fear into the heart of the non-mathematician. Unfortunately, in TEM most practical imaging situations involve dynamical scattering. The terminology derives from X-ray theory (where it is not nearly so important). The reason it is very important in electron diffraction is that the electron beam interacts so strongly with the atoms in the crystal. For most purposes, dynamical diffraction can be thought of in quite a simple manner, as you can see in Figure 11.9.

DYNAMICAL DIFFRACTION

The beam, which has been strongly Bragg diffracted once, is necessarily in the perfect Bragg orientation to be diffracted back into the direct beam by the same set of planes.

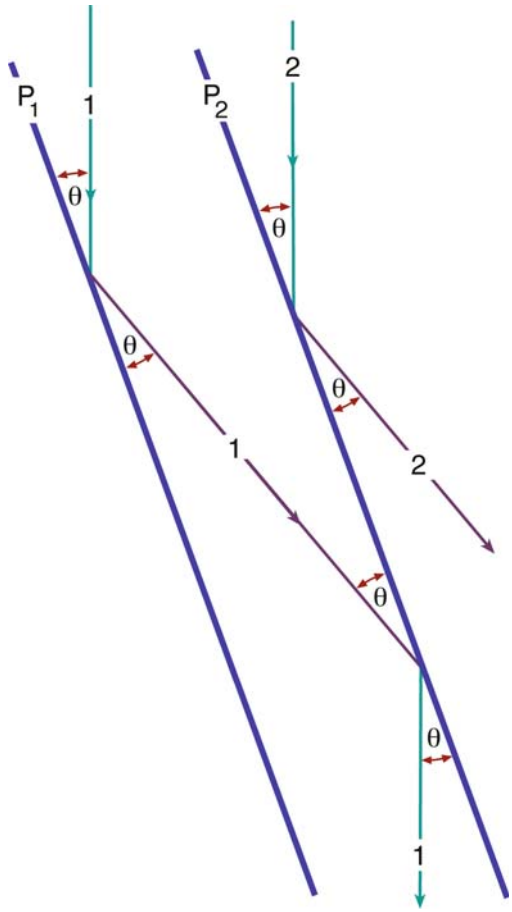


FIGURE 11.9. The beam can be scattered more than once. Any beam which is oriented so as to be Bragg scattered once is automatically in the ideal orientation to be rescattered. This gives rise to the phenomenon of dynamical scattering because the beam can be scattered again and again.

This beam is then said to be a rediffracted beam. The likelihood of this process occurring will increase as the thickness of the specimen increases. Clearly, the rediffracted beam is also perfectly oriented to be diffracted again, and so on. The two beams in Figure 11.9 are said to be dynamically coupled.

11.7 USE OF INDICES IN DIFFRACTION PATTERNS

In Chapter 18 we'll teach you how to index DPs, i.e., how to associate a spot in the DP with a diffracting plane in the specimen. For the time being it will be useful if we just introduce the conventions, rather than the methods, of indexing patterns.

First remember that a particular crystal plane is defined by the Miller indices (hkl); crystallographically equivalent (hkl) planes are denoted as the $\{hkl\}$ planes. We define the direct beam as the 000 reflection and each diffracted beam as a reflection with different hkl indices. It is a crystallographic convention to refer to the diffraction spot from a specific (hkl) plane, as hkl, i.e., without the parentheses. If we assign hkl to **g**, then the second order (2**g**) spot is $2h\ 2k\ 2l$, the $\underline{3}\mathbf{g}$ spot is $3h\ 3k\ 3l$, etc. Similarly, the $-\mathbf{g}$ reflection is $\overline{hkl}$. We'll discuss these points further in Section 12.3.

Now we can explain why we see so many spots in the DP. If we look along a zone axis in a crystal, we will see sets of planes in the edge-on orientation. Remember that a zone axis is the direction along the intersection of two or more planes.

> *Notation*: The zone axis, $[UVW]$, is a direction which is common to all the planes of the zone. So $[UVW]$ is perpendicular to the normal to the plane (hkl) if the plane is in the $[UVW]$ zone. Later, we will see that $[UVW]$ is defined as the incident beam direction. This result applies to all crystal systems and gives the Weiss zone law $hU + kV + lW = 0$.

If there are many planes close to the Bragg orientation, then we will see spots from many different planes. We still have not explained why we can see the 200 spot and the 400 spot in the same pattern: they clearly can't both satisfy the Bragg condition at the same time. We see both spots because of the physical shape of the TEM specimen, as we'll discuss in Chapters 12 and 17.

11.8 PRACTICAL ASPECTS OF DIFFRACTION-PATTERN FORMATION

Remember from Chapter 9, we can form diffraction patterns in the TEM in two complementary ways, SAD and CBED patterns.

SADPs are sharply focused spot patterns that we use to select reflections for all imaging modes. We can easily associate the sharp spots with our diffraction vectors, **g**.

CBDPs are arrays of disks. We can associate a **g** vector with each disk but the location of **g** requires more extensive consideration. For this reason, we'll delay more detailed discussion of CBDPs while we develop diffraction theory and then devote Chapters 20 and 21 to the topic, because it is very important.

11.9 MORE ON SELECTED-AREA DIFFRACTION PATTERNS

We discussed how you form a DP in the SAD mode in Chapter 9. Now we will discuss some of the practical implications and drawbacks of the method.

Why do we want to select a specific area to contribute to the DP? All foils are distorted to some extent so that diffraction conditions change as we move around the specimen; hence we need to select areas of constant orientation. Also, we may wish to determine the orientation relationship between two different crystals, which we can do by selecting the interfacial region. Alternatively, we

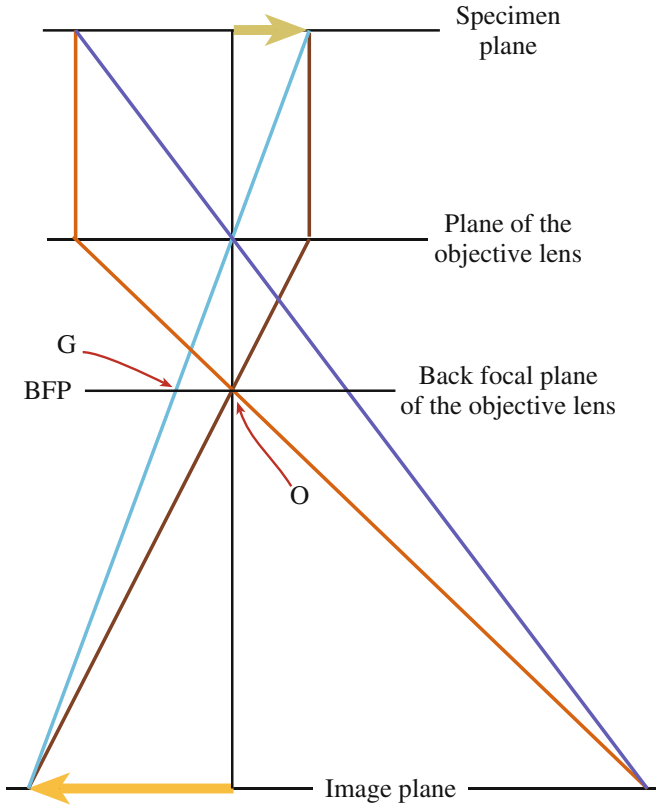


FIGURE 11.10. The diffraction pattern is formed at or close to the back focal plane of the objective lens. O is the direct beam and G is the diffracted beam.

may want to study the DP from a small particle within the foil. Figure 11.10 is a reminder that the DP is formed at the back focal plane (BFP) of the objective lens. A similar diagram was shown in Figure 9.13.

The SAD method for selecting an area is to place an aperture in the first image plane below the objective lens. In this case we really are selecting an area, which is the area in an image, but we always refer back to the volume of the diffracting specimen. Since we are working at an image, plane we do not need to focus the condenser lens, in fact we generally weaken (underfocus) this lens to give more parallel illumination so that all the rays are focused at the same plane, i.e., the BFP. The spots in the DP then become sharper. In practice, you will generally need to 'fine-tune' the focus of the DP since its focus depends on the excitation of the condenser lens.

The key practical steps in forming an SAD pattern were described in procedure #4 in Section 9.3.A. The two key points to remember are (i) be sure that you are at the eucentric focus position, with an image of the area of interest focused on the screen and (ii) remember to focus the DP with the intermediate lens (diffraction focus).

Remember that using an aperture to select an area in the image plane gives an additional advantage: the area has already been magnified, typically 25×. Thus a 50-μm diameter aperture will select a 2-μm diameter area on the specimen.

You might ask: why can't we just use a smaller SAD aperture to select a smaller area? We can provide the answer by looking at Figure 11.11: the objective lens is not perfect. As we saw in Chapter 6, the beams which are farther away from the optic axis are bent more strongly as they pass through the objective lens. For rays entering the objective lens at an angle β to the optic axis, the image formed at magnification M is translated a distance r_M given by

$$r_M = MC_s\beta^3 \tag{11.17}$$

So the area we select using the SAD aperture corresponds to the area PP_1 in the object plane *only* for the direct beam. The error increases as β increases, so that it's larger for a larger Bragg angle or for a larger **g**. The result is illustrated schematically in Figure 11.12 with values given in Table 11.1. (Note that we divide r_M by M to give the distance at the specimen.) The values in the middle column were calculated for a C_s of 3.3 mm and 100-keV electrons. If you use a smaller aperture, selecting an area of less than 1-μm diameter, even the fourth-order 111 reflection, i.e., the 444 reflection, from this area would not contribute to the SADP. Instead, a different area, possibly even an adjacent crystal, would contribute.

We will produce another selection error if the aperture is not located at the image plane. This effect can be

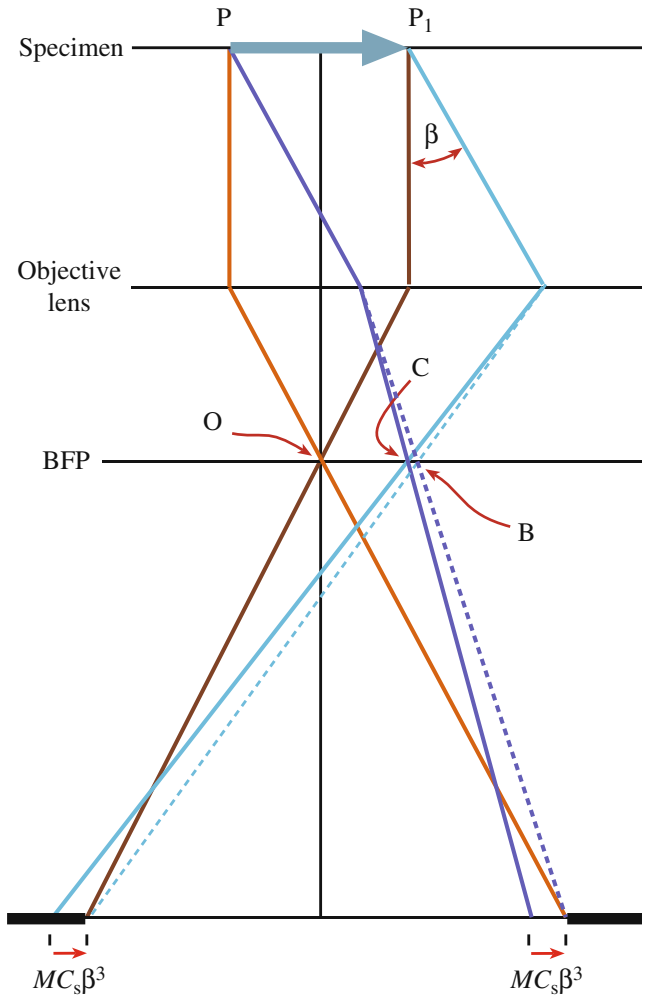


FIGURE 11.11. Formation of an SAD pattern showing that there is an error in selecting the area if the beams do not travel at the same angle to the optic axis. This difference is due to spherical aberration in the objective lens. B is the diffraction spot position for a perfect lens and C is the spot position with spherical aberration.

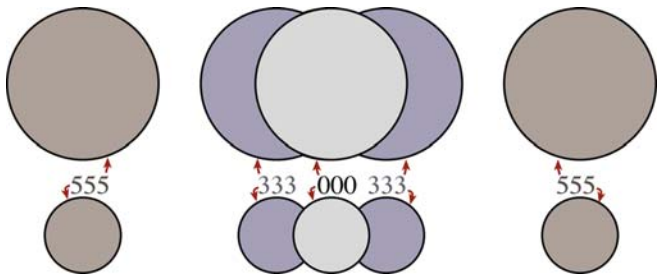


FIGURE 11.12. Schematic diagram showing the effective error in area selection, due to spherical aberration, for different reflections in the 111 systematic row for Al (a_0 = 0.404 nm) assuming 100-keV electrons and C_s = 3 mm. The 000 and 111 disks almost exactly overlap (the translation is 13 nm). The diameter of each disk in the top row is 1 μm, and the diameter of each disk in the bottom row is 0.5 μm.

TABLE 11.1 The Displacement Distance of the Image Formed by 'Reflection G' Due to Spherical Aberration

Reflection in SAD pattern	$C_s\beta^3$ (nm) old TEM	$C_s\beta^3$ (nm) now IVEM
111	13	1.2
222	100	9.1
333	350	31.9
444	760	69.3
555	1620	150
666	2800	250

seen clearly in Figure 11.13 where the objective lens is focused on plane P_f rather than on the specimen. The effect is seen by simple geometry if you extend the diffracted rays back to the specimen plane. The displacement at the first image plane (where the SAD aperture is located) corresponds to a distance y at the specimen plane where y is given by

$$y = D\beta \tag{11.18}$$

On some older machines, a 'click' on the medium image focus control (i.e., of the objective lens) corresponded to a change in focus, D of ~3 μm. You will still find on many TEMs that the aperture in the SAD plane is not always in focus when the DP is in focus. You might also consider the implications when we study very thick specimens. Remember that these two sources of error may be additive and therefore quite substantial.

You may still sometimes want to use an aperture which conventional wisdom tells you is 'too small for SAD.' Perhaps the best advice when this is the case is, if possible, use CBED. However, you should remember that 'conventional wisdom' is often based on Table 11.1, which was first calculated by Hirsch et al. in ~1963 and applied to a machine built in the 1950s! A modern 300-kV machine may have a C_s of ~1 mm and a λ (at 300 kV) of 1.968 pm. The values for $C_s\beta^3$ then become much smaller as shown in the right-hand column of Table 11.1. Clearly you could now use a much smaller SAD aperture, but 10 μm is about the smallest that can be manufactured.

One question which is often asked is: if the SAD aperture is placed at the first image plane, how can it affect the DP which is formed above it? The relationship between the SAD pattern and the image(s) can be illustrated by forming a multiple dark-field image of the type illustrated in Figure 11.14A. To do this, you must first form the SAD pattern in the usual way. Then increase the strength of the intermediate lens so that it's focused below the BFP in Figure 11.14B. Instead of a point we then see a disk, because the beam is convergent at the BFP. To understand what is

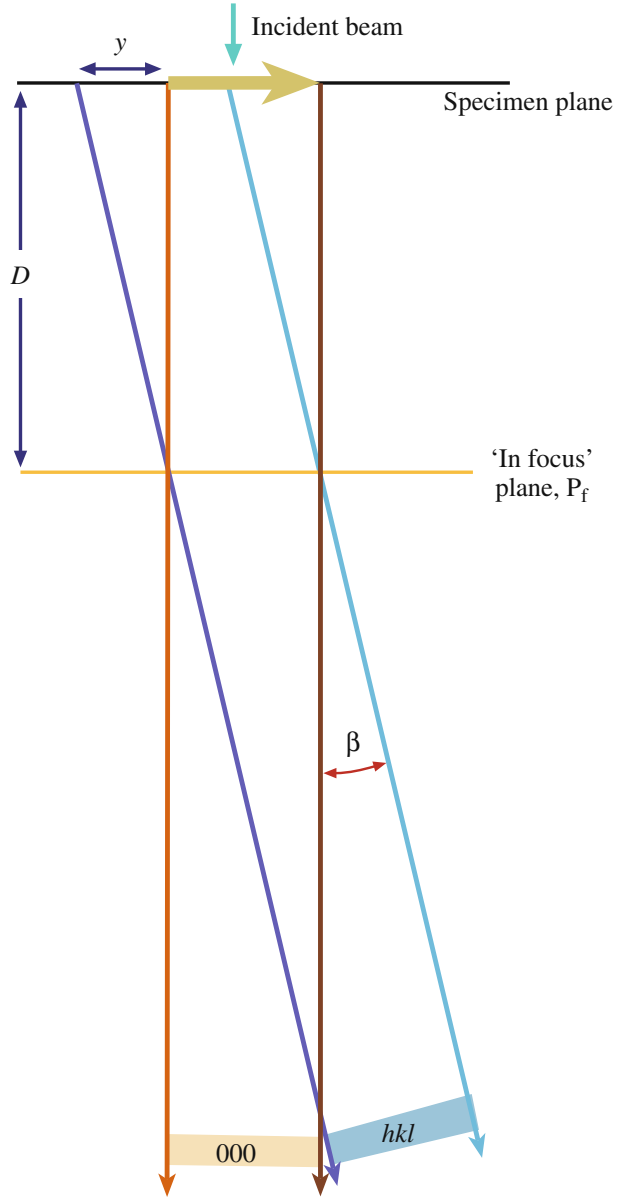


FIGURE 11.13. If the lens is not focused on the SAD plane, images associated with the different **g** vectors will be shifted with respect to one another. D is the defocus. The shift in the selected area is given by $y = D\beta$.

happening we must realize that the magnification of the specimen at the BFP is zero (i.e., when 'X' in Figure 11.14B is in the BFP plane)! As we increase the strength of the intermediate lens, staying in diffraction mode, we increase the magnification of these images (one bright-field image and many dark-field images). Of course, these images are not in focus but this can be corrected by adjusting the strength of the objective lens, which is just conventional focusing.

Now you can appreciate directly that each disk corresponds to a reflection in the SADP. The reflections that were bright now correspond to bright disks; the

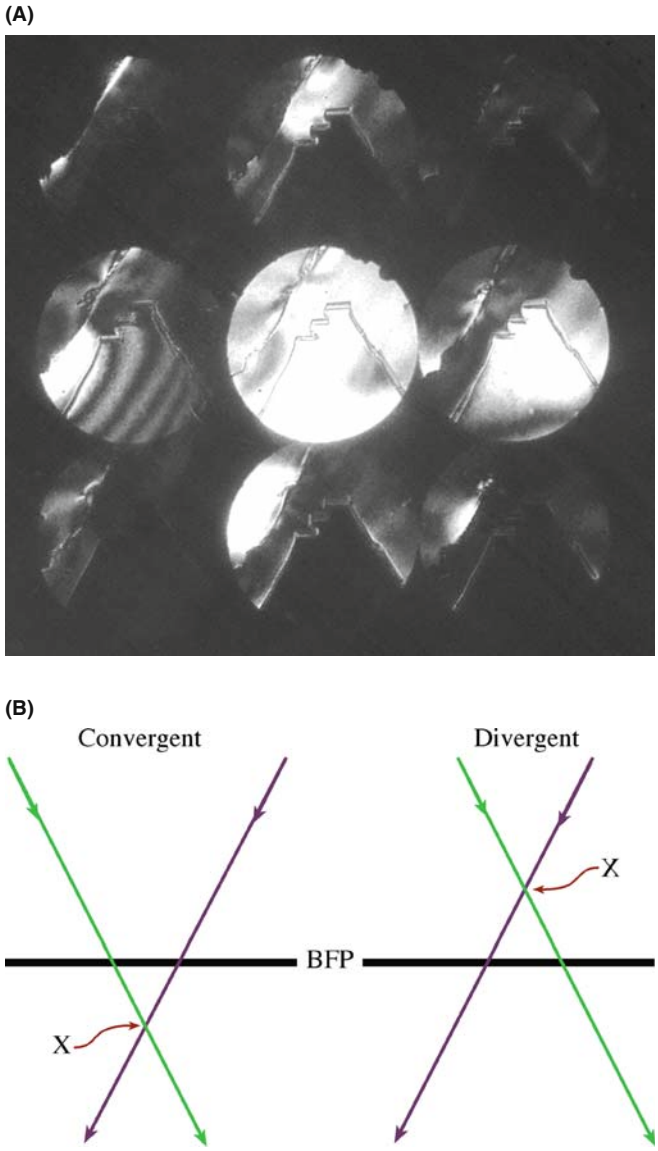


FIGURE 11.14. (A) Multiple dark-field images formed by defocusing the SADP revealing dark-field images in each diffraction disk defined by the objective aperture. Close inspection reveals that each image (of a twin boundary) is slightly shifted from the adjacent images, reflecting the increased error in area selection for higher-order reflections; (B) formation of a disk occurs because a defocused beam is either convergent or divergent at the BFP. An underfocused convergent beam is preferred, since it is more parallel than an overfocused divergent beam (see Figure 6.5).

area was close to the Bragg condition for that reflection. It is at first surprising to realize that none of the disks is uniformly bright. Conversely most of the disks are partly bright! We'll examine the reasons for this variation in Chapter 13.

This uncertainty in the area selected to form the SADPs is one reason that CBDPs can have some advantage whenever you want to get crystallographic information about specific regions of your specimen.

We'll end with some more practical points.

You can change the detail present in your DP simply by changing the C2 lens setting and the exposure time.

To record the SADP you should never use an exposure of <10 seconds (but you will). You don't need to use a 1-second exposure to limit drift! If you're interested in the details in the diffraction pattern you should take as many as three exposures; on film, you might use 10, 30 and 100 seconds. So spread the beam with C2 and remove that beam stop (better still, don't use it; you're damaging your specimen if the beam is that intense). Correct the astigmatism in the intermediate lens after you've spread the beam; this astigmatism becomes noticeable when your spots are small (not all microscopes allow you to do this). Focus the spots to sharp points with the diffraction (intermediate) lens; now you've focused the DP. Just for the exercise, focus the spots in the SADP as you generally view it with the beam condensed down to a minimum diameter. Now spread the beam with C2 and refocus the spots; you will see quite a difference in how sharp the spots are. Use the binoculars for focusing the spots after you've spread the beam. Unless the pattern is well focused, you will miss many of the fine details that make SAD so useful.

Deciding which pattern is best really depends on what information you require. If you would like to see fine detail in your SADP, you will probably need to underfocus the beam using C2. If the beam of interest is of low intensity, you may need to increase the exposure time at the risk of broadening the more intense spots, hence the need to record patterns with a range of exposures. DPs can be recorded on video or sent directly to the computer using a video camera. The use of a CCD camera can give a much greater range of intensities than the photographic film; this will become the preferred method of recording DPs in the future. However, you may have no alternative (i.e., no darkroom).

Cooling your specimen can reduce the thermal-diffuse scattering and thus reduce the background intensity considerably. Changes in the lattice parameter will not usually be a problem in SAD since we are not looking for that level of accuracy, but they will be noticeable in the HOLZ CBDPs (see Chapter 21).

Finally, if your specimen charges, you'll probably have to coat it with a thin film of carbon. Do practice this. Repeat several thin coatings if necessary and be sure that the charging is not due to your specimen not contacting the specimen holder or the holder not contacting the ground.

CHAPTER SUMMARY

DPs are the basis of all image formation in the TEM as well as all crystallographic analysis and defect characterization. We can understand DPs in terms of Bragg reflection from planes of atoms in the specimen and we can define the diffraction vector **g** associated with each Bragg reflection and associate each **g** with a crystal plane *hkl*. The diffracting planes are all in a specific zone axis *UVW*, which we can define as parallel to the incident beam direction.

DIFFRACTION AND OPTICS

Bragg, WL 1965 *The Crystalline State* **I** Ed. WL Bragg Cornell University Press Ithaca NY (first published in 1933). Find it in your library and browse.

Hecht, E 2001 *Optics* 4th Ed. Addison-Wesley Reading MA. Very readable and current.

James, RW 1965 *The Optical Principles of the Diffraction of X-Rays*, *The Crystalline State* **II** Ed. WL Bragg Ed. Cornell University Press Ithaca NY (first published in 1948). Ditto.

Schwartz, LH and Cohen, JB 1987 *Diffraction from Materials* 2nd Ed. Springer New York. One of the two standard texts.

THE COMPANION TEXT

We'll use this theory throughout the companion text but won't refer to it specifically.

SELF-ASSESSMENT QUESTIONS

Q11.1 List the three major differences between electron diffraction and X-ray diffraction.
Q11.2 If TEM imaging, coupled with electron diffraction, is so wonderful, why do some researchers use X-rays instead for both purposes?
Q11.3 Why do we usually not measure intensities of spots in TEM DPs?
Q11.4 Why do we need to specify planes in a lattice?
Q11.5 What are the major differences between SADPs and CBDPs?
Q11.6 Formation of an SADP shows that there are errors in selecting the area. Why do the errors occur?
Q11.7 Why do the spots in the DP vary as we move across the specimen?
Q11.8 Why do we need to change the C2 setting and exposure time as we change location on the specimen?
Q11.9 How can we limit thermal-diffuse scattering and thus reduce background intensity?
Q11.10 Summarize the advantages of TEM over SEM and visible-light microscopes.
Q11.11 What is necessary for constructive interference to occur?
Q11.12 Write down Bragg's law.
Q11.13 How does the specimen thickness affect dynamical scattering?
Q11.14 Why do you underfocus the condenser lens when obtaining SAD patterns?
Q11.15 Why don't you always use the smallest selected-area aperture possible when obtaining an SADP?
Q11.16 When recording an SADP pattern, how long should the exposure time be?
Q11.17 How does the SAD approach work if the aperture is not actually in the specimen plane?
Q11.18 What critical information can you readily obtain from a TEM DP?
Q11.19 In the case of Bragg diffraction, how is the semiangle between the incident beam and the Bragg-diffracted beam related to the Bragg angle?
Q11.20 What is a 'zone axis'?
Q11.21 Give a pictorial explanation for the meaning of *n* in Bragg's law.
Q11.22 What is the most basic characteristic of your specimen indicated by the presence of spots in a DP?
Q11.23 Prove that there is no path difference for scattering from atoms located on the same plane.

TEXT-SPECIFIC QUESTIONS

These questions are not difficult since we are just introducing DPs.

T11.1 Using equations in the book, deduce Table 11.1. Summarize all the parameters you use in making an extended version of the table.
T11.2 Examine Figure 11.14. If the specimen is Si, index the DP. If it's a 300-kV TEM, deduce the camera length.
T11.3 Explain why Figure 11.14 could not be from Cu.

T11.4 Consider Figure 11.10. There are five rays and four planes. Number them in the order in which you would draw these lines. Discuss how you would modify this figure so as to draw it to scale.

T11.5 The DP in Figure 11.1 was obtained from a polycrystalline sample. Identify the single-crystal DPs in this figure.

T11.6 In Figure 11.14, the disks all appear to show the same area of the specimen. Do you expect this to be exactly the same area? Discuss.

T11.7 Consider Figure 11.14. Using a reasonable value for the beam divergence, estimate the value for Δf using information in the figure.

T11.8 Consider Figure 11.12. Assess the accuracy of this drawing. Do a similar drawing for a value of $C_s = 0.7$ mm with everything else being the same.

T11.9 We will see later that a critical challenge for high-resolution imaging is image delocalization (MS Word suggests demoralization is a better word). How has this concept been introduced in Chapter 11? If we could correct C_s, how would SAD in the TEM be improved?

T11.10 Deduce the Weiss zone law using vector multiplication.

12

Thinking in Reciprocal Space

CHAPTER PREVIEW

In the previous chapter, you've already encountered vectors **k** and **g** and seen that they have lengths with units of nm^{-1}. These vectors are referred to as reciprocal-lattice vectors. Now we are going to discuss what this reciprocal lattice is. The reciprocal lattice is simply a lattice in reciprocal space. Note that this lattice is just as real as the 'real lattice' in 'real' space. It's like a new world in *Gulliver's Travels* but the relationship to 'our' world is not a linear scaling factor but a reciprocal one. If something (an object or a length) is large in real space, then it's small in reciprocal space.

When you see an object in real space you need to think, "What would it look like in reciprocal space?"

The reciprocal lattice is a purely geometrical construction. We'll separate the discussion into two parts: (i) the math and (ii) the properties of this lattice. The first is the same as you will meet in any text on condensed-matter physics; the second relates to how we use this construction in TEM. What we will find is that the lattice gives us a method for picturing the geometry of diffraction; it gives us a 'pictorial representation' of diffraction. It helps us to visualize how DPs will vary as the orientation and physical characteristics of the specimen vary.

12.1 WHY INTRODUCE ANOTHER LATTICE?

If you're new to the field of diffraction, the concept of reciprocal space may seem a daunting theoretical proposal. You must persevere. This model gives a physical picture of diffraction geometries that is extremely helpful to you, the experimentalist. The best approach is to think of any crystal as having two lattices. The first describes the arrangement of the unit cells of atoms in the crystal (your specimen). The second is an array of points which is uniquely defined for any given crystal but does not correspond to arrays of atoms; instead, each point is associated with a particular set of planes in the crystal. Of course, the reciprocal lattice is just as real as the 'real' lattice; both are simply geometrical constructions. We'll use the reciprocal lattice to give a physical picture of what happens when a crystal diffracts.

HISTORICAL NOTE

The reciprocal lattice was rediscovered independently by Ewald and Laue in 1911–1914, but it had been described by Gibbs in 1881 and by Bravais (in a somewhat less useful form) in 1850! The discussion in 1962 of Ewald's contribution to the subject is recommended reading.

In Chapter 11 we showed that Bragg diffraction of electrons by crystals occurs when **K** is equal to **g**. The reciprocal-lattice concept allows us to define a lattice where all the lattice points correspond to the possible **g** vectors.

RECIPROCAL

Think of any crystal as having two lattices, one real and the other reciprocal. In the reciprocal lattice, sets of parallel (*hkl*) atomic planes are represented by a single point located a distance $1/d_{hkl}$ from the lattice origin.

To understand why we use the reciprocal lattice, remember that we can always write Bragg's law (equations 11.2 and 11.3) as

$$\frac{2\sin\theta_B}{\lambda} = \frac{n}{d} = |\mathbf{K}| \qquad (12.1)$$

Thus the vector **K** is reciprocally related to d and vice versa. Before using this new lattice, however, we must work through its formal definition.

12.2 MATHEMATICAL DEFINITION OF THE RECIPROCAL LATTICE

In this section we will go through the definition of the reciprocal-lattice as a mathematical construction and prove some of the special mathematical properties of the vector, **g**. You don't need to learn the proofs but you will need to know these equations.

The mathematics of the reciprocal-lattice construction is simple vector algebra.

In real space, we can define any lattice vector, $\mathbf{r}_n$, by the equation

$$\mathbf{r}_n = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \tag{12.2}$$

where the vectors **a**, **b**, and **c** are the unit-cell translations in real space and n_1, n_2, and n_3 are all integers.

Any reciprocal-lattice vector, **r***, can be defined in a similar manner

$$\mathbf{r}^* = m_1\mathbf{a}^* + m_2\mathbf{b}^* + m_3\mathbf{c}^* \tag{12.3}$$

where **a** *, **b***, and **c*** are the unit-cell translations in reciprocal space and m_1, m_2, and m_3 are all integers. The directions of these new vectors are defined by the relations

$$\mathbf{a}^* \cdot \mathbf{b} = \mathbf{a}^* \cdot \mathbf{c} = \mathbf{b}^* \cdot \mathbf{c} = \mathbf{b}^* \cdot \mathbf{a} = \mathbf{c}^* \cdot \mathbf{a} = \mathbf{c}^* \cdot \mathbf{b} = 0 \tag{12.4}$$

In words, **a*** is normal to both **b** and **c**, etc.

We define the lengths of these vectors by another set of relations

$$\mathbf{a}^* \cdot \mathbf{a} = \mathbf{1};\ \mathbf{b}^* \cdot \mathbf{b} = \mathbf{1};\ \mathbf{c}^* \cdot \mathbf{c} = \mathbf{1} \tag{12.5}$$

Equation 12.5 then uniquely defines the length of the vector **a*** in terms of the length of the vector **a**. Therefore, these equations give the scale or dimension of the reciprocal lattice. The product of the projection of **a*** on the vector **a** multiplied by the length of **a** is unity. Be careful; this result does not mean that **a*** is parallel to **a** (think about this!). The direction of **a*** is actually completely defined by equation 12.4. It is perpendicular to both **b** and **c** and must therefore be the normal to the plane containing **b** and **c**.

The vector, **a***, is always perpendicular to the plane (100) even when **a** is not.

We can see that if **a**, **b**, and **c** are large, then the corresponding reciprocal-lattice vectors will be small if we choose conventionally shaped unit cells.

Since V_c, the volume of the unit cell, is given by $\mathbf{a}\cdot\mathbf{b}\wedge\mathbf{c}$, then from equation 12.5 we can write **a*** as

$$\mathbf{a}^* = \frac{\mathbf{b} \wedge \mathbf{c}}{V_c} \tag{12.6}$$

This definition emphasizes that the vector **a*** is orthogonal to the vectors **b** and **c**. However, just as **a**, **b** and **c** need not be normal to one another, **a***, **b***, and **c*** are also not necessarily normal to one another. We use the usual clockwise convention in defining the vector product in equation 12.6.

12.3 THE VECTOR g

We can generalize our definition of **g** a little more. Any vector in reciprocal space can be defined as a combination of the vectors **a***, **b***, and **c***. In particular, we can write **K** in a form for use later

$$\mathbf{K} = \xi\,\mathbf{a}^* + \eta\,\mathbf{b}^* + \zeta\,\mathbf{c}^* \tag{12.7}$$

Here ξ, η, and ζ are any three numbers, not necessarily integers. A particularly important reciprocal-lattice vector is the vector $\mathbf{g}_{hkl}$ which is defined as

$$\mathbf{g}_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{12.8}$$

where h, k, and l are all now integers and together define the plane (hkl).

The definition of the plane (hkl) is that it cuts the a, b, and c axes at $1/h$, $1/k$, and $1/l$, respectively. If you look at Figure 12.1, you'll see that the vector **AB** can be written as $\mathbf{b}/k - \mathbf{a}/h$. This vector and all vectors in the (hkl) plane

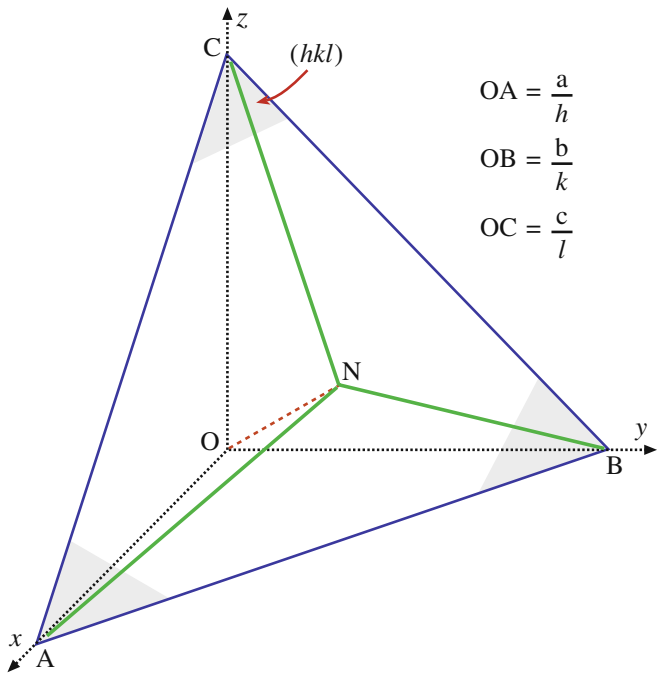


FIGURE 12.1. The plane ABC has Miller indices (hkl). The vectors **OA**, **OB**, and **OC** have lengths a/h, b/k, and c/l. The vector **ON**, which may be written as **n,** is normal to the plane (hkl). In the text we see that the reflection, **g**, which is associated with diffraction from the (hkl) planes, is parallel to **n** and normal to all vectors in (hkl).

are normal to the vector $\mathbf{g}_{hkl}$ defined in equation 12.8. You can prove this by taking the dot product of **AB** and **g** and using equations 12.4 and 12.5. Therefore, the vector $\mathbf{g}_{hkl}$ must be *normal* to the plane (*hkl*)

$$\left(\frac{\mathbf{b}}{k}-\frac{\mathbf{a}}{h}\right)\cdot(h\mathbf{a}^*+k\mathbf{b}^*+l\mathbf{c}^*)=0 \qquad (12.9)$$

The vectors **AB**, **BC**, and **CA** all lie in the plane (*hkl*) and each is normal to $\mathbf{g}_{hkl}$. All that we now have to prove is that the length of the vector, $|\mathbf{g}_{hkl}|$, is given by $(d_{hkl})^{-1}$. To show this relationship, consider a unit vector, **n**, normal to the plane (i.e., parallel to $\mathbf{g}_{hkl}$) and take the dot product with any unit vector inclined to this plane (e.g., $\mathbf{a}/h$ or $\mathbf{b}/k$).

The unit vector, **n**, parallel to **g** is simply $\mathbf{g}/|\mathbf{g}|$. Therefore, the shortest distance from the origin O to the plane is the dot product of **n** with vector **OB** (or **OC**, etc.)

$$\mathbf{n}\cdot\frac{\mathbf{a}}{h}=\frac{\mathbf{g}}{|\mathbf{g}|}\cdot\frac{\mathbf{a}}{h}=\frac{(h\mathbf{a}^*+k\mathbf{b}^*+l\mathbf{c}^*)}{|\mathbf{g}|}\cdot\frac{\mathbf{a}}{h}=\frac{1}{|\mathbf{g}|} \qquad (12.10)$$

where we again used equations 12.4 and 12.5. Since the origin, O, by definition lies on a plane in this family of planes, equation 12.10 gives the distance between parallel (*hkl*) planes, so that

$$d_{hkl}=\frac{1}{|\mathbf{g}|} \qquad (12.11)$$

as we required.

- The definition of the (*hkl*) indices is OA = a/h; OB = b/k; OC = c/l.
- The plane ABC can then be represented as (*hkl*).

We should emphasize a few points before moving on.

- Remember: the reciprocal lattice is so called because all lengths are in reciprocal units.
- If you are familiar with the derivation of band-gap concepts in elementary solid-state physics, you will have already used these ideas. The difference is that the energies of the electrons being produced in the microscope are $\geq$100 keV, whereas those in solids are $\sim$1 eV. This will affect the magnitudes of **k** but the **a***, etc., will not change with kV.
- Reciprocal-space notation. We introduced the use of brackets in Section 11.7. Now we'll extend this notation to the reciprocal lattice: (*hkl*) is shorthand notation for a particular vector in reciprocal space because it is normal to the (*hkl*) plane in real space; {*hkl*} is then the general form for these reciprocal-lattice vectors. [*UVW*] is a particular plane in reciprocal space, e.g., it may contain many {*hkl*} points so that in real space it would be a direction—the zone axis for the {*hkl*} real-space planes (see Table 12.1).

TABLE 12.1. Notation for Planes, Directions, and Reflections

Real space	Reciprocal space	
Particular direction	Particular plane	[*UVW*]
General direction	General plane	<*UVW*>
Particular plane	Particular direction	(*hkl*)
General plane	General direction	{*hkl*}
Diffracting plane	Indexed reflection	*hkl*

When indexing diffraction spots, you will often find that the brackets have been entirely omitted; this is a sort of convention. You should use brackets if there is any ambiguity or for emphasis.

- In non-cubic material, some special vectors may be parallel to one another, but most pairs will not be parallel. This difference can surprise even the experienced microscopist, particularly if you're used to studying cubic metals. For example, if you orient the electron beam to be along the [123] zone axis in an orthorhombic crystal such as olivine, the beam will *not* be normal to the (123) plane.

WARNING

Real-lattice vectors and reciprocal-lattice vectors with the same indices (e.g., [123] and plane normals (123)) are parallel *only* in the case of cubic materials.

12.4 THE LAUE EQUATIONS AND THEIR RELATION TO BRAGG'S LAW

To understand the value of the reciprocal lattice, we will now reconsider some of the terms we discussed previously. We use Bragg's law (Section 11.5) because it is so useful. It gives us a physical picture of the constructive-interference phenomenon, but it does not really correspond to the actual situation in TEM. Our justification in using Bragg's law is that we can derive it as a special form of the Laue equations, which really do describe diffraction in the TEM.

So we'll now derive Bragg's law from the Laue equations using simple vector algebra. For much of our discussion we assume that the crystal is infinitely large; we can always take the reciprocal lattice to be infinite. We can then use intuition to see that constructive interference will only occur when

$$\mathbf{K}=\mathbf{g} \qquad (12.12)$$

From Figure 12.2 we can see that the magnitude of **K** is always $(2\sin\theta)/\lambda$. At the Bragg condition it is also

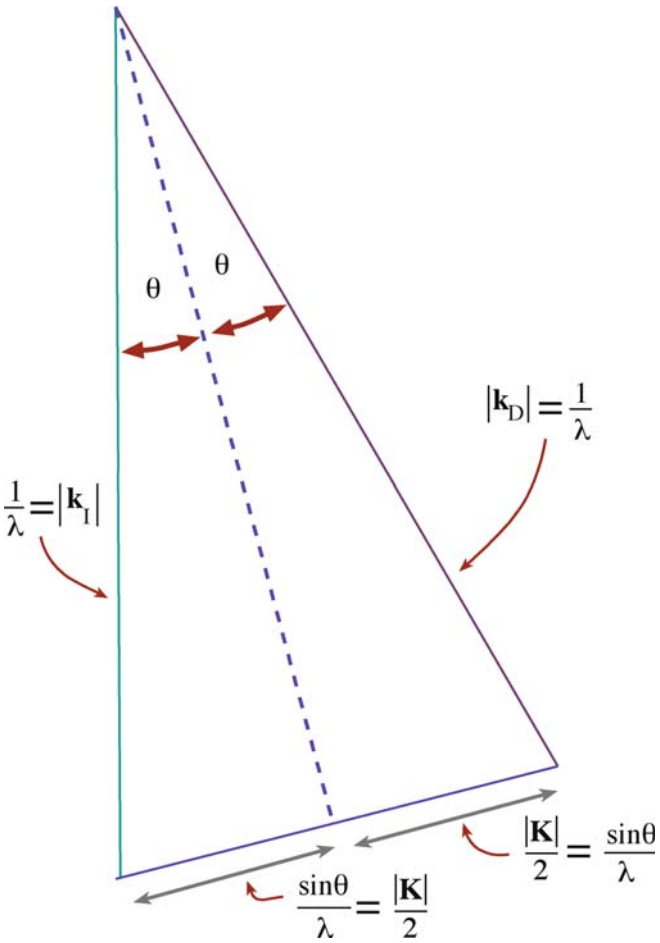


FIGURE 12.2. The geometric relationship between $\mathbf{k}_I$, $\mathbf{k}_D$, $\mathbf{K}$, θ, and λ.

equal to the magnitude of **g**, i.e., $1/d$. Therefore, at the Bragg condition we can write

$$\frac{2\sin\theta}{\lambda} = \frac{1}{d_{hkl}} \tag{12.13}$$

Hence we can write

$$\lambda = 2d\sin\theta \tag{12.14}$$

which is Bragg's law.

Equation 12.12 represents the Laue conditions for constructive interference; so we will refer to this as the condition for Laue, or Bragg, diffraction. Prove for yourself that $\mathbf{g}\cdot\mathbf{r}_n$ is always an integer, N. Then we can use equation 12.2 to write the Laue conditions

$$\mathbf{K}\cdot\mathbf{r}_n = N \tag{12.15}$$

This equation tells us that we must satisfy certain conditions on **K** in order to have Bragg (or Laue) diffraction.

Using equation 12.7 and multiplying this dot product we can see that this equation only holds when $\{n_1\xi + n_2\eta + n_3\zeta\}$ is an *integer*; $\mathbf{K}\cdot\mathbf{r}_n = N$ when ξ, η, and ζ are the integers h, k, and l.

Note: this is a very special case. By setting $\mathbf{r}_n$ equal to the three unit vectors in turn, equation 12.15 gives three relationships

$$\mathbf{K}\cdot\mathbf{a} = h \tag{12.16}$$

$$\mathbf{K}\cdot\mathbf{b} = k \tag{12.17}$$

$$\mathbf{K}\cdot\mathbf{c} = l \tag{12.18}$$

Of course, these equations are the same Laue diffraction conditions which we introduced back in Section 3.10.B, as given in equation 12.15. In Section 11.5 we quoted Bragg's law, with an 'n,' as

$$n\lambda = 2d\sin\theta \tag{12.19}$$

We also discussed the physical reason for n. We can now treat the same situation mathematically. If the integers h, k, and l have a common factor then we can write

$$nd_{nh,nk,nl} = d_{hkl} \tag{12.20}$$

So the n is implicit in the d used in equation 12.14. You will find that there are many other methods for treating this problem. We have chosen this approach to emphasize the underlying geometric principles.

12.5 THE EWALD SPHERE OF REFLECTION

The reciprocal lattice is a 3D array of points, each of which we will now associate with a reciprocal-lattice rod, or relrod for short, which is centered on the point. Furthermore, we will arrange each relrod to be normal to the thin foil, but to have a finite thickness parallel to this foil normal. This geometry of the relrods holds even when we tilt the specimen. The fact that we have rods is the result of the shape of our TEM specimen. At this stage this is purely an empirical construction to allow us to explain why we see spots in the DP even when the Bragg condition is not exactly satisfied. We will examine the shape of these rods and their origin in Chapter 16.

We now construct a sphere of radius $1/\lambda$. The sphere is known as the sphere of reflection or generally, and more simply, the 'Ewald sphere' in honor of its inventor P.P. Ewald. Due to Ewald's German origins, Ewald is pronounced 'A. Valt' rather than 'E. Walled.' Ewald's paper which first described the sphere was published in 1913 and was entitled 'Contributions to the Theory of Interferences of X-rays in Crystals.' It appears, in translation, in the monograph edited by Cruickshank et al., along with several of his other papers; the articles collected in this review

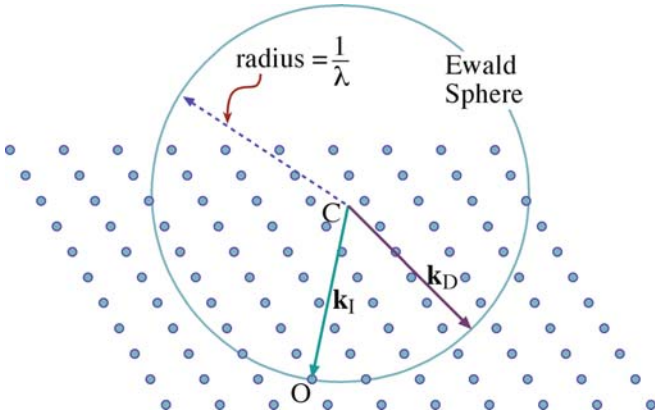


FIGURE 12.3. The Ewald sphere of reflection is shown intersecting a non-cubic array of reciprocal-lattice points. The vector **CO** represents $\mathbf{k}_I$, the wave vector of the incident wave, and O is the origin of the reciprocal lattice. $\mathbf{k}_D$ is any radius vector. When the radius of the sphere is similar to the spacing between the points in the reciprocal lattice, as is the case for X-rays, the sphere can only intersect a few points. When λ is much smaller, as for 100-keV electrons, the radius is much larger, the sphere is flatter, and it intersects many more points.

give a wonderful insight into the whole development of the theory of diffraction.

The sphere is usually represented in two dimensions by a circle and in most figures is drawn together with a two-dimensional section through the reciprocal lattice as shown in Figure 12.3.

The key point is that when the sphere cuts through the reciprocal-lattice point the Bragg condition is satisfied. When it cuts through a rod you still see a diffraction spot, even though the Bragg condition is not satisfied.

We combine the concept of the reciprocal lattice, the relrods, and the Ewald-sphere construction to picture how the intensity of each diffracted beam varies as we tilt the specimen or the electron beam. You may see the position of a spot in the DP move when the Ewald sphere is moved relative to the reciprocal lattice.

We can draw a sphere of radius $1/\lambda$ in reciprocal space so that it passes through the origin of the reciprocal lattice, point O, as defined in Chapter 11. If any point in the reciprocal lattice intersects the surface of the sphere, the set of planes corresponding to that point must satisfy the Bragg equation and hence the planes will diffract strongly. Equation 12.11 suggests that we define a vector **g** which can represent the quantity d^{-1}. The vector has a length and a direction. We choose the obvious length for **g** to be d^{-1} and make **g** the only unique vector for the plane (*hkl*), i.e., parallel to the normal to this plane.

Of course, the diagram drawn in Figure 12.3 shows a cut through the Ewald sphere. We usually draw such a diagram to include the vector describing the incident beam **CO** but this is not a requirement; in fact it is the exception, since our diagram is a two-dimensional cut through a 3D sphere. When we draw such a diagram we usually choose the plane of the diagram to contain the point O, since this point represents the direct beam. A common cause for confusion concerns the location of the center of the Ewald sphere, C. The point C is not the origin; the origin is the point O. In fact C will probably not coincide with a reciprocal-lattice point.

INTENSITY AND RECIPROCAL SPACE

We can associate an 'intensity' with any position in reciprocal space, and in particular with any position along one of these rods.

The value for this intensity is such that if the Ewald sphere cuts through that point in reciprocal space, then the diffracted beam, **g**, will have that intensity.

In general, if the Ewald sphere moves, the intensity will change. The important idea to keep in mind is that the reciprocal lattice is just a construction we use to give us a pictorial way of looking at diffraction.

WHERE IS C?

The vector **CO** is $\mathbf{k}_I$ and has length $1/\lambda$; this defines where C is located, i.e., we start with O and measure back to C.

Now you can appreciate that it is only when the incident beam lies in our chosen plane that the vector **CO** will lie in that plane. For example, we may choose the plane to be parallel to the optic axis of the microscope but tilt the incident beam off this axis; in such cases we will still often be interested in the plane containing both the optic axis and the incident beam. Also notice that $\mathbf{k}_D$ could be any vector which begins at C and ends on the sphere.

Consider the relative dimensions of d_{hkl} and λ. We can see that for X-rays where λ is ~0.2 nm and $1/\lambda$ is ~5 nm^{-1}, the Ewald sphere can only intersect a small number of relrods because $1/d$ is only ~3 nm^{-1}. This explains why it is necessary in X-ray diffraction to use white radiation (giving a wide range of λ) or to oscillate, rotate, or powder the specimen (thus producing many variations of d and θ) in order to produce enough diffraction spots to analyze the structure. For 100-keV electrons, however, λ is 3.7 pm and $1/\lambda$ is 270 nm^{-1}. So the surface of the Ewald sphere is almost planar (but fortunately, as we will see in Section 12.6, not quite) in comparison with the array of reciprocal-lattice spots. Therefore, in a TEM, the Bragg condition is nearly satisfied for many planes and, as we saw in Figure 11.1, many diffraction spots are observed from a thin specimen corresponding to a section through the reciprocal lattice.

Rather than carry out the exercise of identifying arrays of spots for every orientation of the specimen, it is common practice to orient the specimen such that

the beam is incident almost parallel to a low-index zone (U, V, and W are all small numbers), and then to compare the observed zone-axis pattern (ZAP) with standard ones. We'll show you some standard patterns in Chapter 18. This approach is fine if you already know the crystal structure of your material. However, you'll need to know the full procedure if you have a material whose structure you don't know or if you are not able to rotate it to a low-index zone axis. This situation might arise, for example, when you are characterizing a grain boundary.

12.6 THE EXCITATION ERROR

We'll now introduce a new quantity, **s**, known as the excitation error or the deviation parameter. Always use these terms carefully! If the beam is exactly parallel to any zone axis then, according to the Laue conditions, there should be no spots in the DP. Clearly there are many spots (e.g., see Figures 1.6 and 2.13B), so there is intensity in the diffracted beams even when the Bragg condition is not exactly satisfied. The actual intensity will depend on how far we are away from the Bragg condition. This distance is measured by a vector, **s**, in reciprocal space such that

$$\mathbf{K} = \mathbf{g} + \mathbf{s} \tag{12.21}$$

The Ewald sphere intersects the reciprocal-lattice point at the center of a relrod when $\mathbf{s} = 0$. Equation 12.21 is very imprecise! Although **g** is well defined, **K** is not, because it depends on $\mathbf{k}_D$, which could be any vector terminating on the Ewald sphere. In Figure 12.4, we show two special values of **s** by choosing two special values of $\mathbf{k}_D$. In one, $\mathbf{k}_D$ lies along the vector **CG** so $\mathbf{s}_c$ is also parallel to **CD**; in the second, $\mathbf{s}_z$ is chosen to be parallel to vector **CO**, the incident-wave vector. A third special situation would be to define $\mathbf{s}_m$ as being perpendicular to the surface of the specimen, but we don't know where that is. Actually, we will often assume that $\mathbf{s}_m$ is perpendicular to **OG**, but this need not be the case. We will refer to **s** in several ways: $\mathbf{s}_g$ will emphasize that **s** is defined for a particular **g** while $\mathbf{s}_z$ will emphasize that **s** lies along the z-axis, which often corresponds to the incident-beam direction and the foil normal. We write **s** when we are not being specific.

THE VECTOR s

This vector, **s**, is a measure of how far we deviate from the exact Bragg condition.

When we drew Figure 12.4, you noticed that we placed the point G outside the Ewald sphere. Note that we are using G to emphasize that we are referring to the point, not the vector, **g**, from the origin to the point. In

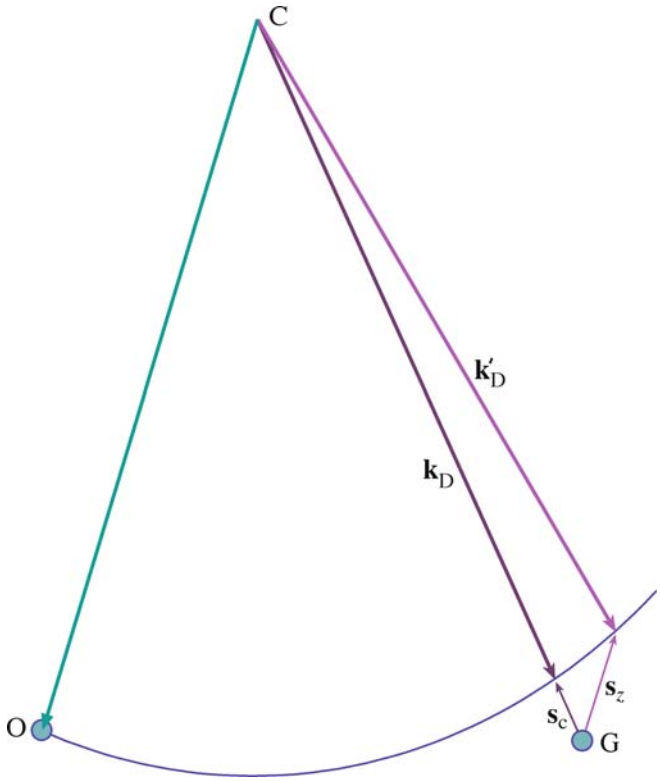


FIGURE 12.4. Two special values of **s** are illustrated. When $\mathbf{k}_D$ lies along **CG** then $\mathbf{s}_c$ is parallel to **CD**. Alternatively, we can choose **s** to be parallel to the incident beam direction **CO**; then $\mathbf{s} = \mathbf{s}_z$. and $\mathbf{k}_D$ becomes $\mathbf{k}'_D$. In each case, $\mathbf{k}_D$ ends on the Ewald sphere.

Figure 12.4, the row of reciprocal-lattice points is essentially at 90° to the incident beam. If we take all such rows, we define a plane of points (only G is shown) which are all at 90° to the incident beam. This plane of points is called the zero-order Laue zone (ZOLZ). We can now number all the planes of points which are parallel to the ZOLZ but do not contain the point O, and call these the higher-order Laue zones or HOLZ. The first of these (going toward C) is the FOLZ, the second is the SOLZ, and the rest are just HOLZ.

SIGN CONVENTION

We define the sign of **s** to be negative when G is outside the sphere, while **s** is positive when G is inside the Ewald sphere.

If we now draw the Ewald sphere as shown in Figure 12.5, you can see that it will intersect points in the FOLZ and other HOLZ. We'll see examples of these kinds of DPs in Chapters 20 and 21.

We can change the value of **s** in two ways.

- First, if we tilt the specimen, the row of spots moves but the Ewald sphere does not.
- Second, if we tilt the beam above the specimen, the Ewald sphere moves, because $\mathbf{k}_I$ tilts, because C moves!

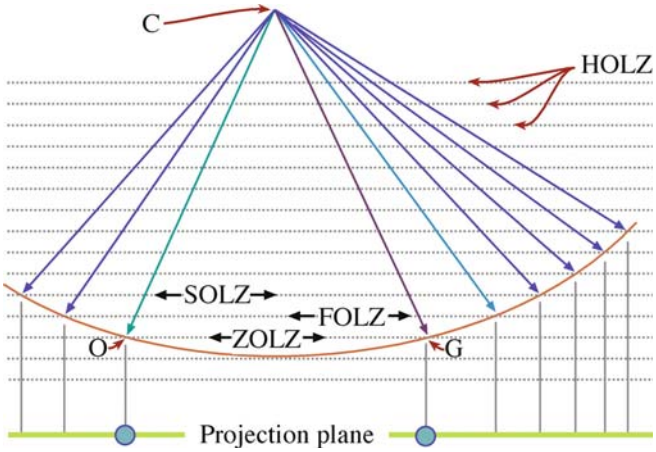


FIGURE 12.5. The Ewald sphere intercepts points in higher-order Laue zones (HOLZ) at large angles to the incident-beam direction. If the radius of the sphere increases (higher kV beam) then the sphere flattens and the HOLZ interception is at still larger angles.

Convince yourself of this. The DPs with different values of **s** may appear identical, but be cautious (more about this in the next chapter). The difference between these two processes is shown in Figure 12.6.

We'll conclude this section by giving you an experimental DP to think about. Figure 12.7 is from a slightly misoriented twin boundary: all you need to know is that different grains are diffracting to give two different DPs.

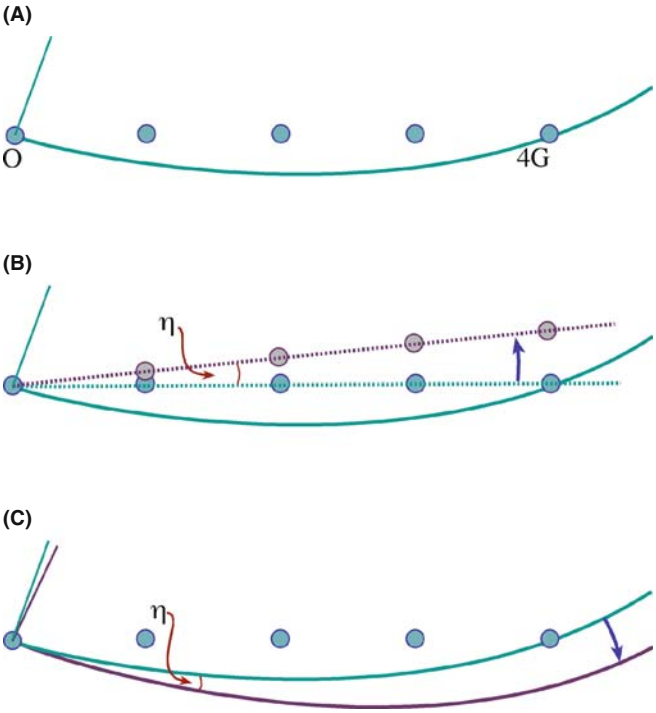


FIGURE 12.6. In (A) $\mathbf{s}_z = 0$ for 4G. We can change $\mathbf{s}_z$ in two ways; (B) if we tilt the specimen through angle η, the row of spots moves inside the sphere; (C) if we tilt the beam through η above the specimen, in the opposite direction, the sphere moves outside the row of spots.

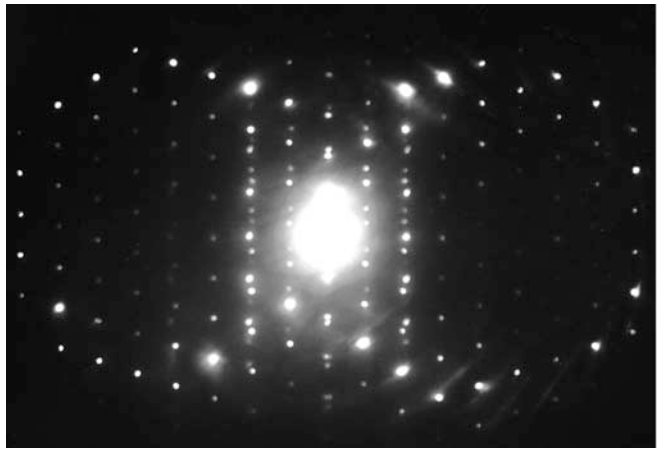


FIGURE 12.7. DP taken across a near-twin boundary in $MgAl_2O_4$ spinel. The rings of bright spots show where the Ewald sphere intercepts the reciprocal lattice of the crystals on either side of the twin boundary.

You can identify a ring of bright spots from each crystal. The question is: why are the rings displaced from one another? Yes, you're right, there is much more to this pattern than first meets the eye, as we'll see in Section 17.3.

12.7 THIN-FOIL EFFECT AND THE EFFECT OF ACCELERATING VOLTAGE

We will return to this topic in detail in Chapter 17 after we've examined a little more of the underlying theory. Here, we will briefly remind you that the radius of the Ewald sphere changes as we change kV. As the kV increases, the surface of the sphere becomes flatter. In a way, we were lucky with the initial choice of 100-keV electrons for TEMs since the sphere for 100-keV electrons has a very useful curvature. How does this curvature affect the DP? Well, we know that $\mathbf{k}_I - \mathbf{k}_D = \mathbf{K} = \mathbf{g}$ where $|\mathbf{g}|$ is d^{-1}. Therefore, **g** does not change as we change λ. Since d does not change but λ does, then Bragg's law tells us that θ must decrease as the kV increases. Therefore, if you keep the camera length constant, it will appear that the length of **g** in the DP decreases as λ decreases. Notice that the key word here is 'appear.' If you look back at

TABLE 12.2. Particular Values of λ and λ^{-1} as a Function of Beam Energy

E	λ (pm)	Radius, λ^{-1} (nm^{-1})	$(v/c)^2$
100 keV	3.701	270.2	0.3005
120 keV	3.349	298.6	0.3441
200 keV	2.508	398.7	0.4834
300 keV	1.969	508.0	0.6030
400 keV	1.644	608.3	0.6853
1 MeV	0.8719	1147.	0.8856

Section 9.6.B, you'll realize that the problem is that you must recalibrate the camera length for the new accelerating voltage.

The specimen is unchanged so the reciprocal lattice is the same. However, as the kV increases, the radius of the Ewald sphere increases and the diffraction spots appear to move closer together.

What is very important for TEM is that because λ is small, the radius of the Ewald sphere, λ^{-1}, is large and hence the Ewald sphere is quite flat. Note that this is very different from what we find in LEED or a typical back-reflection Laue X-ray pattern. The result is that we see many spots in the DP. Some values of the radius of the Ewald sphere are given in Table 12.2. You'll find it a useful exercise to generate this table yourself using a spreadsheet. Use the values from Chapter 1: $m_0 = 9.109\times10^{-31}$ kg, $c = 2.998\times10^{8}$ m/s, $h = 6.626\times10^{-34}$ Nm s and $1\,eV = 1.602\times10^{-19}$ N m.

CHAPTER SUMMARY

When combined with the Ewald-sphere construction, the reciprocal lattice gives us a very simple way of thinking about diffraction. When the sphere exactly cuts through a point, Bragg's law or the Laue equations are exactly satisfied. When the sphere just misses a point, we define a distance **s** to quantify this excitation error. In other words, **s** is a measure of where we cut the relrod. Ideally you will become as familiar with tilting reciprocal lattices in space as you are with tilting real lattices in your specimen holder. Remember that the lattices are rigidly connected to one another: when one turns the other does by exactly the same amount. Although Lilliput does not exist, reciprocal space does—at least for the electron microscopist!

Keep in mind the geometry and the dimensions.

- The Ewald sphere has a radius of $1/\lambda$ and always passes through the point O in the reciprocal lattice.
- Reciprocal-lattice dimensions are nm^{-1} although you'll still see $Å^{-1}$. You know 10 Å = 1 nm; it's not as easy to remember that $1\,Å^{-1} = 10\,nm^{-1}$.

HISTORY

Cruickshank, DWJ, Juretschke, HJ, and Kato, N (Eds.) 1992 *PP Ewald and His Dynamical Theory of X-ray Diffraction* Oxford University Press New York. When you have time.

Ewald, PP 1962 *Fifty Years of X-ray Diffraction* NVA Oosthoek's Uitgeversmaatschappij Utrecht Netherlands. Includes a description of how Ewald developed his description of reciprocal space. PP Ewald was Hans Bethe's father-in-law.

RECIPROCAL SPACE

Cullity, BD and Stock, SR 2001 *Elements of X-Ray Diffraction* 3rd Ed. Addison-Wesley Reading MA. The latest edition of the standard known simply as 'Cullity.'

James, RW 1965 *The Optical Principles of the Diffraction of X-Rays, The Crystalline State* **II** Ed. WL Bragg Cornell University Press Ithaca NY (first published in 1948). Special.

Schwartz, LH and Cohen, JB 1987 *Diffraction from Materials*, 2nd Ed. Springer New York. An alternative to Cullity.

Suryanarayana, C and Norton, MG 1998 *X-Ray Diffraction: A Practical Approach* Springer New York. Practical and different.

THE COMPANION TEXT

The contents of this chapter are used throughout the companion text but not explicitly.

SELF-ASSESSMENT QUESTIONS

Q12.1 How does the size of an object in real space relate to its size in reciprocal space?
Q12.2 When are (*hkl*) and [*hkl*] parallel for all values of *h*, *k*, and *l*?
Q12.3 Why does the incident-beam direction always point toward the 000 reflection of reciprocal lattice in the Ewald-sphere construction?
Q12.4 Why does electron diffraction in the TEM show planes of the reciprocal lattice?
Q12.5 Define excitation error and explain why we want to change its value.
Q12.6 How are the radius and the surface of the Ewald sphere affected if kV increases?

Q12.7 What happens to the diffraction spots and the reciprocal lattice when kV increases?
Q12.8 How are the higher-order Laue (HOLZ) lines affected by increasing the kV?
Q12.9 Define what we mean by 'reciprocal lattice.'
Q12.10 Define the vector **g**.
Q12.11 What is the sphere of reflection and where is its center?
Q12.12 Define what we mean by 'zero-order Laue zone.'
Q12.13 What is the simple relationship between **K** and **g** that represents the Laue conditions for constructive interference and how does it relate to Bragg's law?
Q12.14 How is $d_{nh,\,nk,\,nl}$ related to d_{hkl}?
Q12.15 Is d_{123} the same as d_{321} for an orthorhombic crystal?
Q12.16 What happens to the Ewald sphere when the specimen is tilted?
Q12.17 What happens to the Ewald sphere when the incident beam is tilted?
Q12.18 In real space, the following notation [*UVW*] and (*hkl*) signifies a particular crystalline direction and a particular plane, respectively. What does the same notation mean in reciprocal space?
Q12.19 Define the reciprocal vector **a**.
Q12.20 What is the sign convention for the deviation parameter?

TEXT-SPECIFIC QUESTIONS

T12.1 Consider Figure 12.7. Determine the angle of misorientation between the two grains and the axis of rotation. Take the perfect twin alignment to be zero misorientation.
T12.2 Consider Figure 12.6A. If G is the 220 reflection for Ge and you are using a 200-kV TEM, what is the value of **s** for the O, G, 2G, and 3G reflections. Repeat the question for 400-kV electrons and comment on the difference.
T12.3 For question 12.2, what is the Bragg angle for the excited reflection? (Give your answer in degrees and radians.)
T12.4 Consider Figure 12.4. The two values of **s** differ significantly and the vectors are inclined to one another. Choosing the 220 reflection in Cu and 100-keV electrons, which effect (magnitude or direction) is likely to be more important?
T12.5 Consider Figure 12.5. How has this figure been exaggerated? Describe a material which could give the geometry shown here.
T12.6 Generate Table 12.2 given the data in the text.
T12.7 If G in Figure 12.6 is Cu 220, estimate λ (ignoring relativistic effects) and hence suggest the kV.
T12.8 Assuming two major grains give rise to the DP in Figure 12.7, identify them (a tracing is fine). Then explain why there are two different rings of bright spots.
T12.9 More challenging and specialized. Use historical sources to describe the relationship between Laue, Ewald, and the Braggs and to explain their interactions.

13

Diffracted Beams

CHAPTER PREVIEW

In Chapter 11 we discussed why diffraction occurs; in this chapter we give a more detailed mathematical treatment. It may be more detailed than you need at this stage. Diffraction is one of those phenomena that lends itself directly to a detailed mathematical modeling, but there is a danger: *don't become so engrossed in the math that you miss the principles involved; conversely, don't ignore the subject because it is mathematically daunting!* The topic of this chapter is one which causes major problems for many microscopists. The treatment we will follow is known as the 'dynamical theory.' Later we will make some gross simplifications, partly because this is instructive and partly because these simplifications do apply to some important special cases; the kinematical approximation is one such simplification. Many other texts begin with the so-called 'kinematical' treatment and then advance to the more realistic, more general dynamical case. We will not do this but we will introduce the words and assumptions in Chapter 27.

The main principle of dynamical scattering was discussed in Chapter 11: an electron beam can be strongly scattered by a set of planes of atoms. When these planes are suitably oriented with respect to the beam, they produce a diffracted beam. This diffracted beam can then be rediffracted by a second set of planes in the same specimen, and so on. The physical reason for this repeated, or dynamical, diffraction, is that the electron beam and the atoms in the crystal interact strongly due to Coulomb forces. (X-rays are much less strongly affected by atoms and are thus more likely to be only scattered once, i.e., kinematical scattering.) This repeated scattering between the diffracted beams and the direct beam is the persistent topic of this chapter.

If you have a strong background in physics, you may find the simplifications used in this treatment somewhat unsatisfactory because we should be considering Bloch waves in a periodic object (our crystalline sample). We will discuss the analysis of Bloch waves in Chapter 14. Remember that *experimentally* we will associate arrays of spots in DPs with Bragg beams. Then we will relate these beams to images. We see both images and 'beams' on the screen of the TEM.

In future chapters, we will always discuss the thickness of the specimen in terms of the *extinction distance*. This is a term which we introduce here as a *characteristic length* for a *particular diffracted beam*. So, even in a rigorous Bloch-wave analysis, it is still important to understand the origin of the terminology introduced here (Table 13.1). Remember that the reason for looking at these equations is that they are directly useful to you when you are using the microscope, because they *describe* both the intensity of the electron beam in DPs and the contrast seen in TEM images of crystalline materials.

13.1 WHY CALCULATE INTENSITIES?

In this chapter, we will consider only scattering from perfect, defect-free, crystalline materials.

Ultimately we want to understand the images we see in the microscope. The detail we see in these images is determined by the intensity of the electron beam or beams and this varies for different positions in the image. Our motivation for calculating the intensity of diffracted beams is, therefore, to understand contrast features in TEM images.

In general, the analysis of the intensity of diffracted beams in the TEM is not simple because a beam which is diffracted once will easily be rediffracted. We call this repeated diffraction 'dynamical diffraction.' In a perfect crystal, imagine dividing the crystal into two halves, one

TABLE 13.1 Terminology and Notation

Ψ^T at P	The *total* wave function of the electron beam as measured at a point P at the bottom of the specimen. This wave function is a solution to the Schrödinger equation both inside and outside the specimen. What interests us is not Ψ^T but $\phi_{\mathbf{g}}$ and $\phi_{\mathbf{0}}$
$\phi_{\mathbf{g}}$	The amplitude of the *diffracted* beam for reflection G. The intensity is $\lvert\phi_{\mathbf{g}}\rvert^2$
$\phi_{\mathbf{0}}$	The amplitude of the *direct* beam. Don't use the term 'transmitted' beam; all the beams we are studying are transmitted. Don't call it the 'forward-scattered' beam; diffracted beams can also be forward scattered. $\phi_{\mathbf{0}}$ is a special value of $\phi_{\mathbf{g}}$ for the case where $\mathbf{g} = \mathbf{0}$
θ	The angle between a particular set of lattice planes and the direction of the beam scattered constructively by those planes
θ_B	The Bragg angle; a specific value of θ when $\mathbf{s} = \mathbf{0}$
dz	The thickness of a diffracting slice. This thickness can be as small as we wish to make it; it is not limited to atomic planes
$\xi_{\mathbf{g}}$	A characteristic length for reflection **g**; it is called the *extinction distance*
D, G	D is a diffracted beam; G is a special D and indicates that it is a Bragg-diffracted beam (neither is bold) (see Section 11.5)
χ	The electron wave vector in a vacuum
k	The electron wave vector in the specimen

above the other. The upper half diffracts the direct beam. The lower half further diffracts the direct beam but also rediffracts the diffracted beam. Don't confuse this rediffraction with the term 'double diffraction,' which has a special meaning, described in Chapter 23. If instead of cutting the specimen into two, you cut the specimen into many thin slices, you have multiple, instead of just double, diffraction. We call this effect dynamical diffraction.

Because of dynamical diffraction, we cannot use the intensities of spots in electron DPs (except under very special conditions such as CBED) for structure determination, in the way that we use intensities in X-ray patterns. Actually, a more important practical consideration is that the intensity of the electron beam varies strongly as the thickness of the specimen changes; the thickness may change across distances which are much smaller (as small as 1.5 nm or less) than the lateral dimensions of the electron beam (typically >1 μm in the TEM imaging mode). As we will see in Chapters 24–27 when we discuss images, the beam intensity also changes when lattice defects are present which is why we can 'see' defects in the TEM.

13.2 THE APPROACH

The approach we take here is to develop the basic equations describing the diffraction process and to identify parameters which will be important in understanding the contrast in the image. The different images will then be discussed in Part 3.

Inside a crystalline material, we should think in terms of Bloch waves because only certain wave-propagation vectors are allowed in infinite periodic structures: fortunately you don't need to have a thorough understanding of Bloch waves to understand contrast features in the microscope. However, we will consider them in Chapter 14, because a full understanding of the fundamental principles of diffraction from crystals will require this knowledge. What we 'see' in a DP relates directly to 'beams' because the DP, whether in the microscope or on a print, is *outside* the crystal. In this chapter, we will follow the analysis of Chapter 11, considering the amplitudes of beams simply because this gives a good intuitive understanding of the images—what we *see* in the TEM is the intensity, which is directly related to the amplitude ($I \propto |\phi|^2$).

A NOTE ON TERMINOLOGY

In Figure 13.1 we have labeled both the diffracted beams and the spot in the DP, G_i ($i = 1, 2$, etc.). When discussing images we will often refer to $\mathbf{g}_1$, the diffraction vector for the beam G_1. Then colloquially we will call **g** the 'reflection **g**'; the origin for this terminology goes back to the diagram for Bragg diffraction: geometrically it looks like 'reflection.'

So, what do we need to calculate? We need to calculate the intensity of the beam at the exit surface of the specimen, e.g., at all points, such as P in Figure 13.1, because this becomes the 'image' after suitable magnification.

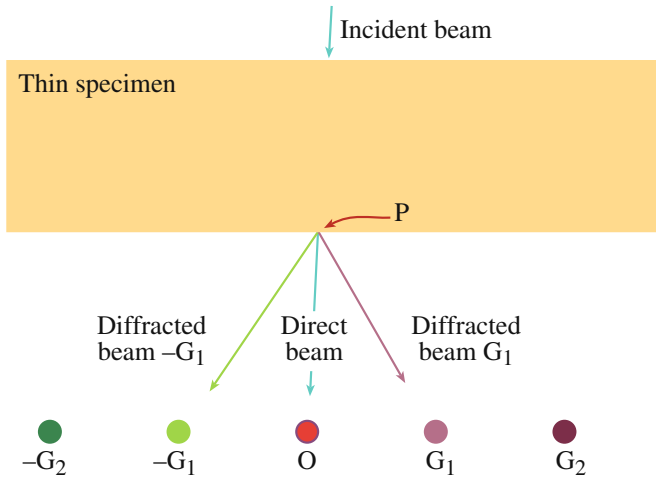


FIGURE 13.1. Defining the point P. The incident beam is scattered inside the thin sample. We want to know the intensities of the direct beam (O) and the diffracted (G_i) beams for each point P at the bottom surface of the specimen (the exit surface).

Before concluding this topic, we will briefly discuss the approximations we are making. One of the most important of these is the column approximation, which is introduced almost without being noticed. It is not a necessary assumption but it simplifies calculations and again aids intuitive understanding. You will also recognize many similarities to visible-light microscopy, but be wary, there are also many differences.

13.3 THE AMPLITUDE OF A DIFFRACTED BEAM

In the analysis of diffracted beams we will consider only crystalline materials. Since any crystal can be constructed by stacking unit cells, we begin by remembering the amplitude scattered by a single unit cell. We can rewrite equation 3.18 so that the amplitude of the electron beam scattered from a unit cell is

$$A_{\text{cell}} = \frac{e^{2\pi ikr}}{r} \sum_i f_i(\theta) e^{2\pi i \mathbf{K}.\mathbf{r}} \qquad (13.1)$$

where the summation is over all i atoms in the unit cell and θ is the angle at which the diffracted beam is traveling relative to the incident beam. We have included the term outside the summation because of how the wave propagates; the r^{-1} term is present because we have a constant flux of electrons traveling through an expanding spherical surface, radius r. The quantities **k**, **K,** and **r** were defined in Chapter 11 and $f(\theta)$ is the atomic-scattering factor. You will often see the sign of the exponent after $f(\theta)$ reversed. Unfortunately, there are two conventions! These conventions are discussed in Section 13.12; we will use the positive convention to be consistent with most materials-science texts.

Figure 13.2 reminds us that $\mathbf{K} = \mathbf{k}_D - \mathbf{k}_I$. The vectors **r** and $\mathbf{r}_i$ are different: **r** is the distance from a point P on the bottom of the specimen to the scattering center and $\mathbf{r}_i$ defines the position of an atom in the unit cell. Remember that $f_i(\theta)$ is the *scattering strength* for the 'i' atom ($f_i(\theta)$ is greater for Au than for Al, etc., as we saw in Figure 3.5). Since we are summing over all the atoms in the unit cell, we can rename this sum as $F(\theta)$, the *structure factor* of the unit cell. (It's just the scattering factor of the unit cell.) Notice that $F(\theta)$ depends on the nature of all the atoms in the unit cell, their positions, and the direction in which the beam is propagating (related to **K** and hence θ). Go back and check Section 3.9.

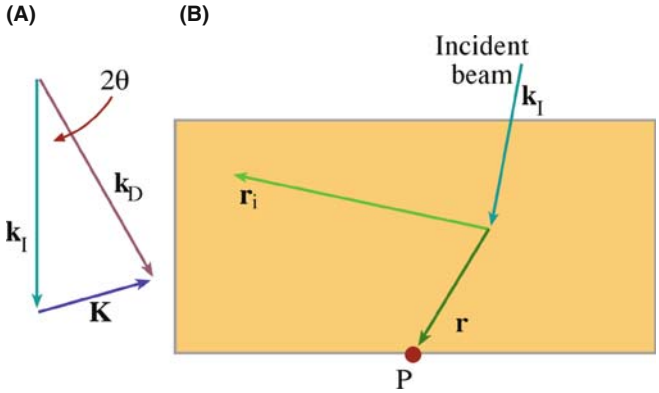


FIGURE 13.2. (A) A reminder that $\mathbf{K} = \mathbf{k}_D - \mathbf{k}_I$. The vector $\mathbf{k}_D$ represents the propagation vector for *any* wave. It does not have to be a diffracted beam but it will only give a spot in the DP when it does correspond to a diffracted beam. (B) shows the relation between the radius of the spherical wavefront, r, the position vector of the i^{th} atom, $\mathbf{r}_i$, and the point where the intensity is calculated, P.

Therefore, expression 13.1 can be rewritten as

$$A_{\text{cell}} = \frac{e^{2\pi i \mathbf{k}.\mathbf{r}}}{r} F(\theta) \qquad (13.2)$$

To find the intensity at some point P, we then sum over all the unit cells in the specimen. For simplicity, we will not solve this problem mathematically here but simply quote the result and discuss its meaning. Let's say we have n unit cells per unit area on a plane parallel to the crystal surface and a is the distance between these planes. The volume of a unit cell, V_c, is simply a/n. The amplitude in a *diffracted beam* (in the direction identified by θ) is denoted as $\phi_{\mathbf{g}}$ and is given by

$$\phi_{\mathbf{g}} = \frac{\pi a i}{\xi_{\mathbf{g}}} \sum_n e^{-2\pi i \mathbf{K}.\mathbf{r}_n} e^{-2\pi i \mathbf{k}_D.\mathbf{r}} \qquad (13.3)$$

Here $\mathbf{r}_n$ denotes the position of each unit cell. (Think about the signs in this equation.) In this analysis, the quantities $f(\theta)$ and $F(\theta)$ both have dimensions of length. We'll now explain what the length $\xi_{\mathbf{g}}$ means in equation 13.3; it is a length because $\phi_{\mathbf{g}}$, the scattering amplitude, is dimensionless (ξ is pronounced 'ksi,' rhyming with 'sigh').

The derivation of these equations involves some tricky manipulation which we will return to later. Some analyses actually make the unrealistic assumption that the intensity of the direct beam, $|\phi_{\mathbf{0}}|^2$, remains unchanged. This assumption is usually not justified, especially when the specimen has a finite thickness! If $|\phi_{\mathbf{g}}|^2$ is not zero then $|\phi_{\mathbf{0}}|^2$ cannot still be 1.

13.4 THE CHARACTERISTIC LENGTH $\xi_{\mathbf{g}}$

At this stage in our analysis it is best to think of the quantity $\xi_{\mathbf{g}}$ as a 'characteristic length' for the diffraction vector **g** so as not to have any preconceived ideas of what it represents. A detailed analysis shows that the magnitude of $\xi_{\mathbf{g}}$ can be expressed as

$$\xi_{\mathbf{g}} = \frac{\pi V_c \cos\theta_B}{\lambda F_{\mathbf{g}}} \qquad (13.4)$$

where $F_{\mathbf{g}}$ is the $F(\theta)$ for reflection **g** (i.e., $F_{\mathbf{g}}$ is a special value of $F(\theta)$ when θ is the Bragg angle).

The quantity ξ_g is an extremely important one; it gives us a way of thinking about nearly all diffraction-contrast phenomena. It is measured in nanometers (or Å) and is known as the 'extinction distance' for reasons that will become obvious.

From equation 13.4, you can see that the magnitude of ξ_g is related to F_g (and through V_c to the lattice parameter) and the wavelength of the electrons, λ. If the structure factor (F_g) is large, ξ_g will be small. Therefore, ξ_g will be small for Au but large for Si. F_g is large when the atomic number is large, because the Coulomb interactions are larger and $f(\theta)$ is large. Similarly, as the accelerating voltage is increased, ξ_g, for a particular material, will increase because the wavelength of the electrons decreases. Table 13.2 lists some useful extinction distances (all for 100-keV electrons).

SUMMARIZE ξ_g

ξ_g is the characteristic length for the diffraction vector **g.** We call it the **extinction distance** because of a property we'll discover later.

Note that ξ_g is a scalar quantity. ξ_g depends on the lattice parameters (through V_c), the atomic number (through F_g), and the kV used (through λ).

The effect of the lattice parameter on ξ_g is illustrated nicely by comparing values of ξ_{111} for diamond, Si, and Ge: the value for Si is larger than for Ge, as expected, because of the smaller atomic number but note that ξ_g for Si is also larger than that for diamond, which has a lower atomic number! Diamond has a particularly small lattice parameter, hence there are more atoms in a given volume.

13.5 THE HOWIE-WHELAN EQUATIONS

The direct and diffracted beams are detected outside the crystal and we see them on the viewing screen. Now we can think of the wave function inside the crystal as being the sum of the beams passing through the crystal. The direct beam has amplitude ϕ_0 (bold **0** to emphasize that the diffraction vector has zero length) and the amplitudes of the diffracted beams can be written as ϕ_{g_1}, ϕ_{g_2}, etc. Each beam has an appropriate phase factor. We write ψ^T, the total wave function, as a series

$$\psi^T = \phi_0 e^{2\pi i \chi_o \cdot \mathbf{r}} + \phi_{g_1} e^{2\pi i \chi_{G_1} \cdot \mathbf{r}} + \phi_{g_2} e^{2\pi i \chi_{G_2} \cdot \mathbf{r}} + \ldots\ldots \quad (13.5)$$

where the wave vectors are χ_O and χ_D (χ (chi) is pronounced 'kai' and rhymes with sky); χ_O is often written simply as χ. We use χ_O here to emphasize that it is a vector which terminates on the point O in reciprocal space; χ_{G_1} terminates on the 'point' G_1, etc. At this stage, we are using wave vectors χ_O and χ_D which describe the wave in the vacuum rather than in the crystal. We will change to being inside the crystal shortly. Most of the time, you could write χ as **k**, but there are occasions when the difference is important so we start with χ and then change over.

First, we simplify equation 13.5 by considering only one diffracted beam G, i.e., we make a 'two-beam approximation' (O is the other beam). This is a very important approximation, which we'll use often. Two-beam conditions mean that we tilt the crystal so there is only one strong diffracted beam (with $\mathbf{s} = 0$). All other diffracted beams are weak ($\mathbf{s} >> $ or $<< 0$), and we ignore their contribution to ϕ_g. Then if the amplitude ϕ_g changes by a small increment as the beam passes through a thin slice of material which is *dz* thick we can write down expressions for the *changes* in ϕ_g and ϕ_0 by using the concept introduced in equation 13.3 but replacing *a* by the short distance *dz*

$$d\phi_g = \left\{ \frac{\pi i}{\xi_g} \phi_0 e^{2\pi i (\chi_O - \chi_D) \cdot \mathbf{r}} + \frac{\pi i}{\xi_0} \phi_g \right\} dz \quad (13.6)$$

$$d\phi_0 = \left\{ \frac{\pi i}{\xi_0} \phi_0 + \frac{\pi i}{\xi_g} \phi_g e^{2\pi i (\chi_D - \chi_O) \cdot \mathbf{r}} \right\} dz \quad (13.7)$$

Here $\chi_O - \chi_D$ is the change in wave vector as the ϕ_g beam scatters into the ϕ_0 beam. Similarly $\chi_D - \chi_O$ is the change in wave vector as the ϕ_0 beam scatters into the ϕ_g beam. Now the *difference* $\chi_O - \chi_D$ is identical to $\mathbf{k}_O - \mathbf{k}_D$ although the individual terms are not equal. Then remember that $\mathbf{k}_D - \mathbf{k}_O$ ($= \mathbf{K}$) is $\mathbf{g} + \mathbf{s}$ for the perfect crystal.

You might wonder why we have introduced the wave vector χ when it appears to be the same as the **k** we used in equation 13.1. The reason is that equation 13.1 is a very general equation describing scattering from any group of atoms, but we are now going to consider two special cases, namely, an electron in the vacuum (wave vector χ) and one in a crystal (wave vector **k**). Incidentally, the excitation error, **s**, should really be written as $\mathbf{s}_g$, since it refers to a particular **g** vector. You can think

TABLE 13.2 Examples of Extinction Distances (in nm)*

Material					
hkl =	110	111	200	220	400
Al	–	56.3	68.5	114.4	202.4
Cu	–	28.6	32.6	47.3	76.4
Au	–	18.3	20.2	27.8	43.5
MgO	–	272.6	46.1	66.2	103.3
Fe	28.6	–	41.2	65.8	116.2
W	18.0	–	24.5	35.5	55.6
Diamond		47.6	–	66.5	121.5
Si		60.2	–	75.7	126.8
Ge		43.0	–	45.2	65.9

*For two-beam condition at 100 kV.

of the parameter ξ_0 as the characteristic length for forward scattering, i.e., scattering from any beam into itself, whereas ξ_g corresponds to scattering through an angle corresponding to a diffraction vector **g**.

The change in ϕ_g depends on the magnitude of **both** ϕ_g and ϕ_0.

These two equations (13.6 and 13.7) can then be rearranged to give a pair of coupled differential equations. We say that ϕ_0 and ϕ_g are 'dynamically coupled.' The term *dynamical diffraction* thus means that the amplitudes (and therefore the intensities) of the direct and diffracted beams are constantly changing, i.e., they are dynamic

$$\frac{d\phi_g}{dz} = \frac{\pi i}{\xi_g}\phi_0 e^{-2\pi i s z} + \frac{\pi i}{\xi_0}\phi_g \tag{13.8}$$

and

$$\frac{d\phi_0}{dz} = \frac{\pi i}{\xi_0}\phi_0 + \frac{\pi i}{\xi_g}\phi_g e^{2\pi i s z} \tag{13.9}$$

Microscopists usually refer to this pair of equations as the '*Howie-Whelan*' equations after Howie and Whelan (1961), who laid the foundations for understanding diffraction contrast in the TEM; you may also see them referred to as the 'Darwin-Howie-Whelan equations' since Darwin (1914) developed the dynamical theory for X-rays! Note that we are further simplifying the expression by writing

$$e^{-2\pi i \mathbf{s}\cdot\mathbf{r}} = e^{-2\pi i s z} \tag{13.10}$$

In doing so, we are making the approximation that **s** and **r** are both parallel to z, i.e., at this time, we ignore components of **s** that are not parallel to the electron beam. The approximation may be written as

$$|\mathbf{s_g}| = s_z \tag{13.11}$$

We then drop the z subscript; just remember it is still there. There are situations where the difference can become important.

Although this approach is totally phenomenological (i.e., we haven't really given any physical justification for the assumptions we have made and actually we know we should use Bloch waves), you will see that it provides enormous insight into the interpretation of your images and DPs. In Chapter 25, we will use these ideas to understand why we see defects in the TEM.

The fundamental idea is that, at any given position in the specimen, the change in the amplitudes of *both* the direct beam and the diffracted beam depends on the amplitude of *both* beams. The fact that part of the change in ϕ_0 is due to the magnitude of ϕ_0 itself, gives rise to the term *forward scattering*; remember the origin of scattering from Section 2.2. Note that scattering from ϕ_g to ϕ_g is also forward scattering, although it takes place in a different forward direction (i.e., $\theta = \theta_B$ and scattering is parallel to $\mathbf{k}_D$ rather than $\mathbf{k}_O$). So forward scattering does occur but it does not change the direction of the beam. However, it does have a characteristic length, ξ_0; this length is another way of saying we have a refractive-index effect for electrons which we'll address later in Section 14.4.

DIRECT

Remember: don't refer to the direct beam as the unscattered or the transmitted beam!

13.6 REFORMULATING THE HOWIE-WHELAN EQUATIONS

From here on, the math is quite straightforward. What we are going to do may seem like a lot of work to derive one equation (13.48) but the result will allow you to picture more clearly what is happening. If you don't want to bother with the math, you can skip to equations 13.47 and 13.48 but you must not miss those two equations; they are essential for understanding images of crystalline materials.

The pair of equations, 13.8 and 13.9, can be simplified by making the substitutions (i.e., a transformation of variables)

$$\phi_{0(\mathrm{sub})} = \phi_0 e^{\frac{-\pi i z}{\xi_0}} \tag{13.12}$$

and

$$\phi_{g(\mathrm{sub})} = \phi_g e^{2\pi i s z - \frac{\pi i z}{\xi_0}} \tag{13.13}$$

Then equations 13.8 and 13.9 become

$$\frac{d\phi_{g(\mathrm{sub})}}{dz} = \frac{\pi i}{\xi_g}\phi_{0(\mathrm{sub})} + 2\pi i s \phi_{g(\mathrm{sub})} \tag{13.14}$$

and

$$\frac{d\phi_{0(\mathrm{sub})}}{dz} = \frac{\pi i}{\xi_g}\phi_{g(\mathrm{sub})} \tag{13.15}$$

Since ϕ_0 and $\phi_{0(\mathrm{sub})}$ only differ by a phase factor, we will ignore the difference in calculating intensities since only the amplitude is then important; similarly for ϕ_g and $\phi_{g(\mathrm{sub})}$. The result of our substitution is that we have removed the phase factor involving ξ_0, i.e., we've removed the refractive-index effect. Equations 13.14 and 13.15 can be combined to give the second-order differential equation for ϕ_0

$$\frac{d^2\phi_0}{dz^2} - 2\pi i s\frac{d\phi_0}{dz} + \frac{\pi^2}{\xi_g^2}\phi_0 = 0 \qquad (13.16)$$

We can obtain a similar equation for ϕ_g and then obtain solutions for these reformulated expressions.

Note that the only other quantities appearing in this equation for ϕ_0 are z, s, and ξ_g: z and s are geometric parameters; the nature of the material only enters through ξ_g.

13.7 SOLVING THE HOWIE-WHELAN EQUATIONS

If we can solve the Howie-Whelan equations, then we can predict the intensities in the direct and diffracted beams (i.e., $|\phi_0|^2$ and $|\phi_g|^2$ in the two-beam case). If we take it step by step, then we know that solutions to equation 13.16 (a second-order differential equation in one variable, ϕ_0) must have the form

$$\phi_0 = C_0 e^{2\pi i\gamma z} \qquad (13.17a)$$

So we can write that

$$\frac{d\phi_0}{dz} = 2\pi i\gamma C_0 e^{2\pi i\gamma z} \qquad (13.17b)$$

and

$$\frac{d^2\phi_0}{dz^2} = -4\pi^2\gamma^2 C_0 e^{2\pi i\gamma z} \qquad (13.17c)$$

What we need to determine is the phase γ and the amplitude C_0. Note that since z is a distance in real space, then γ must be a distance in reciprocal space. Substituting this expression into equation 13.16 shows that γ must be a solution to the algebraic equation

$$\gamma^2 - s\gamma - \frac{\xi_g^{-2}}{4} = 0 \qquad (13.18)$$

Now ϕ_g is related to ϕ_0 through equation 13.15. By substituting equation 13.17a into equation 13.15 we find that for each ϕ_0, we also have a ϕ_g given by

$$\phi_g = 2\xi_g\gamma C_0 e^{2\pi i\gamma z} \qquad (13.19)$$

To emphasize the similarity to equation 13.17a we can define

$$\phi_g = C_g e^{2\pi i\gamma z} \qquad (13.20)$$

Then we can see directly that

$$\frac{C_g}{C_0} = 2\xi_g\gamma \qquad (13.21)$$

We've actually got this far without solving any equation! There are two solutions to the quadratic equation (13.18), using the standard formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad (13.22)$$

to give

$$\gamma^{(1)} = \frac{\left(s - \sqrt{s^2 + \frac{1}{\xi_g^2}}\right)}{2} \qquad (13.23a)$$

and

$$\gamma^{(2)} = \frac{\left(s + \sqrt{s^2 + \frac{1}{\xi_g^2}}\right)}{2} \qquad (13.23b)$$

We have now found two solutions to the Howie-Whelan equations.

There are two different values for ϕ_0 and two corresponding values for ϕ_g.

Now we need to understand what these solutions mean physically. Specifically, what can we learn about $\gamma^{(1)}$ and $\gamma^{(2)}$? Note that they are always real but may be positive or negative depending on the sign and size of s and that they are *independent* of z.

13.8 THE IMPORTANCE OF $\gamma^{(1)}$ AND $\gamma^{(2)}$

Since $\gamma^{(1)}$ and $\gamma^{(2)}$ are solutions of equation 13.18 from the properties of quadratic equations or by combining equations 13.23a and b, we know that

$$\gamma^{(1)} + \gamma^{(2)} = s \qquad (13.24)$$

which is a purely geometric quantity, and

$$\gamma^{(1)} \times \gamma^{(2)} = -\frac{1}{4\xi_g^2} \qquad (13.25)$$

which is a property of the material. Remember that γ is a length in reciprocal space.

In order to make the equations easier to work with, it is useful to define another quantity, w, which is *dimensionless* but has the same sign as s.

$$w = s\xi_g \qquad (13.26)$$

In practical situations w may vary from 0 to ± 10. We can then express the two forms of equation 13.21 (because there are *two* values of γ) in terms of γ or, more conveniently, in terms of w

$$\frac{C_g^{(1)}}{C_0^{(1)}} = 2\xi_g\gamma^{(1)} = w - \sqrt{w^2 + 1} \qquad (13.27)$$

and

$$\frac{C_{\mathbf{g}}^{(2)}}{C_{\mathbf{0}}^{(2)}} = 2\xi_{\mathbf{g}}\gamma^{(2)} = w + \sqrt{w^2+1} \qquad (13.28)$$

(the superscripts on $C_{\mathbf{g}}^{(1)}$, etc., correspond to the superscripts on $\gamma^{(1)}$ and $\gamma^{(2)}$, i.e., the two solutions to the original quadratic equation). Now it is useful to make another substitution (or transformation) to simplify these relationships. We define β by

$$w = \cot\beta \qquad (13.29)$$

Now we can impose a restriction on the absolute magnitudes of $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$ so that they satisfy the relations

$$C_{\mathbf{0}}^{(1)2} + C_{\mathbf{g}}^{(1)2} = 1 = C_{\mathbf{0}}^{(2)2} + C_{\mathbf{g}}^{(2)2} \qquad (13.30)$$

By normalizing these values for C separately for each value of γ, we are restricting the intensity of the beam to values between 0 and 1 (see below). Then if we substitute equation 13.29 into equation 13.27 and then into equation 13.28 we find (using $1-\cos\beta = 2\sin^2(\beta/2)$ and $\sin\beta = 2\sin(\beta/2)\cos(\beta/2)$) that the C values have the following simple forms

$$\begin{aligned} C_{\mathbf{0}}^{(1)} &= \cos\tfrac{\beta}{2} \quad & C_{\mathbf{g}}^{(1)} &= -\sin\tfrac{\beta}{2} \\ C_{\mathbf{0}}^{(2)} &= \sin\tfrac{\beta}{2} \quad & C_{\mathbf{g}}^{(2)} &= \cos\tfrac{\beta}{2} \end{aligned} \qquad (13.31)$$

Now you can understand why we introduced β in equation 13.29. The two independent solutions to the reformulated Howie-Whelan equation for $\phi_{\mathbf{0}}$ (13.16) are then $\phi_{\mathbf{0}} = C_{\mathbf{0}}^{(1)}\exp(2\pi i\gamma^{(1)}z)$ and $\phi_{\mathbf{0}} = C_{\mathbf{0}}^{(2)}\exp(2\pi i\gamma^{(2)}z)$ and *each value* has a corresponding value for $\phi_{\mathbf{g}}$.

We can already see that the ratio of the amplitudes of the diffracted and direct beams, $C_{\mathbf{g}}$ to $C_{\mathbf{0}}$ (and therefore the intensities), in equation 13.21 depends on γ, the phase of the wave, and hence on s, the excitation error. Hence the ratios in equations 13.27 and 13.28 depend on how close the specimen is to the Bragg orientation. We are particularly concerned about *the* Bragg condition because we have chosen a two-beam situation.

THE SUBSTITUTION

Because of this simple substitution you can easily confirm that, for this two-beam situation, the probability of finding the electron in one beam or the other remains unity ($|\psi^{\mathrm{T}}|^2 = 1$). Hence the reason we use a normalized intensity in equation 13.30.

In the two-beam approximation, equation 13.5 is expressed in terms of $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$, both of which depend on γ (equation 13.17), so equation 13.5 can then be written in terms of both values of γ (and hence $C_{\mathbf{0}}^{(1)}$, $C_{\mathbf{0}}^{(2)}$, etc.), giving two independent quantities $b^{(1)}$ and $b^{(2)}$. Either of these two functions could be ψ^{T}, the total wave function. Alternatively, the total wave function could be some combination of them, i.e., part $b^{(1)}$ plus part $b^{(2)}$. Both of these wave functions are dependent on $\mathbf{r}$ and have their own values of $\mathbf{k}$ that we identify as $\mathbf{k}^{(j)}$.

Each value of γ gives a different value of $\mathbf{k}$ that we call $\mathbf{k}^{(j)}$.

Thus we can write expressions for $b^{(1)}$ and $b^{(2)}$.

$$b^{(1)}\left(\mathbf{k}^{(1)},\mathbf{r}\right) = C_{\mathbf{0}}^{(1)}e^{2\pi i\,\mathbf{k}^{(1)}.\mathbf{r}} + C_{\mathbf{g}}^{(1)}e^{2\pi i(\mathbf{k}^{(1)}+\mathbf{g}).\mathbf{r}} \qquad (13.32)$$

and

$$b^{(2)}\left(\mathbf{k}^{(2)},\mathbf{r}\right) = C_{\mathbf{0}}^{(2)}e^{2\pi i\,\mathbf{k}^{(2)}.\mathbf{r}} + C_{\mathbf{g}}^{(2)}e^{2\pi i(\mathbf{k}^{(2)}+\mathbf{g}).\mathbf{r}} \qquad (13.33)$$

Remember: each of these Bloch-wave functions could be a wave in the crystal—each one depends on only one of the $\mathbf{k}$ values. In general, the total wave function will be a combination of these two waves. We'll return to the important relationship between $\mathbf{k}$ and γ in Section 13.9. We use the letter 'b' here because we've actually obtained expressions for the Bloch waves mentioned in Section 13.2, which we'll discuss in the next chapter.

13.9 THE TOTAL WAVE AMPLITUDE

We have now found two different wave functions which can both propagate in the crystal. We still have to determine what $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$ are. The total wave vector, ψ^{T}, is a combination of the two (Bloch) waves, $b^{(1)}$ and $b^{(2)}$

$$\psi^{\mathrm{T}} = \mathcal{A}^{(1)}b^{(1)} + \mathcal{A}^{(2)}b^{(2)} \qquad (13.34)$$

where the constants $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ determine the relative contribution of each (Bloch) wave. We can now combine the last few equations (13.31–13.33 and 13.34) to give

$$\begin{aligned} \psi^{\mathrm{T}} = \mathcal{A}^{(1)}\{&\cos\tfrac{\beta}{2}e^{2\pi i\mathbf{k}^{(1)}.\mathbf{r}} - \sin\tfrac{\beta}{2}e^{2\pi i(\mathbf{k}^{(1)}+\mathbf{g}).\mathbf{r}}\} \\ +\mathcal{A}^{(2)}\{&\sin\tfrac{\beta}{2}e^{2\pi i\mathbf{k}^{(2)}.\mathbf{r}} - \cos\tfrac{\beta}{2}e^{2\pi i(\mathbf{k}^{(2)}+\mathbf{g}).\mathbf{r}}\} \end{aligned} \qquad (13.35)$$

All that now remains is to determine the magnitudes of $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ which we can do by remembering that we have a thin TEM specimen. In mathematical terminology the constants $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ must now be determined using the boundary conditions.

It is helpful to rearrange equation 13.35 first

$$\psi^{T}=\{\mathcal{A}^{(2)}\sin\tfrac{\beta}{2}e^{2\pi i\mathbf{k}^{(2)}.\mathbf{r}}-\mathcal{A}^{(1)}\cos\tfrac{\beta}{2}e^{2\pi i\mathbf{k}^{(1)}.\mathbf{r}}\}+\{\mathcal{A}^{(2)}\cos\tfrac{\beta}{2}e^{2\pi i\mathbf{k}^{(2)}.\mathbf{r}}-\mathcal{A}^{(1)}\sin\tfrac{\beta}{2}e^{2\pi i\mathbf{k}^{(1)}.\mathbf{r}}\}e^{2\pi i\mathbf{g}.\mathbf{r}} \quad (13.36)$$

Only the second term depends on **g**, so this must be the $\phi_{\mathbf{g}}$ term. We know that at the top of the specimen ($\mathbf{r} = 0$), $\phi_{\mathbf{0}}$ is unity and $\phi_{\mathbf{g}}$ is zero (independent of γ)—the amplitude of the diffracted beam is zero before it's diffracted! It follows directly that

$$\mathcal{A}^{(1)} = \cos\frac{\beta}{2} \quad (13.37)$$

$$\mathcal{A}^{(2)} = \sin\frac{\beta}{2} \quad (13.38)$$

These equations (13.37 and 13.38) tell us that $\mathcal{A}$ in equation 13.34 is just determined by the value of **s**, i.e., the deviation from the Bragg condition. So you can adjust the values of $\mathcal{A}$ by changing **s**, which we now know just involves tilting the specimen.

Now, finally, we can write down the general expressions for $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$, each as a function of z. First we need to modify equation 13.5 by using the substitution of equations 13.12 and 13.13, so it becomes

$$\psi^{T} = \phi_{\mathbf{0}}e^{2\pi i\,\mathbf{k}.\mathbf{r}} + \phi_{\mathbf{g}}e^{2\pi i(\mathbf{k}+\mathbf{g}).\mathbf{r}} \quad (13.39)$$

(Remember that $\chi_D = \chi_O + \mathbf{g} + \mathbf{s}$ (or $\mathbf{k}_D = \mathbf{k}_O + \mathbf{g} + \mathbf{s}$), where $\mathbf{k}_O$ is written as **k** and D is G_1 in equation 13.5; then you'll see that the term containing **s** in equation 13.13 drops out.) The $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$ components in equation 13.36 are easily recognized by the presence of $\exp(2\pi i\mathbf{g}\cdot\mathbf{r})$. Comparing equations 13.36 and 13.39 (having replaced $\mathcal{A}$ using equations 13.37 and 13.38) we see that

$$\phi_{\mathbf{g}} = \sin\frac{\beta}{2}\cos\frac{\beta}{2}\left\{e^{2\pi i(\mathbf{k}^{(2)}-\mathbf{K}).\mathbf{r}} - e^{2\pi i(\mathbf{k}^{(1)}-\mathbf{K}).\mathbf{r}}\right\} \quad (13.40)$$

Since we are only considering the z component, we know, from equations 13.17 and 13.19, that the exponential term must have the phase $2\pi i\gamma z$, i.e.,

$$(\mathbf{k}^{(2)} - \mathbf{K})_z = \gamma^{(2)} \text{ and } (\mathbf{k}^{(1)} - \mathbf{K})_z = \gamma^{(1)} \quad (13.41)$$

What we are interested in is the magnitude of $\gamma^{(1)}$ and $\gamma^{(2)}$.

We can now manipulate equation 13.40 using equation 13.41 and the expression $e^{i\theta} = \cos\theta + i\sin\theta$ to give

$$\phi_0 = \{\cos(\pi z\Delta\mathbf{k}) - i\cos\beta.\sin(\pi z\Delta\mathbf{k})\}e^{\pi i s z} \quad (13.42)$$

and

$$\phi_{\mathbf{g}} = +i\sin\beta.\sin(\pi z\Delta\mathbf{k}).e^{\pi i s z} \quad (13.43)$$

In these equations Δk is simply $|\mathbf{k}^{(2)} - \mathbf{k}^{(1)}|$. Leaving the term $e^{\pi i s z}$ in these equations does not affect the amplitudes of $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$ or the beam intensities but it will make it easier for you to check that these expressions satisfy, for example, equation 13.16.

A KEY RESULT

We have shown directly that $\phi_{\mathbf{0}}$ in equation 13.39 is a mixture of terms containing $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$. This is why $\phi_{\mathbf{g}}$ depends on Δk.

13.10 THE EFFECTIVE EXCITATION ERROR

We can now write down the intensity at the bottom (exit surface) of the specimen ($z = t$) and manipulate the equations by substituting for Δk and w. The term Δk in equations 13.42 and 13.43 is the same as $\Delta\gamma$, i.e., $\gamma^{(2)} - \gamma^{(1)}$ (see equation 13.41). We can therefore write down Δk by considering equations 13.27 and 13.28

$$\Delta k = \frac{\sqrt{w^2+1}}{\xi_{\mathbf{g}}} \quad (13.44)$$

The intensity in the diffracted beam, $|\phi_{\mathbf{g}}|^2 = \phi_{\mathbf{g}}\phi_{\mathbf{g}}^*$, is obtained from equation 13.43

$$I_{\mathbf{g}} = \left|\phi_{\mathbf{g}}\right|^2 = \sin^2\beta.\sin^2(\pi t\Delta k) \quad (13.45)$$

$$I_{\mathbf{g}} = \left|\phi_{\mathbf{g}}\right|^2 = \frac{1}{w^2+1}\sin^2\frac{\pi t\sqrt{w^2+1}}{\xi_{\mathbf{g}}} \quad (13.46)$$

We can make this equation look more familiar by defining an effective excitation error, s_{eff}

$$s_{\text{eff}} = \sqrt{s^2 + \frac{1}{\xi_{\mathbf{g}}^2}} = \frac{\sqrt{w^2+1}}{\xi_{\mathbf{g}}} \quad (13.47)$$

Now the equation becomes

$$\left|\phi_{\mathbf{g}}\right|^2 = \left(\frac{\pi t}{\xi_{\mathbf{g}}}\right)^2.\frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} \quad (13.48)$$

This equation gives us the intensity in the Bragg-diffracted beam. In writing down equation 13.47, we have defined another important new quantity s_{eff}, so labeled because it's the *effective* excitation error.

THE REALLY IMPORTANT EQUATION

It's so important, we'll repeat it

$$\left|\phi_{\mathbf{g}}\right|^2 = \left(\frac{\pi t}{\xi_{\mathbf{g}}}\right)^2.\frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2}$$

One important result shown directly by equation 13.45 is that the intensity, I_g, in the diffracted beam emerging from the specimen is proportional to $\sin^2(\pi t \Delta k)$ and thus I_0 is proportional to $\cos^2(\pi t \Delta k)$. I_g and I_0 are both periodic in both t and s_{eff}. As ϕ_g increases and decreases, ϕ_0 behaves in a complementary manner so that

$$I_0 = 1 - I_g \tag{13.49}$$

Remember when testing this formula that $I = \phi\phi^*$ (ϕ^* is the complex conjugate of ϕ).

The effective excitation error, s_{eff}, is a very important quantity. We can summarize some important properties

- The quantity s_{eff} is never zero.
- When s is zero, s_{eff} is ξ_g^{-1}.
- When s is very large, then s_{eff} becomes essentially the same as s.

13.11 THE COLUMN APPROXIMATION

When we form an image, we try to focus the objective lens on a plane in or below the specimen (remember that here, below means underfocus). One special plane we can choose is the plane which corresponds to the bottom of the specimen, assuming that this plane is perpendicular to the direction of the propagating beam. Whatever plane we choose, what we see depends on the beams that finally leave the bottom of the specimen, so let's concentrate on this one plane. Look at Figure 13.3A; P is the point at the bottom of the specimen, and we are calculating the values of ϕ_0 and ϕ_g at this point to construct our image. Where do the electrons come from in order to contribute to ϕ_0 and ϕ_g? The answer is the cone APB where the angle APB is $\sim 2\theta_B$. In other words, we don't just have a diffracted beam which propagates through the specimen from the top to point P. There is actually a cone of material which contributes to the intensity at point P. The shape of the cone can be calculated using the Fresnel-zone construction, which was actually developed nearly 200 years ago for visible-light optics. Figure 13.3B, which is how the cone is usually drawn, summarizes the relevant parameters; don't forget that a cone, not a triangle, of material contributes to the intensity at P. A clear derivation is given by Hecht. Why is it a Fresnel diffraction? The answer is that we form an image, i.e., look at a plane, which is very close to where the diffraction 'event' occurred, we are in the near-field, or Fresnel, regime (see Section 2.9).

FOCUSING

As usual, we should focus only by moving the specimen up and down, but we don't.

Let's consider some actual numbers: At 100 kV, $\lambda = 3.7$ pm, $\theta_B \sim 0.01$ radians or $\sim 0.5^\circ$. So if the thickness of the specimen is 100 nm, then AB is $\sim$ 2 nm. If we increase the t, then the width of the column will also increase. However, if we increase the accelerating voltage so as to increase the thickness we can penetrate, the wavelength decreases causing the Bragg angle also to decrease. This allows us to make the approximation shown in Figure 13.4A and B when calculating ϕ_0 and ϕ_g. This model is known as the column approximation.

The great advantage of this approximation is that it allows us to calculate the scattering from slices which have a constant width as we pass down the column, which itself lies in a well-defined direction (generally parallel to $\mathbf{k}_D$). We might anticipate problems with very small defects on a very

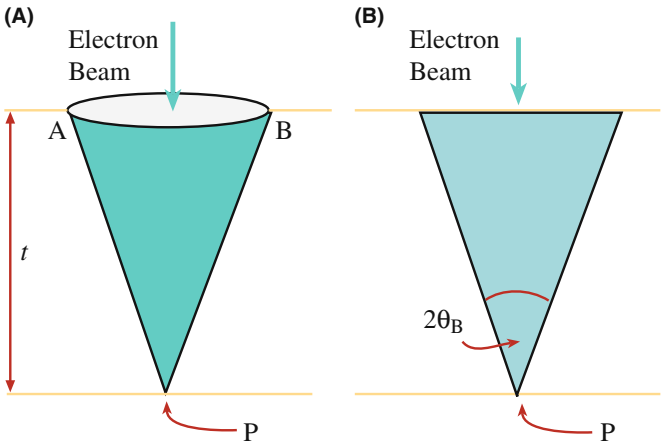


FIGURE 13.3. (A) The intensity of the beams at point P at the bottom of the specimen is influenced by all the scattering within a cone of material. The solid angle of the cone is determined by the diameter of the Fresnel zones which, in turn, are principally determined by λ. The cross section (B) is the more typical view of the cone.

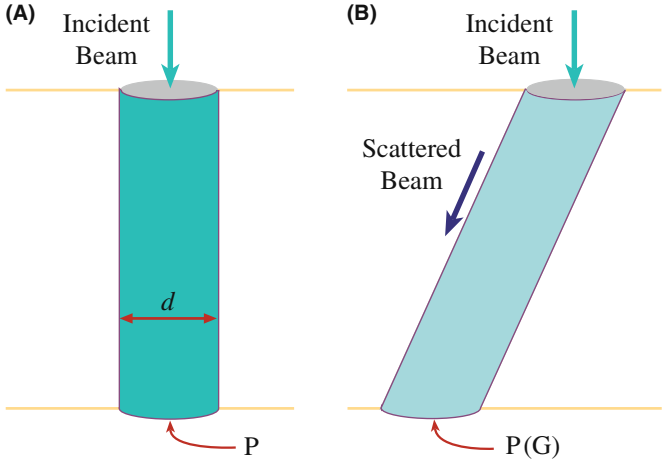


FIGURE 13.4. The column approximation for (A) the direct beam and (B) a diffracted beam. A column replaces the cone. The diameter of the column (d) should be the *average* diameter of the cone it replaces (AB/2 in Figure 13.3). (B) This value will depend on the thickness of the sample. In practice it is usually taken to be ~2 nm.

fine detail, especially when these features can vary their positions in the foil. The more correct, non-column treatment was introduced by Takagi; the analysis by Howie and Basinski is what we use in computer programs.

COLUMN APPROXIMATION

The column approximation often hides itself very well, but it is actually used in many calculations of images.

13.12 THE APPROXIMATIONS AND SIMPLIFICATIONS

In order to minimize the mathematics and to emphasize the underlying physical principles involved in the analysis of diffracted beams, we have made a number of assumptions, simplifications, and approximations. Although we are not going to cover all of these points, you should be aware of some of them

- We have completely neglected any effects due to backscattering of the electrons. This approximation is reasonable since we are dealing with electrons which have very high energies. However, if you are familiar with SEM, you will have encountered backscattered electron (BSE) imaging and possibly, rocking-beam channeling patterns (RCPs) or electron backscattered diffraction (EBSD) patterns. So some electrons must be backscattered.
- In some parts of the discussion, it is an implicit assumption that the crystal has a center of symmetry. This assumption is hidden in our use of $\xi_{\mathbf{g}}$. If the material is non-centrosymmetric, then the BF image and images formed using only a systematic row of reflections will not be affected. Differences will occur in some DF images or when non-systematic reflections contribute to the image. In these cases, you will need to use a computer program to predict or interpret the contrast.

TWO OPPOSITE CONVENTIONS

Two conventions are used to describe the exponential dependence on **k** and **r**; both are commonly used.

$$e^{2\pi i\mathbf{k}.\mathbf{r}} \quad \text{or} \quad e^{-2\pi i\mathbf{k}.\mathbf{r}}$$

These conventions have been discussed by Spence. In our analysis, we have chosen to use $e^{2\pi i\mathbf{k}.\mathbf{r}}$ which Spence has termed the 'quantum-mechanical' convention.
Note that Spence uses the alternative 'crystallographic' convention, except when he discusses Bloch waves.

- From Chapter 11, you know that it is impossible to set up a true two-beam condition for a thin TEM specimen. There will always be more than one diffracted spot visible. So how do we measure $\xi_{\mathbf{g}}$ exactly? The answer is that we don't, but we can make a very good estimate.
- Remember the use of z and t. When we consider the diffracted beam, then z and t are measured along the direction of the diffracted beam. In general, this distance will be different for each beam. The saving feature is that we are usually concerned with small Bragg angles. As a thought exercise, you might consider the effect of having a steeply inclined wedge or a specimen which, although parallel sided, is steeply inclined to the electron beam.
- The full analysis of scattering includes a term in r^{-1} which says that the intensity falls off as r^{-2}. This is just the standard flux relation—the number of electrons passing through a spherical surface around the scattering point is constant. (The surface area of a sphere is proportional to r^2.) This term has been omitted throughout our discussion since it only affects the absolute intensity. A practical lesson from this is that you should use the lowest magnification that will give you the desired resolution; remember that the highest useful magnification in a TEM image is about 10^6 (see Section 6.6.B).

In the quantum mechanical convention, the time-dependent Schrödinger equation is written as

$$\frac{\mathrm{h}^2}{8\pi^2 m}\nabla^2\psi = -\mathrm{i}\frac{\mathrm{h}}{2\pi}\frac{d\psi}{dt} \tag{13.51}$$

with the full solution being

$$\psi(\mathbf{r}, t) = Ae^{+\mathrm{i}(\mathbf{k}.\mathbf{r}-\omega t)} \tag{13.52}$$

- The concept of a refractive-index effect for electron waves is directly analogous to that for visible-light waves, or any other electromagnetic radiation, in that the potential of the crystal causes a change in the kinetic energy of the electrons (because their total energy is unchanged) and therefore, their velocity is changed. Normally, of course, we think of this as a change in the wavelength of the electrons. The magnitude of **k** will always be larger than that of χ.
- We have not mentioned the absorption of Bragg beams, yet we know that this must occur since we can only examine thin specimens in the TEM. Absorption of beams is considered in Section 14.6 and Section 23.8.

13.13 THE COUPLED HARMONIC OSCILLATOR ANALOG

The expression for the intensity of the diffracted beam is particularly simple when $s = 0$. Then from equation 13.48 we can write

$$|\phi_{\mathbf{g}}|^2 = \sin^2\left(\frac{\pi t}{\xi_{\mathbf{g}}}\right) \tag{13.53}$$

and similarly

$$|\phi_{\mathbf{0}}|^2 = 1 - \sin^2\left(\frac{\pi t}{\xi_{\mathbf{g}}}\right) \tag{13.54}$$

Both equations now only have one variable, the thickness of the specimen. We will refer to these equations when we discuss images in Chapter 23, but we can note immediately that $I_{\mathbf{g}}$ is zero at $t = 0$ and again at $t = \xi_{\mathbf{g}}$ (or in general at $t = n\xi_{\mathbf{g}}$, where n is an integer). This is the reason we call $\xi_{\mathbf{g}}$ the extinction distance. This situation corresponds to two coupled simple-harmonic oscillators with energy (i.e., intensity, $I_{\mathbf{0}}$ and $I_{\mathbf{g}}$) being continuously transferred from one to the other and back again. Notice that $I_{\mathbf{g}}$ can only increase to unity when $\mathbf{s} = 0$.

CHAPTER SUMMARY

In this chapter, we have derived equations and introduced terminology that will form the basis for our discussion of diffraction-contrast images. It is not necessary for you to be able to reproduce the mathematical derivations but equations 13.47 and 13.48 are crucial and must be understood. Our analysis was quickly limited to two beams, the direct beam and one Bragg-diffracted beam. In deriving the Howie-Whelan equations, we must consider both forward scattering and Bragg diffraction. We introduced a new parameter, the critical length $\xi_{\mathbf{g}}$, and explained why this parameter is called the extinction distance. This length was defined in equation 13.4 which shows that $\xi_{\mathbf{g}}$ depends on the *material*, the *reflection,* and the *wavelength of the electrons*. Two particular points you need to remember are

- If the voltage increases then λ decreases and $\xi_{\mathbf{g}}$ increases.
- The contribution of each Bloch wave is determined by **s**.

In Section 24.3, we'll show how the two-beam analysis can be extended using the concept of the scattering matrix.

REFERENCES

This treatment of diffracted beams follows that given by Hirsch and Whelan and the textbook by Hirsch et al. that built on Darwin's treatment of X-ray diffraction.

HISTORY

Darwin, CG 1914 *Röntgen-Ray Reflection I;II* Phil, Mag. **27** 315–333 and 675-690. Charles Galton Darwin was a grandson of Charles Robert Darwin and, like (*the*) Darwin (and Hirsch), became a Fellow of Christ's College.

Hecht, E 1987 *Optics*, 4th ed., Addison-Wesley, Reading MA.

Howie, A and Whelan, MJ 1961 *Diffraction Contrast of Electron Microscope Images of Crystal Lattice Defects. II The Development of a Dynamical Theory* Proc. Roy. Soc. **A263** 217–237.

THE COLUMN APPROXIMATION

Howie, A and Basinski, ZS 1968 *Approximations of the Dynamical Theory of Diffraction Contrast* Phil. Mag. **17** 1039–1063.

Takagi, S. 1962 *Dynamical Theory of Diffraction Applicable to Crystals with Any Kind of Small Distortion* Acta Cryst. **15** 1311–1312.

NOTATION

Spence, JCH 2003 *Experimental High-Resolution Electron Microscopy* 3rd Ed. Oxford University Press New York. Uses the quantum-mechanical convention rather than the crystallographic one.

THE COMPANION TEXT

The H-W equations are the basis of many home-written programs including those available in the text by Head et al. If you are taking the time to simulate diffraction-contrast images, you might as well use a package which will allow you at least to assess the influence of beams in the systematic row and then see how such multi-beam images compare to those predicted by this analytical treatment. A complete chapter is devoted to this topic in the companion text.

SELF-ASSESSMENT QUESTIONS

Q13.1 What does $\xi_{\mathbf{g}}$, the characteristic length, depend on?
Q13.2 How is the structure factor related to the characteristic length, $\xi_{\mathbf{g}}$?
Q13.3 What do we mean by the phrase *two-beam condition* and when is it satisfied?
Q13.4 Can forward scattering change the direction of the beam?
Q13.5 What does dynamical diffraction mean in the two-beam case?
Q13.6 What is the difference relationship between the constant $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ and $\mathbf{s}$?
Q13.7 How can we change the value of the constants $\mathcal{A}$?
Q13.8 Explain in words how changing the accelerating voltage affects the characteristic length, $\xi_{\mathbf{g}}$?
Q13.9 Why do we use the column approximation?
Q13.10 What will happen to the width of the column if the thickness of the specimen increases. How can we change the width of the column?
Q13.11 How does dynamical diffraction affect the interpretation of DPs?
Q13.12 What is the structure factor and what does it depend on?
Q13.13 Why is $\xi_{\mathbf{g}}$, the characteristic length, better known as the extinction distance?
Q13.14 What is the total wave function and why should we really be talking about Bloch waves?
Q13.15 What is the relationship between the intensity of the incident beam and the intensity of the diffracted beam? Give the equations and the name they go by.
Q13.16 Write down an expression for the effective excitation error.
Q13.17 Give an equation for the intensity in the Bragg-diffracted beam using the effective excitation error s_{eff} for a specimen thickness t.
Q13.18 What happens to the effective excitation error when $s = 0$? Can the effective excitation error ever equal zero? (Explain your reasoning.)
Q13.19 What value does the effective excitation error approach as the excitation error, s, becomes large?
Q13.20 Define what $\phi_{\mathbf{g}}$ and $\phi_{\mathbf{0}}$ mean in the context of this chapter and how they relate to the beam intensities $I_{\mathbf{g}}$ and $I_{\mathbf{0}}$?
Q13.21 In principle, could the intensities predicted by the Howie-Whelan equations affect what you see in DPs? Is this likely to occur in practice?

TEXT-SPECIFIC QUESTIONS

T13.1 Examine Table 13.2. (a) Discuss why MgO, 111 looks odd. (b) Explain why ξ_{220} for diamond is less than ξ_{220} for Si and why this appears unexpected at first.
T13.2 Taking reasonable values for the different parameters: V_c, θ_B, λ, and F, deduce ξ_{111} for Cu and W using equation 13.4.
T13.3 Starting with equations 13.6 and 13.7, deduce equations 13.8 and 13.9 showing where we make the approximations.
T13.4 We state that we know we should use Bloch waves. Why is this so? Would this be true if the specimen were amorphous? If not, what would you recommend (assuming that you have to work with amorphous materials)?
T13.5 Derive equation 13.18.
T13.6 We show that there are two solutions to the Howie-Whelan equations. Explain in words why this is so. What is the physical significance of this result?
T13.7 Show that equations 13.42 and 13.43 do satisfy equation 13.16.
T13.8 Consider the Cu_{220} reflection in a 200-kV TEM. Increase s in six equal increments to a value of 2×10^{-3} Å^{-1} (six values excluding 0 but including 2×10^{-3} Å^{-1}). Construct a table to show how s_{eff} varies. Give your results in Å^{-1} and nm^{-1}.

T13.9 In Figure 13.4 d is said to be ~2 nm. Consider five samples of Si with thickness 20, 50, 100, 200 nm, and 1 μm in a 200-eV TEM with 220 excited. What value should we use for d in each case?

T13.10 Repeat question 13.9 for Cu and 111.

T13.11 We state that the highest useful magnification in a TEM is about 10^6. How is this changed if we add a CCD camera? What is the lowest useful magnification with and without a CCD camera (excluding the DP)? How is this changed if you use a FEGTEM?

T13.12 For the perfect crystal situation, what is the ratio of $|\phi_{\mathbf{g}}|^2$ to $|\phi_{\mathbf{0}}|^2$ when s is 10^{-1} Å^{-1} and 10^{-2} Å^{-1} for the 220 reflection in Cu and Si at 100 kV. Explain any assumptions you make.

T13.13 Do you think Darwin's papers are really relevant to the Howie-Whelan equations?

T13.14 Using reasonable values for f, deduce values for $\xi_{\mathbf{g}}$ for

a) Cu with 220 and 100 keV
b) Si with 111 and 200 keV
c) W with 100 and 300 keV.

What assumptions have been made in applying this formula?

T13.15 Are the values given in Table 13.2 for Al, Fe, and Ge consistent?

T13.16 We differentiate between χ and $\mathbf{k}$. Do you expect there is a significant difference? Discuss the magnitude of this difference, how it varies with kV and how you might assess this magnitude.

T13.17 Using earlier equations, deduce equations 13.8 and 13.9.

T13.18 How will dynamical scattering of X-rays compare to that for electrons?

T13.19 After equation 13.16 we state that the nature of the material only enters through $\xi_{\mathbf{g}}$. Argue against this statement.

T13.20 Consider Figure 12.6A for Cu G = 220 and 100-keV electrons. Deduce w in equation 13.26. Hence, determine β in equation 13.2.

T13.21 Use the results of the previous question to deduce $C_0^{(1)}$, etc., in equation 13.31.

14

Bloch Waves

CHAPTER PREVIEW

This topic is rather mathematical, with long sequences of differential equations. The discussion of Bloch waves given here follows the treatment of Hirsch et al. which, in turn, was based on the original analysis of electron diffraction by Bethe (1928). The notation we will use closely follows that used by Bethe. Remember that **g** can be any reciprocal-lattice vector, although we will also use it to represent a specific vector.

This analysis leads directly to one of the most important concepts used to understand images of defects in thin foils: it explains the physical origin of the *extinction distance,* $\xi_{\mathbf{g}}$, and thus shows why it is so important. So again it is worth persevering. However, many successful microscopists have skipped this topic. We suggest you first skim through this chapter. Then, when you've recognized its importance and seen the key equations, go back to the beginning and work your way through.

We make certain assumptions about the materials we are considering and what voltages are used. You must keep these assumptions in mind when applying these concepts. The most important point is that, within the limits of our approximations, the analysis is rigorous and we can really understand the meaning of $\xi_{\mathbf{g}}$. If you've previously come across the idea of kinematical diffraction, this chapter will make it clear why this theory is, at best, only an approximation to reality.

We start by considering the property of a crystal which we know quite well, namely, the inner potential. You should remember that, strictly speaking, everything we are about to go through in this section applies only to perfect crystals; crystals with surfaces are not 'perfect.' The periodic nature of the crystal potential leads to the concepts of Bloch functions and Bloch waves.

We include a discussion of the two-beam case, since this can easily be solved analytically and can be related directly to the results discussed in Chapter 13 on diffracted beams. In Chapter 15 we will discuss a graphical representation of the equations we are deriving here. As with the Ewald sphere and reciprocal lattice, the diagrams make for an easier understanding and give a useful guide when you are actually using the TEM. We will consider absorption of Bloch waves here but when we use it in, e.g., Section 24.7, the physical significance will be more obvious.

14.1 WAVE EQUATION IN TEM

We are going to modify the Schrödinger equation for use in TEM to explain why the Bloch waves have the form they do. We are not going to try to be mathematically rigorous in deriving this modified equation; reference texts for this purpose are listed at the end of the chapter. Let's start with the time-independent Schrödinger equation

$$\left[-\frac{\mathrm{h}^2}{8\pi^2 m}\nabla^2 + V(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}) \qquad (14.1)$$

The first term (in ∇^2) represents the kinetic energy and the second term, the potential energy; E represents the total energy. In TEM we usually talk in terms of the accelerating voltage and the crystal potential and, therefore, we rearrange this equation in terms of voltages. In doing so, we have to be careful about signs since the charge on the electron is negative and the applied electric field (associated with the accelerating voltage) points toward the gun! All that equation 14.1 says is that the electron has a kinetic energy due to the acceleration it is given in the gun. Initially, this is the total energy of the electron. When the electron passes

through the crystal, it will have a potential energy due to the periodic potential associated with the atoms in the crystal.

THE SIGNS

- The charge on the electron, q, is a negative number, –e, where e is a positive number.
- The accelerating voltage, $-V$ (usually between 100 kV and 1 MV), is negative for a positive charge leaving the gun. This quantity, $-V$, is really the 'electric-field potential.'
- The initial energy given to the electron is a positive number, E (in eV); it is just the charge times the accelerating voltage. We can write this as eV where both e and V are positive numbers.
- The potential inside the crystal, $V(\mathbf{r})$, is a positive number reaching a local maximum at the nucleus of an atom; the nucleus is positive.
- The potential energy, $\mathcal{V}$, of the electron outside the crystal is zero; it decreases when the electron is inside the crystal ($\mathcal{V}$ is q times $V(\mathbf{r})$, i.e., $-eV(\mathbf{r})$) and is therefore always a negative number.

Now we can rearrange equation 14.1 in terms of the accelerating voltage and the crystal potential

$$\nabla^2\psi(\mathbf{r}) + \frac{8\pi^2 me}{h^2}[V + V(\mathbf{r})]\psi(\mathbf{r}) = 0 \qquad (14.2)$$

The task before us is obvious: we have to solve equation 14.2. In general, however, this is a difficult problem! What makes it possible for us is that $V(\mathbf{r})$ has special properties because we are only considering crystalline materials.

14.2 THE CRYSTAL

The basic property of a crystal is that its inner potential, $V(\mathbf{r})$, is *periodic*. We can therefore express this property as

$$V(\mathbf{r}) = V(\mathbf{r}+\mathbf{R}) \qquad (14.3)$$

where $\mathbf{R}$ represents any lattice vector of the crystal and, as usual, $\mathbf{r}$ represents any real-space vector. Equation 14.3 is the fundamental definition of a perfect crystal: the environment at point $\mathbf{r}$ is identical to that at point $\mathbf{r} + \mathbf{R}$. We can draw this inner potential as shown in Figure 14.1 for the one-dimensional case and you can imagine it in 3D. The atomic nuclei are positively charged; the surrounding electrons gradually screen this charge and the atom appears neutral from the outside. In a crystal, a nucleus is never far away, so an electron which we 'shoot' through the crystal will always

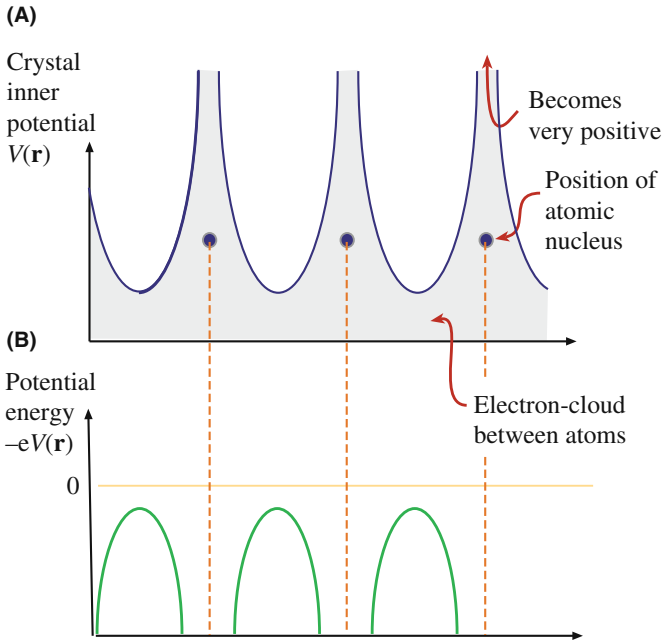


FIGURE 14.1. (A) The local charge sensed by the beam electron as it passes through a metal, represented as a row of 'ion' cores (black circles) in a sea of electrons. The local charge is very large and positive in the vicinity of the ion and becomes small, but not zero between the ions. The difference between the minimum charge and zero corresponds to the mean inner potential of the crystal, which is a few eV (positive). So the beam electron experiences a small positive attraction as it enters the crystal, hence its kinetic energy (velocity) increases. (B) $V(\mathbf{r})$ is the potential of the electrons, so their potential *energy* is negative and becomes more so, the closer they pass by the ions.

see a positive potential; hence $V(\mathbf{r})$ is always positive as noted in Section 14.1 and in Figure 3.1.

The electron beam can be described by its total wave function ψ^{tot} which must always be a solution of the Schrödinger equation; i.e., this equation describes how an electron behaves both inside and outside the crystal.

In the discussion which follows, we will use potentials so the units will be volts. You can always change to the energy formalism, but remember that the charge on the electron is a negative number.

We know that for any crystal the inner potential must be real, i.e., the potential energy must be real, so that $V(\mathbf{r})$ and its complex conjugate, which we denote as $V^*(\mathbf{r})$, are identical

$$V(\mathbf{r}) = V^*(\mathbf{r}) \qquad (14.4a)$$

Now to make the treatment simple, we consider the case of crystals with a *center of symmetry*

$$V(\mathbf{r}) = V(-\mathbf{r}) \qquad (14.4b)$$

The case of non-centrosymmetric crystals, such as GaAs, could be considered, but the equations would

become much more complicated. Since $V(\mathbf{r})$ is periodic, we can express it as a Fourier series in which we sum over all the lattice points in reciprocal space

$$V(\mathbf{r}) = \sum_{\mathbf{g}} V_{\mathbf{g}} e^{2\pi i \mathbf{g}\cdot\mathbf{r}} \tag{14.5}$$

Here $V_{\mathbf{g}}$ is, by definition, the $\mathbf{g}$ component of V in the Fourier series. Now, in order to make future equations simpler we define a parameter $U_{\mathbf{g}}$ related to $V_{\mathbf{g}}$ by

$$V_{\mathbf{g}} = \frac{h^2}{2me} U_{\mathbf{g}} \tag{14.6}$$

In the Fourier series given in equation 14.5 and modified by equation 14.6, $V_{\mathbf{g}}$ and $U_{\mathbf{g}}$ are referred to as the Fourier coefficients. Equation 14.5 becomes

$$V(\mathbf{r}) = \frac{h^2}{2me} \sum_{\mathbf{g}} U_{\mathbf{g}} e^{2\pi i \mathbf{g}\cdot\mathbf{r}} \tag{14.7}$$

Now $V(\mathbf{r})$ has been expanded as a Fourier sum; all the conditions on $V_{\mathbf{g}}$ also apply to each $U_{\mathbf{g}}$ so that

$$U_{\mathbf{g}} = U_{\mathbf{g}}^* = U_{-\mathbf{g}} \tag{14.8}$$

You can check these relationships by just replacing $\mathbf{r}$ by $-\mathbf{r}$, etc. Before continuing, however, you may find it useful to review the relative magnitudes of the energies which are summarized in Table 14.1.

Much of what we are now discussing is mathematically the same as you may have seen in condensed-matter physics. The big difference is that we are injecting electrons with kinetic energies which are 5–6 orders of magnitude greater than the band gap of Si. Notice that the mean inner potential energy is $\mathcal{V}$ in equation 14.1. The actual value of $\mathcal{V}$ is not as precise as it might sometimes appear. You should remember that it is the average background potential energy and is directly related to the characteristic length ξ_0 that we introduced in Chapter 13. More values of $\mathcal{V}$ are given in Table 14.2. One interesting feature of this table is that the magnitude of $\mathcal{V}$ only varies by a factor of 3 when the atomic number changes from 4 to 74.

TABLE 14.1 A Comparison of the Orders of Magnitudes of the Energies Being Discussed in This Chapter

Quantity	Energy (eV)
kT (room temp.: $T = 293$ K)	0.025
Band gap of Si	1.1
Mean inner potential energy for Si	~11
Energy of electrons in TEM	$\geq$ 100,000

TABLE 14.2 Comparison of Mean Inner Potential Energies for Different Elements

Element	Inner Potential Energy (eV)
Be	7.8 ± 0.4
C	7.8 ± 0.6
Al	12.4 ± 1
Cu	23.5 ± 0.6
Ag	20.7 ± 2
Au	21.1 ± 2
Si	11.5
Ge	15.6 ± 0.8
W	23.4
ZnS	10.2 ± 1

14.3 BLOCH FUNCTIONS

Since the electron is in a periodic potential its wave function must have the symmetry of the crystal. The solutions to the Schrödinger equation which always have the required translation property are known as Bloch waves. Since these wave functions, $\psi^{(j)}(\mathbf{r})$, are special, we'll define them as

$$\psi^{(j)}(\mathbf{r}) = b\left(\mathbf{k}^{(j)}, \mathbf{r}\right) = b^{(j)}(\mathbf{r}) \tag{14.9}$$

The reason for the 'j' is that each Bloch wave has a single value of $\mathbf{k}$ (each Bloch wave is a plane wave) which we can denote as $\mathbf{k}^{(j)}$; in general, there will be more than one Bloch wave for a particular physical situation. The notation we will use is such that, whenever we have $\mathbf{k}^{(j)}$ in an expression, we will identify this by the superscript which implies that the function varies with $\mathbf{k}^{(j)}$. Bloch's theorem states that this wave function in a periodic potential can be written as

$$b^{(j)}(\mathbf{r}) = b\left(\mathbf{k}^{(j)}, \mathbf{r}\right) = \mu\left(\mathbf{k}^{(j)}, \mathbf{r}\right) e^{2\pi i \mathbf{k}^{(j)}\cdot\mathbf{r}} = \mu^{(j)}(\mathbf{r}) e^{2\pi i \mathbf{k}^{(j)}\cdot\mathbf{r}} \tag{14.10}$$

such that the *Bloch function*, $\mu^{(j)}(\mathbf{r})$, can itself be expressed as a Fourier series since $\mu(\mathbf{r})$ is also a periodic function of $\mathbf{r}$

$$\mu^{(j)}(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)}\left(\mathbf{k}^{(j)}\right) e^{2\pi i \mathbf{g}\cdot\mathbf{r}} \tag{14.11}$$

We'll call $C_{\mathbf{g}}^{(j)}$ the j-sub-g plane-wave amplitude and generally refer to the C values as the plane-wave amplitudes; they depend on which $\mathbf{k}^{(j)}$ we are considering, but not on $\mathbf{r}$. We combine these definitions to give b which is itself dependent on j

$$b^{(j)}(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i (\mathbf{k}^{(j)} + \mathbf{g})\cdot\mathbf{r}} \tag{14.12}$$

Using our notation, the superscript on C indicates that $C^{(j)}$ depends on j because it is associated with $\mathbf{k}^{(j)}$. We can now write the expanded expression for $b^{(j)}(\mathbf{r})$, which is a solution to the Schrödinger equation

$$b^{(j)}(\mathbf{r}) = C_{\mathbf{0}}^{(j)} e^{2\pi i \mathbf{k}^{(j)}\cdot\mathbf{r}} + C_{\mathbf{g}}^{(j)} e^{2\pi i(\mathbf{k}^{(j)}+\mathbf{g})\cdot\mathbf{r}} + \cdots \quad (14.13)$$

The first term in this series is $C_{\mathbf{0}}$; the subscript is zero is because the length of this $\mathbf{g}$ vector is 0. Much of the following analysis is exactly the same as you may have encountered in studying semiconductor band-gap theory. The difference will be that we can make certain approximations which are only valid because the electrons used in TEM have much higher energies (100 keV to 1 MeV) than the inner potential of the crystal (~7–24 eV). It is always important to keep in mind the magnitude of the quantities we are considering and remember that the Bloch function has the periodicity of the lattice. When you are reading other texts, you'll see that physics textbooks will tend to omit the term 2π in such expressions so that $|\mathbf{k}|$ becomes $2\pi/\lambda$ instead of $1/\lambda$.

EACH BLOCH WAVE

The main point to remember is that each Bloch wave is associated with just one $\mathbf{k}^{(j)}$ but it is a continuously varying function of $\mathbf{r}$. Each Bloch wave is a sum over all the points in reciprocal space. In other words, each Bloch wave depends on every $\mathbf{g}$, and conversely, each $\mathbf{g}$ beam depends on every Bloch wave!

We haven't done anything yet, just restated the problem and remembered Bloch's theorem. The analysis we've just completed follows the original treatment of Bethe (1928). We can now express ψ^{tot} using equation 14.9 to give

$$\psi^{\text{tot}} = \sum_{j=1}^{n} \mathcal{A}^{(j)}\psi^{(j)} = \sum_{j=1}^{n} \mathcal{A}^{(j)} b\left(\mathbf{k}^{(j)}, \mathbf{r}\right) \quad (14.14)$$

In this equation, $\mathcal{A}^{(j)}$ will be determined by the specimen type, the specimen orientation, etc., i.e., the boundary conditions. The $\mathcal{A}$s are known as the Bloch-wave excitation coefficients, since they tell us the relative contributions of each Bloch wave, i.e., how strongly each Bloch wave is excited.

14.4 SCHRÖDINGER'S EQUATION FOR BLOCH WAVES

What we are now going to do is to rewrite the Schrödinger equation to incorporate the properties of Bloch waves automatically. If you wish, you can skip this section and just accept the result given in equation 14.27. The way we include the periodicity is to express the inner potential in equation 14.2 as the Fourier series given in equation 14.7

$$\nabla^2\psi(\mathbf{r}) + \frac{8\pi^2 me}{h^2}\left(E + \frac{h^2}{2me}\sum_{\mathbf{g}} U_{\mathbf{g}} e^{2\pi i \mathbf{g}\cdot\mathbf{r}}\right)\psi(\mathbf{r}) = 0 \quad (14.15)$$

Now we simplify the algebra to give

$$\nabla^2\psi(\mathbf{r}) + 4\pi^2\left(\frac{2me}{h^2}E + \sum_{\mathbf{g}} U_{\mathbf{g}} e^{2\pi i \mathbf{g}\cdot\mathbf{r}}\right)\psi(\mathbf{r}) = 0 \quad (14.16)$$

and hence

$$\frac{1}{4\pi^2}\nabla^2\psi(\mathbf{r}) + \left(\frac{2me}{h^2}E + \sum_{\mathbf{g}} U_{\mathbf{g}} e^{2\pi i \mathbf{g}\cdot\mathbf{r}}\right)\psi(\mathbf{r}) = 0 \quad (14.17)$$

Next, we can introduce a new quantity K which is defined by the equation

$$K^2 = \frac{2meE}{h^2} + U_{\mathbf{0}} = \chi^2 + U_{\mathbf{0}} \quad (14.18)$$

With this definition we have removed the $U_{\mathbf{0}}$ term from the sum over all $\mathbf{g}$, so that equation 14.15 is now

$$\frac{1}{4\pi^2}\nabla^2\psi(\mathbf{r}) + K^2\psi(\mathbf{r}) + \sum_{\mathbf{g}\neq 0} U_{\mathbf{g}} e^{2\pi i \mathbf{g}\cdot\mathbf{r}}\psi(\mathbf{r}) = 0 \quad (14.19)$$

The reason for doing this is that we are going to be concerned with different diffraction vectors, $\mathbf{g}$. The $U_{\mathbf{0}}$ term does not depend on $\mathbf{g}$. We call $U_{\mathbf{0}}$ the (scaled) mean inner potential of the crystal; this potential is thus a 'background' or continuum property of the crystal; it does not directly depend on the crystal structure. (You may recognize this manipulation as the refractive-index idea reappearing.)

When $V(\mathbf{r})$ is 0, then $U_{\mathbf{0}}$ is 0 so that K^2 takes on a special value which we have already called χ^2

$$\chi^2 = \frac{2meE}{h^2} \quad (14.20)$$

The mass, m, is actually the relativistic mass, not the rest mass; eE is the kinetic energy of the electron (in the vacuum between the gun and the specimen). We know that

$$\frac{1}{2}mv^2 = \frac{(mv)^2}{2m} = \frac{p^2}{2m} = \frac{(hk)^2}{2m} \quad (14.21)$$

where v is the velocity, $\mathbf{p}$ the momentum, and $\mathbf{k}$ a wave vector. Thus χ is the wave vector of the electron outside the crystal as we had in Chapter 13.

The meaning of K is now clear: $\mathbf{K}$ is the wave vector of the electron *inside* the specimen, i.e., after correcting for the refractive-index effect. Since $U_{\mathbf{0}}$ is a positive number, K is always larger than χ. Hence the kinetic energy of the electrons in the crystal is greater than in the vacuum. The potential energy inside the crystal is negative so, even

though it may be counterintuitive, you now know that electrons travel faster in the crystal! The wavelength of the electrons in the crystal is therefore smaller than the wavelength outside (λ is the reciprocal of k).

ELECTRON SPEED IN A CRYSTAL
Electrons travel faster in the crystal. Light slows down in a crystal.

Remember that light is electromagnetic radiation. The refractive index for light is $n = c/v$ and is always ≥ 1; c is the velocity of light in a vacuum and v is the velocity in any other material. This is one of those cases where we have to be wary when applying ideas derived for light waves to electron waves.

DISPERSION
When discussing light, the word dispersion means *separation of electromagnetic radiation into constituents of different wavelength*. In electron optics, the meaning is exactly the same but we emphasize different **k** vectors or different energy.

Equations 14.18 and 14.20 are *dispersion relations*. Such equations relate the magnitude of the wave vector, K or χ, to the energy of the electron.

We want to simplify equation 14.19. We know that $\psi(\mathbf{r})$ is a Bloch wave (given by equation 14.12) so we can obtain an expression for $\nabla^2\psi(\mathbf{r})$ by differentiating $b^{(j)}(\mathbf{r})$. Remember that $C_{\mathbf{g}}^{(j)}$ does not depend on **r**

$$\nabla^2\psi(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} \nabla^2 \left(e^{2\pi i \left(\mathbf{k}^{(j)} + \mathbf{g}\right) \cdot \mathbf{r}} \right) \tag{14.22}$$

Thus we can write

$$\nabla^2\psi(\mathbf{r}) = -(2\pi)^2 \sum_{\mathbf{g}} \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 C_{\mathbf{g}}^{(j)} e^{2\pi i \left(\mathbf{k}^{(j)} + \mathbf{g}\right) \cdot \mathbf{r}} \tag{14.23}$$

Now we insert this expression in equation 14.19

$$\frac{1}{4\pi^2} \left(-4\pi^2 \sum_{\mathbf{g}} \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 C_{\mathbf{g}}^{(j)} e^{2\pi i \left(\mathbf{k}^{(j)} + \mathbf{g}\right) \cdot \mathbf{r}} \right) +$$
$$K^2 \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i \left(\mathbf{k}^{(j)} + \mathbf{g}\right) \cdot \mathbf{r}} + \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} e^{2\pi i \mathbf{h} \cdot \mathbf{r}} \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i \left(\mathbf{k}^{(j)} + \mathbf{g}\right) \cdot \mathbf{r}} = 0 \tag{14.24}$$

In doing so we replace the **g** in the summation in equation 14.19 by **h** just for clarity (!); both are called 'dummy' variables. If we sum over all the values of a variable we can 'center' the variable wherever we wish. We can further simplify the third term in equation 14.24 by combining the exponential terms and renaming **g**

$$\sum_{\mathbf{g}} \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}}^{(j)} e^{2\pi i \left(\mathbf{k}^{(j)} + \mathbf{g} + \mathbf{h}\right) \cdot \mathbf{r}} = \sum_{\mathbf{g}-\mathbf{h}} \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}-\mathbf{h}}^{(j)} e^{2\pi i \left(\mathbf{k}^{(j)} + \mathbf{g}\right) \cdot \mathbf{r}} \tag{14.25}$$

Now the sum over all **g**–**h** vectors is the same as the sum over all **g** vectors so we replace **g**–**h** by **g**. (Remember, all we are doing is renaming these dummy variables in a consistent way.) Then equation 14.24, and hence 14.19, becomes much simpler

$$\sum_{\mathbf{g}} \left(\left\{ -\left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 + K^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}-\mathbf{h}}^{(j)} \right) e^{2\pi i \left(\mathbf{k}^{(j)} + \mathbf{g}\right) \cdot \mathbf{r}} = 0 \tag{14.26}$$

We can obtain a useful relation by noting that the coefficients of each term in $\exp(2\pi i \mathbf{g}.\mathbf{r})$ must separately be equal to zero. The only way that equation 14.26 can be true is if the term inside the bracket is always zero. The result is a series of equations (one for each value of **g**)

$$\left\{ -\left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 + K^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}-\mathbf{h}}^{(j)} = 0 \tag{14.27}$$

This is another really important set of equations; they restate the Bloch-wave expression of the Schrödinger equation.

Notice that we are not summing over **g** in equation 14.27. The reason for excluding $\mathbf{h} = 0$ from the sum is that we have already included it in the first term.

14.5 THE PLANE-WAVE AMPLITUDES

We can rewrite and reorder equation 14.27 by, yet again, renaming the variable **h** as **g**–**h**. When we do this, we must exclude $\mathbf{h} = \mathbf{g}$ in the sum

$$\left\{ K^2 - \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq \mathbf{g}} U_{\mathbf{g}-\mathbf{h}} C_{\mathbf{h}}^{(j)} = 0 \tag{14.28}$$

The reason for making this change is that it emphasizes that the 'U' terms are the features which couple together the 'C' terms. In other words, this equation tells us how the potential of the crystal, the U terms, mixes the different Bloch waves. The C terms are the Bloch-wave amplitudes. This is the *dynamical coupling* concept.

This equation represents a set of equations which are the fundamental equations of the dynamical theory. (They are called the secular equations in condensed-matter physics texts.) This equation also links the concepts of Bragg beams and Bloch waves.

$U_{\mathbf{g}-\mathbf{h}}$ is the component of the inner potential which couples the Bragg beams with reciprocal-lattice vectors **g** and **h** to one another.

Now we again simplify the situation by limiting the treatment to two beams, O and P; i.e., we consider the case where the only values of $C_{\mathbf{g}}$ which are *non-zero* are $C_0^{(j)}$ and $C_{\mathrm{p}}^{(j)}$ but U_{p} and $U_{-\mathrm{p}}$ are both allowed.

Remember that the superscript on C indicates that $\mathbf{k}^{(j)}$ is a variable. Note that P could be any diffracted beam. Letting $\mathbf{g} = \mathbf{0}$ in equation 14.27 gives

$$\left(K^2 - \left|\mathbf{k}^{(j)}\right|^2\right) C_{\mathbf{0}}^{(j)} + U_{-\mathbf{p}} C_{\mathbf{p}}^{(j)} = 0 \qquad (14.29)$$

In deriving this and the following equation, we consider all the possible values of $\mathbf{h}$ which would give us $C_{\mathbf{0}}^{(j)}$ or $C_{\mathbf{p}}^{(j)}$.

Next let $\mathbf{g} = \mathbf{p}$ in equation 14.28 and reverse the order of terms to emphasize that we have two equations in $C_{\mathbf{0}}$ and $C_{\mathbf{p}}$

$$U_{\mathbf{p}} C_{\mathbf{0}}^{(j)} + \left(K^2 - \left|\mathbf{k}^{(j)} + \mathbf{p}\right|^2\right) C_{\mathbf{p}}^{(j)} = 0 \qquad (14.30)$$

There are no other possible equations, so to solve these two equations we set the determinant of the coefficients equal to zero

$$\begin{vmatrix} K^2 - \left|\mathbf{k}^{(j)}\right|^2 & U_{-\mathbf{p}} \\ U_{\mathbf{p}} & K^2 - \left|\mathbf{k}^{(j)} + \mathbf{p}\right|^2 \end{vmatrix} = \left(K^2 - \left|\mathbf{k}^{(j)}\right|^2\right)\left(K^2 - \left|\mathbf{k}^{(j)} + \mathbf{p}\right|^2\right) - U_{\mathbf{p}} U_{-\mathbf{p}} = 0 \qquad (14.31)$$

The mean inner potential of the crystal is usually $\leq$20 V while the energy of the electrons is $\geq$100,000 eV. Because $|\mathbf{k}^{(j)} + \mathbf{p}|$ and $|\mathbf{k}^{(j)}|$ are both very close to K, it's the difference that is important. Since $\mathbf{P}$ could be any diffracted beam, we can rename it $\mathbf{G}$ to make it look more familiar!

$$\begin{vmatrix} K^2 - \left|\mathbf{k}^{(j)}\right|^2 & U_{-\mathbf{g}} \\ U_{\mathbf{g}} & K^2 - \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 \end{vmatrix} = \left(K^2 - \left|\mathbf{k}^{(j)}\right|^2\right)\left(K^2 - \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2\right) - U_{\mathbf{g}} U_{-\mathbf{g}} = 0 \qquad (14.32)$$

Now we can use the simple algebraic relation

$$x^2 - y^2 = (x - y)(x + y) \qquad (14.33)$$

and make the high-energy approximation that $|\mathbf{k}^{(j)}|$, $|\mathbf{k}^{(j)} + \mathbf{g}|$, and K are all similar in magnitude. Then equation 14.32 becomes

$$\left(\left|\mathbf{k}^{(j)}\right| - K\right)\left(\left|\mathbf{k}^{(j)} + \mathbf{g}\right| - K\right) = \frac{U_{\mathbf{g}} U_{-\mathbf{g}}}{4K^2} = \frac{\left|U_{\mathbf{g}}\right|^2}{4K^2} \qquad (14.34)$$

It is important not to confuse $\mathbf{k}^{(j)}$ with $\mathbf{k}_{\mathrm{I}}$ or $\mathbf{k}_{\mathrm{D}}$ and to remember that $|\mathbf{K}|$ ($= |\mathbf{k}_{\mathrm{D}} - \mathbf{k}_{\mathrm{I}}|$) is not K. Incidentally, it is not until we write this equation that we use the assumption that the crystal has a center of symmetry (see equation 14.4b).

Equation 14.34 is a more complex dispersion relation than equations 14.18 and 14.20. Since $\mathbf{k}^{(j)}$ can point in any direction, this dispersion relation defines a surface, known as the dispersion surface, which is just the locus of all allowed $\mathbf{k}^{(j)}$ vectors for a particular fixed energy. (See Chapter 15.) The simpler relations given in equations 14.18 and 14.20 each defined a sphere; the vectors K and χ can point in any direction.

From equation 14.29 (renaming $\mathbf{p}$ as $\mathbf{g}$), we have

$$\frac{C_{\mathbf{g}}^{(j)}}{C_{\mathbf{0}}^{(j)}} = \frac{\left|\mathbf{k}^{(j)}\right|^2 - K^2}{U_{-\mathbf{g}}} \qquad (14.35)$$

which we can rewrite as

$$\frac{C_{\mathbf{g}}^{(j)}}{C_{\mathbf{0}}^{(j)}} = \frac{\left(\left|\mathbf{k}^{(j)}\right| - K\right)\left(\left|\mathbf{k}^{(j)}\right| + K\right)}{U_{-\mathbf{g}}} \approx \frac{2K\left(\left|\mathbf{k}^{(j)}\right| - K\right)}{U_{-\mathbf{g}}} \qquad (14.36)$$

Thus we can, in principle, say how $C_{\mathbf{0}}^{(j)}$ and $C_{\mathbf{g}}^{(j)}$ are related.

Now we could extend this analysis to show how all the values of C are related in a many-beam situation. If we did that we could write a new expression

$$\mathcal{A}^{(\mathbf{j})}\left\{\mathbf{C}_{\mathbf{g}}^{(\mathbf{j})}\right\} = \mathbf{0} \qquad (14.37)$$

where $\left\{C_{\mathbf{g}}^{(j)}\right\}$ now denotes a column vector with elements $C_{\mathbf{g}}^{(j)}$. $\mathcal{A}^{(j)}$ is a matrix defined by

$$a_{gg} = K^2 - \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 \qquad (14.38)$$

with the off-diagonal elements given by the Fourier coefficients of the crystal potential

$$a_{gh} = U_{g-h} \qquad (14.39)$$

Here, g refers to rows and h to columns in the $\mathbf{A}$ matrix. Except in special cases, such as the two-beam case in equation 14.31, you'll only encounter this formalism in computer programs! A particularly clear case is given by Metherell and is adapted here for five beams, comprising $\mathbf{g}$, $\mathbf{0}$, $\mathbf{g}$, $2\,\mathbf{g}$, and $3\mathbf{g}$ beams. The 5×5 matrix can be written out (using g and h rather than $\mathbf{g}$ and $\mathbf{h}$) as

$$\mathbf{A} = \begin{pmatrix} a_{-g-g} & U_{-g-0} & U_{-g-g} & U_{-g-2g} & U_{-g-3g} \\ U_{0-(-g)} & a_{00} & U_{0-g} & U_{0-2g} & U_{0-3g} \\ U_{g-(-g)} & U_{g-0} & a_{gg} & U_{g-2g} & U_{g-3g} \\ U_{2g-(-g)} & U_{2g-0} & U_{2g-g} & a_{2g2g} & U_{2g-3g} \\ U_{3g-(-g)} & U_{3g-0} & U_{3g-g} & U_{3g-2g} & a_{3g3g} \end{pmatrix} \qquad (14.40)$$

In the first column h is $-g$; in the second, h is zero, etc. In the first row 'g' is $-g$; in the second, 'g' is zero. So we can simplify this matrix as

$$\mathbf{A} = \begin{pmatrix} a_{-g} & U_{-g} & U_{-2g} & U_{-3g} & U_{-4g} \\ U_{g} & a_{0} & U_{-g} & U_{-2g} & U_{3g} \\ U_{2g} & U_{g} & a_{g} & U_{-g} & U_{-2g} \\ U_{3g} & U_{2g} & U_{g} & a_{2g} & U_{-g} \\ U_{4g} & U_{3g} & U_{2g} & U_{g} & a_{3g} \end{pmatrix} \qquad (14.41)$$

Some points to notice are

- The terms $U_{\mathbf{g}}$, $C_{\mathbf{g}}^{(j)}$, and a_g are related by a set of linear equations (the matrix in 14.37).
- We can't solve for actual values of the $C_{\mathbf{g}}^{(j)}$ terms, but we can find the ratios $C_{\mathbf{g}}^{(j)}/C_{\mathbf{0}}^{(j)}$.

We won't take this topic much further here but refer you again to the excellent article by Metherell who shows that equation 14.37 can be expressed as an eigenvalue equation where $\left\{C_{\mathbf{g}}^{(j)}\right\}$ appears as the eigenvectors and the wave vectors $\mathbf{k}^{(j)}$ appear as the eigenvalues. He expresses this equation as

$$M\{C_{\mathbf{g}}^{(j)}\} = \gamma^{(j)}\left\{C_{\mathbf{g}}^{(j)}\right\} \tag{14.42}$$

where the matrix **M** has diagonal elements m_{gg} and off-diagonal elements m_{gh}. The reason we mention this fact here is that the m_{gg} terms correspond to the excitation errors, s_g, and the m_{gh} terms correspond to the extinction distance $\xi_{\mathbf{g}-\mathbf{h}}$. Remember that h is the column and notice that the subscript here is **g**–**h**; this extinction distance is related to the interference between the **g** beam and the **h** beam. Now if you're intrigued and your math is strong, see Metherell's article.

If you're familiar with this math approach, you'll recognize that eigenvectors must satisfy certain relations for normalization and orthogonality. If you look back to Chapter 13, you'll see that we normalized $C_{\mathbf{g}}^{(j)}$ in writing equation 13.30.

As you can see, the math is beginning to become tricky! In the next chapter, we will derive explicit expressions for $\xi_{\mathbf{0}}$ and $\xi_{\mathbf{g}}$ in the two-beam case, namely

$$\xi_{\mathbf{0}} = \frac{2K\cos\theta_B}{U_{\mathbf{0}}} \tag{14.43}$$

and

$$\xi_{\mathbf{0}} = \frac{2K\cos\theta_B}{U_{\mathbf{g}}} = \frac{1}{\Delta\mathbf{k}} \tag{14.44}$$

In our derivation we will use a graphical representation of the dispersion equations. This approach has much in common with the Ewald-sphere/reciprocal-lattice approach to understanding diffraction. It's particularly useful since it gives you, the microscopist, another picture, this time related to imaging.

14.6 ABSORPTION OF BLOCH WAVES

When we have just two beams excited, O and G, we showed in Section 13.9 that we can express the wave function ψ as

$$\psi(\mathbf{r}) = \mathcal{A}^{(1)}b^{(1)}(\mathbf{r}) + \mathcal{A}^{(2)}b^{(2)}(\mathbf{r}) \tag{14.45}$$

where

$$\mathcal{A}^{(1)} = \cos\frac{\beta}{2}; \mathcal{A}^{(2)} = \sin\frac{\beta}{2} \tag{14.46}$$

We can plot these curves for $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ in relation to the positions of the atoms in a simple-cubic crystal where the electron beam is close to the [001] zone axis. Figure 14.2 shows that the intensity in Bloch wave 1 is centered on the column of atoms (Figure 14.2A) while that in Bloch wave 2 is centered between the atoms (Figure 14.2B). (If you read Hirsch et al., you should note that they have 1 and 2 reversed.) Therefore Bloch wave 1 interacts more strongly with the column of atoms and will be 'absorbed' preferentially. Conversely, Bloch wave 2 will be channeled through the specimen. The intensity in the **g** beam depends on the thickness of the specimen because of the interference between these two Bloch waves. This preferential absorption means that we may expect to 'lose' this thickness dependence even though we can still 'see' through the specimen. We'll return to this topic in Chapter 24.

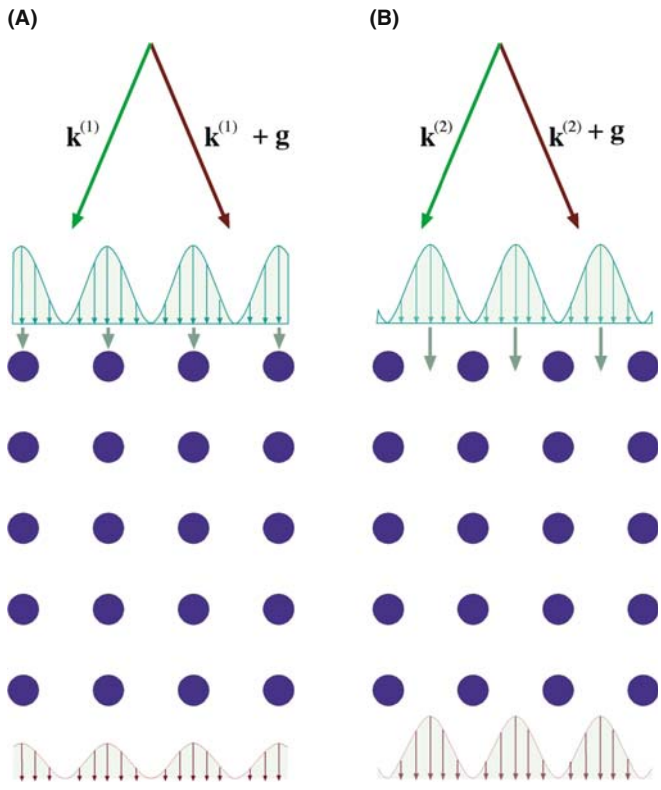


FIGURE 14.2. The two types of Bloch wave in the crystal aligned at the Bragg condition: (A) the maximum lies along the ion cores and Bloch wave 1 interacts strongly; (B) the maximum lies between the ions so that the interactions are weaker.

CHAPTER SUMMARY

We told you at the beginning of the chapter that this discussion would seem to be just theory or manipulating equations. There are, however, some really important ideas

- A basic property of a crystal is that its inner potential, $V(\mathbf{r})$, is periodic and positive.
- An electron in a crystal can be described by a sum of Bloch waves which themselves are solutions to the Schrödinger equation.
- The wave functions $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$ are not solutions to this equation and therefore don't actually exist as waves in the crystal.
- All Bloch waves have the same total energy.

Therefore, if we really want to understand what goes on in the crystal, we must be able to understand the concept of Bloch waves. However, you can understand how to relate images to the structure of the specimen without considering Bloch waves. You just have to accept that the analysis using beams (hence $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$) is phenomenological. Equations 14.27 and 14.28 give you the essential clue to what really happens: each set of equations tells us how the Bloch waves are coupled. When you've worked through this chapter, go through Sections 13.8 and 13.9 again.

There are many possible solutions to the Schrödinger equation, and each Bloch wave is a plane wave; that is, it can be associated with well-defined propagation vector $\mathbf{k}^{(j)}$ as shown in equation 14.9.

The Bloch waves are generally different because the $U_{\mathbf{g}}$ terms are different, i.e., they have different potential energies. Therefore, they have different kinetic energies and different wave vectors.

Finally, a word on relativity. We've kept our treatment as simple as possible, but you should remember that the equations should be relativistically corrected; most texts have ignored relativistic effects when discussing this topic.

REFERENCES

This chapter follows the treatment given by Hirsch et al. in Chapter 9 of their classic text; the details are provided by Metherell.

BLOCH WAVES

Ashcroft, NW and Mermin, ND 1976 *Solid State Physics* W.B. Saunders Co. Philadelphia PA. Chapter 8 ($2\pi/\lambda$ is used).

Bethe, HA 1928 *Theorie der Beugung von Elektronen an Kristallen* Ann. Phys. Lpz. **87** 55–129. Another classic reference (in German).

Howie, A 1971 in *Electron Microscopy in Materials Science* 275–305 Ed. U Valdré Academic Press New York.

Kittel, CJ 2004 *Solid-State Physics* 8th Ed. John Wiley & Sons New York. For the physicists.

Metherell, AJF 1975 in *Electron Microscopy in Materials Science* **II** 397–552 Eds. U Valdré and E Ruedl CEC, Brussels. This is perhaps the clearest and most comprehensive article available on this subject (over 150 pages long). It is strongly recommended reading if you've made it through this chapter and want to begin programming.

THE COMPANION TEXT

EMS can be used to simulate HRTEM images using the Bloch-wave approach.

SELF-ASSESSMENT QUESTIONS

Q14.1 Summarize the reasons for assigning the sign we do to the accelerating voltage, the sign of the energy of the electron, the sign of the potential, and the sign of the potential energy.

Q14.2 Even when the atomic number varies from 4 to 74, the inner potential energy only changes by a factor of ~3. Why?

Q14.3 You're told that the inner potential energy of W is 23.4. Comment on this value.

Q14.4 What is a Bloch wave?

Q14.5 What is a Bloch function?

Q14.6 What is the Bloch theorem?
Q14.7 Why are we talking about plane waves throughout this chapter?
Q14.8 Write down an equation relating K and χ.
Q14.9 Write down an expression relating K to E outside the crystal.
Q14.10 Write down the Bloch-wave expression for the Schrödinger equation. What does this equation tell you?
Q14.11 Explain in words the meaning of the term $U_{\mathbf{g}-\mathbf{h}}$.
Q14.12 In a crystal, light slows down but electrons speed up. What is the fundamental difference between these two phenomena?
Q14.13 The terms $U_{\mathbf{g}}$, $C_{\mathbf{g}}^{(j)}$ and $a_{\mathbf{g}}$ are related by a set of linear equations through the matrix **A**. Write down the matrix **A** for the situation where only $\overline{\mathbf{g}}$, 0, **g**, 2**g**, and 3**g** are important.
Q14.14 In the two-beam diffraction situation, how many Bloch waves are important?
Q14.15 The 'U' terms are the features which couple together the 'C' terms. What does this mean?
Q14.16 $U_{\mathbf{g}-\mathbf{h}}$ is the component of the inner potential which couples together the Bragg beams with reciprocal-lattice vectors **g** and **h**. We don't really have Bragg beams in a crystal. Explain this apparent inconsistency.
Q14.17 How is it possible for all Bloch waves to have the same energy?
Q14.18 In the TEM image, we lose thickness dependence even though we can still see through the specimen. How does Bloch wave theory explain this effect?
Q14.19 If your sample is YAG (bcc with $a = \sim 1$ nm) rather than a model, mono-atomic, simple-cubic crystal, how will your Bloch wave analysis be affected?
Q14.20 If you have a $Si/SiO_2/Si$ sandwich of equally thick (20 nm) layers such that the beam passes through the layers in succession, how will this affect your discussion of Bloch waves?

TEXT-SPECIFIC QUESTIONS

T14.1 For discussion, compare how the velocity of electrons and light change when they enter a crystal. Consider (i) how this will be affected if the specimen is amorphous; (ii) if the accelerating voltage is increased from 100 kV to 1 MV; (iii) if the microscope were a TPM (p: positron) instead.
T14.2 Consider equation 14.44. Calculate a value for $\Delta\mathbf{k}$ when $s = 0$ for 100 kV, Cu 220. How does this value compare to **g** and λ^{-1} and why do we ask?
T14.3 Evaluate equation 14.20 for 100- kV electrons.
T14.4 How does Figure 14.1A differ for [001], [110], and [111] in Si and Cu?
T14.5 How would you modify Figure 14.1B to explain the values of Table 14.2 more clearly?
T14.6 Consider Figure 14.2. Imagine that the crystal is an extinction distance thick. Draw the intensity plots for the Bloch waves at depth $\xi/4$, $\xi/2$, $3\xi/4$, and ξ.
T14.7 Excluding the other III–V compounds which are closely related to GaAs, list five other crystal types that do not have a center of symmetry.
T14.8 Consider equations 14.1 and 14.2. Where does the relativistic correction enter?
T14.9 In equations 14.45 and 14.46, $\mathcal{A}^{(1)^2} + \mathcal{A}^{(2)^2}$ is 1. Why is this?
T14.10 The vectors **k** and $\boldsymbol{\chi}$ differ because of the refractive-index effect. How large is the refractive index and how does this quantity relate to (if it does) the refractive index for light?
T14.11 (Challenging) If the extinction distance depends on $U_{\mathbf{g}}$ for a crystal, can amorphous specimens show thickness fringes?
T14.12 (Challenging) Why does $\xi_{\mathbf{g}}$ decrease when both 220 and $2\bar{2}0$ are excited?

15

Dispersion Surfaces

CHAPTER PREVIEW

The analysis of Bloch waves given in the previous chapter is closely related to the classic analysis of waves that you've seen in condensed-matter physics or semiconductor theory. In semiconductors in particular, we often talk of band diagrams and indirect or direct band gaps. We use terms like conduction bands, valence bands, and Brillouin-zone boundaries (BZBs). We visualize these quantities by drawing diagrams of $E(\mathbf{k})$, the electron energy (which is a function of **k)** versus **k**, the wave vector. This plot of $E(\mathbf{k})$ versus **k** is known as a dispersion diagram. To remind you of the magnitudes involved, the band gap in Si is 1.1 eV and that of Ge is 0.7 eV. In a good insulator, it can be 10 eV. We now follow the same approach to represent pictorially what we described in equations in Chapters 13 and 14. Remember that the big difference from the solid-state physics approach is that the energy of the electrons in the beam in the TEM is $\geq$100 keV.

In this chapter we will see the real origin of the extinction distance $\xi_{\mathbf{g}}$, which we introduced in equation 13.4. We will discuss how it relates to particular materials and why it varies with the diffraction vector being used. We will then discuss the physical origin of the concept of the effective extinction distance: i.e., the value which the extinction distance appears to have when $s \neq 0$. This discussion of dispersion surfaces is included as a separate chapter, so that you can omit it without affecting your understanding of the rest of the text. We should give you a warning: this is a subject which has probably turned off many potential microscopists. It can be very mathematical, pure theoretical physics, or it can provide many useful insights into image formation. We are trying for the latter. If we aren't completely successful, take heart; nearly every established microscopist has survived without completely mastering this concept!

15.1 INTRODUCTION

The analysis of Bloch waves as they apply to electrons in solids is well documented in the condensed-matter physics literature. However, what we want from the theory is different from what an electrical engineer might want: we want to understand how it applies to the formation of contrast in TEM images and DPs. With this aim in mind, we will again follow the treatment given in Metherell's classic and well-hidden article, already referenced in Chapters 13 and 14. In Chapter 14, we derived equations relating **k** to $U_{\mathbf{g}}$. (See Section 14.2 for the definition of $U_{\mathbf{g}}$.) Specifically, we found that there are two Bloch waves if there are two Bragg beams, **0** and **g**. We can rewrite equation 14.35 incorporating equation 14.32 as

$$\frac{C_{\mathbf{g}}^{(j)}}{C_{\mathbf{0}}^{(j)}} = \frac{\left(\mathbf{k}^{(j)}\right)^2 - K^2}{U_{-\mathbf{g}}} = \frac{U_{\mathbf{g}}}{\left(\mathbf{k}^{(j)} + \mathbf{g}\right)^2 - K^2} \qquad (15.1)$$

where $C_{\mathbf{0}}^{(j)}$ is the amplitude of the plane wave with wave vector $\mathbf{k}^{(j)}$, and $C_{\mathbf{g}}^{(j)}$ is the amplitude of the plane wave with wave vector $\mathbf{k}^{(j)} + \mathbf{g}$. The Bloch wave was given in equation 14.12 as

$$b^{(j)}(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i\left(\mathbf{k}^{(j)}+\mathbf{g}\right)\cdot\mathbf{r}} \qquad (15.2)$$

Equation 15.1 says that the values of $C_{\mathbf{g}}^{(j)}$ and $C_{\mathbf{0}}^{(j)}$ are directly related to $k^{(j)2}$–K^2 and thus to $k^{(j)}$–K.

In the general many-beam case (actually, in any situation where we have more than two beams), the situation is more complicated. However, we can separate the problem into two parts

- Determine all the allowed wave vectors $\mathbf{k}^{(j)}$ in a crystal, including all possible orientations of the crystal.
- Determine which set of the allowed $\mathbf{k}^{(j)}$ wave vectors is actually present when you fix the orientation of your crystal.

The first statement fixes the total energy of the electron and selects the crystal. The second statement applies the boundary conditions for the particular situation you are considering, as we'll illustrate in Sections 15.5 and 15.6.

The solution to the first part of the problem is found by setting $|\mathcal{A}^{(j)}| = 0$. (We defined $\mathcal{A}^{(j)}$ in Section 14.3 and gave an expression for it in Section 14.5.) If you multiply the determinant, you get a polynomial to the power $2n$ in $\mathbf{k}^{(j)}$

$$\mathcal{A}_{2n}\left(\mathbf{k}^{(j)}\right)^{2n} + \mathcal{A}_{2n-1}\left(\mathbf{k}^{(j)}\right)^{2n-1} + \mathbf{K} = 0 \qquad (15.3)$$

The coefficient $\mathcal{A}_n$ depends on K^2 (i.e., the energy) and $\mathbf{g}$ (i.e., the crystal).

So, the polynomial in $\mathbf{k}^{(j)}$ relates $\mathbf{k}^{(j)}$ to the total energy. This is a dispersion relation as we defined the term in Section 14.4. The equation has $2n$ roots and some might be complex. To quote Metherell, "at first sight therefore, the situation appears to be a complicated one!" So in following Metherell we make two simplifications

- We consider only the high-energy case.
- We assume that we only excite reflections in the ZOLZ.

There are three reasons for reminding you of these simplifications

- If you want to make a Bloch-wave calculation where you include more than two Bragg beams, then you will need a computer.
- The diagrams we're considering in this chapter are a pictorial representation. The diagrams help us think about what is actually happening to the Bloch waves. If we just did the calculation, we would lose the physical 'feel' for the problem.
- None of the diagrams we will draw will consider HOLZ reflections; if we make the beam energy high enough, we don't need to consider them. However, the energy is not really that high and HOLZ reflections are not only seen experimentally, but can also provide valuable information, as we'll see in Chapters 20 and 21. The saving factor is that modern computers have no problems in handling these equations, especially since they are so amenable to matrix manipulation.

THE DISPERSION SURFACE
. . .is a pictorial representation of the relationship between **k** and energy.

15.2 THE DISPERSION DIAGRAM WHEN $U_g = 0$

We start with equation 14.34, namely

$$\left(\left|\mathbf{k}^{(j)}\right| - K\right)\left(\left|\mathbf{k}^{(j)} + \mathbf{g}\right| - K\right) = \frac{|U_g|^2}{4K^2} \qquad (15.4)$$

Remember that this equation was derived for the two-beam case. When the electrons are in the vacuum, i.e., outside the specimen, the Fourier coefficients U_g are 0. When $U_g = 0$, the left side of this equation is zero and the equation has two solutions.

$$K = \left|\mathbf{k}^{(j)}\right| \text{ or } K = \left|\mathbf{k}^{(j)} + \mathbf{g}\right| \qquad (15.5)$$

where j is 1 or 2. If we plot out these two solutions we find, as shown in Figure 15.1, that we have two interpenetrating spheres, since both $\mathbf{k}_I$ and $\mathbf{k}_D$ can lie in any direction. Since these two **k** vectors have the same length, the two spheres represent surfaces of constant energy, called dispersion surfaces, one centered on O and the other centered on G.

Of course, we already know that the energy of the electron in a vacuum is related to its wave vector by

$$E = \frac{p^2}{2m} = \frac{\mathrm{h}^2\chi^2}{2m} \qquad (15.6)$$

where **p**, the momentum, is related to the wave vector in a vacuum, $\boldsymbol{\chi}$, by $\mathbf{p} = \mathrm{h}\boldsymbol{\chi}$. Here, $\boldsymbol{\chi}$ is the **K** when the electron is in a vacuum.

Rearranging, we have

$$\chi = \left\{\frac{2m}{\mathrm{h}^2}E\right\}^{\frac{1}{2}} \qquad (15.7)$$

The dotted line drawn in Figure 15.1 represents a plane that is defined by the circle created by the intersecting spheres. You will probably be very familiar with the BZB from condensed-matter physics.

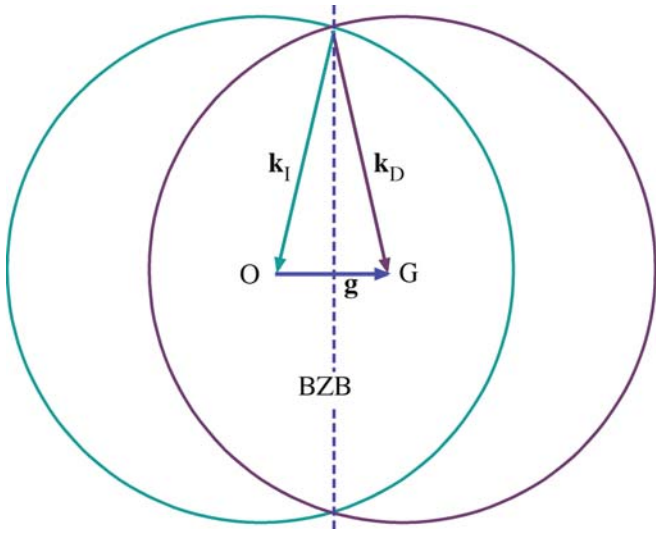


FIGURE 15.1. Cross section through two spheres of radii $\mathbf{k}_I$ and $\mathbf{k}_D$ centered on O and G, respectively. The spheres represent surfaces of constant energy and the dotted line is the trace of the diffracting plane (and is also equivalent to the BZB).

While you work through the diagrams in this chapter, you must remember that for high-energy electrons the scattering angles, e.g., $2\theta_B$, are usually small and the region of interest in reciprocal space is, therefore, close to the BZB. We can redraw part of Figure 15.1 to show an enlarged view of the region close to the BZB in Figure 15.2. At high energies, we approximate the surfaces as a pair of straight lines because λ is very small.

15.3 THE DISPERSION DIAGRAM WHEN $U_g \neq 0$

When $U_g \neq 0$ we know from equation 15.4 that K can never be equal to $|\mathbf{k}_I|$ or $|\mathbf{k}_D|$. Since equation 15.4 is quadratic we must have two values for $|\mathbf{k}|$. So, the two 'spheres' can't intersect if $U_g \neq 0$. You noticed that equation 15.4 *resembles* that for a hyperbola, $xy = a$, where the x and the y axes are shown in Figure 15.2. We can draw these two hyperbolae with their asymptotes as shown in Figure 15.3. These surfaces (remember we are in three dimensions) are known as *branches* of the dispersion surface. The upper branch (identified here by the '1') corresponds to $\mathbf{k}^{(1)}$ and the lower (identified by the '2') to $\mathbf{k}^{(2)}$. We now have vectors $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$ where we used to just have $\mathbf{K}_I$. There are some critical points to remember in this discussion from Chapters 13 and 14

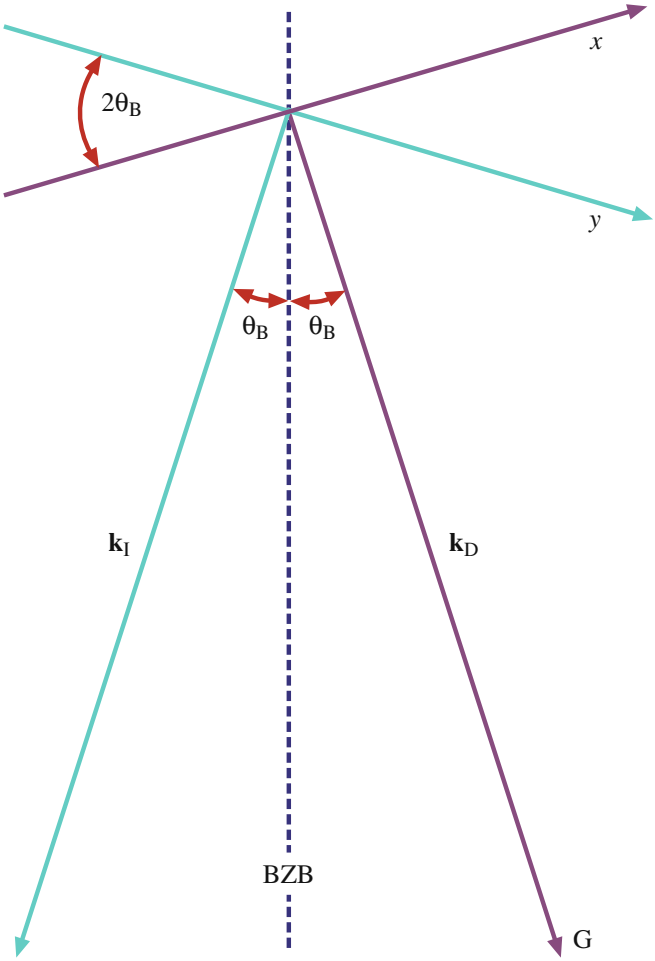


FIGURE 15.2. An enlarged view of the interception of the two dispersion spheres at the BZB. The projections of the two dispersion surfaces approximate to straight lines x and y which are normal to $\mathbf{k}_D$ and $\mathbf{k}_I$, respectively.

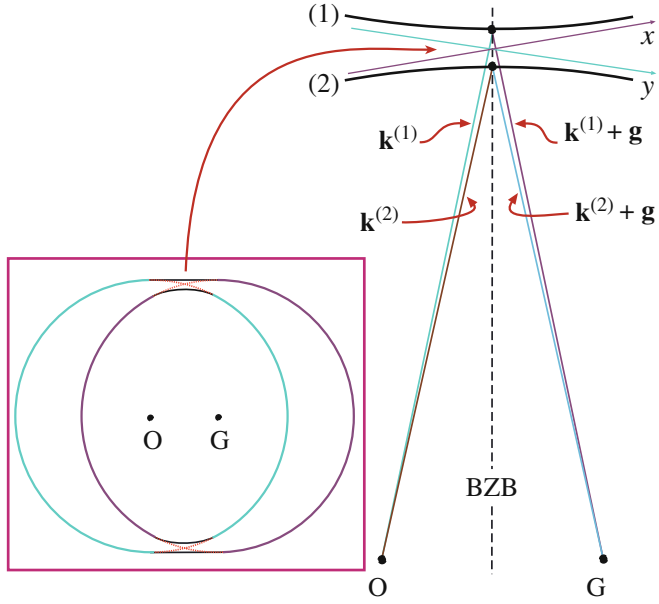


FIGURE 15.3. When the electron is inside the specimen (i.e., $U_g \neq 0$) there are two values of $\mathbf{k}$. The two dispersion spheres can't intersect and two branches of the dispersion surface (1) and (2) are created. The enlarged view of the 'non-intersection' shows the vectors, $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$, and $\mathbf{k}^{(1)} + \mathbf{g}$ and $\mathbf{k}^{(2)} + \mathbf{g}$.

- The Bloch wave $b^{(1)}(\mathbf{k}^{(1)}, \mathbf{r})$ is associated with $\mathbf{k}^{(1)}$.
- The Bloch wave $b^{(2)}(\mathbf{k}^{(2)}, \mathbf{r})$ is associated with $\mathbf{k}^{(2)}$.
- The intensity of the Bragg beam is a function of thickness $|\phi_g(t)|^2 \propto \sin^2(\pi t \Delta k)$ (from equation 13.45).

The difference between Figures 15.1 and 15.3 is the gap between the two branches in Figure 15.3. This gap is present because U_g is not zero; U_g is not zero because we have a periodic array of atoms, i.e., a crystal. This gap is directly analogous to the band gap in semiconductor theory where there are forbidden electron energies within the crystal.

15.4 RELATING DISPERSION SURFACES AND DIFFRACTION PATTERNS

We can gain a lot of physical insight into Bloch waves using the dispersion-surface construction rather than solving the Bloch-wave equations on the computer. Our approach is relatively simple.

- We start with the dispersion surface shown in Figure 15.4A and draw an initial line to represent the incoming beam traversing the thin foil. We assume an idealized thin specimen with parallel surfaces that are perpendicular to the vertical, optic axis of our TEM and we choose **g** to be parallel to the surface. The incident beam is allowed to be inclined to the surface of the specimen.
- We then draw a line normal to any surface that the initial line encounters. This allows us to match the components of wave vectors parallel to that surface. This is the wave-matching construction.

WAVE MATCHING
If we were doing this mathematically, this matching would be the boundary condition.

- We then extend the points M_1 and M_2 back to the '**k**' spheres in Figure 15.4B; these spheres are the **χ** spheres when we are in the crystal.
- The last part of the process is always to relate the waves in the crystal to the beams in the vacuum since our recording film, CCD camera, etc., is always outside the crystal.

That's the plan—now we go through it step by step. As shown on the enlarged view in Figure 15.5, each of the **k** vectors has an associated wave amplitude $C_{\mathbf{g}}^{(j)}$ associated with it.

In this discussion, we will limit ourselves to the two beams, O and G. As we know from Section 13.8, the only values of C (the coefficients of the Bloch waves) that will then be non-zero are $C_{\mathbf{0}}^{(1)}$, $C_{\mathbf{0}}^{(2)}$, $C_{\mathbf{g}}^{(1)}$, and $C_{\mathbf{g}}^{(2)}$.

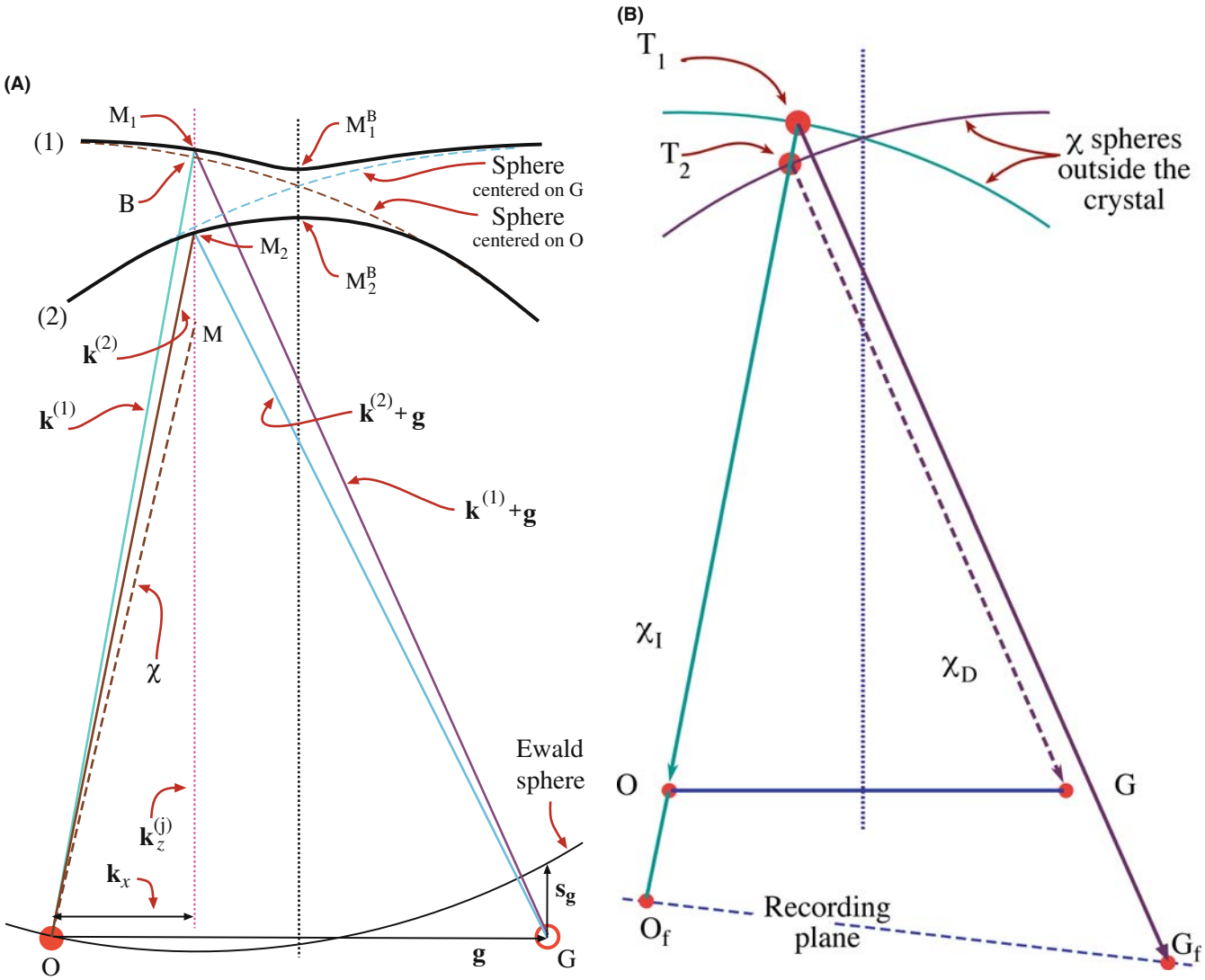


FIGURE 15.4. (A) Combination of the dispersion surfaces (1) and (2), centered on O and G, with the Ewald sphere construction. The surface of the specimen has been set to be parallel to **g**, so points $M_1{}^B$ and $M_2{}^B$ on the branches (1) and (2) are excited. The incident beam direction is then parallel given by the vector **MO**. If we tilt the beam so χ (as shown) becomes more vertical, the excited points move to M_1 and M_2 giving the tie line M_1M_2. The vectors $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$ start at M_1 and M_2, respectively, and end on O. (B) Extension of the lines OM_1 and OM_2 in (A) back to the χ spheres at T_1 and T_2, respectively, relates the waves in the crystal to the beams outside. The points O_f and G_f are what you record on the photographic film.

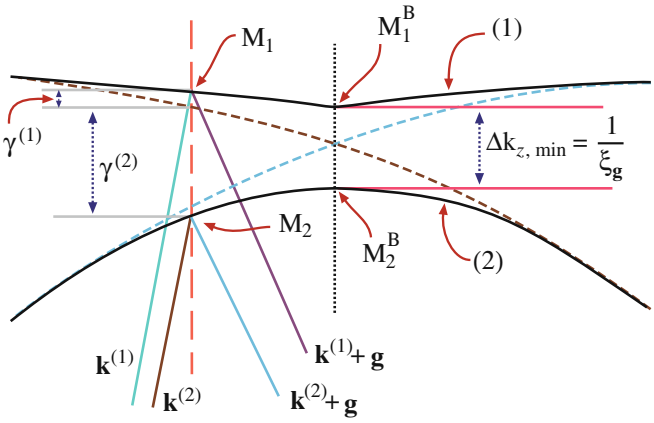


FIGURE 15.5. An enlarged region of Figure 15.4A showing how the vectors $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$ are related to the quantities $\gamma^{(1)}$ and $\gamma^{(2)}$ and the distance Δk_z.

First, we need to know which points on the dispersion surface will actually correspond to the diffraction condition we have chosen. We also need to know the orientation of the specimen relative to the beam and the orientation of the Bragg planes (which is why we start by fixing this).

To begin. We consider the case where the surface of the specimen is parallel to **g**; we will explain why we are so specific on this point in a moment.

Now we have fixed the specimen and **g** relative to the optic axis. Next we orient the incident beam. Note that the beam is not parallel to the (*hkl*) planes; its $\mathbf{k}_1$ is determined by $\boldsymbol{\chi}$ but isn't shown in the figure. We will excite points M_1^B and M_2^B on separate branches of the dispersion surface because this is where the dotted line cuts the two surfaces. The extinction distance will then correspond to Δk^{-1} for $s = 0$ as in Section 13.10. If we now tilt the incident beam so that $\boldsymbol{\chi}$ moves closer to the vertical (keeping the specimen fixed), then the excited points become M_1, and M_2 and, as we see in Figure 15.4, s becomes negative.

We define the lines $M_1^B M_2^B$ and M_1M_2 to be *tie lines* because they tie together points on the different branches of the dispersion surface. Both tie lines are parallel to the BZB because we chose the top surface of the specimen to be parallel to **g**.

Each of these tie lines is normal to the surface that produces it.

The diagrams of the dispersion surface in Figures 15.4 and 15.5 contain lots of reminders

- For this orientation, $\mathbf{k}_x$ is the same for all **k** vectors ending on O.
- You can recognize $\gamma^{(1)}$ and $\gamma^{(2)}$ from Section 13.7.
- The vacuum wave vector $\boldsymbol{\chi}$ is always shorter than **K** or **k**.

We can understand these changes from the following argument. The O beam is always excited so $C_0^{(1)}$ and $C_0^{(j)}$ will always be relatively large. Which other values of C are large will depend on where the Ewald sphere cuts the systematic row of relrods.

Now we can consider what happens when the surface of the specimen is *not* parallel to **g**. Here, the normal to the surface, **n**, is not parallel to the BZB (since the BZB is normal to **g**). However, the tie line is always parallel to **n** so the tie line is no longer parallel to the BZB. Remember: this construction is graphically matching the components of the **k** vector which are parallel to the surface of the specimen. We saw this clearly in Figure 15.4 where we commented that $\mathbf{k}_x$ is the same for all the vectors ending on O because we chose **g** to be parallel to the surface and the surface to be normal to the optic axis in that case.

We don't need tie lines in solid-state physics if the electrons are always moving in a perfect lattice where we don't consider surfaces.

TIE LINES

The tie line is a graphical method of satisfying the boundary conditions imposed by the TEM specimen.

We are now ready to consider the more common TEM wedge specimen shown in Figure 15.6A and then we'll see how these excited Bloch waves relate to the DP.

In this figure we have drawn the wedge with the top surface horizontal. Thus we have tie line $\mathbf{n}_1$ along the optic axis. When the electrons exit the crystal at the inclined bottom surface, we again match components parallel to this surface so we have tie line $\mathbf{n}_2$. Notice that we must draw $\mathbf{n}_2$ through both M_1 and M_2. These tie lines don't excite extra points on the dispersion surface because we are leaving the crystal.

Once we're outside the crystal, we know that the wave vector must be $\boldsymbol{\chi}$ and that $\boldsymbol{\chi}$ defines a pair of spheres centered on O and G. So we extend the $\mathbf{n}_2$ tie lines until they reach the $\boldsymbol{\chi}$ spheres. Now we have excited four points, as we see graphically in Figure 15.6A. The points on the O circle are labeled O_1 and O_2; those on the G circle are D_1 and D_2. We have labeled the subscripts this way because they correspond to the plane waves $\boldsymbol{\chi}_0^{(1)}$, $\boldsymbol{\chi}_0^{(2)}$, etc., as also shown in Figure 15.6A.

To conclude. Now we have reached the final step. We have to relate these beams to the DP. Yes, they are real beams, not Bloch waves, because we are now outside the specimen and in a vacuum. We show this in Figure 15.6B. All of the $\boldsymbol{\chi}$ beams have been related to point O_1 because $\boldsymbol{\chi}_0^{(1)}$ is the incident beam. Remember: $\boldsymbol{\chi}_0^{(1)}$ is not vertical because we made **g** horizontal and tilted the incident beam. The vectors $\boldsymbol{\chi}_0^{(1)}$ and $\boldsymbol{\chi}_0^{(2)}$ are not quite parallel because although they are both radii of the same sphere of radius χ, they actually originate at different points on the circle (see Figure 15.6A).

The conclusion is that we will have two spots at O and two spots at G. In other words, the fact that we have

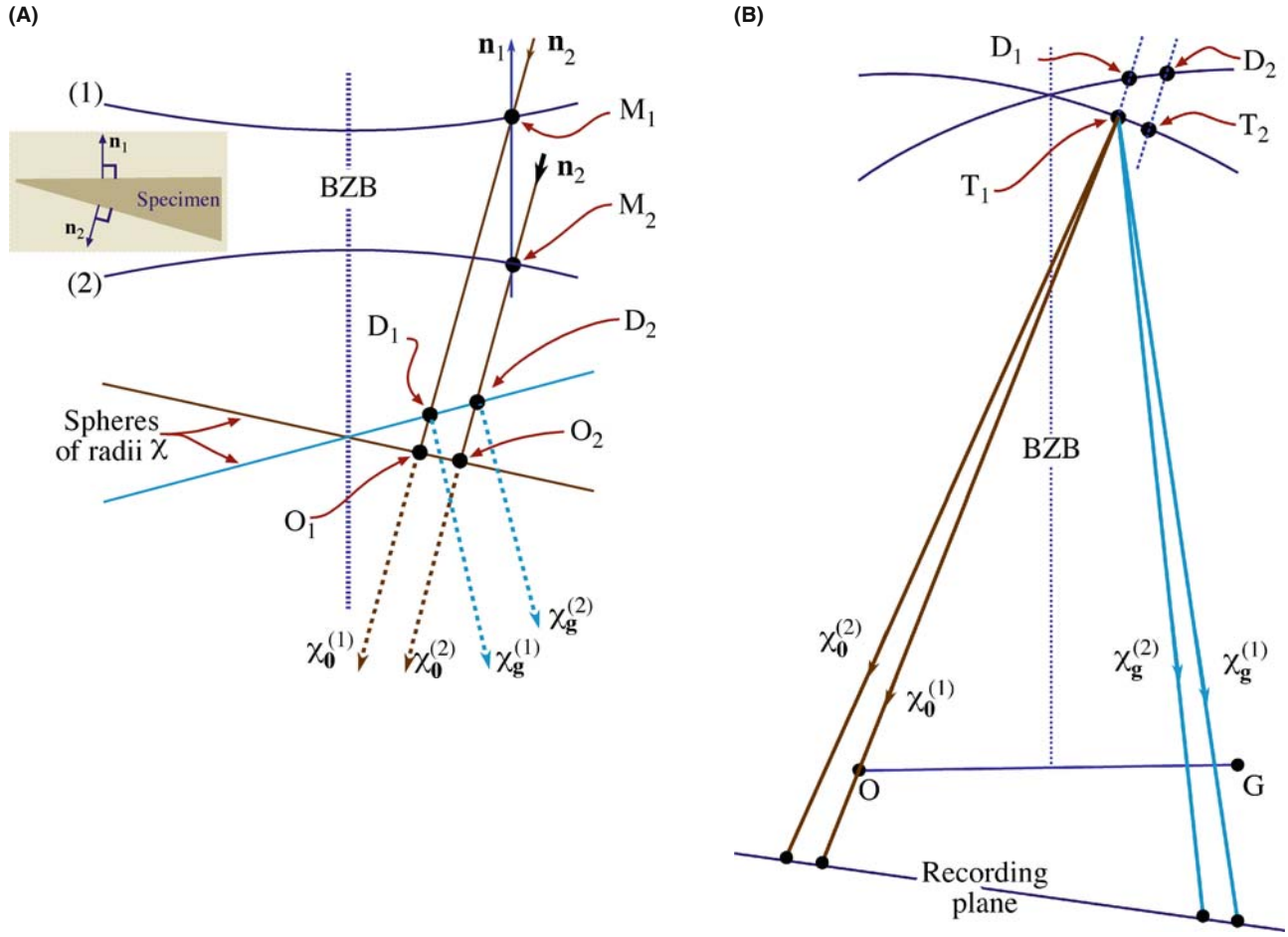


FIGURE 15.6. (A) The same diagram as Figure 15.4B, but for a wedge specimen with the top surface parallel to **g** (normal $\mathbf{n}_1$) and the bottom surface normal $\mathbf{n}_2$. Instead of exciting two points, O_1 and O_2, we excite two more, D_1 and D_2, which correspond to the plane waves $\chi_0^{(1)}$, $\chi_0^{(2)}$, outside the crystal. In (B) we relate all the beams to the point O_1 and we produce two beams at O and two at G. Thus we can predict that a wedge foil will give doublets at O and G.

a wedge specimen has split the spots at G. We will see these split spots in Chapter 18 and we will return to this topic in Chapter 23 when we discuss images.

It can be useful to extend the wedge case to the double wedge. For example, imagine an inclined planar defect in a parallel-sided slab with **g** parallel to the slab surface as shown in Figure 15.7. Everything is as before at the top surface. At the inclined interface then, tie lines do create new excited points B_1 and B_2 on the 1 and 2 branches of the dispersion surface.

Now, $\mathbf{n}_3$ is the tie line due to the bottom surface and $\mathbf{n}_3$ is parallel to $\mathbf{n}_1$. We extend the $\mathbf{n}_3$ tie lines to the χ spheres and find that now we have three χ_0 vectors and three χ_D vectors. Translating these χ vectors to O_1 as the common origin produces the beam diagram shown in Figure 15.7B. Now we have three spots at O and three spots at G. We will return to this topic in Chapter 24 when we discuss images of planar defects, but here let's summarize the new concepts they give us

- The dispersion surface is a graphical approach to thinking about Bloch waves.
- We have to match the components of any wave entering and leaving any surface, internal or external.
- We use the exit-surface tie line to link to the χ spheres.
- Having two inclined surfaces causes a splitting of the Bragg beams.
- An internal interface, such as a stacking fault, increases the number of points excited on the dispersion surfaces and the number of spots at reflection G!

To understand the importance of these ideas, try to imagine what will happen when a defect, which is not abrupt, is present in the crystal (more on this in Section 15.8).

15.5 THE RELATION BETWEEN U_g, ξ_g, AND s_g

We can best appreciate the importance of the dispersion-surface construction by looking at Figure 15.4. This figure shows the original spheres as dashed lines: they are nearly flat close to the BZB. The electron beam is

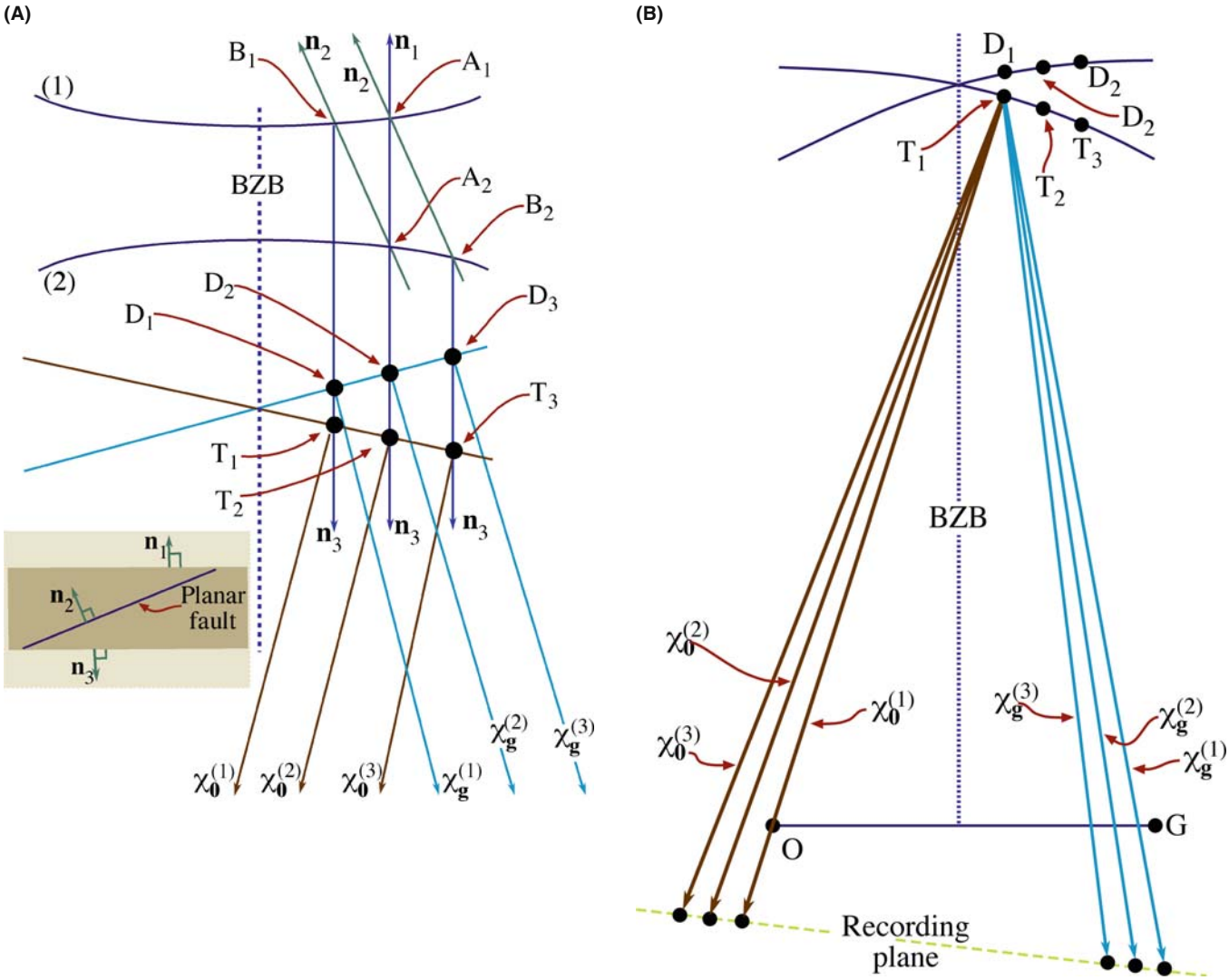


FIGURE 15.7. (A) An enlarged view of the dispersion surface in Figure 15.6 close to the BZB, but this time for a specimen in which both surfaces are parallel to **g** but there is an inclined fault which produces a third wave $\chi_0^{(3)}$ and $\chi_g^{(3)}$. If we then move all the vectors to O_1 again, we predict there will be three spots at O and three at G.

initially traveling with wave vector χ outside the crystal. When the beam enters the crystal the z component of this wave vector changes (this is the refraction effect we saw in Chapters 11 and 13), but the xy component is unchanged. Therefore, the allowed **k** vectors in the crystal are $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$. One **k** vector begins on branch 1 and ends at O, while the other begins at branch 2 and ends at O.

BRANCHES AND BEAMS

There are only two **k** vectors because there are only two branches of the dispersion surface. There are two branches of the dispersion surface because we have a crystal potential (hence $U_{\mathbf{g}}$). There are only two branches because we are considering only two beams.

Clearly we can draw in $\mathbf{k}_g^{(2)}$ and $\mathbf{k}_g^{(1)}$ by adding **g**. Now, how does $\mathbf{k}_0^{(1)}$, say, relate to **K**? The point K is also determined by the tie line through $\boldsymbol{\chi}$, and lies on the circle centered on O. Most importantly, neither $\mathbf{k}_1$ nor $\mathbf{k}_2$ is equal to **K**. If you look back at equation 13.41 you can see that

$$k_z^{(i)} - K_z = \gamma^{(i)} \tag{15.8}$$

So $\gamma^{(i)}$ is simply the distance of the point M_j from the K sphere centered on O. We can write this relationship explicitly

$$\mathbf{k}^{(i)} = \mathbf{k}_z^{(i)} + \mathbf{k}_x^{(i)} \tag{15.9}$$

$$= \left(K + \gamma^{(i)}\right)\mathbf{u}_z + k_x\mathbf{u}_x \tag{15.10}$$

Notice that the last term here is independent of i. Look again at Figure 15.4. We can see that Δk_z is a minimum when M_1 and M_2 lie on the BZB. In that situation

$$\Delta k_{z_{\min}} = \gamma^{(1)} - \gamma^{(2)} \tag{15.11}$$

Simply by looking at the diagram, and as expected from Chapter 13, you also know that

$$\gamma^{(1)} - \gamma^{(2)} = \frac{U_{\mathbf{g}}}{k} = \frac{1}{\xi_{\mathbf{g}}} \tag{15.12}$$

So

$$\Delta k_z = \frac{1}{\xi_{\mathbf{g}}} \tag{15.13}$$

The origin of the thickness oscillations that we will see in the two-beam TEM image is the difference in wavelength of the two Bloch waves. It's the beating between the two Bloch waves.

Thus we see that the gap Δk_z at the BZB is given by the reciprocal of the extinction distance. We'll summarize again to be quite sure that it is clear!

- We have a crystal, therefore $U_{\mathbf{g}} \neq 0$.
- Since $U_{\mathbf{g}} \neq 0$ we have two branches to the dispersion surface and hence a band gap.
- The bandgap is Δk_z.
- Hence we have a finite extinction distance (i.e., $\xi_{\mathbf{g}}$ is not infinite).

An aside: think how s_{eff} and s would be related if $\xi_{\mathbf{g}}$ were infinite. (Go back to equation 13.47.)

If the tie line M_1M_2 does not lie on the BZB then when we draw the Ewald sphere centered just below M_1 (with radius of length $1/\lambda$ or $|K|$) we see that $s_{\mathbf{g}}$ is nonzero. We can easily see from the equations in Section 13.10 that, in general, Δk_z is given by

$$\Delta k_z = s_{\text{eff}} = \frac{1}{\xi_{\text{eff}}} \tag{15.14}$$

This equation is the key to understanding the origins of the extinction distance and why the effective extinction distance depends on the size of the excitation error, **s**. It says that the band gap increases as we increase **s**. Looking at it another way, as we move the tie line off the BZB, the band gap $\Delta\mathbf{k}$ increases.

Some questions raised here are

- What is the physical reason that Δk_z is related to **s**?
- What happens if **g** is not parallel to the foil surface or, indeed, if the foil surfaces are not parallel to one another?

You can also appreciate why we had a problem defining **s** when we first encountered it!

15.6 THE AMPLITUDES OF BLOCH WAVES

In Section 13.9, we found that the total wave function for the two-beam case can be expressed as the sum of two Bloch waves

$$\psi(\mathbf{r}) = \mathcal{A}^{(1)}b^{(1)} + \mathcal{A}^{(2)}b^{(2)} \tag{15.15}$$

We showed that the relative contributions of the two Bloch waves $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ are $\cos\beta/2$ and $\sin\beta/2$, respectively; in addition, $w = \cot\beta = s\xi_{\mathbf{g}}$.

We also showed in Section 13.8 that

$$b^{(1)}\left(\mathbf{k}^{(1)}, \mathbf{r}\right) = C_{\mathbf{0}}^{(1)} e^{2\pi i \mathbf{k}^{(1)} \cdot \mathbf{r}} + C_{\mathbf{g}}^{(1)} e^{2\pi i \left(\mathbf{k}^{(1)} + \mathbf{g}\right) \cdot \mathbf{r}} \tag{15.16}$$

and

$$b^{(2)}\left(\mathbf{k}^{(2)}, \mathbf{r}\right) = C_{\mathbf{0}}^{(2)} e^{2\pi i \mathbf{k}^{(2)} \cdot \mathbf{r}} + C_{\mathbf{g}}^{(2)} e^{2\pi i \left(\mathbf{k}^{(2)} + \mathbf{g}\right) \cdot \mathbf{r}} \tag{15.17}$$

The Bloch wave coefficients were given by equation set 13.31

$C_0^{(1)}$	$C_0^{(2)}$	$C_g^{(1)}$	$C_g^{(2)}$
$\cos\beta/2$	$\sin\beta/2$	$-\sin\beta/2$	$\cos\beta/2$

Now we can consider some special cases and examine the actual values for $C_{\mathbf{0}}^{(1)}$, $\mathcal{A}^{(1)}$, etc. (Table 15.1).

For the Bragg case, $s_{\mathbf{g}} = 0$, **g** is exactly excited and $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ are both equal to $1/\sqrt{2}$. In other words, the two Bloch waves are equally excited.

For the case where $s_{\mathbf{g}} < 0$, we now have $\cos(\beta/2) > \sin(\beta/2)$ so that $\mathcal{A}^{(1)}$ is greater than $\mathcal{A}^{(2)}$. If we reverse the sign of s, $\cos(\beta/2) < \sin(\beta/2)$ and $\mathcal{A}^{(1)}$ is less than $\mathcal{A}^{(2)}$.

TABLE 15.1 Values of Bloch Wave Variables

s	w	β	$\beta/2$	$\cos(\beta/2)$	$\sin(\beta/2)$
0	0	$\pi/2$	$\pi/4$	$1/\sqrt{2}$	$1/\sqrt{2}$
+0.01	$+\Delta$	$(\pi/2)-\delta$	$(\pi/4)-(\delta/2)$	$(1/\sqrt{2})+\epsilon$	$(1/\sqrt{2})-\epsilon$
−0.01	$-\Delta$	$(\pi/2)+\delta$	$(\pi/4)-(\delta/4)$	$(1/\sqrt{2})-\epsilon$	$(1/\sqrt{2})+\epsilon$

AMPLITUDE OF BLOCH WAVE

Whether Bloch wave 1 or Bloch wave 2 has the largest amplitude depends on the sign of s.

Now, let's relate this information to the dispersion surface shown in Figure 15.4. When $s_{\mathbf{g}} < 0$, as shown here, the M_1M_2 tie line is to the left of the BZB, which

is associated with reflection G. When the tie line is closer to O than G, Bloch wave 1 is more strongly excited; the reverse is true when the tie line crosses the BZB. We should remember that the analysis in Chapter 13 was for a two-beam case, where we were close to the Bragg condition. So this discussion of $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ only applies to small values of **s**.

15.7 EXTENDING TO MORE BEAMS

If we allow more beams to contribute to the image, we can imagine the dispersion surface for the case where $U_{\mathbf{g}} = 0$ by constructing more spheres, shown in Figure 15.8. If we have n beams then we will have n spheres. Note that each sphere is centered on its corresponding reciprocal-lattice point and neighboring spheres intersect periodically spaced BZBs. The gap in Figure 15.3 always occurs at the BZB. The BZB itself always corresponds to a plane which is the perpendicular bisector of a **g** vector. Thus the diagram for >2 beams shown in Figure 15.8 will become more complicated with many band gaps and many branches as shown in Figure 15.9. The magnitude of the band gap does decrease as the rank of the neighboring branches increases.

In Chapter 27, we'll discuss what happens in images when 3**g** is excited. We will actually consider the two-beam condition where **0** and 3**g** are the two beams.

We follow the convention used by Metherell and number the branches of the dispersion surface from top down.

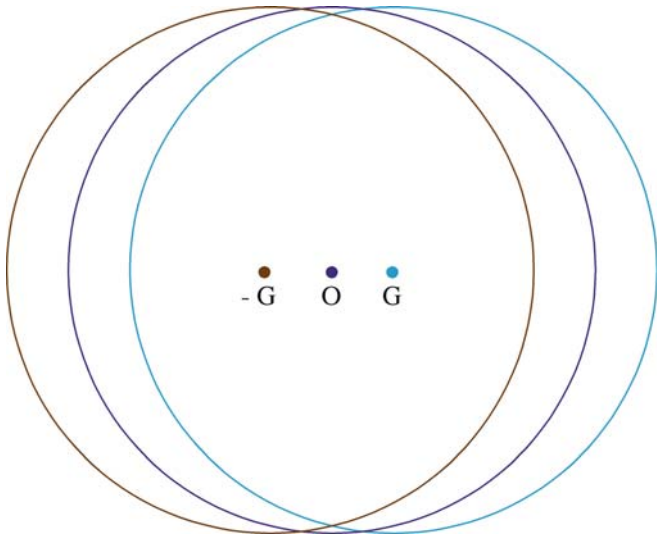


FIGURE 15.8. Three dispersion spheres due to three reflections, –G, O, and G. If we had n spots we would have n spheres.

Then $i = 1$ corresponds to the branch with the highest kinetic energy. Remember that all the electrons have the same total energy in this treatment. You must also be aware that some earlier texts number the top branch two and the second branch one, following Hirsch et al. This was fine when only two branches were considered.

We can still associate the amplitudes $C_{\mathbf{0}}$, $C_{\mathbf{g}}$, etc., with the sphere centered on **0**, **g**, etc. The result is shown by the labels $C_{\mathbf{0}}$, $C_{\mathbf{g}}$, etc., in Figure 15.9. For

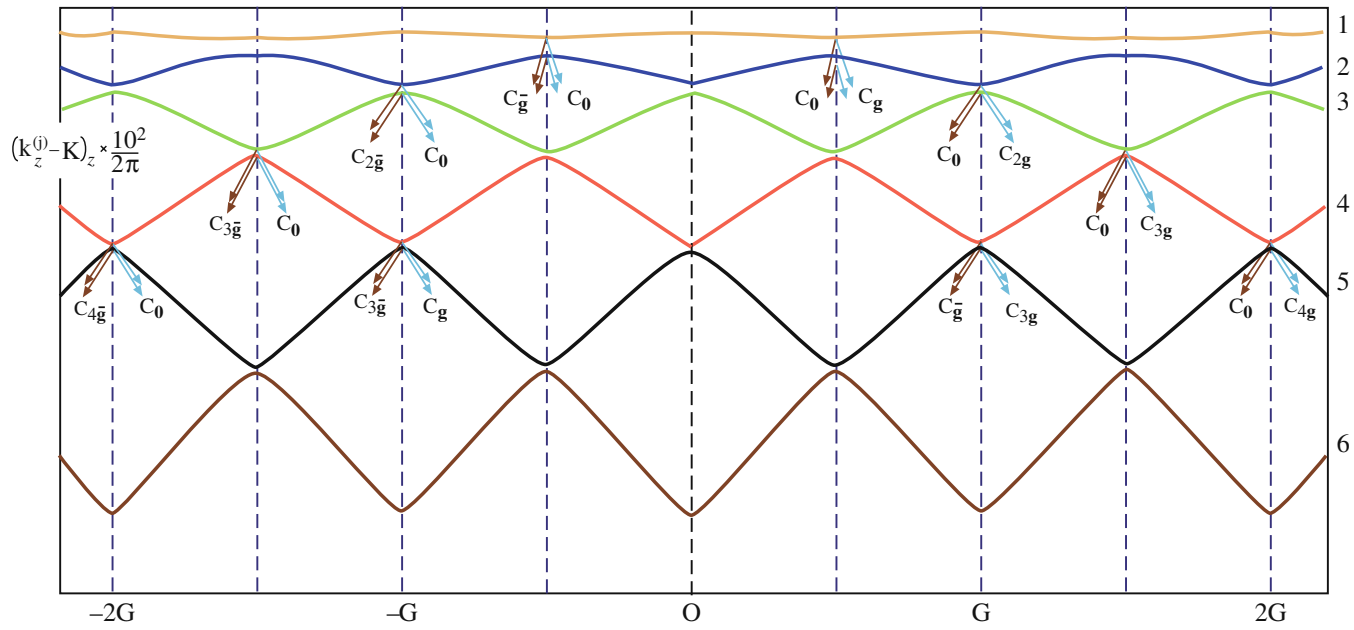


FIGURE 15.9. Six branches of the dispersion surfaces. The two branches $i = 1$ and $i = 2$ have the highest energy and give the largest band gap; notice that these branches give the terms in $C_{\mathbf{0}}$ and $C_{\mathbf{g}}$; smaller gaps occur between branches with lower energy. The diagram can be approximated to a set of spheres centered on O, ±G and ±2G, etc.; $C_{\mathbf{0}}$ is 'normal' to the sphere centered on O, while $C_{\mathbf{g}}$ is 'normal' to the sphere centered on **g**, etc.

example, imagine the original spheres centered on **0** and **g**; they intersect on the BZB which passes through **g**/2 so the $C_{\mathbf{0}}$, $C_{\mathbf{g}}$ are labeled as shown.

Similarly, the spheres centered on **0** and 3**g** intersect on the BZB which passes through 3**g**/2 so $C_{\mathbf{0}}$, $C_{3\mathbf{g}}$ are labeled. As a general rule, $C_{n\mathbf{g}}$ will be largest for the pair of reflections which are excited, i.e., **0** and n**g** and will be related by the n**g**/2 BZB.

We now extend these arguments to the situation where many beams are excited. Values of C other than $C_{\mathbf{0}}$ and $C_{n\mathbf{g}}$ will be non-zero since it's no longer a two-beam case. So the tie line M_1M_2 will then intersect many branches of the dispersion surface. The reason these contributions are smaller when **g** is excited is that they do not intersect the **0** circle. However, they can contribute to the image. Figure 15.9 shows how this can be visualized. (Remember the dispersion surface is a way of visualizing Bloch wave coefficients.) If we satisfy reflection 2G, then $C_{\mathbf{0}}^{(1)}$, $C_{\mathbf{0}}^{(2)}$, $C_{2\mathbf{g}}^{(1)}$, and $C_{2\mathbf{g}}^{(2)}$ are all large. The gap $\Delta k_{4,5}$ between branch 4 and branch 5 at G (on the BZB) is small; the 'circles' would have intersected in the vacuum. If we think about the Ewald sphere we can show that the **s** values for $\bar{\mathbf{g}}$ and 3**g** are identical. We'll see later (Chapter 26) that these reflections will actually couple strongly, although both are weakly excited and the extinction distance is large (because the gap Δk_{4-5} is small). The extinction distance for the coupling of $\bar{\mathbf{g}}$ and 3**g** when 2**g** is strongly excited is $\xi_{4\mathbf{g}}$ ($\xi_{3\mathbf{g}-(-\mathbf{g})}$). We can see this is true by looking at the branch 4/5 gap on the BZB for G.

15.8 DISPERSION SURFACES AND DEFECTS

The original reason for introducing the concept of Bloch waves was that only Bloch waves can exist in a periodic potential, i.e., there are no beams in the crystal. So what happens when a defect is present? We'll discuss this situation in some detail in Chapters 23–26 but will mention the basic ideas here, emphasizing the Bloch waves rather than the defects.

In Section 15.4, we discussed the effect that a planar fault can have on the Bloch waves using the dispersion-surface representation. What we were actually doing was matching the components parallel to the planar defect, so the effect of the planar fault was to create new tie lines $\mathbf{n}_2$. The general result is that, when a defect is present, energy is transferred from one Bloch wave to the other along the tie line; this is known as *interband scattering*. This concept is not only important for our understanding of images of planar defects but also illustrates a general principle for defects.

The difficulty with non-planar defects is that the tie lines are not so well defined. You can, however, imagine the result: instead of having points on the dispersion surface, we will have a distribution of points. We then relate this distribution to the DP. We do this with the tie lines normal to the exit surface and then translate to O_1 in the usual way. So, our distribution of points on the dispersion surface will become a distribution of spots in the DP; this distribution is what we will call a streak in Chapter 17.

CHAPTER SUMMARY

Dispersion surfaces allow us to draw diagrams to represent the equations given in Chapter 14. These surfaces are essentially plots of the **k** vector of the Bloch waves (which is directly related to the energy) versus the **K** vector. They correspond directly to the band diagrams, which are used extensively to represent energy levels in semiconductors; the difference is that in semiconductors, we emphasize energy by plotting energy versus reciprocal-lattice vector (our **K** vector). The **k** vectors themselves vary because, although the total energy of each electron is a constant, the potential energy decreases when the electron is close to the nucleus, causing the kinetic energy to increase.

The most important equation for imaging theory is 15.14, which relates Δk_z, s_{eff}, and ξ_{eff}. Notice that Δk_z is defined for two Bloch waves but is only small when the Bragg equation is nearly satisfied. This relationship links Bloch waves and Bragg beams. Δk is non-zero because we have a crystal which produces a periodic potential. Δk gives rise to thickness fringes and all thickness effects. Thus we see that thickness variations are due to the interference, or beating, of pairs of Bloch waves. As we increase n, $\xi_{\mathbf{g}}$ increases because the gap between the two relevant branches of the dispersion surface becomes narrower. Defects present in the crystal cause a mixing or coupling of the Bloch waves: they 'tie' the branches of the dispersion surface and cause interband scattering.

We've emphasized throughout this chapter that the dispersion surface is a pictorial representation of the **k** versus K relationship. We'll close by quoting the result derived by Kato.

In any wave field, the direction of energy flow is along the normal to the surface of the dispersion surface. This result is equally valid for 'electron wave packets' and other waves. The physicist might say that the Poynting vector is normal to the dispersion surface.

Although there are many texts that discuss dispersion surfaces and band gaps in semiconductors, beware of the $2\pi/\lambda$ versus $1/\lambda$ problem since many of the texts are written by, and for, physicists. Defect analysis using Bloch waves has generally been the preserve of the physicist. However, there are some excellent programs available which use a Bloch-wave approach analysis.

We give the usual caveat: beware of black boxes. Metherell's article goes to greater depth than covered here. However, it has been an inspiration for much of this chapter and is highly recommended for advanced study. It is beautifully written and explained, but is certainly more advanced than our text. If you want to delve deeper into this topic, this is *the* article. Note that Metherell uses the e^{ikr} notation.

REFERENCES

As in Chapter 14 we follow the treatment of Hirsch et al. as extended and illustrated by Metherell. For anyone familiar with Mathematica™ (or the corresponding) MatLab it would be an interesting challenge to construct (and share) notebooks for these diagrams.

BLOCH WAVES

Ashcroft, NW and Mermin, ND 1976 *Solid State Physics* W.B. Saunders Co. Philadelphia. Chapter 8 ($2\pi/\lambda$ is used).

Kato, N 1957 *The Flow of X-rays and Materials Waves in Ideally Perfect Single Crystals* Acta Cryst. **11** 885–887.

Kittel, CJ 2004 *Solid-State Physics* 8th Ed. John Wiley & Sons New York.

Metherell, AJF 1975 in *Electron Microscopy in Materials Science* **II** 397–552 Eds. U Valdré and E Ruedl CEC Brussels. *The reference*.

THE COMPANION TEXT

The companion text doesn't extend this topic much, but there is hope to have a discussion of the use of Mathematica™ notebooks in the future.

SELF-ASSESSMENT QUESTIONS

Q15.1 What does the word dispersion mean in the context of this chapter?

Q15.2 Write down an expression for $b^{(j)}(\mathbf{r})$ in terms of $\mathbf{k}^{(j)}$ and $C_{\mathbf{g}}{}^{(j)}$.

Q15.3 We restrict our discussion to the case where only reflections in the ZOLZ are excited. What is the simple physical reason for doing this?

Q15.4 Draw the dispersion surface diagram for the two-beam case where the crystal has zero inner potential; then explain the term 'dispersion surface.'

Q15.5 The vectors $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$ are not parallel to one another. What is the physical reason for this and does it have any implications?

Q15.6 Why do $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$ end (or begin) on branches (1) and (2) and not on the 'curves' x and y?

Q15.7 We used to have $\mathbf{K}_D$. Now we have $\mathbf{k}^{(1)} + \mathbf{g}$ and $\mathbf{k}^{(2)} + \mathbf{g}$. Why don't we have $\mathbf{K}_D$? Does $\mathbf{K}_D$ still exist?

Q15.8 What is a BZB? Write out the name and explain what causes it.

Q15.9 Draw the dispersion surface for two beams where $\mathbf{g}$ is not satisfied. (Draw Figure 15.4 without looking at it).

Q15.10 Consider Figure 15.4 when the incident beam, χ_1, is parallel to the diffracting planes. Why is this not a good example of two-beam diffraction?

Q15.11 In Figure 15.4, M is 'below' the dispersion surface. Can it be above this surface? Justify your answer.

Q15.12 In Figure 15.4, $\mathbf{k}^{(2)} + \mathbf{g}$ is longer than $\mathbf{k}^{(2)} + \mathbf{g}$. Does this mean that λ has changed?

Q15.13 In Figure 15.4, why is the tie line M_1M_2 normal to $\mathbf{g}$?

Q15.14 The wave vector $\boldsymbol{\chi}$ is always shorter than $\mathbf{K}$ or $\mathbf{k}$. Explain why this is so.

Q15.15 The tie line is a graphical method of satisfying the boundary conditions imposed by the specimen. Explain why this is so.

Q15.16 In Figure 15.6, are points O_1, O_2, D_1, and D_2 really relevant since they don't lie on the dispersion surface for the crystal?

Q15.17 Explain in words why there are three χ_0 vectors ($\chi_0{}^{(1)}$, $\chi_0{}^{(2)}$, and $\chi_0{}^{(3)}$) in Figure 15.7.

Q15.18 In the two-beam case when G is on the Ewald sphere, what are the magnitudes of the Bloch wave coefficients?

Q15.19 If $s_{\mathbf{g}}<0$, we can show that $\mathcal{A}^{(1)} > \mathcal{A}^{(2)}$. What does this mean physically?

TEXT-SPECIFIC QUESTIONS

The following questions contain the manufacturers' warning label.

T15.1 Why are there two expressions for $C_{\mathbf{g}}^{(j)}/C_{\mathbf{0}}^{(j)}$ in equation 15.1?

T15.2 What is the physical reason that Δk_z is related to s?

T15.3 If **g** is not parallel to the foil surface, how will the Bloch-wave construction change?

T15.4 If the two surfaces of the specimen are not parallel to one another, how will the Bloch-wave construction change?

T15.5 When the specimen is oriented so that G is on the Ewald sphere, we again have a situation like that shown in Figure 15.9; many beams are excited but one does have $s_{\mathbf{g}} = 0$. You form a BF image. How does the periodicity of the thickness fringes relate to this diagram? In particular, how is the extinction distance related to the two-beam value?

T15.6 When the specimen is oriented with the beam parallel to the diffracting planes, we have a situation like that shown in Figure 15.9; many beams are excited but none have $s_{\mathbf{g}} = 0$. You form a BF image. How does the periodicity of the thickness fringes relate to this diagram?

T15.7 We initially follow Metherell's simplification of assuming that only reflections in the ZOLZ are excited. Discuss what complications might arise if we did not make this assumption. Are we likely to encounter this in practice, and if so, under what conditions?

T15.8 We also made the approximation that the accelerating voltage is high. Discuss when this approximation might break down.

T15.9 In Figure 15.4, we draw OG so that it is not parallel to O_fG_f. Could this really happen? Discuss why T_1G_f is parallel to T_2G and should this be so?

T15.10 In Figure 15.6B, why is $\chi_0^{(2)}$ not parallel to $\chi_0^{(1)}$ and why are there two such vectors when we are outside the crystal? What is the relevance of D_2 and T_3 in this figure?

16

Diffraction from Crystals

CHAPTER PREVIEW

Since our emphasis is on crystalline materials, we will first discuss how the details of the crystal symmetry affect the DPs we expect to see. What we're doing here is taking the concepts of the reciprocal lattice and applying it to particular examples. There are two basic lessons.

- You must learn some of the rules that we will derive for particular crystal structures; one example will be to determine which reflections are allowed for an fcc crystal.
- The other lesson is more general and is really concerned with why we have these rules. Why are certain reflections absent or weak and how can you use this information to learn more about your material?

We can deduce some selection rules for different crystal structures that tell you which reflections are allowed. We suggest you learn the most common ones by heart. Throughout this chapter, we'll assume that the crystal is perfect and infinite, which it never is. In Chapter 17, we will examine what happens when we include defects or allow the diffracting crystal to become relatively small. In Chapter 18, we'll go through the process of indexing experimental DPs.

16.1 REVIEW OF DIFFRACTION FROM A PRIMITIVE LATTICE

In Chapters 11–15, we examined diffraction from a regular array of lattice points. We will now define such an array as a primitive lattice where there is only one lattice point in the unit cell. Actually, we did begin to consider the present topic when we discussed the meaning of n in the Bragg equation $2d \sin \theta_B = n\lambda$ in Sections 11.5 and 12.4. We showed that the diffraction from the (200) planes would give rise to a 200 reflection even when there were only atoms on the (100) planes.

By combining equations 13.3 and 13.4, we can see that the amplitude of the diffracted beam is given by

$$\phi_{\mathbf{g}} = \frac{a\,\mathrm{i}\lambda F_{\mathbf{g}}}{V_c \cos\theta} \sum_n e^{-2\pi \mathrm{i}\mathbf{K}\,.\,\mathbf{r}_n} e^{2\pi \mathrm{i}\mathbf{K}_D\,.\,\mathbf{r}}$$

where $F_{\mathbf{g}}$ is the structure factor for the material. Since the same type of atom was at each lattice point, we only needed to consider one atomic scattering factor f in Chapter 13. Now we are going to include different types of atoms as we build up real crystal unit cells. From Section 3.7 we know that f varies with the scattering angle. However, in this chapter we are going to restrict ourselves to small values of θ (excluding zero) and will assume that we have fixed values of f; you can easily extend this analysis to other scattering angles. For convenience, we've summarized some useful values of f in Table 16.1.

If you study the original paper of Ibers, from which these data were taken, you will appreciate that these numbers are not really well known. This is unfortunate since much of our analysis depends on the values of f. Furthermore, we have an additional reason for choosing θ not to be zero in Table 16.1 because these values are even less reliable. Fortunately, what saves us is that we are only interested in the details of the intensities in some special cases and then the effects are really insensitive to the precise value of f.

We are just going to take these numbers and move on, but you may want to investigate a little further. Some points you should consider are

- Why are these numbers not better known? We discussed this topic in Chapters 2 and 3. The atomic scattering factor is related to the differential scattering cross section (Section 3.7)

TABLE 16.1 Selected Values of $f(\theta)$, the Atomic Scattering Amplitude at $\theta = \theta_B$

Element	$f(\theta)$ (Å)	Element	$f(\theta)$ (Å)
H	0.31	Ca	3.40
Li	0.75	Cr	3.56
Be	1.16	Mn	3.55
B	1.37	Fe	3.54
C	1.43	Co	3.51
N	1.44	Ni	3.48
O	1.42		
		Cu	3.44
Na	1.59	Zn	3.39
Mg	1.95	Ga	3.64
Al	2.30	As	4.07
P	2.59	Ag	5.58
		W	7.43

These are values were derived using a self-consistent field theory ($\sin\theta/\lambda$ = 0.2 Å^{-1}) and are based on the rest mass. The $f(\theta)$ value must be multiplied by $(1 - (v/c)^2)^{-1/2}$ for electrons with velocity v.

$$|f(\theta)|^2 = \frac{d\sigma(\theta)}{d\Omega} \tag{16.2}$$

and the cross section is not well known at typical TEM voltages.

- If the crystal is ionic, do we use $f(\theta)$ for the atom or for the ion?
- If the material is covalently bonded, how can we incorporate the bonds into our scattering model?

How we calculate $f(\theta)$ depends on the model we use to describe the atom. You can find more details in the references at the end of the chapter, but beware, this is not an easy topic.

The simplest method is just to ignore any ionic character! If you look at Table 16.1, you'll see that if the atomic number is large enough, then the change in f caused by removing an electron may not be great. In ionic materials, we form ions by removing or adding outer electrons so the interaction of the electron beam with the nucleus is not significantly affected. However, you should remember that this argument applies only to f. We'll see in Part 4 that we can detect differences between differently bonded atoms using EELS. So, bonding does affect the electron beam.

The overall effect of the covalent, i.e., directional, component of the bonding is usually ignored. However, as you realize, all the bonds in Si, for example, are aligned along one particular type of crystallographic direction so you may indeed be able to detect some special features in the DPs.

16.2 STRUCTURE FACTORS: THE IDEA

In this section, we are building on Chapter 12. To keep things simple, we will illustrate the concept of the structure factor for cubic crystals. If we have a simple-cubic crystal, then all possible values of $\mathbf{g}$ can give a reflection in the DP. Each reciprocal-lattice point will then correspond to a possible beam. The next step will be to add the basis (i.e., the group of atoms associated with each lattice point) to the primitive lattice. Since we still have the primitive lattice, all of these points will still exist in the reciprocal lattice but the reflections will be weighted. You will find that there are three different ways to look at the situation, which in fact are all equivalent

- *Selection rules*: This is perhaps closest to physics. The structure of the crystal imposes certain selection rules which determine which beams are allowed.
- *Weights* (or *weighting factors*): We can assign a weight (which may be zero) to each of the points in the reciprocal lattice. This is the terminology used by Ewald. The nice feature about weighting factors is that they are analogous to scattering factors.
- *Structure factors* (F): These are the unit-cell equivalents of the atomic-scattering amplitude, $f(\theta)$; they can be thought of as unit-cell scattering amplitudes. This is the terminology favored in materials science.

There are two ways to address this topic:

- We can examine the physical idea of interference as we did in Chapters 2 and 3. This approach can give some useful guidelines to you, the experimentalist. For example, we'll see that the 200 reflection in Si should usually be absent; it should always be present, though weak, in GaAs. Similarly, in Ni_3Al, the 100 reflection is weak, but in Ni it is absent.
- Some materials have a special lattice in real space, for example, fcc or bcc lattices. In these cases, we can describe a corresponding special lattice in reciprocal space. What this means is that certain reflections are always forbidden for these particular structures; these are known as 'kinematically forbidden' reflections. (We'll see, however, that they can be present due to dynamical scattering events, and structure factors do not take any account of dynamical scattering.) The reciprocal lattice (of allowed reflections) of an fcc crystal is bcc, and vice versa.

In equation 13.1 we described the scattering from the unit cell by the expression

$$A_{\text{cell}} = \frac{e^{2\pi i k r}}{\mathbf{r}} \sum_i f_i(\theta) e^{2\pi i \mathbf{K} \cdot \mathbf{r}_i} \tag{16.3}$$

What this equation says is that the atoms within the unit cell all scatter with a phase difference given by $2\pi i \mathbf{K} \cdot \mathbf{r}_i$ where $\mathbf{r}_i$ is a vector which defines the location of each atom within the unit cell

$$\mathbf{r}_i = x_i\,\mathbf{a} + y_i\,\mathbf{b} + z_i\,\mathbf{c} \tag{16.4}$$

We'll start by considering only the case where $\mathbf{K} = \mathbf{g}$ since this is an infinite, perfect crystal

$$\mathbf{K} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \tag{16.5}$$

So we can write

$$F_{hkl} = \sum_i f_i e^{2\pi i(hx_i + ky_i + lz_i)} \tag{16.6}$$

This is our key equation; it is completely general.

This equation applies whether there is one atom or one hundred atoms in the unit cell, no matter where they are located, and it applies to all crystal lattices. What we do next is simply insert the atomic coordinates into equation 16.6 and calculate F_{hkl}.

16.3 SOME IMPORTANT STRUCTURES: BCC, FCC AND HCP

We will now calculate the structure factor for bcc and fcc crystals, because they illustrate the points we just made in Section 16.2 and because, as a materials scientist, you must know these results. You can regard the reciprocal lattice in two ways

- The reciprocal lattice for bcc and fcc are themselves special lattices.
- All reciprocal lattices of cubic materials are simple cubic, but some of the lattice points have a zero structure factor.

Body-centered cubic: The bcc structure is particularly easy. If we set the origin on one lattice point at (0, 0, 0), the other lattice point is at (1/2, 1/2, 1/2) and we substitute these values of (x, y, z) into equation 16.6, then

$$F = f\left\{1 + e^{\pi i(h+k+l)}\right\} \tag{16.7}$$

Now, since h, k, l are all integers, if we define the sum $h + k + l = N$, then the exponential can take two values: +1, for N even, and –1, for N odd.

Thus, we can say that

- $F = 2f$ if $h + k + l$ is even,
- $F = 0$ if $h + k + l$ is odd.

There are **no** other possibilities. The resulting indexed bcc reciprocal lattice is shown in Figure 16.1. This lattice of allowed reflections is face-centered cubic. The reason it may not look like the familiar fcc lattice in real space is that the indices in reciprocal space must all be integers.

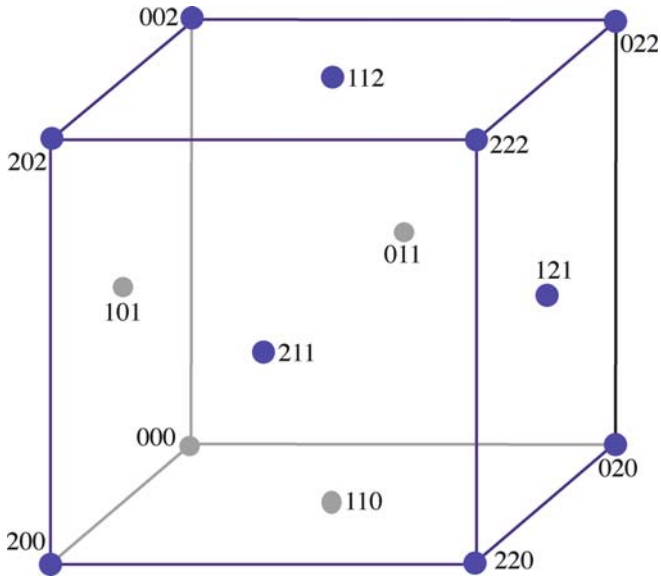


FIGURE 16.1. The reciprocal lattice for the bcc crystal structure. The lattice points that correspond to systematic absences have been removed, so the actual arrangement of points is an fcc lattice.

Face-centered cubic: If we take the same approach for the fcc structure, we now have to include four atoms in the unit cell. We can view this cell as simple cubic with a four-atom basis. The coordinates of the atoms are

$$(x, y, z) = (0, 0, 0), \left(\frac{1}{2}, \frac{1}{2}, 0\right), \left(\frac{1}{2}, 0, \frac{1}{2}\right), \left(0, \frac{1}{2}, \frac{1}{2}\right) \tag{16.8}$$

Substituting these values for $\mathbf{r}_i$ into equation 16.6 gives

$$F = f\left\{1 + e^{\pi i(h+k)} + e^{\pi i(h+l)} + e^{\pi i(k+l)}\right\} \tag{16.9}$$

Again, we consider the possible values of the integers h, k, l. If all three are either odd or even, then all of the exponential terms are $e^{2n\pi i}$. Therefore, all the phases of the diffracted waves are multiples of 2π and are in phase. However, if one of h, k, or l is odd but the other two even, or vice versa, then two of the three phase factors will be odd multiples of π giving two terms of –1 in equation 16.9. Therefore

- $F = 4f$ if h, k, l are all even or all odd,
- $F = 0$ if h, k, l are mixed even and odd.

The resulting lattice is shown in Figure 16.2. This time the reciprocal lattice of allowed reflections is bcc with all the indices being integers.

Hexagonal close-packed: Generally DPs from hcp crystals are more difficult to index for two reasons

- Except for (0001), the patterns can be different for every material because the c/a ratio is different.

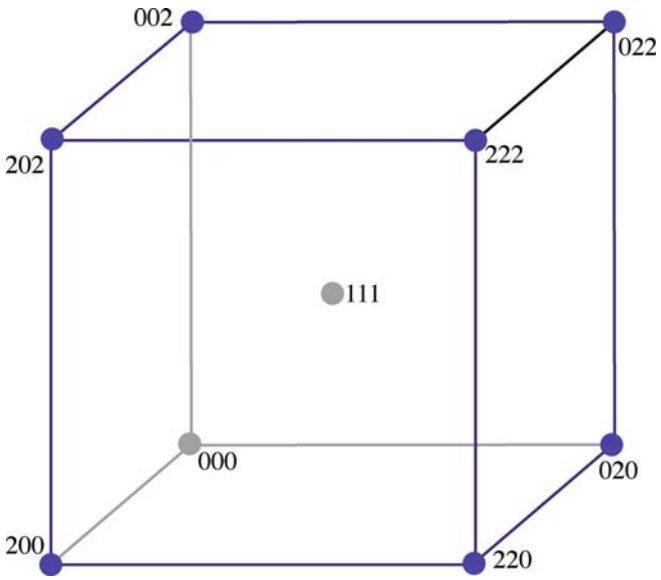


FIGURE 16.2. The reciprocal lattice for the fcc crystal structure. The lattice points that correspond to systematic absences have been removed, so the actual arrangement of points is a bcc lattice.

- We use the three-index notation to derive the structure-factor rules.
- We use the four-index Miller-Bravais notation to index the lattice planes and thus the DPs.

For the hcp structure, we only have to include two atoms in the unit cell. We can view this cell as a simple rhombohedral cell with a two-atom basis. The coordinates of the atoms are

$$(x, y, z) = (0, 0, 0), \left(\frac{1}{3}, \frac{2}{3}, \frac{1}{2}\right) \tag{16.10}$$

Substituting these values for $\mathbf{r}_i$ into equation 16.7 gives

$$F = f\left\{1 + e^{2\pi i\left(\frac{h}{3} + \frac{2k}{3} + \frac{l}{2}\right)}\right\} \tag{16.11}$$

We simplify the notation by setting $h/3 + 2k/3 + l/2 = X$; the complication is simply that X may be a fraction. The analysis is quite straightforward if we consider $|F|^2$, which is what we need in the expression for intensities. Then we can rearrange as follows.

$$|F|^2 = f^2\left(1 + e^{2\pi iX}\right)\left(1 + e^{-2\pi iX}\right) = f^2\left(2 + e^{2\pi iX} + e^{-2\pi iX}\right) \tag{16.12}$$

$$|F|^2 = f^2(2 + 2\cos 2\pi X) = f^2\left(4\cos^2 \pi X\right) \tag{16.13}$$

Now we can write down the rules for hcp which depend mainly on whether or not $h + 2k$ is a multiple of 3.

- $|F|^2 = 0$ if $h + 2k = 3m$ and l is odd,
- $|F|^2 = 4f^2$ if h $+ 2k = 3m$ and l is even,
- $|F|^2 = 3f^2$ if h $+ 2k = 3m \pm 1$ and l is odd,
- $|F|^2 = f^2$ if h $+ 2k = 3m \pm 1$ and l is even.

Thus the $11\bar{2}0$ and $11\bar{2}6$ reflections will be strong but the $11\bar{2}3$ reflection will be absent. Likewise $10\bar{1}0$ and $20\bar{2}0$ are weak but $30\bar{3}0$ is strong. Most importantly, 0001 is absent. You can see that the four-index Miller-Bravais notation takes some time to master. The third index is only included to emphasize the symmetry; if the third index were not included, you might not realize that, e.g., the (110) and $(1\bar{2}0)$ are crystallographically equivalent.

You need to know a few other expressions for this system. If you are working with hcp materials, you *must* have a copy of Frank's (1965) paper on indexing this system.

If the direction [*uvtw*] lies in the plane (*hkil*), then we can show that

$$uh + vk + ti + wl = 0 \tag{16.14}$$

The normal to the plane (*h,k,i,l*) is actually the Cartesian vector [*h,k,i,l*/λ], and likewise the crystallographic direction [*u,v,t,w*] is actually the vector [*u,v,t,*λ*w*] in the Cartesian system. So using the four-index Cartesian vector notation, equation 16.14 can be written as

$$[u, v, t, \lambda w][h, k, i, l/\lambda] = 0 \tag{16.15}$$

In cubic crystals, the direction [*hkl*] is always normal to the plane (*hkl*), but this is not the case for hcp crystals. You can show using some simple geometry that

$$\lambda^2 = \left(\frac{2}{3}\right)\left(\frac{c}{a}\right)^2 \tag{16.16}$$

Thus the Cartesian vector [*HKIL*], which is normal to the plane (*hkil*), is the vector

$$\left(h, k, i, \left(\frac{3}{2}\right)^{\frac{1}{2}}\left(\frac{a}{c}\right)l\right) \tag{16.17}$$

$[11\bar{2}0]$ is normal to the $(11\bar{2}0)$ plane because l is zero but $[01\bar{1}2]$ is not normal to the $(01\bar{1}2)$ plane.

We can now write down an expression for the angle, ϕ, between two planes (*hkil*) and (*defg*). We use equation 16.17 to deduce the normals to the planes, then take the dot product of these two four-index vectors to deduce cos ϕ in the form

$$\cos\phi = \frac{hd + ke + \frac{1}{2}(he + kd) + \frac{3}{4}lg\left(\frac{a}{c}\right)^2}{\left\{h^2 + k^2 + hk + \frac{3}{4}l^2\left(\frac{a}{c}\right)^2\right\}^{\frac{1}{2}}\left\{d^2 + e^2 + de + \frac{3}{4}g^2\left(\frac{a}{c}\right)^2\right\}^{\frac{1}{2}}} \tag{16.18}$$

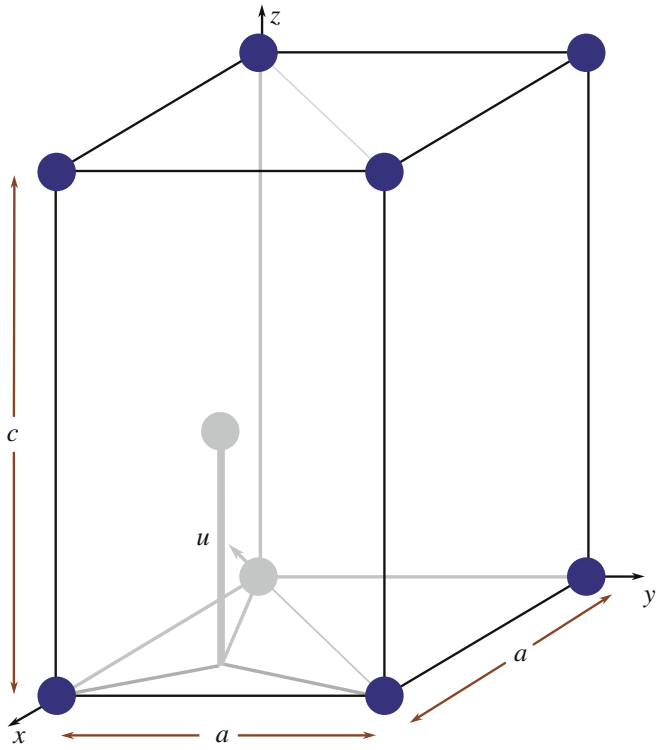


FIGURE 16.3. The hcp unit cell showing the four axes used in the Miller-Bravais indexing system. The three axes in the basal plane, x, y, and u, are all crystallographically equivalent and the z-axis is normal to the basal plane.

The hcp unit cell is shown in Figure 16.3. Remember that there are three crystallographically equivalent axes x, y and u, and that the indices of any plane can be written as $(uviw)$ where $i = (u + v)$.

16.4 EXTENDING FCC AND HCP TO INCLUDE A BASIS

What we did in the previous section was to calculate the reciprocal lattice of a simple-cubic crystal with a basis of four atoms in fcc and two atoms in bcc. We can take this analysis one step further by starting with fcc and adding a basis. This extension both illustrates the technique and deduces structure-factor rules for three important materials.

NaCl, GaAs, and Si: Each of these three crystal structures is an fcc lattice with a basis. In other words, we can separate out the atoms lying on the fcc lattice and those which make up the basis.

NaCl: Let's locate each of the Na atoms on an fcc site; although NaCl is ionic, we'll refer to the ions as atoms since we generally do not take account of the charge on the ion.

We usually say that for every Na atom, there is a Cl atom related to it by the vector [1/2, 0, 0]. However, to emphasize the cubic symmetry, we can choose the alternate basis vector [1/2, 1/2, 1/2]. The phase factor for the Cl atom will be the same as for the Na atom, but with an additional phase of $\pi i(h + k + l)$. Of course, the atomic scattering amplitudes, f, are also different for the two atoms. We can write this expression for F as

$$F = \left\{f_{\mathrm{Na}} + f_{\mathrm{Cl}} e^{\pi i(h+k+l)}\right\}\left\{1 + e^{\pi i(h+k)} + e^{\pi i(h+l)} + e^{\pi i(k+l)}\right\} \qquad (16.19)$$

This again gives rise to some rules:

- $F = 4(f_{\mathrm{Na}} + f_{\mathrm{Cl}})$ if h, k, l are all even,
- $F = 4(f_{\mathrm{Na}} - f_{\mathrm{Cl}})$ if h, k, l are all odd,
- $F = 0$ if h, k, l are mixed.

Clearly, the third condition is the same as for any fcc structure because the factor with four terms is then zero, exactly as we deduced for fcc. You can check this if you imagine that f_{Cl} is zero. Whether the sign in $(f_{\mathrm{Na}} \pm f_{\mathrm{Cl}})$ is positive or negative is the new feature. What this means in practice is that reflections with h, k, l all even will appear much more intense in the DP than those with h, k, l all odd. Look at the values given for f in Table 16.1. LiF, KCl, MgO, NiO, FeO, and ErAs all have the NaCl structure. Since they have different pairs of atomic-scattering amplitudes, the term corresponding to $4(f_{\mathrm{Na}} - f_{\mathrm{Cl}})$ will be different in each case. Reflections with h, k, and l all odd are thus sensitive to the chemistry of the compound and we call them 'chemically sensitive reflections.' We will see further examples in Chapter 29 of how this sensitivity can be used in imaging.

GaAs: You should repeat the above exercise with the Ga located on the fcc lattice and the As related to it by the basis vector [1/4, 1/4, 1/4]. (Crystallographers will immediately note that this puts the As atom in the tetrahedron instead of the octahedron as found in NaCl.) Now the expression for F becomes (see equation 16.9 for F_{fcc})

$$F = \left\{f_{\mathrm{Ga}} + f_{\mathrm{As}} e^{\frac{\pi}{2} i(h+k+l)}\right\} F_{\mathrm{fcc}} \qquad (16.20)$$

So the rules are slightly more complicated:

- $F = 0$ if h, k, l are mixed as always for fcc,
- $F = 4(f_{\mathrm{Ga}} \pm i f_{\mathrm{As}})$ if h, k, l are all odd,
- $F = 4(f_{\mathrm{Ga}} - f_{\mathrm{As}})$ if h, k, l are all even and $h+k+l = 2N$ where N is odd (e.g., the 200 reflection),
- $F = 4(f_{\mathrm{Ga}} + f_{\mathrm{As}})$ if h, k, l are all even and $h+k+l = 2N$ where N is even (e.g., the 400 reflection).

You can appreciate the difference between the 200 reflection and the 400 reflection by drawing a projection onto the (001) plane and applying the physical ideas we discussed in Chapter 11. The case where all three indices are odd is interesting. However, for this case remember that we only see intensities (i.e., $|F|^2$ not F) so $|F|^2$ is $16(f_{\mathrm{Ga}}{}^2 + f_{\mathrm{As}}{}^2)$ and is independent of the

sign initially present. Of course, the structure factor is still different from the others derived here.

Si: Now we can easily extend this analysis to Si, Ge, or diamond. Just replace both f_{Ga} and f_{As} in our results by f_{Si}. The major change is that F is zero when $h+k+l = 2N$ and N is odd. The best known example of this is again the 200 reflection. For Si and Ge, it has $F = 0$, but F is finite for GaAs.

Now for 'hcp.'

Wurtzite: The wurtzite structure is to hcp what GaAs (or zinc blende) is to fcc! It is an important structure because it includes BeO, ZnO, AlN, and GaN, all of which have been widely studied. We can think of it as adding a second hcp lattice displaced by [1/3, 1/3, 1/8] or [0, 0, 3/8] relative to the first. The problem is that we now have a four-atom basis because the second atom in the hcp cell does not lie at a lattice site. This is a good exercise for Section 16.8, if you look ahead.

16.5 APPLYING THE BCC AND FCC ANALYSIS TO SIMPLE CUBIC

Extending bcc to NiAl (B2): For this material, we can easily modify the original treatment of the bcc structure, since now the centering atom is different, so NiAl is simple cubic. If we choose to place the Ni atom at (0, 0, 0) and the Al atoms at [1/2, 1/2, 1/2], then

$$F = \left\{ f_{Ni} + f_{Al} e^{\pi i(h+k+l)} \right\} \qquad (16.21)$$

This leads to two values for F, *neither* of which is zero.

- $F = f_{Ni} + f_{Al}$ if $h + k + l$ is even,
- $F = f_{Ni} - f_{Al}$ if $h + k + l$ is odd.

This would, of course, be the bcc result if f_{Ni} and f_{Al} were the same. The result of this difference is that all of the reflections for NiAl will be present in a DP because F is never zero. This result *is*, of course, exactly what we would expect, because NiAl really is simple cubic. Other materials with this structure include CsCl, CoGa, FeAl, and CuZn. Reflections like (100) are chemically sensitive for NiAl.

The Cu_3Au ($L1_2$) structure: There are many important ordered intermetallics with this structure such as Al_3Li and Fe_3Al. The most important is Ni_3A1 (because of its role in Ni-base superalloys). We can treat Ni_3Al in a similar manner to NiAl. Here the Al atom sits on the (0, 0, 0) site and the three Ni atoms center the faces. The expression for F now becomes

$$F = f_{Al} + f_{Ni}\left\{ e^{\pi i(h+k)} + e^{\pi i(h+l)} + e^{\pi i(k+l)} \right\} \qquad (16.22)$$

The rules for Ni_3Al are

- $F = (f_{A1} + 3f_{Ni})$ if h, k, l are all even or all odd.
- $F = (f_{A1} - f_{Ni})$ if h, k, l are mixed.

Again, all of the possible reciprocal-lattice points for Ni_3Al will give rise to Bragg reflections because the structure is really simple cubic. The mixed *hkl* reflections are now the chemically sensitive reflections. This material is particularly interesting since it can be heat-treated so as to randomize the distribution of the two elements; then each site will be occupied by 75% Ni, 25% Al, and F for mixed *hkl* will be zero. For this reason, reflections with mixed *hkl* are referred to as superlattice reflections (see Section 16.7).

16.6 EXTENDING HCP TO TiAl

The TiAl structure is not as well known as the previous two cases, but illustrates a similar class of materials. We noted in Section 16.4 that the two atoms in the hcp structure are not equivalent. In TiAl, we actually make them chemically distinct too. This means that the rules for hcp will be modified again. Using equation 16.11, we find that

$$F = f_{Ti} + f_{Al} e^{2\pi i\left(\frac{h}{3}+\frac{2k}{3}+\frac{l}{2}\right)} \qquad (16.23)$$

The most important result is that the (0001) reflection is now allowed since $F = f_{Ti} - f_{Al}$. TiAl really does have a primitive hexagonal unit cell.

16.7 SUPERLATTICE REFLECTIONS AND IMAGING

The reciprocal lattices for Ni_3Al and NiAl are shown in Figure 16.4; the small circles indicate the chemically sensitive reciprocal-lattice points. The terminology which has developed calls the chemically sensitive reflections *superlattice reflections*; the idea is that the fcc lattice is viewed as the lattice and the chemically sensitive reflections then lie on a lattice with a finer scale in reciprocal space. The chemically sensitive superlattic reflections are all forbidden in the disordered fcc structure.

Superlattice reflections are those present because the material is ordered such that the actual real-space unit cell is larger and thus the reciprocal-space cell is smaller.

For many years, these superlattice reflections were regarded as a special feature in some unusual materials. However, ordered materials, particularly the ordered intermetallics which we mentioned in Section 16.4, are finding increased uses. We will illustrate the wide variety of superlattice effects by selecting some examples.

Figure 16.5 shows an image from Cu_3Au, the archetypal A_3B ordered fcc structure. The crystal has been

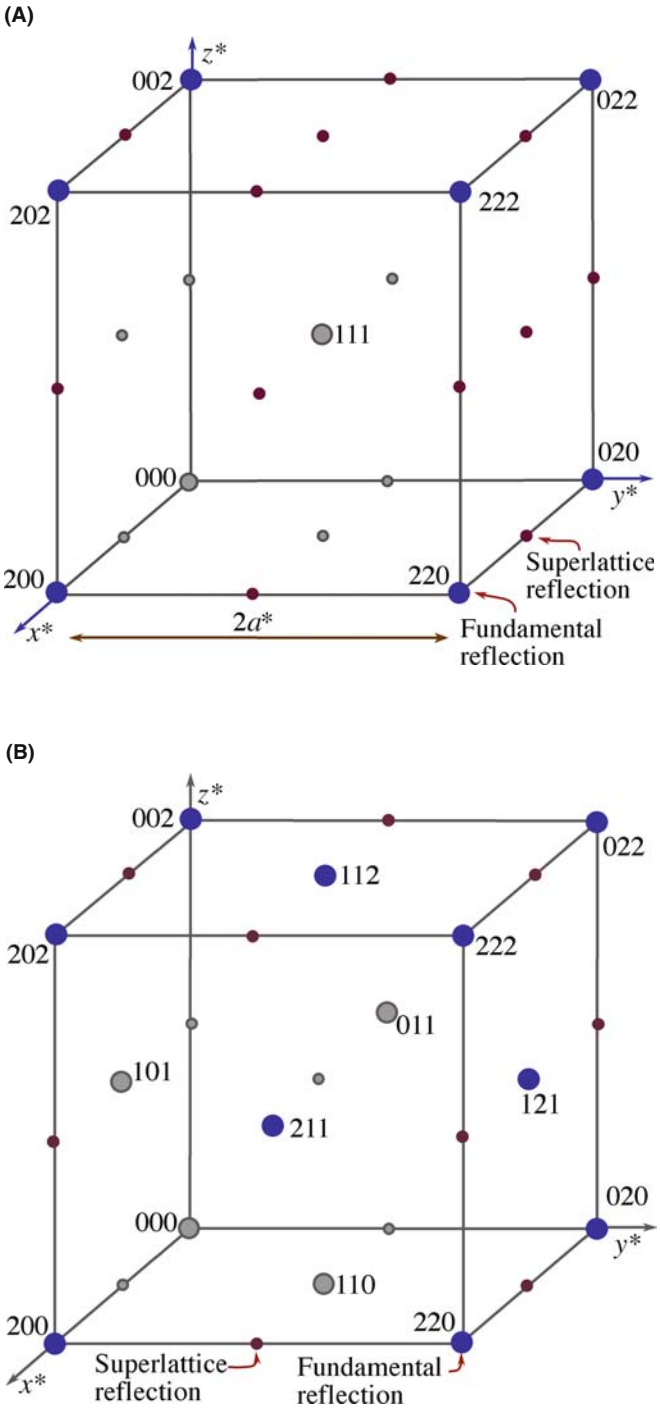


FIGURE 16.4. The reciprocal lattices for (A) the Ni_3Al and (B) the NiAl structures. In (A) Ni_3Al is fcc, so the fcc-forbidden reflections (h, k, l mixed even and odd) are allowed and become chemically sensitive (superlattice) reflections. In (B) NiAl is bcc, so the bcc-forbidden reflections (if $h + k + l$ odd) are now allowed superlattice reflections.

irradiated with ions so that small regions known as cascades have been damaged just enough that the Cu and Au have been mixed up, i.e., the ordering has been destroyed locally (Jenkins et al.). The DF image has been formed using the 110 reflection, which we know is a superlattice reflection. By destroying the ordering,

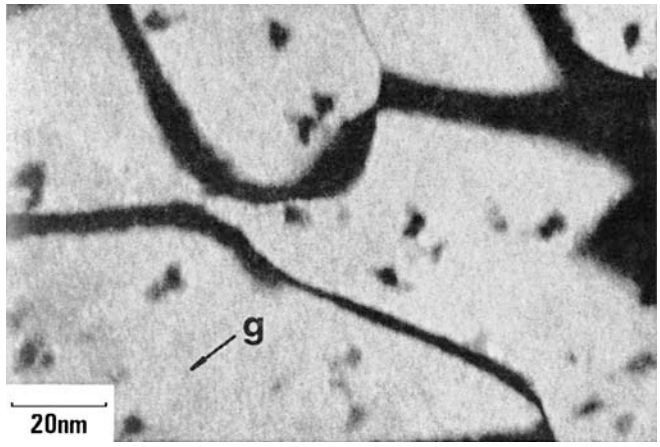


FIGURE 16.5. DF image from a chemically sensitive 110 reflection showing bright ordered domains in Cu_3Au. The dark areas in the bright domains are regions of local disorder induced by ion beam damage.

we 'destroy' the superlattice reflection for the disordered region, so the disordered region appears black when the ordered matrix appears bright. Thus, we can 'see' the disordered region, measure its size, etc., even though it is not diffracting electrons. The dark bands between the domains are inclined anti-phase domain boundaries (APBs), a specific kind of planar defect which we'll examine in more detail in Section 24.6.

Figure 16.6A and B shows a 002 DF image and the corresponding DP from a GaAs/$Al_xGa_{1-x}As$ quantum-well structure. The $Al_xGa_{1-x}As$ appears lighter than the GaAs because the 002 reflection is a superlattice reflection; remember, it would be forbidden for GaAs if f_{Ga} and f_{As} were equal. The reason the Al_xGa_{1-x} As appears lighter is that we have replaced a fraction x of the Ga atoms with the lighter Al atoms, thus increasing the difference $f_{III} - f_V$. Clearly, this is a classic example of chemically sensitive reflections. At this point we should

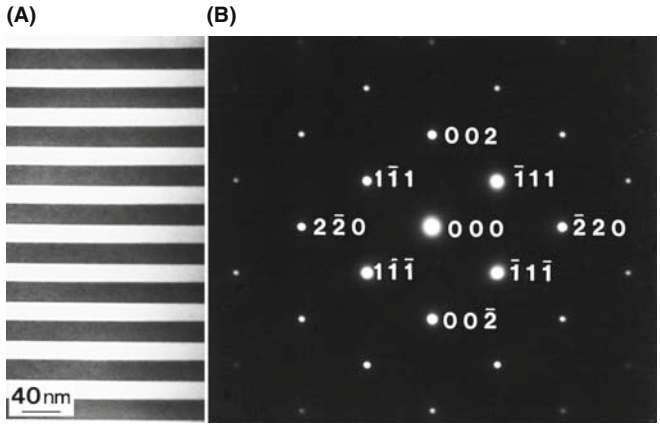


FIGURE 16.6. (A) DF image from a 002 chemically sensitive reflection in GaAs/$Al_xGa_{1-x}As$. The $Al_xGa_{1-x}As$ is the lighter region because Al has replaced Ga in the GaAs (darker regions). (B) DP showing the less intense 002 and other superlattice reflections.

remind you about intensities in images and DPs. The discussion we have just gone through assumes that we have a thin specimen, so that we are within the first thickness zone (i.e., the specimen is thinner than one extinction distance). In other words, be wary of trying to be quantitative about these intensities since superlattice beams are also dynamically diffracted.

Our third example is from a ceramic, vanadium carbide. The structure of VC is the same as for NaCl so we already have the rules. However, this carbide is usually non-stoichiometric, having the composition V_xC_y, where $x > y$. The two images and DPs shown in Figure 16.7 were taken from well-ordered V_6C_5 and V_8C_7 where 1/6 and 1/8 of the carbon sites are not occupied by C: we say these sites are occupied by vacancies and the vacancies have formed ordered arrays. Clearly since we only have four atoms of each element in the unit cell, the vacancies must be distributed over more than one cell so the new lattice parameter must be greater than the lattice parameter (a) of the VC fcc lattice. So, we expect to see extra spots that are closer to the origin than (001). This is the case in both patterns shown here. The ordering actually destroys the cubic symmetry, so we have several orientations of the ordered carbides that are related to one another by the way they break the symmetry. By forming DF images, we can identify which region of the specimen corresponds to which variant (Dodsworth et al.).

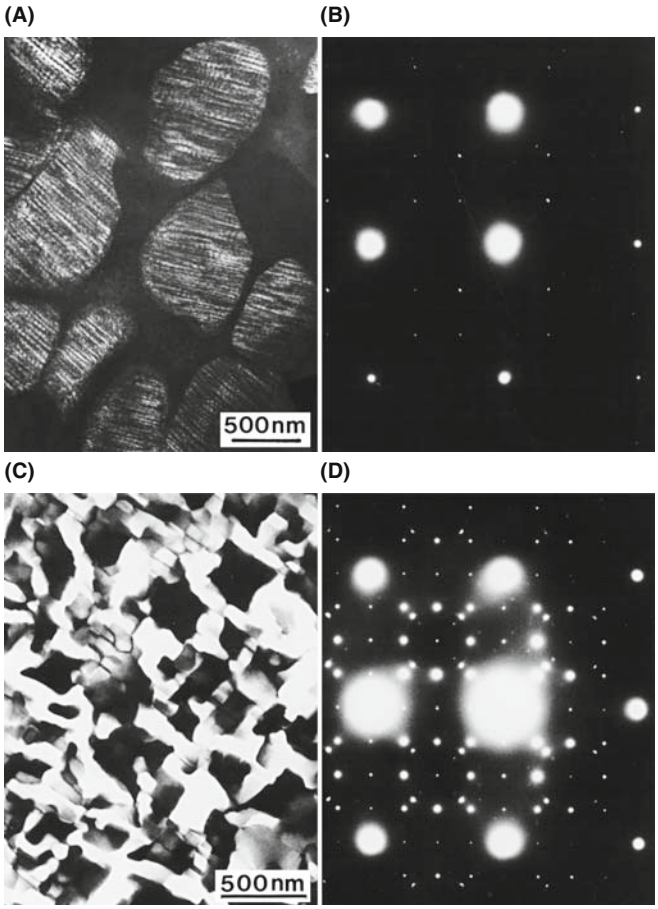


FIGURE 16.7. (A) DF image of ordered V_6C_5 and (B) accompanying DP. (C) DF image of V_8C_7 and (D) DP. In both carbides the ordering is due to vacancies on the C sublattice.

16.8 DIFFRACTION FROM LONG-PERIOD SUPERLATTICES

In the previous section, the atoms or vacancies in the different structures essentially arranged themselves to increase the lattice parameter and therefore give rise to superlattice reflections. In this section, we will discuss several examples where either we (or nature) have arranged the materials to give much larger superlattices. We will begin by considering the image and DP shown in Figure 16.8, which are from an artificial GaAs/Al_xGa_{1-x}As superlattice. The superlattice is created chemically by changing from four layers of GaAs to four of (Al_xGa_{1-x})As. So we see a series of three closely spaced extra spots in the DP which correspond to the new long lattice parameter in real space.

Another example is shown in Figure 16.9. This is a very long period (~10 nm) artificial superlattice of alternating layers of Si and Mo. The extra reflections are very close and are not as useful as they were in Figure 16.6, but they do allow us to check the periodicity of the real-space structure very easily and quickly and without needing to use HRTEM (Chapter 28). This can be useful, particularly for artificially grown superlattices, since the superlattice periodicity is 'internally calibrated' in the DP by the lattice spacing of the material. (Remember that, for comparison, the magnification of a TEM image is usually subject to a ±10% uncertainty.)

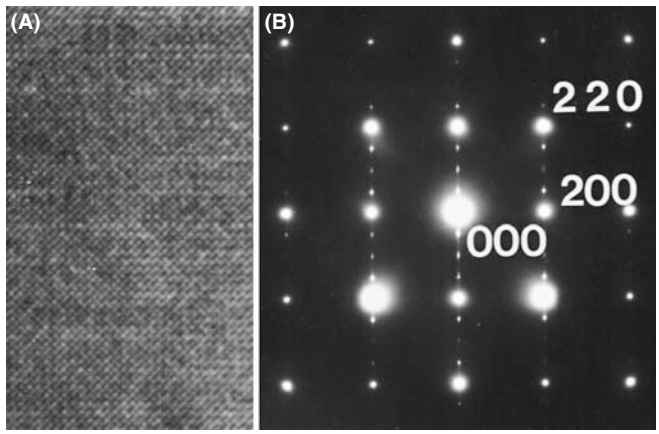


FIGURE 16.8. (A) GaAs/Al_xGa_{1-x}As structure in which order is created by alternating four layers of GaAs and four of (Al_xGa_{1-x})As. (B) DP showing three superlattice spots between the fundamental reflections in the 020 direction.

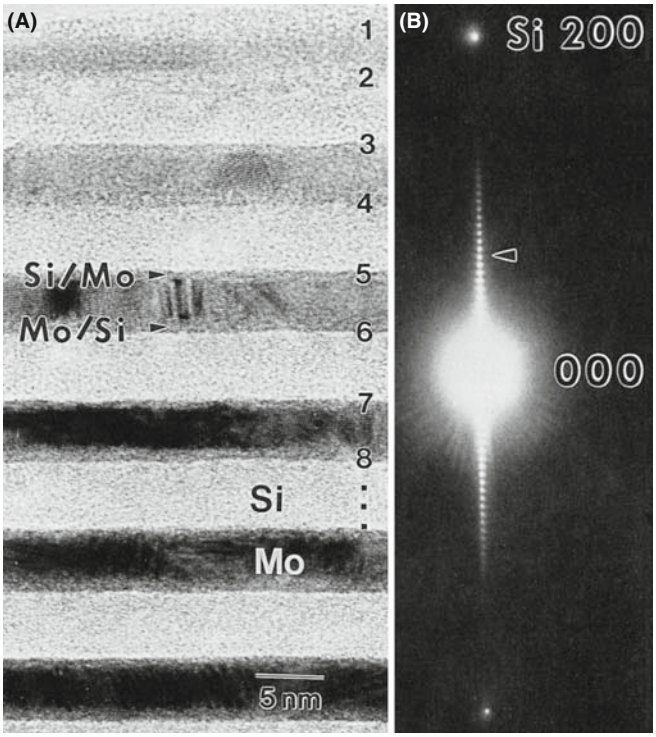


FIGURE 16.9. (A) Artificial superlattice of Si and Mo layers ~5 nm thick. (B) Expanded DP around 000 showing many superlattice spots (arrowed). The large spacing of the superlattice in real space results in very small spacing of the superlattice reflections in the DP in reciprocal space. Compare with Figure 16.8.

16.9 FORBIDDEN REFLECTIONS

We mentioned in Section 16.2 that certain reflections are always forbidden for some structures because they have $F = 0$. They are known as kinematically forbidden reflections, because such reflections can sometimes actually be present due to dynamical scattering events. This process is illustrated in Figure 16.10. The DP is the [011] pattern in Si so that the 200 reflection should be absent according to Section 16.4. The reason it is usually present is that, since we are oriented at the zone axis, the $11\bar{1}$ beam, which has $F \neq 0$, acts like a new incident beam and is rediffracted by the $(1\bar{1}1)$ plane. The result is that we appear to excite the 200 reflection since

$$(11\bar{1}) + (1\bar{1}1) = (200) \qquad (16.24)$$

From this example, you can appreciate the use of the phrase 'kinematically forbidden.'

16.10 USING THE INTERNATIONAL TABLES

As long as you work with fcc or bcc metals or the other special structures listed here, you can use the simple rules derived in this chapter. Once you venture further,

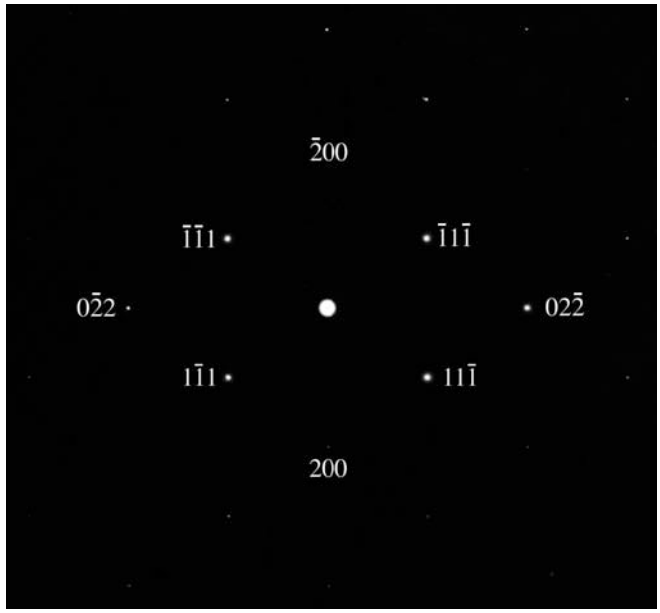


FIGURE 16.10. The [011] DP from Si. The 200 reflection is forbidden, but it is present because the allowed $11\bar{1}$ diffracted beam acts like a new incident beam and is rediffracted by the $(1\bar{1}1)$ plane. The sum of the two allowed reflections $(11\bar{1}) + (1\bar{1}1)$ results in a 200 reflection, which is so weak you may not see it.

you should quickly become familiar with the International Tables for Crystallography (Hahn), in particular with the introductory booklet. You must know the crystal structure of your material; if not, you will in principle be able to determine it after studying Chapter 21. If, for example, you were working with α-Al_2O_3, you

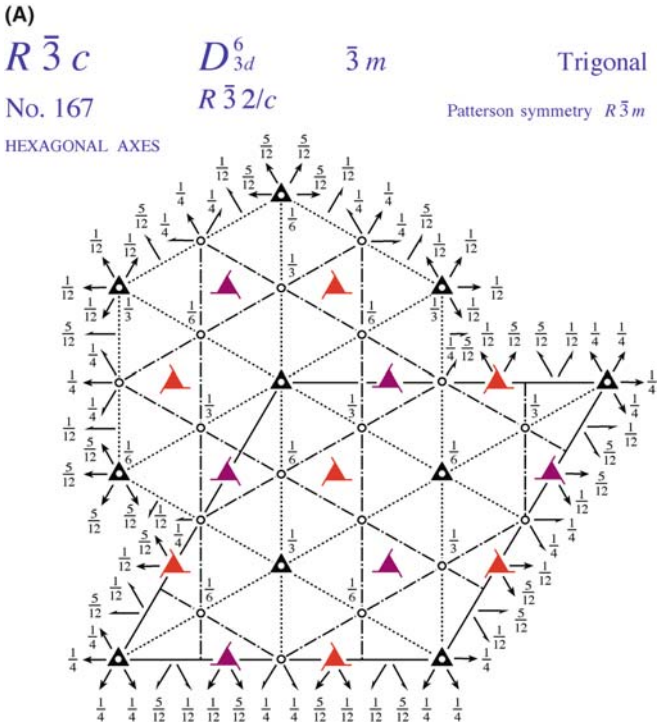


FIGURE 16.11. (continued).

would know that the space group is R$\bar{3}$c or No. 167. Looking this up in the International Tables, you would find the information shown in Figure 16.11A. In this case, you'd have to decide whether you want to use rhombohedral axes or hexagonal axes; you'll notice that there are three times as many atoms in the hexagonal cell. The tables in Figure 16.11B tell you which reflections are allowed, although you can work out or look up the values of F if you want them. You know the chemical formula of your material, but you still need to know which sites are occupied. Look up the positions from X-ray diffraction data. The paper by Lee and Lagerlof summarizes the analysis for this particular example.

(B)

Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates			Reflection conditions
			(0,0,0)+	$(\frac{2}{3},\frac{1}{3},\frac{1}{3})+$	$(\frac{1}{3},\frac{2}{3},\frac{2}{3})+$	General:
36	f	1	(1) x,y,z	(2) $\bar{y},x-y,z$	(3) $\bar{x}+y,\bar{x},z$	$hkil$: $-h+k+l=3n$
			(4) $y,x,\bar{z}+\frac{1}{2}$	(5) $x-y,\bar{y},\bar{z}+\frac{1}{2}$	(6) $\bar{x},\bar{x}+y,\bar{z}+\frac{1}{2}$	$hki0$: $-h+k=3n$
			(7) $\bar{x},\bar{y},\bar{z}$	(8) $y,\bar{x}+y,\bar{z}$	(9) $x-y,x,\bar{z}$	$hh\bar{2}hl$: $l=3n$
			(10) $\bar{y},\bar{x},z+\frac{1}{2}$	(11) $\bar{x}+y,y,z+\frac{1}{2}$	(12) $x,x-y,z+\frac{1}{2}$	$h\bar{h}0l$: $h+l=3n$, $l=2n$
						$000l$: $l=6n$
						$h\bar{h}00$: $h=3n$
						Special: as above, plus no extra conditions
18	e	.2	$x,0,\frac{1}{4}$	$0,x,\frac{1}{4}$	$\bar{x},\bar{x},\frac{1}{4}$	$\bar{x},0,\frac{3}{4}$ $0,\bar{x},\frac{3}{4}$ $x,x,\frac{3}{4}$
18	d	$\bar{1}$	$\frac{1}{2},0,0$	$0,\frac{1}{2},0$	$\frac{1}{2},\frac{1}{2},0$	$0,\frac{1}{2},\frac{1}{2}$ $\frac{1}{2},0,\frac{1}{2}$ $\frac{1}{2},\frac{1}{2},\frac{1}{2}$
12	c	3.	$0,0,z$	$0,0,\bar{z}+\frac{1}{2}$	$0,0,\bar{z}$	$0,0,z+\frac{1}{2}$
6	b	$\bar{3}$.	$0,0,0$	$0,0,\frac{1}{2}$		
6	a	32	$0,0,\frac{1}{4}$	$0,0,\frac{3}{4}$		

Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates			Reflection conditions
12	f	1	(1) x,y,z	(2) z,x,y	(3) y,z,x	General:
			(4) $\bar{y}+\frac{1}{2},\bar{x}+\frac{1}{2},\bar{z}+\frac{1}{2}$	(5) $\bar{x}+\frac{1}{2},\bar{z}+\frac{1}{2},\bar{y}+\frac{1}{2}$	(6) $\bar{z}+\frac{1}{2},\bar{y}+\frac{1}{2},\bar{x}+\frac{1}{2}$	hhl : $l=2n$
			(7) $\bar{x},\bar{y},\bar{z}$	(8) $\bar{z},\bar{x},\bar{y}$	(9) $\bar{y},\bar{z},\bar{x}$	hhh : $h=2n$
			(10) $y+\frac{1}{2},x+\frac{1}{2},z+\frac{1}{2}$	(11) $x+\frac{1}{2},z+\frac{1}{2},y+\frac{1}{2}$	(12) $z+\frac{1}{2},y+\frac{1}{2},x+\frac{1}{2}$	
						Special: as above, plus no extra conditions
6	e	.2	$x,\bar{x}+\frac{1}{2},\frac{1}{4}$	$\frac{1}{4},x,\bar{x}+\frac{1}{2}$	$\bar{x}+\frac{1}{2},\frac{1}{4},x$	
			$\bar{x},x+\frac{1}{2},\frac{3}{4}$	$\frac{3}{4},\bar{x},x+\frac{1}{2}$	$x+\frac{1}{2},\frac{3}{4},\bar{x}$	
6	d	$\bar{1}$	$\frac{1}{2},0,0$	$0,\frac{1}{2},0$	$0,0,\frac{1}{2}$ $\frac{1}{2},0,\frac{1}{2}$ $0,\frac{1}{2},\frac{1}{2}$ $\frac{1}{2},\frac{1}{2},0$	hkl : $h+k+l=2n$
4	c	3.	x,x,x	$\bar{x}+\frac{1}{2},\bar{x}+\frac{1}{2},\bar{x}+\frac{1}{2}$	$\bar{x},\bar{x},\bar{x}$ $x+\frac{1}{2},x+\frac{1}{2},x+\frac{1}{2}$	hkl : $h+k+l=2n$
2	b	$\bar{3}$.	$0,0,0$	$\frac{1}{2},\frac{1}{2},\frac{1}{2}$		hkl : $h+k+l=2n$
2	a	32	$\frac{1}{4},\frac{1}{4},\frac{1}{4}$	$\frac{3}{4},\frac{3}{4},\frac{3}{4}$		hkl : $h+k+l=2n$

FIGURE 16.11. (A) Symmetry information, as given in the International Tables for trigonal α-Al_2O_3, with space group $R\bar{3}c$ (No. 167), showing the two possible unit cells based on the rhombohedral and hexagonal cells. The symmetry elements at specific lattice points are also indicated. (B) The atomic positions for the two choices of unit cells in (A).

That was the traditional approach. Now, you should have access to EMS or Crystal Kit. Alternatively, use jEMS over the WWW (Section 1.6). In all these software packages you can just type in your space group or pull down a menu to find the structure-factor information.

CHAPTER SUMMARY

When we introduced the primitive lattice at the beginning of this chapter, we only considered the lattice sites which actually define the unit cell. If there are other lattice points, these would give us the Bravais lattices. We will conclude by summarizing some of the selection rules for the different structures in Table 16.2

TABLE 16.2 Examples of Selection Rules for Several Crystal Structures. *F* is the Structure Factor

Crystal type	Reflection present for	*F*	No. of lattice points per cell
Primitive	Any *h*, *k*, *l*	*f*	1
Body centered	$(h + k + l) = 2n$	2*f*	2
Face centered including GaAs and NaCl	*h*, *k* and *l* all odd or all even	4*f*	4
Diamond	As fcc but if all even and $h + k + l \neq 4n$ then absent, anyway		
Base centered	*h*, *k* and *l* all odd or all even	2*f*	2
			Example reflection
Hexagonal close-packed	$h + 2k = 3n$ with *l* odd	0	0001
	$h + 2k = 3n$ with *l* even	2*f*	0002
	$h + 2k = 3n \pm 1$ with *l* odd	*f*3	$01\bar{1}1$
	$h + 2k = 3n \pm 1$ with *l* even	*f*	$01\bar{1}0$

In practice, it will become important that you simply *know* some of the DPs for your material. You can, however, look up these patterns in some of the textbooks listed in Chapter 1, but the best sources are the texts by Andrews et al. and Edington; we reproduce some of them in Figures 18.17–18.19. Alternatively, software (e.g., jEMS) available on the WWW (Section 1.5) will print out standard spot patterns of most important crystal structures. When you're sitting at the TEM, you don't have time to index a pattern from first principles and then decide whether or not you are at a pole that contains the reflection you want to use. To do this you'll have to be able to index the DPs and determine the beam direction, which we'll describe in detail in Chapter 18.

REFERENCES

Examples of indexed DPs are given in the texts by Edington and Andrews et al. As usual, we recommend your checking the books by Kelly and Groves, Cullity.

DPs and Crystal Data

Andrews, KW, Dyson, DJ and Keown, SR (1971) *Interpretation of Electron Diffraction Patterns* 2nd Ed. Plenum Press New York. Stereographic projections, angles, spacings, and much more.

Massalski, T, Okamoto, H, Subramanian, PR and Kacprzak, L (Eds.) (1990) *Binary Alloy Phase Diagrams* 2nd Ed. ASM International Materials Park OH. Appendix A1 gives a complete list of the Pearson Symbols ($L1_2$, etc.) with their space group and Strukturbericht designation.

Misell, DL and Brown, EB (1987) *Electron Diffraction: An Introduction for Biologists*, Volume 12 of the series *Practical Methods in Electron Microscopy* Ed. AM Glauert Elsevier New York. Materials science students should not be put off by the title: this is an invaluable practical guide to indexing DPs and more.

SOME CRYSTALLOGRAPHY

Dodsworth, J, Kohlstedt, DL and Carter, CB 1983 *Grain Boundaries in Transition Metal Carbides* Adv. Ceram. **6** 102–109.

Frank, FC 1965 *On Miller-Bravais Indices and Four-Dimensional Vectors* Acta Cryst. **18** 862–866. A favorite.

Hahn, T 1988 *International Tables for Crystallography. Brief teaching edition of volume A, space-group symmetry*, Kluwer Academic Publishers Dordrecht Netherlands.

Ibers, JA 1957 *New Atomic Form Factors for Beryllium and Boron* Acta Cryst. **10** 86. One page.

Jenkins, ML, Katerbau, K-H and Wilkens, M 1976 *Transmission Electron Microscopy Studies of Displacement Cascades in Cu_3Au* Phil. Mag. **34** 1141–1153.

Lee, WE and Lagerlof, KPD 1985 *Structural and Electron Diffraction Data for Sapphire (α-Al_2O_3)* J. Electron Microsc. Tech. **2** 247–258.

THE COMPANION TEXT

You'll find jEMS useful for simulating DPs.

SELF-ASSESSMENT QUESTIONS

Q16.1 What is the key assumption that is made in the amplitude calculation in this chapter?
Q16.2 Give one reason why DP selection rules are sometimes broken.
Q16.3 In a bcc Fe specimen, will any kinematically forbidden reflections occur in the DP?
Q16.4 In a bcc specimen, is the 110 reflection allowed?
Q16.5 In fcc, will you see the 111 and/or 100 reflections?
Q16.6 In NaCl, which *hkl* reflections are chemically sensitive?
Q16.7 In GaAs, which reflections are forbidden?
Q16.8 When determining if a reflection is forbidden or allowed, need we only consider the structure factor, *F*?
Q16.9 If a material is very well ordered, what does it mean to have formed a superlattice in the DP?
Q16.10 List some advantages in determining superlattice spacings from a DP rather than an image.
Q16.11 Why is the 200 reflection of Si seen in Figure 16.9?
Q16.12 Why are hcp DPs more difficult to index than cubic DPs?
Q16.13 Explain the concept of structure factor in words?
Q16.14 What two things should you keep in mind when looking up tabulated values of the atomic scattering factor?
Q16.15 In Figure 16.8, what is the true periodicity in the [020] direction?
Q16.16 What do we mean by the term 'chemically sensitive reflection'? Give two examples.
Q16.17 When is it particularly important to use the International Tables?
Q16.18 Why does the uncertainty in *f* values not usually affect our analysis of DPs?
Q16.19 Why are *f* values not precisely known?
Q16.20 Give the general equation for the structure factor.
Q16.21 Write down the structure factor rules for a bcc crystal and describe the reciprocal lattice.
Q16.22 Write down the structure factor rules for an fcc crystal and describe the reciprocal lattice.
Q16.23 Write down the structure factor rules for an hcp crystal. Can you describe the reciprocal lattice?
Q16.24 What is the Cartesian vector which is normal to the plane (*hkil*)?
Q16.25 How many atoms are needed (minimum number) to calculate the structure factor for simple cubic, fcc, bcc and hcp structures?
Q16.26 How do the structures of NaCl and NiAl differ?
Q16.27 In Figure 16.5, why can we see the ordered domains but not the disordered domains?
Q16.28 In Figure 16.6, why is the $Al_xGa_{1-x}As$ region brighter than the GaAs region?
Q16.29 In Figure 16.7, why is the symmetry of the two DPs changed?
Q16.30 In Figure 16.8, why are there rows of fainter spots between bright spots?
Q16.31 In Figure 16.9, why can we see superlattice reflections even though half the material is amorphous?
Q16.32 If the crystal is ionic, do we use *f* for the atom or for the ion?
Q16.33 What is a basis?
Q16.34 Describe three equivalent descriptions of systematic absences in crystals.
Q16.35 How can you easily distinguish GaAs from Si given two (001) DPs?
Q16.36 Why didn't we see superlattice spots in the DP in Figure 16.6B?

TEXT-SPECIFIC QUESTIONS

T16.1 Starting with the fcc case, deduce the structure factor rules for NaCl.
T16.2 Deduce the structure factor rules for NiAl.

T16.3 Consider Figure 16.6. Draw a schematic of the DP. Index all the remaining reflections. Look at the 002 row of reflections; indicate the location of the extra reflections which should be present due to the superlattice. Justify your answer.

T16.4 Sketch the DP shown in Figure 16.10 with a box as large as that shown in this figure. Add all the other reflections you would expect to see. Justify your answer.

T16.5 Look at Figure 16.9. Explain all the contrast features you see in this image (after you've read Part 3).

T16.6 Using Figure 16.11, explain which positions are occupied in both the rhombohedral and hexagonal cells for Al_2O_3. Neglecting the possibility of dynamical scattering, explain why you will, or will not, see the 0001, $1\bar{1}00$ and/or $11\bar{2}0$ reflections. Will you see the 112, 111 and/or 100 reflections?

T16.7 A simple orthorhombic lattice has the following lattice constants: $a = 0.30$ nm; $b = 0.40$ nm; $c = 0.50$ nm. Draw the DP when a crystal is observed along the [100], [010] and [001] directions, respectively. The DPs should be drawn in the same scale. *NB*: There are no 'forbidden' reflections in this structure. (Courtesy Anders Thølen.)

T16.8 The alloy $CuAu_3$ has an ordered structure. What does the DP look like when the crystal is viewed along the [110] direction? (Courtesy Anders Thølen.)

T16.9 Electron diffraction from a powder material gives a ring pattern. When the incoming beam is tilted an angle θ relative to the optical axis the ring pattern moves. Calculate the angle θ the incoming beam has been tilted when the center of the DP is moved so it falls on the {222} ring of the original pattern. The material is Pt (fcc) with a lattice constant 0.392 nm. The acceleration voltage is 100 kV. (Courtesy Anders Thølen.)

T16.10 What reflections would you expect to observe in an electron DP(s) from a c-centered monoclinic cell? (Courtesy Lucille Giannuzzi.)

T16.11 The unit cell of a crystal of $Ti_2Nb_{10}O_{29}$ is orthorhombic (space group *Amma*) with $a = 28.5$ Å, $b = 3.8$ Å, $c = 20.5$ Å. The DP of 1a is from one of the principal axis patterns. Give the Miller indices for this pattern. The pattern 1b is from a slightly tilted crystal. Find the tilt axis, tilt angle and the excitation error from the spot circled. A further tilt of the crystal gave the pattern of 1c printed to the same scale. What is the Miller index of the spot arrowed? ($\lambda = 0.037$ Å) (See J. Appl. Phys. **42** 5891.) (Courtesy John Spence.)

T16.12 $LaMnO_3$ has a perovskite structure shown below. Calculate the structure factors for {110}, {100}, {200} and {220} in terms of the atomic-scattering factors (such as f_{La}, f_{Mn}, f_O). (A) Use the size of the spot to represent the intensity, sketch the [001] DP of this structure. (B) Index the first eight diffraction spots. (Courtesy ZL Wang.)

T16.13 A ternary phase has the following atoms in a cubic unit cell: A at (0,0,0), B at (1/2, 1/2,0) and C at (1/2,0,1/2), (0,1/2,1/2). Calculate the following structure factors in terms of f_A, f_B and f_C: (a) (001) and (010) and (b) (100) and (200). (Courtesy ZL Wang.)

T16.14 A ternary alloy (A_2BC) has the fcc structure shown in (a) below at elevated temperatures. As the temperature is lowered, all the C atoms go to the corners of the lattice while the A and B atoms randomly occupy the face centers, as shown in (b). At still lower temperatures, the B atoms occupy the (1/2,1/2,0) sites, indicated in (c). Sketch the [100] DPs for each of these structures, plotting the positions and relative intensities of the different reflections. (Courtesy ZL Wang.)

T16.15 Consider a face-centered orthorhombic lattice. By means of the definition of the reciprocal lattice, obtain expressions for the three basic vectors of the reciprocal lattice of this real-space lattice, in terms of the vectors **a, b** and **c**. (Courtesy ZL Wang.)

T16.16 A phase in the Fe-C-Al system as the following structure: Al at (0,0,0), Fe at (1/2,1/2,0), (1/2,0,1/2), (0,1/2,1/2) and C at (1/2, 1/2,1/2). The atomic-scattering factors for the three elements are shown in the plot below. (a) Derive the expression for the structure factor in terms of f_{Al}, f_{Fe} and f_C. (b) Calculate the relative intensity ratios for the following reflection in the DP: (i) $I_{(001)}/I_{(002)}$ and (ii) $I_{(011)}/I_{(002)}$. (c) Sketch and index the [100] DP of this phase. (Courtesy ZL Wang.)

17

Diffraction from Small Volumes

CHAPTER PREVIEW

A very important concept in TEM is that we only ever diffract from small volumes. These volumes are now called nanoparticles, nanograins, nanobelts, etc. By definition, no TEM specimen is infinite in all directions and all defects are small. Of course, the beam is also never infinitely wide! This chapter therefore discusses how the size of what we are examining influences the appearance of the DP. Although we will discuss many different aspects of diffraction, there are three important ideas that underlie all this discussion

- We are diffracting from small volumes.
- We are diffracting from crystals.
- We need to index the DPs we see and relate the patterns to the image.

The fact that it is possible to obtain diffraction from several planes in a zone at the same time is due to the effect of the specimen shape on the diffracted-intensity distribution. The diffraction spot is only a mathematical point if the specimen is perfect and infinite in all directions. For example, a TEM specimen is effectively infinite ($\sim$3 mm) relative to the unit-cell dimensions in the plane of the specimen, but very thin ($< 0.5\,\mu$m) parallel to the electron beam. This means that the diffracted intensity can be represented in the reciprocal lattice as a relrod (see Chapter 12) stretched parallel to the electron beam in reciprocal space, rather than as a point, and the relrod does have a width. Therefore, over a range of angles, the Ewald sphere will still intercept the relrod and diffracted intensity will still be generated. This is equivalent to saying that the Laue condition is relaxed in one dimension in the TEM owing to the specimen shape. For this reason, accurate structural analysis of unknown specimens is very difficult in conventional TEM diffraction, and X-rays are usually the most accurate method for structure determination if your specimen is large. However, we will reconsider this statement in Chapter 21.

17.1 INTRODUCTION

In Chapter 12, we stated that each point in the reciprocal lattice can actually be associated with a rod. This construction allowed us to discuss the geometry of DPs, taking account of the experimental fact that we see spots in the DP even when **s** is not exactly zero. In fact, without this construction, there is no reason to discuss **s**. Now we are going to show quantitatively why we have rods. As we suggested earlier, the reason is that we have a thin specimen: a small thickness in real space gives a large length in reciprocal space. This concept is valid in all directions, not just parallel to the electron beam. Hence, we call this the 'shape effect.' The intensity in the diffracted beam is strongest for a given thickness when $\mathbf{K} = \mathbf{g}$, but we still have intensity when **K** is not exactly equal to **g**, or when

$$\mathbf{K} = \mathbf{g} + \mathbf{s} \tag{17.1}$$

Then we can write, from equation 13.48

$$\left|\phi_{\mathbf{g}}\right|^2 = \left(\frac{\pi t}{\xi_{\mathbf{g}}}\right)^2 \cdot \frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} \tag{17.2}$$

We model the specimen as a thin rectangular slab as shown in Figure 17.1. To keep the math simple, we will assume that we have a rectangular unit cell with sides a, b, c and that there are N_x cells in the x direction, N_y in

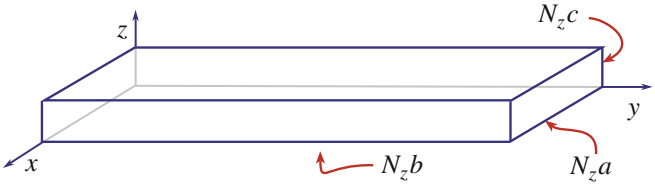


FIGURE 17.1. An idealized thin-foil specimen modeled as a rectangular slab made up of rectangular unit cells of sides *a, b, c*. There are N_x cells in the *x* direction, N_y in the *y* direction and N_z in the *z* direction.

the *y* direction and N_z in the *z* direction. All that we have to do to determine the total diffracted amplitude is to add the amplitudes from each cell, allowing for the phase factor that is present, because the cells are displaced from one another. Each cell has the same structure factor *F*.

We can do the addition of amplitudes in two ways. The first way is to do the summation. In the second, we will show how the same result follows if you start with the integral expression for $\phi_{\mathbf{g}}$. These expressions lead to the important idea of a relrod and subsidiary maxima; in DPs we can see the effects of the relrods, but you may never see the subsidiary maxima.

What we are going to do is derive equations for the shape of the relrods which were introduced in Section 12.5 and which we used in Chapter 13 to explain why we 'see' spots in the DP even when $\mathbf{s} \neq 0$. This whole approach gives us a pictorial aid to understanding diffraction from small volumes. After developing the theory for the simple case, we will go on to discuss the complications introduced because we look at real materials, and specimens of real materials are usually not flat platelets.

17.1.A The Summation Approach

This approach starts with expressing the total amplitude, *A*, of the diffracted beam as the sum of contributions from all the individual cells in a parallel-sided specimen. (Note: this is a kinematical approach and ignores dynamical scattering.)

$$A = F \sum_{n_x} e^{\mathrm{i}\,2\pi n_x \mathbf{K}.\mathbf{a}} \sum_{n_y} e^{\mathrm{i}\,2\pi n_y \mathbf{K}.\mathbf{b}} \sum_{n_z} e^{\mathrm{i}\,2\pi n_z \mathbf{K}.\mathbf{c}} \tag{17.3}$$

Here n_x, n_y and n_z have their usual meanings and all are integers; we have $N_x \times N_y \times N_z$ cells in the specimen. As shown in Figure 17.1, we will let n_x vary from 0 to N_x–1 and similarly with n_y and n_z. The location of each unit cell is then defined by the vector $\mathbf{r}_n$

$$\mathbf{r}_n = n_x\mathbf{a} + n_y\mathbf{b} + n_z\mathbf{c} \tag{17.4}$$

To simplify the first summation we set *X* equal to $e^{\mathrm{i}2\pi\mathbf{K}.\mathbf{a}}$. Then each separate summation term is a geometric series, so we can sum the n_x series as

$$\begin{gathered} S = \sum_{n_x=0}^{n_x=N_x-1} X^n = X^0 + X^1 + \ldots X^{N_x-1} \\ SX = X^1 + X^2 + \ldots X^{N_x} = X^{N_x} - X^0 + S \\ S = \frac{1 - X^{N_x}}{1 - X} \end{gathered} \tag{17.5}$$

Now replace *X* by $e^{\mathrm{i}2\pi\mathbf{K}.\mathbf{a}}$ to find the sum.

$$\sum_{n_x=0}^{n_x=N_x-1} e^{\mathrm{i}\,2\pi n_x \mathbf{K}.\mathbf{a}} = \frac{1 - e^{\mathrm{i}\,2\pi N_x \mathbf{K}.\mathbf{a}}}{1 - e^{\mathrm{i}\,2\pi \mathbf{K}.\mathbf{a}}} \tag{17.6}$$

Since we are interested in the intensities, we multiply this sum by its complex conjugate. To do this we use some simple trigonometric relationships

$$\begin{aligned} \left(1-e^{-\mathrm{i}\alpha}\right)\left(1-e^{\mathrm{i}\alpha}\right) &= (1-\cos\alpha+\mathrm{i}\sin\alpha)(1-\cos\alpha-\mathrm{i}\sin\alpha) \\ &= \left(1-2\cos\alpha+\cos^2\alpha\right)+\sin^2\alpha = 2(1-\cos\alpha) = 4\sin^2\frac{\alpha}{2} \end{aligned} \tag{17.7a}$$

or

$$\left(1 - e^{-\mathrm{i}\,\alpha}\right)\left(1 - e^{\mathrm{i}\,\alpha}\right) = 4\left(\frac{e^{\mathrm{i}\frac{\alpha}{2}} - e^{-\mathrm{i}\frac{\alpha}{2}}}{2}\right)\left(\frac{e^{-\mathrm{i}\frac{\alpha}{2}} - e^{\mathrm{i}\frac{\alpha}{2}}}{2}\right) \tag{17.7b}$$

The intensity is then related to

$$\left|\sum_{n_x=0}^{n_x=N_x-1} e^{\mathrm{i}2\pi n_x \mathbf{K}.\mathbf{a}}\right|^2 = \frac{\left(1 - e^{-\mathrm{i}2\pi N_x \mathbf{K}.\mathbf{a}}\right)}{1 - e^{-\mathrm{i}2\pi \mathbf{K}.\mathbf{a}}} \frac{\left(1 - e^{-\mathrm{i}2\pi N_x \mathbf{K}.\mathbf{a}}\right)}{1 - e^{-\mathrm{i}2\pi \mathbf{K}.\mathbf{a}}} \tag{17.8a}$$

or

$$\left|\sum_{n_x=0}^{n_x=N-1} e^{\mathrm{i}2\pi N_x \mathbf{K}.\mathbf{a}}\right|^2 = \frac{4\sin^2(\pi N_x \mathbf{K}.\mathbf{a})}{4\sin^2(\pi \mathbf{K}.\mathbf{a})} \tag{17.8b}$$

Then we can write

$$\begin{aligned} I = |A|^2 = |F|^2 &\left(\frac{\sin^2(\pi N_x \mathbf{K}.\mathbf{a})}{\sin^2(\pi \mathbf{K}.\mathbf{a})}\right) \\ &\left(\frac{\sin^2(\pi N_y \mathbf{K}.\mathbf{b})}{\sin^2(\pi \mathbf{K}.\mathbf{b})}\right)\left(\frac{\sin^2(\pi N_z \mathbf{K}.\mathbf{c})}{\sin^2(\pi \mathbf{K}.\mathbf{c})}\right) \end{aligned} \tag{17.9}$$

If the dot product $\mathbf{K}.\mathbf{a}$ is an integer, then the first of these terms is unity. This is, of course, the Bragg condition and the intensity is then a maximum. There are also subsidiary maxima or minima when

$$\pi N_x \mathbf{K}.\mathbf{a} = \frac{\pi}{2} C \tag{17.10}$$

where C = an integer. Reordering this equation, we have

$$\mathbf{K}.\mathbf{a} = \frac{C}{2N_x} \tag{17.11}$$

Equation 17.9 is the basis of the shape effect and leads to the idea of the relrod, which you recall is the name we give to a reciprocal-lattice rod (look back at Section 12.5).

17.1.B The Integration Approach

If we take equation 13.2, which is the amplitude diffracted by a single unit cell, and sum this over all the cells in the specimen, the amplitude of the diffracted beam can be written as

$$\phi_{\mathbf{g}} = \frac{e^{2\pi i \mathbf{k}.\mathbf{r}}}{r} \sum_n F_n e^{(-2\pi i \mathbf{K}.\mathbf{r}_n)} \qquad (17.12)$$

Since we have defined $\mathbf{K}$ to be $\mathbf{g} + \mathbf{s}$, we can rewrite this equation as

$$\phi_{\mathbf{g}} = \frac{e^{2\pi i \mathbf{k}.\mathbf{r}}}{r} \sum_n F_{\mathbf{g}} e^{\left(-2\pi i (\mathbf{g}+\mathbf{s}_{\mathbf{g}}).\mathbf{r}_n\right)} \qquad (17.13)$$

Now we know that $\mathbf{g}.\mathbf{r}_n$ is an integer by the definition of $\mathbf{g}$ and $\mathbf{r}_n$ and we will refer to $\mathbf{s}_{\mathbf{g}}$ as $\mathbf{s}$. Hence we can write equation 17.13 as

$$\phi_{\mathbf{g}} = \frac{e^{2\pi i \mathbf{k}.\mathbf{r}}}{r} \sum_n F_{\mathbf{g}} e^{(-2\pi i \mathbf{s}.\mathbf{r}_n)} \qquad (17.14)$$

where $\mathbf{s}$ is the deviation parameter for reflection $\mathbf{g}$. If we make the approximation that the crystal contains many unit cells, we can replace this sum by an integral to give

$$\phi_{\mathbf{g}} = \frac{e^{2\pi i \mathbf{k}.\mathbf{r}}}{rV_c} F_{\mathbf{g}} \int_{\text{crystal}} e^{(-2\pi i \mathbf{s}.\mathbf{r}_n)} dv \qquad (17.15)$$

This is where the present treatment differs from the first. If we now express $\mathbf{s}$ and $\mathbf{r}_n$ as the vectors

$$\mathbf{s} = u\mathbf{a}^* + v\mathbf{b}^* + w\mathbf{c}^* \qquad (17.16)$$

and

$$\mathbf{r}_n = h\mathbf{a} + k\mathbf{b} + l\mathbf{c} \qquad (17.17)$$

then we can write

$$\phi_{\mathbf{g}} = \frac{e^{2\pi i \mathbf{k}.\mathbf{r}}}{rV_c} F_{\mathbf{g}} \int_0^C \int_0^B \int_0^A e^{-2\pi i (ux+vy+wz)} dx dy dz \qquad (17.18)$$

where $A = N_x a$, etc. This integral is straightforward.

$$\int_0^A e^{-2\pi i u x} = \frac{e^{-2\pi i u A} - 1}{-2\pi i u} = \left(\frac{e^{-\pi i u A}}{\pi u}\right)\left(\frac{e^{\pi i u A} - e^{-\pi i u A}}{2i}\right)$$

$$= \frac{e^{-\pi i u A}}{\pi u} \sin(\pi u A) \qquad (17.19)$$

$$\phi_{\mathbf{g}} = \frac{e^{2\pi i \mathbf{k}.\mathbf{r}}}{rV_c} F_{\mathbf{g}} \frac{(\sin \pi A u)}{(\pi u)} \frac{(\sin \pi B v)}{(\pi v)} \frac{(\sin \pi C w)}{(\pi w)} e^{iD} \qquad (17.20)$$

(D is an unimportant phase factor.) The intensity is then as given by equation 17.9, but we have explicitly kept the r^{-2} and V_c^{-2} dependence for the intensities.

You should recognize the form of equations 17.9 and 17.20. These equations have the same form as that given back in equation 2.12 for diffraction from a diffraction grating. The corresponding diffraction grating has N_x lines which are spaced a distance a apart. The physical similarity is that the grating, just like our crystal, has a finite size.

17.2 THE THIN-FOIL EFFECT

Equation 17.9 is very important for TEM. It tells us why the relrods we introduced in Chapter 12 have a finite length if we measure them to the first minimum. It also tells us that the diffracted intensity does depend on the value of $\mathbf{s}$; it is not a constant for any position along the rod.

RELRODS AND INTENSITY

Just remember that when we said "the intensity" we meant "the intensity which the diffracted beam will have if s takes a particular value; i.e., if the Ewald sphere cuts the relrod at that point."

We can better appreciate this variation along the rod if we plot the intensity and draw the Ewald sphere, as shown in Figure 17.2. We only draw the intensity plot for one direction at a time. This diagram shows the Ewald sphere cutting the relrod on one side while showing the intensity along the relrod on the right-hand plot.

Figure 17.2 is an extension of Ewald's 'pictorial representation' of diffraction. We can now draw the reciprocal lattice as shown for a simple-cubic crystal in Figure 17.3, such that every point is replaced by a relrod and every relrod is described by equation 17.9. If the surface of the crystal is exactly parallel to the (112) plane, but we orient the specimen slightly off the [001] pole, then the Ewald sphere cuts the relrod as different positions relative to the square array of spots, which is the projection of the spots at zero tilt (Figure 17.3B). The DP will appear as shown in Figure 17.3C. In Figure 17.3C, C is the projected position of the center of the

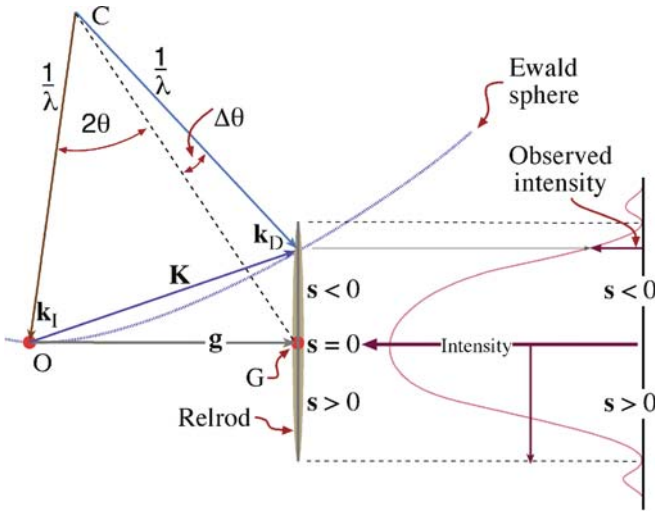


FIGURE 17.2. The relrod at $\mathbf{g}_{hkl}$ when the beam is $\Delta\theta$ away from the exact Bragg condition. The Ewald sphere intercepts the relrod at a negative value of **s** which defines the vector $\mathbf{K} = \mathbf{g} + \mathbf{s}$. The intensity of the diffracted beam as a function of where the Ewald sphere cuts the relrod is shown on the right of the diagram. In this case the intensity has fallen to almost zero.

Ewald sphere. As an exercise, consider whether the pattern would differ if the surface were cut slightly off (001) but oriented at the [001] pole. Then repeat the first exercise but instead of tilting the specimen, tilt the electron beam through the same small angle.

Remember that we deduced equation 17.9 by simply adding the amplitudes from all the unit cells, taking the position of the cells into account.

We calculated a 'structure factor' for the whole volume which contributes to $\phi_{\mathbf{g}}$: we call this calculated factor the shape factor.

We should then use this shape factor rather than the structure factor (since F is included in equation 17.9) in our dynamical calculations of $\phi_{\mathbf{g}}$. The problem is, of course, that the shape factor can be different for every specimen we examine.

We have just deduced a method for picturing how the shape of a perfect parallelepiped (of sides $N_x\,a$, $N_y\,b$ and $N_z\,c$) affects the DP. Now for the next step, we will use this concept of the shape factor to examine how the DPs will be affected by more complex shapes, such as the wedge shape of many real TEM specimens or the perfect parallelepiped of the stacking fault. Then we will consider defects which themselves do not have sharp boundaries; the dislocation is a perfect example of such an imperfection.

17.3 DIFFRACTION FROM WEDGE-SHAPED SPECIMENS

Most TEM specimens do not have parallel surfaces but are wedge-shaped. In drawing the relrods for such a wedge-shaped specimen, we extend the results of Section 17.2

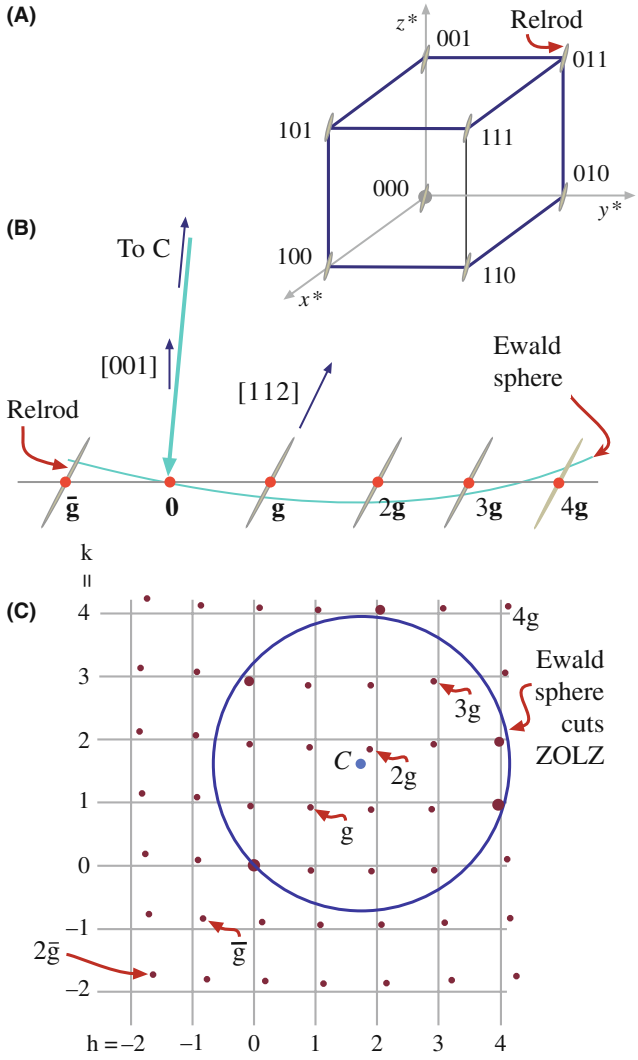


FIGURE 17.3. (A) For a thin specimen, every point is replaced by a relrod. (B) The Ewald sphere cutting the relrods in (A) when the crystal is tilted slightly off the (001) axis. (C) The effect of the tilt in (B) on the DP. Notice that all of the spots in the DP are displaced relative to their positions on the square grid (the projection of the spots at zero tilt), but that the magnitude of the displacement varies depending on the sign and size of **s**. Of course, spots on the Ewald sphere must be the 'correct' distance from 000.

by saying that the relrod will always be normal to the surface. So, for a wedge-shaped specimen (Figure 17.4A) we must have two relrods as shown in Figure 17.4B. What we see in the DP is determined by how the Ewald sphere cuts these two relrods. As shown in Figure 17.4C and D, we will see two spots which lie along a line which is normal to the edge of the wedge. Notice that all the pairs of spots are aligned in the same direction as we expected and that their separation is larger for larger values of **s**. This simple relrod model predicts that we would see only one spot if $\mathbf{s} = 0$. In fact, we should see two or more spots because the relrod model fails when we are in a strongly dynamical diffraction condition. We will return to this point in the next section and again in Chapter 24.

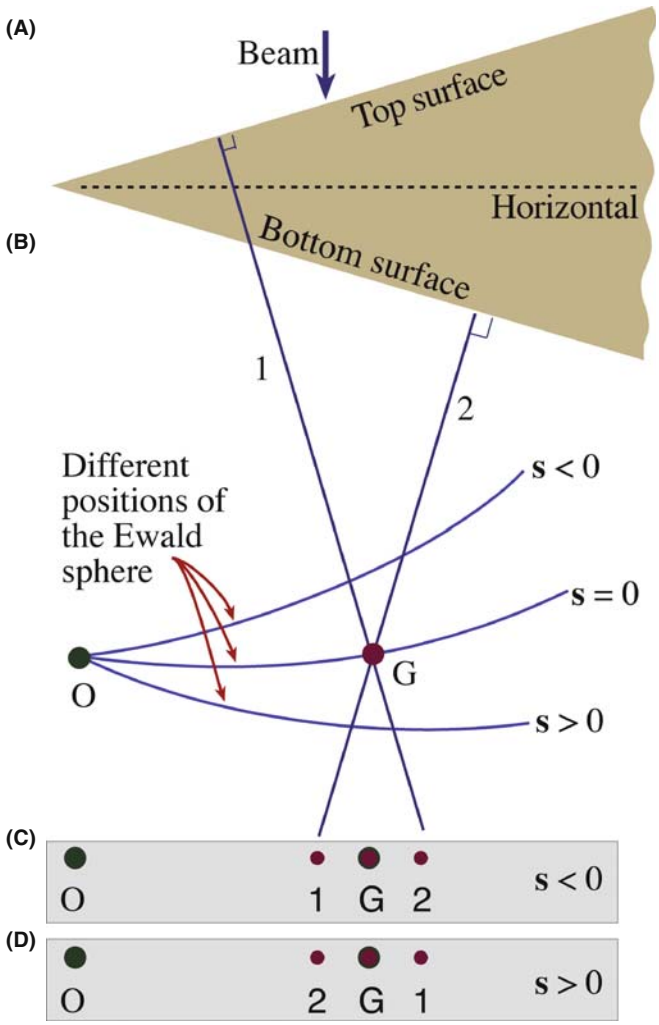


FIGURE 17.4. (A) Diffraction from a wedge-shaped crystal. (B) Notice that when $\mathbf{s} < 0$, relrod 1 is on the left of relrod 2 but the order reverses when $\mathbf{s}$ becomes > 0. The effect of this pair of relrods is to create a doublet shown in (C) and (D). The middle spot is the matrix relrod for a parallel-sided thin foil and is absent for the wedge in A.

17.4 DIFFRACTION FROM PLANAR DEFECTS

The shape factor concept can be readily applied to understand diffraction from a flat platelet or planar fault. The geometry is shown in Figure 17.5. The idea is that the platelet is itself a thin parallelepiped which is inclined to the specimen parallelepiped (Figure 17.5A). The result is that we have two relrods, one normal to the specimen surface and a much longer one normal to the thin platelet (Figure 17.5B and C). When we cut these relrods with the Ewald sphere we produce two spots in the DP and, as for the wedge specimen, the separation of the spots increases with increasing $\mathbf{s}$. The line MN lies normal to the trace of the platelet. There are, however, some differences in this case. Although the *m* and *n* relrods are very different in length and actual intensity,

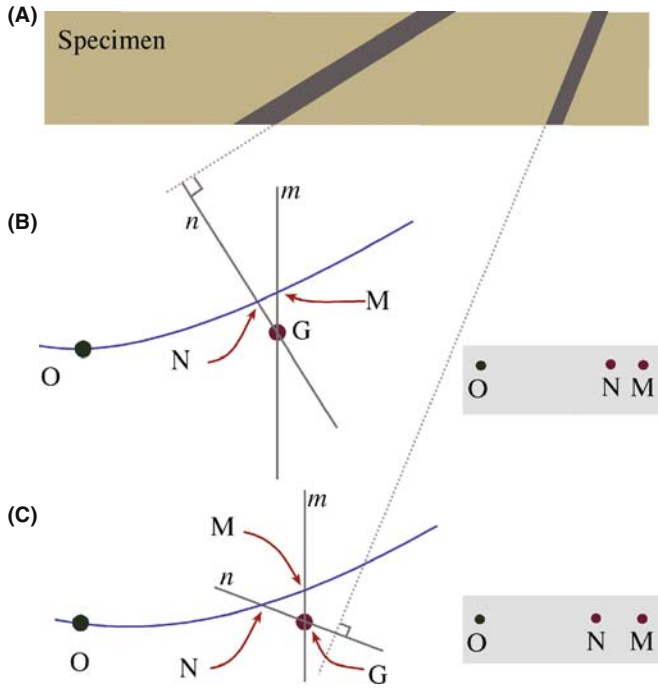


FIGURE 17.5. The effect of a thin inclined plate in a thin specimen. (A) Two plates are shown to illustrate the effect of changing the inclination of the plate relative to the foil surface. When $\mathbf{s} \neq 0$ we see two spots in the DP because there are two relrods for the two different planar-defect inclinations in (B) and (C).

the diffracting volume is much greater for the specimen than for the platelet. Thus, we can usually distinguish reflections M and N.

Providing we know the orientation of the specimen relative to the DP, we can tell whether the inclination angle is less than or greater than 90°; i.e., we can determine the inclination of the planar defect without moving the specimen or using any theory of image contrast (see Chapter 24). As in Section17.3, we would actually see two spots when $\mathbf{s} = 0$ if we could make the spots small enough. We'll return to this topic in Section 17.7.

A stacking fault in an fcc crystal can be thought of as a very thin platelet of hcp material as shown in Figure 17.6; so it really is a platelet with perfect lattice matching parallel to its surface.

We can understand diffraction effects from other planar interfaces by considering two cases

- If the grains on either side of the interface contain a common reflection, then the diffraction effects can be modeled by the thin platelet.
- In the case where a reflection is not common to the two grains, then for that reflection the diffracting crystal behaves like a wedge specimen with one surface parallel to the planar defect. We can ignore the crystal that is not diffracting.

The two DPs in Figure 17.7 show that you really do see pairs of spots for these two types of boundary. As

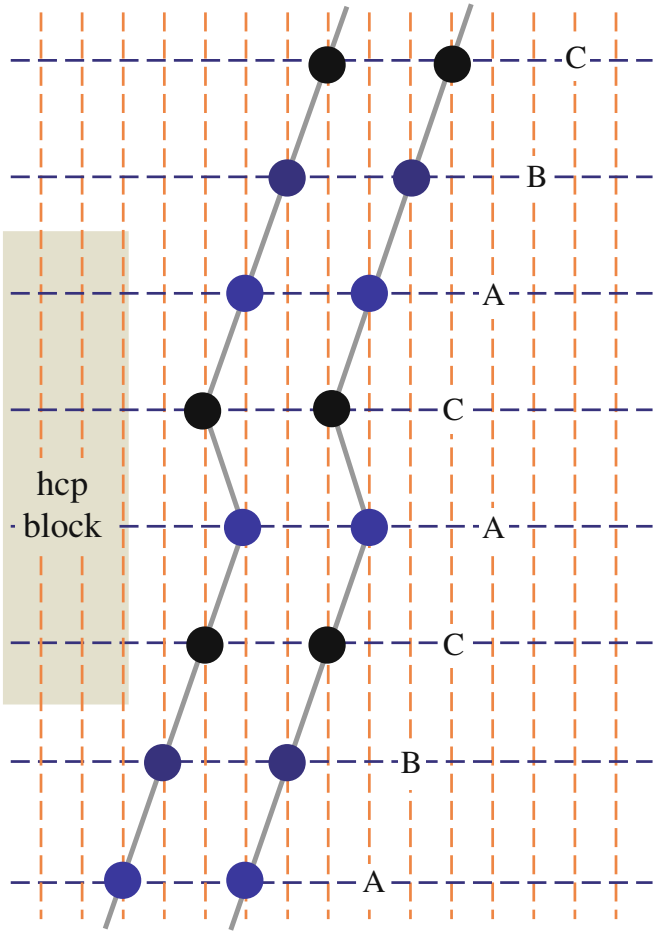


FIGURE 17.6. Schematic of the stacking sequence of close-packed planes A, B, C, in an fcc crystal showing that the SF is similar to a thin layer of hcp material, stacking ACAC.

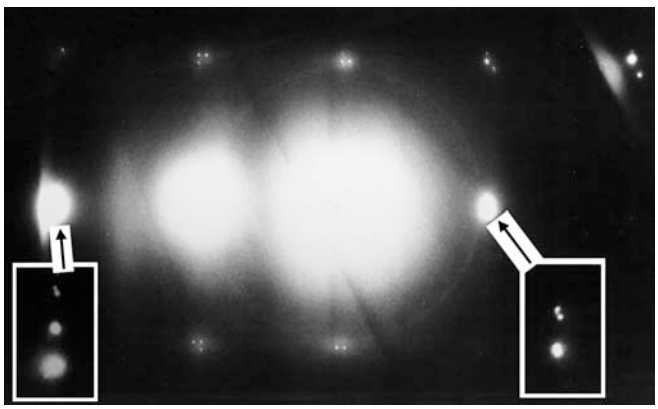


FIGURE 17.7. Pairs of spots in a DP from a grain boundary.

before, the two spots lie normal to the boundary traces, i.e., the intersection of the boundary with the surface of the specimen.

There are two reasons for emphasizing the extra spots which are present because of the interface

- You should always check that any extra spots you see cannot be explained in this way.
- You must be careful when determining spot spacing (as when estimating lattice parameters). You must set **s** to zero for this purpose and that can usually only be done for a few reflections at any time (one reflection always being O, not D, of course).

Twin boundaries are often found to consist of flat segments in particular orientations. The first-order twin boundary in fcc crystals tends to facet parallel to the common {111} plane as shown schematically in Figure 17.8A. This means that if we orient the specimen so that this common plane is nearly parallel to the beam, we will excite the common {111} reflection. Now our platelet is parallel to the beam so that its relrod is normal to the beam. If the specimen is also thin, we can arrange that the Ewald sphere cuts along the length of the relrod. Now, as we can see in Figure 17.8B, we see a 'streak' in the DP rather than a spot. The streak actually extends from in the [111] direction because, as you can appreciate from Figure 17.8A, the twin is a *very* thin platelet.

FROM NANOPARTICLES

Since we'll only detect any effect when the particle is $\leq$100 nm, these particles could be called nanoparticles. Anyone working with such particles must understand this effect which is closely related to Debye-Scherrer analysis in XRD.

