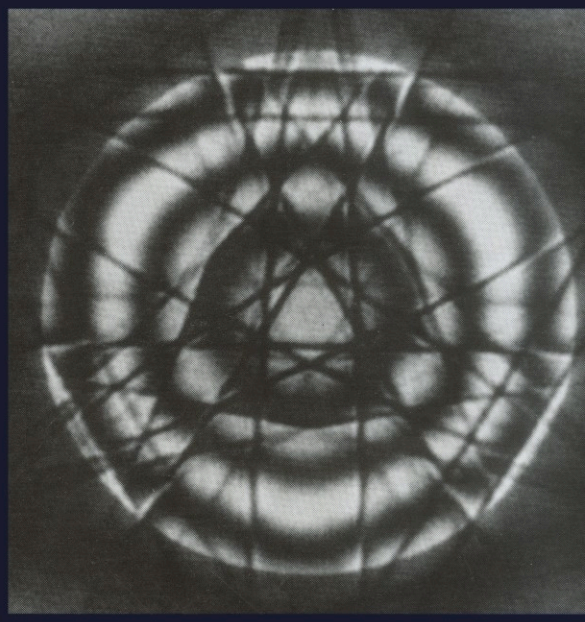
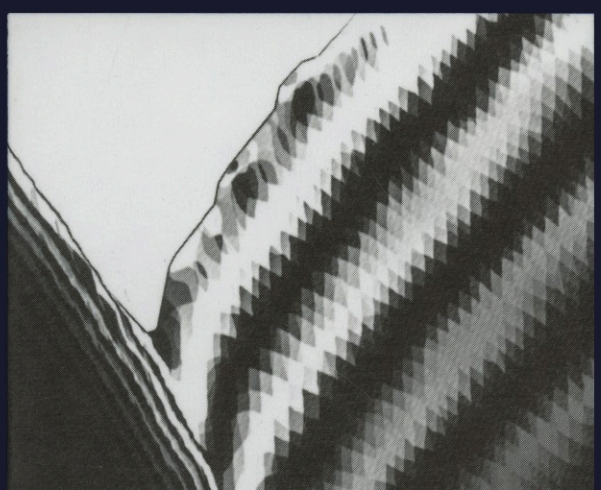
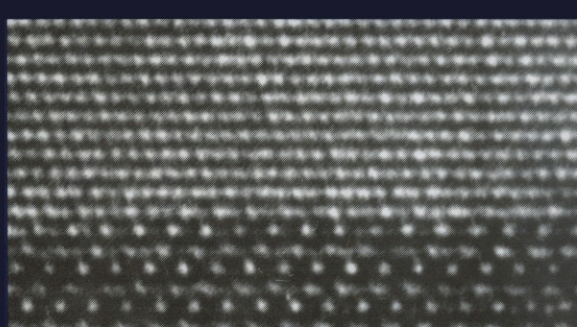
TRANSMISSION ELECTRON MICROSCOPY

A Textbook for Materials Science

David B. Williams
and C. Barry Carter







Transmission Electron Microscopy

A Textbook for Materials Science

Transmission Electron Microscopy

A Textbook for Materials Science

David B. Williams

Lehigh University
Bethlehem, Pennsylvania

and

C. Barry Carter

University of Minnesota
Minneapolis, Minnesota

Springer Science+Business Media, LLC

Library of Congress Cataloging in Publication Data

Williams, David B. (David Bernard), 1949–
Transmission electron microscopy: a textbook for materials science / David B. Williams and C. Barry Carter.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-306-45324-3 ISBN 978-1-4757-2519-3 (eBook)
DOI 10.1007/978-1-4757-2519-3
1. Materials—Microscopy. 2. Transmission electron microscopy. I. Carter, C. Barry. II. Title.
TA417.23.W56 1996 96-28435
502'.8'25—dc20 CIP

ISBN 978-0-306-45324-3

Originally published by Plenum Press, New York in 1996
Softcover reprint of the hardcover 1st edition 1996

10 9 8 7 6 5 4 3

To our parents,
Walter Dennis and Mary Isabel Carter,
and Joseph and Catherine Williams,
who made everything possible

Foreword

Electron microscopy has revolutionized our understanding of materials by completing the *processing–structure–properties* links down to atomistic levels. It now is 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 of 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 TEM correctly to perform imaging, diffraction, and spectroscopy—it's all there! This is, to my knowledge, 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 developments 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* of 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

Preface

How is this book any different from the many other books that deal with TEM? It has several unique features, but the most distinguishing one, we believe, is that it can really be described as a "textbook"—that is, one designed to be used primarily in the classroom rather than in the research laboratory. We have constructed the book as a series of relatively small chapters (with a few notable exceptions!). The contents of many chapters can be covered in a typical lecture of 50 to 75 minutes. The style is informal for easier reading; it resembles an oral lecture rather than the formal writing you would encounter when reading research papers.

In our experience, the TEM books currently available fall into three major categories. They may be too theoretical for many materials science students; they attempt to cover all kinds of electron microscopy in one volume, which makes it difficult to include sufficient theory on any one technique; or they are limited in the TEM topics they cover. The rapid development of the TEM field has meant that many of the earlier books must automatically be placed in the third category. Although these books are often invaluable in teaching, we have not found them generally suitable as the course textbook in a senior-year undergraduate or first-year graduate course introducing TEM, so we have endeavored to fill this perceived gap.

Since this text is an introduction to the whole subject of TEM, we incorporate *all* aspects of a modern TEM into an integrated whole. So, rather than separating out the broad-beam and convergent-beam aspects of the subject (the traditional structural analysis or imaging versus the "chemical" analysis or "new" techniques), we treat these two aspects as different sides of the same coin. Thus scanning-beam (STEM) imaging is just another way to form an image in a TEM. There is no reason to regard "conventional" bright-field and "conventional" dark-field imaging as any more fundamental ways of imaging the specimen than annular dark-field imaging—or even secondary-electron or STEM *Z*-contrast modes. Similarly, convergent-beam and scanning-beam diffraction are integral parts of electron diffraction, and are complementary to selected-area diffraction. Inelastic electron scattering is the source of both Kikuchi lines and characteristic X-rays. So we don't deliberately split off "conventional" microscopy from "analytical" microscopy.

Our approach is to thread two fundamental questions throughout the text.

Why should we use a particular technique?

How do we put the idea into practice?

We attempt to establish a sound theoretical basis where necessary, although not always giving all the details. We then use this knowledge to build a solid understanding of how we use the instrument. The text is illustrated with examples from across the fields of materials science and engineering and, where possible, a sense of the history of the technique is introduced. We keep references to a minimum and generally accepted concepts are not specifically credited, although numerous classical general references are included.

We both have extensive teaching and research backgrounds in all aspects of TEM comprising diffraction, imaging, and microanalysis. Our research in TEM of materials spans metals, ceramics, composites, and semiconductors. We each bring more than 25 years of TEM experience to the book, and have contributed to the training of a generation of (we hope) skilled electron microscopists. We found that writing the book broadened our own knowledge considerably and was actually fun on some occasions. We hope you experience the same reactions after reading it.

Lastly, we encourage you to send us any comments (positive or negative). We can both be reached by email: dbwl @lehigh.edu and carter@cems.umn.edu

Acknowledgments

We have spent the best part of a decade in the conception and gestation of this text and such an endeavor can't be accomplished in isolation. Our first acknowledgments must be to our wives, Margie and Bryony, and our families, who have borne the brunt of our absences from home (and occasionally the brunt of our presence, also!).

We have both been fortunate to work with other microscopists, post-doctoral associates, and graduate students who have taught us much and contributed significantly to the examples in the text. We would like to thank a few of these colleagues directly: Dave Ackland, Ian Anderson, Charlie Betz, John Bruley, Dov Cohen, Ray Coles, Vinayak Dravid, Joe Goldstein, Brian Hebert, Jason Heffelfinger, John Hunt, Matt Johnson, Vicki Keast, Ron Liu, Charlie Lyman, Stuart McKernan, Joe Michael, Grant Norton, Sundar Ramamurthy, René Rasmussen, Kathy Repa, Al Romig, David A. Smith, Changmo Sung, Caroline Swanson, Ken Vecchio, and Mike Zemyan.

In addition, many other colleagues and friends in the fields of microscopy and microanalysis have helped with the book (even if they weren't aware of it). These include: Ron Anderson, Jim Bentley, Geoff Campbell, Graham Cliff, David Cockayne, the late Chuck Fiori, Peter Goodhew, Ron Gronsky, Peter Hawkes, David Joy, Roar Kilaas, Gordon Lorimer, Harald Müllejans, Dale Newbury, Mike O'Keefe, John Steeds, Peter Swann, Gareth Thomas, Patrick Veyssière, Nestor Zaluzec, and Elmar Zeitler. In addition, many other microscopists kindly provided the figures that we acknowledge individually in the list at the end of the book.

We have received financial support for our microscopy studies through several 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 acknowledge the National Science Foundation (Division of Materials Research) for almost 20 years of continuous funding, the National Aeronautics and Space Administration, the Department of Energy (Basic Energy Sciences), Sandia National Laboratories, and the Materials Research Center at Lehigh, which supports the microscopy laboratory. Portions of the text were written while DBW was on sabbatical or during extended visits to Chalmers University, Göteborg, with Gordon Dunlop and Hans Nordén; the Max Planck Institut für Metallforschung, Stuttgart, with Manfred Rühle; and Los Alamos National Laboratory, with Terry Mitchell. 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. This text was started while CBC was with the Department of Materials Science and Engineering at Cornell University.

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 classical texts by these renowned microscopists are referred to throughout this book. They influenced our own views of TEM tremendously, unavoidably contributing to any bias in our opinions, notation, and approach to the whole subject.

List of Acronyms

The field of TEM is a rich source of acronyms, behind which we hide both simple and esoteric concepts. While the generation of new 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
A/D analog to digital (converter)
ADF annular dark field
AEM analytical electron microscope/microscopy
AES Auger electron spectrometer/spectroscopy
AFF aberration-free focus
ALCHEMI atom location by channeling-enhanced microanalysis
ANL Argonne National Laboratory
APB anti-phase domain boundary
ASU Arizona State University
ATW atmospheric thin window

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

C(1, 2, etc.) condenser (1, 2, etc.) lens
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
CFEG cold field-emission gun
CL cathodoluminescence
CRT cathode-ray tube
CS crystallographic shear
CSL coincident-site lattice

DF dark field
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
EELS electron energy-loss spectrometry
EFI energy-filtered imaging
ELNES energy-loss near-edge structure
ELP energy-loss program (Gatan)
EMMA electron microscope microanalyzer
EMS electron microscopy image simulation
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

FCF fluorescence correction factor
FEG field-emission gun
FET field-effect transistor
FFT fast Fourier transform

FOLZ first-order Laue zone
FSE fast secondary electron
FTP file transfer protocol
FWHM full width at half maximum
FWTM full width at tenth maximum

GB grain boundary
GCS generalized cross section
GIF Gatan image filter
GOS generalized oscillator strength

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

IDB inversion domain boundary
IEEE International Electronics and Electrical Engineering
IG intrinsic Ge
IVEM intermediate voltage electron microscope/microscopy

K–M Kossel–Möllenstedt

LEED low-energy electron diffraction
LLS linear least-squares
LUT look-up table

MC minimum contrast
MCA multichannel analyzer
MDM minimum detectable mass
MLS multiple least-squares
MMF minimum mass fraction
MSDS material safety data sheets

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

OR orientation relationship
OTEDP oblique-textured electron diffraction pattern

PB phase boundary
P/B peak-to-background ratio
PEELS parallel electron energy-loss spectrometer/spectrometry
PIMS Precision Ion-Milling System™
PIPS Precision Ion-Polishing System™
PM photomultiplier
POA phase-object approximation

QHRTEM quantitative high-resolution transmission electron microscopy

RB translation boundary (yes, it does!)
RCP rocking-beam channeling patterns
RDF radial distribution function
REM reflection electron microscope/microscopy
RHEED reflection high-energy electron diffraction
RHF relativistic Hartree–Fock
RHFS relativistic Hartree–Fock–Slater

SAD selected-area diffraction
SE secondary electron
SEELS serial electron energy-loss spectrometer/spectrometry
SEM scanning electron microscope/microscopy
SF stacking fault
SHRLI simulated high-resolution lattice images
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
TMBA too many bloody acronyms

UHV ultrahigh vacuum
UTW ultrathin window

V/F voltage to frequency (converter)
VLM visible-light microscope/microscopy

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
XRD X-ray diffraction

YBCO yttrium–barium–copper oxide
YAG yttrium–aluminum garnet

ZAF atomic number, absorption, fluorescence (correction)
ZAP zone-axis pattern
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 relative transition probability
a_0 Bohr radius
a, b, c lattice vectors
a*, b*, c* reciprocal lattice vectors
A amplitude of scattered beam
A amperes
A absorption correction factor
A active area of the detector
A Richardson's constant
A atomic weight
Å Ångstrom
$\mathcal{A}$ Bloch wave amplitude
$A(\mathbf{U})$ aperture function

b beam-broadening parameter
$\mathbf{b}_e$ 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 velocity of light
C contrast
C composition
C_a astigmatism aberration coefficient
C_c chromatic aberration coefficient
C_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

d beam (probe) diameter
d diameter of spectrometer entrance aperture
d spacing of moiré fringes
d interplanar spacing
d_c effective source size
d_d diffraction-limited beam diameter
d_g Gaussian beam diameter
$d_{hk\ell}$ $hk\ell$ interplanar spacing
d_{im} smallest resolvable image distance
d_{ob} smallest resolvable object distance
d_s spherical-aberration limited beam diameter
d_{eff} effective entrance aperture diameter at recording plane
dz thickness of a diffracting slice
$d\sigma/d\Omega$ differential cross section of one atom
D change in focus
D distance from projector crossover to recording plane
D electron dose
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
E Young's modulus
E_a spatial coherence envelope
E_c chromatic aberration envelope
E_c critical ionization energy
E_d displacement energy
E_K ionization energy for K-shell electron
E_L ionization energy for L-shell electron
$\mathcal{E}$ total energy
$\mathcal{E}$ energy loss
$\mathcal{E}_m$ average energy loss
$\mathcal{E}_P$ plasmon energy loss
E_P plasmon 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/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_B fraction of B alloying element
F_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

g diffraction vector (magnitude of **K** at the Bragg angle)
$g(\mathbf{r})$ 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(**r**)

h Planck's constant
h distance from specimen to the aperture
$h(\mathrm{r})$ contrast transfer function
$(hk\ell)$ Miller indices of a crystal plane
$hk\ell$ indices of diffraction spots from $hk\ell$ plane
H spacing of the reciprocal-lattice planes parallel to the electron beam
$H(\mathbf{u})$ Fourier transform of $h(\mathbf{r})$

I intensity
I intrinsic line width of the detector
i_e emission current
i_f filament heating current
I_g intensity in the diffracted beam
I_K K-shell intensity above background
$I(\mathbf{k})$ kinematical intensity
I_P intensity in the first plasmon peak
I_T total transmitted intensity
I_0 intensity in the zero-loss peak
I_0 intensity in the direct beam
$I(\ell)$ low-loss spectrum intensity

J current density

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
K bulk modulus
K Kelvin
K Kramers constant
K sensitivity factor
K inner core shell/characteristic X-ray line/ionization edge
K change in **k** due to diffraction
$\mathbf{K}_B$ magnitude of **K** at the Bragg angle
K_o kernel

L camera length

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 nano
n vector normal to the surface
n_s number of electrons in the ionized subshell
N noise
N number of counts
N number of atoms per unit area
N $h + k + \ell$

$N(E)$ number of bremsstrahlung photons of energy E
N_0 Avogadro's number

O direct beam

p pico
p momentum
p integer
P peak intensity
P FWHM of a randomized electronic pulse generator
P_K probability of K-shell ionization
$P(z)$ scattering matrix for a slice of thickness z

Q number of scattering events per unit distance
Q cross section

r radius
r distance a wave propagates
r power term to fit background in EELS spectrum
r_M image translation distance
$\mathbf{r}_n$ lattice vector
$\mathbf{r}^*$ reciprocal lattice vector
r_{ast} astigmatism disk radius
r_{chr} chromatic-aberration disk radius
r_{sph} spherical-aberration disk radius
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
$\mathbf{R}$ crystal lattice vector
R count rate
R resolution of XEDS detector
R distance on screen between diffraction spots
$\mathbf{R}_n$ lattice displacement vector

s excitation error or deviation parameter
$\mathbf{s}_R$ excitation error due to defect
$\mathbf{s}_z$ ($\mathbf{s}_g$) excitation error in the z direction
s_{eff} effective excitation error
S distance from the specimen to detector
S signal
S standard deviation for n measurements
sr steradians

t shift vector between the ZOLZ and the HOLZ
t' absorption path length
$T(\mathbf{u})$ objective-lens transfer function
$T_{eff}(\mathbf{u})$ effective transfer function

u object distance
u unit vector along the dislocation line
u* vector normal to the ZOLZ
u_k displacement field
U overvoltage
U_g Fourier component of the perfect-crystal potential

v image distance
v velocity of an electron
V accelerating voltage
$\mathcal{V}$ potential energy
V_c the volume of the unit cell
V_c inner potential of cavity
V_t projected potential through the thickness of the specimen
$V(\mathbf{r})$ crystal inner potential

T absolute temperature
T Tesla
T_C period of rotation

(UVW) indices of a crystal direction
UVW indices of beam direction

w $s\xi_g$ (excitation error × extinction distance)

× times
x distance
× times (magnification)
x, y, z atom coordinates
X FWHM due to detector

y parallax shift in the image
y displacement at the specimen

z specimen height (distance along the optic axis)
Z atomic number/atomic number correction factor

Greek symbols

α phase shift due to defect
α semiangle of incidence/convergence
α X-ray take-off angle
α_{opt} optimum convergence semiangle

β brightness
β ratio of electron velocity to light velocity
β semiangle of collection
β_{opt} optimum collection semiangle

γ degree of spatial coherence
γ phase of direct beam

Δ change/difference
Δ width of energy window
$\Delta\phi$ phase difference
$\Delta\theta_i$ angles between Kossel–Möllenstedt fringes
Δ_{AB} difference in mass-absorption coefficients
ΔE energy spread
$\Delta\mathcal{E}_P$ plasmon line width
Δf maximum difference in focus
Δf_{AFF} aberration-free (de)focus
Δf_{MC} minimum contrast defocus
Δf_{opt} optimum defocus

Δf_{sch} Scherzer defocus
Δh relative depth in specimen
ΔI change in intensity
Δp parallax shift
ΔV change in the inner potential
Δx path difference
Δx half-width of image of undissociated screw dislocation
Δx_{res} resolution at Scherzer defocus
Δz change in height
δ angle between detector normal and line from detector to specimen
δ diameter of disk image
δ diffuseness of interface
δ fluorescence enhancement ratio
δ precipitate/matrix misfit

ε angle of deflection
ε detector efficiency
ε energy to create an electron–hole pair
ε strain
ε_0 permittivity of free space (dielectric constant)

$\eta(\theta)$ phase of the atomic scattering factor
η phase change
ν Poisson's ratio

$\mathbf{\Phi}$ work function
$\mathbf{\Phi}_A^{\Delta\rho t}$ X-ray emission from element A in an isolated thin film
ϕ angle between Kikuchi line and diffraction spot
ϕ angle between two Kikuchi line pairs
ϕ angle between two planes
ϕ angle of tilt between stereo images
ϕ phase of a wave
ϕ^* complex conjugate of ϕ
ϕ_g amplitude of the diffracted beam
ϕ_0 amplitude of the direct beam
ϕ_x angle of deflection of the beam
$\mathbf{\phi}(\rho t)$ depth distribution of X-ray production

χ wave vector outside the specimen
χ_G wave vector which terminates on the point G in reciprocal space
χ_O wave vector which terminates on the point O in reciprocal space
$\chi(\mathbf{u})$ phase-distortion function

κ thermal conductivity

ξ_g extinction distance for diffracted beam
ξ_g' absorption parameter
ξ_0 extinction distance for direct beam
ξ_g^{abs} absorption-modified ξ_g

λ mean-free path
λ wavelength
λ_c coherence length
λ_P plasmon mean free path
λ^{-1} radius of Ewald sphere

μ micro
μ refractive index
μ/ρ mass absorption coefficient
$\mu^{(j)}(\mathbf{r})$ Bloch function

ν frequency

ψ amplitude of a wave
ψ^T total wave function
ψ^{tot} total wave function

ρ angle between directions
ρ density
ρ_c information limit due to chromatic aberration
$\rho(\mathbf{r})$ 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
σ_K ionization cross section for K-shell electron
σ_T total scattering cross section
$\sigma_K(\beta\Delta)$ partial ionization cross section

θ scattering semiangle
θ_B Bragg angle
θ_C cut-off semiangle
θ_E characteristic scattering semiangle
θ_0 screening parameter

τ detector time constant
τ dwell time

ω fluorescence yield
ω_c cyclotron frequency
ω_p plasmon frequency
Ω solid angle of collection

$\otimes$ convolution (multiply and integrate)

Contents

IV ■ Spectrometry

32 ■ X-ray Spectrometry

33 ■ The XEDS–TEM Interface

34 ■ Qualitative X-ray Analysis

35 ■ Quantitative X-ray Microanalysis

36 ■ Spatial Resolution and Minimum Detectability

37 ■ Electron Energy-Loss Spectrometers

Transmission Electron Microscopy

A Textbook for Materials Science

Basics

The Transmission Electron Microscope

1

CHAPTER PREVIEW

A typical commercial transmission electron microscope (TEM) costs about $2 for each electron volt of energy in the beam, and if you add on all the options, it can cost about $4–5 per eV. As you'll see, we use beam energies in the range from 100,000–400,000 eV, so a TEM becomes 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 TEM appeared as the instrument developed. Unfortunately, coupled with the advantages are some serious drawbacks, which limit the microscope performance, 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 acronym to denote both the technique (microscopy) and the instrument (microscope). We regard all of the dif-

ferent types of TEM as simply variations on a basic theme and that is why only "TEM" is in the book title. We will 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. The reason for this is 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, and how electrons interact with materials.

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 role of the computer in today's TEM analysis.

The Transmission Electron Microscope

1

1.1. 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.

1.1.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 Davisson and Germer (1927) and Thompson and Reid (1927) 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, in 1986. 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. TEMs were developed by commercial companies only four years later. The Metropolitan-Vickers EM1 was the first commercial TEM. It was built in the UK in 1936, but apparently it didn't work very well and regular production 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 late 1940s when Heidenreich (1949) first thinned metal foils to electron transparency. This work was followed up by Bollman in Switzerland and Hirsch and co-workers in Cambridge. Because so much of the early TEM work examined metal specimens, the word "foil" has come to be synonymous with "specimen." 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 theoretical 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 (Thomas 1962). Other materials-oriented texts followed, e.g., Edington (1976) and Thomas and Goringe (1979).

Today, TEMs constitute arguably the most efficient and versatile tools for the characterization of materials. If you want to read a history of the TEM, the book by Marton (1968) is a compact, personal monograph and that edited by Hawkes (1985) contains a series of individual reminiscences. Fujita (1986) emphasizes the contribution of Japan 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, and Hashimoto 1986) 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.)

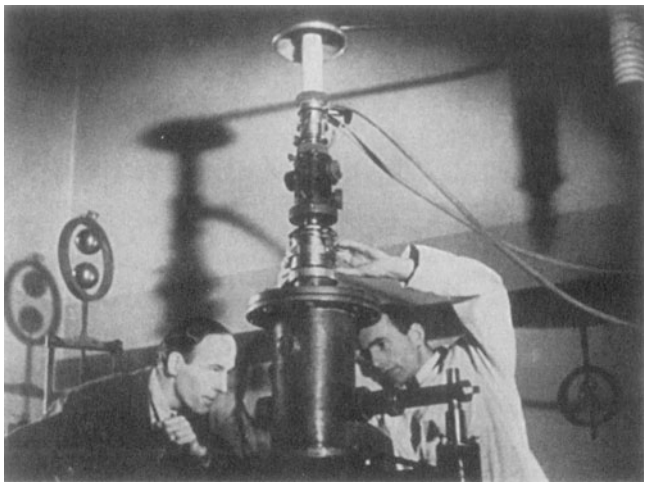


Figure 1.1. The electron microscope built by Ruska and Knoll in Berlin in the early 1930s.

1.1.B. Microscopy and the Concept of Resolution

When asked what a "microscope" is, 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. Because of the general familiarity with the concept of the light microscope, we will draw analogies between electron and visible-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 to see by. This distance is the *resolution* or *resolving power* of our eyes. So any instrument that can show us pictures (or "images" as we'll 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 would 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 the retinal cells, but you wouldn't see anything! So an integral part of any electron microscope is a viewing screen of some form, which translates electron intensity to light intensity, and which we observe or record photographically. We'll discuss these screens and other ways of recording electron images in Chapter 7.

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 classical Rayleigh criterion for light microscopy, which states that the smallest distance that can be resolved, δ, is given approximately by

$$\delta = \frac{0.61\lambda}{\mu \sin \beta} \qquad [1.1]$$

In equation 1.1, λ is the wavelength of the radiation, μ the refractive index of the viewing medium, and β is 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 (5500 Å), and so the resolution of a good light microscope is about 300 nm. In TEMs we can approximate the resolution in equation 1.1 to $0.61\lambda/\beta$ which, as we'll see later, is very small.

> We'll try to use nanometers throughout this book, but you'll find that many microscopists still insist on using Ångstroms rather than the SI units. However, the Ångstrom is close to the atomic diameter and so is a more convenient unit because it saves us using convoluted phrases like "three tenths of a nanometer."

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 light microscope. So there's a real need to image detail down to the atomic level if we want to understand 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 this 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.) Now 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

$$\lambda \sim \frac{1.22}{E^{1/2}} \qquad [1.2]$$

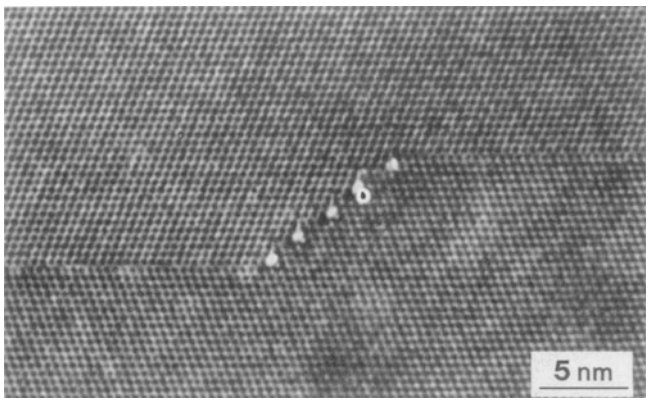


Figure 1.2. A twin boundary in spinel stepping from one {111} plane to another parallel plane. The white dots are columns of atoms. The change in atomic orientation across the twin boundary can be readily seen, even if we do not know what causes the white dots or why, indeed, they are white.

In this equation E is in electron volts (eV) and λ in nm. Remember that we should be precise in our use of the units V and eV: the former represents the *accelerating voltage* of the microscope while the latter refers to the *energy* of the electrons in the microscope. So for a 100-keV electron, we find that $\lambda \sim 4$ pm (0.004 nm), which is much smaller than the diameter of an atom.

We'll see later that we are nowhere near building TEMs that approach this wavelength limit of resolution, because we can't make perfect electron lenses (see Chapter 6). But 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 1.2. The advantages of shorter wavelengths led in the 1960s to the development of high voltage electron microscopes (HVEMs), with accelerating potentials between 1 MV and 3 MV. In fact, most of these instruments were used to introduce controlled amounts of radiation damage into specimens in an attempt to simulate nuclear reactor environments, but changes in the emphasis of energy research mean there is not much call for such instruments today. While we can still improve the resolution by incremental amounts, the drive for much better resolution is now no longer paramount and the TEM is developing in other ways. In fact, 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 300 or 400 kV, but still offer very high resolution, close to that achieved at 1 MV.

1.1.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 one of the tightly bound inner-shell electrons from the attractive field of the nucleus.

One of the advantages to 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," or AEM, giving us chemical information and a lot of other detail about our samples. AEM uses X-ray energy dispersive spectrometry (XEDS) and electron energy-loss spectrometry (EELS). For example, Figure 1.4A is an X-ray spectrum from a very small region of a TEM specimen showing characteristic peaks which identify the elements present. We can transform such spectra into quantitative data describing elemental changes associated with inhomogeneous microstructures as also shown in Figures 1.4B and C. This aspect comprises Part IV of the book. In contrast, microscopes using nonionizing radiation such as visible light usually only generate light (but not much heat, which is good). AEMs generally offer improved performance at intermediate voltages, similar to HRTEMs.

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, so modern TEMs are very good

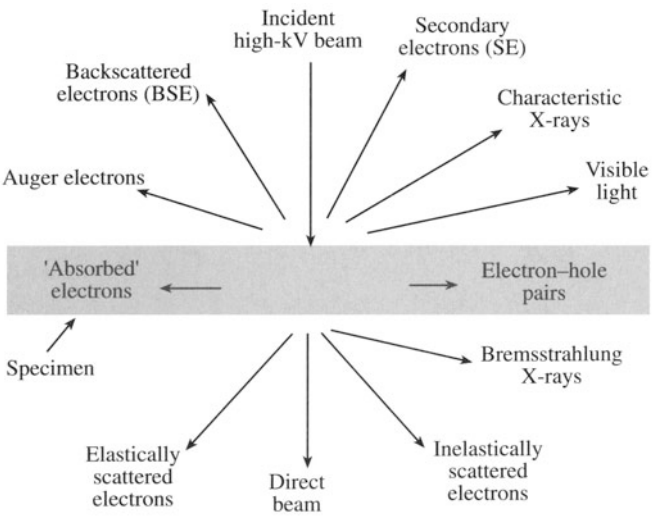
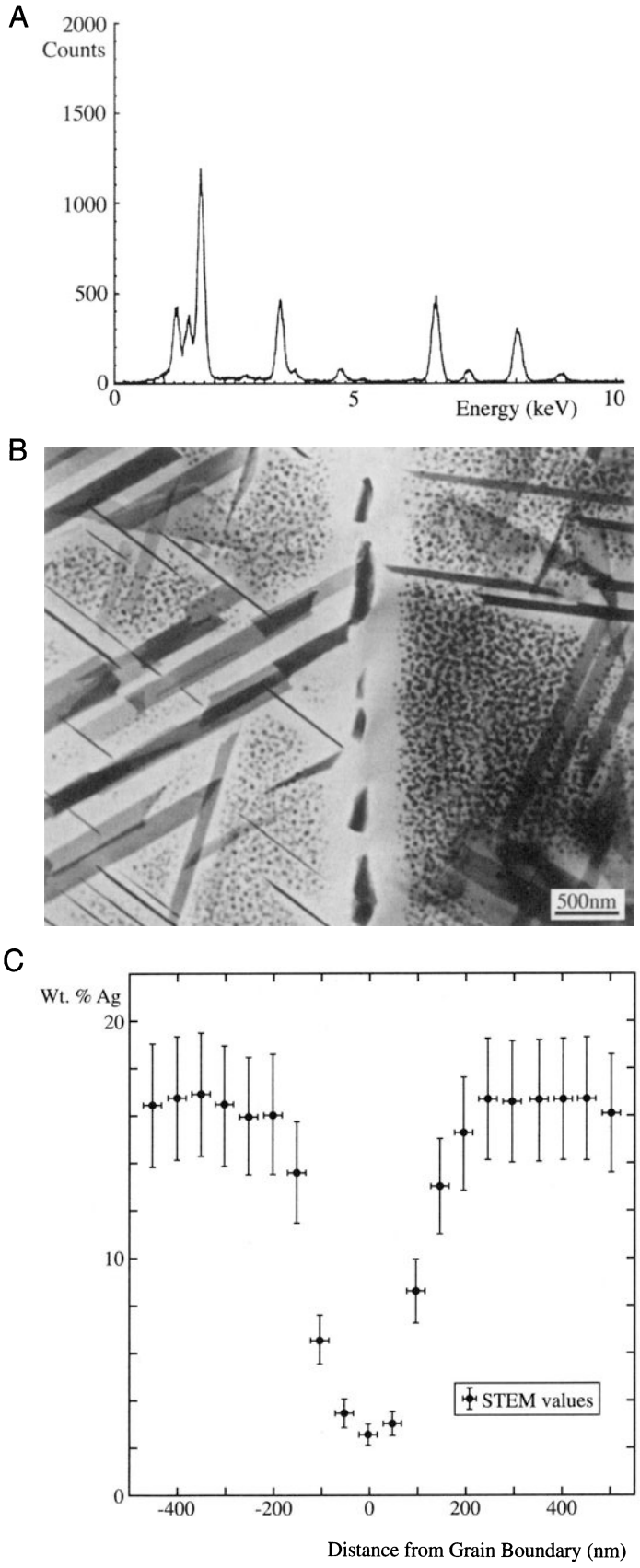


Figure 1.3. Signals generated when a high-energy beam of electrons interacts with a thin specimen. Most of these signals can be detected in different types of TEM. The directions shown for each signal do not always represent the physical direction of the signal but indicate, in a relative manner, where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



signal-generating instruments. To localize these signals we need to get our TEM to form a very fine electron beam, typically <10 nm and at best <1 nm in diameter. We accomplish this by combining 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 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.

1.1.D. Depth of Field

The depth of field of a microscope is a measure of how much of the *object* we are looking at remains "in focus" at the same time. Like the resolution, this property is governed by the lenses in the microscope. The best electron lens is not a very good one, as we've already mentioned, and has been compared to using the bottom of a Coca-Cola bottle as a lens for light microscopy. To minimize this problem we have to use very small limiting apertures in the lenses, narrowing the beam down to a thin "pencil" of electrons which at most is a few micrometers across. These apertures cut down the intensity of the electron beam, but also act to increase the depth of focus of the images that we produce. Remember that "depth of field" refers to the specimen while "depth of focus" refers to the image.

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, all of the specimen is usually in focus at the same time, independent of the specimen topography, as long as it's electron transparent! Figure 1.5 shows a TEM image of some dislocations in a crystal. 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, and they remain in sharp focus at all times. Furthermore, we can record the final image at different positions below the final lens of the instrument and it will still be in focus. Compare this with

Figure 1.4. (A) An X-ray spectrum from a small biotite crystal showing peaks at energies that are characteristic of the elements present in the region that interacts with the electron beam. The major peaks from left to right are for Mg, Al, Si, K, Fe, and the Cu support grid. (B) A TEM image of a precipitate-free zone (PFZ) in an aged Al–16 wt% Ag alloy. (C) The Ag profile across the PFZ in (B), obtained through X-ray spectrometry in the TEM showing the depletion of Ag responsible for the PFZ formation.

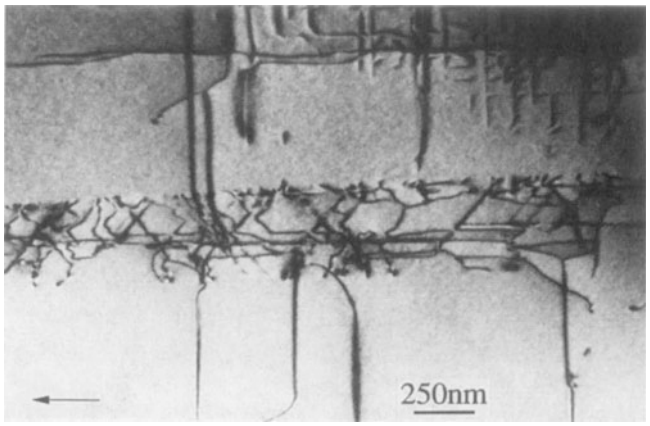


Figure 1.5. TEM image of dislocations in GaAs. A band of dislocations threads through the thin specimen from the top to the bottom but remains in focus through the foil thickness.

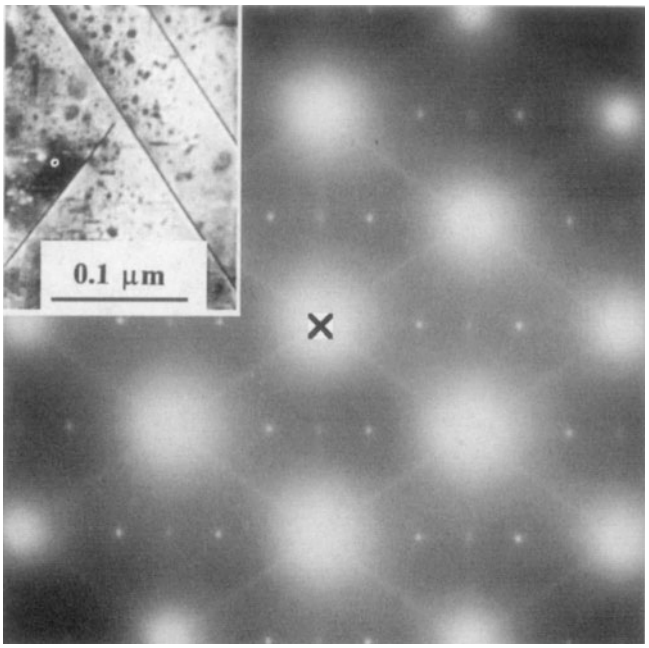


Figure 1.6. TEM diffraction pattern from a thin foil of Al-Li-Cu containing various precipitate phases, shown in the inset image. The central spot (X) contains electrons that come directly through the foil and the other spots and lines are diffracted electrons which are scattered from different crystal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

the visible-light microscope where, as you probably know, unless the surface of the specimen is flat to 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 visible-light microscope.

1.1.E. Diffraction

Thompson and Reid showed that electrons could be diffracted when passing through thin crystals of nickel, and the possibility of combining electron diffraction into TEMs was realized by Kossel and Möllenstedt (1939). Today, electron diffraction is an indispensable part of TEM and is arguably the most useful aspect of TEM for materials scientists. Figure 1.6 shows a TEM diffraction pattern which 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 II that, in addition to the things we just listed, you can conduct a complete crystallographic symmetry analysis of minuscule crystals, including such esoteric aspects as point-group and space-group determination, and at all times the crystallography can be related to the image of your specimen. There is no similar capability on a light microscope because of the relatively large wavelength of visible light.

So an electron microscope can produce atomic level images, can generate a variety of signals telling you about your sample 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.

1.2. LIMITATIONS OF THE TEM

1.2.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 in around 1970 estimating that all TEMs, since

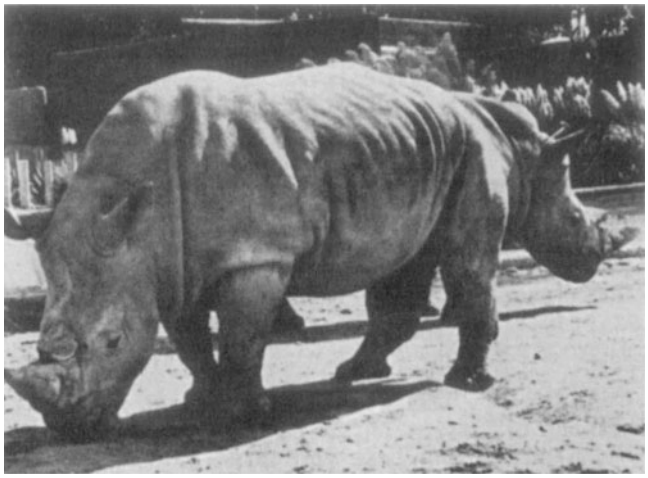


Figure 1.7. Photograph of two rhinos taken so that, in projection, they appear as one two-headed beast. Such projection artifacts in reflected-light images are easily discernible to the human eye but similar artifacts in TEM images are easily mistaken for "real" features.

they first became available commercially, had only examined 0.3 mm^3 of material! Extending that calculation to the present time at best doubles the volume to 0.6 mm^3. So we have an instrument that is a terrible sampling tool. This only serves to emphasize that 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 visible-light microscope, and the scanning electron microscope. In other words, know the forest before you start looking at the leaves on the trees.

1.2.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 (1980) illustrates this problem well by showing a picture of two rhinos, 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 is that, generally, all the TEM information that we talk about in this book (images, diffraction patterns, spectra) is *averaged through the thickness of the specimen.* In other words, a single TEM image has no depth sensitivity, as is apparent from Figure 1.5. So other techniques which are surface-sensitive or depth-sensitive, such as field ion microscopy, scanning-probe microscopy, Auger spectroscopy, Rutherford backscattering, etc., are necessary complementary techniques if you want a full characterization of your specimen.

1.2.C. Electron Beam Damage and Safety

A side effect of ionizing radiation is that it can damage your specimen, particularly in materials such as polymers and some ceramics. Some aspects of beam damage are exacerbated at higher voltages and, with commercial instruments offering up to 400 kV, beam damage now limits much of what we can do in the TEM, even with refractory metals. Figure 1.8 shows an area of a specimen damaged by high-energy electrons. The combination of high-kV beams with the intense electron sources that are available means that we can destroy almost any specimen, if we 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 could kill you. 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!

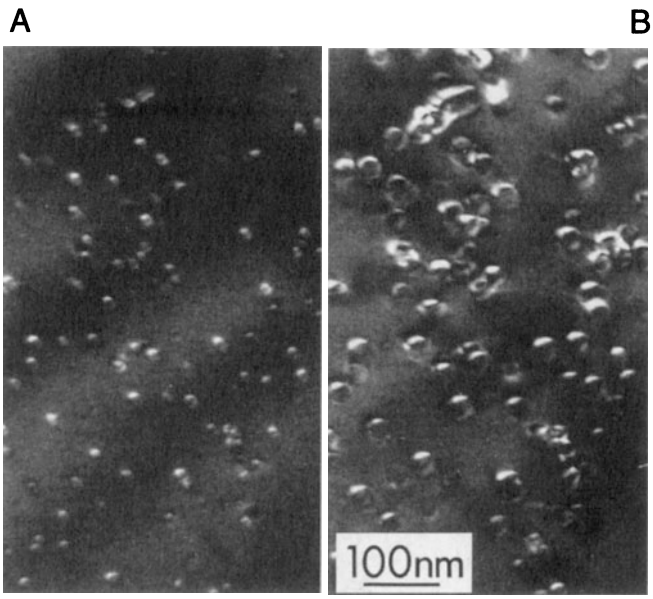


Figure 1.8. Beam damage in quartz after bombardment with 125-keV electrons. With increasing time, from (A) to (B), the damaged regions increase in size.

1.2.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 or photographic film to give us an interpretable image in a reasonable time. Generally this requirement is a function of the electron energy and the average atomic number of the specimen. Typically for 100-keV electrons, specimens of aluminum alloys almost up to 1 μm would be thin, while steel would be thin up to about several hundred nm. However, it is an axiom in TEM that thinner is better, and specimens below 100 nm should be used wherever possible, and in extreme cases, such as when doing HRTEM or electron spectrometry, specimen thicknesses <50 nm are essential. These demands become less strict as the beam voltage increases, but this is offset by the danger of beam damage.

The requirement for thin specimens is a *major* limitation of the TEM. 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 specimen, changing both its structure and its chemistry. So you need to be aware of the dangers of specimen preparation and learn to recognize the artifacts introduced by standard preparation methods.

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 visible-light microscopy and scanning electron microscopy (SEM) before venturing into TEM studies. Otherwise you may fall foul of some of the limitations we have just listed.

1.3. 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 300 or 400 keV TEM can combine aspects of *all* the above microscope types. Figure 1.9 shows four 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-story room; the scale of each instrument can be judged from the common height of the operator's console. A modern machine essentially is an electron optic 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 CRT displays. There is no viewing screen. Furthermore, the electron source is at the base of the column rather than at the top, as we will assume in all of our discussions.

1.4. 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, which 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 G. I. Taylor's famous experiment in which he demonstrated Young's slits interference patterns despite using such a weak light source that only one photon passed through the apparatus at any one time. A typical electron beam current in a TEM is about 0.1–1 μA, 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 $0.5c$ (1.6×10^8 m/s), so they are separated by 0.16 cm 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 imply interaction between the different electron beams. 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.

There are a few important equations which 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

$$\lambda = \frac{h}{p} \qquad [1.3]$$

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, so

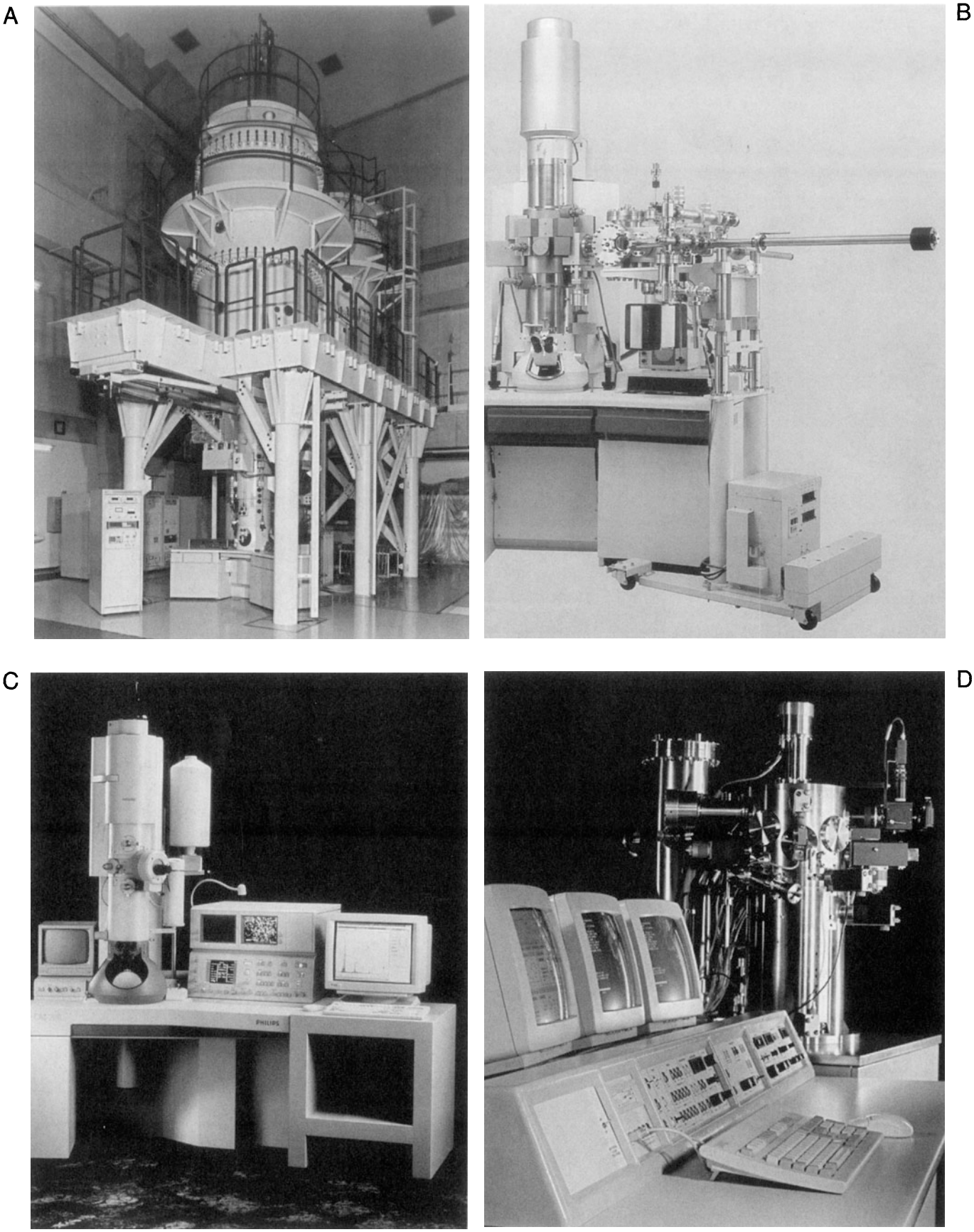


Figure 1.9. Different TEMs: (A) a JEOL 1.25-MV high voltage microscope, used for high-resolution imaging; (B) a Hitachi specialized ultrahigh vacuum TEM for high-resolution surface imaging; (C) a Philips 200-kV analytical microscope with an X-ray spectrometer attached to the stage (the liquid-N_2 dewar cools the detector); and (D) a VG dedicated 100-kV ultrahigh vacuum scanning transmission microscope. Comparison with Ruska's instrument (Figure 1.1) which is 50–60 years older is instructive.

Table 1.1. Fundamental Constants and Definitions

Charge (e)	$(-)\ 1.602 \times 10^{-19}$ 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)
Planck's constant (h)	6.626×10^{-34} N m s
1 ampere	1 C/sec
Speed of light in vacuum (c)	2.998×10^{8} m/sec

$$eV = \frac{m_0 v^2}{2} \qquad [1.4]$$

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 we obtain

$$p = m_0 v = \left(2m_0 eV\right)^{1/2} \qquad [1.5]$$

What all this leads to is the relationship between the electron wavelength λ and the accelerating voltage of the electron microscope, V

$$\lambda = \frac{h}{\left(2m_0 eV\right)^{1/2}} \qquad [1.6]$$

This expression is identical to equation 1.2. This relationship between λ and the accelerating voltage introduces a very important concept: by increasing the accelerating voltage we decrease the wavelength of the electrons.

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 nonrelativistic 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 100-keV energies and above because the velocity of the electrons (as particles) becomes greater than half the speed of light! (The speed of light in vacuum is 2.998×10^8 m/s.) So to be exact we must modify equation 1.6 to give

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

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 greatest for higher accelerating voltages, as shown in Table 1.2.

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).

1.5. MICROSCOPY ON THE INTERNET/WORLD WIDE WEB

TEM users are well integrated into the Internet/WWW and this is a source of useful information (and also some useful knowledge!) about what's going on in the field. Already

Table 1.2. Electron Properties as a Function of Accelerating Voltage

Accelerating voltage (kV)	Nonrelativistic wavelength (nm)	Relativistic wavelength (nm)	Mass (× m_0)	Velocity (×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

you can view research TEMs in real time on the Internet and in due course you'll not only see other instruments but be able to operate them remotely. Such "telepresence microscopy" will represent an extraordinary leap in our ability to characterize materials, since advanced instruments will effectively be available to you in your own laboratories without the need to travel to special sites.

In addition, specialized software packages that allow you to carry out many of the advanced analyses that we will introduce in this text (e.g., diffraction pattern analysis and image/diffraction/spectral simulation) are also available through the Web. In many cases access to this software is limited and any serious microscopy operation should have the software on site, but sometimes it is useful to see the possibilities before you purchase. A list of useful sites is included below but, as with all aspects of the Web, this list is already out of date and the number of actual sites is growing daily.

1.5.A. Microscopy and Microanalysis-Related WWW Sites

http://www.amc.anl.gov

This is the best source for TEM information on the Web in the United States. It is run by N.J. 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 access to Software/Image Libraries. Other useful connections through this site include

http://146.139.72.10/Docs/nonanl/Meetings.html

List of Meetings/Conferences on Microscopy/Microanalysis

http://146.139.72.10/Docs/nonanl/ShortCourses.html

List of Short Courses/Workshops on Microscopy/Microanalysis

http://146.139.72.10/Docs/nonanl/msa/MSA.html

Microscopy Society of America information

http://146.139.72.10/Docs/nonanl/aust/aust.html

Australian Microscopy Societies information

http://146.139.72.10/Docs/nonanl//msc/MSc.html

Microscopical Society of Canada information

http://146.139.72.10/Docs/nonanl/rms/RMS.html

Royal Microscopical Society information

http://146.139.72.10/Docs/nonanl/mas/MAS.html

Microbeam Analysis Society information

http://146.139.72.10/Docs/NonAnl/EduSites.html

University/Educational Sites

http://www.amc.anl.gov/Docs/NonAnl/GovSites.html

Governmental Microscopy Sites

http://146.139.72.10/Docs/NonAnl/ComSites.html

Commercial Sites—microscopy-related manufacturers/suppliers

http://cimewww.epfl.ch/Welcometext.html

A similar operation to the ANL site, but based at the Ecole Polytechnique Fédérale de Lausanne in Switzerland, run by P.-H. Jouneau and P. Stadelmann. A very useful array of software is available including Electron Microscopy Image Simulation (EMS) software, which allows you to perform the following tasks:

> Draw a crystal in perspective view or in projection, stereographic projections, list $(hk\ell)$ distances, structure factors, and extinction distances, draw the microscope contrast transfer function, with $(hk\ell)$ crystal planes, draw kinematical, dynamical, or powder diffraction patterns, draw the amplitude and phase of diffracted beams as a function of specimen thickness, do auto indexing of diffraction patterns, draw Kikuchi patterns, draw high-order Laue zone line patterns, draw convergent beam diffraction patterns (Bloch-wave calculation), draw HRTEM image maps of the crystal (Bloch-wave calculation), do conventional image calculation of dislocations for a cubic crystal

http://cimewww.epfl.ch/emyp/

Another operation based at the Ecole Polytechnique Fédérale de Lausanne in Switzerland, also run by Jouneau and Stadelmann; EM Yellow pages. Contents include:

> Software for Electron Microscopy, Professional Societies, Instruments, Equipment and Consulting, Education in Electron Microscopy, Data and Databases, News and Publications, Related Sources of Information, Conferences, Workshops and Schools, Getting Somewhere Else on the Web

http://www-personal.engin.umich.edu/~jfmjfm/newsgroup.html

Microscopy users newsgroup run by J.F. Mansfield at the University of Michigan

http://www.bocklabs.wisc.edu/imr/microscopists.html

Directory of microscopists on the net

1.5.B. Microscopy and Microanalysis Software

While there is a lot of software available on the WWW, much of it freeware or shareware, you sometimes get what you pay for, so as a serious microscopist you should have access to the best commercial programs, which are not free. Again, 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 microanalysis. 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.

- Comis runs on a UNIX system. A version of the program is being implemented on a Macintosh computer. The graphics are currently supported for Tektronix 4010 and 4100 compatible terminals, but the program may be run from any terminal (without the menu-interface). A version based on the X-window system is also being developed. The interface to the framestore is based on a small set of routines, similar to those used by the SEMPER image processing software (see below). Output to laser printers is supported through the PostScript language. MaComis is available from ComiSoft at 70404.1710@CompuServe.com. Contact the authors for information on the UNIX version
- CRISP is a commercial package running under Windows on a PC. It is designed for image processing 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)
- Desktop Microscopist: software for the Macintosh which allows you to calculate diffraction patterns. Available from Virtual Laboratories at http://www. Rt66.com/~virtlabs/ (505 828 1640)
- Differential Hysteresis Imaging: software that enables you to extract the full range of contrast information out of any digitized TEM (or SEM) image. From Klaus Peters at the University of Connecticut, on the Web at http://panda.uchc.edu/htklaus/index.html
- Digital Micrograph: a complete system for the acquisition, control, and processing of digital images from any electron microscope. In principle it can convert an old analog instrument to a digital one, if beam scan coils are available, but it's not worth doing this for a TEM. Use with a CCD camera that provides digital images from the TEM or interface to any STEM system. From Gatan Inc., 6678 Owens Drive, Pleasanton, CA 94588 (510 463 0200)
- DTSA: (Desk-Top Spectrum Analyzer) simulates energy (and wavelength) dispersive X-ray spectra; can also be used as a multichannel analyzer to acquire, interpret, and process spectra. Absolutely essential for the X-ray microanalyst. From NIST, Standard Reference Data Program, 221/A123 Gaithersburg, MD 20899 (301 975 2208)
- ELD is a commercial package from the producers of CRISP running under Windows on a PC. It is intended for quantitative analysis of diffraction patterns and is available from Calidris, Manhemsvägen 4, S-191 46 Solltuna, Sweden (46 8 625 00 41)
- ELP: the energy loss program that runs all Gatan EELS systems. You can't do EELS without it. From Gatan Inc., 6678 Owens Drive, Pleasanton, CA 94588 (510 463 0200)
- EMS: image simulation program, diffraction analysis, basic crystallographic data, and much more developed by P. Stadelmann at Ecole Polytechnique Fédérale de Lausanne and available from him through the Web (http://cimewww.epfl.ch/Welcometext.html). Some would argue that if you have this software, there is little else you need, which is evident from the listing of its capabilities in Section 1.5.A
- Head *et al.* (1973): The book includes a listing of the original source code for the simulation of diffraction contrast images. You can download the source code via the MSA Web page
- Maclispix: a Macintosh-based image processing program, which works in conjunction with NIH-Image (see below). Developed by David Bright at NIST and can be downloaded (free) from the WWW at http://www-sims.nist.gov/WWW/Internet/InternetResources. The FTP command is ftp://enh.nist. gov/mac/mx30a.bin
- MacTempas and CrystalKit: a Macintosh-based image analysis program from Roar Kilaas for the simulation of high-resolution images, diffraction patterns, and crystal structures. Total Resolution, 20 Florida Avenue, Berkeley, CA 94707. roar@totalresolution.com
- Monte Carlo Simulations: software to simulate electron beam trajectories through materials for estimating the spatial resolution of X-ray microanalysis or the backscattered electron yield. Available from David Joy at the University of Tennessee. A full listing is given in his textbook (Joy 1995)
- NCEMSS: (NCEM Simulation System) for HRTEM image simulations, symmetry operators for all 230 space groups, scattering factors for 98 elements. From the National Center for Electron Microscopy, Lawrence Berkeley Laboratories, University of California, Berkeley, CA 94729, or on the Web at http://ncem.lbl.gov/ncem.html
- NIH-Image: public domain software from NIH, developed by Wayne Rasband, for general image manipulation with a limited set of image processing tools. It is useful for grayscale enhancing and Fourier filtering. It can acquire, display, edit, enhance, analyze, print, and animate images. It

reads and writes TIFF, PICT, PICS, and MacPaint files, including programs for scanning, processing, editing, publishing, and analyzing images. It supports many standard image processing functions, including contrast enhancement, density profiling, smoothing, sharpening, edge detection, median filtering, and spatial convolution with user-defined kernels up to 63×63. Available from the Internet by anonymous ftp from zippy.nimh.nih.gov or on floppy disk from NTIS, 5285 Port Royal Rd., Springfield, VA 22161, part number PB93-504868. Details can be found on the WWW at http://rsb.info.nih.gov/nih-image/

- (Adobe) Photoshop and page layout programs for presentations and labeling your figures
- SEMPER (Synoptics, Ltd.): a general-purpose image processing package. It includes many operations suitable for processing of electron microscope images, and has been extended locally. It can be used for Fourier-space filtering of experimental images (including the background-subtraction method for small particles supported on amorphous substrates), for cross-correlation and lattice averaging, and for spatial-frequency enhancement. Image Processing Systems, Synoptics, Ltd., 271 Cambridge Science Park, Milton Road, Cambridge CB4 4WE, UK. Fax: (1223) 420020
- SHRLI: (Simulated High-Resolution Lattice Images) simulations from models up to 3 nm × 3 nm in area (perfect crystal or small defect structures). From the National Center for Electron Microscopy, Lawrence Berkeley Laboratories, University of California, Berkeley, CA 94729, or on the Web at http://ncem.lbl.gov/ncem.html
- Microdiffraction Programs: the listings for several programs are included in Appendix 5 of the book by Spence and Zuo (1992).

CHAPTER SUMMARY

TEMs comprise a range of different instruments which make use of the properties of electrons, both as particles and as waves. The TEM offers a tremendous range of signals from which we can obtain images, diffraction patterns, 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 electron microscopy, 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 the TEM, but we can explain in a generic sense, in many cases, what you have to do to get your microscope to deliver the enormous amounts of information that it generates. Not least of course, we also describe what you need to know to *interpret* the images, diffraction patterns, and spectra that you obtain.

In addition to the WWW, 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, most of which remain in print, as well as some standard journals and regular conference proceedings.

REFERENCES

General References for TEM

In the reference sections throughout the book, we will list general references that amplify the overall theme of the chapter, as well as specific references that are the source of information referenced in the chapter. If a general reference is referred to specifically in the chapter, we will *not* duplicate it in the specific references.

Books

Amelinckx, S., Gevers, R., and Van Landuyt, J., Eds. (1978) *Diffraction and Imaging Techniques in Material Science,* **1** and **2**, 2nd edition, North-Holland, New York. A collection of excellent individual review articles.

Cowley, J.M., Ed. (1992) *Electron Diffraction Techniques,* **1** and **2**, Oxford University Press, New York. Another collection of excellent individual review articles.

Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science,* Van Nostrand Reinhold, New York. The original out-of-print 1976 edition has been reprinted by TechBooks, 4012 Williamsburg Court, Fairfax, Virginia 22032. It is an essential text, if somewhat outdated.

Goodhew, P.J. and Humphreys, F.J. (1988) *Electron Microscopy and Analysis,* 2nd edition, Taylor and Francis, New York. A succinct summary of SEM, TEM, and AEM.

Hall, C.E. (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, P.W. and Kasper, E. (1989, 1994) *Principles of Electron Optics,* **1–3**, Academic Press, New York. 1900 pages, 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 modern resource.

Heidenreich, R.D. (1964) *Fundamentals of Transmission Electron Microscopy,* Interscience Publisher, New York. Another wonderful but sometimes forgotten classic.

Hirsch, P. B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, New York. For many years, the "Bible" for TEM users!

Loretto, M.H. (1994) *Electron Beam Analysis of Materials,* 2nd edition, Chapman and Hall, New York. A concise overview of the subject.

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

Reimer, L. (1993) *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis,* 3rd edition, Springer-Verlag, New York. Essential reference text. Strong physics background required.

Sarikaya, M., Ed. (1992) *Resolution in the Microscope, Ultramicroscopy,* **47**. Actually, a collection of reviews which all concern resolution in different instruments.

Sawyer, L.C. and Grubb, D.T. (1987) *Polymer Microscopy,* Chapman and Hall, New York. A broad-based qualitative introduction to TEM and SEM of polymers.

Thomas, G. (1962) *Transmission Electron Microscopy of Metals,* Wiley, New York. A historical volume—the first hands-on book for the materials scientist.

Thomas, G. and Goringe, M.J. (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, 4012 Williamsburg Court, Fairfax, Virginia 22032.

von Heimendahl, M. (1980) *Electron Microscopy of Materials,* Academic Press, New York. An introductory-level text, no significant AEM or HRTEM component.

Watt, I.M. (1985) *The Principles and Practice of Electron Microscopy,* Cambridge University Press, New York. A basic, practical introduction to SEM and TEM.

Wenk, H.-R. (1976) *Electron Microscopy in Mineralogy,* Springer-Verlag, New York. Required reading for microscopy of geological or ceramic materials.

Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd edition, Philips Electron Optics Publishing Group, Mahwah, New Jersey. A basic introduction to AEM. The original out-of-print 1987 edition has been reprinted by TechBooks, 4102 Williamsburg Court, Fairfax, Virginia 22032.

Journals

Advances in Imaging and Electron Physics, Academic Press, New York; formerly *Advances in Optical and Electron Microscopy.*

Journal of Microscopy, Blackwell Science, Oxford, United Kingdom.

Microscopy and Microanalysis, Springer, New York (formerly *Journal of the Microscopy Society of America,* Jones and Begell Publishing, Boston.)

Microscopy, Microanalysis, Microstructure (formerly *Journal de Microscopie et Spectroscopie Electronique*), Les Editions de Physique, Les Ulis Cedex A, France.

Microscopy Research and Technique (formerly *Journal of Electron Microscopy Technique*), Wiley-Liss, New York.

Ultramicroscopy, Elsevier Science Publishers, Amsterdam, the Netherlands.

Conference Proceedings

International Congress for Electron Microscopy—every four years (1994).

European Electron Microscopy Congress—every four years (1996).

Microscopy Society of America, San Francisco Press, San Francisco—annual.

Microbeam Analysis Society, VCH, Deerfield Beach, Florida. From 1996, San Francisco Press, San Francisco—annual.

Electron Microscopy and Analysis, Institute of Physics, Bristol, United Kingdom—odd years (1995).

Scanning Electron Microscopy, Scanning Microscopy International, AMF O'Hare, Illinois—annual.

Useful Sources of Numerical Data and Constants

Fischbeck, H.J. and Fischbeck, K.H. (1987) *Formulas, Facts and Constants,* 2nd edition, Springer-Verlag, 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, A.G. (1991) *Handbook for Crystallography for Electron Microscopists and Others,* Springer-Verlag, New York. Ideal for the microscopist, but see the review by A. Eades (*Microsc. Res. Technique* **21**, 368).

Specific References in This Chapter

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

Cosslett, V.E. (1979) The Cosslett Festschrift, *J. Microsc.* **117**, 1.

Davisson, G. and Germer, L.H. (1927) *Phys. Rev.* **30**, 705.

de Broglie, L. (1925) *Ann. de Physiques* **3**, 22.

Ewald, P.P. (1962) *Fifty Years of X-ray Diffraction,* International Union of Crystallography, 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.

Hashimoto, H. (1986) *J. Electron Microsc. Tech.* **3**, 1.

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

Hayes, T.L. (1980) SEM-1980 **1,** 1 (Ed. O. Johari), SEM Inc., AMF O'Hare, Illinois.

Head, A.K., Humble, P., Clarebrough, L.M., Morton, A.J., and Forwood, C.T. (1973) *Computed Electron Micrographs and Defect Identification,* North-Holland, New York.

Heidenreich, R.D. (1949) *J. Appl. Phys.* **20**, 993.

Joy, D.C. (1995) *Monte Carlo Modeling for Electron Microscopy and Microanalysis,* Oxford University Press, New York.

Knoll, M. and Ruska, E. (1932) *Z. Physik* **78**, 318.

Kossel, W. and Möllenstedt, G. (1939) *Ann. Phys.* **36**, 113.

Marton, L. (1994) *Early History of the Electron Microscope,* 2nd edition, San Francisco Press, San Francisco.

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

Spence, J.C.H. and Zuo, J.M. (1992) *Electron Microdiffraction,* Plenum Press, New York.

Thompson, G.P. and Reid, A. (1927) *Nature* **119**, 890.

Scattering and Diffraction

2

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 the 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 didn't scatter, then there would be no mechanism to create TEM images or diffraction patterns 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 three 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?
- 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 ideas of cross sections and mean free paths, so we define these concepts in this chapter. The angle of scattering, usually determined through the differential cross section, is also important because it allows you as 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 III of the book. To answer the third question we must distinguish elastic and inelastic scattering. The former constitutes most of the useful information in diffraction patterns obtained in the TEM, discussed in Part II, while the latter is the source of X-rays and other spectroscopic signals discussed in Part IV. The distinction between electrons that lose energy and those that don't is important enough that we devote the subsequent two chapters to each kind of electron and expand on the basic ideas introduced here.

The electron beam is treated in two different ways: 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, we must always remember that electrons are charged particles and that Coulomb forces are very strong.

Scattering and Diffraction

2

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 eye cannot see anything unless it interacts with visible light in some way, for example through reflection or refraction, which are two forms of scattering. Similarly, we cannot see anything in electron microscope images unless the specimen interacts with, and scatters, the electrons in some way. Thus any nonscattering object is invisible, and we will come across situations where "invisibility" is an important criterion. 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, including backscattered and secondary electrons, are also of interest and we will not neglect them, although they are of the greatest interest to SEM users and give surface-sensitive information, such as topography.

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 electrons in which the electron 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. For example, the electron beam that comes through the specimen, parallel to the direction of the incident beam, is an important beam, which we will term the *direct beam.*

As the electrons travel through the specimen, they are either scattered by a variety of processes or they may remain unaffected by the specimen. The end result is that a nonuniform distribution of electrons emerges from the exit surface of the specimen, as shown schematically in Figure 2.1. It is this nonuniform distribution that contains all the structural and chemical information about our specimen. So everything we learn about our specimen using TEM can be attributed to some form of electron scattering.

We'll see in Chapter 9 that the electron microscope is constructed to display this nonuniform distribution of electrons in two different ways. First, the *angular distribution* of scattering can be viewed in the form of scattering patterns, usually called diffraction patterns, and the *spatial distribution* of scattering can be observed as contrast in images of the specimen. 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. Thus, you as operator have the ability to choose which electrons you want to use and thus you control what information will be present in the image. Therefore, to comprehend these images, you have to understand what causes electrons to scatter in the first place.

We devote the whole of Part II to diffraction phenomena and the whole of Part III to images. Then, Part IV deals with ways in which we use inelastic scattering to

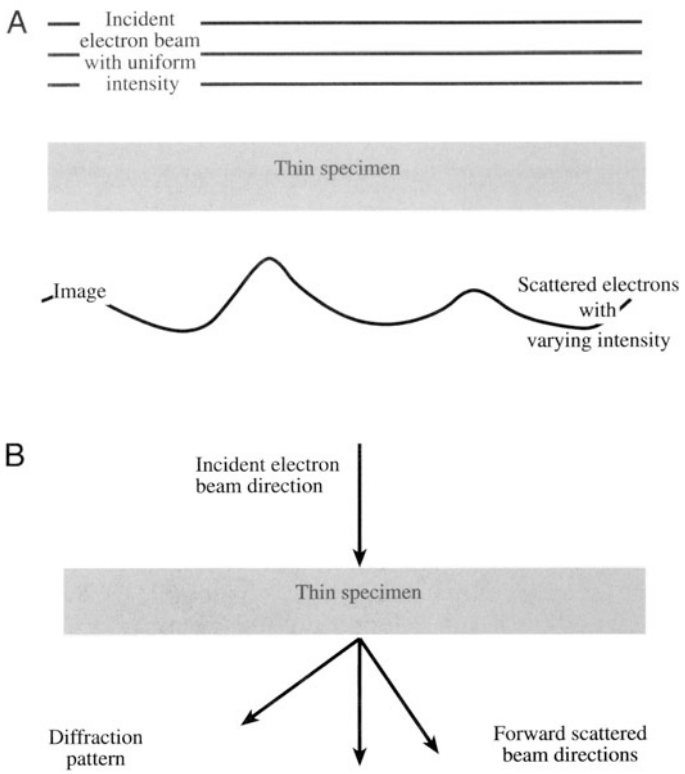


Figure 2.1. (A) A uniform intensity of electrons, represented by the flat line, falls on a thin specimen. Scattering within the specimen changes both the spatial and angular distribution of the emerging electrons. The spatial distribution (intensity) is indicated by the wavy line. (B) The change in angular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

study the chemistry of the specimen. So electron scattering is the theme that permeates this text and connects all aspects of TEM.

2.2. TERMINOLOGY OF SCATTERING

Electron scattering can be grouped in different ways. We've already used the most important terms *elastic* and *inelastic* scattering. These terms are simply descriptions of scattering that results in no loss of energy and some measurable loss of energy, respectively. In this case we tend to consider the electrons as particles, and scattering to involve some interaction like billiard balls colliding. However, we can also separate scattered electrons into *coherent* and *incoherent*, which refers of course to their wave nature. These distinctions are related, since elastic electrons are usually coherent and inelastic electrons are usually incoherent. 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. (We'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 normal to the beam. (Note: you will sometimes see the term forward scattering used in another sense.) If an electron is scattered through $< 90°$ then it is forward scattered, and if $> 90°$ 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.

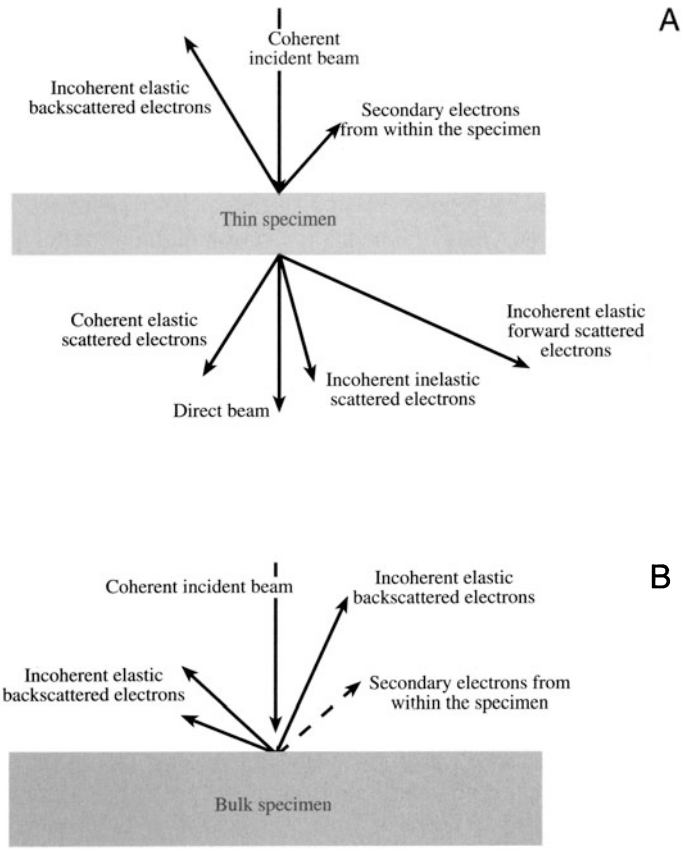


Figure 2.2. Different kinds of electron scattering from (A) a thin specimen and (B) a bulk specimen: a thin specimen permits electrons to be scattered in both the forward and back directions while a bulk specimen only backscatters the incident beam electrons.

- Elastic scattering usually occurs at relatively low angles (1–10°), i.e., in the forward direction.
- At higher angles (> ~10°) elastic scattering becomes more incoherent.
- Inelastic scattering is almost always incoherent and relatively low angle (<1°) forward scattering.
- As the specimen gets thicker, less electrons are forward scattered and more are backscattered until primarily incoherent backscattering is detectable in bulk, nontransparent specimens.

The notion that electrons can be scattered through different angles is related to the fact that electrons can also be scattered more than once. Generally, the more scattering events, the greater the angle of scatter, 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 the specimen to a single scattering event (i.e., an electron either undergoes a single scattering event or it suffers no scattering). We'll see that this is often 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 grim specimen, multiple scattering will not occur. 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, diffraction patterns, and spectra that we gather. So once again we emphasize the importance of creating thin specimens so that the single scattering assumption is plausible.

In the transmission electron microscope 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. We might ask what percentage of the electrons are forward scattered and how does this vary with the thickness 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 causes most of the signals used in the TEM.

Forward scattering includes elastic scattering, Bragg scattering, the events called diffraction, refraction, and inelastic scattering. Because of forward scattering through our thin specimen, we see a diffraction pattern or an image on the viewing screen, and detect an X-ray spectrum or an electron energy-loss spectrum outside the optical column. But don't neglect backscattering; it is an important imaging mode in the SEM.

2.3. THE CHARACTERISTICS OF ELECTRON SCATTERING

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. 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). 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°.

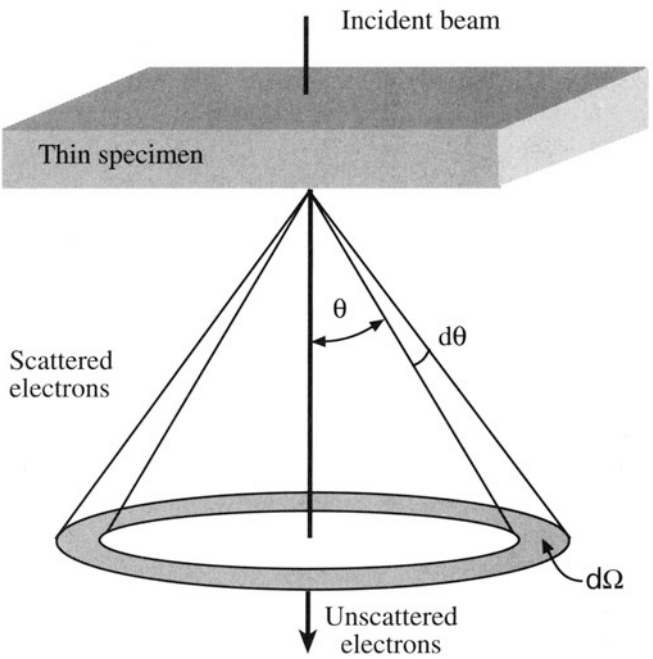


Figure 2.3. Electron scattering by a single isolated atom. The electrons are scattered through a semiangle θ and the total solid angle of scattering is Ω. An incremental increase in scattering angle $d\theta$ gives an incremental increase in a solid angle $d\Omega$.

A convenient definition of a small angle is 10 mrads.

The characteristics of the scattering event are controlled by factors such as the electron energy and the atomic number/weight of the scattering atom. When we consider a specimen rather than a single atom, factors such as the thickness, density, and crystallinity of the specimen also become important. To understand these variables, we need to examine the physics of scattering in more detail.

2.4. THE INTERACTION CROSS SECTION

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 letter Q or the Greek σ) has units of area, not square feet as used in Peierls's analogy, but a tiny fraction of a square centimeter termed a "barn." One barn is 10^{-24} cm^2 [that's $(10^{-4}\text{Å})^2$] and the name arises because of the perverse sense of humor of some of the early atomic physicists who considered that this area is "as big as a barn door." Note again the use of non-SI units (cm^2) which persists in the literature, although we can easily define the barn as 10^{-28} m^2. 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. So the larger the cross section, the better the chances of scattering.

We can look at scattering in two different, but equivalent, ways. First, following Hall (1953), since we are ignoring different kinds of scattering, we can talk about the *total* scattering cross section for the isolated atom, σ_T, which is simply the sum of all elastic and inelastic scattering cross sections such that the total σ_T is

$$\sigma_T = \sigma_{elastic} + \sigma_{inelastic} \qquad [2.1]$$

We can define the cross section (an area) in terms of the *effective radius* of the scattering center, r

$$\sigma = \pi r^2 \qquad [2.2]$$

where r has a different value for each of the scattering processes. For example, in the case of elastic scattering, which we'll see is most important in TEM image and diffraction pattern formation, the radius is given as

$$r_{elastic} = \frac{Ze}{V\theta} \qquad [2.3]$$

where V is the potential of the incoming electron, charge e, which is scattered through an angle greater than θ by atoms of atomic number Z. At first sight this equation seems dimensionally incorrect, since Z and θ are numbers. But V is in volts, so e has to be defined in esu such that the whole term has units of distance. This expression is useful because it indicates the general behavior of electrons in the TEM; i.e., usually electrons scatter less at high kV and high angles, and are scattered more by heavier atoms than light atoms.

Second, following Heidenreich (1964), if we instead consider that the specimen contains N atoms/unit volume, we can define the total cross section for scattering from the specimen (in units of cm^{-1} or m^{-1}) as

$$Q_T = N\,\sigma_T = \frac{N_0\,\sigma_T\,\rho}{A} \qquad [2.4]$$

where N_0 is Avogadro's number (atoms/mole), and A is the atomic weight (g/mole) of the atoms in the specimen which has density ρ (so $NA = N_0\rho$). Thus Q can be regarded as 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

$$Q_T\,t = \frac{N_0\,\sigma_T\,(\rho\,t)}{A} \qquad [2.5]$$

The product of ρ and t 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 2.5 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.

Expressions for the cross section become more complicated as they are modified to give better approximations for the scattering in a real specimen. For example, the expression for $r_{elastic}$ which we would substitute in the cross-section equation for elastic scattering (equation 2.2) neglects any screening effects of the electron cloud around the nu-

cleus, which obviously acts to reduce the Z effect. However, the more complex equations don't alter the basic scattering behavior predicted by the simple equations we've just given. If you want to see a fuller description of scattering, then read Chapter 1 by Newbury in Joy *et al.* (1986). If you're a glutton for punishment, the classical text on scattering is by Mott and Massey (1965).

Because of all the variables that affect σ and Q, it is only possible to give a ball-park value for the cross section. For TEM electron energies, 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 ~10^{-22} m^2 (~10^{-18} cm^2). This is a good number to remember especially when you are considering the probability that a 100-keV electron will be elastically scattered. Inelastic cross sections range from ~10^{-22} m^2 down to 10^{-26} m^2 (100 barns), depending on the specific type of scattering and the material.

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 an important concept. The total cross section for scattering can be expressed as the inverse of the mean free path, λ. 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. Because the dimensions of Q are (length)$^{-1}$ there is a simple expression for the mean free path λ which has units of length

$$\lambda = \frac{1}{Q} = \frac{A}{N_0 \sigma_T \rho} \qquad [2.6]$$

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 this equation we can define a probability of scattering p as the electron travels through a specimen thickness t

$$p = \frac{t}{\lambda} = \frac{N_0 \sigma_T (\rho t)}{A} \qquad [2.7]$$

which is just $Q_T t$ of equation 2.5.

2.6. THE DIFFERENTIAL CROSS SECTION

Because of the importance of the angle of scattering we need to introduce the concept of the *differential cross section* $d\sigma/d\Omega$. This term 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 the θ and Ω

$$\Omega = 2\pi(1 - \cos\theta) \qquad [2.8]$$

and therefore

$$d\Omega = 2\pi \sin\theta \, d\theta \qquad [2.9]$$

So the differential scattering cross section can be written as

$$\frac{d\sigma}{d\Omega} = \frac{1}{2\pi \sin\theta} \frac{d\sigma}{d\theta} \qquad [2.10]$$

Now, we can calculate σ for scattering into all angles which are greater than θ by integrating equation 2.10. This yields

$$\sigma_\theta = \int_\theta^\pi d\sigma = 2\pi \int_\theta^\pi \frac{d\sigma}{d\Omega} \sin\theta \, d\theta \qquad [2.11]$$

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 we measure experimentally, equation 2.11 gives us an easy way to determine σ for an atom in the specimen: σ for all values of θ is simply the integral from 0 to π. From this we can use equation 2.4 to give us the total scattering cross section from the whole specimen, which we will see later allows us to calculate the TEM image contrast. So we can now appreciate, through a few simple equations, the relationship between the physics of electron scattering and the information we collect in the TEM.

Our knowledge of the values of σ and λ is very sketchy, 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 to predict the electron paths as a beam is scattered through a thin foil.

This process is called Monte Carlo simulation because of the use of random numbers in the computer programs; the outcome is always predicted by statistics!

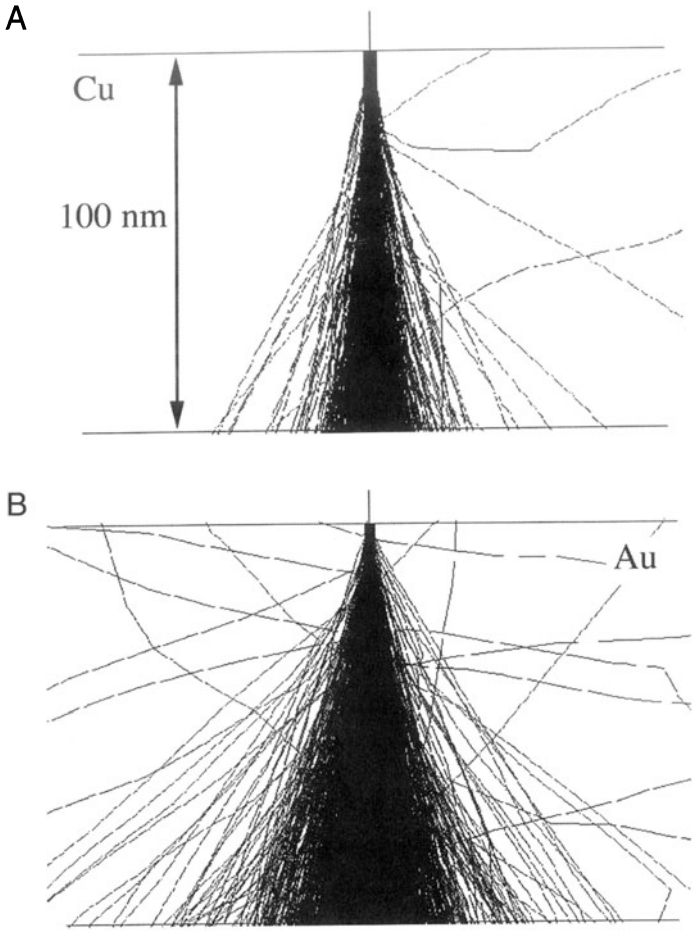


Figure 2.4. Monte Carlo simulation of the paths followed by 10^3 100-keV electrons as they pass through thin foils of (A) copper and (B) gold. Notice the increase in scattering with atomic number.

The Monte Carlo calculation was first developed by two of the United States' foremost mathematicians, J. von Neumann and S. 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 more often used in SEM image calculations (see, e.g., Newbury *et al.* 1986, Joy 1995), but they have a role in TEM in determining the expected spatial resolution of microanalysis. Figure 2.4 shows two Monte Carlo simulations of electron paths through thin foils.

2.7. OTHER FACTORS AFFECTING SCATTERING

By selecting electrons of a certain scattering 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 of the scattered electrons are within ±5° of the unscattered beam.

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.

> Q decreases as E_0 increases. Electron scattering at 300 kV will be smaller than at 100 kV.

We shall see later that the effect of the atomic number of the specimen 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.

2.8. 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, dynamical scattering, elastic scattering, inelastic scattering, etc., 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 diffraction pattern, or collect a spectrum. 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 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 from the processes by which X-rays are scattered.

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 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 in the case of X-rays. Consequently, electrons are scattered much more strongly than X-rays.

2.9. FRAUNHOFER AND FRESNEL DIFFRACTION

Diffraction of visible light is well understood, so we should carry over as much of the analysis as possible. 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.

So why discuss these topics now? 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 to many slits. There are two reasons for reviewing this analysis:

- It reminds us that coherent interference is purely a matter of physical optics.
- We can review 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.

2.10. DIFFRACTION OF LIGHT FROM SLITS

When we place a pair of very narrow slits in front of a wavefront, we 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 **r** are out of phase by $2\pi L/\lambda$. If d and λ are such that this phase difference is actually a multiple of 2π, then the rays are again in phase. Therefore, there is an inverse relationship between d and θ for a given d; as d decreases, $\sin \theta$ increases.

> The inverse relationship between d and θ occurs solely due to the positions of the slits. We'll come across an identical relationship when we talk about electron diffraction in Section 3.11.

When we extend this analysis to more than two slits we see the same 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 26.) These diagrams plot the amplitude and phase of the scattered wave as illustrated, for the case of five slits, by the polyhedra in Figure 2.6; in other words, when we add the amplitudes of beams we must take account of their phase. When θ is zero, the rays experience no phase shift and we simply add all of the ampli-

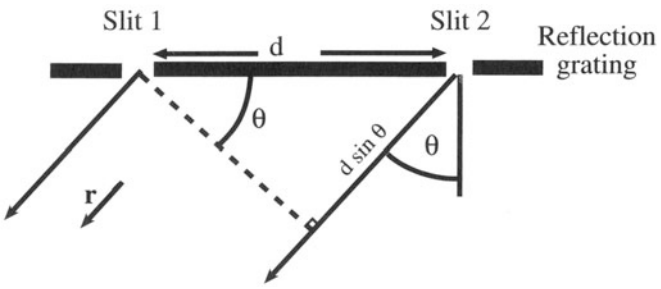


Figure 2.5. An incident plane wave is scattered by two slits, distance d apart. The scattered waves are in phase when the path difference $d \sin \theta$ is $n\lambda$.

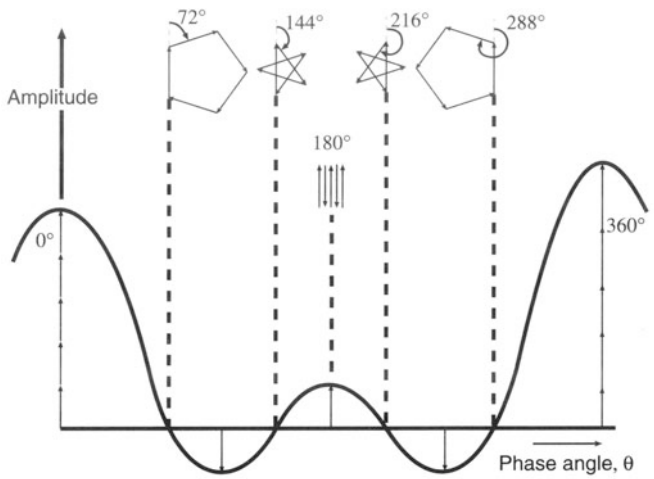


Figure 2.6. A phasor diagram showing how the total amplitude produced by summing five waves produced by five slits varies with the phase angle (how much each wave is out of phase) between the different waves. The arrows show the individual phasors from each slit.

tudes; as θ increases the rays become out of phase, but the phasors can still add to give a large resultant vector. When θ is exactly 72° (360°/5 for five slits), the phasor diagram is a closed pentagon and the resultant amplitude is zero. This process repeats at 144° (2 × 360°/5) and 216° (3 × 360°/5). In between these values at 108° (1.5 × 360°/5) we produce a local maximum in amplitude which is repeated at 180° (2.5 × 360°/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.

Now what happens if we allow the slit to have some width as shown in Figure 2.7A? Each slit produces a phasor diagram as shown in Figures 2.7B and C; i.e., the rays from within a single slit will interfere with each other to modify the polyhedra in Figure 2.6. The amplitude from a single slit varies as $A = A_0\, \phi^{-1} \sin \phi$, where ϕ is the phase $\lambda^{-1}\pi w \sin \theta$ for a single slit of width w. If we imagine just one slit, we would see a zero in the phasor diagram when $\phi = \pm n\pi$ as shown in Figure 2.7D.

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 of $1.22\lambda D^{-1}$ which is shown in Figure 2.8.

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.

The disk of diameter $1.22\lambda D^{-1}$ is named after Airy and will be one of the fundamental limits on the achievable resolution in TEM, as we discuss in Chapter 6. If we introduce *any* aperture into *any* microscope, we will limit the ultimate resolution of the instrument. As the diameter of the aperture, D, decreases, the minimum resolvable spacing, r, increases. This equation also suggests that decreasing λ (increasing the accelerating voltage) will improve resolution: as λ decreases, r decreases.

The final step is to consider the amplitude scattered from many slits which each have a width, w. The result is shown in Figure 2.9, where we've increased λ/w relative to Figure 2.7.

The important point about this analysis for TEM is that we'll see the same relationship in later chapters, where the slits will be replaced by an aperture, by many atoms, or by a thin specimen.

2.11. COHERENT 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 for this wave by the expression

$$\psi = \psi_0 e^{i\phi} \qquad [2.12]$$

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

$$\Delta\phi = \frac{2\pi}{\lambda}\Delta x \qquad [2.13]$$

Figure 2.10 will help you visualize the relationships between the path difference and phase difference for monochromatic waves. Coherent interference between waves relies on the fact that the waves add amplitudes with attention to phase. If all waves scattered by all of the atoms in the specimen are to interfere coherently, 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

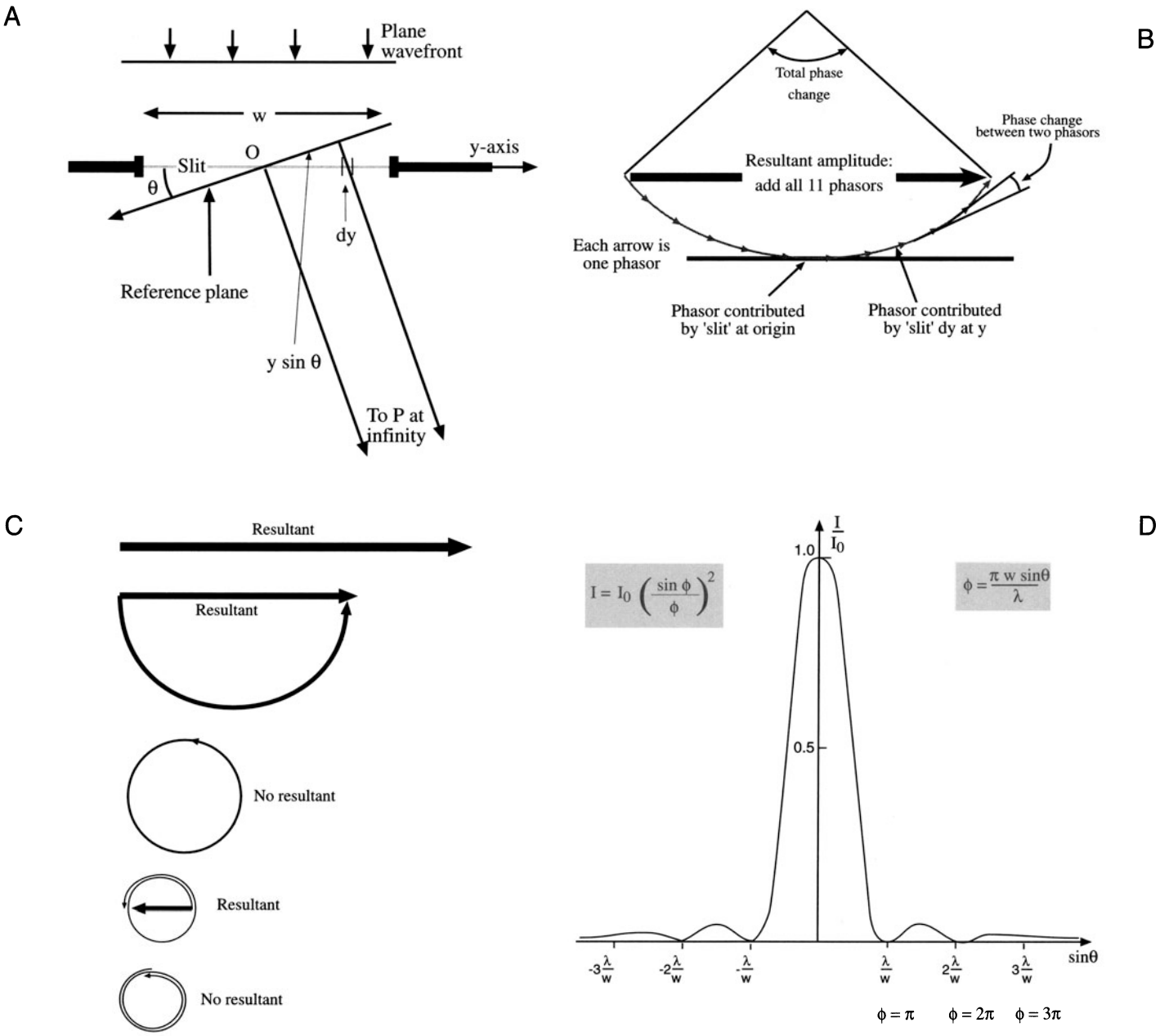


Figure 2.7. (A) Geometry for the scattering from an individual slit. (B) How the phasors from within an individual slit can be added to give the total phasor for the slit shown in (A). (C) How a single slit can produce a beam which has zero amplitude for certain values of θ in (A). The circles are directly comparable to the polyhedra in Figure 2.6. The total length of the phasor increments (from each *dy*) is the same in each figure. (D) A plot of the resulting intensity for scattering from the slit shown in Figure 2.5; this is known as the Fraunhofer diffraction pattern from a single slit; *w* is the slit width defined in (A).

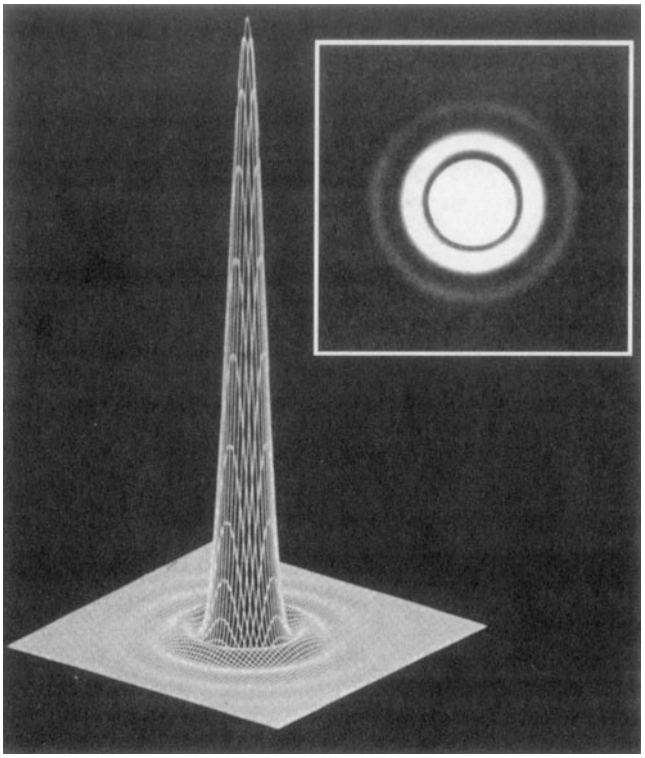


Figure 2.8. The visible-light intensity produced by a 0.5-mm-diameter circular aperture and the observed Airy rings (inset).

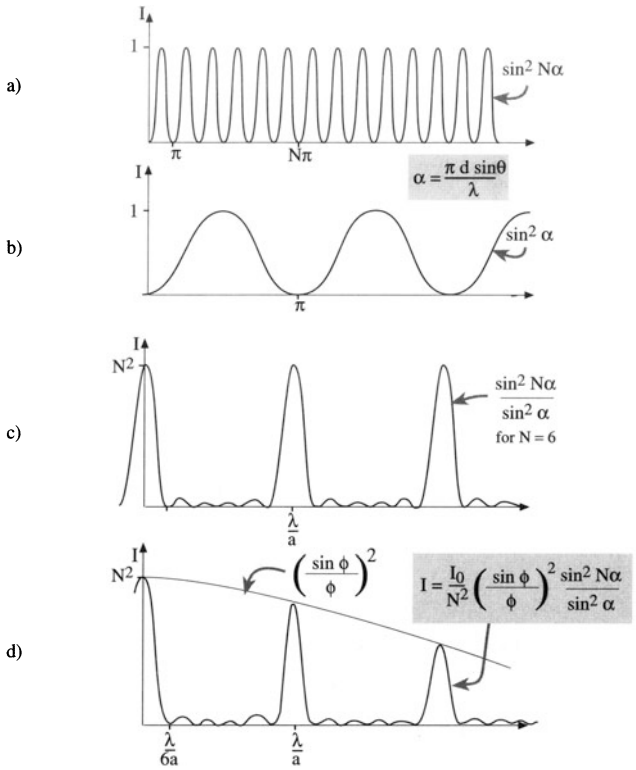


Figure 2.9. The scattered intensity from N slits (shown here for $N = 6$) where each slit would give the intensity shown in (B). (C) is the curve in (A) divided by curve in (B) and (D) is the curve in (C) multiplied by the curve in (Figure 2.7D). The distance d, the separation of the slits, and ϕ are defined in Figures 2.5 and 2.7. (λ/w has been increased compared to Figures 2.7 and 2.8 for simplicity.)

be periodically spaced. Fortunately this is true for all crystals, and the mathematical description of coherent interference is simplified (Part II).

2.12. A WORD ABOUT ANGLES

Since 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 semiangle of incidence as α, as summarized in Figure 2.10.
- In the TEM we use apertures or detectors to collect a certain fraction of the scattered electrons and we will define any semiangle of collection as β.
- We will define all scattering semiangles 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 semiangle θ.

2.13. ELECTRON DIFFRACTION PATTERNS

We've mentioned a couple of times that the TEM is uniquely suited to take advantage of electron scatter because it can form a picture (diffraction pattern) of the distribution of scattered electrons, which we'll discuss in Part II in much more detail. To understand fully how a diffraction pattern 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 lenses to form the TEM imaging system. But before we take you through these concepts it is worth just showing a few of the many kinds of diffraction patterns that can be formed in the TEM. At this stage, all you have to do is imagine that a photographic film is

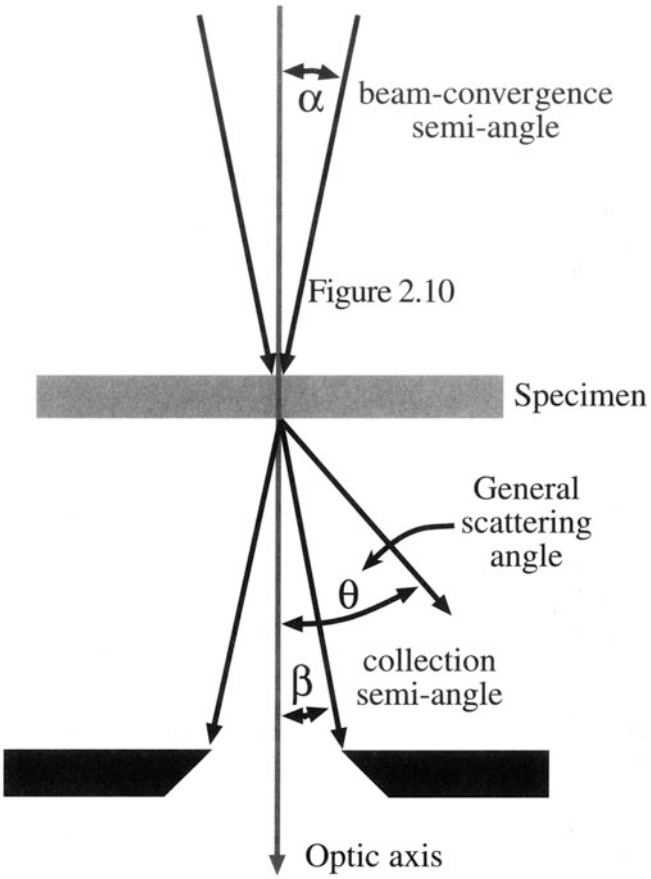


Figure 2.10. Definition of the major semiangles in TEM. Any incidence/convergence semiangle of the beam is termed α; any collection semiangle is β and general scattering semiangles are θ.

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. Thus distances on the film correspond to angles of scatter at the specimen. This relationship is different than the usual interpretation of images in which distances correspond to distances in the specimen, but it is critical to our understanding of diffraction patterns.

Even using this simple description, however, you can comprehend some of the basic features of diffraction patterns. Figure 2.11 is a montage of several kinds of diffraction patterns, 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 are *not* scattered but travel straight through the specimen. Second, the scattered intensity falls with increasing θ (increasing distance from the direct beam), which reflects the decrease in the scattering cross section with θ. Third, the scattering intensity varies strongly with the structure of the specimen. You'll see much more of this in Part II.

So far, in fact, we've only considered the amplitude of the electron wave and we've 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 unscattered. The phase of the scattered wave is most important in the specific topic of phase-contrast images, which are 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. Thus 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 classical calculation for the Coulomb force.)

We have defined three important parameters in this chapter:

σ	the scattering cross section of one atom
λ	the mean free path (average distance between scattering events)
$d\sigma/d\Omega$	the differential scattering cross section of one atom

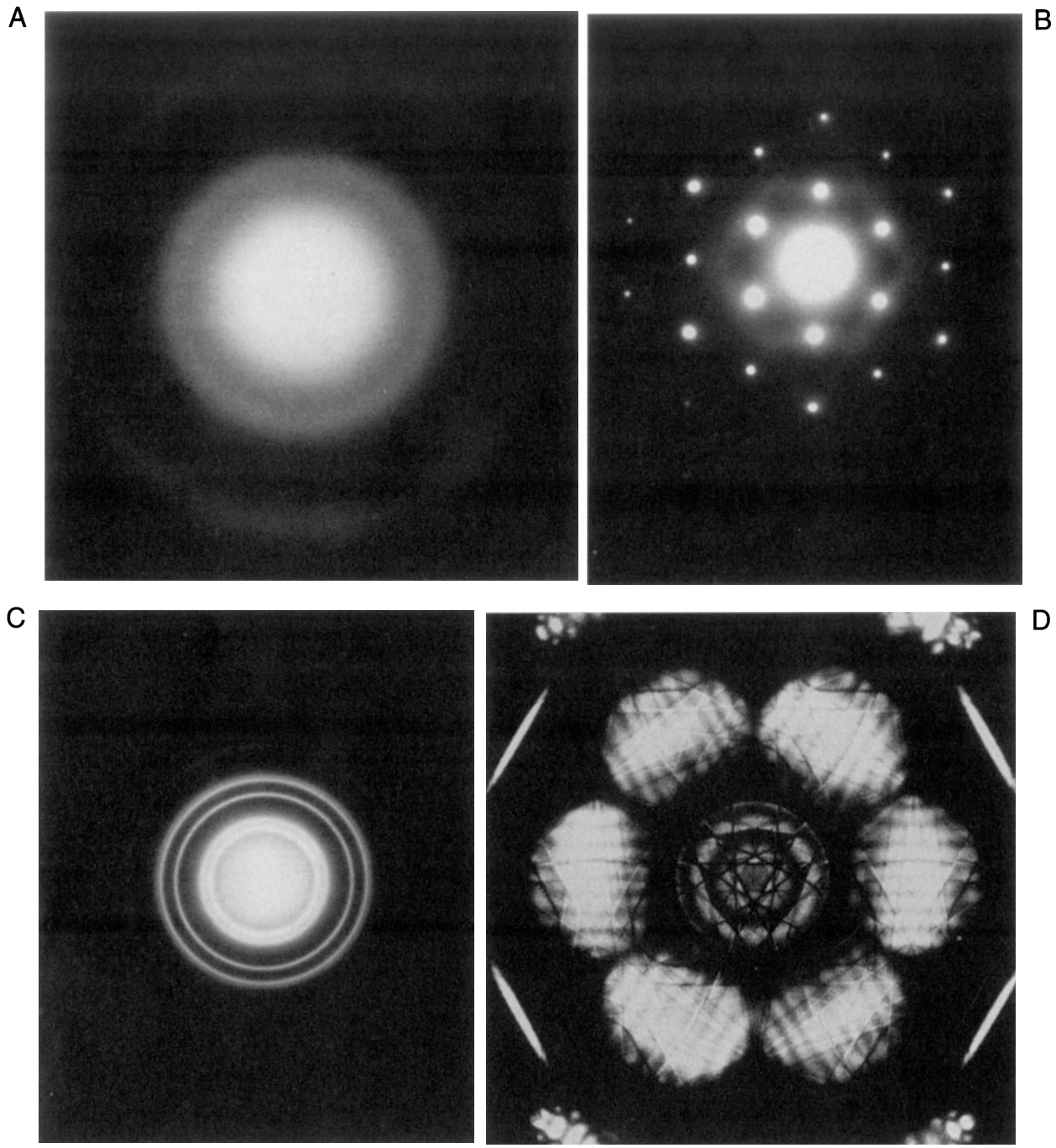


Figure 2.11. Several kinds of diffraction patterns obtained from a range of materials in a conventional 100 kV TEM: (A) amorphous carbon, (B) an Al single crystal, (C) polycrystalline Au, (D) Si illuminated with a convergent beam of electrons. In all cases the direct beam of electrons is responsible 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.

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, we'll use electron "scattering" to denote the effect and to be consistent with the popular usage which goes back to the early work of Bragg and others.

REFERENCES

General References

Fishbane, P.M., Gasiorowicz, S., and Thornton, S.T. (1993) *Physics for Scientists and Engineers,* Prentice Hall, Englewood Cliffs, New Jersey.

Goodman, J.W. (1968) *Introduction to Fourier Optics*, McGraw-Hill, New York. Reissued in 1988 and an excellent source for the advanced student.

Hecht, E. (1987) *Optics*, 2nd edition, Addison-Wesley, Reading, Massachusetts.

Klein, M.V. and Furtak, T.E. (1985) *Optics*, 2nd edition, Wiley, New York.

Mott, N.F. and Massey, H.S.W. (1965) *The Theory of Atomic Collisions*, Oxford University Press, New York.

Smith, F.G. and Thomson, J.H. (1988) *Optics*, 2nd edition, Wiley, New York.

Specific References

Hall, C.E. (1953) *Introduction to Electron Microscopy,* McGraw-Hill, New York.

Heidenreich, R.D. (1964) *Fundamentals of Transmission Electron Microscopy,* Interscience, New York.

Joy, D.C. (1995) *Monte Carlo Modeling for Electron Microscopy and Microanalysis*, Oxford University Press, New York

Joy, D.C., Romig, A.D. Jr., and Goldstein, J.I., Eds.(1986) *Principles of Analytical Electron Microscopy*, Plenum Press, New York.

Newbury, D.E., Joy, D.C., Echlin, P., Fiori, C.E., and Goldstein, J.I. (1986) *Advanced Scanning Electron Microscopy and X-ray Microanalysis*, Plenum Press, New York.

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

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

Elastic Scattering

3

CHAPTER PREVIEW

Elastically scattered electrons are the major source of contrast in TEM images and they also create the intensity distributions in diffraction patterns, so we need to understand what controls this process. We'll consider elastic scattering first 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 aspects of the character of the electron.

Scattering from isolated atoms can occur either as a result of electrons interacting with the negatively charged electron cloud and scattered through small angles of a few degrees, or attracted to the positive nucleus and scattered through large angles up to 180°. The scattering from the nucleus can be interpreted in terms of simple 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, we'll use this Z dependence in different ways to form images that reflect the chemistry of the specimen. We can also treat the electrons as waves, in which case their *coherency* becomes important. The coherency of the scattered electrons is related to their *semi-angle* of scattering (θ). As the scattering angle becomes larger, the degree of coherency becomes less and Rutherford-scattered electrons are incoherent.

In contrast to Rutherford high-angle scattering, electrons which are elastically scattered through less than ~3° are coherent. The intensity of this low-angle scattering is strongly affected by the arrangement of atoms within the specimen. Such collective scattering by the atoms is referred to as *diffraction* and can only be

understood if we treat the electron as a wave and ignore particle concepts such as cross sections. Diffraction is controlled mainly by the angle of incidence of the electron beam to the atomic planes in the specimen and the spacing of atoms or planes of atoms. So this low-angle scattering is invaluable in characterizing the crystallography of the specimen and is undoubtedly the most significant scattering phenomenon in the TEM.

When we think about the scattering of electrons, we often imagine a beam of particles which hits a target and is deflected to emerge as a beam in another direction, termed the scattered beam, much as we might imagine a beam of light being a group of photons. However, the scattering of light does not always follow the rules of geometric optics because light has a wave character. Similar considerations apply to the diffraction of electrons and this is one of the fundamental concepts of TEM. So you will find the wave-particle duality being used simultaneously, because of both lines of thought.

Elastic Scattering

3

3.1. PARTICLES AND WAVES

We have two different ways of looking at how an electron beam interacts with our specimen in the TEM. 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. We can summarize the two viewpoints:

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

- They have a scattering cross section and differential scattering cross section.
- They can be scattered through particular angles.
- The electrons interact with the nucleus 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.

When we discuss *X-ray* and *electron spectrometry* you'll see that we have to use a particle 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.
- We can relate the process to the scattering of X-rays, so lots of analysis already exists.

When we discuss *imaging, HRTEM,* and *diffraction patterns* you'll see that we use a wave description.

The terminology is sometimes confusing if you look at it closely. A clear definition of diffraction is given by Taylor (1987):

An interaction between a wave of any kind and an object of any kind.

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; it refers to any interaction involving a wave.

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.

It's convenient to divide elastic scattering mechanisms into two principal forms: electron scattering from isolated single atoms and collective scattering from many atoms together within the specimen. We'll start in the same way as 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

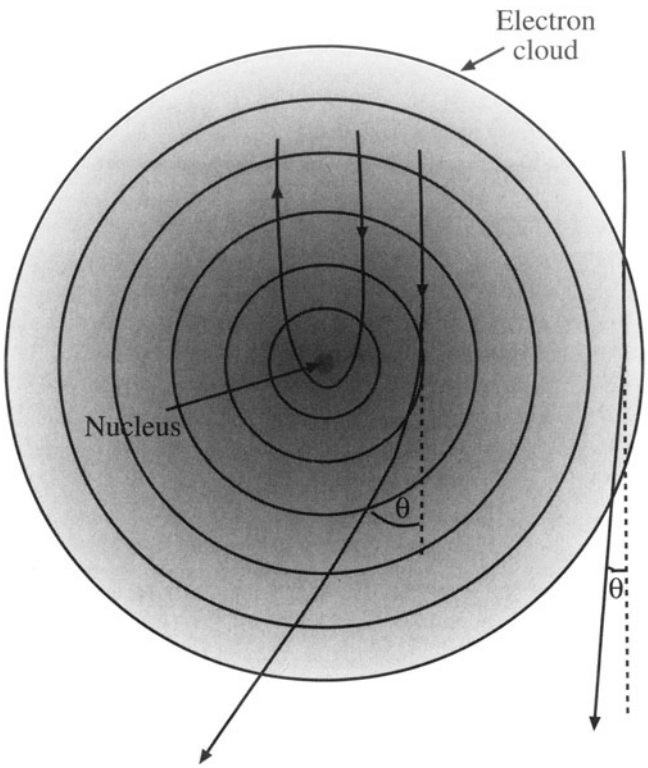


Figure 3.1. Two mechanisms by which a high-energy electron is scattered by an isolated atom. Coulombic interaction within the electron cloud results in low-angle (θ) scatter while Coulombic attraction by the nucleus causes high θ scatter and perhaps complete backscatter. The potential within the electron cloud is always positive.

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° (complete backscattering).

> 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 ignore any inelastic effects in this chapter.

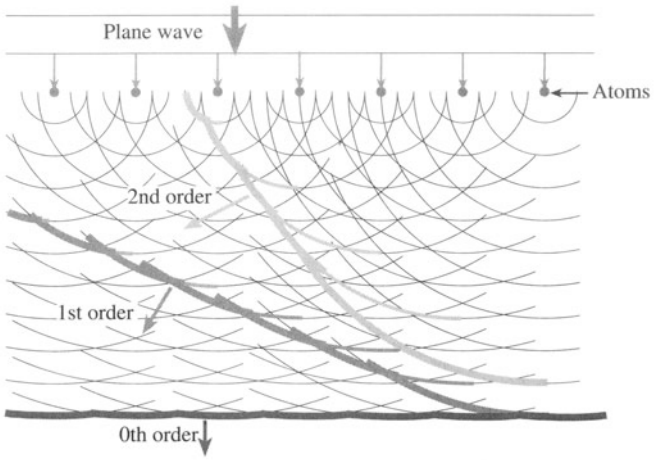


Figure 3.2. A plane coherent electron wave generates secondary wavelets from a row of scattering centers (e.g., atoms in the specimen). The secondary wavelets interfere, resulting in a strong direct (zero-order) beam and several orders of coherent beams scattered (diffracted) at specific angles.

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 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 acting as a source of "secondary" spherical wavelets as illustrated in Figure 3.2. These wavelets reinforce one another in certain angular directions and cancel in others. Thus the low-angle elastic scattering distribution is modified by the crystal structure of the specimen, and intense diffracted beams emerge at certain specific angles; we'll discuss these higher-order effects in Chapters 11 and 12. We'll now go on to examine these two forms of elastic scattering in more detail, starting with the simplest concept which is sometimes referred to as the billiard-ball model. We will briefly describe the scattering of a wave to show how it relates to this particle-scattering treatment and use this as the basis later for a full analysis of diffraction.

3.3. SCATTER FROM ISOLATED ATOMS

Consider the two paths for an electron passing close to an atomic nucleus shown in Figure 3.1. In either case, the di-

rection traveled by the electron changes; the electron is scattered through an angle θ.

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, scattering from an isolated atom, the electron–electron and electron–nucleus scattering cross sections can be easily expressed by two very simple equations (Hall 1953)

$$\sigma_{\text{electron}} = \pi r_e^2 = \pi\left(\frac{e}{V\theta}\right)^2 \qquad [3.1]$$

$$\sigma_{\text{nucleus}} = \pi r_n^2 = \pi\left(\frac{Ze}{V\theta}\right)^2 \qquad [3.2]$$

Remember that we are using a billiard-ball model where r_e and r_n are the radii of the electron cloud and the nucleus, respectively. Hopefully no one will ask you to prove these simple expressions.

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 in volts). 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. For the time being, we'll ignore the low-angle electron cloud scattering and concentrate only on scattering by the nucleus.

3.4. THE RUTHERFORD CROSS SECTION

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, Rutherford, to deduce the existence of the nucleus (never overlook undergraduate research results!). Rutherford (1911), describing backscattering as "the most incredible event that has ever happened to me," derived the following expression for the differential cross section for this kind of scattering

$$\frac{d\sigma(\theta)}{d\Omega} = \frac{e^4 Z^2}{16(E_0)^2 \sin^4\frac{\theta}{2}} \qquad [3.3]$$

All the terms in this equation were defined in Chapter 2. The expression 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. This is generally a good assumption in the TEM. We can substitute appropriate values for the various constants and integrate the differential cross section from 0 to π to obtain the total elastic nuclear cross section (in scattering events per electron per atom per m^2) in a more accurate form than that given in equation 3.2

$$\sigma_{\text{nucleus}} = 1.62 \times 10^{-24}\left(\frac{Z}{E_0}\right)^2 \cot^2\frac{\theta}{2} \qquad [3.4]$$

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 nucleus. As in Chapter 2, we can modify this expression for a single isolated atom to take into account the scattering from atoms in a TEM specimen of thickness t

$$\begin{aligned} Q_{\text{nucleus}}\, t &= \left(N_0 \frac{\rho}{A} t\right)\sigma \\ &= 1.62\, x\, 10^{-24} \left(N_0 \frac{\rho}{A} t\right)\left(\frac{Z}{E_0}\right)^2 \cot^2\left(\frac{\theta}{2}\right) \end{aligned} \qquad [3.5]$$

Notice that we still have the mass-thickness dependence, ρt, but that the strong dependence on Z is now obvious (compare to equation 2.5). See Joy *et al.* (1986) for further discussion of these calculations.

3.5. MODIFICATIONS TO THE RUTHERFORD CROSS SECTION

You'll often see the Rutherford differential cross section (equation 3.3) in different, but mathematically similar, forms. The expression given here neglects the screening effect of the surrounding electron cloud, which acts to reduce the differential cross section, thus lowering the amount of scattering. In other words, when the electron does not pass close to the nucleus, the scattering angle will be small (say < 5°) because screening is important. 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

$$\theta_0 = \frac{0.117\, Z^{1/3}}{E_0^{1/2}} \qquad [3.6]$$

(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 nuclear interaction is dominant. The value of θ_0 at 100 keV is only ~ 2° for Cu and less for lighter elements, so above this angle all scattering can be approximated to Rutherford-type high-angle scattering.

You should note that equation 3.3 is nonrelativistic, which is unfortunate since relativistic effects are significant for electrons with energies greater than 100 keV, as in the TEM. However, we can correct for relativity to give a more accurate differential cross section which is expressed using λ_R, the relativistically corrected wavelength (see equation 1.7), and a_0, the Bohr radius of the scattering atom

$$a_0 = \frac{h^2 \varepsilon_0}{\pi m_0 e^2} \qquad [3.7]$$

where ε_0 is the dielectric constant. Using the other constants listed in Table 1.1 we find a_0 is 0.0529 nm (remember this as 0.5 Å). The result is

$$\frac{d\sigma(\theta)}{d\Omega} = \frac{\lambda_R^4 Z^2}{64\pi^4 (a_0)^2 \left(\sin^2 \frac{\theta}{2} + \left(\frac{\theta_0}{2} \right)^2 \right)^2} \qquad [3.8]$$

This is called the screened relativistic Rutherford cross section.

This cross section is the one most widely used for TEM calculations, although it has particular limitations at the highest TEM voltages (300–400 kV) and for the heavier elements (Z>30), which cause large scattering 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 (Chapter 2).

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 with scattering angle for three different elements and two different electron-beam energies. As you can see, the cross section decreases by several orders of magnitude from ~10^6 barns to about 10 barns as the scattering angle increases from 0–180°; so, as expected, scattering is most likely to occur in the forward (θ close to 0°) direction. Doubling the electron-beam energy can lower the cross section by a factor of two or three, which confirms that higher-energy electrons are less likely to be scattered, 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 keep your specimen below 100-nm thickness and you can then approach the ideal of single scattering that we'll assume many times throughout this text.

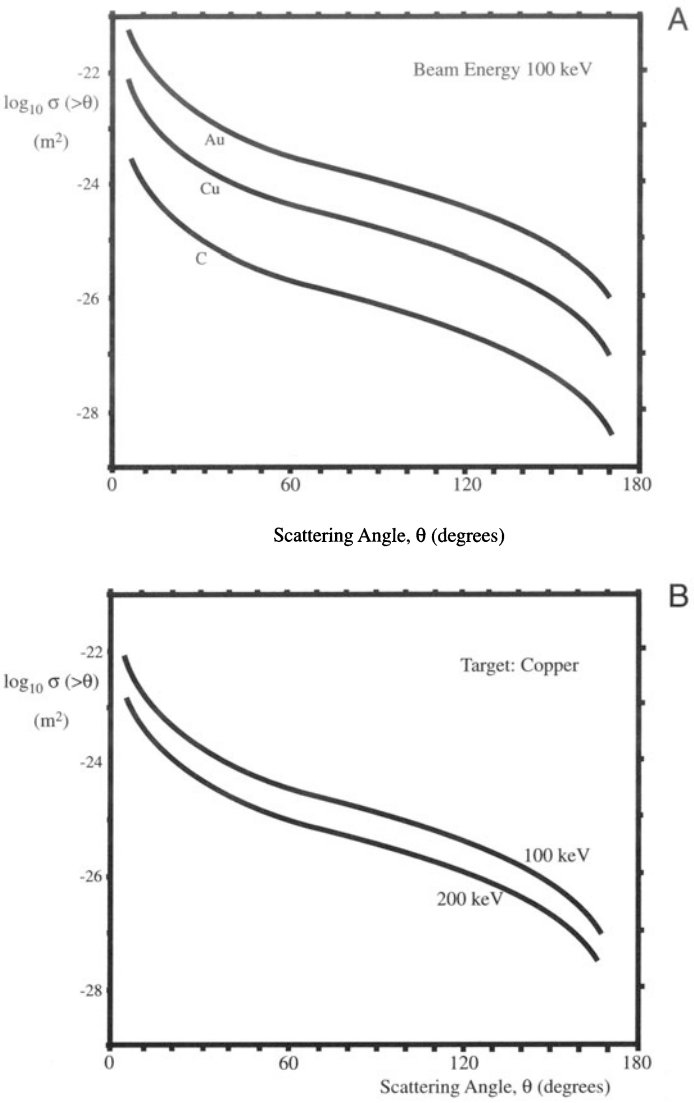


Figure 3.3. The variation of the logarithm of the screened relativistic Rutherford cross section with scattering angle: (A) for different elements at 100 keV and (B) for Cu at different TEM voltages.

3.6. COHERENCY OF THE RUTHERFORD-SCATTERED ELECTRONS

Up to now 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. The coherency of the scattered electron wave is a distinguishing characteristic. High-angle

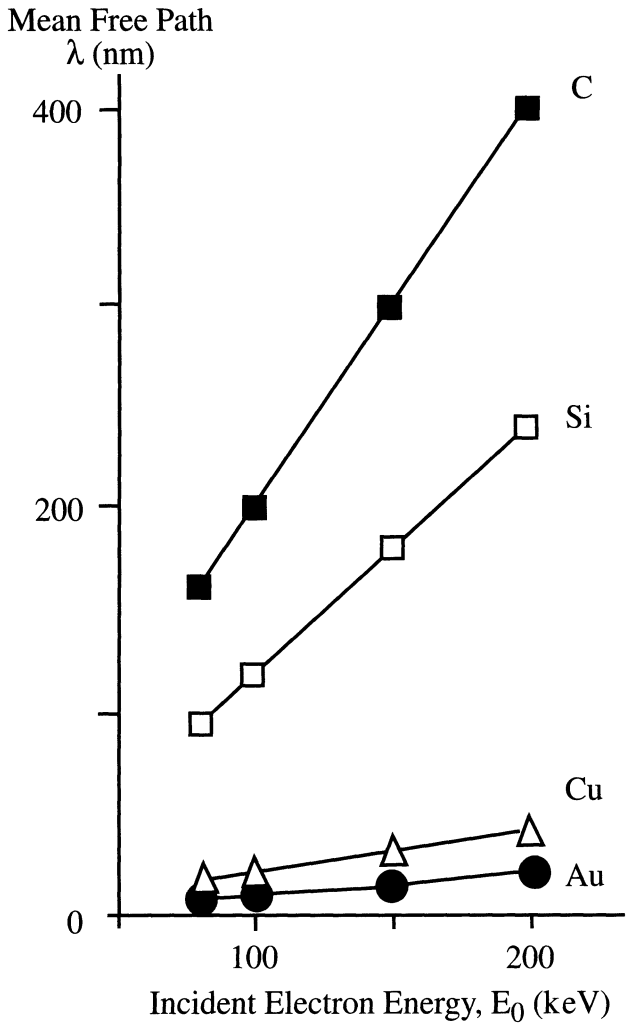


Figure 3.4. The variation of the mean free paths of elastic scatter for four different elements as a function of the beam energy, calculated assuming a screened, relativistic Rutherford cross section.

Rutherford-scattered electrons are *incoherent:* the phases of the electron waves are not in step. 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 to the value of *Z*, not the orientation of the specimen, as we'll see is the case for low-angle coherent scattering. Such *Z-contrast images,* as we'll see in Chapter 22, provide qualitative atomic resolution microanalysis in addition to showing atomic resolution detail at interfaces between regions of different *Z*. This is a relatively new imaging technique which may revolutionize our understanding of materials. Second, the high-angle backscattered electrons 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. Backscattered electron images are rarely used in the TEM because the backscattered 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, and only one of these escaped from the foil. Therefore, the quality of the signal is very poor and the images are noisy. This contrasts with bulk specimens in an SEM in which many electrons are backscattered (about 30% in Cu).

3.7. THE ATOMIC SCATTERING FACTOR

The classical 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 a cross section is the concept of the atomic scattering factor $f(\theta)$. The atomic scattering factor is related to the differential elastic scattering cross section, thus

$$|f(\theta)|^2 = \frac{d\sigma(\theta)}{d\Omega} \qquad [3.9]$$

What we will do is to highlight some of the important features that lead to this result by outlining the 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 you can appreciate why $f(\theta)$ is such an important parameter.

The scattering-factor approach is complementary to the Rutherford differential cross section, because it is most useful for describing the low-angle elastic scattering where the Rutherford model is inappropriate. Usually, the atomic scattering factor is defined in the following manner

$$f(\theta) = \frac{\left(1 + \frac{E_0}{m_0 c^2}\right)}{8\pi^2 a_0}\left(\frac{\lambda}{\sin\frac{\theta}{2}}\right)^2 (Z - f_x) \qquad [3.10]$$

where all the terms have been previously defined. If you need a more detailed approach you could consult the text by Reimer (1993). The wavelength is λ and f_x is the scattering factor for X-rays, which is well known. The best source of electron scattering factors is that due to Doyle and Turner

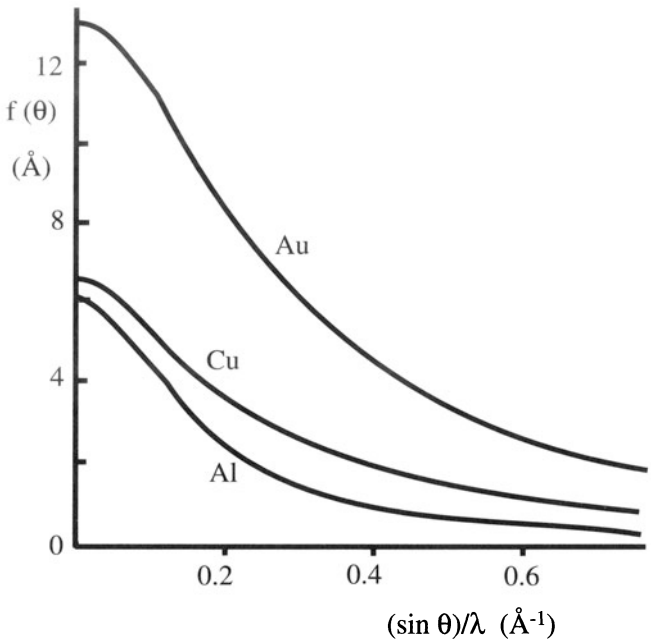


Figure 3.5. Change in the atomic scattering factor $f(\theta)$ with scattering angle θ (calculated from equation 3.10) showing that elastic scattering decreases with angle away from the incident beam direction ($\theta = 0°$) and increases with Z.

(1968), and you can also find values in the NCEMSS software (Section 1.5). The appearance of f_x in this formula is a reminder that $f(\theta)$ is a fundamental result of the wave nature of the electron.

> $f(\theta)$ 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 elastic scattering:

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

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

This expression (equation 3.10) 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 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.

3.8. THE ORIGIN OF $f(\theta)$

Since $f(\theta)$ relates to the amplitude of a scattered wave, we will briefly consider 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.

To find the total elastic scattering cross section, we have to integrate $d\sigma/d\Omega$. We note that this is a particle model, but we should note how the wave nature of the electrons is brought in. We can consider the wave nature by looking at Figure 3.6.

The incident beam can described as a wave of amplitude ψ and phase $2\pi kr$

$$\psi = \psi_0 e^{2\pi i kr} \qquad [3.11]$$

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 in Chapter 11. When it is scattered

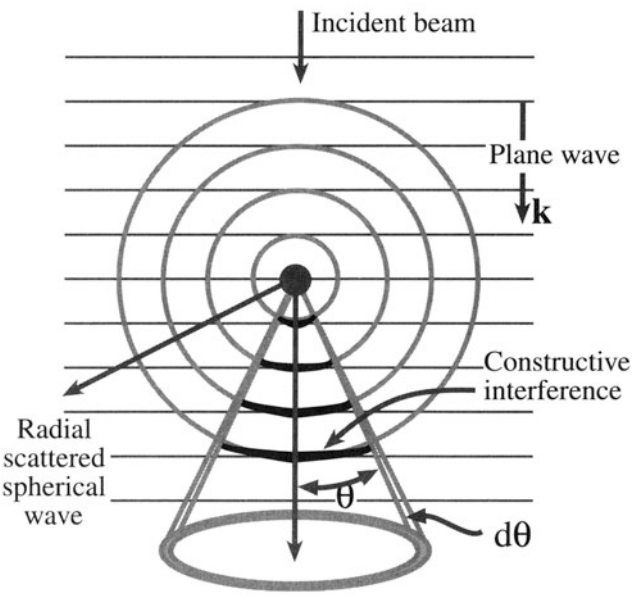


Figure 3.6. The generation of a scattered wave by the interaction of a plane wave (horizontal lines) with a point charge. The circles represent the scattered spherical wavefronts which are in phase (same λ). The in-phase constructive interference between the plane and spherical waves is shown by the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

by the point charge, a spherical scattered wave is created which has amplitude ψ_{sc} but the same phase

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

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

Now we need to know the atomic scattering amplitude. Up to this point, our treatment has been quite rigorous. So we need a model for $f(\theta)$ to make the problem manageable. Ideally, the model would distinguish between a neutral atom in a metal, a covalently bonded atom, and an ion. In practice, we usually use a simple approximation.

The quantity $f(\theta)$ can always, in principle, be calculated from the Schrödinger equation. If we write down the expression for the scattering process shown in Figure 3.6, then we have

$$\psi_{sc} = \psi_0 \left[e^{2\pi i \mathbf{k}_1 \cdot \mathbf{r}} + i f(\theta) \frac{e^{2\pi i k r}}{r} \right] \qquad [3.13]$$

Note 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 second, that $f(\theta)$ can be expressed as

$$f(\theta) = |f(\theta)| e^{i\eta(\theta)} = |f(\theta)| \left(\cos\eta(\theta) + i \sin\eta(\theta) \right) \qquad [3.14]$$

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

First aside: In writing equation 3.13, we have introduced two wave propagation parameters: the vector $\mathbf{k}_1$ 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.

Second aside: The 90° phase change for the scattered wave component in equation 3.14 can be easily understood by considering the following. If the amplitude of the wave is initially $\psi_0 \sin(2\pi k r)$, then after it has passed through the specimen it will be ψ_{tot}; after scattering, the phase is increased by ϕ. We can express the new ψ_{tot} as

$$\psi_{tot} = \psi_0 \sin(2\pi k z + \phi) = \psi_0 \sin(2\pi k z)\cos\phi + \psi_0 \cos(2\pi k z)\sin\phi \qquad [3.15]$$

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

$$\psi_{tot} = \psi_0 \sin(2\pi k z) + \psi_0 \phi \sin\left(2\pi k z + \frac{\pi}{2}\right) \qquad [3.16]$$

As Reimer notes, the $\pi/2$ would arise if we used the exponential rather than the sine to denote the phase, so we can then write equation 3.16 as

$$\psi_{tot} = \psi + i\psi_{sc} \qquad [3.17]$$

which we see then has the form given in equation 3.13.

3.9. THE STRUCTURE FACTOR $F(\theta)$

The next introductory step in discussing electron scattering is to take the idea of individual atoms scattering electrons (the atomic scattering factor), which we've just discussed in some detail, and consider what happens when the atoms are stacked together in crystals. We will deal with this point 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 it is an amplitude like $f(\theta)$ it also has dimensions of length. We can define $F(\theta)$ in terms of the sum of the atomic scattering factors from all the i atoms in the unit cell (with atomic coordinates x_i, y_i, z_i) multiplied by the phase factor that takes account of the difference in phase between waves scattered from atoms on different planes with Miller indices $(hk\ell)$. The scattering angle θ is the angle between the incident and scattered electron beams. So we can write

$$F(\theta) = \sum_i f_i e^{i\varphi_i} = \sum_i f_i e^{2\pi i (hx_i + ky_i + \ell z_i)} \qquad [3.18]$$

All this means is that the amplitude (and hence the intensity) of scatter 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 $(hk\ell)$ 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 scatter is zero, which is often a very useful diagnostic test when determining crystal structures in the TEM. We'll return to this 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 importance arises because, as we'll show you in Chapters 11 and 12, we can use diffraction to determine the spacing of planes in crystals. The interplanar spacings in different crystal structures are characteristic of that structure. As a result we can distinguish between different crystal structures by observing and measuring diffraction patterns. 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 how the diffraction process leads to contrast in TEM images which is related to the orientation of a crystalline specimen.

The combination of the diffraction pattern and the image is a most powerful tool for characterizing crystals and 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 similar plot as for 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, which we see as rings of intensity shown in Figure 3.7A, and in the diffraction pattern in Figure 2.11a. If the specimen is crystalline, then the intensity of the diffracted beams is a maximum at specific angles because the interplanar spacings are fixed (Figure 3.7B). The variation of $f(\theta)$ with θ in Figures 3.7A and B is equivalent to the intensity variation across the diffraction patterns in Figures 2.11a and b/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.

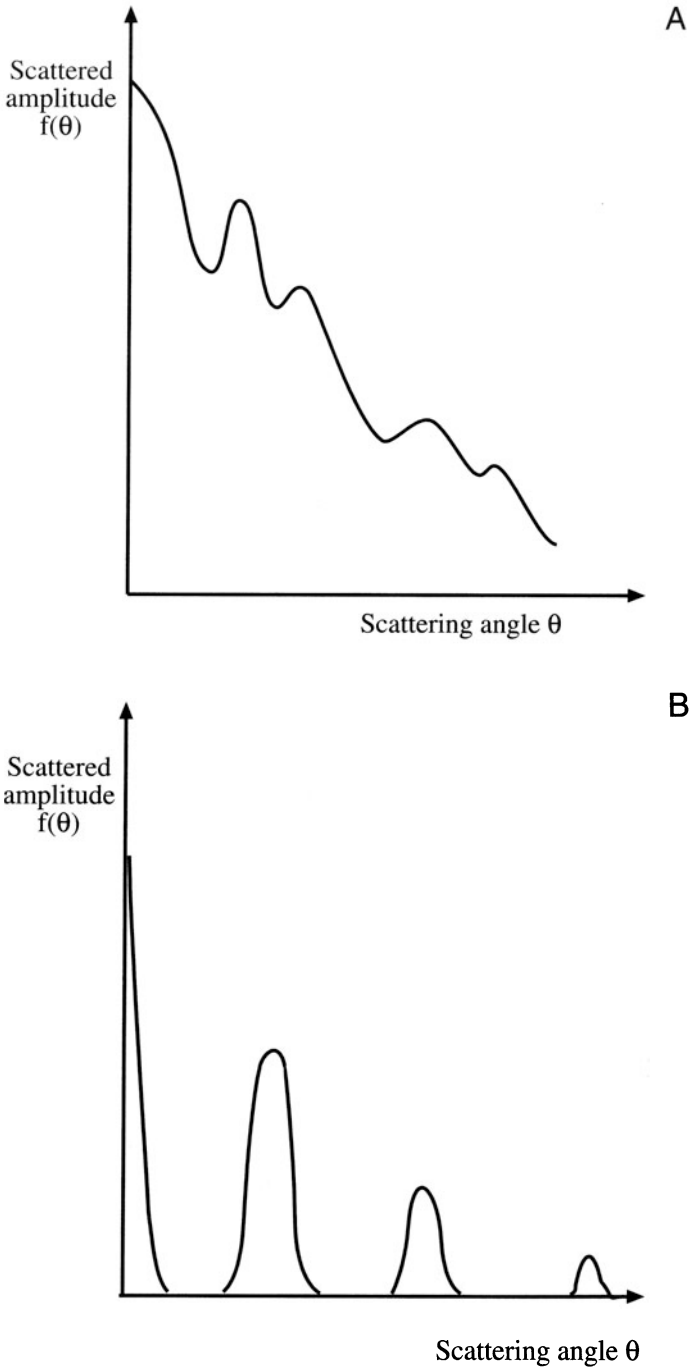


Figure 3.7. Change in $f(\theta)$ with θ (A) for an amorphous specimen and (B) for a crystalline specimen. The amplitude (and therefore the intensity) of scatter generally decreases with increasing θ, but the smooth decrease is modified at certain scattering angles (compare with the intensity variation along a line from the middle of the diffraction patterns in Figures 2.11a and b).

3.10.A. Interference of Electron Waves

To interpret this 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 interactions that characterize high-angle scattering. If you go back and look at Figure 3.2, you see a periodic one-dimensional array of scattering centers (slits); a monochromatic wave (fixed λ) is advancing toward the centers. Each slit 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, i.e., when they are coherent. 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* which proceeds in the same direction as the incident wave, which in the TEM we'll refer to as the direct beam of electrons. There are also *higher-order waves* which propagate in directions that are at some fixed 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, we'll find ways to measure these angles and relate them to the spacing of the scattering planes.

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 accredited to von Laue in Germany in 1912, although others such as Ewald were working on similar ideas at the same time. At von Laue's suggestion, his colleagues Friedrich and Knipping irradiated 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 which 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, $h\lambda$, and 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 then

$$a(\cos\theta_1 - \cos\theta_2) = h\lambda \qquad [3.19]$$

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

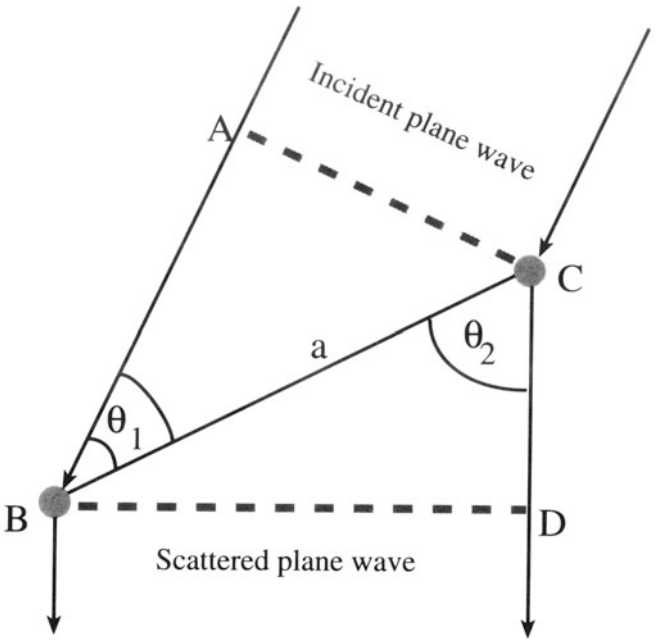


Figure 3.8. The approach used by von Laue to calculate the path difference for a wave (wavelength λ). In this one-dimensional figure the wave is incident at angle θ_1 and scattered at angle θ_2 from two atoms (B and C) spaced distance a apart. The path difference between scattered waves is AB – CD.

$$b(\cos\theta_3 - \cos\theta_4) = k\lambda \qquad [3.20]$$

$$c(\cos\theta_5 - \cos\theta_6) = \ell\lambda \qquad [3.21]$$

These three simultaneous equations bear von Laue's name, and for this work he received the Nobel Prize. If, in a TEM specimen, all three Laue equations are satisfied simultaneously, we will show in Chapter 11 that a diffracted beam is produced. We'll also show you in Chapters 11 and 12 that the letters $hk\ell$ are the indices of the diffracted beam and are equivalent to the Miller indices $(hk\ell)$ of a crystal plane, or some multiple thereof.

Usually in TEM, we use a simpler approach to describe diffraction. Von Laue's approach was simplified by Bragg (the elder) in England who argued that the waves behaved as if they were reflected off atomic planes as shown in Figure 3.9. Bragg showed 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 waves reflected from the upper and lower planes in Figure 3.9 is (AB + BC). Thus, if the "reflecting" $hk\ell$ 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 $2d \sin \theta_B$. Then we have Bragg's law

$$n\lambda = 2d \sin \theta_B \qquad [3.22]$$

We'll reserve θ_B for the Bragg angle, which is the most important scattering angle in TEM and you'll come across it many more times in this text. Bragg also received a Nobel Prize for this one equation, even though the idea of reflected electrons, while mathematically correct, is physically wrong! We'll continue to use the term "Bragg reflection" to describe diffraction because everyone does so, even though it's inaccurate. We will demonstrate, in a rigorous fashion, the equivalence of the Bragg and 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 we know λ for the incident electron and we can measure θ experimentally, we can work out the interplanar spacings in the crystal. It is this crystallographic information that makes diffraction such an important aspect of the TEM.

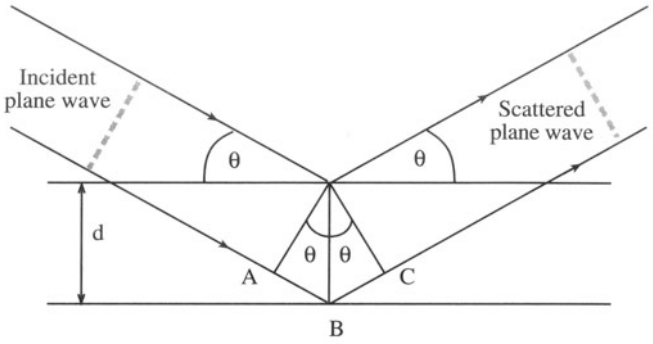


Figure 3.9. The Bragg description of diffraction in terms of the reflection of a plane wave (wavelength λ) incident at an angle θ to atomic planes of spacing d. The path difference between reflected waves is AB + BC.

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
$\frac{d\sigma(\theta)}{d\Omega}$	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.
- The relationships between $f(\theta)$ and the intensity in a diffraction pattern are 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.

Yes, a really rigorous treatment of scattering would take into account the wave nature of the electron (wave mechanics), relativity, and the electron charge at the same time. We won't do this! Fortunately we can do very well using compiled tables.

- The strength of the scattering, $f(\theta)$, depends inversely on the scattering angle, θ.
- $F(\theta)$ is a measure of the amplitude scattered by a unit cell.

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 is covalently bonded? We'll answer these questions as we go on.

REFERENCES

General References

Andrews, K.W., Dyson, D.J., and Keown, S.R. (1967) *Electron Diffraction Patterns,* The Institute of Physics, Bristol, United Kingdom.

Henoc, J. and Maurice, F., Eds. (1976) *Use of Monte Carlo Calculations*, NBS Special Publication 460, Washington, DC.

Kyser, D., Ed. (1984) *Electron-Beam Specimen Interactions for Microscopy, Microanalysis and Lithography,* SEM Inc., AMF O'Hare, Illinois.

Specific References

Doyle, P.A. and Turner, P.S. (1968) *Acta Cryst.* **A24**, 390.

Hall, C.E. (1953) *Introduction to Electron Microscopy,* p. 229, McGraw-Hill, New York.

Joy, D.C., Romig, A.D. Jr., and Goldstein, J.I., Eds. (1986) *Principles of Analytical Electron Microscopy,* p. 6, Plenum Press, New York.

Reimer, L. (1993) *Transmission Electron Microscopy,* 3rd edition, Springer-Verlag, New York.

Rutherford, E. (1911) *Phil. Mag.* **21**, 669.

Taylor, C. (1987) *Diffraction*, Adam Hilger, Bristol, United Kingdom.

Inelastic Scattering and Beam Damage

4

CHAPTER PREVIEW

In the previous chapter, we discussed elastic scattering of the electron beam in which the incident electron lost no energy as it interacted with the specimen. Inelastic or energy-loss electrons are equally important and we'll discuss the processes here, but leave the applications till later. Why are we interested in inelastic scatter? Well, inelastic scattering generates a whole range of signals, each of which can tell us more about the specimen than we can find out from the elastic electrons. The most important signals are the X-rays, inelastic electrons, and secondary electrons, and so we'll emphasize how these signals arise. We will also discuss why these specific signals are useful to materials scientists.

So how do we use these other signals? First we have to detect the electrons and X-rays and we'll describe electron detection in Chapter 7. In Chapter 31 we will explain how we use some of the signals to form images of the specimen. We will discuss how to detect X-rays and get information from the spectra that are created in Chapters 32–36. Then in Chapters 37–40 we'll talk about detecting and analyzing the electrons that lose energy when they are scattered in the specimen. In all cases we get complementary information to that

gained in conventional TEM images and diffraction patterns. Obviously then 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 the specimen which can damage beam-sensitive specimens. So we must also look at the down side of the signal-generating process and we end the chapter by discussing this problem under the general topic of beam damage or radiation damage.

A warning: This chapter contains some quite difficult theoretical concepts. However, it does form the basis of AEM, which is the topic of much of Part IV of the book. You can safely delay studying much of this material in detail until you reach Chapter 32.

Inelastic Scattering and Beam Damage

4

4.1. WHICH INELASTIC PROCESSES OCCUR IN THE TEM?

Historically, the conventional TEM only used two *elastic* signals, namely, the direct beam and the diffracted beams. As we've seen, these signals constitute the diffraction pattern 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 those were shown back in Figure 1.3. These signals are sometimes sought 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 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 signal constitute analytical electron microscopy (AEM), which we discuss in detail in Part IV of this book. 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 analytical TEMs exist which, in one form or another, can detect all the signals shown in Figure 1.3. So in this chapter we'll cover all the signals that are detectable and what use (if any) they are to the materials scientist. We need to know:

- What are the inelastic scattering interactions?
- What is the energy loss associated with each process?
- What is the likelihood that each energy-loss process will occur?

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

A general rule of thumb: The deeper the electron penetrates into the atom, the higher 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.

We know the first two rather well, but the third is usually poorly defined. Figure 4.1 shows the cross sections for some of the more important inelastic processes that we'll talk about. 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 scatter in more detail as we describe each individual inelastic event.

4.2. X-RAY EMISSION

We'll consider X-ray emission first because it's the most important secondary signal generated in the specimen.

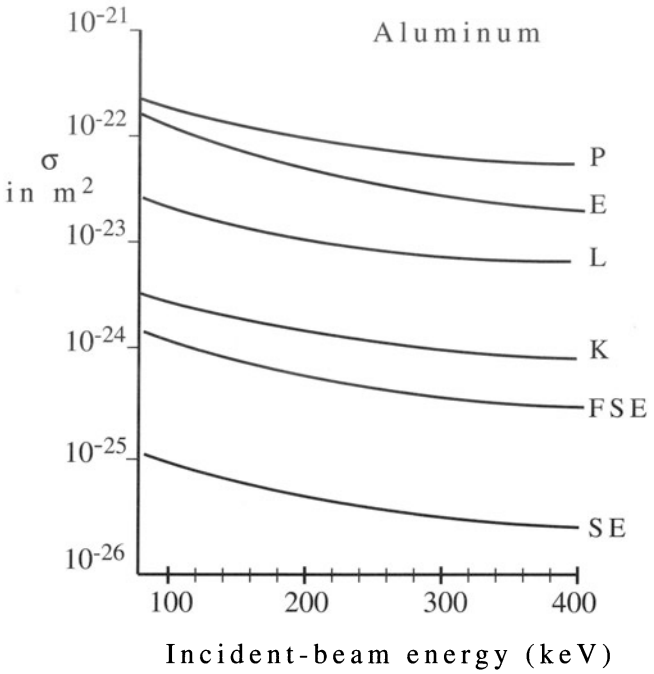


Figure 4.1. Cross sections for the various inelastic scattering processes in Al as a function of the incident electron energy, assuming a small angle of scatter ($\theta \sim 0°$); plasmon (P), K and L-shell ionization (K, L), fast and slow secondary electron generation (FSE, SE). For comparison purposes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

From X-rays we can find out easily 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 IV.) Two kinds of X-rays are produced:

- Characteristic X-rays which are useful to the materials scientist.
- Bremsstrahlung X-rays which are useful to the biologist but generally regarded as a nuisance by most materials scientists.

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 electron shells 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; that is, it escapes the attractive field of the nucleus, leaving a hole in the inner shell. When this happens the atom is left in an excited state because it has a higher 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 missing electron with one from the outer shells. It is this transition which is accompanied either by the emission of an X-ray or an Auger electron. This latter process was first described by the Frenchman Pierre Auger in 1925 and won him the Nobel Prize for Physics. Since the discoverer was French, we pronounce his name to sound like "Ogay" with a soft *g* as in "beige." In either case the energy of the emission is *characteristic* of the difference in energy of the two electron shells involved and so 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 4.7.

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 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 auger electron spectrometer (AES). More recently, however, we've found ways to detect the Auger signal in ultra-high vacuum (UHV) TEMs and so we'll discuss it in Section 4.3.C below.

You need to know several aspects of the ionization process to understand why the characteristic X-ray signal is so useful and what it takes to generate it:

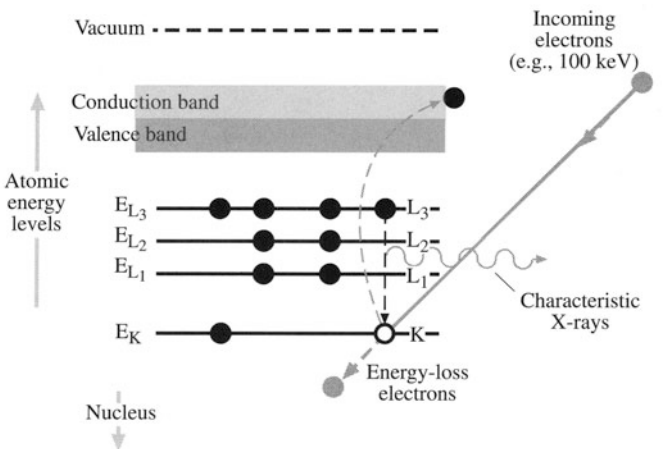


Figure 4.2. The ionization process. An inner (K) shell electron is ejected from the atom by a high-energy electron. When the hole in the K shell is filled by an electron from the L shell, characteristic K_α X-ray emission occurs. The beam electron loses energy but continues on through the specimen.

- What are *electron shells?*
- 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, and so on; we used this terminology in Figure 4.2. All the shells (except the K shell) may themselves have subshells. We name the characteristic X-rays in terms of the shell being filled and the shell from which the electron comes. (The K, L, etc. terminology was first introduced by Charles Barkla, an early X-ray spectroscopist. 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 in 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 more complex because we differentiate the α X-rays in terms of α_1 and α_2, depending from which subshell of the outer shell the electron falls 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. But for X-ray detection in the TEM you don't need to worry about too many details because, as you'll see later, we 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 need to remember. Much more detail is given in books on X-rays and X-ray spectrometry, e.g., Williams (1990).

Not all electron transitions are equally probable and this is taken into account by the "weights" of the lines which are given in Table 4.1. These weights are only important within a given K, L, or M family and not between families, because experimental conditions affect each family differently. In microanalysis we only use the most intense lines, usually the α lines (or, if we can't resolve them, we use the α and β lines). This will become more obvious when you've learned about X-ray qualitative analysis in Chapter 34.

Critical ionization energy: The electron beam has to transfer an amount of energy greater than a critical value to the inner-shell electron to ionize the atom. This energy is called the critical ionization energy (E_c); if we're going to generate X-rays, then the beam energy E_0 must be greater than E_c. The value of E_c increases as the electrons are more tightly bound to the nucleus, so the innermost shell (K) has a higher E_c than the L shell, and so on. Atoms with higher Z have more protons and therefore have a higher E_c. You can see this effect if you go and look at Figure 1.4, 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. For a complete list you have to find a reference text such as Bearden's Tables (Bearden 1964). Such a list is also invaluable in EELS since the E_c values correspond to the positions of peaks in the energy-loss spectrum which, as we'll see in Chapter 38, can be used to identify uniquely the presence of a particular ionized atom in the specimen.

The cross section for ionization (σ) 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. What we have to take into account, however, is a parameter called the overvoltage, U, which is the ratio of the beam energy E_0 to E_c. The cross section varies with U as shown in Figure 4.4, and what this figure tells you is that if E_0 is close to E_c then there isn't much chance of ionization. Usually in the TEM, E_0 is ≥ 100 keV and E_c is < 20 keV, so U is greater than 5 and the ionization cross section is pretty constant. Despite this 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:

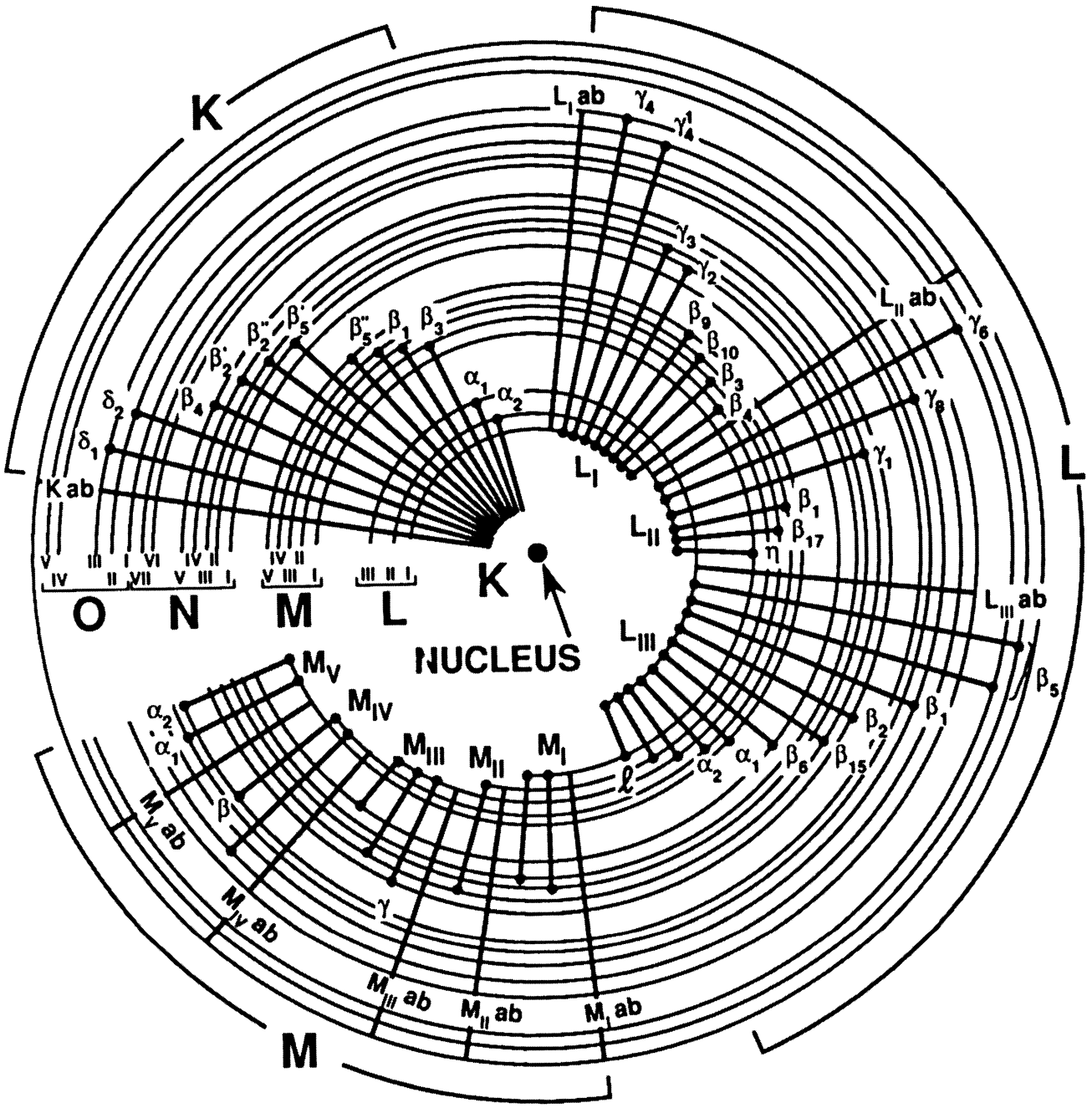


Figure 4.3. The complete range of possible electron transitions that give rise to K, L, and M characteristic X-rays. Not all these X-rays are detectable by EDS in the TEM.

- The electron that ionized the atom is deviated only through a small angle (<~10 mrads).
- The resultant characteristic X-ray is emitted uniformly over 4π sr.

As with the Rutherford cross section, the simple Bethe expression needs to be corrected for the effect of relativity at TEM electron energies, and this means substituting the term $m_0v^2/2$ for the beam energy and introducing a standard relativistic factor, β ($=v/c$)

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

Table 4.1. Weights of Lines (Approximate)[a]

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)		

[a]The weights are given in parentheses.

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, and Powell (1976) gives a good review.

The X-ray energy/wavelength: X-rays are electromagnetic radiation and so we usually think of them 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 should 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; it 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 electron volts we equate it to E, where E is the X-ray energy.

Thus

$$E = h\nu = \frac{hc}{\lambda} \qquad [4.3]$$

Now since h and c are constants we can substitute, and we find that

$$\lambda = \frac{1.24}{E} \qquad [4.4]$$

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 in eV) that we derived back in Chapter 1 and you can easily confuse the two, so beware!

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. 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, H. Moseley (1914). 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

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

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 out in small "slide rules" by the manufacturers of X-ray spectrometers.

If you compare E_c and the X-ray energies you'll see that they are not identical. The X-ray energy E_K or E_L is invariably less than E_c. This is 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 shell. This hole must also be filled by another electron, and so on, until eventually a free electron from the conduction or valence band fills the last hole in one of the inner shells. So the atom returns to ground state by a cascade of transitions, depending on the complexity of the electronic structure of the atom.

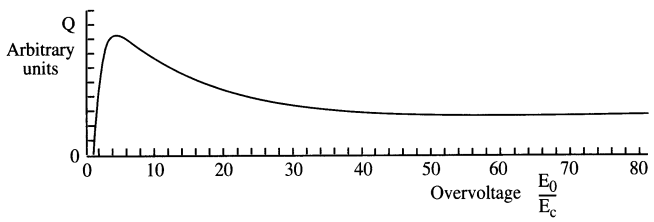


Figure 4.4. The variation of the ionization cross section with overvoltage. Ionization is most probable if the beam energy is ~5× the critical ionization energy. The cross section decreases, but not substantially, at higher overvoltages, typical of a TEM.

An example: A Cu K shell electron requires 8.98 keV of energy for ionization (E_c = 8.98 keV). One possible sequence by which this energy 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). The X-ray energies therefore total 8.978 keV and the remaining 2 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 a phonon (see below).

The possible variations are enormous and affected by such things 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 another 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, but for any masochists among you the book by Dyson (1990) goes into the explicit details of this complicated subject (and our knowledge isn't complete by any means).

Fluorescence yield: Remember that an ionized atom does not have to lose energy by giving off a characteristic X-ray but can also emit an Auger electron. 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 approximate expression for ω gives

$$\omega = \frac{Z^4}{a + Z^4} \qquad [4.6]$$

where $a \approx 10^6$ for the K shell. This is only an approximation but is still a formidable dependence on Z. It predicts that, for carbon (Z = 6), ω is ~ 10^{-3} and, for Ge (Z = 32), ω is ~ 0.5. This means you have to ionize 1000 carbon atoms before you get a single C K_α X-ray generated but only 2 atoms for Ge. 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 detect light elements; you should use EELS (see Part IV).

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 is decelerated by the Coulomb (charge) field of the nucleus, it emits an X-ray. Since the electron can suffer any amount of deceleration depending on the strength of its interaction, then these X-rays can have any energy up to the beam energy. Such X-rays produced as the electron decelerates 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 (1923). This expression is often used for thin TEM specimens, although it was originally derived for bulk specimens. It is common to use the Kramers cross section to predict the bremsstrahlung production rather than the probability of interaction. Thus

$$N(E) = \frac{KZ(E_0 - E)}{E} \qquad [4.7]$$

where $N(E)$ is the number of bremsstrahlung photons of energy E, produced by electrons of energy E_0; K is the

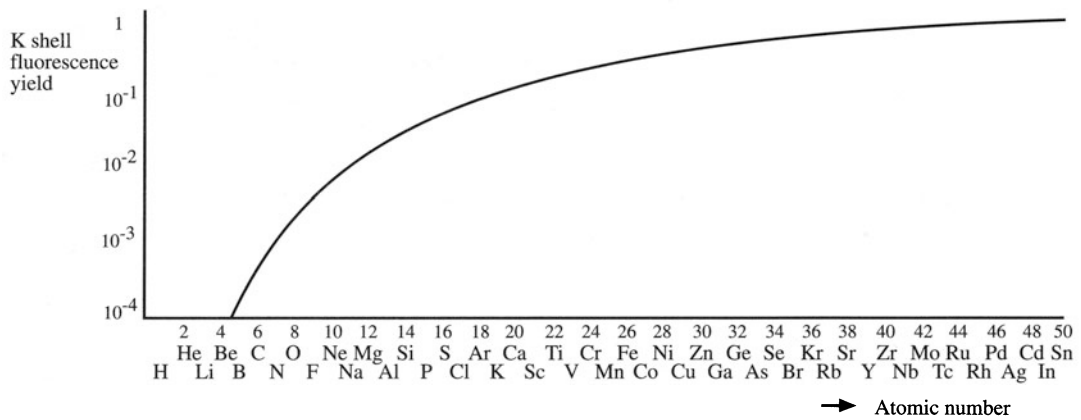


Figure 4.5. Fluorescence yield for K shell X-rays as a function of atomic number. Note the rapid decrease at low atomic numbers. X-rays from elements below Be are undetectable.

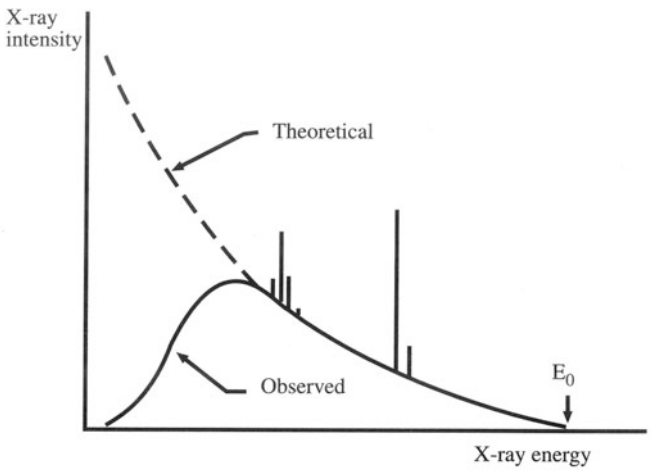


Figure 4.6. The bremsstrahlung X-ray intensity as a function of energy. The generated intensity increases rapidly at low energies but very low energy bremsstrahlung is absorbed in the specimen and the detector, so the observed intensity drops to zero. E_0 is the energy of the electrons that cause the X-ray emission. Two families of characteristic lines at specific energies are also shown superimposed on the bremsstrahlung.

Kramers constant and Z is the atomic number of the atom. This relationship predicts that it is far more likely that the braking event 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 is a function of energy as shown in Figure 4.6. In contrast to the isotropic emission of the characteristic X-rays, the bremsstrahlung is highly anisotropic, showing strong forward scatter which increases as E_0 increases.

The bremsstrahlung has a continuous energy spectrum on which the characteristic X-rays we just talked about are superimposed, as also shown schematically in Figure 4.6 and realistically in the experimental spectrum back in Figure 1.4. Since the characteristic X-rays have a narrow energy range, they appear as peaks in the spectrum at specific energies. The bremsstrahlung intensity depends on the average Z of the specimen and this is most useful to biologists who are interested in this aspect of their specimen. But materials scientists generally dismiss the bremsstrahlung as a useless background signal which only obscures characteristic lines. We'll come back to the X-ray spectrum in more detail in Chapter 32.

4.3. SECONDARY ELECTRON EMISSION

Secondary electrons (SEs) are electrons in the specimen that are ejected by the beam electron. They can be discussed as three distinct groups:

- If the electrons are in the conduction or valence bands then it doesn't take much energy to eject them, and they are called "slow SEs" with energies typically below about 50 eV.
- If the electrons are strongly bound inner-shell electrons they are less readily ejected, but when they are thrown out of their shells they can have a significant fraction (up to about 50%) of the beam energy, and they are then called "fast secondary electrons," or FSEs.
- If the electrons are ejected from an inner shell by the energy released when an ionized atom returns to ground state, then these secondary electrons are called Auger electrons. The process is often termed a "nonradiative transition" and energy undergoes "internal conversion" (which is not quite a religious experience).

Until quite recently SEs were only considered in relation to SEM, where they are used to form the images which are so sensitive to surface topography. We'll now discuss each of these signals and their relative importance in the TEM.

4.3.A. Slow Secondary Electrons

Slow 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. But 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 finding increasing use in STEMs, where they 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 31.

SE images in a STEM have much better resolution than SE images in low-kV SEMs.

We'll discuss several reasons for this in Chapter 31. Recent developments in high-resolution field emission gun (FEG) SEMs have produced SE image resolution better than 1 nm at 30 kV, and a STEM at 100 kV can offer simi-

lar or better resolution even without an FEG, so the slow SEs are very useful. (We discuss FEGs in Chapter 5.)

The number of slow SEs with energies >~50 eV is close to zero and rises to a maximum at about 5 eV. 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 SE 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 reflected light images we are accustomed to seeing with our eyes.

4.3.B. Fast Secondary Electrons

Fast secondary electrons (FSEs) are high-energy electrons which are generated in the specimen; they are high-energy because they receive a large fraction of the beam energy (Joy 1984). From the cross-section data in Figure 4.1 you can see that they should be an order of magnitude more probable than slow SEs. At the low beam energies we use in an SEM, FSEs aren't a problem, so nobody bothers about them. However, in a TEM, FSEs can have energies of ~50–200 keV, in which case they not only travel significant distances within the specimen, but they may also escape from deep within the specimen. As a result, FSEs degrade the spatial resolution of microanalysis in AEMs and they also generate significant numbers of X-rays which can cause problems in quantifying X-ray data, particularly at intermediate voltages. So FSEs aren't an image resolution problem, but rather a problem for chemical analysis.

> FSEs are generally both unavoidable and undesirable. We don't use them to form images or to give us spectroscopic data, but they may degrade the quality of the latter.

This phenomenon is only just beginning to be understood, but it may well turn out to be a major limitation of intermediate voltage microanalysis.

4.3.C. 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 ground state. Figure 4.7 shows how an ionized atom ejects an outer-shell (Auger) electron, and it's instructive to compare with Figure 4.2 for

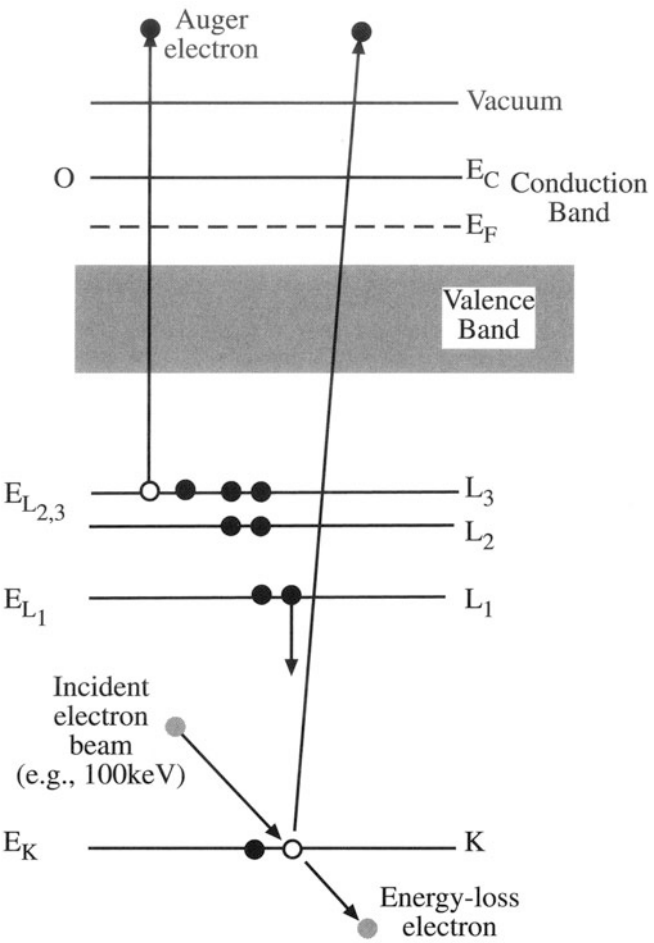


Figure 4.7. The process of inner (K) shell ionization and subsequent Auger electron emission. The energy released when the L_1 electron fills the hole in the K shell is transferred to an electron in the $L_{2,3}$ shell which is ejected as a $KL_1L_{2,3}$ Auger electron.

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. So the Auger electron has a characteristic energy which is dependent on the electronic structure of the ionized atom and is almost identical to the energy of the alternative characteristic X-ray.

> The Auger process is favored in atoms with a small binding energy, 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.

The Auger electrons that do escape come from very close to the surface. Consequently they contain surface chemical information and AES is a recognized surface

chemistry technique. Now you might ask why light-element X-ray analysis in the TEM is not just a surface technique, because of the similarity in energy between Auger electrons and characteristic X-rays. 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.

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 common, the Auger signal is of more interest and a few microscopists are reporting combined Auger/STEM results. However, since it is not simple to attach an Auger system to a STEM these instruments are still quite 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 which 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.

If you don't do anything, the electrons and holes will recombine, and in doing so give off light; this process is termed cathodoluminescence (CL).

The process is shown schematically in Figure 4.8. The photon has a frequency equal to the energy of the gap (E_G) divided by Planck's constant (h), and so if the band gap varies for some reason there will be a spectrum of light given off, or 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.

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 under the internal bias. You can pick up the signal if you ground the specimen through a picoammeter. In this situation, the specimen is acting as its own detector! The current you then detect is sometimes called the "electron beam induced current" or EBIC signal, and if you detect it and use it to form an image then you are doing "charge-collection microscopy" or CCM.

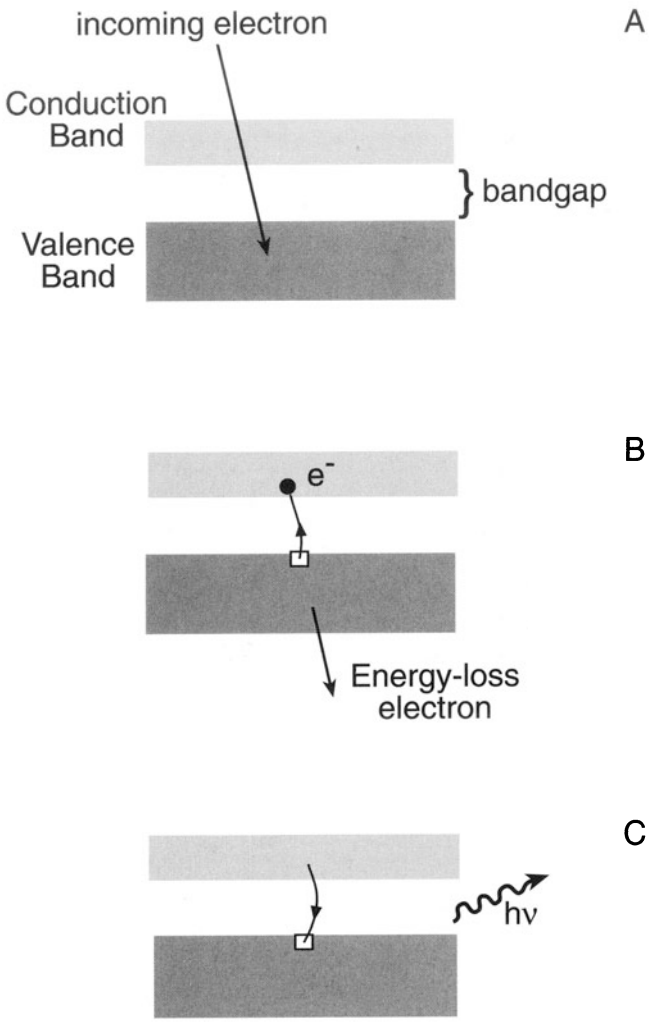


Figure 4.8. Schematic illustration of CL: (A) Initial state before a beam electron interacts with valence-band electrons. (B) A valence-band electron is excited across the gap into the conduction band, leaving a hole in the valence band. (C) The hole is filled by a conduction-band electron falling back into the valence-band hole. Upon recombination a photon of light is emitted, with a frequency determined by the band gap.

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, in general, all these imaging modes are rare and mainly limited to studies of semiconductors (although some minerals also exhibit CL). We'll describe CL detectors in Chapter 7 and the images in Chapter 31. Just remember that CL and CCM are powerful, but rather specialized techniques.

4.5. PLASMONS AND PHONONS

We can link these two phenomena because they are both examples of what we call "collective oscillations."

> Plasmons are collective oscillations of free electrons that occur when the beam electron passes through the free electron "gas."

We can consider plasmons as analogous to sound waves, since they are longitudinal oscillations 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 less than ten nanometers. If you go back to Figure 4.1 you'll see that the plasmon process has the largest cross section and it's by far the most common inelastic interaction occurring in materials. 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 thus a high free-electron density. The plasmon oscillation is quantized and the mean free path for plasmon excitation is of the order of 100 nm. As we'll see in Chapter 39, this makes the number of plasmon excitations a useful way to measure the specimen thickness. Also, the plasmon energy is a function of the free-electron density and this changes with composition, so the plasmon excitation process is chemically dependent, although we rarely use it for microanalysis.

A differential cross section for plasmon excitation was given by Ferrel (1956)

$$\frac{d\sigma_\theta}{d\Omega} = \frac{1}{2\pi a_0}\left(\frac{\theta_E}{\theta^2 + \theta_E^2}\right) \qquad [4.8]$$

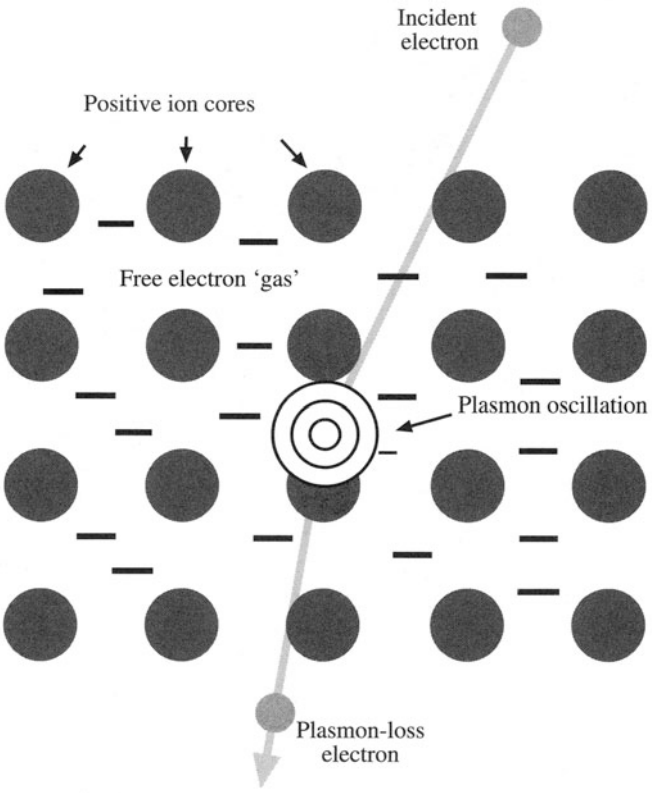


Figure 4.9. Schematic diagram of a high-energy beam electron exciting a plasmon oscillation in a free electron gas that permeates the ion cores in a metal.