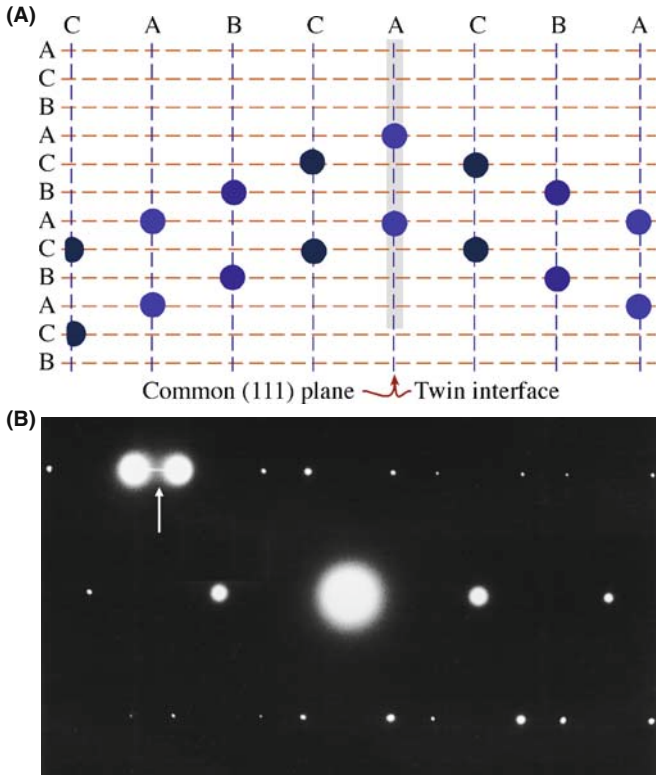


FIGURE 17.8. (A) Schematic of twin and (B) DP with a streak (arrowed) normal to the twin plane. Note that $\mathbf{s} = 0$ for the two bright diffracted spots.

If we regard the surface as a planar defect, we can also observe extra spots in the DP due to a reconstruction of the surface. One factor to be cautious about is that the apparent reconstruction might be influenced by contamination since the TEM is not generally a UHV system.

17.5 DIFFRACTION FROM PARTICLES

Particles come in all shapes and sizes, so we will not try to be exhaustive. Actually, the principle involved in determining the shape factor in reciprocal space is simply 'small becomes large' and vice versa. The shape factors are shown schematically for several particles in Figure 17.9. You should be aware that you will probably never see the subsidiary minima shown in these diagrams.

One example which is common is the platelet shown in Figure 17.10; these can occur as GP zones or other

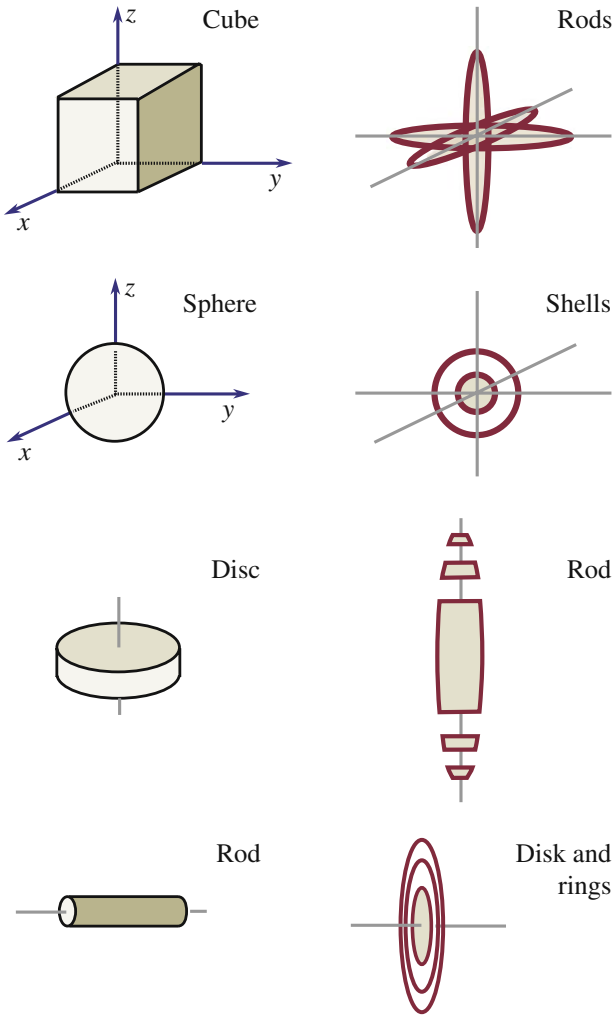


FIGURE 17.9. Examples of how spots in reciprocal space have different shapes depending on the shape of the particles which are diffracting.

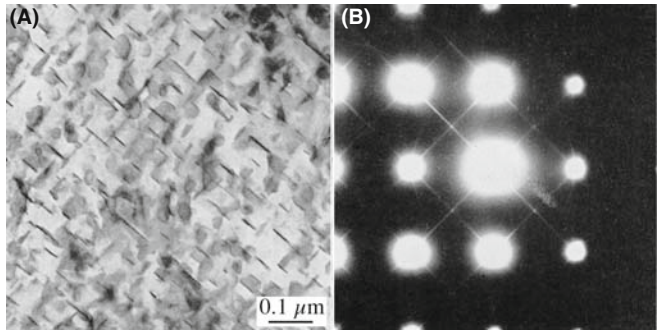


FIGURE 17.10. Very thin plate-like precipitates (A) cause long streaks in the DP (B). In this example, the precipitates are GP zones in an Fe-2.9 at.% Mo alloy.

thin disk-shaped precipitates. When the platelets are oriented parallel to the beam, we see streaks in the DP just as we saw them in Figure 17.8B. The difference in this figure is that the platelets can lie on all the crystallographically equivalent planes in the crystal. For these GP zones they lie on {001} planes so the streaks run in <001= directions for the cubic crystal connecting, for example, 000 and 200. You should notice that these spots would still be connected if the crystal were not cubic. You'll also see that there is a sharp point at the 100 position even though 100 is not an allowed reflection for bcc crystals. The reason we see this spot is that we are cutting the relrod which runs parallel to the electron beam in the [001] direction.

The smallest 'particle' can be thought of as a vacancy, a substitutional atom or an interstitial atom. We will not expect to see any clear effect of a single point defect but, as we saw in Section 16.7, these point defects can order to give a clear superlattice, and therefore extra spots.

As you might expect then, if we have many point defects but not enough to give long-range order, we might expect short-range ordering. Perhaps the clearest example of this phenomenon again occurs in the metal carbides. The effect is shown in Figure 17.11. The short-range ordering gives rise to diffuse scattering in the DP which at first appears quite random, sometimes as circles around the spots and appearing at other times as circles between spots or not circles at all! By combining many different patterns, Sauvage and Parthé proposed that the diffuse scattering could be mapped out as shown in Figure 17.11D. This figure strongly resembles a Fermi surface diagram which you may have encountered in condensed-matter physics. We will discuss some aspects of imaging using diffusely scattered electrons in Section 29.4, but the important points to recognize are

- Point defects really can cause diffraction effects, especially if they interact with one another.
- Diffuse scattering can still be interpreted by the Ewald-sphere construction.

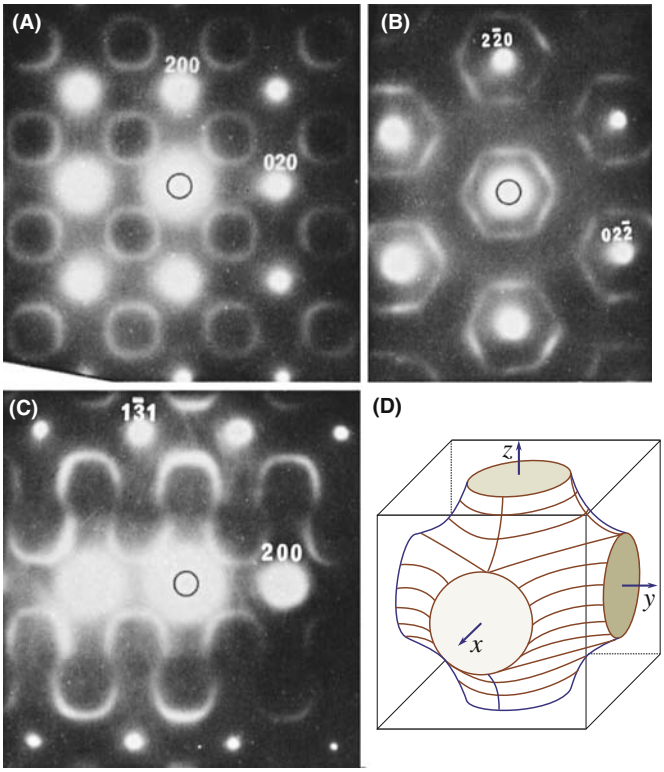


FIGURE 17.11. Short-range ordering can cause diffuse scattering in the DP (A–C). The DPs in this example were obtained from a vanadium carbide. In this case, the 3D map of diffuse intensity has a shape which strongly resembles a Fermi surface shown in (D).

If you are intrigued by this topic, you will find the literature on discommensurate structures in intercalated material a complementary challenge. A library/web search on 'discommensurate' and 'intercalated' will quickly net more recent papers.

17.6 DIFFRACTION FROM DISLOCATIONS, INDIVIDUALLY AND COLLECTIVELY

In Chapter 25 we will discuss images of dislocations. A dislocation is a line defect that is characterized by its line direction and its Burgers vector. The crystal around the defect is distorted or strained.

For a single dislocation, this strain is not expected to cause new spots in the DP, but we do expect diffuse scattering since the dislocation is a line defect. If a region from 0.2 to 1 nm around the core is greatly distorted (we'll see the effect of this strain in Chapter 25), then the diffuse scattering will extend from 1 nm^{-1} to perhaps 5 nm^{-1} from the reciprocal-lattice points, giving a diffuse disk (the reciprocal shape of a long needle). Some planes are essentially unaffected by the dislocations, so we might expect the diffuse scattering to vary in magnitude for the different reciprocal-lattice points. (We'll examine this $\mathbf{g}.\mathbf{b} = 0$ effect in Chapter 24.)

With this simple discussion and without ever seeing this diffuse scattering, we can draw an important conclusion: if we want to learn about the structure of a dislocation core, we must include the diffuse scattering in the image-formation process. We must include that intensity in the objective aperture and the corresponding image calculations.

DIFFUSE SCATTERING

The diffuse intensity from a dislocation is *not* located at the reciprocal-lattice point.

Because the distorted volume associated with a single dislocation is so small, we do not expect to see this intensity in the DP unless we have many dislocations in an ordered array (just like the point defects in Section 17.5). We can demonstrate that this intensity is present by diffracting from an ordered array of dislocations as shown in Figure 17.12. The specimen used to form this image was rather special. Dislocations are present in region A, but not in region B. The array actually forms a structured grain boundary in A, but a layer of glass is present in B. The insets show the same part of the SADPs from the two regions. In B, you can see three spots. The top two are from one grain, the bottom one is from the other grain. The reason for the pair is that **s** is

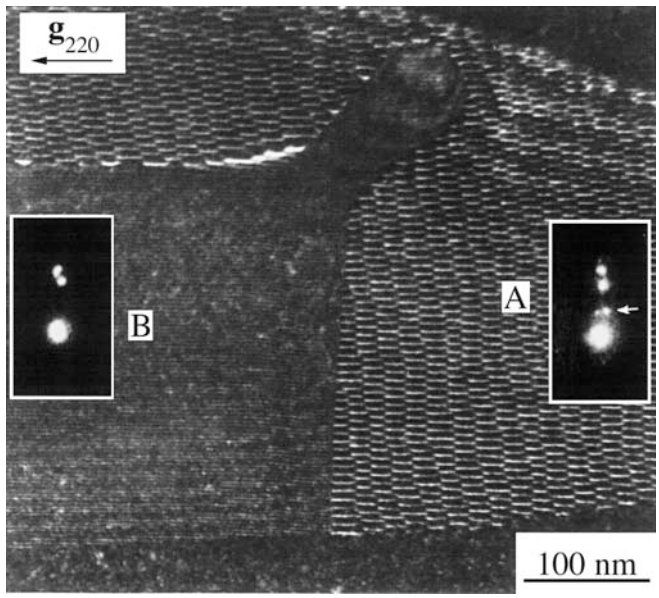


FIGURE 17.12. Diffraction from an ordered array of dislocations. Dislocations are present in region A, but not in region B. The insets show a small part of the DP from the two regions. The extra spots present in A are caused by the visible array of dislocations; these spots are a doublet because there is also a second, nearly orthogonal, set of dislocations present which acts as a separate grating. The other pair is due to the wedge shape and so is common to both DPs.

large for that grain, but almost zero for the other. This is an example of the application of Section 17.3.

In A you see the same three spots (because the two grains are still present) but now there are two extra spots. The reason we see two extra spots is that we have two arrays of dislocations. You are seeing the scattering from the dislocations because they have formed an array with long-range ordering, just like the vacancies in V_8C_7 in Chapter 16.

If you look at the DP when the array of dislocations lies parallel to the beam, you may be able to see a set of streaks as shown in Figure 17.13. The separation of the streaks is the inverse of the actual separation of the dislocations. You see streaks because you have relrods in reciprocal space and we are cutting along these rods with the Ewald sphere. The length of the relrods gives you a measure of how far the strain field of the dislocations extends out into the two grains. In other words, we are seeing a thickness of the strain-field regions. The object of this discussion is not to examine grain boundaries, but to show that the strain field from an array of dislocations causes scattering in the DP and thus to infer that one dislocation will also cause scattering, but it will just be much more diffuse (and very weak).

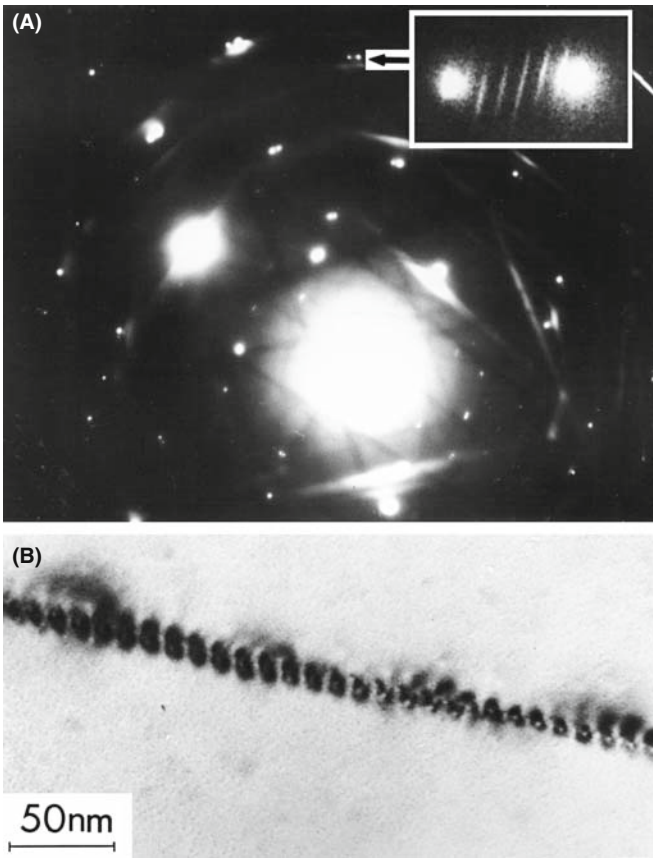


FIGURE 17.13. (A) The set of streaks from an array of dislocations in Al_2O_3 lying parallel to the electron beam. The distance between the streaks is inversely related to the spacing of the dislocations shown in the image (B).

Before moving on, consider the diffraction spots in Figure 17.12 again. Why are the pairs of dislocation spots (arrowed) located where they are? Put another way: which of the two spots in region B corresponds to the N relrod and which corresponds to the M relrod? (See Figure 17.5 for the definition of M and N.)

THE PERIODICITY RULE

If there is a structural periodicity in real space, then there will be an array of points or relrods in reciprocal space and an array of spots or streaks in the DP.

We then ask a simple question: how many objects are required in order to produce a detectable effect in the DP? The answer is two! This point is illustrated in Figure 17.14, which shows a DP and an image of two twin boundaries which are ~15 nm apart. The spacing of the new spots between the twin spots in the DP (expanded in the inserts) is 0.067 nm^{-1}, as expected. Now, why can this occur? The analogy is the Young's slits experiment in visible-light optics. The illustration also reminds us of a special feature of the TEM, namely, that even without a FEG, the electron beam is remarkably coherent.

17.7 DIFFRACTION AND THE DISPERSION SURFACE

Several times in this chapter, we have said "actually, we will see two spots when $\mathbf{s} = 0$," even though the relrod model says that you will only see one. The origin of two spots from a wedge specimen (there may be more for more complicated defects) is due to the dynamical nature of the scattering process. The theory has been derived by Amelinckx and his co-workers in a series of papers. Unfortunately, this group used a different notation, but they did summarize their results graphically. We will also return to this topic when we discuss images in Chapter 24. As an example, the relrod diagram given for the stacking fault in Figure 17.5 should be drawn so that the relrods are the asymptotic to two straight lines, as shown in Figure 17.15. When the Ewald sphere cuts these curves at $\mathbf{s} = 0$, we see that there are two spots which move apart as we increase $\mathbf{s}$ (either positive or, as shown here, negative) until they are at the points defined by the straight lines. So, will there not be a vector that exactly corresponds to $\mathbf{g}$? The answer, of course, is yes, because of the adjacent perfect crystal so we must have three spots, but these are very difficult to see because $\mathbf{s}$ must be very close to zero. Without going into any theory, we can guess the origin of these curves: they look remarkably like the curves of the dispersion

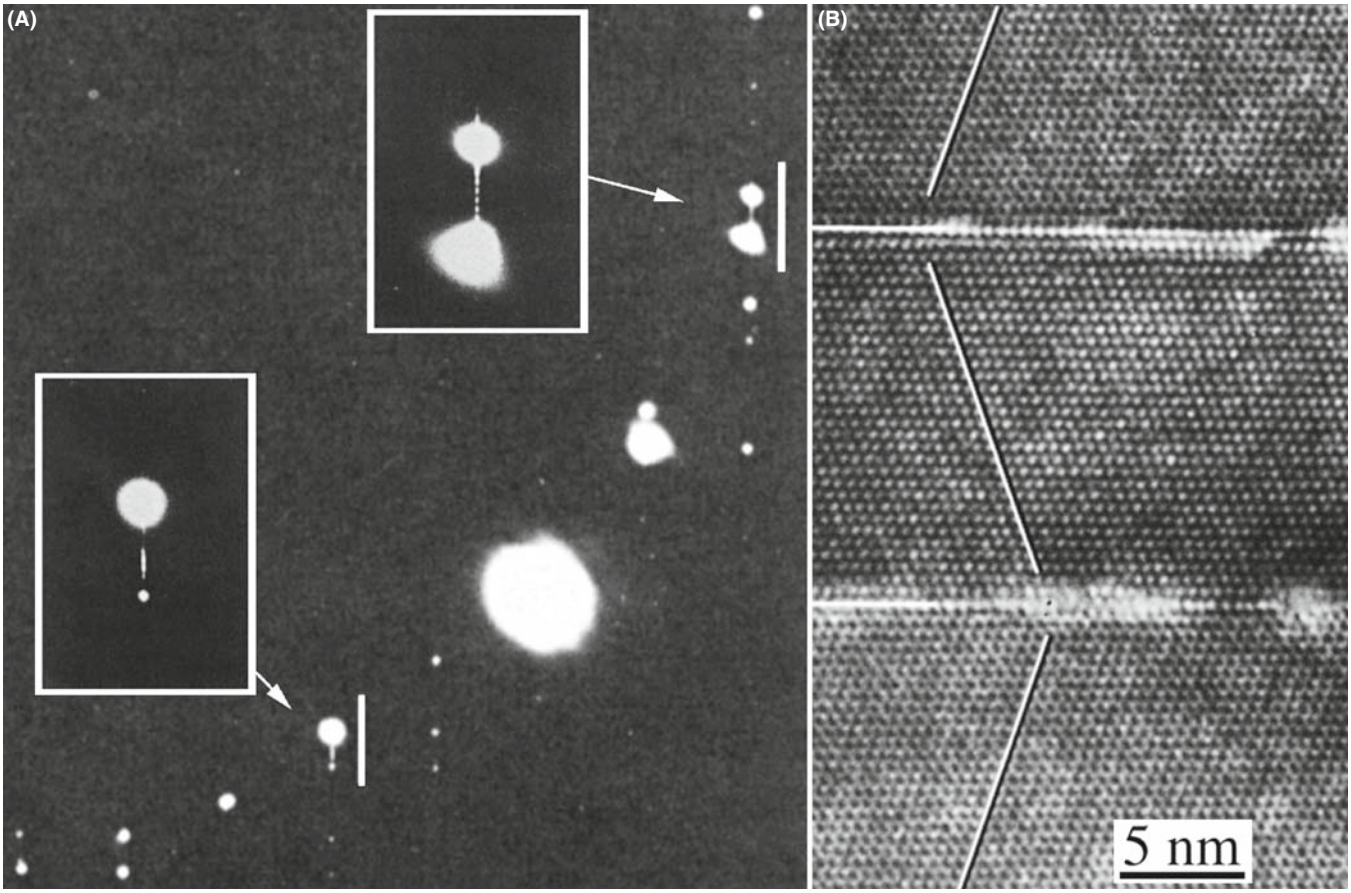


FIGURE 17.14. Extra spots can be formed in the DP (A) when only two defects are scattering in phase. The separation of the extra spots is related to the inverse of the separation of the two twin boundaries seen in the image (B).

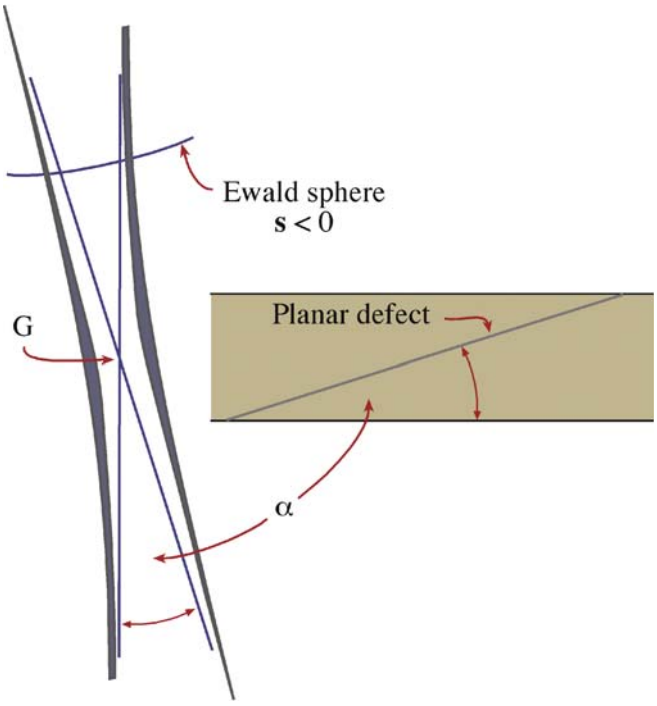


FIGURE 17.15. The relrods from two planes inclined at angle α are actually the asymptotes to two straight lines, so that they don't cross at G; when $\mathbf{s} = 0$, the distance between these two curves is $\xi_{\mathbf{g}}^{-1}$.

surface which also had asymptotes (see Figure 15.3). These curves are indeed directly related.

When you increase **s,** you move out of the dynamical regime and into the kinematical one, where the simple relrod model applies (see Chapter 26). At $\mathbf{s} = 0$, the distance between the curves is inversely proportional to $\xi_{\mathbf{g}}$, the extinction distance for reflection **g**.

You can understand why this is so in the following pictorial way. What you see in the image will be determined by the DP. What you see in the DP is determined by which relrods, or surfaces, the Ewald sphere intersects. All the information about extinction distances and coupling of diffracted beams is fundamentally contained in the dispersion surface ($\xi_{\mathbf{g}}$ is just Δk^{-1} at $\mathbf{s} = 0$). Both the dispersion surface and the reciprocal-lattice/Ewald-sphere models are just pictorial representations of the same diffraction process. So, all the information in the dispersion-surface model should also be present in the reciprocal-lattice/Ewald-sphere model.

The relrods are the asymptotes to these two hyperbolas. Alternatively, we could say that the relrods and the asymptotes are a result of the kinematical diffraction

approximation. There is a one-to-one correlation between what happens at the dispersion surface in the vicinity of the BZB to what happens when the Ewald sphere cuts the relrods in the vicinity of the reciprocal-lattice point, G. Imagine rotating the dispersion-surface diagram through 90°. These ideas have been extensively studied by van Landuyt, de Ridder, Gevers, Amelinckx et al., as summarized in the general references at the end of this chapter. What Amelinckx's group has done is to give us the rules on how to transfer this information from the dispersion surface to the reciprocal lattice and hence to the DP. In Section 24.9, we'll relate this concept to images. If you thought dispersion surfaces were difficult, make **s** large and stick to relrods!

CHAPTER SUMMARY

In this chapter, we have begun to examine the unique features of diffraction in the TEM. These features arise because we are always diffracting from small volumes. The sizes of both our specimen and the special features present in our specimen are always small, so that we must take into account the shape effect. This is particularly important for nanoparticles and nanograins—you must keep this in mind when imaging. Of course, the same considerations will also apply to other forms of diffraction; it's just that only TEM can examine the diffraction information from the vicinity of crystal defects. In other words, the shape effect is not a limitation due to the fact that we are using high-energy electrons. By understanding the concept of the shape effect you can actually learn more about defects in crystals; conversely, you can make some major errors if you do not understand the shape effect. Two points to remember are

- When a platelet is parallel to the beam its relrod is normal to the beam. If the specimen is also thin, we can arrange that the Ewald sphere cuts along the length of the relrod. Now you see a 'streak' in the DP rather than a spot.
- Beam splitting at $\mathbf{s} = 0$ and the dispersion surface both arise because of dynamical scattering.

DIFFRACTION FROM INTERFACES

The relation between diffraction and images from planar defects has been the subject of a long series of papers from the group led by Prof. Severin Amelinckx. The examples below from Phys. stat. sol. will give you a start to your study.

Carter, CB (1984) *Electron Diffraction from Microtwins and Long-Period Polytypes* Phil. Mag. A **50** 133–141. The Young's slit experiment in the TEM.

de Ridder, R, Van Landuyt, J, Gevers, R and Amelinckx, S (1968) *The Fine Structure of Spots in Electron Diffraction Resulting from the Presence of Planar Interfaces and Dislocations. IV. Wedge Crystals* Phys. stat. sol. **30** 797–815; See also: (1970) *ibid.* **38** 747; (1970) *ibid.* **40** 271; (1970) *ibid.* **41** 519; (1970) *ibid.* **42** 645.

Gevers, R (1971) in *Electron Microscopy in Materials Science* (Ed. U. Valdrè) p302–310, Academic Press, New York. An introduction to the work of Amelinckx' group.

Gevers, R, Van Landuyt, J and Amelinckx, S (1966) *The Fine Structure of Spots in Electron Diffraction Resulting from the Presence of Planar Interfaces and Dislocations. I. General Theory and Its Application to Stacking Faults and Anti-phase Boundaries* Phys. stat. sol. **18** 343–361; See also (1967) *ibid.* **21** 393; (1967) *ibid.* **23** 549; (1968) *ibid.* **26** 577.

Van Landuyt, J (1964) *An Electron Microscopic Investigation of Phenomena Associated with Solid Solution of Oxygen in Niobium* Phys. stat. sol. **6** 957–974.

Van Landuyt, J, Gevers, R and Amelinckx, S (1966) *On the Determination of the Nature of Stacking Faults in fcc Metals from the Bright Field Image* Phys. stat. sol. **18** 167–172.

DIFFUSE SCATTERING AND INTERCALATION

Carter, CB and Williams, PM, 1972, *An Electron Microscopy Study of Intercalation in Transition Metal Dichalcogenides* Phil. Mag. **26** (2), 393–398. To encourage new students!

Sauvage, M and Parthé, E (1972) *Vacancy Short-Range Order in Substoichiometric Transition Metal Carbides and Nitrides with the NaCl structure. II. Numerical Calculation of Vacancy Arrangement* Acta Cryst. **A28** 607–616.

Wilson, JA, Di Salvo, FJ and Mahajan, S (1975) *Charge-Density Waves and Superlattices in the Metallic Layered Transition Metal Dichalcogenides* Adv. Phys. **24** 117–201. An early review of intercalation.

THE COMPANION TEXT

Although there is great interest in nanomaterials and XRD is often used to characterize them, diffraction from such materials in the TEM is used much less because imaging is then almost routine. Hence there have not been recent advances in electron diffraction except as CBED, which is treated in the companion text.

SELF-ASSESSMENT QUESTIONS

Q17.1 When the Bragg condition is not satisfied exactly, why do spots appear in the DP?
Q17.2 Reproduce the summation and integral calculations of amplitudes to explain the existence of relrods.
Q17.3 When the specimen is tilted slightly off a particular zone, why might spots in the DP move away from their ideal positions in the DP?
Q17.4 Is a relrod a 'real' object?
Q17.5 In a specimen containing thin platelets, how is it possible to distinguish the DP spots associated with the thinness of the specimen from those due to the platelets?
Q17.6 If a single dislocation is observed in the specimen, should the DP change?
Q17.7 When is it possible to view along the entire length of a relrod?
Q17.8 The spots in the right inset of Figure 17.14 are very clear. Why did we show this region rather than the region between 000 and the common 111?
Q17.9 What is the definition of a twin boundary?
Q17.10 For what conditions might you see diffuse circles surround the DP spots?
Q17.11 How does the thickness of the TEM foil affect how (when) the Laue conditions are satisfied?
Q17.12 How can the thickness of the specimen affect the accuracy of lattice-parameter determination?
Q17.13 How can the wedge shape of a specimen influence the appearance of the diffraction spots?
Q17.14 When $\mathbf{s} = 0$, the (curved) relrod models predicts two spots will be seen in the DP. Why?
Q17.15 In Figure 17.8, we can see a streak between two spots in the DP. What does this tell us?
Q17.16 It is only possible to see a streak in the DP when the specimen is thin. Explain.
Q17.17 Does short-range ordering give rise to effects in the DP?
Q17.18 In Figure 17.5, what determines the distance between spots M and N?
Q17.19 When viewing a twist boundary nearly flat on, we look at the DP and see that there are three spots at $-2\mathbf{g}$ and two spots at $+\mathbf{g}$? Why is there an extra spot at $-2\mathbf{g}$?
Q17.20 Why is the $2\mathbf{g}$ reflection inside the square but $4\mathbf{g}$ reflection outside for the DP in Figure 17.3C?
Q17.21 For a wedge specimen, we say we have two sets of relrods because there are two surfaces. How does this fit with our idea of relrods arising from thin plates?
Q17.22 Can you see an effect in the DP if a twin boundary is viewed flat on?
Q17.23 Why do we see the (100) reflection in Figure 17.10?
Q17.24 How many objects are required in order to produce a detectable periodicity in the DP?

TEXT-SPECIFIC QUESTIONS

T17.1 Consider Figure 17.13 and the relationship of streak length to defect (interface) width. What is the width of the grain boundary according to the DP?
T17.2 Examine the extra spots in inset A of Figure 17.12. Compare the spacing of dislocations giving rise to these spots with the periodicity of the spots.
T17.3 Explain why there are spots in inset B of Figure 17.12.
T17.4 Consider Figure 17.15. How would this figure change if the top half of the figure (above the planar defect) were removed?
T17.5 Consider Figure 17.14B. Construct the DP you expect to see, assuming it's Si.
T17.6 Consider Figure 17.7. If this DP is from an 001 twist boundary, what is the angle of misorientation?
T17.7 Consider Figure 17.7. How can the lower reflection be the brighter in both insets?
T17.8 Consider Figure 17.3. Redraw the figure with the Ewald sphere cutting through (−2, −2) and (0, −2).
T17.9 Fully index the DP in Figure 17.10. Explain why this material must be BCC and not FCC.
T17.10 Why is the streak seen where it is in Figure 17.8?
T17.11 Figure 17.8A shows two grains. Which spots in Figure 17.8B correspond to which grains?
T17.12 Look at Figure 17.14. Is the spacing of the fine spots in the inset consistent with the image? Justify your answer.
T17.13 Consider Figure 17.14. Why are there several spots in a row rather than just one extra spot?
T17.14 Why do the streaks in Figure 17.13 shift parallel to their length as we look from one spot to the other in the insert?
T17.15 A thin Al metal foil contains needle-shaped precipitates which are 2 nm in diameter and 20 nm long. The needle lies in the (001) plane and its axis is parallel to the [110] direction. Sketch the [001] DP of the foil. (Courtesy ZL Wang.)

18

Obtaining and Indexing Parallel-Beam Diffraction Patterns

CHAPTER PREVIEW

The core strength of TEM is that you can obtain both a DP and an image from the same part of your specimen (not to mention various spectra). To obtain the crystallographic data, a method for interpreting and indexing the DP is essential and this aspect is the theme for the next four chapters. We'll start in this chapter by considering classic selected-area diffraction (SAD) patterns (SADPs) and how to index them, but also introduce other related, if less widely used, parallel-beam diffraction methods.

You can proceed to index your pattern in several ways, depending on how much information you already know about your specimen. We will begin the chapter by considering the experimental approach with the aim of being able to identify shortcuts whenever possible. The experienced microscopist will readily identify many patterns just by looking at them, but will still need to index new patterns or to identify unfamiliar ones. The fastest and most efficient experimental approach may take advantage of several concepts covered in the preceding two chapters and the following three. Today, sophisticated computer software is available which takes much of the tedium out of the indexing process.

Most DPs in a TEM are single-crystal (spot) patterns because the area from which we can routinely obtain SADPs (< 1 μm) is smaller than typical grain sizes of most engineering materials. However, with increased emphasis on nanocrystalline materials (grain size < 0.1 μm), it is more usual for the DPs to contain contributions from many crystals and so ring/textured SADPs are increasingly common. (We'll cover convergent-beam (CBED) and other forms of micro/nanodiffraction patterns, which can come from regions < 10 nm, in Chapters 20 and 21.) Using the DP, we can identify the crystal (which we often already know) and its orientation (which we probably don't) with respect to both the beam and to any adjacent crystals. The positions of the allowed *hkl* reflections are characteristic of the crystal system. Indexing associates each spot or ring in the DP with a plane (*hkl*), or set of planes {*hkl*}, in the crystal. From the indexing of the spots, you can deduce the orientation of the crystal in terms of the zone axis [*UVW*] in which the indexed planes lie.

BEAM DIRECTION

It is convention to define [*UVW*] as the beam direction. This direction is *normal* to the plane of the DP and *anti-parallel* to the electron beam.

If you want to know the orientation relationship between two crystals, you need to know more than one [*UVW*] for each crystal and, as we'll show, it is the determination of orientation relationships between different phases or differently oriented crystals that is the most useful information to come out of DP indexing. Orientation determination in the TEM is so important that we devote a complete chapter to the topic in the companion text. Now, computer control of both DP formation and DP indexing is very much the norm, but if you don't understand the principles, you shouldn't believe the computer output (GIGO). So we deliberately confine our discussion of computer-assisted indexing to the end of this chapter.

18.1 CHOOSING YOUR TECHNIQUE

The technique you choose to study your specimen will depend on what you *want* to learn and what you *can* learn. For example, if you want to learn about the crystal structure of a particular region, diffraction is generally the way to go although you may find moiré fringes (Chapter 23) or HRTEM (Chapter 28) more appropriate. Basically there are three diffraction approaches:

- You can spread the beam to give nearly parallel illumination and then use an aperture to select an area in the first image formed by the objective lens (giving the SADPs of Chapters 9 and 11). We emphasize SAD indexing in this chapter.
- SADPs often contain Kikuchi lines which give more accurate orientation determination (see Chapter 19).
- You can focus the beam on a small area of your specimen to form a CBED pattern (we won't use CBDP) (see Chapters 20 and 21).

SAD ACRONYMS

SADP versus SAD pattern, SAED versus SAD and CBEDP versus CBDP versus CBED pattern. We try to use DP throughout the book but some conventions are very well established!

We can also summarize the possibilities as a function of the grain size of the material. Let's consider this particular specimen characteristic

- The grain size may be very small, $<<10$ nm, typical of many nanocrystalline thin films. This is a problem because it's rare that a single grain will extend through the thickness of your specimen, in which case you can't easily index an individual crystal pattern. (This is the rhino problem (Figure 1.7) for DPs!) However, in this case you probably won't want to know the *orientation* of a particular grain but will instead be interested in knowing the *texture* of the material.
- The grain size is between 10 and ~100 nm. Here CBED may be useful because it gives you a small probe. However, much of the benefit of CBED comes from having specimens which are >100 nm thick; the best thickness depends on the structure factor (atomic number) of your specimen. If you're careful and are using a recent-model TEM in which C_s and λ are small, you might be able to use SAD in this range of grain size, as we saw in Table 11.1.
- The grain size is in the range 100 nm to ~2 μm. In this situation, SAD can be used quite routinely in a modern TEM. You must be aware of the limitations and be prepared to unravel a complex DP. Because of errors due to C_s and Δf, the problem will be distinguishing which spots arise from the area you selected and which spots arise from neighboring areas.
- The specimen is uniformly thin with grain size > 2 μm. This type of specimen is just a simpler version of the last case. You should have no problem in applying SAD techniques even at lower voltages and in older microscopes. Now CBED will be very useful in examining local changes *within* a grain.
- The grains of interest are large (> 2 μm, even better if they are > 5 μm) with both thin areas (< 100–300 nm thick, depending on the material) and areas which are sufficiently thick for Kikuchi lines to be visible (see next chapter). Now you can use any of these techniques, except texture analysis, which becomes more difficult! For the latter, you should now consider the electron-backscatter diffraction (EBSD) technique using a bulk specimen in an SEM to give better statistics (Schwartz et al.).

In this chapter, we'll concentrate on the hands-on approach to SAD analysis and leave CBED to Chapters 20 and 21. We'll also introduce hollow-cone and precession diffraction which make your SADPs more useful at times. We can't give you a foolproof guide since the best technique(s) will depend on your specimen.

18.2 EXPERIMENTAL SAD TECHNIQUES

By now you should know how the experimental camera length (L) compares to the value you read from the microscope. You also know how the SADP is rotated with respect to the image as the magnification changes (unless your particular TEM automatically compensates for this rotation). You've checked that you haven't missed a 180° inversion; leading researchers have missed this in the past. Go back to Sections 9.3–9.6 and 11.9 if you need to refresh the details of the practical steps involved in obtaining SADPs.

You can vary L but your pattern may rotate as you do so. We generally use a value of ~500 mm for SAD, but that will depend on your TEM, whether you want to see detail in the HOLZ and on the interplanar spacings in your specimen. It's good practice to choose a particular value of L and always use that value for your SADPs with a particular instrument/specimen combination. You may want to increase L for special high-resolution diffraction, but you'll give up a large number of other reflections and enlarging the photographic film will almost always provide the magnification you need. This is where a wide-angle CCD camera (Chapter 7) that captures a much larger area of reciprocal space than standard TEM film can be really useful.

Tilting and rotating your specimen. One of the biggest assets of the TEM is that you can monitor the DP as you tilt or rotate your specimen. Rotating the pattern requires a rotation holder (Chapter 8), which is ideal if you want to align a particular reflection parallel to the tilt axis, especially for a side-entry holder. This alignment is particularly helpful in stereomicroscopy (see Chapter 29). Tilting the specimen is far more common than rotating, since all side-entry holders automatically have one tilt axis parallel to the specimen rod. We discussed the importance of eucentricity in Chapter 9.

TILTING
Tilting the specimen changes the diffraction conditions and may change the focus.

It is good practice to note the tilt settings whenever you are recording images. If you want to use these settings to give a rough estimate of how far you're tilting the specimen, you should remember that there may be some backlash due to mechanical hysteresis. So you will always need to approach a particular setting from the same tilt direction if you need to be exact. In the next chapter, we'll describe how we use Kikuchi maps to guide us as we tilt the specimen. If you don't have Kikuchi lines because your specimen is too thin, or too bent, you can still use the idea. Select a particular strongly diffracted beam and then tilt the specimen so that that particular beam remains excited. What you are doing is tilting the specimen so that the same plane remains nearly parallel to the electron beam (think about what this means and the crystallographic information that it can give you).

Tilting the beam. If you are really interested in examining the detail present in the DP and the image is less important, you can change the diffraction conditions in a very controlled and reversible way by tilting the beam using the DF deflection coils. You can be much more precise than mechanical tilting with the goniometer since there is no problem with backlash. To increase your accuracy, you may want to increase L. This technique is particularly helpful when you want to examine the effect of small changes in $\mathbf{s}$ on the appearance of diffraction spots.

For example, if you want to excite the third-order reflection 3**g** in BF, you could use the approach shown in Figure 18.1. (You wouldn't, but this exercise is useful!)

- Use the beam tilt (dark-field deflection coils) to put $\bar{\mathbf{g}}$ on the optic axis (where **0** was) (Figure 18.1A). Now tilt the sample so that **g** is excited. Then tilt the beam to put **0** back on the optic axis. 3**g** is now excited.

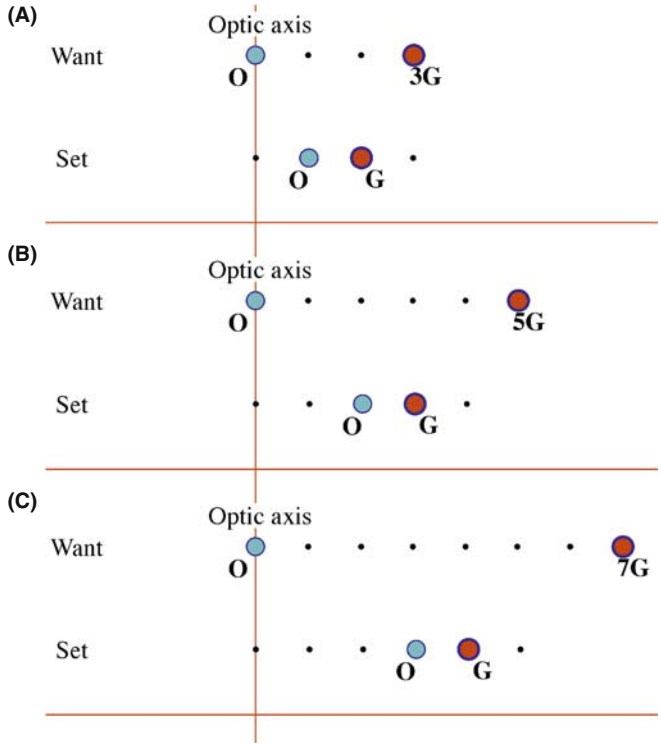


FIGURE 18.1. Exercises in exciting high-order reflections. (A) To excite 3G in BF, tilt the beam so –G is on axis with G strongly excited, then tilt O back onto the optic axis. (B) To excite 5G in BF, tilt the beam so –2G is on axis with G strongly excited, then tilt O back onto the optic axis. (C) To excite 7G in BF, tilt the beam so −3G is on axis with G strongly excited, then tilt O back onto the optic axis.

Next, repeat the exercise but with the aim of exciting 5**g**.

- Use the beam tilt (dark-field deflection coils) to put $2\bar{\mathbf{g}}$ on the optic axis (where **0** was) (Figure 18.1B). Now tilt the sample so that **g** is excited (as before). Then tilt the beam to put **0** back on the optic axis. 5**g** is now excited.

Repeat the exercise again but with the aim of exciting 7**g**.

- Use the beam tilt (dark-field deflection coils) to put $3\bar{\mathbf{g}}$ on the optic axis (where **0** was) (Figure 18.1C). Now tilt the sample so that **g** is excited (as before). Then tilt the beam to put **0** back on the optic axis. 7**g** is now excited.

Now you can see that if you want to excite 11**g**, but you won't be able to see 11**g** on the DP, start by moving $5\bar{\mathbf{g}}$ to the optic axis. You should now appreciate the possibilities.

We'll develop other variations of this technique in Chapter 19, and we'll see in Chapter 27 that the situation in Figure 18.1 does arise in weak-beam microscopy at higher voltages. Computerized control of the beam tilt is essential to hollow-cone and precession diffraction methods which we'll discuss toward the end of the chapter.

18.3 THE STEREOGRAPHIC PROJECTION

DPs not only tell us the direction of the electron beam but also the complete orientation of that region of the specimen illuminated by the beam. If we have a grain boundary or interphase interface (or indeed any planar defect) present in the specimen, we can determine the orientation of both grains and the plane of the interface. As we show in great detail in the companion text, what we often want to know is how the two crystals are related to one another. But first, we need a method for visualizing this relationship; this is where the stereographic projection or stereogram is an invaluable aid. Unfortunately the classic materials texts by Johari and Thomas and by Smaill are out of print, so to get an in-depth appreciation of this technique, you'll have to resort to crystallography texts, such as those listed in the general reference section or a dedicated stereographic-projection text in another field (e.g., Lisle and Leyshon). Like other tools, you'll have to understand it and use it before you fully appreciate its value. We strongly recommend that you take time out to do this if you're not already familiar with the construction. Any introductory crystallography text is a good place to start and several are listed in the references.

The construction. Imagine a crystal located inside a sphere as shown in Figure 18.2. Draw a line normal to each crystal plane from the center of the sphere (the sphere of projection) to intersect the sphere at point P in the northern hemisphere; the cross section view may be easier to visualize. Now draw a second line from the south pole to point P.

This second line cuts the equatorial plane at the point P′. The disk of the equatorial plane is the stereographic projection and the point P′ *uniquely* represents the plane whose radial normal cuts through P. If P is in the southern hemisphere, we draw the line from the north pole instead and identify this P′ on the projection with a circle instead of a dot.

Look again at the crystal; it's cubic to keep it simple but the construction is completely general. The normals to the planes (100), (111), (011) and $(\bar{1}11)$ all lie on the circumference of a circle around the sphere of projection. In this special case, all of the points on this circumference project onto the same great circle on the stereographic projection whose circumference we call the 'primitive' great circle. The Wulff net in Figure 18.3 shows 90 such great circles all passing from the north pole to the south pole and another around the equator: a great circle always passes through opposite ends of a diameter in the projection. These are the familiar lines of longitude on the globe. Circles on the sphere which do not contain the center of the sphere are smaller; they also project as circles called the small

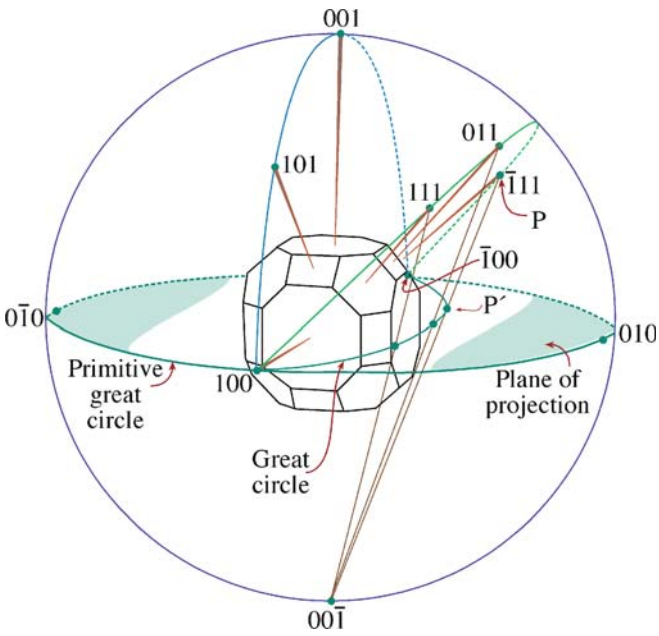


FIGURE 18.2. The stereographic projection. The crystal is at the center of the sphere. Normals to the crystal planes are projected until they intercept the sphere at P, then projected back to the south pole $(00\bar{1})$ of the sphere. Where this projected line crosses the equatorial plane at P' is the point that uniquely represents the original plane in the crystal. Note that planes in the same zone on the crystal project as a great circle in the stereographic projection. An example of such a great circle is its diameter (or line of longitude); another is the circumference in the equatorial plane.

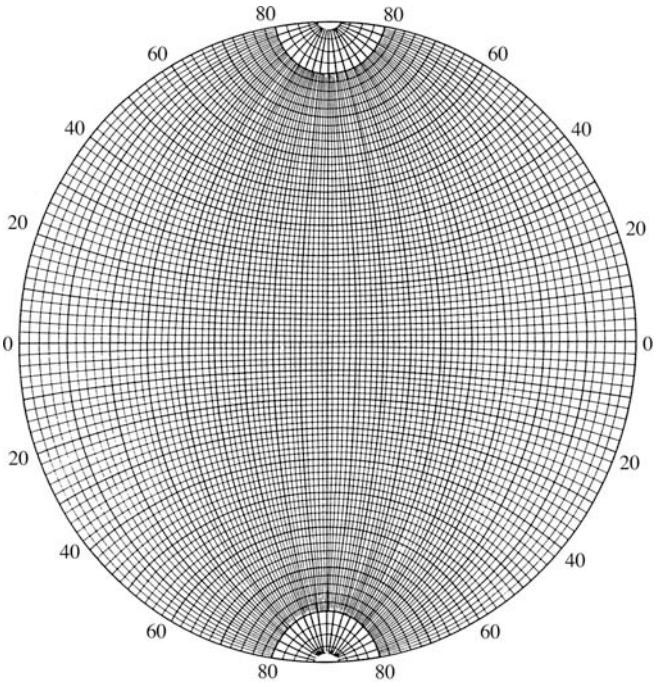


FIGURE 18.3. A Wulff net which contains 90 great circles like the one in Figure 18.2; each great circle is 2° apart so the net covers 180°. The only great circle that actually appears as a circle in the net is the circumference of the projection, called the primitive great circle. Points on the primitive represent planes whose normals are 90° from the north pole (which projects in the center of the Wulff net). Thus all distances on the net are proportional to angles in real space but only correspond exactly to angles around the primitive.

circles which, if concentric with the primitive, are familiar as lines of latitude. (Note, however, that most small circles are not concentric with the primitive.) We can then rotate the Wulff net, as we wish, to realign our great circles.

- We can represent plane normals (also called poles) and directions on the same projection even if they are not parallel to one another. Better still, we can read off the angles between them. Remember that, in general, the normal to the plane (*hkl*) is parallel to the direction [*hkl*] only for cubic materials.
- The zone axis is always 90° away from any plane normal that is in its zone. All the plane normals in a particular zone, [*UVW*], will lie on a single great circle.
- The angle between any two planes is the angle between their plane normals, measured along a great circle using the Wulff net.
- We can use the same construction to summarize all the symmetry elements of any particular crystal system.

These poles at the *UVW* zone axis represent the *possible diffracting planes* for that zone; [*UVW*] is the beam direction. So if [*UVW*] is in the center of the projection, the *hkl* reflections will be around the circumference of the projection (the primitive great circle). Now you should appreciate why stereograms can be so useful in interpreting DPs.

Several examples of stereographic projections are shown in Figure 18.4. Look at the Wulff net and check some simple facts. For example, for the cubic system, check which poles are 90° away from the [001] direction. How large is the angle between $(0\bar{1}1)$ and (011)? How would this angle change if the material were forced to be tetragonal with $c/a > 1$? What happens in this case to the $(1\bar{1}1)$ pole or the $(1\bar{1}0)$ pole? Now consider the more extensive plot shown in Figure 18.5. If the specimen is cubic with the [001] foil normal, what pole would you tilt to if you wanted to form an image with the $0\bar{2}2$ reflection? (One answer is the [011] zone axis, but why?). For the same specimen, if you want to excite the $\bar{1}11$ reflection, you could tilt toward the $[0\bar{1}1]$ zone axis keeping the 200 reflection excited, not toward the [011] zone axis.

- You could work this out using equations, but the stereographic projection tells you what to do while you are sitting at the microscope.
- If you are working with a non-cubic material, buy a large Wulff net and construct your own stereographic projection; you can buy standard projections for cubic materials so, as usual, it's easier to be a metallurgist than a ceramist or mineralogist.
- Use a program like EMS to help you plot the points, or download appropriate software from the Web (e.g., URL #1).
- Wulff nets and plotting software are also available on-line (e.g., URL #2).

18.4 INDEXING SINGLE-CRYSTAL DPs

Remember the fundamental relationship in a DP (Section 9.6.B).

$$Rd = \lambda L \tag{18.1}$$

Any distance between the direct beam and a specific diffraction spot or the radius of a diffraction ring, *R*, which we measure on the DP, is related to a specific spacing between planes in the crystal, *d*. Since λL is a constant, we can measure several values of *R* and know that

$$R_1 d_1 = R_2 d_2 = R_3 d_3 = R_4 d_4 = \ldots \tag{18.2}$$

If you know the lattice parameter of your crystal, then you know the allowed reflections and only certain *d*-spacings will be associated with diffraction spots.

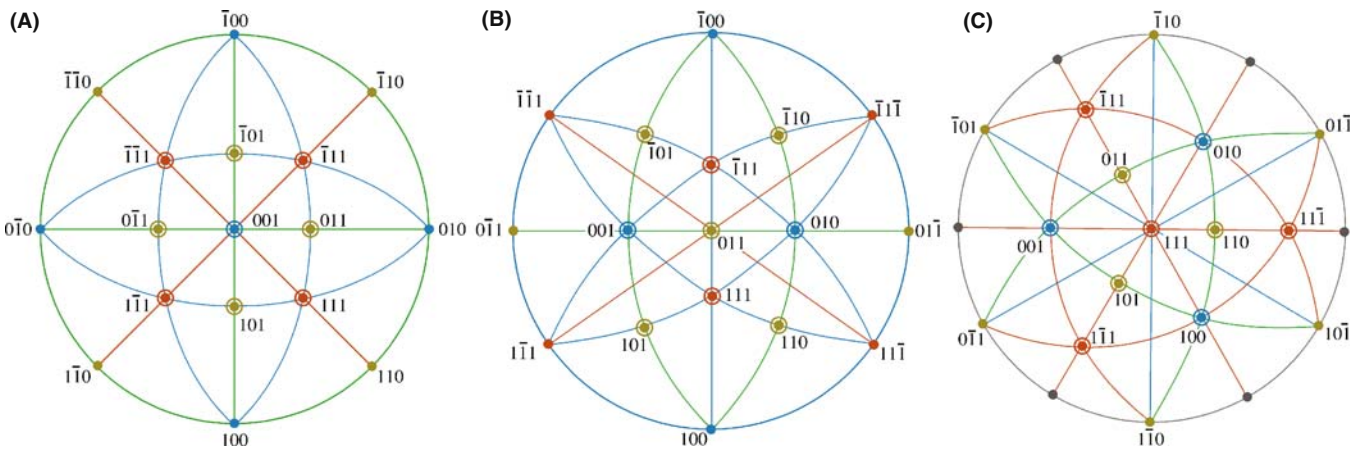


FIGURE 18.4. Some standard cubic stereographic projections. The pole in the center defines each projection, so these are 001, 011 and 111.

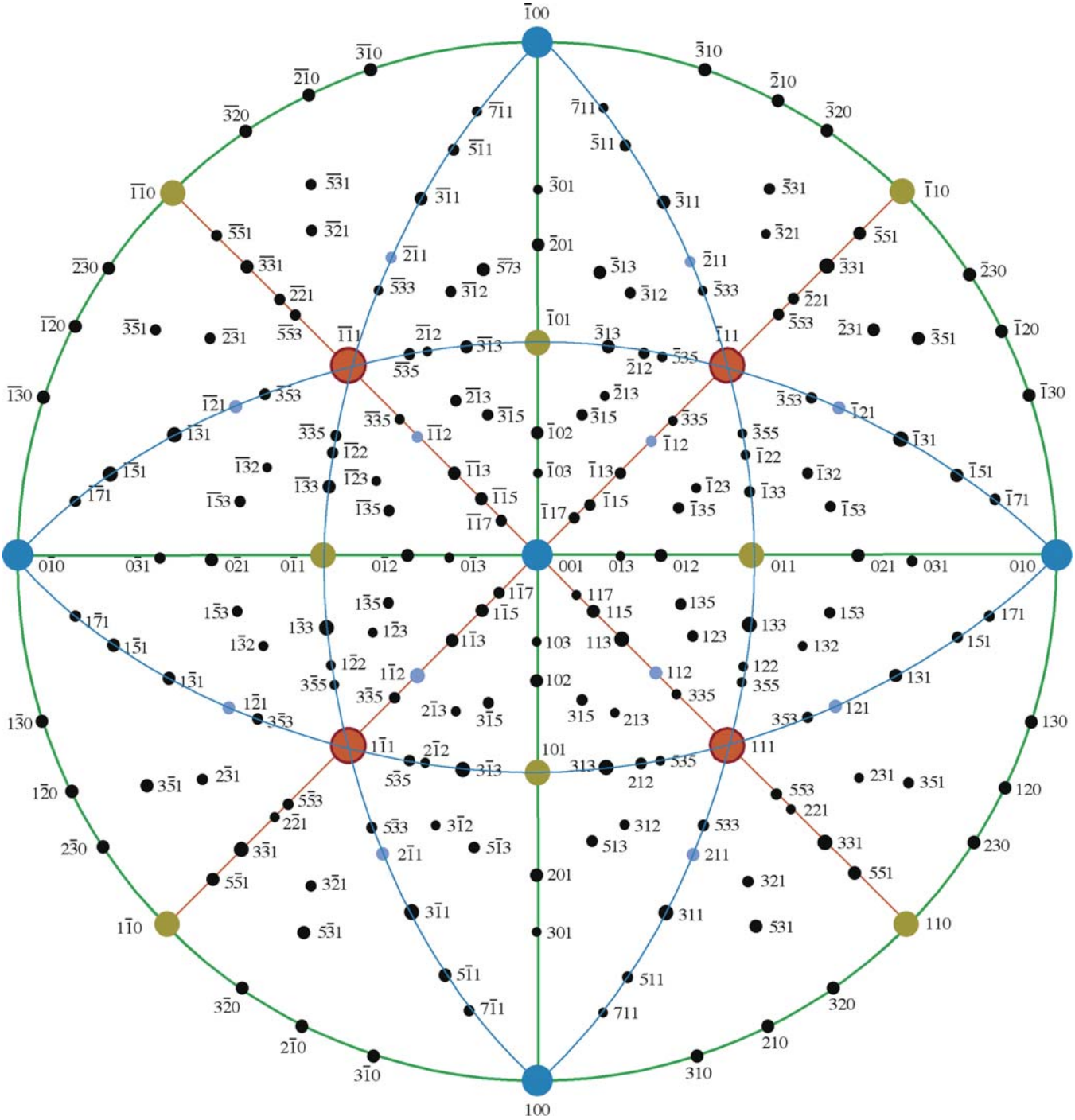


FIGURE 18.5. The stereographic projection for a cubic foil with a [001] normal, assuming the beam is down [001] also. If you want to form an image with the $0\bar{2}2$ reflection, you need to tilt the specimen so the $0\bar{1}1$ pole rotates until it is on the primitive, i.e., it is 90° from the beam direction. To do this, you need to tilt about a pole that is 90° from the $0\bar{2}2$ reflection, such as the [100], [111] or [311] zone axes.

Table 18.1 lists allowed and forbidden reflections for some cubic systems. Rules for more crystal systems are given in Table 16.2.

R AND *d*-SPACINGS

The ratio of any two *R* values gives the inverse ratio of the *d*-spacings.

Once you have tentatively identified possible values for $\mathbf{g}_1$ and $\mathbf{g}_2$, you need to cross-check your answers using the angles between the $\mathbf{g}$ vectors (i.e., the angles between the plane normals). The fully indexed patterns at the end of this chapter (Figures 18.19–18.21) show the principal interplanar angles and the principal ratios of $\mathbf{g}_1/\mathbf{g}_2$. Hence, in practice, you will rarely have to measure more than two or three spacings in order to index a

TABLE 18.1. The Selection Rules for Cubic Crystal Structures

bcc		fcc		Diamond cubic	
$h^2 + k^2 + l^2$	hkl	$h^2 + k^2 + l^2$	hkl	$h^2 + k^2 + l^2$	hkl
2	110				
		3	111	3	111
4	200	4	200	4	200
6	211				
8	220	8	220	8	220
10	310				
		11	331	11	331
12	222	12	222		
14	321				
16	400	16	400	16	400
18	411				
	330				
		19	331	19	331
20	420	20	420		
22	332				
24	422	24	422	24	422
26	431				
		27	511	27	511
		27	333	27	333
30	521				
32	440	32	440	32	440

particular zone-axis DP. However, if your specimen is not oriented close to a zone axis, you'll need to look ahead to Section 18.10.

WEISS ZONE LAW

Check the consistency of your indexing using the Weiss zone law. Each hkl reflection must lie in the $[UVW]$ zone, i.e., $hU + kV + lW = 0$.

The Weiss zone law only holds for diffraction into the zeroth layer of the reciprocal lattice and the diffraction spots that we see here are called the zero-order Laue zone (ZOLZ). There are circumstances when we can see diffraction from higher-order Laue zones (HOLZ) and we'll talk a lot more about this phenomenon in Chapters 20 and 21. All SADPs, particularly those that we talk about and index in this and the next chapter, are ZOLZ patterns.

The angle between normals to the planes $(h_1k_1l_1)$ and $(h_2k_2l_2)$ is ϕ; the angle between directions $[U_1V_1W_1]$ and $[U_2V_2W_2]$ is ρ. You can work these out and cross-check them with your DPs. These are standard equations in many texts (e.g., those by Edington and Andrews et al.). You'll probably find that the equations for the cubic system are the most useful.

$$\cos\phi = \frac{h_1h_2 + k_1k_2 + l_1l_2}{\left(h_1^2 + k_1^2 + l_1^2\right)^{1/2}\left(h_2^2 + k_2^2 + l_2^2\right)^{1/2}} \qquad (18.3)$$

$$\cos\rho = \frac{U_1U_2 + V_1V_2 + W_1W_2}{\left(U_1^2 + V_1^2 + W_1^2\right)^{1/2}\left(U_2^2 + V_2^2 + W_2^2\right)^{1/2}} \qquad (18.4)$$

Remember that you can always work out these expressions for any crystal system using the equation for the dot product of the two appropriate vectors.

In principle, if we don't know the crystal structure, we can still work out the d-spacings of the diffracting planes using equation 18.1. However, you should remember that SAD is not the most accurate method for determining the spacing of lattice planes, d_{hkl}, or the angles between them, ϕ. SAD is generally very good at distinguishing patterns, but it completely fails when the difference between the two patterns is a 180° rotation, as occurs in some patterns of polar material, like GaAs or GaN.

To summarize

- Tilt your specimen to a low-index pole.
- Set $\mathbf{s} = 0$ for the innermost reflections.
- Record the SADP.
- Repeat the exercise using higher-order reflections after tilting the specimen to set $\mathbf{s} = 0$. These measurements will be more accurate, but only if you make sure that $\mathbf{s} = 0$.

The discussion on relrods in Chapter 16 told you that both d and ϕ could be seriously in error if reflections are not set to have $\mathbf{s} = 0$, especially since you've probably tilted the specimen.

So far, you have only indexed one DP. You'll probably need more than one to determine orientation relationships. While you're at the microscope, tilt to pole #2 keeping $\mathbf{g}_1$ (see Figure 18.6) strongly excited. Repeat the

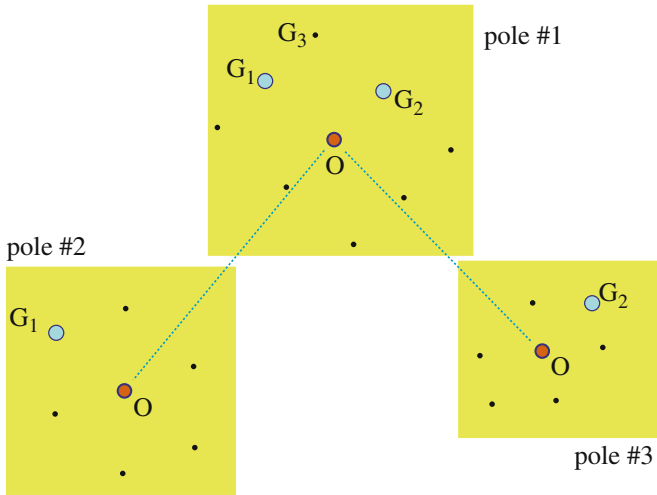


FIGURE 18.6. How to confirm your indexing of reflections and poles by tilting to other poles. Start with $\mathbf{g}_1$ and $\mathbf{g}_2$ strongly excited at pole #1. Tilt to pole #2 keeping $\mathbf{g}_1$ strong, then go back to pole #1 and tilt to pole #3, keeping $\mathbf{g}_2$ strong. Index all the strong reflections each time, measure the tilt angles between each reflection and estimate the tilt between poles.

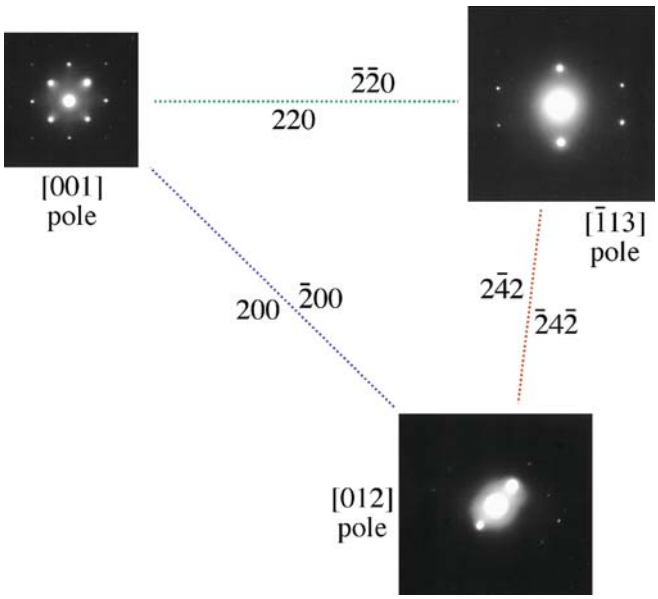


FIGURE 18.7. A practical illustration of the procedure described in Figure 18.6 for an fcc material, in this case MgO.

indexing procedure. Go back to pole #1 and tilt to pole #3 keeping $\mathbf{g}_2$ strongly excited. You can repeat this indexing as many times as you wish. The important idea is that you now have angular measurements allowing you to cross-check your determination of both $\mathbf{g}_1$ and $\mathbf{g}_2$ and the zone axes. Of course the task is simple for an fcc crystal, as you can see in Figure 18.7, which is an experimental illustration of this procedure. The challenge comes when the crystal has less symmetry. If you already know the crystal structure, then you should plot out the most important poles, relating their orientations to one another (more on this in Section 18.11 and Chapter 19) and pay particular attention to the information from systematic absences which occur when the structure factor is zero (go back and check Section 3.9 and Chapter 13).

THE GOLDEN RULE

Make the task as easy as possible. Record *ALL* the DPs you might need and how they relate to one another (draw road maps) while you're at the microscope.

18.5 RING PATTERNS FROM POLYCRYSTALLINE MATERIALS

Diffraction from polycrystalline specimens (especially when the grain sizes are at the nanoscale) can be viewed in much the same way as X-ray diffraction from powders. For a completely random polycrystal, we rotate the reciprocal lattice about all axes and produce a set of nested spheres. When we intersect these spheres with the Ewald sphere (which, in the TEM, approximates to a plane) we will see the rings which are recorded in powder patterns.

NANO AND DPs

Smaller grains give broader spots in the DP. Nanocrystals should give the largest spots, or. . .

If a polycrystal is textured, then there is usually one special plane which is common to nearly all the grains. Since the grains are small, all the reciprocal-lattice points will be broadened by the shape effect; so will the sphere or circles for the polycrystals. This situation will be exacerbated for nanocrystals.

If we then rotate the reciprocal lattice about the lattice vector normal to the texture plane, we produce a set of circles in reciprocal space, as shown in Figure 18.8. If we are examining cubic materials, the reciprocal-lattice vector $\mathbf{g}_{hkl}$ will be parallel to the direction $[hkl]$ in real space. Otherwise this will not generally be the case.

The DP in either of these examples appears as shown in Figure 18.9A and B which differs because the grain

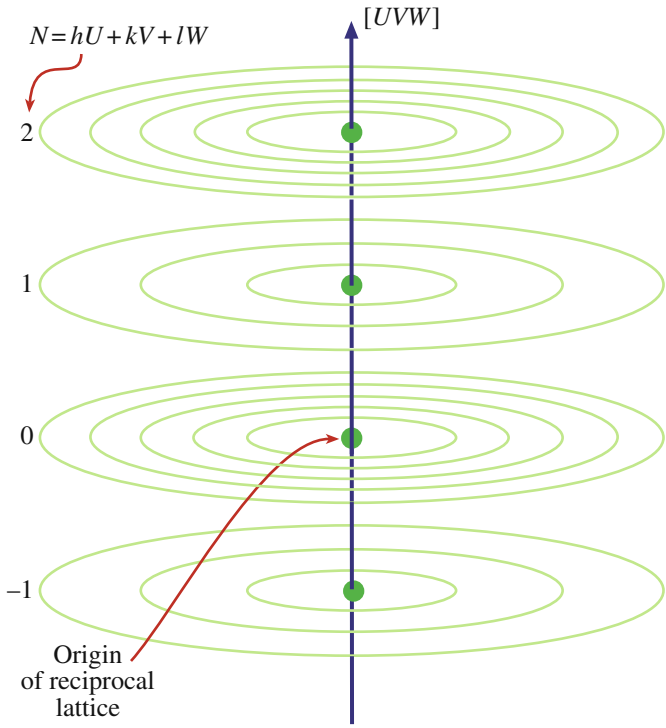
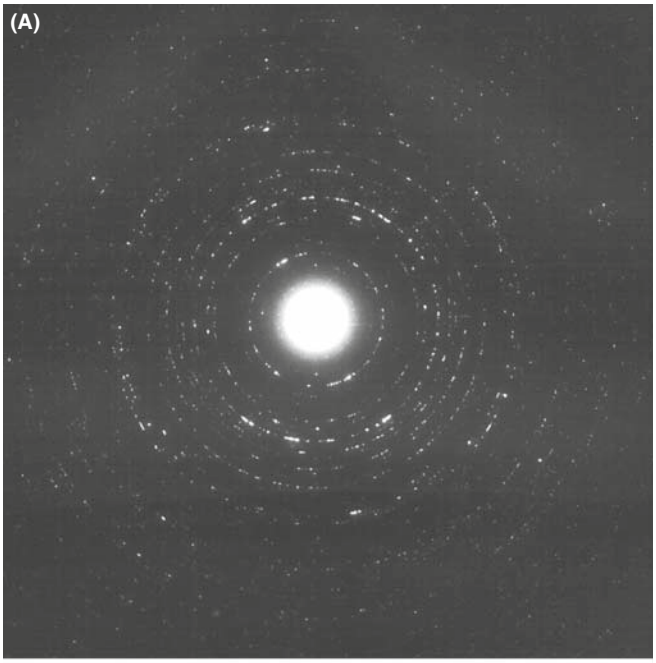


FIGURE 18.8. The generation of a set of circles in reciprocal space by a textured polycrystal. When the reciprocal lattice is rotated about a particular direction $[UVW]$ (in this case the normal to the texture plane) each Laue zone ($N = 1$, 2, etc.) produces a set of concentric circles for each allowed reflection in each zone.



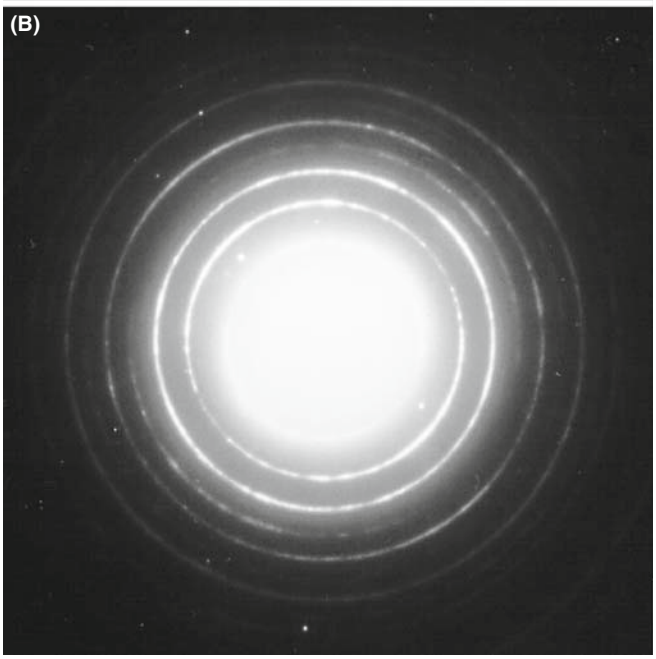


FIGURE 18.9. (A) and (B) Ring DPs from polycrystalline foils. In (A) the grain size is larger than in (B) so the rings are made up of discrete spots. A finer grain size, as in (B) produces a more continuous ring pattern, but the width of the rings of diffracted intensity in fact becomes broader and can be used as an inverse measure of the grain size.

size is different. A larger grain size gives a more speckled pattern.

You can distinguish the pattern produced by a textured specimen from one produced by a random polycrystal by tilting. If your specimen is textured, the rings become arcs as shown in the pattern in Figure 18.10A together with the Ewald-sphere construction in Figure 18.10B. You can locate the grains which give rise to the arcs by forming a CDF image with the arc of diffracted intensity. In Figure 18.10C, these oriented grains are uniformly distributed, but you might encounter a situation where this is not the case. Think how the pattern might differ in this case.

Figure 18.10D and E emphasizes that these ring patterns can be quite varied. In this case the specimen is α-Ag_2Se, which is textured about an axis *inclined* to the beam. When the Ewald sphere cuts the circles now, it produces elongated spots which lie on an ellipse. Vainshtein et al. point out that all the 'spots' on one ellipse can be indexed with the same *hk* indices but a different *l*, and call such a pattern an oblique-textured electron DP. You should also be careful in indexing these textured patterns since not all possible d_{hkl} values need be present, depending on the texture plane.

There is more information in ring patterns than spot patterns. Like a powder pattern, you could estimate the grain size from the width of the rings, but it's more direct (and more reliable) to just look at the DF image. You can see kinematically forbidden rings because you don't necessarily have single scattering from each grain but there are ways around this problem.

Nanocrystalline materials, which fall into our smallest range of ~10-nm grain size, are very challenging because there's probably always more than one grain through the foil thickness. Careful DF imaging combined with HRTEM or CBED with the smallest possible probe (often called nanodiffraction, see Section 21.8.B), is probably optimal, but you need to look for clustering of similarly oriented grains.

18.6 RING PATTERNS FROM HOLLOW-CONE DIFFRACTION

We can combine the advantages of ring patterns from small-grained materials with individual spot patterns from larger-grained materials, As we noted back in Section 9.3.D, if you have a nano/microcrystalline specimen, it is rather inefficient to get a DP from one grain at a time by carefully tilting your specimen and then getting individual DF images of single areas, or of phases within a single grain, by carefully tilting the incident beam. So we can use a computer to continuously change the beam orientation so it samples many angles of incidence and gather DPs from all the crystals superimposed on one another (i.e., get a ring pattern, even from a relatively large-grain material). This is the principle behind hollow-cone (sometimes called 'conical') diffraction (go back and look at Figure 9.15). In the same section we also introduced the idea of hollow-cone DF imaging. In earlier TEMs, before computer control of the beam, hollow-cone illumination was achieved by using an annular C2 aperture rather than a circular hole. Thus the beam that came through the annulus irradiated the specimen from a fixed angle

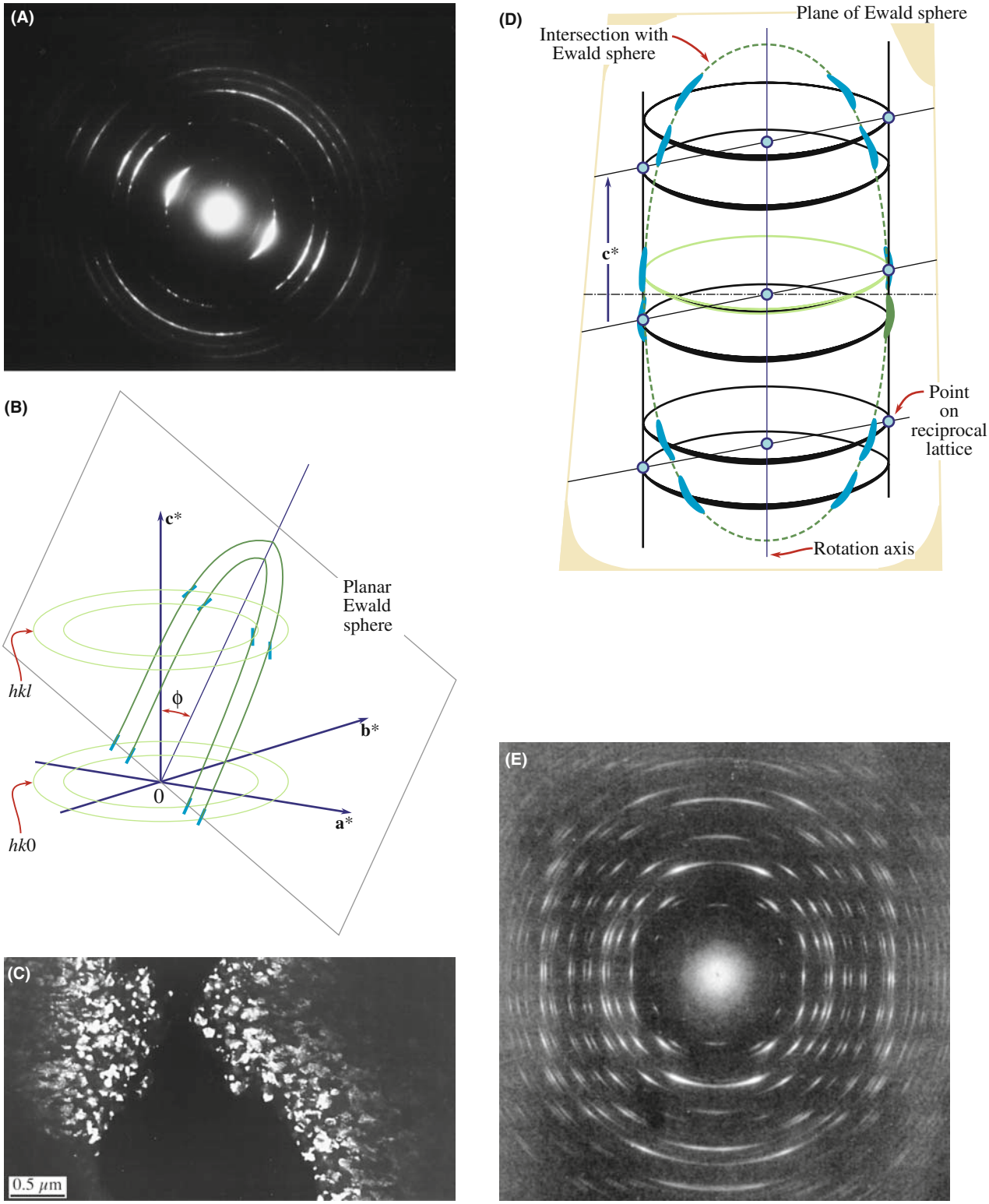


FIGURE 18.10. (A) A textured ring pattern where the rings are more intense over a certain angular range. (B) The corresponding interception of the Ewald sphere (plane) with the reciprocal lattice. (C) A DF image of the textured grains, taken from a brighter potion of one of the *hkl* rings, showing an equiaxed structure. In (D) the specimen is textured about a direction at an angle to the beam, so the Ewald sphere creates elongated spots or arcs in the DP (E).

around the optic axis, as shown schematically back in Figure 9.15A. It's much easier and more flexible to use the computer to control the pre-specimen scan coils and bring the beam onto the specimen at a fixed (or a range of) conical scan angle and a typical conical-scan DP is shown in Figure 9.15B. (If you look ahead, you can see another example of a hollow-cone DF image from an amorphous DP in Figure 18.14C.) Of course, we index the rings in exactly the same way as we described in Section 18.5 for ring patterns obtained from a polycrystalline specimen where the grain size is much smaller than the beam size.

With hollow-cone illumination, we can create a DF image of *all* the crystals in a film that are diffracting from their {*hkl*} planes, rather than imaging only that small fraction diffracting into the part of the diffraction ring that happens to be selected by the aperture. The difference in information content between these two approaches is shown back in Figure 9.15B and C. The hollow-cone technique is widely used in CBED (see Chapter 21) and has more recently been used as the basis for precession diffraction (see Section 18.8.)

18.7 RING PATTERNS FROM AMORPHOUS MATERIALS

Amorphous materials used to represent somewhat of a fringe field in materials science but the rapid growth in glass technology over the last decade, fueled by the optical-communications bubble, the development of bulk metallic glasses and the increased awareness of the role of glassy films at interfaces (particularly the gate oxide in semiconductors) have all conspired to crystallize the following question: *is the material really amorphous or is it (sub) nanocrystalline?* Actually this question is still debated when discussing both amorphous materials and, more intensely, oxide and metallic glasses because there's still no broadly accepted definition of when grains become small enough that a determinable crystal structure no longer exists. At best we should think of a continuum of states from nanocrystalline to amorphous.

DPs AND AMORPHOUS MATERIALS
The DP from an amorphous material looks similar to that from polycrystalline material but the rings are broader and there is no speckle.

Rudee and Howie showed that electron scattering from regions of $\leq$1.5 nm diameter could be coherent. Graczyk and Chaudhari proposed modeling these materials as random networks. If we are careful, we can learn quite a lot about the structure of amorphous materials, but we should first say what we mean by 'amorphous.'

> An amorphous material is one where the locations of the neighboring atoms are defined by a probability function such that the probabilities are never unity.

This idea is best illustrated by a plot of the probability which we call the *radial distribution function* (RDF). The RDF, $\rho(r)$, is the probability, per unit element of volume, that an atom will be found at a distance r from another atom. The first example in Figure 18.11A compares the curves for liquid sodium and crystalline sodium; the numbers on the crystalline curve remind us that in the crystal each sodium has eight nearest neighbors, etc. The second plot, Figure 18.11B, shows the RDF for vitreous silica. This time the peaks are associated with distances

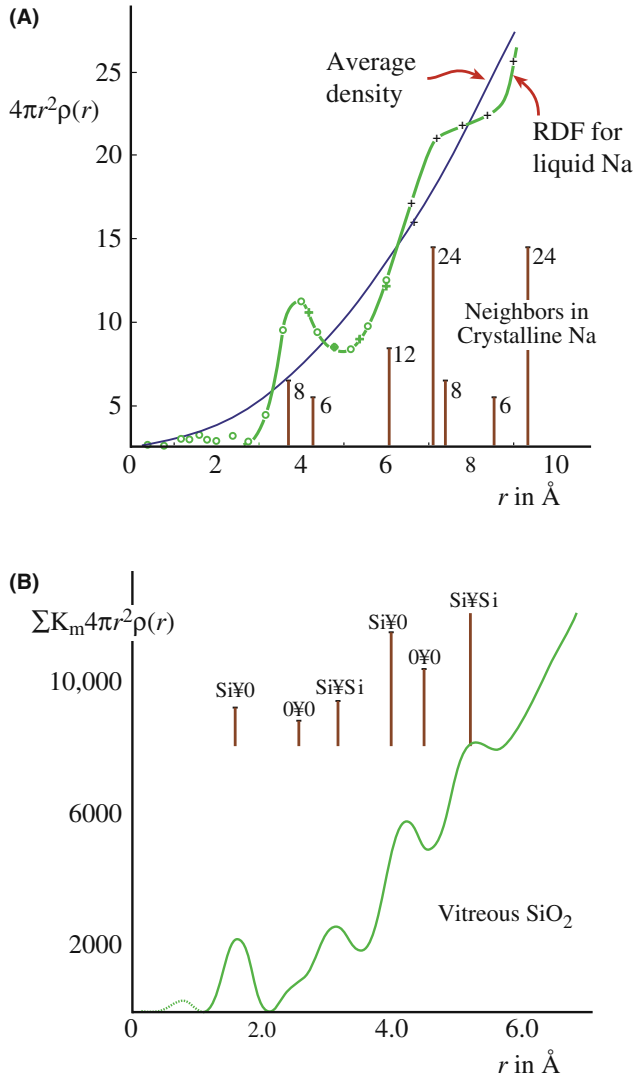


FIGURE 18.11. (A) Radial distribution functions for liquid Na and the average density curve superimposed on the distribution of the nearest neighbors in crystalline Na (vertical lines). (B) The RDF for vitreous SiO_2 is peaked at a distance that represents spacings between Si and O atoms.

between different pairs of Si and O atoms. The features to notice are

- The two curves both show definite peaks.
- The two curves are different.

Some diffraction theory. Since these materials are so different, we'll give a brief introduction to the theory of scattering from amorphous materials. We make the assumption that the electron beam is only scattered once; this is kinematical, but it's more realistic than for crystals at the Bragg condition. Following Howie, we express the kinematical intensity, $I(\mathbf{k})$, by the expression

$$I(\mathbf{k}) = |f(\mathbf{k})|^2 \sum_{i,j} e^{i2\pi\mathbf{k}.(\mathbf{r}_i - \mathbf{r}_j)} \tag{18.5}$$

Here we assume that there are N identical atoms contributing to the scattered intensity and they are located at the different positions ($\mathbf{r}_i$ or $\mathbf{r}_j$).

The $f(\mathbf{k})$ terms are the atomic scattering amplitudes, with $\mathbf{k}$ reminding us that there is an angular dependence to f. If the material is isotropic, we can simplify equation 18.5 as follows

$$I(\mathbf{k}) = N|f(\mathbf{k})|^2 \left(1 + \frac{F(\mathbf{k})}{k}\right) \tag{18.6}$$

where

$$F(\mathbf{k}) = \sum_{i \neq j} e^{i2\pi\mathbf{k}.(\mathbf{r}_i - \mathbf{r}_j)} \tag{18.7}$$

$$F(\mathbf{k}) = k \int \rho(r) e^{i2\pi\mathbf{k}.\mathbf{r}} dV \tag{18.8}$$

$$F(\mathbf{k}) = 4\pi \int_0^\infty \rho(r) \sin(2\pi k r) r \; dr \tag{18.9}$$

The term $\rho(\mathbf{r})$ is the RDF. Equation 18.9 can be inverted to give an expression for ρ

$$\rho(\mathbf{r}) - \rho_0 = \frac{1}{r} \int_0^\infty F(\mathbf{k}) \sin 2\pi k r \; dk \tag{18.10}$$

This equation means that, in principle, the RDF can be obtained directly from DPs. This process is enhanced if the patterns are energy-filtered to remove inelastic contributions (see Chapter 37 and the work of Cockayne et al.) as shown graphically in Figures 18.12 and 18.13. Compare these figures with Figure 2.13A. Other ways to determine the RDF include extended X-ray absorption fine structure (EXAFS) studies in a synchrotron X-ray and, less expensively but more noisily, via EXELFS (see Chapter 40). Alternatively, we can rearrange equation 18.6 again to give a 'reduced-intensity function' as illustrated by the work of Graczyk and Chaudhari who

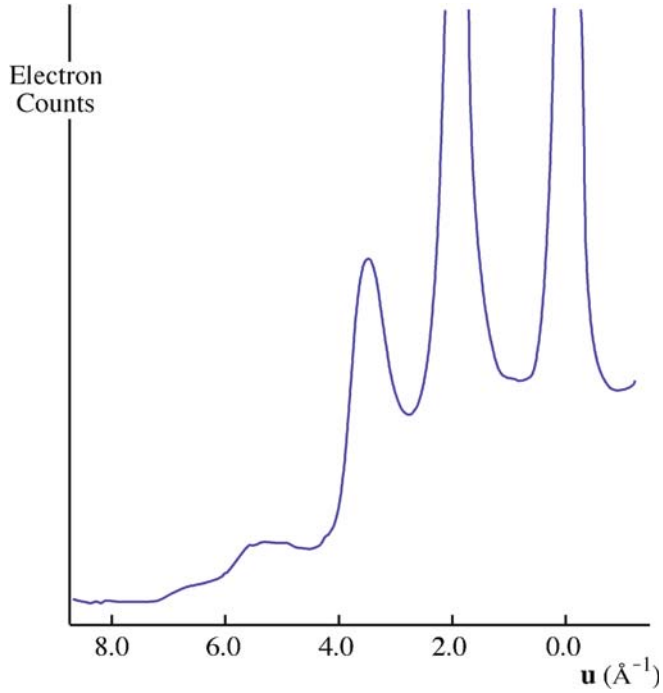


FIGURE 18.12. An intensity profile across an energy-filtered DP from amorphous Si obtained by scanning the pattern across the entrance slit to a serial EELS spectrometer and recording only the elastic (on-axis) electrons.

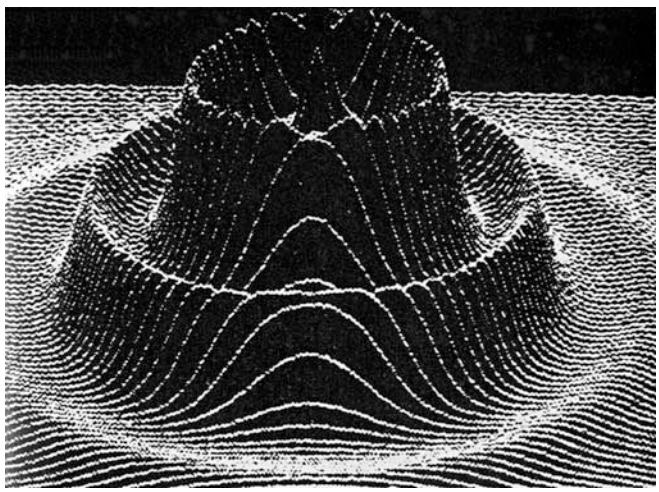


FIGURE 18.13. A computer plot of the diffracted intensity distribution from an amorphous structure, showing diffuse rings of intensity. The direct-beam intensity is off-scale.

showed clearly that the structure correlation can extend to 1.5 nm or more. Life gets more interesting as we explore the structure out beyond the nanometer level and Treacy et al. review a new imaging technique called fluctuation microscopy which is a hybrid imaging/diffraction technique sensitive to the presence of medium-range order in amorphous materials. We'll discuss this technique more in Chapter 29. If you get the idea that

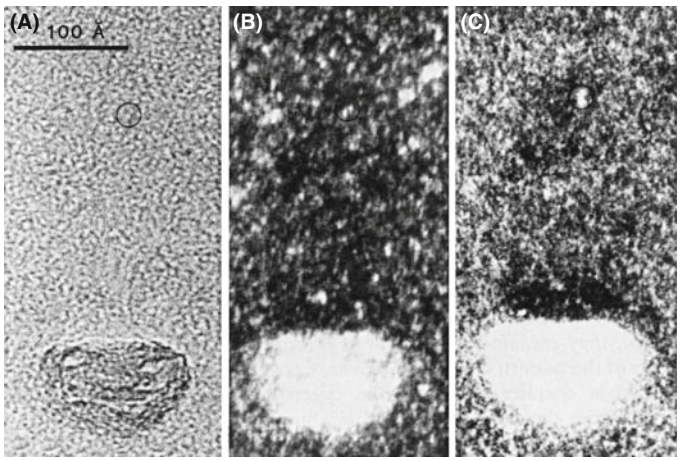


FIGURE 18.14. (A) BF image of amorphous carbon. (B) DF image from the diffuse diffracted intensity taken with a defocused beam and (C) hollow-cone image showing more structure.

the study of amorphous materials is at the frontiers of TEM materials-structure determination, you are right!

To summarize this discussion, the scattering theory is well known but the capability for routinely removing the inelastic contribution is only now becoming available and is still not commonplace. Probably the best way to answer whether a material is nanocrystalline or amorphous will come from a combination of SAD and EELS. A BF image of amorphous material is generally uninformative (Figure 18.14A), but if you try to form a DF image you will see a speckle of white spots against a dark background, as shown in Figure 18.14B. The size of the speckle increases as the defocus increases, so be wary of interpreting the image in terms of the size of regions in the amorphous structure. Hollow-cone DF imaging, as shown in Figure 18.14C, gives even more and finer 'structure' in the image. The fact that you can produce this type of speckled contrast is important because you may well want to study small particles (e.g., catalysts) supported by an amorphous film. In such a case, you need to know what the image of the support film looks like before you add a new component.

Glass at interfaces and grain boundaries. Another area where it is important to know whether or not an amorphous material is present, occurs in the analysis of grain boundaries in ceramic materials or gate oxides in semiconductor devices. Answering this problem severely taxes the limits of HRTEM when the film thicknesses are < 1 nm, which is the case for the gate oxide in the latest semiconductor devices. For somewhat thicker films we can use conventional HRTEM and another technique, known as diffuse-dark-field (DDF) imaging, which essentially forms an image from the region in the SADP where the amorphous ring would be, if glass were present. We'll return to DDF imaging in Chapter 29.

18.8 PRECESSION DIFFRACTION

Precession is a relatively new approach to performing electron diffraction; it is similar to the X-ray method which goes by the same 'name' (although in XRD the crystal, rather than the beam, is precessed) and it is closely related to hollow-cone diffraction that we just talked about. The main advantage that precession brings is to remove strong dynamical effects from the patterns leaving high-quality kinematical data, eliminating, e.g., extra spots due to dynamical effects. Precession diffraction can also be used for crystal structure and full symmetry determination. While the method was first tried in a TEM by Vincent and Midgley, recent advances in both CCD cameras and (you've guessed it) aberration correction make both the operational aspects and the interpretation of precession DPs much easier.

In precession diffraction we double-deflect the incident beam (either parallel or convergent) using the usual DF scan coils in a circular hollow cone (radius G and angle C) about a centered zone-axis direction (Figure 18.15) and de-scan the beam onto the plane of the DP. So it is really a double conical beam-rocking system.

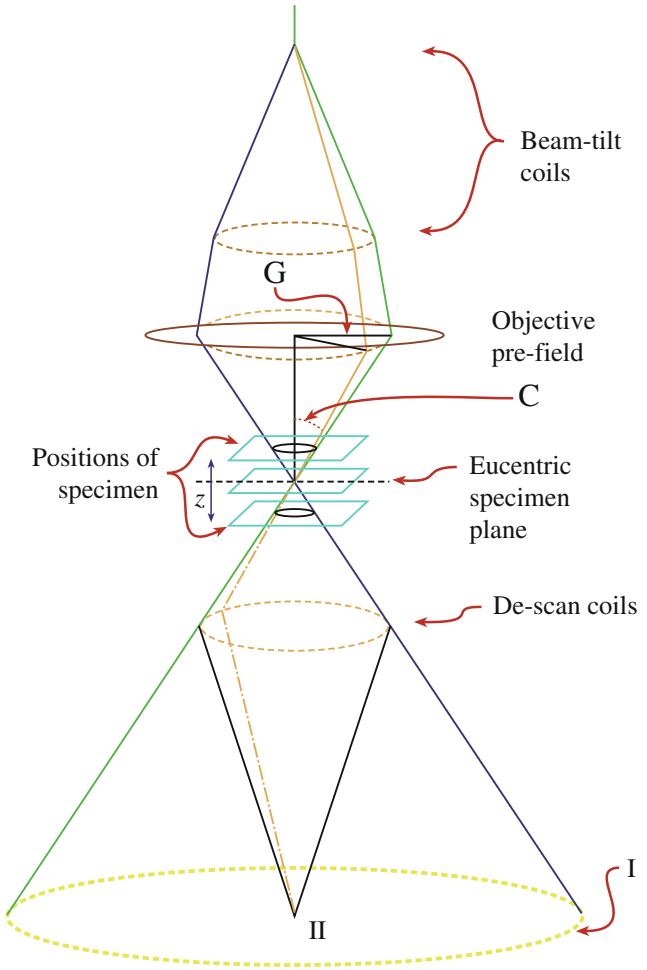


FIGURE 18.15. Schematic ray diagram for precession diffraction.

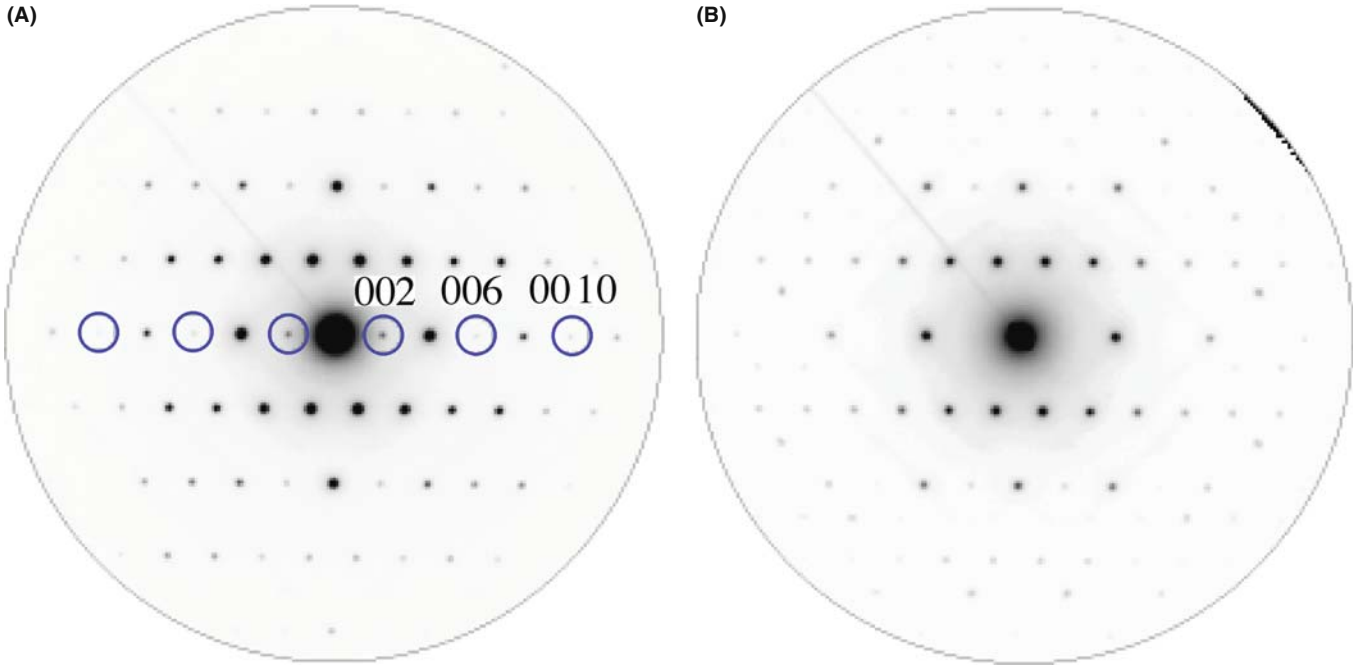


FIGURE 18.16. Si <130> DP taken (A) with a small precession angle (1°) and (B) with a large precession angle (3°). The kinematically forbidden reflections 002, 006, 0010 disappear at the larger precession angle. Note also the appearance of strong differences in the intensity of individual *hkl* maxima.

Unlike standard TEM hollow-cone where we are seeking to get multiple (i.e., ring) patterns sampling many grains, in precession mode we are usually trying to average out the diffraction conditions within a single grain and obtain a single-crystal pattern in which dynamical effects are reduced. If the hollow-cone angle is large enough, all the diffraction data correspond to a two-beam condition and show reduced dynamical diffraction because not many reflections are simultaneously excited off the zone-axis condition. When we precess the beam, we integrate the diffracted intensities through the Bragg condition, so small specimen-tilt changes are averaged out and we can more reliably interpret crystal-symmetry information in the DP. Continuous changing of the precession angle over several degrees can help with point and space-group determination (see Chapter 21, the companion text and Morniroli et al.'s paper). Figure 18.16A and B shows, respectively, the differences between small and large precession-angle DPs from a Si single crystal. The larger angle removes the kinematically forbidden reflections that are present in the smaller-angle DP because the multiple-diffraction paths to them are unlikely to occur during the precession movement. Most of the time, only the direct beam and one diffracted beam are strongly excited as the beam precesses. This does not mean that the dynamical interactions disappear, they just occur between the direct beam and one diffracted beam (as in the two-beam case).

Precession is an area of electron diffraction that is seeing a rapid growth in applications across many fields. Commercial retrofits are available for any TEM (e.g., URL #3) but, if you want to do it on your own, then Own et al. give good detailed instructions.

18.9 DOUBLE DIFFRACTION

Double diffraction occurs when a diffracted beam traveling through a crystal is rediffracted either within the same crystal or when it passes into a second crystal. If the initial diffraction vector of the beam is $\mathbf{g}_1$ and it is rediffracted by reflection $\bar{\mathbf{g}}_2$, then the resultant diffraction vector of the double-diffracted beam is $(\mathbf{g}_1 - \mathbf{g}_2)$. If $\mathbf{g}_2$ is not an allowed reflection in the first crystal, the double-diffracted beam is characteristic of neither the first nor the second crystal.

Reflections attributable to double diffraction are a common feature of DPs recorded from two-phase materials exhibiting epitaxy or topotaxy including, e.g., oxidized metallic specimens. Quite complicated patterns may be formed, requiring careful analysis to distinguish the 'real' reflections from the double-diffraction reflections. Double diffraction is directly responsible for the moiré effect in the electron images that we will discuss in Chapter 23. As an example of this effect, we'll consider small α-Fe_2O_3 (hematite) islands grown on a single-crystal α-Al_2O_3 (alumina or sapphire) substrate as

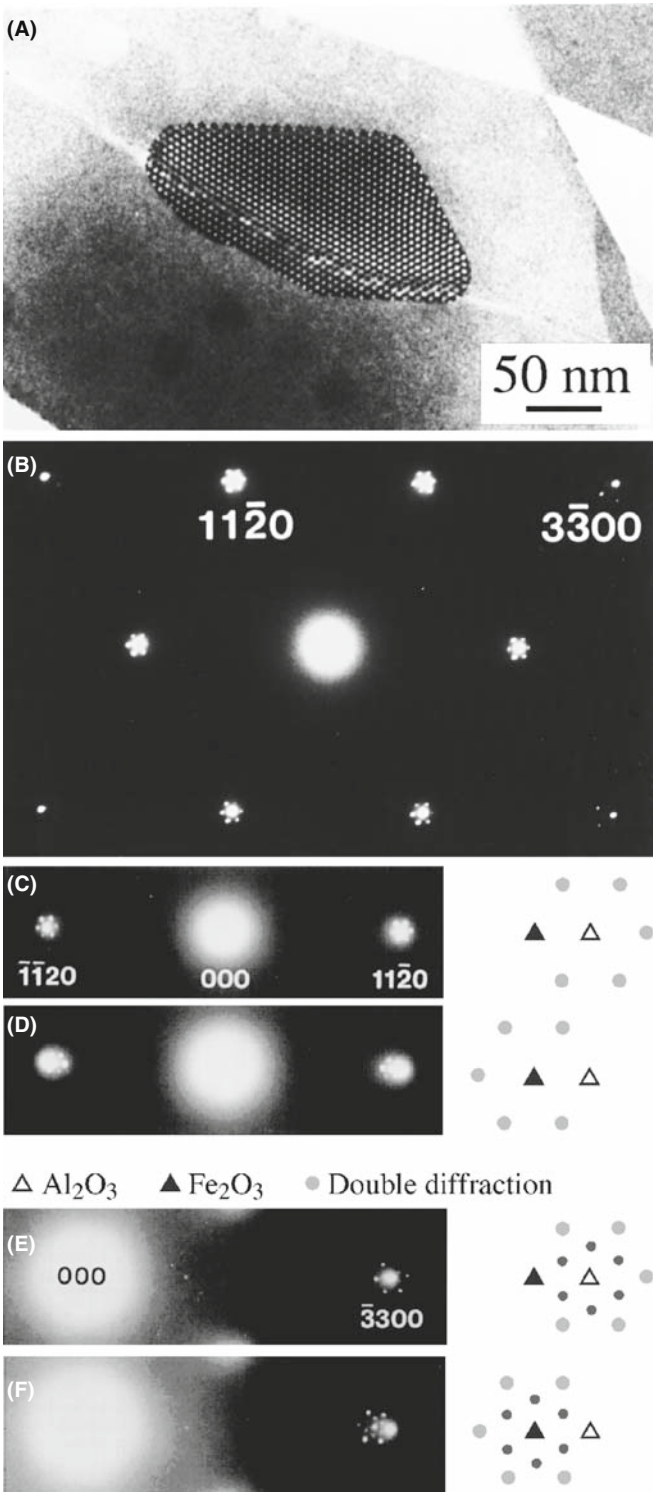


FIGURE 18.17 (A) BF on-axis image of a particle of α-Fe_2O_3 on α-Al_2O_3. (B) [0001] SADP from α-Fe_2O_3 showing double-diffraction spots around the $\{11\bar{2}0\}$ and $\{3\bar{3}00\}$ reflections. (C) Enlargements of regions near the $\{11\bar{2}0\}$ reflections when the hematite island is on the top surface. (D) Enlargements of regions near the $\{11\bar{2}0\}$ reflections when the hematite island is on the bottom. (E) Enlargements of regions near the $\{3\bar{3}00\}$ reflections when the hematite island is on the top surface. (F) Enlargements of regions near the $\{3\bar{3}00\}$ reflections when the hematite island is on the bottom.

shown in Figure 18.17A. The position of the double-diffraction spots relative to the hematite and alumina reflections actually changes depending on whether the islands were on the top or bottom surface of the specimen. This particular top–bottom effect can be derived from simple geometry; however, dynamical diffraction effects must also be considered when the materials are thicker.

Figure 18.17B shows the [0001] SADP recorded from one of these α-Fe_2O_3 particles. The closest reflections to the direct beam are the six $\{11\bar{2}0\}$ reflections. The next closest reflections are the six $\{3\bar{3}00\}$ reflections, only four of which are visible in the figure. Double-diffraction spots are visible around each of these primary reflections. They also surround the direct beam, although they are hidden by the flare from that beam in Figure 18.17B.

Figure 18.17C and D shows enlargements of regions near the $\{11\bar{2}0\}$ reflections in the [0001] SADPs recorded when the hematite island was on the top surface in (C) and on the bottom surface of the sapphire in (D). Both $\mathbf{g}$ and $\bar{\mathbf{g}}$ reflections are shown for the two cases. In (C) the ring of six double-diffraction spots surrounds the Al_2O_3 reflection while in (D) the double-diffraction spots surround the Fe_2O_3 reflection.

The same observation can be made for the $\{\bar{3}300\}$ regions of the SADPs as shown in Figure 18.17E and F. In this case, an inner ring of double-diffraction spots (small filled circles) with the same spacing and orientation as the double-diffraction reflections in Figure 18.17C and D are still visible, as are the outer rings of spots (large filled circles). In general, the outer ring of double-diffraction spots is more intense than the inner ring.

This top-bottom effect in particular, and double diffraction in general, can be explained by the simple geometric analysis we show in Figure 18.18; the bottom crystal is Al_2O_3, which has the smaller lattice parameter and therefore has the larger reciprocal-lattice vectors. Double-diffraction spots can be formed around the primary hematite reflection, $\mathbf{g}_H$, by two different routes

- $2\mathbf{g}_H + \bar{\mathbf{g}}_A$ (A: alumina, H: hematite) giving the double-diffraction spot just inside $\mathbf{g}_H$.
- $\bar{\mathbf{g}}_H + 2\mathbf{g}_A$ gives a double-diffraction spot just outside $\mathbf{g}_H$.

These two routes at first appear to be equivalent. However, if we take into account the curvature of the Ewald sphere, then the deviation parameters of the two routes are very different. In the case of diffraction through the upper crystal, the deviation parameter of the $2\mathbf{g}$ beam is slightly more than twice that of the $\bar{\mathbf{g}}$ beam. This difference will not significantly affect the intensities from a very

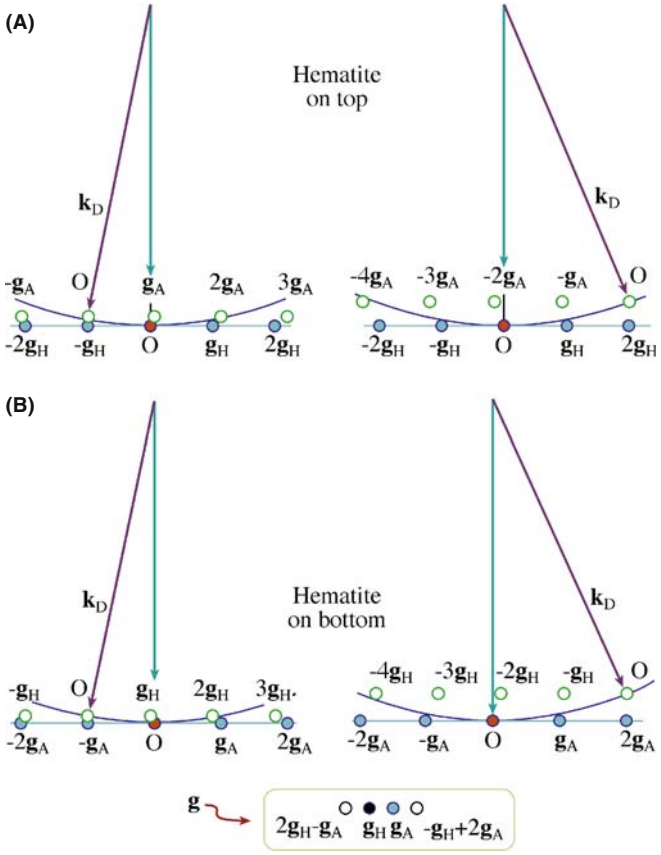


FIGURE 18.18. Top-bottom effect in double diffraction. The pattern depends on which of the two crystals is on top. In this case α-Fe_2O_3 particles are on top of the Al_2O_3 and the two (non-equivalent) paths (A) and (B) for double diffraction are shown. NOTE: $\mathbf{g}_A$ is for alumina and $\mathbf{g}_H$ for hematite.

thin epilayer due to streaking of the reciprocal-lattice spots parallel to the beam direction (the shape-factor effect).

Now we can analyze the effects of diffraction through the lower crystal

- Draw the reciprocal lattice with the origin of the Ewald sphere at $2\mathbf{g}_H$ for the first case and on $\bar{\mathbf{g}}_H$ in the second.
- Keep the radius of the Ewald sphere unchanged since only elastic interactions are considered.
- The incident beam for the lower crystal is in the $2\mathbf{g}$ or $\bar{\mathbf{g}}$ directions for the two cases.
- The height of the ZOLZ is slightly different in the two cases since the deviation parameter at the origin must be zero.

You can see from Figure 18.18A that the deviation parameter for $2\mathbf{g}_A$ is approximately zero, whereas for $\bar{\mathbf{g}}_A$ it is of the same order as $\mathbf{g}_H$. The total deviation parameter is thus much smaller for the second route than the first. A similar analysis for the inverted structure is shown in Figure 18.18B. In both cases, the deviation parameter for the route $\bar{\mathbf{g}}$ (upper) plus $2\mathbf{g}$ (lower) produces a much smaller deviation parameter than the route $2\mathbf{g}$ (upper) plus $\bar{\mathbf{g}}$ (lower). So the double-diffraction spot, which occurs on the same side of the diffraction spot from the upper crystal as the diffraction spot from the lower crystal, will be more intensely excited than the double-diffraction spot which occurs on the opposite side. In two dimensions, for thin films, the strongest double-diffraction spots will always be those arranged symmetrically around the diffraction spot from the lower crystal.

For thicker layers, the relative intensity of the $\bar{\mathbf{g}}$ and $2\mathbf{g}$ beams will vary as dynamical diffraction effects occur. We can simulate the DPs from these structures using the MacTempas program (see Chapter 30 and Section 1.6.B) The top-bottom effect is evident in the case of 2.7 nm of hematite on 13 nm of alumina, but only just discernible for the case of 2.6 nm of alumina on 13.5 nm of hematite. In the latter case, the dynamical diffraction effects are stronger.

We will meet this topic in Chapter 23 when we discuss moiré fringes. We have made this analysis a little more complicated than usual since we have considered the details of where the spots will actually be found. You can make this process simpler

- Trace the patterns from each crystal (if you know what tracing paper is).
- Then construct a new pattern using each diffracted beam from the upper crystal as an incident beam for the lower crystal.

The extent of the moiré pattern gives you an idea of just how strong dynamical scattering is, even for thin films!

18.10 ORIENTATION OF THE SPECIMEN

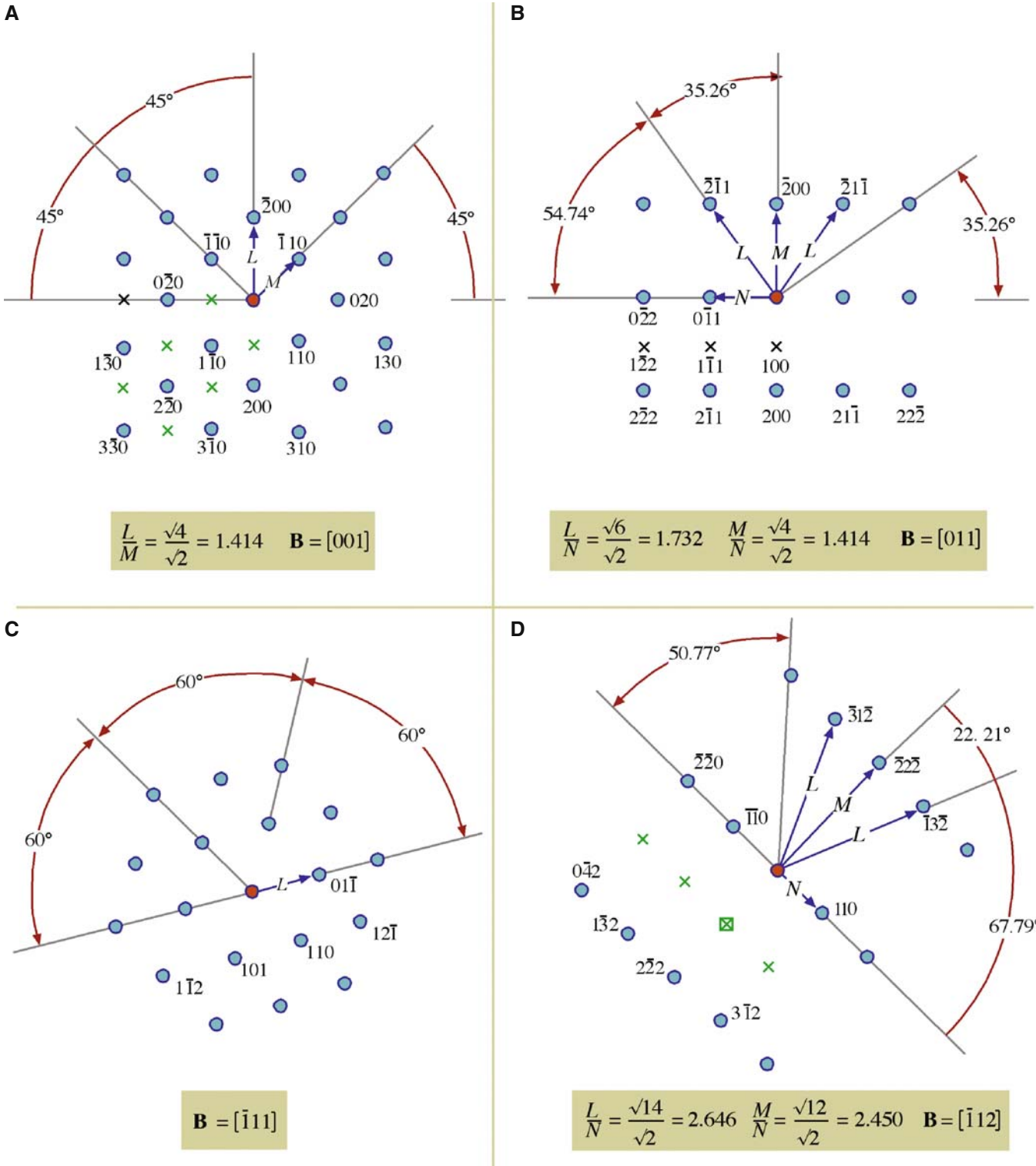
Once you have identified three $\mathbf{g}$ vectors $\mathbf{g}_1$, $\mathbf{g}_2$ and $\mathbf{g}_3$ in a single-crystal DP, you can calculate the direction of the beam $\mathbf{B}$. You can actually estimate $\mathbf{B}$ to within about 10° from the vector cross product as follows

$$\mathbf{B} = \mathbf{g}_1 \times \mathbf{g}_2 = \begin{bmatrix} \mathbf{i}_1 & \mathbf{i}_2 & \mathbf{i}_3 \\ h_1 & k_1 & l_1 \\ h_2 & k_2 & l_2 \end{bmatrix} \tag{18.11}$$

$$= (k_1 l_2 - k_2 l_1, l_1 h_2 - l_2 h_1, h_1 k_2 - h_2 k_1) \tag{18.12}$$

For the three-beam case, you can determine $\mathbf{B}$ with an accuracy of $\sim 3^\circ$. You first need to make sure that the three vectors are taken in the correct order. Draw a circle through these three reflections: if O is inside the circle, then the $\mathbf{g}$ vectors should be numbered counterclockwise; if O is outside, number them clockwise. Check your labeling; the determinant of the matrix in equation 18.13 should be positive

$$\mathbf{g}_1 \cdot (\mathbf{g}_2 \times \mathbf{g}_3) = \frac{1}{V} \begin{bmatrix} h_1 & k_1 & l_1 \\ h_2 & k_2 & l_2 \\ h_3 & k_3 & l_3 \end{bmatrix} \tag{18.13}$$



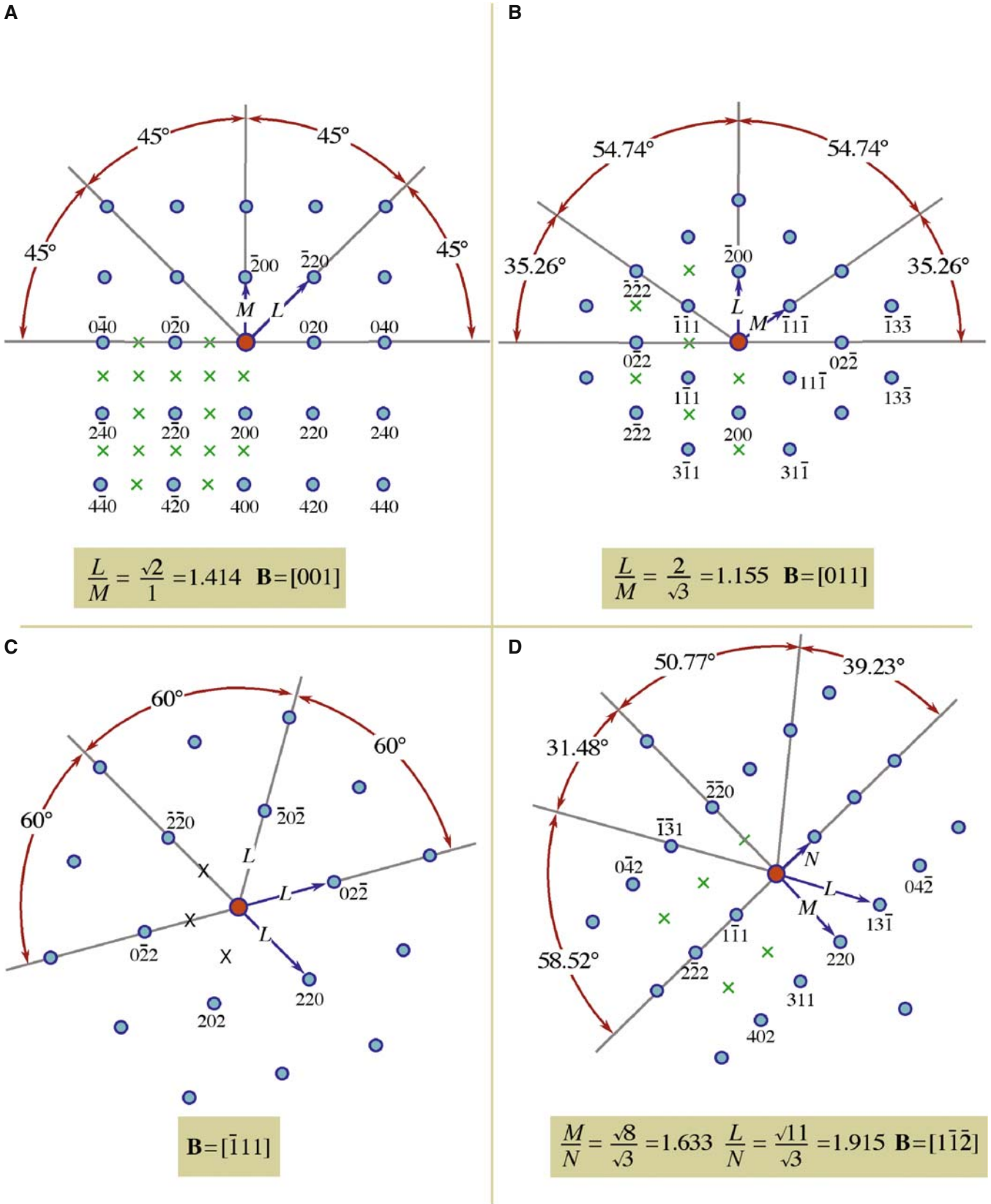
FIGURES 18.19. Four standard, indexed, DPs for bcc crystals in the [001], [011], $[\bar{1}11]$ and $[\bar{1}12]$ beam directions. Ratios of the principal spot spacings are shown as well as the angles between the principal plane normals. Forbidden reflections are indicated by x.

Now we can write a weighted-average expression for **B**

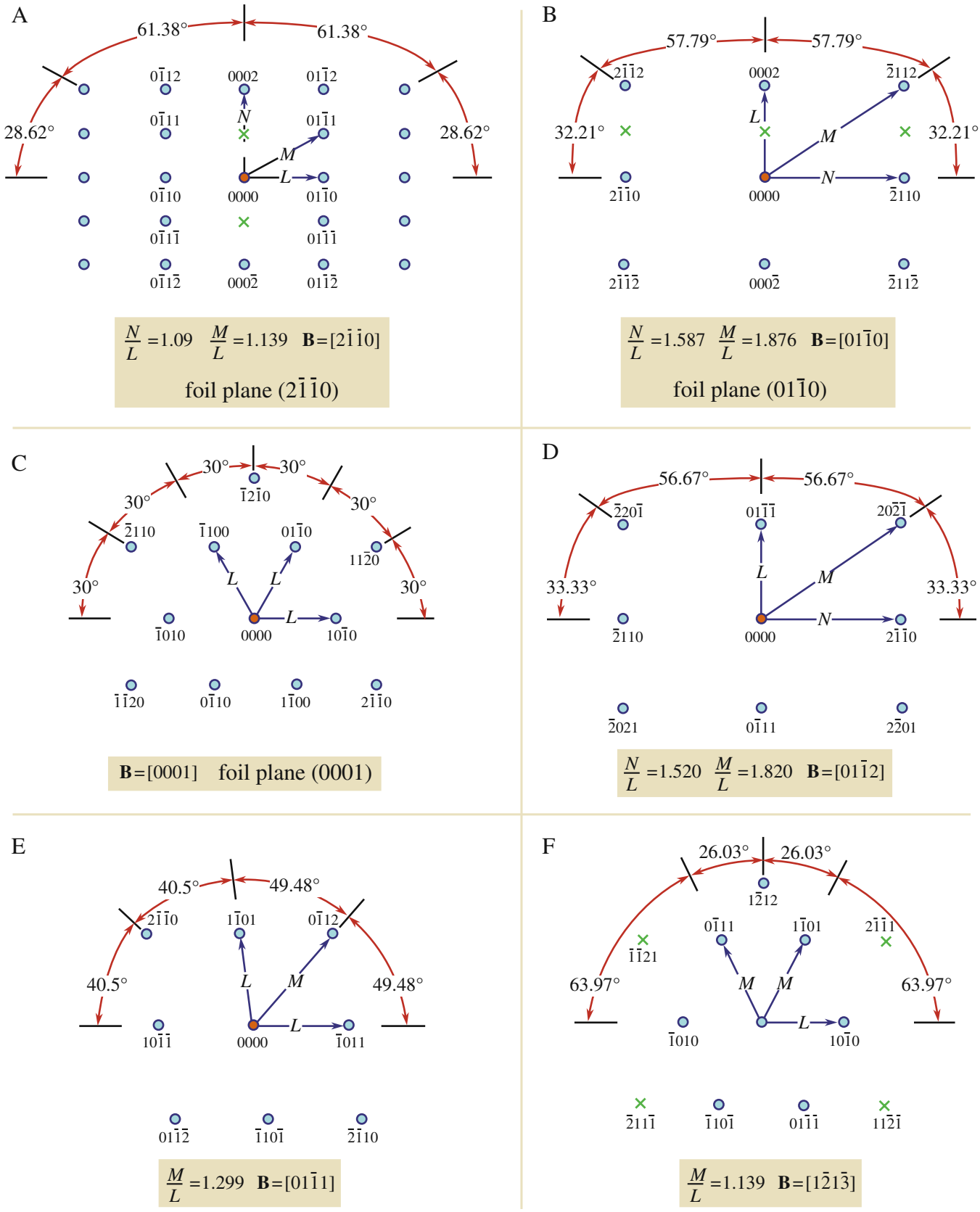
$$\mathbf{B} = \frac{\mathbf{g}_2 \times \mathbf{g}_3}{|\mathbf{g}_1|^2} + \frac{\mathbf{g}_3 \times \mathbf{g}_1}{|\mathbf{g}_2|^2} + \frac{\mathbf{g}_1 \times \mathbf{g}_3}{|\mathbf{g}_3|^2} \qquad (18.14)$$

OUR CONVENTION

The vector **B** points up the column. It is normal to the emulsion side of a photographic negative. The electron beam travels along the direction –**B.**



FIGURES 18.20. Four standard, indexed, DPs for fcc crystals in the [001], [011], [$\bar{1}$11] and [$\bar{1}$12] beam directions. Ratios of the principal spot spacings are shown as well as the angles between the principal plane normals. Forbidden reflections are indicated by x.



FIGURES 18.21. Six standard, indexed, DPs for hcp crystals in the [$2\bar{1}\bar{1}0$], [$01\bar{1}0$], [0001], [$01\bar{1}2$], [$01\bar{1}1$] and [$1\bar{2}1\bar{3}$] beam directions. Ratios of the principal spot spacings are shown as well as the angles between the principal plane normals. Forbidden reflections are indicated by x.

In Figures 18.19–18.21, we illustrate some of the most useful DPs for bcc, fcc and hcp crystals. You can extend these patterns as far as you wish using vector addition; remember the reflections correspond to reciprocal-lattice *vectors*. For example, in Figure 18.19C

$$(12\bar{1}) = (110) + (01\bar{1}) \qquad (18.15)$$

You can extend the patterns in this way and then apply the selection rules to find the corresponding patterns for Si, etc., using the specific examples as a guide.

- bcc real space —> fcc reciprocal space.
- fcc real space —> bcc reciprocal space.

Take the example used by Edington, as shown in Figure 18.22 for an fcc crystal. Measure the distances to the reflections x, y and z. Since the material is fcc, we can ratio d^2 values to find suitable indices or use a calibrated camera length. Thus we find that plane $A = (4\bar{2}0)$, $B = (111)$ and $C = (\bar{3}31)$; check that the angles are correct using

$$\cos(\phi_{AB}) = \frac{\mathbf{g}_A \cdot \mathbf{g}_B}{|\mathbf{g}_A||\mathbf{g}_B|} \qquad (18.16)$$

and so on for ϕ_{BC} and ϕ_{CA}. You should immediately recognize that this is the $\pm[1\bar{2}\bar{3}]$ pole, but continue. Now you can plug pairs of these indices into equation 18.11 or all of them 18.14 to show that $B = [\bar{1}\bar{2}3]$.

Finally, use the [001] stereographic projection. Draw a great circle that passes through the (111), $(2\bar{1}0)$ and $(\bar{3}31)$ points using your Wulff net: they all lie on one great circle because they are in the same zone. Now identify the zone axis directly by measuring 90° from all the poles. The result is, of course, the same in each case.

- Notice that if you used the stereographic technique with a non-cubic material, you would locate a direction, not a plane normal.
- You can make the determination of **B** more accurate by making $\mathbf{s} = 0$ for each reflection you use and then estimating your deviation from this idealized orientation. If the specimen is thicker, use Kikuchi lines (Chapter 19).

18.11 ORIENTATION RELATIONSHIPS

Once you've learned how to index a DP and determine **B**, you can determine *orientation relationships (ORs)*, which are one of the most useful aspects of diffraction in the TEM for the materials scientist and nanotechnologist because the orientation between different grains, phases or crystals controls many properties of engineering materials. For example, the OR will determine how well the atomic planes in two different phases or grains fit together and this controls the nature of the interphase interface (e.g., coherent or incoherent) or grain boundary (special or random) which, in turn, governs the ways dislocations interact with the planar defect, thus controlling the mechanical behavior of the material. There are many other such effects of ORs; for example, we often want to know how a fiber is oriented to the surrounding matrix, a nanoparticle or a thin film to its supporting substrate. So the OR between two different crystals is important in many materials and can be described in one of two ways

- Two directions or plane normals (or two sets of parallel planes) can be parallel in the two crystals (the parallel-plane/direction relationship). We use this description for precipitate-matrix ($\beta - \alpha$) orientation relationships where the crystal systems may be different.
- The two crystals have a common direction (axis) so that one crystal can be rotated through some angle

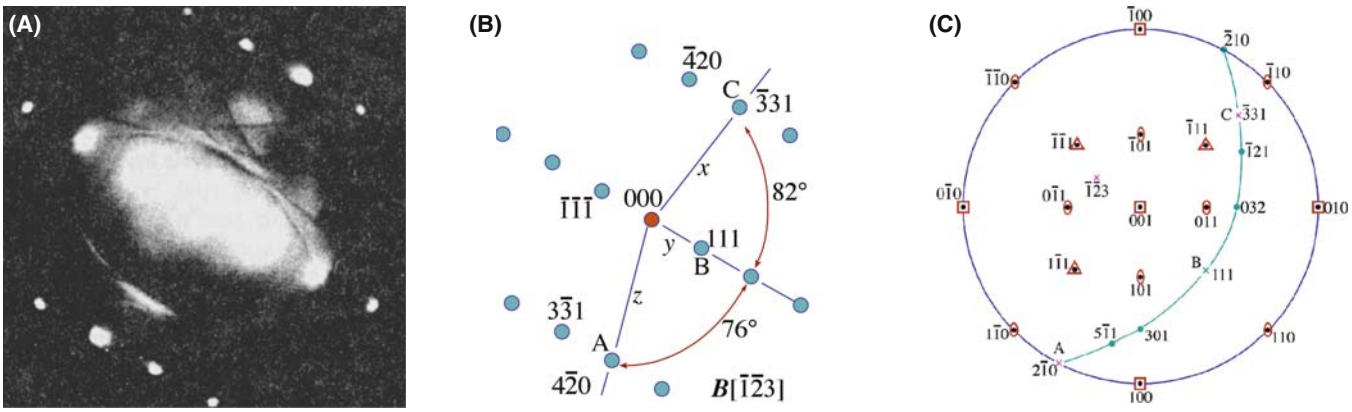


FIGURE 18.22. (A) An fcc pattern, indexed in (B) with the major indexed poles plotted on a stereographic projection in (C), identifying the pole of the great circle as $\bar{1}23$ which is therefore the beam direction for the pattern in (A).

into exact alignment with the other (an axis-angle pair). We use this for GBs where the same material is present, either side of the boundary.

Record a set of three DPs, one from each crystal and one including the interface. If you're lucky you'll be able to index both single-crystal patterns directly. If one of them shows too few spots, you should try to record a complementary Kikuchi pattern (next chapter) or CBED pattern (subsequent two chapters) to provide more information. With CBED patterns from very small regions, you'll have to take a pattern in one crystal, translate the specimen or traverse the beam and take another pattern from the other grain.

We'll go through the experimental steps for analyzing the parallel-plane/direction relationship for two phases α and β

- Tilt to the zone-axis pattern (ZAP) 1 in phase α, the matrix phase. Record and index it to determine $\mathbf{B}_1(\alpha)$.
- Translate the precipitate, β, onto the axis without touching the beam-tilt controls and record another DP. This pattern may not be exactly on a zone axis, so it may be more difficult to index; then Kikuchi lines may help considerably. Nevertheless, you need to determine a parallel-beam direction, $\mathbf{B}_1(\beta)$, for the precipitate.
- Translate back to the matrix. Tilt the specimen in a known direction until you find a different ZAP (again, Kikuchi maps will help you do this). Record and index ZAP 2 to give $\mathbf{B}_2(\alpha)$.
- Translate back to the precipitate, record the DP and index it, giving you $\mathbf{B}_2(\beta)$.
- Plot the position of $\mathbf{B}_1$ and $\mathbf{B}_2$ for both α and β on a stereogram and construct the poles of the important planes that are normal to each **B**. These will be the low-index planes that you indexed in each pattern.

So now you know that $\mathbf{B}_1(\alpha)$ is parallel to $\mathbf{B}_1(\beta)$ and $\mathbf{B}_2(\alpha)$ is parallel to $\mathbf{B}_2(\beta)$. You can also see which plane normals are parallel (if any) from the stereogram. So you can quote the OR in terms of these two pairs of parallel directions, or a pair of directions and a pair of plane normals in the zone of each **B**. It may well be the case that you can't find two low-index planes or directions that are parallel, in which case the orientation relationship is not a strong one. However, there are some well-known ORs between phases that you should know

- Best known is the *cube*/*cube* OR. If an fcc precipitate forms inside an fcc matrix (e.g., Al_3Li (δ') in an Al-Li (α) solid solution), then we find:

  $[100]_{\delta'}$ is parallel to $[100]_{\alpha}$,
  $(010)_{\delta'}$ is parallel to $(010)_{\alpha}$.

  Obviously, in these circumstances, any two $<UVW>$ directions or $\{hkl\}$ planes in the cubic system would be parallel. It's just convention to choose the lowest-index planes or directions to define the OR. When the lowest-index planes and directions align, the surface energy between the phases tends to be lowest, so this configuration is thermodynamically favored.
- The Kurdjumov-Sachs OR is often found relating fcc and bcc crystalline grains. The close-packed planes (or closest packed in bcc) and close-packed directions are parallel, but these are not now identical

  $(111)_{fcc}$ is parallel to $(011)_{bcc}$ (the closest-packed planes),
  $[10\bar{1}]_{fcc}$ is parallel to $[11\bar{1}]_{bcc}$ (the close-packed directions),
  $(\bar{1}2\bar{1})_{fcc}$ is parallel to $(\bar{2}1\bar{1})_{bcc}$.
- The Nishiyama–Wassermann OR is related to the Kurdjumov–Sachs OR

  $[0\bar{1}1]_{fcc}$ is parallel to $[001]_{bcc}$,
  $(\bar{1}11)_{fcc}$ is parallel to $(\bar{1}10)_{bcc}$ (the closest-packed planes),
  $(211)_{fcc}$ is parallel to $(110)_{bcc}$.

If you plot this out on a stereogram, you'll see it's only a few degrees away from the Kurdjumov–Sachs relationship.

- The fcc and hcp systems also share an OR in which the close-packed planes and directions are parallel:

  $(111)_{fcc}$ is parallel to $(0001)_{hcp}$ (the close-packed planes),
  $[1\bar{1}0]_{fcc}$ is parallel to $[1\bar{2}10]_{hcp}$ (the close-packed directions).

If you want to determine an axis-angle pair you proceed in a similar way. Obtain two indexed beam directions, $\mathbf{B}_1$ and $\mathbf{B}_2$, in each crystal, and plot them on a stereogram. Then you need to determine from the stereogram which angle brings the directions and planes from one crystal into coincidence with the other crystal.

There's a full discussion of these methods in Edington's text; Randle and Ralph summarize the various methods available to determine boundary crystallography and many more examples of ORs are described in the companion text.

18.12 COMPUTER ANALYSIS

Although you must be able to analyze and index DPs 'by hand,' it's likely that you'll use one of the many available software packages, especially if your specimen is not cubic (see, e.g., Section 1.6 and URLs #1, 4–7). The main challenge comes when you have to index the

DP of a new material. Your laboratory should have the standard reference sources listed at the end of the chapter. The approach simply requires that you collect all the data you can and then search through the ICDD powder diffraction files, or better still the NIST/Sandia/ICPD electron-diffraction database (see URL #5), until you find a match. Yes, it is a lot of work and you have to remember some rules

- Measurements made on calibrated SADPs will be accurate to 1–2%. If you think you're more accurate, you may eliminate the material you're seeking from your database search!
- Check for multiple domains and double diffraction first. A schematic of such a DP is shown in Figure 18.23. As you can appreciate from the schematic, you must be careful not to confuse such patterns with those showing systematic absences.

A strategy for search-and-match procedures has been given by Lyman and Carr. The goal of the exercise is to identify all the possible compounds that could produce your DP. Then you can use other data (e.g., the chemistry deduced by XEDS or EELS) to make the final identification. Computers not only give us the speed to make such

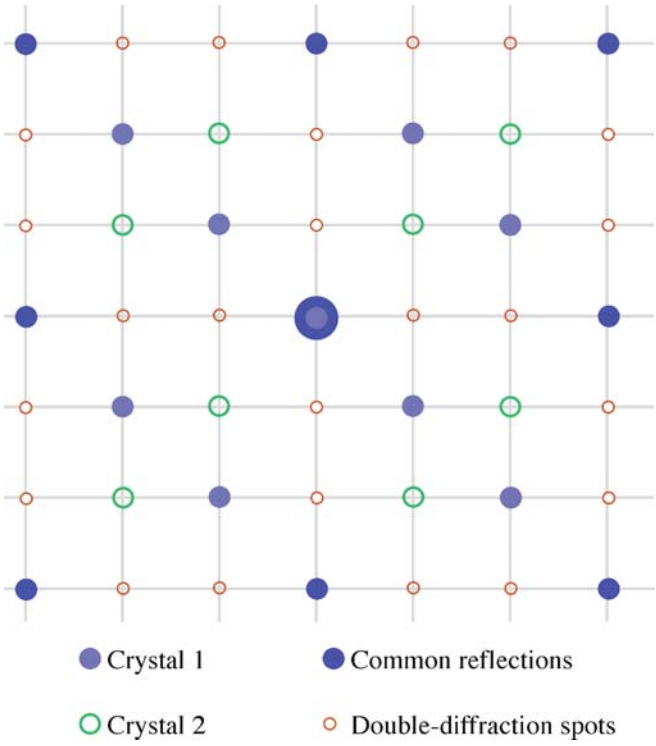


FIGURE 18.23. Care is needed to recognize diffraction from two similar domains which appears identical to diffraction from a real structure with a different symmetry. All the spots lie on a square array which may lead to erroneous indexing as a 100 pattern. The DP actually consists of separate patterns from two overlapping crystals plus double-diffraction spots as indicated.

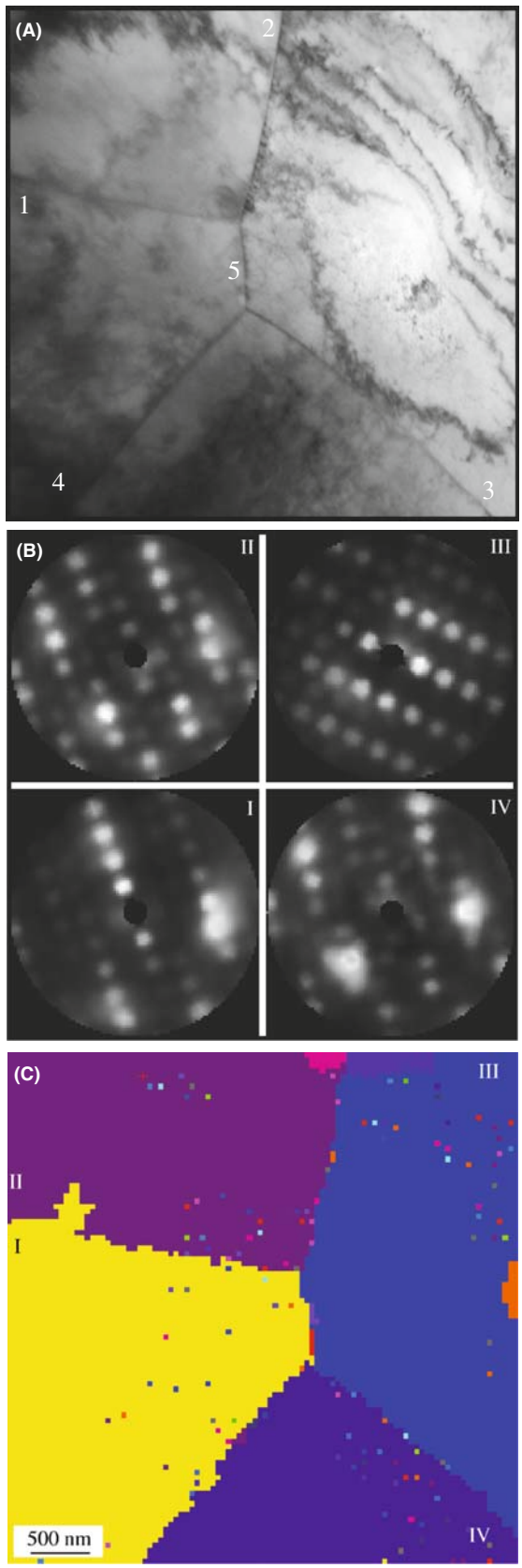


FIGURE 18.24. (A) STEM BF image of four grains in Cu. (B) A set of DPs produced from the grains as the beam scans across them. (C) A grain-orientation map in which different colors relate to different crystal orientations with respect to beam direction.

searches possible, but are also more objective. The procedure has four simple steps

- Obtain reliable data (and do not be too optimistic or overconfident in your accuracy).
- Search the database for possible matches. With the right database, chemical information will help.
- Test the matches you find. Are any of them possible given what you know about your specimen?
- Confirm the identification. Now you can go back to the microscope and use CBED to explore symmetry elements, improve your lattice-parameter measurements, etc. (Chapters 20 and 21).
- You should also simulate the DPs to confirm that the popular software packages do reproduce what you see. It's not a bad idea to carry out this exercise up front using known areas of your specimen or standard specimens before you attack the identification of unknowns.

18.13 AUTOMATED ORIENTATION DETERMINATION AND ORIENTATION MAPPING

In a parallel way to EBSD patterns in the SEM, a sequence of computer-indexed DPs can be transformed into an orientation map, which is an image in which the contrast or color links areas of similar orientation. One way to do this is via Automated Crystallography for TEM (ACT) developed by Dingley and available commercially via EDAX. ACT gives you on-line orientation determination using hollow-cone DF imaging. As shown back in Figure 9.15A, the beam is tilted and rotated by the scan coils and strong diffracted beams travel sequentially down the optic axis as individual Bragg reflections occur. Digital DF images are collected at each beam position by an on-axis CCD camera as each grain diffracts at different orientations and a parallel-beam SADP is built up.

When the incident beam satisfies the Bragg condition for a given grain, the corresponding area in the DF image appears bright. After scanning, the recorded DF images are examined. In the DF images, recorded at different incident-beam positions, a specific pixel always corresponds to the same area in the specimen. The intensity of each pixel can be drawn as a function of beam tilt and rotation angle (i.e., a DP). Adjacent pixels having the same DP are from the same grain, and thus a grain map can be built. Figure 18.24A shows four grains with five grain boundaries in a specimen of Cu. Figure 18.24B shows a set of reconstructed DPs from four grains (I–IV). Once the DPs for all the grains are reconstructed, the orientation relationships across the boundaries (1–5) between the neighboring grains can be extracted and Figure 18.24C shows the orientation map for the four grains. As in the SEM it is easy to contemplate combining the diffraction data with elemental data gathered by XEDS and/or EELS and carrying out on-line phase identification.

CHAPTER SUMMARY

This chapter has been concerned almost entirely with experimental technique.

- The stereographic projection is a very helpful aid. It's similar to projections we use to map the earth. Diffraction space (like global space) is three-dimensional. The stereographic projection gives us a two-dimensional map to guide us from pole to pole!
- How do you obtain the best DP from your specimen? Use the right exposure, always focus the DP and use the best technique (CBED or SAD ± Kikuchis) for the size of the area of interest.
- Take the trouble and time always to get good DPs. You never know when you'll really need that information and an extra 9 or 29 seconds exposure time is not long, considering how long you'll spend analyzing the results!
- Which type of DP should you use? This depends on the characteristics of your specimen and what you want to know.
- Remember that reflections with moderately large values of $\mathbf{g}$ should give you the best value for both d and ϕ, but *be absolutely sure* that $\mathbf{s} = 0$ for your chosen $\mathbf{g}$.
- DPs from polycrystalline, nanocrystalline and amorphous materials contain a wealth of information. The added value that the various TEM techniques bring over X-ray diffraction is the spatial resolution and the accompanying images. TEM may not give the best statistics; XRD only gives statistics.

- Computer indexing of DPs should be the norm and should be automatic *if you know your material*. If you understand the principles discussed here, you will avoid a few pitfalls.
- Computer control of the beam and computer indexing can also be combined to form orientation maps of the distribution of different grain orientations or textures.

Finally, we'll repeat our word of caution: there is a very famous paper on interstitial defects in a ceramic and a follow-up paper on vacancy defects. The first paper missed the 180° ambiguity in the DP! Don't fall into such good company.

REFERENCES

We list more references than usual in this chapter. You'll want to see many different examples so go to the originals where possible.

CRYSTALLOGRAPHY AND DIFFRACTION

Andrews, KW, Dyson, DJ and Keown, SR 1971 *Interpretation of Electron Diffraction Patterns* 2nd Ed. Plenum Press New York. An essential resource for anyone using electron diffraction.

Burger, MJ 1978 *Elementary Crystallography* MIT Press Cambridge Massachusetts. One of the classics. Find it in the library.

Champness, PE 2001 *Electron Diffraction in the TEM* Bios (RMS) Oxford UK. Superb, concise introductory text; full of great examples.

Cullity, BD and Stock, SR 2001 *Elements of X-ray Diffraction* 3rd Ed. Prentice-Hall New York. The standard text on XRD.

Edington, JW 1976 *Practical Electron Microscopy in Materials Science* Van Nostrand-Reinhold New York. Part 2 of the book is full of useful hints and examples on analyzing DPs.

Giacovazzo, C, Monaco, HL, Artioli, G, Viterbo, D, Ferraris, G, Gilli, G, Zanotti, G and Catti, M 2002 *Fundamentals of Crystallography* 2nd Ed. Oxford University Press and IUCr Oxford. A comprehensive reference book.

Glazer, AM 1987 *The Structure of Crystals* Adam Hilger Bristol United Kingdom. The essentials condensed into a 50-page monograph. Find it in the library

Hammond, C 1992 *Introduction to Crystallography* 2nd Ed. Royal Microscopical Society Oxford United Kingdom. An excellent compact introduction to the subject with a nice section on the biographies of crystallographers.

Johari, O and Thomas, G 1969 *The Stereographic Projection and Its Applications in Techniques of Metals Research* Ed. RF Bunshah Interscience New York. A very helpful book if you can find a copy in your library (it's out of print).

Kelly, A, Groves, GW and Kidd, P 2000 *Crystallography and Crystal Defects* Wiley New York. Update of the classic 1970 edition. All materials scientists should already have one on their shelves. Not only the standard introductory text on this subject but also gives a good review of the stereographic projection.

Klein, C and Hurlbut, CS 1985 *Manual of Mineralogy* Wiley New York. This is the modern version of the original classic by James D Dana. It gives an excellent readable review of the stereographic projection and its relation to the globe, plus basic crystallography.

Lisle, RJ and Leyshon, PR 2004 *Stereographic Projection Techniques for Geologists and Civil Engineers* 2nd Ed. Cambridge University Press New York.

Smaill, JS 1972 *Metallurgical Stereographic Projections* Adam Hilger Ltd London. Chapter 20 is another source for stereographic projections and the Wulff net if you can find a copy.

Vainshtein, BK 1981 *Modern Crystallography* **1**–**IV** Springer-Verlag New York. No longer modern but more of a classic.

Villars, P and Calvert, LD 1991 *Pearson's Handbook of Crystallographic Data for Intermetallic Phases* 2nd Ed. ASM Metals Park Ohio. This text is now in many volumes covering an ever-growing number of materials.

Wells, AF 1984 *Structural Inorganic Chemistry* 6th Ed. Oxford University Press New York. The source for crystal-structure data in inorganic materials.

AMORPHOUS MATERIALS

Graczyk, JF and Chaudhari, P 1973 *A Scanning Electron Diffraction Study of Vapor-Deposited and Ion Implanted Thin Films of Ge. I* Phys. Stat. Sol. **B58** 163–179. Early modeling; you should check other papers by these authors if you're working in this area.

Howie, A 1988 in *High-Resolution Transmission Microscopy and Associated Techniques* p 607 Eds. P Buseck, J Cowley and L Eyring Oxford University Press New York. On amorphous materials.

McCulloch, DG, McKenzie, DR, Goringe, CM, Cockayne, DJH, McBride, W, Green, DC 1999 *Experimental and Theoretical Characterization of Structure in Thin Disordered Films* Acta Cryst. **A55**(2) 178–187.

Rudee, ML and Howie, A 1972 *The Structure of Amorphous Si and Ge* Phil. Mag. **25** 1001–1007.

Treacy, MMJ, Gibson, JM, Fan, L, Paterson, DJ and McNulty, I 2005 *Fluctuation Microscopy: a Probe of Medium Range Order* Rep. Prog. Phys. **68** 2899–2944.

DIFFRACTION TECHNIQUES

Dingley, DJ 2000 in *Electron Backscatter Diffraction in Materials Science* p1 Eds. AJ Schwartz, M Kumar and BL Adams Kluwer New York.

Lyman, CE and Carr, MJ 1992 in *Electron Diffraction Techniques* **2** p 373 Ed. JM Cowley Oxford University Press New York.

Morniroli, JP, Redjaimia, A and Nicolopoulos, S 2007 *Contribution of Electron Precession to the Identification of the Space Group from Microdiffraction Patterns* Ultramicroscopy **107** 514–522.

Own, CS, Marks, LD and Sinkler, W *Electron Precession: a Guide for Implementation* Rev. Sci. Instrum. 2005 **76** 33703-1-13.

Randle, V and Ralph, B 1986 *A Practical Approach to the Determination of the Crystallography of Grain Boundaries* J. Mater. Sci. **21** 3823–3828.

Schwartz, AJ, Kumar, M and Adams, BL (Eds.) 2000 *Electron Backscatter Diffraction in Materials Science* Kluwer New York.

Tietz, LA, Carter, CB and McKernan, S 1995 *Top-Bottom Effects in Double Diffraction* Ultramicroscopy **60** 241–246. A challenge on 'double diffraction.'

Vainshtein, BK, Zuyagin, BB and Avilov, AV 1992 in *Electron Diffraction Techniques* **1** p 216 Ed. JM Cowley Oxford University Press New York.

Vincent, R and Midgley, PA 1994 *Double Conical Beam-Rocking System for Measurement of Integrated Electron Diffraction Intensities* Ultramicroscopy **53** 271–282. Precession diffraction.

URLs

1) Diffraction-Pattern Indexing – Start with the EM Yellow Pages and the EMS software http://cimewww.epfl.ch/EMYP/emyp.html and also the Web site for the International Union of Crystallography http://journals.iucr.org/iucr-top/comm/ced/index.html. Also try SingleCrystal™; part of the extensive CrystalMaker® suite of DP and crystal structure software. For a free demo download go to www.crystalmaker.co.uk/singlecrystal/index.html.
2) www.jcrystal.com/products/winwulff/index.htm – WinWulff: a program for plotting stereographic projections onto a Wulff net.
3) www.nanomegas.com – Commercial retrofit for precession diffraction.
4) www.icdd.com – ICDD Powder Diffraction File is produced by the International Center for Diffraction Data (12 Campus Boulevard, Newtown Square, PA 19073). The latest installment of the historic Powder Diffraction File, Release 2006 contains data from the ICDD experimental powder data collection and data collected, edited and standardized from NIST and ICSD database in CD/DVD formats. Expensive, but your university or research lab should buy access.
5) http://icsd.ill.fr/icsd/index.html – Inorganic Crystal-Structure Database (ICSD): the world's most extensive database on inorganic crystal structures and contains information on all inorganic crystal structures (pure elements, minerals, metals and intermetallic compounds, including their atomic coordinates) that have been published since 1913. It is updated twice a year, each time adding approximately 3000 new records. Unlicensed users only have access to a demonstration version, with a 3592 structure subset of the 93,720 inorganic structures in release 2006-2. Access also comes with the ICDD CD.
6) http://icsd.ill.fr/icsd/index.html – NIST Crystal Data comprise standard cell parameters, cell volume, space-group number and symbol, calculated density, chemical formula, chemical name and classification by chemical type. The file includes reliable data from solid-state materials including inorganics, organics, minerals, intermetallics, metals and alloys. Comprehensive chemical, crystallographic and identification search software comes with the database. Access is also available via the ICDD CD.
7) http://www.nist.gov/srd/nist15.htm – NIST/Sandia/ICDD Electron Diffraction Database is thanks to the tireless efforts of M Carr. Designed for phase characterization by electron diffraction, this database and software permit highly selective identification of microscopic and macroscopic crystalline materials.

The database contains chemical, physical and crystallographic information on over 81,534 minerals, metals, intermetallics and general inorganic compounds. Available in CD-ROM format.

THE COMPANION TEXT

Your skill in indexing diffraction patterns always benefits from practice. Energy filtering can improve the clarity of your DP but the geometry will be the same.

SELF-ASSESSMENT QUESTIONS

Q18.1 At what grain size does SAD begin to become useful?
Q18.2 How would you excite a higher-order reflection?
Q18.3 Why is it useful to use stereographic projections when analyzing DPs?
Q18.4 In a stereographic projection, are the small or great circles similar to (a) latitude, (b) longitude?
Q18.5 On the stereographic projection, if the zone axis is at the center of the circle, where are the poles of the planes in the zone?
Q18.6 What kind of ring is seen in the DP from a large-grained polycrystal?
Q18.7 How can we distinguish polycrystalline rings in the DP from those that would be caused by amorphous materials?
Q18.8 Why does double diffraction complicate understanding a DP?
Q18.9 When would you prefer SAD over CBED and vice versa?
Q18.10 Why might grains smaller than 10 nm pose a problem for diffraction analysis in the TEM?
Q18.11 How can you determine if a specimen is nanocrystalline (~1 nm grain size) or amorphous?
Q18.12 Define the radial distribution function.
Q18.13 How can you determine if a polycrystalline specimen is textured?
Q18.14 When is it a good idea to tilt the beam rather than tilt the specimen?
Q18.15 What is an orientation relationship (OR)?
Q18.16 What is the big advantages of the TEM for studying DPs?
Q18.17 How many DPs do you need to determine an OR?
Q18.18 How can you distinguish between the pattern produced by a textured specimen from one produced by a random polycrystal?
Q18.19 What is the best way to image an amorphous material?
Q18.20 Will the DP always rotate as we change the camera length?
Q18.21 Why will SAD not distinguish two patterns with 180° rotation?
Q18.22 When will double diffraction not occur?
Q18.23 What is the easiest way to examine the detail present in a DP?
Q18.24 Why do we tilt from one direction to set a selected specimen orientation?
Q18.25 Distinguish hollow cone and conical diffraction.
Q18.26 Distinguish hollow cone and precession diffraction.

TEXT-SPECIFIC QUESTIONS

T18.1 Describe how you would excite 13**g** when O is on the optic axis if you can't see 13**g** on the screen.
T18.2 Index the DPs in Figure 18.7 to be consistent with the (invisible) Kikuchi lines and with each other (after you've read Chapter 19).
T18.3 We note in Section 18.6 that, unlike those from polycrystalline materials, DPs from amorphous materials do not show speckle. However, speckle from amorphous materials is the basis for the technique of fluctuation microscopy. How can these two statements be consistent? (Use the literature and Figure 18.14 to illustrate your answer.)
T18.4 We note that indexing an OTEDP might be confusing if certain rings are missing due to the texturing. If you tilt the sample 30° these rings should reappear. Explain their absence in terms of a three-dimensional diagram like that shown in Figure 18.10D.
T18.5 Discuss which of the techniques illustrated in Figure 18.14 is better for imaging amorphous carbon. Suggest regions for the large white area in B. Why are the images in B and C different?
T18.6 Using the diagrams shown in Figure 18.17, explain the difference between Figure 18.17E and F. The argument is given in the text so try to put it into your own words or use a diagram to summarize the argument in the text.
T18.7 Is the correct B given in each of Figures 18.19A–D?
T18.8 Sketch Figures 18.21A and C on a sheet of paper so that you can describe how you would tilt from one pole to the other.
T18.9 Notice that in Figure 18.21D–F you are only told B, not the foil plane. Assuming that B is normal to the foil plane, determine the foil plane.
T18.10 Can the DP shown in Figure 18.23 actually happen? If so, give an example.

T18.11 Draw the SADP that you would obtain from a 45° [100] twist grain boundary in Cu when viewed normal to the GB plane.

T18.12 Generate the following DPs (to scale). Show all work and label reflections to a distance of 2**g** in all directions about 000. For the following, assume only 1 atom per lattice site and each lattice site contains the same atom. (A) [123] beam direction; face-centered cubic, $a = 3.68$ Å, $\lambda L = 50$ mm Å. (B) [011] beam direction; c-centered orthorhombic; $a = 4.12$ Å, $b = 3.15$ Å, $c = 5.42$ Å, $\lambda L = 50$ mm Å. (C) [113] beam direction; body-centered tetragonal; $a = 3.41$ Å, $c = 3.0$ Å, $\lambda L = 50$ mm Å. Generate the following DPs (to scale). Show all your work and label reflections to a distance of 2**g** in all directions about 000. For the following, assume only 1 atom per lattice site and each lattice site contains the same atom. (Courtesy Lucille Giannuzzi.)

T18.13 Draw the same cubic stereograms as in Figure 18.4 but rotated 90° along the horizontal axis in each case.

T18.14 Identify on Figure 18.5 the following poles: $22\bar{1}$, $01\bar{1}$, $21\bar{2}$. What is different about these three poles compared with all the others shown in Figure 18.5?

T18.15 Why are the rings of diffracted intensity in the patterns in Figure 18.9 of varying intensity? Under what conditions could you infer something about the different thicknesses of foils that gave rise to each pattern?

19

Kikuchi Diffraction

CHAPTER PREVIEW

In this chapter and the following two, we will discuss two special cases of electron diffraction. We'll see that incoherently scattered, divergent beams of electrons give rise to paired arrays of lines in SADPs, known as Kikuchi patterns. In the next two chapters, we will form DPs with a convergent rather than a divergent (or, as in the previous chapter, parallel) beam. These two techniques have a lot in common. In the first, the electrons are initially being scattered by the atoms in the crystal so that they 'lose all memory of direction' and may also lose energy. We can then think of these electrons as traveling in diverging 'incident' directions into the specimen. When the direction is appropriate, these electrons can be scattered again, this time by Bragg diffraction. In the second technique, we intentionally form a convergent beam in the illumination system to make the electrons incident on the crystal over a range of different angles and create convergent-beam electron diffraction (CBED) patterns. In this case, we have another advantage in that we can focus the beam on a much smaller area of the specimen than in SAD. In both cases the information gained is enhanced if the specimen is thicker; in the case of Kikuchi patterns it has to be thick enough for inelastic scattering to occur and in CBED it has to be thick enough for dynamical scattering. So these next three chapters are particularly useful if you can't make your specimen thin enough for almost all other TEM techniques, which generally produce better quality information if the specimen is thinner.

In this chapter, we will show that these Kikuchi patterns can be used to give us much more accurate information on the beam direction than SADPs and can also give a direct link in reciprocal space to the stereographic projection. The topics we'll cover are basically experimental (although the phenomenon is well understood theoretically) and software for computer simulation of Kikuchi patterns is readily available and very useful. The ideas we develop in this chapter will carry over to the next two chapters when we discuss higher-order Laue-zone (HOLZ) lines in CBED patterns (where Kikuchi lines can also appear).

19.1 THE ORIGIN OF KIKUCHI LINES

The reason Kikuchi patterns form is that, if the specimen is thick enough, it will generate a large number of scattered electrons which travel in all (but mainly forward) directions; i.e., they have been *incoherently* scattered but not necessarily *inelastically* scattered (although obviously some of them will have lost energy). They are sometimes referred to as diffusely scattered electrons. These electrons can then be Bragg diffracted by the crystal planes. The rest of the story is merely geometry.

We'll discuss a little of the theory in Section 19.5, but for now we'll note the following experimental facts

- Since typical energy losses are small (15–25 eV) compared to E_0 (100–400 keV) the diffusely scattered electrons can be assumed to have the same λ as the incident electrons. This assumption holds as long as the specimen is not too thick.
- When first formed, most of the diffusely scattered electrons travel close to the direction of the incident beam. You learned back in Chapter 3 that inelastic scattering is 'peaked in the forward direction.'

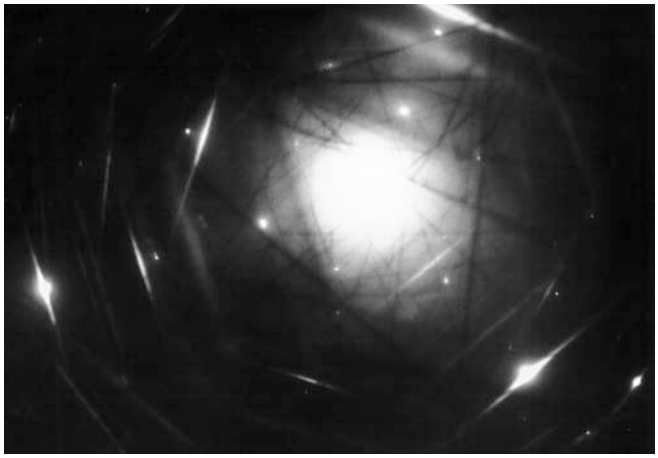


FIGURE 19.1. An ideal DP containing both well-defined spots and clearly visible pairs of bright (excess) and dark (deficient) Kikuchi lines.

- The ideal specimen thickness will be such that we can see both the spot pattern and the Kikuchi lines as illustrated in Figure 19.1. As noted, this is one of the few situations when thinner is not necessarily better.
- Although this phenomenon is related exclusively to electron scattering, Kikuchi described it in 1928, *before* the development of the TEM; it can occur in any crystalline specimen.

Diffuse scattering will again be important when we discuss forming images from these electrons in Section 31.5. We can select a region of reciprocal space containing diffusely scattered electrons to form the image and these electrons can be separated from the inelastically scattered electrons with an energy filter (see Sections 38.2 and 40.5). Your specimen needs to be thick enough but, if it is too thick, then there will be no Kikuchi lines because inelastic scattering then dominates and there is no detectable Bragg diffraction of these electrons. There will also be no Bragg spots and no useful DP!

19.2 KIKUCHI LINES AND BRAGG SCATTERING

The geometry of Kikuchi patterns can be understood from Figure 19.2 which relates what happens in the specimen to what we see in the DP. Let's imagine (Figure 19.2A) that electrons have been generated at the point shown and are scattered in all directions (but

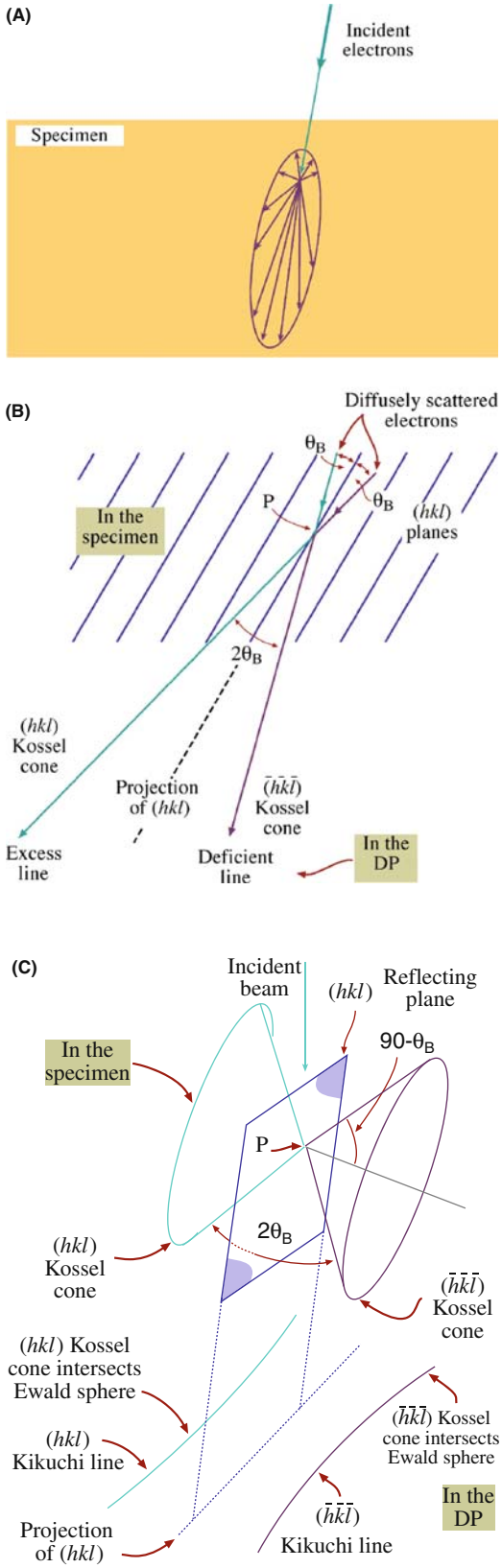


FIGURE 19.2. (A) Schematic representation of all electron scattering localized at a single point in the specimen. In (B) some of the scattered electrons are diffracted because they travel at the Bragg angle θ_B to certain *hkl* planes. The diffracted electrons form Kossel cones centered at P on the diffracting planes. The lines closest to the incident beam direction are dark (deficient) and the lines farthest away are bright (excess). In (C) the cones intercept the Ewald sphere, creating parabolas which approximate to straight lines in the DPs because θ_B is small.

mainly forward). So we draw them as diverging from a point even though, in fact, they'll be scattered at different points throughout the specimen thickness. Some of these electrons will travel at an angle θ_B to the *hkl* planes as shown in Figure 19.2B and then be Bragg diffracted by these specific planes. Since the scattered electrons are traveling in all directions, the diffracted beam will lie on one of two cones (Figure 19.2C). In other words, we see cones of diffracted electrons rather than well-defined beams because there is a *range* of incident **k**-vectors rather than a single **k**-vector. We construct the cones by considering all the vectors oriented at angle θ_B to the *hkl* plane; these are called Kossel cones and the cone angle $(90–\theta_B)$ is very small (remember, angle really means semi-angle). There is a pair of Kossel cones for $\pm\mathbf{g}$, another pair for $\pm 2\mathbf{g}$ and so on.

SEEING KOSSEL CONES

What we see in the DP is the intersection of these two cones with the screen or detector.

Since the screen/detector is flat and nearly normal to the incident beam, the Kossel cones appear as parabolas. If we consider regions close to the optic axis, these parabolas appear as two parallel lines. (Remember how close to 90° the cone angle is.) We'll sometimes refer to this pair of Kikuchi lines as a 'Kikuchi band' to include the lines and the region between them; the contrast associated with the region between the lines is actually quite complex (Section 19.6).

FOR ANY PAIR OF KIKUCHI LINES

One line corresponds to θ_B and the other to $-\theta_B$; one is the **g** Kikuchi line and the other the $\bar{\mathbf{g}}$ Kikuchi line. *Neither* of them is the **0** Kikuchi line.

We can make another important observation on the intensity of these lines by considering Figure 19.2 again. In Figure 19.2B you can see that the scattered beam which was initially closest to the optic axis, and therefore the more intense, is farther away from the axis after being Bragg diffracted. This beam then gives the excess (bright) line and the other the deficient (dark) line. You can see that this simple idea really does work in Figure 19.1.

The value of this result comes when we want to index a pair of Kikuchi lines: if you find a bright line, its partner must not only be parallel to it but must also be closer to O, and dark. The pair is separated by $2\theta_B$.

The cones shown in Figure 19.2C act as if they are rigidly fixed to the plane *hkl*; they are thus 'fixed' to the crystal. We can draw a line half way between the two Kikuchi lines to represent the trace of the plane (*hkl*). Remember our angles are all small. This simple observation explains why we have a whole chapter on Kikuchi lines.

TILTING AND KIKUCHI LINES

If we tilt the crystal through a very small angle, the Kikuchi lines will move but the intensities of the diffraction spots will hardly change and the positions of the spots will not change. So Kikuchi lines are much more sensitive to beam/specimen tilts than spots in SADPs.

The location of the Kikuchi line will also tell us whether **s** is positive or negative. We can't usually deduce that from the spot pattern.

The distance in reciprocal space between the $\bar{\mathbf{g}}$ and **g** Kikuchi lines is **g** (not 2**g**) because the angle between the two Kossel cones is $2\theta_B$. This relationship is very valuable for the following reasons

- When the **g** Kikuchi line passes through the reflection G, $\mathbf{s_g} = 0$ (i.e., the Bragg condition is exactly satisfied) and the $\bar{\mathbf{g}}$ Kikuchi line passes through O. So we can use the Kikuchi lines to set up specific diffraction conditions exciting specific reflections as we tilt the specimen (see Figure 19.3). We'll see later that we can also use Kikuchi lines to determine the exact value of $\mathbf{s_g}$ when we are close to, but not exactly at, an exact Bragg condition.
- A corollary: if the direct beam is exactly parallel to the plane *hkl*, the **g** and $\bar{\mathbf{g}}$ Kikuchi lines are symmetrically displaced about O with the **g** Kikuchi line 'passing through' $\mathbf{g}/2$ and the $\bar{\mathbf{g}}$ line 'passing through' $\bar{\mathbf{g}}/2$.

In this latter case, our simple explanation of Kikuchi-line formation breaks down, because Figure 19.2 predicts equal intensity in both excess and deficient Kikuchi lines and they would both, therefore, be indistinguishable from the diffuse-scattered background. So no Kikuchi lines should be visible if the beam is exactly down a zone axis, and this is patently not true. So the full Kikuchi-line explanation is more complex and (unfortunately) requires Bloch-wave theory. But we do understand the process in great detail.

19.3 CONSTRUCTING KIKUCHI MAPS

If instead of just taking pictures of Kikuchi-line pairs, we assemble a montage of all Kikuchi lines and spot patterns

KIKUCHI MAP

Making a Kikuchi map is a highly recommended exercise if you are going to be doing detailed diffraction-contrast images/SAD experiments because familiarity with the Kikuchi map will help you to immediately identify the orientation of your specimen on your TEM screen.

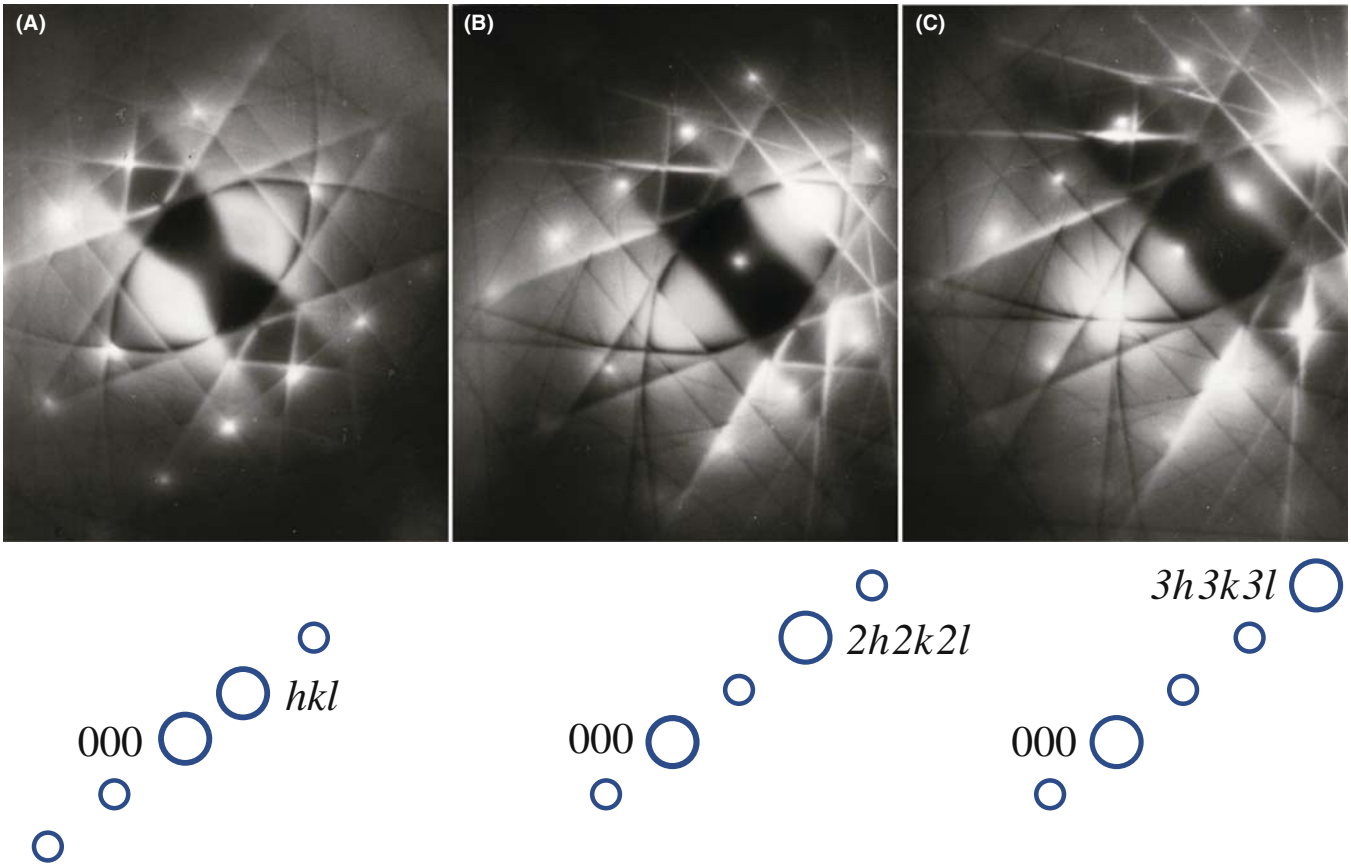


FIGURE 19.3. Three two-beam DPs from pure Al, obtained under different tilting conditions. As shown schematically below each figure, in (A) the *hkl* spot is at the exact Bragg condition (the excess Kikuchi line goes through *hkl*). In (B) the *2h2k2l* and in (C) the *3h3k3l* spots, respectively, are strongly excited. Note that although we refer to these as 'two-beam' DPs, many other diffraction spots are visible.

over a wide array of reciprocal space, then we create a Kikuchi map.

The method for constructing Kikuchi maps is illustrated in Figure 19.4A. First, we draw the lines for the case where the [001] pole is exactly on the optic axis. The lines are then the perpendicular bisectors of every **g**-vector you can find in the ZOLZ. The distance between each pair of lines is then automatically $|\mathbf{g}|$. We can then give each line a unique label **g**.

Next, we can construct the map for the [101] pole. We start as shown in Figure 19.4B, keeping the common 020 **g**-vector pointing in the same direction. So, the 020 and $0\bar{2}0$ Kikuchi lines are common to the two patterns. Although the angle between the [001] and [101] poles is 45°, we draw the 020 lines as parallel and straight because we are always looking at a small segment of the Kikuchi pattern. Notice that we can define all the distances in terms of their equivalent angles, as in any DP.

Now we add the [112] pattern. This pattern shares the $2\bar{2}0$ and $\bar{2}20$ reflections with the [001] pole and shares the $\bar{1}\bar{1}1$ and $11\bar{1}$ reflections with the [101] pole. The corresponding pairs of Kikuchi lines will then also be common, so we produce the triangle shown in Figure 19.5A. We can add other poles and pairs of Kikuchi lines as shown in Figure 19.5B to get the full pattern.

A Kikuchi map for an fcc material is illustrated in Figure 19.6. The maps are available in the literature for fcc, bcc, diamond cubic and some hcp materials. Such maps are mainly from Thomas and co-workers (Levine et al. 1966, Okamoto et al. 1967, Johari and Thomas 1969), who developed the technique. Edington (1976) presents several Kikuchi maps in the appendices. Maps can also be downloaded from the Web using EMS (URL #1).

You can appreciate the value of Kikuchi maps in non-cubic materials from the map shown in Figure 19.7. The map has been drawn for Ag_2Al, which has the same c/a ratio as Ti. The Kikuchi bands are labeled (remember that they correspond to planes). The zone axes are also labeled: they correspond to directions. Thinking back to our brief discussion of Frank's paper on four-index notation in Chapter 16, you should see an obvious application here.

- For cubic materials you need only the [001], [101], [111] triangle shown in Figure 19.5B.
- For hcp materials, the angles will generally depend on the c/a ratio of your material and you'll need a larger area of the map.

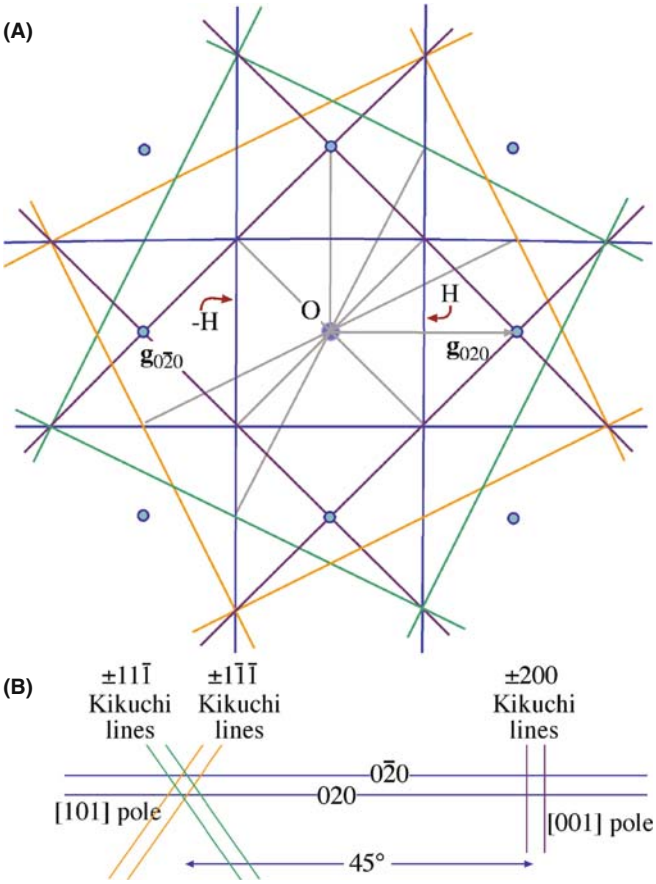


FIGURE 19.4. (A) To construct a Kikuchi pattern, draw pairs of lines each bisecting the ±**g**-vectors. For example, when the [001] fcc pole is on axis, the vector $\mathbf{g}_{020}$ is bisected by the vertical line at H (020); the other Kikuchi line in the pair is at −H (020). All other Kikuchi-line pairs can be constructed for any **g**-vector. (B) From one Kikuchi pattern we can extend the lines to create a second pattern. For example, knowing the [001] pattern we can construct the [101] pattern since a pair of lines is common to both. So we draw the 0$\bar{2}$0 and 020 lines from the [001] pole 45° to the [101] pole.

- For most non-cubic materials and particularly if you are working with monoclinic or triclinic crystals, it's not practical to construct the complete map experimentally. It's probably easier to become a metallurgist!

We can use the following procedure to generate a valuable experimental aid for any material.

- Construct segments of the map to scale as we've illustrated in Figure 19.5B. You can use one of the software packages to help you in this task. Make two copies of each map.
- While you're at the TEM, record the Kikuchi pattern for several special low-index poles along with the spot pattern.
- Index the DPs consistently.
- Print both DPs for each zone axis at the scale you used in your line drawing of the Kikuchi map.

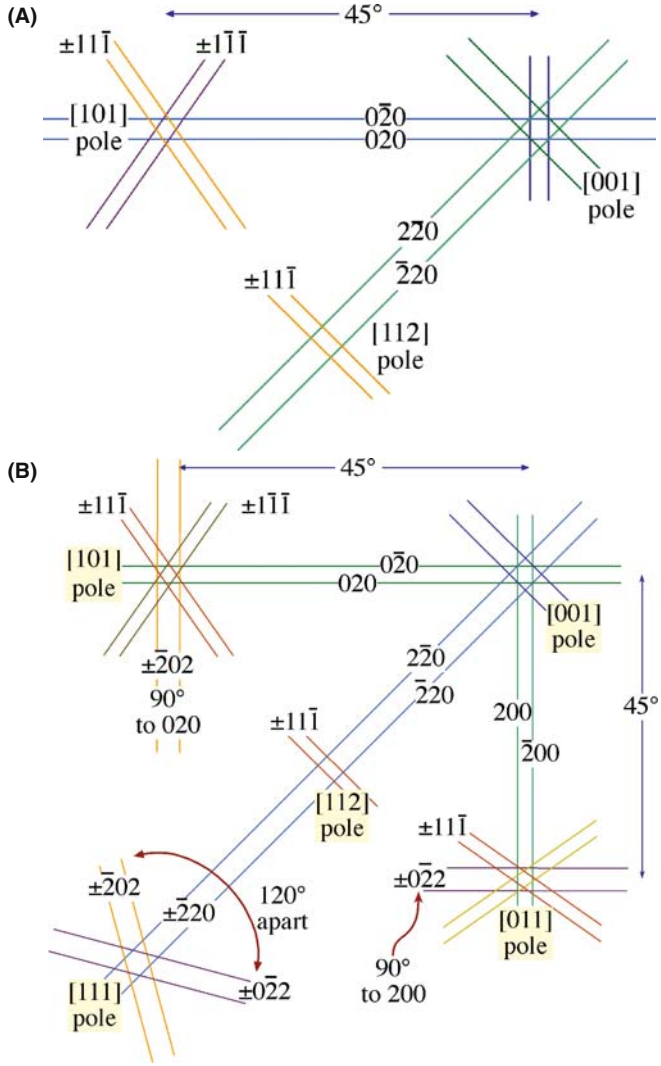


FIGURE 19.5. (A) Construction of the [112] pattern from the [001] and [101] patterns by extending the Kikuchi lines common to each pair of patterns. The [111] pair is common to the [001] and [112] patterns and the $\bar{2}$20 pair is common to [001] and [112] patterns. (B) Other poles can be added such as [011] and [111]. Note that the Kikuchi-line pairs are not straight lines connecting poles. They are curved because over large angles their parabolic shape is evident. Nevertheless we draw them as straight lines where possible.

- Now add the experimental patterns to the line diagrams and you have two very useful experimental aids. An illustration is given in Figure 19.8.

When discussing Kikuchi maps, we like to use the road-map analogy. (Repeatedly!) What we just recommended is that you record the maps of the towns with pictures so that you'll recognize them. When you're on the highway traveling from town to town, you don't much care what the road looks like although you do want to know how far you've traveled and how much farther you have to go.

By now you will appreciate even more the value of the stereographic projection we introduced in Chapter 18.

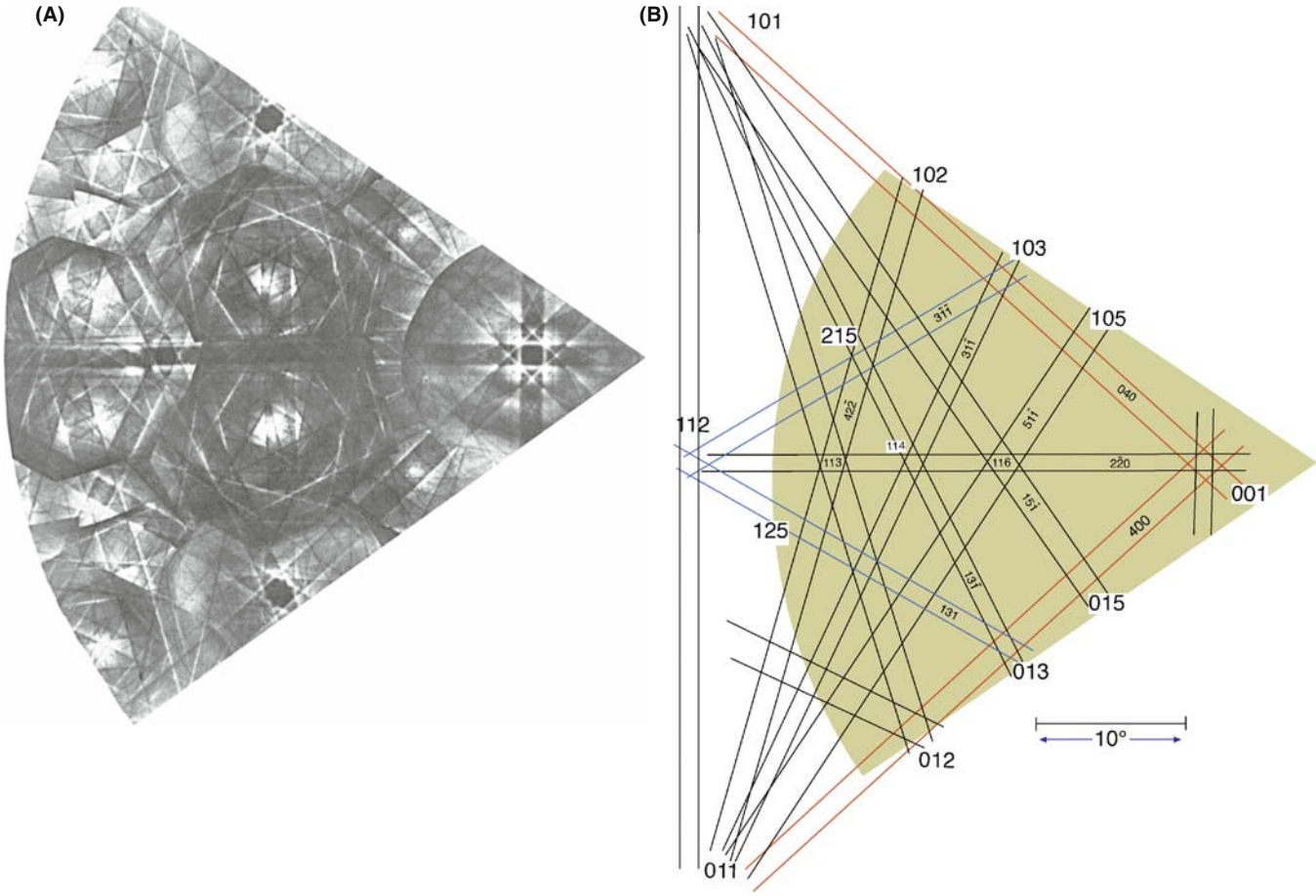


FIGURE 19.6. (A) Experimental Kikuchi map for fcc crystals and (B) indexed Kikuchi lines in the schematic map.

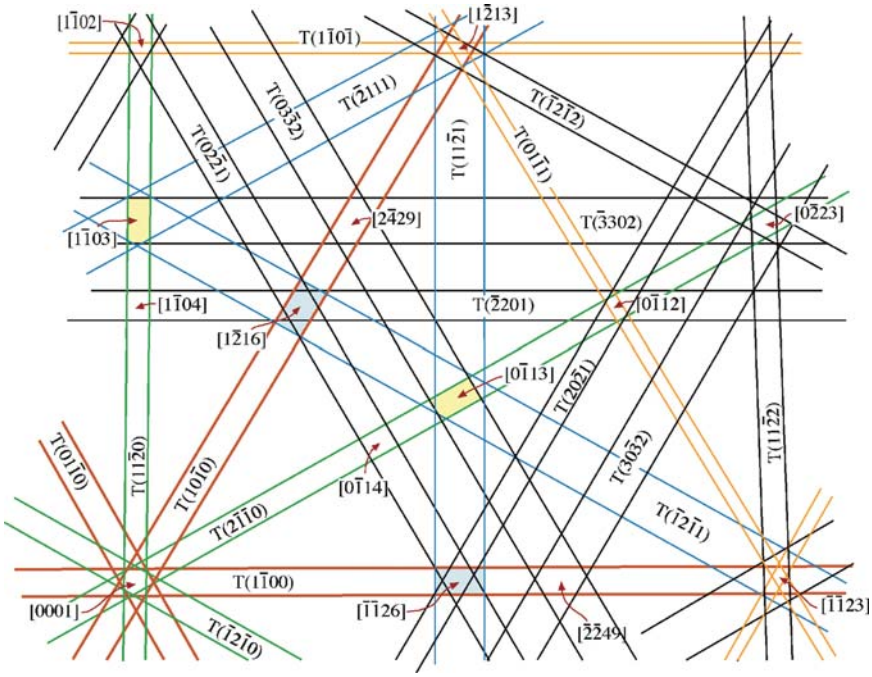


FIGURE 19.7 Part of the Kikuchi map for hexagonal Ag_2Al with the principal poles and pairs of lines indexed.

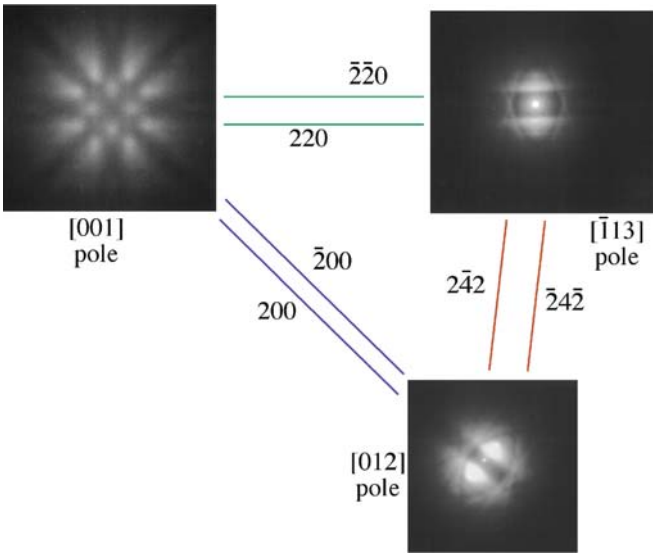


FIGURE 19.8. Experimental Kikuchi patterns around three principal poles in MgO with the common Kikuchi lines between each pole drawn in. You should compare this figure with the DPs in Figure 18.7.

Use the stereographic projection and the Kikuchi map together. The stereographic projection concisely summarizes all the relative locations of all the plane normals and the zone axes. Use the stereographic projection to relate Storrs and Huntsville, but use the Kikuchi map to locate the Benton Museum of Art and the Von Braun Center.

19.4 CRYSTAL ORIENTATION AND KIKUCHI MAPS

In the previous chapter we showed how you could use SADPs to estimate the orientation of the beam relative to the crystal with an accuracy of $\pm 3^\circ$. Using Kikuchi patterns you can increase this accuracy to $\pm 1^\circ$.

A routine method for orientation determination was developed by Thomas and co-workers (e.g., Okamoto et al.); they pioneered the use of Kikuchi maps in TEM analysis. The beam direction $[UVW]$ lies along the optic axis O in Figure 19.9. A, B and C are major poles (i.e., zone axes) which we can determine by observation and measurement. Let the indices of A = $[p_1\ q_1\ r_1]$, B = $[p_2\ q_2\ r_2]$ and C = $[p_3\ q_3\ r_3]$. Having indexed these poles, we can check our result by measuring the angles α, β and γ between the traces of the planes in Figure 19.9A (which equals the angle, ϕ, between the plane normals in all systems); we must calculate each angle using equation 18.3 if the specimen is cubic.

SPECIMEN ALIGNMENT
Kikuchi lines transform SADPs from an approximate to a very precise alignment technique.

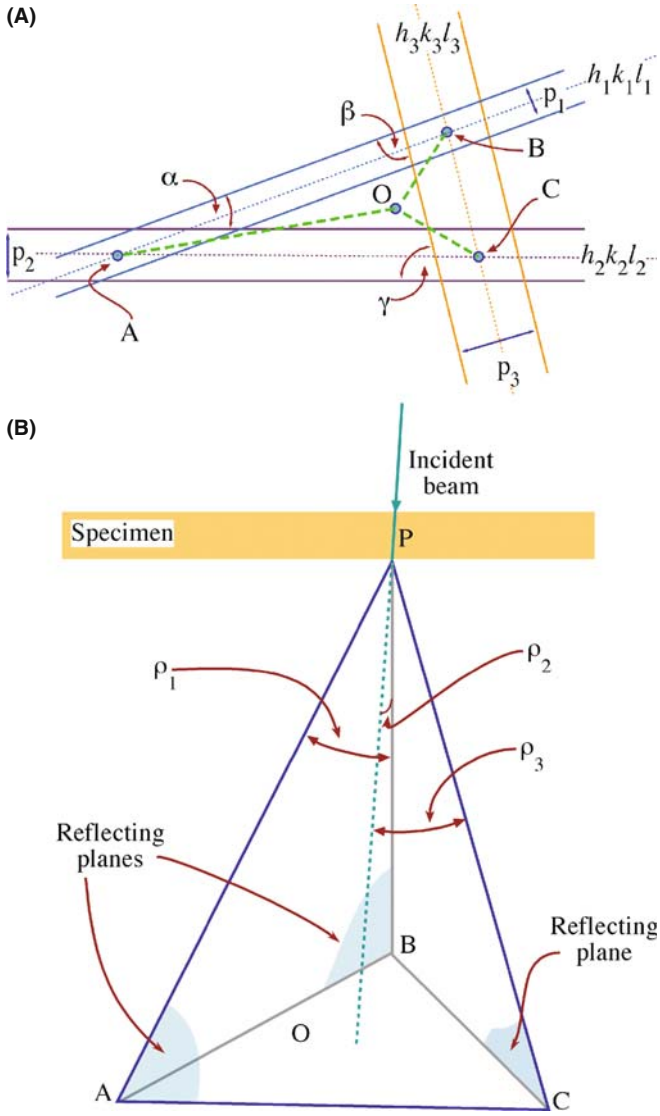


FIGURE 19.9. In (A) pairs of Kikuchi lines from the reflecting planes also intercept at points A, B, C. The distances from O to the points A, B, C correspond to the angles between the beam direction and the three zone axes while the angles α, β, γ correspond to the angles between pairs of plane normals. The angle α is between the $(h_1k_1l_1)$ and $(h_2k_2l_2)$ plane normals, etc. (B) Three reflecting planes in the specimen with traces AB $(h_1k_1l_1)$, AC $(h_2k_2l_2)$ and BC $(h_3k_3l_3)$, around the direct beam, O; the traces of pairs of planes intercept at A (AB, AC), B (AB, BC) and C (AC, BC).

If we measure the distances OA, OB and OC in Figure 19.9A, we can convert them into angles ρ_1, ρ_2 and ρ_3 (which are defined in Figure 19.9B) by using our calibrated camera length. If $[UVW]$ is the direction of the beam, then we can use the same vector-dot-product approach (equation 18.4 for the cubic case) to give equations for ρ_1, ρ_2 and ρ_3. Notice we are distinguishing between ρ and ϕ (see Section 18.4). The angles α, β and γ in Figure 19.9A are slightly distorted values of $(90 - \phi)$.

We can solve these three equations for the three unknowns U, V and W and hence we have **B**. Finally, you should always check the sign of **B**, as we described in Section 18.10.

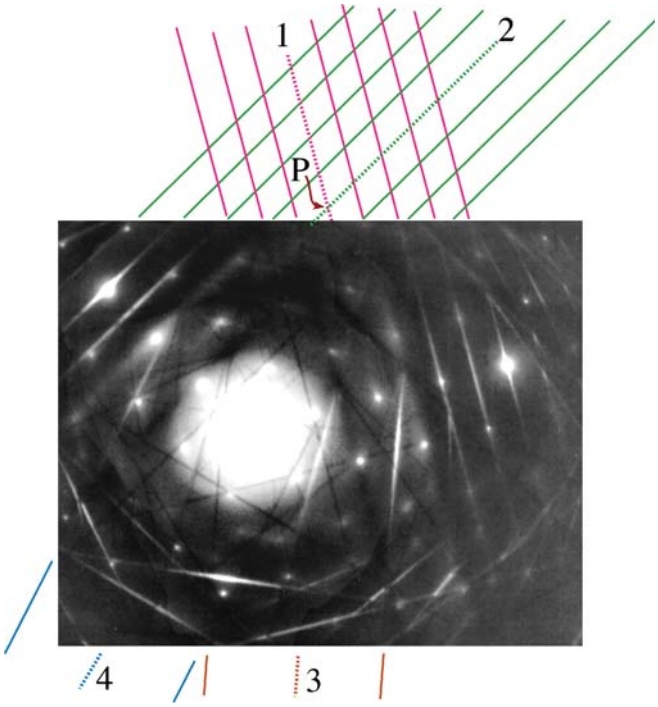


FIGURE 19.10. To index a DP well away from a low-index zone axis, extend the Kikuchi lines. The dark lines 1–4 represent the traces of the diffracting planes which intercept at a pole (P). For Kikuchi lines 1 and 2, the higher-order extensions are also drawn. From the *d*-spacings, index the Kikuchi-line pairs. The angles between the beam direction and the poles, P, can then be measured directly.

It is possible that the DP you have to work with is not obviously near a zone axis. All is not lost if you can just find pairs of Kikuchi lines as shown in Figure 19.10. If you see an excess line you will find the deficient line quite easily, closer to 000. Now trace these lines in both directions and you have found the poles. Use your knowledge of the *d*-spacings to index the pairs of Kikuchi lines. Remember that the zone axis lies parallel to each plane so it's defined by where the two plane traces meet. Now, if you can index three poles, you can obtain **B** as in Figure 19.9.

DO REAL-TIME CRYSTALLOGRAPHY

Normally, while you are at the TEM, tilt along the different Kikuchi bands until you find the appropriate poles to ease your task later.

19.5 SETTING THE VALUE OF S_g

Since the Kikuchi lines are 'rigidly attached' to the crystal, they give us a very accurate measure of the excitation error $\mathbf{s}_g$. The diffraction geometry is shown in Figure 19.11 following Okamoto et al. When $\mathbf{s}_g$ is negative, the **g** Kikuchi line is on the same side of **g** as O;

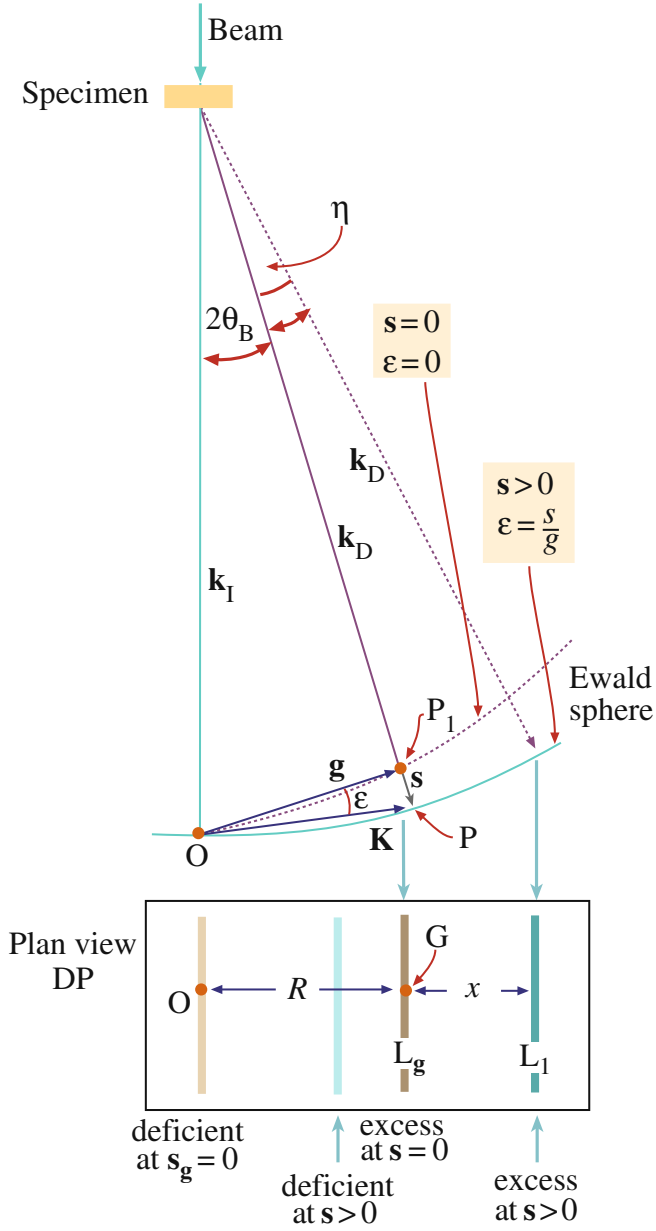


FIGURE 19.11. The distance between the diffraction spot and its Kikuchi line gives a direct measure of **s**. The angle ε is s/g and is zero at the exact Bragg condition. In the box: measure x, the spacing between O and the deficient line (or G and the excess line) to determine **s**.

when $\mathbf{s}_g$ is positive, the line lies on the opposite side of **g**. If you look at Figure 19.10 you'll see that the bright diffraction spot at the top LH corner of the DP is under $\mathbf{s}_g$ positive conditions (excess Kikuchi line outside the spot), the bright spot at the middle/upper RH side is under $\mathbf{s}_g$ zero conditions (the excess Kikuchi line through the spot) and the bright spot toward the bottom of the DP is under $\mathbf{s}_g$ negative conditions (the excess Kikuchi line inside the spot).

For high-energy electrons, and knowing the camera length L, we can write an expression for the angle η in Figure 19.11

$$\eta = \frac{x}{L} = \frac{x\lambda}{Rd} \tag{19.1}$$

where d is $|\mathbf{g}|^{-1}$. The distances x and R are measured on the photographic negative.

The angle ε is given by

$$\varepsilon = \frac{s}{g} \tag{19.2}$$

Now we can set $\varepsilon = \eta$ to give

$$s = \varepsilon g = \frac{x}{L} g = \frac{x}{Ld} \tag{19.3}$$

Again, with our small-angle approximation, the distance between the excess and deficient Kikuchi lines, R (the distance g measured on the DP), is equivalent to $2\theta_B L$. So, using Bragg's law, we have

$$\frac{R}{L} = 2\theta_B = \frac{\lambda}{d} \tag{19.4}$$

Hence the expression for s is

$$s = \frac{x}{Ld} = \frac{x}{d} \cdot \frac{\lambda}{Rd} \tag{19.5}$$

$$s = \frac{x}{R}\frac{\lambda}{d^2} = \frac{x}{R}\lambda g^2 \tag{19.6}$$

We'll reconsider this equation when we discuss weak-beam microscopy in Chapter 27.

Ryder and Pitsch have given a method for determining **B** using the approach we described in Section 19.4 with the accuracy given by equation 19.6. Their expression for **B** is

$$\mathbf{B} = \alpha_1|\mathbf{g}_1|^2(\mathbf{g}_2 \times \mathbf{g}_3) + \alpha_2|\mathbf{g}_2|^2(\mathbf{g}_3 \times \mathbf{g}_1) + \alpha_3|\mathbf{g}_3|^2(\mathbf{g}_1 \times \mathbf{g}_2) \tag{19.7}$$

where α_i is given by

$$\alpha_i = \frac{R_i + 2x_i}{R_i} \tag{19.8}$$

where R and x are defined in Figure 19.11.

19.6 INTENSITIES

We'll conclude with a few remarks for further thought

- Tan et al. have shown experimentally that the distance between a pair of Kikuchi lines may change at larger specimen thicknesses due to dynamical scattering.
- Kikuchi lines can also be produced by the backscattered electrons. In the SEM these patterns are termed electron-backscatter DPs (sometimes EBSPs) and the technique as EBSD. They were regarded as a curiosity until it was shown (see Dingley's paper) that you can map out the texture of polycrystalline materials using these patterns, without thinning the sample. Not much happened for a decade or more until new detection systems, using CCD cameras and fast computer algorithms led to the development of orientation imaging microscopy (OIM) (see Dingley's review). As we saw in Section 18.13, similar techniques are available for automatic indexing of TEM DPs and Kikuchi maps. They aren't as automated, nor can they index as many patterns as EBSD software because there are usually far fewer thin grains in our specimens than crystals in an SEM specimen. But TEM can, in principle, give the interface plane much more accurately, so the two techniques will become more complementary with time.
- In the next chapter, we'll discuss HOLZ lines; HOLZ lines are very closely related to Kikuchi lines but are a little more complicated, since the Bragg planes are always inclined to the direct beam.
- In Chapter 24, we'll discuss ZAPs, or zone-axis patterns, in *images*; these ZAPs are, in many respects, the real-space version of Kikuchi lines. However, you'll see that their physical origin is *completely different*; the most important features of ZAPs are *not* associated with incoherent, inelastic or diffuse scattering.
- Bloch waves with vector $\mathbf{k}^1$, for example, are more strongly scattered than those corresponding to branch 2 of the dispersion surface. Therefore, we can expect anomalous absorption (see Chapter 24) to influence the intensity of Kikuchi patterns. Such effects do in fact lead to excess and deficient Kikuchi bands. Since we haven't yet found any use for the information in these bands we'll leave them as an exercise for further reading!
- We mentioned earlier that the contrast between the Kikuchi lines, i.e., the Kikuchi band, is complex. The contrast is strongly influenced by anomalous absorption of the Bloch waves which are formed by coherent scattering of the incoherently scattered electron, so all is clear.
- There are strong similarities between the Kikuchi process and the operation of a monochromator in optics: both select and diffract a particular wavelength or frequency.
- You can appreciate that the scattering is quite complex by considering what happens when the diffracting plane is exactly parallel to the incident beam: as we've mentioned, the two Kikuchi lines will both be visible although you might have guessed otherwise.

Back in Chapter 6 we noted that electron ray paths rotate through the objective-lens field, but in all our

discussion on diffraction (including Kikuchi lines and the CBED patterns in the next two chapters) we draw all the electron paths as straight lines, ignoring any rotation. However, particularly in a modern condenser-objective lens TEM, the lens field is relatively strong and can introduce a significant rotation into the off-axis incident and diffracted electrons (we described c/o lenses and their effects on ray paths back in various sections of Chapter 9). An interesting consequence of this effect is that Kikuchi lines in modern TEMs may be less sharp than in older TEMs, unless you illuminate only a very small area of the specimen. If you're intrigued by this problem then you must read "Skew thoughts on parallelism" by Christenson and Eades.

CHAPTER SUMMARY

Pairs of Kikuchi lines define the road and, taken together, the roads make up a map of reciprocal space. However, the key is different from real-space road maps because in our Kikuchi maps, narrow roads are the most important! What is the relevance of the roadside curbs? Well, they define the roads and tell us when we are standing on them, but we are not too interested in their detailed appearance. You should by now be under the (correct) impression that we view Kikuchi lines and maps as an invaluable tool for the microscopist. The key points are

- The Kikuchi lines arise from Bragg diffraction of divergent, incoherent electrons scattered within the specimen.
- The Kikuchi lines consist of an excess (bright) line and a deficient (dark) line. In the DP, the excess line is farther from the direct beam than the deficient line.
- The Kikuchi lines are effectively fixed *to the crystal* so we can use them to determine orientations accurately.
- The trace of the diffracting planes is midway between the excess and deficient lines.
- We can use the Kikuchi lines to set up specific (e.g., two-beam) diffracting conditions which are central to diffraction-contrast imaging.
- We can control and determine the value of $\mathbf{s_g}$ by measuring the separation between the **g** Kikuchi line and the G reflection (the separation is 0 when $\mathbf{s_g} = 0$). The precise value of $\mathbf{s_g}$ is also very important in controlling diffraction contrast.

Kikuchi lines and maps are two of the most important aids you have when orienting (or determining the orientation of) crystalline specimens. Knowing the orientation of your specimen is essential for any form of quantitative TEM, whether you're analyzing dislocation Burgers vectors by diffraction contrast, imaging grain boundaries with lattice resolution or measuring chemistry variations by EELS or XEDS. Kikuchi maps are especially useful when combined with the map of zones and poles (directions and plane normals) on the stereographic projection. Use a computer to check or to assist you in constructing such a map for your material but if you're doing any serious crystallography, never leave home or sit at your TEM without a map to guide you.

IN GENERAL

Edington, JW (1976) *Practical Electron Microscopy in Materials Science*, Van Nostrand-Reinhold New York. Part 2 of the book is still an excellent source of guidance if you are doing hands-on DP acquisition and analysis and the Appendix has great Kikuchi maps.

Schwartz, AJ, Kumar, M and Adams, BL (Eds.) (2000) *Electron Backscatter Diffraction in Materials Science*, Springer NY. Kikuchi patterns in the SEM: an insight into what could really be done in terms of orientation determination in TEM, if we put our minds to it.

Thomas, G (1978) in *Modern Diffraction and Imaging Technique in Materials Science* p 399 Eds. S Amelinckx, R Gevers and J Van Landuyt North-Holland Amsterdam. Still the best reference because no one has really done anything of significance with Kikuchi patterns in TEM since Gareth Thomas, 40 years ago (check out the other references below).

SOME HISTORY AND APPLICATION

Christenson, KK and Eades, JA 1988 *Skew Thoughts on Parallelism* Ultramicroscopy **26** 113–132.

Dingley, DJ 1984 *On-Line Determination of Crystal Orientation and Texture Determination in an SEM*. Proc. Royal Microsc. Soc. **19** 74–75. The idea of texture mapping developed for the SEM can be applied to the TEM.

Dingley, DJ 2004 *Progressive Steps in the Development of Electron Backscatter Diffraction and Orientation Imaging Microscopy* J. Microsc. **213** 214–224.
Johari, O and Thomas, G 1969 *The Stereographic Projection and its Applications* in *Techniques of Metals Research* Ed. RF Bunshah Interscience New York.
Kikuchi, S 1928 *Diffraction of Cathode Rays by Mica* Japan J. Phys. **5** 83–96.
Levine, E, Bell, WL and Thomas, G 1966 *Further Applications of Kikuchi Diffraction Patterns; Kikuchi Maps* Appl. Phys. **37** 2141–2148.
Okamoto, PR, Levine, E and Thomas, G 1967 *Kikuchi Maps for H.C.P. and B.C.C. crystals* J. Appl. Phys. **38** 289–296.
Ryder, PL and Pitsch, W 1968 *On the Accuracy of Orientation Determination by Selected Area Electron Diffraction* Phil. Mag. **18** 807–816.
Tan, TY, Bell, WL and Thomas, G 1971 *Crystal Thickness Dependence of Kikuchi Line Spacing* Phil Mag. **24** 417–424.

URLS

1) http://cimewww.epfl.ch/EMYP/comp_sim.html

THE COMPANION TEXT

The main chapters in the companion text that relate to this topic are those on CBED and EFTEM.

SELF-ASSESSMENT QUESTIONS

Q19.1 When viewing different regions of your specimen, the intensity of the Kikuchi lines changes. Could the reason be the variation in thickness?
Q19.2 Why is the Kikuchi line nearer 000 brighter than its partner that is farther away?
Q19.3 You record a series of DPs as you tilt along a Kikuchi band. You paste the DPs together but the Kikuchi lines appear to be slightly curved Kikuchi lines, but Kikuchi lines are drawn straight. Explain.
Q19.4 When viewing another DP, a pair of parallel lines is present where one line is dark and the other is bright. Why does this happen, and how do we name each line?
Q19.5 You're writing your dissertation on the beach in Aruba so you don't have crystallography software available but the crucial DP of a certain fcc specimen stored on your laptop is only showing Kikuchi lines. Can you determine the orientation? If so, how can you do it?
Q19.6 Where is the **g** Kikuchi line relative to O and G if the excitation error, $\mathbf{s}_\mathbf{g}$, is less than zero?
Q19.7 How does the distance between Kikuchi lines vary with specimen thickness?
Q19.8 How accurately can you determine orientations using Kikuchi lines?
Q19.9 Will a conventional TEM with a LaB_6 filament give better Kikuchi lines than a new FEGTEM at the same orientation and thickness?
Q19.10 We say that Kikuchi lines arise due to incoherently scattered electrons? This statement is a little oversimplified. Why?
Q19.11 Why are Kikuchi patterns used for setting the value of **s**?
Q19.12 What are Kossel cones?
Q19.13 What is the distance between a pair of Kikuchi lines?
Q19.14 Can you just read off **s** from where the Kikuchi line cuts the systematic row?
Q19.15 Why is there an ideal thickness for a specimen when viewing Kikuchi diffraction?
Q19.16 How can we trace the location of a plane from Kikuchi lines?
Q19.17 Briefly describe how you can find poles using Kikuchi lines.
Q19.18 Is EBSD possible in the TEM?
Q19.19 What is the distance in reciprocal space between the $-5\mathbf{g}$ and $+5\mathbf{g}$ Kikuchi lines?
Q19.20 How do HOLZ lines and Kikuchi lines differ? (Read Chapters 20 and 21 before answering this.)
Q19.21 Construct a Kikuchi map where the [001] pole is exactly on the optic axis.
Q19.22 Why do we only need the Kikuchi map for the [001], [101] and [111] triangle in fcc crystals but have to map much more of reciprocal space for hcp, and what is the largest area we would have to map for any crystal?

TEXT-SPECIFIC QUESTIONS

T19.1 Assuming that Figure 19.1 was obtained from a cubic material, determine the approximate orientation of the specimen.
T19.2 Redraw Figure 19.4A for the [011] and [111] poles but arrange for two low-index reflections to be excited in each figure.
T19.3 Consider Figure 19.6. Sketch and label the Kikuchi bands which pass through the 115 pole.
T19.4 Using Figure 19.6 draw consistent DPs for the 102, 116 and 013 poles as you would expect to see them as you tilt your fcc sample. (So you have to align them appropriately too.)

T19.5 Using Figure 19.7 draw consistent DPs for the $1\bar{2}13$, $1\bar{1}02$ and $0\bar{1}12$ poles as you would expect to see them as you tilt your hcp sample. (So you have to align them appropriately too.)

T19.6 In Figure 19.10, determine what the pole is and how far it is inclined relative to the electron beam. How would you tilt the sample to bring P onto the optic axis (assuming O is on the optic axis)?

T19.7 Consider Figure 19.10. Identify and index 10 pairs of Kikuchi lines (even if you can only see one of the lines).

T19.8 In Figure 19.11, where are the diffracting planes? Why can we see the Ewald sphere in two places? Which one is correct as the diagram is drawn?

T19.9 Sketch Figure 19.7 for Be showing the $1\bar{1}02$, $1\bar{2}13$, $0\bar{2}23$ and $\bar{1}\bar{1}23$ poles. Now superimpose on this figure the one for Ti as if the Ti had grown in perfect alignment on 0001 Be.

T19.10 Choose a 010 sample of olivine and construct the Kikuchi pattern to a radius of 45° about this pole. Label the Kikuchi bands and sketch the main DPs you will find. (You should use the Web or ICDD files to help you with this question.)

T19.11 An Al-crystal (fcc) is observed exactly along a [011] direction. Draw the corresponding DP with the associated Kikuchi lines (bands). After a small tilt of the crystal, the Kikuchi lines pass through three of the low-index reflections. How far has the crystal been tilted and what was the tilt axis? The lattice constant for aluminum is 0.405 nm. The wavelength for the electrons is 0.0025 nm. (Courtesy Anders Tholen.)

T19.12 Explain how you would deduce the accelerating voltage of the TEM using a DP from Si.

20

Obtaining CBED Patterns

CHAPTER PREVIEW

We know that SAD, while giving us useful information about the specimen, has two severe limitations.

- We have to be very cautious in interpreting SADPs from areas which are $< \sim 0.5\,\mu m$ in diameter because the information in the pattern may not be limited to that region. This scale is still large compared to the dimensions of many crystalline features that interest us in materials science and much larger than nanotechnology dimensions.
- SADPs contain only rather imprecise 2D crystallographic information because the Bragg conditions are relaxed for thin specimens and small grains within the specimen (see Chapter 17).

The technique of CBED overcomes both of these limitations and also generates much new diffraction information which we will introduce to you in Chapter 21 and expand in greater depth in the companion text.

In this chapter, we will show you how simple it is to use the versatility of modern TEMs to create a range of CBED patterns containing a variety of very useful contrast effects such as higher-order Laue-zone (HOLZ) spots and lines. In Chapter 21 and the companion text, you will see why these HOLZ features are so useful. They can give us a complete 3D crystallographic analysis of our specimens. By now you are well aware that a major advantage of doing anything in TEM is that we can get the information at high spatial resolution at the same time as we are looking at the TEM image. CBED is no exception. For most TEM operations, thinner specimens are better, but CBED patterns, like Kikuchi patterns, are generally more useful from thicker specimens in which dynamical scattering is occurring. Finally, as with many other sophisticated analytical techniques, CBED uses various obscure initials and acronyms which we will attempt to clarify as we introduce them.

20.1 WHY USE A CONVERGENT BEAM?

Historically, CBED is the oldest electron-diffraction technique used in the TEM. It was originally developed by Kossel and Möllenstedt, well before LePoole developed SAD. While SAD is the classic way to relate the diffraction-contrast information in the TEM image to the specimen orientation, it has a notable disadvantage. Remember, we saw back in Chapters 9 and 11 that the diameter of the smallest area you can select by SAD is $\sim$ 0.5 μm, with an uncertainty of similar dimensions. However, if you have an intermediate-voltage HRTEM with a very low C_s you may be able to extend the range of SAD to analyze areas ~100 nm in diameter, which is still too large for examining any nanoscale material (particles, films, devices) as well as all crystal defects and most second-phase precipitates which influence the properties of conventional engineering materials. All these features are much smaller than 0.1 μm. As we described back in Figure 9.4, one way we can overcome this limitation is to use a convergent beam of electrons. The region sampled by the convergent beam is a function of its size and the beam-specimen interaction volume, which increases with specimen thickness but is generally a lot smaller than the spatial limitations of SAD. In fact, several so-called micro- and nanodiffraction methods have been developed over the past 40 years to overcome the spatial-resolution limitations of SAD in a TEM. But CBED is by far the most simple and versatile of these techniques and easily

penetrates the nanotechnology domain (< 100 nm), in fact all the way down to single unit-cell dimensions (truly nanodiffraction).

In addition to offering improved spatial resolution, CBED patterns contain a wealth of crystallographic and other data which are difficult or impossible to obtain via SAD and we'll describe much of this new information in the next chapter and in the companion text.

With such capabilities, CBED has transformed diffraction in the TEM from the 'poor relative' of X-ray and neutron diffraction to a more versatile and, in some senses, a unique technique. CBED has already begun to intrude significantly into XRD's traditional domain of quantitative crystallography, as a cursory reading of the contents of *Acta Crystallographica* will show.

Some would argue persuasively that CBED should be the diffraction method of first choice when operating a TEM and this is not surprising given all the new information you can access via CBED. However, much of diffraction theory, indexing methods, etc., were historically developed for parallel-beam SADPs, and standard BF/CDF and other imaging techniques build on such patterns. So we have chosen to construct the text starting with SAD and treat CBED as an 'enriched' variant of SAD.

THE CBED ADVANTAGE

. . .over all other diffraction techniques is that most of the information is generated from minuscule regions beyond the reach of other diffraction methods.

In this chapter, we will concentrate on how you can control the experimental variables to acquire and index CBED patterns. In the next chapter, we will introduce to you, among other things, how to perform the basics of what is known as 'electron crystallography.' All the advantages of CBED can simultaneously be coupled with TEM images and spectrometry data thus allowing you to achieve a remarkable degree of characterization of the same region of your specimen at the same time.

There are two potential drawbacks which you should always keep in mind

- The focused probe may generate contamination which can cause localized stresses.
- The convergent beam may heat or damage the chosen region of your specimen.

In early probe-forming TEMs, the operator only had a few seconds to observe and record the CBED pattern before carbon contamination built up to a thickness which masked all the information. Modern TEMs, particularly UHV instruments, should not suffer from this problem (see Chapter 8). You should be able to study small regions of a clean specimen for many minutes or even hours without contamination.

CONTAMINATION

Most contamination is caused by the specimen-preparation process or your careless handling of your specimen.

As we described in Chapter 4, beam heating/damage may be a problem in materials with poor thermal conductivity but this can be minimized by applying a thin conductive coating or, preferably, by using a liquid-N_2 cooling holder. This latter approach has other advantages for CBED, as we'll see.

You will probably find that it is experimentally easier to do CBED rather than SAD. SAD is used if your primary interest is the image. In CBED, there's no inserting and removing of diaphragms and more information is almost always obtained, invariably from smaller regions. You might not find this fact reflected in your reading of the TEM literature, but remember that researchers (including the authors) often use the technique with which they are most familiar, rather than the best one available. It's also appropriate here to mention why we include the 'E' in CBED but rarely in SAD; it's only so we can pronounce the initials as "see-bed." (Happily no one ever says "sad" or "said" for that matter.) But if the world were logical then all men would ride sidesaddle.

THE CBED HOLDER

A double-tilt, cooling, low-background holder is really useful for CBED.

20.2 OBTAINING CBED PATTERNS

First, which holder should you use? As with SAD you will need to do a lot of tilting so a double-tilt holder is required. Some of the diffraction phenomena we will be describing become more visible if the specimen is cooled to liquid-N_2 temperatures. If you want to carry out XEDS and CBED simultaneously then you'll also need a low-background holder.

Single-tilt rotation holders can sometimes be advantageous (e.g., to line up interfaces in certain directions) and again are best if they can be cooled and are low background (for XEDS).

Second, which kV should you use? In diffraction, unlike most other TEM techniques, there is often an advantage to using lower kVs since the elastic-scattering cross section is higher. So the patterns are more intense. Also the Ewald sphere is more curved at lower kVs and so it intercepts more electrons scattered at higher angles than those in the ZOLZ. However, higher kVs give higher-resolution information from smaller volumes of your specimen and you can 'see' through thicker specimens. So you may have to compromise and gather patterns over a range of kVs.

Third, how thick should your specimen be? As we already mentioned, unlike most other TEM techniques, thicker specimens can be more useful than thinner specimens since more dynamical scattering occurs. Much of the undesirable diffuse scattering that hides the dynamical information can be filtered out in an EFTEM (much more about this later and in Section 38.2).

20.2.A Comparing SAD and CBED

Now let's consider the differences in the electron optics of SAD and CBED. In SAD the electron beam incident on your specimen is parallel (fixed incident vector **k**) and relatively large (usually ~1–10 μm in diameter). In CBED, the beam is convergent (range of **k** vectors) and relatively small (usually ~1–100 nm in diameter) as shown in Figure 20.1. (Compare this situation with Kikuchi patterns which are generated by a divergent beam of electrons within the specimen.) We've already seen in Chapters 11 and 16 that parallel illumination means that the SADP consists of an array of sharp maxima (or spots) in the BFP of the objective lens and, as mentioned in Chapter 9, the beam convergence in CBED gives rise to a pattern of disks of intensity. Figure 20.2A shows an SADP from pure Si and Figure 20.2B is a CBED pattern from a much smaller region of the same Si specimen. While it isn't obvious that the CBED pattern comes from a smaller region of the specimen (i.e., has better *spatial* resolution), you can certainly see that it contains a wealth of contrast detail not present in the SADP. The dark lines within the 000 disk and between the diffraction disks are HOLZ effects which we'll return to many times throughout this and the next chapter. Converging the beam, in effect, opens up a greater angular range of reciprocal space for you to look into and in doing so it reduces the *angular* resolution of the DP (which is not something we worry about too much).

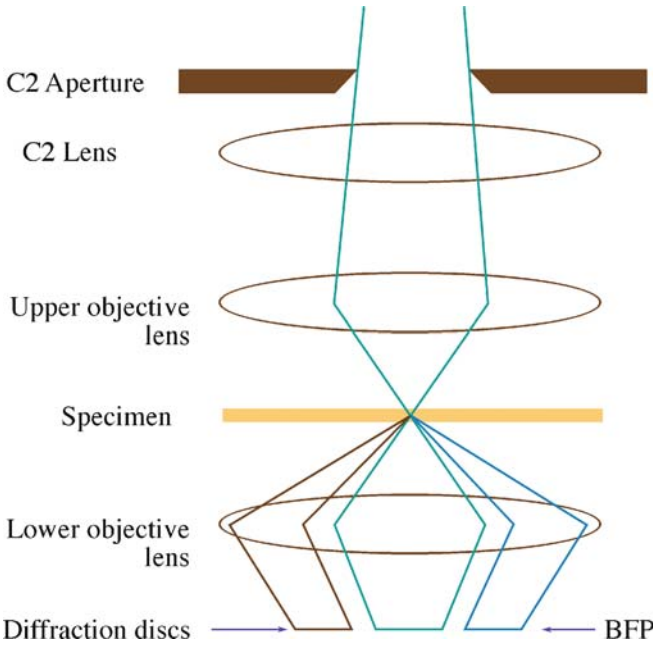


FIGURE 20.1. Ray diagram showing CBED pattern formation. If the c/o lens system focuses the beam at the specimen, the illuminated area is very small compared with parallel-beam SADP formation. A convergent beam at the specimen results in the formation of disks in the BFP of the objective lens.

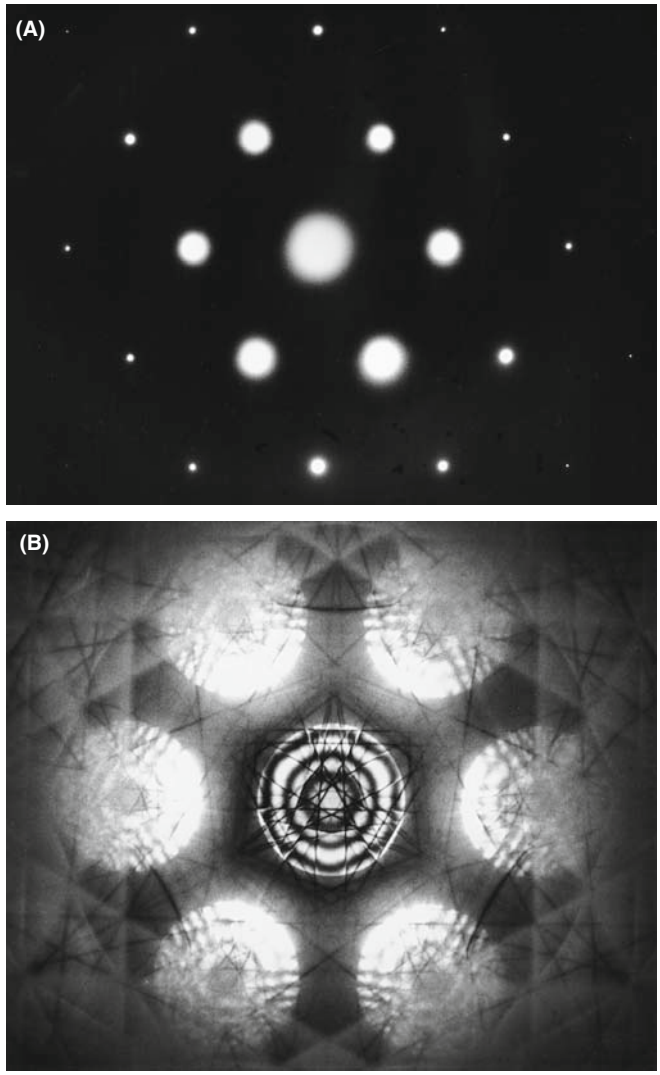


FIGURE 20.2. (A) SADP from [111] Si showing the first few orders of diffraction spots and no visible Kikuchi lines. (B) CBED pattern from [111] Si showing dynamical contrast within the disks as well as diffuse Kikuchi bands and sharp, deficient HOLZ lines.

RELATING SAD AND CBED

You can imagine CBED as magnifying the information within the spots in the SAD. Like SAD, CBED is most useful when you either orient the beam along a zone axis in the crystal to give a symmetrical zone-axis pattern (ZAP) or when you tilt to strong two-beam conditions. Unlike SAD, CBED is also even more useful when the probe is deliberately under- or overfocused.

From Figure 20.1 you can see that we need a condenser/objective (c/o) lens (i.e., a strong upper-objective polepiece) to create a convergent beam, so any probe-forming TEM or dedicated STEM can generate such patterns. We've already described the lens systems and ray diagrams associated with forming a convergent beam in Chapters 6 and 9, so here we will emphasize the experimental variables that you can control. The various steps we describe below are covered in far greater detail, both operationally and in terms of extraordinarily detailed ray diagrams, in the essential text by Morniroli. We'll start with TEM mode and then describe STEM operation.

20.2.B CBED in TEM Mode

It's a good idea to start practicing CBED with a thin, single-crystal Si specimen or a stainless steel foil with large grains, since these specimens give immediately useful patterns containing many of the features that we'll be describing (without the need for cooling). The Si is a little easier since the parent wafer will be in a specific orientation such as <111> and you'll invariably get a pattern like Figure 20.2B.

When you converge the beam in image mode, you'll lose any useful image contrast and you'll just see a bright spot on the TEM screen or computer display. But if you aligned the condenser system correctly then the beam will be focused on the region you chose. You will develop your own procedure as you gain experience. Basically the approach is as follows:

- Start with your specimen in the eucentric plane, as usual. Weaken C2 to give a broad, parallel beam, then focus an image with the feature from which you want to obtain the CBED pattern in the middle of the screen. It helps if you can do this without the objective diaphragm inserted. If you need it, then you'll have to take it out later (see below).
- Select a large C2 aperture about 100–200 μm in diameter, carefully center it, then adjust the C2 lens to converge the beam to a focused spot on the area of interest.
- Keep C1 weakly excited to give a relatively large spot, about 100 nm FWTM (see Chapter 5) which should contain sufficient current to give a high-intensity pattern.
- Select a small camera length, < 500 mm to give a wide-angle view of the pattern.
- To observe the CBED pattern, just switch to diffraction mode, making sure the objective and SAD diaphragms are retracted.

Remember that you control the minimum illuminated area on the screen (i.e., the beam diameter at the specimen) by the strength of the C1 lens, so after these initial steps, you might want to select a smaller probe size by increasing the C1 strength.

20.2.C CBED in STEM Mode

You should first get a focused STEM image of the specimen as we described in Section 9.4.

The procedure is quite simple because in STEM you're always operating with a focused probe and you don't have imaging-system diaphragms to worry about

- First, stop the beam scanning (i.e., select 'spot' mode on the STEM control system).
- Second, position the spot on the STEM screen on the region of interest.

A CBED pattern should then be visible on the TEM screen, but to see it you may have to remove the STEM detector if it sits above the TEM screen, or lower the TEM screen if the detector is below. The CBED pattern is present because the TEM is operated in the diffraction mode during STEM operation. As before, you may have to reduce L to ensure that several orders of diffraction maxima are visible on the screen. The other variables are the same, except that in STEM, the C2 lens in some TEMs is automatically switched off. This means that the C2 aperture alone governs α and you can only focus the pattern with the objective lens.

STEM AND CBED

In TEM you can't see the area of the specimen you have chosen without spreading the convergent beam; in STEM you can always scan the convergent beam to see the image.

In a DSTEM you can see both image and DP at the same time because the CBED pattern can be viewed by introducing a screen after the last post-specimen lens and viewing this screen with a TV camera. A hole in the screen allows any selected portion of the pattern to travel through the EELS to the BF detector and thus both image and DP can be viewed simultaneously. If you don't have post-specimen lenses then you can't vary L; the CBED pattern is then viewed either directly using a TV camera to image the BFP of the objective lens or by scanning the pattern across the BF detector using post-specimen scan coils (see Section 21.8).

If you send the CBED pattern through the EELS before viewing it, then you can switch on the EELS and use it to filter out energy-loss electrons, or form the patterns with electrons of a specific energy. Energy-filtered CBED (which you can do in a TEM as well as a STEM) is a most powerful technique, as we'll see (look ahead to Figure 20.10).

The choice of operating mode then is really up to you; TEM and STEM both have their advantages.

20.3 EXPERIMENTAL VARIABLES

To form CBED patterns, you have to create a small (< 100 nm) beam with a convergence semi-angle (α) $\geq$10 mrads. There are at least five microscope variables you need to control when forming a CBED pattern

- The beam-convergence angle α (remember, we mean semi-angle)
- The camera length L (i.e., the magnification of the pattern)
- The size of the beam (the probe diameter)
- The thickness of the specimen
- The focus of the pattern (under/over or exact)

The last variable is the most complex because there are CBED techniques which require focused patterns and those where we deliberately defocus the beam out of the specimen plane to gain other advantages. So we'll deal with this last variable separately. You can also change the kV if you want and as we noted above, low kV gives better contrast, but this is more than offset by the drop in gun brightness and the increased beam spreading. So, for materials specimens, most CBED is done at the highest kV, as with most other TEM techniques.

20.3.A Choosing the C2 Aperture

After you've read the next few sections you'll probably conclude that any full CBED study should include patterns taken with a range of camera lengths, a range of exposures, a range of C2 apertures and that the more C2 apertures you have in your TEM the better. You'll be right on all counts!

Once the CBED pattern is visible on the screen you can adjust α by changing the C2 aperture, making sure to center the aperture you finally choose. The size of the diffraction disks depends on α as shown in Figure 20.3 (go back and check equation 5.6). Let's start by choosing a small aperture so we get a pattern of discrete disks, similar to an SADP and we'll use the terminology first introduced by Steeds.

KOSSEL AND MÖLLENSTEDT

The pattern of non-overlapping disks is a Kossel-Möllenstedt (K-M) pattern. The pattern of completely overlapping disks is a Kossel pattern.

To get a K-M pattern you must select a C2 aperture such that the disk size (governed by α) is less than the spacing of the disks (governed by θ_B) for the particular specimen and orientation. Typically, the Bragg angle is a few milliradians, and C2 apertures in the 10–50 μm range will usually ensure that you have satisfied the K-M conditions.

If α is large enough for substantial overlap of the disks to occur such that individual diffraction maxima are no longer discernible, then the term 'Kossel pattern' is used (although this can cause confusion with the classic use of the term for geometrically similar X-ray patterns). Figure 20.3A–C shows a series of ray diagrams illustrating the transition from a K-M pattern to a Kossel pattern by increasing 2α. Equivalent experimental patterns from pure Al are shown in Figure 20.3D–F. The patterns in Figure 20.3 were all taken at a small camera length and, although the contrast is not particularly strong, you can clearly see rings of intensity which arise from electrons scattered to quite high angles ($\pm 10^\circ$). We'll continue to read more about these HOLZ diffraction effects throughout this chapter.

Kossel patterns are most useful when viewed with a small camera length (see next section) because they display a relatively extensive area of reciprocal space, and the large value of α gives rise to Kikuchi bands. For reasons we'll describe in Section 20.7, Kikuchi lines are much more prevalent in CBED than in SADPs and don't require the usual thick specimen. As we described in Chapter 18, the Kikuchi bands intersect in the center of the pattern when the beam is incident down a zone axis, as you can see in Figure 20.3F. Thus it is very easy for you to tilt to a particular zone axis, simply by following the bands until they intersect. So, to form a ZAP, it is best to start at very small L with a large α. Later, you can worry about the best choice of C2 aperture, best choice of L and focusing the pattern. Because these Kossel patterns cover a large angular range in reciprocal space they are also an example of a whole class of large-angle CBED or LACBED patterns (see Section 20.4.B below) in which we defocus the beam to enhance the contrast compared with the Kossel patterns.

Because we need to be able to vary α, a range of C2 apertures from about 10 μm up to 200 μm is desirable, consistent with the needs of other techniques. A reasonable choice if you have only three C2 apertures: one of about 200 μm for routine TEM, EELS and Kossel patterns, a 50–70 μm ultra-thick aperture for XEDS (which can also be used for STEM imaging and some K-M patterns) and a 10–20 μm aperture for most K-M patterns. Some TEMs provide more than three apertures. More is better!

Because the C2 lens is excited in TEM mode, you can use it to change α; if you do, the objective lens has to be changed to maintain a focused pattern. You need to adjust C2 if you change the beam size with the C1 lens or if you want a value of α between those given by the fixed C2 apertures.

STAY IN FOCUS!

Use the specimen-height (z) control to maintain the specimen in the eucentric plane as you tilt. A computer-controlled stage is a great advantage.

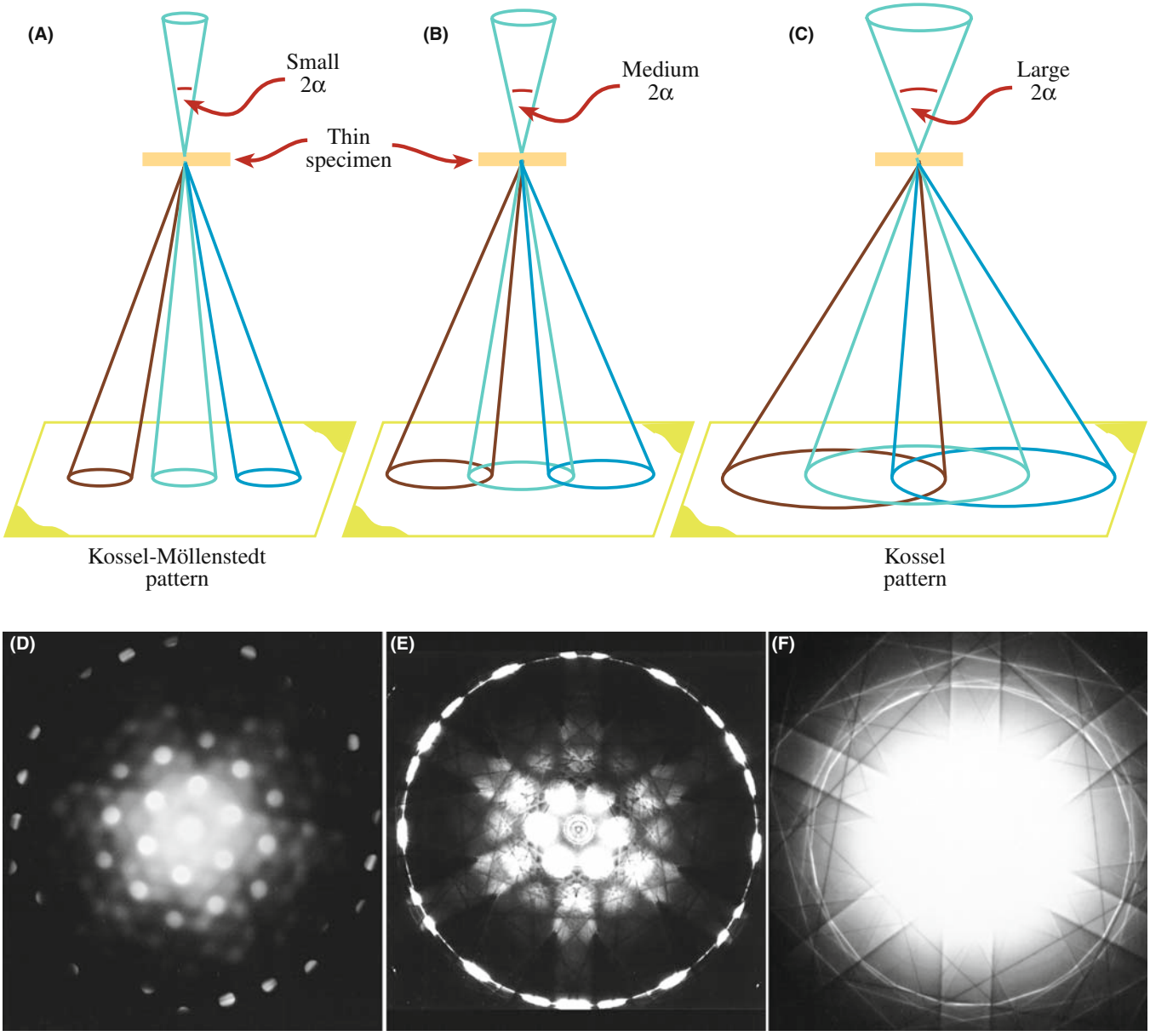


FIGURE 20.3. (A–C) Ray diagrams showing how increasing the C2 aperture size causes the CBED pattern to change from one in which individual disks are resolved (K-M pattern) to one in which all the disks overlap (Kossel pattern). (D–F) You can see what happens to experimental CBED patterns on the TEM screen as you select larger C2 apertures.

If you need to know the value of α, you should use a known crystal to calibrate its variation with C2 aperture size for typical C2 lens excitations, as we described back in Section 9.1 and in equation 5.6.

20.3.B Selecting the Camera Length

The choice of L depends on the information that you want to obtain from the pattern and it's easy to be confused because L controls the magnification of the DP.

Typically we adjust the post-objective lenses in the imaging system to give $L > 1500$–6000 mm when we want to observe detail in the 000 (BF) disk at the highest possible magnification. We reduce L to < 500 mm to view the low-magnification pattern, sometimes called (not surprisingly) the 'whole pattern' (WP) that contains electrons scattered out to high angles. Figure 20.4 shows three CBED patterns obtained over a range of L and you can see that if we start at a high L we can only see the 000 disk (Figure 20.4A), then we see the array of ZOLZ disks that is equivalent to an SADP (Figure 20.4B) but at the smallest L the HOLZ diffraction effects that we just mentioned become visible as a ring of intensity at high angles (Figure 20.4C). So it's often necessary to record your CBED patterns over a range of L (in addition to a range of α).

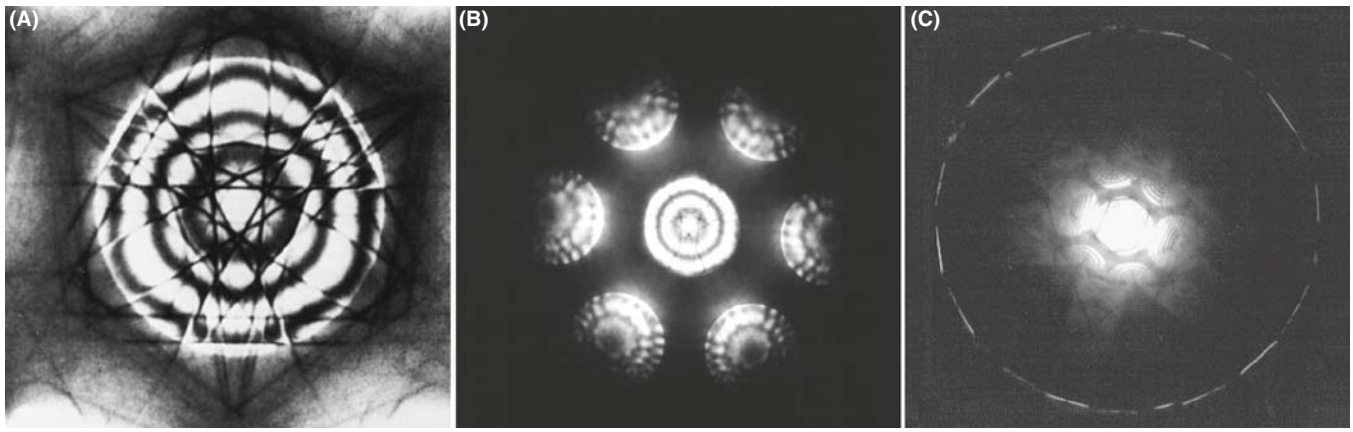


FIGURE 20.4. *Decreasing* the camera length, *L*, *increases* our view of reciprocal space. (A) Starting at high *L*, we see a CBED pattern containing only the 000 diffraction disk. As *L* decreases, we see in (B) the distribution of electrons in the ZOLZ, similar to a typical SADP. At the shortest camera length, (C) a ring of HOLZ intensity is faintly visible surrounding the bright ZOLZ disks. Typically, we can record electrons scattered over an angular range of $\pm 10^{\circ}$.

LARGE *L*, SMALL ANGLE

A *large L* gives a *high*-magnification view of the pattern but only spans a *small* angular range of reciprocal space.

In dedicated STEMs that don't have post-specimen lenses, the CBED pattern is projected onto the BF and/or ADF detector at a fixed magnification.

20.3.C Choice of Beam Size

We've already mentioned that you should start with a reasonably large beam with enough current to give a good intense pattern on the screen. Of course, a large beam size doesn't help if the crystal you're trying to analyze is small. The volume sampled by the beam defines the spatial resolution and so it is important to control the beam diameter. For the thinnest specimens, spatial resolution is close to the beam size but, in thicker specimens, elastic scattering will spread the beam and degrade the resolution in a manner similar to XEDS (see Chapter 36). Using the thinnest specimens and an FEG, CBED patterns can be obtained from extraordinarily small regions; we'll see in Section 21.8.B that sub-nanometer diffraction is possible. However, there is a drawback to using the thinnest specimens because they don't exhibit dynamical-diffraction effects which, as you'll soon see, are really useful.

20.3.D Effect of Specimen Thickness

If your specimen is very thin you may have kinematical-diffraction conditions. Then the diffraction disks have a uniform intensity and are devoid of contrast, as shown in the ZAP in Figure 20.5A. Moving to a thicker area of the specimen in the same orientation transforms the pattern from an array of kinematically bland disks to a display of striking dynamical contrast (Figure 20.5B), which we'll discuss in great detail. So to get the most out of a CBED pattern, your specimen should be thicker than one extinction distance (see Chapter 16). As we've now mentioned many times, this requirement differs from that of almost all other TEM techniques, such as HRTEM, XEDS and EELS where the best information is obtained from the thinnest foils. So if your specimens are too thick for anything else, you can almost always get something useful out of them with CBED! As we'll discuss in Section 20.5, you can always energy-filter your CBED patterns from thick specimens and reduce the diffuse-scattering background to enhance the useful dynamical contrast.

20.4 FOCUSED AND DEFOCUSED CBED PATTERNS

There are times when you need to focus the CBED pattern and times when you have to defocus it. If you think about it, a focused CBED pattern will always come from the smallest possible region of your specimen and if you don't focus your patterns you will miss a lot of the fine detail! However, if the beam is defocused then what happens is that some image information appears in the CBED pattern (in an equivalent way to creating multiple DF images, which we used to calibrate the SADP rotation back in Section 9.6). Thus, in defocused CBED techniques we sample both direct and reciprocal space. In addition, slightly defocusing the objective lens increases the contrast in the patterns. Both of these advantages turn out to be particularly useful for direct analysis of crystal defects such as dislocations and grain boundaries using large-angle CBED (LACBED).

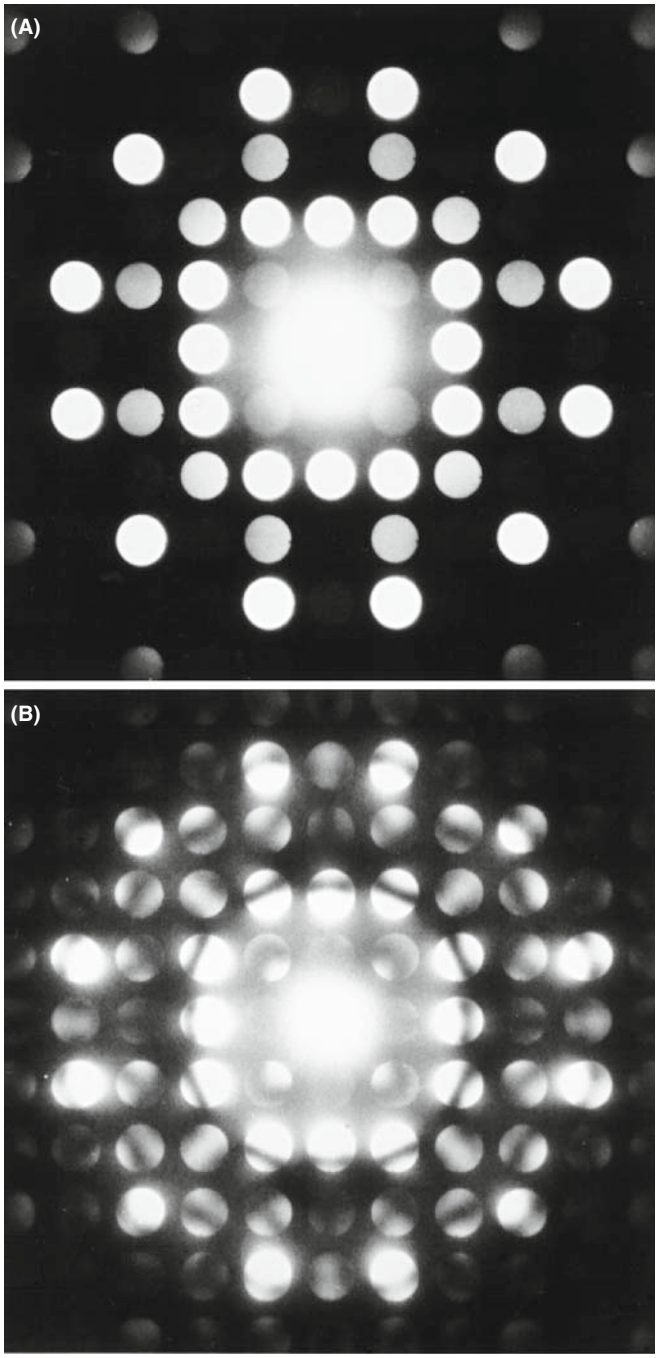


FIGURE 20.5. (A) CBED pattern under kinematical conditions. Such patterns give us no more information than SADPs, and their only advantage over SAD is that they come from a smaller region of the sample. (B) CBED pattern from a thicker area of the same specimen in (A) showing detailed dynamical-contrast phenomena.

20.4.A Focusing a CBED Pattern

Your TEM image is focused if your specimen is at the eucentric height and you can adjust the C2 lens to form the smallest spot on the TEM screen before switching to diffraction mode. To focus the CBED pattern

- Select K-M conditions (choice of C2 aperture) and choose a value of L so you can clearly see the 000 disk.
- Deliberately underfocus (weaken) the objective lens until a BF image is visible in the disk. This is because the beam is now spread at the plane of the specimen; see Figure 20.6A and the equivalent ray diagram back in Figure 6.5C.
- Strengthen the objective lens. As the beam crossover moves toward the specimen plane the image expands to higher magnifications until it goes through an inversion point at exact focus (see Figure 20.6B, which is equivalent to Figure 6.5B).
- Overfocus and again you can see a BF image in the 000 disk, inverted with respect to the underfocused image (Figure 20.6C). As you can see in Figure 20.6B there is non-spatial (i.e., diffraction-contrast) information in the 000 disk when you are at focus. (As we'll see in the next section, there are in fact several CBED techniques which deliberately use either an under- or overfocused beam).

If you leave the objective-lens current fixed and focus the beam on the specimen by adjusting C2 you'll see a similar effect to that shown in Figure 20.6, because the two lenses are coupled in a c/o system. If you use the second (non-eucentric) tilt axis or move to another region of the specimen, you will probably have to refocus the pattern with the z control, unless you have a stage that is fully computer controlled.

OBJECTIVE LENS CURRENT

Know the value of the objective-lens current that focuses the beam at the eucentric plane in your TEM. If your CBED pattern is focused at a different value, then adjust the lens current and refocus with the z control to maintain eucentricity.

The CBED pattern also has to be correctly focused in the BFP and you can do this in the conventional manner using the intermediate-lens fine focus to sharpen the image of the C2 aperture. As you'll see, you can also defocus the pattern by moving the specimen above or below the eucentric plane and this is just fine, unless you need to tilt the specimen to set up specific diffraction conditions, in which case the loss of eucentricity can make tilting tedious because the image will shift as you tilt.

20.4.B Large-Angle (Defocused) CBED Patterns

The reason we want to defocus the objective lens is that in some cases, as with TEM images, the contrast is minimized

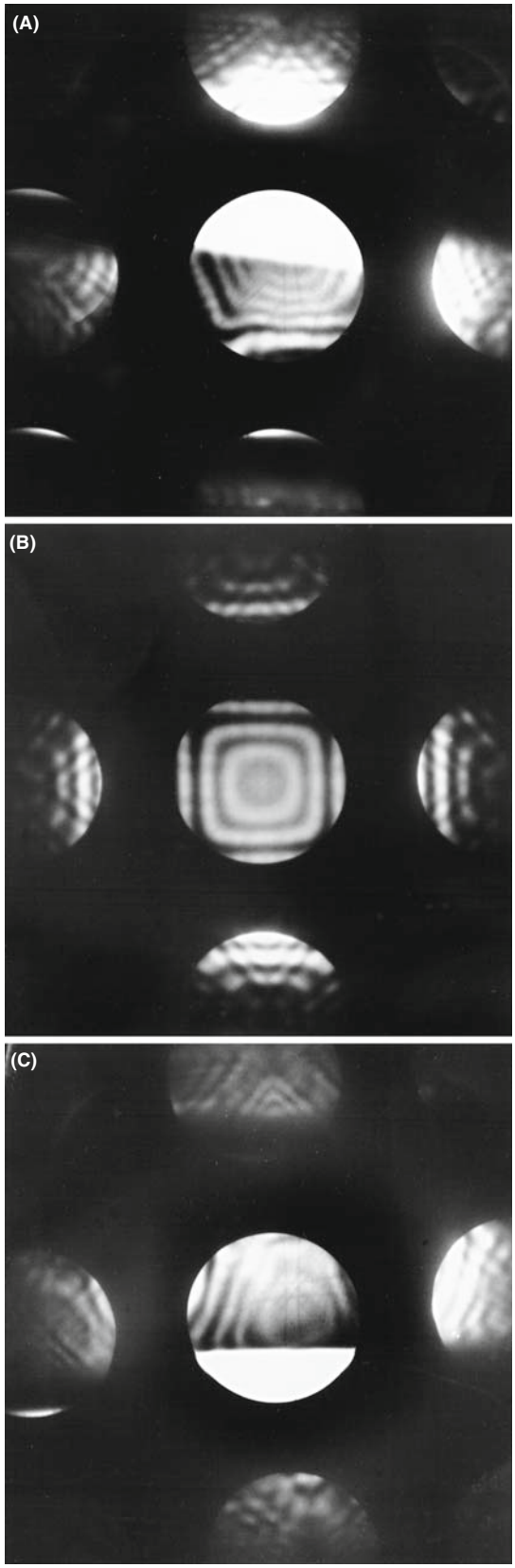


FIGURE 20.6. The procedure for correctly focusing the CBED pattern by adjusting the strength of the c/o lens through focus. In both underfocus (A) or overfocus (C) conditions, you see a BF image in 000 and DF images in the *hkl* disks, but at exact focus (B) the disks contain non-spatial, dynamical, diffraction contrast. Compare with Figure 6.5.

at exact focus. This is clear if you go back and look at the focused Kossel pattern in Figure 20.3F. At exact focus, the defect and excess Kikuchi lines overlap and reduce the overall contrast as shown in Figure 20.7A (although at the edge of the pattern, (i.e., where the electrons are traveling far off axis.) spherical aberration improves the contrast; a rare advantage for this otherwise-limiting lens defect). However, if you defocus the objective lens (either over or under) then the contrast in that same Kossel pattern is increased remarkably, as shown in Figure 20.7B. This

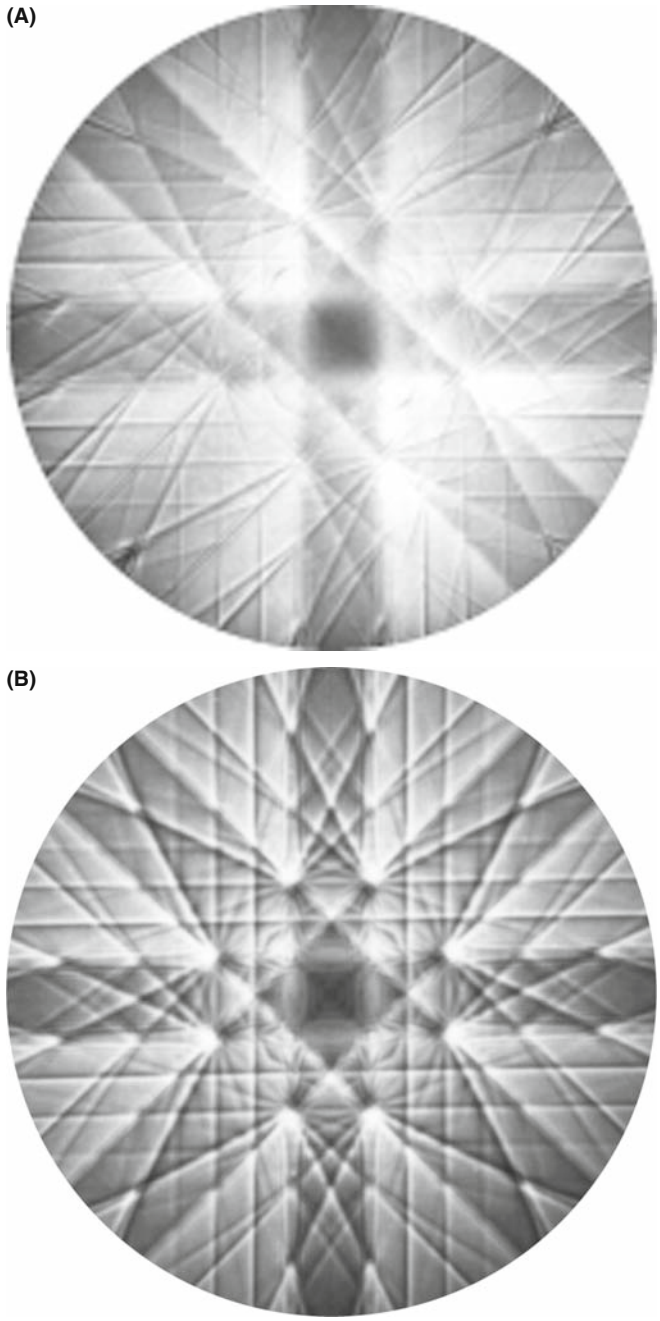


FIGURE 20.7. Kossel pattern from Si with the beam down a <001> zone axis at (A) exact focus and (B) with the diffraction lens weakened to view a plane before the BFP of the objective lens (i.e., overfocus).

novel approach was pioneered by Tanaka and co-workers in Japan (see the references). There are several forms of defocused CBED patterns and we'll summarize three here and show you some applications of the techniques in the companion text.

The first and most common defocus technique is large-angle CBED or LACBED (Tanaka et al. 1980). These patterns are extraordinarily useful, as a perusal of Morniroli's book will confirm, and they are invaluable in analyzing line and planar crystal defects such as dislocations and interfaces (e.g., Spence and Zuo). There are a couple of ways that you can form LACBED patterns but they are based on Kossel patterns which, as we've noted, display the largest area of reciprocal space because they use the largest possible C2 aperture. Now if you want to defocus the pattern, as we just described in Figure 20.6, it's simple enough: you can either change the height of your specimen by adjusting the z control on the stage or you can defocus the C2 lens so that the beam does not form a crossover in the plane of the specimen as shown schematically in Figure 20.8A. This is a beautiful but initially daunting diagram, characteristic of many by Morniroli, but well worth taking the time to understand.

Start by looking at the cone of incident electrons in which two planes (ABE and CDE) are indicated where the electrons satisfy the Bragg condition for $\pm hkl$ and these planes are bisected by the trace of the hkl plane itself. The beam is focused on the object plane of the objective lens and the specimen is raised a distance Δh out of that plane so it is illuminated by a disk of intensity. This has the effect of separating out the $\pm\ \mathbf{g}_{hkl}$ diffracted electrons. The electrons incident along the lines A_E-B_E and C_E-D_E are Bragg diffracted and form $\pm\mathbf{g}_{hkl}$ diffraction spots at K and L, respectively, in the object plane, and the direct-beam electrons form the 000 spot at E. All the electrons diverge through the spots and are re-focused by the objective lens to form spots in the image plane (K′, L′, and E′), respectively, and Figure 20.8B shows such a spot pattern. In the BFP there is a LACBED pattern consisting of a disk of intensity crossed by deficient lines (like Kikuchi lines but arising from elastic scattering) which are called Bragg lines (see Section 20.7 below). These deficient lines (A_F-B_F and C_F-D_F) correspond to the lines in the specimen along which strong diffraction occurred from the $\pm hkl$ diffracting planes as shown in Figure 20.8C and arise from the superpositioning of the $\overline{hkl}$ excess lines on the hkl deficiency line (A_F-B_F) and vice versa (C_F-D_F). As you can see, there are lines from many diffracting planes in this figure while, for simplicity, Figure 20.8A only includes the ray paths for two strongly diffracting planes exactly at $\mathbf{s} = 0$.

Figure 20.8A shows that the key to getting the best LACBED patterns is careful use of the SAD aperture. If you insert an SAD diaphragm into the image plane and permit only the direct beam to go through the aperture, the (BF) LACBED pattern you see on the screen is remarkably enhanced, as shown in Figure 20.8D (compare this with Figure 20.8C which is a LACBED pattern without the SAD aperture inserted). A DF LACBED pattern can be acquired simply by selecting one of the $\mathbf{g}_{hkl}$ reflections in the image plane with the SAD aperture and such a pattern consists of just a single $\pm hkl$ excess line of intensity. It is possible to improve the contrast in the LACBED pattern by choosing a smaller SAD aperture to cut out some of the inelastic, higher-angle scattering. If you look carefully at Figure 20.8D, you see that, as in all LACBED patterns, there is some real-space information about the specimen but the reciprocal-space information clearly dominates. Figure 20.9 shows a montage of BF and multiple DF LACBED patterns from Si in the [111] orientation. The beauty of such a pattern is only enhanced by the fact that the symmetry information can be used to determine the point group of the crystal directly; more about that in the companion text.

The second technique is called convergent-beam imaging or CBIM, which is a little younger than LACBED (Humphreys et al.). CBIM saw relatively little use because all it really did was give you an image of the area of your specimen from which you were obtaining the CBED pattern. If you include an objective aperture, a very small spot size and an energy filter the quality of the CBIM patterns is close to that of LACBED patterns although invariably CBIM emphasizes real-space over reciprocal-space information.

The last defocus technique in this acronym soup is parallel recording of dark-field images (PARODI) as a function of specimen thickness. The intensity variation in the multiple DF images visible in a systematic row of diffraction disks of a defocused CBED pattern (Wu et al.) can be recorded in a single exposure and, if you understand a lot of physics, you can use this approach to determine various properties of your crystalline specimens such as structure factors and valence-electron distributions.

There are several other mechanisms for obtaining LACBED patterns summarized by Morniroli, such as eucentric LACBED, specimen rocking, BF and DF LACBED and montages of CBED patterns, the last of which is illustrated in Figure 20.9.

20.4.C Final Adjustment

Sometimes, in either defocused or focused CBED patterns, it is quite difficult to make the ZAP exactly symmetrical as in Figure 20.2B. It often seems as if your last

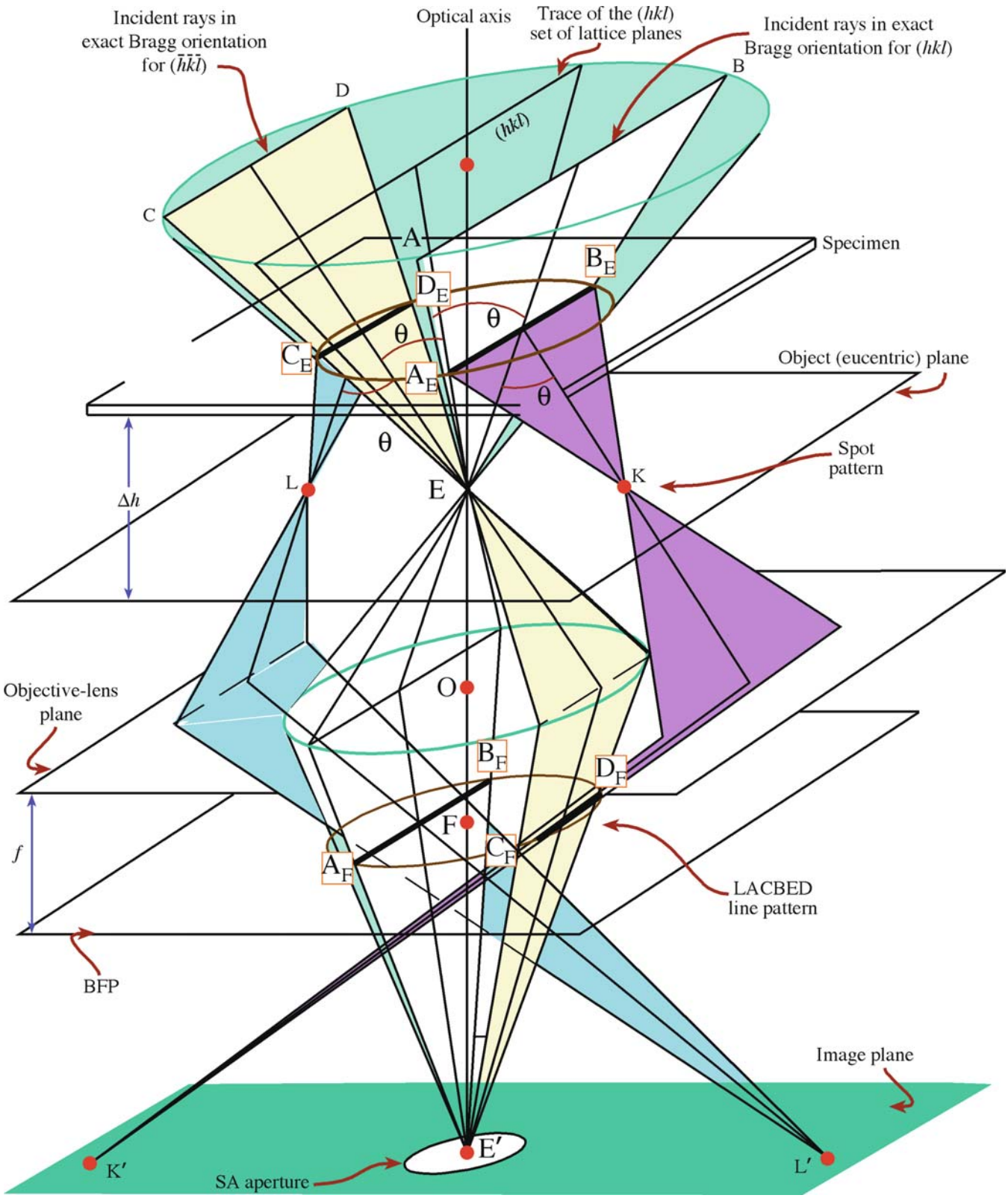


FIGURE 20.8 (A) Ray diagram for LACBED formation under $\pm h$ (three-beam) conditions. With the specimen raised by Δh from the eucentric plane in the objective lens, specific electrons in the incident cone are at the exact Bragg condition for diffraction from both sides of the *hkl* plane indicated, creating a spot pattern in the eucentric plane (B) and a LACBED pattern of lines in the BFP (C). The SAD diaphragm can be inserted to enhance contrast in the LACBED pattern by selecting only one beam to create (D) a BF LACBED pattern.

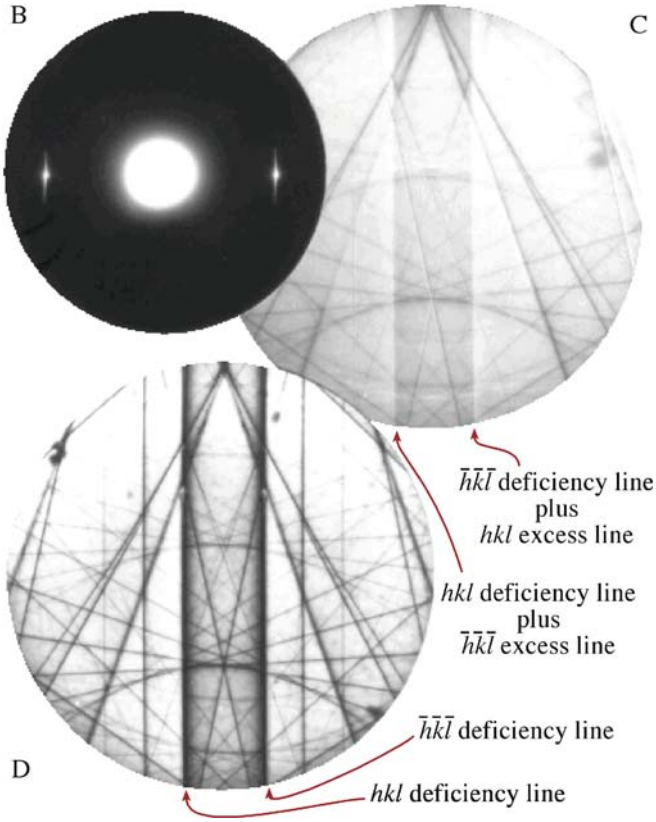


FIGURE 20.8. (Continued).

minor tilt or traverse of the specimen is not precise enough, or mechanical backlash occurs. If this is the case, use the beam tilts or deflectors to make your final adjustments to obtain a symmetrical pattern. In Section 18.2, we used the same method to excite high-order reflections in SAD. You can also move the C2 aperture and center it on the zone axis, but this misaligns the illumination system so it should be the last resort.

As with SAD, a range of exposure times for all CBED patterns will give you the most information.

We can now summarize the experimental steps to obtain a CBED pattern

- Focus the beam to a crossover on your specimen at the eucentric plane and go to diffraction mode in TEM (or stop the beam from scanning in STEM).
- Decrease L to see the full pattern including HOLZ scattering and tilt to the desired orientation.
- Adjust the convergence angle with the C2 aperture.
- Increase the beam size if necessary with the C1 lens to make the pattern brighter.
- Decrease the beam size and/or go to a thinner portion of the specimen to select a smaller region (which generally increases the quality of the pattern).
- Increase L to examine the 000 disk and focus the pattern.
- Defocus the pattern with the objective lens or raise/lower the specimen from the eucentric plane if you want to do LACBED, PARODI, CBIM or some form of defocused CBED.

20.5 ENERGY FILTERING

Every kind of DP contains electrons that have lost energy in going through the specimen. As we've seen already, these inelastically scattered electrons can be very useful if they cause Kikuchi lines, but if your specimen is thick enough, then the diffuse, inelastic scattering can raise the background intensity to levels that mask the useful contrast in the pattern. This is particularly the case in CBED patterns because, as we noted back in Section 20.2 and Figure 20.5, CBED is one of the few TEM techniques wherein more useful information is present in patterns from thicker specimens because they give rise to much more interesting dynamical contrast within the CBED disks.

So there's a balance here: we want thicker specimens to enhance dynamical effects but if the specimen is too thick the diffuse, inelastic scattering hides the useful contrast. If this is the case we can, in fact, have our cake and eat it at the same time. All we have to do is remove the diffuse-scattered electrons from the pattern using an energy filter.

TO FILTER OR NOT?
If you can energy-filter your CBED patterns, you should always do so.

Now we'll talk a lot more about energy filtering when we discuss EELS in the last four chapters of Part 4 and the seminal text on filtering is edited by Reimer. Filtering can be achieved either with an in-column or post-column filter. If we remove inelastically scattered electrons, the CBED pattern contains electrons that have lost no energy. Thus, in effect we have removed any specimen-induced chromatic-aberration effects from the pattern, and all the contrast phenomena in the pattern appear much sharper because all the electrons are focused in the (same) BFP (or image plane if it is a LACBED pattern). All this text pales in comparison with the example shown in Figure 20.10. This improvement is so dramatic that it really never makes sense to do anything other than filter your CBED patterns if you have the necessary instrumentation.

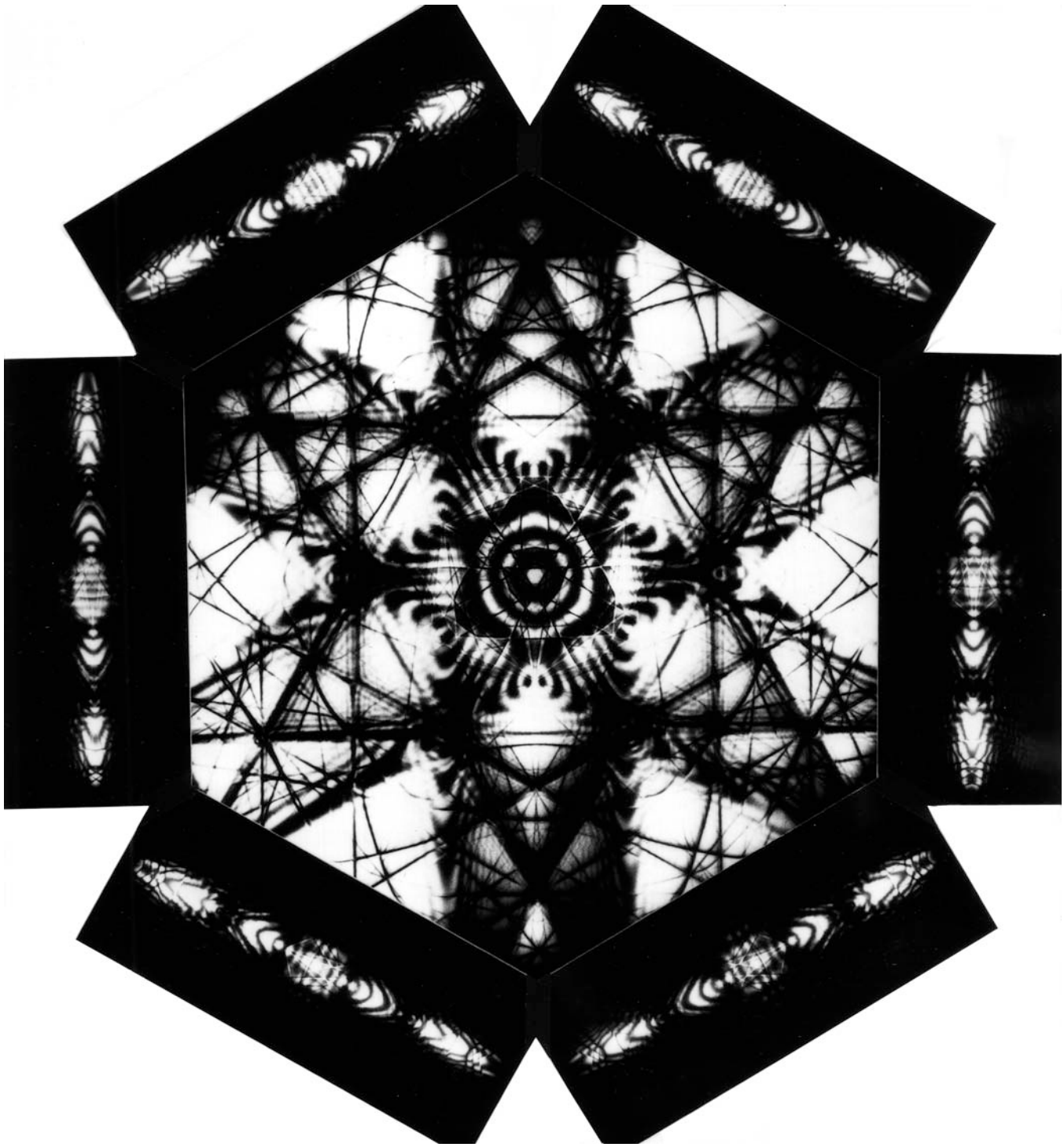


FIGURE 20.9. Montage of Si [111] BF LACBED pattern (center) surrounded by six $\{2\bar{2}0\}$ DF LACBED patterns.

20.6 ZERO-ORDER AND HIGH-ORDER LAUE-ZONE DIFFRACTION

20.6.A ZOLZ Patterns

If you increase L above ~800 mm you will magnify the pattern on the screen and only see the first few diffraction maxima, as shown in Figure 20.4A. Assuming the C2 aperture is small enough, the CBED pattern consists of disks, similar to the array of spots in an SADP, i.e., discrete diffraction maxima surrounding the central 000 disk. Remember that such a pattern is termed a ZOLZ pattern (see Section 18.4) since the permitted *hkl* diffraction maxima must all satisfy the Weiss zone law $hU + kV + lW = 0$, where UVW is the beam direction. Remember also that all the *hkl* maxima in the DP

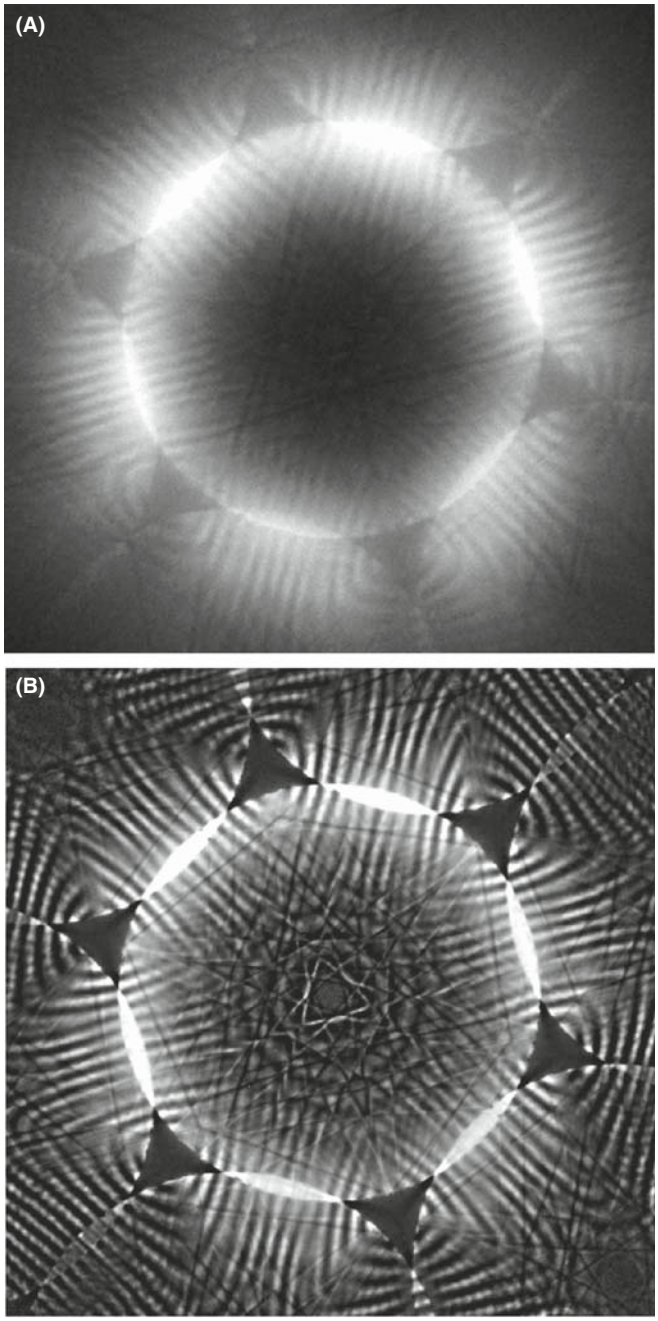


FIGURE 20.10. CBED pattern from a thick specimen of Si (A) without energy filtering and (B) with energy filtering. This figure is included only because it looks so striking, and similar patterns have been the basis of many TEM lab Christmas cards.

correspond to points in the reciprocal-lattice plane containing the origin, 000, of the reciprocal lattice, and this plane is also called the ZOLZ. So in fact SADPs are usually ZOLZ patterns, although we don't always describe them as such. From ZOLZ patterns, we can obtain the usual interplanar spacings and angles, the *hkl* maxima can be indexed and *UVW* identified, in exactly the same manner as we described in Section 18.4 for an SADP. The two options are the method of ratios or using a calibration standard to determine the camera length.

Because of the finite size of the diffraction disks, you must take care to select equivalent points in each disk when measuring the *hkl* spot spacings. If α is too large (Kossel conditions) you might not see individual maxima and you should then select a smaller C2 aperture (K-M conditions).

20.6.B HOLZ Patterns

The central portion of the CBED pattern is bright due to the relatively intense low-angle scattering (go back and check Figure 3.5). At higher angles, the ZOLZ intensity drops because the atomic-scattering amplitude, $f(\theta)$, decreases and the Ewald sphere no longer intercepts the relrods from the ZOLZ. However, the intensity increases again when the Ewald sphere intercepts the relrods from the HOLZ reciprocal-lattice points and a circle or ring of diffracted intensity is observed around the ZOLZ pattern as in Figures 20.3D–F and 20.4C. (You should recall that a circle can be defined geometrically as the interception of a sphere and a plane).

If you've chosen a small enough aperture for K-M conditions you'll see a ring of discrete HOLZ spots as in Figure 20.3D while a large C2 aperture gives a HOLZ ring of intersecting lines as in Figure 20.3F. The HOLZ intensity arises from relatively weak high-angle diffraction from crystal planes that are *not* parallel to the beam. Low temperatures increase the HOLZ scattering and also minimize the thermal-diffuse (phonon) scattering that, in some materials with a large Debye-Waller factor, masks the weak HOLZ intensity. So this is why you'll find a liquid-N_2 cooling holder very useful at times. You can't reduce the diffuse scattering via an energy filter because phonon scattering has a very small energy loss ($\ll$1 eV) and most of these electrons still contribute to the filtered pattern.

SMALL *L*, LARGE ANGLE

Remember that the radial distance from 000 in a DP is related to the angle of scattering; use a smaller *L* to see higher-angle scattering.

Consider the intersection of the Ewald sphere with the reciprocal lattice. The plane of HOLZ reciprocal-lattice points/relrods intercepts the sphere (unlike the zero layer which is tangential to the Ewald sphere). The first ring is called the FOLZ because the possible *hkl* reflections satisfy the relationship $hU + kV + lW = 1$, and so on. Where the Ewald sphere intersects the HOLZ relrods, diffracted intensity is expected, taking into account the usual structure-factor effects (see the next section).

Because the beam converges on the specimen over an angular range 2α, the Ewald sphere is effectively rotated 2α about the origin, and thus a range of angles along each HOLZ relrod is sampled, as shown in Figure 20.11A. This range of angles manifests itself as the CBED disk reflecting the effective angular broadening of the relrod, as shown in Figure 20.11B. Different interception points on the relrod correspond to different points in the disk as shown in Figure 20.11C. Figure 20.11D shows a K-M CBED pattern containing the expected distribution of diffraction maxima from the Ewald-sphere construction in Figure 20.11A.

3D

The most important point to remember is that there is 3D crystallographic information in the CBED pattern whenever significant HOLZ diffraction intensity is present.

We'll make use of this 3D information in the next chapter and in the companion text.

To observe HOLZ rings in addition to the ZOLZ pattern, choose a very small L (< 500 mm) so that you can see the full angular range of the BFP permitted by the imaging system ($\sim \pm 10^\circ$). As shown schematically in Figure 20.11A, the Ewald sphere only intercepts relrods in HOLZs many orders of diffraction maxima away from the direct beam. Because of the large scattering angle, the intensity in the HOLZ spots is relatively low and the exposure time to reveal HOLZ maxima is usually long enough to ensure that the ZOLZ is overexposed on the negative (go back and see Figure 20.3). You might be lucky with your thin area; sometimes you can produce reasonable ZOLZ and HOLZ intensity on the same exposure as in Figure 20.11D. But generally, if you're still stuck with recording your patterns on film you will probably have to record at least two DPs: a relatively short exposure for the ZOLZ pattern containing only 2D crystallographic information and a longer exposure for the weak HOLZ reflections containing the 3D information. As we've already said, a range of exposures is useful for *all* DPs.

There are some alternatives

- A CCD camera will give a greater dynamical range making it easier to record a good pattern showing detail at both low and high scattering angles.
- You can use image-processing techniques on your computer to combine differently exposed patterns (see Chapter 31).

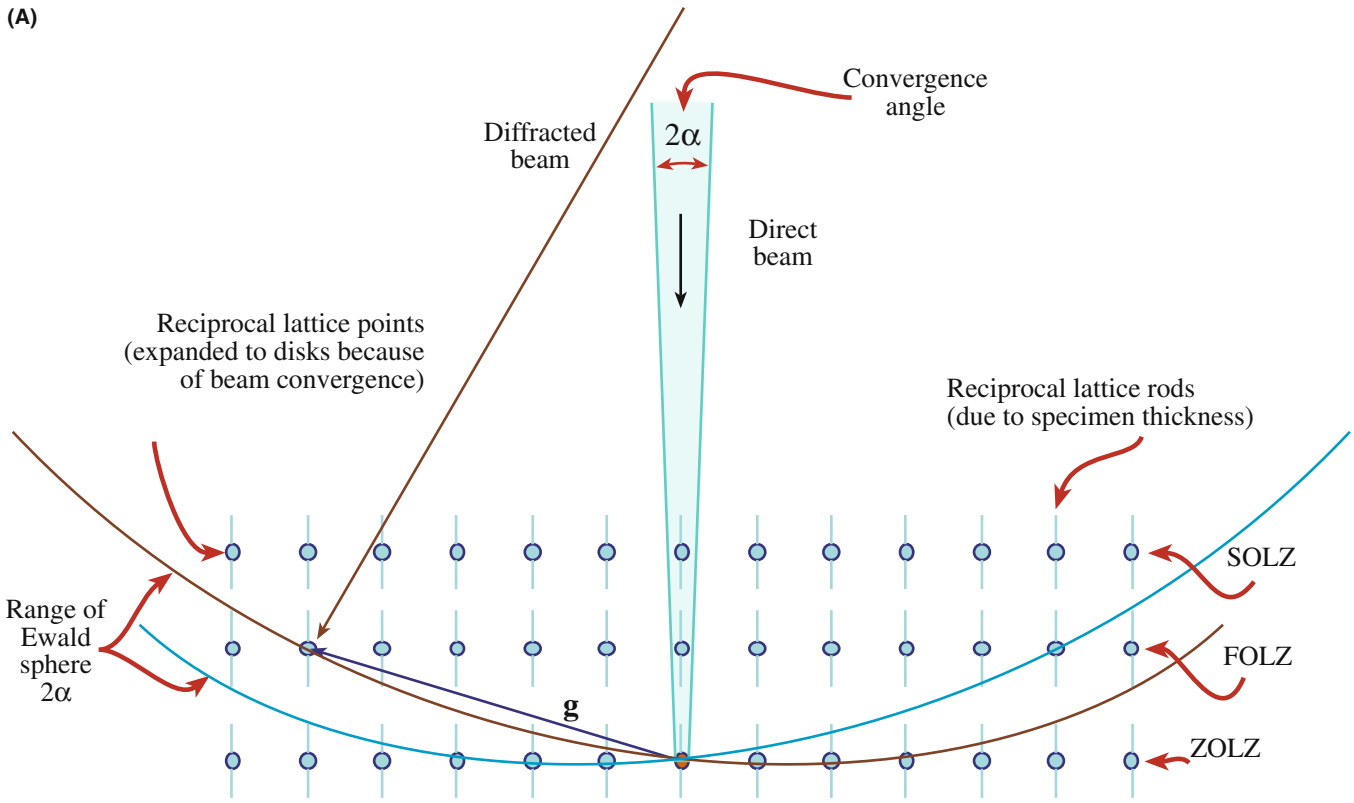
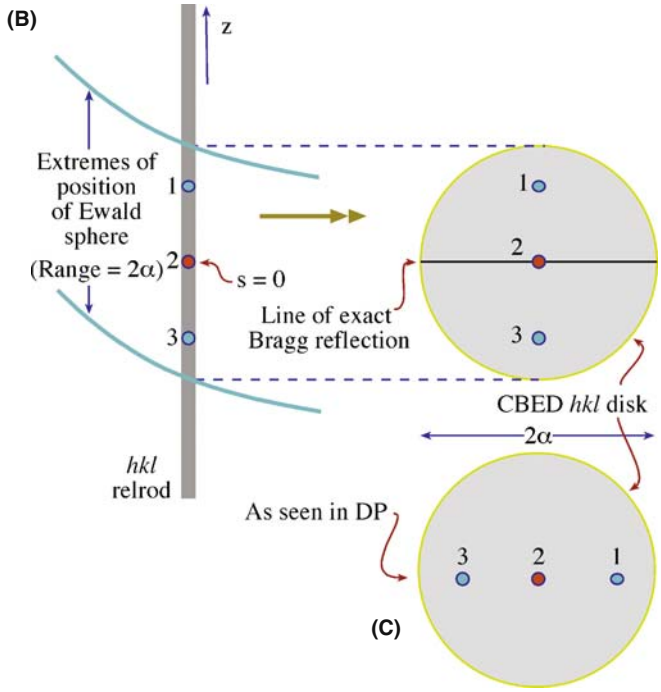


FIGURE 20.11 (A) The Ewald sphere can intercept reciprocal-lattice points from planes not parallel to the electron beam whose **g** vectors are not normal to the beam. The sphere has an effective thickness of 2α because of beam convergence and so intercepts a range of these HOLZ reciprocal-lattice points.



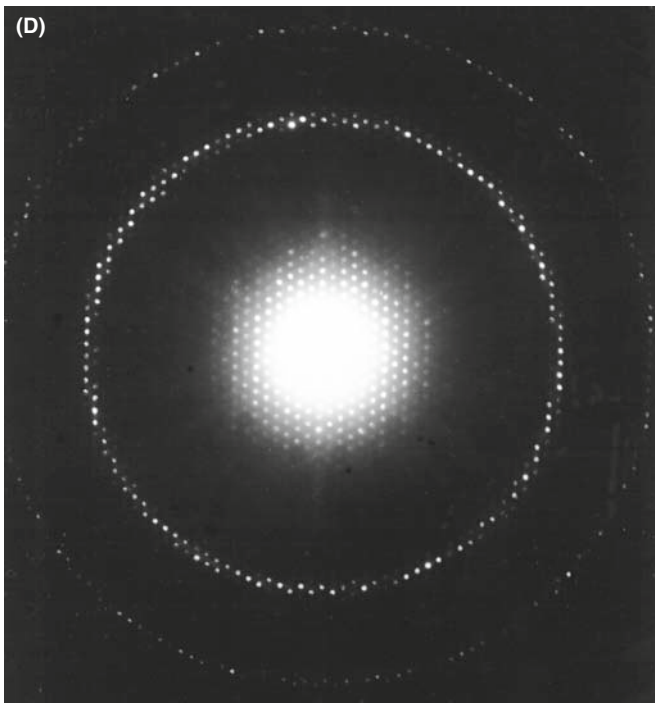


FIGURE 20.11. (Continued) The relrod has a shape shown in (B) and the intensity at specific points x_i in the relrod is directly related to equivalent points in the *hkl* disk (C). The interception of the Ewald sphere with the HOLZ layers gives rings: the first ring is called the FOLZ, the second the SOLZ and so on shown in the experimental pattern in (D).

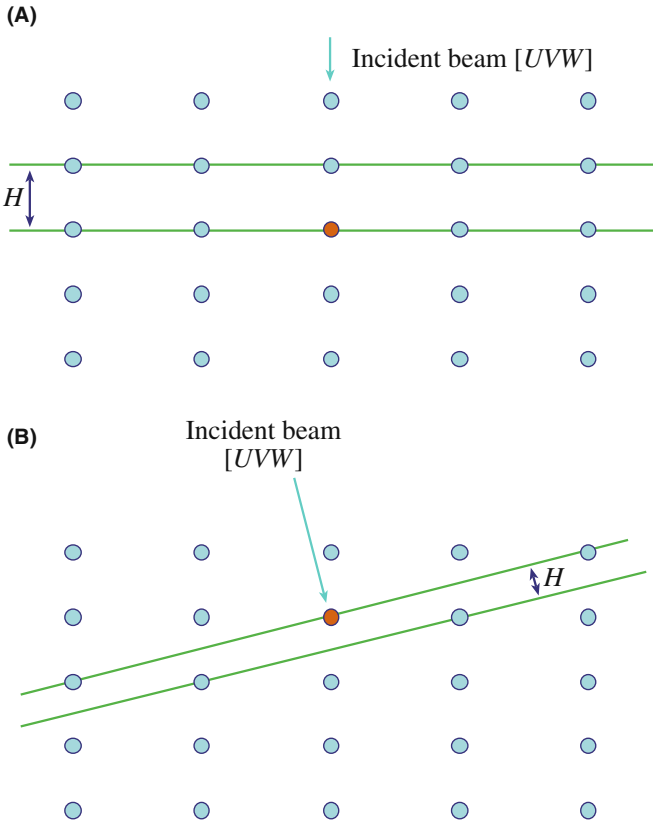


FIGURE 20.12. (A) The reciprocal-lattice spacing (**H**) is large if the beam is down a major zone axis in the crystal. (B) The spacing is small if the beam is down a low-symmetry direction.

The HOLZ ring radius is defined by the interception of the Ewald sphere with the allowed HOLZ relrods in the reciprocal lattice and so depends on the interplanar spacing in the crystal, the electron wavelength (i.e., the kV), L and any off-axis lens distortion. Depending on the crystallography of the specimen, the HOLZ rings may have very large diameters making them difficult to observe experimentally even at very small L. Under these circumstances, you should tilt to a *low*-symmetry zone axis (e.g., <114>) since this gives you a better chance of observing the FOLZ than a high-symmetry zone axis such as <001>. (If the reason for this is not clear, then look at Figure 20.12.) Cooling your specimen will reduce thermal-diffuse scattering that can mask HOLZ effects. If you still can't see a HOLZ ring, then the last thing you can try is increasing λ by lowering the kV. If all this fails, you need to pick another specimen!

In the next chapter, we will show you how to index HOLZ patterns and how HOLZ ring measurements can be used to deduce the lattice-repeat vector of the crystal parallel to the beam direction. You'll see that you can then determine the unit cell, the crystal system and also the type of lattice centering.

20.7 KIKUCHI AND BRAGG LINES IN CBED PATTERNS

In CBED patterns you almost invariably see sharp Kikuchi lines, while in SADPs Kikuchi lines are often rather diffuse or absent (see Chapter 19). This difference arises

in part because the convergent beam samples a much smaller region of the specimen than that selected by the SAD aperture. So, in the volume of specimen contributing to the CBED pattern, there is usually little or no strain, either elastic (due to specimen bending) or plastic (due to lattice defects). As a result, CBED Kikuchi lines will in general be sharper than SAD lines. This effect is shown in Figure 20.13A which is a conventional SADP containing very diffuse Kikuchi lines. This pattern was obtained from a large region of heavily deformed copper. By comparison, Figure 20.13B shows a CBED pattern from a much smaller region of the same specimen showing several pairs of well-defined Kikuchi lines. So you can use Kikuchi lines in CBED patterns to attack problems which are beyond the capability of SAD, for example, to determine accurate misorientation relationships between small grains in deformed materials (see Heilman et al.). However, if you put the probe right on a defect which has significant strain associated with it then even the CBED pattern may be blurred. On the positive side, this effect is used in LACBED patterns to characterize the crystallography of defects (e.g., the Burgers vectors) and this is explained in detail in the companion text. So, unless you want to study defects, always check the TEM image and tilt around to make sure that the area is clear of defects before getting your CBED patterns.

If the CBED pattern is not a ZAP, as in Figure 20.13B, the Kikuchi lines appear as pairs of excess (bright) and deficient (dark) lines, as in SADPs. But when you obtain a ZAP, the ZOLZ Kikuchi lines appear as bright bands. These bands increase in intensity and definition as you increase the convergence angle, as shown back in Figure 20.3D–F. A similar effect is seen in channeling patterns in the SEM that are generated by rocking a parallel beam around the optic axis. If you need to understand the difference between Kikuchi lines in SADPs and Kikuchi bands in CBED patterns, Reimer gives a clear discourse in his 1997 text.

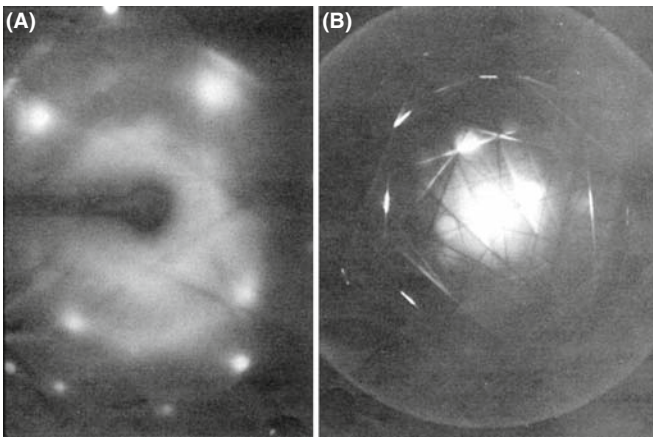


FIGURE 20.13. (A) Comparison of the poor quality of Kikuchi lines in an SADP and (B) the relatively clear distribution in a CBED pattern from deformed copper.

The generation of Kikuchi lines in a CBED pattern is marginally more complex than in an SADP. Remember how Kikuchi lines arise in a specimen illuminated by a parallel beam (see Chapter 19) and how a divergent beam of scattered electrons samples the various crystal planes (see Figure 20.14A). In Figure 20.14B you can see what happens when a convergent beam is used. In this case the incident rather than the scattered electrons span an angular range and therefore, just as in Kikuchi-line generation, some electrons in the beam will probably be at the exact Bragg angle to a ZOLZ plane (this is identical to the argument we used to explain the Kikuchi lines in LACBED patterns back in Figure 20.8A). Thus, there will be an elastic-scattering contribution to the Kikuchi lines where they cross the ZOLZ disks in CBED patterns. If you choose Kossel conditions (i.e., $2\alpha > 2\theta_B$) as in Figure 20.3F, there will *always* be electrons in the beam with the correct trajectory for exact Bragg diffraction from the planes in the *UVW* zone and so there will *invariably* be an elastic contribution to the Kikuchi lines.

Strictly speaking, we should only use the term 'Kikuchi lines' when inelastic scattering alone is responsible for their formation (i.e., the lines *between* any *hkl* disks). However, the term is used rather loosely in the literature to describe the ZOLZ intensity bands, despite the elastic contribution to the scattered intensity. Morniroli has proposed that this uncertainty be resolved by using the term 'Bragg lines' for the deficiency lines in CBED and LACBED patterns; this terminology is gaining wider acceptance.

20.8 HOLZ LINES

20.8.A The Relationship Between HOLZ Lines and Kikuchi Lines

As we just noted, Kikuchi lines can also arise from inelastic scattering by the HOLZ planes and HOLZ Kikuchi lines exist in many CBED patterns. You can see the array of deficient HOLZ Kikuchi lines between the ZOLZ maxima in Figure 20.2B. These HOLZ Kikuchi lines are, in principle, more useful than ZOLZ Kikuchi lines because they come from planes with much larger Bragg angles (and **g** vectors) so they are even more sensitive to changes in lattice parameter than the ZOLZ lines. Since

$$|\mathbf{g}| = \frac{1}{d}, |\Delta\mathbf{g}| = -\frac{\Delta d}{d^2} \tag{20.1}$$

then for smaller d values, the value of $|\Delta\mathbf{g}|$ is much larger at the same Δd. We take advantage of this fact, not by using HOLZ Kikuchi lines specifically, but by seeking out a closely related phenomenon called HOLZ lines. HOLZ lines are simply the elastic part of the HOLZ Kikuchi lines, that is, they are the segments of the lines

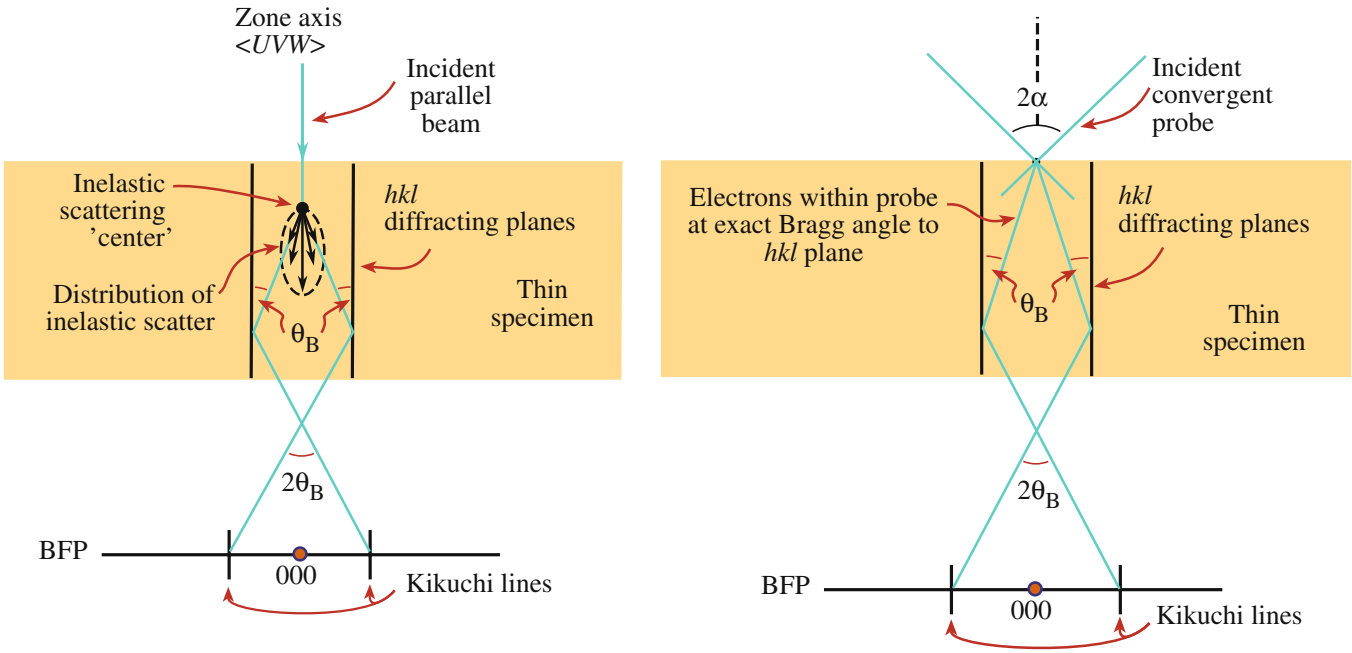


FIGURE 20.14. Comparison of the generation of Kikuchi lines (A) by inelastic scattering of electrons in a parallel beam and (B) by elastic scattering of electrons in a convergent beam when the convergence angle, α is greater than the Bragg angle, θ_B.

which appear *within* the diffraction maxima. By analogy with the production of Kikuchi lines and, as already shown in Figure 20.8A for LACBED patterns (although both these situations referred to ZOLZ diffraction), the HOLZ lines arise when electrons within the cone of the incident probe are at the correct Bragg angle for diffraction by a *HOLZ* plane. Therefore, these electrons are diffracted out to very high angles compared with ZOLZ diffraction. The result of this scattering is a bright line through the HOLZ disk and a dark line across the 000 disk. Not surprisingly, given what you've learned in the preceding nine chapters, the theory for the origin of HOLZ lines is much more complicated than this sparse summary. When you have time and the necessary inclination (tilt?) you should read the paper by Jones et al.

PAIRS OF HOLZ LINES

HOLZ lines come in pairs, like Kikuchi lines, with the bright (excess) lines within the HOLZ *hkl* maxima and the dark (deficient) lines within the 000 disk.

An example of HOLZ lines is shown schematically in Figure 20.15, which compares with the experimental pattern in Figure 20.32B. Because these HOLZ lines contain 3D information they show the true, fcc, 3D, threefold {111} symmetry, while the ZOLZ Kikuchi lines and spots show sixfold, 2D, {111} symmetry. We will make use of these differences when we discuss indexing such patterns and crystal-symmetry determination in Chapter 21 and in the companion text.

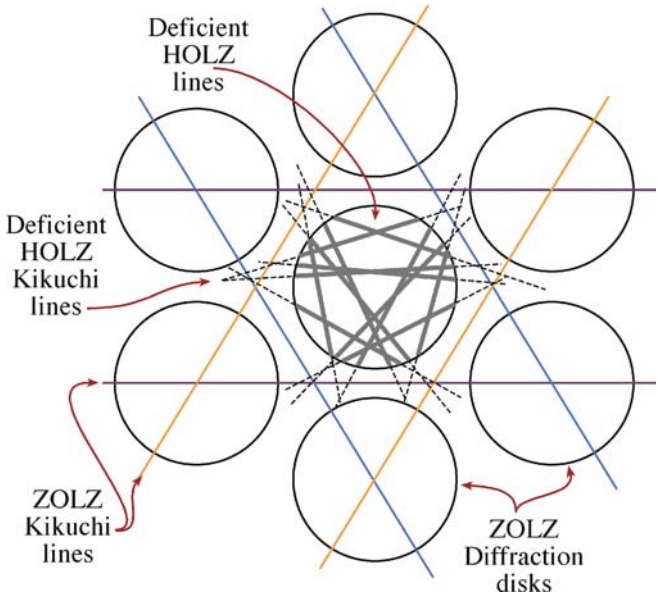


FIGURE 20.15. The relationship between Kikuchi lines and HOLZ lines is shown in this schematic of a <111> CBED pattern from a cubic crystal. The three principal pairs of $2\bar{2}0$ ZOLZ Kikuchi bands show sixfold symmetry (characteristic of the 2D 111 planes) and bisect the **g** vectors from 000 to the six $2\bar{2}0$ ZOLZ maxima. The *inelastic* deficient Kikuchi lines from the HOLZ planes are shown in the regions between the ZOLZ diffraction disks and the *elastic* deficient lines from the HOLZ planes are present within the 000 disk. In both cases, the HOLZ lines show threefold symmetry characteristic of looking down the <111> direction in a 3D crystal.

20.8.B Acquiring HOLZ Lines

Steeds (1981) described the practical problems of recording HOLZ lines in some detail. Nothing much has changed in the intervening decades and the main points you have to consider are

- The lines are often only visible 'on the negatives' and not on the screen, so you should record all the DPs, not just ones on which you can see the lines. This poor visibility is less of a problem with a CCD because you can easily enhance the gain.
- You may have to make small changes in the operating voltage or the orientation in order to view the HOLZ rings, especially if the crystal has a small lattice-repeat spacing parallel to the beam and the angular view of the BFP is poor.
- Strains in the specimen from bending and thermal stresses smear out the HOLZ-line intensity. Choosing the smallest region (i.e., the smallest beam) may help this problem and will also minimize local thickness variations.
- Planar or point disorder as well as thermal (Debye-Waller) effects can reduce high-angle scattering. In practice, this means that cooling the specimen and reducing the kV can help to increase HOLZ-line visibility by reducing contamination and beam heating.
- Minor adjustments in HOLZ-line positions can help to distinguish HOLZ lines that overlap. To do this you need to change the kV by a small amount.

CONTROLLING kV

Continuous kV control is an essential accessory to your TEM for serious CBED work.

The experimental procedure for observing HOLZ lines is quite straightforward but, since the lines themselves can be rather elusive, as we suggested at the start of the chapter, you should practice with a specimen such as Si or stainless steel in which the lines are almost always visible. The best way to search for the lines is

- Select the largest C2 aperture and go to the smallest L ($\sim$ 3–500 mm) at which you can see the full angular view of the BFP.
- Examine the Kossel/LACBED pattern (make 2α large) which should reveal Kikuchi bands intersecting at many poles, spanning a good fraction of the stereographic triangle as shown in Figure 20.16A.
- If you tilt to such a pole, as shown in Figure 20.16B, you should see the ring of HOLZ intensity.
- Tilt to a suitable zone axis for optimizing HOLZ effects. Remember, the best orientation for seeing the HOLZ lines in the 000 disk is *not* a low-index, high-symmetry pole such as <100> or <111> , but a higher-index, lower-symmetry one such as <114>.
- To see the deficient lines, increase L to look in detail at the 000 region of the pattern, and if necessary, put in a

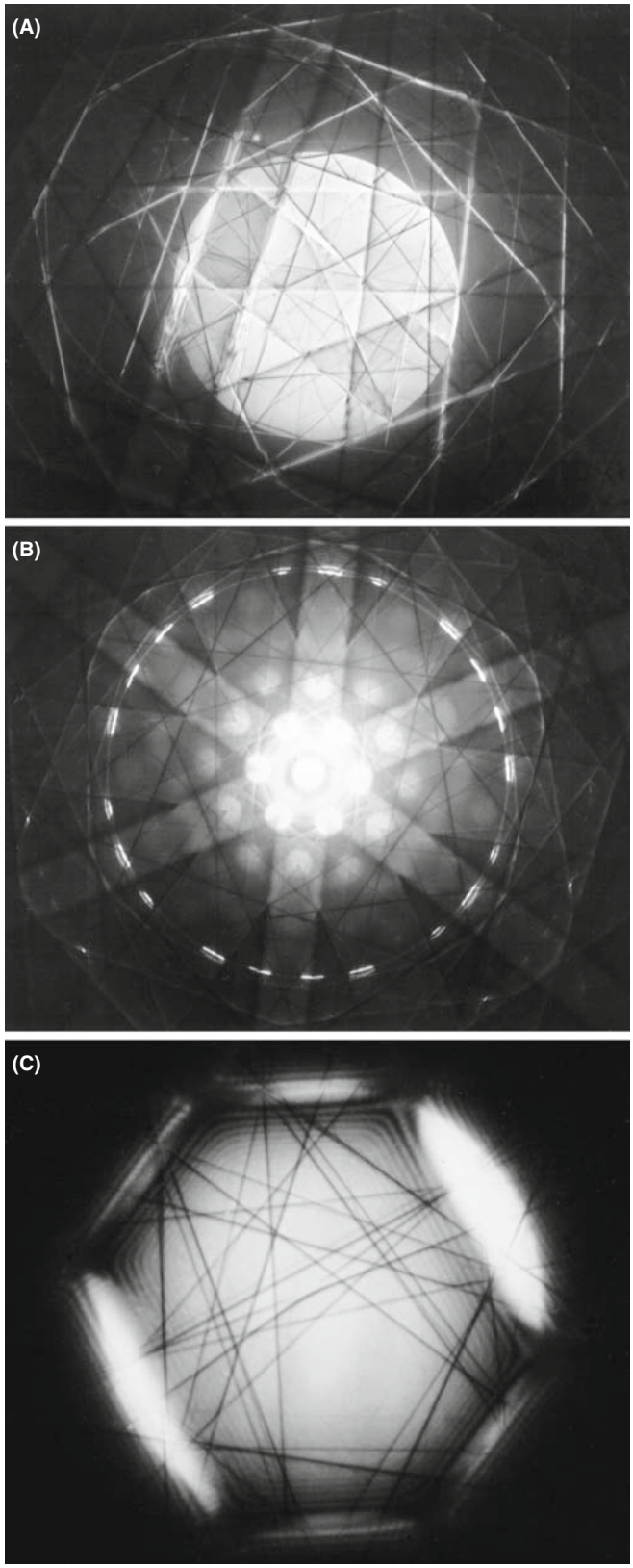


FIGURE 20.16. (A) Low-L, large-α CBED pattern showing a wide area of reciprocal space, away from ZAP conditions. (B) When the specimen is tilted to a high-symmetry <111> ZAP and a smaller C2 aperture is inserted, a ring of excess HOLZ lines appears. In (C) taken at high L, the deficient HOLZ lines are visible in the central 000 disk of the lower symmetry <114> pattern.

smaller C2 aperture, center it and look for the fine dark lines crossing the bright disk as in Figure 20.16C. Usually, you'll just need this deficient-line distribution.

You have to use a range of L (from ~300 to 1500 mm) to obtain all this information and this is easy in any post-1980 TEM. DSTEMs with sufficient post-specimen lenses are as versatile. Energy filtering improves everything that we've just described.

20.9 HOLLOW-CONE/PRECESSION CBED

As we described in Chapter 18, for SAD, we can acquire CBED patterns by rotating the incoming convergent beam about the optic axis in a hollow cone. Just as with SAD, hollow-cone CBED supplements the information we get in a stationary-beam pattern (i.e., all the patterns we've discussed so far). Remember, hollow-cone operation deflects the incident beam and rotates the pattern in the BFP while precession de-scans the beam so that the actual pattern is stationary. A good review of hollow-cone CBED and its relationship to other CBED methods is given in Tanaka's 1986 paper. Hollow-cone patterns do not show specific diffraction spots like CBED patterns since the rotating disk intensity is averaged out and the only stationary features are diffraction events at specific angles such as HOLZ lines. We can take advantage of this because if, for example, the cone angle is set to that of the FOLZ ring then the pattern consists predominantly of an array of excess HOLZ lines as shown in Figure 20.17. The ZOLZ disks, which would normally be so bright that they might mask the excess lines, are removed because the rotation of the incident beam averages out the intensity.

Precession of CBED patterns has all the advantages of precession of SADPs as discussed in Section 18.8. A comparison of a precession DP and a standard DP under K-M conditions is shown in Figure 20.18. Precession extends the FOLZ ring into an annulus of width of ~10 mrad. This dynamical background averaging can improve intensity measurements from DPs by considerably simplifying the problem of background subtraction.

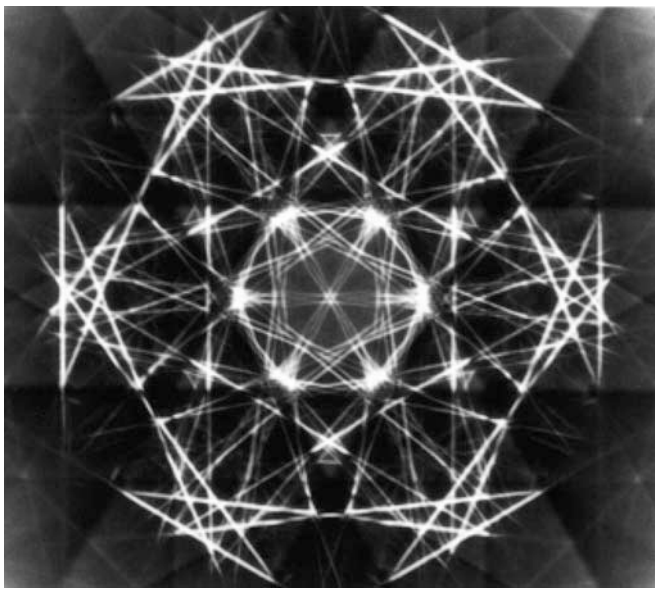


FIGURE 20.17. Hollow-cone CBED pattern showing bright, excess HOLZ lines only. All the usual diffraction-disk intensity is averaged out (Graphite 0001; 200 kV).

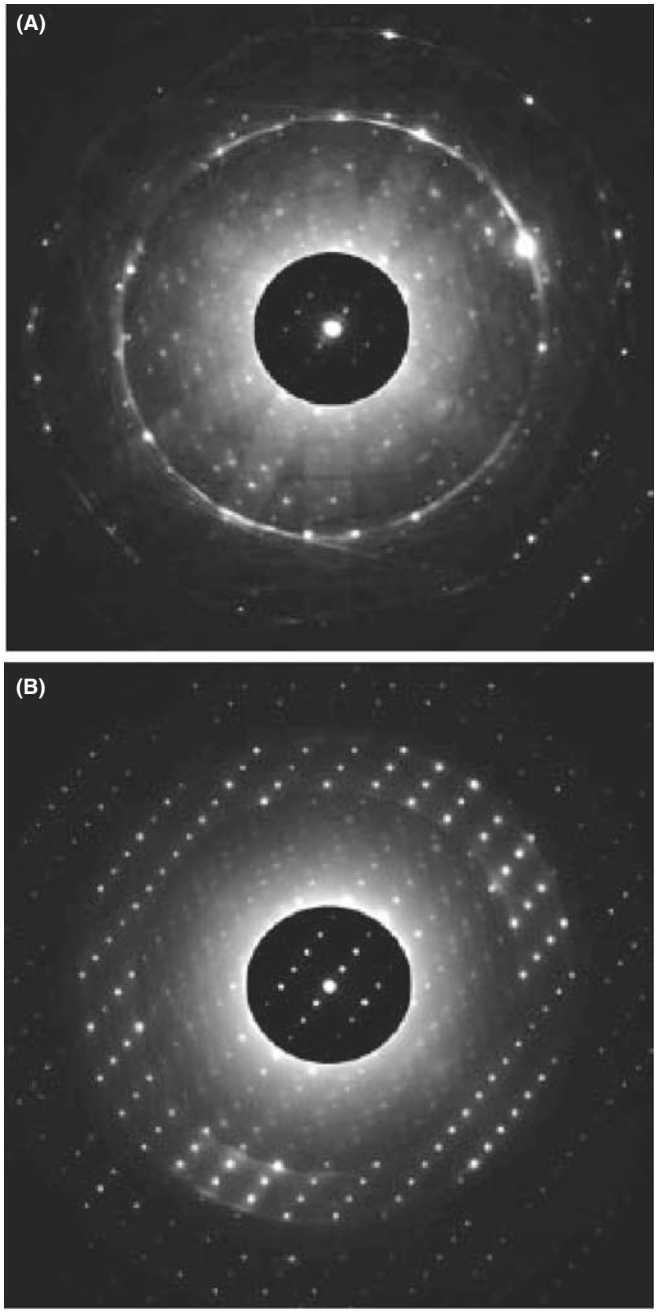


FIGURE 20.18. Comparison of (A) unprecessed and (B) precessed patterns from a very thick specimen of $Mg_3V_2O_8$. A small precession angle (~ 5 mrad) was used to form (B) and the reflections in the ZOLZ are more clearly defined. Dynamical scattering still contributes to the pattern so a thinner specimen would be better.

CHAPTER SUMMARY

In this chapter we've covered how to obtain different CBED patterns experimentally. Particular points and terms that you should know are

- If you vary the specimen thickness, α, L, and the focus you can obtain CBED and LACBED patterns showing many different features. As with SAD, two-beam and zone-axis conditions are the standard operating conditions to get the most useful information.
- You'll probably have to record patterns at different values of α and L, each with different exposure times.
- If you have an EELS then always filter your patterns.
- Even if you can filter the patterns, we strongly recommend using a double-tilt holder, cooled to LN_2 temperature.
- You must be able to change the kV by very small steps, if you are studying HOLZ lines.
- You can precess your CBED patterns or form them with hollow cones of incident electrons.
- Learn the meaning of such terms as ZAP, ZOLZ, FOLZ, HOLZ, K-M, and Kossel patterns.

In the next chapter, we'll show you how to use this contrast information to get the maximum amount of crystallographic information from your specimen.

TEXTS

Champness, PE 2001 *Electron Diffraction in the Transmission Electron Microscope* BIOS Oxford UK. A concise and essential summary textbook full of useful examples often from non-cubic systems, since the author is a mineralogist not a metallurgist.

Morniroli, J-P 2002 *Large-Angle Convergent-Beam Electron Diffraction* SFμ (Société Française des Microscopies) Paris France. Highly recommended reading for serious students for whom CBED will be an important tool.

Reimer. L (Ed.) 1995 *Energy-Filtering Transmission Electron Microscopy* Springer-Verlag New York. For some specifics.

Reimer, L 1997 *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis* (4th Ed.) Springer New York. For more specifics.

Spence, JCH and Zuo, JM (Eds.) 1992 *Electron Microdiffraction* Kluwer New York. *The* book for the dedicated CBED researcher.

Steeds, JW 1979 *Convergent Beam Electron Diffraction* in *Introduction to Analytical Electron Microscopy* 387–422 Eds. JJ Hren, JI Goldstein and DC Joy Plenum Press, New York. *The* first book-chapter dedicated to the technique and still a superb introduction.

THE JEOL CBED ATLAS

This set of four volumes is perhaps the definitive work on CBED and describes an alternative approach to much of what is discussed in Chapters 20 and 21 and also in the companion text. The texts contain hundreds of beautiful and useful patterns.

Tanaka, M and Terauchi, M 1985 *Convergent Beam Electron Diffraction* JEOL Tokyo.
Tanaka, M, Terauchi, M and Kaneyama, T 1988 *Convergent Beam Electron Diffraction* **II** JEOL Tokyo.
Tanaka, M, Terauchi, M and Tsuda, T 1994 *Convergent Beam Electron Diffraction* **III** JEOL Tokyo.
Tanaka, M, Terauchi, M and Tsuda, T 2002 *Convergent Beam Electron Diffraction* **IV** JEOL Tokyo.

HISTORY

Also see the references in Chapter 9.

Kossel, W and Möllenstedt, G 1938 *Elektroneninterferenzen im konvergenten Bündel* Naturwissenschaften **26** 660–661.

APPLICATIONS AND TECHNIQUES

Heilmann, P, Clark, WAT and Rigney, DA 1983 *Orientation Determination of Subsurface Cells Generated by Sliding* Acta Met. **31**(8) 1293–1305.

Humphreys, CJ, Maher, DM, Fraser, HL and Eaglesham, DJ 1988 *Convergent Beam Imaging – A Transmission Electron Microscopy Technique for Investigating Small Localized Distortions in Crystals* Phil. Mag. **58A** 787–798. CBIM.

Jones, PM, Rackham, GM and Steeds, JW 1977 *Higher Order Laue Zone Effects in Electron Diffraction and Their Use in Lattice Parameter Determination* Proc. Roy. Soc. **A354** 197–222. When you have the time to really study HOLZ lines.

Steeds, JW 1981 *Microanalysis by Convergent Beam Electron Diffraction in Quantitative Microanalysis with High Spatial Resolution* 210–216 Eds. GW Lorimer, MH Jacobs and P Doig The Metals Society London. For the terminology in Section 20.3.A.

Tanaka, M, Saito, R, Ueno, K and Harada, Y 1980 *Large Angle CBED* J. Electr. Microsc. **29** 408–412.

Tanaka, M 1986 *Conventional Transmission-Electron-Microscopy Techniques in Convergent-Beam Electron Diffraction* J. Electr. Microsc. **35** 314–323. Includes a good review of hollow-cone CBED.

Wu, L, Zhu, Y and Taftø, J 1999 *Towards Quantitative Measurements of Charge Transfer in Complex Crystals Using Imaging and Diffraction of Fast Electrons* Micron **30** 357–369.

APPLICATIONS AND TECHNIQUE

The companion text has a complete chapter devoted to CBED. Some of that material was in the original edition; it includes techniques that are still very powerful but that are not so widely used.

SELF-ASSESSMENT QUESTIONS

Q20.1 If CBED was invented before SAD, why did it only come into widespread use many years later?

Q20.2 What advantages accrue when using a convergent rather than a parallel beam of electrons to create a diffraction pattern?

Q20.3 What disadvantages are there to using CBED rather than SAD?

Q20.4 How do you control the beam size in CBED?

Q20.5 What are the advantages and disadvantages of using smaller beams to form CBED patterns?

Q20.6 How do you control the size of the disks in the CBED pattern?

Q20.7 Why would you want to change the camera length when viewing a CBED pattern?

Q20.8 Define HOLZ, ZOLZ, FOLZ, K-M, ZAP, LACBED. Find any other acronyms from this chapter that are not in this list and define them.

Q20.9 Explain why there is image information inside a CBED pattern disk when the pattern is not focused.

Q20.10 Explain why this image effectively expands to infinite magnification when the pattern is focused.

Q20.11 Why is it so important to align the C2 aperture when performing CBED?

Q20.12 What is the role of thickness in CBED-pattern contrast?

Q20.13 How can 3D information find its way into a 2D CBED pattern?

Q20.14 Why is a higher-order ZAP preferable to a lower-order one for viewing a CBED pattern?

Q20.15 Why do hollow-cone CBED patterns not show any diffraction disks, just HOLZ lines?

Q20.16 What advantages accrue from precessing a CBED pattern?

Q20.17 HOLZ lines within the 000 disk are continuous with lines outside this disk. What is the difference between the electrons contributing to the line within the disk and those in the line outside the disk?

Q20.18 Why can we see Kikuchi lines in CBED patterns from specimens that are too thin to show such lines in SADPs or in specimens which are too deformed to show such lines in SADPs?

Q20.19 Why do HOLZ lines appear more clearly at low temperatures?

TEXT-SPECIFIC QUESTIONS

T20.1 Look at Figure 20.1 and work out what would happen to the probe convergence angle if you switch on the C2 lens (a) weakly and (b) strongly enough to introduce a crossover before the upper objective lens. Sketch the ray diagrams for each case. Do the same for changing the size of the C2 aperture.

T20.2 Look at the symmetry of the spots in Figure 20.2A and B and compare with the symmetry of the lines within the 000 disk in B. They are not the same. From what you know of the symmetry down the [111] direction in an fcc crystal, which pattern symmetry is correct? Why is the other symmetry observed?

T20.3 Why does the intensity of the overall pattern increase from Figure 20.3D—F? Do you think the probe size has stayed constant as the C2 aperture has increased in size, and if not, why not? From what you know of lens limitations, what prevents us from keeping a very small probe size to localize the CBED pattern while increasing the current in the probe to generate a high-quality pattern? Are there experimental methods to overcome this problem?

T20.4 Look at Figure 20.4 and explain why most of the intensity is in the 000 disk while it is barely possible to see the intensity scattered out to high angles in C. Is there anything you can do to increase the intensity scattered out to higher angles?

T20.5 Draw schematic ray diagrams to explain the reversal in the image direction contained in the 000 disk in Figure 20.6 as the objective lens goes through focus.
T20.6 List the pros and cons of forming CBED patterns in TEM versus STEM mode.
T20.7 Why do you think that dynamical scattering causes regions of low (or zero) intensity to occur in CBED disks as in Figure 20.5B?
T20.8 Explain what will happen to your CBED HOLZ intensity as you go to higher kVs. (Hint: look at Figure 20.11 and remember the effect of kV on λ.)
T20.9 Compare and contrast HOLZ lines and Kikuchi lines in terms of their generation, intensity and distribution in low camera-length CBED patterns.
T20.10 Since we know that the intensity of elastic scatter decreases rapidly with increasing angle away from the direct beam, where does the high intensity come from in the outer ring in Figure 20.16B?
T20.11 Look at Figure 20.15. Explain what happens as the deficient HOLZ Kikuchi lines are extended such that they intercept other *hkl* disks in the ZOLZ.
T20.12 Explain, with reference to Figure 20.8A, why it doesn't matter if you raise or lower the specimen out of the eucentric plane when forming a LACBED pattern. What experimental operation becomes increasingly difficult when you remove the specimen from the eucentric plane?
T20.13 Why is the LACBED BF pattern so extensive while a LACBED DF pattern covers only a very small region of reciprocal space (e.g., see Figure 20.9)?

21

Using Convergent-Beam Techniques

CHAPTER PREVIEW

In the preceding chapter, we described how to obtain a variety of CBED patterns under various experimental conditions. In this chapter you will find out why these patterns are so useful: they contain a wealth of quantitative data, much of which you can't obtain by any other technique and many of which augment standard X-ray crystallographic methods (but always at higher spatial resolution). The established techniques largely depend on simple observation of the patterns whereas newer techniques involve quantitative simulations of the patterns.

CBED gives quantitative data on

- Specimen thickness
- Lots of crystallographic data such as the unit cell and associated lattice parameters, the Bravais lattice, crystal system, and full 3D crystal symmetry (point group and space group)
- Precise lattice-strain measurement (invaluable for semiconductors and other multilayer nanostructures)
- Enantiomorphism and polarity
- Valence-electron distribution, structure factors, and chemical bonding
- Characterization of line and planar defects

Between this chapter and in the companion text, we'll cover most of these applications of CBED. First, we'll show how to index the patterns and the HOLZ lines when they occur within those patterns. We'll also show you how to measure the specimen thickness, which is very useful in many aspects of TEM, particularly the spectrometry that we describe in Part 4. Next, we'll introduce the steps for a complete symmetry analysis of your specimen, which is fully described in the companion text. Then, we'll introduce you to methods of determining extremely small changes in lattice parameter which can be used to measure lattice strain and, indirectly, composition. Like any other TEM technique, we can simulate the contrast in CBED patterns in extraordinary detail and we'll tell you a little about how this is done. Other CBED methods are also available as well as different micro- and nanodiffraction methods, all of which offer better spatial resolution than SAD and we'll briefly summarize these at the end of the chapter.

We should warn you at this stage that some of the analyses we introduce here require a very good understanding of crystallography. Both the learning and the doing can be time-consuming processes, depending on your background. If you're a metallurgist, in a predominantly cubic world, then life is relatively easy, but in the rest of materials science, in functional materials, and in much of the nano-world, more complex crystal structures are common. In studying these materials, CBED is absolutely invaluable, but also really challenging. We suggest that you skim the chapter and see how many of the topics you can already relate to, given your current knowledge of crystallography and crystal defects. If you draw a blank, then it's time to read some of the basic texts listed in Chapter 18 before moving ahead.

21.1 INDEXING CBED PATTERNS

You've seen in the preceding chapter that, in addition to improved spatial resolution, CBED patterns often contain a wealth of new and sometimes spectacular contrast information not available in SADPs. However, before you do anything quantitative, you have to index the patterns. Indexing CBED ZOLZ patterns is no different to indexing SADPs, which we've already described back in Chapter 18. With indexed CBED patterns you can do exactly the same kind of orientation determination as with SADPs, but on a much smaller dimensional scale, suitable for crystallographic analysis of nanostructures, as shown in Figure 21.1, which also shows the power of CBED-pattern simulation.

What you need to learn now is how to index HOLZ spots, disks, and lines and understanding these requires a little extra thought since such diffraction effects come from lattice planes that, unlike those contributing to SADPs, are *not* parallel to the electron beam and, thus, are the key to obtaining 3D information about your specimen.

21.1.A Indexing ZOLZ and HOLZ Patterns

Remember that in a ZAP the HOLZ reflections appear as a ring of spots (or a ring of intensity) (see Figures 20.3, 20.4, 20.11, and many others). If you're not at a zone-axis orientation, the HOLZ spots might appear as an arc or somewhat randomly distributed, *but always farther away from the direct beam than the ZOLZ reflections*. As with SAD, it's easiest to index ZAPs, so tilt around under Kossel conditions until you find a ZAP where Kikuchi bands intersect, then select a C2 aperture size to form diffraction disks (K-M conditions) in the HOLZ ring (like Figures 20.11D and 20.16B), which you can index as follows

- Index the ZOLZ for which $hU + kV + lW = 0$ (see Section 18.4).
- Consult a stereographic projection to identify the poles of the principal planes constituting the FOLZ ($hU + kV + lW = 1$) and SOLZ ($hU + kV + lW = 2$), etc.
- Alternatively, you can just solve the Weiss zone law (see Section 18.4) for the appropriate UVW.
- Check to see if the poles on the stereographic projection constitute allowed reflections.
- Index the HOLZ maxima.

If you want to make use of a stereographic projection, see Chapter 18. Remember that the stereographic projection just gives you the major, low-index, *hkl* planes and ignores any systematic absences.

Examples of indexed ZOLZ, FOLZ, and SOLZ patterns are shown in Figure 21.2A–C for the fcc lattice under (a) [001], (b) [110], and (c) [111] beam directions; Figure 21.3A–C show similar patterns for the bcc lattice. Only the first few orders of diffraction maxima are shown in these figures but, paradoxically, you'll rarely find these reflections, and you'll never see them if you're under zone-axis conditions. This is because the Ewald sphere intercepts the HOLZ at large scattering angles (go back and check Figure 20.11) so relatively high orders of diffraction maxima are present in the HOLZs. Therefore, the schematic patterns in Figures 21.2 and 21.3 should be extended accordingly to match up with the experimental patterns. From the schematic HOLZ patterns in Figures 21.2 and 21.3, you can see that the symmetry of each $<UVW>$ zone is retained in each HOLZ pattern, but they are often shifted by a displacement vector relative to the ZOLZ because there is no allowed HOLZ reflection on the zone axis (i.e., directly above 000 in the ZOLZ). Note that in the HOLZ patterns we do not index the central spot. This displacement can be calculated for any zone axis using equations which we'll discuss in a while. So you should generate similar diagrams for the major zone axes of any specimen that you are going to study by CBED. You may have to generate less symmetrical patterns than those in Figures 21.2 and 21.3 (e.g., [233] or [114]), since these give rise to FOLZ rings with smaller radii (Figure 20.12), in which HOLZ phenomena are easier to see.

In some circumstances, the first ring of spots observed is from the second layer of the reciprocal lattice *but it is still called the FOLZ*. Well-known examples, as you can see in Figures 21.2 and 21.3, are the fcc (110) and bcc (111) patterns. You can't predict the total

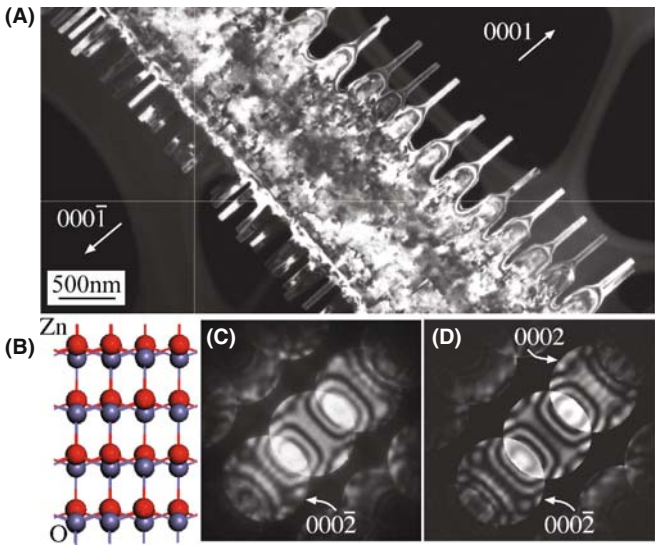


FIGURE 21.1 (A) Dark-field TEM image recorded from a comb-like ZnO nanostructure. (B) A structural model of ZnO projected along 0110, showing polarity on the (0001) surfaces. (C) Experimental CBED pattern recorded from the ribbon. (D) Simulated CBED pattern using dynamical electron-diffraction theory. The simulation includes 127 beams in both the ZOLZ and HOLZ and the Debye-Waller factor at room temperature. The diffraction intensities are very sensitive to thickness. The best thickness match at 200 kV and beam direction <0110> was 165 nm.

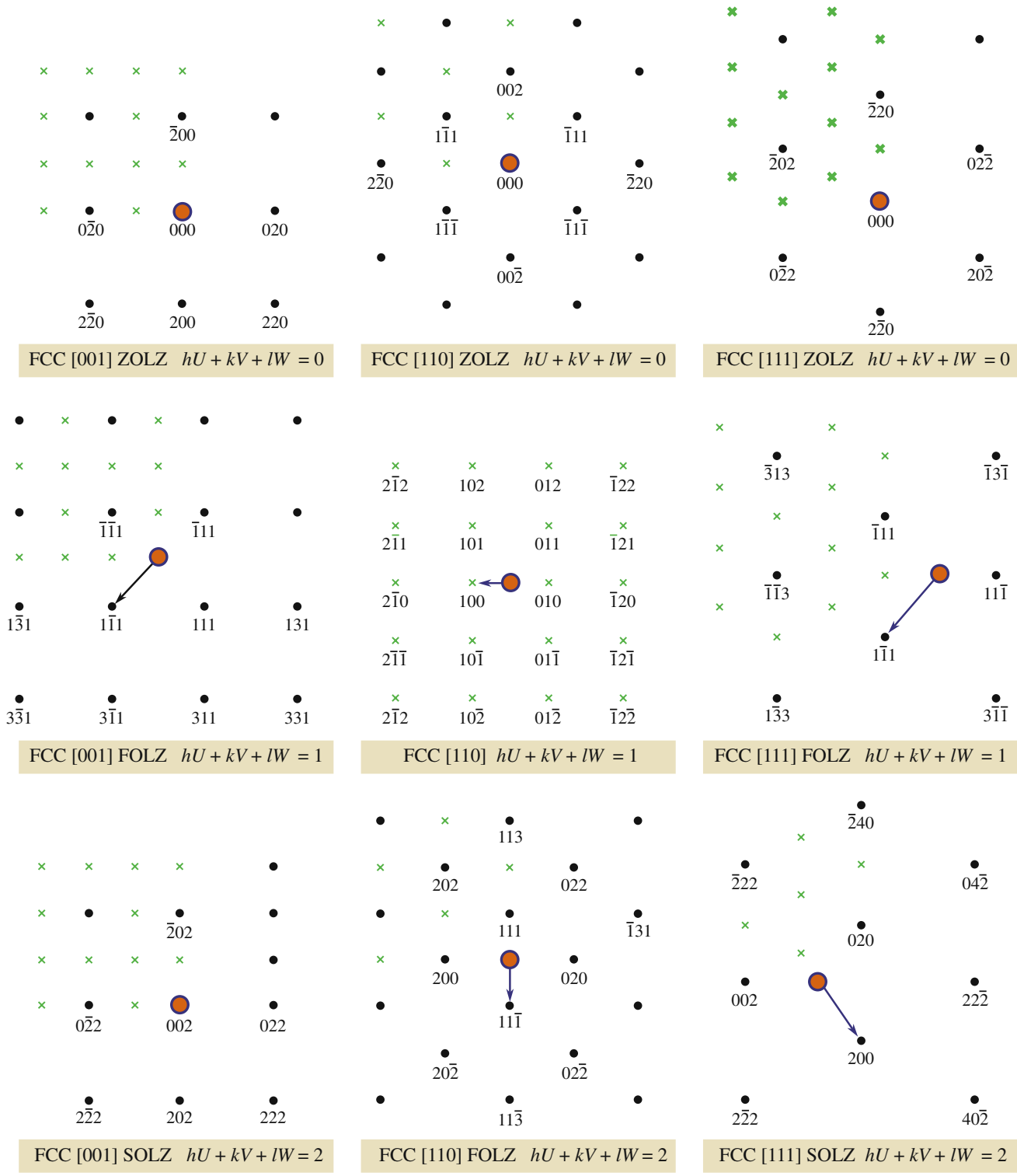


FIGURE 21.2 The possible ZOLZ, FOLZ, and SOLZ reflections for the three principal zone axes of an fcc specimen. Allowed reflections are shown as black dots, forbidden reflections as crosses, the direct beam direction is a red spot, and the arrow indicates the displacement vector between the ZOLZ and FOLZ.

absence of a ring of HOLZ reflections just from the crystal symmetry, but it does vary with orientation. For example, in rhombohedral α-Al_2O_3, which has trigonal symmetry, all HOLZ layers are present for the [001] beam direction but, in other directions in this system, e.g., [121], [141], and [542] only every third Laue zone is present. (Ceramists would use 4-index notation for α-Al_2O_3 but remember the symmetry is

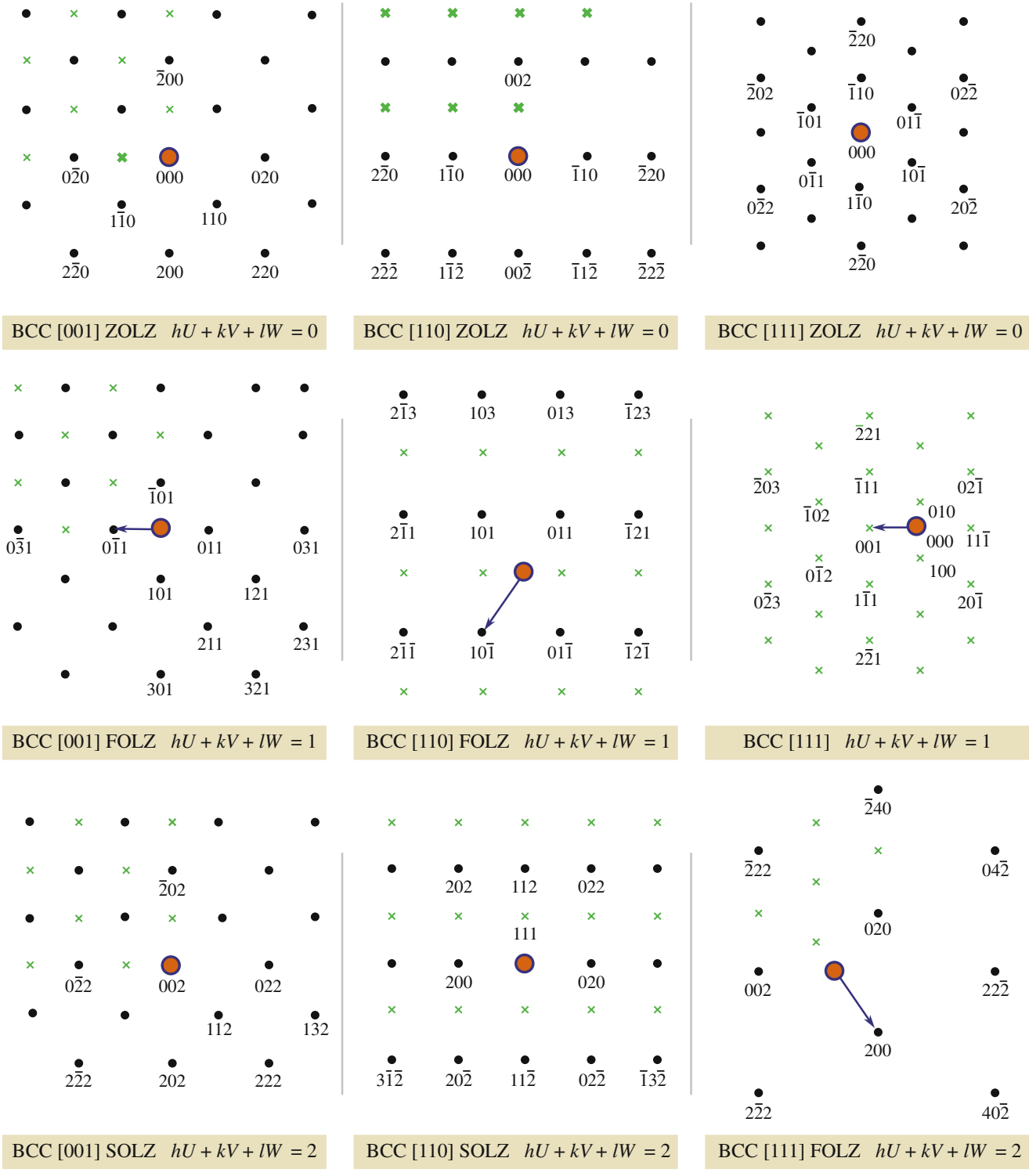


FIGURE 21.3 The possible ZOLZ, FOLZ, and SOLZ reflections for the three principal zone axes of a bcc specimen. Allowed reflections are shown as black dots, forbidden reflections as green crosses, the direct beam direction is a red spot, and the arrow indicates the displacement vector between the ZOLZ and FOLZ.

only threefold.) A detailed explanation of this behavior is given by Raghavan et al. (1984). As with SADPs, computer indexing of HOLZ diffraction maxima is available on the Web. Once you are sure you can extend the principles you learned in Chapter 18 from the ZOLZ to the FOLZ, then it's easy to use the commercial software available at any of the URLs #1–3 listed at the end of the chapter. There is also freeware available (URLs #4–7). One particularly good site is *WebEMAPS* (URL #4); Figure 21.4 illustrates computerized indexing of

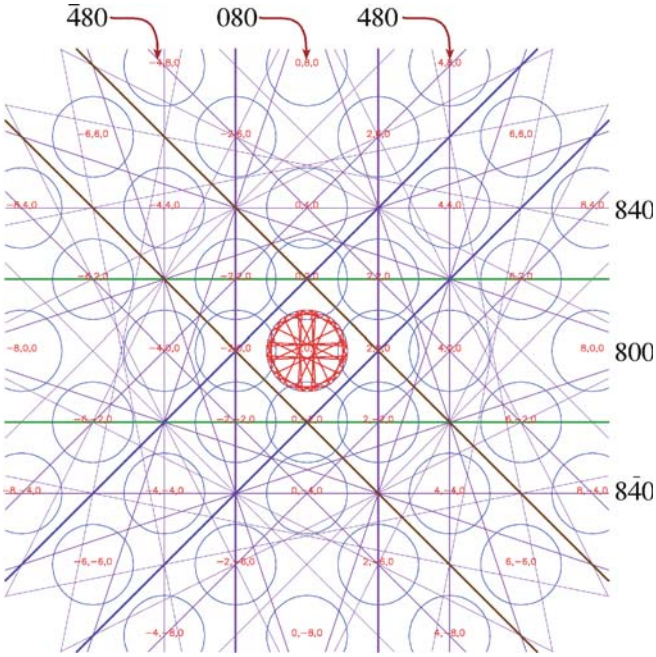


FIGURE 21.4 Computerized indexing of ZOLZ disks and simulation of deficient HOLZ lines in the 000 disk (red) and Kikuchi lines in a 001 ZAP of GaAs (at 200 kV) using WebEMAPS. The indices for the ZOLZ disks are difficult to read because of the computer-output font size so several of the outer disks are indexed more clearly.

HOLZ and Kikuchi-line simulations of a GaAs ZOLZ CBED pattern.

FOLZ WARNING

Be aware when you are indexing HOLZ rings that structure factors can cause *every* reflection in a FOLZ to be forbidden. So then we call the SOLZ the FOLZ!

21.1.B Indexing HOLZ Lines

When you're recording the pattern containing the deficient HOLZ lines, you must also record another pattern with a small L and a small α to show the FOLZ disks. The first thing you have to do is index the FOLZ *hkl* maxima in the manner we just described. Then observe which maxima show the clearest HOLZ excess lines. Each HOLZ-line pair will be perpendicular to the **g** vector from 000 to the FOLZ disk. There should be a parallel deficient HOLZ line in the 000 disk, and this line must be assigned the same indices as the FOLZ disk. If you repeat this exercise around the FOLZ ring, most of the HOLZ lines should be indexed, as shown in Figure 21.5. Below the schematic pattern is a magnified drawing of the 000 disk and you should be able to see the association of each HOLZ line with a corresponding FOLZ maximum (Ecob et al.).

In some cases you may find it difficult to associate a specific HOLZ deficient line in 000 with a specific FOLZ reflection due to strong excitation of two diffraction

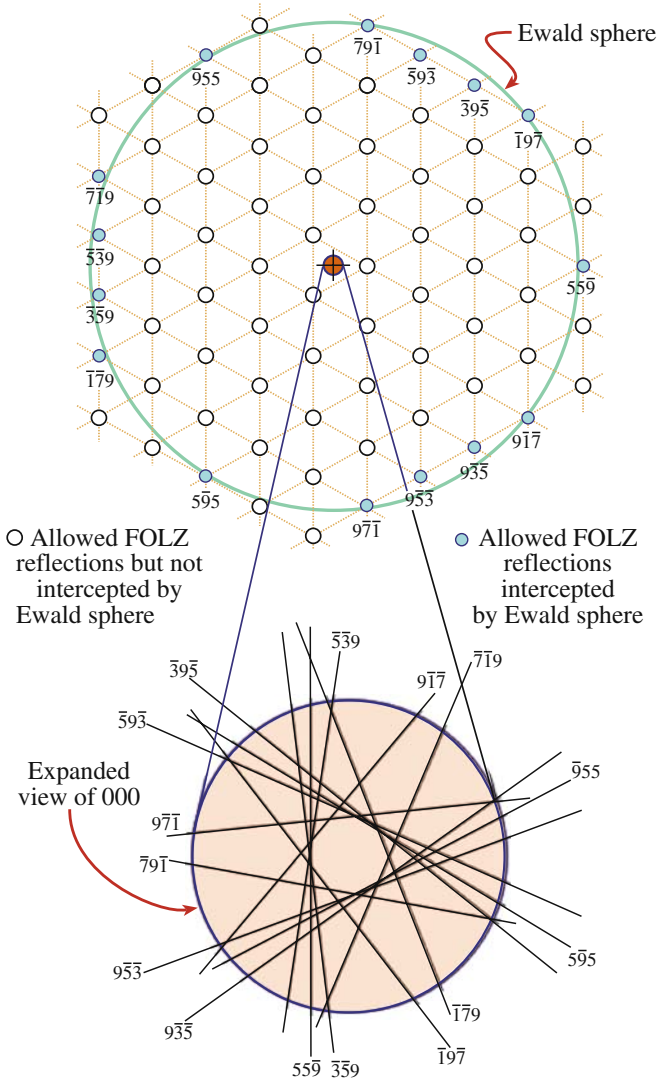


FIGURE 21.5 How to relate deficient HOLZ lines to the HOLZ maxima; the indexed FOLZ reflections in this [111] pattern are shown as full circles and the open circles are the rest of the FOLZ reciprocal lattice points that don't intercept the Ewald sphere. The **g** vector from 000 to each *hkl* FOLZ disk is normal to the *hkl* HOLZ line and the lines are shown in the expanded 000 disk below.

maxima. Under these circumstances, the two deficient lines may merge and appear to form a hyperbola. If you make small changes in the kV, the overlaps may resolve into two discrete lines. You should also be aware that faint HOLZ lines in the 000 disk may sometimes arise from second-order or even third-order Laue zones; these very high order lines are even more sensitive than first-order HOLZ lines to changes in kV and lattice parameter. As with the HOLZ patterns themselves, the indexing of HOLZ lines also lends itself to computer assistance. As we just showed in Figure 21.4, such programs can generate simulated HOLZ line patterns for a given orientation, lattice parameter, and kV. Matching of the computer simulation with the experimental pattern then allows direct indexing. This procedure is also

the first step in measuring the lattice parameter of the specimen and we will discuss this, along with other applications such as composition and strain measurements later in this chapter. Fournier et al. give a clear introduction to the step-by-step indexing and you should read this and then practice indexing some patterns by hand once or twice, just to get the hang of it. New software continues to become available to refine the various methods of HOLZ-line indexing (e.g., Morawiec).

21.2 THICKNESS DETERMINATION

As you read through this book, you will become aware that a direct and accurate measure of the specimen thickness is essential for many aspects of TEM and AEM, such as the correction of X-ray intensities for absorption within the specimen (Chapter 35), determining the attainable X-ray spatial resolution (Chapter 36), and obtaining EELS data with reasonable P/B (Chapters 37 and 38). A most useful application of CBED patterns is that you can use them to measure the thickness of a crystal.

When you record a ZAP under conditions where $2\alpha_S < 2\theta_B$ such as shown back in Figure 20.2B, the 000 disk usually contains concentric diffuse fringes known as Kossel-Möllenstedt (K-M) fringes. If you move the specimen under the beam and it is not too bent then you will see that the number of these fringes changes. In fact the number of fringes increases by one every time the thickness increases by one extinction distance, ξ_g; if the specimen is $< \xi_g$ in thickness, then you'll see no fringes and the 000 and *hkl* disks will be kinematical and uniformly bright and boring, as shown back in Figure 20.5A. So these fringes clearly contain thickness information. In fact, because the foil thickness can be measured at precisely the point you are doing diffraction and analysis, and because the method is very amenable to computerization, it has become a most popular use for CBED patterns. The region of the foil you select should be flat and undistorted (this is relatively easy when you are using a small focused probe, compared to a broad beam for SAD) and the beam must be focused at the plane of the specimen. The method is, of course, limited to crystalline specimens and it can be a bit tedious, but it is one of the best and, certainly for fully crystalline materials, the most accurate method of thickness determination.

In practice, to simplify the interpretation, we don't make thickness measurements under zone-axis conditions. You need to tilt to two-beam conditions with only one strongly excited *hkl* reflection. If you do this you will see that the CBED disks contain parallel rather than concentric intensity oscillations, as shown in Figure 21.6.

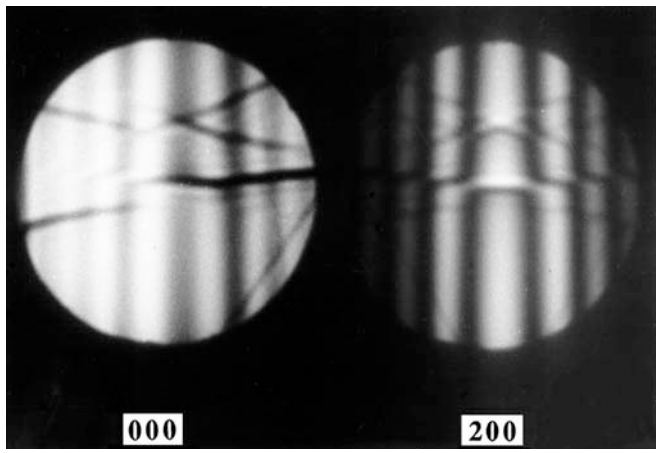


FIGURE 21.6 Parallel Kossel-Möllenstedt fringes in a ZOLZ CBED pattern from pure Al taken under two-beam conditions with (200) strongly excited.

FRINGES IN THE DISK
These oscillations are symmetric in the *hkl* disk and asymmetric in the 000 disk for two-beam conditions.

We'll see in Chapter 23 that these fringes are the reciprocal-space analog of rocking-curve intensity oscillations that occur across a bend contour in two-beam BF and DF images. We'll also see in Chapter 23 that bend contours arise when elastic deformation bends the diffraction planes and so an incoming parallel beam 'sees' a range of scattering angles across the bent region (see Figure 21.7A). In a similar manner, when you use a convergent beam and the illuminated region is undeformed, then the convergent beam provides a range of incidence angles to the diffracting *hkl* planes (see Figure 21.7B). The procedure to extract the thickness from the fringe pattern was first described by Kelly et al., refined by Allen, and is easily computerized. Energy filtering also helps to improve the clarity of the K-M lines.

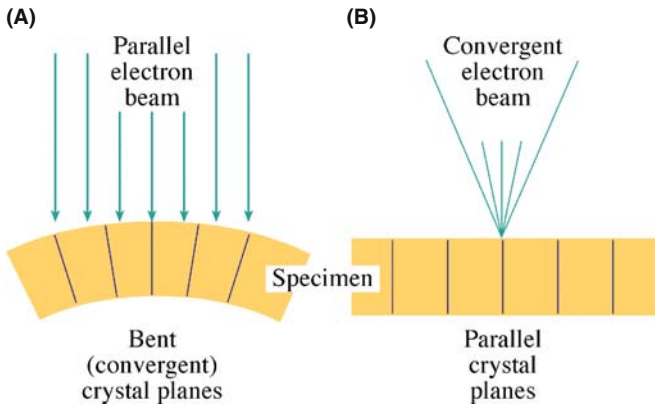


FIGURE 21.7 The reciprocal relationship between electron ray paths during (A) the formation of bend contours in BF images and (B) the formation of K-M fringes in CBED disks.

But before you click on your Web browser, let's do it first by hand so you understand the process. If you look at the *hkl* disk through a 10× Lupe containing a graticule then it is easy to measure the distances between the middle of the central bright fringe and each of the dark fringes with an accuracy of about ±0.1 mm. The central bright fringe is at the exact Bragg condition where $\mathbf{s} = 0$. The fringe spacings correspond to angles $\Delta\theta_i$ as shown schematically in Figure 21.8A, and from these spacings you can obtain a deviation $\mathbf{s}_i$ for the i^{th} fringe (where i is an integer) from the equation

$$s_i = \lambda \frac{\Delta\theta_i}{2\theta_B d^2} \tag{21.1}$$

where θ_B is the Bragg angle for the diffracting *hkl* plane, d is the *hkl* interplanar spacing, and we'll use the magnitude of **s**, ignoring its sign. The angle $2\theta_B$ in the CBED pattern is, of course, just the separation of the 000 and *hkl* disks. The specimen in Figure 21.6 is pure Al and the 200 reflection is excited. For Al, d_{200} is 0.2021 nm. If the extinction distance ξ_g is known, then you can determine the foil thickness t since

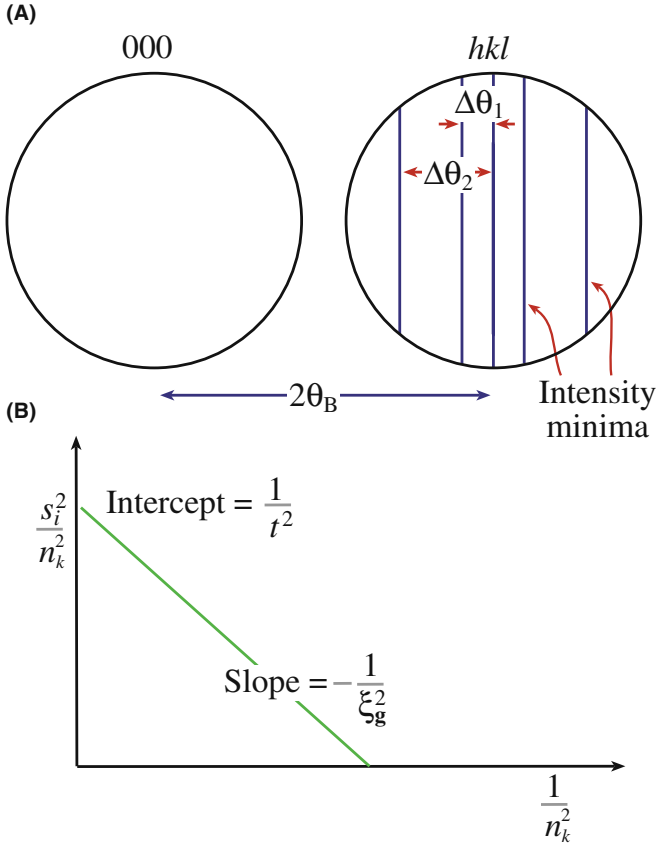


FIGURE 21.8 (A) The measurements necessary to extract thickness (t) from K-M fringes. From n_i measured spacings of $\Delta\theta_i$, determine the deviation parameters s_i, then (B) plot $(s_i/n_k)^2$ against $(1/n_k)^2$. If the plot is a straight line, extrapolate to the ordinate to find t^{-2} and hence t.

$$\frac{s_i^2}{n_k^2} + \frac{1}{\xi_g^2 n_k^2} = \frac{1}{t^2} \tag{21.2}$$

where n_k is an integer (k is an integer identical to i or differing from i by a constant integer not related to λ). If you don't know ξ_g, then you have to use a graphical method, plotting the measurements for several fringes as follows.

- Arbitrarily assign the integer $n = 1$ to the first fringe, which corresponds to an excitation error s_1.
- Then assign $n = 2$ to the second fringe, s_2, etc.
- Plot $(s_i/n_k)^2$ versus $(1/n_k)^2$. If the result is a straight line, your arbitrary assignment was good. That is, the relationship between i and k is given by $k = i + j$ where j is the largest integer $< (t/\xi_g)$.
- If your plot is a curve, then repeat the procedure by re-assigning $n = 2$ to the first fringe.
- Continue to iterate until you find a straight line, as shown in Figure 21.8B.

You have to do all this because the minimum thickness may be $> \xi_g$. From the straight-line plot, the intercept is t^{-2} and the slope is $-\xi_g^{-2}$. We will now go through an example in detail.

An example: If we apply this method to Figure 21.6, we find that the first set of values of s_i for the three dark fringes are s_1, s_2, and s_3 given in Table 21.1. Now we guess the values of n as shown in column 2 to give the values for $(s_j/n_j)^2$ in column 3.

These data do not plot as a straight line because both $(s_1/1)^2$ and $(s_3/3)^2$ are less than $(s_2/2)^2$. So we then assign the integer 2 to the first fringe, etc. We then find a second set of values as shown in Table 21.2, and these numbers plot as a straight line as shown in Figure 21.8B. The intercept of the line with the ordinate is $1/t^2$, and this

TABLE 21.1. CBED Data for Thickness Determination

s_i (nm^{-1})	n_i	$\frac{s_i^2}{n_i^2}$ (nm^{-2})
$s_1 = 0.84 \times 10^{-2}$	$n_1 = 1$	0.7×10^{-4}
$s_2 = 2.1 \times 10^{-2}$	$n_2 = 2$	1.1×10^{-4}
$s_3 = 3.0 \times 10^{-2}$	$n_3 = 3$	1.0×10^{-4}

TABLE 21.2 Alternative CBED Data for Thickness Determination

s_i (nm^{-1})	n_i	$\frac{s_i^2}{n_i^2}$ (nm^{-2})
$s_1 = 0.84 \times 10^{-2}$	$n_1 = 1$	1.7×10^{-5}
$s_2 = 2.1 \times 10^{-2}$	$n_2 = 2$	4.9×10^{-5}
$s_3 = 3.0 \times 10^{-2}$	$n_3 = 3$	5.6×10^{-5}

equals $6.1 \times 10^{-5}\,\text{nm}^{-2}$. Therefore, we find that $t^2 = ((6.1)^{-1} \times 10^5)\,\text{nm}^2 = (1.64 \times 10^4)\ \text{nm}^2$ and so $t = 128$ nm.

Now you can see why this procedure lends itself to computerization. You can digitize the fringes by scanning the pattern on your TEM screen across the STEM detector or do it directly via a CCD camera. As we noted, software is available to do this analysis but this isn't difficult and you may even want to try writing the program yourself or incorporate equation 21.2 into a spreadsheet/graphing program. A good example of the relative strengths of CBED and EELS thickness measurements for a nanostructured material is given by Berta et al.

21.3 UNIT-CELL DETERMINATION

Before you start on the more esoteric aspects of crystal-structure determination, such as the analysis of point groups and space groups, you can make life much easier for yourself by determining the unit cell of your specimen. Remember

> The unit cell is the smallest group of atoms or molecules which, when repeated in 3D, generates the lattice structure of the crystal.

The unit cell possesses the symmetry and properties of the lattice (assuming we could ever test a single, isolated unit cell!). In fact, determination of the unit cell is only possible if you already know the crystal system of the specimen. So don't start with an unknown structure! Remember

> There are seven different axial/coordinate systems into which all crystals are classified: cubic, orthorhombic, tetragonal, hexagonal, trigonal, monoclinic, and triclinic.

Now in TEM investigations it is rare that we look at a totally unknown specimen and so, in this chapter, we'll assume that you know the crystal structure of your specimen. If in fact you don't know the structure, then you have to start with symmetry determination to find the point group/space group (also see the companion text) and then you can deduce the crystal system. Remember

> The point group defines a set of 2D symmetry operations (rotations or reflections), acting through a point, which leaves the crystal unchanged. There are 32 unique point groups.

We can extend the definitions of symmetry to 3D. Remember

> The space group is the complete set of 3D symmetry operations (rotations, reflections, translations, and combinations thereof) that leave the crystal unchanged. There are 230 space groups.

WARNING

If you don't recognize any of these definitions then stop, do not pass this section. Return to the basic crystallography texts (referenced in Chapter 18).

So, assuming you know the crystal system, let's determine the unit-cell dimensions. We saw in the previous chapter that a CBED pattern at small L often reveals one or more rings of HOLZ intensity and you've just learned how to index the diffraction disks that make up these rings and the pairs of HOLZ lines that attach themselves to the HOLZ reflections. If you don't know the crystal structure, then, of course, it will be rather difficult to index the pattern since you don't know the appropriate systematic absences. But these rings are useful, even if you haven't indexed the individual disks.

If you measure the radii of the rings (G), you can deduce the lattice-repeat vector of the crystal parallel to the beam direction.

So by tilting to an orientation in which the beam is coming down an axis of the unit cell such as [001] in an orthorhombic crystal, the disk spacings in the ZOLZ pattern will give you the [100] and [010] lattice parameters and the HOLZ-ring radius will give you [001]. (Now you see why you can't do this if you don't know the crystal system to start with, since you don't know what lattice spacings and angles to expect.)

So you should be able to determine all the lattice parameters of the unit cell from a single ZOLZ/HOLZ pattern. If you're not sure which pattern to choose, any low-index (i.e., high-symmetry) pattern is a good starting point. There are appropriate analytical expressions for calculating the spacing between atomic planes parallel to the beam and we'll discuss them next. These expressions give you the lattice parameters, since the lattice spacing is related to the lattice parameter by standard equations, given in basic crystallography texts (see Chapter 18). Then you have to look at differences between the ZOLZ and HOLZ disk patterns to determine the type of lattice centering. So we'll now show you how to utilize a unique aspect of CBED patterns, namely, that from a single 2D pattern, you can obtain 3D information about your crystalline specimen.

21.3.A Experimental Considerations

The first thing you have to do is get CBED patterns containing clear ZOLZ and HOLZ maxima. The patterns should have a small L to reveal one or more rings.

TWO PATTERNS

It is good practice to record two patterns, one with a large C2 aperture (therefore large α, Kossel conditions) and one with a small C2 aperture to show the individual disks (K-M conditions).

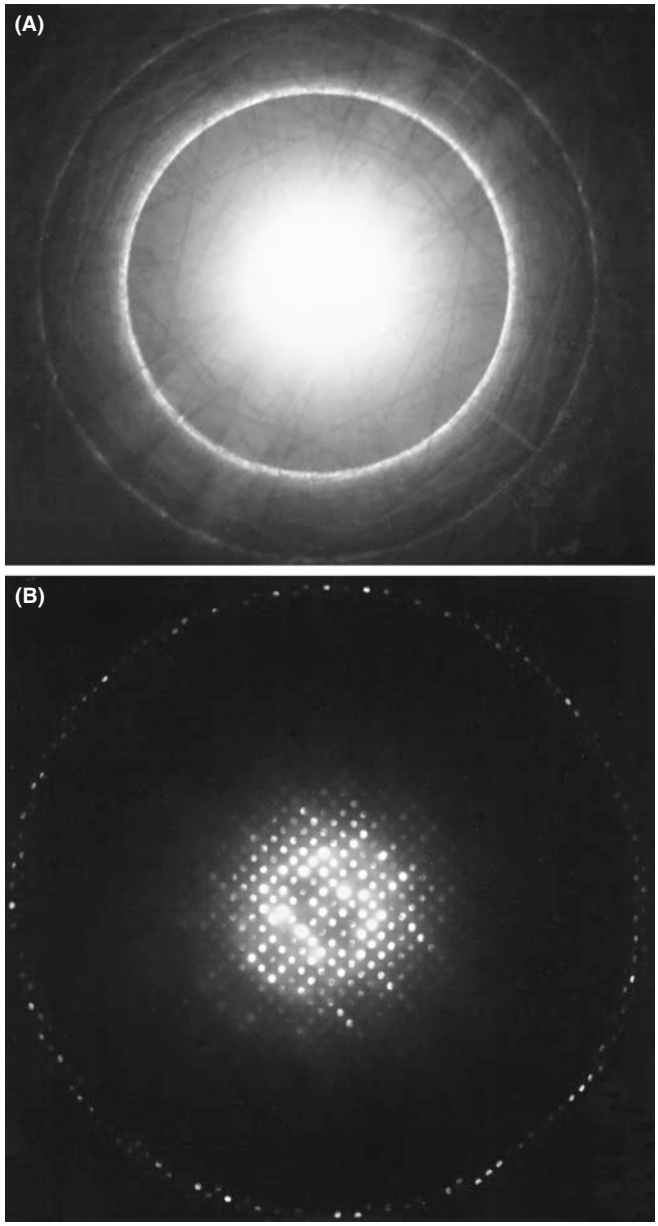


FIGURE 21.9 (A) CBED Kossel pattern from a carbide particle taken with a 150-μm C2 aperture showing the FOLZ ring of intensity surrounding the overexposed ZOLZ region. In (B) the same pattern taken with a 20-μm aperture reveals individual reflections in both the ZOLZ and FOLZ.

Such a pair of patterns is shown in Figure 21.9A and B. We will use the ring pattern to measure G and the disk pattern to index individual HOLZ reflections and observe both the relative spacings and positions of ZOLZ and HOLZ reflections.

21.3.B The Importance of the HOLZ-Ring Radius

If you go back and look at Figures 20.11A–D and 20.12 you will be able to work out the simple geometrical relationship between H and G.

- H in Figure 20.12 is the spacing of the reciprocal-lattice planes parallel to the electron beam.
- G_n is the projected radius of a HOLZ ring that you measure on the recording of the pattern. If the HOLZ ring is split, always measure G_n using the innermost ring.
- If the order of the ring is high, then the electrons are scattered out to large angles ($\geq \sim 10^\circ$), and your measurements may suffer from the effects of lens distortion because the electrons are traveling close to the polepiece. If this is happening, you must calibrate the distortion in reciprocal space using a known specimen from which G_1, G_2, etc., can be calculated and compared with the values you obtain experimentally. Of course, C_s correction will minimize such distortion.

Experimentally, you'll find it much easier to measure G from a Kossel pattern because the HOLZ intensity appears as one or more rings as in Figure 21.9A. Since the radius of the Ewald sphere increases with decreasing electron wavelength λ, the value of G for any given orientation will increase as the accelerating voltage is raised, so it becomes increasingly difficult to see HOLZ rings at intermediate voltages.

From the geometry of Figures 20.11 and 20.12, and assuming that terms in H^2 are negligible, the radii of the FOLZ and SOLZ rings, G_1 and G_2, are

$$G_1 = \left(\frac{2H}{\lambda}\right)^{1/2} \tag{21.3}$$

And from a theorem you learned in high-school geometry

$$G_2 = 2\left(\frac{H}{\lambda}\right)^{1/2} \tag{21.4}$$

where both G and H are in reciprocal-space units (nm^{-1} or $Å^{-1}$). Similar expressions can be developed, if you need them, for third- and higher-order zones. In practice, most people find it easier to think in real space rather than reciprocal space and so we rewrite these equations in terms of the spacing between Laue zones (H^{-1}) in real-space units. We use the inverse relationship between real and reciprocal spaces to give, for the FOLZ

$$\frac{1}{H} = \frac{2}{\lambda G_1^2} \tag{21.5}$$

The value of H^{-1} can be expressed in real-space units (nm) through the measured radius R (mm) and the camera constant λL (nm mm)

$$\frac{1}{H} = \left(\frac{2}{\lambda}\right)\left(\frac{\lambda L}{R}\right)^2 \qquad (21.6)$$

You must take the time to measure λL carefully (see Section 9.6) because this will minimize errors in H^{-1} which could be quite large due to the $(\lambda L)^2$ dependence in equation 21.6. From the above equations and from Figure 20.12, you can see that a low-symmetry zone axis with a small H will give rise to a smaller HOLZ ring than for a high-symmetry zone. The smaller ring will be easier to observe at any chosen L.

Summarizing the story so far: by measuring R values, you can determine the real-lattice spacing (H^{-1}) parallel to the beam direction. The next thing to do is compare this measured value, H_m^{-1} with calculated values H_c^{-1} assuming a certain unit cell. Now H^{-1} is directly related to the magnitude of the real-space direction vector

$$\frac{1}{H} = |[UVW]| \qquad (21.7)$$

and so this magnitude can be calculated for a specific beam direction $[UVW]$.

An example: For an fcc crystal (Steeds 1979)

$$\frac{1}{H} = \frac{a_0(U^2 + V^2 + W^2)^{1/2}}{p} \qquad (21.8)$$

where a_0 is the lattice parameter, $p = 1$ when $(U+V+W)$ is odd, and $p = 2$ when $(U+V+W)$ is even. For bcc crystals $p = 2$ if U, V, and W are all odd, otherwise $p = 1$. These conditions for p just take account of structure-factor effects which cause systematic absences of some reflections, or in some cases whole rings. If a whole ring is absent, the calculated reciprocal-lattice layer spacing H_c^{-1} must be an integer multiple of the measured spacing H_m^{-1}. Thus

$$\frac{1}{H_c} = n\left(\frac{1}{H_m}\right) \qquad (21.9)$$

where n must be an integer. If n is non-integral then your indexing is wrong. A generalized method for determining which Laue zone you should see has been given in Jackson's 1990 paper. If you have indexed the ZOLZ (i.e., $[UVW]$ is known), R is measured and λ is known, then you can determine H without the need to index individual spots in the HOLZ ring of intensity.

There are more generalized equations for H^{-1} than equation 21.8 (see Raghavan et al. and Ayer).

Other examples: In a crystal system with orthogonal axes (orthorhombic, tetragonal, or cubic systems, with lattice-repeat spacings a, b, c) if there are no absences of HOLZ layers ($p = 1$) then for a given zone axis UVW

$$\frac{1}{H} = \left(a^2U^2 + b^2V^2 + c^2W^2\right)^{1/2} \qquad (21.10)$$

Similarly for hexagonal or rhombohedral systems using a three-index system

$$\frac{1}{H} = (a^2(U^2 + V^2 - UV) + c^2W^2)^{1/2} \qquad (21.11)$$

and for the four-index system

$$\frac{1}{H} = (3(U^2 + V^2 + UV)^2 + c^2W)^{1/2} \qquad (21.12)$$

and for the monoclinic system with a unique b axis

$$\frac{1}{H} = (U^2a^2 + V^2b^2 + W^2c^2 + 2UWac\cos\beta^2)^{1/2} \qquad (21.13)$$

If you are working with a low-index, high-symmetry zone axis it may be just as easy to determine H^{-1} directly from reciprocal-lattice constructions rather than using equations. However, for lower-symmetry crystallographic directions, such constructions are effectively impossible to visualize and then you should use these equations.

So in summary we can give some guidelines

- Measure the radius of the HOLZ ring to give a value of the reciprocal of the spacing between the HOLZ and the ZOLZ, H_m^{-1}.
- Compare the measured spacing with the spacing calculated assuming a given unit cell, H_c^{-1}.
- The measured value should agree with, or be a multiple of, the calculated value. For example, if given a square ZOLZ DP, you assume a cubic crystal then the unit-cell repeat vector should be identical in all three dimensions. So the FOLZ ring diameter should give the same value of H^{-1} as that determined for the other two axes from the square [100] pattern. If H^{-1} is different, then the crystal is not cubic but another system, such as tetragonal.

21.3.C Determining the Lattice Centering

When you have measured H^{-1} from the Kossel pattern, the next thing to do is to compare the ZOLZ and FOLZ reflections in the K-M pattern obtained with a small C2 aperture such as Figure 21.9B. The superposition of the FOLZ and ZOLZ gives you information on the type of lattice you are dealing with, since centered lattices of all types will give different superposition patterns compared with a primitive lattice.

In a primitive lattice, the FOLZ will superimpose directly on the ZOLZ because there are no systematic absences. However, face-centered and body-centered lattices will give rise to displacements of the FOLZ pattern with respect to the ZOLZ in certain beam directions as illustrated schematically in Figure 21.10A. You can quite easily work out the displacement in terms of a shift vector for cubic crystal patterns in low-index orientations and we showed examples back in Figure 21.2 and 21.3. It is not so simple in more complex crystals but Jackson (1987) has developed a generalized method of determining the shift vector **t** for all crystal systems and all orientations

$$\mathbf{t} = \mathbf{g} - \mathbf{u}^* \left(\frac{HN_L}{|\mathbf{u}^*|} \right) \tag{21.14}$$

where **g** is the vector for the *hkl* HOLZ reflection, **u*** is the vector normal to the ZOLZ and parallel to *H*, and N_L is the number of the Laue zone containing *hkl*. To determine **t** then all you do is look up values of *H*, **u***, and $H/|\mathbf{u}^*|$, tabulated by Jackson.

An example: We can illustrate the shift due to lattice type, by looking at Figure 21.10B, which is a series of schematic patterns for an orthorhombic cell (given by Ayer) oriented along the [001] axis. In each pattern the experimentally observed distribution of ZOLZ and FOLZ reflections is shown and adjacent to it is the same pattern but containing the FOLZ reciprocal-lattice points. So the FOLZ ring of spots is always coincident with the FOLZ reciprocal-lattice points. In the top pattern (P) the ZOLZ and FOLZ superimpose exactly and this would be the case for a primitive unit cell. In (A) the FOLZ lattice is displaced from the ZOLZ reflections by half the spacing of the ZOLZ reciprocal-lattice points in the [010] direction; this is the situation expected for an A-face-centered lattice. The next two patterns (B and I) show the expected displacements for a B-face-centered and an I (body-centered) lattice.

So you now know how to measure the lattice-repeat vectors in three dimensions and determine the type of lattice centering. This information should be sufficient to allow you to determine the correct unit cell of your specimen, particularly if you have further information, such as elemental analysis by XEDS or EELS.

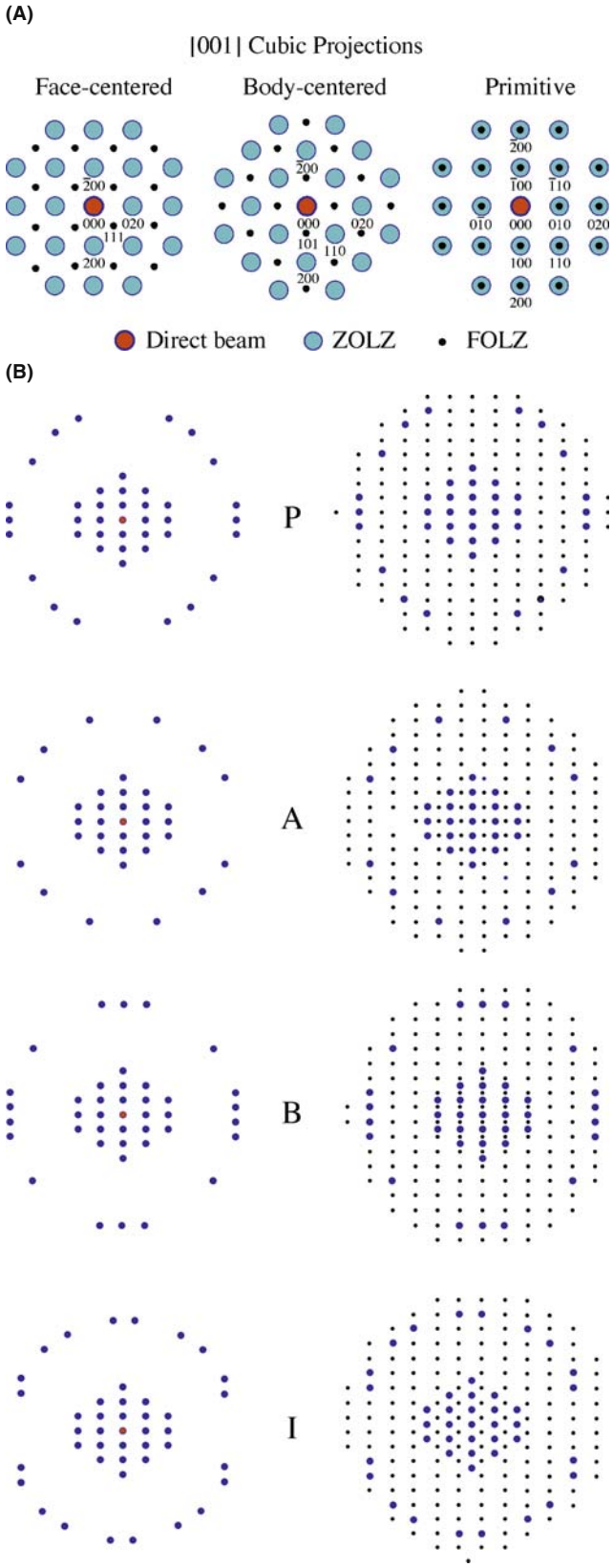


FIGURE 21.10 (A) The overlap between the ZOLZ and the FOLZ when looking down the [001] axis of cubic crystals. In the fcc pattern, 111 is a FOLZ index, likewise 101 in bcc. In the primitive pattern, only the ZOLZ is indexed. (B) Schematic illustration of the superposition of the FOLZ pattern on the ZOLZ pattern for an orthorhombic crystal with the electron beam down [001] showing the differences in the superposition for (P) primitive, (A) A-centered, (B) B-centered, and (C) I-centered lattices.

21.4 BASICS OF SYMMETRY DETERMINATION

21.4.A Reminder of Symmetry Concepts

Before you study the following two sections in the companion text, you must have a basic understanding of crystal-symmetry elements (both rotational and

translational) and be familiar with the standard international notation for point groups and space groups (see Section 16.10). You will also need to know how to represent the point-group symmetry of a crystal using the stereographic projection we discussed in Section 18.4. In Figure 21.11 we reproduce the standard point-group table familiar to any student of crystallography.

Historically, point-group determination has been the domain of X-ray crystallographers and electron microscopists have gladly avoided such concepts. However, the point group is not only useful for classifying crystals with common symmetry elements, but it is also an important indicator of many of the properties of the crystal, such as anisotropy in the electrical resistivity or the refractive index (all described in glorious detail in the classic text by Nye). With the availability of CBED you can determine the point group of a crystal directly in the TEM, simply by recording two or three low-index ZAPs.

This process has a tremendous advantage over classic X-ray techniques because

- We obtain this information from much smaller regions than is possible using X-rays.
- We can also distinguish all 32 possible point groups unambiguously, which is a non-trivial process using X-rays.

So as an electron microscopist, you must now master the details of point-group determination if you are to take full advantage of the capabilities of a modern TEM. This is described in detail in the companion text and you'll see that it's not too difficult an exercise, but there is no escaping the need to comprehend some of the basic principles of crystal symmetry. The exercises will require quite a lot of careful work and time. Therefore, if you can uniquely identify the unknown phase in your specimen through some other technique, such as XEDS or EELS, do so.

21.4.B Friedel's Law

Symmetry determination in crystals has evolved from the early work of Friedel and von Laue on the kinematical theory of XRD. We can summarize a fundamental aspect of Friedel's work by Friedel's law.

FRIEDEL'S LAW

Under kinematical diffraction conditions, the intensity of a reflection hkl is equal to the intensity in its opposite reflection $\bar{h}\bar{k}\bar{l}$.

If this law holds, then it is not possible to tell by diffraction whether an inversion center is present in the crystal or not. This is the case for most XRD patterns from single crystals, because most XRD occurs under kinematical conditions.

So, under kinematical diffraction conditions, we cannot readily distinguish a mirror plane from a twofold rotational axis (diad) parallel to the mirror plane. That is equivalent to saying that *we cannot distinguish point groups m and 2*. Similarly, the presence of a fourfold rotation axis (tetrad) in a crystal, parallel to the c axis, results in $I_{hkl} = I_{\bar{h}kl} = I_{\bar{h}\bar{k}l} = I_{h\bar{k}l}$ where I_{hkl} denotes the diffracted intensity of a reflection of type hkl. This DP cannot be distinguished from one containing two perpendicular mirror planes whose intersection is parallel to the rotation axis.

MIRROR PLANE

The presence of a mirror plane in a crystal, parallel to axes a and b, makes the intensity of all reflections of type hkl equal to the intensity of the corresponding $\bar{h}\bar{k}l$ reflection.

XRD is thus severely limited for point-group determination because of Friedel's law. Since crystals which do not possess true centers of symmetry (non-centrosymmetric crystals) still appear in X-ray DPs to possess a center of symmetry; they cannot be readily distinguished from centrosymmetric crystals. If you go back and look at the 32 point groups in Figure 21.11 and remove all those which do not contain a center of symmetry, then you are left with only 11 centrosymmetric point groups: $\bar{1}$, $2/m$ (equivalent to mm), mmm, $\bar{3}$, $\bar{3}m$, $4/m$, $4/mmm$, $6/m$, $6/mmm$, $m\bar{3}$, and $m\bar{3}m$. These 11 point groups are known as the *Laue classes* in XRD. Except under 'anomalous' scattering conditions, XRD can only determine these 11 symmetry groups.

WHEN FRIEDEL FAILS

In CBED patterns, *Friedel's law* breaks down because of dynamical scattering.

So, to get the full symmetry information, the crystal must be thick enough for you to see dynamical diffraction contrast within the CBED disks. If you then examine the intensity distributions within individual hkl reflections you can distinguish centrosymmetric and non-centrosymmetric crystals. Thus, in CBED patterns, the 32 crystal point groups are not reduced to the 11 Laue classes, as occurs in XRD.

21.4.C Looking for Symmetry in Your Patterns

You can't perform point-group determination without learning how to discern symmetry elements in

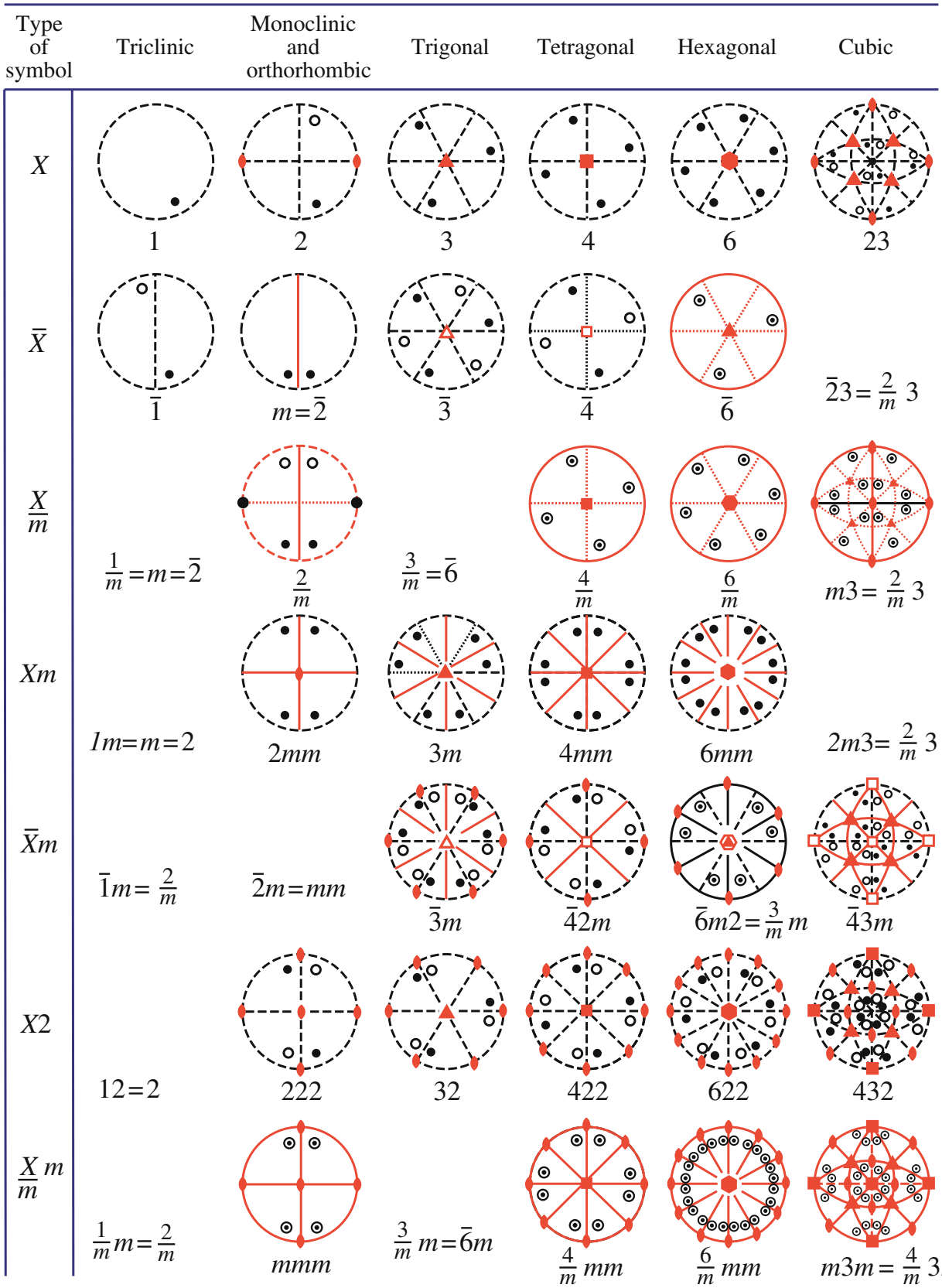


FIGURE 21.11 The 32 crystal point groups represented by stereograms showing the operation of rotational, mirror, and inversion symmetry elements on a general pole *hkl*. The international notation describing the point groups is given under each of the stereograms.

CBED patterns. When we look at CBED-pattern symmetry we use the same notation as for point groups, i.e., a number X (= 1, 2, 3, 4, or 6) for a rotation axis and m for a mirror plane parallel to the rotation axis and a second m for any independent mirror plane. Inversion symmetry or a mirror normal to the beam direction cannot be discerned and so the terms of the form $\overline{X}$ or X/m are not used. The only combinations we can get are the same as for the 10 2D point groups: 1, 2, m, 2mm, 3, 3m, 4, 4mm, 6, or 6mm. These symbols refer to the observable symmetry in the pattern, and four examples of different pattern symmetries are shown schematically in Figure 21.12.

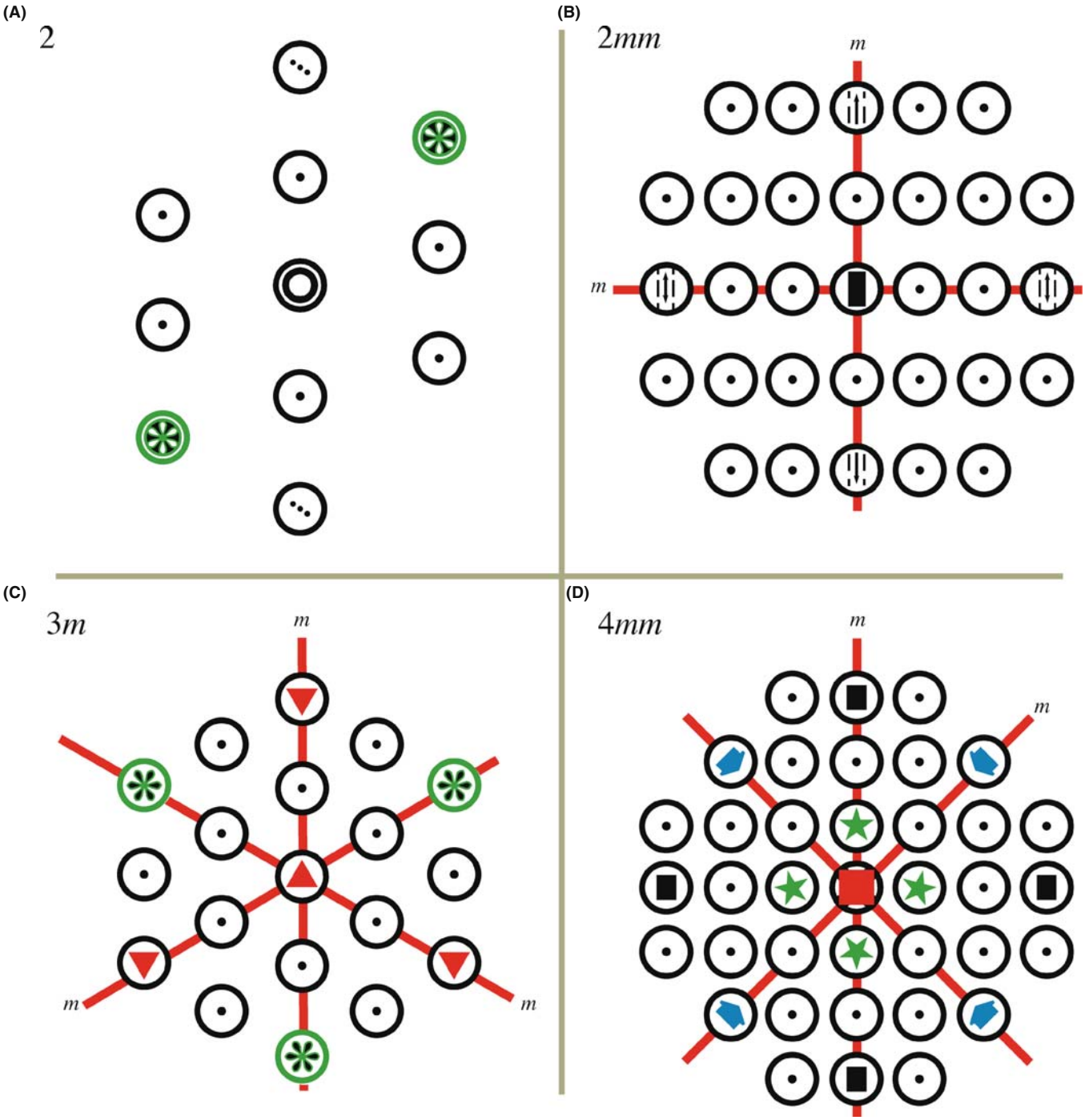


FIGURE 21.12 Four examples of symmetry in CBED patterns. (A) Symmetry 2 refers to a twofold (diad) rotation axis; i.e., the pattern has identical symmetry when rotated 180°. (B) 2mm is a diad symmetry with two independent mirror planes parallel to the diad. (C) 3m indicates threefold rotation (triad) symmetry with one mirror plane, i.e., rotational symmetry every 120° with one mirror plane present at each 120°. (D) 4mm indicates a fourfold rotational symmetry (tetrad) with two independent mirror planes parallel to the tetrad.

Symmetry determination is generally carried out using ZAPs although there is useful information in off-axis patterns. Remember, to get a well centered ZAP

- You may find it easier to do the final adjustments to get an exact ZAP using the beam tilt/shift controls.
- As a last resort, you may have to displace the C2 aperture slightly off axis to center it precisely around the center of symmetry in the ZAP.

Once you've got a centered ZAP, there are specific kinds of symmetry for which to search

- Whole-pattern (WP) symmetry.
- Bright-field (BF) symmetry.
- Projection-diffraction symmetry.

The first and most important is the *WP symmetry*.

To be sure you get the correct symmetry, you should take a small camera-length pattern to include the HOLZ rings since HOLZ effects are not always visible in the ZOLZ (e.g., they may be hidden by the intense 000 disk). The pattern can be either a Kossel or a K-M pattern. The WP symmetry at any orientation must belong to one of the 10 2D point groups listed.

WP SYMMETRY

The WP symmetry is just what it says: it's the symmetry of the whole pattern, including the relative positions of the HOLZ reflections and any HOLZ Kikuchi lines.

The second kind of symmetry is the *BF symmetry*.

In this case, the presence of the HOLZ lines ensures that the BF-disk symmetry contains 3D information. Take care to ensure that the C2 aperture is small enough so that you can see the 000 disk without any overlap from other diffraction disks. For example, the 000 disk back in Figure 20.2B has an array of deficient HOLZ lines which displays 3*m* BF symmetry.

If there is only 2D diffuse intensity within the disk, or if you ignore the HOLZ lines, then the symmetry is more correctly called the *BF projection symmetry*. The symmetry in either of these latter cases should again be classified into one of the 10 2D groups just listed. You'll see in the companion text that combination of the WP and BF symmetry in three ZAPs is often sufficient to determine the point group.

BF SYMMETRY

BF symmetry is the symmetry of the 000 disk only, *when HOLZ lines are present*.

On some occasions the *projection-diffraction symmetry* may be all that is available.

The diffuse contrast within these disks arises from dynamical interactions within the zero-order reciprocal-lattice layer of the crystal that give rise to K-M fringes, which we used above for thickness determination. The projection-diffraction symmetry is simply the symmetry displayed in SADPs. Since this symmetry is only 2D, it is not as useful as the WP and BF symmetry. If you go back and examine Figure 20.2B, the projection-diffraction symmetry is visible in the ZOLZ Kikuchi bands which show a sixfold rotational symmetry with two independent mirror planes, one within the Kikuchi bands and one between them, giving 6*mm* symmetry. Similar symmetry is shown in the SADP from Si in Figure 20.2A.

PROJECTION-DIFFRACTION SYMMETRY

This is displayed by the intensity in the direct 000 beam plus the *hkl* diffracted beams in the ZOLZ layer and corresponds to the projected 2D symmetry of the crystal down the zone axis that you have selected. It ignores any contributions from HOLZ layers, such as HOLZ lines and HOLZ reflections, but includes any diffuse intensity within these disks.

Now that you've learned the basics you should be able to breeze through the companion text which describes the process for point-group determination, step by step. After point-group determination you'll also be shown how to determine space groups. This exercise is a little more challenging and space groups are somewhat less important than point groups in terms of relating crystal symmetry to materials properties. Nevertheless, the space group of your specimen is an important crystallographic characteristic and since there are 230 of them it is a much more selective characteristic than the point group.

We'll now go on to discuss some more information that is available in HOLZ lines but, before you do this and before you read the CBED chapter in the companion text, we must warn you that symmetry determination from CBED patterns requires that you use your eyes to make a judgment on the presence or absence of certain symmetry elements and so this is not an exact science. There are 'error bars' in CBED symmetry and guidance on how to judge whether imperfections in the symmetry of the patterns can be ignored or not is also given in the companion text.

21.5 LATTICE-STRAIN MEASUREMENT

Strain arises from local changes in the lattice spacings of the material. Strain may arise from the presence of defects such as dislocations and coherent precipitates

or through local changes in chemistry. Such chemical changes can arise naturally or can be carefully introduced by controlling the deposition parameters for layered nanostructures formed by MBE or CVD methods. Either way, knowing the exact value of the strain is important in areas as diverse as mechanics, electronics, and optics. We can get a reasonably accurate measurement of the lattice constants (~2%) by indexing the reflections in the ZOLZ and/or by measurement of the HOLZ-ring diameter. However, the best method is to use the positions of HOLZ lines in the BF disk which, because they arise from very high order reflections, are very sensitive to changes in lattice parameter. Results that are an order of magnitude more precise (~0.2%) can be obtained by computer simulation of the position of the HOLZ lines using different lattice constants. The values that produce the best match with the experimentally observed HOLZ-line positions can be identified as the lattice constants of the crystal. So by measuring the HOLZ-line positions we can measure changes in lattice parameters. Materials scientists and nanotechnologists, particularly in the electronics industry, are most interested in changes due to lattice strain.

Local measurement of lattice strain has always been of interest to materials scientists and indeed we devote the whole of Chapter 25 to imaging of strain effects. Typically, direct lattice-strain measurements were performed by X-ray methods but the need for higher spatial resolution in semiconductor technology and nanostructured materials has created a tremendous market for CBED-based methods. For example, in semiconductor manufacturing and optical technologies, strain-layer superlattices (e.g., epitaxial layers of GaAs and AlGaAs which differ slightly in their precisely controlled lattice parameters) are widely used. Specific infrared sensors can be tuned, optical waveguides can be modulated, the threshold voltages of lasers can be tweaked, and so on. Introducing local strain into the channel in semiconductor junctions to enhance carrier mobility and reduce the resistance between the source and drain has increased chip performance markedly, permitting Moore's law to be extended for yet a few more years. This concept is also referred to as 'strain engineering' and is central to the success of nanostructured electronics and optical devices.

Early CBED work in this area involved the measurement of strain around precipitates in thin foils (Rozeveld and Howe) and this has been extended down to nanometer-size probes coupled with sophisticated algorithms for matching the HOLZ-line patterns and extracting the strain measurement. Figure 21.13 shows the extraordinary match that can now be achieved between experiment and theory using Zuo's software which is a development of the early algorithms given in Spence and Zuo. Similar software is continuously being refined by many researchers and various forms are available on the Web at the URLs listed at the end of the chapter. As always, check the validity of the software by checking how well it simulates patterns from known specimens in known orientations: caveat emptor even if it's free!

In early computer programs, the position of the HOLZ lines in the simulation was derived from kinematical diffraction theory only, but dynamical effects may be important and should always be included, otherwise the match is incorrect (Eades et al.). The method is summarized in three steps.

- Start with a standard specimen of known lattice parameter to establish the exact electron wavelength for subsequent simulations.
- Adjust the continuous kV control. Since you know the lattice parameter, you can determine the exact kV.

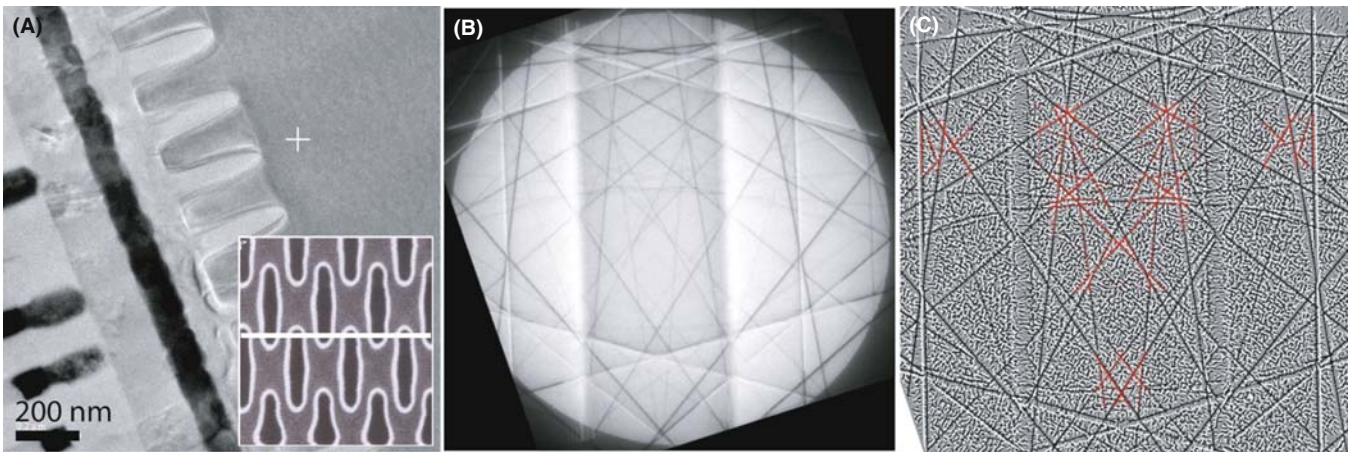


FIGURE 21.13 (A) Cross-sectional BF image from a high-density, plasma-filled, shallow-trench isolation structure prepared along the line indicated on the SEM image (inset). (B) Experimental CBED HOLZ pattern taken at the cross in (A) ~10 μm away from the trench structure. (C) Simulation of a selected area of (B) where the HOLZ lines (highlighted in red) intersect showing the best fit for strain analysis. Both experimental and theoretical patterns were processed to enhance the contrast.

- At this predetermined kV setting, obtain a HOLZ-line pattern from the unknown and compare the experimental pattern with simulated patterns generated for a range of lattice parameters, until good matching is achieved between the simulated and experimental patterns (see Figure 21.13). Theoretically, an accuracy of 0.02% should be achievable but in practice an accuracy of 0.2% is generally obtained.

LATTICE PARAMETERS FROM CBED

For determining precise values, compare the experimental HOLZ-line patterns to the computer-simulated version.

While this approach has demonstrated reasonable success in measuring lattice-parameter shifts as reviewed by Randle et al., you should be wary of other possible causes of HOLZ-line shifts and the difficulties of exact matching between theory and experiment. Often the HOLZ-line patterns display asymmetries which make matching difficult. Remember, you'll probably need to cool the specimen so you can see HOLZ lines in some materials. There are some necessary precautions you should take

- Compare your standard and unknown under identical conditions.
- Take account of any differences in thermal-contraction coefficient if you cool the specimen to liquid-N_2 temperatures.
- Watch for problems such as surface relaxation, or local strain, or the presence of dislocation lines which can give rise to spurious measurements of lattice parameter.

With this same approach we can also make localized measurements of strain around precipitates or defects, but the data are always averaged through the foil thickness and in one dimension and so the interpretation can be difficult. Furthermore, strain relaxation due to the presence of the surfaces of the specimen is an ever-present challenge. However, particularly in planar, thin-film layered structures it is often possible to tilt the specimen so that the local strain is projected along the beam direction. It is also feasible to use this HOLZ-line shift approach to infer, indirectly, the composition of a specimen by comparing the lattice parameter of an unknown with a standard of known composition and in the case of binary solid solutions, extrapolating to other compositions assuming that Végard's law applies. Such an analysis is not limited by elemental considerations or absorption, as is the case for XEDS, and does not require ultra-thin specimens as for EELS. However, this method is indirect and makes several assumptions that may not always apply. It's a last resort.

21.6 DETERMINATION OF ENANTIOMORPHISM

The terms 'handedness,' 'chirality,' and 'enantiomorphism' all refer to the same thing. We say that a crystal or molecule is 'handed' if two forms of the structure occur, and if we describe the atomic coordinates of one of them in a right-handed set of axes, the atomic coordinates of the other structure will be identical in a left-handed coordinate system. These two structures, termed 'enantiomorphic' or 'chiral,' are mirror images of each other and we refer to them as the right-handed and left-handed forms. (Look at your hands and try and rotate them into exact coincidence and all this should be clear.) If you go back and look at the section on symmetry concepts, you should be able to convince yourself that any crystal that contains a mirror (m), inversion center ($\bar{1}$), or a fourfold inversion axis ($\bar{4}$) cannot show chirality.

By the way, if you don't think such an esoteric topic as enantiomorphism is important then a salutary fact is that the drug thalidomide (2-(2,6-dioxo-3-piperidyl) isoindole-1,3-dione) comes in enantiomorphic crystal forms. One enantiomer (the *S*-form) causes catastrophic birth defects in children and the other (*R*-form) provides the desired therapeutic effects for women suffering difficult pregnancies and is particularly efficacious as a treatment for disfiguring skin lesions such as those arising from leprosy. Thousands of seriously deformed children were born because of the failure to distinguish the different effects of the two enantiomorphs. This tragedy gave rise to the whole field of single-enantiomer drug development (currently a $100B/yr market in the USA alone), avoiding the common racemic mixture of both forms that is produced in most chemical processes.

So, although we don't often need to know which of the two-handed forms we are looking at, there are occasions when it is of the utmost importance, and several groups have worked hard on the determination of handedness from CBED.

Enantiomorphism is an essentially 3D phenomenon (again, look at your hands). Since the projection of a crystal is a 2D object, a projection of the structure cannot tell you anything about the handedness. This means that, in ZAPs, to see handedness, we must look at effects of HOLZ diffraction or, more generally, if your specimen is not at a zone axis, there must be diffraction involving at least three *non-coplanar* diffraction vectors in addition to the direct beam. This condition does not necessarily mean that we have to look at a HOLZ pattern. It means that we have to look at HOLZ diffraction effects which, as you now know well, may appear in the ZOLZ as well as in HOLZ directly. In the example given in Figure 21.14, the asymmetric effects are visible only in the reflections in the first HOLZ ring.

Looking at HOLZ effects, we can readily see the difference between enantiomorphs. But just observing

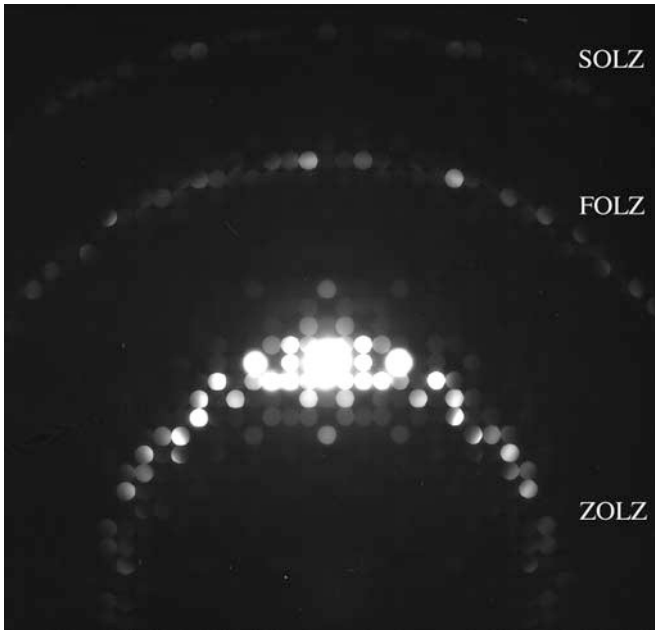


FIGURE 21.14 CBED pattern from a specimen of $Ho_2Ge_2O_7$, tilted slightly away from the zone axis to improve the contrast. The ZOLZ (the bright spots just below the center of the figure) shows mirror symmetry but in the HOLZ ring (above the center) the reflections are clearly asymmetric. If the specimen had the other chirality, the HOLZ ring would be mirrored across the vertical axis of the pattern.

the difference does not let us to say which structure is which hand. It does allow us to say whether two crystals have the *same* or *opposite* handedness.

The determination of whether a particular crystal is right-handed or left-handed requires a dynamical calculation to simulate CBED patterns which you must then compare with the experimental patterns. But note

- You'll find that such a comparison with experiment is not always easy because there is often ambiguity in the indexing of the pattern and, if your indexing is wrong, the handedness will be wrong.
- To perform the simulation and thus to determine the handedness, you need to know the full crystal structure.
- Some authors have suggested that the dynamical calculations (needed to say which enantiomorph is which) can be done by hand, e.g., by calculating the phases of the structure factors and working out the effects of multiple-diffraction paths. However, this calculation requires considerable understanding of dynamical diffraction and it is surely easier, as well as more reliable, to perform a full dynamical *simulation* using some of the software we've already referenced.

Recently Inui et al. have systematically tabulated appropriate zone axes for the determination of handedness for all the point groups that can show handedness. In studies of enantiomorphism, it is crucial to determine whether or not a CBED pattern contains mirror symmetry or not. You'll find that doing this is not always easy but, if you use off-axis patterns rather than ZAPs, the asymmetries tend to be stronger, as described by Jones (and also shown in Figure 21.14).

21.7 STRUCTURE FACTOR AND CHARGE-DENSITY DETERMINATION

When you read the literature about CBED and its applications, you'll often come across references to extraction of structure factor (F_{hkl}) values via CBED. (Go back to Chapter 3 to remind yourself of the importance of structure factors in electron diffraction.) You should remember that $|F(\theta)|^2$ is direct measure of the diffracted intensity. Direct, reproducible, and quantitative measurement of the diffracted intensity is basically impossible in SADPs but is quite feasible in CBED patterns, especially when they are energy-filtered. When we determine a specific crystal structure for our specimen, it is, in fact, equivalent to assigning a structure-factor phase to the diffraction disks in the pattern. There are

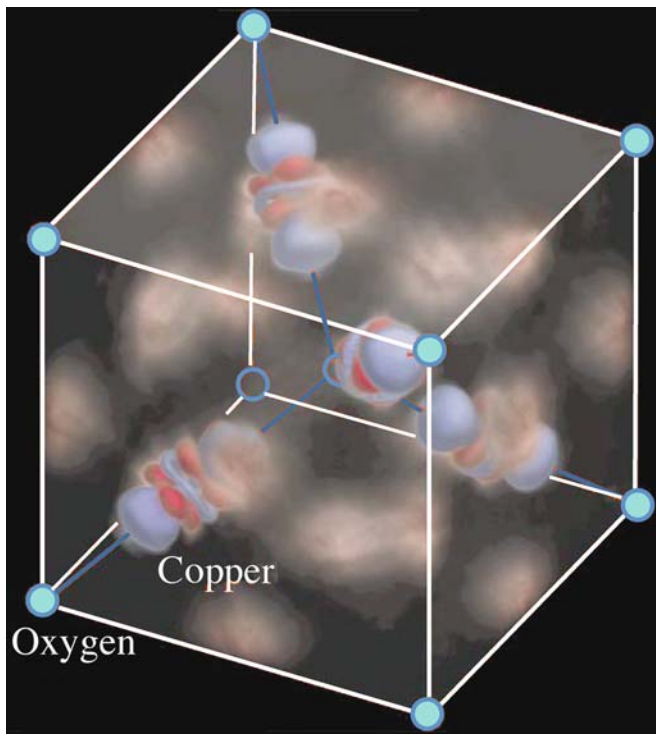


FIGURE 21.15 A three-dimensional rendering of the experimental difference map between the static crystal charge density (determined from quantitative analysis of CBED patterns) and the superimposed spherical charge density of O^{2-} and Cu^{+} ions. The color scheme is blue where the difference in charge density is negative ($\Delta\rho < 0$), white where it is zero ($\Delta\rho = 0$), and red where it is positive ($\Delta\rho > 0$). The difference would be zero everywhere if cuprite were purely ionic (i.e., the ions were indeed spherical). The nonspherical charge distortion on the Cu atoms shows the characteristic shape of d orbitals. The positive charge in the tetrahedral interstitial regions suggests Cu–Cu covalent bonding.

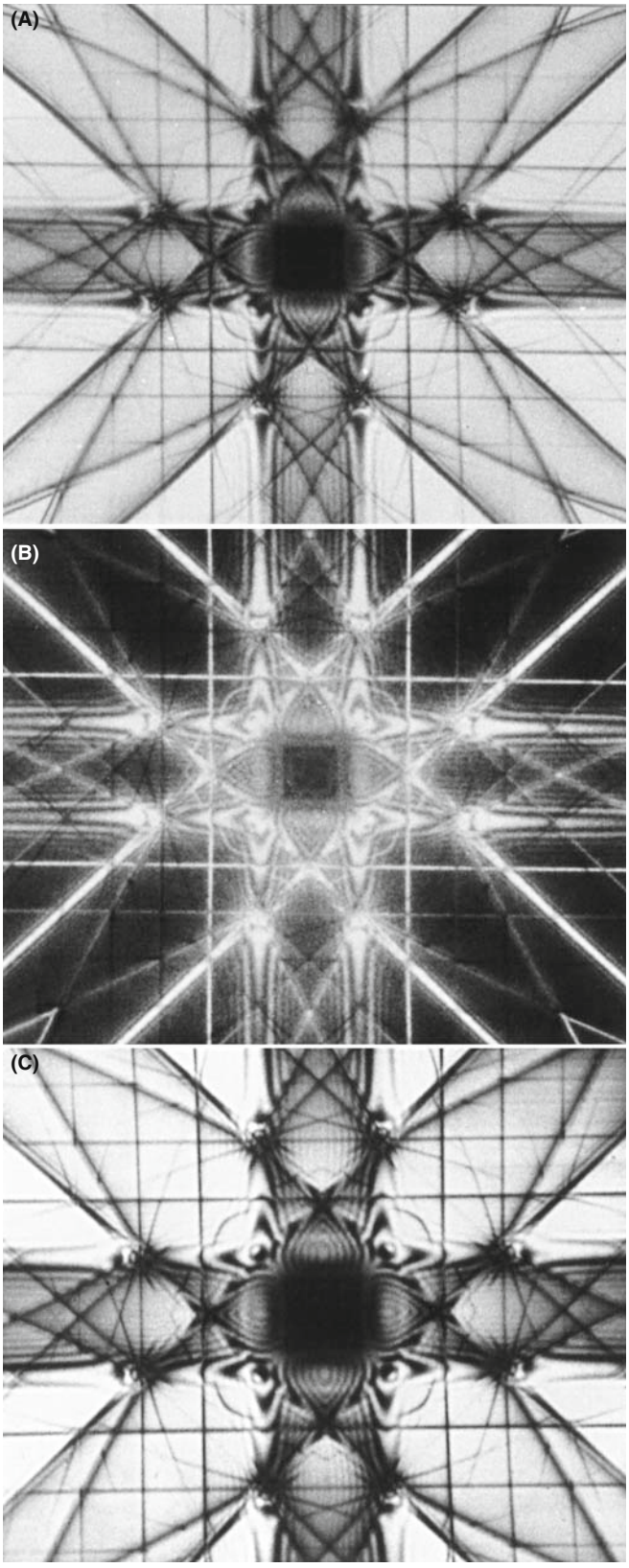


FIGURE 21.16 Series of Eades double-rocking ZAPs obtained from a thin Al foil in the [001] orientation. (A) is the BF image, (B) is the 200 DF image, and (C) is an energy-filtered version of (A). The removal of energy-loss electrons sharpens the image.

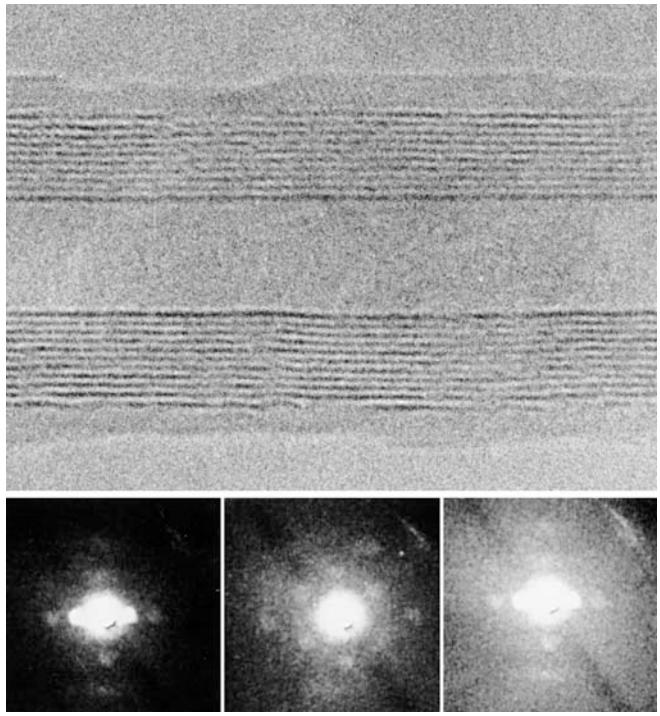


FIGURE 21.17 Nanodiffraction patterns from a multi-walled carbon nanotube (top).

several ways to do this and advances in computer simulation and CCD cameras make the direct, quantitative interpretation of diffracted intensity much more tractable, but the essential steps were all described by Spence and Zuo. Basically, we match computer simulations of intensity variations in the CBED disks with the experimental patterns, assigning specific values of F_{hkl} to each of the reflections until the best fit is obtained. So it is computer-intensive, but that is not a problem today.

Once the structure-factor data are obtained, a very useful side effect is that the valence-electron density can be measured from the low-order structure factors and this characteristic is otherwise inaccessible experimentally for crystals without an inversion symmetry (Spackman et al.). Once measured, the valence-electron density can be mapped out and the culmination of much of this work was the first-ever direct imaging of an atomic bond (Zuo et al.) (Figure 21.15).

21.8 OTHER METHODS

21.8.A Scanning Methods

Diffraction is as useful in scanning-beam instruments as in TEM. As we discussed back in Chapter 18, electron-backscatter diffraction (EBSD) studies are as ubiquitous today as SAD and CBED and very complementary. Many more crystals can be studied by EBSD and the specimen does not need to be thinned. EBSD patterns contain HOLZ lines just like CBED patterns.

Historically, electron channeling was another way to obtain crystallographic information in the SEM but the patterns are of such low contrast that EBSD has rendered the channeling technique obsolete (for now).

Scanning DPs can be obtained in STEMs using either one or two sets of coils both before and after the specimen. In both cases the beam is stationary at the plane of the specimen and rocks back and forth in a manner similar to the hollow-cone or precession methods we described in Chapters 18 and 20. Using only one scan coil below the specimen partially 'de-rocks' the beam but two coils fully 'de-rock' the beam such that, instead of the DP scanning across the BF STEM detector and being recorded sequentially, the 000 disk is always on the optic axis and thus on the BF detector. However, the HOLZ lines, etc., move continuously across the detector because, like Kikuchi lines, they are 'fixed' to the specimen. The resulting double-rocking patterns, which are often named after their inventor, Alwyn Eades, cover several degrees in comparison with the fractions of a degree visible in a normal CBED disk. These patterns are very striking as you can appreciate from Figure 21.16; such patterns can be viewed in BF or DF and, as with all CBED patterns, can be sharpened by energy filtering. These scanning methods can be used to study the occurrence of forbidden reflections which are important in crystal-symmetry determinations and more detail is given in the companion text.

21.8.B Nanodiffraction

Nanodiffraction, as the term implies, involves diffraction with a nanometer-scale beam and has been pioneered by Cowley. Nanometer beams have only a few picoamps of current so the patterns are noisy and require a FEG-TEM (or DSTEM), which also provides a coherent beam. If you scan the literature, you'll find that nanodiffraction permits determination of local variations of crystallography within large individual unit cells, dislocation cores, grain-boundary segregant films, individual nanotubes or nanoparticles such as catalysts. Figure 21.17 shows nanodiffraction patterns from a single, multi-walled, carbon nanotube.

In addition, it is possible to assess crystallographic effects associated with medium-range ordering by correlating atom positions over distance of a few nanometers. Nanodiffraction methods have helped develop the new field of fluctuation microscopy which we mention again in Chapter 29; they are also essential for diffractive imaging with atomic resolution.

CHAPTER SUMMARY

CBED patterns contain contrast information which gives a complete crystallographic characterization of small crystals, as well as other information such as thickness. To make the most of this capability you need detailed knowledge of crystal-symmetry concepts, stereographic projections, and the ability to produce ZAPs from a variety of orientations. In addition, the determined operator is often rewarded with patterns that are both very useful and stunningly beautiful. Before you go and read about CBED in the companion text, remember that, if you want to determine the point group and/or the space group from your CBED patterns, you need to determine

- WP symmetry
- BF symmetry
- Projection-diffraction symmetry

CBED symmetry information from the HOLZ-line distribution in the 000 disk is also very useful for measuring other characteristics of our specimens which cannot be determined, on a nanoscale, by other techniques. The examples we showed were lattice strain, an essential variable in the electronics and optical industries, enantiomorphism, structure-factor and charge-density determinations. If you've successfully worked your way through these two chapters, then you're ready for the companion text and the more challenging aspects of CBED, such as measurement of defect crystallography, structure factors, Debye-Waller factors, polarity, and charge density which are discussed in depth in the essential texts by Spence and Zuo and by Morniroli.

THE JEOL CBED ATLAS

By Tanaka, M et al, 1985-2002 *Convergent Beam Electron Diffraction* I-IV JEOL Tokyo. The details are in Chapter 20.

TECHNIQUE

Eades, JA 1989 (Ed.) Journal of Electron Microscopy Techniques **13** (I & II). Special issues on CBED with many useful papers covering much of the content of this chapter.

Jones DN 2007 *Sensitive Detection of Mirror Symmetry by CBED Applied to $LaAlO_3$ and $GdAlO_3$* Acta Cryst. **B63** 69–74. Using off-axis CBED patterns.

Mansfield, JF 1984 *Convergent Beam Electron Diffraction of Alloy Phases* Adam Hilger Ltd. Bristol UK. A great collection of CBED patterns and proof that steel can still be beautiful.

Morniroli, J-P 2002 *Large-Angle Convergent-Beam Electron Diffraction (LACBED)* SFμ. (Société Française des Microcopies) Paris, France. The essential instructional text full of great diagrams and practical examples along with an excellent bibliography

Spence, JCH and Zuo, JM 1992 *Convergent Beam Electron Diffraction* Plenum Press New York. You will want a copy of this book if you use CBED extensively. You may find a reference to earlier CBED studies on your material here in the selective bibliography organized by material. Although now considered obsolete by some, the original code for plotting HOLZ lines is included in the appendices along with the Fortran code for one Bloch wave and one multislice program.

THICKNESS DETERMINATION

Allen, SM 1981 *Foil Thickness Measurements from Convergent-Beam Diffraction Patterns* Phil. Mag. **A43** 325–335.

Kelly, PM, Jostons, A, Blake, RG and Napier, JG 1975 *The Determination of Foil Thickness by Scanning Transmission Electron Microscopy* Phys. Stat Sol. **A31** 771–780.

APPLICATIONS

Ayer, R 1989 *Determination of Unit Cell* J. Electron Microsc. Tech. **13** 16–26.

Berta, Y, Ma, C and Wang, ZL 2002 *Measuring the Aspect Ratios of ZnO Nanobelts* Micron **33** 687–691.

Cowley, JM 2001 *Electron Nanodiffraction Methods for Measuring Medium-Range Order* Ultramicrosc. **90** 197–206. Nanodiffraction.

Kim, M, Zuo, JM and Park, G-S 2004 *High-Resolution Strain Measurement in Shallow Trench Isolation Structures Using Dynamic Electron Diffraction* App. Phys. Lett. **84** 2181–2183.

Mansfield, JF 1985 *Error Bars in CBED Symmetry?* Ultramicrosc. **18** 91–96. As it says.

Raghavan, M, Scanlon, JC and Steeds, JW 1984 *Use of Reciprocal Lattice Layer Spacing in Convergent Beam Electron Diffraction Analysis* Metall. Trans. **15A** 1299–1302.

Randle, V, Barker, I and Ralph, B 1989 *Measurement of Lattice Parameter and Strain Using Convergent Beam Electron Diffraction* J. Electron Microsc. Tech. **13** 51–65. Accuracy in using HOLZ lines for determining lattice parameters.

Steeds, JW 1979 *Convergent Beam Electron Diffraction* in *Introduction to Analytical Electron Microscopy* 387–422 Eds. JJ Hren, JI Goldstein and DC Joy Plenum Press, New York.

CRYSTALLOGRAPHY

Eades, JA, Moore, S, Pfullman, T and Hangas, J 1993 *Discrepancies in Kinematic Calculations of HOLZ Lines* Microscopy Research and Technique **24** 509–513. The importance of dynamical scattering for HOLZ line calculations.

Inui, H, Fujii, A, Sakamoto, H, Fujio, S and Tanaka, K 2007 *Enantiomorph Identification of Crystals Belonging to the Point Groups 321 and 312 by Convergent-Beam Electron Diffraction* J. Appl. Cryst. **40** 241–249.

Jackson, AG 1987 *Prediction of HOLZ Pattern Shifts in Convergent Beam Diffraction* J. Electron Microsc. Tech. **5** 373–377. Collected data.

Jackson, AG 1990, *Identification of the Laue Zone Number in HCP Systems in Convergent Beam Electron Diffraction* Ultramicrosc. **32** 181–182.

Nye, JF 1985 *Physical Properties of Crystals* (2nd Ed.) Oxford University Press New York.

INDEXING AND LATTICE PARAMETERS

Ecob, RC, Shaw, MP, Porter, AJ and Ralph, B 1981 *Application of Convergent-Beam Electron Diffraction to the Detection of Small Symmetry Changes Accompanying Phase Transformations – I. General and Methods* Phil. Mag. **A44** 1117–1133. Early HOLZ/FOLZ.

Fournier, D, L'Esperance, G, Saint-Jacques, RG 1989 *Systematic Procedure for Indexing HOLZ Lines in Convergent Beam Electron Diffraction Patterns of Cubic Crystal* J. Electr. Microsc. Tech. **13** 123–149.

Morawiec, A 2007 *A Program for Refinement of Lattice Parameters Based on Multiple Convergent-Beam Electron Diffraction Patterns* J. Appl. Cryst. **40** 618–622.

Rozeveld, SJ and Howe, JM 1993 *Determination of Multiple Lattice Parameters from Convergent-Beam Electron Diffraction Patterns* Ultramicrosc. **50** 41–56.

BONDS

Spackman, MA, Jiang, B, Groy, TL, He, H, Whitten, AE and Spence, JCH 2005 *Phase Measurement for Accurate Mapping of Chemical Bonds in Acentric Space Groups* Phys. Rev. Lett. **95** 085502–05. Inversion symmetry.

Zuo, JM, Kim, M, O'Keeffe, M and Spence, JCH 1999 *Direct Observation of d-Orbital Holes and Cu-Cu Bonding in Cu_2O* Nature **401** 49–52. 'Imaging' atomic orbitals

URLs

1) http://cimewww.epfl.ch/people/stadelmann/jemsWebSite/DiffractionPatterns.html (Stadelmann's EM Yellow Pages at EPF Lausanne)
2) www.gatan.com/imaging/dig_micrograph.php (Gatan's software suite)
3) www.soft-imaging.net (Olympus's Soft Imaging System)
4) http://emaps.mrl.uiuc.edu (Maintained by Zuo at UIUC)
5) http://cimesg1.epfl.ch/CIOL (Stadelmann's JEMS for students) and http://cimewww.epfl.ch/people/stadelmann/jemswebsite/jems.html, the main site
6) www.amc.anl.gov (Zaluzec's Argonne National Laboratory site)
7) www.public.asu.edu/~jspence/ElectrnDiffn.html (the IUCr site)

SELF-ASSESSMENT QUESTIONS

Q21.1 \Why would you want to measure your specimen thickness with CBED?
Q21.2 Why do you need a foil of a certain thickness before you see K-M fringes?
Q21.3 Why do the K-M fringes increase in number as the specimen gets thicker?
Q21.4 List the experimental factors that can limit the accuracy of your thickness determination.
Q21.5 How can you create a continuous ring of HOLZ intensity rather than a distribution of HOLZ disks? Why would you want to do this, thus losing the distribution of individual disks in the pattern?
Q21.6 Why might you have to record several CBED patterns to get all the useful information?
Q21.7 Why is the HOLZ-ring radius important and what microscope parameters should you use to ensure that you can easily see the ring(s)?
Q21.8 Distinguish A-, B-, C-, and I-lattice centering.
Q21.9 What is Friedel's law and why is it important when trying to infer symmetry from CEBD patterns?
Q21.10 Distinguish the BF symmetry and the whole-pattern symmetry.
Q21.11 Distinguish diffraction groups and point groups.
Q21.12 What distinguishes space groups from point groups?
Q21.13 Define WP and CBIM.
Q21.14 How can compositional changes manifest themselves in CBED patterns?
Q21.15 Why would lattice strain effects appear in HOLZ-line patterns?
Q21.16 What is enantiomorphism and why would this property affect the symmetry of DPs?
Q21.17 Why might you want to form a DP while the beam is scanning rather than stationary?
Q21.18 Why might it be useful to see both image and diffraction information in the same picture at the same time, as in CBIM?
Q21.19 Why is the nanodiffraction pattern in Figure 21.17 noisy?
Q21.20 List the main applications of CBED that cannot be performed with standard SAD techniques.

TEXT-SPECIFIC QUESTIONS

T21.1 We get parallel fringes in CBED patterns under two-beam conditions and also in TEM images under two-beam conditions. Explain why this is so and why the symmetry of the fringes changes in the 000 and the *hkl* disk in a similar manner to changes in BF and DF images. (Hint: go and look in the imaging chapters.)
T21.2 Reproduce the data in Tables 21.1 and 21.2 from the information given in the chapter.
T21.3 Given that the lattice parameter of stainless steel is 0.405 nm, work out the lattice spacing in the [111] direction from the HOLZ ring in Figure 20.16B.
T21.4 Draw schematic diagrams like in Figure 21.12 to show threefold (triad) and 6*mm* symmetry.
T21.5 Go back and look at Figures 20.4A and 20.16C and state the BF symmetry in each case.
T21.6 Why is the CBED technique seeing increasing usage for crystal structure determination, displacing some of the more conventional XRD techniques?
T21.7 Why are there diffraction spots in Figure 21.17 when there are no visible crystal planes in the center of the image from where the pattern is taken?
T21.8 Why does energy-filtering sharpen the DP in Figure 21.16C? (Hint: look ahead to the EELS chapters.)
T21.9 If lattice strain disturbs the crystal plane spacing in three dimensions, how is it possible to deduce this strain from 2D HOLZ-line distributions?

Part 3

Imaging

22

Amplitude Contrast

CHAPTER PREVIEW

We've already mentioned in Chapters 2–4 that TEM image contrast arises because of the scattering of the incident beam by the specimen. The electron wave can change both its amplitude and its phase as it traverses the specimen and both types of change can give rise to image contrast. Thus a fundamental distinction we make in the TEM is between *amplitude contrast* and *phase contrast*. In most situations, both types of contrast actually contribute to an image, although we usually select conditions so that one will tend to dominate. In this chapter, we'll discuss only amplitude contrast and we'll see that there are two principal types, namely, *mass-thickness contrast* and *diffraction contrast*. This kind of contrast is observed in both TEM and STEM and in both BF and DF images. We'll discuss the important differences between the images formed in each of these two modes of operation. Then we'll go on to discuss the principles of diffraction contrast, which are sufficiently complex that it takes Chapters 24–27 to show you how this form of contrast is used to identify and distinguish different crystal defects. Diffraction-contrast imaging came into prominence in about 1956 when it was realized that the intensity in a diffracted beam depends strongly on the deviation parameter, **s**, and that crystal defects rotate the diffracting planes near the defect. Therefore, the diffraction contrast from regions close to the defect will depend on the properties (in particular, the strain field) of the defect. We'll then consider phase contrast and how it can be used to image atomic-level detail in Chapters 23 and 28. Other forms of TEM imaging and variations on these major types of contrast are gathered in the catch-all Chapter 29.

22.1 WHAT IS CONTRAST?

Before we start to describe specific types of contrast it's worth a quick reminder of what exactly we mean by the word 'contrast.' We can define contrast (C) quantitatively in terms of the *difference* in intensity (ΔI) between two adjacent areas

$$C = \frac{(I_2 - I_1)}{I_1} = \frac{\Delta I}{I_1} \qquad (22.1)$$

In practice your eyes can't detect intensity changes $<5\%$ and even $<10\%$ is difficult. So unless the contrast from your specimen exceeds >5–10% you won't see anything on the screen or on the recorded image. However, if your image is digitally recorded, you can enhance low-contrast images electronically to levels at which your eyes can perceive the contrast. We'll return to image processing and contrast enhancement in Chapter 31.

So we see contrast in TEM images as different levels of green light coming from the viewing screen or the computer display. On the photograph or computer screen, contrast is seen as different gray levels and our eyes can only discern about 16 of these. If we want to quantify the contrast, we need to make direct intensity measurements, e.g., via a microdensitometer if we're using film or directly from the CCD, but usually it's only necessary to see qualitative differences in intensity. Be careful not to confuse *intensity* with *contrast* when you describe your images. We can have strong or weak contrast but not bright or dark contrast. The terms 'bright' and 'dark' refer to density (number/unit area) of electrons hitting the screen or detector, and the subsequent light emission that we see. In fact, you generally get the strongest contrast under illumination conditions that lower the overall intensity. Conversely, you can try to increase the number of electrons falling on the screen by condensing the beam onto a reduced area of the specimen, but then you'll usually lower the image contrast. These points are summarized in Figure 22.1, which defines intensity and contrast.

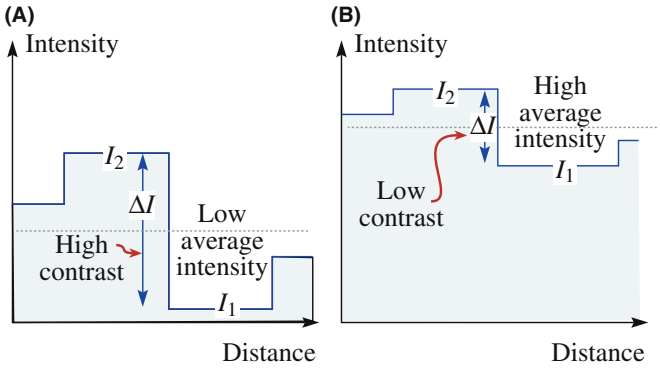


FIGURE 22.1. Schematic intensity profiles across an image showing (A) different intensity levels (I_1 and I_2) and the difference (ΔI) between them which defines the contrast. Generally, in a TEM, if the overall intensity is increased (B) the contrast decreases.

Before we discuss the two forms of amplitude contrast in detail, we need to remind you of the operational principles for creating amplitude contrast in your image. We obtain contrast in our images either by selecting specific electrons or excluding them from the imaging system. We have two choices: we can form either BF or DF images by selecting the direct or scattered electrons, respectively. So this chapter builds on what you learned about electron scattering in Chapters 2–4 and how to operate the TEM, described in Chapter 9.

22.2 AMPLITUDE CONTRAST

Amplitude contrast results from variations in mass or thickness or a combination of the two: the thickness variation can produce contrast because the electron interacts with more material (hence, more mass). Alternatively, diffraction can vary locally because the specimen is not a perfect, uniformly thin sheet.

22.2.A Images and Diffraction Patterns

If you look back at Figure 2.1, you'll see that the uniform electron intensity in the incident beam is transformed into a non-uniform intensity after scattering by the specimen. So a variable electron intensity hits the viewing screen or the electron detector, which translates into contrast on the screen. Now you also know that the DP shows you this non-uniformity because it separates the diffracted and direct beams. Therefore, a fundamental principle of imaging in the TEM is: first, *view the DP*, since this pattern tells you how your specimen is scattering. The relationship between the image and the DP is most critical for crystalline specimens showing diffraction contrast. However, you should view the DP first, no matter which contrast mechanism you want to use, or the specimen you are studying.

22.2.B Use of the Objective Aperture or the STEM Detector: BF and DF Images

In order to translate the electron scatter into interpretable amplitude contrast, we select *either* the direct beam or a diffracted beam in the SADP to form BF and DF images, respectively. (Remember that a small aperture will enhance the contrast but might decrease the resolution.) Note we are justified in using the 'beam' terminology, since the electrons have left the specimen. We've already seen in Section 9.3 that in a TEM we select the direct or a scattered electron beam with the objective aperture. Remember, if we form an image without the aperture, the contrast will be poorer (lower) because many beams then contribute to the image so we lose diffraction contrast. Furthermore, aberrations due to the off-axis electrons will make it impossible to focus your image. Your choice of the aperture size governs which electrons contribute to the image and you thus control the contrast.

Figure 22.2 shows a DP from a single-crystal Al specimen with two possible positions of an objective aperture indicated by white circles. In this figure, if the aperture is in position A it selects the direct beam only and thus a BF image will be formed in the image plane of the lens. This arrangement will produce amplitude contrast whether the specimen is crystalline (as in this case) or amorphous. If the aperture is in position B, it will select only electrons scattered in that specific direction. Thus a DF image will be formed. Traditionally (for 50 years), we have tilted the incident beam such that the scattered electrons remain on axis when forming a DF image. We thus create a centered dark-field (CDF) image, which we described back in Section 9.3. We'll discuss CDF techniques later in Section 22.5 and we'll usually assume CDF is the operational mode in DF imaging. However, you already know that it would be much better if we did not tilt the specimen but instead displaced the aperture (so the BF and DF are formed with the same diffraction conditions). We'll see later

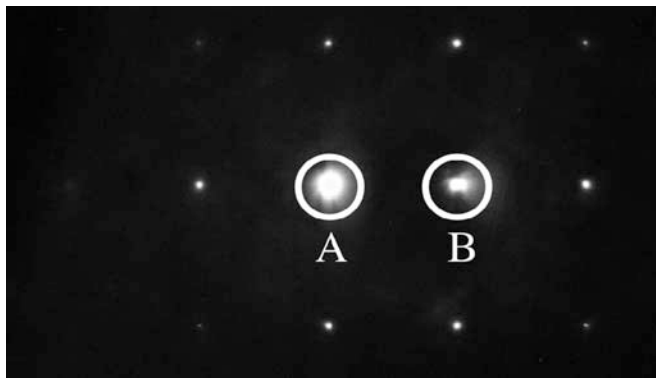


FIGURE 22.2. The relationship between the objective aperture and the diffraction pattern for forming (A) BF and (B) DF images. The circles show the location of the objective aperture.

that if you want to observe phase contrast, you have to use an objective aperture that is large enough to gather more than one beam.

BF AND DF
The two basic ways to form amplitude-contrast images.

In a STEM we select the direct or scattered beams in an equivalent way but use detectors rather than apertures. We compare the two different operational modes in Figure 22.3. Again, we saw back in Section 9.4 that we insert a BF on-axis detector, or an annular DF (ADF) detector, in a plane that's conjugate with the back focal plane. We control which electrons fall on which detector and thus contribute to the image by adjusting the post-specimen (imaging) lenses to change the camera length. Clearly, for DF imaging, the ADF detector will probably gather many more electrons than the objective aperture, which is good for imaging some specimens and bad for imaging others, as we'll see.

So, in summary, we can create BF or DF images with the direct beam or scattered beams, respectively. In order to understand and control the contrast in these images you need to know what features of a specimen cause scattering and what aspects of TEM operation affect the contrast.

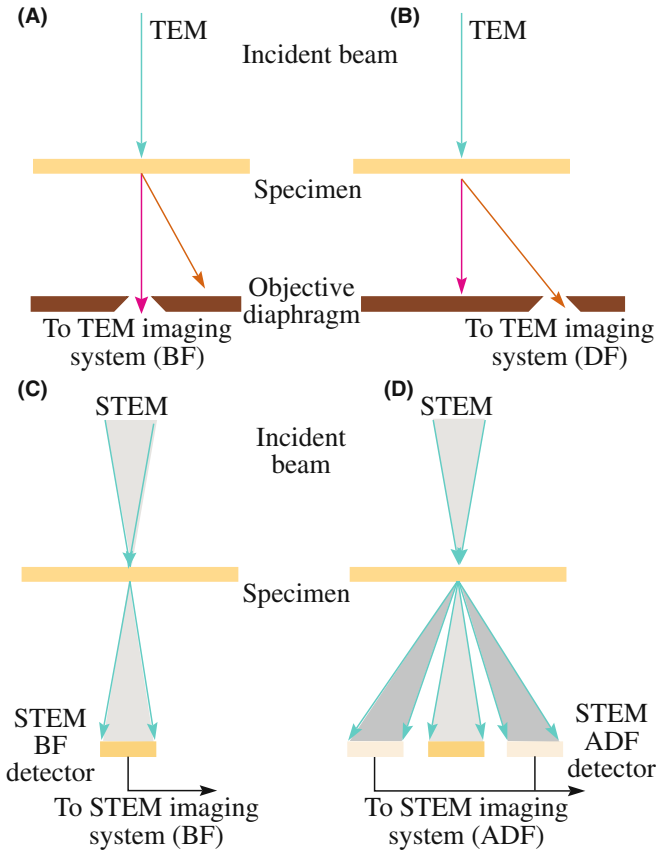


FIGURE 22.3. Comparison of the use of an objective aperture in TEM to select (A) the direct or (B) the scattered electrons forming BF and DF images, respectively. In STEM we use (C) an on-axis detector or (D) an off-axis annular detector to perform equivalent operations.

22.3 MASS-THICKNESS CONTRAST

Mass-thickness contrast arises from incoherent elastic scattering (Rutherford scattering) of electrons. As we saw back in Chapter 3, the cross section for Rutherford scatter is a strong function of the atomic number Z (hence the mass or the density, ρ) and the thickness, t, of the specimen. Rutherford scattering in thin specimens is strongly forward peaked. Therefore, if we form an image with electrons scattered at low angles ($<\sim 5^\circ$), mass-thickness contrast dominates (but it also competes with Bragg-diffraction contrast). However, we'll also see that at high angles ($>5^\circ$), where Bragg scattering is usually negligible, we can pick up low-intensity, scattered beams. (The former is often referred to as coherent and the latter as incoherent; we'll discuss these terms later.) The intensity of these beams depends on atomic number (Z) only. Thus we can also get so-called Z-contrast, which contains elemental information like that in BSE images in the SEM. We can obtain these images with atomic resolution, particularly in a DSTEM. It is also feasible to form BSE images in a TEM but, because the specimen is thin, the number of BSEs is so small that the images are noisy and of poor quality, so no one does it. You shouldn't waste your money buying a BSE detector (but a secondary-electron detector can be very useful).

Mass-thickness contrast is most important if you are looking at non-crystalline materials, such as polymers and it is *the* critical contrast mechanism for biological scientists. But as we'll see, any variations in mass and thickness will cause contrast. As you learned in Chapter 10, it's almost impossible to thin a bulk sample uniformly (although a FIB can get close) and so nearly all real specimens will show some mass-thickness contrast. In some cases this will be the only contrast you can see.

In this section, we'll assume that there is no contribution to the image from diffraction contrast. This is automatically so if the specimen is amorphous. If the specimen is crystalline, then remove the objective aperture or use the ADF detector to minimize any diffraction contrast. You should still use an objective aperture to enhance the mass-thickness contrast to minimize the effects of lens aberrations. You'll still create BF and DF images of amorphous materials.

22.3.A Mechanism of Mass-Thickness Contrast

The mechanism by which differences in mass and thickness cause contrast is shown in Figure 22.4 and at this

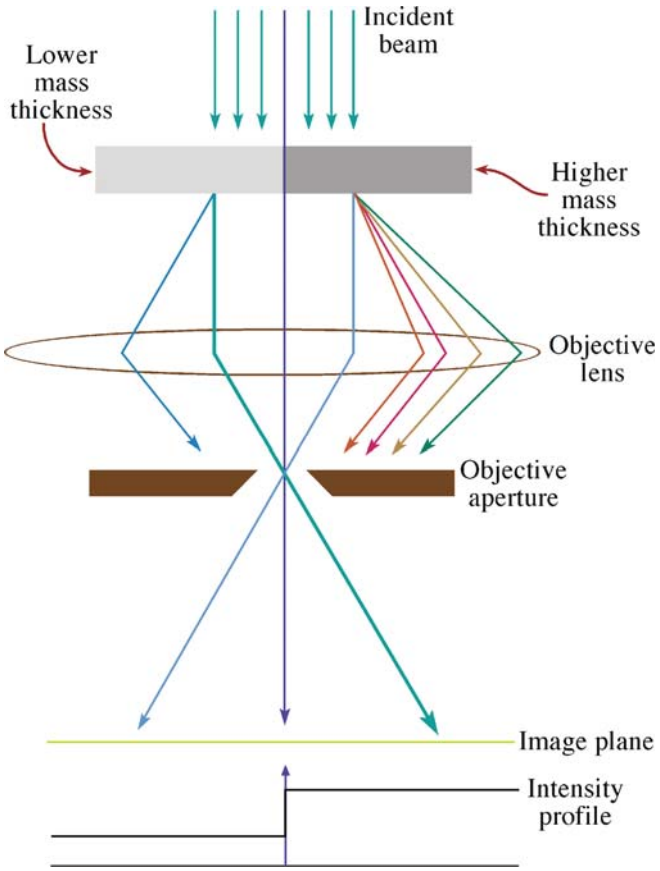


FIGURE 22.4. Mechanism of mass-thickness contrast in a BF image. Thicker or higher-Z areas of the specimen (darker) will scatter more electrons off axis than thinner, lower mass (lighter) areas. Thus fewer electrons from the darker region fall on the equivalent area of the image plane (and subsequently the screen), which therefore appears darker in BF images.

stage we'll talk about the process qualitatively. As electrons go through the specimen, they are scattered off axis by elastic nuclear interactions, i.e., Rutherford scattering. You know two factors from Chapter 3

- The cross section for elastic scattering is a function of Z.
- As the thickness of the specimen increases, there will be more elastic scattering because the mean free path remains fixed.

So, using a very simple, qualitative argument you would expect high-Z (i.e., high-mass) regions of a specimen to scatter more electrons than low-Z regions of the same thickness. Similarly, thicker regions will scatter more electrons than thinner regions of the same average Z, all other factors being constant. Usually, mass-thickness contrast images are interpreted in such a purely qualitative fashion, although we'll see a little later that it is possible to quantify the scattering intensity. So, as you can see from Figure 22.4, for the case of a BF image, thicker and/or higher-mass areas will appear darker than thinner and/or lower-mass areas. The reverse will be true for a DF image.

This is all you need to know for the simplest interpretation of mass-thickness contrast images. Sometimes mass-thickness contrast is explained in terms of different amounts of electron absorption within the specimen and so you may come across the expression 'absorption contrast.' We think that this term is misleading, because in thin foils the actual amount of electron absorption is small; scattering outside the aperture or the detector, not absorption within the specimen, causes the contrast. For much the same reason, we prefer not to use the term 'structure-factor contrast' that is sometimes used to describe this phenomenon, since this implies a Bragg contribution, which may or may not be present.

However, you should be aware that if there are small crystals of different atoms in a given foil thickness, differences in their structure factor (F) from that of the matrix will cause contrast changes, since $I \propto |F|^2$. For example, you can detect the presence of nanometer-size clusters of Ag atoms in very thin foils of Al alloys in this way. Conversely, an absence of atoms (e.g., a void) will also scatter differently, although Fresnel contrast (see Chapter 23) is a better way to detect voids and bubbles.

Let's first look at a few images showing mass-thickness contrast and see which TEM variables you can control.

22.3.B TEM Images

Figure 22.5A is a TEM BF image of some latex particles on an amorphous-carbon support film. Assuming the latex is predominantly carbon, we have a constant Z and varying t. So the latex particles are darker than the support film since they are thicker. What you are basically seeing is a shadow projection image of the latex particles. Because it is a projection image, you cannot say that the particles are spheres (which in fact they are). They could equally well be disks or cylinders. One way to reveal the shape in a single image is to shadow-coat the particles, i.e., evaporate a thin heavy metal (Au or Au-Pd) coating at an oblique angle as shown in Figure 22.5B. The shape of the shadow reveals the true shape of the particles.

Shadowing introduces some mass contrast to what was just a thickness-contrast image. If we assume the Au-Pd film is very thin compared to the carbon support film, then the contrast across the edge of the shadow is predominantly mass contrast, due to the difference in average Z of the Au-Pd and the carbon film. If the spheres are small, there may also be an intensity change across the latex spheres due to the preferential deposit of Au-Pd on the side of the sphere towards the source of the evaporated metal.

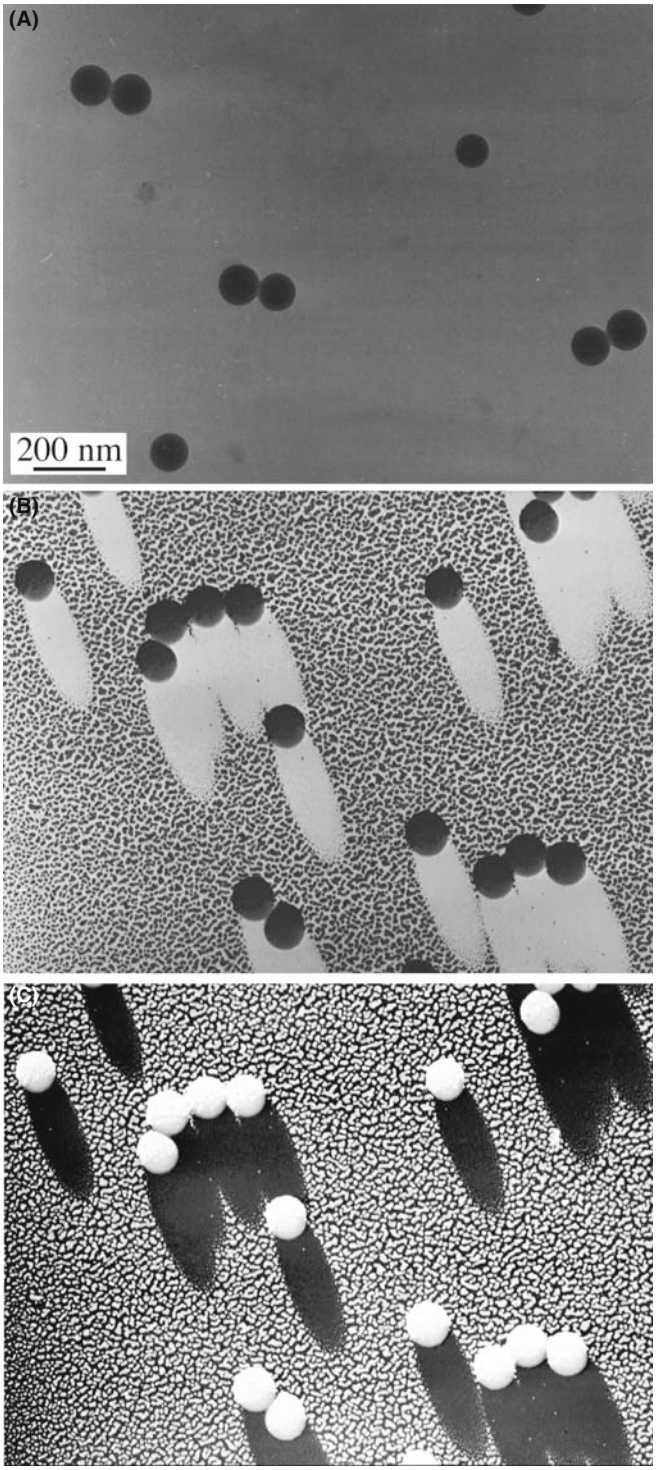


FIGURE 22.5. (A) TEM BF image of latex particles on a carbon support film showing thickness contrast only. (B) Latex particles on a carbon film shadowed to reveal the shape of the particles through the addition of selective mass contrast to the image. (C) Reverse print of (B) exhibits a 3D appearance.

It is an intriguing exercise to print Figure 22.5B in reverse (or take a DF image) as shown in Figure 22.5C. In this image, the latex spheres now appear to stand proud of the surface, even though you're still viewing a two-dimensional projected image. Because the shadows are now dark your brain interprets the picture as it would a reflected-light image and endows it with a 3D nature. While the interpretation in this case is correct, it may not always be so. Once again we stress that you must be careful when interpreting 2D images of 3D specimens.

In addition to the use of shadowing to enhance mass-thickness contrast, it is common practice to stain different areas of polymer and biological specimens with heavy metals such as Os, Pb, and U. The stain leaves the heavy metal in specific regions of the structure (e.g., at unsaturated C = C bonds in a polymer and cell walls in biological tissue) and therefore these areas appear darker in a BF image. Figure 22.6 shows a BF image of a stained two-phase polymer. Since the specimen is of constant thickness (it was ultramicrotomed) the image shows mass contrast only.

TEM VARIABLES

The TEM variables that affect the mass-thickness contrast for a given specimen are the objective aperture size and the kV.

If you select a larger aperture, you allow more scattered electrons to contribute to the BF image. So the contrast between scattering and non-scattering areas is lowered, although the overall image intensity increases. If you choose a lower kV, both the scattering angle and the cross section increase. Hence, more electrons will be scattered outside a given aperture, hitting the diaphragm, and contrast will increase at the expense of intensity. The decrease in intensity will be worse for TEMs with a thermionic source because the gun brightness decreases as the kV is lowered. Figure 22.7 shows how a smaller aperture size results in improved contrast.

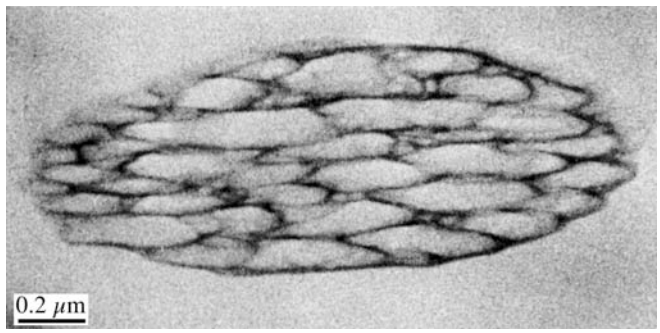


FIGURE 22.6. BF image of stained two-phase polymer exhibiting mass contrast due to the segregation of the heavy metal atoms to the unsaturated bonds in the darker phase.

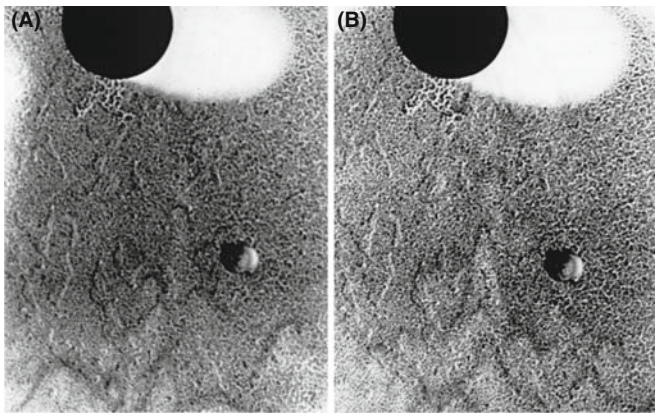


FIGURE 22.7. The effect of objective aperture size on mass-thickness contrast; the images of the shadowed latex particle were taken with an aperture size of (A) 70 μm and (B) 10 μm. A smaller aperture enhances the contrast, in a similar manner to lowering the kV.

Of course, any decrease in intensity can be offset by increased exposure times until specimen drift becomes a limiting factor.

Now for DF images, there isn't much more to be said: the images will generally show regions of contrast that are complementary to those seen in the BF (analogous to the reversed contrast in Figure 22.5B and C) but be prepared for exceptions (in Chapter 24). The overall intensity of the DF image will be much lower than the BF image (hence the relative terms 'dark' and 'bright') because the objective aperture will select only a small fraction of the scattered electrons. It's easy to remember that the BF image of a hole in your specimen will be bright and a DF image will be dark. However, remember that the corollary of low intensity is high contrast and DF images generally show better contrast.

22.3.C STEM Images

In a STEM you have more flexibility than in a TEM because by varying *L*, you change the collection angle of your detector and create, in effect, a variable objective aperture. (We'll mention the old TEMs later where you could do this and the SAD aperture was a triangle or a square.) So you have more control over which electrons contribute to the image. Even so, STEM BF images offer little more than TEM BF images. Generally, STEM images are noisier than TEM images (unless you've got an FEG STEM). Figure 22.8 shows a noisy STEM BF image of the same two-phase polymer as shown in the TEM image in Figure 22.6. The STEM images have generally shown poorer resolution because, with good thin specimens, the beam size dominates the resolution. To get reasonable intensity in a scanning image in reasonable time we have to use a large beam, as we discussed when we compared scanning and static images back in Chapter 9. Figure 22.9 shows the difference between (A) TEM and (B) STEM BF images from a low-contrast

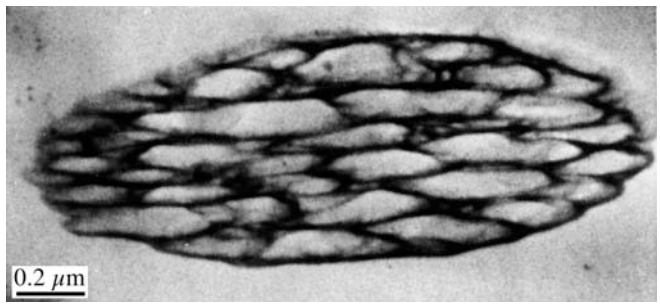


FIGURE 22.8. STEM BF image of a stained two-phase polymer. Comparison with the TEM image in Figure 22.6 shows that while the contrast is higher in STEM, the image quality is poorer.

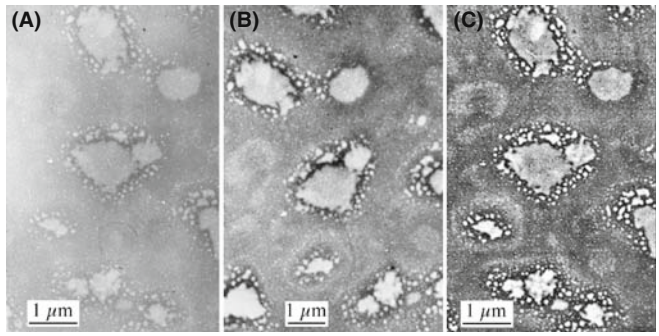


FIGURE 22.9. Comparison of TEM (A) and STEM (B) images of an amorphous SiO_2 specimen containing Cl-rich bubbles. The low mass contrast in the TEM can be enhanced in a STEM image through signal processing. (C) A similar effect can be achieved by digitizing the TEM image (A) and applying contrast-enhancement software.

specimen. The STEM image contrast has been enhanced and is considerably greater than in the TEM image, but the noise in the image is also more visible. However, if you record your TEM image using a CCD camera, or digitize the negative, you can enhance the contrast (see Chapter 31). A good way to do this is with one of the several kinds of image processing software (see Chapter 31) as you can see in Figure 22.9C.

STEM

STEM in a TEM is now a routine high-resolution technique.

In a STEM DF image, the scattered electrons fall onto the ADF detector. This gives rise to a fundamental difference between the TEM and STEM DF modes

- DF TEM images are usually formed by permitting only a fraction of the scattered electrons to enter the objective aperture.
- STEM images are formed by collecting most of the scattered electrons on the ADF detector.

CONTRAST IN STEM

Remember, you can always increase the contrast in the STEM image by adjusting the signal-processing controls, such as the detector gain and black level and the contrast and brightness controls on the computer screen; such options aren't available for analog TEM images.

Therefore, STEM ADF images are less noisy than TEM DF images, as shown in Figure 22.10. Because lenses aren't used to form the STEM image (although they are used to form the probe), the ADF images don't suffer aberrations, as would the equivalent off-axis TEM DF image.

STEM ADF image contrast is greater than TEM DF contrast: in STEM, L can be adjusted to maximize the ratio of the number of scattered electrons hitting the detector to the number of electrons going through the hole in the middle of the detector. You can thus improve the contrast quite easily, just by watching the computer screen and adjusting L.

However, as you can see from Figure 22.10, while the TEM DF image shows poorer contrast and is noisier, it still shows better resolution. Thermionic-source STEM images generally only show better resolution than TEM images when thick specimens are being imaged, because the chromatic aberration effects from thicker specimens do not affect the STEM images. If contrast is more important than resolution, then STEM is more useful. Indeed, in a STEM, you can study unstained polymer specimens which would show negligible contrast in a TEM.

DP AND STEM

The STEM must be well aligned so the DP expands and contracts on axis.

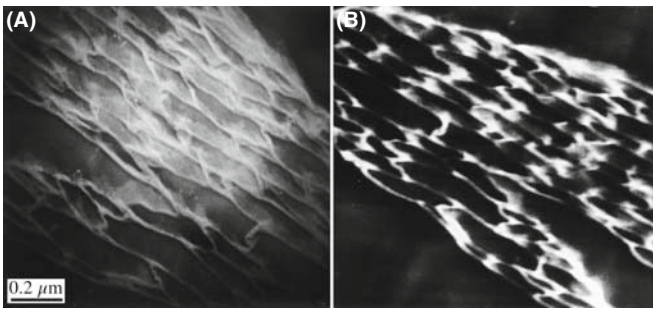


FIGURE 22.10. Comparison of (A) TEM DF and (B) STEM ADF images of the same two-phase polymer as in Figures 22.6 and 22.8. As in BF the STEM image shows higher contrast but lower resolution. Also, the ADF aperture collects more signal than the TEM objective aperture so the STEM image is less noisy.

STEM imaging is also useful if your specimen is beam sensitive, e.g., some polymers. A scanning beam lets you precisely control the irradiated region of the specimen, so it's a form of low-dose microscopy (see Section 4.6). You'll lose some image resolution unless you have access to a FEG STEM.

The comparison we've made of TEM and STEM images here is qualitative, but there have been many quantitative comparisons of STEM and TEM contrast, particularly for biological specimens. When STEMs were first introduced in the 1970s, the absence of chromatic aberration effects led to prophesies that STEM image resolution would invariably be better than TEM; there were even predictions of the end of classic TEM imaging! This hasn't happened because, as we'll see, there is more than just the chromatic aberration factor that governs the image quality, particularly for crystalline specimens. In summary, then, there are occasions when you might want to use STEM mass-thickness contrast images

- The specimen is so thick that chromatic aberration limits the TEM resolution.
- The specimen is beam-sensitive.
- The specimen has inherently low contrast in TEM and you can't digitize your TEM image or negative.
- Your specimen is ideally suited for HRTEM by Z-contrast imaging.

We'll examine the latter in the next section and in Chapter 28, and in much more detail in the companion text.

This said, be aware that STEM in the TEM has greatly improved since many of the comparison studies were carried out. Second, low-dose techniques in TEM will improve in the future. Third, you must be able to digitize your negatives now—negatives will be obsolete in the near future.

22.3.D Specimens Showing Mass-Thickness Contrast

Mass-thickness is the primary contrast source in amorphous materials, which is why we've illustrated this chapter mainly with polymer specimens. Replicas also display thickness contrast (see Figure 22.11A). Remember from Chapter 10 that replicas recreate the specimen topography, e.g., for a fracture surface. The amorphous-carbon replica can be as prepared (without shadowing) (Figure 22.11A) or shadowed (Figure 22.11B). The uneven metal shadowing increases the mass contrast and thus accentuates the topography, as for the latex particles in Figure 22.3. An extraction replica (Figure 22.11C) or a collection of particles dispersed on a support film will also show mass-thickness contrast; shadowing could be useful to reveal the shape of these particles too. If the particles are crystalline there will also be a component of diffraction contrast. Don't shadow if you want to do elemental analysis of the particles.

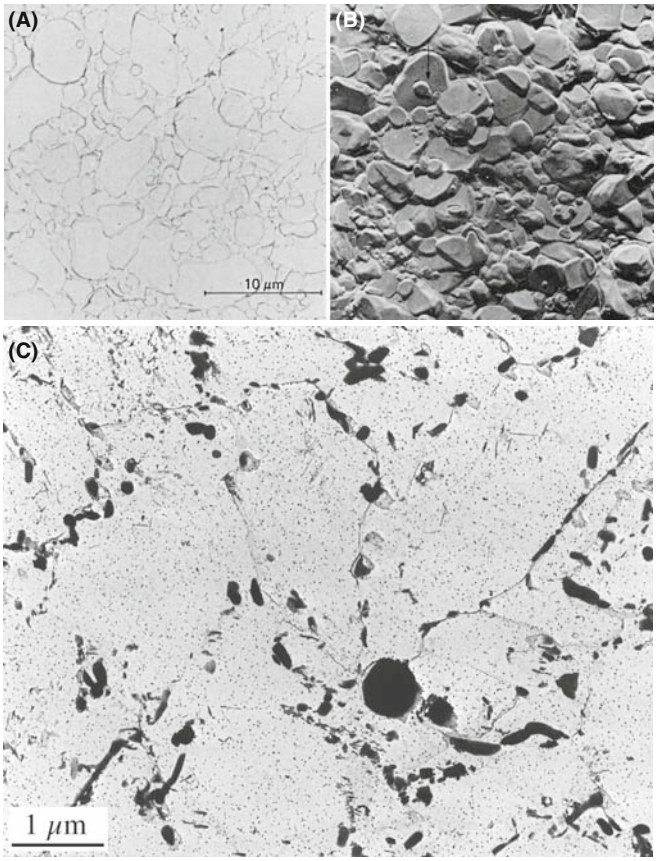


FIGURE 22.11. More examples of mass-thickness contrast: (A) a carbon replica of a fracture surface doesn't show much of either form of contrast until (B) oblique shadowing enhances the topography. (C) An extraction replica of small precipitate particles in a Cr–Mo steel weld shows both mass and thickness contrast.

22.3.E Quantitative Mass-Thickness Contrast

Because mass-thickness contrast is governed by Rutherford scattering, we can use the equations given back in Chapter 3 to predict the effect of Z and t on the scattering angle, θ, and the effect of kV on the cross section. We assume that the atoms scatter independently. (We then say that the scattering is truly incoherent.) This is actually not really the case, since even DPs from amorphous specimens show diffuse rings rather than uniform intensity (Figure 2.13A). Nevertheless, we'll still assume incoherent scattering.

As we stated at the start of this chapter, the contrast C is given by $\Delta I/I$ and it can be shown that a change in thickness, Δt, at constant atomic number Z creates contrast

$$\frac{\Delta I}{I} = 1 - e^{-Q\Delta t} \tag{22.2}$$

Q is the total elastic scattering cross section. So C becomes $Q\Delta t$ for $Q\Delta t < 1$. If 5% is the minimum contrast we can see, then the minimum Δt that we can see is

$$\Delta t \cong \frac{5}{100Q} = \frac{5A}{100\mathrm{N}_0\sigma\rho} \tag{22.3}$$

where A is the atomic weight, N_0 is Avogadro's number, σ is the single-atom scattering cross section, and ρ is the density.

A similar argument can be made if there is a ΔZ (in which case σ or ρ changes). So, if we want to calculate the contrast, we need to know σ. As we've seen in equation 3.8, for low-angle scattering, the differential Rutherford cross section is equal to $f(\theta)^2$ where $f(\theta)$ is the atomic scattering factor, given by equation 3.9

$$f(\theta) = \frac{\left(1 + \frac{E_0}{\mathrm{m}_0\mathrm{c}^2}\right)}{8\pi^2 a_0}\left(\frac{\lambda}{\sin\frac{\theta}{2}}\right)^2 (Z - f_x) \tag{22.4}$$

The Z term represents the Rutherford scattering. For unscreened Rutherford scattering (where we ignore the effects of the electron cloud), σ is therefore proportional to Z^2. This unscreened behavior is quite a good approximation for electrons scattered through angles (remember, we mean semiangles when we talk about scattering) above $\sim 5^\circ$ (e.g., for Cu) although it is dependent on E_0 and Z. At lower angles, scattering becomes increasingly screened (less dependent on Z), and more dominated by inelastic scattering and diffraction. There is no precise angle which we use to define the transition from low- to high-angle scatter but the effect of screening effectively disappears at angles $>\theta_0$, the screening parameter, defined back in equation 3.4.

We can use the atomic scattering factor (equation 22.4) to determine the probability that an electron will be scattered through greater than a given angle. To do this, we integrate $|f(\theta)|^2$ from an angle β (defined by the angle of collection of the objective aperture) to infinity. Thus

$$\sigma(\beta) = 2\pi \int_\beta^\infty |f(\theta)|^2 \theta d\theta \tag{22.5}$$

which can be evaluated to give

$$\sigma(\beta) = \frac{\left[Z\lambda\left(\frac{a_0}{Z^{0.33}}\right)\left(1 + \frac{E_0}{\mathrm{m}_0\mathrm{c}^2}\right)\right]^2}{\pi(a_0)^2\left(1 + \left(\frac{\beta}{\theta_0}\right)^2\right)} \tag{22.6}$$

where a_0 is the Bohr radius and θ_0 is the characteristic screening angle; all the other terms have their usual meaning (see Chapter 3). So in equation 22.6 you can see directly the effect of Z and kV on electron scatter and hence on contrast. As we've already described, higher-Z specimens scatter more while lowering E_0 increases scattering. The effect of thickness is deduced

from the mean free path for elastic scatter, λ (which is inversely proportional to σ). So, thicker specimens scatter more.

Let's assume that n_0 electrons are incident on the specimen and dn electrons are scattered through an angle $>\beta$. Then, from equation 22.6, ignoring any inelastic scattering (which isn't really reasonable, but we'll do it to simplify matters), the reduction in the number of electrons going through the objective aperture to form the BF image is given by

$$\frac{dn}{n} = -N\sigma(\beta)dx \tag{22.7}$$

where $N = N_0/A$ and N_0 is Avogadro's number, $\sigma(\beta)$ is given by equation 22.6, and $x = \rho t$. So this expression gives the dependence of the contrast on Z and t. If we integrate

$$\ln n = -N\sigma x + \ln n_0 \tag{22.8}$$

and if we rearrange this expression

$$n = n_0 e^{-N\sigma x} \tag{22.9}$$

which describes the exponential decrease in the number of scattered electrons (n) as the specimen mass-thickness ($x = \rho t$) increases.

As you'll have gathered, this equation is something of an approximation but it does give you a feel for the factors that control mass-thickness contrast. For a given specimen, the variables are local changes in Z and t; within the microscope, the variables are β and E_0, which you can control to change the contrast, as we saw in Figure 22.7.

In principle, you could use these equations and equation 22.1 to calculate the expected contrast arising from differences in Z or t and see if they were detectable at the 5% contrast level. In practice, however, image-contrast calculations are not carried out for simple mass-thickness contrast in materials specimens.

22.4 *Z*-CONTRAST

Z-contrast is the name given to a high-resolution (atomic), mass-thickness (Z), imaging technique. We'll talk about it here because it represents the limit of mass-thickness contrast where detectable scattering arises from single atoms or column of atoms.

Back in the 1970s, early FEG STEMs demonstrated the remarkable capability of imaging single heavy atoms (e.g., Pt and U) on low-Z substrates as shown in Figure 22.12. These images were formed by the ADF detector collecting low-angle elastically scattered electrons only. Single atoms scatter incoherently and the image intensity is the sum of the individual atomic scattering

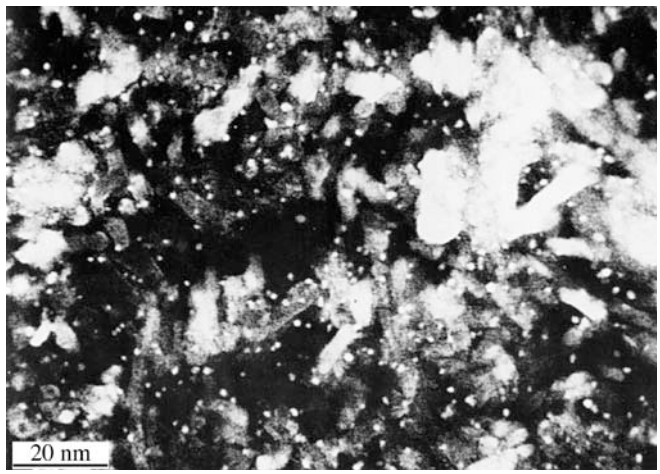


FIGURE 22.12. *Z*-contrast ADF image of individual Pt atoms or groups of atoms on a crystalline Al_2O_3 film obtained using an FEG STEM.

contributions. There was sometimes a problem with thickness changes in the substrate and contributions to the ADF signal from inelastically scattered electrons. This problem was overcome by dividing the digital ADF signal by the inelastic (energy-loss) signal from the EELS system. A drawback to this technique is that diffraction contrast (e.g., from a crystalline substrate) is preserved in the low-angle EELS signal, which can confuse the image interpretation. In Figure 22.12, the large bright regions arise from the Al_2O_3 substrate diffracting onto the ADF detector, obscuring the scatter from the Pt atoms.

HAADF

The detector is called a high-angle ADF or HAADF detector. *Z*-contrast images are also termed HAADF images. The outer diameter of the Fischione HAADF detector is 28 mm; the inner diameter is 4 mm.

Because of Bragg scattering, this early approach to *Z*-contrast was not suited to the study of crystalline specimens. Since the normal ADF detector will always collect some Bragg electrons, it was necessary to design an ADF detector with a very large central aperture. *Z*-contrast images could then be formed from thin crystals (Figure 22.13). You can decrease the camera length with the post-specimen lenses to ensure that the Bragg electrons (including any HOLZ scatter) don't hit the detector. Thus only the electrons scattered through very high angles contribute to the image.

Bragg effects are avoided if the HAADF detector only gathers electrons scattered through an angle of >50 mrad (~3°). Remember that cooling your specimen has the effect of increasing coherent HOLZ scatter, so don't cool it unless you must. Electron channeling effects

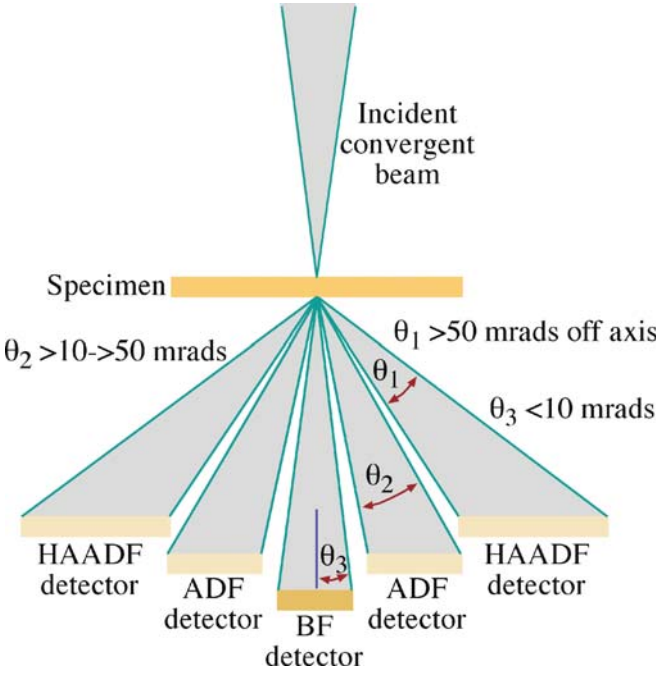


FIGURE 22.13. Schematic of the HAADF detector setup for *Z*-contrast imaging in a STEM. The conventional ADF and BF detectors are also shown along with the range of electron scattering angles gathered by each detector.

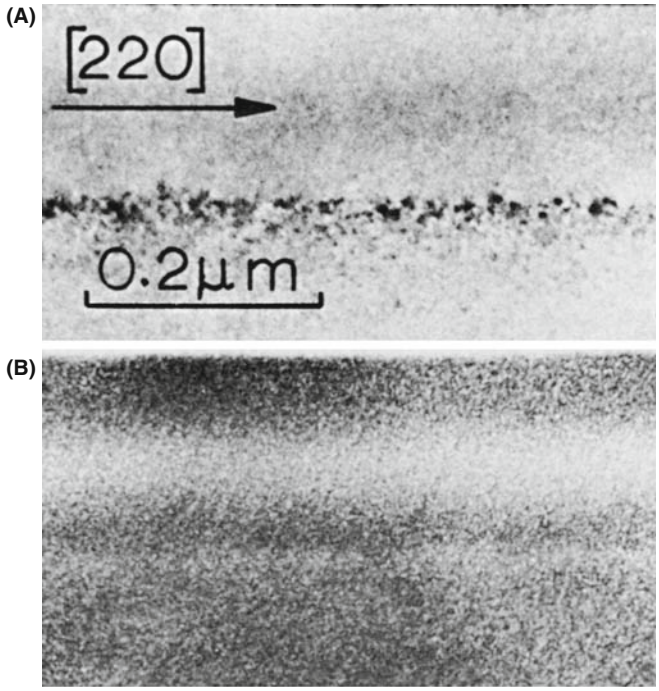


FIGURE 22.14. (A) Low-resolution TEM BF image showing a row of defects in Bi-implanted Si. In (B) obtained under *Z*-contrast conditions, the defects associated with the implant are invisible but the specimen is bright in the region implanted with Bi.

remain at high scattering angles, so imaging away from strong two-beam conditions and closer to zone-axis orientations is wise.

So, what do these *Z*-contrast images of crystals look like? Figure 22.14 shows a TEM BF image of Bi-implanted Si and below is a *Z*-contrast image. In the TEM BF image, formed from the direct beam, defects associated with the Bi implant are shown (we'll talk about such diffraction contrast from defects in Chapter 26) but otherwise there is no contrast associated with the Bi. In the *Z*-contrast image the Bi-implanted area is bright, but note that defect contrast isn't preserved in this image. You can relate the intensity differences in Figure 22.14 to an absolute measure of the Bi concentration. To do this you need to choose a suitable elastic scattering cross section. The contrast is related directly to the cross section for elastic scatter by the matrix (σ_A) and the alloying or dopant element (σ_B)

$$C = \left(\frac{\sigma_A}{\sigma_B} - F_B\right) c_B \qquad (22.10)$$

where c_B is the atomic concentration of the alloying element and F_B is the fraction of the alloying element that substitutes for matrix atoms. The intensity can be quantified to an absolute accuracy of better than $\pm 20\%$.

In an FEG with probe sizes of <0.3 nm, *Z*-contrast image resolution of close to the probe diameter is possible. Figure 22.15A shows a high-resolution phase-contrast TEM image of Ge on Si with an amorphous SiO_2 surface layer. The Si and Ge are indistinguishable by phase contrast. In Figure 22.15B, which is a STEM *Z*-contrast image of the same region, the higher-*Z* Ge crystal region is clearly visible and the lower-*Z* SiO_2 layer appears very dark. The atomic structures of the Si and Ge crystals are visible in both phase-contrast and *Z*-contrast images, although the *Z*-contrast image is noisier. Phase-contrast TEM images can show similar *Z*-contrast effects, as we'll detail in Chapter 28. Figure 22.15C shows a model of a grain boundary superimposed on a *Z*-contrast image which has been refined and processed to reduce the noise via a maximum-entropy approach. You can easily see atomic-level detail.

HAADF

HAADF has the advantage that the contrast is generally unaffected by small changes in objective lens defocus (Δf) and specimen thickness.

We'll see in Chapter 28 that interpretation of atomic-resolution phase-contrast images requires knowledge of t and Δf. Some microscopists claim that *Z* contrast will become the principal method of high-resolution imaging in the future as more FEG STEMs become available; others strongly disagree!

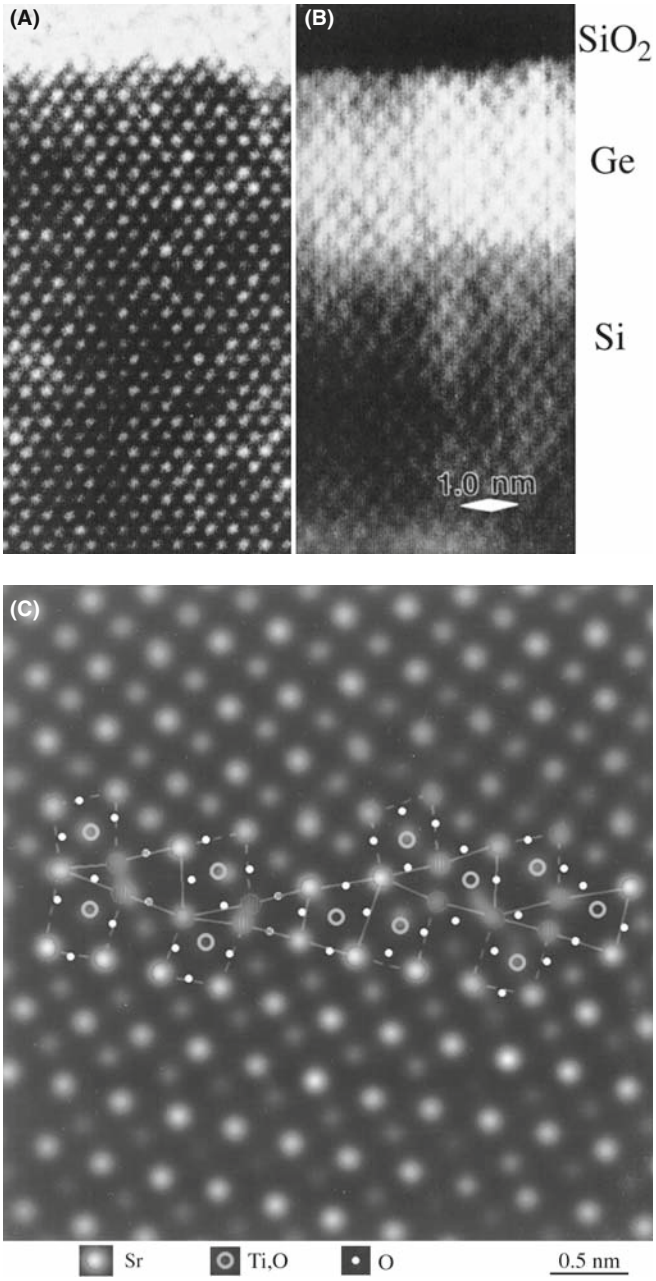


FIGURE 22.15. (A) High-resolution phase-contrast image of epitaxial Ge on Si with an amorphous SiO_2 surface. The bright array of dots common to the crystalline region represents atomic rows and the Ge and Si regions are indistinguishable. (B) The high-resolution *Z*-contrast STEM image shows the atom rows but with strong contrast at the Si–Ge interface and low intensity in the low-*Z* oxide. (C) Model structure of a boundary in $SrTiO_3$ superimposed on a processed *Z*-contrast image.

We can think of the image in Figure 22.15B as a direct map of the $f(\theta)$ variation in the specimen. In that respect, it is similar to an X-ray map showing the distribution of a certain element.

RESOLUTION
The $f(\theta)$ map can have atomic-level resolution, which XEDS imaging can't provide (yet!).

So why do we need a STEM for *Z*-contrast imaging? We are constrained in TEM if we use an analog screen rather than a digital detector to form the image. Nevertheless, we can do *Z*-contrast imaging in a TEM but we have to create electron-optical conditions which are equivalent to those used in STEM. So the beam-convergence angle in TEM must equal the collection angle of the HAADF detector. This is an example of the so-called 'principle of reciprocity' which we'll discuss in more detail in the next section. To converge the TEM beam to the required angular range, we use so-called 'hollow-cone' illumination, which requires an annular C2 aperture. However, the highest incidence angles possible in hollow-cone illumination are typically a few mrad rather than the 50–150 mrad (up to ~9°) collected by the STEM HAADF detector. So TEM *Z*-contrast images are not equivalent to STEM and will always contain some diffraction contrast from crystalline specimens. This leads us into the topic of diffraction contrast, which is the other form of amplitude contrast we see in TEM images.

The future (see Section 39.10): *Z*-contrast images can be collected at the same time and from the same location as the EELS data. There is no other combination of techniques to match this, if it gives the information you need.

22.5 TEM DIFFRACTION CONTRAST

Bragg diffraction, as we discussed in Part 2, is controlled by the crystal structure and orientation of the specimen. We can use this diffraction to create contrast in TEM images. Diffraction contrast is simply a special form of amplitude contrast where the scattering occurs at special (Bragg) angles. We've just seen how incoherent elastic scattering causes mass-thickness contrast. Now we'll see how coherent elastic scattering produces diffraction contrast. As you know, crystalline specimens usually give a single-crystal DP, such as in Figure 22.2. So, as for mass-thickness contrast, we can form BF images by placing the objective aperture around the direct beam (Figure 22.2A) and DF images by selecting any of the diffracted beams (Figure 22.2B). Remember that the incident electrons must be parallel in order to give sharp diffraction spots and strong diffraction contrast. So, if you can, underfocus C2 to spread the beam.

22.5.A Two-Beam Conditions

There is one major difference between forming images to show mass-thickness contrast or diffraction contrast. We can use *any* scattered electrons to form a DF image showing mass-thickness contrast. However, to get good strong diffraction contrast in both BF and DF images

we tilt the specimen to *two-beam conditions*, in which only one diffracted beam is strong. Of course, the direct beam is the other strong spot in the pattern.

Remember: the electrons in the strongly excited *hkl* beam have been diffracted by a *specific* set of *hkl* planes and so the area that appears bright in the DF image is the area where the *hkl* planes are at the Bragg condition. Hence the DF image contains *specific* orientation information, not just general scattering information, as is the case for mass-thickness contrast.

We can tilt the specimen to set up several different two-beam conditions. Figure 22.16A includes a zone-axis DP from a single-crystal specimen in which the beam direction is [011]. The surrounding patterns are a series of two-beam conditions in which the specimen has been tilted slightly so that different *hkl* spots are strongly excited in each pattern. We can form DF images from each strongly diffracted beam after tilting the specimen, and each will give a different image.

TWO-BEAM CONDITIONS

If you're working with crystalline materials, you'll spend a lot of time tilting the specimen to set up different two-beam conditions.

As you can see in Figure 22.16B and C, the BF and DF images show complementary contrast under two-beam conditions. We'll explain the image contrast in detail in Chapter 23. Obviously, to set up a series of two-beam conditions, we need precise tilt control, which explains why a double-tilt eucentric holder is the holder of choice for viewing crystalline specimens.

We'll see in the following chapters that two-beam conditions are not only necessary for good contrast but they also greatly simplify interpretation of the images. This is why we emphasized two-beam theory in our discussion of diffraction in Part 2.

22.5.B Setting the Deviation Parameter, s

Setting up two-beam conditions is very simple. While looking at the DP, tilt around until only one diffracted beam is strong, as in Figure 22.16. As you can see, the other diffracted beams don't disappear because of the relaxation of the Bragg conditions, but they can be made relatively faint. Now if you just do as we've described, the contrast might still not be the best. For reasons we'll discuss in detail in the next chapter, to get the best contrast from defects, your specimen shouldn't be *exactly* at the Bragg condition ($\mathbf{s} = 0$) as in Figure 22.17A. Tilt your specimen close to the Bragg condition but make **s** small and positive. (The excess *hkl* Kikuchi line, just outside the *hkl* spot; go back and check the text for Figures 19.10 and 19.11.) This will give you the best possible strong-beam image contrast as in Figure 22.17B. If you tilt the specimen slightly, so **s** increases further as shown in Figure 22.17C, the defect images become narrower but the contrast is reduced.

USE $\mathbf{s}>0$

Never form strong-beam images with **s** negative; it's easier to see the defects when $\mathbf{s}>0$.

22.5.C Setting Up a Two-Beam CDF Image

We described the basic mechanism of forming BF and DF images back in Chapter 9 (Figure 9.14A). To produce the best BF diffraction contrast, tilt to the desired two-beam condition as in Figure 22.18A, and insert the objective aperture on axis as in Figure 22.2A. A two-beam CDF image is not quite as simple. You might think it involves just tilting the incident beam so the strong *hkl* reflection moves onto the optic axis. If you do that, you'll find that the *hkl* reflection becomes weaker as you move it onto the axis and the $3h3k3l$ reflection becomes strong, as shown in Figure 22.18B. What you've just done is in fact set up a *weak-beam* image condition, which we'll discuss in Chapter 27. To set up a *strong-beam* CDF image, tilt in the $\bar{h}\bar{k}\bar{l}$ reflection which was initially weak, and it becomes strong as it moves on axis, as shown in Figure 22.18C. The CDF technique is absolutely crucial for obtaining and interpreting diffraction-contrast images, so we will take you through it in detail

- Look at the SADP and tilt the specimen until the desired *hkl* reflection is strong. Make sure the incident beam is well underfocused.
- Now tilt the specimen until the $\bar{h}\bar{k}\bar{l}$ reflection is strong: *hkl* will now be weak.
- Use the DF tilt controls to move the 000 reflection towards the strong $\bar{h}\bar{k}\bar{l}$ reflection. The weak *hkl* reflection will move towards the optic axis and become strong.
- When *hkl* is close to the axis, switch off the DF deflectors, insert and carefully center the objective aperture around 000.
- Switch the DF tilt coils on and off while looking through the binoculars. Check that the *hkl* and 000 reflections appear in the same position. Make fine adjustments to the DF coils until you can see no shift between 000 and *hkl* when the deflectors are off and on, respectively.
- Switch to image mode. If necessary, condense the beam slightly with C2 until you can see the CDF image. If you can't see an image, either the *hkl* reflection is too weak (unlikely) or your tilt coils are misaligned (common). In the latter case, realign the coils (see the manufacturer's handbook).

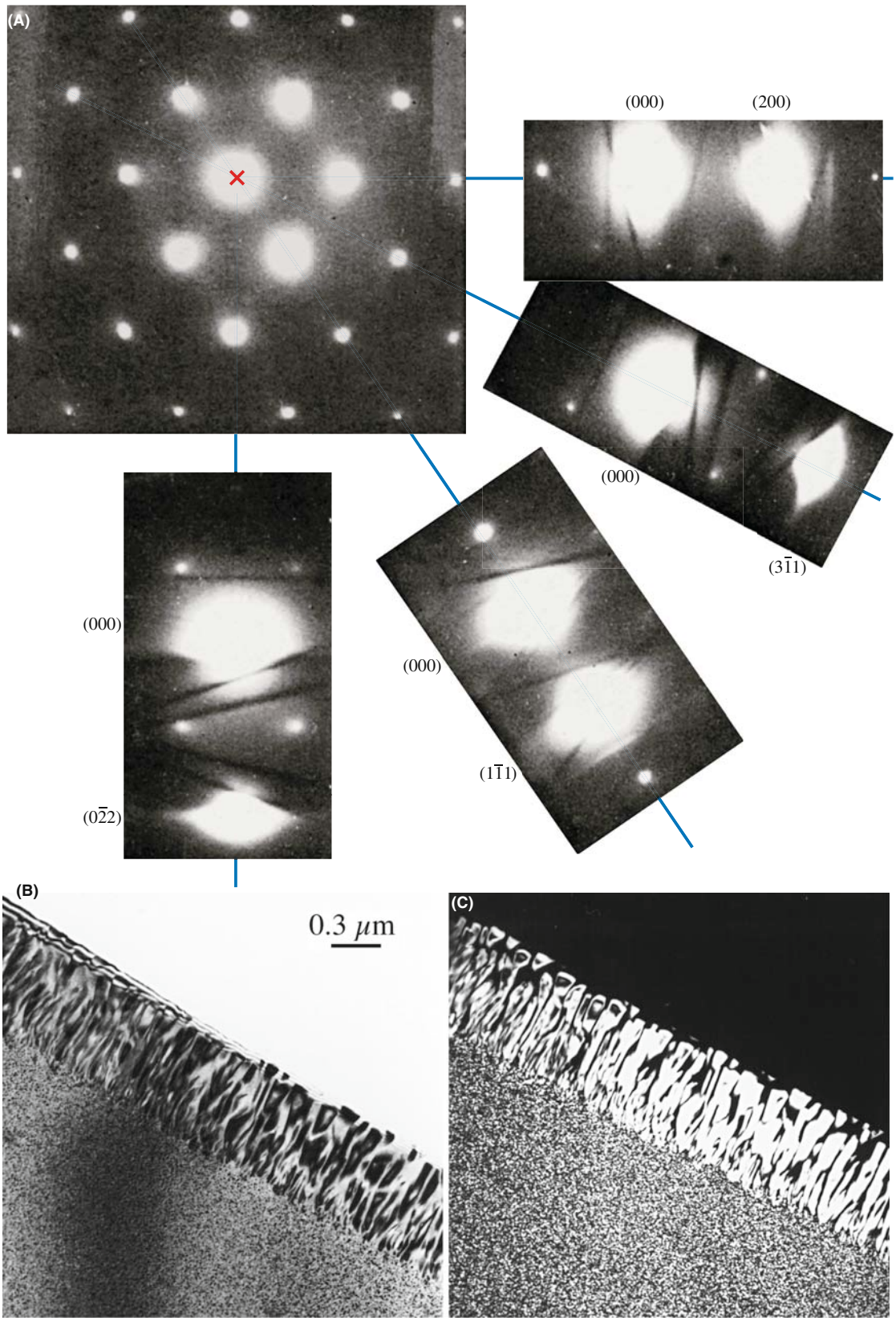


FIGURE 22.16. (A) The [011] zone-axis diffraction pattern has many planes diffracting with equal strength. In the smaller patterns, the specimen is tilted so there are only two strong beams, the direct 000 on-axis beam and a different one of the *hkl* off-axis diffracted beams. Complementary (B) BF and (C) DF images of Al-3 wt% Li taken under two-beam conditions are shown also. In (B) the Al_3Li precipitate phase (present as tiny spheres in the grain and coarse lamellae at the boundary) is diffracting strongly and appears dark. In (C), imaged with a precipitate spot, only the diffracting precipitates appear bright.

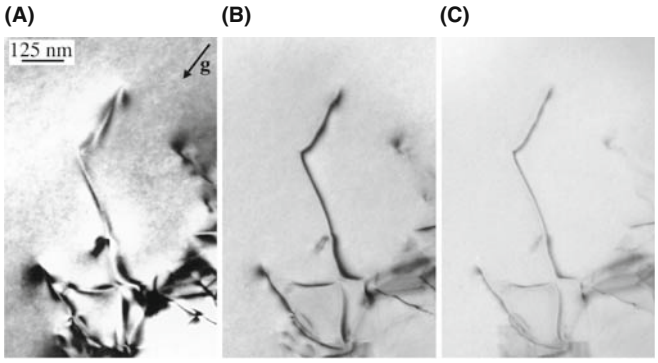


FIGURE 22.17. Variation in the diffraction contrast when **s** is varied from (A) zero to (B) small and positive and (C) larger and positive.

Now go back and study Figure 9.14C carefully. You'll see that the beam was tilted through an angle $2\theta_B$ to bring the weak beam in Figure 9.14B onto the optic axis.

Notice that $\bar{\mathbf{g}}$ is excited in the DF condition; **g** was excited in the BF condition. If you want to form a CDF image with **g** excited, you must tilt the beam *and* the specimen. Alternatively, just displace the aperture but be cautious.

22.5.D Relationship Between the Image and the Diffraction Pattern

From what we've just described, there is clearly an important relationship between the DP and a diffraction-contrast image. If we change the DP in any way, the contrast in the image will change. So it is critical to relate the DP to the image. We need to indicate the direction of the **g** vector in the image. To relate the two, remember that you may have to calibrate the rotation between the image and the DP if, whenever you change magnification, your image rotates but your DP does not. We described this calibration in Section 9.6.C. You should usually show the **g** vector in any BF or DF two-beam, diffraction-contrast image after correcting for any rotation between the image and the DP. Remember: you should always check the calibration of any microscope, especially a TEM, and be careful with possible 180° rotations.

ROTATION CALIBRATION
Do it even if only to convince yourself that you don't need to!

We will expand on diffraction contrast in far more detail in the subsequent chapters, making use of the fundamental operational principles we have just described.

22.6 STEM DIFFRACTION CONTRAST

The principle of forming BF and DF images in STEM is just the same as for mass-thickness contrast; i.e., use the BF detector to pick up the direct beam and the ADF

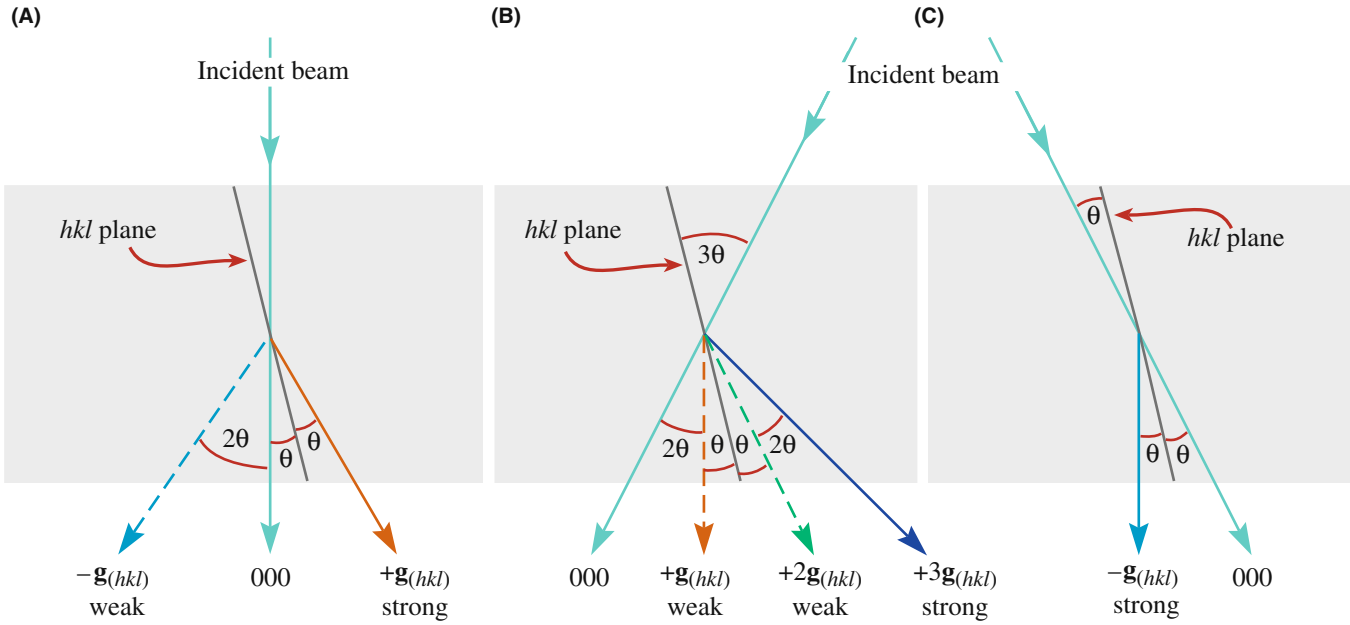


FIGURE 22.18. (A) Standard two-beam conditions involve the 000 spot and the *hkl* spot bright because one set of *hkl* planes are exactly at the Bragg condition. (B) When the incident beam is tilted through 2θ so that the excited $\mathbf{g}_{hkl}$ spot moves onto the optic axis, the $\mathbf{g}_{hkl}$ intensity decreases because the $\mathbf{g}_{3h3k3l}$ spot becomes strongly excited. (C) To get a strong *hkl* spot on axis for a CDF image, it is necessary to set up a strong $-\mathbf{g}_{hkl}$ condition first of all, then tilt the initially weak $\mathbf{g}_{hkl}$ maximum onto the axis.

detector to pick up the diffracted beams. To preserve two-beam conditions, the ADF detector must only pick up one strong diffracted beam and this can be ensured by inserting the objective aperture and selecting only one diffracted beam. Alternatively, the DP could be displaced so the chosen *hkl* reflection falls on the BF detector. Either way the computer screen will display a DF image.

However, the diffraction contrast observed in the STEM image will generally be much poorer than TEM contrast; the normal STEM operating conditions are not equivalent to the TEM conditions that ensure strong diffraction contrast. To understand the contrast in STEM images you need to know the beam convergence and detector collection angles. It's rare in fact that you'll need to do this, but we showed you how to determine the beam convergence angle back in Section 5.5. To calculate the collection angle, you need to carry out a similar exercise as we use to determine the EELS spectrometer collection angle in Section 37.4.

Remember, there are three conditions that must be fulfilled for strong contrast in your image

- The incident beam must be coherent, i.e., the convergence angle must be very small.
- The specimen must be tilted to a two-beam condition.
- Only the direct beam or the one strong diffracted beam must be collected by the objective aperture.

This condition is shown schematically in Figure 22.19A. We define the TEM convergence angle as α_T and the objective aperture collection angle as β_T. In a STEM, the equivalent angles are the beam convergence angle α_S and the STEM detector collection angle β_S as shown in Figure 22.19B. Therefore, we have identical operating conditions if

$$\alpha_T = \alpha_S \tag{22.11a}$$

$$\beta_T = \beta_S \tag{22.11b}$$

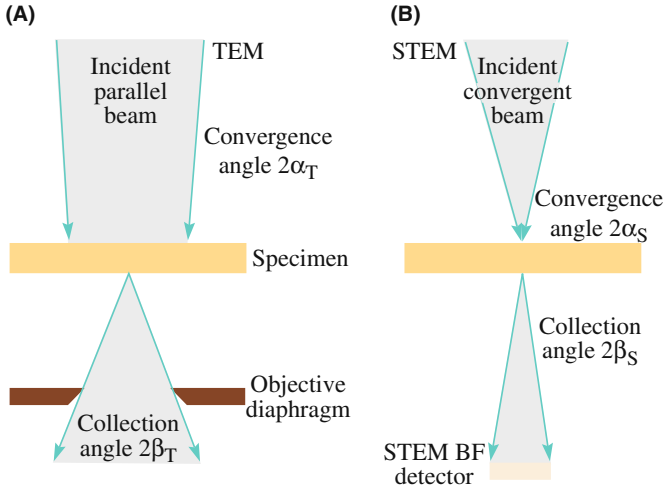


FIGURE 22.19. Comparison of the important beam-convergence and divergence angles (A) in TEM and (B) in STEM. (Note that we show the full convergence and divergence angles, not the semiangles that are usually described in the text.)

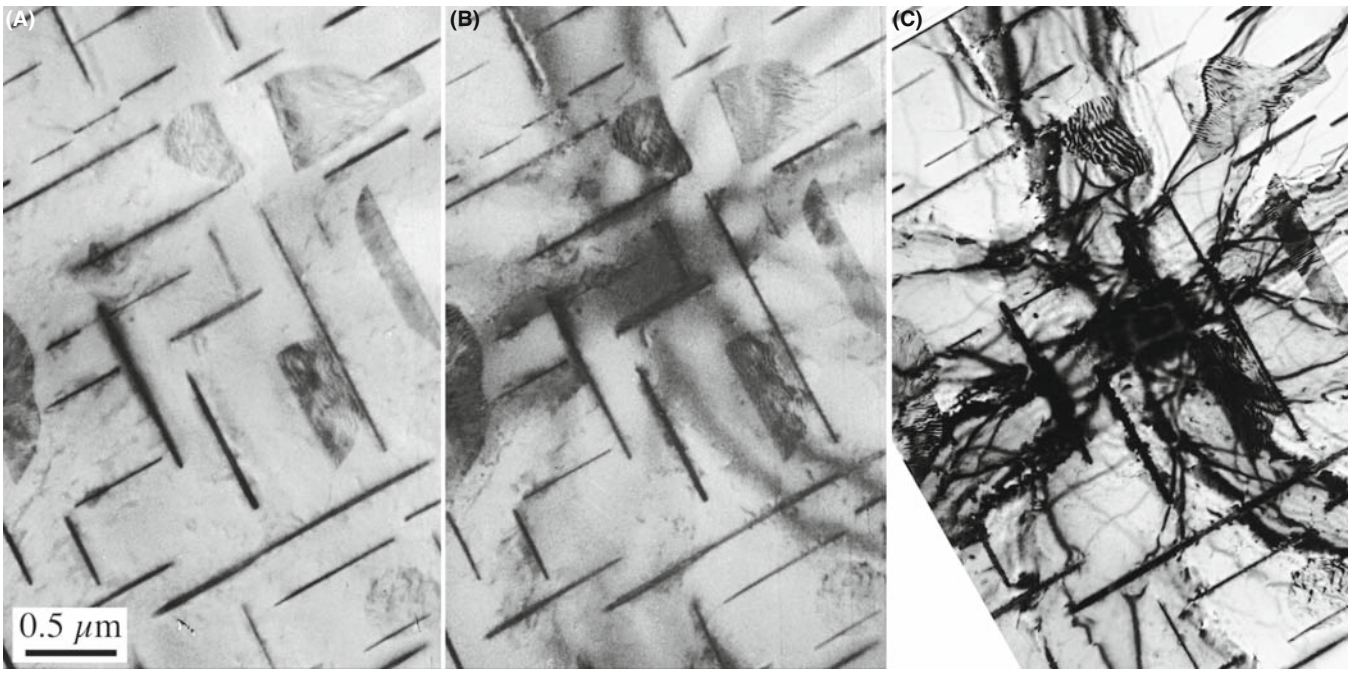


FIGURE 22.20. (A) A BF STEM image of an Al-4 wt% Cu specimen showing weak diffraction contrast in the form of bend contours. As the STEM detector collection angle is lowered (B), the diffraction contrast increases slightly at the expense of increased noise in the image. Even at the smaller collection angle, comparison with the contrast in the TEM image (C) is unfavorable. Note that the Cu-rich θ′ precipitates maintain strong mass contrast in all the images.

Now it should be immediately clear that we can't get such equivalence in a STEM because the convergence angle of the beam is very much greater than in a TEM (since in STEM we deliberately create a convergent rather than a parallel beam). However, there is a way around this dilemma and it depends on a theorem that is often used in electron optics called the *principle of reciprocity*. In essence this principle says that so long as the electron ray paths contain equivalent angles (of convergence and collection) at some point in the electron optical system, the image contrast will be identical.

In other words, while the conditions in equations 22.11 can't be fulfilled, we can create conditions such that

$$\alpha_S = \beta_T \tag{22.12a}$$

$$\alpha_T = \beta_S \tag{22.12b}$$

Under these circumstances the electrons in TEM and STEM do see equivalent angular constraints, although not at the equivalent points of convergence and collection.

- Since the objective-aperture collection angle in TEM is about equal to the convergence angle in STEM, the first of this pair of equations is easily satisfied.
- To satisfy the second pair, we have to make a very small STEM collection angle β_S.

We can't simply increase α_T, because we must keep a parallel beam to get good TEM diffraction contrast and making the beam non-parallel (large α_T) destroys the contrast.

There is an obvious drawback to making β_S small. The signal falling on the STEM detector becomes very small and the STEM image becomes noisy. So STEM diffraction-contrast images become noisier as we attempt to increase the amount of diffraction contrast, as in Figure 22.20. (See Chapter 24 for an explanation of the contrast (bend contours) in this figure.) Having an FEG helps to offset this increase in noise, but in general, STEM diffraction-contrast images (in both BF and DF) compare so unfavorably with TEM images (see Figure 22.20C) that, while they may be useful if you're performing analysis, they are rarely used to show diffraction-contrast images of crystal defects. This is solely the domain of TEM, as we'll discuss in detail in the next few chapters.

CHAPTER SUMMARY

Mass-thickness contrast and diffraction contrast are two forms of amplitude contrast. Both arise because the specimen scatters electrons. The operational procedures to produce BF and DF images are identical. Interpretation of mass-thickness contrast is generally simpler than interpretation of diffraction contrast. In fact, the interpretation of diffraction contrast is sufficiently complex that we need to devote several subsequent chapters to the various forms arising in perfect and imperfect crystals.

We can summarize the characteristics of mass-thickness contrast

- Areas of greater Z and/or t scatter electrons more strongly (in total), and therefore appear darker in BF images and brighter in DF images. The contrast can be quantified if necessary.
- TEM mass-thickness contrast images are better quality (lower noise and higher resolution) than STEM images, but digital STEM images can be processed to show higher contrast than analog TEM images.
- STEM mass-thickness contrast images are most useful for thick and/or beam-sensitive specimens.
- Z-contrast (HAADF) images can show atomic-level resolution.

We can summarize the characteristics of diffraction contrast

- Diffraction contrast arises when the electrons are Bragg scattered.
- To form a diffraction-contrast image in TEM, the objective aperture selects one Bragg-scattered beam. Often, the STEM detectors gather several Bragg beams which reduce diffraction contrast.
- Diffraction-contrast images in TEM always show better contrast than in STEM images, which are always noisier and almost never used.

Some of the statements made in the First Edition are not now valid! FEGs are now common. The original DSTEM was discontinued in ~1985; but DSTEMs are now made by Hitachi, JEOL, and Nion. C_s correctors are becoming more available. Modern TEMs all use digital recording. Z-contrast imaging has greatly improved on commercial machines and rivals CTEM in many applications.

REFERENCES

References on diffraction contrast are included in Chapters 24–27.

GENERAL IMAGING AND SOME HISTORY

The topic of staining is discussed and used much more for biological samples.

Cosslett, VE 1979 *Penetration and Resolution of STEM and CTEM in Amorphous and Polycrystalline Materials* Phys. Stat. Sol. **A55** 545–548. Early review of using STEM in biological specimens by one of the pioneers of TEM.

Heidenreich, RD 1964 *Fundamentals of Transmission Electron Microscopy* p 31 John Wiley & Sons New York. More details on the derivation of equation 22.2.

Humphreys, CJ 1979 *Introduction to Analytical Electron Microscopy* p305 Eds. JJ Hren, JI Goldstein and DC Joy Plenum Press New York.

Reimer, L 1997 *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis* 4th Ed. Springer-Verlag New York. Includes a more detailed discussion of equation 22.6.

Sawyer, LC, Grubb, DT and Meyers, DT 2008 *Polymer Microscopy* 3rd Ed. Springer New York. Gives a good introduction to staining for the materials scientist.

Watt, IM 2003 *The Principles and Practice of Electron Microscopy* 2nd Ed. Cambridge University Press New York. See in particular the discussion on shadowing.

EARLY HAADF

Brown, LM 1977 *Progress and Prospects for STEM in Materials Science* Inst. Phys. Conf. Ser. No. **36** 141–148. Diffraction effects in low-angle ADF.

Donald, AM and Craven, AJ 1979 *A Study of Grain Boundary Segregation in Cu-Bi Alloys Using STEM* Phil. Mag. **A39** 1–11.

Howie, A 1979 *Image Contrast and Localized Signal Selection Techniques* J. Microsc. **117** 11–23.

Isaacson, M, Ohtsuki, M and Utlaut, M 1979 in *Introduction to Analytical Electron Microscopy* p 343 Eds. JJ Hren, JI Goldstein and DC Joy Plenum Press New York. Includes discussion of early *Z*-contrast imaging.

Jesson, DE and Pennycook, SJ 1995 *Incoherent Imaging of Crystals Using Thermally Scattered Electrons* Proc. Roy. Soc. (Lond.) **A449** 273–293.

Pennycook, SJ 1992 *Z-Contrast Transmission Electron Microscopy: Direct Atomic Imaging of Materials* Annu. Rev. Mat. Sci. **22** 171–195.

Treacy, MMJ, Howie, A and Pennycook, SJ 1980 *Z Contrast of Supported Catalyst Particles on the STEM* Inst. Phys. Conf. Ser. No. **52** 261–264.

Treacy, MMJ, Howie, A and Wilson, CJ 1978 *Z Contrast of Platinum and Palladium Catalysts* Phil. Mag. **A38** 569–585. Early paper noting the advantage of higher-angle scattering in *Z*-contrast imaging.

THE COMPANION TEXT

The chapter on simulating images is particularly relevant but you must understand what you are simulating first. The companion text also includes a full chapter on HAADF.

SELF-ASSESSMENT QUESTIONS

Q22.1 What do we mean by the term minimum contrast and can we quantify it?
Q22.2 When considering contrast (either qualitatively or quantitatively), why do we immediately consider the DP?
Q22.3 Will a TEM be able to resolve two atoms next to each other in the periodic table?
Q22.4 You are shown a polymer sphere in a TEM image. Why would you be cautious?
Q22.5 How would you prepare an unstained polymer sample for TEM imaging?
Q22.6 Why is the resolution in STEM not as good as in TEM on most TEMs?
Q22.7 Why is resolution on some STEMs as good as, or better than, many TEMs?
Q22.8 How are diffraction-contrast images different from mass-thickness contrast images?
Q22.9 No matter how hard a particular student tries to obtain a two-beam condition, he or she still sees the direct beam, a strongly diffracted beam, and several weak spots in the DP. Could the student be at the two-beam condition?
Q22.10 What is the difference between a phase-contrast image and an amplitude-contrast image?
Q22.11 What causes mass-thickness contrast?
Q22.12 What influences the intensity of mass-thickness contrast?
Q22.13 For what materials is mass-thickness contrast most useful?
Q22.14 What microscope controls can affect the mass-thickness contrast?
Q22.15 What features of a STEM give more flexibility in mass-thickness contrast imaging?
Q22.16 What is diffraction contrast?

Q22.17 How can you obtain good diffraction contrast in BF and DF images?
Q22.18 How can you obtain the best strong-beam contrast from defects?
Q22.19 How are *Z*-contrast imaging and HAADF imaging related?
Q22.20 Does a BF STEM image show strong contrast from bend contours? Explain after you've read Chapter 24.

TEXT-SPECIFIC QUESTIONS

T22.1 Where is the Ewald sphere in Figure 22.2? Give a full explanation.
T22.2 How was the specimen prepared in Figure 22.12 and why is this information important?
T22.3 Redraw Figure 22.13 showing all the angles and distances to scale and with typical values in mrad and degrees. Are the values of θ correct/reasonable/sensible? Explain your reasoning carefully.
T22.4 What do you guess is the geometry of the specimen shown in Figure 22.14?
T22.5 Draw an atomistic diagram to explain the contrast seen in Figure 22.15B.
T22.6 Draw and label direction vectors on Figure 22.15C for both grains.
T22.7 Index all the spots in the diffraction pattern in Figure 22.16.
T22.8 Explain why the bend contours in STEM images are much less pronounced than in TEM diffraction-contrast images.
T22.9 In Figure 22.17, why are some of the defects in C almost out of contrast?
T22.10 Suggest how the contrast in Figure 22.16C is influenced by the fact that the Al_3Li precipitate on the grain boundary is a lamella.

23

Phase-Contrast Images

CHAPTER PREVIEW

We see phase contrast any time we have more than one beam contributing to the image. In fact, whenever we say "fringes," we are essentially referring to a phase-contrast phenomenon. Although we often distinguish phase and diffraction contrast, this distinction is generally artificial. For example, in Chapters 24 and 25, we will examine thickness fringes and stacking-fault fringes; both types of contrast result from interference of waves so both are phase-contrast images although we usually think of them as two-beam, diffraction-contrast images.

Phase-contrast imaging is often thought to be synonymous with high-resolution TEM. In fact, phase contrast appears in most TEM images even at relatively low magnifications. We will draw your attention to its role in the formation of moiré patterns and Fresnel contrast at defects. This Fresnel contrast has the same origin as the Fresnel fringes that we used in Chapter 9 to correct the astigmatism of the objective lens.

As with many of the topics we've discussed, we can approach the problem at several different levels. One danger is that you may be tempted to use one of the prepackaged simulation programs to predict phase-contrast images, without first learning about the limitations of such packages. The second danger is that with a modern HRTEM you can very easily obtain phase-contrast images that show great detail without you fully knowing/understanding the conditions used to obtain the image. This may impress your advisor but you should know better.

We will begin this chapter by discussing some simple approaches to understanding phase-contrast effects as they relate to lattice-fringe imaging.

23.1 INTRODUCTION

Contrast in TEM images can arise due to the differences in the phase of the electron waves scattered through a thin specimen. This contrast mechanism can be difficult to interpret because it is very sensitive to many factors: the appearance of the image varies with small changes in the thickness, orientation, or scattering factor of the specimen, and variations in the focus or astigmatism of the objective lens. However, its sensitivity is the reason phase contrast can be exploited to image the *atomic structure* of thin specimens. Of course this also requires a TEM with sufficient resolution to detect contrast variations at atomic dimensions and the proper control of instrument parameters that affect the phases of the electrons passing through the specimen and the lenses. If you know what you are doing, the procedures can be straightforward; the level of operator skill that is necessary to obtain such images can be acquired with practice; the challenge is in getting enough practice and knowing when you have!

The most obvious distinction between phase-contrast imaging and other forms of TEM imaging is the number of beams collected by the objective aperture or an electron detector. As described in the previous chapters, a BF or DF image requires that we select a single beam using the objective aperture. A phase-contrast image requires the selection of *more than one* beam. In general, the more beams collected, the higher the resolution of the image. However, we will see that there are reasons why some beams, which are apparently admitted through the aperture, might not contribute to the image. The details of this process depend on the performance of the electron-optical system. We'll first examine the theory and then consider the practical aspects.

23.2 THE ORIGIN OF LATTICE FRINGES

We can understand the origin of lattice fringes by extending the analysis of Chapter 13 to allow the two beams, **0** and **g**, to interfere; i.e., use the objective aperture to select only two beams. We begin by rewriting equation 13.5

$$\psi = \phi_{\mathbf{0}}(z)\exp 2\pi i(\mathbf{k}_{\mathrm{I}} \,.\, \mathbf{r}) + \phi_{\mathbf{g}}(z)\exp(2\pi i\mathbf{k}_{\mathrm{D}} \,.\, \mathbf{r}) \qquad (23.1)$$

where we know

$$\mathbf{k}_D = \mathbf{k}_I + \mathbf{g} + \mathbf{s}_g = \mathbf{k}_I + \mathbf{g}' \tag{23.2}$$

We are thus using a two-beam approximation but allowing $\mathbf{s}_g$ to be non-zero. Now we will make some simple substitutions setting $\phi_0(z) = A$ and take $e^{2\pi i \mathbf{k}_I . \mathbf{r}}$ out as a factor. We will also represent the expression for ϕ_g from equation 13.5 as

$$\phi_g = B \exp i\delta \tag{23.3}$$

where

$$B = \frac{\pi}{\xi_g} \frac{\sin \pi t s_{eff}}{\pi s_{eff}} \tag{23.4}$$

and

$$\delta = \frac{\pi}{2} - \pi t s_{eff} \tag{23.5}$$

The $\pi/2$ in the expression for δ takes care of i in equation 13.5 and we'll pretend that the specimen is so thin that we can replace s_{eff} by s. Thus equation 23.1 becomes

$$\psi = \exp(2\pi i \mathbf{k}_I . \mathbf{r})[A + B \exp i(2\pi \mathbf{g}' . \mathbf{r} + \delta)] \tag{23.6}$$

The intensity can then be expressed as

$$I = A^2 + B^2 + AB[\exp i(2\pi \mathbf{g}' . \mathbf{r} + \delta) + \exp -i(2\pi \mathbf{g}' . \mathbf{r} + \delta)] \tag{23.7}$$

$$I = A^2 + B^2 + 2AB \cos(2\pi \mathbf{g}' . \mathbf{r} + \delta) \tag{23.8}$$

Now $\mathbf{g}'$ is effectively perpendicular to the beam so we'll set it parallel to x and replace δ giving

$$I = A^2 + B^2 - 2AB \sin(2\pi g' x - \pi s t) \tag{23.9}$$

Therefore, the intensity is a sinusoidal oscillation normal to $\mathbf{g}'$, with a periodicity that depends on s and t. (Note that g and s are not bold in equation 23.9 because they represent the magnitude of the vectors, not the vectors themselves.) We can, with care, relate these fringes to the spacing of the lattice planes normal to $\mathbf{g}'$. Although we have obtained this equation using a very simple model, it gives us some useful insight, which will also be helpful when we talk about many-beam images in Chapter 28.

The intensity varies sinusoidally with different periodicities for different values of $\mathbf{g}'$. This model will be equally valid even if the incident beam is tilted slightly off the optic axis.

DELOCALIZATION

This simple analysis shows that the location of a fringe does not necessarily correspond to the location of a lattice plane.

23.3 SOME PRACTICAL ASPECTS OF LATTICE FRINGES

23.3.A If s = 0

If we just have **0** and **g** in the objective aperture and we then set $\mathbf{s} = 0$ for reflection G (so $\mathbf{g}' = \mathbf{g}$), we will see fringes in the image (Figure 23.1A) which have a periodicity of $1/g$ in the x direction; i.e., the fringe periodicity is the same as the spacing of the planes which give rise to **g**. This result holds wherever $\mathbf{s} = 0$ no matter how **0** and **g** are located relative to the optic axis, even if the diffracting planes are not parallel to the optic axis.

Figure 23.1B shows the ideal geometry for producing images like Figure 23.1A. It is called the 'tilted-beam condition' and it means that the planes of interest lie parallel to the optic axis. If we use the geometry shown in Figure 23.1B we have $\mathbf{s} = 0$ and the planes are parallel to the optic axis but not parallel to the incident beam. Therefore the fringes cannot correspond directly to the individual planes. If we use the on-axis geometry shown in Figure 23.1C, the planes are viewed edge on, but $\mathbf{s} \neq 0$ for reflection G; so we must also consider reflection –G.

CONFUSING

The trouble is that the fringes look so like atomic planes that we can be easily misled into thinking that they are atomic planes.

23.3.B If s ≠ 0

If the specimen is not exactly flat, then **s** will vary across the image; even if you set $\mathbf{s} = 0$ in the DP, it will not be

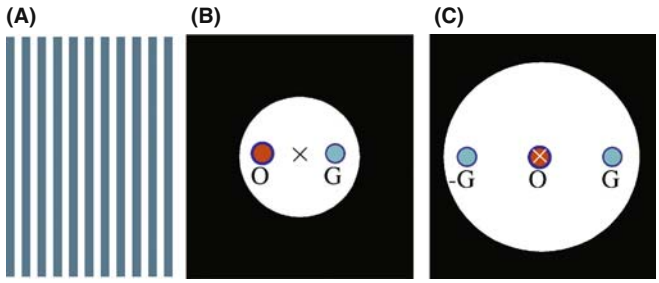


FIGURE 23.1. (A) Schematic tilted-beam 111 lattice fringes in Si formed using the O and G beams symmetrically displaced relative to the optic axis; **g** is normal to the fringes. (B) Ideal diffraction geometry to produce tilted-beam fringes. (C) On-axis, three-beam geometry.

zero everywhere. If **s** is not zero, then the fringes will shift by an amount which depends on both the magnitude of **s** and the value of t, but the periodicity will not change noticeably. We expect this **s** dependence to affect the image when the foil bends slightly, as is often the case for thin specimens. We also expect to see thickness variations in many-beam images, since **s** may be non-zero for all of the beams; **s** may also vary from beam to beam.

23.4 ON-AXIS LATTICE-FRINGE IMAGING

We've just seen that two beams can interfere to give an image with a periodicity related to $|\Delta\mathbf{g}|^{-1}$. Since one beam is the direct beam, $|\Delta\mathbf{g}|^{-1}$ is just d, the interplanar spacing corresponding to **g**. If you align your beam parallel to a low-index zone axis then you'll see fringes running in different directions; these fringes in the image must correspond to an array of spots in the DP. The spacings of the spots may be inversely related to the lattice spacings, as shown in Figure 23.2 which extends Figure 23.1 to the many-beam case. In general, this array of spots bears no direct relationship to the *position* of atoms in the crystal.

We'll see more on this when we discuss image simulation in Chapter 30. In case you are in doubt, compare the beautiful image shown in Figure 23.3A with the projected structure of Si in Figure 23.3B. The Si dumbbells are a pair of atoms which are 1.4 Å apart in this projection of the structure. The aperture used to form the image included 13 reflections, as shown in Figure 23.3C. The difficulty is that in the image, the spots in the dumbbell image are really only 1.3 Å apart but the point resolution of the TEM was only ~2.5 Å. You can see from the

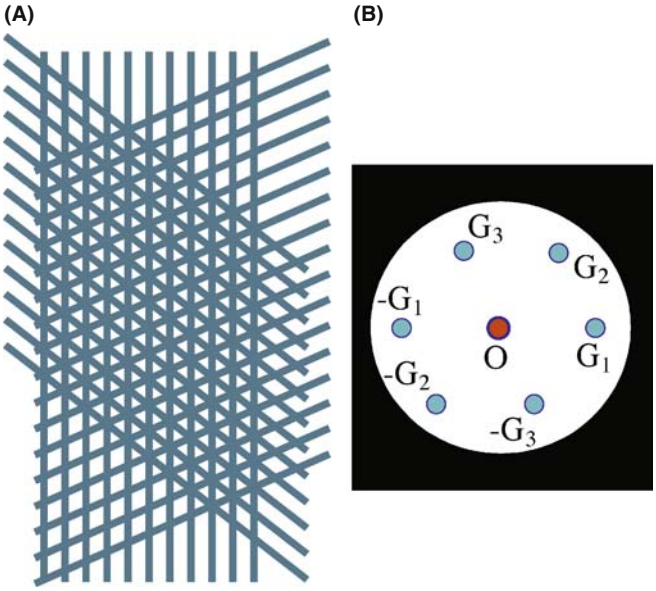


FIGURE 23.2. (A) Schematic many-beam image showing crossing lattice fringes and (B) the diffraction pattern.

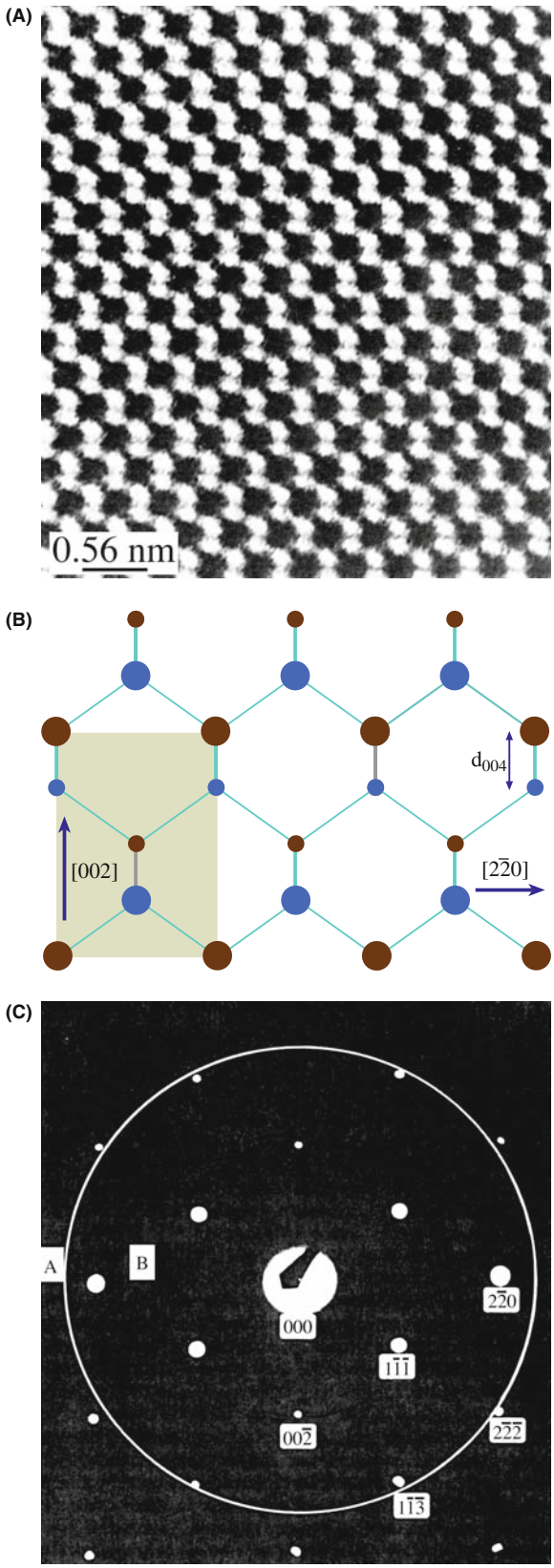


FIGURE 23.3. (A) On-axis image of a perfect Si crystal; (B) the projected structure; (C) the diffraction pattern showing the 13 spots used to form the image inside the aperture (ring). The Si dumbbells do not correspond to the closely spaced pairs of spots in the image.

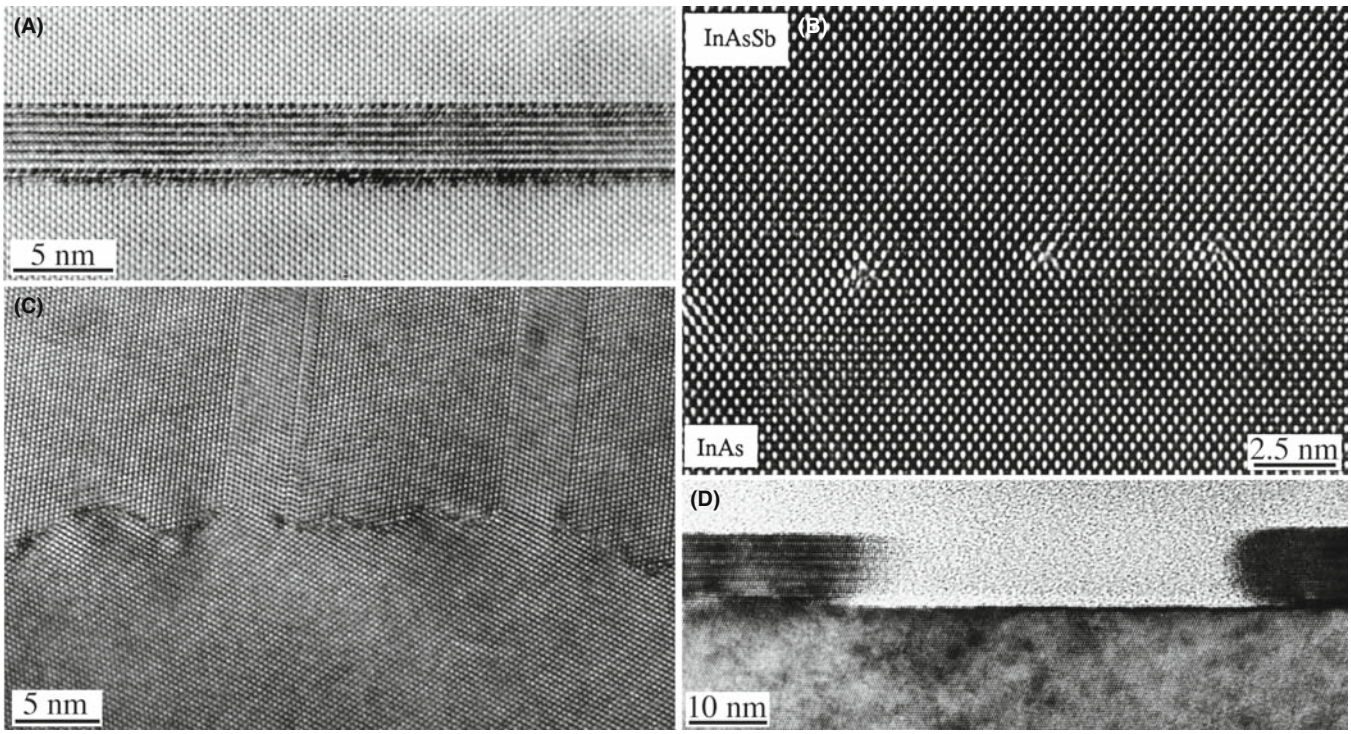


FIGURE 23.4. Illustrations of lattice images which contain easily interpreted information. (A) The spinel/olivine interface; (B) dislocations at a heterojunction between InAsSb and InAs; (C) a grain boundary in Ge faceting on an atomic scale; (D) a profile view of a faceted surface.

structure that the real dumbbell spacing corresponds to the (004) plane spacing, but the 004 reflection was *not* used to form the image. The explanation was given by Krivanek and Rez; the dumbbells in the image are caused by the crossing {113} fringes—so the contrast is an artifact, if we are interested in the real structure.

The lesson is: we only knew the image did not correspond to the structure because we knew the structure! Taking this example as your guide, consider the case where a defect is present in an image where the perfect crystal spots are all in the 'correct' position. Could you still be certain that the detail in the image close to the defect gives you a true picture of the location of the atoms close to the defect? The answer is, of course, "no."

FRINGES NOT PLANES

So lattice fringes are not direct images of the structure, but just give you information on lattice spacing and orientation.

On-axis lattice-fringe images are perhaps best used as a measure of the local crystal structure and orientation. The exception, as we'll see in the next chapter, is when these images can only be interpreted using extensive computer simulation. Figure 23.4 illustrates some typical applications of the imaging mode, where we can learn a lot about our material by intuitive interpretation without the need for simulating their images. Our guess is that at least 99% of all lattice-fringe images are interpreted without using image simulation.

Figure 23.4A shows the interface between a spinel particle and an olivine matrix; Figure 23.4B shows how we can locate dislocations at a heterojunction; Figure 23.4C shows the atomic-scale faceting of a grain boundary in Ge; and Figure 23.4D illustrates the faceting of a surface.

23.5 MOIRÉ PATTERNS

Moiré (pronounced mwa-ray) patterns can be formed by interfering two sets of lines that have nearly common periodicities. We can demonstrate two fundamentally different types of interference: the rotational moiré and the translational (often referred to as misfit) moiré. It's easy to understand moirés if you make three transparent sheets of parallel lines (two with the same spacing and one slightly different): you can generate such sets of lines readily using any computer, choosing the line widths to be similar to the gaps between them. Then try these three exercises. (You should 'do' the experiment by making your own 'specimen' or using those on the Web site.)

- Take two misfit sets and align them exactly. This gives a set of moiré fringes which are parallel to the lines forming them, as shown in Figure 23.5A.

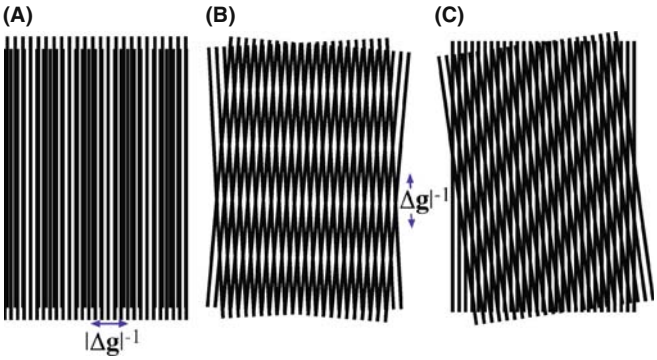


FIGURE 23.5. (A) Translation moiré fringes; (B) rotation moiré fringes; (C) mixed moiré fringes; note the relationship between the fringes and their constituent lattices.

- Take two identical sets of lines and rotate them. Now, we produce a set of moiré fringes which is perpendicular to the average direction of the initial lines (Figure 23.5B).
- Take the first two sets and rotate them so we produce moiré fringes as in Figure 23.5C; but note that their alignment to your reference sets is not obvious.

When the misfit or misorientation is small, the moiré fringe spacing is clearly much coarser than that of the lines themselves. In particular, if the sets of lines in Figure 23.5 are actually lattice planes in a crystal, the moiré fringes may give information about the crystals even if you cannot resolve the lattice planes. The simplest way to analyze the spacings and orientation of the moiré fringes is to consider the diffraction vectors from the two 'lattices.' Incidentally, the term 'moiré' originated in the textile industry; it's related to the French word for 'mohair,' the silky hair of the Angora goat, hence the watery or wavy pattern seen in silk fabrics—and the lowercase 'm'!

23.5.A Translational Moiré Fringes

In this case, since the planes are parallel, the **g**-vectors will also be parallel. If we write these as $\mathbf{g}_1$ and $\mathbf{g}_2$, we produce a new spacing $\mathbf{g}_{tm}$ given by

$$\mathbf{g}_{tm} = \mathbf{g}_2 - \mathbf{g}_1 \tag{23.10}$$

In Figure 23.6A, we have assigned $\mathbf{g}_2$ to the smaller 'lattice' spacing and tm indicates 'translational moiré' fringes. The vector $\mathbf{g}_{tm}$ corresponds to a set of fringes with spacing d_{tm}, as shown by the following simple manipulation

$$d_{tm} = \frac{1}{g_{tm}} = \frac{1}{g_2 - g_1} = \frac{\frac{1}{g_2}\cdot\frac{1}{g_1}}{\frac{1}{g_1} - \frac{1}{g_2}} = \frac{d_2 d_1}{d_1 - d_2} = \frac{d_1}{1 - \frac{d_2}{d_1}} \tag{23.11}$$

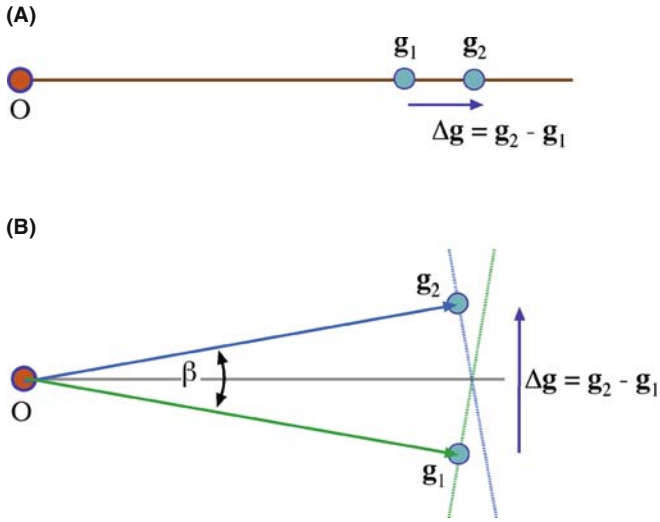


FIGURE 23.6. (A) Relationship between **g**-vectors and translational moiré fringes; (B) relationship for rotational moiré fringes.

23.5.B Rotational Moiré Fringes

We follow the same procedure as above, but now the two **g**-vectors are identical in length and rotated through an angle β so that the new **g**-vector, $\mathbf{g}_{rm}$, has length $2\mathbf{g}\sin\beta/2$, as shown in Figure 23.6B. The fringe spacing is then

$$d_{rm} = \frac{1}{g_{rm}} = \frac{1}{2g\sin\beta/2} = \frac{d}{2\sin\beta/2} \tag{23.12}$$

23.5.C General Moiré Fringes

If we use the same approach to locate $\mathbf{g}_{gm}$ (gm: 'general moiré') we can readily show that, for small misorientation, the spacing d_{gm} of our fringes is given by

$$d_{gm} = \frac{d_1 d_2}{\left((d_1 - d_2)^2 + d_1 d_2 \beta^2\right)^{1/2}} \tag{23.13}$$

23.6 EXPERIMENTAL OBSERVATIONS OF MOIRÉ FRINGES

Moiré fringes in TEM images were first reported early in the history of the microscope. They were used by Minter to identify a dislocation before lattice imaging was possible. Later, they were regarded as an imaging artifact that obscured the true dislocation structure in twist boundaries. Most recently, there has been renewed interest due to the widespread development of thin films grown on different substrates.

You must be wary of the limitation or pitfall of using moiré fringes to learn about interfaces and defects. Moiré patterns result purely from the interference of two 'sets of planes.' Their appearance will be essentially the same even if the two 'crystals' are not in contact.

In TEM the moiré patterns correspond to interference between a pair of beams, $\mathbf{g}_1$ and $\mathbf{g}_2$. If $\mathbf{g}_1$ is generated in the upper crystal and $\mathbf{g}_2$ in the lower, then each reflection $\mathbf{g}_1$ in crystal 1 acts as an incident beam for the lower crystal and produces a 'crystal-2 pattern' around each $\mathbf{g}_1$ reflection, as shown in Figure 23.7A. This process is another example of double diffraction, as discussed in Section 18.9. Figure 23.7A is from a pair of perfectly aligned but misfitting cubic crystals viewed along their common [001] zone axis; the pattern is indexed in Figure 23.7B. When we have many planes diffracting at a zone axis, as in this pattern, we expect to see crossed moiré fringes.

In the following three sections we will discuss examples of the use of moiré fringes.

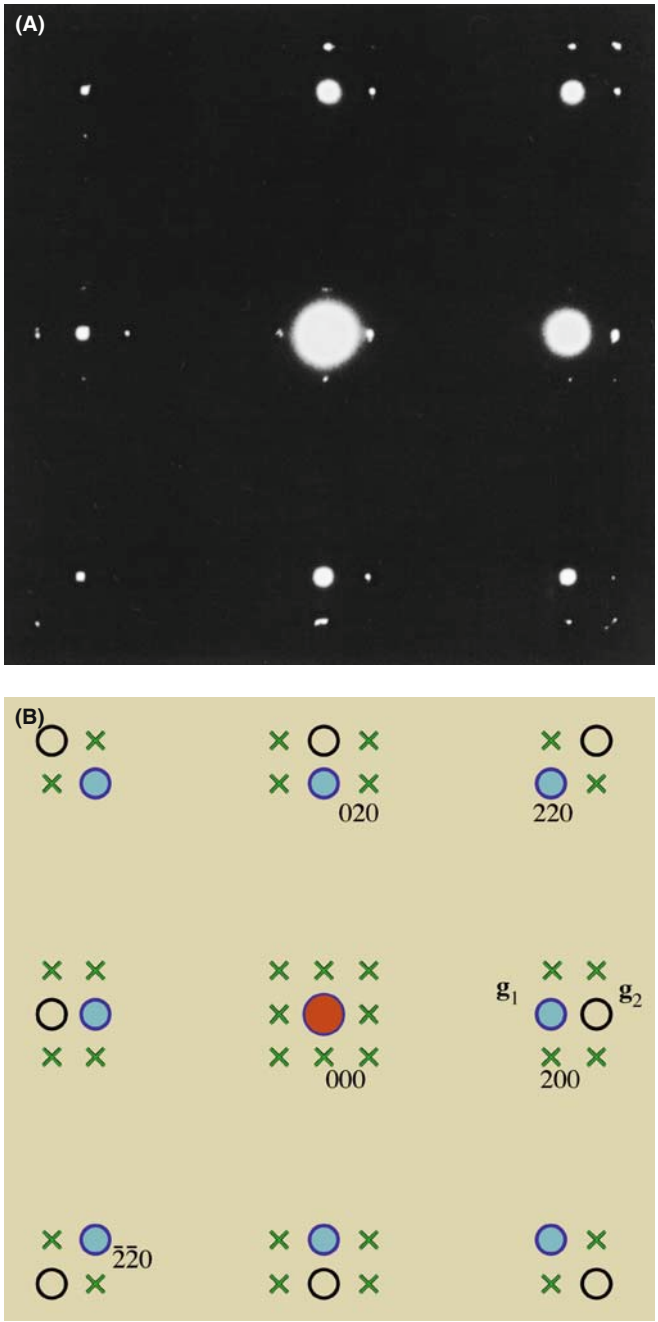


FIGURE 23.7. (A) Experimental diffraction pattern from perfectly aligned Ni and NiO. Brighter spots are from NiO, which has the larger lattice parameter. (B) Schematic which explains translational moiré fringes. Closed circles ($\mathbf{g}_1$) correspond to crystal 1, open circles ($\mathbf{g}_2$) to crystal 2 and × double diffraction of $\mathbf{g}_1$ beams by crystal 2. Only × reflections close to $\mathbf{g}_1$ and $\mathbf{g}_2$ have appreciable intensity.

23.6.A Translational Moiré Patterns

When a continuous film is grown on a thick substrate, one question which is asked is: "Do the lattice parameters of the thin film correspond to the values of the same material in bulk form?" For example, a thin film of a cubic material on an (001) substrate may be tetragonally strained so that the a_{film} lattice parameter is smaller than a_{bulk}, but the c_{film} parameter is larger. If the bulk material has its bulk lattice parameter, then the measurement of d_{tm}, the translational moiré spacing, can give a very accurate value for a_{film}. Furthermore, we can tilt the specimen 45° or 60° and deduce a value for c_{film} to estimate the tetragonal distortion directly.

Tilting the specimen can also give us information about misfitting islands, as illustrated in Figure 23.8. In this case, we see a hexagonal array of fringes when the two pseudo-hexagonal materials are viewed parallel to their common *c*-axis. The variation in the contrast of the moiré fringes around the edge of the particle occurs because the particle facets on inclined planes, as is confirmed when we tilt the specimen. In this system, when the islands are grown on different substrates, they may still grow as platelets. In Figure 23.8B and C the platelet is thick in the direction of the beam but, when tilted over, we again see moiré fringes. In particular, we can see moiré fringes at the *top* of the platelet.

INTERFERENCE NOT STRUCTURE

We know that the top of an inclined island is not in contact with the substrate yet it shows fringes; so this reminds us that moiré fringes do not tell us about the interface structure!

23.6.B Rotational Moiré Patterns

We often see rotational moiré fringes at twist boundaries, as illustrated for Si in Figure 23.9. A complicating factor is that the misfit may be accommodated by an array of dislocations having a periodicity which is related to the moiré-fringe spacing. The periodic strain field from the dislocations is, of course, only present if the two materials are in intimate contact. This image is actually recorded under rather special conditions, which we will discuss in more detail in Chapter 27.

23.6.C Dislocations and Moiré Fringes

Since the moiré pattern can often be thought of as a magnified view of the 'structure' of the materials, such

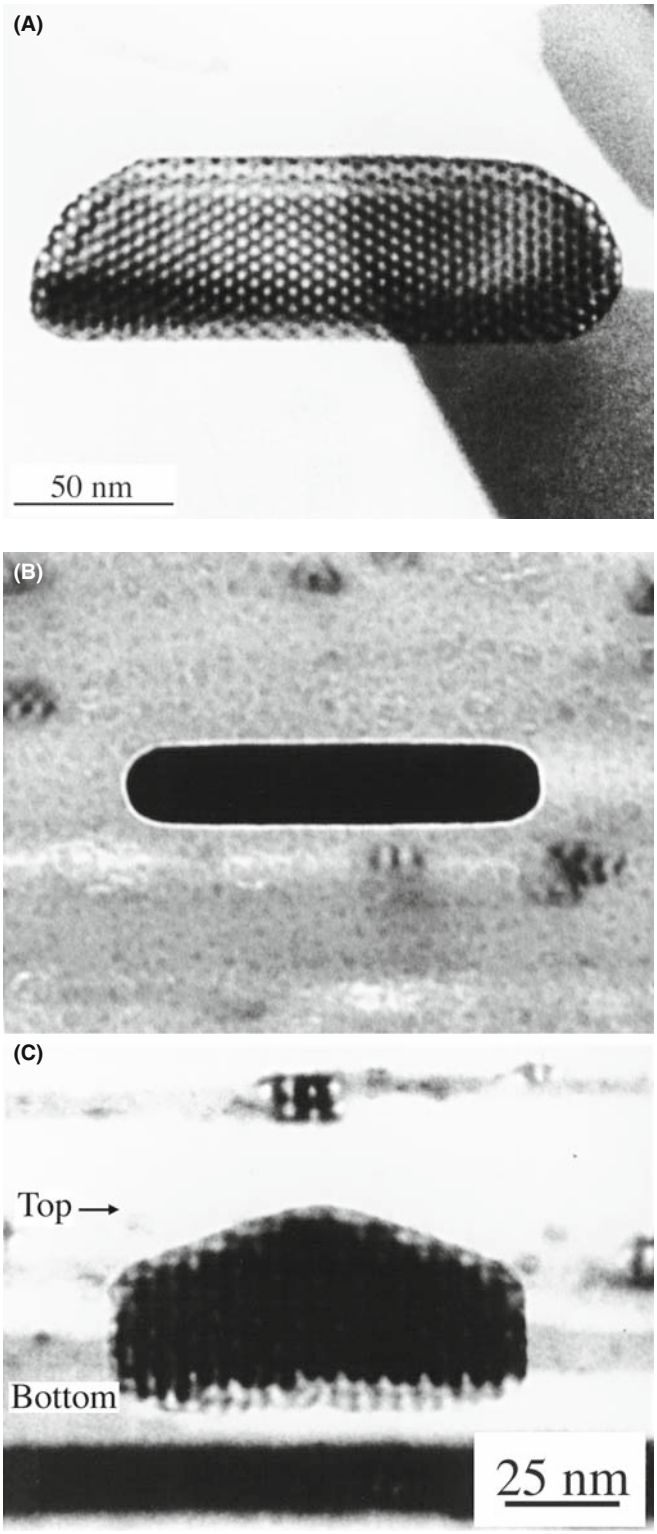


FIGURE 23.8. (A) The appearance of moiré fringes depends on the thickness of the specimen, as you can see where the edges of this island (Fe_2O_3) are inclined relative to the surface of the substrate (Al_2O_3). (B) The particle is too thick to show moiré fringes when edge on. (C) When this thick particle is tilted over, moiré fringes are seen at both the top and the bottom.

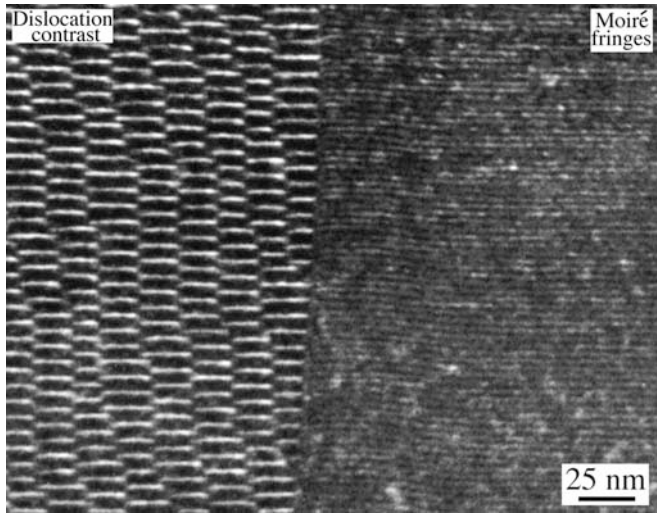


FIGURE 23.9. WBDF image of moiré fringes at a grain boundary showing very different contrast to the region containing dislocations.

patterns can be used to locate and give information on dislocations which are present in one material but not the other (but "which?" is the question). We can form an image which contains information on the dislocation if it is associated with a terminating lattice plane in one material, but we don't actually 'see' the dislocation. This effect is illustrated in Figure 23.10; the image appears as a magnified view of the projection of the dislocation.

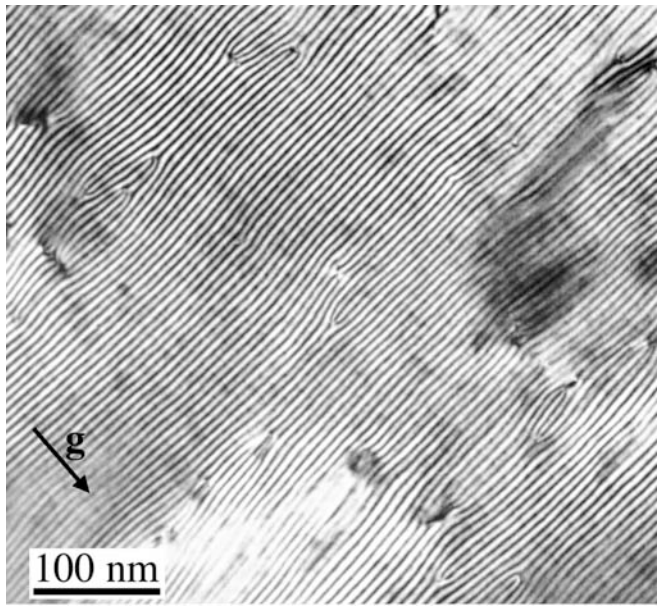


FIGURE 23.10. Moiré fringes reveal the presence of dislocations in a thin film of CoGa grown on a GaAs substrate. The (001) interface lies parallel to the specimen surface. Although the images contain much detail, most of it cannot readily be related to the structure of the defects.

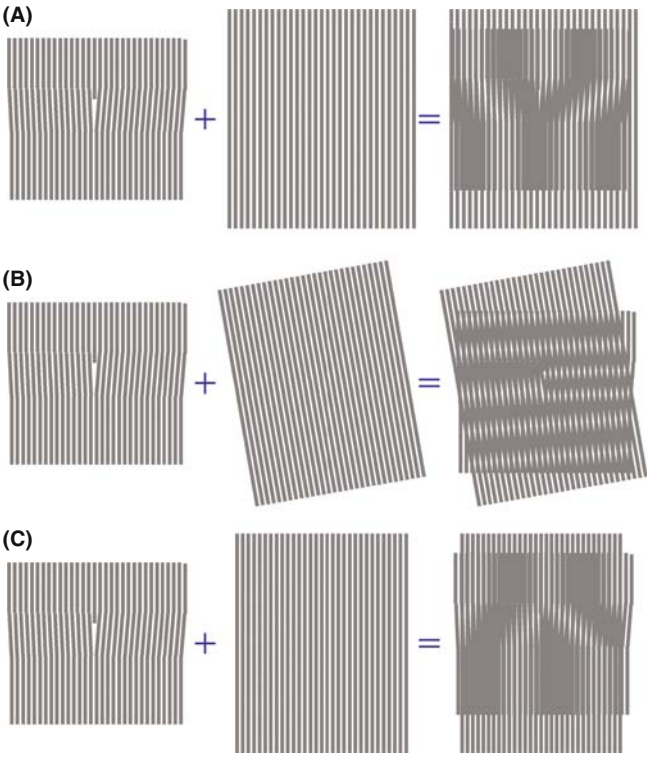


FIGURE 23.11. Schematic diagrams showing why moiré patterns from regions containing dislocations cannot be readily interpreted: (A) a dislocation image formed by interference between a regular lattice and one containing an extra half-plane. (B) In comparison with (A), a small rotation of the lattice of either grain can cause a large rotation of the dislocation fringes. (C) A small spacing change of either lattice can cause the dislocation image to reverse.

This result can be deceptive, as you can see in Figure 23.11, where we have rotated the perfect grain slightly and changed the dislocation fringe spacing.

The images can always be related directly to the projected Burgers vector of the dislocation, but you must know which planes give rise to the fringes. So make some models and experiment.

This analysis even works if you have two or more terminating fringes, but don't put too much emphasis on the actual location of the fringes. Remember, the dislocation may not be parallel to the beam. Moiré fringes may be related to a dislocation in the plane of the interface, since these locally relax the misfit. One example of such an application comes from the work of Vincent who showed that as Sn islands grew on a thin film of SnTe, the moiré-fringe spacing around the perimeter of the islands gradually increased. Suddenly, the strain at the interface was so large that a dislocation was nucleated to relax the strain and the process began again. The analysis of the changes in moiré-fringe spacing is shown in Figure 23.12.

Since the spacing of moiré fringes essentially gives a magnified view of the misfit between aligned particles

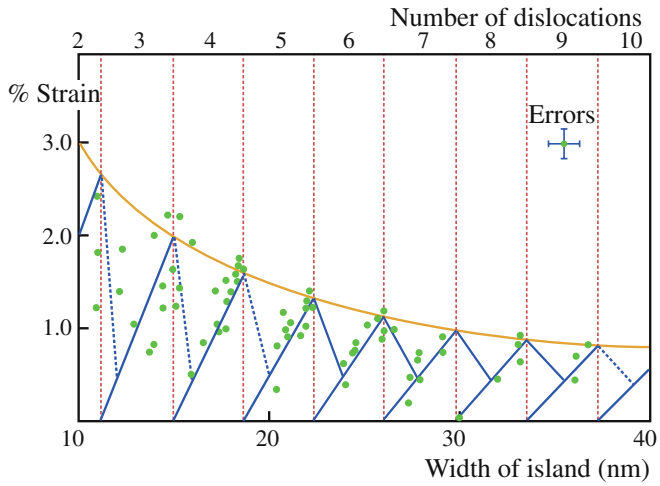


FIGURE 23.12. Moiré-fringe spacings can be used to monitor the change in lattice parameter as small islands of Sn grow in size on a thin film of SnTe. This plot shows how the strain (measured from the moiré-fringe spacing) can be related to the width of the misfitting island and then to the number of dislocations in the interface.

and a substrate, we can use them to measure the strain in such particles. In its simplest form, in one dimension, the strain is given by

$$\varepsilon = \frac{a_1 - a_0}{a_0} \tag{23.14}$$

where a_1 and a_0 are the lattice spacings of the particle and the substrate, respectively. You may need to modify this equation if the alignment is not simple cube-on-cube.

23.6.D Complex Moiré Fringes

Since moiré fringes can occur whenever $\Delta\mathbf{g}$ is small enough to be included in the objective aperture, we can have a situation where the relative rotation is rather large (45° or even 90°) so that $\mathbf{g}_1$ and $\mathbf{g}_2$ correspond to different sets of planes. This is illustrated in Figure 23.13 for YBCO grains rotated 45° on an MgO substrate. You can see that, as a bonus, the moiré fringes allow you to locate the 45° boundaries directly. Small rotations of the diffracting planes cause small rotations of $\mathbf{g}$ but large rotations of $\Delta\mathbf{g}$.

Two overlapping lattices produce a pattern of interference fringes that is much coarser than the original pattern and is very sensitive to differences in lattice spacing and relative orientations. We can use this sensitivity to provide an optical method for examining small rotations or lattice-parameter differences in HRTEM images. Make transparencies of the 'distorted' image such as that shown in Figure 23.14A and a reference lattice; the reference lattice could be the perfect-crystal image or a template you have created on the computer. Now overlay the two and rotate/translate them relative

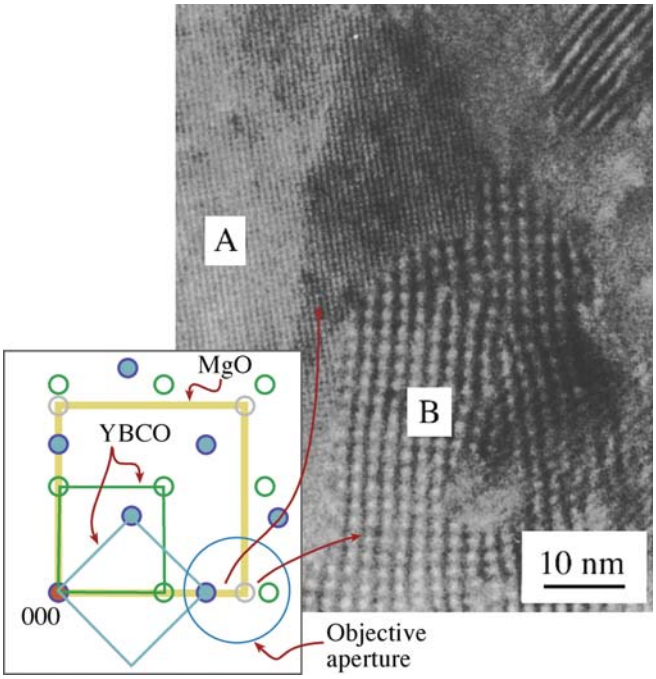


FIGURE 23.13. Moiré fringes formed when grains of YBCO grown on a single crystal of MgO are aligned to the substrate (B) or rotated through 45° (A); the spacing of the fringes is different so the position of the grain boundary can be identified. The circle in the DP shows the spots that cause the fringes. Small rotations of the fringe away from the perfect alignment are exaggerated because the spots are close together.

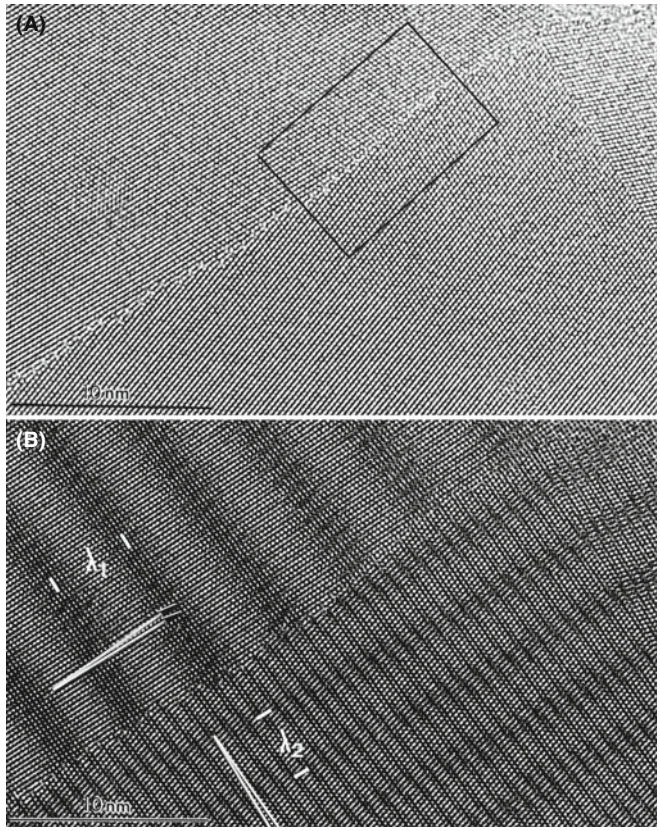


FIGURE 23.14. The use of 'artificial' moiré fringes to analyze a special grain boundary in Al. (A) An experimental image. (B) The same image overlaid with a perfect-crystal lattice transparency, producing moiré fringes of different spacings, λ_1 and λ_2.

to one another. You will have created a new artificial moiré image similar to that shown in Figure 23.14B, which was formed for a special grain boundary in Al by Hetherington and Dahmen. This boundary is special because one set of {111} planes in the upper grain is nearly normal to one set in the lower grain. How near is near? Hetherington and Dahmen overlapped their experimental image with a template that was drawn to have two sets of lines normal to one another. Overlaying the two images gave moiré fringes, which were not quite perpendicular to one another. Careful measurements of the rotation and fringe spacing showed that the fringes in the experimental image were actually 89.3° apart, not 90°.

23.7 FRESNEL CONTRAST

We saw in Chapter 9 that we can use Fresnel-contrast images of holes in carbon films to correct the astigmatism of the objective lens. We'll now discuss how we can use this same contrast mechanism to learn more about particular features in the specimen. In the classic demonstration of Fresnel contrast using visible light, bright fringes can appear in the geometric shadow of an opaque mask or dark fringes can appear in the illuminated region. The complication introduced in the TEM version is that the 'mask' is not opaque but simply has a different inner potential. Therefore, in any situation where the inner potential changes abruptly, we can produce Fresnel fringes if we image that region out of focus. Since we still focus the objective lens on a plane that is close to the specimen, we are in the near-field or Fresnel regime. Since we often study lines, planes, or platelets by this technique, we'll often see Fresnel fringes.

23.7.A The Fresnel Biprism

We can demonstrate a particularly simple interference phenomenon by placing a wire at a position F on the optic axis, as shown in Figure 23.15A. Since the beam is narrow, the wire should be less than 1 μm in diameter and can be made of a drawn glass fiber coated with Cr or Au. If we apply ~10 V to the wire, it will bend the electron beam on either side in opposite directions. The resulting interference fringes can be recorded on photographic film, as shown in Figure 23.15B, or on the CCD. The wire here is acting as a beam splitter; we'll encounter it again when we discuss holography in Section 29.11. The visible-light analog is the prism. Notice how the wire acts to produce two virtual sources s_1 and s_2, which are D_s apart. Horiuchi gives

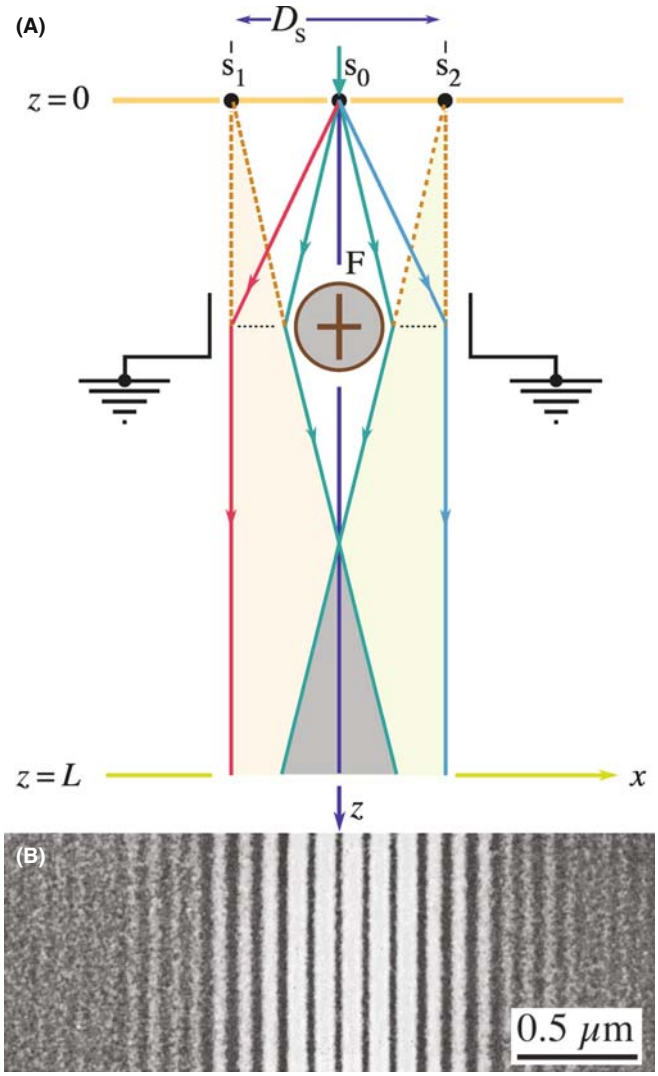


FIGURE 23.15. (A) A Fresnel biprism formed using a charged wire placed in the path of the beam; (B) the resulting interference fringes in the image.

the following equation to define a measure of the degree of spatial coherence, γ, which, as we discussed in Section 5.2 and Figure 5.13, is a function of the source size

$$\gamma = \frac{I_{\text{Max}} - I_{\text{Min}}}{I_{\text{Max}} + I_{\text{Min}}} \qquad (23.15)$$

Here I_{Max} is the intensity of the central fringe and I_{Min} is the intensity of the first minimum in Figure 23.15B.

23.7.B Magnetic-Domain Walls

Although we'll discuss imaging magnetic materials in Chapter 29, it is appropriate here to consider briefly the similarity of Lorentz microscopy of magnetic-domain walls to other interference images. We know from our discussion of magnetic fields in the electron lens in Chapter 6 that the Lorentz force acting on an electron with velocity **v** is proportional to $\mathbf{v} \times \mathbf{B}$. If the sign of **B** is opposite in two adjacent domains, then the electrons will be deflected in opposite directions as shown in Figure 23.16. The 'converging' domain wall is remarkably similar to the electron interferometer in the previous section. We can indeed produce a series of interference fringes. You should consult the original analysis of Boersch et al.; the basics are given in the book by Hirsch et al. who show that we can calculate the fringe spacing Δx using

$$\Delta x = \frac{\lambda(L + \ell)}{2\ell\beta_{\text{m}}} \qquad (23.16)$$

where β_{m} is the angle of deflection of the beam, λ is the electron wavelength, ℓ is the 'source'-to-specimen distance and L is the specimen-to-'detector' distance. Quotation marks are used to emphasize that these are 'effective' distances like the 'camera length.' The value of Δx can be ~20 nm. You'll only see such interference fringes if you form the image using parallel illumination. We'll return to the imaging of magnetic specimens in Chapter 29.

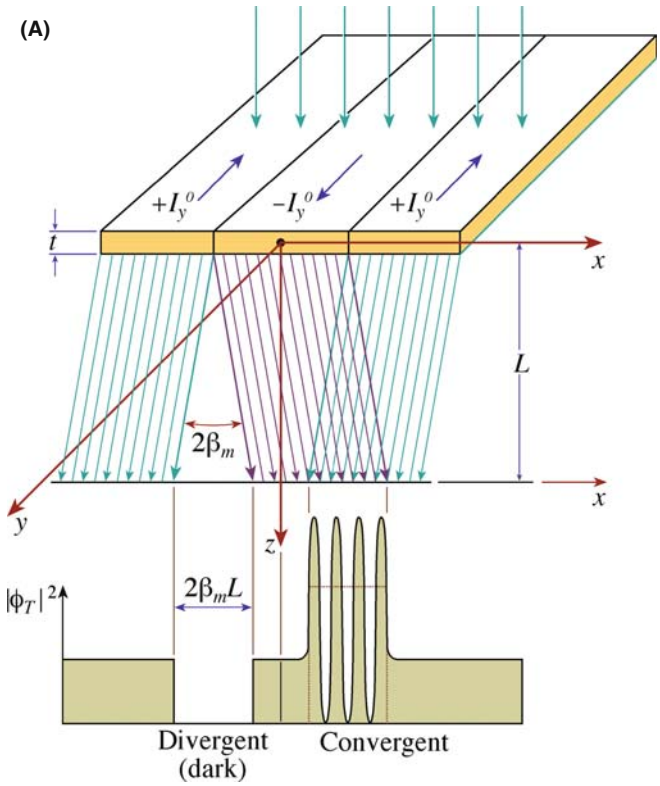


FIGURE 23.16. (A) Deflection of the electron beam by magnetic-domain walls; compare with Figure 23.15A. (B) Experimental interference fringes from one such wall; compare with Figure 23.15B.

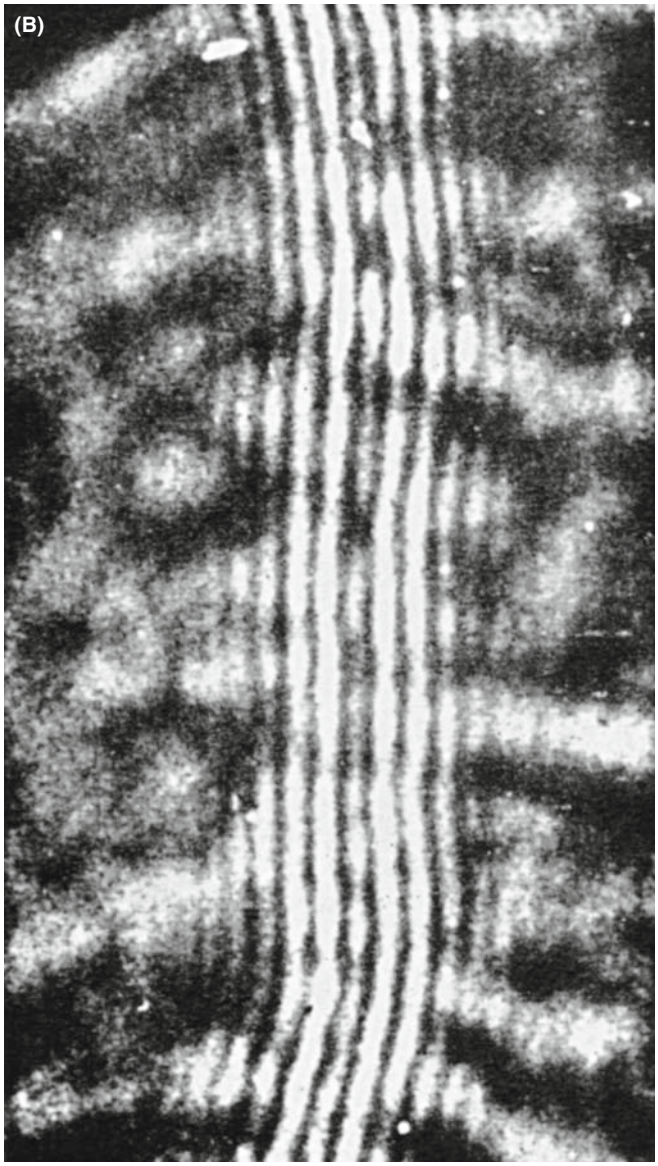


FIGURE 23.16. (Continued).

23.8 FRESNEL CONTRAST FROM VOIDS OR GAS BUBBLES

You might think that it would be difficult to image voids or small gas-filled cavities when there is no associated strain field, because voids or cavities do not scatter electrons. However, we can image holes that are fully enclosed inside the specimen by defocusing the image and observing the special phase contrast; this is the Fresnel contrast that we introduced back in Sections 2.9 and 9.5. In principle, we can apply this technique to holes that contain a liquid or even a solid (i.e., a second phase). In the latter case, however, the Fresnel contrast is likely to be hidden by strain contrast in the specimen.

You can image small voids or gas bubbles in two ways

- By orienting the region of interest so that $\mathbf{s} = 0$; the cavity then reduces the 'thickness' of material locally.
- By using Fresnel contrast.

CAUTION

Small particles can give similar contrast to small voids. The Fresnel contrast can easily be misinterpreted as a core-shell structure!

In the Fresnel technique, the image shows contrast whenever the objective lens is not focused on the bottom surface of the specimen.

Wilkens expressed the wave function as

$$\psi(t, \mathbf{r}') = \psi_0(t)[1 + \Delta_r(\mathbf{r}') + i\Delta_i(\mathbf{r}')] \tag{23.17}$$

Here $\psi_0(t)$ is the wave function in the absence of the cavity; Δ_r and Δ_i are real functions which depend on

- The location and dimensions of the cavity.
- The extinction distance and absorption parameter of the matrix (ξ_g and ξ_g').
- The potential difference, ΔV, between the inner potential of the matrix, V_0, and that of the cavity, V_c (it could be filled or empty).

FRESNEL FOCUS?

Fresnel-contrast images are always out of focus.

In the case of thick foils where z_c, the size of the cavity in the direction of the beam, is $<0.1\xi_g$, the wave function can be expressed as

$$\psi(t, \mathbf{r}') = \psi(t)[1 + i\Delta_i(\mathbf{r}')] \tag{23.18}$$

where Δ_i (using $w = s\xi_g$) is given by

$$\Delta_i = -\left(2\varepsilon_0 - \frac{1}{\varepsilon_g}\frac{1}{(1+w)^{1/2}}\right) z_c(\mathbf{r}')p_i(z) \tag{23.19}$$

The difference in inner potential is included in ε_0, which is defined by the equation

$$\varepsilon_0 = -\frac{\Delta V}{E}k \tag{23.20}$$

Here, k is the magnitude of the wave vector and E is the energy of the electron beam. When the thickness dependence is damped out (the foil is thick), the intensity can be expressed quite simply as

$$|\psi(t,\mathbf{r}')|^2 = |\psi_0(t)|^2\left(1+\Delta_i^2\right) \qquad (23.21)$$

We can summarize some results from this analysis

- When the image is in focus, the cavity is invisible so we view it out of focus to observe the Fresnel-fringe contrast.
- The contrast depends on the difference in the inner potential of the matrix and the cavity; we usually see the most contrast if the content of the cavity is vacuum, because then ε_0 is greatest.
- The contrast does depend on the wavelength of the electrons through both k and E.
- Cavities as small as 1–2 nm in diameter can be imaged using Δf values of 0.5–1.0 μm.
- In the case where $w = 0$ and $2\varepsilon_0 > \xi_g^{-1}$ (so Δ_i is < 0), if $\Delta f < 0$, the image is a bright dot surrounded by a dark fringe; if $\Delta f > 0$ the dot is dark and the fringe is bright.
- This is the same behavior as we saw in Figure 9.21, where we had a dark fringe at underfocus and a bright fringe at overfocus.

The contrast is illustrated in Figure 23.17. You should note that it is not the same as the black-white contrast from small precipitates that result from strain fields as we'll discuss in Chapter 26. You'll find a more detailed analysis in the article by Rühle and Wilkens, but you can already appreciate the relevance for imaging nanoparticles.

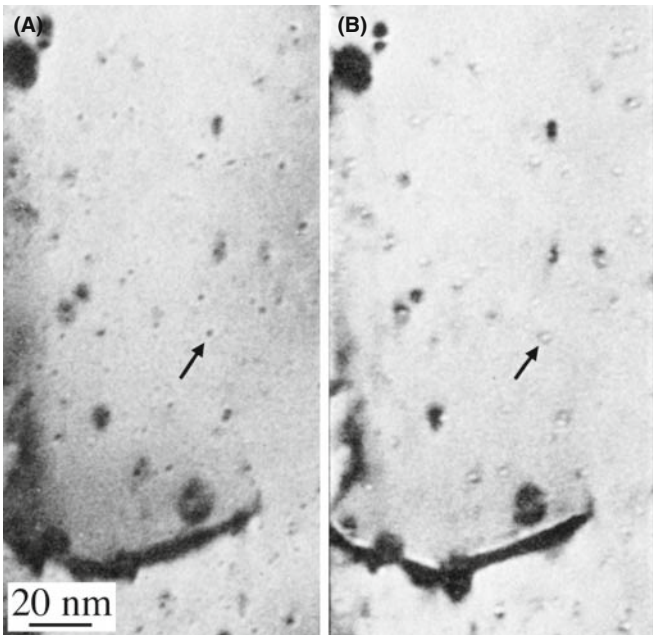


FIGURE 23.17. Fresnel contrast from He bubbles in Au. (A) Overfocus image. (B) Underfocus image.

23.9 FRESNEL CONTRAST FROM LATTICE DEFECTS

This topic is one that is receiving more attention as computers and simulation programs become, respectively, more powerful and more user-friendly. The reason for this increased attention is clear, as Bursill et al. showed in their pioneering studying of Fresnel fringes from edge-on defects. They demonstrated that if you take great care in determining all the electron-optical

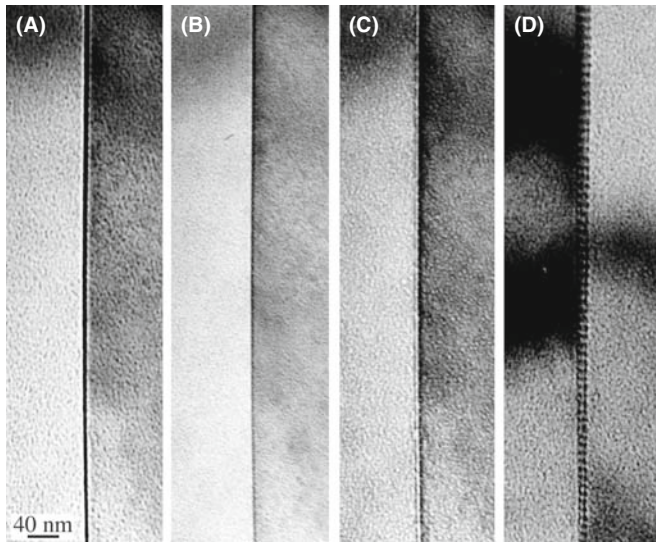


FIGURE 23.18. (A–D) A through-focus series of images from an edge-on GB showing the changes in Fresnel contrast. The image in (D) shows the boundary tilted over to reveal its periodic structure more clearly.

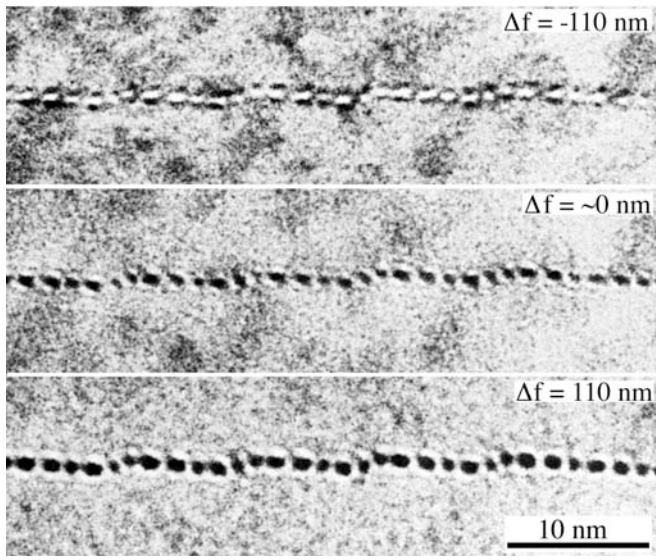


FIGURE 23.19. Series of experimental images recorded at different values of Δf for a low-angle grain boundary in NiO. Each white or black spot corresponds to one end-on dislocation.

parameters (particularly what defocus steps you are using), you can obtain new information on edge-on defects. The defect images are very sensitive to the model used to simulate them, but you will need other information for a full analysis. We'll refer you to their paper on the {100} platelets in diamond and concentrate on two more widely applicable situations, namely, end-on dislocations and edge-on grain boundaries; in both cases, we now have many techniques, such as XEDS, EELS, and HRTEM, to complement the Fresnel-fringe studies.

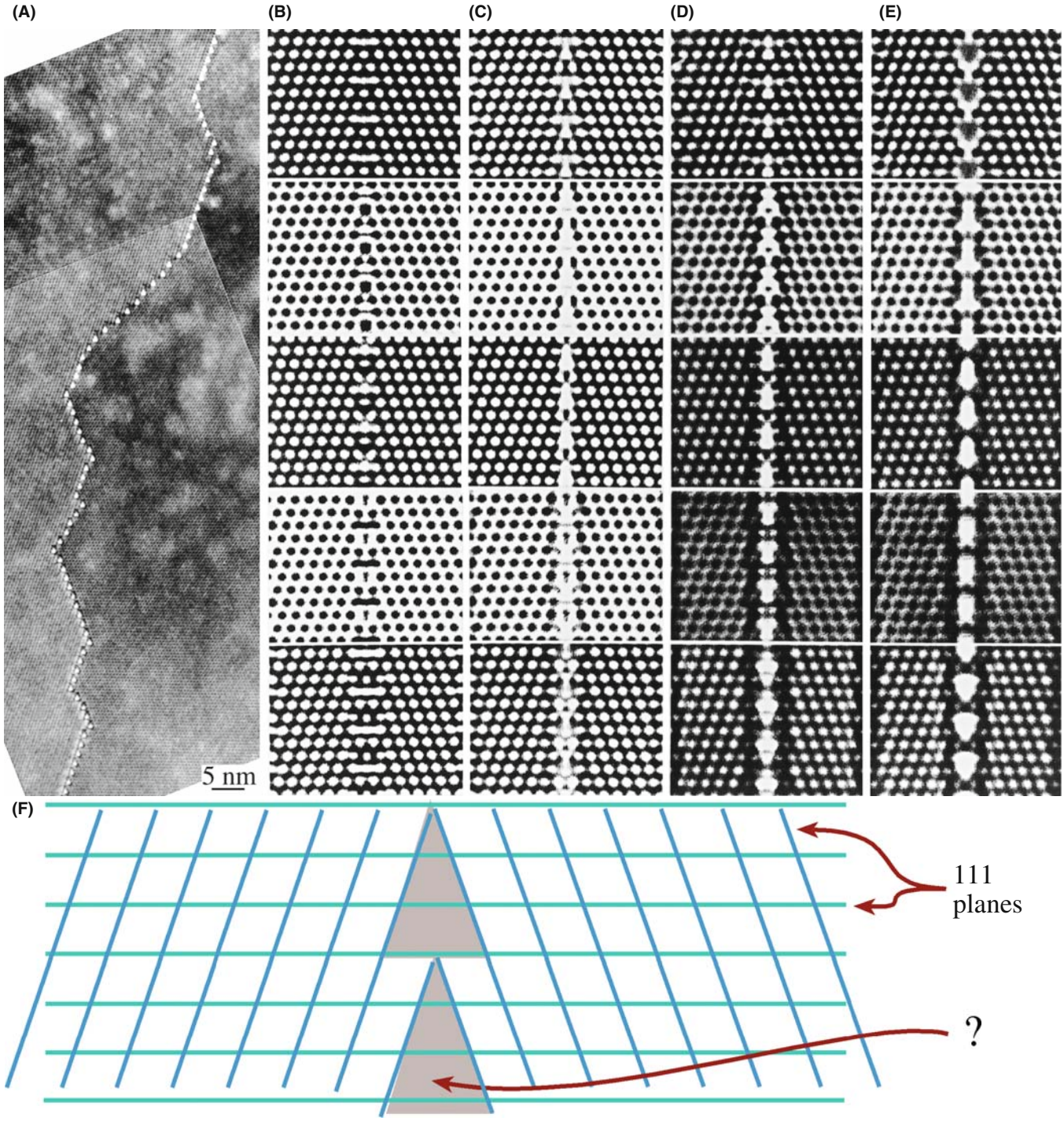


FIGURE 23.20. The structure of the (112) lateral twin boundary in a thin-foil of spinel consists of triangular prisms with a density lower than the bulk crystal as shown in the diagram at the bottom. (A) Fresnel contrast is seen when these prism-like defects are imaged out of focus. (B–E) Simulated images; each column is a different model: (B) no ions removed, (C) ion sites in the prisms half occupied, (D, E) all ions removed. The defocus in each row increases from the top (–10, –70, –130, –160, –210 nm), and the thickness is 5.7 nm. (F) Schematic of the prism 'tubes.'

23.9.A Grain Boundaries

We might expect almost any grain boundary to show a localized change in the inner potential. However, following the original suggestion by Clarke (1979), Fresnel-contrast imaging has been used most extensively to study those interfaces which are thought to contain a thin layer of glass. Part of the reason for this emphasis is simply that other techniques tend to give ambiguous results for such interfaces.

When you use the Fresnel-fringe technique to study grain boundaries or analyze intergranular films, you must orient the boundary in the edge-on position so that you can probe the potential at the boundary. Later, in Section 30.11, we will consider the actual shape of this 'potential well.'

In a real TEM specimen, the specimen thickness is likely to change at the grain boundary even if only by a nanometer or so. Since the specimen will be quite thin, this change can give an appreciable contribution to the difference in the 'effective inner potential' seen by the electron beam. You can defocus the image to see the Fresnel contrast shown in Figure 23.18.

The Fresnel-contrast technique can equally well be applied to phase boundaries, with perhaps the most thoroughly studied example being the Si/SiO_2 interface. Since the details of the contrast are sensitive to the abruptness of the change in the inner potential, the technique can also produce information on this aspect of the interface. Nevertheless, you must always look for associated changes in the real geometry that can occur even when you're just forming Fresnel fringes from the edge of the specimen.

23.9.B End-On Dislocations

We've just seen that we can detect Fresnel-fringe contrast from edge-on high-angle grain boundaries. We might then ask: "is it possible to detect similar contrast from low-angle grain boundaries, i.e., grain boundaries which consist of arrays of distinct dislocations?" It is indeed possible, as shown in the series of images from a tilt boundary in NiO in Figure 23.19. This through-focus series, and images of other grain boundaries, was analyzed by assuming that there is a change in $\Delta V(\mathbf{r})$ in the mean inner potential at the core of the dislocation. Two models were proposed for $\Delta V(\mathbf{r})$. In model 1, when $r < r_0$

$$\Delta V(\mathbf{r}) = \Delta V_0 \left\{ 1 - e^{\frac{-(r - r_0)}{a}} \right\} \tag{23.22}$$

but when $r > r_0$

$$\Delta V(\mathbf{r}) = 0 \tag{23.23}$$

The constant a is $\sim 0.1 r_0$. In model 2

$$\Delta V(\mathbf{r}) = \Delta V_0 \exp\left(-r^2/r_0^2\right) \tag{23.24}$$

In both cases, ΔV_0 is negative. As an example of the quantities involved in these equations, if the Burgers vector of the dislocations is $^1/_2$ [110], Rühle and Sass found that $\Delta V_0 = 0.09 V_0$ for $r_0 = 3.2$ Å. They could not distinguish between the two models for $\Delta V(\mathbf{r})$ but two clear points come out of this study

- You must know the inclination of your foil surface. If the lower surface is inclined to the horizontal, then thicker parts of the specimen can be much closer to the objective lens than in the thin area; you can do a quick calculation to prove this point.
- The inner potential at a dislocation core is not the same as the bulk value. You should expect the value of ΔV_0 to be influenced by a change in stoichiometry or impurity segregation.

Before leaving this topic, we should point out that the inner potential at the grain boundary may not be uniform, perhaps because the width of the interface varies or the interface facets on a mesoscopic scale. You can still see Fresnel effects which relate to the periodicity in the grain boundary even if this periodicity is not associated with dislocations. A particularly clear example of such a variation is shown in Figure 23.20, where a twin boundary in spinel is essentially constructed of parallel triangular tubes; the inner potential inside the tube is much lower than the matrix value and the tubes are only about 1.2 nm high.

CHAPTER SUMMARY

Phase contrast will occur whenever we have more than one beam contributing to the image. The clue is: if you see fringes (periodic contrast) of any sort, then you are almost certainly observing a phase-contrast image. This conclusion even applies to thickness fringes (Chapter 24) and stacking-fault fringes (Chapter 25) in what are traditionally called two-beam diffraction-contrast images.

Phase-contrast images are widely used in three forms

- Images which relate directly to the structural periodicity of the crystalline specimen
- Moiré-fringes
- Fresnel-contrast

It is even possible for an image to show all three effects at the same time. So you must remember that phase-contrast effects don't just occur when you are forming high-resolution images. You will create Fresnel contrast whenever your specimen is thick or you are working out of focus. We should note that it is difficult, but in principle not impossible, to be quantitative in your analysis of Fresnel fringes.

The usefulness of moiré fringes continues to surprise even experienced users of the TEM. However, you still have to exercise caution when interpreting what they are telling you about defects in your material.

The appearance of the Fresnel image varies with small changes in the thickness, orientation, or scattering factor of your specimen, and variations in the focus or astigmatism of the objective lens.

REFERENCES

We encourage you to delve into the early papers on moiré fringes. Search the literature for those by Pashley, Stowell, et al. on Au islands on mica.

The books by Hirsch et al. (1977), Horiuchi (1994), and De Graef (2003) are useful references. That by Spence (2003) is essential follow-up reading (and described in Chapter 1).

SPECIAL TOPICS

Boersch, H, Hamisch, H, Wohlleben, D and Grohmann, K 1920 Z. Phys. **159** 397–404. The original analysis of interference fringes from domain walls.

Bursill, LA, Barry, JC, Hudson, PRW 1978 *Fresnel Diffraction at {100} Platelets in Diamond* Phil. Mag. **A37** 789–812.

Heavens, OS and Ditchburn, RW 1991 *Insight into Optics* p 73 John Wiley & Sons New York. Gives a clear treatment of Fresnel fringes.

Hetherington, CJD and Dahmen, U 1992 *Scanning Microscopy Supplement* **6** p 405 Scanning Microscopy International AMF O'Hare IL.

Krivanek, OL and Rez, P 1980 Proc. 38th Ann. EMSA Meeting p 170 Ed. GW Bailey Claitors Baton Rouge LA.

MOIRÉ FRINGES

Menter, JW 1956 *The Direct Study by Electron Microscopy of Crystal Lattices and Their Imperfections* Proc. Roy. Soc. (London) **A236** 119–135. Classic early work.

Norton, MG and Carter, CB 1995 *Moiré Patterns and Their Application to the Study of the Growth of* $YBa_2Cu_3O_{7-\delta}$ *Thin Films* J. Mater. Sci. **30** 381–389. Islands of YBCO.

Rühle, M and Wilkens, M 1975 *Defocusing Contrast of Cavities; Theory* Cryst. Lattice Defects **6** 129–400. Voids—the definitive work.

Vincent, R 1969 *Analysis of Residual Strains in Epitaxial Thin Films* Phil. Mag. **19** 1127–1139. Showing the magical size effect using moirés.

Wilkens, M 1975 in *Electron Microscopy in Materials Science* **II** p 647 Eds. U Valdré and E Ruedl CEC Brussels. A clear explanation of the use of Fresnel fringe theory to analyze voids.

FRESNEL FRINGES, SURFACES, AND LAYERS OF GLASS

Clarke, DR 1979 *On The Detection of Thin Intergranular Films by Electron Microscopy* Ultramicrosc. **4** 33–44.

Fukushima, K, Kawakatzu, H and Fukami, A 1974 J. Phys. **D7** 257. Just surfaces.

Longworth, S 2006 Ph.D. Thesis, Cambridge University. Concludes that the Fresnel fringe technique is not reliable for measuring amorphous films.

Ness, JN, Stobbs, WM and Page, TF 1986 *A TEM Fresnel Diffraction-Based Method for Characterizing Interfacial Films* Phil. Mag. **54** 679–702.

Rasmussen, DR and Carter, CB 1990 *On the Fresnel-Fringe Technique for the Analysis of Interfacial Films* Ultramicrosc. **32** 337–348.

Rasmussen, DR, Simpson, YK, Kilaas, R and Carter, CB 1989 *Contrast Effects at Grooved Interfaces* Ultramicrosc. **30** 52–55.

Simpson, YK, Carter, CB, Morrissey, KJ, Angelini, P and Bentley, J 1986 *Identification of Thin Amorphous Films at Grain-Boundaries in* Al_2O_3 J. Mater. Sci. **21** 2689–2696. Discussion of different techniques for characterizing these layers.

FRESNEL FRINGES INTERFACES

Carter, CB, Elgat, Z and Shaw, TM 1987 *Lateral Twin Boundaries in Spinel* Phil. Mag. **55** 21–38. Tubes along GBs.

Ross, FM and Stobbs, WM 1991a *Study of the Initial Stages of the Oxidation of Silicon Using the Fresnel Method* Phil. Mag. **A63** 1–36.

Ross, FM and Stobbs, WM 1991b *Computer Modelling for Fresnel Contrast Analysis* Phil. Mag. **A63** 37–70. The inner potential.

Rühle, M and Sass, SL 1984 *Detection of the Change in Mean Inner Potential at Dislocations in Grain Boundaries in NiO* Phil. Mag. **A49** 759–782. Individual dislocations with different Burgers vectors show different Fresnel contract.

Tafto, J, Jones, RH and Heald, SM 1986 *Transmission Electron Microscopy of Interfaces Utilizing Mean Inner Potential Differences Between Materials* J. Appl. Phys. **60** 4316–4318.

THE COMPANION TEXT

The companion text includes several chapters that build on phase contrast as it applies to HRTEM and holography.

SELF-ASSESSMENT QUESTIONS

Q23.1 What are the differences between electron diffraction and X-ray diffraction?

Q23.2 Deduce a simple expression for the intensity when both beams are admitted through the objective aperture in the two-beam geometry.

Q23.3 What is the key advantage of using the three-beam condition where **g** and are equally strongly excited and O stays on the optic axis?

Q23.4 Might you see a different fringe periodicity if you use the three-beam rather than two-beam condition?

Q23.5 If you want to resolve the dumbbells in Si, what reflections must you include when forming the image?

Q23.6 Deduce the spacing for translational moiré fringes by using the DP construction.

Q23.7 Deduce the spacing for rotational moiré fringes by using the DP construction.

Q23.8 What do moiré fringes tell you about the structure between the two materials which overlap to cause the fringes?

Q23.9 The image of a dislocation seen in a moiré pattern may suggest a Burgers vector which is 90° off or even of the wrong sign. What is the most important factor to know when you interpret moiré patterns of dislocations?

Q23.10 Islands of Sn growing on a thin film of SnTe show a magic number effect. What is this effect?

Q23.11 Looking at a particular specimen in the TEM we see fringes when we are out of contrast and say this is a Fresnel image. Why?

Q23.12 Why is the Fresnel biprism so named?

Q23.13 What is Lorentz microscopy?

Q23.14 You could image a small gas bubble if the specimen were in focus or you could intentionally make the image out of focus. Why do both techniques work and which approach would you recommend?

Q23.15 What information are we obtaining when we image grain boundaries using Fresnel contrast?

Q23.16 Could we use Fresnel contrast to examine twist boundaries? (Be careful.)

Q23.17 Would you prefer underfocus or overfocus when imaging a grain boundary using the Fresnel method?

Q23.18 Approximately how large is the defocus value used when imaging end-on dislocations?

Q23.19 You could use the Cornu spiral in visible-light optics to calculate the location and intensity of fringes from a semi-infinite opaque sheet. Why is the TEM specimen more complex?

TEXT-SPECIFIC QUESTIONS

T23.1 Consider Figure 23.1B. Why does this not give a true image of the lattice planes? Hence why is 'C' better?

T23.2 Consider Figure 23.3. Why do the dumbbells in the image not correspond to those in the structure and why are they there? Give a full clear explanation.

T23.3 Consider Figure 23.4B. Why would *Z*-contrast imaging be useful here? Explain in detail.

T23.4 Consider Figure 23.5A and B. (A) If one set of these 111 planes (smaller *d*) corresponds to Si and the other set is 111 planes from a different fcc material. What is the lattice parameter of the second material. (B) If the fringes are the 111 fringes of Si, show that the rotation is consistent with the moiré fringe spacing.

T23.5 Consider Figure 23.16. Use reasonable values of the experimental parameters to suggest a spacing for the fringes in B. Justify your choice of values.

T23.6 Consider Figure 23.16. Estimate the value of *Z*. Justify your choice of other values.

T23.7 Consider Figure 23.12. Deduce a value for the dislocation periodicity and compare this value with the expected misfit for this system.

T23.8 Consider Figure 23.13. Index both DPs and then compare the observed fringe spacing with that which you would expect from the crystal parameters.
T23.9 Consider Figure 23.11. What is the misfit in A and in C? Does the observed angle of rotation in B really produce the fringes (and their spacing) in B?
T23.10 Assuming the hematite is fully relaxed, what would be the error in the magnification bar in Figure 23.8A.
T23.11 Show quantitatively with the aid of a diagram, the image you would obtain using (a) the 220 and (b) the 200 regions of the DP in Figure 23.7.
T23.12 If the InAsSb film in Figure 23.4B is fully relaxed, what is the As:Sb ratio?
T23.13 Fully describe the geometry of the grains shown in Figure 23.4C.
T23.14 Figure 23.4A actually shows two spinel/olivine interfaces. Which material is where and what are the orientations?
T23.15 In Figure 23.4B, what is the character and Burgers vectors of each dislocation?
T23.16 Confirm that the moiré fringe spacings in Figure 23.5A and B are as predicted by the formula.
T23.17 Confirm that the dumbbells in Figure 23.3A are not identical to those in Figure 23.3B and explain how they arise.
T23.18 Work through the equations leading to equation 23.9. Show quantitatively how you might see different 'lattice fringes' from the same set of lattice planes.
T23.19 Consider Figure 23.9. Deduce the **g**-vector used to form the image, the rotation angle, and explain why the spacing of the short bright lines is not the same as that of the moiré fringes.
T23.20 Through what angle was the specimen in Figure 23.8 tilted between image B and C?
T23.21 In Figure 23.10, what is **g** and what is the misfit between the two materials?
T23.22 Consider the terminating fringes near the middle of Figure 23.10. Assume the threading dislocation lies in the thin film and hence deduce its Burgers vector. If it were in the substrate, would the image be the same?
T23.23 Consider Figure 23.17. Estimate the size of the arrowed void and the difference in focus between the two images.
T23.24 If the specimen shown in Figure 23.17 were imaged in focus, what would you see in the image? Justify your answer.
T23.25 Discuss why the contrast in Figure 23.19 doesn't disappear when $\Delta f = 0$.
T23.26 Consider the simulated images in Figure 23.20. Discuss the change in each set of four images going first across, and then down, the set. Do the changes agree with your expectations?

24

Thickness and Bending Effects

CHAPTER PREVIEW

We see diffraction contrast in an image for two reasons: either the thickness of the specimen varies or the diffraction conditions change across the specimen: the *t* effect and the *s* effect!

The *thickness* effect: when the thickness of the specimen is not uniform, the coupling (interference) of the direct and diffracted beams occurs over different distances, thus producing a thickness effect. Don't confuse diffraction contrast due to thickness changes with mass-thickness contrast discussed in the previous chapter. The effects are very different. The diffraction contrast changes with small changes in tilt, but the mass-thickness contrast doesn't.

The *bending* effect: Whenever the orientation of the diffracting planes changes, i.e., when the diffracting planes tilt relative to the beam, the contrast changes. To interpret changes in image contrast we need to understand how the contrast is related to thickness and bending.

We call these two important contrast phenomena 'thickness fringes' and 'bend contours.'

The present chapter is particularly important for three reasons

- All TEM specimens are thin but their thickness invariably changes.
- Because the specimens are so thin they also bend elastically, i.e., the lattice planes physically rotate.
- The planes also bend when lattice defects are introduced.

We can see the effects of these rotations even when they are $<0.1^\circ$ since they still have a significant effect on the image. Therefore, the bending may arise because the specimen is thin (i.e., giving possible artifacts of the technique) or it may be caused by strains which were present in the bulk material. The result is that, in real specimens, bending and thickness effects often occur together.

24.1 THE FUNDAMENTAL IDEAS

To understand the origin of thickness fringes and bend contours, we limit the discussion to the two-beam situation and recall equations 13.46 and 13.47, which we derived from the Howie-Whelan equations. The intensity of the Bragg-diffracted beam is then given by equations 13.48 and 13.49 and summarized as

$$I_{\mathbf{g}} = |\phi_{\mathbf{g}}|^2 = \left(\frac{\pi t}{\xi_{\mathbf{g}}}\right)^2 \cdot \frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} = 1 - I_{\mathbf{0}} \qquad (24.1)$$

where s_{eff} is the effective excitation error in equation 13.47

$$s_{\text{eff}} = \sqrt{s^2 + \frac{1}{\xi_{\mathbf{g}}^2}} \qquad (24.2)$$

Although we will concentrate on $I_{\mathbf{g}}$ (the DF image intensity) for most of this discussion, the direct beam (BF image) behaves in a complementary manner (neglecting, for now, the effect of absorption and other diffracted beams). The diffracted intensity is periodic in the two independent quantities, *t* and s_{eff}. If we imagine the situation where *t* remains constant but *s* (and hence s_{eff}) varies locally, then we produce bend contours. Similarly, if *s* remains constant while *t* varies, then thickness fringes will result.

This chapter is simply concerned with the physical understanding of equation 24.1 and how you can relate the image to the information contained in the DP. Although these effects are often a hindrance to systematic analysis of lattice defects, they can, in certain situations, be useful. The most important reason for understanding them is that they are essentially unavoidable!

24.2 THICKNESS FRINGES

As a result of the way that we thin TEM specimens, very few of them (only evaporated thin films or ideal ultramicrotomed, cleaved, or FIBbed sections) have a uniform thickness over their entire area. A BF/DF pair of images from the same region of the specimen is shown in Figure 24.1; the thin area is generally in the form of a wedge.

Consider again equation 24.1. You should remember that, in this calculation, t is not the 'thickness' of the foil; it is actually the distance 'traveled' by the diffracted beam, but we'll usually just say 'thickness.' If we try to treat the many-beam situation rigorously, then the value of t would, in general, be different for each beam. If you are actually viewing the foil flat-on (i.e., one surface normal to the beam), then t will be close to the geometric thickness of the foil. However, it is more difficult to analyze the image thoroughly when the foil is wedge-shaped and inclined to the beam. We almost invariably make the approximation that t is fixed with the justification being that the Bragg angles are small.

TWO BEAMS

The situation is complicated by absorption and the presence of other diffracted beams: we are never truly in a two-beam situation.

Equation 24.1 tells us that intensities of both the **0** and the **g** beams oscillate as t varies. Furthermore, these oscillations are complementary for the DF and BF images, as we show schematically in Figure 24.2. You can, of course, confirm this observation at the microscope by forming the image without using an objective aperture; there is then minimal contrast when you're in focus. The intensity, I_0, of the incident beam starts equal to unity and gradually decays, while the intensity of the diffracted beam, I_g, gradually increases until it becomes unity; I_0 is then zero; the process then repeats itself.

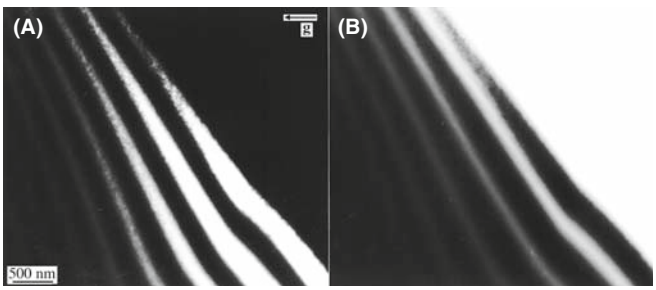


FIGURE 24.1. (A) BF and (B) DF images from the same region of a wedge-shaped specimen of Si at 300 kV tilted so that **g**(220) is strong. The periodicity and contrast of the fringes are similar and complementary in each image.

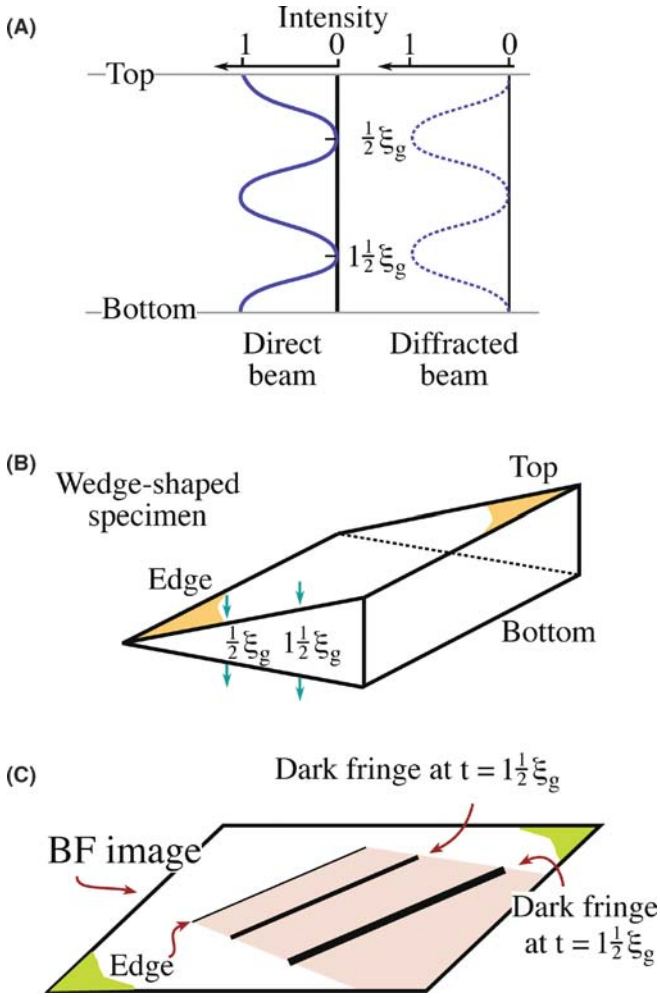


FIGURE 24.2. (A) At the Bragg condition ($\mathbf{s} = 0$), the intensities of the direct and diffracted beams oscillate in a complementary way. (B) For a wedge specimen, the separation of the fringes in the image (C) is then determined by the angle of the wedge and the extinction distance, ξ_g.

THICKNESS, ξ_g, AND ABSORPTION

As a rule of thumb, when other diffracted beams are present the effective extinction distance is reduced. At greater thicknesses, absorption occurs and the contrast is reduced.

These oscillations in I_0 or I_g are known as thickness fringes, though they are often not fringes. We sometimes call them thickness contours, because they denote the

contours where the specimen has constant thickness; you will only see these fringes when the thickness of the specimen varies locally, otherwise the contrast will be a uniform gray. As we'll see, the actual contrast can quickly change if the specimen is tilted through a small angle.

BLACK AND WHITE

It is important to realize that the image may appear to be black or white depending on the thickness of the specimen.

For example, in BF images, thicker areas are often brighter than thinner areas, which really is counter-intuitive.

Several examples of how thickness fringes might appear in your image are shown in Figure 24.3. Although it is often helpful to think of these fringes as thickness contours analogous to height or depth contours on a map, with the hole at sea level, remember that there are *two* surfaces to the TEM specimen. A DF image will usually appear to give greater contrast. This is partly because the hole now appears dark but also because many-beam effects are less important in DF. In Figure 24.3A, the narrow fringe pattern in this DF image is due to the grain boundary region being thinner than the matrix. In the DF image in Figure 24.3B, the reflection used to form this image is only excited in the right grain so the left grain is black; the diffracting grain exhibits strong thickness fringes in the regions where there are microtwins. This image introduces the idea that images of defects can also show thickness effects.

In Figure 24.3C, the specimen is an almost flat, parallel-sided film of MgO with holes formed by preferential chemical etching at defects; after thinning, the surfaces were faceted by heating the specimen at 1400°C. The holes in the image are white, so it is a BF image. The contours, like the holes, are angular because of the faceting, but they are not uniformly spaced because it's not a uniform wedge. Notice that the center of the hole is faceted and that the first fringe is a very narrow dark line. We know that this surface is different because it is curved. We also know from Figure 24.2 that, in a BF image, the first fringe must be bright if the thickness actually decreases to zero. We can therefore conclude from this one image that the specimen is not *tapering* to zero thickness at the center of this hole.

Although we've talked about wedges or specimens with gently curving surfaces so far, the way we actually calculate and analyze the contrast from such wedges is shown in Figure 24.4. We imagine that the specimen has two parallel surfaces which are normal to the electron beam, so that we have a fixed thickness for each calculation. We also assume that the beam is normal to the surface. We then change t and recalculate the intensity. Finally, we plot the different values for the intensity against t (as in Figure 24.2A) but we never actually inclined the surfaces!

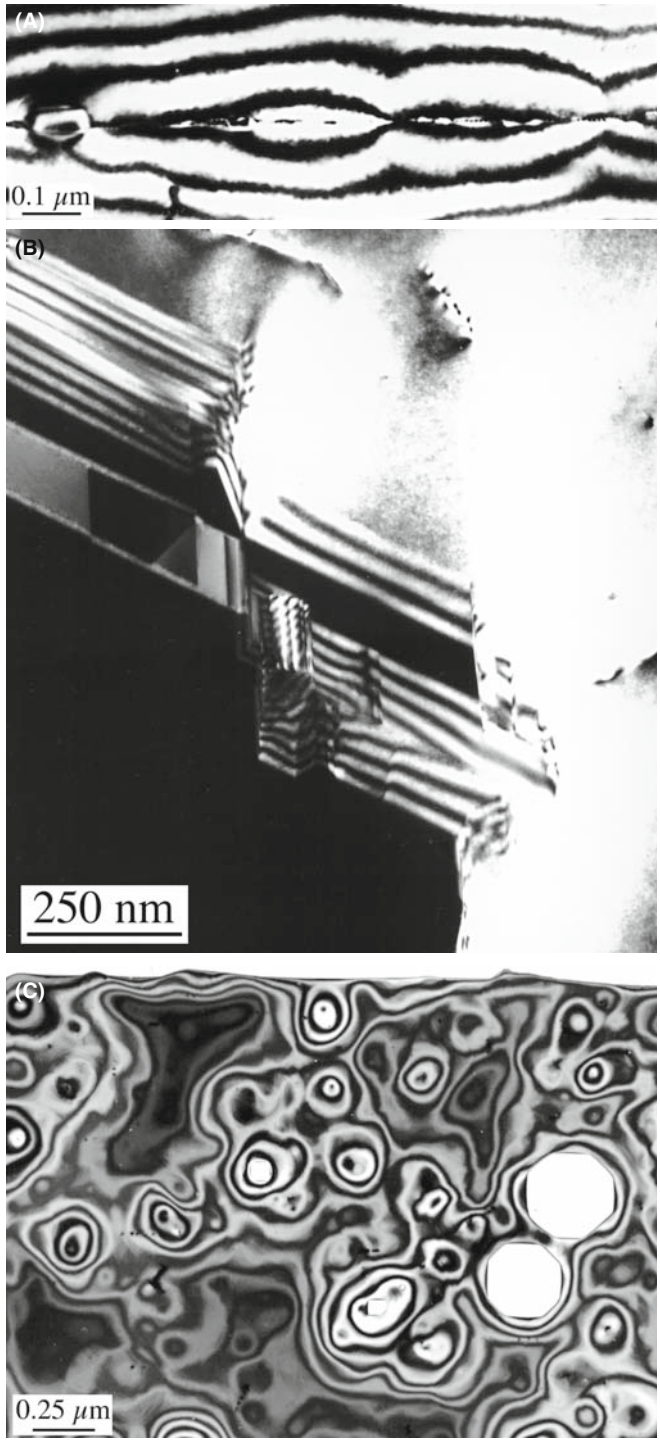


FIGURE 24.3. Examples of thickness fringes in (A) DF image of a preferentially thinned grain boundary; (B) a strong 220 DF image of microtwinned GaAs taken with only the right-hand grain diffracting, and (C) BF image of a chemically etched thin film of MgO. The white regions in (C) are holes in the specimen.

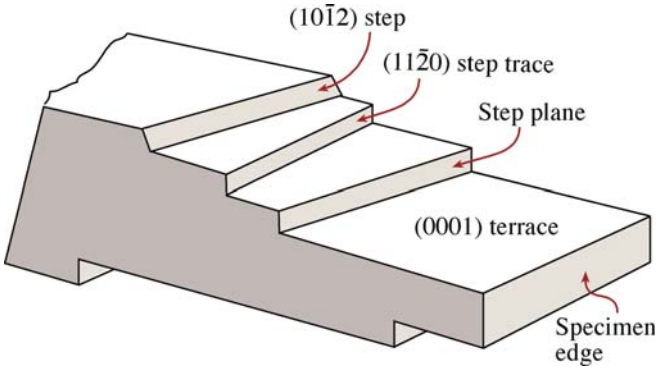


FIGURE 24.4. Schematic cross-sectional view of a specimen with terraces parallel to the surface and steps connecting terraces.

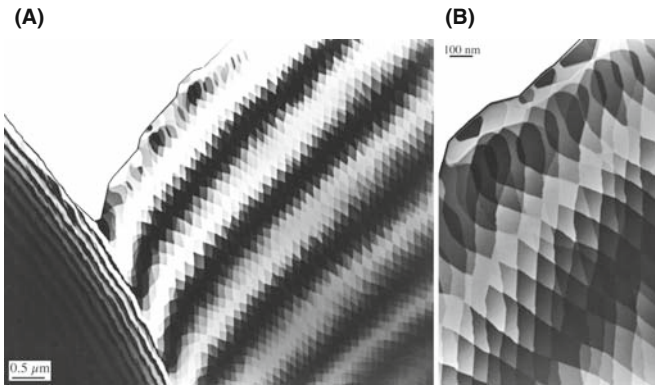


FIGURE 24.5. Thick fringes from an annealed Al_2O_3 specimen with the geometry shown in Figure 24.4. (A) At low magnification, the fringes are well defined and continuous, even when the wedge angle and wedge axis change. (B) At higher magnification, the contrast is seen to be quantized within a given fringe.

Figure 24.5 shows an image of a wedge-shaped specimen of Al_2O_3 which has been heat treated so that the surface has faceted parallel to certain low-index planes. The thickness fringes can then be seen to be discrete regions of different shades of gray; the fringes are, in general, quantized. You can form similar specimens by cleaving layer materials (e.g., graphite or mica) but the specimens tend to bend, which obscures these abrupt contrast changes.

24.3 THICKNESS FRINGES AND THE DP

A general rule in TEM is that, whenever we see a periodicity in real space (i.e., in the image), there must be a corresponding array of spots in reciprocal space; the converse is also true. If we image a specimen with a constant wedge angle, then we will see a uniform spacing of thickness fringes in both the BF and DF two-beam images even when $\mathbf{s_g} = 0$. We must therefore have more than one spot 'at G' when $\mathbf{s_g} = 0$ otherwise we would not see fringes. We already know that if we increase $\mathbf{s}$ or if the wedge angle were larger, then the fringe separation would decrease and the spacing of these spots must therefore increase.

To understand why there is more than one spot at G, go back to Chapter 17 where we showed that, because the specimen is thin, any spot in the DP will be elongated normal to the surface. When the specimen is wedge shaped, there will be two surfaces and we can imagine the spot being elongated normal to both surfaces, as was shown in Figure 17.4. Actually we have two curved relrods which do not intersect at $\mathbf{s} = 0$; we related this curvature to the dispersion surface in Chapter 15. The diffraction geometry close to G is shown in Figure 24.6.

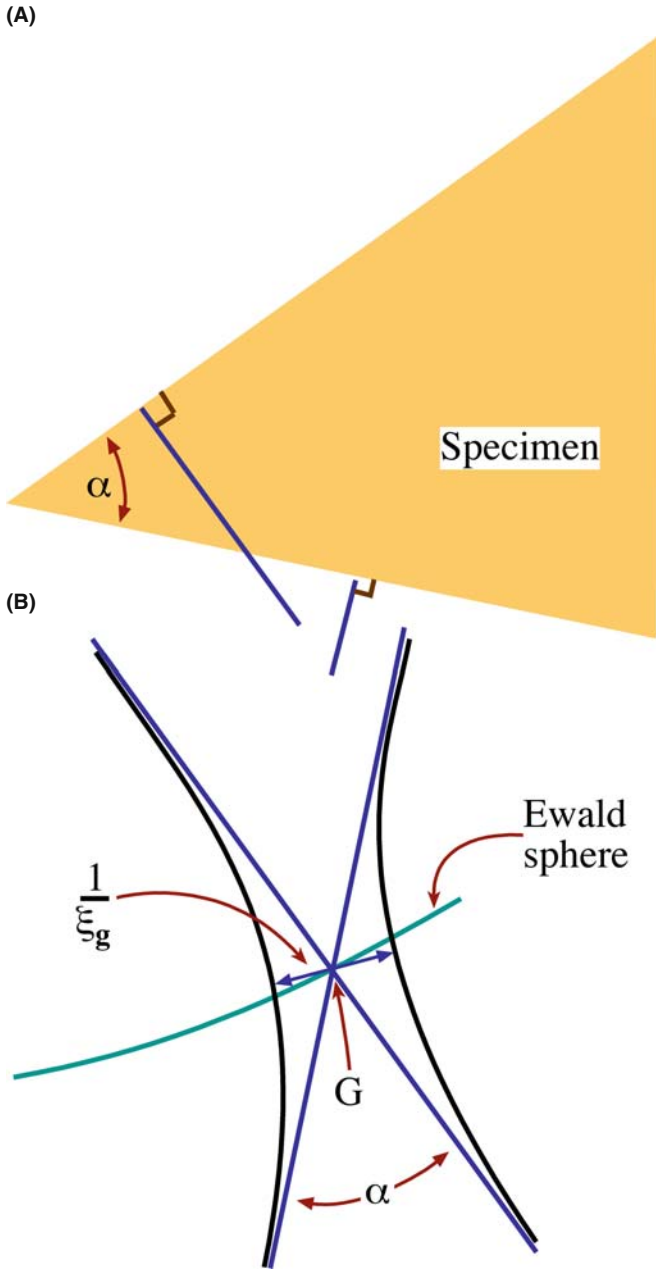


FIGURE 24.6. Relrods in (B) aligned normal to both surfaces of a wedge-shaped specimen in (A). In practice, the relrods (blue lines) don't actually cross (they are the black curves) so there are always two spots in the DP.

This spot spacing is thus related to ξ_g^{-1} and the wedge angle. While the spacing of the thickness fringes depends on ξ_g, it is not equal to ξ_g. As we referenced back in Chapter 17, Amelinckx's group has shown that we can describe this geometric relationship as shown in Figure 17.4. As we tilt the crystal away from $\mathbf{s} = 0$, the Ewald sphere will move up or down (as $\mathbf{s}$ becomes negative or positive) to cut the two 'rods.' So, you can see that there will be two spots instead of one at G and their separation will increase as $\mathbf{s}$ increases. As the separation increases, the spacing of the fringes decreases; the thickness fringes move closer together because the ξ_{eff} has decreased. The change in the fringe spacing is similar either side of $\mathbf{s} = 0$. So be wary of trying to make accurate thickness measurements for wedge-shaped crystals.

FRINGE PERIODICITY

The minimum spot spacing in the DP corresponds to the periodicity of the thickness fringes, which at $\mathbf{s} = 0$ is given directly by the extinction distance and the wedge angle.

We refer to thickness fringes as being an example of amplitude contrast because, in the two-beam case, they are associated with a particular reflection, $\mathbf{g}$. They actually occur due to interference between two beams, both of which are located close to $\mathbf{g}$, so they are really an example of phase contrast although we rarely think of them as such.

24.4 BEND CONTOURS (ANNOYING ARTIFACT, USEFUL TOOL, INVALUABLE INSIGHT)

This is a particularly satisfying topic, because you can understand it by considering a simple physical picture and yet the concept involved is the basis for understanding most aspects of defect contrast. Bend contours (don't call them extinction contours) occur when a particular set of diffracting planes is not parallel everywhere; the planes rock into, and through, the Bragg condition.

The specimen shown schematically in Figure 24.7 is aligned so that the *hkl* planes are exactly parallel to the incident beam at the center of the figure and always lie normal to the specimen surface even when it bends. We imagine that the foil bends evenly, so that the *hkl* planes are exactly in the Bragg condition at A and the $\bar{h}\bar{k}\bar{l}$ planes are exactly in the Bragg condition at B. We can draw the systematic row of reflections as you see below the bent crystal. Notice that −G is now on the left and G on the right (with no lens present). Now if we form a BF

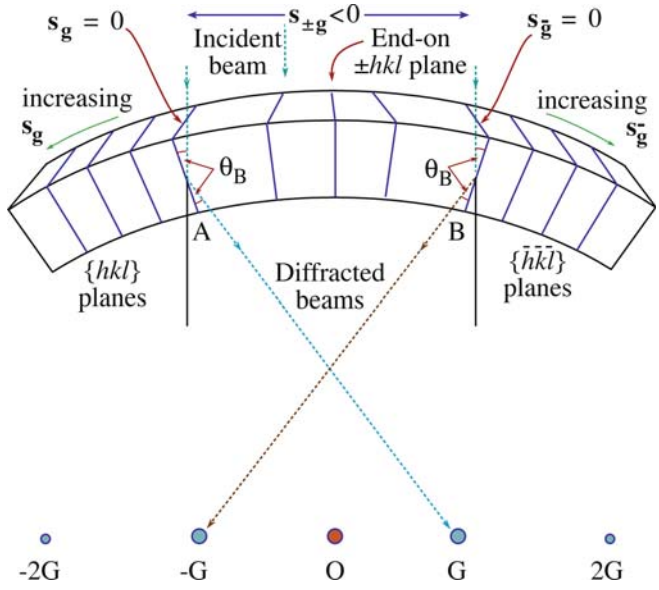


FIGURE 24.7. The origin of bend contours shown for a foil symmetrically bent either side of the Bragg conditions. For this geometry, when the *hkl* planes are in the Bragg condition, the reflection G is excited. Notice that G and the diffracting region are on opposite sides of O; if the foil were bent upwards, they would be on the same side.

image we will see two dark lines. Next, we form the DF image using reflection $\mathbf{g}$. We see a bright band on the left because that's where $\mathbf{g}$ is excited. Now use $\bar{\mathbf{g}}$ to form the image and the bright band is on the right. These bands are referred to as bend contours. A set of experimental images is also shown in Figures 24.8 and 24.9.

Remembering Bragg's law, the (2*h* 2*k* 2*l*) planes diffract strongly when θ has increased to $\sim 2\theta_B$. So we'll see extra contours because of the higher-order diffraction. As θ increases, the planes rotate through the Bragg condition more quickly (within a small distance Δx) so the bend contours become much narrower for higher-order reflections.

DISPLACE THE SAD APERTURE

In actually doing this imaging experiment, you should translate the objective aperture to form the DF images; you'll lose some resolution but don't move (neither translate nor tilt) the specimen.

Tyro-microscopists occasionally have difficulty in distinguishing higher-order bend contours from real line defects in the crystal. The solution is very simple: tilt your specimen. Bend contours are not fixed to any particular position in the specimen and quickly move as you tilt.

Bend contours are true amplitude contrast, not phase contrast.

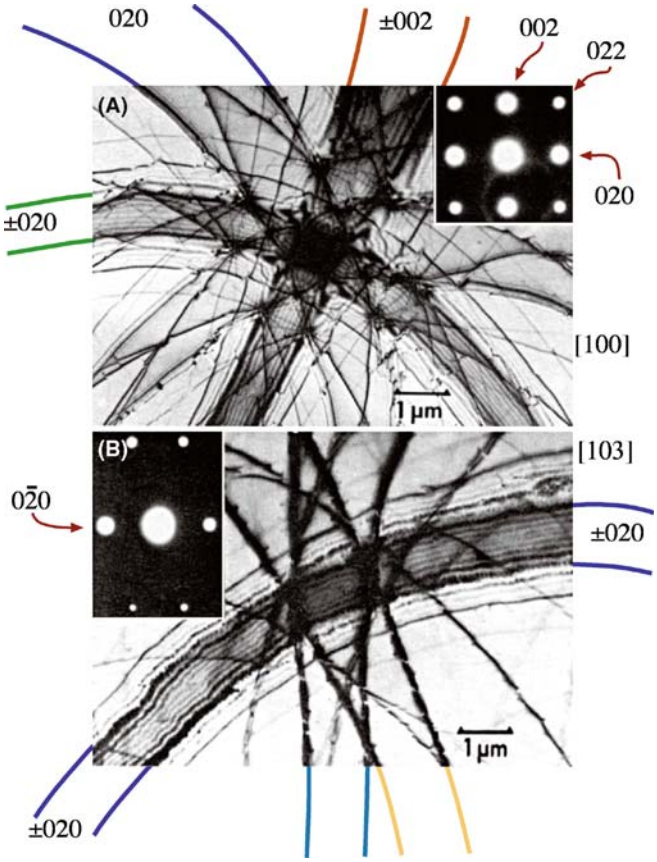


FIGURE 24.8. BF images of a bent Al specimen oriented close to the (A) [100] and (B) [013] zone axes. These images are known as (real-space) zone-axis patterns, or ZAPs, and are shown with their respective zone-axis DPs (insets). Each diffracting plane produces two bend contours, depending on whether θ_B or $-\theta_B$ is satisfied. Note that the separation of the bend contours is not uniform for any particular pair of planes because the curvature of the bending is not, in general, the same.

24.5 ZAPs AND REAL-SPACE CRYSTALLOGRAPHY

In the above discussion, we only considered bending about one axis. In real specimens, the bending will be more complex. This complexity will be important when the bent area is oriented close to a low-index pole, because the bend contours then form a zone-axis pattern or ZAP. Two examples of these ZAPs are shown in Figure 24.8. Although the ZAP is distorted, the symmetry of the zone axis is clear and such patterns have been used as a tool for real-space crystallographic analysis. Each contour is uniquely related to a particular set of diffracting planes, so the ZAP does not automatically introduce the twofold rotation axis that we are used to in SAD patterns. These contours are the real-space analog of the symmetry seen in large-angle CBED patterns.

In fact it's the exception that a $\pm\mathbf{g}$ pair of bend contours is straight and parallel. In case you are having

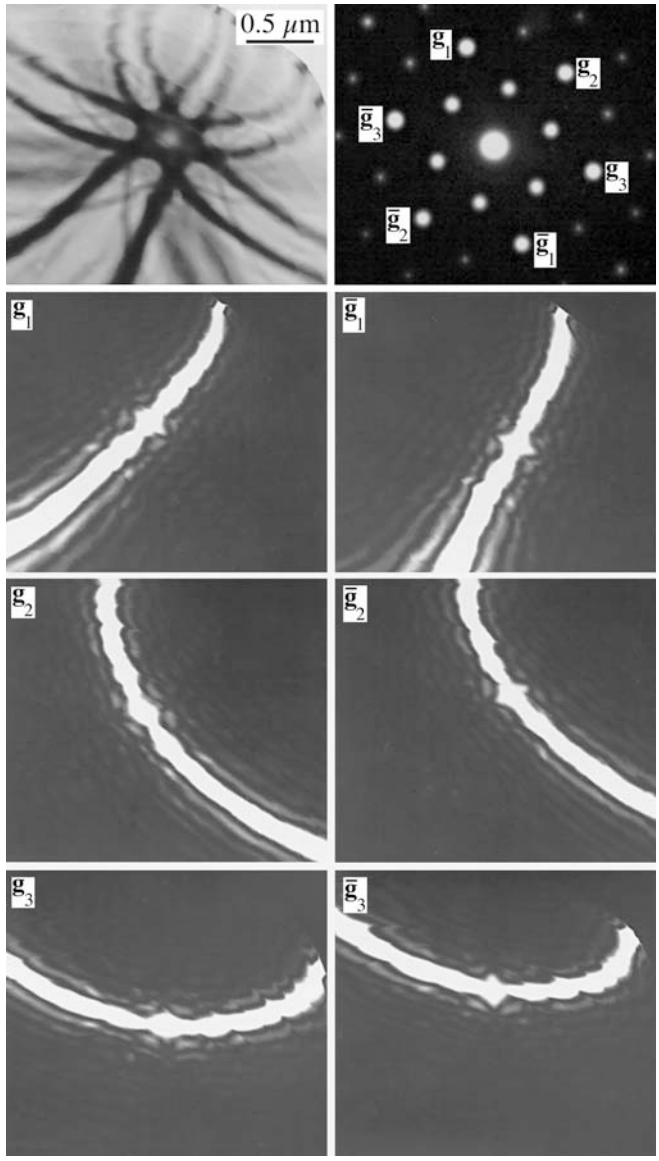


FIGURE 24.9. An 0001 real-space ZAP of Al_2O_3: (A) BF image and (B) corresponding DP. (C–H) Displaced-aperture DF images taken from the spots indicated in (B) identifying the principal dark bend contours in (A); (C, D) $\pm(\bar{3}030)$, (E, F) $\pm(\bar{3}300)$, (G, H) $\pm(03\bar{3}0)$. Note in (A) that the inner $\{11\bar{2}0\}$ spots produce fainter bend contours than the $\{\bar{3}300\}$.

a problem visualizing how a pair of contours might diverge, go back to the bent specimen in Figure 24.7, hold the $\bar{h}\,\bar{k}\,\bar{l}$ plane fixed at $x = x_0$, and then as you move along the foil (going into the page) gradually decrease the bend in the foil. The position where the hkl planes are in the Bragg condition gradually moves to the left, so $-x_0$ becomes more negative. Since x_0 is fixed the contours move apart in the image.

Notice how at the zone axis, the main 020 and 002 contours in the [100] ZAP are closely spaced, while in the [103] pattern, only one pair of contours is closely spaced; the others are more clearly defined and farther apart.

When the foil curvature is equal, this effect allows you to recognize a low-index ZAP. Since you can tilt the crystal, you can form different ZAPs from exactly the same area of your specimen just as you can for SAD and CBED (where we also use the term ZAP but then refer to a DP; really a ZADP). You index the contours in the manner described in Section 18.4 but use all the spots in the ZADP.

ZAPs IN REAL SPACE
In this case, a small **g** in the DP gives a small spacing in the image, contrary to the usual inverse relationship between image and DP.

If your specimen is buckled, you can tilt it so that a particular bend contour stays at the position you're studying in the image. You're then doing the same operation as we did using Kikuchi lines in Chapter 19. Tilting in image mode is more tricky, but if the specimen is very buckled or too thin, you can't use Kikuchi lines. The ZAP and bend contours let you work in real space. You can even set the value of **s** for a particular **g** at a particular location on your specimen!

24.6 HILLOCKS, DENTS, OR SADDLES

The simplest use for bend contours is in determining whether an area is a hillock or a dent. This information is useful when analyzing particles grown on a substrate, particularly if the substrate is a thin film.

Figure 24.9A and B shows a ZAP in a thin specimen of Al_2O_3 and the associated SADP; the dark bands are $\{3\bar{3}00\}$ bend contours. Figure 24.9C–H shows DF images recorded using each of these reflections; the area is identical in each of these images. Using Figure 24.7 you can determine the sense of the bending. You can see directly that the bend contour from one set of *hkl* planes does not necessarily lie parallel to those planes but instead both curves and changes in width.

Remember, for these contrast experiments especially, it is important to move the objective aperture, not the specimen. The resolution of the image will be somewhat lower but this is not critical for this application.

24.7 ABSORPTION EFFECTS

When your specimen is very thick you won't see an image, so we can say that the electrons have then all been *absorbed*. The absorption process is more important than this obvious statement might suggest. Much of our thinking about this topic is, however, just as empirical. In fact it is common to define an imaginary component $\xi'_{\mathbf{g}}$ to the extinction distance, so

$$\xi_{\mathbf{g}}^{\text{abs}} = \xi_{\mathbf{g}}\left(\frac{\xi'_{\mathbf{g}}}{\xi'_{\mathbf{g}} + i\xi_{\mathbf{g}}}\right) \tag{24.3}$$

Then we use $\xi_{\mathbf{g}}^{\text{abs}}$ in the Howie-Whelan equations. The term $\xi'_{\mathbf{g}}$ is found to be approximately $10\xi_{\mathbf{g}}$. The reason for choosing this expression for $\xi_{\mathbf{g}}^{\text{abs}}$ is that the $1/\xi_{\mathbf{g}}$ in the Howie–Whelan equations can then be replaced by $(i/\xi'_{\mathbf{g}} + 1/\xi_{\mathbf{g}})$. We do the same for $\xi_{\mathbf{0}}$. The result is that γ in the Howie-Whelan equations has an imaginary component. Consequently, we now have an exponential decay of the diffracted amplitude. It's a completely phenomenological treatment, but you will see reference to it. When we discuss EELS in Part 4, you'll appreciate the difficulties in modeling the effects of inelastic scattering on the image by a single parameter.

We did briefly discuss absorption of Bloch waves in Chapter 14. We showed that Bloch wave 2 (smaller **k**) is less strongly absorbed than Bloch wave 1; Bloch wave 1 travels along the atom nuclei while Bloch wave 2 channels between them. As the crystal becomes thicker we lose Bloch wave 1. Since thickness fringes result from a beating between the two beams, we will lose the thickness fringes but will still be able to 'see through' the specimen as you can appreciate from Figure 24.10. Now you can see why equation 24.3 must be phenomenological—there are no beams in the crystal!

Bend contours in thicker parts of the specimen will also show the effect of this anomalous absorption. Looking back at Chapter 15 on the dispersion surface, you'll see that when $\mathbf{s}_{\mathbf{g}}$ is negative, the tie line D_1D_2 would be closer to **0** than **g** and Bloch wave 2 contributes to $\phi_{\mathbf{g}}$ more strongly. When $\mathbf{s}_{\mathbf{g}}$ becomes positive, Bloch wave 1 is the more strongly excited. We lose Bloch wave 1 because the specimen is thick. So as we rock through the Bragg condition on the bend contour, we'll lose the thickness fringes faster where Bloch wave 1 was weaker already, i.e., when $\mathbf{s}_{\mathbf{g}}$ is negative or inside the $\pm\mathbf{g}$ pair of bend contours.

ANOMALOUS ABSORPTION
Absorption due to the loss of Bloch wave 1 is called anomalous absorption for historical reasons, not because it is unexpected.

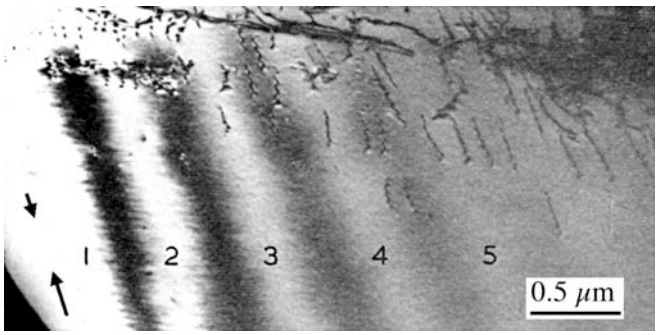


FIGURE 24.10. The contrast of thickness fringes in a two-beam BF image decreases when the effect of anomalous absorption is included. Note that the defects are still visible when the fringes have disappeared at a thickness of $-5\,\xi_g$.

We can summarize this discussion of absorption with some conclusions

- We can define a parameter ξ'_g which is usually about $10\,\xi_g$ and is really a fudge factor which modifies the Howie-Whelan equations to fit the experimental observations.
- The different Bloch waves are scattered differently. If they don't contribute to the image, we say that they were absorbed. We thus have anomalous absorption which is quite normal!
- Usable thicknesses are limited to about $5\xi_g$, but you can optimize this if you channel the less-absorbed Bloch wave.

24.8 COMPUTER SIMULATION OF THICKNESS FRINGES

Thickness fringes can be simulated. We'll talk more about the programs in Chapter 25, but be wary: don't use a program as a 'black box.' Why do we need or want to simulate thickness fringes? As an illustration, let's look at a 90°-wedge specimen (see Section 10.6) so that we know how the thickness changes with position. The actual thickness will be *very* sensitive to the orientation of the specimen since the specimen is so thick, as shown in Figure 24.11A. The specimen is a GaAs/$Al_xGa_{1-x}As$ layered composite grown on (001). Since the cleavage surface is {110} it can be mounted at 45° so the beam is nearly parallel to the [100] pole. The value of ξ_g is different for the two materials, so they can be readily distinguished. Clearly, a quantitative simulation of this situation is non-trivial, especially if you have to consider the effect shown in Figure 24.6.

Because the fringe spacing changes as ξ_g changes, it will also change if you vary the accelerating voltage. You can see this effect clearly in Figure 24.11B and C which compares the same region of a wedge specimen imaged at 300 and 100 kV.

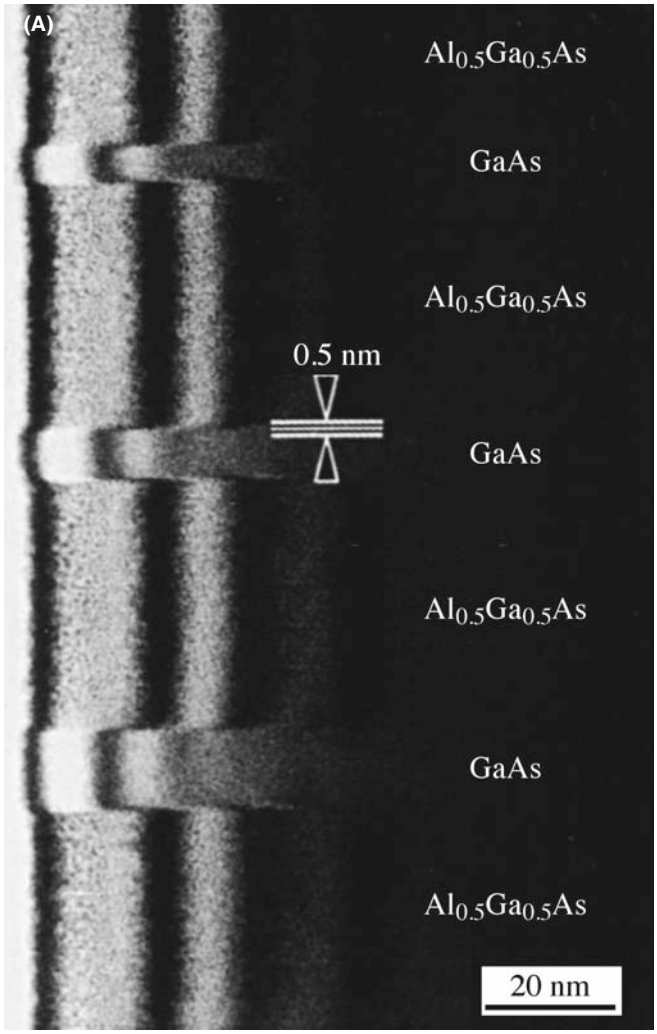


FIGURE 24.11. (A) Thickness fringes in a 90° wedge of alternating GaAs and AlGaAs. The extinction distance changes in each phase so the fringe spacing changes. Strong beam BF images ($\mathbf{s} = 0$) for (B) 300-kV and (C) 100-kV electrons. The extinction distance increases as the accelerating voltage increases, and you can see through thicker areas; compare with the images in Figure 24.1.

24.9 THICKNESS-FRINGE/BEND-CONTOUR INTERACTIONS

It's clear from equation 24.1 that both bending and thickness effects can occur together. This combined effect is shown in Figure 24.12, where the axis of bending runs normal to the edge of the wedge specimen. When $\mathbf{s} = 0$, the value of ξ_{eff} is largest. As we bend away from the Bragg condition, on either side, ξ_{eff}

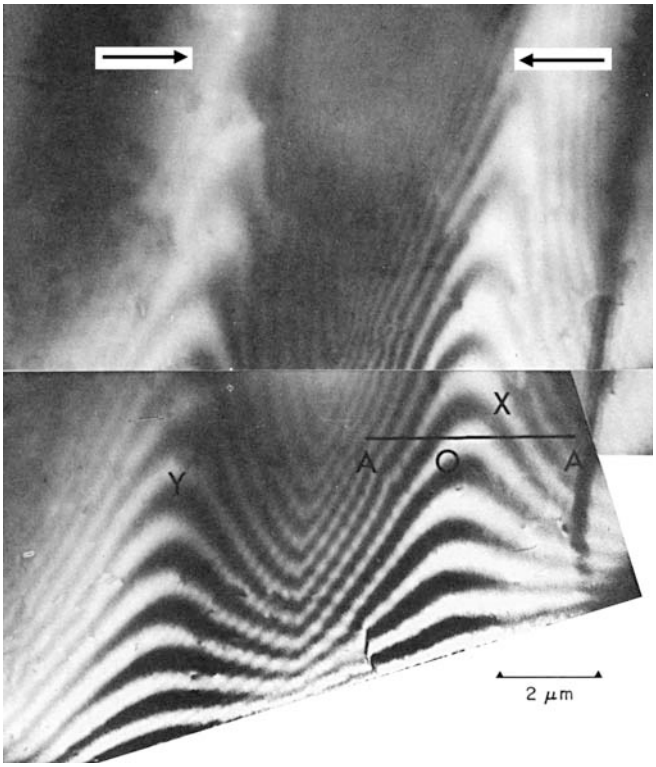


FIGURE 24.12. Since both thickness fringes and bend contours (X and Y) affect the contrast seen in the image, and both can occur in the same part of the specimen, they can affect, or couple with, one another to give the complex contrast shown in this BF image. Along the line A–A, **s** changes in sign, being approximately zero at O and negative between the contours.

decreases so the thickness contours curve towards the edge of the specimen. This image actually shows the $\mathbf{g}(111)$ and $\bar{\mathbf{g}}(\bar{1}\bar{1}\bar{1})$ contours (arrowed). As an exercise you can calculate the value of **s** at any point between the contours in this image. Assume the wedge angle is constant and $t = 0$ at the edge; then compare the thickness you deduce using ξ_{eff} with the thickness value extrapolated from the regions where $\mathbf{s} = 0$.

If a defect causes the specimen to bend, then the contrast from the defect and that from thickness variations will be linked.

Since the effective thickness is s_{eff}^{-1}, it will change as we increase the deviation parameter, **s**. You can use this fact to determine the thickness of an area quite accurately, provided you have a reference value, such as zero thickness at a hole in the specimen. You initially tilt the specimen so that the chosen reflection is at the Bragg condition. (Remember that this analysis assumes that we have only two beams.) You can determine $\mathbf{s_g}$ quite accurately if you can see the Kikuchi lines. Then you can determine s_{eff}^{-1} at different positions on the specimen. The maximum value of s_{eff}^{-1} is $\xi_{\mathbf{g}}$ and occurs at $\mathbf{s_g} = 0$. As you tilt the specimen to increase $\mathbf{s_g}$ in the positive or negative sense, you'll see the thickness fringes move closer together. We'll examine the situation where $\mathbf{s_g}$ is very large in Chapter 27, and where the foil also bends in Section 24.10.

ON THICKNESS DETERMINATION

Be careful when using fringes for determining t to use in XEDS analysis, since only the thickness of the diffracting (crystalline) material is determined. There may be amorphous material on the surface which has similar or different composition.

24.10 OTHER EFFECTS OF BENDING

In some situations, the bending of the foil may be more subtle. For example, strains in TEM specimens may relax at the surface of the thin specimen. A particularly important example of this effect was found in the study of superlattices in semiconductors.

We'll generalize the situation a little. Imagine that two cubic materials, which normally have slightly different lattice parameters, are grown on one another to form an artificial superlattice with an (001) (i.e., cube-on-cube) interface plane. One crystal must expand and the other contract normal to this interface (the tetragonal distortion). When we prepare a cross section TEM specimen, we might then imagine it relaxing at the surface as shown in Figure 24.13B. The reason for the relaxation is simply that this allows the one material to expand while the other

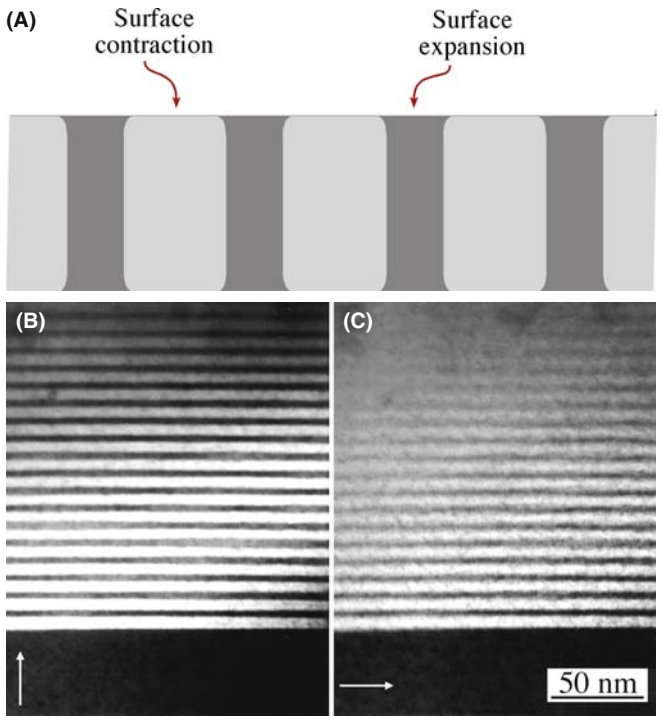


FIGURE 24.13. (A) A schematic of how an interface might relax at the surface of a thin specimen. (B, C) DF images of a GaAs/AlGaAs superlattice imaged in two orthogonal reflections, 200 and 020, with the specimen oriented at the 001 pole. (B) The [200] vector is parallel to the interface while (C) [020] is normal to it. If planes parallel to the interface bend to relax the strain caused by the lattice misfit, then only the 020 image will be affected, giving a more abrupt contrast change.

contracts; the constraint at the surface has been removed during the specimen preparation process. This argument is admittedly simplistic, but Figure 24.13B and C shows that images recorded with $\mathbf{g} = 020$ normal to this interface do indeed appear *sharper* than images formed when $\mathbf{g} = 200$ is parallel to the interface.

So, no matter whether **g** is 020 or $0\bar{2}0$, the Bragg planes are bent closer to $\mathbf{s} = 0$ at one surface or the other.

Here the bending only occurs within a short distance of the interface but it significantly affects the appearance of the DF image. The bending is actually making the image appear sharper than it should.

This example is special but emphasizes the point: relaxation at the surface can cause the diffracting planes to bend and this bending will affect the appearance of your image.

CHAPTER SUMMARY

The effects of changes in thickness and specimen bending are both explained by equation 24.1. Although this equation was derived for a two-beam geometry, you'll see similar effects when more strongly excited beams are present but the simple $\sin^2$ dependence will be lost.

- Varying t while keeping **s** constant gives thickness fringes.
- Varying **s** while keeping t constant gives bend contours.

Thickness fringes are an interference effect and, with care, can be used to calculate the foil thickness and reveal the topography.

Note that if the two surfaces of the specimen are parallel, then we don't see thickness fringes, even if the specimen is tilted. However, the contrast of that region will depend on the projected thickness.

Bend contours are very useful because they map out the value of **s** in the specimen. If your foil is bent around more than one axis, bend contours can combine to produce beautiful ZAPs which reflect the true symmetry of the material.

However, if you want to keep the defect analysis simple, then you need to avoid specimen bending and work in relatively thin regions of nearly constant thickness. There are exceptions to this rule: in a few special cases where you want to do exactly the opposite! So the message is that bending and thickness variations give you extra parameters that you can use in your study as long as you can control these parameters. This control comes from mastering the BF/DF/SAD techniques in Chapter 9.

Last, be aware that anomalous absorption is not anomalous. It can best (only) be explained by Bloch wave interactions.

SOME SPECIFIC READING

Gibson, JM, Hull, R, Bean, JC and Treacy, MMJ 1985 *Elastic Relaxation in Transmission Electron Microscopy of Strained-Layer Superlattices* Appl. Phys. Lett. **46**(7) 649–651. Shows the effect of surface relaxation on the contrast from superlattices.

Rackham, GM and Eades, JA 1977 *Specimen Contamination in the Electron Microscope When Small Probes are Used* Optik **47**(2) 226–232. Example of using CBED for real-space crystallographic analysis.

Susnitzky, DW and Carter, CB 1992 *Surface Morphology of Heat-Treated Ceramic Thin Films* J. Am. Ceram. Soc. **75**(9) 2463–2478. An overview of the surface morphology of heat-treated ceramic thin films studied using TEM.

THE COMPANION TEXT

The simulation programs discussed in the companion text allow you to do the contrast experiments in a computer.

SELF-ASSESSMENT QUESTIONS

Q24.1 In a BF image, there is a set of dark lines on either side of a particular bend contour. What causes these lines?

Q24.2 Is it always correct to say "my sample must vary in thickness a lot because I see a lot of thickness fringes"? (Apart from the grammar.)

Q24.3 In a TEM image, you see a pair of dark lines indicating a bend contour. Is this BF or DF?
Q24.4 When would a wedge sample give two closely spaced spots in the DP and when would it not?
Q24.5 Imagine that you can continuously increase the accelerating voltage on the TEM from 200 to 300 kV. How will the thickness fringes change?
Q24.6 If your specimen is exceptionally thin but bent, how can you find a zone axis?
Q24.7 A set of bend contours gradually broadens and the contours move apart. What is happening in the specimen?
Q24.8 If the calculated extinction distance of a sample is 50 nm, what is the maximum thickness of the specimen that you can use to produce reasonable BF images?
Q24.9 How will thickness fringes curve as the specimen bends away from the Bragg condition?
Q24.10 Are thickness fringes caused by amplitude contrast or phase contrast?
Q24.11 Can thickness fringes tell you about surface topography?
Q24.12 What causes bend contours?
Q24.13 How do you distinguish between bend contours and dislocations?
Q24.14 Can a void cause bend contours?
Q24.15 What is a ZAP and why does it occur?
Q24.16 Write down the empirical expression for the extinction distance when absorption is included.
Q24.17 How reliably can you determine the thickness of a specimen using thickness fringes?
Q24.18 When imaging a GaAs/AlGaAs superlattice grown on (001), which {200} reflection will appear to give the sharpest image and why is this so?
Q24.19 If you look along [100] at a 90°-wedge specimen of a (001) GaAs quantum well in a AlGaAs matrix, which material will show more thickness fringes in two-beam conditions?
Q24.20 How thick is the crystal at the center of the third bright fringe in a two-beam DF image (as a multiple of $\xi_{\mathbf{g}}$)?

TEXT-SPECIFIC QUESTIONS

T24.1 If the two grains in Figure 24.3A have a common [001] foil normal, describe the shape of the specimen assuming the electron had 120-keV energy.
T24.2 Is there an error in the caption for Figure 24.3? Estimate the thickness of the sample. Explain all your assumptions.
T24.3 What is the wedge angle of the two segments of specimen in Figure 24.5A. Explain your assumptions.
T24.4 In Figure 24.8A, the ±002 bend contours show that the specimen bends. What is the axis of bending that produces this pair of contours? Deduce the variation in the angle through which the specimen bends.
T24.5 In Figure 24.8B, index the DP. Estimate the radius of curvature around the $[30\bar{1}]$ axis.
T24.6 Explain the shape of the specimen in Figure 24.9.
T24.7 Why does the contrast of the thickness fringes in Figure 24.10 decrease moving left to right?
T24.8 Why is there a contraction and expansion in Figure 24.13? Does the shading therefore correspond to that in the image?
T24.9 In Figure 24.11, consider the GaAs region. What voltage was used to form this image? Explain how you would deduce this (and do so) even if you could not look up the original publication.
T24.10 In equation 24.3, we introduce ξ_g', the anomalous extinction distance. What is this distance for Cu 220 at 100 and 300 kV? Is this a reasonable value to explain the observations in Figure 24.10 (assuming Cu 220 and 100 kV)? Why is ξ_g' a fudge factor?
T24.11 A crystal of nickel is prepared in the form of a wedge with its upper surface parallel to (100) and its lower surface to (110). Electrons of energy 100 keV enter through the upper surface and travel in the (001) plane perpendicular to the edge of the wedge, falling on the (020) planes at the exact Bragg angle. If the extinction distance for this reflection is 25 nm, calculate the angular splitting of the reflected beam due to refraction as it leaves through the lower surface. (Courtesy Mike Goringe.)
T24.12 A specimen of thickness t contains a small inclusion of thickness Δz which scatters electrons differently from the matrix. The crystal is viewed under two-beam dynamical conditions at the Bragg position ($s = 0$) and the effective extinction distances are ξ_{gm} for the matrix and ξ_{gt} for the inclusion, which may be assumed to have the same crystal structure as the matrix. Calculate the visibility of the inclusion as a function of (a) its depth in the specimen and (b) the specimen thickness. How would you expect these results to be modified if the specimen is not precisely at the Bragg position (i.e., $s \neq 0$)? (Courtesy Mike Goringe.)
T24.13 A bcc sample has been badly bent due to mishandling. You have imaged a bend center and the corresponding ZADP. The zone axis is [011]; draw and index the main diffraction spots you would expect at this zone axis. If the sample is locally bent in the shape of a horse saddle, explain (using a sketch, if necessary) the relationship between the DP and the bent contour lines that form the bend center. How would you confirm the relationship you suggest? (Courtesy Ian Robertson.)

25

Planar Defects

CHAPTER PREVIEW

Internal interfaces (grain boundaries, phase boundaries, stacking faults) or external interfaces (i.e., surfaces) are surely the most important defects in crystalline engineering materials. Their common feature is that we can usually think of them as all being two-dimensional, or planar, defects (even though they're not really). The main topics of this chapter will be

- Characterizing which type of internal interface we have and determining its main parameters.
- Identifying lattice translations at these interfaces from the appearance of the diffraction-contrast images.

Rotations are usually associated with line defects and they will be discussed in Chapter 26. We can't usually identify the details of the local structure of an interface unless we use HRTEM, so we will return to that topic in Chapter 28. We'll talk a little about simulating images of these defects because it is relatively direct and instructive.

25.1 TRANSLATIONS AND ROTATIONS

This section is a brief summary of some fundamentals of materials science. If the concepts are not immediately familiar to you, then go and study an introductory textbook on defects in materials, several of which are in the reference section.

An interface is simply a surface which separates any two distinct regions of the microstructure. For most of our discussion, we will make the assumption that the surface is flat and is thus a planar (parallel to a plane) defect, though this is rarely the case. We can sketch a general interface as shown in Figure 25.1

With this general definition, we can summarize the different classes of planar defects

- *Translation boundary, RB*. Any translation **R**(**r**) is allowed, θ is zero, and both regions are identical and thus perfectly aligned. Stacking faults (SFs) are a special case. We'll denote the translation boundary as RB so as to avoid confusing it with the twin boundary (TB).
- *Grain boundary, GB*. Any values of **R**(**r**), **n**, and θ are allowed (**n** is the normal to the defect plane), but the chemistry and structure of the two grains must be the same. The SF is again a special case, but this class also includes TBs.
- *Phase boundary, PB*. As for a GB, but the chemistry and/or structure of the two regions can differ.
- *Surface*. A special case of a PB where one phase is vacuum or gas.

DEFINE R(r)

The upper crystal is held fixed while the lower one is translated by a vector **R**(**r**) and/or rotated through some angle θ about any axis, **v**.

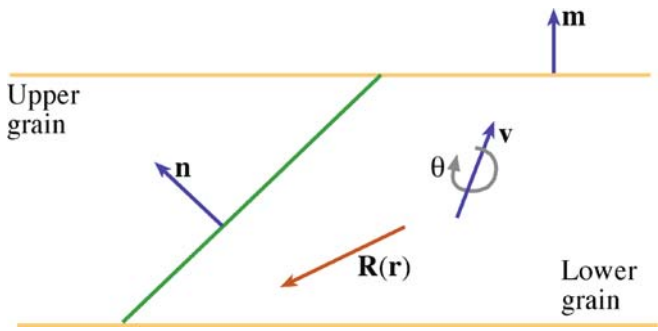


FIGURE 25.1. A specimen containing a planar defect. The lower grain is translated by a vector **R**(**r**), and rotated through an angle θ about the vector **v**, relative to the upper grain. The defect plane normal is **n**, the foil normal is **m**.

TABLE 25.1 Examples of Internal Planar Defects

Planar defect	Structure	Example	Example
SF	Diamond-cubic, fcc, zinc blende	Cu, Ag, Si, GaAs	$\mathbf{R} = \frac{1}{3}[111]$ or $\mathbf{R} = \frac{1}{6}[112]$
APB/IDB	Zinc blende, wurtzite	GaAs, AlN	Inversion
APB	CsCl	NiAl	$\mathbf{R} = \frac{1}{2}[111]$
APB/SF	Spinel	$MgAl_2O_4$	$\mathbf{R} = \frac{1}{4}[110]$
GB	All materials	Often denoted by Σ where Σ^{-1} is the fraction of coincident lattice sites	Rotation plus **R**
PB	Any two different materials	Sometimes denoted by Σ_1,Σ_2 which are not equal	Rotation plus **R** plus misfit
RB	Extra translation	{112} Twin boundary in Al	**R** not related to lattice

Now with each of these groups, we can have special examples. We will list some of the most common examples in Table 25.1, including those that we will consider in this chapter. For a more detailed discussion, we refer you to the general references at the end of the chapter.

RBs include the familiar SFs found in fcc, hcp, diamond-cubic, and layer materials. They have been widely studied because they play an important role in the mechanical properties of the fcc metals, e.g., Cu and stainless steel. They are also found in more complex materials such as spinels, Ni_3Al, Ti_3Al, etc., where the lattice parameters, and therefore the dislocation Burgers vectors, are large.

The anti-phase boundary (APB) in ordered CuAu (which we can describe as two interpenetrating simple-cubic superlattices) is produced by translating one superlattice by $\frac{1}{2}<111>$ with respect to the other. It is called an APB because one superlattice is out of phase with the other. If the crystal were disordered and the Cu and Au occupied the bcc sites randomly, then $\frac{1}{2}<111>$ would be a lattice vector and no defect would exist. This particular APB can thus be regarded as an SF. We will find that the methods used to characterize RBs can often be used to determine $\mathbf{R}(\mathbf{r})$ in other interfaces.

STACKING FAULT PLANES

Although we know that {111} is the favored SF plane for fcc metals, SFs in other materials lie on different planes.