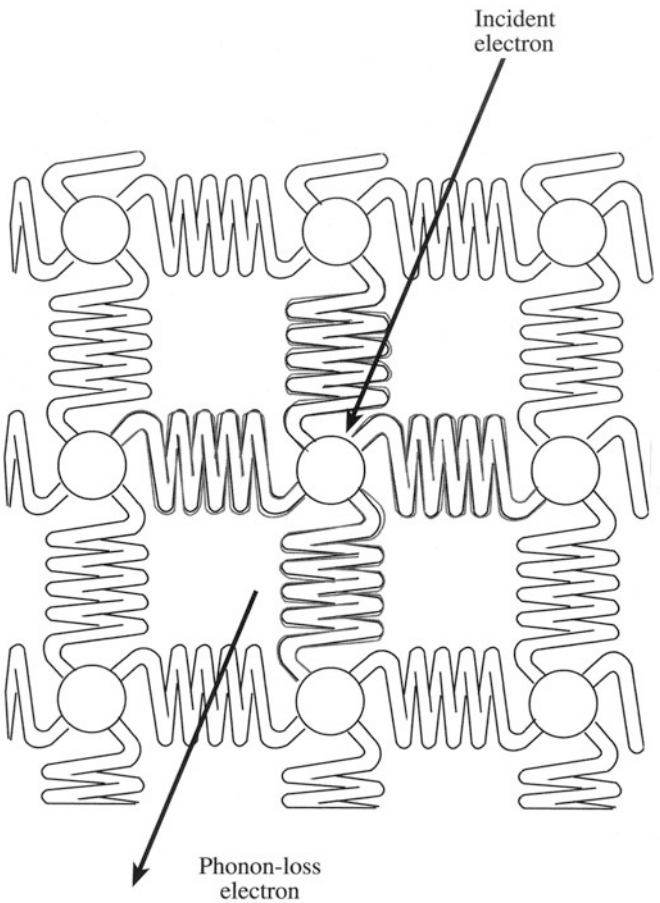


Figure 4.10. An illustration of the crystal lattice as a group of atoms linked elastically by springs. The bonds vibrate when struck by a high-energy electron creating lattice oscillations or phonons. The lattice absorbs heat by creating phonons, so phonon excitation is equivalent to heating the specimen.

where a_0 is the Bohr radius, θ is the scattering angle, and θ_E is the so-called "characteristic scattering angle" given by $E_P/2E_0$. 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 the 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, we can visualize the atoms as all linked 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 interband transitions 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 sample, even amorphous ones in which a crystal "structure" as such does not exist. Typically, a phonon vibration causes a very small energy loss of <0.1 eV but the phonon scattered electrons are scattered out to quite large angles (5–15 mrads), and these electrons account for the diffuse background intensity present between the Bragg spots in diffraction patterns. Phonon scattered electrons carry no useful microchemical information, nor do they carry contrast useful to the microscopist.

The phonon scattering cross section is not important to know exactly, but it is useful to remember that phonon scattering increases with Z with a dependence of approximately $Z^{3/2}$, which is rather less strong than for true elastic scattering. Also, because of the effect of temperature on atomic vibration, the phonon scatter is increased 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 clear diffraction patterns. The mean free path for phonon scatter 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 ~ 2–3×.

Phonons are oscillations where all the atoms in the crystal lattice vibrate collectively. Such vibrations of the atoms are equivalent to specimen heating. You can reduce the number of phonons by cooling the specimen.

We don't use either plasmons or phonons directly to form images, but we do detect the electrons that caused them, and we'll discuss the (rather limited) uses of plasmon energy-loss electrons in Chapter 40.

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, that of 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 mainly on the beam energy. Certain materials are more susceptible than others, but in the end, you can damage virtually anything that you put into the TEM. 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 our specimen changes it. Once the structure or the chemistry is changed, the specimen is not representative of the parent material and interpreting TEM images, diffraction patterns, and spectra becomes more difficult and eventually impossible. 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 use electron damage to emulate other forms of radiation damage. Generally, however, beam damage is undesirable.

Damage takes one of two principal forms:

- *Radiolysis:* Inelastic scattering (mainly ionization) breaks the chemical bonds of certain materials such as polymers and alkali halides.
- *Knock-on damage:* Direct displacement of atoms from the crystal lattice creates point defects.

We will see that, paradoxically, the former is reduced at higher beam energies while the latter is increased, so there is sometimes no way around the problem.

Phonons represent heat in the specimen and heat is a major source of damage to polymers. Electron–electron interactions can give rise to chemical bonding changes through *radiolysis;* this process is common in polymers and alkali halides. Atomic displacement is termed "knock-on damage" within the specimen or "sputtering" if it occurs at the surface of the thin foil, and these processes are ubiquitous if E_0 is high enough. 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 is describe the fundamental processes in different materials, explain how you can determine if your specimen is being damaged, and how you can minimize or eliminate the problem. First, however, we need to know the terms we use to measure beam damage.

4.6.A. Electron Dose

In the TEM we define the electron dose as the charge density (Cm^{-2}) hitting the specimen. It is easy to convert this to the number of electrons/unit area (usually e/nm^2) knowing that $e = 1.6 \times 10^{-19}$ 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 dose is defined by the Gray (Gy), which is the absorption of 1 joule 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 in the TEM are well above lethal doses for human tissue. While this is another warning about the dangers inherent in the 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 experimental variables that can affect the result, such as the thermal conductivity, thickness, and surface condition of the specimen and the beam size, energy, and current. Hobbs (1979) 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.

In addition to these practical steps, beam heating is 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.

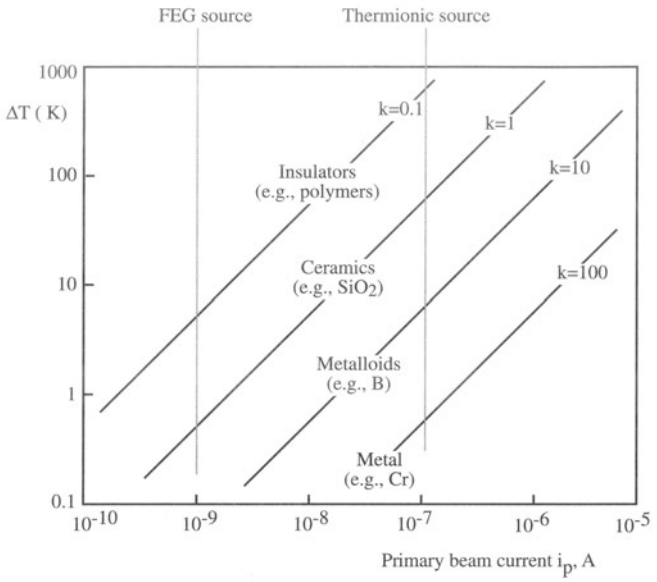


Figure 4.11. The increase in specimen temperature as a function of the beam current and the thermal conductivity k, in $Wm^{-1}K^{-1}$ of the specimen. Typical materials are noted, but should not be considered representative, since k varies substantially in any class of materials.

Beam heating for metals is negligible under standard TEM conditions, but if thermal conduction is poor, then heating can be quite substantial. Small ceramic particles may be heated by the beam to temperatures of ~1700°C.

4.6.C. Beam Damage in Polymers

Polymers are most sensitive to the electron–electron interactions which, by one means or another, generate phonons or lattice vibrations. These phonons heat the specimen and break the bonds, creating new structures. This process is called 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 cross link to form a new structure.

A break formed this way in the polymer chain is known as scission. Generally, polymers show a tendency either to break down or to cross link. In the former case the polymer will continue to lose mass under irradiation, while in the latter the polymer eventually becomes mainly a mass of carbon. Mass loss can sometimes be measured directly in the TEM by electron spectrometry, and it can also manifest itself as major dimensional changes in the specimen.

Mass loss ultimately results in a hole forming in the illuminated area of the specimen; the image contrast will usually change before the hole appears!

If the polymer was crystalline originally, then radiation damage results in a loss of crystallinity, and you can measure this quantitatively either from the loss of diffraction contrast in the image or the loss of intensity in the diffraction pattern and the gradual appearance of an amorphous pattern. Sometimes the crystal structure can be preserved by staining. However, whenever you stain a specimen you affect its structure and mask the chemistry, so this isn't ideal.

There are several methods you can use to minimize beam damage in polymers (Sawyer and Grubb 1987):

- Use low-dose imaging techniques (see Chapter 30).
- Operate at the highest kV.
- Cool the specimen to liquid-N_2 temperatures or lower.
- Coat the specimen with a conducting metal film.
- Use STEM imaging (Section 22.3).
- Do all of the above if necessary.

4.6.D. Beam Damage in Covalent and Ionic Crystals

In covalent and ionic materials such as ceramics and minerals, radiolysis can occur which changes the specimen chemistry and possibly the structure through a series of reactions driven by the electron beam. The major inelastic interaction is that of interband transitions similar to those responsible for CL. The interband 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, can create an anion vacancy and a cation interstitial. In a similar process crystalline quartz can be amorphized. Often the process can result in the formation of new compounds which can be identified by electron diffraction and AEM. The formation of Ag from Ag halides in the photographic plate is an example of radiolysis.

We can't stop radiolysis simply by cooling or coating the specimen, since it isn't affected by heat transfer considerations. We can use higher voltages to lower the cross section for the electron–electron interactions. The best way is to use both higher voltages and thin specimens. Nevertheless, radiolysis remains a major limitation when looking at certain ceramics and minerals, and many polymers in the TEM.

4.6.E. Beam Damage in Metals

The primary way that metals are affected is by knock-on or displacement 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).

Knock-on damage is directly related to the beam energy.

How strongly the atoms are bonded to their neighbors will also be a factor. A simple expression given by Hobbs (1979) 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 5–50 eV, but varies with bonding type. If we assume that a typical value of E_d is ~25 eV, we can determine the threshold potentials for a range of elements from Figure 4.12. From this figure it is quite evident that if you have a 400-kV intermediate voltage TEM, you can displace atoms with atomic weight below about Ti. 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, or those with particularly strong covalent bonds such as diamond. The only way to avoid displacement damage is to operate below threshold.

How can you identify displacement damage? It usually manifests itself as small vacancy clusters which appear as black-white lobe contrast or dot contrast as we showed back in Figure 1.8, or sometimes damage is discernible as dislocation loops. 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. So depending on your specimen there may in fact be no way to avoid damage of one form or another in the TEM.

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

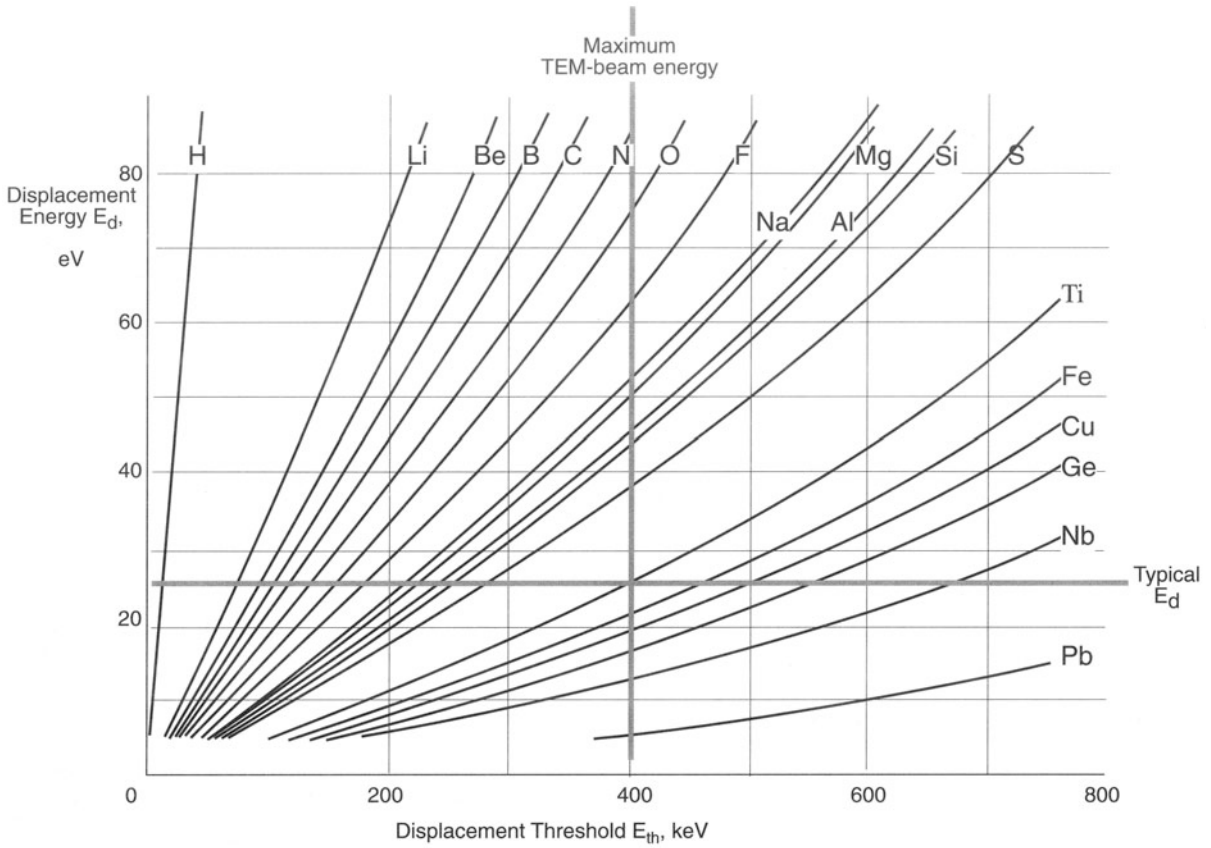


Figure 4.12. The displacement energy for a range of atoms as a function of the threshold energy (i.e., the beam energy) required for displacement damage. In a typical material E_d is ~15–25 eV, but it can vary substantially with bond strength.

Table 4.2. Comparison of Maximum Transferable Kinetic Energy (T) with Displacement and Sputtering Energies at 100, 200, 300, and 400 kV (from Zaluzec and Mansfield 1987)

	T (eV)				E_d (eV)	E_s (eV)
Element	100 kV	200 kV	300 kV	400 kV		
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
Pt	1.23	2.69	4.37	6.26	33	8–16
Au	1.22	2.67	4.32	6.2	36	9–18

occurring in nuclear reactors. A general rule of thumb was that a few minutes' exposure in an HVEM was equivalent to many years in a nuclear reactor and so accelerated studies of materials degradation 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 have seriously reduced the number of such studies, but if you want to find out about it there are reviews in the literature, such as Laidler and Mastel (1975).

Vacancies caused by displacement damage can enhance diffusion processes, and this in turn can speed diffusional transformations when they're being studied *in situ* in the HVEM. There are many other problems that can arise when doing this, and other *in situ* observations, so interpretation isn't always straightforward. The book by Butler and Hale (1981) is recommended for more facts.

4.6.F. Sputtering

The displacement of surface atoms, or sputtering, occurs in the TEM, at voltages which are about 50% less than knock-on thresholds. If your specimen is quite thick then this problem is minor, but often the specimen has to be very thin if you want the best images and the best microanalytical resolution. In these circumstances sputtering may substantially change the surface chemistry of the specimen and affect quantitative microanalysis. Table 4.2 lists typical sputtering threshold energies (E_s) compared with displacement thresholds (E_d(eV)) and, as you can see, there is cause for concern even at 100 kV.

CHAPTER SUMMARY

Inelastic scatter transfers energy to the specimen, generating a lot of useful signals which we can use to form different images of the specimen or get spectroscopic information about its chemistry and electronic structure. Unfortunately, the same processes transfer heat to the specimen which can be disastrous for certain materials such as polymers. To minimize heat transfer, higher voltages should be used, but eventually knock-on and sputtering damage occur which create defects and change the surface chemistry of all materials.

REFERENCES

General References

Inokuti, M. (1971) *Rev. Mod. Phys.* **43**, 297.

Mott, N.F. and Massey, H.S.W. (1965) *The Theory of Atomic Collisions,* Oxford University Press, New York.

Wang, Z.L. (1995) *Elastic and Inelastic Scattering in Electron Diffraction and Imaging*, Plenum Press, New York. Covers more than inelastically scattered electrons using a more rigorous mathematical approach than in this chapter.

Specific References

Auger, P. (1925) *J. Phys. Rad.* **6**, 20.

Bearden, J.A. (1964) NYO-10586, US Atomic Energy Commission, Oak Ridge, Tennessee.

Bethe, H.A. (1930) *Ann. der Phys. (Leipzig)* **5**, 325.

Butler, E.P. and Hale, K.F. (1981) "Dynamic Experiments in the Electron Microscope," in *Practical Methods in Electron Microscopy,* **9** (Ed. A.M. Glauert), Elsevier, Amsterdam.

Dyson, R. (1990) *X-rays in Atomic and Nuclear Physics,* Cambridge University Press, Cambridge, United Kingdom.

Ferrel, C.R. (1956) *Phys. Rev.* **101**, 554.

Hobbs, L.W. (1979) in *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 437, Plenum Press, New York.

Joy, D.C. (1984) *J. Microsc.* **136**, 241.

Joy, D.C., Goldstein, J.I., and Romig, A.D. Jr., Eds. (1986) *Principles of Analytical Electron Microscopy,* p. 1, Plenum Press, New York.

Kramers, M.A. (1923) *Phil. Mag.* **46**, 836.

Laidler, J.J. and Mastel, B. (1975) in *Physical Aspects of Electron Microscopy and Microbeam Analysis* (Eds. B.M. Siegel and D.R. Beaman), p. 103, Wiley, New York.

Moseley, H.E.G. (1914) *Phil. Mag.* **26**, 1024.

Powell, C.J. (1976) *Rev. Mod. Phys.* **48**, 33.

Sawyer, L.C. and Grubb, D.T. (1987) *Polymer Microscopy,* Chapman and Hall, New York.

Williams, K.L. (1990) *X-ray Spectrometry,* Allen and Unwin, London.

Zaluzec, N.J. and Mansfield, J.F. (1987) in *Intermediate Voltage Microscopy and Its Application to Materials Science* (Ed. K. Rajan), p. 9, Philips Electron Optics Publishing Group, Mahwah, New Jersey.

Electron Sources

5

CHAPTER PREVIEW

All microscopes need a source of electrons to illuminate the specimen. 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 for the beam of electrons and these are best met by only two types of source: thermionic and field-emission sources. Thermionic sources are either tungsten filaments or lanthanum hexaboride (LaB_6) crystals, and field emitters are fine tungsten needles. In this chapter we'll first explain briefly the physics of these two 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 three sources and show you that no one source is best for all aspects of TEM, but all three have their roles. Finally, we'll explain ways to check that a particular source meets your specification.

Because the source is so critical to the performance of the microscope, the technology is advancing rapidly to the point of complete computer control, which would leave you, the operator, with precious 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 vast majority of TEMs still use thermionic sources, and these 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 most of this chapter talking about thermionic sources, although there's a good chance that field emission will be the source of choice in the future.

Electron Sources 5

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 an intense electric field is applied to it. 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. However, from a practical standpoint, you don't have to know too much of the physics, and we can summarize the essential points very briefly, using a few simple equations. Keep in mind two points as you read about sources:

> Your TEM will use a thermionic source or a field-emission source and the two cannot be interchanged.
> Field-emission sources give "monochromatic" electrons; thermionic sources are less monochromatic and give "whiter" electrons.

The analogy here is to X-rays or visible light. You don't always want to use "monochromatic" electrons, even if the 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. This barrier is termed the "work function" (Φ) and has a value of a few electron volts.

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

$$J = AT^2 e^{-\frac{\Phi}{kT}} \qquad [5.1]$$

where k is Boltzmann's constant (8.6×10^{-5} eV/K) and A is Richardson's "constant" ($A/m^2 K^2$), which depends on the source material. From this equation then you can see that we need to heat the source to a temperature T such that energy greater than Φ 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 work function. In practice we use both types: tungsten has the necessary high melting temperature (3660 K) and lanthanum hexaboride (LaB_6) has a low work function. If you look ahead to 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 sometimes call tungsten sources "filaments," because tungsten can be drawn into fine "thread" which is about 0.1 mm in diameter and is similar to the filament used in an incandescent light bulb. The wire is bent into a V shape so they're also called "hairpin" filaments, or they may be sharpened to a fine point. For decades these have been the standard source in most electron-beam instruments. LaB_6, or other rare-earth boride 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.

Thermionic sources:	W hairpin Pointed W LaB_6 and other low-Φ materials, e.g., CeB_6

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, and we achieve this by operating under a condition called "saturation," which we'll discuss in Section 5.3.A.

5.1.B. Field Emission

Field-emission sources operate on a fundamentally different principle than thermionic sources. The principle behind field emission 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

$$E = \frac{V}{r} \qquad [5.2]$$

The technique of field-ion microscopy is another well established experimental tool. It requires specimens with a very fine needle shape, and so there's a lot of expertise available to help produce field-emission electron sources. One of the easiest materials to produce with a fine tip 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. This process imposes quite severe stress on the tip and the material has to be strong. Field emission, like thermionic emission from LaB_6, depends on the crystallography of the tungsten tip; the <310> orientation is found to be best.

To allow field emission, the surface has to be pristine, that is, free of contaminants and oxide. We can achieve this by operating in UHV conditions (<10^{-11} Torr), and in this case the tungsten is operated at ambient temperatures and the process is called "cold field emission." 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 so much that, in fact, the electrons don't tunnel through the barrier. For such "thermal field emission," surface treatments with ZrO_2 improve the emission characteristics, particularly the stability of the source, and such "Schottky" emitters are becoming popular. New sources such as semiconductor p-n field emitters are also provoking some interest.

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 terms 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 microscope and you'll see that there's no "best" source for all applications, but for specific applications one source or another is clearly the best.

Before we define the electron beam characteristics needed in a TEM, it is worth summarizing here a few of the properties of electron beams in general and how these vary with accelerating voltage.

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.

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 angles, so we can't ignore the angular distribution of the electrons. Brightness is particularly important when we are using very fine electron beams, as we do in analytical and scanning microscopy. 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 diffraction patterns.

So we can consider an electron source as having the following characteristics:

- a diameter d_0,
- giving off a certain cathode emission current i_e,
- the electrons diverging from the source with a semiangle α_0.

The actual way in which this is achieved we'll talk about in Section 5.3, where we discuss the complete gun assembly, but if you look ahead 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 (current per unit area) is $i_e / \pi(d_0/2)^2$ and the solid angle of the source is $\pi\alpha_0^2$, so we define the brightness, β, as

$$\beta = \frac{i_e}{\pi\left(\frac{d_0}{2}\right)^2 \pi(\alpha_0)^2} = \frac{4\, i_e}{(\pi\, d_0\, \alpha_0)^2} \qquad [5.3]$$

This equation is an important one which you should remember. The units of β are usually A/(cm^2 sr) or A cm^{-2} sr^{-1}. Again we see that microscopists are not comfortable using SI units, which would be A m^{-2} sr^{-1}, increasing the traditional brightness number by a factor of 10^4. What is not shown in this equation is the important fact embodied in equation 5.1 that β increases linearly with increasing accelerating voltage for thermionic sources. This is 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 an electron beam of a given size, and so the more information we can extract from the specimen and the more we can damage sensitive specimens. The beam current is an important part of the brightness equation. Having some way of 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 the source brightness, but for the time being you can again look ahead to Table 5.1 to see how the three sources compare in brightness, which we have given in non-traditional SI units.

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 at a maximum. 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. The output of a typical electric power generating turbine is only about 600 MW. Clearly, we can change our specimen when we look at it in the TEM as we discussed in relation to beam damage. The energy density we just calculated means that an electron source is the brightest continuous radiation source known; it is considerably brighter than a supernova.

The brightness is particularly important in AEM, which is the technique of quantitative analysis of the many signals that come from a specimen irradiated by an electron beam, shown back in Figure 1.3. Similarly, as we go to higher magnifications in HRTEM, the screen intensity becomes less because we are viewing only a fraction of the illuminated area of the specimen. The electron density can be increased by using the brightest available source. Then images can be recorded with reasonably short exposure times.

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 wavelengths (colors), and so to get a coherent beam of electrons we must create one in which all the electrons have the same 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 they are all identical they have the same coherence length. A definition of the coherence length λ_c is

$$\lambda_c = \frac{v\, h}{\Delta E} \qquad [5.4]$$

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's sometimes called for historical reasons) so that all the electrons have a small ΔE, thus giving a well-defined wavelength. Now in practice it's impossible to create a truly monochromatic beam and we have to live with a certain range of electron energies/wavelengths, although the stability of electronic components has improved substantially over the years. Again if you look at Table 5.1 you'll see typical ΔE values for the three sources and they're in the range 0.1 to 3 eV (which is remarkably small compared with a total energy of 100 to 400 keV). So it isn't really correct to imply that thermionic sources give "white" electrons since ΔE is still 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-tension power supplies, this rarely limits any aspect of TEM except perhaps high energy-resolution electron spectrometry (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.

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 source sizes 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

$$d_c << \frac{\lambda}{2\alpha} \qquad [5.5]$$

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 field-emission source.
- 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.

Spatial coherency is more important practically than temporal coherency. 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. The more coherent and parallel the beam is, the better the quality of the phase-contrast images (see Part III), the sharper the diffraction patterns (see Part II), and the better the diffraction contrast in images of crystalline specimens (see Part III). An in-depth and rather mathematical description of coherency in the TEM is given in the review by Hawkes (1978).

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 will vary, making it difficult for you to take correctly exposed images, and also making microanalysis impossible in many cases. 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 less than 1% per hour in the current. For cold field emission sources, however, the emission current is not very stable, and electrical feedback circuits are required to maintain stability to better than 5%. Stability does improve with better UHV conditions.

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 off the source. The design of the gun is different for thermionic sources and field-emission sources.

5.3.A. Thermionic Guns

Both tungsten and LaB_6 sources are 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-tension cable, which in turn connects to the high-

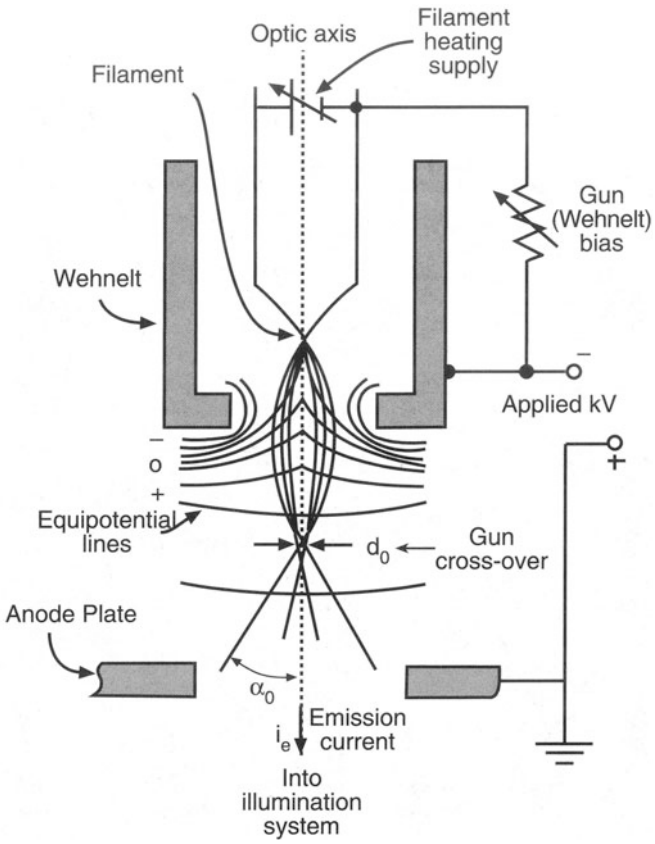


Figure 5.1. Schematic diagram of a thermionic electron gun. A high voltage is placed between the filament and the anode, modified by a potential on the Wehnelt which acts to focus the electrons into a crossover, with diameter d_0 and convergence/divergence angle α_0.

voltage power supply. This cable also connects to the tungsten filament to supply a current to heat the filament resistively to the operating temperature. LaB_6 sources are indirectly heated usually by bonding them to a metal filament such as rhenium, which is resistively heated.

As the filament current (i_f) increases the temperature increases until thermionic emission occurs, and an emission current from the cathode i_e can be measured. Sometimes you'll find this current referred to as the "beam current," but this is misleading, because 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 microscope.

When the electrons leave the cathode they have a negative potential of 100 kV with respect to the earthed anode, so they accelerate through this potential difference acquiring an energy of 100 keV, and a velocity of 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 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 cathode heating and the Wehnelt bias controls independently, but the electronic circuitry of the gun is designed so that as the emission current increases the Wehnelt bias increases, and this arrangement is called a "self-biasing" gun. The result is shown in Figure 5.3, which plots the filament emission current (i_e) against the current used to heat the filament (i_f). As you can see, i_e reaches a maximum such that further increase in i_f doesn't increase the current going into the microscope. This is the *saturation condition* and all thermionic sources should be operated at or just below saturation. Operating above saturation reduces filament life without any compensating ad-

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 onto the cathode (filament) support and both are attached to the high-tension cable which contains power supplies for heating the filament and biasing the Wehnelt. The anode sits just below the Wehnelt, in the top of the TEM column.

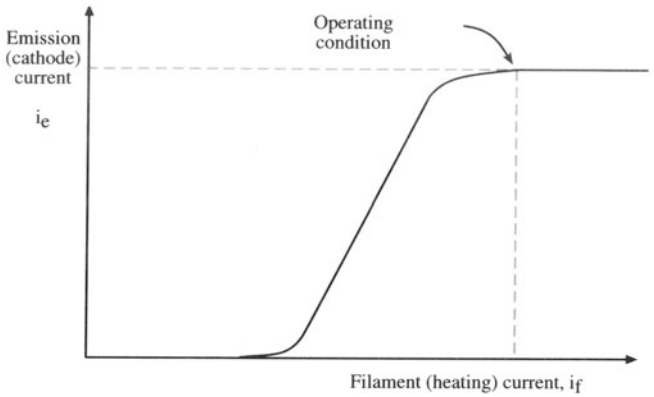


Figure 5.3. The relationship between the current emitted by the electron source (i_e) and the filament heating current (i_f) for a self-biasing gun. Increasing the filament current results in a maximum emission current termed saturation.

vantage; operating significantly below saturation reduces the current into your specimen, thus reducing the intensity of all the signals coming out of your specimen.

The Wehnelt acts as a simple electrostatic lens: the first lens in the microscope.

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

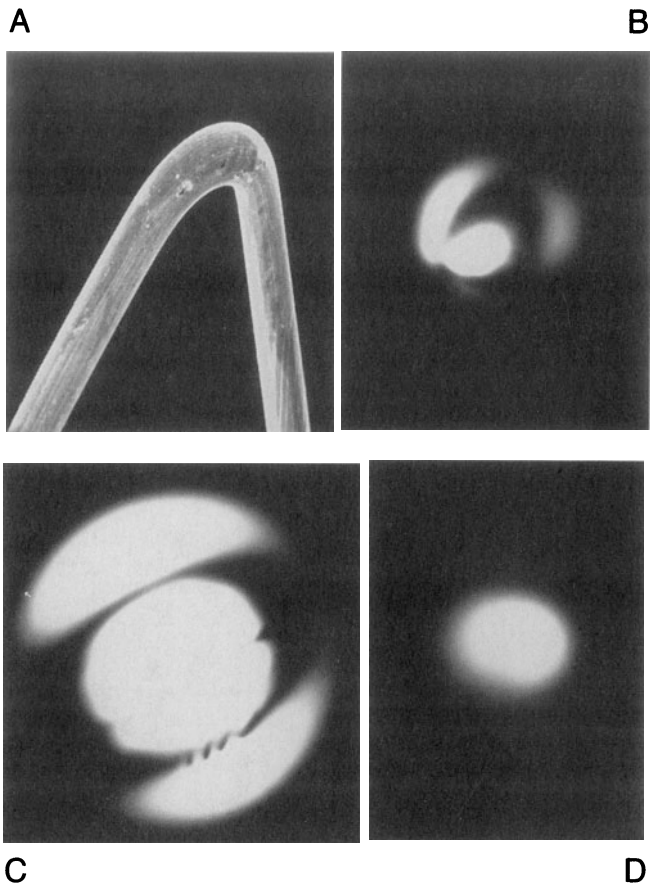


Figure 5.5. (A) The tip of a tungsten hairpin filament and the distribution of electrons when the filament is (B) undersaturated and misaligned, (C) undersaturated and aligned, and (D) saturated.

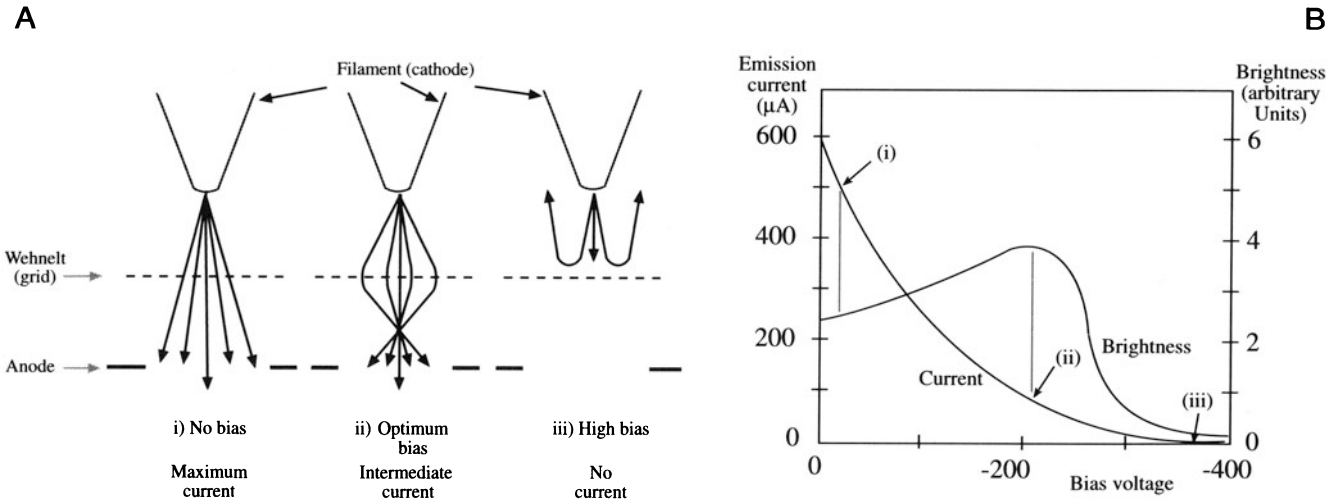


Figure 5.4. (A) The effect of increasing Wehnelt bias (i–iii) on the distribution of electrons coming through the anode. (B) The relationship between the bias and the emission current/gun brightness. Maximum brightness is achieved at an intermediate Wehnelt bias, and an intermediate emission current [condition (ii) in A].

back in the brightness equation (equation 5.3) and the convergence/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 (i) d_0 would not be very small, and if the bias were too high (iii) the cathode emission current would be suppressed. In either case β would be low. The optimum β is at an intermediate bias setting (ii), as summarized in Figure 5.4B. You might think that the small bias on the Wehnelt acts against the accelerating voltage, so the true beam voltage is 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 how do we achieve saturation? One way is to look at the meter which displays the 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 filament crossover on the TEM screen; this image shows you the distribution of electrons coming off the filament. As thermionic emission starts the electrons come from both the central tip of the filament and a region surrounding the tip (Figure 5.5A), and so the filament image is as shown in Figure 5.5B or C, and is characteristic of an unsaturated tungsten filament. With increasing emission the halo of emission collapses in on the central bright disk, although some structure may still be visible. The filament is truly saturated when no structure is visible (Figure 5.5D).

Since LaB_6 sources have well-defined crystal facets (Figure 5.6A) they show a slightly different undersaturated image, as you'll see in Figure 5.6B, but in essence the process is identical. It is probably best to operate an LaB_6 source at conditions just below saturation, since this will extend the source life without undue loss of signal. We'll find that there are a few occasions when undersaturated operation can be useful, because the electrons in the halo are more coherent than those in the central bright region. LaB_6 crystals are more susceptible to thermal shock than tungsten, and so you should take care when heating *and cooling* an LaB_6 source. Increasing the heating current should be done slowly, with 10 to 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, such as we show in Figures 5.5 and 5.6, can also be used to align the gun assembly so that the beam is aligned along the optic axis of the microscope. This is the only other thing 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, the undersaturated source image is asymmetrical as in Figure 5.5B

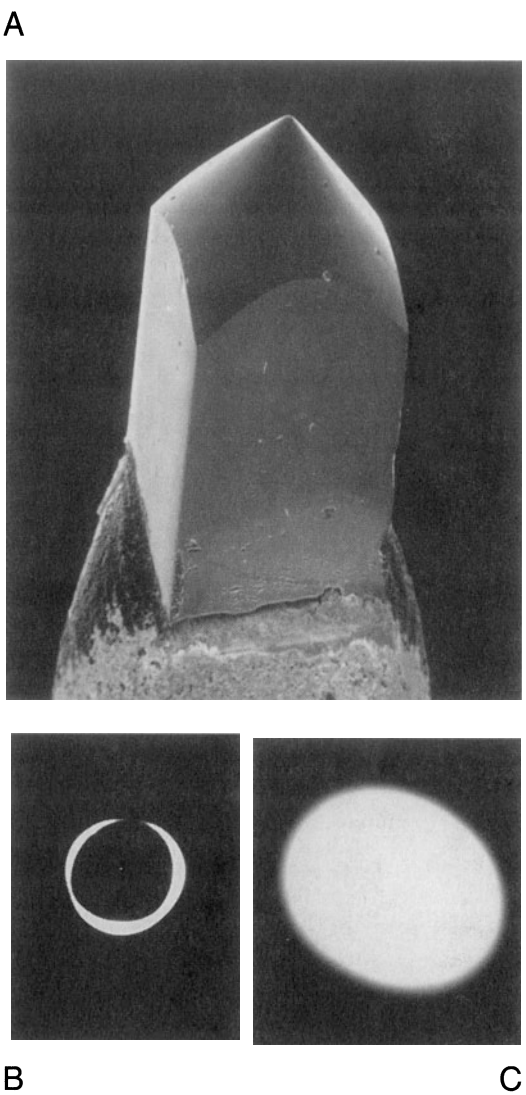


Figure 5.6. (A) An LaB_6 crystal and the electron distribution when the source is (B) undersaturated and aligned and (C) saturated.

and in those circumstances all you have to do is tilt the gun assembly to make it symmetrical as in Figure 5.5C. Detailed instructions will be in the manufacturer's handbook.

> Achieving optimum β is critical in any operations that require a fine beam (<0.1 μm).

In an SEM, which always requires a small probe, the gun is carefully adjusted by the manufacturer to produce optimum β at saturation, and 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 "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 is not important if you're operating with a broad beam, but 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 that 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 an 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 is called the "extraction voltage" since it generates the intense electric field-extracting electrons by enabling them to tunnel out of the tip. Increasing the extraction voltage when you first switch on has to be done slowly, so the mechanical shock doesn't fracture the tip. This is the only practical step you have to carry out to run an FEG, and it can easily be 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 applied potential by the second anode. The combined fields of the anodes act like a more refined electrostatic lens to produce a crossover, as shown in Figure 5.7A. This 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.

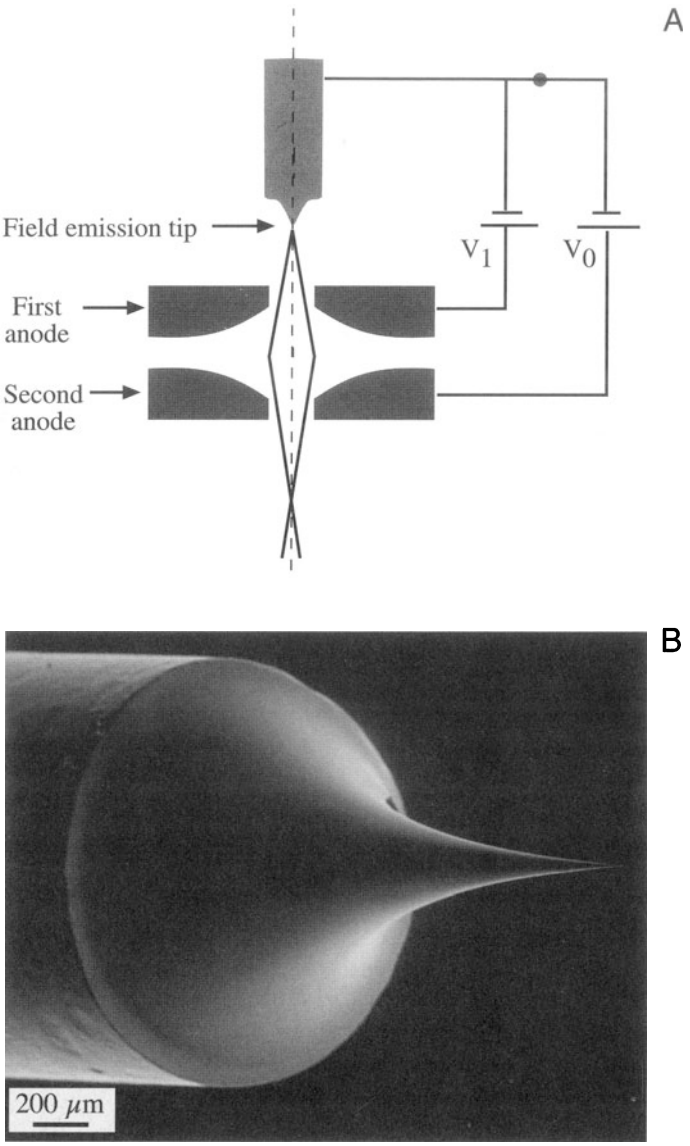


Figure 5.7. (A) Electron paths from a field-emission source showing how a fine crossover is formed by two anodes acting as an electrostatic lens. Sometimes an extra (gun) lens is added below the second anode. (B) An FEG tip, showing the extraordinarily fine W needle.

In a vacuum of 10^{-7} Torr, one monolayer of contaminants will form on a substrate in less than a minute. At 10^{-10} Torr, it will take 7 hours to form a monolayer.

We have already noted that field emission 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 has to be increased to compensate. But eventually it becomes necessary to remove the contamination by "flashing" the tip. This just means reversing the potential to the tip and "blowing off" a surface layer of atoms, and/or heating the tip quickly to ~5000 K to evaporate the contaminants. In most FEGs flashing occurs automatically, when the extraction voltage increases to a certain predetermined level. Thermally assisted FEGs do not form the same surface contamination

layer and so don't need flashing. A typical FEG tip is shown in Figure 5.7B.

5.4. COMPARISON OF GUNS

All the important characteristics of the three guns we've talked about are summarized in Table 5.1. Tungsten sources are the worst in most respects, but for routine TEM applications they are excellent, reliable sources and are cheap, robust, and easily replaceable.

LaB_6 is a 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. The brightness is typically 10 times that of tungsten, 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 can be as little as 1 eV.

The drawback to LaB_6 is purely economic. LaB_6 sources cost several hundred dollars each while tungsten filaments are so cheap that the manufacturer often provides them free. Because LaB_6 is a highly reactive material, the gun vacuum has to be 10–100 times better than for tungsten, and is correspondingly more expensive to construct. So if the cost is not the criterion, LaB_6 guns are *the* recommended thermionic source, for all aspects of TEM, but particularly AEM. The increased brightness, higher coherency, and longer life are tremendous advantages. But you as the operator have the most control over its performance and you can most easily destroy it by careless heating and cooling and oversaturation. So treat LaB_6 sources gently and you will be well rewarded. If users are not careful, your TEM supervisor may try to 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 the tungsten and LaB_6 sources, β increases linearly with kV, so there are advantages to using 300 and 400 kV instruments, although the thermionic source brightness at 400 kV still does not approach β of an FEG at 100 kV. The extremely small source size means that the beam is highly spatially coherent and the resulting energy spread is minuscule for cold FEGs; thermally assisted FEGs give a larger energy spread. 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!

For routine TEM, an FEG is far from ideal because the source size is so small. It is thus not possible to illuminate large areas of the specimen without losing current density, and therefore intensity, on the screen. Under these circumstances, a thermionic source is better. This limitation to FEG applications may be overcome by the larger *p-n* FE sources, which use small (≈ 1–10 μm) Si semiconductor crystals, but this is still a new and developing technology.

Another drawback to FEGs is the need for UHV conditions. UHV technology is expensive and requires a much higher level of operator competence. As a result, FEG TEMs are relatively rare. But in the SEM field there is a whole new generation of computer-controlled low-voltage instruments, and it will only be a matter of time before FEG TEMs are common.

TABLE 5.1. Characteristics of the Three Principal Sources Operating at 100 kV

	Units	Tungsten	LaB_6	Field Emission
Work function, Φ	eV	4.5	2.4	4.5
Richardson's constant	A/m^2K^2	6×10^5	4×10^5	
Operating temperature	K	2700	1700	300
Current density	A/m^2	5×10^4	10^6	10^{10}
Crossover size	μm	50	10	<0.01
Brightness	A/m^2sr	10^9	5×10^{10}	10^{13}
Energy spread	eV	3	1.5	0.3
Emission current stability	%/hr	<1	<1	5
Vacuum	Pa	10^{-2}	10^{-4}	10^{-8}
Lifetime	hr	100	500	>1000

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 gun and then ignore it. 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 poor. Similarly, knowing the energy spread of your source is important for electron spectroscopy, 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 semiangle of convergence, we can determine β. However, while we can easily get a measure of the emission current at the gun, it is more difficult to measure d_0 and α_0 there. 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.

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., about 50 μm) and the metal block deep enough (about 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 should be available permanently in the column of a TEM, and this would permit constant monitoring of the beam current. You can also calibrate the Faraday cup measurement against the TEM screen exposure meter or the electron energy-loss spectrometer shield current. This procedure permits you to make a more rapid estimate of i_b at any time you need it.

As we'll show in Chapter 9, 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.

- The beam current is usually in the range from nanoamps to picoamps.
- The emission current is typically several microamps.

Most of the current from the gun is lost in the illumination system, as we'll see in Chapter 9.

5.5.B. Convergence Angle

You can easily measure the convergence semiangle α from the convergent-beam diffraction pattern, which you can see directly on the TEM screen. (You will need to read Chapter 21 on convergent-beam diffraction in order to find out how to generate such patterns.) In the schematic diagram in Figure 5.8, the total convergence angle 2α is proportional to the width of the diffraction disks, 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 disk to the $hk\ell$ disk. Thus

$$2\alpha = 2\theta_B \frac{a}{b} \qquad [5.6]$$

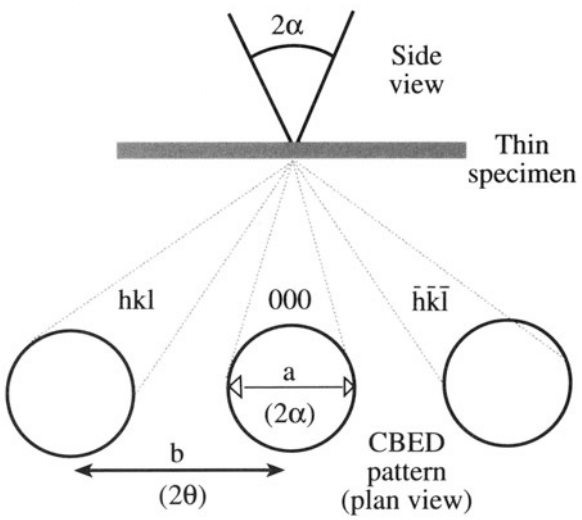


Figure 5.8. The distances on a convergent-beam diffraction pattern from which you can measure the beam-convergence semiangle, α, which is proportional to the width of the diffraction disk.

The convergence semiangle is not only important in the brightness equation, but we'll see that it also plays a major role in convergent-beam patterns, in STEM imaging, and in EELS. Knowledge of α is useful 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 where we use a fine focused beam, such as AEM and STEM imaging. We can either calculate d or measure it experimentally.

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 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 effects will then broaden the beam size beyond a true Gaussian. If you select too small an aperture, then the intensity distribution will be truncated at a fraction of the full Gaussian curve.

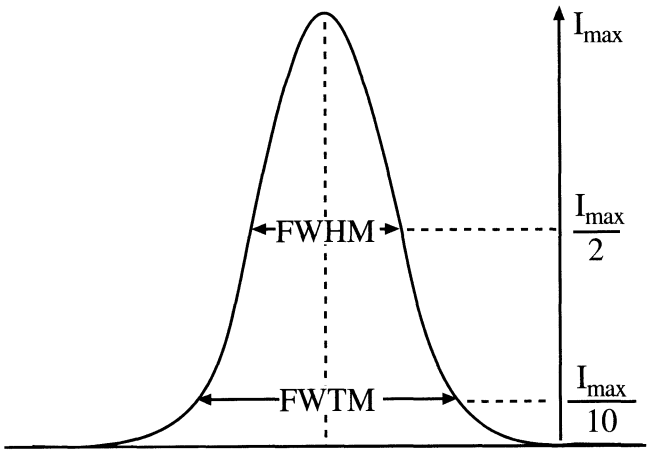


Figure 5.9. The definition of the full width at half maximum (FWHM) and the full width at tenth maximum (FWTM) of a Gaussian intensity distribution which is typical of a well-aligned beam.

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 should be added in quadrature to give a total, calculated beam size, d_t

$$d_t = \left(d_g^2 + d_s^2 + d_d^2\right)^{1/2} \qquad [5.7]$$

This equation gives us only a first-order estimate, since it is not clear that all the contributions are 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

$$d_g = \frac{2}{\pi}\left(\frac{i}{\beta}\right)^2 \frac{1}{\alpha} \qquad [5.8]$$

We have already defined i, β, and α.

The disk of minimum confusion caused by spherical aberration has a diameter given by

$$d_s = 0.5\, C_s\, \alpha^3 \qquad [5.9]$$

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. The calculated diameter due to diffraction is

$$d_d = 1.22\, \frac{\lambda}{\alpha} \qquad [5.10]$$

which is the Rayleigh criterion which we discussed in Chapter 1. Although all these definitions do not define the same diameter of the electron distribution, they are all combined to give a first approximation of the FWHM of the beam. Clearly, it is more reliable, but more time-consuming, to measure d experimentally. Figure 5.10 shows the result of calculations of the three contributions to the beam diameter in a VG HB501 STEM.

5.5.D. Measuring the Beam Diameter

To measure the beam size in a TEM/STEM, you must form an image of the beam on the TEM viewing screen under conditions where you know, or can calibrate, the magnification. This is a nontrivial exercise and you may need to

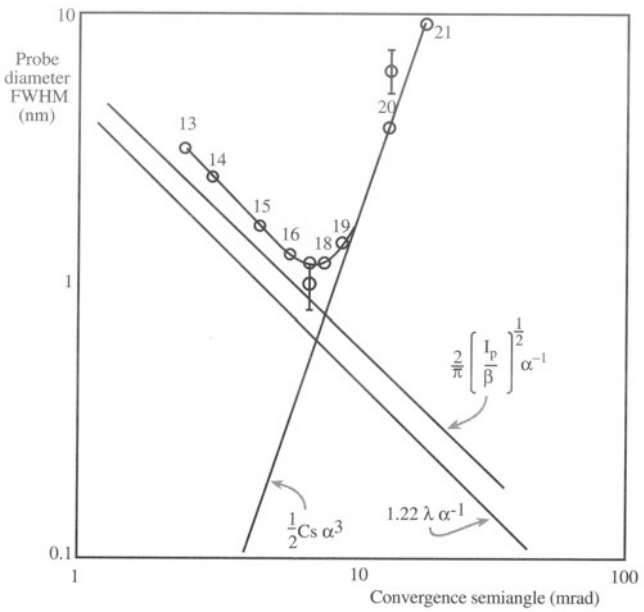


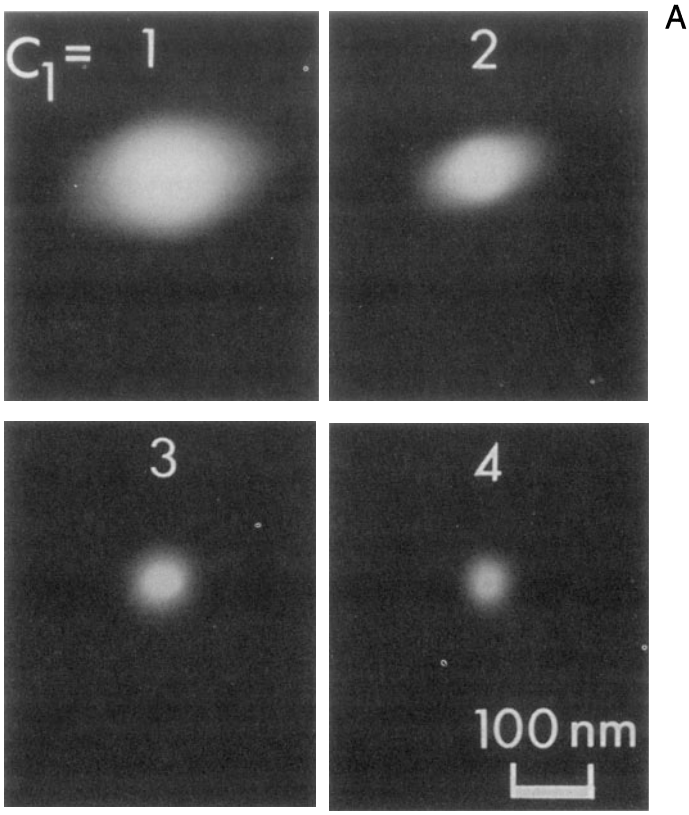
Figure 5.10. Calculations of the three contributions to the probe size as a function of the convergence semiangle α in an FEG STEM with a probe current I_p of 0.85×10^{-8} A. Two experimental measurements are shown, at condenser 1 lens settings 17 and 20. The minimum probe dimension is ~1 nm with α < 10 mrads.

consult the manufacturer 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 image, as shown in Figure 5.11.

- The FWHM contains 50% of the integrated intensity. It is the value used by the manufacturers when they report beam sizes. It is also the important dimension when considering the effect of *d* on the STEM image resolution.
- The full width at tenth maximum (FWTM) contains 90% of the integrated intensity. It is a more relevant dimension because the Faraday cup measures the current in the total beam which is closer in size to the FWTM.

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 x FWHM and this is also shown in Figure 5.9. You should note, therefore, that you overestimate β if you use the smaller FWHM. Use of the FWTM is also the preferred beam size when calculating the spatial resolution of microanalysis, as we describe in Chapter 36.

In a dedicated STEM you can't image the beam directly, since there are no post-specimen lenses to magnify



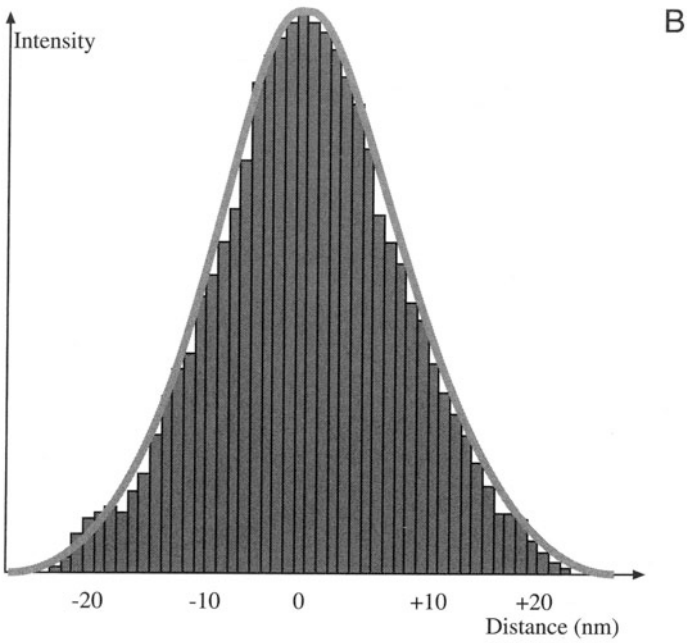


Figure 5.11. (A) Four images of the beam formed on the TEM screen at different condenser 1 lens settings and (B) the corresponding microdensitometer trace across spot #3, confirming the Gaussian nature of the intensity distribution.

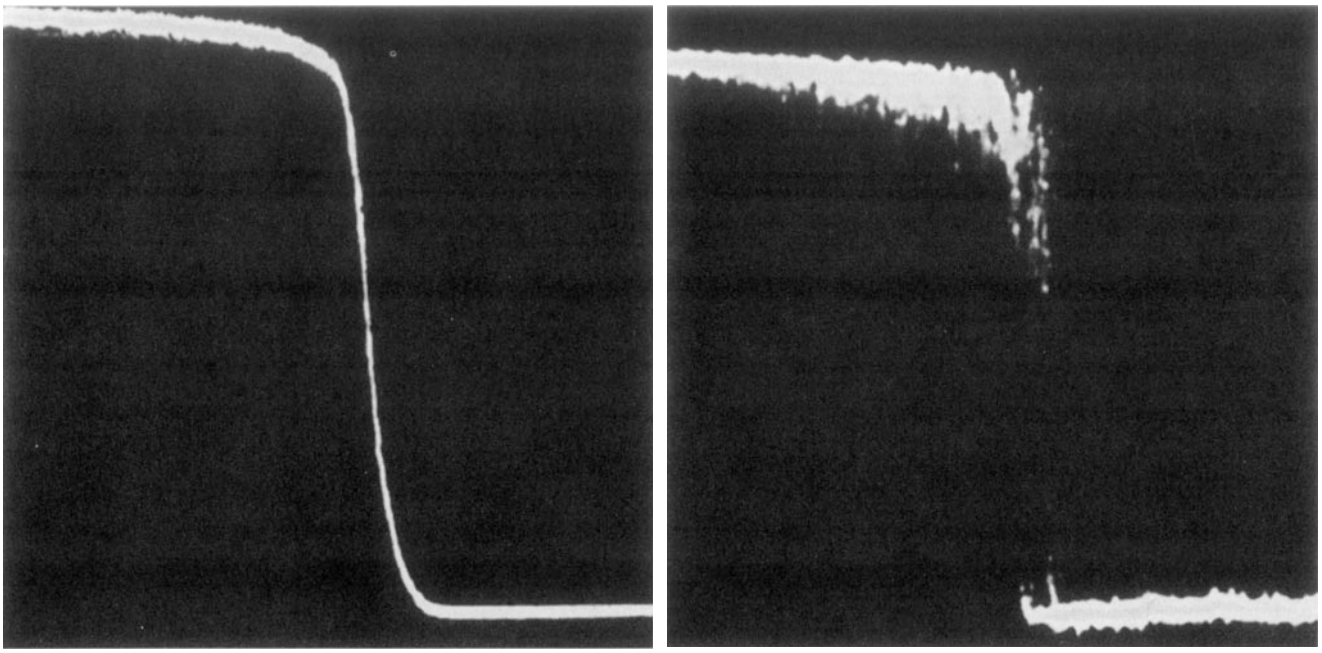


Figure 5.12. Intensity profiles obtained by scanning a fine beam across a sharp edge of a cube of MgO. The measured probe size (FWTM) in (left) is 7.4 nm (magnification 1×10^6) and in (right) 1.8 nm (magnification 11×10^6). The smaller probe contains a much smaller current and is therefore a noisier trace.

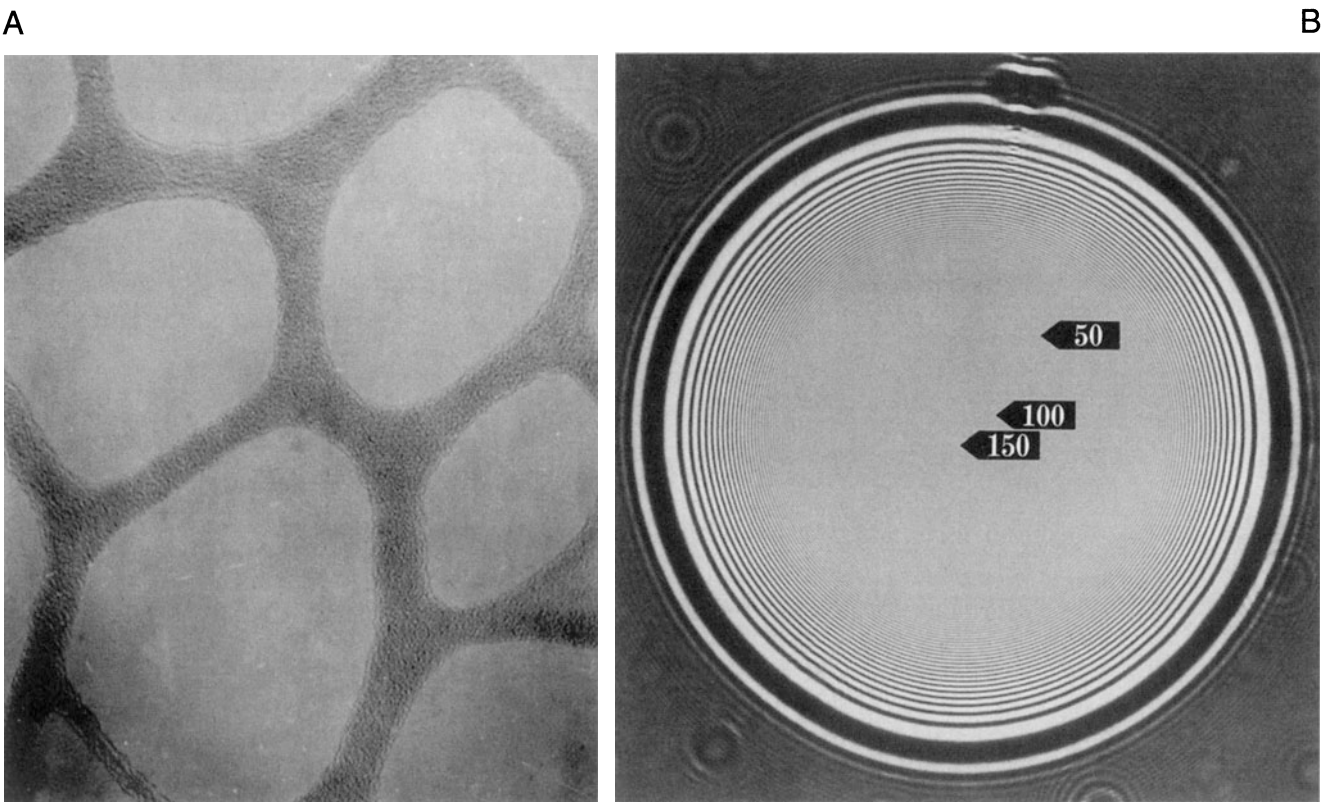


Figure 5.13. Fresnel fringes from (A) a thermionic source with poor coherency and (B) an FEG with high coherency.

its image and no photographic film to record it. The value of d must be determined indirectly, as in other scanning instruments. The best method involves scanning the beam across a knife-edge specimen and monitoring the intensity change which occurs, for example, by recording the output from the annular dark-field detector. 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. In Figure 5.10, two experimental beam-size measurements are shown; they show reasonable agreement with the calculated values from the brightness equation.

The measurement of d is clearly not a simple procedure. You can find a full description of the problems in the paper by Michael and Williams (1987).

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. The spectrum then consists of a single Gaussian peak and the resolution of the spectrum is defined as the FWHM of this peak. You can find out how to do this in detail in Chapter 37. Typical values of ΔE for the various electron sources are also given in Table 5.1.

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 (see Part III). We can also use them to correct the astigmatism in the objective lens, as we'll see in Chapter 9. The number of visible fringes is a measure of the beam coherency. Figure 5.13B shows the enormous number generated by an FEG.

5.6. WHAT kV SHOULD YOU USE?

For the materials scientist, this is usually an easy question to answer. You always operate at the maximum available kV, unless there is a definite reason to use a lower kV. Of these reasons, the most obvious is avoiding beam damage, but we'll see others later in the book, so don't forget that you can always operate a 300-kV machine at 100 kV. Remember, it's like being able to change the wavelength of a monochromatic light source in a visible-light microscope (VLM). The threshold for beam 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 voltages may be better, but there is not much use going below 100 kV since the images will be rather dim and you'll have to make extraordinarily thin specimens to see anything useful. The reasons for choosing the highest kV are:

- The gun is brightest.
- The wavelength is shortest; the resolution is potentially better.
- The cross section for inelastic scatter is smaller; the heating effect is smaller.

CHAPTER SUMMARY

Most TEMs use thermionic sources and, if you have the choice, use an LaB_6 source and run at the highest kV. Take care when heating and cooling the LaB_6 crystal and always operate just below saturation to maximize the lifetime of a source. If you're going to be doing AEM, get some idea of the beam current that you can get from your source under typical operating conditions. Also, measure the beam size and convergence angle to give a measure of β, and if you're doing EELS then the energy spread is essential information. If you have an FEG you'll most likely be doing fine probe analytical work, in which case all the above characteristics must be measured, and if you're going to do high-resolution imaging, then the degree of coherency is important too. Always treat the source carefully when changing it, aligning it, saturating it, or switching it off. There's nothing more annoying than losing your source, since it usually happens at some critical point during your work.

REFERENCES

General References

Broers, A.N. (1974) SEM 1974 (Ed. O. Johari), p. 9, IITRI Chicago.
Joy, D.C. (1973) SEM 1973 (Ed. O. Johari), p. 743, IITRI Chicago.
Orloff, J. (1989) *Ultramicroscopy* **58**, 88.
Reimer, L. (1993) *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis*, 3rd edition, Springer-Verlag, New York.
Veneklasen, L.H. (1975) *Optik* **36**, 410.

Specific References

Hawkes, P.W. (1978) *Advances in Optical and Electron Microscopy* **7**, 101.
Michael, J.R. and Williams, D.B. (1987) *J. Microsc.* **147**, 289.

Lenses, Apertures, and Resolution

6

CHAPTER PREVIEW

Electron lenses are the magnetic equivalent of the glass lenses in an optical microscope and, to a large extent, we can draw comparisons between the two. For example, the behavior of all the lenses in a 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:

- either take all the rays emanating from a point in an object and recreate a point in an image,
- or 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 magnetic 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.

The 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 aberrations, which we control by inserting limiting apertures. You need to understand these aberrations, since they play a major role in deciding what we can and cannot do with the microscope. In particular, the resolution of an electron lens (rather than the wavelength of the electrons) limits the resolution of the TEM. Since resolution is usually the single most important reason for buying a TEM, you need a firm understanding of this concept. Unfortunately, we electron microscopists aren't very firm in our definitions of resolution. Finally, we describe how the apertures we put in the lenses aid both the depth of field and the depth of focus of the instrument.

Lenses, Apertures, and Resolution

6

6.1. WHY LEARN ABOUT LENSES?

Why should we learn about electron lenses? As in a visible-light microscope, the lenses in a TEM control all the basic operational functions of the instrument. We physically move glass lenses up and down in a light microscope to control the intensity of the illumination and the focus and magnification of the image. The focal length of a glass lens is fixed. In a TEM the positions of the lenses are fixed and we focus, etc., by changing the strength of the lenses. As you'll see, in most cases the lenses we use are magnetic, so that we change their strength by changing the 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 diffraction patterns.

We also use apertures in the lenses to control the beam current and the convergence of the beam hitting the specimen.

These factors are critical in imaging, diffraction, and microanalysis. 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 diffraction pattern.

In essence then, we control the quality of our images, diffraction patterns, and analytical signals by adjusting the lenses and their apertures. 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.

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 light microscope?

We'll see that the answer to these questions lies 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.

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. (As children, we used this latter property to set fire to a piece of paper by focusing the 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 that we use act, to a reasonable approximation, like convex glass lenses; in detail, they're often equivalent to more complex combinations of convex lenses. Remember that, at present, all magnetic lenses are convex lenses.

6.2.A. How to Draw a Ray Diagram

In traditional light optics it's customary to draw ray diagrams of the path of light rays through the lens, and we do the same for electrons and their lenses. These ray diagrams are usually drawn horizontally because the traditional optical bench on which light optics experiments are carried out is a horizontal setup. But since the electron microscope is usually a vertical instrument, we will draw all our ray diagrams vertically.

Let's start by drawing ray diagrams to illustrate the two fundamental actions of image formation and focus 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 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 later that all these assumptions are wrong, to a degree, but that these traditional illustrations are nonetheless 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.

> The optic axis is an imaginary line down the column of the TEM passing through the center of each lens.

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 semiangle 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 semiangle of collection defined as α, but we will reserve α for convergence semiangles (see Section 2.7).

> So, all lenses are imperfect insofar as they cannot gather all the radiation emitted by an object and so we can never create a perfect image.

As you know from Chapters 2–4, most electrons are strongly forward scattered, so we can in practice gather a high fraction of the scattered electrons.

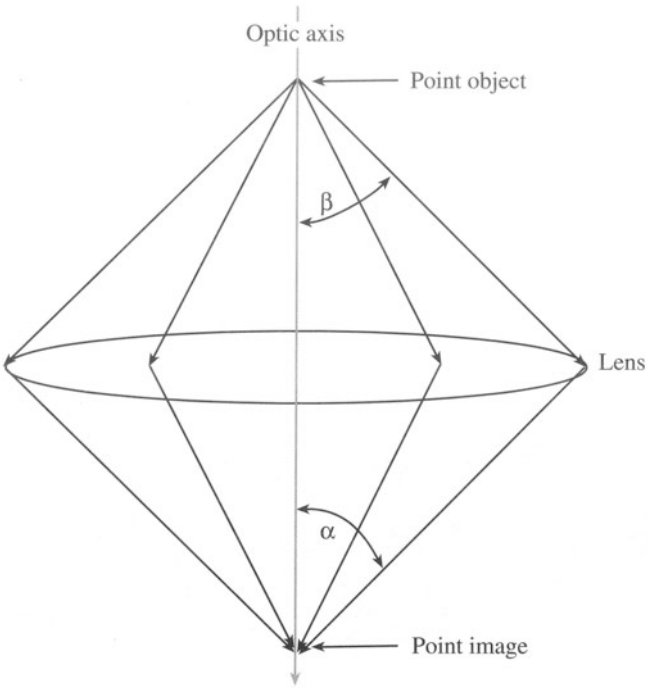


Figure 6.1. Image formation by a convex lens. A point object is imaged as a point and the collection semiangle of the lens is defined relative to the object (β) or the image (α).

> The angles in Figure 6.1 and in the other ray diagrams we'll draw are all greatly exaggerated. In practice, a typical value of β is maybe a few tens of milliradians (10 mrads = 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. (Of course, this is a fundamental principle of how a lens works.) The second step is to draw line 2, which is a ray from the arrowhead that is parallel to the optic axis. The farther away that rays are from the optic axis, the more strongly they are bent by a convex lens, so we take line 2 and bend it toward the optic axis as it passes through the

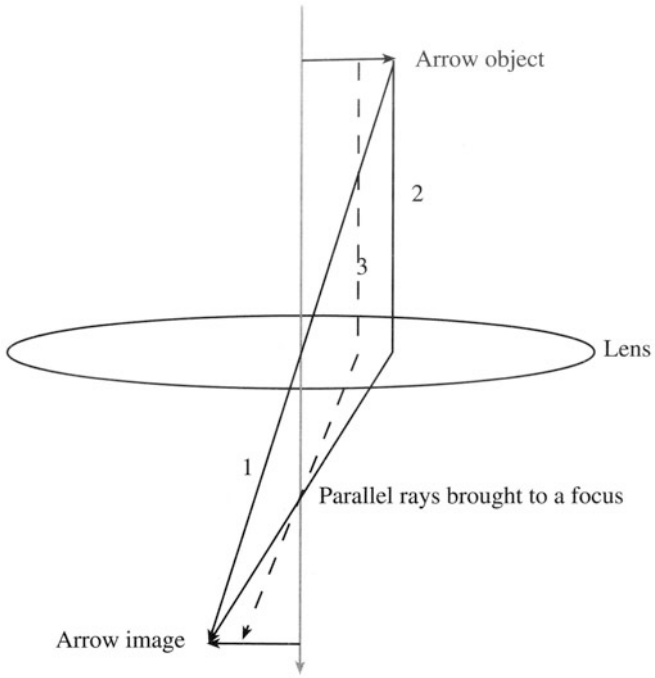


Figure 6.2. How to draw a ray diagram: first construct ray 1 through the middle of the lens, then ray 2, parallel to the optic axis, to determine the lens strength. Finally, draw line 3 parallel to 2 to define the focal plane where the parallel rays are focused. Thus an asymmetric object is imaged off axis and rotated through 180°.

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. We could draw as many rays as we wished from any point on the object arrow to an equivalent point on the image arrow, such as line 3. Note that parallel rays 2 and 3 both cross the axis at the same point, illustrating the second fundamental action of a convex lens, i.e., bringing parallel rays to a focus. Again, the strength of the lens determines where the parallel electrons are focused.

> The image formed after each 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 diffraction patterns in their focal plane. We'll use this diagram to introduce you to the principal terms used in lens optics.

6.2.B. The Principal Optical Elements

From the above diagrams, we can define three important planes to which we will often refer. The first plane is the object plane, which 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 second plane is the image plane (sometimes called the Gaussian image plane), which 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. The third plane is the focal plane of the lens, and this 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"

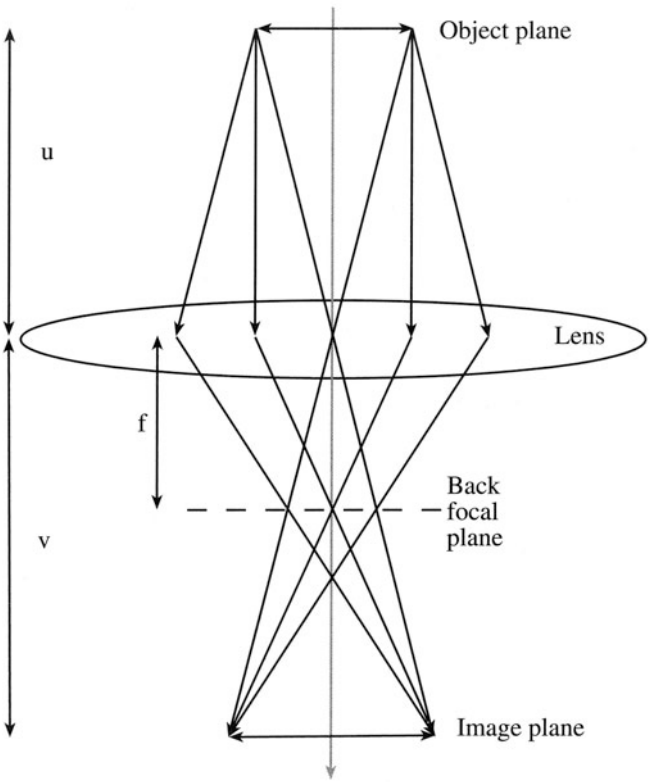


Figure 6.3. A complete ray diagram for a finite object, symmetrically positioned around the optic axis. All rays emerging from a point in the object (distance u from the lens) that are gathered by the lens converge to a point in the image (distance v from the lens) and all parallel rays are focused in the focal plane (distance f from the lens).

the lens and so the plane is sometimes called the "back focal plane" (BFP). There is also an equivalent "front focal plane," 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, u), the distance from the lens to the image plane (the image distance, v), 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

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \qquad [6.1]$$

which is known as Newton's lens equation, and its proof can be found in a standard optics text such as Jenkins and White (1976). In thick lenses u and v are measured from different principal planes in the 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 the convex lens as

$$M = \frac{v}{u} \qquad [6.2]$$

M is also approximately equal to the ratio of the collection semiangles 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.

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 finest possible beam at the specimen). If that is the case, we define the demagnification as $1/M$. In an optical microscope 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 an electron microscope, 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 u is unchanged, then v must be correspondingly shorter and the image magnification is smaller, or the demagnification is larger. Under these conditions which normally occur in the TEM, strong lenses magnify less and demagnify more, which is counter to our understanding of light microscopes in which stronger lenses produce greater magnifications.

How do we get the high magnification that we need to form images of atom rows such as Figure 1.2? What we do is put the object close to the lens, making u small and M large, and repeat this for several lenses in tandem one after the other. So we end up with a multilens system like a com-

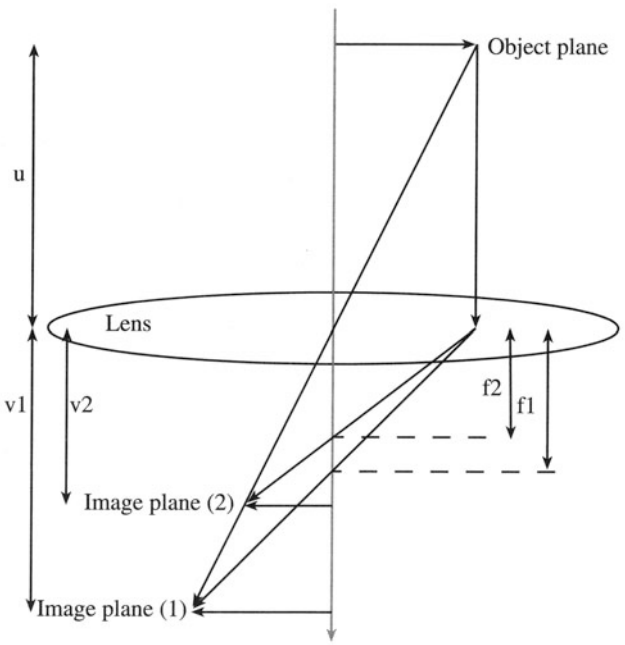


Figure 6.4. Strengthening the lens shortens the focal length f. So a weaker lens ($f1$) produces a higher magnification of the object than a stronger lens ($f2$) since the image distance v increases, but the object distance is unchanged.

pound optical microscope. (We'll discuss the details of the lens combinations in the illumination and imaging systems of the TEM in Chapter 9.) In these circumstances, the image plane of the upper lens acts as the object plane for the lower lens, assuming we want the lower lens to further magnify the image produced by the upper lens. In principle, there's nothing to stop us magnifying as much as we wish.

Don't confuse magnification with resolution.

Above a certain magnification, we will see no more information because other factors limit the image detail and therefore the resolution of the microscope. We'll discuss this point 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 diffraction pattern). To do this, the back focal plane of the upper lens must become the object plane for the subsequent lenses in the imaging system. When discussing focus we need another convention.

- If the lens is too weak and the image forms below the desired image plane, the image will be out of focus and the lens is said to be *underfocused.*
- If the lens is too strong and the image forms above the image plane, then we say the lens is *overfocused.*

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.

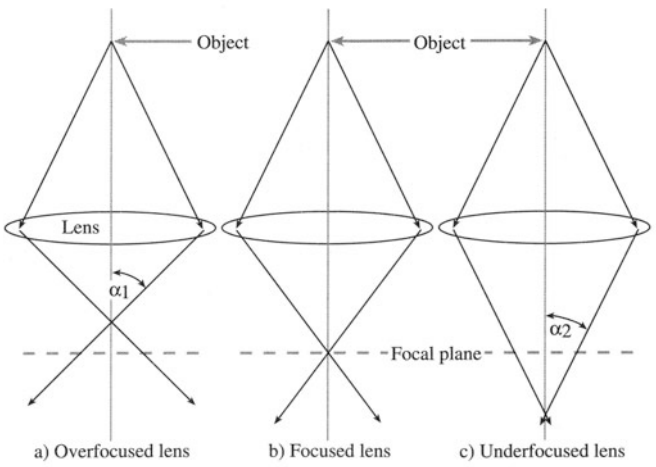


Figure 6.5. (a) Ray diagram illustrating the concepts of overfocus, in which a strong lens focuses the rays before the image plane, and (c) underfocus, where a weaker lens focuses after the image plane. It is clear from (c) that at a given underfocus the convergent rays are more parallel than the equivalent divergent rays at overfocus ($\alpha_2 < \alpha_1$).

6.3. ELECTRON LENSES

Electrons were first successfully focused by Busch in 1927; he used an electromagnet of the sort that Ruska incorporated into the first TEM. Busch also showed that it's 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. So TEMs use magnetic lenses exclusively and we won't discuss electrostatic lenses further.

6.3.A. Polepieces and Coils

To make a magnetic electron lens you need two parts, and both are drawn schematically 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. 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 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 and 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

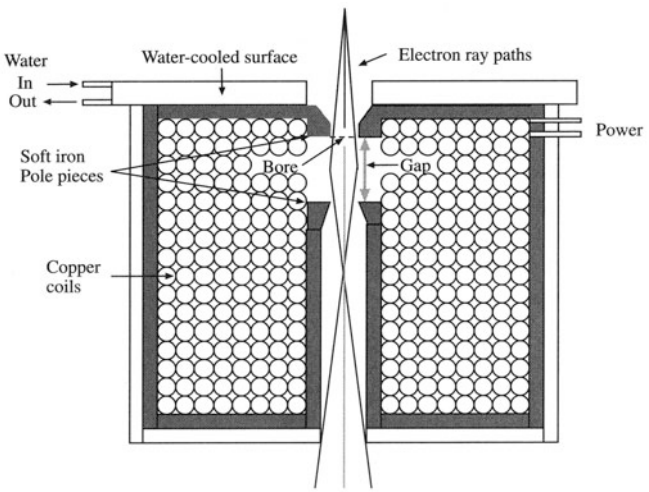


Figure 6.6. Schematic diagram of a magnetic lens. The pole pieces surround the coils and, when viewed in cross section, the bore and the gap between the polepieces are visible. The magnetic field is weakest on axis and increases in strength toward the side of the polepiece, so the electrons are more strongly deflected as they travel off axis.

through the coil, a magnetic field is created in the bore. This field is inhomogeneous along the length of the lens, but axially symmetric. The strength of the field in a magnetic lens controls the 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 microscope is shown in Figure 6.7.

Practical hint: *You should be able to get a readout (on your TEM console) of the current through any lens coil and it is a useful thing to know the standard lens currents for your common operating modes such as imaging and diffraction and for various beam sizes.*

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 the diffraction pattern and project it onto the viewing screen in ways that we'll see in Chapter 9. Typically, these lenses are of the sort shown schematically in Figure 6.6 and an aperture can be introduced into the bore of the lens, as we'll discuss later.

The objective lens is the most important lens in the TEM, since it forms the images and diffraction patterns that will be magnified by all the other lenses. It is also the most difficult to construct, since the specimen must be located so close to the "plane" of this lens.

The objective lens is a 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 coil, as shown in Figure 6.8A. This geometry gives the space needed to allow us to insert the specimen and the objective 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, and straining, and this versatility accounts for the popularity of the split polepiece lens in TEMs.

With split polepieces it is possible to make the upper polepiece behave differently than the lower polepiece.

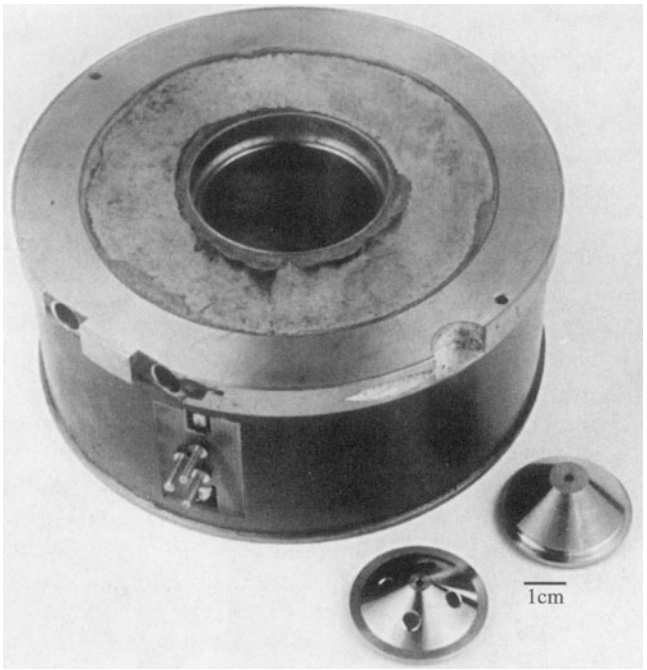


Figure 6.7. A real lens: the cylindrical shape conceals the copper wire coils. The two conical polepieces beside the lens sit inside the central hole in the lens. The three-pin electrical connections provide current to the coil to magnetize the polepieces, and cooling water is circulated in and out of the two holes in the top plate of the lens to dissipate the resistive heat generated.

The most common application of this is to excite the upper objective polepiece very strongly. This kind of 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 needed. This is traditionally accomplished by an immersion lens in which 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 than 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 more recent lens designs such as the snorkel lens, as shown

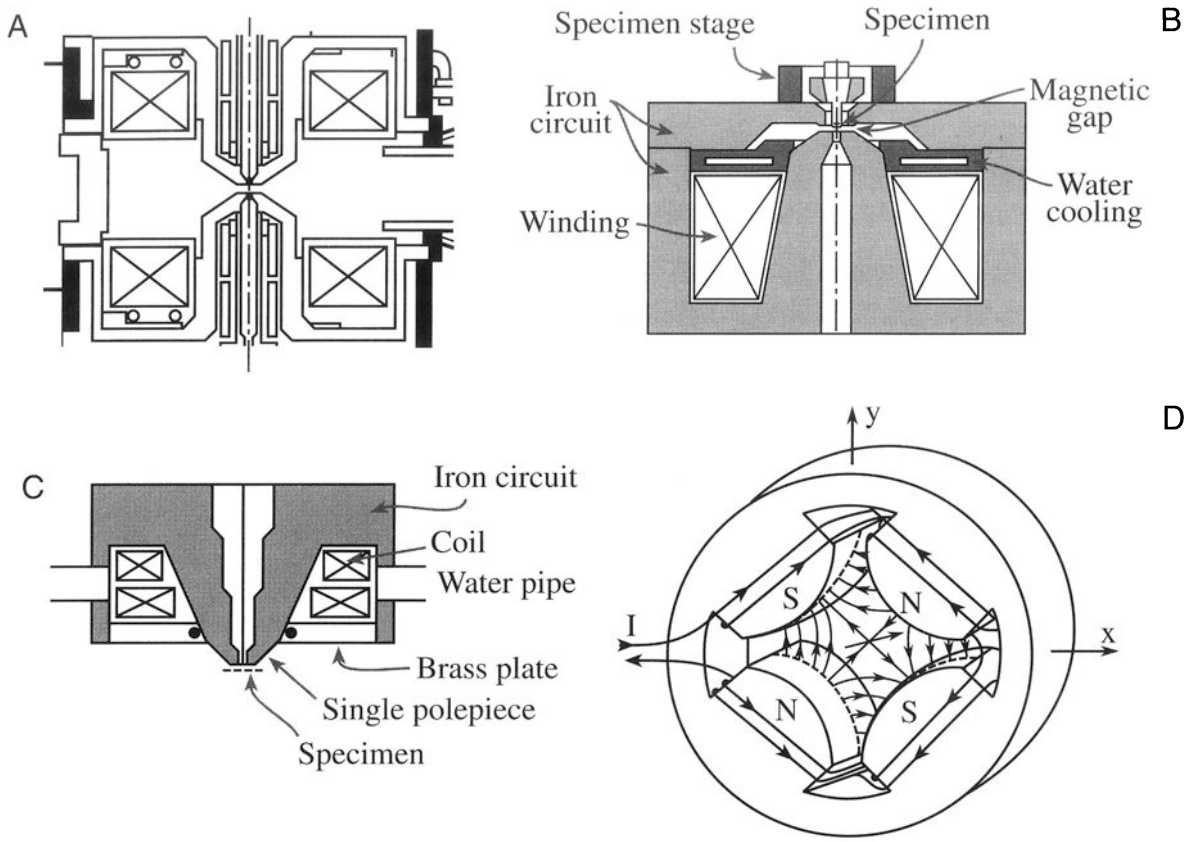


Figure 6.8. A selection of different lenses: (A) a split polepiece objective lens, (B) a top-entry immersion lens, (C) a snorkel lens, and (D) a quadrupole lens.

in Figure 6.8C, which is a single polepiece lens with a small bore to give a strong 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. Nevertheless, there is a lot of interest in these lenses because they are small, they don't need water cooling, and they cool the area around the specimen which both improves the vacuum and helps minimize contamination. They can generate intense fields (>100 T) 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 reduced to the level where resolutions < 0.1 nm are feasible and they may be used in the future to construct compact TEM columns.

In addition to these variations on the theme of a single or double polepiece, it is also possible to design a quadrupole and octupole lens in which the focusing action is achieved by four or eight polepieces. Adjacent polepieces are of opposite polarity as shown in Figure 6.8D. These lenses are not used in TEMs as magnifying lenses but they are used to correct lens defects such as astigmatism, and they are used as lenses in electron spectrometers (see 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 the magnetic lenses actually work. When an electron with charge q ($= -e$) enters a magnetic field of 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

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \qquad [6.3]$$

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 the thumb.

Field: Forefinger; Velocity (Speed): Second finger; Thrust: Thumb (Right-hand rule.)

The force on an electron entering a uniform magnetic field, nearly 90° to **B** is

$$F = evB \sin\theta = evB = \frac{mv^2}{r} \qquad [6.4]$$

where r is the radial distance of the electron from the optic axis and m is the mass of the electron.

We can rearrange equation 6.4 to give an expression for

$$r = \frac{mv}{eB} \qquad [6.5]$$

Since v is a relativistic velocity, we should write this equation as

$$r = \frac{\left[2m_0E\left(1 + \frac{E}{2E_0}\right)\right]^{1/2}}{eB} \qquad [6.6A]$$

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)

$$r = \frac{3.37 \times 10^{-6}\left[V(1 + 0.9788 \times 10^{-6}\, V\right]^{1/2}}{B} \qquad [6.6B]$$

For example: if V = 100 kV and the magnetic field B is 1 Tesla, the radius is less than 1 mm.

In deriving equation 6.4, we made a rather gross oversimplification. If θ does equal 90°, the electron is traveling straight across the optic axis and is not focused! 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 $z = v_2t$, while the perpendicular component produces circular motion with radius given by equation 6.5. Hence we see the very important result:

The electron spirals through the lens field with a helical trajectory.