GBs fall into two groups, low angle and high angle. Low-angle boundaries necessarily involve a rotation through a small angle that is usually accommodated by arrays of dislocations; we'll consider these defects in Chapter 26. High-angle boundaries can adopt some special values of **n** and θ; a large fraction of lattice sites in one grain is then shared by (is common to) the other grain. We characterize the fraction by the inverse of this fraction, which we call Σ. For example, the common twin boundary in fcc metals is the $\Sigma = 3$ grain boundary (the fraction is $\frac{1}{3}$). The reason this is important to our discussion is that if a set of lattice points is common to two grains (as implied by the Σ coincidence-site lattice concept), then certain planes may also be common and may give rise to common reflections. These reflections will remain common even if one grain is translated relative to the other. In that case, we'll have a special type of RB, called the rigid-body translation. Rigid-body translations in grain boundaries behave just like other SFs except **R** is usually small and is not directly related to the lattice parameters.

There is a second group of APBs where the two grains cannot be related by a translation. These occur in GaAs, ZnO, AlN, and SiC, for example. One lattice can always be related to the other by a rotation of 180° to give the equivalent of an inversion; they are sometimes known as inversion domain boundaries (IDBs). (They are $\Sigma = 1$ GBs.) These special interfaces can often be seen in the image because there is usually a small associated translation. We analyze this translation as if it were a simple RB because all the planes on one side of the IDB are parallel to their counterparts in the other; the (*hkl*) plane on one side is parallel to the $\bar{h}\bar{k}\bar{l}$ on the other. We can't distinguish **g** from $\bar{\mathbf{g}}$ unless we use CBED or we have a great specimen for HRTEM.

PBs are rarely (if ever) analyzed fully using diffraction contrast. If the orientation, chemistry, and structure can all change on crossing the boundary, then not only will the reflections change, but all extinction distances will change too. Some special examples of such interfaces are hcp-Co/fcc-Co, bcc-Fe/fcc-Fe, NiO/$NiFe_2O_4$, and GaAs/$Al_xGa_{1-x}As$. Of course, the number of other such interfaces is countless.

Surface studies using TEM have quite recently become very important although the experimental tools have been available for some time. We will discuss surfaces in this chapter insofar as they are imaged by diffraction contrast. So-called profile imaging will follow in Chapter 28 on HRTEM. The other two surface-sensitive techniques are plan-view and reflection electron microscopy (REM, see Chapter 29).

25.2 WHY DO TRANSLATIONS PRODUCE CONTRAST?

As usual, we will start our analysis considering only two beams, O and G. Our approach will use hand-waving arguments, which are not perfect, to justify adapting the Howie-Whelan equations for specimens containing interfaces. We'll use the same approach for other defects in Chapters 26 and 27. Because the Howie-Whelan equations for perfect crystals assume two-beam conditions, we are able to solve them analytically. We'd like to be able to do the same when defects are present, because this gives us a physical understanding of the processes which produce the contrast. There are two important features that we will need to keep in mind

- Diffraction contrast only occurs because we have Bloch waves in the crystal. However, our analysis will initially only consider diffracted beams.
- We make the column approximation so we can solve the equations; we must be wary whenever the specimen or the diffraction conditions change within a distance comparable to the column diameter.

TEM OF SURFACES
The techniques are (i) profile imaging, (ii) plan-view imaging, and (iii) reflection electron microscopy (REM).

A unit cell in a strained crystal will be displaced from its perfect-crystal position so that it is located at position $\mathbf{r}'_n$ instead of $\mathbf{r}_n$ where n is included to remind us that we are considering scattering from an array of unit cells; we'll soon omit the n (and we still have the column approximation)

$$\mathbf{r}'_n = \mathbf{r}_n + \mathbf{R}_n \tag{25.1}$$

In this expression, $\mathbf{R}_n$ is actually $\mathbf{R}_n(\mathbf{r})$; it can vary throughout the specimen. The term $e^{2\pi i\mathbf{K}.\mathbf{r}}$ in equation 13.3 (it's actually $\mathbf{Kr_n}$) now becomes $e^{2\pi i\mathbf{K}.\mathbf{r}'}$ so we need to examine the term $\mathbf{K}.\mathbf{r}'$. We know that $\mathbf{K}$ is $\mathbf{g} + \mathbf{s}$, so we can write

$$\mathbf{K}.\mathbf{r}'_n = (\mathbf{g}+\mathbf{s}).(\mathbf{r}_n+\mathbf{R}_n) = \mathbf{g}.\mathbf{r}_n + \mathbf{g}.\mathbf{R}_n + \mathbf{s}.\mathbf{r}_n + \mathbf{s}.\mathbf{R}_n \tag{25.2}$$

Now since $\mathbf{r}_n$ is a lattice vector, $\mathbf{g}.\mathbf{r}_n$ is an integer as usual. The third term, $\mathbf{s}.\mathbf{r}_n$ gives our usual sz term so the new terms are $\mathbf{g}.\mathbf{R}_n$ and $\mathbf{s}.\mathbf{R}_n$.

When we discuss strong-beam images we know that $\mathbf{s}$ is very small. Since we are using elasticity theory $\mathbf{R}_n$ must be small. Hence we ignore the term $\mathbf{s}.\mathbf{R}_n$. Remember that we have made a special assumption which may not be valid in two situations

- When $\mathbf{s}$ is large; we'll encounter this when we discuss the weak-beam technique in Chapter 27.
- When the lattice distortion, $\mathbf{R}$, is large; this occurs close to the cores of some defects.

We now modify equation 13.8 intuitively to include the effect of adding a displacement from equation 25.2

$$\frac{d\phi_\mathbf{g}}{dz} = \frac{\pi i}{\xi_\mathbf{0}}\phi_\mathbf{g} + \frac{\pi i}{\xi_\mathbf{g}}\phi_\mathbf{0}\exp[-2\pi i(sz + \mathbf{g}.\mathbf{R})] \tag{25.3}$$

and

$$\frac{d\phi_\mathbf{0}}{dz} = \frac{\pi i}{\xi_\mathbf{0}}\phi_\mathbf{0} + \frac{\pi i}{\xi_\mathbf{g}}\phi_\mathbf{g}\exp[+2\pi i(sz + \mathbf{g}.\mathbf{R})] \tag{25.4}$$

Next, we simplify these equations just as we did in Chapter 13 by setting

$$\phi_\mathbf{0}(z)_{(\text{sub})} = \phi_\mathbf{0}\exp\left(\frac{-\pi i z}{\xi_\mathbf{0}}\right) \tag{25.5}$$

and

$$\phi_\mathbf{g}(z)_{(\text{sub})} = \phi_\mathbf{g}\exp\left(2\pi i s z - \frac{\pi i z}{\xi_\mathbf{0}}\right) \tag{25.6}$$

Then the Howie-Whelan equations become

$$\frac{d\phi_{\mathbf{0}(\text{sub})}}{dz} = \frac{\pi i}{\xi_\mathbf{g}}\phi_{\mathbf{g}(\text{sub})}\exp(2\pi i\mathbf{g}.\mathbf{R}) \tag{25.7}$$

and

$$\frac{d\phi_{\mathbf{g}(\text{sub})}}{dz} = \frac{\pi i}{\xi_\mathbf{g}}\phi_{\mathbf{0}(\text{sub})}\exp(-2\pi i\mathbf{g}.\mathbf{R}) + 2\pi i s\phi_{\mathbf{g}(\text{sub})} \tag{25.8}$$

These equations are just as before (equations 13.14 and 13.15) but with the addition of the $2\pi i\mathbf{g}.\mathbf{R}$ term. This additional phase is termed α, hence planar defects are seen when $\alpha \neq 0$.

$$\alpha = 2\pi\mathbf{g}.\mathbf{R} \tag{25.9}$$

These expressions will be particularly useful in two cases

- When $\mathbf{R}$ = constant.
- Understanding phasor diagrams when defects are present.

We start with a simple stacking fault lying parallel to the surface as shown in Figure 25.2. In this situation, the beams propagate through the upper layer just as if no fault were present. At a depth $z = t_1$, the beams may experience a phase change due to the effect of the

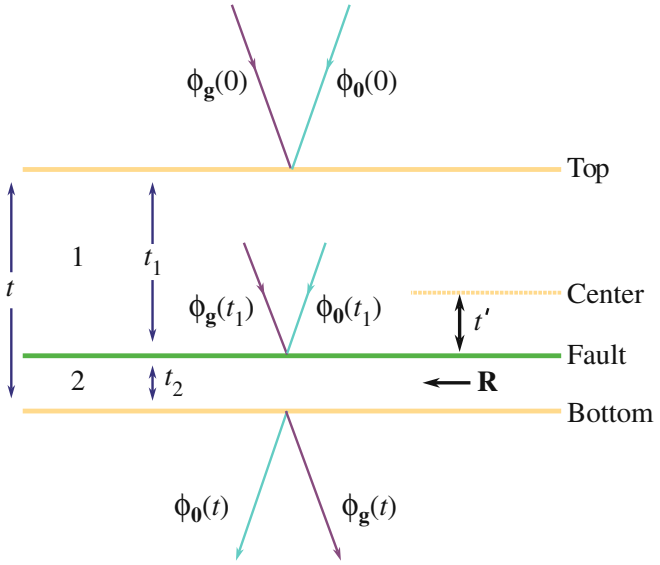


FIGURE 25.2. A stacking fault lying at depth t_1 in a parallel-sided uniformly thick specimen. The total thickness is t and $t_2 = t - t_1$.

translation **R**, but after that they again propagate as if in perfect crystal.

In this chapter, we'll see several values of α. A special case occurs when $\alpha = \pm 120°$. This value of α is often encountered since it occurs for fcc SFs. We'll also encounter the case where $\alpha = \pm 180°$; this value arises for some special APBs which are really SFs.

25.3 THE SCATTERING MATRIX

This discussion of the scattering matrix introduces no new concepts. It is just a different way of writing the equations so that if you are calculating the image contrast, you can program the computer more easily, especially when you have complicated arrays of lattice defects. Our reason for delaying the introduction of the scattering matrix until now is that it is much easier to understand when you can apply it to a specific problem.

In equations 13.17–13.20, we showed that in the two-beam case, we can write these simple expressions for ϕ_0 and ϕ_g

$$\phi_0 = C_0 e^{2\pi i \gamma z} \tag{25.10}$$

and

$$\phi_g = C_g e^{2\pi i \gamma z} \tag{25.11}$$

Since there are two values for γ, we can express both the **0** and **g** beams as the combination of these two contributions to give

$$\phi_0(z) = C_0^{(1)}\psi^{(1)} \exp\left(2\pi i \gamma^{(1)} z\right) + C_0^{(2)}\psi^{(2)} \exp\left(2\pi i \gamma^{(2)} z\right) \tag{25.12}$$

and

$$\phi_g(z) = C_g^{(1)}\psi^{(1)} \exp\left(2\pi i \gamma^{(1)} z\right) + C_g^{(2)}\psi^{(2)} \exp\left(2\pi i \gamma^{(2)} z\right) \tag{25.13}$$

where the $\psi^{(i)}$ terms tell us the relative contributions of the $\gamma^{(1)}$ and $\gamma^{(2)}$ terms. (We are really saying that both Bloch waves contribute to both the **0** and **g** beams.) We can rewrite equations 25.12 and 25.13 in a matrix form. (This is the key step.)

$$\begin{pmatrix} \phi_0(z) \\ \phi_g(z) \end{pmatrix} = \begin{pmatrix} C_0^{(1)} & C_0^{(2)} \\ C_g^{(1)} & C_g^{(2)} \end{pmatrix} \begin{pmatrix} \exp(2\pi i \gamma^{(1)} z) & 0 \\ 0 & \exp(2\pi i \gamma^{(2)} z) \end{pmatrix} \begin{pmatrix} \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} \tag{25.14}$$

We can express our boundary conditions as

$$C_0^{(1)}\psi^{(1)} + C_0^{(2)}\psi^{(2)} = \phi_0(0) \tag{25.15}$$

and

$$C_g^{(1)}\psi^{(1)} + C_g^{(2)}\psi^{(2)} = \phi_g(0) \tag{25.16}$$

which we can now rewrite as

$$\begin{pmatrix} C_0^{(1)} & C_0^{(2)} \\ C_g^{(1)} & C_g^{(2)} \end{pmatrix} \begin{pmatrix} \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} = \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \tag{25.17}$$

(We actually saw in Section 13.9 that $\phi_0(0)$ is 1 and $\phi_g(0)$ is 0 because $z = 0$ is the top surface.) We know from Chapters 13 and 14 that the elements of the matrix C are determined by the two-beam conditions that we set up. The matrix C does not depend on z. Now we can use matrix algebra to solve equation 25.17. First rewrite it as

$$C \begin{pmatrix} \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} = \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \tag{25.18}$$

then rewrite equation 25.18 as

$$\begin{pmatrix} \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} = C^{-1} \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \tag{25.19}$$

where C^{-1} is just the inverse matrix. Remember that the order is important in matrix multiplication and that $C^{-1}C = I$, the unit matrix.

Therefore we can rewrite equation 25.14 as

$$\begin{pmatrix} \phi_{\mathbf{0}}(z) \\ \phi_{\mathbf{g}}(z) \end{pmatrix} = C \begin{pmatrix} \exp(2\pi i \gamma^{(1)} z) & 0 \\ 0 & \exp(2\pi i \gamma^{(2)} z) \end{pmatrix} C^{-1} \begin{pmatrix} \phi_{\mathbf{0}}(0) \\ \phi_{\mathbf{g}}(0) \end{pmatrix} \tag{25.20}$$

Finally, we can define a new matrix $P(z)$ as the scattering matrix for a slice of perfect material with thickness z

$$P(z) = C \begin{pmatrix} \exp(2\pi i \gamma^{(1)} z) & 0 \\ 0 & \exp(2\pi i \gamma^{(2)} z) \end{pmatrix} C^{-1} = C\,\Gamma C^{-1} \tag{25.21}$$

The matrix $P(z)$ thus gives us the values of the exit wave amplitudes at the bottom of the slice in terms of the incident values. In other words, the matrix $P(z)$ includes all the information to describe the propagation of the beams through the crystal; $P(z)$ is a *propagator* matrix.

P = C Γ C^{-1} AND VARIES WITH z

Notice that z only enters the 'P equation' through the Γ matrix.

25.4 USING THE SCATTERING MATRIX

Now we illustrate the real strength of the scattering matrix approach by considering the effect of a planar fault lying parallel to the foil surface, as we saw in Figure 25.2. The idea is that we now have two slices of material of thickness t_1 and t_2. We can easily calculate $\phi_{\mathbf{0}}(t_1)$ and $\phi_{\mathbf{g}}(t_1)$ using equation 25.20. These values for $\phi_{\mathbf{0}}$ and $\phi_{\mathbf{g}}$ then become the incident values for slice 2. The effect of the translation $\mathbf{R}$ is to multiply the terms in $C_{\mathbf{g}}$ in the lower slice by a phase factor $\exp(-i\alpha)$ where $\alpha = 2\pi \mathbf{g} \cdot \mathbf{R}$ as usual. The matrix C for slice 2 is then written as

$$C_2 = \begin{pmatrix} C_{\mathbf{0}}^{(1)} & C_{\mathbf{0}}^{(2)} \\ C_{\mathbf{g}}^{(1)} \exp(i\alpha) & C_{\mathbf{g}}^{(2)} \exp(i\alpha) \end{pmatrix} \tag{25.22}$$

We can write down the expression for $\phi_{\mathbf{0}}(t)$ and $\phi_{\mathbf{g}}(t)$ as

$$\begin{pmatrix} \phi_{\mathbf{0}}(t) \\ \phi_{\mathbf{g}}(t) \end{pmatrix} = C_2 \Gamma(t_2) C_2^{-1} C_1 \Gamma(t_1) C_1^{-1} \begin{pmatrix} \phi_{\mathbf{0}}(0) \\ \phi_{\mathbf{g}}(0) \end{pmatrix} \tag{25.23}$$

Here the subscripts on C_1 and C_2 just identify the slices. Normally, this equation goes straight into the computer, which handles matrices very easily. However, we'll go back and consider a few special points

- Look at equation 25.22 and set $\mathbf{R} = 0$ so that $C_2 = C_1$. You can see that $P(t) = P(t_1)P(t_2)$. Clearly we could cut the perfect-crystal specimen into many slices and $P(t)$ would always be the product of the scattering matrices for each slice. (We could call this a multislice approach to scattering from a crystal but we don't to avoid confusion with HRTEM simulation.)
- How do we prove equation 25.22? From equation 14.12 we know that a Bloch wave can be written as

$$b(\mathbf{k}) = \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k}) \exp(2\pi i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{r}) \tag{25.24}$$

- If the lower crystal is displaced by a vector $\mathbf{R}$ then we replace $\mathbf{r}$ by $\mathbf{r} - \mathbf{R}$ (notice the sign). (We have just used a 'hidden' column approximation.) Equation 25.24 is then written as

$$b(\mathbf{k}) = \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k}) \exp(2\pi i(\mathbf{k} + \mathbf{g}) \cdot (\mathbf{r} - \mathbf{R})) \tag{25.25}$$

$$b(\mathbf{k}) = e^{-2\pi i \mathbf{k} \cdot \mathbf{R}} \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k}) e^{(-2\pi i \mathbf{g} \cdot \mathbf{r})} e^{2\pi i(\mathbf{k}+\mathbf{g}) \cdot \mathbf{r}} \tag{25.26}$$

$C_{\mathbf{0}}$ is not affected by $\mathbf{R}$ since than $2\pi \mathbf{0} \cdot \mathbf{R} = 0$, but $C_{\mathbf{g}}$ is multiplied by $e^{-i\alpha}$.
- If you choose the coordinates appropriately then C is a unitary matrix. In this case, you can find C^{-1} just by reflecting across the diagonal and taking the complex conjugate of each term. This *trick* will allow you to express equation 25.23 explicitly, as given by Hirsch et al. (omitting a phase factor).

$$\begin{aligned} \phi_{\mathbf{0}}(t) = &[\cos(\pi \Delta k t) - i \cos(\beta) \sin(\pi \Delta k t)] \\ &+ \frac{1}{2}\left(e^{i\alpha} - 1\right) \sin^2 \beta \cos(\pi \Delta k t) \\ &- \frac{1}{2}\left(e^{i\alpha} - 1\right) \sin^2 \beta \cos(\pi \Delta k t') \end{aligned} \tag{25.27}$$

$$\begin{aligned} \phi_{\mathbf{g}}(t) = &\, i \sin(\beta) \sin(\pi \Delta k t) \\ &+ \frac{1}{2} \sin \beta \left(1 - e^{(-i\alpha)}\right) [\cos \beta \cos(\pi \Delta k t) - i \sin(\pi \Delta k t)] \\ &- \frac{1}{2} \sin \beta \left(1 - e^{(-i\alpha)}\right) [\cos \beta \cos(\pi \Delta k t') - i \sin(\pi \Delta k t')] \end{aligned} \tag{25.28}$$

In equations 25.27 and 25.28, t' is the distance of the fault below the center of the slice, i.e., we define $t' = t_1 - t/2$ where t_1 lies between 0 and t. (It's a good, but tedious, exercise to derive these equations for yourself.) The right-hand side of equations 25.27 and 25.28 each contains three terms.

- The first term is just what we found in Chapter 13 where the phase factor $\alpha = 0$, i.e., it's just like the perfect crystal.
- The second term is independent of the position of the planar fault because it doesn't depend on t'.

- The third term depends on t' such that both ϕ_0 and ϕ_g change with a periodicity in t' given by Δk^{-1}. So these amplitudes show the same dependence on ξ_g^{eff}. They will both show thickness variations.

THE GEOMETRY
You should keep in mind that we derived these equations for a planar defect lying parallel to the parallel surfaces of our specimen and normal to the beam.

We can now take these ideas and apply them to planar defects which are inclined to the surface, by calculating the contrast for all values of t' between 0 and t. The important points to remember are

- The model used in the calculation was a flat interface parallel to the surface of a plate-like specimen. You'll see fault fringes when t' varies across the fault but you don't usually have to consider the fact that either the surface or the fault may be inclined to the beam.
- The concept of the scattering matrix allows you to identify very clearly the effect of the defect on ϕ_0 and ϕ_g.

25.5 STACKING FAULTS IN FCC MATERIALS

We'll begin our discussion of actual examples with the SF in fcc materials. Before we discuss the details of contrast from SFs in fcc materials, we'll summarize the important results which hold for all planar defects

- The appearance of the image depends on the specimen thickness.
- Pairs of BF/DF SF images are not generally complementary even though we are using a two-beam approximation. Compare to the complementary behavior of the thickness fringes discussed in Chapter 24.
- Planar defects always really have a thickness. We'll illustrate this concept using overlapping faults in fcc materials (see also Section 26.6).

NOT ALL MATERIALS ARE FCC
Do not assume all faults are the same as in fcc materials!

25.5.A Why fcc Materials?

There are several reasons for emphasizing the analysis of stacking faults in fcc crystals

- Many important materials are fcc, including the metals Cu, Ag, Au, and austenitic stainless steel, and the semiconductors Si, Ge, and GaAs.
- Most of the analysis of SFs derives from the study of fcc materials.
- The translations are well known and directly related to the lattice parameter: **R** is either $1/6<\bar{1}\bar{1}2>$ or $1/3<111>$. Notice that these definitions differ by the lattice vector, $1/2<110>$. (Actually there may be small deviations from these ideal values but we'll ignore them for now.)

We want to learn how to extend this analysis to other fault vectors and avoid making unfounded assumptions when we do extend it. The geometry often encountered when studying defects in fcc materials is shown in Figure 25.3.

You should note that $(11\bar{1})$ is one of three possible planes for an inclined SF. In this case, the translation at the stacking fault will be $\mathbf{R} = \pm 1/3[11\bar{1}]$; the phase factor, α, is $2\pi\,\mathbf{g}.\mathbf{R}$. If we form an image with the $\mathbf{g} = (2\bar{2}0)$ reflection strongly excited then $\mathbf{g}.\mathbf{R} = 0$ and the fault is out of contrast in both BF and DF. If instead we use the reflection $\mathbf{g} = (02\bar{2})$ then $\mathbf{g}.\mathbf{R} = 4/3$ or $-4/3$ and $\alpha = 8\pi/3 = 2\pi/3 = 120°$ or $-8\pi/3 = -2\pi/3 = -120°$ (modulo 2π in each case). Notice that if the stacking fault lies parallel to the surface of this (111)-oriented specimen, you must tilt the specimen to see *any* contrast from the SF, i.e., $\mathbf{g}.\mathbf{R} = 0$ for all values of **g** lying in the fault plane.

INVISIBILITY CRITERION
When $\mathbf{g}.\mathbf{R} = 0$ the defect is 'out of contrast'; it's invisible.

Figure 25.4A and B shows two typical BF/DF pairs of $\pm\mathbf{g}$ strong-beam images from the same SF. In the BF images, the outer fringes are the same on both sides of the fault, both gray or both white while in the DF images one outer fringe is white but the other is gray. The questions which arise are

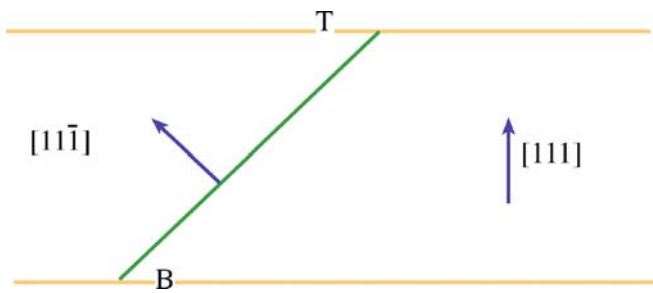


FIGURE 25.3. A stacking fault in a parallel-sided fcc specimen. The normal to the specimen is [111] and the normal to the SF is [11$\bar{1}$]. T and B indicate the top and bottom of the foil.

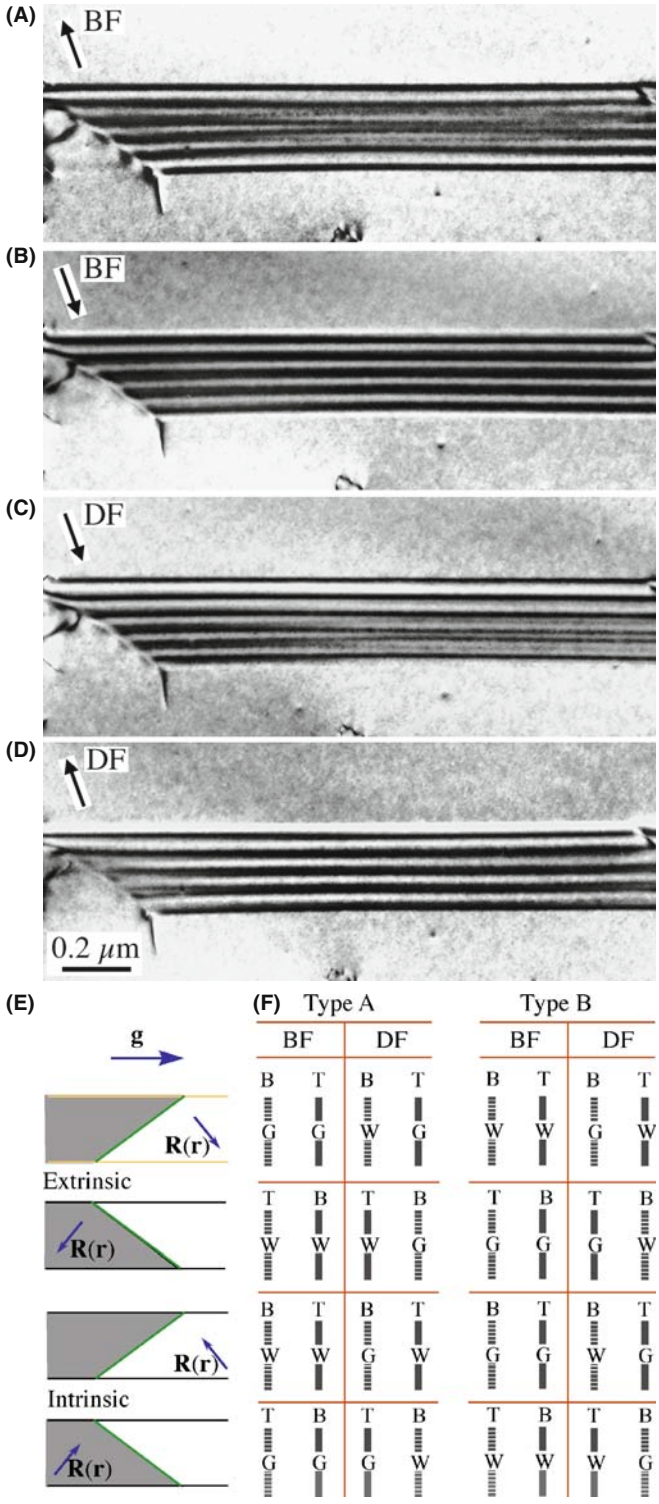


FIGURE 25.4. (A–D) Four strong-beam images of an SF recorded using $\pm\mathbf{g}$ BF and $\pm\mathbf{g}$ DF. The beam was nearly normal to the surfaces; the SF-fringe intensity is similar at the top surface but complementary at the bottom surface. The rules are summarized in (E) and (F) where G and W indicate that the first fringe is gray or white; (T, B) indicates top/bottom.

- What determines whether a fringe will be gray or white?
- Why are the two images not complementary?

Note that in the summary schematic shown in Figure 25.4E, 200, 222, and 440 are type A reflections while 111, 220, and 400 are type B reflections. Also we'll call the first fringe gray (G), not black (i.e., B), to reduce confusion!

25.5.B Some Rules

There are some experimental rules

- Be very careful when you record such a pair of images: record the DP for each image. Be sure to note which of the two bright spots corresponds to the direct beam.
- Use the same strong *hkl* reflection for BF and DF imaging. Therefore, to form the CDF image using a strong *hkl* reflection, you must first tilt the specimen so that $\bar{h}\bar{k}\bar{l}$ is strong and then use the beam tilts to move *hkl* onto the optic axis where it will become strong (see Section 22.5). This is confusing, so we recommend that you sacrifice a little image resolution and compare the BF image with a displaced-aperture DF (DADF) image, rather than a CDF image.

This is exactly the opposite of the approach used by Edington, who advocates tilting in $\bar{h}\bar{k}\bar{l}$ for the DF image, which reverses the DF contrast in Figure 25.4E and F. Our approach ensures that diffraction from the same *hkl* planes causes the contrast in both DF and CDF images but DADF beats both! (TEMs were not as good in 1975 as they are now.)

Then there are some rules for interpreting the contrast

- In the image, as seen on the screen or on a print, the fringe corresponding to the top surface (T) is white in BF if $\mathbf{g}\cdot\mathbf{R}$ is > 0 and black if $\mathbf{g}\cdot\mathbf{R} < 0$.
- Using the same strong *hkl* reflection for BF and DF imaging, the fringe from the bottom (B) of the fault will be complementary whereas the fringe from the top (T) will be the same in both the BF and DF images.
- The central fringes fade away as the thickness increases. If this seems anomalous, the explanation is in Section 25.10.
- The reason it is important to know the sign of $\mathbf{g}$ is that you will use this information to determine the sign of $\mathbf{R}$.
- For the geometry shown in Figure 25.3, if the origin of the $\mathbf{g}$ vector is placed at the center of the SF in the DF image, the vector $\mathbf{g}$ points away from the bright outer fringe if the fault is extrinsic and toward it if it is intrinsic (200, 222, and 440 reflections); if the reflection is a 400, 111, or 220 the reverse is the case.

- Don't forget that, as we said at the start of Chapter 22, any contrast must be >~5–10% to be visible to the eye, so we traditionally say an intensity change due to $\mathbf{g}.\mathbf{R}$ effects is only detectable if $\mathbf{g}.\mathbf{R} > 0.02$; of course, now you can digitally record the TEM image or digitize your analog version and then process this so the limit could be lower. With experience, you'll find there is an optimum thickness to view defect contrast, before absorption effects make it difficult. You must also carefully select **s** so that the background intensity in the matrix around the defect is gray and this maximizes visibility of lighter and darker fringes.

DISPLACE THE APERTURE

To avoid the possibility of confusion about which reflection to use for the DF image, just displace the aperture to the strong *hkl* reflection in every case. In a modern IVEM, there is almost no loss of resolution between DADF and CDF. DADF is always preferred but CDF is used to give much better images.

As we said, these complex rules are summarized in Figure 25.4C and D. Although they are very useful, in practice you should remember that they were derived for a very special combination of **R** and **g** in fcc materials. Some important examples of $\mathbf{g}.\mathbf{R}$ are given in Table 25.2. As we'll describe in Section 25.11, if this is your research, you should use a computer program to check the contrast.

INVISIBILITY

Just because you can't see a defect doesn't mean it isn't there or that $\mathbf{g}.\mathbf{R} = 0$.

25.5.C Intensity Calculations

Now let's consider intensity calculations using the column approximation, which we briefly discussed in Section 13.11. If the fault cuts the column at a depth t_1 we can deduce from equations 25.22 and 25.23 that

$$\phi_{\mathbf{g}} = \frac{i\pi}{\xi_{\mathbf{g}}}\left\{\int_0^{t_1} e^{-2\pi i s z}dz + e^{-i\alpha}\int_{t_1}^{t} e^{-2\pi i s z}dz\right\} \qquad (25.29)$$

TABLE 25.2. Values of $\mathbf{g}.\mathbf{R}$ for Some Common R Plus g Combinations

	R	g	$\alpha = 2\pi\,\mathbf{g}.\mathbf{R}$ (mod 2π)
SF in fcc	$\frac{1}{3}[111]$	(111), (220), (113)	$2\pi/3$
SF in fcc	$\frac{1}{3}[111]$	(113)	$4\pi/3$
Translation at APB in Fe_3Al	$\frac{1}{2}[110]$	(100)	π
Small **R**, e.g., NiO	Any	**g** or **s** or $\xi_{\mathbf{g}}$ differ slightly	δ

which gives

$$\phi_{\mathbf{g}} = \frac{i\pi}{s\xi_{\mathbf{g}}} e^{-2\pi i s t_1}\left\{\sin(\pi s t_1) + e^{-i\alpha}\sin(\pi s(t-t_1))\right\} \qquad (25.30)$$

We rearrange equation 25.30 to give an expression for the intensity, $I_{\mathbf{g}}$ $(=\phi_{\mathbf{g}}\cdot\phi_{\mathbf{g}}^*)$. This rearrangement involves a little manipulation.

$$I_{\mathbf{g}} = \frac{1}{(s\xi_{\mathbf{g}})^2}\left\{\sin^2\left(\pi s t_1 + \frac{\alpha}{2}\right) + \sin^2\left(\frac{\alpha}{2}\right) - \sin\left(\frac{\alpha}{2}\right)\sin\left(\pi s t + \frac{\alpha}{2}\right)\cos(2\pi s t')\right\} \qquad (25.31)$$

where $t' = t_1 - t/2$ as before. So the contrast depends on both the thickness and the depth. Note that $t/2$ is the center of the foil. Since α is fixed for a particular defect, let's fix t. Then equation 25.31 becomes

$$I_{\mathbf{g}} \propto \frac{1}{s^2}\{A - B\cos(2\pi s t')\} \qquad (25.32)$$

Now we have cosine depth fringes or defect thickness fringes, just as we did for the perfect crystal

- The thickness periodicity depends on s^{-1}.
- The intensity varies as s^{-2}.

We could have derived this equation from equation 25.28 with more work. However, the value of the scattering matrix approach is that we don't derive the analytical expression but just run the computer program.

In Chapter 27, we will discuss this SF contrast in terms of phasor diagrams which give a graphical way to represent these equations.

25.5.D Overlapping Faults

It is interesting to extend this analysis to the case of overlapping faults. Taking the analytical approach, we can extend equation 25.29 to the case of two overlapping faults, the first at depth t_1 and the second at depth t_2.

$$\phi_{\mathbf{g}} = \frac{i\pi}{\xi_{\mathbf{g}}}\left\{\int_0^{t_1} e^{-2\pi i s z}dz + e^{-i\alpha}\int_{t_1}^{t_1+t_2} e^{-2\pi i s z}dz + e^{-i(\alpha_1+\alpha_2)}\int_{t_1+t_2}^{t} e^{-2\pi i s z}dz\right\} \qquad (25.33)$$

An experimental illustration of a somewhat more complex situation, involving several overlapping SFs, is shown in Figure 25.5

- It sometimes appears that there is no contrast, even when we know that there are overlapping SFs. This can happen if, e.g., three SFs overlap on adjacent (or nearly adjacent) planes; then the effective R can

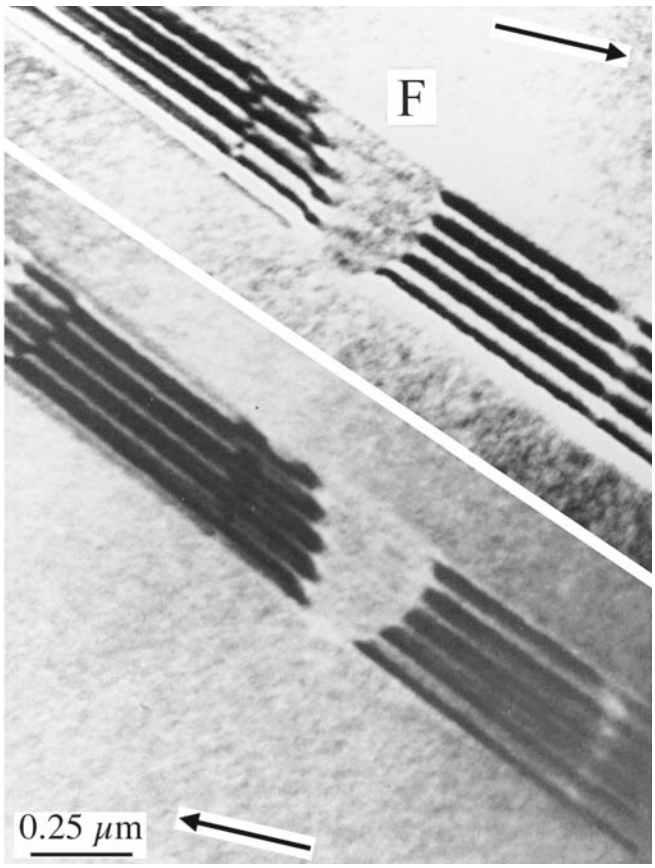


FIGURE 25.5. Two BF images of overlapping SFs in fcc steel with the direction **g** indicated. The faults are very close together. When three faults overlap at F the effective value of **R** is 0 so the contrast disappears.

be $= 3 \times \frac{1}{3}[11\bar{1}]$, which is a perfect lattice vector and can therefore appear to give $2\pi\mathbf{g}\cdot\mathbf{R} = 0$.

We will return to this topic in Section 27.8, where we'll show that some planar defects, such as the extrinsic SF in Si, or the dissociated {112} twin boundary in some fcc metals, really have a significant thickness. We can then analyze the contrast from such interfaces using the overlapping-fault model.

25.6 OTHER TRANSLATIONS: π AND δ FRINGES

We discussed the $L1_0$ structure of NiAl in Section 16.5. This intermetallic is an example of a large group of materials which can contain a different type of RB. If the Ni atoms sit at the corners of the cell in one crystal region but the Al atoms sit at the corners in another part of the crystal, then the two crystal regions are related by a translation of $\frac{1}{2}[111]$. The two crystals would otherwise still be perfectly aligned but are separated by this RB, which we call an APB (but which could be called an SF or an IDB).

Similarly in the $L1_2$ structure of the intermetallic Ni_3Al, we could have the Al atoms on the corners of the unit cell in one part of the crystal and displaced by $\mathbf{R} = \frac{1}{2}[110]$ in the adjacent region. (We can actually form six non-equivalent APBs in this structure and these are *not* IDBs.) The crystal structure looks like fcc but the Al atoms are at the corners of the unit cell (forming the simple-cubic superlattice) with the Ni at the face-centered positions. The easy way to appreciate this RB is to think what would happen if the alloy were completely disordered: there would be no planar defect. This RB can be imaged using the (100) reflection. Notice again that for a disordered structure, the {100} reflections would be absent if the alloy were disordered; the {100} planes are said to give rise to superlattice reflections; these reflections would be forbidden if the material were disordered. For this case we can readily show that the phase factor $\alpha = \pi$, so the fringes we see are called π fringes. The structure of this interface is shown schematically in Figure 25.6. These π fringes can give symmetric fringes in DF and BF and complementary DF/DF pairs.

SUPERLATTICES

We usually use the term *superlattice* in the sense that it is the *real* lattice, because we have ordering of the components. If the same material was disordered (but with the atoms on the same sites), the lattice parameters, in real space, would be smaller.

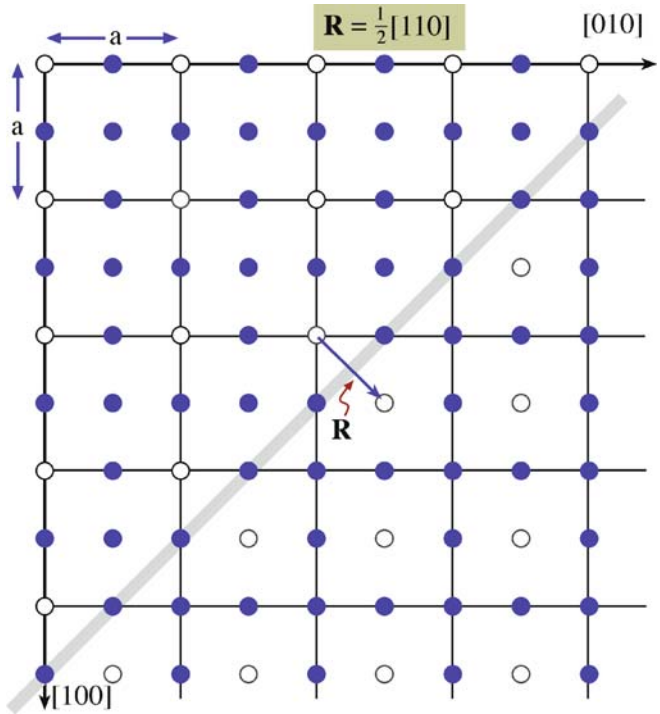


FIGURE 25.6. Schematic of an interface in the intermetallic Ni_3Al showing how the two structures link coherently but are displaced by the vector **R**. The phase factor at such an interface is π and the fringes seen in the image are called π fringes.

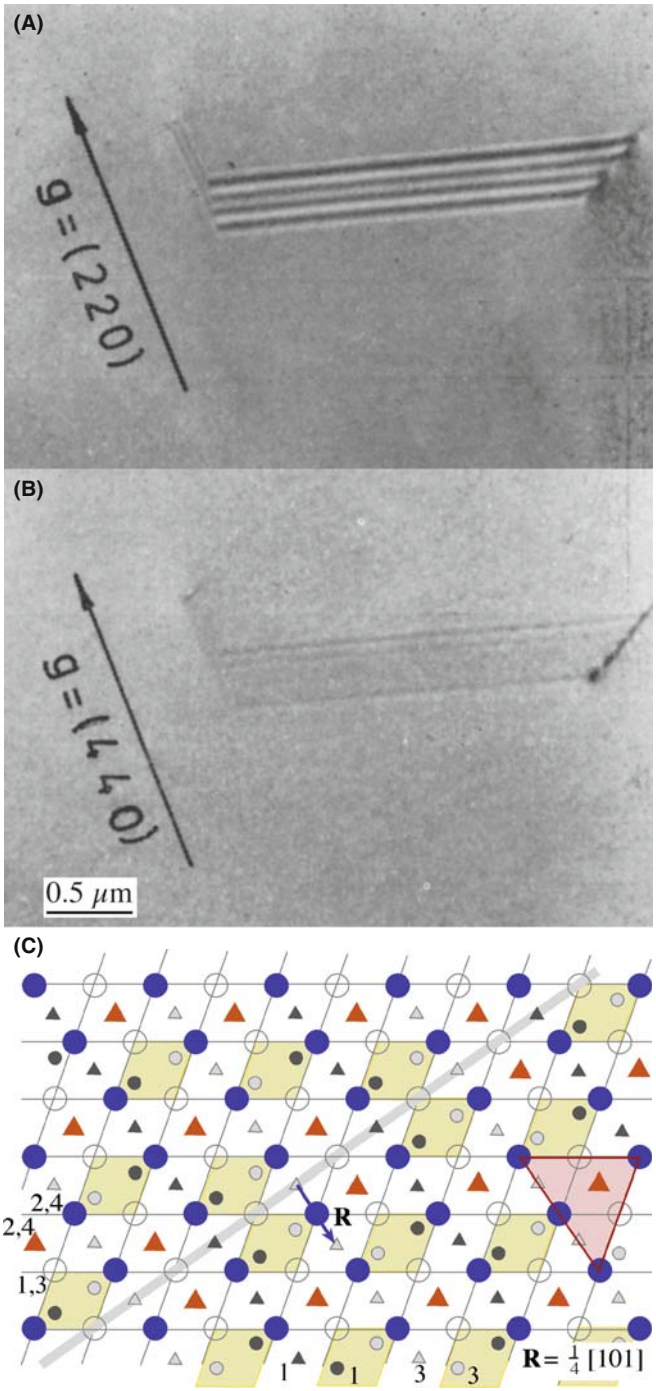


FIGURE 25.7. (A, B) Pair of BF images and (C) schematic of an SF in spinel; this interface is also known as an APB since the SF translation vector is a perfect (sub)lattice vector in the underlying fcc oxygen sublattice as shown in (C). Large circles are O anions at different heights (1, 3) and (2, 4); the small circles and triangles are cations at different heights (1–4 as indicated). (D) APBs can overlap just as SFs can as shown by these faceted APBs in TiO_2. Many of the facets give quite similar contrast but those near the center are strikingly different because of overlap. The schematic (E) shows a series of APBs, each of which is formed by a translation, $\mathbf{R} = {}^1/_2[011]$ which has little effect of the oxygen sublattice (open circles).

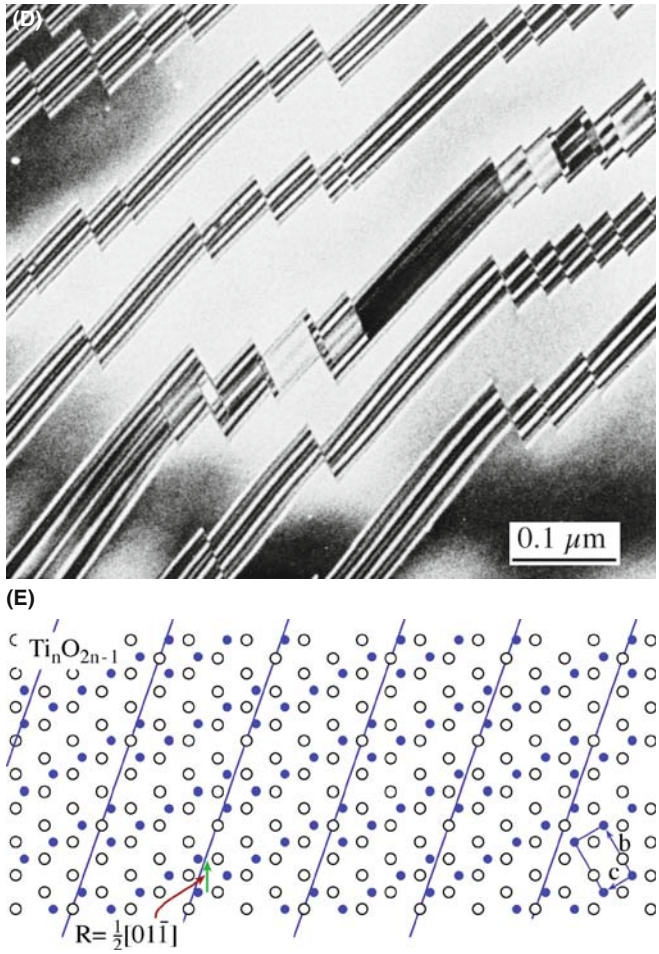


FIGURE 25.7. (Continued).

Similar RBs are very common in oxides because the unit cell is often quite large, giving more opportunities to form such interfaces. The interface shown in Figure 25.7A and B has been called both an SF and an APB in spinel. These interfaces can show all the features we discussed in Section 25.5 for SFs in fcc materials, and those we've just discussed depend on which reflection you use. You can again see a change in contrast in Figure 25.7D when APBs in TiO_2 overlap as shown schematically in Figure 25.7E (Amelinckx and Van Landuyt 1978). In Figure 25.7A and B, if you image the fault using the 220 reflection $2\pi\mathbf{g}\cdot\mathbf{R} = \pi$ and so you'll see SF fringes. If, however, you image using 440, $2\pi\mathbf{g}\cdot\mathbf{R} = 2\pi$, so you'll only see residual contrast (because **R** is not exactly $\frac{1}{4}[101]$.

The APB shown in Figure 25.8A is different yet again. This planar defect in GaAs is also known as an IDB (Section 25.1). The fringes you see are caused by a translation, but **R** is not related in a simple way to the structure of the crystal or to the inversion symmetry. Rasmussen et al. showed that the translation is present because there is a small relaxation of the Ga–Ga and As–As bonds at this {110} interface. The value of **R** was determined to be 0.19 Å with a statistical uncertainty of ±0.03 Å, so it's a translation that's almost parallel to the

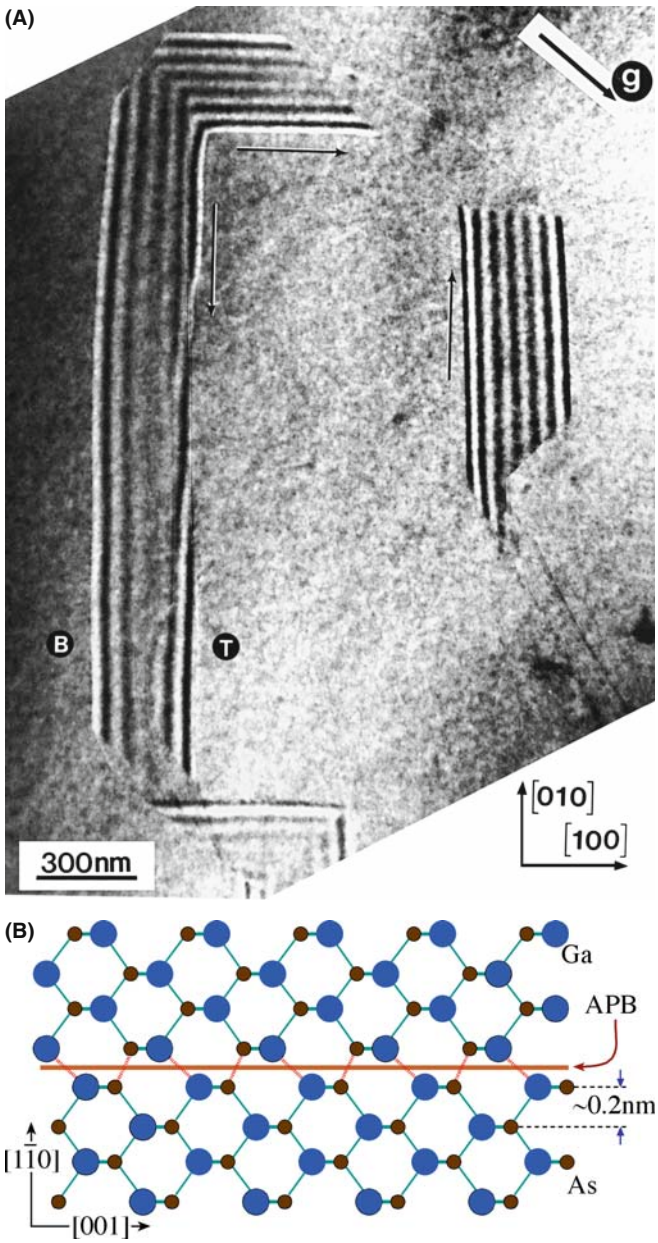


FIGURE 25.8. (A) A faceted APB (or IDB) in GaAs with (B) a schematic of the (110) facet. The translation is caused by the difference in length of the Ga–Ga and As–As bonds and is much smaller than a lattice vector in the GaAs lattice.

interface, as shown schematically in Figure 25.8B. These fringes are then known as δ fringes (because there is only a small translation and α is not related in a simple way to 2π). The use of image-simulation programs, which are necessary to determine **R** (remember that the wavelength of the 200-kV electrons used for the measurement is itself 0.025 Å or 2.5 pm), is discussed in Section 25.13.

25.7 PHASE BOUNDARIES

We'll list a few special phase boundaries in Table 25.3.

TABLE 25.3 Examples of Special Phase Boundaries

Boundary	Example of material	Features
Ferromagnetic domain boundaries	NiO	
Ferroelectric and piezoelectric boundaries	$BaTiO_3$	Small tetragonal distortion
Composition boundary	GaAs/AlGaAs	$\xi_{\mathbf{g}}$ is different on two sides of boundary, even for perfect lattice matching
Structure boundaries	α-SiC/β-SiC hcp-Co/fcc-Co	
Composition/structure	NB/Al_2O_3 Al/Cu α-Fe/Fe_3C	

An example of a PB is shown in Figure 25.9. In NiO, which is ferromagnetic, some of the planes rotate slightly when the structure changes from cubic symmetry below the Curie temperature. Now we can also define the cubic structure as rhombohedral with $\alpha = 60^\circ$ in the rhombohedron. Below the Curie temperature, the rhombohedral angle is distorted by only 4.2′ (yes, minutes of arc) from the true 60°. Therefore, most **g**-vectors will rotate through a very small angle and hence produce a change in the value of **s**. However, as you can see in Figure 25.9, this small rotation can readily be detected by the change in contrast and the faint fringes at the phase boundary.

We can have overlapping PBs, so, the warning is the same: be very wary and use tilting experiments while you are at the TEM.

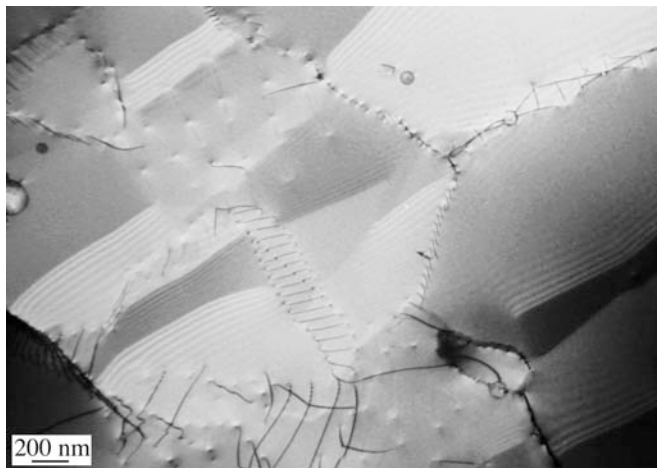


FIGURE 25.9. The ferroelectric material NiO undergoes a structural change from cubic to distorted rhombohedral at the Curie temperature. Although the distortion in the rhombohedral structure is very small, it does cause a detectable rotation of the lattice planes that results in the δ fringes in the image.

25.8 ROTATION BOUNDARIES

What can we learn about rotation boundaries when the rotation angle is greater than about 0.1°? Unfortunately, the answer is "not a lot," unless we have defects which accommodate the rotation. Then we are into the subject of diffraction contrast of line defects in interfaces. However, with care you may be able to excite **g** in one grain or in both by tilting the specimen. The difficulty, of course, is that $\mathbf{s_g}$ is likely to be different in each material. Complications will also arise if other defects are present, since you may or may not see those defects. Examples of such inclined, rotation interfaces are shown schematically in Figure 25.10A and experimentally in Figure 25.10B and C.

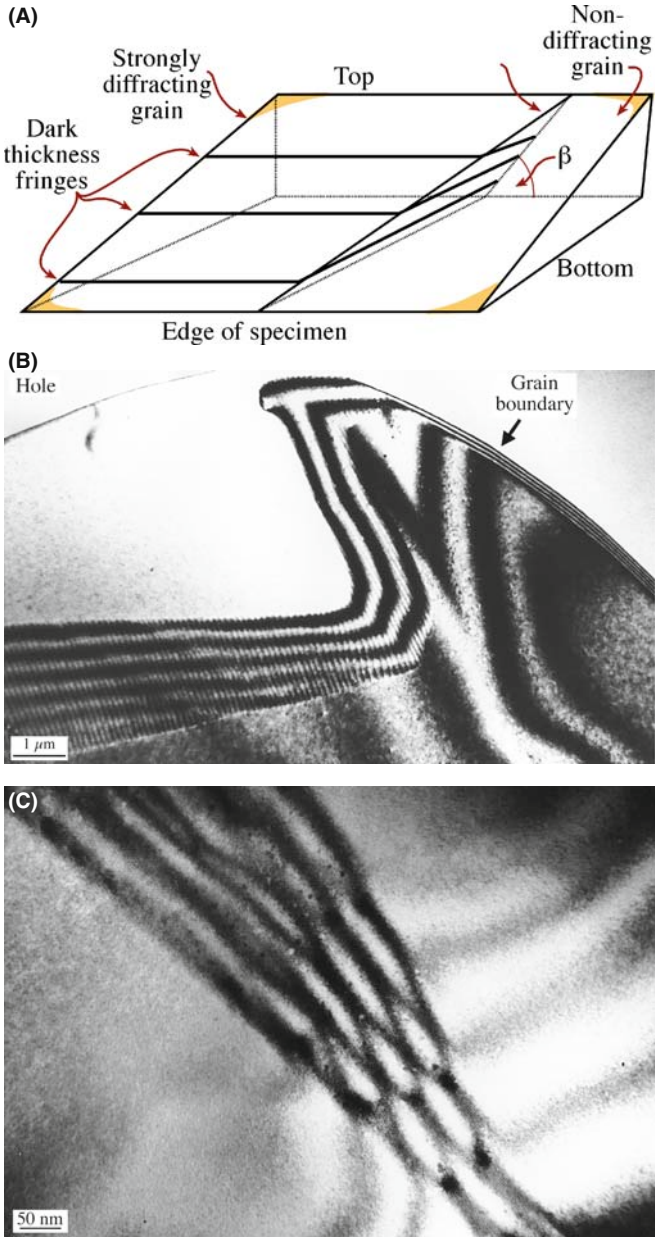


FIGURE 25.10. (A) If the adjoining grains are rotated so that they do not share a common reflection, images can be formed where only one of the grains diffracts. As shown in (B), the thickness fringes associated with the wedge merge into the thickness fringes associated with the inclined interface. (C) If the foil is tilted so that the same (though not coincident) reflection is excited in both grains, the number of fringes in the interface increases with each incremental increase in the wedge thickness.

25.9 DIFFRACTION PATTERNS AND DISPERSION SURFACES

You read in Chapter 17, that what you see in an image must be related to what happens in the DP which in turn is determined by how the Ewald sphere intersects the reciprocal lattice. Figure 17.5 showed that a planar defect which is inclined to the surface of a parallel-sided specimen will give rise to relrods. Therefore, a planar defect that is inclined to the surface of a parallel-sided specimen will produce at least two spots in the DP. Since most specimens are wedges (see Figure 17.4), and the planar defect will, in general, be inclined to both surfaces, the relrod geometry is actually even more complex. Figure 25.11 shows lines normal to each interface and their associated relrods. You can appreciate that when the Ewald sphere cuts these relrods, several spots may appear in the DP. Now we need to relate these relrods to the fringes we see in the image. This model would predict that we would not produce fringes when $\mathbf{s} = 0$, so we should modify what we did in Figure 17.15. The periodicity of the fringes in the image is inversely related to the distances (M_1N and M_2N) between the spots in the DP.

FRINGE SPACING

At $\mathbf{s} = 0$, the thickness fringe spacing is determined by $\Delta k = \xi_{\mathbf{g}}^{-1}$ and the wedge angle: the spacing of the fringes is proportional to $\xi_{\mathbf{g}}$.

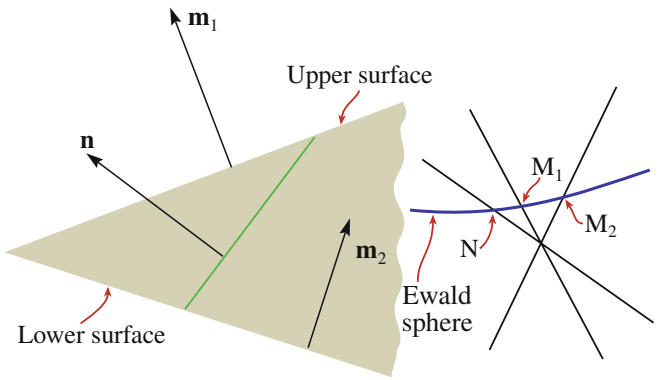


FIGURE 25.11. In a wedge specimen, the planar defect will, in general, be inclined to both surfaces and the relrod geometry is complex. The fringe spacing in the image is related to the reciprocal of the distances M_1N and M_2N.

When a planar defect is present in the specimen, the two branches of the dispersion surface are not only coupled along a tie-line normal to the surface of the specimen but also along the normal to the planar defect. However, when $\mathbf{s} = 0$, the thickness periodicity in the image corresponds to the extinction distance. When we relate this to the region G in the reciprocal lattice, the two relrods (which result from a kinematical construction) must actually separate to give the two hyperbolas shown in Figure 25.12 which is why we drew Figures 17.15 and 23.6 as we did.

FRINGES IMPLY SPOTS
If you see fringes in the image, spots will be present in the DP.

The spots in the DP are associated with points M and N in Figures 25.11 and 25.12.

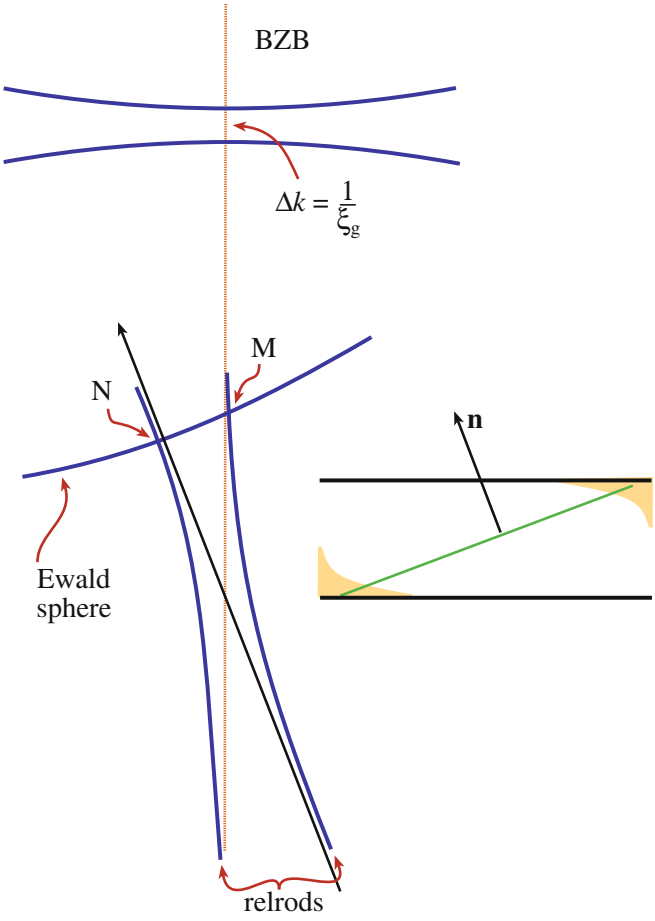


FIGURE 25.12. The dispersion-surface construction for an inclined planar defect in a parallel-sided specimen. (Compare Figures 17.15 and 24.6.) For simplicity, we show the hyperbolas due to the defect alone, not the extra effects that would arise in a wedge specimen.

25.10 BLOCH WAVES AND BF/DF IMAGE PAIRS

In Chapter 14 we saw that, in a crystal, the electron must propagate as Bloch waves, and yet we have not mentioned Bloch waves in our discussion of thickness and bending so far. Most of the analysis of this topic is beyond the scope of this text, but it is important to understand the basic ideas, particularly since they will also apply to scattering from defects in the crystals. Remember that ξ_g is a direct consequence of having two Bloch waves. The important message here is: don't let the words overawe you.

The idea is quite simple. Since we have two Bragg beams excited, then we must have two Bloch waves in the crystal. The propagation vectors of these two waves are $\mathbf{k}_1$ and $\mathbf{k}_2$ with the difference $|\Delta\mathbf{k}|$ being related to s_{eff} (and $\gamma_1 - \gamma_2$ in Section 13.10). We see a thickness dependence in the image because the two waves are interfering. The only two waves which are really present in the crystal are the two Bloch waves. It's the beating of these two waves which gives rise to thickness effects.

In the two-beam case, the Bloch waves, 1 and 2, are channeled along and between the atom columns (see Figure 14.2). A fault may change the channeled wave into the non-channeled one, as you can see in Figure 25.13. So, the effect of the planar defect is to couple the

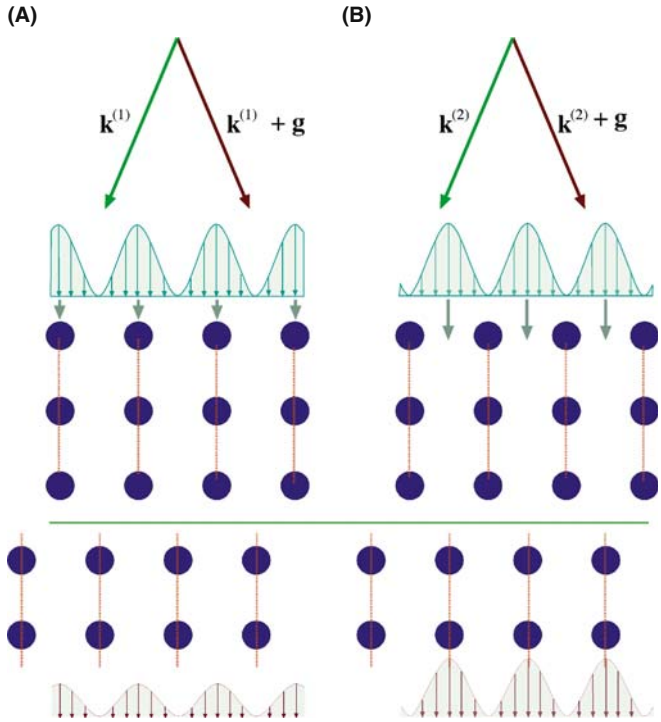


FIGURE 25.13. Bloch waves 1 and 2 are channeled (A) along and (B) between the atom columns, respectively, until they meet the fault. There the atomic columns are translated so that the channeled Bloch wave may become the non-channeled one and vice versa.

Bloch waves; in other words, the defect links (or connects or ties) the different branches of the dispersion surface (along the tie line). The non-complementary contrast at SFs in fcc metals is directly explained by this coupling.

As soon as the beam enters the specimen we excite Bloch waves 1 and 2. Therefore, in Figure 25.14A, the two Bloch waves 1 and 2 are shown everywhere at the top surface of the foil. The planar defect links points D_1 and D_2 on the two branches of the dispersion surface, as shown in Figure 25.14B, along the tie line, D_1D_2' and $D_1'D_2$. We'll analyze the three situations shown in Figure 25.14A, which correspond to the planar defect being close to the top, the middle, and the bottom of the specimen. The key feature is that, as we saw in Section 14.6, Bloch wave 1, which has the larger **k**-vector, will be preferentially absorbed. It is actually totally absorbed in thicker specimens.

- When the planar defect is close to the top surface (as occurs near T in Figure 25.14A), waves 1 and 2 are both coupled (or scattered) to the other branch of the dispersion surface so we form four Bloch waves (but with only two **k**-vectors). Both Bloch waves which are associated with the upper branch of the dispersion surface (wave vector $\mathbf{k}_1$) are preferentially absorbed, but the waves D_2 and D_2' both reach the lower surface. There, they interfere to give the thickness fringes even though they are both associated with the lower branch of the dispersion surface; D_2' retains a 'memory' of D_1.
- When the fault is close to the middle of the specimen in Figure 25.14A, the branch-1 Bloch wave is absorbed before it reaches the planar defect but a new Bloch wave D_1' is formed at the defect. However, while traversing the other half of the foil, this wave is also absorbed so that only wave D_2 reaches the lower surface. Thus the electrons can propagate through the specimen (we can see through it) but there are no thickness fringes because only one Bloch wave survives. However, we can still image defects in these thicker areas, as you'll see if you look back at Figure 23.10.
- At the lower surface, B in Figure 25.14A, only wave D_1 survives to reach the planar defect but it now produces a new wave D_1' which can reach the lower surface, recombine with Bloch wave D_2, and produce thickness fringes. The resulting contrast is summarized in Figure 25.14C.

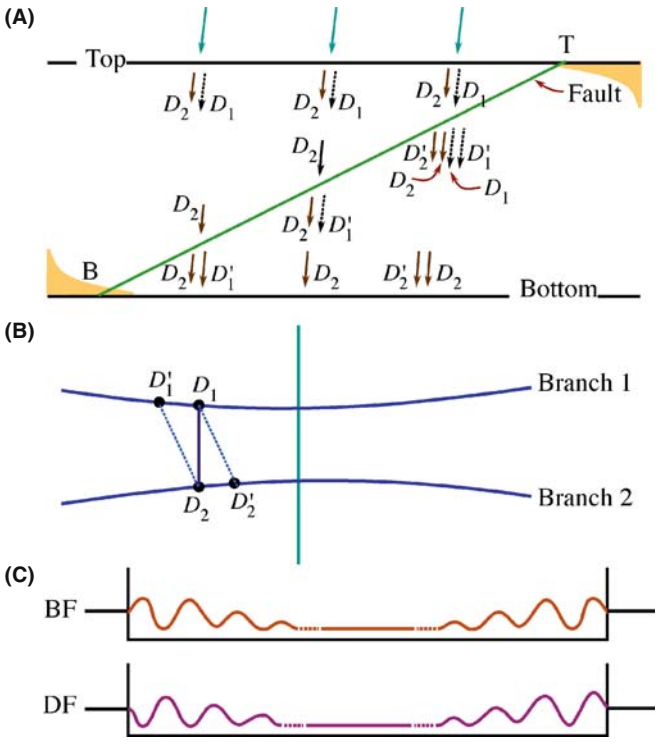


FIGURE 25.14. The absorption of the branch-1 Bloch wave near the top surface, T, of the specimen and its re-creation when the planar defect is near the bottom, B, determines the contrast we see. (A) shows which Bloch waves are present at the different depths in the specimen, (B) shows how the Bloch waves are coupled along tie lines joining the two branches of the dispersion surface, and (C) shows the resulting contrast profiles.

Bloch-wave absorption is a critical factor in explaining the appearance of contrast from planar defects. The part of this argument which is not intuitive is the fact that D_2' retains a memory of D_1; this memory allows it to interfere with D_2 to produce the thickness fringes near the top of the specimen, even though no Bloch wave from branch 1 reaches the bottom of the specimen. We'll refer you to the article by Hashimoto et al. for further discussion on this topic.

25.11 COMPUTER MODELING

From the discussion in Sections 25.5 and 25.6, you will realize that α and π fringes are usually understandable as long as you know what the defect is and as long as it's not actually a set of overlapping defects. The contrast from δ fringes is much more complex and combinations of α, π, and δ fringes are difficult! The situation will become even more complicated if you want to understand the contrast occurring when other defects interact with these planar faults.

A computer program is then really the only way to analyze the contrast from these defects. The problem here is that many programs have been written for computers that now only exist in museums, by researchers who became professional programmers and never saw a TEM image again.

The first program to attempt the task of simulating two-dimensional images (rather than line profiles) of planar defects is described in the book by Head et al. Two current software packages are CuFour and

TEMACI. CuFour is the subject of an entire chapter in the companion text; TEMACI incorporated the ideas from Comis that we'll refer to here. We will mention some of the features of these programs to help you select one, but leave the detailed descriptions to the appropriate manuals. The most important reason for using any program must be your desire to understand the contrast and thus characterize the defect.

These programs are tools to assist you toward the goal of quantitative analysis of diffraction contrast, but you always need the fully quantitative experimental data, too. Few TEM users have ever collected such data.

You need an accurate simulation of the image you see in the microscope. The fact that the image varies with depth, thickness, **g**-vector, etc., is actually to your advantage, since you then have many variables, all of which you must be able to measure to achieve a good match with your experimental image. From the point of view of quantitative analysis, a one-dimensional line (intensity) profile is as valid as a two-dimensional image. Of course, if you can compare the contrast in an image and a simulation point by point, then you can have much greater confidence in the matching. The two-dimensional simulated image is also more viewer-friendly!

A great advantage of more powerful computers is that you can also test the effect of specimen geometry more readily. Thus, for example, Viguier et al. showed, using the CuFour simulation package, that the rules for fringe contrast given by Gevers et al. will not work if the specimen is tilted such that it intersects the bottom of the foil above the point where it intersects the top! You can understand this situation more easily by looking at the specimen geometry shown in Figure 25.15.

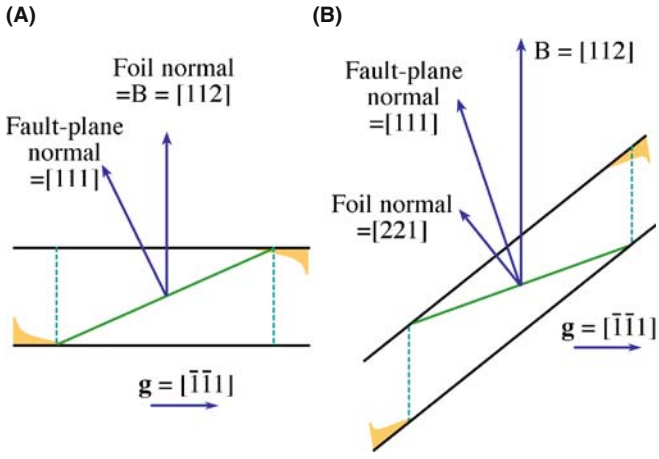


FIGURE 25.15. Many of the 'rules' for predicting the contrast from planar 'defects' make certain assumptions about the geometry of the defect relative to the surface of the specimen that may not always hold. The 'usual' situation is shown in (A) but the intersection of the planar defect with the upper specimen surface may be lower than the intersection with the bottom surface as in (B). This geometry can cause a reversal in the rules.

Image simulation tells us that $\mathbf{g}.\mathbf{R}$ must be > 0.02 to produce visible fringes, and you don't need to know the local structure at the planar defect when determining this condition. You could, in principle, detect smaller values of **R** by using larger **g**-vectors but in practice it then becomes more difficult to set up a well-defined diffraction geometry.

The next two sections are rather specialized and you may wish to leave them until much later, especially if you don't have access to a suitable program or until you are prepared to write your own. Do consult the key references and list of available programs in Section 1.6 before writing your own program. The subject is just as relevant to the topics of Chapters 26 and 27, but we include it in this chapter mainly because the analysis of planar defects is the most straightforward application.

25.12 THE GENERALIZED CROSS SECTION

Head et al. presented a method and a computer program for the computation of BF and DF images of line and planar defects. The source code is given in their book and is available from the WWW. You should note several important features of this program

- It uses the two-beam theory of electron diffraction.
- It uses the column approximation.
- The simulated image can be displayed as a halftone image rather than as intensity profiles.

This program was so successful, in part, because Head et al. were able to calculate the images quite quickly in spite of the fact that the computers available to microscopists were often not particularly powerful in ~1970. The calculations used a concept which they called the generalized cross section (GCS). (The GCS is not a scattering cross section, it is actually a slice through the specimen.) The GCS can be used when the displacement field, $u_{\mathbf{k}}$, satisfies the requirement that

$$u_{\mathbf{k}}(x, y, z) = u_{\mathbf{k}}(x, 0, z + cy) \qquad (25.34)$$

Here c is a constant and the foil is imagined to be laterally infinite. When this requirement is satisfied, the calculation of $u_{\mathbf{k}}$ is greatly simplified. One such situation is the important case where several dislocations and their associated fault planes are all parallel to one another. Then you only need to calculate the many-beam, Howie-Whelan equations on the plane $y = 0$. The displacement field for two columns y_1 and y_2 will only differ by a translation along the column, i.e., the z-direction. You don't want to repeat calculations you've already done; just calculate the image on a mesh in the x–y plane.

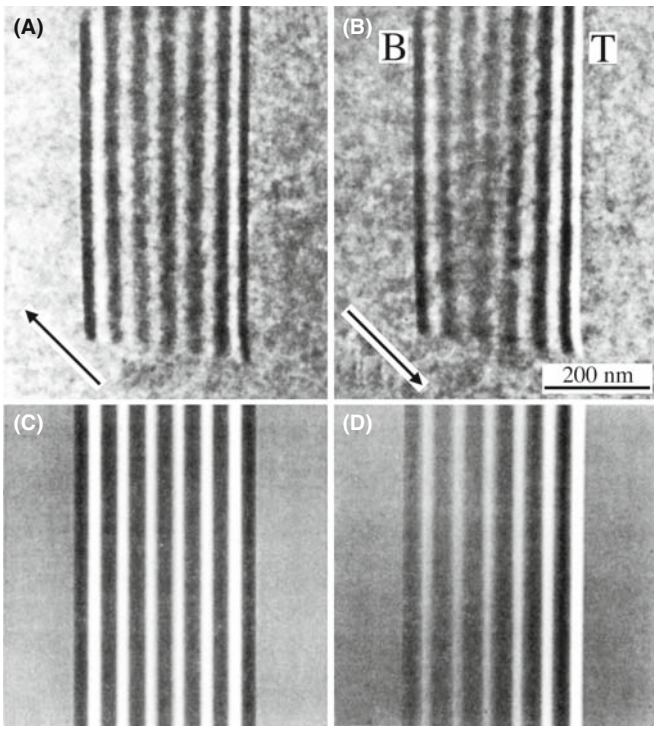


FIGURE 25.16. (A) Experimental BF image of an APB with $\mathbf{g} = 220$. (B) DF image of the same defect, $\mathbf{g} = \bar{2}20$. (C, D) Corresponding simulated images.

Examples of experimental and simulated images are shown in Figure 25.16. The simulation package used here (Comis again) was particularly attractive since it also performed elasticity calculations for simple defect configurations. It could simulate the effect of changing different parameters, and hence

- Change the accelerating voltage to see the extinction distances change.
- Change the absorption parameters to see the loss of SF-fringe contrast near the middle of the foil ($z = 0.5t$).
- Change the number of beams contributing to the image; how good is your two-beam assumption?
- Look at how reversing **g** changes the geometry of the image.
- Compare BF and DF as you vary the value of $\mathbf{s_g}$.

The Comis program ran on a Convex mini-supercomputer with an intelligent terminal, then on a Mac with OS6. Zhou used the concepts in the new and advanced package, TEMACI.

25.13 QUANTITATIVE IMAGING

One of the important applications of diffraction-contrast images is the detailed characterization of defects. With the improvements in the TEM, particularly in resolution and drift, we are now able to pay more attention to the fine structure of defects and this requires *quantitative* image analysis. In particular, we need to use the actual intensity levels in the image. One obstacle for quantitative analysis has been the uncertainty in the background-level intensity caused by inelastic scattering. As energy-filtered images become more widely available (see Chapter 37), we hope this problem will disappear. Direct digital recordings of the intensities, using a CCD camera are making quantitative analysis more tractable, eliminating uncertainties associated with the calibration of the response of the emulsion of the photographic film.

With these new applications of diffraction-contrast images in mind, improved simulation programs have become essential. An ideal program will be versatile, but user-friendly; it will allow us not only to calculate the image but also to give defect interaction geometries. In simulating images of crystal defects, you'll encounter several problems that are almost independent of one another. You must be able to do the following

- Define the geometry of the defects and of the specimen (the diffracting conditions).
- Calculate the displacement field associated with the defects.
- Propagate and scatter the electron beams throughout the foil (i.e., solve the Howie-Whelan equations).

We've already discussed the theoretical basis of the diffraction process, so we'll now illustrate some of the numerical methods which we can employ for different defect geometries. In Chapter 26, we'll consider other types of defects which may be analyzed, methods for defining them, and how we can calculate the displacement field.

25.13.A Theoretical Basis and Parameters

We'll use the Comis program as an example of the considerations that go into a simulation program. One message that you should certainly understand from this discussion is that you must be very cautious when using any program to simulate images. All such programs make assumptions and simplifications.

BLACK BOX

As always, when using the computer to simulate TEM images: beware of the black box. Don't automatically believe everything that comes out of it.

Comis was based on the Howie-Whelan dynamical theory of electron diffraction and thus neglects diffuse scattering but it could simulate without the column approximation. The basis of the approach used was given by Howie and Basinski.

- Use the deformable-ion approximation to describe how the crystal is influenced by the displacement field, **R**. In this model, the potential at **r** in the deformed crystal is assumed to be equal to the potential at the point **r** – **R**(**r**) in the perfect crystal. The model is good unless **R**(**r**) varies too rapidly.
- Extend the Howie-Whelan approach to many beams and avoid the column approximation.

The resulting equations are basically the same as those we derived in Chapter 13, so don't be put off by their appearance.

In treatments avoiding the column approximation, we include terms which allow for a variation in x and y: these terms were specifically excluded in Chapter 13.

The equations are now written as

$$\frac{\partial\phi_{\mathbf{g}}(r)}{\partial z} = \mathrm{i}\pi\sum_{\mathbf{h}}\left(\frac{1}{\xi_{\mathbf{g}-\mathbf{h}}}+\frac{\mathrm{i}}{\xi'_{\mathbf{g}-\mathbf{h}}}\right)\phi_{\mathbf{h}}e^{2\pi\mathrm{i}\left(\left(s_{\mathbf{h}}-s_{\mathbf{g}}\right)z-(\mathbf{h}-\mathbf{g})\,.\,\mathbf{R}\right)} - \theta_x\frac{\partial\phi_{\mathbf{g}}}{\partial x}-\theta_y\frac{\partial\phi_{\mathbf{g}}}{\partial y}+\frac{\mathrm{i}}{4\pi\chi_z}\left(\frac{\partial^2\phi_{\mathbf{g}}}{\partial x^2}+\frac{\partial^2\phi_{\mathbf{g}}}{\partial y^2}\right) \tag{25.35}$$

As usual, $\boldsymbol{\chi}$ is the incident-beam wave vector in vacuum, **g** is a particular diffraction vector, and **h** represents all the other possible diffraction vectors; you might like to compare this equation with equation 13.8! We have defined two new parameters to take account of the direction of the beam

$$\theta_x = \frac{(\boldsymbol{\chi}+\mathbf{g})_x}{\chi_z} \quad \text{and} \quad \theta_y = \frac{(\boldsymbol{\chi}+\mathbf{g})_y}{\chi_z} \tag{25.36}$$

The x–y plane in the reciprocal lattice contains the dominant reflections, and z is almost parallel to the incident beam. The number of beams you can include in the calculation is limited only by the capacity of your computer. The standard default for a program would be to select only beams on the systematic row. However, non-systematic beams can substantially influence your image, so it is useful if you can include sets of beams which are coplanar with the systematic row. We define the deviation of the crystal orientation from the exact Bragg condition by specifying the wave-vector components χ_x and χ_y; the latter applies when reflections outside the systematic row are included. Now we can calculate *all* the deviation parameters, $\mathbf{s}_{\mathbf{h}}$; there are many beams and each **s** can be different.

Each extinction distance $\xi_{\mathbf{g}}$ is defined as the ratio $\chi_{\mathbf{g}}/|U_{\mathbf{g}}|$, as usual with $U_{\mathbf{g}}$ being a Fourier component of the perfect-crystal potential. The Fourier components can be calculated from X-ray scattering factors, using the Mott expression. For most situations the scattering angle is small enough, so the X-ray scattering factors may, in turn, be calculated using the nine-parameter Gaussian fit given by Doyle and Turner.

You'll need to know the unit cell for your material and the Debye-Waller factor, B. Comis could then automatically calculate $\xi_{\mathbf{g}}$ using a built-in table of the Doyle-Turner parameters. You need to know B if you want to convert X-ray structure factors to electron structure factors (or vice versa) at a given temperature, or if you want to compare structure-factor measurements taken at different temperatures. When you calculate extinction distances, the Debye-Waller factors are essential to determine the effect of temperature.

DEBYE-WALLER FACTOR

B is related to the (mean-square) vibrational amplitude of an atom on a lattice site. It is a temperature-sensitive term.

Equation 25.35 is only valid if your crystal has a center of symmetry, otherwise we have to redefine $\xi_{\mathbf{g}}$. Simulations involving non-centrosymmetric crystals can be performed, but we have to replace the $\xi_{\mathbf{g}}$ and $\xi'_{\mathbf{g}}$ with complex quantities and then defining all the parameters is really difficult (see equations 14.2–14.8).

You can take account of absorption effects in the usual way by adding imaginary Fourier components, $U'_{\mathbf{g}}$ the absorption distances, $\xi'_{\mathbf{g}}$, are then defined as we discussed in Section 24.7. There are equations which a program can use to estimate the absorption distances, e.g., the linear relation $|U'_{\mathbf{g}}|/|U_{\mathbf{g}}| = a + b|\mathbf{g}|$, as suggested by Humphreys and Hirsch, or you could specify each individual absorption distance directly.

25.13.B Apparent Extinction Distance

The program developed by Head et al. was based on the two-beam approximation. The success of such calculations relies on the fact that the $\xi_{\mathbf{g}}$ may be replaced by an apparent extinction distance, $\xi^a_{\mathbf{g}} < \xi_{\mathbf{g}}$. This substitution compensates for scattering into beams that are not included in the two-beam calculation. The term $\xi^a_{\mathbf{g}}$ depends on t and must be estimated in each individual situation, e.g., by fitting the simulated image to your experimental image. For a quantitative image analysis it is important that you should have as few adjustable parameters as possible; using the many-beam program eliminates the need to use the parameter $\xi^a_{\mathbf{g}}$. Alternatively, you may determine $\xi^a_{\mathbf{g}}$ by comparing simulated thickness fringes calculated using many-beam and two-beam approximations.

25.13.C Avoiding the Column Approximation

Now you know we can perform simulations with or without the column approximation. With the column

approximation, you only keep the first term on the right-hand side of equation 25.35. The equations are reduced to a system of ordinary differential equations which the program must solve at each image point (x, y). In practice, the equation is solved on the nodes of a mesh (columns) using a fifth-order Runge-Kutta integration routine (which you, or the program, can look up when you need it). You need to choose the size and 'resolution' of the mesh. As we'll see in Chapter 27, there are situations where the column approximation will not be acceptable.

Without the column approximation, equation 25.35 gives us a system of coupled partial differential equations. The boundary conditions (at $z = 0$) can be generally written in the form

$$a_{\mathbf{g}}\phi_{\mathbf{g}} + b_{\mathbf{g}}\frac{\partial\phi_{\mathbf{g}}}{\partial x} = c_{\mathbf{g}} \qquad (25.37)$$

where we're ignoring changes in the y-dimension.

We can use fixed boundary conditions again following Howie and Basinski

- The foil is divided into thin slices of thickness Δz. You should not confuse this with the multislice method for lattice-image simulation which we'll see in Chapter 30; we are still using Howie-Whelan equations. Then, equation 25.35 is integrated, using the column approximation, through the first slice, i.e., from $z = 0$ to $z = \Delta z$, at all the mesh-points.
- The corrections to the column approximation, i.e., the terms containing derivatives with respect to x and y, are then evaluated by interpolation and included.
- The procedure is repeated until the exit surface of the foil is reached.

With this procedure, you are actually applying the column approximation to the outer boundary of the mesh. So, in equation 25.37 $a_{\mathbf{g}} = 1$, $b_{\mathbf{g}} = 0$, and $c_{\mathbf{g}} = \phi_{\mathbf{g}}$ at the initial surface. In order to avoid distortion of the image, we must choose the step size, Δz, carefully and be sure that the distance between columns (mesh size) is small enough (see Anstis and Cockayne).

25.13.D The User Interface

You'll want to run your program interactively so it should include commands which allow you to change parameters easily. Ideally, it will allow you to access each command through the keyboard using a menu. In Comis certain standard menus were available for special purposes. The user could also build (and save) menus interactively. This allowed all the relevant parameters and commands for a particular problem to be present within a single menu. At any time, all the commands were available through the keyboard.

Although typical simulations may be performed in a matter of seconds, many-beam calculations including several dislocations may require more CPU time. For this situation, Comis included a 'submit' command that would start a batch job based on your current data and parameters. Thus, the interactive mode could be used as a convenient way of submitting several jobs with varying parameter values.

For many problems, a purely visual comparison of experimental and simulated images is sufficient to allow you to interpret your image. In these situations, you can often find a $\xi^a_{\mathbf{g}}$ such that the simulations can be carried out with only two beams (à la Head et al.). However, since many parameters are involved in the image-matching process, it is best to eliminate as many unknown variables as possible. Many-beam calculations are even more important for quantitative analyses.

CHAPTER SUMMARY

The key points discussed in this chapter are

- We see contrast from planar defects because the translation, **R**, causes a phase shift $\alpha = 2\pi\mathbf{g}.\mathbf{R}$.
- In the two-beam case, we can derive analytical expressions to describe the contrast.
- We can use the scattering matrix method in the two-beam case and can readily extend it to more complicated multibeam situations.

Many different types of planar defect can be studied. You should be careful not to assume that all defects behave the same as SFs in fcc materials.

There is a direct relationship between the information in the images and that in the DPs, which you can understand using the concept of the relrod.

You need to understand how Bloch waves behave to explain why BF/DF pairs of images are not complementary and why the contrast from planar defects can disappear in the 'middle' of the image. The latter is a result of preferential absorption (really scattering) of certain Bloch waves.

We can now use computer modeling of diffraction-contrast images of planar defects to perform quantitative analysis and image matching.

REFERENCES

We suggest a few books or chapters of books for background reading. You'll realize that careers have been built on this topic. As usual, we also recommend that you read some of the original papers.

INTERFACES

Christian, JW 1975 *The Theory of Transformations in Metals and Alloys*, Part 1, 2nd edition, Pergamon Press New York.

Carter, CB and Norton, MG 2007 *Ceramic Materials* Springer New York. Chapters illustrating interfaces in ceramics.

Forwood, CT and Clarebrough, LM 1991 *Electron Microscopy of Interfaces in Metals and Alloys*, Adam Hilger New York. Invaluable for anyone studying interfaces by TEM.

Howe, JM 1997 *Interfaces in Materials: Atomic Structure, Thermodynamics and Kinetics of Solid-Vapor, Solid-Liquid and Solid-Solid Interfaces* Wiley New York.

Matthews, John Wauchope 1975, *Epitaxial Growth* Academic Press/Elsevier New York. Unfortunate use of 'epitaxial' instead of epitactic by a pioneer in the subject. (Yes, the father of Dave Matthews.)

Sutton, AP and Balluffi, RW 1995 *Interfaces in Crystalline Materials*, Oxford University Press New York.

Wolf, D and Yip, S, Eds. 1992 *Materials Interfaces, Atomic-level Structure and Properties*. Chapman and Hall New York.

THINKING ABOUT CONTRAST

The book by Head et al. (see Section 1.5) is the starting text for simulating diffraction-contrast images.

Amelinckx, S and Van Landuyt, J 1978 in *Diffraction and Imaging Techniques in Materials Science*, **1** and **2**, (Eds., S Amelinckx, R Gevers and J Van Landuyt), 2nd Ed., p. 107, North-Holland New York.

Anstis GR and Cockayne DJH 1979 *The Calculation and Interpretation of High-Resolution Electron Microscope Images of Lattice Defects* Acta Cryst. A**35**, 511–524.

Edington, JW 1976 *Practical Electron Microscopy in Materials Science*, Van Nostrand Reinhold New York. A classic practical guide but sometimes not suitable for modern TEMs.

Gevers R, Art, A and Amelinckx S 1963 *Electron Microscopic Images of Single and Intersecting Stacking Faults in Thick Foils – 1. Single faults* Phys. Stat. Sol. **3**, 1563–93.

Gevers R, Blank H and Amelinckx S 1966 *Extension of the Howie-Whelan Equations for Electron Diffraction to Non-Centro Symmetrical Crystals* Phys. Stat. Sol. **13**, 449–465. See this paper for discussion of imaginary extinction distances, $\xi'_{\mathbf{g}}$.

THE COLUMN APPROXIMATION

Howie A and Basinski ZS 1968 *Approximations of the Dynamical Theory of Diffraction Contrast* Phil. Mag. **17**, 1039–63. The non-column approximation paper.

Howie A and Sworn H 1970 *Column Approximation Effects in High Resolution Electron Microscopy Using Weak Diffracted Beams* Phil. Mag. **22**, 861–4. Example of a situation where the column approximation gives the wrong answer.

EXTINCTION DISTANCES

Doyle PA and Turner PS 1968 *Relativistic Hartree-Fock X-ray and Electron Scattering Factors* Acta Cryst. A**24**, 390–7.

Hashimoto, H, Howie, A and Whelan, MJ 1962 *Anomalous Electron Absorption Effects in Metal Foils: Theory and Comparison with Experiment* Proc. Roy. Soc. London A **269**, 80–103. Bloch waves and planar defects.

Hirsch PB, Howie A, Nicholson RB, Pashley DW and Whelan MJ 1977 *Electron Microscopy of Thin Crystals*, 2nd edition, p. 225, Krieger, Huntington, New York. Deduces equation 25.27.

Humphreys CJ and Hirsch PB 1968 *Absorption Parameters in Electron Diffraction Theory* Phil. Mag. **18**, 115–122.

Mott NF and Massey HSW 1965 *The Theory of Atomic Collisions*, 3rd Ed., Clarendon Press Oxford. Source of the Mott expression for X-ray scattering factors.

Taftø, J and Spence, JCH 1982 *A Simple Method for the Determination of Structure-Factor Phase Relationships and Crystal Polarity Using Electron Diffraction* J. Appl. Cryst. **15**, 60-4. Shows how to distinguish $\mathbf{g}$ and $\bar{\mathbf{g}}$ using CBED.

Yoshioka, H 1957 *Effect of Inelastic Waves on Electron Diffraction* J. Phys. Soc. Japan **12**, 618–628. See this paper for discussion of imaginary Fourier components, $U'_{\mathbf{g}}$.

SIMULATION

The original citation is the text by Head et al. (Section 1.5). It is still useful reading but needs a trip to the library.

Rasmussen, DR and Carter, CB 1991 *A Computer Program for Many-Beam Image Simulation of Amplitude-Contrast Images* J. Electron Microsc. Techniques **18,** 429–36. An idea of what we need in a user-friendly (Comis) program.
Rasmussen, DR, McKernan, S and Carter, CB 1991 *Rigid-Body Translation and Bonding Across {110} Antiphase Boundaries in GaAs* Phys. Rev. Lett. **66**, 2629–32.
Schaublin, R and Stadelmann, P 1993 *Method for Simulating Electron Microscope Dislocation Images* Mater. Sci. Engng. A**164**, 373–8. The paper describing CuFour.
Thölen AR 1970 *A Rapid Method for Obtaining Electron Microscope Contrast Maps of Various Lattice Defects* Phil. Mag. **22**, 175–182.
Thölen AR 1970 *On the Ambiguity Between Moiré Fringes and the Electron Diffraction Contrast from Closely Spaced Dislocations* Phys. Stat. Sol. (A) **2**, 537–550.
Viguier, B, Hemker, KJ and Vanderschaeve, G 1994 *Factors Affecting Stacking Fault Contrast in Transmission Electron Microscopy: Comparisons with Image Simulations* Phil. Mag. A**69**, 19–32.
Zhou, Z 2005 *Electron Microscopy and Elastic Diffuse Scattering of Nanostructures* D.Phil. Thesis, Oxford University. Zhongfu developed TEMACI.

THE COMPANION TEXT

Chapters on CuFour and more on TEMACI.

SELF-ASSESSMENT QUESTIONS

Q25.1 Describe the distinguishing features of the following interfaces: (a) phase boundary; (b) anti-phase boundary; (c) stacking fault.
Q25.2 What is α and for what values of α will planar defects not be seen?
Q25.3 Deduce the value of α for an fcc stacking fault.
Q25.4 A series of overlapping stacking faults show no contrast in the BF image. Explain how this could occur.
Q25.5 Which SF fringe is the same in both BF and DF images?
Q25.6 What interface gives complementary BF/DF fringes?
Q25.7 You can calculate the extinction distance. Can you confirm this value experimentally from an image?
Q25.8 What is the difference between a low-angle GB and a high-angle GB?
Q25.9 What does the concept of the scattering matrix allow to describe?
Q25.10 Why are BF/DF SF images usually not complementary?
Q25.11 Is the column approximation relevant to the study of planar defects?
Q25.12 Why might we see no contrast from the SF in the middle of the specimen even when we see fringes from the part close to each surface?
Q25.13 How does the planar defect affect the Bloch wave when it is close to the lower surface?
Q25.14 We describe SF contrast in terms of the inclination of the planar defect in the foil. When does this description clearly fail?
Q25.15 What does the term 'deformable-ion approximation' mean?
Q25.16 In deducing the propagation of Bloch waves in the presence of a SF, we use the tie line concept. What is the physical reason for using this concept in this situation?
Q25.17 Why do we recommend that you use the displaced-aperture DF approach for comparing BF and DF images?
Q25.18 Why is the displaced-aperture DF approach more useful today than in 1960?

TEXT-SPECIFIC QUESTIONS

T25.1 Consider Figure 25.5. At the far left of the lower image, the fringes are shifted. Explain this shifting using words and equations.
T25.2 Consider Figure 25.7 and if possible, the original paper. Confirm that the translation vector given in the text is correct. (Draw a larger diagram and deduce the direction and magnitude of **R**.)
T25.3 Consider Figure 25.8. Is this DF or BF and what is **g**? Explain your answer.
T25.4 In Figure 25.9, the contrast changes abruptly between regions that we might think of as 'grains.' Is this change real and why is it occurring?
T25.5 By considering Figure 25.10A and B, explain how the image will differ if the angle β is 30° or 150°.
T25.6 Is the shape of the sample in Figure 25.10c a perfect wedge? Discuss the reasons for your answer and the possible implications.
T25.7 Suggest a geometry for the anti-phase domain in Figure 25.8 using information from the fringes. Be as complete as possible.
T25.8 Explain why π fringes in Ni_3Al are so named.

T25.9 Extra challenge. By referring to the literature or otherwise, discuss the origin of the δ fringes in Figure 25.9. Are these fringes SF-type fringes? Is it coincidental that they end on arrays of dislocations? How would the fringes in CoO differ?

T25.10 Extra challenge. Explain the contrast in Figure 25.4 and relate them to one of the geometries. One difficulty you will encounter is identifying the first W fringe. Discuss why this is so.

T25.11 Sketch the possible SFs in Si for which the **g** vector is 131 and the beam direction is [2, 3, 11]. The planes on which the faults lie have been determined and are (111), (111), (111), and (111) for faults A, B, C, and D, respectively. How would you determine the nature of the SFs. Comment on why any intensity is observed for the faults on the $(\bar{1}\bar{1}\bar{1})$ and $(\bar{1}\bar{1}\bar{1})$ planes. (Adapted from Ian Robertson.)

T25.12 Consider a DP from two thin crystals and the corresponding moiré pattern. Both crystals are fcc structures. If you know that the magnification of the moiré pattern is 10^6 times and that the moiré pattern and the DP are correctly oriented with respect to each other, show how you would calculate the corresponding lattice parameters (a_1 and a_2) of the two fcc crystals. (Courtesy Matt Halvarson.)

26

Imaging Strain Fields

CHAPTER PREVIEW

As we discussed in Chapter 24, bending of the lattice planes causes a change in the diffraction conditions and therefore a change in the contrast of the image. The presence of a lattice defect in the specimen causes the planes to bend close to the defect. The special feature here is that the bending varies not just laterally, but also through the specimen. Since the details of the bending generally depend on the characteristics of the defect, we can learn about the defect by studying the contrast in the TEM image. This simple principle has led to one of the main applications of TEM, namely, the study of defects in crystalline materials. We can claim that our understanding of the whole field of dislocations and interfaces, for example, has advanced because of TEM. We have even discovered new defects using TEM—like the stacking-fault tetrahedron, the faulted dipole, and the multipole.

Usually we want to learn two things about these defects: we want to know where they are and then understand what they are. So the idea underlying this chapter is the same as for bend contours: we use different reflections corresponding to different sets of lattice planes. We see how the defects affect the image contrast from those different lattice planes and thus characterize the defects. In case you are worried, we would like to emphasize that this is *not* a chapter about defects; it is concerned with understanding contrast in the TEM. We will introduce the necessary terminology and notation concerning defects, but we won't try to give you a comprehensive discussion of them. You should consult the standard references on dislocations at the end of the chapter if you need more details. However, we will show lots of pictures because now we are concerned with the appearance of images. We finish with some discussion of image simulation.

26.1 WHY IMAGE STRAIN FIELDS?

First, we should review our terminology. When we displace the atom at position **r** a distance **R**(**r**) from its site in the perfect crystal, we say the crystal is under a strain (ε). If the crystal is strained, then it must be subject to a stress which we'll call σ. (Metallurgists traditionally use these symbols and although σ means 'cross section' to a microscopist, we'll stick with it.) Since **R**(**r**) varies with position in the crystal, ε and σ will in general also vary with **r**. We will assume that ε and σ can each be defined at a point. Then we will refer to these quantities as the displacement field, **R**(**r**), the strain field, ε(**r**), and the stress field, σ(**r**). You will notice that these terms are used interchangeably in the literature although there is a clear causal sequence. What we image is the effect of the **R**(**r**).

To have an intuitive feel for why we see contrast from dislocations, consider the geometry shown in Figure 26.1. The diffraction geometry has been set up so that the specimen is slightly tilted away from the Bragg condition. The distortion due to the dislocation will then bend the near-diffracting planes back into the Bragg-diffracting condition. We have relrods so there will still be some intensity in the electron beam even when we are not at the exact Bragg condition. The figure shows planes bending at a dislocation; compare this to Figure 24.7 showing bend contours. Regions far from the dislocation are tilted well away from the Bragg condition, while the regions or either side of the dislocation core are at the Bragg condition for $\pm\mathbf{g}_{hkl}$. It is more difficult to recognize the diffracting planes for a screw dislocation (Figure 26.2) but the planes are bending just the same.

When studying a particular dislocation (edge or screw), we want to determine the following parameters

- The direction and magnitude of the Burgers vector, **b**, which is normal to the *hkl* diffracting planes (Figures 26.1 and 26.2B).

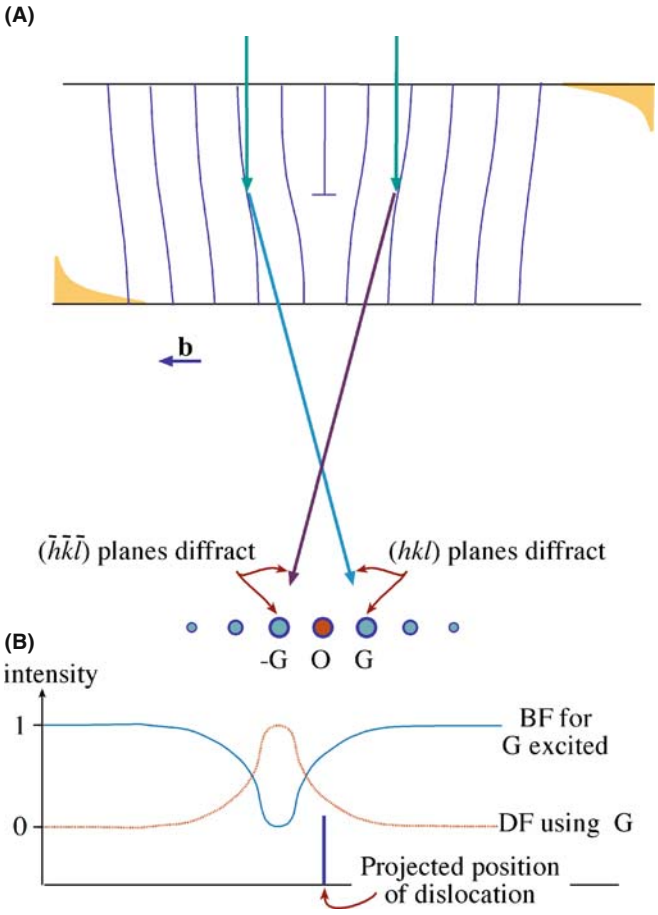


FIGURE 26.1. (A) The specimen is tilted slightly away from the Bragg condition ($\mathbf{s} \neq 0$). The distorted planes close to the edge dislocation are bent back into the Bragg-diffracting condition ($\mathbf{s} = 0$), diffracting into G and –G as shown. (B) Schematic profiles across the dislocation image showing that the defect contrast is displaced from the projected position of the defect. (As usual for an edge dislocation, **u** points into the paper.)

- The line direction, **u** (a vector), and therefore, the character of the dislocations (edge, screw, or mixed).
- The glide plane: the plane that contains both **b** and **u**.

There are other questions we want to answer

- Is the dislocation interacting with other dislocations, or with other lattice defects?
- Is the dislocation jogged, kinked, or straight?
- What is the density of dislocations in that region of the specimen (and what was it before we prepared the specimen)?
- Has the dislocation adopted some special configuration, such as a helix?

In many of these questions, you may find that stereomicroscopy or tomography (Section 29.1) can be very helpful although we will not emphasize those techniques here. The basic requirement if you do use stereomicroscopy is that you must form all of your images using the same **g** vector.

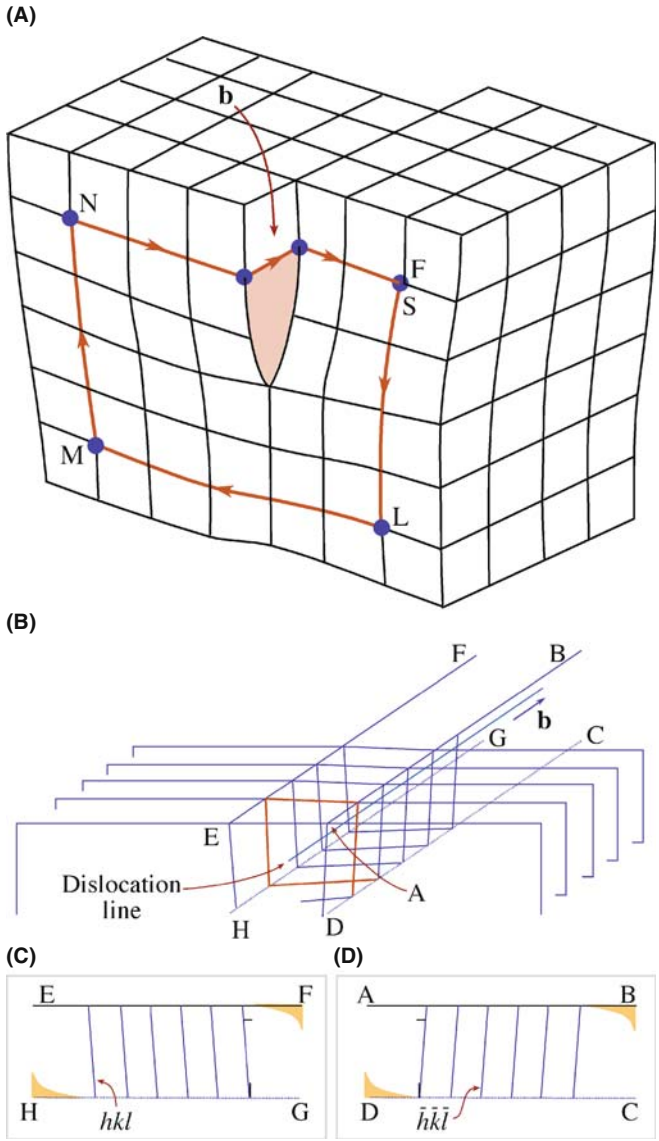


FIGURE 26.2. (A) Distortion of planes around a screw dislocation. The circuit SLMNF is used to define the Burgers vector, **b** (see Figure 26.5). (B) Schematic showing the rotation of the diffracting planes by a screw dislocation. The planes are rotated in opposite directions on either side of the dislocation. (C) and (D) show sections (ABCD and EFGH) through these diffraction planes.

26.2 HOWIE-WHELAN EQUATIONS

Let's start with the two-beam phenomenological approach because it worked so well in Chapter 25. An important assumption is that we have linear elasticity. What this means is that if we have $\mathbf{R}_1$ due to one defect and $\mathbf{R}_2$ due to a second defect, then at any point in the specimen we can just add these two values to determine the total displacement field, **R**. We will not consider anisotropic elasticity although this can readily be included in calculations if you're using a computer.

In Chapter 25, we showed that we could modify the Howie-Whelan equations to include a lattice distortion **R**. So for the imperfect crystal

$$\frac{d\phi_{\mathbf{g}}}{dz} = \frac{\pi i}{\xi_{\mathbf{0}}}\phi_{\mathbf{g}} + \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{\mathbf{0}} \exp[-2\pi i(sz + \mathbf{g}.\mathbf{R})] \qquad (26.1)$$

Now we make a different substitution of variables (compare with equations 25.5 and 25.6). Set

$$\phi_{\mathbf{0}}(z)_{(\text{sub})} = \phi_{\mathbf{0}}(z) \exp\left(\frac{-\pi i z}{\xi_{\mathbf{0}}}\right) \qquad (26.2)$$

and

$$\phi_{\mathbf{g}}(z)_{(\text{sub})} = \phi_{\mathbf{g}} \exp\left(2\pi i s z - \frac{\pi i z}{\xi_{\mathbf{0}}} + 2\pi i \mathbf{g}.\mathbf{R}\right) \qquad (26.3)$$

The justification for this substitution is the same as always. You'll notice that $\phi_{\mathbf{0}}(z)_{(\text{sub})}$ is the same as before, but $\phi_{\mathbf{g}_{(\text{sub})}}$ now includes a **g**.**R** term. The reason for this substitution is that it will give us a simple expression for $d\phi_{\mathbf{g}}/dz$.

The equations become

$$\frac{d\phi_{\mathbf{0}}(z)_{(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{\mathbf{g}}(z)_{(\text{sub})} \qquad (26.4)$$

and

$$\frac{d\phi_{\mathbf{g}(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{\mathbf{0}}(z)_{(\text{sub})} + \left[2\pi i\left(s + \mathbf{g}\cdot\frac{d\mathbf{R}}{dz}\right)\right]\phi_{\mathbf{g}}(z)_{(\text{sub})} \qquad (26.5)$$

which can be rewritten, while dropping the subscript

$$\frac{d\phi_{\mathbf{g}}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{\mathbf{0}} + 2\pi i s_{\mathbf{R}}\phi_{\mathbf{g}} \qquad (26.6)$$

This equation looks just like equation 13.14 but with $s_{\mathbf{R}}$ instead of s, where

$$s_{\mathbf{R}} = s + \mathbf{g}\cdot\frac{d\mathbf{R}}{dz} \qquad (26.7)$$

The concept of $s_{\mathbf{R}}$ is new.

The importance of this result is that although we have a new 's.' we have the same equation so we can use the rest of the analysis of Chapter 13 and obtain the same results with a modified value of s, i.e., $s_{\mathbf{R}}$. Therefore, we'll have the same thickness dependence so that the contrast of the defects will depend on both **s** and $\xi_{\mathbf{g}}$. The big change is that we can now treat the case where **R** is a continuous function of z.

We will examine how the $\mathbf{g}.d\mathbf{R}/dz$ and **g**.**R** terms are used to understand dislocations. Since the equations we have just derived have the same form as those we discussed in Chapters 13 and 25, we can expect many of the same properties in the images. In particular, the images of defects will show the same sort of thickness dependence. We can also use the equations we derived in Chapter 25, so we have two ways of looking at the defects

- **g**.**R** contrast is used when **R** has a single value,
- $s_{\mathbf{R}}$ contrast is used when **R** is a continuously varying function of z, which in turn is associated with $\mathbf{g}.d\mathbf{R}/dz$.

Now let's consider the principles of this analysis. Remember, we are not trying to be quantitative or totally rigorous. We will generalize the two-beam treatment for the imperfect crystal. Note that we still have beams, it's a dynamical situation, and we assume that the column approximation is valid. So how does the column approximation relate to the theory? The model relates **R** to the column as shown in Figure 26.3 and the calculation is for a continuum even though we have atoms. The important point is that the displacement field, **R**, varies with position, **r**; we can define the origin as the core of the defect. We'll go through the calculation for a dislocation parallel to the foil surface.

As we saw in Section 13.11, the column approximation is equivalent to assuming that the crystal can be divided into narrow columns. We then calculate the amplitudes of the beams in any such column as if the whole crystal consisted of an infinite number of identical columns. The approximation is valid when we don't need to see image detail below 2–3 nm. The actual diameter of the column depends on the diffracting conditions. We can include the effect of distortions due to strains from lattice defects by imagining that the column consists of slabs of perfect crystal, each displaced by an amount $\mathbf{R}(z)$ (like a mini-SF that's different on every plane). Remember that z is actually measured along the column.

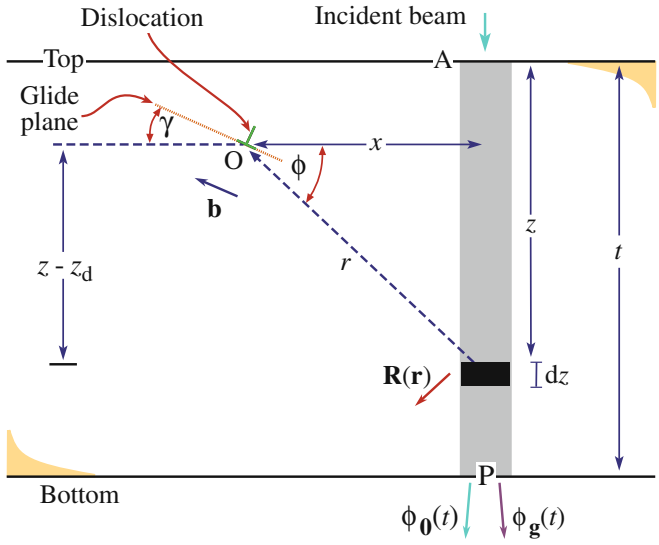


FIGURE 26.3. The effect of a dislocation with Burgers vector, **b**, at O on a column, distance x away. The effect of the strain field on the electron waves in the column is integrated in increments dz over its total length t, giving amplitude $\phi_{\mathbf{0}}(t)$ and $\phi_{\mathbf{g}}(t)$ at P.

26.3 CONTRAST FROM A SINGLE DISLOCATION

When we study dislocations, we usually want to know how many there are (the density) and whether they are edge, screw, or mixed in character. The displacement field in an isotropic solid for the general, or mixed, case can be written as

$$\mathbf{R}=\frac{1}{2\pi}\left(\mathbf{b}\phi+\frac{1}{4(1-\nu)}\{\mathbf{b}_e+\mathbf{b}\times\mathbf{u}(2(1-2\nu)\ln r+\cos 2\phi)\}\right) \tag{26.8}$$

For convenience, **R** is given here in polar coordinates (r and ϕ) shown in Figure 26.3; **b** is the Burgers vector, $\mathbf{b}_e$ is the edge component of the Burgers vector, **u** is a unit vector along the dislocation line (the line direction), and ν is Poisson's ratio.

It was particularly important to be able to write down this expression when we did the calculations by hand. However, when we have a computer available, it's quite straightforward to use anisotropic elasticity or just feed in displacements calculated from a computer model of the atom structure.

The amplitude of the diffracted beam, $\phi_{\mathbf{g}}$, is directly influenced by the value of **R**. We can consider two particular cases, namely, the screw and edge dislocations. For the screw dislocation, $\mathbf{b}_e = 0$ and **b** is parallel to **u** so that $\mathbf{b}\times\mathbf{u} = 0$. Then the expression for **R** in equation 26.8 simplifies to

$$\mathbf{R} = \mathbf{b}\frac{\phi}{2\pi} = \frac{\mathbf{b}}{2\pi}\tan^{-1}\left(\frac{z-z_d}{x}\right) \tag{26.9}$$

Here, z is the distance traveled down the column and z_d is the distance of the dislocation core below the top surface (again, refer to Figure 26.3). The dependence on $(z - z_d)$ emphasizes that the displacement field is present above and below the dislocation; it affects the whole column. From these two equations we see that **g**.**R** is proportional to **g**.**b**. For this reason, we often discuss images of dislocations in terms of **g**.**b** (g-dot-b) contrast. Examples of **g**.**b** values for some dislocations lying on a (111) plane in an fcc material with a [011] beam direction are given in Table 26.1.

THE SPECIMEN

If the sample is a single crystal, then you need to prepare a specimen with a [111] foil normal, so that you can image long segments of the dislocations lying in the plane of the foil on their (111) glide plane.

The second special case arises when the dislocation is pure edge in character. Then $\mathbf{b} = \mathbf{b}_e$ and **g**.**R** involves two terms **g**.**b** and **g**.**b** × **u**. (The latter term is read as 'g-dot-b-cross-u.') The displacement field causes the Bragg-diffracting planes associated with **g** to bend. Incidentally, the origin of **g**.**b** × **u** is interesting; it arises because the glide plane is buckled by the presence of an edge dislocation as illustrated in Figure 26.4. This buckling can be important because it complicates the analysis of **b** for some dislocations with an edge component, as we'll see below.

TABLE 26.1. Different Burgers Vectors and Different Reflections Give Different g.b = *n* Values

g \ b	$\frac{1}{6}[11\bar{2}]$	$\frac{1}{6}[1\bar{2}1]$	$\frac{1}{6}[\bar{2}11]$	$\frac{1}{3}[111]$
$\pm(1\bar{1}1)$	$\pm\frac{1}{3}$	$\pm\frac{2}{3}$	$\pm\frac{1}{3}$	$\pm\frac{1}{3}$
$(\bar{1}\bar{1}1)$	$\pm\frac{2}{3}$	$\pm\frac{1}{3}$	$\pm\frac{1}{3}$	$\pm\frac{1}{3}$
$(0\bar{2}2)$	± 1	± 1	0	0
(200)	$\pm\frac{1}{3}$	$\pm\frac{1}{3}$	$\pm\frac{2}{3}$	$\pm\frac{2}{3}$
$(3\bar{1}1)$	0	± 1	± 1	± 1
$(\bar{3}\bar{1}1)$	± 1	0	± 1	± 1

The dislocations all lie on a (111) plane in an fcc material; the beam direction is [011].

- Always remember: **g**.**R** causes the contrast and for a dislocation, **R** changes with z.
- We say that **g**.**b** = n. If we know **g** and we determine n, then we know **b**.

An experimental point: you usually set s to be greater than 0 for **g** when imaging a dislocation in two-beam conditions. Then the dislocation can appear dark against a bright background in a BF image. Of course, you still need to think about $s_{\mathbf{R}}$ and $d\mathbf{R}/dz$ since these will vary with z, as we saw in Figure 26.1.

The + and – signs in Table 26.1 are very important. If the sign of **R**, and hence **g**.**R** or **g**.**b**, reverses, then the image of the dislocation will move to the other side of the projected

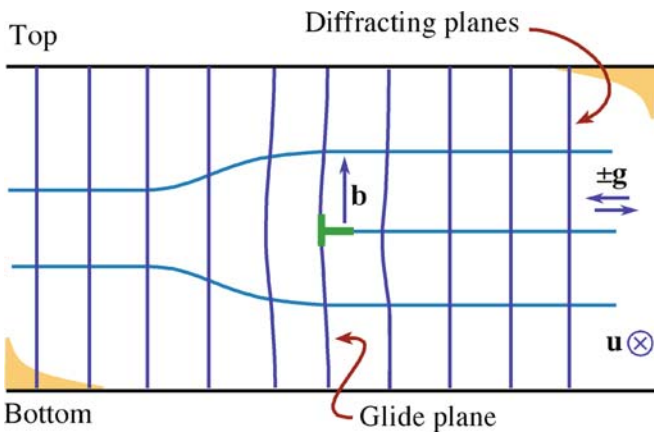


FIGURE 26.4. Buckling of the glide planes arises because of the term **g**.**b** × **u** and is important because it complicates the analysis of **b**.

position of the dislocation core. If you look carefully at Figure 26.1, you can appreciate that reversing the sign of s produces the same effect as reversing the sign of **g**. We can summarize these two ideas in terms of the quantity $(\mathbf{g}.\mathbf{b})s$ ('g-dot-b-times-s'), as shown in Figure 26.5.

If we identify two reflections $\mathbf{g}_1$ and $\mathbf{g}_2$ for which $\mathbf{g}.\mathbf{b} = 0$, then $\mathbf{g}_1 \times \mathbf{g}_2$ is parallel to **b**. This identification of **b** is actually a little more complicated because dislocations appear out of contrast when $\mathbf{g}.\mathbf{b} < 1/3$; similarly, dislocations need not be invisible even if $\mathbf{g}.\mathbf{b} = 0$ when $\mathbf{g}.\mathbf{b} \times \mathbf{u} \neq 0$.

INVISIBILITY CRITERION
If $\mathbf{g}.\mathbf{b} = 0$, then you won't see any contrast because the diffracting planes are then parallel to **R**.

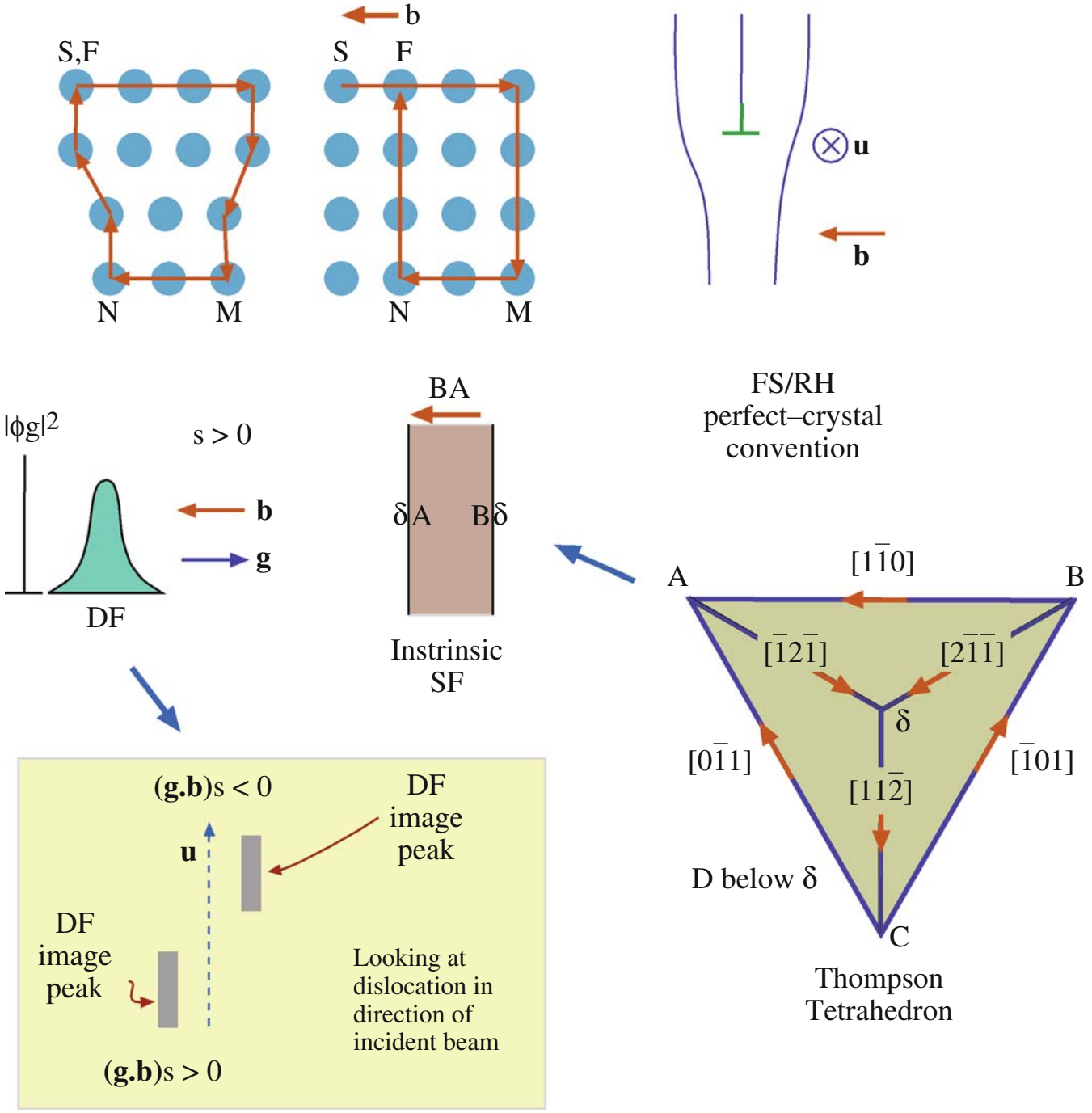
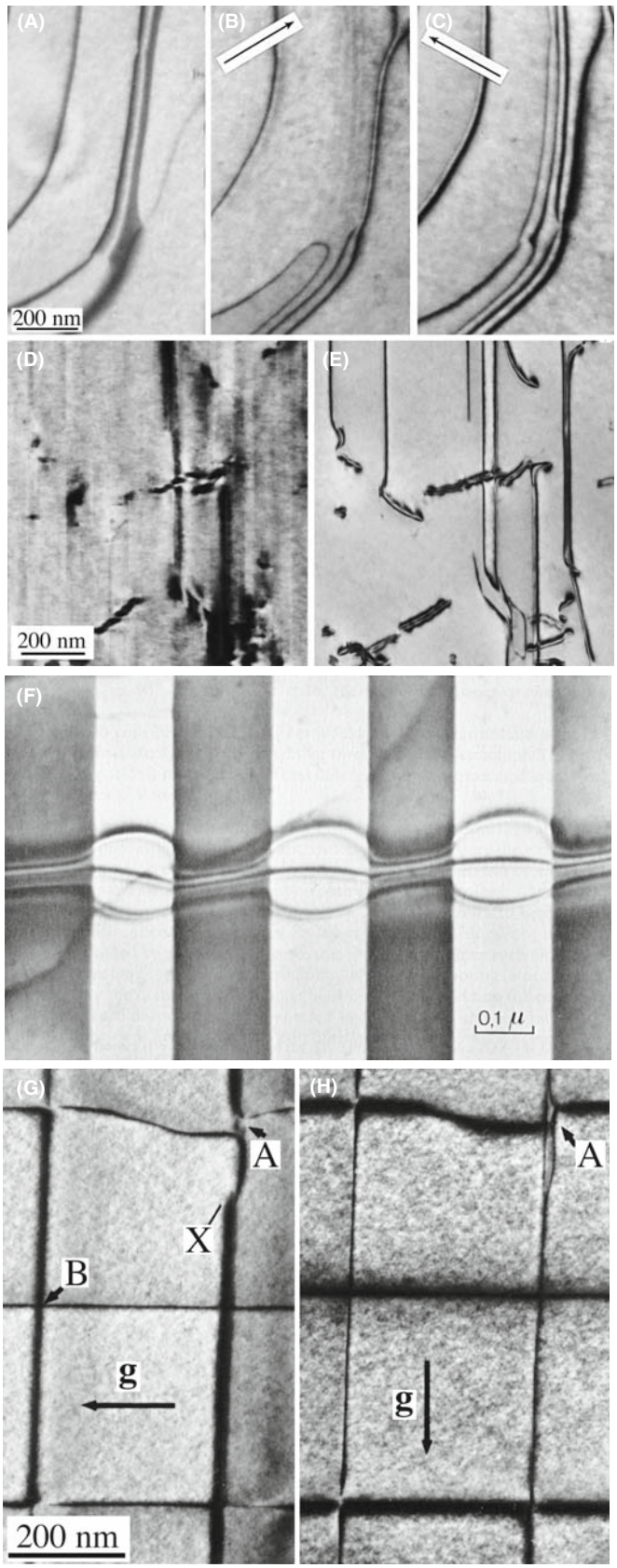


FIGURE 26.5. A brief summary of dislocations in an fcc crystal: **b** is defined by the finish- (F) to-start (S) vector in a right-hand (RH) circuit that comes to closure around the dislocation but fails to close in the perfect crystal. The location of the diffracted intensity $|\phi_{\mathbf{g}}|^2$ relative to the core depends on the sign of **b**, **g**, and s for the FSRH convention. If any sign is reversed, the contrast shifts across the core. When a perfect dislocation splits into Shockley partial dislocations, the order of the partial dislocation is given by the Thompson tetrahedron.



If we compare the contrast from a dislocation with that from a SF, the difference is that now α is a continuously varying function of z. The image of the dislocation itself shows thickness fringes but it may be 'out of contrast' at some depths or thicknesses, as you can see (look carefully) in the experimental image shown in Figure 26.6E.

Some points to remember from this discussion are

- The sign of s affects the image.
- The sign of x affects the image; the image is asymmetric.
- The magnitude of s affects the image.
- The depth of the dislocation and the thickness of the specimen affect the image.
- The appearance of the image depends on $\mathbf{g}.\mathbf{b}$ or, more completely, on $(\mathbf{g}.\mathbf{b})s$ and $\mathbf{g}.\mathbf{b} \times \mathbf{u}$.
- If we repeat this analysis for other values of $\mathbf{g}.\mathbf{b}$ $(= n)$ and plot intensities, we would find that the image width becomes broader as n increases.
- Note where the dislocation image 'comes from': the position of the line in the image only rarely corresponds to the projected position of the dislocation; it is usually displaced to one side of the core.
- As a complication, remember that the dislocations will probably be found in wedge specimens, not ideal parallel-sided ones.

A final 'rule of thumb' which you may find useful (from computer modeling and early analytical calculations) is that if $\mathbf{g}.\mathbf{b} = 0$, you can still 'see' dislocations when $\mathbf{g}.\mathbf{b} \times \mathbf{u} \geq 0.64$. For fcc materials, this rule can be useful when the foil is not parallel to a {111} plane.

Other examples of dislocation images are illustrated in Figure 26.6. Remember that partial dislocations are not only present in fcc metals; they also occur in many fcc semiconductors and many layer materials. Such materials may have a very low stacking-fault energy allowing the partial dislocations to separate, forming wide ribbon-like defects, as shown in Figure 26.6A–C. The single 'line' (that is actually a dark line and a less-dark line) below the arrow in (C) is a dislocation having its Burgers vector parallel to $\mathbf{g}$ (so $\mathbf{g}.\mathbf{b} = 2$); you can see two 'peaks' in the image, one darker and broader than

FIGURE 26.6. (A–C) Three strong-beam BF images from the same area using (A) $\{11\bar{1}\}$ and (B, C) {220} reflections to image dislocations which lie nearly parallel to the (111) foil surface in a Cu alloy which has a low stacking-fault energy. (D, E) Dislocations in Ni_3Al in a (001) foil imaged in two orthogonal {220} reflections. Most of the dislocations are out of contrast in (D). (F) A complex dislocation crossing a (rotational) domain boundary; the character of the dislocation changes and thus its dissociation width changes. (G, H) Dislocations of a (001) interface between two slightly lattice-mismatched III–V compounds.

the other. Notice that one of those dark lines in (C) has nearly disappeared in (B) and the dark peak is on the other side of the dislocation. A group of three parallel lines is present in (C) but is almost out of contrast in (B). These are Shockley partial dislocations all having the same **b** and thus giving **g**.**b** = 0 in (B) (the three lines actually form by the dissociation of a perfect dislocation with Burgers vector $\frac{1}{2}<112>$). The $11\bar{1}$ image (A) is formed by tilting the specimen to a 112 pole (~20° from the 111 pole) and shows contrast from the stacking faults themselves (imagine them filling the area between the dislocations); these faults will never give contrast at the 111 pole since **g**.**R** is then always 0 (or an integer). Intermetallics tend to have large unit cells so that the superlattice dislocations dissociate into partial dislocations, which would have been perfect dislocations in the disordered crystal (Figure 26.6D and E). These super-partial dislocations can dissociate further as they might have in the disordered lattice, or they can separate differently in different ordered domains (Figure 26.6F). Dislocations in interphase boundaries can be revealed by imaging with different reflections (Figure 26.6G and H). Since the dislocations are present to accommodate the mismatch, they must lie at, or close to, the (001) phase boundary. It can be difficult to analyze their Burgers vectors unambiguously, because the adjoining materials have different extinction distances, etc. The topic is 'image interpretation': now you know why. One of the extra challenges is determining the plane on which this dissociation occurs. We'll illustrate how we can see **g**.**b** × **u** contrast when we examine dislocation loops in Zn in Section 26.6.

A QUESTION

How many dislocations can you really see in Figure 26.6H? Answer: zero.

26.4 DISPLACEMENT FIELDS AND EWALD'S SPHERE

In Section 26.2, we showed that when a displacement **R**(**r**) is present, we can think of s as being replaced by $s_{\mathbf{R}}$ (equation 26.7). This new s should be written more completely as

$$s_{\mathbf{R}} = s + \mathbf{g}.\frac{\partial R}{\partial z} + \theta_B \mathbf{g}.\frac{\partial \mathbf{R}}{\partial x} \qquad (26.10)$$

The point is that, as you can see in Figure 26.7, **R** causes the lattice planes to bend through an angle δϕ. So two other parameters, namely **g** and s, also change. The diffraction vector is actually lengthened by Δ**g** but, more

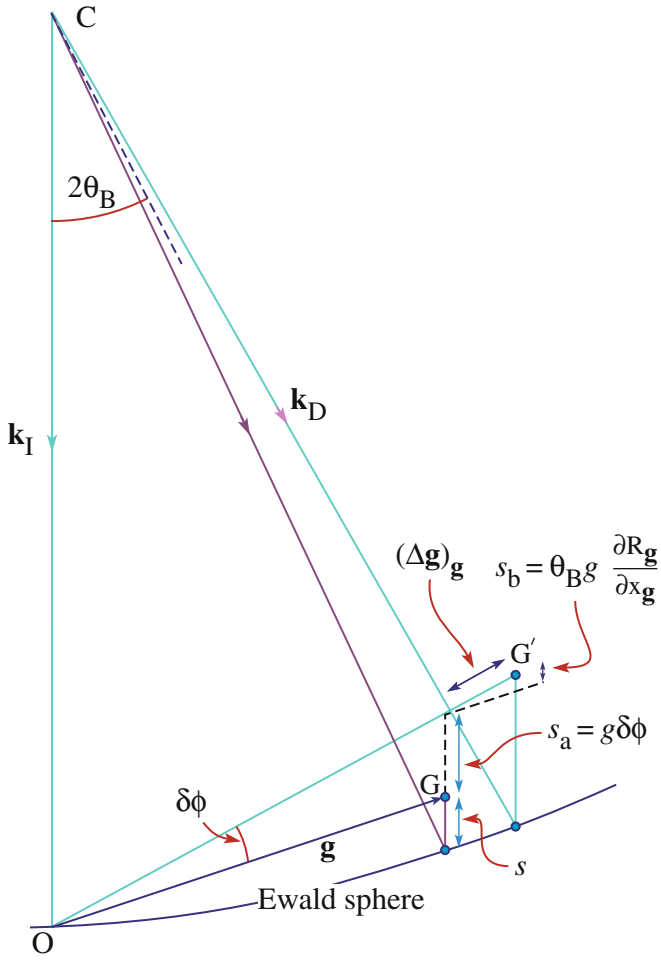


FIGURE 26.7. The strain field of the dislocation causes the lattice planes to bend through an angle δϕ. So **g** and s also change. The diffraction vector is lengthened by Δ**g** and **g** is rotated. So s increases by the two components of $s_{\mathbf{R}}$, i.e., s_a and s_b.

importantly, **g** is rotated. The result is that **s** increases by the two components s_a and s_b shown in the figure, to give $s_{\mathbf{R}}$. If you manipulate this equation for small angles you can produce equation 26.7. We usually neglect the third term because θ_B is small but it can become important when screw dislocations intersect the surface.

An alternative way of looking at this deformation is to think of **g** as changing by Δ**g**. We can define this change by the equation

$$\mathbf{g}.(\mathbf{r} - \mathbf{R}(\mathbf{r})) = (\mathbf{g} + \Delta\mathbf{g}).\mathbf{r} \qquad (26.11)$$

so that

$$-\mathbf{g}.\mathbf{R}(\mathbf{r}) = \Delta\mathbf{g}.\mathbf{r} \qquad (26.12)$$

The implication is that the information about the displacement field, **R**(**r**), is present in the region around **g** but not actually at **g**. Remember that the reflection **g** is present because we have a perfect crystal. It is difficult to image this type of scattering (but it becomes easier using

EFTEM). If you displace the objective aperture, you will still see the dislocation, but other inelastic scattering will complicate image interpretation. We saw that scattering does indeed occur between Bragg reflections in Section 17.6. An analogy for scattering from dislocations is the scattering of light from a single slit which we discussed in Chapter 2.

In the deformable-ion approximation (Section 25.13), we make the assumption that the atom doesn't know it has moved. If **R**(**r**) varies rapidly, as it does near the core of a dislocation, the approximation must fail. You can draw the same conclusion whenever the density of the material changes rapidly. So what we should do is use a better model for the atomic potential, one that also takes account of what happens to the valence electrons at such a defect. Of course, linear elasticity theory also fails when the strains, and hence **R**(**r**), are large, as at dislocation cores.

26.5 DISLOCATION NODES AND NETWORKS

You can analyze the Burgers vectors of dislocations which form networks directly and easily, if all the dislocations lie in a plane parallel to the surface of the specimen, as illustrated in Figure 26.8 for the case of graphite. The idea is simple: you form a series of images using different **g**-vectors. Don't forget that you can tilt to other poles; in fact, you'll often need to tilt the specimen just to image SFs which lie parallel to the foil surface, as in Figure 26.8A (but you won't see SF fringes then). When the diffraction vector is parallel to the plane of the SF you'll only see contrast from the dislocations as in Figure 26.8B–D. Such tilting experiments are essential if you're examining networks of misfit dislocations since the dislocations will then often have a component of their Burgers vector out of the plane of the network.

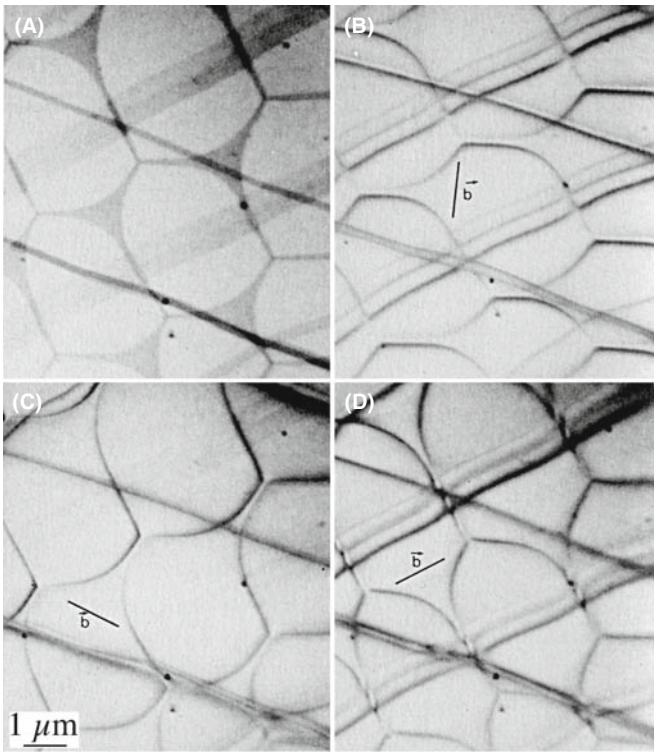


FIGURE 26.8. Dislocation networks in graphite. In (A) the stacking faults parallel to the foil surface are imaged by using a **g** that is inclined to the fault (so **g**.**R**≠ 0); in (B–D) two of the three dislocations at each node are in contrast but the third is invisible. Knowing **g** for each image, the Burgers vector of the dislocations can be determined as shown in (B–D) (**g**.**b** = 0 for the *missing* dislocations).

26.6 DISLOCATION LOOPS AND DIPOLES

Loops have been studied extensively because they can form when point defects coalesce. There are probably thousands of papers describing TEM studies of radiation damage and the formation of dislocation loops. In fact, many HVEMs were built in the 1960s just to study this problem. Questions which were answered led to a greatly improved understanding of irradiation processes (but failed to justify the construction of more nuclear power stations). We found that

- The loops can form by coalescence of interstitials or vacancies.
- The rate of growth, critical size, and nucleation time for different loops can be measured.
- Some of the loops are found to be faulted (containing a SF) while others are not faulted. The faulting should be related to the size of the loop and the stacking-fault energy of the material.

These studies were particularly instructive illustrations of the value of diffraction contrast.

- Dislocation loops can have either positive or negative **b,** and can be inclined to the beam on different planes as shown in Figure 26.9. Many person-years have been spent analyzing the inside-outside contrast from such defects to determine if they were caused by vacancies or interstitials.
- Loops can be present that show no **g**.**b** contrast.
- Loops can enclose single or multiple stacking faults, and so exhibit SF contrast as shown in Figure 26.10.
- The dislocation dipole is a special case and gives an important example of interacting dislocations. TEM is the best way to image dipoles because they have no long-range strain fields; the Burgers vector of the complete dipole is zero! (More on this in Chapter 27.)

Dislocations in Zn provide a particularly nice illustration of **g**.**b** × **u** contrast. If the specimen surface is parallel

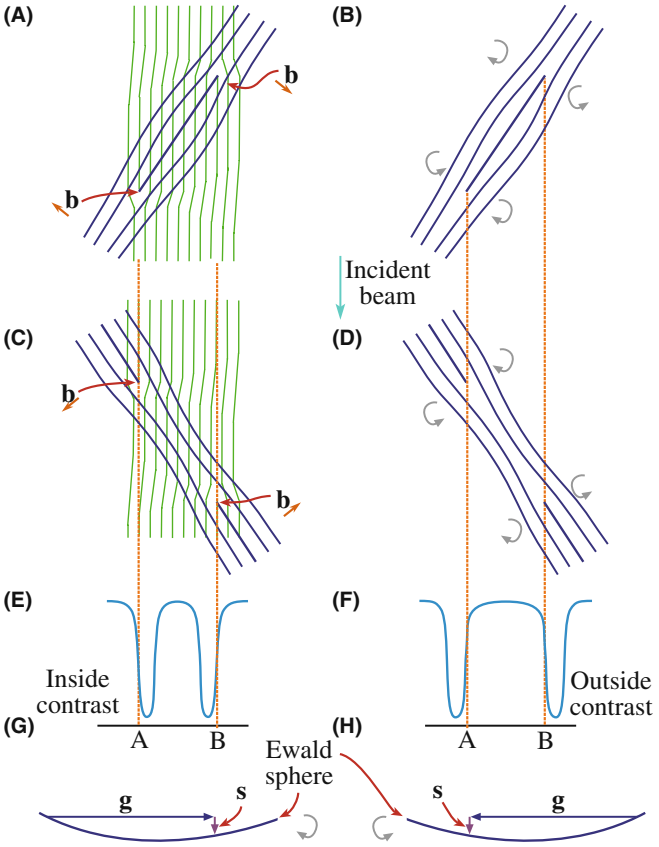


FIGURE 26.9. (A) Structure of an interstitial loop relative to the diffracting plane (faint lines). (B) Arrows show the rotation of the diffracting planes around the dislocation. (C, D) Vacancy loops. (E, F) Position of the image contrast relative to the projected dislocation position. Inside contrast occurs when clockwise rotation of the diffracting planes brings them into the Bragg condition. Outside contrast occurs for the counterclockwise case. (G, H) The relationship between **g**, **s**, and the sense of rotation. Everything is reversed if the loops are tilted in the opposite direction relative to the beam (i.e., reflect this figure in a mirror).

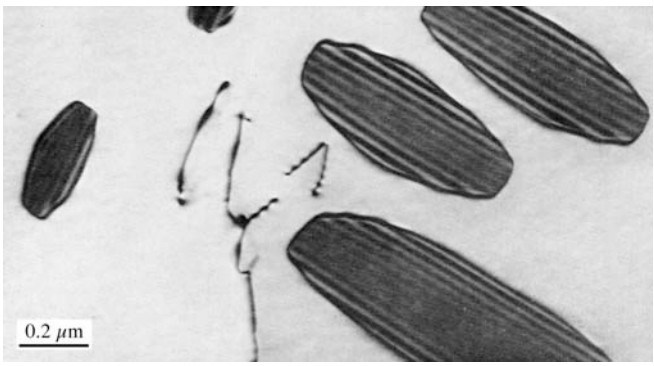


FIGURE 26.10. Dislocation loops in irradiated Ni showing SF contrast.

to the (0001) basal plane, then dislocation loops can readily form by coalescence of vacancies. In Figure 26.11, **b** is normal to **g** so that **g**.**b** = 0. These loops give a clear illustration of how the appearance of the image

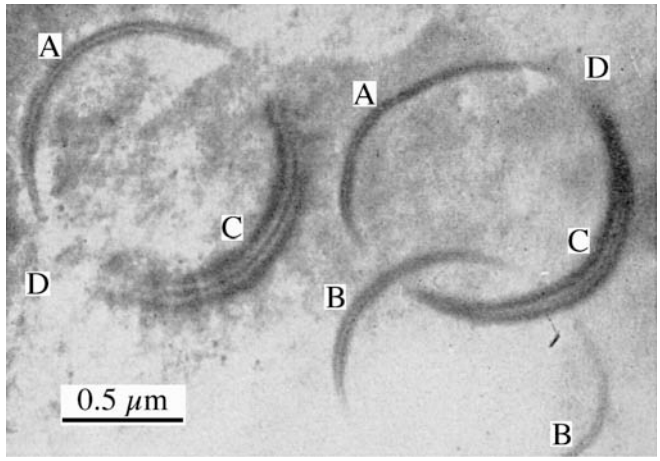


FIGURE 26.11. Prismatic loops in Zn parallel to the (0001) surface of the specimen with **b** = *c*[0001]. All round the loop, **b** is normal to **g** so that **g**.**b** = 0 and the vector **b** × **u** lies in the plane of the loop. At A, B, and C, **b** × **u** is parallel to **g** so that we see strong contrast. However, at D, **b** × **u** and **g** are mutually perpendicular so that **g**.**b** × **u** = 0 and the loop disappears.

depends on the line direction, **u**, of a dislocation. Note that you can see the dislocation contrast, even though **g**.**b** is zero, so this is not an absolute criterion for invisibility.

The above discussion is fine if the loops are large, but a problem arises when they are small. You must then consider the details of the contrast mechanism.

This was, of course, true for all these images, but now the size of the defect is small compared to the extinction distance. The schematic shown in Figure 26.12 summarizes the contrast that arises from small vacancy loops; if the loops were interstitial in nature, the contrast would be reversed. Not only does the black/white contrast change as the position of the defect changes in the specimen, but its size also appears to change. When the nature of the loops becomes more complex, the appearance of the image may also become more difficult to interpret with 'butterflies,' 'lozenges,' and 'peanuts' being seen. Notice that the behavior of the contrast differs in the BF and DF images; this effect is similar to that which we discussed in Chapter 25 and is again related to anomalous absorption.

IMAGES AND THICKNESS

The basic idea is that the appearance of the image is now dependent on the thickness of the specimen.

Dislocation dipoles can be present in great numbers in heavily deformed metals, but can also be important in the degradation of some semiconductor devices. Dipoles can be thought of as loops which are so elongated that they look like a pair of single dislocations of opposite Burgers vector, lying on parallel glide planes. As a result, they are best recognized by their

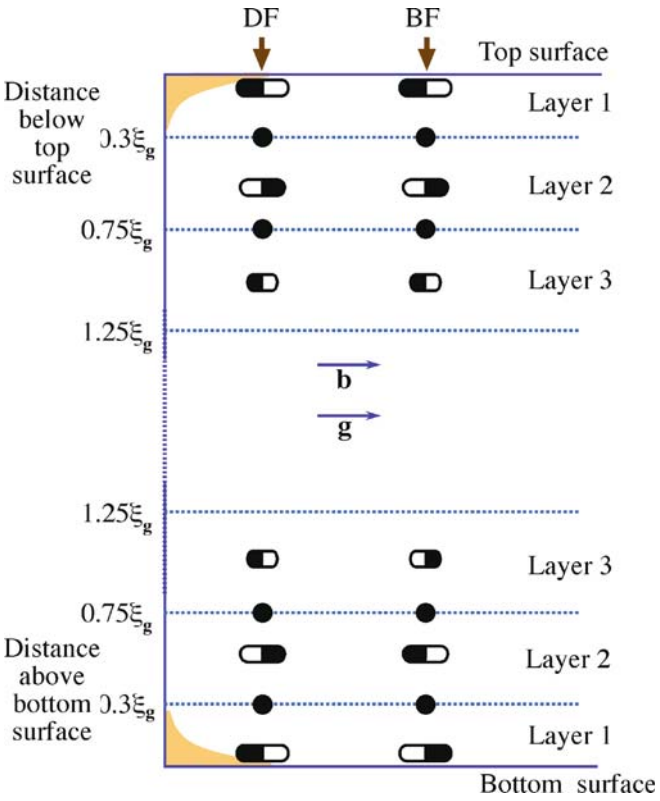


FIGURE 26.12. Changes in the black-white contrast from small dislocation loops at different depths in the specimen. The DF shows the same contrast at top and bottom while the BF contrast is complementary at the two surfaces.

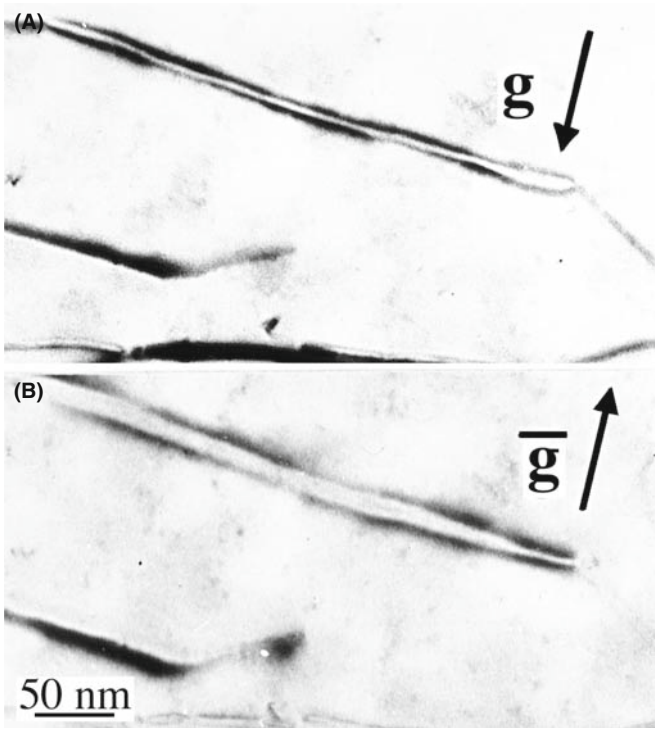


FIGURE 26.13. Images of dislocation dipoles in Cu showing inside-outside contrast on reversing **g** (±220). (A) is inside; (B) is outside.

'inside-outside' contrast, as illustrated in Figure 26.13. You can appreciate the origin of the term by looking at the projection of the images of the two dislocations when you reverse the sign of **g**: since the two dislocations have opposite Burgers vectors, Figure 26.9 tells you that one image will lie on one side of the core and the other on the opposite side. The order reverses when we reverse **g**.

26.7 DISLOCATION PAIRS, ARRAYS, AND TANGLES

Remember, you are not limited to **g**-vectors which are parallel to the foil surface; hence you can tilt the specimen to see SF contrast. As we saw in Figure 26.8A, this is often helpful if you have SFs associated with the dislocations; you can then produce **g**.**R** contrast for the fault. We will discuss dislocation dissociation more in Chapter 27. If you look back at Figure 26.6 or 26.8, you will see the benefit of being able to see the SF. This figure also illustrates the effect of n on the dislocation contrast.

Consider a dislocation in an fcc metal which can dissociate into two Shockley partial dislocations on the (111) plane. We can write down the dislocation reaction as

$$\tfrac{1}{2}[1\bar{1}0] = \tfrac{1}{6}[1\bar{2}1] + \tfrac{1}{6}[21\bar{1}] \text{on } (111) \qquad (26.13)$$

If we image this dislocation using the $(2\bar{2}0)$ reflection, then $\mathbf{g}.\mathbf{b} = 2$. If, instead, we use the $(20\bar{2})$ reflection, then $\mathbf{g}.\mathbf{b} = 1$. The appearance of the image is very different even if we cannot see the individual partial dislocations.

The advantage of using high voltages to study arrays of dislocations is illustrated in Figure 26.14; everything we said in Chapter 11 applies when we study dislocations. We see thickness fringes at the surface, but these disappear in the central region of the foil—the contrast becomes constant. When the foils are this thick, you may find stereomicroscopy helpful in giving a 3D view of the defect arrangement; you can imagine its value in interpreting an image such as that shown in Figure 26.15A. The defects may be very close together in heavily deformed materials, as shown in Figure 26.15B; of course, the interpretation of such an image depends on the thickness of the specimen so you also need to know the voltage used to form the image. Ideally, make the specimen thin over very large areas to minimize image overlap. If the density of defects is too large, the weak-beam technique may be the only way to 'look into' the walls (Chapter 27) but then the specimen needs to be even thinner.

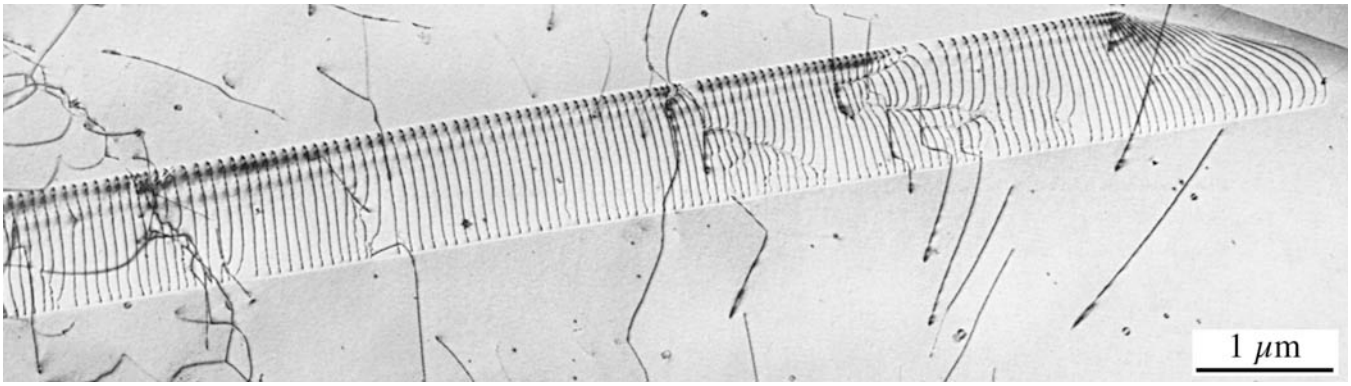


FIGURE 26.14. Dislocations threading through a very thick specimen in an image recorded using a (very) high-voltage (3 MeV) TEM.

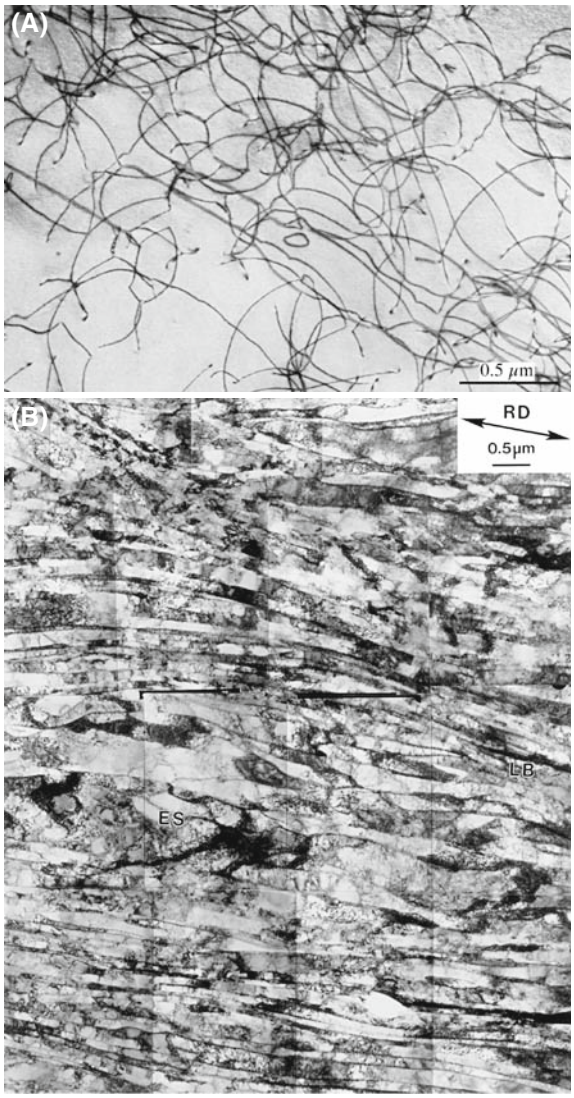


FIGURE 26.15. (A) Dislocation tangles in an Fe-35% Ni-20% Cr alloy, creep tested at 700°C; the dislocations have moved by glide and climb and do not lie on well-defined planes. (B) Dislocation walls in Al which has been heavily deformed by directional rolling. (200-keV electrons but super specimen.)

A CALCULATION

$\sin\theta_D/2 = \mathbf{b}/2t$.

If $\mathbf{b} \sim 0.25$ nm and $t = 50$ nm $\sin\theta_D/2 = 0.0025$ and $\theta_D = 0.29°$; compare to a Bragg angle from $\sin\theta_B = (n\lambda/2d)$ (0.0037 nm/0.05 nm=0.0074). θ_B=0.42°. Notice how θ_D increases for thinner foils and θ_B (and λ) decreases as voltage increases.

26.8 SURFACE EFFECTS

In TEM, we always have thin foils. Dislocation strain fields are long range, but we often assign them a cut-off radius of ~50 nm. However, the thickness of the specimen might only be 50 nm or less, so we can expect the surface to affect the strain field of the dislocation, and vice versa.

When an edge dislocation lies parallel to the surface of a very thin specimen, it causes the specimen to bend. The effect is not large, but large enough compared to the Bragg angle, as illustrated schematically and with an example in Figure 26.16.

Similarly, if the dislocation is dissociated, the proximity of the surface causes its width to decrease (as in Figure 26.16C). We can model this situation using 'image dislocations' as shown in Figure 26.16D. The main point is that we can think of these image dislocations as forcing the partial dislocations closer together; the proximity of the surface can really change the structure of the defect, not just its contrast. A similar effect can occur when the dislocation is inclined to the surface and can result in a V-shaped dissociated dislocations, as we'll show in Section 27.8.

A special interaction between dislocations and surfaces occurs when a dislocation tries to glide out of the material but can't penetrate a surface layer (which might even be amorphous, as in the case of oxide films on metals), as shown in Figure 26.17.

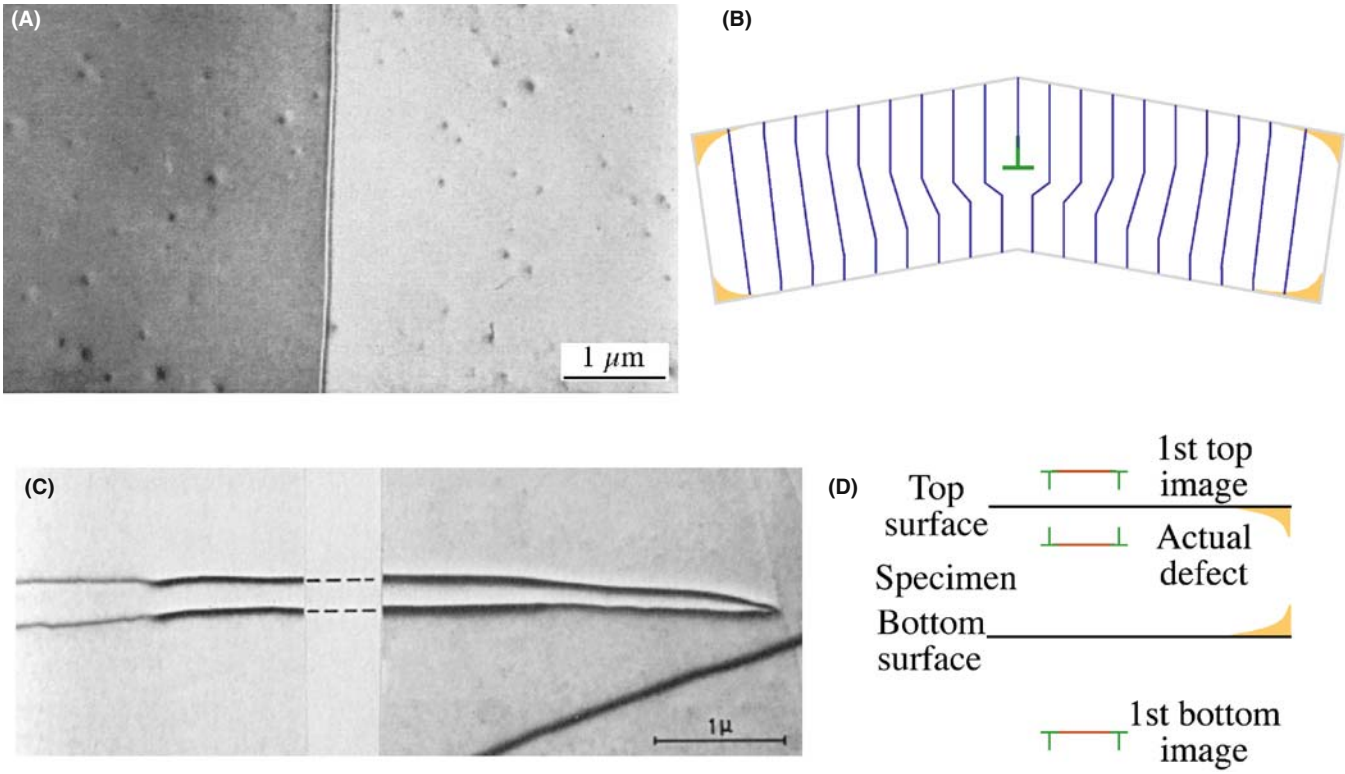


FIGURE 26.16. (A) BF image of a single-edge dislocation lying parallel to the surface of a very thin foil of SnSe, causes the diffracting planes to bend as shown schematically in (B) so we see different contrast intensity in the matrix on either side of the defect. (C, D) If the dislocation is dissociated, the image forces due to the surface cause its width to decrease as it approaches the surface. The schematic in (D) shows two image dislocations included to represent the effect of the nearest surface.

We can have dislocations at the surface just because the structure of the surface layer is different from that of the bulk material. The surface of materials can actually reconstruct. The surface of a very clean (111) Au film is actually more dense than the rest of the film and this misfit is accommodated by surface dislocations, as shown in Figure 26.18. We see the contrast because the strain field extends into the bulk layer. The identification of these dislocations has been confirmed using STM, which also gives more information on the detailed surface structure. However, they were observed first by TEM. The difficulty in TEM studies is that the surface contaminates unless you operate under UHV conditions. The same comment is true for STM of metals but we are used to having a poorer vacuum in TEM and faster/easier specimen exchanges.

GOLD

The dislocation in Figure 26.18 is one atomic plane below the surface but $\mathbf{g}.\mathbf{b}$ still works.

Dislocations can be viewed nearly parallel to their line directions, when we still see contrast even for screw dislocations, as you can see in Figure 26.19A. Initially, this contrast is surprising since $\mathbf{g}.\mathbf{b}$ and $\mathbf{g}.\mathbf{b} \times \mathbf{u}$ must be zero for any dislocation. However, the screw dislocation can relax at the surface, as shown in Figure 26.19B; it gives a twist. You see the screw dislocation even though $\mathbf{g}.\mathbf{b} = 0$ because of this twist.

26.9 DISLOCATIONS AND INTERFACES

Interfaces are, of course, important in all polycrystalline materials. In metals, semiconductors, and thin films on substrates, the interaction between dislocations and interfaces is critical. So now we'll briefly examine the special features we see when combining line and planar defects, as illustrated in Figure 26.20. This is one topic where image simulation, which we'll discuss in Section 26.12, is invaluable. The location of the GB is known because we can see (just) the thickness fringes and the GB dislocations. Notice that there are also 'extrinsic' (lattice) dislocations running into the GB. This is a great figure for student questions!

When we have an array of dislocations, the strain fields overlap so that the value of $\mathbf{R}(\mathbf{r})$ for each dislocation tends to be reduced. This is the GB model of an interface.

Dislocations can be present at interfaces where the composition, or structure, or both change

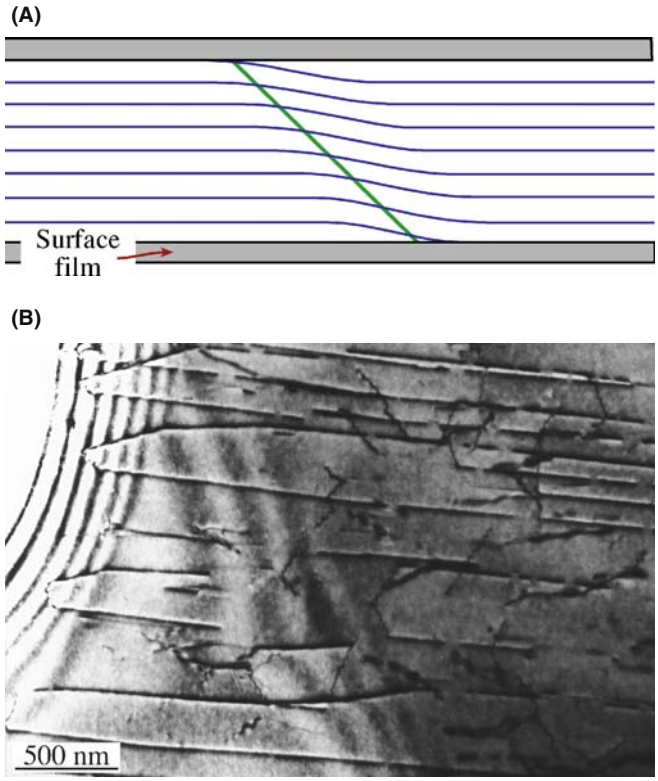


FIGURE 26.17. (A) Schematic diagram of dislocations pinned at the surface of the specimen by surface films such as oxides. (B) A reduced (i.e., metal) film on NiO pins dislocations. Such films may be introduced during or after thinning to electron transparency.

- Misfit dislocations accommodate the difference in lattice parameter between two well-aligned crystalline grains. Surface dislocations (as we saw in Section 26.8) are a special subgroup of misfit dislocations.
- Transformation dislocations are the dislocations that move to create a change in orientation or phase. The $\frac{1}{6}<112>$ dislocations in twin boundaries in fcc materials are an example of transformation dislocations (twinning dislocations).

A complication in the analysis of images of interfacial dislocations is that they are often associated with steps in the interface. There are many examples of such steps in Figure 26.21, including an example of 'step bunching.' (Notice also that **s** is different for the two grains!) Sometimes, as is the case for the $\frac{1}{6}<112>$ twinning dislocations, the dislocations must introduce a step. In other situations, steps are present but there is no dislocation. The difficulty is that we often encounter all three of these situations at the same time. We will also examine these defects using weak-beam conditions in Chapter 27 and using HRTEM in Chapter 28.

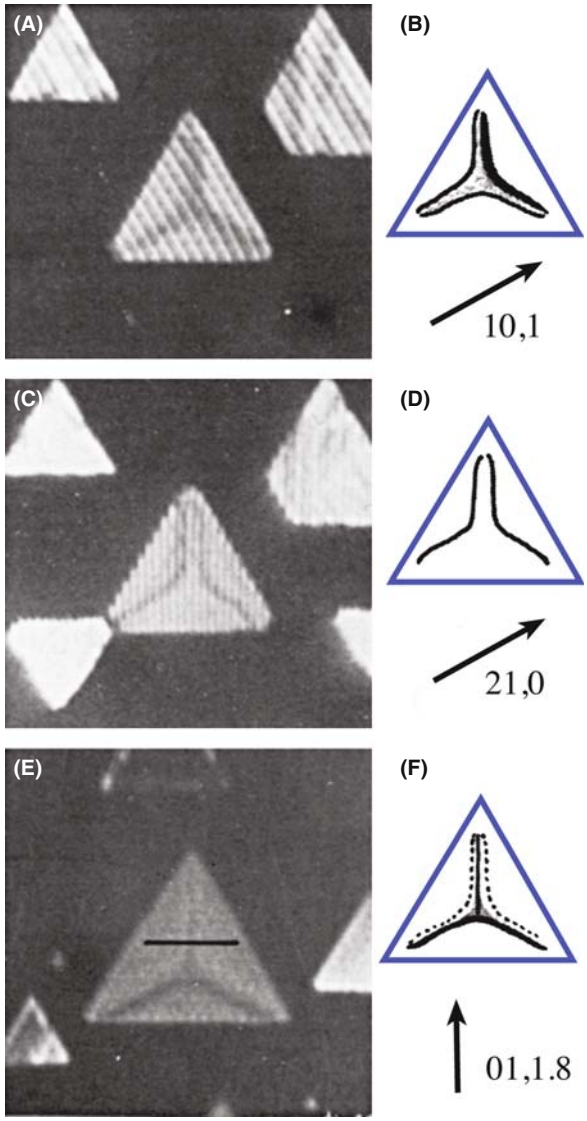


FIGURE 26.18. Dislocation networks can form at the surface of (111) Au islands because the surface layer relaxes to a 'lattice' parameter that is different from that of the bulk material. Different dislocations are visible under different diffracting conditions. The edge misfit dislocations form nodes in the triangular islands and show the usual **g**.**b** contrast behavior. (A), (C), and (E) are experimental images; (B), (D), and (F) are schematics showing **g** and summarizing the contrast.

We will discuss the images first and then, remembering that information must also be present in the DP, we will relate the two.

In many cases that interest us, GBs appear as arrays of dislocations. In general, the grains are misoriented. There are some special cases as we saw in Chapter 24.

- Two grains may have a near-common plane and therefore a nearly common, but different, **g**-vector.
- In small-angle GBs, θ is small, so the separation of the dislocations is large ($\sin\theta/2 = \mathbf{b}/2d$).

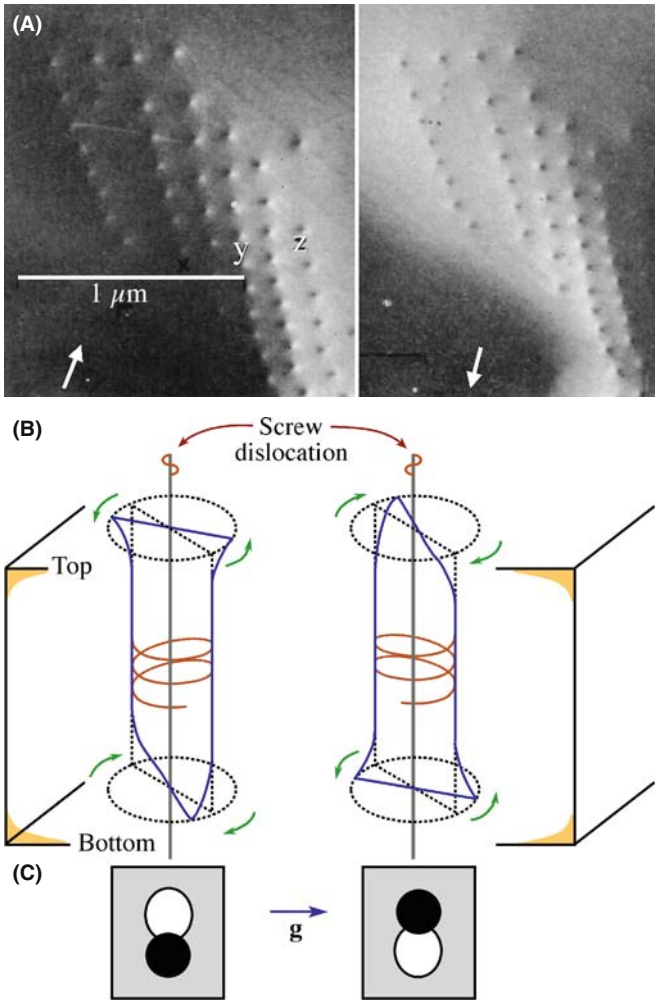


FIGURE 26.19. Screw dislocations viewed end on $\pm\mathbf{g}$. (A) Two images in $\pm\mathbf{g}$. (B) The twist relaxation at the surface. (C) Representations of the resulting contrast. The schematics in (B) show the diffracting planes rotating in the same direction away from the edge-on orientation at both surfaces.

The $\Sigma = 3$ twin boundary in fcc materials is an example of an interface where you can use common, but different, **g** vectors. Here, e.g., the $(3\bar{3}\bar{3})$ plane in one grain is parallel to the $(\bar{5}11)$ plane in the other grain (because these two reflections superimpose), and the lengths of these two **g** vectors are identical, as you can see in Figure 26.22. However, this common reflection would not normally be used to form an image because **g** is rather large. This type of coincidence of dissimilar planes can also occur for other GBs.

In the case of small-angle boundaries, we can pretend that the reflection is common to both grains. The image of the twist GB in Figure 26.23 was formed using such a 'reflection' (i.e., two nearby reflections). (Look back at Figure 17.7 to get an idea of what the DP would be like.) What we are really doing is treating the dislocations as if they were isolated lattice defects; actually, the **g** vectors for the two grains will be rotated relative to one another.

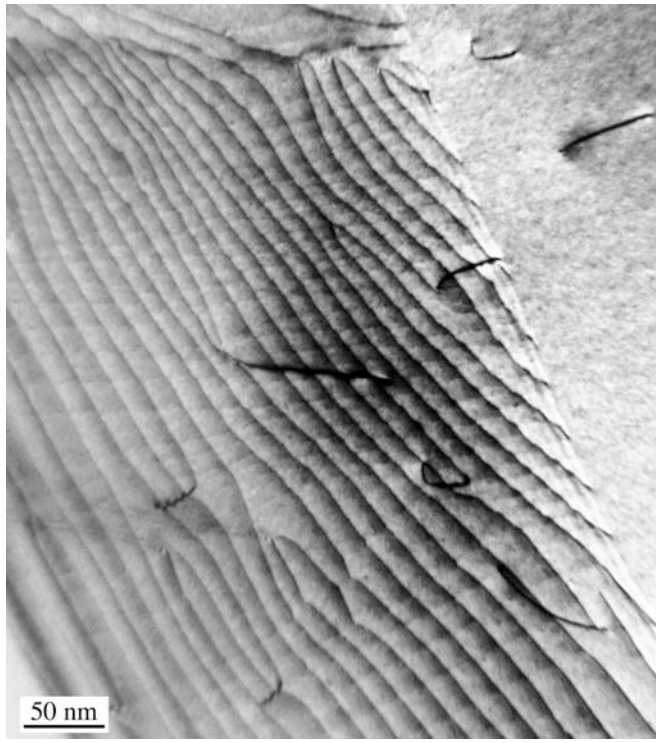


FIGURE 26.20. Dislocations interacting with a GB; the dislocation contrast changes because its strain field changes when it enters the boundary and becomes part of the dislocation structure.

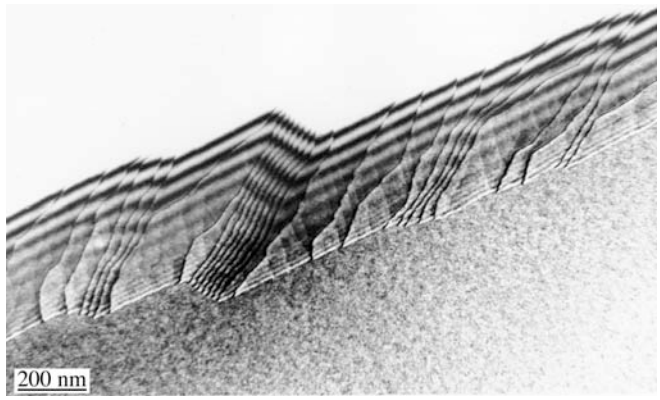


FIGURE 26.21. Steps at interfaces may also cause diffraction contrast when associated with strain. In this Ge specimen, the steps displace the thickness fringes in the GB so they are readily visible. The fringe spacing is different at the top and bottom of the boundary because the diffraction conditions are different at each grain.

Lattice misfit is very important whenever we are studying thin films; dislocations are often present to accommodate misfit. An example is shown in Figure 26.24, where dislocations are present between spinel and NiO; these two materials each have an fcc crystal structure. Although you can easily appreciate the change in lattice parameter, there is also a less obvious change in the elastic constants.

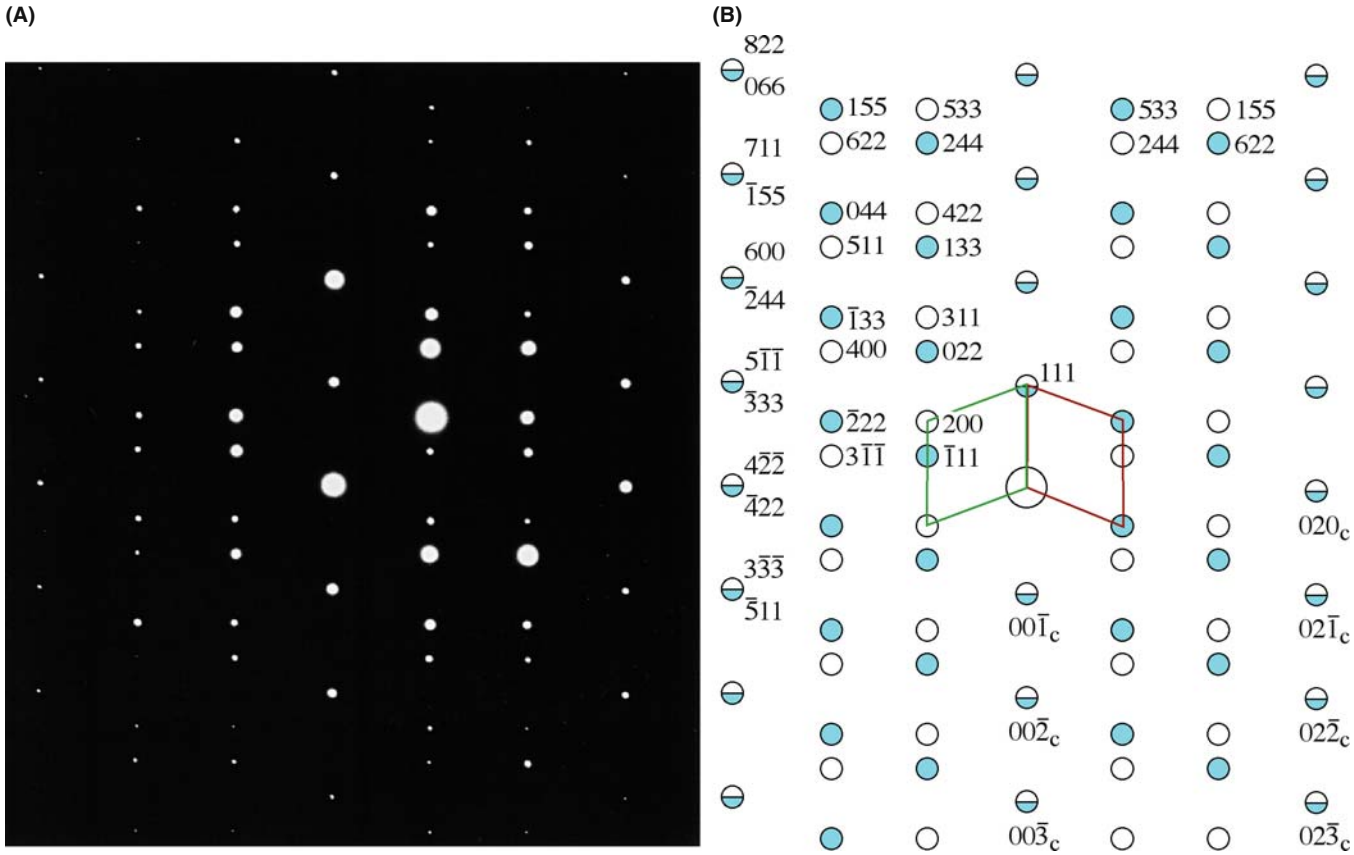


FIGURE 26.22. (A) DP and (B) its indexed schematic for a $\Sigma = 3$ twin boundary in an fcc material. Notice that many pairs of **g**-vectors exactly overlap but have very different indices.

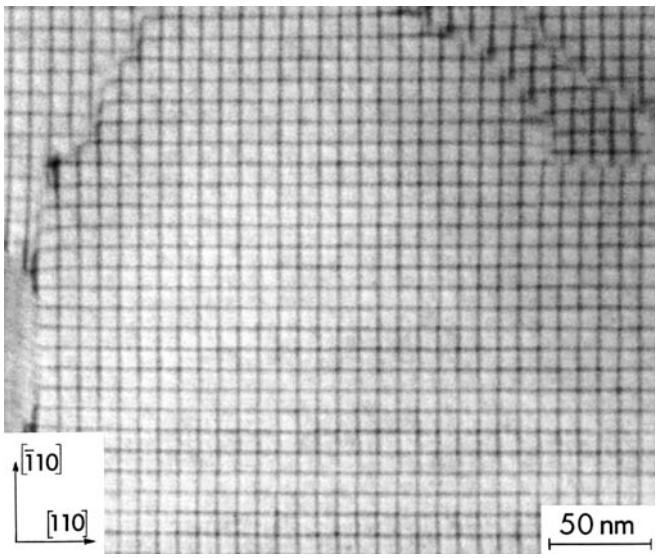


FIGURE 26.23. A low-angle (001) twist boundary in Si oriented almost exactly parallel to the specimen surface. Two (040) reflections were excited to form this BF image, but for small misorientations these are so close that we treat them as one reflection.

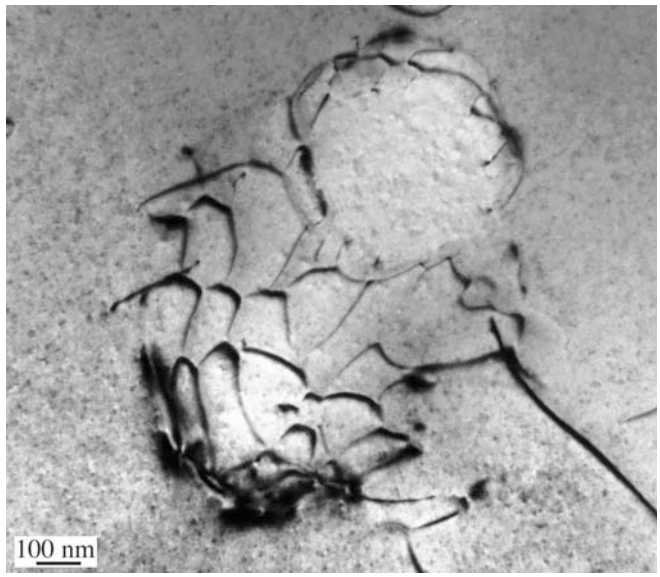


FIGURE 26.24. An irregular array of misfit dislocations at the interface between a spinel particle and a NiO matrix. The lattice mismatch is very small as you can appreciate from the scale. Although you can 'see' a distorted hexagonal array of dislocations, you have to remember that this interface is actually curving within the specimen so that we are only seeing a projection of the structure.

The TEM beam 'sees' yet another change: the extinction distance is different. The result is that, if the crystal is inclined to the electron beam, you will see thickness fringes associated with the interface. Not much work has been done on this, but you may find that it is more difficult to use the **g**.**b** criterion for determining Burgers vectors, especially when the misfit is large.

GB VERSUS PB

This means that the strain field at phase boundaries is *not* the same as at a GB.

Phase transformations often involve the movement of dislocations, generally at semicoherent interfaces. All the conditions discussed above may hold; however, now the dislocations will certainly be associated with a step on the interface to allow the transformation to proceed, so as to physically translate the interface. However, you'll find it difficult to model the contrast from such dislocations, especially when you have a thin layer of the new phase enclosed by the matrix, as is the case when a precipitate grows, as illustrated in Figure 26.25.

The main effect of steps on such interfaces is that they cause a shift in the thickness fringes. It is often difficult to tell if there is also a dislocation present.

We'll summarize some features you should remember when studying dislocations in interfaces

- If the orientation of the grains is different, the distribution of strain from the dislocation may be different in the two grains; the diffraction contrast is determined by this strain field.
- If the chemistry of the two grains is different or if you use different but equal **g** vectors, the extinction distances will be different and the image of the dislocations must therefore be affected.
- Be careful not to confuse moiré fringes with dislocations (we discussed moiré fringes in Chapter 23). The guide is that the dark and light moiré fringes have approximately equal widths; if there is any ambiguity, you should use weak-beam imaging (Chapter 27) and carefully examine the DP.

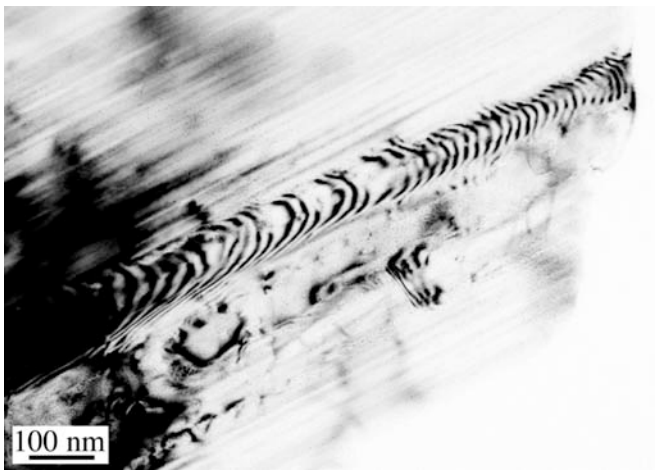


FIGURE 26.25. Transformation dislocations in the interface between a growing lath of hematite (pseudo-hexagonal alumina structure) in a ferrite (cubic spinel structure) matrix. The dislocations are curved because they were moving while heating the thinned specimen, which is why we know they are transformation dislocations, not simply misfit dislocations.

Humble and Forwood have shown using computer simulation of dislocations in interfaces that it is best to use diffraction conditions where a reflection is satisfied in both grains, otherwise the dislocation images tend to be rather featureless relative to the interface thickness fringes.

26.10 VOLUME DEFECTS AND PARTICLES

When the defects are small, the image may be dominated by the strain-field contrast; that is the aspect we are considering here. You have to remember, though, that these defects may have a different structure, lattice parameter, and composition. The theory for a spherical particle in a matrix was given over 40 years ago and is known as Ashby-Brown contrast. The theory works well for coherent particles but as soon as the first interface dislocation forms, analysis becomes much more difficult.

Lattice-strain effects around spherical precipitates appear as lobes of low intensity with a line of no contrast perpendicular to **g**, as shown schematically in Figure 26.26. If you measure the size of the precipitates from a DF image and the size of the strain-contrast lobes in BF, you can get a direct measure of the lattice strain surrounding a single precipitate, which is really quite remarkable. The process requires some specific experimental conditions and careful calibration of the image recording procedure. It is quite a specialized study, so you should read the original references for details. If your precipitates aren't spherical, intuitive interpretation of the images is unreliable and you have to resort to computer simulation.

Figure 26.26B shows how a spherical particle might strain the lattice. Notice that in this case, all the planes continue across the particle so it is coherent and there are no misfit dislocations. The figure here assumes that all the strain occurs in the matrix, which is only true for a hard particle in a soft matrix. The displacement field used to model this situation is

$$\mathbf{R} = C_\varepsilon \mathbf{r} \qquad (26.14)$$

when $\mathbf{r} \leq \mathbf{r}_0$, and

$$\mathbf{R} = C_\varepsilon \frac{r_0^3}{r^3}\mathbf{r} \qquad (26.15)$$

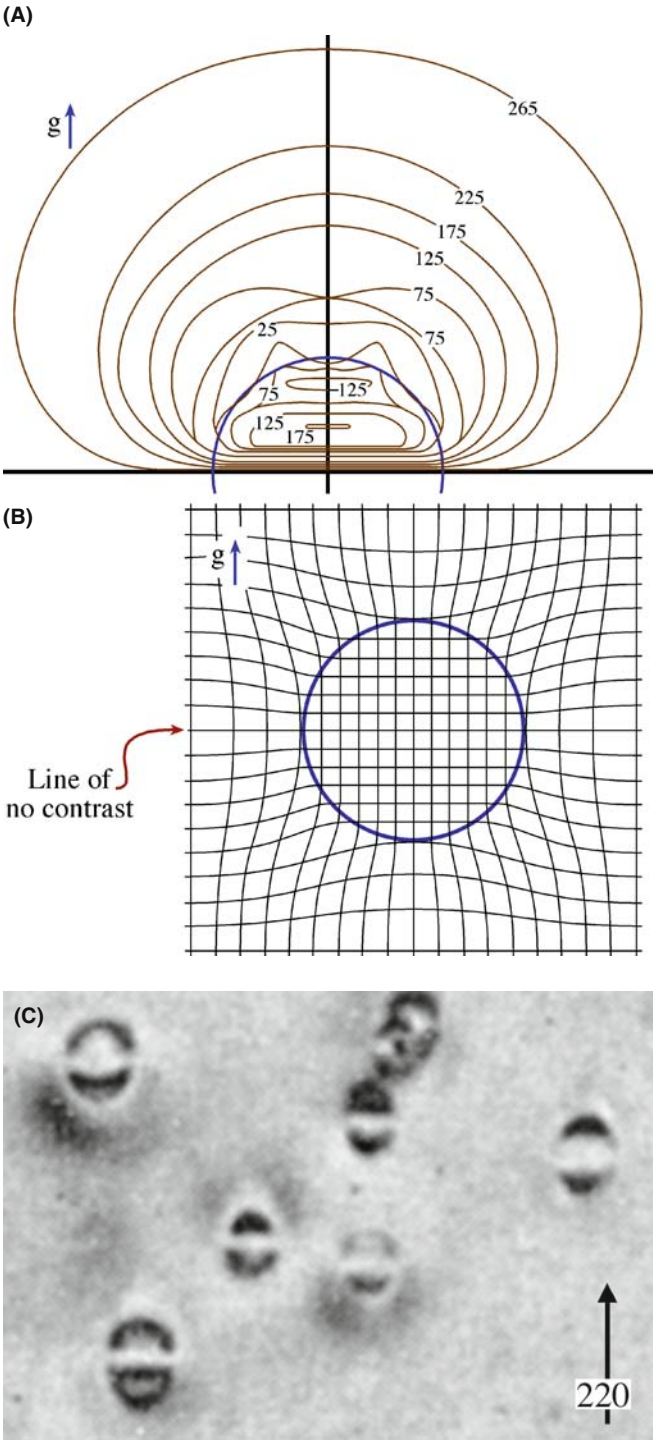


FIGURE 26.26. (A) Intensity contours from a simulated image of a particle like that shown schematically in (B). Notice the line of no contrast which corresponds to the plane which is not distorted by the strain field of the particle. (C) Experimental image of coherent particles in Cu-Co showing strain contrast and a line of no contrast as predicted.

when $\mathbf{r} \geq \mathbf{r}_0$. C_ε is an expression for the elastic constants, given by

$$C_\varepsilon = \frac{3K\delta}{3K + 2E(1+\nu)} \qquad (26.16)$$

K is the bulk modulus of the precipitate; E and ν are Young's modulus and Poisson's ratio, respectively, for the matrix. The important feature is that $\mathbf{R}$ always has radial symmetry. Thus, when we consider the Howie-Whelan equations, we realize that when $\mathbf{g}.\mathbf{R} = 0$ we will see no contrast. So, there will be a 'line of no contrast' normal to $\mathbf{g}$.

The strain can be plotted using the equations given by Ashby and Brown and the image simulated (see below), as shown in Figure 26.26A. In the image from a specimen of a Cu-Co alloy containing small Co precipitates shown in Figure 26.26C, we can see that the images of the particles resemble butterflies or coffee beans. With the improvement in computers, the image contrast expected from much more complex particle geometries can now be calculated and can even consider statistical structural fluctuations.

26.11 SIMULATING IMAGES

It is important that you understand the origins of diffraction contrast from strain fields before you try to simulate this contrast using a computer. Having said that, few students would want to calculate image intensities by hand. The Howie-Whelan equations can be used to simulate images of dislocations, which is especially important when the dislocations are close together. The principal approaches used to simulate diffraction-contrast images were discussed in Sections 25.11–25.13.

Although the algorithm employed by the Head et al. programs allows very fast computation of the image, it does so by restricting the geometry of the defects. To cope with more general geometries, e.g., the strain field from end-on screw dislocations or non-parallel dislocations, we can use Thölén's matrix algorithm. As we saw in Chapter 25, the Howie-Basinski treatment extends the two-beam calculations to include several beams on the systematic row and provides a method for circumventing the column approximation.

PHOTOGRAPHIC FILM

If you want to make quantitative comparisons with real images on film, you must correct for the non-linearity of the film (see Chapters 30 and 31).

26.11.A The Defect Geometry

When choosing the optimal simulation method, depending on the defect geometry, the problem of calculating the image belongs to one of three categories

- *Two-dimensional problem*: including the most general geometries where integration of the full two-dimensional (x,y) grid is necessary.
- *One-dimensional problem*: geometries where the image depends only on either x or y and can be represented by a profile, e.g., problems involving a dislocation parallel to the foil surfaces.
- *GCS problem*: geometries where the method of generalized cross sections (GCS), developed by Head et al., can be applied. Situations where the dislocations and fault planes are parallel to each other, but inclined to the foil surface, are included in this group.

Choosing the best method can speed up the simulations considerably, as we'll show later. The Head et al. program automatically determines the category and selects the appropriate calculation method.

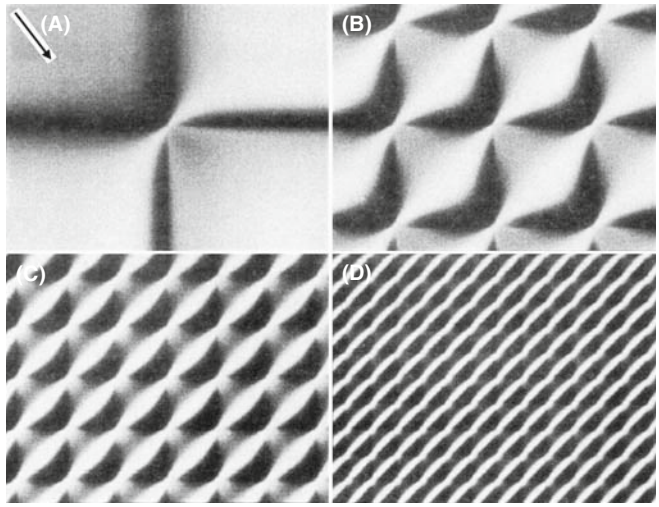


FIGURE 26.27. Comis simulations of two-beam BF images of networks of screw dislocations, located in the middle of a foil with thickness equal to four times the extinction distance, $\xi_{\mathbf{g}}$; $\mathbf{g}.\mathbf{b} = 1$ for both dislocation types. The separation between the dislocations is (A) ∞, (B) $1\xi_{\mathbf{g}}$, (C) $0.5\xi_{\mathbf{g}}$, (D) $0.25\xi_{\mathbf{g}}$.

26.11.B Crystal Defects and Calculating the Displacement Field

The program Comis could simulate amplitude contrast from any number of defects consisting of fault planes and straight, infinite dislocations. You just needed to define the Burgers vector, line direction, and relative position; planar faults were defined by the plane normal, the displacement vector, and the relative position. You could then predefine certain standard geometries to ease the process of defining the defect system.

Once you've defined the defect geometry, you must consider the region of the crystal you want to simulate. In situations where the 'interesting' region is well defined (as in the case of inclined dislocations or intersecting dislocations), Comis determined this region and provided it as the default. However, you could always set the image region manually in Ångström units, to obtain a desired magnification.

The displacement field for the dislocations was calculated in Comis using linear, anisotropic elasticity theory and was based on the algorithms of Head et al., so you had to specify the elastic constants of your crystal. The displacement field then corresponds to straight, infinite dislocations in an infinite medium with no account taken of surface relaxations. In any simulation, you should be able to introduce image dislocations outside the crystal in order to include surface effects.

26.11.C The Parameters

An example which shows simulated images of an orthogonal network of screw dislocations is given in Figure 26.27. Comis could calculate the equilibrium configuration of certain types of interacting dislocations using anisotropic elasticity theory, and then directly incorporates the resulting geometry in subsequent image simulations. As you can appreciate from equation 26.5, in such simulation studies, you will need *all* the parameters for the defects, the specimen, and the diffraction conditions

- The foil thickness.
- The stacking-fault energy.
- The absorption parameters, usually using $|U'_{\mathbf{g}}|/|U_{\mathbf{g}}| = 0.1$.
- The number of beams included in the calculation.
- The zone-axis and the diffracting vectors.
- Also required are the electron energy, the elastic constants, the normal to the foil surface, the Burgers vectors, and the line direction of the dislocations.

The exact beam direction can then be specified by defining the 'center' of the Laue zone, giving the coordinates in terms of the $\mathbf{g}$ vector and $\mathbf{g}_z$. Here $\mathbf{g}_z$ is a specially defined vector in reciprocal space, which is automatically set to lie in the ZOLZ and to be perpendicular to $\mathbf{g}$. Thus, if you place the center of the ZOLZ at (0,0), the specimen is oriented on the zone-axis; if you place the center of the ZOLZ at (0.5,0), it corresponds to being at $\mathbf{g}/2$, i.e., at the Bragg position with the $\mathbf{0}$ and $\mathbf{g}$ beams excited. If you change the second coordinate to be, say, (0.5,0.5), you would include beams from off the systematic row.

CHAPTER SUMMARY

The central idea of this chapter is that the strain field moves atoms off their perfect-crystal positions. We've concentrated on dislocations because the edge dislocation gives the clearest illustration of how the deformation produces the contrast and its structure can be understood with a two-dimensional projection. We can summarize the topics of the chapter as follows

- There is a new feature to the column approximation. The displacement moves atoms out of the column and brings others into the column.
- The basis of the **g**.**b** analysis of a dislocation is simply that the contrast is determined by **g**.**R**(**r**) and that **R**(**r**) is linearly related to **b**. For the screw dislocation, **R**(**r**) is directly proportional to **b**. For the edge dislocation, the image can also be affected by a $\mathbf{g}.\mathbf{b} \times \mathbf{u}$ component which is caused by the buckling of the dislocation glide plane.
- Dislocation images are usually asymmetric. The contrast depends on the sign of $(\mathbf{g}.\mathbf{b})s$.
- As a practical rule, we usually set s to be >0. Then the distortion due to the defect will bend the near-diffracting planes back into the Bragg-diffracting condition to give strong contrast. When $s > 0$, detail in the image is more localized relative to the defect than if we use the $s = 0$ condition.

There are many other situations which are closely related to the topics we've discussed in this chapter. For example, we have not discussed strain contrast associated with crack tips or the analysis of buckling of thin specimens. Although these are rather specialized situations, they do illustrate the growing applications of diffraction contrast in the TEM.

REFERENCES

The text by Hirsch et al. summarizes the early work by the Cambridge group including the derivation of equation 26.10.

BACKGROUND TO DISLOCATIONS AND INTERFACES

Amelinckx, S 1964 *The Direct Observation of Dislocations*, Academic Press New York. A fascinating summary of the early studies by TEM by the man who influenced many. Source of Figure 26.16.

Amelinckx, S 1979 in *Dislocations in Solids* 2, (Ed. FRN Nabarro), North-Holland New York. If you're interested in dislocations, there are many other volumes in this set.

Eshelby, JD, Read WT and Shockley W 1953 *Anisotropic Elasticity with Applications to Dislocation Theory* Acta Metall. **1**, 251–9. An early paper on anisotropic elasticity theory (used by Comis).

Hirth, JP and Lothe, J 1982 *Theory of Dislocations*, 2nd Ed., John Wiley & Sons New York. The definitive textbook, but not for the beginner. Source for equation 26.8, buckling of the glide plane, etc. Not a TEM book.

Hull, D. and Bacon, DJ 2001 *Introduction to Dislocations*, 4th Ed., Pergamon Press New York. A great introductory text.

Matthews, JW, Ed. 1975 *Epitaxial Growth*, Parts A and B, Academic Press New York. Our hero. Unfortunate grammar in the title.

Nabarro, FRN 1987 *Theory of Dislocations*, Dover Publications New York.

Porter, DA and Easterling, KE 1992 *Phase Transformations in Metals and Alloys*, 2nd Ed., Chapman and Hall New York.

Smallman, RE 1985 *Modern Physical Metallurgy*, 4th Ed., Butterworth-Heinemann Boston.

Steeds, JW 1973 *Anisotropic Elastic Theory of Dislocations*, Clarendon Press Oxford, UK. As readable as this subject can be: written by a microscopist. See also Eshelby et al.

Sutton, AP and Balluffi, RW 1995 *Interfaces in Crystalline Materials*, Oxford University Press New York.

Wolf, D and Yip, S, Eds. 1992 *Materials Interfaces, Atomic-level Structure and Properties*. Chapman and Hall New York. A collection of review articles.

IMAGE SIMULATION

Head, AK, Humble P, Clarebrough LM, Morton AJ and Forwood, CT 1973 *Computed Electron Micrographs and Defect Identification* North-Holland New York.

Morton, AJ and Forwood CT 1973 *Equilibria of Extended Dislocations* Cryst. Lattice Defects **4** 165–177. TEM of arrays of interacting dislocations (Section 26.11.C).

Humble, P and Forwood, CT 1975 *Identification of Grain Boundary Dislocations I and II* Phil. Mag. **31**, 1011–23 and 1025–48. Simulating images of interfaces.

Rasmussen, DR and Carter, CB 1991 *A Computer Program for Many-Beam Image Simulation of Amplitude-Contrast Images* J. Electron Microsc. Techniques **18**, 429. Description of Comis, a great program that died with the advances in operating systems.

Thölén, AR 1970a *Rapid Method for Obtaining Electron Microscope Contrast Maps of Various Lattice Defects* Phil. Mag. **22** 175–182. Matrix algorithm for more complex geometries.

Thölén, AR 1970b *On the Ambiguity between Moiré Fringes and the Electron Diffraction Contrast from Closely Spaced dislocations* Phys. stat. sol. *(a)* **2** 537–550. Applying the algorithm to a network of orthogonal dislocations (Section 26.11.C).

Thölén, AR and Taftø, J 1993 *Periodic Buckling of the Lattice Planes in the Thin Regions of Wedge-Shaped Crystals* Ultramicrosc. **48** 27–35. TEM of buckled specimens: challenging exercise.

CONTRAST THEORY

Amelinckx, S 1992 in *Electron Microscopy in Materials Science*, (Eds PG Merli and MV Antisari), World Scientific River Edge NJ. Includes discussion of dislocation strain field relaxing at the surface.

Amelinckx, S and Van Dyck, D 1992 in *Electron Diffraction Techniques* **2** (Ed. J.M. Cowley), p. 1, Oxford University Press New York.

Ashby, MF and Brown, LM 1963 *Diffraction Contrast from Spherically Symmetrical Coherency Strain* Phil. Mag. **8** 1083–1103 and *On Diffraction Contrast from Inclusions* Phil. Mag. **8** 1649-1676. Ashby-Brown contrast.

de Graef, M and Clarke, DR 1993 *Strain Contrast at Crack Tips for in-situ Transmission Electron Microscopy Straining Experiments* Ultramicrosc. **49**, 354–365. TEM of crack tips: challenging exercise.

Goringe, MJ 1975 in *Electron Microscopy in Materials Science* (Eds. U Valdré and E Ruedl), p. 555, Commission of the European Communities Luxembourg. Further discussion of equation 26.10 and much more.

Hirsch, PB, Howie, A and Whelan, MJ 1960 *Kinematical Theory of Diffraction Contrast of Electron Transmission Microscope Images of Dislocations and Other Defects* Phil Trans Roy Soc. **A252** 499–529. Early paper you should read, with more on the column approximation.

The original series of papers by H Hashimoto, PB Hirsch, A Howie, MJ Whelan in Proc. Roy. Soc. London A **252** 499 (1960), **263** 217 (1960), **267** 206 (1962), and **268** 80 (1962) are strongly recommended.

ON THE NON-COLUMN APPROXIMATION

Howie, A and Basinski ZS 1968 *Approximations of the Dynamical Theory of Diffraction Contrast* Phil. Mag. **17** 1039–1063.

Howie, A and Sworn H 1970 *Column Approximation Effects in High Resolution Electron Microscopy using Weak Diffracted Beams* Phil. Mag. **31** 861–864.

EXAMPLES

Hughes, DA and Hansen, N 1995 *High Angle Boundaries and Orientation Distributions at Large Strains* Scripta Met. Mater. **33** 315–321. Particularly clear illustration of the value of producing a large specimen area.

Karth, S, Krumhansl, JA, Sethna, JP and Wickham, LK 1995 *Disorder-Driven Pretransitional Tweed Pattern in Martensitic Transformations* Phys. Rev. B **52** 803–822. Early work considering statistical structural fluctuations on image contrast (in Section 26.10).

Takayanagi, K, Tanishiro, Y, Yagi, K, Kobayashi, K and Honjo, G 1988 *UHV-TEM Study on the Reconstructed Surface of Au(111): Metastable* $p'' \times p''$ *and Stable* $p \times 1$ *Surface Structure* Surf. Sci. **205** 637–651. They showed surface dislocations by TEM before they were discovered by STM.

Tunstall, WJ, Hirsch, PB and Steeds, JW 1964 *Effects of Surface Stress Relaxation on Electron Microscope Images of Dislocations Normal to Thin Metal Foils* Phil. Mag. **9** 99–119. Classic paper on the contrast seen when a screw dislocation intersects a surface.

Wilkens, M 1978 in *Diffraction and Imaging Techniques in Materials Science*, 2nd Ed. (Eds S Amelinckx, R Gevers and J Van Landuyt) p.185 North-Holland New York. A detailed analysis of butterflies, lozenges and peanuts (Figure 26.12).

THE COMPANION TEXT

The programs TEMACI and Cufours are available for simulating these defects. TEMACI was designed for the study of small defects.

SELF-ASSESSMENT QUESTIONS

Q26.1 Give the physical reason for the change in contrast near a dislocation.
Q26.2 We are given an assignment to perform a computer simulation of contrast due to defects. The assignment includes areas of a hypothetical specimen with dimensions of about 1 nm. Is the column approximation valid?
Q26.3 A screw dislocation in your specimen appears to be out-of-contrast. Does this mean $\mathbf{g}.\mathbf{R}=0$?
Q26.4 An edge dislocation has $\mathbf{g}.\mathbf{b} = 0$ but it still shows strong contrast. Why?
Q26.5 Is the image of a dislocation located directly at the projected position of the dislocation?
Q26.6 Is the information in the image of dislocation core located coming from the reflection **g**?
Q26.7 You are looking at a region in the image that is about 1 μm in diameter. Could this be the strain field from a defect?
Q26.8 You see abrupt steps in the thickness fringes at different locations in the specimen. What might cause this?
Q26.9 How can you 'see' a dislocation and why do we use quotes in this question?
Q26.10 What parameters are needed to describe a particular dislocation?
Q26.11 How do we determine the Burgers vector of a dislocation?
Q26.12 How is the Burgers vector determined for dislocation nodes or networks?
Q26.13 What complications arise for small dislocation loops?
Q26.14 Why do dislocation dipoles show 'inside-outside' contrast?
Q26.15 What happens to $\mathbf{g}.\mathbf{b}$ dislocation contrast when **g** is reversed?
Q26.16 How does using a higher kV help in 'viewing' dislocations?
Q26.17 How does using a lower kV help in 'viewing' dislocations?
Q26.18 Give two ways that the surface can affect the image of a dislocation.
Q26.19 Why do we talk about $\mathbf{g}.\mathbf{b}$ rather than $\mathbf{g}.\mathbf{R}$?
Q26.20 How does the term $\mathbf{g}.\mathbf{b}\times\mathbf{u}$ relate to $\mathbf{g}.\mathbf{R}$?

TEXT-SPECIFIC QUESTIONS

T26.1 Consider Figure 26.1A. How wide do you expect the dislocation image (the peak in 'B') to be? Just give your thought process without doing displacement-field calculations, etc.
T26.2 Consider Figure 26.3. Why do we always like to have **b** lying in the plane of the specimen? (Give at least three different valid reasons.)
T26.3 Figure 26.5 suggests that if we keep $\mathbf{g}.\mathbf{b}$ the same but reverse the sign of s, the image of the dislocation will arise from the other side of the dislocation core. Use your analysis of the Howie-Whelan equations to explain this result.
T26.4 Give a physical explanation of equation 26.12 and discuss how it differs from using **R**.
T26.5 Consider Figure 26.6A. Give a consistent set of **b** vectors for all the dislocations you see.
T26.6 Consider Figure 26.6C. By consulting the original source or otherwise, explain the variation in the separation of the dislocation images.
T26.7 Consider the lower half of Figure 26.6G,H. Explain why all the dislocations are visible in the two **g** vectors.
T26.8 Consider Figure 26.8. Suggest a set of consistent **g** vectors for the four images.
T26.9 Consider Figure 26.10. Why are the fringes on one loop running in a different direction? Suggest what **g** might be and include it on the diagram.
T26.10 Consider Figure 26.11. Explain how $\mathbf{g}.\mathbf{b}\times\mathbf{u}$ can lie in the plane of the loop and thus confirm the interpretation of the contrast in this image.
T26.11 Consider Figure 26.16A and B. Estimate the angle through which the specimens bend.
T26.12 How can we increase the information on dislocations parallel to the specimen surface if dislocations with values of $\mathbf{g}.\mathbf{b}$ which are <0.3 can't be distinguished from those with $\mathbf{g}.\mathbf{b} = 0$?

27

Weak-Beam Dark-Field Microscopy

CHAPTER PREVIEW

The term 'weak-beam microscopy' refers to the formation of a diffraction-contrast image in either BF or DF where the useful information is transferred by weakly excited beams. The DF approach has been more widely used, in part because it can be understood using quite simple physical models. It also gives stronger contrast; we see white lines on a dark gray background. This chapter will be concerned only with the DF approach. Historically, the weak-beam dark-field (WBDF often abbreviated to WB) method became important because under certain special diffraction conditions, dislocations can be imaged as narrow lines which are approximately 1.5 nm wide. Equally important is the fact that the positions of these lines are well defined with respect to the dislocation cores; they are also relatively insensitive to both the foil thickness and the position of the dislocations in the specimen. The technique is particularly useful if you are studying dissociated dislocations where pairs of partial dislocations may only be ~4 nm apart and yet this separation greatly affects the properties of the material.

We first choose a particular **g** and bring this onto the optic axis as if intending to form a regular on-axis DF image. We then tilt the specimen to make $\mathbf{s_g}$ large and examine the DF image using reflection **g**. If a defect is present, the diffracting planes may be bent locally back into the Bragg-diffracting orientation to give more intensity in the DF image. The problem is that, as we increase $\mathbf{s_g}$, the average intensity decreases as $1/s^2$; in the DP the beam appears as a weak spot, hence the name. When $\mathbf{s_g}$ is large, the coupling between **g** and the direct beam becomes small and the diffracted beam is said to be 'kinematically diffracted.' So, this chapter is where we will discuss the 'kinematical approximation'.

- You will sometimes see reference to the **g**(3**g**) WB condition. Beware! Sometimes you don't need to be this weak; sometimes this is not weak enough.
- It is often not the fact that **s** is large that is important; what is important is that ξ_{eff} is small.

This chapter is unusual in that it deals with a special imaging technique, rather than a concept or theory. Also, WBDF is only really useful when the specimen is not perfect, i.e., when you are interested in defects in the specimen or small changes in thickness. Therefore, you can skip this chapter if crystal-lattice defects are not relevant to your microscopy study. If you are interested in defects, you will find that this chapter really covers much more than WB microscopy. For example, we will use concepts developed for diffracted beams and carefully set the excitation error, $\mathbf{s_g}$, by referring to the Kikuchi-line pattern. In Section 27.9, we will discuss some of the ways that new developments in TEM design are changing the way we do WB microscopy and how we interpret the images.

27.1 INTENSITY IN WBDF IMAGES

We showed in Chapter 13 that, in a two-beam situation, the intensity of the diffracted beam **g** in a perfect crystal can be written as

$$|\phi_{\mathbf{g}}|^2 = \left(\frac{\pi t}{\xi_{\mathbf{g}}}\right)^2 \cdot \frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} \qquad (27.1)$$

Remember that when we derived this expression we assumed that only two beams, O and G, are important.

We will consider complications which arise when more beams are present in Section 27.9. The important variables in equation 27.1 are the thickness, t, and the effective excitation error, s_{eff}, which was given by equation 13.47

$$s_{\text{eff}} = \sqrt{s^2 + \frac{1}{\xi_{\mathbf{g}}^2}} \qquad (27.2)$$

In the WB technique we increase s to about $0.2\,\text{nm}^{-1}$ so as to increase s_{eff}. (In most WB papers you will see this value as 2×10^{-2} Å^{-1}; remember 50 Å is 5 nm.) This large value of s means that s_{eff}, and therefore the intensity, $I_{\mathbf{g}}$, becomes independent of $\xi_{\mathbf{g}}$ except as a scaling factor for t (in the prefactor in equation 27.1). The actual value of **s** can be set by carefully positioning the Kikuchi lines for the systematic row of reflections, which includes **g**. You can best appreciate this effect by calculating a range of values for **s**. Remember when doing this that you must specify **g** and the kV because, as we saw in Chapter 13, $\xi_{\mathbf{g}}$ varies with both the reflection used to form the WB image and the energy of the electrons, as implied in equation 13.47

$$\xi_{\text{eff}} = \frac{\xi_{\mathbf{g}}}{\sqrt{w^2 + 1}} \qquad (27.3)$$

Practical Considerations: As the value of s_{eff} increases, equation 27.1 shows that the intensity of the G beam decreases very rapidly. The result is that the exposure time needed to record the image on the CCD or on photographic film also rapidly increases and has, in the past, been the factor which limited the usefulness of the technique. Although manufacturers may guarantee a drift rate of less than 0.5 nm per minute for new machines, values of six times this rate are common on many older instruments. In early WB studies, this problem was partly overcome by using photographic film with a fast emulsion or by modifying photographic processing conditions. In either case, the grainy appearance of the photographic emulsion would be increased. The problem of drift can, in principle, be overcome by using a video system to record the image and capturing frames from the video. You could then, in principle, reduce the noise by frame averaging, particularly if you can take account of any drift. The causes of drift (specimen and thermal effects) and their correction or minimization are discussed in Chapter 8, but it is worth reminding you that changes in the temperature of the water in the objective lens are a major cause of drift. Although WBDF imaging may aim for 0.5 nm resolution rather than 0.2 nm in HRTEM, exposure times are often 10 times greater for WBDF than HRTEM, so drift may be even more important.

ON EQUATION 27.2

If $s >> \xi_{\mathbf{g}}^{-2}$ then $s \approx s_{\text{eff}}$ so that equation 27.2 becomes what is known as the 'kinematical equation'; the kinematical equation cannot be applied for small **s** unless the thickness, t, is also very small.

27.2 SETTING S_g USING THE KIKUCHI PATTERN

Since the contrast in the WB image is so dependent on the value of $\mathbf{s}_{\mathbf{g}}$ we need a method for determining $\mathbf{s}_{\mathbf{g}}$. We draw a line through the **g**-systematic row and let the Ewald sphere cut through this line at $n\mathbf{g}$ where n is not an integer. Figure 27.1 illustrates this situation. How can we 'see' what the value of n is? Of course we can't, since we are looking approximately normal to the ZOLZ. We can't just judge n by looking at the intensities of the spots except in special circumstances.

The solution to the problem can be appreciated by looking at the Kikuchi pattern shown in Figure 27.2. When **g** is exactly at the Bragg condition, the **g** Kikuchi line passes through the **g** reflection; when 3**g** is exactly satisfied, the 3**g** Kikuchi line passes through the 3**g** reflection. We might guess that n is ~3.2 in Figure 27.2, but we don't have a 3.2**g** Kikuchi line; we have to deduce this value of n from the position of the 3**g** Kikuchi line. Remember (from Chapter 19) that when the 3**g** Kikuchi line passes through 3.5**g** on the **g**-systematic row, the 4**g** Kikuchi line and the Ewald sphere pass through the 4**g** reflection as shown in Figure 27.3. Therefore, when the Ewald sphere passes through 3.2**g**,

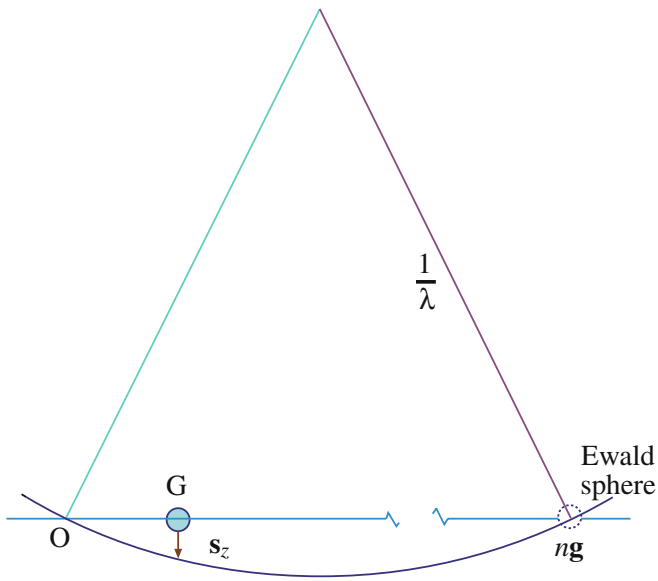


FIGURE 27.1. The Ewald-sphere construction showing the diffraction conditions used to obtain weak-beam images. The sphere cuts the row of systematic reflections at 'n**g**' where n is not necessarily an integer.

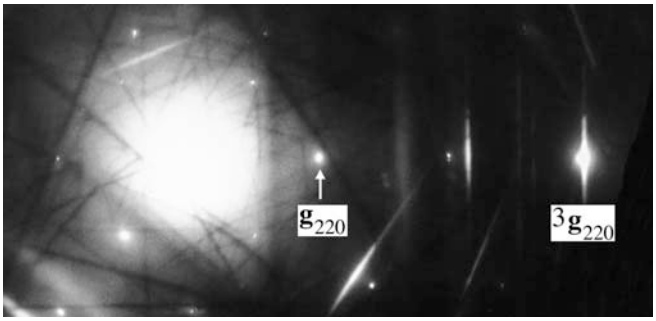


FIGURE 27.2. DP obtained when the specimen is tilted to a suitable orientation for WB microscopy. Here **g** is a 220 reflection and 3**g** is strong.

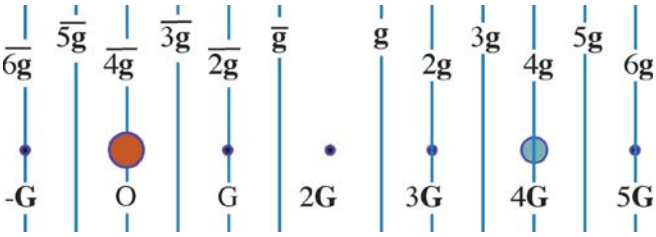


FIGURE 27.3. A schematic diagram showing the positions of the Kikuchi lines for the systematic row of reflections when 4**g** is excited.

the 3**g** Kikuchi line will pass through 3.1**g**; we can express this simple geometric result as

$$n = 2m - N \tag{27.4}$$

where N**g** refers to the Kikuchi line closest to n**g** (N is an integer) and m**g** is the location of the Kikuchi line as we measure it. In the example above, we can choose N to be 3 so that if m is 3.1 then n is 3.2; if, instead, we choose N to be 4, then m is 3.6 (because we measure the position of the 4**g** Kikuchi line) and n is still 3.2. Having determined n we need to estimate **s**. This we do using the expression

$$\mathbf{s} = \frac{1}{2}(n-1)|\mathbf{g}|^2\lambda \tag{27.5}$$

which you can derive from Figure 27.4 using the intersection chord theorem ($ab = cd$) and the fact that $1/\lambda$ is much larger than **s**.

You can immediately appreciate some important results from this expression

- Setting $n = -1$ gives the same value of **s** as for $n = 3$ but the sign is reversed.
- The magnitude of **s** is more strongly dependent on $|\mathbf{g}|$ than on λ but it depends on *both*.
- The specific nature of the material enters through **g**, the microscope affects **s** through λ.

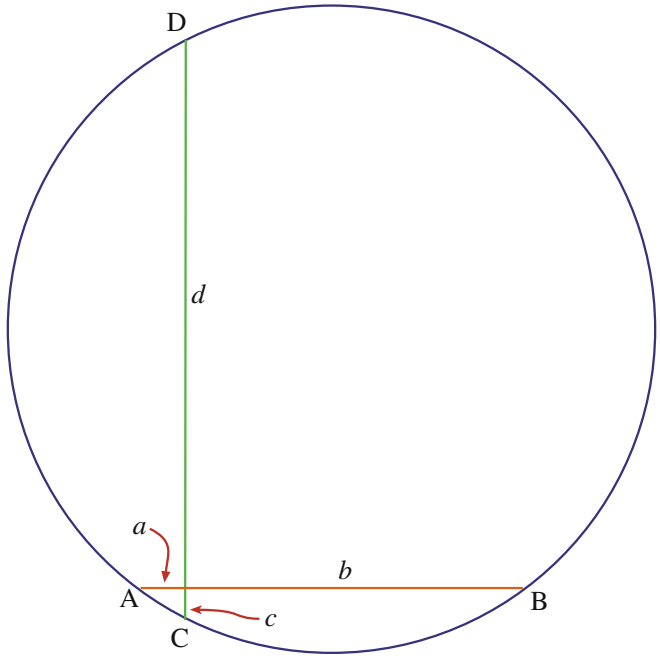


FIGURE 27.4. The intersection chord construction used to deduce the value of $\mathbf{s_g}$: we approximate d to $2/\lambda$, c is **s**, a is $|\mathbf{g}|$, and b is $(n-1)|\mathbf{g}|$.

Here we recommend that you use a spreadsheet to calculate different values of **s** as you vary **g** or λ. A selection of these for Cu and Si is given in Table 27.1.

As an exercise, you can use equations 27.2 and 27.5 to calculate s_{eff} with $3\mathbf{g} = 42\,\text{nm}$ for the 220 reflection in Cu at 100 kV. You can then see when s_{eff} becomes 'independent' of $\xi_{\mathbf{g}}$. Next, repeat the exercise with other values of λ or for other reflections and materials.

One point you should bear in mind is that none of the above discussion requires a particular value for **s** and yet you will often read that **s** must be $\geq 0.2\,\text{nm}^{-1}$ for a WB image. This value of **s** is recommended when you are studying defects quantitatively, because computer calculations show that the position of the image can then be directly related to the position of the defect. You will find that smaller values of **s** will often give you WB images which contain the information you want and you can more easily see and record the image!

TABLE 27.1. Values of s (nm^{-1}) for Various kV

Si	Cu	Accelerating voltage			
n_{Si}	n_{Cu}	100 kV	200 kV	300 kV	400 kV
4.9	2.8	0.20	0.14	0.11	0.09
6.9	3.6	0.30	0.20	0.16	0.13
8.5	4.3	0.38	0.25	0.20	0.17
9.9	4.9	0.44	0.30	0.24	0.20

For Cu, $a = 0.3607\,\text{nm}$; for Si, $a = 0.534\,\text{nm}$. The values of n are accurate to better than 0.1. In both cases the calculation is for $\mathbf{g} = 220$.

27.3 HOW TO DO WBDF

The nature of the WB image imposes a restriction on the maximum specimen thickness which you can use because the visibility of such images decreases as the thickness increases (due to a corresponding increase in inelastic scattering). However, the orientation of the specimen is accurately set by reference to the Kikuchi lines observed in the DP (Chapter 19), and these are not visible in specimens which are too thin. Therefore, your specimen thickness must be greater than a certain minimum value. Also, if the observations you made on certain defects, in particular, dislocation ribbons and nodes, are to be interpreted as representative of the bulk and not influenced by the surface of the foil, then again the foil must not be too thin. You can generally satisfy these requirements by selecting defects for detailed study which, in the case of Cu and its alloys, lie in areas which are about 70 nm thick.

Due to the very low intensity of the WB images, exposure times required are typically on the order of 4–30 seconds using Kodak SO-163 film (CCD cameras are better!). The main factor limiting the exposure time is the inherent instability of any specimen stage. To minimize the exposure time, you can usually use a highly convergent (or divergent) beam, contrary to the assumption implicit in the simple theory. The effect of this convergence is that oscillations in image intensity and position, which result from variations in depth and thickness parameters, are diminished.

WB STEP-BY-STEP

We'll go through how to set up the **g**(3.1**g**) diffraction mode condition since this is widely used in practice. Actually, it is generally referred to as '**g**(3**g**) with $\mathbf{s}_{3\mathbf{g}}$ positive' because we guess the value of n by estimating m. This condition ensures that the 3**g** reflection is not satisfied and also that you can use the BF **0**(**g**) image, with $\mathbf{s}_{\mathbf{g}}$ slightly positive, to locate the defect and to focus your image.

The first two steps in setting up the WB conditions are illustrated in Figure 27.5, relating what happens in the Ewald-sphere model with what you see happening to the Kikuchi lines.

- Orient the specimen in BF so that **g** is excited and $\mathbf{s}_{\mathbf{g}}$ is just greater than zero. Make sure that no other reflections are excited (see upper diagram in Figure 27.5).
- Use the DF beam-deflecting coils to bring the reflection **g** on to the optic axis. Use the binoculars because **g** becomes very weak; underfocus the beam before you use the high-resolution screen.
- Insert the objective aperture. In BF, check that the aperture is centered, then switch to DF and check that the spot G is centered in the aperture.
- Fine-tune your conditions looking at the DP with G centered.
- Go to imaging mode; you now have a WB image with the required **g**(3.1**g**) condition (see lower diagram in Figure 27.5).

Since you have inserted a small objective aperture, you should now check the objective astigmatism. We use a small objective aperture so as to remove inelastic scattering; remember that this aperture will then limit our potential resolution. If you focus the beam, you may change the position of the beam and probably the astigmatism! Remove the objective aperture and check that no other reflections are strongly excited when you are in DF. Then repeat the process starting at the third step (insert the objective aperture).

After finely focusing the image, record it together with its SADP.

If you're not sure why this 'trick' for setting the **g**(3.1**g**) condition works, go back to Chapter 19 and draw the systematic row and the corresponding Kikuchi lines. Then move the spots, while keeping the crystal, and thus the Kikuchi lines, fixed. You may see the conditions **g**($\bar{\mathbf{g}}$) used. This was the original condition suggested by Cockayne et al. (1969); it does give you the same values of **s** as the **g**(3.1**g**) condition, but the interband scattering processes are different and it is not so convenient to change from BF to WBDF. However, you may find variations on this condition useful when you need to use a **g**(n**g**) condition with a large value of n.

WB USING CDF

Note that we are using CDF imaging—we are not comparing contrast in BF and DF images.

Weak-beam microscopy becomes much easier when you are comfortable with using a CCD camera. Just being able to see whether the image is moving due to specimen drift can save boxes of photographic film. As we mentioned earlier, you could use frame averaging to reduce the noise. However, you will realize that the extra magnification from the video system tends to limit the area which can be viewed, so that film is still preferred by many WB practitioners but this is changing as higher resolution CCD cameras become affordable. You will find that 30 k× magnification on the plate is a good compromise; without video, you will use 50–60 k× to allow you to focus with the binoculars.

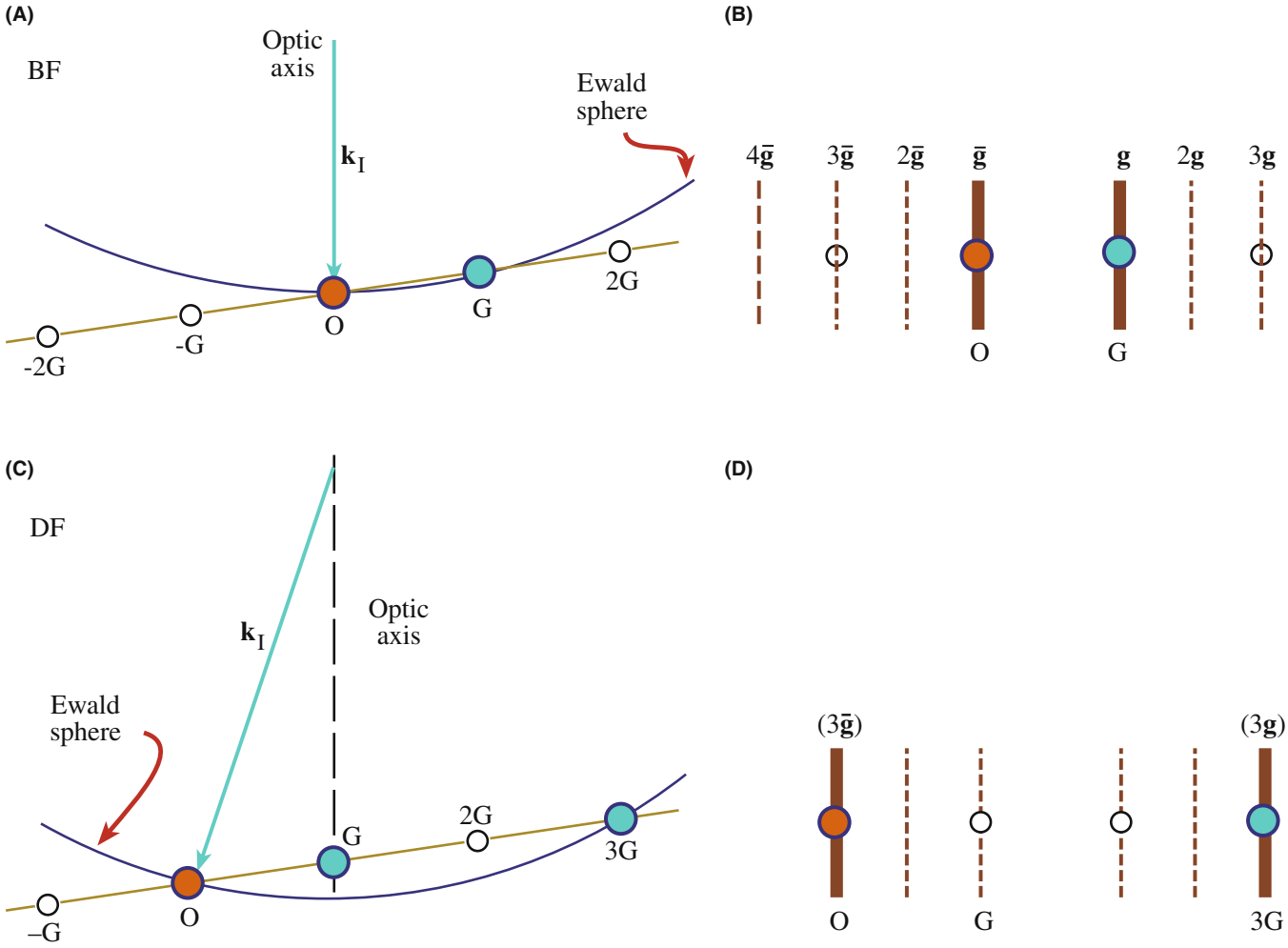


FIGURE 27.5. Relationship between the orientation of the Ewald sphere and the position of the Kikuchi lines for the **0**(**g**) (A and B) and **g**(3**g**) (C and D) diffraction conditions. The two pairs of diagrams are related by tilting the beam; the specimen has not tilted so the position of the Kikuchi lines is unchanged.

27.4 THICKNESS FRINGES IN WEAK-BEAM IMAGES

Thickness fringes in WB images are just like thickness fringes in strong-beam images but the effective extinction distance, ξ_{eff}, is much smaller. From equation 27.1 we can see that the intensity minimum occurs at thicknesses of $\mathcal{N}(s_{\text{eff}}{}^{-1})$ with maxima at $(\mathcal{N} + \frac{1}{2})\mathcal{N}(s_{\text{eff}}{}^{-1})$. The effective extinction distance for $s = 0.2\,\text{nm}^{-1}$ is 5 nm; this value is rather sensitive to the precise value of s so that the fringes will change if the foil bends. Using WB images we can form a rather detailed contour map of the specimen, but you must remember that both surfaces may be inclined to the beam, as shown in Figure 27.6.

The thickness effect is illustrated in Figure 27.7. These images were recorded with $s = 0.2\,\text{nm}^{-1}$. The MgO specimen has been heat-treated so that there are

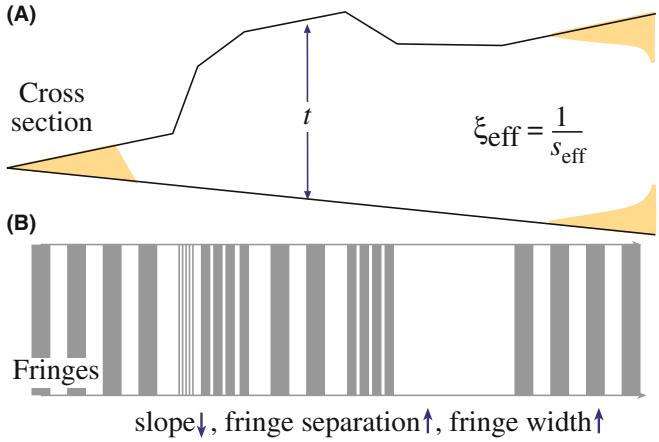


FIGURE 27.6. (A) In WB imaging, the thickness periodicity depends on the effective extinction distance, ξ_{eff}. (B) The separation of the fringes varies accordingly.

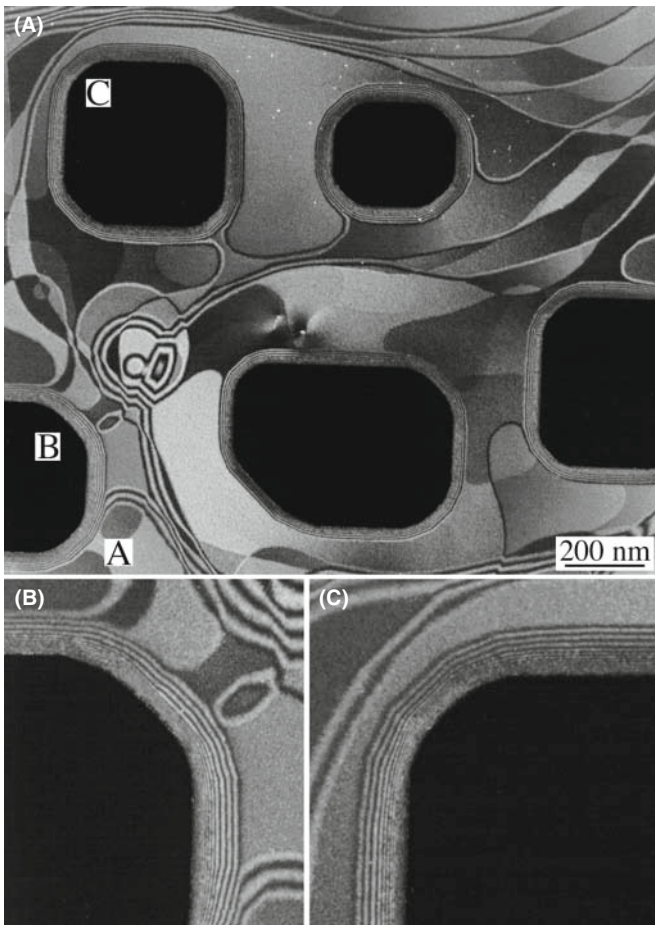


FIGURE 27.7. (A) WB thickness fringes from annealed MgO. (B, C) Higher magnifications of regions B and C. Compare to Figure 23.3.

large regions where the surface is atomically flat on both sides. Before heating, the specimen had been acid etched, which caused the holes seen in this image (they're black because we're in DF); you can see the defects causing the etch pits in some pits. The specimen shows inclined steps which curve across the surface. Where we see wide, uniform gray regions the surface is atomically flat. At A there is a large inclined step which runs into the hole B. Notice how the number of fringes around the hole B increases at A. Around the holes (Figure 27.7B and C) we see much more closely spaced fringes because the thickness changes more quickly here. Now if we look at the edge of any hole such as C, we see that the spacing of the fringes has one value away from the hole but another, smaller value close to the hole. What we find is that the inclined surface facets on different planes, with each facet becoming steeper closer to the hole. This topology is a result of the way the specimen has been prepared and would not normally be found, say, in electrochemically polished specimens, but it does illustrate the possibility of 'profiling' using thickness fringes.

27.5 IMAGING STRAIN FIELDS

The principle of the technique is very simple. When the area of your specimen in which the defect of interest lies is oriented away from the Bragg position, the reflecting planes may be bent back into the reflecting position close to the defect. The region over which this occurs is very small because the strain has to be quite large to cause this bending. For the (220) planes in Cu (which fixes the plane spacing, d), the planes must rotate through an angle of $\sim 2^\circ$ to change $\mathbf{s}$ locally from $0.2\,\mathrm{nm}^{-1}$ to zero.

WHAT IS WEAK?

The intensity of the reflection that we see in the DP is still small even though a relatively intense peak may occur in the image close to the defect core because the DP averages over a large area.

When you look at dislocations in the WB image, you see bright lines on a dark background. Let's compare the WB image to a BF image of the same defect in Figure 27.8. You can see that the WBDF image is much narrower; you could make the comparison look even better if you make $\mathbf{s}$ very close to zero in the BF image.

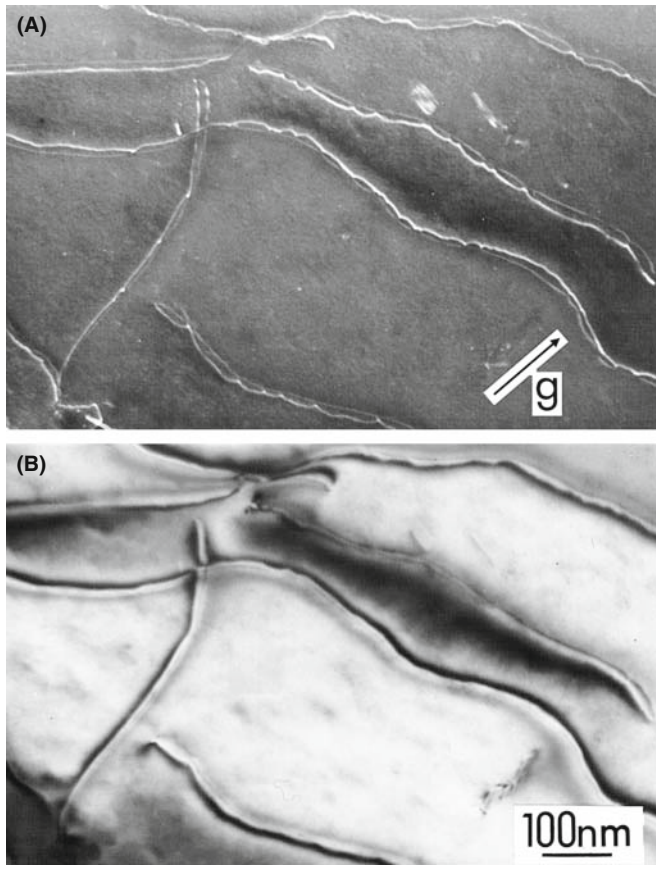


FIGURE 27.8. A comparison of dislocation images in a Cu alloy formed using (A) WB and (B) strong-beam ($\mathbf{s}_\mathbf{g} > 0$) conditions.

We'll keep our discussion of dislocations brief but draw your attention to a few particular points

- In the WB technique, most of the specimen is tilted so that **s** is large; the lattice planes in most of the specimen are then rotated away from the Bragg condition. However, as you can see in Figure 27.9, near the core of the dislocation the planes are locally bent back into the Bragg condition.
- This bending is only large close to the core of the dislocation (i.e., at the same depth from the surface).
- The peak you see in the WB image is always displaced to one side of the dislocation core. If you reverse the sign of **g**, the peak moves to the other side of the core. If you reverse the Burgers vector, **b** (rotate the diagram in Figure 27.9 through 180°), but keep **g** the same, the peak again moves to the other side.
- If you increase **s** in the crystal, then the planes must bend more to satisfy the Bragg condition, which means the observed peak will move closer to the dislocation core.

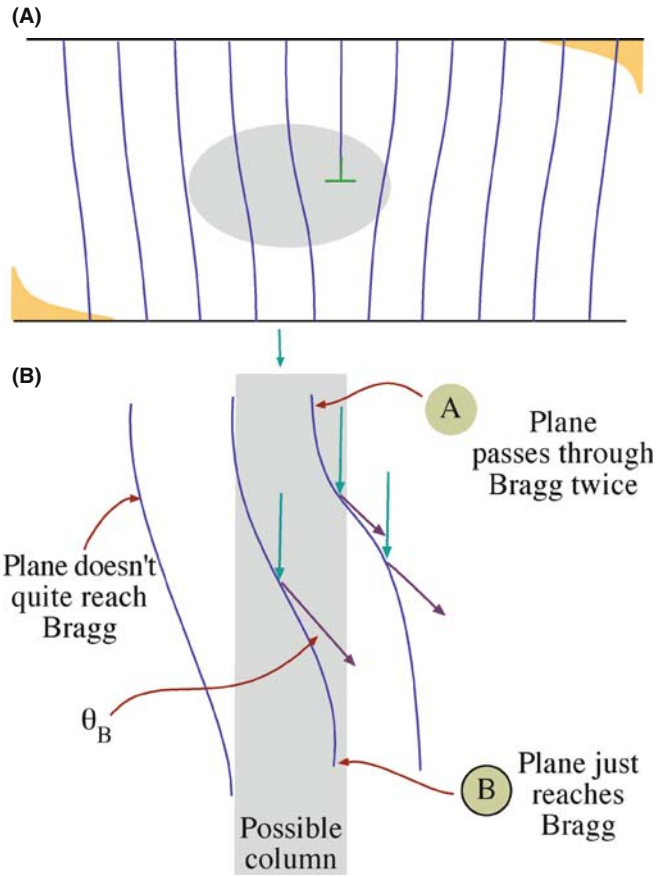


FIGURE 27.9. WB images from strained regions of the specimen show high intensity only from those regions where the diffracting planes are bent back into the Bragg condition. This illustration is for an edge dislocation. (B) shows the core region selected in (A).

- When we say 'position of the peak,' we are always talking about a projected position where the projection is along $\mathbf{k}_D$.
- There will be some situations where the local strain at the defect is not large enough to compensate for the **s** you have chosen. Then you will only see poor contrast in the image.

27.6 PREDICTING DISLOCATION PEAK POSITIONS

There are three ways to calculate the contrast in a WB image. Since each teaches us something new, we'll go through them in turn.

Method 1: The WB Criterion states that the largest value of $\phi_{\mathbf{g}}$ in the WB image occurs when $s_{\mathbf{R}}$, which we derived in equation 26.7, is zero. We can express this result as

$$s_{\mathbf{R}} = s_{\mathbf{g}} + \mathbf{g} \cdot \frac{d\mathbf{R}}{dz} = 0 \qquad (27.6)$$

Equation 27.6 tells us that if the effective value of s (i.e., s_{eff}) is zero, even though $s_{\mathbf{g}}$ is not zero, then the direct beam and the diffracted beam, **g**, are strongly coupled. In this situation the strain field effectively rotates the lattice planes into the Bragg-reflecting position. Therefore, the crystal can be oriented so that $\phi_{\mathbf{g}}$ is small for all columns except those near the dislocation core where it can attain a considerable magnitude due to the strong coupling with the transmitted beam as it passes through the region close to the core of the dislocation where s_{eff} is zero. This increased amplitude is then retained below the core when the coupling between the two beams is decreased again. The intensity is expected to be largest for that column where s_{eff} remains closest to zero over the longest length, and this occurs for the column where there is an inflection in the curve of **R** versus z. Therefore, the position of the WB peak should occur when equation 27.6 is satisfied at a turning point of $\mathbf{g} \cdot d\mathbf{R}/dz$.

Method 2: The Kinematical Integral. An alternative criterion for defining the position of the WB peak was derived with the approximation that only two beams are considered and **s** is sufficiently large. Cockayne showed that the maximum scattering from the transmitted to the diffracted beam occurs where the kinematical integral, defined as

$$\int_{\text{column}} e^{\{-2\pi i(s_{\mathbf{g}} z + \mathbf{g}.\mathbf{R})\}} dz \qquad (27.7)$$

is maximized. This maximum, in general, occurs for a column which is closer to the dislocation core than predicted by *Method 1*. The reason for this difference

is interesting: because the planes are bent, the reciprocal lattice point is, on average, nearer to the Ewald sphere. Therefore, the integral has a larger value over the length of the column.

Without doing all the math, we can illustrate how these two approaches are related. What we want to do is determine when $\phi_{\mathbf{g}}$ is large, but still kinematical (i.e., $\mathbf{s}$ is large); we want to maximize the kinematical integral in equation 27.7. We can do this using the stationary-phase method described by Stobbs. We write the integral as

$$\int_0^t \exp\left(-2\pi i\left[\frac{z^2}{2}\cdot\frac{d^2}{dz^2}(\mathbf{g}\,.\,\mathbf{R}) - \frac{z^3}{3}\cdot\frac{d^3}{dz^3}(\mathbf{g}\,.\,\mathbf{R})\right]\right)dz \quad (27.8)$$

where we have set $s + d/dz(\mathbf{g}.\mathbf{R}) = 0$. If we also set $d^2/dz^2(\mathbf{g}.\mathbf{R}) = 0$ (at the inflection), we go further to ensuring that the term in the square brackets is zero. This condition is what we guessed for the first method of defining the WB criterion.

Method 3: Compute the Contrast. Now that personal computers are widely available, we can calculate the position of the WB peak and graph the results. What we then find is that the WB peak actually lies between the two values predicted by the two criteria deduced using the first two methods. We also find, using the computer, that the position and width of the image peak are affected by any strongly excited diffracted beams so these *must* be avoided. Some apparently weak beams can also affect the image. A practical point is that the computer sometimes gives a rather pessimistic view of the variability in the peak position, so we have to weight the results carefully. Remember that the important advantages of this approach are that we can include the effects of the other diffracted beams which are always present, and we can take account of other effects, such as the convergence of the beam.

In the kinematical approximation, the half-width, Δx, of the image of an undissociated screw dislocation with $|\mathbf{g}.\mathbf{b}| = 2$ is given approximately by the relation derived by Hirsch et al. (1960)

$$\Delta x = \frac{1}{\pi s_{\mathbf{g}}}\cdot\frac{\xi_{\text{eff}}}{3} \quad (27.9)$$

This expression is a very useful rule of thumb. You will realize that this WB image width is special for three reasons, which arise because it doesn't depend on $\xi_{\mathbf{g}}$. So, once $\mathbf{s}$ is fixed in a WB image, we can make several surprising statements regarding the width of the dislocation peak

- It does not depend on the material.
- It does not depend on the reflection.
- It does not depend on the kV.

Take the example of the 220 reflection in Cu at 100 kV: $\xi_{\mathbf{g}}$ is 42 nm, and the width Δx is 14 nm. So, even if equation 27.9 is slightly wrong, the image width is greatly reduced in WB. If we make $s_{\mathbf{g}} = 0.2\,\text{nm}^{-1}$, then the half-width is 1.7 nm.

Computed many-beam images confirm that dislocations in other orientations give rise to similar narrow peaks when this value of $s_{\mathbf{g}}$ is used. A series of peak profiles for different values of t is shown in Figure 27.10. Notice that although the intensity of the peak may be only about 0.1% of the incident beam, it is still much higher than the background.

THE WB 'CRITERION'

$s_{\mathbf{g}} = 0.2\,\text{nm}^{-1}$ is always a useful guide; it satisfies the requirements that the image should have a narrow width and show a high contrast between the defect and background regions.

Equation 27.9 indicates that as we increase the value of s, the half-width of the image peak decreases. However, a maximum is imposed on s by the fact that the intensity of the diffracted beam varies as s^{-2}. If we make s much larger, the contrast of the image therefore becomes too small to be of practical use.

The basic requirements governing the value of s which you must use for quantitative imaging are

- $s \geq 2 \times 10^{-2}\,\text{Å}^{-1}$ to give sufficiently narrow peaks for fine detail to be studied.
- $s \leq 3 \times 10^{-2}\,\text{Å}^{-1}$ because the intensity varies as s^{-2} in the kinematical limit.
- $s\xi_{\mathbf{g}} \geq 5$ to give sufficient contrast in the WB image.

If you use the $\mathbf{g}(3\mathbf{g})$ condition for Cu with $\mathbf{g} = 220$ and 100-keV electrons, then the value of $\mathbf{s}_{\mathbf{g}}$ will be $0.238\,\text{nm}^{-1}$.

27.7 PHASOR DIAGRAMS

We sometimes find it useful to demonstrate the depth dependence of the contrast in the WB image using phasor, or amplitude-phase, diagrams which we introduced in Chapter 2. You can generally use such diagrams whenever the kinematical approximation holds; they are equivalent to a graphical integration of the two-beam equations for the case where $\mathbf{s}$ is large. In fact, many of the early calculations of defect contrast were made using this approach before computers became widely available. We recommend that you glance at the original paper by Hirsch et al. (1960): the contrast profiles were calculated by plotting the phasor diagram.

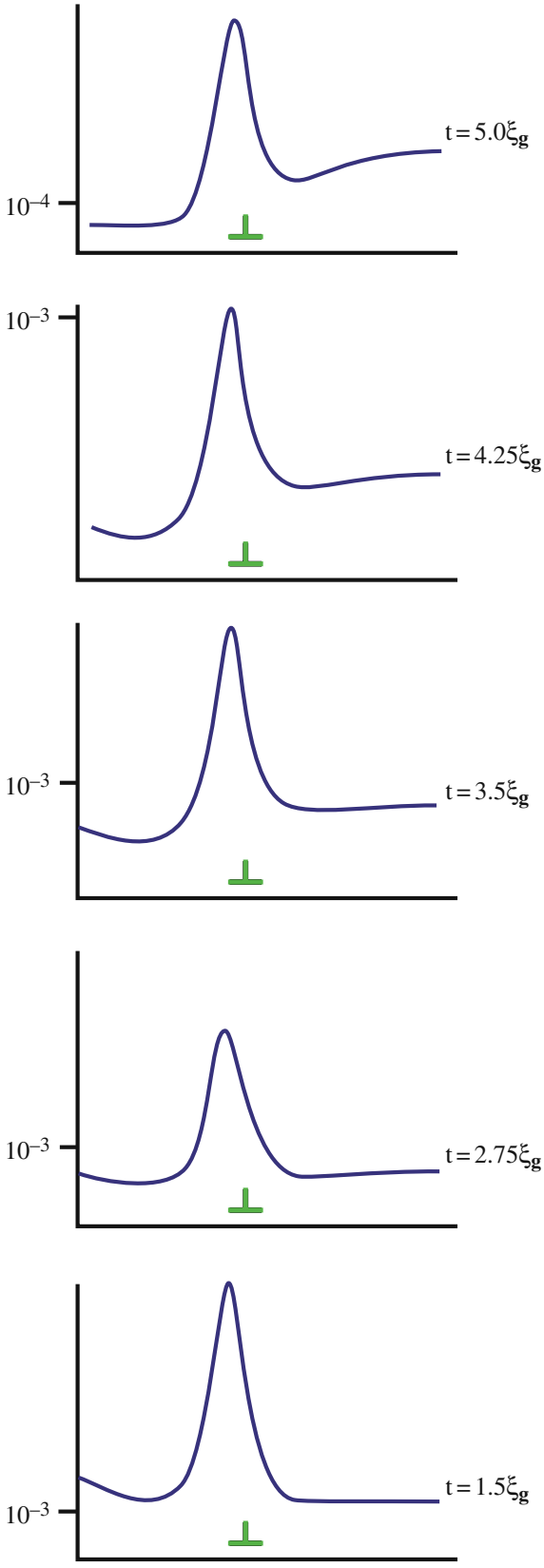


FIGURE 27.10. Examples of computer-calculated intensity peaks in WB images of an edge dislocation in Cu for different values of t. The intensity is relative to the unit incident-beam intensity. Note that the dislocation position and the peak intensity never coincide.

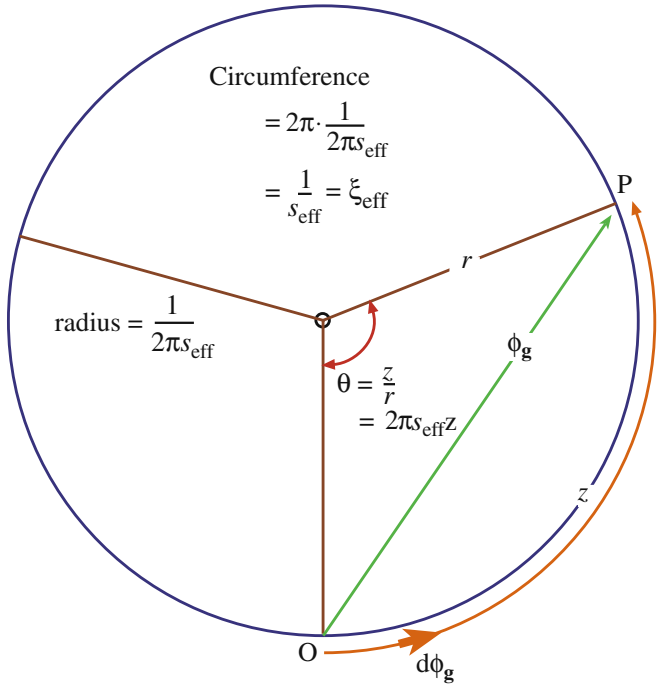


FIGURE 27.11. A phasor diagram for the WB case. The distance z is the arc OP measured around the circumference and the radius of the circle is $(2\pi s_{eff})^{-1}$. An increase in θ of 2π means the thickness increased by the ξ_{eff}. The amplitude in the WB is ϕ_g.

The basic idea is shown in Figure 27.11. We simply add all the $d\phi_g$ increments to ϕ_g. In doing so, we take account of the phase changes which occur as the beam passes through the crystal. Remember that in this approximation, no electrons leave the **g** beam! If the crystal is perfect and our increments are sufficiently small, we will produce a smooth circle.

The circumference of this circle is ξ_{eff} as we require for the depth periodicity, and the radius is $\xi_{eff}/2\pi$ or $(2\pi s_{eff})^{-1}$. Notice that as we increase **s**, we decrease ξ_{eff} and the circle becomes smaller. Thus we move around the circle more quickly if **s** is large. In other words, our effective extinction distance is reduced, as we knew from Chapter 13.

If the diffracted beam passes through a stacking fault, it will experience an extra phase shift given by $2\pi \mathbf{g} \cdot \mathbf{R}$. Using the familiar example for an fcc crystal, we take an example where $\mathbf{R} = \frac{1}{3}[11\bar{1}]$ and $\mathbf{g} = (20\bar{2})$ which gives $\alpha = 2\pi/3 = 120°$ (modulo 2π). The abrupt phase change is shown at P_3, in Figure 27.12. Now, the value of ϕ_g (P_1P_2) can be much larger than in a perfect crystal. The locus of ϕ_g still travels around the first circle until it meets the planar fault at $z = n\xi_{eff} + t_1$, where z and t_1 are measured parallel to $\mathbf{k}_D$. It then moves onto the second circle until it reaches $z = t$. You can readily see that if we keep the depth of the fault, t_1, fixed, we then see depth fringes which vary with periodicity ξ_{eff} as we vary the total thickness, t. (The value of α is still 120°.) The situation is a little

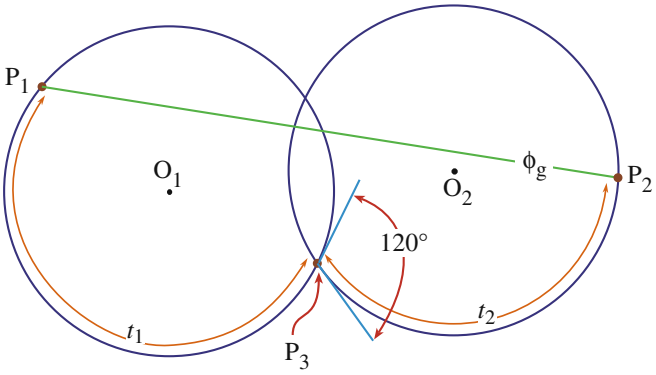


FIGURE 27.12. A phasor diagram used to explain the WB contrast from a SF at depth t_1 in a foil of thickness $t_1 + t_2$. P_1 is the top of the foil (on a circle centered on O_1); P_2 is the bottom of the foil (on a circle centered on O_2). Here, the phase change at the SF (P_3) is 120°. The total amplitude in the WB is ϕ_g.

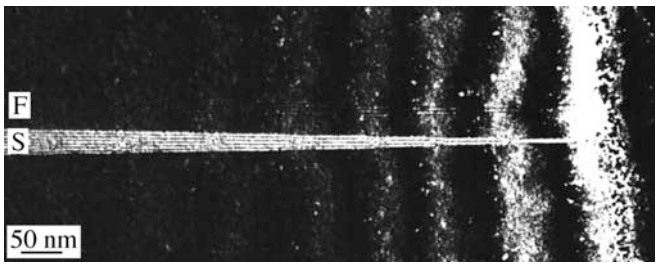


FIGURE 27.13. Illustration of thickness fringes in an experimental image of two inclined translation boundaries (S and F) in a wedge specimen of spinel. The thickness fringes from the wedge and from the inclined boundaries can be clearly seen and counted. Boundary 'F' shows very little contrast because it has a small **R** value.

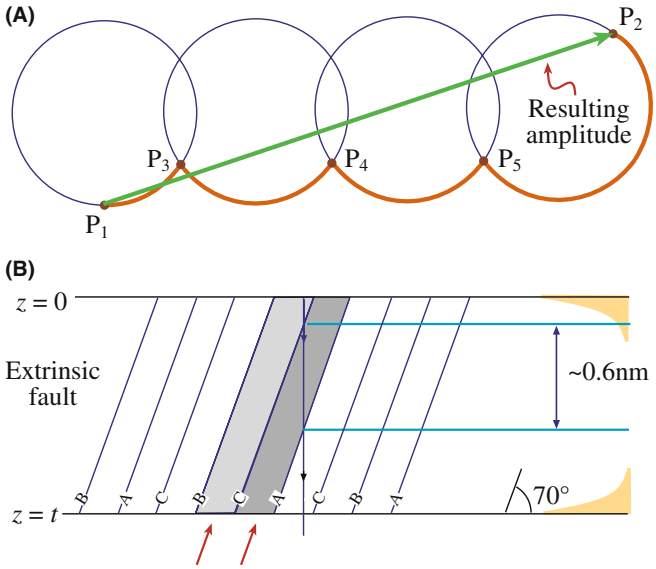


FIGURE 27.14. Phasor diagram (A) for WB encountering a series of overlapping planar defects (B) at P_3, P_4, and P_5. For a 111 foil normal, the inclined $11\bar{1}$ planes lie at 70° to the surface so that the spacing between adjacent planes in the direction of the beam is 0.627 nm. Compare with Figure 27.12.

more difficult to envision if we keep t fixed but vary t_1, but the principle is the same. Images of WB fringes from inclined SFs are illustrated in Figure 27.13. The thickness fringes of the wedge specimen and those from the inclined stacking fault can both be clearly seen and counted. Notice that the number of bright fringes on the planar defect really does increase by one for every increment ξ_{eff} in the thickness of the wedge and you don't need to know **R**.

You can imagine applying this analysis to the situation where you have several overlapping planar faults, as shown in Figure 27.14. ϕ_g (P_1P_2) can become very large. (Does our approximation of unit incident intensity still hold?) This situation does occur in practice, as shown in Figure 27.15. Here we have several overlapping faults. The bright ones are bright even when ϕ_g reaches its minimum, as you would expect from Figure 27.14. If you compare the WB image with its strong-beam counterpart in Figure 25.5, you will notice that there is much more detail in the WB image; as we saw in Chapter 25, overlapping faults on nearly adjacent planes can essentially give no contrast in the BF image. You can easily check such effects in WB by adjusting **s** as in Figure 27.15C–E. It's interesting to realize that this effect can occur even when two intrinsic SFs lie on adjacent planes to give the extrinsic SF. This approach can be used to image other planar defects such as the {112} twin boundary which can have a significant thickness. The key factor is that, under WB conditions, ξ_{eff} can become comparable to the distance that the beam travels between encountering successive planes of atoms, particularly when the interface is quite steeply inclined relative to the beam.

We can also use a phasor diagram to describe the contrast from a dislocation, but now the phase change occurs over a length of the column rather than at a particular depth in the sample. As illustrated in Figure 27.16, the phase can either add or subtract depending on the sign of $\mathbf{g}.\mathbf{R}$. When the phase changes quickly with a change in t, as at the center of Figure 27.16B, it means that we are strongly coupling the incident and diffracted beams.

We'll summarize our discussion of phasor diagrams by two points

- They should only be used when the kinematical approximation holds.
- They then give us a graphical method for understanding the variation of ϕ_g with thickness; the method is especially useful when crystal defects are present.

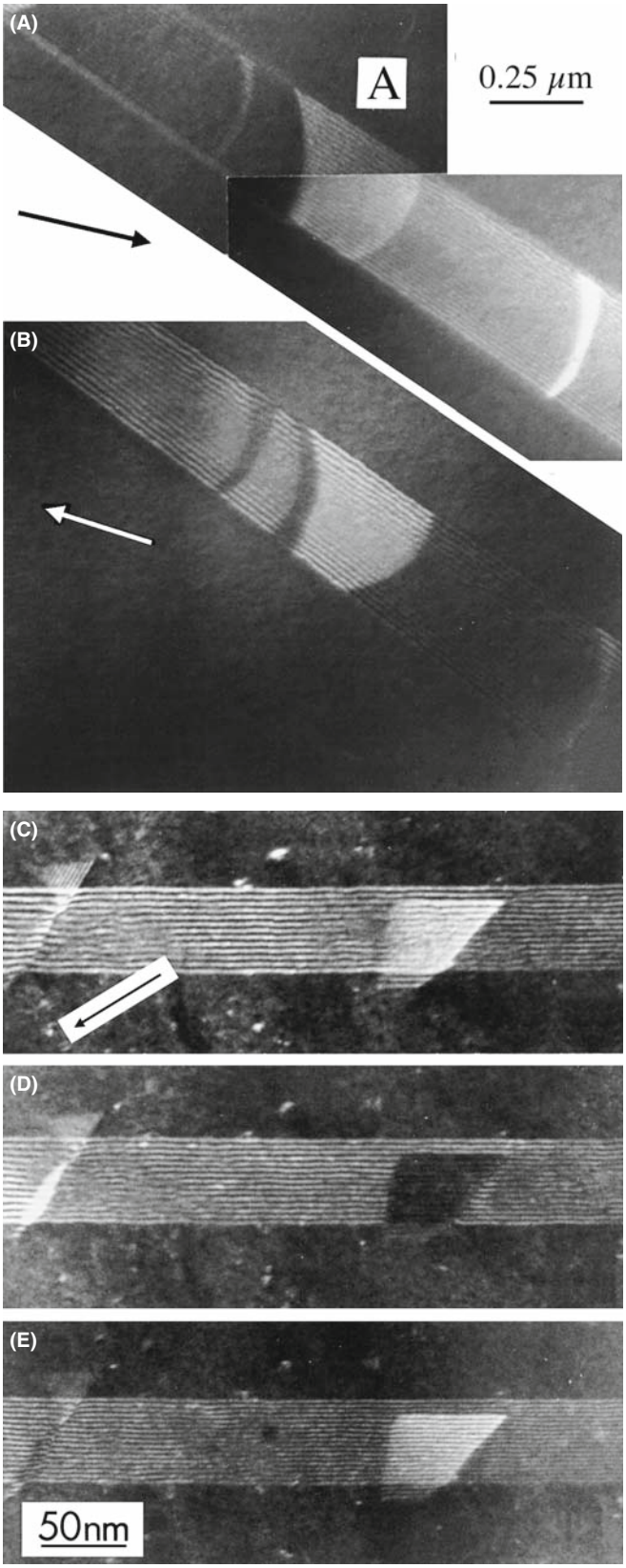


FIGURE 27.15. (A, B) Overlapping SFs imaged in WBDF using two different conditions (±**g**). Many more fringes occur than in the corresponding BF image in Figure 24.5. Fringes are present at A where the BF image showed no contrast. (C–E) Changes in fringe spacing and intensity when one SF overlaps another; **s** increases from C to E. (Note the decrease in the fringe spacing.)

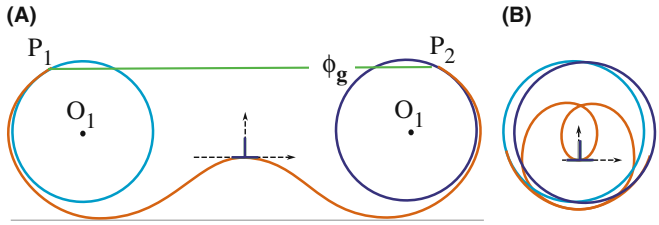


FIGURE 27.16. Phasor diagrams for a dislocation for ±**g**. The phase change is not abrupt but instead occurs over an extended distance along the column. (A) The phase change causes an increase in the amplitude scattered: ϕ_g is larger than in the perfect crystal (compare to Figure 27.12). (B) When **g** is reversed the phase change has the opposite sign and the resulting ϕ_g is much smaller (so it's not shown in this confusing figure).

27.8 WEAK-BEAM IMAGES OF DISSOCIATED DISLOCATIONS

Although the study of dislocations is a very specialized topic, it beautifully illustrates the potential of the WB technique. Dissociated dislocations are common in face-centered-cubic (fcc) materials (including Si) and ordered intermetallics such as Ni_3Al. The geometry of a dissociated dislocation in Cu is summarized in Figure 27.17. We gave some general references on the theory of dislocations in Chapter 26.

Since the computed many-beam images show that the position of the dislocation image lies close to the position predicted by the WB criterion, this criterion is used in practice because it allows us to deduce an equation for the position of the image. We can thus directly relate the separation of Shockley partials, for example, to the measured separation of the two peaks observed in $|\mathbf{g}.\mathbf{b}_T| = 2$ images of dissociated dislocations. Hence we can estimate the stacking-fault energy (SFE) of

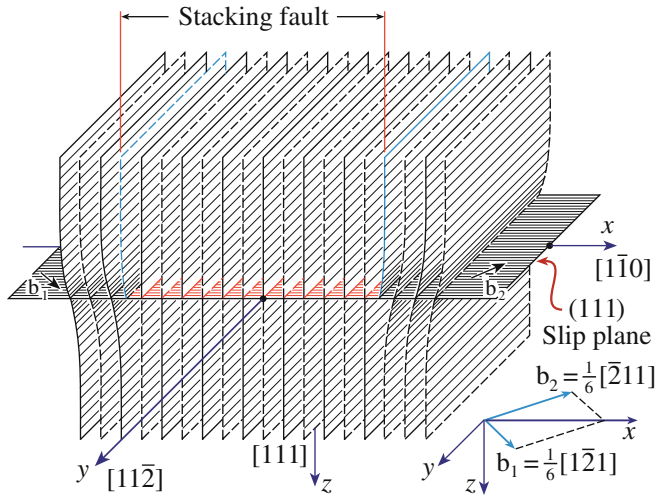


FIGURE 27.17. The geometry of a perfect dislocation in Cu. The perfect dislocation separates into two Shockley partial dislocations with Burgers vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ separated by a SF on a {111} plane. Although it might appear that partial dislocations are edge type, they are actually 60° dislocations (shown in the figure) with an edge component.

semiconductors and several fcc metals. In order to interpret WB images of extended dislocation configurations, we may need to know how the position of the image peak is related to the position of the dislocation core. This information is essential whenever we use the WB technique to collect *quantitative* data. We will now answer two questions for quantitative WB analysis

- What factors determine the position of the WB peak?
- What methods are used to determine the Burgers vectors of the dislocations?

When the WB conditions are satisfied, dissociated dislocations can be imaged in a {220} reflection for which $|\mathbf{g}.\mathbf{b}_T| = 2$, so that each partial dislocation gives rise to an intensity peak having a half-width of 1–1.5 nm. The separation of the partials can be deduced to within ±0.7 nm providing that the peak separation is greater than ~2.5 nm. For many materials, we need to use anisotropic elasticity theory to relate the atomic displacements to the image, so the computer becomes essential.

Why not just use HRTEM since, as we'll see in Chapter 28, HRTEM can give detail down to below 0.2 nm, whereas WB is often limited to ~1 nm? If you want to interpret an HRTEM image the defect must be *absolutely straight*, parallel to the beam, and located in a very thin region of the specimen. The segment of dislocation studied by HRTEM will thus be no longer than 20 nm and less than 10 nm for the highest resolution. In a WB image the defect can be micrometers long and, in a relatively thick foil, it can even change direction. If you look back at Figure 27.8 you'll see pairs of lines which correspond to partial dislocations in the WB image. You can see other features such as constrictions. The strong-beam image may also show two or more lines for a particular dislocation but as we saw in Chapter 26, these lines are generally not related to the detailed structure of the dislocation but rather to n in the equation $\mathbf{g}.\mathbf{b} = n$.

Ideally, for quantitative analysis, you should choose long, nearly straight dislocations. As shown in Figure 27.17, the two Shockley partial dislocations lie in the (111) plane of the foil. The Burgers vectors of the total and partial dislocations can be determined by imaging the dislocation in the WB mode using the $\{2\bar{2}0\}$ reflections. A sharp peak is found for a partial dislocation with $|\mathbf{g}.\mathbf{b}_p| = 1$, and either no peak or a diffuse one (arising from the anisotropy of the lattice) if the partial dislocation has $|\mathbf{g}.\mathbf{b}_p| = 0$. When $|\mathbf{g}.\mathbf{b}_T| = 2$, the diffraction vector $\mathbf{g}$ and the Burgers vector $\mathbf{b}_T$ are parallel and, in the image of a dissociated dislocation, two sharp peaks are formed, one corresponding to each of the partial dislocations (both now have $|\mathbf{g}.\mathbf{b}_p| = 1$).

We can explain this difference in intensity if one peak (the weaker) arises from the region between the partial dislocations and the other from outside the dissociated dislocation. This effect cannot occur in $|\mathbf{g}.\mathbf{b}_T| = 1$ images, when $|\mathbf{g}.\mathbf{b}_p| = 1$ for one partial and $|\mathbf{g}.\mathbf{b}_p| = 0$ for the other, and it can be used to identify the reflection for which $|\mathbf{g}.\mathbf{b}| = 2$. Confirmation of the Burgers vector is always obtained using the BF mode, observing characteristic $|\mathbf{g}.\mathbf{b}_T| = 2$ or $|\mathbf{g}.\mathbf{b}_T| = 0$ images.

DISSOCIATED DISLOCATIONS

One of the peaks is on average more intense than the other; the order reverses when $\bar{\mathbf{g}}(3\bar{\mathbf{g}})$ is used instead of $\mathbf{g}(3\mathbf{g})$.

In a WB image with $|\mathbf{g}.\mathbf{b}_T| = 2$, each of the partial dislocations will generally give rise to a single peak in the image which is close to the dislocation core. You can calculate the approximate positions of these peaks using the criterion from equation 27.6. Then, you can relate the separation of the peaks in the image to the separation Δ of the partial dislocations.

We can write the displacement, using isotropic elasticity theory, as the sum of the displacements due to the individual partial dislocations. If the Burgers vector of a straight, mixed dislocation lies in the (111) plane parallel to the surface of a foil, then at a distance x from the dislocation core

$$-s_g = \frac{|\mathbf{g}|}{2\pi}\left[\left(|\mathbf{b}_1| + \frac{|\mathbf{b}_{1e}|}{2(1-\nu)}\right)\frac{1}{x} + \left(|\mathbf{b}_2| + \frac{|\mathbf{b}_{2e}|}{2(1-\nu)}\right)\frac{1}{x-\Delta}\right] \qquad (27.10)$$

Here x defines an axis perpendicular to both the dislocation line and the beam direction, and e refers to the edge component of the Burgers vectors of the partial dislocations, 1 and 2. This relation is particularly simple because, for the geometry we have chosen, the term $\mathbf{g}.\mathbf{b} \times \mathbf{u}$ is zero. Using the notation

$$a = -s_g\left[\frac{|\mathbf{g}|}{2\pi}\cdot\left(|\mathbf{b}_1| + \frac{|\mathbf{b}_{1e}|}{2(1-\nu)}\right)\right]^{-1} \qquad (27.11)$$

and

$$b = -s_g\left[\frac{|\mathbf{g}|}{2\pi}\cdot\left(|\mathbf{b}_2| + \frac{|\mathbf{b}_{2e}|}{2(1-\nu)}\right)\right]^{-1} \qquad (27.12)$$

equation 27.10 reduces to

$$1 = \frac{1}{ax} + \frac{1}{b(x-\Delta)} \qquad (27.13)$$

which has two solutions, x_+ and x_-, given by

$$x_{\pm} = \frac{ab\Delta + a + b \pm \left[(ab\Delta + a + b)^2 - 4a^2b^2\Delta\right]^{1/2}}{2ab} \quad (27.14)$$

These values of x define the positions of the peaks in the image. The separation between these peaks is then given by

$$\Delta_{obs} = \left[\Delta^2 + \frac{(a+b)^2}{a^2b^2} + \frac{2(a+b)\Delta}{ab} - \frac{4\Delta}{a}\right]^{1/2} \quad (27.15)$$

We can rearrange this equation to make it appear more symmetric in a and b. Of course, it will not be symmetric because the peak is always located on one side of the dislocation.

$$\Delta_{obs} = \left[\left(\Delta + \frac{1}{b} - \frac{1}{a}\right)^2 + \frac{4}{ab}\right]^{1/2} \quad (27.16)$$

Computed images confirm that this relation is accurate for $\Delta_{obs} > 2.5$ nm, to within ± 0.7 nm. This uncertainty is due to the variation of the peak position with the depth of the dislocation in the foil and the foil thickness. A small uncertainty arises when you have not determined the actual direction of $\mathbf{b}_T$, i.e., whether it is in the direction of $\mathbf{g}$ or $\bar{\mathbf{g}}$. Stobbs and Sworn found that the relation (equation 27.16) subject to the ± 0.7 nm uncertainty is still a good approximation when anisotropic elasticity theory is used.

As a simple exercise, consider the WB images of a dissociated screw dislocation and a dissociated edge dislocation. You should pay particular attention to 'a' and 'b' because in one case, $\mathbf{b}_{1e}$ and $\mathbf{b}_{2e}$ have the same sign, while in the other the sign is opposite. Does the image always have the same width when you reverse $\mathbf{g}$?

Example 1. Even if you never want to calculate the actual separation of two dislocations from observations of two peaks, you can learn new ideas about dislocations from such images. Figure 27.18 is a famous set of images showing a dislocation in Si which is constricted along part of its length and dissociated along the rest. Even if you don't know the precise details of the dislocation structure, you know that it can adopt two variants; the rest of the task is modeling the defect.

Example 2. The WB image of the node pair in Figure 27.19A, tells you very quickly that the two nodes are different; if we form images using other $\mathbf{g}$ vectors (Figure 27.19B–D), we find that one of the partial dislocations is out of contrast in the image. The extended node contains the same type of intrinsic stacking fault that is present in the dissociated dislocation; $\mathbf{g}.\mathbf{R}$ is zero in this image for the stacking fault. The other node is constricted,

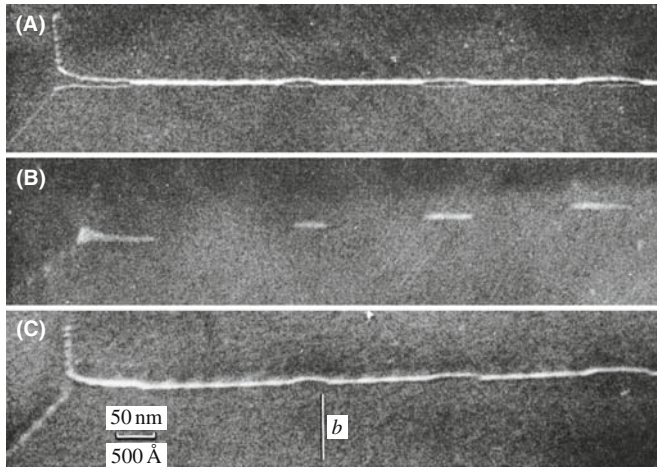


FIGURE 27.18. WB image of a dislocation in Si which has both dissociated and constricted segments: (A) $\mathbf{g}.\mathbf{b} = 2$; both partial dislocations are visible. (B) $\mathbf{g}.\mathbf{b}_T = 0$ showing SF contrast. (Note: no fringes.) (C) $\mathbf{g}.\mathbf{b} = 1$; only one partial dislocation is visible.

within the detection limit of the WB technique. Comparison with the BF images in Figure 26.8 is instructive. Note that in Figure 27.19A the SF in the node does not show fringes because the SF plane is parallel to the surface of the specimen; we just tilted the specimen.

Example 3. We mentioned that the peak moves to the other side of the dislocation if we reverse $\mathbf{b}$. This is exactly what happens for a dislocation

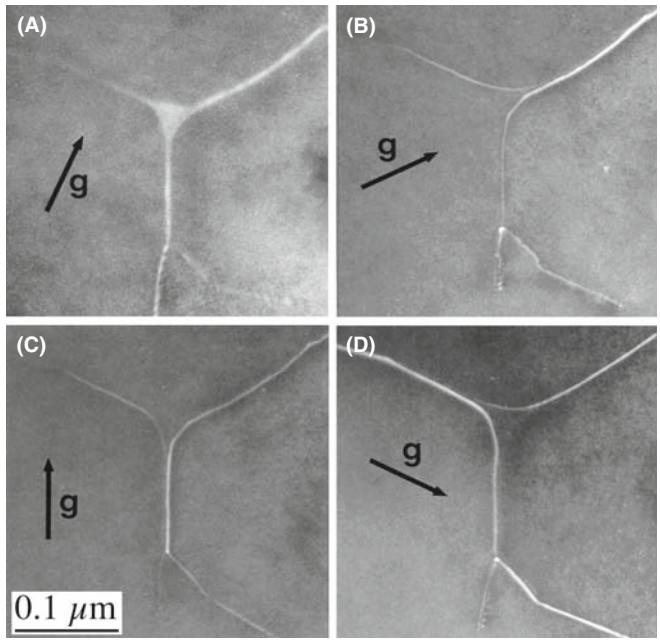


FIGURE 27.19. WB images of a pair of dislocation nodes formed by the dissociation of interacting dislocations lying on a {111} plane in a Cu alloy. The SF is imaged in (A), and imaging with different reflections (B–D) the partial dislocations are out of contrast when $\mathbf{g}.\mathbf{b} = 0$. Compare with the BF image in Figure 26.8 and note the difference in magnification.

dipole as you can see in Figure 27.20. This is a complicated figure except that you can use it not only to see the inside-outside contrast in WB images of dislocation dipoles but also as an exercise in $\mathbf{g}.\mathbf{b}$ analysis. A dislocation dipole is a pair of dislocations identical in every way, apart from the sign of the Burgers vector. If we now reverse $\mathbf{g}$, then both peaks move to the other side of their respective dislocations. This change in contrast is referred to as inside-outside contrast and is commonly seen on dislocation loops, which are themselves closely related to these dipoles. Loops are just more 'equiaxed' than dipoles and always consist of one dislocation (which may be dissociated). The images shown in Figure 27.20B and C illustrate the dramatic change in contrast which you can see on reversing $\mathbf{g}$. Some of the dipoles completely disappear in Figure 27.20A–C because they are a special form of defect known as a faulted dipole. Such dipoles usually give very low contrast in strong-beam BF images because the dislocations are always very close together so that their strain fields overlap and cancel, and the lattice is thus only distorted over very small distances. All the dislocations are in contrast in Figure 27.20B: it's the $\mathbf{g}.\mathbf{b} = 2$ image. When we use the WB techniques, we are probing the structure on these very small dimensions and the contrast can be high. Again compare with the BF image in Figure 26.13.

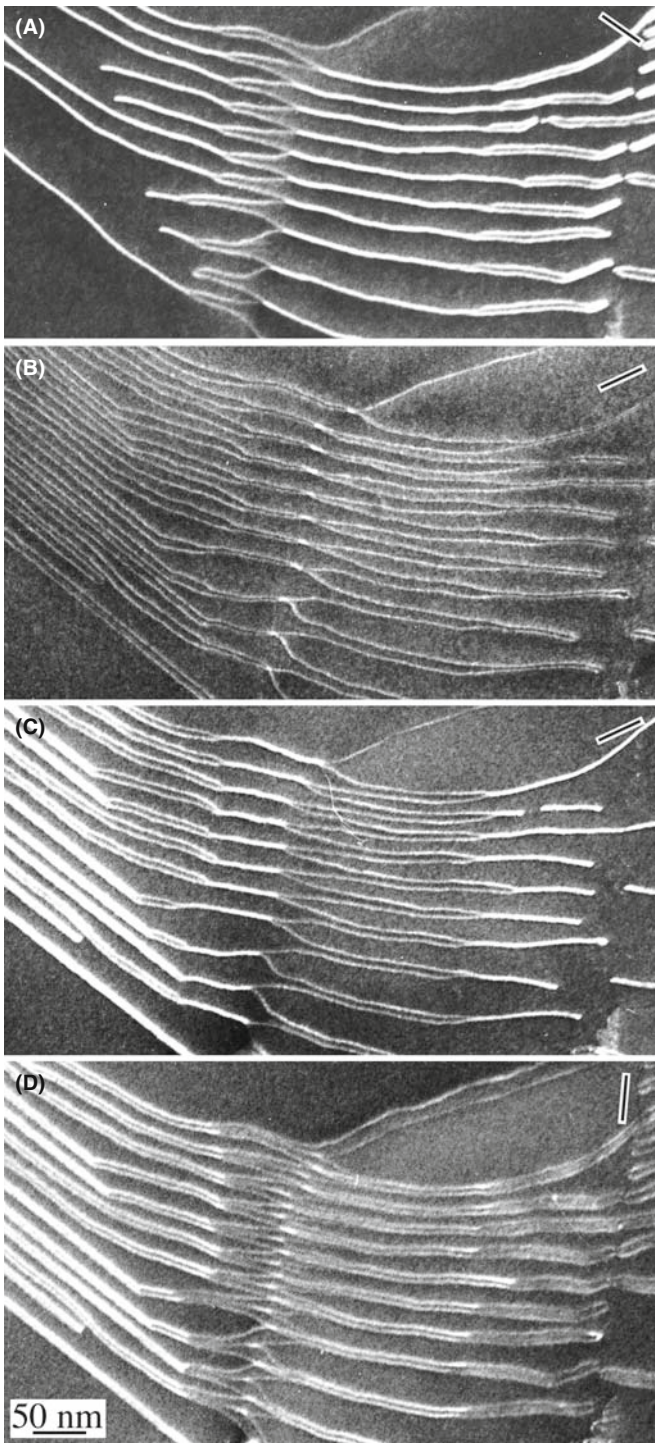


FIGURE 27.20. Four WB images showing an array of dislocation dipoles in a Cu alloy having a low stacking-fault energy. The reflections are all 220-type and the dislocations all lie on (111) planes which are nearly parallel to the surface of the specimen. All the dislocations are dissociated. All the Shockley partial dislocations are in contrast in (D) while half are out of contrast in (A–C). Notice that the narrower images are brighter than the wider ones; the strain is large in between the dislocations but decreases rapidly outside the dipole since the total Burgers vector of a dipole is zero.

Examples 4 and 5. The WB technique allows us to see features which would be hidden if we used strong-beam imaging. Figure 27.21 gives a simple example where an inclined SF is cutting through several dissociated dislocations. The interaction of the two defects would be masked by the SF fringes in the strong-beam image but is clearly visible in the WB image. Seeing small particles close to dislocations is difficult in strong-beam imaging. Although not easy in WB, Figure 27.22 does illustrate that it can be done. This image shows, for example, that the behavior of the dissociated dislocation is different on each side of the particle.

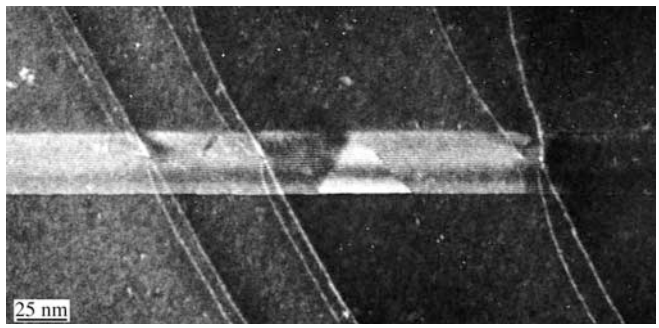


FIGURE 27.21. WB image of an inclined SF cutting though a series of dissociated dislocations lying parallel to the surface of the Cu-alloy specimen.

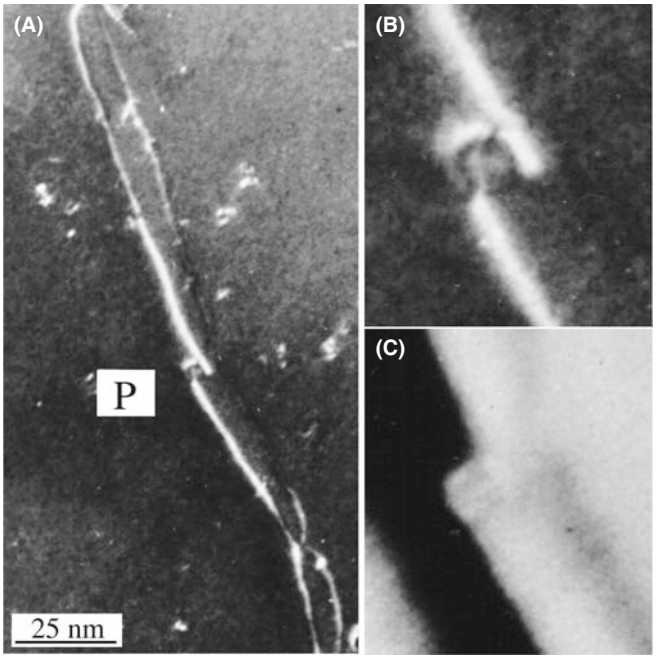


FIGURE 27.22. (A) WB image of a dissociated dislocation interacting with a particle (P) in a Cu alloy. (B) and (C) are enlargements of the WB and corresponding BF images, respectively. Note the difference in detail.

FIGURE 27.23. WB image showing the dissociation of a group of dislocations which are inclined to the foil surface and give wedge-shaped SFs. The shape of the SFs is caused by surface stresses.

Figure 27.22B and C compares magnified WB and BF images from the region close to the particle, and again stresses the advantage of WB over BF imaging.

Example 6. We noted in Section 26.8 that the surface of the specimen can affect the geometry of the defects we are examining. In general, the specimen needs to be thinner for WB imaging than for strong-beam imaging. Therefore, the surface can affect not only the appearance of the image but also the actual geometry of the defect. Figure 27.23 shows an example where this effect is particularly clear. Dislocations that were uniformly separated in the bulk material now appear wedge-shaped: the effects of the two surfaces are different in this case.

27.9 OTHER THOUGHTS

27.9.A Thinking of Weak-Beam Diffraction as a Coupled Pendulum

We can illustrate the principle which underlies the increase in intensity in the WB image close to a dislocation using the mechanical analog of a coupled pendulum. A diagram is shown in Figure 27.24. The two pendula are connected (coupled) by a third string. If we start the left pendulum swinging but hold the connecting string, the right-hand pendulum remains stationary. Now release the connecting string. You will see that the right-hand pendulum now begins to swing. If we let the process continue, eventually the right-hand pendulum is swinging as much as the original one did, but the original one is stationary: this is the strong-beam analog! All the kinetic energy has been transferred from one pendulum to the other. Given more time, the pendulum will return to the original condition. Now repeat the exercise but hold the connecting string again after the right-hand pendulum has begun to swing; you will notice that both pendula continue to swing, each with a constant amplitude. The role of the connecting string is to couple the two pendula (beams) so that we transfer energy from one beam to the other. In WB TEM, the defect acts as the connecting string. The two beams are only coupled over a short length as they travel past the defect. We can plot this amplitude (or intensity); try this as an exercise.

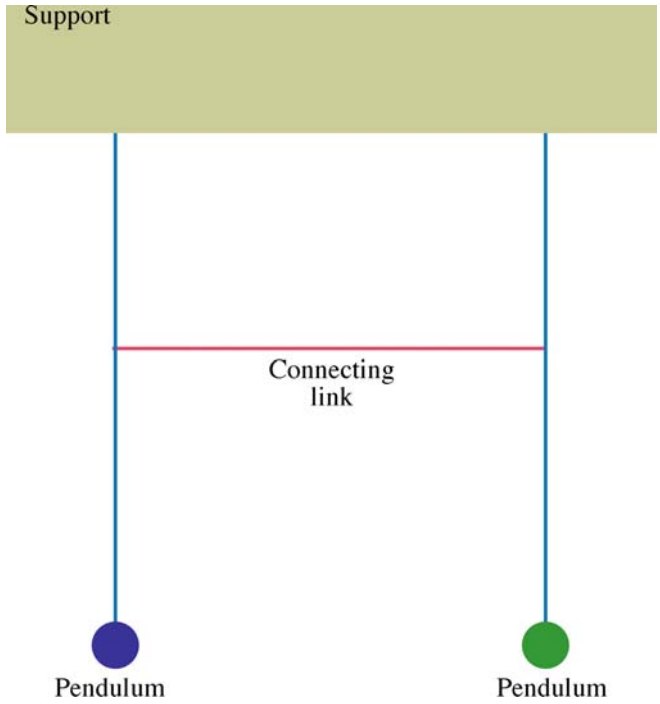


FIGURE 27.24. The coupled pendulum: a thought experiment.

27.9.B Bloch Waves

We discussed Bloch waves in Chapter 14. The difficulty in applying Bloch-wave analysis to the WB situation is that we are now usually interested in defects while Bloch waves are properties of the perfect crystal. However, we can offer some basic comments. For the reflection **g** to give a WB image, $|\phi_{\mathbf{g}}|$ must be much smaller than unity in the regions of perfect crystal but, in strained regions, a change $\Delta\psi^{(j)}$ in the amplitude of the Bloch wave j can give rise to a change $\Delta\phi_{\mathbf{g}}$. Cockayne has shown that the appreciable contrast which can then be present in the WB image is due, in the two-beam approximation, to the interband scattering from Bloch wave 1 to Bloch wave 2. In the general case, the scattering is from the branch of the dispersion surface with the largest $\psi^{(j)}$ to those branches with the largest $C_{\mathbf{g}}^{(j)}$, i.e., from the Bloch wave with the largest amplitude to the one which is most strongly excited. The dispersion surface for the **g**(3**g**) diffraction geometry is shown in Figure 27.25. It's an instructive exercise to reread this paragraph thinking how each statement relates to this figure and to consider other diffraction geometries, e.g., **0**(2**g**).

The Bloch-wave analysis of the problem leads to two further points which simplify the interpretation of WB images

- The diffraction conditions should be such that only one interband scattering process is important.
- In the two-beam approximation, in order for the image peaks to show sufficient contrast, it is generally found that w ($= s\xi_{\mathbf{g}}$) is greater than ~5.

You can satisfy the first requirement by ensuring that no reflections are strongly excited. The second condition is usually already satisfied because of the more stringent requirement that s should be greater than $0.2\,\mathrm{nm}^{-1}$. For example, for a $\{2\bar{2}0\}$ reflection in copper with 100-keV electrons, w is automatically greater than 8 since $\xi_{\mathbf{g}}$ is 42 nm.

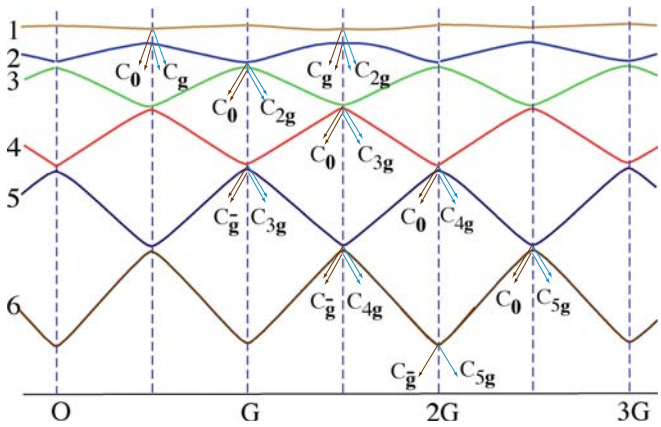


FIGURE 27.25. Dispersion-surface construction which is used to discuss the **g**(3**g**) geometry. The BZB is then at 1.5G and tells you which reflections are strongly coupled; see Figure 15.9.

The **g**(3.1**g**) condition may also be preferable to **g**(–1.1**g**) on theoretical grounds if images are to be compared with computed profiles made using the column approximation; i.e., it simplifies your interpretation. The basis for this suggestion is that the region of the dispersion surface from which the scattering occurs is flatter for the **g**(3.1**g**) diffraction geometry than for the **g**($\bar{\mathbf{g}}$) case.

27.9.C If Other Reflections are Present

Several times in the previous discussion, we have said that no reflections should be strongly excited. Much of our thinking has been based on the two-beam approximation we introduced in Chapter 13. When you are using WB conditions, you must be even more careful. Consider the **g**,3**g** geometry shown in Figure 27.5. We form the WB image using reflection **g** so electrons are weakly scattered from the O beam into the G beam. However, once in the G beam, they can be strongly scattered into the 2G beam. We can picture this process by drawing the new Ewald sphere for the 'new' incident beam, G; this sphere passes through 2G!

For the mathematically inclined, you can go back to the many-beam equations, which we introduced briefly in Chapter 14. The coupling of beams **g** and **h** is determined by ($s_{\mathbf{g}}$–$s_{\mathbf{h}}$) and has an extinction distance given by ($\xi_{\mathbf{g}-\mathbf{h}}$). If $s_{\mathbf{g}}$ and $s_{\mathbf{h}}$ are equal, then the coupling between these beams will be strong. Furthermore, the characteristic length for the coupling in this example will be $\xi_{2\mathbf{g}-\mathbf{g}}$ or $\xi_{\mathbf{g}}$ which is what you would have guessed by thinking about Figure 27.5A and C.

27.9.D The Future Is Now

Several new developments are changing how we practice the WB technique. The main point here is that you should remember the principles because they will not change.

- Slow-scan CCD cameras give a very linear response and therefore make quantitative analysis of WB images possible. For this to happen, computer modeling of the defect and simulation of the image will be needed. We will return to this topic in Chapter 30 but mainly for HRTEM.
- An FEG and energy filtering allow us either to minimize the effect of variations in the energy (decrease diffuse scattering) or to form WB images using particular sections of the energy-loss spectrum. To make use of this extra information, we need to extend the theory.
- Image processing and frame averaging should allow us to reduce the noise, aid quantification, and use WB microscopy for in-situ studies.

Image simulation, as we described in Section 26.11, will allow us to be more quantitative in our interpretation of WB images when we actually have quantitative recording.

CHAPTER SUMMARY

The basic idea of the WB technique is very simple: using a large value of **s** gives a small ξ_{eff} and hence a narrow image of most defects, since the width of a dislocation is related to $\xi_{eff}/3$. What you should remember is that the value of s for a particular diffraction condition **g**(n**g**) depends not only on n and **g**, but also on the lattice parameter of the crystal and the wavelength of the electrons. You will see the 'magic number' $s = 0.2\ \text{nm}^{-1}$ quoted often.

- Remember that this number gives a rule of thumb if you want to do quantitative analysis. It does not usually correspond to **g**(3**g**).
- Don't use the **g**(3**g**) condition without calculating the value of $\mathbf{s}_g$.
- The term **s**.**R** has been neglected in this analysis; we usually assume that the deformable-ion model from Section 25.13 is valid.

You can often get all the information you need with less effort using a somewhat smaller value of **s**. As always, the longer you take to perfect the image, the more likely you are to alter your specimen, especially the defect structure.

Finally, remember that the diffracted beam travels parallel to $\mathbf{k}_D$. Therefore, the image of any defect is also projected in this direction. Even though the Bragg angle is small, this means, for example, that the apparent separation of defects in the image may not be equal to their horizontal separation relative to their glide planes if the defects are located at different heights in the specimen. This projection error can vary depending on the **g** and **s** used to form the image and the orientation of the specimen.

THE TECHNIQUE

Cockayne, DJH 1972 *A Theoretical Analysis of the Weak-Beam Method of Electron Microscopy [Defectoscopy]* Z. Naturf. **27a**, 452–460. Derives equation 27.2.

Cockayne, DJH 1981 *Weak-Beam Electron Microscopy* Ann. Rev. Mater. Sci. **11**, 75–95. A review of WB.

Cockayne, DJH, Ray, ILF and Whelan, MJ 1969 *Investigations of Dislocation Strain Fields Using Weak Beams* Phil. Mag. **20**, 1265–1270. The original paper; derives equations 27.10 and 27.15.

Hirsch, PB, Howie, A and Whelan, MJ 1960 *A Kinematical Theory of Diffraction Contrast of the Electron Transmission Microscope Images of Dislocations and Other Defects* Proc. Roy. Soc. London A **252**, 499–529. Always worth another look.

Stobbs, WM 1975 *The Weak Beam Technique in Electron Microscopy in Materials Science*, vol. **II** (Eds. U Valdrè and E Ruedl), p. 591–646, CEC Brussels. A masterful review which also compared WB and a Stradivarius.

Stobbs, WM and Sworn, C 1971 *The Weak Beam Technique as Applied to the Determination of the Stacking-Fault Energy of Copper* Phil. Mag. **24**, 1365–1381. WB calculations using anisotropic elasticity.

SOME DEFECTS

Carter, CB 1984 *What's New in Dislocation Dissociation?* in *Dislocations –1984* (Eds. P Veyssière, L Kubin, and J Castaing), p. 227, Editions du CNRS Paris. A 50th birthday review of dissociated dislocations.

Carter, CB and Holmes, SM, 1975 *The Study of Faulted Dipoles in Copper Using Weak-Beam Electron Microscopy* Phil. Mag. **32** (3), 599–614.

Carter, CB, Mills, MJ, Medlin, DL and Angelo, JE 1995 *The 112 Lateral Twin Boundaries in FCC Metals* in *7th International Conference Intergranular and Interphase Boundaries in Materials*, Lisbon, Portugal. The 'thickness' of a plane is important in WB! (Section 27.7.)

Föll, H, Carter, CB and Wilkens, M 1980 *Weak-Beam Contrast of Stacking Faults in Transmission Electron Microscopy* Phys. stat. sol. (A) **58**, 393–407. Discusses the anomalous WB contrast from SFs.

Gerthsen, D and Carter, CB 1993 *Stacking-Fault Energies of GaAs* Phys. stat. sol. (A) **136**, 29–43. An experimentalist's comparison of WB and HRTEM.

Hazzledine, PM, Karnthaler, HP and Wintner, E 1975 *Non-parallel Dissociation of Dislocations in Thin Foils* Phil. Mag. **32**, 81–97. Used WB to show the unambiguous effect of surfaces on dislocation core splitting: a paper with broad implications for the microscopists.

Wilson, AR and Cockayne, DJH 1985 *Calculated Asymmetry for Weak Beam Intrinsic Stacking Fault Images* Phil. Mag. **A51**, 341–354. More on the WB contrast from SFs.

SELF-ASSESSMENT QUESTIONS

Q27.1 Write down the equation relating s_{eff} to s.
Q27.2 Write down the equation relating ξ_{eff} to $\mathbf{s_g}$.
Q27.3 Write down the equation relating s to n, $\mathbf{g}$, and λ.
Q27.4 If the $3\mathbf{g}$ reflection is excited, where are the $\mathbf{g}$ and $3\bar{\mathbf{g}}$ Kikuchi lines?
Q27.5 What is the optimum value of s for WB and why is this a compromise value?
Q27.6 If an Si specimen is set to $\mathbf{g}$,$3\mathbf{g}$ conditions at 100 kV, how wide will the image of a dislocation be?
Q27.7 How will the image differ if we use 300 kV rather than 100 kV for $\mathbf{g}$,$3\mathbf{g}$ conditions?
Q27.8 Name the principal methods used for calculating WB images.
Q27.9 What is the radius of the phasor diagram of a perfect crystal in the WB condition?
Q27.10 Why is the separation of the two lines in a $\mathbf{g}.\mathbf{b}=2$ WB image not equal to the separation of the two partial dislocations which cause these lines?
Q27.11 A WB image can show two lines when $\mathbf{g}.\mathbf{b}=2$ for a non-dissociated dislocation. How can this happen?
Q27.12 Why is the circumference of the circle in a phasor diagram of a perfect crystal imaged under WB conditions interesting/instructive?
Q27.13 Why might the contrast from an extrinsic stacking fault differ from that from an intrinsic SF under WB conditions?
Q27.14 Sketch the phasor diagrams which show the difference behavior of a dislocation image in $\pm\mathbf{g}$.
Q27.15 What extinction distance do you expect to have for the coupling of $\mathbf{g}$ and $2\mathbf{g}$ when $3\mathbf{g}$ is excited?
Q27.16 In WB images, we usually set $w \geq 5$. Why?
Q27.17 Why do we avoid exciting other reflections when using the WB technique? Is it more important than when forming BF images?
Q27.18 What is the classic value for $\mathbf{s_g}$ in WB images and why was this value originally chosen?
Q27.19 Why is the condition with s positive preferred over that with s negative even though both conditions may have the same value of s?
Q27.20 Why is WB microscopy more difficult as we go to higher kV or to lower kV? (Higher and lower indicate relative to 100 kV.)

TEXT-SPECIFIC QUESTIONS

T27.1 Starting from Figures 27.1 and 27.4, deduce equation 27.5.
T27.2 Check Table 27.1 (any errors?) and add a column for 120-keV electrons. (Why?)
T27.3 Consider the lowest part of Figure 27.7C (a line at and parallel to the bottom of the figure). Draw an accurate sketch of what the cross section of the sample might be and justify your answer.
T27.4 If the sample in Figure 27.8 is close to the (111) plane, deduce the character of the dislocations in this image which lie in the (111) plane. Explain your answer very carefully.
T27.5 If we have $\mathbf{g}(3\mathbf{g})$ conditions, through what angle must the planes near the dislocation bend in order to be in the Bragg condition in Figure 27.9?
T27.6 Using equation 27.9, deduce the width of the dislocation image for $\mathbf{g}(3\mathbf{g})$ and $\mathbf{g}(4\mathbf{g})$ conditions if $\mathbf{g}$ is $2\bar{2}0$ and the sample is parallel to the (111) plane.
T27.7 Consider Figure 27.15. Deduce how many {111} planes apart the continuous and dark (in D) SFs are. Explain all the assumptions and steps in your answers.
T27.8 Consider Figure 27.19. Deduce the character of all of the perfect and partial dislocations in this image assuming that they all lie in the (111) plane.
T27.9 Consider Figure 27.23. If the sample is (111) and the $\mathbf{g}(3\mathbf{g})$ conditions were used to form the image, deduce the thickness of the sample. Explain all your assumptions and arguments.
T27.10 Consider Figure 27.21. What and where (indices and direction) is the most likely $\mathbf{g}$ used to form this image? Explain your reasoning.
T27.11 (An extra challenge question) Consider Figure 27.16. When will you see the left phasor and when will you see the one on the right, assuming you are looking at the same dislocation.
T27.12 Determine the value of s for 220 taking 100 kV and Cu (Figure 27.2).
T27.13 Figure 27.8. Estimate the value for the SFE. Assume a Cu fcc alloy, $\mathbf{g}$,$3\mathbf{g}$ conditions.
T27.14 Estimate the maximum thickness of the sample in Figure 27.13. Spinel with 220, $\mathbf{g}$,$3\mathbf{g}$, 100 kV.
T27.15 Figure 27.14. Phasor diagram-100 kV. For Cu, $\mathbf{g}=220$ and $\mathbf{g}$,$3\mathbf{g}$ conditions. How thick is the sample?
T27.16 Challenging. Derive equation 27.8 explaining all your assumptions.
T27.17 Challenging: Figure 27.9 greatly exaggerates the curvature of the dislocations. Estimate the length of the column that contributes to $\mathbf{g}$ for $\mathbf{g}$,$3\mathbf{g}$ and 100-kV electrons.
T27.18 What value of n would you use to produce WB images of dislocation for these reflections assuming 100-, 200-, and 300-kV electrons? (a) $(11\bar{2}0)$ in Al_2O_3 (alumina); (b) (220) in $MgAl_2O_4$ (spinel); (c) (111) in $MgAl_2O_3$; (d) (222) in YAG; (e) (220) in Si; (f) (111) in Si; (g) (110) in W; (h) (200) in Au.

T27.19 If the specimen is oriented for **g**,3**g** (Cu, 100 kV) throughout, what angle must you tilt to set up **g**,4**g** conditions? (Answer in degrees and radians.)

T27.20 Imagine four overlapping SFs in a crystal. What is the maximum intensity you could produce in the image? Give your answer as a multiple of the intensity possible for a perfect crystal. Use Cu 100 kV and **g**,5**g** for **g** = 111.

T27.21 Discuss how you might use weak-beam imaging to profile the topography of a thin sample even though the thickness can be changed by undulations on either surface. Hint: this question is related to electron tomography.

T27.22 Discuss how weak-beam images will be improved using a new energy-filtered TEM equipped with a drift-compensated stage. Will you still use the same value of s?

T27.23 Professor J wants to use weak-beam imaging to obtain information on the core of dislocations. What factors will ultimately limit the information she can obtain? How will these limitations depend on the nature of the sample?

28

High-Resolution TEM

CHAPTER PREVIEW

We will now rethink what we mean by a TEM, in a way that is more suitable for HRTEM, where the purpose is to maximize the useful detail in the image. (Note the word *useful* here.) You should think of the microscope as an optical device that transfers information from the specimen to the image. The optics consists of a series of lenses and apertures aligned along the optic (symmetry) axis. What we would like to do is to transfer *all* the information from the specimen to the image, a process known as mapping. There are two problems to overcome and we can never be completely successful in transferring *all* the information. First, as you know from Chapter 6, the lens system is not perfect so the image is distorted and you lose some data because the lens has a finite size (Abbe's theory). The second problem is we have to interpret the image using an atomistic model for the material. Ideally, this model will include a full description of the atomic potential and the bonding of the atoms, but we don't know that either. We will also need to know exactly how many atoms the electron encountered on its way through the specimen. So most of our task will be concerned with finding the best compromise and producing models for the real situation. To conclude our discussion of the theory, we will introduce the language of *information theory,* which is increasingly used in HRTEM.

Perhaps the biggest challenge for HRTEM now is interpretation; getting an image that shows detail finer than 0.2 nm is now routine: correctly interpreting it may not be routine. This caution is particularly critical when you're examining nanomaterials. We close the chapter with a review of the experimental applications of HRTEM to include periodic and non-periodic materials, mixtures of the two, or just single atoms.

A few words of caution: HRTEM is one of the most important aspects of TEM and is usually an essential consideration in obtaining funds for a new TEM. This chapter introduces some standard concepts and some ideas that are just beginning to be explored. Some topics are straightforward; others are exceedingly difficult.

28.1 THE ROLE OF AN OPTICAL SYSTEM

What the microscope does is to transform each point on the specimen into an extended region (at best, a circular disk) in the final image. Since each point on the specimen may be different, we describe the specimen by a specimen function, $f(x,y)$. The extended region in the image which corresponds to the point (x,y) in the specimen is then described as $g(x,y)$ as shown schematically in Figure 28.1; note that both f and g are functions of x and y.

If you consider two nearby points, A and B, they will produce two overlapping images g_A and g_B as shown in Figure 28.2. If we extend this argument, we can see that each point in the image has contributions from many points in the specimen. We express this result mathematically by

$$g(\mathbf{r}) = \int f(\mathbf{r}')h(\mathbf{r}-\mathbf{r}')d\mathbf{r}' \tag{28.1}$$

$$= f(\mathbf{r}) \otimes h(\mathbf{r}-\mathbf{r}') \tag{28.2}$$

POINT-SPREAD FUNCTION

Since $h(\boldsymbol{r})$ describes how a point spreads into a disk, it is known as the point-spread function or smearing function, and $g(\mathbf{r})$ is called the convolution of $f(\mathbf{r})$ with $h(\mathbf{r})$.

Here $h(\mathbf{r}-\mathbf{r}')$ is a weighting term telling us how much each point in the specimen contributes to each point in the image.

Spence calls $h(\mathbf{r})$ the impulse response function, and notes that it can only apply to small patches of specimen

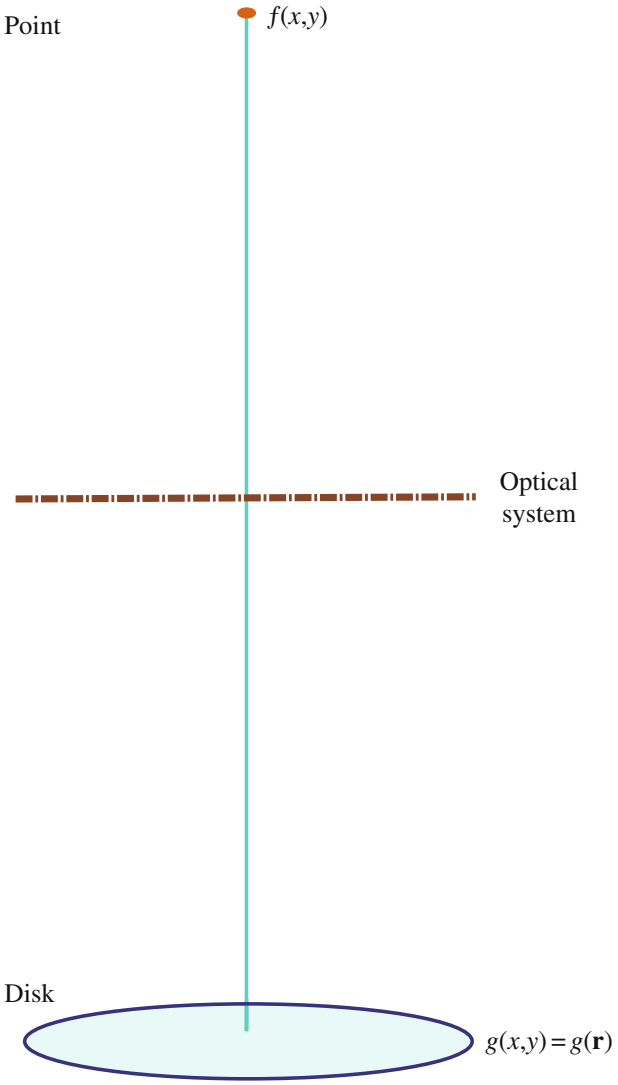


FIGURE 28.1. An optical system transforms a point in the specimen (described by $f(x,y)$) into a disk in the image described by $g(x,y)$. The intensity in the image at point (x,y) can be described by the function $g(x,y)$ or $g(\mathbf{r})$. It has a unique value for each value of (x,y), so we say it is a representation of the image.

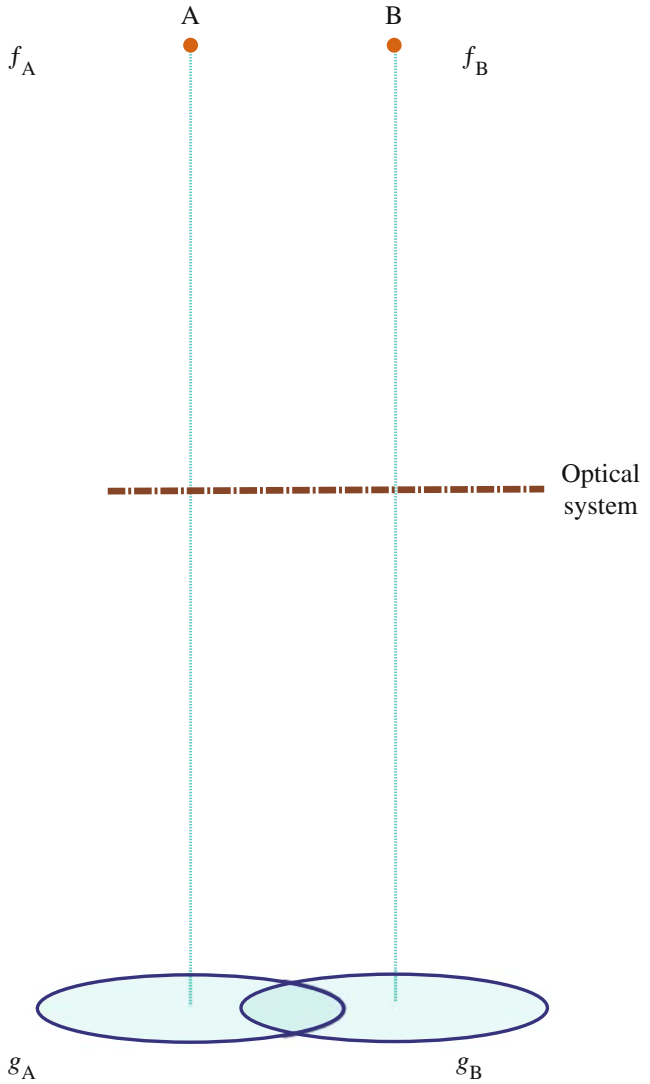


FIGURE 28.2. Two points, f_A and f_B, in the specimen produce two disks, g_A and g_B, in the image.

which lie in the same plane and are close to the optic axis. The symbol $\otimes$ indicates that the two functions, f and h, are 'folded together' (multiplied and integrated) or 'convoluted with one another.'

28.2 THE RADIO ANALOGY

We can compare this imaging process with the task of recording the sound of an orchestra on a record/tape/CD or even transmitting to the brain directly or via a radio. We want to hear the loud drum and quiet flute (large amplitude and small amplitude); we want to hear the high note on the violin and the low note on the double bass (high frequency and low frequency). Our audio amplifier has limits on both the low and high frequencies, so we won't achieve perfect reproduction. The importance of amplitude is obvious as discussed in Chapter 22, but how do we define frequency in a TEM image? High frequency in audio is related to $1/t$; frequencies in lattice images are related to $1/x$. So the high spatial frequencies simply correspond to small distances. What we are looking for in high-resolution work are the high spatial frequencies. Notice our use of high/low and large/small.

SPATIAL FREQUENCY
High resolution requires high spatial frequencies.

Figure 28.2 shows two points A and B in the specimen and their disk images on the screen. We see disks (see our discussion of the Rayleigh disk in Chapter 6)

because the lens system is not perfect. We can also write $g(x,y)$, the intensity of an image at point (x,y), as $g(\mathbf{r})$, and in the simplest case, these disks have uniform intensity. We can always represent any function in two dimensions as a sum of sine waves.

$$g(x,y) = \sum_{u_x,u_y} G(u_x - u_y)\exp\left(2\pi i\left(xu_x + yu_y\right)\right) \qquad (28.3)$$

$$g(x,y) = \sum_{\mathbf{u}} G(\mathbf{u})\exp(2\pi i\ \mathbf{u}\,.\,\mathbf{r}) \qquad (28.4)$$

Here $\mathbf{u}$ is a reciprocal-lattice vector, the spatial frequency for a particular direction. We have expressed $g(\mathbf{r})$ in terms of a combination of the possible values of $G(\mathbf{u})$, where $G(\mathbf{u})$ is known as the Fourier transform of $g(\mathbf{r})$. We can now define two other Fourier transforms

$F(\mathbf{u})$ is the Fourier transform of $f(\mathbf{r})$,

and

$H(\mathbf{u})$ is the Fourier transform of $h(\mathbf{r})$.

FOURIER TRANSFORM

The Fourier transform of a function is an expression of that function as a 'sum' of frequencies; it is the frequency-domain representation of the function.

Since $h(\mathbf{r})$ tells us how information in real space is transferred from the specimen to the image, $H(\mathbf{u})$ tells us how information (or contrast) in $\mathbf{u}$ space is transferred to the image.

$H(\mathbf{u})$ is the contrast transfer function.

Now these three Fourier transforms are related by

$$G(\mathbf{u}) = H(\mathbf{u})F(\mathbf{u}) \qquad (28.5)$$

So a convolution in real space (equation 28.1) gives multiplication in reciprocal space (equation 28.5).

The factors contributing to $H(\mathbf{u})$ include

Apertures → The aperture function $A(\mathbf{u})$
Attenuation of the wave → The envelope function $E(\mathbf{u})$
Aberration of the lens → The aberration function $B(\mathbf{u})$

We write $H(\mathbf{u})$ as the product of these three terms

$$H(\mathbf{u}) = A(\mathbf{u})E(\mathbf{u})B(\mathbf{u}) \qquad (28.6)$$

The aperture function says that the objective diaphragm cuts off all values of $\mathbf{u}$ (spatial frequencies) greater than (higher than) some selected value governed by the radius of the aperture. The envelope function has the same effect but is a property of the lens itself, and so may be either more or less restricting than $A(\mathbf{u})$. $B(\mathbf{u})$ is usually expressed as

$$B(\mathbf{u}) = \exp(i\chi(\mathbf{u})) \qquad (28.7)$$

The term $\chi(\mathbf{u})$ can be written as

$$\chi(\mathbf{u}) = \pi\Delta f\lambda u^2 + \frac{1}{2}\pi C_s\lambda^3 u^4 \qquad (28.8)$$

We will give a simple (simplified) derivation of this equation in Section 28.6. It builds on the concepts we discussed in Chapter 6 when we examined the origin of C_s.

OVERFOCUS

$\Delta f > 0$ is known as overfocus. It means we have focused the objective lens on a plane above the specimen. (By above, we mean before the electrons reach the specimen; the story is the same if the microscope is upside down!)

Summarizing so far: High spatial frequencies correspond to large distances from the optic axis in the DP. The rays which pass through the lens at these large distances are bent through a larger angle by the objective lens. They are not focused at the same point by the lens, because of spherical aberration, and thus cause a spreading of the point in the image. The result is that the objective lens magnifies the image but confuses the fine detail. The resolution we require in HRTEM is limited by this 'confusion'

- Each point in the specimen plane is transformed into an extended region (or disk) in the final image.
- Each point in the final image has contributions from many points in the specimen.

LINEAR

What we need for our analysis to 'work' is a 'linear relationship between the image and the weak specimen potential.'

We now have to go back and look at how we can represent the specimen. That is, what is $f(\mathbf{r})$ in equation 28.1? (We'll use the coordinates $\mathbf{r}$ and x,y interchangeably in this discussion; the former is more compact but we can extend the latter notation to emphasize the possibility of a z component.)

28.3 THE SPECIMEN

Since we are using a TEM, we call the specimen function, $f(\mathbf{r})$, the specimen transmission function. Here you have to be very careful to remember that we are going to

use a model to represent the specimen and the model will make certain assumptions. A general model would describe $f(\mathbf{r})$ as

$$f(x,y) = A(x,y)\exp(-i\phi_t(x,y)) \tag{28.9}$$

where $A(x,y)$ is the amplitude (not the aperture function) and $\phi_t(x,y)$ is the phase which depends on the thickness of the specimen.

For our application to HRTEM, we simplify our model further by setting $A(x,y) = 1$; i.e., we set the incident wave amplitude to be unity. We can show that the phase change only depends on the potential $V(x,y,z)$ which the electron sees as it passes through the specimen (by following Van Dyck's argument). We will assume that the specimen is so thin that we can write down a projected potential $V_t(x,y)$ with t being the thickness of the specimen, as usual.

$$V_t(x,y) = \int_0^t V(x,y,z)dz \tag{28.10}$$

What we are doing is creating a two-dimensional projection of the crystal structure; this approach is critical to much of our interpretation of HRTEM images.

We can relate the wavelength, λ, of the electrons in vacuum to the energy. (Ideally, λ should have its relativistic value, but the principle is correct.)

$$\lambda = \frac{h}{\sqrt{2meE}} \tag{28.11}$$

(We'll give the analysis in a simple non-relativistic form for simplicity.) When the electrons are in the crystal, λ is changed to λ'

$$\lambda' = \frac{h}{\sqrt{2me(E + V(x,y,z))}} \tag{28.12}$$

so we can say that, when passing through a slice of material of thickness dz, the electrons experience a phase change given by

$$d\phi = 2\pi\frac{dz}{\lambda'} - 2\pi\frac{dz}{\lambda} \tag{28.13}$$

$$d\phi = 2\pi\frac{dz}{\lambda}\left(\frac{\sqrt{E + V(x,y,z)}}{\sqrt{E}} - 1\right) \tag{28.14}$$

$$d\phi = 2\pi\frac{dz}{\lambda}\left(\left(1 + \frac{V(x,y,z)}{E}\right)^{\frac{1}{2}} - 1\right) \tag{28.15}$$

$$d\phi = 2\pi\frac{dz}{\lambda}\frac{1}{2}\frac{V(x,y,z)}{E} \tag{28.16}$$

$$d\phi = \frac{\pi}{\lambda E}V(x,y,z)dz \tag{28.17}$$

$$d\phi = \sigma\, V(x,y,z)dz \tag{28.18}$$

So the total phase shift is dependent only on $V(x,y,z)$ since

$$d\phi = \sigma\int V(x,y,z)dz = \sigma V_t(x,y) \tag{28.19}$$

where $V_t(x,y)$ is the potential projected in the z-direction.

We call σ the interaction constant. (See more discussion of σ in Chapter 3 but be careful—all σs are not the same.) This σ tends to a constant value as V increases, since the energy of the electron is proportional to E or λ^{-1} (i.e., changes in the two variables, λ and E, tend to compensate for one another).

THE INTERACTION CONSTANT

This σ is not the stress or the scattering cross section. It is another elastic interaction.

Now, we can take account of absorption by including a function $\mu(x,y)$ so that our specimen transfer function $f(x,y)$ is now given by

$$f(x,y) = \exp[-i\sigma V_t(x,y) - \mu(x,y)] \tag{28.20}$$

The effect of this model is that, apart from $\mu(x,y)$, we have represented the specimen as a 'phase object.' This is known as the phase-object approximation or POA. We are actually lucky because the absorption will usually be small in the regime where the rest of the approximation holds.

POA

In general, the phase-object approximation only holds for thin specimens.

We can simplify the model further if the specimen is *very* thin, so that $V_t(x,y)$ is $\ll 1$. Then we expand the exponential function, neglecting μ and higher-order terms, so that $f(x,y)$ becomes

$$f(x,y) = 1 - i\sigma V_t(x,y) \tag{28.21}$$

Now we have reached the weak phase-object approximation or the WPOA. We see that the WPOA essentially says that, for a very thin specimen, the amplitude of a transmitted wave function will be linearly related to

the projected potential of the specimen. Note that in this model the projected potential is taking account of variations in the z-direction, and is thus very different for an electron passing through the center of an atom compared to one passing through its outer regions.

Fortunately, there are software packages that allow us to calculate what an image will look like for a particular specimen geometry. However, you must always remember that a model has been used to represent the specimen and have a clear understanding of its limits. To emphasize this last point, bear in mind that the WPOA fails for an electron wave passing through the center of a single uranium atom! An atomic layer of U would be too thick for the WPOA. As a second example, Fejes has shown that, for the complex oxide $Ti_2Nb_{10}O_{27}$, the WPOA is only valid if the specimen thickness is <0.6 nm! The good news is that the approach appears to be more widely applicable than these particular estimates would suggest.

28.4 APPLYING THE WPOA TO THE TEM

So far our treatment has been quite general, but now we use our WPOA model. If we use the expression for $f(\mathbf{r})$ given by equation 28.21, then equation 28.2 tells us that the wave function as seen in the image is given by

$$\psi(x,y) = [-i\sigma V_t(x,y)] \otimes h(x,y) \qquad (28.22)$$

If we represent $h(x,y)$ as $\cos(x,y) + i\sin(x,y)$, then $\psi(x,y)$ becomes

$$\psi(x,y) = 1 + \sigma V_t(x,y) \otimes \sin(x,y) - i\sigma V_t(x,y) \otimes \cos(x,y) \qquad (28.23)$$

As usual, the intensity is given by

$$I = \psi\psi^* = |\psi|^2 \qquad (28.24)$$

Multiplying this out and neglecting terms in σ^2 (because σ is small), we find that

$$I = 1 + 2\sigma V_t(x,y) \otimes \sin(x,y) \qquad (28.25)$$

Knowing this result we can say that, in the WPOA, only the imaginary part of $B(\mathbf{u})$ in equation 28.7 contributes to the intensity in equation 28.24 (because it gives the imaginary part of $h(x,y)$). Thus, we can set $B(\mathbf{u}) = 2\sin\chi(\mathbf{u})$ rather than $\exp(i\chi(\mathbf{u}))$. (Notice the 2!)

We can now define a new quantity, $T(\mathbf{u})$, which we could call the intensity transfer function to distinguish it from $H(\mathbf{u})$. It's given by

$$T(\mathbf{u}) = A(\mathbf{u})E(\mathbf{u})2\sin\chi(\mathbf{u}) \qquad (28.26)$$

Note that $T(\mathbf{u})$ is not identical to $H(\mathbf{u})$, which we defined in equation 28.6. The '2' in equation 28.26 is the '2' in equation 28.25 and arises because we are interested in the intensity in the beam, and therefore we multiplied ψ by its complex conjugate in equation 28.24. You may also see authors use a negative sign in equation 28.26 (in particular, in Reimer's text). This has the effect of inverting the graph of $B(\mathbf{u})$ versus $\mathbf{u}$ and making $B(\mathbf{u})>0$ for positive phase contrast.

A note on terminology. You will often see $T(\mathbf{u})$ rather than $H(\mathbf{u})$ called the contrast transfer function in the HRTEM literature. This terminology comes from the analysis of the imaging process for incoherent light in visible-light optics. With coherent illumination, $T(\mathbf{u})$ and $H(\mathbf{u})$ are identical. The smearing function (point-spread function) for that case is the Fourier transform of the CTF. The equation describing $T(\mathbf{u})$ was derived for the situation where we have *coherent* imaging. For incoherent light the smearing function would be

$$\cos^2(x,y) + \sin^2(x,y) \qquad (28.27)$$

which is just unity.

So the CTF in HRTEM would be different from $T(\mathbf{u})$, and therefore we will call $T(\mathbf{u})$ the *objective lens transfer function*.

28.5 THE TRANSFER FUNCTION

You must note two things here. First, as we just said, the transfer function, $T(\mathbf{u})$, formulation applies to any specimen, and second, $T(\mathbf{u})$ is *not* the CTF of HRTEM. The problem with this formulation is that the image wave function is not an observable quantity! What we observe in an image is contrast, or the equivalent in optical density, current readout, etc., and this is not linearly related to the object wave function. Fortunately, there is a linear relation involving observable quantities under the special circumstances, where the specimen acts as a WPO.

TRANSFER FUNCTION

When $T(\mathbf{u})$ is negative, positive phase contrast results, meaning that atoms would appear dark against a bright background. When $T(\mathbf{u})$ is positive, negative phase contrast results, meaning that atoms would appear bright against a dark background. When $T(\mathbf{u}) = 0$, there is no detail in the image for this value of $\mathbf{u}$. (Note that we assume here that $C_s > 0$.)

If the specimen acts as a weak-phase object, then the transfer function $T(\mathbf{u})$ is sometimes called the CTF, because there is no amplitude contribution, and the output of the transmission system is an observable quantity (image contrast). The transfer function appropriate for this image formation process has the form

which we derived above (equation 28.26) and if we ignore $E(\mathbf{u})$,

$$T(\mathbf{u}) = 2A(\mathbf{u}) \sin \chi(\mathbf{u}) \qquad (28.28)$$

where we know that $A(\mathbf{u})$ is the aperture function and might call $\chi(\mathbf{u})$ the phase-distortion function.

$\chi(\mathbf{u})$

In other words, the phase-distortion function has the form of a phase shift expressed as $2\pi/\lambda$ times the path difference traveled by those waves affected by spherical aberration (C_s), defocus (Δz), and astigmatism (C_a).

Assuming that astigmatism can be properly corrected, the phase-distortion function is the sum of two terms. If the CTF is now compared to the phase-distortion function, a number of observations can be made. Note that the CTF is oscillatory; there are 'bands' of good transmission separated by 'gaps' (zeros) where no transmission occurs.

The CTF shows maxima (meaning maximum transfer of contrast) whenever the phase-distortion function assumes multiple odd values of $\pm\pi/2$. Zero contrast occurs for $\chi(\mathbf{u})$ = multiple $\pm\pi$.

The reason that negative $T(\mathbf{u})$ gives positive phase contrast is that there is a phase shift of $-\pi/2$ due to diffraction. If a diffracted beam is further phase shifted by $-\pi/2$, it subtracts amplitude from the forward scattered beam, causing atoms to appear dark (positive contrast). If the same beam is instead phase shifted by $+\pi/2$, it adds amplitude to the forward scattered beam (they are 'in phase'), causing atoms to appear bright (negative contrast).

28.6 MORE ON $\chi(\mathbf{u})$, $\sin \chi(\mathbf{u})$, AND $\cos \chi(\mathbf{u})$

The ideal form of $T(\mathbf{u})$ would be a constant value as $\mathbf{u}$ increases, as shown in Figure 28.3; $T(\mathbf{u})$ must be zero at $\mathbf{u} = 0$ but, since small values of $\mathbf{u}$ correspond to very large values of x (i.e., long distances in the specimen), this is not a problem. If $T(\mathbf{u})$ is large, it means that information with a periodicity or spatial frequency corresponding to that value of $\mathbf{u}$ will be strongly transmitted, i.e., it will appear in the image. What we then need is that the different values of $\mathbf{u}$ give the same contrast. Then all the atoms in a crystal appear as black spots, say, rather than some as black spots and others as white spots; if the latter occurs, interpretation will be difficult!

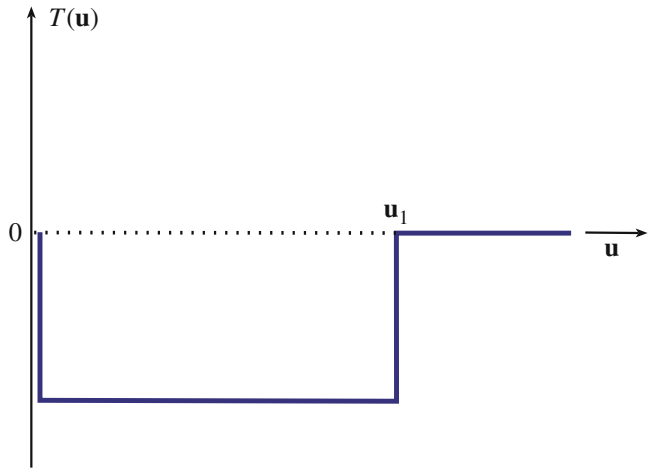


FIGURE 28.3. The ideal form of the transfer function, $T(\mathbf{u})$. In this example, $T(\mathbf{u})$ is large and negative between $\mathbf{u} \neq 0$ and $\mathbf{u} = \mathbf{u}_1$.

USEFUL u

$T(\mathbf{u})$ becomes zero again at $\mathbf{u} = \mathbf{u}_1$; what we would like is for $\mathbf{u}_1$ to be as large as possible. If $T(\mathbf{u})$ crosses the $\mathbf{u}$-axis the sign of the transfer function reverses. This means that $\mathbf{u}_1$ defines the limit at which our image may be quite directly interpreted; it is a very important parameter.

We will now go through a simple exercise to produce an expression for $\chi(\mathbf{u})$. If we combine the effects of the spherical aberration (equation 6.14) and the defocus (equations 11.18) of the objective lens, we find that a point at the specimen will actually be imaged as a disk with radius $\delta(\theta)$.

$$\delta(\theta) = C_s\theta^3 + \Delta f\theta \qquad (28.29)$$

Due to the spherical aberration of the objective lens and the finite value of Δf, the rays which pass through the objective lens at angle θ are not focused on the Gaussian image plane. If we only had one value of θ, we would still be all right! Of course, we have a range of values, so we average (integrate) these with respect to θ to give

$$D(\theta) = \int_0^\theta \delta(\theta) d\theta = \frac{C_s\theta^4}{4} + \Delta f\frac{\theta^2}{2} \qquad (28.30)$$

Now, Bragg's law tells us that

$$2d \sin \theta_B = n\lambda \qquad (28.31)$$

or since θ_B is small

$$2\theta_B \cong \lambda g \qquad (28.32)$$

So, we can replace θ in equation 28.30 by λu where **u** is a general reciprocal-lattice vector. (Remember that the scattering angle is $2\theta_B$, not θ_B.)

We are interested in the phase $\chi(\mathbf{u})$, so we write

$$\chi(\mathbf{u}) = \text{phase} = \frac{2\pi}{\lambda} D(\mathbf{u}) = \frac{2\pi}{\lambda}\left(C_s \frac{\lambda^4 u^4}{4} + \Delta f \frac{\lambda^2 u^2}{2}\right) \quad (28.33)$$

and we have

$$\chi = \pi \Delta f\, \lambda u^2 + \frac{1}{2}\pi C_s \lambda^3 u^4 \quad (28.34)$$

which we quoted as equation 28.8. Clearly $\sin\chi(\mathbf{u})$ will be a complicated curve which will depend on the values of C_s (the lens quality), λ (the accelerating voltage), Δf (the defocus value you choose to form the image), and **u** (the spatial frequency). This topic is addressed in a novel way in the companion text; see also Section 3.3 of John Spence's book; most of us just start with equation 28.34 and this reasonable justification.

The best way to appreciate the importance of χ is to use one of the simulation packages discussed in Chapter 30 and vary each of the parameters one by one. The plot of $T(\mathbf{u})$ ($= 2\sin\chi$) versus **u**, shown in Figure 28.4, illustrates the main features. The curve has been drawn for $C_s = 1$ mm, $E_0 = 200$ kV, and a defocus value of –58 nm.

The important features of this curve are shown in Figures 28.4–28.6

- $\sin\chi$ starts at 0 and decreases. When **u** is small, the Δf term dominates.
- $\sin\chi$ first crosses the **u**-axis at $\mathbf{u}_1$ and then repeatedly crosses the **u**-axis as **u** increases.
- χ can continue forever but, in practice, it is modified by other functions which we discuss in Section 28.8.

Once you've selected your microscope and its objective lens, you have fixed C_s (unless you have a TEM with a C_s corrector); C_s does depend to some extent on the λ you choose. The curve of $T(\mathbf{u})$ versus **u** does not depend on your specimen. Figure 28.5 shows a series of $\sin\chi$ curves for an imaginary 200-kV microscope where C_s has been changed. In each case, the 'best' curve (we'll discuss this in a moment) has been chosen. You can appreciate that the smaller C_s values give the larger $\mathbf{u}_1$ values; so a small C_s means we can achieve a higher spatial resolution.

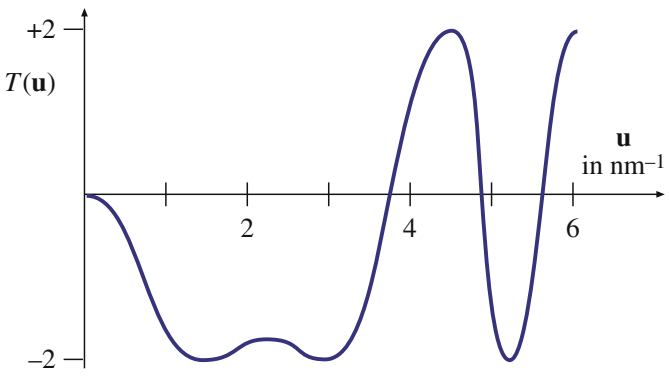


FIGURE 28.4. A plot of $T(\mathbf{u})$ versus **u** (C_s = 1 mm, E_0 = 200 kV, Δf = –58 nm).

C_s, Δf, AND β

High spatial frequencies ⇒ large diffraction angles ⇒ larger effect of objective lens (C_s).

So for a large objective aperture semi-angle β, the β^4 term wins, i.e., C_s wins. We can all vary Δf; some can vary C_s too.

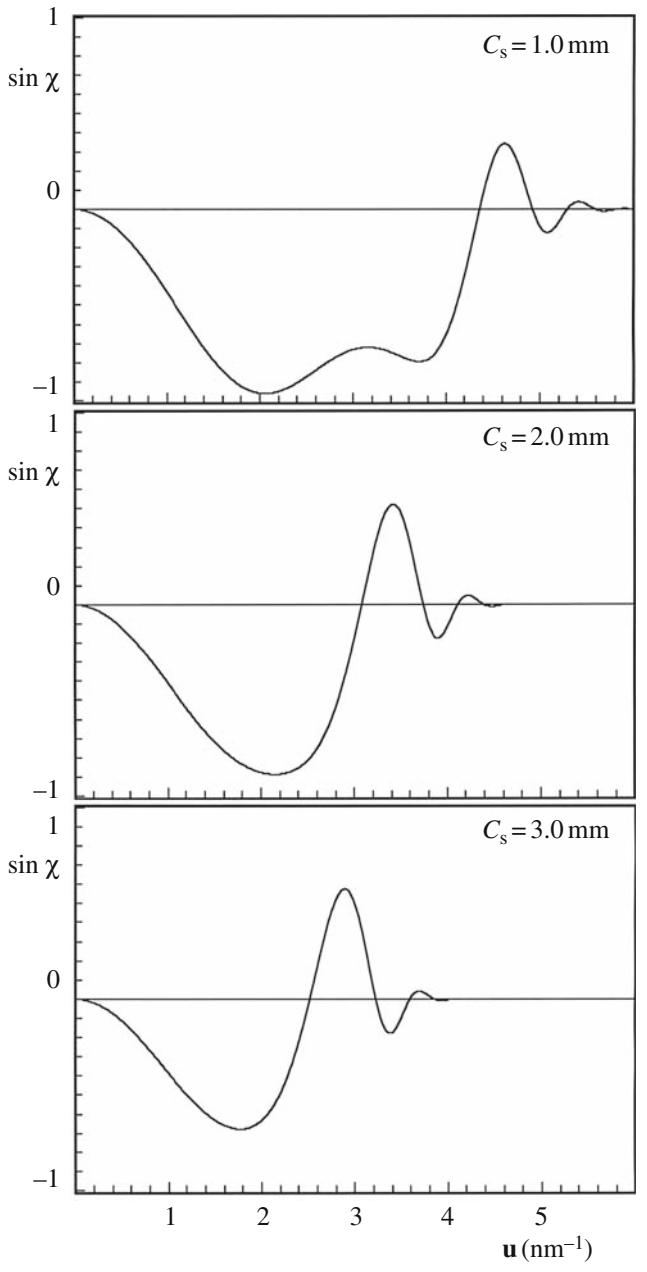


FIGURE 28.5. A series of $\sin\chi$ curves calculated for different values of C_s. Remember $2\sin\chi = T(\mathbf{u})$. ($E_0$ = 200 kV, Δf = –60 nm).

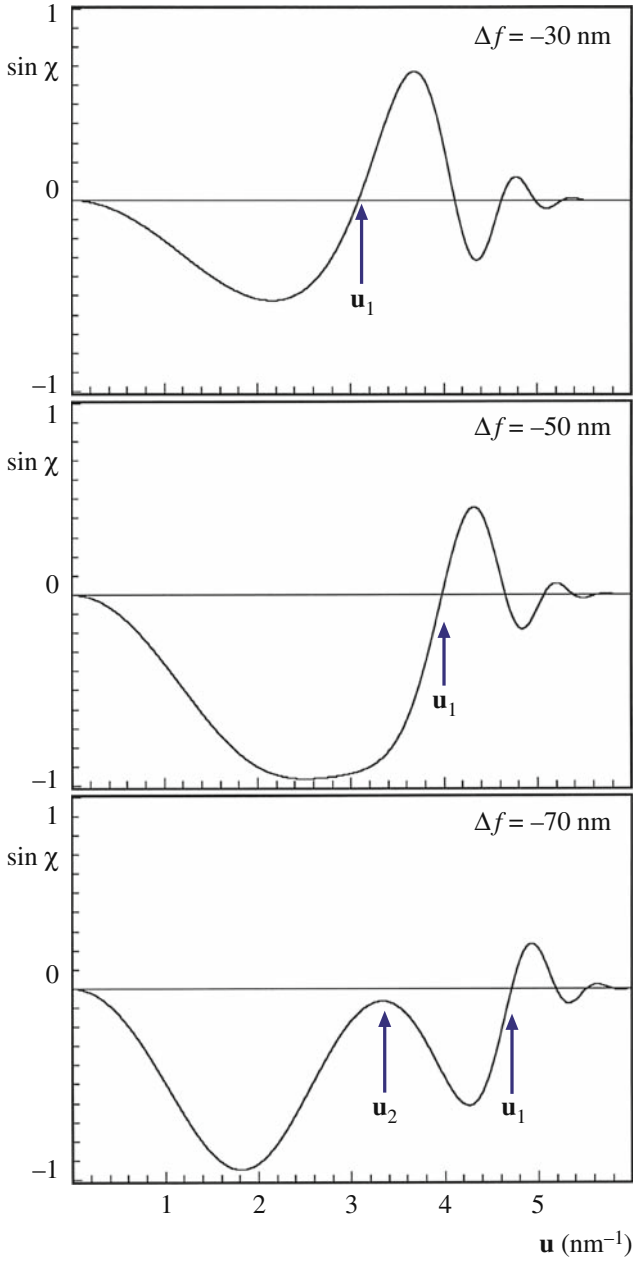


FIGURE 28.6. A series of sin χ curves calculated for different values of Δf. (E_0 = 200 kV; C_s = 1.0 mm.)

If instead we fix C_s and again plot the best curves, but this time varying λ, you can see that the smallest value of λ allows us to achieve a higher spatial resolution. The result is not surprising; we want a small C_s and a small λ or a high voltage. So we choose the microscope to optimize C_s and λ. Now we only have Δf to vary. The set of curves shown in Figure 28.6 illustrates the effect of varying Δf. Notice how the bump in the curve at u_2 will eventually increase as Δf increases until it crosses the **u**-axis so that $\mathbf{u}_1$ is suddenly much smaller. If we just make Δf smaller, then $\mathbf{u}_1$ steadily decreases. In the next section we will discuss the optimum value for Δf.

28.7 SCHERZER DEFOCUS

The presence of zeros in the CTF means that we have gaps in the output spectrum which do not contribute to the output signal: it's as if these frequencies were filtered out. Obviously, the best transfer function is the one with the fewest zeros, which would be the case for a perfect lens, for example. What Scherzer did back in 1949 was to notice that the CTF could be optimized by balancing the effect of spherical aberration against a particular negative value of Δf. This value has come to be known as 'Scherzer defocus,' Δf_{Sch} which occurs at

$$\Delta f_{Sch} = -1.2(C_s\lambda)^{1/2} \qquad (28.35)$$

At this defocus (which we'll derive below) all the beams will have nearly constant phase out to the 'first crossover' of the zero axis. This crossover point is defined as the instrumental resolution limit. This is the best performance that can be expected from a microscope unless we use image processing schemes to extract more information. In other words, this is not the information limit but it is the limit where we can use nearly intuitive arguments to interpret what we see. Again, as we discussed in Chapter 6 when we defined image resolution, you will see other authors give different values for the constant rather than the 1.2 given in equation 28.34; remember that this number is a calculated value, so it does depend on the details of your approximations.

This definition of resolution has new implications. The Rayleigh criterion which we used in Chapter 6, was only concerned with the ability to distinguish closely spaced point objects by eye. Our new definition requires a flat response in the object spectrum, and the goal is to have as many beams as possible being transferred through the optical system with identical phase, i.e., within the flat response regime. This is the underlying principle governing phase-contrast imaging in HRTEM.

DETAIL AND INFORMATION

A TEM image with detail of 0.66 Å was demonstrated in 1970 when the *interpretable* resolution was about 3.3 Å. So just because you can see detail in the image does not mean that you can gain useful information about your specimen.

The closest we can get to the ideal curve in Figure 28.6 occurs when χ(**u**) is close to –120°; then sin χ will be near –1 when χ is between –120° and –60°. We know that when χ = π, sin χ = 0 so we want sin χ to be as large as possible over a large range of **u**; sin χ will be a nearly flat function if $d\chi/du$ is zero. So we look for the value of

Δf when $d\chi/du$ is zero and χ is -120° (you should consider why we choose this value of χ). Differentiating equation 28.29 gives

$$\frac{d\chi}{du} = 2\pi\Delta f\lambda u + 2\pi C_s\lambda^3 u^3 \qquad (28.36)$$

Set the left-hand term equal to 0

$$0 = \Delta f + C_s\lambda^2 u^2 \qquad (28.37)$$

When $\chi = -120^\circ$, equation 28.29 becomes

$$-\frac{2\pi}{3} = \pi\Delta f\,\lambda u^2 + \frac{1}{2}\pi C_s\lambda^3 u^4 \qquad (28.38)$$

Combining equations 28.37 and 28.38 gives a special value for Δf

$$\Delta f_{Sch} = -\left(\frac{4}{3}C_s\lambda\right)^{\frac{1}{2}} \qquad (28.39)$$

The subscript denotes the Scherzer defocus value. Since $(1.33)^{1/2} = 1.155$ ($\sim$1.2), we have deduced equation 28.35. At this value of Δf (in equation 28.37) we find by substituting Δf_{Sch} into equation 28.37 (and using $(1.155)^{1/2} = 1.51$) that we next cross the axis at

$$u_{Sch} = 1.51 C_s^{-\frac{1}{4}}\,\lambda^{-\frac{3}{4}} \qquad (28.40)$$

The resolution at the Scherzer defocus can then be defined as the reciprocal of u_{Sch}.

$$r_{Sch} = \frac{1}{1.51}C_s^{\frac{1}{4}}\,\lambda^{\frac{3}{4}} = 0.66 C_s^{\frac{1}{4}}\,\lambda^{\frac{3}{4}} \qquad (28.41)$$

You will often see this expression with different values for the constant (0.66) for reasons discussed back in Section 6.6.B (here we are essentially summing the effects of Δf and C_s). The value of the constant can be increased, thus lowering r_{Sch} (i.e., giving higher resolution) if we are less restrictive about the value we choose for χ.

The quantities $(C_s\lambda)^{\frac{1}{2}}$ and $\left(C_s\lambda^3\right)^{\frac{1}{4}}$ seen in equations 28.39 and 28.41 are so important in HRTEM that Hawkes has designated them to be the units 1 Sch and 1 Gl (the scherzer and the glaser) in honor of two of the most noted pioneers of HRTEM. Notice that these *units* vary depending on the microscope you're using!

You'll find it interesting to plot the phase shift due to varying of Δf and C_s using EMS (Section 1.6). An excellent, though advanced, discussion of such diagrams is given by Thon, who describes how they can be used to design phase plates for the TEM. Spence shows how you can use a plot of nu^{-2} versus u^2 to help you determine experimental values of Δf and C_s; see Figure 31.6A.

28.8 ENVELOPE DAMPING FUNCTIONS

The plots of $\chi(\mathbf{u})$ as a function of $\mathbf{u}$ could extend out as far as you want to plot them. In practice, they don't because of the envelope damping function. In other words, the $\chi(\mathbf{u})$ plot stops where it does because the microscope is incapable of imaging the finest detail due to reasons other than the simple transfer characteristics of a linear system.

We know from Chapters 5 and 6 that resolution is also limited by the spatial coherence of the source and by chromatic effects. We can include these effects in our analysis of images by imposing an envelope function on the transfer function. The result is that higher spatial frequencies that might normally pass through higher-order windows are in fact damped out, as shown in the plot in Figure 28.7B.

The exact mathematical form of these envelope functions is complex. In general, the result is described by multiplying the (objective lens) transfer function $T(\mathbf{u})$ by both the chromatic aberration envelope E_c and the spatial coherence envelope E_a to yield an effective transfer function $T_{eff}(\mathbf{u})$.

$$T_{eff}(\mathbf{u}) = T(\mathbf{u})E_cE_a \qquad (28.42)$$

The effect of the envelope functions is to impose a virtual aperture in the back focal plane of the objective

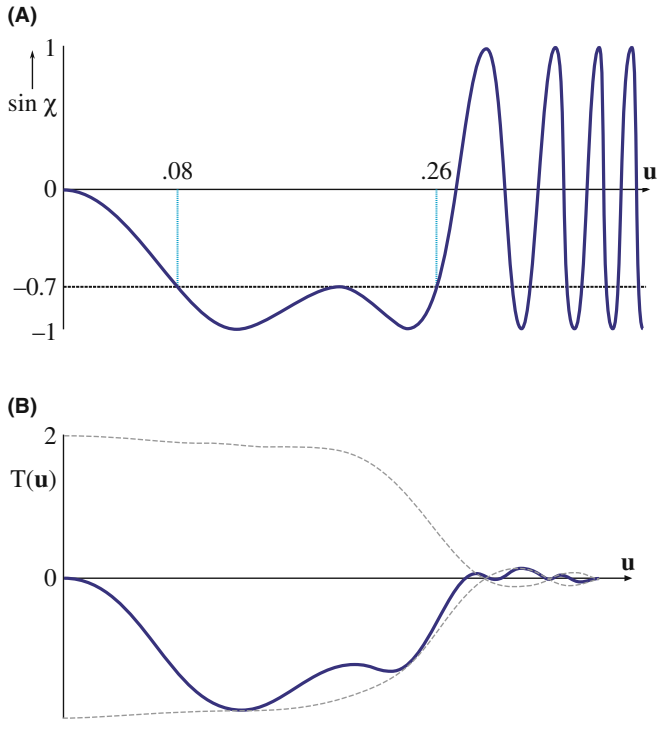


FIGURE 28.7. (A) $\sin\chi(\mathbf{u})$ versus $\mathbf{u}$ without damping of the higher spatial frequencies. (B) $T(\mathbf{u})$ versus $\mathbf{u}$ modified by the damping envelope (dashed line); $\Delta f = -100$ nm, $C_s = 2.2$ nm.

lens, *regardless* of the setting of focus. If we are going to use a physical aperture to remove unwanted noise, we should make it no larger than the 'virtual aperture' present due to this envelope. The presence of this virtual aperture means that higher-order passbands are simply not accessible. This cut-off thus imposes a new resolution limit on the microscope. This is what we earlier called the 'information retrieval limit' or simply the 'information limit.'

If we keep these restrictions in mind then we can say that, up to the instrumental resolution limit, phase-contrast images are directly (i.e., intuitively) interpretable; this limit is set by the crossover at Scherzer defocus or the envelope function, i.e., whichever equals zero first. If the information limit is beyond the Scherzer resolution limit, we need to use image-simulation software (see Chapter 30) to interpret any detail beyond the Scherzer limit.

So you can image columns of atoms along the incident-beam direction and their positions are faithfully rendered with respect to one another up to Scherzer resolution. If the microscope is operated at different defocus values, the crossovers in the transfer function make image interpretation more indirect and you have to resort to using computer simulation.

28.9 IMAGING USING PASSBANDS

Because of the focus dependence of the CTF, you, the microscope operator, have control over its overall form. For example, the worst case of contrast transfer is where all contrast is minimized. This minimum contrast (MC) defocus condition (Δf_{MC}) is also known as the dark-field focus condition in STEM imaging and occurs for a special value of χ.

$$\sin \chi(\mathbf{u}) = 0.3 \tag{28.43}$$

or

$$\Delta f_{MC} = -0.44(C_s\lambda)^{1/2} \tag{28.44}$$

The importance of this focus setting is that when you are actually working on the TEM, you can recognize this focus setting visually on the TEM screen, since it occurs when you can't see anything! If you adjust the focus to this condition visually, you then have a reference point from which you can change to the Scherzer defocus. The procedure is actually quite simple (after lots of practice), since you can minimize the contrast easily, providing you have correctly aligned the microscope and corrected the astigmatism.

Some other special settings of the CTF may also be useful. The idea is to make use of *passbands* or large 'windows' in the CTF to allow higher spatial frequencies to contribute to the image. As you see in Figure 28.8, what this requires is that χ is constant, or $d\chi/du$ small,

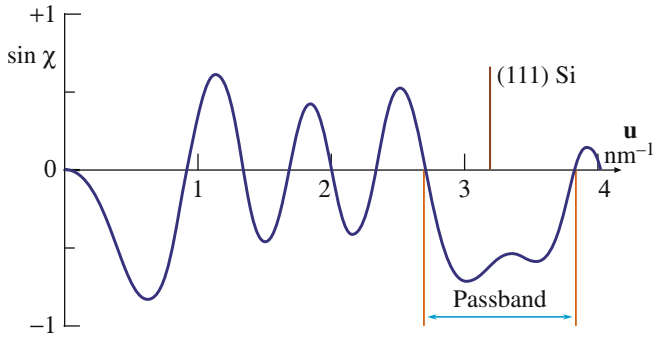


FIGURE 28.8. Special settings of the CTF to make use of passbands or 'windows' in the transfer function, here optimized to image Si(111).

over a range of **u** which includes the reflection of interest. These passbands occur periodically with underfocus at values set by

$$\Delta f_p^n = -\{([8n+3]/2)(C_s\lambda)\}^{1/2} \tag{28.45}$$

This formula is not an exact relationship but it gives us a good guide; its derivation is given by Spence. The $n = 0$ passband is, in fact, equivalent to the Scherzer defocus setting. This technique gives us access to higher spatial frequencies and thus finer detail in real space. The price we pay is that there are now zeros in the transfer function at lower spatial frequencies. For some applications, the presence of these zeros may be a problem, but for others, useful information can be obtained in these higher passband settings. For a microscope like the JEM 200 CX, these passband settings are –66 nm (Scherzer, or $n = 1$), –129 nm ($n = 2$), –169 nm ($n = 3$), –202 nm ($n = 4$), etc. Note that all are negative values of focus.

Hashimoto and Endoh defined an 'aberration-free focus' (AFF) condition for any specific crystal. The idea is to set the transfer function so that the gaps will only occur between Bragg reflections. All Bragg reflections would then see a window in the CTF out to very high order. This aberration-free focus setting is defined by

$$\Delta f_{AFF} \cong \{2(4m \pm 0.23)C_s\lambda^3/d^4\}(d^2/2\lambda) \tag{28.46}$$

where $m = 0, 1, 2, 3$, etc., and d is the fundamental lattice spacing of the first-order Bragg beam to be resolved. Applying this equation to a Au crystal in the [001] orientation where $d(020) = 0.2035$ nm, using a 100-kV microscope with $C_s = 0.75$ mm, gives $\Delta f_{AFF} = -53.3$ nm. At this setting of focus, the transfer function peaked at –2 for beams 020, 220, 040, 420, 440, and 060.

There is, of course, a catch. We can only use this technique when we know which spatial frequencies we are interested in. In other words, it is great for perfect crystals since we are only concerned with Bragg peaks. If defects are present we will then lose much of the information about the defect, since defects scatter between

the Bragg peaks. Any information falling in the gap of the CTF is lost to the image: in effect, the defect will be invisible!

You should therefore be very cautious in using higher passband settings. You may obtain a pretty picture which does not give a true image of your specimen. If you do use higher-order passbands, you must realize that you are imaging the specimen beyond the instrumental resolution limit so you can't use the intuitive approach for image interpretation. You must know exactly where the zeros are in the CTF. You can only know that by very careful evaluation of your images using diffractograms, computer simulation, and image processing.

28.10 EXPERIMENTAL CONSIDERATIONS

Whenever you are using HRTEM imaging, you must first ask what information you are hoping to obtain. Lattice-fringe images which show lots of straight lines but tell you nothing of where the atoms are located may be all that you need. These fringes are giving you information about the crystal orientation on a very fine scale. Another situation is illustrated by early studies of spinel. You would like to obtain information at, say, 0.23 nm (the spacing of the oxygen 111 planes), but your point-to-point resolution is 0.27 nm. You could still learn a lot about the spinel from the 46 nm spinel (111) planes, so you might even use an aperture to remove information which only adds uninterpretable detail below 0.46 nm. The difficulty comes when you want to relate your HRTEM image to the atomic structure of your specimen. Then you must remember that all of the above treatment is based upon the TEM specimen behaving as a WPO. Most specimens of interest do *not* satisfy this criterion.

MULTIPLE AND PLURAL

Note that the HRTEM community uses the term 'multiple scattering' to denote >1 scattering event. This terminology differs from that used by analytical microscopists, who define 'multiple' as >20 scattering events and reserve 'plural' for 2–20 events. In HRTEM you never hear of 'plural scattering.'

If you look at a typical HRTEM specimen, there will be a wedge-shaped region near the thinnest edge, and thickness extinction contours will be visible. As soon as the first contour is visible, the specimen is already much too thick to behave as a WPO! Multiple scattering limits most phase-contrast imaging conditions for crystalline materials.

Thicker specimens also are susceptible to Fresnel effects associated with spreading of the wave front as it is transmitted through more of the specimen along the beam direction. Inelastic scattering effects, etc., will also become important as the thickness increases. These effects are not easy to simulate in the computer, although the techniques we will discuss in Chapter 30 are very helpful.

To be really sure that you have correctly interpreted the image, the match between experimental and simulated images should be good over a range of thicknesses and defocus values, as we'll see more clearly in Chapter 30.

We can now summarize the ten steps you need to take to obtain a phase-contrast image with atomic resolution

- Choose an instrument of low C_s and small λ.
- Align it well; it will take time for the electronics and moving parts to become stable.
- Work with an undersaturated LaB_6 filament and a small condenser aperture (unless you have an FEG: see later).
- Perform current and voltage centering of the objective lens routinely and frequently at high magnification.
- Work in thin, flat, and clean regions of the specimen.
- Orient the specimen using small SAD apertures or bend contours in the image, so the beam is aligned along a zone axis.
- Correct the astigmatism, using optical diffractograms if necessary, but ideally on-line (Chapter 31).
- Find the minimum-contrast focus setting and record a through-focus series.
- Record the DP at the same setting of the condenser; calculate α, the convergence angle, and remember that angle means semiangle!
- Simulate and/or process the images using available computer codes (Chapter 30).

A comment on alignment: You'll find that it's relatively straightforward to align the electron beam with the current center or voltage center. The result is an image which does not shift as the current changes in the objective lens or the accelerating voltage fluctuates. As you'll appreciate more from Chapter 31, for the highest resolution, it is also critical that the incident beam is precisely parallel to the optic axis of the microscope. If the incident beam is not exactly aligned with the optic axis we can see the coma aberration, which is only important at the highest resolution. (A discussion of other aberrations is given in the companion text.) We refer to the process of aligning the beam with the optic axis as 'coma-free alignment.' The process involves alternately applying equal and opposite beam tilts to the incident beam; you choose the magnitude of the tilt to match the periodicity in the image. If there is a residual beam tilt of the incident beam away from the optic axis, then one image will look more distorted than the other. Adjust the beam tilt controls until both tilted images look equally distorted. Repeat this procedure for the orthogonal direction. You will need a lot of practice to do this successfully (see Section 30.5).

COMA

If a point object on the optic axis is imaged as a point, a similar point object off the axis may appear distorted. This distortion is known as coma or comatic aberration. In a telescope, the point (a star) would look like a comet—a cometary coma. (Coma: Latin for hair.)

Some final remarks on experimental techniques: Always remember that specimen orientation is very critical for HRTEM. Always be aware of contamination and damage caused by the electron beam; the specimen will have changed long before you can see the change by eye. Although HRTEM is now so much easier because high-quality CCD video cameras are available to give you an image at TV rates, don't spend any longer than you have to with the beam on the area of interest in your specimen. Get used to using the CCD and the computer. If you are going to do quantitative HRTEM, you'll have to be comfortable with both. Of course, if you use a remote microscope you aren't even sitting near the screen.

28.11 THE FUTURE FOR HRTEM

The historical approach to HRTEM was: be pleased if you recorded what you saw. Now machines are sufficiently stable that we can reliably record images at different values of Δf. Certainly as important is the availability of computers, as we will discuss in Chapters 30 and 31, since we can 'predict' the image for model structures and quantify the contrast of the image. We are thus able to do quantitative HRTEM (QHRTEM or HRQTEM!).

Another approach to improve resolution is provided by the FEG TEM. The beam in an FEG TEM is now more highly coherent, so the envelope function shown in Figure 28.7 extends to greater values of **u**. The computer becomes indispensable because we have to interpret images which have contrast reversals beyond Scherzer defocus. If a set of carefully designed multipole lenses is inserted into an HRTEM (by the manufacturer), it is possible to correct C_s or even make it variable like Δf.

If you use such a C_s-corrected TEM, you have to rethink your approach to HRTEM. Think what will happen to the scherzer and the glaser. What will the Gl/Sch be? The corrector proposed by Rose was shown schematically in Figure 6.12B. The actual corrector, incorporated into a 200-kV JEOL TEM, is shown in Figure 28.9 (with the covers removed); it significantly increased the height of the column. It is a combination of round lenses and hexapoles, all of which are magnetic elements. The hexapoles don't affect the paraxial path of the rays and only need to be stabilized to an accuracy

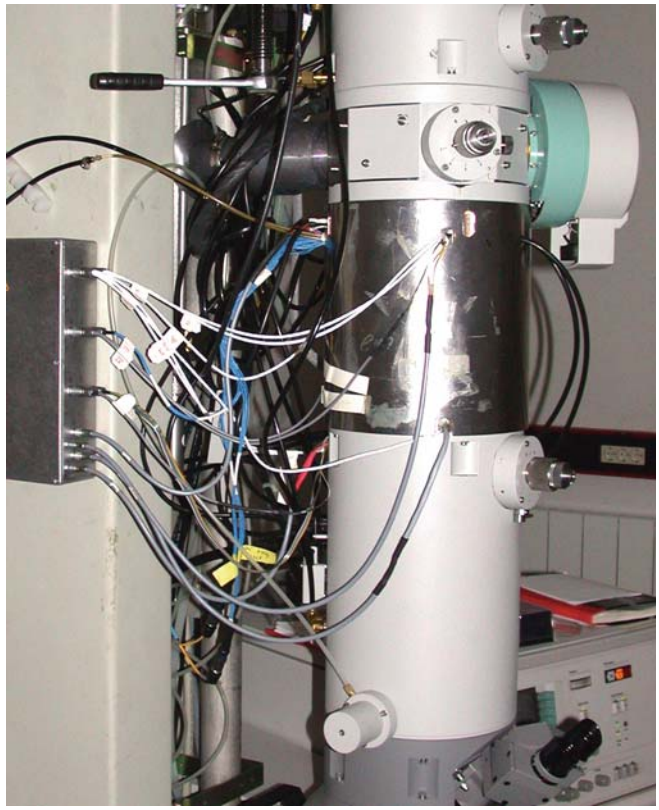


FIGURE 28.9. The post-objective lens corrector system as incorporated into a 200-kV JEOL TEM; it looks sleeker in the finished product. (See also Figure 6.12B.)

of 1 in 10^4 to give atomic resolution. When C_s is zero, the specimen resolution limit will be determined by C_c

$$d_{C_s=0} \approx \left[\left(\frac{\Delta E}{E}\right)\lambda C_c\right]^{\frac{1}{2}} \tag{28.47}$$

If $C_c = 2\,\text{mm}$ and $\Delta E \sim 0.3\,\text{eV}$, a 200-kV FEG TEM could achieve a resolution of 0.8 Å. If C_c is also corrected, which we'll see in Chapter 40, it is possible that the resolution will become limited by the fifth-order spherical aberration constant. In practice, it will be important to correct C_c in the lens design first. In Rose's original C_s-correcting lens, $C_s = 3\,\text{mm}$ and the resolution limit is 0.28 Å for this 200-kV FEG TEM. Other lens defects will limit this to ~0.5 Å but with a price of $12 M for a 1.25-MeV machine which changes your specimen in seconds, the Rose corrector will be quite important!

28.12 THE TEM AS A LINEAR SYSTEM

The discussion we went through above is an example of a much larger topic known as information theory. The concept of a 'phase-contrast transfer function' is central to this field. So you can understand the practice of

phase-contrast imaging at high resolution, we will briefly discuss the way an information specialist might view this process. We will define the transfer function in elementary terms, and make detailed reference to phase-contrast imaging in the TEM.

Remember, the purpose of the TEM is to transmit information about the specimen to the image. We can thus consider the microscope to be an 'information channel' and use the concepts of information theory.

- The input signal comes from the specimen.
- The output signal is the image.

If we neglect the effects of noise, there is a unique relation between the input signal and the output signal, determined by the optical system of the microscope.

Most information theory treats linear systems. A linear system is one which is characterized by the property that if

$$S_0(r_0) \rightarrow \text{Transmission System} \rightarrow S_1(r_1),$$

and if

$$S'_0(r_0) \rightarrow \text{Transmission System} \rightarrow S'_1(r_1),$$

(the prime here denotes the derivative) then the system is linear if

$$a(S_0) + b(S'_0) \rightarrow \text{Transmission System} \rightarrow a(S_1) + b(S'_1)$$

for any values of a and b. (It's like linear elasticity where we can simply add stresses and strains.)

The linear relation between input and output signals can be described by the concept of the CTF. Overall, the transfer function relates an input spectrum to an output spectrum, and it operates only in the frequency domain.

LINEAR SYSTEM

Schrödinger's wave equation is linear. Therefore, the amplitudes of an electron wave in the specimen are linearly related to the amplitudes of an electron wave in the image.

In general, for a linear system, if we know the CTF, then the relation between S_0 and S_1 is uniquely defined. On the other hand, if the relation between S_0 and S_1 could be empirically determined, then we can deduce the CTF.

One of the best examples of a linear system is an electrical transmission cable. The transfer of electrical signals through transmission lines can be made linear enough for the above theory to apply. Conversely, the transfer of mass-thickness information from a specimen to the optical density of a developed photographic negative is far from linear, and the above theory does not apply. Then why bother to discuss this in HRTEM? The answer lies in finding an appropriate linear relation between the object and the image.

28.13 FEG TEMs AND THE INFORMATION LIMIT

We've mentioned that an FEG reduces the instrumental contribution to chromatic aberrations and extends the envelope function to larger values of **u**. This means that information with higher spatial frequencies is transferred to the image. We've just analyzed the Scherzer defocus problem, so now we'll consider the information limit. The reason for emphasizing the FEG here is that it really makes a difference and we're just beginning to learn how to use this information: the contrast reversals mean that any image interpretation is not intuitive. The topic has been laid out in two papers by Van Dyck and de Jong but be warned—this topic is exceedingly tricky!

Since the information limit is determined by the envelope function, this is split into its separate terms. The total envelope function, $E_T(\mathbf{u})$, is the product of all of these (like equation 28.5).

$$E_T(\mathbf{u}) = E_c(\mathbf{u})E_s(\mathbf{u})E_d(\mathbf{u})E_v(\mathbf{u})E_D(\mathbf{u}) \qquad (28.48)$$

The individual envelope functions in equation 28.48 are

$E_c(\mathbf{u})$: for chromatic aberration.
$E_s(\mathbf{u})$: for the source dependence due to the small spread of angles from the probe.
$E_d(\mathbf{u})$: for specimen drift.
$E_v(\mathbf{u})$: for specimen vibration.
$E_D(\mathbf{u})$: for the detector.

As you can see, some of these envelope functions are new, some are old. We won't discuss all these functions; we'll only mention a couple of the key points.

The chromatic aberration is well known, and its envelope function $E_c(\mathbf{u})$ can be expressed by the equation

$$E_c(\mathbf{u}) = \exp\left[-\frac{1}{2}(\pi\lambda\delta)^2 u^4\right] \qquad (28.49)$$

where c reminds us that this is a chromatic aberration and δ is the defocus spread due to this aberration

$$\delta = C_c\left[4\left(\frac{\Delta I_{obj}}{I_{obj}}\right)^2 + \left(\frac{\Delta E}{V_{acc}}\right)^2 + \left(\frac{\Delta V_{acc}}{V_{acc}}\right)^2\right]^{\frac{1}{2}} \qquad (28.50)$$

The terms $\Delta V_{acc}/V_{acc}$ and $\Delta I_{obj}/I_{obj}$ represent the instabilities in the high-voltage (accelerator) supply

and the objective lens current. $\Delta E/V_{acc}$ is the intrinsic energy spread in the electron gun. Notice that ΔE and ΔV are different: ΔV depends on how well we can control the voltage supply whereas ΔE depends on our choice of electron source (see Chapter 5). If we neglect any other contributions to the envelope function, then we can define an information limit due to instrument chromatic aberrations by ρ_c

$$\rho_c = \left(\frac{\pi\lambda\delta}{\sqrt{2\ln(s)}}\right)^{\frac{1}{2}} \tag{28.51}$$

where e^{-s} is the cut-off value for the envelope. If we take $\ln_e s$ to be 2, (s is not the deviation parameter) then

$$\rho_c = \left(\frac{\pi\lambda\delta}{2}\right)^{\frac{1}{2}} \tag{28.52}$$

The source-dependent envelope function is new because, until we have a, FEG, we don't usually consider the 'source of a probe.' If we imagine that the source has a Gaussian distribution, we have an envelope function $E_s(\mathbf{u})$ given by

$$\begin{aligned} E_s(\mathbf{u}) &= \exp\left[\left(\frac{\alpha}{2\lambda}\right)^2\left(\frac{\partial\chi(\mathbf{u})}{\partial u}\right)^2\right] \\ &= \exp\left[-\left(\frac{\pi\alpha}{\lambda}\right)^2\left(C_s\lambda^3u^3+\lambda u\right)^2\right] \end{aligned} \tag{28.53}$$

Here α is the semi-angle characterizing the Gaussian distribution. What this equation tells us is that if α is too large ($\leq$1 mrad) it can limit the information limit. If we say that u must lie between u and some maximum value u_{max}, we can maximize the argument of the exponential in equation 28.53 to give an optimum focus

$$\Delta f_{opt} = -\frac{3}{4}C_s\lambda^2u_{max}^2 = -\frac{3}{4}\frac{C_s\lambda^2}{\rho_i^2} \tag{28.54}$$

In this equation ρ_i is the information limit of the microscope (because of how we chose u_{max}). This defocus value will be important later when we discuss holography in a FEG TEM. The two curves shown in Figure 28.10 illustrate graphically how this envelope function varies within Δf. It can also be optimized by decreasing the angle α. With a little more manipulation, de Jong and Van Dyck show that the information limit due to the limited coherence of the source is given by

$$\rho_\alpha = \left(\frac{6\pi\alpha a}{\lambda\sqrt{\ln(s)}}\rho_s^4\right)^{\frac{1}{3}} \tag{28.55}$$

The envelope functions for the drift/translation and vibration of the specimen represent a new method for

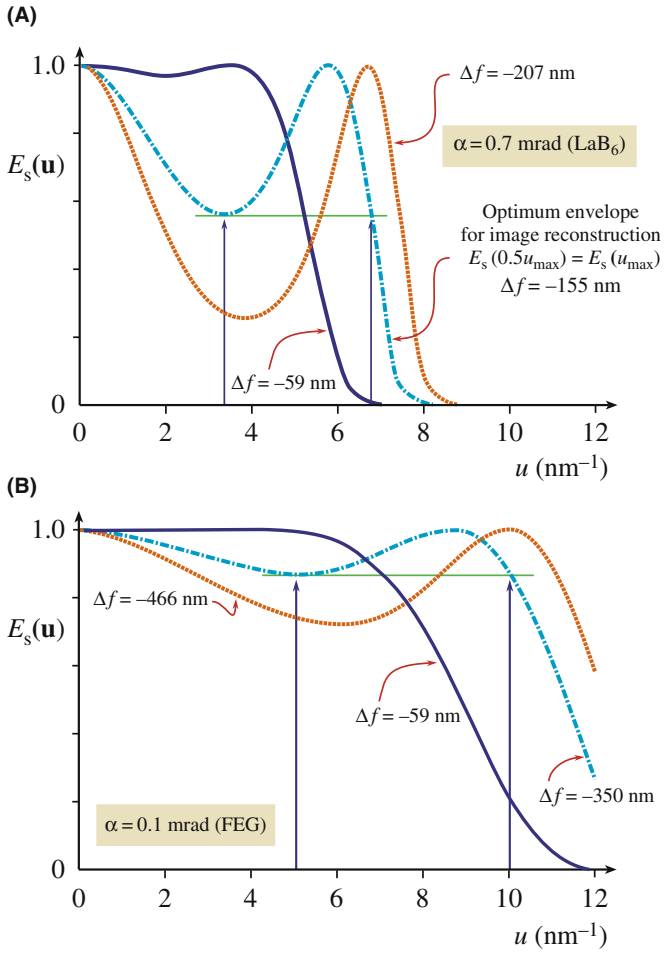


FIGURE 28.10. Variations in the envelope function, $E_s(\mathbf{u})$, for different objective lens defocus: (A) LaB_6 source, (B) FEG.

taking account of these two unavoidable quantities. We'll just quote the results for the two 'information limits' which are the crossover values for the two envelope functions $E_d(\mathbf{u})$ and $E_v(\mathbf{u})$

$$\rho_d = \frac{\pi d}{\sqrt{6\ln(s)}} \tag{28.56}$$

and

$$\rho_v = \frac{\pi v}{\sqrt{\ln(s)}} \tag{28.57}$$

In these equations, d is the total specimen drift during the exposure, t_{exp}, so $d = v_d t_{exp}$ for a drift velocity v_d; v is the amplitude of the vibration.

The detector envelope function, $E_D(\mathbf{u})$, is something we never worried about with film, but CCD cameras have a limited number of pixels, i.e., we only have a limited number of resolved image points. This envelope function results from two effects

- Delocalization of the information in the image.
- The finite pixel size.

DELOCALIZATION

Image delocalization depends on **u**. It is large when $\partial\chi(\mathbf{u})/\partial u$ oscillates rapidly as it does for large **u**, i.e., where we are placing the information limit.

The idea is simple but the math is more difficult. Let's assume that the image which is actually captured by the camera is circular. We can then say that if R is less than R_w, the radius of the window, then we capture the information; if it's greater than R_w, we don't. So the CCD detector is acting like an aperture! Now de Jong and Van Dyck show how u_{max} is related to R_w

$$\alpha C_s \lambda^3 u_{max}^3 = R_w \tag{28.58}$$

The important result is that the delocalization of the information in the image must be less than the half-width of the CCD detector array.

The value of R_w is related to the number of pixels, N, and their size, D

$$R_w = \frac{1}{2} ND \tag{28.59}$$

The information limit due to the detector (i.e., the cross-over values of the detector envelope functions $E_D(\mathbf{u})$) is

$$\rho_D = \left(\frac{12\sqrt{2}\pi a}{N\sqrt{\ln(s)}} \right)^{\frac{1}{4}} \rho_s \tag{28.60}$$

Clearly we can decrease ρ_D by increasing N, but not quickly. With this analysis in mind, we can summarize the conditions necessary for ρ_i to be limited by chromatic aberration.

$$\alpha \leq \frac{\lambda}{6\pi a \rho_s} \left(\frac{\rho_c}{\rho_s} \right)^3 \tag{28.61}$$

$$N \geq 12\sqrt{2}\pi a \left(\frac{\rho_s}{\rho_c} \right)^4 \tag{28.62}$$

$$d \leq \frac{\sqrt{6}}{\pi} \rho_c;\ 0.8\rho_c \tag{28.63}$$

$$u \leq \frac{1}{\pi} \rho_c;\ 0.3\rho_c \tag{28.64}$$

Table 28.1 gives some numerical examples of what these equations mean.

To see whether we will ever reach the information limit, we have to consider the effect of the noise. We know that the signal-to-noise ratio is proportional to $\beta^{1/2}$, where β is the brightness of the electron gun. The smallest image element we need to examine has an area ρ_i^2. Then the background signal I_0 is given by

$$I_0 = D\rho_i^2 = \beta\pi\alpha^2 t \rho_i^2 \tag{28.65}$$

where D is the electron dose, α is the angle of convergence, and t is the time.

For white noise, the noise in an element will be related to $I_0^{1/2}$. The total contrast in our small pixel can be written as $DEF\rho_i^2$ where D is the dose, E is the envelope function, F is the structure factor, and ρ_i^2 is the area of the pixel. Now we can say that the minimum detectable signal-to-noise ratio is k, which gives

$$DEF\rho_i^2 = kI_0^{1/2} = k\rho_i D^{1/2} \tag{28.66}$$

Therefore for $\alpha = 1$ mrad and $t = 1$ second, we can express the signal-to-noise ratio as

$$s_0 = 443\rho_i \frac{F}{k} \beta^{1/2} \tag{28.67}$$

Now you can use some real numbers: take $k = 2$ (think what this means for the minimum contrast), and assume that β is 10^{10} A/m^2 sr for a LaB_6 gun and 10^{13} A/m^2 sr for a Schottky FEG. You can show that for $\rho_i = 0.15$ nm (a LaB_6 gun) $\ln_e s_0$ is 1.2–2.2 whereas for a FEG, $\rho_i = 0.1$ nm and $\ln_e s_0$ is 4.5–5.2. You can also appreciate why s_0 depends on your material: low atomic numbers mean weak scattering. For our last two equations we'll again quote de Jong and Van Dyck. We can deduce optimum values for both the angle of convergence α and the exposure t by differentiating the envelope equations.

$$\alpha_{opt} = \frac{1}{k_s\sqrt{2}} \left(\frac{\rho_i}{\rho_c} \right)^3 = \frac{1}{6\pi a\sqrt{2}} \frac{\lambda}{\rho_s} \left(\frac{\rho_i}{\rho_s} \right)^3 \tag{28.68}$$

$$t_{opt} = \frac{1}{2k_d} \left(\frac{\rho_i}{\rho_c} \right) \simeq 0.39 \left(\frac{\rho_i}{v_d} \right) \tag{28.69}$$

TABLE 28.1. Maximum Convergence Angle α and Minimum Number of Unusable Image Points N for Different Values of the Point Resolution to (Chromatic) Aberration Limit Ratio (ρ_s/ρ_c)

	α (mrad)		N (pix)	
ρ_s/ρ_c	ε_0	ε_{opt}	ε_0	ε_{opt}
1	0.58	2.3	53	13
1.5	0.17	0.69	270	67
2	0.07	0.30	853	213
2.5	0.04	0.15	2082	521
3	0.02	0.09	4320	1080

ε_0: Gaussian focus, ε_{opt}: optimum focus; $\lambda = 0.011\ \rho_s$ (de Jong and Van Dyck 1993).

Notice that α_{opt} depends not only on ρ_s and ρ_i, but also on λ (of course ρ_s and ρ_i also depend on λ) and that t_{opt} only depends on the drift rate; fortunately v_d will never be zero!

We can now summarize these new concepts

- Microscopy is much more complex when you try to use the information limit rather than the Scherzer limit!
- If you want to use a computer, the size of the CCD camera will also affect the actual information limit; this is the effect of $E_D(\mathbf{u})$.
- Drift and vibrations must be minimized or they will determine your resolution; these contributions were described by $E_d(\mathbf{u})$ and $E_v(\mathbf{u})$.
- When everything else is perfect, your resolution will be controlled by the signal-to-noise ratio of the detector and the coherence functions, $E_D(\mathbf{u})$ and $E_c(\mathbf{u})$.
- An FEG improves the information limit because of the large increase in the brightness, β. This increase allows us to decrease α, increase the dose, and increase the signal-to-noise ratio.

This section has been quite long and not what you traditionally find in a chapter on HRTEM. This is quite deliberate. The TEM produces a great deal of information but many researchers want numbers, quantities, not data encoded in images. So we need to convert our data into numbers; we have to be quantitative.

MORE DELOCALIZATION

To fully interpret an HRTEM image, you must understand the implications of this concept. See the papers by (i) Coene and Janssen and (ii) Lichte.

28.14 SOME DIFFICULTIES IN USING AN FEG

We've discussed the advantages of using an FEG for HRTEM, but there are some practical difficulties which have been analyzed by Otten and Coene at FEI. A cold FEG (CFEG) allows us to extract a very high current per unit area, but the total area of the emitting region is very small so that the extraction current is <5 nA. This current can be increased if we thermally assist the field emission by heating the Schottky emitter to ~1500°C. It gives the same high brightness, but a larger maximum current because of the larger emitting area. So what are the difficulties?

- The emitter area may be so small that we have to 'fan' the beam in order to illuminate the area used in TEM. This fanning may actually increase the effect of coma aberration (a radial aberration as noted in Section 28.10). If a CFEG has a source size of ~3 nm, we can study ~15 nm with a 5× magnification. The Schottky source has a source diameter ~10× greater, and the price you pay for this is a decrease in spatial coherence, and a larger energy spread.
- Correcting astigmatism is very tricky with an FEG. As shown in Figure 28.11, if the image is astigmatic you'll see that at all defocus settings with an LaB_6 source. In an FEG, when astigmatism is present, all the images look similar and you can't use the technique of finding the minimum-contrast defocus (at ~0.4 sch) to determine Δf. If you try to use the wobbler to do coma-free alignment, that fails too, because you can't interpret the focus difference between two FEG images for the two wobbler directions. There is a solution to finding Δf, fortunately; either use on-line processing (Chapter 31) or converge the beam! The latter way deteriorates the spatial coherence and you've made your $2 M FEG behave like an old $200 K LaB_6 machine.
- Focal series of images are a challenge, because you can now use a very large range of Δf values, and it becomes a major task just to determine your value of Δf.
- Image delocalization occurs when detail in the image is displaced relative to its 'true' location in the specimen. The effect is emphasized by the graph shown in Figure 28.12 and becomes worse as you go away from Scherzer defocus. The effect is illustrated in Figure 28.13, where fringes from the gold particles can appear outside the particle. If we rewrite equations 28.36, we can express the delocalization as

$$\Delta R = \lambda u\left(\Delta f + C_s\lambda^2 u^2\right) \tag{28.70}$$

You may notice a similarity between this equation and that for the SAD error (Chapter 11): there is a good reason for this similarity. Two values have been proposed for Δf_{opt}, the optimum defocus setting to minimize delocalization. They give an optimum value for the defocus of

$$\Delta f_{opt} = -MC_s\lambda^2 u_{max}^2 \tag{28.71}$$

where M is a factor between 0.75 and 1. The value for ΔR_{min} is then close to

$$\Delta R_{min} = \frac{1}{4}C_s\lambda^3 u_{max}^3 \tag{28.72}$$

The actual value of M is determined by where you define the cut-off value for $\mathbf{u}$. There are three important conclusions on delocalization

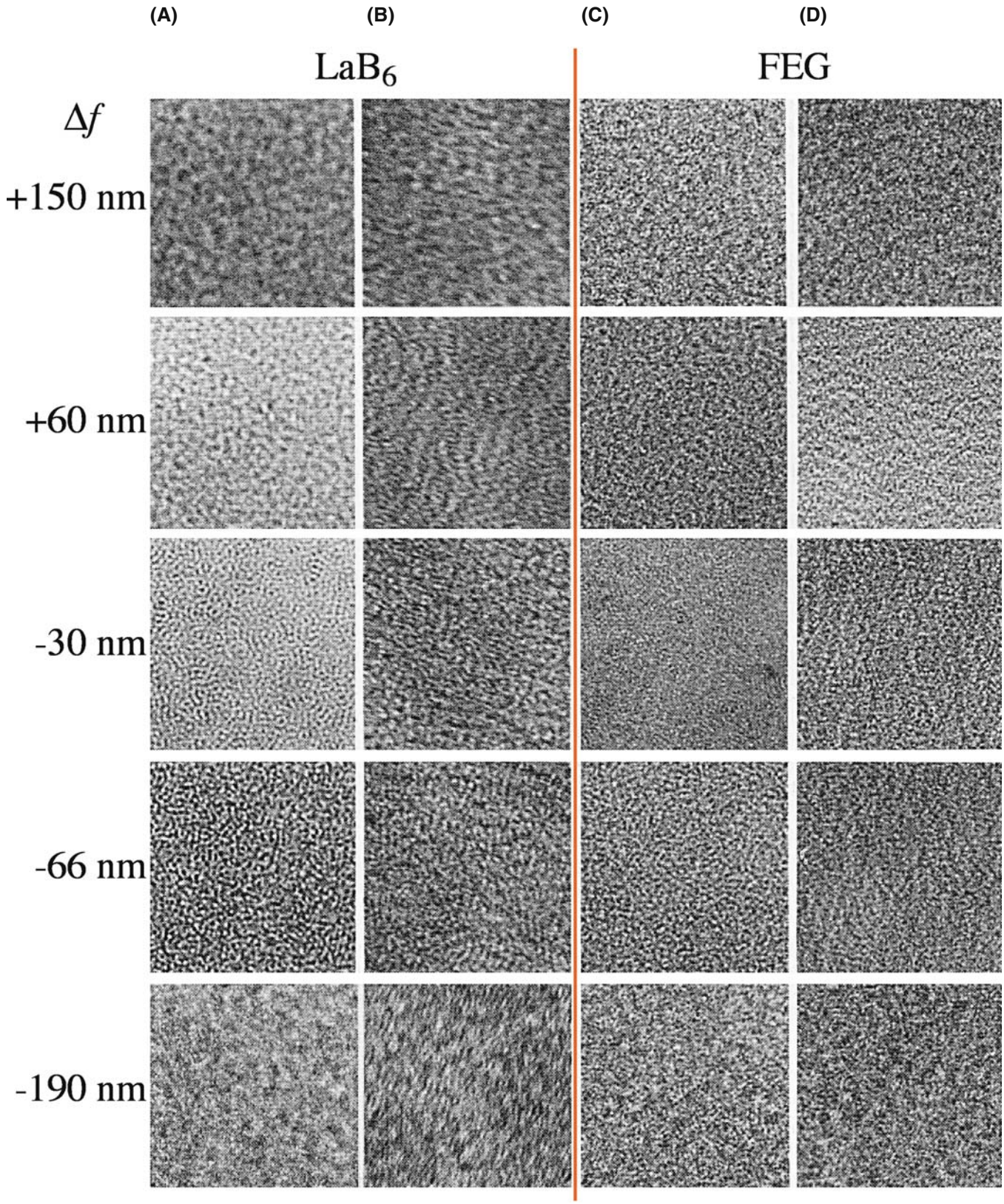


FIGURE 28.11. A tableau of images from an amorphous film at varying defocus. (columns A and B) LaB_6; (columns C and D) FEG. (columns A and C) Without astigmatism; (columns B and D) with astigmatism. With LaB_6 you can easily see the astigmatism, while with an FEG, you can't.

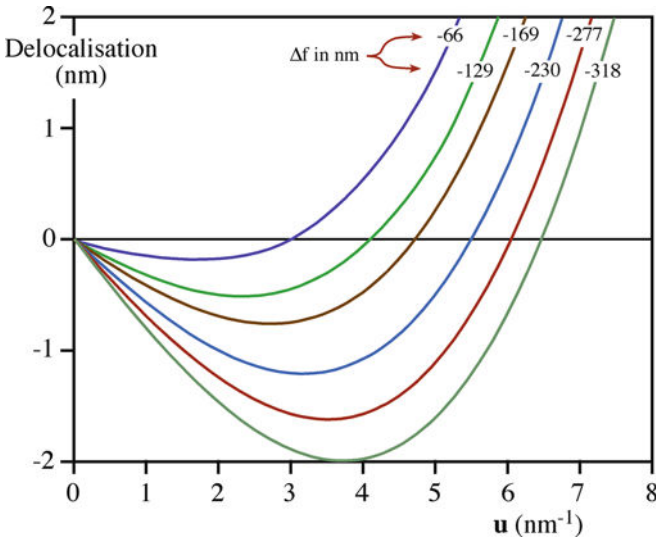


FIGURE 28.12. Image delocalization plotted against **u** as Δf is changed for a Philips CM20 FEG with $C_s = 1.2$ nm.

- As C_s decreases, delocalization decreases.
- As λ decreases (accelerating voltage increases), delocalization decreases.
- Delocalization cannot be avoided in an FEG, except by greatly reducing C_s!

28.15 SELECTIVELY IMAGING SUBLATTICES

In Chapter 16, where we discussed ordered intermetallic alloys, we saw that many materials with a large-unit cell are closely related to a material with a smaller unit cell. If the two structures don't have the same symmetry, then the two unit cells can show several different orientation relationships, as was the case for vanadium carbide.

We can use this information to form different high-resolution images instead of different DF images. Two [001] DPs from an ordered alloy of Au_4Mn are shown in Figure 28.14, together with a schematic of one pattern. Two domains are present in the combined pattern. Both patterns have fourfold symmetry, but they are rotated relative to one another. If we use the DF lattice-imaging mode and exclude all the fcc reflections using the objective aperture, we form an image like that shown in Figure 28.15. The two variants are not only easily recognized, but we know where they are with an accuracy of atomic dimensions. If you're used to grain-boundary theory, the original cell has become the coincident-site lattice (CSL) in reciprocal space and the two sublattices are like grains related by a small Σ. This approach has been used to estimate the size of very small particles of $NiFe_2O_4$ spinel which are completely contained in a matrix of NiO, as illustrated in Figure 28.16. The lattice

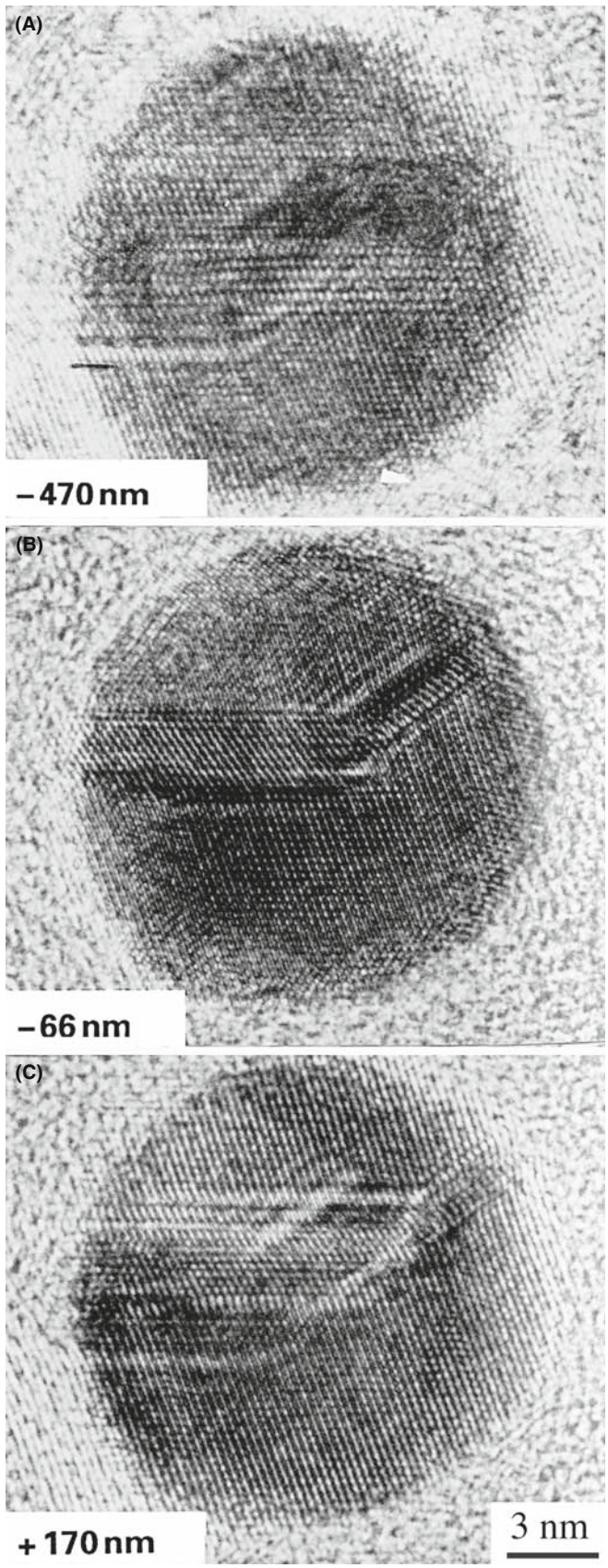


FIGURE 28.13. Experimental images showing delocalization in HRTEM images of an Au particle: (A) underfocus, (B) Scherzer focus, (C) overfocus.

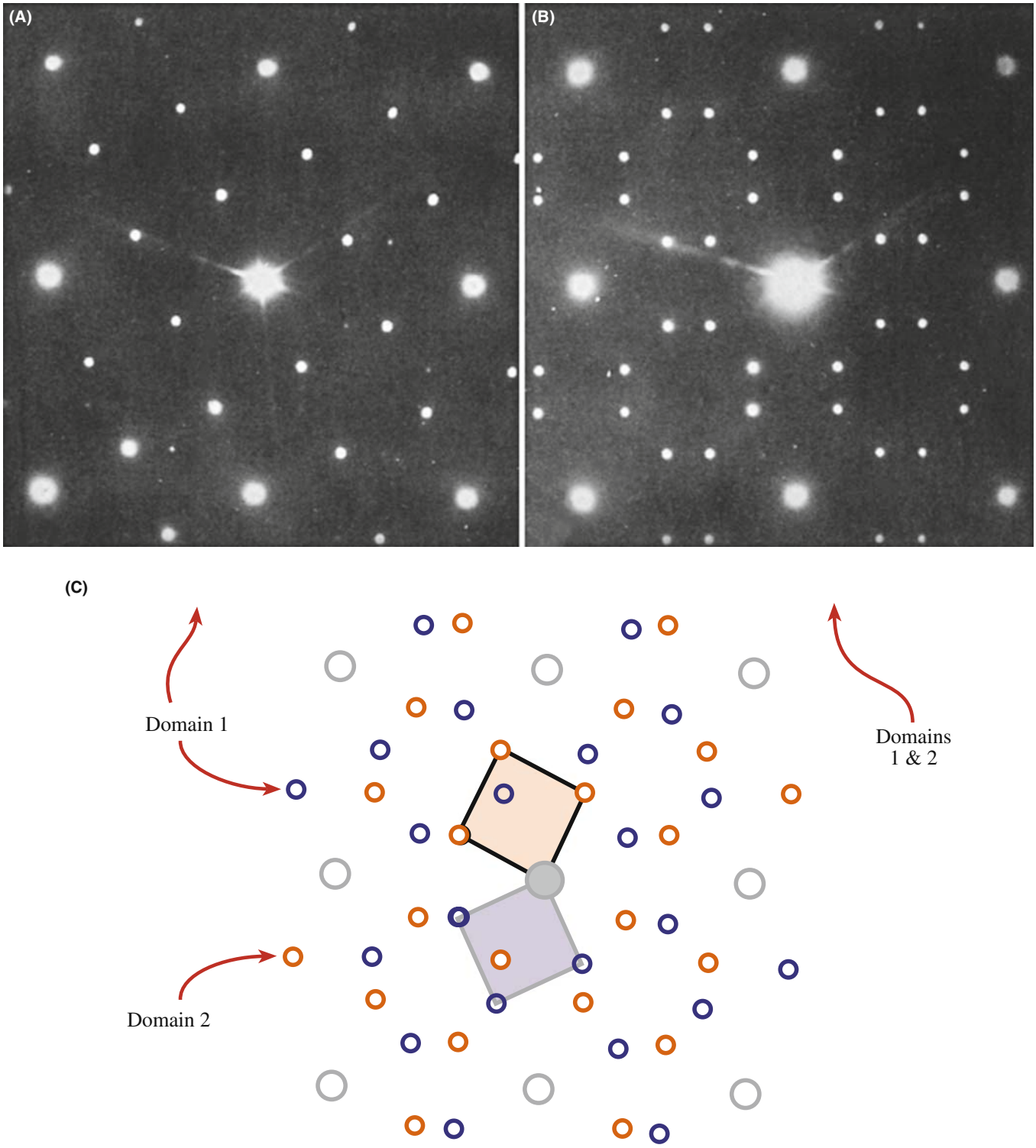


FIGURE 28.14. Two [001] DPs from an ordered alloy of Au_4Mn with schematics. (A) One domain. (B) Two symmetry-related domains. (C) Schematic diagram showing how (B) arises from the relative rotation of the two domains.

parameter of the spinel is twice that of the NiO but the latter is generally above and below the particle. This approach is therefore quite difficult, especially if, as in the analysis of Figure 28.16, the shape of the particle is important. Then you need to resort to simulation and processing, as we'll discuss in Chapter 30.

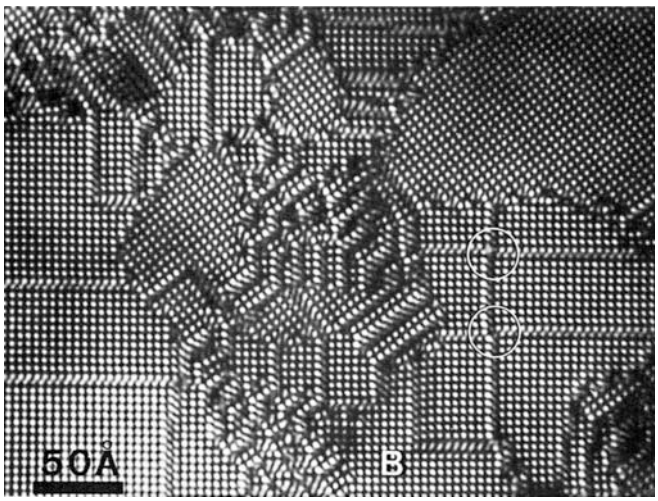


FIGURE 28.15. DF lattice image of Au_4Mn using an objective diaphragm to exclude all the fcc reflections. The two differently oriented domains correspond to the two orientations in Figure 28.14C.

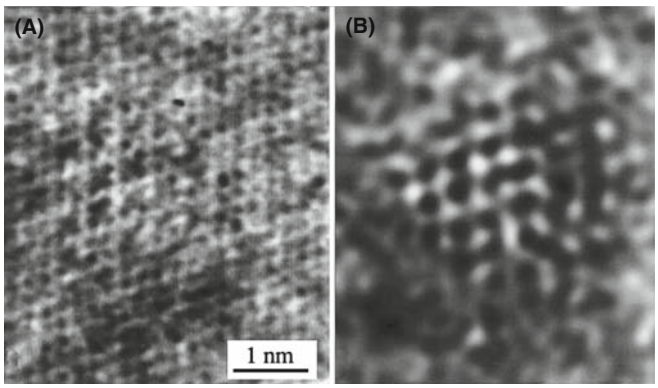


FIGURE 28.16. (A) The experimental image of a small spinel particle in a NiO matrix. The NiO is thicker and dominates the image. (B) After filtering out the NiO contribution (its lattice parameter is twice that of the spinel), we can see the spinel particle and estimate its size. See also Figures 31.2 and 31.3.

28.16 INTERFACES AND SURFACES

Interfaces of all kinds have been extensively studied by HRTEM. Of course, we want the near-atomic resolution. Sometimes it's because they make ideal subjects for study! Point defects require extensive image processing and simulation, dislocations tend to move, but interfaces seem to remain stationary if you're careful. However, we are always limited as to which interfaces we can study.

The fundamental requirement is that the interface plane must be parallel to the electron beam.

If a low-index plane in one grain (but preferably in both grains) is parallel to the interface, you're in business. The problem is that we are rarely sure that this is the case, but because you are looking at a very thin specimen, the projected width of even a slightly tilted boundary is small.

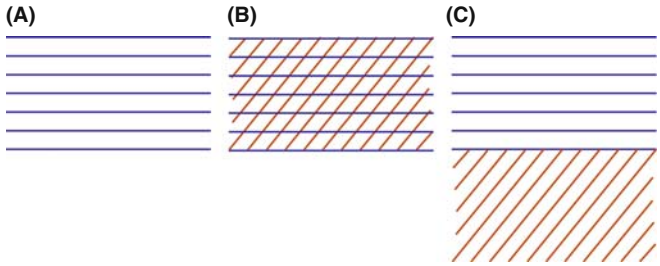


FIGURE 28.17. Schematic HRTEM images of grain boundaries showing (A) one set of fringes in one grain; (B) specimen tilted to give crossing fringes in one grain; (C) one set of fringes in each grain boundary remains parallel to the beam.

So you can tilt the specimen to look down a pole in one grain or to make the beam parallel to a low-index plane in the second grain as shown in Figure 28.17.

If you are lucky (and we often are because we only study tilt boundaries by HRTEM), you can produce crossed fringes in both grains. A selection of images is shown in Figure 28.18. Here you can see structured boundaries, boundaries with an amorphous layer between the grains, interfaces between two different materials, and a surface profile image. We can make some general comments about these images

- Even a 'low'-resolution, lattice-fringe image gives you information on the local topology of your interface.
- If the layer of amorphous material in the boundary is quite thick (>5 nm), you can see it directly.
- You can quite easily see detail like five-membered rings in grain boundaries but you should be wary of interpretation until you've covered Chapter 30.
- You can see abrupt interfaces at near-atomic dimensions.

Now we can also list our concerns

- Has grooving at the interface affected the appearance of the image? The answer is "maybe, but does it affect what you wanted to learn?"
- Is the phase boundary as abrupt chemically as it is structurally? It is very difficult to answer this. The appearance of the image changes at the interface in Figure 28.18C, mainly because the total number and location of the cations (Fe^{3+} and Ni^{2+}) changes, not because there is a 2:1 ratio of Fe:Ni.
- Are all of the black spots in Figure 28.18D complete columns? The next question is "Complete columns of what?"

We will address some of these problems in Chapter 30 but we can make some comments now.

- The quality of your imaging data will be governed by how well you prepare your specimen. Nearly all subsequent analysis will assume that it has a uniform

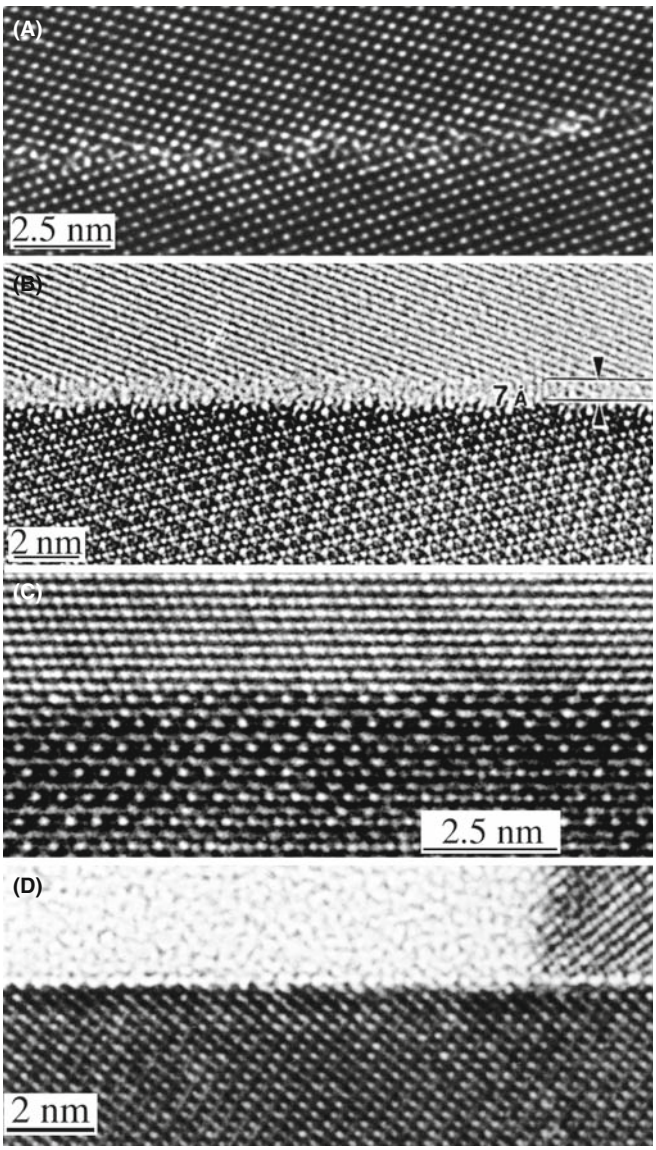


FIGURE 28.18. Examples of HRTEM images of planar interfaces. (A) Grain boundary in Ge; (B) grain boundary in Si_3N_4 with a layer of glass along the interface; (C) phase boundary separating NiO and $NiAl_2O_4$; (D) profile images of the (0001) surface of Fe_2O_3.

thickness across the interface. If you do not *know* that this is so, then your interpretation may and should be questioned.

- Crystalline grains also thin at different rates if they have different orientations or different structures or different chemistries. The grain-boundary layer, whether crystalline or amorphous, will also thin at a different rate. Why? Because the bonding and density are different. So careful specimen preparation is absolutely critical.
- You can learn a lot about your interface using HRTEM without trying to use atomic-resolution imaging.
- The longer you look at your specimen, the more it will differ from what you started with. Use at least a pseudo-low-dose approach, if possible. Figure 28.19 illustrates an extreme example. Here, the oxide has been completely reduced to the metal at the edge of the foil. Of course, this now provides a method for studying the reduction of oxides under the electron beam in the presence of hydrocarbons and in a good vacuum!

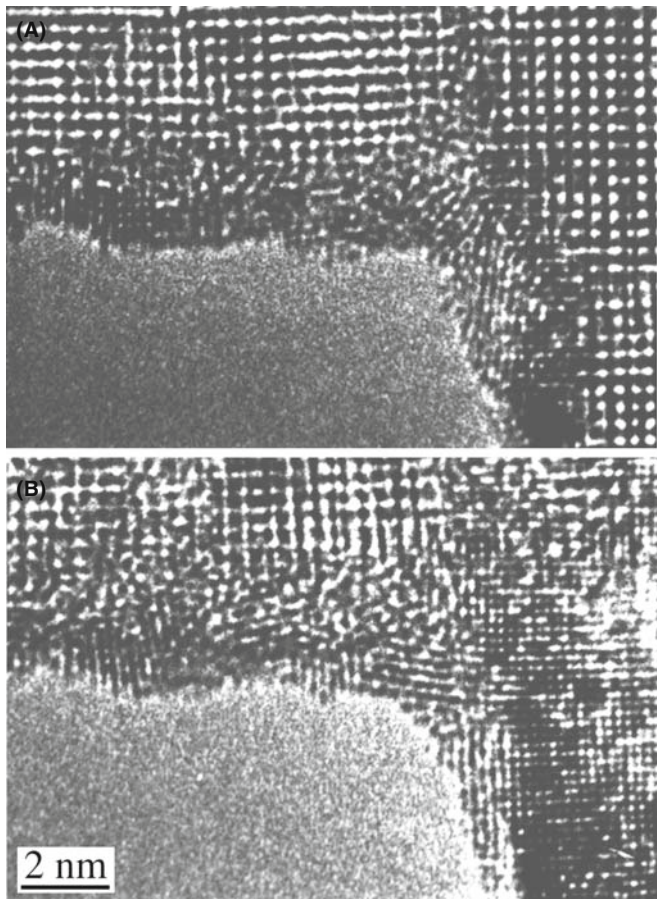


FIGURE 28.19. Reduction of Nb oxide to the metal by beam-induced loss of oxygen during observation of the edge of the foil. The reduction increases with time.

28.17 INCOMMENSURATE STRUCTURES

We'll illustrate this topic by considering several types of incommensurate (modulated) structures. In each case, the structure consists of a 'parent' structure to which we then add a periodic modulation by means of an internal planar defect. Van Landuyt et al. have characterized three different types of incommensurate structures.

- Periodic modules of the parent separated by interfaces. The interface may be a stacking fault (SF), twin boundary (TB), anti-phase boundary (APB), inversion domain boundary (IDB), crystallographic shear (CS) plane, or discommensuration wall.
- A parent structure with a superimposed periodic deformation wave with a larger periodicity.

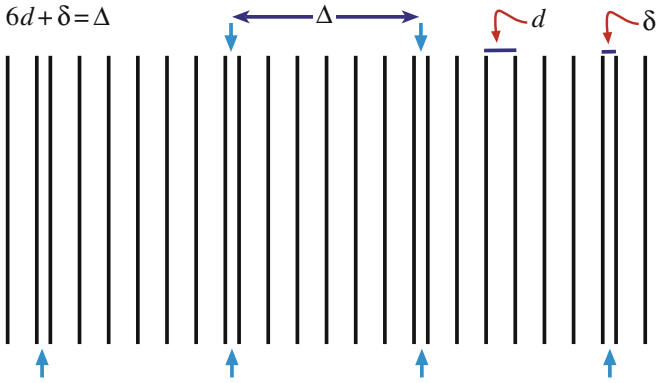


FIGURE 28.20. An incommensurate structure formed by inserting a planar defect after every seventh layer to expand the lattice by δ every seventh plane.

- A parent structure where the composition or site occupancy changes periodically.

The next complication is that we can find commensurate and incommensurate structures, and also structures where the modulation is variable. To understand how a structure can be incommensurate, consider Figure 28.20, where we've placed a planar defect after every seventh layer so that it expands the lattice by δ every seventh plane. The parent lattice will show a spot spacing in the DP proportional to d^{-1} but the 'superlattice' will have a periodicity of Δ^{-1}, which means that we need not have a simple relationship between the two arrays of spots.

These different kinds of modulation can be combined! We'll illustrate this type of specimen with two examples. The Bi-Sr-Ca-Cu-O superconductor provides a good illustration of this type of structure. The parent is a perovskite-like cube; we may have two, three, or four layers of the perovskite with each group separated from the next by a bismuth oxide layer. The formula can be written as $Bi_2Sr_2Ca_nCu_{n+1}O_{2n+\delta}$ so for $n = 1$ we have a sequence of layers (planes) described as BiO-SrO-CuO_2-Ca-CuO_2-SrO-BiO, as shown in Figure 28.21. The DPs depend on the particular value of n in the chemical formula and show rows of satellite reflections due to the modulation of the basic structure. When we form the HRTEM image, the lattice planes appear wavy although we can recognize an orthorhombic pattern. The wavy modulation in the image is probably due to excess oxygen in the BiO layers: the BiO layers don't fit very well to the perovskite block but the misfit stresses can be relaxed by introducing excess oxygen. There are several clear lessons from this example

- HRTEM is essential if we are to understand such structures.
- You need supplementary information, such as the chemistry of the specimen.
- Images of this structure produced with the beam along the orthogonal direction would be difficult to explain.

Modulated structures are not confined to the superconductors. In fact, such structures are ubiquitous in the materials world. Many useful engineering alloys exhibit spinodal decomposition and the spinodal wavelength can be directly measured from the extra spots in the DP. Many ceramics are described as polytypes, e.g., SiC or polytypoids, which are just polytypes with the composition fluctuating from layer to layer (e.g., SiAlONs). These structures consist of random or locally ordered stacking of specific atomic layers, which often give predictable effects in the DP. All these materials are particularly amenable to HRTEM analysis, because you can imagine the individual modulations, and still characterize the overall structure with conventional amplitude contrast.

28.18 QUASICRYSTALS

The study of quasicrystals continues to be a challenge for HRTEM, since these materials do not have the translational symmetry that we associate with crystals. However, they are strongly ordered, as you can appreciate from Figure 28.22. The HRTEM image shows many sharp white spots from a stable decagonal quasicrystal of Al-Mn-Pd. The DP from another specimen also shows very strong, clear, well-defined spots. In our earlier discussion of DPs, we associated each spot with a single set of planes, which were present throughout the specimen. Although the quasicrystals do not contain such planes, there is clearly far more order than in an amorphous material. You can indeed see that the spots in the HRTEM image are aligned in certain well-defined directions, but the spacing is difficult to identify. We have a growing understanding of these materials and it appears that the spots in the HRTEM image are this sharp because, at least in decagonal quasicrystals (but not in icosahedral ones), they really do correspond to columns; we don't need translational periodicity along the column. In fact we can rotate the quasicrystal (they can be grown as large as 1 mm) to reveal twofold and threefold axes, as illustrated in Figure 28.23. You can see how this might arise by looking along the rows of spots in Figure 28.22. We can draw some interesting lessons from the use of TEM to study quasicrystals

- HRTEM excels when materials are ordered on a local scale.
- For HRTEM, we need the atoms to align in columns because this is a 'projection technique,' but the distribution along the column is not so critical, and we

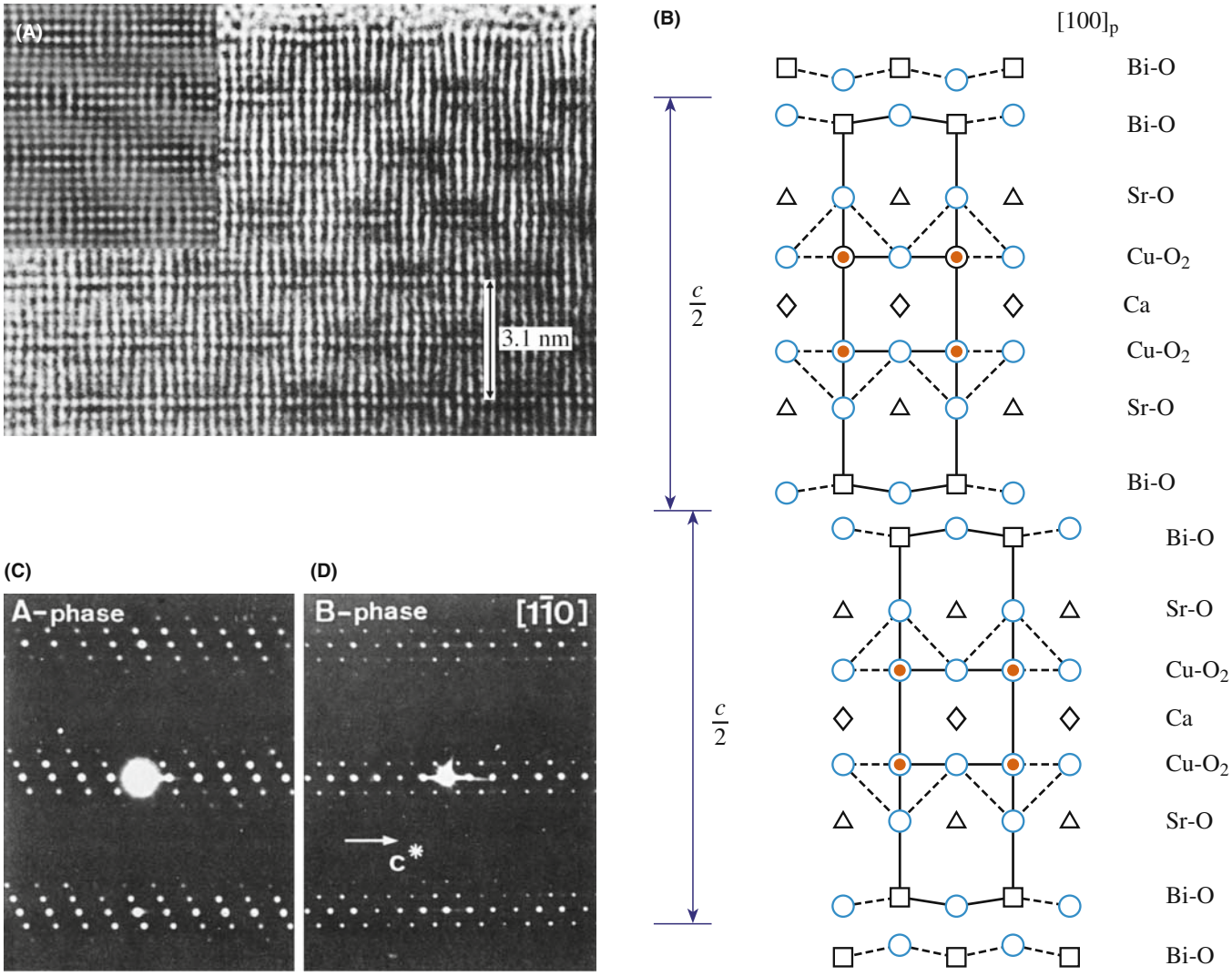


FIGURE 28.21. (A) HRTEM image of the superconductor $Bi_2Sr_2Ca_nCu_{n+1}O_{2n+\delta}$; this simulation (inset) assumes that the lattice relaxation occurs in the Bi-O layer. (B) For $n = 1$, the structure is built up from blocks which are shifted relative to one another. The DPs are from (C) the $n = 0$ phase and (D) the $n = 1.2$ phase. Notice that the spacings of the spots (the satellite sequence) are different.

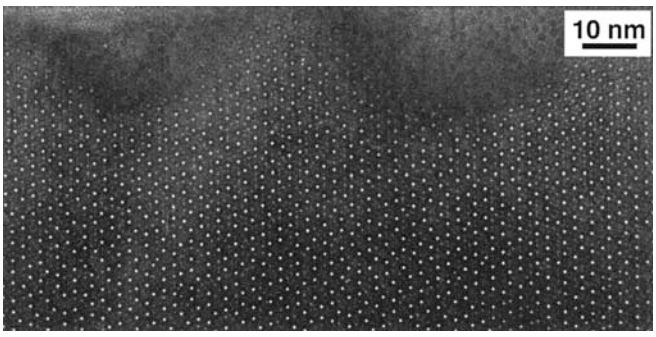


FIGURE 28.22. Tenfold symmetry in a decagonal Al-Mn-Pd quasicrystal.

can't determine it without tilting to another projection in the perfect crystal.

- SAD and HRTEM should be used in a complementary fashion.

28.19 SINGLE ATOMS

You have read that it has become possible to study materials at atomic resolution in the TEM quite recently. So you may be surprised to find that many groups have been reporting studies of individual atoms since about 1970! The techniques used include phase contrast and amplitude contrast in a conventional TEM, and (see Section 22.4) a dedicated STEM. Parsons et al. used mellitic acid molecules stained with uranyl ions from uranyl acetate so the atoms that were imaged are heavy. Parsons et al. then knew that the uranium atoms would be 1 nm apart at each apex of an equilateral triangle and they knew that there were 10^{13} of these per cm^2 supported on a thin (0.8 nm) film of evaporated carbon. One challenge is recognizing that the contrast from the individual uranium atoms reverses as you change defocus, just as we've seen for columns of

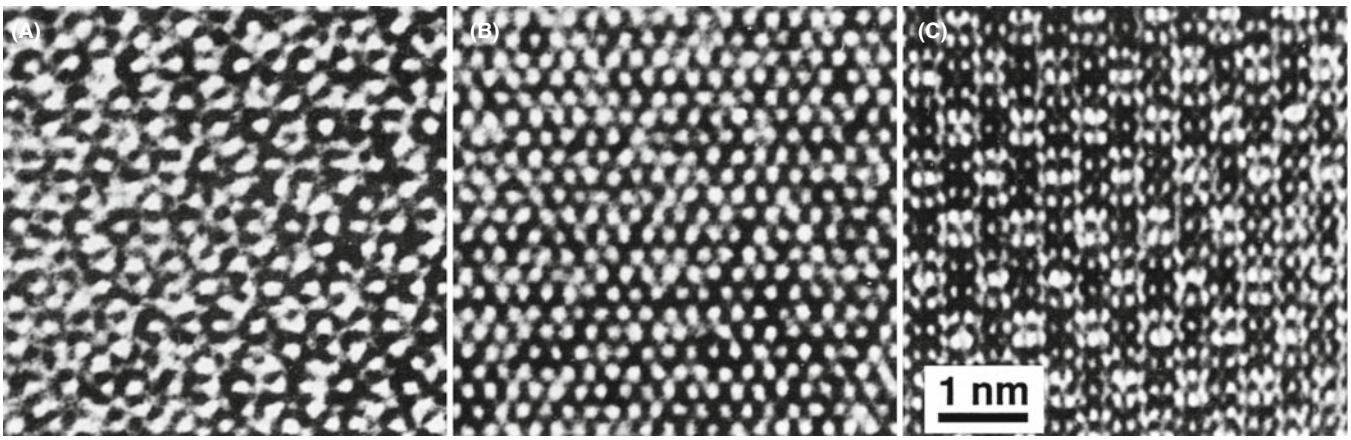


FIGURE 28.23. Images of (A) fivefold, (B) threefold, and (C) twofold projections of an Al-Cu-Li quasicrystal; Δf –27 nm.

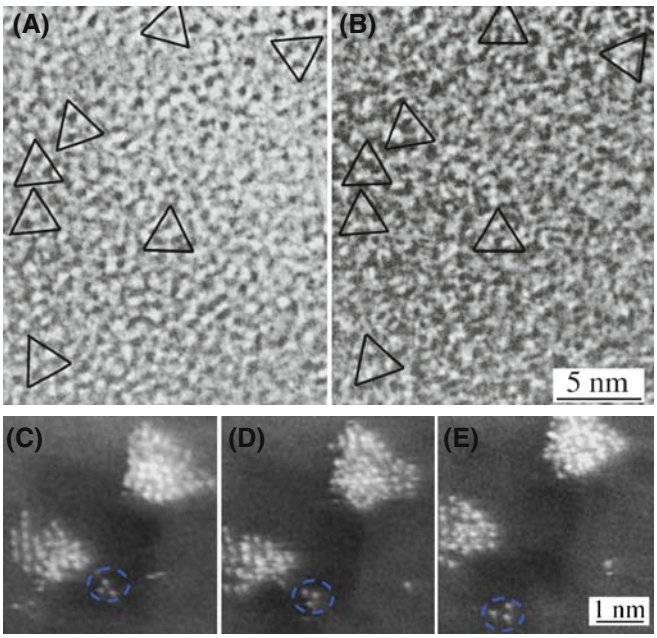


FIGURE 28.24. Images of atoms. (A, B) Triangular arrangements of uranium atoms at different values of defocus in conventional TEM. (C–E) STEM images of Pt/Fe catalyst supported on MgO; time between images is 10–15 seconds. These images have been processed to reduce noise and optimize visibility of the atoms.

atoms (and voids). You can see this effect in Figure 28.24A and B; for comparison, a series of *Z*-contrast STEM images is shown in Figure 28.24C–E. In this case the contrast is much higher and the movement of the cluster is clear.

Some points to notice

- This is a case where we really do have 'white atoms or black atoms!'
- Parsons et al. used a Siemens 101 TEM operating at 100 kV with a point-to-point resolution of about 0.33 nm; this is not today's state-of-the-art machine!
- The specimen in the TEM study was so stable that they could do 'through-focus' imaging.

Z-contrast imaging heralded the arrival of the STEM as a real research tool: atoms are now seen to move on the surface, agglomerate, etc. The imaging mode is essentially a high-angle dark-field technique so that the heavy atoms are by far the strongest scatterers and appear bright, as we discussed in Chapter 22. The difficulty with *Z*-contrast imaging in the STEM is that it requires an FEG, whereas almost any TEM operating today can produce images like those first demonstrated by Parsons et al.

CHAPTER SUMMARY

The major problem that separates this chapter from Chapter 23 is the language. In HRTEM, the language is that of physics or electrical engineering; you can get so involved with the language and the equations that you miss the point. Having said that, you must know the following terms and understand what they mean or don't mean

- Point-spread function.
- Contrast transfer function (CTF) (and transfer function).
- Weak-phase object approximation (WPOA).

With this understanding of the restrictions that our model involves, we can now consider the simulation of high-resolution images. If you want to delve further into the theory, we

recommend starting with the companion text, then going to the books by Cowley and Spence. You will need to have a strong background in math and physics to appreciate fully the further subtleties of more complex models. Always keep in mind that all of the above discussion was concerned with arriving at models or approximations

- We *model* the effect of the lens.
- Then we *model* the specimen.
- Finally, we combine the two models.

We will take the next two chapters to achieve these three tasks. Although by 2008 ~25 TEMs with C_s correctors had been installed, we leave the discussion of this exciting topic to the companion text since it is certainly not an instrument for someone learning TEM.

REFERENCES

One of the pioneers in the interpretation of HRTEM images was the late John Cowley. Much of our analysis of the specimen transfer function, $f(x,y)$, follows directly from his teaching. When pronouncing names, don't confuse Lord Rayleigh (born John William Strutt) with Walter Raleigh. Otto Scherzer was professor in Darmstadt and actually built an aberration corrector for his TEM. He was succeeded at Darmstadt by Harald Rose who with his former student, Max Haidar, made aberration correction work for the rest of us. Ondrej Krivanek and Nicolas Delby did the same for STEMs. Articles by Shannon and Weaver 1964, Van Dyck 1992 on information theory will start you on this topic; for HRTEM, you must then have access to John Spence's book.

ESSENTIAL FURTHER READING

Buseck, PR, Cowley, JM and Eyring, L Eds. 1988 *High-Resolution Electron Microscopy and Associated Techniques*, Oxford University Press New York.

Horiuchi, S 1994 *Fundamentals of High-Resolution Transmission Electron Microscopy*, North-Holland Amsterdam.

Spence, JCH 2003 *High-Resolution Electron Microscopy*, 3rd Ed., Oxford University Press New York.

NEW INSIGHTS

Haider, M, Müler, H, Uhlemann, S., Zach, J, Loebau, U and Hoeschen, R 2008 *Prerequisites for a Cc/Cs-Corrected Ultrahigh-Resolution TEM* Ultramicrosc. **108** 167–178.

Hawkes, PW 1980 *Units and Conventions in Electron Microscopy, for Use in Ultramicroscopy* Ultramicrosc. **5**, 67–70. An enjoyable and informative diversion. Includes definitions of the Glaeser and the Scherzer units (very non-SI).

Vladár, AE, Postek, MT, and Davilla, SD 1995 *Is Your Scanning Electron Microscope Hi-Fi?* Scanning **17**, 287–295. Looks at the SEM as a hi-fi instrument.

DELOCALIZATION

Coene, W and Janssen, AJEM 1992 in *Signal and Image Processing in Microscopy and Microanalysis*, Scanning Microscopy Supplement **6** (Ed. PW Hawkes), p. 379, SEM inc. O'Hare IL.

Lichte, H 1991 *Optimum Focus for Taking Electron Holograms* Ultramicrosc. **38**, 13–22.

INFORMATION LIMIT AND INFORMATION THEORY

Van Dyck, D 1992 in *Electron Microscopy in Materials Science*, p 193, World Scientific, River Edge New Jersey. Gives more detail on the derivation of equations 28.19 and 28.55 and on information theory for the TEM.

Van Dyck, D and De Jong, AF 1992 *Ultimate Resolution and Information in Electron Microscopy: General Principles* Ultramicrosc., **47** 266–281. On the information limit—part 1.

de Jong, AF and Van Dyck, D 1993 *Ultimate Resolution and Information in Electron Microscopy. II. The Information Limit of Transmission Electron Microscopes* Ultramicrosc. **49**, 66–80. On the information limit—part 2.

Shannon, CE and Weaver, W 1964 *The Mathematical Theory of Communication*, University of Illinois Press Urbana IL. Early text on information theory.

MORE THEORY

Cowley, JM 1992 in *Electron Diffraction Techniques* **1**, (Ed. JM Cowley), p. 1, IUCr. Oxford Science Publication. More extensive discussion of equations 28.20 and 28.21.

Cowley, JM 1995 *Diffraction Physics*, 3rd edition, North-Holland, Amsterdam.

Fejes, PL 1977 *Approximations for the Calculation of High-Resolution Electron-Microscope Images of Thin Films* Acta Cryst. **A33**, 109–113.
Hall, CE 1983 *Introduction to Electron Microscopy*, Krieger New York. An early clear discussion of other lens aberrations.
Hashimoto H and Endoh H 1978 in *Electron Diffraction 1927–1977*, (Eds. PJ Dobson, JB Pendry and CJ Humphreys), p. 188, IoP, Bristol. The aberration-free focus paper.
Otten, MT and Coene, WMJ 1993 High-Resolution Imaging on a Field Emission TEM Ultramicrosc. **48**, 77–91. Some practical difficulties, including the use of the wobbler, when using FEGs.
Rose, H 1990 *Outline of a Spherically Corrected Semiaplanatic Medium-Voltage Transmission Electron Microscope* Optik **85**, 19–24. One of two early articles by the father of today's aberration correctors.
Rose, H 1991 in *High Resolution Electron Microscopy: Fundamentals and Applications*. (Eds. J. Heydenreich and W Neumann), p. 6, Institut für Festkörperphysik und Elektronmikroskopie, Halle/Salle, Germany.
Thon, F 1975 in *Electron Microscopy in Materials Science*, (Ed. U Valdrè) p. 570, Academic Press New York.

MATERIALS APPLICATIONS

Amelinckx, S, Milat, O and Van Tendeloo, G 1993 *Selective Imaging of Sublattices in Complex Structures* Ultramicrosc. **51**, 90–108.
Bailey, SJ 1977 *Report of the International Mineralogical Association (IMA)-International Union of Crystallography (IUCr) Joint Committee on Nomenclature* Acta Cryst. **A33**, 681–684. Definitions of polytypes, polytypoids, etc.
Butler, EP and Thomas, G 1970 *Structure and Properties of Spinodally Decomposed Cu-Ni-Fe Alloys* Acta Met. **18**, 347–365. Classic study of DPs from spinodal decomposition.
Carter, CB, Elgat, Z and Shaw, TM 1986 *Twin Boundaries Parallel to the Common-{111} Plane in Spinel* Phil. Mag. **A55**, 1–19. Early study of spinel showing that you don't always need the highest resolution. See also p21–38.
Nissen H-U, and Beeli, C 1991 in *High Resolution Electron Microscopy: Fundamentals and Applications*, (Eds. J Heydenreich and W Neumann) p. 272, Institut für Festkörperphysik und Elektronmikroskopie Halle/Salle Germany. HRTEM from quasicrystals (Figure 28.22).
Parsons, JR, Johnson, HM, Hoelke, CW and Hosbons, RR 1973 *Imaging of Uranium Atoms with the Electron Microscope by Phase Contrast* Phil. Mag. **29**, 1359–1368. Individual Pt atoms in 1973!
Rasmussen, DR, Summerfelt, SR, McKernan, S and Carter, CB 1995 *Imaging Small Spinel Particles in an NiO Matrix* J. Microsc. **179**, 77–89.
Van Landuyt, J, Van Tendeloo, G and Amelinckx, S 1991 in *High Resolution Electron Microscopy: Fundamentals and Applications,* (Eds. J Heydenreich and W Neumann), p. 254, Institut für Festkörperphysik und Elektronmikroskopie Halle/Salle Germany.

THE COMPANION TEXT

The companion text includes a complete chapter by Van Aert and Van Dyck on direct imaging methods for HRTEM. Aberrations are discussed by Kruit (on lenses) and Haider (on aberration correction). Since HRTEM is changing rapidly, the companion text is our way of keeping more current.

SELF-ASSESSMENT QUESTIONS

Q28.1 If $f(\mathbf{r})$ describes the specimen function and $g(\mathbf{r})$ describes the image, write down the relationship between the two.
Q28.2 The function $h(\mathbf{r})$ is given two names in the text, state the two names and explain why each is used.
Q28.3 Describe in 30 words or less how a TEM is like a radio.
Q28.4 $H(\mathbf{u})$ is described as a product of three or more terms. What are they and why do we use the product?
Q28.5 Write down an expression for $\chi(\mathbf{u})$.
Q28.6 Justify the definition of the term 'overfocus.'
Q28.7 In the weak-phase object approximation, the electrons experience a phase change in passing through a sample thickness dz. Write down an expression for the phase shift.
Q28.8 Write down Cowley's expression for the transfer function of a weak-phase object with absorption.
Q28.9 Give a numerical value for the maximum thickness for $Ti_2Nb_{10}O_{27}$ in the WPOA.
Q28.10 Write down an expression defining $T(\mathbf{u})$, the objective lens transfer function.
Q28.11 What is the problem with using the term 'contrast transfer function' in TEM?
Q28.12 A point in the specimen is imaged as a disk of radius $\delta(\theta)$ in the image. Write down an expression for $\delta(\theta)$ and use it to deduce an expression for $\chi(\mathbf{u})$.
Q28.13 Sketch the graph of $T(\mathbf{u})$ versus $\mathbf{u}$ with and without C_c.
Q28.14 Write down an equation for Scherzer defocus and explain the sign of Δf.

Q28.15 Write down an expression for u_{Sch} and hence write an equation for r_{Sch}.
Q28.16 What is coma-free alignment?
Q28.17 What is the special feature of the Rose corrector?
Q28.18 What is image delocalization and why is it so important in an FEG TEM?
Q28.19 Atoms were seen in ~1972. How can this be true if the resolution of the TEM used was only 0.33 nm?
Q28.20 Can we image incommensurate structures by HRTEM? Explain and justify your answer.

TEXT-RELATED QUESTIONS

T28.1 Explain fully why the WPOA is only valid if the specimen thickness is <0.6 nm for $Ti_2Nb_{10}O_{27}$.
T28.2 Explain fully the connection between equations 28.29 and 28.34.
T28.3 In Section 28.7, we say "and χ is –120°." Why do we say this?
T28.4 Explain why it is important that the TEM is a linear system.
T28.5 Choose two TEMs, one 100 kV from 1970 and the other 300 kV from 2008. Plot χ for each using equation 28.34 and Excel (or similar).
T28.6 Consider a group of TEMs with C_s = 3, 2.8, 2.6,. . .,0.6 mm (or just one with variable C_s) operating at 200 kV. Plot the values for the resolution at Scherzer defocus.
T28.7 By delving into the literature, explain how equations 28.43 and 28.44 were deduced.
T28.8 Using equation 28.47 construct a table for d, varying ΔE from 0.1 to 1.0 eV in increments of 0.1 eV for TEMs operating at 100, 200, and 300 kV.
T28.9 Discuss (using the literature) why we write $E_s(\mathbf{u})$ as shown in equation 28.53.

29

Other Imaging Techniques

CHAPTER PREVIEW

Much of what we've discussed in the preceding imaging chapters is what we might call 'classical' TEM imaging. It began with BF and DF techniques and quickly expanded to include many beams. Diffraction contrast, phase contrast, and to a lesser extent, mass-thickness contrast are the mechanisms we use to characterize our specimens. We control the contrast by inserting the objective aperture, or a STEM detector, and excluding or collecting electrons that have been scattered by the different processes. However, there are variations to the standard ways in which we can extract more information from a TEM image; in this chapter, we'll present a brief overview of some of them. Most of these operational modes that we'll discuss here are somewhat esoteric and have rather specialized applications. Nevertheless, you should know that they exist because they may be just what you need to solve your particular problem. There's no importance to the order in which we go through the various modes, but we'll cover modifications to conventional parallel-beam TEM imaging as well as those techniques that require STEM and use some of the electron detectors we discussed in Chapter 7. It turns out, however, that the various procedures are often feasible in either TEM or STEM mode.

This is a bit of a potpourri of a chapter but the techniques are not, to our knowledge, gathered together in any other text. The descriptions will, of necessity, be brief but we'll reference suitable source material so you can follow up if you really want to try the technique for yourself.

29.1 STEREO MICROSCOPY AND TOMOGRAPHY

This section is an excellent example of why we've included this catch-all chapter. In the first edition of this textbook, stereo microscopy was an old technique only used by biologists and tomography was used in an MRI. Now tomography fills an entire chapter in the companion text. The principle is, of course, the same as stereo microscopy: record images from more than one direction to give a 3D view of the sample.

You've have realized by now that any TEM or STEM image is a two-dimensional projection of a 3D specimen and this is a fundamental limitation. Sometimes we can discern differences in diffraction-contrast images of certain defects, depending on whether the defect is intersecting the top or the bottom of the foil, but generally we lose the depth dimension. To regain this depth information we use stereo microscopy, but only for features showing mass-thickness or diffraction contrast. We cannot use stereo for phase-contrast imaging because the essential experimental step, tilting the specimen, changes the phase contrast and the projected potential of the specimen. So any stereo effect in the image is lost. You may need stereo microscopy if, for example, you want to know whether precipitates have formed on your specimen surface rather than in the interior, or if you want to see how dislocations are interacting with each other.

Stereo imaging works because your brain gauges depth by simultaneously interpreting signals from both your eyes, which view the same scene from slightly different angles (about 5°), giving a parallax shift. So in the TEM, if you take two pictures of the same area but tilted a few degrees relative to each other, then present the two images simultaneously to your brain using a stereo viewer, you'll see a single image in which the different depths of the features are apparent. In fact, some people are able to see the stereo effect without the aid of a viewer and some people are incapable of discerning the stereo effect at all.

To see in stereo, the two images should be separated by ~60 mm, but in practice it's often sufficient just to move the pictures relative to each other until your eye and brain seize on the effect.

STEREO RULES

You have to make sure that the field of view, the contrast, and the magnification in each of the two images are the same.

2D TOMOGRAPHY

The stereo technique of the future: the data is processed by a computer, not by your brain.

A couple of points are worth noting before we describe the method. First, if the features you want to observe show diffraction contrast, then the only way to maintain contrast is to tilt along a Kikuchi band, keeping both **g** and **s** fixed; so tilt while looking at the DP. This procedure almost invariably requires a double-tilt stage and may be difficult, or impossible, if your specimen is heavily deformed. If you just want to measure the foil thickness, any tilt is sufficient and contrast does not have to be maintained. Second, if you want to be pedantic, there is a right way and a wrong way to view the stereo images. You have to present the images in the same way that your eyes would see a scene, that is, the two images have to be correctly positioned, otherwise the brain will interpret depth the wrong way round. If you are trying to perceive the true surface topography (such as with SE images) using SEM or STEM images, then the choice of which image goes into the left eye and which into the right is crucial. (See any SEM text, such as that by Goldstein et al., for more details on stereo viewing.) Of course, for TEM images this difference is irrelevant. TEM applications are reviewed by Hudson and a whole set of related papers appears in that same issue of *Journal of Microscopy*.

So if you want to take a stereo pair, follow these steps

- Select the region of interest, making sure that the specimen is eucentric.
- Record an image (BF, DF, or WBDF, it doesn't matter, although usually the BF image is used).
- Tilt the specimen by at least 5° (much higher tilts give a larger parallax shift but it's more difficult to keep the focus and the diffraction contrast constant).
- Ensure that the whole field of view didn't move while you tilted. If it did, translate it back to its original position (using the beam stop as a point of reference if you wish). All features in the image will be shifted slightly relative to each other: it is this parallax shift which your brain interprets in stereo.
- If the area is now out of focus (it will be if you had to use the second, non-eucentric, tilt axis), refocus *using the specimen-height (z) control*; otherwise you'll change your image magnification. Obviously, computerized stages will help in this respect.
- Record another image.
- Develop or display the images and observe them under a stereo viewer.

Figure 29.1 shows a pair of BF images showing precipitates. If the images are correctly spaced and you look through a stereo viewer then you should be able to see the relative depth of the precipitates. You can purchase cheap cardboard stereo viewers through any EM supplier. Although a proper stereo viewer is an expensive optical tool, you can use it to calculate the relative depth (Δh) of a feature in a stereo pair, since

$$\Delta h = \frac{\Delta p}{2\,M \sin\frac{\phi}{2}} \tag{29.1}$$

where Δp is the parallax shift between the same feature in the two images tilted by ϕ at a magnification M. Be careful how you define ϕ because some microscopists define the tilt angle as $\pm\phi$, in which case $\sin\phi/2$ becomes $\sin\phi$. For true depth determination you need to deposit some recognizable feature on the surface, such as gold islands, but that is not usually important in TEM images and relative depth is often sufficient. If you need quantitative stereo measurements, you can either learn about the field of stereology, which is an old established discipline or you can learn the tomography technique.

We will discuss electron tomography in the companion text. This is a very large topic and an emerging activity in TEM. It has been limited until recently by the need for fast computers and large amounts of memory, and specimen holders that allow you to tilt the specimen through nearly the full 360°. Obviously, disk specimens are not ideal for such tilting experiments. The principle involved is just as we've described for stereo, but we use about 100 times as many images and a computer is essential to analyze and view the results.

29.2 $2\frac{1}{2}$D MICROSCOPY

This is an example of imaginative terminology in the TEM field. If there are diffraction spots in the SADP which are too close together to give separate DF images, center the objective aperture around them all. Then, if you view two DF images taken at different focus settings through a stereo viewer, you see features at different apparent depths. However, you're using a pseudo-stereo technique because the 'depth' difference is due to a difference

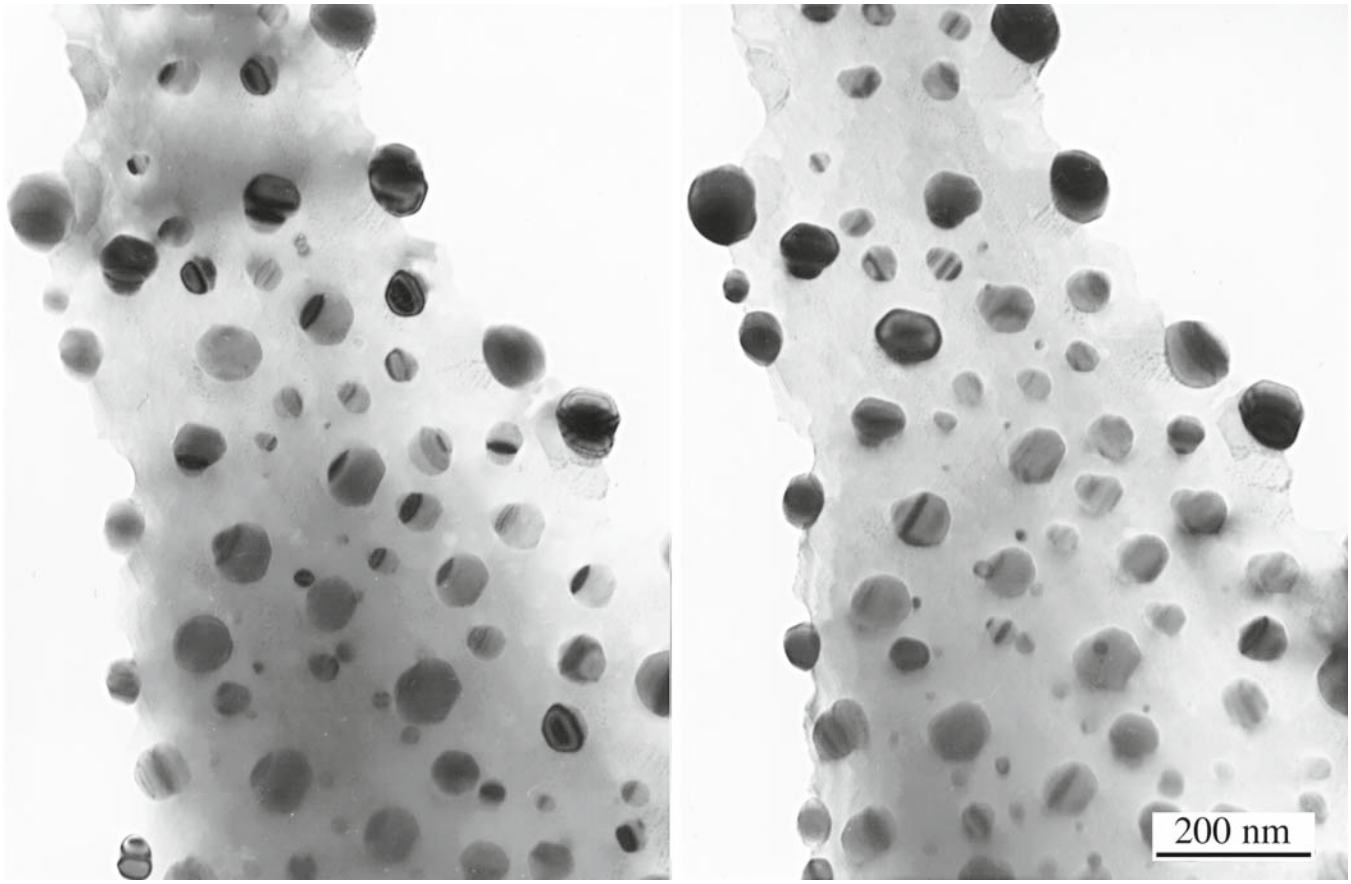


FIGURE 29.1. A stereo pair of spinel precipitates in (and on) a NiO-Cr_2O_3 specimen which, when combined, show the relative depth of the precipitates. You can see the small parallax shift between the two images. Note also the mixture of mass-thickness and diffraction contrast.

in **g**, not a true depth difference at all: hence the term $2\frac{1}{2}$D or 'not quite 3D.' The technique is also known by the rather more staid name of 'through-focus DF.'

Bell developed a simple theory explaining how a change in focus Δf introduces a parallax shift in the image, y.

$$y = M\Delta f\,\lambda\mathbf{g} \tag{29.2}$$

$$\Delta y_{12} = \Delta f\,\lambda\mathbf{g}_{12} \tag{29.3}$$

where M is the magnification, λ is the electron wavelength, and **g** is the diffraction vector. So, as you change focus, the parts of the image coming from different diffracted beams shift relative to each other by an amount, $M\Delta y_{12}$.

The term $\mathbf{g}_{12}$ ($= \mathbf{g}_1 - \mathbf{g}_2$) is the vector between the spots arising from features 1 and 2. It is this parallax shift which introduces a stereo effect in the same manner as one introduced through tilting.

When would you need to use such a technique? Well, if you're dealing with a multiphase specimen in which many of the phases have similar structures and/or lattice parameters, then there would be many closely spaced diffraction spots. In these circumstances, if you succeed in separating the spots with a small objective aperture, the selected spot will undoubtedly be very close to the aperture and the image will display serious aberrations. It's more likely you will be unable to select only one spot in the aperture.

You must start by setting up a CDF imaging configuration, but one in which several diffraction maxima are contained within the objective aperture. In the normal CDF image, therefore, all the parts of the specimen diffracting into the several spots will appear bright. However, if you take two images at different defocus conditions, the features diffracting into different spots will shift by slightly different amounts given by equation 29.3. This relative shift between the two images is perceived by your brain as a relative depth difference when you view the images simultaneously through a stereo viewer. You might argue that your images are going to be out of focus, and strictly speaking you're right, but remember that in the TEM there's a large depth of field (see Section 6.7) and it's easy to keep both images reasonably in focus over quite a range of objective lens excitation, so long as the magnification isn't too high. The greater the change in the objective lens current, the better the ability to resolve similarly diffracting features, so this process works better at lower magnifications where the depth of field is greater.

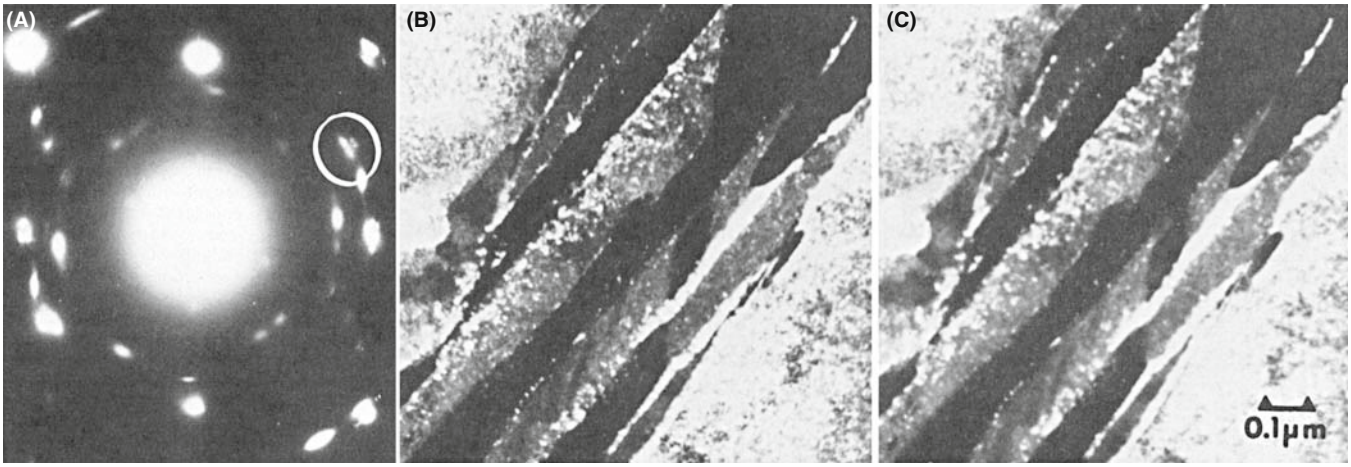


FIGURE 29.2. (A) SAD pattern of retained austenite and carbide precipitates in steel. (B, C) Stereo pair showing $2\frac{1}{2}$D effect in which the relative depth of austenite and carbide is related to the position of their diffraction spots in the objective aperture.

So the experimental steps are as follows

- Select the area of interest making sure the specimen is eucentric.
- Tilt the beam so that the several diffraction spots cluster around the optic axis; center the objective aperture around this group of spots.
- Return to image mode and *underfocus* the objective lens until you see the clarity of the image begin to degrade; then overfocus one click back to maintain focus.
- Record an image.
- *Overfocus* the objective lens until the image begins to lose clarity again, then underfocus back one click.
- Record another image.
- Develop or display the images and view them in a stereo viewer.

Figure 29.2 shows a pair of CDF images taken under these conditions. The DP, as you can see, has many closely spaced spots. This example and many more are given in a well-illustrated review by Sinclair et al. If you look at the images through a stereo viewer, different bright regions will appear at different 'heights.' If you read the review, you'll find that it is quite easy to determine which features in the image are responsible for which diffraction spot.

29.3 MAGNETIC SPECIMENS

If you happen to be looking at a magnetic material, then this can cause major problems for both you and the TEM and you might want to switch to studying aluminum. However, if you're patient and want a challenging task, then you can learn to correct for the magnetic disturbance introduced by the specimen. You can also get more information by making use of the interaction of the electron beam with the magnetic field of the specimen. The use of TEM to study magnetic effects has seen a resurgence since the discovery of high-T_c superconductors, and the interest in magnetic recording media capable of storing a high density of information.

MAGNETIC DISKS

You should not use self-supporting magnetic disk specimens if you can avoid doing so.

First, we'll look at how to get the best images from your magnetic specimen and then we'll describe a couple of specialized imaging techniques that allow you to see either the domains or the domain walls, which are relevant to both ferromagnetic and ferroelectric materials. This topic will be extended in the companion text.

29.3.A The Magnetic Correction

If your specimen is magnetic, its magnetic field will deviate the electron beam as it passes through and then the electrons that you use to form images will not be on the optic axis. All your images will be severely aberrated and shift when you try to focus them. You can minimize these effects as we'll now describe.

The most important step is to make the magnetic part of your specimen as thin and small as possible to reduce its total magnetic field strength.

However, small flake specimens are more easily pulled out from between support grids when you put the specimen into the objective lens. So, make sure your foil is well clamped into the specimen holder and use oyster grids for thin flakes with conducting glue if possible. Insert the holder into the lens with the objective lens strength as low as possible; look at the current running through the lens and minimize it. Ideally, you will use a TEM that has been optimized for the task.

Such a TEM may, for example, have a modified objective lens. Notice that none of the images shown in this section are HRTEM.

MAGNETIC MATERIAL

If you lose any specimen, or part of a specimen, in the TEM it is essential that you stop work and get help; if the specimen is magnetic this is doubly essential—and you can be identified if you don't get help!

You'll find it very difficult to set a magnetic specimen to the eucentric height because, as you tilt it through the 0° position, the specimen tends to rotate, or it is pulled out of the holder (if you're unlucky) or it shifts slightly in height and position. If you lose the specimen, the column must be split and the lost specimen retrieved, otherwise you may introduce a fixed, non-correctible, astigmatism into the microscope. In the extreme, you may end up with your specimen welded to the polepiece, which gets to be quite expensive. It is better to incur the wrath of your technical support staff when the objective polepiece is still undamaged.

So when you have found a general area of interest, set up the eucentric height by tilting the specimen, taking care *not* to tilt through 0°; keep the tilt range to one side of zero. When the eucentricity is reasonable (and it rarely gets perfect for magnetic materials), focus and center a recognizable feature on the screen, remove the objective diaphragm, then carry out the following steps

- Underfocus C2 and the objective lens by one coarse step. The feature should move off center. If you lose illumination you haven't underfocused C2 enough.
- Return the feature to the screen center using the DF beam tilt potentiometers. This brings the imaging beam onto the optic axis.
- Refocus the image, recondense the beam, and if the feature shifts, recenter it with the stage translators.
- Repeat the procedure for objective lens overfocus, until the feature stays centered as the objective lens is moved through focus

If you then go to the DP, you'll see that the 000 beam should now be on axis and you can put the objective diaphragm back in again and check the astigmatism. However (and here's the catch), if you tilt or traverse the specimen, then you will have to redo the whole correction since the magnetic field will change as you move the specimen relative to the beam. You'll rapidly get accomplished at this procedure if you practice for a while. Now, if you want to do CDF imaging at the same time as doing the magnetic correction, then you need an extra set of external DF controls, which may be an option or may be built into your microscope. Then you center the image feature with one set of potentiometers and tilt in the required *hkl* maximum with the other set. Figure 29.3 shows the image of a magnetic specimen before and after the magnetic correction. It is more difficult to image magnetic specimens in STEM mode, because the scanning beam may interact differently with different areas of the specimen and so the image quality will be variable.

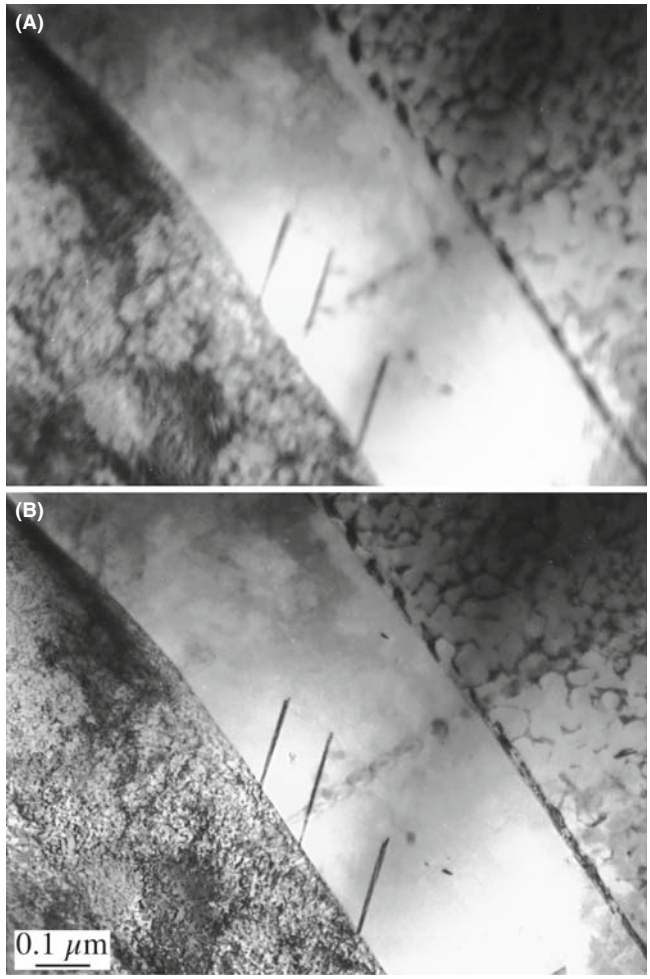


FIGURE 29.3. The effect of the magnetic correction on the quality of the image of an Fe-Ni meteorite: (A) before magnetic correction, (B) after magnetic correction.

29.3.B Lorentz Microscopy

When you've corrected for the magnetic field introduced by the specimen, you can image the magnetic domains if they're on the right scale to see in the microscope. This is a form of phase-contrast microscopy which we saw back in Chapter 23. The general term for this kind of imaging is called Lorentz microscopy and it comes with two options, and a special lens.

THE LENSES

The objective lens produces a magnetic field around the sample. A Lorentz lens replaces the objective lens: it gives a lower magnification and is farther from the sample.

Lorentz? Because we use the Lorentz force.

Foucault Images: If there are several domains in the illuminated area, the electron beam will be deviated in different ways by different domains. This will result in a splitting of the diffraction spot, as shown in Figure 29.4A. It is useful to know immediately that the direction of magnetization is normal to the direction of spot splitting (the effect of the Lorentz force). You can then take a BF image using either of the split 000 reflections. If you use all the spots, you can just see the domain walls (Figure 29.4B). The domains bending electrons into the chosen spot will then appear bright and the other domains will appear darker, as in Figure 29.4C. These images of the like domains are called Foucault images (perhaps after the French inventor of the pendulum used to demonstrate the earth's rotation) for reasons unknown to the authors. It is possible to see an analogous effect in STEM images if you use a detector that is cut into electrically isolated segments so different portions (usually halves or quadrants) pick up electrons coming through different domains. In STEM you can also add, subtract, or divide the various signals, as shown in Figure 29.5, because you are picking up each signal digitally. However, the intensity in Foucault images cannot be related quantitatively to the magnetic induction, so their only use is to give a rapid estimate of the domain size.

Fresnel Images: This option allows you to see the domain walls rather than the domains. As we discussed in Chapter 23, Fresnel imaging is named for another famous Frenchman whose TEM images would never be in focus. If you over- or underfocus the objective lens, then the electrons coming through different domains will produce images in which the walls appear as bright or

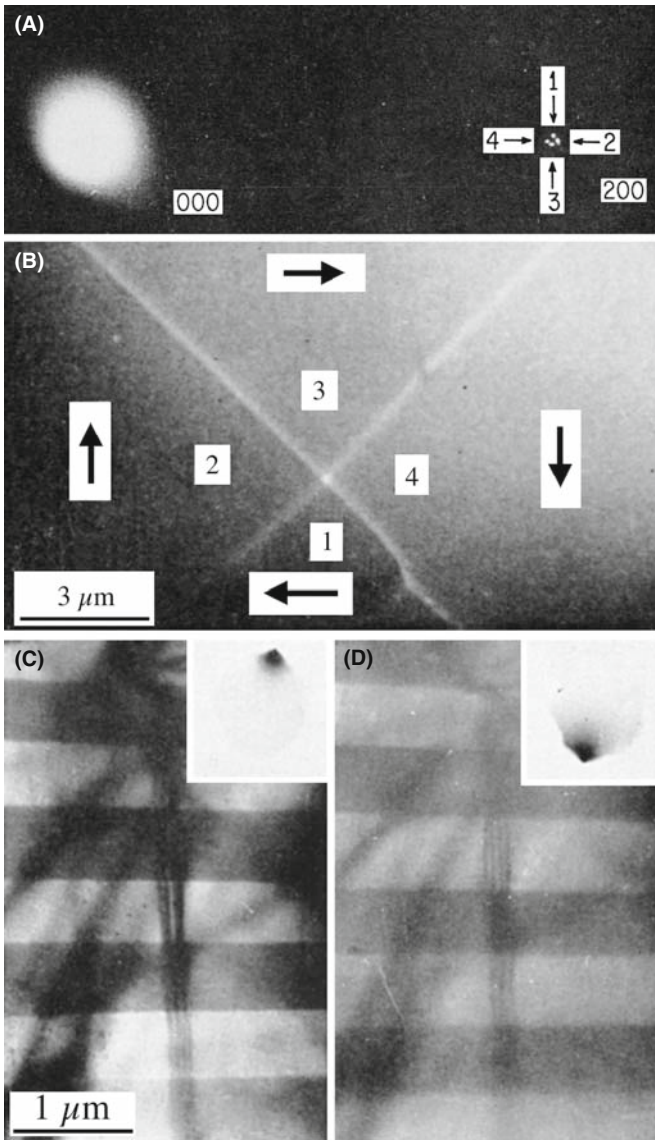


FIGURE 29.4. (A) Splitting of the 200 spot from Ni_3Mn due to the presence of magnetic domains. (B) Image taken from all of the split spots in (A) showing the four domains which scattered electrons into the various spots. (C, D) Foucault images of domains in Co formed by displacing the objective aperture to select one of the two split spots as shown in the insets.

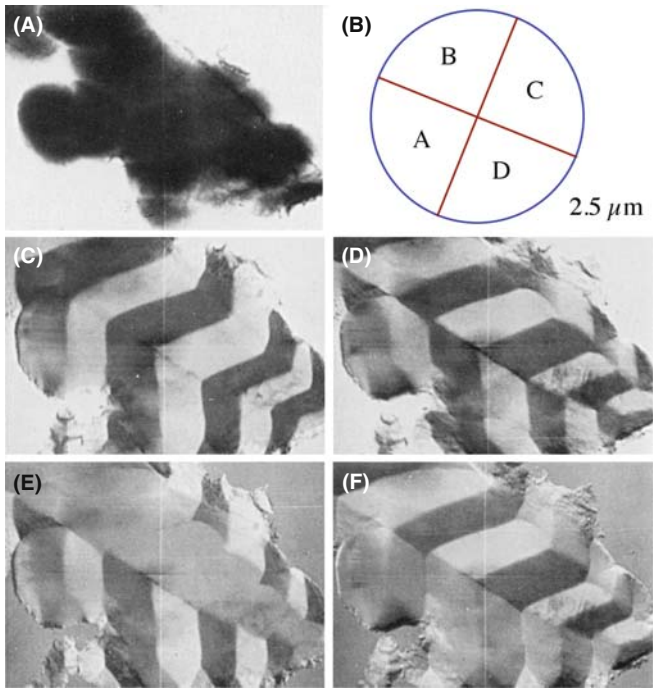


FIGURE 29.5. The use of a quadrant detector in STEM to differentiate regions of different magnetic induction in pure Fe. (A) The BF image from all four quadrants is shown schematically in (B). (C–F) Images formed by different quadrant combinations. Regions showing the same intensity are regions of like magnetic induction.

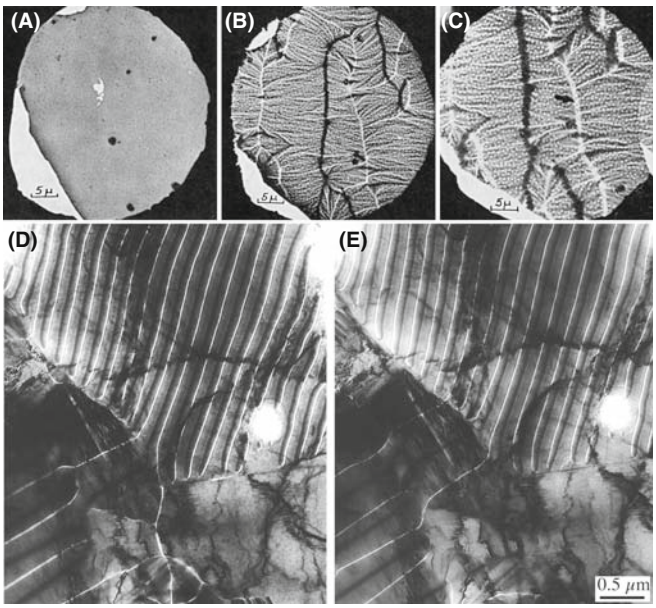


FIGURE 29.6. (A) In-focus image showing no magnetic contrast. (B, C) Fresnel-defocus images showing the magnetic domain walls which image as bright and dark lines and increase in width with defocus. (D) Underfocus and (E) overfocus images showing reversal of domain-wall contrast.

dark lines, which reverse contrast as you go through focus, as in Figure 29.6. The contrast depends on whether the electrons going through the domains on either side of the wall were deflected toward or away from each other, as shown schematically in Figure 23.16A. From such images you can work out whether you're looking at Bloch walls, Néel walls, or cross-tie walls. Figure 29.7 explains why this technique is so important.

In an FEG TEM, the highly coherent source means that coherent Fresnel and Foucault (CF) imaging is possible. These techniques give quantitative measurements of the magnetic induction.

If you want to do Lorentz microscopy, you have to decrease the field strength of your objective lens, which will otherwise dominate the internal field and thus control the domain size in the specimen. So, you either switch off the objective lens and use the intermediate lens for focusing, or use a specially designed low-field lens. TEM manufacturers offer appropriate objective lenses for Lorentz microscopy and it can also be done in STEM.

In addition to imaging the domains and the walls, you can also image the flux lines in the specimen if you evaporate Fe on the surface, just as you use Fe filings to delineate flux lines around a bar magnet. If you heat the specimen *in situ,* then the spot splitting decreases linearly to become zero at the Curie temperature.

29.4 CHEMICALLY SENSITIVE IMAGES

In Section 16.7, we showed that in many materials the structure factor, F, for some reflections was sensitive to the difference between the atomic-scattering amplitudes of the constituents. If we form a DF image using such a reflection, it will, in principle, be sensitive to changes in the composition of the material. The material systems which have been studied most extensively are related to GaAs and the other III–V compound semiconductors. There is great interest in partially replacing either the group III element or the group V element locally to produce superlattices and quantum wells, as illustrated in Figure 29.8. The contrast in the DF images will be brighter in thin specimens when the difference in the two atomic-scattering amplitudes is large. Therefore, layers of $Al_xGa_{1-x}As$ appear brighter than the surrounding GaAs matrix.

This imaging technique can, in principle, be applied to many different materials; it is the same, in principle, as the use of superlattice reflections to image ordered regions (see Figures 16.5 and 16.6). We will use the same information when discussing quantitative chemical lattice imaging in Section 31.11; we also use this effect when studying site location by ALCHEMI (see Section 35.9). In practice, it is often important to know how abruptly the composition changes, since this interface affects the properties of the material, so you may be interested in the change in contrast which occurs exactly at the interface. However, you should keep in mind our discussion (Section 24.10) of surface relaxations in the thin specimen which can influence any diffraction-contrast images.

29.5 IMAGING WITH DIFFUSELY SCATTERED ELECTRONS

This is just a variation on the theme of normal DF imaging. If your specimen contains non-crystalline regions which scatter electrons into a different region of reciprocal space away from the diffracted beams, then the non-crystalline regions can be seen in strong contrast. To do this, you perform a DF operation with the objective aperture away from any strong diffraction spots but at a position to intercept a fraction of the diffuse scatter. For example, in silicate glasses the diffuse scatter peaks radially at 3–4 nm^{-1}. As shown in Figure 29.9, a DF image reveals the amorphous regions with high contrast at grain boundary in a ceramic bicrystal. However, you must be very careful when interpreting such images, as shown by Kouh et al.

Diffuse scatter can also arise from short-range ordering in the specimen due either to microdomains of ordered nuclei or local regions of increased order analogous to spinoidal decomposition. These regions produce diffuse intensity maxima at positions that will eventually correspond to a superlattice spot when the short-range order has developed to long-range order (Cowley 1973a,b). A DF image from the diffuse scatter will reveal the short-range ordered regions as diffuse

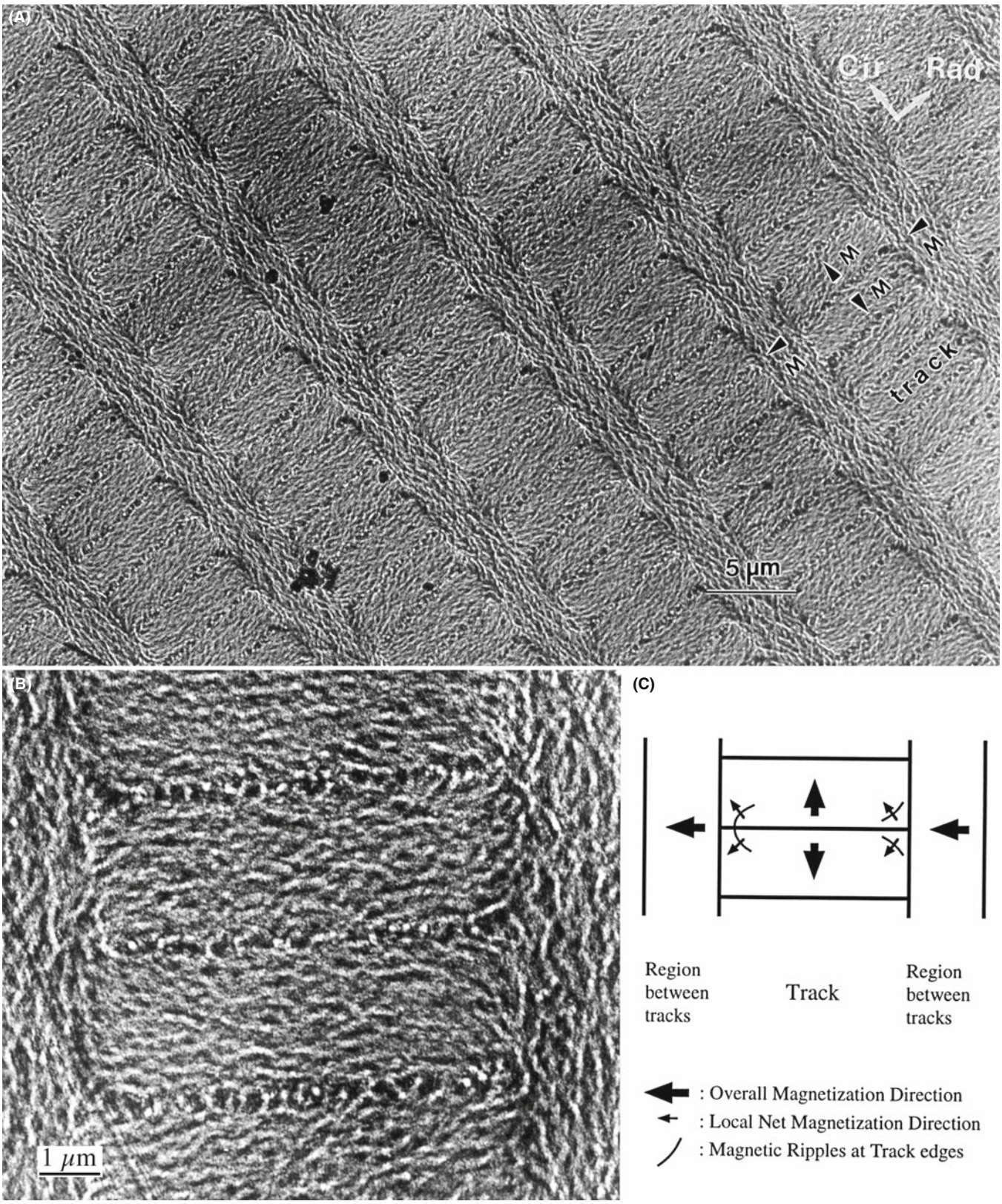


FIGURE 29.7. (A) Fresnel image of $Co_{84}Cr_{10}Ta_6$/Cr film on a smooth NiP/Al substrate used for magnetic data storage, imaged in the bits-written magnetic state. (B) Fresnel image at higher magnification. (C) Schematic of magnetic ripples at the track edges. The bits were in alternating direction of magnetization along the tracks in the circumferential direction of the hard disk, while the inter-track regions had remnant magnetization in the radial direction, perpendicular to that within the bits.

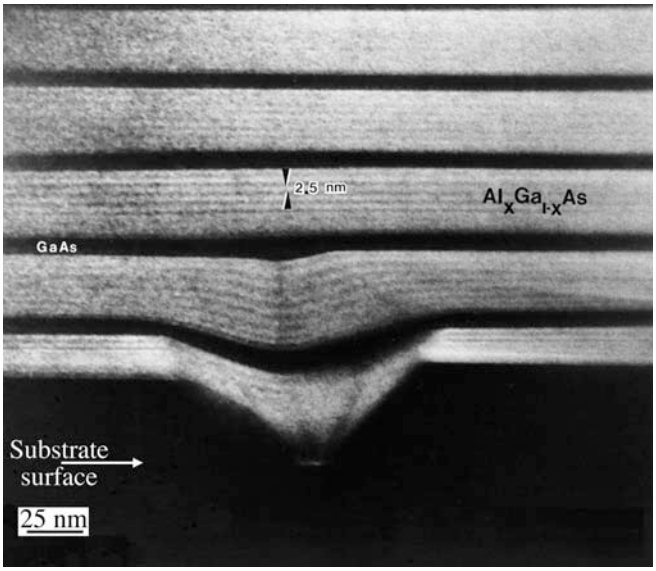


FIGURE 29.8. A chemically sensitive image of a GaAs-AlGaAs quantum-well structure. The composition of the AlGaAs is non-uniform because of growth fluctuations and the substrate surface was imperfectly covered by the first GaAs layer.

FIGURE 29.9. Diffuse-scatter DF image of a glassy material at a GB in a bicrystal of Al_2O_3.

intensity maxima. Remember that we saw similar contrast effects, *but in the DPs*, of short-range ordered materials in Figure 17.11.

The practical suggestion in either of these cases for a modern TEM is to use a displaced aperture approach rather than translating the beam.

29.6 SURFACE IMAGING

We can obtain surface information in the TEM in a variety of ways. We can do reflection electron microscopy (REM), or use a technique called 'topographic contrast.' We can also form an SE image in an SEM or STEM, as we describe later in the chapter. We've already mentioned profile imaging by HRTEM in Section 28.16.

29.6.A Reflection Electron Microscopy

Reflection electron microscopy (REM) of surfaces requires that you mount your specimen in the holder so the beam hits at a glancing angle, as shown in Figure 29.10A. Since the electrons are scattered from the surface, your specimen doesn't need to be thinned. The image is foreshortened by an amount which depends on the reflection used. Different parts of the specimen will be focused at different positions behind the lens. Once you have chosen an image plane, you can move the specimen in the z-direction to focus different regions of the surface. A reflection high-energy electron diffraction (RHEED) pattern is generated by the surface layers of the specimen, as shown in Figure 29.10B. This diffraction geometry is exactly that which we used to derive Bragg's law.

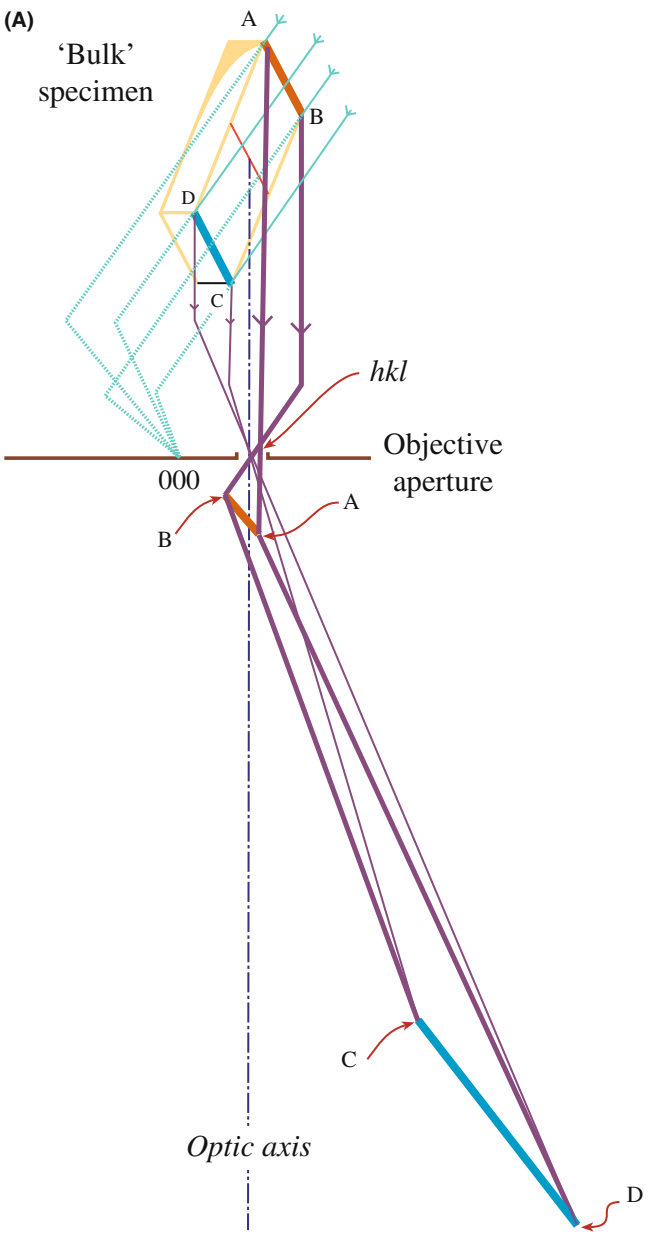
REM

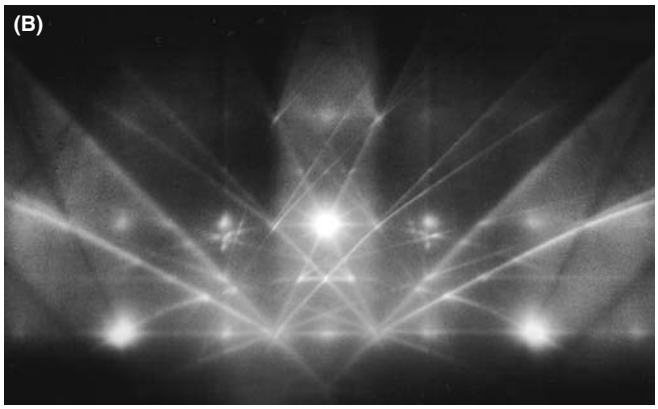
Notice that the surface is actually tilted away from the optic axis so that the diffracted beam will be on the optic axis: The electron source can't 'see' the diffracting surface!

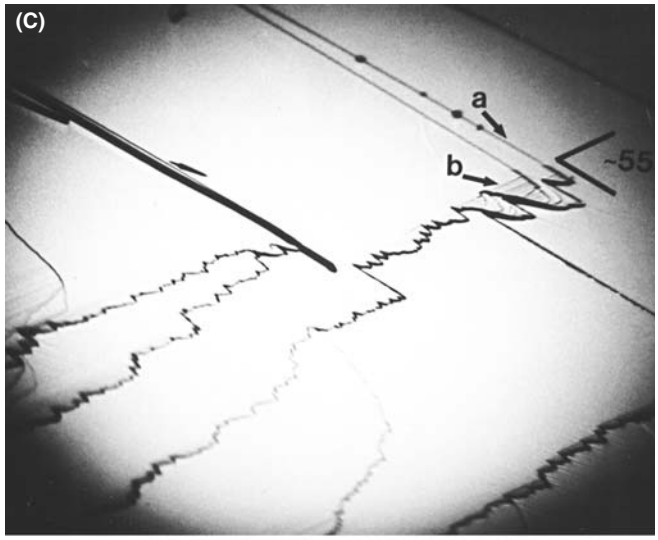
Once you've formed the DP, the experimental procedure is essentially the same as for conventional diffraction-contrast DF imaging. You insert an objective aperture to select a Bragg-reflected beam and form an REM image, as in Figure 29.10C and D. The images are strongly foreshortened in the beam direction but maintain the usual TEM image resolution in the plane normal to the beam where the image is in focus. This can make interpretation difficult; you'll see two different magnification markers in the two orthogonal directions. Note how different the two rotated images appear.

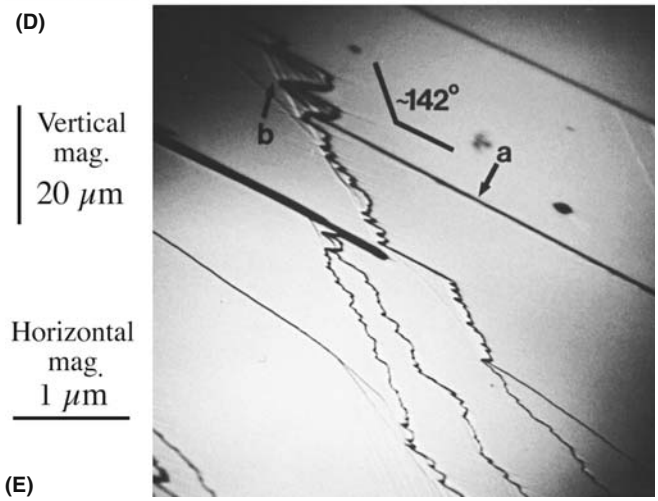
Ideally, you would like to have a very clean surface to simplify interpretation of the contrast. Of course, if the surface is not very flat, or is covered with a thick contamination or oxide layer, you won't learn much from the image. However, with care, you can use REM to study the surface of many different materials. All that you've learned about diffraction-contrast images will apply to REM images. For example, you can also detect chemically sensitive contrast, as shown in Figure 29.10E. Here the dark bands are layers of $Al_xGa_{1-x}As$ in a GaAs matrix; the contrast is sensitive to the actual value of x. Notice that this contrast is the reverse of what we saw in TEM using chemically sensitive reflections.

Since you can do REM using any regular TEM holder, you can, in principle, easily heat or cool your bulk sample. In many ways, this is easier than using a transmission specimen since the sample is much more robust. In situ REM studies of Si provided a leap in our understanding of the reconstruction of Si surfaces. Yagi et al. showed the surface reconstruction taking place in situ. This technique has many potential applications but requires practice.









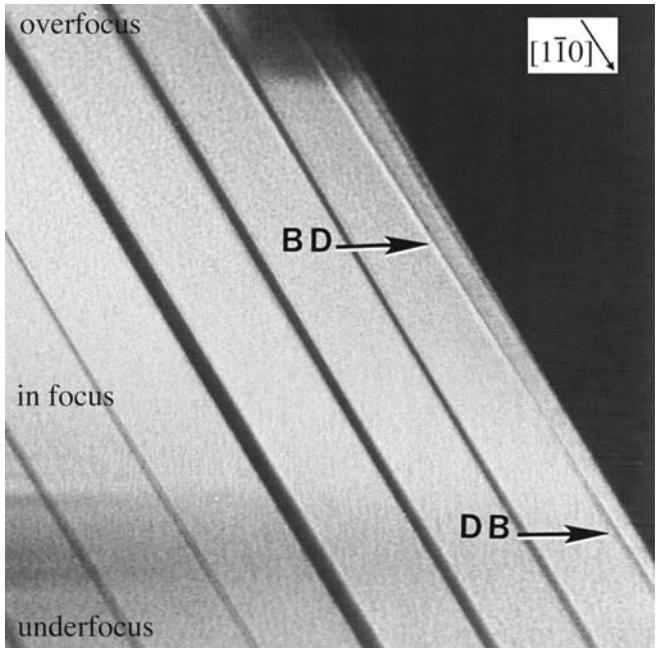


FIGURE 29.10. (A) Schematic diagram of the formation of an image using REM. (B) A 100 REM DP from a Si 001 surface. You can see the square array of spots but the 000 reflection has been blocked by the specimen itself. (C, D) REM images showing surface steps on a cleaved single crystal of GaAs. The two images are rotated with respect to one another and show the effect of the foreshortening. This emphasizes the care needed when interpreting foreshortened REM images. (E) REM image showing chemically sensitive contrast from quantum wells in GaAs/AlGaAs.

29.6.B Topographic Contrast

You can get a sense for the topography of your specimen by a neat technique which simply involves displacing the objective aperture until its shadow is visible across the region of the image that you're looking at. Around this area of the specimen you will see contrast which arises from an electron refraction effect but can be simply interpreted in terms of thickness changes in the specimen. In Figure 29.11, you can easily see that the Fe_3O_4 particles are supported on top of the carbon film and you can also see that the carbon film is not flat. Although the displacement of the aperture introduces astigmatism into the image, it does not limit resolution at the relatively low magnification at which this image was taken. To carry out an equivalent operation in STEM, you just displace the BF detector or shift the 000 diffraction disk so it falls half on and half off the detector. Again, the technique is not widely used, partly because refraction is a difficult topic in TEM and partly because the images don't look so 'nice.'

FIGURE 29.11. Topographic contrast from Fe_3O_4 particles on a carbon film. Ripples in the film are clearly visible. The shadow of the displaced objective diaphragm crosses the field of view.

29.7 HIGH-ORDER BF IMAGING

As we've taken some pains to point out, you get the strongest diffraction contrast in your images when **s** is small and positive. Under these circumstances, it's easiest to see crystal defects and interpret their contrast. This strong dynamical contrast can mask the details of the defect and sometimes it is more important to lower the contrast. You can do this by operating under kinematical conditions. Rather than having one strong low-index spot in the DP, if you tilt to make a high-order spot bright then the overall contrast is less and the background is not dominated by bend contours or thickness fringes, and the effects are shown in Figure 29.12. You can get a similar effect called 'kinematical diffraction imaging' if you form a BF image when no *hkl* spot is strong.

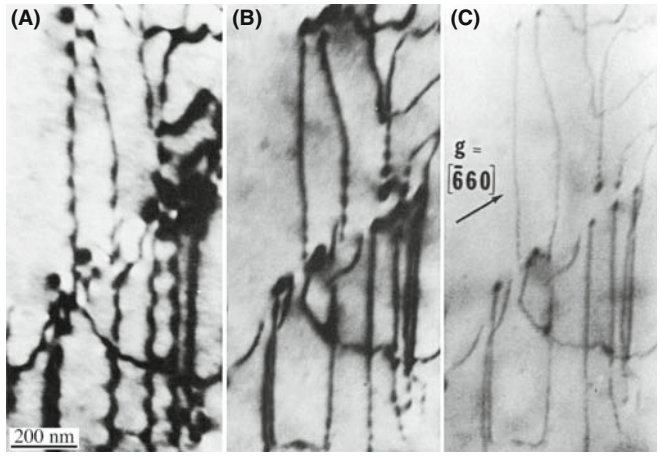


FIGURE 29.12. BF images of dislocations in GaAs taken with (A) **g**, (B) 2**g**, and (C) 3**g**. In the higher-order BF images the width of the defects decreases.

We get a similar effect in STEM imaging because, as we saw back in Section 22.6, dynamical-contrast effects are reduced when the reciprocity conditions between TEM and STEM are not fulfilled and all STEM images are more kinematical in nature.

Notice that Figure 29.12 was recorded at 650 kV. This tells you that high-order BF imaging can be a very useful technique when you're using an IVEM. You'll see there are subtle contrast effects that can occur at higher voltages. For example, in addition to improving the spatial localization of defect images, there are times when the 2**g** reflection gives better contrast.

PLAN YOUR EXPERIMENT

The point is, as usual, be prepared to experiment when you are sitting at the microscope.

Many of the early texts assume that you will be working at 100 kV, but only because that was all that was routinely available when they were written. The world is now different.

29.8 SECONDARY-ELECTRON IMAGING

SE images reveal the surface topography, which isn't much if you've polished your sample well, but is very important if you're looking at particulate specimens such as catalysts. You have to use a STEM if you're going to form SE images. If you look back to Figure 7.2, you'll see how the SE signal is detected by a scintillator-PM detector situated in the upper objective polepiece of a STEM. SEs generated in the top few nanometers of the specimen surface are confined by the strong magnetic field of the upper polepiece and spiral upward until they see the high voltage (~10 kV) on the aluminized surface of the scintillator. This design differs from the conventional SEM, in which the SE detector is situated under the final polepiece. In the STEM we get SE images of superior resolution and quality, which is why this design has been adopted for HRSEM. There are several reasons for this improvement in resolution

- A conventional SE image in an SEM invariably has noise contributions from BSEs which can enter the scintillator directly. In the STEM there is no line of sight to the SE detector for the BSE, and so the SE signal in STEM lacks the BSE noise that exists in an SEM image.
- The brightness of a thermionic source in a STEM is higher than in an SEM, because of the higher kV, and so the SE signal will be correspondingly stronger. An FEG will be brighter still.
- The C_s of STEM objective lenses is usually a lot smaller than for conventional SEMs. Therefore, SE images in a STEM are invariably of better quality than in an SEM because these two factors increase the S/N ratio.

The fourth advantage depends on a knowledge of the different types of SE which we'll now discuss.

The presence of remote SE signals in an SEM also decreases the S/N ratio. As shown in Figure 29.13, the SE detector in an SEM can pick up four different types of secondary electrons labeled SE I –SE IV generated by the primary electron (PE) beam

- The SE I signal is the only signal we want, since it emanates from the region around the probe and contains high-resolution topographic information from that region only.
- The SE II signal comprises SEs generated by BSEs that emerge from the specimen some distance away from the beam. The only way to reduce SE II is to reduce the BSE fraction and we can do this by using thin specimens, as in STEM.
- The SE III component arises from parts of the microscope stage that are struck by BSE from the specimen. So, reducing the BSE fraction by using thin specimens also reduces SE III.

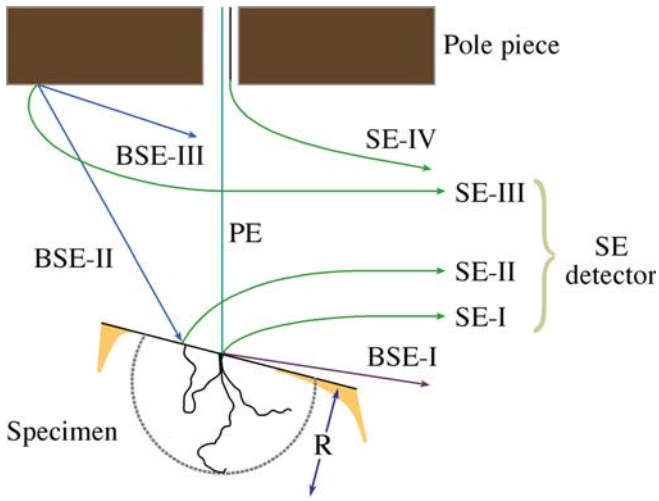


FIGURE 29.13. The four possible sources of SEs that enter the detector in a conventional SEM. The SE I signal is the only desirable one since it comes from the probe region, but SE II from BSE electrons, SE III from the stage of the microscope, and SE IV from the final aperture all combine to reduce the S/N ratio in an SE image. In a STEM, the relative contributions are different.

- The SE IV signal comes from SEs generated at the edges of the final probe-forming aperture, which in an SEM, often sits in the final lens. In a STEM, however, the C2 aperture performs this function and it is well away from the stage and the SE detector.

So, in a STEM with a thin specimen, the ratio of desirable SE I to undesirable SE II–IV is higher than in a conventional SEM. If you look at a bulk sample in the STEM, then the SE II signal will be the same as in an SEM since the BSE yield does not vary with beam energy. Furthermore, the SE III signal may be slightly worse than in an SEM because of the smaller STEM stage. Experimentally, the net result of all these differences is that the STEM still provides higher-resolution SE images of the same bulk sample than a conventional SEM, as shown in Figure 29.14. Now, the highest resolution FEG SEMs incorporate many of the design aspects of the STEM stage and can produce SE image resolutions <1 nm at 30 kV; some SEMs even encourage you to use thin specimen and give a simple STEM image.

A thermionic source STEM cannot better that SEM resolution; however, an FEG STEM at 100 kV or higher should, in theory, offer the best possible SE images and may even show atomic-level topographic information, especially with an aberration-corrected probe.

Despite these obvious advantages and the ready availability of STEMs with SE detectors, very few high-resolution studies of the surface topography of specimens have been carried out with a STEM. It appears that one of the major uses of the SE image is simply to find the hole in the TEM disk! Why this is so remains unclear, but we strongly encourage you to use

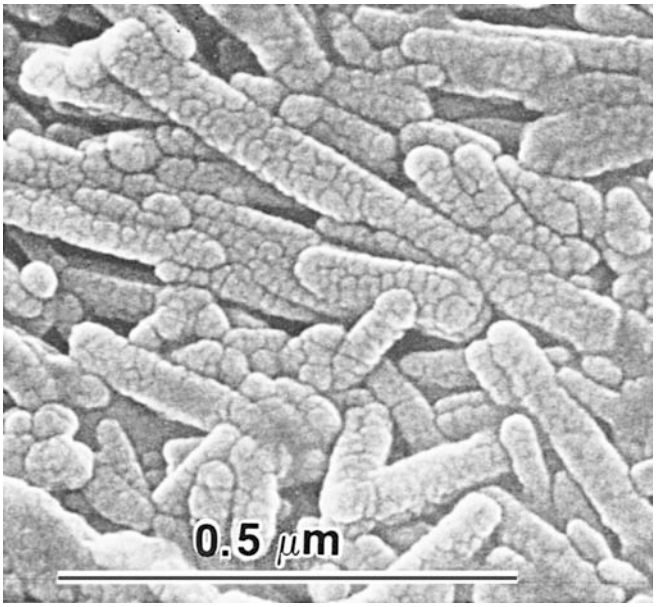


FIGURE 29.14. High-resolution SE image of coated magnetic tape in a TEM/STEM with an LaB_6 source at 100 kV showing ~2 nm spatial resolution.

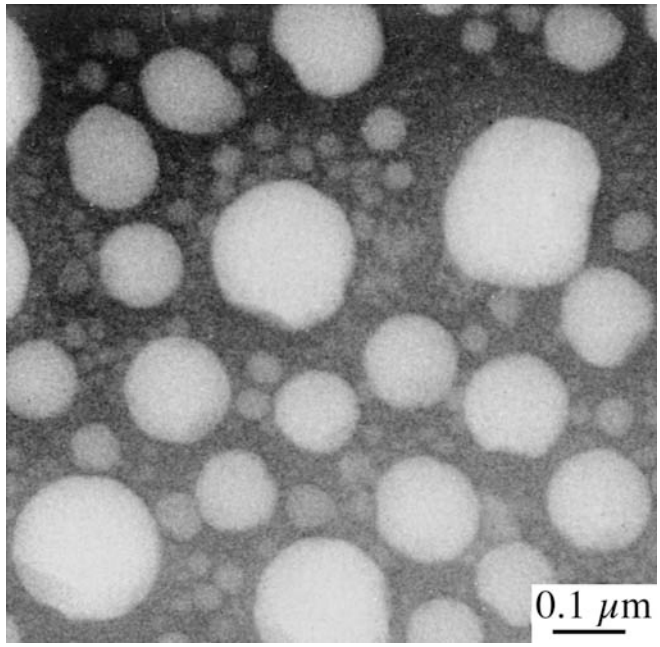


FIGURE 29.15. High-resolution STEM BSE image of Au islands on a carbon support film obtained in a TEM/STEM at 100 kV with an LaB_6 source. Resolution of ~7 nm is obtained, although the image quality is not very good because of the poor signal level.

SE images if they are available in your STEM. Of course, your specimen surface must be prepared carefully and kept clean, otherwise you'll image contamination, oxide, or some other artifact of your preparation process. If you have to coat your specimen because it's an insulator, then you should use a modern high-resolution, high-vacuum coater to generate a continuous thin film of a refractory metal such as Cr. Don't use the more conventional Au-Pd coatings, which merely mask the fine detail of the specimen surface.

29.9 BACKSCATTERED-ELECTRON IMAGING

Remember that the BSE detector in a STEM is situated directly under the upper objective polepiece. In an SEM, the detector is in the same position and we get equal collection efficiency. Higher kV gives a brighter source, but specimen thickness dominates the yield and it is abysmal in thin specimens (0.4% in 100 nm of Au). (Look at Figures 2.4, 36.1, and 36.2 and check the number of BSE in the Monte-Carlo simulations.) So there is a very low total signal and the S/N is poor. Nevertheless, in a thin specimen we can still form BSE images in the STEM that show resolution approaching the probe size, but only if a high-contrast specimen is used, e.g., Au islands on a C film, as shown in Figure 29.15. So in general, the BSE signal in a STEM offers no obvious advantages over a modern SEM with a high-efficiency scintillator or semiconductor detector. In fact, the manufacturers of dedicated STEMs do not even offer a BSE detector as an option, and it is doubtful if it is worth installing one on a TEM/STEM.

For both SE and BSE imaging in STEM, you can look at both electron transparent or bulk samples of the sort normally examined in a conventional SEM. A major drawback to the STEM compared with the SEM is the relative volume of the microscope stage. In an SEM stage you can insert samples up to several centimeters in diameter and several centimeters thick, which means that sample preparation from the original object is often minimal. The confined-stage region of a STEM means that, even with a specially designed holder for bulk samples, the largest specimen will be about 10 mm × 5 mm and not more than a couple of millimeters deep.

29.10 CHARGE-COLLECTION MICROSCOPY AND CATHODOLUMINESCENCE

Charge-collection microscopy (CCM), otherwise known as electron-beam induced conductivity (EBIC), and the related phenomenon of cathodoluminescence (CL) are common techniques for the characterization of semiconductors in the SEM. It is possible to do the same things in a STEM, but few have been brave enough to try. You learned back in Section 7.1 about semiconductor electron detectors in which the incident beam generates electron-hole pairs which are swept apart by the internal field of the p-n junction and not allowed to

recombine. So if your *specimen* is a semiconductor, electron-hole pairs will be formed during the normal imaging process. You have to separate out the pairs by applying an external voltage through ohmic contacts evaporated on the surface of your foil. Then you can use the electron charge pulse to generate a signal on the STEM screen. The signal is strong wherever the pairs are separated, and weak at recombination centers such as dislocations and stacking faults. You can also measure the minority carrier diffusion length. Now in a STEM you can easily see the recombination centers by standard imaging techniques, so CCM is not really any great advantage, but it is essential in an SEM because the defects are subsurface.

If we don't separate the electron-hole pairs, they recombine and give off visible light. This light is an extremely weak signal, but it can be detected and dispersed by mirrors and spectrometers shown back in Figure 7.6. Again, recombination centers appear dark in CL images because most of the recombination is deep-level and non-radiative. The advantage of CL over CCM is that you don't have to coat your specimen to produce ohmic contacts, and you get a spectrum of light which contains information about doping levels and band-gap changes. However, you have to dedicate your STEM primarily to this imaging mode and it is a difficult and tedious technique. Early work in this area was pioneered by Petroff et al. and more recently by Batstone. It is not obvious, however, that doing the work in transmission offers major advantages over studies of bulk samples in the SEM (see the textbooks by Yacobi and Holt and by Newbury et al.) and this is reflected in the limited TEM applications of CL in the literature.

29.11 ELECTRON HOLOGRAPHY

Although the technique of electron holography became widely known in the early 1990s, Gabor had originally proposed the technique in 1948 as a way to improve the resolution of the TEM. The delay in its wide implementation was due to the lack of affordable, reliable FEG TEMs. An FEG is required so that the source will be sufficiently coherent; the FEG is then the electron equivalent of the laser. The topic is broader than you might have guessed, unless you've read the 1992 article by John Cowley entitled 'Twenty forms of electron holography' and the text by Völkl et al. We'll discuss the general principles of the technique and refer you to the companion text and the references for the details and more ideas.

The key feature is that, unlike conventional TEM imaging techniques, both the amplitude and the phase of the beam can be recorded. We can use this feature in two ways

- The effects of C_s can be partially corrected; thus we can improve the resolution of the TEM.
- We can examine other phase-dependent phenomena, such as those associated with magnetism.

Several different forms of holography are possible; the three best known are all based on the approach described in Chapter 23

- In-line holography.
- Single-sideband holography.
- Off-axis holography.

The approach which is mainly used in the TEM is the off-axis variation, as we'll discuss later. Figure 29.16 shows a hologram of a wedge-shaped crystal of Si oriented close to the [110] pole. The boxed

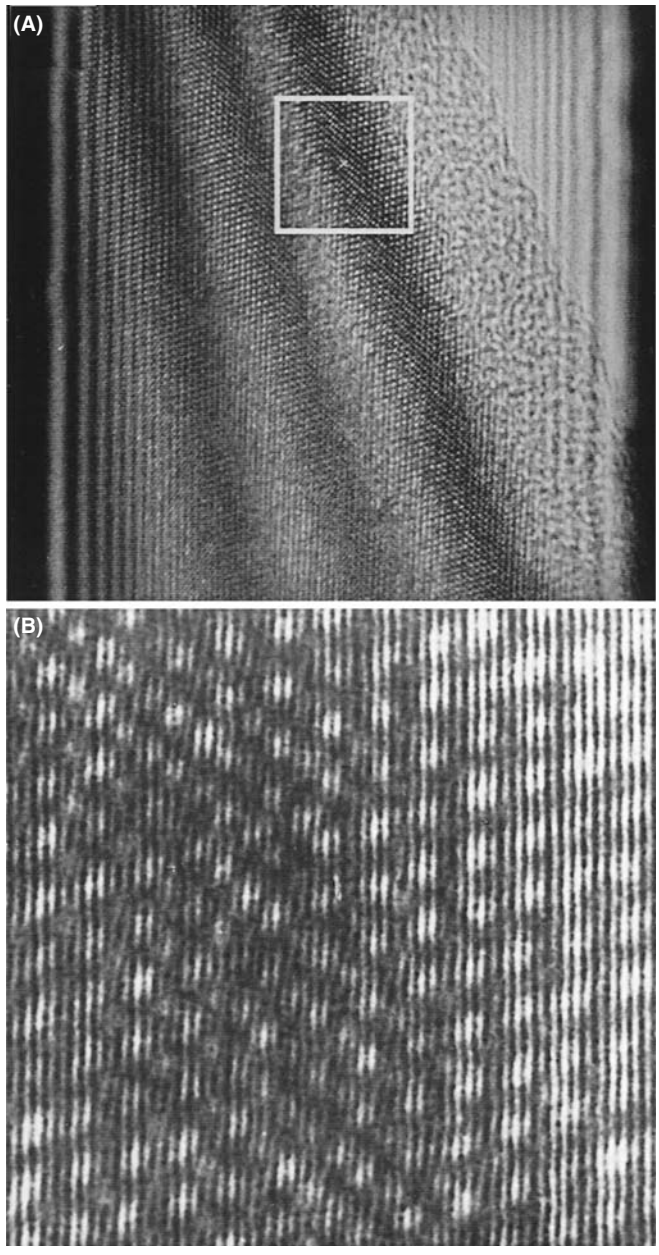


FIGURE 29.16. (A) Hologram of a [110]-oriented Si specimen; the broad bands are thickness fringes where the specimen is thickest at the lower right corner. (B) An enlargement of the boxed region in (A) showing fringes with a 70 pm spacing.

region of the figure shows a set of fringes which are only 70 pm apart. These are not lattice fringes but do contain phase information relating to the diffracted beams. In order to analyze such images, you'll need the appropriate software although much of the original research used optical processing techniques; this is such a specialized branch of image processing and image reconstruction that we will not cover it in Chapter 31, but it is discussed in the holography chapter in the companion text.

The principle of holography is shown by the schematic in Figure 29.17. Here $\chi(\mathbf{u})$ is the function we used in Chapter 28 to describe the effect of the objective-lens aberrations and defocus. In conventional TEM, we choose conditions so that the specimen acts as a pure phase object and the imaginary part of $e^{i\chi(\mathbf{u})}$, i.e., the sine term, converts this phase information into an amplitude which we record as the image. With holography, we can use the real part of the exponential too. A nice way to think of the process is that, for a real specimen, $\chi(\mathbf{u})$ mixes the amplitude, A, and phase, ω, from the specimen to give amplitude, A, and phase, Ω, in the image. In conventional imaging we record A^2 and lose all the information on Ω. With holography, we don't lose any information but we have to work hard to recover it. The original proposal by Gabor, to use holography to improve the resolution of the TEM, is being actively pursued. Although holography has still not reached its full potential in this field, you can see the future by looking again at Figure 29.16.

REFERENCE BEAM

The essential feature of the experiment is that a reference beam from the same coherent source passes outside your specimen and is then deflected by the biprism so that it interferes with the beam that passed through the specimen.

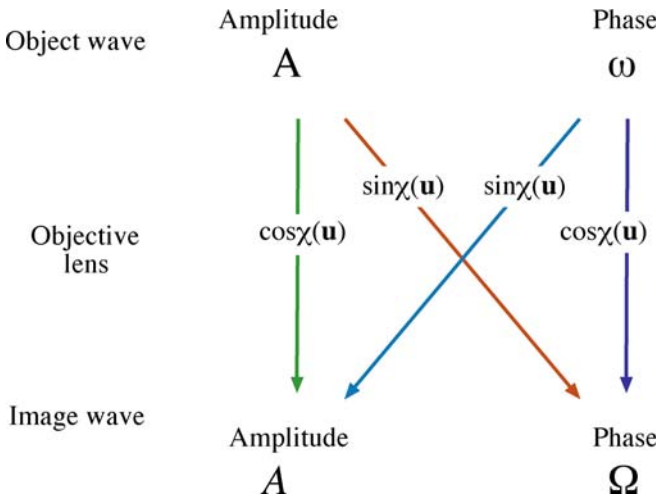


FIGURE 29.17. A schematic of how the objective lens mixes the amplitude and phase components of the specimen to form the amplitude and phase components of the image.

In practice, electron holography is carried out using an FEG TEM which has been fitted with a beam splitter; not every FEG TEM is so equipped. The beam splitter is made by coating a thin glass fiber with metal to prevent it charging, and assembled to give the biprism; i.e., it's the biprism we discussed in Section 23.7 and can be $<0.5\,\mu m$ in diameter. Part of the beam passes through the specimen while the other part forms the reference beam, as shown in Figure 29.18. You must be able to rotate either the biprism or the specimen. For holography, the biprism is placed below the objective lens, e.g., at the position usually occupied by the SAD aperture rod.

HOLOGRAM

A 3D image of an object; the image is made by holography.

HOLOGRAPHY

A method of creating a 3D image of an object. The process requires recording both the intensity and the phase information of the waves emitted or scattered by the object.

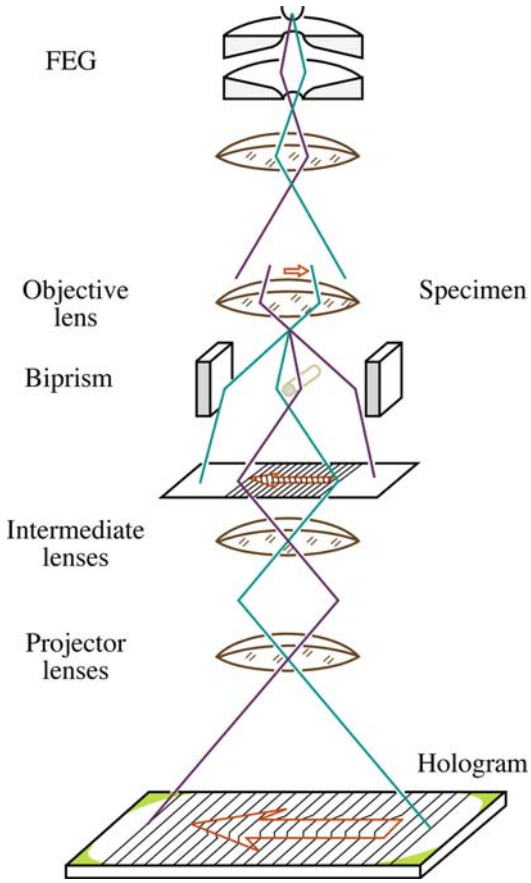


FIGURE 29.18. The operation of the biprism in electron holography.

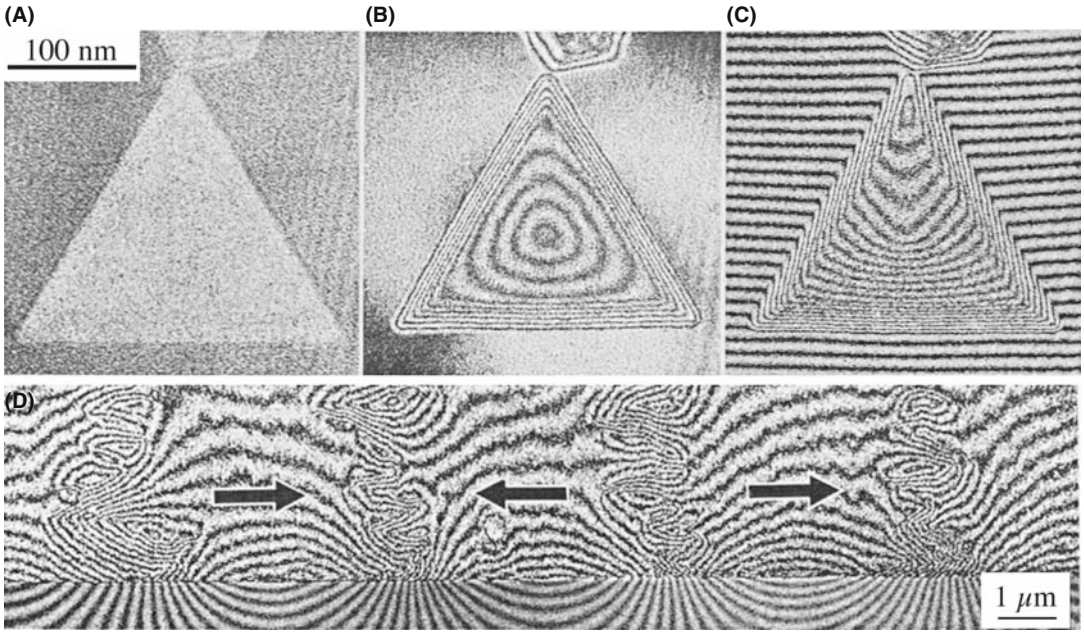


FIGURE 29.19. Holographic images of a magnetic Co particle showing (A) the reconstructed image, (B) the magnetic lines of force, and (C) the interferogram. (D) Line of force in a magnetic medium.

All that you then have to do is interpret the interference pattern, i.e., the electron hologram, but as you'll appreciate after reading Chapter 31, this process is not trivial. However, the technique does provide the possibility for 'coherent processing' of the electron wave. In principle, all the required data can be recovered from one hologram but the technique is unlikely to replace conventional HRTEM in most applications.

We'll conclude this section by illustrating one of the unique applications of the technique, namely, the imaging of magnetization characteristics and flux lines which has been developed by Tonomura. Figure 29.19 shows a series of images illustrating how magnetization patterns can be imaged both for single particles and for an actual recording medium. Tonomura's classic studies of the quantization of flux lines in superconducting materials is illustrated in Figure 29.20A. As the ring is gradually cooled, the phase shift due to the magnetization changes until it becomes superconducting: the phase shift inside the ring then becomes exactly π. The other images in Figure 29.20B and C show flux lines constricting as they enter the superconductor. Notice that you can also clearly see these flux lines outside the specimen: the magnetic field outside the specimen influences the electron beam.

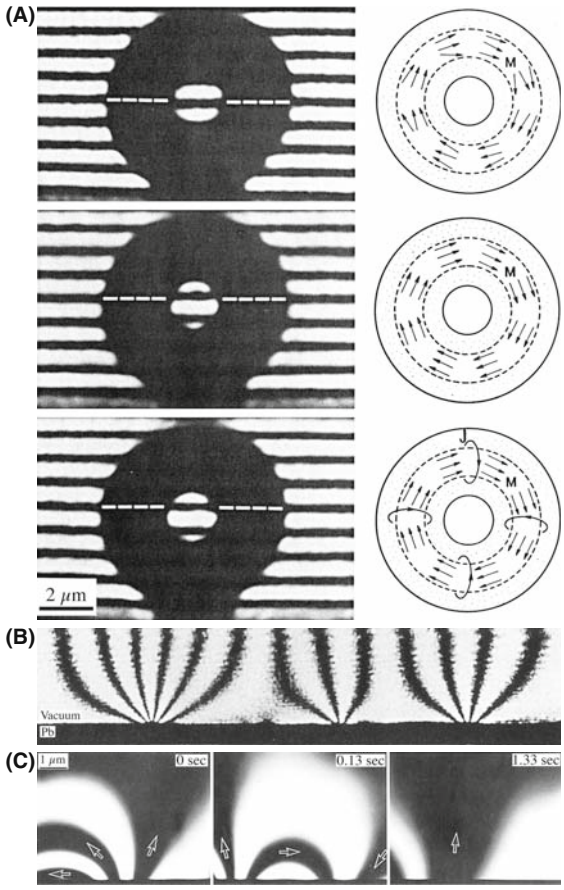


FIGURE 29.20. (A) Gradually cooling a ring (torus) of a superconducting material from 300 to 15 to 5 K to demonstrate the quantization of flux. (B) Magnetic flux lines penetrating a superconducting Pb film. (C) Interference of fluxons trapped in superconducting Pb as time increases from 0 to 0.13 to 1.3 seconds.

29.12 IN-SITU TEM: DYNAMIC EXPERIMENTS

We've mentioned several times that we can use the TEM for in-situ experiments. In Chapter 8, we described holders you can use to heat the specimen, e.g., causing phase transformations, or to strain it to change the defect

structure. In Section 7.3.C, we discussed CCD video cameras, which are the best way to record dynamic changes in the microstructure, and in this chapter, we've discussed imaging moving flux lines in magnetic materials. In situ experiments remove the doubts that exist whenever you observe materials after heat treatment or after deformation and then try to infer what actually happened at temperature or during deformation. It is implicitly assumed in most TEM investigations that cooling a heat-treated specimen to room temperature or removing the applied stress does not change the microstructure. Having made this assumption, we draw conclusions about what happened during our experiment. However, this assumption is clearly not valid for many situations. Nevertheless, we generally view our specimen at ambient temperatures and not under load.

In fact, there are good reasons why we rarely do in-situ experiments. The main reason is that such studies are difficult to perform on thin specimens. As we indicated in Section 26.8, when the surface properties dominate the bulk properties, as is often the case in thin specimens, TEM images and their analysis can be misleading. Surface diffusion is much more rapid than bulk diffusion and defects are subject to different stress states in thin foils where surface relaxation effect may dominate. The best way to overcome this limitation is to use much thicker specimens and this requires higher voltages. In-situ experimentation was widely used in the 1960s and 1970s when 1–3 MV instruments were first constructed and used to look through foils which were >1 μm thick (see the book by Butler and Hale and Figures 26.14 and 26.15). So in-situ experiments can be expensive: currently there is no longer a TEM operating at >400 kV in the USA.

However, the advent of 300–400 kV IVEMs, the construction of megaVolt instruments in Stuttgart and Japan and the extra space (because of the larger pole-piece gap) that is potentially available in C_s-corrected TEMs have brought a resurgence of interest in using in situ experimentation. Developments in electron optics, stage design, and recording media since the 1960s mean that combined HRTEM and in-situ experimentation is now possible. This combination is a powerful tool, permitting the observation of reactions at the atomic level, such as the motion of individual ledges at interfaces, as in Figure 29.21, although lower-resolution images are no less impressive.

You must bear in mind that the experiments you perform *in situ* are taking place under conditions that still do not approach the bulk conditions experienced by many engineering materials. In particular, the high-keV

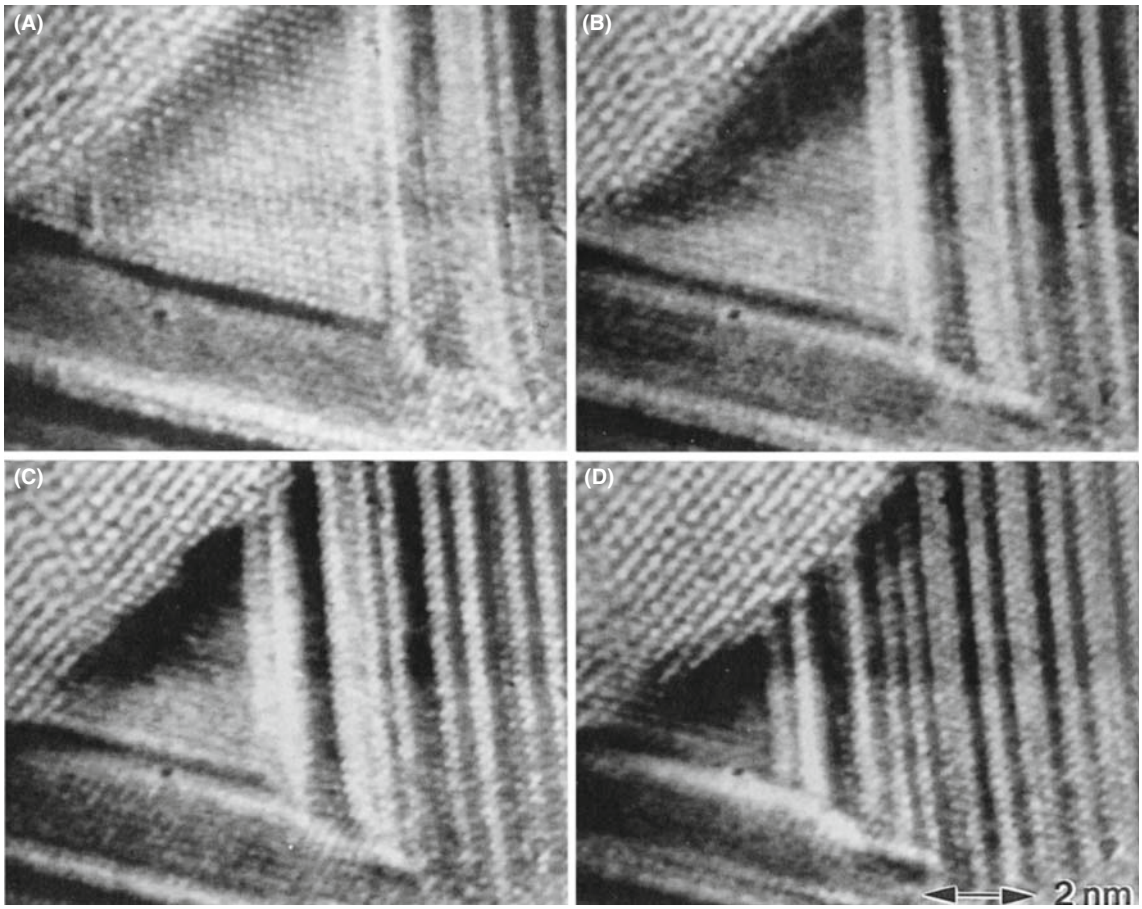


FIGURE 29.21. (A–D) Four images taken from a video showing the reaction front of a Ge/Ag/Ge trilayer while heating the specimen *in situ* at 250°C. The time interval between the images was 8 seconds. The Ge crystal grows from the upper left without moving the Ag lattice.

electron flux brings an uncertainty to all in-situ experiments, particularly if you are trying to infer kinetic data from measurements taken of reactions. So while in-situ experiments can give you powerful demonstrations of real-time changes in materials, be wary of direct interpretation. You have to ensure you are controlling all the variables, and you must cross-check kinetic data with calculations to verify that, e.g., diffusion fluxes are consistent with the known temperature and not influenced by surface or vacancy effects, having said that, it is nonetheless very useful to have access to an IVEM for this kind of experiment.

29.13 FLUCTUATION MICROSCOPY

Fluctuation electron microscopy (FEM) is a technique invented by Treacy and Gibson that has come into prominence in the last decade to study medium-range order (~0.5–2 nm) in amorphous materials, such as glasses, amorphous silicon and carbon. We saw back in Section 18.7 how to use different kinds of diffraction and imaging to study these increasingly important materials and FEM is a hybrid imaging and diffraction technique. The 'fluctuation' refers to changes in the local structure and orientation of the glass where 'local' means not quite down at the atomic level.

You can do FEM experiments in two equivalent ways. Either you can form a hollow-cone DF image (go back and check Figure 18.14) or you can do STEM diffraction with a fine probe, scanning the pattern across the STEM detector. The normalized variance of the intensity distribution in the microdiffraction patterns gives equivalent information to the hollow-cone image. In both modes, you insert a small objective aperture which, in either mode, limits the resolution of the fluctuation image to ~1–2 nm. This resolution matches the dimensional scale of local structural fluctuations, which appear as speckles in the image (see Figure 29.22A and B). Depending on the degree of local coherence in the random network structure, the speckles (both bright and dark regions) vary in intensity. More random structures give gray regions; more ordered structures give locally brighter or darker regions depending on whether the local electron scatter is on axis or off axis (intercepted by the diaphragm), respectively (Figure 29.22C).

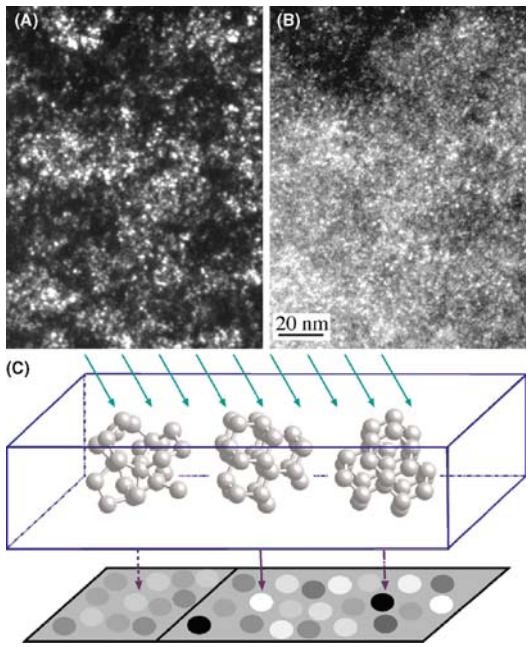


FIGURE 29.22. Two images of the same region of shungite. (A) The speckle corresponding to fluctuations on the graphite 002 reflection (3.0 nm^{-1}). (B) Ditto but for 7.1 nm^{-1}. Some of this speckle corresponds to fullerenes (and some to noise). The presence of fullerenes is inferred from the fact that there is a peak (such as that shown in the plot) due to the C–C bond distance, which is ordinarily forbidden. Structure noise from truly amorphous sample will not give a significant peak corresponding to ~1.5 nm. (C) Schematic diagram showing how the local coherence influences the intensity in the image.

29.14 OTHER VARIATIONS POSSIBLE IN A STEM

We can form images with the characteristic X-rays and the energy-loss electrons that we detect with the appropriate spectrometers. Rather than discuss these here, we'll save the topics for the chapters devoted to XEDS and EELS in Part 4.

In a STEM then, we can, in theory, detect and measure all of the signals generated in a thin specimen and shown way back in Figure 1.3. We can reproduce all the conventional TEM imaging methods and some unconventional ones too, such as *Z*-contrast and most of the specialized techniques in this chapter. Remember that for all the STEM signals we use a detector of one form or another. We can make detectors of different shapes and sizes, as we already described for looking at directional scattering from magnetic specimens. Remember also that the detectors permit us to digitize the signal so we can process it, manipulate it, and present it for viewing in ways that are impossible with analog TEM images. Most STEM systems come with relatively basic image processing, such as black level, gain, contrast, brightness, gamma, *Y* modulation, and signal addition and subtraction. The computer system for X-ray and EELS analysis often has a lot of image analysis software thrown in. So there's a lot you can do with the STEM images that really fall into the category of digital image processing which you may study in Russ's specialist texts, after you've read Chapter 31.

CHAPTER SUMMARY

We've described only a few of the many ways that we can manipulate the electron beam to produce different images and different contrast phenomena, and it is more than likely that there are still unknown methods awaiting discovery. For example, the introduction of a continuously variable detector collection angle in STEM is equivalent to using a continuously variable iris-type objective aperture in TEM and offers new imaging possibilities. Similarly, there is much to be learned from mutually complementary techniques in SEM and TEM.

We scarcely mentioned the use of annular apertures, conical illumination, pre- and post-specimen scanning, or rocking. All of these are possible, so you should be prepared to try new things in the microscope and see what the effects are. Many advances of the sort we have described only came about by accident, but the microscopist was wise enough to see the effect and try and understand it, rather than dismiss it as unimportant.

REFERENCES PHILOSOPHY

Dennis Gabor (Gábor Dénes; 1900–1979) worked in Rugby, England with the British Thomson-Houston company. This company was already merged with the Metropolitan-Vickers company and in 1960 became part of the holding company, Associated Electrical Industries (AEI). The early TEMs, including the HVEMs, in England were manufactured by AEI. He was awarded the Nobel Prize for physics in 1971 for discovering holography.

REM

De Cooman, BC, Kuesters, K-H and Carter, CB 1984 *Reflection Electron Microscopy of Epilayers Grown by Molecular Beam Epitaxy* Phil. Mag. A **50** 849–56. Using chemically sensitive reflections.

Hsu, T, Ed. 1992 *J. Electron Microsc. Tech.* **20**, part 4. A special issue using REM for imaging surfaces.

Wang, ZL 1996 *Reflection Electron Microscopy and Spectroscopy for Surface Analysis* Cambridge University Press New York. The latest book—reflecting that not much has happened recently.

Yagi, K, Ogawa, S and Tanishiro, Y 1987 *Reflection High-Energy Electron Diffraction and Reflection Electron Imaging of Surfaces* (Eds. PK Larsen and PJ Dobson), p. 285, Plenum Press, New York.

HOLOGRAPHY

Electron holography: The following key articles will give you a sense of the possible applications but you also have the monograph by Tonomura. An electron holography plug-in is available for digital micrograph.

Gabor, D 1949 *Microscopy by Reconstructed Wave-Fronts: I* Proc. Roy. Soc. London **197A**, 454–487; (1951) *Microscopy by Reconstructed Wave Fronts: II* Proc. Phys. Soc. B**64**, 449–69. Not easy to find.

Harscher, A, Lang, G and Lichte, H 1995 *Interpretable Resolution of 0.2 nm at 100 kV Using Electron Holography* Ultramicroscopy **58**, 79. Pursuing higher resolution in TEM using a FEG.

Lehmann, M and Lichte, H 2005 *Electron Holographic Material Analysis at Atomic Dimensions* Crystal Research and Technology **40** 149–160.

Tonomura, A 1987 *Application of Electron Holography* Rev. Mod. Phys. **59**, 639–669; 1992 *Electron-Holographic Interference Microscopy* Adv. Phys. **41**, 59–103. Overview of his work on imaging flux lines.

Tonomura, A 1999 *Electron Holography* 2nd Ed Springer Heidelberg. A focused text on the subject.

Völkl, E, Allard, LF, Joy, DC Eds. 1999 *Introduction to Electron Holography* Springer NY. A comprehensive text on the subject

MAGNETIC MATERIALS

Chapman, JN, Johnston, AB, Heyderman, LJ, McVitie, S and Nicholson, WAP 1994 *Coherent Magnetic Imaging by TEM* IEEE Trans. Magn. **30** 4479–4484. Qualitative measurement of magnetic induction.

Petford-Long, AK and Chapman JM 2005 *Lorentz Microscopy* in *Magnetic Microscopy of Nanostructures* (Eds. H Hopster and HP Oepen) pp. 67–86 Springer New York. A recent review in a broad-ranging review text.

CL ETC.

Newbury, DE, Joy, DC, Echlin, P, Fiori, CE and Goldstein, JI 1986 *Advanced Scanning Electron Microscopy and X-ray Microanalysis*, p. 45 Plenum Press New York. Review of CL imaging. Compare to Section 29.10.

Yacobi, G and Holt, DB 1990 *Cathodoluminescence Microscopy of Inorganic Solids* Plenum Press New York.

SEM IN THE TEM

Goldstein, JI, Newbury, DE, Echlin, P, Joy, DC, Lyman, CE, Lifshin, E, Sawyer, LC and Michael, JR 2003 *Scanning Electron Microscopy and X-ray Microanalysis*, 3rd Ed., Springer New York. Whenever you think of SEM, this is *the* text.

IN SITU TEM

Butler P, Hale KF 1981 *In Situ Studies of Gas–Solid Reactions* in *Practical Methods in Electron Microscopy*, Ed. AM Glauert **9** North-Holland New York. A collection of impressive images and techniques.

Gai, PL Ed. 1997 *In-Situ Microscopy in Materials Research* Springer New York. Comprehensive set of review articles.

Rühle, M, Phillipp, F, Seeger, A and Heydenreich, J, Eds. 1994 Ultramicroscopy **56** (1-3). An issue on in-situ TEM.

Sinclair, R and Konno, TJ 1994 *In-Situ HREM: Application to Metal-Mediated Crystallization* Ultramicroscopy **56** 225–232. HRTEM in situ studies.

Wang, ZL, Poncharal, P and de Heer, WA 2000 *Measuring Physical and Mechanical Properties of Individual Carbon Nanotubes by In-Situ TEM* J. Phys. Chem. Sol. **61** 1025–1030. The combination of in situ and nano.

Special collections of papers:

J. Mater. Res. **20** 2005 *In-Situ TEM* Eds I Robertson, J Yang, R Hull, M Kirk and U Messerschmidt.

J. Mater. Sci. **41** 2006 *Characterization of Real Materials and Real Processing by Transmission Electron Microscopy* Ed. H Saka.

FLUCTUATION MICROSCOPY

Treacy, MMJ and Gibson, JM 1996 *Variable Coherence Microscopy: a Rich Source of Structural Information from Disordered Systems* Acta Cryst **A52** 212–220. The first description of fluctuation microscopy.

Treacy, MMJ, Gibson, JM, Fan, L, Paterson, DJ and McNulty, I 2005 *Fluctuation Microscopy: a Probe of Medium Range Order* Rep. Prog. Phys. **68** 2899–2944. The latest in-depth review.

COMBINED TECHNIQUES

Williams, DB and Newbury, DE 1984 *Recent Advances in the Electron Microscopy of Materials* in Advances in Electronics and Electron Physics (Ed. PW Hawkes) **62** 161–288 Academic Press New York. Lots of techniques, some of which are still relevant.

GENERAL

Bell, WL 1976 $2\frac{1}{2}D$ *Electron Microscopy: Through-Focus Dark-Field Image Shifts* J. Appl. Phys. **47** 1676–82. *The* paper on this topic.

Bell, WL and Thomas, G 1972 *Applications and Recent Developments in Transmission Electron Microscopy* Electron Microscopy and Structure of Materials (Ed. G Thomas), p. 23–59 University of California Press Berkeley CA. High-order BF imaging; not widely used but interesting.

Clarke, DR 1979 *On the Detection of Thin Intergranular Films by Electron Microscopy* Ultramicroscopy **4** 33–44. The original paper on using diffuse DF to study intergranular films.

Cowley, JM 1973a *High-Resolution Dark-Field Electron Microscopy. I. Useful Approximations* Acta Cryst. **A29** 529–536.

Cowley, JM 1973b *High-Resolution Dark-Field Electron Microscopy. II. Short-Range Order in Crystals* Acta Cryst. **A29** 537–540.

Cowley, JM 1992 *Twenty Forms of Electron Holography* Ultramicroscopy **41** 33–348.

Franke, F-J, Hermann, K-H and Lichte, H 1988 in *Image and Signal Processing for Electron Microscopy* (Eds. PW Hawkes, FP Ottensmeyer, A Rosenfeld and WO Saxton), p. 59 Scanning Microscopy Supplement **2** AMF O'Hare IL.

Hudson, B 1973 *Application of Stereo-Techniques to Electron Micrograph* J. Microsc. **98** 396–401. A review of stereomicroscopy in the TEM.

Imeson D 1987 *Studies of Supported Metal Catalysts Using High Resolution Secondary Electron Imaging in a STEM* J. Microsc.**147** 65–74. More on SE imaging in the STEM.

Joy, DC, Maher, DM and Cullis, AG 1976 *The Nature of Defocus Fringes in Scanning-Transmission Electron Microscope Images* J. Microsc. **108** 185–193. Topographic contrast in the TEM (Section 29.6.B).

Russ, J 1990 *Computer-Assisted Microscopy: The Analysis and Measurement of Images* Plenum Press New York.

Simpson, YK, Carter, CB, Morrissey, KJ, Angelini, P and Bentley, J 1986 *The Identification of Thin Amorphous Films at Grain-Boundaries in* Al_2O_3 J. Mater. Sci. **21** 2689–96. An early paper comparing different techniques for studying intergranular films.

Spence, JCH 1999 *The Future of Atomic-Resolution Electron Microscopy for Materials Science* Mat. Sci. Eng. **R26** 1–49.

COMPANION TEXT

Four complete chapters in the companion text are devoted to topics described in this chapter: electron holography (Lichte and Lehmann), magnetic specimens (Petford-Long and de Graef), and in situ TEM (CBC). Chemically sensitive imaging is becoming much more commonplace now but using more AEM combined with imaging. A fourth chapter is dedicated to tomography.

SELF-ASSESSMENT QUESTIONS

Q29.1 Make a list of all the imaging techniques in this chapter and determine the electron-scattering mechanism(s) operating in each case.
Q29.2 Make a list of all the imaging techniques in this chapter and determine the contrast mechanism(s) operating in each case.
Q29.3 Make a list of all the imaging techniques in this chapter and determine the TEM operating mode(s) operating in each case.
Q29.4 List the pros and cons of stereo versus tomographic approaches to the TEM image-projection problem. Why is stereo more difficult for diffraction-contrast experiments than mass-thickness contrast? Why is a phase-contrast, atomic-resolution stereo experiment effectively impossible?
Q29.5 Why is SE imaging much more useful than BSE imaging in STEM while both are equally useful in SEM?
Q29.6 Search the Web and determine how many different kinds of experiments can be performed *in situ* in a TEM.
Q29.7 List the pros and cons of a typical in-situ experiment.
Q29.8 Why is $2\frac{1}{2}$D imaging not called 2D?
Q29.9 If your specimen is magnetic, how does this affect your choice of preparation method?
Q29.10 If your specimen is magnetic, how does this affect your choice of TEM versus STEM mode?
Q29.11 If your specimen is magnetic, how does this affect your choice of DF imaging method?
Q29.12 Magnetism is becoming increasingly important in nanotechnology. Why is nanomagnetism easier to study in a (S)TEM than a conventional ferromagnetic steel specimen?
Q29.13 Who were Lorentz, Fresnel, and Foucault (yes; go and study some history of science) and why are their names given to specific forms of TEM of magnetic specimens?
Q29.14 Discuss the factors limiting the resolution in Foucault and Fresnel images. Why is this resolution important in nanomagnetism studies?
Q29.15 Distinguish Bloch, Néel walls, and cross-tie walls and how best to image them in a TEM.
Q29.16 Why would you use chemically sensitive imaging techniques when XEDS or EELS spectrum imaging (see Part 4) can reveal the distribution of every element in your specimen down to the atomic level?
Q29.17 Find out every way possible to explore the structure and chemistry of glasses that is discussed in this book. List the pros and cons of each approach.
Q29.18 Why is DADF preferable to CDF when forming images with diffusely scattered electrons?
Q29.19 Why do REM in your TEM when you can do SEM in your STEM?
Q29.20 How can we image atomic-level detail in REM images when the TEM is not being operated in phase-contrast HRTEM mode?
Q29.21 Compare and contrast HOBF and HODF imaging. What is the more common name for HODF?

TEXT-SPECIFIC QUESTIONS

T29.1 Why would you be surprised to find any examples in the literature of BSE imaging in a STEM?
T29.2 If you want to watch nanotube electronic circuits operating *in situ* in a TEM what technique(s) would you use? Could you do this in a SEM?
T29.3 Why would Denis Gabor be delighted if he were still alive? Why was he a 'microscopist before his time' and what experimental development helped most to realize his dream?
T29.4 Explain how a biprism works and what you give up if you put one into your TEM column.
T29.5 Suggest why fluctuation microscopy is not common in the Far East.
T29.6 Why in fluctuation microscopy is the *variance* of the intensity distribution used to form images rather than the intensity distribution itself?
T29.7 Why is fluctuation microscopy not simply described as a subset of diffuse-scatter imaging?
T29.8 Search the Web to find at least one new TEM imaging or diffraction technique *not* described in this chapter or anywhere else in the text. What does your search tell you about TEM as an experimental materials-characterization tool?
T29.9 Look at Figure 29.1 and try and cross your eyes until the left and right images overlap. You then may be able to see the stereo effect without using a stereo viewer. To see the effect, you may have to tilt your head to get exact image overlap (why?) and you will still see de-focused image of the two originals (why?).
T29.10 Look at Figure 29.2. Use this to explain why $2\frac{1}{2}$D microscopy effects can better be achieved with more recent developments in high spatial resolution diffraction.

T29.11 List all the information about magnetic specimens that is discernible in Figures 29.3–29.7. What can you not find out about magnetism in TEM images?
T29.12 Why do the dislocation images get (a) sharper and (b) fainter across the sequence shown in Figure 29.12?
T29.13 Explain why you can see the structure inside the amorphous carbon support film in Figure 29.11.
T29.14 What are the lines in Figure 29.10B? Explain how they arise.
T29.15 Draw a similar diagram to Figure 29.13 for a thin specimen and explain how the different SE contributions vary between SEM and STEM SE images.
T29.16 Find examples in the literature of SE imaging in a STEM for small particles on a substrate and explain the advantages over conventional TEM imaging.
T29.17 Some scientists believe strongly that Tonomura's work in Figure 29.20 is worthy of the Nobel Prize. Justify this claim.
T29.18 Explain why Figure 29.21 is an extraordinary sequence of images.

30

Image Simulation

CHAPTER PREVIEW

When we need to obtain information about the specimen in two directions, we need to align the specimen close to a low-index zone axis. If the HRTEM image information is going to be directly interpretable, the specimen must be oriented with the incident beam exactly aligned with both the optic axis of the TEM and the zone axis of the specimen. Thus, we will have many reflections excited and the simple two-beam analysis of Chapter 27 cannot be used.

A method for modeling the contrast of images obtained under these conditions was developed by Cowley, Moodie, and their co-workers, principally at Melbourne and Arizona State University (ASU), in a series of classic papers beginning with that by Cowley and Moodie in 1957. Fortunately, the growing interest in HRTEM has coincided with the availability of increasingly powerful computers which can handle the extensive calculations.

There are several software packages available commercially, so there is little reason for most users to re-invent the wheel. However, the different packages do not necessarily perform the calculations in the same way: one may be more appropriate for your application than others. Since these packages essentially operate as 'black boxes' it is also reassuring to simulate images from the same structure using different packages (unless they don't give the same answers).

One point to keep in mind as you work through the literature is that this subject already has a lot of history. We will point out some of the things that have been done, and in some cases continue to be done, for historical reasons.

30.1 SIMULATING IMAGES

The idea of simulating HRTEM images arose because of the realization that the loss of phase information when we form an experimental intensity map means that we can't go back from the image to the structure. Instead, we assume a structure (perfect crystal or crystalline material containing defects), simulate the image, see how closely the simulated image resembles the experimental image, modify the structure, and repeat the process. The only difficulty is that the image is sensitive to several factors

- The precise alignment of the beam with respect to both the specimen and the optic axis.
- The thickness of the specimen (as we saw in Chapter 28).
- The defocus of the objective lens.
- Chromatic aberration which becomes more important as t, the thickness, increases.
- Coherence of the beam.
- Other factors: one example would be the intrinsic vibration in the material which we try to take account of through the Debye-Waller factor.

In principle, we could have the same image from two different structures. So obviously, this is the tricky part!

30.2 THE MULTISLICE METHOD

The basic multislice approach used in most of the simulation packages is to section the specimen into many slices, which are normal to the incident beam.

There are different methods for actually performing the multislice calculation. The different approaches have been developed for several reasons. Some try to optimize the use of available hardware. Others were written with the intention of providing a convenient method of simulating DPs using the same program. At least one package was written to make use of a popular

TABLE 30.1. Software

Cerius2	Runs on UNIX; by Accerlys. Look for c2_hrtem www.accelrys.com/products/cerius2/cerius2products/hrtem.html
EMS and jEMS	By Pierre Stadleman. Very widely used and user-friendly. Does Bloch wave and conventional multislice calculations. Is used by CuFour (see Chapter 24). Multi-platform including Mac, Unix, Windows. The 'j' denotes the java version
Kirkland	By Earl Kirkland. Well described in his book on the subject
MacTEMPAS	By Roar Kilaas. Runs on a Mac so it's very user-friendly. NCEMSS was the Unix version that was produced by NCEM
SHRLI81	By Mike O'Keefe. This program made image simulation freely available to everyone and introduced a generation to HRTEM simulation. Unix only, but no longer supported
WinHREM and MacHREM	www.hremresearch.com/Eng/download/documents/HREMcatE.html by HREM Research Inc. (Kazuo Ishizuka)

personal computer with a user-friendly interface (see Table 30.1). The principal methods for performing these calculations are

- The reciprocal-space formalism.
- The FFT formalism.
- The real-space approach.
- The Bloch-wave approach.

We'll go through the special features of each approach. The software packages which are readily available are listed in Section 1.6 and in Table 30.1.

30.3 THE RECIPROCAL-SPACE APPROACH

We project each slice onto a plane somewhere in the slice (usually top, bottom, or middle) giving a projected potential for that slice, and we call this the phase grating. We then calculate the amplitudes and phases for all the beams which will be generated by the incident beam interacting with the first projection plane. We could think of this as being a many-beam image calculation for a single slice. We then allow all these beams to propagate down the microscope in free space until they meet the next phase grating, which need not be the same as the first. The scattering calculation is now repeated for all the beams incident on this plane. This calculation produces a new set of beams which propagate through free space to the next phase grating, and so on. The process is summarized in Figure 30.1.

One point which you must remember: scattering by the phase grating does not just produce Bragg beams. It is crucial to keep track of the scattering in *all* directions. All of these beams will be incident on the next phase grating. So we don't just have Bragg beams, we *sample* all of reciprocal space.

A calculation based on a 128 × 128 array will impose a limit of ~4096 on the number of 'beams' which can be included in the calculation. This number might appear large, especially when you form a [110] HRTEM image of Si with six Bragg beams (plus the O beam) but, particularly for imperfect crystals, this number will be inadequate.

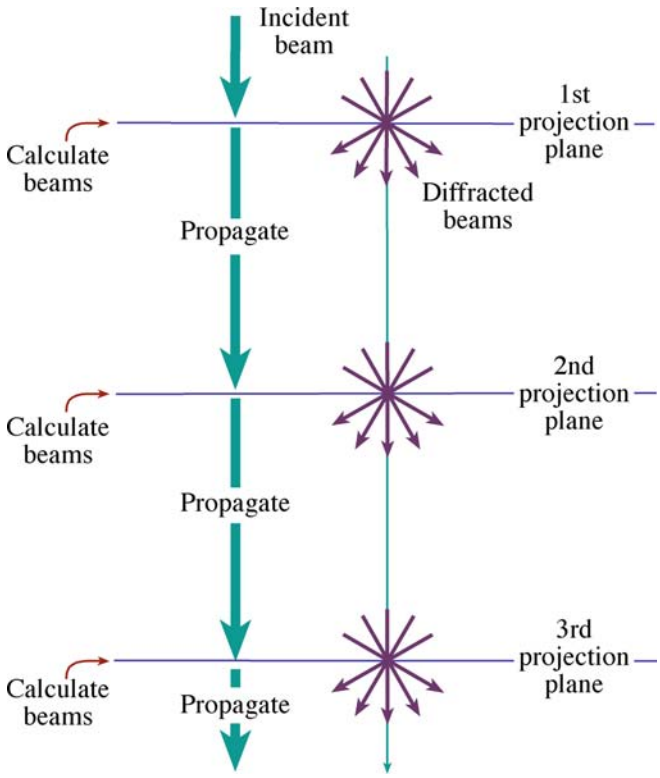


FIGURE 30.1. The potential within a slice is projected onto the first projection plane; this is the phase grating. We calculate the amplitudes and phases for all the beams generated by interacting with this plane and then propagate all the diffracted beams through free space to the next projection plane, and repeat the process.

k-SPACE BETWEEN THE BRAGG BEAMS

Why do we need to consider regions of **k**-space between the Bragg beams? In other words, why do we need to sample all of reciprocal space? The answer is that the Bragg beams contain information about the periodic structure, but all of the information from defects, i.e., non-periodic structure, is contained *between* the Bragg spots, though it will generally be quite close to them.

Essentially the multislice method considers three components

- ψ describes the *electron* wave.
- P is the propagation of the electron wave in free space: the *microscope*.
- Q is the phase grating: the *specimen*.

The process can be described by this equation

$$\psi_{n+1}(\mathbf{k}) = [\psi_n(\mathbf{k}) P_{n+1}(\mathbf{k})] \otimes Q_{n+1}(\mathbf{k}) \qquad (30.1)$$

where $\psi_{n+1}(\mathbf{k})$ is the wave function in reciprocal space at the exit of the $n+1$ slice and the symbol $\otimes$ denotes a convolution as before. $P_{n+1}(\mathbf{k})$ is the propagator for the $n+1$ slice. In other words, this is expressing the Fresnel diffraction phenomenon for this one slice because we are making a near-field calculation. (Look back to Section 2.8 for a discussion of near-field versus far-field.) Similarly, $Q_{n+1}(\mathbf{k})$ is the phase-grating function; it is a transmission function, for the $n+1$ slice.

The three functions $\psi(\mathbf{k})$, $P(\mathbf{k})$, and $Q(\mathbf{k})$ are all functions in reciprocal space, so this approach is referred to as the reciprocal-space formulation. Notice that they are all two-dimensional arrays. We can think of the different terms as being diffracted beams within the specimen. We can easily insert a circular objective aperture of radius $\mathbf{r}$; we just require that all values of $\psi(\mathbf{k})$ are zero for $\mathbf{k} > \mathbf{k_r}$.

To give you an idea of the complexities involved, consider what values of $Q(\mathbf{k})$ you must use in the calculation. $Q(\mathbf{k})$ must go out twice as far as $\psi(\mathbf{k})$ or $P(\mathbf{k})$ in reciprocal space. You can understand why by considering Figure 30.2. If you represent the number of

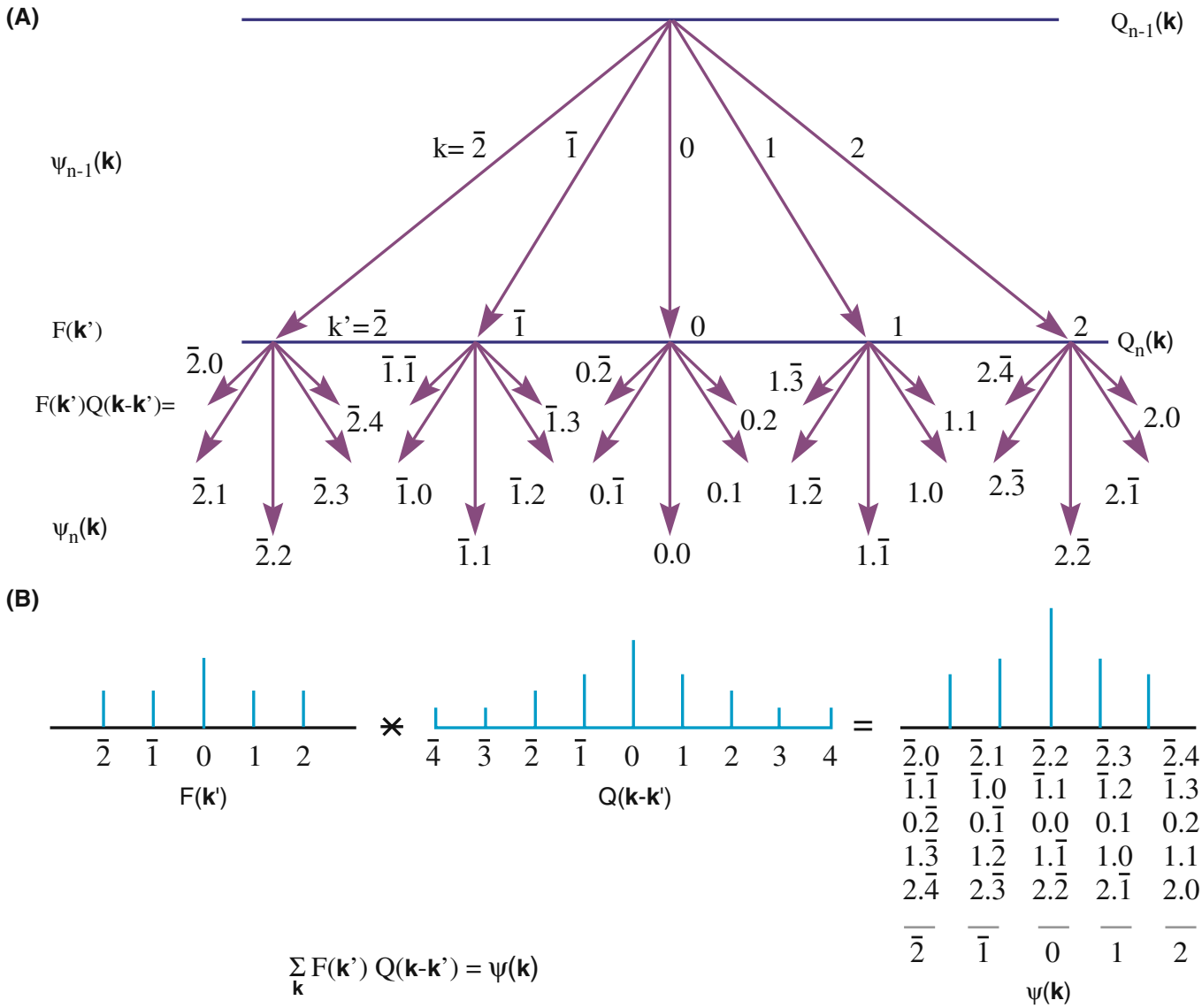


FIGURE 30.2. (A) Schematic used to explain why, in the one-dimensional case, $Q(\mathbf{k})$ must take account of twice as many $\mathbf{k}$ values as $\psi(\mathbf{k})$ or $P(\mathbf{k})$. Consider wave $k = \bar{2}$ from $Q_n(\mathbf{k})$: to produce wave at $+2$ at this point you need to add 4 to $\bar{2}$ and similarly for every possible wave in slice $Q_n(\mathbf{k})$. As summarized in (B) $Q(\mathbf{k}–\mathbf{k}')$ extends from $\bar{4}$ to $+4$ so that $\psi(\mathbf{k})$, which we want, extends from $\bar{2}$ to $+2$, including all possible combinations of $\mathbf{k}'$ and $\mathbf{k}$.

beams from slice $Q_{n-1}(\mathbf{k})$ as $F(\mathbf{k}')$ then $Q(\mathbf{k}–\mathbf{k}')$ must go out to $k = -4$ because when you multiply these two functions to give $\psi(\mathbf{k})$ you can produce $k = -2$ by using $k = -4$ in Q and $k = +2$ in F as in Figure 30.2B. Putting this into an equation we have

$$\sum_{\mathbf{k}'} F(\mathbf{k}')Q(\mathbf{k}-\mathbf{k}') = \psi(\mathbf{k}) \qquad (30.2)$$

where

$$F(\mathbf{k}) = \psi(\mathbf{k})P(\mathbf{k}) \qquad (30.3)$$

The function $Q(\mathbf{k})$ is a 'probability map.' What we are doing here is using the convolution to describe multiple scattering.

We can illustrate the complexity of the calculation by considering a 128 × 128 array for $Q(\mathbf{k})$ using SHRLI81 (see Table 30.1). The maximum value for (k_x,k_y) is only (31,31) but even so, the number of diffracted beams is nearly 4096. Remember, we usually just use the seven inner Bragg beams in, e.g., the Si <110> DP, as we saw in Figure 27.3; most of the beams in our calculation are not Bragg beams. However, you will remember that the information concerning defects in crystals is contained in the regions between the Bragg spots in the DP, so it does make sense. Specific examples of $Q(\mathbf{k})$, including numerical computations of the phase change per slice, are given by John Barry.

30.4 THE FFT APPROACH

We can recast equation 30.3 to maximize the efficiency of the computer in using fast Fourier transform (FFT) routines. In equation 30.4, the notation F and F^{-1} tell us to take the Fourier transform or the inverse transform of the function inside the brackets.

$$\psi_{n+1}(\mathbf{k}) = F\{F^{-1}[\psi_n(\mathbf{k})P_{n+1}(\mathbf{k})]q_{n+1}(\mathbf{r})\} \qquad (30.4)$$

In this equation, $q_{n+1}(\mathbf{r})$ is the real space form of $Q_{n+1}(\mathbf{k})$, i.e., it is the inverse Fourier transform of $Q_{n+1}(\mathbf{k})$. So $q(\mathbf{r})$ is a real-space phase grating. Now we can look at some numbers for the calculation and take $Q(\mathbf{k})$ as a 128 × 128 array to keep the calculation small. The main steps carried out by the computer are

- Multiply $\psi_n(\mathbf{k})$ by $P_{n+1}(\mathbf{k})$: that is a 64 × 64 array times another 64 × 64 array. Remember that we are limited to 64 points, if we have a 128 × 128 array, because the Q array must be twice as large in all directions in $\mathbf{k}$ space.
- Take the inverse Fourier transform of the result.
- Multiply this new result by $q_{n+1}(\mathbf{r})$ which is the 128 × 128 array.
- Fourier transform the final result and set all values outside the inner 64 × 64 array equal to zero so that we can repeat the process for the next slice.

You will notice that this example used a square array. In modern programs, we are not restricted even to using powers of 2 but this helped the original FFT routines. You will see the value of this advance when we examine some defect calculations later. If you are interested in the mechanics of the FFT routine and other aspects of this simulation approach, the article by Mike O'Keefe and Roar Kilaas is required reading.

30.5 THE REAL-SPACE APPROACH

As we noted earlier, image simulation used to be limited by your budget, i.e., by your computer. The real-space approach was developed, in part, to decrease the time needed for the calculations by using our knowledge that $P(\mathbf{r})$ is strongly peaked in the forward direction. In our notation, the method developed by Coene and Van Dyck for calculating $\psi(\mathbf{r})$ can be expressed by the equation

$$\psi_{n+1}(\mathbf{r}) = [\psi_n(\mathbf{r}) \otimes P_{n+1}(\mathbf{r})]q_{n+1}(\mathbf{r}) \qquad (30.5)$$

where $P_{n+1}(\mathbf{r})$ is now the propagator in real space and $q_{n+1}(\mathbf{r})$ is again the real-space phase grating. Once you have written this, it's all computing, which is a substantial task since the size of the multislice calculation is the size of the largest array, i.e., $Q(\mathbf{k})$ or $q(\mathbf{r})$.

30.6 BLOCH WAVES AND HRTEM SIMULATION

Although we saw in Chapters 14 and 15 that electrons propagate through crystalline specimens as Bloch waves, the multislice method we've described so far is essentially a 'diffracted-beam' approach. In two classic papers Fujimoto (1978) and Kambe (1982) showed that, for the perfect crystal, the HRTEM image may be understood simply in terms of images of Bloch waves. The key point is that, although a large number of diffracted waves are formed, only a small number of Bloch waves determine the appearance of the image, providing the crystal has a sufficiently high symmetry. Following Kambe's 'simple' example, we consider the case where only three Bloch waves i, j, and k are significant. Let's assume that Bloch waves i and j are in phase at a thickness $z = D$. Then we have

$$e^{ik_z^{(i)}z} = e^{ik_z^{(j)}D} \qquad (30.6)$$

Using our expression for ψ, namely

kTH BLOCH WAVE
Don't confuse the *k*th Bloch wave with the **k**-vector!

$$\psi(\mathbf{r}) = \sum_i C^{(i)}\phi^{(i)}(x,y)e^{ik_z^{(i)}z} \tag{30.7}$$

and the normalization rule

$$\sum_i C^{(i)}\phi^{(i)}(x,y) = 1 \tag{30.8}$$

we can therefore express ψ at $z = D$ in terms of our three Bloch waves.

$$\psi(x,y,D) = \left[C^{(i)}\phi^{(i)} + C^{(j)}\phi^{(j)}\right]e^{ik_z^{(i)}D} + C^{(k)}\phi^{(k)}e^{ik_z^{(k)}D} \tag{30.9}$$

We rearrange this equation so that we can extract the phase factor $e^{ik_z^{(i)}z}\left(= e^{ik_z^{(i)}D}\right)$. We write

$$\psi(x,y,D) = \left[1 - C^{(k)}\phi^{(k)}\right]e^{ik_z^{(i)}D} + C^{(k)}\phi^{(k)}e^{i\left(k_z^{(k)}-k_z^{(i)}\right)D}e^{ik_z^{(k)}D} \tag{30.10}$$

$$\psi(x,y,D) = e^{ik_z^{(i)}D}\left[1 + \beta_{ik}(D)C^{(k)}\phi^{(k)}\right] \tag{30.11}$$

where we've defined a new parameter β given by

$$\beta_{ik}(D) = e^{i\left(k_z^{(k)}-k_z^{(i)}\right)D} - 1 \tag{30.12}$$

These equations tell us that if any two of the Bloch waves (here they are i and j) are in phase then the amplitude of the wave at the exit surface is determined by the third Bloch wave.

If the third Bloch wave is also nearly in phase, we have a relation like equation 30.6 but with i, j, and k all equal. Then we can approximate $\beta_{ik}(D)$ by

$$\beta_{ik}(D) = i\left[\left(k_z^{(k)} - k_z^{(i)}\right)D + 2n\pi\right] = i\gamma_{ik}(D) \tag{30.13}$$

Now we've defined another factor γ_{ik}. If you plug this expression back into equation 30.11, you see we have a pure phase object. All the diffracted beams will be shifted in phase by $\pi/2$.

Now you can test the effects of how we change k. Consider what conditions this will really correspond to using equations 30.11 and 30.13

- If k is such that the phase of Bloch wave k is ahead of i and j (which were equal), then you'll see a 'negative' image of $C^{(k)}\phi^{(k)}$. A 'delayed' k gives us the 'positive' image.
- For the Ge <110> zone axis, HRTEM image at 100 kV, only three Bloch waves are strongly excited.

The relationship to the Bloch-wave contours in Chapter 14 is clear. Using this information and the projected potential shown in Figure 30.3, Kambe calculated the Bloch-wave amplitudes and the two ideal images of the Bloch waves: one positive and the other negative. In the calculation of different images for increasing thickness, several images corresponding to a single Bloch wave can be predicted and identified, as shown in the figure. At other thicknesses, the images form by a combination of Bloch waves. So, what can we learn?

- For a perfect crystal, you may need as few as three Bloch waves to give the essential features of an HRTEM zone-axis image.
- There is a direct connection between the WPOA (go back to the definition if you don't remember) and the propagation of Bloch waves.

THE MULTISLICE APPROACH
We usually use this to simulate HRTEM images. It is actually a very elegant form of brute force.

We saw in Chapter 14 that the electron propagates as Bloch waves inside the crystal. The reason we don't use Bloch waves is that our specimens are not perfect. However, EMS does give you the option of using this approach.

30.7 THE EWALD SPHERE IS CURVED

When you are using the TEM, some other complications arise because the Ewald sphere is curved

- If you align the beam exactly parallel to a zone axis, **s** will be non-zero for every Bragg reflection. In fact, it will also be different for each type of reflection.
- If you do not align the beam exactly parallel to the zone axis, then **s** will also be slightly different for each reflection in that zone.
- If you change the wavelength of the electrons, the radius of the sphere changes.
- If you converge the beam, then you'll add a thickness to the Ewald sphere.

The point is, knowing precisely what the correct values are to put in the program will also require thought and work.

30.8 CHOOSING THE THICKNESS OF THE SLICE

So far, we've just cut the specimen into slices in the computer without considering how thick each slice should be, or even whether they should all be the

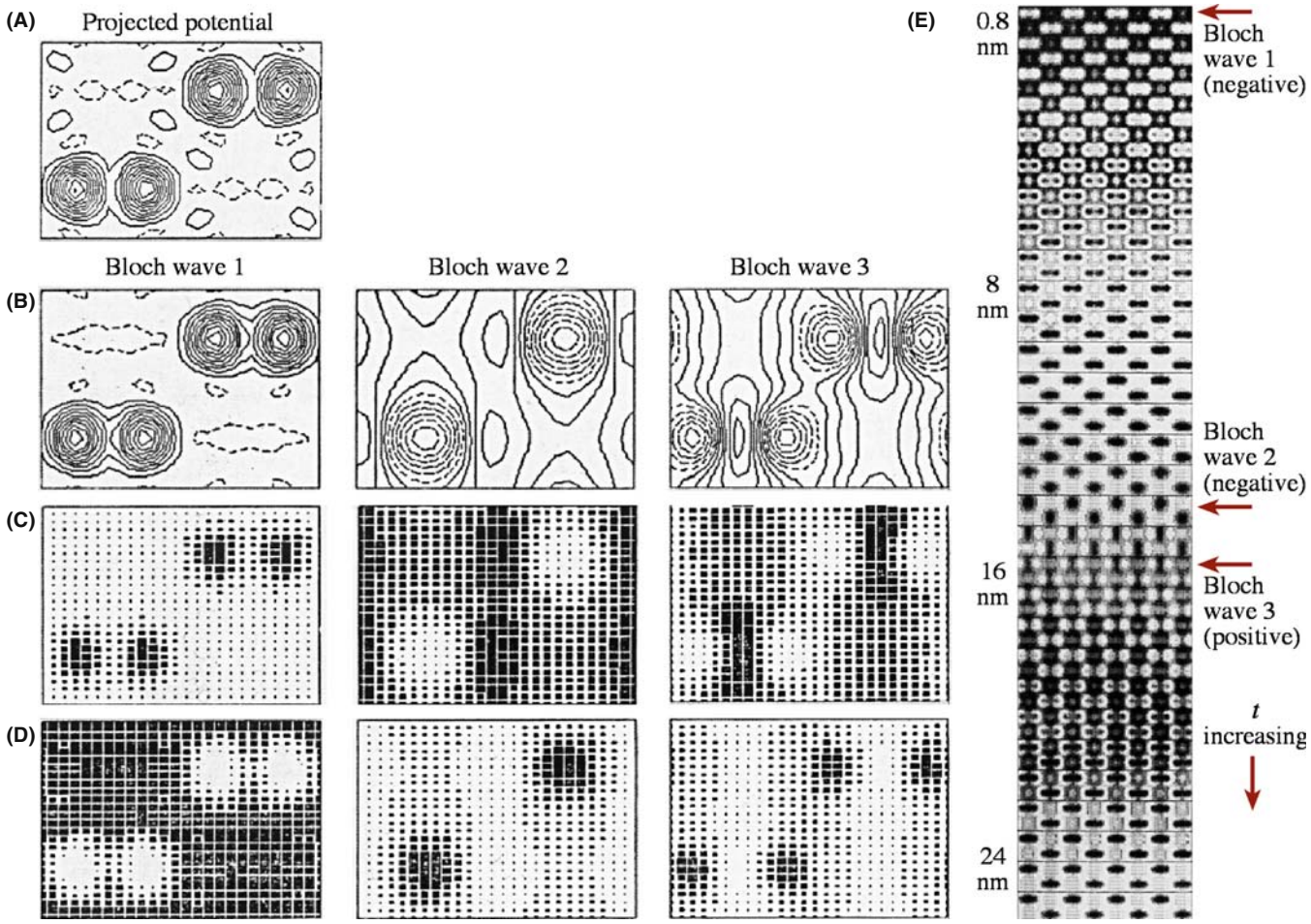


FIGURE 30.3. (A) The projected potential for Ge where the contour lines represent changes in potential of –10 eV, and the dashed lines are positive values. (B–D) are for the three Bloch waves 1, 2, and 3 for 100-keV electrons: (B) the amplitudes, (C) ideal positive images, (D) ideal negative images. (E) Calculated lattice images showing thickness dependence.

same. If all the slices are the same, then there can be no information about the z-direction. Although HOLZ lines are not important for the simulation of HRTEM images, some of the programs we are discussing can now just as readily be used to simulate CBED patterns and HOLZ lines. So, following the philosophy of attacking problems with different techniques, you should be aware of these limitations, since it is easy to overlook the simplifications you made once you see the computed image. You should remember that when you are studying a material with a large unit cell, the reciprocal lattice spacing will be short in the beam direction, so HOLZ effects come into play sooner.

Consider the different methods for making the slice

- You could calculate the projected potential for a thick slice and then do n calculations with slices which are $1/n$ times this thickness.
- A better approach would be to subdivide the cell into layers of atoms, create a different grating for each of these layers, and then run the program with the sequence.

For example, if the beam is aligned along the [111] direction of an fcc crystal, then you would have three identical gratings displaced relative to one another, giving the ABC stacking of close-packed planes. This approach would allow you to test for the effect of a real error in the stacking sequence normal to the beam. Even this point can be a bit difficult. In general, you orient the beam to be parallel to a particular zone axis $[UVW]$ so that the planes in that zone are parallel to the beam (so our projection works). If the material is not cubic, we will not generally have a low-index plane normal to the beam to make this slice.

30.9 BEAM CONVERGENCE

When you are recording HRTEM images, you need to keep exposure times short. So, if you don't use parallel illumination, you have to take account of the beam convergence when simulating the images. O'Keefe and Kilaas (see also the paper by Self and O'Keefe) have developed one approach to address this problem. If

the beam actually has some convergence, then the diffraction spots will be disks, as illustrated in Figure 30.4, so you need to simulate disks in the DP. Experimentally, the large objective aperture admits many disks, so in the simulation routine you should sample each disk at many points. This means the program needs to calculate the image at each of these convergence angles and average all the resulting images. Of course, the objective aperture is easily applied in the computer. If you choose 49 points, you can make the sampling interval in reciprocal space $\leq 0.1\,\mathrm{nm}^{-1}$. It is instructive to examine just how much work is necessary to sample the 49 points.

We can start by writing the usual expression for the χ, the phase change due to the objective lens

$$\chi = \pi \Delta f \lambda u^2 + \pi C_s \lambda^3 \left(\frac{u^4}{2}\right) \tag{30.14}$$

Then differentiate this with respect to the variable u

$$\frac{d\chi}{du} = 2\pi\left(\lambda u \Delta f + C_s \lambda^3 u^3\right) \tag{30.15}$$

This equation tells us that if u changes by δu, then χ changes by

$$\delta\chi = 2\pi\lambda\left(u\Delta f + C_s\lambda^2 u^3\right)\delta u \tag{30.16}$$

Now we choose $\delta\chi$ so that

$$\delta\chi < \frac{2\pi}{n} \tag{30.17}$$

where n will allow us to determine the maximum change in χ between two points in the disk. For example, if $n = 12$, then the maximum value of $\delta\chi$ is 30°. Combining equations 30.15 and 30.17, we can write

$$\delta u = \left[n\lambda u\left(\Delta f + C_s\lambda^2 u^2\right)\right]^{-1} \tag{30.18}$$

If we plot χ versus u (or play with equation 30.15 and its derivative), then we find a minimum in χ at

$$\Delta f = -C_s\lambda^2 u^2 \tag{30.19}$$

and an inflection at

$$\Delta f = -3C_s\lambda^2 u^2 \tag{30.20}$$

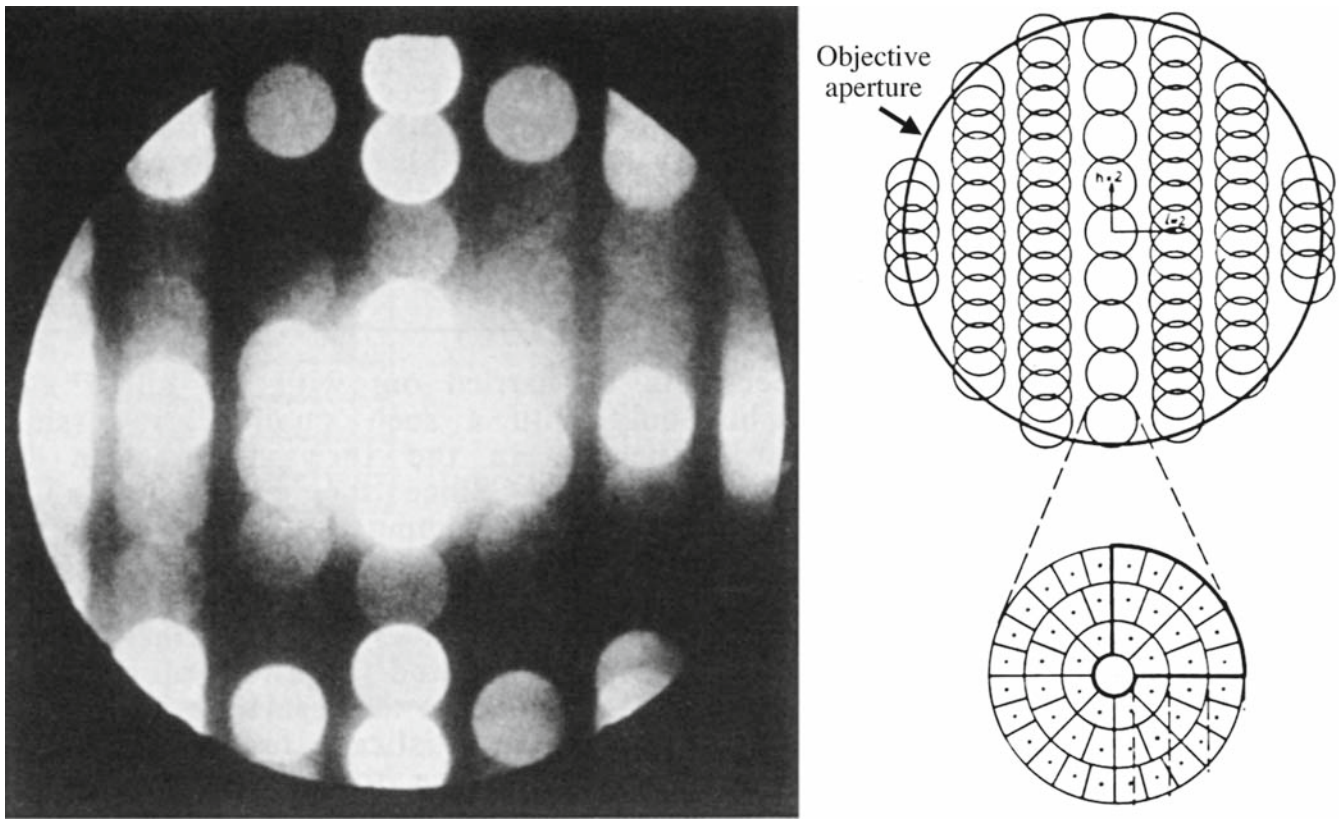


FIGURE 30.4. Disks in the DP from a crystal of $Nb_{12}O_{29}$. The computer simulation can divide each disk into many sectors and simulate the image for each sector, as shown in the schematic, excluding sectors which are intersected by the objective aperture.

So the simulation program can check to find the smallest δu at an inflection point, which equations 30.18 and 30.20 tell us is

$$\delta u = -\left[\frac{27C_s}{(\Delta f)^3}\right]^{1/2}\left(\frac{1}{2n}\right) \qquad (30.21)$$

The value of δu therefore depends on both C_s and Δf.

THE BLACK BOX
Remember where all these calculations takes place!

You can also appreciate the relevance of this type of approach if your disks actually intersect the objective aperture, as shown in Figure 30.4. Put another way, you can learn two lessons from this analysis

- Always try to minimize the convergence of the beam when recording HRTEM images.
- Use an aperture which does *not* cut through the diffraction disks.

30.10 MODELING THE STRUCTURE

To simulate any HRTEM image, you need a unit cell. If you are concerned only with perfect crystals, then your program should have all the space groups already included so that you only need to add the lattice parameters (lengths and angles) and the occupied sites for your material. If you are interested in simulating images from defects, then you have to create a new unit cell which must be sufficiently large that it will not add effects due to the edges. There are many ways to create this defect unit cell. You can input from other programs, such as those performing atomistic modeling of defects, or create your own starting structure. In either case, you will need to move atoms, either manually or following a rule you've selected for image matching, to optimize the match between your experimental series of through-focus images and the simulated images.

At some stage, you will find it useful to combine different slices, as when simulating GBs with or without a surface groove, or modeling large complex unit cells using a multilayer approach. We'll now go through some specific features of this task and return to modeling in Chapter 31 when we discuss quantitative HRTEM.

30.11 SURFACE GROOVES AND SIMULATING FRESNEL CONTRAST

The analysis of interfaces by the Fresnel-fringe technique, which we introduced in Chapter 23, illustrates the importance of image simulation and emphasizes that it is not just for HRTEM. The calculation is complicated for several reasons, as shown in Figure 30.5A

- The potential change at the interface is probably not abrupt.
- The potential depends on the detailed structure of the interface.
- During preparation, TEM specimens may be preferentially damaged at GBs, giving rise to surface grooves.

If you use a thicker specimen, you'll reduce the effect of surface grooves on any Fresnel fringes, but in practice, your foil thickness is usually limited (~20 nm), since you need to view the boundary exactly edge-on.

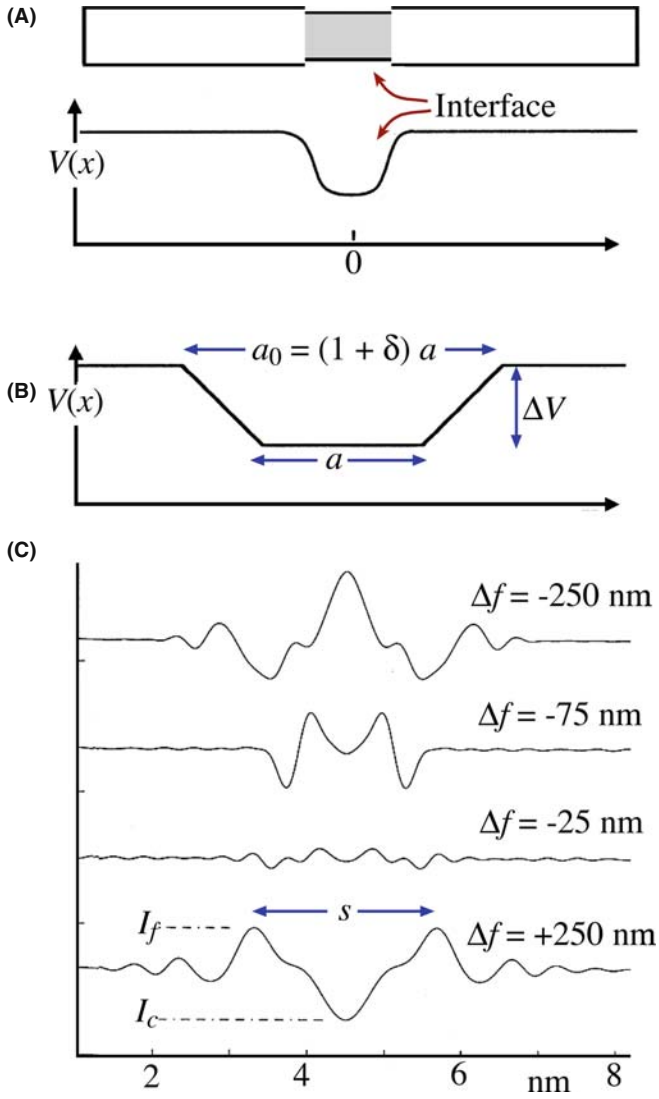


FIGURE 30.5. (A) Schematic of a GB containing a layer of material with a different inner potential; (B) one model used to represent such a GB giving variable parameters a, a_0, and δ; (C) a typical set of simulated Fresnel-fringe intensity profiles at increasing Δf: s is the distance between the first two fringes, I_c and I_f are the intensities of the central and first fringe, respectively.

Even for foils this thick, surface grooves can influence the projected potential considerably. If we assume that the bulk has a mean inner potential $V = 20$ V, and take a typical potential drop for an intergranular film to be ~1 V, then the total projected potential drop for a 20 nm thick foil would be the same as that caused by a pair of grooves at the top and bottom surfaces which are only 0.5 nm deep. Although the surface groove may be partly filled with a second phase, the effect on the Fresnel fringes can still be substantial.

We can examine Fresnel fringes using different methods. In all of them, we describe the potential at the interface in terms of the projected potential drop $\Delta V_p = t\Delta V$, an inner width a, and an outer width a_0, and a 'diffuseness,' δ, defined by

$$a_0 = (1 + \delta)a \tag{30.22}$$

These parameters are shown in Figure 30.5B. Then we construct models of a foil with a surface groove at the edge-on interface by combining such potentials.

The models: Values of $\delta = 0.5$ and $\delta = 0.2$ represent shallow and steep surface grooves, respectively. The total projected potential drop can be due to a real change in V or a change in t. A groove without a film implies $a = 0$. If $a = 1$ nm and $a_0 = 1.5$ nm, the model could correspond to two different situations

- If the atoms at the interface relax, then the atomic density at the interface will usually be reduced. This occurs at both structured interfaces and those where a layer of glass is present.
- Surface grooves at the interface.

What image simulation shows is that the relative shapes and sizes of these models are more important than the actual dimensions. Therefore, we can give most of the following analysis in terms of dimensionless quantities. Inner potentials are typically 5–10 eV. Except for very small defocus values ($\Delta f < \approx t$), we find that the distribution of the potential through the foil is not important. Usually, the projected potential at the interface is lower than that in the bulk. However, the opposite situation can occur, e.g., when a Bi_2O_3-rich phase is present at interfaces in ZnO. When we discuss the calculated profiles, the term 'interface width' will be used for the parameters a and a_0, whether they actually correspond to an intergranular film, a surface groove, or otherwise.

SPACING OF FRESNEL FRINGES

The distance from the edge to the first fringe is proportional to $(\lambda\Delta f)^{1/2}$. The fringe spacing, s_f, can be extrapolated to zero defocus to obtain the interface width, based on the relation $(s_f - a) \propto \sqrt{(\lambda\Delta f)}$.

This relation was first described by Clarke and only holds when a is large and Δf is relatively small; then the fringes from each 'edge' at the interface are independent. We observe the minimum fringe spacing at small values of defocus and this spacing can be used to provide a measure of the interface width. For more details on the simulation of Fresnel fringes, you should check the original articles.

In practice, the analysis of Fresnel fringes is impaired, not only by specimen artifacts such as surface grooves, but also by various sources of noise, which all add to the uncertainty of measurements, especially at small values of Δf. For diffuse interfaces, the contrast decreases rapidly as Δf approaches zero (Figure 30.5C), and measurements of the fringe spacing for $\alpha > 0.7$ are increasingly susceptible to noise and artifacts. You can always use larger defocus values and thus obtain higher contrast. However, without prior knowledge about the shape of the potential drop (its 'diffuseness'), we can't reliably determine the interface width by measuring the fringe spacing alone. Since the fringe spacing is dominated by the outer width, a_0, we may easily overestimate the interface width. The atomic density in a region close to the boundary is also often reduced, even if the boundary is structured, so you can easily misinterpret the image as showing the presence of an intergranular film when actually there is no amorphous film.

The region of defocus, where the central fringe shows little contrast, provides complementary information to the fringe spacings, so it is more sensitive to the inner width.

From this discussion, you'll appreciate that, before you can completely understand the effect of any intergranular films, you must estimate the extent to which surface grooves are present in your specimen. Shadowing (e.g., using platinum or gold) may provide evidence for surface grooves, but in the case where the surface groove is already filled (e.g., if your specimen was coated with carbon or contaminated without your knowing), this technique won't work.

CHARACTERIZING A POTENTIAL

The conclusion is that you must use all the information in the image to characterize the shape of a potential well and you need to know what is in it!

To summarize, this discussion gives us a method for analyzing Fresnel fringes from a GB. We can draw some conclusions

- To interpret the contrast from Fresnel fringes at GBs, you must simulate images of many different interface models. In particular, it is essential that you

consider the possibility of artifacts such as surface grooving. Even a rather 'flat' or diffuse surface groove may influence the fringes in some range of defocus values.

- Both the fringe spacing and the central fringe intensity depend on the shape of the potential well and are sensitive to surface grooving.
- The interface width, which you can infer from the fringe spacing, is dominated by the outer width of a diffuse interface.
- A direct match with the $s_f - a$ curve (or with similar simulated curves when the assumptions employed here fail) leads to a better estimate of the average interface width, but cannot give you much information on the shape of the potential well.
- Determining when the central fringe is weak (the range Δf) gives complementary information on the interface width which, in combination with the estimate based on the fringe spacing, you can use to evaluate the diffuseness of the potential well.

30.12 CALCULATING IMAGES OF DEFECTS

When we simulate HRTEM images of perfect crystals, we only need to input the unit cell and the program generates the rest of the specimen. If we want to calculate the image of a defect, we have to use the same approach: we set up a unit cell to contain the defect and the program treats it like any other unit cell. This is known as the *periodic continuation* method for defect calculation. What we've actually done is shown in Figure 30.6: there is an array of defects throughout our specimen in all directions. We need to know two things

- To what extent does this ordered array introduce artifacts in the image?
- Have we created interfaces where the 'cells' join which may influence our image?

An example of a supercell for a GB is shown in Figure 30.6. This figure illustrates clearly how we can create a cell which is more suitable for this periodic continuation by including two defects in a single supercell. As shown in this figure, the periodic continuation then not only creates many other GBs but also makes them very long. If we don't match the crystals exactly at the edges of the supercell we create a different 'ghost' boundary.

SIMULATING DEFECTS
The periodic continuation is most widely used. Real-space patching avoids the artificial cell.

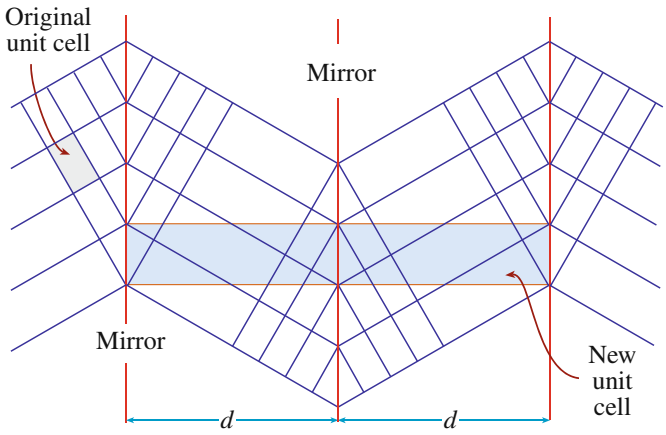


FIGURE 30.6. The periodic continuation technique illustrating how an artificial unit cell can be constructed to contain two GBs, thus allowing the HRTEM image to be simulated. The distance (d) between the two interfaces can be varied to check for overlap artifacts.

You can see that this really can be a problem by considering the DP which our new cell would produce. We are calculating the image of a small part of a periodic array of interfaces. Periodic arrays in real space produce rows of extra spots in reciprocal space. If we include these spots in forming the image, we should change the image. The solution for image simulation is quite simple, make the supercell wider and wider until the change in the image detail is less than some specified limit. However, don't try to interpret the data in the calculated DP without consulting the paper by Wilson and Spargo.

An alternative approach to the periodic continuation approach has been developed by Coene et al. and is called the real-space patching method. This method uses the 'real-space' image simulation approach to perform the calculation. The structure you want to simulate can be divided into a number of different 'patches' as illustrated in Figure 30.7; the image from each patch is

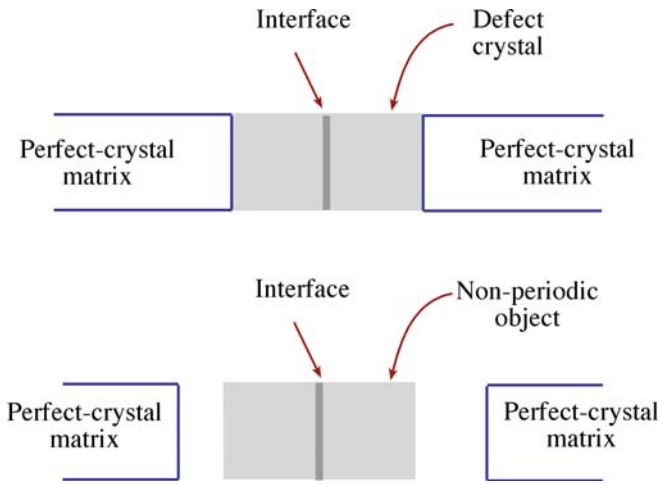


FIGURE 30.7. In the real-space patching method, the defect crystal (in this case the interface and several adjacent layers) is a non-periodic object that is surrounded by a perfect crystal matrix.

calculated for a slice and then the patches are joined together. The key, of course, is that you must correctly take account of what happens at the edge of each patch. This means each patch needs some information about the neighboring patches. Assuming (correctly) that this can be done, you can appreciate the nice feature of this approach: we avoid the artificial interference effects due to the array of defects that would be produced by the periodic continuation technique. The defect does not now 'see' its own image; it only sees the perfect matrix on all sides.

30.13 SIMULATING QUASICRYSTALS

There are several problems in simulating HRTEM images of quasicrystals, not least of which concerns which model you should use. (Remember that quasicrystals are not periodic objects anyway!) Several models have been reviewed by Shoemaker, and the possibilities are illustrated by the work of Beeli and Horiuchi, who used a combination of 10 layers in the multislice calculation. The layers are made up from the planar structures shown in Figure 30.8. The final structure (shown in Figure 30.8A) is made up of two sets of five layers. The first set of layers is B-C-D-C-B in this figure. The second set of five layers is constructed from the first by using the screw symmetry of the structure; the screw axis has 10_5 screw symmetry. The supercell used was 3.882 nm by 3.303 nm which was chosen to contain a complete decagonal cluster which is 2.04 nm in diameter and the center part of a pentagon tile. The calculation was then carried out for thicknesses up to 10 nm.

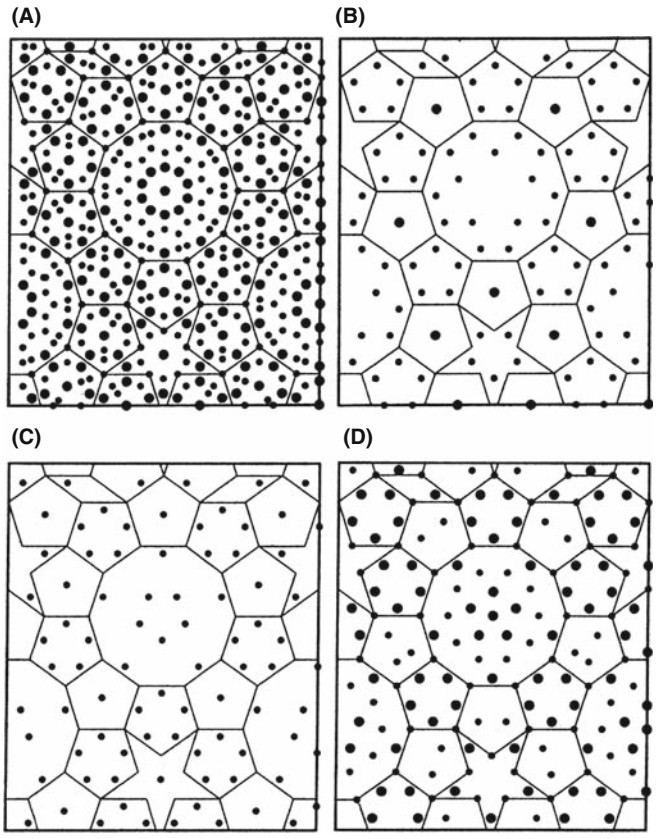


FIGURE 30.8. Projections used to simulate images of $Al_{70}Mn_{17}Pd_{13}$ quasicrystals. (A) Combination of all the layers; (B–D) layers are used to contribute (repeatedly) to (A). The edges of all the tiles are 0.482 nm. The large circles denote Al atoms.

Z-CONTRAST

Since the atoms lie along columns for phase-contrast imaging we can also use *Z*-contrast imaging.

The results of such calculations with only Al and Mn atoms are illustrated in Figure 30.9. The edges of the cells are essentially artificial because, as we just noted, the structure used in the calculation is a 'unit cell' and a quasicrystal doesn't have one. In spite of these difficulties Beeli and Horiuchi could conclude that the image match was much improved when Pd atoms were included to replace some of the Mn atoms in the D layer and Al atoms in the B-C layers with the results shown in Figure 30.10.

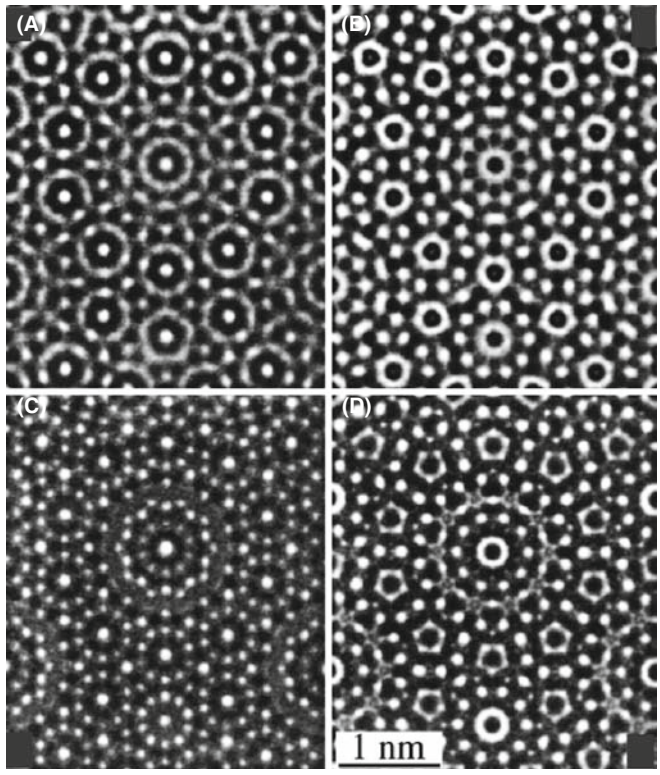


FIGURE 30.9. Four simulated images of the model constructed from the layers shown in Figure 30.8 using only Al and Mn atoms. The thickness is 3.77 nm, which corresponds to three periods in the beam direction. The values of Δf are (A) 0 nm, (B) 46 nm, (C) 88 nm, and (D) 124 nm.

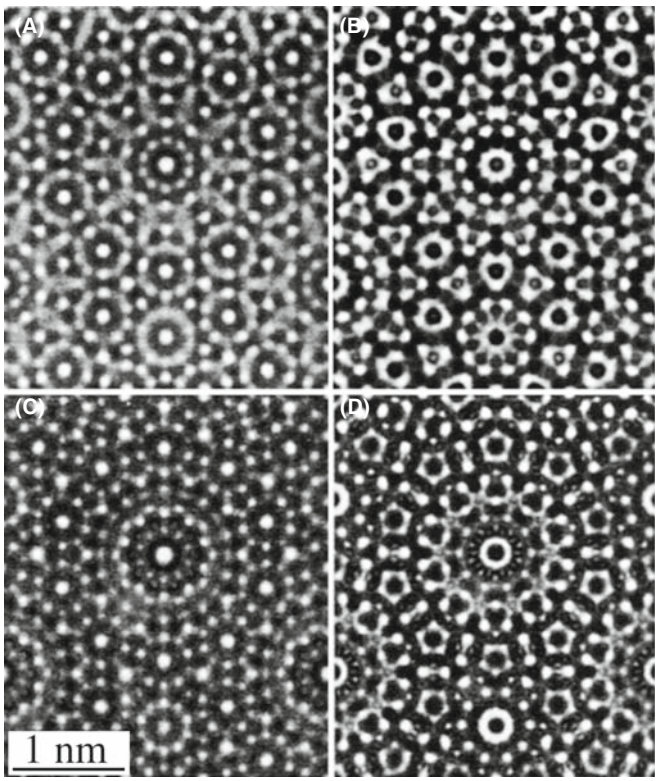


FIGURE 30.10. Examples of simulated images of the quasicrystal shown in Figure 30.8 but substituting Pd atoms for Mn atoms. The values of Δf are (A) 0 nm, (B) 48 nm, (C) 88 nm, and (D) 128 nm.

SIMULATING IMAGES OF A QUASICRYSTAL

As you realize, the problems in such an image calculation are increased because the quasicrystal does not have translational symmetry but we must impose such a symmetry to do the calculation.

Another illustration of the success of HRTEM comes from the work of Jiang et al. on quasicrystals with eightfold symmetry. Here the multislice calculation could again be made using a relatively simple sequence of four layers ABAB′, where the layers are at $z = 0$, 0.25, 0.5, and 0.75. The structures of the A and B layers are shown in Figure 30.11 with the B′ layer being a 45°-rotated B layer; i.e., the B and B′ layers are again related by a screw axis, this time, it's an 8_4 screw axis.

- In each of these examples, it is possible to view the same structure parallel to an orthogonal axis.
- Quasicrystals do not have translational symmetry, but we pretend they do for thickness calculations and for the periodic continuation of the unit cell.

Our reason for showing so much detail on these rather esoteric materials is that they show what can be done using image simulation. Furthermore, they emphasize the important fact that although we can construct the crystal using different layers and different sequences of layers, we always use a projection of the structure, to compare with the experimental image.

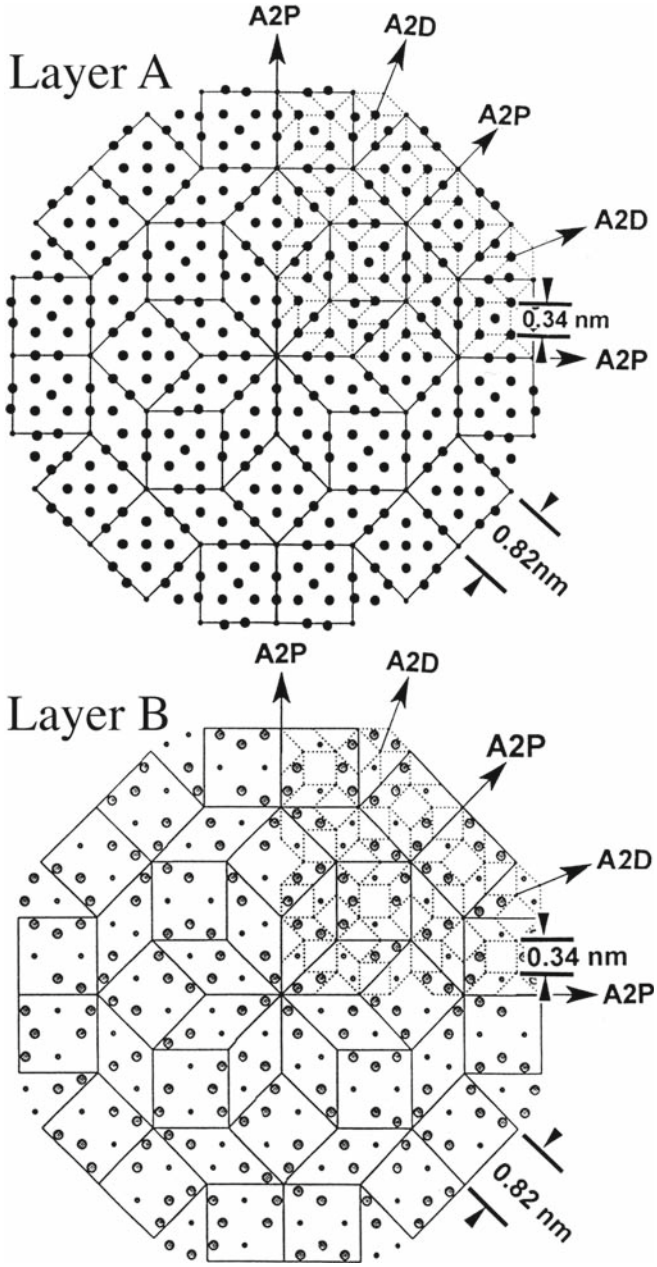


FIGURE 30.11. The model used to simulate quasicrystals with eightfold symmetry. The structure for the simulation was constructed as a four-layer sequence ABAB′ where the B and B′ layers are related by an 8_4 screw axis.

30.14 BONDING IN CRYSTALS

We mentioned early on that one problem we have with simulation concerns the fact that atoms are bonded in different ways in different materials. The standard

approach has been to use values for structure factors tabulated by Doyle and Turner and by Doyle and Cowley. These values were calculated using a relativistic Hartree-Fock (RHF) model for the atomic potential. An alternative approach is to relate the scattering factor for electrons (f_e) to that for X-rays (f_x) using the Mott equation, or to use a more sophisticated atomic potential known as the relativistic Hartree-Fock-Slater (RHFS) model. Carlson et al. give tabulated results while Tang and Dorignac have made detailed comparisons for HRTEM imaging.

O'Keeffe and Spence have re-examined the meaning of the mean inner potential. One of the reasons that we need to understand this concept is that we often link data from X-ray diffraction and data from electron diffraction. As usual, computers are making it possible to do more elaborate calculations using other potentials.

While this is an evolving study, some important results have been obtained

- The inner potential is very sensitive to bonding effects. O'Keeffe and Spence discuss this result for MgO (large ionic component), Si (covalent), and Al (metallic).
- We are still not able to take full account of bonding effects, which could be important for HRTEM images.

This paper by O'Keeffe and Spence is a highly recommended reading for those who have a strong physics background but think TEM is a 'known' subject! For a very different approach, see the paper by Zuo and Spence which uses the DP to deduce information about bonding.

30.15 SIMULATING *Z*-CONTRAST

This is a topic where you can make an impact! The clues are in Chapters 2 and 3 and in the companion text.

30.16 SOFTWARE FOR PHASE-CONTRAST HRTEM

This is a short section because it is summarized in Table 30.1. HRTEMs are priced at up to $4 M as of 2009. Image simulation is essential for HRTEM. Very few students can, or should, write their own programs to simulate HRTEM images—it has already been done many times. The catch is that no manufacturer supports any software package for simulating HRTEM images. Your lab must have at least two programs for you to use. In future we will see much more use of Mathematica and Matlab in this field.

CHAPTER SUMMARY

If you are going to do HRTEM imaging, you must be prepared to use image simulation to assist you in interpreting your images. If you want to do quantitative imaging, simulation is an essential component of the process. Most materials scientists using TEM will want to use one of the established software packages in Table 30.1. There are several important conclusions contained in this chapter

- Make sure that you know all you can about your specimen. We illustrated the dangers with our discussion of grooved GBs. You can waste too much time looking at artifacts caused by specimen preparation.
- Make sure that you know all you can about your TEM. You now have some idea of how many parameters are required by the simulation routines. Beware of the parameters which you did not measure for your machine. The program will need to use some value: *you* should make sure it is appropriate.
- Make sure that you accurately align your TEM before you record any images.
- If possible, use more than one program to simulate the images. At least try this once.
- Record a through-focus series and check for changes in Δf by repeating the first image.
- The fact that the thickness of your specimen varies can be a great asset provided you can determine that thickness; i.e., it gives you another variable.

If you can afford a C_s-corrected HRTEM, then C_s becomes another variable.

The traditional method of using simulated images has often involved looking at a series of simulated images for different values of Δf and t and finding the best match with your experimental image. Clearly this is not the ideal approach! Remember that the interpretation of HRTEM images may not be straightforward or unique. We must next compare the simulated images with those generated experimentally. This is the subject of the next chapter and is the basis of quantitative HRTEM.

GENERAL

Buseck, PR, Cowley, JM and Eyring, L, Eds. 1988 *High-Resolution Transmission Electron Microscopy and Associated Techniques* Oxford University Press New York. An excellent resource covering more than HRTEM.

Horiuchi, S 1994 *Fundamentals of High-Resolution Electron Microscopy* North-Holland, New York. An ideal complement to Spence's book.

Kihlborg, L, Ed. 1979 *Direct Imaging of Atoms in Crystals and Molecules* Nobel Symposium 47 The Royal Swedish Academy of Sciences Stockholm. A classic collection of papers on HRTEM.

Krakow, W and O'Keefe, M, Eds. 1989 *Computer Simulation of Electron Microscope Diffraction and Images* TMS, Warrendale, PA. A collection of focused review articles.

O'Keefe, MA and Kilaas, R 1988 in *Advances in High-Resolution Image Simulation in Image and Signal Processing in Electron Microscopy*, Scanning Microscopy Supplement 2. (Eds. PW Hawkes, FP Ottensmeyer, WO Saxton and A Rosenfeld) p. 225 SEM Inc. AMF O'Hare Il. This article is another excellent introduction to the subject.

Spence, JCH 2003 *High-Resolution Electron Microscopy* 3rd edition Oxford University Press New York. This is *the* text for users of the HRTEM.

Zuo, JM and Spence, JCH 1992 *Electron Microdiffraction* Springer, NY

SCATTERING FACTORS

Carlson, TA, Lu TT, Tucker, TC, Nestor, CW and Malik, FB 1970 *Report ORNL-4614*, ORNL Oak Ridge TN. Tabulated values of scattering factors.

Doyle, PA and Turner, PS 1968 *Relativistic Hartree-Fock X-ray and Electron Scattering Factors* Acta. Cryst. **A24** 390–7.

Tang, D. and Dorignac, D 1994 *The Calculation of Scattering Factors in HREM Image Simulation* Acta. Cryst. **A50** 45–52. Detailed comparisons of scattering factors.

QUASICRYSTALS

Beeli, C and Horiuchi, S 1994 *The Structure and its Reconstruction in the Decagonal $Al_{70}Mn_{17}Pd_{13}$ Quasicrystal* Phil. Mag. **B70** 215–40. Calculations of HRTEM images.

Shoemaker, CB 1993 *On the Relationship between* μ-*$MnAl_{4.12}$ and the Decagonal Mn-Al Phase* Phil. Mag. **B67** 869–81. A review of several models.

MODELS

Clarke, DR 1979 *On The Detection of Thin Intergranular Films by Electron Microscopy* Ultramicrosc. **4** 33–44.

Coene, W and Van Dyck, D 1984 *The Real Space Method for Dynamical Electron Diffraction Calculations in High Resolution Electron Microscopy: II. Critical Analysis of the Dependency on the Input Parameters* Ultramicrosc. **15** 41–50.

Coene, W and Van Dyck, D 1984 *The Real Space Method for Dynamical Electron Diffraction Calculations in High Resolution Electron Microscopy: I. Principles of the Method* Ultramicrosc. **15** 29–40.

Rasmussen, DR and Carter, CB 1990 *On the Fresnel-fringe Technique for the Analysis of Interfacial Films* Ultramicrosc. **32** 337–348.

Ross, FM and Stobbs, WM 1991 *A Study of the Initial Stages of the Oxidation of Silicon Using the Fresnel Method* Phil. Mag. **A63** 1–36.

Ross, FM and Stobbs, WM 1991 *Computer Modelling for Fresnel Contrast Analysis* Phil. Mag. **A63** 37–70.

Simpson, YK, Carter, CB, Morissey, KJ, Angelini, P and Bentley, J 1986 *The Identification of Thin Amorphous Films at Grain-Boundaries in Al_2O_3* J. Mater. Sci. **21** 2689–96.

IMAGE SIMULATION

Barry, J 1992 in *Electron Diffraction Techniques* **1** (Ed. JM Cowley) p. 170 I.U.Cr. Oxford Science Publication New York.

Coene, W, Van Dyck, D, Van Tendeloo, G and Van Landuyt, J 1985 *Computer Simulation of High-Energy Electron Scattering by Non-Periodic Objects. The Real Space Patching Method as an Alternative to the Periodic Continuation Technique* Phil. Mag. **52** 127–143. Real-space patching for Section 30.12.

Cowley, JM and Moodie, AF 1957 *The Scattering of Electrons by Atoms and Crystals. I. A New Theoretical Approach* Acta Cryst. **10** 609–19.

Kambe, K 1982 *Visualization of Bloch Waves of High Energy Electrons in High Resolution Electron Microscopy* Ultramicrosc. **10** 223–227.

Kirkland EJ 1998 *Advanced Computing in Electron Microscopy* Springer, NY

Stadelmann, PA 1979. *EMS - a Software Package for Electron Diffraction Analysis and HREM Image Simulation in Materials Science* Ultramicrosc. **21** 131–145. The original description of EMS.

SOME PHYSICS

Doyle, PA and Cowley, JM 1974 in International Tables for X-ray Crystallography, Vol. **IV**, pp. 152–173, Kluwer Academic Publ. Dordrecht Netherlands.

Fujimoto, F 1978 *Periodicity of Crystal Structure Images in Electron Microscopy with Crystal Thickness* Phys. stat. sol. (a) **45** 99–106.

Jiang, J-C, Hovmöller, S and Zou, X-D 1995 *A Three-dimensional Structure Model of Eight-Fold Quasi-crystals Obtained by High-Resolution Electron Microscopy* Phil. Mag. Lett. **71** 123–129.

O'Keeffe, M and Spence, JCH 1994 *On the Average Coulomb Potential (Φ_0) and Constraints on the Electron Density in Crystals*. Acta Cryst. **A50**(1) 33–45.

Self, PG and O'Keefe, MA 1988 in *High-Resolution Transmission Electron Microscopy and Associated Techniques* (Eds. PR Buseck, JM Cowley and L Eyring) p. 244 Oxford University Press New York.

Taftø, J, Jones, RH and Heald, SM 1986 *Transmission Electron Microscopy of Interfaces Utilizing Mean Inner Potential Differences between Materials* J Appl. Phys. **60** 4316–8.

Wilson, AR and Spargo, AEC 1982 *Calculation of the Scattering from Defects Using Periodic Continuation Methods* Phil. Mag. **A46** 435–49.

Zhu, Y, Taftø, J, Lewis, LH and Welch, DO 1995 *Electron Microscopy of Grain Boundaries. An Application to RE-Fe-B (RE = Pr or Nd) Magnetic Materials* Phil. Mag. Lett. **71** 297–306.

THE COMPANION TEXT

The companion text extends the topic of simulation in chapters on EMS (jEMS is the java version that Pierre Stadleman supports) and on simulating diffraction contrast images. One idea which you might examine immediately is the use of Mathematica or similar packages to graph the equations.

SELF-ASSESSMENT QUESTIONS

Q30.1 List the four principal methods of calculating multiple images.
Q30.2 What is a phase grating?
Q30.3 Why do we sample all of reciprocal space, not just the Bragg beams?
Q30.4 Write down an equation relating ψ_{n+1} and ψ using a propagator and a phase grating.
Q30.5 Why are we using Fresnel diffraction in this multislice treatment rather than Fraunhofer?
Q30.6 We use convolution to describe what sort of scattering and why?
Q30.7 Why must $Q(\mathbf{k})$ take account of nearly twice as many k values as $\psi(\mathbf{k})$ or $P(\mathbf{k})$?
Q30.8 Write down an equation for $\psi_{n+1}(\mathbf{k})$ used in the FFT method.
Q30.9 Extra challenge. Why is it FFT not just FT?
Q30.10 Write down an equation for $\psi_{n+1}(\mathbf{r})$ in the Coene-Van Dyck method.
Q30.11 Why do we use ' ' when describing Kambé's 'simple' example?
Q30.12 In Kambé's 'simple' example, we consider the case where only three Bloch waves are significant. Why choose just three, not more (or fewer)?
Q30.13 We mention the direct connection between WPOA and Bloch wave propagation. Why is this?
Q30.14 Why is it relevant that λ is finite (not zero)?
Q30.15 Why does beam convergence alter a HRTEM image according to the Ewald construction?
Q30.16 Why will GB grooving influence HRTEM images?
Q30.17 What is the periodic continuation method?
Q30.18 What is the problem in imaging crystals with 5- and/or 10-fold symmetry?
Q30.19 How do most multislice images simulation programs take care of ionic and covalent bonding?
Q30.20 Why do many researchers prefer a through-focus series to just a focal series of images?

TEXT-SPECIFIC QUESTIONS

T30.1 Assuming that the first Fresnel fringe is $\delta = (\lambda \Delta f)^{1/2}$ away from the edge of the sample, calculate Δf for Figure 9.21B.
T30.2 How well does the relationship $\delta = (\lambda \Delta f)^{1/2}$ for the position of the first Fresnel fringe predict the images in Figure 30.5C assuming the structure given in Figure 30.5A?
T30.3 Construct a unit cell to use in HRTEM simulation of the $\Sigma = 3$ 111 twin boundary in Si.
T30.4 Construct a unit cell to use in HRTEM simulation of the $\Sigma = 3$ 112 twin boundary in GaP. Discuss the information you would obtain for an HRTEM giving point (structural) resolution of 0.3, 0.2, and 0.1 nm.
T30.5 Show the smallest reproducible unit for Figure 30.11 given the symmetry shown in the figure.
T30.6 Plot equation 30.14 for different values of Δf. How does this plot change as we change the kV of the microscope?

T30.7 By playing with equation 30.15 deduce equations 30.19 and 30.20.
T30.8 Deduce equation 30.21 using equations 30.18 and 30.20.
T30.9 Explain why the 'images' from Bloch wave 1 differs from those in Bloch waves 2 and 3, as shown in Figure 30.3.
T30.10 Index the DP in Figure 30.4 and thus determine the radius (in nm^{-1}) of the objective aperture.

31

Processing and Quantifying Images

CHAPTER PREVIEW

In this chapter we will equate processing with the use of the computer to analyze our data. We will simply use image processing to extract more information from the data than we can obtain by eye. The data will generally be an HRTEM image but could be other images or DPs. We'll quantify spectra after we describe them in Part 4 of this text. In the past, the optical bench was also used for this purpose, but the number of optical benches is negligible compared to the number of computers now found in every TEM lab. Optical benches did allow us to form DPs which we could then modify to produce a processed image. This analog approach has now largely been replaced by its digital counterpart. The computer can be much cheaper than the optical bench and is far more flexible. The number of software packages which are designed for, or can easily be adapted to, TEM is also growing.

We can use image processing to produce a clearer view of the image, for example, by subtracting unwanted background detail, correcting for noise or drift, or removing artifacts. The big warning, though, is that, when removing one artifact, you must be very careful not to introduce others.

Although it's nice to see information more clearly, the unique feature of the computer approach is that we can *quantify* the data in any image and then normalize these data. Now we can directly compare the quantified experimental image with computer-simulated images. Although throughout this chapter we will be concerned with HRTEM images, most of what we say can be transferred directly to the analysis of diffraction-contrast images.

The other general point is that the ideas we'll discuss are also applicable to images derived from different sources. Once the data are in the computer, i.e., in digital form, the source becomes unimportant as far as processing possibilities are concerned. Examples of 'images' which might be obtained from the TEM include X-ray or EELS maps, STEM images, TEM images, and CBED or BSE patterns.

Most of the discussion will concern the use of computers. All you need to know is how best to get the data into the computer, how to process it, what to do with the data, how to display the result, and how to report what you've done. Most importantly, when we have the image in a numerical form we can start to consider statistics, which means much more than adding 'error' bars.

31.1 WHAT IS IMAGE PROCESSING?

Image processing is essentially manipulating images. The topic arises in many fields so we need to understand the words/jargon; we'll discuss the language of image processing as it is applied to TEM.

Image processing is not only becoming more common, but is also finding new applications in many fields. Faster, more powerful computers and increased memory storage are making tasks possible which could not previously have been considered. As a result of this increasing user base, there are now many software packages available which can be used in microscopy; we listed some in Section 1.5. These range from programs used widely in desktop publishing to those which have been custom designed for EM. The goals of image processing include that of quantitative microscopy. You must choose among the different packages, commercial and freeware, and match them to the computer available in your lab. One point to remember is that some very simple optical methods which don't rely on a computer can often be very helpful. The other point is that the eye is hard to beat.

THE BASIC IDEA

Change images into numbers and manipulate the numbers.

There are many specialized books on this topic for the beginner or the expert; a selection is given in the references. The purpose of this chapter is to give a generalized overview. One problem in discussing this topic is that it is a very rapidly changing field. We will try to avoid specifics concerning particular programs but will mention these programs at the end of this chapter.

31.2 PROCESSING AND QUANTIFYING IMAGES

We process images primarily for two reasons

- We may want to improve the appearance of an image, make it look sharper, more even in contrast, higher contrast, etc. This processing can be unwise.
- We may want to quantify the information contained in the image. This processing may be essential: many physicists can read equations but not images.

Processing for improving the appearance of TEM images has been practiced for many years using such photographic techniques as 'dodging,' using 'filters,' selecting different emulsions or varying the developer, etc. It is only recently that relatively powerful personal computers have become widely available, but the term 'image processing' almost automatically implies the use of computers. Computer image processing will be the emphasis of our discussion. We have three requirements.

- We must be able to create a digital form of the image in the computer.
- We need appropriate software for processing the image.
- We need a computer which can perform the processing in an acceptable period of time with the required resolution.

THE DIFFICULTY

When image processing, the difficulty may be describing what you have done.

Many comments here are similar to those we made in discussing the microscope itself. For example, you may have to work with the built-in system or the system that's already available in the lab. The difference is that some of the free programs are extremely powerful, so that all you need is the desktop computer. Many programs designed for desktop publishing are relatively inexpensive. Thus, you can almost always find a way to extend your processing capabilities.

The motivation is that we need to obtain more information from an image than we can get by just looking at it. This principle applies to more than HRTEM; we are discussing it here because HRTEM is where at present it is most needed/used in TEM. However, any TEM, X-ray map, or energy-filtered image or DP may benefit from processing and quantification. We need to quantify the TEM parameters, in particular C_s. One unique aspect of image processing in the TEM is that we have a choice between on-line and off-line processing. In fact, we often use on-line processing (frame averaging and background subtraction on the video image) to see the image even though the image we record may not be processed. However, it may indeed already be processed—so you should know and report this.

31.3 A CAUTIONARY NOTE

For most of our discussion we will consider only processing techniques using computers. To a large extent we can simulate a TEM using the computer. As we saw in Chapter 30, we can model a crystal, insert apertures, define the electron beam, including its broadening in the specimen, and then calculate the image. What we do in image processing is start with the image, add apertures and special filters, and then create a new image, the processed image. This image is a *real* image. What we must be careful about is explaining just what processing procedures we have used, since these may affect the interpretation of the data. Reporting the processing procedure is particularly critical when the raw data (the 'original' image) is not being reported at the same time.

ALWAYS

Report how you have treated your image, so that the reader can compare your data with related data that may have been processed differently or not at all.
Be very careful. One honest (or dishonest) mistake may change your career!

31.4 IMAGE INPUT

There are several methods you can use to put the TEM image on the computer. The choice depends in part on how much detail you want in your digitized image, but also depends on how much work you're prepared to do. In this discussion, we'll only consider images which you have looked at on a video monitor, a computer display, or the fluorescent screen. Your basic choices are

- Transfer the image directly from the TEM to the computer.
- Record the image on film, then digitize it using a microdensitometer.

- Record the image on videotape.
- Record the image on film, then print it and use a flatbed (or other) scanner.

There are many methods for creating a digital form of an image in your computer. The simplest is to use a slow-scan CCD camera, which we discussed in Chapter 7. The drawback of CCD cameras is that high-quality CCD chips are very expensive for 2 k × 2 k arrays and astronomically expensive for 4 k × 4 k arrays (but they are getting cheaper). Such cameras may become routine add-ons for all TEMs in the future. You will probably also want to use video. With film, you can record a larger area than you can using a CCD unless you have the 4 k × 4 k version. You should use a digital video-recorder for in situ studies when using a heating or straining holder.

We can transfer the image from a videotape or a video camera to the computer using a frame grabber. Frame-grabber boards are readily available for most computers. You can use a high-resolution scanner for photographs or negatives. At this time, scanners cost about the same as a digital camera with comparable resolution. The purist's approach is to use a microdensitometer to measure the intensity of the film point by point and read this directly into the computer. The advantage of the microdensitometer is that it is very precise and can achieve the highest resolution for a very large area. The main problem is that it is slow, being a serial-collection technique. If you use it to its best advantage, your image will require a large amount of computer memory, which in itself is not a problem, but manipulating such images will still be slow.

31.5 PROCESSING TECHNIQUES

31.5.A Fourier Filtering and Reconstruction

The principle involved in filtering is that a mask is used to remove some information from an image in order to enhance or emphasize other information. As an extra complication we can process the image, e.g., Fourier transform a HRTEM image, then apply a mask, and then reverse the processing.

We can vary the size of the apertures and the sharpness of their edges, something you can't do in a TEM with normal fixed-diameter objective apertures. A single variable SAD aperture was used on some early TEMs; it was triangular or square in shape and used three or two movable blades. You can best understand the procedure by an example. A square mask was used to select the region in Figure 31.1A from a much larger region of the HRTEM image, and its Fourier transform (i.e., the DP effectively from a few nanometers) is shown in Figure 31.1B.

THE SUPPORT FILM IS IMPORTANT

You will often see similar diffractograms obtained using a film of amorphous carbon. While such films are easier to make, they give little diffracted intensity for the range of **u** values between 6 and 8.5 nm^{-1}, which is important in HRTEM.

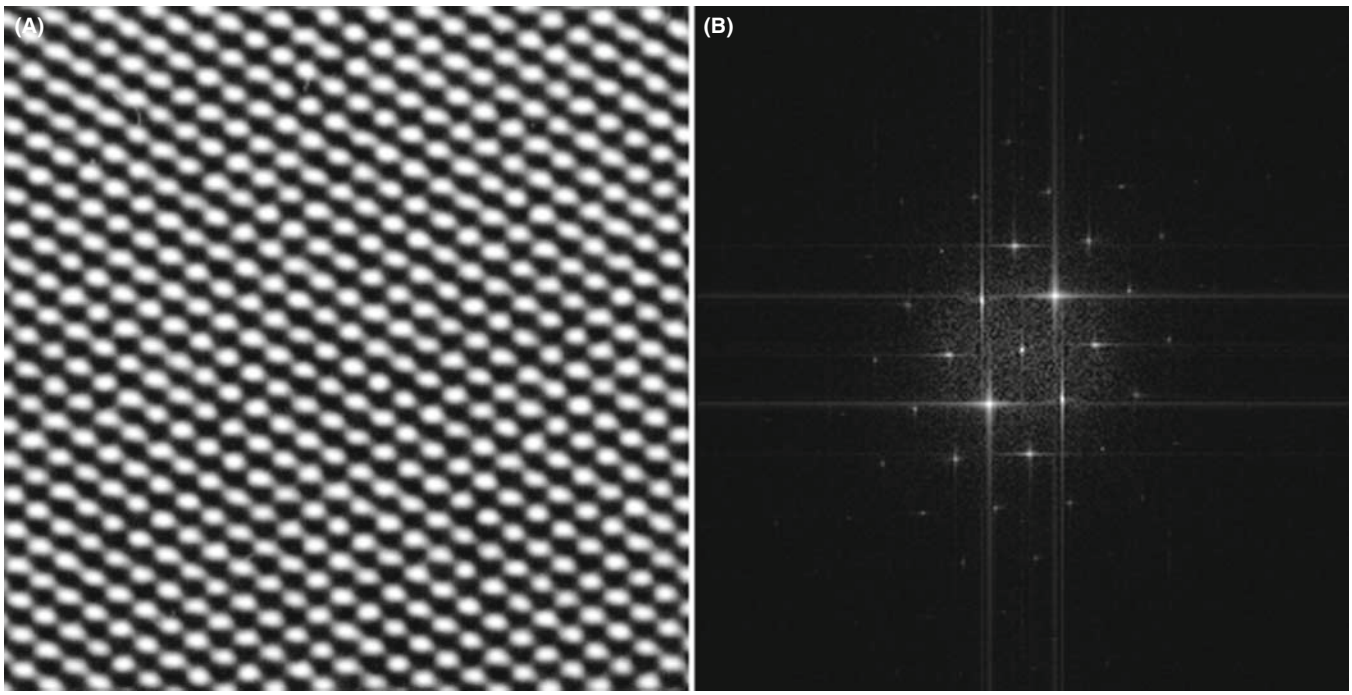


FIGURE 31.1. A square mask has been used to select the area shown in (A) from a much larger print of the image. The Fourier transform of this region is shown in (B) where you can see not only the spots in the 110 DP but also long streaks (artifacts of the processing) that run normal to the edges of the mask.

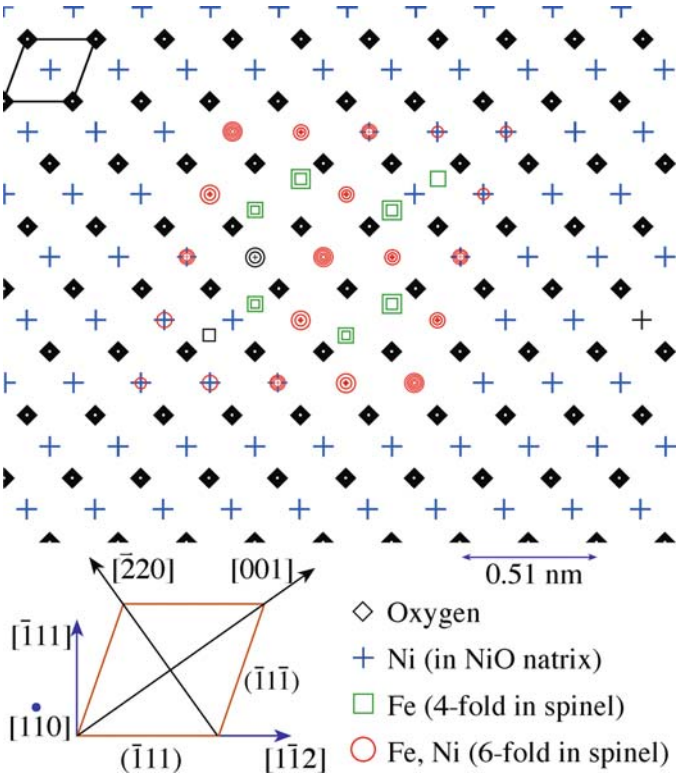


FIGURE 31.2. A model of an octahedron of spinel fully enclosed within a matrix of NiO. The rest of the specimen could then be modeled by adding extra layers of NiO above or below the defect layer.

What this technique does is to allow you to do microscopy in the computer. Your image becomes the specimen. You form the DP, then you can use apertures to select one or more beams to form the image; these apertures are the computer version of the objective aperture. Small apertures limit resolution just as in the 'real' TEM because the information about anything other than the perfect lattice is carried between the reciprocal lattice points. Figure 31.2 gives an illustration of how a model can be constructed of a particle in a matrix, which can be useful when simulating HRTEM and conventional BF/DF images in the computer, as shown in Figure 31.3. This model was made using the Digital Micrograph package (see Section 1.6).

31.5.B Analyzing Diffractograms

In Chapter 28 we showed that the transfer function could be plotted out as shown schematically in, e.g., Figure 28.4. Another way of thinking about this plot is to imagine what would happen if we have a specimen which generated equally every possible value of **u**, i.e., every possible spatial frequency.

An amorphous film of Ge can provide just such a plot, but it is difficult to record the result because the scattered intensity is low.

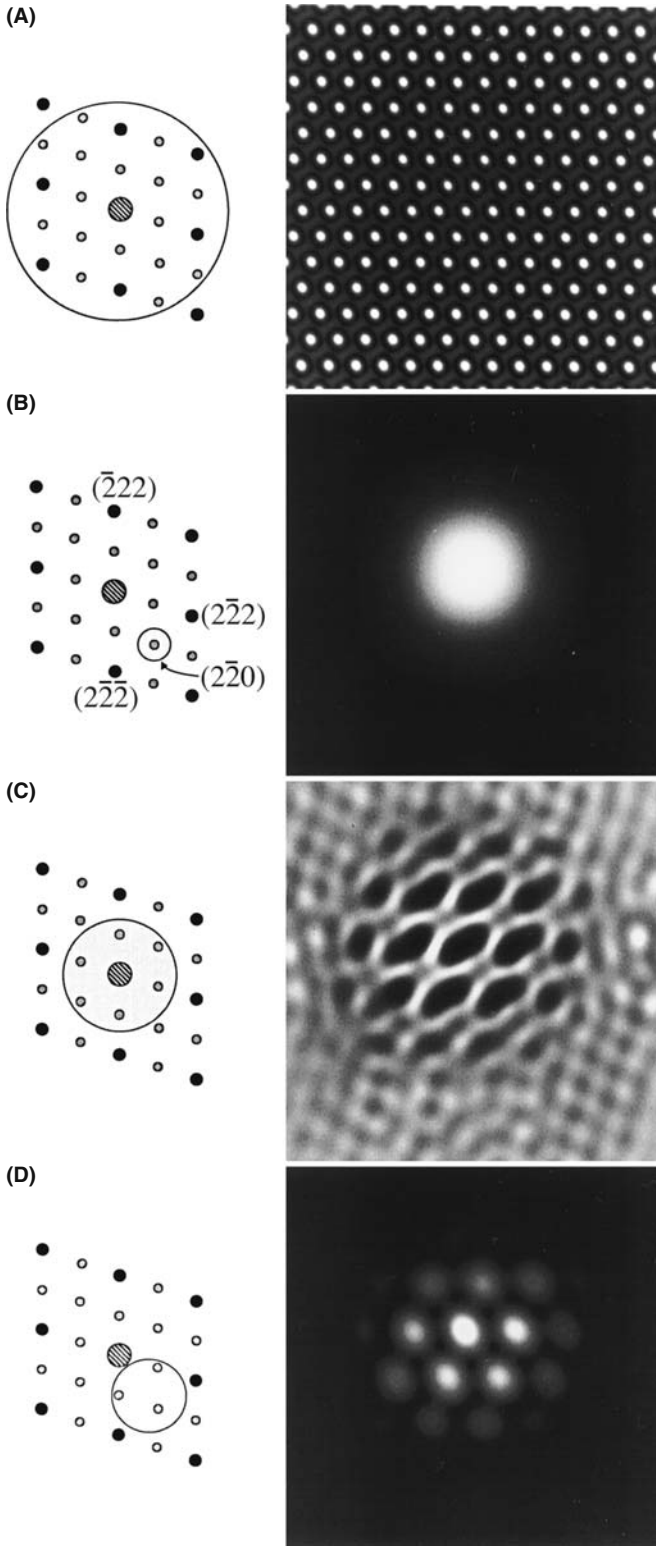


FIGURE 31.3. The DP from a model specimen containing a particle such as that shown in Figure 31.2 is shown schematically in (A) together with the resulting lattice image. The other three pairs of diagrams illustrate how we can use the computer to produce different masks and thus generate different images, such as the DF image in (B). The image in (C) corresponds to the image shown in Figure 28.26, while the DF lattice image shown in (D) is analogous to that discussed in Figures 28.14 and 28.15.

We therefore record the image at high resolution, preferably directly, using a slow-scan CCD camera, although digitizing the negative is fine. Then by comparing the experimental plot of I versus $\mathbf{u}$ with those calculated for different values of Δf and C_s we can determine the astigmatism, the defocus, Δf, and the value of C_s (as we'll see below). It helps if you have a few particles of Au on the Ge film since the Au spots then give an internal calibration. Such a set of images and their corresponding diffractograms is shown in Figure 31.4. Notice that as the defocus of the objective lens increases the number of rings increase but they become narrower. The contrast transfer gradually extends to larger values of $\mathbf{u}$.

Determining astigmatism. you can use such diffractograms to correct the astigmatism, since a perfectly stigmated image will give a DP with circular symmetry. As you can see in Figure 31.5, even a small amount of astigmatism can be detected by eye. The computer can readily measure and provide feedback to the TEM to correct the lens defect, as we will see shortly. This set of diffractograms also shows that the computer can distinguish astigmatism and drift in the image while the eye can easily mistake one for the other. Drift produces a circular pattern but the higher spatial frequencies are lost in the direction of drift.

Determining Δf and C_s. You can determine Δf for any image by measuring the radii of the bright and dark rings in the diffractogram, since bright rings correspond to $\sin \chi(\mathbf{u}) = 1$ and dark rings correspond to $\sin \chi(\mathbf{u}) = 0$.

$$\sin \chi(\mathbf{u}) = 1 \quad \text{when } \chi(\mathbf{u}) = \frac{n\pi}{2} \text{ and } n \text{ is odd} \tag{31.1}$$

$$\sin \chi(\mathbf{u}) = 0 \quad \text{when } \chi(\mathbf{u}) = \frac{n\pi}{2} \text{ and } n \text{ is even} \tag{31.2}$$

Since C_s will also influence the location of the rings, you need at least two rings. Krivanek has given a simple procedure for finding both C_s and Δf. If we start with our definition of χ

$$\chi(\mathbf{u}) = \pi \Delta f \lambda u^2 + \frac{1}{2}\pi C_s \lambda^3 u^4 \tag{31.3}$$

then, inserting the values given in equations 31.1 and 31.2 leads to

$$\frac{n}{u^2} = C_s \lambda^3 u^2 + 2\Delta f \lambda \tag{31.4}$$

All we now have to do is plot nu^{-2} versus u^2 to obtain a straight line with slope $C_s\lambda^3$ and with an intercept on the nu^{-2} axis of $2\Delta f\lambda$. Assign $n = 1$ to the intensity maximum of the central bright ring, $n = 2$ to the first dark ring, etc. The analysis can be trickier if you have used an underfocus condition or if you are very close to Scherzer defocus, but you will know when you have not found a straight line! Hopefully, you'll find that your value of C_s is close to that given by the manufacturer! A rather neat result is that if you plot nu^{-2} versus u^2 for different diffractograms (i.e., different values of Δf), then the points corresponding to each particular value of n will lie on a hyperbola, as shown in Figure 31.6A. You can use these hyperbola to determine C_s for any microscope and Δf for any diffractogram.

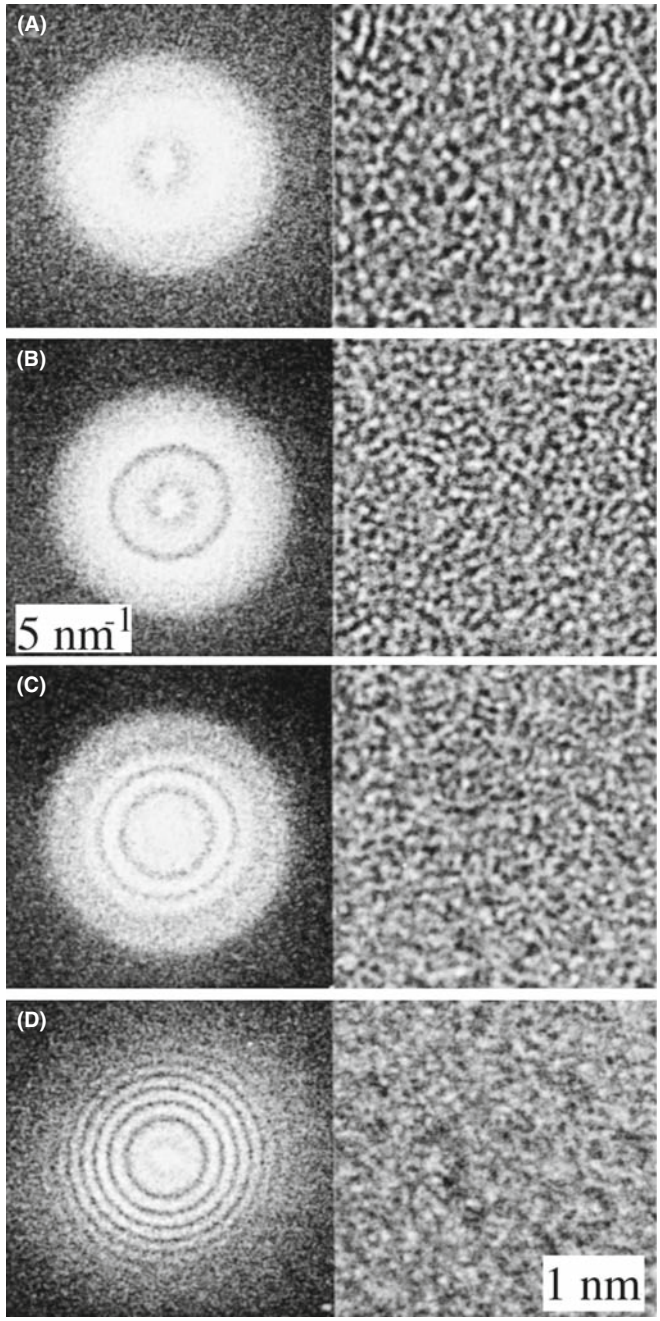


FIGURE 31.4. Four images of an amorphous Ge film and their corresponding diffractograms. Δf has the following values: (A) 1 sch; (B) 1.87 sch; (C) 2.35 sch; (D) 3.87 sch. Remember that 1 sch = $-(C_s\lambda)^{1/2}$.

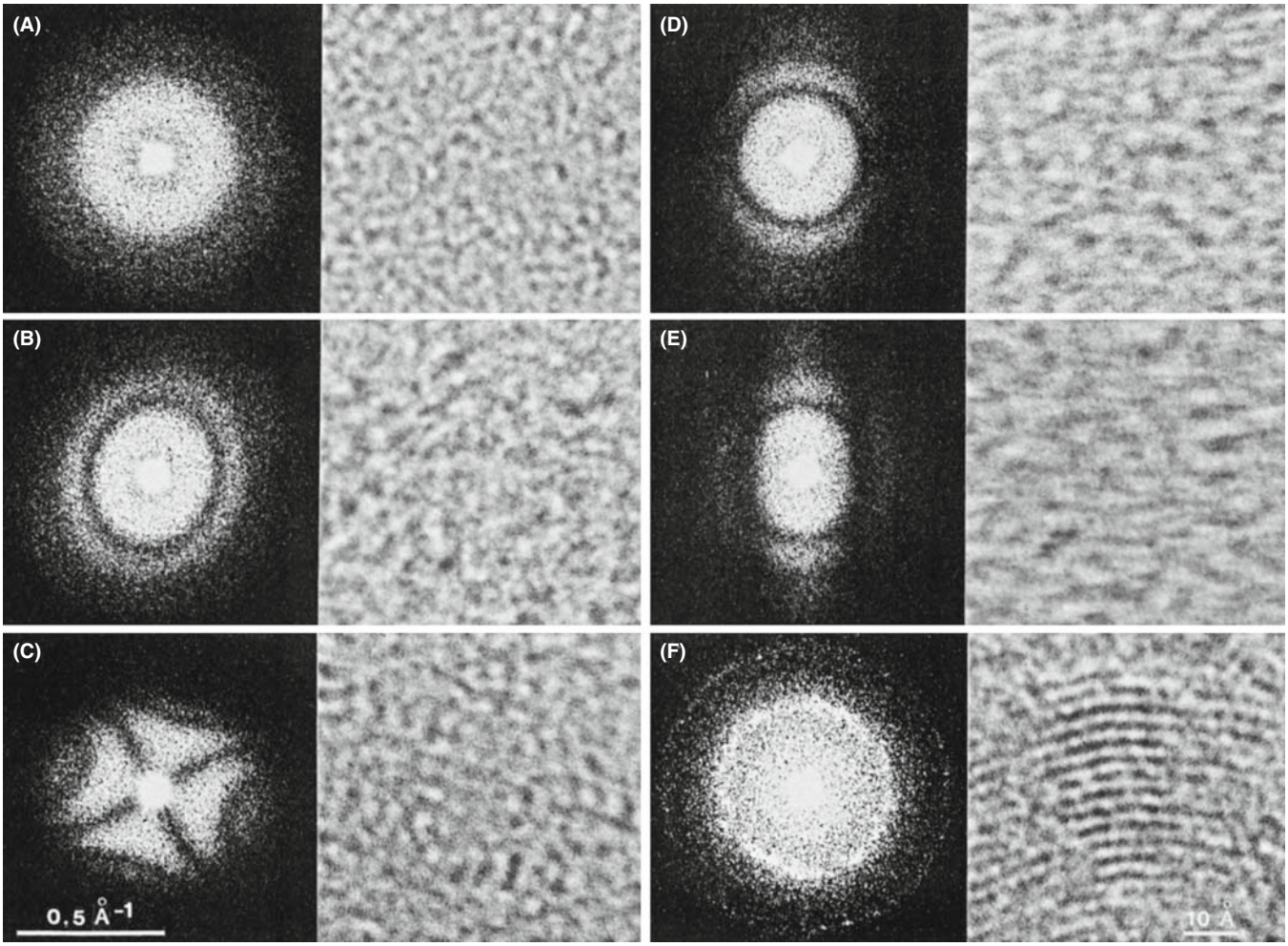


FIGURE 31.5. Six images of an amorphous carbon film and their corresponding diffractograms illustrating different misalignments of a 300-kV HRTEM. (A) Well aligned and no drift; (B) some astigmatism (C_a = 14 nm); (C) more astigmatism (C_a = 80 nm); (D) no astigmatism but drifted 0.3 nm; (E) no astigmatism but drifted 0.5 nm; (F) well aligned and no drift showing graphite calibration fringes of 0.344 nm spacing. (B,D,E) Δf = 2.24 sch; (F) Δf = 0.

THE ZEMLIN TABLEAU

This is a display of diffractograms recorded by systematically changing a parameter, such as tilt or defocus, as in Figure 31.5.

Diffractograms and beam tilt. Beam tilt is very difficult to correct by eye; even worse, it can cause the diffractogram to look astigmatic, so you correct the astigmatism instead. In the image, as we saw earlier, beam tilt can improve the appearance but confuse the interpretation! The set of diffractograms shown in Figure 31.6B shows you how to overcome the problem. You have to compare diffractograms taken at different beam tilts to determine the zero-tilt condition. A pair of diffractograms taken at $\pm\theta^\circ$ tilt will only look the same (though rotated) if the beam had zero tilt at $\theta = 0^\circ$. In the example shown, the diffractograms above and below the horizontal line are similar, so θ_y was very close to zero for the central condition. However, the pairs of diffractograms on opposite sides of the vertical axis differ slightly, so the alignment of θ_x was not perfect.

31.5.C Averaging Images and Other Techniques

If you have recorded a series of images using a video camera, for example, you can average them over several frames as your eye does automatically. The result of such a process is illustrated in Figure 31.7. Different methods can be used to average the images. The easiest approach appears to be as good as any and simply involves taking the unweighted average of your best images, i.e., in the video example, just average over a series of frames. If you know that the object you're studying has a certain symmetry, you can use that information to improve the image further. The article by Trus et al. will give you a start on this process. If you want to remove the blur due to motion of the image, then you will really need to delve much more into this subject.

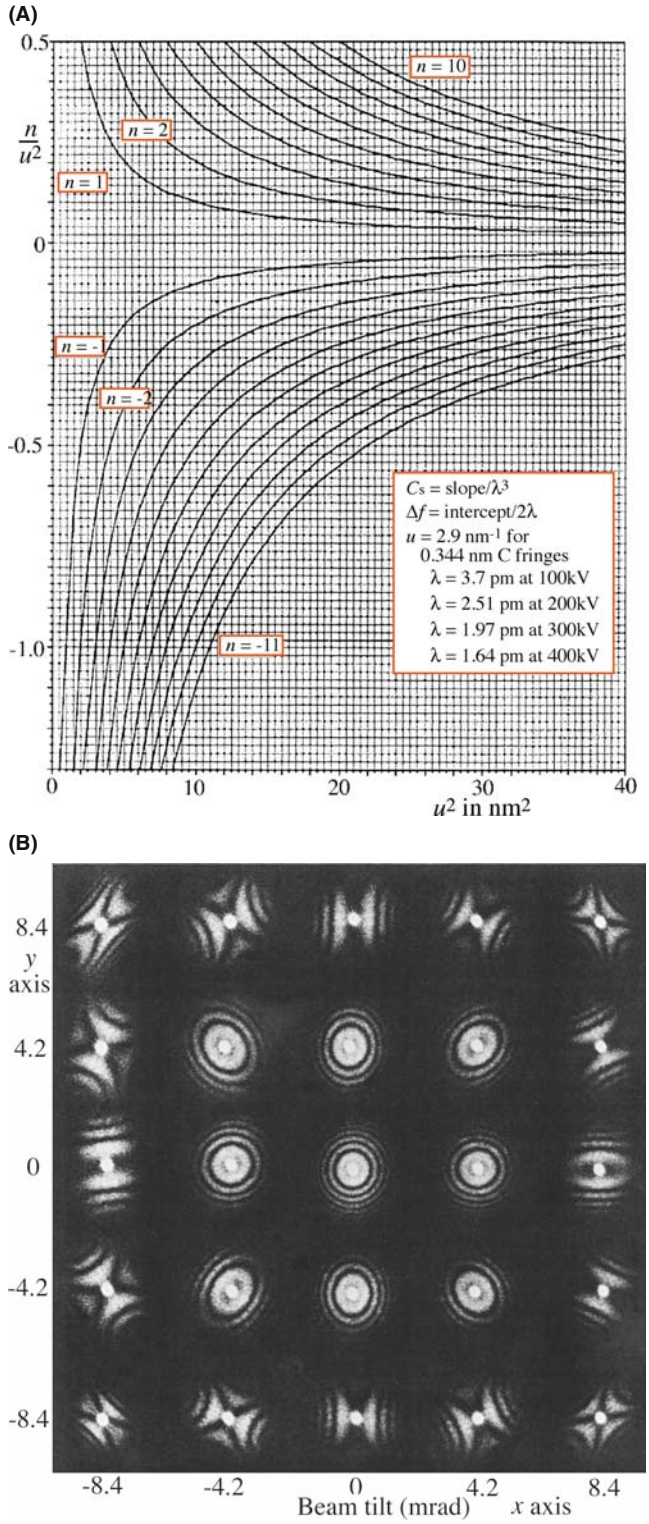


FIGURE 31.6. (A) Plot of nu^{-2} versus u^2. The rings in any diffractogram correspond to a series of n values which allow you to draw straight lines on this figure and thus determine the slope and the ordinate intercept, giving C_s and Δf, respectively. (B) Set of diffractograms showing the effect of the incident beam tilt.

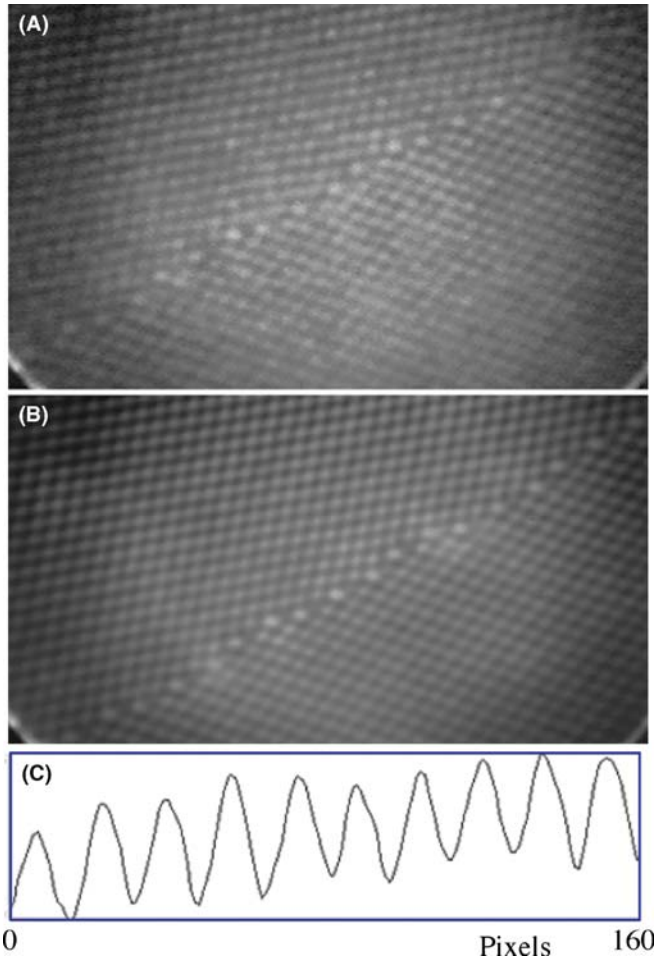


FIGURE 31.7. An example of the benefit of frame averaging to improve information from a video recording: (A) one frame; (B) 16 frames; (C) intensity profile along a (111) plane in (B).

If you use a TV-rate video, you'll almost certainly use background subtraction routinely. For example, you can record an image of the honeycomb pattern of the YAG detector, store it, and then automatically subtract it from all subsequent images in real time.

You may find it useful to add artificial color (pseudocolor) to your TEM images, as illustrated in Figure 31.8. Although it is often assumed that this is done just to make the images even more appealing to the non-microscopist (or non-scientist), there is actually a valid reason for the practice. Our eyes are much more sensitive to small variations in color than they are to small variations in gray level. You might therefore find color useful if you have a wide range of gray levels and want the viewer to be able to 'see' some subtle variations. Similarly, you can use color to emphasize a particular gray level in an image. However, you have to be very careful in your choice of look-up table (LUT), the table which relates each gray level to a particular color. To get a feel for the dangers, play with Photoshop™ and your favorite TEM image.

As we've already noted, in TEM, all of our apertures have relatively sharp edges but in the computer, you have the possibility of using multiple apertures, apertures with different shapes and apertures with diffuse

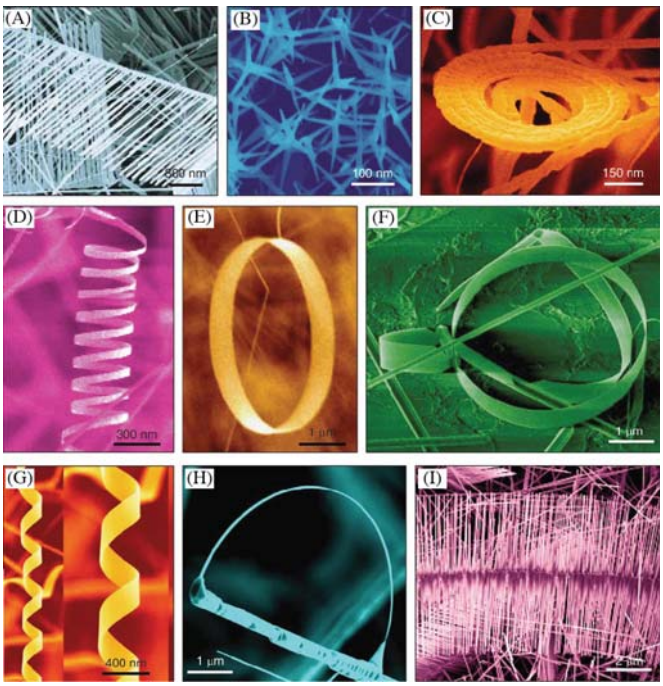


FIGURE 31.8. Illustrating the effectiveness of using false color.

edges. Apertures with diffuse edges will help eliminate the streaking which will otherwise be present (go back and look at Figure 31.1). You can also use the computer to do 'unsharp masking,' which is not the same as simply using a diffuse mask. The technique comes from the photographic process whereby we first print an image out of focus onto film, thus making a complementary image except where there is fine detail present in the original image; in digital processing this is called Laplacian filtering. Many more examples are given in Russ's two books.

31.5.D Kernels

A kernel is simply an array of numbers which we can use to perform operations on a digital image. If we have the 3×3 kernel, K (we can have 5×5, 7×7, etc., but the computation time becomes too long, especially for real-time situations)

$$K = \begin{matrix} -1 & -1 & -1 \\ -1 & +8 & -1 \\ -1 & -1 & -1 \end{matrix} \qquad K_o = \begin{matrix} A & B & C \\ D & E & F \\ G & H & I \end{matrix}$$

we can apply it to every 3×3 group of pixels in our image, e.g., K_o and put the result in a new digital image. If we call our new 3×3 image, K_i then

$$K_i = \begin{matrix} A' & B' & C' \\ D' & E' & F' \\ G' & H' & I' \end{matrix}$$

The new image will have, for example, $E' = 8E - A - D - G - B - H - C - F - I$. This kernel then gives us a digital Laplacian (an approximation to the second linear derivative, ∇^2). What this kernel is doing is subtracting the brightness value of each neighboring pixel from the center pixel. If the area is a uniform gray it will become white so changes in contrast will be exaggerated. We can design a wide range of kernel operators. For example, the edge enhancer kernel has the effect of digitally differentiating the image. (We'll see a related digital-processing procedure applied to spectra in Chapters 35 and 39.) The Sobel and Kirsch operators are examples of such edge detectors; each can be thought of as the sum of several kernel operators. We can also use binary morphological operators which make binary features become larger or smaller. All of these operations can be carried out in any standard image-processing package. In general, you should be very careful when using such techniques in TEM; their value is in displaying data which might otherwise be missed, rather than helping you quantify an image.

REPORT IT

We've said it before. Check HREM papers in the literature. Have the images been processed? Are you told how? Can you reproduce the 'experiment'? The last test is the essential component of science.

31.6 APPLICATIONS

This section will give you a taste of how image processing is being used now. It is just part of a rapidly growing list, so we are not going to be detailed or inclusive. We can separate the applications into two groups

- Noise reduction or improving the signal/noise ratio.
- Quantifying images.

Of course, the first topic is included in the second.

31.6.A Beam-Sensitive Materials

Low-dose microscopy necessarily implies that the signal-to-noise ratio will not be large; if it is large, the dose could have been smaller. This problem has been extensively addressed in biological EM and led to Klug's Nobel prize for 'Development of crystallographic electron microscopy and the structural elucidation of biologically important nucleic acid–protein complexes' in 1982 (see the 1971 paper by Erickson and Klug). In materials science, we have tended to accept 'beam damage' as a fact of life, but this attitude will not be acceptable for future quantitative HRTEM. Most modern microscopes will allow you to perform all your

alignments on one area and then translate the beam to a predetermined distance in a predetermined direction before recording the image of a pristine area. Clearly, the CCD camera will not only let you see your image without waiting to develop the plates, but you can take a series of images for noise reduction purposes and/or assess whether the imaging conditions were what you had intended. (Sometimes, by taking a chronological sequence of images you can get a sense of the 'undamaged' image by extrapolating information back to zero time). The image shown in Figure 31.9 illustrates the possibilities. If you read the review by van Heel et al., you will get some idea of how far you can already go in this field.

BEAM DAMAGE

The electron beam will invariably change your specimen. Historically, we have worried about beam damage. Beam healing is just as bad for your image interpretation!

31.6.B Periodic Images

In discussing quantitative analysis, we have already noted how we can use the computer to identify similar features and combine them in order to reduce the noise. This technique has many possible variations. Again biological applications are leading the way with 3D crystallographic reconstruction, as illustrated by Downing and by Dorset and even correcting for distortions in the specimen, as shown by Saxton et al.

31.6.C Correcting Drift

Although drift is not as limiting on new machines, many older TEMs are still in use. Drift can be corrected now if the rate and direction of movement are constant. The computer can calculate the relative translation of two images and change the current in the image translation coils appropriately (which avoids moving the specimen). The difficulty is that the drift may not be linear. When implemented, such routines are already particularly valuable for frame averaging using a video camera. There will also be many applications for diffraction-contrast imaging as well as for X-ray or EELS analysis.

31.6.D Reconstructing the Phase

Although we are studying phase contrast, the image intensity doesn't directly give us phase information. Kirkland et al. showed as early as 1982 that the phase can be reconstructed by processing a defocus series. In their approach they use an iterative non-linear image-processing technique to reconstruct the complex

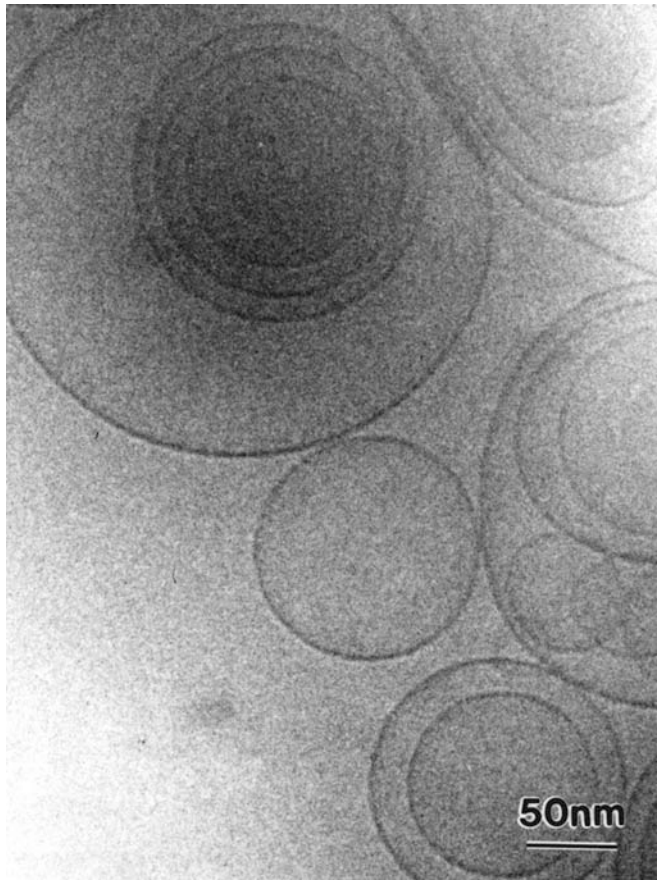


FIGURE 31.9. The image is from a highly beam-sensitive solution of surfactants in water. The solution has been frozen by plunging a film into liquid ethane and then transferring it to the TEM. The large circles are the surfactants that have aggregated to form vesicles; the concentration of the surfactants in the solution is just right for them to form lamellar structures (the circles seen flat on). Texture starts to appear in the image as soon as the beam interacts with the specimen due to crystalline ice formation and/or beam damage.

electron-transmission function. The technique was demonstrated using images of $CuCl_{16}PC$ (hexadecachlorophthalocyanine copper).

Five images from the experimental defocus series are shown in Figure 31.10A–E, together with the reconstructed transmission function plotted both as a real and imaginary part and then as an amplitude and phase part in Figure 31.10F–I. The projected structure of the known unit cell is also shown in Figure 31.10J. The phase image contains most of the structural information: it corresponds to the projected potential while the amplitude image contains features due to inelastic scattering. Notice in particular that we can now identify the benzene ring. This is one of the earliest published examples of the full-phase reconstructions. You must record such a series of defocus images if you want to do quantitative HRTEM. The companion text includes an extensive chapter giving details and an overview of how the process is practiced today.

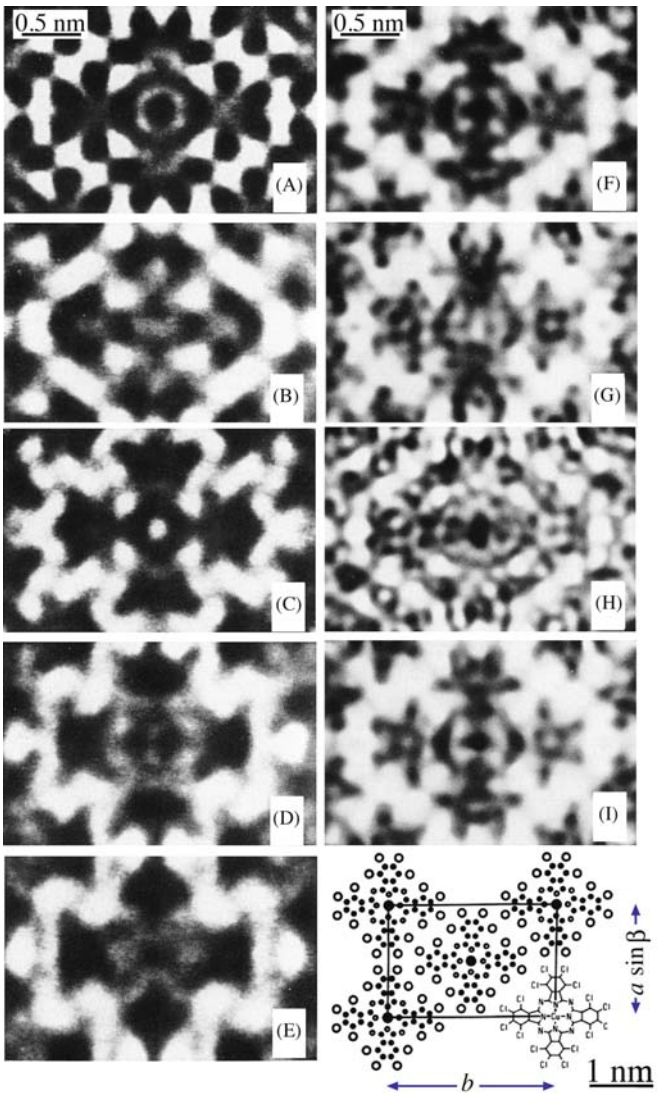


FIGURE 31.10. (A–E) Images from an experimental defocus series of $CuCl_{16}PC$; the reconstructed transmission function plotted as both a real and imaginary part (F,G) and then as an amplitude and phase part (H,I); (J) the projected structure of the crystal.

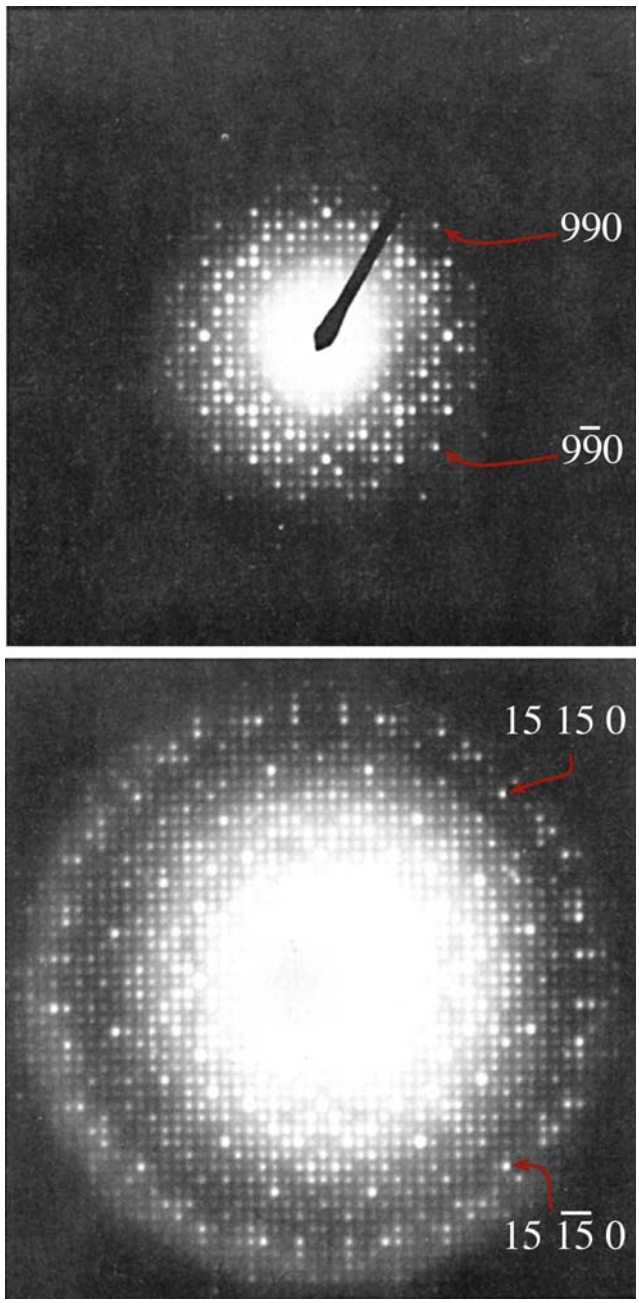


FIGURE 31.11. SADPs from $K_2O \cdot 7Nb_2O_5$ recorded using two different exposures. More than one exposure is needed to get all the information in the DP. The space group is P4bm and $a = b = 2.75$ nm. The (15, 15, 0) reflections correspond to spacing of 0.13 nm.

31.6.E Diffraction Patterns

We have generally ignored the intensities in DPs because they are so strongly influenced by dynamical scattering. However, if the specimen is very thin, we can use the intensity in the SADP to carry out electron crystallography in the same way as in classic X-ray crystallography. As you can appreciate from Figure 31.11, particularly if the unit cell is large and the specimen examined is thin there is a great deal of information in the SADP but you can't get it all in one exposure. Hovmöller's group (see, e.g., Zou and the ELD program in Section 1.5) has provided a routine for analyzing such patterns and getting structure-factor information. For example

- Several patterns are recorded using exposure times of 0.5 s to >15 s.
- The patterns are digitized directly from the negatives using a CCD camera and a light-box for backlighting.
- The intensity of each film is calibrated using a calibration strip with 20 equal exposure steps.
- The intensity is measured for all the spots and the processing begins.

Thus far it is just like the classic analysis of XRD intensities. This digitization process is particularly

demanding, because each reflection typically covers an area of <0.5 mm diameter on the photographic film. You will need to be able to index three strong, but clear, reflections. The computer can then perform a series of functions

- Optimize the location of these points using a center-of-gravity approach, locate the origin, and index the rest of the pattern.
- Extract the intensities of each peak, taking care not to be misled by any shape effects of the specimen.
- Use reflections which are present on two successive negatives (since the intensities are now in digital form) to calibrate films recorded with different exposures and thus develop a very large dynamic range.

A cooled slow-scan CCD camera will give you a large dynamic range and better linearity than its room-temperature counterpart and should simplify this type of analysis. There are other complications in using electrons rather than X-rays for this kind of crystallography. While the Ewald sphere is still curved, as with X-rays, electrons can easily damage your specimen. However, the technique clearly has potential! Like all TEM techniques it can be applied to much smaller regions of the specimen than is possible for X-ray beams. We can also use the symmetry present in the SADPs. Because the specimen is very thin, this technique could be described as 'kinematical' crystallography and complements the 'dynamical' electron crystallography that we described for CBED patterns from thicker specimens in Chapter 21. The process of extracting intensities from DPs can now be carried out using the ELD software package (Section 1.5).