The period of rotation (T_c) through the field gives rise to a so-called "cyclotron frequency" ω

$$\omega = \frac{2\pi}{T_c} = \frac{eB}{m} \qquad [6.7]$$

From these various relationships, we can calculate the complete ray paths through the lens. The most important equations are called the "paraxial" (near-axis) ray equations, which describe both r and the angle of rotation (θ) about the axis as the electron moves down the axis in the direction z, rotating under the influence of the cylindrically symmetrical field, B. These equations, which neglect electron trajectories far off axis, are derived in texts on electron optics. The account by Hawkes (1972) is particularly clear if you have an interest in the physics of electron lenses. As Hawkes succinctly states, "a straightforward, but quite lengthy calculation yields"

$$\frac{d^2r}{dz^2} + \frac{\eta^2B^2r}{2V^{1/2}} = 0 \qquad [6.8]$$

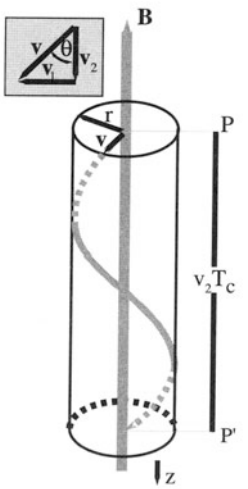


Figure 6.9. Electron trajectories in a homogeneous magnetic field, strength **B**. The electrons have velocity components parallel and perpendicular to the field, so long as they are not traveling at 90° to the direction of **B**. The Lorentz force causes electrons passing through point P on the optic axis to spiral through the field and intersect the axis again at P'. The electron's helical path defines the cyclotron radius, r.

$$\frac{d\theta}{dz} = \frac{\eta B}{2V^{1/2}} \qquad [6.9]$$

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 is smaller for more energetic electrons and larger for more intense field strengths. From equation 6.9 the rate of angular rotation increases with increasing field strength and decreases for more energetic electrons.

When the energy of the electrons increases, we must use stronger lenses (larger **B**).

When we increase **B**, the pitch of the helical path becomes steeper.

Both of these conclusions are intuitively obvious, but the implication is often missed. When we change our operating kV, we change the lenses in the microscope! (Think what this would mean in a visible-light microscope.) Thus the calibration of the microscope and lens "constants" change as we change the accelerating voltage.

While all these ray equations are approximations, they form the basis of more detailed mathematical models of electron motion through lenses. The more complete models are used in a computer program (Munro 1974), 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 modern TEMs. We also use non-paraxial rays to explain the effect of spherical lens aberrations on resolution in Section 6.5.A.

6.3.D. Image Rotation and the Eucentric Plane

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. Its effects are seen in the routine operation of the TEM because the image, or diffraction pattern, rotates on the viewing screen as you try and focus or if you change magnification. This rotation may require calibration; as we'll see later, the manufacturer may have 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, 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 sit in the eucentric plane because an image of an object in this plane will not move as you tilt the specimen. All other planes in the imaging system are defined with reference to the 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. We'll return to this point later in the book.

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 exceedingly 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 drawing the lens as having zero thickness along the optic axis, the magnetic field actually acts over a length L. The angle of deflection ε is (for small ε)

$$\varepsilon = \frac{eLB}{mv} \qquad [6.10]$$

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 1 cm, giving a field of 0.01 T. For electrostatic blanking we need about 20 kV/cm.

6.4. APERTURES AND DIAPHRAGMS

We mentioned earlier that an aperture is often inserted into a lens. The aperture limits the collection angle of the lens as shown schematically in the diagram in Figure 6.10, and such an aperture in the objective lens thus 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 diffraction pattern, etc. Physically, the aperture may reside above, in, or below the plane of the lens as we draw it in ray diagrams. Apertures can also perform other functions which we'll come across later, such as protecting the specimen from stray radiation

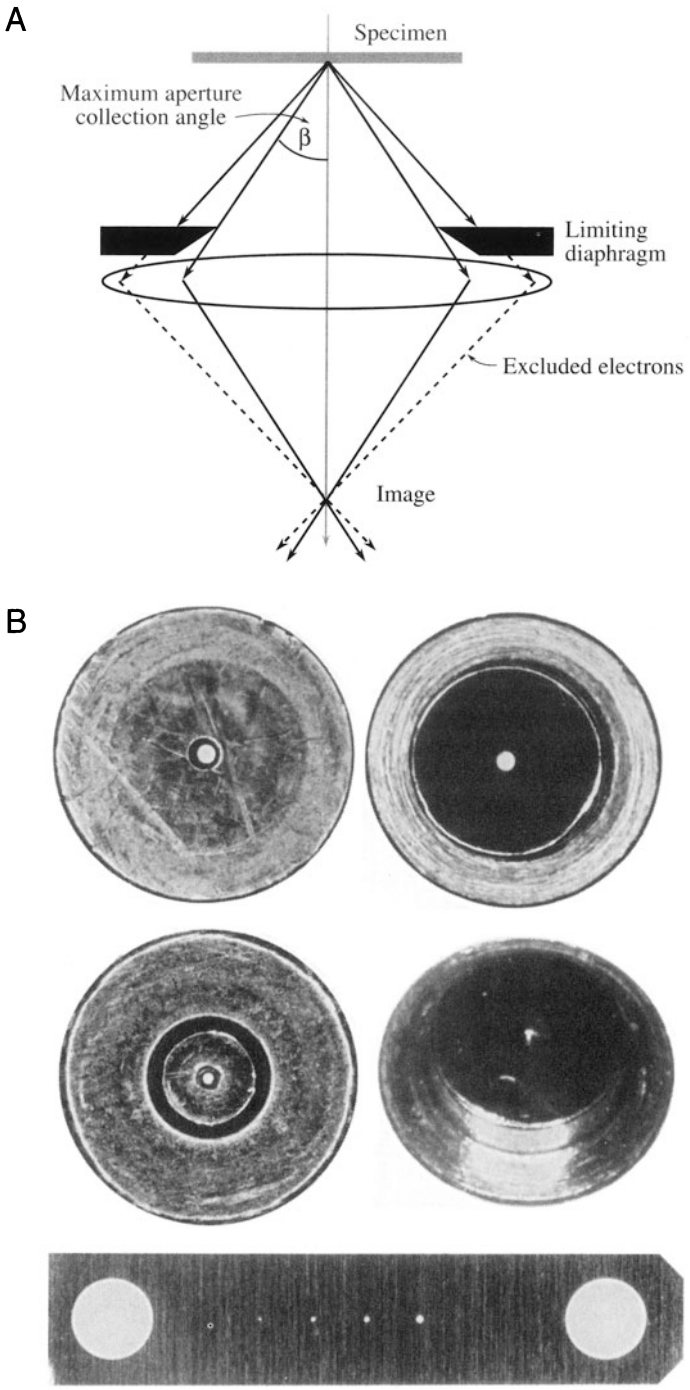


Figure 6.10. (A) Ray diagram illustrating how a diaphragm restricts the angular spread of electrons entering the lens. Only electron paths less than a semiangle β subtended by the aperture at the object are allowed through the lens (full ray paths). Electrons from the object scattered at angles >β are stopped by the diaphragm (dashed ray paths). (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.

in the illumination system. So they are really important. 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 an optical microscope 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 just use the term "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 said 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 with a varying diameter aperture, 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 the smallest circular aperture we can make consistently, up to ~ 0.3 mm (300 μm). Usually the individual diaphragms or the strips are about 25–50 μm thick, but if their job is also to prevent X-rays from hitting the specimen they may be several millimeters thick, which means they can be quite expensive if they're 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 gather on the edges of the aperture, destroying its circular shape, and this causes 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 of very thin metal foil (e.g., Au or Mo) because the foil gets hot in the electron beam and boils off any contamination. 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: *X-rays with energies up to the beam energy are generated within the lens wherever the electron beam hits a surface (particularly the limiting diaphragm). So substantial and 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 service engineers should be allowed to dismantle or in any way take apart and repair the lenses.*

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 70 years since Busch's first magnetic lens, we haven't progressed so far and our lenses are very imperfect. We've already compared the best electromagnetic lens to the equivalent of using the bottom of a Coke bottle as a magnifying glass. Another common description is that if your own eye lens was 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 ten kinds of lens defects (see Reimer 1993) and at one time or another all their effects can be seen. In practice, however, we 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 is caused by the lens field acting inhomogeneously on the off-axis rays. 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. Ultimately, it is this defect that limits the resolution of modern 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 as an intense central bright region with a surrounding halo in the Gaussian image plane. This image goes by the delightful term of the "disk of least confusion" or sometimes the "disk of minimum confusion." Spherical aberration is most important in the objective lens because it degrades the image that we view in a TEM. Also, it is equally important in the condenser lenses in an AEM or STEM which we use to form the finest beam with the most current. What we can accomplish is almost always limited by spherical aberration.

From Figure 6.11 you can see why we use the term "spherical" to describe the aberration. The effect is to take

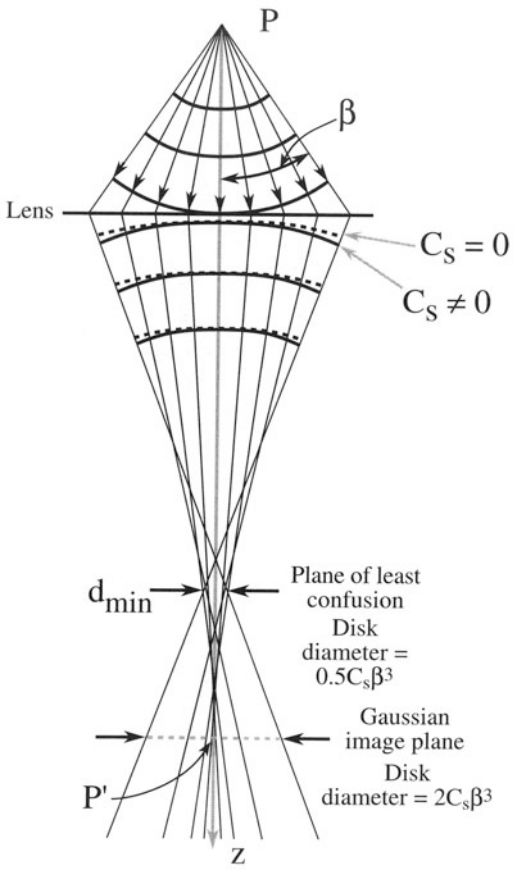


Figure 6.11. Spherical aberration in the lens causes wavefronts from a point object P to be spherically distorted. The point is thus imaged as a disk with a minimum radius in the plane of least confusion and a larger disc at P' in the Gaussian image plane.

the curved wavefronts and further curve them. 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 (Reimer 1993)

$$\begin{aligned} PP' &= v_2 T_c = vT_c \cos\theta = 2\pi\frac{mv}{eB}\left(1 - \frac{\theta^2}{2} + \cdots\right) \\ &= L_0\left(1 - \frac{\theta^2}{2} + \cdots\right) \end{aligned} \quad [6.11]$$

In this relationship $L_0 = PP_0'$, where P_0' is the Gaussian image of the point P for very small θ (i.e., paraxial conditions). As θ increases, the distance PP' decreases because of spherical aberration and we can write

$$PP' = PP'_0 - \Delta z \quad [6.12]$$

where $\Delta z = 0.5L_0\theta^2$. Thus we get an expression describing the error, δ, in the Gaussian image position due to spherical aberration

$$\delta = \Delta z \tan\theta \sim \Delta z\theta = 0.5L_0\theta^3 \quad [6.13]$$

So the diameter of the Gaussian image of a point *formed by paraxial rays* is given by this expression, which we usually write as

$$\delta = C_s\theta^3 \quad [6.14]$$

where C_s is a constant for a particular lens called the spherical aberration coefficient.

Now this equation is for paraxial rays only, and in a real microscope the apertures are usually large enough so that paraxial conditions don't apply. As a result the spherical aberration error in the Gaussian image is broadened to a diameter of $2C_s\theta^3$. We'll see in a while that this equation is *very* important because of its effect on the resolution of the TEM, and sometimes it is written $2C_s\theta^3M$ when referred to the image plane. In a real lens the value of θ in equation 6.14 is replaced by the maximum semiangle of collection of the objective lens aperture, β. We'll use β to define the objective lens semiangle of collection in the forthcoming discussion of resolution. This contrasts with our use of α when discussing beam size in Chapter 5, since α defined a semiangle of beam convergence. Most other TEM texts use α indiscriminately for collection and convergence angles. When we discuss resolution in the TEM later in this chapter, we will use the radius rather than the diameter, and since all discussion of resolution refers back to the object, the magnification term is customarily ignored.

So we end up with an expression for the radius of the spherical aberration disk r_{sph} in the Gaussian plane under nonparaxial (i.e., realistic) conditions, given by

$$r_{sph} = C_s\beta^3 \quad [6.15]$$

From this derivation you can see that C_s has the dimensions of length. Typically, it is approximately equal to the focal length, which in most TEMs is about 3 mm, but in high-resolution instruments may be well below 1 mm. So one way to minimize this aberration is to put your money into buying a TEM with a short focal length lens.

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. The smallest dimension is called the disk of least confusion and has a radius of $0.25C_s\beta^3$. As we'll discuss in Section 6.6.C, some texts use this smaller dimension to define the resolution limit imposed by spherical aberration.

6.5.B. Chromatic Aberration

This term is related to the "color" (i.e., wavelength and hence energy) of the electrons. We've assumed the electrons are monochromatic, but they aren't. However, we can now make very good high-tension supplies and the variation of the electron energy is usually smaller than one part in 10^6, which is 0.1 eV for a 100-keV beam. In fact, this is so good that we don't have to worry about chromatic aberration in the illumination system. In TEM images, chromatic aberration could also be safely 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 form a disk image (Figure 6.12). The radius r_{chr} of this disk is given by

$$r_{chr} = C_c\frac{\Delta E}{E_0}\beta \quad [6.16]$$

where C_c is the chromatic aberration coefficient of the lens (like C_s it is a length, approximately equal to the focal length), ΔE is the energy loss of the electrons, E_0 is the initial beam energy, and β is the semiangle of collection of the lens. While ΔE in the incident electron beam is < 1 eV, it is typically 15–25 eV for a good fraction of the electrons coming through a typical thin foil 50–100 nm thick. Chromatic aberration gets worse for thicker foils and better for thinner ones. So you can do something about this aberration—make good thin specimens!

6.5.C. Astigmatism

Astigmatism occurs when the electrons sense a nonuniform magnetic field as they spiral round the optic axis.

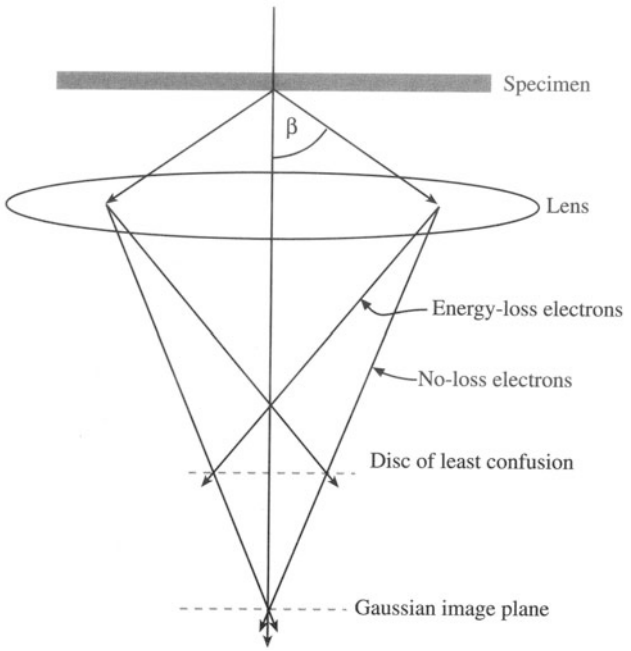


Figure 6.12. Chromatic aberration results in electrons with a range of energies being focused in different planes. Electrons emerging from the specimen with no loss of energy are less strongly focused than those that suffered energy loss in the specimen, so a point is imaged as a disk.

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 is a variety of contributions to astigmatism, which distorts the image by an amount r_{ast} where

$$r_{ast} = \beta \Delta f \qquad [6.17]$$

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.

These then 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 close to the bore of the polepiece appear in the image. Other effects, such as coma, field curvature, etc., we will ignore.

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 light optical microscopy, which is concerned with the imaging of incoherent waves through amplitude contrast. High-resolution performance in the electron microscope is a different matter and 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 light microscopy (Bradbury *et al.* 1989) the term "resolution" strictly applies to the act of displaying fine detail in an *image.* The resolving power of the microscope is the ability to make points which are closely adjacent in the object, distinguishable in the *image.* Now 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 microscope in terms of distance in the *object* (usually a few Å), we should then use the term "minimum resolvable distance," but instead everyone says "resolution."

> We will use the term "resolution," but we define it to mean the "minimum resolvable distance" in the object.

Because the lens defects that we've 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 image of the electron source demagnified onto the specimen. In either case, the aberrations limit the resolution.

6.6.A. Theoretical Resolution

If there are *no* aberrations at all, the resolution of *any* lens (glass, electromagnetic) is customarily defined in terms of

the Rayleigh criterion, which we introduced back in Chapter 1. Rayleigh devised a criterion for resolution which 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 the eyes' ability to distinguish separate images of two self-luminous incoherent point sources. 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 a limiting aperture. This diffraction results in a point being imaged as a disk (called the Airy disk) which has a cross-section intensity profile as shown in Figure 6.13a. This should be familiar to anyone who has encountered basic light optics. Rayleigh stated that if the maximum from one source lies over the first minimum from the other source, as shown in Figure 6.13c, then the overall intensity profile exhibits a dip in the middle at about 80% I_{max}. The eye can discern this dip as two overlapping images, thus indicating the presence of two separate objects. Under these circumstances the distance apart of the two incoherent point sources is defined as the theoretical resolution of the lens r_{th} and is given by the *radius* of the Airy disk

$$r_{th} = 0.61\frac{\lambda}{\beta} \qquad [6.18]$$

Beware! Sometimes the diameter is used (hence the term 1.22 in equation 5.10) because it is the beam diameter which defines image resolution in SEM and STEM. Any standard text on light optics (which we've already referenced) will derive this criterion if you're interested. Now, strictly speaking, we should not use this equation for electron sources because they are not incoherent, and when dealing with true high resolution a different approach is used (see Chapter 28). But for our purposes here, we will be content with this approximation.

Obviously, then we can get higher resolution (smaller r) if we lower λ or increase β. Unfortunately, microscopists often use the expression "higher resolution" when in fact they mean "better resolution" and this can be confusing since the term "higher" is then associated with a lower number. The vacuum is also "higher" if its magnitude is smaller! The improvement in resolution with lower λ accounts for much of the drive toward intermediate and high voltage microscopes since λ decreases with keV, as we saw back in Chapter 1. But 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, and the lens aberrations increase as we increase β.

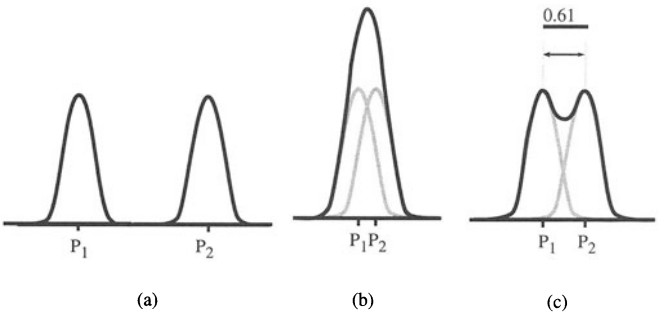


Figure 6.13. (a) The Airy disk intensity profile from two point sources P_1 and P_2 defines the resolution of a lens. In (b) the two Airy disks are so close that they cannot be distinguished, but in (c) the two are separated such that the maximum in the image of P_1 overlaps the minimum in P_2. This is the definition of resolution defined by the Rayleigh criterion.

6.6.B. Spherical Aberration-Limited Resolution (The Practical Resolution)

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 (1972) 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* figure of merit used when investing hundreds of thousands of dollars in a TEM, it is essential that you understand that the definition is not exact. We will follow Hawkes' treatment in detail.

We take the combination of the Rayleigh and spherical aberration disks in quadrature

$$r = \left(r_{th}^2 + r_{sph}^2\right)^{1/2} \qquad [6.19]$$

We can thus find how r varies with β

$$r(\beta) = \left[\left(0.61\frac{\lambda}{\beta}\right)^2 + \left(C_s\beta^3\right)^2\right]^{1/2} \qquad [6.20]$$

Since the two terms vary differently with the aperture collection semiangle β, a compromise value exists when

$$\frac{dr(\beta)}{d\beta} = 0 = -2\frac{(0.61\,\lambda)^2}{\beta^3} + 6C_s^2\beta^5 \qquad [6.21]$$

From this equation the optimum (compromise) value of β is given by

$$\beta_{opt} = 0.77\frac{\lambda^{\frac{1}{4}}}{C_s^{\frac{1}{4}}} \qquad [6.22]$$

(Sometimes, this compromise value is determined approximately by simply equating the equations for r_{th} and r_{sph}. 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} is substituted into equation 6.18, we get a minimum value of $r(\beta)$

$$r_{min} \approx 0.91\left(C_s\lambda^3\right)^{1/4} \qquad [6.23]$$

This is the expression we want; it gives the practical resolution of the microscope.

Typically, the value for r_{min} is ~0.25–0.3 nm, but the high-resolution instruments have r_{min} ~0.15 nm. 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 the 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 ~10^6. 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 leads to expressions of the form of equation 6.22 for β_{opt} and 6.23 for r_{min}. So in many textbooks you will see the use of letters A and B for the constants in these two equations, and often A and B are put equal to unity.

Remember, however, that we ignored any contribution of chromatic aberration. If you have a thick specimen in which most electrons lose 15–25 eV of energy, the chromatic aberration resolution limit given by equation 6.16 will dominate. For example, at 100 keV with β_{opt}, from equation 6.16 the value of r_{chr} is ~2 nm, and you can see all the available information at a magnification of 10^5. Under these circumstances, it doesn't matter what voltage you use or how low your C_s is. If you have thick specimens, then you'll have poor resolution. So how thick is thick? Well, it depends on the voltage and the mean free path for inelastic scatter and many other factors. For good high resolution at 100 keV the specimen should be ~30 nm, while at 300 keV you can probably get away with ~50 nm. A rule of thumb given by Sawyer and Grubb (1987) is that, for biological and polymeric specimens, the resolution limit is about one-tenth the specimen thickness.

6.6.C. 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. Now 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. Also, Reimer (1993) further points out that 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, and so the disk of least confusion is the relevant feature from which to define the 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 (1972), who notes that defocusing the image slightly, to bring the disk of least confusion to the image plane, will indeed lead to a decrease in the value of B 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 B ~ 1 (from the Gaussian image) is "a more prudent choice" (i.e., closer to reality) than B = 0.43 (from the disk of least confusion) even though the disk of least confusion refers strictly to the conditions operating in the TEM.

Beware! Electron microscopists are generally rather careless in their definition of the term (r_{sph}) that they use to describe the effects of spherical aberration on resolution. We use the Gaussian image radius referred back at the object plane, i.e., $r_{sph} = C_s\beta^3$.

Part of this confusion arises out of an uncertainty whether to use the Gaussian image or the disk of least confusion for the definition of r_{sph} and it seems to be a matter of opinion which is more correct. Fortunately, it doesn't really matter too much since, in the end, the choice only alters the value of the constants A and B, which are often approximated to unity. For example, the value of A will depend on exactly which of the several quoted expressions was used for r_{sph}, i.e., if there was 0.25, 0.5, or 1 in front of $C_s\beta^3$. After these various terms are fed into equations and the value of β_{opt} is extracted, A only varies by about ±15%. A small variation in B will occur also, for the same reason.

We have consistently used the radius of the Airy disk and the radius of the aberration error. Obviously, it doesn't really matter whether you use the radius or the diameter, so long as you are consistent. Occasionally, however, you will find the Airy disk *radius* is used in combination with the *diameter* of the disk of least confusion or the Gaussian image, so this also contributes much to the discrepancy between various TEM texts.

6.7. DEPTH OF FOCUS AND DEPTH OF FIELD

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 intense 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 and 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 light 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 irrevelant in SEM and STEMs where we don't use lenses to form the image, so both terms are equivalent.) In TEM both terms are important.

The depth of field, D_{ob}, is the depth of sharpness in *object space.* It's the distance along the axis on both sides of the object plane within which the object can be moved without detectable loss of sharpness in the image. The depth of focus, D_{im}, is the depth of sharpness in the *image plane,* i.e., the distance along the axis on both sides of the image plane within which the image appears sharp (assuming the object plane and objective lens are fixed).

We can derive expressions for these definitions using Figure 6.14. 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 furthest 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. Now we can show that angles α_{im} and β_{ob}, which are both small, are given by

$$\alpha_{im} \sim \tan \alpha_{im} = \frac{\frac{d_{im}}{2}}{\frac{D_{im}}{2}} \qquad [6.24]$$

and

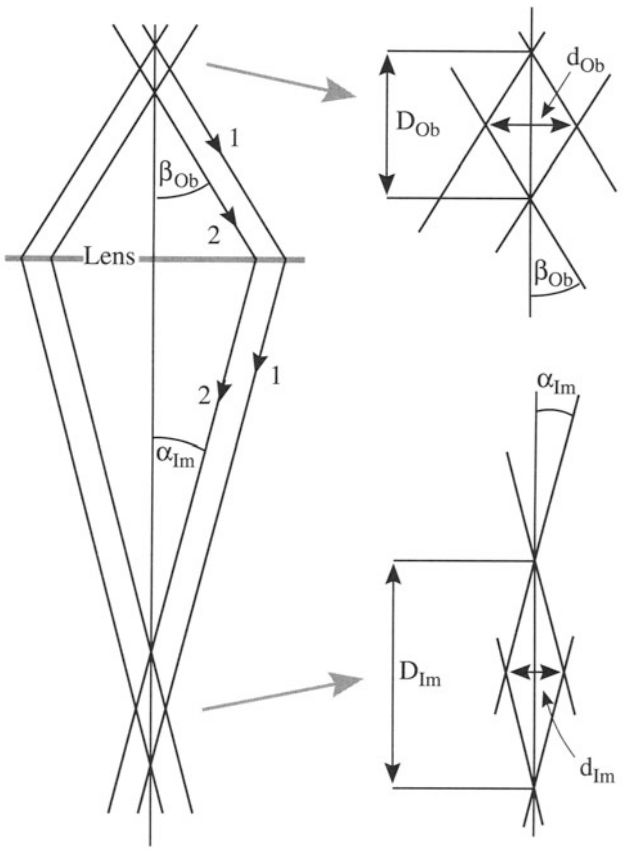


Figure 6.14. The definition of the depth of field and the depth of focus. Rays 1 and 2 represent the extremes of the ray paths that remain in focus when emerging ± $D_{Ob}/2$ either side of the specimen. The same rays define the depth of field over which the image is in focus ± $D_{Im}/2$ either side of the image plane. The resolution in the object is d_{Ob} and that in the image is d_{Im}.

$$\beta_{ob} \sim \tan\beta_{ob} = \frac{\frac{d_{ob}}{2}}{\frac{D_{ob}}{2}} \quad [6.25]$$

The angular magnification is thus

$$M_A = \frac{\alpha_{im}}{\beta_{ob}} \quad [6.26]$$

and the transverse magnification (which we simply call the magnification) is

$$M_T = \frac{d_{im}}{d_{ob}} \quad [6.27]$$

If these two magnifications are related in the usual way by

$$M_T = \frac{1}{M_A} \quad [6.28]$$

then we can say that the depth of focus is given by

$$D_{im} = \frac{d_{ob}}{\beta_{ob}} M_T^2 \quad [6.29]$$

and the depth of field is

$$D_{ob} = \frac{d_{ob}}{\beta_{ob}} \quad [6.30]$$

So we get a much increased depth of focus and field by using small apertures which reduce β. 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 semiangle 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 small ($\sim 10^{-4}$ rad). In a thicker, more strongly scattering specimen, the objective aperture defines the angle and β_o is usually 10^{-2} rad.

For a collection semiangle, β_{ob}, of 10 mrad and a d_{ob} of 2 Å, 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 all be in focus.

If we want to see detail at the 2Å 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 plate (or other recording media) because it would still record a focused image many meters either side of the screen. This explains why we can use a 35-mm camera which often sits below the final projector lens, and still get a focused image with a TV camera well below the plate camera. In fact, the TEM image would be in focus on the floor under the microscope if you projected it there but M_T would be different!

CHAPTER SUMMARY

We've introduced you to the principles of how an electromagnetic lens works, and how we describe its functions via simple ray diagrams. There are two principal operations: either we use the lens to create an image of an object, or we use it to 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 and also used to create fine electron probes (demagnified images of the electron source) at the plane of the specimen in a STEM or SEM. The latter operation is used to create diffraction patterns 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 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.

Points to be wary of when reading about definitions of resolution:
Grundy and Jones (1976), Watt (1985), and Sawyer and Grubb (1987) use the Gaussian image radius referred back at the object plane, as we just did, i.e., $r_{sph} = C_s\beta^3$.
Beware: *Edington (1976) implies, and Thomas (1962), Murr (1970) and Hirsch* et al. *(1977) state, that* $C_s\beta^3$ *is the radius of the disk of least confusion, which it is not, since by definition it must be less than the Gaussian image radius (see Figure 6.11).*
Beware: *von Heimendahl (1980) defines the diameter of the disk of least confusion as* $C_s\beta^3$, *which is also incorrect.*

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.*
Grundy and Jones (1976), Thomas and Goringe (1979), and Sawyer and Grubb (1987) use the conventional definition given here.
Reimer (1993) uses the term "depth of focus" for the "depth of field," a rare inconsistency!
The terms are used interchangeably in SEM because there is no lens between the object and the image.

REFERENCES

General References

Grivet, P. (1972) *Electron Optics,* Pergamon Press, New York.

Hawkes, P.W. (1982) *Magnetic Electron Lenses,* Springer-Verlag, New York. A collection of review articles.

Hawkes, P.W. 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.

Klemperer, O. and Barnett, M.E. (1971) *Electron Optics,* Cambridge University Press, New York.

Specific References

Bradbury, S., Evennett, P.J., Haselmann, H., and Piller, H. (1989) *Dictionary of Light Microscopy,* Royal Microscopical Society Handbook #15, Oxford University Press, New York.

Busch, H. (1927) *Arch. Elektrotechnik* **18**, 583.

Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science,* Van Nostrand Reinhold, New York.

Grundy, P.J. and Jones, G.A. (1976) *Electron Microscopy in the Study of Materials,* E. Arnold, London.

Hawkes, P.W. (1972) *Electron Optics and Electron Microscopy,* Taylor and Francis Ltd., London.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, NewYork.

Jenkins, F.A. and White, H.E. (1976) *Fundamentals of Optics,* 4th edition, McGraw-Hill, New York.

Munro, E. (1974) Proc. 8th Int. Cong. for Electron Microscopy, **1**, p. 218, The Australian Academy of Sciences, Canberra.

Murr, L.E. (1970) *Electron Optical Applications in Materials Science,* McGraw-Hill, New York.

Reimer, L. (1993) *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis,* 3rd edition, Springer-Verlag, New York.

Sawyer, L.C. and Grubb, D.T. (1987) *Polymer Microscopy,* Chapman and Hall, New York.

Thomas, G. (1962) *Transmission Electron Microscopy of Metals,* John Wiley and Sons, New York.

Thomas, G. and Goringe, M.J. (1979) *Transmission Electron Microscopy of Materials,* Wiley Interscience, New York.

von Heimendahl, M. (1980) *Electron Microscopy of Materials,* Academic Press, New York.

Watt, I.M. (1985) *The Principles and Practice of Electron Microscopy,* Cambridge University Press, New York.

How to "See" Electrons

7

CHAPTER PREVIEW

If we are studying the structure of a material, when all is said and done, all we have to show for our expensive electron microscope, hours of specimen preparation, and careful alignment, etc., is an image, a diffraction pattern, or a spectrum. These images and diffraction patterns, 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 result for future reference, perhaps so we can print it for a 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 into visible-light distributions. 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 advances in electronics, and so this chapter will undoubtedly contain anachronisms by the time you read it.

How to "See" Electrons

7

7.1. ELECTRON DETECTION AND DISPLAY

Images and diffraction patterns are different two-dimensional electron-intensity distributions which can be produced by scattering by the same object. 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 diffraction patterns 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. We will briefly discuss the properties of the viewing screen. The manufacturer controls the initial 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 detect the electron signals by various types of electronic detection. 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 to image 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 *digital* scanning images are presented on a fluorescent screen as an analog image. We often refer to this fluorescent screen as the CRT, which is the acronym for "cathode-ray tube" and a relic from the early days of electron physics.

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 image analysis. The signal from any electronic detector can be digitized and electronically manipulated prior to display on the CRT, 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 30. Because of developments in computer technology, there is great interest in recording analog TEM images via a TV camera in order to digitize them and charge-coupled device (CCD) cameras are already available for on-line viewing and processing, particularly of HRTEM images.

In attempting to compare the properties of detection and recording devices we often use the concept of the "detection quantum efficiency" or DQE. If the 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) in order to provide an interface between electrons and our eyes. Any electron display system that we look at relies on CL. 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.

- *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 with impurities 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 the 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 the 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 about 50 μm, although they may be as small as 10 μm for the highest-resolution screens.

As we've seen in Chapter 4, 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 small focusing screen support is made of a heavy metal, such as Pt, to try and encourage 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, carefully selected to absorb the radiation to levels at or below ambient background. In HVEMs this can amount to several centimeters 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 practical hints 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 diffraction patterns. 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 underfocused, 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.

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 (silicon *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 brief comment 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.

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. We call this region 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

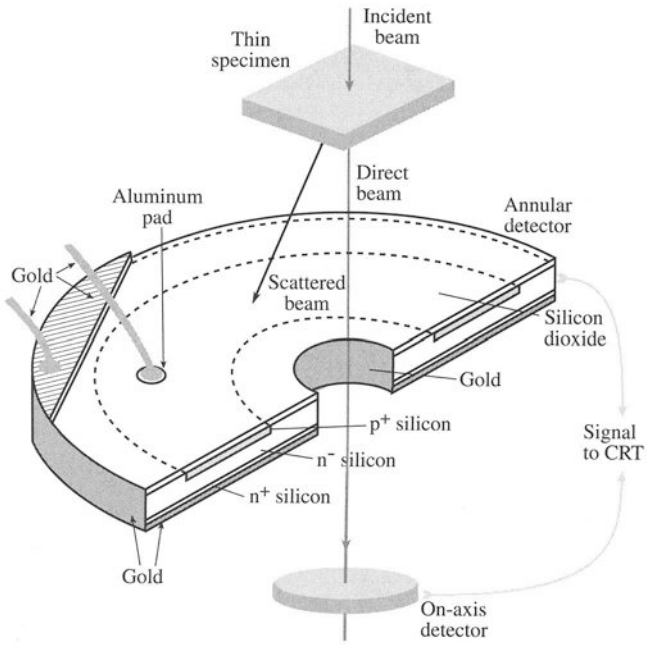


Figure 7.1. Semiconductor detector of the surface-barrier type, shown in a configuration where it would be used to detect high-energy forward-scattered electrons. The direct beam is detected by a small circular detector on the optic axis surrounded by a concentric wide-angle annular detector, which detects any scattered electrons.

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 into the conduction band; we thus create electron–hole pairs. 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 about 1 μm^2. So the semiconductor detector is remarkably responsive to the incoming electron signal. The net result of all this 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 about 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 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, and so 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), and 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, and it is this latter factor which 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 it, but the scattered electrons are very efficiently detected. We thus have a dark-field detector. We can also make detectors that are divided into halves or quadrants. These 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. 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 sig-

nals, 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 the 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 detector, which we will now describe.

7.3.B. Scintillator–Photomultiplier Detectors

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 microseconds 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 set up to detect secondary electrons in a TEM, and the design used in the SEM 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 the column from a thermionic source and reflected from the polished surface of the specimen. If you have a 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 dynodes in the PM. A value of 10^8 is not unusual (compare with 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 room illumination conditions. 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 more susceptible to radi-

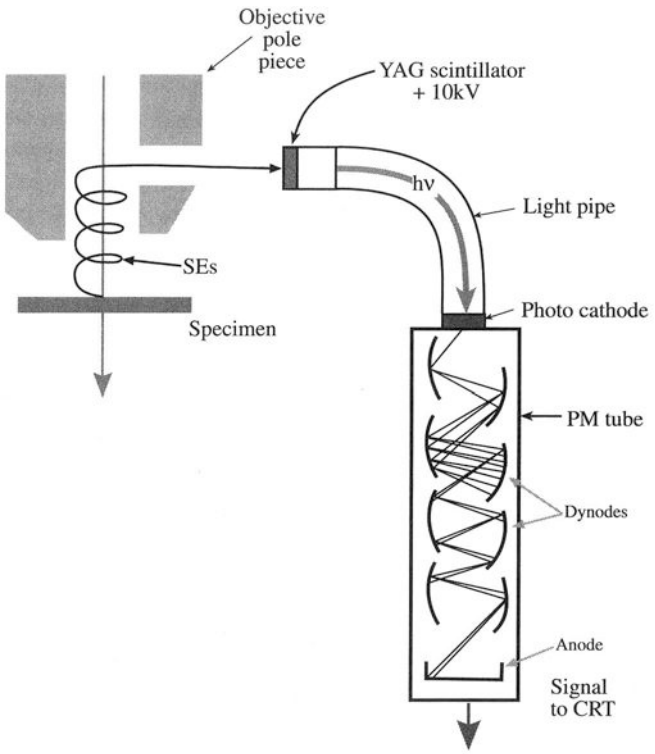


Figure 7.2. Scintillator–photomultiplier detector system in a TEM. SEs from the specimen spiral back up through the polepiece and are accelerated by the high voltage onto the scintillator, generating visible light which travels via fiber optics to a photocathode. There the light is reconverted to electrons. The electron signal is then multiplied by several electrodes in the PM tube.

ation damage, particularly after long-time 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 multidetector configurations. However, plastic scintillators can be shaped to give large-angle collection, such as in the Robinson BSE detector used in many SEMs.
- The energy conversion efficiency of a scintillator is also rather low (about 2%–20%) compared to a semiconductor detector and typically we only get about 4000 photons per incident 100-keV electron, about 7 times less than 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.

7.3.C. TV Cameras and Charge-Coupled Devices

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 doing this for on-line viewing of faint HRTEM images (see Chapter 29), or for recording of dynamic *in situ* events. A standard TV camera is often quite sufficient for this job, although in the US, the TV-image resolution (500 lines/frame) is rather low and a high-resolution camera (1000 lines/frame) is preferred. Video storage and display again requires higher resolution than standard VHS video formats if you want to get the most out of your images. The best TV cameras are CCDs, which are replacing standard plumbicon tube TV cameras.

CCDs are MOS devices that store charge generated by light or electron beams. CCD arrays consist of thousands or millions of pixels which are electrically isolated from each other by creating potential wells under each CCD cell so they can accumulate charge in proportion to the incident beam intensity, as shown in Figure 7.3A. The cells currently can be as small as 6 μm. To create a picture we have to read out the array, which can be done by changing the applied potentials to transfer the charge serially

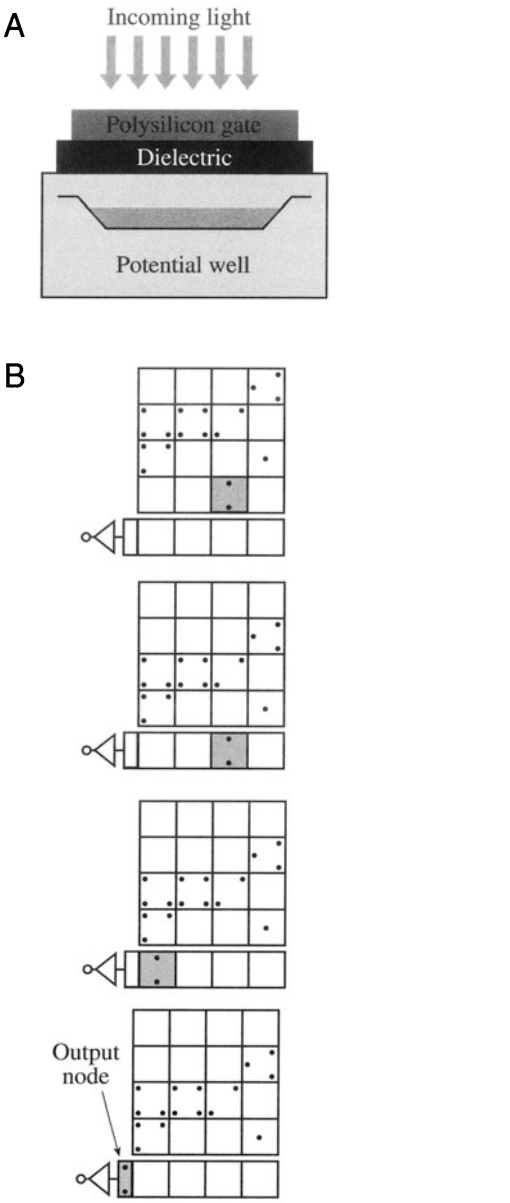


Figure 7.3. (A) A single cell in a CCD array showing the storage of charge in the potential well under one pixel. If we vary the applied potential to rows of pixels in sequence, as in (B), one pixel row is shifted to the parallel register, and is read out pixel by pixel, after which the next row is moved to the parallel register, and so on. The stored charge in each pixel is thus fed into an amplifier and digitized.

from each potential well along a line into an output amplifier as shown in Figure 7.3B. Once all the cells are empty the array can be re-exposed. We thus have a frame time for reading the display which can be as short as 0.01 s, well below standard TV rates (0.033 s). Rather than serial readout, it is also possible to have full frame CCDs in which the whole frame is transferred to an adjacent storage array leaving the main array free to collect a new signal flux. We get an analog output, i.e., a charge current, which we then

digitize, usually through an 8-bit A to D converter to give 256 gray levels.

CCD arrays have several advantages:

- When cooled, they have a very low noise and a good DQE (>0.5) even at low input signal levels.
- The dynamic range of a CCD is high, making it ideal for recording diffraction patterns which can span an enormous intensity range.

The major disadvantages are the speed and the expense. These devices can also be used as two-dimensional arrays for parallel-collection electron energy-loss spectrometers rather than the more conventional one-dimensional silicon diode arrays described in Chapter 37.

7.3.D. Faraday Cup

In conventional TEM there isn't much need to know the beam current, but 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.

A Faraday cup is a black hole for electrons.

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. Fortunately, 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 (Figure 7.4B). These cups are not ideal because they don't trap all the electrons. 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.

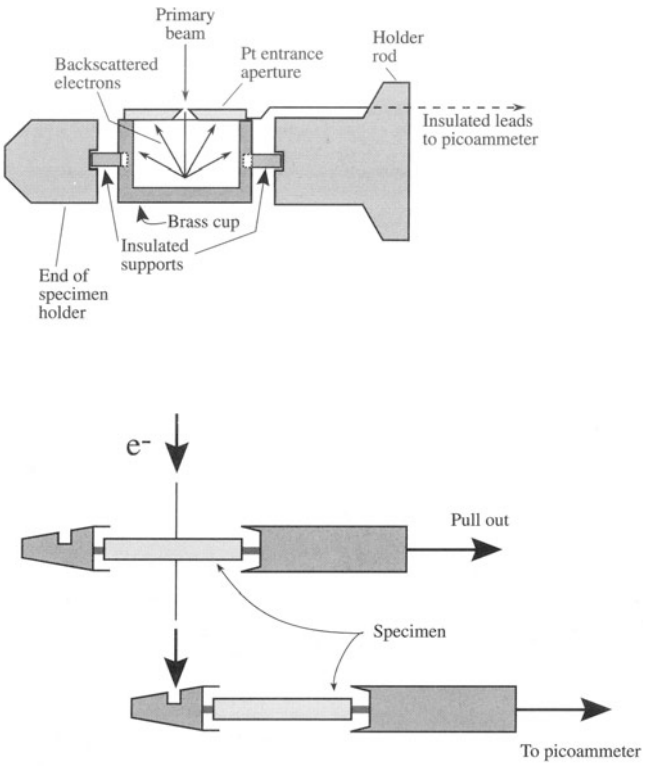


Figure 7.4. (A) Schematic diagram of a Faraday cup in the end of a side-entry specimen holder. The entrance aperture has to be found using SEs or BSEs. In (B) the holder is retracted slightly so the electrons fall into a cup on the tip of the holder. The electron current is measured as it goes to ground 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 the 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 it is about 0.15. It is also possible to deflect the beam onto the last beam-defining diaphragm and measure the current via an insulated feed-through (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 are the most common TEM images, and the BSE and SE signals from the beam-entry surface of the specimen.

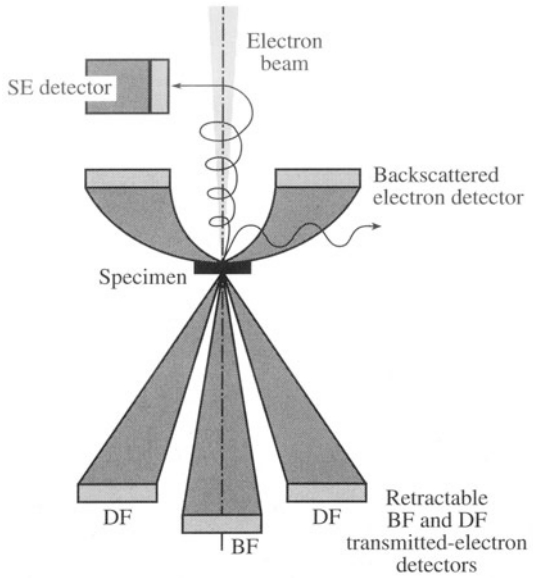


Figure 7.5. The various electron detectors in a STEM. Scintillator–PM detectors are invariably used for SE detection and semiconductor detectors for the BSE. The on-axis and annular forward-scattered detectors may be either type, depending on the microscope.

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 and a scintillator–PM system is required. 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., where 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.

Sometimes we examine specimens which themselves exhibit cathodoluminescence under electron bombardment. We'll discuss why this is done in Chapter 31. A mirror is used to focus the light into a scintillator–PM system, and one such design is shown in Figure 7.6. This setup effectively prevents detection of all 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.

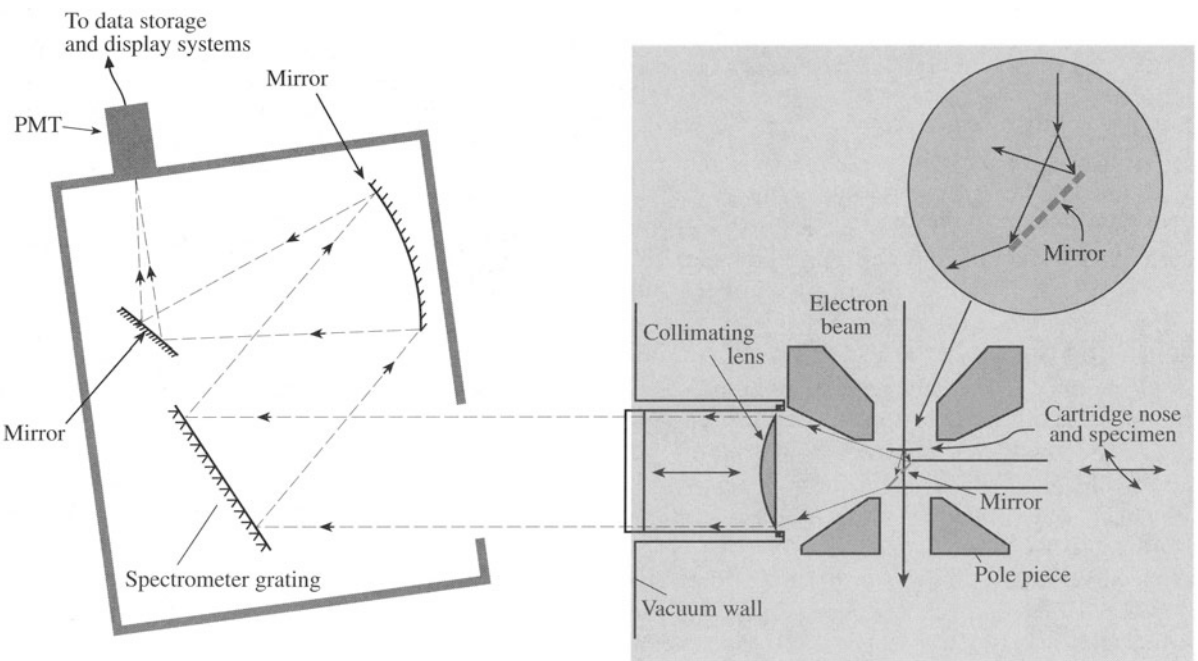


Figure 7.6. Cross section of a mirror detector below a thin cathodoluminescent specimen that collects light and focuses it into a spectrometer–PM system. The CL signal is usually very weak and so the detector has to be as large as possible, and it takes up much of the free volume in the TEM stage.

7.5. IMAGE RECORDING

7.5.A. Photographic Emulsions

Although film is the oldest recording medium, it still retains sufficient advantages that we continue to use it in virtually all TEMs. Photographic emulsions are suspensions of silver halide grains in a gel. Electrons strike the halide, ionize it, and transform it to silver. The emulsions are usually supported on a polymer film or (very rarely) glass plates. Unlike polymer film, glass plates do not outgas and do not shrink during prepumping or processing. However, glass plates are heavy and occupy an enormous volume compared to polymer film, so you can't load as many into the microscope at once and you need more storage space in general. Also, Murphy's law means that your best plates will invariably break because you spend more time looking at those than the others. Most microscopists use film rather than plate, but we still sometimes call them "plates."

You do have a choice of photographic emulsion, just as you do for your own camera. 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 the faster film grain sizes 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 mean that the overall dose to the specimen is minimized. The only time you may need to use the 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 10 cm × 10 cm 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). CCDs will eventually replace the photographic plate because the device size is already comparable with silver halide grains. As we've already noted, CCDs have high dynamic range. In fact, the latest CCDs boast > 2.5×10^7 pixels, and while the cost is currently prohibitive, it is falling as is the cost of storing the data.

We use instant film for recording CRT images in STEM mode. Different film speeds are available. For example, Polaroid instant film comes in positive–negative form (Type 55) or just a positive print (Type 52), which is also slightly faster. Instant color is also available for recording false-color images from a digital CRT display (such as X-ray maps, as we discuss in Chapter 36). The main drawback to instant film is its expense. Although you can see the image "instantly," you still need to coat the print chemically to prevent degradation of the image with time.

7.5.B. Other Image Recording Methods

Digital images can be stored and retrieved magnetically, for example on floppy disks and hard disks or optically on compact disks. These devices are cheaper and easier to use than photographic recording and images on an optical disk will not degrade with time even after years of storage. To present a stored image for publication you still have to print it in some way and photographic methods still dominate. However, alternative devices such as thermal printers, laser printers, and image plates are approaching the quality required for published images.

So currently the photographic method still dominates both in TEM analog recording and hard copy output from STEM or any digital imaging system. If we can remove photographic film from the TEM it will be a major improvement, because the absorbed water degrades the vacuum.

> 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 pressure of hydrocarbons and water vapor in the instrument, and in turn causes contamination of the specimen. We are often caught in a compromise between drying the film before putting it in the TEM (prepumping) and avoiding the static electricity discharge which can cause white lines on your film if it is too dry. Anything that gives you a high-quality print from digital storage and doesn't require instant film recording from a CRT display will also make the microscopist's life considerably easier and microscopy cheaper.

7.6. COMPARISON OF SCANNING IMAGES AND STATIC TEM 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 the 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 CRT, and this limits the amount of information in the image. Typically, the viewing CRT 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. The recording CRT may have up to 2×10^3 lines, giving a total of 2×10^6 pixels in the 10 cm $\times$ 10 cm final image. In contrast, as we just noted, a TEM image recorded directly onto photographic emulsion will have a higher information density, with more than 10^7 pixels of information available in the same 10 cm $\times$ 10 cm 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 CRT) 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 time means that, almost invariably, STEM images are poorer in quality than static TEM images. Only in FEG STEMs does the picture quality compare with analog TEM images.

CHAPTER SUMMARY

The TEM is still in the age of analog images. We look at fluorescent screens and CRTs and we record our pictures on photographic film. However, the whole area of electron detection is in a state of rapid flux as electronic systems develop. 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. 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 diffraction patterns will eventually be replaced by digital methods.

REFERENCES

General References

Chapman, J.N., Craven, A.J., and Scott, C.P. (1989) *Ultramicroscopy* **28**, 108.

Knoll, G.F. (1989) *Radiation Detection and Measurement,* John Wiley and Sons, New York.

Nicholson, W.A.P. (1981) *J. Microsc.* **121**, 141.

Reimer, L. (1985) *Scanning Electron Microscopy,* Springer-Verlag, New York.

Yang, E.S. (1978) *Fundamentals of Semiconductor Devices,* McGraw-Hill, New York.

Pumps and Holders

8

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 the parts that, if you are not careful, can seriously degrade the quality of the information you generate. 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 that you can degrade the vacuum.

The vacuum in the stage of a typical TEM is ~10^{-5} Pa, which compares with atmospheric pressure of ~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. Through the holder you have to transmit all the experimental variables that you want to inflict on your

specimen. The most basic requirement is that you should be able to move the specimen laterally to look at different areas. In addition we'll describe how you can tilt, rotate, heat, cool, strain, and bias the materials that you are studying. In addition to transferring useful external variables to the specimen, the holder also transmits vibrations, drift, and contamination to the specimen and may be a source of X-rays that can confuse any microanalysis 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.

Pumps and Holders

8

8.1. VACUUMS

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 send coherent, controlled electron beams through the atmosphere, 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 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. 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. For a full exposition of vacuum technology for TEMs, read Bigelow (1995) or the equally informative user's guide by O'Hanlon (1980).

First of all, a word on units—which as usual are in disarray. The SI unit of pressure is the pascal (Pa); 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. One bar is atmospheric pressure (~760 Torr) and is equivalent to ~10^5 Pa and so

One Torr is ~ 130 Pa, 1 Pa is 7.5×10^{-3} Torr.

We'll mainly use Pa, but since Torr is still very common terminology, we'll occasionally put highly 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 can divide vacuums into rough, low, high, and ultrahigh. Pressures between 100 and 0.1 Pa (~1 and 10^{-3} Torr) are rough vacuums, 0.1–10^{-4} Pa (~10^{-3}–10^{-6} Torr) are low, 10^{-4}–10^{-7} Pa (~10^{-6}–10^{-9} Torr) are high vacuums (HV). If the pressure is <10^{-7} Pa (~10^{-9} Torr) you have an ultrahigh vacuum (UHV). These are approximate, not standardized definitions. A typical modern TEM has a pressure inside the column of ~10^{-7} Torr (1.3×10^{-5} Pa), which is in the HV range. UHV TEMs operate below 10^{-7} Pa and the gun region of an FEG TEM operates at ~10^{-9} Pa (10^{-11} Torr). To get an electron beam inside the TEM that is not scattered by the air molecules in the column, the pressure has to be < ~0.1 Pa and 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.

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 under repair or service. 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

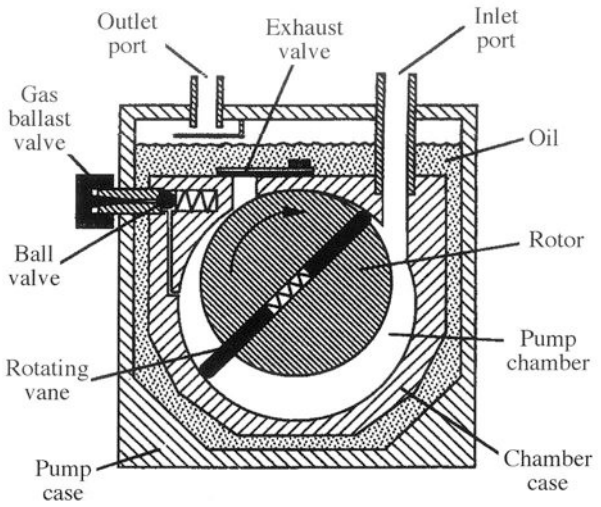


Figure 8.1. A mechanical pump for roughing vacuums. The eccentric motion of the pump creates a vacuum in the RH side when it rotates and the vacuum sucks air into the inlet valve. As the cylinder rotates further, it cuts off the inlet and forces the air through the outlet on the LH side, creating a vacuum 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.

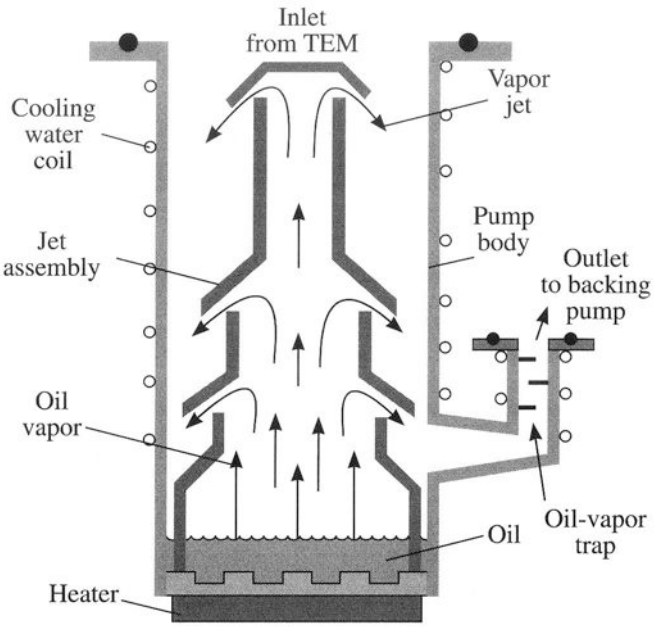


Figure 8.2. Principles of diffusion pump operation. A heater plate at the base of the pump boils synthetic oil. The expansion of the oil vapor on boiling creates a pressure, which forces the vapor up the central column and out of several holes. The stream of oil vapor pulls gas molecules out of the top 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.

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 ~10^{-1} Pa (~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 be well 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.

8.3. HIGH/ULTRA-HIGH 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 hundreds of 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 water cooling and burn-out 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™. 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 ~10^{-1} to ~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.

8.3.B. Turbomolecular Pumps

Turbomolecular pumps, or turbopumps, as the name implies, use a turbine to force gases from the microscope.

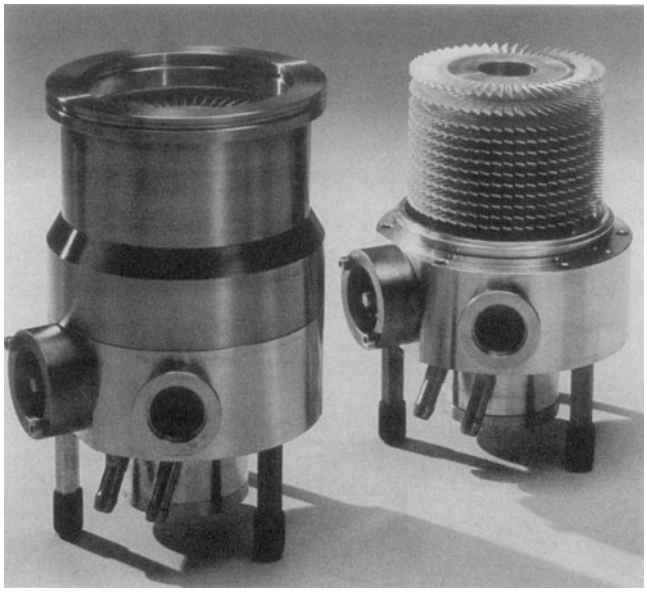


Figure 8.3. A turbopump (with and without its casing) which is nothing more than a small turbine that rotates at high speed. Like a jet turbine it pulls air in at the front end and forces it out of the back. The blades are designed like airfoils to enhance the flow of gas through the system.

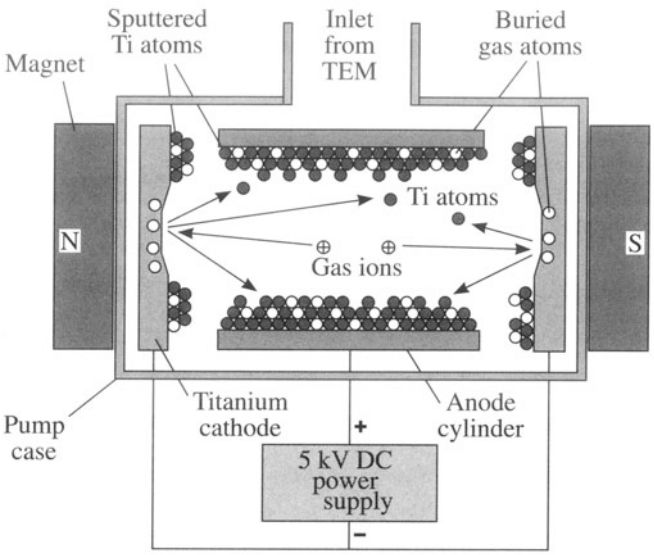


Figure 8.4. Schematic diagram showing how ion pumps trap ionized gas atoms by layers of Ti atoms at electrodes. Once trapped, the ions cannot escape until the pump is turned off.

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 they would destroy the image resolution. The turbopump can start (slowly) at 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.

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, and these 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 system, 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.

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 out 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 (to your specimen) for condensation of residual components in the vacuum.

Ion pumps and cryopumps are trapping pumps. They keep the air molecules within them and release them when turned off or warmed up, respectively.

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.

8.4. THE WHOLE SYSTEM

As shown schematically in Figure 8.5 the TEM has 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 a pressure below which the HV/UHV pumps can 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).
- #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.

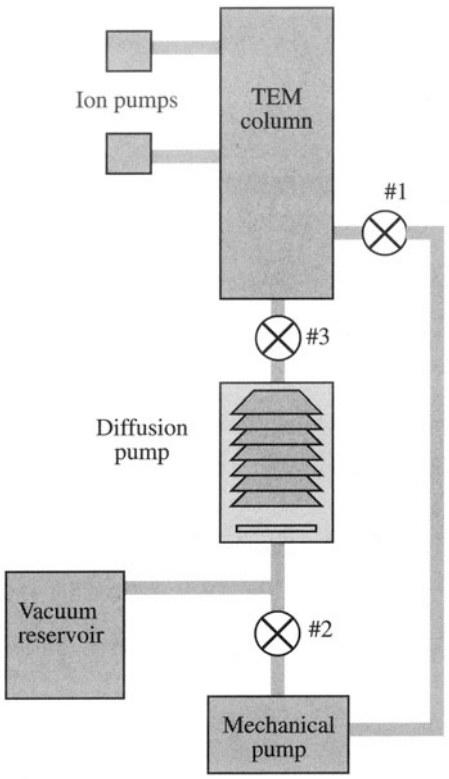


Figure 8.5. The principles of the TEM vacuum system. Often, the console display on the TEM will show a similar diagram. The mechanical pump can pump the column directly or back out the diffusion pump, which is connected directly to the base of the microscope. Ion pumps are often interfaced directly to the stage and gun areas. Computer-controlled valves separate the pumps from the column and from each other.

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 also continuously backed by the mechanical pump. Alternatively, a vacuum reservoir is attached 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., the diffusion pump is 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 sig-

nificantly 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 got to pass through it and the diffraction pattern 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 Francois Rabelais said in 1534, and that's the reason why the pumps have to keep pumping: the TEM leaks. But some leaks are too large for the pumps to handle, and then the instrument performance degrades. For example, you might not be able to run the electron gun, so your TEM is useless. Under these circumstances, you have to find the leak, cure it, and repump out the instrument (this is usually a job for your service engineer). 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 (e.g., the stage airlock, which sees a lot of use, is a common point of failure). The small helium 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, 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 bake 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 can be a source of contamination, particularly residual hydrocarbons from the pump oil which 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 appropriate traps on the pumps and synthetic oils.

However, even if you've paid dearly for a clean vacuum system, contamination often occurs and 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 ashing of the specimen holder and specimen prior to insertion in the TEM has proven a very successful way to ensure a clean specimen, but this is expensive. 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 ashing 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 in Minnesota).

There will be more about this in the ensuing sections on specimen holders.

In addition to the specimen, you personally can be a major source of contamination unless you 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). 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, you 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 TEM, which is already appearing.

The cold trap, cold finger, or cryoblades are a critical part of the stage. Ideally, this cold finger will completely surround 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.

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, the new 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 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 mm 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 pole-piece (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.

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. The traditional design is shown in Figure 8.6. The key parts of the holder are:

- *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 gener-

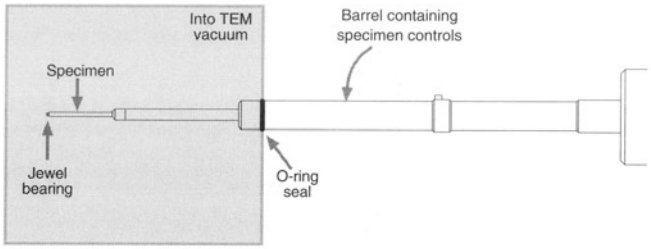


Figure 8.6. Principal parts of a side-entry holder that is held in the goniometer stage. The specimen is clamped into the cup at the end of the rod. A small jewel at the end of the rod (usually sapphire) fits into another jewel bearing in the stage to provide a stable base for manipulating the specimen. The O-ring seals the end of the holder inside the vacuum. Manipulating the specmen is accomplished from outside the column via controls within the rod.

ation of X-rays that would interfere with microanalysis.

- *The clamping ring or screw,* which holds the specimen in the cup. This ring, 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, since brittle disks may break as you are loading them. There are two kinds: 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 makes Be spring clips!

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, all TEMS up to 400 kV use side-entry holders.

Another drawback of such holders is that the bore of the objective lens must be asymmetric, which actually limits the ultimate resolution by constraining the lens designer. Figure 8.7 shows a schematic diagram of such a holder.

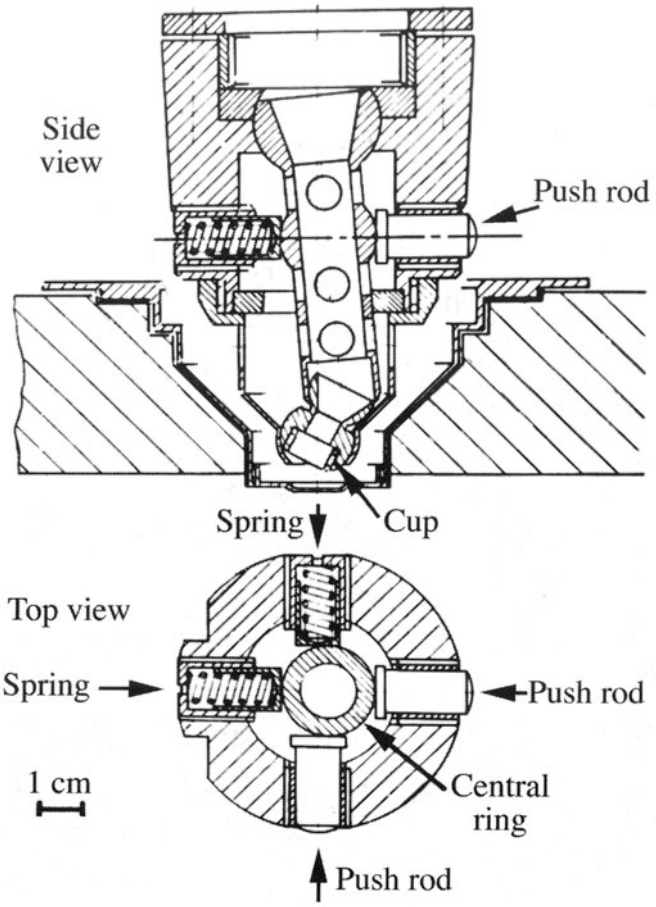


Figure 8.7. Cross section of a top-entry holder. The cartridge has a cone shape which fits into the tapered bore of the objective lens polepiece. The specimen sits in a cup at the base of a column through the cone down which the incident beam travels. Simple manipulations such as tilting or rotating 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 the upper diagram, push rods are pressed against springs in two orthogonal directions, displacing a ring around the column (lower diagram), thus tilting the specimen cup.

8.10. DIFFERENT TYPES OF HOLDERS

One feature of TEMs which may surprise you if you are a new user is the wide variety of holders which is available. Figure 8.8 shows illustrations of some 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.

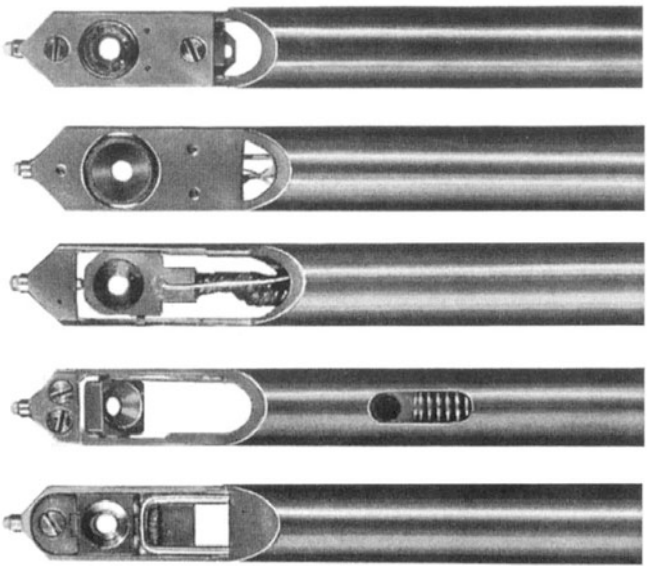


Figure 8.8. Examples of different designs for the side-entry holder. From the top, they are: a rotation holder, a heating holder, a cooling holder, a double-tilt holder, and a single-tilt holder.

- *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's great for ceramics. Different retainers can be substituted for the clamp 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 modern TEMs, specimen exchange is relatively quick, except in UHV instruments where the multiholder is probably more valuable.
- *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 31 for more about these techniques). The bulk specimen is larger than the traditional 3-mm disk (usually ~10 mm x 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).

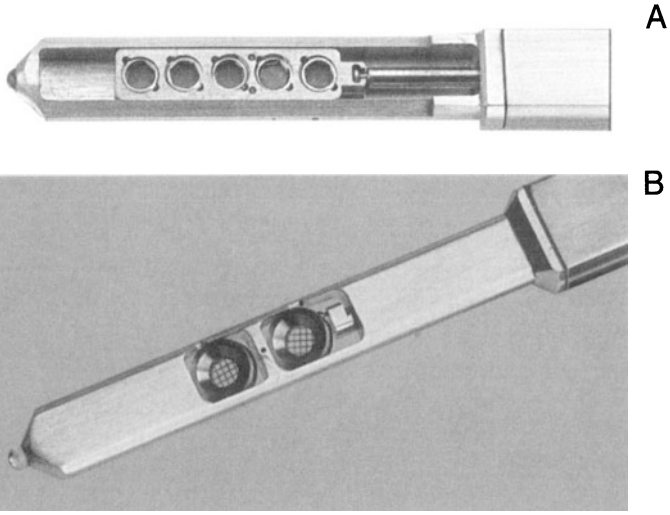


Figure 8.9. Multiple-specimen holders. (A) Five-specimen single-tilt and (B) two-specimen double-tilt.

So don't always think that you are limited to 3-mm 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

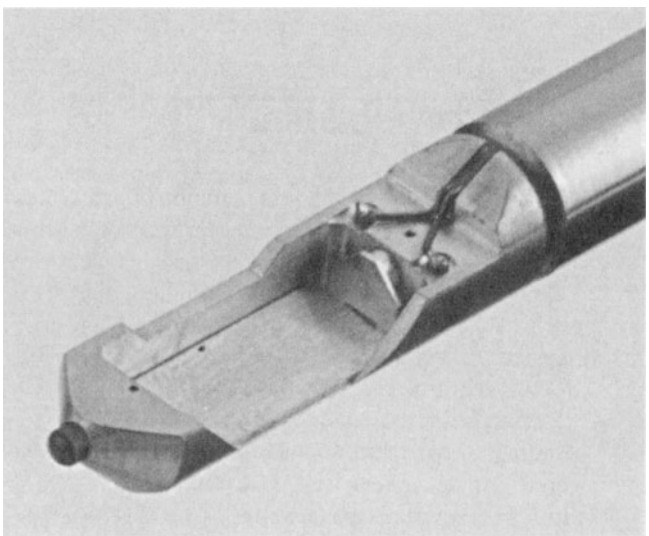


Figure 8.10. A bulk holder for large specimens.

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 select the tilt axis. This holder lets you do just that and is particularly advantageous for the side-entry holder, since the tilt axis is then 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.
- *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 (both high and low T_c) and ideal for polymers or biological tissue. However, you should remember that the cold holder can also act as a small cryopump, 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.
- *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 and then applies a load to one end, via a load cell or screw-thread mechanism, as shown in Figure 8.11. The sample is often in the shape of a small tensile specimen and 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.
- *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 specimen. 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

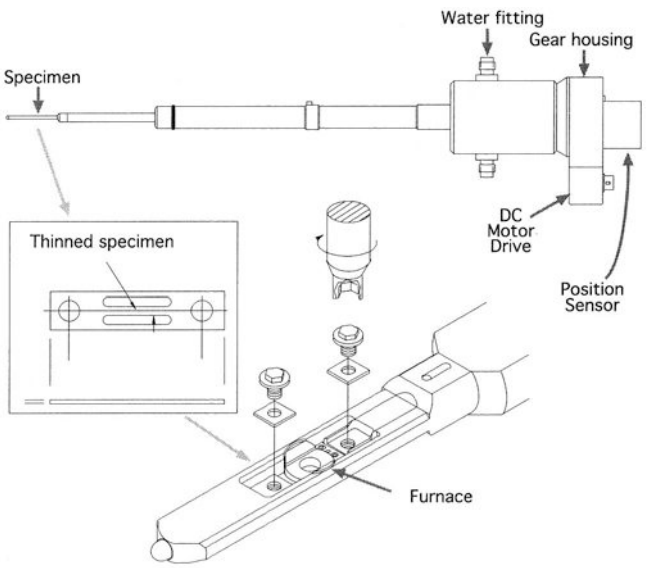


Figure 8.11. A side-entry combined straining and heating holder. The specimen looks like a miniature tensile specimen (inset) and is clamped at either end by hex screws. There is a screw-thread arrangement for pulling the specimen contained within the rod. The furnace surrounds the central thin portion of the specimen.

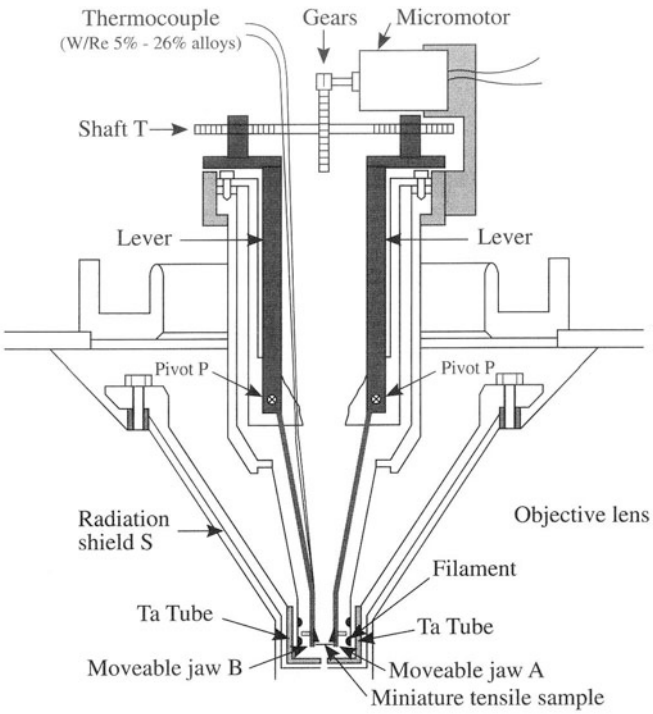


Figure 8.12. A top-entry, heating-straining holder which can be used at temperatures up to 2300 K in a 3-MV HVEM.

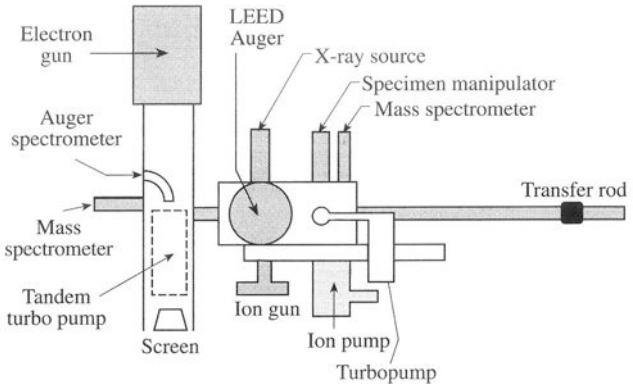


Figure 8.13. Schematic diagram showing the Hitachi H9000 UHV TEM. This instrument is equipped with a prechamber with LEED, Auger, and an ion gun which can be used to clean the specimen, allowing UHV surface analysis to be carried out on the TEM specimen. The holder has to transfer the specimen through a prepump chamber where it is ion-cleaned before going into the column.

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 31.12). 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-straining holder which is reported to be capable of operating at temperatures up to 2300 K (Komatsu *et al.* 1994). The heat is provided by a coaxial Ta tube which 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 on into the TEM column for transmission studies. A similar prechamber has been used elsewhere to provide a method to clean the sample before growing thin films on the surface 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.

The article by Valdrè and Goringe (1971) gives a detailed description of several TEM holders.

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; for example, if you don't bother to prepump your film, and if you handle the specimen holder without gloves. 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!

With the range of holders available today, you can conduct many standard 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 remains tilting in two orthogonal directions to orient different crystal planes parallel to the electron beam. You'll understand why this is important after you've finished reading Parts II and III.

REFERENCES

General References

Watt, I.M. (1985) *The Principles and Practice of Electron Microscopy,* Cambridge University Press, New York. See Appendix 1.

Bigelow, W.C. (1995) *Vacuum Methods in Electron Microscopy,* in the series *Practical Methods in Electron Microscopy,* **15** (Ed. A.M. Glauert), Portland Press, London. An essential reference.

O'Hanlon, J.F. (1980) *A User's Guide to Vacuum Technology,* John Wiley and Sons, New York.

Specific References

Butler, E.P. and Hale, K.F. (1981) *Dynamic Experiments in the Electron Microscope,* in the series *Practical Methods in Electron Microscopy,* **9** (Ed. A.M. Glauert), Elsevier, Amsterdam.

Komatsu, M., Mori, H., and Iwasaki, K. (1994) *J. Am. Ceram. Soc.* **77**, 839.

Valdrè, U. and Goringe, M.J. (1971) in *Electron Microscopy in Materials Science* (Ed. U. Valdrè), p. 208, Academic Press, New York.

The Instrument

9

CHAPTER PREVIEW

We've introduced all the essential components of the TEM. Now it's time to see how the guns, lenses, and detectors are combined to form the microscope. Just as we do for the visible-light microscope (VLM), it's convenient to divide the TEM up 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 for TEM imaging and diffraction, while the second is used for scanning (STEM) imaging, microanalysis, and microdiffraction.

The objective lens/stage system is the heart of the TEM. The critical region usually extends over less than 1 cm along the length of the column. Here is where all the beam–specimen interactions take place and here we create the bright-field and dark-field images and selected-area diffraction patterns (SAD) that are the fundamental TEM operations. Likewise, it is here that we manipulate a scanning beam to form STEM images and diffraction patterns. The quality of the image formed by the objective lens controls the resolution of the image that you view and record. We therefore often say that the objective lens is the most important lens. The reason is not simply because the objective lens forms the first image, but rather because the specimen must be placed so close to the center of this short-focal-length lens that it is impossible to make a perfect lens (or even a very good lens by visible-light standards).

The imaging system uses several lenses to magnify the image or the diffraction pattern produced by the objective lens and to focus these on the viewing screen. We'll refer to the magnifying lenses as the intermediate and diffraction lenses and the final lens as the projector lens (it projects an image on the viewing screen). Alternatively, an electron detector coupled to a TV/CRT can be used to display the STEM image. So all TEM operations involve observing the electrons on a viewing screen of some form, with or without a specimen in place. In many modern-day TEMs you will have a button for focus, another for magnification, and another for diffraction (or a slide on the computer screen).

The purpose of this chapter is to go through the principal functions of each of the three components and give you some feel for what is happening in the microscope when you "press the button." The more you understand the operation of the TEM, the better you can be sure that you are getting the most out of the instrument.

The Instrument 9

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. We can think of these two cases as the equivalent of wide-field illumination 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 illumination system which consists of two or three condenser lenses (which we'll call C1, etc.). We will discuss the two different ways to use the illumination system: we'll refer to these as forming a parallel beam (it is almost never truly parallel) or a convergent beam.

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 a demagnified image of the gun crossover. In the case of a thermionic source, the original crossover may be several tens of micrometers across, and this is demagnified by an order of magnitude or more: in the case of an FEG, the source size may be less than the desired illumination area on the specimen so it may be necessary to magnify the crossover—the condenser lenses don't always condense! To produce a parallel beam you adjust the C2 lens to produce an underfocused image of the C1 crossover.

> Remember that the convergence angles (α) are so small that the ray diagrams are drawn with highly exaggerated angles, and 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.

In TEMs used for generating the very small electron beams we need in STEM and AEM, the upper polepiece, of the objective lens is also used to control the beam hitting the specimen as shown in Figure 9.1B. Now the C2 lens is focused to produce an image (of the crossover) at the front focal plane of the upper objective polepiece, which then generates a broad parallel beam of electrons incident on the specimen.

> Consider the question: How is this argument consistent?

When the beam is parallel, it is as coherent as possible. We'll see in Chapter 18 that parallel illumination is essential to get the sharpest diffraction patterns as well as the best image contrast. (It is also usually assumed in the interpretation of our images that the beam is parallel.) Usually you underfocus C2 until the illuminated area on the specimen fills the viewing screen. A higher magnification means strengthening C2 so the beam illuminates less of the specimen (so you see, it isn't really "parallel," just not very "convergent").

In the parallel-beam TEM mode, 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 falling on 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

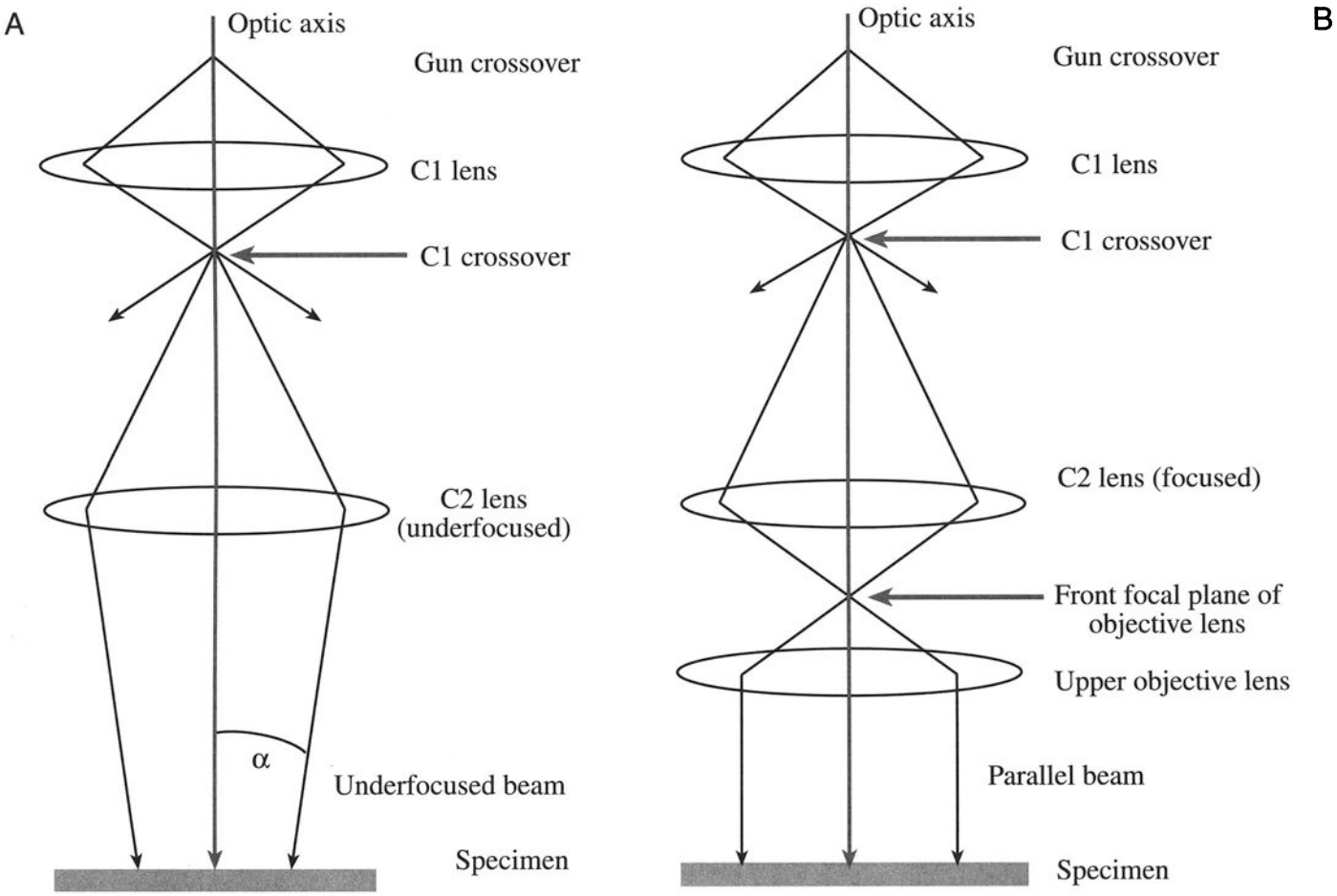


Figure 9.1. Parallel-beam operation in the TEM (A) using just the C1 and an underfocused C2 lens and (B) using the C1 and C2 lenses to image the source at the front focal plane of the upper objective lens.

focused rather than underfocused, and you form an image of the C1 crossover at the specimen, as shown in Figure 9.3. This is the condition under which you can view the source image to adjust its saturation or to measure the dimensions of the beam. 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. 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 for poor transmission through a thick specimen.

There are times when we need to deliberately create a focused convergent beam at the specimen. We then use the other principal way to operate the illumination system: the convergent-beam mode. When you use this mode you won't immediately see a useful image of your specimen; the convergence destroys the parallelism and 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.

> The convergent beam is a probe. We use such a probe when we want to localize the signals coming from the specimen, as in microanalysis or convergent-beam (also known as micro or nano) diffraction.

Now, unless you have an FEG, it isn't possible to use just the C1 and C2 lenses as in Figure 9.3 to converge the beam to as small a probe as you would like (<10 nm). This is because the C1 and C2 lenses can't demagnify the gun crossover sufficiently. So the usual solution is to con-

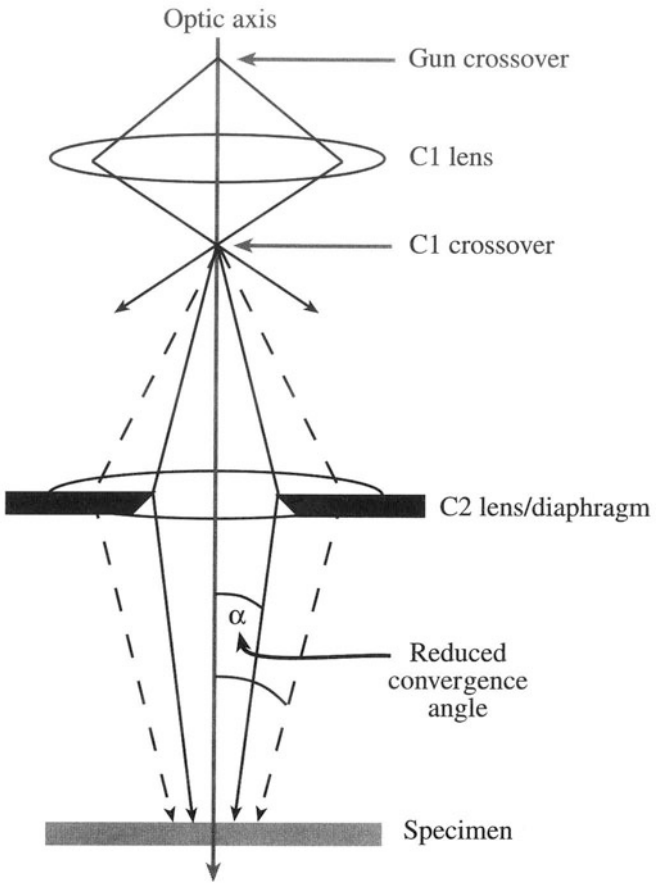


Figure 9.2. Effect of the C2 aperture on the parallel nature of the beam: a smaller aperture creates a more parallel beam.

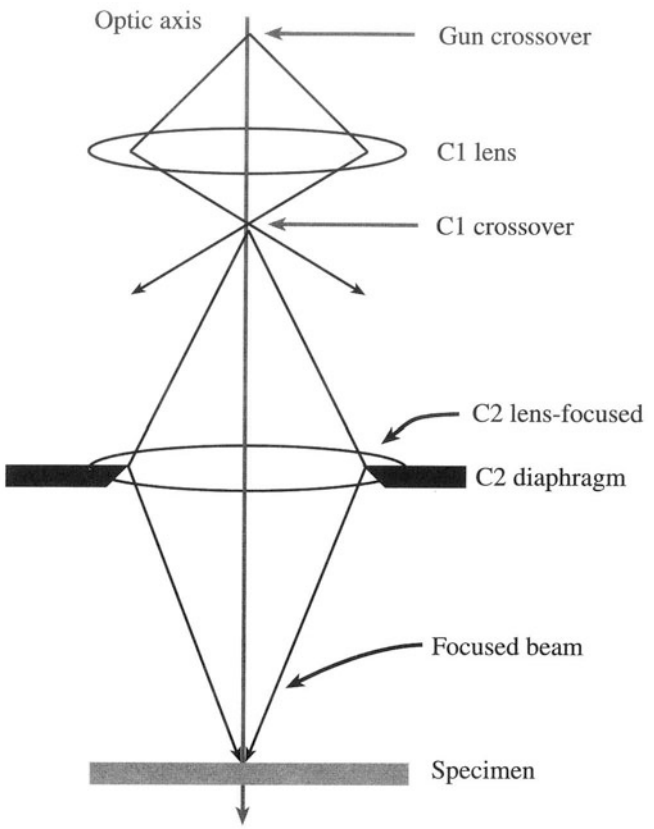


Figure 9.3. A focused C2 lens illuminates a small area of the specimen with a nonparallel beam.

vert the upper polepiece of the objective lens into a third condenser lens, which we then call a condenser-objective lens (C3). We make the upper part of the objective lens much stronger than usual and weaken C2 or turn it off, as shown in Figure 9.4. 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 (v) is much less than the object distance (u), which gives a large demagnification of the C1 crossover (see equation 6.2). From Figure 9.4 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 that the correct choice of C2 aperture is important in convergent-beam electron diffraction (CBED) and also in defining the exact dimensions of the probe for X-ray microanalysis (see Chapter 36).

The role of C1 here is fundamentally different from its role in parallel-beam TEM; now the C1 lens is used directly to form the probe because we adjust it to change the probe size at the specimen. As shown in Figure 9.5, 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 v, thus lengthening u for the probe-forming C3 lens and therefore increasing the C3 demagnification. When convergent probe TEMs were first constructed, it was not possible to design a C3 lens that would give both a parallel and a convergent beam with the same polepiece. This problem was overcome in the mid-1970s by the introduction of an extra lens between C2 and C3 (not shown in the diagrams) and this auxiliary lens is now standard on TEMs that also operate as STEMs.

9.1.C. 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 microanalysis). Similarly, there are times when we need to tilt the beam off axis so it impinges on the specimen at a specific angle (e.g., for centered dark-field imaging using a specific diffraction spot which we describe in Section 9.3.C). Ray diagrams to explain translating and tilting are shown in Figures 9.6A and B. Both operations are accomplished by varying the

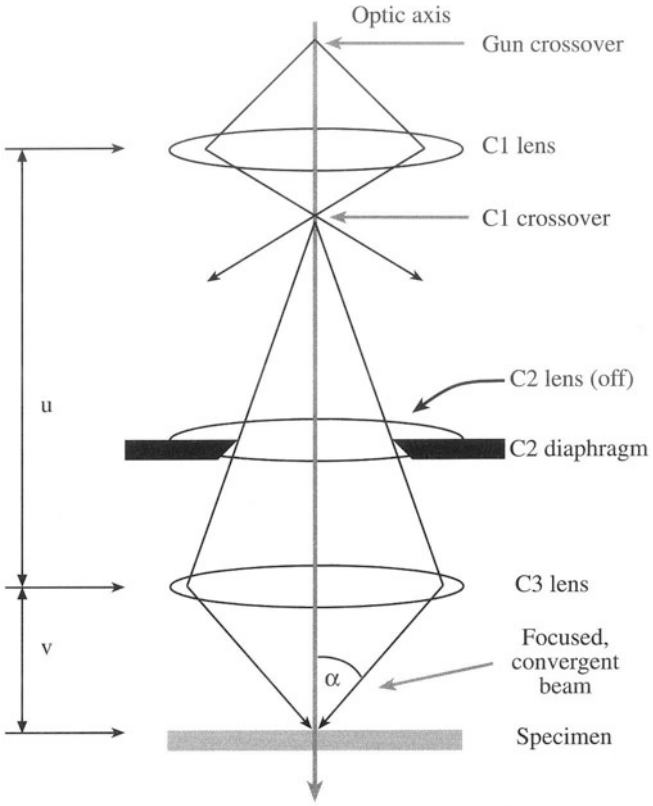


Figure 9.4. Use of the objective polepiece as a third condenser lens (also called a condenser-objective, or C3, lens) gives the smallest possible probe and large convergence angles. The large *u/v* ratio gives the maximum demagnification of the image of the gun crossover.

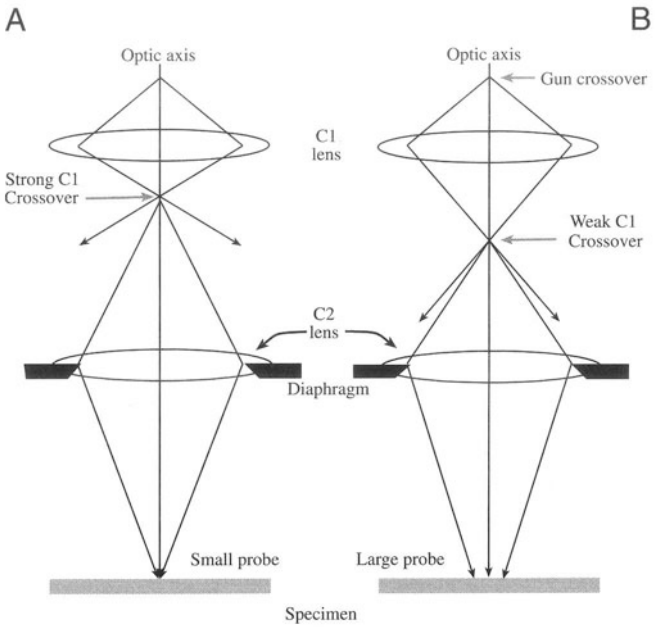


Figure 9.5. Effect of the C1 lens strength on probe size: a stronger C1 lens (A) results in greater demagnification by any subsequent lens (C2 or C3), giving a smaller electron beam at the specimen. A weaker lens (B) gives a broader probe.

current through potentiometers, which we'll call *scan coils*. We use these scan coils (of which there are several in the column) to apply a local magnetic field to deflect (rather than focus) the beam. To translate the beam we use deflector scan coils. To tilt the beam just before it reaches the specimen we use tilt scan coils situated between C2 and C3.

When we create a scanning beam for STEM imaging, the beam must always move parallel to the optic axis. 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 front focal plane of C3. Then, wherever the beam enters the C3 lens field, it is bent to follow a path parallel to the optic axis. You can see how this is done if you look ahead to Figure 9.15. This rather complex adjustment is computer-controlled. Like many other procedures on a modern TEM, this adjustment is made automatically when you select STEM mode.

9.1.D. Alignment

If you correctly align the illumination system, the gun crossover is on the optic axis and the electrons can then

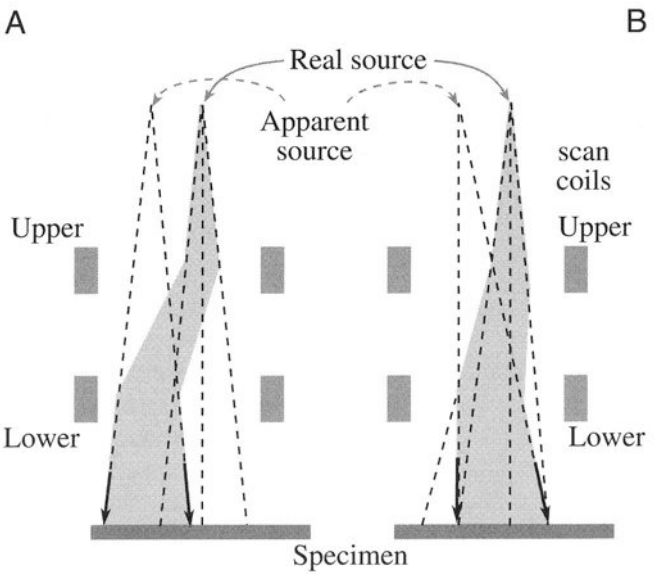


Figure 9.6. The use of pre-specimen scan coils for (A) traversing the beam and (B) tilting the beam. Traversing moves the beam to a different area of the specimen but it stays parallel to the optic axis. Conversely, tilting the beam illuminates the same area of the specimen, but from a different angle.

follow a straight line through the lenses and apertures until they hit the specimen. Alignment used to be a tedious manual affair involving tilting and translating the gun and the condenser lenses and centering the apertures on axis. Now most of the components are machined accurately enough that minor electronic adjustment is all that is needed. Nevertheless, manual centering of the C2 aperture remains a most critical step in obtaining the best performance out of the TEM, particularly if you intend to operate in scanning mode for STEM imaging and microanalysis.

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 it 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.

Gun Alignment

First, you have to undersaturate the filament so structure can be seen in the image, as shown back in Figure 5.5. If the gun is very badly misaligned, you may have to turn the condenser lenses off, before you use the gun traverses to center the filament image. Then use the gun tilts to make the source image symmetrical and repeat the whole procedure. If this alignment is very bad, then either there is a major problem with the gun or with the previous user; in either case expert help is required to correct the fault.

Alignment of the C2 Aperture

You must 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 align the aperture on axis, you need first to overfocus C2 so the image of the beam is spread and the outline of the C2 aperture is visible on the screen (make sure any other apertures in the imaging system are out of the column). Then use the external drives to center the aperture on the screen. Next, you must adjust C2 so the image of the beam is focused. Then, center the beam with the deflector controls. Now underfocus the C2 lens until you can again see the aperture and center it again with the external drives. 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. 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 saves you from manually underfocusing and overfocusing the lens.

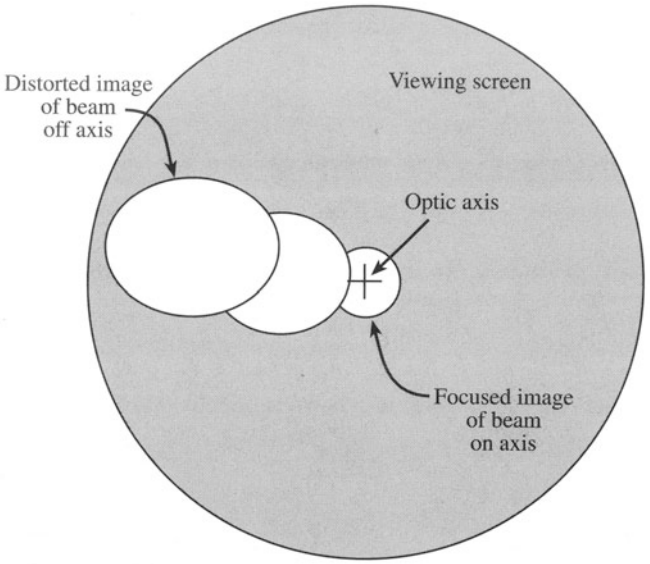


Figure 9.7. If the C2 aperture is misaligned, underfocusing or overfocusing the C2 lens causes the image of the beam to sweep off axis (i.e., across the viewing screen) and to become distorted.

9.1.E. Condenser Lens Defects

The illumination system lenses suffer from the standard lens defects, such as aberrations and astigmatism. These defects don't really limit the operation of the TEM in paral-

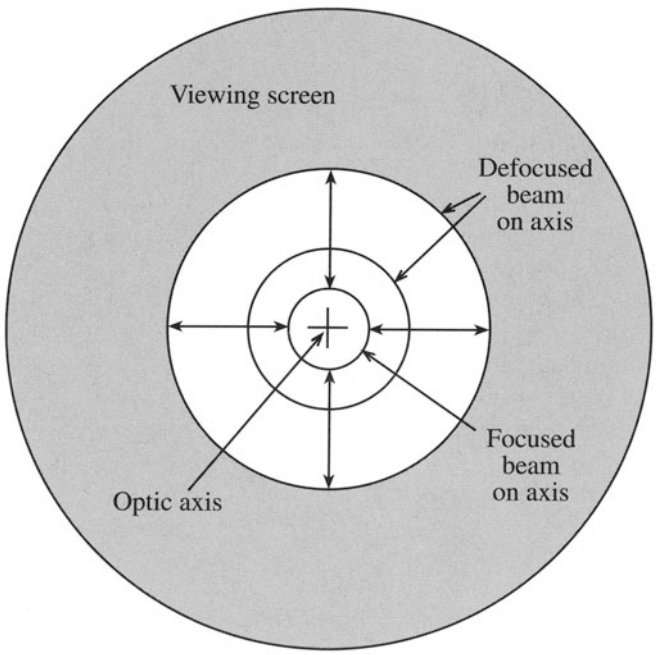


Figure 9.8. If the C2 aperture is aligned, the image of the beam remains circular and expands or contracts about the optic axis as the lens is underfocused or overfocused.

lel-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 6.23) of $r_{min} \sim 0.91(C_s\lambda^3)^{1/4}$. This is why 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 6.22) 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.

Chromatic Aberration

Remember this aberration depends on the energy spread of the electrons. Since the electrons in the illumination system have such a small energy spread, you can regard them as monochromatic and there is no detectable degeneration of the probe dimensions.

Astigmatism

This is the most common defect in the TEM illumination system and arises either because the final C2 limiting aperture is misaligned or it is contaminated and charging up, thus deflecting the beam. Let's assume you've centered the C2 aperture as we just described and talk about correcting any residual astigmatism due to contaminated apertures.

You can discern astigmatism in the illumination system if you look at an image of the electron source on the screen; focus C2 so the beam is a minimum diameter and the image of the beam is circular, as you did when aligning the aperture. If you then wobble the C2 lens either side of the focal setting, 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. The condenser stigmators introduce a compensating field which you use to correct this distortion. You first overfocus the beam so you can see the effect of the astigmatism (i.e., the beam appears elliptical). 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.

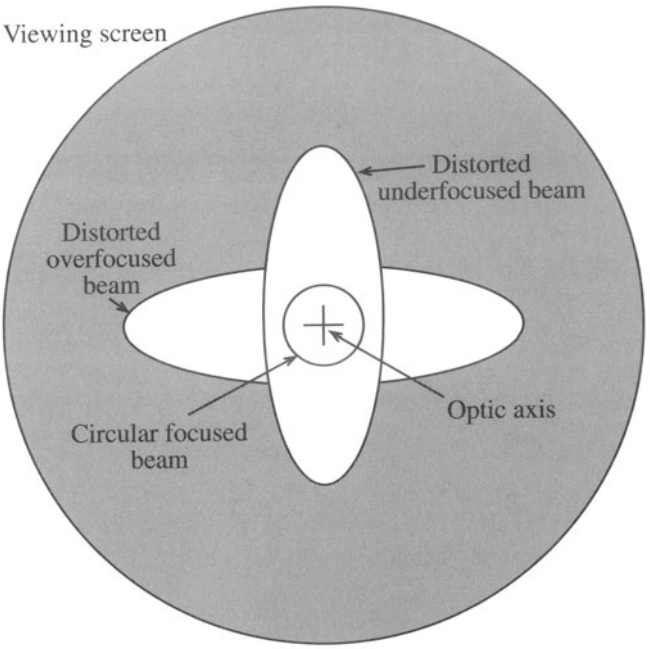


Figure 9.9. The effect of astigmatism in the illumination system is to distort the image of the beam elliptically as the C2 lens is underfocused or overfocused. Correction of this astigmatism results in an image that remains circular as the C2 lens is defocused.

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 you need to remove the source of the astigmatism by flame cleaning the condenser diaphragms, as we described in Chapter 6.

9.1.F. Calibration

We've already seen what it takes to calibrate the performance of the electron gun and optimize the brightness so that the maximum beam current goes into the minimum beam size. All that's left is to calibrate the condenser system. The major variables are the probe size 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. Figure 9.10 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

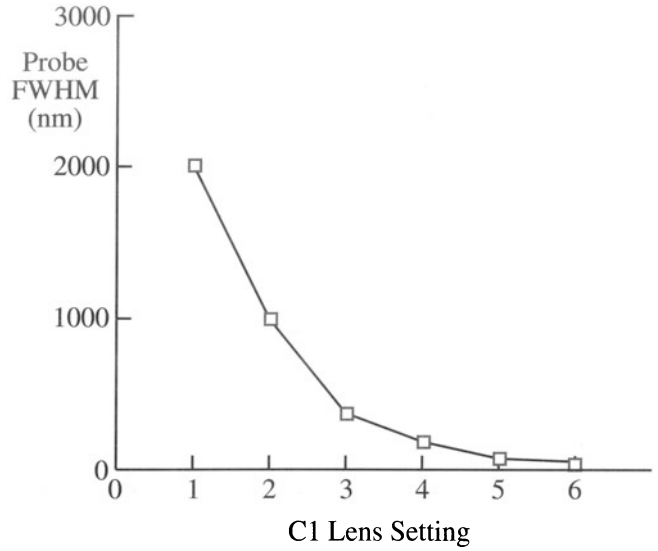


Figure 9.10. Calibration of the illumination system requires determining the variation of the probe size with C1 lens strength.

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 C2 aperture size governs the convergence semiangle 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 2α by increasing the C2 aperture size (Figure 9.11).

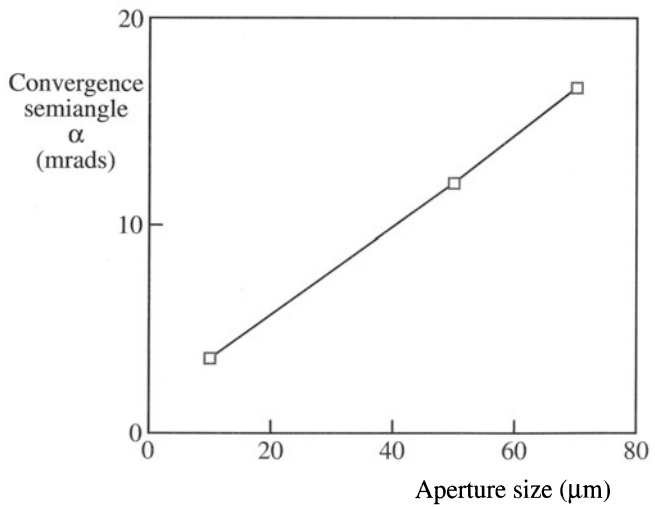


Figure 9.11. Variation of the beam-convergence semiangle, α, with the C2 aperture dimensions.

Different TEMs will have different responses. On some instruments the C2 aperture is virtual (so you have an effective aperture size), which makes it rather difficult to measure 2α. (See Goldstein *et al.* 1992 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 as a function of both 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 diffraction patterns 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 with which you have to make adjustments. Our discussion will emphasize the side-entry holder since this is becoming the standard, but top-entry holders require the same adjustment of the z-control or specimen height.

> We need to fix the height of the specimen on the optic axis. This will allow us to work at the same objective lens current and thus at a fixed objective lens magnification.

As a practical consideration, you would like to be able to tilt the sample without changing its height on the optic axis. Otherwise, you would be continuously using the z-control when you tilt the sample. 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.

The central requirement 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 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 the specimen is located at this plane and the image is in focus, the objective lens current is an optimum 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. Of course, if you tilt your specimen normal to the holder axis, or rotate it off axis, then the point you're examining almost invariably moves out of the eucentric plane.

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, you tilt the specimen and adjust the height of the specimen holder until the image of the specimen remains stationary when you tilt the sample through ±30° either side of zero.

With computer control and auto-focusing techniques becoming common, this operation can be automated. As a result we now see completely eucentric stages in which 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 a computer-controlled stage, be cautious.

The eucentric plane should also be coincident with the plane that is symmetrically positioned with respect to the upper and lower objective polepiece fields. This means that 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 diffraction patterns 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 diffraction pattern or an image in TEM and STEM modes. You should not have to refocus the image or diffraction pattern with the objective lens when you change from one mode to the other.

9.3. FORMING DIFFRACTION PATTERNS AND IMAGES: THE TEM IMAGING SYSTEM

You know that the objective lens takes the electrons emerging from the exit surface of the specimen, disperses them to create a diffraction pattern (DP) in the back focal plane, 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 diffraction patterns 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 will 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. The first operation that you need to master when using the TEM is viewing the diffraction pattern. In all the subsequent imaging, we'll use this pattern to select electrons that have suffered particular angles of scatter to form our images.

- To see the diffraction pattern you have to adjust the imaging system lenses so that the back focal plane of the objective lens acts as the object plane for the intermediate lens. Then the diffraction pattern is projected onto the viewing screen, as shown in Figure 9.12A.
- 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, 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 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 diffraction pattern 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. So we perform a basic TEM operation both to select a specific area of the specimen to contribute to the diffraction pattern and to reduce the intensity of the pattern falling 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 diffraction pattern.

- We could make the beam smaller.
- We could somehow 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 diffraction pattern with a parallel beam of electrons, the standard way is to use a selecting aperture. This operation is called se-

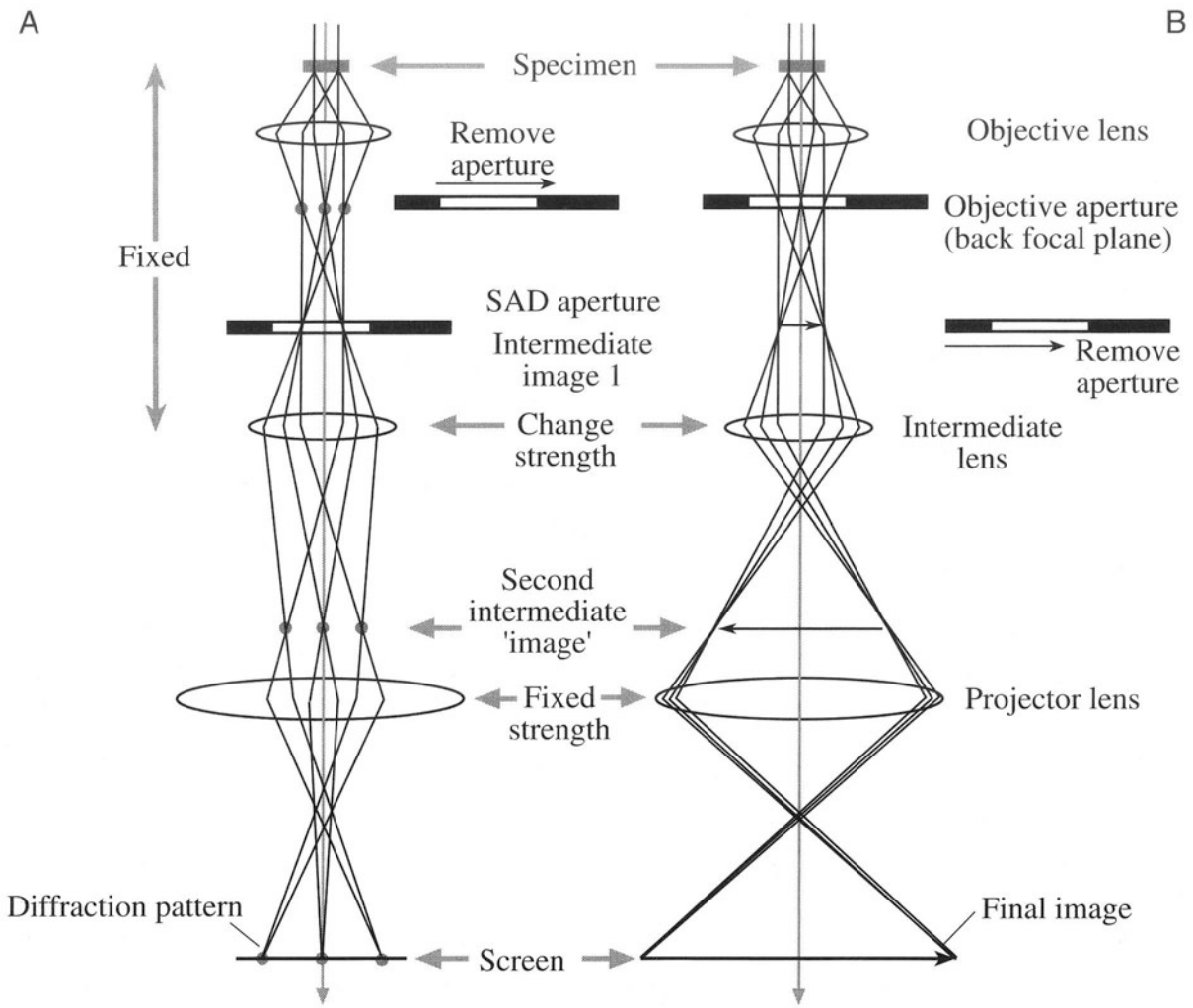


Figure 9.12. The two basic operations of the TEM imaging system involve (A) projecting the diffraction pattern on the viewing screen and (B) projecting the image onto the screen. In each case the intermediate lens selects either the back focal plane or the image plane of the objective lens as its object.

lected-area diffraction, or SAD, and was invented by LePoole (1947). Now, we can't insert an aperture at the specimen plane, because the specimen is already there! If we insert an aperture in a plane conjugate with the specimen, i.e., in one of the image planes, then it creates a virtual aperture at the plane of the specimen. This is exactly what we do.

The conjugate plane that we choose is the image plane of the objective lens, as shown in Figure 9.13. We insert the *SAD aperture* into the image plane of the objective lens and center the aperture on the optic axis in the middle of the viewing screen. You can see the image of this aperture on the viewing screen. It must be focused 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. Then any electron that hits the specimen outside the area defined by the virtual aperture will hit the real diaphragm when it travels on to the image plane. It will thus be excluded from contributing to the diffraction pattern that is projected onto the viewing screen. In practice, we can't make apertures smaller than about 10 μm, and the demagnification back to the plane of the specimen is only about 25×, which gives a minimum selected area of ~0.4 μm—which isn't as small as we'd like. We'll discuss in Chapter 11 whether or not smaller values would be useful.

The SAD pattern is often displayed on the viewing screen at a fixed magnification.

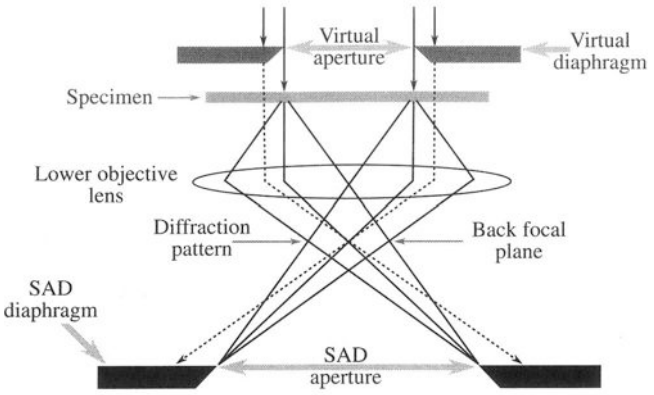


Figure 9.13. Ray diagram showing SAD pattern formation: the insertion of an aperture in the image plane results in the creation of a virtual aperture in the plane of the specimen. Only electrons falling inside the dimensions of the virtual aperture at the specimen will be allowed through into the imaging system. All other electrons will hit the SAD diaphragm.

By analogy with the hand-held camera we define a distance called the "camera length" (*L*).

This distance corresponds to the distance of the "film" from the diffraction pattern. We choose the value of *L* such that the spacings in the diffraction pattern are easily discernible on the screen and on the photographic plate. This magnification can be changed by adjusting the intermediate lenses. We'll describe how we calibrate this magnification later.

It is a basic principle of TEM operation that when you want to look at the diffraction pattern (i.e., the *back focal plane* of the objective lens), you put an *SAD aperture* into the *image plane* of the objective lens.

You can see this aperture if you want to change it or center it, by projecting the image plane onto the viewing screen, which we'll now discuss.

Beware: In most TEM books, SAD is the only standard diffraction technique. As a result, some microscopists use only SAD to obtain diffraction information. However, you should know that CBED, which we discuss in Chapters 20 and 21, can provide complementary diffraction information and must also be used by all TEM operators in the materials field. There are still certain times when you'll need to form an SAD pattern.

- 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 diffraction pattern, such as streaks (see Chapter 17).

On all other occasions, when the diffraction maxima provide the most important information in the pattern, then we strongly recommend that you use CBED.

9.3.B. Bright-Field and Dark-Field Imaging

When the SAD pattern is projected onto the viewing screen, we can use the pattern to perform the two most basic imaging operations in the TEM. No matter what kind of specimen you're are looking at, the SAD pattern will contain a bright central spot which contains the direct electrons and some scattered electrons (as shown in Figures 2.11a–c). 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 to insert an aperture into the back focal plane of the objective lens, thus blocking out most of the diffraction pattern except that which is visible through the aperture. We use the external drives to move the aperture so that either the direct electrons or some scattered electrons go through it. If the direct beam is selected as shown in Figure 9.14A, we call the resultant image a bright-field (BF) image, and if we select scattered electrons of any form, we call it a dark-field (DF) image, as shown in Figure 9.14B. Typical magnification ranges will be 25,000× – 100,000×.

The BF and DF images can be viewed at any magnification simply by adjusting the intermediate lenses of the microscope. It is necessary to calibrate the actual magnification and also to be able to relate directions in the image at any magnification to directions in the diffraction pattern at a fixed camera length. These are the two basic calibrations required for any TEM.

It is another 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 into the *back focal plane* of the objective lens. This is called the *objective aperture* and is most important in the TEM, since its size controls the collection angle (β) and hence determines the effect of all the aberrations and resolution of the most important lens in the instrument.

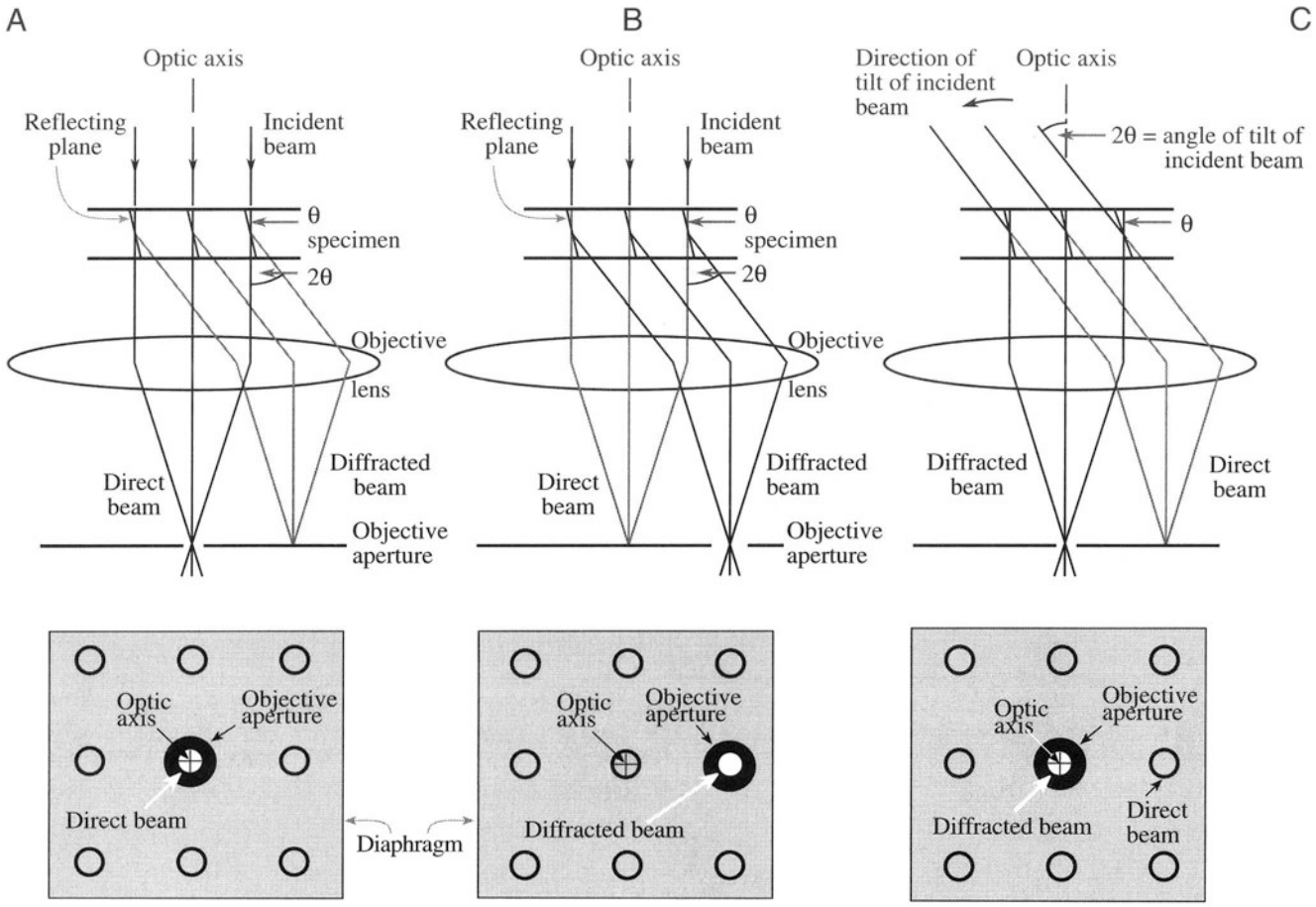


Figure 9.14. Ray diagrams showing how the objective lens/aperture are used in combination to produce (A) a BF image formed from the direct beam, (B) a displaced-aperture DF image formed with a specific off-axis scattered beam, and (C) a CDF image where the incident beam is tilted so that the scattered beam remains on axis. The area selected by the objective aperture, as seen on the viewing screen, is shown below each ray diagram.

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 SAD patterns 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 back focal plane, 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 diffraction pattern, 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.

9.3.C. Centered DF 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. These off-axis electrons suffer aberrations and astigmatism and the DF image is difficult to focus, since it will move on the screen as you adjust the objective lens strength. To avoid this you have to adjust the beam tilt potentiometers above the objective lens so that the incident beam hits 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. This operation is called centered dark-field (CDF) imaging and is the way to do DF imaging in the TEM, if you want to record the best, focused image. However, there are situations where you will want to form a displaced-aperture DF image. You do this by physically moving the aperture rather than by tilting the incident beam.

We'll return to BF, CDF, and SAD operations when we discuss specific contrast mechanisms that occur in TEM images in Chapter 22.

9.4. FORMING DIFFRACTION PATTERNS 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.

> The beam has to scan parallel to the optic axis at all times so that it mimics the parallel beam in a TEM even though it's scanning.

As we show in Figure 9.15, the way we do this is to use two pairs of scan coils to pivot the beam about the front focal plane 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, then a stationary diffraction pattern is formed in the back focal plane (so called because this pattern does not move even though the beam is scanning, since it is conjugate with the front focal plane, as shown in Figure 9.16). If we stop the beam from scanning, then we have a CBED pattern in the back focal plane and we can project that onto the TEM screen if we wish. Now let's discuss how to form STEM images.

One potentially very big advantage of forming images this way is that we don't use lenses, as in an SEM. So defects in the *imaging lenses* do not affect your image resolution, which is controlled by the beam only. Hence chromatic aberration, which can limit TEM images, is absent in STEM images, which is advantageous if you're dealing with a thick specimen. However, there are drawbacks also, as we'll discuss below, and STEM images aren't widely used, particularly for crystalline specimens.

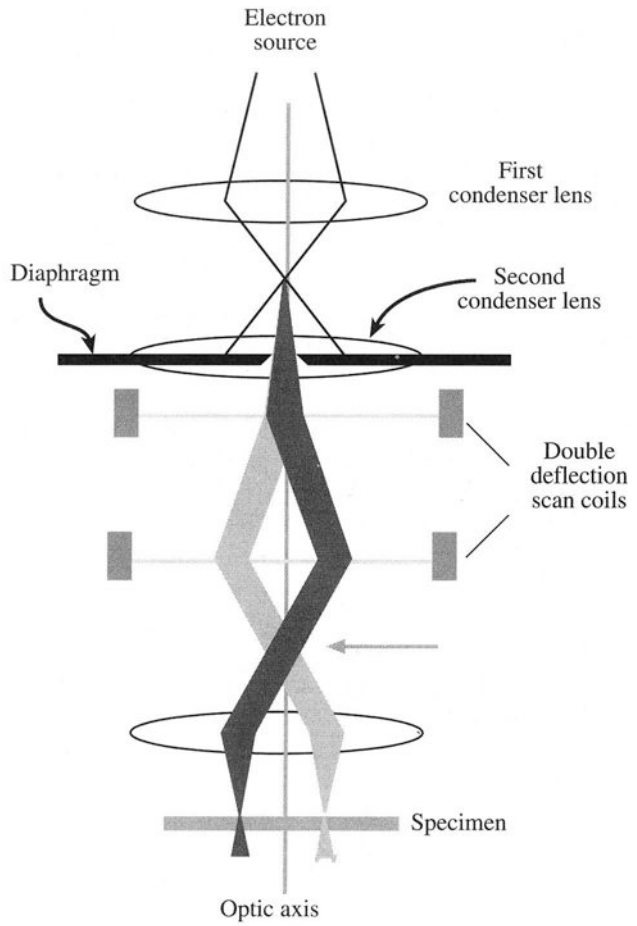


Figure 9.15. Scanning the convergent probe for STEM image formation using two pairs of scan coils between the C2 lens (usually switched off) and the upper objective polepiece. The probe remains parallel to the optic axis as it scans.

9.4.A. Bright-Field STEM Images

The basic principle of image formation in the scanning mode is fundamentally different from that for a static-beam TEM image. As you've just seen, in the TEM we se-

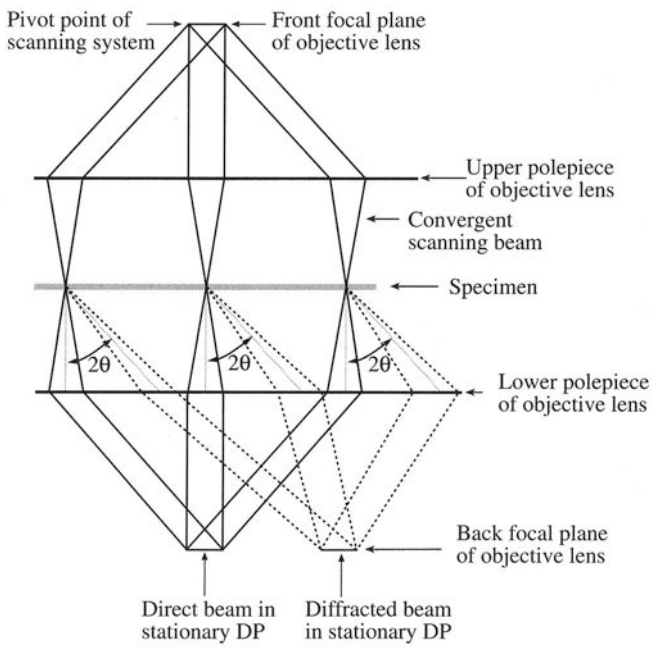


Figure 9.16. The creation of a stationary (convergent-beam) diffraction pattern in the back focal plane of the objective lens is a necessary prerequisite for STEM imaging. Note that electrons scattered through 2θ at different points in the specimen are focused at the same point in the focal plane.

lect 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 9.17. Simply stated, the beam is scanned on the specimen by adjusting the scan coils; these same coils are used to scan the CRT synchronously. The electron detector acts as the interface between the electrons coming from the specimen and the image viewed on the CRT. Since it takes up to 2048 scan lines to build up an image on the CRT, the whole process of creating a STEM image is much slower than TEM imaging: it's serial recording instead of parallel recording.

> The STEM signal generated at any point on the specimen is detected, amplified, and a proportional signal is displayed at an equivalent point on the CRT. The image builds up over several seconds or even minutes.

This process is exactly the same principle as used in any scanned-probe microscope, 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 diffraction pattern 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 put a BF detector (either a semiconductor or scintillator detector) into the direct beam in the diffraction pattern in Figure 9.18. Thus, only direct electrons hit that detector from wherever the beam is scanning on the specimen. This variable signal travels from the detector via an amplification system to modulate the signal on a CRT, thus creating a BF image as also shown in Figure 9.18.

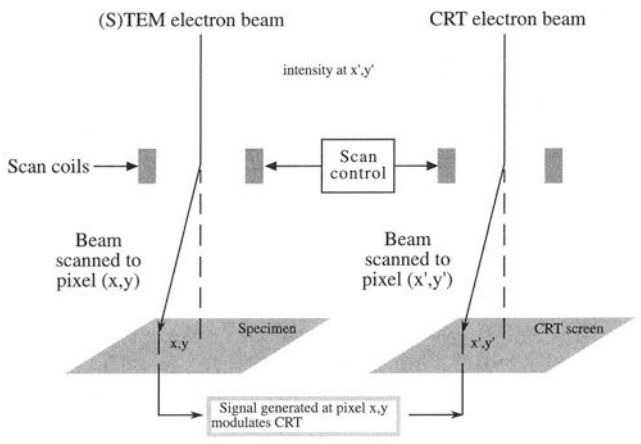


Figure 9.17. The principle of forming a scanning image, showing how the same scan coils in the microscope control the beam-scan on the specimen and the beam-scan on the CRT. Thus no lenses are required to form the image.

> The variable signal which emerges from the BF detector depends on the intensity in the direct beam from that point on the specimen.

Now, in a TEM we can't physically put the detector in the back focal plane 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 diffraction pattern, below the projector lens. 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 the screen. The stationary diffraction pattern falls on the detector and the signal goes to the CRT. In a DSTEM, there may not be any imaging system lenses, in which case the detector is 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 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 of the scattered electrons, rather than the direct electrons. Remember, in a TEM we tilt the incident beam so the scattered electrons that 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.

> If we want a specific beam of scattered electrons to fall on the BF detector, we can simply shift the stationary diffraction pattern so that the scattered beam is on the optic axis.

It's simple to do this with the diffraction pattern 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 DF 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 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.

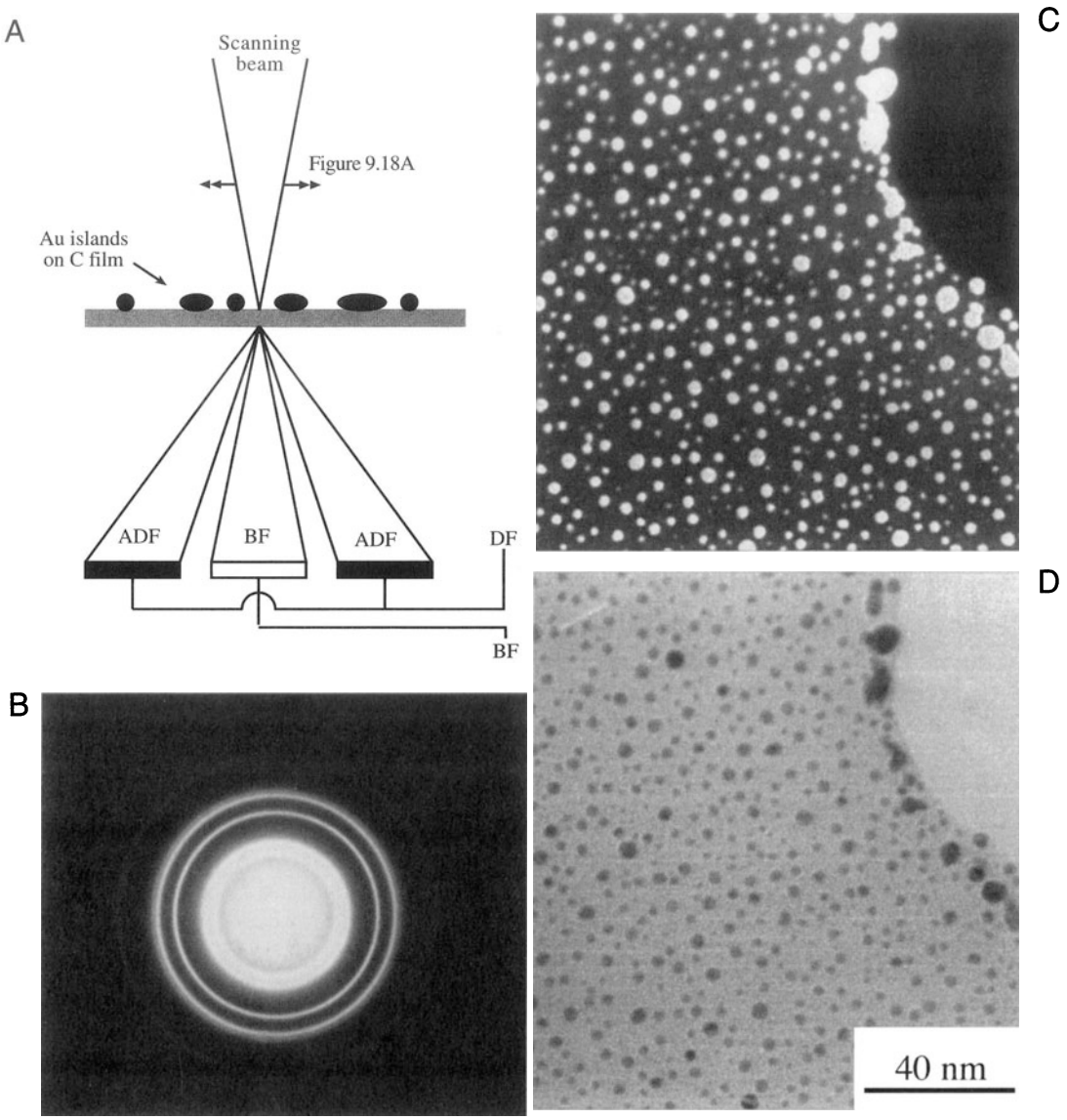


Figure 9.18. STEM image formation: A BF detector is placed in a conjugate plane to the back focal plane to intercept the direct beam (A) and a concentric annular DF detector intercepts the diffracted electrons (B). The signals from either detector are amplified and modulate the STEM CRT. The specimen (Au islands on a C film) gives complementary ADF (C) and BF (D) images.

As is shown in Figure 9.18, 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 is complementary to the BF image.

Now, of course, you can make the detector any size or shape you wish. For example, you can split the annulus into two halves or four quadrants and electrically isolate each part of the detector. Then you can form 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. 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

Any of the STEM images that we have just described appear on the CRT 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. Scanning images are *not* magnified by lenses (and are thus not affected by aberrations in the imaging lenses as we stated at the start of this

section). If the scanned area on the specimen is 1 cm × 1 cm, and the resultant image is displayed on a CRT with an area 10 cm × 10 cm (which is the standard size of the *record* CRT, though rarely the size of the *viewing* CRT screen) then the magnification is 10×. If the scan dimension is reduced to 1 mm, the magnification is 100×, and so on, up to magnifications in excess of 10^6×, which are common in dedicated STEMs. As with the TEM, we have to calibrate the STEM magnification and the camera length of the diffraction pattern 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 diffraction pattern on the optic axis. To get the best out of your TEM, you must master these two fundamental alignments.

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 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 and your image will rotate about a position off axis as you change the objective lens (focus), as shown schematically in Figure 9.19.

To center the objective rotation, you start at a relatively low magnification (say 10,000×), select an obvious reference point in the image, and observe the way the point rotates as you wobble the objective lens over and under focus. Then use the beam tilts to move the center of rotation to the middle of the screen and repeat the process at higher magnifications. Above ~100,000× 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."

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.

The diffraction center is aligned by adjusting the projector lens until the central spot in the diffraction pattern is on axis. If you change the diffraction pattern magnification (the camera length) the pattern will move off axis, which can easily be compensated for in a computer-controlled column.

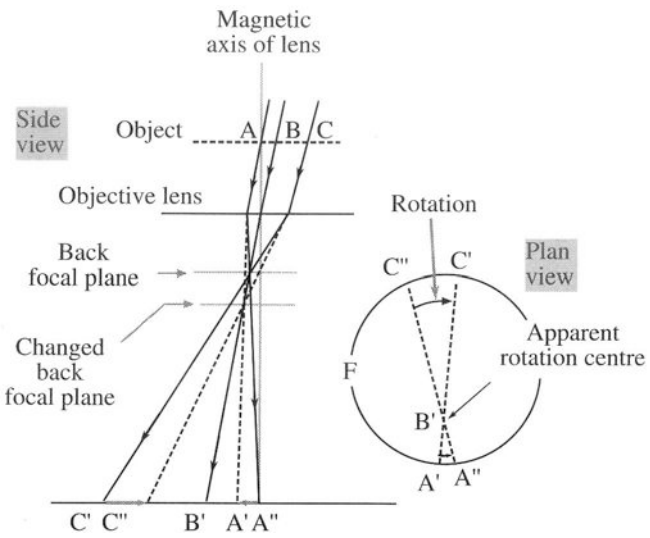


Figure 9.19. When the objective lens center of rotation is misaligned, the image appears to rotate about a point away from the center of the viewing screen when the lens is wobbled about focus.

Centering the diffraction pattern is useful in STEM image formation, since you use it to center the diffraction pattern on the STEM detector 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 diffraction pattern, the main cause of problems in the imaging system is objective and intermediate lens astigmatism.

Objective lens astigmatism occurs mainly if the objective aperture is misaligned, so you must carefully center the aperture on the optic axis, symmetrically around the electron beam used to form the BF or DF image. Despite careful centering, however, residual contamination may 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. Often you can only see objective astigmatism at the highest magnifications, where it manifests itself as a streaking in the image that rotates through 90° as you alternately underfocus and overfocus the objective lens. So again, you have to wobble the objective lens, but if the magnification is too high then manual wobbling is required.

If you then overfocus C2 to ensure a parallel beam, and also overfocus the objective lens, then a Fresnel fringe will be visible at the thin edge of the specimen.

Either look for a small hole in the specimen or look at the edge of the specimen. Ideally, you might use a holey carbon film to correct residual astigmatism *before* you put in your specimen, especially while learning this procedure. In practice, you have to check your astigmatism throughout your TEM session so you use the same approach on a thin curved edge of your specimen.

As shown in Figure 9.20A, when you underfocus the objective lens, there is a bright 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 9.20D, then the focus of the lens is changing around the hole because of astigmatism. So you then 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 9.20B). At exact focus, there is no fringe and the image contrast is minimized (Figure 9.20C).

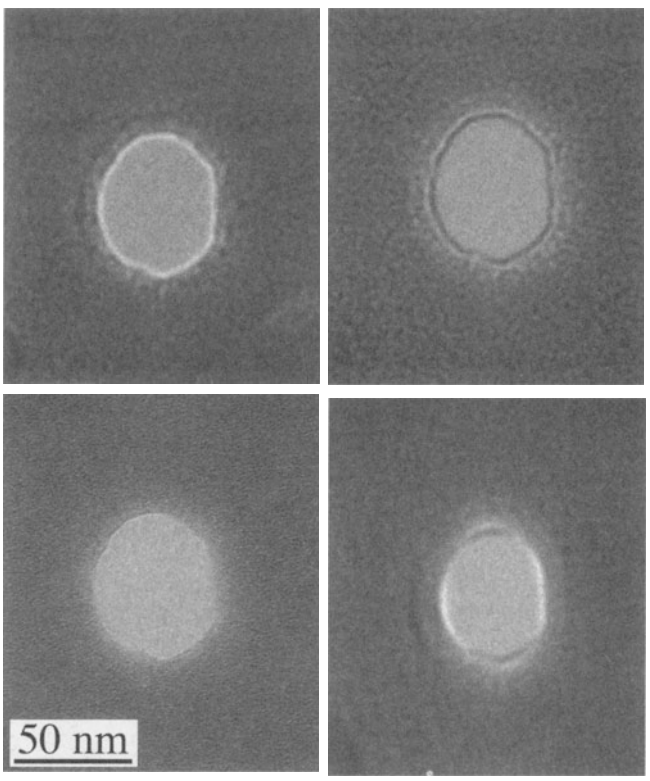


Figure 9.20. The image of a hole in an amorphous carbon film illuminated with a parallel beam showing that (A) with the beam underfocused, a bright Fresnel fringe is visible; (B) with the beam overfocused a dark fringe is visible; (C) at exact focus there is no fringe; and (D) residual astigmatism distorts the fringe.

This method of correcting the astigmatism is reasonable at magnifications up to several hundred thousand times. For high-resolution imaging at magnifications of >300,000×, we actually 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. CALIBRATION OF THE IMAGING SYSTEM

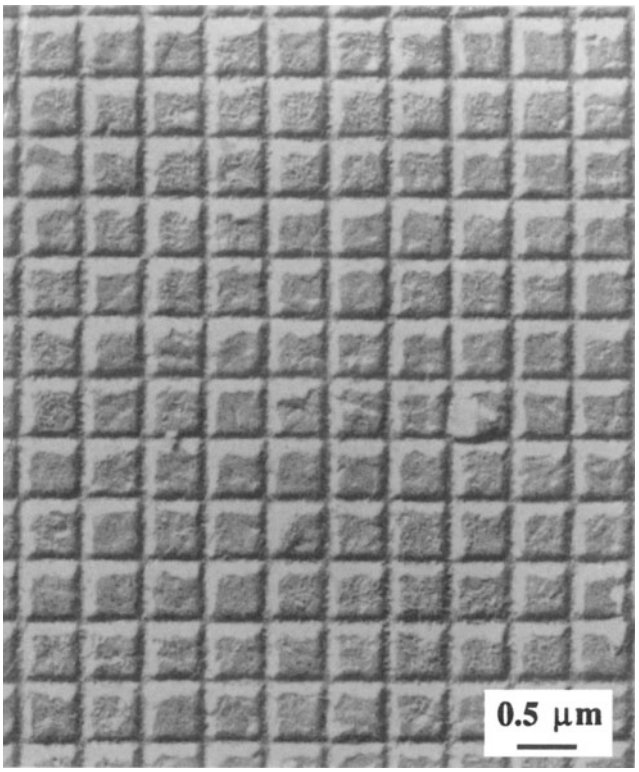
Your TEM should be calibrated when it is first installed and then periodically throughout its life, especially if you wish to carry out accurate measurements from images or diffraction patterns. If the instrument is modified substantially, then it must be recalibrated. In all cases you must specify a set of standard conditions under which the calibrations are carried out (e.g., objective lens current and other lens settings, eucentric height, etc.).

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 specimens we use are thin carbon film replicas of optical diffraction gratings of known spacing, such as shown in Figure 9.21A. The typical linear density

A



B

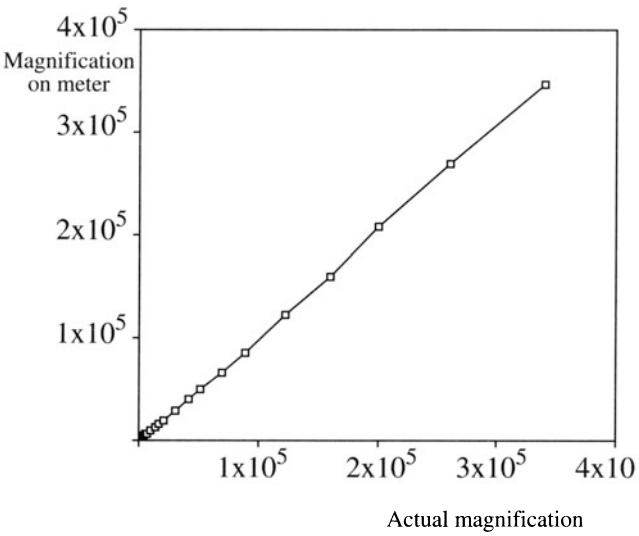


Figure 9.21. (A) An image of a diffraction grating replica in which the actual spacing of the grating is known. (B) The TEM magnification can thus be calibrated, relating specific magnification settings to be assigned specific magnifications.

of lines in the replica is 2160 lines per mm (giving a line spacing of 0.463 µm), which enables calibration up to magnifications of about 200,000×. Above this magnification, individual grating spacings are wider than the film. So we 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, the images of known crystal spacings, such as the 0002 spacing in the graphite structure (0.344 nm), can be used. 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 later in the text.

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 on the same negative.

Basically, you set the TEM to its standard conditions with the specimen at the eucentric height and focus the image. Next, record images of the diffraction grating at all magnification settings and calculate the magnification experimentally from the image. Figure 9.21B shows the magnification calibration for a Philips CM30 TEM, using both a diffraction grating and latex spheres.

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, with the efficiency of the cooling system of the lenses, and with lens hysteresis. Therefore, if you want to make accurate measurements from 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 distortions in the image, particularly at low magnification (<5000×). You can find a full description of all the details concerned with TEM calibration in Edington (1976).

Because of the magnification error in the TEM, it is not the best instrument for absolute measurement of parti-

Table 9.1. Magnification Calibration for a Philips EM400T in STEM Mode at 120kV

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

cle 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 this (e.g., 52,700×). It may indicate that the microscopist does not understand the instrument's limitations and the work should be interpreted with due caution.

Remember that the electromagnetic lenses have hysteresis and the area of the specimen you are working on must be at exactly the right "height" in the column.

You can use an identical procedure to calibrate the STEM CRT image magnification. This is equally important 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 readout of the STEM magnification and the experimentally determined magnification using a diffraction grating replica.

9.6.B. Camera-Length Calibration

We describe the magnification of the diffraction pattern 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 these cameras magnification is increased by moving the recording film further away from the specimen. This principle can be applied in the TEM, as shown in Figure 9.22. This figure represents the imaging system, but without the lenses drawn in.

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 in the TEM is thus 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), then the separation of the direct and diffracted beams as measured on the screen (R) is determined by L since

$$\frac{R}{L} = \tan 2\theta \sim 2\theta \qquad [9.1]$$

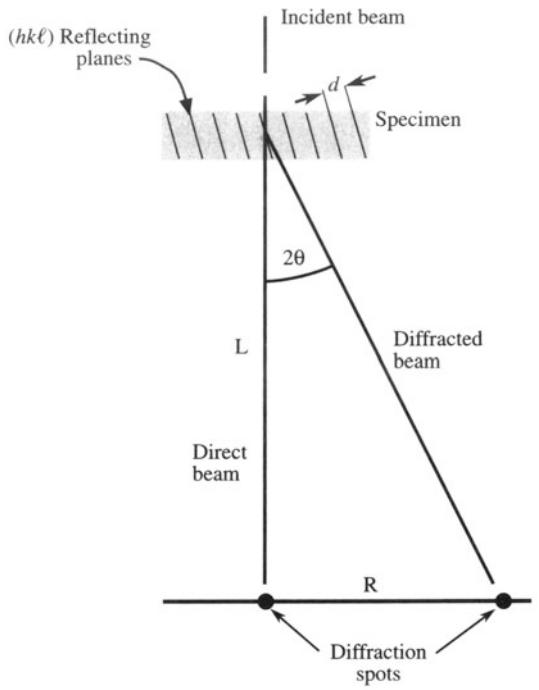


Figure 9.22. The relationship between the spacing R of diffraction maxima and the camera length, L. Increased magnification corresponds to effectively increasing L, although in practice this is accomplished with lenses.

From the Bragg equation we know that $\lambda/d = 2 \sin \theta \sim 2\theta$, and so we can write

$$Rd = \lambda L \qquad [9.2]$$

Thus to calibrate the magnification of the diffraction pattern we need to record patterns from a specimen with known crystal spacing (d), such as a thin film of a polycrystalline Au or Al. This gives a ring pattern (see Figure 2.11). We know the lattice parameter of the specimen, we can measure the ring radius R on the photographic film for any plane that is diffracting (see Chapter 18 to find out exactly how we do this), and since we know λ we can determine L. A typical TEM camera length calibration is shown in Table 9.2. The STEM camera length calibration may be different than 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.

As a general rule, you should *always* do the calibrations yourself, and not rely on any factory calibrations, because the conditions you use in your laboratory may differ from those of the manufacturer.

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 = 0.0335Å)

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

9.6.C. Rotation of the Image Relative to the Diffraction Pattern

Anyone studying crystalline materials must determine the angle between directions in the image and directions in the diffraction pattern. At a fixed camera length, the diffraction pattern 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 diffraction pattern. (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 common directions.)

To determine this rotation, we 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. You must take care to ensure that, as usual, the image is focused with the specimen at the eucentric plane. Then insert the SAD aperture and ensure that it is focused using the intermediate lenses to coincide with the image plane. Finally, switch to diffraction mode with the beam underfocused and adjust the diffraction focus to give sharp diffraction maxima. Then, take a double exposure of the diffraction pattern and the image as shown in Figure 9.23A. Repeat the whole exercise for different magnifications and plot out the variation of the angle Φ as shown in Figure 9.23B. You can do the same if necessary for different values of L, which introduce a systematic change in Φ. It is recommended that you carry out all your SAD work at a standard value of L; 500–1000 mm is usually optimum.

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 diffraction pattern 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 9.24.

9.6.D. Analysis of TEM Images and Diffraction Patterns

If you don't use a double exposure when comparing images and diffraction patterns (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.

> Whenever you're comparing images and diffraction patterns, 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 diffraction patterns.

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

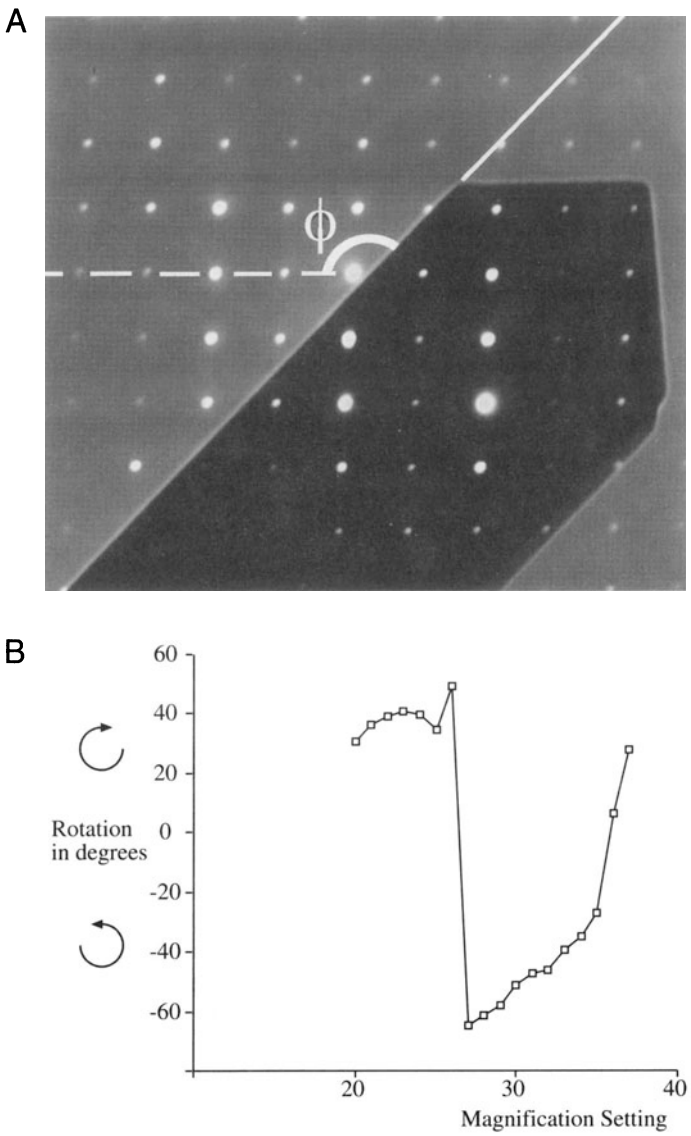


Figure 9.23. (A) A double exposure showing the superposition of an image of a MoO_3 crystal on a diffraction pattern from the same crystal, defining the rotation angle Φ. (B) The rotation calibration gives the angle Φ between equivalent directions in the image and the diffraction pattern as the magnification is varied. The calibration assumes a constant camera length.

measure the electron wavelength λ by measuring the angle ϕ between two Kikuchi line pairs (see Chapter 19) which intersect at a distance R from the direct beam

$$\tan\phi = \frac{R}{L} = \frac{\lambda}{d} \qquad [9.3]$$

Alternatively, you can match simulations of higher-order Laue-zone lines to experimental lines in CBED patterns (see Chapter 21) and determine which λ gives the best match. Finally, if you can get your X-ray computer system

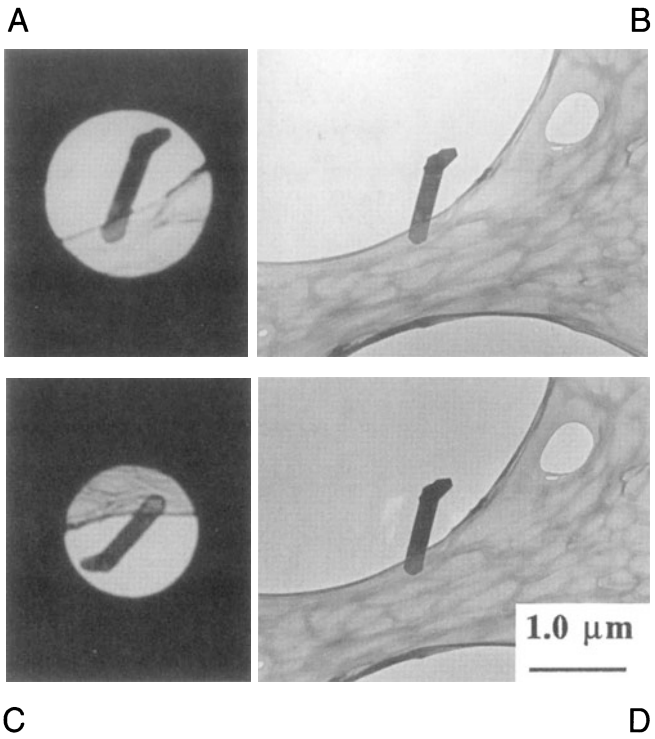


Figure 9.24. Defocused direct beam in a diffraction pattern from MoO_3 compared with a BF image, showing how to determine if a 180° inversion exists or not. If the image in the spot is rotated with respect to the image on the screen, as in (C) and (D), then the 180° inversion is required. In (A) and (B), no rotation occurs between the DP and BF image.

to display the X-ray spectrum (Chapter 34) out to E_0, the beam energy, the bremsstrahlung intensity vanishes to zero at the exact beam energy (this is called the Duane-Hunt limit).

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 often rotated relative to the specimen, how can you locate the rotation axis? Move the specimen. From this movement, you can determine the sense of tilt for a specimen of known geometry.

> If you gently push on the end of your side-entry specimen holder, the image moves parallel to the principal tilt axis.

If you are looking at the diffraction pattern, defocus the pattern so you can see the BF image in the zero spot, as in Figure 9.25, 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.

Focal increments of the objective lens: If you're going to do high-resolution phase-contrast imaging, then

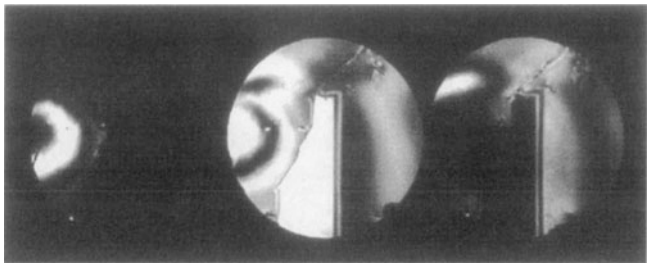


Figure 9.25. Defocused multiple DF image showing how it is possible to determine simultaneously the direction of features in the image (e.g., the vertical twin boundary) and directions in the DP (e.g., the horizontal vector between the diffraction disks). If the specimen holder is moved in the direction of the principal tilt axis, the image will move and identify the relationship between that tilt axis and the DP.

you need to know the value of each defocus step of the objective lens. 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

$$\Delta f = \frac{\Delta x}{2Mm\theta} \qquad [9.4]$$

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 use 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.

CHAPTER SUMMARY

We've now shown you how a TEM is put together. While the manufacturer does a pretty good job, 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 is 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 the astigmatism at high magnification. Diffraction and STEM operation require a centered diffraction pattern.

If you want to make any quantitative measurements from your images and diffraction patterns (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 diffraction patterns are relatively useless unless you know their magnification (camera length) 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.

REFERENCES

General References

Chapman, S.K. (1980) *Understanding and Optimizing Electron Microscope Performance 1. Transmission Microscopy,* Science Reviews Ltd., London.

Chapman, S.K. (1986) *Maintaining and Monitoring the TEM,* Royal Microscopical Society Handbook, No. 8, Oxford University Press, New York.

Chescoe, D. and Goodhew, P.J. (1990) *The Operation of Transmission and Scanning Microscopes,* Royal Microscopical Society Handbook, No. 20, Oxford University Press, New York.

Specific References

Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science,* Van Nostrand Reinhold, New York.

Goldstein, J.I., Newbury, D.E., Echlin, P., Joy, D.C., Romig, A.D. Jr., Lyman, C.E., Fiori, C.E., and Lifshin, E. (1992) *Scanning Electron Microscopy and X-ray Microanalysis,* Plenum Press, New York.

LePoole, J.B. (1947) *Philips Tech. Rundsch.* **9**, 93.

Specimen Preparation

10

CHAPTER PREVIEW

There are many ways to prepare specimens for the TEM. The method you choose will depend on both the type of material and the information you need to obtain. 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.

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."

The TEM specimen, when you've made it, must be electron transparent and representative of the material you want to study. In most cases (not all) you would like your specimen to be uniformly thin, stable under the electron beam and in the laboratory environment, conducting, and nonmagnetic (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 is usually Cu but could be Au, Ni, Be, 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.

It is often assumed that preparation of the TEM specimen will take several hours. Actually, this time could be as short as five minutes or as long as two 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 mortar and pestle using a nonaqueous solvent, catch the small particles on a carbon film, and put the specimen in the TEM; time required, about ten 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, one or two days. Which method you choose would depend on what you want to learn about your material.

Specimen Preparation

10

10.1. SAFETY

Either the specimen itself or the best method for preparing it for viewing in the TEM may require extreme care. Even specimens 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 penetrates the body to dissolve 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 material safety data sheets (MSDS) before you begin specimen preparation. You might also save a lot of time.

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 which 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 which have large thin areas. If you use HF, completely cover any ex-

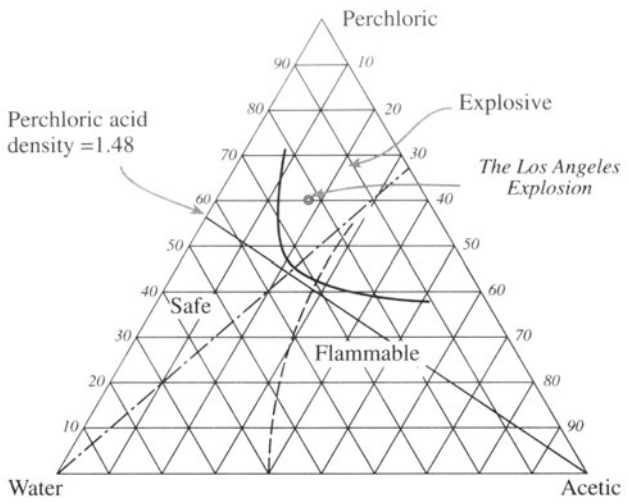


Figure 10.1. Perchloric–acetic–water phase diagram showing the hazardous regions and the recommended density line for safe use of all perchloric solutions. Always operate to the left of this line. (After Medard *et al.* 1949)

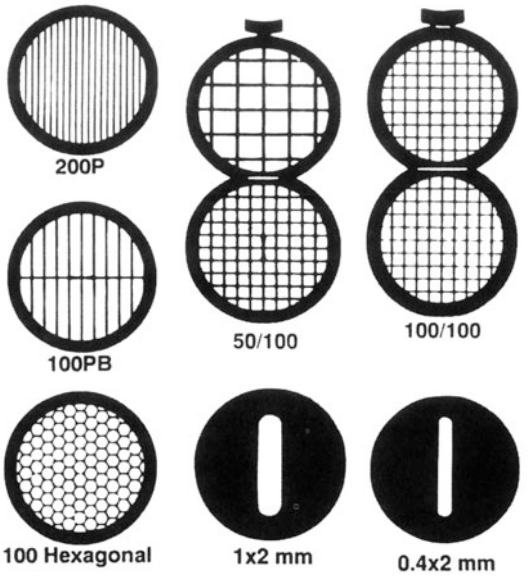


Figure 10.2. A variety of specimen support grids of different mesh size and shape. At the top right is the oyster grid, useful for sandwiching small slivers of thin material.

posed skin; HF rapidly penetrates the flesh and dissolves bone.

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 self-supporting specimen is one where the whole specimen consists of one material. Other specimens are supported on a grid or on a Cu washer with a single slot. Several grids are shown in Figure 10.2. 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. 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 then you "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 where no Cu is present in the specimen. We'll talk later 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.

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 specimen which is 5 mm x 5 mm will give one 3-mm disk or four 2.3-mm disks. The second advantage of such specimens relates to tilting; the specimen holder can be manufactured to allow a greater tilt angle. Don't forget that if you only need one

axis of tilt you may find the bulk holder useful. Then you can use a specimen which may be up to 1 cm long and 3 mm wide.

10.3. PREPARING A SELF-SUPPORTING DISK FOR FINAL THINNING

Final thinning involves three parts:

- Initial thinning to make a slice of material between 100 μm 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. Creation of 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 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.

(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 rocksalt. 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 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 (Figure 10.3). 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, ultrasonic drilling, and 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 which 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

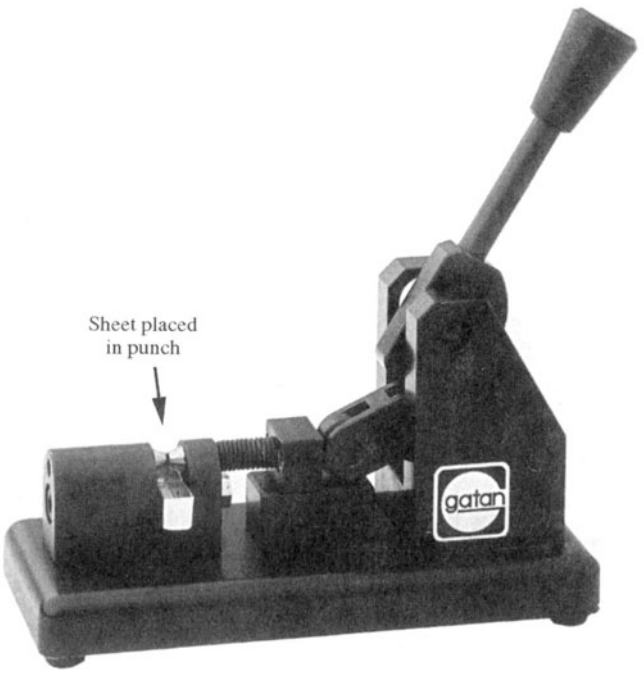


Figure 10.3. A mechanical punch for stamping disks from thin sheets of ductile materials. A sheet sample is placed in the punch as indicated and the handle on the right is pushed down, ejecting a 3-mm-diameter disk suitable for thinning

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 2× the damage depth or else mechanical damage will be visible in the final specimen.

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 introduces no 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.

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 10.4) 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.

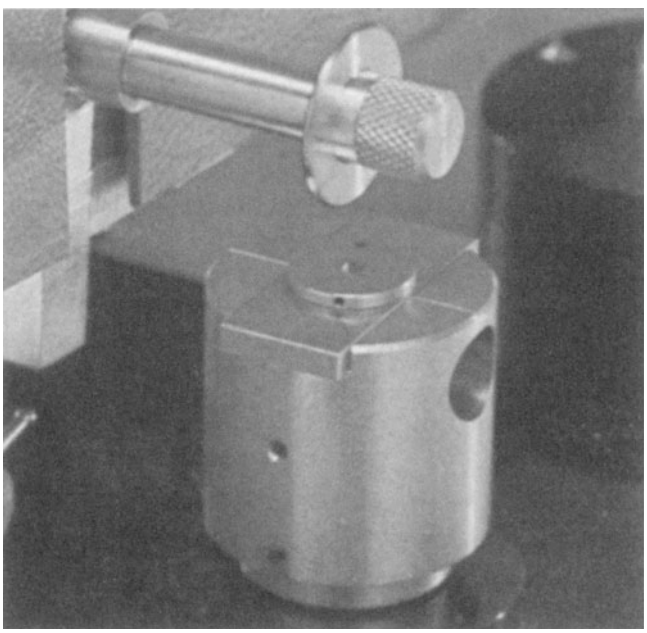


Figure 10.4. A dimpling apparatus showing grinding tool, and specimen support block.

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 10.5) 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.

Anderson has revolutionized TEM specimen preparation through the development of the tripod polisher (Klepeis *et al.* 1988); this tool can help you 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. A polisher can be purchased commercially or you can build your own.

> 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.

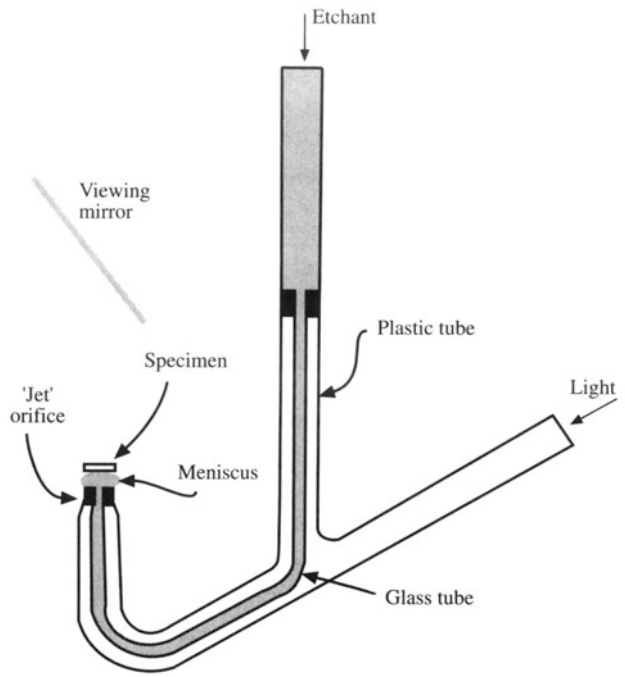


Figure 10.5. Surface dimpling using a chemical solution, e.g., to remove Si from one side of a disk. The light pipe permits visual detection of perforation using the mirror.

- You need a supply of fine diamond lapping films; these are not inexpensive but it is 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 attach them to the glass platen by water tension 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, you 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 DISK

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 10.6. The classical jet polish is shown in Figure 10.7A. By keeping the volume of the

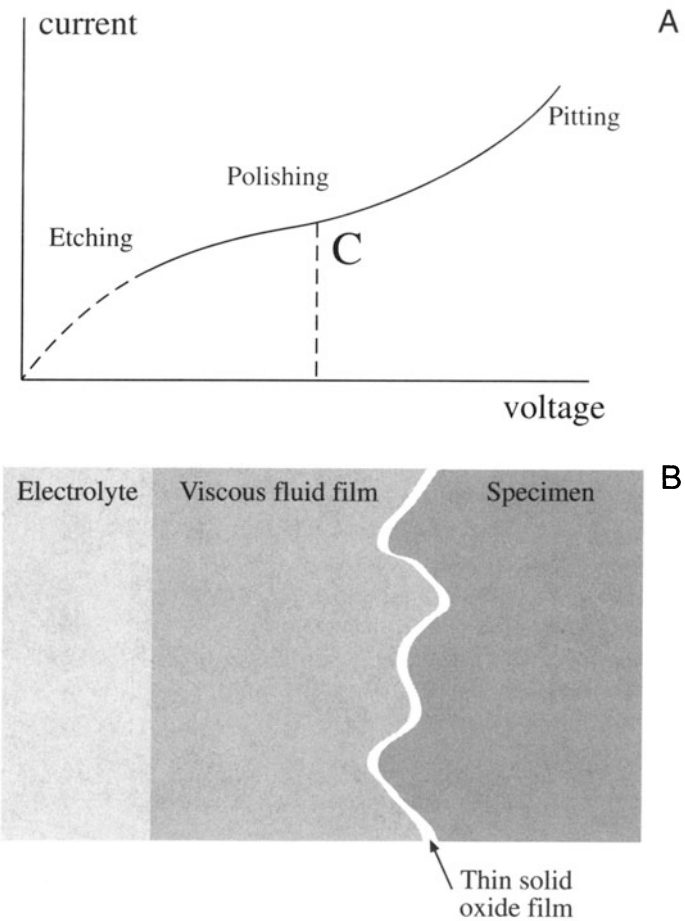


Figure 10.6. (A) Electropolishing curve showing the increase in current between the anode and the cathode as the applied voltage is increased. Polishing occurs on the plateau, etching at low voltages, and pitting at high voltages. (B) The ideal conditions for obtaining a polished surface require the formation of a viscous film between the electrolyte and the specimen surface.

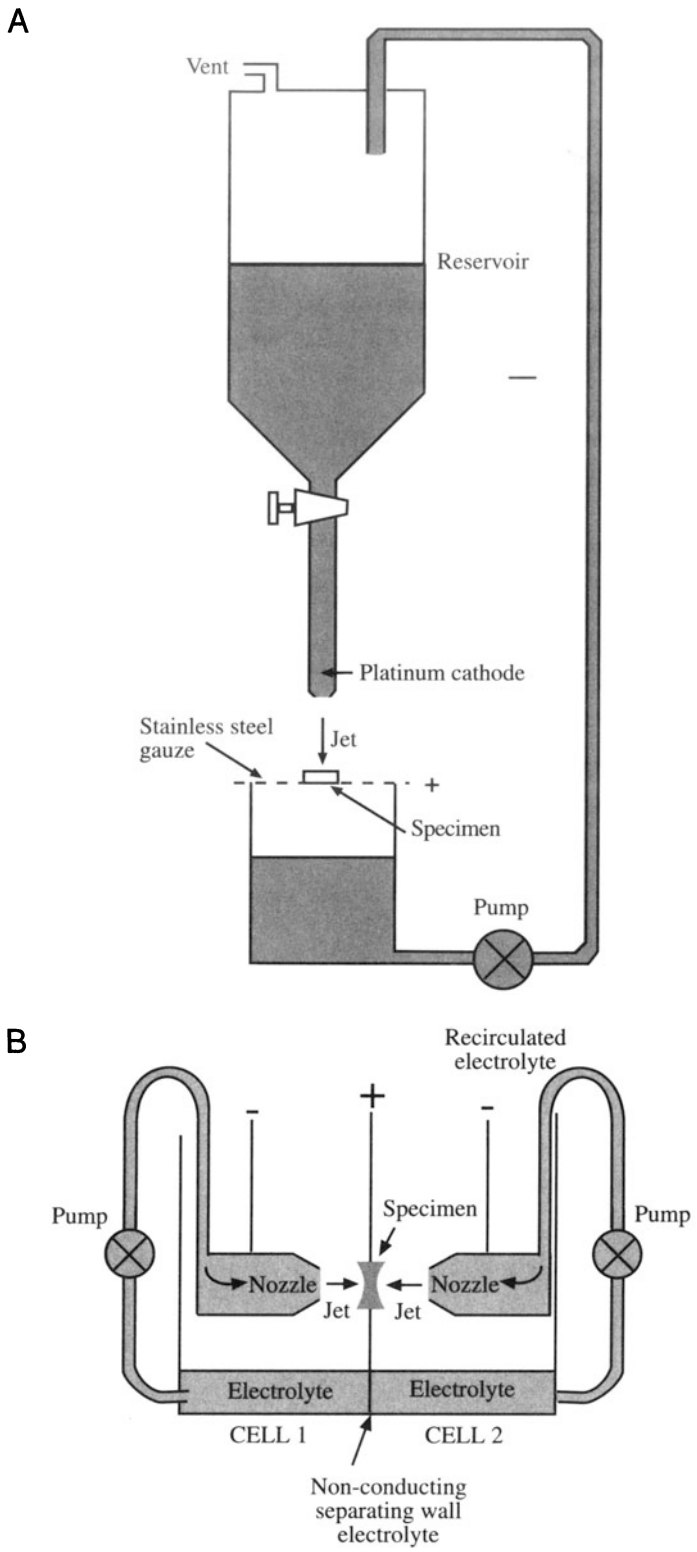


Figure 10.7. (A) Jet electropolishing by allowing a single jet of gravity-fed electrolyte to thin a disk supported on a positively charged gauze. The disk has to be rotated periodically. (B) Schematic of a twin-jet electropolishing apparatus. The positively charged specimen is held in a Teflon holder between the jets. A light pipe (not shown) detects perforation and terminates the polishing.

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 10.7B. 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.

Electropolishing is a "black art." Undoubtedly you get better 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 is shown in Figure 10.8. The variables which you control include the voltage, temperature of the specimen [e.g., cold milling (liquid N_2)], the nature of the ion [Ar, He, or a reactive ion (iodine)], and the geometry (the angle of incidence).

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

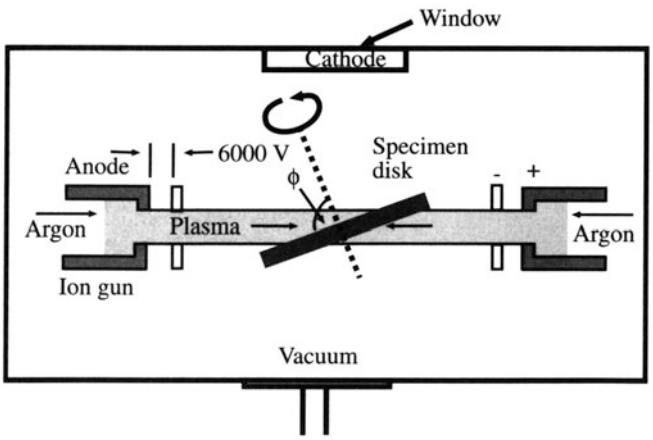


Figure 10.8. Schematic diagram of an ion-beam thinning device: Ar gas bleeds into an ionization chamber where a potential up to 6 keV creates a beam of Ar ions that impinge on a rotating specimen. Although not shown, the whole apparatus is under vacuum, the specimen may be cooled to liquid-N_2 temperatures, and perforation is detected by the penetration of ions through the specimen.

beam to the surface of the specimen. We usually align the ion beam at an angle of 15–25° to the surface. However, Barna (1992) has shown that this angle of incidence should be avoided in many cases since it leads to compositional thinning; use an inclination of ≤5° to avoid preferential thinning. Some implantation will occur so that the chemistry of the near-surface region is changed and the 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 10.9)

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 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:

- *Ion:* mass, energy, charge, and angle of incidence.
- *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 of 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 10.9. 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.

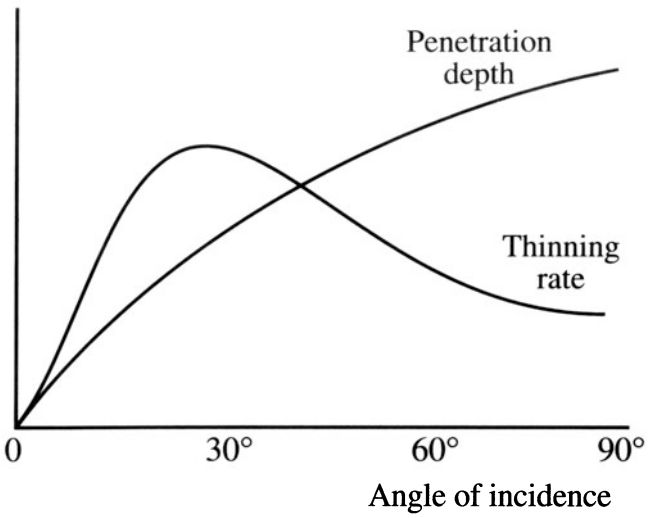


Figure 10.9. Variation in ion penetration depth and thinning rate with the angle of incidence. High incidence angles promote implantation, which is undesirable. The rate of thinning reaches a maximum at ~20° incidence, after which the beam penetrates rather than sputters the sample surface. Initial thinning should start at 20–30°, reducing to <5° as perforation approaches.

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 ceramic 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 first to 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 10.10. 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 (1985) reports that Ar bubbles form in silicon at a depth of ~10 nm after 5-keV thinning. Chemical analysis (EDS) 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 accommodate large amounts of Ar). Cooling the specimen can often reduce contamination and surface damage. It is best to use two ion 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.