USING THE DP

The structure deduced from the HRTEM approach should generate the experimental SADP so it should be possible to use these diffraction data to further refine the structure.

If we can use the quantitative information available in DPs, we could combine this information with our experimental and simulated HRTEM images. The quantitative analysis of the DP is known as structure-factor-modulus restoration or reconstruction (Tang et al.). One limitation of this approach is that the specimen be sufficiently thin that diffraction is kinematical. Of course, this requirement is necessarily similar to the HRTEM requirement of the WPOA.

31.6.F Tilted-Beam Series

Having gone to great trouble to remove any beam tilt, we will mention how beam tilt can be used to extend the resolution of your microscope! The basic idea goes back to the tilted-beam lattice-fringe imaging we discussed in Section 23.3. Now you use a computer to combine information in different tilted-beam images. The method proposed by Kirkland et al. assumes that you know when the beam tilt is zero. You tilt the beam through different angles in well-defined directions so that you transfer information in overlapping regions of reciprocal space, as shown in Figure 31.12A; you also need the on-axis image, as shown in the tableau in

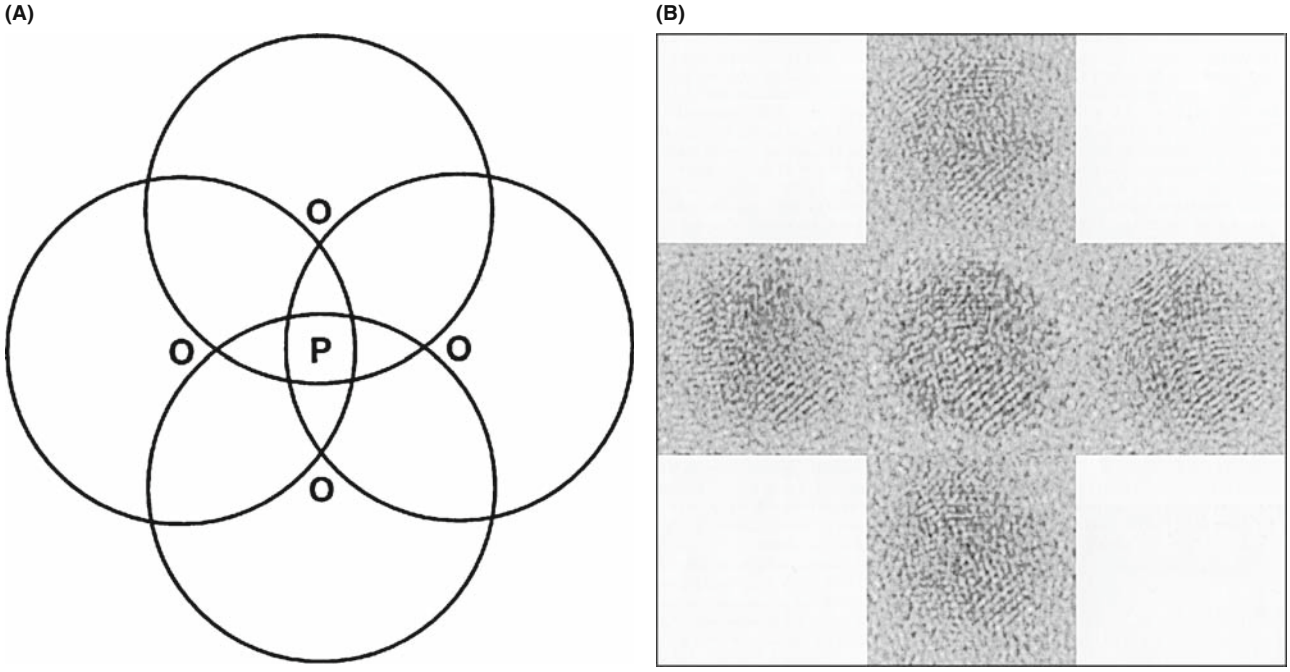


FIGURE 31.12. A method for extending the resolution of your TEM. Set the beam tilt to zero, then tilt the beam through different angles. (A) The four regions of Fourier space are shown by the four circles; O is each position of the tilted beam, P is the optic axis, and PO corresponds to the angle of tilt. (B) The five images used in the restoration arranged according to the beam tilt used in (A) with the on-axis image at the center.

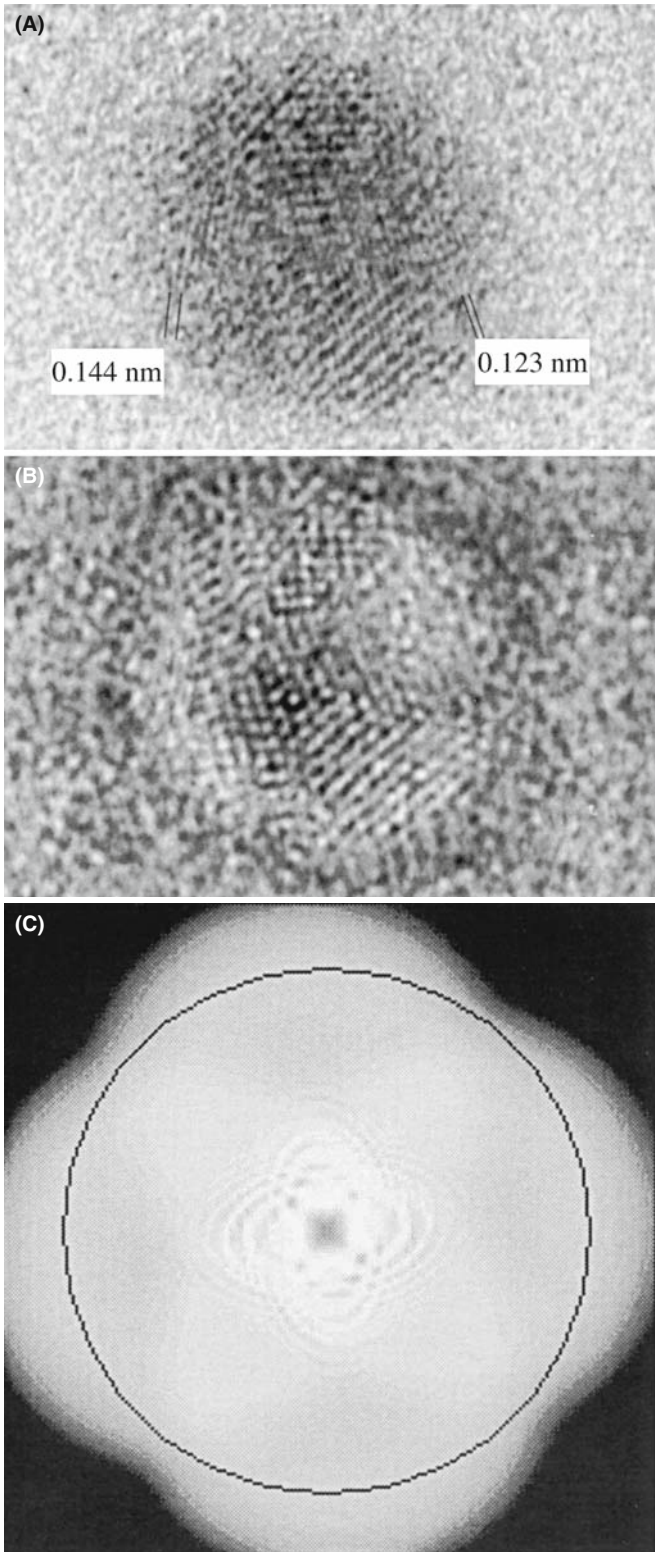


FIGURE 31.13. The restored image of a gold particle on amorphous Ge: (A) the amplitude (modulus) image showing 0.123-nm fringes, (B) the corresponding phase image, (C) the transfer function after restoration, plotted in two dimensions. The circle corresponds to 0.125-nm detail successfully transferred to the image. For such thin specimens, atom-dense positions have a reduced modulus [black in (A)] and an increased phase [white in (B)].

Figure 31.12B. Since it is important that the same area is imaged, a sixth (on-axis) image is recorded and correlated with the first (on-axis) image to check for drift and specimen degradation. You now need to restore the modulus and phase to create a higher-resolution image.

Kirkland's paper is a beautiful demonstration of the care needed in image processing. Even aligning the images is not trivial. However, the resulting restoration shown in Figure 31.13, demonstrates the potential of the technique: detail is present in the image at a resolution of 0.123 nm using a 400-kV microscope.

31.7 AUTOMATED ALIGNMENT

In the not-too-distant future, all TEMs will have automatic beam alignment, astigmatism correction, and readout for Δf. What makes this possible is the diffractogram analysis, a slow-scan CCD camera to digitize the image, and computer control of all the microscope functions. By microscope functions, we mean all lens currents, deflector currents, specimen drive, and aperture drives. The slow-scan camera is needed because the computer needs to make measurements on more than one ring in the diffractogram. So you will not actually have to sit in front of the microscope once you have loaded the specimen.

THE ADVANTAGE OF REMOTE TEM

The big advantage in remote control will not be that you can sit in Huntsville and operate a microscope in California but that you can locate the microscope in its controlled environment with no one opening the door to check if it is working (it was) or entering the room and thus changing the heat-load. If your specimen is not ideal or the TEM breaks down, you won't have to go for a walk on the beach in California but can continue word processing in Alabama.

Autoalignment is now available to everyone with a digital camera and the plug-in for Digital Micrograph (see Section 1.6). The early developments are summarized in the references. Your role is to select a suitable region of the specimen close to the area of interest; the area that you are interested in should ideally only be examined at low magnification. You will make the initial alignment manually and then ideally turn the process over to the computer. The computer will then adjust the astigmatism and correct the beam tilt independently and quickly.

Figure 31.14 shows how well and quickly this procedure can now be done. The different diffractograms in each tableau correspond to incremental changes in the beam tilt of 6 mrad in the x and y directions. The computer showed that the initial tilt error was 4 mrad, which

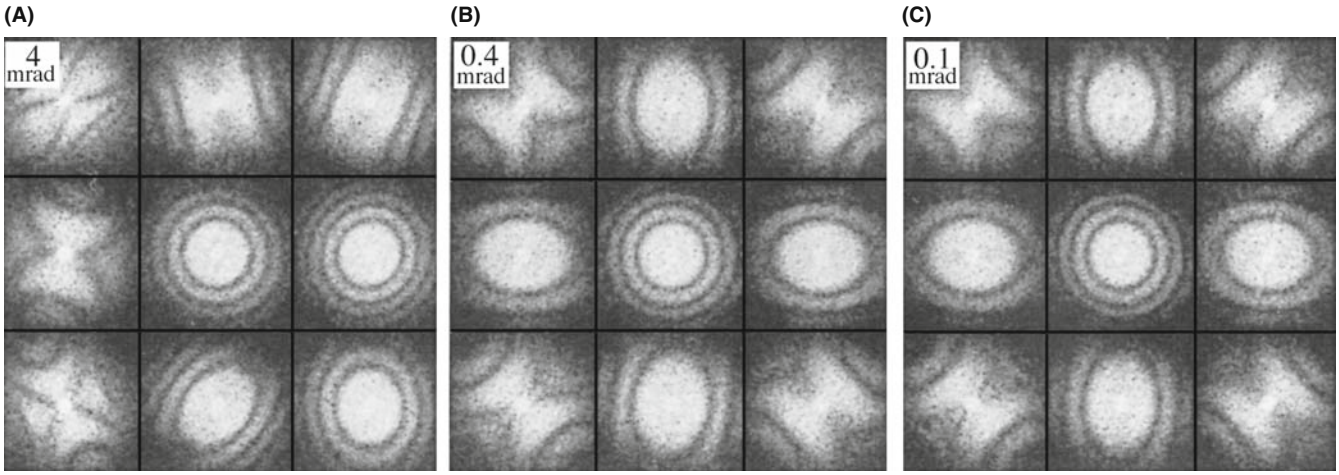


FIGURE 31.14. Using the computer to correct the beam tilt. (A) is the initial setting and (C) is the final. The incremental changes in the beam tilt are 6 mrad in the x and y directions away from the initial beam tilt in the (B). Note the central diffractogram is almost unchanged, emphasizing the need for computer-controlled tilting to give correct alignment. The misalignment in each setting is given in the upper left corner. (A) is the initial setting and (C) is the final.

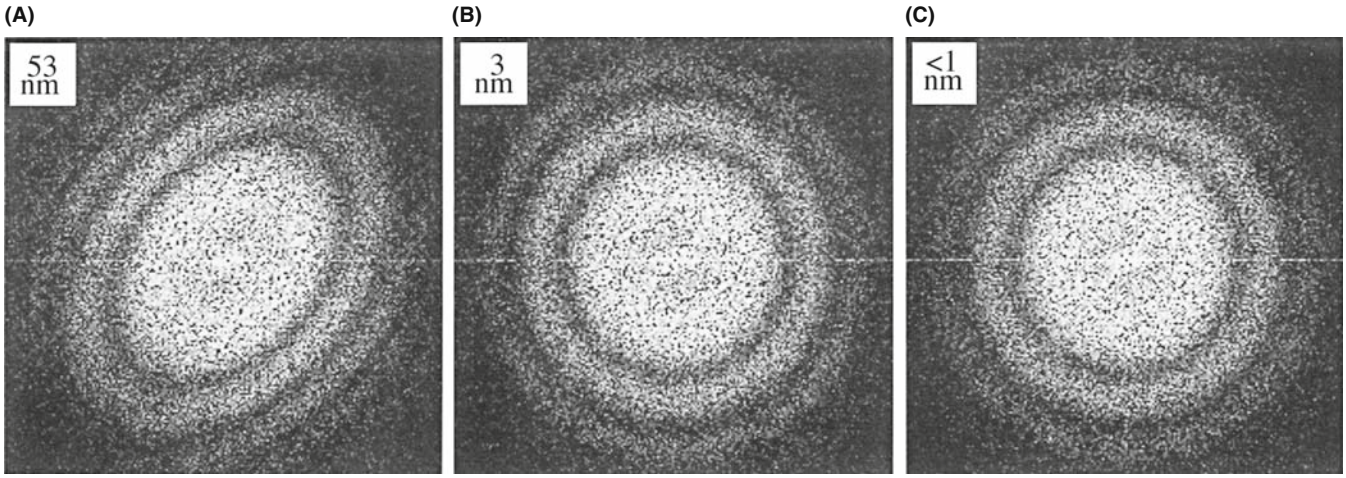


FIGURE 31.15. Diffractograms showing the astigmatism corrections made by the computer following a similar procedure to that shown in Figure 31.13. (A) is the initial setting and (C) is the final (and is not the same as (B)). (C) shows that the HRTEM is now well stigmated. The misalignment in each setting is given in the upper left corner.

was reduced to 0.4 mrad after one pass and <0.1 mrad after the second pass. Each pass took only 28 s! The astigmatism shown in Figure 31.15 was initially 53 nm. It was reduced to 3 nm after one pass and to <1 nm after the second pass. For this correction, each pass took only 8 s. Even the most experienced operator can't match this speed or accuracy for either correction, and both corrections are now *quantitative*.

The defocus value is then found by calibrating the image with minimum contrast occurring at Δf_{MC}. The value of Δf_{Sch} can then be found when the image contrast is a maximum. Although the method described here uses the diffractogram, a corresponding approach can be followed by analyzing variations in the contrast of the image. This technique has been described by Saxton et al. and uses a method of cross-correlating pairs of images recorded at each focus setting of the microscope. The reason for cross-correlating images is to remove the effects of electron shot noise; variations due to the photographic emulsion are avoided by using the slow-scan camera.

31.8 QUANTITATIVE METHODS OF IMAGE ANALYSIS

In the next six sections, we will go through several special topics and illustrations of image processing in HRTEM.

- Pattern recognition
- Parameterizing the image
- Chemical information from images

- Measuring the fit
- Quantitative comparison of simulation and experiment
- Fourier techniques

Our discussion will draw heavily on the work of a few pioneers in this field; we will also emphasize that, although this subject is still in its infancy, it is developing rapidly. The main cause for the delay in its application in materials science has been the lack of affordable fast computers and the feeling that everyone must write their own image-processing program; the latter is not true and is certainly not recommended. At this time, we can summarize the situation as follows

- Quantitative analysis is difficult, often tedious, and invariably time consuming.
- You must understand the basic ideas of image theory before you start processing images.
- Your analysis is only as good as your image and your image is only as good as your specimen.

We gave some information on software in Section 1.6 and expand on this in Section 31.16.

31.9 PATTERN RECOGNITION IN HRTEM

The most obvious feature of most HRTEM images is that we see patterns of white, gray, and black dots or other shapes. If the pattern is perfect everywhere, your specimen is probably a single crystal with no defects, no thickness variations, no variation in atomic composition, and no use. If it is not perfect, then we can use pattern recognition to quantify the variations.

PATTERN RECOGNITION

The idea is to take/make a template, move it across your image, and measure how closely the image resembles your template.

Clearly you need a computer for this! Your template needs to match the magnification and rotation of the pattern you are examining. Then you need a method to say how close your match is, i.e., you need to know your 'goodness of fit.' We will go through some basics here, but strongly recommend that you consult the list of original papers given at the end of this chapter when you are ready to apply this technique.

We can illustrate the approach following Paciornik et al. The large rectangle shown in Figure 31.16 represents your digitized image and could be 1 k × 1 k; remember the numbers indicate pixels. The small rectangle represents your template. This template might be a small area of the pattern or a simulated image, in which case it might be a 128 × 128 pixel template. If the template is taken from your image, then you have

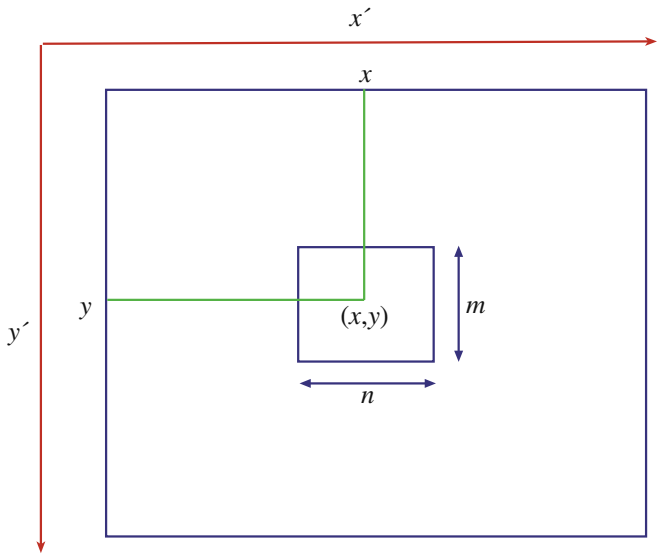


FIGURE 31.16. The large rectangle represents the digitized image, size ($x' \times y'$); the small rectangle, ($m \times n$) pixels, represents the template used in the cross-correlation calculation. The small rectangle is moved to different (x,y) positions during the process.

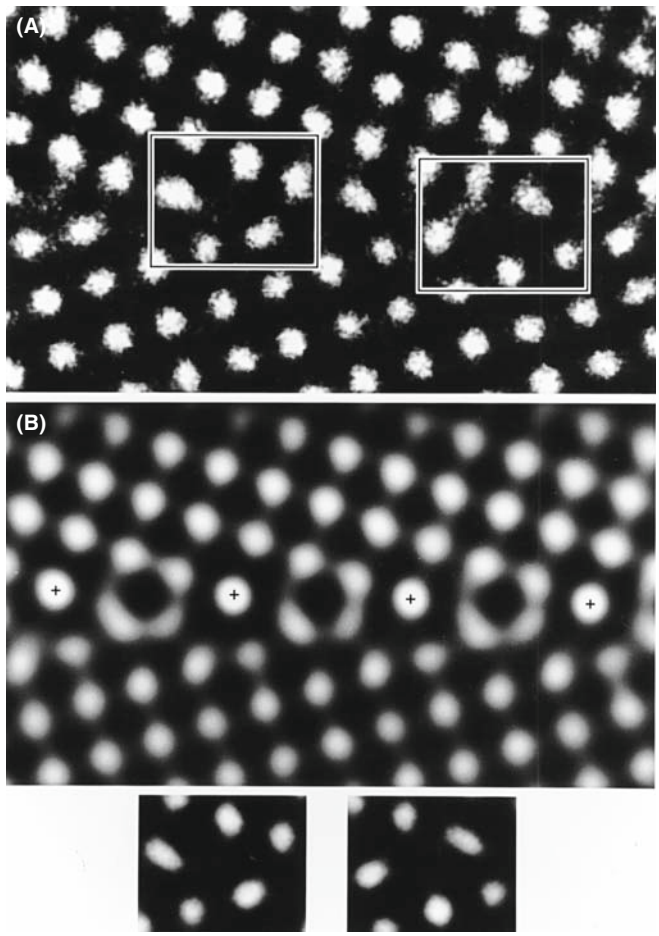


FIGURE 31.17. Analysis of a small region of a Σ = 5 tilt boundary in TiO_2. The two small boxed regions in (A) are only present at the boundary; these are used as the templates for the cross-correlation method. (B) The cross-correlated image. The small rectangles at the bottom of the figure are the low-noise-averaged images of the GB templates.

already got the right magnification and rotation. If not, you have to set these first, but remember that the lens will still cause distortions. We will return to this problem shortly. Klug actually used pattern recognition in the work that led to his Nobel Prize.

REAL-SPACE APPROACH
We only look at the image.

This is a real-space approach. The process is best understood by an example. Figure 31.17A shows an HRTEM image of a small region of a $\Sigma = 5$ tilt boundary in TiO_2. The two small boxed regions appear only at the boundary and are selected as templates. The matching process has then been carried out and the new image is shown in Figure 31.17B. Having found all the regions which match the template, we could then take the average of these to produce low-noise images of the GB templates. The final step is the comparison of these templates with models of the GB structure. There are two important points to remember

- When you average images, you implicitly assume that all the images are the same except for random noise.
- Don't forget our discussion in Chapter 30 of interface grooving and the problems associated with interfacial segregation.

31.10 PARAMETERIZING THE IMAGE USING QUANTITEM

In general, the thickness or chemistry will vary as you cross the specimen, i.e., the projected potential varies across the specimen. This means that one template will only match a small area, so you have to use many templates. These templates could, in principle, be totally empirical, but to be quantitative we must derive them from image simulations. This approach has been described for two special cases in papers by Kisielowski et al. and Ourmazd et al.

31.10.A The Example of a Specimen with Uniform Composition

In QUANTITEM, the results of Chapter 30 are summarized by a general equation linking the intensity and all the imaging (S_i) and materials (P) parameters

$$I(x,y) = F(P(x,y), S_i) \tag{31.5}$$

This equation just tells us that the intensity depends on the imaging conditions and on the specimen. For a particular set of imaging conditions, S_i will be known (more or less) and we'll call it S_i^0. Then we can write that

$$I(x,y) = F\left(P(x,y), S_i^0\right) = F^0(P(x,y)) \tag{31.6}$$

The basis of this approach is quite straightforward, potentially powerful but almost never used

- Define the function F^0 for each image that you may obtain.
- Then construct a set of templates for your matching process.

Providing you stay within one extinction band, F^0 will be directly related to the projected potential of the specimen. A helpful simple analogy for this process is illustrated in Figure 31.18. The function F^0 describes the path of a swinging pendulum as it varies with time (Figure 31.18A). Each value of F^0 corresponds to a snapshot of the pendulum, so if you plot F^0 you can 'see' the path of the pendulum (Figure 31.18B). The velocity of the pendulum is related to the density of points along the path. So it should be possible to plot out the function F^0 from a single lattice image even if you don't know the microscope parameters used to form the image.

Yes, there are limitations and conditions and we'll discuss them later. All we need now is a method for representing each image by a snapshot of the pendulum: we have to *parameterize* the image. This process is the key to the technique. Manipulating and quantifying, in principle, thousands of images, each requiring 4 M byte of memory is not a fast process, even if you do have that much memory. If we could characterize each image by a few numbers (a vector or parameter) the comparison process could be much faster.

We separate the image into unit cells and digitized these to give many templates which are n pixels by m pixels, as shown here for one cell (Figure 31.18C–F). If we define N to be $n \times m$, then we have N numbers for the N pixels, where each number represents a gray level. Now the N numbers are regarded as the N components of an N-dimensional vector. (The math is not complicated but don't try to visualize this vector.) So now all the information in each unit cell is represented by a vector in N-dimensional space. The function F^0 describes how these N-dimensional vectors (the path) change as the projected potential changes.

PARAMETERIZING IMAGES
This may remind you of using the Hough transform to quantify EBSD patterns from the SEM.

The next step is to define a reference frame for these vectors. Three basis vectors are derived from the experimental image. Ourmazd et al. argue that three basis vectors will be sufficient, as we can show in the

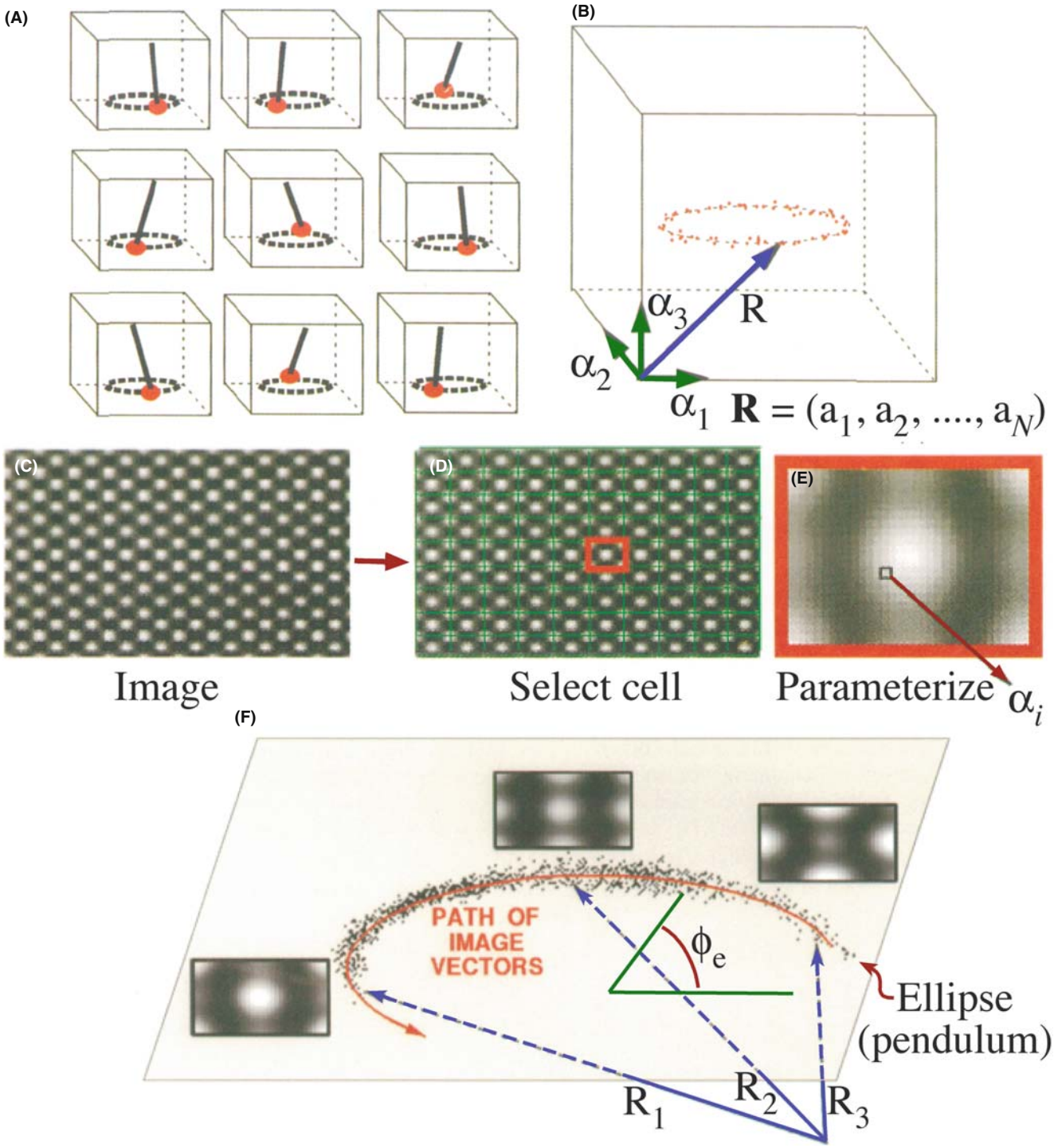


FIGURE 31.18. The principle of vector parameterization used in QUANTITEM. Each HRTEM image is represented by a single vector **R** which has N dimensions. (A) and (B) show the 'swinging pendulum'; (B) shows its 'path'; (C–E) the image is separated into unit cells and digitized to give ($n \times m$) pixel templates; (F) three vector-parameterized images ($\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{R}_3$) of a wedge-shaped specimen of Si at different thicknesses.

following way. We will be using a low-index zone axis for any HRTEM analysis. Then we have three types of images

- The background, $\mathbf{R}^{B}$, due to the direct beam, O.
- A single-period image, $\mathbf{R}^{S}$, due to the interference between O and the strongest reflections, G_i.
- A double-period image, $\mathbf{R}^{D}$, due to the interference between these strong G_i reflections.

Each of these **R** terms is a vector which represents an image. Any image we can form must be a combination of these three types of image, so a general image, G, can be written as

$$\mathbf{R}^{G} = a_{G}\mathbf{R}^{B} + b_{G}\mathbf{R}^{S} + c_{G}\mathbf{R}^{D} \tag{31.7}$$

Each of the basis vectors (images) can be expressed in the same manner

$$\mathbf{R}_{i}^{T} = a_{i}\mathbf{R}^{B} + b_{i}\mathbf{R}^{S} + c_{i}\mathbf{R}^{D} \tag{31.8}$$

giving three vectors for $i = 1$, 2, and 3.

We can, in principle, turn these equations around (invert them) to define any vector $\mathbf{R}^{G}$ in terms of the basis vectors

$$\mathbf{R}^{G} = \alpha_{G}\mathbf{R}_{1}^{T} + \beta_{G}\mathbf{R}_{2}^{T} + \gamma_{G}\mathbf{R}_{3}^{T} \tag{31.9}$$

which is what we wanted to show.

Ourmazd et al. point out that this treatment gives three important results

- The vector notation allows us to parameterize the lattice image.
- Projecting the vectors onto planes and/or paths aids noise reduction.
- Any noise which remains can be quantified.

The result of vector-parameterizing an experimental image of a wedge-shaped specimen of Si is shown by the three images corresponding to $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$ in Figure 31.18F.

31.10.B Calibrating the Path of R

In order to relate any image to the projected potential, we have to calibrate the curve showing the path of $\mathbf{R}^{G}$. This is where the image simulation comes in. We start with the vector-parameterized analysis of the series of simulated images from the Si wedge. Each point on the curve corresponds to an image of the unit cell, and thus to a vector $\mathbf{R}^{G}$. The ellipse has been fitted empirically, and the thickness of the cell has been increased by 0.38 nm for successive calculations. The points are closer together in some parts of the plot because, as we saw in Chapter 30, some characteristic images appear for a wider range of thicknesses. Now we have a way to quantify this 'experimental' observation. What the ellipse does is to allow us to parameterize the path in terms of the phase angle of the ellipse ϕ_e, shown by the ellipse in Figure 31.18F. Thinking back to the pendulum analogy, the path parameters are the image version of the coordinates for the harmonic oscillator.

Now the parameters of the curve (ϕ_e) can be obtained from a series of images. We can vary the material and in each case examine three other variables

- The orientation of the specimen (i.e., the zone axis).
- The defocus, Δf, of the objective lens.
- The specimen thickness.

The remarkable result is that when we plot φ_e versus the thickness, normalized by the extinction distance, we obtain a straight line. The explanation for this result is related to the fact that only a small number of Bloch waves usually contribute to the image, as we saw in Section 30.6. In materials such as YBCO, this is not the case, and the curve is not a straight line.

This analysis gives more detail than most students will ever need, but it introduces a concept that could (should) be more widely recognized: the power of parameterizing the image.

31.10.C Noise Analysis

Noise in the image moves the vector off the ellipse; so, we have a method for analyzing the noise. If the noise moves the vector exactly along the ellipse, we can't analyze the noise, but that is quite unlikely since the noise would then be accurately mimicking a change in projected potential. So this parameterization gives a way to reduce the noise by a factor of $\sqrt{N}$ which for a 10 pixel $\times$ 10 pixel cell is a factor of 10!

The analysis given by Ourmazd et al. then shows that, in the case where only two Bloch waves are excited, the image intensity, I, can be expressed as

$$I = B + S + D \tag{31.10}$$

where B, S, and D are the contributions from the background, single interaction, and double interaction as we defined them above. The point (B, S, D) does indeed describe an ellipse which lies on a plane independent of Δf.

The value of this approach can be appreciated if you look at the examples shown in Figure 31.19. In the first example (Figure 31.19A and B), the technique has been used to provide a map of the roughness of the Si surface. The experimental image looks really uniform until you analyze it using this method, when you can discern the roughness at the 0.5-nm level; remember that there are two surfaces.

As you know from earlier discussions, changes in chemistry produce effects which are similar to changes in thickness, because they change the projected potential. In terms of the present analysis, the effects are different: composition changes cause changes in the ellipse and in ξ (notice that there is no subscript, since this ξ is a many-beam value).

The method is more limited, in this respect, to change in composition but can be used if the thickness and roughness are known, i.e., if you can measure the roughness elsewhere on your specimen (using a known reference cell) and infer it for the area you want to analyze. (Warning lights should be flashing.) The approach is as follows

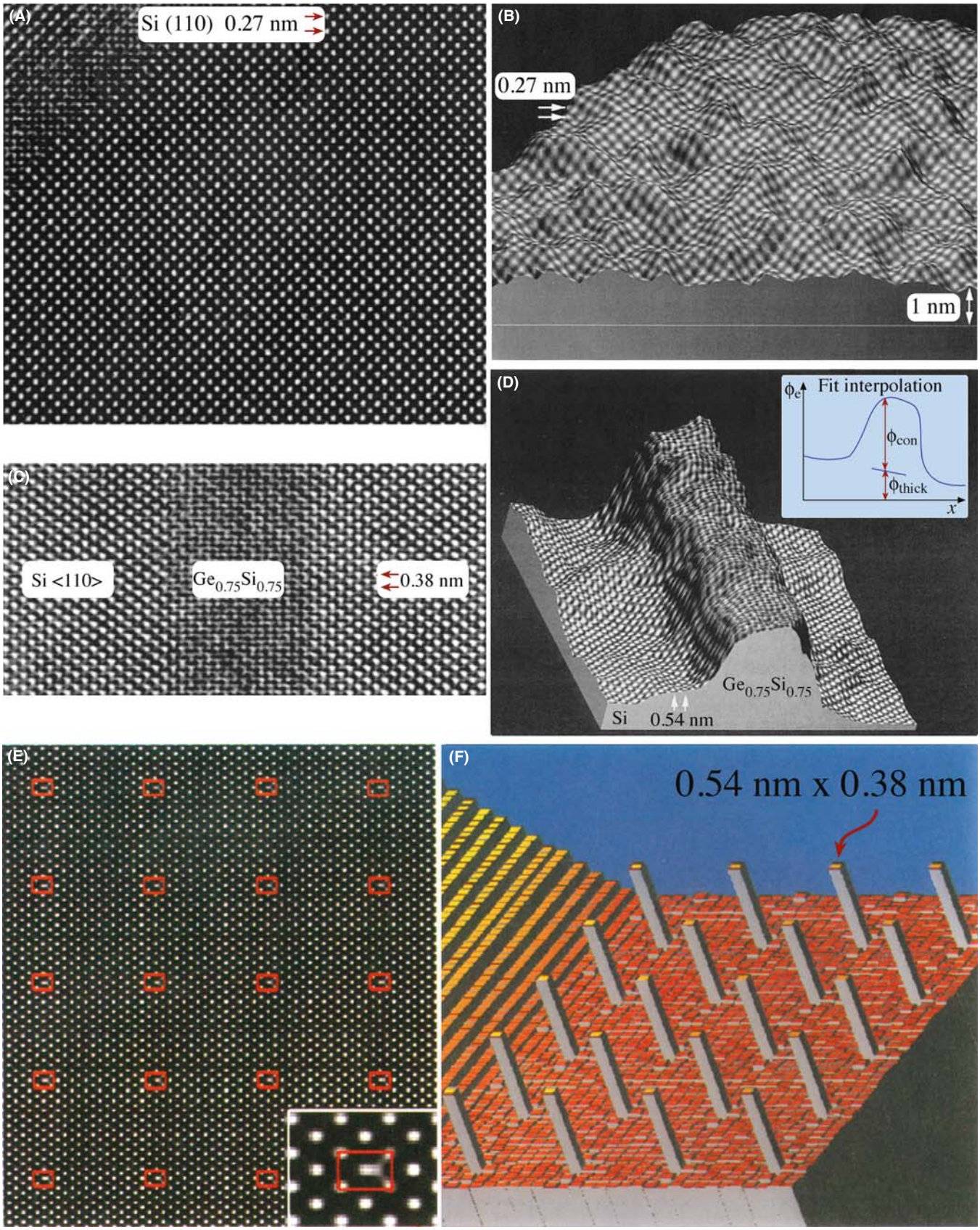


FIGURE 31.19. Examples of the application of QUANTITEM: on the left is the image, on the right is the QUANTITEM image. (A, B) Mapping of the roughness of the Si surface covered by SiO_2; (C, D) a layer of Ge_xSi_{1-x} in a matrix of Si, the inset shows the plot of ϕ_e versus x; (E, F) analyzing a simulated image of columns of Ge (a δ-function in concentration) in Si.

- Use QUANTITEM to measure the advance in ϕ_e at your target cell relative to your reference cell.
- Subtract $\Delta\phi_e$, which is due to a thickness change.
- Then the rest of the change in ϕ_e must be due to changes in ξ. If you know how ξ varies for different compositions, you have determined the local composition.

The second example shown in Figure 31.19C and D is a near-perfect application for the technique, since the elements in the alloy Ge_xSi_{1-x} are randomly located on the lattice sites. The slope tells us how abruptly the composition varies.

NOISE

You can test the potential resolution of the technique, its sensitivity to the alignment of the beam, bending of specimen, and beam divergence in the usual way by creating model structures, simulating the images, and then analyzing them.

The third example in Figure 31.19E and F shows that the potential resolution is superb (we can see the quantum dots) but beam tilt can cause 10% errors in thickness measurement. The conclusion is clear: as always, you only get the best results if your specimen is ideal and both the microscope and the specimen are perfectly aligned. Note, however, that the technique has not yet been successfully applied to a wide range of materials, but it is complementary in many ways to STEM *Z*-contrast (see Figure 22.15).

31.11 QUANTITATIVE CHEMICAL LATTICE IMAGING

This technique is an instructive tool but is perhaps less necessary now that STEM is commonly available. The technique has not been widely used since the software was not readily available.

This technique uses the approach described in Section 30.4, but can only be applied to materials where we have chemically sensitive reflections, which we discussed in Section 16.4. We used these reflections in Section 29.4 to produce chemically sensitive DF images. In HRTEM, the chemically sensitive reflections not only contribute to the overall image but they will generally have a different dependence on thickness, too.

This effect is shown in Figure 31.20A–D for AlAs and GaAs, which have identical structures: the 002 reflection is allowed for both, but is stronger for AlAs since *F*, the structure factor, is proportional to $f_{III}-f_V$. You can see that, under the conditions chosen for this comparison, the intensity of the 022 reflections is also very different for the different thicknesses. The sort of image we can analyze with this approach is illustrated here by the quantum wire. We want to know how abruptly the composition changes at the interface. In this example, the ideal GaAs and $Al_{0.4}Ga_{0.6}As$ unit-cell images are first characterized by the two vectors, $\mathbf{R}_{GaAs}$ and $\mathbf{R}_{AlGaAs}$, following the approach described in Section 31.10. In this case the cells were simulated, divided into 30 × 30 pixel arrays (so $N = 900$), and then **R** was plotted. The information content is contained in θ_C. As before, we can directly assess the noise in such an image. So how is the direction of **R** dependent on composition?

The technique is explained using the three known simulated templates (GaAs, AlGaAs, and AlAs); each one produces a vector $\mathbf{R}^t$. Although the vector for the intermediate composition does not lie in the plane, it can be projected onto this plane to give a unique vector for certain ranges of thickness. Since this is a complex procedure, you'll find the 'image flow chart' shown in Figure 31.21 helpful.

- The experimental image is first digitized; the image contained approximately 25 × 25 unit cells and used a 514 × 480 frame buffer.
- Next, the image must be separated into individual cells.
- The pair of templates shown is then used to calculate the angular positions of the **R** vectors for all the unit cells. Such templates can be calculated or taken from known areas of the specimen.
- These **R** vectors are characterized in terms of where they cut through a plane (D) (see Figure 31.21D also).

The maximum chemical difference determines how far apart the two principal distributions can be (see Figure 31.21E). Since the image is now fully parameterized we can do the statistics and finally invert the angular data to give the compositions.

This technique has enormous potential, but you must also remember that it is susceptible to all the drawbacks inherent in HRTEM. The advance is that now you can put numbers on those effects. The technique is material-specific, but if you know your material, you can combine image simulation and this processing method to examine what will be the limiting factors for your material. You can construct a test image like that shown in Figure 31.19E. If your specimen is ideal, you could, in principle, easily detect a column of Al in a mainly GaAs matrix without any 'spreading' due to the electron beam and in this respect alone makes QUANTITEM better than atomic-resolution X-ray mapping and equivalent to atomic-column EELS (see Part 4). This technique could be applied to other materials.

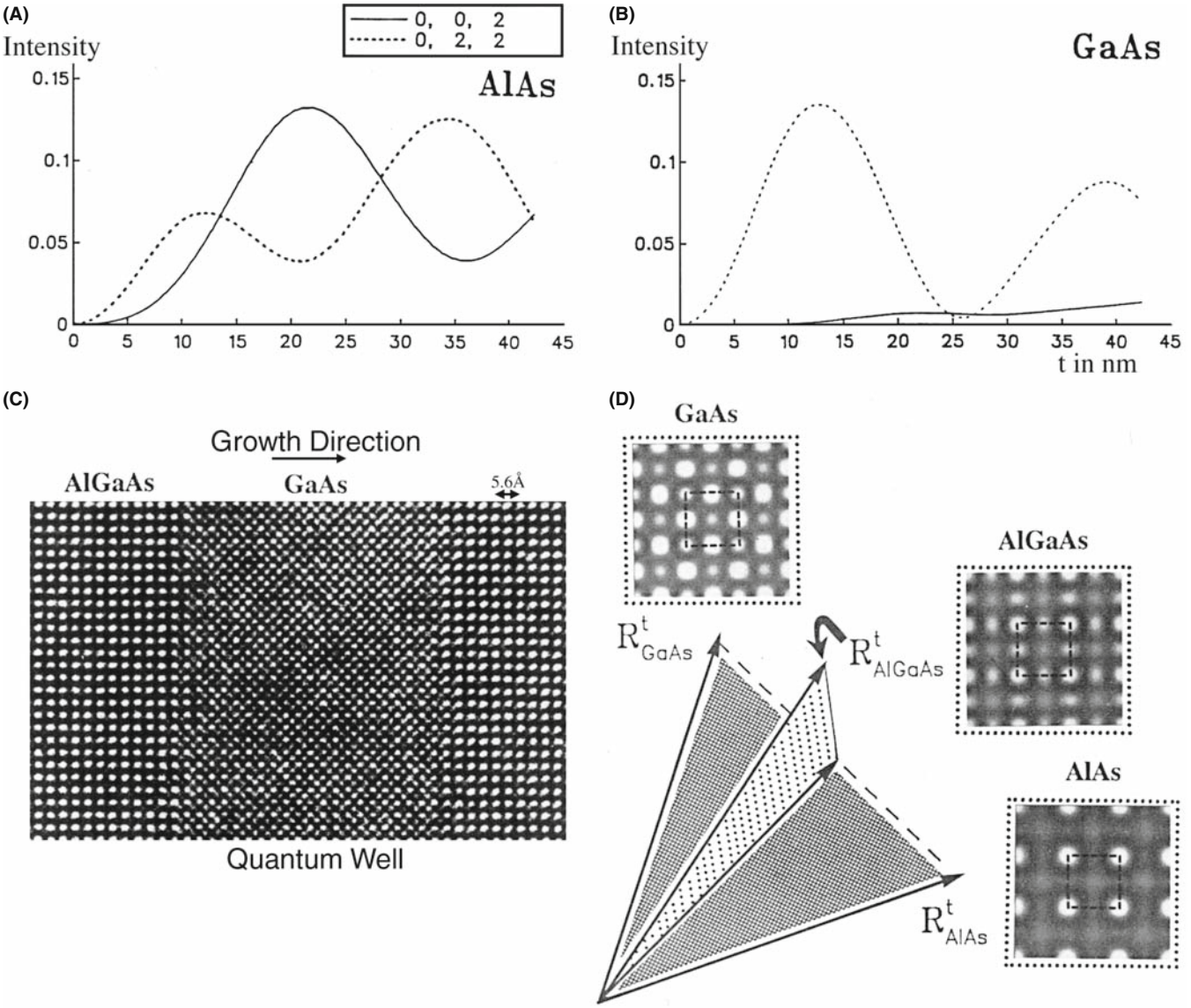


FIGURE 31.20. (A,B) Variation in intensity of the (002) and (022) beams along [100] in AlAs and GaAs (400 keV); (C) chemical lattice image of a layer of GaAs between two layers of $Al_xGa_{1-x}As$ ($x = 0.4$); (D) templates simulated for different values of x each produce a vector $\mathbf{R}^t$.

31.12 METHODS OF MEASURING FIT

There are two methods presently used to obtain a measurement of how well the data and the model match. In statistics, we determine this 'goodness of fit' using cross correlation or a least-squares refinement (Section 31.13).

CROSS CORRELATION

The cross correlation gives a measure of how similar two images (or signals) are. The math process looks very much like a convolution of two functions but we don't take the Fourier transform. What's special is that the computer does it easily as long as our images are numbers!

LEAST-SQUARES

A method of fitting a curve to a series of values.

In this section we'll use the cross-correlation method to compare an $n \times m$ pixel template (see Section 31.9) with every possible $n \times m$ rectangle in the image. The computer moves the template across the image one pixel column at a time, then shifts down one pixel row, and repeats the exercise. The cross-correlation function (CCF) gives the goodness of fit or a 'measure of similarity' between the template and each $n \times m$ image

$$CCF(x,y) = \frac{\sum_{x'}\sum_{y'}[i(x',y') - \langle i(x',y')\rangle] \cdot [t(x'-x,y'-y) - \langle t\rangle]}{\sqrt{\left\{\sum_{x'}\sum_{y'}[i(x',y') - \langle i(x,y)\rangle]^2 \sum_{x'}\sum_{y'}[t(x'-x,y'-y) - \langle t\rangle]^2\right\}}} \tag{31.11}$$

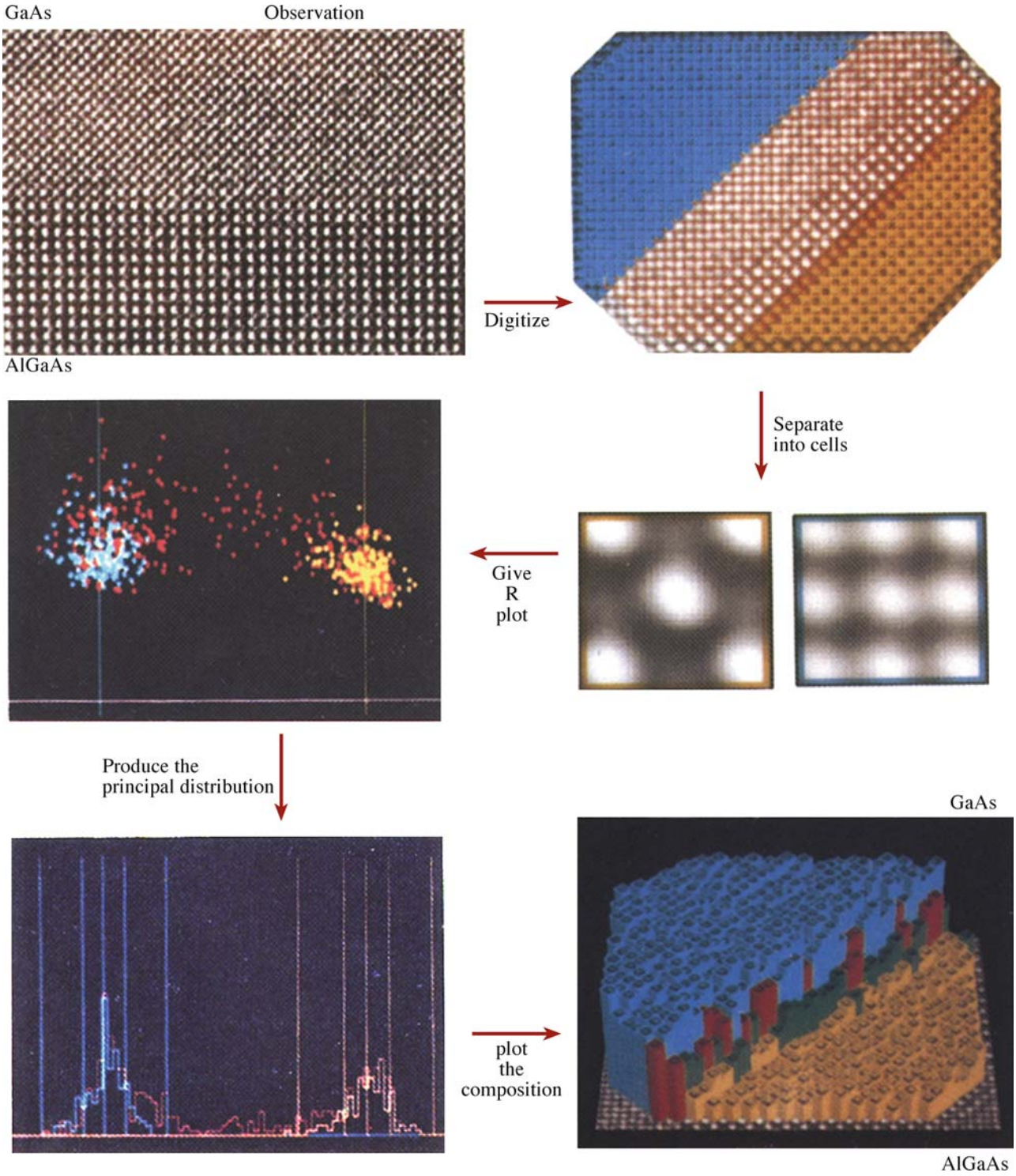


FIGURE 31.21. A flow chart summarizing the pattern-recognition procedure.

In this equation x varies from 0 to x_{max}, y varies from 0 to y_{max}.

- $i(x', y')$ represents the image.
- $t(x', y')$ represents the template.
- $<t>$ is the average value of the pixels in $t(x', y')$; it is computed just once.
- $<i(x, y)>$ is the average of $i(x', y')$ in the region coincident with the current location of t.

The summations are taken over the coordinates common to both i and t. The origin of the image is at its top left corner and the origin of the template is at its center. In this equation, the denominator is a

normalization factor, so that the CCF will not depend on any difference in the intensity scale between the template and the image.

We can rewrite equation 31.11 as the dot product of two vectors **t** and **i**, which gives us the $n \times m$ component of the template

$$\mathrm{CCF}(x, y) = \cos(\theta)\frac{\mathbf{t}\cdot\mathbf{i}}{|\mathbf{t}||\mathbf{i}|} \qquad (31.12)$$

Now we can plot the CCF as a map of our image and then examine it to deduce where there is a particularly good match. Since the CCF value varies from 0 to 1, we can plot out the number of times each particular CCF value occurs, as illustrated in Figure 31.22. The two peaks in this curve correspond to the best fit and the worst fit, so the distance between them gives a measure of the 'discrimination signal.' From the width of the peaks we have a measure of the noise and hence a signal-to-noise ratio. The regions of good fit can be combined to produce a better template, and the process repeated, giving the dashed line. A second measure of the noise is then given by how far the good peak differs from unity. A particularly nice feature of this approach is that the procedure is available as a plug-in module for Digital Micrograph (see Section 1.6). The template could alternatively be a simulated image and the process repeated for a series of different thicknesses and/or defocus values. When you want to learn more about correlation techniques, see the article by Frank.

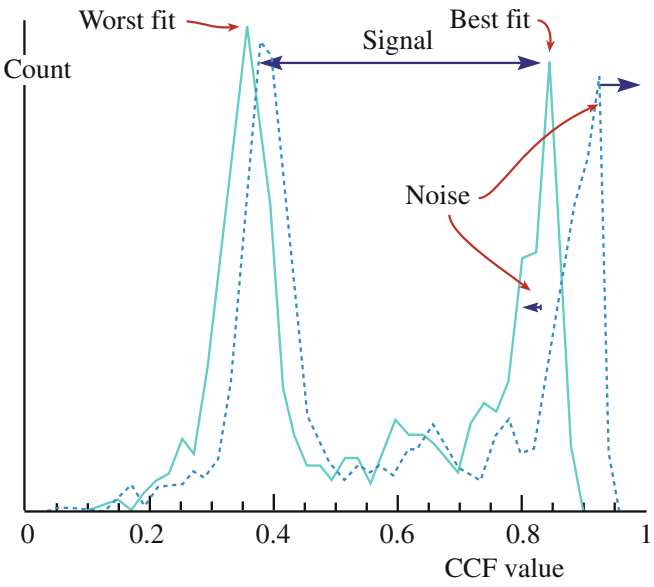


FIGURE 31.22. A plot of how often a particular CCF value occurs. The two peaks in the continuous curve are the best fit and the worst-fit: their separation gives a measure of the discrimination signal; the width of the peaks gives a measure of the noise and hence a signal-to-noise ratio. The plot can be redrawn after repeating the process (dashed line) to estimate the improvement in signal-to-noise ratio.

31.13 QUANTITATIVE COMPARISON OF SIMULATED AND EXPERIMENTAL HRTEM IMAGES

If we want to compare simulated and experiment images quantitatively, we really should modify our usual approaches to both simulation and experiment (King and Campbell 1993 and 1994). When doing the simulation, most programs automatically adjust the gray scale for each image so that darkest is 1 and brightest is 0 (or vice versa). This means that two simulated images might appear similar even though you would hardly see the pattern in one if both appeared on the same negative. In a similar way, we usually print an image to be as clear as possible using the full range contrast of the photographic paper.

We need methods for normalizing these procedures if we want to make quantitative comparisons. The solution for the simulation is simple. For the experimentalist, it means recording extra data while you're at the microscope. After recording the image, you record another image with the specimen removed. You then use this image to scale the lattice image such that you correct for variations in intensity across the field of view and the non-linearity of the response from the photographic film. Figure 31.23 illustrates the experimental transmittance for Kodak SO-163 film, 400-keV electrons, plotted against the digital value on a CCD array. Of course, you must process both images at the same time. This is called the 'flat-field' correction; a slow-scan CCD camera would simplify this procedure at the cost of reducing the area you examine.

When you analyze the image, you'll find out if the area you photographed was correctly aligned. Since your image takes only 2 seconds or so to record, you

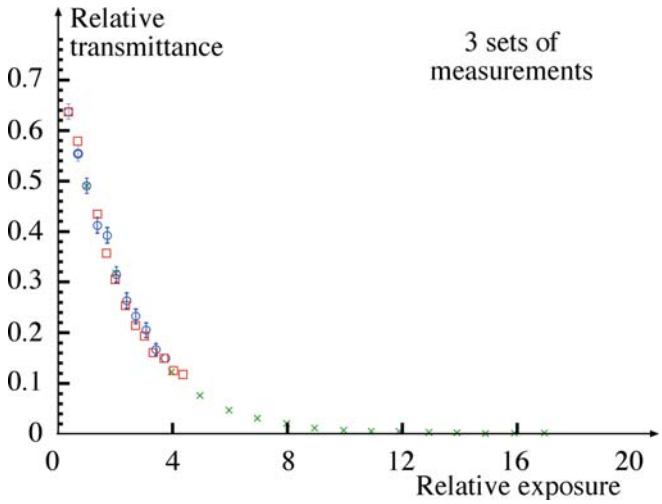


FIGURE 31.23. Plot of transmittance versus relative exposure measured using a CCD camera to digitize images from Kodak SO-163 film. (The symbols indicate three different microscopes.)

may risk several exposures using this technique. You are now comparing numbers, so you can use a least-squares fit where the residual $f_i(x)$ is defined as

$$f_i(x) = \frac{\left[f_i^{\text{obs}} - f_i^{\text{calc}}(x)\right]}{W_i} \tag{31.13}$$

and your task is to minimize $f_i(x)$. The difference between the intensity in the experimentally observed ith pixel and its calculated value will be zero if everything has been scaled correctly, the imaging conditions (Δf, C_s, etc.) are correct, and you have the right structure.

HRTEM AND LOW DOSES

Don't forget that since HRTEM uses higher voltages, the perfect image will only be recorded from an area of the specimen that has only seen the beam while you recorded the image! So you should always use low-dose techniques for quantitative imaging (but others don't).

Let's say W_i is the image which represents the error bars for pixel i. Then we can write that

$$W_i = \min\left[\sum_{i=1}^{N} f_i(x)^2\right] \tag{31.14}$$

This equation defines the non-linear least-squares problem. We use x to summarize a set of parameters (Δf, C_s, the model, etc.); N is the number of pixels in the image. Fortunately, this analysis is now routine statistics. You'll need a computer program to tell you how good the first guess was, make an improvement and continue until it meets our specific criterion for matching. [King and Campbell used MINIPACK-1. (See articles by Moré et al.)]

In their demonstrations of this approach to analyze a [001] tilt GB in Nb, King and Campbell varied four parameters: thickness, defocus, x-tilt, and y-tilt. The steps were as follows

- They first optimized the electron-optical parameters using a 64 × 64 pixel image, giving $N = 4096$ and an image computational cell of 3.303 nm by 3.303 nm. Using the EMS program (Section 1.6), the optimization took 20 iterations and 80 multislice calculations.
- Next they had to optimize the structure of the GB. This process required defining 84 atomic positions in a unit cell of 4.16 nm × 1.04 nm and a 512 × 128 (= 65,536) pixel image. Now the optimization required 16 iterations and 1300 multislice calculations.

These numbers are instructive. First they tell you that this computation can be done, which wasn't obvious. Second they tell you that this is a computer-intensive process; that we could have guessed!

You'll need to take enormous care in this type of analysis

- Align the simulated cell with the experimental cell and measure the unit cell in pixels.
- Choose a number of cells and relate them by the translation vector parallel to the rows of the image array.
- Calculate the standard-deviation images.
- Rotate the unit cell and repeat the exercise several times.

The orientation which gives the smallest standard deviation is your alignment. You must now adjust the magnification of the experimental image to fit the simulation, in a similar way to what you did for rotation. Next, you have to match the origins of both cells; the procedure is the same as we just described but translating the unit cell not rotating it. For a bicrystal, you now repeat this exercise for the other grain and then for the GB. You can improve the fit further if you take account of a constant background contribution, which probably arises due to the amorphous layer on both surfaces. Comparing experimental and calculated images quantitatively, we define f_i^{obs} as the corresponding value in the simulated image. We then calculate the residual $f_i(x)$ as follows

$$f_i(x) = \frac{\left(f_i^{\text{obs}} - \left(f_i^{\text{calc}}(x) + b^{\text{fit}}\right)\right)}{W_i} \tag{31.15}$$

where b^{fit} is included as a free parameter in the optimization procedure. King and Campbell's calculations showed that W_i could be expressed as

$$W_i = \sigma_i^{\text{obs}} + 0.05 f_i^{\text{obs}} \tag{31.16}$$

where σ_i^{obs} is the standard deviation of the ith pixel. Examples of the experimental, best fit, and normalized residuals are shown in Figure 31.24 for images from a $\Sigma = 5$, (310), [001] GB in Nb.

31.14 A FOURIER TECHNIQUE FOR QUANTITATIVE ANALYSIS

Möbus et al. proposed using what is referred to as an adaptive Fourier-filtering technique. The HRTEM image is digitized in the usual manner and then a special spatial-frequency filter is applied. This type of mask is designed for analyzing regions which contain defects.

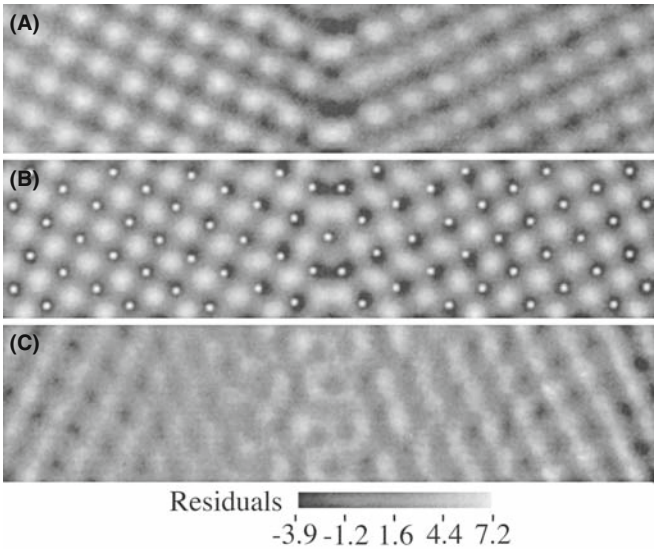


FIGURE 31.24. (A) Experimental image, (B) best fit simulation, and (C) normalized residuals of a $\Sigma = 5$ symmetric tilt boundary in Nb.

So the idea is that the computer automatically optimizes the mask to maximize the separation of the signal and the noise. This approach has not been widely practiced in TEM but clearly holds enormous promise. By varying the mask, this approach can prevent the analysis of a defect layer being dominated by the bulk information. Since the approach is quite straightforward signal processing, we will just illustrate an example found in the analysis of a simulated $\Sigma = 5$ GB with an extra period along the boundary. To test the analysis, white noise was added to a calculated image to give the image shown in Figure 31.25A. The power spectrum (the computer-generated DP) of the micrograph is shown in Figure 31.25B. The adaptive filter and the filtered image are shown in Figure 31.25C and D. The important feature of the adaptive filter in this example is that it was created as such because the computer detected the doubling of the periodicity which is *only* present in the GB. Second the mask consists of elongated openings which we know we need when analyzing the GB because of the shape effect that we discussed when analyzing DPs (see Chapter 17).

AN ADAPTIVE FILTER
One where the shape of the filter, or mask, is adapted to fit the shape of the 'image' it's filtering.

31.15 REAL OR RECIPROCAL SPACE?

In principle, we could equally well compare two images in reciprocal space rather than real space. However, because the Fourier transforms can generally be carried out much faster, the real-space approach has several advantages

- Fourier analysis separates local information into sine and cosine functions which are delocalized. When we reassemble the real-space image, higher parts of the frequency spectrum will be lost which will degrade the resolution.
- We want to maintain information on the absolute value of the intensities.
- Real-space methods are visually more intuitive for most of us. We can easily see what we have removed in the process.
- The real-space approach allows us to choose any values of n and m in defining our templates. Fourier space prefers aspect ratios given by 2^n.

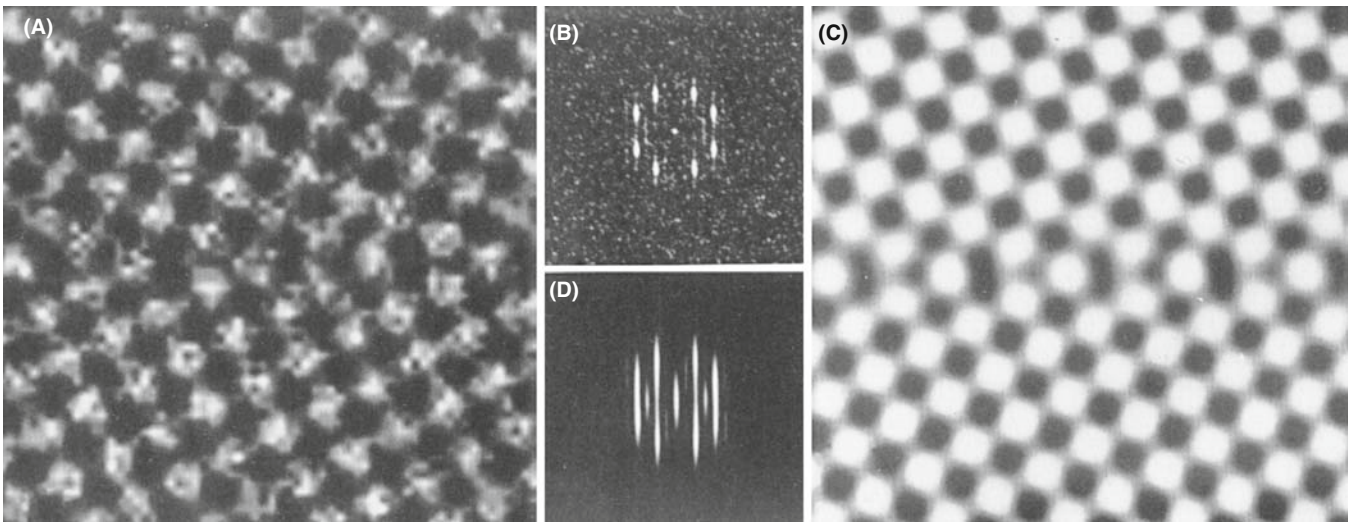


FIGURE 31.25. (A) White noise added to a calculated image of a $\Sigma = 5$ GB; (B) the power spectrum of (A); (C) the adaptive filter; (D) the filtered image.

31.16 SOFTWARE

Several software packages are widely used by the TEM community and have been listed in Section 1.6. We summarize the ones we know in Table 31.1. In its earliest application, image processing in TEM was almost exclusively applied to HRTEM images. This is no longer the case. Remember: always start with the best possible data. You can't always obtain a perfect image because your specimen might be beam sensitive, or coated with oxide, and you need to be aware of these limitations when processing or quantifying the image. This chapter has given you a hint of what is possible and where the subject is developing. We recommend that you obtain the software and start experimenting.

In this chapter we have discussed several different techniques used for processing TEM images. Different software is suitable for processing different types of image. The problem is that there may not be sufficient financial benefit unless a package supports another product; a similar problem was found for simulation software in Chapter 30. Many programs have been written to run on particular computer systems and support can disappear even though the TEM has a lifetime of 15+ years.

TABLE 31.1. Software for Image Processing

Package	Supplier	Cost
Digital Micrograph	Gatan	Commercial
NIH Image	NIH	Freeware

31.17 THE OPTICAL BENCH—A LITTLE HISTORY

Although not widely used now, the optical bench is a useful instructional tool but you need to be old fashioned and record the image on film! A typical experimental set-up is shown in Figure 31.26. The laser provides a coherent source of illumination representing the electron beam. The negative acts as the specimen. If it contains a set of lattice fringes, these act as a diffraction grating and give rise to a row of spots on the screen placed at the back focal plane of the 'objective' lens. The lens is thus performing an optical Fourier transform of the photograph. If you 'move the screen' to the image plane, the fringes reappear. You can make different masks and place them at the back focal plane or even create an 'adaptive filter' by exposing a photographic film and using this as the template for your mask. These masks correspond to the objective aperture in TEM. Students will find it instructive to transform their instructor or another suitable photograph, examine the frequency spectrum, and investigate the resulting spatial effect of different masks. The detail in the image is quickly lost as you remove the high spatial frequencies. This corresponds to inserting a smaller aperture in the back focal plane of the objective lens, as illustrated in Figure 31.27. So Figure 31.27D is effectively a BF image: clearly, you discard a lot of information in such images!

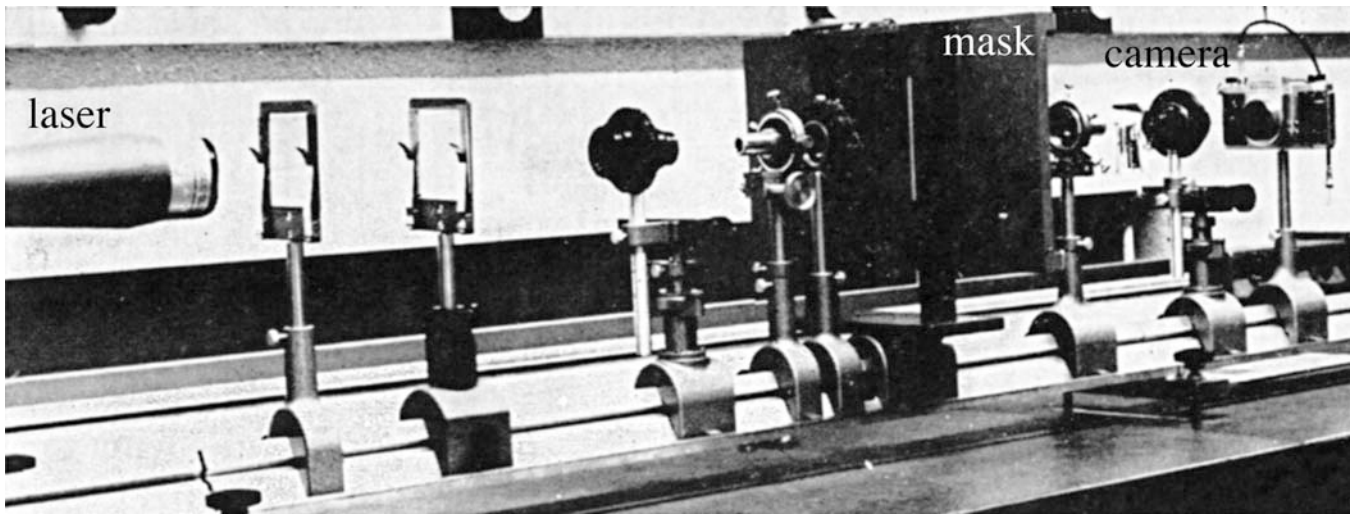


FIGURE 31.26. A typical experimental setup for an optical bench with the mask in the back focal plane.

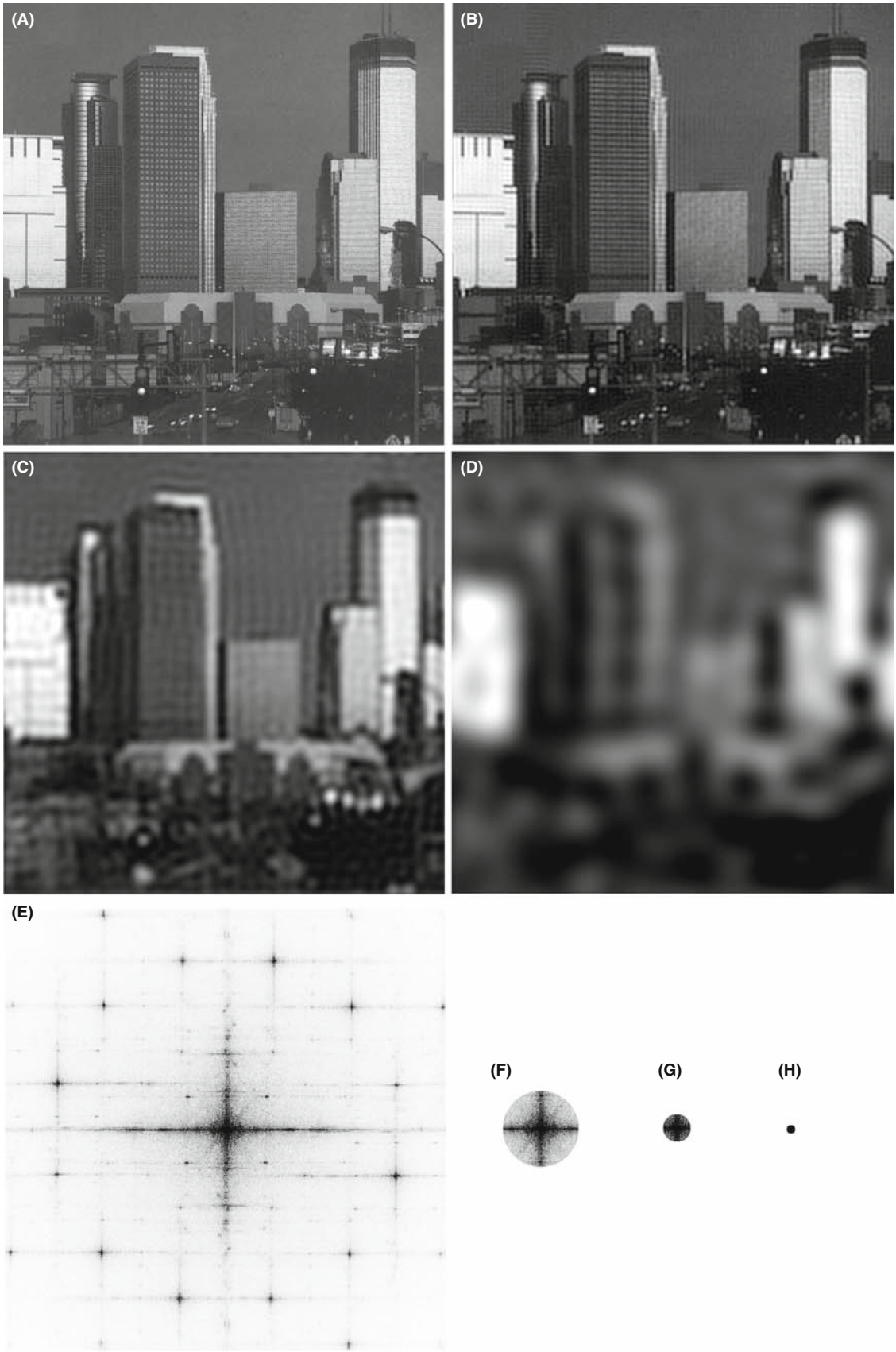


FIGURE 31.27. The effect of a mask (aperture) size on a non-periodic image of the Minneapolis skyline. (A–D) Reducing the aperture size, as indicated in the corresponding optical transform (diffractograms) (E–H) reduces the image detail. The streaks in (E) arise from the edges of the photographs.

CHAPTER SUMMARY

We have been doing image processing for many years; it's called 'dodging' in the photographic darkroom. You can even do this automatically with a special commercial enlarger. However, we have done very little quantitative imaging in materials science and even less in nanotechnology. The points you should remember when starting in the field are the following

- Quantitative comparison of simulated and experimental images depends on both the simulation program and the experimental parameters.
- If you are going to use reciprocal-space techniques for quantitative analysis, you should consider the shape of the optimum mask; it may not be a circular mask especially if you are studying interfaces! The process can, in principle, be assisted by the computer.
- The potential for image restoration is no longer limited by how the signals are mixed due to C_s and Δf. You can unscramble those effects if you have the right equipment.
- You will notice the repeated use of the word 'potential' where we don't mean $V(\mathbf{r})$! In many ways this chapter is a guide to the future of HRTEM and TEM in general. Some of the features won't be commonly available or optimized until the manufacturers realize their importance to the user and the user realizes that stability (for example) costs money.

Ultimately, the resolution limit is set by the signal-to-noise ratio in your image.

When you are processing images, there is always the possibility of removing information which is important. For example, Fresnel fringes often *should* be there! Beware of making reality match simulation, rather than the reverse. In the same vein, we draw your attention to the conclusion of Hÿtch and Stobbs, who found that they could only match their experimental and simulated images if they used a value for the specimen thickness which they knew was wrong! Their study emphasizes that, wherever possible, you should obtain independent measurements of the characteristics of your specimen and your machine. Remember the double-headed rhino in Figure 1.7; don't publish artifacts, even well-processed ones. Even if it's just for entertainment, images can reappear in unexpected places. Remember that an early image of Si dumbbells was later interpreted correctly because the authors gave the necessary data on how the image was formed when they published it. You should *always* do likewise.

PROCESSING IMAGES

Harrel, B *et al.* 1995 *Using Photoshop for Macintosh* Que Indianapolis. A detailed description of the possibilities using Photoshop; this volume should also serve as a warning for all microscopists. Ask the question: has the image you are examining been processed, and if so, how?

Hawkes, PW, Ed. 1980 *Computer Processing of Electron Microscope Images*, Springer-Verlag New York. Despite its age, this collection of articles is for advanced students.

ANALYZING C_S, Δf, ETC.

Krivanek, OL 1976 *Method for Determining the Coefficient of Spherical Aberration from a Single Electron Micrograph* Optik **45**, 97–101. An early paper giving a method for analyzing C_s and Δf using a diffractogram.

Zemlin, F, Weiss, K, Schiske, P, Kunath, W and Herrmann, K-H 1978 *Come-Free Alignment of High Resolution Electron Microscopes with the Aid of Optical Diffractograms* Ultramicrosc. **3** 49. The Zemlin tableau paper.

FOCAL SERIES AND TILT SERIES

Kirkland, EJ, and Siegel, BM, Uyeda, N and Fujiyoshi, Y 1982 *Nonlinear high resolution image processing of conventional transmission electron micrography: II. Experiment* Ultramicrosc. **9**, 65–74.

Kirkland, AI, Saxton, WO, Chau, K-L, Tsuno, K and Kawasaki, M 1995 *Super-resolution by aperture synthesis: tilt series reconstruction in CTEM* Ultramicrosc. **57**, 355–374.

Kirkland, AI, Meyer, RR and Chang L-YS 2006 *Local Measurement and Computational Refinement of Aberrations for HR TEM* Microsc. Microanal. **12** 461–468.

Kirkland, AI and Meyer, RR 2004. *'Indirect' High-Resolution Transmission Electron Microscopy: Aberration Measurement and Wavefunction Reconstruction* Microsc. Microanal. **10** 401–413.

ADVANCED TEM PROCESSING

Dorset, DL 1995 *Structural Electron Crystallography* Plenum New York. An early example of 3D reconstruction.

Downing, KH 1992 Scanning Microscopy Supplement **6** AMF O'Hare IL p. 405.

Hawkes, PW, Ed. 1992 *Signal and Image Processing in Microscopy and Microanalysis* Scanning Microscopy Supplement **6**, AMF O'Hare IL. This volume and its 1988 and 1996 companions are required reading for this subject.

Hawkes, PW, Ottensmeyer, FP, Rosenfeld, A and Saxton, WO (Eds) 1988 *Image and Signal Processing for Electron Microscopy* Scanning Microscopy Supplement **2** AMF O'Hare IL.

Hawkes, PW, Saxton, WO and Frank, J Eds. 1996 *Image Processing* Scanning Microscopy Supplement X AMF O'Hare IL.

Russ, J.C. (1990) *Computer-Assisted Microscopy* Plenum Press New York. Chapter 3 is particularly relevant. Subsequent chapters give detailed analyses of the topics introduced here.

Russ, JC 1995 *The Image Processing Handbook* 2nd edition CRC Press Boca Raton. A beautiful, comprehensive, and an essential component of any EM lab (or home).

Saxton, WO 1992 Scanning Microscopy Supplement **6**, AMF O'Hare IL p. 405. Correcting for distortion in the specimen (Section 31.6**B**).

Trus, BL, Unser, M, Pun, T and Stevens, AC 1992 Scanning Microscopy Supplement **6**, AMF O'Hare IL p. 441.

CHEMISTRY FROM IMAGES (SECTION 31.10)

Kisielowski, C, Schwander, P, Baumann, FH, Seibt, M, Kim, Y, and Ourmazd, A 1995 *An Approach to Quantitative High-Resolution Transmission Electron Microscopy of Crystalline Materials* Ultramicrosc. **58**, 131–155.

Ourmazd, A Baumann, FH, Bode, M and Kim, 1990 *Quantitative Chemical Lattice Imaging: Theory and Practice* Ultramicrosc. **34**, 237–255.

PATTERN RECOGNITION AND FILTERING ON INTERFACES

King, WE and Campbell, GH 1993 *Determination of Thickness and Defocus by Quantitative Comparison of Experimental and Simulated High-Resolution Images* Ultramicrosc. **51**, 128–135.

King, WE and Campbell, GH 1994 *Quantitative HREM Using Non-Linear Least-Squares Methods* Ultramicrosc. **56**, 46–53.

Möbus, G, Necker, G and Rühle, M 1993 *Adaptive Fourier-Filtering Technique for Quantitative Evaluation of High-Resolution Electron Micrographs of Interfaces* Ultramicrosc. **49**, 46–65. The adaptive Fourier-filtering technique.

Paciornik, S, Kilaas, R, Turner, J, and Dahmen, U 1995 *A Pattern Recognition Technique for the Analysis of Grain Boundary Structure by HREM* Ultramicrosc. **62**, 15–27. Pattern recognition at interfaces.

SPECIFICS

Erickson, HP and Klug, A 1971 *Measurement and Compensation of Defocusing and Aaberrations by Fourier Processing of Electron Micrographs* Phil. Trans. Roy. Soc. London B. **261**, 105–118. An example from the Nobel Prize winner.

Frank, J 1980 in *Computer Processing of Electron Microscope Images*, Springer-Verlag New York. p. 187. Correlation techniques (Section 31.12).

Hÿtch, MJ and Stobbs, WM 1994 *Quantitative Comparison of High Resolution TEM images with image simulations* Ultramicrosc. **53** 191–203. A very interesting conclusion on the thickness of the specimen.

Koster, AJ, van den Bos, A and van der Mast, KD 1988 Scanning Microscopy Supplement **2** AMF O'Hare IL p. 83. Autoalignment of the TEM.

Koster, AJ and de Juijter, WJ 1992 *Practical Autoalignment of Transmission Electron Microscopes* Ultramicrosc. **40** 89–107. Autoalignment of the TEM.

Krivanek, OL and Mooney, PE 1993 *Applications of Slow-Scan CCD Cameras in Transmission Electron Microscopy* Ultramicrosc. **49** 95–108. Autoalignment of the TEM.

Moré, JJ 1977, in *Lectures Notes in Mathematics*, Ed. G.A. Watson p. 630 Springer Berlin.

Moré, JJ, Garbow, BS and Hillstrom, KE 1980 *User Guide for MINIPACK-1.*

Saxton, WO and Koch, TL 1982 *Interactive Image Processing with an Off-Line Minicomputer: Organization, Performance and Applications* J. Microsc. **127** 69–83. Autoalignment of the TEM.

Saxton, WO, Smith, DJ and Erasmus, SJ 1983 *Procedures for Focusing, Stigmating and Alignment in High Resolution Electron Microscopy* J. Microsc. **130** 187–201. Cross-correlating pairs of images.

Tang, D, Jansen, J, Zandbergen, HW and Schenk, H 1995 *The Estimation of Crystal Thickness and the Restoration of Structure-Factor Modulus from Electron Diffraction: A Kinematical Approach* Acta Cryst. **A51** 188–197. (31.6E) Quantitative analysis of DPs.

van Heel, M, Winkler, H, Orlora, E and Schatz, M 1992 Scanning Microscopy Supplement **6**, AMF O'Hare IL p. 23. Review of TEM of beam-sensitive materials.

Zou, XD 1995 *Electron Crystallography of Inorganic Structures, Chemical Communications* Stockholm University, Stockholm, Sweden.

THE COMPANION TEXT

The topic of image processing is often associated only with HRTEM images. This aspect is the subject of an entire chapter on focal-series reconstruction in the companion text.

One point you must always keep in mind is that the data are what you recorded originally. You will usually modify these data in some way before it is reported to the scientific community. The question you must always keep in mind is how much detail on the processing must you report so that other scientists will be able to check and repeat your analysis.

SELF-ASSESSMENT QUESTIONS

Q31.1 Define the term 'quantifying HRTEM images.'
Q31.2 Why is Fourier filtering so named?
Q31.3 Sketch a pair of diffractograms from an amorphous Ge film with $\Delta f = 1$ sch and 4 sch, respectively.
Q31.4 Why is amorphous Ge used rather than amorphous C to support Au nanoparticles in testing the resolution of an HRTEM?
Q31.5 If we plot $\frac{n}{u^2}$ against u^2, what is the slope of the curve and what is n?
Q31.6 What is unsharp masking?
Q31.7 How would you obtain the best TEM still image from a TEM video?
Q31.8 What is a kernel in image processing?
Q31.9 Give an example of a kernel for enhancing the contrast in image processing.
Q31.10 Why do we say that the signal-to-noise ratio will always be small in low-dose TEM?
Q31.11 Why is it helpful to record a focal series of images if you want to do quantitative HRTEM?
Q31.12 Will Kikuchi diffraction influence electron crystallography using DPs?
Q31.13 What is the reason for using a tilted-beam series to enhance resolution in HRTEM?
Q31.14 What is the most important use of remote microscopy?
Q31.15 We use a diffractogram tableau to bring the beam alignment to within 0.1 mrad of the optic axis. What aberration are we correcting?
Q31.16 What is the use of pattern recognition in HRTEM and why would you like to have a new 4 k $\times$ 4 k camera for this purpose?
Q31.17 What is the basic principle behind QUANTITEM?
Q31.18 There are two methods for estimating the 'goodness of fit'? What are they and how do they differ?
Q31.19 For the best quantitative data, you should use a focal series and low-dose imaging. What fundamental principle of quantum physics does this statement remind you of?
Q31.20 What is the adaptive Fourier-filtering technique?

TEXT-SPECIFIC QUESTIONS

Some of the questions for this chapter require that you have access to image-processing software.

T31.1 Identify another particle/matrix pair that might be studied like that in Figures 31.2 and 31.3. Describe how you would use image processing in your analysis to show it is a spherical particle, not a cylinder.
T31.2 Describe how you would extract the best image from a 30-second video clip of a stable grain boundary. (Hint: use several sections in this chapter.)
T31.3 Use your scanner and Digital Micrograph to check the correlation between the images and diffractograms in Figures 31.4 and 31.5. You can consider the original publications. Comment on your results and the scientific process.
T31.4 By considering the original paper or just from scratch, reproduce the graphs shown in Figure 31.6A.
T31.5 Explain how you would actually use the kernel for a digital Laplacian in analyzing a TEM image. What type of image might this be useful for? Find (in the literature) and describe two other kernels.
T31.6 Low-dose microscopy should be used much more routinely. Describe in detail the routines that are actually used by two different cryo-TEM manufacturers.
T31.7 Consider Figure 30.10. Summarize how you would carry out this type of study in your lab. Pay special attention to the important features of the specimen (e.g., is it applicable to all specimens).

T31.8 By delving into the literature, write a description of how you would use a program like QUANTITEM. Identify other groups who have used similar programs. Give a critical review of the applications of the concept of parameterization in TEM (with special attention to the range of materials that can be studied).

T31.9 Use a laser pointer and a small-mesh TEM grid to produce a diffraction pattern of the grid on a nearby wall. (See Figure 31.27.) Relate the spacing of the spots to the distances (e.g., mesh spacing) that you used.

T31.10 Use Digital Micrograph, or a similar product, to reproduce the images shown in Figure 31.27 using your home town or a face as the object. Comment on the reduction of resolution as you decrease the aperture diameter.

Part 4

Spectrometry

32

X-ray Spectrometry

CHAPTER PREVIEW

To make use of the X-rays generated when the beam strikes the specimen, we have to detect them and identify from which element they originated. This is accomplished by X-ray spectrometry, which is one way to transform the TEM into a far more powerful instrument, called an analytical electron microscope (AEM). Currently, the only commercial spectrometer that we use on the TEM is an X-ray energy-dispersive spectrometer (XEDS), which uses a Si semiconductor detector or sometimes a Ge detector. New detector technologies are emerging, which we'll describe briefly. While some of these may render the Si detector obsolete, we'll nevertheless emphasize this particular detector.

The XEDS is a sophisticated instrument that utilizes the fast processing speeds made possible by modern semiconductors. The detector generates voltage pulses that are proportional to the X-ray energy. Electronic processing of the pulses translates the X-ray energy into a signal in a specific channel in a computer-controlled storage system. The counts in the energy channels are then displayed as a spectrum or, more usefully, transformed into a quantitative compositional profile or, better still, a compositional image or 'map.'

COUNTS
We'll see over and over again that maximizing the number of X-ray counts is paramount.

The Si detector is compact enough to fit in the confined region of the TEM stage and, in one form or another, is sensitive enough to detect all the elements above Li in the periodic table. We'll start with the basic physics you need to understand how the detectors work and give you a brief overview of the processing electronics. We then describe a few simple tests you can perform to confirm that your XEDS is working correctly and the choices you have to make due to the way the XEDS is interfaced to the AEM column.

It is really most important from a practical point of view that you know the limitations of your XEDS and understand the spectrum. Therefore, we'll describe these limitations in detail, especially the unavoidable artifacts (Chapter 33). In Chapter 34, we'll show how the spectra can easily give a qualitative elemental analysis of any chosen feature in your image and, in just a little more time (Chapter 35), a full quantitative analysis. In Chapter 36, we'll show that this information can be obtained with a spatial resolution approaching a nanometer or below and offers detection limits close to a single atom. So '*micro*analysis' is not a good term; 'nanoanalysis' is more accurate but sounds worse. 'Analysis' is how we'll describe it.

32.1 X-RAY ANALYSIS: WHY BOTHER?

The limitations of only using TEM imaging should, by now, be obvious to you. Our eyes are accustomed to the interpretation of 3D, reflected-light images. However, as we have seen in great detail in Part 3, the TEM gives 2D projected images of thin 3D specimens and you, the operator, need substantial experience to interpret these images correctly. For example, Figure 32.1 shows six images, taken with light and electron microscopes (can you distinguish which images are from which kind of microscope?). The scale of the images varies over 6 orders of magnitude from nanometers to millimeters and yet they all appear similar. Without any prior knowledge it would not be possible, even for an experienced microscopist, to identify the nature of these specimens simply from the images.

Now if you look at Figure 32.2, you can see six X-ray spectra, one from each of the specimens in Figure 32.1. The spectra are plots of X-ray *counts* (imprecisely

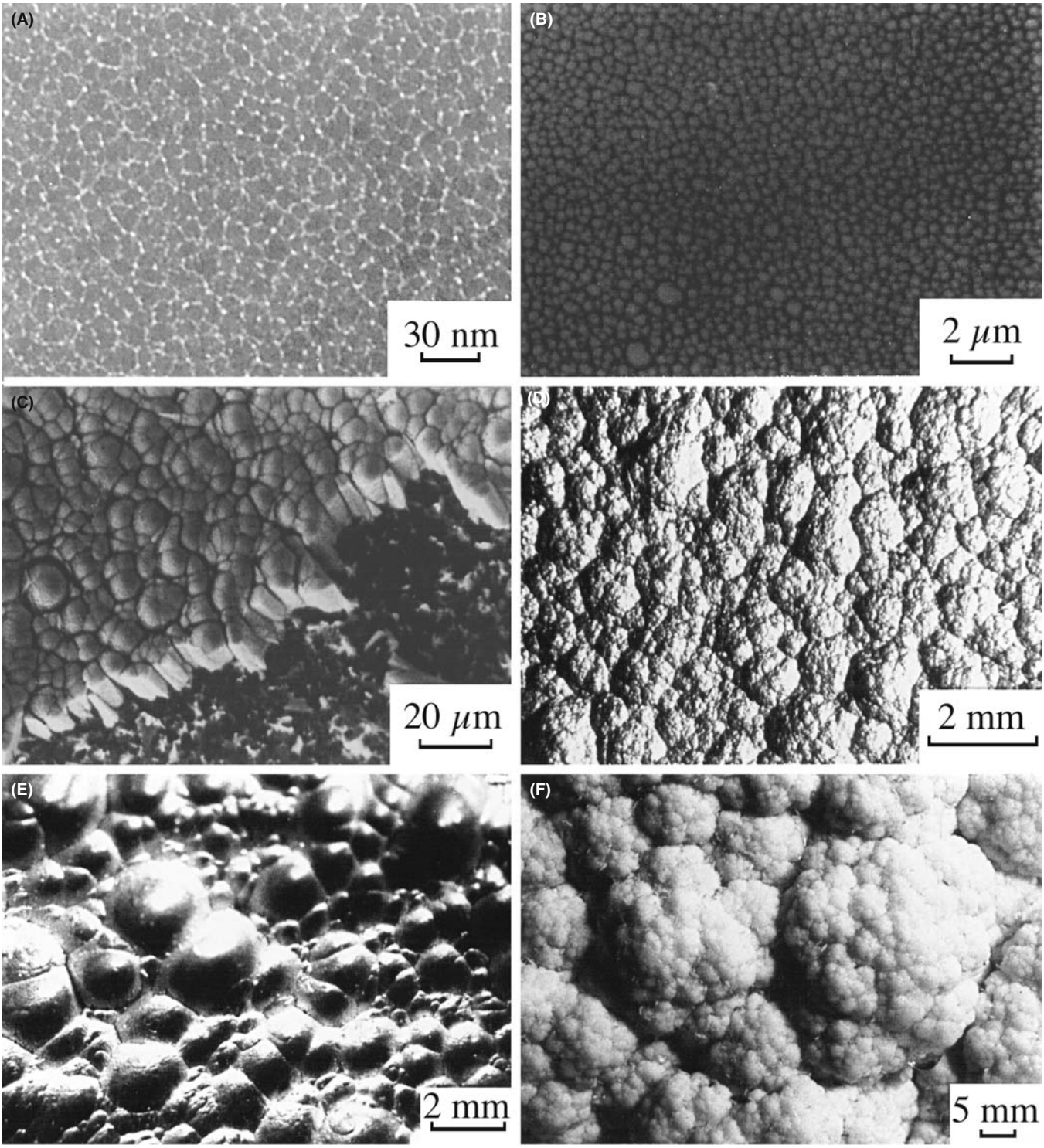


FIGURE 32.1. Six images of various specimens, spanning the dimensional range from nanometers to millimeters. The images were taken with TEMs, SEMs, and light microscopes, but the characteristic structures are very similar, and it is not possible, without prior knowledge, to identify the specimens.

termed 'intensity') versus X-ray *energy* and basically consist of Gaussian-shaped peaks on a slowly changing background. From Chapter 4 you already know that the peaks are characteristic of the elements in the specimen and the background is also called the bremsstrahlung. But even with no knowledge of XEDS, you can easily see that each specimen gives a different spectrum.

Different characteristic peaks mean different elemental constituents; it is possible to obtain this information in a matter of minutes or even seconds.

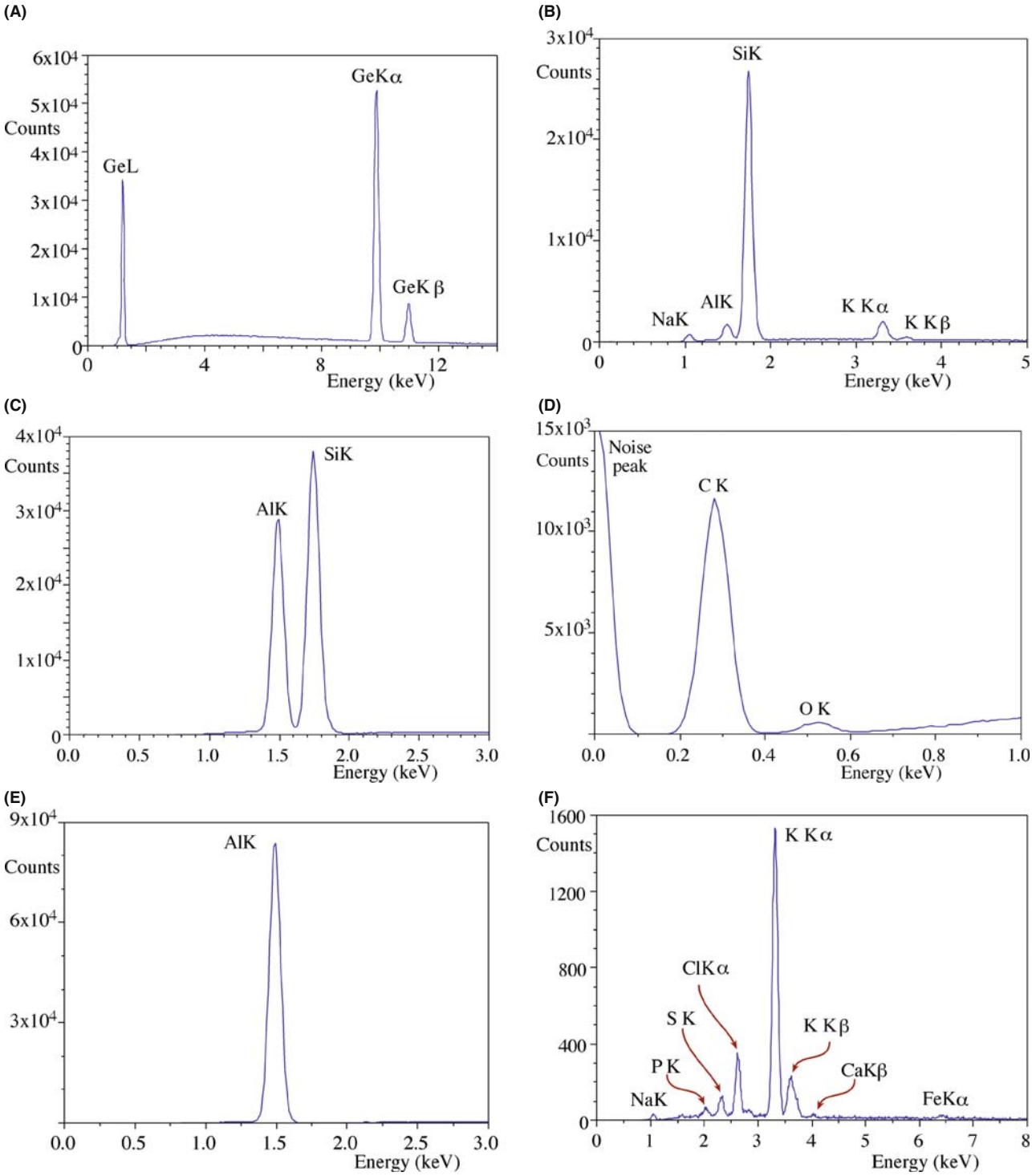


FIGURE 32.2. XEDS spectra from the six specimens in Figure 32.1. Each spectrum is clearly different from the others, and helps to identify the specimens as (A) pure Ge, (B) silica glass, (C) Al evaporated on a Si substrate, (D) pyrolitic graphite, (E) pure Al, and (F) a cauliflower.

When you have such elemental information, any subsequent image and/or diffraction analysis is greatly facilitated. For your interest, the identity of each specimen is given in the caption to Figure 32.2. While Figure 32.2A–E is from common inorganic materials, Figure 32.2F is from a cauliflower which, once you get it into the electron microscope, provides a very distinctive spectrum, albeit from a somewhat carbonized relic of the original vegetable. The familiar morphology of this specimen, now obvious in Figure 32.1F, also accounts for the generic term 'cauliflower structure' which is given to these and similar microstructures.

The main message you should get from this illustration is that the combination of imaging and spectroscopy transforms a TEM into the much more powerful AEM.

32.2 BASIC OPERATIONAL MODE

To produce spectra such as those in Figure 32.2, you first obtain a TEM or STEM image of the area you wish to analyze. In TEM mode, you then have to condense the beam to an appropriate size for analysis. This means exciting the C1 lens more strongly, decreasing the C2 aperture size and adjusting the C2 lens strength. These steps will misalign the illumination system and it can be tedious to move between TEM-image and focused-spot analysis modes, unless you are driving a fully computer-controlled (S)TEM. So, we recommend that you operate in STEM mode. First, create your STEM image as we described back in Section 9.4. Then simply stop the scanning probe and position it on the feature you wish to analyze and switch on the XEDS. In STEM mode, digital software can also check for specimen drift during your analysis.

STEM MODE

Use STEM for AEM. It makes it easier to change from image to analysis mode, easy to form compositional images, and easier to compensate for drift.

In this 'spot' mode you can simply move the beam around the specimen and get a sense of the elemental chemistry of different features you select. However, this approach is very limited from a statistical sampling standpoint and highly biased toward what you think looks interesting in the image. It is now feasible to gather not merely a spectrum from a feature in your specimen, as in Figure 32.2, but a spectrum at *every pixel* in a digital STEM image. From such 'spectrum images' we can extract maps showing the distribution of each element in the specimen and its relationship to the features in the electron image, thus adding another dimension (literally) to the power of the AEM (go back and check the X-ray map in Figure 1.4). We'll talk more about this in Chapter 33 and discuss both qualitative and quantitative maps in Chapters 34 and 35.

For reasons that we'll describe in detail later, you should *always* perform XEDS with your specimen in a low-background (Be) holder. Unless you have an UHV AEM, the holder should be cooled to liquid-N_2 temperature to minimize contamination, and we recommend a double-tilt version, so you can simultaneously carry out diffraction and/or imaging along with your analysis.

32.3 THE ENERGY-DISPERSIVE SPECTROMETER

The XEDS was developed in the late 1960s and by the mid-1970s was an option on many TEMs and even more widespread on the SEM. This rapid spread testifies to the fact that the XEDS is really quite a remarkable instrument, embodying many of the most advanced features of semiconductor technology. It is compact, stable, robust, easy to use, and you can quickly interpret the readout. Several books have been devoted to XEDS on electron-beam instruments and these are listed in the general references. Figure 32.3A shows a schematic diagram of the complete XEDS system and we'll deal with each of the major components as we go through this chapter.

The computer controls all three parts. First, it controls whether the detector is on or off. Ideally, we only want to process one incoming X-ray photon at one time. So the detector is switched off when an X-ray photon is detected and switched on again after that signal is processed (notice we use the particle description of an X-ray here; other detectors work in ways that assume the X-ray is a wave). Second, the computer controls the processing electronics, assigning the signal to the correct energy channel in the storage system. Third, the computer calibrates the spectrum display and tells you the conditions under which you acquired the spectrum, the peak identity, the number of X-rays in a specific channel

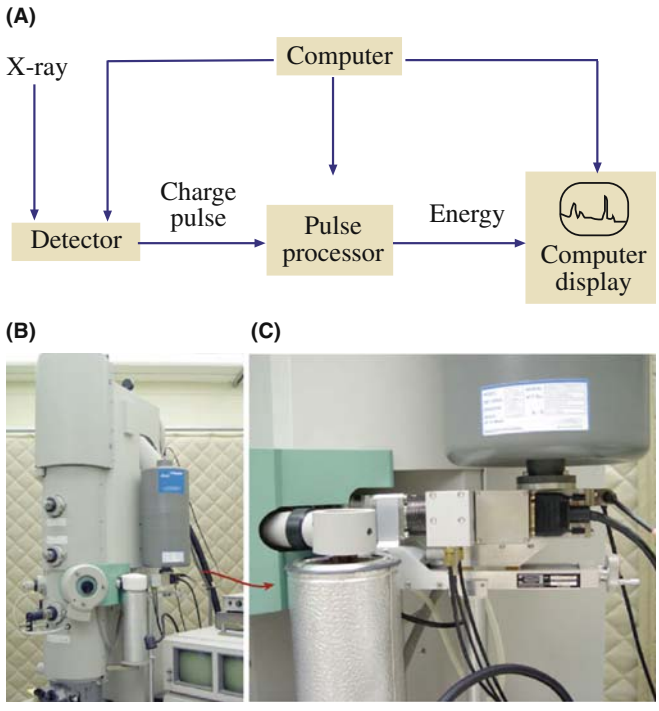


FIGURE 32.3. (A) Schematic diagram of the principle of XEDS; the computer controls the detector, the processing electronics and the display. (B) An XEDS system interfaced to the stage of an AEM. Even in close-up (inset), all that is visible is the large liquid-N_2 dewar attached to the side of the column.

or 'window' of several channels, etc. Any subsequent data processing is also carried out using the computer.

3 COMPONENTS

The three main parts of an XEDS system are
(i) the detector
(ii) the processing electronics
(iii) the computer

We can summarize the working of the XEDS as follows

- The detector generates a charge pulse proportional to the X-ray energy.
- This pulse is first converted to a voltage.
- The voltage is amplified through a field-effect transistor (FET), isolated from other pulses, further amplified, then identified electronically as resulting from an X-ray of specific energy.
- A digitized signal is stored in the channel assigned to that energy in the computer display.

The speed of this process is such that the spectrum appears to be generated in parallel with the full range of X-ray energies detected simultaneously, but the process actually involves very rapid serial processing of individual X-ray signals. Thus, the XEDS both detects X-rays and separates (*disperses*) them into a spectrum according to their *energy*; hence the name of the spectrometer.

Figure 32.3B shows an XEDS interfaced to an AEM. In fact, you can't see the processing electronics, the display, or even the detector itself because it sits close to the specimen within the column. The only feature that you can see is the dewar containing liquid-N_2 to cool the detector and even this is disappearing from the latest detectors.

32.4 SEMICONDUCTOR DETECTORS

The Si detector in an XEDS is a reverse-biased p-i-n diode and since this is still, by far, the most common detector, we will take this as our model. Later in this section we'll discuss the role of other semiconductor detectors, such as intrinsic-Ge (IG) and Si-drift detectors (SDDs).

32.4.A How Does an XEDS Work?

While you don't need to know precisely how the detector works in order to use it, a basic understanding will help you optimize your system and it will also become obvious why certain experimental procedures and precautions are necessary.

When X-rays deposit energy in a semiconductor, electrons are transferred from the valence band to the conduction band, creating electron-hole pairs, as we saw back in Section 4.4. The energy required for this transfer in Si is ~3.8 eV at liquid-N_2 temperature. (This energy is a statistical quantity, so don't try to link it directly to the band gap.) Since characteristic X-rays typically have energies well above 1 keV, thousands of electron-hole pairs can be generated by a single X-ray. The number of electrons or holes created is directly proportional to the energy of the X-ray photon. Even though all the X-ray energy is not, in fact, converted to electron-hole pairs, enough are created for us to collect sufficient signal to distinguish most elements in the periodic table, with good statistical precision. Figure 32.4 is a schematic

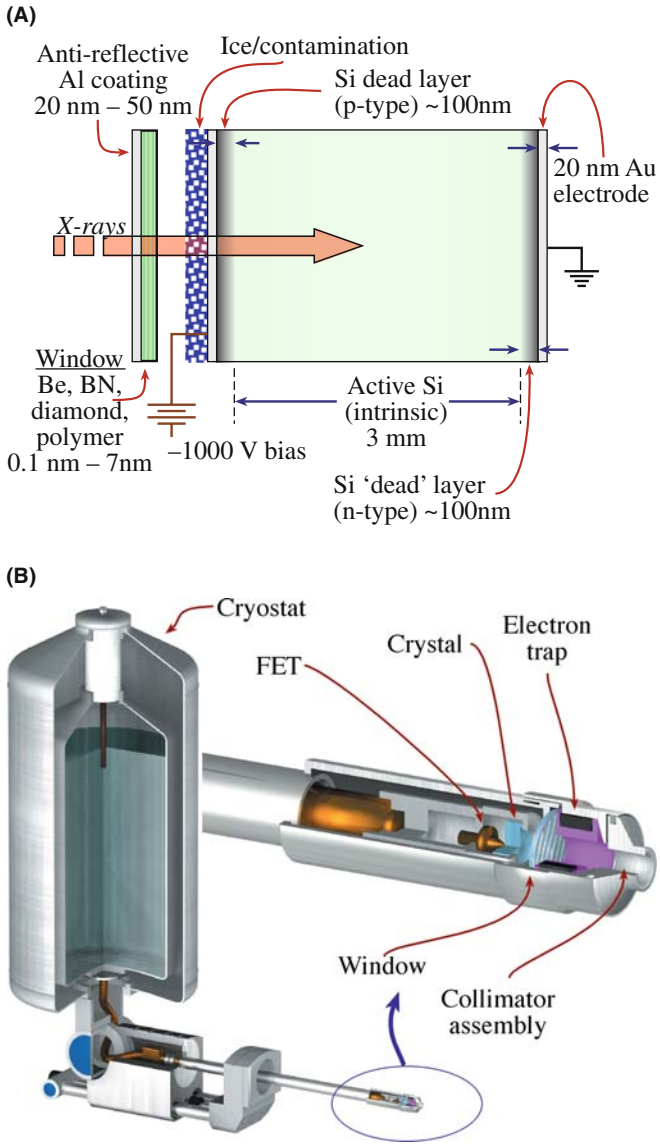


FIGURE 32.4. (A) Cross section of a Si(Li) detector with dimensions indicated (not to scale). In the intrinsic Si region the incoming X-rays generate electron-hole pairs which are separated by an applied bias. A positive bias attracts the electrons to the rear ohmic contact and this charge pulse is amplified by an FET. (B) Exploded diagram of how the individual parts fit together.

diagram of a Si detector and it is similar to the semiconductor electron detectors we discussed back in Chapter 7.

Electron detectors separate the electrons and holes by an internal reverse bias across a very narrow p-n junction, but we need a much thicker detector for X-rays to generate electron-hole pairs since X-rays penetrate matter much more easily than electrons.

Even the purest commercial Si contains acceptor impurities and exhibits p-type behavior. So we compensate for the impurities, which would aid recombination of electron-hole pairs by 'filling' any recombination sites with Li, thus creating intrinsic Si in which the electrons and holes can be separated. Henceforth, we'll refer to Si(Li) (often pronounced "silly") detectors.

The thousands of electrons and holes generated by an X-ray still constitute a very small charge pulse ($\sim 10^{-16}$ C), and so we apply a 0.5–1 keV bias between evaporated Au or Ni ohmic contacts to separate most of the charge. The metal film on the front face creates a p-type region and the back of the crystal is doped to produce n-type Si under a thicker rear contact. So the whole crystal is now a p-i-n device, with shallow junctions on either side of an intrinsic region.

THE ENERGY OF A PULSE

Remember that the magnitude of the charge pulse is proportional to the energy of the X-ray that generated the electron-hole pairs.

When a reverse bias is applied (i.e., a negative charge is placed on the p-type region and a positive charge on the n-type), the electrons and holes are separated and an electron pulse can be measured at the rear contact.

In the p and n regions at either end of the detector, the Li compensation is not completely effective. These regions are effectively unresponsive to the X-ray because most of the electron-hole pairs recombine, and don't contribute to the pulse. These so-called 'dead layers' are an inevitable result of the fabrication process and reduce the detector efficiency. In practice, it is the p-type dead layer at the entrance surface that is most important since the X-rays must traverse it to be detected and we will refer to this as *the* dead layer.

DEAD ACTIVE

The p and n regions are called 'dead layers'; the intrinsic region between them is referred to as the 'active layer.'

The dead layer has become thinner as the detector technology has improved and its effects on the spectrum continue to be reduced (although not to zero, as we shall see).

32.4.B Cool Detectors

Why do we have to cool the detector? Well, if the detector were at room temperature, three highly undesirable effects would occur

- Thermal energy would activate electron-hole pairs, giving a noise level that would swamp the X-ray signals we want to detect.
- The Li atoms would diffuse under the bias, destroying the intrinsic nature of the detector.
- The noise level in the FET would mask signals from low-energy X-rays.

So the detector and the FET are usually cooled with liquid N_2, hence the characteristic dewar shown in Figure 32.3B. The weight of the dewar and the need for constant filling with liquid N_2 are major drawbacks. While the majority of XEDS systems on AEMs still use liquid-N_2 cooled Si(Li) detectors, alternatives are available, such as compact dewars (which use much less N_2), cryo-cooling, compressor-based devices which attain liquid-N_2 temperature mechanically, non-compressive technologies and Peltier-cooled systems, which cool sufficiently (and very rapidly) to deliver reasonable energy resolution with fewer problems. Liquid-N_2 cooling has other drawbacks. Residual hydrocarbons and water vapor in the column form carbon contamination or ice films on the cold detector surface, causing absorption of low-energy X-rays. There are obvious solutions to this problem. We can either isolate the detector from the vacuum, or remove hydrocarbons and water vapor from the column. The latter is a more desirable solution but the former is far easier and much less expensive.

THE XEDS DEWAR

The cylinder hanging on the side of the AEM is the dewar holding liquid N_2 that quietly boils away.

32.4.C Different Kinds of Windows

Liquid-N_2 cooled detectors are usually isolated from the AEM stage in a pre-pumped tube with a sealed 'window' which allows most X-rays through into the detector. There are three kinds of detector; those with a Be window, those with an ultra-thin window and those without a protective window.

Let's examine the pros and cons of each window; a good review has been given by Lund.

Beryllium-window detectors use a thin Be sheet. The best foil is $\sim$7 μm which is transparent to most X-rays, and can withstand atmospheric pressure when the stage

is vented to air. But 7 μm Be is expensive (~$3 M/pound!), rare, and slightly porous, so a thicker sheet (~12–25 μm) is more commonly used. Rolling such a thin Be sheet is a remarkable metallurgical achievement but it still absorbs X-rays with energies < ~1 keV. Therefore, we cannot detect K_α X-rays from elements below about Na ($Z = 11$) in the periodic table, preventing analysis of B, C, N, and O, which are important in the materials, biological, and geological sciences. Other factors such as the low fluorescence yield and increased absorption within the specimen make light-element X-ray analysis somewhat of a challenge, and EELS is often preferable (see Chapters 38 and 39).

KLM TIME

Remind yourself now of the energies involved for K, L, and M for different elements.

Ultra-thin window (UTW) detectors use <100 nm polymer films, diamond, boron nitride, or silicon nitride, all of which can withstand atmospheric pressure while transmitting 192-eV boron K X-rays and the best UTW windows can even analyze Be K X-rays (110 eV). Early polymer UTWs would break if you accidentally vented the column to air without withdrawing the detector and window behind a valve. This problem was overcome by strengthening the polymers with Al films and this accounts for the term 'atmospheric thin window' (ATW) which you may hear. You should remember that different window materials absorb light-element X-rays differently, so you need to know the characteristics of your window. For example, carbon-containing windows absorb nitrogen K_α X-rays very strongly, nitrogen absorbs oxygen, etc.

Windowless detectors only make sense in UHV AEMs such as old VG instruments and Nion dedicated STEMs which minimize hydrocarbons and keep the partial pressure of water vapor by operating with a stage vacuum < ~10^{-8} Pa. Windowless systems routinely detect Be K X-rays as shown in Figure 32.5, which is a remarkable feat of electronics technology.

The relative performance of the various windows is summarized in Figure 32.6. Here we plot the detector efficiency as a function of X-ray energy. You can clearly see the rapid drop in efficiency at low energies and the improved performance of windowless/UTW systems. In fact, Si(Li) detectors absorb (i.e., detect) X-rays with almost 100% efficiency over the range from ~2 to 20 keV, as shown in Figure 32.7. Within this range are X-rays from all the elements in the periodic table above P. This uniform high efficiency is a major advantage of the XEDS detector. Table 32.1 is a concise summary of the pros and cons of each kind of window.

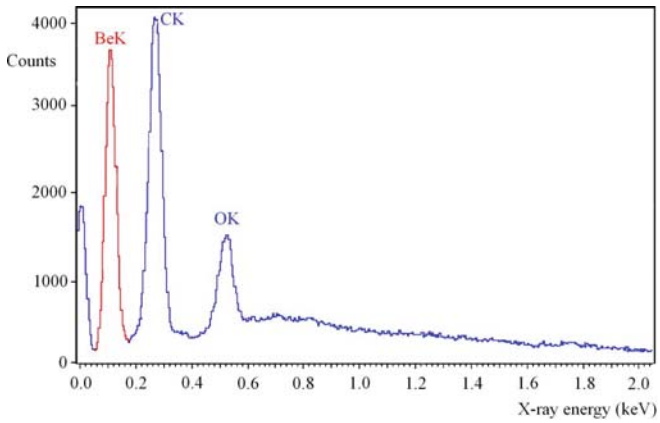


FIGURE 32.5. XEDS spectrum showing the detection of Be in an oxidized Be foil in an SEM at 10 keV. The Be K_α line is not quite resolved from the noise peak.

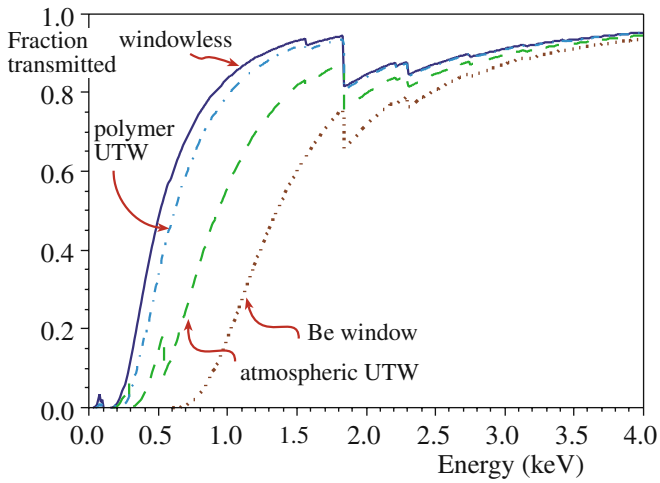


FIGURE 32.6. Low-energy efficiency calculated for a windowless detector, an UTW (1-μm Mylar coated with 20 nm of Al) detector, an ATW detector, and a 13-μm Be-window detector. Note that the efficiency is measured in terms of the percentage of X-rays *transmitted* by the window.

3.8 eV

It takes ~3.8 eV to generate an electron-hole pair in Si, so a Be K_α X-ray will create at most ~29 electron-hole pairs, giving a charge pulse of ~5×10^{-18} C!

32.4.D Intrinsic-Germanium Detectors

You can also see in Figure 32.7 that Si(Li) detectors show a drop in efficiency >~20 keV. This is because such high-energy X-rays can pass through the detector without creating electron-hole pairs. This effect limits the use of Si(Li) in 300–400 keV AEMS in

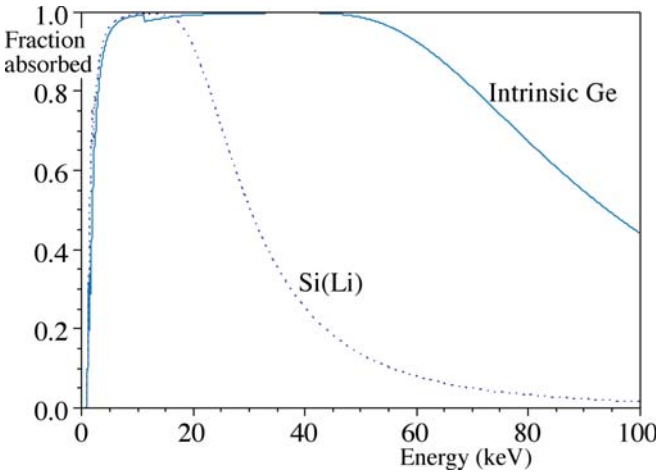


FIGURE 32.7. High-energy efficiency up to 100-keV X-ray energy calculated for Si(Li) and IG detectors, assuming a detector thickness of 3 mm in each case. Note the effect of the Ge absorption edge at about 11 keV. In contrast to Figure 32.6, the efficiency is measured by the percentage of X-rays *absorbed* within the detector.

which we can generate K_α X-rays from all the high-Z elements, e.g., 75 keV Pb K_α X-rays are easily formed at 300 keV. As we'll see in Chapter 35, there are advantages to using the K lines for quantification rather than the lower-energy L or M lines and with a Si(Li) detector, the K lines from elements above Ag ($Z = 47$) are barely detectable. One possible solution is to use an intrinsic Ge (IG) detector which more strongly absorbs high-energy X-rays, as detailed by Sareen.

We can manufacture Ge of higher purity than Si and such Ge is inherently intrinsic, so, Li compensation is not necessary. IGs are more robust and can be warmed up repeatedly which, as we'll see, can solve certain detector problems.

PROTECT YOUR DETECTOR

The intense doses of high-energy electrons or X-rays which can easily be generated in an AEM (e.g., when the beam hits a grid bar) can destroy the Li compensation in a Si(Li) detector, but there is no such problem in an IG crystal.

Furthermore, the intrinsic region can easily be made ~5 μm thick, giving 100% detection of Pb K_α X-rays. Figure 32.7 compares the efficiency of Si(Li) and IG detectors up to 100 keV. There is an even more fundamental advantage to IG detectors. Since it takes only ~2.9 eV to create an electron-hole pair in Ge, compared with 3.8 eV in Si, a given X-ray produces more electron-hole pairs in Ge, and so the energy resolution and signal to noise are better. The only drawback is that the ionization cross sections for high-energy K X-ray excitations are very small for 300–400 keV electrons, so the peak intensities are low. So, why aren't we all using IG detectors? Well, there's no good technical answer but the facts that Si(Li) detectors are easier to manufacture and have a long history of dependable operation are sound commercial reasons why IG detectors have never seriously penetrated the market.

32.4.E. Silicon-Drift Detectors

Si-drift detectors (SDD) may eventually displace traditional Si(Li) detectors; although relatively new, they are already the detector of choice in the much larger SEM and XRF markets. The SDD is basically a CCD (go back and look at Figure 7.3) consisting of concentric rings of p-doped Si implanted on a single crystal of n-Si across which a high voltage is applied to pick up the electrons generated as X-rays enter the side opposite the p-doped rings (Figure 32.8A–C). Applying a voltage from inside to outside the detector (rather than front to back as in a Si(Li)) permits collection of the electrons generated in the n-Si with a 4× lower voltage. Because the central anode in the middle of the p-doped rings has a much smaller capacitance than the large anode at the rear face of a Si(Li) detector (Figure 32.3A–C), a very high throughput of counts is possible, peaking at output rates of many hundred of kcps (Figure 32.8D and look ahead to Figure 32.12B)). If we could in fact generate such enormous count rates in an AEM, we could reduce quantification errors (see Chapter 35), increase analytical sensitivity (see Chapter 36), and seriously improve the statistics of X-ray mapping (see Chapters 33 and 35). In addition to a high throughput, the SDD can operate with no cooling or minimal thermoelectric (Peltier) cooling, while maintaining energy resolutions competitive with Si(Li) detectors (see Section 32.8). This is possible because modern Si-processing technology has reduced thermal-electron generation to extremely low values.

TABLE 32.1. Comparison of Windows

Type	Name	Thickness	Material	Advantage	Disadvantage
Be	Beryllium	~7 μm	Be	Robust	Absorption
UTW	Ultra-thin window	300 nm	Polymer	Low absorption	Breaks easily
ATW	Atmospheric thin window	300 nm	Polymer on grid	Low absorption, robust	Less effective area
None	Windowless	0 nm	None	No absorption	Contamination, light transmitted, need UHV

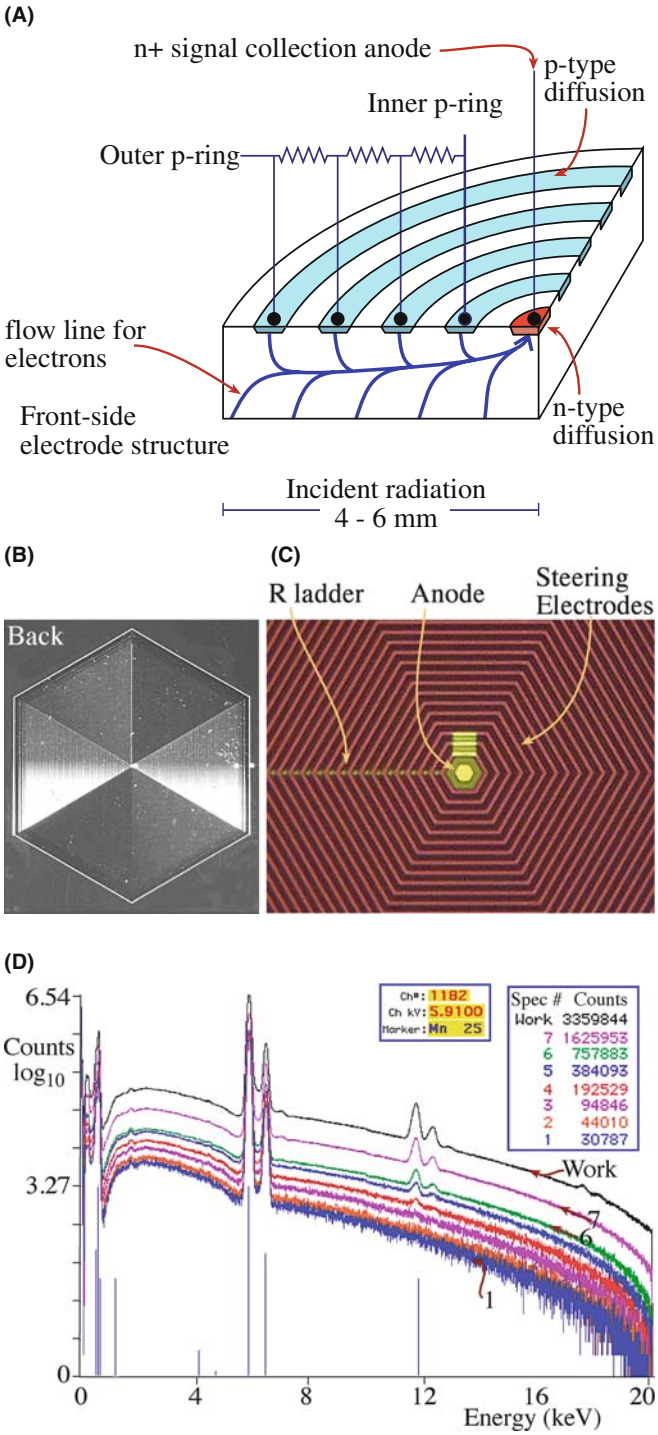


FIGURE 32.8. (A) Schematic diagram showing the back of a quadrant of an SDD consisting of concentric rings of p-type Si on a single crystal of n-type Si. The FET is integrated onto the back of the detector and the bias is applied between the outside p-type ring and the anode inside the inner ring giving electron paths inside the detector as shown in blue. (B) Low-magnification and (C) high-magnification image of the electrode structure on the back of the SDD. (D) SDD spectra from a bulk Mn sample showing no degradation in resolution with increasing output count rates. Be careful: the counts are displayed on a log scale which distorts the usual linear vertical counts scale. The maximum counts are in the Mn Ka peak at 5.91 keV. The black spectrum has over 3.3×10^6 counts in a single channel at that energy. The colored spectra have less counts down to the blue one with 30×10^3 counts in the same channel. But all spectra display the same shape. A Si(Li) spectrometer cannot handle such count rates.

The only problem here is that, as we discuss at length throughout the subsequent chapters, the use of small probes and thin specimens means that total X-ray count rates in AEMs are usually small, negating a principal advantage of SDDs. However, all is not lost and with the advent of intermediate-voltage FEGs and, more recently, C_s correctors, we can create electron probes of < 0.2 nm with > 1 nA of current. If you are prepared to sacrifice spatial resolution by increasing the size of the probe-limiting aperture, then several nA can be generated in probes of a few nm. If you are prepared to sacrifice spatial resolution still further by using thicker specimens, then it is easy to reach the current signal-processing limits (see next section) of Si(Li) electronics (>50 kcps), in which case SDDs might become attractive alternatives. But the jury is still out.

Because an SDD is made up of arrays of individual cells, each one like Figure 32.8A, it is also feasible to consider designing specially shaped SDDs that conform to the inside of the AEM stage. Thus, we could increase the collection angle well beyond typical flat Si(Li) collection angles of a few tens of mrads, and overcome the count-rate limitation (see Section 32.9.A). Count rates are, of course, no problem in SEM-based, bulk-specimen, X-ray microanalysis, which explains the rapid increase of SDDs on those instruments. If you want to learn more, Newbury gives a thorough analysis of the pros and cons of SDDs for mapping in the SEM.

32.5 DETECTORS WITH HIGH-ENERGY RESOLUTION

The poor energy resolution (typically ~135±10 eV) is a fundamental limitation of any of the semiconductor detectors. This limitation arises because, as we discuss in Section 32.7, the detection and processing steps are statistical processes. This poor resolution gives rise to significant peak overlaps (see Section 34.4) and fundamentally limits the sensitivity (detection limits) of analyses (see Section 36.4). However, there are X-ray spectrometers available with significantly better resolution than EDS (<1–10 eV), which may provide better options in the future, so it is worth noting these potential technologies.

32.6 WAVELENGTH-DISPERSIVE SPECTROMETERS

32.6.A Crystal WDS

Before the invention of the XEDS, the wavelength-dispersive (WDS) or crystal spectrometer was widely used. The WDS uses crystals of known interplanar

spacing which (see Part 2) disperse X-rays of wavelength (λ) through different angles (θ) according to Bragg's law ($n\lambda = 2d \sin\theta$). So a WDS treats electrons as waves and this gives considerable advantages over XEDS

- Better energy resolution (~5–10 eV) to minimize peak overlaps.
- Higher peak to background (P/B) ratio to improve detection limits.
- Better detection of light elements (minimum $Z = 4$, Be) by careful choice of analyzing crystal rather than solely through a dependence upon electronics, as in the XEDS.
- No artifacts in the spectrum from the detection and signal processing, except for higher-order lines from fundamental reflections (when $n \geq 2$ in the Bragg equation).
- Higher throughput count rate using a gas-flow proportional counter.

So why don't we have WDS systems on our AEMs? Well, as you'll see in Chapter 35, the forerunner of the AEM, back in the 1960s and early 1970s, was the electron microscope micro-analyzer (EMMA), which did indeed use WDS. However, the WDS was a large and inefficient addition to the TEM, and never attained general acceptance by TEM users for two reasons

- The crystal has to be moved to a precise angle where it collects a tiny fraction of the X-rays from the specimen, whereas the XEDS detector subtends a relatively large solid angle.
- The WDS collects a single λ at any time while the XEDS detects X-rays over a large range of energies. WDS is a very slow, serial collector; XEDS is a fast, effectively parallel, collector.

The geometrical advantage of XEDS (remember, we need to maximize X-ray counts) combined with rapid detection over a wide energy range, without the mechanical motion of the WDS, accounts for the dominance of Si(Li) XEDS systems in all AEMs. Two alternative approaches to overcoming the poor energy resolution of XEDS are being explored. One is a development of the traditional WDS and the other is a totally new approach to X-ray detection.

32.6.B CCD-Based WDS

In an attempt to detect ultra-soft (i.e., very low energy) X-rays, Terauchi and Kawana have designed a WDS using an aberration-corrected, concave, diffraction grating (instead of the usual bent crystal) with a CCD detector in place of the proportional counter. The

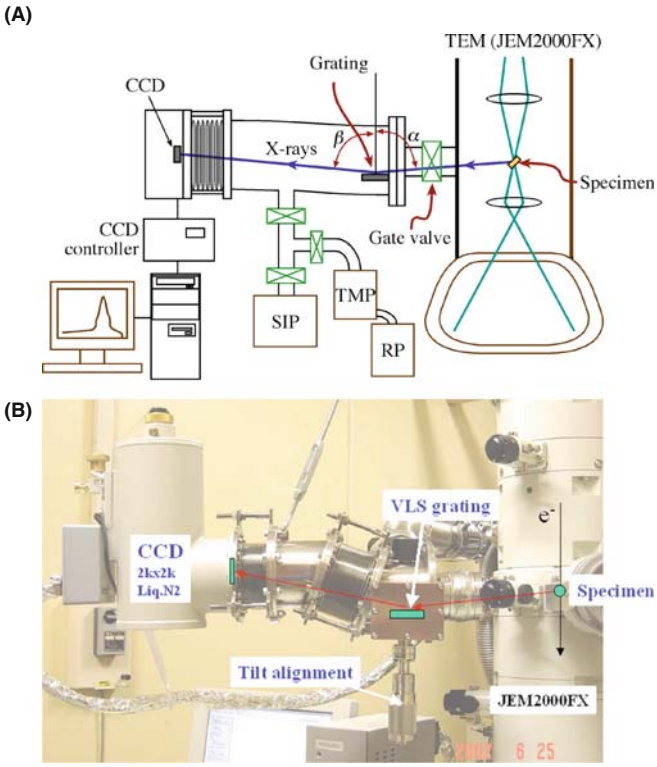


FIGURE 32.9. High-energy resolution X-ray spectrometry; (A) schematic diagram and (B) image of the diffraction-grating WDS system fixed to a TEM column. (C) High-resolution X-ray spectra from hexagonal, cubic, and wurtzite forms of BN showing differences in the B K_α peak shapes due to differences in bonding. (D) Comparison of Si(Li) and bolometer spectra obtained in an SEM illustrating the tremendous difference in energy resolution.

detector is much more compact than a standard WDS, as shown in Figure 32.9A and B, and delivers an energy resolution of 0.6 eV, which is 200× better than a typical EDS resolution and sufficient to resolve intensity variations in the characteristic peaks that arise due to changes in the density of states (DOS) of the valence band (see Figure 32.9C). Study of the DOS and related bonding effects is usually the role of EELS (see Chapter 40), so CCD-WDS is a real breakthrough for X-ray spectrometry. The energy resolution would improve further if the CCD pixel size were decreased, and the grating dispersion increased. Unfortunately, the count rate is very low and to get the spectra shown in Figure 32.9C, a probe size of ~1 μm was used. A larger CCD would permit smaller probes and these will become available via digital-camera technology. Also, recent developments in capillary optics may increase the WDS collection angle, so the technology is worth watching.

32.6.C Bolometers/Microcalorimeters

A totally different approach to detecting X-rays is to measure the heat emitted when an X-ray is absorbed. At first sight this might appear ridiculous, but the

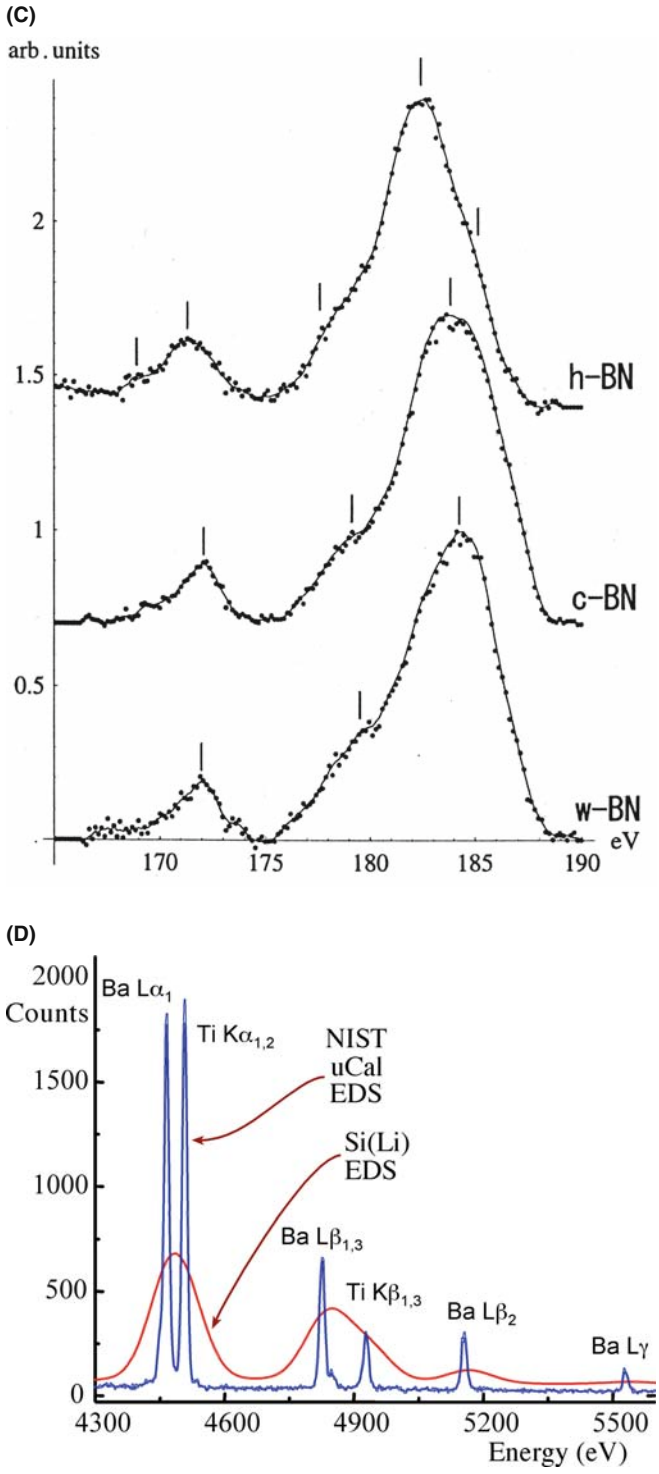


FIGURE 32.9. (Continued).

concept has been demonstrated (Wollman et al.) by creating a microcalorimeter or bolometer (i.e., a sensitive thermometer). As with WDS, the bolometer is limited by its small area, so the solid angle is small and the count rate correspondingly low, but it delivers a resolution comparable with WDS (~5–10 eV), while offering the effectively parallel-collection aspects of XEDS (see Figure 32.9D). Bolometers have been installed on SEMs where their small area is offset by the large count rate. Despite initial enthusiasm, bolometers are not commercially available because they have to be cooled to a few mK with liquid He, which increases the cost close to that of a new SEM and their size to the point where attaching one to a TEM is a major mechanical-engineering feat. The technology to create bolometer arrays is available but until compact, large collection angle WDS or cheap bolometer systems are available, we'll have to live with the poor resolution and other limitations of Si(Li) detectors, to which we'll devote the rest of the chapter.

Table 32.2 below summarizes much of the preceding and following discussion.

32.7 TURNING X-RAYS INTO SPECTRA

The electronics attached to a Si(Li) or SDD convert the charge pulse created by the incoming X-ray into a voltage pulse, which is stored in the appropriate energy channel of the computer display (which used to be called a multi-channel analyzer or MCA). The pulse-processing electronics must maintain good energy resolution across the spectrum without peak shift or distortion, even at high count rates. To accomplish this, all the electronic components beyond the detector crystal must have low-noise characteristics and employ some means of handling pulses that arrive in rapid succession. This whole process used to rely on analog pulse processing, but many of the problems inherent in the analog process have been solved by digital techniques (Mott and Friel), and all current XEDS systems process the charge pulses digitally.

Let's consider first of all what happens if a single, isolated X-ray enters the detector and the electron-hole pairs are separated and captured to create a charge pulse.

- The charge enters the FET preamplifier and is converted into a voltage pulse.
- The pulse is digitized and the X-ray energy that generated the pulse is computed.
- The computer assigns the signal to the appropriate energy channel on the display.

The accumulation of pulses or counts entering each channel at various rates produces a histogram of counts versus energy that is a digital representation of the X-ray spectrum. The computer display offers multiples of 1024 channels in which to display the spectrum and various energy ranges can be

TABLE 32. 2. Comparison of X-ray Spectrometers

Characteristic	IG	Si(Li)	SDD	WDS	Bolometer
Energy resolution (typical/on column)	135 eV	150 eV	140 eV	10 eV	10 eV
Energy resolution (best)	114 eV	128 eV	127 eV	5 eV	5 eV
Energy to form electron-hole pairs (77 K)	2.9 eV	3.8 eV	3.8 eV	n.a.	n.a.
Band gap energy (indirect)	0.67 eV	1.1 eV	1.1 eV	n.a.	n.a.
Cooling required	LN_2 or thermoelectric	LN_2/thermoelectric	None/ thermoelectric	None	100 mK
Detector active area	10–≥50 mm^2	10–≥50 mm^2	≥50 mm^2	n.a.	1 mm^2
Detector arrays available	No	No	Yes	No	Yes
Typical output rates	5–10 kcps	5–20 kcps	1000 kcps	50 kcps	1 kcps
Time to collect full spectrum	~1 min	~1 min	few secs	~30 min	~30 min
Collection angle (sr)	0.03–0.20	0.03–0.30	0.3	10^{-4}–10^{-3}	10^{-4}–10^{-3}
Take-off angle	0°/20°/72°	0°/20°/72°	20°	40°–60°	40°–60°
Artifacts	Escape, sum peaks Ge K/L peaks	Escape, sum peaks Si K peak	Multiple sum peaks	High-order lines	

Data in this table come from the Web sites of the leading XEDS manufactures. For the latest information, check the URLs listed in the reference section.

assigned to these channels. For example, a 10-, 20-, or 40-keV energy range can be used (or even 80 keV for an IG detector on an intermediate-voltage AEM). The display resolution chosen depends on the number of channels available.

A typical energy range that you might select for a Si(Li) and SDD detector is 10 or 20 keV and in 2048 channels this gives you a display resolution of 5 or 10 eV per channel.

DISPLAY RESOLUTION

You should keep the resolution at 10 eV per channel or better. Smaller values use more memory, but memory is cheap. Larger values mean fewer channels for each characteristic peak, giving the peak a serrated step-like appearance rather than a smooth Gaussian shape.

Details of the pulse-processing electronics are not important except for two variables over which you have control. These are the time constant and the dead time. The time constant (τ) is important only if you have an old analog system; it is the time (~5–100 μs) allowed for the analog processor to evaluate the magnitude of the charge pulse.

- Choosing the shortest τ (typically, a few μs) will allow more counts per second (cps) to be processed but with a greater error in the assignment of a specific energy to the pulse, and so the energy resolution (see Section 32.8 below) will be poorer.
- Choosing a longer τ will give you better resolution but the count rate will be lower.

LONG OR SHORT τ?

For most routine thin-foil analyses, you should maximize the count rate (shortest τ), unless there is a specific reason why you want to get the best possible energy resolution (longest τ).

With an analog system you can't have a high count rate and good resolution, so for most routine thin-foil analyses you should maximize the count rate (shortest τ), unless there is a specific reason why you want to get the best possible energy resolution (longest τ). This recommendation is based on a detailed argument presented by Statham.

If you have a digital system, the individual pulses are monitored and τ is varied for each pulse, depending on how close they are together. We call this 'adaptive pulse processing'; it gives a continuous variation of output count rate with input rate rather than a discrete range for each value of τ.

Now, in reality, there are many X-rays entering the detector but, because of the speed of modern electronics, the system can usually discriminate between the arrival of two, almost simultaneous, X-rays. You can find details of the electronics in Goldstein et al. When the electronic circuitry detects the arrival of a pulse, it takes less than a μs before the detector is effectively switched off for a period of time called the dead time, while the pulse processor analyzes that pulse. The dead time is clearly closely related to τ, and is so short that you should expect your XEDS system to process outputs up to 10 kcps quite easily (if it is a later analog system) and 30 kcps or more for a digital system. Even higher outputs (up to 70–100 kcps) need to be handled with an SDD (but in AEMs the beam current and/or specimen thickness is rarely great enough). The dead time increases as more X-rays enter your detector, because it shuts down more often. The dead time can be defined in several ways. If you take the ratio of the output count rate (R_{out}) to the input count rate (R_{in}), which you can usually measure, then

$$\text{Dead time in } \% = \left(1 - \frac{R_{out}}{R_{in}}\right) \times 100\% \qquad (32.1a)$$

An alternative definition is

$$\text{Dead time in } \% = \frac{(\text{clock time} - \text{live time})}{\text{clock time}} \times 100\% \quad (32.1b)$$

These different 'times' can be confusing, but it helps to think of equation 32.1b as follows: if you ask the computer to collect a spectrum for a live time of 100s, then the detector must be live and receiving X-rays for this amount of time. If the detector is actually dead for 20s because it is processing X-rays, it will actually take 120s of 'clock time' to accumulate a spectrum, so the dead time (from equation 32.1b) will be 20/120 = 16.7%. As the input count rate increases, the output count rate will drop and the clock time will increase accordingly. Dead times in excess of 50–60% (or as little as 30% in very old systems) mean that your detector is saturated with X-rays and collection becomes increasingly inefficient. Then you should turn down the beam current or move to a thinner area of the specimen to lower the count rate; but this is a rare situation for a thin-foil analyst to face. Just remember these times.

- Dead time is when the detector is not counting X-rays but processing the previous photon.
- Live time is when the detector is ready to detect an X-ray and not processing any signal.
- Clock time is what it says.

32.8 ENERGY RESOLUTION

The natural line width of the emitted X-rays is only a few eV but the measured widths are usually >> 100 eV. The electronic noise in the XEDS system is a major source of the difference between the practical and theoretical energy resolutions and the width of the electronic noise is described as the 'point-spread function' of the detector. Since the poor energy resolution of XEDS is a major limitation, we need to examine this concept more closely.

We can define the energy resolution R of the detector as

$$R^2 = P^2 + I^2 + X^2 \quad (32.2)$$

where P is a measure of the quality of the processing electronics, defined as the full width at half maximum (FWHM) of a randomized electronic-pulse generator. X is the FWHM-equivalent attributable to detector leakage current and incomplete charge collection (see Section 32.9.A). I is the intrinsic line width of the detector which is controlled by fluctuations in the numbers of electron-hole pairs created by a given X-ray and is given by

$$I = 2.35(F\varepsilon E)^{1/2} \quad (32.3)$$

Here F is the Fano factor of the distribution of X-ray counts from Poisson statistics, ϵ is the energy to create an electron-hole pair in the detector, and E is the energy of the X-ray line. Because of these two factors, the experimental resolution can only be defined under standard analysis conditions defined by the IEEE.

THE IEEE STANDARD FOR *R*

This is the FWHM of the Mn K_α peak, generated (not in the microscope) by an Fe^{55} source which produces 10^3 cps with an 8 μs pulse-processor time constant.

Rather than using IEEE-required radioactive Fe^{55} we recommend measuring the R on your AEM column! Now since Mn is not a common specimen to have lying around, you will find it useful to keep a thin NiO specimen (Egerton and Cheng) to check the resolution when the detector is on the column. You can also use the O K peak to measure the low-energy resolution of your detector. Suitable NiO films < 50 nm thick are available from commercial companies that provide supplies for EM laboratories. Ni is close enough to Mn in the periodic table that you can get good measure of resolution (although it will be slightly worse than that at Mn since resolution degrades with increasing X-ray energy). Others have used thin Cr films instead of NiO. You should be more concerned with changes in R over time than the absolute value since changes indicate that the detector is responding differently to the X-ray flux.

COUNT-RATE EFFECT

All detectors lose resolution as their temperature increases and the count rate increases, although digital electronics handle higher count rates better.

Your computer system will have software that calculates R rather than directly measuring the FWHM of the Mn or Ni peak. It's good to measure the FWHM yourself and you do this by determining the energy width between the channels either side of the peak that contain half the maximum counts in the central (peak) channel, as shown in Figure 32.10.

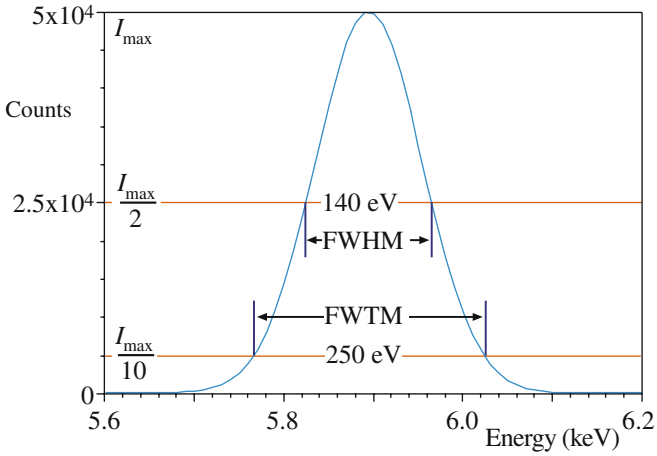


FIGURE 32.10. Measurement of the energy resolution of an XEDS detector by determining the number of channels that encompass the FWHM of the Mn K_α peak. The number of channels multiplied by the eV/channel gives the resolution which, typically, should be about 130–140 eV on the column. You can also measure the FWTM to give an indication of the degree of the ICC which distorts the low-energy side of the peak. The FWTM should be ~1.83× the FWHM if ICC is insignificant.

DETECTOR RESOLUTION

Typically, Si(Li) detectors have a resolution of ~140 eV at Mn K_α with the best being <130 eV. The best reported IG resolution is 114 eV. SDDs offer about 140 eV but can get down to ~130 eV with Peltier cooling.

Because the value of ϵ is lower for Ge (2.9 eV) than for Si (3.8 eV), IG detectors have higher R than Si(Li). The resolution is also a function of the detector area, and the best values we just gave are for 10 mm^2 detectors. The 30 mm^2 or 50 mm^2 detectors, which are typically installed on AEMs to increase the count rate, have resolutions ~5–10 eV worse than the figures just mentioned. So you should be aware that, when you measure R on your AEM, there will be a further degradation in resolution. It is rare to find a 30 mm^2 Si(Li) detector delivering a resolution <~140 eV on the AEM column, even though quoted values are typically ~10 eV less.

How close are XEDS detectors to their theoretical resolution limit? If we assume that there is no leakage and the electronics produce no noise, then $P = X = 0$ in equation 32.2, so $R = I$. For Si, $F = 0.1$, $\varepsilon = 3.8$ eV, and the Mn K_α line occurs at 5.9 keV, which gives R = 111 eV. So it seems that there is not much more room for improvement. The resolution of semiconductor detectors won't approach that of crystal spectrometers or bolometers (< 1 to ~10 eV). However, because of the dependence of I on X-ray energy, light-element K lines do have FWHMs well below 100 eV. We'll see in Section 34.5 that there are signal-processing methods to improve the resolution.

32.9 WHAT YOU SHOULD KNOW ABOUT YOUR XEDS

There are several fundamental parameters which you can specify, measure, and monitor to ensure that your XEDS is performing acceptably. Many of these tests are standard procedures (e.g., see the XEDS laboratories described by Lyman et al.) and are summarized by Zemyan and Williams. In an SEM, which is relatively well behaved, Si(Li) detectors have been known to last 10 years or more before requiring service or replacement. In contrast, an AEM is a hostile environment and the life of a detector can be considerably shortened unless there is a protective shutter (see Section 32.11), which should *always* be closed unless you are acquiring a spectrum. If you analyze a thick portion of your specimen (a waste of time), traverse a grid bar across the field of view (easy to do), or accidentally insert the objective diaphragm while the shutter is open (in which case you should be sentenced to memorize this chapter) you can 'flood' the detector and close down the electronics. If this happens, ask for help, but be prepared to wait a while because the system takes time to recover. If it happens too often, you can permanently damage your detector.

DAMAGING THE DETECTOR

High X-ray or electron fluxes can damage the detector; it is particularly important to monitor the detector performance on your AEM, so that quantitative analyses you make at different times may be compared in a valid manner.

You need to know the operating specifications for your own XEDS, and how to measure them. We can break these specifications down into detector variables and signal-processing variables.

For all of the tests/actions that we describe below, you must discuss the procedure with the instrument technician/laboratory manager **before** doing anything because it is easy to damage the detector if you don't do it right.

32.9.A Detector Characteristics

The *detector resolution* that we just defined may degrade for a variety of reasons. Two are particularly common

- Damage to the intrinsic region by high-energy fluxes of radiation.
- Bubbling in the liquid-N_2 dewar due to ice crystals building up.

We've just told you how to avoid the first problem. If part of your responsibility in the lab is to

top up the liquid N_2 then it should be filtered before putting it into the dewar and *never* re-cycled. If the N_2 is bubbling, you must ask that the problem be corrected and you can suggest warming up the detector.

Or you can suggest that the AEM needs a detector cooled by means other than liquid N_2!

A WARNING ON WARMING

Warming the detector to room temperature should only happen after consultation with the manufacturer, and after turning off the applied bias. (Think what happens to the Li otherwise.)

Incomplete-charge collection (ICC): because of the dead layer, the X-ray peak will not be a perfect Gaussian shape. Usually the peak will have a low-energy tail, because some X-ray energy will be deposited in the dead layer and will not create electron-hole pairs in the intrinsic region. You can measure this ICC effect from the ratio of the full width at tenth maximum (FWTM) to the FWHM of the displayed peak, as shown schematically in Figure 32.10.

In Si(Li) detectors, the phosphorus K_α peak shows the worst ICC effects because this X-ray fluoresces Si very efficiently. ICC will also occur if the Si has a large number of defects arising, for example, from damage via a high flux of backscattered electrons. The crystal defects act as recombination sites, but they can be annealed by warming the detector, as we just described. An IG detector should meet the same FWTM/FWHM criterion as a Si(Li) detector. If the ratio is higher than 2 for the Ni K_α peak there is something seriously wrong with the detector and it needs to be replaced. An SDD has an extremely thin dead layer, so ICC should be minimal.

Detector contamination: over a period of time, even in a UHV STEM, ice and/or hydrocarbons will eventually build up on the detector surface, or the window. When contamination occurs, it reduces the detection efficiency for low-energy X-rays. Ultimately, any detector will contaminate because of residual water vapor in the detector vacuum or because the window may be slightly porous. In all cases, the problem is insidious because the effects develop over time and you might not notice the degradation of your spectrum quality until differences in light-element quantification occur in the same specimen analyzed at different times. Therefore, you should regularly monitor the quality of the low-energy spectrum. The ratio of the Ni K_α/Ni L_α can be used (as described by Michael) as a signal to warn of icing/contamination.

IDEAL GAUSSIAN

An ideal Gaussian shape gives a ratio FWTM/FWHM of 1.82 (Mn K_α or Ni K_α) but this will be larger for lower-energy X-rays that are more strongly absorbed by the detector.

The K/L ratio will differ for different dead layers, different UTWs or ATWs, and different specimen thicknesses, so we can't define an acceptable figure of merit. The best you can do is to measure the ratio (ideally when you first use your detector) and monitor any changes, being aware that, as the ratio increases, quantification of low-energy X-ray lines becomes increasingly unreliable. When the ratio become unacceptable, the detector should be warmed up. Automatic, in situ, heating devices which raise the detector temperature sufficiently to sublime off the ice, without warming the dewar up to ambient temperature, make this process routine. If your detector doesn't have such a device then you should again ask that it be warmed. An SDD doesn't have icing problems if operated at ambient temperatures and, even if Peltier cooling is used, it is still nowhere as cold as liquid N_2.

ICE ON THE DETECTOR

If contamination or ice builds up on the detector, the K_α/L_α ratio rises; the ice selectively absorbs the lower-energy L line.

In summary, you should measure and continually monitor changes in

- *The detector resolution* on the column using the Mn or Ni K_α line (typically 150 eV for Si(Li) and 140 eV for an IG or SDD).
- *The ICC* defined by the FWTM/FWHM ratio of the Ni K_α line (ideally 1.82).
- *The ice/contamination build-up* reflected in the Ni K_α/L_α ratio.

If any of these figures of merit get significantly larger than *your* baseline values, then warming up the detector, if necessary to room temperature, may help.

In summary, you must be very careful with your XEDS.

- DO NOT generate high fluxes of X-rays or backscattered electrons unless your detector is shuttered.
- DO NOT ever warm up the detector yourself; get help. But make sure that the bias is turned off and the manufacturer is consulted.
- DO NOT use unfiltered or re-cycled liquid N_2.

All these DO NOTs make liquid-N_2-free and minimal dead-layer SDDs rather attractive.

32.9.B Processing Variables

You need to ensure that the output counts (i.e., the spectrum) reflect the input X-ray counts. There are three things you can check to make sure the pulse-processing electronics are working properly

- First, check the calibration of the energy range of the spectrum.
- Second, check the dead-time correction circuitry.
- Third, check the maximum output count rate.

The energy resolution should not change significantly from day to day, unless you change the (analog) time constant. Electronic-circuit stability has improved to a level where such checks need only be done a couple of times a year. Will you do the check or rely on someone else?

Calibration of the energy-display range: this process is quite simple; collect a spectrum from a specimen which generates a pair of X-ray lines separated by about the width of the display range (e.g., Cu for 0–10 keV, Mo for 0–20 keV). Some systems use an internal electronic strobe to define zero, and in this case you only need a specimen with a dominant line at the high end of the energy range. Having gathered a spectrum, see if the computer markers are correctly positioned at the peak centroid (e.g., Cu L_α at 0.932 keV and Cu K_α at 8.04 keV). (You'll learn more about the specific energies in Chapter 34 and you can also go back to Chapter 4 to remind yourself of the relationship between E and Z.)

RECALIBRATE
If the peak and the marker are >1 channel (10 eV) apart, then you should recalibrate your display using the commercial software.

Checking the dead-time correction circuit: if the dead-time correction electronics are working properly, the electronics will give an increase in output counts directly proportional to the increase in input counts, for a fixed live time. This behavior is absolutely essential for valid quantification.

- Choose a pure element, say our favorite NiO foil which we know will give a strong K_α peak.
- Choose a live time, say 50s, and a beam current to give a dead-time readout of about 10%.
- Measure the total Ni K_α counts that accumulate in about 30–60s (longer is better).

Then repeat the experiment with higher input count rates (e.g., dead times of 30, 50, 70%).

MUST BE LINEAR
The processing electronics must show linear behavior for valid quantification.

To increase the count rate, increase the beam current by choosing a larger diameter beam or larger C2 aperture. The dead time should increase as the input count rate goes up, but the live time remains fixed (by your choice). If you plot the number of output counts against the beam current, measured with a Faraday cup, or a calibrated exposure-meter reading, then it should be linear, as shown in Figure 32.11. But you will see when you do the experiment that it takes an increasingly longer (clock) time to attain the preset live time as the dead time increases. If you don't have a Faraday cup, you can use the input count rate as a measure of the current; remember that the Faraday cup is useful for many other functions, such as characterizing the performance of the electron source, as we saw in Chapter 5. Remember

Determination of the maximum output count rate: again the procedure is simple

- Gather a spectrum for a fixed clock time, say 10–30s, with a given dead time, say 10%.
- Increase the dead time by increasing the beam current, C2 aperture, or specimen thickness.
- See how many counts accumulate in the Ni K_α peak in the same fixed clock time.

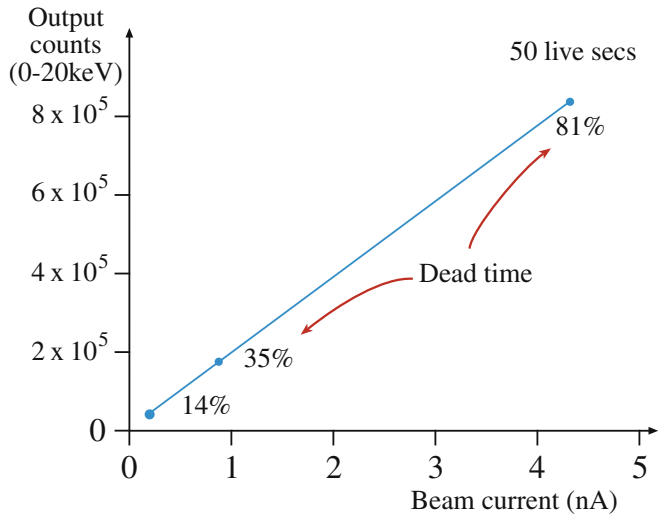


FIGURE 32.11. A plot of the output counts in a fixed live time as a function of increasing beam current showing good linear behavior over a range of dead times, demonstrating that the dead-time correction circuitry is operating correctly.

The number of counts should rise to a maximum and then drop off above a certain dead time, which depends on the system electronics. Beyond the maximum, the detector will be dead more than it is live and so the counts in a given clock time will decrease. In Figure 32.12A, this maximum is at about 60% dead time, typical of modern systems, although in older XEDS units this peak can occur at as little as 30%. If you have analog electronics, you can repeat this experiment for different time constants, τ, and the counts should increase as τ is lowered (at the expense of energy resolution), as also shown in Figure 32.12A. Clearly, if you operate at the maximum in such a curve (if you can generate enough input counts) then you will be getting the maximum possible counts in the shortest possible time. As we've already said, it is almost always better to have more counts than to have the best energy resolution, so select the shortest τ unless you have a peak overlap problem. Digital processing is easier than analog for this reason.

If your specimen is too thin it might not be possible to generate sufficient X-ray counts to reach dead times in excess of 50%, so the curve may not reach a maximum, particularly if τ is very short or you have digital processing. In this case, just use a thicker specimen.

While it is rare that a good thin foil produces enough X-rays to overload the detector electronics, there are situations (e.g., maximizing analytical sensitivity—see Chapter 36) when you need to generate as many counts as possible. However, if you use a thick specimen and high beam current, you may produce too many counts for analog processing, or even digital, systems. If high count rate is your primary mode of operation, you might want to consider an SDD. Figure 32.12B shows the prospects for high output count rates with an SDD. The problem in TEM, as we've noted, is to generate sufficient counts to make use of this extraordinary processing capability.

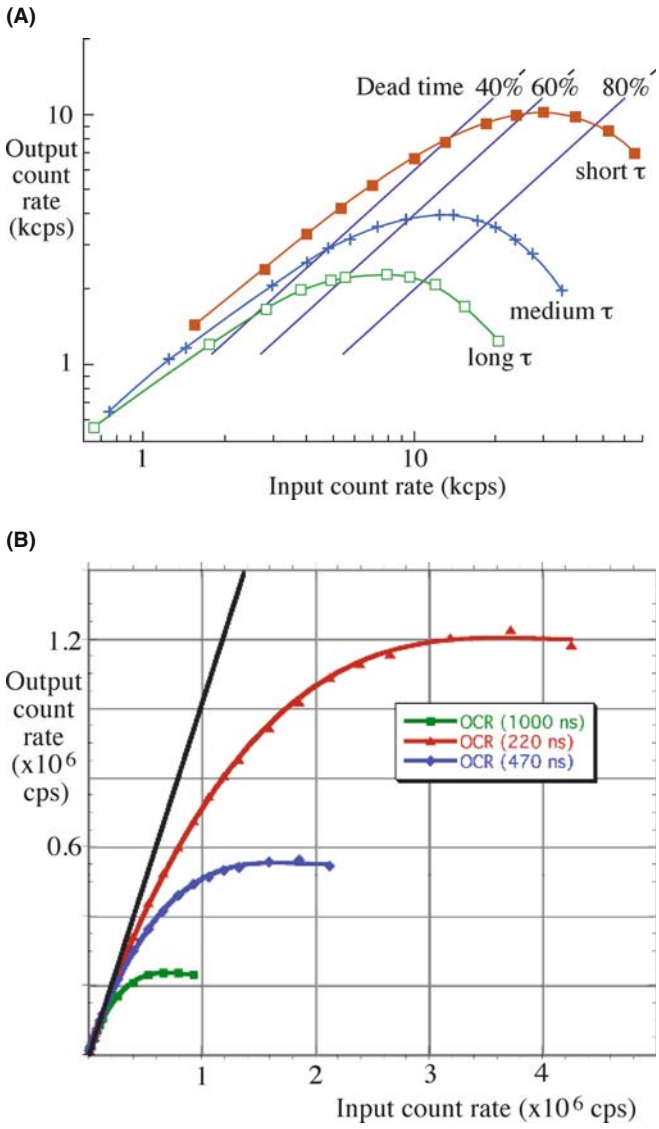


FIGURE 32.12. (A) The output count rate in a given clock time as a function of dead time. The maximum processing efficiency is reached at ~60% dead time. It is very inefficient to use the system above the maximum output rate because of the very long (clock) times needed to gather a spectrum. Increasing the (analog) time constant results in fewer counts being processed and a drop in the output count rate. (B) Data from an SDD on an SEM at three different time constants: note the enormous increase in output count rate up to 1.2 million cps.

TOO THIN

In XEDS, your specimen may be too thin: you need enough counts.

As shown in Figure 32.13, digital processing permits a higher throughput over a continuous range of energy resolution than the fixed ranges available from each specific (in this case six) τ for an analog system. We've already mentioned that C_s correction permits larger apertures to be used and therefore, significantly more current can be put into small probes, but it's unlikely that if you have a C_s-corrected AEM it will be still interfaced to an old analog XEDS system.

In summary, you should occasionally

- Check the energy calibration of the computer display.
- Check the dead-time circuitry by the linearity of the output count rate versus beam current (Figure 32.11).
- Check the counts in a fixed clock time as a function of beam current to determine the maximum output count rate (Figure 32.12).

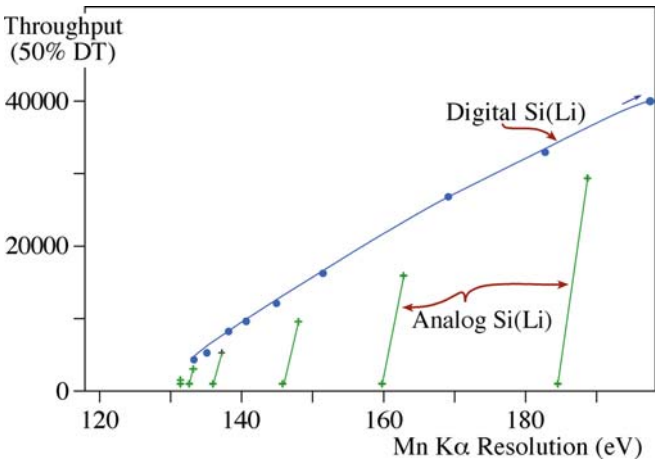


FIGURE 32.13. Digital pulse processing gives a continuous range (blue line) of X-ray throughput at 50% dead time, compared with a set of fixed throughput ranges (green lines) for specific (analog) time constants.

32.10 THE XEDS-AEM INTERFACE

In your TEM, an intense beam of high-energy electrons bombards your specimen, which scatters many electrons. The specimen *and any other part of the TEM* that is hit by these electrons emit both characteristic and bremsstrahlung X-rays (which have energies up to that of the electron beam). X-rays of several tens or hundreds of keV can penetrate long distances into the material and fluoresce characteristic X-rays from anything that they hit. Ideally, the XEDS should only 'see' the X-rays from the beam-specimen interaction volume. However, as shown in Figure 32.14, it is not possible to prevent radiation from the stage and other areas of the specimen from entering the detector. As you can see from Figure 32.14, the XEDS has a collimator in front of the detector crystal. This collimator is the last line of defense against the entry of undesired radiation from the stage region of the microscope.

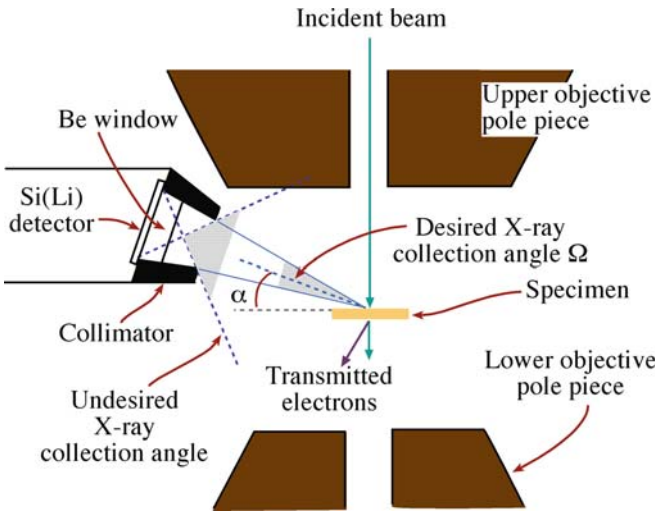


FIGURE 32.14. Schematic diagram of the interface between the XEDS and the AEM stage showing how the detector can 'see' X-rays from regions other than the beam-specimen interaction volume over the (relatively large) undesired collection angle. The (relatively small) desired collection angle Ω and take-off angle α are also shown.

It's worth checking if you have an ideal collimator, constructed of a high-Z material, such as W, Ta, or Pb, coated externally and internally with a low-Z material, such as Al, C, or Be. The low-Z coating minimizes the production of X-rays from backscattered electrons that happen to spiral into the collimator. The high-Z material absorbs high-energy bremsstrahlung. The inside of the collimator should be baffled to prevent any backscattered electrons from generating X-rays that then penetrate the detector. No collimator is entirely successful and we'll describe the contribution of system-produced X-rays in the next chapter.

THE COLLIMATOR
The collimator defines both the (desired) collection angle of the detector and the average take-off angle of X-rays entering the detector.

32.10.A Collection Angle

The detector collection angle (Ω) is the solid angle subtended at the analysis point on the specimen by the active area of the front face of the detector. The collection angle is shown in Figure 32.14 and is defined as

$$\Omega = \frac{A \cos \delta}{S^2} \tag{32.4}$$

where A is the active area of the detector (usually 30 mm^2), S is the distance from the analysis point to the detector face, and δ is the angle between the normal to the detector face and a line from the detector to the specimen. In many XEDS systems, the detector crystal is tilted toward the specimen so $\delta = 0$, then $\Omega = A/S^2$. It is clear that to maximize Ω the detector should be placed as close to the specimen as possible.

As we'll see, in most AEM experiments, it is the low X-ray count rate that limits the accuracy of the experiment. Commercial Si(Li) crystals have A values from 10 to 30 mm^2 and 50 mm^2 is becoming more common. As a result, values of Ω dictated by the closest distance between the specimen and detector/collimator lie in the range from 0.3 down to 0.03 sr. ATW detectors invariably have lower Ω values than Be-window or windowless detectors because the polymer window has to be supported on a grid which reduces the collection angle by ~20%. IG detectors need a reflective window to prevent IR radiation from generating noise in the

detector. So, even with the largest detectors, Ω is a small fraction of the total solid angle of characteristic X-ray generation, which is, of course, 4π sr.

Ω

The value of Ω is the most important parameter in determining the quality of your X-ray analysis. You need three things for good AEM X-ray analysis: counts, counts, and more counts.

This value of Ω is calculated from the dimensions of the stage and the collimator. There is no way, however, that you can measure this critical parameter directly, although you can compare X-ray count rates between different detector systems using a standard specimen such as our thin NiO film and a known beam current. A useful figure of merit is the X-ray counts per second detected from a standard specimen, with a given beam current and a given detector collection angle (cps/nA/sr). Typically, for an AEM with a nominal Ω of 0.13 sr and a beam energy of 300 keV the figure of merit is >8000. For a beam energy of 100 keV, it is about 13,000 (Zemyan and Williams). The increase at lower keV is due to the increased ionization cross section.

The magnitude of Ω is limited because the upper polepiece of the objective lens gets in the way of the collimator, thus limiting *S*. To avoid this limitation, we could increase the polepiece gap but doing so would lower the maximum beam current and degrade the image resolution, both of which are highly undesirable. The advent of C_s-corrected AEMs has removed this limitation and will open up a new chapter in high count-rate X-ray analysis in AEMs (Watanabe and Williams). Arrays of SDDs have already been used to give solid angles >1 sr on PIXE systems (Doyle et al.), so the future of XEDS in the AEM has exciting possibilities.

32.10.B Take-Off Angle

The take-off angle α is the angle between the specimen surface (at 0° tilt) and a line to the center of the detector, as shown in Figure 32.14. Sometimes, it is also defined as the angle between the transmitted beam and the line to the detector, which is simply (90° + α). In the SEM/EPMA, we keep α high to minimize X-ray absorption in bulk specimens. Unfortunately, in an AEM, if we maximize α, the price we pay is lowering Ω. A high-angle detector has to be positioned above the upper objective polepiece and will be much farther from the specimen. In the EPMA, low Ω is not a problem because there are always sufficient X-rays from a bulk specimen, but in the AEM the highest possible Ω is essential, as we've already emphasized (counts, counts...).

In those AEMs where the detector has a high take-off angle but a low Ω, the poor X-ray count rate makes quantitative analysis much more time consuming. Keeping the detector below the polepiece restricts α to a maximum of about 20°. In most cases you will find that such a small value of α is not a problem because one of the major advantages of thin-specimen AEM compared to EPMA is that absorption can usually be neglected. If absorption is a problem in your particular specimen, you can reduce the X-ray path length by tilting it toward the detector, thus increasing α (see Section 35.5). We don't recommend tilting because it increases spurious effects (see Chapter 33) and also lowers the P/B (peak to background) ratio in the spectrum.

TAKE-OFF ANGLE AND COUNTS

We would like to maximize the take-off angle *and* maximize the count rate, but we can't.

32.10.C Orientation of the Detector to the Specimen

There are two simple questions that you must be able to answer.

(a) *Is the detector pointing on axis?* The detector is inserted to a point where it is almost touching the objective polepiece, and you hope that it is 'looking' at the region of your specimen that is on the optic axis when the specimen is eucentric and at zero tilt. To find out if your system is thus aligned, take a low-magnification X-ray map from a homogeneous specimen, such as our thin NiO film. If the detector is not pointing on axis, the map will show an asymmetric intensity. Alternatively, if you can't map at a low enough magnification, simply see how the Ni K_α intensity varies from area to area on the foil with the stage traverses set at zero and different areas selected using the beam deflectors. The maximum intensity should be recorded in the middle grid square and for some distance around. It is also instructive to do the same test with the specimen moved above or below the eucentric plane. Again, the maximum intensity should be recorded at eucentricity. If the intensity is asymmetric then the detector or the collimator is not well aligned and some of your precious X-ray counts are being shadowed from the detector, probably by the collimator. So you need to ask for technical help.

(b) *Where is the detector with respect to the image?* It is best if the detector is 'looking' toward a thin region of the specimen, rather than toward a thicker region, as shown in Figure 32.15A. This alignment minimizes the X-ray path length through your specimen. If the BF image rotates when changing magnification in your TEM then the apparent detector orientation with respect to the image (on the screen) will vary with

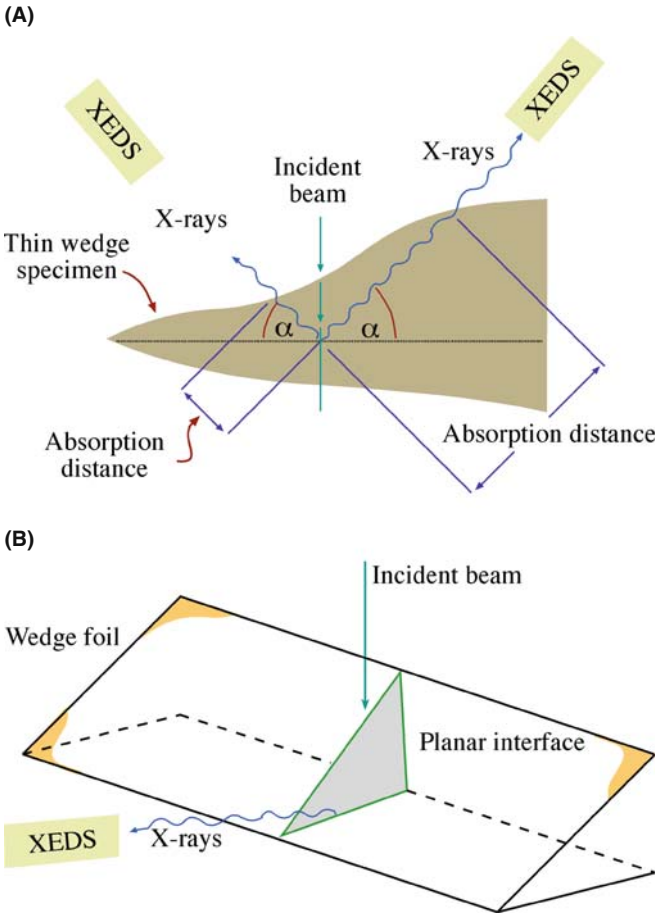


FIGURE 32.15. (A) The position of the XEDS relative to a wedge-shaped thin foil results in different X-ray path lengths. The shortest path length with the detector 'looking' at the thinnest region of the foil minimizes any X-ray absorption. (B) The preferred orientation of the XEDS when analyzing a planar defect: the interface plane is parallel to the detector axis and the incident-beam direction.

magnification. A STEM BF doesn't rotate so the relative orientation of the detector to your image is fixed. It is simple to find this orientation if the detector axis (y axis) is normal to a principal traverse axis (x) of the stage. If you press gently on the end of your holder, the image will move in the $+x$ direction. Then you can determine geometrically the direction ($+y$ or $-y$) along which the detector is 'looking' with respect to that $+x$ direction in the TEM image. In STEM, the image might be rotated 180° with respect to the TEM image so you have to take this into account. But it's all a good exercise in 3D geometry.

If you're doing analysis or mapping across a planar interface, which is a common AEM application, then you should also orient your specimen such that the interface is parallel to the detector axis and the beam, so the detected X-rays have come through regions of similar composition and don't cross the interface. A tilt-rotation holder is ideal for this but, if a low-background version is not available, then you may need to take out the holder and re-position your foil manually a few times until the interface is in the right orientation (see Figure 32.15B).

LOOKING AT THE SPECIMEN

The XEDS detector must be 'looking' at the thin edge of your specimen and aligned with any planar interface you are studying.

32.11 PROTECTING THE DETECTOR FROM INTENSE RADIATION

If you are not careful, the XEDS electronics can be temporarily saturated if high doses of electrons or X-rays hit the detector. The detector itself may also be damaged, particularly in intermediate-voltage AEMs. As we've noted, these situations occur when you inadvertently move thick areas of your specimen under the beam, or if you are traversing around a thin specimen and the support grid hits the beam. To avoid these problems, shutter systems are built into most XEDS systems which automatically protect the Si(Li) crystal if the AEM is switched to low magnification or if the pulse processor detects too high a flux of radiation.

If you don't have a shutter then you can physically retract the detector to lower Ω (if it is retracted along a line of sight to the specimen) or remove the detector from out of view of the specimen. The drawback to this approach is that constant retraction and reinsertion of the detector may cause undue wear on the sliding 'O'-ring seal. Also you may reposition your detector slightly differently each time, unless the system is designed so you can push the detector to a fixed stop, thus insuring a constant Ω and α. A shutter is highly recommended!

SHUT THE SHUTTER

To avoid reliance on the automatic system, it is best to have the shutter closed until you have decided which area you want to analyze and it is thin enough that the generated X-ray flux doesn't saturate the detector. Also, never have the objective aperture inserted.

CHAPTER SUMMARY

The XEDS (usually of the Si(Li) variety) is the only X-ray spectrometer currently used in TEMs. It is remarkably compact, efficient, and sensitive. A combination of Si(Li)/SDD and IG detectors can detect K_α lines from all the elements from Be to U. The XEDS is limited by its poor energy resolution, artifacts in the spectra, and the need for cooling, but it is simple to run and maintain. You must take care to perform certain basic procedures and refrain from certain others that can damage the detector. Sometimes, it may be too simple; beware. You need to

- Measure your detector resolution every 6 months at the Mn or Ni K_α line (at best 130–140 eV for Si(Li) or SDD and 120–130 eV for IG).
- Measure the ICC (FWTM/FWHM ratio of the Ni K_α line: ideally 1.82), every 6 months.
- Unless you have an SDD or a Peltier-cooled Si(Li), monitor ice build-up via in the Ni K_α/L_α ratio on a monthly basis.
- Check the calibration of the energy range of your computer display every 6 months especially if you have an SDD.
- Check the dead-time correction circuitry by the linearity of the output count rate versus beam current, every 6 months.
- Check the counts in a fixed clock time as a function of beam current to determine the maximum output count rate, every 6 months.
- Always operate with the shutter closed until you are ready for analysis.
- Always retract the objective diaphragm prior to analysis.
- Ensure the XEDS is pointing toward the thin edge of any wedge/disk specimen.

Interfacing your XEDS to the AEM is crucial since it determines the count rate, the need for an absorption correction, and the intrusion of spurious X-rays into your spectrum. In any decision involving XEDS in the AEM, you should always choose the option that optimizes the count rate.

For the sake of completeness, Table 32.2 shows you the relative merits of the various detectors that we have discussed in this chapter.

GENERAL TEXTS

Garratt-Reed, AJ and Bell, DC 2002 *Energy-dispersive X-ray Analysis in the Electron Microscope* Bios (Royal Microsc. Soc.) Oxford UK. Similar in scope to the XEDS chapters in this textbook.

Goldstein, JI, Newbury, DE, Echlin, P, Joy, DC, Romig, AD Jr, Lyman, C., Fiori, CE and Lifshin, E 2003 *Scanning Electron Microscopy and X-ray Microanalysis* 3rd Ed. Springer New York. In-depth treatment of all aspects of XEDS in the SEM/EPMA. Includes details of the electronics (Section 3.2.7)

Goodhew, PJ, Humphreys, FJ, and Beanland, R 2001 *Electron Microscopy and Analysis* 3rd Ed. Taylor and Francis New York. A broad introduction covering SEM, TEM, and AEM.

Jones, IP 1992 *Chemical Microanalysis Using Electron Beams* Institute of Materials London. Quantitative AEM; lots of calculations to illustrate the analytical principles; essential for the serious X-ray analyst.

Williams, DB, Goldstein, JI and Newbury, DE, Eds. 1995 *X-Ray Spectrometry in Electron Beam Instruments* Plenum Press New York. Tells all you need to know and more about X-ray detection and processing in SEM/EPMA (mainly) and TEM.

REFERENCES FOR THE CHAPTER

Doyle, BL, Walsh, DS, Kotula, PG, Rossi, P, Schülein, T and Rohde, M 2004 *An Annular Si Drift Detector μPIXE System Using AXSIA Analysis* X-Ray Spectrom. **34** 279–284. Illustrating the use of an array of SDDs.

Egerton RF and Cheng SC 1994 *The Use of NiO Test Specimens in Analytical Electron Microscopy* Ultramicrosc. **55** 43–54. As it says!.

Lund, MW 1995 *Current Trends in Si(Li) Detector Windows for Light Element Analysis* in *X-Ray Spectrometry in Electron Beam Instruments* DB Williams, JI Goldstein and DE Newbury, Eds. 21–31 Plenum Press New York. Detailed review of windows.

Lyman, CE, Newbury, DE, Goldstein, JI, Williams, DB, Romig, AD Jr, Armstrong, JT, Echlin, PE, Fiori, CE, Joy, D., Lifshin, E and Peters, KR 1990 *Scanning Electron Microscopy*, *X-Ray Microanalysis and*

Analytical Electron Microscopy; *A Laboratory Workbook* Plenum Press New York. Includes some standard tests for XEDS.

Michael, JR, 1995 *Energy-Dispersive X-ray Spectrometry* in *Ultra-High Vacuum Environments* in *X-Ray Spectrometry in Electron Beam Instruments* Eds. DB Williams, JI Goldstein and DE Newbury p83 Plenum Press New York. Using the Ni K_α/Ni L_α ratio.

Mott, RB and Friel, JJ, 1995 *Improving EDS Performance with Digital Pulse Processing* in *X-Ray Spectrometry in Electron Beam Instruments* Eds. DB Williams, JI Goldstein and DE Newbury 127–155 Plenum Press New York. Digital processing as we discuss in Section 32.7.

Newbury, DE 2006 *The New X-ray Mapping: X-ray Spectrum Imaging Above 100 kHz Output Count Rate with the Silicon Drift Detector* Microscopy and Microanalysis **12** 26–35. Using SDDs for mapping.

Sareen, RA, 1995 *Germanium X-ray Detectors* in *X-Ray Spectrometry in Electron Beam Instruments* Eds. DB Williams, JI Goldstein and DE Newbury 33–51 Plenum Press New York. IG detectors.

Statham, PJ 1995 *Quantifying Benefits of Resolution and Count Rate in EDX Microanalysis* in *X-Ray Spectrometry in Electron Beam Instruments* Eds. DB Williams, JI Goldstein and DE Newbury 101–126 Plenum Press New York. Balancing count rate and resolution.

Terauchi, M and Kawana, M 2006 *Soft-X-ray Emission Spectroscopy Based on TEM—Toward a Total Electronic Structure Analysis* Ultramicrosc. **106** 1069–1075. CCD-based WDS.

Terauchi, M, Yamamoto, H and Tanaka, M 2001 *X-ray Emission Spectroscopy, Transmission Electron Microscope, DOS of the Valence Band, Soft-X-ray Spectrometer, B K-emission Spectra, Hexagonal Boron-Nitride* J. Electr. Microsc. **50**(2) 101–104. Development of a sub-eV resolution soft-X-ray spectrometer for a transmission electron microscope.

Watanabe, M and Williams, DB 2006 *Frontiers of X-ray Analysis in Analytical Electron Microscopy: Toward Atomic-Scale Resolution and Single-Atom Sensitivity* Microscopy and Microanalysis **12** 515–526. C_s-corrected AEM.

Wollman, DA, Nam, SW, Hilton, GC, Irwin, KD, Bergren, NF, Rudman, DA, Martinis, JM and Newbury DE 2000 *Microcalorimeter Energy-Dispersive Spectrometry Using a Low Voltage Scanning Electron Microscopy* J. Microsc **199** 37–44. The bolometer for XEDS.

Zemyan, SM and Williams, DB 1995 *Characterizing an Energy-Dispersive Spectrometer on an Analytical Electron Microscope* in *X-Ray Spectrometry in Electron Beam Instruments* Eds. DB Williams, JI Goldstein and DE Newbury 203–219 Plenum Press New York. Summary of standard tests for XEDS.

URLs: SELECTION OF XEDS MANUFACTURERS

1) www.bruker-axs.de/
2) www.edax.com/
3) www.oxinst.com
4) www.pgt.com
5) www.thermo.com/

SELF-ASSESSMENT QUESTIONS

Q32.1 Define: XEDS, IG, SDD, AEM.
Q32.2 Explain in four steps how X-rays from the specimen are converted into a spectrum.
Q32.3 Distinguish between dead time, live time, and clock time.
Q32.4 Distinguish between dead layer(s) and active regions of detectors.
Q32.5 List four ways that you can damage your detector while you're operating the AEM.
Q32.6 What is more important during X-ray microanalysis and why: X-ray counts, X-ray energy resolution, X-ray take-off angle, or specimen tilt?
Q32.7 What is the best accelerating voltage to use for XEDS?
Q32.8 Why must Si detectors be doped with Li while IG detectors are not doped (Hint: what does the 'I' mean?)?
Q32.9 How does this Li affect the operation of the detector? Why is there no Li in a SDD?
Q32.10 What type of detector is better for detecting (a) high-Z materials, (b) lower-Z materials and why?
Q32.11 Why is pulse processing required to translate X-ray photons into a spectrum?
Q32.12 What is a reasonable dead time and how is this affected by the vintage of your detector electronics?
Q32.13 Why is digital pulse processing preferred over analog processing? Give one exception to this preference.
Q32.14 What advantage does XEDS have over WDS?
Q32.15 What are some of the disadvantages to XEDS?
Q32.16 Why is a large collection angle, Ω, useful? What limits the value of Ω in practice?
Q32.17 Why is a high take-off angle, α, useful? What limits the value of α in practice?
Q32.18 Why is a larger Ω more important than a higher α?
Q32.19 Why is the collimator so important in front of your detector?
Q32.20 What aspects of TEM design restrict the use of large arrays of detector crystals such as SDDs?
Q32.21 Why is it not a major issue that we cannot optimize the take-off angle α?

Q32.22 What materials are used in the collimator to avoid X-ray generation within the collimator and what might cause the generation of these X-rays in the first place?

Q32.23 You switch on your XEDS detector but you don't register any X-rays coming through the system. Explain what may be causing this situation.

TEXT-SPECIFIC QUESTIONS

T32.1 Explain what factors control the shape of the characteristic peaks and the background in a typical energy-dispersive spectrum such as Figure 32.2A.

T32.2 List three reasons why we cool Si(Li) or IG detectors with liquid N_2 and three undesirable consequences. Search the Web sites of the major EDS manufacturers to see if there are alternatives to liquid-N_2 cooling. Why do we not have to cool an SDD to liquid-N_2 temperatures?

T32.3 Explain why, in comparison to a Si(Li) detector, an IG detector exhibits (a) better energy resolution, (b) more resistance to high-energy electron damage, (c) less susceptibility to accidental loss of liquid N_2 and (d) better detection of high-energy X-rays. Given these advantages, why are IG detectors not more common on AEMs?

T32.4 Explain why, in comparison to a Si(Li) detector, a wavelength-dispersive (crystal) spectrometer (WDS) offers (a) better resolution, (b) more throughput of counts, (c) fewer artifacts. Given these advantages, why is the WDS not more common on AEMs?

T32.5 Why does the output X-ray count rate rise to a maximum and then fall as the input count rate continues to rise (as in Figure 32.12A)?

T32.6 So why don't we always operate the AEM at the maximum count rate (which would, inter alia, improve counting statistics, thus reducing errors), reduce the time required to acquire the spectrum (thus reducing damage to beam-sensitive specimens)?

T32.7 Consider the curves in Figure 32.12. Explain why the statements in the caption are true. (Are they?)

T32.8 Give examples of how the limitations of the curves in Figure 32.12A might be overcome?

T32.9 Explain why SDDs can operate at much higher count rates than other XEDS detectors.

T33.10 An engineer decided to remove the collimator baffles from her AEM. What does she put up with as a consequence?

T33.11 Why would it be good to have both the largest detector take-off angle and the highest detector collection angle? Why do we have to make a compromise choice in practice? Explain why correction of the spherical aberration coefficient of the objective lens would remove this compromise.

T32.12 Figure 32.9 shows a WDS system on a TEM. Why is it not commonly used now? (A careful discussion is required.)

T32.13 Look at Figure 32.9D. Why are the EDS peaks so much broader than those from the bolometer and could this possibly be improved?

T32.14 An image produced using a backscattered-electron detector shows the presence of three regions of significantly different contrast. Where should you begin to further analyze this specimen?

T32.15 A strong background appears in the 0–10 keV spectrum of a geological specimen. What is this specimen, and how can the high background be compensated for?

T32.16 Using a windowless detector it becomes apparent by comparing past and previous spectra of the same specimen that there is likely a coating of ice or hydrocarbons on the detector. Should you just thaw out the detector or not?

T32.17 After measuring XEDS spectra from a biological or polymeric specimen, you close your AEM session but find yourself physically unable to open the chamber to remove the specimen. What is likely to be the cause?

T32.18 You have decided to switch an IG detector in for the Si(Li) already in your microscope to identify X-ray lines above 20 keV. After the switch, you notice what appears to be a large increase in the number of escape peaks in your spectra. Name one reason why this might be happening.

T32.19 Barry and David are laughing about something you did in lab last week. You were attempting to optimize the energy resolution of the XEDS spectra you obtained. What is their point?

T32.20 Are there hidden peaks obscured by the low-energy edge of several of your more intense peaks and what can you do about it if there are?

T32.21 A novice at using the XEDS on the TEM decides to employ WDS instead, with the notion in mind that it wouldn't be such a bad thing to get more resolution for the spectra. Which rude awakening is this person close to discovering?

T32.22 A geological specimen is elementally analyzed using XEDS. This turns out to contain Si, Al, O, and Fe. Out of curiosity, an X-ray map in search of Cu is performed on this same region of the specimen. Multiple locations of Cu spring up in the map. Explain.

33

X-ray Spectra and Images

CHAPTER PREVIEW

The X-ray spectrum generated within your specimen consists of element-specific characteristic peaks with well-defined energies superimposed on a non-characteristic background. While the XEDS system is a remarkable piece of technology, we've already described its limited resolution and we will see in this chapter that it is also prone to creating small artifact peaks in the spectrum. Furthermore, the unavoidable presence of scattered electrons and X-rays within the AEM conspire to degrade the quality of the generated spectrum and increase the number of false peaks in the displayed spectrum. The AEM illumination system and specimen stage are rich sources of powerful radiation, not all of it by any means coming from the area of interest in your specimen. So you have to take precautions to ensure that the X-ray spectrum you collect comes predominantly from the area of the specimen that you want to analyze and we describe several tests you should perform to ensure that the XEDS-TEM interface is optimized. Once you understand the desirable and undesirable contents of a spectrum, we'll show you the various ways to gather your spectra, display them, and also form X-ray images from them. In particular, digital spectrum imaging is the most powerful technique for optimizing the information that you can gather from the low number of counts generated in thin foils.

33.1 THE IDEAL SPECTRUM

Back in Chapter 4 we learned that electrons generate two kinds of X-rays. When electrons ionize an atom which returns to ground state, the emitted characteristic X-ray *energy* is unique to the ionized atom (Figure 4.2). When electrons are slowed by electrostatic interaction with the nucleus they produce a continuum of bremsstrahlung X-rays and together the characteristic peaks and the bremsstrahlung background comprise the X-ray spectra we detect and display via the XEDS (as shown schematically back in Figure 4.6 and experimentally in Figure 32.2 and many others).

33.1.A The Characteristic Peaks

Some more detailed revision: a beam electron ionizes an atom in your specimen by ejecting an inner or core-shell electron, leaving a hole in the shell. The probability of this event occurring is governed by the ionization cross section. Then a cascade of electron transitions occurs, with each transition filling the hole with an electron from a more weakly bound shell (leaving a hole in that shell, and so on) and ultimately the last electron falls into a core-shell from the conduction band. Depending on the fluorescence yield, each transition results in either a characteristic X-ray or an Auger electron. The characteristic X-rays have a well-defined energy and a natural 'line width' (the FWHM of the Gaussian distribution of X-ray energies) of typically 1–5 eV. But, as you already know from the previous chapter, the XEDS degrades this width to a Gaussian-shaped peak with a FWHM of about 135 ± 10 eV. We'll use the term 'line' to denote the actual X-ray energy at the peak of the Gaussian and we'll talk about K, L, and M families of lines (and indeed the XEDS computer display includes lines superimposed on the spectra at the K, L, M, etc., peak energies). The actual number of counts (intensities) of the characteristic peaks from a given element, and the relative differences between spectra from different elements, are really quite complex and we'll go into this more in the next chapter when we discuss details of spectra and qualitative analysis. For the time being, all you need to know is that, generally speaking, lower-energy X-ray peaks are more intense than higher-energy ones and the heavier the element,

the more complex the characteristic spectrum. Also X-rays with energies $< \sim 1$ keV are absorbed both within your specimen and within the detector and ultimately, the combination of absorption and low fluorescence yield means it is not possible to detect X-rays with energies $< \sim 110$ eV (the Be K line).

33.1.B The Continuum Bremsstrahlung Background

The background in the X-ray spectrum is bremsstrahlung (braking) radiation arising as beam electrons are slowed down or stopped by electrostatic interactions with nuclei in the specimen. The intensity of the bremsstrahlung is zero at the beam energy (since we can't create more energy than the beam has) and rises until it is effectively infinite at zero energy. This distribution is well understood and described mathematically by variations on Kramers' law as we also noted back in Figure 4.6. Modifications to Kramers' law account well for the absorption of the X-rays and accurately describe the experimentally detected bremsstrahlung distribution, even from thin specimens. So the net result is that X-ray spectra consist, as we have seen, for example, in Figures 1.4A and 32.2, of characteristic, Gaussian-shaped peaks superimposed on a background of bremsstrahlung X-rays, most clearly visible in Figure 32.2F. Many more spectra will appear throughout this and subsequent chapters and, as already noted, we can create elemental images from specific characteristic peaks.

Unfortunately, what makes X-ray spectrometry challenging is that the spectrum generated within your thin specimen, the spectrum detected by your Si(Li) detector, and the spectrum displayed on your computer screen are all quite different. So understanding what controls the X-ray counts that you measure in the characteristic peaks in your spectra is absolutely essential. Otherwise it will be difficult to first identify (Chapter 34) and then quantify (Chapter 35) the peak intensities in terms of the presence and the amount, respectively, of the elements in the analysis volume of your specimen.

We'll first describe how the XEDS system generates artifact (escape, sum, and internal fluorescence) peaks and then we'll show how the AEM itself also contributes unwanted X-rays, all of which can confuse the unwary operator into thinking certain elements are in their specimen when, in fact, they are not. Misinterpretation of the presence or absence of certain elemental signals could cause real problems if, for example, you are making decisions on the suitability or otherwise of a specific material for a given application, or even more seriously, making a forensic or medical decision based on the chemistry of your specimen: more about this in Chapter 34.

33.2 ARTIFACTS COMMON TO Si(Li) XEDS SYSTEMS

The XEDS system introduces its own artifacts into the spectrum. Fortunately, we understand all these artifacts, but they still occasionally mislead the unwary operator; see the review by Newbury. We can separate the artifacts into two groups

- Signal-detection artifacts: examples are escape peaks and internal fluorescence peaks.
- Signal-processing artifacts: e.g., the sum peaks.

Escape peak: Because the detector is not a perfect sink for all the X-ray energy, it is possible that a small fraction of the energy is lost and not transformed into electron-hole pairs. The easiest way for this to happen is if the incoming photon of energy E fluoresces a Si K_α X-ray (energy 1.74 keV) which escapes from the intrinsic region of the detector. The detector then registers an apparent X-ray energy of $(E - 1.74)$ keV, as shown in Figure 33.1.

ESCAPE PEAK
Si escape peaks appear in the spectrum 1.74 keV below the true characteristic peak position.

The magnitude of the escape peak depends on the design of the detector and the energy of the fluorescing X-ray. The most efficient X-ray to fluoresce Si K_α X-rays is the P K_α, but in a well-designed detector even the P escape peak will only amount to <2% of the P K_α intensity. This fact explains why you can only see escape peaks if there are major characteristic peaks in the spectrum. More escape peaks occur in IG spectra because we can fluoresce both Ge K_α (9.89 keV) and L_α (1.19 keV)

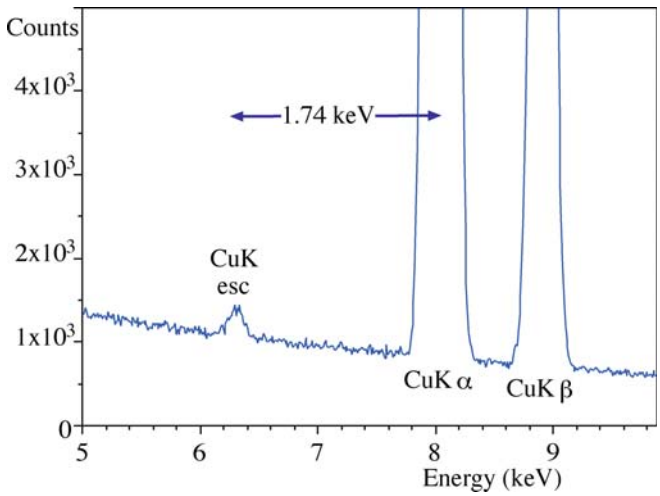


FIGURE 33.1. The escape peak in a spectrum from pure Cu, 1.74 keV below the Cu K_α peak. The intense K_α peak is truncated in the display because it is 50–100× more intense than the escape peak.

characteristic X-rays in the detector. Each of these will cause corresponding escape peaks. SDDs are thinner than typical Si(Li) crystals so there may be a slightly enhanced possibility for an escape peak. The analysis software should be able to recognize any escape peak, remove it, and add the intensity back into the characteristic peak where it belongs. Because the escape-peak intensity is so small it is rarely a problem unless you misinterpret it as coming from an element which is not present.

The internal fluorescence peak: This is a characteristic peak from the Si (or Ge) in the detector dead layer. Incoming photons can fluoresce atoms in the dead layer and the resulting Si K_α or Ge K/L X-rays enter the intrinsic region of the detector which cannot distinguish their source and, therefore, register a small peak in the spectrum. As semiconductor detector design has improved and dead layers have decreased in thickness, the internal fluorescence peak artifact has shrunk but it has not yet disappeared entirely. So beware.

Obviously if you are looking for small amounts of Si in your specimen you'll always find it in a Si(Li) spectrum after long enough counting time! Depending on the dead-layer thickness, the Si K peak intensity corresponds to ~0.1–1% of the specimen composition (see Figure 33.2) so it is hardly a major problem if you are aware of it. Similar effects are observed in IG detectors also, but the very thin dead layer in SDDs should be an advantage.

LONG COUNTING TIMES

A small Si K_α peak will occur in ALL spectra from Si(Li) detectors after long counting times.

Sum peak: As we described earlier, the processing electronics are designed to switch off the detector while each pulse is analyzed and assigned to the correct energy channel. The sum peak arises when the count rate exceeds the electronics' ability to discriminate all the individual pulses and so-called 'pulse pile-up' occurs. This is likely to occur when

- The input count rate is high.
- The dead times are $>\sim 60\%$.
- There are major characteristic peaks in the spectrum.

The electronics simply cannot be perfect. Occasionally, two photons will enter the detector at exactly the same time. The analyzer then registers an energy corresponding to the sum of the two photons. Since this coincidence event is most likely for the X-ray giving the major peak, a sum peak first appears at twice the energy of the major peak, as shown in Figure 33.3.

If you are using an SDD because you have a C_s-corrected AEM and are seeking to maximize count rates, you can generate multiple sum peaks, causing serious interpretation problems in your analyses. There is evidence (see Newbury's paper) that digital processing is, in fact, worse than analog processing in terms of generating sum peaks in SDD systems, and hybrid analog/digital systems are being considered in the SEM field where this is a bigger problem.

THE SUM PEAK

The sum peak should be invisible if you maintain a reasonable input count rate, typically <10,000 cps, which should give a dead time of <60%.

Since you can't usually generate very high count rates, unless your specimen is really thick, there is little need to worry about sum peaks in AEM spectra but, as always,

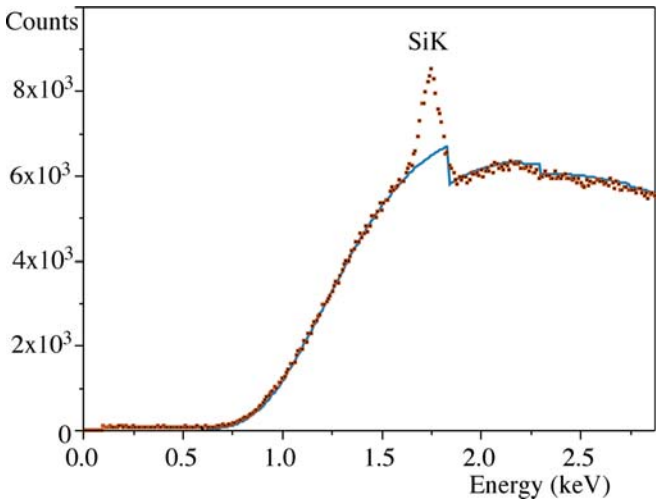


FIGURE 33.2. The Si internal fluorescence peak in a spectrum from pure carbon obtained with a Be-window Si(Li) detector. The ideal spectrum is fitted as a continuous line that only shows the Si absorption edge.

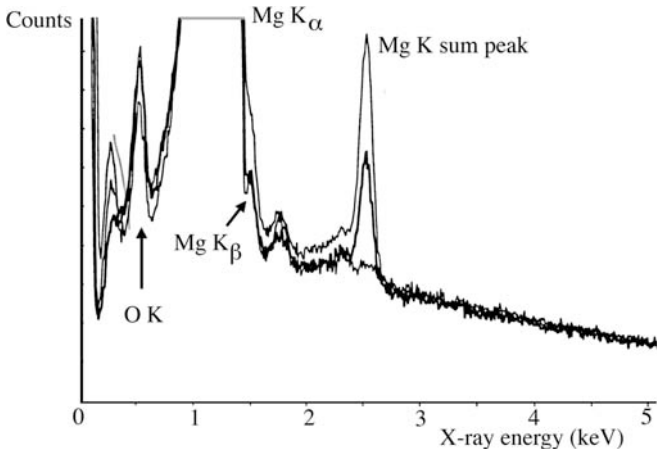


FIGURE 33.3. The Mg K sum (coincidence) peak occurs at twice the Mg K_α peak in this spectrum from a bulk specimen of (oxidized) pure Mg. The sum peak decreases change rapidly with decreasing dead times; upper trace 70%, middle trace 47%, lower trace 14% dead time. The sum-peak artifact is close to the background intensity at 14% dead time.

you should at least be aware of the danger. (For example, the Ar K_α energy is almost exactly twice the Al K_α energy and the sum peak has led some researchers to report Ar being present in Al specimens when it wasn't, and has caused others to ignore Ar, which actually was present in ion-milled specimens!) One exception to this lack of concern is if you have intense low-energy peaks where the residual noise in the electronic circuitry interferes with the pile-up rejection. So if you're analyzing elements lighter than Mg, take care to use low input count rates. Reducing the dead time to 10–20% should remove even the Mg K_α sum peak, as shown in Figure 33.3.

Much of what we have just discussed and most of what we'll cover in Chapters 34 and 35, can be observed experimentally on the AEM. But it is often just as instructive, and certainly easier, to simulate the spectra. To this end, we strongly advise you to download the public-domain simulation software 'Desktop Spectrum Analyzer' (DTSA) from NIST (Fiori and Swyt and URL #1); it is listed in the recommended software in Section 1.6 and its use is described in depth in the companion text. In fact, almost all the spectra in this and the subsequent chapter were generated by DTSA (in part, to ensure a uniform appearance). This software permits realistic simulation of XEDS data in TEMs (and SEMs) and introduces you to all the aspects of spectral processing, artifacts, modeling, etc., that are discussed in this and the next three chapters.

33.3 THE REAL SPECTRUM

In the perfect AEM, all spectra would be characteristic only of the chosen analysis region of your specimen with which the beam interacts. The analysis of bulk specimens in the SEM/EPMA approaches this ideal but, in the AEM, two factors combine to introduce false information which we call system and spurious X-rays. These X-rays introduce small errors into both qualitative and quantitative analysis, unless you are aware of the dangers and take appropriate precautions to identify and minimize the problems. The two factors that make the AEM different to the EPMA and that are responsible for these problems are

- The high-voltage electrons which generate intense doses of stray X-rays and scattered electrons in the illumination system.
- The thin specimen which scatters high-energy electrons and X-rays around the limited confines of the AEM stage.

Modern AEMs are designed to minimize some of these problems, but they have not disappeared entirely. It is important for you to understand that these problems are *not* major distortions of the X-ray spectrum, and only introduce small peaks or changes in intensity equivalent to composition changes at the 1% level or less. Nevertheless, identification and quantification of such small elemental amounts are often the *raison d'être* for an analysis (e.g., impurity segregation), so you have to be wary of these artifacts which we'll now discuss in some detail. Remember, these artifacts are in addition to the XEDS artifacts, described in the previous section.

33.3.A Pre-specimen Effects

The TEM illumination system produces high-energy bremsstrahlung X-rays and electrons scattered outside the main beam, both of which may strike the specimen, producing spurious X-rays.

In inhomogeneous specimens (which are usually just the kind that we want to analyze) the presence of significant amounts of spurious X-rays means that the quantification process could give the wrong answer. There are reviews (e.g., Williams and Goldstein or Allard and Blake), which describe in detail how to identify and minimize these artifacts from the illumination system, so we will just describe the precautions necessary to insure that the AEM is operating acceptably. Since these artifacts are primarily a result of the high-energy electrons interacting with column components, such as diaphragms and polepieces, you must take extra care when using intermediate-voltage instruments.

SPURIOUS X-RAYS

We define **spurious X-rays** as those that come from the specimen but are **not** generated by the electron probe in the chosen analysis region.

The standard way to detect stray radiation from the illumination system is to position the focused electron beam to pass through a hole in your specimen and see if you can detect an X-ray spectrum characteristic of the specimen.

If the major peak in the hole-count spectrum contains more than a few percent of the characteristic intensity in the same major peak obtained from a thin area of your specimen, under similar conditions, then the illumination system needs attention.

You can easily determine whether stray electrons or X-rays are the problem, as illustrated in Figure 33.4, which shows spectra obtained on a silver disk specimen (A) and down the hole (B). Ag has a high-energy K_α line at ~23 keV and a low-energy L_α line at ~ 3 keV. With the beam through the hole in the disk, stray X-rays will fluoresce the high-energy peak more efficiently and stray electrons will preferentially excite the low-energy peak. So changes in the K/L ratio from a heavy metal test specimen such as Ag are determinative experiments. The NiO specimen that we described in Chapter 32 is available on Mo support grids and the Mo K/L ratio is equally diagnostic.

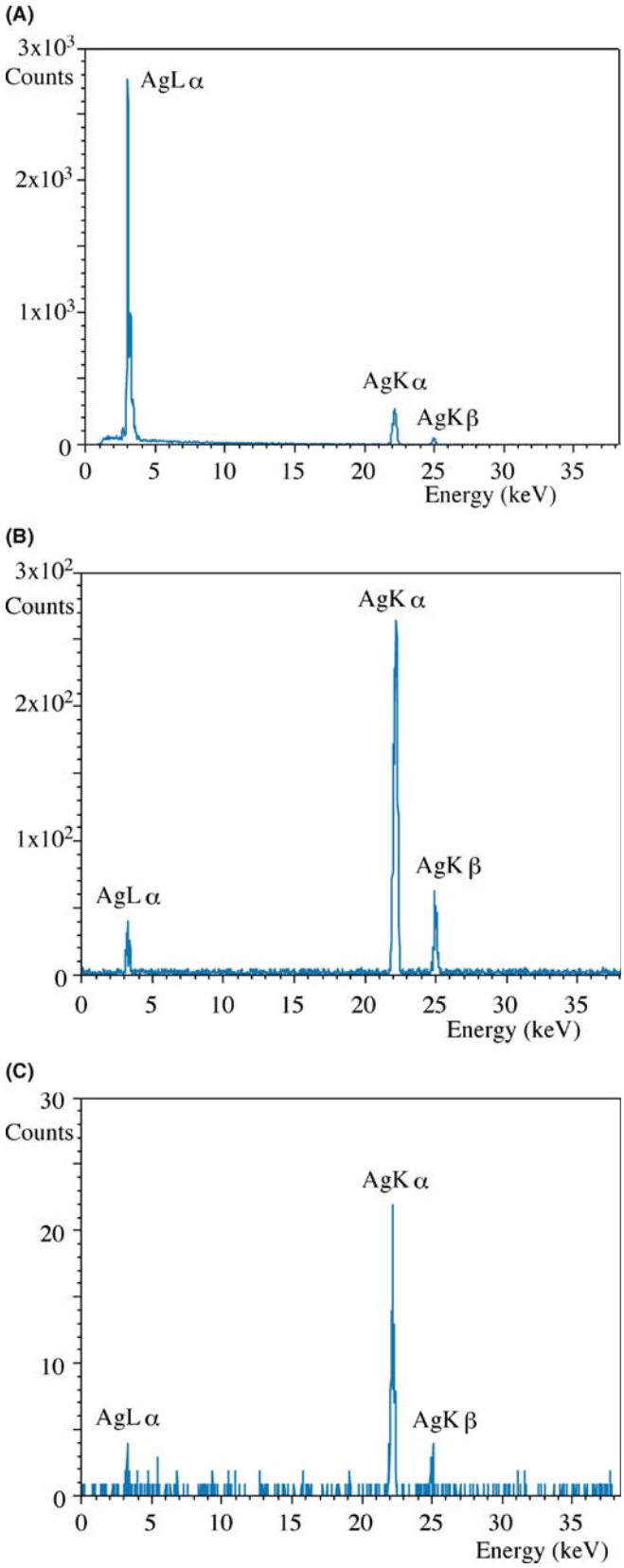


FIGURE 33.4. The 'hole count.' (A) A Ag self-supporting disk produces an electron-characteristic (high L/K ratio) spectrum when struck by the primary beam. (B) Without a thick C2 diaphragm, a reasonably intense Ag spectrum is also detected when the beam is placed down a hole in the specimen. This spectrum has a high K/L ratio which indicates high-energy

THE HOLE COUNT

We cannot reduce spurious radiation sources to zero, so a spectrum, sometimes termed a 'hole count,' is **invariably** obtained in all AEMs if you count for long enough.

Let's consider first the problem of high-energy bremsstrahlung X-rays generated in and penetrating through a standard Pt C2 diaphragm (go back and look at Figure 6.10 to remind yourself why diaphragms stop beam electrons). These X-rays flood your specimen and fluoresce characteristic X-rays from a large area around the analysis region. These stray X-rays would give a high Ag (or Mo) K/L ratio in the spectrum from your test specimen (see Figure 33.4B).

These diaphragms should be a standard fixture in your AEM (check with your lab manager) but they are expensive and they cannot be flame-cleaned in the usual way. When the thick diaphragms do contaminate, you have to discard them, otherwise the contamination itself will become a source of X-rays and will also deviate electrons out of the main beam by charging.

Pt TOP HAT

The solution to this problem is to use very thick (several mm) platinum diaphragms which have a top-hat shape, and a slightly tapered bore to maintain good electron collimation (see Figure 33.4C).

Alternatively, some AEMs incorporate a small diaphragm just above the upper objective lens to shadow the thicker outer regions of the specimen from stray X-rays.

Another way you can minimize the effects of this undesirable bremsstrahlung is to use an evaporated film or window-polished flake on a Be grid rather than a self-supporting disk, but you don't always have this choice. If your specimen is thinner than the path length for fluorescence (a few tens of nm in many cases), spurious X-rays will not be generated. Of course, it may not be possible to prepare such thin specimens, or it may take a great deal of effort, while self-supporting disks are relatively easy and quick to produce; so this isn't a popular suggestion with graduate students. However, with increasing interest in the properties of nano-scale

FIGURE 33.4 (Continued) bremsstrahlung fluorescence. Note that the (spurious) Ag K_α hole-count intensity (B) is ~ 50% of the (real) Ag K_α intensity (A) recorded from the specimen! (C) Use of a thick Pt C2 diaphragm reduces the intensity of the Ag K_α hole count substantially. The K_α intensity in (C) is about 30 times less than in (B). (Note the scale change.)

thin films you may be fortunate enough to be studying such ideal specimens.

For a quantitative, reproducible measure of the hole count, you should use a uniform thin specimen, such as the NiO film we have described. This film should be supported on a bulk material that has a low-energy (<~3 keV) L line and a high-energy (>~15 keV) K line. A thick Mo or Au washer or support grid is ideal. Any high-energy bremsstrahlung X-rays penetrating through the C2 diaphragm will strongly fluoresce the Mo K or Au L line while stray electrons will excite the Mo L or Au M lines preferentially.

With thick diaphragms and thin foils, the remaining stray X-rays will not influence the accuracy of quantification or introduce detectable peaks from elements not in the analysis region. For more details on this test, see Lyman and Ackland. If you don't want to go to the trouble of this test, then the least you should do is measure the in-hole spectrum on your specimen and subtract it from your experimental spectrum if the hole spectrum contains peaks with >1% of the counts in the characteristic peaks that come from your specimen.

RULE OF THUMB

The ratio of Mo K_α or Au L_α intensity detected (when the beam is down the hole) to the Ni K_α intensity obtained with the beam on the specimen should be less than 1%.

Now that we've minimized stray X-rays, let's consider the possibility that all the electrons are not confined to the beam. Stray electrons will generate a hole spectrum with a low K/L ratio (the opposite of Figure 33.4B). If your microscope has a non-beam defining spray diaphragm below the C2 diaphragm it will eliminate such stray electrons without generating unwanted X-rays. Then the main source of poorly collimated electrons is usually the 'tail' of electrons around the non-Gaussian-shaped probe that arises from spherical aberration in the C2 lens, as shown in Figure 33.5 (Cliff and Kenway). The best way to minimize this effect is to image the beam on the TEM screen under the conditions that you will use for analysis, and select the best C2 aperture size to define the probe, as we discussed way back in Section 5.5. It is a simple test to move your probe closer and closer to the edge of your specimen and see when you start generating X-rays. Do this with different size, top-hat C2 diaphragms. If you are fortunate enough to have a C_s corrector in your probe-forming lens then the probe will have much smaller tails, but this is a very expensive solution, not available to all.

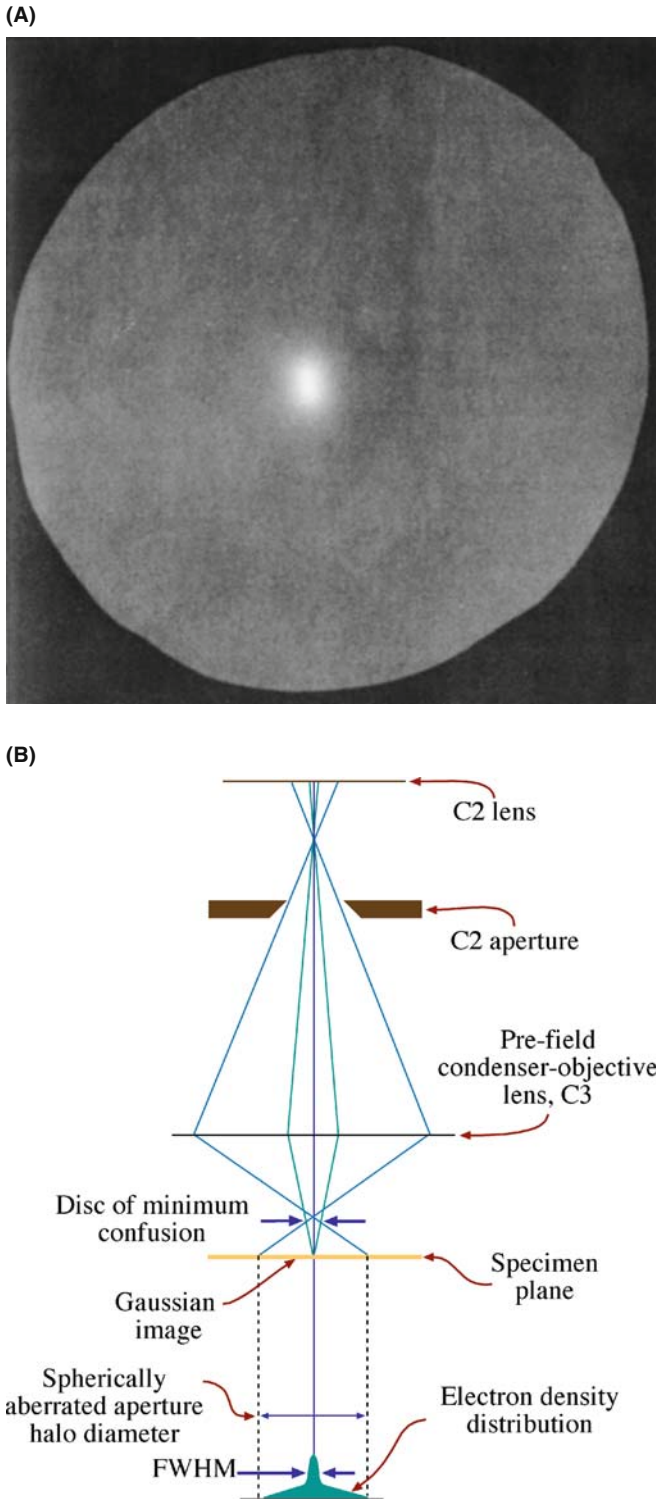


FIGURE 33.5. (A) The shadow of the C2 diaphragm defines the extent of the halo of electrons which excites X-rays well away from the chosen analysis region. (B) Ray diagram showing how the STEM probe formed with a large C2 aperture generates a broad halo around the intense Gaussian-shaped probe. Such beam tailing is the major source of uncollimated electrons and arises due to spherical aberration in the probe-forming lens. Choosing a smaller aperture will limit the tailing but cut down the probe current.

In summary

- Stray X-rays will give a high Ag (or Mo) K/L ratio.
- Stray electrons will give a low Ag (or Mo) K/L ratio.
- Always operate with clean, thick, top-hat C2 diaphragms.
- Use very thin flake specimens or uniform thin films, if possible.
- Always image the electron beam on the TEM screen prior to analysis, to insure that it is well collimated by the C2 aperture.
- Use a C_s-corrected AEM if you can find one.

Now the only significant X-ray source will be the region where you position the probe.

33.3.B Post-specimen Scatter

The electrons interact with the specimen and are scattered elastically or inelastically. It is fortunate for us that the intensity of such scatter from a thin specimen is greatest in the forward direction. So most of the scattered electrons are gathered by the field of the lower objective polepiece and proceed into the imaging system of the AEM away from the XEDS detector. Unfortunately, some electrons are scattered through high enough angles that they strike other parts of your specimen, the support grid, the holder, or the objective lens polepiece or other material in the AEM stage.

SYSTEM X-RAYS
X-rays that come from parts of the AEM other than the specimen.

It is instructive to insert the objective diaphragm during an analysis (just once!) to see the enormous increase in spurious and system X-rays that result. (Actually, you'll do this experiment by mistake anyway.) Usually the X-ray flux is so great that the pulse-processing electronics are saturated, the dead time reaches 100%, and the automatic shutter will activate. However, even when you remove the diaphragm, electrons scattered by the specimen may still create X-rays characteristic of the materials in the holder (brass), the polepiece (Fe and Cu), and the collimator (e.g., Al, W) and any of these X-rays could be picked up by the XEDS detector.

Cu IS EVERYWHERE
Remember the post-specimen scatter will still generate specimen-characteristic X-rays remote from the area of interest, even if a Cu peak is not present.

Furthermore, despite the strong field of the upper objective polepieces in probe-forming STEMs (another good reason to always operate in STEM mode) some backscattered electrons may travel directly into the XEDS detector, generating electron-hole pairs. Other scattered electrons may hit your specimen at some point well away from the area of interest where they will still produce specimen-characteristic (and therefore) spurious, X-rays. All these possibilities are undesirable but unavoidable because, without the beam-specimen interactions that produce this scattering, we would get no information at all from the specimen. Figure 33.6 summarizes all the possible sources of spurious and system X-rays from post-specimen scatter.

In addition to electron scatter, there will be a flux of bremsstrahlung X-rays produced in the specimen. The intensity of these X-rays is also greatest in the forward direction (see gray shaded area in Figure 33.6). Since they possess a full spectrum of energy, the bremsstrahlung will fluoresce some characteristic X-rays from any material that they strike. The easiest way to discern the magnitude of this problem is to use a uniformly thin foil (such as our standard NiO film) on a Cu grid. When you position the probe on the film in the middle of a grid square, many micrometers from any grid bar, the collected spectrum will invariably show a Cu peak arising from the grid, as a

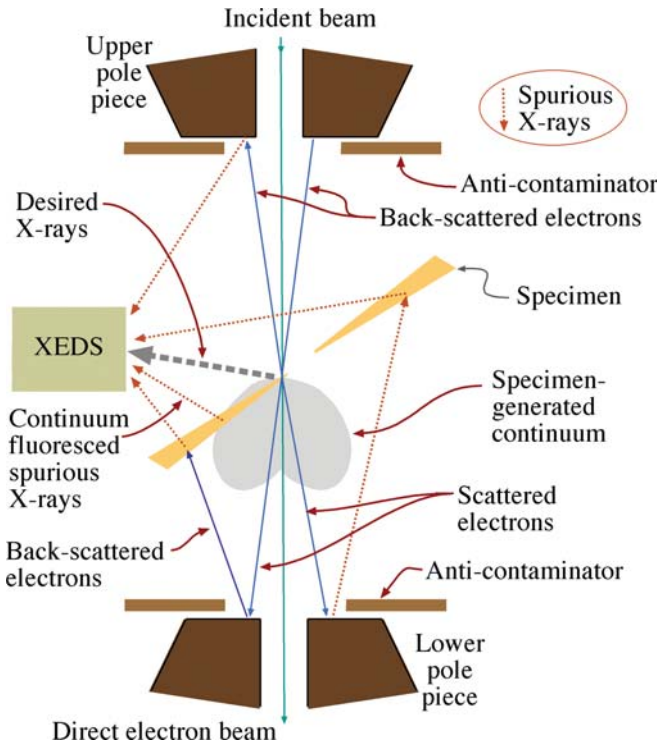


FIGURE 33.6. Sources of system and spurious X-rays generated when the incident beam (green) is scattered by the specimen. BSEs and forward-scattered electrons (blue) excite system X-rays in the stage and spurious X-rays (red dotted lines) elsewhere in the specimen. Bremsstrahlung (gray-shaded region) fluoresces the specimen away from the analysis region (also red dotted lines). The gray dotted line represents the desired X-rays from the analysis region.

result of interactions with electrons or X-rays scattered by the film. An example of this effect in a Cr film is shown in Figure 33.7. The presence of the Cu peak can be removed by using a Be grid, since Be K_α X-rays are much more difficult to detect. However, using Be grids merely removes the observable effect in the spectrum, not the cause.

To minimize the effects of the scattered radiation, you should keep your specimen close to zero tilt (i.e., normal to the beam). If you tilt $< \sim 10^\circ$ then the background intensity is not measurably increased. Under these conditions, your specimen will undergo minimum interaction with both the forward-directed X-rays and any backscattered electrons. Both of these phenomena have only a small horizontal component of intensity. The effects of your specimen interacting with self-generated X-rays will be further reduced by using specimens, such as evaporated films or window-polished flakes rather than self-supporting disks, just as we suggested in the previous section. In self-supporting disks, the bulk regions will interact more strongly with the bremsstrahlung. It is not known what fraction of the post-specimen scatter consists of electrons and what fraction is X-rays, because this will vary with both specimen and microscope conditions. However, there is no evidence to suggest that this X-ray fluorescence limits the accuracy of quantitative analyses (which is at best ±3–5%, as we'll see in Chapter 35).

In addition to keeping your specimen close to zero tilt, you can further reduce the effects of post-specimen scatter by surrounding the specimen with low-Z material. Use of these materials will also remove from the spectrum any characteristic peaks due to the microscope constituents. Be is the best material for this purpose and, as we said right at the beginning of this part of the book, Be specimen holders and Be support grids are essential for X-ray analysis. Ideally, all solid surfaces in the microscope stage region that could be struck by scattered radiation should also be shielded with Be. Unfortunately, such modifications are rarely available commercially.

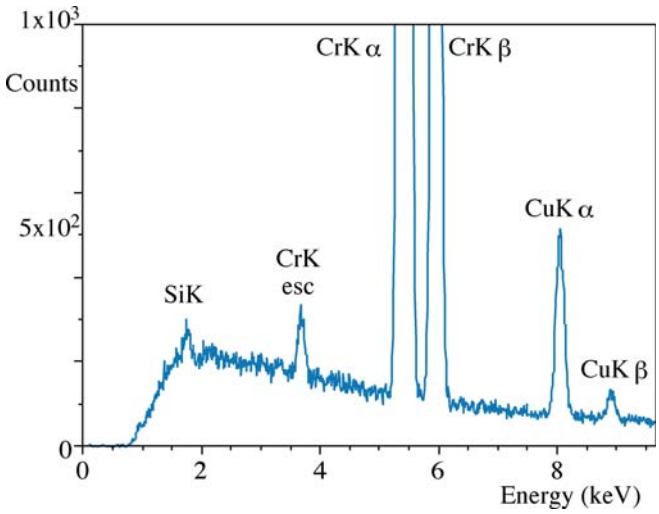


FIGURE 33.7. Cu peaks in a spectrum from a thin Cr film on a Cu grid. Although the beam was many micrometers from the grid, Cu X-rays are excited by electron scatter from within the specimen, and their intensity generally increases with specimen tilt. The Cr escape peak and the Si internal-fluorescence peaks are also visible.

Be

Be oxide is highly toxic if inhaled, so if you have to handle Be grids or other Be components, use gloves and tweezers and don't breathe!

The narrow polepiece gap, required to produce high probe currents, and the cold finger, used to reduce hydrocarbon contamination, both tend to increase the problems associated with post-specimen scatter. In the ideal AEM, the vacuum would be such that a cold finger would not be necessary and the polepiece gap would be chosen to optimize both the detected peak to background ratio and the probe current. When an AEM stage was substantially modified with low-Z material (e.g., by Lyman et al.) a large reduction in bremsstrahlung intensity was reported and X-ray peak to background ratios were produced that are still unmatched by most commercial AEMs. We'll discuss this more in Section 33.5.

You must note, however, that whatever precautions you take, the scattered electrons and X-rays, which are invariably present, result in a specific limitation to X-ray analysis. (See the below box on this.)

SMALL AND LARGE

If you are seeking small amounts (<2%) of element A in a specific region of your specimen, and that same element A is present in large amounts, either elsewhere in your specimen or in the microscope stage, then you ***cannot*** unequivocally identify the presence of that element A in the specific region of your specimen! If you count for long enough, a small peak from A will *invariably* be present in all spectra, just as surely as the Si internal-fluorescence peak from your detector will be present.

Obviously then, you must determine the contributions to the X-ray spectrum from your microscope, and this is best achieved by inserting a low-Z specimen in the beam that generates mainly a bremsstrahlung spectrum, such as an amorphous-carbon film, supported on a Be grid or a pure B foil. If a spectrum is accumulated for a substantial period of time (say 10–20 minutes, or even over lunch), then in addition to the C or B peak (if your XEDS can detect them) the various instrumental

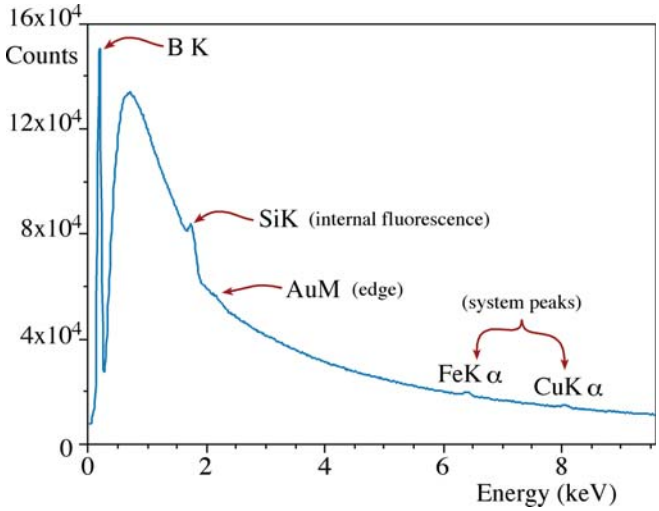


FIGURE 33.8. An XEDS spectrum from high purity boron, showing system peaks. The Si K_α peak and the Au M absorption edges are detector artifacts but the small peaks at 4.6 and 7.5 keV are system peaks from elements in the microscope stage.

contributions to the spectrum should become visible. Such an 'instrument spectrum' (see Figure 33.8) should only exhibit the internal-fluorescence peak and possibly the Au absorption edge from the detector. Any other peaks will be from the TEM itself, assuming the specimen is pure. These peaks will tell you which elements it is *not* possible to seek in small quantities in your specimen because of their presence in your AEM.

We can summarize the methods used to minimize the effects of post-specimen scattering quite simply

- Always remove the objective diaphragm.
- Operate as close to zero tilt as possible.
- Use a Be specimen holder and Be grids.
- Use thin foils, flakes, or films rather than self-supporting disks.

Remember, that even with these precautions, you will still have to look out for artifacts in the spectrum, particularly those from the XEDS system.

33.3.C Coherent Bremsstrahlung

As we noted earlier, the bremsstrahlung spectrum is sometimes referred to as the continuum because the intensity is assumed to be a smooth, slowly varying function of energy. This assumption is perfectly reasonable when the bremsstrahlung is generated in bulk polycrystalline materials by electrons with energies $<\sim$30 keV, such as in a SEM. However, in thin single-crystal specimens illuminated by high-energy electrons, it is possible to generate a bremsstrahlung X-ray spectrum that contains small, Gaussian-shaped peaks known as 'coherent bremsstrahlung' (CB). The phenomenon of CB is well known from high-energy physics experiments, but no one thought it would occur at AEM voltages until it was clearly demonstrated by Reese et al. Figure 33.9A shows a portion of an X-ray spectrum

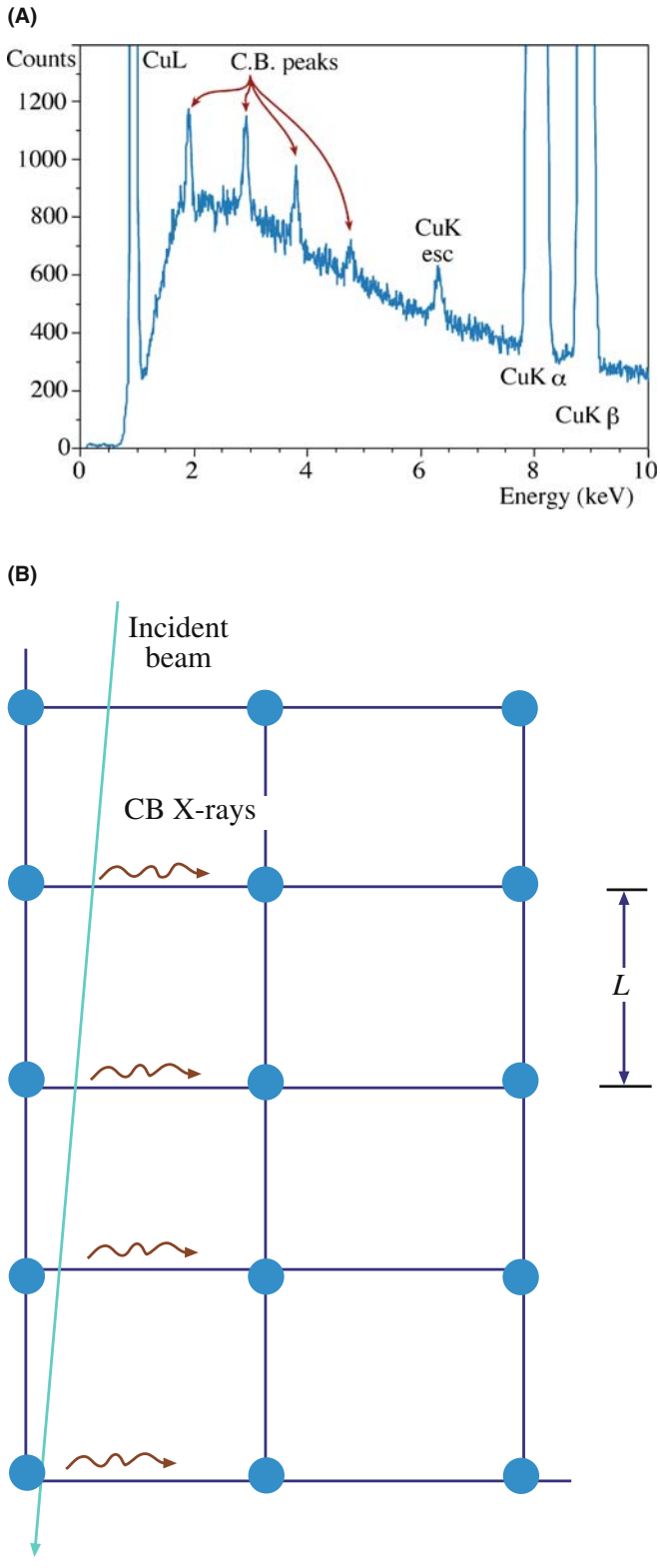


FIGURE 33.9. (A) CB peaks in a spectrum from pure Cu and (B) the schematic generation of CB when the beam passes close to a row of atoms in the specimen.

from a thin foil of pure copper taken at 120 keV. The primary peaks, as expected, are the Cu $K_{\alpha/\beta}$ and the L family of lines. In addition, the escape peak is identified. The other small peaks are the CB peaks. They arise, as shown in Figure 33.9B, by the nature of the coulomb interaction of the beam electrons with the regularly spaced nuclei in the crystal specimen. As the beam electron proceeds through the lattice, close to a row of atoms, each bremsstrahlung-producing event is similar in nature and so the resultant radiation tends to have the same energy. The regular interactions result in X-ray photons of energy E_{CB} given by

$$E_{CB} = \frac{12.4\beta}{L(1 - \beta\cos(90 + \alpha))} \qquad (33.2)$$

where β is the electron velocity (v) divided by the velocity of light (c), L is the real lattice spacing in the beam direction, equal to $1/H$ in a zone-axis orientation (go back and check Section 21.3.B), and α is the detector take-off angle. More than one CB peak arises because different Laue zones give different values of L. The CB peak intensity seems greatest when the beam is close to a low-index zone axis, and these conditions should be avoided if possible. Operating with a convergent beam reduces the intensity of the CB peaks and this will be helped with a C_s-corrected AEM since larger convergence angles can be used to give more probe current without degrading the probe size.

CB PEAKS

You may mistakenly identify these CB peaks as characteristic peaks from a small amount of some element in the specimen, but fortunately, you can easily distinguish CB peaks from characteristic peaks.

Unfortunately, you can't remove the CB effects entirely, even by operating far from a major zone axis, since some residual peaks are invariably detectable if you count for long enough.

As predicted by equation 33.2, the CB peaks will move depending on both the accelerating voltage (which will alter v and hence β) and the specimen orientation, which will change the value of L. Of course, characteristic peaks show no such behavior and are dependent only on the elements present in your specimen. While CB peaks are a problem, you should only be concerned if you are seeking to detect a small amount of specific element in your specimen. More about this problem in the next chapter.

33.4 MEASURING THE QUALITY OF THE XEDS-AEM INTERFACE

In the end, what we need is some measure of how well our XEDS system is working and to be able to compare it with values from other systems. There are two ways to do this, both of which use a thin film, such as our standard NiO (although in these examples, we'll use Cr since that was how the original experiments were done. The principles are the same).

33.4.A Peak-to-Background Ratio

The first test of how well your XEDS is interfaced to your TEM is to measure the peak to background (P/B) ratio in the film.

FIORI P/B

There are several definitions of P/B ratio, but the best one, termed the Fiori definition, is shown in Figure 33.10A.

For the Ni K_α peak, you should integrate the peak intensity from 7.1 to 7.8 keV and divide this by the average background intensity in a 10 eV window (i.e., one or two channels depending on the display resolution) under the peak. In the Cr thin-film example shown in Figure 33.10A, the Cr K_α peak is summed from 5.0 to 5.7 keV. In a well-constructed AEM, the P/B ratio will increase with keV. The P/B values shown in Figure 33.10B (measured again on a Cr film rather than a NiO film (Zemyan and Williams)) should be achievable in any modern AEM. This value is an important test of the XEDS-AEM interface and the design of the stage. Higher is better!

33.4.B Efficiency of the XEDS System

The relative detector efficiency is a measure of how many X-ray counts per second (cps) are collected, detected and processed by the XEDS system. This is very important because of the overwhelming need to gather the most X-ray counts possible, given all that we've described about the inherent inefficiencies of X-ray generation and detection in XEDS of thin films.

In a fixed live time the detector efficiency will be affected by the specimen thickness, the probe current, and the solid angle of collection of the detector. So you should use the standard NiO specimen again, to fix the thickness variable and measure the probe current with a Faraday cup, as we've mentioned several times before (this gives cps/nA). Last, you need to factor in the collection angle given by the XEDS manufacturer (which is in fact calculated, not measured) to give the

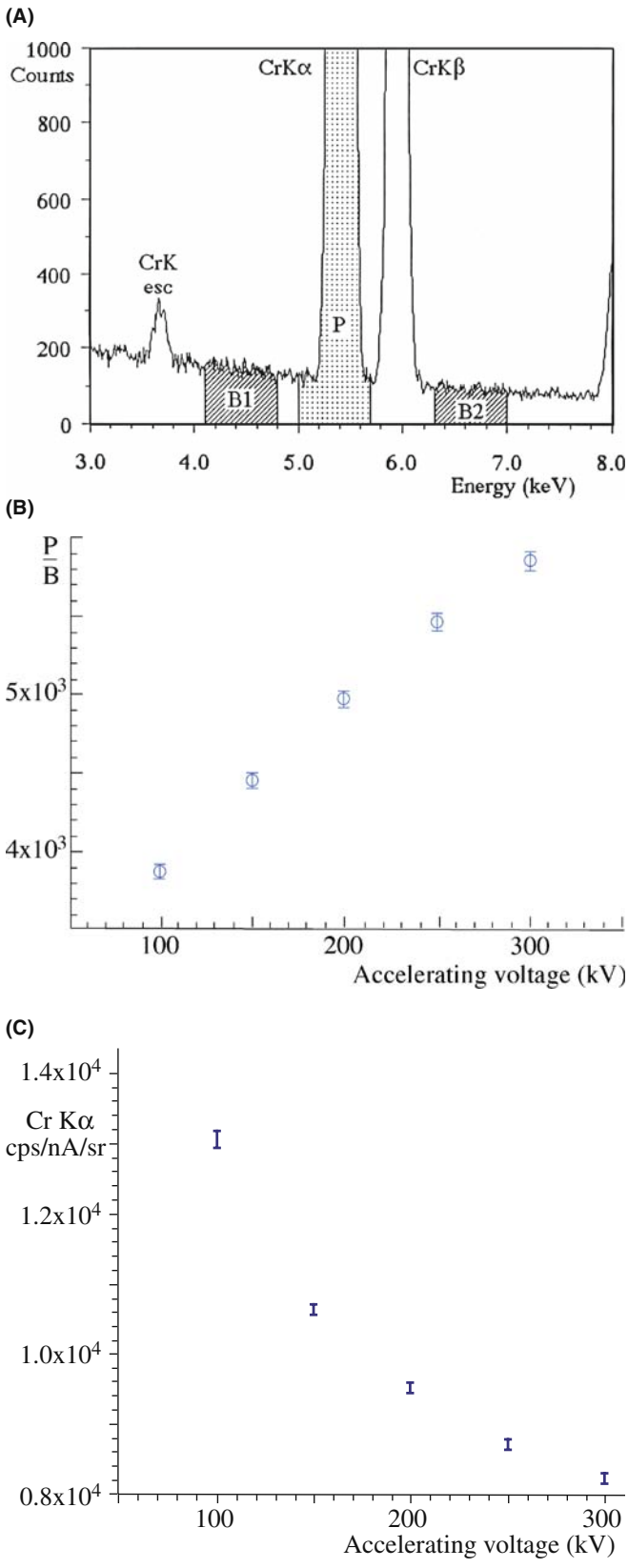


FIGURE 33.10. (A) Definition of the Fiori P/B ratio in a Cr thin film. (B) Change in the Fiori P/B as a function of kV for a high-performance 300-kV FEG AEM. (C) Decrease in the cps/nA/sr from a thin Cr film demonstrating decreasing detector collection and processing efficiency with increasing kV.

best figure of merit in terms of cps/nA/sr as described by Zemyan and Williams. Typical values are shown (again for Cr) in Figure 33.10C and, as you can see, the efficiency decreases with increasing kV because of the decrease in ionization cross section. But at any given kV, a higher number indicates higher efficiency.

It is possible to determine both the P/B ratio and the efficiency from direct spectral measurement and software to do this is described in the companion text.

33.5 ACQUIRING X-RAY SPECTRA

There are many commercial systems for acquiring spectra and images and many of these differ in terms of the file formats used within the computer system. Consequently, you might find it difficult to exchange spectra between different XEDS systems (e.g., for comparison of data gathered on two different AEMs). There has been a concerted attempt by the various manufacturers and professional societies to rectify this by creating the so-called 'EMSA/MAS standard file format' (Egerton et al.). You should check to ensure that your XEDS system supports this format.

So from what we've told you so far, you now know what information is likely to be in your spectra; which peaks might be real, which are most likely artifacts, and which ones you can and can't interpret as coming from the analysis region. Now we can concentrate on how best to actually gather the spectra before proceeding with analyses.

33.5.A Spot Mode

The standard way of gathering spectra, from the earliest days of AEM (and preceded by SEM/EPMA) was simply to use the beam deflectors to position the spot on a feature in your image and switch on the XEDS. We call this 'spot mode.' You could do this in a TEM by condensing the beam down with the C2 lens and adjusting C1 iteratively until the beam is small enough to interact only with the feature you wanted to analyze, such as a precipitate. In STEM spot mode you simply stop the beam from scanning and move the probe onto the feature, hoping to get it in the right position before the image fades from the STEM screen. In either case you also have to hope that both the probe position and the feature in your image stay stationary for long enough to gather a spectrum with sufficient counts to give you the composition information you need. On vintage AEMs, this method ensures that any carbon contamination buildup precisely buries the feature of interest! However, that same contamination spot would also show you if the beam or specimen had drifted during the analysis.

So spot mode is time consuming, the statistical confidence is appalling, and the pre-selection of which feature(s) to analyze introduces serious operator bias (and if you're not careful, all that you might analyze is an artifact of the specimen preparation rather than a key microstructural feature). Despite all this, it is still common to see publications in which this method is precisely how the analysis was performed and, in terms of giving a gross indication of local variations in chemistry, say for a complex extraction replica containing multiple phases, it certainly has some benefit, as shown in Figure 33.11. So, be careful when you build on someone else's results.

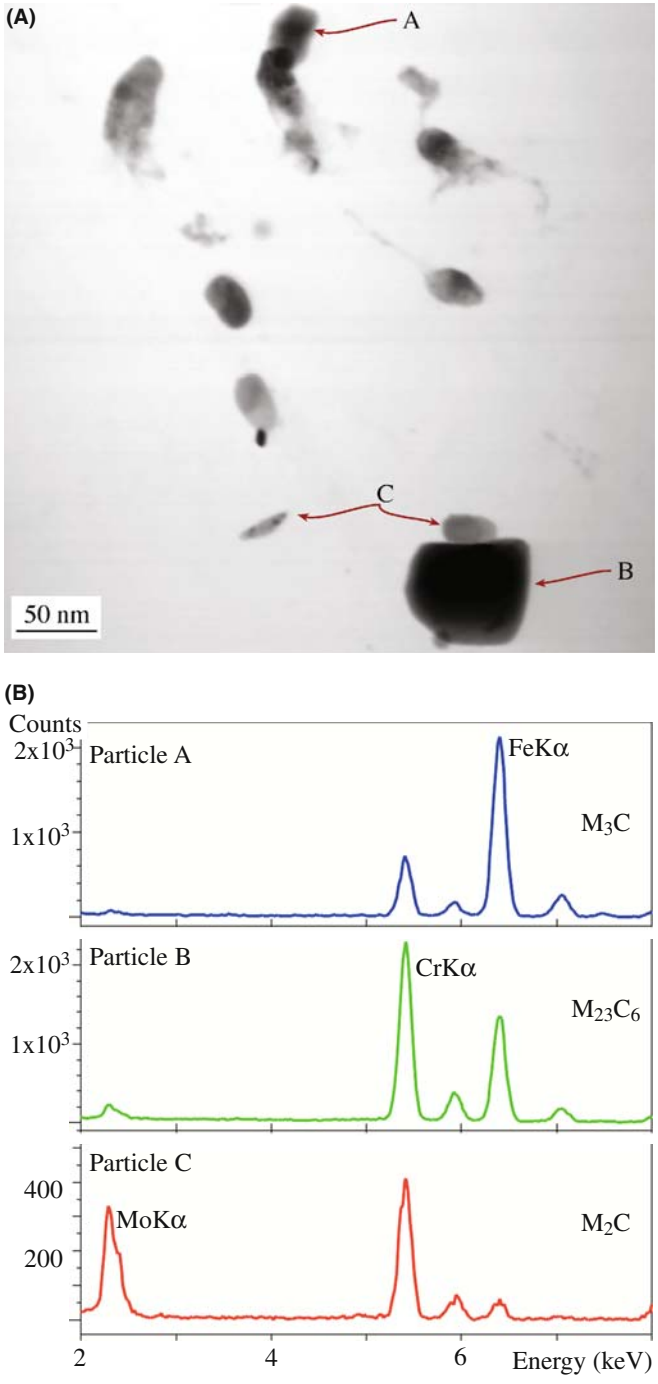


FIGURE 33.11. (A) STEM image of different carbide particles on an extraction replica from stainless steel specimen. (B) Multiple spectra taken from an array of points in the micrograph illustrating the variable chemistry of the different carbides.

33.5.B Spectrum-Line Profiles

A variation of the spot mode of analysis is to take a series of spot analyses across a linear feature of interest in your specimen, such as a grain boundary or interphase interface, as shown in Figure 33.12A, and build up a set of spectra which, when analyzed, will reveal the composition profile across the interface. The information can be displayed as a set of superimposed spectra, termed a spectrum-line profile, as shown in Figure 33.12B, and the changes in the characteristic-peak intensities reflect significant composition changes. Since any kind of interface is a major planar defect in an engineering material, this approach at least removes some of the operator bias of the single-spot mode. However, line profiles still only reveal the composition across a single point on the boundary and many such profiles are required if you are to determine composition variations along the defect. So this approach is also tedious.

The solution to the limitations of both spot and line profiles approaches is to gather parts of spectra or preferably full spectra at every pixel in the STEM image, producing compositional images or maps.

This method is by far the best way to gather X-ray information with some semblance of statistical significance and without operator bias, so we'll now spend the rest of the chapter describing this method.

33.6 ACQUIRING X-RAY IMAGES

Mapping, or compositional imaging, used to be rarely used for analysis in the AEM because of the low X-ray count rate due to the small probe currents and collection angles. The overall inefficiency of the process meant that to gather a map with sufficient X-ray counts to be able to draw any conclusions about the variations in chemistry in the thin foil, you would have to scan the area for many hours. During this time, specimen drift, contamination or damage would occur and the resultant information would be compromised. However, with the developments of intermediate-voltage FEG sources, detectors with higher collection efficiency, drift-correction software, cleaner stage vacuums, and, most recently, C_s correctors, it is now possible, on the best AEMs, to gather quantitative X-ray maps in a matter of minutes. So it's worth looking at the various imaging options which have been developed, along with these improvements in AEM design and computer technology. In all of these approaches it will take significant

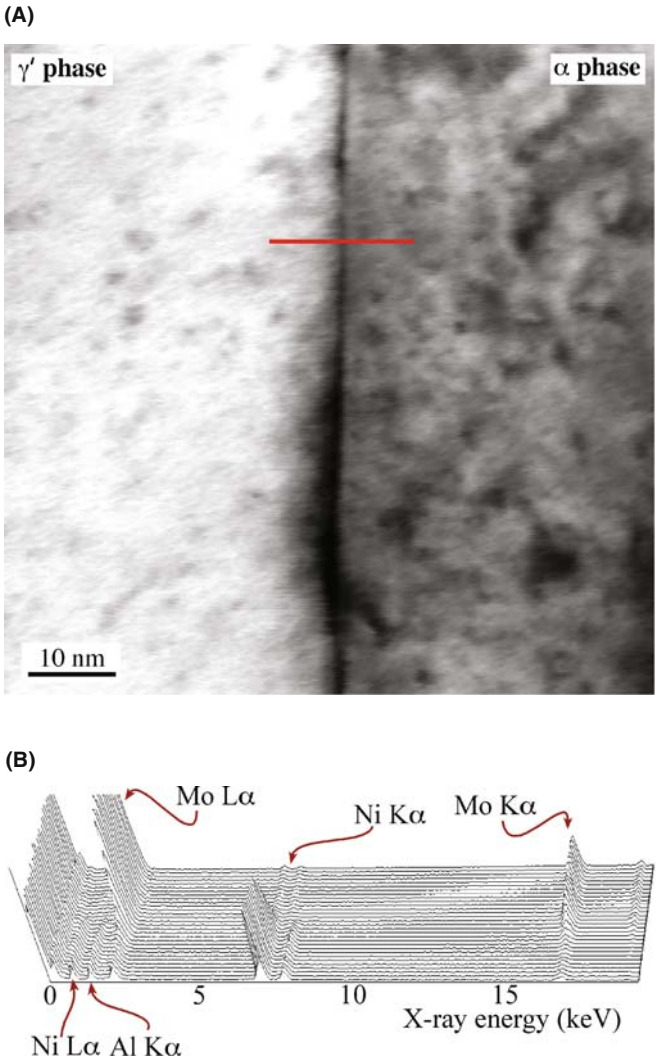


FIGURE 33.12. (A) STEM image of an α/γ' interphase interface in a Ni-base superalloy with a line showing from where X-ray spectra were obtained. (B) A spectrum-line profile taken along the line in (A) showing clearly the change in Ni and Mo composition across the interface.

time to acquire a map and the biggest danger is that the specimen will drift during that time or the beam current will change (particularly, if you have a cold-FEG source). So learn about the drift-correction software that should be available in any standard commercial XEDS software package. Unless you can measure the probe current *on the fly* in a cold-FEG AEM, a Schottky FEG is recommended for mapping.

Analysis of bulk specimens in the SEM or EPMA is not limited by X-ray counts so the X-ray imaging techniques we'll discuss were generally pioneered on the SEM/EPMA. Likewise, as we'll see in Chapter 37 and beyond, EELS measurements have many millions or even billions of counts and so EELS mapping of thin films was used long before thin-foil X-ray analysts were able to benefit. But in this book, we'll rewrite history a little and talk first about all these techniques in reference to X-ray mapping.

33.6.A Analog Dot Mapping

Dot maps are the original method of acquiring qualitative X-ray images and, somewhat surprisingly, were first created more than half a century ago by Cosslett and Duncumb and progress since then has been reviewed by Friel and Lyman. The approach is simple: you select a specific energy (or wavelength) channel (or a window (range) of channels) in your X-ray spectrum, scan the beam across the area of interest and when the X-ray detector registers an X-ray of the selected energy (range), record and display it. So as the beam scans, the display intensity builds up and the changes in intensity reflect the changes in the number of X-rays detected. For example, if you select the Pd L_α peak channel or a window covering the peak, then regions showing lots of dots on the screen are high in Pd, as shown in Figure 33.13 (we'll go into more detail about how to do this in Section 34.7). This qualitative, analog

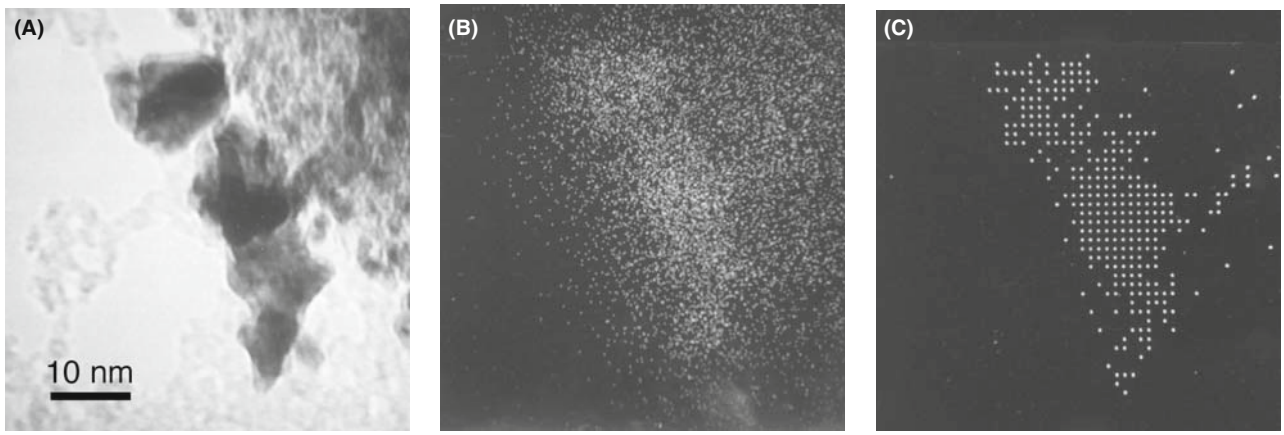


FIGURE 33.13. (A) STEM BF image of a Pd catalyst particle on a support film. (B) Analog dot map using the Pd L_α signal. Note the correlation between thicker (darker) regions of the specimen and increased Pd signal. (C) Early digital map taken with Pd L_α signal but subtracting the background signal.

approach is not directly quantifiable since the background can't be removed, unless you simultaneously gather a dot map from a bremsstrahlung window adjacent to the peak window and subtract the one from the other (more about this in Chapter 35). Furthermore, in thin foils, changes in thickness will also produce changes in intensity in both the peak and background maps and this is clear in Figure 33.13B. It is possible to refine this approach by gathering multiple X-ray maps, assigning colors to different X-rays, and overlaying the maps to give an indication of relative composition changes. Again, this process is *not* quantitative.

33.6.B Digital Mapping

As computers became more powerful, it was possible to collect X-rays from multiple channels or windows and thus acquire several maps simultaneously. If one or more of the maps was the bremsstrahlung intensity then you could produce quantitative maps, on the fly, as the data accumulates. The first success in digitizing thin-foil X-ray maps was carried out by Hunneyball et al. almost 30 years ago. It was only possible to build up 128×128 pixel images with 256 X-ray counts at each pixel using a 100-nm probe with 5 nA of probe current over 200s. But the maps were fully quantitative, removed the effects of foil-thickness variation, and revealed relatively small composition variations around GBs in aged Al alloys, as shown in Figure 33.14A. Because of the central role of the computer in the acquisition, this approach has the advantage of post-acquisition processing and comparison of different quantification routines.

Also after acquisition it was possible to use such processing techniques as false coloring, computerized image overlays, scatter diagrams, etc. (Bright and Newbury). All these advantages combine to make digital imaging a most attractive approach for displaying XEDS data. This process really came of age in the mid-1990s with the availability of faster computers, improved data storage, intermediate-voltage FEG sources, reasonable X-ray collection angles, and stable AEMs permitting long collection times without drift. An example from such an AEM is shown in Figure 33.14B which is 256×256 pixels and was acquired using a 1-nm probe with 0.9 nA of current for 5400 s. Comparing Figure 33.14A and B is instructive and

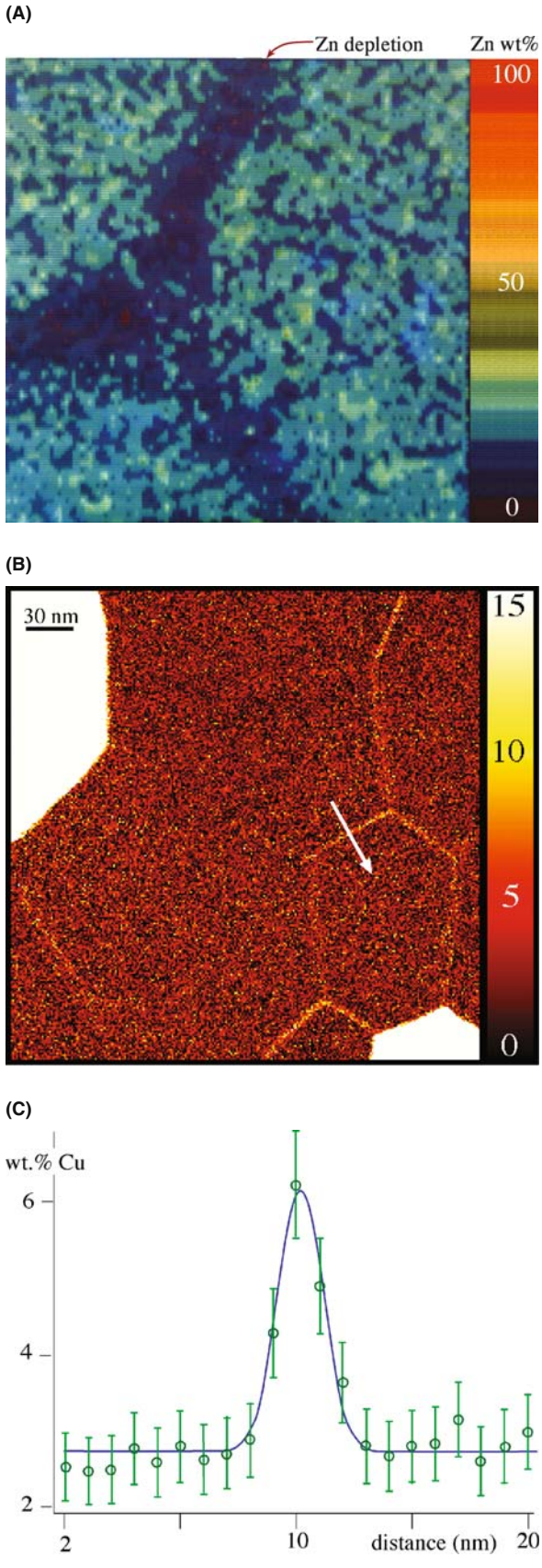


FIGURE 33.14. (A) The first quantitative digital map (128 × 128 pixels) obtained in an old thermionic-source AEM generating a low count rate. The image from a thin foil of Al-Zn shows depletion of Zn around a triple point (compare the colors in the map with the quantitative look-up 'table' on the right side of the image). (B) More recent digital map from a 300-kV FEG AEM showing enrichment of Al at GBs in an electro-migrated specimen of Al-4% Cu. The bright regions are $CuAl_2$ intermetallics. In both (A) and (B) quantification was achieved by subtracting the bremsstrahlung intensity at each pixel. (C) Quantitative Cu line profile taken across a GB indicated in (B), by arrow.

shows the enormous progress in both the quality of mapping and the spatial resolution. Once you have a digital map such as this you can go into the map and extract quantitative data from any region, e.g., the line profile taken across the GB in Figure 33.14C. However, you still have to select the window(s) or peak(s) in the spectrum that you map, thus introducing your own bias in terms of what you expect to be in the specimen. Furthermore, once the map is recorded and stored, you cannot return to re-check the data, or map another element from the same area, since all the other information in the spectrum was lost. Likewise, unless you store the image from the mapped area, it too cannot be re-examined. So you have to get everything right, collect all the X-ray images you need, as well as the background spectra to permit subsequent quantification. Gathering multiple maps was limited by the computer memory, but that is no longer a problem since memory is cheap. Now, as we discuss below, you don't even have to bother about preselecting which X-rays you want to map, you just gather them all and later on decide which you want to image. How do we do this? It's called 'spectrum imaging.'

SPECTRUM IMAGING

SI is *the* preferred method for X-ray (and EELS) mapping.

33.6.C Spectrum Imaging (SI)

As the term implies, SI collects a full spectrum at every pixel in the digital image (so you can only do this in STEM mode (although there are analog versions in energy-filtered TEM, as we'll see in Chapter 39)). The result of the SI process is a 3D data cube, as shown in Figure 33.15A, with the electron image constituting the x–y plane and all the XEDS spectra in the z direction. The SI term was first used for EELS in the late 1980s and we'll mention this topic again in Chapter 37. Only much later did SI became feasible for X-ray mapping although it is now common enough to be used in materials problem solving (e.g., Wittig et al.). On a historical note, you should be aware that SI methods have been practiced in other fields, such as radio astronomy for several decades and, indeed, Legge and Hammond took the output of their EDS and WDS spectrometers and synchronized the

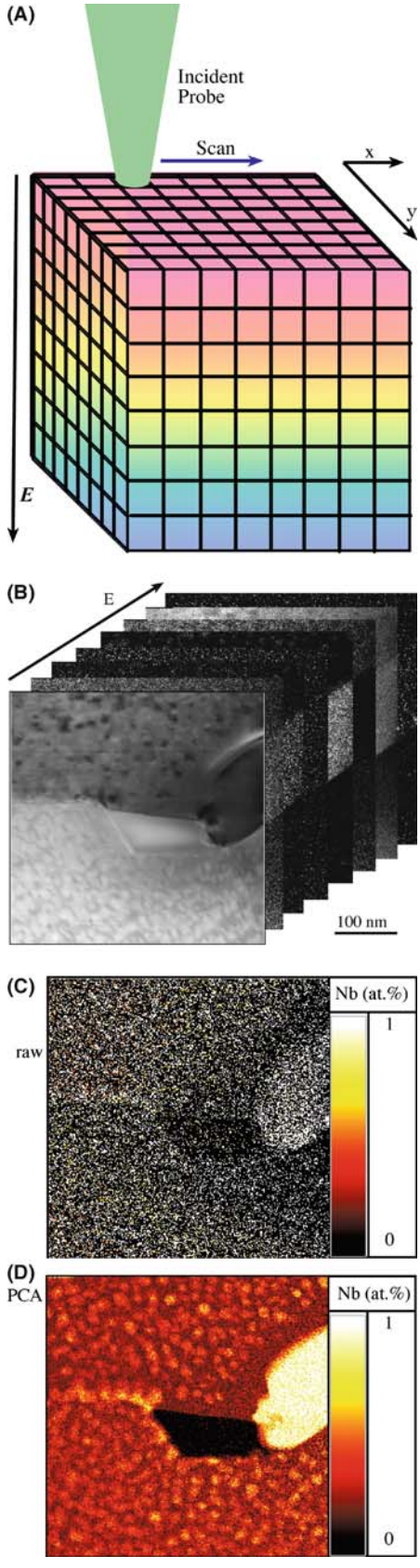


FIGURE 33.15. (A) Schematic spectrum-image data cube showing how as the beam stops at each pixel in the x–y plane as it is creating a STEM image, a full X-ray spectrum is gathered at each pixel. The different colors at different energy (z) values indicate different signals from different elements that appear at different energies. (B) A series of X-ray maps of a GB region in a Ni-base superalloy taken at specific energies from an SI data cube. (C) An example of a map from a single channel in the X-ray spectrum (i.e., a single image plane) in the SI in (B) coinciding with the Nb K_α peak energy. (D) The application of multivariate statistical analysis and principal-component analysis to remove noise and enhance the Nb signal.

detectors' output pulses with the position of the beam 30 years ago. They collected the data on magnetic tape, and later reconstructed the data into a 3D file, so it can be argued that this was the first demonstration of the SI concept, at least in electron-beam instruments, so we aren't really revising the historical record!

The beauty of the SI technique is that, once you have this data cube stored, you can go back to it at any time, recheck the data, and re-do any analysis, search for other spectral features that you might not originally have thought present or important, look at different images at different energies, and yet always have the original image and spectra at your disposal.

If you think about the diagram in Figure 33.15A, you'll see that there are many ways to slice and dice the data cube which reproduce all the other methods of analysis that we've described. If you select a single pixel in the x–y (image) plane, then there is a full spectrum attached to it and so, if you wish, you can select a set of individual pixels and do as many spot analyses as you wish. Likewise, you can select a line of pixels in the image, effectively slicing the cube along the x–z, y–z, or some combination of these directions and thus produce spectrum-line profiles. You can (and this is where it gets really useful) slice the cube at any plane in the z (labeled E for energy) direction and, at each plane, you'll get a different image consisting of X-rays of that particular energy. You can add planes, subtract planes, and sum image pixels only in certain features (e.g., a strangely shaped precipitate or along a boundary plane). You can also envisage gathering spectra at certain pixels as a function of time (chrono-spectrometry?) and there are surely other options which you can think of. Now if you do the math and consider a 1 k $\times$ 1 k pixel image with a 2048-channel spectrum at each pixel, you'll find that the data cube is about 2 GB in size. In fact, in gathering such an enormous amount of data, the bigger challenge is to find ways both to search this enormous database efficiently and extract meaningful data. Various advanced software methods, such as multivariate statistical analysis, principal-component analysis, and maximum-pixel spectrum analysis are available to remove noise and extract rare events from the data cube. Such processes are closely related to the parameterizing of HRTEM images discussed in Chapter 31, and are discussed in far more detail in the companion text.

Just to give you a hint of what can be done, Figure 33.15B shows a series of X-ray images cutting the x–y plane at different energy values, displayed 'behind' the original STEM ($E = 0$) image. If we select a single x–y plane such as that at the Nb K_α line energy (Figure 33.15C), that image is very noisy because of the limited signal captured in that one plane. However, we can use sophisticated software to remove all the noise components in the plane and also add in all the other Nb signals from other energy planes to produce the image showing the Nb distribution (Figure 33.15D). Before you think that such an extraordinary change is unreasonable and that information is being created where it did not exist, compare the processed image (D) with the original STEM electron image (the top slice in B). Clearly the small matrix precipitates that are imaged in (A) are also mapped in (D). Extracting information like this from the SI data cube tells you how much more powerful the SI approach is compared with spot or line-profile analyses. Now it is routine to gather both XES and EELS SI simultaneously, thus completely optimizing the acquisition of analytical data.

33.6.D Position-Tagged Spectrometry (PTS)

PTS is a specific commercial version of generic SI, developed by Mott and Friel at Princeton Gamma Tech (PGT, now Bruker AXS). PTS eliminates the conflict between your wishing to view full X-ray images quickly versus the analytical advantages of having complete spectra saved at each pixel (which even on the best AEMs still takes 30–45 minutes if you want to quantify the data). In PTS, the beam is scanned rapidly relative to traditional mapping, and the X-rays are counted in the analysis computer, preserving both spatial information from the image and spectral information. Sophisticated processing software can be used to interrogate the data *during* acquisition, which is not possible in conventional SI. Alternatively, this software can be used after the complete spectrum is stored, as with conventional SI, where you acquire the full spectrum at a single pixel and then move to the next pixel and gather another spectrum. PTS also permits relatively easy monitoring of such phenomena as specimen drift, contamination, or damage during the acquisition.

CHAPTER SUMMARY

XEDS in the AEM is a challenge because the detection/processing system creates artifact peaks in the spectrum and X-rays are generated and detected from sources in the AEM other than the region of your specimen where you put the beam. Nevertheless, there are well-defined precautions you can take so that you are sure that the artifacts, the spurious and system X-rays are minimized and that your subsequent interpretation and quantification are not compromised. There are also several standard tests you can carry out to compare your AEM system performance with other instruments.

In summary, to understand your AEM and to acquire meaningful spectra, you should

- Buy your own NiO standard thin film on a Mo grid.
- Test your XEDS to determine the artifact peaks that it produces.
- Gather a spectrum down a hole in the NiO film to see what spurious X-rays the AEM illumination system produces. Use top-hat C2 apertures at all times.
- Gather an XEDS spectrum from a light-element film to see what system peaks your AEM introduces into the spectrum. Use thin foils, flakes, or films rather than self-supporting disks if possible.
- Be aware of CB if you count for a long time to detect small peaks from trace elements.
- Image the electron beam on the TEM screen to ensure that it is Gaussian.
- Always remove the objective diaphragm.
- Operate with the specimen as close to zero tilt as possible.

Check that

- The hole count is <1% of your experimental spectrum under the same operating conditions.
- The P/B ratio and the detector efficiency data are acceptable and do not change with time.
- If you're looking for characteristic peaks from trace or minor elements in your specimen, understand that such peaks are much more likely to be confused with artifacts or system peaks (much more about this in the next chapter).

Once you are sure of what's in your spectrum, decide whether you want to do quick and dirty point analyses of some features in your image or, if you are confident of the importance of the area that you want to analyze completely, gather X-ray maps or, ideally, acquire a SI data cube which will really give you the full picture of your specimen composition.

GENERAL REFERENCES

Garrett-Reed, AJ and Bell, DC 2003 *Energy-dispersive X-ray Analysis in the Electron Microscope* Bios (Royal Microsc. Soc.) Oxford UK. Different descriptions of many of the issues discussed in this chapter.

Lyman, CE (Ed.) 2006 Microscopy and Microanalysis **12** 1. A commemorative edition celebrating 50 years of X-ray mapping; great history and examples of good practice.

Williams, DB, Goldstein, JI and Newbury, DE (Eds.) 1995 *X-Ray Spectrometry in Electron Beam Instruments* Plenum Press New York. Still the best available source of background information on XEDS hardware and software, even though increasingly dated in content.

SPECIFIC REFERENCES

Allard, LF and Blake, DF 1982 *The Practice of Modifying an Analytical Electron Microscope to Produce Clean X-ray Spectra* in *Microbeam Analysis-1982* 8–20 Ed. KFJ Heinrich San Francisco Press San Francisco CA. Early review of system artifacts.

Bright, DS and Newbury, DE 1991 *Concentration Histogram Imaging* Analytical Chemistry **63** 243A–250A. Processing maps for appearance and more.

Cliff, G and Kenway, PB 1982 *The Effects of Spherical Aberration in Probe-forming Lenses on Probe Size and Image Resolution* in *Microbeam Analysis-1982* 107–110 Ed. KFJ Heinrich San Francisco Press San Francisco CA. The tail on the probe.

Cosslett, VE and Duncumb, P 1956 *Microanalysis by a Flying-spot X-ray Method* Nature **177** 1172–1173. First dot maps by XEDS—more than 50 years ago!!

Egerton, RF, Fiori, CE, Hunt, JA, Isaacson, MS, Kirkland EJ and Zaluzec, NJ 1991 *EMSA/MAS Standard File Format for Spectral Data Exchange* EMSA Bulletin **21** 35–41. The EMSA/MAS standard file format.

Fiori, CE, Swyt, CR and Ellis, JR 1982 *The Theoretical Characteristic to Continuum Ratio in Energy Dispersive Analysis in the Analytical Electron Microscope* in *Microbeam Analysis-1982*, 57–71 Ed. KFJ Heinrich San Francisco Press San Francisco CA. The Fiori P/B definition.

Fiori CE and Swyt, CR 1994 *Desk Top Spectrum Analyzer (DTSA),* U.S. Patent 5 299 138. The original description of the essential software.

Friel JJ and Lyman CE 2006 *Tutorial Review: X-ray Mapping in Electron-Beam Instruments* Microscopy and Microanalysis **12** 2–25. A good place to start before you map.
Hunneyball, P D, Jacobs, MH and Law, TJ 1981 *Digital X-ray Mapping from Thin Foils* in *Quantitative Microanalysis with High Spatial Resolution* 195–202 Eds. GW Lorimer, MH Jacobs and P Doig, The Metals Society London. First digital XEDS maps.
Legge, GJF and Hammond, I 1979 *Total Quantitative Recording of Elemental Maps and Spectra with a Scanning Microprobe* J. Microsc. **117** 201–210. A little history.
Lyman, CE and Ackland, DW 1991 *The Standard Hole Count Test: a Progress Report* in *Microbeam Analysis-1991*, 720–721 Ed. DG Howitt San Francisco Press San Francisco CA. Measuring hole count.
Lyman, C.E, Goldstein, JI, Williams, DB, Ackland, DW, von Harrach, S, Nicholls, AW and Statham, P.J 1994 *High Performance X-ray Detection in a New Analytical Electron Microscope* J. Microsc. **176** 85–98. Modifying the stage to use only low-Z materials.
Mott RB and Friel, JJ 1999 *Saving the Photons: Mapping X-rays by Position-Tagged Spectrometry* J. Microsc. **193** 2–14. The first commercial SI software
Newbury DE 1995 *Artifacts in Energy Dispersive X-Ray Spectrometry in Electron Beam Instruments; Are Things Getting Any Better?* in *X-Ray Spectrometry in Electron Beam Instruments* DB Williams, JI Goldstein and DE Newbury (Eds.) 167–201 Plenum Press New York. Discussion of artifacts and how digital imaging changes them.
Newbury, DE 2005 *X-ray Spectrometry and Spectrum Image Mapping at Output Count Rates above 100 kHz with a Silicon Drift Detector on a Scanning Electron Microscope* Scanning **27** 227–239. SDD performance.
Reese, GM, Spence, JCH and Yamamoto, N 1984 *Coherent Bremsstrahlung from Kilovolt Electrons in Zone Axis Orientations* Phil. Mag. **A 49** 697–716. Early demonstration of CB.
Williams, DB and Goldstein, JI 1981 *Artifacts Encountered in Energy Dispersive X-ray Spectrometry in the Analytical Electron Microscopy* in *Energy Dispersive X-ray Spectrometry* 341–349 Eds. KFJ Heinrich, DE Newbury, RL Myklebust and CE Fiori NBS Special Publication 604 U.S. Department of Commerce/NBS Washington D.C. Early review of system artifacts.
Wittig, JE, Al-Sharaba, JF, Doerner, M, Bian, X, Bentley, J and Evans, ND 2003 *Influence of Microstructure on the Chemical Inhomogeneities in Nanostructured Longitudinal Magnetic Recording Media* Scripta Mater. **48** 943–948. Early example of SI.
Zemyan, SM and Williams, DB 1995 *Characterizing an Energy-Dispersive Spectrometer on an Analytical Electron Microscope* in *X-Ray Spectrometry in Electron Beam Instruments* DB Williams, JI Goldstein and DE Newbury (Eds.) 203–219 Plenum Press New York. Basic understanding of XEDS performance.

URLs

1) www.cstl.nist.gov/div837/Division/outputs/DTSA/DTSA.htm To download DTSA

SELF-ASSESSMENT QUESTIONS

Q33.1 What's the difference between spurious X-rays and system X-rays?
Q33.2 What are the best steps you can take to minimize spurious X-rays?
Q33.3 What causes incomplete charge collection (ICC) in an XEDS?
Q33.4 Name three common artifacts in the spectra generated by XEDS systems.
Q33.5 Which of the various artifacts should you be particularly aware of while operating your AEM and why?
Q33.6 Distinguish the sources of desired and undesired radiation impinging on the detector.
Q33.7 How can you find out what X-rays are generated by your AEM-EDS system rather than by your specimen?
Q33.8 Why is it important to know what your system X-rays are?
Q33.9 What's the danger of not minimizing any spurious X-rays?
Q33.10 Why would you not use a Cu grid to analyze diffusion profiles in brass using XEDS in the AEM?
Q33.11 Why is it challenging to detect trace (~0.1 wt%) amounts of Si unambiguously in an Fe alloy using XEDS?
Q33.12 Why is a top-hat C2 aperture/diaphragm so called and why is it essential to have such apertures in your AEM?
Q33.13 Ideally how many C2 apertures should you have in an AEM?
Q33.14 How can one minimize post-specimen scattering?
Q33.15 What causes coherent bremsstrahlung (CB)?
Q33.16 If the hole count is >>1%, what is likely causing this?
Q33.17 Is STEM well suited for AEM? If so, why?
Q33.18 From the extra information in this chapter, explain why it is so important to align your X-ray detector with any planar interface.
Q33.19 List four ways by which you can minimize post-specimen scattering.

Q33.20 Why might CB be a problem that could cause you to misinterpret your analysis?
Q33.21 Give a simple way to distinguish a CB peak from a characteristic peak.
Q33.22 Why is it better to form X-ray images rather than individual spectra or line profiles?
Q33.23 Why is it much more difficult to acquire good X-ray images than individual spectra?

TEXT-SPECIFIC QUESTIONS

T33.1 A P/B ratio is measured on the Ni K_α line at 300 kV using the Fiori method and is determined to be 1000. What should you probably do?
T33.2 Why don't microscope manufacturers just coat the insides of the AEM with Be on all surfaces?
T33.3 Why coat any of the insides of the microscope at all with a low-Z material?
T33.4 To decrease the X-ray background for a particular specimen, the kV should be increased. Why?
T33.5 In an experiment, some unidentified X-ray peaks are present. Name some possible causes.
T33.6 Upon measuring a hole count for a particular specimen, you discover that the illumination system is not 'clean.' Besides being floored by this discovery, you want to know what you can do to fix it. So what do you do?
T33.7 Having become quite the expert on XEDS in the SEM, your lab partner tilts the specimen in the TEM chamber to 45° 'just like we did in the SEM.' Before smacking your lab partner repeatedly, what should you remind him of?
T33.8 Upon analyzing the XEDS spectra of your material, you cannot conclude what several peaks correspond to. You have tried sum peaks, escape peaks, and the internal fluorescence peak with no luck. Have you discovered a new element, or is there another answer that is more likely?
T33.9 David and Stuart are working at the 1210 when we arrive one morning. They are using XEDS, and all their spectra have peaks below 1 keV. What in the world are they doing?
T33.10 How would you discern if your X-ray spectrum contains a significant amount of spurious or system X-rays? How can you distinguish the spurious contributions from stray X-rays and stray electrons?
T33.11 Calculate the energies of the principal coherent bremsstrahlung peaks generated from a thin foil of Cu in a <001> orientation by 120-kV electrons when the EDS detector has a 20° take-off angle. Compare your answer with Figure 33.9 and comment on any discrepancies. (Hint: you'll have to find the lattice parameter of pure Cu.)
T33.12 How wide do you think the probe in Figure 33.5 is? Explain your reasoning.
T33.13 If the electron probe that you use to do analysis is as shown in Figure 33.5, what effect would this have, for example, on a line-profile analysis across an interface?
T33.14 Which effect(s) in Figure 33.6 is (are) causing the extra peaks in Figure 33.7?
T33.15 Are any other peaks in Figure 33.7 not caused by effects in Figure 33.6 and if so what is causing them?
T33.16 Explain the reasons for each of the directives at the end of Section 33.3.B.
T33.17 Distinguish analog and digital mapping and explain why one is so much better than the other.
T33.18 Distinguish SI and PTS.
T33.19 Make a list of the different ways to cut up the SI data cube and explain the different information that you would get from each cut. See if you find one that has not been published and make a name for yourself.

34

Qualitative X-ray Analysis and Imaging

CHAPTER PREVIEW

It is a waste of time to proceed with *quantitative* analysis of your XEDS spectrum or image without first carrying out *qualitative* analysis. Qualitative analysis requires that *every* peak in the spectrum be identified unambiguously, with statistical certainty, otherwise it should be ignored for both subsequent quantitative analysis and imaging. We emphasize this point because of the many opportunities for the misidentification of small peaks in the spectrum. In this chapter, we'll deal initially with acquisition and identification of the elemental information in spectra and images. First, we will show you how to choose the best operating conditions for your particular AEM and XEDS system. Then we'll explain the best way to obtain a spectrum for qualitative analysis. You have to acquire a spectrum with sufficient X-ray counts to allow you to draw the right conclusions with a given degree of confidence. There are a few simple rules to follow which allow you to do this.

A QUALITATIVE MUST!
Although such an approach may seem time-consuming and unnecessarily tedious, the need for initial qualitative analysis of the spectrum cannot be stressed too strongly.

Two advantages are gained from rigorous qualitative analysis. First, you may be able to solve the analytical problem at hand without needing to perform full quantification. Second, when quantification is carried out (see the next chapter), you will not spend an inordinate amount of time analyzing an element that isn't there, and you can be confident that your results are valid. We'll go over the many ways to misidentify peaks in your spectra, particularly small ones, which may, in fact, arise from important trace elements but, which might be artifacts, could be peaks from another element, or are possibly statistically insignificant. Commercial peak-identification software, while improving all the time, is not error-free. We'll end with a few words about qualitative X-ray imaging.

34.1 MICROSCOPE AND SPECIMEN VARIABLES

When you first acquire a spectrum, the operating conditions should maximize the X-ray count rate to give you sufficient intensity in the characteristic peaks in your spectrum, in the shortest time, with the minimum number of artifacts. You need sufficient counts so you can detect, unambiguously, the presence of *all* the elements in your specimen (within the limitations of your XEDS detector) *with statistical certainty*. As we'll explain, the best conditions for such qualitative analyses require that you obtain the spectrum from a reasonably thick, large area of your specimen, using a large probe and a large aperture to give the most current, but in doing this you'll compromise other desirable analysis qualities, particularly high spatial resolution. So right up front you need to know two key points

- There are only three requirements for good qualitative analyses; counts, counts and more counts
- The conditions for the best qualitative (and quantitative) analysis (which are also those that give the best analytical sensitivity) are precisely the worst for obtaining the best spatial resolution.

A more complicated factor in getting the most X-ray counts in your spectrum is choosing the right operating voltage. Remember, back in Figure 33.10.C, we showed that you get a higher detection and collection efficiency if you decrease the kV because the scattering cross section (σ) increases when the kV decreases; that was a specimen effect. Now we are talking about a gun effect;

as the kV increases, the gun brightness increases (go back and check this in Chapter 5). While the two effects counter each other somewhat, the added advantage of an increased P/B ratio (which helps with detecting small peaks) with higher kV, as well as improved spatial resolution (due to less beam spreading), tips the scales in favor of using the maximum kV at all times. Only choose a lower voltage if knock-on damage is a problem, as might be the case, for example, in a 200–400 kV AEM with a beam-sensitive specimen such as a ceramic, a mineral, most semiconductors or a low-Z (<15) metal/alloy.

Pick a portion of your specimen that is single phase in the area of interest and make sure it is tilted well away from strong diffraction conditions to minimize crystallographic effects (more on this in Chapter 35) and coherent bremsstrahlung. Ideally, you will need a probe current of several tens of nanoamps. The necessary combination of probe size and final aperture depends on the type of source in your AEM. To get several tens of nanoamps from a LaB_6 source, you have to select a relatively large probe size, say, a few tens of nanometers and a large C2 aperture. An FEG source will give much less total current than a thermionic source and it won't be possible to generate more than a few nanoamps at best. So, for reasonable qualitative analysis and subsequent quantitative analysis, you'll have to accumulate your spectrum for a longer time with an FEG compared to a thermionic source. A C_s-corrector on your FEG-AEM will help tremendously, since a larger probe-forming aperture can be used without compromising the quality of the probe, and tens of nanoamps can be generated in a C_s-corrected probe of a couple of nanometers dimension. So if your institution has enough money to spend on your AEM you can minimize the otherwise necessary compromise between needing lots of counts and getting high spatial resolution

You can always gather more counts in your spectrum by choosing a thicker region of the specimen. There is nothing wrong with doing this when you are carrying out *qualitative* analysis. A thick specimen degrades your spatial resolution, but we've already agreed to compromise that aspect of the analysis during this initial qualitative procedure. The only danger is that, if you are interested in finding a few weight percent of a light element, those weak X-rays may be absorbed in your specimen and so may not be detected. However, from an experimental standpoint, having carried out qualitative analysis of a relatively large, thick region, you can always do further analyses of smaller, thinner areas, under conditions that optimize spatial resolution, which we'll discuss in Chapter 36.

Remember that we have been talking about several different 'resolutions'. Don't confuse them

- spatial resolution: distances measured in nm (see Chapter 36).
- chemical resolution: analytical sensitivity/detection limits depending on P/B (see Chapter 36).
- energy resolution: identifying elements by distinguishing their spectral peaks at different energies (see this chapter and Chapter 33).

So, just in case you haven't got the message by now, good qualitative analysis (and subsequent quantification) requires a large number of X-ray counts in the spectrum (just how many we'll tell you in a while). These counts might take a long time to generate, so you run the danger of damaging, or changing the chemistry of any beam-sensitive specimens. You may also contaminate your chosen area if your AEM is not UHV and/or your specimen is not clean. So it's always good to use a plasma cleaner before analysis, unless doing so will destroy your specimen. (Go and check Chapter 10 or the companion text where we talk about specimen preparation.) To minimize damage and contamination, you should spread the beam over as large an area as possible, either by overfocusing C2 if you're in TEM mode or by rastering the beam in STEM mode, remembering that, in doing so, your analysis will be an average over the chosen area. Use a liquid-N_2 cooled, low-background holder, especially if contamination is still a problem.

34.2 BASIC ACQUISITION REQUIREMENTS: COUNTS, COUNTS, AND MORE CAFFEINE

The first and most important step in qualitative analysis is to acquire a spectrum across the complete X-ray energy range. Analysis can often be accomplished using X-rays with energies from $\sim$1 to 10 keV, and this is the typical range used in the SEM. However, the TEM has a much higher accelerating voltage, and the consequent increase in available overvoltage (remind yourself what this is. Hint: see Chapter 4) means that you can easily generate and detect much higher-energy X-rays. If you are using an intermediate-voltage AEM and a windowless IG detector, we noted in Chapter 32 that all the possible K_α lines from all the elements above Be in the periodic table can be detected.

ENERGY RANGE

The first thing to do is to adjust your computer display to the widest possible energy range. For a Si(Li) detector or an SDD, 0–40 keV is sufficient and for an IG detector, 0–80 keV may be more useful.

Of course, if you know the specimen you are analyzing, such a step may not seem essential, but it is still a wise initial precaution since unanticipated contaminants or trace impurities may be present. The next steps are the basics for acquiring a spectrum for qualitative analysis

- Collect a spectrum over, say, 0–40 keV for several hundred seconds (take a coffee break) and ascertain the actual energy range over which all the detectable characteristic peaks occur.
- If all the peaks that you can see are present in an energy range < 40 keV, re-gather the spectrum over that reduced range (take another coffee break), thereby improving the resolution of the computer display by lowering the number of eV per channel.
- The spectrum that you finally gather for qualitative analysis must be displayed with *no more than 10 eV per channel* resolution or better. A display range of 0–20 keV should be possible under these conditions (i.e., 2048 channels in total) in all but the most ancient of XEDS computer systems.
- If you have analog processing electronics, you can increase the counts in your spectrum by reducing the detector time constant to maximize the throughput. This step degrades the energy resolution of the XEDS but, for many qualitative analyses, this is not important. A digital system will automatically optimize the throughput of counts.
- Watch the dead-time readout while acquiring the spectrum to make sure you haven't chosen a combination of probe current and specimen thickness that overloads the detector electronics. Remember that you want to keep the dead time below about 50–60%, and an output count rate of around 10 kcps (analog) to 30 kcps (digital) is about the best that can be handled by current detector electronics under these conditions. This will rarely be a problem in thin-foil analysis!
- The total counts in this qualitative spectrum should exceed 1,000,000 *over the full energy range*. While this may seem a lot, at 3 kcps, it will take you just over 15 minutes to accumulate this number of counts (now it's probably time for a bathroom break, anyhow). So adjust the probe size/current, the size of the C2 diaphragm and (if possible) the specimen thickness until the count rate is sufficient.

When you've got a good high-count spectrum over a suitable energy range, there is a well-defined sequence of steps developed for analysis of spectra from the SEM (see Goldstein et al. as usual) that should be followed to ensure that you correctly identify each peak in the spectrum and disregard those peaks that are artifacts or are not statistically significant and we'll describe a modified form of this procedure for thin-foil specimens in the next section.

Figure 34.1 shows the effects of increasing acquisition time (i.e., increasing counts) on the visibility of small peaks. The longer the time, the better the quality of the spectrum. So coffee breaks can be very beneficial.

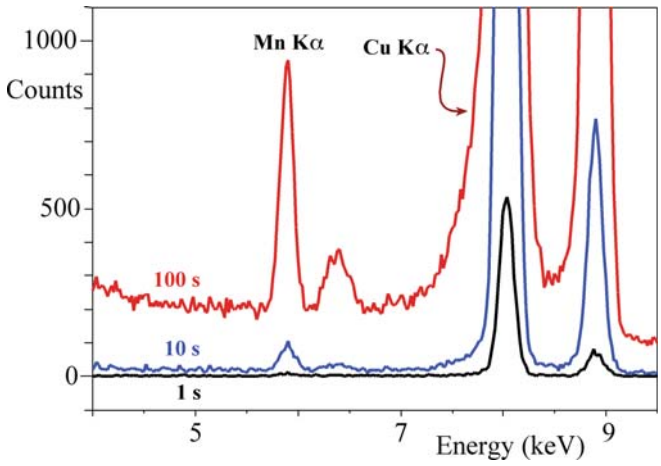


FIGURE 34.1. The improvement in the quality of a spectrum from a Cu-1% Mn thin foil with increasing acquisition time. After 1 s (black) the small Mn $K_{\alpha/\beta}$ peaks are not visible; after 10 s (blue) the MnK_{α} is detectable but the K_{β} is barely visible. After 100 s (red) all peaks are clearly visible. With increasing acquisition time it is generally much simpler to discern peaks from the background and specimen peaks from artifact peaks, so peak identification will proceed a lot more easily.

34.3 PEAK IDENTIFICATION

There are four key steps to running a successful qualitative analysis.

First: let's assume that you have read, understood, and applied the contents of Chapters 32 and 33. So you know what artifacts are likely to arise from your XEDS system and what system peaks occur in your AEM and you're aware of CB, etc. Now, ensure that the computer display is calibrated to be as accurate as the display resolution, over the energy range you've selected. If your spectrum is displayed at 10 eV per channel, the characteristic peak centroids must all be within ±10 eV of their true position on the energy scale.

Second: the computer system can be used to run an automatic identification check on the peaks in the spectrum, assuming the energy display is well calibrated. If the spectrum is simple, containing a few well-separated peaks, this automatic step may be all that is required. However, as we'll discuss in Section 34.5, misidentification occasionally occurs during such an 'auto-search' or 'peak ID' software routine. The more complex the spectrum, the more likely this is to happen, e.g., if the spectrum contains many peaks, particularly if peak overlap is occurring (e.g., Zn L_{α} confused with Na K_{α}) or if the spectrum contains complex peak families from heavier and/or rarer elements (e.g., Ta M_{α} confused with Si K_{α}). This problem is exacerbated if you don't understand the complexities of X-ray families or don't follow all the precautions that we'll go through below. In addition, even in the best software, small peaks may sometimes be missed and phenomena such as CB are often not taken into account.

THE KEY

For good qualitative analysis be suspicious. Don't just seek the peaks you expect, but be prepared to find peaks that you don't expect.

LABEL THE PEAKS

Take care to label each peak on the computer display or note it in your lab notebook when you have decided which element it comes from.

Remember, each X-ray is emitted with a very well-defined line energy (~ 1–5 eV wide) but the XEDS system degrades the line to a broad peak (FWHM ~ 80–180 eV over the energy range detected by a typical Si(Li) detector). So we'll talk about peaks in the spectrum, which correspond to X-ray lines emitted from specific elements and which are identified as belonging to families of lines that are superimposed by the computer software on the displayed spectrum.

Third: go back to Chapter 4 and remind yourself about things like critical ionization energy, X-ray line energy, K, L, M families of X-ray lines, relative weights of lines, fluorescence yield, etc., because we're assuming you know all of this backwards.

Our peak analysis will always include the following steps

- Look first at the most intense peak since this should be easiest to identify as a K, L, or M_α peak; then work on down through the associated family lines. If you can't identify the most intense peak easily then you've got a problem; e.g., the calibration is off or the electronics or software are not functioning correctly, so it's time to ask for technical help and go for yet more coffee.
- The most intense peak is also the most likely to have associated artifacts, such as an escape peak (1.74 keV below that peak energy in a Si(Li) system) and a sum peak (at twice the peak energy), so you can quickly remove such small peaks from the unknown list (most software now automatically notes the energy where these artifact peaks should appear).
- Go to the next most intense peak not included in the above step and repeat the search. Then repeat this exercise until all peaks are identified.
- Always think about pathological overlaps; look for spurious peaks, system peaks, and artifact peaks.

Fourth: the bookkeeping; in choosing the possible K, L, or M lines that could be present at a specific energy, you can either use the computer-generated X-ray line markers on the display or consult an appropriate source such as the 'slide rules' offered by most commercial manufacturers, or find them on the Web at, e.g., URLs 1 and 2. We introduced DTSA in Chapter 33 and it is great for comparing simulated spectra with acquired spectra, as well as for checking the specific energies of X-ray lines, particularly in the more complex families of lines in spectra from heavier elements. There's much more about DTSA in the companion text.

Good bookkeeping is essential during the identification sequence we will now describe, particularly if your spectrum contains many peaks.

So now we've outlined the principles, let's get down to the specifics. Follow this 8-step process

1. If a K_α line matches the peak, look for the K_β line which has about ~10% of the K_α intensity (this 10% is a line 'weight'; go back and check Table 4.1). With a modern XEDS, the K_β line *must* be present at X-ray energies above ~1.74 keV (Si K_α), so long as it is not overlapped by a more intense peak from another element. Below this energy, your detector may not be able to resolve the two lines.
2. If a K_α and K_β pair fits the peaks and the K_α energy is $>$ ~8 keV (Ni K_α), look for the L lines at ~0.9 keV if you are using a Be-window detector. For an UTW/windowless detector the L_α lines from Cl and above ($>$ ~ 0.2 keV) may be detectable but only if there's a lot of Cl, because (a) these relatively weak X-rays will be strongly absorbed and (b) their fluorescence yield is pretty abysmal. Ni L_α = 849 eV, Cl L_α = 200 eV.
3. If a K_α line does not fit, check for an L_α or M_α line fit since these are the most intense lines in the L and M families.
4. If an L_α line fits, there *must* be accompanying lines in the L family. The number of visible lines will vary depending on the intensity and energy of the L_α line, with more lines resolvable at higher line energies. The other lines in the family are all of lower intensity than the L_α line, and the following lines may be detectable (the number in parentheses is the weight relative to the L_α line); $L_{\beta 1}$ (0.7), $L_{\beta 2}$ (0.2), and $L_{\gamma 1}$ (0.08) lines at higher energies and possibly the Ll (0.04) line at lower energy. Other, even less intense, lines ($L_{\gamma 3}$ (0.03) and L_η (0.01)) may be visible if the L family is extraordinarily intense, but this is rare.
5. If the L lines fit, there *must* be a higher energy K_α/K_β pair, since the AEM beam energy is usually sufficient ($>$ 200 keV) to generate the K lines from all the elements in the periodic table. Make sure you choose a broad enough energy range to display them on your computer.
6. The M lines are usually only visible for elements above La in the periodic table if a Be-window detector is used, and above about Nb, if a UTW detector is used. Again there has to be a lot of Nb to pick up the weak M line. La M_α = 833 eV, Nb M_α = 202 eV.

7. The M_α/M_β line overlap is difficult to resolve because all the M lines are < 4 keV. If an M_α/M_β line fits, look for three very small lines M_ζ (0.06), M_γ (0.05), and $M_{II}N_{IV}$ (0.01) lines, which will be more visible if more of the element is present.
8. If the M_α line fits there *must* be a higher energy L line family and possibly the very high energy K lines may be detectable; again, this depends on the detector (IG is better for high kV lines), computer display (out to ~ 80 keV), and the accelerating voltage (higher is better).

Figure 34.2 shows the families of lines expected in the display range from 0 to 20 keV, giving you some idea of the distribution of families of elemental lines that you

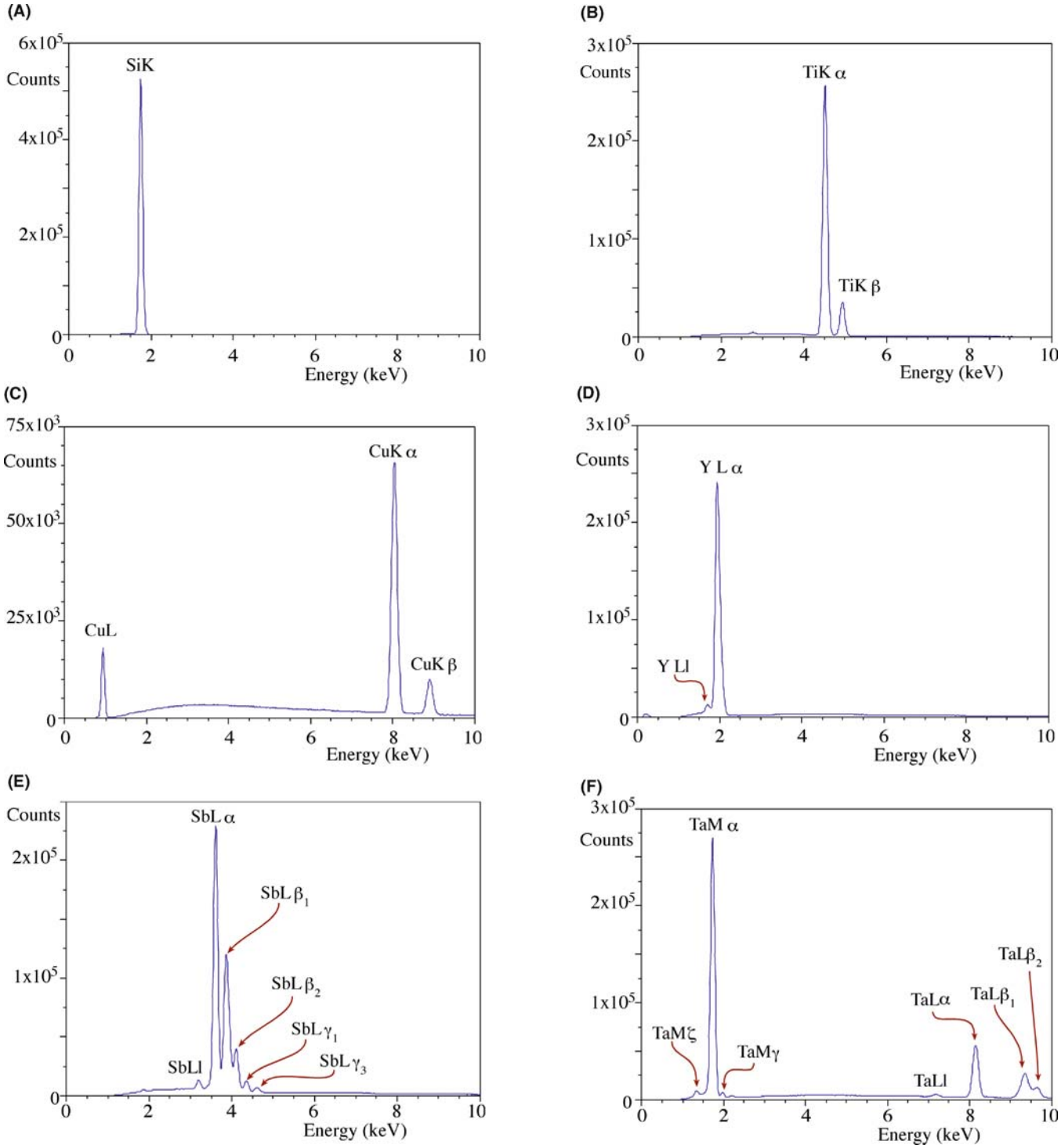


FIGURE 34.2. X-ray spectra from elements spanning much of the periodic table showing the families of characteristic lines for (A) Si, (B) Ti, (C) Cu, (D) La, (E) Sb, and (F) Ta. Starting with a single Si K_α line at low Z and low X-ray energy, the series progresses through the appearance of the families of L (Cu and La) and M (Sb and Ta) lines. Note the increasing separation of the peaks of a given family as both *Z* and keV increase.

should find when you follow the procedure outlined above. For example, you can see for which elements you should expect to see only a single K line or resolve the K_α/K_β pair, and for which elements you should expect to see both K and L families or L and M families.

FAMILIES OF PEAKS
Looking for **families** of peaks. If a family member is missing your identification may be wrong.

Reasons for a missing family member

- It may be overlapped by another peak. This is the most likely cause and you can possibly resolve it by
  (a) re-gathering the spectrum with a longer time constant (analog system only)
  (b) using peak deconvolution software (see Section 34.4)
  (c) using a higher-resolution technique, such as EELS (see Chapter 38)
- Your computer display range may be too small so the peak is cut off (easy to solve, although if you followed our instructions this should never happen)
- The keV of the beam is too low to excite the line (should never be a problem in an AEM, only in an SEM)

Repeat the exercise: Go to the next most-intense peak that has not been identified by the eight steps in the first search. Continue this process *until all the major peaks are accounted for*. As you go, remember again to look for the escape peak(s) and sum peak associated with each *major* characteristic peak that you have conclusively identified. However, these artifacts and any CB peaks will be very small and before you worry about them you should make sure that the small peaks are statistically significant and we discuss how to do this for all minor peaks in Section 34.5 below. If you have a Si(Li) detector you'll always find a small Si-K internal-fluorescence peak, the Si escape peaks will lie at 1.74 keV below major peaks in the spectrum, and will not occur for elements below phosphorus. For an IG detector there will be Ge internal-fluorescence peaks (possible K and L) and there may be both Ge K and L line escape peaks at the appropriate energy below major peaks (9.89 keV for the Ge K_α escape and 1.19 keV for the Ge L_α escape). If you suspect a sum peak at twice the energy of any major peaks then re-acquire the spectrum at a much lower dead-time ($< 20\%$) and see if the suspected sum peak disappears. If you suspect a CB peak then re-acquire the spectrum at a different accelerating voltage or specimen orientation and see if the small peak shifts.

Check for special cases: The relatively poor energy resolution of the XEDS detector means that there are several pairs of peaks that occur quite commonly in materials science specimens that cannot be resolved. These go by the delightful name of 'pathological overlaps' and include, inter alia

(a) the K_β and K_α lines of neighboring transition metals, particularly Ti/V, V/Cr, Mn/Fe, and Fe/Co
(b) the Ba L_α line (4.47 keV) and the Ti K_α line (4.51 keV)
(c) the Pb M_α (2.35 keV), Mo L_α (2.29 keV), and S K_α (2.31 keV) lines
(d) the Ti, V, and Cr L_α lines (0.45–0.57 keV) and the K lines of N (0.39 keV) and O (0.52 keV) detected in UTW/ATW or windowless XEDS systems.

PATHOLOGICAL OVERLAP
When it is impossible to separate two peaks even when you know they are both there.

These problems can often be solved by careful choice of the energy range on the computer display. For example, if you are only observing from 0 to 10 keV the S K/Mo L line overlap would be clarified by the presence or absence of the Mo K lines around 18 keV which, again, you should have seen in your first, broad energy-range, spectrum acquisition. If you suspect that any pathological peak overlaps are occurring in your spectrum, then re-gather under conditions that maximize the energy resolution of the detector system (i.e., longest (analog) time constant and low count rate (< 5 kcps)), and also maximize the display resolution to at most 5 eV per channel (you'll need more coffee).

34.4 PEAK DECONVOLUTION

If the overlap is still not resolvable, then you should run a peak deconvolution routine in your computer software (these are pretty standard and not much has changed since; see, e.g., Schamber 1981). Such routines are capable of detecting and resolving many of the classic materials science overlaps, such as the transition metal L lines and low-Z K lines and an example of such a deconvolution is shown in Figure 34.3.

In addition to deconvoluting any peak overlaps, it can be very useful to deconvolute the point-spread function of the XEDS detection system using a process called zero-peak deconvolution (which is analogous to the zero-loss peak deconvolution process used in EELS (Sections 37.5.A and 39.6)). Such a process is becoming increasingly popular in many imaging and spectroscopy techniques because of the development of various robust mathematical procedures (e.g., Janssen 1997)

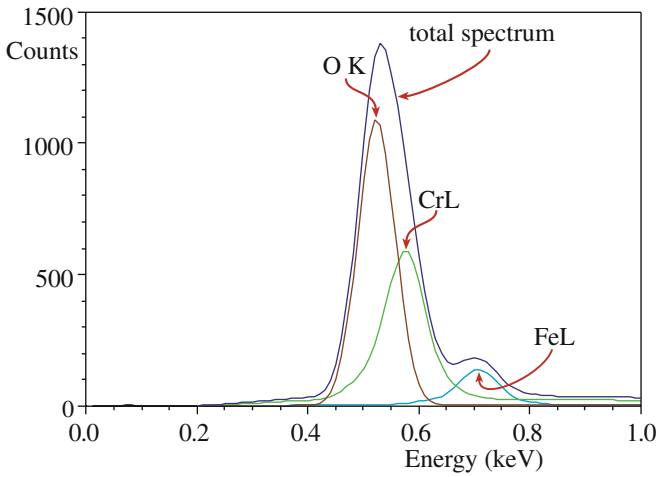


FIGURE 34.3. The total spectrum (dark blue) arises from the overlap of three Gaussian spectral peaks (the L_α lines of Fe (light blue) and Cr (green) and the O K_α line (brown)) from a mixed Fe-Cr oxide. Deconvolution of the individual contributions to the total spectrum is essential to determine the intensities in the three constituent peaks prior to any quantification attempt.

some of which are described in the companion text. In effect, this deconvolution removes the electronic-noise component of the characteristic peak width, giving a spectrum with nearer to noise-free resolution (Watanabe and Williams). This process requires that your XEDS system must display the noise peak at 0 eV in the spectrum (which not all manufacturers do) and this is called the 'strobe peak'. An iterative process can, in effect, remove the noise in your spectrum. As shown in Figure 34.4, this process obviously improves energy resolution but also improves the P/B ratio (and thus the minimum detection limit, as we'll see in Chapter 36) and can reveal peaks that would otherwise be masked by adjacent, more-intense, peaks.

Now, you always have to be careful with deconvolution, since any mathematical manipulation can introduce its own artifacts, while otherwise improving the spectrum quality. So it's best to practice deconvolution on both simple and complicated spectra that you know and understand well, until you feel confident that you understand the strengths and limitations of the procedure available in your particular software package.

Given all these steps and the multiple decisions that you have to make, it's clear that qualitative analysis can be an extraordinarily difficult procedure even for the

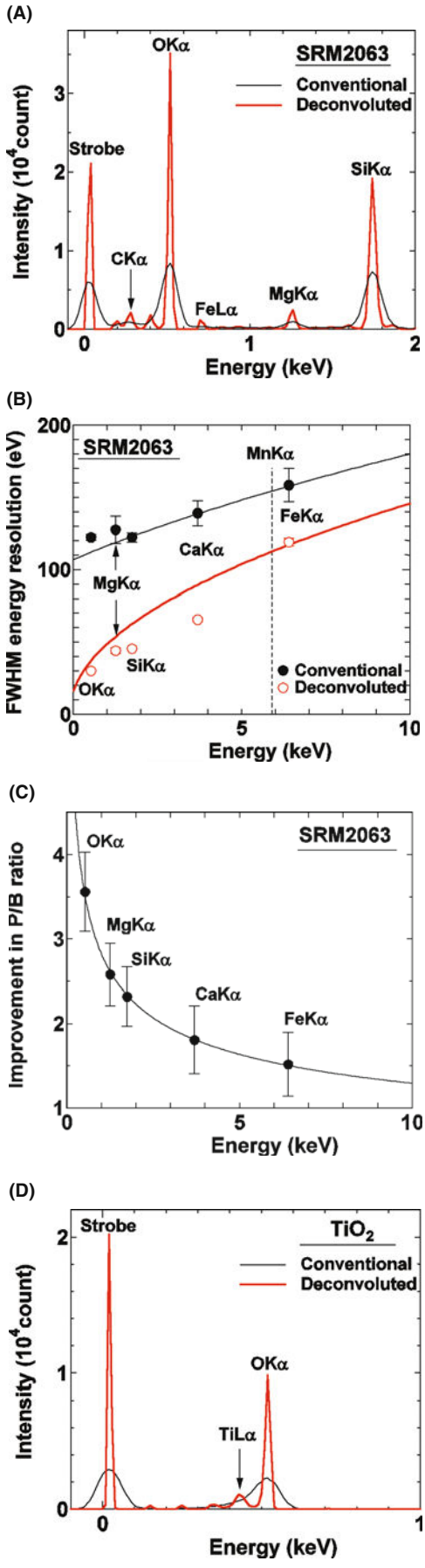


FIGURE 34.4. (A) The effects of deconvoluting the zero-energy strobe peak from an experimental XEDS spectrum of NIST SRM 2063. (B) The consequent reduction in peak FWHM (i.e., improvement in energy resolution) across the energy range 0–10 keV. (C) The improvement in P/B ratio of the major peaks in the SRM 2063 spectrum as a result of deconvolution. (D) Revealing the L_α peak from Ti in TiO_2, which is usually hidden by the O K_α.

most experienced analyst. This complexity is one reason why even the best software gets it wrong sometimes (see Section 34.6). So, as we said at the start, be suspicious!

In summary, applying the 8-step process, with appropriate deconvolution where needed, should permit you to identify all the major peaks in your spectrum. There might still be minor peaks, which may or may not be statistically significant, and you have to decide whether you are going to identify or ignore these peaks. Now we'll tell you how to make this decision.

34.5 PEAK VISIBILITY

Small intensity fluctuations are often present in your spectrum that you cannot clearly identify as peaks. In this case, there is a simple statistical criterion (Liebhafsky et al.) that you can apply to ascertain if the peak is statistically significant or if it can be dismissed as random noise. You must count for a long enough time so that the bremsstrahlung intensity is smooth and any peaks are clearly visible, as summarized in Figure 34.5.

- Increase the display gain until the average background intensity is half the total full scale of the display, so the small peaks are more easily observed.
- Get the computer to draw a line under the peak to separate the peak and background counts.
- Integrate the peak (I_A) and background ((I_A^b)) counts over the same number of channels; use FWHM if it can be discerned with any confidence; if not, then the whole peak integral will do.

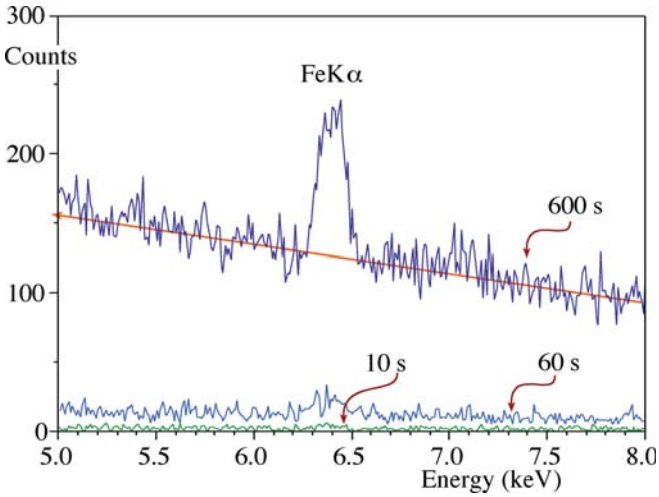


FIGURE 34.5. With increasing counting time, a clear characteristic Fe K_α peak develops above the background in these spectra from Si-0.2% Fe, thus demonstrating the need to acquire statistically significant counts before deciding if a peak is present or absent. As indicated (orange line), the background in the 600 s spectrum approximates to a straight line making the peak clearly visible. Note that Fe K_β is beginning to appear at 7.05 keV, although it is not yet statistically significant.

If $I_A > 3\sqrt{I_A^b}$ then the peak is statistically significant at the 99% confidence limit and must be identified. Remember, you'll still make an erroneous peak identification in ~ 1% of analyses using this criterion; just hope it isn't the one that derails your PhD or gets you fired!

If $I_A < 3\sqrt{I_A^b}$ then the peak is not significant and should be ignored.

If the insignificant peak is at an energy where you expect a peak to be present, but you think there is only a small amount of the suspected element in your specimen, then *count for a longer time* to see if the statistical criterion can be satisfied in a reasonable length of time. If this peak is a critical one, and it is often the minor or trace elements that are most important, then take *whatever time is necessary* to detect the peak. There is no reason not to gather the spectrum for many minutes or even an hour or more, so long as doing so does not change/damage or contaminate your specimen. Lunch breaks now become beneficial.

TIME

When you count for long times to search for characteristic peaks of low intensity, you will also begin to detect more easily the small peaks from the various spurious effects; e.g., CB peaks, Si or Ge internal-fluorescence peaks, and system peaks such as Fe and Cu. You also increase the possibility of contamination and beam damage to your specimen.

However, do *not* obtain more counts by raising the count rate above that which the processing electronics can handle, because you may introduce extra sum peaks and also degrade the energy resolution of the spectrum.

If you're worried about damaging the analysis area, as we stated at the beginning, it is best to spread the probe over as large an area as possible, either by defocusing the C2 lens in TEM mode or by using a scan raster in STEM mode.

Identifying the statistically significant peaks by the above method is one thing. Quantifying the amount of the element responsible for the peak is another matter and usually many more counts are required, as we'll see when we talk about detection limits in Chapter 36. However, once you're happy with the peak ID, you may be able to identify the phase/nanoparticle/precipitate that is being analyzed without any further work. For example, in the material that you are investigating, thermodynamics may tell you that there are only a few possible phases that can exist after the processing/thermal treatment given to it, and these phases may have very different chemistries. A glance at the relative peak intensities may be sufficient to conclude which phase you have just analyzed because, as we'll see below and in

more detail in the next chapter, one of the marvelous advantages of thin-foil analysis in the AEM is that the peak intensities are often directly proportional to the elemental concentrations. As a result, quantification can be extremely simple.

To conclude this section, we'll look at two examples before going into the details of quantification in the next chapter.

The oxide-glass example: let's run a qualitative analysis on the spectrum in Figure 34.6. The spectrum is from a thin NIST oxide-glass film on a carbon support film on a Cu grid. X-rays were accumulated for 1000 s with a Be-window, IG detector at an accelerating voltage of 300 kV. Because of the Be window, we do not expect to see lines below ~0.8 keV and so the O K_α (0.52 keV) will not be detectable. The spectrum only contained peaks in the range from 0 to 10 keV and the first peak to be examined was the most intense high-energy peak, line #1, which was consistent with the Si K_α line at 1.74 keV. (The K_α/K_β pair cannot be resolved by the detector at this energy.) A similar treatment of the next most intense high-energy line (#2) at 3.69 keV produced a match with the Ca K_α and it is also possible to identify the associated K_β at 4.01 keV (the Ca L line is not detectable). The third most intense line #3 is at an energy of 6.4 keV. The K-line markers identified it, along with the smaller one to its right, as being the Fe K_α and K_β pair. No L line fit was reasonable (Dy L_α at 6.5 keV being the only alternative) and there are no M lines above about 4 keV. The Fe L line at 716 eV will not be detectable because of the Be window.

Next, the smaller peaks were tackled and the Cu K_α (and K_β) was identified at 8.04 keV, the Ar K_α at 2.96 keV (the K_β was too small to be visible), and the Mg K_α was the last to be identified at 1.25 keV.

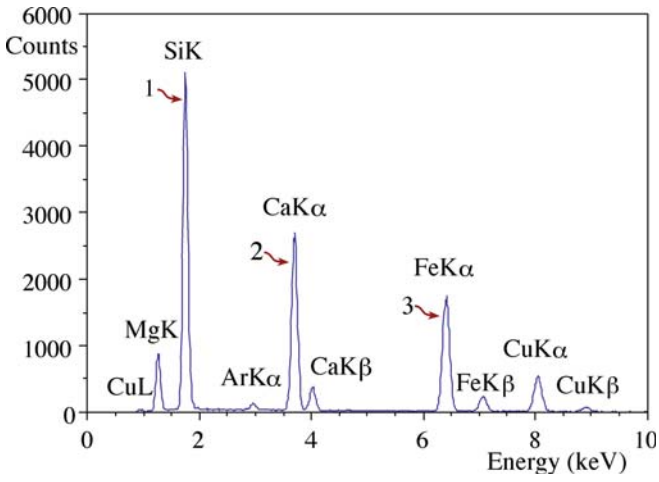


FIGURE 34.6. Energy-dispersive spectrum obtained at 300 kV from a thin oxide-glass film. The characteristic peaks were identified through the procedure outlined in the text.

Mg (as its oxide) is a common element in glasses, so this peak is expected. But Ar isn't such a common element in glasses. It probably arises from Ar implanted during ion-beam thinning and, as we discussed earlier and in detail back in Chapter 10, many specimen preparation processes can affect the chemistry of the specimen surface. You should also be aware that Ar K at 2.96 keV is often confused with the Al K sum peak (2×1.49 keV). But since there is no Al in the glass, and no major peak at half the Ar K energy, then Ar is the best answer.

SPECIMEN PREPARATION

Now you see why it is very useful to have some knowledge about your specimen and how it was prepared.

Likewise, Cu is not a common glass-forming element, but, since the specimen was on a Cu support grid, the Cu peaks are most probably due to post-specimen scatter of electrons or X-rays and so we cannot conclude that there is any Cu within the specimen. No escape or sum peaks were detectable.

ABSENCE OF THE Cu L_α Line

The absence of the line at 0.93 keV in Figure 34.6 is evidence that the thick Cu grid is responsible for a Cu line; the low-energy L X-rays will be absorbed in the grid itself before they can be detected.

The Fe-Cr-Ni example is shown in Figure 34.7 and this spectrum contains six Gaussian peaks, which can easily be identified following the procedure outlined above as the K_α and K_β pairs from Fe, Cr, and Ni. Even the average metallurgist will know that this specimen can only be some kind of stainless steel and this may be all the information that is required, making subsequent quantitative analysis redundant. But if you need more information, e.g., the specific grade of stainless steel, then you have to make measurements of the relative peak intensities, and this is the first step in the quantification procedure. In fact, we will see in Chapter 36 that the thin-foil quantification equation, to a first approximation, predicts that the amount of each element is directly proportional to the peak height. If you measure the relative heights of the K_α peaks in Figure 34.7 with a ruler, you can estimate the composition as ~ Fe-20% Cr-10% Ni which is within 10% of the classic 316 stainless composition of Fe-18% Cr-8 % Ni.

One real advantage of thin foil X-ray analysis is that you can get a good estimation of the composition of the analysis volume just by measuring the relative peak heights with a ruler.

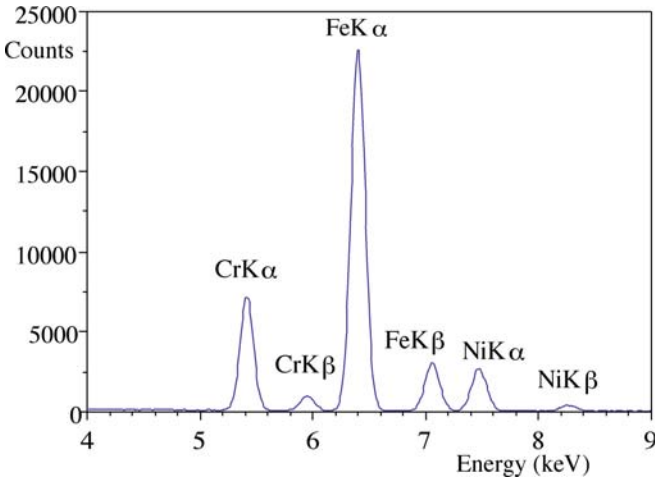


FIGURE 34.7. Spectrum from a stainless-steel foil in which the peaks are resolved and quite close in energy. To a first approximation, quantification is possible simply by measuring the relative heights of the K_α peaks.

Now there must be a good reason why, despite the fact that a 50-cent plastic ruler can give you a reasonable quantitative analysis of your spectrum in a few seconds, all AEMs have tens of thousands of dollars of computer hardware and software attached to the XEDS detector. This is because the stainless-steel spectrum is not really a challenge since all the peaks are close in energy, the peaks have lots of counts in them and there are no small peaks from trace elements. If the peaks were far apart in energy then relative X-ray absorption might occur, which can change the simple relationship between peak height and composition. Also if we really wanted to be sure that we had 304 rather than 316-stainless steel, then we would need a full quantification using the procedures described in the next chapter. If you were to do this, then you'd find that a full quantification would give a very similar result, but you could have much greater confidence in the true composition.

34.6 COMMON ERRORS

As we've seen, there's lots of room to misidentify peaks, particularly small ones that may be one of the many artifacts from the detector, the processing electronics, or the AEM-XEDS system. The first thing to realize is that if you (or even your advisor) can make mistakes, then the software can also do so. Remember, the software is only as good as the programmer. So while commercial automated peak-ID software is generally outstanding and getting better with every iteration, don't automatically believe it is always correct. By all means, as we said right at the start of Section 34.3, push the peak-ID button and get a quick analysis of your spectrum, but after you've done that, you might want to read the paper by Newbury which indicates some errors produced by a range of software systems. (See also the subsequent discussion noted in that reference.) The specimens used by Newbury to show the errors were often quite complex, with multiple peaks from rare, high-Z elements, but the article is instructive and illustrates the point that simply believing the software output is not always wise and suspicion is healthy.

As we've mentioned, it can be really useful to know something about your specimen chemistry *before* you put it into the AEM (and of course the TEM should be just about the last technique you use to study a completely unknown foil). It can also be very useful to remember how your specimen was thinned, since most thinning methods can change the surface chemistry of the foil. For example, electro-thinning methods can preferentially remove, or re-deposit one element from the foil or leave surface residue from the polishing medium. The thinner the foil is the more such changes in surface chemistry are exacerbated. Likewise, ion-beam thinning can result in the implantation of Ar, and FIB thinning can do likewise for Ga. So your foil may oxidize preferentially during thinning or pick up Cl from the perchloric-acid polishing solution, and so on. Ultramicrotomy is about the only method of thin-foil preparation that doesn't change the surface chemistry (although the freshly cut surface may be prone to rapid corrosion in the atmosphere and the defect structure is seriously changed from the bulk sample). So always be suspicious.

34.7 QUALITATIVE X-RAY IMAGING: PRINCIPLES AND PRACTICE

Individual-point analyses or multiple-point profiles across an interface are not the only ways to display X-ray data. As we saw in the previous chapter, we can produce X-ray images by a variety of methods and, if you are careful, such images can be used as compositional maps in which the intensity of the signal in the map is directly proportional to the generated X-ray intensity I_A. Under most circumstances, as we just described, we can take the next step and assume that, to some reasonable degree of accuracy, the X-ray intensity from element A in a thin foil is proportional to the concentration C_A. But there are some limitations to drawing a direct correlation, which we'll deal with in the next chapter on quantitative analysis. While there are obvious advantages to comparing maps of elemental distributions with other TEM images, this process is limited by the relatively poor statistics of X-ray acquisition, as should be eminently obvious to you by now. As you'll see in the next chapter, good quantification requires ~10,000 counts in the characteristic peak from element A, I_A. In early AEMs, such intensity

would easily take you a minute or so to acquire, even if there was a large amount of A present in your specimen. At this acquisition rate, even a 56×56 pixel image would take 50 hours to gather, so imaging didn't see a lot of practitioners and we just had to make do with qualitative, noisy maps, as shown back in Figure 33.13. However, as we've told you already, a lot has changed in recent years to improve the X-ray acquisition rate. First, to generate more X-rays, we have intermediate-voltage FEGs and spherical-aberration correctors to permit larger final apertures while maintaining small probes sizes. Second, to permit greater throughput of X-rays, we have developed digital pulse-processing, larger X-ray collection angles, and SDD arrays. Consequently, particularly if you can get away with using a reasonably thick specimen, X-ray mapping is a viable option on most modern AEMs where digital-display technologies and new image-processing software also makes life much easier.

The older X-ray dot maps, such as in Figure 33.13, simply register a dot when an X-ray is registered in a particular energy window in the XEDS spectrum. It doesn't matter whether the X-ray is a characteristic X-ray, a bremsstrahlung X-ray, a spurious X-ray, or a system X-ray, it is still registered, so problems like thickness or atomic number effects can occur since thicker/higher-Z specimens will generate more bremsstrahlung (as well as more characteristic) X-rays. These problems are a lot greater in bulk-specimen, SEM X-ray imaging, which has grappled with these issues for many decades (summarized in 1990 by Newbury et al.). To account for all these effects, a full quantitative procedure has to be applied to the intensity in each pixel and we'll cover this in detail in the next chapter.

For qualitative mapping, gray-scale images are still acceptable, but for full quantification, there is no way around the use of full-color images because of the many more color signal levels that your eye can discern and the advantages of color overlays to compare maps of different elements, and an example is shown in Figure 34.8, where the final RGB color image shows the chemical inhomogeneity of the Au-Pd nanoparticles, thus giving insight into why they work as catalysts for the peroxide-synthesis process. (Similar qualitative maps were shown in Figures 33.14 and 33.15.)

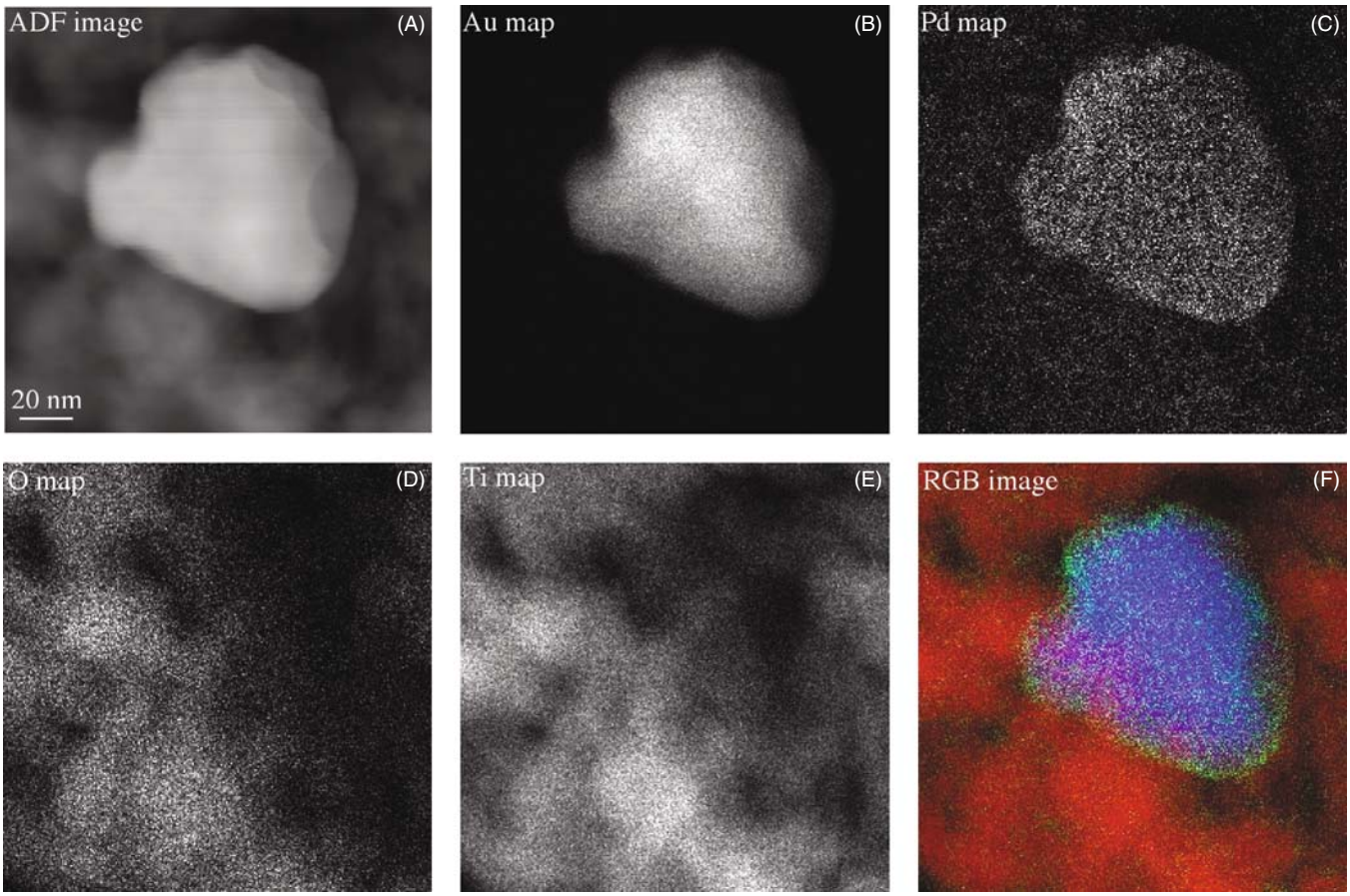


FIGURE 34.8. (A) STEM ADF image and qualitative X-ray maps showing the distribution of the (B) Au, (C) Pd, (D) O, and (E) Ti in a Au-Pd/TiO_2 catalyst nanoparticle. Such particles are used for peroxide synthesis. (F) The overlay of the color images from Ti (red), Pd (green), and Au (blue) reveals the core-shell nature of the particles with Pd on the outside and Au on the inside.