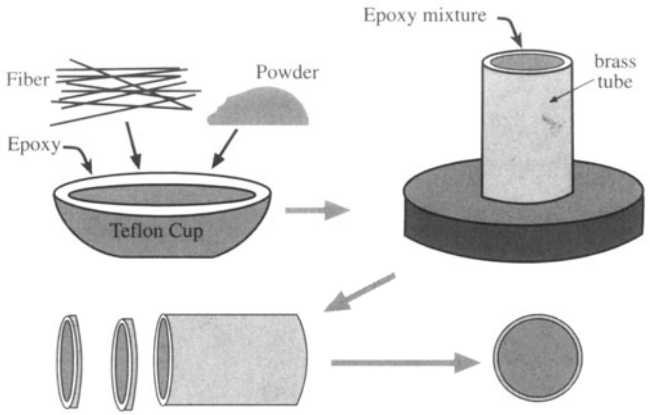


Figure 10.10. Sequence of steps for thinning particles and fibers by first embedding them in epoxy and forcing the epoxy into a 3 mm (outside) diameter brass tube prior to curing the epoxy. The tube and epoxy are then sectioned into disks with a diamond saw, dimpled, and ion milled to transparency.

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 so that it cannot thin, say at 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 10.8 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 $<10^{-5}$ Torr without Ar and 10^{-2}–10^{-3} Torr 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 (1988). Iodine has a clear advantage for InP, where In island formation 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 argon ions (Figure 10.11).
- The PIMS. Gatan's precision ion milling system provides a built-in ion microscope to view the specimen (through SE emission) so that you can then choose a particular area to thin. This is only useful for prethinned specimens where you can locate a very small area (~1 μm × 1 μm) that needs further thinning.
- 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.
- The PIPS. Gatan's precision ion polishing system combines 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

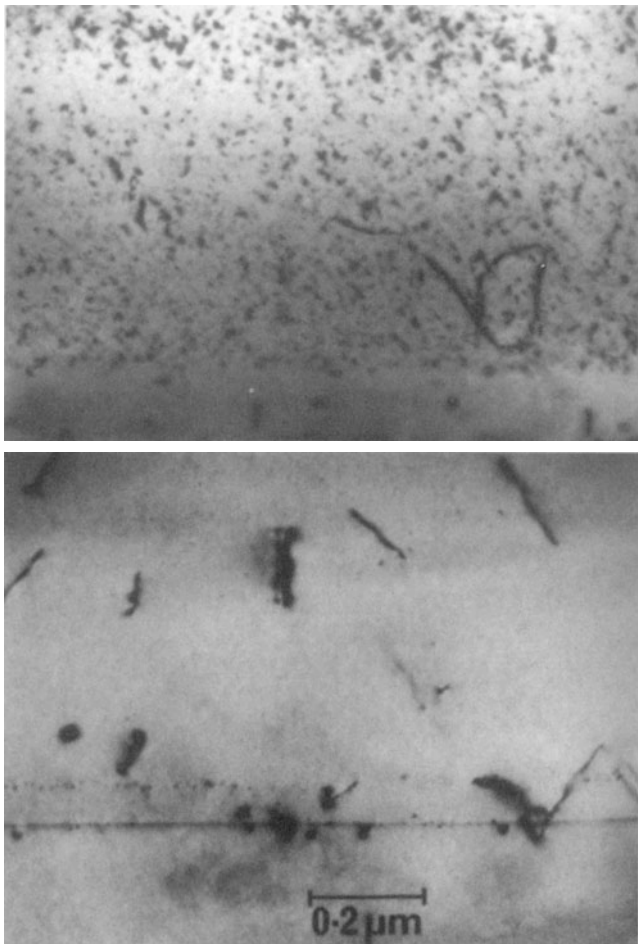


Figure 10.11. BF images of CdTe showing (A) defects (dark spots) in Ar-thinned specimen and (B) undamaged crystal thinned by reactive-iodine ion milling. The residual defects in (B) were formed during CdTe crystal growth.

problems, while the high-power guns ensure reasonable thinning rates (Alani and Swann 1992).

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 the surface which will probably be a combination of amorphous, highly damaged, and implanted material! The chemistry of the layer will be different than 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 used 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 and many details are reported in four Materials Research Society proceedings (Bravman *et al.* 1988; Anderson 1989, 1990; Anderson *et al.* 1992), 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 10.12. 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 samples smaller and encase them in a 3-mm thin-walled tube. Section the 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 metal tube 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.

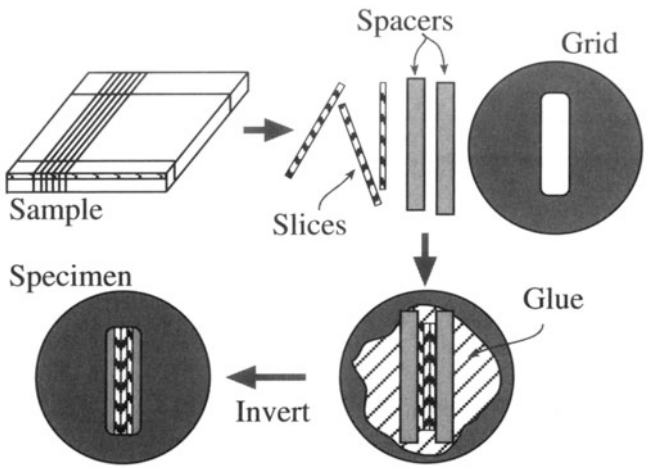


Figure 10.12. Schematic sequence for cross-section specimen preparation; the sample is cut into thin slices normal to the interfaces which are glued together between spacers which could be Si, glass, or some other inexpensive material so that they are wider than the slot in the grid. The "club" sandwich is then itself glued to the grid (over the slot) and ion milled to perforation.

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. 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 10.2) 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 ~1 cm × 1 cm square 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 10.13A. 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 10.13B. If this procedure is done correctly (and this might require several rotations)

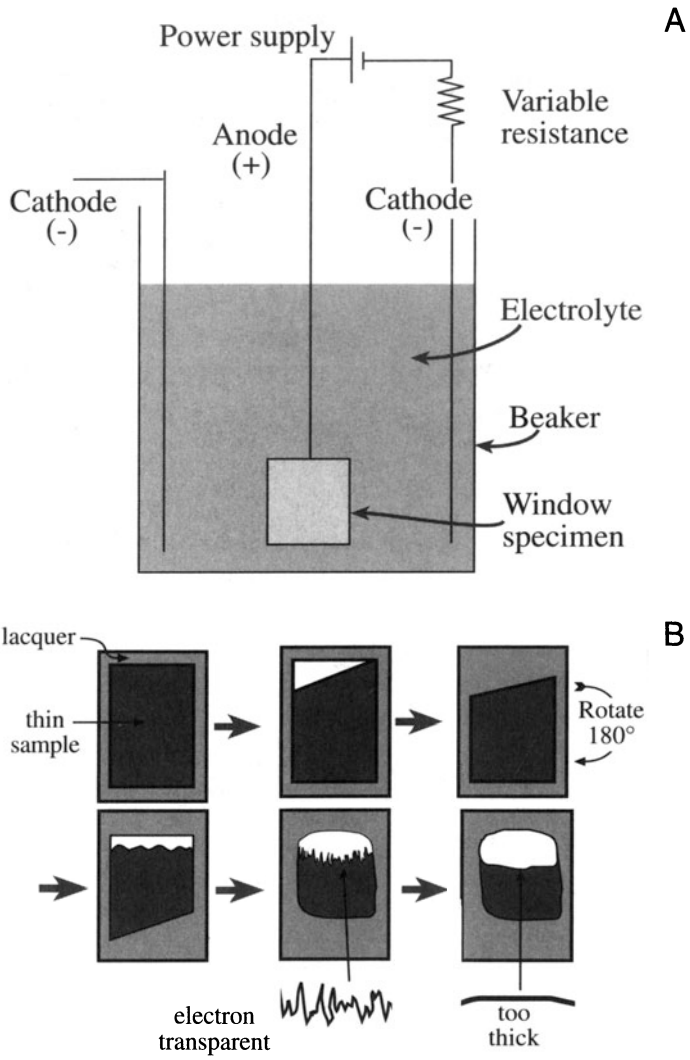


Figure 10.13. Window polishing: (A) A sheet of the metal ~1 cm^2 is lacquered around the edges and made the anode of an electrolytic cell. (B) Progress during thinning: the initial perforation usually occurs at the top of the sheet; lacquer is used to cover the initial perforation and the sheet is rotated 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.

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.

10.6.B. Ultramicrotomy

The microtome has long been used for sectioning biological materials. (A tome is a "piece cut off" but microtome refers to the instrument used to cut a very thin tome, unlike 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 biology or for polymers where the samples tend to be quite soft (Sawyer and Grubb 1987). More recently they have been used for many studies of crystalline materials (Malis 1989). 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 introduces a deformation structure to the materials and therefore is most useful in cases where the defect structure is of secondary 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 the sample if it is soft, or the knife can cause a partly controlled fracture if it is hard/brittle. In either case the limiting process is usually plastic deformation of the specimen. The principles of this technique are shown in Figure 10.14.

You will 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 10.10). We also use epoxy if the sample contains so many interconnected pores that it cannot be thinned mechanically. In this case, 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.

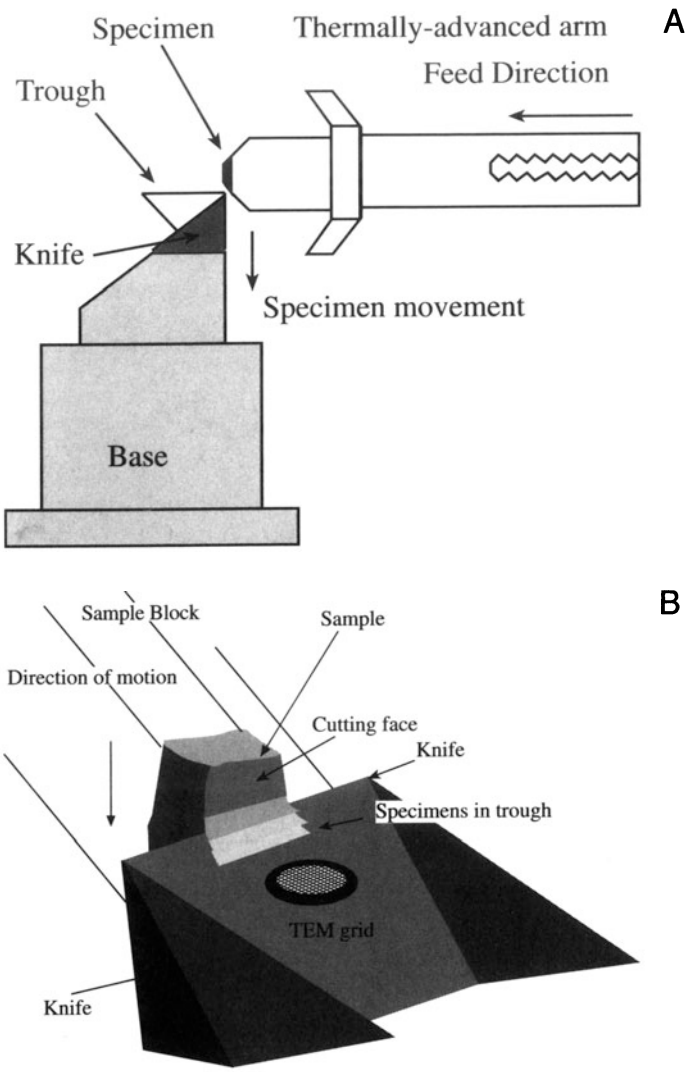


Figure 10.14. Ultramicrotomy: (A) The sample is first embedded in epoxy or some other medium, or the whole sample is clamped and moved across a knife edge. (B) The thin flakes float off onto water or an appropriate inert medium, from where they are collected on grids.

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. The thin particles are too small to be seen 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 distri-

bution 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.

Dust particles found in airborne pollution can be collected simply by leaving the grid and a support film out for a period of time.

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 enhance mass-thickness contrast (see Chapter 22); support the film on a grid for observation. As an alternative (Figure 10.15A) 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 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 10.15B. 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.

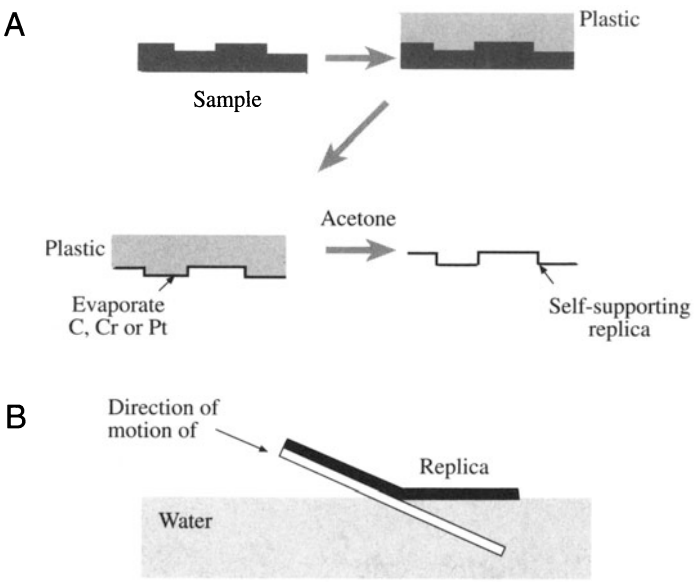


Figure 10.15. (A) Replication of a surface by the two-step method: spray acetone on the surface to be replicated before pressing a plastic (usually cellulose acetate) onto the surface which softens in contact with the acetone; the plastic is removed from the surface when it has hardened and a C, Cr, 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.

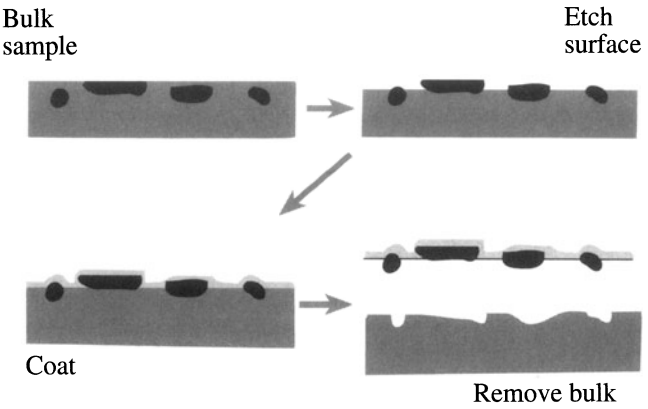


Figure 10.16. Extraction replication: particles embedded in a matrix are revealed by etching the matrix, which leaves the particles standing proud of the surface; a thin amorphous carbon film is evaporated over the particles, then the rest of the matrix is etched away leaving the particles adhering to the carbon film.

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 10.16. 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. Again, oblique shadowing may be useful to enhance image contrast, but not if AEM is to be used.

10.6.E. Cleaving

This is one of the oldest techniques and has been used to make thin specimens of graphite, mica, and other layer ma-

terials which are weakly bonded along one plane. The 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. 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.

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, which are perpendicular to this growth surface. When you are practiced at cleaving the sample as shown in Figure 10.17, 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 thick it is at the position you choose for study. We will find this wedge useful when we discuss image contrast in Part III.

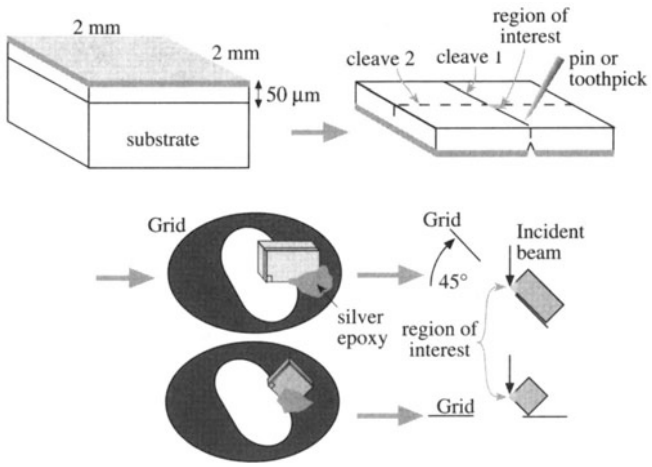


Figure 10.17. The 90°-wedge specimen: prethin to create a 2-mm square of the multilayers on a Si substrate; scribe the Si through the surface layers, turn over, and cleave; inspect to make sure the cleavage is clean, giving a sharp 90° edge; reject if not; mount the 90° corner over the edge of a hole in a Cu grid, then insert in the TEM; note that two different orientations are available from a single cleavage operation.

10.6.G. Lithography

Using 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 10.18 (Brown and Sheng 1988). 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 pro-

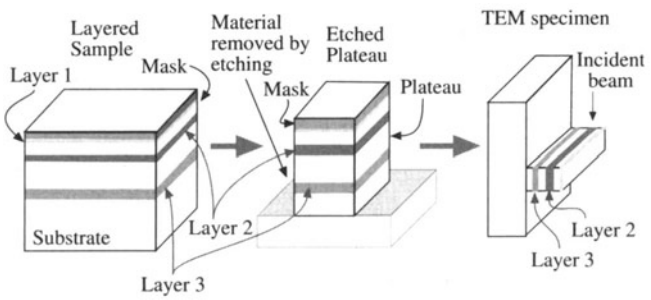


Figure 10.18. Etching of a multilayer sample. Etch away most of the sample, leaving a small etched plateau; mask a region < 50 nm across and etch away the majority of the surrounding plateau. If this thin region is turned 90° and mounted in a specimen holder, the interfaces are now parallel to the electron beam.

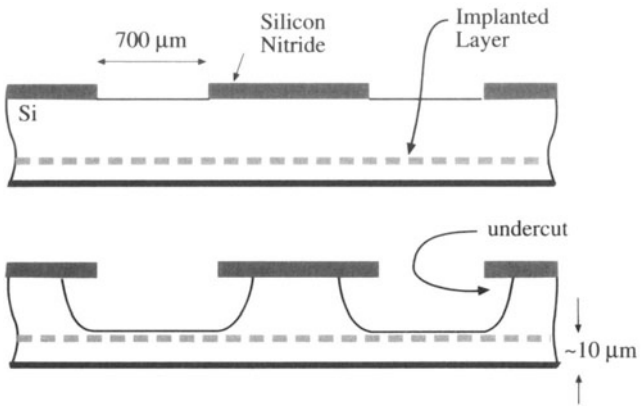


Figure 10.19. Lithographic techniques applied to thinning a multilayer specimen: in the top diagram, the unthinned sample is shown with a grid of Si_3N_4 barrier layers evident. Etching between the barrier layers, shown in the lower figure, produces an undercutting down to the implanted layer (e.g., B) 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.

duced by implanting with boron (Figure 10.19). In both cases, the resulting thin layers may find use as substrate materials for thin-film studies rather than as the subject of study in their own right.

10.7. STORING SPECIMENS

The best advice is to look at your specimen as soon as possible after preparation. If that is not possible, then keep the specimens under optimum conditions. Usually this means keeping them dry (water affects most materials), perhaps in an inert atmosphere (dry nitrogen works well, or a dry-pumped vacuum desiccator) and in an inert container (a 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, your most important specimen is the one most likely to break, bend, interact with sharp tweezers, or jump onto the floor.

Lastly, old specimens can be cleaned by ion polishing. 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.

CHAPTER SUMMARY

Specimen preparation is a craft and there is no substitute for hard work and careful, detailed experimentation as you seek to master it. This is the most tedious aspect of all 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 was 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.

We stress once again that you must know what you want to study in your specimen before you begin specimen preparation. Figure 10.20 is a useful flow chart (Goodhew 1988) which summarizes the various possible options. Be aware of the limitations of the method you choose, particularly the artifacts introduced. In this respect, Table 10.1 (Malis 1989) is a nice summary of the artifacts introduced by various methods.

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.

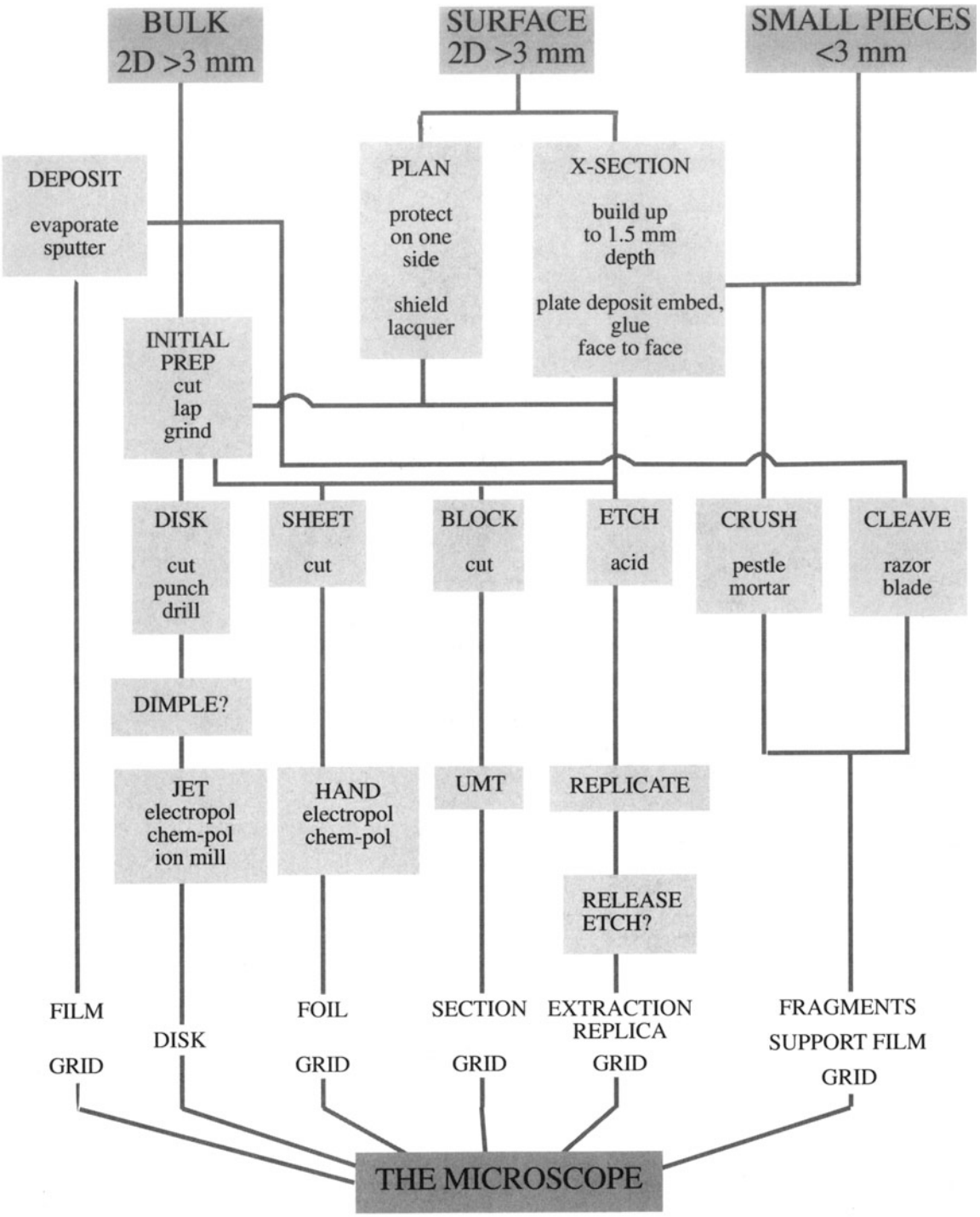


Figure 10.20. Summary flow chart for specimen preparation.

Table 10.1. Artifacts Produced during Specimen Preparation[a]

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)
	■ 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)

[a]EP: electropolished; UM: ultramicrotomed; CD: controlled dimpling; R: extraction replication; IT: ion thinned; TP: tripod polish; C: cleavage (grinding, crushing).

REFERENCES

General References

Anderson, R.M., Ed. (1990) *Specimen Preparation for Transmission Electron Microscopy of Materials, II,* Materials Research Society Symposium Proceedings **199**, MRS, Pittsburgh, Pennsylvania.

Anderson, R.M., Tracy, B., and Bravman, J.C., Eds. (1992) *Specimen Preparation for Transmission Electron Microscopy of Materials, III,* Materials Research Society Symposium Proceedings **254**, MRS, Pittsburgh, Pennsylvania.

Barber, D.J. (1970) *J. Mater. Sci.* **5**, 3.

Bravman, J.C., Anderson, R.M., and McDonald, M.L., Eds. (1988) *Specimen Preparation for Transmission Electron Microscopy of Materials,* Materials Research Society Symposium Proceedings **115**, MRS, Pittsburgh, Pennsylvania.

Echlin, P. (1992) *Low Temperature Microscopy and Analysis,* Plenum Press, New York. Mainly biological but useful for polymers.

Goodhew, P.J. (1984) *Specimen Preparation for Transmission Electron Microscopy of Materials,* Oxford University Press, New York.

Goodhew, P.J. (1985) *Practical Methods in Electron Microscopy, Vol. 11: Thin Foil Preparation for Electron Microscopy,* Elsevier, New York.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, New York. Chapter 2 and Appendix 1, the largest chapter and the largest appendix.

Humphries, D.W. (1992) *The Preparation of Thin Sections of Rocks, Minerals and Ceramics,* Royal Microscopical Society Handbook No. 24, Oxford University Press, New York. Although not aimed at TEM, this handbook contains many helpful ideas.

Kestel, B.J. (1986) "Polishing Methods for Metallic and Ceramic Transmission Electron Microscopy Studies," in ANL-80-120 Rev. 1 (from Argonne National Laboratory). 66 pages of recipes and directions which have all been extensively used by Kestel.

Marcus, R.B. and Sheng, T.T. (1983) *Transmission Electron Microscopy of Silicon VLSI Circuits and Structures,* Wiley, New York.

Reid, N. and Beesley, J.E. (1991) *Practical Methods in Electron Microscopy Vol. 13: Sectioning and Cryosectioning for Electron Microscopy,* Elsevier, Amsterdam. Biological but comprehensive.

Roos, N. and Morgan, A.J. (1990) *Cryopreparation of Thin Biological Specimens for Electron Microscopy: Methods and Applications,* Oxford University Press, New York. Biological again, but a good one to know.

Sawyer, L.C. and Grubb, D.T. (1987) *Polymer Microscopy,* Chapman and Hall, New York. Chapter 4 is 71 pages long, covering specimen preparation for SEM and TEM; ~25% of the book.

Thomas, G. and Goringe, M.J. (1979) *Transmission Electron Microscopy of Metals,* p. 336, Wiley, New York.

Thompson-Russell, K.C. and Edington, J.W. (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science,* Macmillan, Philips Technical Library, Eindhoven, the Netherlands. Many recipes.

Specific References

Alani, R. and Swann, P.R. (1992) in Anderson *et al.,* op. cit., p. 43.

Barna, (1992) in Anderson *et al.,* op. cit., p. 254.

Brown, J.M. and Sheng, T.J. (1988) in Bravman *et al.,* op. cit., p. 229.

Cullis, A.G., Chew, N.G., and Hutchison, J.L. (1985) *Ultramicroscopy* **17**, 303.

Goodhew, P.J. (1988) in Bravman *et al.*, op. cit., p. 51.

Klepeis, S.J., Benedict, J.P., and Anderson, R.M. (1988) in Bravman *et al.,* op. cit., p. 179.

Malis, T.F. (1989) in *Microbeam Analysis—1989* (Ed. P.E. Russell), p. 487, San Francisco Press, San Francisco.

Médard, L., Jacquet, D.A., and Sarlorius, R. (1949) *Rev. Mat.* **46**, 549.

Diffraction

II

Diffraction Patterns

11

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 diffracted 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 which we will examine in Chapters 19 to 21.

We need to know how to use the information which 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 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.

Diffraction Patterns

11

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 that 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.?
- 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?

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 ~0.001°) 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 both 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 SAD patterns are not always the most useful DPs, since CBED (Chapters 20 and 21) can give you other useful information. Nevertheless, we are emphasizing SADs 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

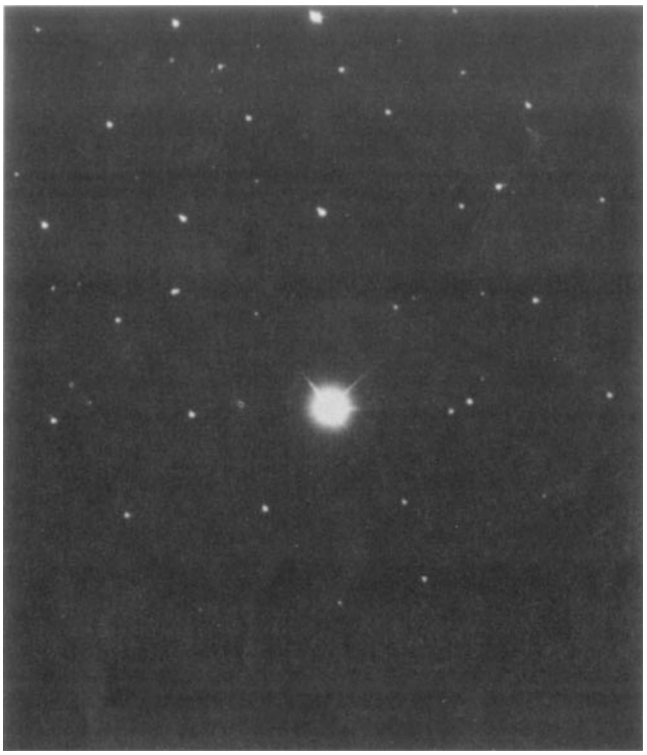


Figure 11.1. An experimentally observed diffraction pattern showing the central, intense, direct beam and an array of diffraction spots from different atomic planes. Such a pattern, with sharply focused spots, is best obtained by underfocusing the beam.

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 usually 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 the television, but remember that in TEM we are using very high energy electrons (≥100 keV rather than 20 keV used in a TV). References to 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, in practice, several seconds to a minute should usually be used.

Much of our 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.C. 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.

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 the Bragg Law. In this chapter we'll derive the Bragg 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 two planes 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 3.9.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.

We'll only consider plane wavefronts, i.e., the wavefront is flat and **k** is normal to this wavefront. The diagram in Figure 11.3a,b defines vectors $\mathbf{k}_I$, $\mathbf{k}_D$, and **K** and gives us the following important equation (which is just vector addition)

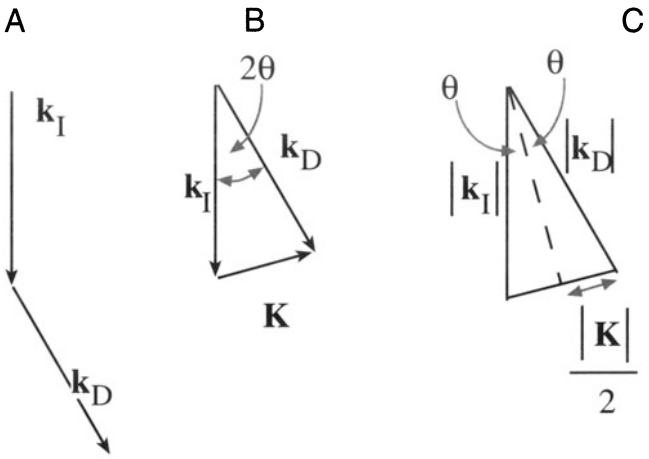


Figure 11.3. Definition of the scattering vectors: (a) the incident wavefront normal is $\mathbf{k}_I$, the diffracted wave normal is $\mathbf{k}_D$; (b) **K** is the difference vector ($= \mathbf{k}_D - \mathbf{k}_I$); (c) sin θ is defined as $K/2k_I$.

$$\mathbf{K} = \mathbf{k}_D - \mathbf{k}_I \qquad [11.1]$$

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

$$|\mathbf{k}_I| = |\mathbf{k}_D| = \frac{1}{\lambda} = |\mathbf{k}| \qquad [11.2]$$

providing the energy of the electron is unchanged during diffraction, i.e., the scattering process is elastic. From Fig-

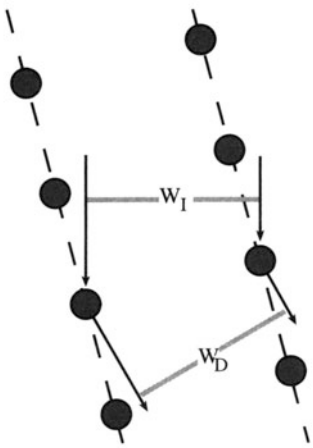


Figure 11.2. Scattering from two planes of atoms. W_I and W_D are the incident and diffracted wavefronts, respectively.

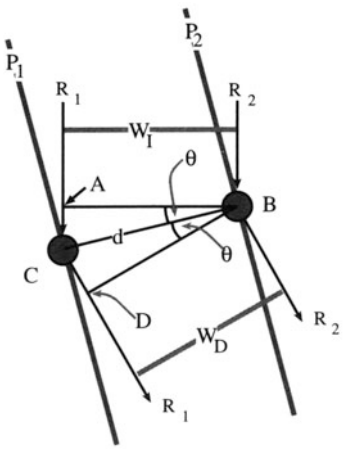


Figure 11.4. Two beams are scattered from two points, C and B, which lie on different planes, P_1 and P_2. The rays travel different distances, giving a path difference of AC + CD.

ure 11.3C we can write down an expression for θ using simple trigonometry

$$\sin\theta = \frac{|\mathbf{K}|/2}{|\mathbf{k}_I|} \qquad [11.3]$$

or

$$|\mathbf{K}| = \frac{2\sin\theta}{\lambda} \qquad [11.4]$$

|K|, like $|\mathbf{k}_I|$, has units $Å^{-1}$ if λ is measured in Å. **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 θ)/λ remember that it is just **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

$$AC + CD = 2d\sin\theta \qquad [11.5]$$

which is the basis for the Bragg 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 in the special case when θ equals the Bragg angle, θ_B, equation 11.4 becomes

$$|\mathbf{K}| = \frac{2\sin\theta_B}{\lambda} \qquad [11.6]$$

When θ is θ_B, the path difference in equation 11.5 is $n\lambda$, where *n* is any integer, and the equation becomes

$$n\lambda = 2d\sin\theta_B \qquad [11.7]$$

which is Bragg's Law (equation 3.22). If *n* is 1

$$2\sin\theta_B = \frac{\lambda}{d} \qquad [11.8]$$

but we already know from equation 11.6 that, at the Bragg angle,

$$2\sin\theta_B = \lambda\,|\mathbf{K}| \qquad [11.9]$$

so when we are at the Bragg angle, the magnitude of the vector **K** has a special value, K_B,

$$|\mathbf{K}_B| = \frac{1}{d} \qquad [11.10]$$

and we define this vector, $\mathbf{K}_B$, to be **g** so that

$$\mathbf{K}_B = \mathbf{g} \qquad [11.11]$$

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.

Don't forget: we are really dealing with diffraction, not reflection, and we derived Bragg's Law by considering just two atoms. The reason that this derivation of Bragg's Law is not valid 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° to 20°) angles; you will find it useful to remember the order of magnitude of these numbers.

> Remember that 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.

> 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 $2d \sin \theta$.

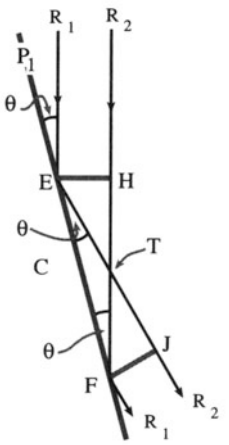


Figure 11.5. Two beams are scattered from two points, E and F, which lie on the same plane P_1. This simple diagram shows that the two beams travel the same distance since triangles EHF and FJE are congruent.

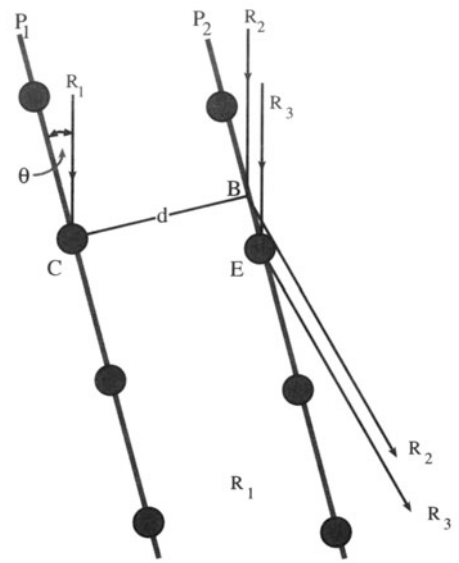


Figure 11.6. Scattering from three points on two planes. The path difference for scattering from points B and C is $2d \sin \theta$, so the path difference for scattering from points C and E is also $2d \sin \theta$. Hence scattering in the direction of the diffracted beam from all points shown will be in phase if $2d \sin \theta = n\lambda$.

This result is summarized in Figure 11.6. Rays R_1, R_2, and R_3 all scatter in phase, if $\theta = \theta_B$.

Next, we extend this analysis to include many parallel planes each a distance d from its neighbors, as is shown in Figure 11.7.

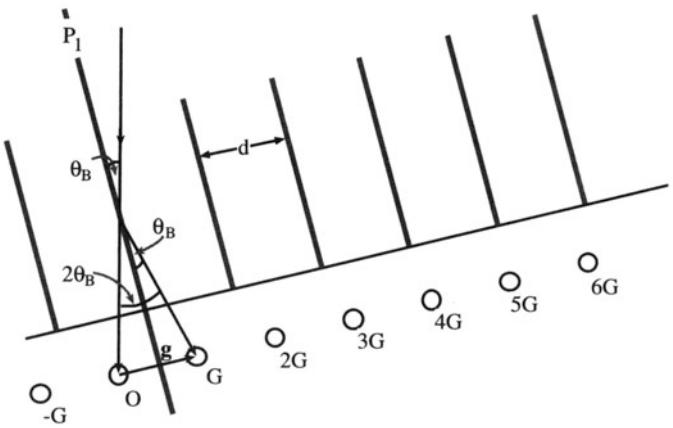


Figure 11.7. Diffraction from a set of planes a distance d apart. The planes have been oriented to be in the Bragg diffracting condition (θ_B is the incident angle). Note that the planes are not parallel to the incident beam. The resultant diffraction spots (reciprocal lattice points) are labeled G, 2G, etc. The vector **g** from the origin (O) to the first diffraction spot G is normal to the diffracting plane.

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: $hU + kV + \ell W = 0$.

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**.

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, O, 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.

The vector $\bar{\mathbf{g}}$ is pronounced "bar g" and is -G, pronounced "minus g" (!); you will also hear $\bar{\mathbf{g}}$ pronounced "g bar."

These other reflections (*n***g**, where $n \neq 1$), called higher-order reflections, are particularly important in TEM. Pictorially, you can imagine them as arising from the interference 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 halfway 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

$$2\left(\frac{d}{2}\right)\sin\theta = \lambda \qquad [11.12]$$

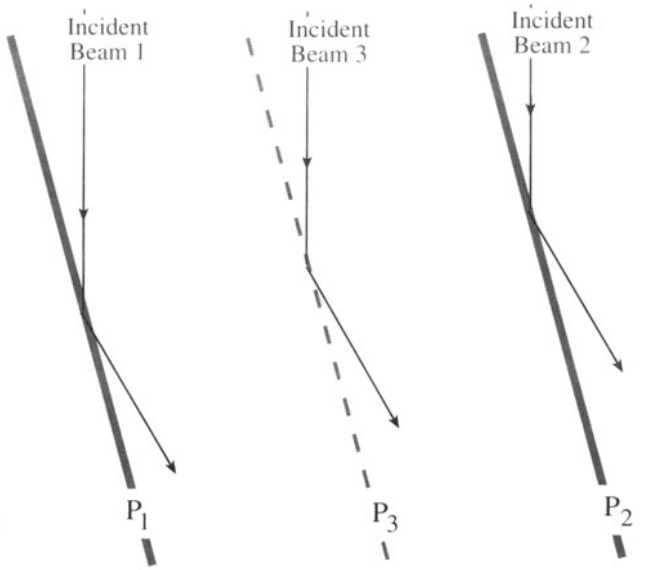


Figure 11.8. Scattering from three planes with plane P_3 positioned exactly halfway between planes P_1 and P_2.

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

$$|\mathbf{g}_2| = 2\,|\mathbf{g}| \qquad [11.14]$$

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

$$2\left(\frac{d}{n}\right)\sin\theta = \lambda \qquad [11.15]$$

or rewrite this as

$$2d\sin\theta = n\lambda \qquad [11.16]$$

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 11.12 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 which 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 nonmathematician. 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, it can be thought of in quite a simple manner. As you can see in Figure 11.9, 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. 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.

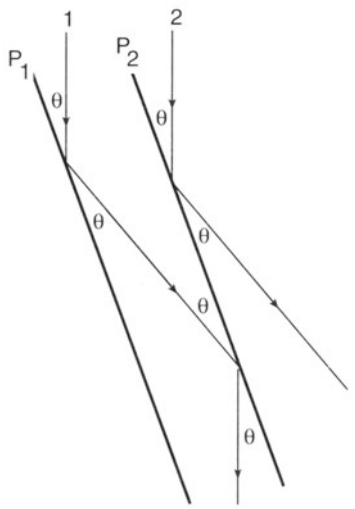


Figure 11.9. The beam can be scattered more than once. Any beam which is oriented so as to be Bragg-diffracted once is automatically in the ideal orientation to be rediffracted. This gives rise to the phenomenon of dynamical scattering.

11.7. USE OF INDICES IN DIFFRACTION PATTERNS

In Chapter 18 we'll teach you how to index diffraction patterns, i.e., how to associate a spot in the diffraction pattern 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 set of parallel crystal planes is defined by the Miller indices $(hk\ell)$ and a set of such planes is $\{hk\ell\}$. We define the direct beam as the 000 reflection and each diffracted beam as a reflection with different $hk\ell$ indices. It is a crystallographic convention to refer to the diffraction spot from a specific $(hk\ell)$ plane as $hk\ell$, i.e., without the parentheses. If we assign $hk\ell$ to **g**, then the second-order (2**g**) spot is $2h\ 2k\ 2\ell$, the 3**g** spot is $3h\ 3k\ 3\ell$, etc. Similarly, the $\bar{\mathbf{g}}$ reflection is $\bar{h}\bar{k}\bar{\ell}$. 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 $(hk\ell)$ 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: $hU + kV + \ell W = 0$.

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). This results from the physical shape of the TEM specimen and will be discussed 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.

SAD patterns are sharply focused spot patterns, which we use to select reflections for all imaging modes. We can easily associate the sharp spots with our diffraction vectors, **g**.

CBED patterns 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 CBED patterns while we develop diffraction theory and then devote two chapters 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 cross the specimen, so 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 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.

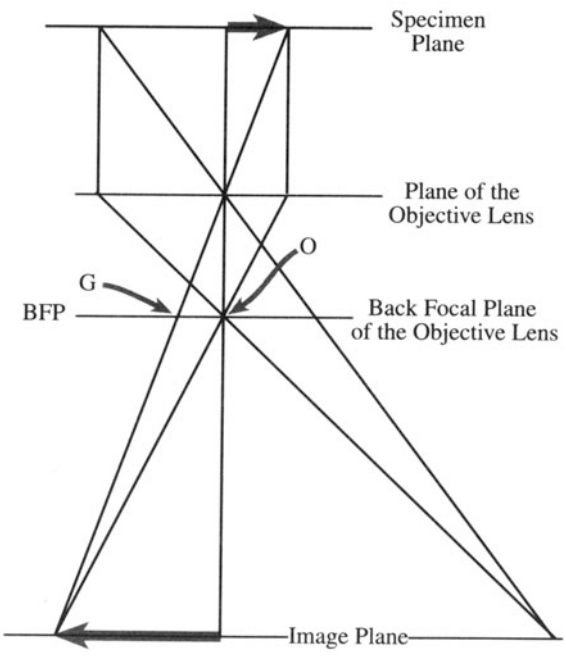


Figure 11.10. The diffraction pattern is formed at or close to the back focal plane of the objective lens. O is the direct beam and G is a diffracted beam.

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 are:

- Be sure that you are at the eucentric focus position, with an image of the area of interest focused on the screen.
- Insert the SAD aperture.
- Remove the objective aperture.
- Focus the SAD aperture.
- Switch to diffraction mode.
- Spread the beam using C2, within the limits imposed by your specimen.
- 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 aperture will select a 2-μm 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, which shows the "real" version of Figure 11.10 since the objective lens is not perfect. As we saw in Chapter 6, the beams which are further away from the optic axis are bent more strongly as they pass through the objective lens. For rays entering the lens at an angle β to the optic axis, the image formed at magnification M, is translated a distance r_M given by

$$r_M = M\, C_s\, \beta^3 \qquad [11.17]$$

So the area we select using the SAD aperture corresponds to the area PP_I 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

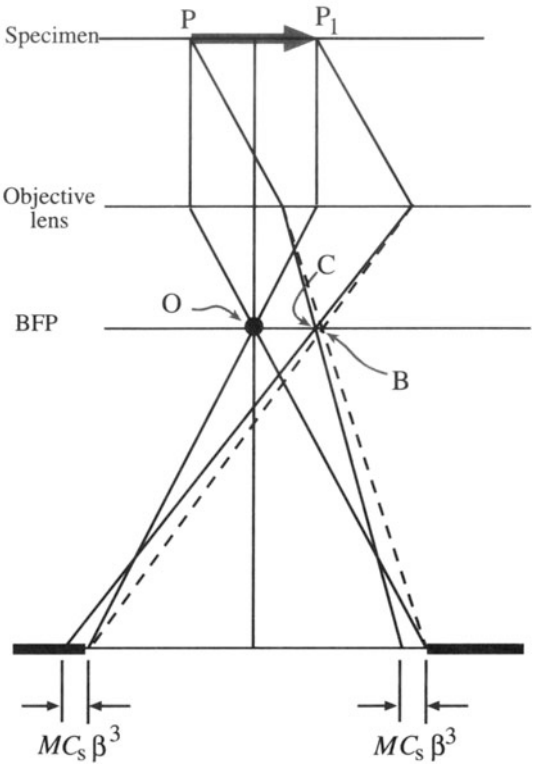


Figure 11.11. Formation of an SAD pattern showing that there is an error in selecting the area if the beams do not travel at the same angle to the optic axis. This difference is due to spherical aberration in the objective lens. B is the diffraction spot position for a perfect lens and C is the spot position with spherical aberration.

reflection, from this area would not contribute to the SAD pattern. 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 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

$$y = D\beta \qquad [11.18]$$

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 ($0.5D_{Ob}$), of ~3 μm (see Figure 6.14). You will still find on many TEMs that the aperture in the

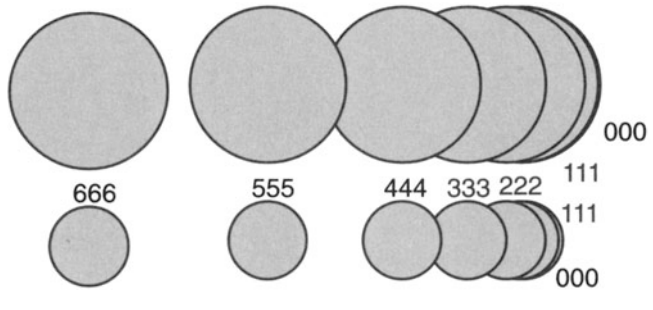


Figure 11.12. Schematic diagram showing the effective error in area selection, due to spherical aberration, for different reflections in the 111 systematic row for Al (a_0 = 4.04 Å) assuming 100-keV electrons and C_s = 3 mm. The 000 and 111 disks almost exactly overlap (the translation is 13 nm). The 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) modern IVEM
111	13	1.2
222	100	9.1
333	350	31.9
444	760	69.3
555	1620	150
666	2800	250

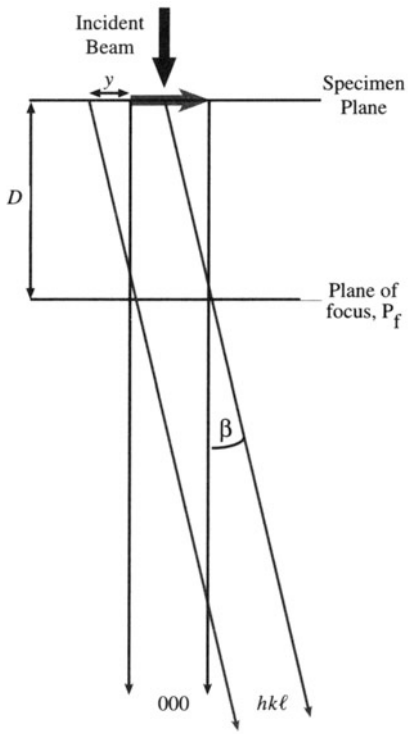


Figure 11.13. If the lens is not focused on the SAD plane, images associated with the different **g** vectors will be shifted with respect to one another. D is the defocus. The shift in the selected area is given by $y = D\beta$.

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 based on the middle column in Table 11.1, which was first given by Hirsch *et al.* (1977) 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 0.1968 nm. The values for $C_s\beta^3$ then become much smaller, as shown in the right-hand column in Table 11.1. Clearly you could now use a much smaller SAD aperture, and 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 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 SAD pattern. The reflections that were bright now correspond to bright disks; the 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 of selection of SAD patterns is one reason that CBED patterns 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.

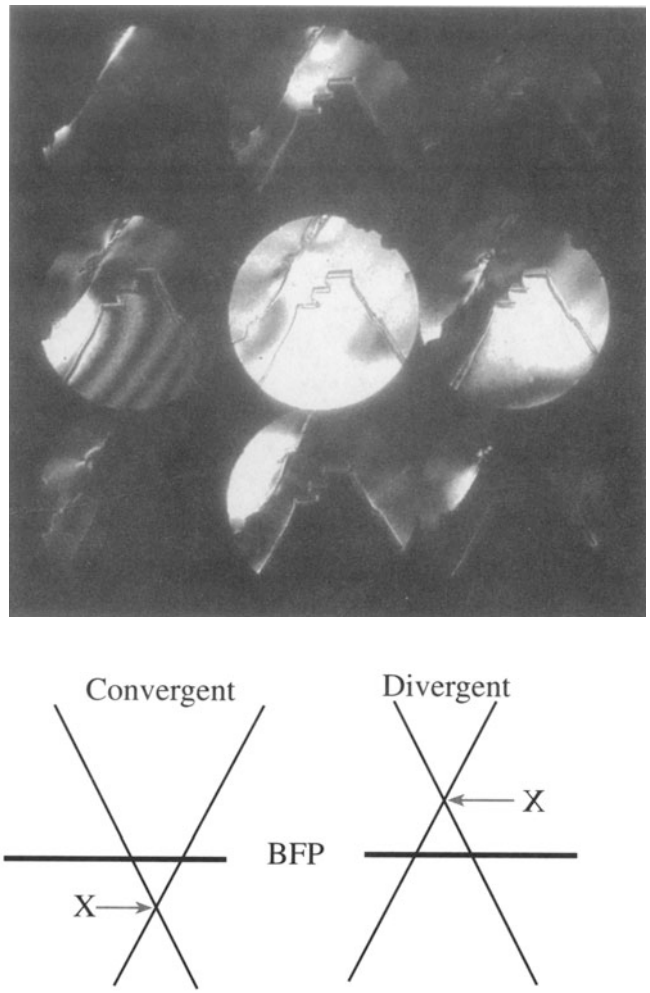


Figure 11.14. (A) Multiple dark-field images formed by defocusing the SAD pattern revealing dark-field images in each diffraction disk. Close inspection reveals that each image (of a twin boundary) is slightly shifted from the adjacent images, reflecting the increased error in area selection for higher-order 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).

To record the SAD pattern you should never use an exposure of <10 s. 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, say 10 s, 30 s, and 100 s. 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 diffraction pattern. Just for the exercise, focus the spots in the

SAD pattern 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 electron diffraction so useful.

Deciding which pattern is best really depends on what information you require. If you would like to see fine detail in your SAD pattern, 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. In fact, it's good practice to record patterns with a range of exposures, from a few seconds to 100 s if necessary. 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 diffraction patterns in the future.

Cooling the 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-line patterns (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 a problem in the specimen contacting the specimen holder or the holder contacting the ground.

CHAPTER SUMMARY

Diffraction patterns 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 $hk\ell$. The diffracting planes are all in a specific zone axis *UVW*, which we can define as parallel to the incident beam direction.

REFERENCES

General References

Bragg, W.L. (1965) *The Crystalline State,* **I** (Ed. W.L. Bragg), Cornell University Press, Ithaca, New York (first published in 1933).

James, R.W. (1965) *The Optical Principles of the Diffraction of X-rays, The Crystalline State,* **II** (Ed. W.L. Bragg), Cornell University Press, Ithaca, New York (first published in 1948).

Schwartz, L.H. and Cohen, J.B. (1977) *Diffraction from Materials*, Academic Press, New York.

Hecht, E. (1987) *Optics*, Addison-Wesley, Reading, Massachusetts.

Specific References

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, New York.

Thinking in Reciprocal Space

12

CHAPTER PREVIEW

In the previous chapter, you've already encountered vectors **k** and **g** and seen that they have lengths with units $Å^{-1}$ or 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 solid-state 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 visualize how diffraction patterns will vary as the orientation and physical characteristics of the specimen vary.

Thinking in Reciprocal Space

12

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 you a physical picture of what happens when a crystal diffracts.

Think of any crystal as having two lattices, one real and the other reciprocal.

Historical Note: The reciprocal lattice was rediscovered independently by Ewald and Laue in 1911–14, but it had been described by Gibbs in 1881 and by Bravais (in a somewhat less useful form) in 1850! The discussion of Ewald's contribution to the subject is recommended reading (Ewald 1962).

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.

In the reciprocal lattice, sets of parallel $(hk\ell)$ atomic planes are represented by a single point located a distance $1/d_{hk\ell}$ 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

$$\mathbf{r}_n = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \qquad [12.2]$$

where the vectors **a**, **b**, and **c** are the unit-cell translations in real space while n_1, n_2, and n_3 are all integers.

Any reciprocal lattice vector, **r***, can be defined in a similar manner

$$\mathbf{r}^* = m_1\,\mathbf{a}^* + m_2\,\mathbf{b}^* + m_3\,\mathbf{c}^* \qquad [12.3]$$

where **a** *, **b***, and **c*** are the unit-cell translations in reciprocal space while m_1, m_2, and m_3 are all integers. These new vectors are defined by the relations

$$\mathbf{a}^* \cdot \mathbf{b} = \mathbf{a}^* \cdot \mathbf{c} = \mathbf{b}^* \cdot \mathbf{c} = \mathbf{b}^* \cdot \mathbf{a} = \mathbf{c}^* \cdot \mathbf{a} = \mathbf{c}^* \cdot \mathbf{b} = 0 \quad [12.4]$$

In words, **a*** is normal to both **b** and **c**, etc. We also define that

$$\mathbf{a}^* \cdot \mathbf{a} = 1; \mathbf{b}^* \cdot \mathbf{b} = 1; \mathbf{c}^* \cdot \mathbf{c} = 1 \quad [12.5]$$

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.

Equation 12.5 then uniquely defines the length of the vector **a*** in terms of the length of the vector **a**. Therefore, this equation gives 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. 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

$$\mathbf{a}^* = \frac{\mathbf{b} \wedge \mathbf{c}}{V_c} \quad [12.6]$$

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 this form for use later

$$\mathbf{K} = \xi\,\mathbf{a}^* + \eta\,\mathbf{b}^* + \zeta\,\mathbf{c}^* \quad [12.7]$$

A particularly important reciprocal lattice vector is the vector $\mathbf{g}_{hk\ell}$, which is defined as

$$\mathbf{g}_{hk\ell} = h\,\mathbf{a}^* + k\,\mathbf{b}^* + \ell\,\mathbf{c}^* \quad [12.8]$$

where h, k, and ℓ are all integers and together define the plane $(hk\ell)$.

The definition of the plane $(hk\ell)$ is that it cuts the a, b, and c axes at $1/h$, $1/k$, and $1/\ell$, 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 $(hk\ell)$ plane, are normal to the vector $\mathbf{g}_{hk\ell}$ 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 $\mathbf{g}_{hk\ell}$ must be *normal* to the plane $(hk\ell)$.

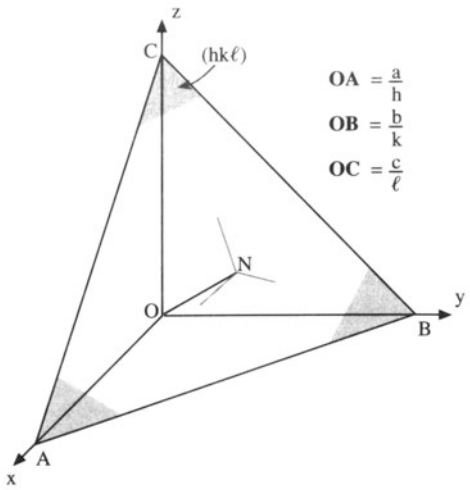


Figure 12.1. The plane ABC has Miller indices $(hk\ell)$. The vectors **OA, OB,** and **OC** have lengths a/h, b/k, and c/ℓ. The vector **ON,** which may be written as **n,** is normal to the plane $(hk\ell)$. In the text we see that the reflection, **g,** which is associated with diffraction from the $(hk\ell)$ planes, is parallel to **n** and normal to all vectors in $(hk\ell)$.

$$\left(\frac{\mathbf{b}}{k} - \frac{\mathbf{a}}{h}\right) \cdot \left(h\mathbf{a}^* + k\mathbf{b}^* + \ell\mathbf{c}^*\right) = 0 \quad [12.9]$$

The vectors **AB**, **BC**, and **CA** all lie in the plane $(hk\ell)$ and each is normal to $\mathbf{g}_{hk\ell}$. All that we now have to prove is that the length of the vector $|\mathbf{g}_{hk\ell}|$ is given by $(d_{hk\ell})^{-1}$. To show this relationship, consider a unit vector, **n**, normal to the plane (i.e., parallel to $\mathbf{g}_{hk\ell}$) 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.)

$$\mathbf{n} \cdot \frac{\mathbf{a}}{h} = \frac{\mathbf{g}}{|\mathbf{g}|} \cdot \frac{\mathbf{a}}{h} = \frac{\left(h\mathbf{a}^* + k\mathbf{b}^* + \ell\mathbf{c}^*\right)}{|\mathbf{g}|} \cdot \frac{\mathbf{a}}{h} = \frac{1}{|\mathbf{g}|} \quad [12.10]$$

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 $(hk\ell)$ planes, so that

$$d_{hk\ell} = \frac{1}{|\mathbf{g}|} \quad [12.11]$$

as we required.

- The definition of the $hk\ell$ indices is OA = a/h; OB = b/k; OC = c/ℓ.
- The plane ABC can then be represented as $(hk\ell)$.

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 ≥100 keV, whereas those in solids are ~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: $(hk\ell)$ is shorthand notation for a particular vector in reciprocal space because it is normal to the $(hk\ell)$ plane in real space; $\{hk\ell\}$ is then the general form for these reciprocal lattice vectors. $[UVW]$ is a particular plane in reciprocal space, e.g., it may contain many $\{hk\ell\}$ points so that in real space it would be a direction—the zone axis for the $\{hk\ell\}$ real-space planes (see Table 12.1). 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.
- *Warning:* the real-lattice vectors and the reciprocal-lattice vectors with the same indices (e.g., [123] and the normal to the plane (123)) are *parallel only* in the case of cubic materials. In other 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.

TABLE 12.1. Notation for Planes, Directions, and Reflections

Real space	Reciprocal space	Indices
Particular direction	Particular plane	$[UVW]$
General direction	General plane	$<UVW>$
Particular plane	Particular direction	$(hk\ell)$
General plane	General direction	$\{hk\ell\}$
Diffracting plane	Indexed reflection	$hk\ell$

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 11.7) 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 equal to the magnitude of **g**, i.e., $1/d$. Therefore, at the Bragg condition we can write

$$\frac{2\sin\theta}{\lambda} = \frac{1}{d_{hk\ell}} \qquad [12.13]$$

i.e.

$$\lambda = 2d\sin\theta \qquad [12.14]$$

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

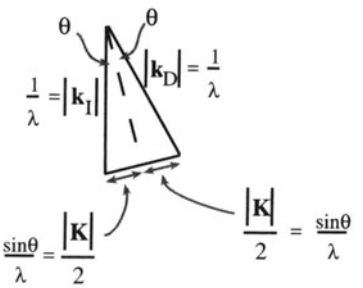


Figure 12.2. The geometric relationship between $\mathbf{k}_I$, $\mathbf{k}_D$, **K**, θ, and λ.

$$\mathbf{K}\cdot\mathbf{r}_n = N \qquad [12.15]$$

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 out 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

$$\mathbf{K}\cdot\mathbf{a} = h \qquad [12.16]$$

$$\mathbf{K}\cdot\mathbf{b} = k \qquad [12.17]$$

$$\mathbf{K}\cdot\mathbf{c} = \ell \qquad [12.18]$$

Of course, these equations are the same Laue diffraction conditions which we introduced back in Section 3.9.B, as given in equation 12.15. In Section 11.5 we quoted Bragg's Law, with an "n," as

$$n\lambda = 2d\sin\theta \qquad [12.19]$$

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

$$n\,d_{nh,\,nk,\,n\ell} = d_{hk\ell} \qquad [12.20]$$

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 rod 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 diffraction pattern 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.* (1992), along with several of his other papers; the articles collected in this review 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 diffraction pattern 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

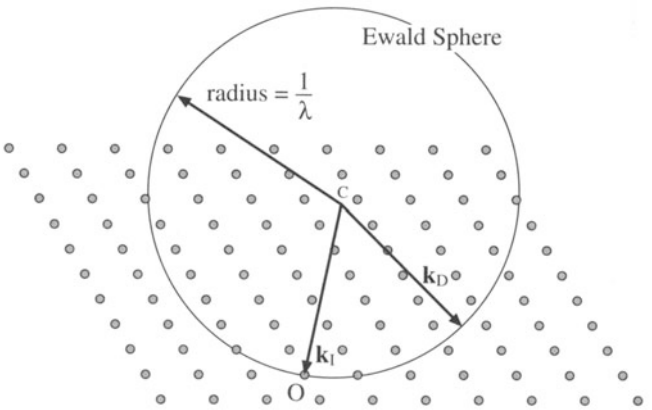


Figure 12.3. The Ewald sphere of reflection is shown intersecting a noncubic array of reciprocal-lattice points. The vector **CO** represents $\mathbf{k}_I$, the wave vector of the incident wave, and O is the origin of the reciprocal lattice. $\mathbf{k}_D$ is any radius vector. When the radius of the 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, as shown. 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.

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 $(hk\ell)$, i.e., parallel to the normal to this plane.

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.

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 point of 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.

The vector **CO** is $\mathbf{k}_1$ 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 $d_{hk\ell}$ 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 0.0037 nm 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 pattern 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 diffraction pattern. Clearly there are many spots, 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

$$\mathbf{K} = \mathbf{g} + \mathbf{s} \qquad [12.21]$$

This vector, **s**, is a measure of how far we deviate from the exact Bragg condition.

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 **CG**; in the second, $\mathbf{s}_z$ is chosen to be parallel to vector **CO**, the incident wave vector. A third special situation would be

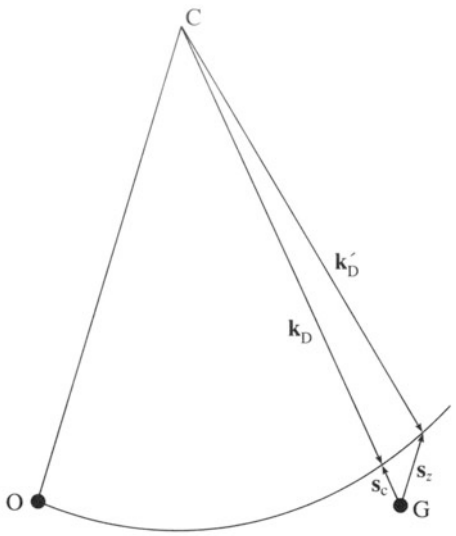


Figure 12.4. Two special values of **s** are illustrated. When $\mathbf{k}_D$ lies along **CG** then $\mathbf{s}_c$ is parallel to **CG**. Alternatively, we can choose **s** to be parallel to the incident beam direction **CO**; then $\mathbf{s} = \mathbf{s}_z$ and $\mathbf{k}_D$ becomes $\mathbf{k}'_D$. In each case, $\mathbf{k}_D$ ends on the Ewald sphere.

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: 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.

When we drew Figure 12.4, you noticed that we placed the point G outside the Ewald sphere. By convention, we define the sign of **s** in this case to be negative, while **s** is positive when G is inside 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, the row of reciprocal lattice points (only G is shown) is essentially at 90° to the incident beam. If we take all such rows, we define a plane of points 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.

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 diffraction patterns 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.

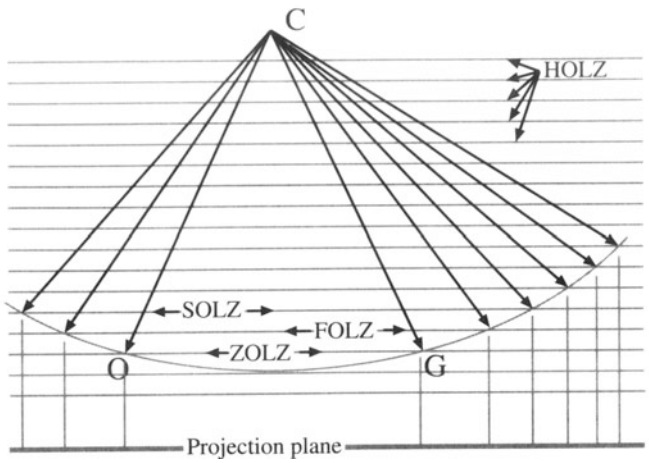


Figure 12.5. The Ewald sphere intercepts points in higher-order Laue zones (HOLZ) at large angles to the incident-beam direction. If the radius of the sphere increases (higher kV beam) then the sphere flattens and the HOLZ interception is at still larger angles.

- Second, if we tilt the beam above the specimen, the Ewald sphere moves, because $\mathbf{k}_I$ tilts, because C moves!

Convince yourself of this. The diffraction patterns with different values of **s** may appear identical, but be cautious (more about this in next chapter). The difference between these two processes is shown in Figure 12.6.

We'll conclude this section by giving you an experimental diffraction pattern to think about. Figure 12.7 shows a DP from a slightly misoriented twin boundary: all you need to know is that different grains are diffracting to give two different DPs. 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 that first meets the eye, as we'll see in Chapter 19.

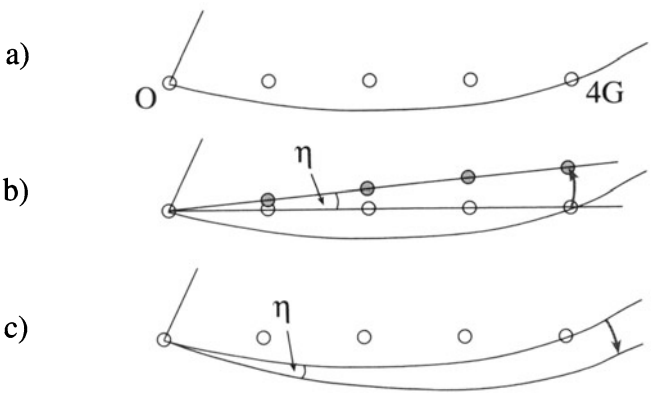


Figure 12.6. In (a) $\mathbf{s}_z = 0$ for 4G. We can change $\mathbf{s}_z$ in two ways: (b) if we tilt the specimen through angle η, the row of spots moves inside the sphere; (c) if we tilt the beam through η above the specimen, in the opposite direction, the sphere moves outside the row of spots.

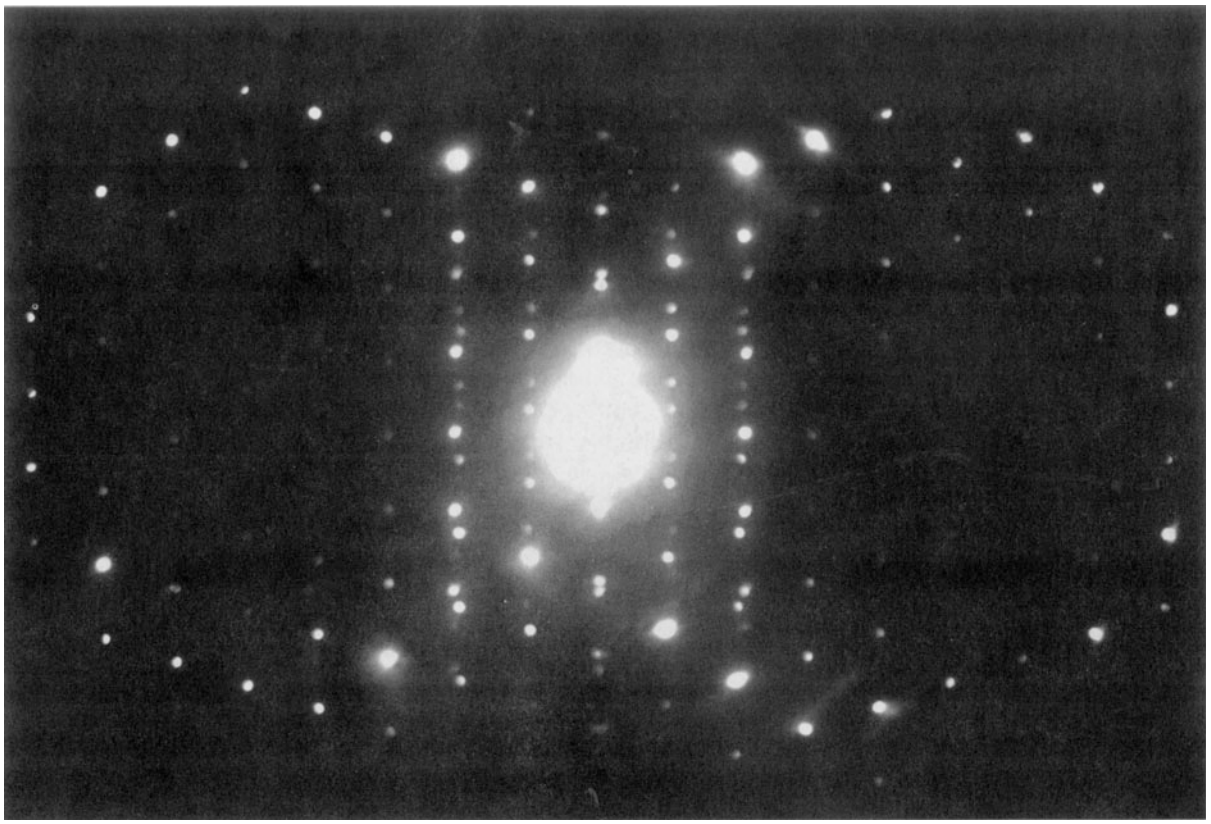


Figure 12.7. Diffraction pattern taken across a twin boundary in $MgAl_2O_4$ spinel. The rings of bright spots show where the Ewald sphere intercepts the reciprocal lattice of the crystals either side of the twin boundary.

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 diffraction pattern? Well, we know that $\mathbf{k}_I - \mathbf{k}_D = \mathbf{K} = \mathbf{g}$ where $|\mathbf{g}|$ is d^{-1}. Therefore, $\mathbf{g}$ 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 $\mathbf{g}$ in the diffraction pattern decreases as λ decreases. Notice that the key word here is "appear." If you look back at 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.

It is very important for TEM 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

Table 12.2. Particular Values of λ and λ^{-1} as a Function of Accelerating Voltage

E	λ (Å)	Radius, λ^{-1} (Å^{-1})	$(v/c)^2$
100 keV	0.03701	27.02	0.3005
120 keV	0.03349	29.86	0.3441
200 keV	0.02508	39.87	0.4834
300 keV	0.01969	50.80	0.6030
400 keV	0.01644	60.83	0.6853
1 MeV	0.008719	114.7	0.8856

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 spread-sheet. Use the values from Chapter l: $m_0 = 9.109 \times 10^{-31}$ kg, $c = 2.998 \times 10^{8}$ ms^{-1}, $h = 6.626 \times 10^{-34}$ Nms, and 1eV = 1.602×10^{-19} Nm.

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 Å^{-1} or nm^{-1}. Since 10 Å = 1 nm, 1 Å^{-1} = 10 nm^{-1}.

REFERENCES

General References

Cullity, B.D. (1978) *Elements of X-ray Diffraction,* Addison-Wesley, Reading, Massachusetts.

James, R.W. (1965) *The Optical Principles of the Diffraction of X-rays, The Crystalline State,* **II** (Ed. W.L. Bragg), Cornell University Press, Ithaca, New York (first published in 1948).

Schwartz, L.H. and Cohen, J.B. (1977) *Diffraction from Materials,* Academic Press, New York.

Specific References

Cruickshank, D.W.J., Juretschke, H.J., and Kato, N., Eds. (1992) *P.P. Ewald and His Dynamical Theory of X-ray Diffraction,* Oxford University Press, New York.

Ewald, P.P. (1962) *Fifty Years of X-ray Diffraction,* N.V.A. Oosthoek's Uitgeversmaatschappij, Utrecht, the Netherlands.

Diffracted Beams 13

CHAPTER PREVIEW

In Chapter 11 we discussed why diffraction occurs; in this chapter we give a more detailed mathematical treatment. It may be more detail than you need at this stage. Diffraction is one of those phenomena which 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 dynamical case. We will not do this but we will introduce the words and assumptions elsewhere.

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 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 *extinction distances*. 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. 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.

Diffracted Beams 13

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 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 27. If instead of cutting the specimen in 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-diffraction patterns (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 15 Å or less) than the lateral dimensions of the electron beam (typically >1 μm in the TEM imaging mode). As we will see in Chapters 23–26 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 III.

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 ($I \propto |\phi|^2$).

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. Terminology and notation are given in Table 13.1.

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 neces-

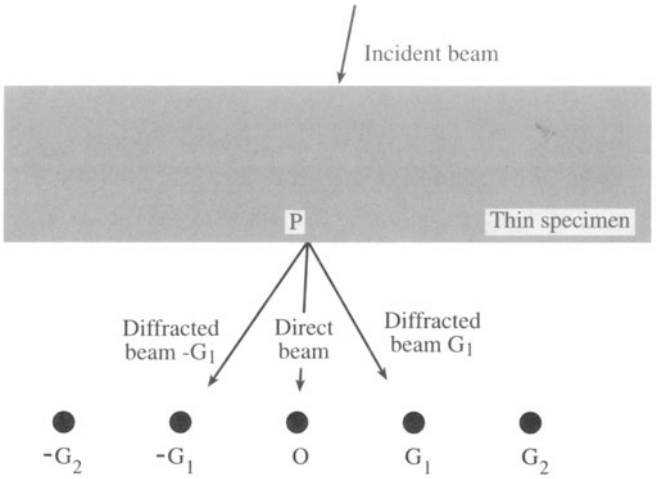


Figure 13.1. Defining the point P. The incident beam is scattered inside the thin specimen. We want to know the intensities of the direct beam (O) and the diffracted (G_i) beams for each point P at the bottom surface of the specimen (the exit surface).

sary assumption, but it simplifies calculations and again aids intuitive understanding. You will recognize many similarities to visible-light microscopy but be wary, there are also many differences.

A note on terminology. *In Figure 13.1 we have labeled both the diffracted beams and the spot in the diffraction pattern, 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."*

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

$$A_{\text{cell}} = \frac{e^{2\pi i\mathbf{k}\cdot\mathbf{r}}}{r}\sum_i f_i(\theta)e^{2\pi i\mathbf{K}\cdot\mathbf{r}_i} \qquad [13.1]$$

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 added 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 from Chapter 3. 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 and we will use the positive convention to be consistent with most materials science texts.

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 ϕ_g and ϕ_0.
ϕ_g	The amplitude of the *diffracted* beam for reflection G. The intensity is $\lvert\phi_g\rvert^2$.
ϕ_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. ϕ_0 is a special value of ϕ_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.
ξ_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 vacuum
k	the electron wave vector in the specimen

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. 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).

Therefore, equation 13.1 can be rewritten as

$$A_{\text{cell}} = \frac{e^{2\pi i\mathbf{k}\cdot\mathbf{r}}}{r}F(\theta) \qquad [13.2]$$

To find the intensity at some point P, we then sum over all the unit cells in the specimen. For simplicity here,

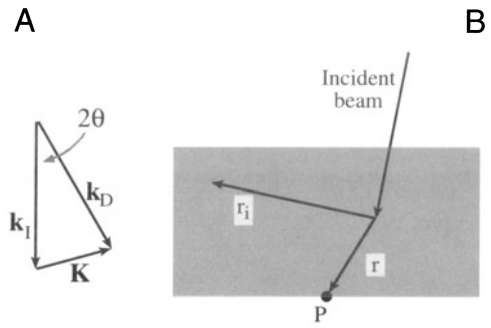


Figure 13.2. (A) A reminder that $\mathbf{K} = \mathbf{k}_D - \mathbf{k}_I$. The vector $\mathbf{k}_D$ represents the propagation vector for *any* wave. It does not have to be a diffracted beam but it will only give a spot in the diffraction pattern when it does correspond to a diffracted beam. (B) shows the relation between the radius of the spherical wavefront, r, the position vector of the ith atom, $\mathbf{r}_i$, and the point where the intensity is calculated, P.

we will not solve this problem mathematically but simply quote the result and discuss its meaning. 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 amplitude in a *diffracted beam* (in the direction identified by θ) is denoted as ϕ_g and is given by

$$\phi_g = \frac{\pi a i}{\zeta_g} \sum_n e^{-2\pi i K \cdot r_n} e^{2\pi i K_D \cdot r} \qquad [13.3]$$

Here $\mathbf{r}_n$ denotes the position of each unit cell. In this analysis, the quantities $f(\theta)$ and $F(\theta)$ both have dimensions of length. We'll now explain what the length ξ_g means in equation 13.3; it is a length because ϕ_g, the scattering amplitude, is dimensionless (ξ (xi) 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, $|\phi_0|^2$, remains unchanged. This assumption is usually not justified, especially when the specimen has a finite thickness! If $|\phi_g|^2$ is not zero then $|\phi_0|^2$ cannot still be 1.

13.4. THE CHARACTERISTIC LENGTH ξ_g

At this stage in our analysis it is best to think of the quantity ξ_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 ξ_g can be expressed as

$$\xi_g = \frac{\pi V_c \cos\theta_B}{\lambda F_g} \qquad [13.4]$$

Table 13.2. Examples of Extinction Distances (in nm)*

Material $hk\ell$ =	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 conditions at 100 kV.

where F_g is the $F(\theta)$ for reflection **g** (i.e., F_g is a special value of $F(\theta)$ when θ is the Bragg angle θ_B). The volume of a unit cell, V_c, is simply a/n.

ξ_g is the characteristic length for the diffraction vector **g**. We call it the *extinction distance*. 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 which will become obvious. Note that ξ_g is a scalar quantity.

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).

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.

ξ_g depends on the lattice parameters (through V_c), the atomic number (through F_g), and the kV used (through λ).

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

$$\psi^T = \phi_0 e^{2\pi i \chi_O \cdot \mathbf{r}} + \phi_{g_1} e^{2\pi i \chi_{G_1} \cdot \mathbf{r}} + \phi_{g_2} e^{2\pi i \chi_{G_2} \cdot \mathbf{r}} + \cdots \quad [13.5]$$

with wave vectors 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*

$$d\phi_g = \left\{ \frac{\pi i}{\xi_g} \phi_0 e^{2\pi i(\chi_O - \chi_D)\cdot \mathbf{r}} + \frac{\pi i}{\xi_0} \phi_g \right\} dz \quad [13.6]$$

and

$$d\phi_0 = \left\{ \frac{\pi i}{\xi_0} \phi_0 + \frac{\pi i}{\xi_g} \phi_g e^{2\pi i(\chi_D - \chi_O)\cdot \mathbf{r}} \right\} dz \quad [13.7]$$

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 $\mathbf{k}_D - \mathbf{k}_O$ (= **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 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 change of 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.

$$\frac{d\phi_g}{dz} = \frac{\pi i}{\xi_g} \phi_0 e^{-2\pi i s z} + \frac{\pi i}{\xi_0} \phi_g \quad [13.8]$$

and

$$\frac{d\phi_0}{dz} = \frac{\pi i}{\xi_0} \phi_0 + \frac{\pi i}{\xi_g} \phi_g e^{2\pi i s z} \quad [13.9]$$

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

$$e^{-2\pi i \mathbf{s}\cdot \mathbf{r}} = e^{-2\pi i s z} \quad [13.10]$$

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

$$|\mathbf{s}_g| = s_z \quad [13.11]$$

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 $\mathbf{k}_0$). 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. 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)

$$\phi_{0(\mathrm{sub})} = \phi_0 e^{\frac{-\pi i z}{\xi_0}} \qquad [13.12]$$

and

$$\phi_{g(\mathrm{sub})} = \phi_g e^{2\pi i s z - \frac{\pi i z}{\xi_0}} \qquad [13.13]$$

Then equations 13.8 and 13.9 become

$$\frac{d\phi_{g(\mathrm{sub})}}{dz} = \frac{\pi i}{\xi_g}\phi_{0(\mathrm{sub})} + 2\pi i s \phi_{g(\mathrm{sub})} \qquad [13.14]$$

and

$$\frac{d\phi_{0(\mathrm{sub})}}{dz} = \frac{\pi i}{\xi_g}\phi_{g(\mathrm{sub})} \qquad [13.15]$$

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 13.17 into 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 13.17 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). Use 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.

$$w = s\xi \qquad [13.26]$$

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

$$\frac{C_g^{(2)}}{C_0^{(2)}} = 2\xi_g \gamma^{(2)} = w + \sqrt{w^2 + 1} \qquad [13.28]$$

(the superscripts on $C_g^{(1)}$, 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 ϕ_0 and ϕ_g so that they satisfy the relations

$$C_0^{(1)^2} + C_g^{(1)^2} = 1 = C_0^{(2)^2} + C_g^{(2)^2} \qquad [13.30]$$

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

$$\begin{aligned} C_0^{(1)} &= \cos\frac{\beta}{2} & C_g^{(1)} &= -\sin\frac{\beta}{2} \\ C_0^{(2)} &= \sin\frac{\beta}{2} & C_g^{(2)} &= \cos\frac{\beta}{2} \end{aligned} \qquad [13.31]$$

Now you can understand why we introduced β in equation 13.29. The two independent solutions to the reformulated Howie–Whelan equation for ϕ_0 (13.16) are then $\phi_0 = C_0^{(1)}\exp(2\pi i\gamma^{(1)}z)$ and $\phi_0 = C_0^{(2)}\exp(2\pi i\gamma^{(2)}z)$ and *each value* has a corresponding value for ϕ_g.

Most importantly, 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 ($|\psi^T|^2 = 1$). This is the reason we use a normalized intensity, in equation 13.30.

We can already see that the ratio of the amplitudes of the diffracted and direct beams, C_g to C_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 concerned about *the* Bragg condition because we have chosen a two-beam situation.

In the two-beam approximation, equation 13.5 is expressed in terms of ϕ_0 and ϕ_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 $C_0^{(1)}$, $C_0^{(2)}$, 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 **r** and have their own values of **k** which we identify as $\mathbf{k}^{(j)}$.

> Each value of γ gives a different value of **k** which we call $\mathbf{k}^{(j)}$.

Thus we can write expressions for $b^{(1)}$ and $b^{(2)}$

$$b^{(1)}\left(\mathbf{k}^{(1)},\mathbf{r}\right) = C_0^{(1)}e^{2\pi i\,\mathbf{k}^{(1)}\cdot\mathbf{r}} + C_g^{(1)}e^{2\pi i\,(\mathbf{k}^{(1)}+\mathbf{g})\cdot\mathbf{r}} \quad [13.32]$$

and

$$b^{(2)}\left(\mathbf{k}^{(2)},\mathbf{r}\right) = C_0^{(2)}e^{2\pi i\,\mathbf{k}^{(2)}\cdot\mathbf{r}} + C_g^{(2)}e^{2\pi i\,(\mathbf{k}^{(2)}+\mathbf{g})\cdot\mathbf{r}} \quad [13.33]$$

Remember: each of these Bloch-wave functions could be a wave in the crystal—each one depends on only one of the **k** values. In general, the total wave function will be a combination of these two waves. We'll return to the important relationship between **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 ϕ_0 and ϕ_g are. The total wave vector, ψ^T, is a combination of the two (Bloch) waves, $b^{(1)}$ and $b^{(2)}$

$$\psi^T = \mathcal{A}^{(1)}b^{(1)} + \mathcal{A}^{(2)}b^{(2)} \quad [13.34]$$

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

$$\begin{aligned}\psi^T = \mathcal{A}^{(1)}&\left\{\cos\frac{\beta}{2}e^{2\pi i\,\mathbf{k}^{(1)}\cdot\mathbf{r}} - \sin\frac{\beta}{2}e^{2\pi i\,(\mathbf{k}^{(1)}+\mathbf{g})\cdot\mathbf{r}}\right\}\\ &+\mathcal{A}^{(2)}\left\{\sin\frac{\beta}{2}e^{2\pi i\,\mathbf{k}^{(2)}\cdot\mathbf{r}} + \cos\frac{\beta}{2}e^{2\pi i\,(\mathbf{k}^{(2)}+\mathbf{g})\cdot\mathbf{r}}\right\}\end{aligned} \quad [13.35]$$

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

$$\begin{aligned}\psi^T = &\left\{\mathcal{A}^{(2)}\sin\frac{\beta}{2}e^{2\pi i\,\mathbf{k}^{(2)}\cdot\mathbf{r}} + \mathcal{A}^{(1)}\cos\frac{\beta}{2}e^{2\pi i\,\mathbf{k}^{(1)}\cdot\mathbf{r}}\right\}\\ &+\left\{\mathcal{A}^{(2)}\cos\frac{\beta}{2}e^{2\pi i\,\mathbf{k}^{(2)}\cdot\mathbf{r}} - \mathcal{A}^{(1)}\sin\frac{\beta}{2}e^{2\pi i\,\mathbf{k}^{(1)}\cdot\mathbf{r}}\right\}e^{2\pi i\,\mathbf{g}\cdot\mathbf{r}}\end{aligned} \quad [13.36]$$

Only the second term depends on **g**, so this must be the ϕ_g term. We know that at the top of the specimen (**r** = 0), ϕ_0 is unity and ϕ_g is zero (independent of γ)—the amplitude of the diffracted beam is zero before it's diffracted! It follows directly that

$$\mathcal{A}^{(1)} = \cos\frac{\beta}{2} \quad [13.37]$$

and

$$\mathcal{A}^{(2)} = \sin\frac{\beta}{2} \quad [13.38]$$

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, as we'll see, just involves tilting the specimen.

Now, finally, we can write down the general expressions for ϕ_0 and ϕ_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

$$\psi^T = \phi_0 e^{2\pi i\,\mathbf{k}\cdot\mathbf{r}} + \phi_g e^{2\pi i\,(\mathbf{k}+\mathbf{g})\cdot\mathbf{r}} \quad [13.39]$$

(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 ϕ_0 and ϕ_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

$$\phi_g = \sin\frac{\beta}{2}\cos\frac{\beta}{2}\left\{e^{2\pi i(\mathbf{k}^{(2)}-\mathbf{K})\cdot\mathbf{r}} - e^{2\pi i(\mathbf{k}^{(1)}-\mathbf{K})\cdot\mathbf{r}}\right\} \quad [13.40]$$

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.

$$\left(\mathbf{k}^{(2)}-\mathbf{K}\right)_z = \gamma^{(2)} \text{ and } \left(\mathbf{k}^{(1)}-\mathbf{K}\right)_z = \gamma^{(1)} \quad [13.41]$$

What we are interested in is the magnitude of $\gamma^{(1)}$ and $\gamma^{(2)}$. We have also shown directly that ϕ_0 in equation 13.39 is a mixture of terms containing $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$. This is a key result. We can now manipulate equation 13.40 using equation 13.41 and the expression $e^{i\theta} = \cos\theta + i\sin\theta$ to give

$$\phi_0 = \{\cos(\pi z\Delta k) - i\cos\beta\cdot\sin(\pi z\Delta k)\}\, e^{\pi i s z} \quad [13.42]$$

and

$$\phi_g = +i\sin\beta\cdot\sin(\pi z\Delta k)\cdot e^{\pi i s z} \quad [13.43]$$

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 ϕ_0 and ϕ_g, but it will make it easier for you to check that these expressions satisfy, for example, equation 13.16.

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.

$$\Delta k = \frac{\sqrt{w^2+1}}{\xi_g} \quad [13.44]$$

The intensity in the diffracted beam, $|\phi_g|^2 = \phi_g\phi_g^\star$, is obtained from equation 13.43

$$I_g = |\phi_g|^2 = \sin^2\beta\sin^2(\pi t\Delta k) \quad [13.45]$$

$$I_g = |\phi_g|^2 = \frac{1}{w^2+1}\sin^2\left(\frac{\pi t\sqrt{w^2+1}}{\xi_g}\right) \quad [13.46]$$

We can make this equation look more familiar by defining an effective excitation error, s_{eff}, where

$$s_{eff} = \sqrt{s^2 + \frac{1}{\xi_g^{\,2}}} = \frac{\sqrt{w^2+1}}{\xi_g} \quad [13.47]$$

Now the equation becomes

$$|\phi_g|^2 = \left(\frac{\pi t}{\xi_g}\right)^2 \frac{\sin^2(\pi t s_{eff})}{(\pi t s_{eff})^2} \quad [13.48]$$

This is the **REALLY** important equation for us.