So how do you acquire a qualitative map like Figure 34.8?

- First you have to be in STEM mode, so your beam is scanning the specimen under digital control. Then select the largest probe size that will be still compatible with the desired X-ray image resolution. (Mapping with < 5 nm spatial resolution as in Figure 34.8 is challenging (see Chapter 36 on the factors controlling spatial resolution)). Only the best intermediate-voltage FEG instruments approach 1–2 nm, so a 5 nm probe would be a good start.
- Second, select a region of the X-ray spectrum that you want to map (e.g., set a window around a principal characteristic peak).
- Third, set a dwell time for the probe at each pixel that will permit sufficient X-ray counts to be acquired. Much more than a few seconds/pixel will make the acquisition time rather long, and < 0.1 s is only worthwhile if you have an intermediate-voltage FEG. (The best aberration-corrected FEG-AEMs can gather acceptable maps with dwell times of < 10 ms/pixel.) But, unless you have access to one of these instruments, start at about 1 s/pixel and gather a 64×64 pixel map and this will take you just over an hour. Do all the things we've told you to maximize the count rate: large beam diameter, large C2 aperture, and shortest (analog) pulse-processing time (this will minimize the chances of the processing electronics rejecting any of the few counts that are generated in such short dwell time).
- Last, start the scan and allow the map to commence building up on the computer display. If the intensity is acceptable and useful gray-scale information is obtained, it might be worthwhile to stop and regather the map for a longer period of time, e.g., for several hours, or even overnight, if it is a crucial map, However, for such long mapping times, you'll need to apply drift-correction software so that the map comes from the intended region of your specimen. Also it is obviously paramount that your specimen and AEM are clean, otherwise carbon contamination will build up to the point where the X-ray detection can be compromised, Also, as we've now told you innumerable times, specimen damage/contamination becomes an issue with longer scan times.

There are many other things you can do with the quantitative-analysis software and, given the time involved for even a simple qualitative map as just described, it's probably only worth doing fully quantitative maps. So we'll return to this technique in a lot more detail after you've learned all the steps necessary to translate the X-ray intensity into the elemental composition.

CHAPTER SUMMARY

One last time; doing the qualitative analysis first is not an option. The following steps are essential

- *Always be suspicious of any small peaks.*
- Get an intense spectrum across the energy range that contains all the characteristic peaks.
- Starting at the high-energy end of the spectrum, identify all the major peaks and any associated family lines and artifacts.
- If in doubt, collect for a longer time to decide if the intensity fluctuations are in fact peaks.
- Beware of pathological overlaps and be prepared to deconvolute any that occur.
- If you have the time, take a qualitative image using any crucial X-ray peaks.

GENERAL REFERENCES

Goldstein, JI, Newbury, DE, Joy, DC, Lyman, CE, Echlin, P, Lifshin, E, Sawyer, L and Michael JR 2003 *Scanning Electron Microscopy and X-ray Microanalysis* 3rd Ed. p355 Kluwer Academic Press New York.

SPECIAL REFERENCES

Jansson, PA, Ed. 1997 *Deconvolution of Image and Spectra* Academic Press San Diego.

Newbury, DE, Fiori, CE, Marinenko, RB, Myklebust, R., Swyt, CR and Bright, DS 1990 (A,B) *Compositional Mapping with the Electron Probe Microanalyzer* A: Anal. Chem. **62** 1159–1166A, B: Anal. Chem. **62** 1245–1254A. More on errors.

Liebhafsky, HA, Pfeiffer, HG, Winslow, EH and Zemany, PD 1972 *X-rays, Electrons and Analytical Chemistry* p349 John Wiley and Sons New York. Some statistics.

Newbury, DE 2005 *Misidentification of Major Constituents by Automatic Qualitative Energy Dispersive X-ray Microanalysis: A Problem that Threatens the Credibility of the Analytical Community* Microsc. Microanal. **11** 545–561. (See discussion by Burgess, S 2006 *idem* **12** 281 and Newbury, DE *idem* **1**, 282.) Discussion of errors produced using commercial software.

Schamber, FH 1981 *Curve Fitting Techniques and Their Application to the Analysis of Energy Dispersive Spectra* in *Energy Dispersive X-ray Spectrometry* 193–231 Eds. KFJ Heinrich, DE Newbury, RL Myklebust and CE Fiori NBS Special Publication 604 US Department of Commerce/NBS Washington DC.

Watanabe, M and Williams, DB 2003 *Improvements to the Energy Resolution of an X-ray Energy Dispersive Spectrum by Deconvolution Using the Zero Strobe Peak* Microscopy and Microanalysis Eds. D Piston, J Bruley, IM Anderson, P Kotula, G Solorzano, A Lockley and S McKernan 124–125.

URLs

1) http://microanalyst.mikroanalytik.de/index_e.phtml
2) http://www.cstl.nist.gov/div837/Division/outputs/DTSA/DTSA.htm

SELF-ASSESSMENT QUESTIONS

Q34.1 Why is it important to know the relative weights of characteristic X-ray family lines?
Q34.2 Why can't we predict the relative weights of the lines in different X-ray families (e.g., the K_α/L_α ratio) while we can predict the ratios within a family (e.g., K_α/K_β?)
Q34.3 Why should you identify the largest, high-energy X-ray peak first?
Q34.4 What factors determine whether or not an X-ray peak is visible?
Q34.5 What units give a good measure of the total efficiency of X-ray generation and detection?
Q34.6 Why should you do qualitative analysis anyhow?
Q34.7 What are the three kinds of resolution we worry about in XEDS analysis and how do they rank in importance?
Q34.8 Why is careful bookkeeping so important when doing qualitative analysis?
Q34.9 Why is it essential to calibrate your XEDS display and how often should you do it?
Q34.10 Why is Ray Dolby equally applauded by TEM and audio enthusiasts?
Q34.11 Why is it necessary to define a minimum criterion for peak visibility?
Q34.12 State the minimum visibility criterion.
Q34.13 Why would you choose a 99% confidence limit for peak visibility rather than a lesser value and what does this 99% confidence limit mean?
Q34.14 Why are you wasting your time if you gather spectra for very long times (such as overnight) in order to maximize the counts in the spectrum?
Q34.15 Why are characteristic-peak overlaps described as 'pathological'?
Q34.16 List a few key pathological overlaps that might affect you if you are a metallurgist.
Q34.17 Why do X-ray spectra get more complex as your specimens get higher in atomic number?
Q34.18 If you can't conclusively identify a small peak in the spectrum, what should you do?
Q34.19 If you can't conclusively identify a large peak in your spectrum, what should you do?
Q34.20 Can you think of an occasion when qualitative analysis might preclude the necessity for future quantitative analysis?

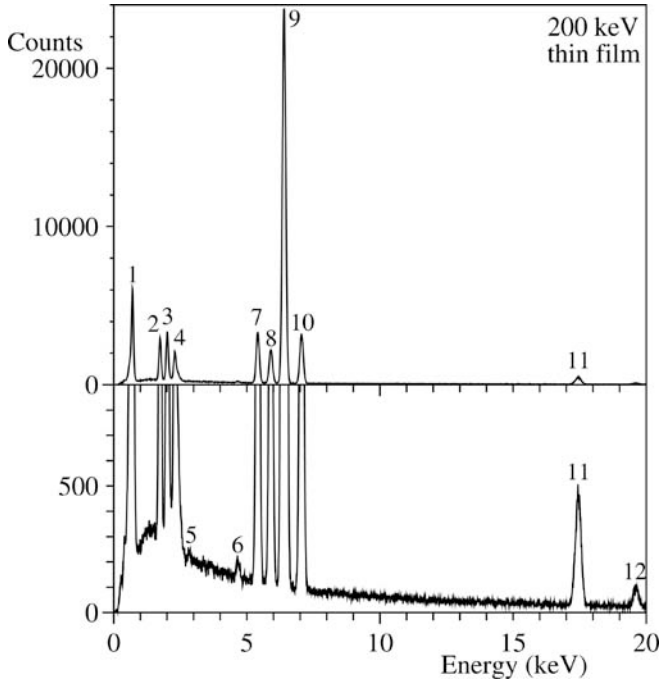
TEXT-SPECIFIC QUESTIONS

T34.1 Using Figure 34.2 as a basis, list the key characteristics of each of the principal families of lines that we see in XEDS spectra between 0 and 10 keV.
T34.2 Why don't we see N X-ray lines?
T34.3 There isn't an Fe K_β peak in Figure 34.5. Should you be concerned? If you are concerned, what should you do to lower your blood pressure?
T34.4 If you acquire spectra from a Ni jet-polished disc, why does the NiK_α/NiL_α ratio vary as you move the probe away from the thin edge surrounding the hole in the specimen?
T34.5 If you are analyzing an electropolished Al-Cu thin foil, why might the Cu signal increase substantially at the thinnest edges of the foil? (Hint: go back and look at Chapter 10.)
T34.6 Why is it difficult to analyze spectra from transition elements adjacent to one another in the periodic table, especially if the lower-Z transition metal is present in significantly greater amounts?
T34.7 When acquiring a qualitative spectrum from an unknown specimen, explain why you use as large a probe size as possible.
T34.8 When acquiring a qualitative spectrum from an unknown specimen, explain why you use as short a time constant as possible.
T34.9 When acquiring a qualitative spectrum from an unknown specimen, explain why you count for as long as possible.
T34.10 No. 13 of Murphy's Law of Microanalysis* states that 'The probability of detecting argon in aluminum decreases with time'. Justify the validity of this law. (*Copyright Kevex Inc.; reproduced with permission.)

T34.11 Look at the spectrum in Figure 34.6 and explain why

- you cannot confuse the Si K_α peak with the Si internal fluorescence artifact peak?
- the Cu K and L lines are from the Cu support grid and not from Cu in the specimen?
- the Cu L_α peak is present but not the Fe L_α peak?
- the Ar peak is not an artifact? Explain from whence it came.
- the Ca K_α and K_β peaks (at 3.69 and 4.01 keV, respectively) are not, in fact, the Sn L_a and L_β peaks (at 3.66 and 4.13 keV)?
- the Ca K_α and K_β peaks are not in fact the Te L_α and L_β peaks (at 3.77 and 4.03 keV)?
- there are no detectable escape and sum peaks?

T34.12 Look at the following spectra (which are really two different 'magnifications' of the vertical (counts) scale of the same spectrum to reveal both the small peaks and more intense peaks). This was taken from a thin-foil specimen in a 200-keV AEM with an ATW-XEDS Si(Li) detector. Carry out a qualitative analysis and identify unambiguously all the peaks labeled as #1–#12. When you have successfully finished the qualitative analysis of the above spectrum, use DTSA and try to reproduce this same spectrum.



.

35

Quantitative X-ray Analysis

CHAPTER PREVIEW

Now you've got an idea of how to acquire XEDS spectra and images from thin foils. You understand what factors limit the useful information they may contain and what false and misleading effects may arise. Also you know how to be very sure that a certain characteristic peak is due to the presence of a certain element and the occasions when you may not be so confident. Having obtained a spectrum or image that is qualitatively interpretable, it turns out to be a remarkably simple procedure to convert that information into quantitative data about the distribution of elements in your specimen; this is what we describe in this chapter.

This chapter is a little longer than the average. You may find that you can skip parts of it as you work through it the first time. We have decided to keep the material together so as to be a more useful reference when you are actually doing your analyses on your microscope. This aspect of XEDS is poised to change significantly in the future and the prospects for much improved quantification are hinted at and covered in detail in the companion text.

35.1 HISTORICAL PERSPECTIVE

Quantitative X-ray analysis in the AEM is a straightforward technique. What is surprising is that, given its simplicity, relatively few users take the trouble to extract quantitative data from their spectra, or produce quantitative images, despite the fact that numerical data are the basis for all scientific investigations. Before we describe the steps for quantification, you should know a little about the historical development of quantitative X-ray analysis, because this will emphasize the advantages of thin-foil analysis over analysis of bulk specimens, which was in fact the driving force for the development of the first commercial analytical TEMs.

Historically, X-ray analysis in electron-beam instruments started with the study of bulk specimens in which the electron beam was totally absorbed, as opposed to 'thin' specimens through which the beam penetrates. The possibility of using X-rays generated by a focused electron beam to give elemental information about the specimen was first described by Hillier and Baker in 1944, and the necessary instrumentation was built several years later by Castaing in 1951. In his extraordinary Ph.D. dissertation, Castaing not only described the equipment but also outlined the essential steps to obtain quantitative data from bulk specimens. The procedures that Castaing proposed still form the basis of the quantification routines used today in the EPMA and may be summarized as follows. Castaing assumed that the concentration C_i of an element i in the specimen generates a certain intensity of characteristic X-rays. However, it is very difficult in practice to measure this generated intensity so Castaing suggested that a known standard of composition $C_{(i)}$ be chosen for element i. We then measure the intensity ratio $I_i/I_{(i)}$

I_i is the measured intensity emerging from (not generated within) the **specimen**.
$I_{(i)}$ is the measured intensity emerging from the **standard**.

Castaing then proposed that, to a reasonable approximation

$$C_i/C_{(i)} = [K]I_i/I_{(i)} \tag{35.1}$$

where K is a sensitivity factor (not a constant) that takes into account the difference between the *generated* and *measured* X-ray intensities for both the standard and the unknown specimen. The contributions to K come from three effects

- Z The atomic number
- A The absorption of X-rays within the specimen
- F The fluorescence of X-rays within the specimen

The correction procedure in bulk analysis is often referred to as the ZAF correction. The necessary calculations, which have been refined over the years since

Castaing first outlined them, are exceedingly complex and best handled by a computer. If you're interested, there are several standard textbooks available which describe the ZAF and related procedures in detail, for example, those by Goldstein et al. and by Reed.

It was soon realized that, if an electron-transparent rather than a bulk specimen was used, the correction procedure could be greatly simplified. To a first approximation, in thin films, the A and F factors could be ignored and only the Z correction would be necessary. Furthermore, in thin specimens, the analysis volume would be substantially reduced, giving a much better spatial resolution. (We discuss this latter point in detail in the next chapter.)

These two obvious advantages of thin-foil analysis led to the development of the so-called Electron Microscope MicroAnalyzer (EMMA), pioneered by Duncumb in England in the 1960s. Unfortunately, the EMMA was far ahead of its time, mainly because the WDS was the only X-ray detector available. As we have seen back in Chapter 32, the classic WDS is handicapped by its poor collection efficiency, relatively cumbersome size, and slow, serial operation. These factors, particularly the poor efficiency, meant that a large probe size (~0.2 μm) had to be used to generate sufficient X-ray counts for quantification of the weak signal from thin foils and, therefore, the gain in spatial resolution over the EPMA was not so great. Also, the poor stability of the WDS meant that it was necessary to measure the beam current to make sure that the X-ray intensities from both standard and unknown could be sensibly compared. As a result of all these drawbacks, the EMMA never sold well and the manufacturer (AEI) soon went out of the EM business.

It is ironic that around this time the commercial developments that would transform TEMs into viable AEMs were all taking place. The XEDS detector was developed in the late 1960s, and commercial TEM/STEM systems appeared in the mid-1970s. However, before the demise of the EMMAs, they were to play a critical role in the development of the thin-foil analysis procedures that, surprisingly, we still use today. The EMMA at the University of Manchester, operated by Graham Cliff and Gordon Lorimer, was re-fitted with an XEDS system and they soon realized that the pseudo-parallel collection mode, the greater collection efficiency, and the improved stability of the XEDS removed many of the problems associated with WDS on the EMMA. Cliff and Lorimer (1975) showed that quantification was possible using a simplification of Castaing's original ratio equation, in which there was no need to incorporate intensity data from a standard, but simply ratio the intensities gathered from two elements simultaneously in the XEDS. This finding revolutionized thin-foil analysis and remains the basis for most quantifications today.

35.2 THE CLIFF-LORIMER RATIO TECHNIQUE

The basis for the Cliff-Lorimer technique is to rewrite equation 35.1 as a ratio of two elements A and B in a binary system. We have to measure the above-background characteristic intensities, I_A and I_B, simultaneously. This is trivial with an XEDS and, therefore, there is no need to measure the counts from a standard

THIN SPECIMENS

We assume that the specimen is thin enough so that we can ignore any absorption or fluorescence. This assumption is called the 'thin-foil criterion.'

The weight percents of each element C_A and C_B can then be related to I_A and I_B thus

$$\frac{C_A}{C_B} = k_{AB}\frac{I_A}{I_B} \tag{35.2}$$

This equation is the Cliff-Lorimer equation and the term k_{AB} is often termed the Cliff-Lorimer factor. As with K in equation 35.1, k_{AB} is a sensitivity factor, not a constant, so don't be fooled by the use of this letter. The k-factor varies according to your TEM/XEDS system and the kV you choose. Because we are ignoring the effects of absorption and fluorescence, k_{AB} is related only to the atomic-number correction factor (Z) in Castaing's original ratio equation. Now to obtain an absolute value for C_A and C_B, we need a second equation and, in a binary system, we simply assume that A and B constitute 100% of the specimen, so

$$C_A + C_B = 100\% \tag{35.3}$$

We can easily extend these equations to ternary and higher order systems by writing extra equations of the form

$$\frac{C_B}{C_C} = k_{BC}\frac{I_B}{I_C} \tag{35.4}$$

$$C_A + C_B + C_C = 100\% \tag{35.5}$$

You should also note that the k-factors for different pairs of elements AB, BC, etc., are related thus

$$k_{AB} = \frac{k_{AC}}{k_{BC}} \tag{35.6}$$

So long as you are consistent, you could define the composition in terms of atomic %, or weight fraction or any appropriate units. Of course the value of the k-factor would change accordingly.

Remember that Cliff and Lorimer developed the ratio approach to overcome the limitations of early

AEMs, particularly the low brightness of thermionic sources, the small collection angle of early detectors, electrical and mechanical instabilities (particularly the beam current), and the tendency of the early AEMs to contaminate the analysis area. Consequently, X-ray count rates were very low, thus limiting quantification. So it was very difficult to use the well-established, pure-element standards approach developed over the preceding 25 years for the EPMA because you'd have to keep changing specimens to measure the standards. In doing so, the probe current would change because, in a TEM, you have to switch off the beam to stop the vacuum changes burning out the gun (unlike in an EPMA where you can measure the current in situ and also isolate the gun automatically). Switching the gun on and off prevents any meaningful comparison of standards and unknowns. The ratio approach cancels out variations in the probe current incident on the analysis area; such variations arise from electron-gun/condenser-system instabilities, drift, and contamination buildup. Despite the fact that most of these problems have been minimized in modern AEMs, the ratio technique still remains the only quantitative thin-film analysis software available on commercial XEDS systems. Clearly, this is not an ideal situation and after describing the Cliff-Lorimer method, we'll discuss the ζ-factor, an alternative, improved approach, which combines the ease of application of the ratio method with the more rigorous aspects of pure-element (or other) thin-film standards (more about this in the companion text). The ζ-factor method requires in-situ measurement of the probe current, which, while standard on EPMAs for almost 50 years, has still not penetrated the design of commercial AEMs! So, despite its vintage, the Cliff-Lorimer equation remains the basis for quantitative analysis on all AEMs. Let's see how we use it in practice.

WT%

The convention is: define the units of composition as wt%.

35.3 PRACTICAL STEPS FOR QUANTIFICATION

First of all, you should try to use K_α lines, where possible, for the measured counts (I). (The K_β peak is combined with the K_α if the two K peaks cannot be resolved.) Use of L or M lines is more difficult because of the many overlapping lines in each family, but may be unavoidable if the K_α lines are too energetic and go right through your detector. (Think why you can't use L or M lines if the K lines are too *weak* for your detector.)

To gather characteristic X-ray intensities for quantification

- Keep your specimen as close to 0° tilt as possible to minimize spurious effects.
- If you have a wedge specimen, orient it so the thin portion of the wedge faces the detector, to minimize X-ray absorption (see Section 35.6).
- If the area of interest in your specimen is close to a strong two-beam dynamical diffraction condition, tilt it slightly to kinematical conditions.
- Accumulate enough counts in the characteristic peaks, I_A, I_B, etc. As we will see below, for acceptable errors, there should ideally be at least 10^4 counts above background in each peak.

While you can't always obtain 10^4 counts in a reasonable time before specimen drift, damage, or contamination limits your analysis, you should always choose the largest probe size which is consistent with maintaining the desired spatial resolution, so you get most current into your specimen. (Remember all the other ways to maximize the X-ray count rate that we discussed in Chapter 34.)

(The reason we worry a little about the diffraction conditions is that anomalous X-ray generation can occur across bend contours or whenever a diffracted beam is strongly excited. This point is not too critical because we quantify using a ratio technique. If the beam has a large convergence angle, which is usually the case, any diffraction effect is further reduced. However, we will see in Section 35.9 that, under certain conditions, there are some advantages to be gained from such crystallographic effects.)

Having accumulated a spectrum under these conditions, how do you quantify it? All you have to do is measure the peak intensities I_A, I_B, etc., and then determine a value for the k_{AB} factor. To determine the peak intensities, you first have to remove the background counts from the spectrum and then integrate the peak counts. Both of these steps are accomplished by various software routines in the XEDS computer system or in DTSA. There are advantages and disadvantages to each approach, so you should pick the one that is most suited to your problem.

35.3.A Background Subtraction

Remember, as we saw back in Section 4.2.B, we are not very precise in the terminology we use for the X-ray background intensity, so it can be confusing. 'Background' refers to the counts under the characteristic peaks in the spectrum displayed on your computer screen. These X-rays are generated by the 'bremsstrahlung' or 'braking-radiation' process as the beam electrons interact with the coulomb field of the nuclei in the specimen. The intensity distribution of the bremsstrahlung decreases continuously as the X-ray energy increases, reaching zero at the beam energy (go back

and look a, Figure 4.6). Thus, the energy distribution can be described as a 'continuum,' although as we've seen, the phenomenon of coherent bremsstrahlung disturbs this continuum.

TERMINOLOGY

We tend to use these three terms 'background,' 'bremsstrahlung,' and 'continuum' interchangeably, although strictly speaking they have these specific meanings.

Remember also that the generated bremsstrahlung intensity is modified at energies below about 1.5 keV by absorption within the specimen and the detector, so we are usually dealing with a background in the spectrum that looks something like Figure 35.1. The best approach to background subtraction depends on two factors

- whether the region of interest in your spectrum is in the low-energy regime, where the intensity decreases rapidly with decreasing energy.
- if the characteristic peaks you want to measure are close together or isolated.

Window methods: In the simplest case of isolated characteristic peaks superimposed on a slowly varying background, you can easily remove the background counts by drawing a straight line below the peak, and defining the background intensity as that present below the line, as shown in Figure 35.2. So you get the computer first to define a 'window' in the spectrum spanning the width of the peak, and then draw the line between

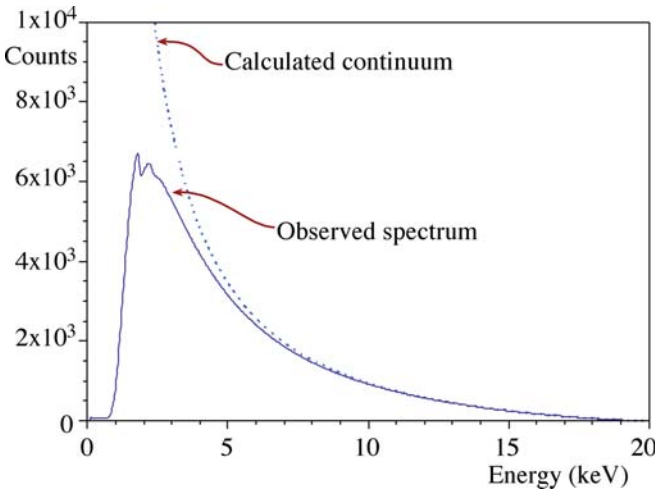


FIGURE 35.1. The theoretically calculated and experimentally observed bremsstrahlung intensity distribution as a function of energy. Both curves are similar until energies below ~2 keV when absorption within the specimen and the XEDS system reduces the detected counts. The best method of background removal depends on where in the spectrum your characteristic peaks are present.

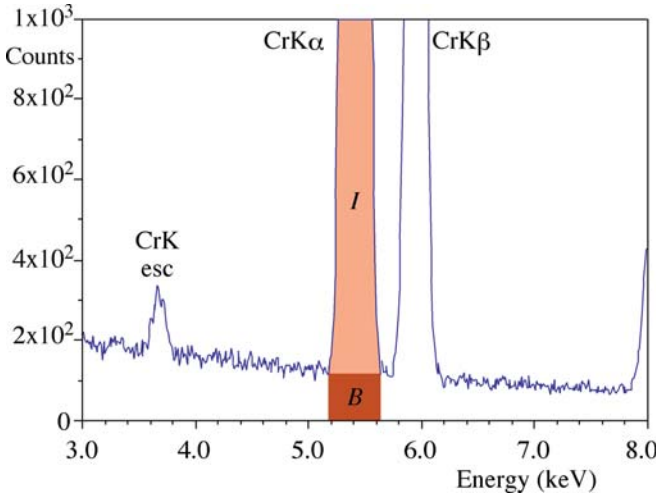


FIGURE 35.2. The simplest method of estimating the background contribution (B) to the counts in the characteristic peak (I); a straight line drawn beneath the Cr K_α peak provides a good estimate, if the counting statistics are good and the intensity approximates to a slowly varying function of energy. There should be no overlap with any other characteristic peak and the peak energy should be > ~2 keV.

the background intensities in the channels just outside the window. As with *all* spectral manipulations, this method gives better results with more counts in the spectrum. The background intensity variation is then less noisy, so it is easier to decide where the peak ends and the background begins and, furthermore, the background variation better approximates to a straight line.

Another, similarly primitive, approach involves averaging the bremsstrahlung counts above and below the characteristic peak by integrating the counts in two identical windows on either side of the peak, as shown in Figure 35.3. We then assume that the average of the two intensities equals the background counts under the

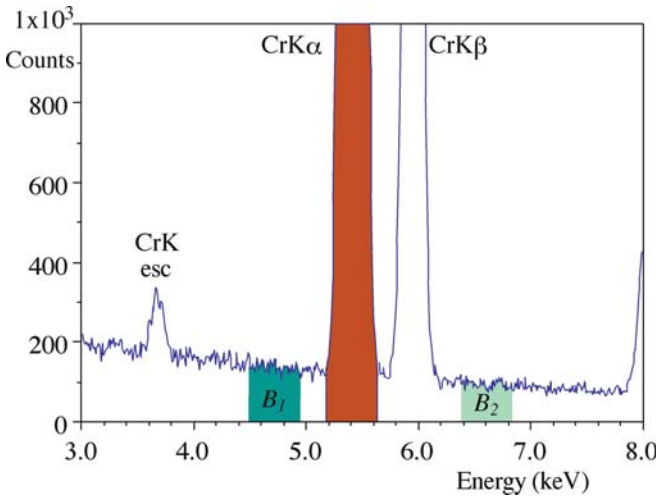


FIGURE 35.3. Background subtraction can be achieved by averaging the bremsstrahlung counts in two identical windows (B_1, B_2) on either side of the characteristic (Cr K_α and K_β) peaks. There should be no overlap with any other characteristic peaks and the peak energy should be >2 keV.

peak. This assumption is reasonable in the higher-energy regions of the spectrum and when the specimen is thin enough so that the bremsstrahlung is not absorbed in the specimen. This absorption happens because the bremsstrahlung X-rays with energies just above the peak energy preferentially fluoresce the characteristic-peak X-rays, resulting in a detectable reduction in the bremsstrahlung counts above the peak energy compared to below.

TOO THICK?

If you see this bremsstrahlung absorption effect in your spectrum, your specimen is too thick for Cliff-Lorimer quantification.

When you use the two-window approach, you must remember the window width you used, because *identical* windows must be used when subtracting the background both in the unknown spectrum and in the spectrum from the known specimen that you will use to determine the *k*-factor (see Section 35.4).

While the two techniques we just described have the advantage of simplicity, you can't always apply them to real specimens because the spectral peaks may overlap. Also, if the peaks lie in the low-energy region of the spectrum where the background is changing rapidly due to absorption, then neither of these two simple methods gives a good estimate of the background and more sophisticated mathematical approaches are required. We'll now discuss these methods.

Modeling the background: the bremsstrahlung distribution can be mathematically modeled, based on the expression developed by Kramers (1923). The number (N_E) of bremsstrahlung photons of energy E produced in a given time by a given electron beam is given by Kramers' law.

$$N_E = KZ\frac{(E_0 - E)}{E} \tag{35.7}$$

Here Z is the *average* atomic number of the specimen, E_0 is the beam energy in keV, and E is the X-ray energy in keV. The factor K in Kramers' law actually takes account of numerous parameters. These include

- Kramers' original constant.
- The collection efficiency of the *detector*.
- The processing efficiency of the *detector*.
- The absorption of X-rays within the *specimen*.

OPTIMUM WINDOW

The typical choice of window width is FWHM, but this throws away a substantial amount of the counts in the peak. FWTM gives better statistics, but incorporates more bremsstrahlung than characteristic counts; 1.2(FWHM) is the optimum window.

All these terms have to be factored into the computer calculation when you use this method of background modeling.

Be wary when using this approach because Kramers developed his law for bulk specimens. However, the expression is still used in commercial software, and seems to do a reasonable job.

Modeling the spectrum produces a smooth curve fit that describes the shape of the complete spectrum. This approach is particularly valuable if many characteristic peaks are present, since then it is difficult to make local measurements of the background counts by a window method. Figure 35.4 shows an example of a spectrum containing many adjacent peaks, with the background counts estimated underneath all the peaks.

Filtering out the background: another mathematical approach to removing the background uses digital filtering. This process makes no attempt to take into account the physics of X-ray production and detection as in Kramers' law. Rather it relies on the fact that the characteristic peaks show a rapid variation of counts as a function of energy (i.e., dI/dE is large), while the background exhibits a relatively small dI/dE. This approximation is valid even in the region of the spectrum below ~ 1.5 keV where absorption is strong. In the process of digital filtering, the spectrum intensity is filtered by convoluting it with another mathematical function. The most common function used is a 'top-hat' filter function, so called because of its shape. When the top-hat filter is convoluted with the shape of a typical X-ray spectrum, it acts to produce a second-difference spectrum, i.e. d^2I/dE^2 versus E. After the top-hat filter, the background with small dI/dE is transformed to a linear

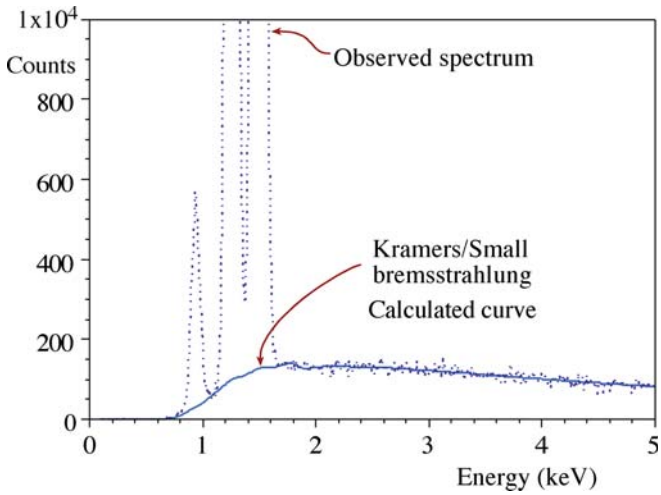


FIGURE 35.4. The bremsstrahlung intensity modeled using Kramers' law, modified by Small to include the effects of absorption of low-energy X-rays in the specimen and the detector. This method is useful when the spectrum contains many overlapping peaks, particularly in the low-energy range, such as the Cu L_α and the Mg and Al K_α lines shown in this spectrum (based on Chapter 34, can you determine which peak is which?).

function with a value of zero (thus it is 'removed'), while the peaks with large dI/dE, although distorted to show negative intensities in some regions, are essentially unchanged as far as the counting statistics are concerned. Figure 35.5A shows schematically the principle behind the filtering process and Figure 35.5B,C shows an example of a spectrum before and after digital filtering.

BACKGROUND REMOVAL

You must always take care to apply the same background-removal process to both the standard and the unknown.

In summary, you can remove the background by selecting appropriate windows to estimate the counts in the peak, or use one of two mathematical-modeling approaches. The window method is generally good enough if the peaks are isolated and on a linear portion of the background. The mathematical approaches are most useful for multi-element spectra and/or those containing peaks below ~1.5 keV. You should choose the method that gives you the most reproducible results (check this on a specimen for which you know the composition).

After removing the background, you have to integrate the peak intensities I_A, I_B, etc.

35.3.B Peak Integration

If you used a window method of background estimation, then the peak counts are obtained simply by subtracting the estimated background counts from the total counts in the chosen window. Therefore, if the computer drew a line under the peak as in Figure 35.2 then the peak intensity is that above the line.

- If you chose, e.g., an ideal window of 1.2 FWHM and averaged the background on either side of the peak then the average value must be subtracted from the total counts in the 1.2 FWHM window; always use the same window width for B and I.

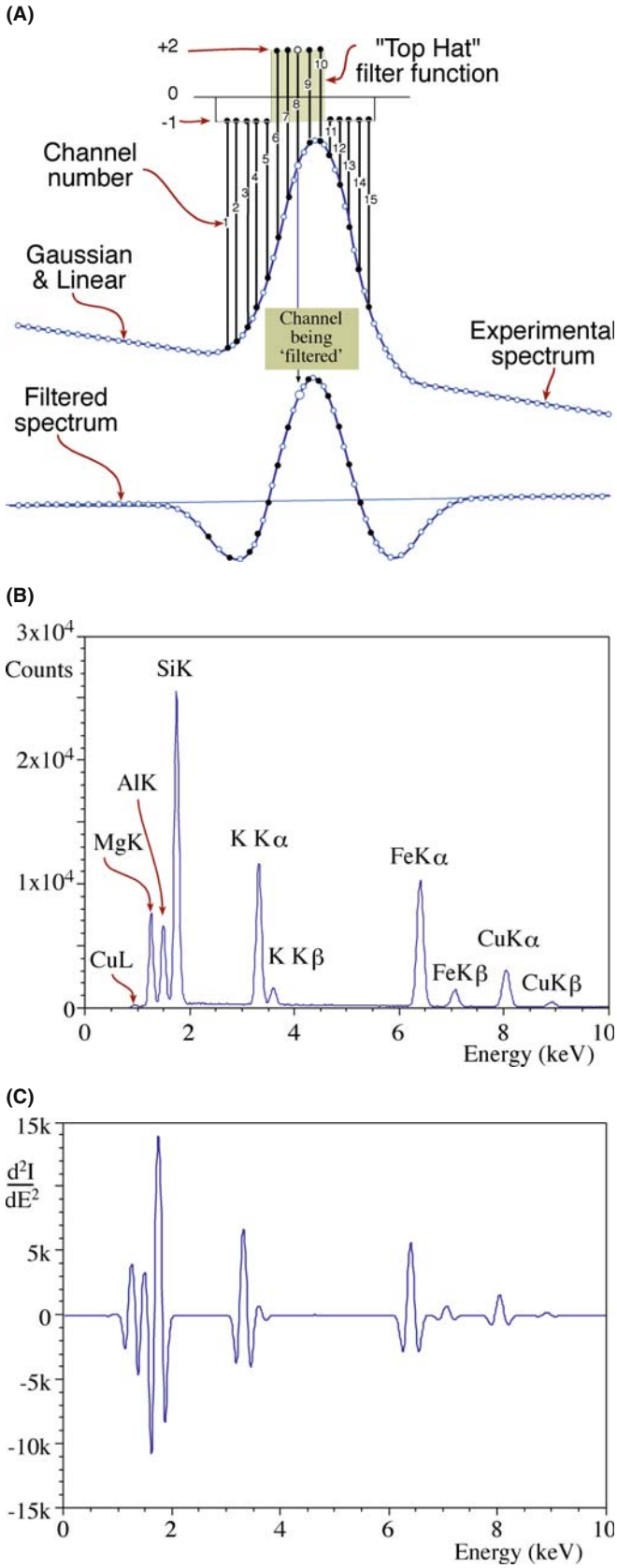


FIGURE 35.5. (A) Digital filtering involves convolution of a top-hat filter function with the acquired spectrum. To obtain the filtered spectrum, each channel has the top-hat filter applied to it. The channels on either side of that being filtered (#8 in this case) are multiplied by the appropriate number in the top-hat function. So channels 1–5 and 11–15 are multiplied by –1 and channels 6–10 by +2. The sum of the multiplications is divided by the total number of channels (15) and allotted to channel #8 in the filtered spectrum at the bottom. The digital filtering process in (A) applied to a spectrum from biotite (B) results in the filtered spectrum (C) in which the background intensity is zero at all places, and the characteristic peaks remain effectively unchanged.

- If you used a Kramers' law fit, the usual method of peak integration is to get the computer to fit the peak with a slightly modified Gaussian, and integrate the total counts in the channels under the Gaussian.
- If a digital filter was used, you have to compare the peaks with those that were taken previously from standards, digitally filtered, and stored in a library in the computer. The library peaks are matched to the experimental peaks via a multiple least-squares fitting procedure and the counts determined through calculation of the fitting parameters.

Each of the two curve-matching processes is rapid. Each can be used to deconvolute overlapping peaks and each uses all the counts in the peak. The Kramers-law fit and the digital filter have much wider applicability than the simple window methods. However, these computer processes are not invariably the best, nor are they without error.

The Gaussian curve fitting must be flexible enough to take into account several variables

- The peak width can change as a function of energy or as a function of count rate.
- The peak distortion due to incomplete charge collection can vary.
- There may be an absorption edge under the characteristic peak if your specimen is too thick.

The creation of a library of spectra gathered under conditions that match those liable to be encountered during analysis (particularly similar count rates and dead times) is a tedious exercise. However, you do get a figure of merit for the 'goodness of fit' between the unknown spectrum and the standard. Usually, a χ^2 value is given which has no absolute significance, but is a most useful diagnostic tool. Typically, the χ^2 value should be close to unity for a good fit, although a higher value may merely indicate that some unidentified peaks were not accounted for during the matching process. What you have to watch out for is a sudden increase in χ^2 compared with previous values. This indicates that something has changed from your previous analyses. Perhaps your standard is not giving a good fit to the experimental spectrum and either a new library spectrum needs to be gathered or the experimental peak should be looked at carefully. For example, another small peak may be hidden under the major peak and would need to be deconvoluted from the major peak before integration proceeds. If you suspect a poor fit, you should make the computer display the 'residuals,' that is the counts remaining in the spectrum after the peak has been integrated and removed. As shown in Figure 35.6, you can easily see if a good fit was made (Figure 35.6A) or if the library peak and the experimental peak do not match well (Figure 35.6B).

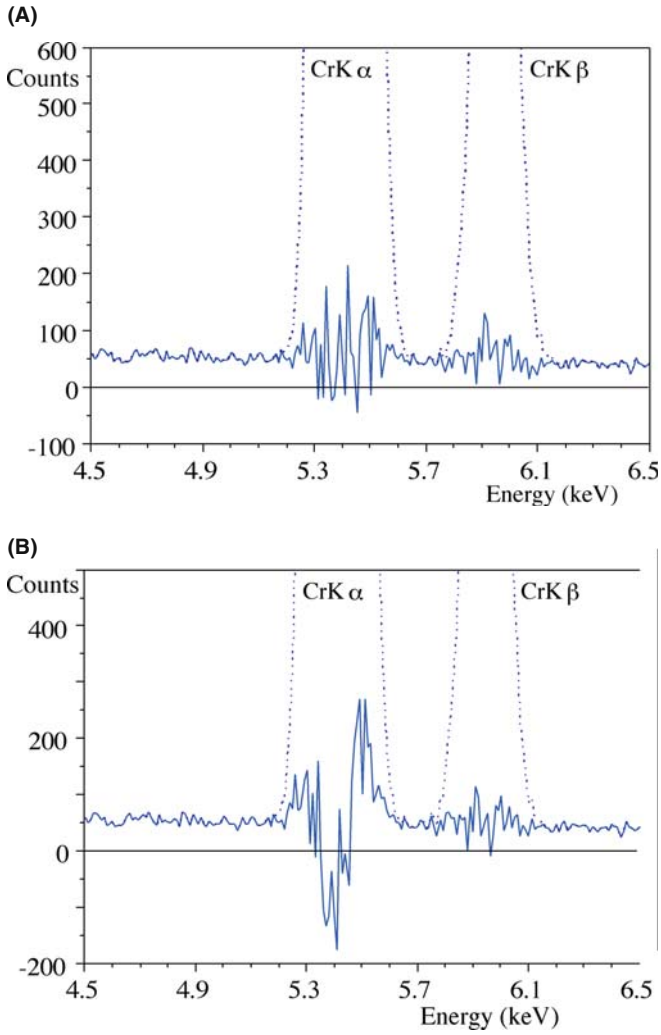


FIGURE 35.6. (A) A filtered Cr K-line family spectrum showing the residual background counts after the peaks have been removed for integration. The approximately linear residual intensity distribution indicates that the peaks matched well with the library standard stored in the computer. (B) A similar filtered spectrum showing the distorted residual spectrum characteristic of a poor fit with the library standard.

χ^2 OR CHI-SQUARED

Don't fear the math or the statistics. You'll rarely ever need to repeat it but you should know what your software is doing.

Any of the above methods is valid for obtaining values of the peak counts. They should all result in the same answer when used to quantify an unknown spectrum, so long as you apply the same method consistently to both the standard and the unknown.

Statham has reviewed the limitations of extracting peak intensities from X-ray spectra, with particular emphasis on low-keV lines. While aimed primarily at the EPMA community, almost all the issues in this paper are relevant to thin-film quantification.

Having obtained the peak counts, the next step is to insert the values into the Cliff-Lorimer equation and know the correct value of the k-factor. So we now need to discuss the various ways to obtain k_{AB}.

35.4 DETERMINING *k*-FACTORS

Remember that the k-factor is **not** a constant. It is a sensitivity factor that will vary not only with the X-ray detector, the microscope, and the analysis conditions, but also with your choice of background subtraction and peak-integration methods. So values of k-factors can be sensibly compared *only* when they were obtained under identical conditions. We will return to this point at the end of this section when we look at various sets of k-factors published in the literature. There are two ways you can determine k-factors

- Experimental determination using standards.
- Calculation from first principles.

The first method is slow and laborious but gives the most accurate values. The second method is quick and painless but the results are less reliable. You might wonder—can k-factors even depend on the previous user of the TEM?

k-FACTOR VERSUS ζ FACTOR

There are no generally accepted standards that meet all the above criteria for ideal k-factor determination, which is a major limitation to this approach and is overcome by the pure-element standards used in the ζ-factor method, which we describe later.

35.4.A Experimental Determination of k_{AB}

If you have a thin specimen of known composition, C_A, C_B etc., then all you have to do, in principle, is place that specimen in the microscope, generate a spectrum, obtain values of I_A, I_B, etc., and insert those values in the Cliff-Lorimer equation 35.2. Since you know C_A and C_B the only unknown is k_{AB}. However, there are several precautions that you must take before this procedure can be used

- The standard must be a well-characterized specimen, and it is usually best if it is single phase.
- The standard must be capable of being thinned to electron transparency. Ideally, when the specimen is thin there should be no significant absorption or fluorescence of the X-rays from the elements A, B, etc., that you wish to analyze.
- You must be sure that the thinning process did not induce any chemical changes (this is discussed in some detail in Chapter 10).
- It must be possible to select thin regions that are characteristic of the chemistry of the bulk specimen.
- You must be sure that the thin foil is stable under the electron beam at the voltage you intend to use for analysis.

This last point may often be the limiting factor in your choice of standards because, as we saw in Section 4.6, you have to take care to avoid not only direct knock-on damage, but also sputtering effects, which occur at voltages substantially below the threshold for direct atom displacement. Obviously, both these problems become greater as the beam voltage increases.

The National Institute of Standards and Technology (NIST) has issued a thin standard containing the elements Mg, Si, Ca, and Fe and O (SRM #2063 and subsequently #2063A). Unfortunately, X-rays from the lighter elements in this standard film are absorbed significantly in the film, and also in the detector and so a correction to the measured k-factor is necessary and NIST has not re-issued the standard.

It is best to use your own judgment in choosing standards, and also make use of the knowledge gained in previous k-factor studies.

Cliff and Lorimer's approach using mineral standards had three advantages.

- Crushing is an easy way to make thin flakes and does not affect the chemistry.
- The mineral stoichiometry is usually well known.
- The minerals chosen all contained Si, thus permitting the creation of a whole series of k_{ASi} factors.

The drawbacks are that the mineral specimens often contain more than one phase, or may be naturally non-stoichiometric. Clearly, some prior knowledge of the mineralogy of the specimen is essential in order to be able to select the right spectrum to use as a standard. Also Si K_α X-rays at ~1.74 keV are liable to be absorbed in the XEDS detector, so there may be a systematic difference in k-factors determined with different detectors. Finally, silicate minerals often exhibit radiolysis i.e., chemical changes due to beam-induced breaking of bonds.

GLASS STANDARDS

Can be made completely homogeneous. Reproducible (one batch). No channeling complications.

Several alternative approaches have been proposed that attempt to avoid the problems with k_{ASi}

- Wood et al. generated a series of k_{AFe} factors to overcome the Si absorption and the beam-sensitivity problems.
- Graham and Steeds used crystallized microdroplets which were routinely thin enough and retained their stoichiometry.
- Kelly used rapidly solidified droplets for the same reasons.
- Sheridan demonstrated the value of the NIST multi-element glasses.
- NIST created its own standard multi-element glass, as already noted.

Note also that there has been no new systematic determination of k-factors for many years, indicating the mature nature of the k-factor approach. But remember that you need factors for your TEM, your kV, etc.

DETERMINING *k*-FACTORS

A typical k-factor determination involves taking many spectra from different parts of the thin-foil standard. You must check both the homogeneity and the stability of the specimen. This is very time consuming, which is why so few analysts bother to do it properly.

In all cases, the bulk chemistry of the k-factor standard has to be determined by some technique with known accuracy, such as EPMA, atomic-absorption spectroscopy, or wet chemistry. Since all these techniques analyze relatively large volumes of material, it is best that the standard be single phase. However, because none of these techniques can determine if the specimen is homogeneous on a sub-micrometer scale, the only way to find out the level of homogeneity is to carry out many analyses within the AEM to confirm that any variation in your answer is within the expected X-ray statistical fluctuations.

Each spectrum should contain sufficient counts in the peaks of interest to ensure that the errors in the k-factor determination are at least less than $\pm 5\%$ relative and, if possible, less than $\pm 3\%$. So, now we need to consider the errors associated with the X-ray spectra.

35.4.B Errors in Quantification: The Statistics

An unfortunate aspect of the simple Cliff-Lorimer ratio equation is that it has relatively large errors associated with it. The very nature of the thin foil minimizes the problems of absorption and fluorescence, but also generates relatively few X-ray photons per incident electron, compared with bulk specimens. This effect is compounded by the small collection angle of the XEDS detector and the end result is that poor counting statistics are the primary source of error in most AEM quantifications. The best way you can limit these errors is to use higher-brightness sources, large electron probes, and thicker specimens (unless absorption is a problem, or spatial resolution is paramount) and, of course, C_s-correction of the probe helps. In any case you should be prepared to count for a long time, assuming that specimen drift and/or contamination don't compromise your data.

The rest of this section is pure statistics. If you know it, then jump ahead.

GAUSSIAN STATISTICS

Experimental results show that the X-ray counts in the spectrum obey Gaussian statistics. Hence, we can apply simple statistics to deduce the accuracy of any quantification.

Given that our characteristic peak is Gaussian, then the standard deviation σ is obtained from

$$\sigma = N^{\frac{1}{2}} \tag{35.8}$$

where N is the number of counts in the peak above the background. For a single measurement, there is a 67% chance that the measured value of N will be within 1σ of the true value of N. This chance increases to 95% for 2σ and 99.7% for 3σ. If we use the most stringent condition, then the relative error in any single measurement is

$$\text{Relative Error} = \frac{3N^{\frac{1}{2}}}{N}100\% \tag{35.9}$$

Clearly, the error decreases as N increases and hence, the emphasis throughout this chapter on the need to maximize the X-ray counts gathered in your spectra. Since the Cliff-Lorimer equation uses an intensity ratio, we can get a quick estimate of the error by summing the errors in I_A, I_B, and k_{AB} to give the total error in the composition ratio C_A/C_B.

Summing the errors in fact gives an overestimate of the error. Strictly speaking, we should add the standard deviations of the various terms in the Cliff-Lorimer equation in quadrature to give the standard deviation in the composition-ratio measurement σ_C using the expression

$$\left(\frac{\sigma_C}{C_A/C_B}\right)^2 = \left(\frac{\sigma_{k_{AB}}}{k_{AB}}\right)^2 + \left(\frac{\sigma_{I_A}}{I_A}\right)^2 + \left(\frac{\sigma_{I_B}}{I_B}\right)^2 \tag{35.10}$$

So we can determine the error for each datum point in this manner. If we are determining the composition of a single-phase region (for example, when determining a k-factor) then we can reduce the error by combining

the results from n different measurements of the intensity ratio I_A/I_B. The total absolute error in I_A/I_B at a given confidence limit is obtained using the student's t distribution. For example, in this approach the error is given by

$$\text{Absolute Error} = \frac{(t_{95})^{n-1} S}{n^{\frac{1}{2}}} \qquad (35.11)$$

where $t_{95}{}^{n-1}$ is the student's t value at the 95% confidence limit for n measurements of k_{AB}. You can find lists of student's t values in any statistics text (see, e.g., Larsen and Marx) or on the Web (e.g., URL #1). Obviously, you could choose a lower or higher confidence level. S is the standard deviation for n measurements where

$$S = \left(\sum_{n=1}^{n} \frac{(N_i - N)^2}{n-1} \right)^{1/2} \qquad (35.12)$$

So by increasing the number of measurements n, you can reduce the absolute error in k_{AB}. With enough measurements and a good homogeneous specimen, you can reduce the errors in the value of k_{AB} to $\pm 1\%$, as we will see in the example below. However, remember that this figure must be added to the errors in I_A and I_B. From equation 35.8 it is easy to determine that if we accumulate 10,000 counts in the peak for element A then the error at the 99% confidence limit is $[3\ (10{,}000)^{1/2}/10{,}000] \times 100\%$, which is $\sim 3\%$. Using equation 35.10 and a similar value for I_B you get a total error in C_A/C_B of $\sim \pm 4.5\%$. Now you see again why 'counts, counts, and more counts' is the mantra for thin-foil analysts.

If you take the time to accumulate 100,000 counts for I_A and I_B the total error is reduced to $\sim \pm 1.7\%$, which represents about the best accuracy that can be expected for quantitative XEDS analysis in the AEM. This is an accuracy that almost no one ever takes the time to achieve.

It is appropriate here to go through an illustration of a k_{AB} determination using experimental data. Before deciding that a particular specimen is suitable, it should be checked for its level of homogeneity: there is a well-established criterion for this. If we take the average value N of many composition determinations, and all the data points fall within $\pm 3(N)^{1/2}$ of N then the specimen is homogeneous. In other words, this is our definition of 'homogeneous.' There are more rigorous definitions but the general level of accuracy in thin-foil analysis is such that there is no need to be so stringent.

An example: A homogenized thin foil of Cu-Mn solid solution was used to determine k_{CuMn}. The specimen was first analyzed by EPMA and found to be 96.64 wt% Cu and 3.36 wt% Mn. Since our accuracy is increased by collecting many spectra, a total of 30 were accumulated ($n = 30$ in equation 35.12). In a typical spectrum the Cu K_α peak contained 271,500 counts above background and the Mn K_α peak contained 10,800 counts. So if we insert these data into the Cliff-Lorimer equation we get

$$\frac{96.67}{3.36} = k_{CuMn} \frac{271{,}500}{10{,}800}$$

$$k_{CuMn} = 1.14.$$

To determine an error on this value of the k-factor, equation 35.11 must be used. The student's t analysis of the k-factors from the other 29 (yes 29!) spectra gives an error of $\pm$ 0.01 for a 95% confidence limit. This error of about $\pm 1\%$ relative is about the best that can be achieved using the experimental approach to k-factor determination, but remember that 30 individual spectra had to be accumulated from different regions of a well-characterized thin foil.

STUDENT'S *t* DISTRIBUTION

More statistics but with much more interesting origins.

So it takes a real effort to get quantitative data with errors $< \pm$ 5–10% relative. Bear this in mind when you read any paper in which compositions are given to better than a rounded $\pm$ 5% (such as the EPMA data in our own example).

THIN-FOIL COMPOSITIONS

Any thin-foil composition data given with decimal points must be really scrutinized to see if they were obtained via a procedure such as that outlined here.

Otherwise they are just plain wrong.

Tables 35.1 and 35.2 summarize many of the available k-factor data in the published literature. You should go and read the original papers, particularly if you want to find out what standards and what conditions were used in their determination.

35.4.C Calculating k_{AB}

While it is clear that many of the values in these k-factor tables are very similar, the differences cannot be

TABLE 35.1 Experimentally Determined k_{ASi} and k_{AFe} Factors for K_α X-rays

Element (A)	k_{ASi} (1) 100 kV	k_{ASi} (2) 100 kV	k_{ASi} (3) 120 kV	k_{ASi} (4) 80 kV	k_{ASi} (5) 100 kV	k_{ASi} (5) 200 kV	k_{AFe} (6) 120 kV	k_{ASi} (7) 200 kV
Na	5.77	3.2	3.57 ± 0.21	2.8 ± 0.1	2.17	2.42		3.97 ± 2.32
Mg	2.07 ± 0.1	1.6	1.49 ± 0.007	1.7 ± 0.1	1.44	1.43	1.02 ± 0.03	1.81 ± 0.18
Al	1.42 ± 0.1	1.2	1.12 ± 0.03	1.15 ± 0.05			0.86 ± 0.04	1.25 ± 0.16
Si	1.0	1.0	1.0	1.0	1.0	1.0	0.76 ± 0.004	1.00
P			0.99 ± 0.016				0.77 ± 0.005	1.04 ± 0.12
S			1.08 ± 0.05		1.008	0.989	0.83 ± 0.03	1.06 ± 0.12
Cl					0.994	0.964		1.06 ± 0.30
K		1.03	1.12 ± 0.27	1.14 ± 0.1			0.86 ± 0.014	1.21 ± 0.20
Ca	1.0 ± 0.07	1.06	1.15 ± 0.02	1.13 ± 0.07			0.88 ± 0.005	1.05 ± 0.10
Ti	1.08 ± 0.07	1.12	1.12 ± 0.046				0.86 ± 0.02	1.14 ± 0.08
V	1.13 ± 0.07			1.3 ± 0.15				1.16 ± 0.16
Cr	1.17 ± 0.07	1.18	1.46 ± 0.03				0.90 ± 0.006	
Mn	1.22 ± 0.07	1.24	1.34 ± 0.04				1.04 ± 0.025	1.24 ± 0.18
Fe	1.27 ± 0.07	1.30	1.30 ± 0.03	1.48 ± 0.1			1.0	1.35 ± 0.16
Co							0.98 ± 0.06	1.41 ± 0.20
Ni	1.47 ± 0.07	1.48	1.67 ± 0.06				1.07 ± 0.006	
Cu	1.58 ± 0.07	1.60	1.59 ± 0.05		1.72	1.50	1.17 ± 0.03	1.51 ± 0.40
Zn	1.68 ± 0.07				1.74	1.55	1.19 ± 0.04	1.63 ± 0.28
Ge	1.92							1.91 ± 0.54
Zr								3.62 ± 0.56
Nb							2.14 ± 0.06	
Mo	4.3		4.95 ± 0.17				3.8 ± 0.09	
Ag	8.49		12.4 ± 0.63				9.52 ± 0.07	6.26 ± 1.50
Cd	10.6				9.47	6.2		
In								7.99 ± 1.80
Sn	10.6							8.98 ± 1.48
Ba					29.3	17.6		21.6 ± 2.6

TABLE 35.2 Experimentally Determined k_{ASi} and k_{AFe} Factors for L X-rays

Element (A)	k_{ASi} (8) 100 kV	k_{ASi} (5) 100 kV	k_{ASi} (5) 200 kV	k_{ASi} (9) 100 kV	k_{AFe} (6) 120 kV	k_{ASi} (7) 200 kV
Cu		8.76	12.2			
Zn		6.53	6.5			8.09 ± 0.80
Ge						4.22 ± 1.48
As						3.60 ± 0.72
Se						3.47 ± 1.11
Sr					1.21 ± 0.06	
Zr					1.35 ± 0.1	2.85 ± 0.40
Nb					0.9 ± 0.06	
Mo				2.0		
Ag	2.32 ± 0.2				1.18 ± 0.06	2.80 ± 1.19
In					2.21 ± 0.07	2.86 ± 0.71
Cd		2.92	2.75			
Sn	3.07 ± 0.2					
Ba		3.38	2.94			3.36 ± 0.58
Ce				1.4		
Sn	3.1 ± 0.2			1.3		
W	3.11 ± 0.2			1.8		3.97 ± 1.12
Au	4.19 ± 0.2	4.64	3.93		3.1 ± 0.09	4.93 ± 2.03
Pb	5.3 ± 0.2	4.85	4.24	2.8		5.14 ± 0.89

All L-line *k*-factors use the total L counts from the L_α and L_β lines.

Sources: (1) Cliff and Lorimer (1975), (2) Wood et al. (1981), (3) Lorimer et al. (1977), (4) McGill and Hubbard (1981), (5) Schreiber and Wims (1981), (6) Wood et al. (1984), (7) Sheridan (1989), (8) Goldstein et al. (1977), (9) Sprys and Short (1976).

accounted for by X-ray statistics alone. Some of the differences arise due to the choice of standard and the reproducibility of the standard. Other differences arise because the data were obtained under different conditions, such as different peak-integration routines. Therefore, the point made at the beginning of this section is worth repeating: the *k*-factors are not standards; they are sensitivity factors.

The only conditions under which you can expect the *k*-factors obtained on different AEMs to be identical are if you use the *same* standard at the *same* accelerating voltage, *same* detector configuration, *same* peak integration, and *same* background-subtraction routines. Even then there will be differences if one or more of the measured X-ray lines are not gathered by the detector with 100% efficiency; the X-ray may be either absorbed by the detector or it may be too energetic and pass straight through the detector.

You may not be able to obtain a suitable standard. For example, you might be working in a system in which no stoichiometric phases exist. You might not need a standard because accuracy might not be critical but you still need a quick analysis. Under such circumstances you can calculate an approximate *k*-factor. The programs necessary to calculate k_{AB} are stored in the XEDS computer and will give you a value of k in a fraction of a second. The calculated value should be accurate to within ±20% relative. This level of accuracy might be all you need to draw a sensible conclusion about the material you are examining (in which case, you can almost rely on a simple peak-height measurement). In general, if you can avoid the tedious experimental approach, then do so.

QUICK

Calculating *k*-factors is the recommended approach when a quick answer is required and the highest accuracy is not essential.

The expression for calculating the *k*-factor from first principles is derived in the paper by Williams and Goldstein. The derivation gives a good illustration of the relationship between bulk and thin-film analysis, and provides insight into the details of X-ray interactions with solids. However, at this stage you don't need to know the details of this derivation so we'll simply state the final expression

$$k_{AB} = \frac{1}{Z} = \frac{(Q\omega a)_B A_A}{(Q\omega a)_A A_B} \qquad (35.13)$$

The subscripts A, B denote the elements A, B of atomic weight A_A and A_B. This expression derives from the physics of X-ray generation. An electron passing close to an atom in the specimen has first to ionize that atom and this is governed by Q_A, the ionization cross section (sometimes given by σ; go back and check equation 4.1). An ionized atom does not necessarily give off a characteristic X-ray when it returns to ground state and the fraction of ionizations that do generate an X-ray is governed by the fluorescence yield for the characteristic X-rays, ω_A (go back and check equation 4.6) The remaining term '*a*' is the relative-transition probability. This term takes account of the fact that if a K-shell electron is ionized and returns to ground state through X-ray emission, it can emit either a K_α or K_β X-ray. You may remember that we listed the relative weights of the various K, L, and M families of X-ray lines in Table 4.1.

THE *k*-FACTOR CALCULATION

Elements A and B; atomic weight A_A and A_B
Ionization cross sections Q_A and Q_B
Fluorescence yield ω_A and ω_B
Relative transition probability, a
Detector efficiency, ε_A and ε_B

As we mentioned at the start of the discussion on quantification, the Cliff-Lorimer *k*-factor for thin-foil analysis is related to the atomic-number correction factor (Z) for bulk specimen analysis. From equation 35.13, we can easily see what experimental factors determine the value of k

- The accelerating voltage is a variable since Q is strongly affected by the kV.
- The atomic number affects ω, A, and a.
- The choice of peak-integration method will also affect a.

Therefore, in order to calculate and compare different *k*-factors, it is imperative to define these conditions very clearly, as we have taken pains to emphasize.

Equation 35.13 assumes that equal fractions of the X-rays generated by elements A and B are collected and processed by the detector. This assumption will only be true if the same detector is used and the X-rays are neither strongly absorbed by, nor pass completely through, the detector. However, as we have already seen in Chapter 32, X-rays below ~1.5 keV are absorbed significantly by Be window and X-rays above ~20 keV pass through a 3 mm Si detector with ease. Under these circumstances, it is necessary to modify the *k*-factor expression, equation 35.13, in the following manner

$$k_{AB} = \frac{1}{Z} = \frac{(Q\omega a)_A}{(Q\omega a)_B}\frac{A_B}{A_A}\frac{\varepsilon_A}{\varepsilon_B} \qquad (35.14)$$

The symbol ε represents simply a detector efficiency term plotted back in Figure 32.7 that we can write as follows.

$$\varepsilon_A = \exp\left(-\left[\frac{\mu}{\rho}\right]_{Be}^{A}\rho_{Be}t_{Be}\right)\exp\left(-\left[\frac{\mu}{\rho}\right]_{Au}^{A}\rho_{Au}t_{Au}\right)$$
$$\exp\left(-\left[\frac{\mu}{\rho}\right]_{Si}^{A}\rho_{Si}t_{Si}\right)\left\{1-\exp\left(-\left[\frac{\mu}{\rho}\right]_{Si}^{A}\rho_{Si}t'_{Si}\right)\right\} \quad (35.15)$$

The first term accounts for absorption (via the mass absorption coefficients, μ/ρ) of X-rays from element A passing through the Be window. It should of course be modified for different windows and it disappears for windowless detectors. The second term covers absorption in the Au contact layer and the third term accounts for the Si dead layer. These two terms will have different values for different contact elements, IG dead layers, and SDDs which have very thin dead layers and side contact layers. The last term adjusts the *k*-factor for X-rays that *do not* deposit their energy in the active region of a detector which has density ρ and thickness t'. Typical values of t' (~ 3 mm for Si(Li) and IG but ~ 1 mm for SDDs) were discussed in Chapter 32. An IG detector will more efficiently stop high-energy X-rays, since it is designed to detect them preferentially, while an SDD will be less efficient since it is generally much thinner than a Si(Li) or IG. In fact, much of the effort over the last 20 years to improve detector technology that we discussed in detail in Chapter 32, minimizes the effects of equation 35.15 on the *k*-factor.

While equations 35.14 and 35.15 are simple for a computer to solve, the values that have to be inserted in the equations for the various terms are not always well known, or cannot be measured accurately. For example, we do not know the best value of Q for many elements in the range of voltages typically used in the AEM (100–400 kV). There are considerable differences of opinion in the literature concerning the best way to choose a value for Q. The two major approaches used are

- Assume various empirical parameterization processes (e.g., Powell).
- Interpolate values of Q to give the best fit to experimental *k*-factors (Williams et al.).

The other major variable in equation 35.15 is the Be-window thickness which is nominally 7.5 μm but in practice may be 3–4× thicker. Tables 35.3A and 35.3B list calculated *k*-factors obtained using various expressions for Q. As you can see, the value of k may easily vary by $>\pm 10\%$, particularly for the lighter and the heavier elements. This variation is due to the uncertainties in the detector-efficiency terms in equation 35.14. The values of k_{AB} for the L lines are even less accurate than for the K lines mainly because the values of Q for the L lines are somewhat speculative. There are no data available for calculated *k*-factors for M lines. Under these circumstances, experimental determination is the only approach (or choose different materials to study). This point again emphasizes the advantages of K-line analysis where possible. If you are unfortunate enough to have heavy elements (say $Z > 60$) in your specimen, the L or M lines, which may be the strongest in a spectrum from a Si(Li) detector, will undoubtedly give rise to greater errors than the K lines, which may only be detectable with an IG system.

TABLE 35.3A Calculated k_{AFe} Factors for K_α X-rays Using Different Theoretical Cross Sections

Element A	k_{MM}*	k_{GC}*	k_P*	k_{BP}*	k_{SW}*	k_Z*
Na	1.42	1.34	1.26	1.45	1.17	1.09
Mg	1.043	0.954	0.898	1.03	0.836	0.793
Al	0.893	0.882	0.777	0.877	0.723	0.696
Si	0.781	0.723	0.687	0.769	0.638	0.623
P	0.813	0.759	0.723	0.803	0.671	0.663
S	0.827	0.776	0.743	0.817	0.688	0.689
K	0.814	0.779	0.755	0.807	0.701	0.722
Ca	0.804	0.774	0.753	0.788	0.702	0.727
Ti	0.892	0.869	0.853	0.888	0.807	0.835
Cr	0.938	0.925	0.917	0.936	0.887	0.909
Mn	0.98	0.974	0.970	0.979	0.953	0.965
Fe	1.0	1.0	1.0	1.0	1.0	1.0
Co	1.063	1.069	1.074	1.066	1.096	1.079
Ni	1.071	1.085	1.096	1.074	1.143	1.23
Cu	1.185	1.209	1.227	1.19	1.31	1.24
Zn	1.245	1.278	1.305	1.255	1.44	1.32
Mo	3.13	3.52	3.88	3.27	3.84	3.97
Ag	4.58	5.41	6.23	4.91	5.93	6.28

TABLE 35.3B Calculated k_{AFe} Factors for L X-rays Using Different Theoretical Cross Sections

Element	k_{MM}*	k_P*	k_{BP}*	k_{SW}*	k_Z*
Sr*	1.73	1.33	1.32	1.64	1.39
Zr*	1.62	1.26	1.24	1.51	1.33
Nb*	1.54	1.21	1.18	1.43	1.28
Ag*	1.43	1.16	1.09	1.26	1.26
Sn	2.55	2.09	1.93	2.21	2.30
Ba	2.97	2.52	2.25	2.49	2.83
W	3.59	3.37	2.68	2.80	3.88
Au	3.94	3.84	2.94	3.05	4.43
Pb	4.34	4.31	3.05	3.34	4.97

All L-line *k*-factors use the total L counts from the L_α and L_β lines.
Cross sections used in the calculations are: MM (Mott-Massey); GC (Green-Cosslett); P (Powell); BP (Brown-Powell); SW (Schreiber-Wims); Z (Zaluzec).

AFTER A SERVICE

If your detector is replaced or serviced, which is not an unusual occurrence on an AEM, then the new detector parameters must be inserted into the software.

The combination of uncertainties in Q and in the detector parameters is the reason why calculated k-factors are not very accurate, usually no better than ±10–20% relative. The computer system attached to the AEM should have predetermined values of all the terms in equations 35.14 and 35.15 stored in its memory. You don't usually have control over which particular parameters are being used. However, you should at least find out from your technical support the sources of the values of Q, ω, and a in your computer. You should then carry out a cross-check calculation with a known specimen to ensure that the calculated k-factor gives a reasonable answer.

BLACK BOXES

All software packages use preset values in their calculations and these may vary from package to package.

We cannot recommend a best set of values for Q, ω, and a, but the values of Q given by Powell, ω from Bambynek et al., and a from Schreiber and Wims have often been used. Also we can't give you specific detector parameters, so you should read the literature from your XEDS manufacturer. The values of μ/ρ that are still widely accepted are those determined by Heinrich although there is still considerable uncertainty in μ/ρ values for low-energy X-rays from the lighter elements. If you use the DTSA program from NIST (see Section 1.6), you may find that it predicts a worse value.

Figure 35.7A and B shows a comparison of the two methods of k-factor determination. The experimental data are shown as individual points with error bars and the solid lines represent the range of calculated k-factors, depending on the particular value of Q used in equation 35.14. The relatively large errors possible in the calculated k-factors are clearly seen and comparison of the K-line data in Figure 35.7A with the L-line data in Figure 35.7B again emphasizes the advantages of using K lines for the analysis where possible. Similar data for M lines are almost non-existent.

We can summarize the k-factor approach to analysis in the following way

- The Cliff-Lorimer equation has the virtue of simplicity; all you have to do is specify all the variables and treat the standard and unknown in an identical manner.
- You are better off calculating k_{AB} if you prefer speed to accuracy; experimental determination is best if you wish to have a known level of confidence in the numbers that you produce.

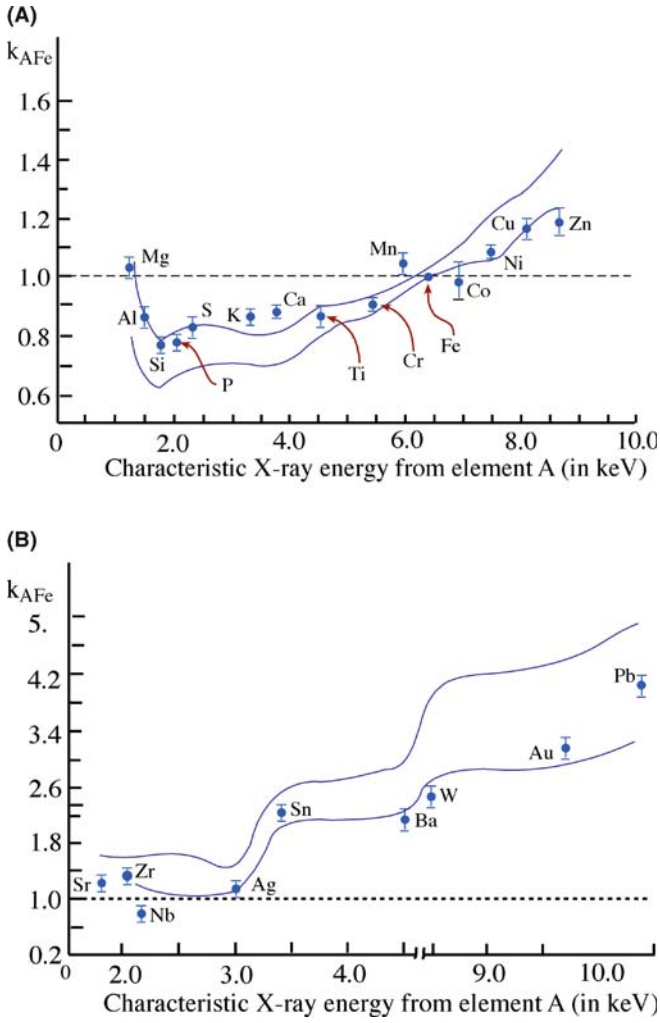


FIGURE 35.7. (A) Experimental k_{AFe} factors as a function of X-ray energy for the K_α X-rays from a range of elements A with respect to Fe. The solid lines represent the spread of calculated k-factors using different values for the ionization cross section. (B) Similar data to (A) for L_α lines from relatively high-Z elements. The errors in the calculated values of k are large, reflecting the uncertainties in L-line ionization cross sections.

35.5 THE ZETA-FACTOR METHOD

The Cliff-Lorimer ratio method is > 30 years old and, while it is simple in concept, the need for a k-factor combined with the difficulties of finding the right standard specimen, and calculating an accurate enough k-factor to solve a specific problem are significant limiting factors. These can be overcome if, instead of using a ratio method, we go back to the fundamentals of X-ray analysis as originally developed for the EPMA (based on equation 35.1) wherein pure-element standards are used. Pure-element standards have the distinct advantages of being easy to fabricate and they don't change composition during thinning or under beam damage. This approach, developed by Watanabe and Horita in Japan, is termed the zeta (ζ) factor method. A review of the development of the method and the pros and cons of

ζ-factors versus *k*-factors was given by Watanabe and Williams in 2006.

The major experimental drawback to ζ-factors is that the method requires in-situ measurement of the probe current hitting your specimen. Unfortunately, despite the fact that beam-current measurement is standard on almost all SEMs and every EPMA, commercial AEMs don't come equipped with this capability, emphasizing the point made some while ago in Chapter 33 that AEMs are really modified TEMs and XEDS analysis is still a bit of an afterthought.

But, if you can measure the current in situ, e.g., with a specimen holder within which a Faraday cup is embedded, life gets much easier. In a thin-foil specimen, we can assume that the characteristic X-ray intensity is proportional to the mass thickness, ρt, if X-ray absorption and fluorescence are negligible. Therefore, we can define the ζ-factor for pure element A as

$$\rho t = \zeta_A \frac{I_A}{C_A} \tag{35.16}$$

Where

$$\zeta \equiv \frac{A}{CN_0 Q \omega a i} \tag{35.17}$$

where the only new terms are the beam current i, and Avogadro's number N_0. Under these conditions, the ζ-factor is dependent on the X-ray energy, the kV, and the beam current. The first two are constant for any experiment and the third you have to measure. The ζ-factor is independent of specimen thickness, composition, and density which, as we shall see later, makes any absorption correction trivial.

So we can write similar equations for all the other pure elements in our specimen

$$\rho t = \zeta_B \frac{I_B}{C_B} \tag{35.18}$$

When we know the ζ-factors for A and B, C_A, C_B, and ρt can be expressed from equations 35.16 and 35.18, assuming $C_A + C_B = 1$ in a binary system:

$$C_A = \frac{I_A \zeta_A}{I_A \zeta_A + I_B \zeta_B},$$
$$C_B = \frac{I_B \zeta_B}{I_A \zeta_A + I_B \zeta_B} \quad \rho t = I_A \zeta_A + I_B \zeta_B \tag{35.19}$$

Therefore, we can determine C_A, C_B, and ρt simultaneously just by measuring X-ray intensities. It is simple to rearrange equation 35.19 to compare with the Cliff-Lorimer ratio equation (equation 35.2) and equation 35.20 is just as simple to apply, but *k*-factors are no longer required, just ζ-factors.

$$\frac{C_A}{C_B} = \frac{I_A \zeta_A}{I_B \zeta_B} \tag{35.20}$$

As we'll see below, absorption and fluorescence-correction terms can be combined directly with equation 35.19. To determine the ζ-factors, you just measure X-ray characteristic intensities (above the background)

TABLE 35.4 ζ-Factor Values Estimated from Experimental X-ray Spectra from the NIST SRM2063a Glass Thin Film in a 200-keV FEG-STEM JEM-2010F with an ATW Detector and in a 300-keV FEG-DSTEM VG HB603 with a Windowless Detector.

	ζ-factor (kg electron/m/photon)			ζ-factor (kg electron/m/photon)	
Element (Z)	200 keV ATW	300 keV Windowless	Element (Z)	200 keV ATW	300 keV Windowless
N	16,505.2±1,537	720.2±58.0	Mn	1,752.2±41.8	706.0±17.3
O	4,092.3±205.5	583.5±31.3	Fe	1,790.5±42.7	721.1±17.7
F	20,548.2±4,852.8	891.6±214.5	Co	1,919.0±45.8	772.0±18.9
Ne	5,345.6±702.1	635.8±85.5	Ni	1,950.0±46.5	783.2±19.2
Na	2,819.6±221.9	571.7±46.3	Cu	2,163.4±51.6	867.3±21.2
Mg	1,847.9±95.0	501.4±26.6	Zn	2,300.1±54.9	920.1±22.5
Al	1,510.2±56.2	483.6±18.6	Ga	2,541.8±60.7	1,014.4±24.8
Si	1,369.5±41.3	467.5±14.6	Ge	2,762.1±65.9	1,099.5±26.9
P	1,691.2±50.4	484.3±13.4	As	3,009.5±71.8	1,194.7±29.2
S	1,488.6±40.0	507.6±14.2	Se	3,328.6±79.4	1,317.7±32.2
Cl	1,465.5±37.3	528.0±13.9	Br	3,531.1±84.3	1,397.7±34.2
Ar	1,532.7±37.8	579.3±14.9	Kr	3,890.2±92.8	1,534.6±37.6
K	1,405.5±34.2	545.7±13.8	Rb	4,182.3±99.8	1,644.2±40.2
Ca	1,377.6±33.2	544.3±13.6	Sr	4,476.8±106.8	1,755.7±43.0
Sc	1,522.5±36.5	607.0±15.0	Y	4,786.6±114.2	1,871.6±45.8
Ti	1,553.4±37.2	622.8±15.4	Zr	5,185.4±123.7	2,019.8±49.4
V	1,632.3±39.0	656.5±16.1	Nb	5,578.9±133.1	2,164.8±53.0
Cr	1,654.6±39.5	666.5±16.4	Mo	6,088.3±145.3	2,353.4±57.6

from pure-element thin films with known composition and thickness, rather than trying to find a suitable set of standards, such as those listed in the references for Tables 35.1 and 35.2. Pure-element standards are more routinely available than the various multi-element, thin-film standards that have been used for *k*-factor determination over the last 30 years or so. You can determine an entire set of the ζ-factors for K-shell X-ray lines from a single spectrum from the National Institute of Standards and Technology (NIST) thin-film glass standard reference material (SRM) 2063. Table 35.4 lists such a set of ζ-factors obtained from the NIST standard SRM 2063a, which is a thinner version of SRM 2063. Its composition is known to a high degree of accuracy, as is its thickness and density but, unfortunately, it is now out of production.

There's much more about the ζ-factor in the companion text. Despite its obvious advantages, the ζ-factor approach is not yet commercially available; however, it can be downloaded from the book Web site (URL #2).

35.6 ABSORPTION CORRECTION

The point at which the simple Cliff-Lorimer approach breaks down is when the thin-foil criterion is invalid. Then the X-ray counts from your specimen are not a function of Z alone and absorption and (very occasionally) fluorescence invalidate the simple criterion. The effects of absorption are much more of a problem than fluorescence, so let's look at absorption first.

Preferential absorption of the X-rays from one of the elements in your specimen means that the detected X-ray counts will be less than the generated counts and so C_A is no longer simply proportional to I_A. So you have to modify the *k*-factor to take into account the reduction in I_A. This problem can arise (a) if your specimen is too thick, (b) if one or more of the characteristic X-rays has an energy less than ~1–2 keV (i.e., light-element analysis) or (c) when you have X-ray lines in your spectrum that differ in energy by > 5–10 keV (because the lower energy X-ray is much more likely to be absorbed than the higher energy one).

If we define k_{AB} as the true sensitivity factor when the specimen thickness $t = 0$, then the effective sensitivity factor for a specimen in which absorption occurs is given by k_{AB}* where

$$k^*_{AB} = k_{AB}(\mathrm{ACF}) \tag{35.21}$$

So we can write

$$\frac{C_A}{C_B} = k_{AB}(\mathrm{ACF})\frac{I_A}{I_B} \tag{35.22}$$

The absorption-correction factor (ACF) is the A term in equation 35.13 and we can write it as

$$\mathrm{ACF} = \frac{\int_0^t \left\{ \varphi_B(\rho t)\; e^{-\left(\frac{\mu}{\rho}\right]^B_{\mathrm{Spec}} \rho t \,\mathrm{cosec}\,\alpha} \right\} d(\rho t)}{\int_0^t \left\{ \varphi_A(\rho t)\; e^{-\left(\frac{\mu}{\rho}\right]^A_{\mathrm{Spec}} \rho t \,\mathrm{cosec}\,\alpha} \right\} d(\rho t)} \tag{35.23}$$

In this expression $\varphi(\rho t)$ is the depth distribution of X-ray production (which is the ratio of the X-ray emission from a layer of element A/B (of thickness $\Delta\rho t$ at depth t in the specimen with density ρ) to the X-ray emission from an identical, but isolated film). The term $\left(\frac{\mu}{\rho}\right]^A_{\mathrm{Spec}}$ is the mass-absorption coefficient of X-rays from element A in the specimen and α is the detector take-off angle. Since the values of μ/ρ are often from old publications, the units are usually given in cm^2/gm rather than kg/m^2, so you may have to use ρ in gm/cm^3 and t in cm, rather than SI units (kg/m^3 and m, respectively). Obviously, the value of the ACF is unity when no absorption occurs. Typically, if the ACF is >10% we define the absorption as significant, since 10% accuracy is routinely attainable in quantitative analysis using experimental *k*-factors. Let's now look at each of the terms and the problems associated with determining their value.

Again, we recommend that you use the values of μ/ρ given by Heinrich. The value of μ/ρ for a particular X-ray (e.g., from element A) within the specimen is the sum of the mass-absorption coefficients for each element times the weight fraction of that element, so

$$\left(\frac{\mu}{\rho}\right]^A_{\mathrm{spec}} = \sum_i \left(\frac{C_i\mu}{\rho}\right]^A_i \tag{35.24}$$

where C_i is the fractional concentration of element i in the specimen such that

$$\sum_i C_i = 1 \tag{35.25}$$

The absorption of X-rays from element A by all elements i in the specimen is summed. The summation includes self-absorption by element A. Elements that may not be of interest in the experiment or that might not be detectable may still cause absorption.

The NiO-MgO example of this phenomenon occurs when Mg is being quantified in homogeneous NiO-MgO. The Mg K_α X-rays will be absorbed by oxygen, even if the O K_α X-ray is not of interest or cannot be detected because a Be-window detector is being used. This effect is shown in Figure 35.8, which shows an increase in the intensity ratio (Ni K_α/Mg K_α) as a function of thickness due to the increased absorption of the

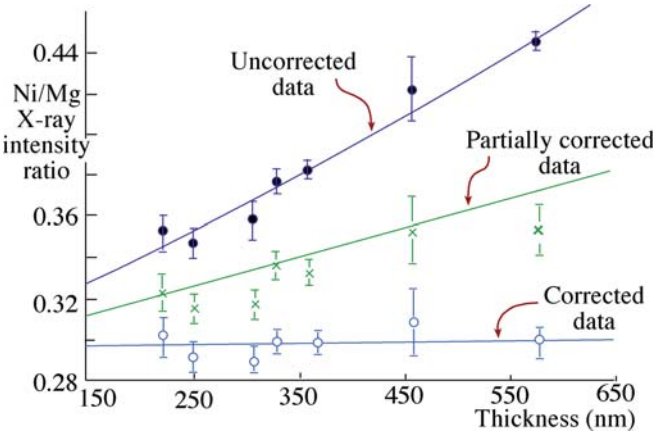


FIGURE 35.8. The upper curve shows the raw Ni K_α/Mg K_α intensity ratio as a function of thickness in a homogeneous specimen of NiO-MgO. The slope indicates strong absorption of Mg K_α X-rays. The middle curve shows the effect of correcting for absorption of the Mg K_α X-rays by Ni and the bottom line shows the effect of a further correction for absorption of the Mg K_α by O to give the expected horizontal line.

Mg K_α X-rays. (Absorption appears in an exponential term.) If we correct for the absorption by Ni, the slope of the line is reduced, but only when the effects of absorption by oxygen are taken into account does the slope become zero, as it should be for a homogeneous specimen.

In equation 35.22, we assume that the depth distribution of X-ray production $\varphi(\rho t)$ is a constant, and equal to unity. That is, a uniform distribution of X-rays is generated at all depths throughout the foil. This is a reasonable first approximation in thin foils, but in bulk specimens $\varphi(\rho t)$ is a strong function of t and so the measurement of $\varphi(\rho t)$ for bulk specimens is a well-established procedure. The few studies in thin specimens show an increase in $\varphi(\rho t)$ with specimen thickness, although the increase is no more than ~5% in foil thicknesses of < 300 nm. Therefore, the assumption appears reasonable since, if your specimen is thicker than 300 nm, you will have problems to worry about other than $\varphi(\rho t)$. The fact that we use a ratio of the two $\varphi(\rho t)$ terms in the absorption equation also helps to minimize the effects of this assumption.

We assume that $\varphi(\rho t)$ equals unity, then we can simply use equation 35.23 to give

$$\mathrm{ACF} = \left(\frac{\left[\frac{\mu}{\rho}\right]^{\mathrm{A}}_{\mathrm{Spec}}}{\left[\frac{\mu}{\rho}\right]^{\mathrm{B}}_{\mathrm{Spec}}}\right)\left(\frac{1 - e^{-\left(\left[\frac{\mu}{\rho}\right]^{\mathrm{B}}_{\mathrm{Spec}} \rho t \operatorname{cosec}\alpha\right)}}{1 - e^{-\left(\left[\frac{\mu}{\rho}\right]^{\mathrm{A}}_{\mathrm{Spec}} \rho t \operatorname{cosec}\alpha\right)}}\right) \qquad (35.26)$$

So we still need to know the values of ρ and t for our specimens.

The density of the specimen (ρ) can be estimated if you know the unit-cell dimensions, e.g., from CBED since

$$\rho = \frac{nA}{VN} \qquad (35.27)$$

where n is the number of atoms of average atomic weight A in a unit cell of volume V, and N is Avogadro's number.

The absorption path length (t') is a major variable in the absorption correction. Fortunately, it is also the one over which you, the operator, have the most control. In the simplest case of a parallel-sided thin foil of thickness t at 0° tilt, the absorption path length, as shown in Figure 35.9, is given by

$$t' = t' \operatorname{cosec}\alpha \qquad (35.28)$$

where α is the detector take-off angle. To minimize this factor it is obvious that your specimen should be as thin as possible and the value of α as high as possible. There are many ways to determine the foil thickness, which we have discussed at various points in this text; they are summarized in Section 36.3. More recently, Banchet et al. (2003) have combined EELS measurements of relative specimen thickness with XEDS peak intensities to give a variation on the traditional, iterative absorption-correction process. No method is universally applicable, and few are either easy or accurate, so it's best to make thin specimens in the first place.

The value of α when your specimen is at the ideal 0° tilt is fixed by the design geometry of your AEM stage and the only way you can vary α is by tilting your specimen. As we have seen, there are good reasons not to tilt beyond about 10°, because of the increase in spurious X-rays, but if there is a severe absorption problem, then decreasing t' by tilting your specimen *toward* the detector is a sensible first step toward minimizing the problem. On some very old AEMs the detector may not be orthogonal to the axis of your specimen holder, in which case you've got a challenging exercise in solid geometry to determine α.

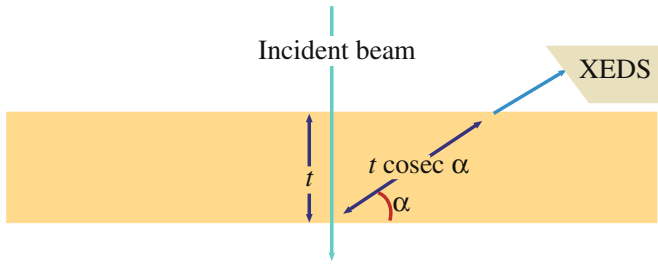


FIGURE 35.9. Relationship between the specimen thickness, t, and the absorption path length, $t \operatorname{cosec}\alpha$, for a take-off angle α.

HOW THICK?
Remember: it's rare that you'll know your specimen thickness as well as you would like.

So far we've assumed that our specimens are parallel sided, but this is uncommon. Many thin-foil preparation methods result in wedge-shaped foils, and under these circumstances the detector must always be looking toward the thin edge of the specimen so that the X-ray path length is minimized, as we already described in Figure 32.15. The only way to ascertain if this is a problem is to measure the thickness at each analysis point. Because this is such a tedious exercise, you should get round it by using the ζ-factor method, as we explain in the next section.

Because the specimen density ρ (and therefore the values of μ/ρ) varies with your specimen composition the complete absorption-correction procedure is an iterative process. The first step is to use the Cliff-Lorimer equation without any absorption correction and thus produce values for C_A and C_B. From these values, the computer performs a first-iteration calculation of μ/ρ and ρ, and generates modified values of C_A and C_B, and iterates again. Usually the calculation converges after two or three iterations.

In summary, there is substantial room for error in determining the various terms to insert into the ACF. For example, the ACF for k_{NiAl} in Ni_3Al, which is a strongly absorbing system, varies from ~5.5 to ~12% when the specimen doubles in thickness from 40 to 80 nm. This change is still quite small and within the limits of all but the most accurate analyses. In FeNi, which is a weakly absorbing system, a similar change in thickness would change the ACF for k_{FeNi} from ~0.6 to ~1.3%, which is negligible. So while we've spent a fair bit of time introducing you to the absorption correction, the final message is clear.

CLIFF-LORIMER ACCURACY?
The Cliff-Lorimer approach only incurs large errors in strongly absorbing systems and/or very thick specimens.

35.7 THE ZETA-FACTOR ABSORPTION CORRECTION

So, if significant absorption is unavoidable, information about your specimen density and thickness is required at each analysis position, in order to apply the absorption correction. Obviously, this is the major limitation, since independent measurements are required for the specimen density and thickness and inaccuracies in such measurements may cause further errors in quantification. In fact, the ζ-factor method was originally proposed in order to overcome these limitations and difficulties associated with the absorption correction because, if we substitute equation 35.16 into equation 35.22, the ρt term can be eliminated

$$\frac{C_A}{C_B} = k_{AB}\left(\frac{I_A}{I_B}\left[\frac{(\mu/\rho)^A_{sp}}{(\mu/\rho)^B_{sp}}\right]\left\{\frac{1-\exp\left[-(\mu/\rho)^B_{sp}\ \zeta_A(I_A/C_A)\,\mathrm{cosec}(\alpha)\right]}{1-\exp\left[-(\mu/\rho)^A_{sp}\ \zeta_A(I_A/C_A)\,\mathrm{cosec}(\alpha)\right]}\right\}\right) \tag{35.29}$$

If you've determined ζ-factors, then the k-factor can be substituted in the above equation since

$$k_{AB} = \frac{\zeta_A}{\zeta_B} \tag{35.30}$$

So the ζ-factor overcomes the two major limitations of the Cliff-Lorimer method by avoiding the tedium of preparing multiple thin-foil standards and making multiple thickness and density measurements at each analysis point when significant X-ray absorption is occurring. That's why we introduce its depth in the companion text and suggest strongly that the serious X-ray analyst use this approach.

35.8 THE FLUORESCENCE CORRECTION

X-ray absorption and fluorescence are intimately related because a primary cause of X-ray absorption is the fluorescence of another X-ray (such as the fluorescence of Si K_α X-rays in the XEDS detector which gives rise to the escape peak). You might think, therefore, that fluorescence corrections should be as widespread as absorption corrections. However, this is not the case for the following reasons. Strong absorption effects occur when there is a small amount of one element whose X-rays are being absorbed by the presence of a relatively large amount of another element. The absorption of Al K_α X-rays by Ni in Ni_3Al is a classic example. In this case, Ni X-rays are indeed fluoresced as a result of the absorption of Al K_α X-rays. However, there is a relatively small

increase in the total number of Ni X-rays because Ni is the dominant element; the relative decrease in the Al K_α intensity is large because Al is the minor constituent. In this particular example there is a further reason why fluorescence of Ni X-rays is ignored; it is the Ni L_α X-rays which are fluoresced by the absorption of Al K_α X-rays. The Ni L X-rays are not the ones that we use for analysis anyhow, since the higher energy Ni K X-rays are not absorbed or fluoresced.

In the rare case that fluorescence occurs to a degree that limits the accuracy of your analysis, read the detailed discussion given by Anderson et al. Practical examples of the fluorescence correction are hard to come by and a classic case is Cr in stainless steels where the minor Cr K_α line is fluoresced by the major Fe K_α line, giving rise to an apparent increase in Cr content as the foil gets thicker.

DON'T WORRY ABOUT FLUORESCENCE

Fluorescence is usually a minor effect and often occurs for X-rays that are not of interest. (So don't worry if you know you needn't!)

35.9 ALCHEMI

We told you early on in this chapter to acquire your X-ray spectra away from strong diffraction conditions. This is because of the Borrmann effect. Close to two-beam conditions, the Bloch waves interact strongly with the crystal planes and so X-ray emission is enhanced compared with kinematical conditions, thus negating the assumptions inherent in the Cliff-Lorimer equation, which assumes emission is constant with specimen tilt. However, we can make use of this phenomenon to locate which atoms lie on which crystal planes. The technique has the delightful (and wholly inappropriate) acronym ALCHEMI, which is a selective abbreviation of the expression 'atom location by channeling-enhanced microanalysis.'

ALCHEMI is a quantitative technique for identifying the crystallographic sites, distribution and types of substitutional impurities in crystals. The technique was first developed for the TEM by Spence (who, with archetypical antipodean humor, coined the acronym) and Taftø. Interestingly, channeling is also used for atom-site location in other analysis techniques (e.g., see Chu et al.).

The way to do ALCHEMI experimentally is to tilt your specimen to a strong two-beam condition and acquire a spectrum under strong channeling conditions, such that the Bloch wave is interacting strongly with a particular systematic row of atoms. You should choose the channeling orientation so that the specific crystal planes interacting strongly with the beam also contain the candidate impurity atom sites. So it helps a lot if you have some a priori ideas about where substitutional atoms are most likely to sit. This technique is therefore particularly well suited to layer structures. When the Bloch wave is maximized on a particular plane of atoms, the X-ray counts from the atoms in that plane will be highest. So start by finding the orientations 1 and 2 that give the most pronounced channeling effects for the atoms A and B, as shown schematically in Figure 35.10A. Usually a very small tilt is all that is necessary to get a different spectrum from the two planes.

If you are looking at two elements A and B and a substitutional element X then follow this procedure

- Measure X-ray intensities from each element in orientations 1 and 2.
- Then find a non-channeling orientation (3) where the electron intensity is uniform for both planes.

In this orientation we define the ratio k (NOT the Cliff-Lorimer factor) as

$$k = \frac{I_B}{I_A} \tag{35.36}$$

I_B is the number of X-ray counts from the element B in the non-channeling orientation. For the two channeling orientations 1 and 2, we define two parameters β and γ such that

$$\beta = \frac{I_B^{(1)}}{kI_A^{(1)}} \tag{35.37}$$

$$\gamma = \frac{I_B^{(2)}}{kI_A^{(2)}} \tag{35.38}$$

Now assuming we know from looking at the relative intensity changes in the spectra that the element X sits on specific sites, say it substitutes for atom B, then we define an intensity ratio term R such that

$$R = \frac{I_A^{(1)} I_X^{(2)}}{I_X^{(1)} I_A^{(2)}} \tag{35.39}$$

Then the fraction of atom X on B sites is given by

$$C_X = \frac{R-1}{R-1+\gamma-\beta R} \tag{35.40}$$

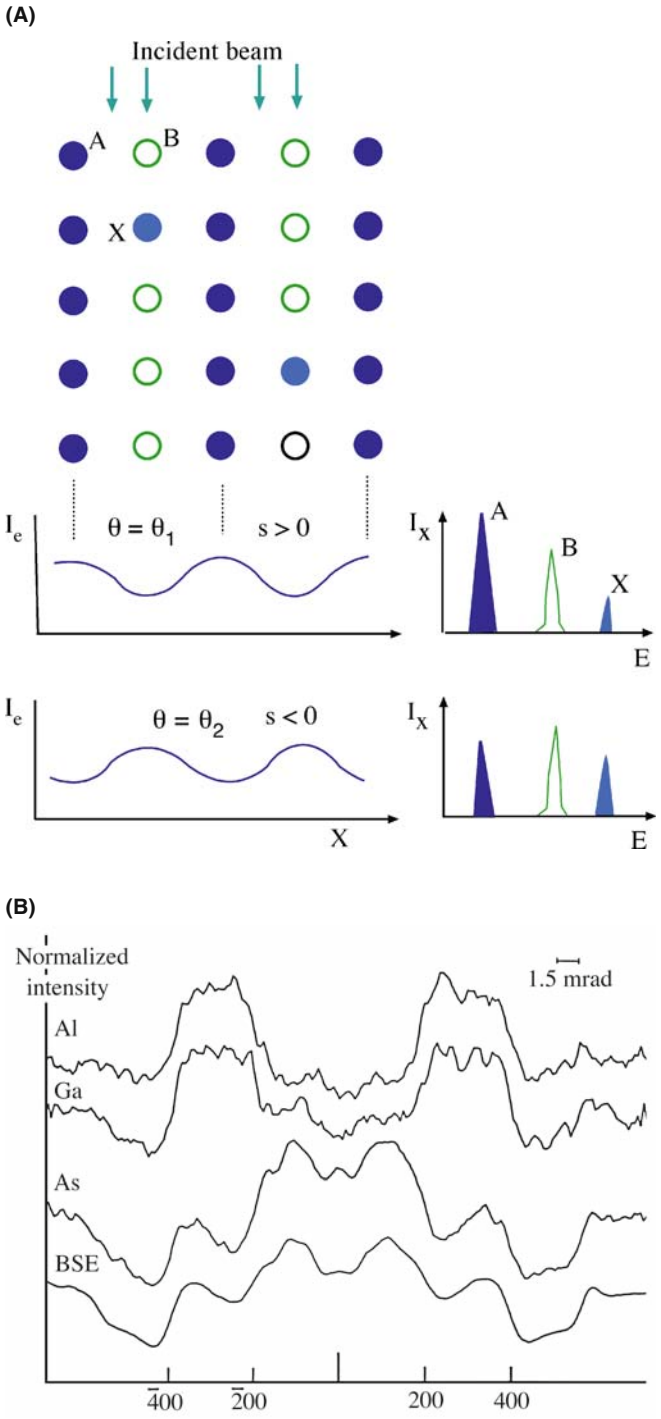


FIGURE 35.10. (A) ALCHEMI allows the determination of the site occupancy of atom X (light blue) in columns of atoms A (dark blue) and B (open circles). By tilting to **s** >0 and then **s** <0, the Bloch waves interact strongly with row A then row B giving different characteristic intensities, shown schematically in the spectra, from which the relative amounts of X in columns of A and B can be determined. (B) The Borrmann effect: the variation in the characteristic X-ray emission close to strong two-beam conditions as the beam is rocked across the 400 planes of GaAlAs. The X-rays from Al, which occupies Ga sites, follow the Ga emission variation while the As varies in an approximately complementary fashion. The BSE signal is inversely proportional to the amount of channeling so the As signal is strongest where the channeling is weakest.

Similar expressions can be generated for X atoms on A sites, but in fact the fraction of X atoms on A sites must be $1–C_X$.

As you see, ALCHEMI can give a direct measure of the occupation of substitutional sites. However, the intensity differences in different orientations are often quite small and you need good X-ray statistics to draw sound conclusions (counts, counts, and more counts!). This makes ALCHEMI difficult to apply if high spatial resolution is also desired because, as we shall see in the next chapter, the conditions to give the best spatial resolution also give the worst counting statistics. Figure 35.10B shows the variation in X-ray emission across a bend contour highlighting the Borrmann effect which is the basis of ALCHEMI. A comprehensive review has been given by Jones, who extends the discussion of the technique and develops it in depth in the companion text.

35.10 QUANTITATIVE X-RAY MAPPING

As we have discussed throughout the previous chapters, gathering X-ray maps rather than individual or lines of spectra makes a great deal of sense in terms of getting unbiased elemental-distribution information about your specimen. The major difficulty with moving from qualitative to quantitative mapping is the need for sufficient counts for quantification. As noted in Section 35.4.B above, we recommend acquiring 10,000 counts in a characteristic peak in order for reasonable quantification (i.e., with errors $\sim \pm$ 10% relative). A simple calculation will show how unrealistic this is if we are to acquire maps in a reasonable time. A minimum map to give a reasonable X-ray image is 128 $\times$ 128 pixels giving $>$ 16,000 total pixels. Even if we acquire for only 1 s/pixel (if we are lucky we will acquire a few tens of total counts rather than a few thousands), then we will be mapping for 4.5 hours minimum and more likely days if we wish to see hundreds or thousands of counts/pixel. As we've stated too often, such long times introduce specimen drift, damage, contamination, and operator boredom and so conspire to make life very difficult. While overnight mapping with lower-resolution EPMA systems (where the drift and stability requirements are much less stringent and the X-ray count rate is far greater) is indeed a common occurrence, STEMs are not yet stable enough to do this while retaining nanometer-level resolution.

Nevertheless, there has been significant progress in quantitative mapping, particularly with the use of 2–300 kV FEG instruments, the development of

higher solid angles of collection of X-rays, and general improvements in instrument design. So, as shown back in Figures 33.14 and 33.15, quantitative X-ray mapping is feasible. This process requires that the counts in the characteristic peak that is being mapped be integrated, subject to background subtraction, then processed via a Cliff-Lorimer or ζ-factor equation to turn the counts into composition. It is often best to exhibit the map as a color image because usually more than one element needs to be mapped or the composition changes need to be emphasized in which case overlays or side-by-side comparisons of the different maps permit easier understanding of the relative distributions of the various elements. Not surprisingly, quantitative mapping is improved by C_s correction since corrected probes have ~3–5× more current without loss of probe size. Thus the count rate is increased, or the acquisition time decreased, or both.

While it is certainly of interest to improve the stability of our TEM and XEDS systems and optimize the gathering of data over long periods of time, there are ways to minimize the limitations of low count rates and long acquisition times and this requires high-level computer control of the X-ray acquisition, implementation of spectrum imaging, or position-tagged spectrometry (check back in Sections 33.6.C and 33.6.D) and then manipulation of the resulting data cube using multivariate statistical analysis (MSA) to extract the maximum signal information and minimize the noise (which constitutes most of the signal in the channels in a spectrum acquired for a short period of time).

When this combination of SI and MSA is implemented, it is possible to gather spectra in as little as 100–500 ms/pixel giving mapping times of a few minutes to several tens of minutes for a 128 × 128 pixel image and it is equally feasible to contemplate larger maps. The SI/MSA combination is dealt with in great detail in the companion text. About the best that can be done with a modern C_s-corrected FEGSTEM, sophisticated data handling is shown in Figure 35.11 where segregation of trace elements to grain boundaries is mapped. A similar example was also shown back in Figure 33.15D which maps out differences in the composition of various small precipitates, a few nanometers in diameter. In Figure 35.11B, the data were acquired on a 300-keV STEM and the improvement in mapping quality obtained after MSA (Figure 35.11C) is clear. Addition of a C_s corrector to the STEM results in a further gain in spatial resolution, as shown in Figure 35.11D. This latter example of Zr segregating to a grain boundary in a Ni-base superalloy mapping reveals a typical enrichment of only ~1–2 atoms/nm^2 with a spatial resolution of ~0.5 nm!

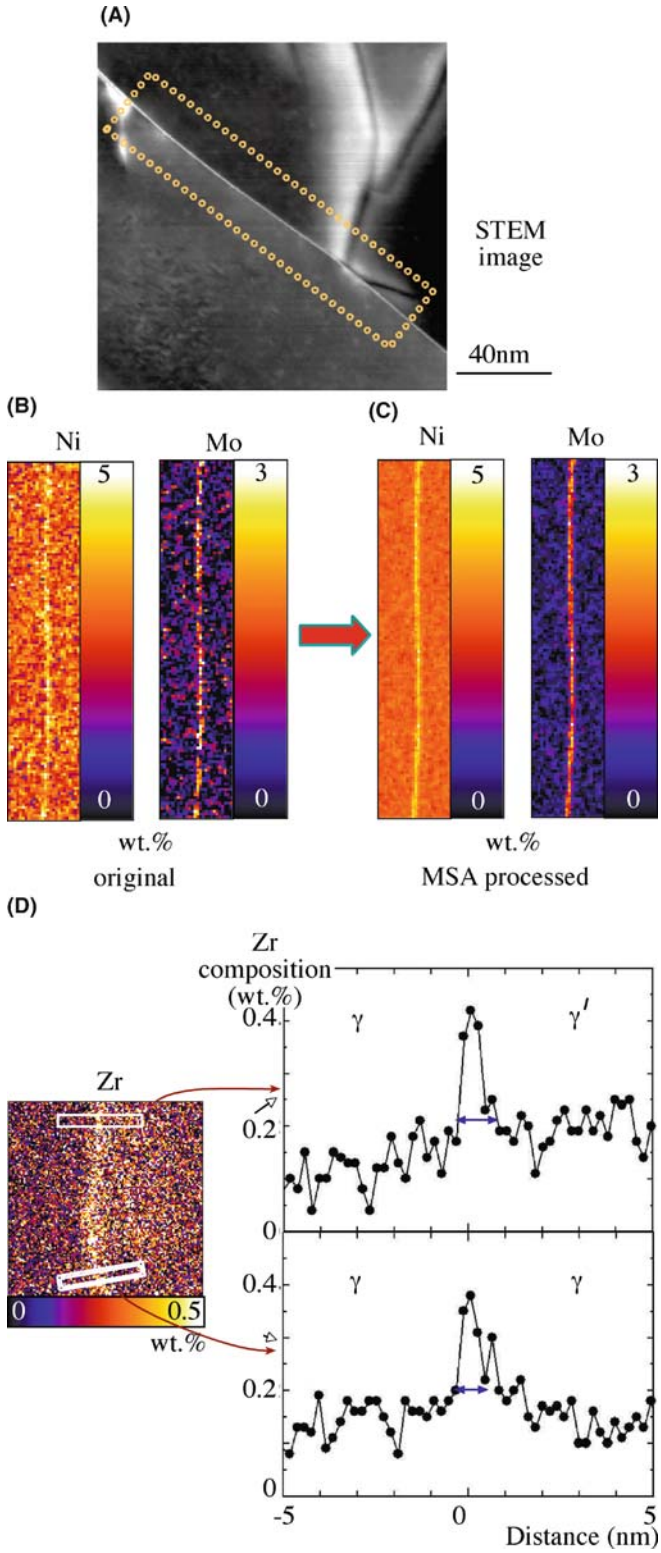


FIGURE 35.11. (A) STEM ADF image and (B) quantitative X-ray maps showing the segregation of trace amounts of Ni and Mo to grain boundaries in a low-alloy steel. (C) Applying MSA improves the quality of the maps. (D). Mapping the segregation of Zr to an interface in a Ni-base superalloy in a C_s-corrected STEM designed to give a 0.4-nm (FWTM) probe containing 0.5 nA. The Zr is present in the bulk alloy at ~0.04 wt% and without MSA processing could not be mapped. The composition profiles show that the Zr is localized to <1 nm at two different positions on the interface.

CHAPTER SUMMARY

The fact that much of this chapter is unchanged from the first edition sends a message that not much has changed in the last decade. This is unfortunate because, while quantitative analysis of spectra from thin foils can be straightforward, the standard Cliff-Lorimer approach has serious limitations. Most of the problems are overcome by the newer ζ-factor method, which is not yet commercially available but can be downloaded from the book web site (URL #2). Perhaps the greatest difficulty remains the need to know the specimen thickness in order to compensate for X-ray absorption and, again, the ζ-factor approach is invaluable in avoiding this. We can minimize absorption by making the thinnest possible specimens but then the number of X-ray counts may be so small that errors in the quantification are unacceptably large. The use of FEG sources, C_s correction, and improved TEM-EDS configurations with detector arrays to maximize the collection angle all help. With these latest advances, we can now perform quantitative X-ray mapping with a spatial resolution of a less than a nanometer and detection limits of a few atoms. There's much more about these exciting new aspects of quantitative analysis in the companion text.

GENERAL REFERENCES

Garratt-Reed, AJ and Bell, DC 2003 *Energy-Dispersive X-ray Analysis in the Electron Microscope* Bios (Royal Microsc. Soc.) Oxford UK.

Goodhew, PJ, Humphreys, FJ and Beanland, R 2001 *Electron Microscopy and Analysis* 3rd Ed. Taylor & Francis London. Instructive comparison of XEDS in SEM and TEM.

Goldstein, JI, Williams, DB and Cliff, G 1986 *Quantification of Energy Dispersive Spectra* in *Principles of Analytical Electron Microscopy* 155–217 Eds. DC Joy, AD Romig Jr. and JI Goldstein, Plenum Press New York. Introduction to many of the concepts in this chapter and the next one, including many worked examples.

Friel JJ and Lyman CE 2006 *X-ray Mapping in Electron-Beam Instruments* Microsc. Microanal. **12** 2–25. Detailed review of qualitative and quantitative mapping.

Jones, IP 1992 *Chemical Analysis Using Electron Beams* The Institute of Materials, London. The best source of examples of quantitative XEDS calculations.

Williams, DB and Goldstein, JI 1991 *Quantitative X-ray Microanalysis in the Analytical Electron Microscope in Electron Probe Quantitation* 371–398 Eds. KFJ Heinrich and DE Newbury Plenum Press New York. Derivations of the essential equations for thin-film quantification.

Zaluzec, NJ 1979 *Quantitative X-ray Microanalysis: Instrumental Considerations and Applications to Materials Science* in *Introduction to Analytical Electron Microscopy* Eds. JJ Hren JI Goldstein and DC Joy 121–167 Plenum Press NY. We've left this in so Nestor doesn't take us off his Web site.

CALCULATIONS

Anderson, IM, Bentley, J and Carter, CB 1995 *The Secondary Fluorescence Correction for X-Ray Microanalysis in the Analytical Electron Microscope* J. Microsc. **178** 226–239.

Bambynek, W, Crasemann, B, Fink, RW, Freund, HU, Mark, H, Swift, CD, Price, RE and Rao, PV *X-ray-fluorescence Yields, Auger and Coster-Kronig Transition Probabilities* 1972 Rev. Mod. Phys. **44** 716–813.

Cliff, G and Lorimer, GW 1975 *The Quantitative Analysis of Thin Specimens* J. Microsc. **103** 203–207. The original paper based on crushed mineral standards (all borrowed from the desk drawer of Pam Champness, Lorimer's wife, who is a well-known mineralogist).

Heinrich, KFJ 1986 *Mass Absorption Coefficients for Electron Probe Microanalysis* in *Proc. ICXOM-11* 67–77 Eds. J Brown and R Packwood University of Western Ontario Canada.

Powell, CJ 1976 *Evaluation of Formulas for Inner-shell Ionization Cross Sections* in *Use of Monte Carlo Calculations in Electron Probe Analysis and Scanning Electron Microscopy* 97–104 NBS Special Publication 460 Eds. KFJ Heinrich, DE Newbury and H Yakowitz U.S. Department of Commerce/NBS Washington DC.

Watanabe, M and Williams, DB 2006 *The Quantitative Analysis of Thin Specimens: a Review of Progress from the Cliff-Lorimer to the New ζ-Factor Methods* J. Microsc. **221** 89–109.

Williams, DB. Newbury, DE Goldstein, JI and Fiori, CE 1984 *On the Use of Ionization Cross Sections in Analytical Electron Microscopy* J. Microsc. **136** 209–218.

Schreiber, TP and Wims, AM 1981 *Quantitative Analysis of Thin Specimens in the TEM Using a $\phi(\rho z)$ Model* Ultramicrosc. **6** 323–334.

Schreiber, TP and Wims, AM 1982 *Relative Intensity Factors for K, L and M Shell X-ray Lines* X-ray Spectrometry **11** 42–45.
Larsen, RJ and Marx, ML 2001 *An Introduction to Mathematical Statistics and its Applications* 3rd Ed. Prentice Hall Upper Saddle River NJ.

CHANNELING

Chu, W-K, Mayer, JM and Nicolet, M-A 1978 *Backscattering Spectrometry* Academic Press Orlando.
Goldstein, JI, Newbury, DE, Echlin, P, Joy, DC, Romig, AD Jr, Lyman, C., Fiori, CE and Lifshin, E 2003 *Scanning Electron Microscopy and X-ray Microanalysis* 3rd Ed. Springer New York.

SOME HISTORY

Castaing, R 1951 *Application des Sondes Électroniques a une Méthode d'Analyse Ponctuelle Chimique et Cristallographique* Théses, Université de Paris ONERA Publication #55 Paris. A widely reproduced and historically significant publication.
Hillier, J and Baker, RF 1944 *Microanalysis by Means of Electrons* J. Appl. Phys. **15** 663–675.
Jones, IP 2002 *Determining the Locations of Chemical Species in Ordered Compounds; ALCHEMI* in *Advances in Imaging and Electron Physics* **125** 63–119.
Kramers, HA 1923 *On the Theory of X-ray Absorption and of the Continuous X- ray Spectrum* Phil. Mag. **46** 836–871.
Lorimer, GW, Al-Salman, SA and Cliff, G 1977 *The Quantitative Analysis of Thin Specimens: Effects of Absorption, Fluorescence and Beam Spreading* in *Developments in Electron Microscopy and Analysis* 369–371 Ed. DL Misell The Institute of Physics Bristol and London.
McGill, R. and Hubbard, FH 1981 *Quantitative Analysis with High Spatial Resolution* p30 Eds. GW Lorimer, MH Jacobs and P Doig The Metals Society London.
Reed, SJB 2005 *Electron Microprobe Analysis and Scanning Electron Microscopy in Geology* 2nd Ed. Cambridge University Press Cambridge UK.
Spence, JCH and Taftø, J 1983 *ALCHEMI - A New Technique for Locating Atoms in Small Crystals* J. Microsc. **130** 147–154.
Sprys, JW and Short, MA 1976 *Quantitative Elemental Analysis of 'Transparent' Particles in the TEM* Proc. 34th EMSA Meeting Ed. GW Bailey Claitors Baton Rouge LA 416–7
Statham PJ 2002 *Limitations to Accuracy in Extracting Characteristic Line Intensities From X-Ray Spectra* J. Research NIST **107** 531–546.
Wood, JE, Williams, DB and Goldstein, JI 1981 *Determination of Cliff-Lorimer k Factors for a Philips EM 400T* in *Quantitative Analysis with High Spatial Resolution* 24–30 Eds. GW Lorimer, MH Jacobs and P Doig The Metals Society London. k_{Fe} factors.

PROBLEMS WITH k_{ASi}

Goldstein, JI, Costley, JL, Lorimer, G, and Reed, SJB 1977 *Quantitative X-ray Microanalysis in the Electron Microscope SEM 1977* **1** 315–325 Ed. O Johari IITRI Chicago IL.
Graham, RJ and Steeds, JW 1984 *Determination of Cliff-Lorimer k Factors by Analysis of Crystallized Microdroplets* J. Microsc. **133** 275–280.
Sheridan, PJ 1989 *Determination of Experimental and Theoretical k_{ASi} Factors for a 200-kV Analytical Electron Microscope* J. Electr. Microsc. Tech. **11** 41–61.
Wood, JE, Williams, DB and Goldstein, JI 1984 *An Experimental and Theoretical Determination of k_{AFe} Factors for Quantitative X-ray Microanalysis in the Analytical Electron Microscope* J. Microsc. **133** 255–274.

URLS

1) mathworld.wolfram.com/Studentst-Distribution.html
2) http://www.TEMbook.com

SELF-ASSESSMENT QUESTIONS

Q35.1 Why do we have to correct the *k*-factor for absorption of X-rays?
Q35.2 Why is the fluorescence correction so small and generally ignored?
Q35.3 Why do we integrate only the K_α peak intensity rather than the $K_\alpha + K_\beta$ peaks if we can resolve them in the spectrum?
Q35.4 Why is the *k*-factor not a constant between different AEM-XEDS systems?
Q35.5 Why is the calculated *k*-factor generally inaccurate?
Q35.6 What's the largest contribution to the absorption correction and what does this tell you about ways to reduce your chance of having to do such?
Q35.7 What's the best way to minimize the errors in X-ray microanalysis?

Q35.8 What are the essential requirements for a good thin-foil standard and what does your list tell you about finding and selecting such standards?
Q35.9 What's the best way to measure the thickness of (a) your glass specimen, (b) your alloy foil, (c) your BN nanoparticle? (Hint: look ahead to Chapter 36.)
Q35.10 Why do quantitative analysis anyhow?
Q35.11 Why is it important to determine the errors in your quantitative analysis?
Q35.12 What is a typical ballpark quantification error in a simple binary (A–B) quantitative analysis?
Q35.13 What do you have to do to improve significantly on this error value?
Q35.14 When would you choose to use calculated *k*-factors rather than experimental ones?
Q35.15 When would you choose to determine your *k*-factors experimentally rather than calculate them?
Q35.16 List three ways to subtract the bremsstrahlung intensity from beneath the characteristic peaks in a spectrum.
Q35.17 Distinguish bremsstrahlung, continuum, and background X-rays.
Q35.18 What does ALCHEMI stand for and why is it anything but alchemy?
Q35.19 Why do we typically use wt% rather than at.% in X-ray microanalysis?
Q35.20 Why is it best to have as thin a window as possible on your XEDS detector but why is it generally impractical to have no window at all?

TEXT-SPECIFIC QUESTIONS

T35.1 If *k*-factors are not constants, what is the use of tables of *k*-factors, such as Tables 35.1 and 35.2?
T35.2 What are 'residuals' in the filtered spectra in Figure 35.6 and why are they useful?
T35.3 Distinguish top-hat apertures and top-hat filters and explain why both are useful in AEM-XEDS.
T35.4 Why would you use the two-window method of background subtraction as shown in Figure 35.5 rather than the one-window method in Figure 35.4?
T35.5 Why does the background intensity go to zero in Figure 35.4 and what experimental and instrumental factors affect the energy at which it goes to zero?
T35.6 Why do the values of the *k*-factors shown in Figure 35.7 decrease with decreasing atomic number and then increase again? (Hint: there's a good physics explanation for both of these trends.)
T35.7 Copy Figure 35.8 and extrapolate the three lines to lower thickness. At what value of thickness do they converge and why is this the case?
T35.8 Give three limitations to the *k*-factor approach that are overcome by the ζ-factor method?
T35.9 What is the single most cautionary lesson you can gain from Figure 35.10A?
T35.10 Why, in the AEM, can we measure elemental segregation phenomena generated in materials at lower temperatures compared with EPMA experiments? (Hint: think abut diffusion kinetics.)
T35.11 You have standard thin foils of Fe_2O_3, NiO, Ni_3Al, $CuSO_4$. Explain how you would determine *k*-factors for analysis of (a) FeS, (b) NiAl, (c) Al_2O_3, and (d) Al-Cu solid solution. List any specific concerns you may have with your determinations.
T35.12 Using Table 35.1, calculate reasonable first approximations for the *k*-factors for (a) Mn-Cr, (b) Mg-Al, and (c) Al-Cu. State any assumptions you make in your calculations.
T35.13 Using the DTSA software, practice generating X-ray spectra for different elements and compounds at different accelerating voltages (e.g., 100–300 kV) and different take-off angles (e.g., $\alpha = 20^\circ, 60^\circ$). Observe the differences in the background and characteristic spectra as a function of kV and α. Use the background-subtraction options to measure peak intensities and run practice quantifications.
T35.14 Using DTSA run quantification routines for a single binary specimen of your choice, but select a range of ionization cross sections with which to calculate a *k*-factor and compare the range of answers that you get for your *k*-factor and your quantification. What does this tell you about the limitations of calculated *k*-factors?
T35.15 Calculate the absorption correction factor (ACF) for Fe-10 wt% Al foils 10, 100, and 300 nm thick for different take-off angles of 20° and 70°. Then, evaluate the specimen thickness necessary to meet the thin-film criterion (> 10% absorption). Use 6.61 g/cm^3 for the specimen density and following data for the mass-absorption coefficients for Fe K_α, Fe L_α, and Al K_α lines (Heinrich 1986). If you wish, after completing the manual calculation compare your result with DTSA calculation of the same correction. (courtesy M. Watanabe)

	Mass-absorption coefficient (cm^2/g)		
Absorber	Fe K_α	Fe L_α	Al K_α
Fe	71.1	2,157	3,626
Al	96.5	2,936	397.5

36

Spatial Resolution and Minimum Detection

CHAPTER PREVIEW

Often when you do X-ray analysis of thin foils you are seeking information that is close to the limits of spatial resolution. Before you carry out any such analysis you need to understand the various controlling factors and in this chapter we explain these. Minimizing your specimen thickness is perhaps the most critical aspect of obtaining the best spatial resolution, so we summarize the various ways you can measure your foil thickness at the analysis point, but the quality of the TEM-XEDS system is also important.

A consequence of going to higher spatial resolution is that the X-ray signal comes from a much smaller volume of the specimen. A smaller signal means that you'll find it very difficult to detect the presence of trace constituents in thin foils. Consequently, the minimum mass fraction (MMF) in TEM is not as small as many other analytical instruments which have poorer spatial resolution. This trade-off is true for any analysis technique, and so it is only sensible to discuss the ideas of spatial resolution in conjunction with analytical detection limits. We'll make this connection in the latter part of the chapter. Despite the relatively poor MMF, it is possible to detect the presence of just a few atoms of one particular element if the analyzed volume is small enough, and so the TEM actually exhibits excellent minimum detectable mass (MDM). With the latest advances in XEDS and TEM technology, particularly C_s correction, X-ray analysis with atomic-column resolution and single-atom detection is now feasible in the same instrument.

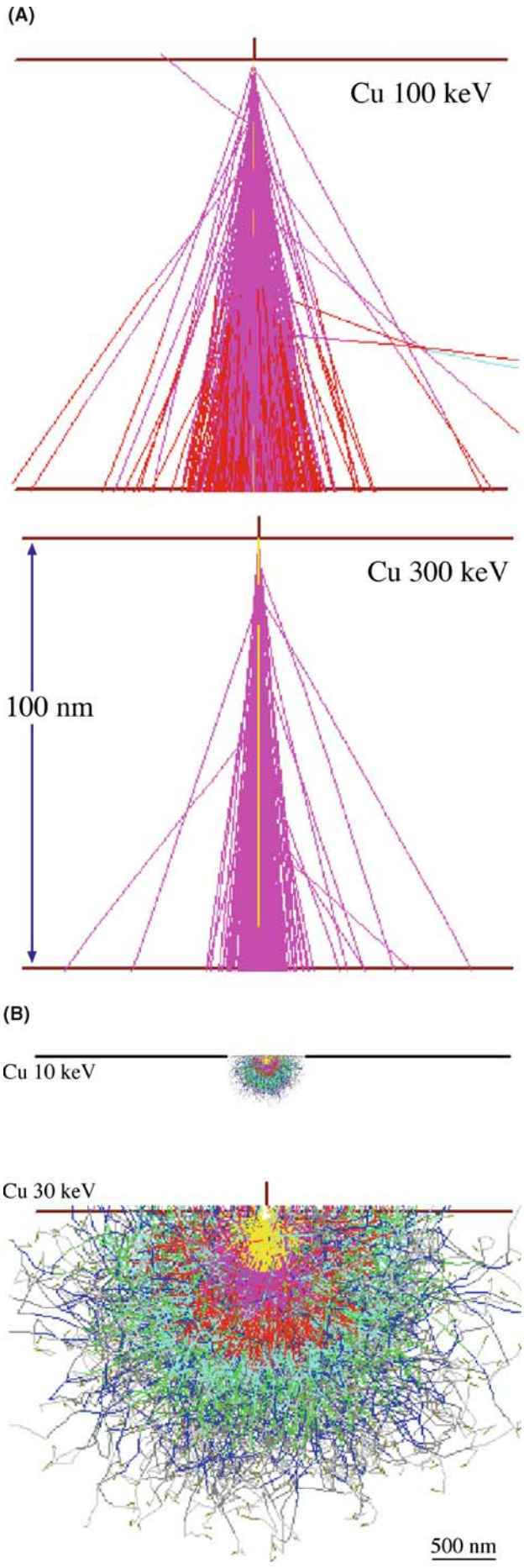
36.1 WHY IS SPATIAL RESOLUTION IMPORTANT?

As we described in the introduction to Chapter 35, the historical driving force for the development of X-ray analysis in the TEM was the improvement in spatial resolution compared with the EPMA. This improvement arises for two reasons

- We use thin specimens, so less electron scattering occurs as the beam traverses the specimen.
- The higher electron energy (>100–400 keV in the TEM compared with 5–30 keV in the EPMA) further reduces scattering.

The latter effect occurs because the mean free path for both elastic and inelastic collisions increases with the electron energy. The net result is that *increasing* the accelerating voltage when using thin specimens *decreases* the total beam-specimen interaction volume, thus giving a more localized X-ray signal source and a higher spatial resolution, which is good (see Figure 36.1A). Conversely, with bulk samples, *increasing* the voltage *increases* the interaction volume and spatial resolution is at best ~0.5–1 μm which is not so good (see Figure 36.1B). There is increasing interest in reducing the spatial resolution of SEM-X-ray analysis using very low voltage electron beams and low-energy X-ray lines. While this is challenging, there has been considerable progress and a spatial resolution <100 nm with E_0 <5 keV is feasible. Aberration correctors and bolometer detectors will help even further but, for the best spatial resolution, there is still no alternative to thinning your specimen.

Much theoretical and experimental work was carried out in the early days of AEM to define and measure the spatial resolution of XEDS in the TEM, and we'll introduce some of these concepts. The ultimate aim, of course, is to push spatial resolution to the atomic scale and detection limits to the single-atom level. Both of these goals have been attained in EELS, as we'll see in the next several chapters, but, as ever, we are limited by the small number of X-ray counts generated in thin foils and the poor collection efficiency of the XEDS. C_s correctors and SDDs are helping here, as we shall see.



36.2 DEFINITION AND MEASUREMENT OF SPATIAL RESOLUTION

It has long been recognized that the analysis volume, and hence the spatial resolution, is governed by the beam-specimen interaction volume, since the XEDS can detect X-rays generated anywhere within that volume (you'll see later that this is different to the situation in EELS). The interaction volume is a function of the incident-beam diameter (d) and the beam spreading (b) caused mainly by elastic scattering within the specimen. Therefore, the measured spatial resolution (R) is a function of your specimen and this has made it difficult to define a generally accepted measure of R. Let's look first at d and b and how we define them.

SPATIAL RESOLUTION FOR XEDS

We can define this spatial resolution as the smallest distance (R) between two volumes in the specimen from which independent analyses can be obtained. The definition of R has evolved as AEMs have improved and smaller analysis volumes have become possible.

We've already discussed how to define and measure d in TEMs and STEMs way back in Chapter 5, so you need only remind yourself that the beam diameter d is defined as the FWTM of the Gaussian electron intensity. We can measure d directly from the TEM image or indirectly by traversing the beam across a sharp edge and looking at the intensity change on the STEM screen.

This definition takes account of only 90% of the electrons entering the specimen, so it is still an approximation. Remember that the electron-intensity distribution in the incident beam is Gaussian only if you are careful in your choice (small) and alignment of the C2 aperture and you restrict the beam to paraxial conditions (go back and read Section 6.5.A to decide if you should be at the Gaussian image plane or the disk of minimum confusion for best resolution). It is a little more difficult to define and measure b, so this needs more explanation.

FIGURE 36.1. (A) Monte Carlo simulations of 10^3 electron trajectories through a 100 nm Cu foil; (upper) 100 kV; (lower) 300 kV. Note the *improved* resolution at higher kV. (B) Conversely, in a bulk sample, the interaction volume at 30 kV is much larger than that at 10 kV, thus giving *poorer* X-ray spatial resolution at higher kV. The color in both sets of simulations reflects the change in energy of the electrons. Note the relatively constant energy in the thin foil compared with the rapid energy loss in the bulk sample.

36.2.A Beam Spreading

The amount that the beam spreads (b) on its way through the specimen has been the subject of much theoretical and experimental work. While results and theories differ in minor aspects, there is a general consensus that b is governed by the beam energy (E_0), foil thickness (t), and atomic number (Z). It turns out that the simplest theory for b gives a good approximation under most analysis conditions. This theory (sometimes called the 'single-scattering' model because it assumes that each electron only undergoes one elastic scattering event as it traverses the specimen) was first given in the seminal paper by Goldstein et al. and re-defined in SI units by Jones.

$$b = 8 \times 10^{-12} \frac{Z}{E_0} (N_v)^{1/2} t^{3/2} \qquad (36.1)$$

where b and t are in m, E_0 is in keV, and N_v is the number of atoms/m^3. In the original derivation, this latter term was given as $(\rho/A)^{1/2}$ and is confusing in that the density may vary considerably from point to point in a multiphase alloy and is generally unknown anyhow. Furthermore, the atomic-weight dependence was the opposite to the atomic-number dependence which is counterintuitive. So using N_v is clearer and you can work the value out from the ratio of the number of atoms/unit cell to the volume of the unit cell, for which you need to know the lattice parameter. This definition again comprises 90% of the electrons emerging from the specimen, so it is consistent with our definition of d.

There is some question as to whether this single-scattering expression adequately describes the behavior of b for either very thin or very thick foils, but it has generally survived the test of time and its strength remains in its simplicity.

You should of course estimate/calculate b *prior* to spending an inordinate amount of time trying to do an experiment that is impossible for lack of sufficient resolution. Prior simulation of the expected resolution versus the necessary resolution to detect the phenomenon of interest can be very useful here, so let's discuss how best to do this, particularly when the specimen geometry is complex (e.g., multiple/overlapping phases) so equation 36.1 is difficult to apply.

FOR YOUR MAC/PC

We recommend that you keep this equation stored in the TEM computer (or your phone) so you can quickly estimate the expected beam spreading in your planned experiment.

When you can't apply equation 36.1, the best alternative is the Monte Carlo computer simulation, which we introduced in Section 2.5, as a way of modeling electron scattering. Such simulations are used in a wide variety of fields, including SEM and EPMA, as well as other nuclear-particle fields, as a quick search of the Web will reveal. A full description of Monte Carlo simulations is beyond the scope of this text but good reference books exist on the topic. In Joy's book, you'll find a code listing for a Monte Carlo simulation program which can be run on a PC. The public-domain Monte-Carlo programs that do the best job are WinCASINO and WinXRAY from Gauvin at McGill (see URLs #1 and #2); thin-foil versions of this software are under development. Until these are available, we recommend Joy's software. These simulations are now extremely rapid, and in a few minutes on a PC or Mac, they can provide all the information you need to estimate the beam spreading in more complex microstructures.

Basically, the Monte Carlo technique simulates, in a random manner (hence the name), a feasible set of electron paths through a defined specimen. After simulating ideally several thousand paths, an approximate value of b can be obtained by asking the computer to calculate the diameter of a disk at the exit surface of the specimen that contains 90% of the emerging electrons. This definition of b is consistent with that described at the start, and is the dimension of b given by equation 36.1. In fact, the schematic trajectories in Figure 36.1A and B are Monte Carlo simulations using Joy's software. Figure 36.2 shows Monte Carlo simulations of electron trajectories at three points across an interface between Cu and Au. Such a complex situation with elements of radically different Z cannot be easily handled by the single-scattering model estimates of b from equation 36.1.

While beam spreading is the main aspect of spatial-resolution theories, we mustn't forget that what we really want to know is the beam-specimen interaction volume, which corresponds to the size of the X-ray source. Monte Carlo simulations can help because in principle they can

- Incorporate the effects of different kVs and beam diameters
- Handle difficult specimen geometries, specimen tilt, thickness variations, and multi-phase specimens
- Automatically calculate the effect of the depth distribution of X-ray production, $\varphi(\rho t)$, on the X-ray source size
- Display the X-ray distribution generated anywhere in your specimen, as a function of all the variable parameters in equation 36.1, N_v, Z, and t. This tells you the relative contributions to your XEDS spectrum from different parts of the microstructure.

In addition to the theories of beam spreading that we've discussed, there are several more in the literature. A common feature of these theories is that they all predict a linear relationship between b and $t^{3/2}$ and an

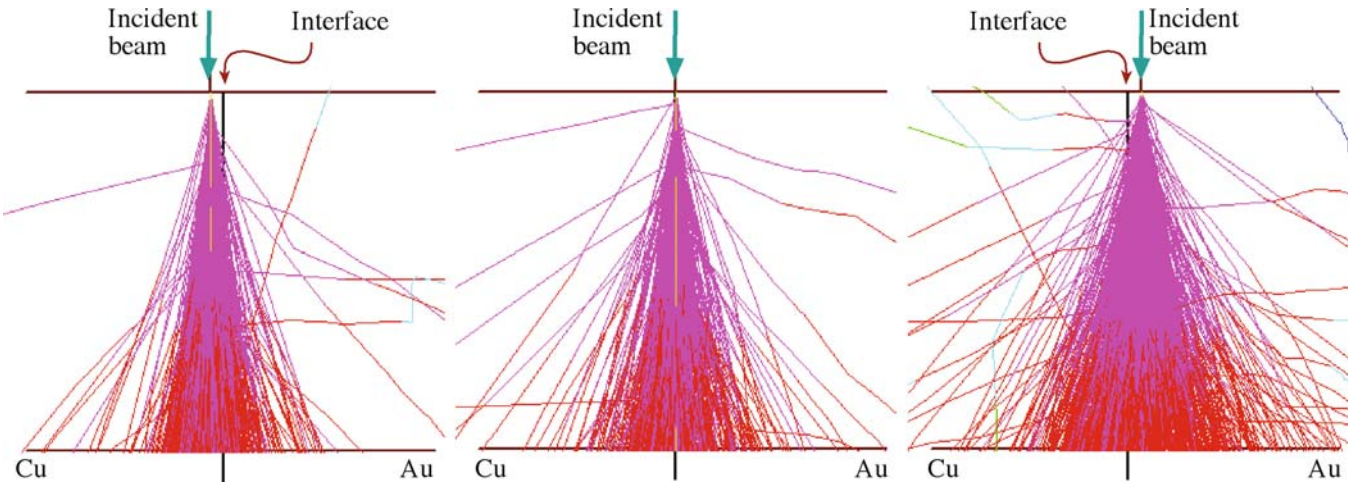


FIGURE 36.2. Monte Carlo simulation of electron trajectories across an interface between two metals of different Z, in which the scattering is very different. Note the rapid increase in the electron scattering in the higher Z region and therefore, X-rays would come from larger regions, thus lowering the local spatial resolution.

inverse relationship between b and E_0. If you're interested in the details of the various theories you'll find a discussion in the Goldstein et al. 1986 paper. However, we'll also see (look ahead to Figure 36.9 in Section 36.3.E) that there are ways to determine the spatial resolution on-line while you're doing your analyses and/or mapping and this is undoubtedly the best approach, since it combines the simple equations we'll now discuss with actual experiments rather than calculations, which make assumptions about your specimens.

36.2.B The Spatial-Resolution Equation

Now we've defined d and b, all we have to do is combine them to come up with a definition of R. If the intensity distribution of the incident beam is Gaussian, and if the beam emerging from the specimen retains a Gaussian form, it is reasonable to add b and d in quadrature (just as we did for image resolution back in Section 6.6.B) to give a value for R

$$R = \left(b^2 + d^2\right)^{\frac{1}{2}} \tag{36.2}$$

Gaussian beam-broadening models are also available, based on equation 36.1, which permit convolution of the Gaussian descriptions of d and b to come up with a definition of R. Based on the Gaussian model and experimental measurements, Michael et al. proposed that the definition of R be modified so as not to present the worst case (given by the exit-beam diameter) but to define R midway through the foil, as shown in Figure 36.3

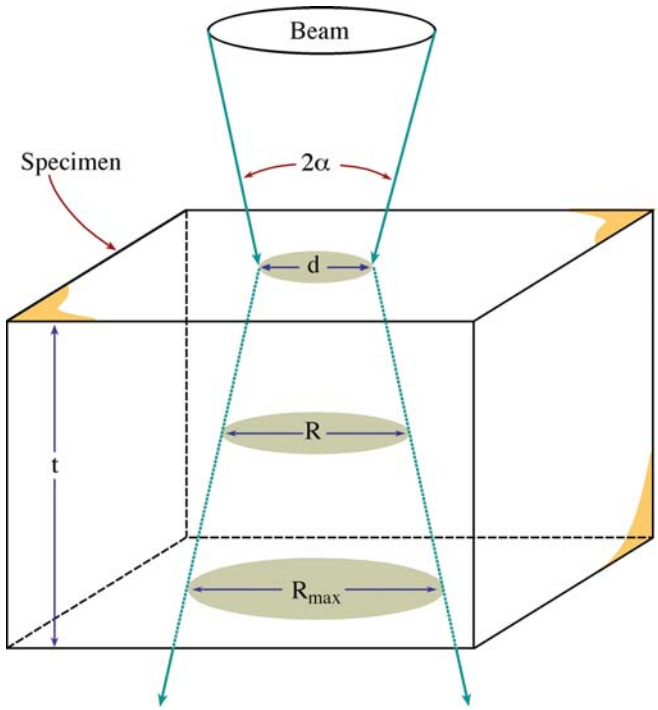


FIGURE 36.3. Schematic diagram of how the incident beam size and the beam spreading combine to degrade the exit-probe diameter to R_{max}, thus defining R.

$$R = \frac{d + R_{max}}{2} \tag{36.3}$$

where R_{max} is given by equation 36.2.

Like all definitions of spatial resolution, there is no fundamental justification for the choice of various factors, such as the FWTM diameter and the selection of the mid-plane of the foil at which to define R. Similarly,

this approach ignores any contribution of electron diffraction in crystalline specimens and beam tailing beyond the 90% limit. Nevertheless, the definition has been shown to be consistent with experimental results and sophisticated Monte Carlo simulations (Williams et al.). Finally, this definition retains the advantage of the original single-scattering model, i.e., it has a simple form and is easily amenable to calculation.

RESOLUTION IN XEDS

Equation 36.3 is the formal definition of the X-ray spatial resolution.

36.2.C Measurement of Spatial Resolution

Experimental measurements of the spatial resolution, such as composition profiles measured across atomically sharp interfaces, are very useful (go back and look at the profiles in Figure 1.4D). Several other kinds of specimens have been proposed but using interphase interfaces retains its validity since there are fewer unknowns than for the other specimens. If thermodynamic equilibrium exists on either side of the interface, the solute content of each phase is well defined. Also, interphase interfaces are common to many engineering materials, as is evident from many images of such defects throughout this text.

In order to compare experimental and calculated measurements of R, you have to understand how we relate the measured composition profile across the interface to the actual discrete profile shape, shown schematically in Figure 36.4. We do this by deconvolution of the beam shape from the measured profile. The finite beam size d and the effect of b degrade the sharp profile to a width L which is related to R by the following equation

$$R = 1.414L \tag{36.4}$$

Assuming this relationship holds, we just measure the distance L between the 2% and 98% points on the profile, as shown in Figure 36.4. This spread contains 90% of the beam electrons, consistent with our assumption of a 90% (FWTM) incident-beam diameter. In practice, you will find it difficult to measure the 2% and 98% points because of the errors in the experimental data. So you should measure the distance from the 10% to the 90% points on your profile, corresponding to the beam spread containing 50% of the electrons (FWHM), then multiply this distance by 1.8 to give the FWTM.

Nevertheless, this definition is easy to remember, relatively easy to measure, consistent with the

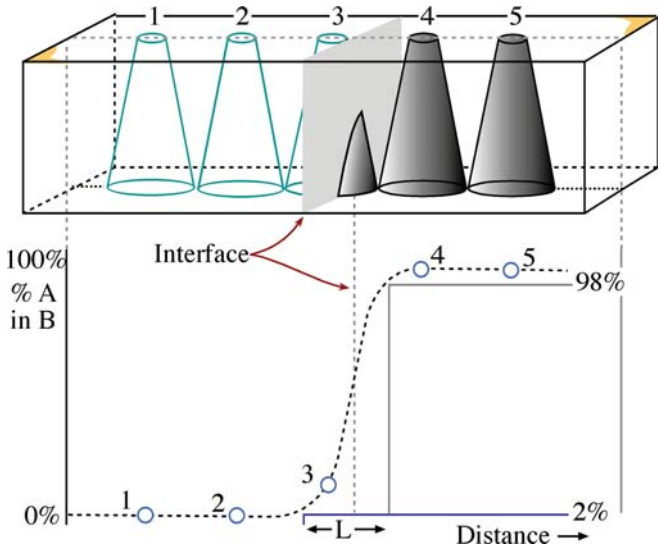


FIGURE 36.4. Schematic diagram showing a composition profile measured across an interface at which an atomically discrete composition change occurs (like the simulation in Figure 36.2). The measured spatial resolution can be defined in terms of the extent (L) of the measured profile between the 2% and 98% points.

definitions of b and d, and, most importantly, gives a number that is close to the experimentally measured degradation of discrete composition changes introduced by the beam-specimen interaction.

It is obvious from equation 36.2, that if we want to improve spatial resolution, then both d and b must be minimized. Unfortunately, if we minimize d we reduce the input beam current: for thermionic sources, if d <10 nm, count rates will become unacceptably low. However, with a FEG, sufficient current (~1 nA) can be generated in a small enough (1 nm) beam to permit quantitative analyses with high spatial resolution.

DEFINING *R, b,* AND *d*

Note that this definition of R, like the definitions of b and d that we have used, is arbitrary.

So if you have a thermionic source TEM

- Your specimen has to be thick enough that sufficient counts are generated for quantification and b will be the main contributor to R.
- Alternatively, you may have to increase the beam size such that d dominates rather than b.

A large beam is needed in that example in order to generate sufficient beam current to get a reasonable X-ray count rate at 100 kV. This is why 200–300 kV FEG TEMs are the best high-resolution analysis instruments

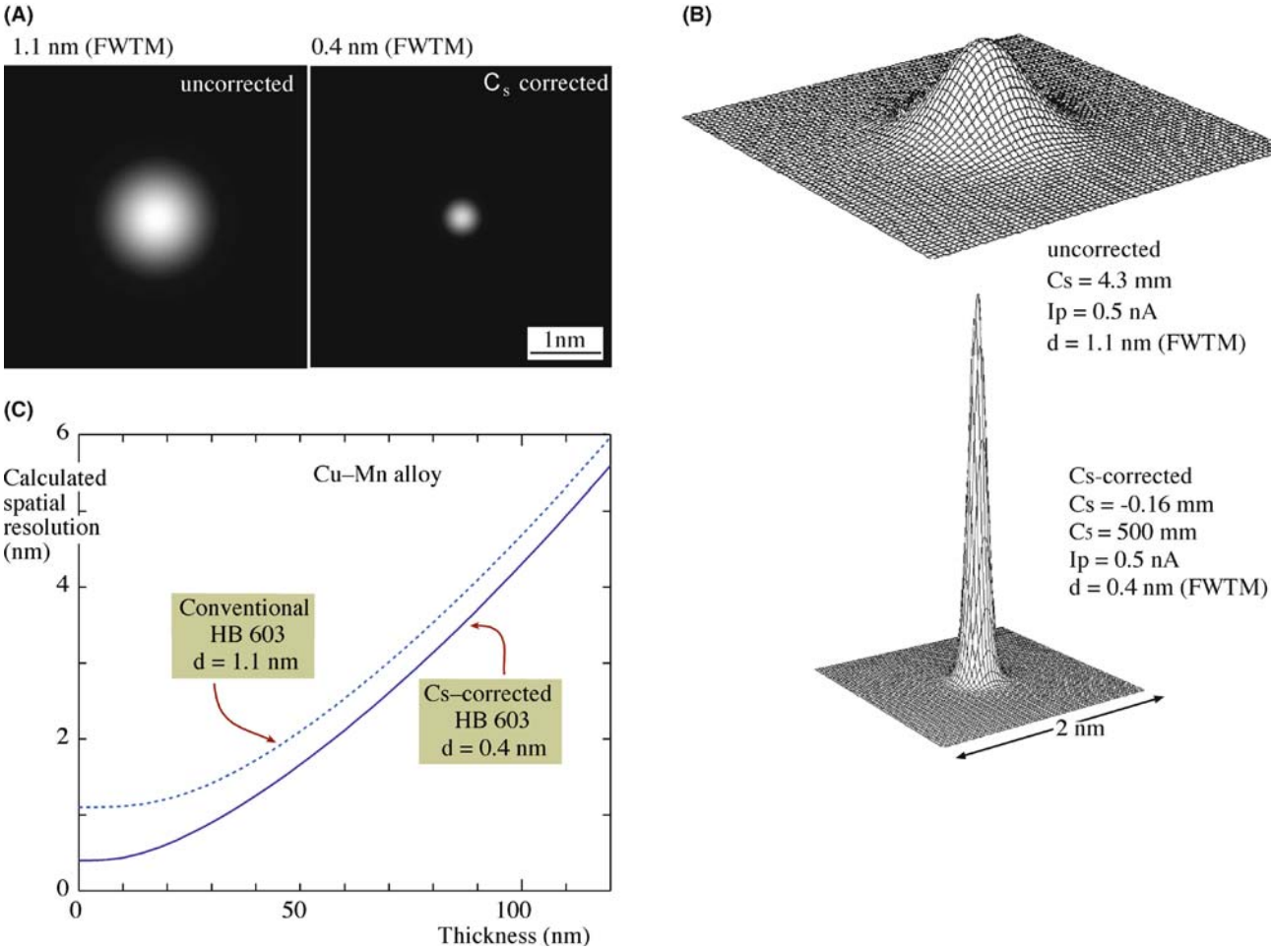


FIGURE 36.5. Simulated probe images for a VG HB-603 300-kV FEG STEM with and without C_s correction; (A) probe dimensions in plan view, (B) 3D intensity distributions. The simulations assume the same probe current (0.5 nA) and C_s correction results in a $\sim 3\times$ decrease in the FWTM probe size from ~1.1 to ~0.4 nm. (C) Calculation of the effect of C_s correction on the spatial resolution in a Cu-Mn alloy as a function of decreasing foil thickness. The resolution at zero foil thickness is improved also by $\sim 3\times$ from just >1 to ~0.4 nm. Compare Figure 36.5B with Figure 2.11 and wonder!

and C_s correction further improves the resolution (go back and compare the segregation-profile widths in Figure 35.11C and D) because you can keep the same probe current while reducing the probe size by a factor of $\sim 3\times$. New X-ray detectors with larger collection angles would also help.

There are some practical factors that can also limit your experimental spatial resolution and the most important is specimen drift. If your specimen or probe drifts for mechanical or electrical reasons, then drift-correction software should be used. If you're planning to carry out analysis at the highest spatial resolution where you're obliged to count for long times to accumulate adequate X-ray intensity, then such software is indispensable.

In summary, the spatial resolution R is a function of both the beam size and the beam spreading. You can get a good estimate of R from equation 36.3. The theories all indicate a $t^{3/2}$ dependence for b, so thin specimens are essential for the best resolution. Intermediate-voltage FEG sources, especially when augmented with C_s correctors, give sufficient beam current in sub-nm probes to generate reasonable counts even from very thin specimens and invariably give the best spatial resolution which, as you can see from the calculations in Figure 36.5, can approach atomic dimensions.

36.3 THICKNESS MEASUREMENT

Given the $t^{3/2}$ dependence of the beam spreading, you can see the importance of knowing t when estimating the spatial resolution. You already know that t is also an essential parameter in correcting for the absorption of characteristic X-rays, as we saw in Section 35.6. Furthermore, you should remember that knowledge of t is important in high-resolution phase contrast imaging and CBED. You'll also see in Chapter 39 that minimizing t is critical to obtaining the best ionization-edge spectra in EELS. So, in almost all TEM techniques, your specimen

has to be as thin as possible to get the best results (although some CBED studies and many in-situ experiments are notable exceptions to this generalization).

So let's take the opportunity here to summarize the methods available for measuring thickness. The methods are many and varied, and a full discussion of the most important techniques will be found in other parts of this book. The first point to consider is, what is t?

THE REAL THICKNESS

The thickness we are interested in is t; this is the thickness through which the beam penetrates. It is not necessarily the same as t_0.

This value of t depends both on the tilt of the specimen γ, and the true thickness at zero tilt, t_0. As shown in Figure 36.6, for a parallel-sided foil

$$t = \frac{t_0}{\cos\,\gamma} \tag{36.5}$$

If your specimen is wedge-shaped, then t and t_0 will vary in an arbitrary fashion depending on the foil shape.

36.3.A TEM Methods

In the TEM you can always make an estimate of your specimen thickness if it is wedge-shaped (and crystalline). By tilting to two-beam conditions for strong dynamical diffraction, the BF and DF images both show thickness fringes, as we saw in Section 24.2. These fringes occur at regions of constant thickness. The intensity in the BF image falls to zero at a thickness of $0.5\xi_g$ at $\mathbf{s} = 0$. Therefore, to determine t, all you have to do is look at the BF image and count the number (n) of dark fringes from the edge of the specimen to the analysis region. At that point $t = (n-0.5)\xi_g$ assuming that the thinnest part at the edge is $< 0.5\xi_g$ thick. (Be very careful with this assumption.) Remember that the value of ξ_g varies with diffracting conditions and so the $\mathbf{g}$ vector has to be specified. You can calculate ξ_g from the expression

$$\xi_g = \frac{\pi\Omega\cos\theta}{\lambda f(\theta)} \tag{36.6}$$

where Ω is the volume of the unit cell, λ is the electron wavelength, and $f(\theta)$ is the atomic scattering amplitude. Remember also that if you're not exactly at $\mathbf{s} = 0$ then the effective extinction distance ξ_{eff} must be used.

A related method relies on the presence of an inclined planar defect adjacent to the analysis region. The projected image of the defect, again under two-beam conditions, will exhibit fringes, which can be used to estimate the local thickness, or the projected width, w, of the defect image using the expression

$$t_0 = w\,\cot\,\delta \tag{36.7}$$

as shown in Figure 36.7 in which δ is the angle between the beam and the plane of the defect. Again, you have to compensate geometrically to measure t rather than t_0 if the foil isn't normal to the beam, and then

$$t = w(\cos\delta - \tan\gamma) \tag{36.8}$$

Of course, neither of these methods is applicable to non-crystalline materials, and it is not always possible to find a suitable inclined defect next to the analysis region in a crystal. Furthermore, two-beam conditions are not recommended for analysis because of the dangers of anomalous X-ray emission (see Section 35.9 on ALCHEMI for both an explanation of, and an exception to, this generalization). More insidious is the fact that oxidation, during or after specimen preparation,

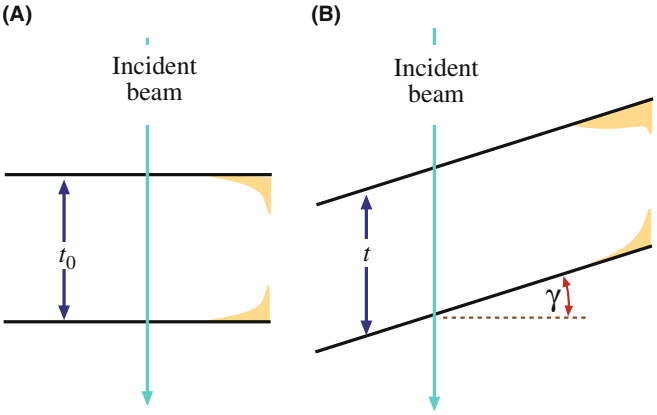


FIGURE 36.6. (A) The specimen thickness t_0 is equal to t, the distance traveled by the beam through a parallel-sided foil at zero tilt. (B) The beam travels a longer distance, t, in a specimen tilted through an angle γ and thus the beam will spread more in a tilted foil.

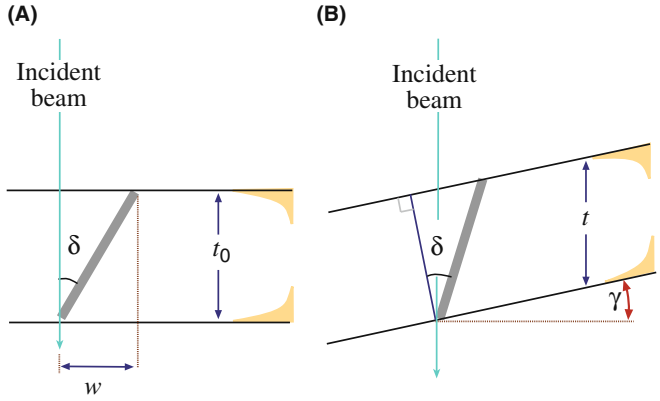


FIGURE 36.7. The parameters required to measure the specimen thickness t_0 from a planar defect (projected width, w), inclined to the incident beam by angle δ. Comparison of (A), a specimen normal to the beam, with (B) a specimen tilted through an angle γ gives some indication of the complexity of determining the appropriate thickness, t, to put into the beam-spreading equation.

means that your crystalline specimen may be coated with an amorphous layer which will not be measured by these diffraction-contrast techniques.

Another method related to the TEM-image contrast involves measurement of the relative transmission of electrons. The intensity on the TEM screen decreases with increasing thickness, all other things being equal. You can use a Faraday cup to calibrate the intensity falling on the screen and from this you can get a crude measure of relative thickness, which can be converted into an absolute measure of t if some absolute method is used for calibration. But you must be careful to make all the intensity measurements on your specimen under the same diffraction conditions and the same incident-beam current but with no objective aperture. The only advantage of this approach is that it is applicable to all materials, both amorphous and crystalline, but it is tedious and not very accurate.

36.3.B Contamination-Spot Separation Method

This method, common in old (S)TEMs, but also commonly used by makers of dirty specimens, relies on the propensity of old instruments or contaminated foils to generate carbon peaks on both top and bottom surfaces of the specimen, at the point of analysis. If you tilt your specimen by a large enough angle (γ), you can see discrete contamination spots (Figure 36.8). Their separation r, at a screen magnification M, is related to t_0 by the following expression

$$t_0 = \frac{r}{M \sin \gamma} \tag{36.9}$$

Matters get a little more complicated if the specimen itself is tilted by an angle ϵ when the contamination is deposited. Then, as in the case of tilted planar defect, you have to be careful to measure the thickness which will determine the beam spreading and, as we've taken pains to point out, this is not t_0 (the thickness at zero tilt) but t

$$t = \frac{r}{M} \frac{\cos \varepsilon}{\sin \gamma} \tag{36.10}$$

Although this method is straightforward, it relies on highly undesirable contamination, which obscures the

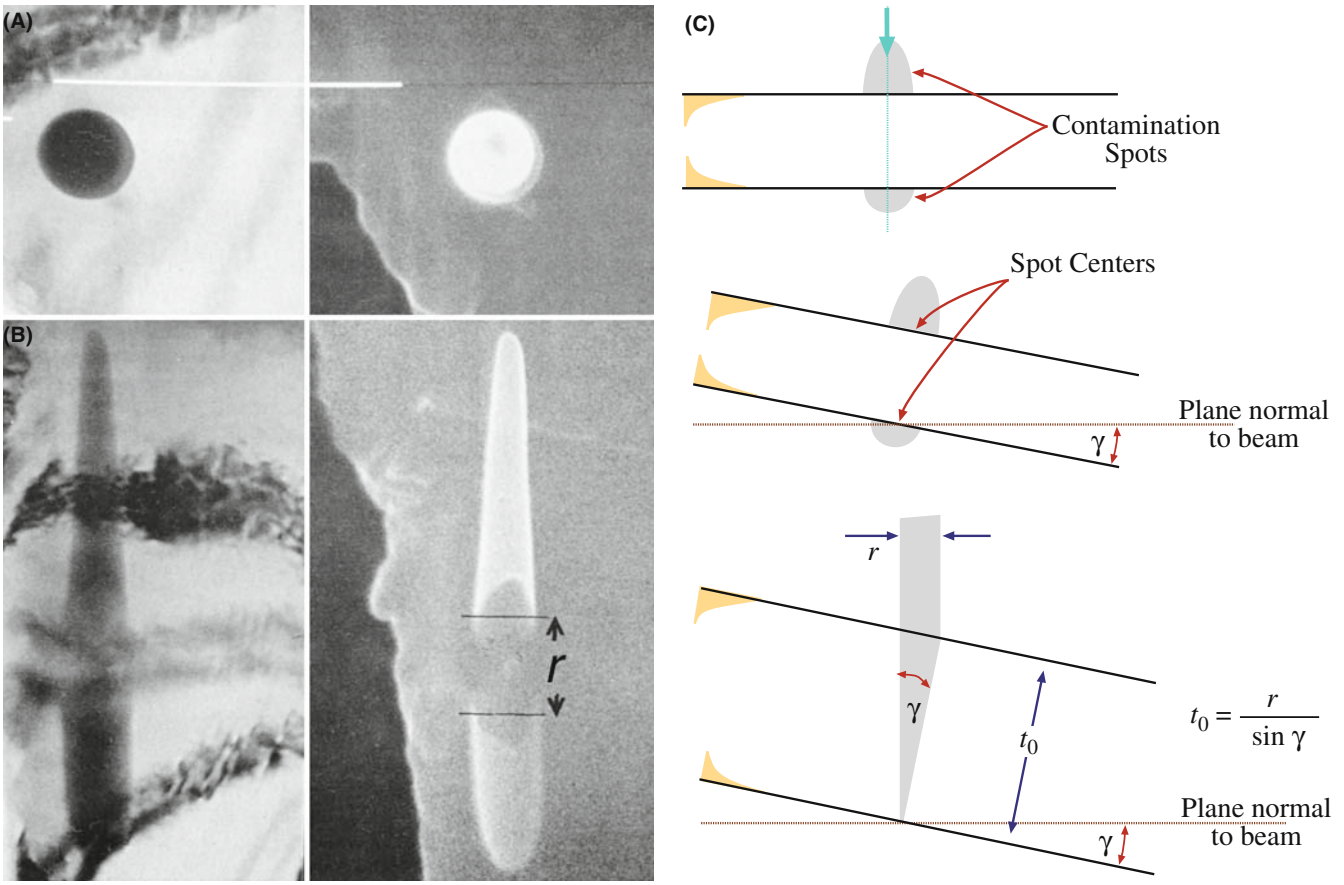


FIGURE 36.8. The contamination-spot separation method for thickness determination; (A) the contamination is deposited on both surfaces of the specimen at zero tilt and the separation (r) is only visible in (B) when the specimen is tilted sufficiently through an angle γ. STEM BF images are on the left and STEM SE images on the right; the SE mode gives the best contrast. (C) Geometrical diagrams of (A) and (B) showing how to determine t_0 from r, the projected separation of the contamination spots.

very area you're looking at. Contamination degrades the spatial resolution and increases the X-ray absorption. In fact, we spend a lot of time and effort trying to minimize contamination, so it would be perverse to propose it as a useful way of determining t. The only redeeming feature is that this method measures t exactly at the analysis point and the shape of the spots can indicate if the beam or the specimen has drifted during analysis. If you find yourself even thinking about using this method, then your TEM should not be used for analysis or you should clean up your specimen-preparation act.

36.3.C Convergent-Beam Diffraction Method

The CBED pattern which is visible on the TEM screen when a convergent beam is focused on the specimen can also be used to determine the thickness of crystalline specimens. In Section 21.2, we described the procedure to extract the thickness from the K-M fringe pattern obtained under two-beam conditions. The CBED pattern must come from a region thicker than $1\xi_g$ or else fringes will not be visible. Also the region of the foil should be relatively flat and undistorted.

Remember that for a totally clean, crystalline specimen, CBED is the way to determine t at specific points in your specimen.

36.3.D Electron Energy-Loss Spectrometry Methods

Thickness information is present in the electron energy-loss spectrum since the intensity of inelastically scattered electrons increases with specimen thickness. In essence, you have to measure the intensity under the zero-loss peak (I_0) and ratio this to the total intensity in the spectrum (I_T). The relative intensities are governed by the mean free path (λ) for energy loss. A parameterization formula for λ (Malis et al. 1988) and other EELS methods, are discussed in detail in Section 39.5.

We can apply the EELS parameterization to any specimen, amorphous or crystalline. But the main advantage is that it is possible to measure the thickness sufficiently quickly that, unlike all the methods described so far, the EELS method can also produce maps of specimen thickness. Given that we have emphasized the value of compositional mapping over point analyses or line profiles, we have to conclude that EELS is best.

EELS FOR *t*

The EELS approach is highly recommended because it is applicable over a wide range of thicknesses and you can produce thickness maps of thin foils.

36.3.E X-ray Spectrometry Method

Since we're talking about X-ray spectrometry, it's good to know that there are X-ray methods which can also determine thickness. We can categorize these approaches, which have been developed to solve the absorption-correction problem (see Section 35.6), into two types: the first is the extrapolation method, which determines the absorption correction by extrapolation of X-ray intensity ratios to zero thickness; the second uses the difference in relative X-ray absorption between two emitted X-ray lines (K and L, or L and M) from the same element. Unfortunately, the extrapolation method is not easily applicable to thin foils where compositions vary locally, since you have to obtain a series of X-ray intensities from different thickness areas (by moving the incident beam or by tilting the specimen). In the intensity-ratio method, the essential requirement of two different X-ray lines from a single element limits the application to specimens which contain elements with $Z > 20$ (Ca).

All of these problems are solved using the ζ-factor approach, as described in detail in 2006 by Watanabe and Williams and in the companion text. The general expression (equation 35.18) for ρt from the ζ-factor analysis can be modified for N different elements in the specimen, thus

$$\rho t = \sum_j^N \frac{\zeta_j I_j A_j}{D_e}, \quad C_A = \frac{\zeta_A I_A A_A}{\sum_j^N \zeta_j I_j A_j}, \cdots$$

$$C_N = \frac{\zeta_N I_N A_N}{\sum_j^N \zeta_j I_j A_j} \tag{36.11}$$

An iterative process is required to solve these equations for both composition and thickness determination. However, the iteration is straightforward and converges rapidly; ~10–15 iterations converge with < 0.001 wt% and 0.01 nm differences in composition and thickness, respectively, which are clearly far more than sufficient tolerance values for termination. Obviously, if X-ray absorption is negligible in a specific material, the initial mass-thickness and compositions are the final values and the iteration is no longer necessary. The essential point here is that in the ζ-factor method, the absorption-corrected compositions can be determined simultaneously with the specimen mass-thickness by *only using X-ray intensity data*.

This method is so quick and versatile that, like the EELS methods, it also permits direct mapping of the thickness at the same time as the composition is being mapped. Therefore, there is nothing to stop you mapping out the spatial resolution at the same time as doing your quantitative mapping, as shown in Figure 36.9. Both the ζ-factor and EELS methods can handle amorphous and crystalline specimens and Ohshima et al. compare the two methods.

In summary, there are many methods for determining t, but none is universally convenient, accurate, and

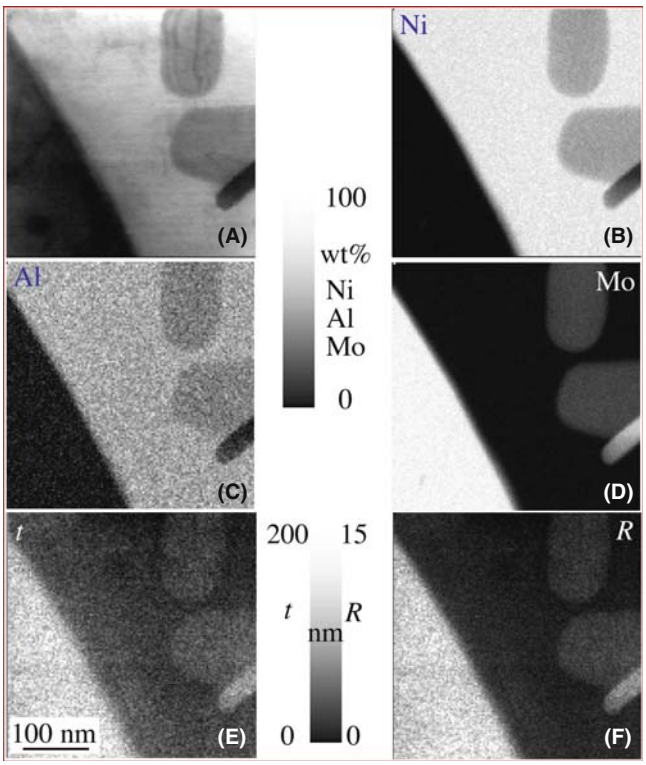


FIGURE 36.9. (A) STEM image and X-ray maps showing the quantitative distribution of (B) Ni, (C) Al, and (D) Mo in precipitates in a Ni-base superalloy. Using the ζ-factor the variation in thickness, t, across the foil can be mapped out (E). Knowing t, the spatial resolution, R, can also be mapped (F). Note the complex interaction of different atomic numbers and the variations in thickness in the resulting variations in R.

applicable. Beware; the various methods also measure different thicknesses, such as only the crystalline thickness, ignoring porosity and/or amorphous surface/oxide films, or the full thickness including porosity and surface films, or just the mass-thickness. Mitchell gives a good case study of how to handle such problems. The EELS and ζ-factor methods both have the possibility of widespread, real-time use and can produce thickness maps, so we recommend these. CBED is very useful for individual point analyses of crystals.

36.4 MINIMUM DETECTION

Minimum detection is a measure of the smallest amount of a particular element that can be detected with a defined statistical certainty. Minimum detection and spatial resolution are intimately related.

A TRUISM

It is a feature of any analysis technique that an improvement in spatial resolution is balanced by a worsening of the detection limit (all other factors being equal).

As the spatial resolution improves, the analyzed volume is smaller and, therefore, the signal intensity is reduced. This reduction in signal intensity means that the acquired spectrum will be noisier and small peaks from trace elements will be less detectable and more easily confused with artifact peaks. Accordingly, in the AEM, the price that is paid for improved spatial resolution is a relatively poor minimum detection. By way of comparison, Figure 36.10 compares the size of the analyzed volume in an EPMA, a TEM/STEM with a thermionic source, and a dedicated STEM with a FEG. The enormous reduction in the beam-specimen interaction volume explains the small signal levels that we obtain in the TEM. However, as we've noted on several occasions, C_s correction gives more current in a smaller probe so it offsets the traditional compromise. But you should now understand why we have spent so much time emphasizing the need to optimize your beam current through use of higher-brightness sources, optimizing the specimen-detector configuration, and so on.

MINIMUM DETECTION

One definition of minimum detection: the minimum mass fraction (MMF) that can be measured in the analysis volume. MMF represents the smallest concentration of an element (e.g., in wt% or ppm).

Alternatively, the minimum detectable mass (MDM) is sometimes used; the MDM describes the smallest amount of material (e.g., in mg or atoms) we can detect.

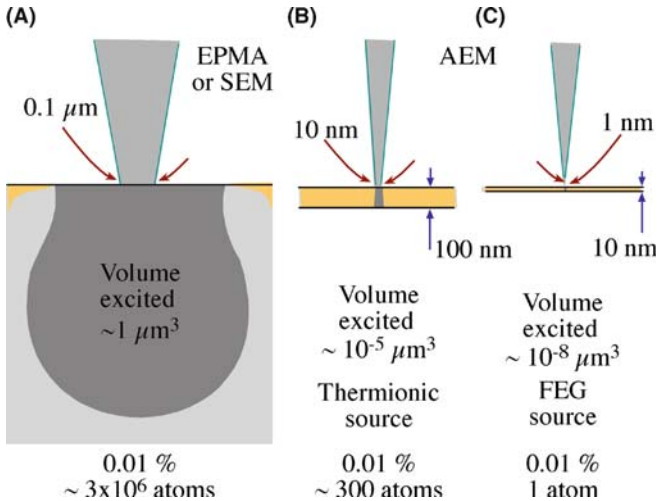


FIGURE 36.10. Comparison of the relative size of the beam-specimen interaction volumes in (A) a SEM/EPMA, (B) a thermionic source AEM, and (C) a FEG-AEM with bulk, thin, and ultra-thin specimens, respectively. The MMF (~0.01%) in each analyzed volume would correspond to ~10^7 atoms, ~300 atoms, and <1 atom, respectively.

We'll use the MMF since materials scientists are more used to thinking of composition in terms of wt% or at.%.

36.4.A Experimental Factors Affecting the MMF

We can relate the MMF to the practical aspects of analysis through the expression of Ziebold

$$\mathrm{MMF} \propto \frac{1}{\sqrt{P(P/B)n\tau}} \tag{36.12}$$

Here P is the X-ray count rate in the characteristic peak (above background) of the element of interest, P/B is the peak-to-background count-rate ratio for that peak (defined here in terms of the same width for both P and B), and τ is the analysis time for each of n analyses.

To increase P you can increase the current in the beam by increasing the probe size and/or choosing a thicker analysis region. To increase P/B you can increase the operating voltage (E_0), which is easy, and decrease instrumental contributions to the background, which is not so easy (Lyman et al.). Improvements in TEM design, such as using a high-brightness, intermediate-voltage source, a C_s corrector if possible, and a larger collection angle for the XEDS will also increase P. To increase P/B, you need a stable instrument with a clean vacuum environment to minimize or eliminate specimen damage and contamination. Improved stage design, to minimize stray electrons and bremsstrahlung radiation, both of which contribute background to the detected spectrum, will also help to increase P/B, as we discussed back in Chapter 33.

Remember that the Fiori definition of P/B is not the one used in Ziebold's equation (36.12). If you actually want to calculate the MMF, go back and check the original references.

The other variables in equation 36.12, are the time of analysis (τ) and how many analyses (n) that you do, which are entirely within your control as operator. Usually both n and τ are a direct function of your patience and the recommended coffee break is usually the maximum time for any one analysis. With computer control of the analysis procedure, however, there should really be no limit to the time available for analysis. Particularly when detection of very small amounts of material is sought, τ should be increased to very long times. As computer control and stage stability improve, acquisitions of several hours or overnight are becoming feasible. Of course, the investment of so much time in a single analysis is dangerous unless you have judiciously selected the analysis region, and you are confident that the time invested will be rewarded with a significant result. Obviously, you should minimize factors that degrade the quality of your analysis with time, such as contamination, beam damage, and specimen drift. Therefore, you should only carry out long analyses if your TEM is clean (preferably UHV) and your specimen is also clean and stable under the beam. Any specimen drift must be corrected by computer control during the analysis, unless your specimen is uniformly thin and homogeneous in composition, in which case why bother analyzing it?

36.4.B Statistical Criterion for the MMF

We can also define the MMF by a purely statistical criterion. We discussed in Section 34.5 that we can be sure a peak is present if the peak intensity is greater than three times the standard deviation of the counts in the background under the peak. From this criterion we can come up with a definition of the detection limit which, when combined with the Cliff-Lorimer equation (assuming Gaussian statistics), gives the MMF (in wt%) of element B in element A as

$$C_B(\mathrm{MMF}) = \frac{3\left(2I_B^b\right)^{1/2}C_A}{k_{AB}\left(I_A - I_A^b\right)} \tag{36.13}$$

where $I^b{}_A$ and $I^b{}_B$ are background intensities for elements A and B; I_A is the raw integrated intensity of peak A (including background); C_A is the concentration of A (in wt%); and $k_{AB}{}^{-1}$ is the reciprocal of the Cliff-Lorimer k-factor. However, if we express the Cliff-Lorimer equation as

$$\frac{C_A}{k_{AB}\left(I_A - I_A^b\right)} = \frac{C_B}{\left(I_B - I_B^b\right)} \tag{36.14}$$

and substitute it into equation 36.13, the MMF is

$$C_B(\mathrm{MMF}) = \frac{3\left(2I_B^b\right)^{1/2}C_B}{I_B - I_B^b} \tag{36.15}$$

Experimentally, low count-rates from thin specimens mean that typical values of MMF are in the range 0.1–1%, which is rather large compared with some other analytical techniques. The best compromise in terms of improving MMF while maintaining X-ray spatial resolution is to use high operating voltages (300–400 kV) and thin specimens to minimize beam broadening. The loss of X-ray intensity, P (or I), a consequence of using thin specimens, can be compensated in part by the higher voltages and/or by using an FEG where a small spot size of 1–2 nm can still be maintained. Obviously C_s-corrected TEMs will help because of their ability to put even more current into the same size probe. Figure 36.11 summarizes the classic

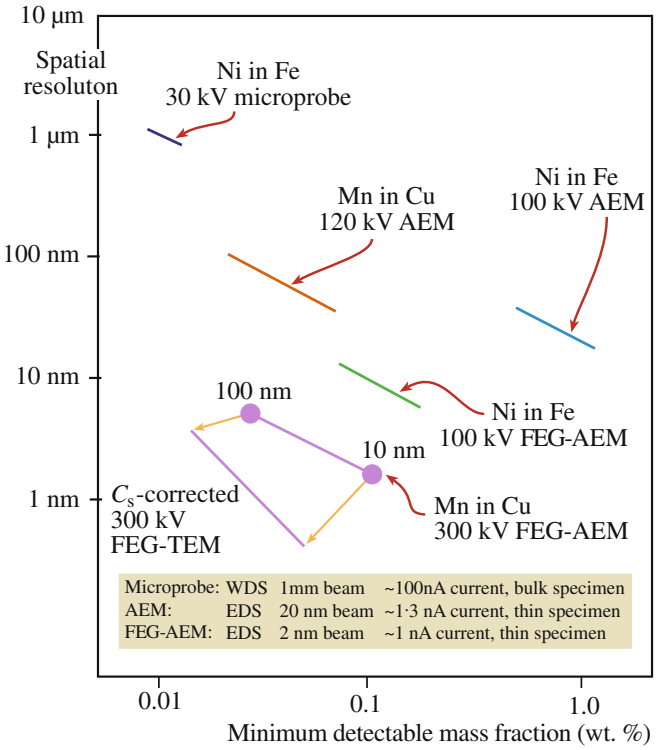


FIGURE 36.11. Calculation of the relationship between MMF and spatial resolution, R, for the EPMA and a range of AEMs. The inverse relationship between the MMF and R is clear, although it is also apparent that the high-brightness sources and high-kV electron beams in the AEM can compensate for the decreased interaction volume in a thin foil. C_s correction results in an enormous improvement in both resolution and sensitivity.

compromise between resolution and detection and how instrumentation improvements have continued to push the limits over the past few decades.

36.4.C Comparison with Other Definitions

The MMF definition is not the only way we can measure detection limits. Currie has noted at least eight definitions in the analytical-chemistry literature. Currie defined three specific limits.

- The decision limit: Do the results of your analysis indicate detection or not (L_C)?
- The detection limit: Can you rely on a specific analysis procedure to lead to detection (L_d)?
- The determination limit: Is a specific analysis procedure precise enough to yield a satisfactory quantification (L_q)?

For I_B counts from element B in a specific peak window and I_B^b in the background it can be shown that

$$L_C = 2.33\sqrt{I_B^b} \tag{36.16}$$

$$L_d = 2.71 + 4.65\sqrt{I_B^b} \tag{36.17}$$

$$L_q = 50\left\{1 + \left(1 + \frac{I_B^b}{12.5}\right)^{\frac{1}{2}}\right\} \tag{36.18}$$

If there are sufficient counts in the background

$$L_d = 4.65\sqrt{I_B^b} \quad \text{when } I_B^b > 69 \tag{36.19}$$

$$L_d = 14.1\sqrt{I_B^b} \quad \text{when } I_B^b > 2500 \tag{36.20}$$

Comparison of these definitions with the statistical criterion in the previous section shows that $C_{MMF} \approx L_d$. So, if you want to quantify an element, not just determine that it is present (L_d), then you need substantially more ($\sim 3\times$) of the element in your specimen. Rather than do the experiment yourself, it is possible to simulate spectra from small amounts of element B in A (or vice versa), using DTSA, as described in the companion text. We recommend that you simulate your analysis before embarking on a time-consuming experiment, which may be futile because the amount of the element you are seeking is below the MMF.

36.4.D Minimum-Detectable Mass

The MMF values of fraction of a percent may seem poor compared with other analytical techniques which report ppm or even ppb detection limits. However, it's a different matter if you calculate what the MMF translates to in terms of the minimum detectable mass (MDM).

Using data for the MMF of Cr in a 304L stainless steel measured in a VG HB-501 AEM with an FEG, Lyman and Michael obtained an MMF of 0.069 wt% Cr in a 164 nm foil with a spatial resolution of 44 nm and a 200 s counting time. The electron beam size was 2 nm (FWTM) with a beam current of 1.7 nA. In this analysis, an estimated 2×10^4 atoms were detected. The MDM was less than 10^{-19} g. If the counting time is increased by a factor of 10 and if the operating voltage is increased to 300 kV, the spatial resolution would improve to ~15 nm and the MMF would improve to ~0.01 wt%. Thus about 300 atoms could be detected. For a foil thickness of 16 nm (1/10th the above measured thickness), the MMF would degrade to ~0.03 wt%. However, the spatial resolution would improve to about 2 nm. For this case, about 20 atoms would be detected corresponding to less than 10^{-22} g, which is an amazing figure by any standard. Experimental verification of this was reported in 1999 by Watanabe and Williams: 2–5 atoms of Mn were detected in a 10-nm thick Cu-Mn alloy film. With the advent of C_s-correction and improved computer data analysis routines, single-atom

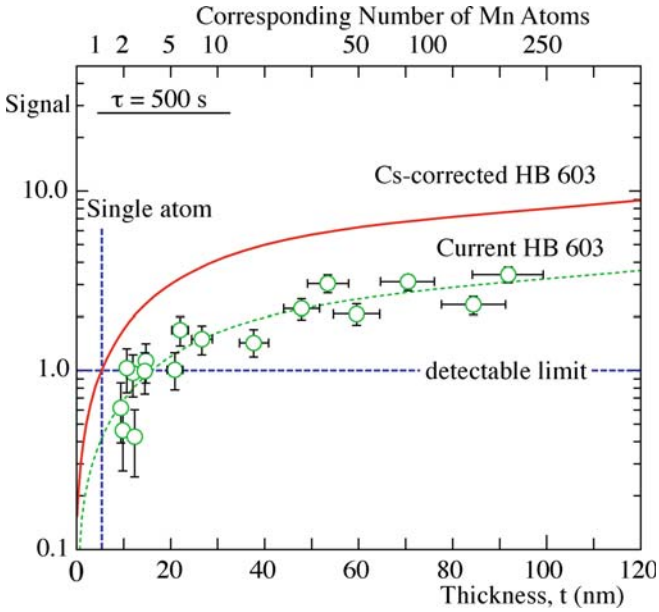


FIGURE 36.12. Calculation of the number of Mn atoms detectable in a Cu-0.1 wt% Mn foil as a function of foil thickness (dotted green line) based on experimental Mn K_α counts (green circles) in a 300-keV FEG STEM. When the Mn K_α signal is undetectable the ordinate axis value = 1 and this occurs when there are between 2 and 5 Mn atoms in the analysis volume (see top axis) which can be calculated knowing the foil thickness and bulk chemistry. A C_s corrector (red line) is calculated to improve the MDM from several atoms to ~1 atom right at the detection limit.

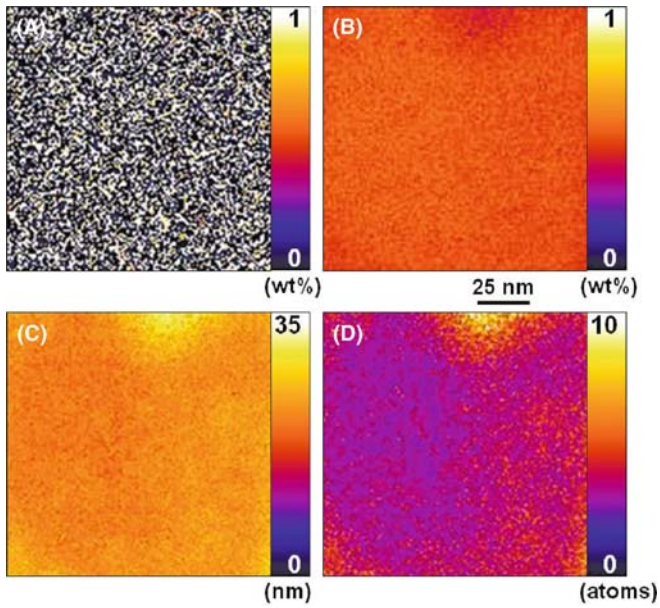


FIGURE 36.13. A series of quantitative maps obtained from a homogeneous Cu-0.5 wt% Mn foil in a C_s-corrected, 300-kV, UHV, FEG STEM. (A) Mn composition map from the original spectrum image, (B) Mn composition map from the spectrum image enhanced by MSA noise reduction, (C) thickness map, and (D) map of a number of Mn atoms. Note the look-up tables with each map. In (D) the dominating purple color corresponds to ~ 2–3 atoms. These maps were quantified by the ζ-factor method.

detection is now a distinct possibility and Figure 36.12 shows calculated improvements in detection limits for Mn atoms in solution in Cu in a C_s-corrected 300-kV FEG TEM. Figure 36.13 shows how mapping of a homogeneous solid solution can detect a few atoms at each pixel. This last figure summarizes just about everything we have discussed for AEM quantification and C_s-corrected mapping: (a) ζ-factor quantification (the image is still noisy), (b) MSA data manipulation (the image is much less noisy), (c) the importance of good thin foils (specimen thickness is very uniform and < 20 nm in most areas), and (d) MDM close to 1–2 atoms is attainable (even when mapping). Just to remind you how good this is, go back and take a look at Figure 36.10. Consider that in the EPMA with a ~1 μm^3 excitation volume and a 0.01 wt% MMF, ~3 million atoms are detected in the analysis volume. So XEDS in the best C_s-corrected, intermediate-voltage, UHV, FEG TEM has an MDM detection limit that is *several million times better* than an EPMA.

MDM

In AEM it is useful to define the MDM as the minimum number of atoms detectable in the analyzed volume.

While the best XEDS-TEM combinations are approaching atomic-level detection and sub-nm spatial resolution, it is not yet possible to detect single atoms within individual atomic columns as is achievable in EELS in similar, C_s-corrected, intermediate-voltage FEG TEMs (see Chapter 39) so there is still room for improvement, e.g., through larger collection angles and more sophisticated data processing.

CHAPTER SUMMARY

Optimizing spatial resolution and minimum detection in the same experiment is always a compromise. You must decide which of the two criteria is more important for the result you're seeking

- To get the best spatial resolution, operate with the thinnest foils and the highest energy electron beam. Use an FEG if possible and a C_s corrector if you're lucky.

- To measure the specimen thickness use the ζ-factor method or the parameterized EELS approach. If neither is possible, use CBED for a crystalline foil. If you're reduced to contamination spots, find a better TEM or make cleaner specimens.
- To get the best MMF, use the brightest electron source, the largest possible beam, and thickest specimen, and count for as long as possible with the shortest time constant.
- If you want the best resolution *and* MMF, a C_s-corrected, intermediate-voltage, UHV, FEG TEM is essential, along with a clean specimen and computer-controlled drift correction; patience is also equally essential.

CLASSICS

Berriman, J, Bryan, R, Freeman, R and Leonard, KR 1984 *Methods for Specimen Thickness Determination in Electron Microscopy* Ultramicrosc. **13** 351–364. Old, but still useful, review of thickness measurements in the TEM.

Goldstein, JI, Williams, DB and Cliff, G 1986 *Quantification of Energy Dispersive Spectra* in *Principles of Analytical Electron Microscopy* 155–217 Eds. DC Joy, AD Romig Jr. and JI Goldstein, Plenum Press New York. Introduction to many of the concepts in this chapter and the preceding one.

Jones IP 1992 *Chemical Microanalysis Using Electron Beams* Institute of Materials London. Redefined Goldstein et al. equation in SI units on p173.

Joy, DC 1995 *Monte Carlo Modeling for Electron Microscopy and Microanalysis* Oxford University Press New York. The only textbook covering this essential topic

THICKNESS AND RESOLUTION

Malis, T, Cheng, SC and Egerton RF 1988 *The EELS Log-ratio Technique for Specimen-Thickness Measurement in the TEM* J. Electr. Microsc. Tech. **8** 193–200.

Michael, JR, Williams, DB, Klein, CF and Ayer, R 1990 *The Measurement and Calculation of the X-ray Spatial Resolution Obtained in the Analytical Electron Microscope* J. Microsc. **160** 41–53.

Mitchell, DRG 2006 *Determination of Mean Free Path for Energy Loss and Surface Oxide Film Thickness Using Convergent Beam Electron Diffraction and Thickness Mapping: a Case Study Using Si and P91 Steel* J. Microsc. **224** 187–196.

Williams, DB, Michael, JR, Goldstein, JI and Romig AD Jr. 1992 *Definition of the Spatial Resolution of X-ray Microanalysis in Thin Foils* Ultramicrosc. **47** 121–132.

P/B, ζ, AND MMF

Currie, LA 1968 *Limits for Qualitative Detection and Quantitative Determination. Application to Radiochemistry* Anal. Chem. **40** 586–593. Detection limits and different definitions.

Goldstein, JI, Costley, JL, Lorimer, G. and Reed, SJB 1977 *Quantitative X-ray Microanalysis in the Electron Microscope SEM 1977* **1** 315–325 Ed. O Johari IITRI Chicago IL. Seminal paper.

Lyman, CE, Goldstein, JI, Williams, DB, Ackland, DW, Von Harrach, S, Nicholls, AW and Statham, PJ 1994 *High Performance X-ray Detection in a New Analytical Electron Microscope* J. Microsc. **176** 85–98. Description of what is still the best X-ray analysis instrument in the world.

Lyman, CE and Michael, JR 1987 *A Sensitivity Test for Energy-dispersive X-ray Spectrometry in the Analytical Electron Microscope* in *Analytical Electron Microscopy-1987* 231–234, Ed. DC Joy, San Francisco Press San Francisco CA. Spatial resolution versus sensitivity limits.

Ohshima, K, Kaneko, K, Fujita, T and Horita, Z 2004 *Determination of Absolute Thickness and Mean Free Path of Thin Foil Specimen by ζ-Factor Method* J. Electron Microsc. **53** 137–142.

Watanabe, M and Williams, DB 1999 *Atomic-Level Detection by X-ray Microanalysis in the Analytical Electron Microscope* Ultramicrosc. **78** 89–101.

Watanabe, M and Williams, DB 2006 *The Quantitative Analysis of Thin Specimens: a Review of Progress from the Cliff-Lorimer to the New ζ-Factor Methods* J. Microsc. **221** 89–109. The ζ-factor approach for quantification, mapping, thickness determination, and everything else.

Ziebold, TO 1967 *Precision and Sensitivity in Electron Micro-probe Analysis* Anal. Chem. **39** 858–861. Just what it says!

URLs

1) http://montecarlomodeling.mcgill.ca/software/winxray/contacts.html

SELF-ASSESSMENT QUESTIONS

Q36.1 Why do we spend so much time discussing the spatial resolution of XEDS?
Q36.2 Define R, b, and d.
Q36.3 Why are there so many variable definitions of the spatial resolution in the literature?
Q36.4 What's the most important factor controlling the spatial resolution?
Q36.5 Why do you sometimes have little control over this specific factor?
Q36.6 Why is it challenging to measure the spatial resolution experimentally?
Q36.7 List the various methods of determining your specimen thickness and under each method list its most important advantage and its greatest disadvantage.
Q36.8 What's the most important factor in controlling the detection limits in any experiment?
Q36.9 What's the difference between the MMF and the MDM?
Q36.10 Why is a high peak (P) intensity more important than a high P/B ratio when trying to improve the detection limit?
Q36.11 Why do we choose 90% of the exit electron distribution to define the spatial resolution? Why not choose 100%? Why not choose 50% which is commonly used when calculating probe-limited image resolution?
Q36.12 Why is an interphase interface often chosen as the ideal feature across which to measure the experimental spatial resolution of analysis?
Q36.13 Can you suggest other specimens that might offer similar advantages?
Q36.14 Why is an FEG the best electron source to use if you want the highest spatial resolution?
Q36.15 Is an FEG necessarily the best source to use if you want to obtain the highest analytical sensitivity? If not, why not?
Q36.16 What improvements might be gained in both spatial resolution and analytical sensitivity from using an aberration-corrected electron probe?
Q36.17 Why isn't it a good idea to rely on the contamination-spot method to estimate your specimen thickness?
Q36.18 Define the decision limit, the detection limit, and the determination limit.
Q36.19 If a 300-keV FEG AEM can detect 0.01 wt% of an element in a foil ~5 nm thick, estimate how many atoms this represents in the analyzed volume. State any assumptions; be brief.

TEXT-SPECIFIC QUESTIONS

T36.1 Why does a higher voltage give higher spatial resolution in a thin specimen in the AEM but lower spatial resolution in a thick, EPMA specimen as shown in Figure 36.1?
T36.2 Figure 36.1A does not take into account electron-diffraction effects; why does this not seriously compromise our estimation of spatial resolution?
T36.3 The situation in Figure 36.3 assumes that all the incident beam is confined on the entrance surface of the specimen in a circular probe of diameter b. List several factors that can make this assumption unreasonable. (Hint: go back and look at Figure 33.5.)
T36.4 Look at Figure 36.4. Is there any advantage to be gained by moving the interaction cones closer together (i.e., taking more point analyses in the profile)? (Hint: go and look at equation 36.3.)
T36.5 Estimate, from Figure 36.4, the maximum angle at which the interface could be tilted before the spatial resolution profile is degraded beyond the usual experimental limits.
T36.6 Tilting the specimen (see Figure 36.6) degrades the spatial resolution. What other disadvantages occur when the specimen is tilted, and under what analytical conditions is there an advantage to tilting the specimen?
T36.7 To a first approximation, calculate what it would take to detect a single atom of element B in the analyzed volume of element A such as in Figure 36.10. State any assumptions.
T36.8 If you compare the left and right diagrams in Figure 36.10, it is clear that if a FEG AEM is to exhibit comparable analytical performance to an EPMA, it has to be millions of times more sensitive, since the analyzed volume is smaller by such a factor. Indicate what technical differences exist between the two techniques such that this extraordinary improvement in signal detection and generation actually occurs.
T36.9 Explain clearly why the trend in Figure 36.11 is common to all microanalysis techniques, i.e., improving spatial resolution invariably results in degrading the minimum detection limit.
T36.10 Which method(s) would you use to determine the thickness of (a) SiO_2 glass, (b) SiO_2 crystal, (c) Cu-4% Al? Justify your choice of method in each case.
T36.11 Use DTSA to determine the minimum amount of P impurity detectable in a spectrum from otherwise pure Fe. Your specimen is 100 nm thick, and you are operating at 200 kV with a take-off angle of 20°. Courtesy M. Watanabe
T36.12 Using the results of the previous question (MMF of P in Fe), calculate a number of detectable P atoms in Fe (MDM) for a LaB_6-AEM (incident beam size, $d = 10$ nm) and a FEG-AEM ($d = 2$ nm). For the density and the atomic weight to calculate the beam broadening, use the values for Fe, since the detection limits of P are low enough to ignore any effects of this element (if you have answered correctly). Courtesy M. Watanabe

37

Electron Energy-Loss Spectrometers and Filters

CHAPTER PREVIEW

Electron energy-loss spectrometry (EELS) is the analysis of the energy distribution of electrons that have come through the specimen. These electrons may have lost no energy or may have suffered inelastic (usually electron-electron) collisions.

These energy-loss events tell us a tremendous amount about the chemistry and the electronic structure of the specimen atoms, which in turn reveals details of their bonding/valence state, the nearest-neighbor atomic structure, their dielectric response, the free-electron density, the band gap (if there is one), and the specimen thickness.

In order to examine the spectrum of electron energies, we invariably use a magnetic-prism spectrometer. As we saw with X-rays, it is now common to form images of the various EELS signals in addition to gathering spectra and to do this we use an energy filter, which is based on the same magnetic-prism concept. Energy-filtered TEM (EFTEM) is perhaps the most powerful AEM technique, as should become apparent. The magnetic prism/energy filter is a highly sensitive device with an energy resolution <1 eV, even when the electron-beam energy is as high as 300 keV.

In this chapter, we'll describe the operational principles of both the spectrometer and filter, how to focus and calibrate them, and how to determine the collection semi-angle (β). This angle affects the quality and interpretation of much of your experimental data. In subsequent chapters, we'll go on to look at the spectra and images in more detail and the information they contain.

The EELS technique is an excellent complement to, and is now more widely used than the somewhat simpler XEDS, since it offers substantially more information than mere elemental identification and is well suited to the detection of light elements, which are difficult to analyze with XEDS.

37.1 WHY DO EELS?

So why should we do EELS when XEDS can identify and quantify the presence of all elements above Li in the periodic table with a spatial resolution approaching a few atoms and analytical sensitivity close to the single-atom level? Well, in fact EELS does even more than XEDS in that it can detect and quantify *all* the elements in the periodic table and is especially good for analyzing the light elements. Furthermore, EELS offers even better spatial resolution and analytical sensitivity (both at the single-atom level) in addition to providing much more than just elemental identification as indicated in the preview. So the next question is, why bother with XEDS at all? The answer to this is that EELS, as you will see, can be a challenging experimental technique; it requires very thin specimens to get the best information and understanding and processing the spectra and images requires somewhat more of a physics background than XEDS.

The main point that you need to understand is that XEDS and EELS are highly complementary techniques; most AEMs come equipped with both kinds of spectrometer.

37.1.A Pros and Cons of Inelastic Scattering

When a high-energy electron traverses a thin specimen it can either emerge unscathed or it loses energy by a variety of processes that we first discussed way back in Chapter 4. EELS separates these inelastically scattered electrons into a spectrum, which we can interpret and

quantify, form images or DPs from electrons of specific energy, and also combine the spectra and images via spectrum imaging, as we've already seen for XEDS (see Section 33.6.C). Throughout the book, we've already seen some contrasting aspects of inelastic scattering

- Kikuchi lines and HOLZ lines occur in DPs; the electrons in these lines are diffracted very close to the Bragg angle, and give us much more accurate crystallographic information than the SAD/CBED (spot/disk) pattern. In thick specimens, many of the electrons in these lines are inelastically scattered, so thicker specimens can be useful.
- Conversely, the background intensity that surrounds the direct beam in DPs, obscuring faint spots in SADPs and fine detail in CBDPs is due to inelastic scatter. So if you can remove (i.e., filter out) these electrons, you can considerably enhance the quality of your DPs.
- Chromatic aberration is due to energy-loss electrons following different paths through the objective lens and limits the image resolution in thick specimens. You can avoid this by using very thin specimens, but if you're stuck with a thick specimen then you can restore the quality of your images by filtering out inelastic electrons thus removing specimen-induced chromatic-aberration effects.
- Specimen damage, which is usually undesirable, is often caused by inelastic interactions and there isn't much you can do about this, as we discussed back in Chapter 4.
- In Chapter 4, we also discussed many of the other ways that electrons could lose energy going through the specimen, producing X-rays and other phenomena such as plasmons and phonons. These and other interactions produce energy-loss electrons, which contain useful information about the electronic structure and elemental makeup of your specimen.

In fact, if your AEM has an energy-loss spectrometer or filter, then inelastic scattering, in general, is something you should want to happen in your specimens because, if it didn't happen, TEM would be a much less useful technique and this book would be several chapters shorter (so we at least are happy).

The technique of EELS predates X-ray spectrometry. In fact, the experimental pioneers of EELS, Hillier and Baker (1944), were the same two scientists who first proposed and patented the idea of X-ray spectrometry in an electron-beam instrument, similar to the EPMA. If you want to read a brief history of the technique, see the classic book by Egerton. We'll refer to the second edition of Egerton's text on many occasions. In contrast to X-ray analysis, EELS was relatively slow to develop, but is now perhaps the dominant spectrometry technique on probe-forming TEMs for the reasons we gave at the start of the chapter.

Once you've finished this set of chapters, you should find Egerton's book to be highly informative and likewise, the multi-author text on energy-filtered imaging, edited by Reimer. Every 4 years since 1990, there has been an international, focused workshop on EELS and related techniques in the TEM (URL #1) and the proceedings have been published in special issues of various EM journals, as noted in the reference list.

37.1.B The Energy-Loss Spectrum

Let's start as we did for XEDS by looking at a typical spectrum such as Figure 37.1. We won't go into detail here but will save that for the subsequent chapters. For the time being, it's worth pointing out that we split up the spectrum into the low-loss and high-loss regions, with ~ 50 eV being the somewhat arbitrary break point. It isn't particularly apparent from the figure, but we'll see later that the low-loss region contains electronic information from the more weakly bound conduction and valence-band electrons, while the high-loss region contains primarily elemental information from the more tightly bound, core-shell electrons and also details about bonding and atomic distribution. For the time being, you should note that

- the zero-loss peak is very intense, which can be both an advantage and a hindrance
- the intensity range is enormous; this graph uses a logarithmic scale as the only way to display the whole spectrum
- the low-loss regime containing the plasmon peak (see Chapter 38) is relatively intense

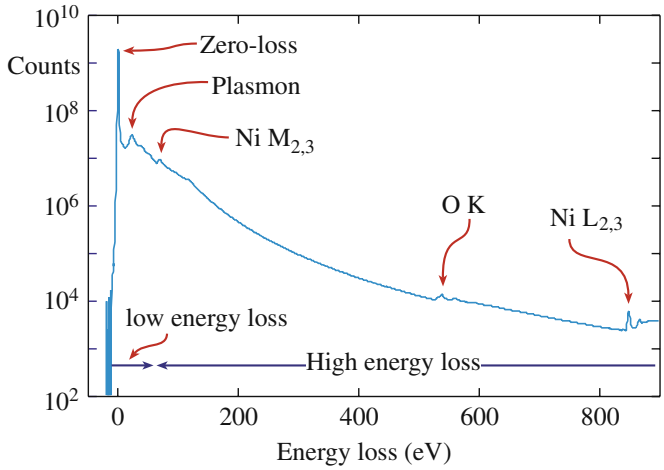


FIGURE 37.1. An EELS spectrum displayed in logarithmic intensity mode. The zero-loss peak is an order of magnitude more intense than the low energy-loss portion (characterized by the plasmon peak), which is many orders of magnitude more intense than the small ionization edges identified in the high energy-loss range. Note the relatively high (and rapidly changing) background.

- the element-characteristic features called ionization edges (see Chapter 39), are relatively low in intensity compared to the background
- the overall signal intensity drops rapidly with increasing energy loss, reaching negligible levels above ~2 keV, which really defines the energy limits of the technique (and this is about the energy when XEDS really comes into its own, emphasizing again their complementarity).

37.2 EELS INSTRUMENTATION

Throughout this and the subsequent chapters, we'll distinguish between spectrometers and energy filters (so called because they filter out electrons of specific energy). The former primarily produce spectra and are great for the dedicated spectroscopist. The latter can produce spectra but are designed to create images and so are more useful to microscopists who find spectra boring and who want filtered images and DPs to compare with standard ones produced by their TEM. There is only one kind of spectrometer commercially available, manufactured by Gatan, Inc. termed a parallel-collection EELS or PEELS. The PEELS is a magnetic-prism (sometimes called a magnetic-sector) system and is mounted on a TEM or STEM after the viewing screen or post-specimen detectors. There are two kinds of filters which result in radically different instruments which, nevertheless, perform similar functions. We'll describe these two types in some detail throughout the chapter.

The post-column Gatan Image Filter (GIF) is a development of their magnetic-prism PEELS. The in-column filter, a magnetic variant of the original Castaing-Henry magnetic prism/electrostatic mirror, is exemplified by the Omega (Ω) filter, pioneered by Zeiss and also now used by JEOL. The in-column filter, as the name implies, is integrated into the TEM and sits between the specimen and the viewing screen/detector (rather than being an optional addition, like the PEELS/GIF). Magnetic spectrometers, along with electrostatic or combined electrostatic/magnetic systems, have been the subject of serious reviews by Metherell and Egerton. If you're an instrument enthusiast, you should try to read these articles but if you just want a more concise summary, then try Egerton's chapter in Ahn's edited book.

ENERGY FILTERS

Two types of commercial filters are presently manufactured: the post-column filter and the in-column filter.

Surface scientists use electron spectrometers to measure exceedingly small (meV) energy losses in low-energy electron beams reflected from the surfaces of samples in UHV instrumentation, such as XPS and Auger systems. We will deal only with transmission EELS studies of high-voltage TEM beams. Recent advances in monochromators for electron guns, combined with spherical and chromatic aberration correction, mean that energy resolution <100 meV is now possible and in the near future this will open a whole new field of TEM-based EELS studies, complementing the established, surface-science techniques but with much better lateral resolution.

Whether you use a Gatan post-column PEELS, a GIF, or an Ω filter, the electron passes through one or more magnetic spectrometers, so we'll start by discussing the principles of this basic tool before moving on to the more complex filtering systems.

37.3 THE MAGNETIC PRISM: A SPECTROMETER AND A LENS

The magnetic prism spectrometer is preferred to an electrostatic or combined magnetic/electrostatic spectrometer, for several reasons

- It is compact and easily interfaced to the TEM. (Remember the WDS problem.)
- It offers sufficient energy resolution to distinguish spectra from all the elements in the periodic table and so is ideal for analysis.
- Electrons in the energy range 100–400 keV, typical of AEMs, can be dispersed sufficiently to detect the spectrum electronically, without limiting the energy resolution.

The basic PEELS-TEM interface and ray paths are shown in Figure 37.2 and a picture of a Gatan spectrometer (actually a combination imaging spectrometer), which would be installed beneath the camera system of a TEM, is shown in Figure 37.3. Because these spectrometers are so widespread, many of the numerical values in this chapter are taken from the Gatan literature (see URL #2). For the details of operation you should, of course, read the instruction manual and Brydson gives a concise summary of all the experimental steps.

From Figure 37.2B, you can see that electrons are selected by a variable entrance aperture (diameters: 1, 2, 3, or 5 mm in the Gatan system). (Obviously, you have to make sure the screen is raised and any on-axis detectors or cameras are removed in order to detect the spectrum.) The electrons travel down a 'drift tube' (Figure 37.2A) through the spectrometer and are deflected through $\geq 90^\circ$ by the magnetic field. Electrons that have lost

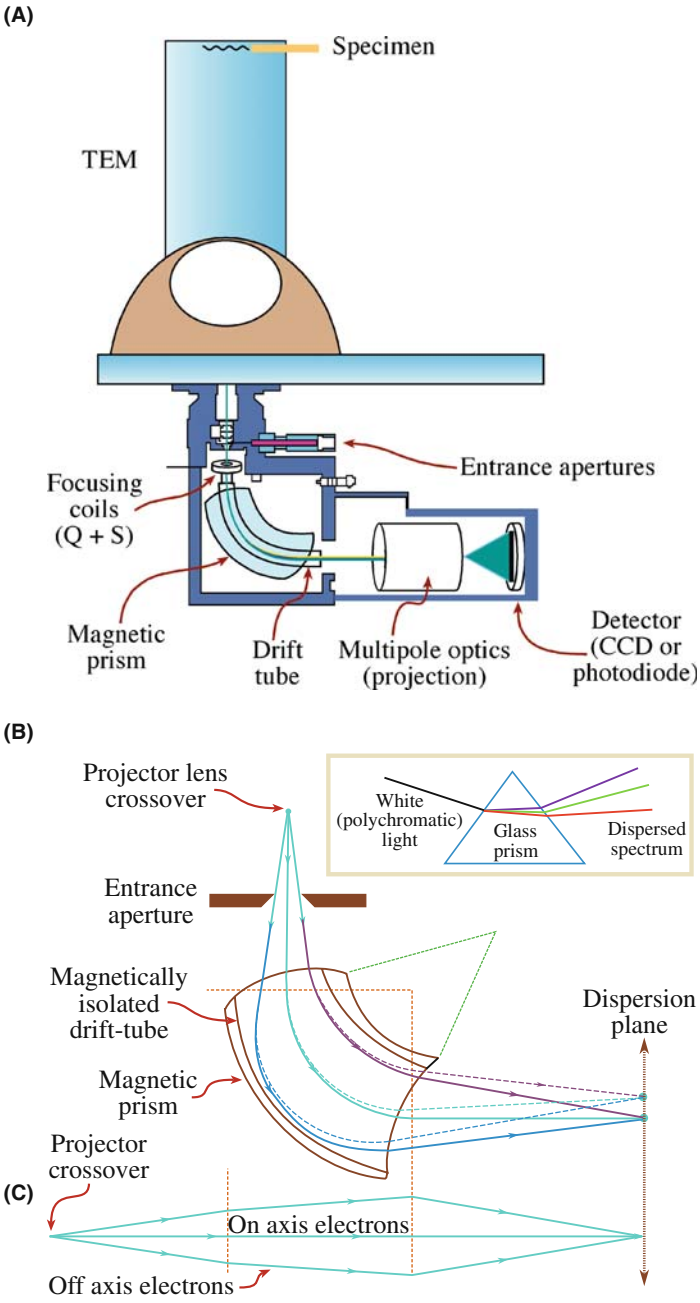


FIGURE 37.2. (A) Schematic diagram showing how a PEELS is interfaced below the viewing screen of a TEM and the position of the various components. (B) Ray paths through a magnetic prism spectrometer showing the different dispersion and focusing of the no-loss and energy-loss electrons in the image (dispersion) plane of the spectrometer. The inset shows the analogy with the dispersion of white light by a glass prism. (C) The lens focusing action in the plane normal to the spectrometer.

energy are deflected further than those suffering zero loss. A spectrum is thus formed in the dispersion plane, consisting of a distribution of electron intensity (I) versus energy loss ($\mathcal{E}$). You can see that this process is closely analogous to the dispersion of white light by a glass prism shown in the inset.

FIGURE 37.3. A Gatan Tridiem PEELS which interfaces below the viewing screen of an AEM.

NOTATION

Although we've consistently used the letter E for energy, energy-loss should, therefore, be denoted by ΔE since it's a change in energy. However, it is a convention in the EELS literature to use E interchangeably for both an energy loss (e.g., the plasmon loss E_p) and a specific energy (e.g., the critical ionization energy E_C). So we will use $\mathcal{E}$ (note the different font) but remember, it really means a **change** in E.

Now if you look at Figure 37.2B, you'll see that electrons suffering the same energy loss but traveling in both on-axis and off-axis directions are also brought back to a focus in the dispersion (or image) plane of the spectrometer. So the prism also acts as a magnetic lens. This focusing action is not seen in the otherwise-analogous glass prism (if you recall your high-school physics lab, you had to use a post-prism convex lens to focus the spectrum and separate the individual colors (i.e., frequencies/energies)). We'll give many examples of EEL spectra in the subsequent chapters.

37.3.A Focusing the Spectrometer

Because the spectrometer is also a lens, you have to know how to focus it, and how to minimize the aberrations and astigmatism that are inherent in any magnetic lens. The latest spectrometers are fully corrected for third-order aberrations and the alignment, compensation for stray-AC fields, and focusing are all software controlled. So read the manual, because we will not reproduce it here, merely describe the principles.

The spectrometer has to focus the electrons because off-axis electrons experience a different magnetic field to on-axis electrons. The spectrometer is an axially *asymmetric* lens unlike the other TEM lenses. The path length of off-axis electrons through the magnet also varies, and the magnet has to be carefully constructed to ensure correct compensation for different electron paths so that focusing occurs. This correction is achieved by machining the entrance and exit faces of the spectrometer so they are not normal to the axial rays, as shown in

Figure 37.2B. These non-normal faces also act to ensure that electrons traveling out of the plane of the paper in Figure 37.2B are also focused in the dispersion plane, as shown in Figure 37.2C (because it focuses in two planes we call it 'double focusing'). The faces of the prism are curved to minimize aberrations and, like their counterparts within the TEM column, the spectrometer lenses continue to improve as higher-order aberrations are minimized. The many quadrupoles, sextupoles, and other focusing electronics and lenses are not shown, but the length of the spectrometer in Figure 37.3 (and looking ahead to Figure 37.15) gives you some idea of its complexity.

As with any lens, the spectrometer takes electrons emanating from a point in an object plane and brings them back to a point in the image (dispersion) plane. Because the spectrometer is an asymmetric lens, we have to fix both the object distance and image distance if we want to keep the spectrum in focus. The object plane of the spectrometer or filter depends on the detail of the machine you are using

- In DSTEMs with no post-specimen lenses the object plane for a post-column spectrometer is the plane of the specimen.
- In a TEM/STEM, or a DSTEM with post-specimen lenses, the object plane for a post-column spectrometer is the back-focal plane of the projector lens, which can contain either an image or a DP.
- In a TEM/STEM, or a DSTEM with post-specimen lenses, the object plane for an in-column filter is the back-focal plane of the first projector (or intermediate) lens, which can contain either an image or a DP.

In a TEM, the projector-lens setting is usually fixed, so the object plane is fixed and the manufacturer usually sets this plane to coincide with the differential pumping aperture separating the column from the viewing chamber. In some DSTEMs there are no post-specimen lenses so the object plane of the spectrometer is the plane of the specimen. In this case, it is essential that you keep your specimen height constant.

In practice, the back-focal plane of the projector may move slightly as you change operating modes (for example, from TEM to STEM) and so you have to be able to adjust the spectrometer. You do this by looking at the electrons that come through the specimen without losing any energy. These electrons have a Gaussian-shaped intensity distribution called the zero-loss peak (ZLP), which we'll talk about more in the next chapter. You can see the ZLP on the computer display of the EELS system and the software focuses it by adjusting a pair of pre-spectrometer quadrupoles until it has a minimum width and maximum height.

The consistency of the information passed through the spectrometer is described by the transmissivity, which is the imaged area (object radius) as a function of the solid scattering angle for a given energy resolution. A perfect spectrometer would uniformly transmit electrons of a specific energy loss over the whole spectrum or, more importantly, over the whole image. If the ZLP is scanned across the spectrometer slit, then a uniform intensity should fall on the detector. Deviations from uniformity reflect the aberrations in the spectrometer; see the paper by Uhlemann and Rose for the details.

37.3.B Spectrometer Dispersion

We define the dispersion as the distance in the spectrum (dx) between electrons differing in energy by dE. It is a function of the strength of the magnetic field (which is governed by the strength (i.e., size) of the spectrometer magnet) and the energy of the incident beam E_0. For the Gatan magnet, the radius of curvature (R) of electrons traveling on axis is about 200 mm, and for 100-keV electrons dx/dE is ~ 2 μm/eV. For PEELS this dispersion value is inadequate and typically electrons with an energy range of about 15 eV would fall on each 25-μm wide diode. Therefore, the dispersion plane has to be magnified $\sim 15\times$ before the spectrum can be detected with resolution closer to 1 eV. This magnification requires post-spectrometer lenses and four quadrupoles are used. The dispersion should be linear across the PDA (photo-diode array); you can check this by measuring the separation of a known pair of spectral features (e.g., zero loss and C K edge) as you displace the spectrum across different parts of the PDA.

37.3.C Spectrometer Resolution

We define the energy resolution of the spectrometer as the FWHM of the focused ZLP and, while it might seem trivial, you should remember to focus the spectrometer every time you acquire a spectrum or filtered image. The type of electron source determines the resolution. As we saw back in Chapter 5 (Table 5.1), at ~ 100 keV a W source has the worst energy resolution ($\sim$3 eV), a LaB_6 is slightly better at $\sim$1.5 eV, a Schottky field emitter can give $\sim$0.7 eV, and a cold FEG gives the best value of $\sim$0.3 eV. These values will all get slightly worse at higher keV. Because of the high emission current from thermionic sources, the energy resolution is limited by electrostatic interactions between electrons at the filament crossover. This electron-electron interaction is called the Boersch effect. We can partially overcome this limit by undersaturating the filament and using only the electrons in the halo. Then a LaB_6 source can attain a resolution of $\sim$1 eV but at the expense of a considerable loss of current, which we can compensate for by increasing the beam size and/or the C2 aperture. There are other ways you can improve the energy resolution, for example, dropping the kV and the probe current. Figure 37.4A shows data from a cold FEG operated at 200 keV delivering a FWHM of 370 meV (0.37 eV) which is about the best that can be obtained under standard operating conditions.

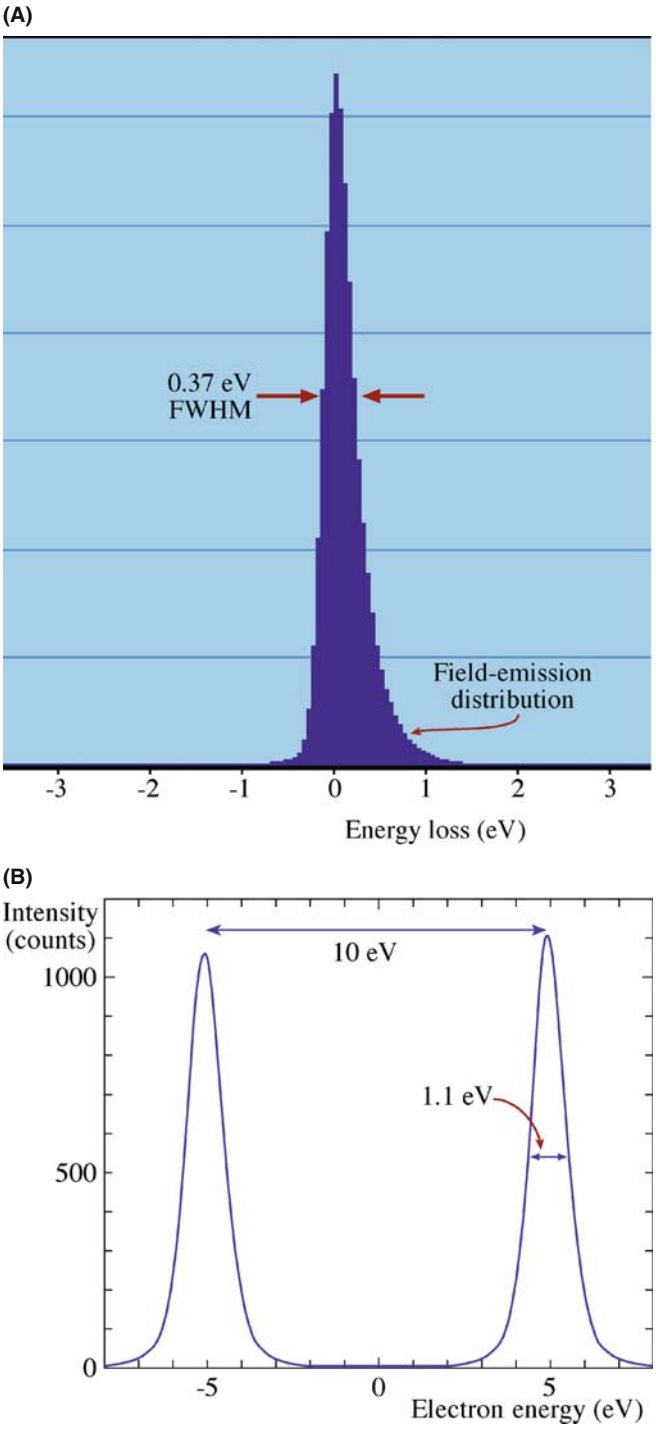


FIGURE 37.4. (A) The energy resolution (0.37 eV) of a cold FEG at 200 keV with 150 pA of current determined from the FWHM of the ZLP. The peak is not symmetrical because of electrons tunneling out from the tip with a slight (< 1 eV) loss of energy. (B) The intensity profile across the ZLP exposed on a CCD camera, then displaced by 10 eV and re-exposed. The resolution, defined by the FWHM, in this case is 1.1 eV and if the number of channels between the peak centroids is counted (e.g., 100 channels) then the dispersion (0.1 eV/channel) is easy to calculate.

The energy resolution decreases slightly as the energy loss increases, but it should be no worse than ~1.5× the ZLP width up to 1000 eV energy loss.

If you operate at a higher voltage you should also expect a degradation of energy resolution as the kV increases, approximately tripling from 100 to 400 kV.

Because the magnetic prism is so sensitive, external magnetic fields in the microscope room may limit the resolution. If you have an older PEELS, you may see a disturbance to your spectrum if you sit in a metal chair and move around or if you open metal doors into the TEM room, so fine, hand-carved chairs are *de rigeur* for old EELS operators.

ENERGY RESOLUTION

For comparison: for XEDS it's >100 eV; for EELS it's <1 eV.

The best energy resolution requires a small projector crossover and a small (1 or 2 mm) entrance aperture giving a collection semi-angle of ~ 10 mrads (see Section 37.4.B below). A larger entrance aperture degrades the resolution because the off-axis beams suffer aberrations. The resolution may change as you deflect the ZLP onto different regions of the PDA, although this should not happen if your spectrometer is properly aligned.

37.3.D Calibrating the Spectrometer

We can calibrate the spectrometer (in terms of eV/channel, just like an XEDS) by placing an accurately known voltage on the drift tube, or changing the accelerating voltage slightly, both of which displace the spectrum by a known, fixed amount. Figure 37.4B shows images of the ZLP displaced by a known amount, thus defining both the resolution and the dispersion of the spectrometer at the same time. Usually, calibration is automatically handled by the software but, if you have a really early system, you can (as in XEDS) look for features in a spectrum from a known specimen that occur at specific energies, such as the ZLP at 0 eV and the Ni L_3 ionization edge at 855 eV. Modern electronics are reasonably stable and the calibration doesn't shift substantially but, unless you have the PEELS, which automatically compensates for energy and current drift, you have to check these regularly throughout an operating session since, if shifts of even a few eV occur, they are of the same order as the energy resolution of the spectrometer.

37.4 ACQUIRING A SPECTRUM

To gather a spectrum, such as shown in Figure 37.1, we need a recording device in the dispersion plane of the spectrometer. Historically this was photographic film and early commercial PEELS had a semiconductor PDA but now a CCD is used, in common with both

GIFs and Ω filters (we talked about CCDs in Chapter 7). CCDs show lower gain variation, ~30× better sensitivity, higher dynamic range, and improved energy resolution compared with PDAs. It's worth a brief aside to mention serial EELS or SEELS which was the first commercial method of recording spectra around 1980. While it offered some advantages, SEELS was slow and tedious since each energy loss channel was recorded sequentially and it was rapidly superseded by PEELS which gathers the whole energy range simultaneously. The Gatan PEELS uses a YAG scintillator coupled via fiber optics to a PDA or CCD in the dispersion plane of the spectrometer, as shown in Figure 37.5. Since many PDA-PEELS systems are still in use, it is worth describing its operation and limitations, The PDA consists of 1024 electrically isolated and thermoelectrically cooled Si diodes, each ~25 μm across. The integration time to gather a spectrum can vary from a few msec to several hundred seconds depending on the intensity of the signal. Because each diode saturates at ~16,000 counts you have to select an integration time that avoids saturation during any single acquisition, and you can then sum as many individual integration times as you need to give a spectrum of the desired quality. The CCD detector is 100 × 1340 array of 20 μm pixels and does not saturate rapidly like the PDA so acquisition is much more straightforward. After integration, the whole spectrum is read out via an amplifier through an A/D converter and into the computer display system. The Gatan software offers a variety of standard acquisition conditions suited to different types of spectra.

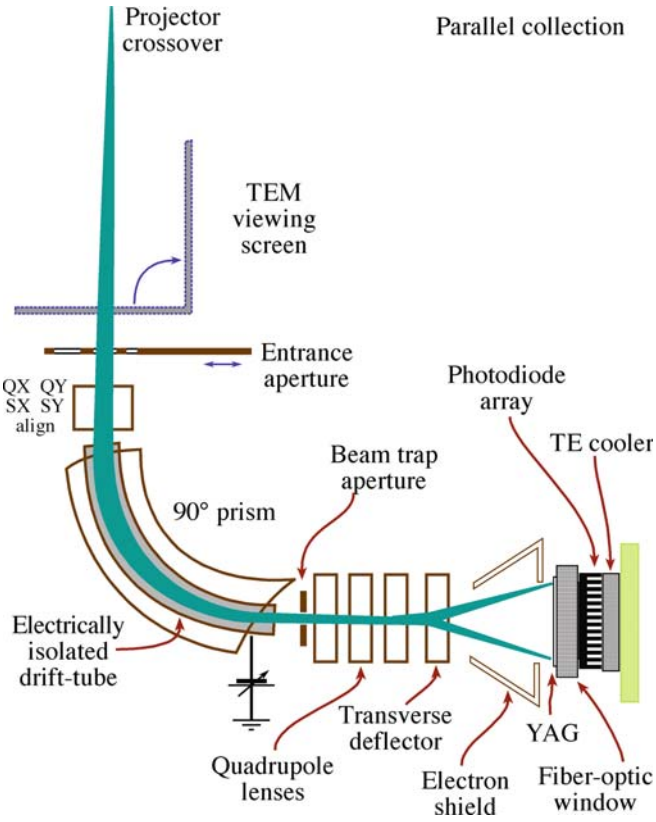


FIGURE 37.5. Schematic diagram of parallel collection of the energy-loss spectrum onto a YAG scintillator fiber-optically coupled to a semiconductor PDA.

37.4.A Image and Diffraction Modes

When using any spectrometer or filter in a TEM/STEM, we can operate in one of two modes, and the terminology for this is confusing. If we operate the TEM such that an image is present on the viewing screen then the back-focal plane of the projector lens contains a DP, which the post-column spectrometer uses as its object. From the spectroscopist's viewpoint, therefore, this is termed 'diffraction mode' or 'diffraction coupling,' but from the microscopist's viewpoint, it is more natural to call this 'image mode' since we are looking at an image on the screen. Conversely, if you adjust the microscope so a DP is projected onto the screen (which includes STEM mode in a TEM/STEM), then the (post-column) spectrometer/GIF object plane contains an image, and the terminology is reversed. Likewise, as we'll see for an in-column filter, the back-focal plane of the intermediate lens on which the filter is focused can contain either an image (diffraction mode; there's a DP on the TEM screen) or a DP (image mode; there's an image on the TEM screen).

- The spectroscopist uses the term *image coupling* and the microscopist says *diffraction mode*.
- In this text, *image mode* means an image is present on the TEM screen; naturally, we use the microscopists' terminology. You should *not* use this mode for any spectroscopy, only for EFTEM imaging.
- You should use *diffraction (or STEM) mode* for all spectroscopy and imaging, except for EFTEM imaging.
- Both sets of terms appear in the earlier literature, often without precise definition, so it can be rather confusing, but the microscopists have generally won this conflict.

37.4.B Spectrometer-Collection Angle

The collection angle (as before, we really mean semi-angle) of the spectrometer (β) is a most important variable in several aspects of EELS, so you should know β for all your usual operating modes. If you do gather spectra with different β it is difficult to make sensible comparisons without considerable post-acquisition processing. Poor control of the collection angle is the most common error in quantification, although it is less important the higher the value of β that you use. The detailed intensity variations in the spectrum depend on the range of electron-scattering angles gathered by the spectrometer. Under certain circumstances, the effective value of β can be modified if the beam-convergence

angle α is > β, but we'll discuss that when we talk about quantification of ionization edges in Chapter 39. Likewise, we'll see that there is a characteristic or most-probable scattering angle for specific loss processes and typically, β should be 2–3× that angle. So, if in doubt, make β larger. The value of β is affected by your choice of operating mode, and so we will describe how to measure β under different conditions that you may encounter and, while the Gatan software can calculate the value for you, it's good to know the principles behind the black box. Let's start by considering the most simple definition for β, as illustrated in Figure 37.6.

SEMI-ANGLE β

β is the semi-angle subtended at the specimen by the entrance aperture to the spectrometer or filter.

Dedicated STEMs. In a basic DSTEM the situation is straightforward if there are no post-specimen lenses because, as shown in Figure 37.6, β can be calculated from simple geometry. Depending on the diameter (d) of the spectrometer entrance aperture and the distance from the specimen to the aperture (h), β (in radians) is given by

$$\beta \approx \frac{d}{2h} \tag{37.1}$$

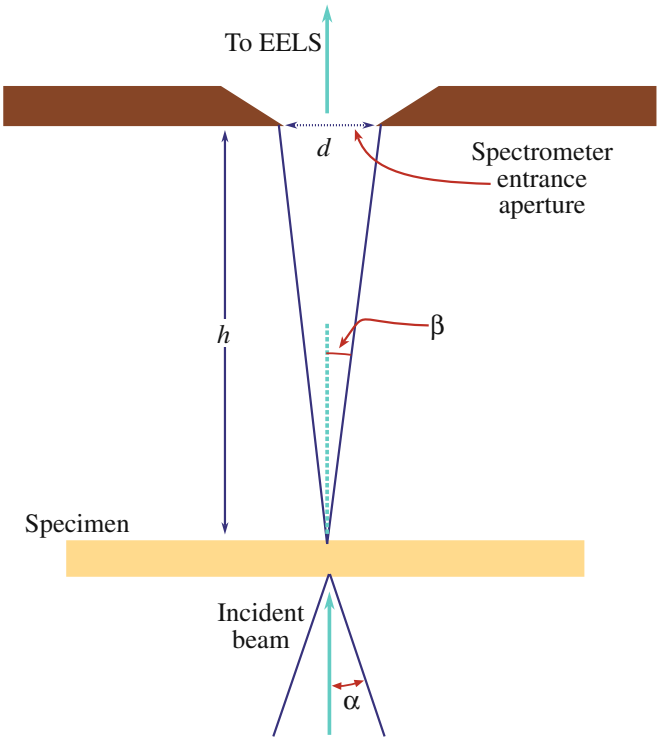


FIGURE 37.6. Schematic diagram showing the definition of β in a DSTEM in which no lenses exist between the specimen and the spectrometer entrance aperture.

This value is approximate and assumes β is small. Since h is not a variable, the range of β is controlled by the number and size of the spectrometer apertures. Therefore, if h is ~100 mm, then for a 1-mm diameter aperture, β is 5 mrads. If there are post-specimen lenses and apertures, the situation is similar to that in a TEM/STEM, as we discuss below.

TEM-image mode. Remember that, in image mode, a magnified image of the specimen is present on the viewing screen or detector. In contrast to what we just described for a dedicated STEM, the angular distribution of electrons entering the spectrometer aperture below the center of the TEM screen is *independent* of the entrance-aperture size. This is because we control the angular distribution of electrons contributing to any TEM image by the size of the objective aperture in the back-focal plane of the objective lens. If we don't use an objective aperture then the collection angle is very large (>~100 mrads) and need not be calculated accurately because we'll see that small differences in a large β do not affect the spectrum or subsequent quantification.

If, for some reason, you do wish to calculate β in image mode with no aperture inserted, you need to know the magnification of the DP in the back-focal plane of the projector lens (which is the front-focal plane of the spectrometer). As you will recall, this magnification is controlled by the camera length L of the DP, and this is given by

$$L \approx \frac{D}{M} \tag{37.2}$$

where D is the distance from the projector crossover to the recording plane and M is the magnification of the image in that plane. So if D is about 500 mm and the screen magnification is 10,000× then L is 0.05 mm. Thus we can show

$$\beta \approx \frac{r_0}{L} \tag{37.3}$$

where r_0 is the maximum radius of the DP in the focal plane of the spectrometer. Typically, r_0 is ~5 μm, and so β is 0.1 rads or 100 mrad which, as we just said, is so large that we rarely need to know it accurately. In fact, in TEM-image mode without an objective aperture, if you just assume β = 100 mrads, any calculation or quantification you do will be effectively independent of β.

If you insert an objective aperture and you know its size and the focal length of the objective lens then β can easily be calculated geometrically. To a first approximation, in a similar manner to equation 37.1, β is the objective-aperture diameter divided by twice the focal length of the objective lens, as shown in Figure 37.7. For example, with a focal length of 3 mm and a 30 μm diameter aperture, β is ~ 5 mrads, which is good if you need high-energy resolution.

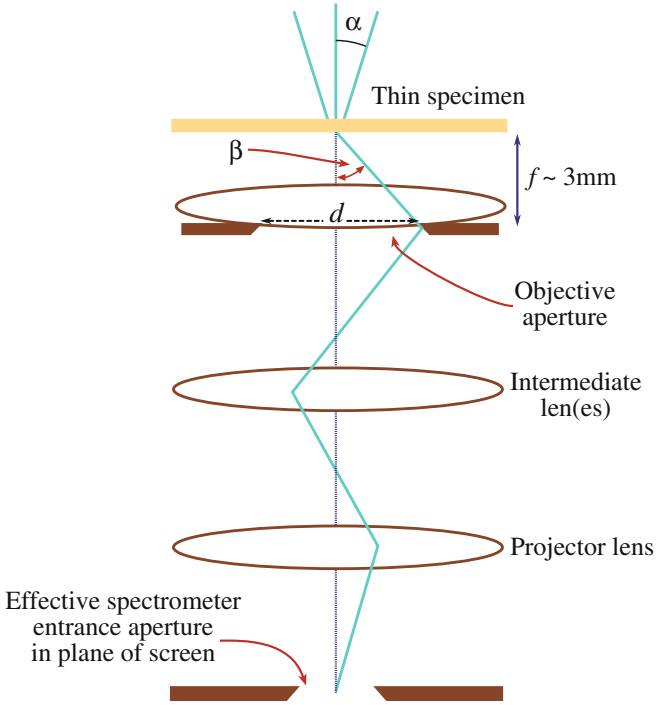


FIGURE 37.7. The value of β in TEM image mode is governed by the dimensions of the objective aperture in the BFP of the objective lens.

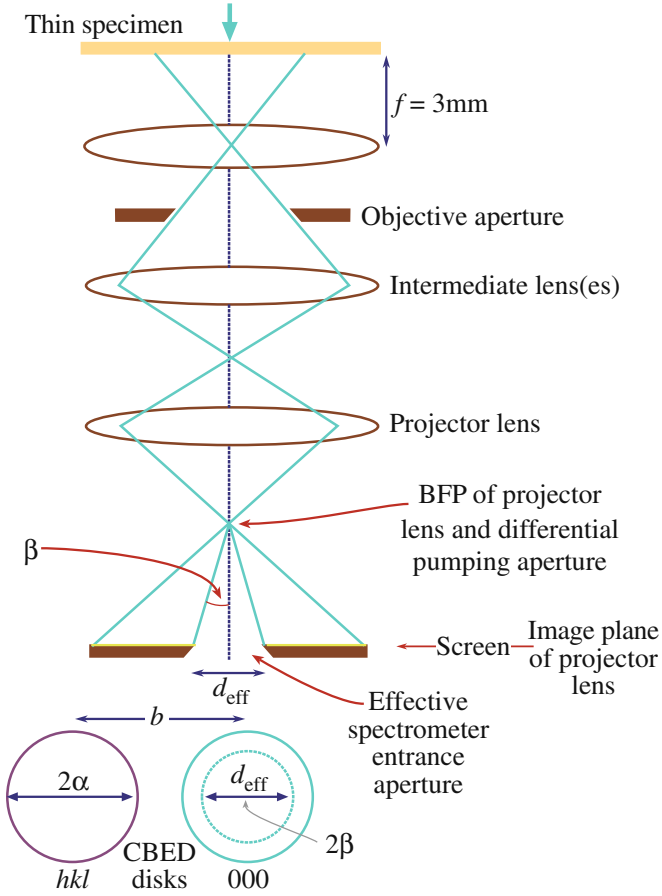


FIGURE 37.8. The value of β in TEM/STEM diffraction mode is determined by the dimensions of the spectrometer entrance aperture, projected into the plane of the DP. The dimensions can be calibrated by reference to a known DP.

If you insert an objective aperture, a normal BF image can be seen on the TEM screen and the information in the spectrum is related to the area of the image that sits directly above the spectrometer-entrance aperture. However, as we'll see in Section 37.4.C, there is some considerable error (~ 100 nm) due to chromatic aberration. We will return to this point in Section 39.10 when we discuss the spatial resolution. Remember also that with the objective aperture inserted, you cannot do XEDS, so simultaneous EELS and XEDS is not possible in this mode.

TEM/STEM diffraction mode. In diffraction (also STEM) mode, the situation is a little more complicated. We focus the spectrometer on an image of the specimen; so we see a DP on the screen and also in the plane of the spectrometer-entrance aperture. Under these circumstances, you control β by your choice of the spectrometer-entrance aperture.

If a small objective aperture is inserted, it is possible that it may limit β; the effective value of β at the back focal plane of the projector lens is β/M where M is the magnification of the image in the back focal plane of the projector lens.

You have to calibrate β from the DP of a known crystalline specimen, as shown in Figure 37.8. Knowing the size of the spectrometer-entrance aperture, the value of β can be calibrated because the distance (b) that separates the 000 spot and a known *hkl* maximum is twice the Bragg angle, $2\theta_B$. If the effective aperture diameter in the recording plane (equivalent to the STEM detector collection angle back in Section 22.6) is d_{eff} and $b = 2\theta$ then

$$\beta = \frac{d_{eff}}{2}\frac{2\theta_B}{b} \tag{37.4}$$

The effective entrance aperture diameter d_{eff} at the recording plane is related to the actual diameter d by

$$d_{eff} = \frac{d\,D}{D_A} \tag{37.5}$$

where D is the distance from the projector crossover to the recording plane and D_A is the distance between the crossover and the entrance aperture. Alternatively, β can be determined directly if the camera length on the recording plane (L) is known since

$$\beta = \frac{D}{D_A}\frac{d}{L} \tag{37.6}$$

D_A is typically 610 mm for Gatan PEELS systems but D varies depending on the TEM; you have control over d and L. For example, if D is 500 mm and L is 800 mm then for a 5 mm diameter aperture, β is ~ 5 mrads.

If we choose a camera length such that the image of the specimen in the back focal plane of the spectrometer is at a magnification of 1×, then, in effect, we have moved the specimen to the focal plane of the spectrometer. This special value of L is equal to D, which you should know for your own microscope. Then β is simply the entrance aperture diameter divided by D_A (610 mm).

In summary, the collection angle is a crucial factor in EELS and the spectrometer/filter software should calculate the values of β in the various operating conditions that you'll encounter.

- Generally, large collection angles will give high intensity but poor energy resolution.
- If you collect your spectrum in image mode without an objective aperture then you won't compromise your energy resolution but spatial resolution is poor, as we'll see below.
- If you're in diffraction mode and you control β with the entrance aperture, then a large aperture (high intensity, high β) will lower the resolution and vice versa.
- Generally, smaller collection angles also give a higher signal-to-noise ratio in the spectrum.

37.4.C Spatial Selection

Depending on whether you're operating in image or diffraction mode, you obtain your spectrum from different regions of the specimen. In TEM-image mode, we position the area to be analyzed on the optic axis, above the entrance aperture. The area selected is a function of the aperture size demagnified back to the plane of the specimen. For example, if the image magnification is 100,000× at the recording plane and the *effective* entrance aperture size at the recording plane is 1 mm, then the area contributing to the spectrum is 10 nm. So, you might think that you can do high spatial resolution analysis without a probe-forming STEM. However, if you're analyzing electrons that have suffered a significant energy loss, they may have come from areas of the specimen well away from the area you selected, because of chromatic aberration. This displacement d is given by

$$d = \theta\ \Delta f \tag{37.7}$$

where θ is the angle of scatter, typically < 10 mrads, and Δf is the defocus error due to chromatic aberration given by

$$\Delta f = C_c \frac{\mathcal{E}}{E_0} \tag{37.8}$$

where C_c is the chromatic-aberration coefficient. So if we take a typical energy loss $\mathcal{E}$ of 284 eV (the energy required to eject a carbon K-shell electron) and we have a beam energy E_0 of 100 keV, then the defocus due to chromatic aberration (with C_c = 3 mm) will be close to 10 μm which gives an actual displacement, d, of 10^{-4} mm or 100 nm. This figure is very large compared to the value of 10 nm, which we calculated without considering chromatic-aberration effects.

In TEM diffraction mode, you select the area of the specimen contributing to the DP in the usual way. You can use either the SAD aperture, which has a lower limit of about 5 μm, or you can form a fine beam as in STEM so that a CBED pattern appears on the screen. In the latter case, the area you select is a function of the beam size and the beam spreading, but is generally < ~50 nm wide. Therefore, this method is best for high spatial resolution EELS, just as for XEDS analysis. But rather than just selecting a single point, it is much better to keep the beam scanning and gather a spectrum at every point (i.e., spectrum imaging; see Section 37.8). This conclusion is true for both in-column and post-column filtering.

THE PRICE

While TEM-image mode might be good for gathering spectra with a large β and high-energy resolution, the price you pay is much poorer spatial resolution, so we don't recommend it.

37.5 PROBLEMS WITH PEELS

There are several standard tests you need to perform to determine that all is well with the PEELS PDA and electronics, just as we described for XEDS back in Chapters 32 and 33.

37.5.A Point-Spread Function

In a PEELS, you can reduce the magnification of your spectrum so that the ZLP occupies a single PDA channel or pixel on the CCD. Any intensity registered outside that channel is an artifact of the system and is called the point-spread function (PSF). This function acts to degrade the inherent resolution of the magnetic spectrometer. The ZLP may spread on its way through the YAG scintillator and the fiber optics before hitting the PDA or CCD. Figure 37.9A shows the PSF of a PEELS and clearly there is intensity well outside a single channel, although a CCD offers considerably better PSF performance than the PDA. The PSF broadens features in your spectrum such as ionization edges, but you need to remove its effect by deconvolution (see Section 39.6) thus restoring the resolution of the spectrum to that inherent in the beam, in the same way we described for X-ray spectra in Section 34.4, except that the commercial software is available and it makes sense to deconvolute the PSF from any spectrum that you gather. The concept is essentially the same as the point-spread function we

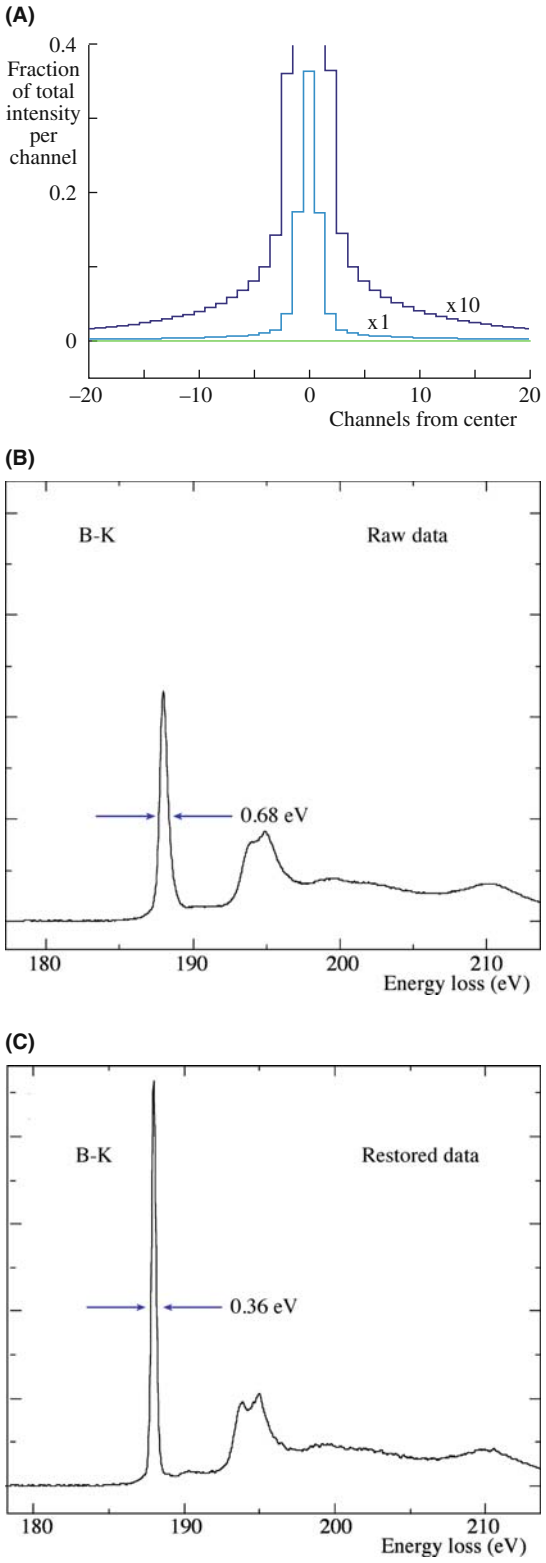


FIGURE 37.9. (A) The point-spread function showing the degradation of the intense well-defined ZLP through spreading of the signal as it is transferred from the scintillator via the fiber optic coupling to the PDA or CCD. The peak should occupy a single channel but is spread across several channels. (B, C) The improvement in energy resolution in the boron K edge from a BN nanotube as a result of ZLP deconvolution. The raw data (B) indicate a resolution of 0.68 eV for the B-K edge and this is improved (C) to 0.36 eV after deconvoluting the point-spread function.

discussed for HRTEM. Figure 37.9B and C demonstrates that such energy deconvolution can produce a significant sharpening of the spectrum and, just as in XEDS, there is no reason not to run a ZLP deconvolution routinely to get the best resolution in your spectrum, so long as you are confident (e.g., by testing the software on spectra from known specimens (see Section 39.6)) that such processing does not introduce its own artifacts; which is a good segue into the next section.

37.5.B PEELS Artifacts

Almost all the artifacts in a PEELS system are a consequence of the PDA, which is why CCD detectors have been introduced. If you have a PDA system, then the individual diodes will differ slightly in their response to the incident electron beam and therefore, there will be a channel-to-channel gain variation in intensity. If you spread the beam uniformly over the array using at least the 3 mm aperture and looking at the diode readouts, you can see any variation, as shown in Figure 37.10. You have to divide your experimental spectrum by this response spectrum to remove any gain variation. Alternatively, and this is recommended, you can gather two spectra with slight energy shifts (~1–2 eV) or spatial shifts between them and superimpose them electronically. Using a 2-D CCD array, as in the GIF, also removes this problem.

If you gather many spectra and superimpose them to avoid saturation of the PDA, you'll generate readout noise. Random readout noise or shot noise from the electronics chain is minimized by taking fewer readouts, and by thermoelectric cooling of the PDA. Individual diodes may fail and give high leakage currents which appear as spikes in the spectrum. The fixed-pattern noise is a function of the three-phase readout circuitry. All these effects will appear when there is no current falling on the diodes and together they constitute the dark current (see Figure 37.11). The dark current is

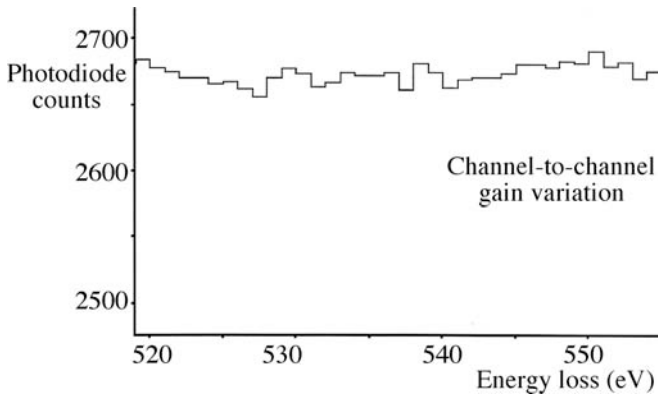


FIGURE 37.10. The variation in the response of individual diodes in the PEELS detection system to a constant incident electron intensity. The channel-to-channel gain variation is clear and each detector array has its own characteristic response function.

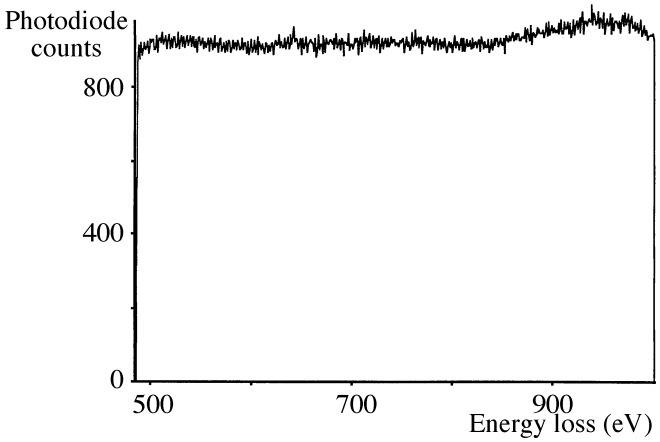


FIGURE 37.11. The intensity of the dark current which flows from the PDA when no electron beam is present.

small unless you have a bad diode and it is only a problem when there are very few counts in your spectrum or you have added together multiple (e.g., 10 or more) spectra. A CCD detector has a higher dark current than a cooled PDA. Figure 37.12 shows some of these effects and how to remove them.

When the diodes are cooled, only ~95% of the signal is read out in the first integration, ~4.5% on the second, ~0.25% on the third, and so on. This incomplete readout can introduce an artifact if you saturated the diodes with an intense signal like the ZLP. This residual peak then shows up as a ghost peak in the next readout and decays slowly over several readouts. So, if a ghost peak appears, just run several readouts and it will disappear; this way you'll never confuse a ghost with a genuine ionization edge or other spectral feature. You can also get a ghost peak on a CCD detector if you overexpose the scintillator.

Increasing the keV means more electrons are generated in the scintillator and the sensitivity should be linearly related to the electron energy. If you're doing quantitative analysis, check that the YAG responds linearly to different intensities by comparing the zero-loss intensity measured in a single 1 s readout with that recorded say in 40 readouts each of 0.025 s. In each case, subtract the dark current. Obviously the ratio of these two intensities should be unity, for all levels of signal falling on the YAG. If you're more intent on low-loss spectrometry or fine-structure studies, this non-linearity is not important.

USING A CCD

If you use a PEELS, GIF, or another filter with a CCD detector, all these artifacts are absent, except the ghost peak and dark current.

We can summarize the PDA artifacts and how we eliminate them in Table 37.1.

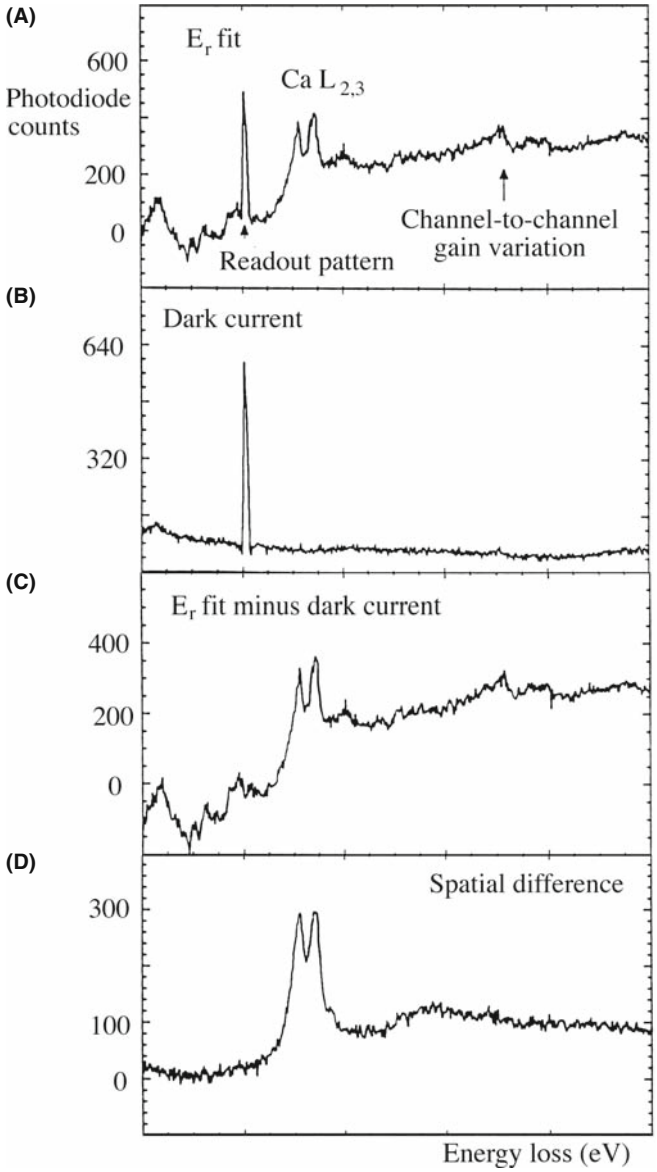


FIGURE 37.12. (A) A Ca $L_{2,3}$ edge spectrum showing both channel to channel gain variation and a faulty diode with a high leakage current which appears as a spike in the spectrum. The spike is referred to as the readout pattern and is present in every recorded spectrum. Subtracting the dark current (shown in (B)) removes the spike (C) and a difference spectrum (D) removes the gain variation, leaving the desired edge spectrum.

37.6 IMAGING FILTERS

Energy-filtered TEM (EFTEM) is sometimes termed energy-filtered imaging (EFI) or electron-spectroscopic imaging (ESI) and is perhaps the most powerful AEM technique, as will become apparent. To perform EFTEM, you basically select (or filter out) electrons of a specific energy coming through the spectrometer and form either an image or a DP. Doing EFTEM is perfectly acceptable in TEM image mode (unlike spectroscopy) so long as you keep the energy slit small to minimize chromatic aberration and, to select the

TABLE 37.1 PEELS Artifacts and How to Eliminate Them

Artifact	Source	Elimination
High leakage current (spike)	Bad diodes	Subtract dark current
Channel to channel gain variation	Different diode responses	Gather spectra on different portions of array and superimpose
Internal scanning noise	Electronics readout	Adjust the electronics and subtract the dark count
Ghost peak	Saturation of diode	Run several readouts
Non-linear response	YAG scintillator is damaged	Zero-loss intensity from different readout numbers with same total integrated time should be identical; if not, replace scintillator

electrons with a specific energy, you change the gun voltage slightly such that those electrons come on-axis and you don't need to keep re-focusing the objective lens.

At the start of the chapter, we noted that there are two types of energy filters: the in-column (Ω) filter and the post-column GIF and both produce EFTEM images. In-column filters are placed in the heart of the TEM imaging system, between the intermediate and projector lenses such that the recording CCD detector only receives electrons that have come through the filter. So all images/DPs consist of electrons of a specific selected energy. (You can, of course, turn off the filter and use the microscope like a normal TEM, but since there are so many advantages to filtered imaging, your reasons for doing this would be questioned.) You can also record a spectrum as we'll show, but the in-column filters are primarily imaging tools and it is both feasible and desirable to operate all the time in filtered mode.

The post-column GIF is added below the TEM viewing screen, just like a PEELS and therefore you can choose to use it or not. The GIF can be seen as either a more flexible instrument or one that limits you to having to decide whether or not to filter your images. Such differences, while perhaps pedantic, have been known to lead to fights in bars at M&M conferences. The best solution, of course, is to have one of each type of filter in your laboratory, but this takes some of the fun out of going to bars at M&M. Let's examine each type of filter in more detail.

37.6.A The Omega Filter

Zeiss first used a mirror-prism system originally devised by Castaing and Henry in 1962. The drawback to the mirror-prism is the need to split the high-voltage supply and raise the mirror to the same voltage as the gun. So Zeiss now uses a magnetic Ω filter, as does JEOL, currently the only other TEM manufacturer with an in-column filter. The filter is placed in the TEM column between the intermediate and projector lenses and consists of a set of magnetic prisms arranged in an Ω shape which disperses the electrons off axis, as shown in Figure 37.13A but, in the end, brings them back onto the optic axis before entering the final projector lens. Although it is not shown, each of the four prisms is machined with curved faces (like Figure 37.2B) to reduce aberrations. There isn't much to see externally on the TEM except an asymmetry on the side of the (somewhat taller) column, which you can also see if you go back and look at the picture of the Zeiss UHRTEM in Figure 1.9.

The multiple steps necessary for ETEM imaging via an in-column filter are shown in Figure 37.13B. As shown in this figure, we usually project an image into the prism which is focused on a DP in the back-focal plane of the intermediate lens (i.e., image mode in our terminology). Therefore, the entrance aperture to the spectrometer selects an area of the specimen and β is governed by the objective aperture. Electrons following a particular path through the spectrometer (the red ones in Figure 37.13B) can be selected by the post-spectrometer slit. Thus, only electrons of a given energy range, determined by the slit width, are used to form the image projected onto the TEM CCD. EFTEM has several advantages over conventional TEM images; there's much more about EFTEM in the companion text.

If you change the focus of the projector lens onto the dispersion plane of the filter (where the energy-selecting slit is located in Figure 37.13B) and then remove this slit, you'll see a spectrum on the viewing screen/CCD. The spectrum appears as a line of varying intensity (see Figure 37.14A), which you can imagine as looking down from above on a traditional spectrum display such as Figure 37.1. Since the spectrum is recorded digitally on the CCD, it is simple to select a line through this spectrum and have the computer display a traditional spectrum of counts versus energy loss, as shown in Figure 37.14B.

As with PEELS, you can also change the TEM optics and project a DP into the prism, thus producing an energy-filtered DP on the CCD, as we've already described back in Chapter 20. If you then use the slit to select a portion of the DP, you get an energy-loss spectrum showing not only the intensity distribution as a function of energy but also the angular distribution of the electrons. Such angular (or momentum)-resolved

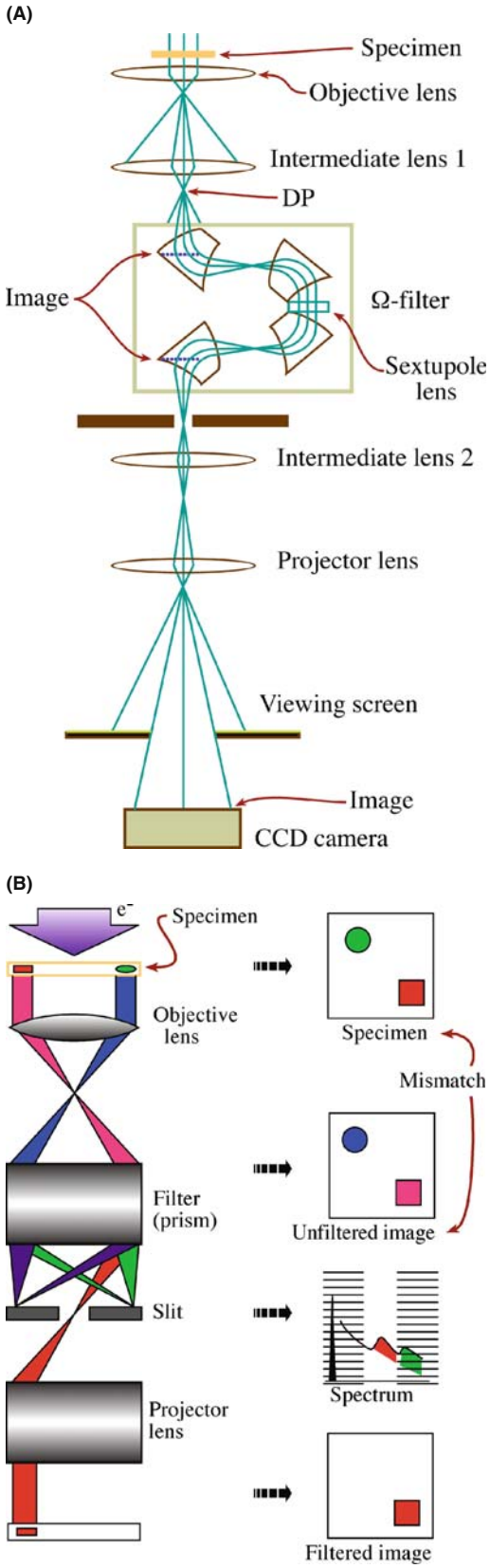


FIGURE 37.13. (A) Schematic diagram of an in-column Ω filter inserted in the imaging lens system of a TEM. (B) Schematic ray diagram of the steps needed to create an EFTEM image.

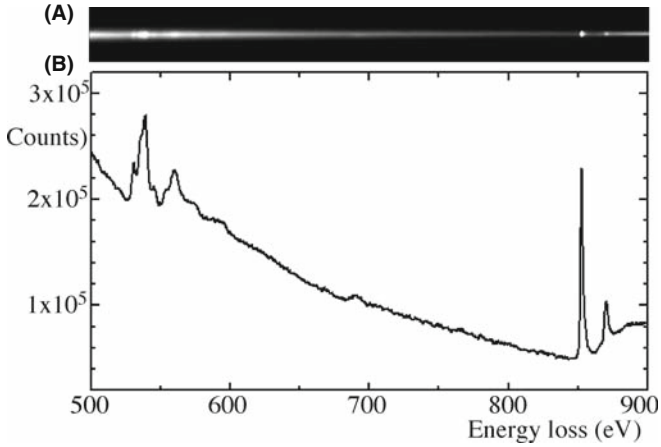


FIGURE 37.14. (A) Spectrum from an Ω filter. The axis of the line is energy loss and the intensity varies along the line. (B) Conventional spectrum of intensity versus energy-loss obtained from (A).

EELS is a whole separate field of study which we'll discuss in Section 40.8.

37.6.B The GIF

The GIF is a Gatan PEELS with an energy-selecting slit after the magnet and a 2D slow-scan CCD detector, rather than a 1D PDA, as the detector. Figure 37.15A shows a schematic diagram of a GIF interfaced to a TEM and Figure 37.15B shows an exploded view of a GIF (which shows the internal parts of the spectrometer shown back in Figure 37.3). Compared to a standard spectrometer, there are many more quadrupoles and sextupoles in the optics of the GIF. The dispersion of the spectrometer onto the slit has to be magnified and the quadrupoles after the slit have two functions. Either we project an image of the spectrum in the plane of the selecting slit onto the CCD or we compensate for the energy dispersion of the magnet and project a magnified image of the specimen onto the CCD. In the first mode, the system is operating like a standard PEELS; in the second, it produces images (or DPs) containing electrons of a specific energy selected by the slit. While the spectrometer is double focusing as we saw, the aberration correction is good only for a single plane so astigmatism is introduced and this has to be corrected by a further combination of sextupoles and octupoles. Obviously, such a large number of variable sextupoles and quadrupoles could be a nightmare to operate without appropriate computer control and this is built into the system software. One potential operational difficulty is that the magnification of the GIF system is such that the actual TEM screen magnification needs to be rather small in order to observe a filtered image with reasonable magnification. More recent AEMs satisfactorily compensate for this magnification differential, but others do not and

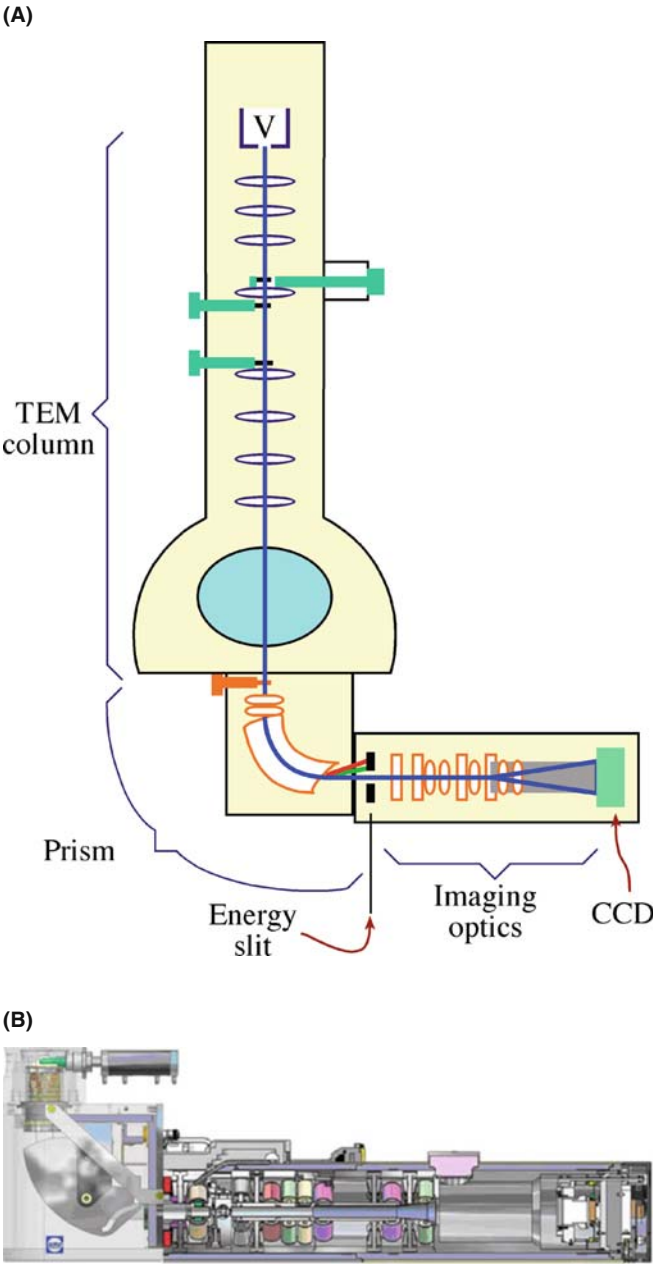


FIGURE 37.15. (A) Schematic diagram of how a post-column imaging filter is attached to the TEM column below the viewing chamber in the same position as a PEELS. (B) Cross section showing the complex inner workings of the Gatan (Tridiem) imaging filter (GIF).

then you have to move between TEM and GIF images and it is not so easy to operate in filtered mode all the time, which would be preferred.

37.7 MONOCHROMATORS

Energy resolution is obviously a key factor in EELS, much more so in fact than in XEDS where we still live with the miserable resolution of the solid-state detector. Resolution of a few eV is more than sufficient for ionization-loss spectrometry, but as we use EELS more to study such aspects as vibrational modes of atoms, inter/intra band excitations, fine structure, and electronic effects, sub-eV resolution is increasingly of interest. For example, Batson at IBM has pioneered high-resolution (HREELS) in a decades-long research project probing ever deeper into the electronic structure of Si and the Si/SiO_2 interface and other materials essential for modern electronics.

If your spectrometer is not a limiting factor, then the gun dictates the ultimate energy resolution. Thermionic sources with ~3 eV (W) and 1.5 eV (LaB_6) resolution, respectively, might be tolerable for ionization-loss spectrometry and basic low-energy plasmon studies. However, EELS really requires that you have a FEG. Either a cold FEG or a Schottky will give you sub-eV resolution, but this is still not sufficient in some cases. So commercial monochromators have become available which offer resolutions ~100 meV although at very low energy losses (<50 eV) it is even possible to get resolutions of ~25 meV. Basically, a monochromator is an EELS system fitted on the FEG source and the selecting slit refines the already narrow energy spread still further giving remarkably fine detail in the spectra. The monochromator is usually a Wien filter which has perpendicular electrostatic and magnetic fields which permit the chosen electrons to travel in a straight line down the TEM column. (Again, if you want to learn more, go back to Metherell's early review which has it all.) Figure 37.16A shows the ZLP at 200 keV obtained with and without a monochromator showing the reduction in FWHM. From Figure 37.16A, you can see that the FWHM of a commercial monochromated ZLP is much better than the 600 meV you can get with a Schottky gun and a little better than the ~ 300 meV delivered by a good cold FEG. The intensity of the beam tail is relatively low in comparison with that of a standard cold FEG. This low-intensity tail results in real improvements for low-loss spectrometry, as we'll see in the next chapter.

Figure 37.16B compares the quality of different spectra of the Co $L_{2,3}$ edges obtained on a variety of instruments and calculated (see Chapter 40 and the companion text for more details on the calculation). Clearly the degree of fine detail in the spectra improves as the energy range of the electron sources gets smaller. The monochromated FEG is similar in quality to the spectrum from the synchrotron source, if a little noisier. The degree of agreement between these latter two spectra and the calculated spectrum is encouraging.

The drawback to monochromation, of course, is that, in filtering out the tails of the Gaussian energy distribution to reduce the FWHM, we reduce the number of electrons significantly. So having spent serious dollars to get the brightest possible electron source for our AEM, we immediately throw away a large amount of it in pursuit of the best energy resolution. This compromise is a real limiting factor and you should rarely contemplate acquiring an AEM with a monochromator.

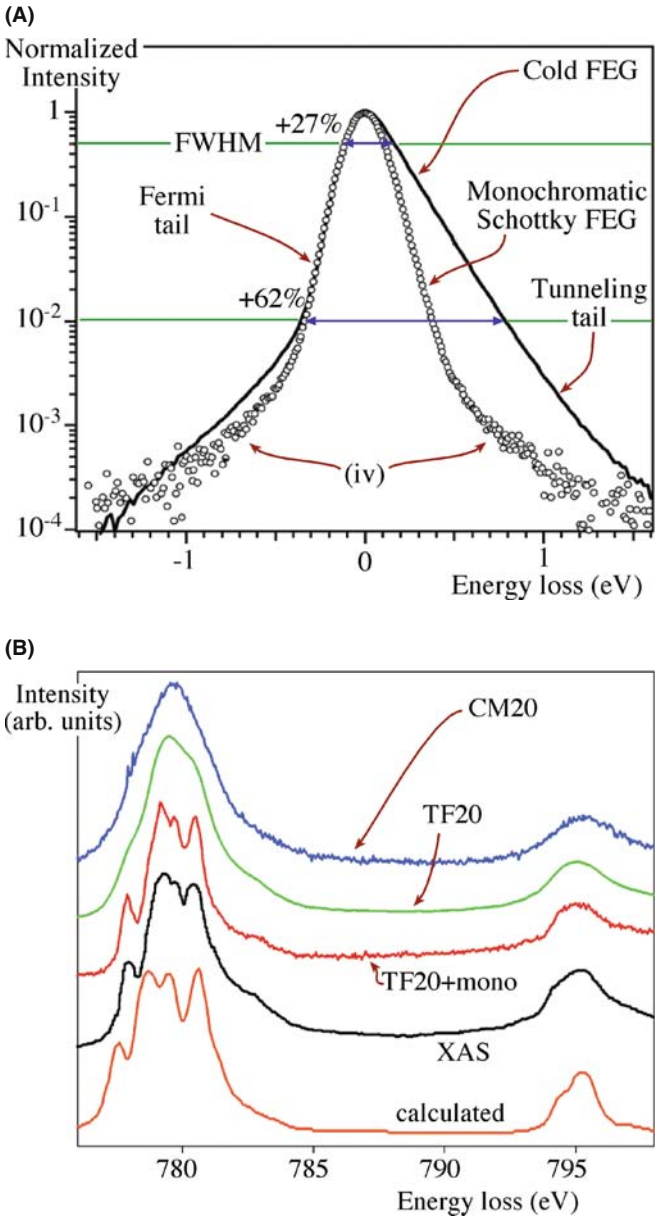


FIGURE 37.16. (A) Typical cold FEG ZLP with and without monochromation. Note that the ordinate is logarithmic so the FWHM is close to the top of the peak. (B) Comparison of Co $L_{2,3}$ spectra from: a Philips CM20 thermionic source, an FEI Tecnai TF20 with a cold FEG, a TF20 with a monochromator, a synchrotron (X-ray absorption spectrum), and a calculated spectrum using crystal-field theory. The improvement in resolution with monochromation is apparent.

WAIT BEFORE YOU MONOCHROMATE

If HREELS is the *raison d'etre* of the instrument remember that just about all other TEM/AEM operations will be degraded if you switch on the monochromator.

The one thing that you get in EELS is lots of electrons, particularly in the low-loss spectrum. In this case, monochromation is truly an advantage. Of course, it would also help if we could develop even brighter sources than cold FEGs.

There are also software alternatives to physical monochromation of the source and by various deconvolution routines it is possible to obtain spectra with a resolution of about 200 meV (e.g., the papers by Kimoto et al. and Gloter et al.). This approach is a very acceptable alternative to using a monochromator, unless you need the absolute best resolution for exploring the finest details of the electronic structure of nanomaterials (e.g., papers by Kimoto et al. in 2005, and Spence in 2006) when you need to lower both the keV and the beam current to produce sub-100 meV resolution.

37.8 USING YOUR SPECTROMETER AND FILTER

So now you've got a thin specimen in your microscope (you'll see later that for most EELS experiments, *very* thin is much better than merely thin) and you want to start acquiring spectra and filtered images. There are several ways to do this and basically they parallel what we've already described for XEDS

1. *Point analyses*: stop the STEM probe from scanning and position it on a selected point in the image over the entrance aperture and record a spectrum (see Figure 37.17A). This was the standard modus operandi for decades, but is (a) biased (because you pick what *you* think ought to be analyzed), (b) statistically poor (only one pixel selected out of a million or so in a typical STEM image), and (c) prone to damaging/contaminating the chosen area of interest by leaving the intense probe on the same point for long periods of time. So don't do point analyses, unless you're just looking quickly around the specimen to see what you can find. If you do, don't necessarily believe that what you find is significant!
2. *Line analyses*: take a series of spectra along a line that traverses some feature of interest (such as a planar interface or defect such as a grain boundary). This process is still biased by your choice of line but at least has the advantage of focusing on something that should provide useful information about the specimen. You can either plot the information from the spectra (e.g., the composition or the dielectric constant or whatever data you extract) or you can display the data as a spectrum-line, as shown in Figure 37.17B which shows changes in chemistry across a nanotube.
3. *TEM filtered images*: using a GIF or Ω filter, gather a filtered image or DP by using the slit to select electrons of a specific energy thus allowing only those electrons to fall on the viewing screen or CCD. If you use this method to filter out all the energy-loss electrons and produce a ZLP image, you can

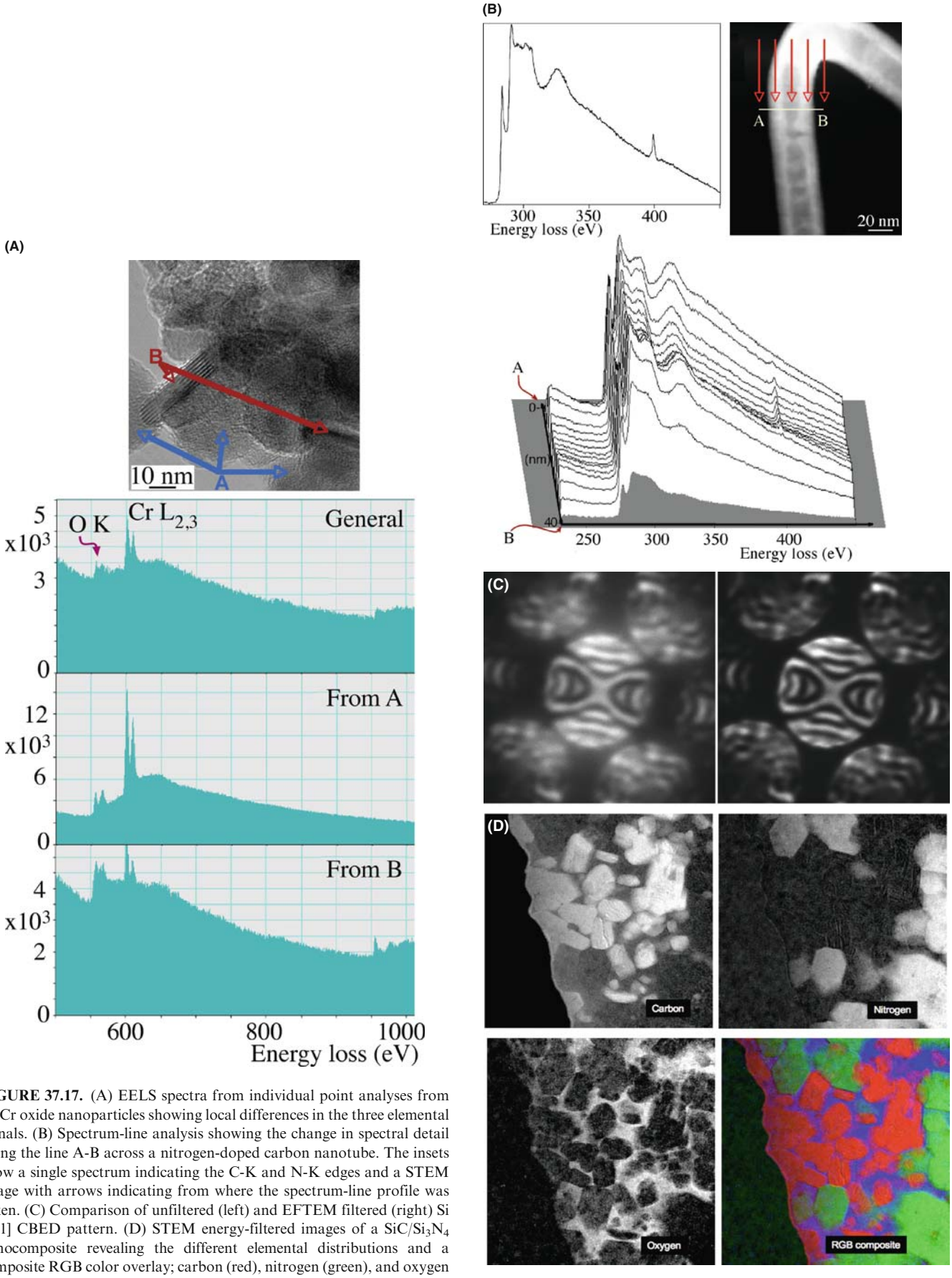


FIGURE 37.17. (A) EELS spectra from individual point analyses from CuCr oxide nanoparticles showing local differences in the three elemental signals. (B) Spectrum-line analysis showing the change in spectral detail along the line A-B across a nitrogen-doped carbon nanotube. The insets show a single spectrum indicating the C-K and N-K edges and a STEM image with arrows indicating from where the spectrum-line profile was taken. (C) Comparison of unfiltered (left) and EFTEM filtered (right) Si [111] CBED pattern. (D) STEM energy-filtered images of a SiC/Si_3N_4 nanocomposite revealing the different elemental distributions and a composite RGB color overlay; carbon (red), nitrogen (green), and oxygen (blue).

FIGURE 37.17. (Continued).

substantially enhance the quality of CBED patterns, as shown in Figure 37.17C (and also back in Figure 20.10). Also, as we'll see at various times, EFTEM improves mass-thickness contrast, phase contrast, and diffraction contrast. So why wouldn't you use this technique if you have a filter on your AEM?

4. *STEM filtered images*: scan the beam over an area of interest using the slit to select electrons of a specific energy thus allowing only those electrons to fall on the viewing screen or CCD. If you select electrons of a specific energy then you can produce composition maps (Figure 37.17D) similar to many of the XEDS composition maps we have already shown.
5. *Spectrum-images*: store a full spectrum at each pixel (obviously, you need to be operating in STEM to do this). It's much easier to do this in STEM because if you try and do this in TEM mode, you'll have to record hundreds of images at different selected energies and comparing complete images is challenging and time consuming, so not everyone does this; however, this situation is changing. In a similar manner to what we showed you for XEDS, after you've acquired a full spectrum image, you can go into the data cube and select whatever information you want. For example, you can view images at specific energies or spectra from specific points or lines in your specimen, thus removing the bias that affects your choice of spot or line analyses. EELS spectrum imaging was more easily implemented than the XEDS version because of the ease of acquiring large numbers of counts and the broad array of possible images was demonstrated by Hunt and Williams.

We'll use examples of these various methods throughout the next three chapters. Generally speaking, as we've emphasized for X-rays, it makes sense for you to form images of specific features in the spectrum rather than gather spectra from specific points, so that they can at least be compared with your TEM images. Since counts are not a problem in EELS (unlike for XEDS) imaging is relatively straightforward and fast. If you really want to optimize your information then acquire spectrum images.

CHAPTER SUMMARY

We use a magnetic-prism spectrometer for PEELS and a combination of one or more prisms plus imaging lenses for GIF/EFTEM. The prism is a simple device and very sensitive but you have to understand how it functions in combination with different TEM modes, You can operate with your TEM in imaging or diffraction (STEM) mode, or use a DSTEM. You have to know how the software focuses and calibrates the system and how it determines the collection angle, β. Once you can understand this you're in a position to acquire and analyze EEL spectra. So in the next chapter we'll tell you what these spectra look like and what information they contain. If you have an Ω filter or GIF you can routinely form images or DPs with electrons of specific E and it makes sense, if you can, to view all your images and DPs in filtered mode for reasons we've already mentioned in past chapters and that we'll reiterate in the subsequent chapters. Energy filtering is a rapidly evolving field and systems with reduced aberrations and better energy resolution such as the Mandoline filter on Zeiss's Sub-eV-Sub-Ångstrom Microscope (SESAMe) and Nion's UltraSTEM will offer far more exciting breakthroughs than we can describe here and Egerton gave a brief synopsis of a range of hardware and software advances in 2003.

THE INSTRUMENT

Brydson, R 2001 *Electron Energy-Loss Spectroscopy* 2001 Bios (Royal Microsc. Soc.) Oxford UK. Good introductory text, similar in content and level to much of this chapter and the subsequent ones.

Castaing, R. and Henry, L 1962 *Filtrage Magnétique des Vitesses en Microscopie Electronique* C. R. Acad. Sci. Paris **B255** 76–78. The original design for the mirror prism.

Egerton, RF 2003 *New Techniques in Electron Energy-Loss Spectroscopy* Micron **34** 127–139 A concise review of recent advances in instrumentation.

Egerton, RF, Yang, YY and Cheng, SY 1993 *Characterization and Use of the Gatan 666 Parallel-Recording Electron Energy-Loss Spectrometer* Ultramicrosc. **48** 239–250.

Hillier, J. and Baker, RF 1944 *Microanalysis by Means of Electrons* J. Appl. Phys. **15** 663–675. History – it's amazing to read that AEM was essentially all conceived more than 60 years ago.

Metherell, AJF 1971 *Energy Analysing and Energy Selecting Electron Microscopes* Adv. Opt. Elect. Microsc. **4** 263–361 Eds. R Barer and VE Cosslett Academic Press New York. Still the best review of spectrometers.

Uhlemann, S and Rose, H 1996 *Acceptance of Imaging Energy Filters* Ultramicrosc. **63** 161–167.

BACKGROUND DATA

Ahn, CC Ed. 2004 *Transmission Electron Energy Loss Spectrometry in Materials Science* 2nd Ed. Wiley-VCH Weinheim Germany. Updated version of the Disko et al. text (below).

Disko, MM, Ahn, CC and Fultz, B Eds. 1992 *Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS Atlas* TMS Warrendale PA. Multi-author, practical text.

Egerton, RF 1986 *Electron Energy-Loss Spectroscopy in the Electron* Plenum Press New York. The bible of EELS; required reading for all serious spectroscopists.

Egerton, RF 1996 *Electron Energy-Loss Spectroscopy in the Electron Microscope* 2nd Ed. Plenum Press New York. Second edition of the EELS bible.

APPLICATIONS

Batson PE 2004 *Electron Energy Loss Studies of Semiconductors* in *Transmission Electron Energy Loss Spectrometry in Materials Science* 2nd Ed. 353–384 Ed. CC Ahn Wiley-VCH Weinheim Germany.

Gloter, A, Douiri, A, Tencé, M and Colliex, C 2003 *Improving Energy Resolution of EELS Spectra: an Alternative to the Monochromator Solution* Ultramicrosc. **96** 385–400.

Hunt, JA and Williams, DB *Electron Energy-Loss Spectrum Imaging* Ultramicrosc. **38** 47–73.

Kimoto, K, Kothleitner, G, Grogger, W, Masui, Y and Hofer, F 2005 *Advantages of a Monochromator for Bandgap Measurements Using Electron Energy-loss Spectroscopy* Micron **36** 185–189.

Spence, JCH 2006 *Absorption Spectroscopy with Sub-Angstrom Beams: ELS in STEM* Rep. Prog. Phys. **69** 725–758.

THE EELS WORKSHOP REPORTS

The quadrennial EELS workshops (1990, 1994, 1998, 2002, and 2006) pioneered and often organized and edited by Krivanek are a rich source of papers by the leading researchers in the field describing cutting-edge aspects of EELS and related techniques. Often the proceedings are published separately and are divided between methodology/instrumentation and practice. The respective journals (either complete volumes or part thereof) and editors are listed below.

Krivanek, OL Ed. 1991 Microsc. Microanal. Microstruct. **2** (# 2–3).
Krivanek, OL Ed. 1995a Microsc. Microanal. Microstruct. **6** 1.
Krivanek, OL Ed. 1995b Ultramicrosc. **59** (# 1–4).
Krivanek, OL Ed. 1999 Ultramicrosc. **78** (# 1–4).
Krivanek, OL Ed. 1999 Micron **30** (#2) 101.
Krivanek, OL Ed. 2003 Ultramicrosc. **96** (#2–4) 229.
Krivanek, OL Ed. 2003 J. Microsc. **210** 1.
Browning, ND and Midgley, P Eds. 2006 Ultramicrosc. **106** (#11–12).
Mayer, J Ed. 2006 Micron **37** 375.

URLs

1) http://www.energyloss.com/index.html
2) http://www.gatan.com/ Gatan's web site

SELF-ASSESSMENT QUESTIONS

Q37.1 What's the difference between a GIF and a PEELS system?
Q37.2 Why is it useful for the spectrometer to act as a lens also? Can we achieve the same combination when dispersing visible light?
Q37.3 What is the dispersion plane of the spectrometer and what is a typical value of the dispersion?
Q37.4 How would you measure the energy resolution of the spectrometer?
Q37.5 What's a typical value of the energy resolution and what governs the minimum possible value?
Q37.6 What factors can degrade (i.e., increase the value of) the energy resolution?
Q37.7 Why is the spectrometer collection angle (β) so important?
Q37.8 What controls β in image mode?
Q37.9 What controls β in diffraction mode?
Q37.10 Why might you have to integrate several spectra in your PEELS rather than just acquire a single spectrum?
Q37.11 What controls the spatial resolution in image mode?
Q37.12 What controls the spatial resolution in diffraction mode?
Q37.13 Why is diffraction mode in TEM equivalent to operating a dedicated STEM?
Q37.14 What is the point-spread function and why should you be concerned about it?
Q37.15 How do you correct for this artifact?
Q37.16 Why do you need to calibrate the spectrometer?
Q37.17 How do you calibrate the spectrometer (if the software can't do it for you)?
Q37.18 Why is it necessary to cool the diode array in a PEELS?

Q37.19 What's the difference between a GIF and an Ω filter?
Q37.20 What are the pros and cons of these two types of filter?
Q37.21 Why would you want to form an image from specific electrons in an EELS spectrum rather than just look at the spectrum?
Q37.22 Why is TEM image mode generally a bad choice for gathering spectra?

TEXT-SPECIFIC QUESTIONS

T37.1 Using Figure 37.1 and Figure 32.2A, contrast the principal characteristics of XEDS and EELS spectra, indicating the relevance of your observations to interpreting/quantifying the spectra.
T37.2 Examine Figure 37.2B. Why is the back focal plane of the projector lens used as the object plane of the spectrometer/lens?
T37.3 What is the object of the spectrometer/lens when an image is on the TEM screen?
T37.4 Electrons that have lost energy can be both an advantage and a disadvantage. Explain how such electrons are used to advantage in diffraction and in imaging, and likewise how they can degrade DPs and images (find examples (especially figures where possible) throughout the book to support your arguments).
T37.5 Using Figure 37.9 estimate what fraction of the single channel intensity is lost through the spreading of the point-spread function. What instrument modification might help reduce the point-spread function by keeping the probe more localized?
T37.6 Contrast the specific roles of the objective aperture and the selected area aperture in TEM and EELS.
T37.7 Describe the steps you would take to gather a spectrum with $\beta = 20$ mrad in (a) DSTEM, (b) TEM image mode, or (c) TEM diffraction mode.
T37.8 Can you think of circumstances in which you might get better energy resolution in a non-monochromated spectrum than in a monochromated one?
T37.9 Examine Figure 37.1 and estimate the absolute intensities of the zero-loss peak, the plasmon peak and the Ni L ionization edge above background. What does this tell you about the difficulties that we will encounter with processing EELS spectra? How does the P/B ratio compare with an XEDS spectrum?
T37.10 Although SEELS systems are no longer sold, can you think of an advantage to serial collection over parallel collection?
T37.11 Why is the camera length an important variable if you are operating in diffraction mode (Figure 37.8)? Choose two reasonable values of the camera length and calculate what effect this has on the operation of the spectrometer.
T37.12 Explain under what circumstances the spectrometer entrance aperture, rather than the objective aperture, might control the value of β in TEM mode (Figure 37.7).
T37.13 If the specimen is really thin so that the electrons suffer no significant energy losses while traversing the thin foil, can we totally ignore chromatic-aberration effects?
T37.14 Explain why the specimen you are looking at can affect the spatial resolution of the EELS analysis (in totally different ways to which the specimen controls the spatial resolution for XEDS).
T37.15 If your lab can't afford a monochromator, list all the other ways that you can improve the resolution of your spectra. Give the pros and cons of each method and list them according to their relative costs.
T37.16 What is a typical count rate in an EELS spectrum versus an XEDS spectrum? (If you can't find these data in the book, try a quick experiment on the AEM.)
T37.17 Would you expect a channel-to-channel gain variation in a scintillator-photomultiplier detector on an old SEELS and in a CCD camera in a GIF?
T37.18 How can you minimize the dark current?
T37.19 List the other common artifact in EELS spectra and explain how you would (a) recognize and (b) correct for each one.
T37.20 Why is it not a problem if one of the detectors in the PEELS diode array dies?
T37.21 When might it be better to perform a spectrum–profile analysis rather than an EFTEM analysis?
T37.22 Can you think of any circumstance when point analysis might be better than a line-profile or EFTEM analysis?

38

Low-Loss and No-Loss Spectra and Images

CHAPTER PREVIEW

The term 'energy loss' implies that we are interested only in inelastic interactions, but the spectrum will also contain electrons which have not lost any discernible energy, so we need to consider elastic scattering as well. In this chapter, we'll focus on the low-energy portion of the EEL spectrum which comprises

- The zero-loss peak, which primarily contains elastic, forward-scattered electrons, but also includes electrons that have suffered small energy losses. Forming images and DPs with the zero-loss electrons offers tremendous advantages over unfiltered images, particularly from thicker specimens.
- The low-loss region up to an (arbitrary) energy loss of $\sim$ 50 eV contains electrons which have interacted with the weakly bound, outer-shell electrons of the atoms. Thus, this part of the spectrum reflects the dielectric response of the specimen to high-energy electrons. We can also form images from these low-loss electrons that reveal information about the electronic structure and other characteristics of the specimen.

The energy-loss spectrum is more useful than an X-ray spectrum which only contains elemental information. However, this kind of spectrum is also far more complex. To understand its content you need a greater understanding of the physics of beam-specimen interactions, so we'll give you some hints about where to get the necessary education.

38.1 A FEW BASIC CONCEPTS

Back in Chapters 2–4, we described the difference between elastic and inelastic beam-specimen interactions and introduced the ideas of scattering cross sections and the associated mean free path. Remember, the cross section is a measure of the probability of a specific scattering event occurring and the mean free path is the average distance between particular interactions. It might be a good idea to re-read about those ideas before starting on this chapter. Briefly, you should recall that elastic scattering is an electron-nucleus interaction; the word elastic implies that there is no energy loss, although a change in direction, and hence in momentum, usually occurs. Elastic scattering occurs mainly as Bragg diffraction in crystalline specimens. Inelastic scattering is primarily an electron-electron interaction and entails both a loss of energy and a change of momentum.

We have to be concerned with both the amount of energy lost and the direction of the electrons after they've come through the specimen.

This latter point is one reason why the collection angle of the spectrometer, β, is so important.

Also you must remember to distinguish between the definitions of scattering that will keep appearing.

- *Single scattering* occurs when each electron undergoes at most one scattering event as it traverses the specimen.
- *Plural (>1* and *< 20) scattering* implies that the electron has undergone a combination of interactions and lost energy from some or all of them.
- *Multiple (> 20) scattering* only occurs in very thick specimens or with very low energy electrons, so is irrelevant to TEM.

We'll see that the energy-loss spectrum is most understandable and more easily modeled when we can approximate everything to single scattering. This ideal is approached when we have a combination of very thin specimens and high accelerating voltages. In practice, most specimens are thicker than ideal and so we usually acquire plural-scattering spectra and we may have to remove the plural-scattering effects via deconvolution routines which are available in commercial and free EELS software packages (see URLs #1–3). We'll address the general topic of spectral simulation and manipulation later and in more depth in the companion text. We've already told you about deconvoluting out the PSF in Chapter 37, we'll see in Section 38.2.B, that you can do similar things if you want to remove the ZLP, and removing plural-scattering effects in high-loss spectra will be discussed in Section 39.6.

DECONVOLUTION WARNING

Whenever we mention deconvolution (and we'll do so quite often), remember that there is the danger of introducing artifacts into the part of the spectrum that you have just tried to simplify.

Typical energy losses: The principal inelastic interactions in order of increasing energy are phonon excitations, interband and intraband transitions, plasmon excitations, and inner-shell ionizations. We've already introduced these processes back in Chapter 4. The energy loss $\mathcal{E}$ of the principal scattering processes that we study in EELS are single-electron scattering (inter/intraband transitions), 2–20 eV, plasmon interactions 5–30 eV, and inner-shell ionizations, 50–2000 eV. Phonon excitations cause losses of $\sim$ 0.02 eV so, even with the best energy resolution in monochromated AEMs, it's not possible to separate these from the ZLP, although the phonon-scattering angle can be quite large and (particularly for heavier elements) these electrons can be seen as the background intensity between the principal spots in an SAD pattern. But that's about all we'll say about phonon scattering, except that you should be able to work out why cooling your specimen will help to reduce their presence. We'll deal with the first three (low-loss) processes in this chapter and the ionization (high-loss) process in the next chapter. There are also ionization events with energy losses >2 keV but it is difficult to detect these because the signal is relatively weak and, at such energies, the X-ray signal is strong, so we tend not to do much EELS > 2 keV.

REMEMBER θ

The symbol θ in all cases refers to the scattering *semi*-angle even if we just say angle.

THE MOST IMPORTANT ANGLE IS θ_E

The so-called characteristic or most-probable scattering angle (for an energy loss $\mathcal{E}$)—it depends on the beam energy.

Typical scattering angles: It's a little difficult to be specific about the values of scattering angle, θ, because the angle varies with beam energy and the energy-loss process. We always assume that the scattering is symmetrical around the direct beam and there are two principal scattering angles that you should know. You can find derivations of the equations in Egerton's text.

$$\theta_E \approx \frac{\mathcal{E}}{2E_0} \tag{38.1}$$

If the beam energy is in eV the angle is in radians. This equation is an approximation (good to about ±10%) and it ignores relativistic effects and doesn't work for phonons, so you should only use it for rough calculations and be particularly suspicious above ~100 keV. We can be more precise and define θ_E as

$$\theta_E \approx \frac{\mathcal{E}}{(\gamma m_0 v^2)} \tag{38.2}$$

Here we have the usual definitions: m_0 is the rest mass of the electron, v is the electron velocity and γ is given by the usual relativistic equation (where c is the velocity of light)

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \tag{38.3}$$

The other useful angle θ_c is the cut-off angle above which the scattered intensity is zero

$$\theta_c = (2\theta_E)^{\frac{1}{2}} \tag{38.4}$$

Be careful to calculate θ_c in radians, not milliradians. Knowing the characteristic scattering angle is obviously important if you want to gather an intense spectrum highlighting a particular energy loss. For example, a plasmon interaction with 100-keV electrons, causing a typical energy loss of 20 eV will have a characteristic scattering angle of ~0.1 mrads. Using a smaller β will cut off intensity in the spectrum which is why we told you in the last chapter to ensure that $\beta > 2–3\theta_E$. Knowing the cut-off angle (which is typically an order of magnitude greater than the characteristic angle at 100 keV) will give you a maximum useful value of β. If you use too large a value of β then there's the chance that you'll get unwanted electrons in the spectrum

(e.g., diffracted beams) but you've got to try really hard to encounter this problem.

38.2 THE ZERO-LOSS PEAK (ZLP)

38.2.A Why the ZLP Really Isn't

If your specimen is thin enough for EELS, the predominant feature in the spectrum will be the ZLP, as shown in Figure 38.1. As the name implies, the ZLP consists primarily of electrons that have retained the incident-beam energy E_0. Such electrons may be forward scattered in a relatively narrow cone within a few mrads of the optic axis and constitute the 000 spot in the DP, i.e., the direct beam. If you were to tilt the incident beam so a diffracted beam entered the spectrometer then it, too, would give a ZLP.

MAGNITUDE OF ANGLES

The scattering angles for diffraction ($2\theta_B$) are relatively large (~20 mrad) compared to the smaller scattering angles in EELS. So the diffracted beams will only enter the spectrometer if you select them.

Actually, we can also measure the intensity and energy of electrons as a function of their angular distribution, and we'll discuss angular- or momentum-resolved EELS later, in Section 40.8.

Now the term ZLP is really a misnomer for two reasons. First, as we've seen, the spectrometer has a finite energy resolution (at best ~ 0.3 eV without monochromation) so the ZLP will also contain electrons that have energy losses below the resolution limit, which are mainly those that excited phonons. This is not a great loss since phonon-scattered electrons don't carry useful information; they only cause the specimen to heat up. However, it does explain why we shouldn't really call this the ZLP. Second, we can't produce a monochromatic (single-color, i.e., single-wavelength/energy) beam of electrons; the beam has a finite energy range about the nominal value E_0 (at best 10–100 meV even with a monochromator). Despite this imprecision, we will continue to use the term zero loss.

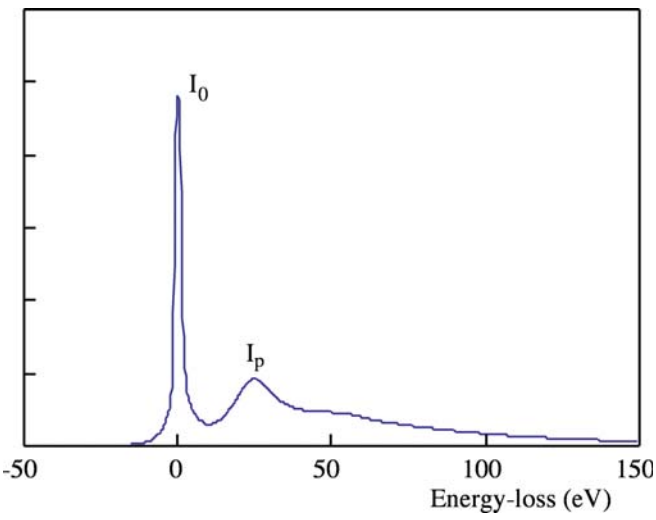


FIGURE 38.1. The low-loss spectrum showing an intense ZLP. The next most intense peak is a plasmon peak and the rest of the spectrum out to the high-loss (> 50 eV) region is relatively low intensity.

From a spectroscopist's point of view, the ZLP is more of a problem than a useful feature in the spectrum because, as we mentioned in the previous chapter, it is so intense that it can saturate the PDA or the CCD detector, creating a ghost peak. So if you don't need to collect the ZLP in the spectrum then deflect it off the detector (or use the attenuator in the Gatan system). Conversely, from the microscopist's standpoint, as we'll see, selecting the ZLP to form an image or DP from which most of the energy-loss electrons have been excluded is a *very* useful technique. Conversely (again) filtering out the ZLP and forming images with selected energy-loss electrons is also extraordinarily useful.

38.2.B Removing the Tail of the ZLP

The intense ZLP has a tail, either side of it (go back and look at Figure 37.4), ultimately limited by the energy resolution. On the low (negative) energy side of the peak, the point-spread function accounts for the tailing, but on the high (positive) energy side there are contributions from the low-loss (e.g., phonon) electrons we just discussed. It is sometimes necessary to remove this tail before studying the (very) low-loss spectrum, e.g., for dielectric-constant determination (see Section 38.3). There are various ways to remove this tail in Gatan's commercial software (e.g., comparison with reference spectra, subtraction versus deconvolution) and the software continues to improve. You must make sure you are displaying the spectrum with a high dispersion and deconvolute the point-spread function before doing anything else.

The best way to remove the tail, if you *really* need to study the spectrum close to the ZLP, is to use a monochromator and cut off the tail of the peak at the source, so that any intensity outside the ZLP is a true low-loss part of the spectrum. In this particular case, since the low-loss spectrum is relatively intense, the principal argument against monochromation (i.e., you throw away a lot of expensive electrons) is seriously weakened. We've already mentioned this in Section 37.7 and Figure 37.16 compares the energy spread in the ZLP before and after monochromation.

If you don't have a monochromator, removal of the ZLP peak is challenging and prone to artifacts. The

main problem is that the shape of the ZLP measured with the beam on the specimen is not always the same as the ZLP measured in the hole because of phonons and elastic-scattering effects.

38.2.C Zero-Loss Images and Diffraction Patterns

If we filter out all the electrons which have lost energy greater than the resolution of the spectrometer (typically $> \sim 1$ eV) then basically we have an elastic image or DP. In doing this we immediately remove chromatic-aberration effects from the image, since it is the imprecise focusing of energy-loss electrons that degrades the resolution of TEM images from thick specimens. You should go back and take a look at Section 6.5.B and equation 6.16 and you'll see that, if many electrons suffer a typical (e.g., plasmon) $\mathcal{E}$ of $\sim$15 eV, your image resolution degrades from a few Å to several nanometers. For such $\mathcal{E}$ to occur, the specimen thickness has to be a goodly fraction of the plasmon mean free path (see Table 38.1), but that is not unusual. In addition to degrading the resolution by adding a diffuse component to otherwise focused TEM images, the energy-loss electrons also account for the diffuse intensity between spots in DPs. So filtering out these electrons should both increase the image contrast and improve the quality of the DPs.

The positive effect of energy filtering on resolution and *all* forms of TEM-image contrast has been known for several decades (see Egerton's early papers). This technique is particularly useful for enhancing the quality of images from thick biological (or polymeric) specimens in which the inelastic scattering is stronger than the elastic scattering (see Figure 38.2). However, diffraction contrast can also be enhanced by filtering of images from thick specimens (see Figure 38.3) but also in thin specimens because inelastically scattered electrons broaden the excitation error, **s**, thus reducing diffraction contrast. Sometimes enhancement of mass-thickness contrast images can be achieved by 'tuning' the spectrometer to select a specific range of energies (Figure 38.4). Contrast tuning (see Egerton's text) involves selecting an energy-loss window and sliding it around the spectrum while watching the image to find the best contrast. Tuning is useful in both the low and high-loss regions of

TABLE 38.1 Plasmon-Loss Data for 100-keV Electrons for Several Elements (from Egerton, 1996)

Material	$\mathcal{E}_P$ (calc)(eV)	$\mathcal{E}_P$ (expt)(eV)	θ_E(mrad)	θ_C(mrad)	λ_P(calc) (nm)
Li	8.0	7.1	0.039	5.3	233
Be	18.4	18.7	0.102	7.1	102
Al	15.8	15.0	0.082	7.7	119
Si	16.6	16.5	0.090	6.5	115
K	4.3	3.7	0.020	4.7	402

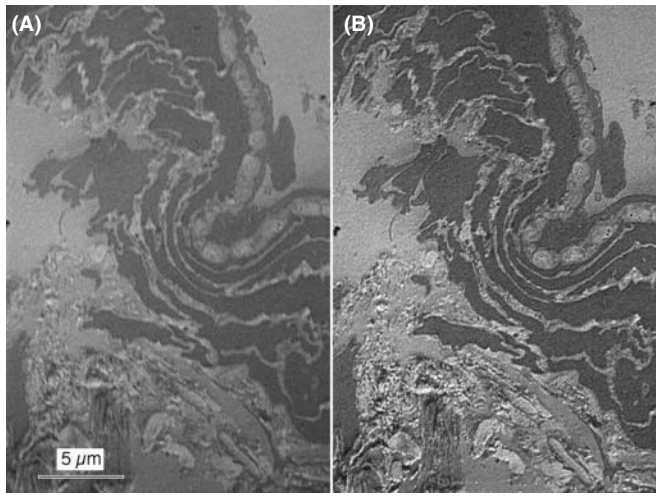


FIGURE 38.2. Comparison of (A) unfiltered and (B) filtered image of a thick biological section showing the enhanced contrast and resolution when the energy-loss electrons are removed.

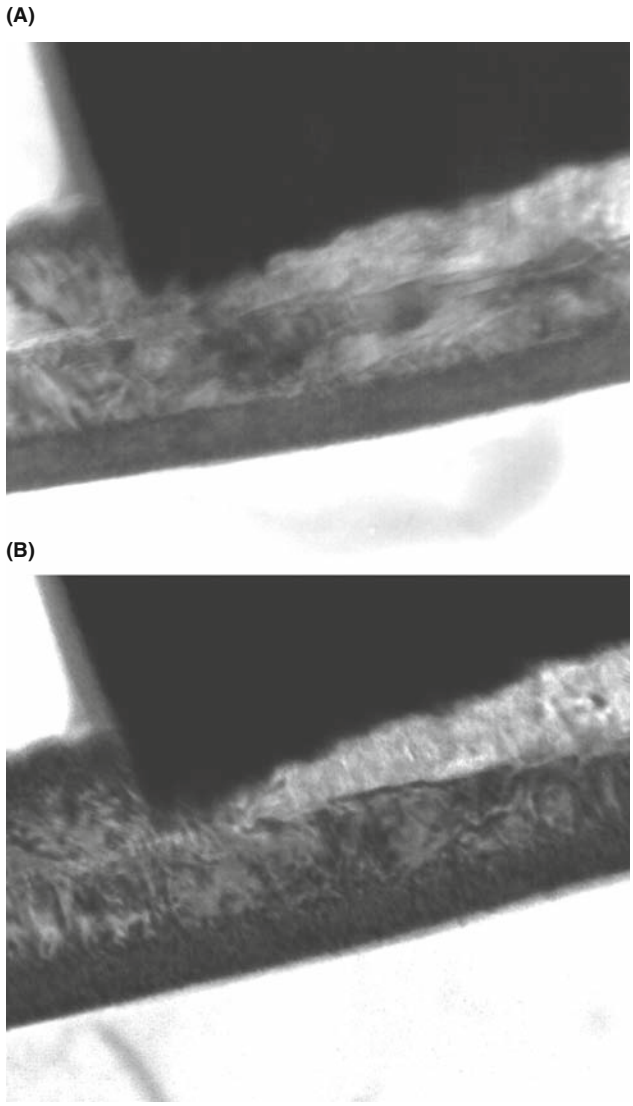


FIGURE 38.3. Comparison of (A) unfiltered and (B) filtered image of a thick crystalline specimen showing enhanced diffraction contrast when the energy-loss electrons are removed.

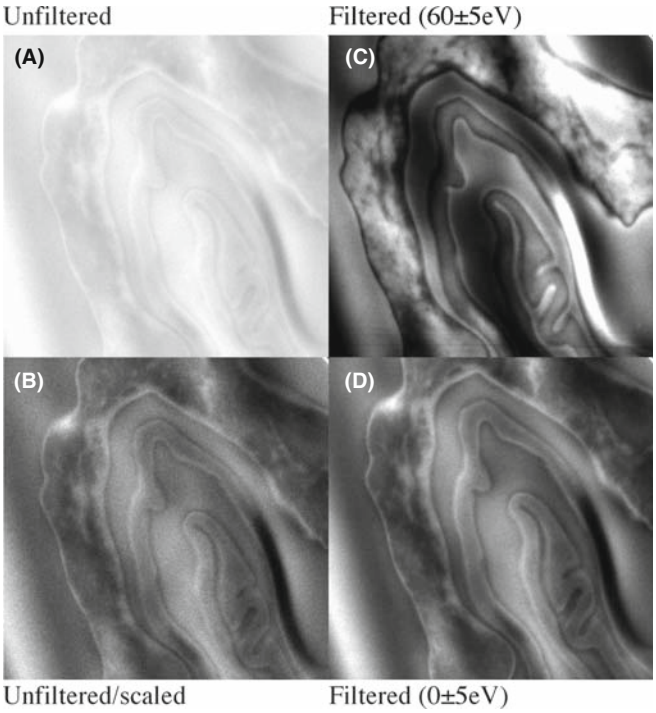


FIGURE 38.4. Contrast tuning of an image from a thick biological specimen to determine the region of the low-loss spectrum that gives the optimum contrast. (A) Unstained mouse epidermis (thickness = 0.1 μm, 100 keV). (B) Unfiltered image, digitally scaled to show the best contrast. (C) Filtered, contrast-tuned at 60±5 eV; the image has much better contrast than (B). (D) Filtered, contrast-tuned at 0±5 eV; the image has improved resolution and better contrast (B), but not as strong contrast as (C). Full width = 1 μm.

the spectrum and is typically done anywhere from 0 to 200 eV. As shown in Figure 38.4, there is significant possibility for contrast improvement in the low-loss region. In the high-loss region, selecting an energy window before an edge tends to reduce contrast from the edge electrons while windows after an edge tend to enhance contrast from the edge (see also jump-ratio imaging in Section 39.9).

High-resolution phase-contrast images (from thin specimens) should be more easily compared with theory if they are filtered to remove the diffuse background because we won't need so many 'fudge factors.' (See Chapter 28.) For a given thickness of specimen, the only alternative to reducing chromatic aberration is to go to higher voltages and this is perhaps an even more expensive option than buying a filter!

As we discussed back in Sections 37.8 and 20.5, if the energy-loss electrons are removed from SADPs and CBDPs it makes them much clearer (e.g., the paper by Midgley et al. and Reimer's chapter in Ahn's text), as shown back in Figures 20.10 and 37.17C. Energy filtering can also reveal extra diffraction information, such as the radial-distribution function for amorphous materials, which we'll tell you about later in Section 40.7.

If you've got a thick (several tens of nm) specimen (which is sometimes all you can manage to create), filtering may give many improvements

- Filtering can improve the image resolution.
- Filtering can enhance the image contrast no matter what mechanism is operating.
- Filtering can improve the contrast in diffraction patterns.
- Filtering can reveal finer detail in images and DPs.

So it would be really great if we could leave the filter switched on all the time, but this isn't always practically feasible.

Perhaps what is more surprising is why, given the tremendous advantage of filtering, simple zero-loss filters haven't been commercially available for decades? One experimental problem is that the energy-loss spectrometer is susceptible to external fields and the ZLP shifts over time, making continuous EFTEM imaging difficult: you have to continually re-align and re-focus the ZLP. Also, the best results, particularly for quantitative imaging (see the next chapter), require that your specimen has a similar thickness over the entire area of the image and that strong diffraction effects are minimized. Another possible reason is that filtering works best for thick specimens, which don't permit the TEM to perform at its (spherical-aberration) resolution limit of a fraction of a nanometer. It is perhaps not advantageous to remind users that most of their specimens are such that their TEM can't deliver anything like its best resolution, and that buying a 'better' TEM will have no beneficial effect on image quality or resolution for the great majority of (thick) specimens. But of course, this is only speculation. . .!

38.3 THE LOW-LOSS SPECTRUM

Look again at Figure 38.1, which is a typical low-loss spectrum and several points are immediately obvious

- After the ZLP, the plasmon peak is the next major peak.
- Apart from the plasmon, the spectrum is relatively featureless (the intensity changes are small).
- Despite the lack of features, there's still a lot of counts (check the units of the ordinate), so extracting useful data is still feasible and imaging should be relatively straightforward.

The cut-off energy for the low-loss spectrum, as we've already noted, is $\sim$ 50 eV and the reason for this is that the other principal features of energy-loss spectra, namely, the ionization edges, don't appear (at least for solids) until $\mathcal{E} > 50$ eV. In the low-loss spectrum, we

are detecting beam electrons that have interacted with conduction and/or valence bands (hence another common term for the low-loss spectrum is 'valence spectrum'). These weakly bound electrons control many electronic properties of the specimen. In general, the low-loss spectrum is not as well understood as the high-loss spectrum and there has not been the same effort put into modeling low-loss spectra, as we'll describe for the higher-loss spectra in Chapters 39 and 40. However, things are beginning to change, as you'll see later in this chapter.

While we have just shown you the tremendous advantages of filtering out the plasmon peak in order to enhance contrast and resolution in TEM images and DPs, there is also much to be gained by imaging the plasmon peak and filtering out the ZLP and high-loss electrons. This approach was pioneered by Batson. Because of the strength of the signal, plasmon imaging is becoming a much more popular technique, particularly for mapping out low-loss properties of nanomaterials (e.g., see the papers by Eggeman et al. and by Ding and Wang).

38.3.A Chemical Fingerprinting

So what can we do with the low-loss spectrum? Because there are a lot of counts, we can use the shape of the spectrum to help identify specific phases or features in the TEM image with some degree of statistical certainty. The low-loss spectrum all the way up to ~50 eV, including any plasmon peaks (see Section 38.3.C), should be used for fingerprinting.

FINGERPRINTING

The low-loss spectrum can only be used for phase identification through a 'fingerprinting' process. You store the spectra of known specimens in a library in the computer.

So you overlay your unknown spectrum on one or more stored library-standard spectra. Figure 38.5 shows how low-loss spectra vary (A) for aluminum and various compounds and (B) for the main constituents of biological specimens. A collection of low-loss spectra from many elements and common compounds (mainly oxides) has been compiled in various databases, such as the EELS Atlas or on the Web at URL #4. Such sources can help considerably with fingerprinting unknown features in your image. As with any fingerprinting technique, including the forensic variety, you must be careful to decide when a 'match' is satisfactory. There is no 'black and white' here, only shades of gray, so don't convict unless the statistics are with you and there is strong supporting evidence from other techniques.

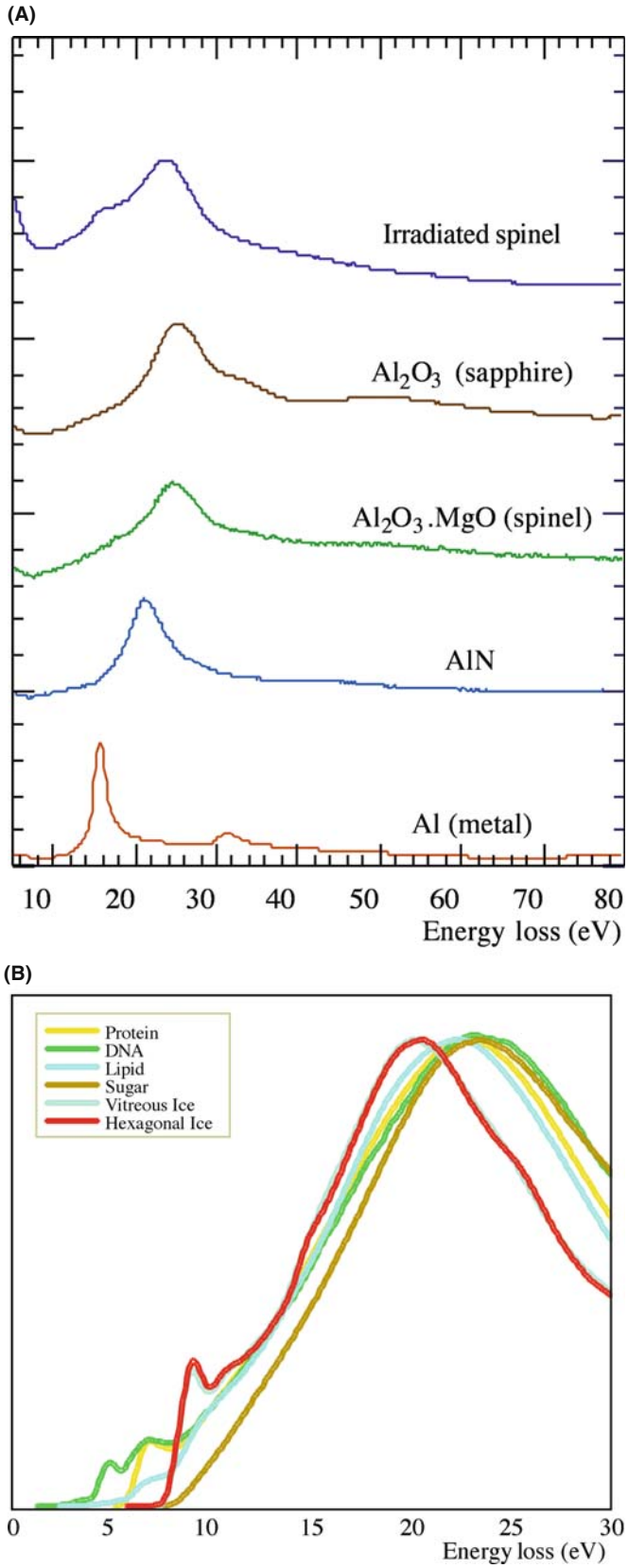


FIGURE 38.5. (A) The low-loss spectrum from specimens of Al and various compounds, showing differences in the intensity variations that arise from differences between the Al-Al, Al-O, and Al-N bonding. (B) Low-loss spectra from the principal components of cellular tissue.

38.3.B Dielectric-Constant Determination

We can view the energy-loss process as the dielectric response of the specimen to the passage of a fast electron. As a result, the very low energy spectrum (up to $\mathcal{E} \sim 20$ eV) contains information about the dielectric constant or permittivity (ε). Localized dielectric-constant measurements are of great interest as the semiconductor industry explores high-dielectric materials, such as HfO_2, for the next generation of nanometer-scale gate oxides.

Assuming a free-electron model, the single-scattering spectrum intensity $I(\mathcal{E})$ is related to the imaginary (Im) part of the dielectric constant ε by the expression (modified from Egerton)

$$I(\mathcal{E}) = I_0 \frac{t}{k} \mathrm{Im}\left(-\frac{1}{\varepsilon}\right) \ln\left[1 + \left(\frac{\beta}{\theta_E}\right)^2\right] \qquad (38.5)$$

where I_0 is the intensity in the ZLP, t is the specimen thickness, k is a constant incorporating the electron momentum and the Bohr radius, β is the collection angle (again note its importance), and θ_E is the characteristic-scattering angle. You can use a Kramers-Kronig analysis to analyze the energy spectrum in order to extract the real part of the dielectric constant from the imaginary part in equation 38.5 and details are given in Egerton's text. As usual, the advantage to doing this kind of measurement in the TEM is the high spatial resolution and the advantages of this are exemplified by the use of low-loss spectroscopy to determine optical gaps on BN nanotubes (Arenal et al.). Since you need a single-scattering spectrum, removing the plural scattering intensity by Fourier-logarithmic deconvolution is the first step (see Section 39.6) when determining the dielectric constant. The Gatan software package has the appropriate programs and public-domain software is also available, e.g., URLs #1, 5, and 6.

KRAMERS-KRONIG ANALYSIS
This analysis gives the energy-dependence of the dielectric constant and other information, which we usually obtain by optical spectroscopy.

The alternatives to EELS for this kind of work are various kinds of optical and other electromagnetic-radiation techniques. The part of the low-loss spectrum from ~1.5 to 3 eV is of great interest and corresponds to optical analysis of the dielectric response from the infra-red (~800 nm) through the ultra-violet (~ 400 nm) wavelength range. (This correspondence between EELS and optical spectroscopy only holds for small angles of scattering so the value of β you choose must be small (< ~ 10 mrads) thus lowering the intensity in the spectrum.) Higher energies correspond to various electronic transitions. Thus, in a single EELS experiment you can, in theory, substitute for a whole battery of optical-spectroscopy instrumentation (although optical spectroscopy techniques do offer even better energy resolution than EELS). Remember that EELS always offers better spatial resolution.

There is a tremendous similarity between the TEM-EELS approach and the valence (surface) EELS approach, including the need for Kramers-Kronig and deconvolution software. TEM-based EELS is in the extremely low-energy (i.e., low-frequency) range around 1 eV and below and corresponds to far infra-red spectroscopy which is well into the energy range of studies of bond vibrations. Higher up the energy-loss range corresponds to the visible and ultra-violet ranges.

If you don't have access to a monochromator, you can use software to remove the contribution of the tail of the ZLP but, as we mentioned above, be careful, because this process may introduce its own artifacts. An example of the correspondence between EELS and optical valence spectra is shown in Figure 38.6. In Figure 38.6A, the importance of initial deconvolution is demonstrated and the deconvoluted valence spectra are compared with ultra-violet spectra in Figure 38.6B. It is straightforward to assign the various peaks in the low-loss spectra to specific interband transitions and also to compare the data with band-structure calculations (e.g., van Bentham et al.). Thus, the electronic and optical properties can be obtained and you can, of course, select any of the features in the low-loss spectrum and form images with those electrons. So dielectric-constant imaging is feasible, as is imaging the various other low-loss signals which we'll now discuss.

FOR THE BEST LOW-LOSS SPECTROSCOPY
You need an FEG, a high-resolution, high-dispersion spectrometer and if you're really going to do it properly, a monochromator, so the tail of the ZLP does not mask the low-energy intensity.

38.3.C Plasmons

Plasmons are longitudinal wave-like oscillations that occur when a beam electron interacts with the weakly bound electrons in the conductance or valence band. You can think of plasmons as being like the ripples that spread out from where a pebble is dropped into a pond. But, unlike in a pond, the oscillations are rapidly damped, typically having a lifetime of about 10^{-15} s and so are quite localized to <10 nm. The plasmon peak is the second most dominant feature of the energy-loss spectrum after the ZLP. The small peak beside the ZLP in Figure 38.1 is a plasmon peak.

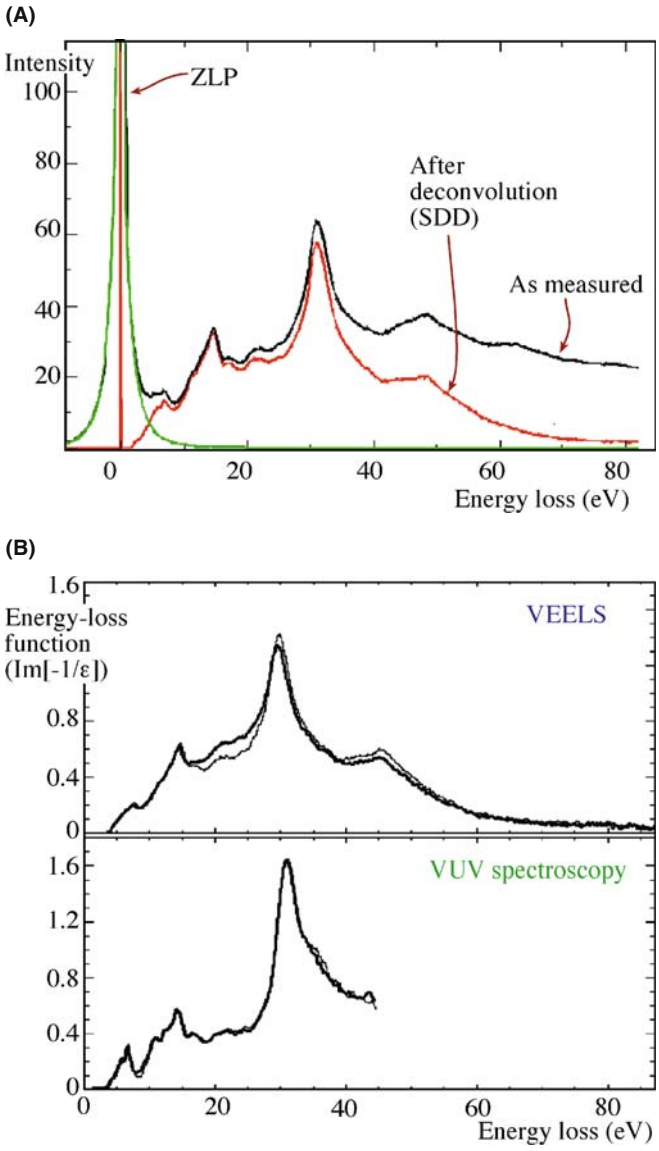


FIGURE 38.6. (A) Low-loss (valence) spectrum from $SrTiO_3$ before (black) and after (red) Fourier-log deconvolution. The extracted ZLP is shown in green. (B) Comparison of the change in the imaginary part of the complex dielectric function obtained from pairs of valence EELS (VEELS) spectra and valence ultra-violet (VUV) spectra from two different regions of $SrTiO_3$. The spectra show similar features but the VUV spectrum cannot be measured beyond ~ 45 eV.

If we assume the electrons are free (i.e., not bound to any specific atom or ion), then the energy $\mathcal{E}_P$ lost by the beam electron when it generates a plasmon of frequency ω_p is given by a simple expression

$$\mathcal{E}_P = \frac{h}{2\pi}\omega_p = \frac{h}{2\pi}\left(\frac{ne^2}{\varepsilon_0 m}\right)^{\frac{1}{2}} \qquad (38.6)$$

where h is Planck's constant, e and m are the electron charge and mass, ε_0 is the permittivity of free space (remember the dielectric constant is the relative permittivity of a polarizable medium), and n is the free-electron density. Typical values of $\mathcal{E}_P$ are in the range 5–25 eV and a summary of plasmon-loss characteristics is given in Table 38.1

- Plasmon losses dominate in materials with free-electron structures, such as Li, Na, Mg, and Al.
- Plasmon-like peaks occur to a greater or lesser extent in the low-energy spectra from all materials, including insulators, such as polymers and biological tissue.

So the 'free-electron' assumption is clearly not rigorous and we don't know everything about how this feature arises.

From equation 38.6, you can see that $\mathcal{E}_P$ is affected by n, the free-electron density. Interestingly, n may change with the chemistry of the specimen. Thus, measurement of the plasmon-energy loss can give indirect analytical information (see the next section).

The characteristic plasmon-scattering angle θ_E is very small, being typically < 0.1 mrads (as listed in Table 38.1), which means that the plasmon-loss electrons are strongly forward scattered. Their cut-off angle $\theta_{c\ is}$ much greater than θ_E so if you use a collection angle β of only 10 mrads, you will easily gather almost all the plasmon-loss electrons (again note the importance of knowing β in your system). Conversely, this means that even a small objective aperture will not stop plasmon-loss electrons entering the TEM imaging system. Plasmon-loss electrons also carry contrast information and, because they are the most intense energy-loss signal, they are the primary contribution to chromatic aberration in TEM images, which is why it is often a good move to filter them out. As we've already seen, Figure 38.2 shows the improvement in image contrast and resolution when the low-loss (primarily plasmon) spectrum is filtered out of the image of a specimen showing predominantly mass-thickness contrast. Likewise, Figure 38.3 shows a thick foil exhibiting primarily diffraction contrast. A similar improvement in resolution occurs when the many plasmon peaks are filtered out.

A typical value of the plasmon mean free path λ_P at AEM voltages is about 100 nm and so it is reasonable to expect at least one strong plasmon peak in all but the thinnest specimens. Likewise, the number of individual losses should increase with the thickness of your specimen and we can use the plasmon-peak intensity to estimate the specimen thickness. If indeed your specimen is so thin that only single scattering occurred, and the only significant scattering was a single plasmon event, then you should be very pleased because it's a great specimen for ionization-loss EELS (see the next chapter). Conversely, if your spectrum shows several plasmon peaks then it is too thick for ionization-loss studies. Under single-scattering circumstances we can assume

$$t = \lambda_p \frac{I_p}{I_0} \qquad (38.7)$$

where λ_P is the plasmon mean free path, I_P (see Figure 38.1) is the intensity in the first (and only) plasmon peak, and I_0 is the intensity in the zero-loss peak.

BALLPARK CALCULATION

A typical ballpark figure: if the intensity in the first plasmon peak is greater than one tenth the zero-loss intensity then your specimen is too thick for EELS quantification.

The method has advantages over other thickness measurement techniques in that you can apply it to any specimen, amorphous or crystalline, over a wide range of thicknesses. We'll tell you more about EELS thickness measurements and their role in ionization-loss spectrometry in Section 39.5.

If plural scattering is significant, then the spectrum becomes more difficult to interpret and other problems arise; e.g., your ionization-loss quantification results (next chapter) become unreliable.

Of course, one way around this problem is to use very thin foils, but often you can't produce thin-enough specimens. Murphy's law says that the area you're interested in will usually be too thick. Then you have to deconvolute the spectra, again using the Fourier-log approach (Section 39.6) to make the single-scattering assumption valid. As we've already noted, deconvolution brings its own problems.

Figure 38.7 shows the plasmon-loss spectra from (A) thin and (B) thick foils of pure Al and (B) also indicates how the Gatan software uses this information to come up with a measure of the local foil thickness. Since Al is a good approximation to a free-electron metal, the plasmon-loss process is the dominant energy-loss event. Plural-plasmon scattering in thicker foils is of concern because it limits the interpretation of high energy-loss spectra containing chemical information from ionization losses in which we are really interested (see Section 39.4).

The plasmon losses which we've just described all arise from interactions with the electrons in the interior of your specimen, but the incident electrons can also set up plasmon oscillations on the specimen surface. We can envisage these surface plasmons as transverse charge waves. Surface plasmons have about half the energy of volume plasmons (because the surface atoms are not so strongly bound). Generally, however, the surface plasmon peak is much less intense than the volume plasmon peak(s), even in the thinnest specimens, but you can still use them for imaging, as shown by Batson. With monochromators and aberration correctors, studies of surface plasmons, along with other low-loss features, will assume more importance in the TEM.

38.3.D Plasmon-Loss Analysis

As we just mentioned, the plasmon peaks contain chemical information, because the composition of the specimen may affect the free-electron density, n, which in turn changes the plasmon-loss peak position. Historically, this technique was the first aspect of EELS to produce quantitative analysis data, and it was used in a limited number of systems, mainly aluminum and magnesium alloys in which the plasmon-loss spectrum is dominant and consists of sharp Gaussian peaks (Williams and Edington). The lack of a more recent review gives some indication of the limitations of this approach (see below)!

The principle of plasmon-loss analysis is based on empirical observation of the shift in the plasmon-peak position ($\mathcal{E}_P$) with composition (C), giving an expression of the form

$$\mathcal{E}_P(C) = \mathcal{E}_P(0) \pm C\left(\frac{d\mathcal{E}_P}{dC}\right) \qquad (38.8)$$

where $\mathcal{E}_P(0)$ is the plasmon loss for the pure component. By creating a series of binary alloys of known composition we can develop a working curve which we can then use to calibrate measurements of $\mathcal{E}_P$ in unknown alloys.

Since plasmon-loss analysis demands the measurement of peak *shifts* rather than peak positions, you need an energy spectrum of the highest resolution and sufficient dispersion to measure the peak centroid accurately. The early plasmon-loss studies did not have

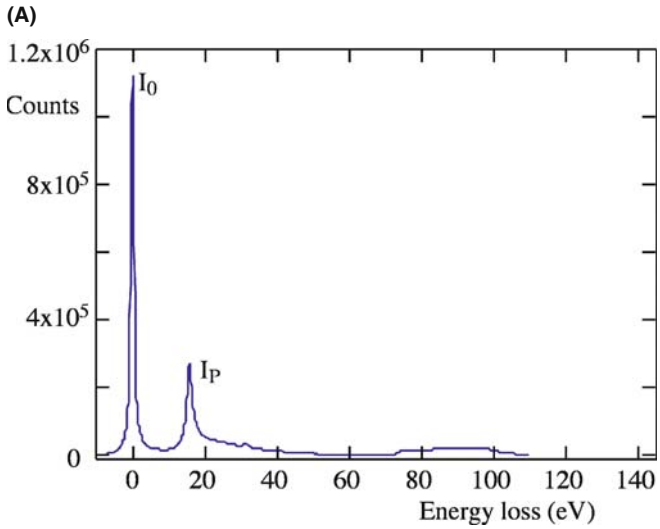


FIGURE 38.7. (A) The low-loss spectrum from a very thin specimen of pure Al showing the intense ZLP (I_0) and a small plasmon peak (I_P) at about 15 eV. (B) The low-loss spectrum from a thick specimen of pure Al showing several plasmon peaks, the first of which is almost as intense as the ZLP. The inset shows the calculation of the thickness from the Gatan software.