It 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.

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

$$I_0 = 1 - I_g \quad [13.49]$$

Remember when testing this formula that $I = \phi\phi^\star$ ($\phi^\star$ 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 $\sim 2\theta_B$.

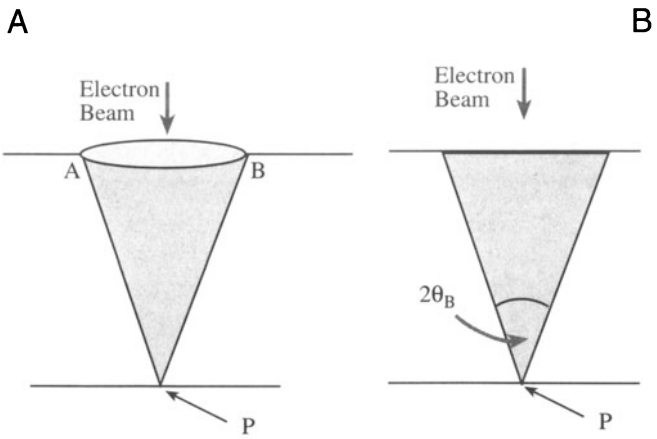


Figure 13.3. (A) The intensity of the beams at point P at the bottom of the specimen is influenced by all the scattering within a cone of material. The solid angle of the cone is determined by the diameter of the Fresnel zones which, in turn, are principally determined by λ. The cross section (B) is the more typical view of the cone.

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 (1987). Why is it 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).

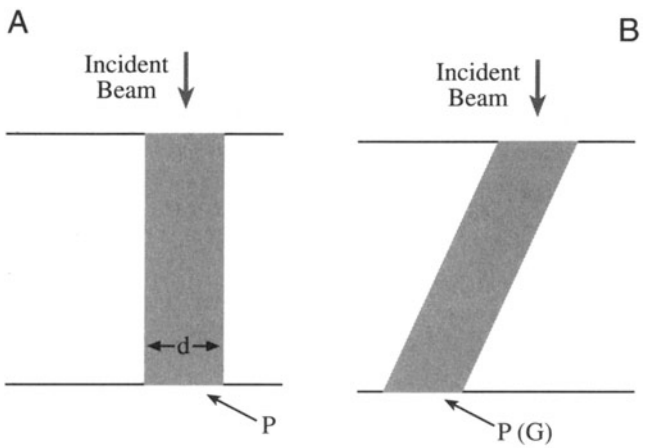


Figure 13.4. The column approximation for (A) the direct beam and (B) a diffracted beam. A column replaces the cone. The diameter of the column (*d*) should be the *average* diameter of the cone it replaces (AB/2 in Figure 13.3). This value will depend on the thickness of the sample. In practice it is usually taken to be ~2 nm.

Let's consider some actual numbers: at 100 kV, $\lambda = 0.0037$ nm, $\theta_B \sim 0.01$ radians or ~0.5°. So if the thickness (t) of the specimen is 100 nm, then AB is ~ 2 nm. If we increase 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 Figures 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 of very fine detail, especially when these features can vary their positions in the foil. The column approximation often hides itself very well, but it is actually used in many calculations of images. The more correct noncolumn treatment was introduced by Takagi (1962); the analysis by Howie and Basinski (1968) is what we use in computer programs.

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 backscattered electron diffraction (BSED) 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 ξ_g. If the material is noncentrosymmetric, 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 nonsystematic reflections con-

tribute to the image. In these cases, you will need to use a computer program to predict or interpret the contrast.

- 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 ξ_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).
- Two conventions are commonly used to describe the exponential dependence on **k** and **r**

$$e^{2\pi i\mathbf{k}\cdot\mathbf{r}} \quad \text{or} \quad e^{-2\pi i\mathbf{k}\cdot\mathbf{r}} \qquad [13.50]$$

These conventions have been discussed by Spence (1988). In our analysis we have chosen to use $e^{2\pi i\mathbf{k}\cdot\mathbf{r}}$, which Spence has termed the "quantum-mechanical" convention. (Note that Spence uses the alternative "crystallographic" convention except when he discusses Bloch waves.) In the quantum-mechanical convention, which is also used by Spence (1988), the time-dependent Schrödinger equation is written as

$$\frac{h^2}{8\pi^2 m}\nabla^2\psi = -i\frac{h}{2\pi}\frac{d\psi}{dt} \qquad [13.51]$$

with the full solution being

$$\psi(\mathbf{r},t) = Ae^{+i(\mathbf{k}\cdot\mathbf{r}-\omega t)} \qquad [13.52]$$

- The concept of a refractive-index effect for electron waves is directly analogous to that for 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.
- 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 13.46 we can write

$$|\phi_g|^2 = \sin^2\left(\frac{\pi t}{\xi_g}\right) \qquad [13.53]$$

and similarly

$$|\phi_0|^2 = 1 - \sin^2\left(\frac{\pi t}{\xi_g}\right) \qquad [13.54]$$

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_g is zero at $t = 0$ and again at $t = \xi_g$ (or, in general, at $t = n\xi_g$, where n is an integer). This is the reason we call ξ_g the extinction distance. This situation corresponds to two coupled simple-harmonic oscillators with energy (i.e., intensity, I_0 and I_g) being continuously transferred from one to the other and back again. Notice that I_g can only increase to unity when $\mathbf{s} = 0$.

CHAPTER SUMMARY

In this chapter we have derived equations and introduced terminology which will form the basis for our discussion of diffraction-contrast images. It is not necessary to be able to reproduce the mathematical deriva-

tions but equations 13.47 and 13.48 are crucial and must be understood. The analysis was quickly limited to two beams, the direct beam and one Bragg-diffracted beam. In deriving the Howie–Whelan equations it is necessary to consider both forward scattering and Bragg diffraction. We introduced a new parameter, the critical length ξ_g, and explained why this parameter is called the extinction distance. This length was defined in equation 13.4, which shows that ξ_g depends on the *material*, the *reflection,* and the *wavelength of the electrons*. Two particular points to remember are:

- If the voltage increases then λ decreases and ξ_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

General References

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, New York.

Specific References

Darwin, C.G. (1914) *Phil. Mag.* **27**, 315 and 675.
Hecht, E. (1987) *Optics,* 2nd edition, Addison-Wesley, Reading, Massachusetts.
Howie, A. and Whelan, M.J. (1961) *Proc. Roy. Soc.* **A263**, 217.
Howie, A. and Basinski, Z.S. (1968) *Phil. Mag.* **17**, 1039.
Spence, J.C.H. (1988) *Experimental High-Resolution Electron Microscopy,* p. 155, Oxford University Press, New York.
Takagi, S. (1962) Acta Cryst. **15**, 1311.

Bloch Waves

14

CHAPTER PREVIEW

This topic is rather mathematical, with sequences of differential equations. The discussion of Bloch waves given here follows the treatment of Hirsch *et al.* (1977) 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 origin of the *extinction distance,* ξ_g, so again you must persevere. 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 ξ_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 only applies 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 we 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 lat-

tice, 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 23.7, the physical significance will be more obvious.

Bloch Waves

14

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

$$\left[-\frac{h^2}{8\pi^2 m}\nabla^2 + \mathcal{U}(\mathbf{r})\right]\Psi(\mathbf{r}) = \mathcal{E}\Psi(\mathbf{r}) \qquad [14.1]$$

The first term (in ∇^2) represents the kinetic energy, and the second term the potential energy; $\mathcal{E}$ 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, $-E$ (usually 100 kV–1 MV), is negative for a positive charge leaving the gun. This quantity, $-E$, is really the "electric field potential."
- The initial energy given to the electron is a positive number, $\mathcal{E}$ (in eV); it is just the charge times the accelerating voltage. We can write this as eE, where both e and E 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

$$\nabla^2\psi(\mathbf{r}) + \frac{8\pi^2 me}{h^2}\left[E + V(\mathbf{r})\right]\psi(\mathbf{r}) = 0 \qquad [14.2]$$

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 nature and environment at point $\mathbf{r}$ is identical to that at point $\mathbf{r} + \mathbf{R}$. We can picture this inner potential as shown in Figure 14.1 for the one-dimensional case. 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 elec-

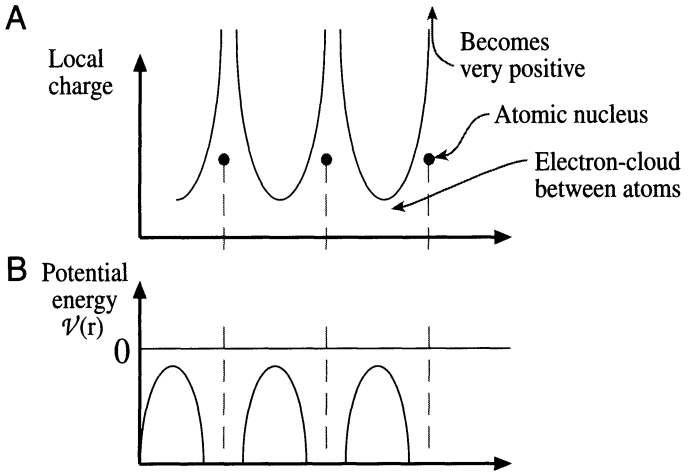


Figure 14.1. (A) The local charge sensed by the beam electron as it passes through a metal, represented as a row of "ion" cores (black circles) in a sea of electrons. The local charge is very large and positive in the vicinity of the ion and becomes small, but not zero, between the ions. The difference 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.

tron which we "shoot" through the crystal will always 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} (the total wave function) 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 the 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 noncentrosymmetric 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

$$V(\mathbf{r}) = \sum_{\mathbf{g}} V_{\mathbf{g}} e^{2\pi i \mathbf{g}\cdot\mathbf{r}} \qquad [14.5]$$

Here, $V_{\mathbf{g}}$ is 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

$$V_{\mathbf{g}} = \frac{h^2}{2me} U_{\mathbf{g}} \qquad [14.6]$$

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

$$V(\mathbf{r}) = \frac{h^2}{2me} \sum_{\mathbf{g}} U_{\mathbf{g}} e^{2\pi i \mathbf{g}\cdot\mathbf{r}} \qquad [14.7]$$

Now $V(\mathbf{r})$ has been expanded as a Fourier sum; all the conditions on $V(\mathbf{r})$ also apply to each $U_{\mathbf{g}}$, so that

$$U_{\mathbf{g}} = U_{\mathbf{g}}^* = U_{-\mathbf{g}} \qquad [14.8]$$

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 energies which are 5 orders of magnitude greater than the band gap of Si. Notice the value of the inner potential energy, which 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 $\mathcal{V}$ is the average background potential energy and is directly related to the characteristic length ξ_0 which we introduced in Chapter 13. More values of $\mathcal{V}$ are given in Table 14.2. The 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 = 293K)	0.025
Band gap of Si	1.1
Inner potential energy for Si	~11
Energy of electrons in TEM	≥ 100,000

Table 14.2. Comparison of Inner Potential Energies for Different Elements (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

$$\psi^{(j)}(\mathbf{r}) = b\left(\mathbf{k}^{(j)},\mathbf{r}\right) \qquad [14.9]$$

The reason for the "j" is that each Bloch wave has a single value of $\mathbf{k}$ (each one 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

$$\begin{aligned} b^{(j)}(\mathbf{r}) &= b\left(\mathbf{k}^{(j)},\mathbf{r}\right) = \mu\left(\mathbf{k}^{(j)},\mathbf{r}\right) e^{2\pi i\mathbf{k}^{(j)}\cdot\mathbf{r}} \\ &= \mu^{(j)}(\mathbf{r})\, e^{2\pi i\mathbf{k}^{(j)}\cdot\mathbf{r}} \end{aligned} \qquad [14.10]$$

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 $\mathbf{r}$

$$\mu^{(j)}(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)}\left(\mathbf{k}^{(j)}\right) e^{2\pi i\mathbf{g}\cdot\mathbf{r}} \qquad [14.11]$$

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 $\mathbf{k}^{(j)}$ but not on $\mathbf{r}$. Combining these definitions gives

$$b^{(j)}(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i\left(\mathbf{k}^{(j)}+\mathbf{g}\right)\cdot\mathbf{r}} \qquad [14.12]$$

Using our notation, the superscript on C indicates that $C^{(j)}$ depends on $\mathbf{k}^{(j)}$. We can now write the expanded expression for $b^{(j)}(\mathbf{r})$, which is a solution to the Schrödinger equation

$$b^{(j)}(\mathbf{r}) = C_{\mathbf{0}}^{(j)} e^{2\pi i\mathbf{k}^{(j)}\cdot\mathbf{r}} + C_{\mathbf{g}}^{(j)} e^{2\pi i\left(\mathbf{k}^{(j)}+\mathbf{g}\right)\cdot\mathbf{r}} + \cdots \qquad [14.13]$$

The first term in this series is $C_{\mathbf{0}}$; the subscript is zero 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 eV to 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 $1/\lambda$

> 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

$$\psi^{\text{tot}} = \sum_{j=1}^{n} \mathcal{A}^{(j)}\psi^{(j)} = \sum_{j=1}^{n} \mathcal{A}^{(j)} b\left(\mathbf{k}^{(j)},\mathbf{r}\right) \qquad [14.14]$$

where $\mathcal{A}^{(j)}$ will be determined by the specimen type, 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.

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 $\mathcal{K}$ which is defined by the equation

$$\mathcal{K}^2 = \frac{2me}{h^2} + U_0 = \chi^2 + U_0 \quad [14.18]$$

With this definition we have removed the U_0 term from the sum over all **g**, so that equation 14.15 is now

$$\frac{1}{4\pi^2}\nabla^2\psi(\mathbf{r}) + \mathcal{K}^2\psi(\mathbf{r}) + \sum_{\mathbf{g}\neq 0} U_{\mathbf{g}}\, e^{2\pi i \mathbf{g}\cdot\mathbf{r}}\psi(\mathbf{r}) = 0 \quad [14.19]$$

The reason for doing this is that we are going to be concerned with different diffraction vectors, **g**. The U_0 term does not depend on **g**. We call U_0 the (scaled) mean inner potential of the crystal; this potential is a "background" or continuum property of the crystal; it does not directly depend on the crystal structure. (You may recognize this as the refractive index idea reappearing.)

When $V(\mathbf{r})$ is 0, then U_0 is 0 so that $\mathcal{K}^2$ 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 value; 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, **p** the momentum, and **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 $\mathcal{K}$ is now clear: it is the wave vector of the electron *inside* the specimen, i.e., after correcting for the refractive index effect. Since U_0 is a positive number, $\mathcal{K}$ is 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).

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.

Equations 14.18 and 14.20 are *dispersion relations*. Such equations relate the magnitude of the wave vector, $\mathcal{K}$ or χ, to the energy of the electron. 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.

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 $b^{(j)}(\mathbf{r})$.

$$\nabla^2\psi(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)}\nabla^2\left(e^{2\pi i\left(\mathbf{k}^{(j)}+\mathbf{g}\right)\cdot\mathbf{r}}\right) \quad [14.22]$$

$C_{\mathbf{g}}^{(j)}$ does not depend on r, thus

$$\nabla^2\psi(\mathbf{r}) = -(2\pi)^2\sum_{\mathbf{g}}\left|\mathbf{k}^{(j)}+\mathbf{g}\right|^2 C_{\mathbf{g}}^{(j)}\, e^{2\pi i(\mathbf{k}^{(j)}+\mathbf{g})\cdot\mathbf{r}} \quad [14.23]$$

Now we insert this expression in equation 14.19

$$\begin{aligned}&\frac{1}{4\pi^2}\left(-4\pi^2\sum_{\mathbf{g}}\left|\mathbf{k}^{(j)}+\mathbf{g}\right|^2 C_{\mathbf{g}}^{(j)}\, e^{2\pi i(\mathbf{k}^{(j)}+\mathbf{g})\cdot\mathbf{r}}\right)\\ &+\mathcal{K}^2\sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i(\mathbf{k}^{(j)}+\mathbf{g})\cdot\mathbf{r}}\\ &+\sum_{\mathbf{h}\neq 0} U_{\mathbf{h}}\, e^{2\pi i\mathbf{h}\cdot\mathbf{r}}\sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i(\mathbf{k}^{(j)}+\mathbf{g})\cdot\mathbf{r}} = 0\end{aligned} \quad [14.24]$$

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**

$$\begin{aligned}&\sum_{\mathbf{g}}\sum_{\mathbf{h}\neq 0} U_{\mathbf{h}}\, C_{\mathbf{g}}^{(j)}\cdot e^{2\pi i(\mathbf{k}^{(j)}+\mathbf{g}+\mathbf{h})\cdot\mathbf{r}}\\ &=\sum_{\mathbf{g}-\mathbf{h}}\sum_{\mathbf{h}\neq 0} U_{\mathbf{h}}\, C_{\mathbf{g}-\mathbf{h}}^{(j)}\cdot e^{2\pi i(\mathbf{k}^{(j)}+\mathbf{g})\cdot\mathbf{r}}\end{aligned} \quad [14.25]$$

Now the sum over all **g**–**h** vectors is the same as the sum over all **g** vectors, so we can replace **g**–**h** with **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

$$\sum_{\mathbf{g}} \left(\left\{ -\left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 + \mathcal{K}^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}-\mathbf{h}}^{(j)} \right) e^{2\pi i(\mathbf{k}^{(j)} + \mathbf{g}) \cdot \mathbf{r}} = 0 \qquad [14.26]$$

We can obtain a useful relation by noting that the coefficients of each term in $\exp(2\pi i \mathbf{g} \cdot \mathbf{r})$ must separately be equal to zero. The only way that equation 14.26 can be true is if the term inside the brackets is always zero. The result is a series of equations (one for each value of **g**)

$$\left\{ -\left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 + \mathcal{K}^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}-\mathbf{h}}^{(j)} = 0 \qquad [14.27]$$

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 **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 **h** = **g** in the sum

$$\left\{ \mathcal{K}^2 - \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq \mathbf{g}} U_{\mathbf{g}-\mathbf{h}} C_{\mathbf{h}}^{(j)} = 0 \qquad [14.28]$$

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 solid-state 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., consider the case where the only values of $C_{\mathbf{g}}$ which are *nonzero* are $C_{\mathbf{0}}^{(j)}$ and $C_{\mathbf{p}}^{(j)}$ but $U_{\mathbf{p}}$ and $U_{-\mathbf{p}}$ 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 **g** = **0** in equation 14.27 gives

$$\left(\mathcal{K}^2 - \left|\mathbf{k}^{(j)}\right|^2\right) C_{\mathbf{0}}^{(j)} + U_{-\mathbf{p}} C_{\mathbf{p}}^{(j)} = 0 \qquad [14.29]$$

In deriving this and the following equation, we consider all the possible values of **h** which would give us $C_{\mathbf{0}}^{(j)}$ or $C_{\mathbf{p}}^{(j)}$.

Next, if we let **g** = **p** in equation 14.28 and reverse the order of terms to emphasize that we have two equations in $C_{\mathbf{0}}$ and $C_{\mathbf{p}}$

$$U_{\mathbf{p}} C_{\mathbf{0}}^{(j)} + \left(\mathcal{K}^2 - \left|\mathbf{k}^{(j)} + \mathbf{p}\right|^2\right) C_{\mathbf{p}}^{(j)} = 0 \qquad [14.30]$$

There are no other possible equations, so to solve these two equations we set the determinant of the coefficients equal to zero

$$\begin{vmatrix} \mathcal{K}^2 - \left|\mathbf{k}^{(j)}\right|^2 & U_{-\mathbf{p}} \\ U_{\mathbf{p}} & \mathcal{K}^2 - \left|\mathbf{k}^{(j)} + \mathbf{p}\right|^2 \end{vmatrix} = \left(\mathcal{K}^2 - \left|\mathbf{k}^{(j)}\right|^2\right)\left(\mathcal{K}^2 - \left|\mathbf{k}^{(j)} + \mathbf{p}\right|^2\right) - U_{\mathbf{p}} U_{-\mathbf{p}} = 0 \qquad [14.31]$$

The inner potential of the crystal is usually ≤20 V while the energy of the electrons is ≥100,000 eV. Because $|\mathbf{k}^{(j)} + \mathbf{p}|$ and $|\mathbf{k}^{(j)}|$ are both very close to $\mathcal{K}$, it's the difference that is important. Since P could be any diffracted beam, we can rename it G to make it look more familiar!

$$\begin{vmatrix} \mathcal{K}^2 - \left|\mathbf{k}^{(j)}\right|^2 & U_{-\mathbf{g}} \\ U_{\mathbf{g}} & \mathcal{K}^2 - \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2 \end{vmatrix} = \left(\mathcal{K}^2 - \left|\mathbf{k}^{(j)}\right|^2\right)\left(\mathcal{K}^2 - \left|\mathbf{k}^{(j)} + \mathbf{g}\right|^2\right) - U_{\mathbf{g}} U_{-\mathbf{g}} = 0 \qquad [14.32]$$

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 $\mathcal{K}$ are all similar in magnitude. Then equation 14.32 becomes

$$\left(|\mathbf{k}^{(j)}| - \mathcal{K}\right)\left(|\mathbf{k}^{(j)} + \mathbf{g}| - \mathcal{K}\right) = \frac{U_{\mathbf{g}} U_{-\mathbf{g}}}{4\mathcal{K}^2} = \frac{|U_{\mathbf{g}}|^2}{4\mathcal{K}^2} \qquad [14.34]$$

It is important not to confuse $\mathbf{k}^{(j)}$ with $\mathbf{k}_I$ or $\mathbf{k}_D$, and you must remember that $|\mathbf{K}|$ (= $|\mathbf{k}_I + \mathbf{g}|$) is not equal to $\mathcal{K}$. 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. The simpler relations given in equations 14.18 and 14.20 each define a sphere and in these equations the vectors $\mathcal{K}$ and χ can point in any direction.

From equation 14.29 (renaming **p** as **g**), we have

$$\frac{C_{\mathbf{g}}^{(j)}}{C_{\mathbf{0}}^{(j)}} = \frac{|\mathbf{k}^{(j)}|^2 - \mathcal{K}^2}{U_{-\mathbf{g}}} \qquad [14.35]$$

which we can rewrite as

$$\frac{C_{\mathbf{g}}^{(j)}}{C_{\mathbf{0}}^{(j)}} = \frac{\left(|\mathbf{k}^{(j)}| - \mathcal{K}\right)\left(|\mathbf{k}^{(j)}| + \mathcal{K}\right)}{U_{-\mathbf{g}}} \approx \frac{2\mathcal{K}\left(|k^{(j)}| - \mathcal{K}\right)}{U_{-\mathbf{g}}} \qquad [14.36]$$

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

$$\mathbf{A}^{(j)}\left\{C_{\mathbf{g}}^{(j)}\right\} = 0 \qquad [14.37]$$

where $\{C_{\mathbf{g}}^{(j)}\}$ now denotes a column vector with elements $C_{\mathbf{g}}^{(j)}$ $\mathbf{A}^{(j)}$ is a matrix defined by

$$a_{gg} = K^2 - |\mathbf{k}^{(j)} + \mathbf{g}|^2 \qquad [14.38]$$

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 **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 (1975) and is adapted here for 5 beams, comprising $\bar{\mathbf{g}}$, 0, **g**, 2**g**, and 3**g** beams. The 5 × 5 matrix can be written out (using g and h rather than **g** and **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 $\{C_{\mathbf{g}}^{(j)}\}$ appears as the eigenvectors and the wave vectors $\mathbf{k}^{(j)}$ appear as the eigenvalues. He expresses this equation as

$$\mathcal{M}\left\{C_{\mathbf{g}}^{(j)}\right\} = \gamma^{(j)}\left\{C_{\mathbf{g}}^{(j)}\right\} \qquad [14.42]$$

where the matrix $\mathcal{M}$ 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 down equation 13.30.

As you can see, the math is beginning to become tricky. In the next chapter, we will derive explicit expressions for ξ_0 and ξ_g in the two-beam case, namely,

$$\xi_0 = \frac{2\,\mathcal{K}\cos\theta_B}{U_0} \qquad [14.43]$$

and

$$\xi_g = \frac{2\,\mathcal{K}\cos\theta_B}{U_g} = \frac{1}{|\Delta\mathbf{k}|} \qquad [14.44]$$

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

$$\psi(\mathbf{r}) = \mathcal{A}^{(1)}b^{(1)}(\mathbf{r}) + \mathcal{A}^{(2)}b^{(2)}(\mathbf{r}) \qquad [14.45]$$

where

$$\mathcal{A}^{(1)} = \cos\frac{\beta}{2};\quad \mathcal{A}^{(2)} = \sin\frac{\beta}{2} \qquad [14.46]$$

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.* (1977), 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 and this phenomenon is visible in many TEM images. We'll return to this topic in Chapter 23.

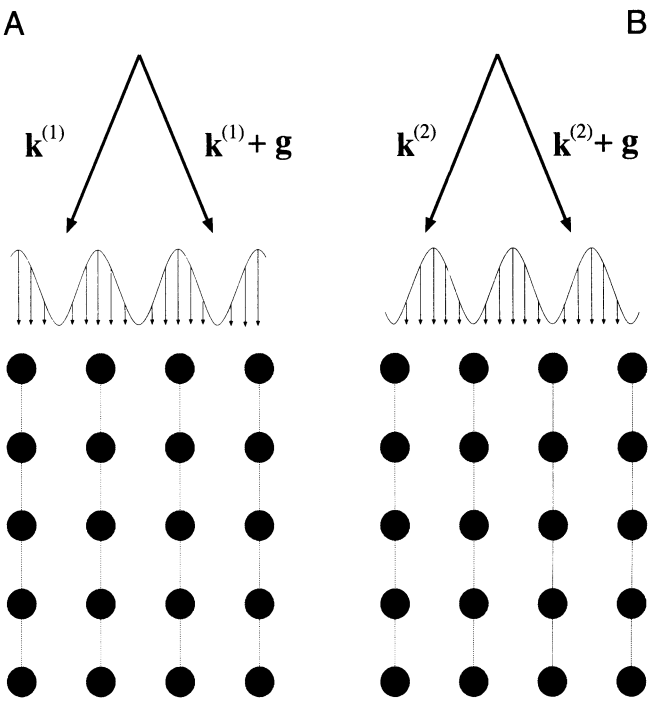


Figure 14.2. The two types of Bloch wave in the crystal aligned at the Bragg condition: (A) the maximum lies along the ion cores and Bloch wave 1 interacts strongly; (B) the maximum lies between the ions so that the interactions are weaker.

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:

- The 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 ϕ_0 and ϕ_g are not solutions to this equation and therefore don't actually exist 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 ϕ_0 and ϕ_g) is often phenomenological. Equations 14.27 and 14.28 give you the essential clue: each set of equations tells us how the Bloch waves are coupled. There are many possible solutions to the Schrödinger equation, and each Bloch wave is a plane wave; that is, it can be associated with a well-defined propagation vector $\mathbf{k}^{(j)}$ as shown in equation 14.9.

The Bloch waves are generally different because the U_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 tried to keep 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, and we have done the same.

REFERENCES

General References

Ashcroft, N.W. and Mermin, N.D. (1976) *Solid State Physics,* W.B. Saunders Co., Philadelphia. Chapter 8 ($2\pi/\lambda$ is used).

Howie, A. (1971) in *Electron Microscopy in Materials Science* (Ed. U. Valdrè), p. 275, Academic Press, New York.

Kittel, C.J. (1986) *Solid-State Physics,* 6th edition, John Wiley & Sons, New York. Chapters 1, 10, and 11 are particularly relevant.

Metherell, A.J.F. (1975) in *Electron Microscopy in Materials Science,* **II** (Eds. U. Valdrè and E. Ruedl), p. 397, CEC, Brussels. This is perhaps the clearest and most comprehensive article available on this subject (over 150 pages long).

Specific References

Bethe, H.A. (1928) *Ann. Phys. Lpz.* **87**, 55. In German.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, New York. Chapter 9.

Dispersion Surfaces

15

CHAPTER PREVIEW

The analysis of Bloch waves we gave in the previous chapter is closely related to the classical analysis of waves that you've seen in solid-state physics or semiconductor theory. In semiconductors in particular, we often talk of indirect and 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 $\mathbf{k}$) versus $\mathbf{k}$, the wave vector. This plot of $E(\mathbf{k})$ versus $\mathbf{k}$ is known as a dispersion diagram. For example, the band gap in Si is 1.1 eV, but the energy of most electrons in this material is somewhat smaller. We now follow the same approach to represent pictorially what we described in equations in Chapters 13 and 14. Remember that the difference to the solid-state physics approach is that, in TEM, the energy of the electrons is ≥ 100 keV.

In this chapter we will see the real origin of the extinction distance ξ_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; many established microscopists have survived without completely mastering this concept!

The dispersion surface is a pictorial representation of the relationship between **k** and energy.

Dispersion Surfaces

15

15.1. INTRODUCTION

The analysis of Bloch waves as they apply to electrons in solids is well documented in the solid-state literature. However, what we want from the theory is different than 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 $\mathbf{k}$ 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, $\mathbf{0}$ and $\mathbf{g}$. We can rewrite equation 14.35 incorporating equation 14.32 as

$$\frac{C_{\mathbf{g}}^{(j)}}{C_{\mathbf{0}}^{(j)}} = \frac{\left(k^{(j)}\right)^2 - \mathcal{K}^2}{U_{-\mathbf{g}}} = \frac{U_{\mathbf{g}}}{\left(k^{(j)} + g\right)^2 - \mathcal{K}^2} \qquad [15.1]$$

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 $k^{(j)2} - \mathcal{K}^2$, and thus to $k^{(j)} - \mathcal{K}$.

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.
- 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 $|A^{(j)}| = 0$. (We defined $A^{(j)}$ in Section 14.3 and gave an expression for it in Section 14.5.) If you multiply out the determinant, you get

$$A_{2n}\left(\mathbf{k}^{(j)}\right)^{2n} + A_{2n-1}\left(\mathbf{k}^{(j)}\right)^{2n-1} + \cdots = 0 \qquad [15.3]$$

The coefficient A_n depends on $\mathcal{K}^2$ (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 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.

15.2. THE DISPERSION DIAGRAM WHEN $U_g = 0$

When the electrons are in the vacuum, i.e., outside the specimen, the Fourier coefficients, U_g, are 0. We start with equation 14.34, namely,

$$\left(\left|\mathbf{k}^{(j)}\right| - \mathcal{K}\right)\left(\left|\mathbf{k}^{(j)} + \mathbf{g}\right| - \mathcal{K}\right) = \frac{|U_g|^2}{4\mathcal{K}^2} \qquad [15.4]$$

Remember that this equation was derived for the two-beam case. When $U_g = 0$, the left side of this equation is zero and the equation has two solutions.

$$\mathcal{K} = \left|\mathbf{k}^{(j)}\right| \quad \text{or} \quad \mathcal{K} = \left|\mathbf{k}^{(j)} + \mathbf{g}\right| \qquad [15.5]$$

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 $\mathbf{k}$ vectors have the same length, the two spheres

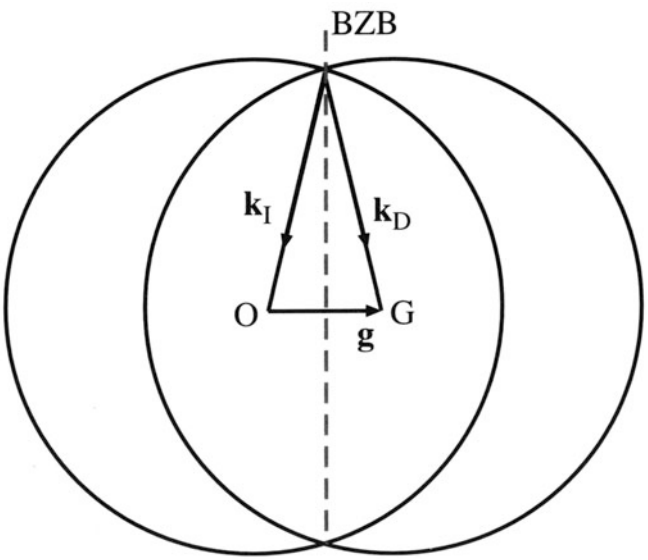


Figure 15.1. Cross section through two spheres of radii $\mathbf{k}_I$ and $\mathbf{k}_D$ centered on O and G, respectively. The spheres represent surfaces of constant energy and the dotted line is the trace of the diffracting plane (and is also equivalent to the Brillouin-zone boundary).

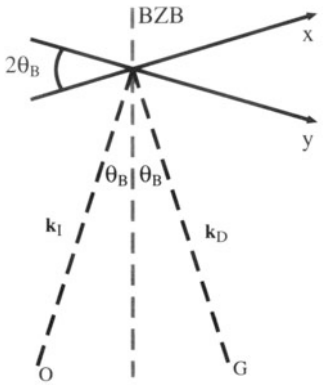


Figure 15.2. An enlarged view of the intersection of the two dispersion spheres at the Brillouin-zone boundary. The projections of the two dispersion surfaces approximate to straight lines x and y, which are normal to $\mathbf{k}_D$ and $\mathbf{k}_I$, respectively.

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

$$E = \frac{p^2}{2m} = \frac{h^2\chi^2}{2m} \qquad [15.6]$$

where $\boldsymbol{p}$, the momentum, is related to the wave vector in a vacuum, χ, by $\boldsymbol{p} = h\chi$. Here, χ is the $\mathbf{k}$ when the electron is in a vacuum.

Rearranging, we have

$$\chi = \left\{\frac{2m}{h^2}E\right\}^{\frac{1}{2}} \qquad [15.7]$$

The dotted line drawn in Figure 15.1 represents a plane which is defined by the circle created by the intersecting spheres. In solid-state physics this plane is known as the Brillouin-zone boundary (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 in projection 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 $\mathcal{K}$ can never be equal to $|\mathbf{k}_I|$ or $|\mathbf{k}_D|$. Since equation 15.4 is quadratic we

must have two values for |**k**|. So, the two "spheres" can't intersect if $U_g \neq 0$. We notice 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 corresponds to $\mathbf{k}^{(1)}$ and the lower to $\mathbf{k}^{(2)}$. We now have vectors $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$ where we used to have $\mathbf{K}_I$ and $\mathbf{K}_D$. There are some critical points to remember in this discussion from Chapters 13 and 14:

- 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.

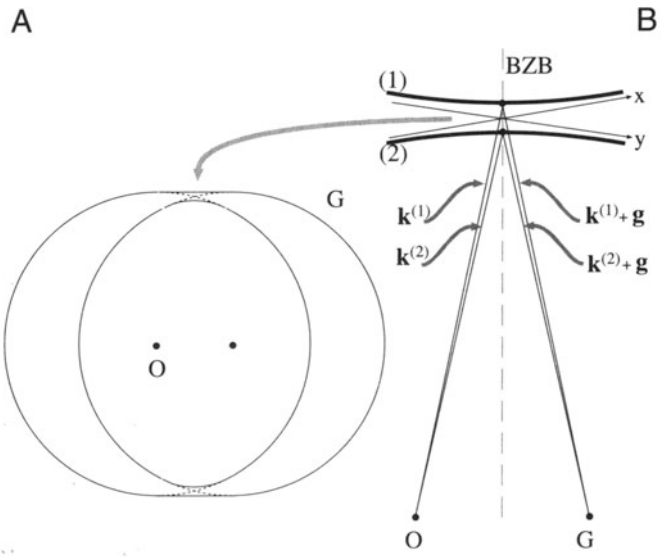


Figure 15.3. When the electron is inside the specimen (i.e., $U_g \neq 0$) and there are two values of **k**, the two dispersion spheres can't intersect and two branches of the dispersion surface, (1) and (2), are created: (A) and (B) show the nonintersecting spheres and an enlarged view showing pairs of vectors, $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$, and $\mathbf{k}^{(1)} + \mathbf{g}$ and $\mathbf{k}^{(2)} + \mathbf{g}$.

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 start by assuming an idealized thin specimen with parallel surfaces. 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.

Finally, we extend the points M_1 and M_2 back to the χ spheres in Figure 15.4B. 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, etc., is always outside the crystal.

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) which will then be nonzero are $C_{\mathbf{0}}^{(1)}$, $C_{\mathbf{0}}^{(2)}$, $C_{\mathbf{g}}^{(1)}$, and $C_{\mathbf{g}}^{(2)}$.

First, we need to know which points on the dispersion surface will actually correspond to the diffraction condition we have chosen. Next, we need to know the orientation of the specimen relative to the beam and the orientation of the Bragg planes.

We begin by considering 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**. If we align the incident beam parallel to the $(hk\ell)$ planes, then we will excite points M_1^B and M_2^B on separate branches of the dispersion surface shown in Figure 15.4A. 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 χ 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 $M_1 M_2$ 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 which produces it.

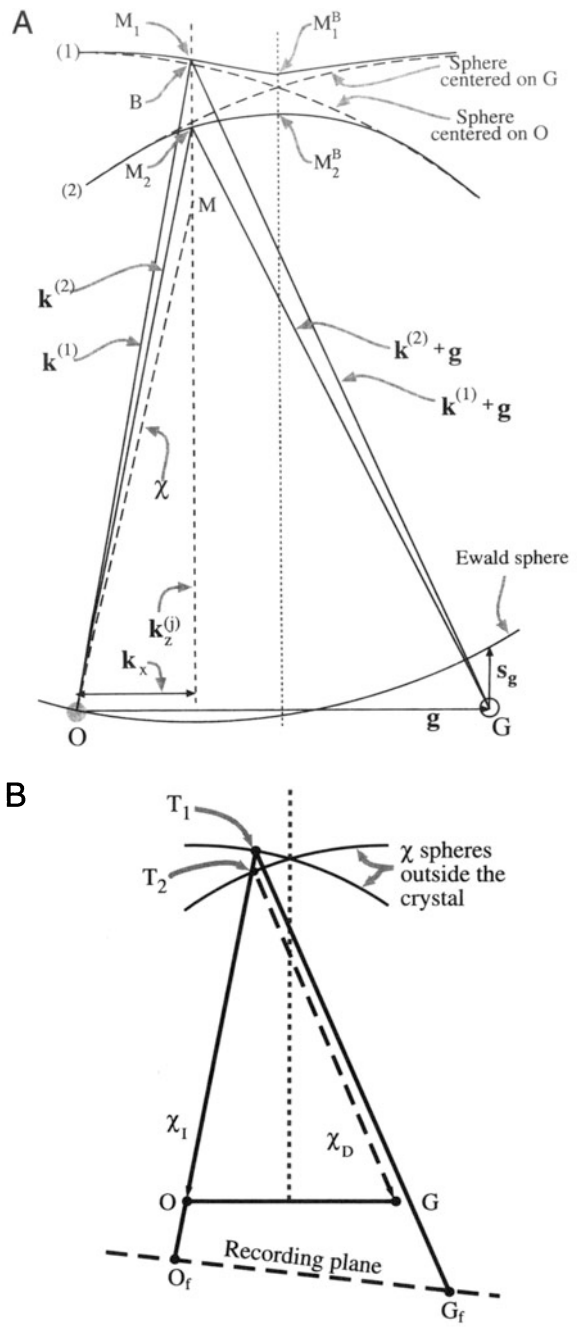


Figure 15.4. (A) Combination of the dispersion surfaces (1) and (2), centered on O and G, with the Ewald sphere construction. The surface of the specimen has been set to be parallel to **g**, so points M_1^B and M_2^B on the branches (1) and (2) are excited. The incident beam direction is given by the 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.

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.

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 χ is always shorter than $\mathcal{K}$ 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. However, the tie line is always parallel to **n** so the tie line is no longer parallel to the BZB. Remember: this construction matches 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 the beam to be normal to the surface in that case.

> The tie line is a graphical method of satisfying the boundary conditions imposed by the TEM specimen.

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.

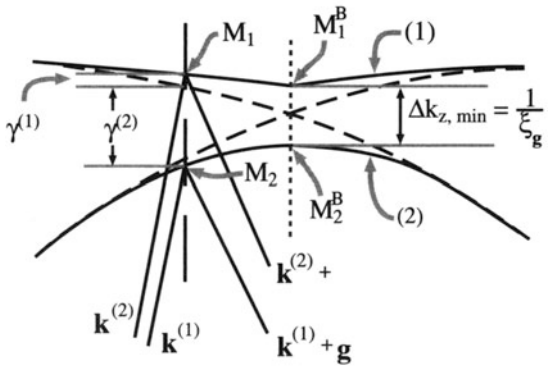


Figure 15.5. An enlarged region of Figure 15.4A showing how the vectors $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$ are related to the quantities $\gamma^{(1)}$ and $\gamma^{(2)}$ and the distance Δk_z.

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.

The wedge has been drawn with the top surface parallel to **g**. Thus we have tie line $\mathbf{n}_1$. 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 wavelength must be χ and that χ defines a pair of spheres centered on O and G. So we extend the $\mathbf{n}_2$ tie lines until they reach the χ 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 $\chi_0^{(1)}$, $\chi_0^{(2)}$, etc., as also shown in Figure 15.6A.

Now we have reached the final step: we have to relate these beams to the DP. Yes, they are real beams because we are now outside the specimen and in a vacuum. We show this in Figure 15.6B. All of the χ beams have been related to point O_1 because $\chi_0^{(1)}$ is the direct beam. Remember: $\chi_0^{(1)}$ is not vertical because **g** is horizontal. The vectors $\chi_0^{(1)}$ and $\chi_0^{(2)}$ are not parallel because they are both radii of the circle χ_0 and originate at different points on the circle.

The conclusion is that we will have two spots at O and two spots at G. In other words, the fact that we have 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.

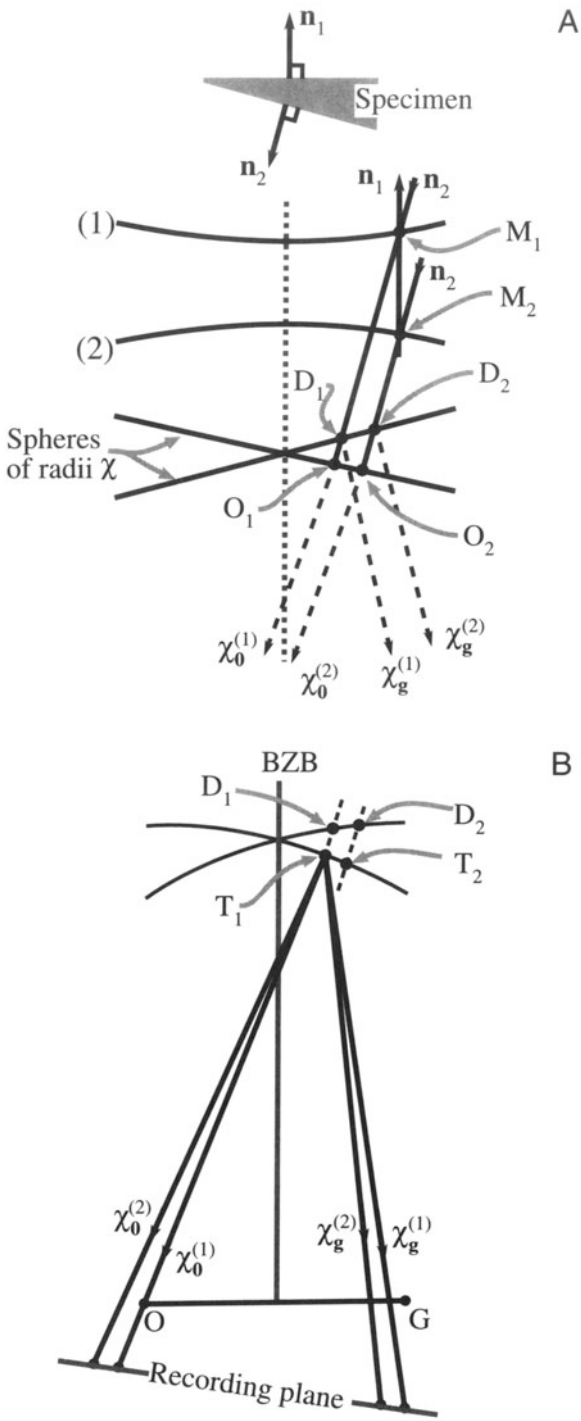


Figure 15.6. (A) The same diagram as Figure 15.4B, but for a wedge specimen with top surface parallel to **g** (normal $\mathbf{n}_1$) and the bottom surface normal $\mathbf{n}_2$. Instead of exciting two points, O_1 and O_2, we excite two more, D_1 and D_2, which correspond to the plane waves $\chi_0^{(1)}$, $\chi_0^{(2)}$, 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.

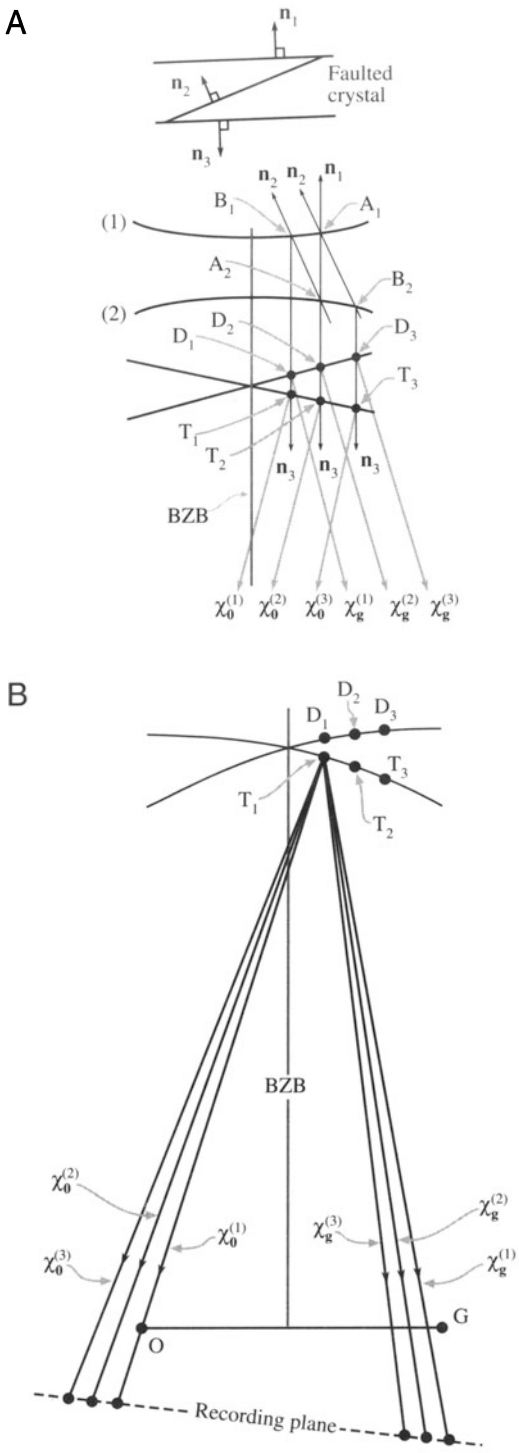


Figure 15.7. (A) An enlarged view of the dispersion surface in Figure 15.6 close to the BZB, but this time for a specimen in which both surfaces are parallel to **g** but there is an inclined fault which produces a third wave $\chi_0^{(3)}$ and $\chi_g^{(3)}$. (B) If we then move all the vectors to O_1 again, we predict there will be three spots at O and three at G.

- 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.

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 initially traveling with wave vector χ outside the crystal. When the beam enters the crystal the z component of this wave vector changes (i.e., 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.

Aside: 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 are considering only two beams. Clearly, we can draw in $\mathbf{k}_{\mathbf{g}}^{(2)}$ and $\mathbf{k}_{\mathbf{g}}^{(1)}$ by adding **g**. Now, how does $\mathbf{k}_{\mathbf{0}}^{(1)}$, say, relate to **K**? The point K is also determined by the tie line through χ, and lies on the circle centered on O. Most importantly, neither $\mathbf{k}_1$ nor $\mathbf{k}_2$ is equal to $\mathcal{K}$. If you look back at equation 13.41 you can see that

$$\mathbf{k}_z^{(i)} - \mathcal{K}_z = \gamma^{(i)} \qquad [15.8]$$

So $\gamma^{(i)}$ is simply the distance of the point M_j from the $\mathcal{K}$ sphere centered on O. We can write this relationship explicitly

$$\mathbf{k}^{(i)} = \mathbf{k}_z^{(1)} + \mathbf{k}_x^{(i)} \qquad [15.9]$$

$$= \left(K_z + \gamma^{(i)}\right)\mathbf{u}_z + \mathrm{k}_x\mathbf{u}_x \qquad [15.10]$$

Notice that the last term here is independent of i. Look again at Figure 15.4. You can see that Δk_z is a minimum when M_1 and M_2 lie on the BZB. In that situation

$$\Delta k_{z\,\min} = \gamma^{(1)} - \gamma^{(2)} \qquad [15.11]$$

Simply by looking at the diagram, and as expected from Chapter 13, you also know that

$$\gamma^{(1)} - \gamma^{(2)} = \frac{U_g}{k} = \frac{1}{\xi_g} \qquad [15.12]$$

So

[15.13]

The origin of the thickness oscillations is the difference in wavelength of the two Bloch waves. It's the beating between the two Bloch waves.

Thus you see that the gap Δk_z at the BZB is given by the reciprocal of the extinction distance.

- We have a crystal, therefore $U_g \neq 0$.
- Since $U_g \neq 0$, we have two branches to the dispersion surface and hence a band gap.
- The band gap is Δk_z.
- Hence we have a finite extinction distance (i.e., ξ_g is not infinite).

An aside: think how s_{eff} and s would be related if ξ_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 $|\mathcal{K}|$) we see that s_g is nonzero. We can easily see from the equations in Section 13.10 that, in general, Δk_z is given by

$$\Delta k_z = s_{\text{eff}} = \frac{1}{\xi_{\text{eff}}} \qquad [15.14]$$

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?

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

$$\psi(\mathbf{r}) = \mathcal{A}^{(1)}b^{(1)} + \mathcal{A}^{(2)}b^{(2)} \qquad [15.15]$$

The relative contributions of the two Bloch waves, $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$, were shown to be cos β/2 and sin β/2, respectively, and $w = \cot\beta = s\xi_g$.

We also showed in Section 13.8 that

$$b^{(1)}\left(\mathbf{k}^{(1)},\mathbf{r}\right) = C_0^{(1)}e^{2\pi i\,\mathbf{k}^{(1)}\cdot\mathbf{r}} + C_g^{(1)}e^{2\pi i\,(\mathbf{k}^{(1)}+\mathbf{g})\cdot\mathbf{r}} \qquad [15.16]$$

and

$$b^{(2)}\left(\mathbf{k}^{(2)},\mathbf{r}\right) = C_0^{(2)}e^{2\pi i\,\mathbf{k}^{(2)}\cdot\mathbf{r}} + C_g^{(2)}e^{2\pi i\,(\mathbf{k}^{(2)}+\mathbf{g})\cdot\mathbf{r}} \qquad [15.17]$$

The Bloch-wave coefficients were given by equation set 13.31

$C_0^{(1)}$	$C_0^{(2)}$	$C_g^{(1)}$	$C_g^{(2)}$
cos β/2	sin β/2	–sin β/2	cos β/2

Now we can consider some special cases and examine the actual values for $C_0^{(1)}$, $\mathcal{A}^{(1)}$, etc. (Table 15.1).

For the Bragg case, $s_g = 0$, **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 $s_g < 0$, we now have cos (β/2) > sin (β/2) so that $\mathcal{A}^{(1)}$ is greater than $\mathcal{A}^{(2)}$. If we reverse the sign of **s**, cos (β/2) < sin (β/2) and $\mathcal{A}^{(1)}$ is less than $\mathcal{A}^{(2)}$.

The result is that whether Bloch wave 1 or Bloch wave 2 has the larger amplitude depends on the sign of **s**.

Now, let's relate this information to the dispersion surface shown in Figure 15.4. When $s_g < 0$, as shown here, the M_1M_2 tie line is to the left of the BZB, which is associ-

Table 15.1. Values of Bloch-Wave Variables

s	w	β	β/2	cos (β/2)	sin (β/2)
0	0	π/2	π/4	$1/\sqrt{2}$	$1/\sqrt{2}$
+0.01	+Δ	π/2 – δ	π/4 – δ/2	$1/\sqrt{2} + \varepsilon$	$1/\sqrt{2} - \varepsilon$
–0.01	–Δ	π/2 + δ	π/4 – δ/4	$1/\sqrt{2} - \varepsilon$	$1/\sqrt{2} + \varepsilon$

ated 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 the 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 picture the dispersion surface for the case where $U_g = 0$ 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 band gap tends to decrease as the rank of the neighboring branches decreases.

In Chapter 26, 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.

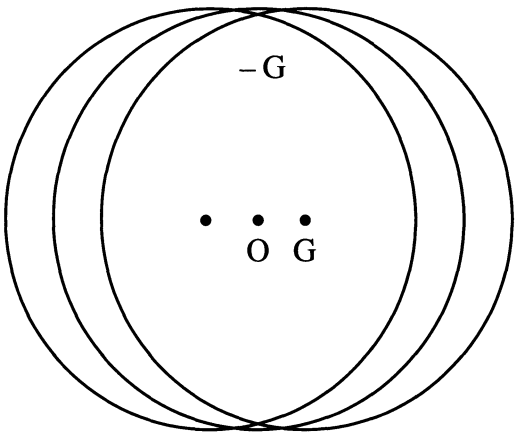


Figure 15.8. Three dispersion spheres due to three reflections, -G, O, and G. If we had n spots we would have n spheres.

We follow the convention used by Metherell (1975) and number the branches of the dispersion surface from the top down. 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.* (1977). This was fine when only two branches were considered.

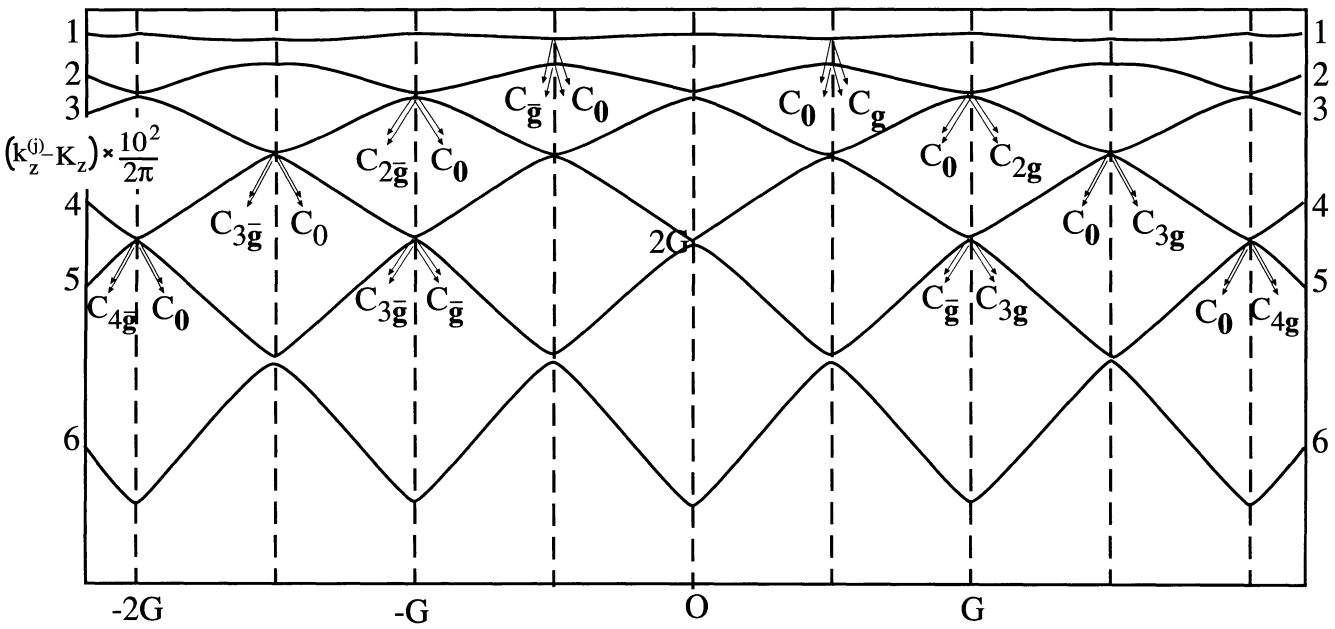


Figure 15.9. Six branches of the dispersion surfaces. The two branches $i = 1$ and $i = 2$ have the highest energy and give the largest band gap; notice that these branches give the terms in C_0 and C_g; smaller gaps occur between branches with lower energy. The diagram can be approximated to a set of spheres centered on O, ±G, and ±2G, etc.; C_0 is "normal" to the sphere centered on O, while C_g is "normal" to the sphere centered on **g**, etc.

We can still associate the amplitudes C_0, C_g, etc. with the sphere centered on **0**, **g**, etc. The result is shown by the labels C_0, C_g, etc. in Figure 15.9. For example, imagine the original spheres centered on **0** and **g**; they intersect on the BZB which passes through **g**/2, so the C_0, C_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_0, C_{3g} are labeled. As a general rule, C_{ng} will be largest for the pair of reflections which are excited, i.e., **0** and n**g**, and will be located on the n**g**/2 BZB.

We now extend these arguments to the situation where many beams are excited. Values of C other than C_0 and C_{ng} will be nonzero, 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 $C_{\mathbf{0}}^{(1)}$, $C_{\mathbf{0}}^{(2)}$, $C_{\mathbf{0}}^{(1)}$, $C_{2\mathbf{g}}^{(1)}$, 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 $\Delta k_{4,5}$ is small). The extinction distance for the coupling of **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 G BZB.

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 which a stacking 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 stacking 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 nonplanar 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 $\mathcal{K}$ 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 $\mathcal{K}$ 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 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 only nonzero because we have a crystal. 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, ξ_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 $\mathbf{k}$ versus $\mathcal{K}$ relationship. We'll close by quoting the result derived by Kato (1957).

> 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 which discuss dispersion surfaces and band gaps in semiconductors, beware of the $2\pi/\lambda$ versus $1/\lambda$ problem since many of these texts are 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

General References

Ashcroft, N.W. and Merman, N.D. (1976) *Solid State Physics,* W.B. Saunders Co., Philadelphia. Chapter 8 ($2\pi/\lambda$ is used).

Kittel, C.J. (1986) *Solid-State Physics,* 6th edition, John Wiley & Sons, New York. Chapters 1, 10, and 11 are particularly relevant.

Metherell, A.J.F. (1975) in *Electron Microscopy in Materials Science,* **II**, (Eds. U. Valdrè and E. Ruedl), p. 397, CEC, Brussels. *The* reference.

Specific References

Kato, N. (1958) *Acta. Cryst.* **11**, 885.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, New York. Chapter 9.

Diffraction from Crystals

16

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.

Diffraction from Crystals

16

16.1. REVIEW OF DIFFRACTION FROM A PRIMITIVE LATTICE

In Chapters 11 to 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

$$\phi_g = \frac{ai\lambda F_g}{V_c \cos \Theta} \sum_n e^{-2\pi i \mathbf{K} \cdot \mathbf{r}_n} e^{2\pi i \mathbf{k}_D \cdot \mathbf{r}} \qquad [16.1]$$

where F_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 (1957), 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)

$$|f(\theta)|^2 = \frac{d\sigma(\theta)}{d\Omega} \qquad [16.2]$$

  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 IV that we can detect differences between differently bonded atoms using EELS.

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

Table 16.1. Selected Values of $f(\theta)$, the Atomic Scattering Amplitude at $\theta = \theta_B$[a]

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

[a]These are values given by Edington (1976) 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.

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 **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

$$A_{\text{cell}} = \frac{e^{2\pi ikr}}{r} \sum_i f_i(\theta) e^{2\pi i \mathbf{K}\cdot\mathbf{r}_i} \qquad [16.3]$$

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

$$\mathbf{r}_i = x_i\,\mathbf{a} + y_i\,\mathbf{b} + z_i\,\mathbf{c} \qquad [16.4]$$

We'll start by considering only the case where $\mathbf{K} = \mathbf{g}$ since this is an infinite, perfect crystal

$$\mathbf{K} = h\,\mathbf{a}^* + k\,\mathbf{b}^* + \ell\,\mathbf{c}^* \qquad [16.5]$$

So we can write

$$F_{hk\ell} = \sum_i f_i e^{2\pi i\left(hx_i + ky_i + \ell z_i\right)} \qquad [16.6]$$

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 $F_{hk\ell}$.

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 lattices 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

$$F = f\left\{1 + e^{\pi i(h+k+\ell)}\right\} \qquad [16.7]$$

Now, since h, k, ℓ are all integers, if we define the sum $h + k + \ell = N$, then the exponential can take two values: +1, for N even; and -1, for N odd.

Thus, we can say that:

- $F = 2f$ if $h + k + \ell$ is even,
- $F = 0$ if $h + k + \ell$ is odd.

There are **no** other possibilities. The resulting 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.

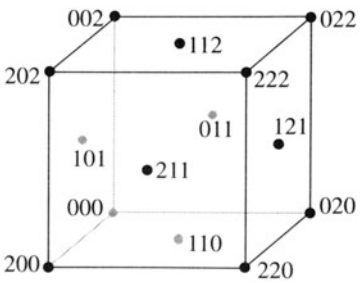


Figure 16.1. The reciprocal lattice for the bcc crystal structure. The lattice points that correspond to systematic absences have been removed, so the actual arrangement of points is an fcc lattice.

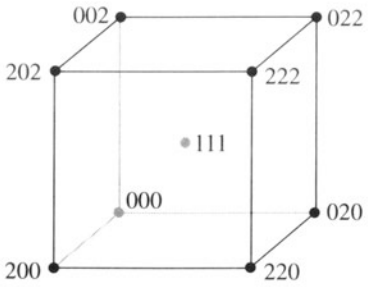


Figure 16.2. The reciprocal lattice for the fcc crystal structure. The lattice points that correspond to systematic absences have been removed, so the actual arrangement of points is a bcc lattice.

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

$$(\mathbf{x}, \mathbf{y}, \mathbf{z}) = (0, 0, 0), (\tfrac{1}{2}, \tfrac{1}{2}, 0), (\tfrac{1}{2}, 0, \tfrac{1}{2}), (0, \tfrac{1}{2}, \tfrac{1}{2}) \qquad [16.8]$$

Substituting these values for $\mathbf{r}_i$ into equation 16.6 gives

$$F = f\left\{1 + e^{\pi i(h+k)} + e^{\pi i(h+\ell)} + e^{\pi i(k+\ell)}\right\} \qquad [16.9]$$

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 are 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 three reasons.

- Except for (0001), the patterns can be different for every material because the *c/a* ratio is different.
- 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

$$(x, y, z) = (0, 0, 0), (\tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{2}) \qquad [16.10]$$

Substituting these values for $\mathbf{r}_i$ into equation 16.6 gives

$$F = f\left\{1 + e^{2\pi i\left(\frac{h}{3} + \frac{2k}{3} + \frac{\ell}{2}\right)}\right\} \qquad [16.11]$$

We simplify the notation by setting $h/3 + 2k/3 + \ell/2 = X$; 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

$$|F|^2 = f^2\left(1 + e^{2\pi iX}\right)\left(1 + e^{-2\pi iX}\right) = f^2\left(2 + e^{2\pi iX} + e^{-2\pi iX}\right) \qquad [16.12]$$

$$|F|^2 = f^2\left(2 + 2\cos 2\pi X\right) = f^2\left(4\cos^2 \pi X\right) \qquad [16.13]$$

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 $h + 2k = 3m + 1$ and ℓ is odd,
- $|F|^2 = f^2$ if $h + 2k = 3m + 1$ 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 $(hki\ell)$, then we can show that

$$uh + vk + ti + w\ell = 0 \qquad [16.14]$$

The normal to the plane (h,k,i,ℓ) is actually the Cartesian vector $[h,k,i,\ell/\lambda]$, and likewise the crystallographic direction $[u,v,t,w]$ is actually the vector $[u,v,t,\lambda w]$ in the Cartesian system. So using the four-index Cartesian vector notation, equation 16.14 can be written as

$$[u,v,t,\lambda w]\cdot[h,k,i,\ell/\lambda] = 0 \qquad [16.15]$$

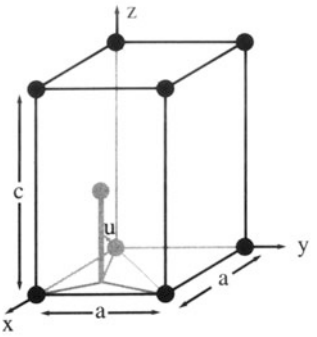


Figure 16.3. The hcp unit cell showing the four axes used in the Miller–Bravais indexing system. The three axes in the basal plane, *x*, *y*, and *u*, are all crystallographically equivalent and the *z*-axis is normal to the basal plane.

In cubic crystals, the direction $[hk\ell]$ is always normal to the plane $(hk\ell)$, but this is not the case for hcp crystals. You can show using some simple geometry that

$$\lambda^2 = \left(\frac{2}{3}\right)\left(\frac{c}{a}\right)^2 \qquad [16.16]$$

Thus the Cartesian vector $[HKIL]$, which is normal to the plane $(hki\ell)$, is the vector

$$\left(h, k, i, \frac{3}{2}\left(\frac{a}{c}\right)^2\ell\right) \qquad [16.17]$$

So, $[11\bar{2}0]$ 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 $(hki\ell)$ 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 $\cos\phi$ in the form

$$\cos\phi = \frac{hd + ke + \frac{1}{2}(he + kd) + \frac{3}{4}\ell g\left(\frac{a}{c}\right)^2}{\left\{h^2 + k^2 + hk + \frac{3}{4}\ell^2\left(\frac{a}{c}\right)^2\right\}^{\frac{1}{2}}\left\{d^2 + e^2 + de^2 + \frac{3}{4}g^2\left(\frac{a}{c}\right)^2\right\}^{\frac{1}{2}}} \qquad [16.18]$$

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 $i = -(u + v)$. We'll come across simpler expressions in the cubic system in Chapter 18.

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 alternative 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 $\pi i(h + k + \ell)$. Of course, the atomic scattering amplitudes, f, are also different for the two atoms. We can write this expression for F as

$$F = \left\{f_{Na} + f_{Cl}\, e^{\pi i(h+k+\ell)}\right\}\left\{1 + e^{\pi i(h+k)} + e^{\pi i(h+\ell)} + e^{\pi i(k+\ell)}\right\} \qquad [16.19]$$

This again gives rise to some rules:

- $F = 4(f_{Na} + f_{Cl})$ if h, k, ℓ are all even,
- $F = 4(f_{Na} - f_{Cl})$ 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 $(f_{Na} \pm f_{Cl})$ 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 $4(f_{Na} - f_{Cl})$ 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 31 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})

$$F = \left\{f_{Ga} + f_{As}\, e^{\frac{\pi}{2} i(h+k+\ell)}\right\} F_{fcc} \qquad [16.20]$$

So the rules are slightly more complicated:

- $F = 0$ if h, k, ℓ are mixed as always for fcc,
- $F = 4\,(f_{Ga} \pm i f_{As})$ if h, k, ℓ are all odd,
- $F = 4\,(f_{Ga} - f_{As})$ if h, k, ℓ are all even and $h + k + \ell = 2N$ where N is odd (e.g., the 200 reflection),
- $F = 4\,(f_{Ga} + f_{As})$ if h, k, ℓ are all even and $h + k + \ell = 2N$ 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, remember that we only see intensities (i.e., $|F|^2$ not F), so $|F|^2$ is $16(f_{Ga}^2 + f_{As}^2)$ and is independent of the sign initially present. Of course, the structure factor is still different than the others derived here.

Si: Now we can easily extend this analysis to Si, Ge, or diamond. Just replace f_{Ga} and f_{As} in our results with f_{Si}. The major change is that F is zero when $h + k + \ell = 2N$ and N is odd. The best known example of this is again the 200 reflection. For Si it has $F = 0$, but F is finite for GaAs.

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, and AlN, 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 ($L1_0$): For this material, we can easily modify the original treatment of the bcc structure, since now the centering atom is different. If we choose to place

the Ni atom at (0, 0, 0) and the Al atoms at [1/2, 1/2, 1/2], then

$$F = \{f_{Ni} + f_{Al}\, e^{\pi i(h+k+\ell)}\} \qquad [16.21]$$

This leads to two values for F, *neither* of which is zero:

- $F = f_{Ni} + f_{Al}$ if $h + k + \ell$ is even,
- $F = f_{Ni} - f_{Al}$ if $h + k + \ell$ 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 a simple-cubic lattice 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 are 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_3Al (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

$$F = f_{Al} + f_{Ni}\{e^{\pi i(h+k)} + e^{\pi i(h+\ell)} + e^{\pi i(k+\ell)}\} \qquad [16.22]$$

The rules for Ni_3Al are:

- $F = (f_{Al} + 3f_{Ni})$ if h, k, ℓ are all even or all odd,
- $F = (f_{Al} - f_{Ni})$ if h, k, ℓ are mixed.

Again, all of the possible reciprocal lattice points of the simple-cubic lattice will give rise to Bragg reflections because the structure is really simple cubic. The mixed $hk\ell$ reflections are now the chemically sensitive reflections. This material is particularly interesting, since it can be heat-treated to randomize the distribution of the two elements; then each site will be occupied by 75% Ni, 25% Al, and F for mixed $hk\ell$ will be zero. For this reason, reflections with mixed $hk\ell$ 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

$$F = f_{Ti} + f_{Al}\, e^{2\pi i\left(\frac{h}{3} + \frac{2k}{3} + \frac{\ell}{2}\right)} \qquad [16.23]$$

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

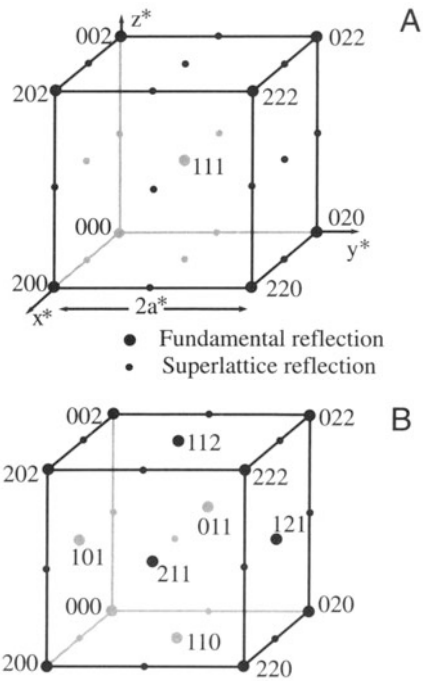


Figure 16.4. The reciprocal lattices for (A) the Ni_3Al and (B) the NiAl structures. In (A) Ni_3Al is fcc, so the fcc forbidden reflections (h, k, ℓ mixed even and odd) are allowed and become chemically sensitive (superlattice) reflections. In (B) NiAl is bcc, so the bcc forbidden reflections (if $h + k + \ell$ odd) are now allowed superlattice reflections.

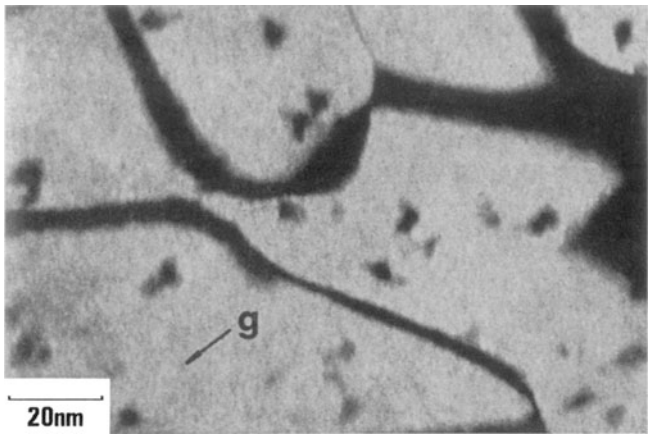


Figure 16.5. DF image from a chemically sensitive 110 reflection showing bright ordered domains in Cu_3Au. The dark areas in the bright domains are regions of local disorder induced by ion beam damage.

chemically sensitive superlattice 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.

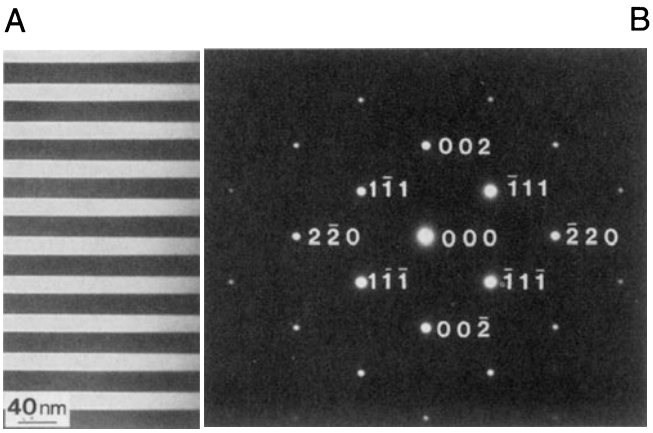


Figure 16.6. (A) DF image from a 002 superlattice reflection in GaAs. The $Al_xGa_{1-x}As$ is the lighter region because Al has replaced Ga in the GaAs (darker regions). (B) Diffraction pattern showing the less intense 002 and other superlattice reflections.

Figure 16.5 shows an image from Cu_3Au, the archetypal A_3B ordered fcc structure. The crystal has been 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.* 1976). The DF image has been formed using the 110 reflection, which we know is a superlattice reflection. By destroying the ordering, 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 show 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 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 which 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 which 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.* 1983).

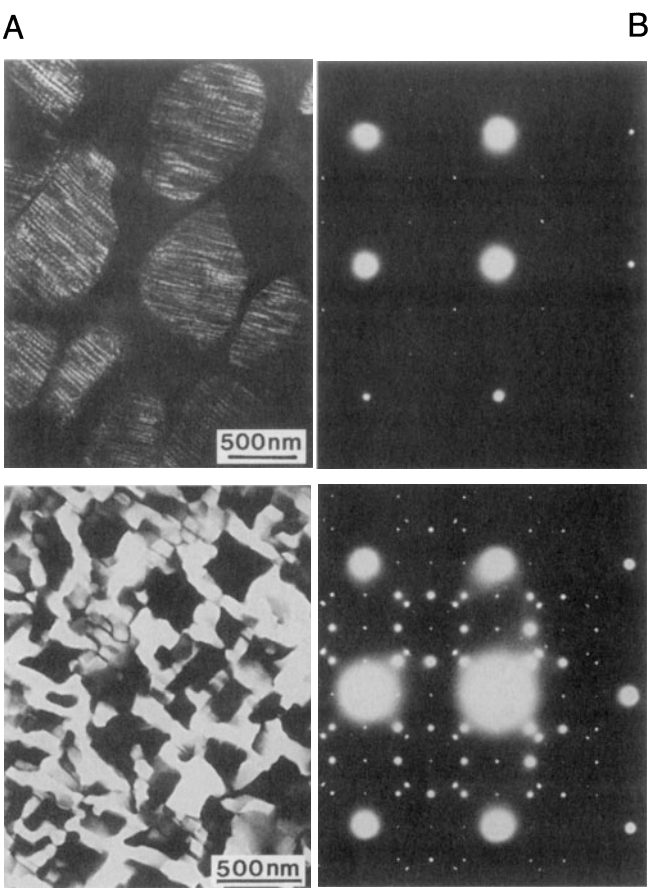


Figure 16.7. (A) DF image of ordered V_6C_5 and (B) accompanying diffraction pattern. (C) DF image of V_8C_7 and (D) diffraction pattern. In both carbides the ordering is due to vacancies on the C sublattice.

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

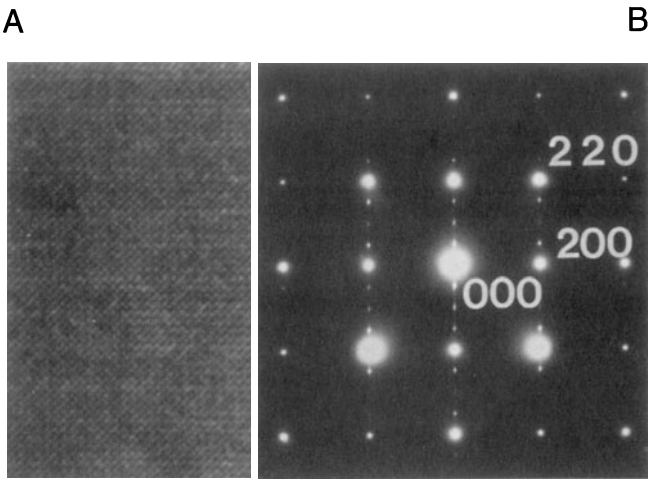


Figure 16.8. (A) Artificial GaAs/Al_xGa_{1-x}As structure in which order is created by alternating four layers of GaAs and four of (Al_xGa_{1-x})As. (B) DP showing three superlattice spots between the fundamental reflections in the 020 direction.

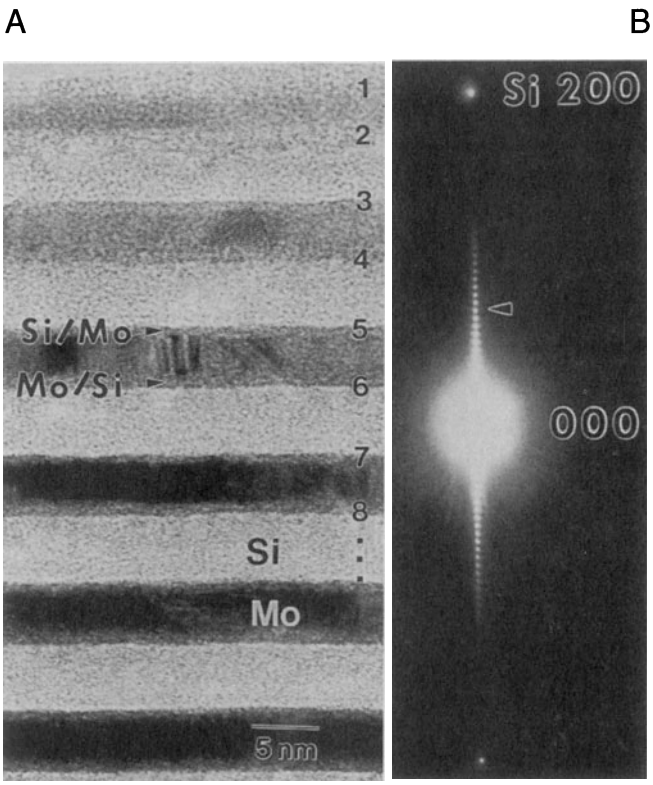


Figure 16.9. (A) Artificial superlattice of Si and Mo layers ~5 nm thick (B) Expanded DP around 000 showing many superlattice spots (arrowed). The large spacing of the superlattice in real space results in very small spacing of the superlattice reflections in the DP in reciprocal space. Compare with Figure 16.8.

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 the magnification of a TEM image is usually subject to a ±10% uncertainty.)

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 diffraction pattern is the [011] in Si so that the 200 reflection should be absent, according to Section 16.4. The reason it is actually present is that, since we are oriented at the zone axis, the 111 beam, which has $F \neq 0$, acts like a new incident beam and is rediffracted by the (111) 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, you should quickly become familiar with the International Tables for Crystallography (Hahn 1988), 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 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 will

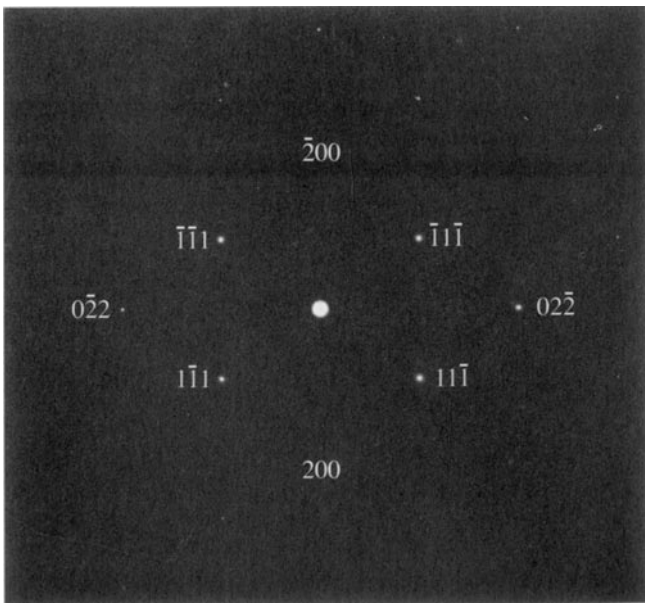


Figure 16.10. The [011] diffraction pattern from Si. The 200 reflection is forbidden, but it is present because the allowed $11\bar{1}$ diffracted beam acts like a new incident beam and is rediffracted by the $(1\bar{1}1)$ plane. The sum of the two allowed reflections, $(11\bar{1}) + (1\bar{1}1)$, results in a 200 reflection, which is so weak you may not see it.

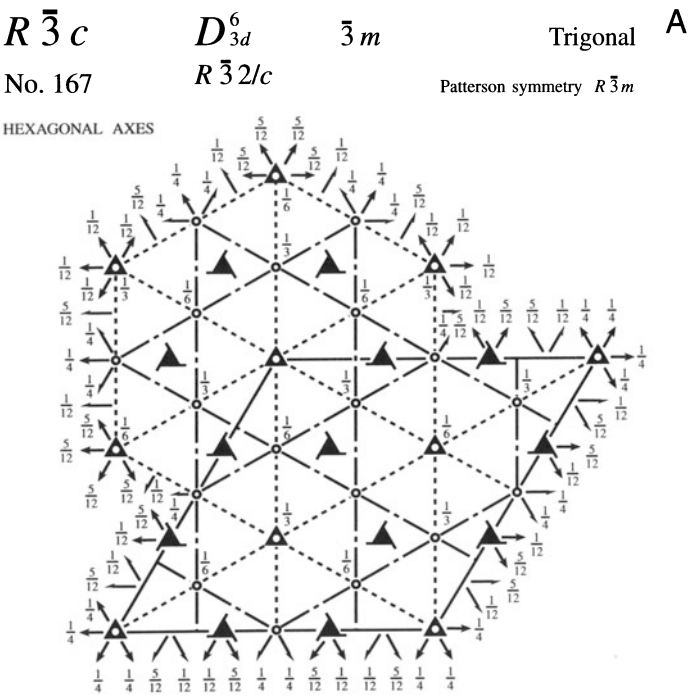


Figure 16.11. (A) Symmetry information, as given in the International Tables for trigonal α-Al_2O_3, with space group $R\bar{3}c$, showing the two possible unit cells based on the rhombohedral and hexagonal cells. The symmetry elements at specific lattice points are also indicated. (B) The atomic positions for the two choices of unit cells in (A).

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
18	e	.2	$x,0,\frac{1}{4}$	$0,x,\frac{1}{4}$	$\bar{x},\bar{x},\frac{1}{4}$	$\bar{x},0,\frac{3}{4}$ $0,\bar{x},\frac{3}{4}$ $x,x,\frac{3}{4}$
						no extra conditions
18	d	$\bar{1}$	$\frac{1}{2},0,0$	$0,\frac{1}{2},0$	$\frac{1}{2},\frac{1}{2},0$	$0,\frac{1}{2},\frac{1}{2}$ $\frac{1}{2},0,\frac{1}{2}$ $\frac{1}{2},\frac{1}{2},\frac{1}{2}$
						$hkil$: $l=2n$
12	c	3.	$0,0,z$	$0,0,\bar{z}+\frac{1}{2}$	$0,0,\bar{z}$	$0,0,z+\frac{1}{2}$
						$hkil$: $l=2n$
6	b	$\bar{3}$.	$0,0,0$	$0,0,\frac{1}{2}$		
						$hkil$: $l=2n$
6	a	32	$0,0,\frac{1}{4}$	$0,0,\frac{3}{4}$		
						$hkil$: $l=2n$

Positions

Multiplicity, Wyckoff letter, Site symmetry			Coordinates			Reflection conditions
						General:
36	f	1	(1) x,y,z	(2) z,x,y	(3) y,z,x	hhl : $l=2n$
			(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) $\bar{z}+\frac{1}{2},\bar{y}+\frac{1}{2},\bar{x}+\frac{1}{2}$	hhh : $h=2n$
			(7) $\bar{x},\bar{y},\bar{z}$	(8) $\bar{z},\bar{x},\bar{y}$	(9) $\bar{y},\bar{z},\bar{x}$	
			(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
6	e	.2	$x,\bar{x}+\frac{1}{2},\frac{1}{4}$	$\frac{1}{4},x,\bar{x}+\frac{1}{2}$	$\bar{x}+\frac{1}{2},\frac{1}{4},x$	
			$\bar{x},x+\frac{1}{2},\frac{3}{4}$	$\frac{3}{4},\bar{x},x+\frac{1}{2}$	$x+\frac{1}{2},\frac{3}{4},\bar{x}$	no extra conditions
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. (Continued)

can work out 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 (1985) summarizes the analysis for this particular example.

That was the traditional approach. Now, you should have access to Desktop Microscopist or Crystal Kit. Alternatively, use EMS over the WWW (Section 1.5). In all these software packages you can just type in your space group or pull down a menu to find the structure-factor information.

Table 16.2. Examples of Selection Rules for Several Crystal Structures Where *F* is the Structure Factor

Crystal type	Reflection present for	F	Lattice points per cell
Primitive	Any h, k, ℓ	f	1
Body centered cubic	$(h + k + \ell) = 2n$	$2f$	2
Face centered cubic including GaAs and NaCl	$h, k,$ and ℓ all odd or all even	$4f$	4
Diamond cubic	As fcc but if all even and $h + k + \ell \neq 4N$ then absent, anyway.		
Base centered	h, k and ℓ all odd or all even	$2f$	2
			Example
Hexagonal close-packed	$h + 2k = 3n$ with ℓ odd	0	0001
	$h + 2k = 3n$ with ℓ even	$2f$	0002
	$h + 2k = 3n \pm 1$ with ℓ odd	$f\sqrt{3}$	$01\bar{1}1$
	$h + 2k = 3n \pm 1$ with ℓ even	f	$01\bar{1}0$

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.

In practice, it will become important that you simply *know* some of the DPs for your material. You can, however, look up schematic indexed patterns in some of the textbooks listed in Chapter 1, but the best sources are Andrews *et al.* (1971) and Edington (1976) and we reproduce some of them in Figures 18.17–18.19. Alternatively, software (e.g., EMS) 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 which contains the reflection you want to use. To do this you'll have to be able to index the diffraction patterns and determine the beam direction, which we'll describe in detail in Chapter 18.

REFERENCES

General References

Andrews, K.W., Dyson, D.J., and Keown, S.R. (1971) *Interpretation of Electron Diffraction Patterns,* 2nd edition, Plenum Press, New York. Stereographic projections, angles, spacings, and much more.

Cullity, B.D. (1978) *Elements of X-ray Diffraction,* 2nd edition, Addison-Wesley, Reading, Massachusetts.

Massalski, T., Okamoto, H., Subramanian, P.R., and Kacprzak, L., Eds. (1990) *Binary Alloy Phase Diagrams,* 2nd edition, ASM International, Materials Park, Ohio. Appendix A1 gives a complete list of the Pearson Symbols ($L1_2$, etc.) with their space group and Strukturbericht designation.

Misell, D.L. and Brown, E.B. (1987) *Electron Diffraction: An Introduction for Biologists,* Volume 12 of the series *Practical Methods in Electron Microscopy* (Ed. A.M. Glauert), Elsevier, New York. Materials science students should not be put off by the title: this is an invaluable practical guide to indexing diffraction patterns and more.

Specific References

Crystal Kit, see Section 1.5.

Desktop Microscopist, see Section 1.5.

Dodsworth, J., Kohlstedt, D.L., and Carter, C.B. (1983) *Adv. Ceram.* **6**, 102.

Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science,* Van Nostrand Reinhold, New York.

Frank, F.C. (1965) *Acta Cryst.* **18**, 862.

Hahn, T. (1988) *International Tables for Crystallography.* Brief teaching edition of volume A, space-group symmetry, Kluwer Academic Publishers, Dordrecht, Netherlands.

Ibers, J.A. (1957) *Acta Cryst.* **10**, 86.

Jenkins, M.L., Katerbau, K.-H., and Wilkens, M. (1976) *Phil. Mag.* **34**, 1141.

Lee, W.E. and Lagerlof, K.P.D. (1985) *J. Electron Microsc. Tech.* **2**, 247 (errata in).