(B)

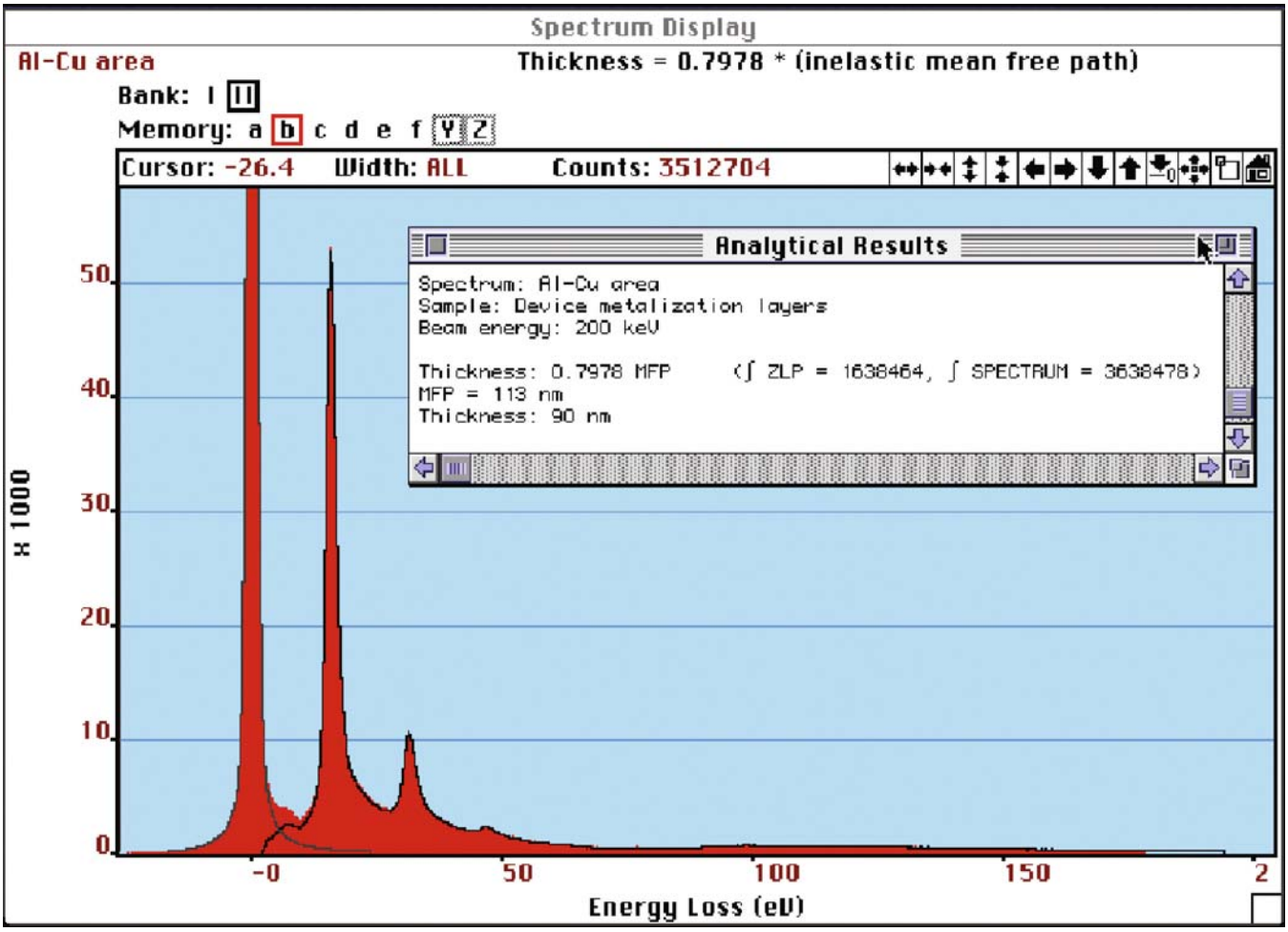


FIGURE 38.7. (Continued).

access to FEGs and so the resolution of the thermionic source was a limiting factor. Figure 38.8 illustrates some early plasmon-loss concentration data and the visible peak shifts that occur and also shows how we can use the plasmon peak shifts in Al-Li alloys and convert them into Li concentration data and also create Li compositions maps, which, given its low Z, is rather difficult to do with other analytical techniques.

Plasmon-loss spectrometry has reasonable spatial resolution and is relatively insensitive to specimen thickness and surface deposits. The spatial resolution is controlled by the localization of the plasmon oscillation which is only a few nm, since the plasmon disturbance is rapidly damped by the free electrons. The specimen thickness only affects the number and intensity of the plasmon peaks, not their position, as you can see in Figure 38.7. In fact you get the best results from plasmon-loss spectrometry when your specimen is about 1–2 mean free paths (λ_P) thick so that several, intense, Gaussian peaks are observable. There are, unfortunately, strong practical disadvantages, which account for the almost complete absence of plasmon-loss data since the advent of ionization-loss techniques in the mid-1970s

- We are limited to specimens showing well-defined peaks, and only binary specimens can be sensibly analyzed.
- The alloying element must produce a detectable change in $\mathcal{E}_P$ and this is not always the case. For example, the addition of 30 at.% Zn to Al scarcely changes $\mathcal{E}_P$.

It is possible that application of modern detection and data processing techniques may improve the quality and ease of analyzing plasmon-loss spectra. While plasmon peak-shift analysis is limited, we can at least use the low-loss plasmon spectra for chemical fingerprinting, as we've already described, and we'll discuss the prospects for more quantitative interpretation of low-loss spectra via modeling in Section 38.4.

With increasing interest in the mechanical properties of nanoscale materials, the fact that strong scaling

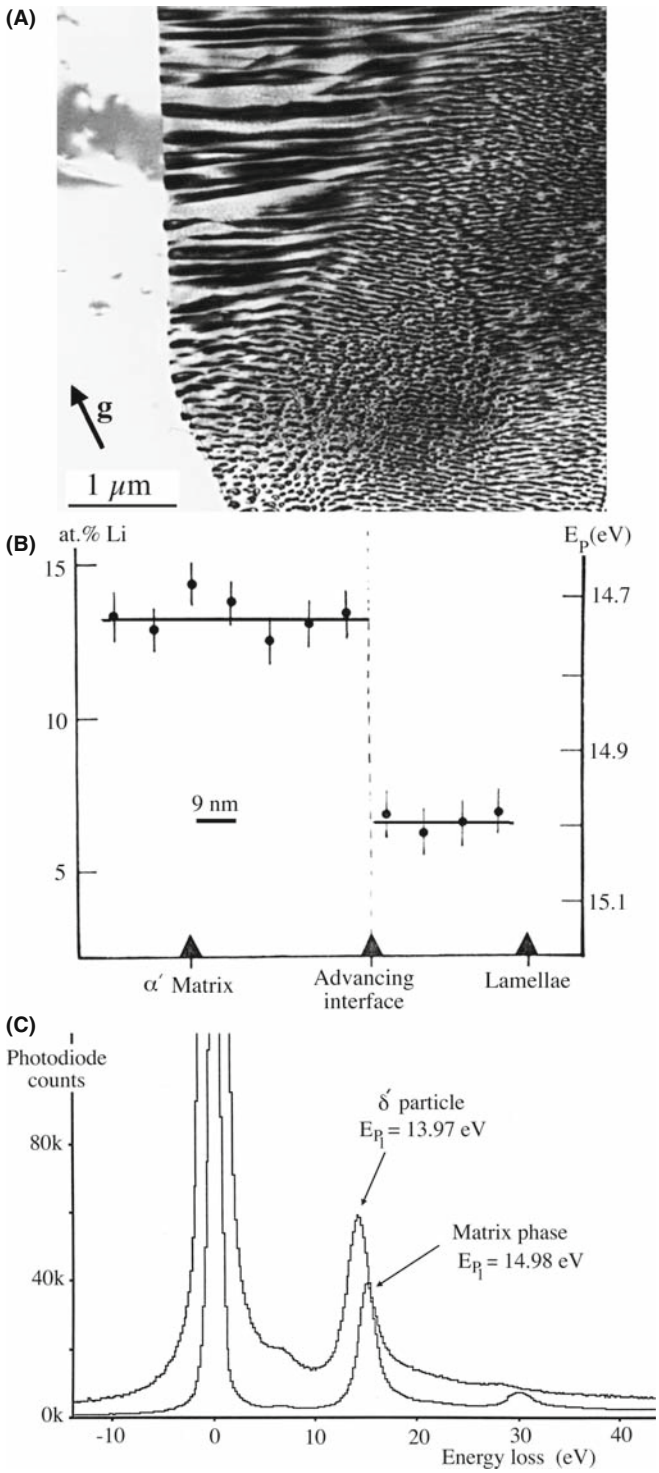


FIGURE 38.8. (A) A discontinuous precipitation reaction front in an Al-11 at.% Li specimen. (B) Experimental plasmon-loss measurements of the Li composition variation across the interface. (C) The shift in the plasmon peak for the matrix (5 at.% Li) and the precipitate (25 at.% Li) is clear in the two spectra.

correlations exist between the plasmon energy and elastic properties, hardness, valence-electron density, and cohesive energy is leading to a resurgence of interest in this part of the spectrum (e.g., Oleshko and Howe).

38.3.E Single-Electron Excitations

A high-energy beam electron may transfer sufficient energy to a single electron in the valence band to change its orbital state, perhaps moving it to an unoccupied state in the conduction band. We call these events single-electron interactions and they result in inter/intraband transitions for the valence electrons, with energy losses of up to ~25 eV. An example of an interband transition is given in the spectra from different polymers which can be distinguished solely by their electronic differences (Hunt et al.), as shown in Figure 38.9A. Interactions with molecular orbitals, such as the π orbitals produce characteristic peaks in this low-energy region of the spectrum, sometimes causing shifts in the plasmon peak (either up or down depending on the relative energy of the interband transition and the plasmon loss), and that is why it is possible to use the intensity variation in this part of the spectrum to fingerprint a particular phase. A more challenging example is given in Figure 38.9B, which shows what can be done with a combination of cryo- and low-dose STEM to image a polymer nano-emulsion in an aqueous medium. The low-loss spectra reveal the electronic differences between the phases (including amorphous ice!) and the filtered images show the lobed shape of the emulsified particles. There is probably no other technique that could image such beam-sensitive material at such high resolution (Kim et al.).

If a beam electron gives a weakly bound, valence-band electron sufficient energy to escape the attractive field of the nucleus, then a secondary electron (SE) is created, of the sort used to give topographic images in the SEM and STEM. Typically, a SE requires < 20 eV to escape the surface and therefore the electrons causing SE emission appear in the same low-energy region of the spectrum as the inter- and intraband transitions.

38.3.F The Band Gap

In the region of the spectrum immediately after the ZLP, and before the rise in intensity preceding the plasmon peak, you can see a region of low intensity. If there are no interband transitions occurring, the intensity in this portion of the spectrum approaches the dark-current (noise) level of the detector. This low intensity implies that there is a forbidden-transition region, which is simply the band gap, between the valence and conduction bands in semiconductors and insulators. To determine the band gap, you need to strip off the tail of the ZLP (with all the consequent difficulties) and measure the energy range of the gap out to the rise in the initial low-loss spectrum. Figure 38.10A illustrates the variable band gap in spectra from specimens of Si, and its oxide and nitride. Mapping this change in the energy range in which no transitions occur gives band-gap images (Figure 38.10B) and several examples of this are given by Kimoto et al. As sub-nanometer-scale semiconductor technology advances,

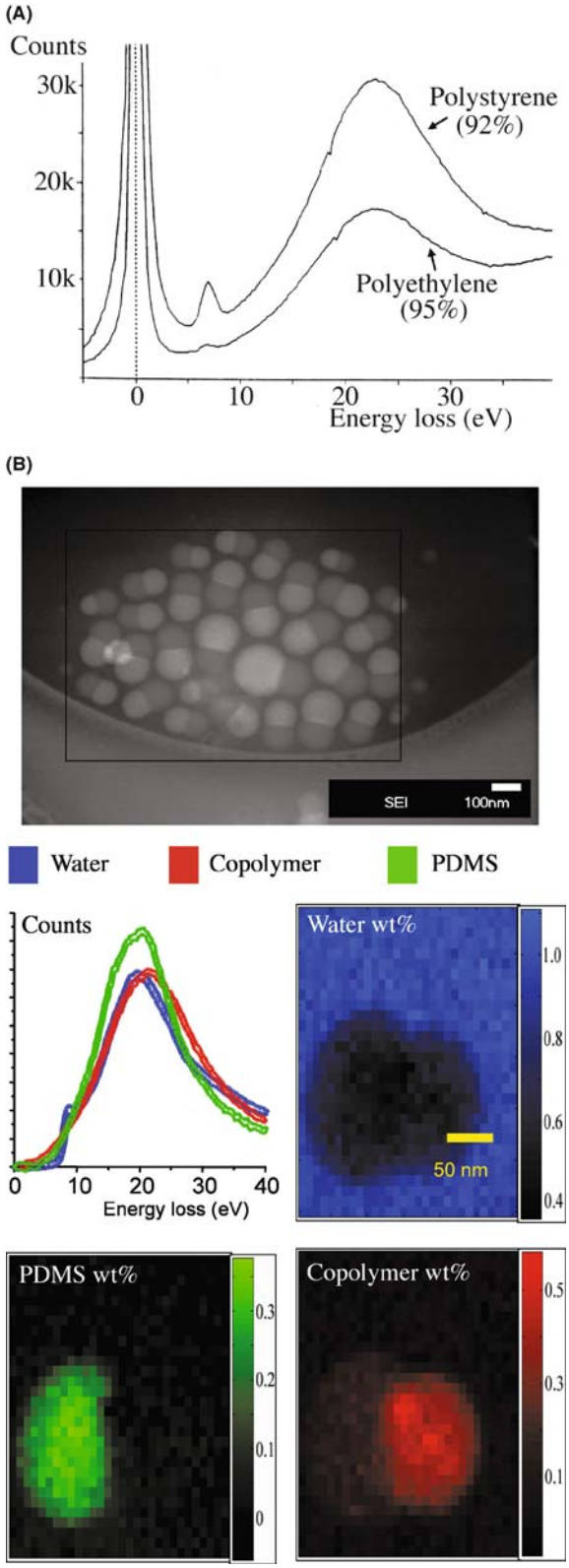


FIGURE 38.9. (A) The interband transition characteristic of polystyrene, clearly visible on the rise of the plasmon peak, compared with the absence of such a transition in polyethylene. (B) (top) A low-dose cryo-HAADF image of a two-phase polymer nano-emulsion in water; (bottom) low-loss spectra from amorphous ice (blue) polydimethyl siloxane (PDMS) (green) and a multi-phase co-polymer (red), together with the corresponding composition maps based on the differences in their low-loss spectra.

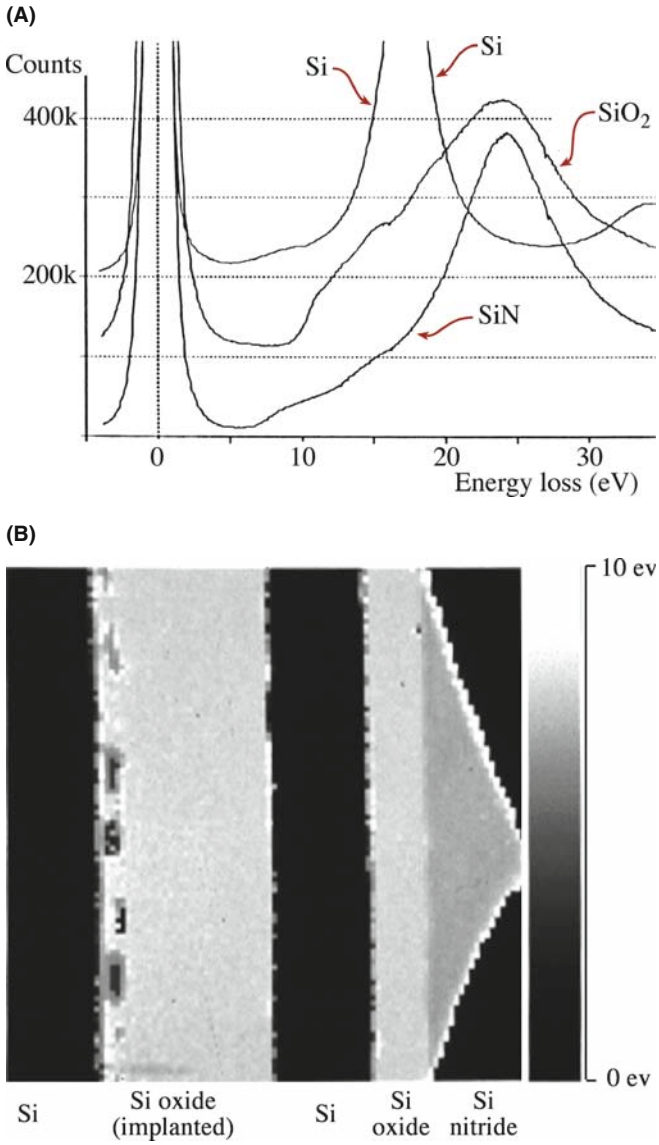


FIGURE 38.10. (A) Band-gap differences evident in the low-loss spectra of a Si semiconductor, SiO_2, and Si nitride (almost Si_3N_4) ceramic insulators and (B) the corresponding band-gap image (with scale on right; recorded at 90 K using 1024 channels and 150 ms dwell time). Note the Si islands in the oxide layer which were not visible in the TEM image.

the need for sub-nanometer resolution imaging of the band gap will increase and low-loss EEL images are the only way to visualize this electronic property.

38.4 MODELING THE LOW-LOSS SPECTRUM

As you now know, the low-loss spectrum has the advantage of significant intensity (so counting statistics are not a problem) and it contains useful data about your specimen, such as composition, bonding, the dielectric constant, the band gap, the free-electron density, and optical properties. With all this information you might have thought that we understood the spectrum very well and

were able to model it with some degree of precision and could use the modeling to predict spectra from different materials. Perversely, we are better able to do this for the much lower intensity, high-loss spectrum, as we'll describe in Chapters 39 and 40. However, significant progress is being made in calculating the plasmon-loss energies and interband transitions. As we've seen, the plasmon peak is basically an oscillation of the free electrons, so equation 38.6 has been used over several decades to calculate the plasmon-loss energy, but this approach can't handle the effects of other low-loss features like interband transitions. French has developed software for low-loss modeling, called Electronic Structure Tools (see URL #1), which consists of a number of programs for the quantitative analysis of optical, VUV, and EELS spectra. Keast has shown good agreement between experimental and calculated low-loss spectra for a range of metals and ceramics, as shown in Figure 38.11, using abinitio methods, which we'll describe in some more detail in Chapter 40, and this topic is dealt with extensively in the companion text. Modeling of such spectra requires careful experimental control and for the data of Figure 38.11 the convergence angle of the (100-kV) beam was 8.3 mrad, the Gatan spectrometer collection angle was 5.8 mrad, 100 spectra (0.05 s per acquisition) were aligned, corrected for dark current and gain variations, and summed. The density-functional theory (see Section 40.5.A) calculations (using the random-phase approximation and neglecting local-field effects) were performed using the WIEN2k code. Exchange and correlation effects were treated using the generalized gradient approximation. The final spectrum was averaged over the different orientation components. So you get the idea that this is not straightforward!

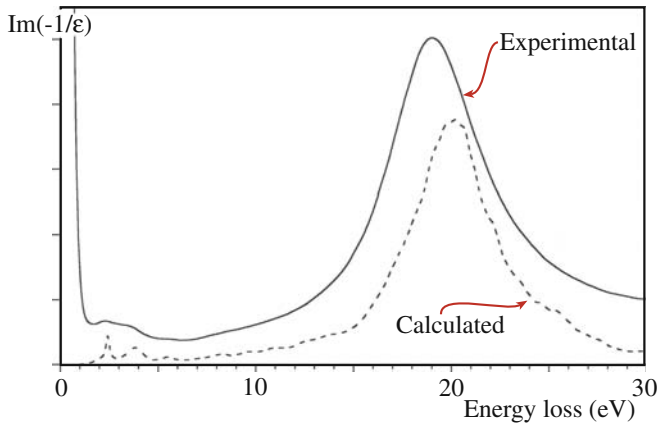


FIGURE 38.11. A comparison between the calculated (dashed) and experimental (full) low-loss spectrum from commercial MgB_2 particles.

Software for all aspects of low-loss analysis can be found at URLs #1 and 6.

CHAPTER SUMMARY

The low-loss (valence/plasmon) portion of the spectrum from 0 to 50 eV contains a wealth of useful information about the specimen.

- The ZLP is the most intense signal. If you filter out all the low-loss electrons apart from the ZLP, you get images and DPs which generally show higher resolution and better contrast than unfiltered ones because they are free of chromatic aberration and diffuse-scattering effects.
- The low-loss spectrum reflects beam interactions with loosely bound conduction and valence-band electrons.
- From different portions of the low-loss spectrum, you can measure the local dielectric constant of your specimen, the free-electron density, the thickness, the band gap, and observe inter/intraband transitions. You can also form images using energy-loss electrons which map out all these phenomena, generally with sub-nanometer resolution.
- The low-loss spectrum can be used to fingerprint (identify) specific elements, compounds, and biological tissue by comparison with the characteristics of standard spectra in databases.
- In some binary alloy systems of light elements, you can determine composition by measuring shifts in the plasmon-peak centroid. Plasmon imaging also has the potential for mapping nanoscale mechanical properties.
- We are getting much better at simulating the low-loss spectrum and understanding the various beam-specimen interactions that contribute to this high-intensity portion of the spectrum.

THE EELS ATLAS

Ahn, CC Ed. 2004 *Transmission Electron Energy-Loss Spectrometry in Materials Science and the EELS Atlas* 2nd Ed. Wiley-VCH Weinheim Germany. Buy this.

Ahn, CC and Krivanek, OL 1983 *EELS Atlas* Gatan Inc., 5933 Coronado Lane Pleasanton CA 94588. Buy this too (if you can find it).

SOME CALCULATIONS AND SPECIAL CONCEPTS

Egerton, RF 1976 *Inelastic Scattering and Energy Filtering in the Transmission Electron Microscope* Phil. Mag. **34** 49–65. One of the earliest indications of the power of EEL techniques.

Egerton, RF 1996 *Electron Energy Loss Spectroscopy in the Electron Microscope* 2nd Ed. Plenum Press New York. Includes the idea of high-contrast tuning.

Eggeman, AS, Dobson, PJ and Petford-Long AK 2007 *Optical Spectroscopy and Energy-Filtered Transmission Electron Microscopy of Surface Plasmons in Core-Shell Nanoparticles* J. Appl. Phys. **101** 024307–10.

Keast, VJ 2005 *Ab Initio Calculations of Plasmons and Interband Transitions in the Low-Loss Electron Energy-Loss Spectrum* J. Electron Spectrosc. Relat. Phenom. **143** 97–104.

Schattschneider, P and Jouffrey, B 1995 *Plasmons and Related Excitations* in Reimer, L Ed. *Energy-Filtering Transmission Electron Microscopy* 151–224 Springer New York. A thorough introduction to plasmons and related excitations.

APPLICATIONS

Arenal, R, Stéphan, O, Kociak, M. Taverna, D. Loiseau, A and Colliex, C 2005 *Electron Energy Loss Spectroscopy Measurement of the Optical Gaps on Individual Boron Nitride Single-Walled and Multi-walled Nanotubes* Phys. Rev. Lett. **95** 127601–127604.

Batson, PE 1982 *Surface Plasmon Coupling in Clusters of Small Spheres* Phys. Rev. Lett. 49 936–940.

Ding, Y and Wang, ZL 2005 *Electron Energy-Loss Spectroscopy Study of ZnO Nanobelts* J. Electr. Microsc. **54** 287–291.

Hunt, JA, Disko, MM, Behal, SK and Leapman, RD 1995 *Electron Energy-Loss Chemical Imaging of Polymer Phases* Ultramicrosc. **58** 55–64.

Kim, G, Sousa, A, Meyers, D, Shope, M and Libera, M 2006 *Diffuse Polymer Interfaces in Lobed Nanoemulsions Preserved in Aqueous Media* J. Am. Chem. Soc. **128** 6570 –6571.

Kimoto, K, Kothleitner, G, Grogger, W, Matsui, Y and Hofer F 2005 *Advantages of a Monochromator for Bandgap Measurements Using Eelectron-Loss Spectroscopy* Micron **36** 185–189.

Midgley, PA, Saunders, M, Vincent, R and Steeds, JW 1995 *Energy-Filtered Convergent-Beam Diffraction: Examples and Future Prospect*s Ultramicrosc. **59** 1–13.

Oleshko, VP and Howe, JM 2007 *In Situ Determination and Imaging of Physical Properties of Metastable and Equilibrium Precipitates Using Valence Electron Energy-Loss Spectroscopy and Energy-Filtering Transmission Electron Microscopy* J. Appl. Phys. **101** 054308–11.

Reimer, L 2004 *Electron Spectroscopic Imaging* in *Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS Atlas* 2nd Ed. 347–400 Ed. CC Ahn Wiley-VCH Weinheim Germany.

Van Bentham, K, Elsasser, C and French, RH 2001*Bulk Electronic Structure of SrTiO3: Experiment and Theory* J. Appl. Phys. **90** 6156–6159.

Williams, DB and Edington, JW 1976 *High Resolution Microanalysis in Materials Science Using Electron Energy Loss Measurements* J. Microsc. **108** 113–145. Historical but not superceded!

URLs

1) http://www.lrsm.upenn.edu/∼frenchrh/index.htm
2) http://www.hremresearch.com/Eng/download/documents/EELScatE2.html
3) http://www.gatan.com/answers2/index.php
4) http://www.cemes.fr/%7Eeelsdb/
5) http://www.cemes.fr/epsilon/home/main.php
6) http://www.deconvolution.com/

SELF-ASSESSMENT QUESTIONS

Q38.1 Distinguish the low-loss and high-loss regions of the spectrum.
Q38.2 What is usually the second most intense peak in any spectrum? What might be the second most intense peak in a spectrum from a very thick specimen?
Q38.3 List the characteristic scattering angles of the principal energy-loss processes and give ballpark values. How do these compare with other important scattering angles in TEM such as typical Bragg angles?
Q38.4 What's a typical value for a plasmon-energy loss?
Q38.5 What are inter- and intraband transitions and why do they result in relatively low energy losses?
Q38.6 Why is it important for the ZLP to be the most intense peak in the spectrum by a factor of 10 or more?
Q38.7 What's another expression for the 'permittivity of free space'?
Q38.8 What is meant by the 'free-electron density' and what role does it play in low-energy losses?
Q38.9 Why is the plasmon peak the most prominent energy-loss peak in the spectrum from a thin specimen?

Q38.10 What is the difference between the characteristic and the cut-off angle? Which is more important in EELS and why?
Q38.11 What electrons are in the ZLP?
Q38.12 Under what conditions would you wish to remove the tail of the ZLP?
Q38.13 Describe one other way to measure the dielectric constant apart from low-loss EELS. What are the relative advantages and disadvantages of the two approaches?
Q38.14 What's the best way to remove the tail of the ZLP?
Q38.15 What is fingerprinting and why should you be cautious about using it?
Q38.16 Why would you ever bother to form an image from which the energy-loss electron have NOT been removed?
Q38.17 Why would you ever bother to form a CBED pattern from which the energy-loss electrons have NOT been removed?
Q38.18 What is a Kramers-Kronig transformation? What information does it extract from the low-loss spectrum?
Q38.19 Why hasn't there been more use of plasmon-shift measurements for composition determination?
Q38.20 Explain why you might want to model the intensity in the low-loss spectrum.
Q38.21 Distinguish single, plural, and multiple scattering. Which is best for EELS and why?

TEXT-SPECIFIC QUESTIONS

T38.1 Distinguish the characteristic scattering angle, the cut-off angle, and the spectrometer collection angle. Explain why large differences in the characteristic scattering angle affect the information in the spectrum.
T38.2 Why does filtering out the energy-loss electrons improve the quality of images of specimens showing mass-thickness contrast?
T38.3 Why does filtering out the energy-loss electrons improve the quality of images of specimens showing diffraction contrast?
T38.4 Why does filtering out the energy-loss electrons improve the quality of diffraction patterns?
T38.5 What is contrast tuning and under what circumstance might you use it?
T38.6 Why do you think there's a residual interband transition peak in the polyethylene spectrum in Figure 38.9?
T38.7 Can you think of any other way to image the distribution of Li shown in Figure 38.8 (Hint: read Chapter 39 first)?
T38.8 Why were we able to use plasmon-peak shift measurements as an analysis technique over 30 years ago and why does nobody use it any more?
T38.9 Why does EELS low-loss determination of the dielectric constant compare with UV spectroscopy in terms of the valence states that can be determined? (Hint: work out the wavelength of electrons with a typical low energy loss.)
T38.10 Given that we typically think of the band gap as a non-spatially localized phenomenon due to overlap of the energy states above the atomic potential wells, explain how we can talk about band-gap imaging and the high spatial resolution of images such as Figure 38.10.
T38.11 Why would we want to calculate the intensity distribution in low-loss spectra?
T38.12 Given that the low-loss spectrum is so much more intense than the high-loss spectrum, why has there been relatively little theoretical and experimental work on this part of the spectrum compared with the high-loss regime?
T38.13 Estimate the relative intensities in the zero-loss and the low-loss regions of Figure 38.1 and then explain why we can approximate the total spectrum intensity to the sum of these two components.
T38.14 Study Figure 38.7, then draw diagrams showing how the spectral peaks continue to change in relative intensity with increasing thickness beyond that in Figure 38.7B.
T38.15 Why would you expect to see differences in the low-loss spectra from different compounds such as shown in Figure 38.5?
T38.16 Why do plasmon-like peaks occur in spectra from biological materials in which there are no free electrons?

39

High Energy-Loss Spectra and Images

CHAPTER PREVIEW

The high energy-loss spectrum ($\mathcal{E} > 50$ eV) consists primarily of ionization or core-loss edges on a rapidly decreasing plural-scattering background. Elemental-composition data and elemental maps can be extracted from these ionization edges. In this chapter, we'll examine how to get this information, quantify it, and image it. A good use for such data is light-element analysis wherein EELS complements XEDS. First, we'll remind you of the experimental variables over which you have control, because these are rather critical. Then we'll discuss how to obtain a spectrum and what it should look like if you're going to quantify it. Next, we'll discuss the various quantification routines which, in principle, are just as straightforward as those for XEDS but in practice require rather more sophisticated software, and we often need to use deconvolution routines. Elemental imaging is a powerful aspect of high-energy-loss EELS, particularly, because both the spatial resolution and minimum detection limits are somewhat better than XEDS and atomic-column spectroscopy and single-atom detection are more easily achievable in EELS.

39.1 THE HIGH-LOSS SPECTRUM

The high-loss portion of the spectrum above ~50 eV contains information from inelastic interactions with the inner or core shells. These interactions provide direct elemental identification in a manner similar to XEDS and other information, such as bonding and atomic position. We'll emphasize quantitative elemental analysis and mapping in this chapter and discuss the other features of the high-loss spectrum in the next chapter.

39.1.A Inner-Shell Ionization

When a beam electron transfers sufficient energy to a core-shell electron (i.e., one in the inner, more tightly bound K, L, M, etc., shells) to move it outside the attractive field of the nucleus, the atom is said to be ionized (go back and look at Figure 4.2). As you know from the earlier chapters on X-ray analysis, the decay of the excited atom back to its ground state may produce a characteristic X-ray or an Auger electron. So high-loss EELS and XEDS detect different aspects of the same phenomenon. Ionization is the primary event and X-ray emission is one of two secondary processes. We are interested in ionization losses precisely because the process is characteristic of the atom involved and so the signal is a direct source of chemical information, just like the characteristic X-ray. We call the ionization-loss signal in the EELS spectrum an 'edge,' rather than a peak, for reasons we'll describe shortly and we use the edges as the basis for elemental mapping.

EELS COMPLEMENTS XEDS

Detection of the high-energy electron that ionized the atom is independent of whether the atom emits an X-ray or an Auger electron. So EELS is **not** affected by the fluorescence-yield limitation that restricts light-element X-ray analysis. These differences explain, in part, the complementary nature of XEDS and ionization-loss EELS but also the much higher efficiency of EELS.

Ionization is a relatively high-energy process. For example, the lightest solid element, Li, requires an input of ~55 eV to eject a K-shell electron, and so the loss electrons are usually found in the high-loss region of the spectrum, above $\mathcal{E} = 50$ eV. K-shell electrons require much more energy for ejection as Z increases because they are more strongly bound to the nucleus. The binding energy for electrons in the uranium K shell is about 99 keV. So we tend to use L and M edges when dealing with high-Z atoms (just like in XEDS) because the intensity of the K edges decreases substantially above ~2 keV. It's worth a short mention here about the nomenclature used

for EELS edges. As for X-rays, where we have K, L, M, etc., peaks in the spectrum, we get ionization edges from K, L, M, etc., shell electrons. However, the much better energy resolution of the magnetic-prism spectrometer means that it is much easier to detect small differences in spectra that arise from the presence of different energy states in the shell. For example

- The K shell electron is in the 1 s state and gives rise to a single K edge.
- In the L shell, the electrons are in either 2s or 2p orbitals, and if a 2s electron is ejected, we get an L_1 edge and a 2p electron causes either an L_2 or L_3 edge.

Depending on the ionization energy, the L_2 and L_3 edges may not be resolvable (they aren't in Al but they are in Ti), and so we call this edge the $L_{2,3}$. The full range of possible edges is shown schematically in Figure 39.1, and you can see that other dual edges exist, such as the $M_{4,5}$.

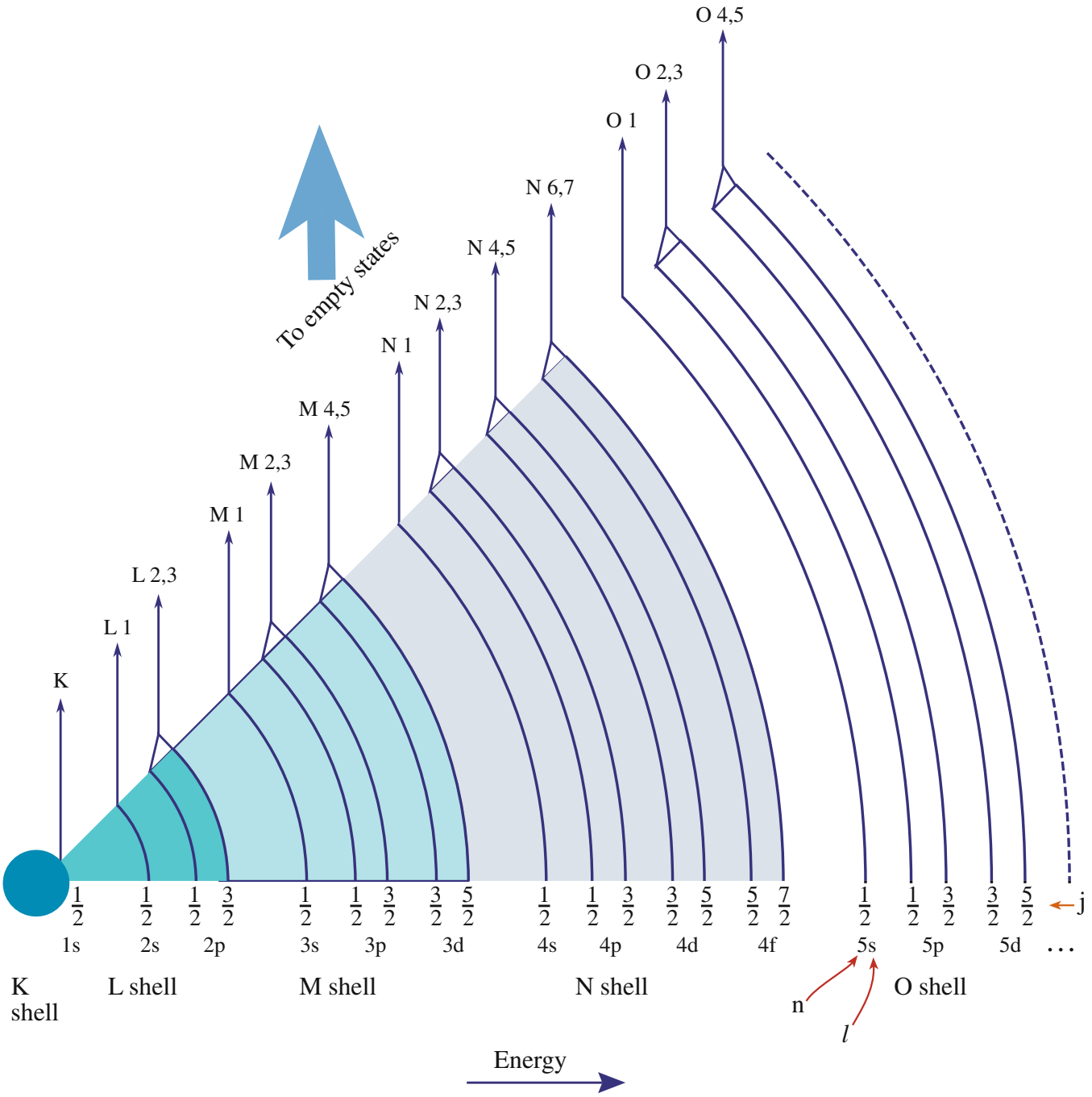


FIGURE 39.1. The full range of possible edges in the energy-loss spectrum due to core-shell ionization and the associated nomenclature.

Compared with plasmon excitation, which requires much less energy, the ionization cross sections are relatively small and the mean free paths relatively large. As a result, the ionization-edge intensity in the spectrum is much smaller than the plasmon peak, and becomes even smaller as the energy loss increases (look back to Figure 37.1). This is another reason for staying with the lower energy-loss (L and M) core edges. While the possibility of plural ionization events being triggered by the same electron is small in a typical thin foil, we'll see that the combination of an ionization loss and a plasmon loss is by no means uncommon. This phenomenon distorts the EEL spectrum and any filtered images.

If you go back and look at Figure 4.2, you can see that a specific minimum-energy transfer from the beam electron to the inner-shell electron is required to overcome the binding energy of the electron to the nucleus and ionize the atom. This minimum energy constitutes the ionization threshold, or the critical ionization energy, E_c.

We define E_C as E_K for a particular K-shell electron, E_L for an L shell, etc. Of course, it is also possible to ionize an atom by the transfer of $\mathcal{E} > E_C$. However, the probability of ionization occurring becomes less with increasing energy above E_C because the ionization cross section decreases with increasing energy transfer. As a result, the ionization-loss electrons have an energy distribution that ideally shows a sharp rise to a maximum at E_C, followed by a slowly decreasing intensity above E_C back toward the background. This triangular shape is called an edge because you'll notice that, as shown in Figure 39.2A, it has a similar intensity profile to the absorption edges in X-ray spectroscopy. Often the term 'hydrogenic' is used for such a sharp edge-onset because this is what would arise from the ionization of the ideal single isolated hydrogen atom.

TRIANGULAR SHAPE

This idealized triangular or sawtooth shape is only found in spectra from isolated hydrogen atoms and is therefore called a hydrogenic ionization edge. Real ionization edges have shapes that approximate, more or less, to the hydrogenic edge.

In reality, because we aren't dealing with isolated atoms, but atoms integrated into a crystal lattice or an amorphous structure, the spectra are more complex. Ionization edges are superimposed on a rapidly decreasing background intensity from electrons that have undergone random, plural inelastic scattering events (Figure 39.2B). The edge may also show fine structure oscillations within ~50 eV of E_C (Figure 39.2C) which are due to bonding effects (termed energy-loss near-edge structure, ELNES). More than 50 eV after the edge, small intensity oscillations may be detectable (Figure 39.2D) due to diffraction effects from the atoms surrounding the ionized atom, and these are called extended energy-loss fine structure (EXELFS), which is analogous to extended X-ray absorption fine structure (EXAFS) in X-ray spectra, particularly, those generated from intense synchrotron sources.

ELNES AND EXELFS

Fine structure around the ionization edge onset is known as ELNES. Small intensity oscillations >~50 eV after the edge due to diffraction effects are called EXELFS.

Finally, as we noted earlier, the ionization-loss electrons may also undergo low-loss interactions. For example, they may create plasmons, in which case the edge contains extra plural-scattering intensity ~15–25 eV above E_C, as shown schematically in Figure 39.2E. So experimental ionization edges are far more complicated than the Gaussian peaks in an XEDS spectrum, but they also contain far more information about the specimen than a characteristic peak. From an XEDS spectrum, you only get *elemental* identification rather than *chemical* information, such as bonding which is contained both in the ELNES and the low-loss structure (although as we showed in Figure 32.9C, if the X-ray spectrometer has sufficiently high-energy resolution, it can detect such differences but the price to pay in the AEM is an unacceptably low count rate). Figure 39.3 shows a spectrum from BN on a C film. The various ionization edges show some of the features drawn schematically in Figure 39.2, in particular, strong ELNES on the B-K edge; we'll discuss these fine-structure effects more in Section 40.1 and how to form fine-structure images from them in Section 40.5.C.

39.1.B Ionization-Edge Characteristics

The angular distribution of ionization-loss electrons varies as $(\theta^2 + \theta_E^2)^{-1}$ and will be a maximum when $\theta = 0$, in the forward-scattered direction. The distribution decreases to a half width at the characteristic scattering angle θ_E given by equation 38.1. This behavior is essentially the same as for plasmon scattering but because we have relatively large values of E_C compared to E_P, we get larger characteristic scattering angles for ionization-loss electrons (e.g., for the typical, maximum core-loss energy that we would use for analysis $\mathcal{E} = 2000$ eV so $\theta_E \sim 10$ mrad when $E_0 = 100$ keV).

The angular distribution varies depending on $\mathcal{E}$, and because of the extended energy range of ionization-loss

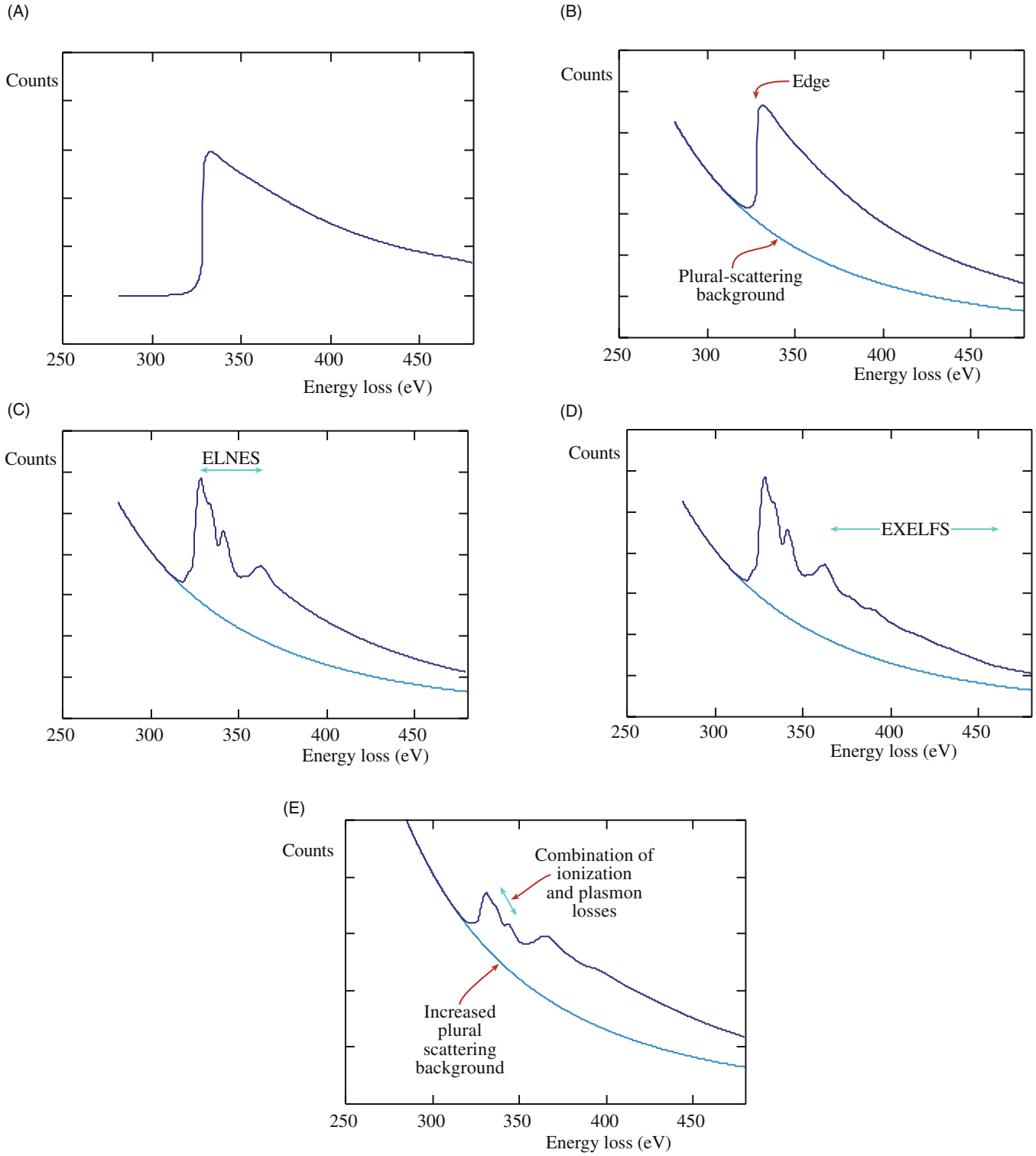


FIGURE 39.2. The characteristic features of an inner-shell ionization edge. (A) The idealized sawtooth (hydrogenic) edge. (B) The hydrogenic edge superimposed on the background arising from plural inelastic scattering. (C) The presence of ELNES. (D) The EXELFS. (E) In a thick specimen, plural scattering, such as the combination of ionization and plasmon losses, adds another peak to the post-edge structure and raises the background level.

electrons above E_C, this can be quite complicated. For $\mathcal{E} \sim E_C$ the scattering intensity drops rapidly to zero over about 10 mrads (at θ_e), but as $\mathcal{E}$ increases above E_C the angular-intensity distribution drops around $\theta = 0°$, but increases at larger scattering angles, giving rise to the so-called Bethe ridge. However, this effect is not really important for the kind of analytical studies that we are emphasizing in this chapter.

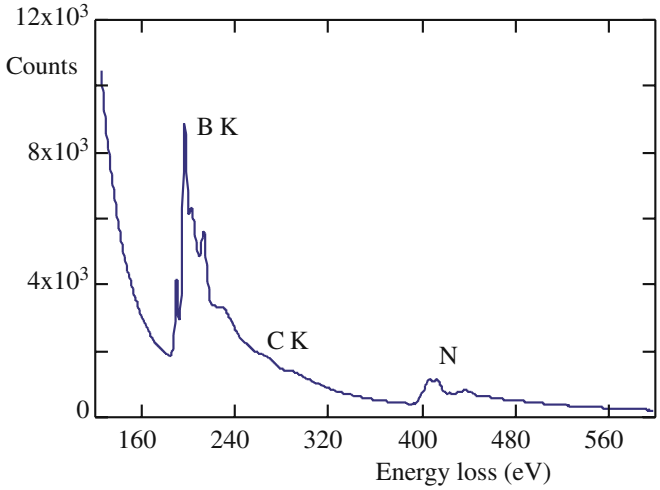


FIGURE 39.3. High energy-loss spectrum from a thin flake of BN sitting on the edge of a hole in an amorphous-carbon support film. The B K and N K edges are clearly visible superimposed on a rapidly decreasing background. A very small C K edge is also detectable at ~280 eV.

So, the distribution of characteristic scattering angles for the core-loss electrons that we use for analysis span the range from ~0.2 to 10 mrads and the scattering cut-off angles range from ~25 to 200 mrads (equation 38.4). In other words, like the plasmon-loss electrons, the ionization-loss electrons are very strongly forward-scattered. Consequently, efficient collection of most inelastically scattered electrons is straightforward, since a spectrometer entrance aperture angle (β) of 10 mrads will collect the great majority of such electrons. As a result, collection efficiencies in the range 50–100% are not unreasonable, which contrasts with XEDS where the isotropic generation of characteristic X-rays results in very inefficient collection. Figure 39.4 compares the collection of X-rays and energy-loss electrons and Figure 39.5 shows the variation in collection efficiency for ionization-loss electrons as a function of both β and energy.

While the K edges in Figure 39.3 show sharp onsets, like an ideal hydrogenic edge, not all edges are similar in shape. Some edges have much broader onsets, spread over several eV or even tens of eV. The edge shape in general depends on the electronic structure of the atom but, unfortunately, we can't give a simple relationship between edge types and specific shapes. The situation is further complicated because the edge shapes change depending on whether or not certain energy states are filled or unfilled. For example, if you go back and look at Figure 37.1, the Ni L edge shows two sharp peaks, which are the L_3 and L_2 edges. (We'll discuss these details much more in Section 40.1.) These sharp lines arise because the ejected L shell electrons don't entirely escape from the atom and have a very high probability of ending up in unfilled d-band states. In contrast, in Cu where the d band is full, the $L_{2,3}$ edge does not show these intense lines. Similar sharp lines appear in the $M_{4,5}$ edges in the rare earths. As if this were not enough, the

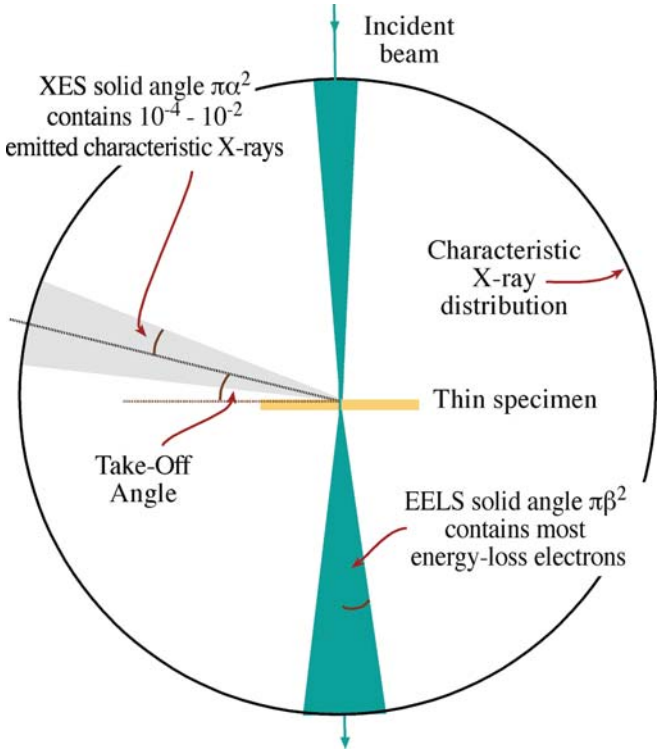


FIGURE 39.4. Comparison of the relative efficiencies of collection of EELS and XES. The forward-scattered energy-loss electrons are very efficiently collected with even a small EELS collection angle. In contrast, only a small fraction of the uniformly emitted (4π sr) characteristic X-rays is detected by the XEDS.

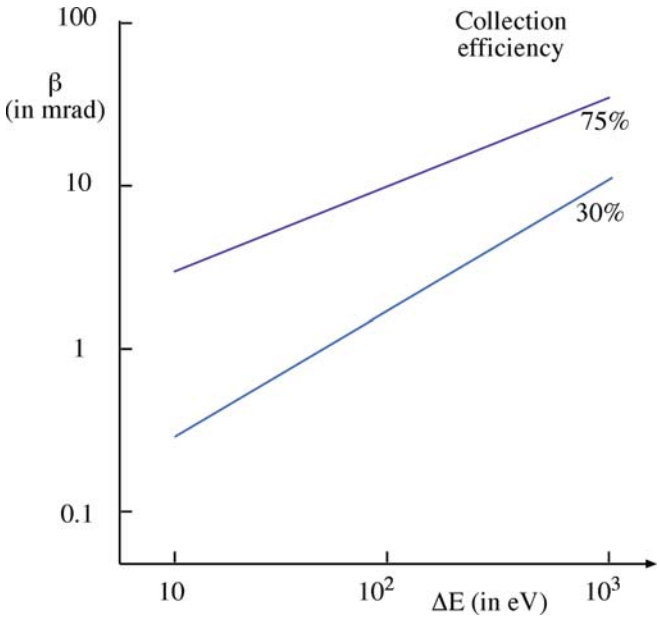


FIGURE 39.5. Variation in the collection efficiency of ionization-loss electrons as a function of energy loss and spectrometer collection angle. A 10 mrad collection angle will gather over 75% of all the incident-beam electrons that ionized C atoms and lost ~285 eV.

details of the fine structure and edge shapes are also affected by bonding. For example, the Si edge in a spectrum from SiO_2 is different from the Si edge from pure Si. To sort all this out it's best if you consult the 2004 EELS Atlas (by Ahn and related references in the previous chapter) which contains representative edges from all the elements and many oxides.

Now that we've covered both the low and high energy-loss processes, we can summarize the characteristics of the energy-loss spectrum by examining a complete spectrum from NiO containing both low and high-loss electrons, as shown in Figure 39.6. In this figure, we also compare the spectrum to the energy-level diagram for NiO. You can see that

- The ZLP is above the potential wells since these electrons don't interact with the atom.
- The plasmon peak comes from interactions with the valence/conduction band electrons just below the Fermi level (E_F).

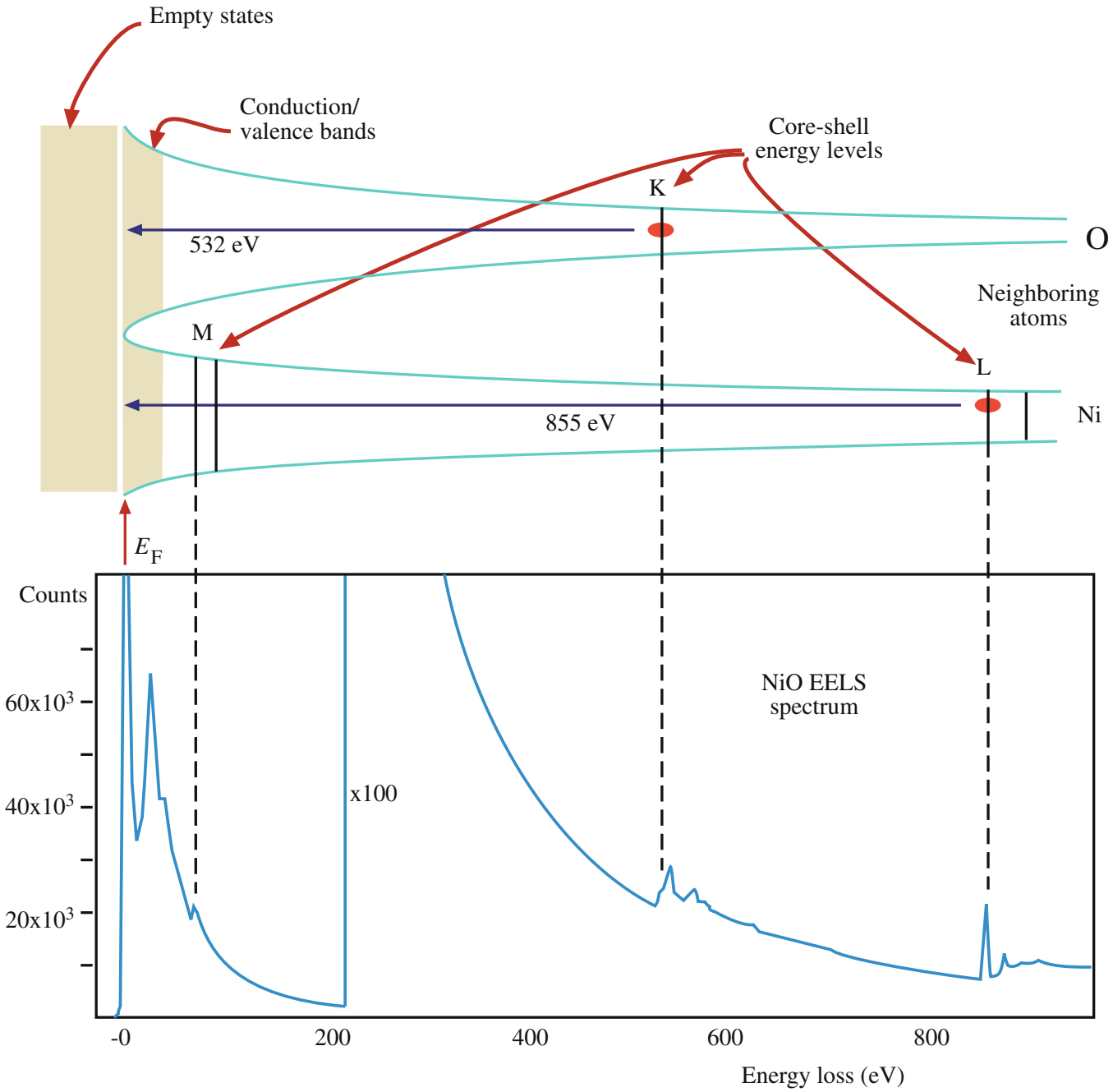


FIGURE 39.6. The correspondence between the energy levels of electrons surrounding adjacent Ni and O atoms and the energy-loss spectrum. The deeper the electrons sit in the potential well the more the energy needed to eject them. The ZLP is above the Fermi energy E_F, the plasmon peak is shown at the energy level of the conduction/valence bands where plasmon oscillations occur in the loosely bound electrons. The critical ionization energy required to eject electrons in specific shells is shown (Ni L: 855 eV and O K: 532 eV).

- The relative energy levels of the ionized shell (K, L, or M) control the position of the ionization edge in the spectrum. The closer to the nucleus, the deeper the potential well and the more the energy required to eject the electron.
- There will be a different density of states in the valence (3d) band of the Ni atom compared to the s band of the O atoms at the top of the potential wells.
- The core electrons could also be given enough energy to travel into the empty states, well above E_F and, in this case, we see ELNES after the ionization edge. We'll discuss more details of such fine structure in the spectrum in Chapter 40.

Despite the very high collection efficiency of the spectrometer, the ionization edges still show relatively low intensity, particularly as $\mathcal{E}$ increases. The edges have an extended energy range well above E_C and ride on a rapidly varying, relatively high background. All these factors, as we shall see, combine to make quantitative analysis using EELS a little more challenging than XEDS. However, for the lighter elements the X-ray fluorescence yield drops to such low values, and absorption becomes so strong, even in thin specimens, that EELS is the preferred technique. Experimentally, the choice between the two is not always simple, but below oxygen in the periodic table, EELS shows better performance than XEDS and for elements below boron, there is no sensible alternative to EELS for nanometer-scale analysis.

39.2 ACQUIRING A HIGH-LOSS SPECTRUM

From what we've described about the various EEL spectrometers and filters and the complexity of the spectra, it should be clear that there are many variables to control when acquiring a spectrum (see Brydson's monograph for a detailed description). Computer control via the Gatan software now makes this process very straightforward. We'll start by summarizing the major parameters relevant to acquiring high-loss spectra and images and indicate reasonable values for each parameter.

- *Beam Energy* E_0: It's best to use the highest E_0, unless doing so causes displacement damage or significant surface sputtering. A higher E_0 reduces the scattering cross section and so you get reduced edge intensity. Conversely, as E_0 increases, the plural-scattering background intensity falls faster than the edge intensity and so the signal to background increases. The increase in signal to background varies with the particular edge but it is never a strong variation; so while we recommend using the highest kV, it's not a good reason (on its own) to justify purchasing a 300-keV TEM.
- *Convergence angle* α: You know how to control α with the C2 aperture and/or the C2 lens, but α is only important in quantification if it is larger than β. So if you operate in TEM image or diffraction mode with a broad, parallel beam, rather than STEM mode, you can ignore α; otherwise, use the correction factor we give later in Section 39.7.
- *Beam size and current*: You control these factors by your choice of electron source, C1 lens, and C2 aperture. As usual, the beam size is important in determining spatial resolution in STEM mode, and the beam current controls the signal intensity.
- *Specimen thickness*: The specimen must be thin because this minimizes plural-scattering contributions to the spectrum and quantification is more straightforward.

SPECIMEN THICKNESS
Making your specimen as thin as possible is the most important part of EELS.

If your specimen is too thick then you'll have to use deconvolution procedures to remove the effects of plural scattering. So we'll tell you more about how to determine your specimen thickness from the spectrum and how to decide if you need to deconvolute the spectrum.

- *Collection angle* β: You know from Section 37.4 how to measure β in all operating modes. If you need lots of intensity and are happy with poor spatial resolution, use TEM image mode with no objective aperture ($\beta > \sim 100$ mrads). A small spectrometer entrance aperture ensures good energy resolution at the same time. If you want a small β to prevent contributions to the spectrum from high-angle scattering, use diffraction mode (TEM or STEM) and keep the small entrance aperture for good energy resolution. In the STEM case you also get good spatial resolution.

ENTRANCE APERTURE
Remember that a 5 mm diameter entrance aperture gives $\beta \sim 5$ mrad at a camera length of ~ 800 mm.

Generally, for analysis $\beta \sim 1$–10 mrads is fine, so long as it's less than the Bragg angle for your specimen orientation; but for EELS imaging, which we discuss in Section 39.9, 100 mrads may be necessary to get the necessary signal intensity.

- *Energy resolution*: ΔE is limited by your electron source unless you have a monochromator. Elemental analysis and imaging (the topic of this chapter) do not require the best ΔE, so ~5 eV would suffice. You really need the best ΔE for low-loss and fine-structure studies, which are probably the most useful and widespread aspect of EELS (see the surrounding chapters). Use an FEG source and a PEELS/imaging filter if you want to do this, especially if you're lucky enough to have access to a monochromator.
- *Energy-loss range and spectrum dispersion*: The full spectrum extends out to the beam energy E_0, but the useful portion only extends to ~2 keV. Above this $\mathcal{E}$, the intensity is very low, and XEDS is both easier and more accurate. Since you rarely need to collect a spectrum above ~2 keV a minimum of 2048 channels in the computer display, giving 1 eV/channel is a good starting dispersion. You can easily select a higher display resolution if you want to look at a more limited region of the spectrum or if you want to see detail with ΔE <1 eV. Typically, you're only examining a limited portion of the spectrum anyhow and you set this by putting the necessary voltage on the drift tube or changing the high voltage.
- *Dwell time*: If you have a PEELS with a PDA, set the integration time so that at the maximum intensity you don't saturate the diodes: i.e., stay below 16,000 counts per acquisition in the most intense channel and sum as many spectra as you need to give sufficient counts for analysis.
- *Number of acquisitions*: Again, if you have PEELS/PDA, multiple acquisitions may be necessary to get sufficient counts in the edge, but remember that multiple acquisitions may give rise to minor artifacts, as we discussed in Section 37.5.

Before you analyze a particular spectrum, you should check four things

- Focus and align the ZLP and check the spectrometer resolution.
- Look at the low-loss (plasmon) portion of the spectrum; this gives you an idea of your specimen thickness.
- Look for the expected ionization edges. If you can't see any edges, your specimen is probably too thick or you need to raise the display gain.
- It's probably worth deconvoluting out the PSF prior to any quantification.

The first of these tasks is not critical, as we noted earlier. Regarding the second task, we noted back in Chapter 38 that, to a first approximation, if the plasmon-peak intensity is less than about one tenth the ZLP, then the specimen is thin enough for analysis. Otherwise, you'll probably have to deconvolute plural-scattering effects from your experimental spectrum. For the third task, you should ideally see discrete edges on a smoothly varying background, but you need to see at least a change in slope in the background intensity at the expected E_C. If the background intensity is too noisy it will make quantification more difficult, so acquire sufficient counts to generate a smoothly varying background.

THE JUMP RATIO

An important parameter in determining the quality of your spectrum is the signal-to-background ratio which in EELS we call the jump ratio.

The jump ratio is the ratio of the maximum edge intensity (I_{max}) to the minimum intensity (I_{min}) in the channel preceding the edge onset, as shown in Figure 39.7 (which is a well-defined edge from a suitably thin film of amorphous carbon). If the jump ratio is above ~5, for the carbon K edge at 284 eV from a 50 nm carbon film at 100 kV, then your system is operating satisfactorily. You should keep a standard thin, amorphous-carbon film available as a standard reference specimen and occasionally check that the jump ratio remains the same. We'll see that jump-ratio imaging is one method of acquiring filtered images from ionization edges. The jump ratio increases as E_0 increases. If you can't get such a jump ratio from a standard, thin, carbon film, then probably you need to realign the spectrometer. The actual ionization-loss edges from your real specimen, that you may wish to quantify or use to form images, will probably be nothing like this ideal edge, but the EELS software programs are more than capable of

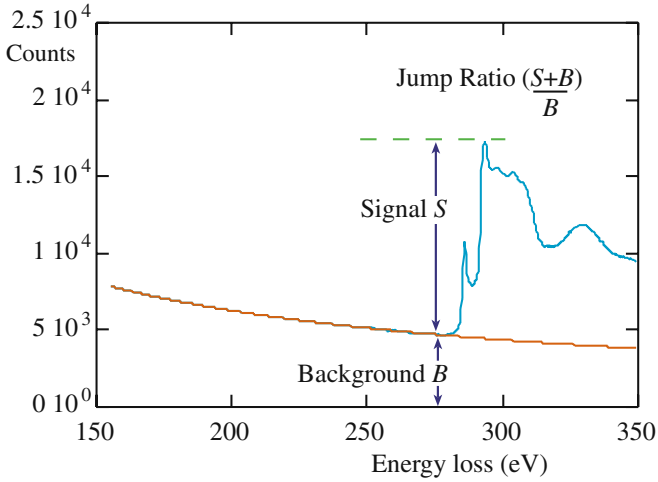


FIGURE 39.7. Definition of the jump ratio of an ionization edge which should be about 5–10 for the carbon K edge if the EELS is well aligned and the specimen really thin. This spectrum shows an adequate jump ratio.

handling much smaller edges riding on much higher backgrounds.

39.3 QUALITATIVE ANALYSIS

As with XEDS, you should always carry out a qualitative analysis first to ensure that you have identified all the features in your spectrum. Then you can decide which edges to use for subsequent quantitative analysis and imaging.

Qualitative analysis using ionization edges is very straightforward. Unlike XEDS, there are actually very few artifacts that can be mistaken for an edge. The most prominent artifact that may lead to misidentification is the ghost peak from diode saturation (see Section 37.5) which is easily removed. So long as you calibrate the spectrum to within a few eV you can unambiguously identify the edge energy.

IONIZATION EDGE

We identify the ionization edge as the energy loss at which there is a discrete increase in the slope of the spectrum; this value is the edge onset, i.e., E_C, the critical ionization energy.

You have to be careful here: sometimes you'll see the edge energy defined somewhat arbitrarily halfway up the edge, e.g., at the π* peak on the front of a C-K edge. There is no strict convention, and very often L and M edges do not have sharp onsets anyhow. Examination of a portion of a spectrum, such as that shown back in Figure 39.3, is usually sufficient to let you draw a definite conclusion about the identity of the specimen, which in this case is BN on a C support film. In addition, it is wise to compare your spectrum with reference spectra from the EELS Atlas that we've mentioned several times before or through an on-line database, such as URL #1.

Remember that there are families of edges (K, $L_{2,3}$, $M_{4,5}$, etc.) just as there are families of peaks in X-ray spectra but, as with X-ray spectra, you might not be able to resolve all the edges in a single family. Given that above ~2 keV the edges are usually too small to be detected, it is in fact very rare that you would expect to see more than one family of lines from a given element (the Si L edge at ~100 eV and the Si K edge at 1.7 keV should both be visible in the same spectrum). As a rule of thumb, quantification is equally easy with K and L edges, but the accuracy of K-edge quantification is slightly better. Up to $Z = 13$ (Al) we usually use K edges because any L edges occur at very low energy and are masked by the plasmon peak. Above $Z = 13$ you can use either K or L edges. Sometimes there is the question of which edge is most visible. The K edge onset is generally a bit sharper than the L edge which consists of both the L_2 and L_3 edges and so may be somewhat broader, but this is not always the case.

L edges for $Z = 19$–28 (e.g., the Ca-L edge in Figure 37.12 and the Cr-L edge in Figure 39.13) and $Z = 37$–45 are characterized by intense near-edge structure called white lines. M edges for $Z = 55$–69 have similar intense lines.

These white lines are so named because they appeared as lines of varying intensity in photographically recorded, energy-loss spectra; they also appear that way if you look at the spectra from in-column filters (see Figure 37.14A). More details will be given in Section 40.1. If you have to use the M, N, or O edges without any white lines, you should know that they are very broad, with an ill-defined threshold, and quantification is best achieved with standards, as we'll see shortly.

The energy-loss spectrum clearly does not lend itself to a quick 'semi-quantitative' analysis; so we can't follow our XEDS approach. For example, the spectrum in Figure 39.3 comes from equal numbers of B and N atoms, but the intensities in the B and N edges are markedly different. This difference arises because of the variation in ionization cross section with $\mathcal{E}$, the strongly varying nature of the plural-scattering background, and the edge shape, which causes the C and N K edges to ride on the tails of the preceding edge(s).

The Ti-nitride and Ti-carbide example: Sometimes qualitative analysis is often all that you need to do. Figure 39.8 shows images and spectra from two small precipitates in an alloy steel. The spectra show a Ti L_{23} edge in both cases and C and N K edges in Figure 39.8A and B, respectively. It does not take much effort to deduce that the first particle is TiC because it is the only known carbide of Ti, but the nitride could be either TiN or Ti_3N. To determine which of the two it is, you have to carry out full quantification, which we'll discuss shortly. You should note that such clear discrimination between TiC and TiN in Figure 39.8B would be difficult using windowless XEDS because the energy resolution is close to the separation of the Ti L ($\mathcal{E}$ = 452 eV) and the N K ($\mathcal{E}$ = 392 eV) X-ray peaks. In addition, the DPs from both phases are almost identical, so this problem is a perfect one for EELS.

39.4 QUANTITATIVE ANALYSIS

To quantify the spectrum or to form a quantitative image, you have to integrate the intensity (I) in the ionization edge(s) by removing the plural-scattering background. Then you have to determine the number of atoms (N) responsible for I. N is related to I by a sensitivity factor termed the partial ionization cross section (σ). We'll see that σ plays a similar role to the k_{AB} factor in X-ray analysis. If you go back and look at Figure 39.2,

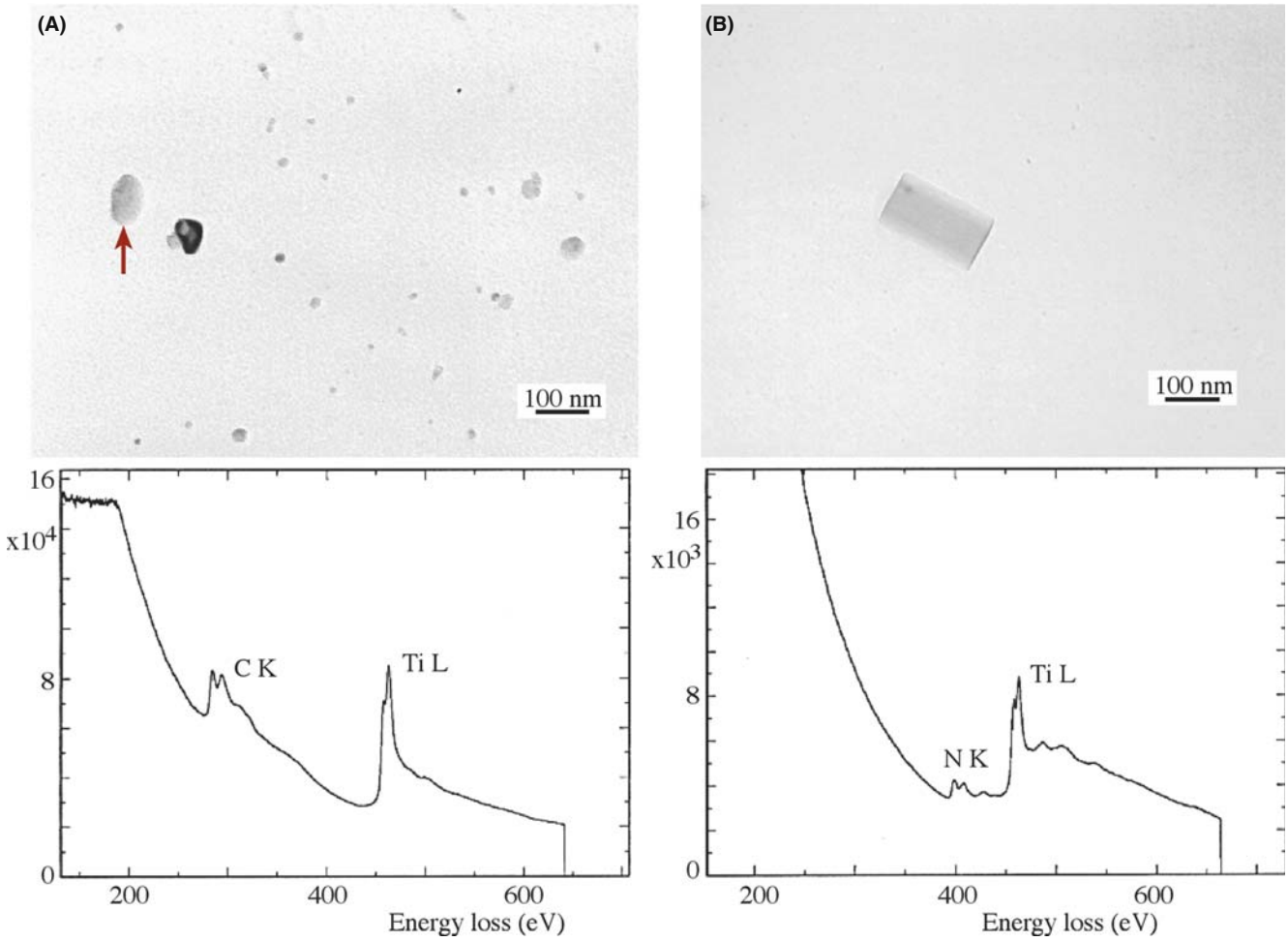


FIGURE 39.8. Images of small precipitates on an extraction replica from a stainless steel specimen, and the corresponding ionization edges showing qualitatively the presence of Ti, C, and N. Thus the precipitates can be identified as (A) TiC and (B) TiN, respectively.

you'll see how an ionization edge is built up from several contributions. The process of quantification in essence involves stripping away (or ignoring) the various contributions until you're left with Figure 39.2A, which contains the single-scattering or hydrogenic-edge intensity.

39.4.A Derivation of the Equations for Quantification

The equations we use for quantitative analysis and imaging have been derived, refined, and applied by Egerton and co-workers. The following derivation is a summary of the full treatment by Egerton given in his textbook.

We'll assume that we are quantifying a K edge, although the basic approach can be used for all edges. The K-edge intensity above background, I_K, is related to the probability of ionization, P_K, and the total transmitted intensity, I_T, by

$$I_K = P_K I_T \tag{39.1}$$

This equation assumes that the intensities are measured over the complete angular range (0–4π sr), which of course, is not the case, but we'll correct this later. In a good thin specimen we can approximate I_T to the incident intensity, neglecting backscatter and absorption effects. Now, this is the *important* point: if we assume also that the electrons contributing to the edge have only undergone a single ionization event, then we can easily obtain an expression for P_k.

$$P_K = N_{\sigma K} \exp\left(-\frac{t}{\lambda_K}\right) \tag{39.2}$$

where N is the number of atoms *per unit area* of the specimen (of thickness t) that contribute to the K edge. The assumption of a single ionization (i.e., scattering) event is reasonable, given the large mean free path (λ_k) for ionization losses; and it explains why you have to make thin specimens. Assuming single scattering also means that the exponential term is very close to unity and so

$$I_K \approx N\sigma_K I_T \tag{39.3}$$

and therefore

$$N = \frac{I_K}{\sigma_K I_T} \tag{39.4}$$

Thus, we can measure the absolute number of atoms per unit area of the specimen simply by measuring the intensity above background in the K edge and dividing it by the total intensity in the spectrum multiplied by the ionization cross section. We can easily extend this expression to two edges from elements A and B, in which case the I_T drops out and we can write

$$\frac{N_A}{N_B} = \frac{I_K^A \sigma_K^B}{I_K^B \sigma_K^A} \tag{39.5}$$

Similar expressions apply to L, M edges, etc., and combinations of edges can be used. So you see that if you are quantifying more than one element, then you don't need to gather the ZLP, which saves hitting the PDA or CCD with this high-intensity signal.

In both equations 39.4 and 39.5 we assumed that we could accurately subtract the background under the edge and that we know σ. Unfortunately, as you'll see, both background subtraction and determination of σ are non-trivial. We will discuss these points later, but first we must take account of the practical realities of spectrum acquisition, and modify the equations accordingly.

- First, you can't gather the whole of the spectrum out to the beam energy, E_0, because above ~2 keV the intensity decreases to a level close to the system noise.
- Second, while ionization-loss electrons can theoretically have any energy between E_C and E_0, in practice the intensity in the edge falls to the background level within about 100 eV of the threshold, E_C.
- Third, the background-extrapolation process becomes increasingly inaccurate beyond ~100 eV.

For all these reasons, it is imperative to restrict integration of the edge intensities to some window, Δ, usually in the range 20–100 eV. So we modify equation 39.4 to give

$$I_K(\Delta) = N\sigma_K(\Delta) I_T(\Delta) \tag{39.6}$$

The term $I_T(\Delta)$ is more correctly written as $I_l(\Delta)$ where I_l is the intensity of the zero-loss (direct beam) electrons combined with the low-loss electrons over an energy-loss window Δ. Only if we have true single scattering can we use I_T and we'll discuss the conditions for this later.

As we discussed, EELS has the tremendous advantage that the energy-loss electrons are predominantly forward scattered and so you can easily gather most of the signal. So because we never manage to collect the full angular range of energy-loss electrons, we must further modify the equation by including the collection angle β and write

$$I_K(\beta\Delta) = N\sigma_K(\beta\Delta) I_l(\beta\Delta) \tag{39.7}$$

This factor $\sigma_K(\beta\Delta)$ is the partial ionization cross section.

From this equation therefore, the absolute quantification for N is given by

$$N = \frac{I_K(\beta\Delta)}{I_l(\beta\Delta)\sigma_K(\beta\Delta)} \tag{39.8}$$

For a ratio of two elements A and B, the low-loss intensity drops out again as in equation 39.5

$$\frac{N_A}{N_B} = \frac{I_K^A(\beta\Delta)\, \sigma_K^B(\beta\Delta)}{I_K B(\beta\Delta)\sigma_K^A(\beta\Delta)} \tag{39.9}$$

We can draw a direct analogy between this equation and the Cliff-Lorimer expression (equation 35.2) used in thin-film XEDS. In both cases, the composition ratio C_A/C_B or N_A/N_B is related to the intensity ratio I_A/I_B through a sensitivity factor which we call the k_{AB} factor in XEDS and which in EELS is the ratio of two partial cross sections, σ^B/σ^A.

Remember that the major assumption in this whole approach is that the electrons undergo *a single-scattering event*. In practice, it's difficult to avoid some plural scattering, although in very thin specimens the approximation remains valid, if errors of ±10–20% are acceptable. If plural scattering is significant, then it must be removed by deconvolution, which we'll discuss in Section 39.6. You should also note when using the ratio equation your analysis is a lot better if the two edges are similar in shape, i.e., both K edges, or both L edges, otherwise the approximations inherent in equation 39.9 will be less accurate.

In summary, equations 39.8 and 39.9 give us, respectively, an absolute value of the number of atoms/unit area of the specimen or a ratio of the number of atoms of the elements A and B either at a given analysis point or within a filtered image. To get this information experimentally, you have to carry out two essential steps

- Background subtraction to obtain I_K (and hence, N) for each element A, B, etc.
- Determination of the partial ionization cross section $\sigma_K(\beta\Delta)$ to get the ratio N_A/N_B.

So again, you can see why it is important to know β.

39.4.B Background Subtraction

The background is a rapidly changing continuum decreasing from a maximum intensity just after the plasmon peak at about 15–25 eV, down to a minimum at which it is indistinguishable from the system noise, typically at $\mathcal{E} > \sim 2$ keV. In addition to plural scattering, there is also the possibility of single-scattering contributions to the background from the tails of preceding ionization edges and perhaps contributions from the spectrometer itself. Because of the complexity of these contributions, it has not been possible to model the background from first principles, as is possible in XEDS using Kramers' law.

Despite the complexity of the various contributions to the background, the methods for subtraction are relatively simple. There are two ways commonly used to remove the background

- Curve fitting.
- Using difference spectra.

Curve Fitting: You select a window δ in the background before the edge onset and fit a curve to the intensity in the window. Then you extrapolate the curve over another window Δ under the edge. This process is shown schematically in Figure 39.9, and experimentally in Figure 39.10.

We assume that the energy dependence of the background has the form

$$I = A\mathcal{E}^{-r} \tag{39.10}$$

where I is the intensity in the channel of energy loss $\mathcal{E}$, and A and r are constants. The fitting parameters are only valid over a limited energy range because they depend on $\mathcal{E}$. The exponent r is typically in the range 2–5, but A can vary tremendously. We can see some trends in how r varies. The value of r decreases as

- The specimen thickness, t, increases.
- The collection angle, β, increases.
- The electron energy loss, $\mathcal{E}$, increases.

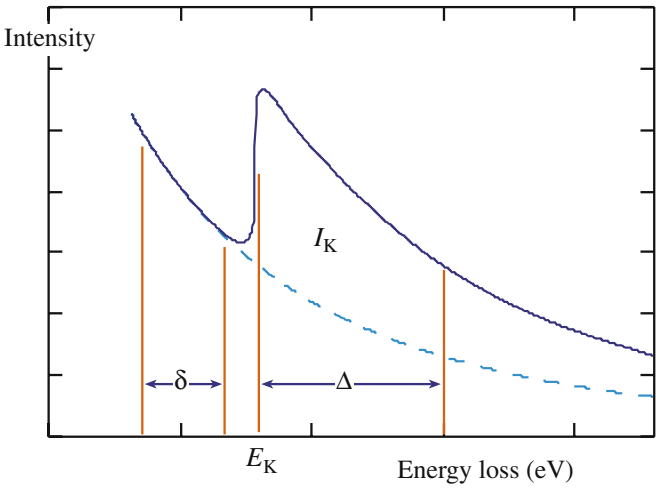


FIGURE 39.9. The parameters required for background extrapolation and subtraction under an ionization edge. The pre-edge fitting window δ is extrapolated over a post-edge window Δ then the intensity under the extrapolated line is subtracted from the total intensity in the window Δ to give the desired edge intensity I_K.

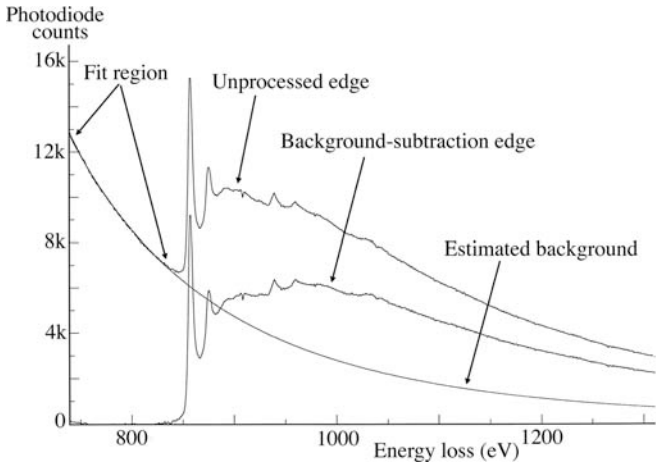


FIGURE 39.10. Comparison of an experimental Ni $L_{2,3}$ edge before and after background subtraction. The fitted region before the unprocessed edge is extrapolated to give the estimated background which is then subtracted leaving the (total) edge intensity (note: there is no edge window Δ shown here).

The fit of the curve to the tail of a preceding edge shows a similar power-law dependence to the background, and may be fitted similarly, i.e., $I = B\mathcal{E}^{-S}$. The fitting window δ should not be <10 channels and should not be >30% of E_K. In practice, however, you might not be able to fit the background over such a wide window if another edge is present within that range, which limits the goodness of fit of the curve.

You should choose the extrapolation window, Δ, such that the ratio of the finish to the start energies, $\mathcal{E}$(finish)/$\mathcal{E}$(start), is < 1.5; so Δ is smaller for lower edge energies. Using larger windows, although improving the statistics of the edge intensity, eventually reduces the accuracy of the quantification because the fitting parameters A and r are only valid over ~100 eV. If there's a lot of ELNES, either use a larger Δ to minimize its effect or avoid it in the extrapolation window unless the quantification routine can handle it.

Instead of the simple power-law fit, you can in fact use any expression, such as an exponential, polynomial, or log-polynomial, so long as it provides a good fit to the background and gives acceptable answers for known specimens. Polynomial expressions can behave erratically if you extrapolate them over a large Δ, so use them cautiously. Generally, the power law seems adequate for most purposes except close to the plasmon peaks ($\mathcal{E}$ <~100 eV). Clearly, the background channels closest

to the edge onset will influence the extrapolation most strongly, and various weighting schemes have been proposed. A noisy spectrum will be particularly susceptible to poor fitting, unless some type of weighting is used.

FITTING WINDOWS

There are two fitting windows: δ before the edge and Δ after the edge. Each has constraints for good background fitting.

We can judge the goodness of fit of a particular power-law expression qualitatively by looking at the extrapolation to ensure that it is heading toward the post-edge background and not substantially under- or over-cutting the spectrum. More quantitatively, we can assign a χ^2 value based on a linear least-squares fit to the experimental spectrum. The least squares fit can be conveniently tied in with a weighting scheme using the expression

$$\chi^2 = \sum_i \frac{(y - y_i)^2}{y^2} \qquad (39.11)$$

where y_i is the number of counts in the ith channel and $y = \ln_e I$. (Look back at Section 35.3.B.) The squared term in the denominator ensures suitable weighting of the channels close to the edge. Alternatively, the Gatan software includes 'smart' feedback which forces the background extrapolation to merge with the experimental background intensity well after the edge.

Difference Spectra: You can also remove the background using a first-difference approach (which is equivalent to differentiating the spectra). This method is particularly suited to PEELS since it simply involves taking two spectra, offset in energy by a few eV, and subtracting one from the other. As shown in Figure 39.11, the difference process results in the slowly varying background being reduced to zero and the rapidly varying ionization edge intensity showing up as classic difference peaks, similar to what you may have seen in Auger spectra. This is the *only* way to remove the background if your specimen thickness changes over the area of analysis and it also has the advantage that it suppresses spectral artifacts common to PEELS, particularly the channel-to-channel gain variation.

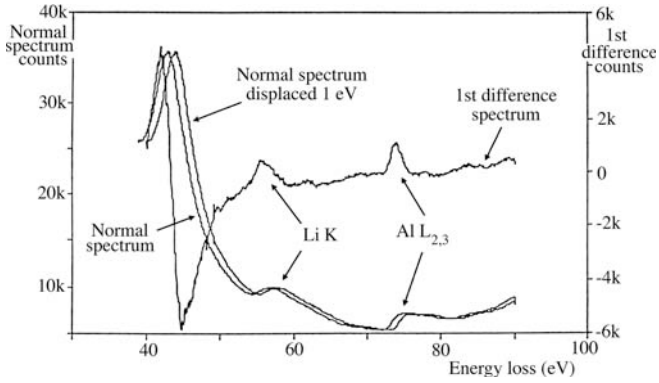


FIGURE 39.11. First-difference method of background subtraction, showing two PEELS spectra from a specimen of Al Li displaced by 1 eV and subtracted from one another to give a (first-difference) spectrum in which the background intensity falls to a straight line of close to zero counts and the small Li K and Al $L_{2,3}$ edges are clearly revealed.

DIFFERENTIATE A SPECTRUM

The difference spectrum is a numerical method of differentiating the spectrum. It emphasizes changes in the spectrum.

The top-hat filter gives a second-difference spectrum. Difference and/or division are all carried out digitally.

Another kind of difference method involves convoluting the experimental spectrum with a top-hat or similar filter function, as we described in Section 35.3 for XEDS (Michel and Bonnet). Top-hat filtering effectively gives a second-difference spectrum which also removes the background but exacerbates some artifacts.

Background subtraction for energy-filtered imaging can be achieved in several ways.

The first and most usual method of background subtraction while filtering, is called the three-window method (Jeanguillaume et al. 1978). Two pre-edge windows are used to calculate the background fit and one post-edge window in which the extrapolated background is subtracted from the total intensity to leave the edge intensity (equivalent to adding another pre-edge window in Figure 39.9). Egerton has shown that the intensity in the background window under the edge (I_b) is related to the intensity in the two pre-edge windows (I_1 and I_2) by

$$I_b = [A/(1-r)][E_h^{1-r} - E_l^{1-r}] \qquad (39.12)$$

where A and r are the usual factors in the background-fitting equation (39.10) determined from I_1 and I_2 and E_h and E_l are the high- and low-energy values defining the extrapolation window under the edge. To factor out the thickness effects, a low-loss image has to be acquired and divided into the K edge image, as in equation 39.8. Alternatively, two edges can be quantified and divided to give relative quantitative images, as in the quantification equation 39.9. Selection of the energy windows and choice of their width are subject to all the limitations we discussed for background subtraction and peak integration in that same section. Because the specimen thickness will often vary over the area being imaged, this method must be applied at every pixel in the image.

METHODS FOR BACKGROUND SUBTRACTION
The three-window method: use two pre-edge windows to calculate a background.
The jump ratio: calculate by dividing two signals.
Maximum-likelihood: when you only have a few channels to use for the subtraction.

The second method that is the commonly used method is to simply divide the signal in the edge by the signal in a background window just preceding the edge. These so-called jump-ratio images are only qualitative but give useful information, as we'll see in Section 39.9.

The third background-subtraction approach often used in ETEM imaging is the maximum-likelihood method (Unser et al.), which is useful when only a few channels are available for estimating the background and the peak intensity.

Kothleiner and Hofer describe the various parameters that control your choice of the best three windows. In general, many of the considerations for the selection of δ and Δ for spectral acquisition also apply during imaging.

39.4.C Edge Integration

The edge integration procedure you use depends on how you removed the background. If you used a power-law approach, then remember that there is a limit over which the edge integration window Δ is valid. The value of Δ should be large enough to maximize the integrated intensity, but not so large that the errors in your background extrapolation dominate. Often the presence of another edge limits the upper end of the integration window. The lower end is usually defined from the edge onset, E_K, but if there is strong (well defined) near-edge structure, such as in the B K edge or the Ca L_{23} edge, then your integration window should start at an energy above these, unless the quantification schemes can handle ELNES (see below). If you subtracted the background using a first-difference approach, then you determine the peak intensity by fitting the experimental spectrum to a reference spectrum from a known standard using multiple least-squares fitting. We'll talk more about this when we discuss deconvolution of spectra.

39.4.D The Partial Ionization Cross Section

There are several ways we can determine the partial ionization cross section, $\sigma(\beta\Delta)$. We either use a theoretical approach or compare the experimental spectra with known standard spectra.

Theoretical calculation: The most common approach is that due to Egerton (the 1979 and 1981 papers) who produced two short computer programs to model the K and L shell partial cross sections. The programs are called SIGMAK and SIGMAL, respectively. They are public-domain software, have been updated regularly, and are stored in your EELS computer system. They are standard parts of the Gatan software. The codes are given in the appendices of Egerton's book. The cross sections are modeled by approximating the atom in question to an isolated hydrogen atom with a charge on the nucleus equal to the atomic number Z of the atom, but with no outer-shell electrons.

At first sight, this so-called hydrogenic cross section is an absurd approximation! The approach is actually tractable because the hydrogen-atom wave function can be expressed analytically by Schrödinger's wave equation, which can be modified to account for the increased charge on atoms above hydrogen. Because this treatment neglects the outer-shell electrons, it is best suited to K shell electrons. Figure 39.12 shows comparison between the measured nitrogen-K intensity and that computed using SIGMAK. As you can see, the SIGMAK hydrogenic model essentially ignores the near-edge and post-edge fine structure (which would be absent in the spectrum from a hydrogen atom), but still gives a reasonably good fit to the experimental edge. Figure 39.13 compares the Cr L edge with the SIGMAK model. The L shell fit is almost as good as the K fit, although the white lines are imperfectly modeled. These programs are very widely used since they are simple to understand and easy and quick to apply. An alternative approach uses empirical parameterized equations to modify σ for the effects of β and Δ; Egerton gives the appropriate codes which you can also download from URL #2.

There are more complex methods which calculate the cross section in a more realistic way than the

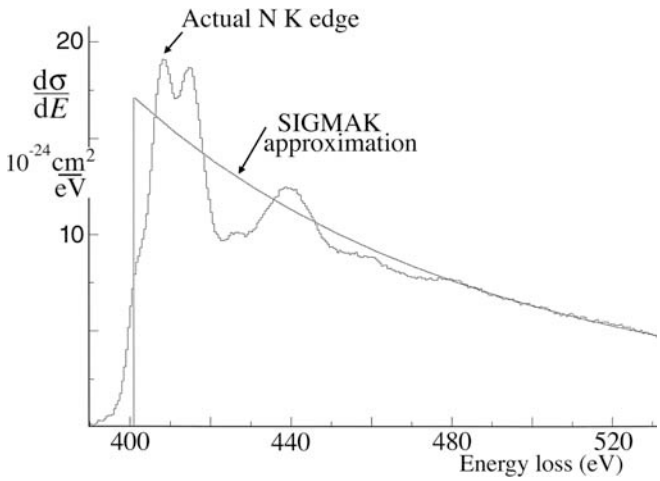


FIGURE 39.12. Comparison between an experimental N-K edge and a hydrogenic fit to the edge obtained using the SIGMAK program. The fit makes no attempt to model the near-edge fine structure but the total area under the fit is still a close approximation to the area under the experimental edge.

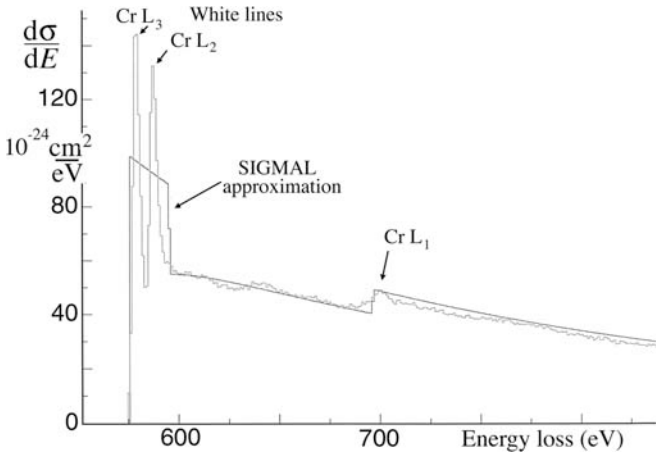


FIGURE 39.13. Comparison between an experimental Cr $L_{2,3}$ edge and a hydrogenic fit obtained using the SIGMAL program. The fit makes no attempt to model the intense white lines, but only makes a rough estimate of their average intensity.

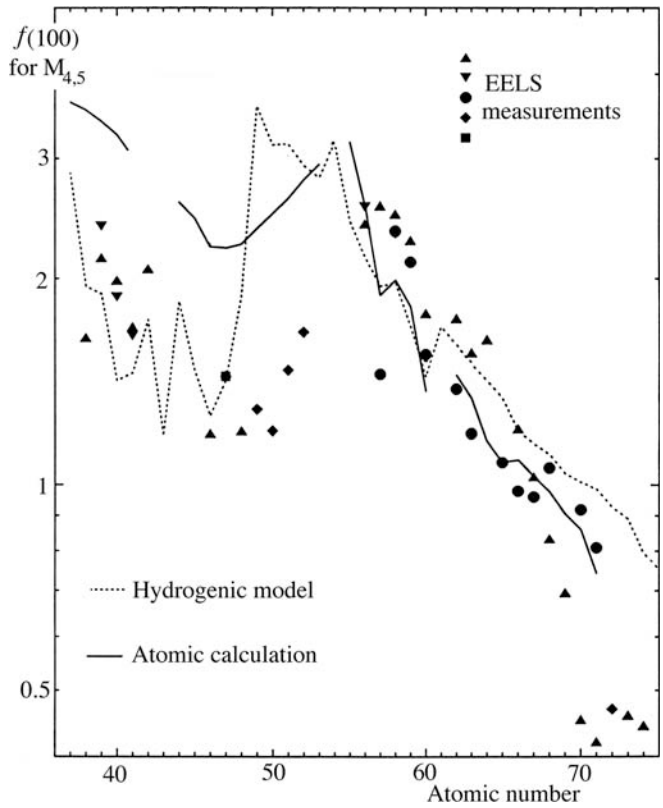


FIGURE 39.14. Comparison of the experimental and theoretical approaches to determination of the $M_{4,5}$ ionization cross section shown in terms of the variation in the dipole oscillator strength (*f*) as a function of atomic number.

SIGMAK/L hydrogenic models, e.g., using the Hartree-Slater model which is available in the Gatan software or atomic-physics approaches which are better for the more complex L and M (and even N) edges (see papers by Rez and Hofer et al.). Egerton has compared experimental and theoretical cross sections; the M shell data (which are the worst case) are shown in Figure 39.14. The data are actually plotted in terms of the oscillator strength γ (which is a measure of the response of the atom to the incident electron). This term is the integral of the generalized oscillator strength (GOS), which is proportional to the differential cross section, so just think of γ as being proportional to σ. There is still relatively poor agreement between experiment and theory for the M shell. Note that the models in Figure 39.14 are all atomic rather than hydrogenic. Similar data in Egerton's paper show better agreement for K and L shells. These models, while more precise, require substantially longer computing time but this is fast becoming less of a problem. Given the other sources of error in EELS analysis, you rarely need to go to such lengths to obtain a better value of $\sigma(\beta\Delta)$ and, unless you are well versed in the physics of ionization cross sections, you should probably stick with the SIGMAK/L methods, particularly for routine quantification.

Experimental Determination: Rather than calculating σ theoretically, you can generate a value experimentally using known standards. This approach is of course exactly analogous to the experimental *k*-factor approach for XEDS quantification in which the cross section is automatically included (along with the fluorescence yield and other factors). It is surprising at first sight that the classic XEDS approach of using standards has not been widely used in EELS, but the reason is obvious when you remember the large number of variables that affect the EELS data. The standard and unknown must have the same thickness, the same bonding characteristic, and the spectra must be gathered under identical conditions; in particular β, Δ, E_0, and t must be the same. So your standard would basically be your unknown!

Again, it is the problem of thickness measurement that appears to be the main limitation to improving the accuracy of analysis.

In summary, there are two approaches to the determination of $\sigma(\beta\Delta)$: theoretical calculation and experimental measurement. In contrast to XEDS, the theoretical approaches dominate. There is good evidence that, particularly for the lighter elements, for which EELS is best suited, the simple and quick hydrogenic model is usually adequate. However, for the heavier elements, where the M shell is used for analysis, the calculations are getting better although for such elements, it's still probably better to revert to X-ray analysis. Leapman has given a lucid and much more detailed description of the various methods of background subtraction and peak integration necessary for quantification.

One last point worth noting is that, before carrying out an experiment or gathering a long spectrum image for quantification, it is worth simulating the spectra to see if the experiment will produce useful data. We

described the advantages of DTSA for similar aspects of XEDS data in Chapter 33 and the companion text chapter. Gatan offers the EELS Advisor software (URL #3) to help take the uncertainty out of your planned experiment. Like DTSA, the EELS advisor allows you to simulate both spectra and images and will let you know if the element you seek will be detectable under your planned experimental conditions. It may tell you that the specimen is too thick or the amount of the element is below the detection or spatial resolution limits and that you need to change one or more of the many experimental variables that you have at your disposal.

39.5 MEASURING THICKNESS FROM THE CORE-LOSS SPECTRUM

While you may think we're now in a position where we have all the data needed to solve the quantification equations 39.8 and 39.9, our assumption all along has been that the spectra were the result of single scattering and we neglected plural scattering. In practice, there will *always* be some plural-scattering contribution to the ionization edges.

COMBINATION

The combination of a plasmon interaction and an ionization will show up as a bump about 15–25 eV past the onset of the edge.

This effect is shown schematically back in Figure 39.2E. So how do we go about correcting this? We can either make our specimens so thin that plural scattering is negligible or we can deconvolute the spectra. The former approach is better but sometimes unrealistic. The latter approach is mathematically simple but, as we've taken pains to point out on several occasions, deconvolution can be misleading and create spectral artifacts if not done properly; so we need to examine deconvolution in more detail. First, let's look at how we determine t because EELS offers us a simple method for this.

We saw back in Section 38.3.C that the plasmon-peak intensity is a measure of the specimen thickness but there is also thickness information in any energy-loss spectrum since the *total* amount of inelastic scatter increases with specimen thickness. So we can write a parallel expression to equation 38.7

$$t = \lambda \ln \frac{I_t}{I_0} \tag{39.13}$$

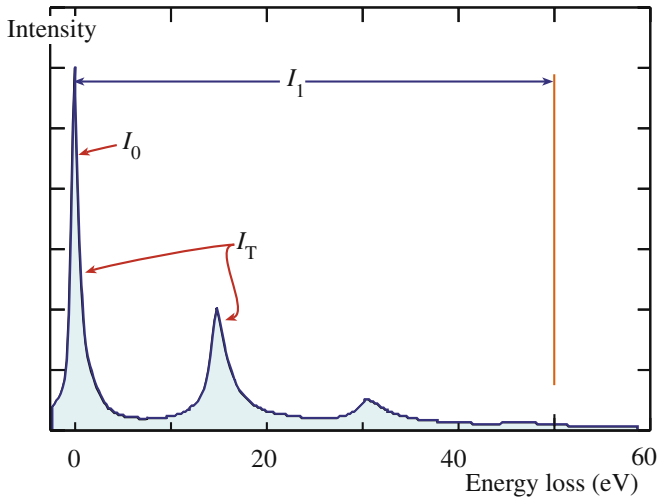


FIGURE 39.15. Definition of the zero-loss counts (I_0) and the total counts (I_T) required for thickness determination. I_T is effectively equivalent to the low-loss counts (I_l) out to ~50 eV, including I_0.

where I_0 is the ZLP intensity, I_t is the total intensity in the low-loss spectrum out to 50 eV, including I_0 (as shown in Figure 39.15) and λ is the average mean free path for these low-energy losses. We ignore any intensity above ~50 eV because, even though that's where all the interesting ionization edges are, it is a negligible fraction of I_t (as is apparent if you go back and look at Figure 37.1). To determine λ in equation 39.13, we use a parameterization based on many experimental measurements (Malis et al.)

$$\lambda = \frac{106F(E_0/E_m)}{\ln(2\beta E_0/E_m)} \tag{39.14}$$

where λ is in nm, E_0 in keV, β in mrad, F is a relativistic correction factor, and $\mathcal{E}_m$ is the average energy loss in eV which, for a material of average atomic number Z, is given by

$$\mathcal{E}_m = 7.6\,Z^{0.36} \tag{39.15}$$

The relativistic factor (F) is given by

$$F = \frac{1 + E_0/1022}{(1 + E_0/511)^2} \tag{3.16}$$

You can easily store these equations in the TEM computer or in your calculator and they give t with an accuracy of ~±20 % so long as the β is < ~15 mrads at 100 keV. In addition to this and the plasmon-peak-intensity approach, there are other methods (in Egerton's book) for determining thickness from various aspects of the EEL spectrum, but Malis' parameterization method is by far the most widely used. From what you've learned so far it should be obvious that if you can determine a thickness by comparing the intensity in two

regions of the spectrum, we can just as easily form an image from each of the two intensities in equation 39.13, and thus extract a thickness image as we did for XEDS in Figure 36.13C.

You may find that, when you measure t, Murphy's law is operating and the area you're interested in is too thick for quantitative core-loss analysis. Then you'll have to deconvolute the spectra to make the single-scattering assumption valid.

39.6 DECONVOLUTION

We saw back in Figure 39.2 that the plural scattering adds intensity to the ionization edge, mainly as a result of combined inner (ionization) and outer-shell (plasmon) losses.

We can approximate the experimental ionization edge as a true single-scattering (hydrogenic) edge convoluted with the plasmon, or low-loss, spectrum.

The aim of deconvolution therefore, as shown schematically in Figure 39.16, is to extract the single-scattering contribution from the plural-scattering intensity in the spectrum. We'll describe two methods, the Fourier-Log and the Fourier-Ratio, both developed in Egerton's book and available for download at URL #1. Both methods are also incorporated in the Gatan software. Using a small β increases the deconvolution error since the plural-scattered electrons have a wide angular distribution and so more of them are excluded as β decreases.

The Fourier-Log method removes the effects of plural scattering from the whole spectrum. The technique describes the spectrum in terms of the sum of individual scattering components, i.e., the zero loss (elastic contribution) plus the single-scattering spectrum plus the double-scattering spectrum, etc. Each term is convoluted with the instrument response function, which is a measure of how much the spectrometer degrades the generated spectrum; in the case of a PEELS, this is the point-spread function we described in Section 37.5. The Fourier transform of the whole spectrum (F) is then given by

$$F = F(0)\exp\left(\frac{F(\mathcal{E})}{I_0}\right) \tag{39.17}$$

where $F(0)$ is the transform of the elastic contribution, $F(\mathcal{E})$ is the single-scattering transform and I_0 is the zero-loss intensity. So to get the single-scattering transform you take logarithms of both sides, hence, the name of the technique.

Extracting the single-scattering spectrum would ideally involve an inverse transformation of $F(\mathcal{E})$, but this results in too much noise in the spectrum. There are various ways around this problem, the simplest of which is to approximate the zero-loss peak to a delta function. After deconvolution, you can subtract the background in the usual way, prior to quantification.

DECONVOLUTION METHODS

The Fourier-Log method: deconvolute then subtract background.
The Fourier-Ratio technique: subtract background then deconvolute.
Multiple least-squares fitting: when the specimen is not uniformly thin.
All three methods are approximations.

The danger of any deconvolution is that you may introduce artifacts into the single-scattering spectrum, e.g., from artifacts in the original spectrum. Despite the

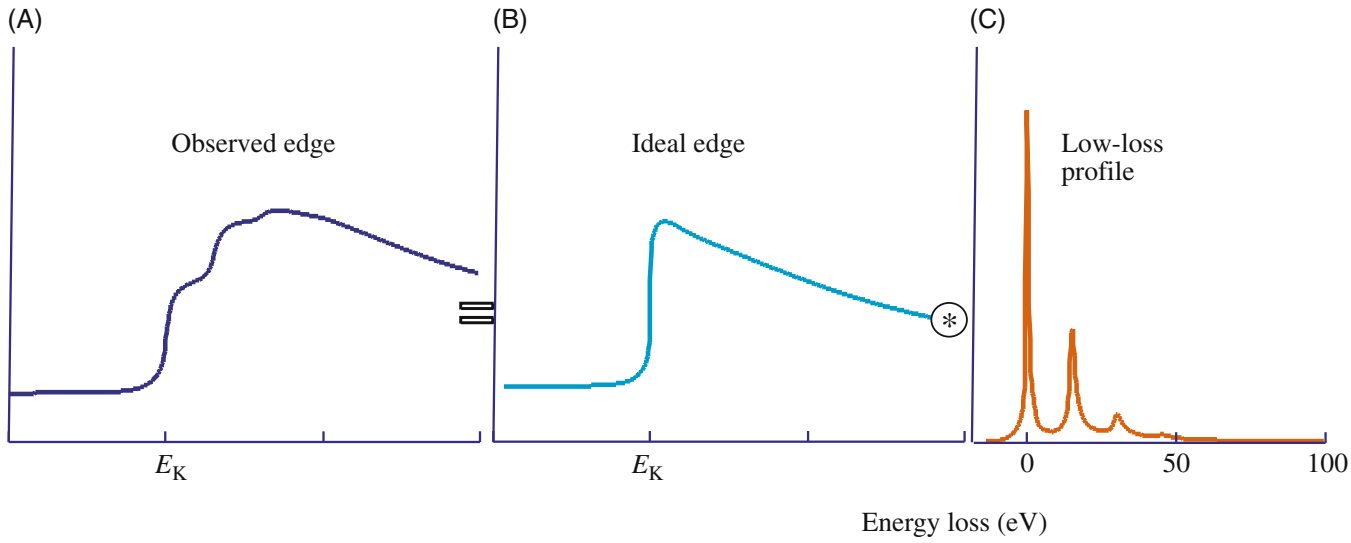


FIGURE 39.16. The experimentally observed ionization-edge intensity (A) consists of the convolution of the hydrogenic single-scattering ionization edge intensity (B) with the low-loss plasmon intensity profile (C).

assumptions and approximations, the net result of deconvolution is often an increase in the ionization-edge jump ratio. This improvement is important when you are attempting to detect small ionization edges from trace elements, or the presence of edges in spectra from thick specimens. An example of Fourier-Log deconvolution is shown in Figure 39.17.

The Fourier-Ratio technique: This approach approximates the experimental spectrum to the ideal single-scattering spectrum $F(\mathcal{E})$, convoluted with the low-loss spectrum. We define the low-loss portion of the spectrum as the region up to ~50 eV, including the ZLP, but before the appearance of any ionization edges. So we can now write

$$F' = F(\mathcal{E}) \cdot F(P) \qquad (39.18)$$

where F' is the Fourier transform of the experimental intensity distribution around the ionization edge and $F(P)$ is the Fourier transform of the low-loss (mainly plasmon) spectrum. In this equation, therefore, the instrument response is approximated by the low-loss spectrum rather than the ZLP. If we rearrange equation 39.18 to give a ratio (hence the name of the technique)

$$F(\mathcal{E}) = \frac{F'}{F(P)} \qquad (39.19)$$

We now obtain the single-scattering distribution by carrying out an inverse transformation. In contrast to the Fourier-Log technique, you must subtract the background intensity before deconvolution. Again, to avoid the problem of increased noise, it is necessary to multiply equation 39.19 by the transform of the ZLP.

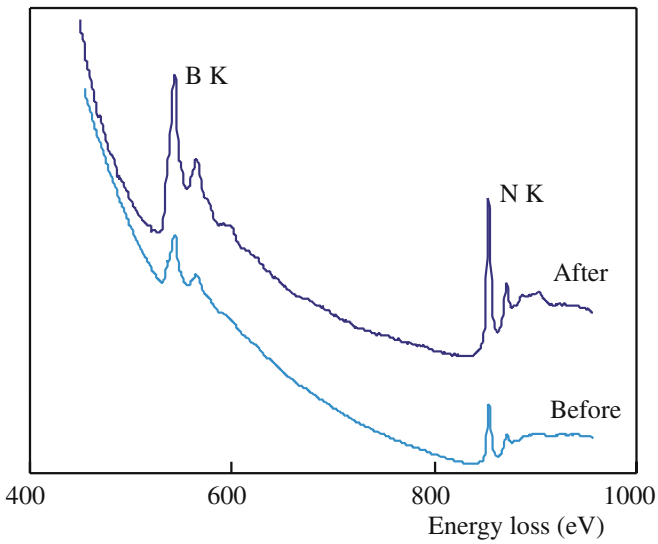


FIGURE 39.17. Spectrum from a thick crystal of BN before and after Fourier-Log deconvolution. The jump ratio in the deconvoluted spectrum (which is displaced vertically for clarity) is clearly increased by the process.

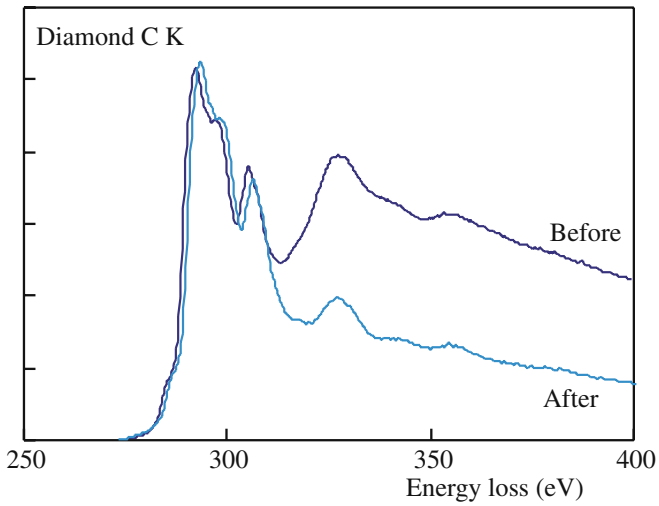


FIGURE 39.18. A carbon-K edge from a thick specimen of diamond before and after Fourier-Ratio deconvolution. You can see how the plural-scattering contribution to the post-edge structure is removed.

Figure 39.18 shows a carbon K edge after Fourier-Ratio deconvolution.

Multiple least-squares fitting: If your specimen is not uniformly thin, Fourier techniques won't work. Then you should use multiple least-squares (MLS) fitting of convoluted standard reference spectra (Leapman 2004). A single-scattering reference spectrum $R0(\mathcal{E})$ in the region of the edge to be quantified is convoluted with the first plasmon-loss portion of the unknown spectrum (P) and the resultant spectrum $R1(\mathcal{E}) = P^*R0(\mathcal{E})$ is used to generate several reference spectra ($R2(\mathcal{E}) = P^*R1(\mathcal{E})$, etc.). These reference spectra are then fitted to the experimental spectrum using MLS routines and specific fitting parameters are obtained. An experimental set of Fe, Co, and Cu reference spectra is shown in Figure 39.19A and the actual fit to part of the experimental spectrum from an intermetallic in a Cu-Be-Co alloy is shown in Figure 39.19B.

In summary, to quantify ionization-loss spectra you need a single-scattering spectrum, which can be approximated if you have very thin specimens or generated by deconvolution of your experimental spectrum. It is arguable that, given the stringency of the single approximation, it might be wise to deconvolute all core-loss spectra prior to quantification, but the uncertain effects of the possible errors introduced by deconvolution mean that you should do this cautiously. Often you'll find it useful to deconvolute the point-spread junction from all PEELS spectra, since this sharpens the edge onset and any ELNES intensity variations.

DECONVOLUTION CAUTION

Always check the validity of the deconvolution routine by applying it to spectra from a known specimen obtained over a range of thickness.

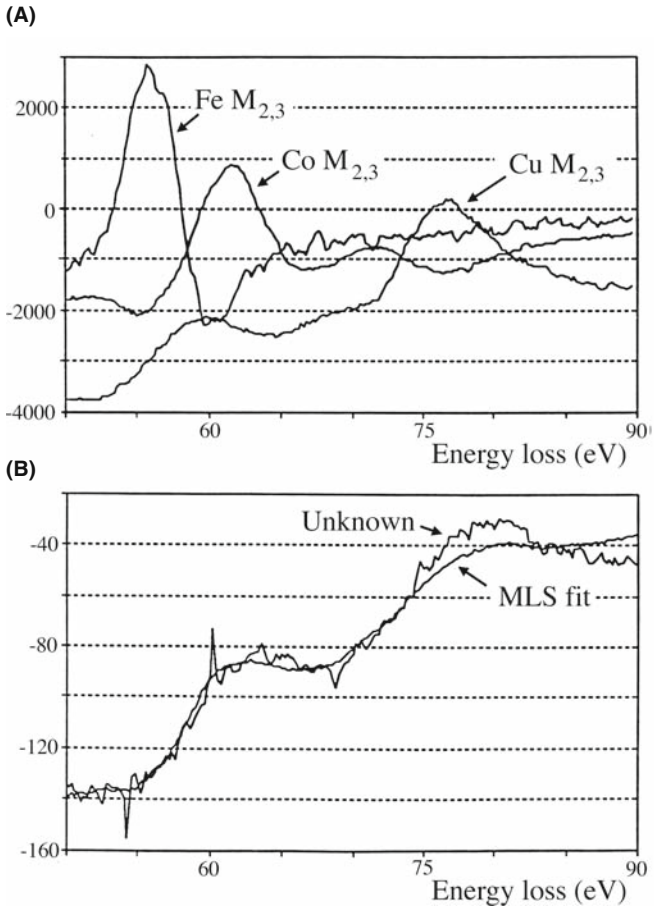


FIGURE 39.19. (A) Three first-difference low-loss, M-edge reference spectra from Fe, Co, and Cu superimposed on a low-energy portion of an experimental (first-difference) spectrum from an intermetallic particle in a Cu-Be-Co alloy. (B) MLS fit of the combined reference spectra to the experimental spectrum showing the good fit that can be obtained.

39.7 CORRECTION FOR CONVERGENCE OF THE INCIDENT BEAM

If you're working in STEM mode to get high spatial resolution, then it is possible that the beam-convergence angle α may introduce an error into your quantification. When α is equal to or greater than β, convergence effects can limit the accuracy because the experimental angular distribution of scattered electrons will be wider than expected (yet again, a good reason to know β). Therefore, you have to convolute the angular distribution of the ionization-loss electrons with the beam convergence angle. Joy (1986b) proposed handling this through a simple equation which calculates the effective reduction (R) in the partial cross section $\sigma(\beta\Delta)$ when $\alpha > \beta$.

$$R = \frac{\left[\ln\left(1 + \frac{\alpha^2}{\theta_E^2}\right)\beta^2\right]}{\left[\ln\left(1 + \frac{\beta^2}{\theta_E^2}\right)\alpha^2\right]} \qquad (39.20)$$

where θ_E is the characteristic scattering angle. A similar reduction factor is incorporated in the Gatan software quantification routines. So you can see that if α is small (particularly if it is smaller than β) then R is $<<1$ and the effect of beam convergence is negligible. Generally, with the typical range of probe-limiting apertures in a STEM, convergence angles should not be larger than 5–10 mrads so it should always be feasible to make sure that β is large enough. However, a note of caution is worthwhile because, with C_s correction, it is now possible to use much larger convergence angles to increase the probe current without degrading the probe size.

39.8 THE EFFECT OF THE SPECIMEN ORIENTATION

In crystalline specimens, diffraction may influence the intensity of the ionization edge. This effect may be particularly large if your specimen is oriented close to strong two-beam conditions and, as we saw back in Chapter 35, this can be used to good effect in ALCHEMI. Both X-ray emission and ionization-loss intensity can change because of electron channeling effects close to the Bragg condition. At the Bragg condition, the degree of beam-specimen interaction increases, compared with zone-axis illumination where no strong scatter occurs and the energy-loss processes behave similarly. This phenomenon (known as the Borrmann effect in XEDS) is not important for low-energy edges, but intensity changes of a factor of 2 have been reported for Al and Mg K edges (Taftø and Krivanek). The use of large α minimizes the problem in XEDS, but beam-convergence effects are themselves a problem in EELS as we just described. Unless you're an alchemist, the easiest way to avoid orientation effects is simply to operate under kinematical conditions and stay well away from any bend centers or bend contours, just as in XEDS.

39.9 EFTEM IMAGING WITH IONIZATION EDGES

There are countless examples of EELS analyses using ionization-loss edges in the general references that we have given at the end of this and the other EELS chapters. As we described back in Section 37.8, you have several experimental options such as point analyses, spectrum-line profiles, and various forms of energy-filtered TEM imaging. By far the most powerful method of analysis, as with XEDS, is to form EFTEM images. You can either select a given edge from which to form a single image or you can gather a spectrum image and select the specific energy later. The former method is the norm for in-column filters while the latter is more

common for post-column GIFs. We'll look at both but refer you to the 1995 text on the subject by Reimer for more theoretical and practical details and the review by Hofer and Warbichler for many illustrative examples. You can also image portions of the fine structure in the edge as we'll show in the next chapter, but here we will emphasize the power of ionization-loss analysis with reference to elemental images only. It is good to draw comparisons with what can be achieved with elemental imaging with XEDS, as we described back in Chapters 32–35.

39.9.A Qualitative Imaging

EFTEM images using the intensity in ionization edges will obviously correspond to elemental maps. The simplest way to get this information is to *subtract* a pre-edge background image from a post-edge image; this two-window subtraction method gives a passable qualitative elemental distribution. Alternatively, you can just *ratio* images of pre- and post-edge windows then what you get is called a jump-ratio image (see Figure 39.20B and C) which, again, is not quantitative and the intensity just reflects the edge to background ratio. Obviously, both of these qualitative methods will work best if (i) the jump ratio is high, (ii) the edge intensity is clearly visible above the background, and (iii) both thickness and diffraction conditions remain reasonably constant across the mapped region. Otherwise, the interpretation of any intensity changes is fraught with danger and artifacts abound. In fact, unless your specimen is very thin ($t \leq 0.1\ \lambda$) it's probably not worth bothering with this qualitative approach and you may as well spend your time forming quantitative EFTEM images.

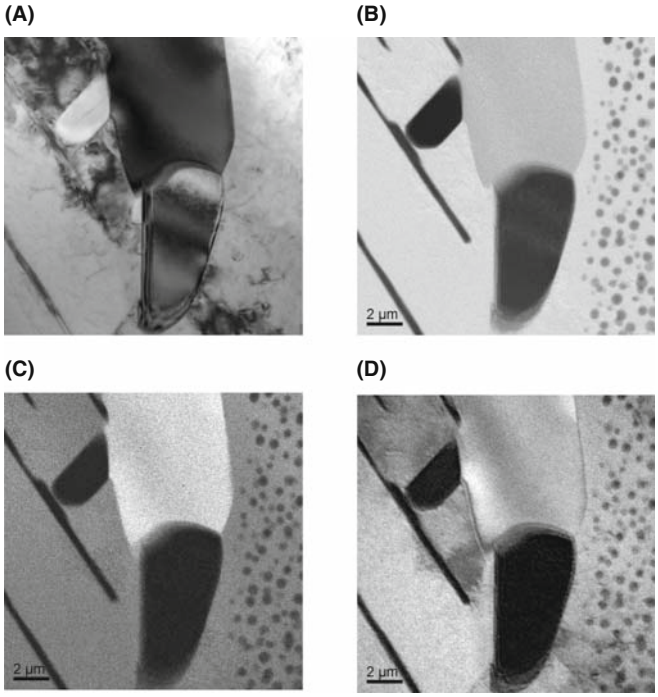


FIGURE 39.20. (A) BF image of precipitates in a stainless-steel foil. The other images were obtained from specific energy-loss electrons and illustrate both jump ratio and fully quantitative images. (B) Fe M jump ratio image; (C) Cr L jump ratio image; (D) quantitative Cr map.

39.9.B Quantitative Imaging

If, at each pixel, we carry out a background subtraction, an edge integration, and multiply the resulting intensity by the partial ionization cross section ratio, we should get quantitative images of the distribution of specific elements. There isn't too much difference, in principle, between quantifying a spectrum as we've just described in equations 39.8 and 39.9 and forming an image. The main difference is in fact in the method of background subtraction, which we've also described already in Section 39.4.B.

The most common method for quantitative EFTEM imaging involves acquiring three images from electrons in selected energy windows: two from the background preceding the edge and one from under the edge as we've already described. This approach works equally well for an in-column filter or for a post-column GIF; you can either operate in TEM mode and acquire three images or operate in STEM mode and choose whether you select just the three energy windows or gather a full spectrum image. The only difference is that the TEM images can be acquired in a few seconds while the STEM approach may require minutes or even hours and is probably only worth doing if you are acquiring a full spectrum image. Figure 39.20 shows ratio images and fully quantitative images pulled out of a spectrum-image data cube.

In both instruments, to form a specific filtered image, you shift the energy spectrum until the desired energy window passes through the energy-selecting slit (see the schematic diagrams in Figures 37.13 and 37.15). The energy shift is actually achieved by changing the accelerating voltage of the TEM so that electrons of different energies stay on-axis and thus in focus through the spectrometer. The quality of your images is governed by the same factors that control the spectrum: good jump ratios to get good signals, well-separated peaks to ensure good background-fitting statistics, and a good thin specimen to permit valid quantification under single-scattering conditions. Once acquired, your filtered images can be subject to advanced post-specimen processing, such as pixel-clustering methods (Cutrona et al.) or the standard methods, such as color assignments for different elements, as shown back in Figure 37.17D. If you're at all uncertain about whether the experiment will work, remember to simulate it first, using the EELS Advisor software (URL #2).

EFTEM imaging will only continue to improve as all the instrumental factors that we have already discussed become common. We will have C_s correction, monochromation, and better spectrometers with more uniform transmissivity achieved via higher-order aberration correction.

39.10 SPATIAL RESOLUTION: ATOMIC-COLUMN EELS

In contrast to the situation in XEDS, beam spreading is not a major factor in determining the source of the EELS signal and so the many factors that influence beam spreading are mainly irrelevant. The spectrometer only collects those electrons emanating from the specimen in a narrow cone, as shown in Figure 39.21 Therefore, energy-loss electrons that are elastically scattered through large angles are excluded from contributing to your spectrum. Remember that for XEDS these same high-angle electrons would still generate X-rays some distance from the incident-probe position, and these X-rays would be detected by XEDS. In the absence of a contribution from beam spreading, the spatial resolution of ionization-loss spectrometry depends on the mode of analysis

- The factor controlling the resolution in STEM mode, or in a probe-forming (diffraction) mode on a TEM, is mainly the size of the probe; because of the strong forward-scattered signal, we can easily get data with probe sizes < 0.2 nm and with aberration correction we can break the Ångstrom barrier.
- When we operate in TEM mode, the spatial resolution is a function of the selecting aperture, i.e., the spectrometer entrance aperture and its effective size at the plane of the specimen. Lens aberrations usually limit the spatial resolution, as we showed back in Section 37.4.C.

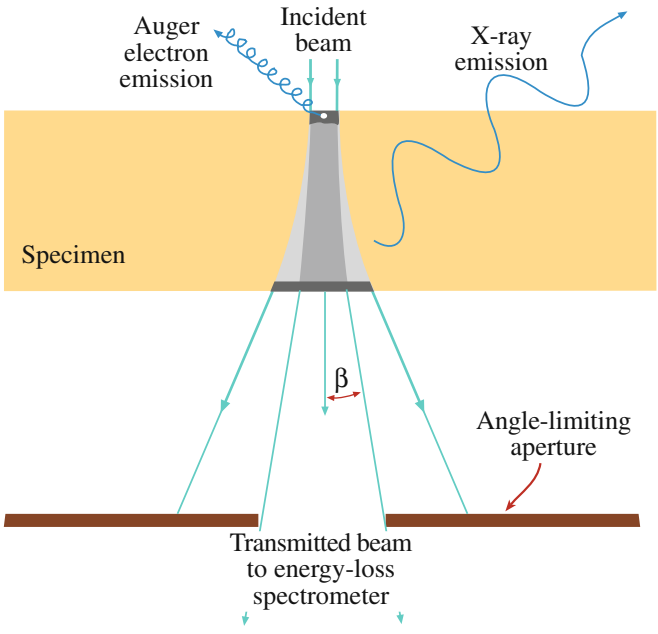


FIGURE 39.21. The effect of the spectrometer collection angle is to limit the contribution to the spectrum from high-angle scattered electrons, thus ensuring high spatial resolution. In contrast, X-rays can be detected from the whole beam-specimen interaction volume.

In addition to the usual factors affecting the probe size, such as the diffraction limit and lens aberrations (go back and read Chapter 5), another factor that we have to consider in EELS, but ignore in XEDS (although it occurs in X-ray generation also) is the phenomenon of delocalization.

DELOCALIZATION IN EELS

Delocalization is the ejection of an inner-shell electron by the passage of a high-energy electron some distance from the atom. It's as if the beam electron scares the core-shell electron sufficiently to eject it without actually laying a finger on it!

The scale of this wave-mechanical effect is inversely proportional to the energy loss over dimensions of a few nm (which, of course, is very large if you're worried about atomic-level spatial resolution). Egerton gives the following simplified expression for the diameter (d_{50}) containing 50% of the inelastic intensity

$$(d_{50})^2 = \left(\frac{0.5\lambda}{\theta_E^{3/4}}\right)^2 + \left(\frac{0.6\lambda}{\beta}\right)^2 \qquad (39.21)$$

where you should recognize all the terms and yet again appreciate the importance of knowing your collection angle. This expression gives a localization of ~1 nm for $\mathcal{E} = 50$ eV and 0.4 nm for $\mathcal{E} \simeq 300$ V (C edge K) and might therefore appear to prevent atomic-resolution EELS. Fortunately, delocalization does not appear to be a factor in STEM images of single, isolated atoms and resolution appears to be primarily determined by the width of the probe, even for light atoms. So it appears that the usual factors, such as probe aberrations, signal to background in the EELS signal, and damage are much more important in terms of EELS spatial resolution. All the experimental evidence seems to agree with the secondary role for delocalization in that HAADF imaging of single atoms has been long established (see Section 22.4 and the companion text); atomic-level changes in chemistry were first demonstrated by ionization loss in the early 1990s (Browning et al. and Batson). The same process that confines the electron beam to channel along the atomic columns in HAADF imaging also means that the spectroscopic

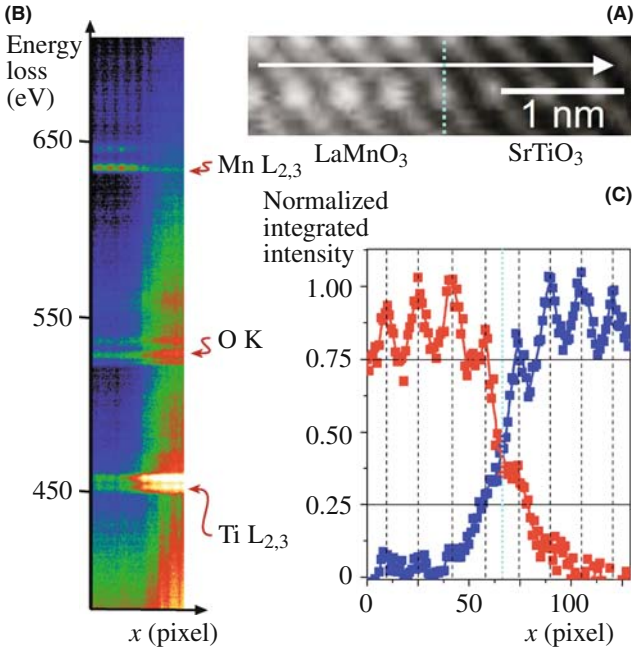


FIGURE 39.22. (A) HAADF image of a $LaMnO_3/SrTiO_3$ interface (blue dashed line). (B) EELS linescan along the arrowed direction in (A). Approximate positions of the Ti $L_{2,3}$, O K and Mn $L_{2,3}$ absorption edges are highlighted. The ripple corresponds to atomic-plane positions along the scan. (C) Normalized integrated intensities (40-eV window) under the Ti $L_{2,3}$ (blue) and the Mn $L_{2,3}$ (red) edges. Black dotted lines show the estimated positions of the respective MnO_2 and TiO_2 atomic planes.

signal is similarly localized. Because the EELS intensities are so much higher than XEDS, we are able to extract EELS spectra which contain information from a single atomic column, as shown in Figure 39.22). Both the Ti and Mn integrated signals in Figure 39.22C show an approximate decrease of 50% (or higher) within an atomic plane of the interface indicating true atomic-level resolution.

It is possible even to detect the presence of single atoms of high-Z elements on individual atomic columns of lower-Z elements (see next section) and C_s correction now makes atomic-resolution EELS almost straightforward (Varela et al.)

39.11 DETECTION LIMITS

The detection limits for ionization-loss spectrometry are governed by the same factors as we discussed for XEDS, so we have to optimize several factors

- The edge intensity.
- The signal to background ratio (jump ratio).
- The efficiency of signal detection.
- The time of analysis.

As is obvious in Figure 39.21, EELS has an inherently higher signal collection efficiency than XEDS. Conversely, as we've seen, it also has a correspondingly poorer signal to background because of the higher plural scattering. But, as for the best spatial resolution, more signal wins in the end. Leapman and Hunt argued in 1991 that, in most situations, PEELS is more sensitive to the presence of small amounts of material than XEDS. This has been borne out in experimental studies over many years and the latest improvements in FEG sources, C_s correction, and spectrometer hardware have brought us to the point where combined imaging and spectroscopic analysis (including fine structure and associated electronic effects) of single atoms is achievable. Figure 39.23 shows single-atom analytical sensitivity with atomic-level spatial resolution. So while, in principle, the inverse relationship between analytical sensitivity and spatial resolution that we described for XEDS in Figure 36.11 applies, when you're able to detect single atoms with atomic resolution you've just about reached the fundamental limit of any analytical technique.

Conclusion: analysis using ionization edges offers both atomic-resolution spatial resolution and single-atom analytical sensitivity.

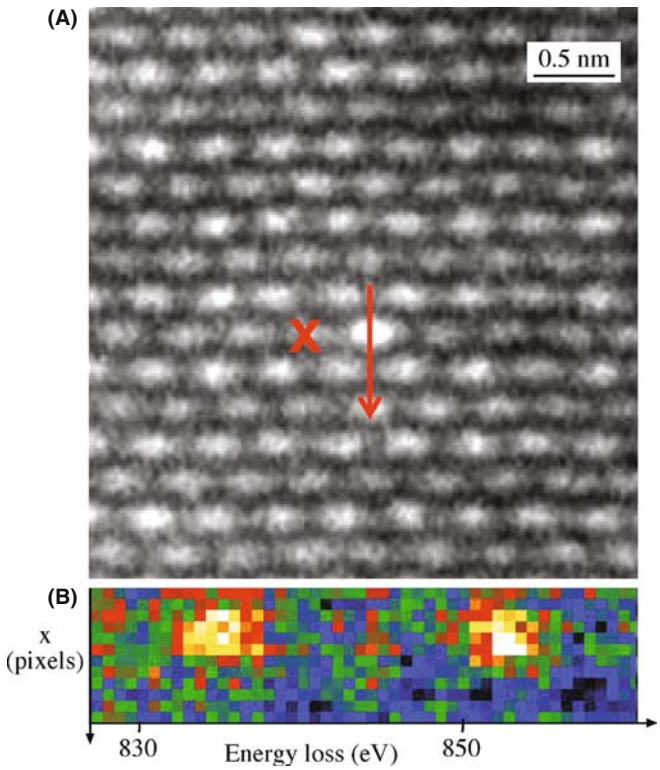


FIGURE 39.23. (A) HAADF image of a La impurity atom (X) in single atomic column in $CaTiO_3$. (B) Spectrum line-intensity profiles showing the white lines in the La $M_{4,5}$ edge as the beam scans across the La atom (along the red arrow). The white lines appear only when the La atom is scanned.

CHAPTER SUMMARY

The ionization edges can be used to give quantitative elemental analyses and quantitative images from all the elements in the periodic table, using a simple ratio equation. Beware, however, of the many experimental variables you have to define for your TEM, PEELS, energy filter, and (most importantly) your very thin specimen. Compared to XEDS there have been relatively few quantitative analyses or composition profiles measured using EELS, but quantitative imaging is becoming much more common.

To use Egerton's ratio equation you have to

- Subtract the background using a power law, or MLS approach. The former is easier. The latter is better for complex spectra.
- Integrate the edge intensity. That's usually straightforward.
- Determine the partial ionization cross section $\sigma_K(\beta\Delta)$. Calculate $\sigma_K(\beta\Delta)$ with SIGMAK and SIGMAL for most K and L edges.
- For M edges use a known standard or, better still, use XEDS.
- For the lightest elements (e.g., Li) use a known standard.

The biggest limitation to quantification is that your specimens have to be much thinner than one mean free path (typically $<<$ 50 nm) otherwise deconvolution routines are needed, which can introduce artifacts of their own.

Ionization-loss imaging is becoming widespread because of the increased availability of in-column and post-column filters and is the recommended method for quantitative analysis.

Spatial resolution and minimum detection are better than for XEDS. Combined atomic-column resolution and single-atom detection has been demonstrated.

BOOKS AND REVIEWS

Ahn, CC Ed. 2004 *Transmission Electron Energy-Loss Spectrometry in Materials Science and the EELS Atlas* 2nd Ed. Wiley-VCH Weinheim Germany. Must-read chapters on quantification and imaging.

Brydson, R 2001 *Electron Energy-Loss Spectroscopy* Bios (Royal Microsc. Soc.) Oxford UK. Great basic introduction; lots of practical tips.

Egerton, RF 1996 *Electron Energy-Loss Spectroscopy in the Electron Microscope* 2nd Ed. Plenum Press New York. Still the EELS bible. Read more here about the Bethe ridge.

Hofer, F and Warbichler, P 2004 *Elemental Mapping Using Energy-Filtered Imaging* in *Transmission Electron Energy-Loss Spectrometry in Materials Science and the EELS Atlas* 2nd Ed. **159**–233 Ed. CC Ahn Wiley-VCH Weinheim Germany. A thorough review of EFTEM/ESI, full of outstanding examples.

Joy, DC 1986a *The Basic Principles of EELS,* 1986b *Quantitative Microanalysis using EELS* in *Principles of Analytical Electron Microscopy* 249–276 and 277–299 Eds. DC Joy, A. Romig Jr. and JI Goldstein Plenum Press New York. Introduction to the principles and in-depth discussion of the experimental details necessary for quantification

Kohler-Redlich, P and Mayer, J 2003 *Quantitative Analytical Transmission Electron Microscopy* in *High-Resolution Imaging and Spectrometry of Materials* 119–187 Eds. F Ernst and M Rühle Springer New York. Integrated review of quantitative EELS and other TEM techniques with an emphasis on interfacial studies.

Reimer, L Ed. 1995 *Energy-Filtering Transmission Electron Microscopy* Springer New York. The first book on EFTEM/ESI with all you need to know about the theory and practice, but very few applications.

APPLICATIONS

Batson, PE 1993 *Simultaneous STEM Imaging and Electron Energy-Loss Spectroscopy with Atomic-Column Sensitivity* Nature **366** 727–728.

Browning, ND Chisholm, MF and Pennycook, SJ 1993 *Atomic-Resolution Chemical Analysis Using a Scanning Transmission Electron Microscope* Nature **366** 143–146.

Cutrona, J, Bonnet, N, Herbin, M and Hofer, F 2005 *Advances in the Segmentation of Multi-Component Microanalytical Images* Ultramicrosc. **103** 141–152.

Egerton, RF 1979 *K-Shell Ionization Cross-Sections For Use in Microanalysis* Ultramicrosc. **4** 169–179.

Egerton, RF 1981 *SIGMAL; A Program For Calculating L-shell Ionization Cross-Sections* in *Proc. 39th EMSA Meeting*198–199 Ed. G.W. Bailey Claitors Baton Rouge LA.
Egerton, RF 1993 *Oscillator-Strength Parameterization of Inner-Shell Cross Sections* Ultramicrosc. **50** 13–28.
Hofer, F, Golob, P and Brunegger, A 1988 *EELS Quantification of the Elements Sr to W by Means of M_{45} Edges* Ultramicrosc. **25** 81–84.
Jeanguillaume, C, Trebbia, P and Colliex, C 1978 *About the Use of Electron Energy-Loss Spectroscopy for Chemical Mapping of Thin Foils with High Spatial Resolution* Ultramicrosc. **3** 237–249.
Leapman, RD and Hunt, JA 1991 *Comparison of Detection Limits for EELS and EDXS* Microsc. Microanal. Microstruct. **2** 231–244.
Leapman RD (2004) *EELS Quantitative Analysis* in *Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS Atlas* 2nd Ed. 49–96 Ed CC Ahn Wiley-VCH Weinheim Germany. A great reference source.
Malis, T, Cheng, S and Egerton, RF 1988 *EELS Log-Ratio Technique for Specimen-Thickness Measurement in the TEM* J. Electron Microsc. Tech. **8** 193–200.
Michel, J. and Bonnet, N 2001 *Optimization of Digital Filters for the Detection of Trace Elements in EELS. III – Gaussian, Homomorphic and Adaptive Filters* Ultramicrosc. **88** 231–242.
Muller, DA and Silcox, J 1995 *Delocalization in Inelastic Scattering* Ultramicrosc. **59** 195–213. If you're physics-oriented and want to read more about delocalization in EELS.
Rez, P 2003 *Electron Ionization Cross Sections for Atomic Subshells* Microscopy and Microanalysis **9** 42–53.
Taftø, J and Krivanek, OL 1982 *Site-Specific Valence Determination by Electron Energy-Loss Spectroscopy* Phys. Rev. Lett. **48** 560–563.
Unser, M, Ellis, JR, Pun, T and Eden, M 1986 *Optimal Background Estimation in EELS* J. Microsc. 145 245–256.
Varela, M, Lupini, AR, van Bentham, K, Borisevich, AY, Chisholm, MF, Shibata, N, Abe, E and Pennycook, SJ 2005 *Materials Characterization in the Aberration–Corrected Scanning Transmission Electron Microscope* Annu. Rev. Mater. Res. **35** 539–569.

URLs

1) http://www.cemes.fr/&7Eeelsdb/
2) http://laser.phys.ualberta.ca/∼egerton/programs/programs.htm3.
3) http://www.gatan.com/software/eels_advisor.php

SELF-ASSESSMENT QUESTIONS

Q39.1 Is it better to do ionization–loss analysis at 100 keV or 200 keV? Justify your answer.
Q39.2 Why should you not operate with too small a collection angle if attempting to do quantitative analysis?
Q39.3 Why should you not operate with too large a collection angle if attempting to do quantitative analysis?
Q39.4 What is the compromise you make when choosing a large or small collection angle?
Q39.5 What kind of energy resolution do you need for quantitative analysis. Does this resolution requirement permit you to improve other aspects of the spectral acquisition?
Q39.6 Define the jump ratio. How can you increase this ratio for a given ionization edge?
Q39.7 Why do we not define the ionization edge energy as the peak-intensity channel of the edge?
Q39.8 What are white lines and why are they called this?
Q39.9 Define N, I, σ, and δ, including the appropriate units.
Q39.10 What relationship exists (if any) between the background-fitting window and the background-extrapolation window?
Q39.11 Under what circumstances might you use a 'goodness of fit' criterion for the background-extrapolation procedure?
Q39.12 Why do we use the term 'partial ' to describe the ionization cross section (σ) used in the Egerton equation?
Q39.13 Why do the simplistic SIGMAK and SIGMAL models still give a reasonable approximation to fitting the edge intensity?
Q39.14 Why isn't there a SIGMAM model and in its absence, how do you quantify M edges?
Q39.15 Why would you deconvolute your spectrum before attempting to quantify it?
Q39.16 Why is the background subtraction procedure in EELS far more important and difficult than in XEDS?
Q39.17 What are the units of quantification of the specimen composition obtained via the ratio method?
Q39.18 What is the greatest uncertainty in the quantification procedure?
Q39.19 Why might you need to correct the partial ionization cross section for too large a convergence angle instead of simply decreasing the C2 aperture?
Q39.20 Why is the spatial resolution of EELS fundamentally better than that of XEDS?

Q39.21 What is delocalization and how might it affect your EELS analysis?
Q39.22 Why do we call an ideal edge 'hydrogenic'?
Q39.23 Why does a hydrogenic ionization edge look like a triangle?
Q39.24 Why don't we see ionization edges from core levels much deeper than about 2 keV?

TEXT-SPECIFIC QUESTIONS

T39.1 Explain the nomenclature in Figure 39.1.
T39.2 Examine Figure 38.1 and compare with Figure 39.3. Why are both the zero loss and plasmon loss peaks effectively Gaussian in shape when all the ionization edges have a much broader intensity distribution extending asymmetrically over several tens of eV?
T39.3 Is the background intensity as drawn in Figure 39.2 realistic and if not, why not?
T39.4 If the energy resolution of the EELS spectrometer is on the order of a few eV at worst, why is it not possible to discriminate ionization edges that are several tens of eV apart without resorting to deconvolution?
T39.5 Given that TiN and TiC have equal numbers of atoms of each elemental constituent, why are the intensities so very different in the EELS spectra in Figure 39.8? What other spectroscopic method might you use to distinguish TiC and TiN, and why can't this method be used in an AEM?
T39.6 Does the first difference approach (Figures 39.10 and 39.19) actually remove the background intensity? If not, why not? How does this method compare with the top-hat filter approach used in XEDS?
T39.7 Why do we recommend using experimental standards for the best *k*-factor determination in XEDS but generally prefer calculation of partial ionization cross section values for EELS quantification?
T39.8 Calculate the thickness of an Fe specimen for which the low-loss intensity is 10% of the zero-loss intensity in a spectrum gathered with a collection angle of 100 mrads at 100 keV.
T39.9 Compare and contrast the two principal methods of deconvolution in Figures 39.17 and 39.18.
T39.10 Why can't we simply expand the integration windows δ and Δ in Figure 39.9 in order to increase the goodness of fit of the background subtraction and increase the total number of counts in the edge, respectively?
T39.11 Compare and contrast the expression for quantification of an ionization-loss spectrum (equation 39.5) with the Cliff-Lorimer expression for quantification of a characteristic X-ray spectrum (equation 35.2).
T39.12 Justify using the gross simplifications of the hydrogenic approach to model ionization edges shown in Figures 39.12 and 39.13.
T39.13 Look at Figure 39.6 and explain the relationship between the penetration of the electron into the potential well and the value of the energy loss.
T39.14 List the pros and cons of deconvoluting out the plural-scattering contributions and the point-spread function from a spectrum.
T39.15 Calculate the maximum convergence angle (α) you should use, such that the cross section for carbon K shell ionization (100 kV, $\beta = 20$ mrad) needs no correction. State any assumptions.
T39.16 Why has the spatial resolution of EELS analysis received so little study compared to that of X-ray analysis?
T39.17 Compare and contrast two different background-subtraction methods used for EFTEM with two different methods used for spectroscopy.
T39.18 Compare and contrast the experimental factors that limit spatial resolution in XEDS and EELS.
T39.19 Compare and contrast the experimental factors that limit analytical detection limits in XEDS and EELS.
T39.20 How do you know if your specimen is too thick for EELS analysis and how do you minimize the problem (apart from making a thinner specimen)?
T39.21 Why is EELS so much more efficient at collecting electrons compared to XEDS collecting X-rays, as summarized in Figures 39.4 and 39.5, given that the actual collection angles are about the same?
T39.22 Why is the B K edge so much more intense than the N K edge in Figure 39.3 when the atomic ratio of B:N in boring nitride is 1:1?
T39.23 What is a typical collection angle for X-rays in an XEDS spectrometer and how does it compare with the collection angle for electrons in an EELS?

40

Fine Structure and Finer Details

CHAPTER PREVIEW

In the previous chapter, we described elemental analysis using ionization edges, but there is much more than just elemental information in the ionization edges and this distinguishes EELS from XEDS. There are detailed intensity variations in the core-loss spectra called energy-loss near-edge structure (ELNES) and extended energy-loss fine structure (EXELFS). From this fine structure, which we can resolve because of the high-energy resolution inherent in EELS, we can obtain data on how the ionized atom is bonded, the coordination of that specific atom, and its density of states. As always, we can use any intensity changes to create filtered images which show the distribution of, e.g., regions of different bonding states. Furthermore, we can probe the distribution of other atoms around the ionized atom (i.e., determine the radial-distribution function (RDF) which is very useful for the study of amorphous materials) and we can study momentum-resolved EELS, observe the anisotropy of chemical bonds, combine EELS with tomography, inter alia. Understanding these phenomena often requires that we use certain concepts from atomic and quantum physics. The non-physicist can skip some sections at this time and just concentrate on the results. The rewards of working through this topic will be an appreciation of some of the more powerful aspects of EELS.

WHY LOOK AT FINE STRUCTURE ?

If high spatial resolution is important, you can't obtain this additional information by any other spectroscopic technique.

This fine structure is all the more useful because we now have the ability to simulate the spectra using atomic-structure calculations, which help us understand the details in the spectra. A full appreciation of the calculations is beyond the scope of the book but this is a growing field that will only assume more significance.

As a wrap-up to EELS and the book as a whole, we'll finish by saying a few words on some of the more esoteric aspects of TEM, such as angular-resolved spectrometry, radial-distribution-function determination, Compton scattering, core-level shifts, and tomographic EELS that are not yet in the mainstream but, with continuing advances in instrumentation and computation, will surely grow in importance.

40.1 WHY DOES FINE STRUCTURE OCCUR?

We saw in Section 39.1 that the ionization edges have intensity variations superimposed on the ideal hydrogenic sawtooth shape. The stronger oscillations occur within about 30–50 eV of the onset of the edge (ELNES) and the weaker ones extend out for several hundred eV as the edge intensity diminishes (EXELFS). This fine structure contains a wealth of useful information, but to understand its origins you have to use some ideas from quantum physics.

One way we can look at this process is to switch from a particle to a wave model of the electron, as we've done before, e.g., when we talked about diffraction in Part 2. Then we can imagine that any excess energy ($>E_c$) that the ejected electron possesses is a wave emanating from the ionized atom. Now, if this wave has only a few eV of excess energy, it undergoes plural, elastic scattering from the surrounding atoms, as shown schematically

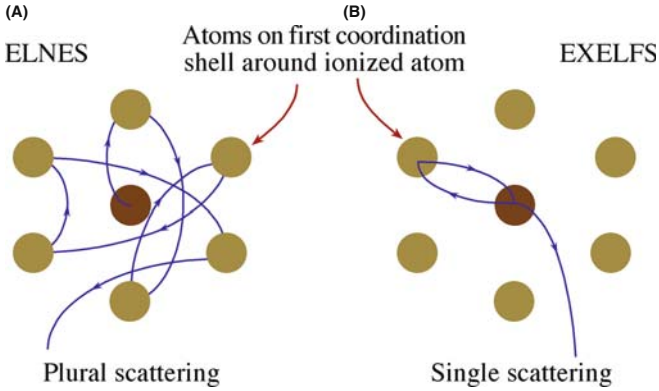


FIGURE 40.1. Schematic diagram showing the source of (A) ELNES and (B) EXELFS. The excess energy retained by the electron escaping above the Fermi level creates a wave radiating from the ionized atom and is scattered by surrounding atoms. The low-energy ELNES arises from plural scatter and is affected by the bonding between the atoms. The higher-energy EXELFS approximates to single scatter and is affected by the local atomic arrangement.

in Figure 40.1A, and this scattering is responsible for the ELNES, as we'll show. If the wave has even more excess energy, then, because of the smaller interaction cross section for higher-energy electrons (as we've already seen many times) it is less likely to be scattered by the surrounding atoms. In fact, we can approximate the cause of the ELNES to a single-scattering event, as shown in Figure 40.1B. Thus, EXELFS and ELNES can be viewed as a continuum of electron-scattering phenomena, with the arbitrary distinction that ELNES is confined to a few tens of eV past the edge onset while EXELFS extends for several hundred eV past the edge onset. There are other ways to explain fine structure and we'll mention some of these later when we talk about modeling the phenomena.

ELNES AND EXELFS I

Both arise because the ionization process can impart more than the critical ionization energy (E_c) needed by the core electron to be ejected from its inner shell.

Energy-Loss Near-Edge Structure

EXtended Energy-Loss Fine Structure

You should know that similar fine-structure effects can occur in X-ray spectra, but are usually not resolvable in the TEM because of the poor resolution of the semiconductor XEDS detector. However, we did note that experimental high-resolution X-ray detectors can resolve bonding effects in terms of shifts in X-ray peaks (see Figure 32.9C). In fact, there is a whole field of X-ray spectrometry that is used for studying atomic bonding (X-ray absorption near-edge structure or XANES) and atom positions and structure (extended X-ray absorption fine structure or EXAFS). These techniques are analogous to ELNES and EXELFS, but require a synchrotron to generate sufficient signal. This is one of the few examples where TEM is the cheaper characterization technique.

Most of this chapter deals with the experimental measurement and basic theoretical simulation of ELNES and EXELFS; the information we give is augmented in the companion text. While ELNES arises from plural scattering and is thus a more complex process than EXELFS, it is much more widely used, because it gives a more intense signal and the information it reveals has been used to study a very wide range of materials. So we'll discuss ELNES first.

40.2 ELNES PHYSICS

40.2.A Principles

As you know well, when an atom is ionized, it is raised from its ground state to an excited state leaving a hole in an inner shell. The core electron must receive enough energy from the beam electron to be ejected from its shell, but it may not receive enough to escape to the vacuum level. So it is still not completely free of the Coulomb attraction to the nucleus. In such circumstances, the final state of the core electron will be in one of a range of possible energy levels above the Fermi energy (E_F). You may recall that the Fermi level (or the Fermi surface in three dimensions) is the boundary between the filled states and the unfilled states in the weakly bound conduction/valence bands (although, strictly speaking, this statement is only true when $T = 0$ K). In a metal, there is no separate valence band and E_F sits somewhere in the conduction band, as shown schematically in the classic energy level diagram of an atom in Figure 40.2. In an insulator or a semiconductor, E_F is between the valence band (in which all the states are filled) and the conduction band (which has no filled states). The possible energy values that can be imparted to the ejected electron are controlled by the energy distribution of these unfilled states and, therefore, the energy lost by the incident electron similarly reflects this distribution of the unfilled states. One philosophical point of quantum uncertainty is that these states don't exist until an electron appears in them, but we'll conveniently ignore this.

So, the excited electron can reside in any of the unfilled states, but what's crucial here is that there is not an equal probability of the electron ending up in each possible unfilled state. Some empty states are more likely to be filled than others because there are more states within certain energy ranges than in others. This uneven distribution of empty energy levels is termed the density of (unfilled) states (DOS) and this is also shown

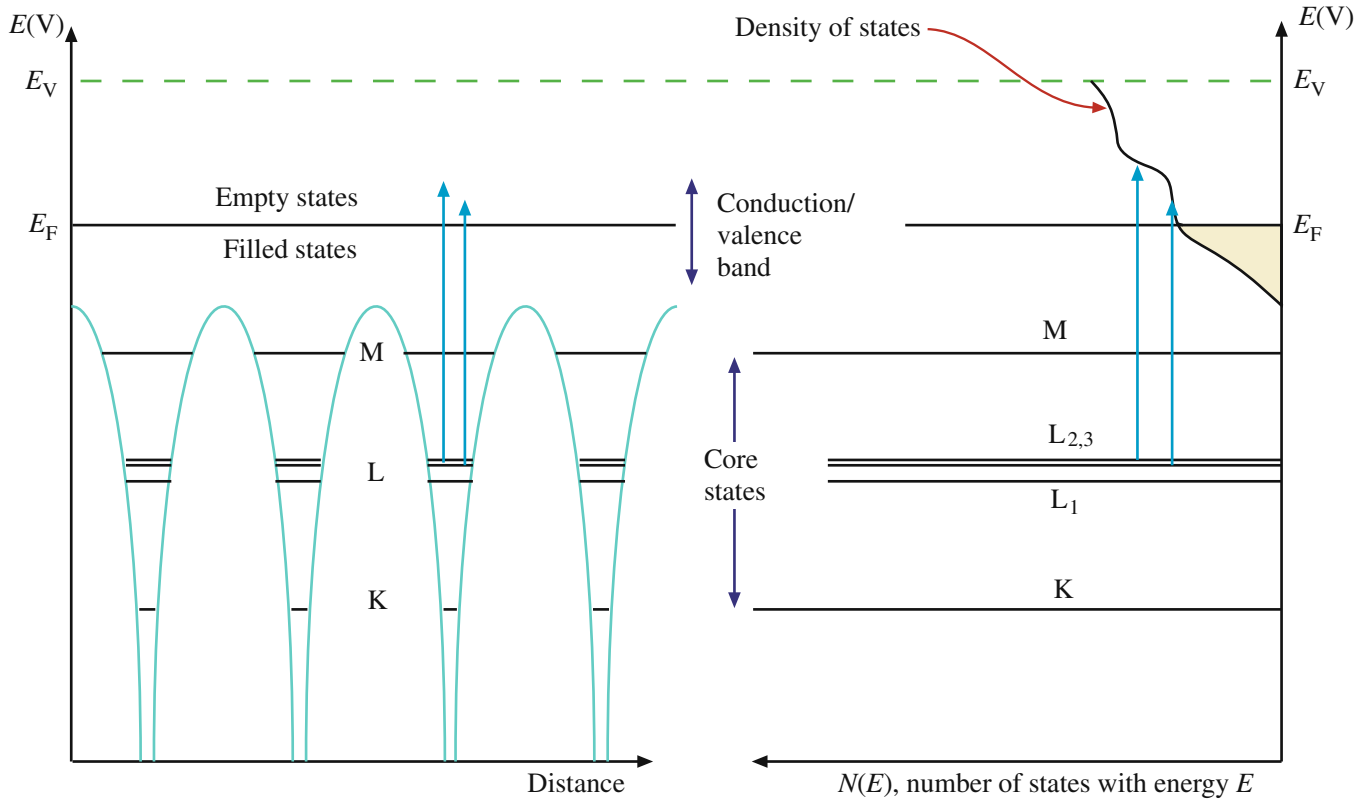


FIGURE 40.2. Relationship between the classic energy diagram of a metal atom (left) and the density of filled (shaded) and empty (unshaded) states in the conduction/valence band (right). The DOS is approximately a quadratic function on which small variations are superimposed. Ionization results in electrons being ejected from the core states into empty states above the Fermi level E_F.

in Figure 40.2. Because of the greater probability of electrons filling certain unoccupied states above E_F, the ELNES intensity is greater at the energy losses corresponding to these high DOS regions above the Fermi energy (which can be thought of as equivalent to the critical ionization energy E_C), as shown in Figure 40.3.

ELNES

This variation in intensity, extending several tens of eV above the ionization edge onset, E_C, is the ELNES, and it effectively mirrors the unfilled DOS above E_F.

The importance of ELNES is that the DOS is extremely sensitive to changes in the bonding or the valence state of the atom. For example, if you look ahead to Figure 40.5, the carbon K ELNES is different for graphite, diamond, and buckyballs and the Cu L ELNES changes when Cu is oxidized to CuO. On an even more detailed level, we can even deduce the coordination of the ionized atom from the shape of the ELNES.

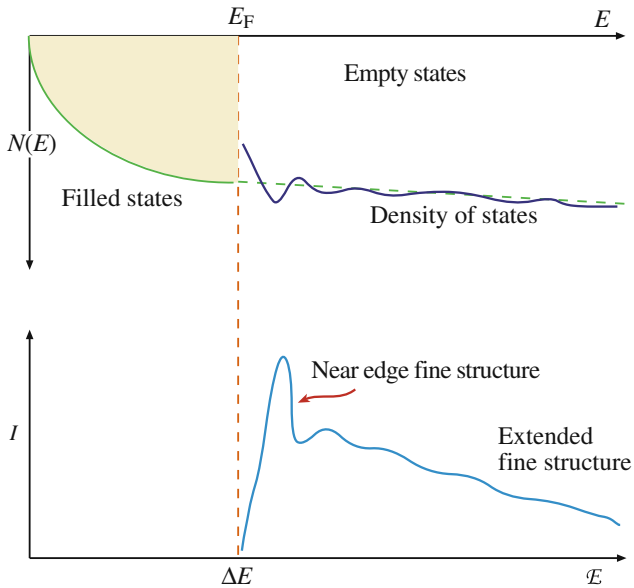


FIGURE 40.3. Relationship between the empty DOS and the ELNES intensity in the ionization edge fine structure. Note the equivalence between the Fermi energy E_F and the ionization edge onset E_C. Electrons ejected from the inner shells reside preferentially in regions of the DOS that have the greatest density of unfilled states. The filled states below E_F are drawn as a quadratic function, but this is an approximation.

DOS AND FERMI SURFACES

Even if you don't understand the intricacies of the DOS and Fermi surfaces, you can still deduce bonding information by comparing your experimental ELNES with that from standard specimens of known valence state or coordination.

We discuss this fingerprinting approach in Section 40.2.D below and you can check it in the EELS Atlas or at URL #1. Remember, we did exactly the same type of fingerprinting of different phases with the low-loss spectra in Section 38.3.A.

40.2.B White Lines

Perhaps the most startling example of ELNES is the presence of the white lines, which we first saw in Chapter 39; these lines are intense sharp peaks on certain ionization edges. These sharp peaks arise because in certain elements the core electrons are excited into well-defined empty states, not a broad continuum, as in Figure 40.3. The $L_{2,3}$ edges of the transition metals and the $M_{4,5}$ edges of the rare-earth elements show such lines. The white lines in the Fe L edge are the L_3 and L_2 edges, respectively, as shown in Figure 40.4, and these specific lines arise because the d shell has unfilled states. (We'll explain what happened to L_1 later.) To explain these lines we need a little more quantum physics, which you can skip if you wish and go to the last paragraph of this section. You should also be aware that there is disagreement as to whether white lines are truly fine structure or strictly ionization edge (atomic) intensity; but we'll leave this somewhat arcane discussion to those who know better (another cause of fracas in bars at M&M meetings).

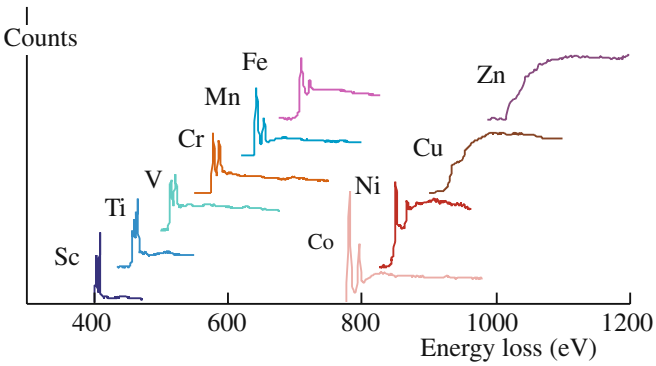


FIGURE 40.4. Spectra from the transition metals show a variation in the L_3 and L_2 white-line intensity ratios reflecting the variation in the number of core L-shell electrons ejected into unfilled d states. Note that Cu and Zn show no white lines because their d shells are full. The L_3 and L_2 white lines in the Fe L edge are the only ones that show the expected L_3:L_2 of ~2:1.

40.2.C Quantum Aspects

First, remember that the various electron energy levels, K, L, M, etc., correspond to principal quantum numbers (n) equal to 1, 2, 3, etc. Within those energy levels, the electrons may have s, p, d, or f states, for which the angular-momentum quantum number (l) equals 0, 1, 2, 3, respectively. The notation s, p, d, f comes from the original description of the atomic-spectral lines arising from these electron states, namely, sharp, principal, diffuse, and fine, although these terms have no counterpart in the EELS spectra we obtain.

As we noted in Section 39.1, the nomenclature $L_{2,3}$ arises from the fact that the L shell, from which the electron was ejected, has different energy levels. Such separation of the energies of the core states is called *spin-orbit splitting*.

Because the L electrons in levels 2 and 3 are in the p state, quantum theory demands that the sum (j) of their spin quantum number (s) and angular-momentum quantum numbers (l) is governed by the Pauli exclusion principle such that $j\,(=s+l)$ can only be equal to 1/2, 3/2, 5/2, etc. The spin quantum number, s (not to be confused with the s state), can only equal $\pm 1/2$. Taking all this into account, along with other quantum-number restrictions, it turns out that in the higher energy (i.e., more tightly bound) L_2 shell, we can have two p electrons with $j = \pm 1/2$ while in the L_3 shell, we can have four p electrons with $j = \pm 1/2, \pm 3/2$. Therefore, we might expect twice as many electrons to be excited from the L_3 shell as from the L_2 shell giving an L_3/L_2 intensity ratio (above the edge intensity, not above background) of 2. While this rule is approximately obeyed in the Fe spectrum shown in Figure 40.4, in practice, the ratio is seen to increase along the transition metal series from 0.8 for Ti to 3 for Ni, as is also seen in the spectral sequence in Figure 40.4.

Now these p-state electrons in the L shell cannot be excited to just any unoccupied state.

So for the p state ($l = 1$) the only permitted final states are either an s state ($l = 0$) or a d state ($l = 2$). Consequently, the core electrons are ejected primarily into the unoccupied d states in the conduction band, since there are few available s states there.

DIPOLE-SELECTION RULE

The change Δl in the angular momentum quantum number between the initial and final states must equal ± 1.

It is because of the dipole-selection rule that we don't see a strong L_1 edge in the spectrum. The L_1 edge sits closer to the nucleus than the L_2 and L_3 edges and its electrons are in the s state ($l = 0$) so they can only be excited to a p state ($l = 1$), but not to a d state ($l = 2$), or to another s state. Since there are few unfilled p states in

the conduction band of transition metals and they are much more spread out in energy than the d states, the L_1 intensity is very low and the peak is broad and may even be invisible in the $L_{2,3}$ post-edge structure.

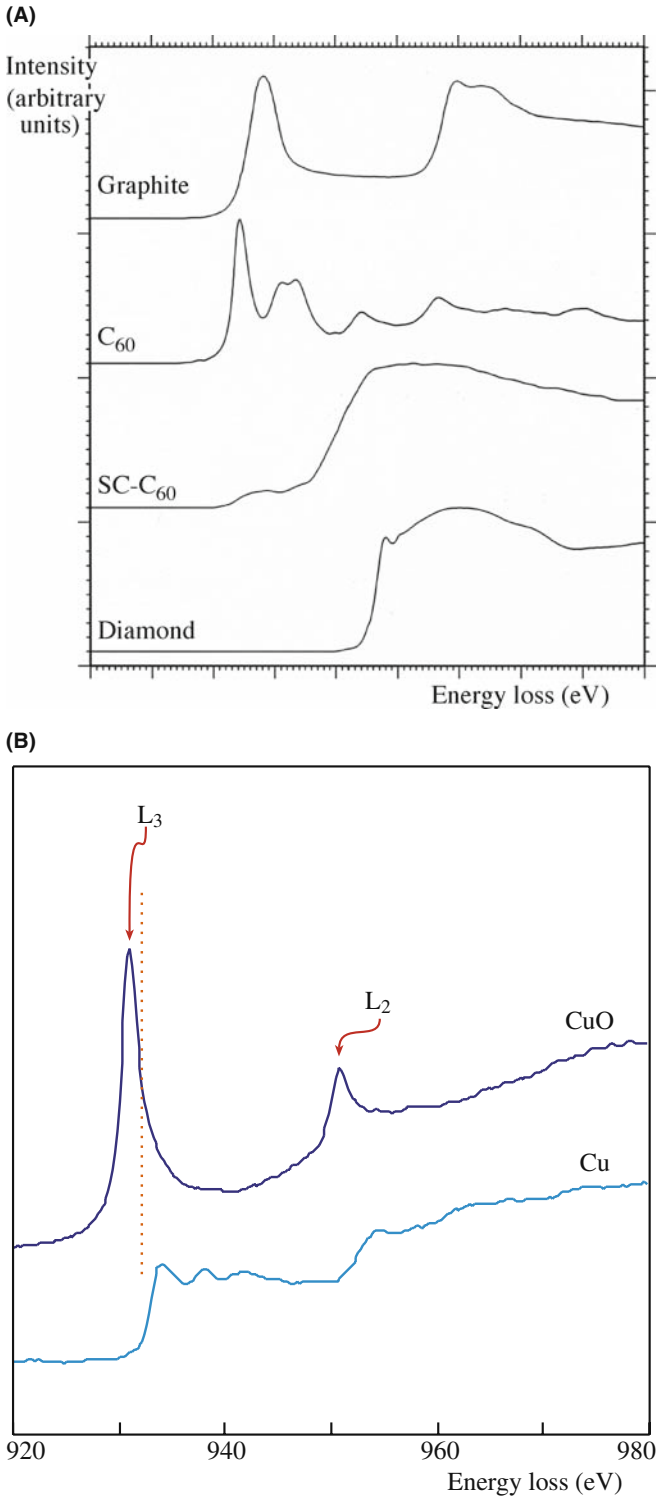


FIGURE 40.5. (A) Differences between the ELNES of the carbon-K edge from various forms of carbon. (B) Change in the Cu $L_{2,3}$ edge ELNES as Cu metal is oxidized and the filled d states lose electrons, thus permitting the appearance of white lines.

The energy width of the white lines is also affected by the time it takes for the ionized state to decay. One form of Heisenberg's uncertainty principle states that $\Delta E \Delta t = h/4\pi$, so a rapid decay gives a wide peak. For example, the Fe L_2 ionization can be rapidly compensated by an electron from the L_3 shell filling the hole and ejecting an Auger electron from the d shell. (This is called a Coster-Kronig transition.) A conduction band electron could also fill the L_2 core hole but the L_3 core hole can *only* be filled from the conduction band. Therefore, because there are two possible ways to fill the L_2 core hole, the L_2 line has a shorter Δt and a larger E than the L_3 line, which is much sharper.

In elements that don't have strong white lines, the ELNES is still present but appears just as weaker oscillations in intensity, which still reflect the DOS, and which, like the white lines, we can calculate and predict (much more of this later) (e.g., look at the pure-Cu ELNES in Figure 40.5C compared with the Fe ELNES in Figure 40.4).

40.3 APPLICATIONS OF ELNES

So let's see how all of this physics can be useful. (Is this an oxymoron?) The ELNES has been found to be dependent on details of the local atomic environment, such as coordination, valence state, and the type of bonding. Measurement of the fine structure, understanding how it is related to the electronic structure and ultimately to materials properties, can answer some hitherto-unsolved problems, particularly those where changes in bonding occur over small distances in your specimen. If you look at Figure 40.5, you'll see the carbon K edges for graphite and diamond. The carbon atom has hybridized s and p orbitals (termed σ and π in molecular-orbital theory). Graphite contains sp2 bonds in the basal plane with Van der Waals bonding between the planes. In contrast, the diamond structure has four directional, hybridized, sp3 covalent bonds and the atoms are tetrahedrally coordinated rather than arranged in graphitic sheets. The strong peak K edge at 284 eV identifies the empty π^* states into which the K shell electrons are transferred in graphite, while the diamond K edge has no π^* peak but shows a strong σ^* peak at about 290 eV. This kind of information is also extremely useful in the study of thin diamond and diamond-like carbon films, which are of great interest to both semiconductor manufacturers and the coatings industry (sunglasses in particular). Carbon films can be made with a continuous range of graphitic and diamond-like character and it is possible to deduce the relative fraction of sp3 (diamond) and sp2 (graphite) bonding from the K edge ELNES (Bruley et al.). In today's world of carbon nanotubes, buckyballs, and graphene, all these newer forms of carbon can easily be

distinguished by their ELNES. For example, carbon K-edge spectra from C_{60} (Buckminsterfullerene of Buckyballs) are also shown in Figure 40.5 in the standard and shock-compressed form. Another useful example is given in Figure 40.5B, where the changes in the Cu $L_{2,3}$ edge with oxidation are shown. This is a classic example. Since Cu metal has all its 3d states filled, there are no white lines in spectra from the metal. Upon oxidation, some 3d electrons are transferred to the oxygen, leaving unfilled states, and the white lines appear in the oxide spectrum. Note also that the onset of the oxide edge is different from that of the metal, because this electron transfer changes the value of E_C.

ELNES changes often occur at interfaces where the bonding changes locally over less than 1 nm. In Figure 40.6 the Si-K edge ELNES is seen to change across a Si-SiO_2 interface because the Si bonding changes. In this example, you can see the extraordinary power of an FEG STEM to provide simultaneous atomic-level images and spectra localized to individual atomic columns (even though this work (from Batson) is now more than 15 years old). The combination of *Z*-contrast imaging (see Section 22.4) and PEELS is arguably the most powerful analytical technique for atomic characterization, as we showed in Figures 39.22 and 39.23.

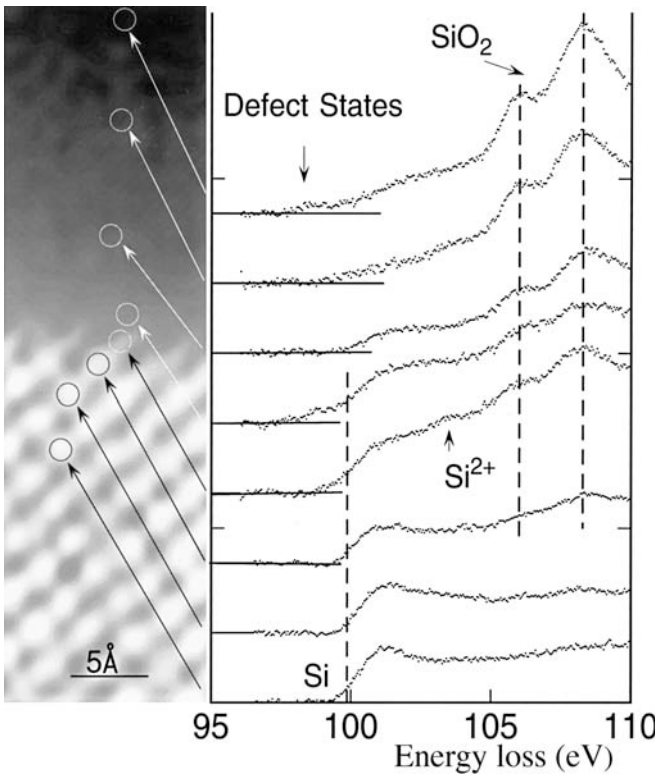


FIGURE 40.6. The change in the ELNES of the Si L edge across an interface between crystalline Si and amorphous SiO_2. Local electronic changes at the atomic level are easily discerned.

THE CHEMICAL SHIFT

This difference in edge-onset energies is called a chemical shift and also helps to fingerprint the specimen. (More on this in Section 40.6.)

Bonding may also be changed by local segregation and one of the more powerful examples of ELNES is the detection of bonding changes associated with elemental segregation to interfaces, which can cause extraordinary changes in the mechanical properties of metals and alloys. For example, Ni_3Al has great potential as a high-temperature intermetallic, but is limited by its inherent brittle behavior resulting in intergranular fracture. It has been known for many years that this brittle behavior can be countered by the addition of a fraction of a percent of B which is known to segregate to the boundaries. Why this segregation results in a major ductility improvement was unknown until it was shown that, at B-containing boundaries, the Ni $L_{2,3}$ edge exhibits slight ELNES changes consistent with the more metallic-like bonding of pure Ni (see Muller et al.). In a complementary study, Keast et al. measured Bi segregation to Cu grain boundaries, and observed ELNES changes in the Cu $L_{2,3}$ edge, consistent with the Cu atoms in the boundary taking on a less-metallic bonding state (see Figure 40.7). This ELNES change, which is equivalent to the transfer of less than 0.3 electrons (whatever that means) from each Cu atom at the boundary to an adjacent Bi atom, may account for the brittle behavior of Cu doped with as little as 20 ppm Bi; an extraordinary change in mechanical behavior, noted first in 1874. Understanding the role of slight electronic bonding effects in such macroscopic behavior as brittle behavior may help to transform the power-generation industry for example, which spends billions of dollars removing impurity elements that cause catastrophic failure of pressure vessels if left to segregate to grain boundaries.

Such studies of ELNES are probably the most widely used aspects of EELS and the literature abounds with ELNES studies of valency determination and atomic coordination. Some examples include bonding changes at oxide interfaces on Si (Botton et al.) and probing the structure of potential next-generation Hf-based gate oxides in Si semiconductors (McComb et al.). Reviews of the potential and practical applications of ELNES have been given by Keast et al. and Brydson et al.

40.4 ELNES FINGERPRINTING

Although the ELNES is directly related to the details of the electronic structure, interpretation of particular features in an experimental spectrum is not always

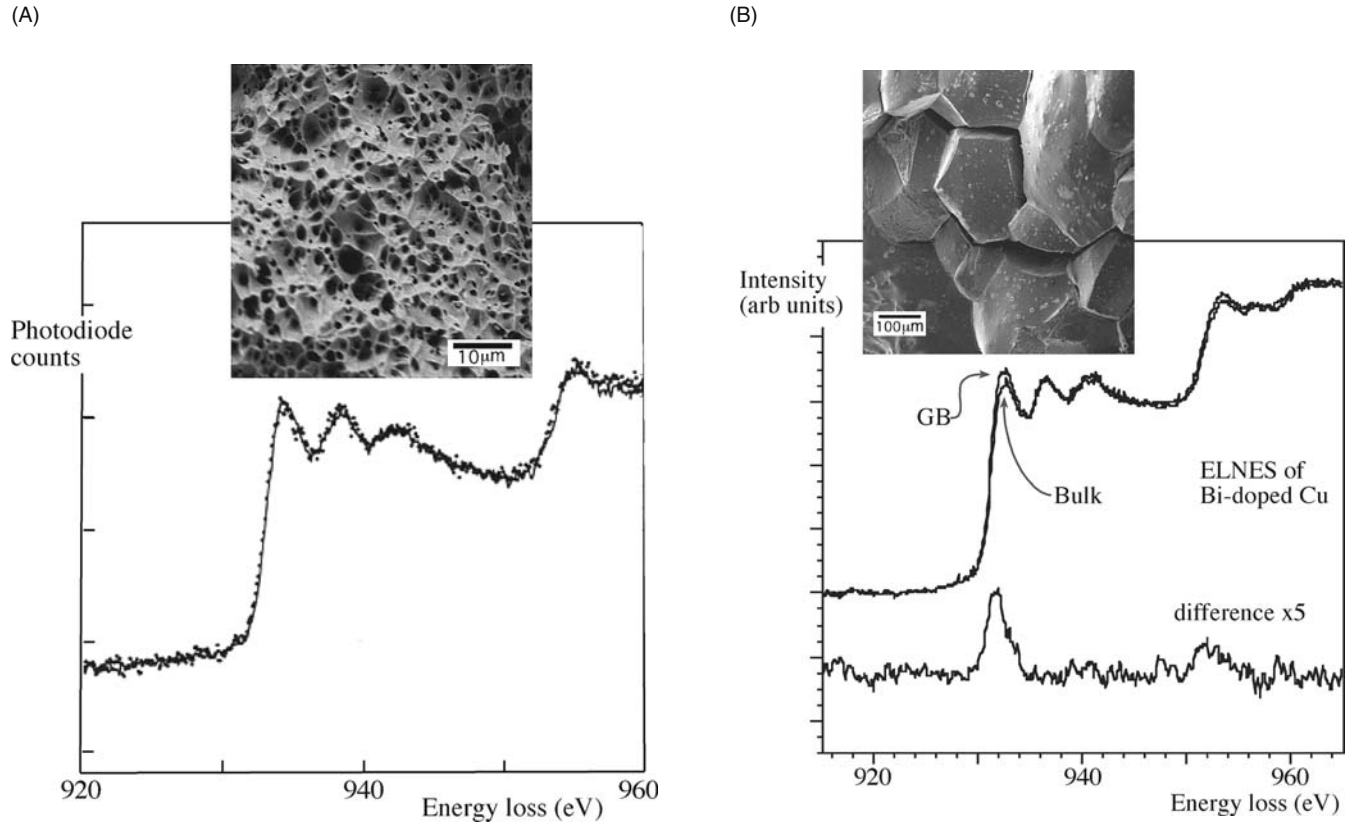


FIGURE 40.7. Change in ELNES due to impurity segregation. (A) Cu $L_{2,3}$ ELNES in pure Cu. (B) Slight Cu $L_{2,3}$ ELNES change between the bulk (grain interior) and a grain boundary to which Bi is segregated. The effect is magnified 5× in the difference plot. The two SE images show the extraordinary change in fracture behavior of ductile, pure Cu and brittle, Bi-doped Cu.

straightforward and you may not have the capability to carry out the atomic-structure calculations that we'll describe below. If this is the case, don't despair because you can still use a fingerprinting approach without fully knowing the details of the electronic structure. The idea behind such fingerprinting, as we've seen for the low-loss spectra, is that the general form of the ELNES is predominantly sensitive to the nearest-neighbor coordination and so it changes with changes in the structure. An example is provided by the Al-$L_{2,3}$ ELNES and Al-K ELNES of aluminum-oxygen materials which are sensitive to the local coordination of Al (i.e., whether octahedral or tetrahedral). Likewise, Figure 40.8 shows the experimental Mn-$L_{2,3}$ ELNES for different minerals in which the valence state of the Mn varies from 2+ to 4+. Again, we use the word fingerprint to emphasize that it is not necessary to understand the details of the DOS of complex materials in order to be able to interpret the ELNES spectra. Direct comparison with spectra from known standards is often all that is required for probable identification of the bonding state of a specific atom in your specimen. But because the matches are rarely perfect (given all the experimental and specimen variables that may affect the detailed intensity in the fine structure) go back and read our caution about low-loss fingerprinting in Section 38.3.A and apply it to your ELNES fingerprinting. Note that you don't need the very best energy resolution to carry out fingerprinting: the data in Figure 40.8 were taken many years ago from a standard PEELS system and, for many cases, a LaB_6 source is fine.

A theoretical calculation of the unoccupied DOS will always be useful in understanding or predicting features in the ELNES. In the next section, we'll show some examples where modeling the ELNES has helped our interpretation.

40.5 ELNES CALCULATIONS

Many attempts have been made to compare the experimental ELNES with calculations of the DOS in simple materials, such as metals and oxides. Great strides have been made in the last few years, mainly in improvements in models of the atomic potentials and in the computing power needed to pursue the calculations. This aspect is transforming the study of ELNES from an esoteric field to one with broad applications in materials science. This topic is also addressed in substantial depth in the companion text.

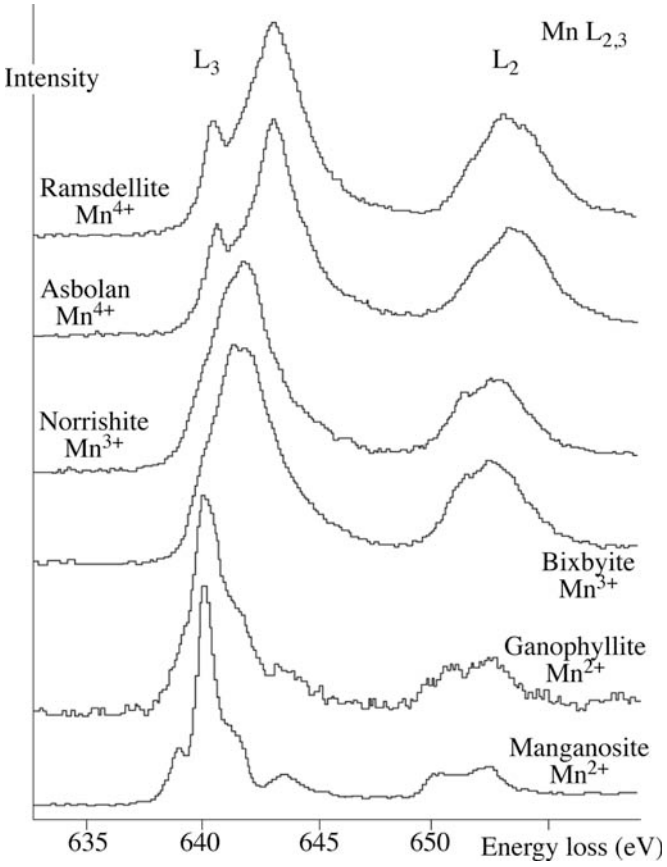


FIGURE 40.8. Comparison of Mn $L_{2,3}$ ELNES from a range of minerals in which the Mn coordination and hence valence state changes. The L_2 and L_3 white lines broaden as the oxidation state increases from +2 to +4 and in some cases the L_3 peak splits into two peaks. Understanding why such changes happen from an electronic standpoint is not necessary for identifying the different minerals or valence states.

40.5.A The Potential Choice

Calculating the electronic structure in solids involves solving the Schrödinger equation for each electron in the potential of the solid, including the Coulomb potential of both the nuclei and all the other electrons. (Now might be a good time to move on to the next section if you haven't recognized too many words in the previous sentence!) We also have to include terms due to the fact that the electrons are affected by the presence of other electrons and their behavior is correlated (i.e., they aren't isolated particles). Given the large number of electrons involved in any calculation, we often use an approach called density-functional theory (DFT) (if you're into physics and want to go to the next level (or two) then you need to read Finnis' (2003) book on atomic-modeling which includes DFT). Out of DFT comes a simplifying assumption, which we call the local-density approximation (LDA). Within the LDA method, we choose one of three different approaches to perform our calculations, which basically come down to a choice of atomic potential

- We can calculate the band structure directly in reciprocal space. This is usually described as the band-theory approach and the electron states are formed in a repeating crystal lattice. If this reminds you of Bloch states, back in Chapter 14, you're right.
- We can describe the electron states in terms of molecular orbitals (MO).
- We can calculate the effect of multiple scattering (MS) of the electron wave in real space based on the model shown in Figure 40.1A.

A range of band-structure methods are used and they go by rather strange names, such as augmented plane wave (APW), full-potential linearized APW (FLAPW) (URL #2), augmented spherical wave (ASW) (URL #3), CASTEP (URL #4), Layer Korringa-Kohn-Rostoker (LKKR), pseudopotentials, and other methods. The URLs will lead you to the sites for the various public-domain or commercial versions of the software. The best name by far is the muffin-tin (MT) potential (URL #5), which is spherically symmetric at the atomic positions and flat between them. (Apparently, to some physicists, this shape looks like a cross section of a tin used to bake muffins.) This model modifies the classic energy diagram, as shown in Figure 40.9. The MT form is useful because it generates wave functions that we can break down into the various angular-momentum components (which describe the partial DOS which is reflected in the ELNES). However, most MT approaches assume the crystal lattice is infinite and you need Bloch's theorem for the wave-function calculations which give the DOS. You'll probably find these techniques computationally challenging and not very flexible. But with the advent of easily available, high-performance, parallel computing, we are no longer so constrained (see the next section). For example, only recently have MO theorists been able model large unit cells, planar interfaces, and those (now-ubiquitous) amorphous materials.

THE TERMINOLOGY

Pseudopotential
MT: Muffin-tin potential
DOS: density of states
DFT: density functional theory
LDA: local density approximation
APW: the augmented plane wave
FLAPW: the full-potential linearized APW

MO theory is just an extension of using molecular orbitals to describe solids. (Not surprisingly, this approach is often used by chemists!) To use this approach, we have to divide our specimen into separate molecular units. If we calculate the MOs for each unit,

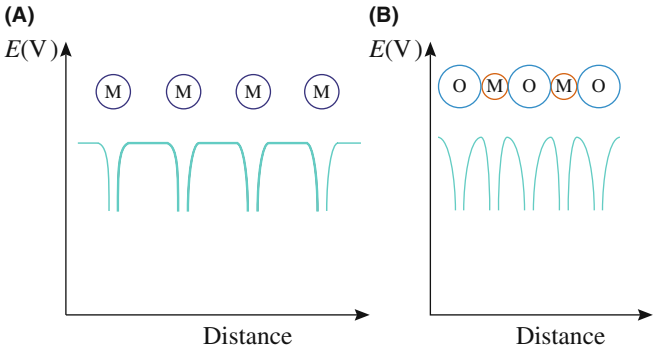


FIGURE 40.9. The muffin-tin potential energy diagram for (A) a metal and (B) an oxide. Note the symmetry of the potential wells for the metal and the asymmetry for the oxide.

we can then interpret the ELNES spectra in terms of core-shell electrons being ejected into unoccupied MOs. These unoccupied MOs arise because the excited-atom orbitals interact (i.e., bond) with nearest-neighbor atoms. (This is the π/π^* and σ/σ^* bonding/antibonding orbital notation that we used to describe the C-K shell ELNES back in Figure 40.5A). We can extend this idea and imagine the various MOs as simply linear combinations of atomic orbitals (which is then called the LCAO approach). LCAO works well if the orbitals are occupied but, for unoccupied orbitals, we have to use the self-consistent field (SCF) method, which basically assumes that the atoms are organized in a localized molecular cluster which then uses a version of the MS method which we'll now describe.

MS (not Word!) calculations are based on the interpretation of the ELNES as scattering of the electron wave that emerges from the ionized atom by atomic shells around that excited atom (which we started with back in Figure 40.1). Still the most elegant MS method is that due to Durham et al. The Durham method first divides the cluster of atoms into shells, each approximately equidistant from the ionized atom. We then solve the scattering within each atomic shell in turn. Finally, we consider scattering between different atomic shells. Since we have to calculate *all* possible scattering paths it's much easier in crystals because we can use their symmetry to facilitate our calculations. A self-consistent version of the MS method called FEFF (now in version 8) is commercially available (URL #6) and highly recommended. You can extend this shell-by-shell MS approach to model much more complex amorphous systems, incommensurate structures, and non-periodic atomic arrangements at planar interfaces and defects. You'll find that the MS calculations predict modulations in the near-edge intensity, which correspond directly to the DOS of the ionized atom. So you should be aware that these calculations are only an *interpretation* of what actually happens to the electron after it emerges above the Fermi level. Also, many calculations of ELNES only show reasonable agreement with experiment when the effect of the core hole is included, so we now need to explain this terminology.

Figure 40.10 gives a comparison of the experimental C-K edge ELNES from TiC with the results of calculations using several different potentials, none of which reproduce the experimental spectrum precisely but all of which capture some aspects of the general shape.

40.5.B Core Holes and Excitons

Having chosen an atomic potential, we actually determine the ELNES (using the MS approach) by calculating all possible inter- and intra-shell scattering events suffered by the electron after it emerges above the Fermi level. One of the problems that confuses this issue is that the ionization event results in a hole in the core shell which, of course, changes the atomic potential.

A bit more physics: The ionization process occurs in the time taken for the beam electron to traverse the diameter of the particular inner shell. We know that a 200-keV electron has a velocity, $v = 2.7 \times 10^8\ \mathrm{ms}^{-1}$, and the K-shell of oxygen, for example, has a diameter of ~0.01 nm, so the ionization process occurs over $\sim 10^{-19}$–10^{-20} s. By comparison, the atom stays in its excited state (which is a

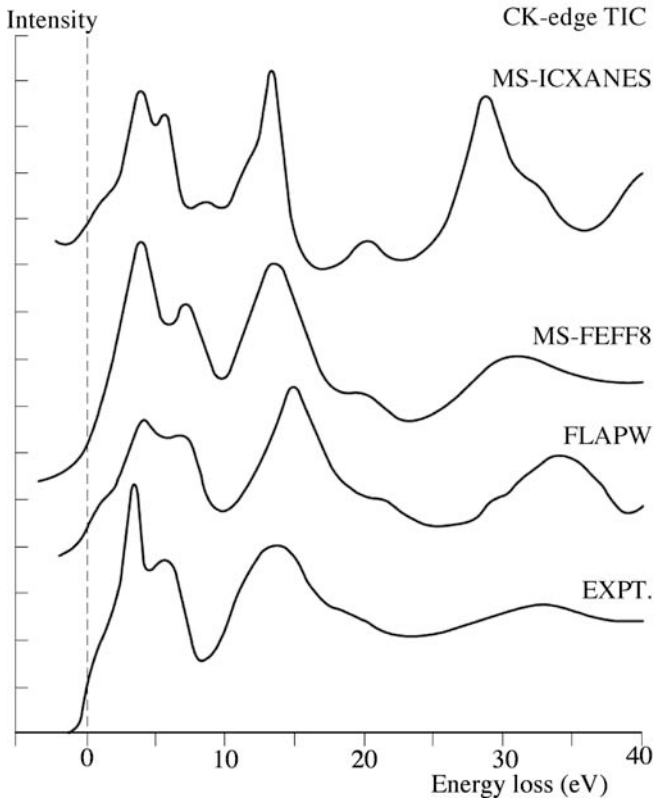


FIGURE 40.10. Comparison of the experimental C K-ELNES from TiC (*Expt.*) with the results of theoretical modeling calculations using both a band-structure code (*FLAPW*), and two different MS codes (*FEFF8* and *ICXANES*).

REAL MATERIALS

In ceramics and semiconductors, the ionized electron remains localized to the ionized atom. It may interact with the hole creating an electron-core hole bound state termed an exciton. Creation of an exciton may influence the ELNES; this remains a matter of some debate.

combination of the lifetimes of the excited electron and the lifetime of the hole in the inner-shell) for much longer, because the hole decays in $\sim 10^{-14}$–10^{-15} s. Because the lifetime of a hole is 10^5 × longer than the excitation process, the outermost electron states, including the final state of the excitation process, will experience an attractive potential because of the core hole, which behaves like an extra nuclear charge on the atom. So, in fact, all we do to compensate for this is assume that the ionized atom now has a nuclear charge of $Z+1$, rather than Z, because the missing electron lowers the shielding affect of the core electrons. This extra positive charge may be shielded by other electrons (e.g., weakly bound valence electrons) which will move in response to the existence of the hole and reduce its effect. Despite this screening (go back and check Section 3.5 to remind yourself what this term means), the core-hole potential will tend to attract the outer electron states more strongly. So the available final states for the ejected electron, in the presence of the long-lived hole, will tend to be more sensitive to the short-range environment of the excited atom and, of course, this will be reflected in the ELNES.

40.5.C Comparison of ELNES Calculations and Experiments

The 1982 seminal paper in the field of ELNES experiments on transition metals and oxides is by Leapman et al. For further examples, you should read the review articles which we mentioned at the end of Section 40.3. We'll just show a couple of examples here but the literature contains many. The difference due to different coordinations is obvious. The sharp peak at the Al L edge onset is thought to be an exciton. This effect is not well modeled by the theory, which otherwise makes a good match with the experimental data.

The electron energy-loss near-edge structure (ELNES) at the O K edge has been studied in yttria-stabilized zirconia (YSZ) (Ostanin et al.). The electronic structure of YSZ for compositions between 3 and 15 mol.% Y_2O_3 has been computed using a pseudopotential-based technique to calculate the local relaxations near the O vacancies. The results showed phase transition from the tetragonal to cubic YSZ at 10 mol.% of Y_2O_3, reproducing experimental observations. Using the relaxed defect geometry, calculation of the ELNES was carried out using the full-potential linear muffin-tin orbital method. The results show very good agreement with the experimental O K-edge signal, demonstrating the power of using ELNES to probe the stabilization mechanism in doped metal oxides.

If an atom exists in two different environments in a structure, then we can make the approximation that the ELNES is simply a linear superposition of the contributions from the two environments and experiments tend to support this simple approach.

It's perhaps best to conclude with the conclusion of Duscher et al.

> "We have reached a level of agreement between theory and experiment not achieved previously in such a range of different materials by including a localized core-hole effect. There was no significant difference between both the methods used, a core hole in an all-electron method and the $Z+1$ approximation. This approach is sufficiently robust to proceed to interface structures."

In other words, theoretical calculations are a well-established and useful field of ELNES research and this is explored much further in the companion text. One last and obvious point to note is that, because the ELNES signals are often quite strong, it is straightforward to map out different portions of the ELNES, thus imaging, in effect, changes in the DOS and localized variations in atomic bonding, with the usual high resolution expected of EFTEM images. An example is shown in Figure 40.11.

40.6 CHEMICAL SHIFTS IN THE EDGE ONSET

We can think of the atoms in our specimen as having different charges with respect to one another (which we otherwise call electronegativity). So any changes in the charge in different systems will lead to changes in the binding energies of the various (occupied and unoccupied) electron states and it's reasonable to ask if we could detect this binding-energy change in EELS. We already know that (for a hypothetical, single, isolated, hydrogen atom) the ionization-edge threshold energy is effectively the critical ionization energy, E_C. However, in a real material, the experimental-edge onset corresponds to the difference in energy between the initial state and the lowest unoccupied final state *in the presence of the core-hole*. More often than not, it is extremely difficult to determine accurately the threshold energy, which often lies above the experimental-edge onset. Changes in the effective charge on the atom affect the energies of both the initial and final states. Unlike the deep-lying core orbitals, the outer orbitals are easily influenced by factors, such as bonding. If we consider changing from a metal to an insulator, the presence of a band gap in the insulator will result in a shift of the edge

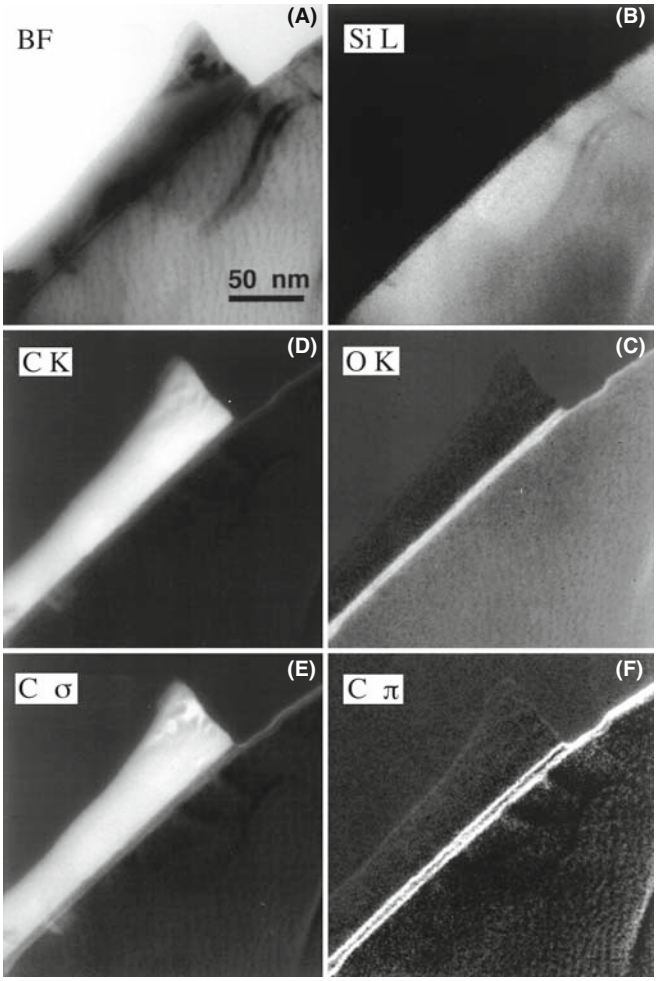


FIGURE 40.11. (A) TEM BF and (B–F) a series of energy-filtered images revealing the Si, C, and O elemental distributions and the carbon bonding maps at the interface between a diamond-like carbon film and a Si substrate. In the oxygen-rich amorphous layer at the interface, there is a double layer of carbon atoms that is primarily π-bonded (and possibly arises from carbon contamination in surface grooves at the interface) (F). The carbon film is predominantly σ-bonded (E), indicating a high degree of diamond-like character.

onset to higher energy loss. For example, the Al-L_{23} edge shifts from 73 eV in the metal to 77 eV in Al_2O_3, and we saw similar behavior for the Cu/CuO spectra in Figure 40.5B. Another example is provided by the C-K edges in Figure 40.5A, in which shifts in the π* peak position are easier to see than changes in the rather ill-defined edge onsets.

Similar edge-onset shifts are well known in XPS and are called chemical shifts. They are reasonably well understood and we can often predict them theoretically. However, the electron-excitation process in EELS is more complicated than X-ray induced ionization (detected in XPS), particularly the unavoidable presence of the core hole and the variable extent to which it is screened by the remaining electrons. Consequently, in comparison with X-rays, little systematic work on EELS chemical shifts has been done, apart from fingerprinting experimental spectra and comparing them with reference materials. While we can correlate the edge-onset energies with such variables as oxidation states, atomic charge, and coordination, there is room for considerable improvement in calculation of the true edge-onset energies. It's worth noting that the possibility of a chemical shift in the edge may make it difficult to interpret ELNES intensity changes detected by difference techniques (such as the example shown back in Figure 40.7B). However, careful experimentation should minimize this danger.

40.7 EXELFS

If, after an atom is ionized, the ejected electron does not fill an empty state but escapes outside the atom, then it acts like a free electron (typically with energy >50 eV). We can interpret this excess energy as an electron wave, which can be diffracted by the atoms in the structure around the original ionization site. Because the electron has higher energy than those which gave rise to multiple-scattering ELNES, the diffraction is assumed to be a single-scattering event, as shown back in Figure 40.1B, and this diffraction causes oscillations on the otherwise smooth DOS. We call these oscillations extended energy-loss fine structure or EXELFS. As with any diffraction event, there is information in these EXELFS ripples about atomic positions and the atomic information comes from a relatively short range (the first few nearest neighbors) since this weak electron doesn't scatter from more distant atoms.

ELNES AND EXELFS II

ELNES is multiple scattering and EXELFS is single scattering, although the two phenomena overlap since, e.g., the L_1 ELNES peak is often far enough past the edge onset to be included in the EXELFS.

The EXELFS modulations start about 50 eV above the ionization-edge energy, are each 20–50 eV wide, as shown in Figure 40.12A, and occur over several hundred eV. EXELFS is closely analogous to the oscillations seen in the extended X-ray absorption (edge) fine structure (EXAFS) in synchrotron X-ray spectra and is one reason why EELS has long been described as a synchrotron in a TEM (albeit much cheaper than your typical synchrotron). One significant difference is that EXAFS results in complete photoabsorption of the incident X-ray while EXELFS involves absorption of only a small fraction of the energy of the beam electron. We can carry this analogy a little further. Both EXAFS and EXELFS give us structural information from materials in which there are strong, local, atomic correlations. Both techniques are atom specific so, in principle, we can solve even the most complex multi-component

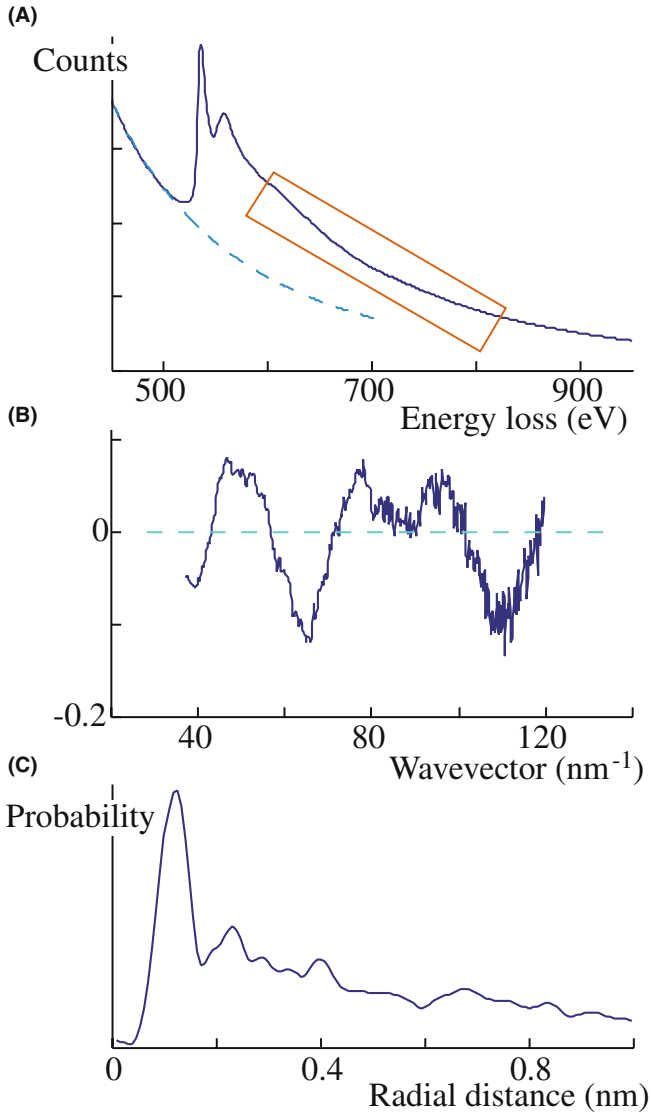


FIGURE 40.12. (A) EXELFS modulations on an ionization edge. (B) From the oscillations in the post-edge spectrum it is necessary to transfer the data to *k*-space before (C) Fourier transforming the data to produce a radial-distribution function.

structures, if the information around all the atoms can be accessed.

There are, however, limitations to conventional EXAFS.

- We can't easily access the K-edges below 3 keV since X-ray absorption at these low energies requires thin specimens for transmission-EXAFS and a low absorption atmosphere over the entire X-ray beam path (source-specimen-detector).
- As you know from the XEDS chapters, X-rays cannot easily be focused to a sub-micrometer spot, so EXAFS has a relatively low spatial resolution, although this is constantly improving. EXELFS offers us the unique ability to obtain atomic and electronic structure with nanometer-scale spatial resolution.
- Since TEMs operate in high vacuum and use thin specimens, EELS is more naturally suited to K-edge analysis of low-*Z* elements (as well as L-edge analysis of higher-Z elements) than low-energy EXAFS.

There is the usual price to pay for high spatial resolution in that the EXELFS signal is noisy and you'll find that extracting high-quality atomic information is much more challenging than for EXAFS where there is no shortage of signal.

40.7.A RDF via EXELFS

With EXELFS we can determine the partial radial distribution function (RDF) around a specific atom, and we are not restricted to the heavier atoms ($Z>18$) needed for EXAFS. So there is great potential for studying materials, such as low-*Z* glasses, amorphous Si, bulk metallic glasses, and quasicrystalline structures (both of the latter two often contain relatively low-Z elements such as Be, Mg, Al, P). In particular, since glasses lack any long-range periodic structure, we are limited in the techniques to determine their atomic structure. As you've already seen back in Section 18.7, diffraction of electrons (or X-rays or neutrons) from glass provides only diffuse information. To get atomic-structure information from glasses, you have to employ resonance signals from the Å-level and EXELFS can do that. The high spatial resolution of EXELFS is obviously advantageous and all your data can be compared with your images and the rest of the TEM-based information that you acquire from the analyzed volume. However, you can't get good EXELFS unless your specimen is very thin and you'll also have to consider phase effects, which are averaged out in EXAFS. Despite these apparent advantages of EXELFS, RDF work continues to be dominated by synchrotron X-ray sources because of the intensity of the signal. If you're interested in pursuing this (and EXELFS for that matter), a good place to start is the text by Koningsberger and Prins.

DECONVOLUTION AGAIN

Deconvolution is always the first step if your specimen isn't thin enough, i.e., if the plasmon peak is greater than 10% of the ZLP.

Experimentally, it's not easy to see the EXELFS modulations because they are only ~5% of the edge intensity, and so you need good counting statistics. This is one of the rare cases where you might find a thermionic source useful because it can deliver more

total current than a FEG and, for this application, energy resolution is not so important. TEM diffraction mode will also increase your total signal intensity. Either way, however, you pay a price in terms of a loss of spatial resolution and an increased chance of specimen damage. If you need the best spatial resolution, a FEG STEM should (as usual) be your instrument of choice.

So, we're interested in EXELFS because of the structural information contained in the weak intensity oscillations. To extract this information, you can use the commercial Gatan software (see Section 1.6) or the EXELFS version of Rehr's FEFF code (URL #6). Public-domain EXELFS software is also available at URL #7.

You first have to ensure that your spectrum contains only single-scattering information, otherwise the plural-scattering intensity may mask the small EXELFS peaks. Deconvoluting the point-spread function may also help sharpen the faint modulations.

Next, you have to remove the background if it wasn't done prior to deconvolution. Then your spectrum intensity has to be converted to an electron-wave function in k-space (reciprocal space) where

$$k = \frac{2\pi}{\lambda} = \frac{[2m_0(E - E_K)]^{\frac{1}{2}}}{h} \qquad (40.1)$$

where E_K is the edge onset energy, E is the energy of the ejected electron, of wavelength λ, and the rest of the terms have their usual meaning. The electron-wave interference gives periodic intensity maxima in k-space when

$$\left(\frac{2a}{\lambda}\right)2\pi + \Phi = 2\pi n \qquad (40.2)$$

Here a is the distance from the ionized atom to the first scattering atom, and Φ is the phase shift that accompanies the scattering. Therefore, we expect to see periodic maxima occurring for $n = 1, 2$, etc., and for different interatomic spacings. We obtain the atomic spacing by Fourier transforming the k-space modulations to give the RDF, originating at the ionized atom. Obviously, these two equations don't tell the full story and there is more about data analysis in Egerton's book.

When we have the RDF, we ought to be able to determine the local atomic environment, if the various interferences can be discriminated and identified. Peaks in the RDF indicate the probability of an atom occurring a certain distance from the ionized-atom site. Figure 40.12 shows a summary of the EXELFS data extraction technique. Despite the low signal problems, EXELFS studies appeared at the earliest stages of EELS research (see, e.g., Leapman and Cosslett) and references to the technique have continued to surface in the literature over the intervening decades. For example, Sikora et al. have compared EXELFS and EXAFS applied to crystalline materials. Alamgir et al. have made comparisons of the two techniques for the study of slow-cooled, bulk metallic glasses (a fascinating new range of materials). Figure 40.13 shows the extraction of EXELFS data from the P-K edge (inaccessible in a synchrotron spectrum) from Pd-Ni-P, a model amorphous metal. The various steps involve first the isolation of the fine structure (Figure 40.13A) and expressing it as a function of momentum transfer (k), $\chi(k)$, from beyond the P K-edge (Figure 40.13B). The Fourier transform of $\chi(k)$, FT[$\chi(k)$], is proportional to the partial radial distribution function of atoms around P (Figure 40.13C). Upon back-Fourier transformation of the first peak FT[$\chi(k)$], the contribution to $\chi(k)$ from the first coordination shell around P in these glasses is determined (Figure 40.13D). Similar exercises could be performed from the second and higher shells although the quality of the signal degrades rapidly. These data are then fitted with calculated $\chi(k)$ functions of various model structures using the multiple-scattering ab-initio code FEFF7 and a possible model of the coordination of Pd and Ni atoms around the P atoms is created.

40.7.B RDF via Energy-Filtered Diffraction

RDF data acquired through EXELFS complement another TEM method of acquiring RDF information. This involves energy filtering of SADPs by scanning the pattern across the entrance aperture to the PEELS using post-specimen scan coils. (See, e.g., McBride and Cockayne; see also Section 18.7.) Effectively, a full spectrum is available at each scattering angle, but, in fact, only the zero-loss (ideally only the elastic) electrons are required. The plot of the ZLP intensity as a function of scattering angle constitutes a line profile across a filtered DP from which the RDF can be extracted. This process does not have the spatial resolution of EXELFS, since typical SADPs are integrated over ~0.2–1 μm^2, but the signal is much stronger. Accuracies of ± 0.001 nm in nearest-neighbor distances can be obtained, and the process is rapid enough to be performed on-line.

40.7.C A Final Thought Experiment

ELNES and EXELFS are really quite remarkable demonstrations of quantum theory and the wave-particle duality. Consider that the EXELFS part of the spectrum only contains electrons that have been scattered by electrons in the specimen atoms, and yet we are able to deduce information about what happened to those atoms *long after the beam-specimen interaction occurred* and also deduce where the scattering atom sits in the structure!

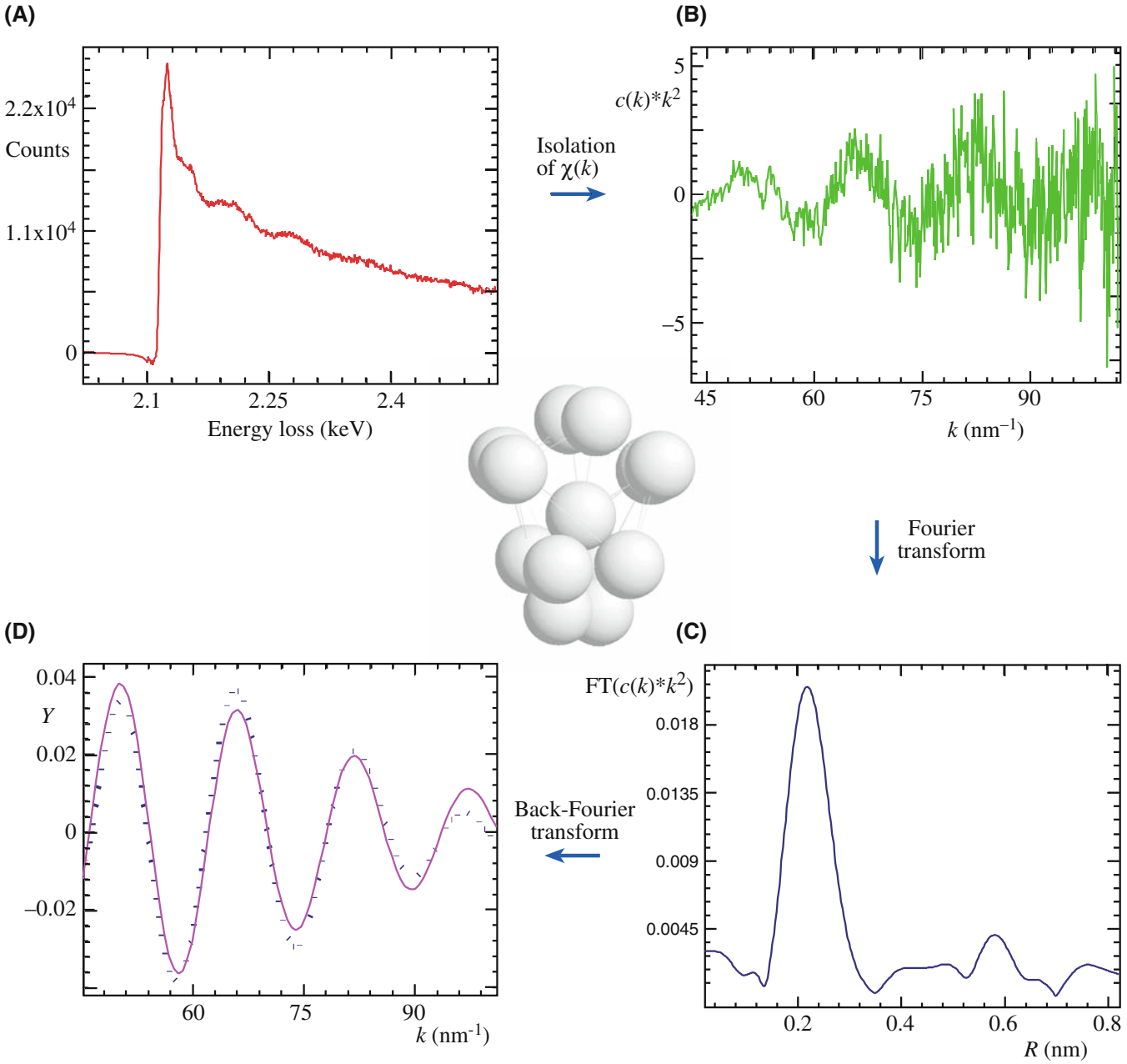


FIGURE 40.13. The EXELFS analysis of $Pd_{30}Ni_{50}P_{20}$ bulk metallic glass. (A) The pre-edge, background-subtracted P K-edge. (B) The isolated $\chi(k)$ data. (C) The Fourier transform of $\chi(k)$ to radial space, FT[$\chi(k)$], and (D) the back-Fourier transform of the first peak of FT[$\chi(k)$] back to k-space (dots) and fitting with a calculated function for a tetragonal dodecahedron (dashed line). The model in the center is a tetragonal dodecahedron with a P atom at the center surrounded by a first nearest-neighbor shell of Pd and Ni atoms deduced from the EXELFS.

An (approximately wrong) particle-based analogy would be to imagine that we are catching bowling balls that have been thrown at pins, arranged in a certain pattern. (Although instructive, this exercise is best carried out as a thought experiment.) From the velocity (energy) of the balls that we catch, we are able not only to identify the weight of the pin that was hit (i.e., identify the characteristic ionization edge), but we can also deduce how the pin fell down and where it rolled (the ELNES). Furthermore, we can also work out the spatial arrangement of the surrounding pins that didn't fall down (the EXELFS).

So how does the beam electron know where the core electron went after it left the inner shell? The answer lies in the fact that the bowling ball (particle) analogy is totally inadequate. In fact, only certain electron transitions are allowed and the beam electron can therefore only transfer certain quantized energies to the core electron, not a continuum of possible energies. So the beam electron does, in effect, 'know' the possible final state of the core electron since it reflects that state in its energy loss. (If you really understand this, then you should pat yourself on the back.)

40.8 ANGLE-RESOLVED EELS

Most of the time so far, we've been talking about gathering spectra and images by sending the direct beam into the spectrometer and splitting it up into its component energies. This is often called *spatially resolved* EELS since we map out a specific region or gather individual spectra from different spatial locations on the specimen. However, we have occasionally mentioned that the *angle* of scatter of the energy-loss electrons is important, and there is a whole field of EELS research that studies angle-resolved spectra. To do this, we just scan the DP across the PEELS entrance aperture and gather spectra at different angles as for the RDF measurements that we just described. However, rather than studying the energy of electrons primarily, this technique emphasizes the determination of the *momentum* of the energy-loss electrons. Momentum-transfer studies were pioneered by Silcox and co-workers. Now with FEG STEMs you can get even more information about the symmetry of electronic states, which complements spatially resolved ELNES (e.g., Wang et al.).

Because of such angular effects, the size of the spectrometer entrance aperture and/or the collection angle β may influence the details of ELNES. If the final state of the ejected electron has a definite directionality, as it will in an anisotropic crystal, the ELNES for such a specimen will depend on both the scattering angle θ and the crystal orientation. The classic paper on orientation or momentum dependence is the study of graphite and BN (Leapman et al.).

There are various ways of performing angle-resolved EELS and Botton and co-workers describe many methods and applications. For a given energy loss and a given specimen orientation, the momentum transfer at zero scattering angle, θ, is parallel to the electron beam ($q_{||}$). As θ is increased, a component perpendicular to the beam is introduced ($q_{\perp}$) and, at approximately θ_E, $q_{\perp}$ becomes dominant. To obtain angle-resolved ELNES, we have to measure the spectra as a function of the angle between a crystal direction and the direction of momentum transfer.

First, you can keep the orientation fixed and change the collection aperture. Figure 40.14 shows angle-resolved ELNES of a very thin (~ 30 nm) graphite flake obtained by changing the size of the collection aperture in a STEM. The π* states are parallel to the *c*-axis, which in this case is parallel to the electron beam. Therefore, the corresponding π* peak intensity is larger, relative to the σ* states, when we have a small collection aperture and $q_{||}$ dominates. Similar effects would be obtained if the orientation of the graphite planes to the electron beam was changed.

Second, you can keep your collection angle fixed at a small angle $< \theta_E$ and measure the ELNES as a function of θ. This is particularly easy with energy filtering which can display angular scattering distributions of specific energy-loss electrons.

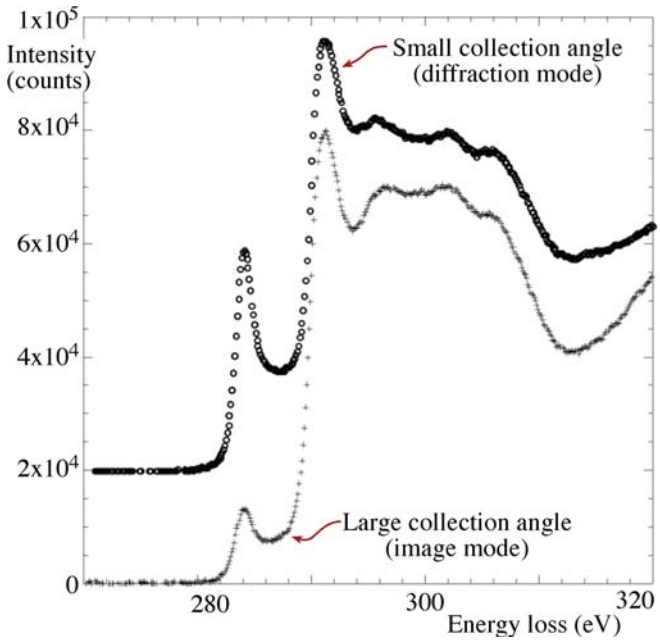


FIGURE 40.14. Image mode EEL spectra of the C-K edge in graphite showing changes in the relative intensities of the π* and σ* peaks under two different collection angles due to the directional scattering variation from the sp2 and sp3 bonds in the graphite.

You can also use the '45° method' (Botton) which combines tilting the specimen at 45° to the principal axis and measuring the ELNES at ±θ, where θ is chosen to select both $q_{\perp}$ and $q_{||}$. If you make the beam convergent (e.g., in STEM mode for high spatial resolution) then the angular resolution of this technique is reduced.

One practical aspect of angle-resolved EELS is the study of Compton scattering, which is the ejection of outer-shell electrons by high-energy photons or electrons. We can detect these Compton-scattered electrons by observing the EELS spectrum at a high scattering angle (θ ~100 mrad), either by displacing the objective aperture to select an off-axis portion of the diffraction pattern or by tilting the incident beam. This process has been used to analyze the angular and energy distribution of Compton-scattered electrons and determine bonding information, since the Compton-scattering process is influenced by the binding energy.

40.9 EELS TOMOGRAPHY

We've talked about tomography in various parts of the book and how taking a series of images at different tilts allows you to reconstruct 3D information about features within your specimen. You should also read the chapter in the companion text which gives an in-depth review of the various experimental challenges for this kind of imaging.

Just as for X-ray images, it is possible to use a full-tilt series of EFTEM images to build up a full 3D image

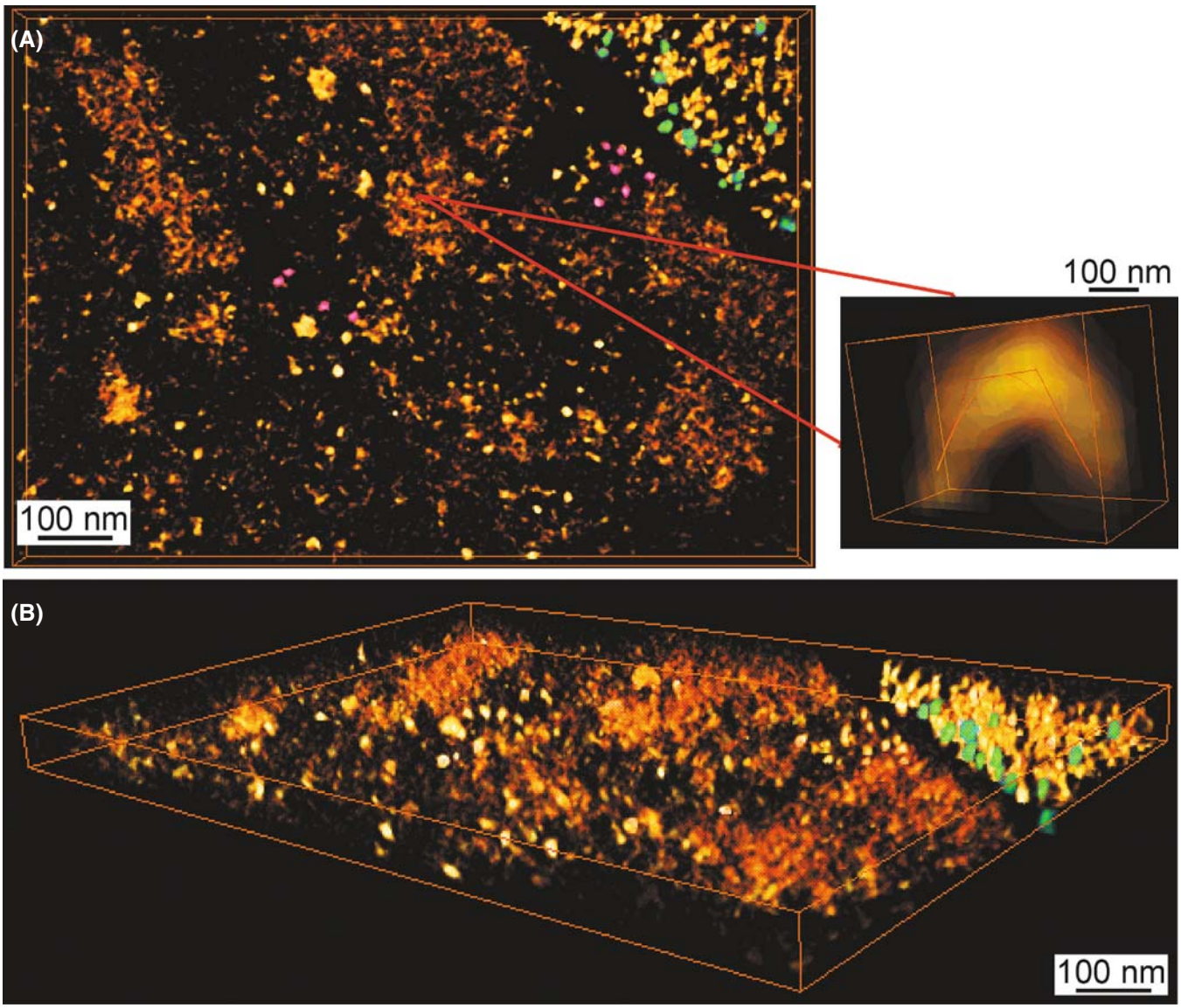


FIGURE 40.15. (A) 2D projection (top view) and (B) energy-filtered, 3D-tomographic image reconstructed from multiple P-$L_{2,3}$ edge filtered images showing the phosphorus distribution in an unstained plastic section of freeze-substituted *Drosophila* larva. The principal region is part of a cell nucleus and the top right is a region of cytoplasm containing ribosomes. The phosphorus distribution reflects the distribution of nucleic acid. Ribosomes (colored green) are known to contain about 7000 P atoms in their RNA. Another series of phosphorus-containing dense particles of unknown origin are present in the nucleus (see enlarged inset).

from a set of 2D projections. (See the paper by Midgley and Weyland.) Because ionization-edge composition images are not susceptible to significant contrast changes as a function of tilt, they are ideal for tomographic reconstruction which reveals the surface features, growth angles/facets, and other features that are not easily obtained from the usual 2D projection image. EFTEM tomography is similar to XEDS tomography which we didn't discuss: EFTEM tomography has the advantage of much quicker generation because XEDS tomography is only feasible via a series of tilted STEM images which take much longer to produce.

Figure 40.15A shows one of the P-L ionization-edge images from a tilt series and Figure 40.15B shows the reconstructed tomographic image of the distribution of P in a cell from a fruit fly. In this area, the biologists are well ahead of the materials scientists and Leapman et al., in 2004, gave a fine example of the application of EELS tomography to discerning the 3-D shape of ribosomes (not quite EELS of eels, but of nematode worms, which look a bit like eels!). In contrast to ionization-edge images, plasmon images retain significant diffraction contrast and so are less useful for tomography.

EFTEM tomography is an area that will see increasing applications, particularly as nanotechnologists pursue their dream of manufacturing devices in a controlled manner from the atom/molecule upward. The ability of EFTEM (and to a lesser extent, XEDS) to reveal the actual shape (combined with the quantitative local chemistry) of quantum dots, gate oxides, and other sub-nanometer fabrications will seriously enhance TEM's role in this growing field.

The ability to extract 3D information by tomographic methods is just one of several examples of the extraordinary advances that have taken place in the TEM field since the first edition of this book was published more than 10 years ago.

It is worth concluding now, as we did then, that we encourage you to experiment with your TEM at all opportunities and never think that there is nothing new to discover. The current generation of students growing up with an expectation of, and familiarity with, full computer control of everything and access to information immediately from anywhere should be able to combine these skills to make TEMs perform in ways that the former generation of more manually oriented TEM operators could never dream of. For example, we haven't even mentioned time-resolved EELS although the strong low-loss signal and efficient collection means that gathering spectra with millisecond or even microsecond resolution is not out of the question and ultra-fast (nanosecond) TEM imaging using laser-excited sources is fast becoming a reality. No doubt there will be a serious need for yet another edition of 'TEM: a text for nanotechnologists' a decade hence, which we sincerely hope will be written by some of those who started their careers by reading this book, rather than by those who wrote it.

CHAPTER SUMMARY

You can appreciate now that there is a wealth of fine detail in the EEL spectrum beyond the relatively strong plasmon peaks and the ionization edges. To extract this information, you need a single-scattering (deconvoluted) spectrum and occasionally some sophisticated mathematical analysis. Interpretation of the data is still limited by our lack of knowledge of the physics of the electron-specimen interactions. However, considerable research is going on into EELS fine-structure studies which are the future of the technique. Both the experimental methods and the theoretical calculations are still developing. We have introduced several specialized topics

- Energy-loss near-edge structure.
- Extended energy-loss fine structure.
- RDF determination.
- Angle-resolved (momentum-transfer) EELS.
- EELS tomography.

However, we have really only given you a suspicion of the potential of these topics. If EELS becomes a technique you use in your research, we recommend watching its development and that of related techniques in the reference sources we have given you in the last four chapters, particularly the quadrennial EELS workshops (see Chapter 37), conferences such as the biannual Frontiers of Electron Microscopy in Materials Science (FEMMS), and the proceedings of the various national and international microscopy and analysis societies which, if you haven't joined by now, you should do so immediately!

BOOKS AND REVIEWS

Brydson, RMD, Sauer, H and Engel, W 2004 *Probing Materials Chemistry using ELNES* in *Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS Atlas* 2nd Ed. 223–270 Ed. CC Ahn Wiley-VCH Weinheim Germany. Comprehensive review of chemical information via the ELNES spectrum; exceptional bibliography.

Egerton, RF 1996 *Electron Energy-Loss Spectroscopy in the Electron Microscope* 2nd edition Plenum Press New York. Everything you need to know on fine structure.

Finnis, MW 2003 *Interatomic Forces in Condensed Matter* Oxford University Press, New York. Essential reading if you are modeling spectra.

Keast, VJ, Scott, AJ, Brydson, R, Williams, DB and Bruley, J 2001 *Electron Energy-Loss Near-Edge Structure – a Tool for the Investigation of Electronic Structure on the Nanometre Scale* J. Microsc. **20** 135–175. Broad-based review with lots of examples.

Koningsberger, DC and Prins, R 1988 *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS and XANES* Wiley New York. Probably more than you'll ever want to know about the X-ray analogs of EELS.

Raether, H 1965 *Electron Energy-Loss Spectroscopy* in Springer Tracts in Modern Physics Springer-Verlag New York. *The* source if you really want to know the physics of EELS.

CALCULATIONS AND TECHNIQUE

Durham, PJ, Pendry, JB and Hodges, CH 1982 *Calculation of X-ray Absorption Near Edge Structure, XANES* Comp. Phys. Comm. **25** 193–205.

Duscher, G, Buczko, R, Pennycook, SJ and Pantelides, ST 2001 *Core-Hole Effects on Energy-Loss Near-Edge Structure* Ultramicrosc. **86** 355–362.

Leapman, RD and Cosslett, VE 1976 *Extended Fine Structure Above the X-ray Edge in Electron Energy Loss Spectra* J. Phys. D: Appl. Phys. **9** L29–L32.

Midgley, PA and Weyland, M 2003 *3D Electron Microscopy in the Physical Sciences: the Development of Z-Contrast and EFTEM Tomography* Ultramicrosc, **96** 413–431.

McBride, W, and Cockayne, DJH 2003 *The Structure of Nanovolumes of Amorphous Materials* J. Non-Cryst. Sol. **318** 233–238.

MOMENTUM TRANSFER STUDIES

Botton, GA, Boothroyd, CB and Stobbs, WM 1995 *Momentum Dependent Energy Loss Near Edge Structures Using a CTEM: the Reliability of the Methods Available* Ultramicrosc. **59** 93–107.

Leapman, RD, Grunes, LA and Fejes, PL 1982 *Study of the L_{23} Edges in the 3d Transition Metals and Their Oxides by Electron-Energy Loss Spectroscopy with Comparisons to Theory*. Phys. Rev. **25**(12) 7157–73.

Leapman, RD and Silcox, J 1979, *Orientation Dependence of Core Edges in Electron-Energy-Loss Spectra from Anisotropic Materials* Phys. Rev. Lett. **42** 1361–1364.

Wang, YY, Cheng, SC, Dravid, VP and Zhang, FC 1995, *Momentum-Transfer Resolved Electron Energy Loss Spectroscopy of Solids: Problems, Solutions and Applications* Ultramicrosc. **59** 109–119.

APPLICATIONS

Alamgir, FM, Jain, H, Williams, DB and Schwarz, R 2003 *The Structure of a Metallic Glass System Using EELFS and EXAFS as Complementary Probes* Micron **34** 433–439.

Batson, PE 1993 *Carbon 1 s Near-Edge-Absorption Fine Structure in Graphite* Phys. Rev. B **48** 2608–2610.

Botton, GA 2005 *A New Approach to Study Bond Anisotropy With EELS* J. Electr. Spectr. Rel. Phen. **143** 129-137.

Botton, GA, Gupta, JA, Landheer, D, McCaffrey, JP. Sproule, GI and Graham, MJ 2002 *Electron Energy Loss Spectroscopy of Interfacial Layer Formation in Gd_2O_3 Films Deposited Directly on Si (001)* J. Appl. Phys. **91** 2921–2924. Bond changes at oxide interfaces.

Bruley, J, Williams, DB, Cuomo, JJ and Pappas, DP 1995 *Quantitative Near-Edge Structure Analysis of Diamond-like Carbon in the Electron Microscope Using a Two-Window Method* J. Microsc. **180** 22–32.

Keast, VJ, Bruley, J, Rez, P, Maclaren, JM and Williams, DB 1998 *Chemistry and Bonding Changes Associated with the Segregation of Bi to Grain Boundaries in Cu* Acta Mater. **46** 481–490.

Leapman, RD, Kocsis, E, Zhang, G, Talbot, TL and Laquerriere, P 2004 *Three-Dimensional Distributions of Elements in Biological Samples by Energy-Filtered Electron Tomography* Ultramicrosc. **100** 115–125.

McComb, DW, Craven, AJ, Hamilton, DA and MacKenzie, M 2004 *Probing Local Coordination Environments in High-k Materials for Gate Stack Applications* Appl. Phys. Lett. **84** 4523–4525.

Muller, DA, Subramanian, S, Batson, PE, Silcox, J and Sass, SL 1996 *Structure, Chemistry and Bonding at Grain Boundaries in Ni_3Al-I. The Role of Boron in Ductilizing Grain Boundaries* Acta Mater. **44** 1637–1645.

Ostanin, S, Craven, AJ, McComb, DW, Vlachos, D, Alavi, A, Paxton, AT and Finnis, MW 2002 *Electron Energy-Loss Near-Edge Shape as a Probe to Investigate the Stabilization of Yttria-Stabilized Zirconia* Phys. Rev. B **65** 224109–117.

Sikora, T, Hug, G, Jaouen, M and Rehr, JJ 2000 *Multiple-Scattering EXAFS and EXELFS of Titanium Aluminum Alloys* Phys. Rev. B **62** 1723–1732.

URLs

1) www.cemes.fr/~eelsdb
2) www.flapw.de
3) www.physik.uni-augsburg.de/~eyert/aswhome.shtml
4) www.castep.org
5) http://hermes.phys.uwm.edu/projects/elecstruct/mufpot/MufPot.TOC.html
6) http://feff.phys.washington.edu
7) www.cemes.fr/epsilon/home/main.php

SELF-ASSESSMENT QUESTIONS

Q40.1 Why does the ionization edge extend beyond the critical ionization energy (the edge onset) to give ELNES and EXELFS, rather than exist simply as a peak at the critical ionization energy?
Q40.2 What is the Fermi level/Fermi surface and why is it crucial to our understanding of the energy-loss process?
Q40.3 What is the density of states (DOS) and why are there both filled and unfilled DOS?
Q40.4 Relate the K, L, etc., core shells to the principal quantum numbers (n).
Q40.5 State the Pauli exclusion principle and explain why this is relevant to ELNES.
Q40.6 What is spin-orbit splitting and why is this relevant to ELNES?
Q40.7 What is the dipole-selection rule and why is this relevant to ELNES?
Q40.8 Why does the ionization edge onset for a specific elemental core loss sometimes shift when that element is bonded differently?
Q40.9 What is XANES, how is it detected, and what is its relation to ELNES?
Q40.10 Why do bonding changes change the ELNES?
Q40.11 What useful information is contained in the EXELFS spectrum?
Q40.12 Why is EXELFS such a challenging technique to apply?
Q40.13 What is an exciton?
Q40.14 What is a core hole?
Q40.15 Why is there bonding information in both the low-loss and high-loss spectrum?
Q40.16 Distinguish angle-resolved and spatial-resolved EELS.
Q40.17 Why is angle-resolved EELS linked to the concept of electron momentum transfer?
Q40.18 What is the RDF, why is it useful, and how can you measure it?
Q40.19 Why would you want to calculate the ELNES intensity?
Q40.20 What is Compton scattering and how can we study this in EELS?
Q40.21 Under which circumstance would you choose to use an ELNES spectrum as a fingerprint and what precautions should you take when drawing conclusions from a potential match?

TEXT-SPECIFIC QUESTIONS

T40.1 Distinguish single, multiple, and plural scattering for EELS. How do these definitions compare with scattering terms used in high-resolution imaging?
T40.2 Figure 40.1 gives an electron-wave description of the generation of ELNES and EXELFS. Can you use a particle analogy to describe the process?
T40.3 Is Figure 40.2 drawn for a crystalline metal or an amorphous semiconductor? Explain your answer and thus indicate how the figure would change if the other kind of material were being illustrated.
T40.4 In Figure 40.3 there appears to be no intensity in the ionization edge corresponding to the filled states. Why is this? In a real spectrum there would indeed be intensity before the ionization edge. What would cause this?
T40.5 Why does the Cu L edge in Figures 40.4 and 40.5B exhibit no intense white lines at the edge onset like the rest of the transition metal series in Figure 40.4?
T40.6 In old specimens and older TEMs, the diamond K edge (like in Figure 40.5A) sometimes shows residual intensity preceding the ionization-edge onset, at roughly the same energy as the π^* sp2 peak in the graphite and C_{60} edges shown above. Since diamond has no sp2-bonded carbon, can you speculate what might be giving rise to this intensity?
T40.7 Why are the muffin-tin potential wells in Figure 40.9 symmetric for the metal but asymmetric for the oxide?
T40.8 Look at the comparison of calculated and experimental spectra in Figure 40.10. These calculations were done more than a decade ago. Go on the Web and see if you can find better examples of calculated edge shapes that show a better fit to experimental spectra. If you can't, what conclusions can you draw about calculating ELNES. If you can, what different conclusions can you draw?
T40.9 How do you think that correcting the spherical aberration in the objective lens will improve the study of energy-loss fine structure? Do you think the addition of electron gun monochromators will affect the study of this same phenomenon?
T40.10 What crucial information can be gained about the behavior of semiconductor interfaces and gate oxides via ELNES? (Hint: Google PE Batson and read his papers.)
T40.11 Under what circumstances would you choose an MO rather than an MS approach to calculating the near-edge spectrum?
T40.12 List the principal differences between FLAPW, ASW, CASTEP, and LKKR.
T40.13 ELNES fingerprinting can distinguish different mineral species as in Figure 40.8. Why should we ever bother to use XEDS to study the same problem? Does the beam-sensitivity of many minerals have a role to play in deciding what technique to use? If so, explain what.
T40.14 Why does the signal in Figure 40.12B become noisier at larger wavevectors?

T40.15 Given what you know about the crystal structures of graphite and diamond, would you expect either of their energy-loss spectra to be sensitive to crystallographic orientation? If so, how do you think the fine scale features of the relevant spectrum in Figure 40.5A might change with orientation? (Hint: look at Figure 40.14.)

T40.16 Compare and contrast EXAFS and EXELFS for studying short-range atomic structures. Why would you use EXELFS when TEM diffraction patterns give similar short-range atomic structural information?

T40.17 In addition to ELNES and energy-filtered diffraction for RDF determination, can you think of other ways to explore the structure of glasses using TEM?

T40.18 For both momentum-resolved and tomographic EELS, we have to tilt the specimen considerably. What are the experimental challenges to doing this and how might they be overcome?

T40.19 If the low-loss spectrum reveals the valence states of the atoms in the specimen why do we not use this part of the spectrum more often for bonding studies but instead use ELNES which only explores the unfilled DOS (i.e., the electrons that aren't there)?

T40.20 Explain why K-shell ionization results in a hydrogenic edge.

T40.21 Explain why L shell ionization gives L_1, L_2, and L_3 edges.

T40.22 Why is the L_1 edge rarely visible, thus leaving the usual L edge as the $L_{2,3}$ in spectra from transition metals?

T40.23 Similarly, why is the $M_{4,5}$ edge the expected M edge in the rare earths?

T40.24 Explain why EELS edges and X-ray absorption edges are effectively the same phenomenon.

Index

A
A_3B ordered fcc, 262–263
Aberration, 6–7, 54, 61, 64, 81, 82, 91, 92, 98, 99, 103, 105, 106, 107, 108, 109, 110–111, 133, 141, 148, 150, 155, 158, 161, 162, 295, 372, 373, 485, 488, 493, 513, 522, 525, 590, 636, 663, 682, 683, 684, 691, 692, 696, 707, 735
  aberration-free focus, 492
  coma, 494, 498
  function, 485, 495
  See also Chromatic aberration; Spherical aberration
Absorption, 241, 413–414, 654–656
  absorption-free intensity ratio, 190
  anomalous, 319, 413, 414, 416, 449
  contrast, 374
  correction, 601, 653, 654–656, 671
  distance, 435
  edge, 588, 607, 613, 645, 717, 736
  of electrons, 118, 374
  extrapolation techniques for correction, 671, 725, 726, 727, 728
  parameters, 434, 458
  path length, 655
  of X-rays, 28, 294, 599, 600, 634, 641, 653, 656, 660, 671, 694, 717, 742, 751, 752
Accelerating voltage
  calibration of, 149–150, 168–169
  continuous kV control for CBED, 217–218, 341
  effect on Bloch waves, 252–253
  effect on EELS, 190, 681–682, 684
  effect on Ewald sphere, 355
  effect on X-rays, 615
Adaptive filter, 572, 573
Airy disk, 32, 33, 107, 109, 110
ALCHEMI, 517, 657–658, 669
Allowed reflections, 258, 259, 265, 287, 348, 349, 350
Amorphous
  carbon, 35, 163, 183, 185, 295, 374, 551, 554, 612, 719, 722
    See also Holey carbon film
  germanium, 587–588
  layer, 502, 571, 670, 751
  materials, 197, 293–295, 373, 415, 502, 504, 528, 703, 741, 748
    specimen, 377–378
Amplitude contrast, 106, 371–386, 411, 458, 504, 505
  See also Contrast
Amplitude of diffracted beam, 223
Amplitude-phase diagrams, 31–32
Analog
  collection, 102
  to digital converter, 607
  images, 115, 125
  pulse processing, 591
Analog dot mapping, 617–618
Analytical electron microscopy (AEM), 7, 25, 53, 54, 62, 66, 75, 76, 80, 81, 82, 83, 97, 99, 103, 111, 121, 132, 133, 138, 143, 144, 150, 184, 185, 186, 352, 581, 584, 586, 588, 589, 590, 592, 593, 594, 595, 597, 598–600, 605, 606, 607, 608, 609, 611, 612, 613, 614, 615, 616, 617, 618, 625, 626, 627, 628, 630, 633, 634, 639, 647, 648, 651, 655, 663, 672, 674, 682, 690, 693, 694, 696, 706, 717
Angle, 26–27, 83, 685–688
  See also Bragg; Collection semiangle; Convergence semiangle; Incidence semiangle
Angle-resolved EELS, 755
Angular-momentum quantum number, 744
Annular condenser aperture, 157
Annular dark field (ADF), 122, 160, 161, 162, 329, 373, 376, 377, 379, 380, 384, 385, 635, 659
  image, 161
  See also Dark field (DF), detector
Anodic dissolution, 178
Anomalous X-ray generation, 641
  See also Absorption
Anticontaminator, 130
Anti-phase (domain) boundaries (APB), 229, 263, 420, 426, 427, 428, 429, 434, 503
Aperture
  alignment of C2, 147
  condenser (C2), 111, 157, 493
  differential pumping, 131, 683, 687
  function, 485, 486, 488
  objective, 101, 109, 111, 152, 154–158, 161, 162, 165, 167, 207, 278, 332, 372–373, 375, 376–379, 381, 382, 385–386, 389, 390, 396, 408, 411, 413, 448, 466, 489, 500, 511, 512–514, 516, 517, 519, 520, 521, 528, 529, 535, 539, 540, 551, 552, 573, 600, 670, 686–688, 691, 706, 721, 755
  virtual, 152, 154, 491, 492
  virtual C2, 152, 154
  See also Diaphragm
Artifact
  in EELS, 730
  in image, 9, 10, 542
  of specimen preparation, 190
  X-ray peak, 605, 606–607, 613, 625, 628, 672
Artificial color, 124, 555
Artificial superlattice, 264, 265, 415
Ashby-Brown contrast, 456
Astigmatism
  condenser, 162
  intermediate, 163
  objective, 162, 163, 164, 169, 466
Atomic
  basis, 259, 260, 262
  correction factor, 640, 650
  number, 11, 16, 24, 26, 29, 30, 39, 41, 42, 57, 58, 59, 60, 122, 224, 237, 258, 284, 373, 378, 497, 635, 639, 640, 643, 650, 665, 672, 728, 729, 730
  scattering amplitude, 40, 45, 49, 258, 261, 294, 336, 517, 669
  scattering factor, 44–45, 223, 257, 378
  structure, 48, 55, 380, 389, 493, 679, 741, 747, 752
Atomic-column EELS, 567, 735–736
Auger electron spectrometer (AES), 53, 55, 61, 62
Augmented plane wave, 748
Automated crystallography, 305
Automated orientation determination, 305
Automatic beam alignment, 560
Automatic peak identification, 627–630
Averaging images, 554–556
Axis-angle pair, 303

B
Back focal plane, 94, 95, 111, 152, 162, 204, 205, 373, 491, 573, 683, 685–688, 691
  See also Lens
Background
  extrapolation, 725, 726, 727, 728
  modeling, 643
  subtraction, 342, 550, 555, 641–644, 646, 650, 659, 725, 726–728, 729, 734
  See also Bremsstrahlung
Backscattered electron (BSE) detection, 115, 230
Baking, Band gap (semiconductor), 709–710
  image, 710
Bandwidth, 118, 119
Bar, 128, 157, 188, 202, 517, 588, 594, 611
Barn, 27
Basal plane, 261, 449, 745
Basis vectors, 563–565
Beam-defining aperture, 122, 329, 380, 466, 599, 610, 615
Beam (electron)
  blanking, 101
  broadening, 84, 87, 666, 673
  coherence, 533
  convergence, 538–540
    See also Convergence semiangle
  current, 13, 65, 76, 78, 81, 82–83, 84, 87, 107, 110, 121, 149, 182, 592, 593, 596–597, 599, 601, 617, 640, 641, 653, 667, 668, 670, 672, 674, 694, 721
  damage, 10–11, 30, 53–68, 76, 86, 123, 164, 263, 556, 557, 632, 652, 673
  deflection, 398
  diameter, 82, 83–85, 107, 326, 329, 636, 664, 665, 666, 667
  diffracted, 17, 24, 40, 41, 47–49, 53, 155, 156, 157, 166, 198, 199, 201, 202, 204, 215, 216, 221–231, 235, 240, 257, 265, 271, 272, 273, 274, 280, 285, 296, 298, 305, 313, 337, 361, 371, 372, 381–382, 383, 385, 407, 408, 421, 444, 463, 469, 470, 471, 472, 488, 513, 517, 519, 525, 534, 535, 536, 537, 641, 701
    See also Diffracted beam
  diffracted amplitude, 272, 413
  direct, 24, 25, 34–36, 48, 53, 116, 117, 152, 155–157, 159, 160, 161, 162, 166, 168, 169, 198, 202, 203, 204, 205, 215, 225, 229, 294, 317, 337, 349, 350
  direction, 26, 44, 45, 204, 216, 217, 248, 283, 288, 299, 300, 301, 302, 304, 312, 317, 318, 348, 349, 350, 543, 600, 614
  energy, *See* Accelerating voltage
  incident, 325, 384
  many-beam conditions, *See* Many-beam
  parallel, 92, 94, 141, 142–143, 145, 146, 147, 148, 152, 158, 163, 283–305, 311, 324, 325, 326, 339, 340, 352, 385, 386, 511, 721
  shape, 667
  splitter, 397, 525
  tilting, 147, 199, 285, 382, 467, 669, 755
  translation, 147
  two-beam conditions, *See* Two-beam approximation
Beam-sensitive materials709
  See also Beam (electron), damage
Beam-specimen interaction volume, 323, 598, 663, 664, 665, 672, 735
Bend contour, 352, 385, 386, 407, 411–412, 413, 415, 441, 493, 521, 641, 658, 733
Beryllium
  grid, 609, 612
  oxide, 612
  specimen holder, 613
  window, 586, 587, 598, 607, 628, 633, 650, 651, 654
Bethe cross section, 58
Bethe ridge, 718
Biprism, 77, 397, 398, 525
Black level, 377, 528
Black/white contrast, 449
Bloch theorem, 237
Bloch wall, 517
Bloch wave
  absorption, 241
  amplitude, 252–253
  coefficient, 238
  kinematical condition, 221
Body-centered lattice, 357
Boersch effect, 683
Bohr radius, 42, 63, 378, 705
Bohr theory, 55
Bolometer, 590, 591, 592, 594, 663
Borrmann effect, 657, 658, 733
Boundary, *See* Grain, boundary; Interface; Phase boundary
Bragg
  angle, 34, 49, 83, 169, 200, 201, 202, 205, 222, 223, 229, 230, 312, 327, 339, 340, 353, 381, 408, 451, 680, 687, 721
  beam, 221, 230, 239, 245, 246, 247, 250, 254, 431, 492, 534, 536
    See also Diffracted beam
  condition, 204, 207, 213–216, 227, 228, 241, 253, 272, 274, 294
  diffraction, 26, 201, 202, 211, 214, 222, 231, 311, 312, 320, 339, 373, 381, 699
  law, 199, 200–202, 211, 213–214, 217, 218, 319, 411, 488, 519, 590
  plane, 249, 319, 416
  reflection, 49, 157, 202, 208, 262, 305, 448, 492, 537
Bravais lattice, 267, 347
Bremsstrahlung, 40, 58, 60, 135, 168, 582, 598, 605, 606, 608–614, 618, 626, 632, 635, 641–643, 673
  coherent, 613–614, 626, 642
  See also Background
Bright field (BF)
  detector, 122, 159, 160, 161, 162, 326, 366, 373, 380, 384, 385, 521
  high-order BF, 521
  image, 155–159, 161, 163, 165, 166, 168, 169, 182, 230, 295, 304, 330, 331, 352, 362, 365, 372, 374–376, 379–381, 386, 407–409, 411, 412, 414, 415, 424, 425, 427, 428, 434, 444, 452, 455, 458, 468, 472, 473, 475–477, 512, 516, 521, 573, 599, 617, 669, 670, 687, 734
  in STEM, 304, 373, 376, 385, 600, 617, 670
  symmetry, 361, 366
Brightness (gun), 79, 116, 150, 327, 375, 626
Brillouin-zone boundary (BZBs), 245, 253
Buckyballs, 743, 745
Bulk holder, 135, 175
  See also Specimen, holder
Bulk modulus, 457
Burgers vector, 278, 320, 339, 396, 402, 420, 441–444, 446–449, 456, 458, 469, 473–474, 476
  See also Dislocation

C
Calibration
  of accelerating voltage, 168–169
  of camera length, 165–166
  of focal increment, 169
  of illumination system, 149–150
  of image rotation, 100–101
  of magnification, 164–165
Camera constant, 166, 355
Camera length, 154, 155, 161, 162, 165–167, 197, 198, 217, 218, 284, 302, 317, 318, 326, 327, 328, 329, 336, 361, 373, 379, 398, 686, 687, 688, 721
c/*a* ratio, 259, 314
Carbon
  amorphous, 35, 163, 183, 185, 295, 374, 551, 554, 612, 719, 722
  contamination, 586
  film, 86, 162, 163, 164, 173, 183–185, 374, 375, 397, 521, 554, 612, 722, 745, 751
  nanotube, 73, 81, 365, 366, 695, 745
  See also Holey carbon film
Cartesian-vector notation, 260
Cathode-ray tube (CRT), 115
Cathodoluminescence, 53, 62–63, 116, 122, 523–524
Cauliflower structure, 583
CCD-based WDS, 590

Center of symmetry, 230, 236, 240, 358, 361, 435
Centrosymmetric point group, 358
Channel-to-channel gain variation, 689, 727
Channeling, 230, 339, 366, 379, 646, 657–658, 733
Characteristic length, 221, 222, 223–224, 225, 237, 478
  See also Extinction distance
Characteristic scattering angle, 63, 700, 705, 717, 719, 733
Charge-collection microscopy, 62–63, 523–524
Charge-coupled device (CCD) camera, 10, 116, 120–121
Charge-density determination, 366
Chemically sensitive images, 517
Chemically sensitive reflections, 261, 262, 263, 519, 567
Chemical resolution, 626
Chemical shift, 746, 750–751
Chemical wire/string saw, 176
Chi-squared, 645
Chromatic aberration, 6, 8, 9, 91, 103, 104–106, 108, 109, 148–149, 159, 334, 377, 491, 495–497, 533, 680, 681, 687–688, 690, 702, 703, 706, 711
  See also Aberration
Chromium-film standard (C film), 160, 189, 717
Clamping ring, 132, 133, 135, 181
Cleavage, 176, 187, 191, 414
Cliff-Lorimer
  equation, 640, 641, 646, 647, 648, 652, 656, 657, 673
  k factor, 646
Coherence, *See* Beam (electron), coherence; Bremsstrahlung, coherent; Spatial coherence
  Foucault imaging, 516–517
  Fresnel imaging, 516
  interference, 31
  particles, 456, 457
  processing, 526
  scattering, 39, 202, 319
Coincident-site lattice (CSL), 500
Cold FEG, 498
Cold trap, *See* Anticontaminator
Collection semiangle, 34
Collimator, 598–599, 611
Column approximation, 223, 229–230, 421, 423, 426, 433, 434–436, 443, 457, 478
Coma-free alignment, 493, 498
Comis, 433–436, 458
Composition
  measurement, 7, 667–668
  profile, 616, 659, 667
Compton scattering, 741, 755
Computer simulation, 311, 351, 362, 365, 392, 414, 456, 492, 493, 539, 665
  See also Image, simulation of
Condenser 1 lens, 83, 84, 85, 142, 144, 145, 148, 149, 150, 158, 167, 326, 327, 334, 584, 721
Condenser 2 lens, 83, 111, 116, 142–150, 158, 167, 207, 326, 327, 328, 330, 332, 584, 610, 615, 632, 721
  aberration, 81, 82, 91, 92, 108, 111, 150, 373, 525, 725
  alignment, 161–162
  aperture, 101–102
  calibration, 154
  defocusing, 95–96
  diaphragm, 101–102
  See also Lens
Condenser objective condition, 326
Condenser-objective lens, 325, 326
Conduction band, 59, 62, 64, 66, 117, 245, 585, 605, 709, 720, 742, 744–745
Confidence limit, 632, 648
Confocal microscopy, 4, 8
Conical diffraction, 157, 291
Conical scanning, 157
Conjugate plane, 94, 152, 160, 161
Constructive interference, *See* Interference
Contamination, 62, 81, 99, 102, 106, 118, 124, 127, 130, 132, 135, 137, 138, 149, 161, 162, 181, 189, 190, 277, 324, 341, 494, 519, 523, 584–586, 588, 595, 609, 612, 615, 620, 626, 632, 636, 641, 647, 658, 670–671, 673, 751
Continuum, 55, 238, 293, 443, 605, 606, 613, 642, 726, 742, 744, 754
  See also Background; Bremsstrahlung
Contrast
  difference, 371, 372–373, 374, 376
  Fresnel, 374, 389, 397, 399–402, 540–541
  inside-outside, 448, 450, 476
  minimum, 378, 492, 493, 497, 498, 561
  minimum defocus, 498
  topographic, 519, 521
  transfer function, 17, 485, 487, 494, 506
  See also Amplitude contrast; Diffraction, contrast; Phase contrast
Convergence semiangle, 84
  correction (Cs correction), 6–8, 10, 62, 68, 76, 82, 84, 104, 108, 124, 355, 597, 647, 659, 663, 668, 672, 674, 733, 735, 736
Convergent beam
  diffraction, 671
    See also Higher-order Laue zone (HOLZ)
  energy-filtered, 77
  imaging (CBIM), 332, 334
Cooling holder, 132, 134, 135–136, 324, 336
  See also Specimen, holder
Core-hole effect, 750
Core-loss image, 715, 717, 719, 730, 731, 732, 741
Coster-Kronig transition, 59, 745
Coulomb force, 24, 36, 39, 40, 199, 221
Count rate, 588–594, 596–599, 601, 607, 608, 616, 618, 625, 627, 630, 632, 636, 641, 658, 659, 667, 673, 717
Coupled harmonic oscillator, 231
Coupled pendulum, 477
Critical energy, *See* Ionization
Cross-correlating image, 561
Cross-correlation function, 568
Cross section
  differential, 23, 27, 39, 41, 42, 44, 45, 63, 729
  elastic, 27, 28, 44, 54, 58
  experimental, 27–28
  generalized, 433–434
  modified Bethe-Heitler, 58
  partial ionization, 723, 725, 728–730, 734
  phonon, 63–64
  plasmon differential, 63–64
  relativistic Hartree-Fock, 545
  Rutherford, 41–43
  screened-relativistic Rutherford, 42–43
Cross-section specimen preparation, 182
Cross-tie wall, 517
Cryogenic pump, 130
Cryo-transfer holder, 136
  See also Specimen, holder
Crystal
  A-face centered, 357
  B-face centered, 357
  cubic, 198, 241, 258, 260, 261, 273, 277, 289, 340, 356, 357, 394
  diamond cubic, 289, 314, 420
  high-symmetry pole, 341
  I-centered, 357
  imperfect, 386, 443, 534
  low-index, 289
  low-symmetry zone axis, 338, 356
  non-centrosymmetric, 230, 236, 358, 435
  orientation of, 304, 317–318, 435, 493, 755
  orthorhombic, 213, 354, 357
  perfect, 197, 224, 235, 236, 259, 279, 386, 392, 396, 397, 421–423, 426, 435, 441, 443, 445, 447, 459, 463, 471, 473, 478, 492, 505, 533, 534, 536, 537, 540, 542
  plane, 9, 48, 49, 138, 204, 208, 286, 311, 336, 339, 657
  pole, 287, 288
  potential of, 230, 236, 238, 239, 240
  primitive, 257–258, 267, 356
  projected potential of, 534
  simple cubic, 241, 258, 261, 273
  tetragonal distortion of, 415
  zone axis, 17, 204, 213
  See also Lattice
Crystal Kit, 267
Crystallographic convention, 204, 230
Crystallographic shear, 503

Curie temperature, 429, 517
Current, 74–79, 82–86
  See also Beam (electron)
Current centering
  dark, 118, 689–690, 709, 711
  density, 162
Curve fitting, 645, 726
Cut-off angle, 700, 706
Cyanide solution, 174
Cyclotron radius, 99, 100

D
Dark field (DF)
  annular, 85, 122, 161
  centered, 155–156
  detector, 160
  diffuse, 295
  displaced-aperture, 155, 156, 412
  external control for, 515
  focus of, 492
  high-angle annular, 122, 144, 379–381, 386, 710, 735, 736
  image, 206, 332
  multiple, 206, 207
  STEM, 376–377
  through focus (2½D), 513
  tilt control, 382
  See also Weak-beam dark field
Dead layer, 118, 585, 586, 595, 596, 607, 651
Dead time, 592–593, 596–598, 601, 607–608, 611, 627, 630
Debye-Waller factor, 336, 348, 435, 533
Decision limit, 674
Deconvolution, 630–632, 667, 688–689, 694, 700, 701, 705, 706, 707, 715, 721, 725, 728, 730, 731–733, 752, 753
Defect
  computer modeling of, 432–433
  core, 421
  unit cell, 361–362
  See also Dislocation; Grain, boundary; Stacking fault; Twin boundary
Defocus condition, 492, 513
Defocused CBED patterns, 329–330
Defocus image, 557
Deformable-ion approximation, 435
Delocalization, 390, 497, 498, 500, 735
Delta (δ) fringe, 427–429
Delta function, 731
Density, 16, 27, 28, 29, 63, 65, 74, 75, 76, 79, 81, 82, 115, 120, 124, 164, 174, 176, 180, 190, 293, 362, 364, 365, 371, 373, 378, 401, 442, 444, 448, 450, 487, 495, 503, 514, 541, 563, 590, 651, 653, 654, 655, 656, 665, 676, 679, 706, 707, 709, 710, 721, 741, 742, 743, 748
Density-functional theory, 711, 748
Density of states, 590, 721, 742, 743, 744, 745, 747, 748, 749, 750, 751
Depth
  distribution of X-ray production, 655, 665
  of field, 8, 91, 92, 101, 103, 110–111, 513
  of focus, 8, 91, 92, 101, 103, 110–111
  fringes, 426, 471
Desktop Microscopist, 685
Desk-Top Spectrum Analyzer (DTSA), 16, 608, 628, 641, 652, 674, 730
Detectability limits, 54, 475, 581, 589, 590, 626, 631, 632, 663, 672, 673, 674, 675, 715, 736
Detection quantum efficiency (DQE), 116, 118, 119, 121, 123
Detector (electron), 10, 24, 27, 117–122, 159, 372, 373, 389, 511, 523, 586
  depletion region of, 117, 118
  envelope function, 485, 491, 492, 494, 495, 496, 497
  gain of, 118, 119
  STEM, 326, 366, 372, 385, 386, 511, 528, 687
  See also Spectrometer (EELS)
Detector (X-ray), *See* Spectrometer (X-ray energy-dispersive); Spectrometer (X-ray wavelength-dispersive)
Determination limit, 674
Deviation parameter, 216, 273, 297, 298, 353, 371, 382–384, 415, 435, 496
  See also Excitation error
Diamond-cubic structure, 420
Diamond window, 187
Diaphragm, 101–102, 122, 132, 144, 145, 147, 148, 149, 153, 154, 155, 156, 324, 326, 332, 333, 373, 375, 385, 485, 502, 515, 521, 528, 594, 601, 608, 609–611, 613, 621, 627
  self-cleaning, 102
  top-hat C2, 611, 621
  See also Aperture
Dielectric constant, 42, 694, 701, 706, 710, 711
  determination of, 705
  image, 705
Dielectric response, 679, 699, 705
Difference spectrum, 643, 690, 727, 733
Differential hysteresis imaging, 165
Differential pumping aperture, 131, 683, 687
Differentiating the image, 556
Diffracted beam, 47–49, 156, 166, 204, 221–231, 265, 337, 383, 408, 519, 534, 701
  amplitude of, 31–33, 371, 444
  intensity of, 47, 215, 273, 274
Diffracting plane, 199, 201, 202, 204, 208, 213, 246, 287, 289, 312, 318, 319, 320, 332, 340, 371, 390, 396, 407, 411, 412, 416, 441, 442, 444, 445, 449, 452, 454, 463, 469
  See also Bragg, plane
Diffraction
  camera, 165, 198–199
  center, 162
  contrast, 197–207, 313, 371–386
  convergent beam, *See* Convergent beam, diffraction
  from dislocations, 278–279
  double, 222, 296–298, 304, 394
  extra reflection, 264
  Fraunhofer, 30–31
  Fresnel, 30–31
  grating, 31–32, 164, 165, 273, 573, 590
  group, 365–366
  indexing, 213
  180° inversion of, 31
  mode, 116, 152, 153, 154, 161, 166, 167, 206, 326, 330, 334, 466, 685, 687, 688, 721, 735, 753
  multiple, 296, 364
  nanodiffraction, 283, 291, 323, 347, 365, 366
  oblique-textured, 291
  pattern, 17, 49, 198, 204, 207, 372, 382, 383, 384, 391, 394, 755
  ring, 155, 287, 293
  rocking-beam, 230
  rotation, 167–168
  scanning-beam, 365
  selected area, *See* Selected area diffraction (SAD)
  shell scattering, 749
  single-crystal, 168
  split spot in, 516
  spot spacing in, 336
  streak in, 254
  systematic absence in, 304
  systematic row in, 332
  vector (g), 201
Diffraction coupling, 685
Diffractogram, 17, 493, 551, 552–554, 555, 560, 561, 574
Diffuse scattering, 329, 711
  See also Scanning transmission electron microscope (STEM)
  Diffusion pump, 129, 130, 131, 180, 191
Diffusion coefficient, 104
Digital
  filtering, 643, 644
  image, 16, 117, 124, 155, 528, 556, 619
  mapping, 618–620
  pulse processing, 598, 635
  recording, 131, 434
Digital Micrograph, 552, 560, 570, 573
Dimpling, 177, 178, 191
Diode array, 683
  saturation of, 691, 723
Dipole selection rule, 744
Direct beam, *See* Beam (electron), direct
Discommensurate structure, 278
Discommensuration wall, 503
Disk of least confusion, 103
Dislocation
  array, 278–279, 394, 450–451, 452, 455
  contrast from, 444–448
  core of, 402, 443, 448, 469

density, 442
dipole, 448–450, 476
dissociated, 451, 463, 473–477
edge, 442, 444, 451, 452, 459, 469, 471, 475
end-on, 400, 401, 402
faulted dipole, 441, 476
faulted loop, 448
inclined, 458
interfacial, 453
intersecting, 458
line direction, 449
loop, 448–450
misfit, 448, 453, 455, 456
network, 448, 453
node, 448
ordered array of, 278
pair, 450
partial, 445–447, 450, 451, 463, 473–476
screw, 441, 442, 444, 447, 452, 454, 457, 458, 459, 470, 475
strain field of, 279, 371, 447, 451, 457
superlattice, 447
transformation, 453, 456
See also Burgers vector; Strain
Disordered/ordered region, 263, 517–518
Dispersion diagram
branches of, 247, 249, 250, 251, 253, 254
plane (of spectrometer), 246, 247, 249
relation, 250–251
surface, 247–250
Displacement
damage, 67, 68, 86, 721
energy, 67
field, 433, 435, 441, 442, 443, 444, 447–448, 456, 458
vector, 348, 349, 350, 458
Display resolution, 592, 614, 627, 630, 722
Double-period image, 564
Double-tilt holder, 134, 324
See also Specimen, holder
Drift
correction, 617, 636
rate, 464, 498
tube, 681, 682, 684, 722
Dwell time, 124, 636, 710, 722
Dynamical
absence, 258
calculation of intensity, 364–365
condition, 362–363
contrast in CBED, 347
coupling, 239
diffraction, 203, 221, 222, 225, 274, 296, 297, 298, 329, 331, 358, 364, 641, 669
scattering, 30, 203, 221, 258, 265, 272, 298, 311, 319, 323, 342, 358, 558
See also Diffraction
Dynamic experiments, 526–528

E

Edge, *See* Ionization
EELS advisor software, 730, 734
EELS, *See* Electron energy-loss spectrometry (EELS)
EELS tomography 755–757
Effective EELS aperture diameter, 688
EFTEM imaging, 685, 703, 733–735
Elastic
coherent, 25
constant, 454, 457, 458
cross section, 41–43
mean-free path, 23
scattering, 25–26, 27, 28, 30, 39–49, 64, 199, 324, 329, 332, 339, 340, 373, 374, 378, 380, 381, 664, 665, 699, 702, 741
See also Scanning transmission electron microscope (STEM)
Elasticity theory, 421, 448, 458, 474, 475
ELD software, 559
Electric-field potential, 236
Electro-discharge machining, 176
Electron
backscatter pattern (EBSP), beam, *See* Beam (electron)
beam-induced current (EBIC), 62, 136, 523
channeling, 366, 379, 733
charge, 523–524
crystallography, 47, 324, 558, 559
detector, *See* Detector (electron)
diffraction, 8–9, 34–36
dose, 65
See also Beam (electron), damage
source, 73–87
Electron-electron interaction, 40, 41, 42, 66, 683, 699
Electron-hole pair, 62–63, 117, 118, 524, 585, 586, 587, 591, 592, 593, 595, 611
Electronic-structure tools, 711
Electron-spectroscopic imaging, 690
Electropolishing, 178, 179, 183
Electrostatic lens, 78, 80, 96
See also Lens
ELP (energy-loss program), 16
Empty state, 751
EMS (electron microscope simulation program), 17, 127, 267, 287, 314, 491, 534, 571
Enantiomorphism, 347, 363–364
Energy
EELS, 715–736
spectrometry (EELS) resolution, 76
spread, 73, 74, 76–77, 79, 81, 82, 85–86, 105, 148–149, 496, 498, 693, 701
window, 635, 703, 734
X-ray, 590, 591–593
See also Electron energy-loss spectrometry (EELS)
Energy-dispersive spectrometry, 7
See also Spectrometer (X-ray energy-dispersive)
Energy filtering, 7, 8, 106, 334, 336, 342, 352, 366, 478, 696, 702, 703, 753, 755
Energy-loss, *See* Electron energy-loss spectrometry (EELS)
Electron energy-loss spectrometry (EELS)
angle-resolved, 755–756
collection efficiency of, 735–736
collection mode, 735
detectability limit of, 736
diffraction mode, 735
image mode, 755
imaging, 735–736
microanalysis by, 589
parallel collection, 681
serial collection, 685
spatial resolution of, 735–736
See also Spectrometer (EELS)
Energy-loss spectrum
artifacts in, 688–689
atlas of, 720, 723
channeling effect in, 733
deconvolution of, 731–733
extended fine structure in, 743
extrapolation window in, 726, 727
families of edges in, 723
fine structure in, 736
gun, 693
See also Gun, holography
interferometer, 398
lens, *See* Lens
microscope microanalyzer, 646
momentum, 741
near-edge structure in, 723
parameterization of, 730
phase, *See* Phase boundary
potential energy, 749
power-law fit, 726
rest mass, 700
scattering, *See* Diffraction, shell scattering
source, 735–736
See also Gun
structure factor, 567
See also Structure factor
velocity of, 700
wavelength of, 705
wave vector, 752
Envelope function, 485, 491, 492, 494, 495, 496, 497
Epitaxy, 138, 168, 296
Errors in peak identification, 634
Errors in quantification, 647–648
Escape peak, 606, 612, 614, 628, 630, 656
Eucentric
height, 151, 164, 165, 166, 169, 330, 515
plane, 100–101, 151, 167, 326, 327, 330, 333, 334, 599
specimen, 295
See also Goniometer
Ewald Sphere, 214–218, 235, 241, 248, 249, 252, 271, 273–275, 279–281, 290–292, 298, 312, 318, 324, 336–338, 351, 355, 411, 430, 431, 447, 449, 464, 466, 467, 470, 478, 537, 559

Excitation error, 216–217, 224, 227, 252, 318, 353, 463, 702
effective, 228–229, 407, 464
See also Deviation parameter
Exciton, 66, 750
Exposure time, 123, 158, 207, 305, 337, 464, 466
Extended energy-loss fine structure (EXELFS), 294, 717, 718, 741, 742, 751–754, 757
Extended X-ray absorption fine structure (EXAFS), 294, 717, 751–753
Extinction distance
apparent, 435
determination of, 435
effective, 245, 252, 408, 467, 471, 669
Extraction replica, 185, 191, 377, 378, 616, 724
Extraction voltage, 80–81
Extrinsic stacking fault, 473

F
Face-centered cubic, 259
See also Crystal; Lattice
Fano factor, 593
Faraday cup, 82, 85, 101, 121–122, 596, 614, 653, 670
Fast Fourier transform (FFT), 534, 536
Fe^{55} source, 593
Fermi
energy, 720, 742, 743
level, 55, 720, 742, 743, 749
surface, 63, 277, 278, 742
FIB (Focused-ion beam), 11, 157, 186, 188–189, 373, 634
Field-effect transistor (FET), 585, 586, 589, 591
Field emission, 61, 73, 74–75, 80–81, 498
See also Gun
Filtered image, 120, 434, 550, 572, 681, 683, 692, 694–696, 699, 702, 703, 709, 717, 722, 725, 734, 741, 751, 756
Filter mask, 573
Fine structure, *See* Electron energy-loss spectrometry (EELS)
Fingerprinting, 704, 708, 744, 746–747, 751
Fiori definition, 614, 673
See also Peak-to-background ratio
First-difference spectrum, 727, 733
First-order Laue zone, 351
See also Higher-order Laue zone (HOLZ)
Fitting parameter, 645
Fixed-pattern readout noise, 689
Flat-field correction, 570
Fluctuation microscopy, 294, 366, 528
Fluorescence (light), 116
Fluorescence (X-ray), 607, 609, 612, 613, 630, 632, 639, 640, 647, 653, 654
correction, 656–657
yield, 55, 59, 587, 605–606, 628, 650, 715, 721, 729
Flux lines, 517, 526, 527
See also Magnetic correction, flux lines
Focus, 110–111, 148, 151, 329–332, 399, 490–491, 682–683, 895–896
See also Lens; Overfocus; Underfocus
Focused-ion beam, *See* FIB (Focused-ion beam)
Focusing circle (WDS), 681
Forbidden electron energies, 247
Forbidden reflection, 258, 263, 265, 288, 296, 299, 300, 301, 349, 350, 366
See also Diffraction, pattern; Systematic absence
Foucault image, 516
Fourier
analysis, 572
coefficient, 17, 237, 240, 246
component, 435
deconvolution (logarithmic, ratio), 705
fast (Fourier) transform (FFT), 534, 536
filtering, 551, 571
inverse transform, 536, 731, 732
reconstruction, 551–552
series, 237, 238
transform, 485, 487, 536, 551, 568, 572, 573, 731, 732, 752–754
Frame averaging, 464, 466, 478, 550, 555, 557
Frame grabber, 551
Frame time, 120, 121
Free electron, 59, 61, 705, 708, 751
Free-electron density, 63, 706, 707, 710
Fresnel
biprism, 77, 397–398
contrast, 374, 389, 397–402, 540–541
diffraction, 30–31, 229, 535
fringe, 86, 87, 163, 389, 397, 400, 401, 402–403, 540–541, 575
image, 403
zone construction, 229
Friedel's law, 358
Full-potential linearized augmented plane wave (FLAPW), 748
Full Width at Half Maximum (FWHM), 83, 84, 85, 149, 150, 593–595, 605, 628, 631, 632, 643, 644, 667, 683, 684, 693, 694
Full Width at Tenth Maximum (FWTM), 84, 85, 86, 326, 594, 595, 643, 659, 664, 667, 668, 674
FWTM/FWHM ratio, 595, 601

G
Gas bubble, 399–400
Gas-flow proportional counter, 590
Gatan image filter (GIF), 681
Gaussian
curve fitting, 645
diameter, 84
image plane, 103–104, 106–110, 488, 664
intensity, 83, 84, 85
statistics, 647, 673
Generalized-oscillator strength (GOS), 729
Generated X ray emission, 600, 612, 628, 634
Germanium detector, 587–588
See also Spectrometer (X-ray energy-dispersive)
Ghost peak, 690, 691, 701, 723
Glaser, 491, 494
Glass layer, 278
Glide plane (dislocation), 443, 444
Glide plane (symmetry), 444
Goniometer, 132, 133, 151, 285
See also Eucentric
GP zone, 277
Grain
boundary, 6, 17, 216, 276, 278, 286, 302, 366, 380, 392, 395, 397, 400, 402, 409, 419, 420, 500, 502, 503, 517, 616, 659, 694, 747
See also Stacking fault; Twin boundary
coincident-site lattice, 500
high-angle, 402
low-angle, 400, 402
rotation, 430
small-angle, 454
size, 116, 123, 197, 283, 284, 290, 291, 293
texture, 292
tilt, 402
twist, 455
Gray level, 371, 555, 563
Gray scale, 123, 570, 635, 636
g.**R** contrast, 443, 450
Great circle, 286–287, 302
See also Stereographic projection
Grid, 75, 77, 133, 134, 173, 174, 175, 182–185, 187, 188, 274, 458, 588, 594, 598, 599, 600, 609, 610, 611, 612, 621, 633
Gun
alignment of, 498
brightness of, 498
crossover, 80
emission current, 79, 82
field-emission, 80–81
filament, 81, 493
flashing of, 81
holography, 81
lanthanum hexaboride, 81–82
saturation of, 81
self-biasing, 78
tungsten, 81
undersaturated image of, 493
Wehnelt cylinder, 77, 78
See also Electron, source; Field emission
G vector, 204, 224–225, 238, 253, 314, 315, 351, 384, 393, 425, 433, 442, 453, 458, 669
See also Diffraction, vector (g)

H
Handedness, 363, 364
Hartree-Slater model, 729

Heisenberg's uncertainty principle, 745
Hexagonal close-packed crystal, 259, 267
  See also Crystal
Higher-order Laue zone (HOLZ)
  indexing, 348–352
  line, 351–352
  plane, 339, 340
  ring, 336, 337, 338, 341, 348, 351, 354, 355–356, 361, 362, 363, 364
  scattering, 334, 336
  shift vector t, 357
  simulation of, 351, 362
  See also Convergent beam, diffraction; Kikuchi diffraction, line
Higher-order reflection, 202, 207, 289
Higher-order waves, 48
Higher-order X-ray lines, 57
High-resolution TEM, 483–506
High voltage, 6, 13, 75, 76, 77, 96, 104–105, 108, 119, 179, 450, 451, 490, 495, 522, 588, 608, 681, 691, 722
High-voltage electron microscope, 6
History of the TEM, 5
Holder, *See* Specimen, holder
Hole-count, 608, 609
Holey carbon film, 86, 162, 163, 183, 184
Hollow-cone diffraction, 157, 291–293, 295
Hollow-cone illumination, 157, 291, 293, 381
Hollow-cone image, 158, 295, 528
Holography, 81, 397, 496, 524–526
Howie-Whelan equations, 224–226, 407, 413, 414, 421, 433, 434, 436, 442–443, 457
Hydrocarbon contamination, 138, 612
  See also Contamination
Hydrogenic edge, 717, 718, 719, 724, 731

I

Ice, 595
Illumination system, 75, 77, 78, 83, 93, 101, 141, 142–150, 161, 163, 311, 334, 584, 605, 608
  See also Condenser 1 lens; Condenser 2 lens
Image
  analysis of, 453, 491
  calculation of, 17, 537
  contrast in, 24, 77, 109, 222, 384, 469, 475
  coupling, 685
  of defects, 235, 409, 443
  delocalization in, 500, 735
  distance, 94, 95
    See also Lens
  drift, 76
  of flux lines, 526
  formation, 92, 93
  lattice-fringe, 392, 502, 559
  matching, 436, 540
  plane, 94, 95, 96
    See also Lens
  processing of, 16, 493, 549–556
  rotation of, 99, 167, 384
  simulation of, 15, 389, 429, 433, 458, 533–545, 563
  of sublattice, 500
Imaging system of TEM, 164–168
Incidence semiangle, 34
  See also Angle; Scanning transmission electron microscope (STEM), incoherent
Incommensurate structure, 503
Incomplete charge collection, 593, 595, 645
Incomplete read-out, 690
Inelastic, *See* Scanning transmission electron microscope (STEM), inelastic
Information limit, 492, 495
Information theory, 495
Infrared sensors, 362
In-hole spectrum, 610
In-line holography, 524
Inner potential, 235, 236, 238, 541, 545
  mean, 237
  scaled mean, 238
Inner-shell ionization, *See* Ionization; Instrument response function
In-situ holders, 135
In situ TEM, 138, 526, 531
Instrument response function, 731
Instrument spectrum, 613
Integration, 42, 273, 458, 470, 644
  total, 691
  window, 728
Integration approach, 273
Intensity, 371
  See also Spectrum, electron energy-loss, X-ray
Interaction constant, 486
Interband scattering, 254, 466, 478
Interband transition, 63, 66, 705, 709, 710, 711
Interface, 44, 275, 295, 430
  contrast, 667–668
  dislocation, 456
  interphase, 286, 302, 616, 617, 667
  semicoherent, 456
  strain at, 396
  See also Grain, boundary; Phase boundary
Interference, 13, 31, 40, 47, 48, 77, 185, 200, 202, 203, 241, 394, 398, 526
  constructive, 33, 34, 45, 213, 214
  destructive, 48, 200, 203
  fringe, 77, 396, 397, 398
Intergranular film, 402, 541
Intermediate-voltage electron microscope, 76, 323, 589, 600, 608, 616, 618, 626, 635, 636, 668, 673, 675
Internal-fluorescence peak, 612, 613, 630
International Tables, 265–266
Internet, 15–17, 117
Intersecting chord construction, 465
Inter-shell scattering, 749
Interstitial atom, 277
Intraband transition, 700, 709
Intra-shell scattering, 749
Intrinsic Ge detector, 585, 587–588
  See also Spectrometer (X-ray energy-dispersive)
Inversion domain boundary, 420, 503
Invisibility criterion, 424, 445
Ion beam blocker, 181
Ionic crystal, 66
Ionization, 57–59, 667, 715–721
  critical energy for, 55, 57, 58, 628, 682, 717, 723, 742, 743, 750
  cross section for, 55, 57, 588, 599, 605, 615, 650, 652, 717, 723, 725, 728, 734
  edge, 717, 723, 733–735
  integration of, 717
  intensity of, 715, 733
  jump ratio of, 722, 732
  onset of, 730
  shape of, 717
Ion milling, 174, 178–181, 182, 186
Ion pump, 130, 131, 137

J

Jump-ratio image, 728, 734

K

Kernel, 556
Kikuchi diffraction, 311–320
  3g, 464, 465
  band, 313, 314, 318, 319, 325, 327, 339, 340, 341, 348, 361, 512
  deficient, 313, 319, 340
  excess, 314, 318
  line, 311–313, 318, 327, 339–340, 366, 382, 413, 464, 466
  map, 303, 313–318, 319
  pair, 313, 314, 317, 318, 340
  pattern, 303, 311, 312, 314, 315, 317, 319, 325, 464–465
Kinematical diffraction, 235, 280, 358, 362, 521
  approximation, 221, 463, 470, 472
  crystallography, 559
  equation, 464
  integral, 469, 470
  intensity, 294
Kinematically forbidden reflection, 258, 265, 296
  See also Forbidden reflection; Systematic absence
Kinetic energy, 14, 68, 230, 235, 236, 238, 253, 477
$K(k_{AB})$ factor, 640, 646
  calculation of, 648–652
  error in, 647–648
  experimental values of, 646–647
Knock-on damage, 65, 66, 67, 68, 626, 646
  See also Beam (electron), damage
Kossel, 8, 323, 327, 336, 339, 348, 352, 354
  cone, 312, 313

Kossel (cont.)
pattern, 327, 328, 331, 332, 341, 355, 356, 361
Kossel-Möllenstedt (K-M), 327, 328, 330, 336, 337, 342, 348, 352, 353, 354, 356, 361, 671
conditions, 327, 330, 336, 342, 348, 354
fringe, 352, 353, 361, 671
pattern, 327, 328, 356, 361
Kramers' cross section, 60
Kramers-Kronig analysis, 705
Kramers' Law, 606, 643, 645, 726
k space, 534, 536, 752, 753, 754
Kurdjumov–Sachs, 303
k vector, 199, 200, 239, 246, 248, 249, 251, 313, 325, 432, 537

L
L_{12} structure, 262, 427
Laplacian filtering, 556
Large-angle convergent-beam electron-diffraction patterns, 330–332, 412
Lattice, 211–212, 257–258, 356–357, 361–363, 389–392, 400–402, 567–568
centering, 338, 354, 356–357
defect, 400–402, 407, 408, 422, 442, 454
See also Dislocation
fringe, 389–392, 502, 525, 559, 573
imaging, 393, 500, 517, 567–568
misfit, 415, 454
parameter, 363, 394, 396, 415, 420, 424, 427, 454, 456, 540
point, 215, 216, 237, 257, 259, 262, 278, 336, 357, 470
strain, 347, 361–363, 456
vector, 200, 211, 212, 239, 297, 302, 420, 421, 424, 427, 428, 429, 485, 489
See also Crystal
Leak detection, 131–132
Least-squares refinement, 568
Lens, 91–112, 145–146, 148–149, 150–152, 161–164, 681–684
aberration of, 485, 488
See also Chromatic aberration; Spherical aberration
astigmatism, 106, 162, 163
See also Astigmatism
asymmetric, 682, 683
auto-focusing, 151
auxiliary, 116
bore of, 97, 145
condenser, *See* Condenser 1 lens; Condenser 2 lens
condenser-objective, 142, 143, 145–146
current, 97, 151, 164, 330, 496, 513, 560
defects, 99, 103, 107, 148, 331, 494, 553
demagnification, 95, 144, 145, 154
focal length of, 104
focal plane of, 91, 92, 94
focus of, 163
See also Overfocus; Underfocus
gap, 145
hysteresis, 165
immersion, 98
intermediate, 152, 153, 154, 155, 162, 163, 164, 166, 167, 205, 206, 207, 330, 517, 683, 685, 687, 691
low-field, 517
mini-, 145
Newton's equation, 95
objective, *See* Objective lens
octupole, 99, 104, 105, 106, 692
optic axis of, 92, 93, 99, 104, 147
pincushion distortion, 106, 165
polepiece of, 143, 144, 599
post-spectrometer, 683
projector, 111, 131, 141, 162, 166, 167, 682, 683, 685, 686, 687, 691
projector crossover, 682, 684, 686, 687
ray diagram, 103, 325, 333
rotation center of, 161–162, 163
sextupole, 99, 683, 692
snorkel, 98
superconducting, 98, 99
symmetric plane of, 151, 152
thin, 92, 94, 95
wobbling of, 148, 162
Library standard, 645, 704
Light element, 59, 61, 587, 590, 594, 595, 626, 679
Line analysis, 651, 695
Linear combination of atomic orbitals, 749
Linear elasticity, 442, 448, 495
Line of no contrast, 456, 457
Liquid N_2, 66, 129, 130, 132, 133, 180, 324, 363, 584, 585, 586, 595, 626
dewar, 133, 584, 594
holder, 132, 324, 336
Lithography, 187
Local-density approximation, 748
Long-period superlattice, 264–265
Long-range ordering, 279
Lorentz force, 99, 100, 398, 516
microscopy, 81, 398, 515–517
Low-dose microscopy, 377, 556
Low-loss, 680, 690, 693, 699–711, 722, 727, 732, 733, 744
intensity, 725
spectrum, 694, 701, 703–711, 730, 731, 732
See also Electron energy-loss spectrometry (EELS); Plasmon, fingerprinting

M
Magnetic correction, 514–515
domain wall, 514, 516
flux lines, 517, 526
induction, 516, 517
prism spectrometer, 679, 681, 682, 716
recording media, 514
specimen, 514–517
Magnification, 5, 7, 8, 11, 53, 76, 82, 86, 91, 95–96, 104, 106, 109, 110, 142, 143, 145, 147, 151, 153, 155, 161, 162, 164–165, 167, 168, 169, 206, 230, 264, 284, 327, 328, 330, 384, 389, 458, 466, 475, 493, 498, 512, 513, 518, 560, 562, 589, 599, 638, 670, 683, 686, 687, 688, 692
Many-beam, 240, 245, 390, 391, 408, 409, 433, 435, 436, 473, 478, 534, 565
calculation, 436
conditions, 478, 565
images, 390, 391, 470, 473, 534
Mask, 174, 187, 324, 334, 336, 338, 342, 397, 476, 521, 523, 551, 556, 572, 573, 586, 631, 705, 723, 753
Mass-absorption coefficient, 654
See also Absorption, of X-rays
Mass-thickness contrast, 185, 371, 373–379, 381, 382, 384, 407, 511, 696, 702
See also Contrast
Materials examples in text
Ag, 59, 66, 258, 374, 420, 424, 527, 588, 608, 609, 611, 649, 651
Ag_2Al, 314, 316
Ag_2Se, 291
Al, 35, 45, 54, 64, 67, 68, 82, 117, 118, 122, 166, 174, 205, 224, 258, 262, 263, 303, 352, 353, 372, 374, 385, 397, 420, 427, 429, 543, 583, 611, 618, 633, 651, 656, 704, 706, 708, 727, 733, 750, 752
Al-Ag, 649
$AI_xGa_{1-x}As$, 263, 264, 420, 517, 519, 568
Al_2O_3, 265, 266, 279, 296, 297, 298, 349, 379, 395, 410, 412, 413, 429, 519, 751
Al_3Li, 262, 303, 383
AlAs, 567, 568
Al-Cu, 506
Al-Li-Cu, 9
Al-Mn-Pd, 504, 505
Al-Zn, 618
Au, 29, 35, 43, 45, 64, 68, 102, 117, 160, 166, 173, 224, 263, 374, 397, 400, 424, 452, 453, 492, 523, 553, 585, 610, 635, 651, 665
Au_4Mn, 500, 501, 502
B, 187, 587, 613, 689, 721
$BaTiO_3$, 429
Be, 59, 67, 173, 237, 258, 587, 588, 612, 633, 702, 732, 733, 752
biotite, 644
Bi-Sr-Ca-Cu-O, 504
BN, 585, 590, 689, 705, 717, 719, 723, 732, 755
Ca, 59, 258, 504, 505, 633, 634, 638, 646, 671, 728
carbon, 35, 43, 59, 86, 163, 173, 183, 185, 207, 264, 295, 324, 365, 374, 375, 378, 397, 505, 521, 523, 528, 541, 551, 587, 612, 615, 633, 636, 670, 688, 695, 719, 722, 745, 751

See also Amorphous; Holey carbon film
carbon nanotube, 73, 81, 365, 366, 695, 745
catalyst particles, 4, 10, 617
CdTe, 181, 182
Co, 68, 258, 429, 457, 526, 630, 649, 651, 653, 732, 733
CoGa, 262, 395
Cr film, 518, 593, 612, 614, 615
CsCl, 262, 420
Cu, 29, 42, 43, 44, 59, 68, 173, 174, 187, 201, 237, 258, 263, 304, 305, 364, 385, 424, 429, 457, 466, 470, 471, 473, 475, 596, 611, 612, 618, 629, 633, 648, 665, 674, 675, 732, 743, 745, 746, 747, 751
Cu-Al, 649
Cu_3Au, 262, 263
CuAu, 420
$CuCl_{16}PC$, 557, 558
Cu-Co, 457
CuZn, 262
diamond, 173, 176, 178, 184, 224, 262, 289, 314, 401, 420, 585, 732, 743, 745, 751
Fe, 67, 68, 224, 258, 420, 502, 517, 630, 631, 632, 646, 652, 674, 732, 734, 744, 745
See also Materials examples in text, stainless steel, steel
FeAl, 262
Fe_3Al, 262, 426
Fe_2O_3, 295, 296, 297, 298, 395, 503
Fe_3O_4, 521
Fe-Cr-Ni, 633
Fe-Cr-O, 631
Fe-Mo, 649
Fe-Ni, 515
ferrite, 456
ferroelectric, 429, 514
GaAs, 8, 175, 176, 177, 186, 187, 236, 258, 263, 289, 351, 414, 420, 428, 517, 519, 521, 567, 568
Ge, 59, 67, 168, 224, 237, 245, 262, 380, 381, 392, 424, 454, 503, 527, 538, 552, 553, 560, 566, 583, 588, 594, 607, 632, 649
glass
metallic, 293, 752, 753, 754
oxide, 633
silica, 583
silicate, 517
graphite, 164, 186, 342, 410, 448, 528, 554, 583, 743, 745, 755
See also Carbon
hematite, 296, 297, 298, 456
high-T_c superconductor, 99, 514
hydrofluoric acid, 173
icosahedral quasicrystal, 504
InAs, 392
$K_2O \cdot 7Nb_2O_5$, 558
latex particle, 374, 375, 376, 377
Mg, 57, 67, 258, 465, 607, 608, 629, 633, 646, 651, 654, 655, 672, 706, 733, 752
MgO, 68, 86, 176, 224, 261, 290, 317, 397, 409, 467, 468, 545, 655
Mo, 68, 101, 102, 264, 265, 277, 596, 608, 609, 610, 611, 617, 649, 659, 672
MoO_3, 167, 168
Na, 67, 258, 261, 293, 589, 627, 651, 653, 706, 725
NaCl, 176, 186, 261, 264
nanocrystals, 157, 283, 284, 290, 291, 293, 295
Nb, 67, 68, 506, 571, 572, 619, 620, 628, 649, 651, 653
Nb-Al, 649
$Nb_{12}O_{29}$, 539
Ni, 68, 173, 258, 262, 353, 394, 427, 449, 451, 502, 515, 586, 589, 593, 595, 596, 599, 619, 628, 633, 653, 655, 656, 657, 672, 680, 720, 726, 744, 746, 753, 754
NiAl, 262, 263, 420, 427, 503, 656
Ni_3Al, 258, 262, 263, 420, 427, 446, 473, 656, 746
Ni-Cr-Mo, 649
$NiFe_2O_4$, 420, 500
NiO, 175, 261, 394, 400, 402, 420, 426, 429, 453, 454, 455, 500, 501, 502, 513, 552, 593, 596, 599, 611, 614, 654, 720
nitric acid, 173, 174
ordered intermetallic alloy, 500
Pb, 66, 67, 375, 588, 598, 630, 649, 651
perchloric acid, 173, 174, 634
perovskite, 504
polymer, 3, 10, 30, 41, 60, 65, 66, 67, 86, 99, 109, 123, 124, 135, 138, 181, 184, 373, 375, 376, 377, 585, 587, 598, 702, 706, 709, 710
polystyrene, 710
polytype, 504
polytypoid, 504
Pt, 101, 102, 116, 173, 185, 188, 379, 506, 609
quantum-well heterostructure, 182
quartz, 10, 66
quasicrystal, 198, 504–505, 506, 543–544, 752
Sb, 629
Si, 120, 129, 181, 197, 528, 587, 588–589
SiC, 420, 429, 504, 695
Sigma (σ) phase, 729
Si/Mo superlattice, 264–265
Si_3N_4, 187, 188, 503, 695, 710
SiO_2, 65, 174, 177, 293, 376, 380, 381, 402, 566, 693, 710, 720, 746
SnSe, 452
SnTe, 396
spinel, 6, 217, 392, 401, 402, 420, 428, 454, 455, 456, 472, 493, 500, 501, 502, 513, 552
$SrTiO_3$, 6, 381, 706, 736
stainless steel, 174, 326, 341, 420, 424, 616, 633, 634, 657, 674, 724, 734
steel, 11, 378, 427, 514, 633, 659, 723
superconductor, 99, 173, 504, 505, 526
Ta, 68, 136, 137, 598, 627, 629
Ti_3Al, 420
Ti, 7, 67, 68, 130, 314, 629, 630, 631, 635, 651, 716, 723, 736, 744
TiAl, 262
TiC, 723, 724, 749
TiN, 723, 724
TiO_2, 428, 562, 563, 631, 635, 736
U, 487, 505, 506, 715
vanadium carbide, 264, 278, 500
wurtzite, 262, 420, 590
Y, 118
YBCO, 396, 397, 565
yttrium-aluminum garnet, 118
Zn, 68, 258, 447, 448, 449, 618, 627, 649, 651, 653, 708, 744
ZnO, 262, 348, 420, 541
ZnS, 116, 118, 237
Materials safety data sheet, 173
Mean-free path, 23
elastic, 39–50
inelastic, 53–69
plasmon, 63–64
Mean-square vibrational amplitude, 435
Mechanical punch, 176
Microanalysis, 132, 133, 589, 657
qualitative, 581
quantitative, 76, 364, 433, 434, 478
See also Spectrometer (EELS); Spectrometer (X-ray energy-dispersive); Spectrometer (X-ray wavelength-dispersive)
Microcalorimeter, 590–591
Microdensitometer, 85, 371, 550, 551
Microdiffraction, 528
See also Convergent beam, diffraction
Microdomain, 517
Miller-Bravais notation, 260
Miller indices, 46, 49, 204, 212
Mini lens, 145, 146
Minimum contrast, 378, 492, 493, 497, 498, 561
detectability, 379, 497
detectable mass, 663, 674
detectable signal-to-noise ratio, 497
mass fraction, 663, 674
resolvable distance, 107
MINIPACK-1, 571
Mirror plane, 358, 360, 361
See also Point group; Symmetry
Mirror prism, 691
Modulated structure, 503, 504
Moiré fringes, 284, 298, 392, 393–397, 456
complex, 396–397
general, 393
rotational, 393, 394
translational, 393, 394
Molecular-orbital theory, 745

Möllenstedt spectrometer, 352
Momentum transfer, 753, 755
Monochromator, 76, 86, 105, 106, 319, 681, 693–694, 701, 705, 707, 722
Monte-Carlo simulation, 523
Moore's Law, 362
Moseley's Law, 58
Muffin-tin potential, 748, 749
Multi-channel analyzer (MCA), 591
Multi-element spectrum, 644, 647, 654
Multi-phase specimen, 665
Multi-photon microscopy, 4
Multiple domains, 304
Multiple least-squares fitting, 645, 731, 732
Multiple scattering, *See* Scanning transmission electron microscope (STEM), multiple
Multislice calculation, 533, 534, 536, 543, 544, 571
Multivariate statistical analysis, 619, 620, 659
Multi-walled nanotube, 365, 366
Murphy's law, 123, 707, 731

N

Nanocharacterization, 4
Nanodiffraction, 283, 291, 323, 324, 347, 365, 366
Nanomaterials, 4, 174, 483, 694, 704
Nanoparticles, 4, 189, 271, 276, 302, 366, 400, 632, 635, 695
Nanostructured electronics, 362
Nanotechnology, 3, 4, 154, 323, 324
Nanotubes, 4, 73, 77, 81, 365, 366, 689, 694, 695, 705, 745
Nanowires, 4
Near-field calculation, 535
Near-field microscopy, 4, 30, 535
Near-field regime, 229, 397
Néel wall, 517
Nematode worms, 756
NIST, 14, 29, 44, 58, 304, 608, 631, 633, 646, 647, 652, 653, 654
  multi-element glass, 647
  oxide glass, 633
  Sandia/ICPD electron diffraction database, 304
  thin-film standard (SRM 2063), 631, 641, 654
Noise, 81, 115, 116, 118, 119, 121, 122, 124, 376, 380, 386, 464, 466, 478, 492, 495, 497, 498, 522, 528, 541, 549, 556, 557, 561, 562, 563, 565, 567, 570, 572, 586, 588, 591, 593, 594, 598, 608, 619, 620, 631, 659, 675, 688, 689, 709, 725, 732
  reduction, 556, 557, 565, 675
  See also Signal-to-noise ratio

O

Objective lens, 101, 111, 152, 154–158, 161, 162, 167, 207, 278, 332, 372–373, 375, 376–379, 382, 385–386, 389, 390, 396, 411, 413, 448, 466, 489, 500, 511, 512–514, 516, 519, 520, 521, 528, 529, 539, 540, 552, 573, 600, 670, 686–688, 691, 706, 721, 755
  aperture, 91, 101–102
  astigmatism, 106, 162, 163
  collection semiangle of, 34
  defocus, 162, 163, 331, 553, 565
  diaphragm, 156
  focal increment of, 169
  instability of, 466
  polepiece, 143, 144, 599
  rotation alignment of, 162
  transfer function of, 485, 486, 487–488, 490, 491, 492, 494, 495, 552, 560
  See also Lens
Oblique-textured electron DP, 291
Omega (Ω) filter, 681, 691–692
On-axis image, 297, 391, 559, 560
Optical bench, 92, 198, 549, 573–574
Optical system, 82, 386, 389, 483–484, 490, 495
Optic axis, 34, 75, 79, 92, 96, 99, 100, 101, 104, 106, 110, 131, 143, 144, 147, 149, 151, 155, 158, 161, 162, 198, 205, 248, 285, 305, 313, 317, 382, 384, 390, 397, 463, 466, 485, 493, 514, 515, 519, 533, 599, 688, 691, 701
  See also Lens
Ordering, 263, 264, 272, 277, 278, 279, 366, 427, 517
  long-range, 279
  short-range, 277, 278, 517
Orientation imaging, 319
Orientation mapping, 305
Orientation relationship, 204, 283, 289, 302–303, 305, 339, 500
  cube/cube, 303
  Kurdjumov–Sachs, 303
  Nishiyama–Wasserman, 303
  precipitate-matrix, 302
O-ring, 131, 132, 133
Overfocus, 96, 116, 143, 145, 147, 148, 149, 162, 163, 164, 165, 207, 325, 330, 331, 400, 485, 500, 514, 515, 517, 626
  See also Lens; Underfocus
Overvoltage, 57, 626
Oxide layer, 504, 519, 710

P

Parallax shift, 511, 512, 513
Paraxial ray condition, 100, 104
Particle on a substrate, 81, 396, 413
Pascal, 128
Passband, 492–493
Path difference, 31, 33, 48, 49, 200, 201, 202, 488
Path length, 33, 100, 599, 600, 609, 655, 656, 682
  See also Absorption, of X-rays
Pathological overlap, 628, 630
Pattern recognition, 561, 562–563, 569
Pauli exclusion principle, 744
Peak-to-background ratio, 614
Peak (X-ray characteristic), 469–470, 605–606, 614, 627–634, 644–646, 701–702
  deconvolution of, 630–632
  integration of, 644–646, 650, 727, 729
  overlap of, 589, 590, 597, 627, 630
  visibility of, 632–634
Periodic continuation method, 542
Phase boundary, 191, 419, 420, 429, 447, 502, 503
  distortion function, 488
  of electron wave, 31, 47
  factor, 191, 419, 420, 429, 447, 502, 503
  grating, 534, 535, 536
  negative, 487
  object approximation, 486
  reconstructed, 557
  shift, 46, 471, 486, 488, 491, 526, 753
  transformation, 136, 456, 526
  See also Contrast, difference; Interface
Phase contrast, 77, 86, 106, 163, 164, 169, 371, 373, 380, 381, 389–403, 411, 487, 488, 490, 492, 493, 494, 495, 505, 511, 515, 543, 545, 557, 668, 696, 703
Phasor diagram, 31, 32, 421, 426, 470–473
Phonon, 59, 63–64, 336, 680, 700, 701, 702
Phosphorescence, 116
Photo-diode array, 683
  See also Diode array
Photographic dodging, 550
Photographic emulsion, 66, 122–124, 197, 464, 561
Photomultiplier, 117, 118–120
  See also Scintillator-photomultiplier detector
p-i-n device, 586
Pixel, 120, 121, 123, 124, 159, 305, 496, 497, 556, 562, 563, 564, 567, 568, 569, 571, 590, 618, 619, 620, 636, 658, 675, 685, 694, 727, 734, 736
Pixel-clustering, 734
Planar defect, 250, 254, 263, 275–277, 286, 302, 347, 419–436, 452, 472, 503, 504, 600, 616, 669, 670
  inclined, 250, 431, 669
  See also Grain, boundary; Stacking fault; Twin boundary
Planar interface, 275, 503, 600, 694, 748, 749
Plane normal, 213, 287, 288, 299, 300, 301, 302, 303, 317, 419, 458, 519, 538, 670, 682
Plane wave, 31, 32, 33, 40, 45, 46, 48, 49, 200, 237, 245, 249, 250, 748
  amplitude, 239–241
Plasma cleaner/cleaning, 132, 137, 138, 189, 626
Plasmon, 63–64, 109, 680, 682, 693, 700, 703, 705–708, 710, 717, 719, 720, 722, 726, 731, 732, 752, 756

energy, 63, 64, 109, 706, 709
excitation, 54, 63, 700, 717
fingerprinting, 704, 708
frequency, 63
loss, 682, 702, 706, 707, 708, 709, 711, 717, 718, 719, 731, 732
See also Low-loss, spectrum
mean-free path, 63–64
peak, 63, 680, 701, 703, 704, 705, 706, 707, 708, 709, 710, 717, 720, 722, 723, 726, 730, 752
Plural elastic scattering, 741
See also Elastic, scattering; Scanning transmission electron microscope (STEM), elastic
p-n junction, 62, 117, 118, 523
Point analysis, 354
Point defect, 65, 277, 278, 448, 502
See also Interstitial atom; Vacancy
Point group, 9, 332, 347, 354, 358, 359, 361, 364, 366
determination of, 358
symmetry of, 237, 358, 361
two-dimensional, 354, 486, 511
Point-to-point resolution, 493, 506
Point-spread function, 483, 593, 630, 688–689, 701, 753
Poisson's ratio, 444, 457
Poisson statistics, 593
Polepiece, *See* Lens
Polycrystalline material, 290–291, 293, 319, 452, 613
Polymer, 3, 10, 30, 41, 60, 64, 65, 66, 67, 86, 99, 109, 123, 124, 132, 135, 138, 181, 183, 184, 373, 375, 376, 377, 585, 587, 588, 598, 702, 706, 709, 710
Polytype, 504
Polytypoid, 504
Position-tagged spectrometry, 620, 659
Post-specimen lens, 85, 110, 161, 326, 329, 342, 379, 683, 686
See also Lens
Potential
inner, 235, 236, 237, 238, 239, 240, 397, 399, 400, 402, 540, 541, 545
periodic, 236, 237, 254
projected, 486, 487, 511, 534, 537, 538, 541, 557, 563, 565
well, 120, 402, 541, 542, 720, 721, 749
Precession CBED, 342
Precession diffraction, 147, 158, 284, 285, 293, 295
Precision ion milling, 181
Precision ion polishing, 181
Primitive great circle, 286, 287
See also Stereogram
Primitive lattice, 257–258, 356, 357
See also Crystal; Lattice
Principal quantum number, 744
Probability map, 536
Probe, 8, 9, 48, 81, 82, 95, 98, 103, 111, 124, 135, 144, 146, 148, 149, 150, 158, 161, 189, 291, 325, 339, 340, 362, 377, 402, 496, 522, 584, 589, 590, 610, 611, 612, 614, 615, 616, 618, 625, 635, 636, 647, 659, 668, 672, 673, 680, 694, 733, 735, 750
current, 84, 85, 110, 124, 143, 149, 150, 610, 612, 614, 616, 617, 618, 626, 627, 641, 653, 668, 683, 733
size, 82, 84, 85, 86, 144, 145, 148, 149, 150, 326, 523, 590, 614, 626, 627, 636, 640, 641, 659, 668, 673, 733, 735
See also Beam (electron)
Processing HRTEM image, *See* Image, processing of
Propagator matrix, 423
Pulse processing, 591, 592, 596, 598, 635, 636
Pump, vacuum, 127–138, 178, 180, 181, 184, 189, 586, 683, 687
cryogenic, 130
diffusion, 129, 130, 131, 180, 191
dry, 189
ion, 130, 131, 137
roughing, 128–129, 130
turbomolecular, 129–130

Q

Quadrupole, 98, 99, 104, 105, 683, 692
See also Lens
Qualitative mapping, 635
Qualitative microanalysis, 581
Quantifying HRTEM images, 549–575
Quantitative chemical lattice imaging, 517, 567–568
defect contrast imaging, 411, 422, 470
HRTEM, 567
image analysis, 561–562
mass-thickness contrast, 373–379, 381, 382, 384, 696, 702
microanalysis, 132, 133, 589, 657
Quantitative mapping, 658, 659, 671
QUANTITEM, 563–567
Quantum-mechanical convention, 230
Quantum number, 744
Quasicrystal structure, 198, 504–505, 506, 543–544, 752

R

Racemic mixture, 636
Radial-distribution function (RDF), 293, 294, 373, 703, 741, 752, 753, 755
Radiation damage, 6, 10, 53, 64, 65, 66, 68, 119, 448
See also Beam (electron), damage
Radiolysis, 64, 65, 66, 68, 646
See also Beam (electron), damage
Ray diagram, 91, 92–94, 100, 101, 102, 103, 104, 105, 111, 143, 147, 152, 154, 156, 157, 198, 295, 325, 326, 327, 328, 330, 333, 610, 692
Rayleigh criterion, 5, 84, 107, 108, 490
Rayleigh disk, 108, 484
Real space, 211, 212, 213, 226, 236, 258, 262, 264, 265, 271, 279, 286, 302, 319, 332, 355, 356, 410, 412–413, 427, 485, 492, 534, 536, 542, 572, 748
approach, 534, 536, 563, 572
crystallography, 412–413
patching method, 542
unit cell, 262
vector, 236
Reciprocal lattice, 200, 202, 211–212, 213–216, 235, 254, 258, 259, 260, 262, 271, 273, 278, 280, 289, 290, 292, 297, 336, 337, 348, 356, 357, 430, 431, 435, 470, 489, 538, 552
formulation of, 535
origin of, 215
point, 202, 215, 216, 253, 258, 262, 278, 290, 336, 351, 357, 470, 552
rod, 214, 273, 337
See also Relrod
spacing, 338, 538
vector, 200, 211, 212, 213, 235, 239, 290, 297, 302, 485, 489
See also Diffraction, vector (g)
Reciprocity theorem, 94, 381, 386, 521
Recombination center, 524
Reference spectra, 701, 723, 732, 733
Reflection electron microscopy, 420, 519–520
Reflection high-energy electron diffraction, 519
Refractive index, 5, 225, 230, 238, 239, 358
Relative-transition probability, 650
Relative transmission, 670
Relativistic effect, 6, 14, 41, 42, 700
Relrod, 214, 215, 216, 249, 271–277, 279, 280, 281, 289, 336, 337, 338, 410, 430, 431, 441
See also Reciprocal lattice, rod
Replica, 164, 165, 185, 377, 378, 616, 724
Resolution, 5–7, 91–112, 483–507, 589, 663–676, 735
atomic level, 381
limit, 4, 6, 103, 104, 109, 323, 490, 492, 493, 494, 594, 701, 703, 730
theoretical, 107–108, 594
Resolving power, 5, 33, 106, 107, 124
Reverse-bias detector, 585
Richardson's Law, 74
Right-hand rule, 99
Rigid-body translation, 420
Rose corrector, 494

S

Safety, 10, 102, 173–174, 175, 176, 178, 189
Scan coil, 147, 149, 157, 158, 159, 165, 293, 295, 305, 326, 366, 753
Scanning image, 101, 115, 116, 118, 122, 124, 159, 161, 376

Scanning transmission electron microscope (STEM), 8, 158–161, 326, 372–373, 376–377, 384–386, 528
annular dark-field image, 161
bright-field image, 122, 159–161
coherent, 373, 379
cross-section, 158, 168
dark-field image, 161
detectors in, 326, 366, 372–373, 385, 386, 511, 528, 687
diffraction contrast in, 384–386
digital imaging, 618
elastic, 373
factor, *See* Atomic, scattering factor
forward, 122, 735
image magnification in, 165
incoherent, 39, 43, 106, 107, 378, 379, 381
inelastic, 319
inter-shell, 749
intra-shell, 749
mass-thickness contrast in, 377
matrix, 185, 374
mode, 124, 151, 159, 166, 326, 584, 626, 632, 685, 721, 733, 734, 735, 755
multiple, 493, 699, 748, 753
multiple-scattering calculations, 753
nuclear, 42, 45
plural, 493, 699, 700, 721
post-specimen, 110, 161, 326, 342
Rutherford, 41–44, 161, 373, 374, 378
semiangle of, 84, 378
single, 29, 43
strength, 83, 135, 326, 354, 584
thermal-diffuse, 64, 336, 338
Z contrast, 44, 379–381, 506, 543, 545, 567, 746
See also Angle; Coherent; Elastic; Z contrast, scattering
Scherzer, 490–491, 492, 494, 495, 498, 500, 553
defocus, 490–491
Schottky, 62, 73, 74, 75, 81, 82, 117, 497, 498, 617, 683, 693
diode, 117
See also Detector (electron)
emitter, 75, 498
Schrödinger equation, 46, 222, 230, 235, 236, 237, 238, 239, 748
Scintillation, 116
Scintillator-photomultiplier detector, 118–120
Screw axis, 543, 544
See also Space Group; Symmetry, screw axis
Secondary electron, 24, 53, 54, 60–62, 115, 118, 188, 373, 522, 709
detector, 373
fast, 705
imaging of, 522–523
slow, 605, 606
types of, 522
Segregation to boundaries, 659
Selected area diffraction (SAD), 141, 152–155, 156, 157, 160, 166, 167, 204–207, 283, 284–285, 289, 295, 311, 313, 323, 324, 325, 326, 330, 332, 333, 334, 339, 342, 347, 348, 352, 365, 376, 411, 412, 413, 493, 498, 505, 525, 551, 680, 688
aperture, 152, 154, 155, 158, 166, 167, 205, 206, 332, 339, 376, 411, 493, 525, 551, 688
error, 498
pattern exposure, 154, 157, 160, 205, 206, 283, 284, 412
Selection rules, 257, 258, 267, 289, 302, 744
Semiangle, *See* Angle; Bragg; Collection semiangle; Convergence semiangle; Incidence semiangle
Semiconductor detector, 117–118, 119, 122, 161, 523, 581, 585–589, 594, 607
Semi-quantitative analysis, 723
Shadowing, 185, 374, 375, 377, 378, 541
Shape effect, 271, 273, 290, 559, 572
Short-range ordering, 277, 278, 517
Side-entry holder, 132, 133, 134, 135, 150, 285
See also Specimen, holder
SIGMAK(L) program, 729
Signal-to-background ratio (jump ratio), 703, 722, 728, 732, 734, 736
Signal-to-noise ratio, 116, 118, 124, 497, 498, 556, 570, 688
Signal processing, 376, 377, 572, 589, 590, 594, 606
Silicon-drift detector, 588–589
Silicon dumbbells, 391, 575
Si(Li) detector, 585, 586, 587, 588, 591, 594, 595, 606, 607, 626, 628, 630, 638, 651
See also Spectrometer (X-ray energy-dispersive)
Simulated probe image, 668
Single-atom detection, 663, 715
Single-atom imaging, 54, 378, 663, 674, 679, 715, 736
Single-electron counting, 700, 709
Single-electron interaction, 709
Single-period image, 564
Single scattering, *See* Scanning transmission electron microscope (STEM), single
Single-sideband holography, 524
SI units, 14, 65, 128, 654, 665
Slow-scan CCD, 478, 553, 559, 570, 692
See also Charge-coupled device (CCD) camera
Small-angle cleaving, 186
Small circle, 262, 287, 428
See also Stereogram
Smearing function, 483, 487
See also Point-spread function
Space group, 9, 17, 47, 266, 267, 296, 347, 354, 358, 361, 540, 558
Spatial coherence, 77, 398, 491, 498
Spatial resolution, 8, 29, 54, 62, 76, 77, 85, 87, 148, 323, 324, 325, 329, 347, 348, 352, 362, 490, 523, 581, 589, 619, 625, 626, 640, 647, 658, 659, 663–676, 688, 705, 721, 730, 733, 735–736, 752, 753, 755
Specimen
90°-wedge, 186, 187, 414
artifacts in, 541
bulk, 25, 44, 135, 136, 284, 520, 589, 599, 607, 608, 617, 635, 639, 640, 643, 646, 647, 650, 655, 674
cooling of, 10
damage to, 10, 24, 64, 636, 673, 680, 753
See also Beam (electron), damage
density of, 656
double-tilt, 134, 135
drift of, 136, 376, 466, 495, 496, 584, 616, 620, 641, 647, 658, 668, 673
EBIC, 136, 523
heating, 64, 65–66
height of, 101, 150, 151, 327, 512, 683
See also z control
holder, 10, 11, 82, 97, 121, 127, 132–133, 134, 135, 150, 151, 169, 175, 187, 207, 512, 514, 612, 613, 653, 655
low-background, 135, 324, 584, 600, 626
multiple, 134, 135
orientation of, 238, 323, 494, 614, 630, 721, 733, 755
preparation of, 11, 134, 173–192, 416, 503, 616, 626, 633, 669, 671
quick change, 134
rotation of, 285
self-supporting, 173, 174, 175
single-tilt, 134, 135, 324
single-tilt rotation, 324
spring clips for, 133
straining, 136, 137
surface of, 61, 62, 65, 136, 158, 179, 275, 395, 411, 433, 448, 455, 522, 523, 599, 633, 707
thickness of, 11, 29, 63, 109, 110, 111, 164, 197, 323, 329, 352, 402, 466, 487, 565, 595, 627, 654, 655, 656, 669, 671, 675, 679, 702, 705, 708, 721, 726, 727, 730
See also Thickness of specimen
tilt axis, 169
tilting of, 134, 181, 187, 228, 274, 285, 289, 382, 394, 430, 447, 511, 515, 671, 755
top-entry, 133, 134, 136, 169
transmission function, 485
vibration, 495
wedge-shaped, 274, 408, 410, 564, 565
Spectrometer (EELS)
aberrations of, 682, 683, 684
artifacts in, 689–690

calibration of, 684
collection semiangle of, 685–688
dispersion of, 683
entrance aperture of, 681, 682, 684, 686, 687, 688
entrance slit of, 294
focusing of, 682–683
object plane of, 683, 685
post-spectrometer slit, 683
resolution of, 683–684, 722
See also Electron energy-loss spectrometry (EELS)
Spectrometer (X-ray energy-dispersive), 7, 581
artifacts of, 689–690
Au absorption edge, 613
Au contact layer, 651
automatic shutter, 611
clock time, 593, 596, 597
collection angle of, 101, 598–599
contamination of, 132
dead layer, 585, 586, 595, 596, 651
dead time, 592, 593, 596, 597, 598
efficiency of, 715, 721
escape peak, 606, 607, 612, 614, 628, 630, 656
incomplete-charge collection, 59
internal-fluorescence peak, 612, 613, 630
leakage current of, 593, 689, 690, 691
live time, 593, 596, 614
performance criteria for, 688
residuals in, 645
shutter, 594, 595, 600, 611
sum peak, 607, 608, 628, 630, 632
system peaks, 613, 627, 628, 632
take-off angle of, 614, 654, 655
time constant, 593, 596, 597, 598, 627, 630
window, 702, 703, 725
Spectrometer (X-ray wavelength-dispersive), 589–591
Spectrum
electron energy-loss, 680–681
X-ray, 26, 57, 60, 175, 591, 605–621, 643, 699
Spectrum imaging, 605, 619–620, 659, 680
Spectrum-line profile, 616, 617, 620, 695, 733
Sphere of projection, 286
See also Stereographic projection
Spherical aberration, 6, 84, 100, 103–104, 108, 148, 205, 331, 488, 490, 494, 610, 635, 703
broadening, 84
coefficient, 84, 104
error, 108
See also Aberration
Spinodal decomposition, 504
Spinorbit splitting, 744
Spin quantum number, 744
Spot mode, 144, 149, 326, 584, 615–616
Spurious peak, 628
Spurious X-ray, 608, 609, 611, 635, 655
Sputtering, 65, 66, 68, 178, 180, 181, 646, 721
Stacking fault, 67, 250, 275, 279, 419, 420, 421, 422, 424–427, 441, 446, 447, 448, 458, 471, 472, 473, 475, 476, 503, 524
contrast, 424, 425, 426
energy, 446, 448, 458, 473, 476
fringes, 402
inclined, 472
intrinsic, 475
overlapping, 426–427
See also Planar defect
Stage, 24, 34, 62, 68, 81, 98, 101, 115, 118, 119, 121, 122, 123, 127, 130, 131, 132–133, 136, 137, 138, 150–152, 154, 169, 177, 188, 197, 214, 223, 224, 327, 330, 347, 374, 466, 512, 515, 522, 523, 540, 584, 586, 587, 589, 598, 599, 605, 611, 612, 614, 616, 650, 655, 673, 753
Staining, 66
Standard Cr film, 518, 593, 612, 614, 615
Standard specimen, 164, 168, 305, 362, 599, 652, 744
Stationary-phase method, 470
Statistical criterion, 632, 673, 674
Statistically significant peak, 632
Stereogram, 286, 287, 303, 359
Stereographic projection, 286–287, 288, 302, 311, 315, 317, 348, 358
Stereology, 512
Stereomicroscopy, 285, 442, 450
Stigmators, 106, 149, 162, 163, 164
See also Astigmatism
Strain, 127, 135, 136, 278, 279, 339, 347, 361–363, 394, 396, 399, 400, 415, 441–442, 443, 444, 447, 448, 451, 452, 456, 457, 468–469, 476, 526
analysis of, 362
contrast, 399, 456, 457
field, 279, 371, 394, 399, 400, 441–459, 468–469, 476
lattice, 347, 361–363, 456
measurement of, 347, 352, 361–363
Strain engineering, 362
Straining holder, 136, 137, 551
See also Specimen, holder
Strain-layer superlattice, 362
Stress, 74, 81, 133, 134, 136, 182, 189, 324, 341, 375, 441, 477, 486, 495, 504, 527, 625
field, 136, 441
Strong-beam image, 382, 421, 424, 425, 467, 474, 476
Structure correlation, 294
Structure factor, 46–47, 223, 224, 257, 258–259, 260, 261, 262, 267, 272, 274, 284, 290, 332, 336, 347, 351, 356, 364–365
Structure-factor determination, 261, 364–365
Structurefactor-modulus restoration, 559
Student *t* value, 648
Substitutional atom, 277, 657
Substitutional site, 658
Substrate, 4, 81, 186, 187, 296, 302, 379, 393, 394, 395, 396, 397, 413, 452, 518, 519, 583, 751
Summation, 84, 108, 223, 239, 272–273, 569, 654
Sum peak, 592, 606, 607, 608, 628, 630, 632, 633, 638
Superlattice, 262–265, 277, 362, 415, 420, 427, 447, 504, 517
dislocation, 447
reflection, 262–265, 427, 517
See also Ordering
Surface barrier detector, 117, 118
diffusion, 527
dislocation, 452, 453
faceting, 392
groove, 540–542, 751
imaging, 519–521
layer, 81, 117, 182, 187, 380, 451, 452, 453, 519
plasmon, 707
reconstruction, 519
relaxation, 363, 458, 517, 527
of specimen, 61, 62, 65, 136, 179, 275, 395, 411, 433, 448, 455, 511, 522, 523, 599, 633, 707
Symmetry
bright-field projection, 155, 159–161, 361
determination of, 295, 340, 354, 357–361, 366
elements of, 266, 287, 305, 357, 358, 359, 361
glide plane, 442, 443, 444, 449
inversion, 359, 360, 365
projection-diffraction, 361
rotational, 360, 361
rotation axis, 358, 360, 412
screw axis, 543, 544
three dimensional, 364
translational, 504, 544
whole-pattern (WP), 361
zone-axis, 348, 352, 380, 383, 412, 458
See also Mirror plane; Point group; Space group
Systematic absence, 259, 260, 290, 304, 348, 354, 356, 357
See also Forbidden reflection
Systematic row, 202, 205, 230, 249, 332, 411, 435, 457, 458, 464, 465, 466, 657
See also Diffraction, pattern
System peaks, 613, 627, 628, 632

T

Template, 305, 396, 397, 562, 563, 564, 567, 568, 569, 570, 572, 573, 659, 693
Temporal coherence, 76–77, 85
See also Spatial coherence

Texture, 283, 284, 290, 291, 292, 319, 557
See also Grain
Thalidomide, 363
Thermal-contraction coefficient, 363
Thermal-diffuse scattering, 64, 338
Thermionic gun, 74, 77–80, 82, 144
See also Gun
Thickness of specimen, 703
dependence, 538, 567
determination of, 352–354, 361, 415, 670, 671, 730
effect on contrast, 373–379
effective, 337, 415
fringe, 389, 407, 408–411, 413, 414–415, 424, 426, 430, 432, 446, 452, 456, 467, 468, 472, 521, 669
image, 731
simulation of, 414
Thin-foil criterion, 640, 654
See also Cliff-Lorimer
Thin-foil effect, 217–218, 273–274
See also Shape effect
Threshold energy, 67, 750
See also Beam (electron), damage
Through-focus dark-field, 513
See also Dark field (DF)
Through-focus image, 402, 493, 506
Tie line, 248, 249, 250, 251, 252, 254, 413, 431, 432
See also Dispersion diagram, surface
Tilted-beam condition, 390
Tomography, 10, 135, 442, 511–512, 741, 755–757
Top bottom effect, 297, 298
Top-hat filter function, 643, 644
Topotaxy, 296
Torr, 128, 129, 130
Trace element, 625, 632, 634, 659, 672, 732
Transfer function, 485, 486, 487–488, 490, 491, 492, 494, 495, 552, 560
effective, 491
See also Contrast, transfer function
Translation boundary, 419
Transmission electron microscope, 3–18
Tripod polisher, 177, 178
TV camera, 111, 118–120, 141, 326
See also Charge-coupled device (CCD) camera
Twin boundary, 6, 169, 207, 217, 276, 402, 419, 420, 427, 454, 455, 472, 503
See also Grain, boundary; Stacking fault
Twin-jet apparatus, 178
Twin lens, 145, 146
Two-beam approximation, 224, 227, 390, 424, 435, 478
calculation, 457
condition, 224, 230, 296, 352, 381–382, 383, 384, 385, 422, 444, 658, 669, 733
Two-photon microscopy, 4
Two-window method, 643

U

Ultrahigh vacuum, 13, 55, 128, 129–130
Ultramicrotomy, 180, 183–184, 634
Ultra-soft X-rays, 590
Ultrasonic cleaning, 176, 182, 184
Ultra-thin window, 586, 587, 588
See also Spectrometer (X-ray energy-dispersive)
Uncollimated electrons, 610
Underfocus, 96, 142, 143, 145, 147, 148, 149, 154, 162, 163, 164, 165, 198, 205, 207, 229, 330, 381, 382, 400, 466, 492, 500, 514, 515, 516, 517, 553
See also Lens; Overfocus
Unfilled states, 55, 742, 743, 744, 746
Unit cell, 46, 47, 212, 223, 257, 258, 261, 262, 264, 272, 273, 274, 324, 338, 347, 354–357, 421, 427, 435, 447, 500, 540, 542, 544, 558, 563, 567, 571, 655, 665, 669, 748
determination of, 354–357
image, 567
scattering amplitude from, 258
volume of, 28, 65
See also Crystal; Lattice

V

Vacancy, 66, 67, 277, 449, 528
loop, 449
See also Dislocation, loop
ordered array, 264, 278, 542
Vacuum
backing valve, 131
high, 128, 129–131, 523, 752
low, 128
pumps, 127
rough, 128
tweezers, 175, 189
ultrahigh, 129–131
wave vector in, 249
Vacuum level, 742
Valence-electron density, 365, 709
Valence state image, 679, 743, 744, 745, 747, 748
Valence-ultraviolet spectrum, 705
Végard's law, 363
Video image, 550
See also Charge-coupled device (CCD) camera
Viewing screen, 5, 11, 26, 75, 76, 82, 85, 97, 115, 116–117, 119, 124, 141, 143, 144, 146, 148, 149, 151, 152, 153, 154, 155, 156, 162, 166, 199, 224, 371, 372, 681, 682, 685, 686, 691, 694, 696
Visible-light microscope, 5, 86, 91, 198
Void, 374
Voltage centering, 162, 493
Von Laue, 48, 49, 358
condition, 199, 214, 216, 271
equations, 48, 49, 201, 203, 213–214
groups, 49
X-ray pattern, 218
zone, 216, 217, 289, 290, 311, 323, 335–338, 349, 351, 355, 356, 357, 458, 614
See also Higher-order Laue zone (HOLZ); Zero-order Laue zone

W

Wafering saw, 176
Water vapor, 124, 127, 132, 136, 189, 586, 587, 595
Wave
diffracted, 48, 199, 200, 259, 536
equation, 235–236, 247, 495, 728
function, 33, 222, 224, 227, 236, 237, 241, 252, 399, 486, 487, 535, 728, 748, 753
incident, 34, 36, 48, 199, 215, 216, 486
matching construction, 248
total amplitude of, 31
total function, 222, 224, 227–228, 236, 252
vector, 45, 199, 200, 215, 216, 222, 224, 227, 238, 239, 241, 245, 246, 248, 249, 251, 399, 432, 435
Wavefront, 30, 31, 32, 45, 103, 199, 200, 223
Wavelength-dispersive spectrometer (WDS), 589–591
See also Spectrometer (X-ray wavelength dispersive)
Weak-beam dark field, 463–479
3g reflection, 464, 466
condition, 453
of dislocation, 469–470
thickness fringes in, 467
See also Dark field (DF)
Weak phase-object approximation, 486
WEBeMAPS, 350, 351
Wedge specimen, 186, 187, 249, 250, 275, 279, 408, 414, 430, 431, 446, 472, 641
Wehnelt, 75, 77, 78, 79, 132
bias, 75, 78, 79
cylinder, 77, 78
See also Gun; Thermionic gun
Weighting factor, 258
Weiss zone law, 204, 289, 335, 348
White line, 463, 723, 728, 729, 736, 744, 745, 746, 748
White noise, 497, 572
White radiation, 198, 215
Wien2k code, 711
Wien Filter, 693
Windowless (X-ray energy-dispersive spectrometer) detector, 587, 630, 723
Window polishing, 183
Window (X-ray energy dispersive spectrometer), 586–587
Wobbling (lens), 148, 162
Work function, 74
World Wide Web, 15
Wulff net, 286, 287, 302

X
X-ray
absorption of, 28, 294, 599, 600, 634, 641, 653, 656, 671, 694, 717, 742, 751, 752
See also Absorption, of X-ray
absorption near-edge structure, 742
atomic-number correction, 640, 650
bremsstrahlung, 40, 55, 60, 135, 598, 605, 606, 608, 609, 610, 611, 613, 635, 643
characteristic, 653
count rate, 589, 598, 599, 616, 625, 641, 658, 673
depth distribution of, 654, 655, 665
detector, *See* Spectrometer (X-ray energy-dispersive)
diffraction, 5, 30, 197, 198, 215, 290, 545
emission of, 55–60, 61, 63, 650, 654, 657, 658, 669, 715, 733, 735
energy of, 7, 55, 58, 60, 581, 582, 585, 587, 588, 591, 593, 594, 595, 605, 606, 626, 629, 641, 643, 652, 653
energy-dispersive spectrometry (XEDS), *See* Microanalysis; Spectrometer (X-ray energy-dispersive)
families of lines, 605, 628, 629
fluorescence, 612, 721
image, 605, 616–620, 634, 636, 658, 755
interaction volume, 663, 664, 665, 672, 674, 735
map, 381, 550, 567, 584, 588, 616, 617, 618, 619, 658–659, 672
peak-to-background ratio, 673
scattering factor, 435
spectra, 87, 581, 590, 605–621, 645, 647, 653, 688, 717, 723, 742, 751
See also Microanalysis; Spectrometer (X-ray energy-dispersive); Spectrometer (X-ray wavelength-dispersive)

Y
YAG scintillator, 685, 688, 691
Young's modulus, 457
Young's slits, 13, 30, 31, 279

Z
ZAF correction, 639
Z contrast, 44, 161, 379–381, 506, 543, 545, 567, 746
scattering, 379, 380
z control, 151, 327, 330, 332, 512
Zero-energy strobe peak, 631
Zero-loss, 630, 671, 680, 683, 690, 691, 699, 701–703, 707, 725, 730, 731, 753
integral, 729
peak, 701–703
Zero-loss filter, 703
Zero-order Laue zone, 216, 289
Zero-order wave, 48
Zeta-factor, 652–654, 656
z-factor, 641, 646, 653, 654, 656, 659, 671, 672, 675
Zone axis, 348, 352, 380, 383, 412, 458
high-symmetry, 338, 356
image, 537
low-symmetry, 338, 356
orientation, 348, 380, 614
pattern, 216, 303, 319, 412
symmetry, 338, 356