Diffraction from Small Volumes

17

CHAPTER PREVIEW

A very important concept in TEM is that we only diffract from small volumes. By definition, no specimens are 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 which 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 (~3 mm) relative to the unit-cell dimensions in the plane of the specimen, but very thin (< 0.5 μm) parallel to the electron beam. This means that the diffracted intensity can be represented in the reciprocal lattice as a rod stretched parallel to the electron beam in reciprocal space, rather than as a point, and the rod does have a width. Therefore, over a range of angles, the Ewald sphere will still intercept the rod 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.

Diffraction from Small Volumes

17

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 generally strongest when $\mathbf{K} = \mathbf{g}$, but we still have intensity when **K** is not exactly equal to **g**, or when

$$\mathbf{K} = \mathbf{g} + \mathbf{s} \qquad [17.1]$$

Then we can write, from equation 13.48

$$|\phi_{\mathbf{g}}|^2 = \left(\frac{\pi t}{\xi_{\mathbf{g}}}\right)^2 \frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} \qquad [17.2]$$

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 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, 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 we usually don't see the subsidiary maxima.

What we are going to do is derive equations for the shape of the relrods which were 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

$$A = F \sum_{n_x} e^{i2\pi n_x \mathbf{K}\cdot\mathbf{a}} \sum_{n_y} e^{i2\pi n_y \mathbf{K}\cdot\mathbf{b}} \sum_{n_z} e^{i2\pi n_z \mathbf{K}\cdot\mathbf{c}} \qquad [17.3]$$

Here n_x, n_y, and n_z have their usual meanings and all are integers. 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 $\mathbf{r}_n$

$$\mathbf{r}_n = n_x\mathbf{a} + n_y\mathbf{b} + n_z\mathbf{c} \qquad [17.4]$$

To simplify the first summation we set X equal to $e^{i2\pi\mathbf{K}\cdot\mathbf{a}}$, then each separate summation term is a geometric series, so we can sum the n_x terms as

$$S = \sum_{n_x=0}^{n_x=N-1} X^n = X^0 + X^1 + \cdots X^{N-1} = \frac{1 - X^N}{1 - X} \qquad [17.5]$$

(Remember, if you multiply S by X you get $S + X^N - X^0$ and $X^0 = 1$.)

Summing from $n_x = 0$ to $n_x = N_x - 1$ we obtain

$$\sum_{n_x=0}^{n_x=N-1} e^{i2\pi n_x \mathbf{K}\cdot\mathbf{a}} = \frac{1 - e^{i2\pi n_x \mathbf{K}\cdot\mathbf{a}}}{1 - e^{i2\pi \mathbf{K}\cdot\mathbf{a}}} \qquad [17.6]$$

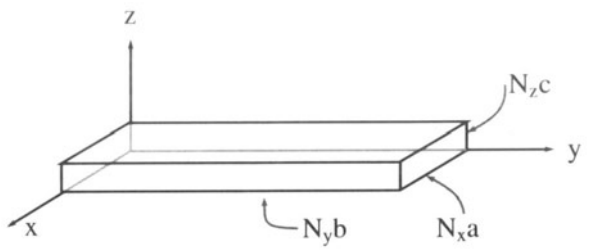


Figure 17.1. An idealized thin-foil specimen modeled as a rectangular slab made up of rectangular unit cells of sides a, b, c. There are N_x cells in the x direction, N_y in the y direction, and N_z in the z direction.

Since we are interested in the intensities, we multiply this sum by its complex conjugate. To do this we use some simple trigonometric relationships

$$\begin{aligned}&\left(1-e^{-i\alpha}\right)\left(1-e^{i\alpha}\right)\\&=\left(1-\cos\alpha+i\sin\alpha\right)\left(1-\cos\alpha-i\sin\alpha\right)\\&=\left(1-2\cos\alpha+\cos^2\alpha\right)+\sin^2\alpha\\&=2\left(1-\cos\alpha\right)=4\sin^2\frac{\alpha}{2}\end{aligned} \qquad [17.7]$$

The intensity is then related to

$$\left|\sum_{n_x=0}^{n_x=N-1} e^{i2\pi n_x \mathbf{K}\cdot\mathbf{a}}\right|^2 = \frac{1-\cos\left(2\pi N_x \mathbf{K}\cdot\mathbf{a}\right)}{1-\cos\left(2\pi\,\mathbf{K}\cdot\mathbf{a}\right)} \qquad [17.8]$$

Then we can write

$$\begin{aligned}&I=|A|^2\\&=|F|^2\left(\frac{\sin^2(\pi N_x\mathbf{K}\cdot\mathbf{a})}{\sin^2(\pi\,\mathbf{K}\cdot\mathbf{a})}\right)\left(\frac{\sin^2(\pi N_y\mathbf{K}\cdot\mathbf{b})}{\sin^2(\pi\,\mathbf{K}\cdot\mathbf{b})}\right)\left(\frac{\sin^2(\pi N_z\mathbf{K}\cdot\mathbf{c})}{\sin^2(\pi\,\mathbf{K}\cdot\mathbf{c})}\right)\end{aligned} \qquad [17.9]$$

If the dot product $\mathbf{K}\cdot\mathbf{a}$ 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

$$\pi N_x\mathbf{K}\cdot\mathbf{a}=\frac{\pi}{2}C \qquad [17.10]$$

where C = an integer. Reordering this equation, we have

$$\mathbf{K}\cdot\mathbf{a}=\frac{C}{2N_x} \qquad [17.11]$$

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.

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

$$\phi_{\mathbf{g}}=\frac{e^{2\pi i\mathbf{k}\cdot\mathbf{r}}}{r}\sum_n F_n e^{\left(-2\pi i\,\mathbf{K}\cdot\mathbf{r}_n\right)} \qquad [17.12]$$

Since we have defined $\mathbf{K}$ to be $\mathbf{g}+\mathbf{s}$, we can rewrite this equation as

$$\phi_{\mathbf{g}}=\frac{e^{2\pi i\mathbf{k}\cdot\mathbf{r}}}{r}\sum_n F_{\mathbf{g}} e^{\left(-2\pi i\,(\mathbf{g}+\mathbf{s}_{\mathbf{g}})\cdot\mathbf{r}_n\right)} \qquad [17.13]$$

Now we know that $\mathbf{g}\cdot\mathbf{r}_n$ 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

$$\phi_{\mathbf{g}}=\frac{e^{2\pi i\mathbf{k}\cdot\mathbf{r}}}{r}\sum_n F_{\mathbf{g}} e^{\left(-2\pi i\,\mathbf{s}\cdot\mathbf{r}_n\right)} \qquad [17.14]$$

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

$$\phi_{\mathbf{g}}=\frac{e^{2\pi i\mathbf{k}\cdot\mathbf{r}}}{r\,V_c}F_{\mathbf{g}}\int_{\text{crystal}} e^{\left(-2\pi i\,\mathbf{s}\cdot\mathbf{r}_n\right)}dv \qquad [17.15]$$

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

$$\mathbf{r}_n=h\mathbf{a}+k\mathbf{b}+\ell\mathbf{c} \qquad [17.17]$$

then we can write

$$\phi_{\mathbf{g}}=\frac{e^{2\pi i\mathbf{k}\cdot\mathbf{r}}}{r\,V_c}F_{\mathbf{g}}\int_0^C\int_0^B\int_0^A e^{\left(-2\pi i\,(ux+vy+wz)\right)}dx\,dy\,dz \qquad [17.18]$$

where $A=N_x a$, etc. This integral is straightforward

$$\begin{aligned}\int_0^A e^{-2\pi iux}&=\frac{e^{-2\pi iuA}-1}{-2\pi iu}=\left(\frac{e^{-\pi iuA}}{\pi u}\right)\left(\frac{e^{\pi iuA}-e^{-\pi iuA}}{2i}\right)\\&=\frac{e^{-\pi iuA}}{\pi u}\sin(\pi uA)\end{aligned} \qquad [17.19]$$

$$\phi_{\mathbf{g}}=\frac{e^{2\pi i\mathbf{k}\cdot\mathbf{r}}}{r\,V_c}F_{\mathbf{g}}\frac{(\sin\pi Au)}{(\pi u)}\frac{(\sin\pi Bv)}{(\pi v)}\frac{(\sin\pi Cw)}{(\pi w)}e^{iD} \qquad [17.20]$$

(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 in equation 2.12 for the 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 **s**; it is not a constant for any position along the rod.

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.

Just remember that when we said intensity in the last sentence, 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.

In other words, 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 at different positions relative to the square array which is the projection of the spots at zero tilt (Figure 17.3B). The DP will appear as shown in Fig-

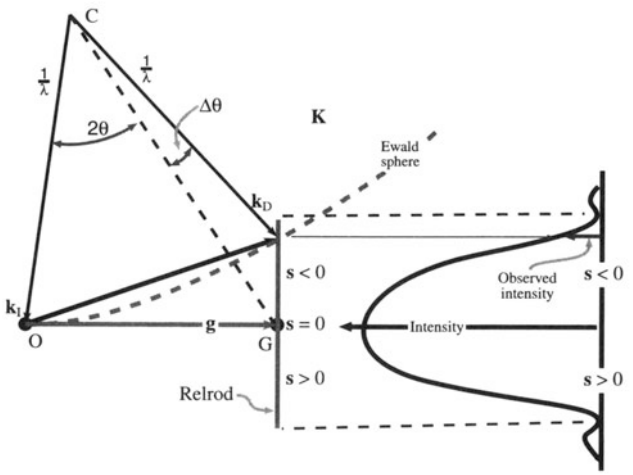


Figure 17.2. The relrod at $\mathbf{g}_{hk\ell}$ when the beam is $\Delta\theta$ away from the exact Bragg condition. The Ewald sphere intercepts the relrod at a negative value of **s** which defines the vector $\mathbf{K} = \mathbf{g} + \mathbf{s}$. The intensity of the diffracted beam as a function of where the Ewald sphere cuts the relrod is shown on the right of the diagram. In this case the intensity has fallen almost to zero.

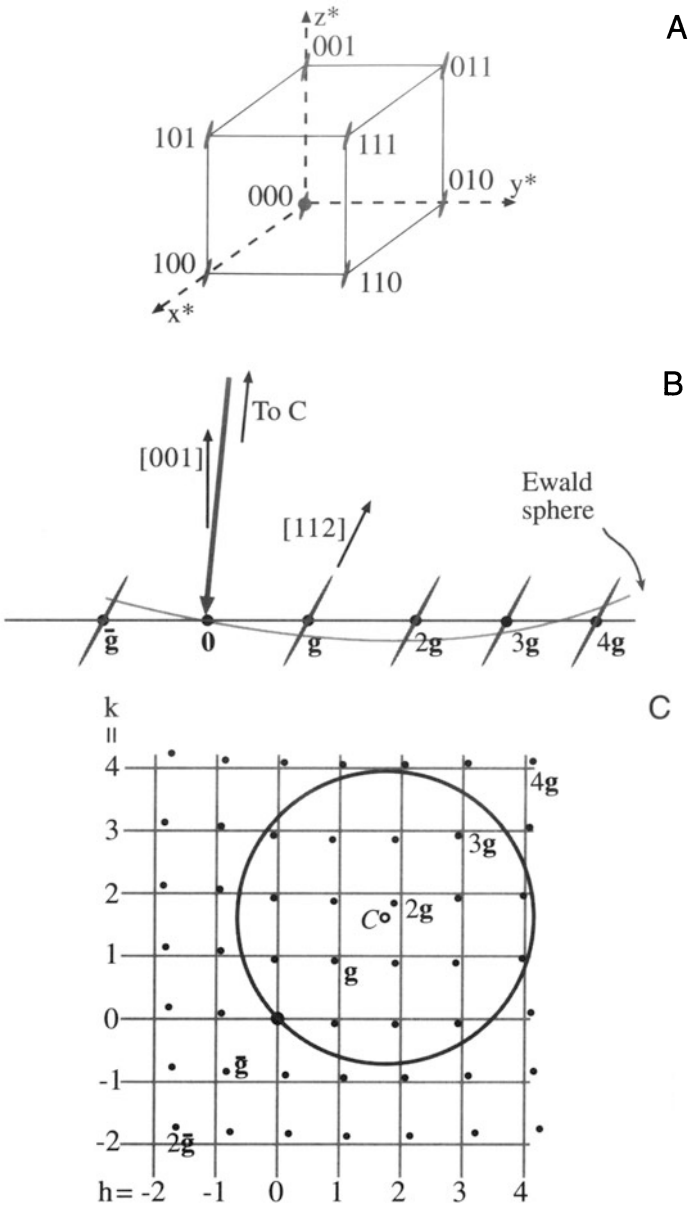


Figure 17.3. (A) For a thin specimen, every point is replaced by a relrod. (B) The Ewald sphere cutting the relrods in (A) when the crystal is tilted slightly off the 001 axis. (C) The effect of the tilt in (B) on the DP. Notice that all of the spots in the DP are displaced relative to their positions 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 still be the "correct" distance from 000.

ure 17.3C. In Figure 17.3C, C is the projected position of the center of the 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 ϕ_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 ϕ_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 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 Figures 17.4C, 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 strong dynamical-diffraction condition. We will return to this point in the next section and again in Chapter 24.

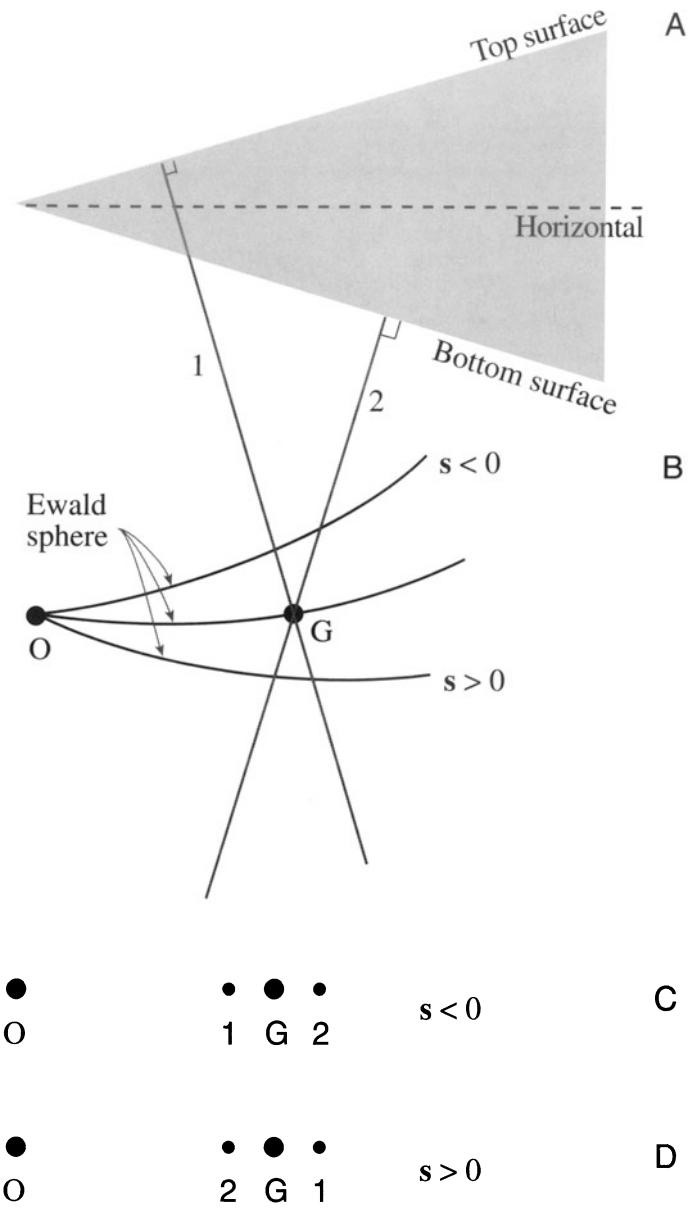


Figure 17.4. (A) Diffraction from a wedged crystal. (B) Notice that when $\mathbf{s} < 0$, relrod 1 is on the left of relrod 2 but the order reverses when **s** becomes >0. The effect of this pair of relrods is to create a doublet shown in (C) and (D). The middle spot is the matrix relrod.

17.4. DIFFRACTION FROM PLANAR DEFECTS

The shape factor concept can be readily applied to understand diffraction from a flat platelet or planar fault, such as the geometry shown in Figure 17.5. The idea is that the platelet is itself a thin parallelepiped which is inclined to

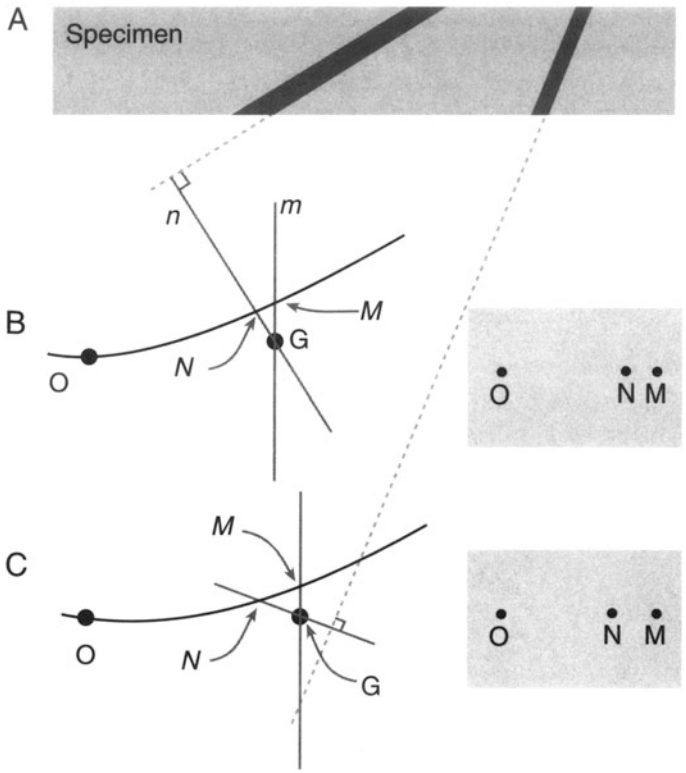


Figure 17.5 The effect of a thin inclined plate in a thin specimen. (A) Two plates are shown to illustrate the effect of changing the inclination of the plate relative to the foil surface. When $\mathbf{s} \neq 0$ we see two spots in the DP because there are two relrods for the two different planar-defect inclinations in (B) and (C).

the 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). 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 **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, 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 Section 17.3, we actually see two spots when **s** = 0, and 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.

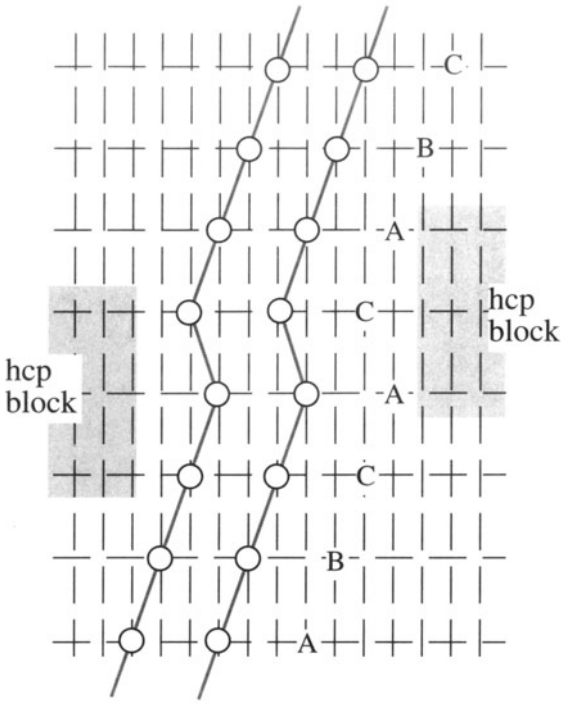


Figure 17.6 Schematic of the stacking sequence of close packed planes A, B, C, in an fcc crystal showing that the SF is similar to a thin layer of hcp material, stacking ACA.

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 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.

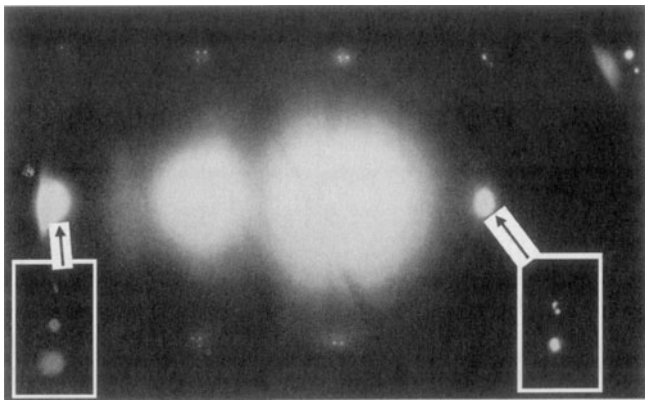


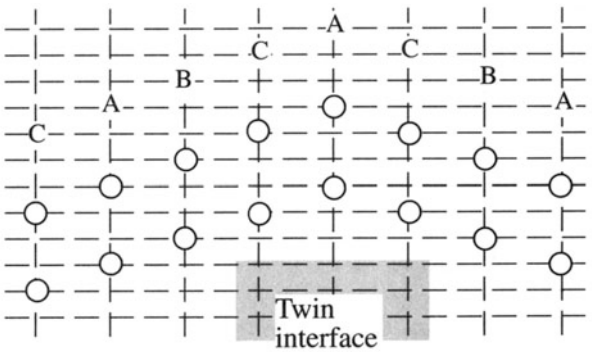
Figure 17.7. Pairs of spots in a DP from a grain boundary.

- 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 **0**, 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 you can see in Figure 17.8B, there is a "streak" in the DP rather than a spot. The streak actually extends in the [111] direction because, as you can appreciate from Figure 17.8A, the twin is a *very* thin platelet.

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.

A



B

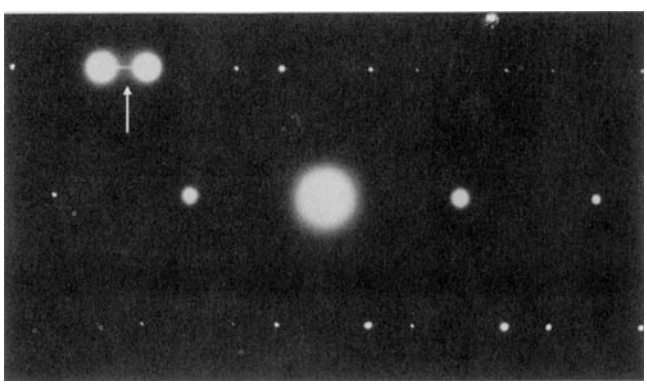


Figure 17.8. (A) Schematic of twin and (B) DP with a streak (arrowed) normal to the twin plane. Note that **s** = 0 for the two bright diffracted spots.

17.5. DIFFRACTION FROM PARTICLES

Particles come in all shapes and sizes, so we will not try to be exhaustive. Actually, the principles involved in determining the shape factor in reciprocal space are 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 maxima shown in these diagrams.

One example, which is common, is the platelets shown in Figure 17.10; these can occur as GP zones or other 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

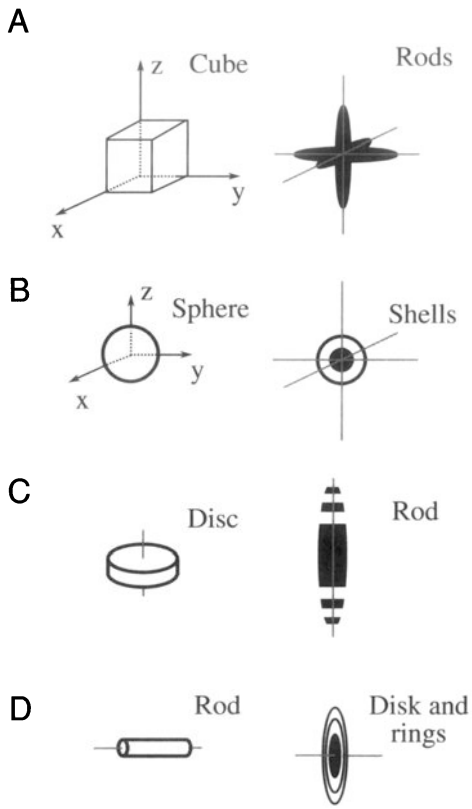


Figure 17.9. Examples of how spots in reciprocal space have different shapes, depending on the shape of the particles which are diffracting.

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 (1972) 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 solid-state physics. We will discuss some aspects of imaging using diffusely scattered electrons in Section 31.4, but the important points to recognize are:

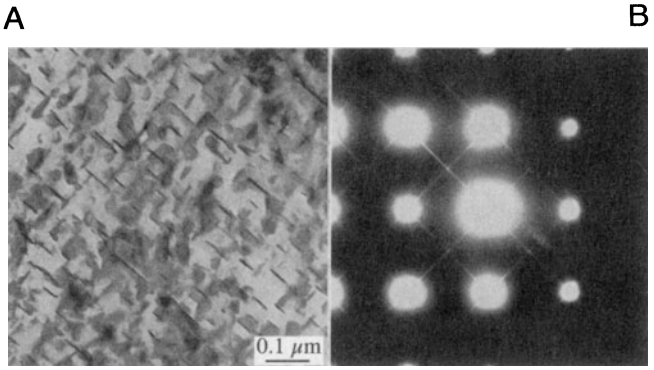


Figure 17.10. Very thin plate-like precipitates (A) cause long streaks in the DP (B). In this example, the precipitates are GP zones in an Fe-2.9 at % Mo alloy.

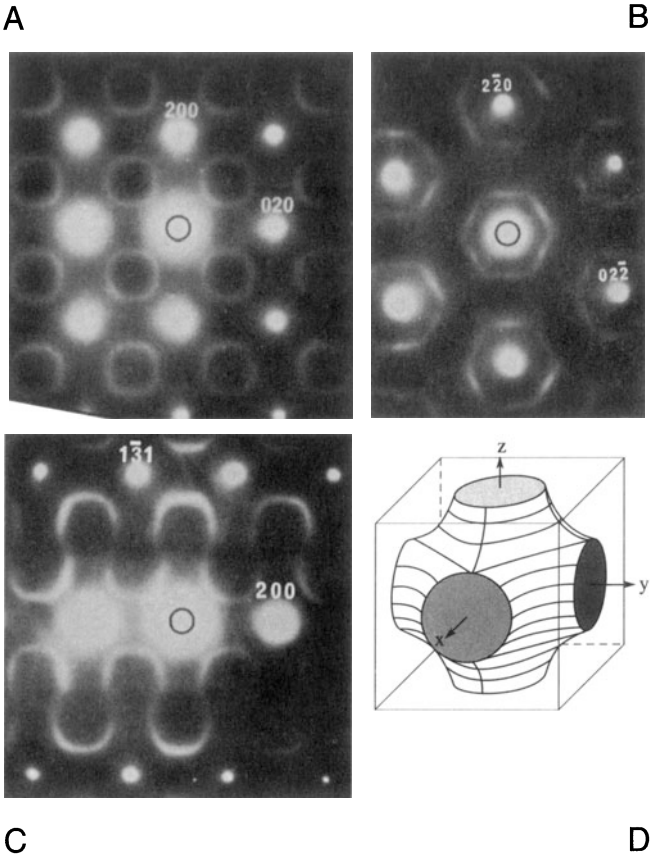


Figure 17.11. Short-range ordering can cause diffuse scattering in the DP (A–C). The DPs in this example were obtained from a vanadium carbide. In this case, the 3D map of diffuse intensity has a shape which strongly resembles a Fermi surface, shown in (D).

- Point defects can cause diffraction effects, especially if they interact with one another.
- Diffuse scattering can still be interpreted by the Ewald sphere construction.

If you are intrigued by this topic, you will find the literature on discommensurate structures in intercalated material a complementary challenge (Wilson *et al.* 1975). A library search on "discommensurate" and "intercalated" should quickly net any 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 which 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 2 Å to 10 Å around the core is greatly distorted (we'll see the effect of this strain in Chapter 25), then the diffuse scattering will extend from 0.1 $Å^{-1}$ to perhaps 0.5 $Å^{-1}$ 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.

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.

This diffuse intensity 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. 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 SAD patterns 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 of spots is that **s** is 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 grains are still present) but now there are two extra spots. The reason you 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

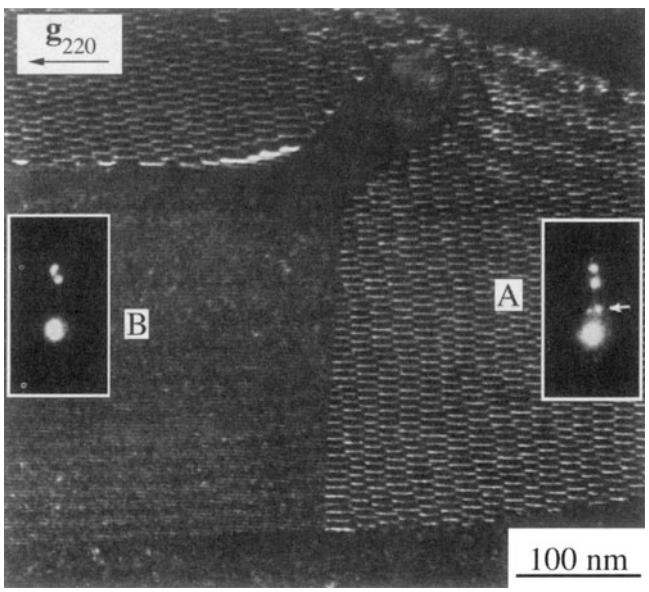


Figure 17.12. Diffraction from an ordered array of dislocations. Dislocations are present in region A, but not in B. The insets show a small part of the DP from the two regions. The extra spots arrowed in A are caused by the visible array of dislocations; these spots are a doublet because there is also 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.

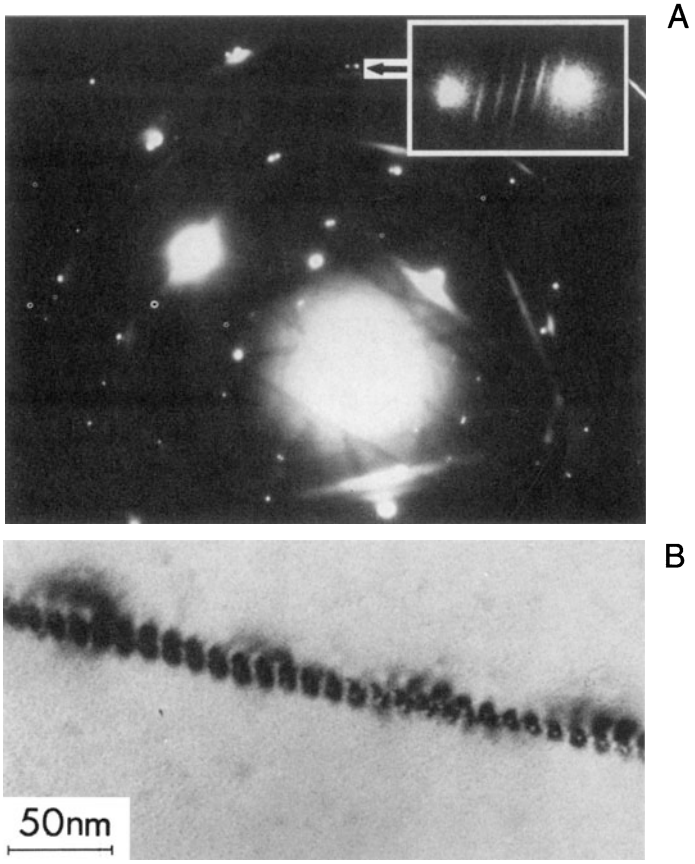


Figure 17.13. (A) The set of streaks from an array of dislocations in Al_2O_3 lying parallel to the electron beam. The distance between the streaks is inversely related to the spacing of the dislocations shown in the image (B).

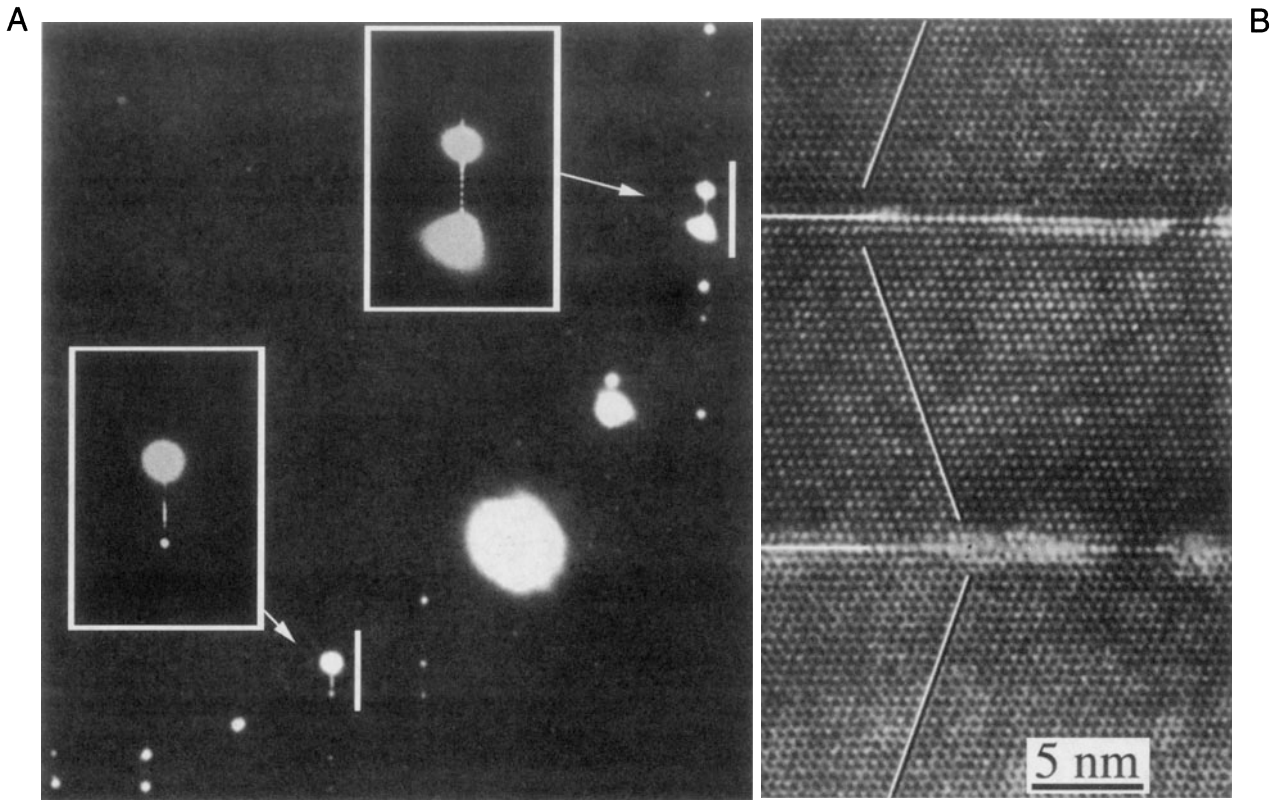


Figure 17.14. Extra spots can be formed in the DP (A) when only two defects are scattering in phase. The separation of the extra spots is related to the inverse of the separation of the two twin boundaries seen in the image (B).

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.

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 general rule is:

> 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 insets) is 0.067 nm^{-1}, as expected. Now, why can this occur? The analogy is Young's slits experiment in visible-light optics (Carter 1984). The illustration also reminds us of a special feature of the TEM, namely, that even without an FEG, the electron beam is remarkably coherent.

17.7. DIFFRACTION AND THE DISPERSION SURFACE

Several times in this chapter, we have said "actually, you 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 (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. For a full introduction, we recom-

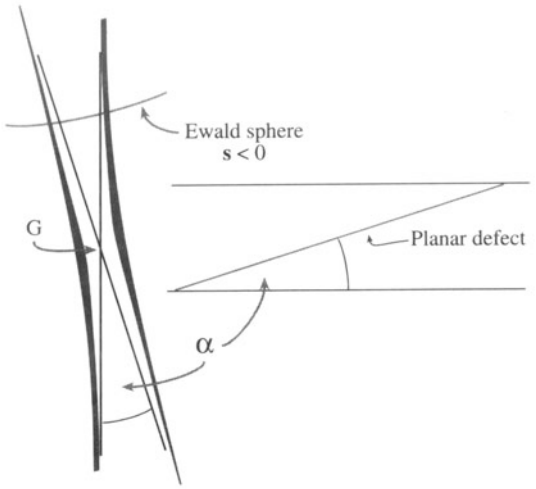


Figure 17.15. The relrods from two planes inclined at angle α are actually asymptotic to two straight lines, so that they don't cross at G; when **s** = 0, the distance between these two curves is ξ_g^{-1}.

mend the review by Gevers (1971). 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 asymptotic to two straight lines, as shown in Figure 17.15. When the Ewald sphere cuts these curves at **s** = 0, we see that there are two spots which move apart as we increase **s** (either positive or, as shown here, negative) until they are at the points defined by the straight lines. So, will there be a vector that exactly corresponds to **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 **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 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 **s** = 0, the distance between the curves is inversely proportional to ξ_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 (ξ_g is just Δk^{-1} at **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 and 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 you must take into account the shape effect. 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, you can arrange that the Ewald sphere cuts along the length of the relrod. Now you'll see a "streak" in the DP rather than a spot.
- Beam splitting at **s** = 0 and the dispersion surface both arise because of dynamical scattering.

REFERENCES

General References

The relation between diffraction and images from planar defects has been the subject of a long series of papers. The following will give you a start to your study.

de Ridder, R., Van Landuyt, J., Gevers, R., and Amelinckx, S. (1968) *Phys. stat. sol.* **30**, 797; (1970) *ibid.* **38**, 747; (1970) *ibid.* **40**, 271; (1970) *ibid.* **41**, 519; (1970) *ibid.* **42**, 645.

Gevers, R., Van Landuyt, J., and Amelinckx, S. (1966) *Phys. stat. sol.* **18**, 343; (1967) *ibid.* **21**, 393; (1967) *ibid.* **23**, 549; (1968) *ibid.* **26**, 577.

Van Landuyt, J. (1964) *Phys. stat. sol.* **6**, 957.

Van Landuyt, J., Gevers, R., and Amelinckx, S. (1966) *Phys. stat. sol.* **13**, 467; (1966) *ibid.* **18**, 167.

Specific References

Carter, C.B. (1984) *Phil. Mag.* **A50**, 133.

Gevers, R. (1971) in *Electron Microscopy in Materials Science* (Ed. U. Valdrè), p. 302, Academic Press, New York.

Sauvage, M. and Parthé, E. (1972), *Acta Cryst.* **A28**, 607.

Wilson, J.A., Di Salvo, F.J., and Mahajan, S. (1975) *Adv. Phys.* **24**, 117.

Indexing Diffraction Patterns

18

CHAPTER PREVIEW

Since the strength of TEM is that you can obtain both crystallographic data and an image from the same part of your specimen, a method for interpreting the DP is essential. The first step in any interpretation is to index your pattern. You can proceed 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.

Using the DP, we can identify the crystal (which we often already know) and its orientation. The positions of the allowed $hk\ell$ reflections are characteristic of the crystal system. Indexing associates each spot in the DP with a plane, $(hk\ell)$, 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. This direction is normal to the plane of the DP and antiparallel to the electron beam. It is convention to define $[UVW]$ as the beam direction. If you want to know the orientation relationship between two crystals, you need to know more than one $[UVW]$ for each crystal.

Indexing Diffraction Patterns

18

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 structure of a particular region, you may find moiré fringes (Chapter 27) or HRTEM (Chapter 28) more appropriate. We can summarize the possibilities as a function of the grain size of the material. Basically there are two approaches:

- You can focus the beam on a small area of your specimen to form a convergent-beam electron-diffraction (CBED) pattern (see Chapters 20 and 21).
- 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 (the SAD patterns of Chapters 9 and 11).

Let's consider the specimen characteristics:

- The grain size may be very small, ≤10 nm. This is a problem! 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. It is also probable that the electron beam will pass through several such grains in a typical thin specimen, in which case you can't analyze an individual grain.
- The grain size is between 10 nm 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 specimen thicknesses which are > 100 nm; this thickness depends on the structure factor (atomic number) of your specimen. With modern TEMs C_s and λ are so small that you might be able to use SAD in this range if you're careful, 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 nm to 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 pattern (EBSP) technique in an SEM (Randle 1993).

In this chapter we'll concentrate on the hands-on approach to SAD analysis and leave CBED to Chapters 20 and 21. We can't give you a foolproof guide since your technique will depend on your specimen.

18.2. EXPERIMENTAL TECHNIQUES

By now you should know how the experimental camera length (L) compares to the value you read from the micro-

scope. You also know how the pattern 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 and 11.10 if you need more details on the practical steps involved in obtaining SAD patterns.

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 microscope, 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 SAD patterns 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.

Tilting and rotating the 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, 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 31). 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 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. Set up a particular excited beam and tilt the specimen so that one particular beam remains excited. What you are doing is tilting the specimen so that the same plane remains nearly parallel to the electron beam.

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 very precise and there is no problem with backlash. To increase your accu-

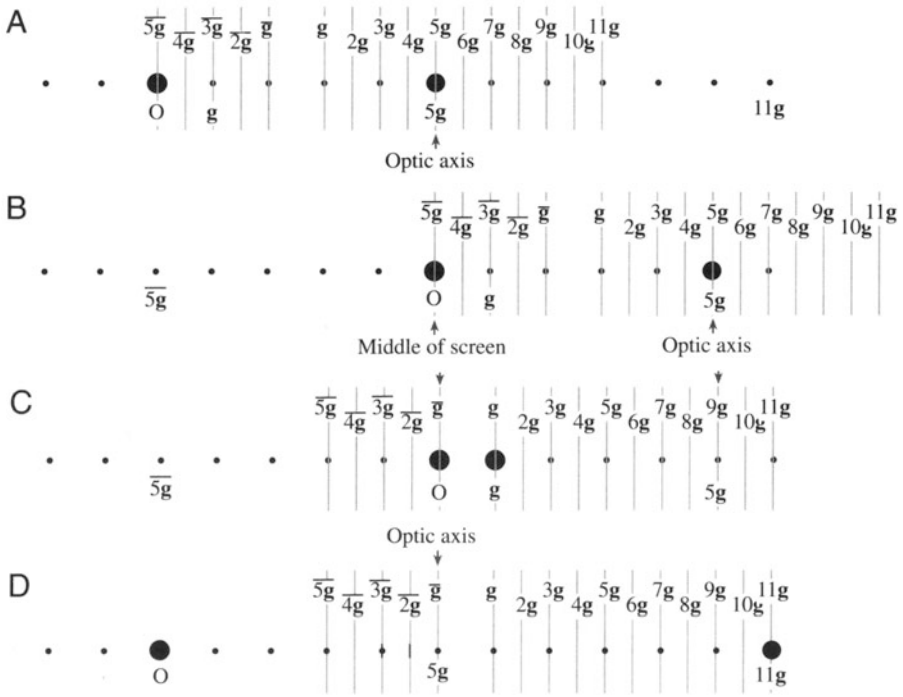


Figure 18.1. The steps used to excite a high-order reflection. (A) Tilt the beam so **5g** is on axis and strongly excited. (B) Translate the pattern so O is in the middle of the screen. (C) Tilt the specimen to excite **g** (the Kikuchi lines move). (D) Tilt the beam so that **5g** is back on axis and **11g** is strong.

racy, you may want to increase L. This technique is particularly helpful when you want to examine the effect of small changes in **s** on the appearance of diffraction spots. For example, if you want to excite the eleventh-order reflection 11**g** but you can't see 11**g**, use the following approach, shown in Figure 18.1:

- Tilt the beam using the dark-field deflection coils so that 5**g** is on axis (Figure 18.1A).
- Use the translation coils on the projector lens to bring the direct beam back to the middle (so that you can see it easily) (Figure 18.1B).
- Tilt the specimen so that **g** is excited (Figure 18.1C).
- Then tilt the beam with the DF coils and translate the pattern with the projector lens so that 5**g** is back on the optic axis; 11**g** should now be excited (Figure 18.1D).

We'll develop other variations of this technique in Chapter 19, and we'll see in Chapter 26 that the situation in Figure 18.1 does arise in weak-beam microscopy at higher voltages.

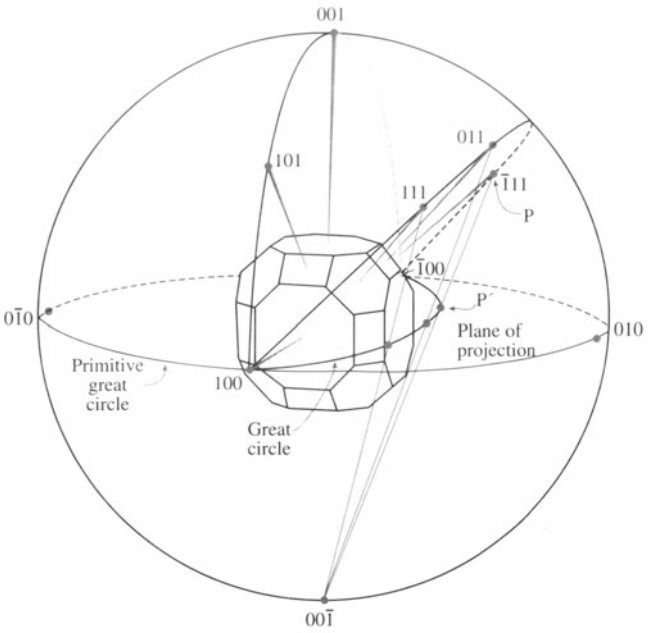


Figure 18.2. The stereographic projection. The crystal is at the center of the sphere. Normals to the crystal planes are projected until they intercept the sphere at P, then projected back to the south pole $(00\bar{1})$ of the sphere. Where this projected line crosses the equatorial plane at P' is the point that uniquely represents the original plane on the crystal. Note that planes in the same zone on the crystal project as a line of longitude on the sphere, called a great circle, and project as the arc of a circle on the equatorial plane, whose circumference is called the primitive great circle.

18.3. THE STEREOGRAPHIC PROJECTION

Diffraction patterns 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 present in the specimen we can determine the orientation of both grains and the plane of the interface. What we often want to know is how the two grains are related to one another. We need a method for visualizing this relationship; this is where the stereographic projection is an invaluable aid (Johari and Thomas 1969). 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 general 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 line from the south pole to point P.

This line cuts the disk containing the equator at the point P'. The disk 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, 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 $(\bar{1}11)$ and 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 a great circle on the stereographic projection whose circumference we call the "primitive" great circle. The Wulff net shown 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 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. Bet-

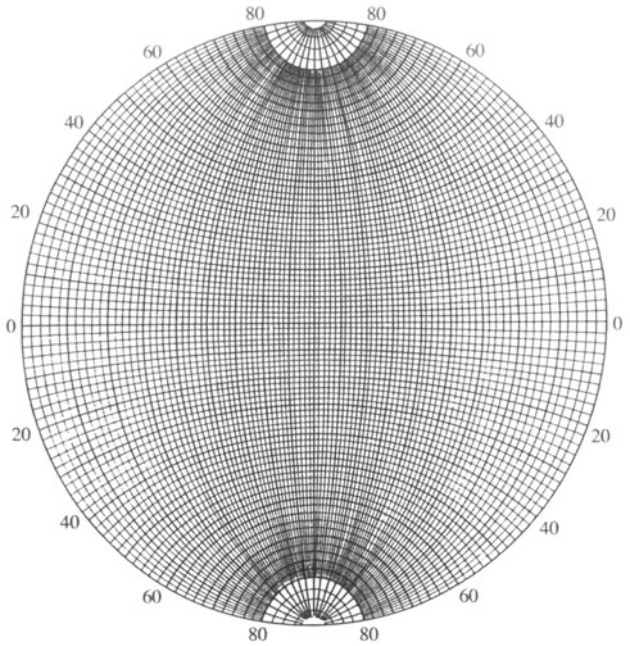


Figure 18.3. A Wulff net which contains 90 great circles like the one in Figure 18.2; each great circle is 2° apart so the net covers 180°. The only great circle that actually appears as a circle in the net is the circumference of the projection, called the primitive great circle. Points on this circle represent planes whose normals are 90° from the north pole of the sphere (which projects in the center of the Wulff net). Therefore, all distances on the net are proportional to angles in real space but only correspond exactly to angles around the primitive great circle.

ter still, we can read off the angles between them. Remember, in general, that the normal to the plane $(hk\ell)$ is parallel to the direction $[hk\ell]$ only for cubic materials.

- The zone axis is always 90° away from any plane normal in its zone. All the plane normals in a particular zone, $[UVW]$, will lie on a single great circle; these are the possible diffracting planes for that zone (i.e., the $[UVW]$ beam direction). So if $[UVW]$ is in the center of the projection, the $hk\ell$ reflections will be around the circumference of the projection (the primitive 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.

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 $[\bar{0}01]$ 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, tilt toward the $[\bar{1}10]$ zone axis with the 220 reflection excited, not toward the $[1\bar{1}0]$ 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 noncubic material, buy a large Wulff net and construct your own stereographic projection; you can buy standard projections for cubic materials. Use a program like Desktop Microscopist or EMS to help you

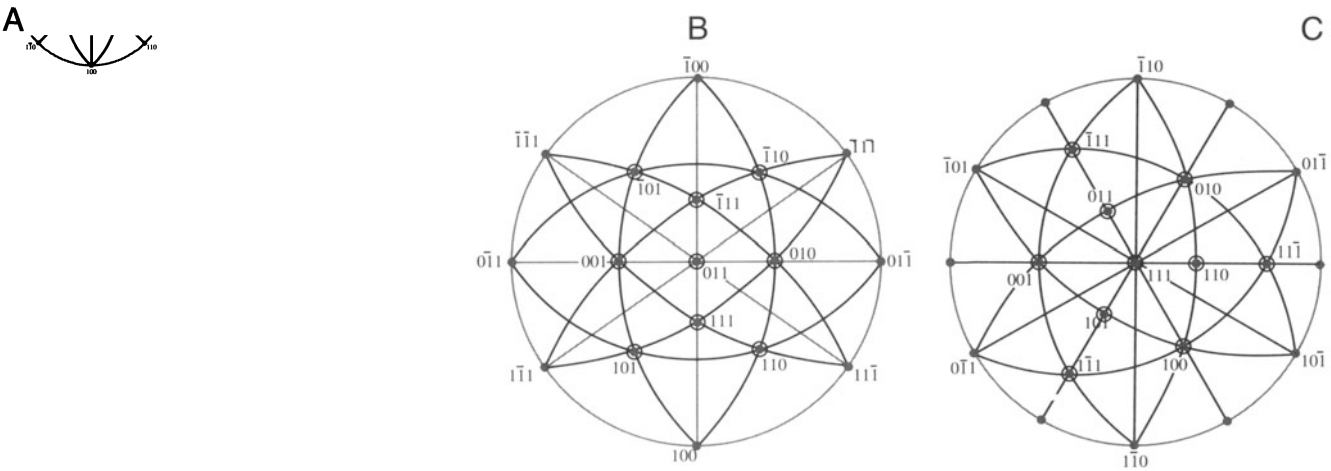


Figure 18.4. Some standard cubic stereographic projections. The pole in the center defines each projection, so these are 001, 011, and 111.

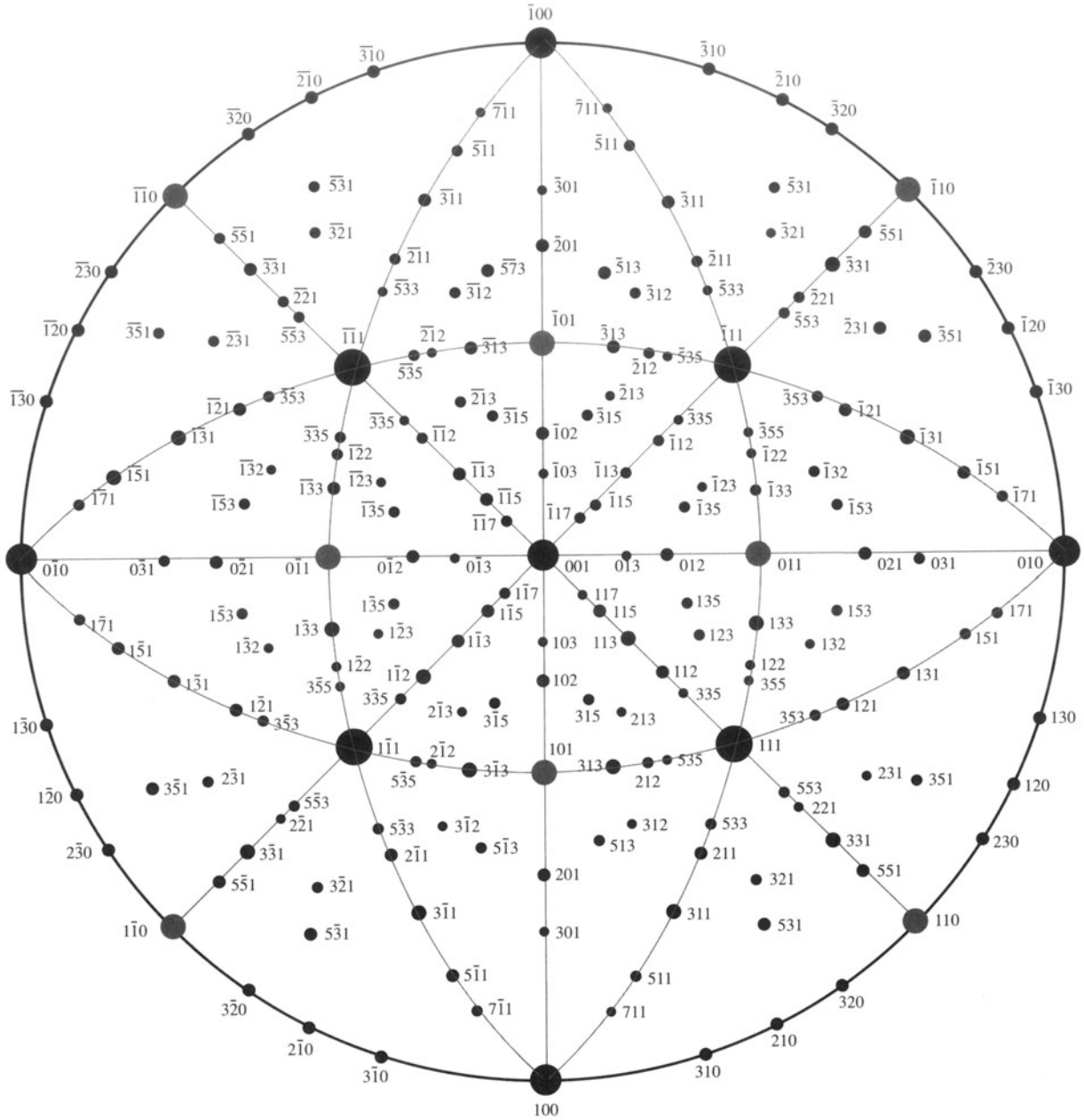


Figure 18.5. The stereographic projection for a cubic foil with a [001] normal, assuming the beam is down [001] also. If you want to form an image with the $0\bar{2}2$ reflection, you need to tilt the specimen so the $0\bar{1}1$ pole rotates until it is on the primitive, i.e., it is 90° from the beam direction. To do this you need to tilt about an axis that is 90° from the $0\bar{2}2$ reflection, such as the [100], [111], [311], zone axes.

plot the points, or download appropriate freeware from the WWW.

18.4. INDEXING SINGLE-CRYSTAL DIFFRACTION PATTERNS

Remember the fundamental relationship in a DP (refer back to Chapter 9)

$$Rd = \lambda L \qquad [18.1]$$

Any distance, R, which we measure on the DP is related to a specific spacing in the crystal, d. Since λL is a constant, we can measure several values of R and know that

$$R_1 d_1 = R_2 d_2 = R_3 d_3 = R_4 d_4 = \cdots \qquad [18.2]$$

Therefore the ratio of any two R values gives the ratio of the d-spacings. If you know the lattice parameter of your

Table 18.1. The Selection Rules for Cubic Crystal Structures

bcc		fcc		Diamond cubic	
$h^2+k^2+\ell^2$	$hk\ell$	$h^2+k^2+\ell^2$	$hk\ell$	$h^2+k^2+\ell^2$	$hk\ell$
2	110				
		3	111	3	111
4	200	4	200	4	200
6	211				
8	220	8	220	8	220
10	310				
		11	311	11	311
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

crystal, then you know the allowed reflections and only certain d-spacings will be associated with diffraction spots. Table 18.1 illustrates allowed and forbidden reflections for some cubic systems. Rules for more crystal systems are given back in Table 16.2.

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 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 particular zone-axis DP. However, if your specimen is not oriented close to a zone axis, you'll need to look ahead to Section 18.8.

> Remember to check the consistency of your indexing using the Weiss zone law. Each $hk\ell$ reflection must lie in the $[UVW]$ zone, i.e., $hU + kV + \ell W = 0$.

The angle between normals to the planes $(h_1k_1\ell_1)$ and $(h_2k_2\ell_2)$ 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., Edington (1976) and Andrews *et al.* (1971). You'll probably find that the most useful are the equations for the cubic system

$$\cos\phi = \frac{h_1h_2 + k_1k_2 + \ell_1\ell_2}{\left(h_1^2+k_1^2+\ell_1^2\right)^{\frac{1}{2}}\left(h_2^2+k_2^2+\ell_2^2\right)^{\frac{1}{2}}} \quad [18.3]$$

$$\cos\rho = \frac{U_1U_2 + V_1V_2 + W_1W_2}{\left(U_1^2+V_1^2+W_1^2\right)^{\frac{1}{2}}\left(U_2^2+V_2^2+W_2^2\right)^{\frac{1}{2}}} \quad [18.4]$$

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, $d_{hk\ell}$, 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.

To summarize, tilt your specimen to a low-index pole, set $\mathbf{s} = 0$ for the innermost reflections, and record the SAD pattern. 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 relation-

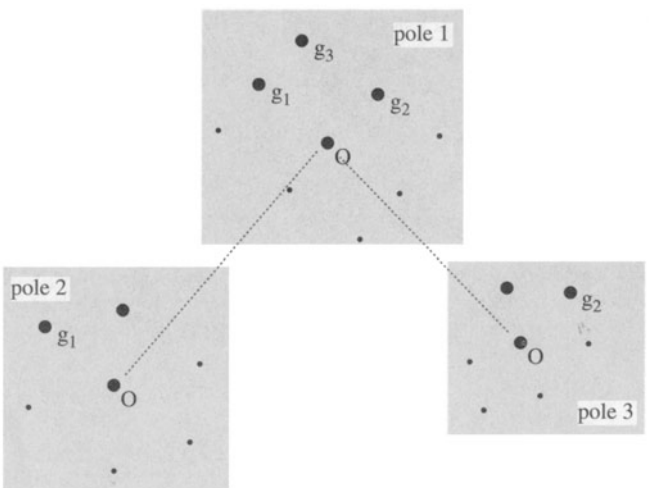


Figure 18.6. How to confirm your indexing of reflections and poles by tilting to other poles: Start with $\mathbf{g}_1$ and $\mathbf{g}_2$ strongly excited at pole #1. Tilt to pole #2, keeping $\mathbf{g}_1$ strong, then go back to pole #1 and tilt to pole #3, keeping $\mathbf{g}_2$ strong. Index all the strong reflections each time, measure the tilt angles between each reflection, and estimate the tilt between poles.

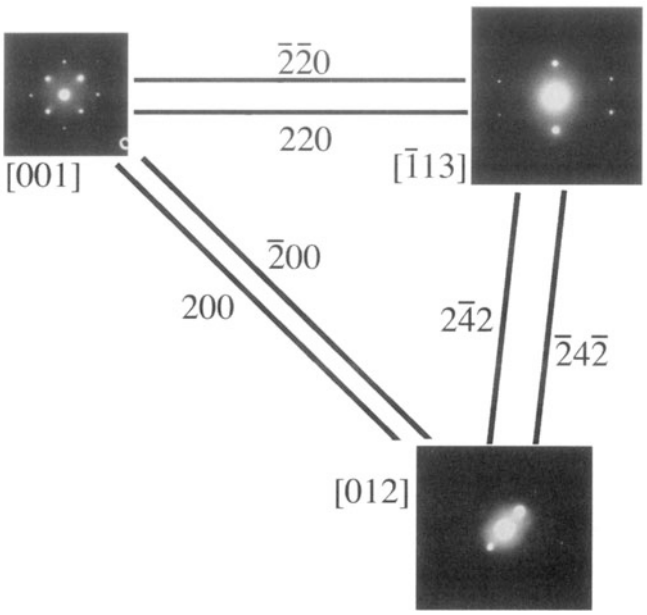


Figure 18.7. A practical illustration of the procedure described in Figure 18.6 for an fcc material, in this case MgO.

ships. While you're at the microscope, tilt to pole #2, keeping $\mathbf{g}_1$ (see Figure 18.6) strongly excited. Repeat the 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 18.9 and Chapter 19), and pay particular attention to the information from systematic absences as occur when the structure factor is zero.

The golden rule is: make the task as easy as possible while you're at the microscope. Record all the DPs you might need during your TEM session.

18.5. RING PATTERNS FROM POLYCRYSTALLINE MATERIALS

Diffraction from polycrystalline specimens 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 that are recorded in powder patterns.

If the polycrystal is textured, then there is usually one special plane which is nearly common to all the grains. We then rotate the reciprocal lattice about the lattice vector normal to this plane and produce a set of circles in reciprocal space, as shown in Figure 18.8. If we are examining cubic materials, the reciprocal lattice vector $\mathbf{g}_{hk\ell}$ will be parallel to the direction $[hk\ell]$ in real space. Otherwise, this will not generally be the case.

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.

The DP in either of these examples appears as shown in Figures 18.9A and B, which differ because the grain 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 the specimen is textured, the rings become arcs as shown 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 DF image with the arc. In the example shown in Figure 18.10C, these

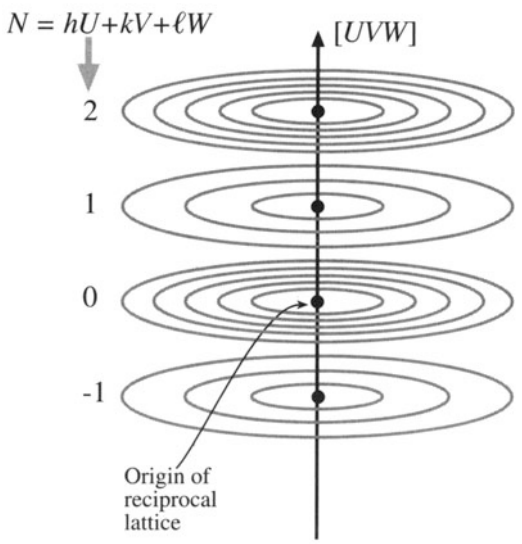
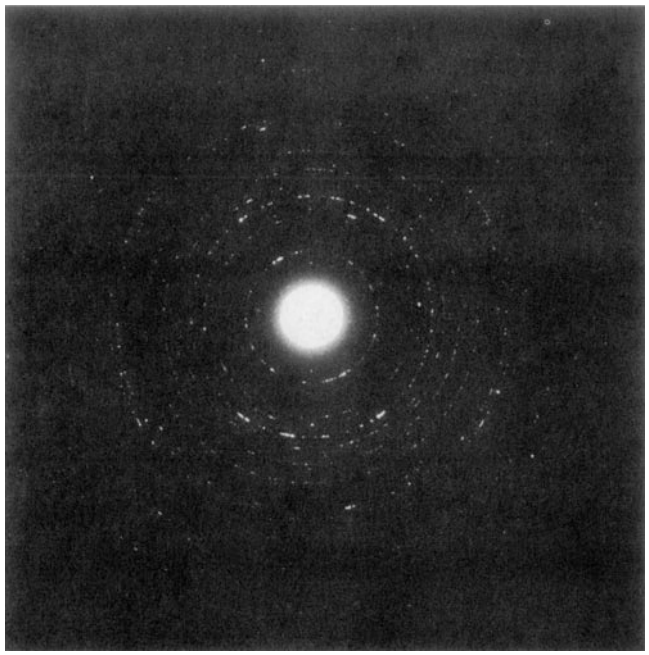


Figure 18.8. The generation of a set of circles in reciprocal space by a textured polycrystal. When the reciprocal lattice is rotated about a particular direction $[UVW]$ (in this case the normal to the texture plane), each Laue zone (N = 1, 2, etc.) produces a set of concentric circles for each allowed reflection in each zone.

A



B

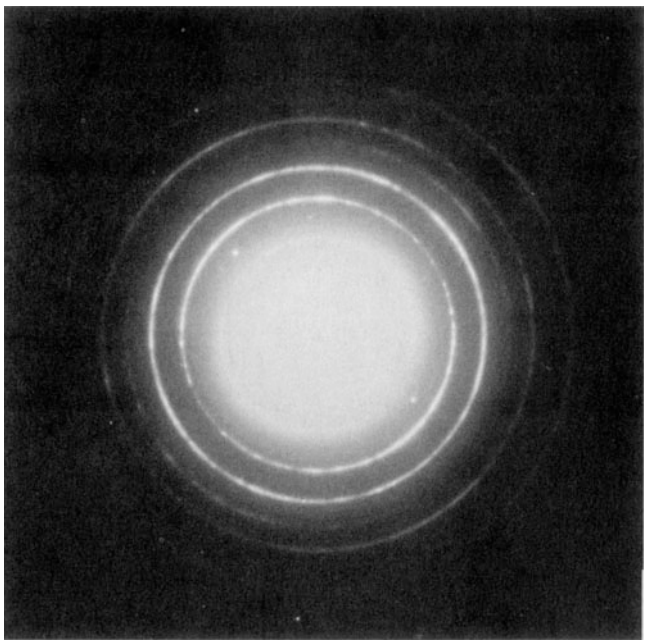


Figure 18.9. Ring diffraction patterns from polycrystalline foils. In (A) the grain size is larger than in (B), so the rings are made up of discrete spots. A finer grain size, as in (B), produces a more continuous ring pattern, but the widths of the rings of diffracted intensity in fact become broader and can be used as an inverse measure of the grain size.

oriented grains are uniformly distributed, but you might encounter a situation where this is not the case.

Figures 18.10D and E emphasize that these patterns can be quite varied. In this case, the specimen is α-Ag_2Se, which is textured about an axis that is *inclined* to the beam. When the Ewald sphere cuts the circles now, it produces elongated spots which lie on an ellipse. Vainshtein *et al.* (1992) 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 (OTEDP). You should also be careful in indexing these textured patterns since not all possible $d_{hk\ell}$ values need be present, depending on the texture plane.

There is more information in these patterns. Like a powder pattern you could estimate the grain size from the width of the rings, but it's more direct to just look at the DF image. You can see kinematically forbidden rings because you don't necessarily have single scattering from each grain.

One challenging application of these patterns concerns nanocrystalline materials, which fall into our smallest range of ~10-nm grain size. Careful DF imaging combined with HRTEM is probably optimal, but you need to look for clustering of similarly oriented grains.

18.6. RING PATTERNS FROM AMORPHOUS MATERIALS

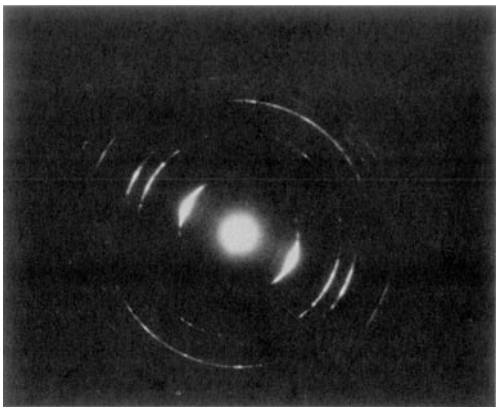
The first question you have to answer on this topic is one of the most difficult, namely, is the material really amorphous or is it (sub)nanocrystalline? Actually, this question is still debated when discussing both amorphous materials and, more intensely, oxide and metallic glasses.

> The DP from an amorphous material looks similar to that from polycrystalline material, but the rings are broader and there is no speckle.

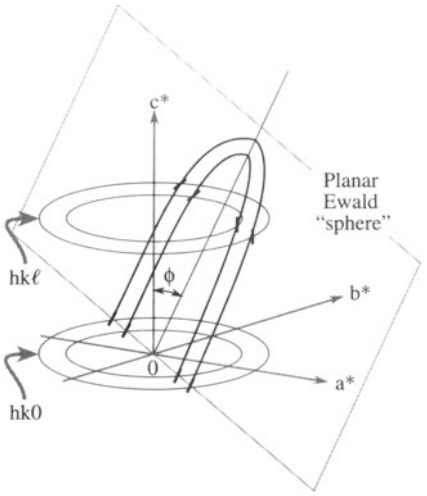
Rudee and Howie (1972) showed that electron scattering from regions of ≤15-Å diameter could be coherent. Graczyk and Chaudhari (1973) proposed modeling these materials as random networks. If we are careful, we can learn quite a lot about the structure of amorphous materials, but we should first say what we mean by "amorphous".

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 portion 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 diffraction pattern (E).

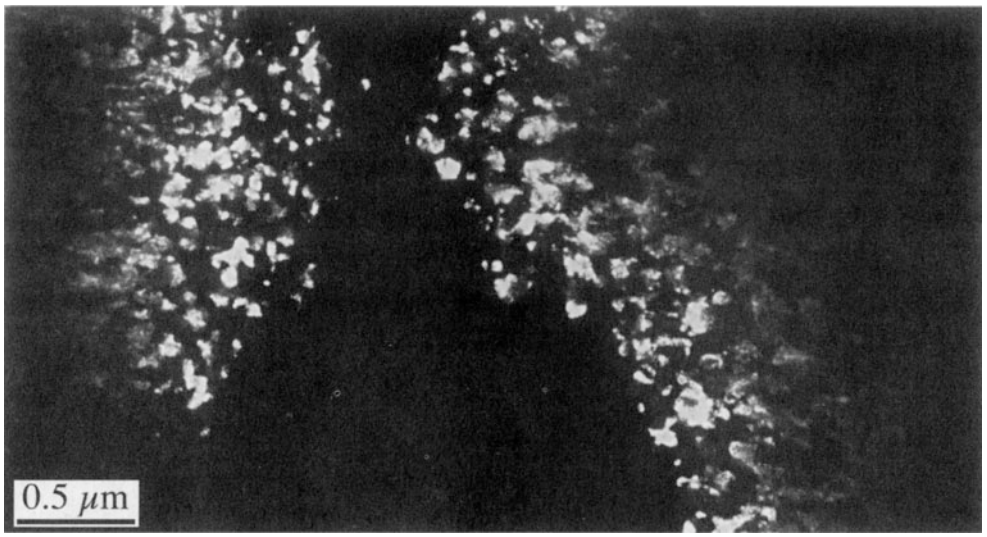
A



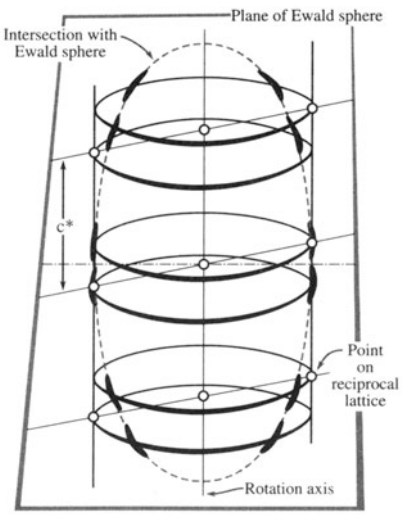
B


c*
Planar
Ewald
"sphere"
hkℓ
ϕ
b*
0
a*
hk0

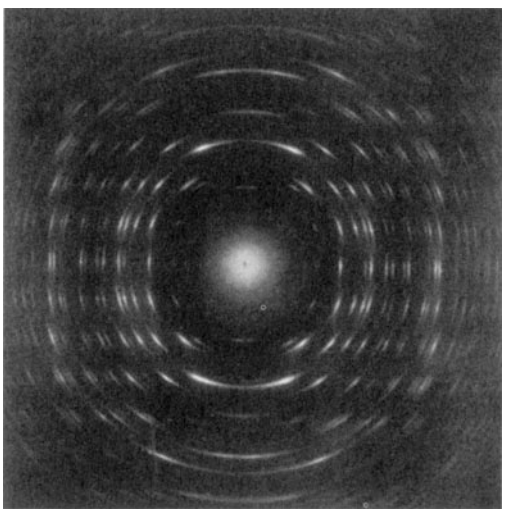
C


0.5 μm

D


Plane of Ewald sphere
Intersection with
Ewald sphere
c*
Point
on
reciprocal
lattice
Rotation axis

E



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 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 more realistic than for crystals at the Bragg condition. Following Howie (1988) we express the kinematical intensity, $I(\mathbf{k})$, by the expression

$$I(\mathbf{k}) = |f(\mathbf{k})|^2 \sum_{i,j} e^{i2\pi\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} \qquad [18.5]$$

Here we assume that there are N identical atoms contributing to the scattered intensity and they are located at the different $\mathbf{r}_i$ positions.

The $f(\mathbf{k})$ terms are the atomic scattering amplitudes, with $\mathbf{k}$ reminding us that there is an angular dependence on f. If the material is isotropic, we can simplify equation 18.5 as follows

$$I(\mathbf{k}) = N|f(\mathbf{k})|^2\left(1 + \frac{F(\mathbf{k})}{k}\right) \qquad [18.6]$$

where

$$F(\mathbf{k}) = \sum_{i \neq j} e^{i2\pi\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} \qquad [18.7]$$

$$F(\mathbf{k}) = k \int \rho(r) e^{i2\pi\mathbf{k}\cdot\mathbf{r}} dV \qquad [18.8]$$

$$F(\mathbf{k}) = 4\pi \int_0^\infty \rho(r) \sin(2\pi kr)\, r\, dr \qquad [18.9]$$

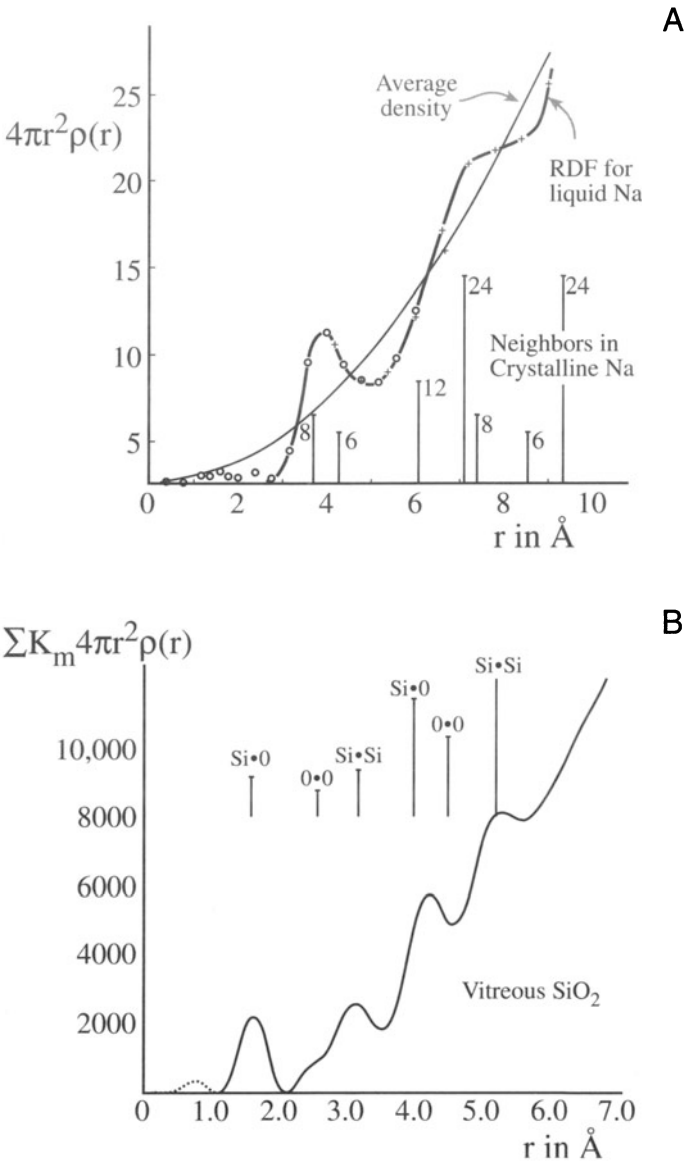


Figure 18.11. (A) Radial distribution function for liquid Na and the average density curve, superimposed on the distribution of the nearest neighbors in crystalline Na (vertical lines). (B) The RDF for vitreous SiO_2 is peaked at distances that represent various spacings between atoms.

The term $\rho(\mathbf{r})$ is the RDF. Equation 18.9 can be inverted to give an expression for ρ

$$\rho(r) - \rho_0 = \frac{1}{r}\int_0^\infty F(\mathbf{k}) \sin 2\pi kr\, dk \qquad [18.10]$$

The RDF can be obtained directly from DPs, and this process is enhanced if the patterns are energy-filtered to remove inelastic contributions (see Chapter 40 and the work of Cockayne *et al.* 1991). Software to determine the RDF

is listed back in Section 1.5. Then we can obtain DPs shown graphically in Figures 18.12 and 18.13. Alternatively, we can rearrange equation 18.6 again to give a "reduced intensity function" as illustrated by the work of Graczyk and Chaudhari (1973). Graczyk and Chaudhari showed clearly that the structure correlation can extend to 15 Å or more.

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 your image in terms of the size of regions in the amorphous structure. Hollow-cone DF imaging, where you use an annular C2 aperture 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.

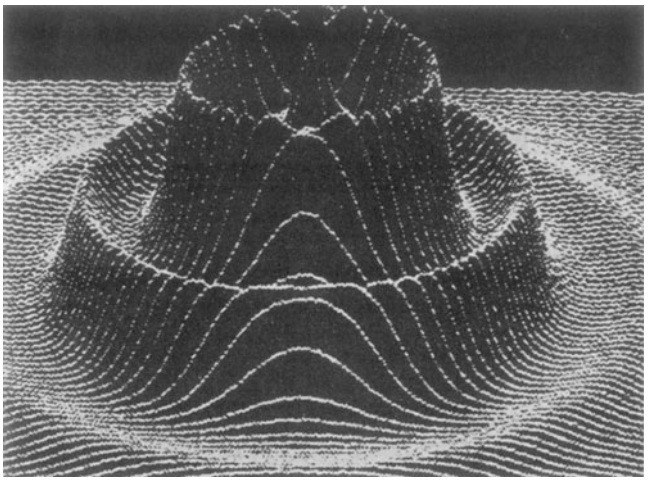


Figure 18.13. A computer plot of the diffracted-intensity distribution from an amorphous structure, showing diffuse rings of intensity. The direct-beam intensity is off scale.

Glass 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. The technique, known as diffuse-dark-field imaging, essentially forms an image from the region in the SAD pattern where the amorphous ring would be, if glass were present. We'll return to this topic in Chapter 31.

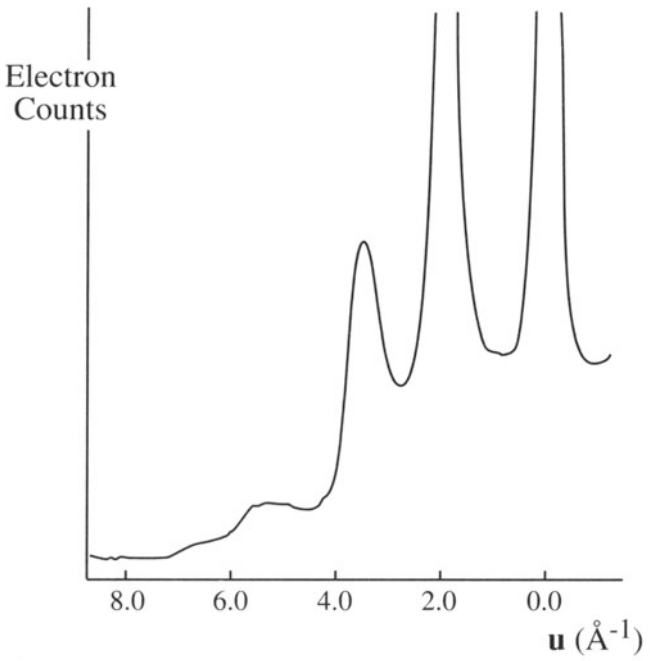


Figure 18.12. An intensity profile across an energy-filtered diffraction pattern from amorphous Si obtained by scanning the pattern across the entrance slit to a serial EELS spectrometer and recording only the elastic (on-axis) electrons.

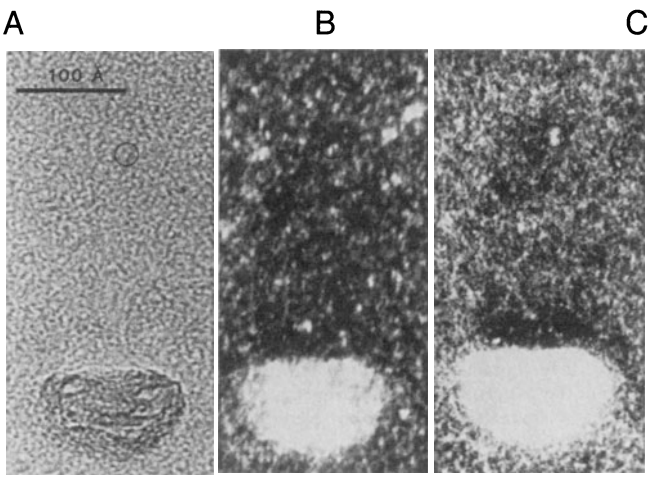


Figure 18.14. (A) BF image of amorphous carbon. (B) DF image from the diffuse diffracted intensity taken with a defocused beam. (C) A hollow-cone image showing more structure.

18.7. 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 $(\mathbf{g}_1\text{-}\mathbf{g}_2)$. 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 which we will discuss in Chapter 27. 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 (Tietz *et al.* 1995), as shown in Figure 18.15A. The position of the double-diffraction spots relative to the hematite and alumina reflections actually changed 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 18.15B shows the [0001] SAD pattern 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 18.15B.

Figures 18.15C and D show enlargements of regions near the $\{11\bar{2}0\}$ reflections in the [0001] SAD patterns 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 SAD patterns, as shown in Figures 18.15E 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 Figures 18.15C and D are still visible, as are the outer rings of spots (large

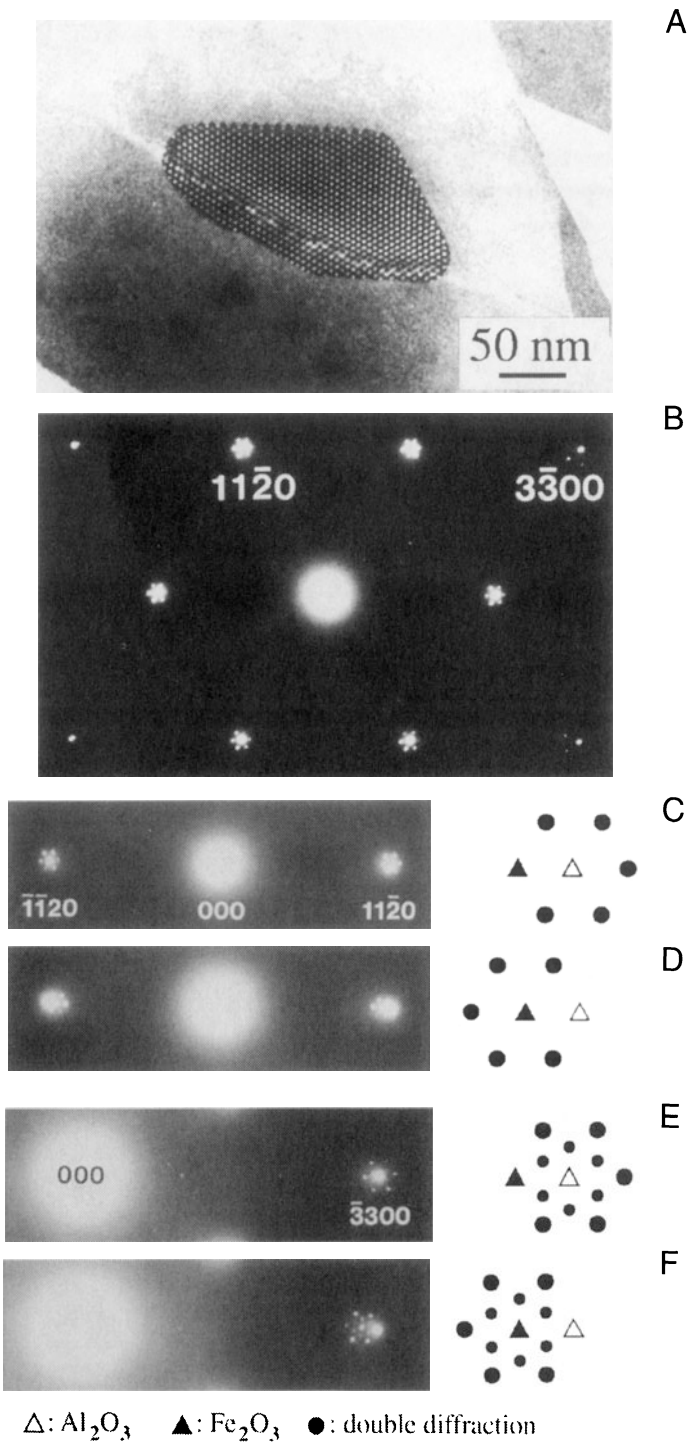


Figure 18.15. (A) BF on-axis image of a particle of α-Fe_2O_3 on α-Al_2O_3. (B) [0001] SAD pattern from α-Fe_2O_3 showing double-diffraction spots around the $\{11\bar{2}0\}$ and $\{3\bar{3}00\}$ reflections. (C) Enlargements of regions near the $(11\bar{2}0)$ reflections when the hematite island was on the top surface. (D) Enlargements of regions near the $(11\bar{2}0)$ reflections when the hematite island was on the bottom. (E) Enlargements of regions near the $(\bar{3}300)$ reflections when the hematite island was on the top surface. (F) Enlargements of regions near the $(\bar{3}300)$ reflections when the hematite island was on the bottom.

filled circles). In general, the outer ring of double-diffraction spots is more intense than the inner ring.

Both this top–bottom effect in particular, and double diffraction in general, can be explained by the simple geometric analysis we show in Figure 18.16; 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) gives a 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 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.

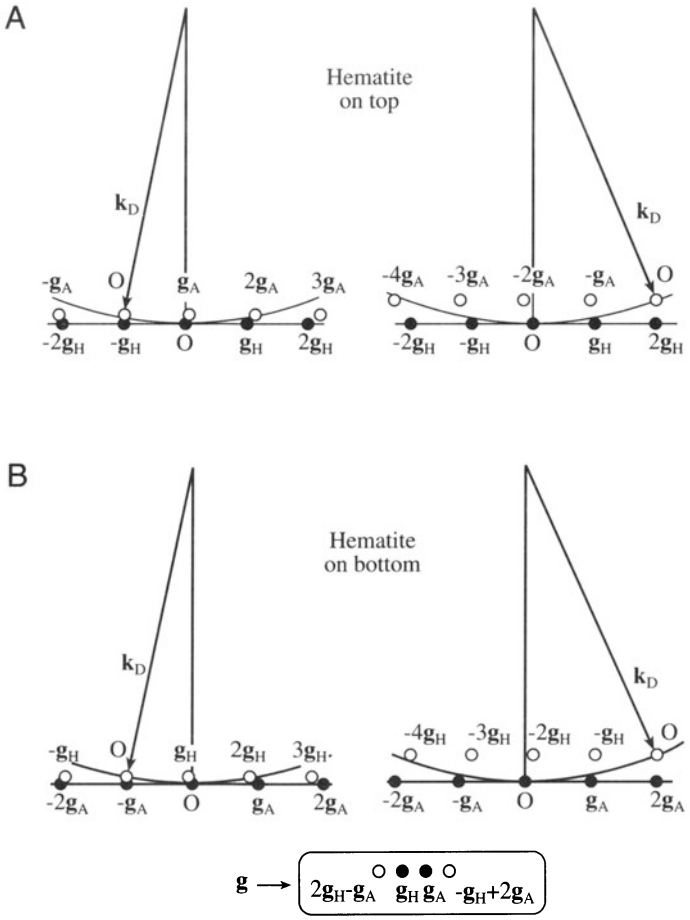


Figure 18.16. Top–bottom effect in double diffraction. The pattern depends on which of the two crystals is on top. In (A) α-Fe_2O_3 particles are on top of the Al_2O_3; in (B) Fe_2O_3 particles are below the Al_2O_3. Two nonequivalent paths for double diffraction are shown.

You can see from Figure 18.16A 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 18.16B. 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 29). 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'll return to this topic in Chapter 27 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.
- 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! More examples of double diffraction are given in Edington (1976).

18.8. ORIENTATION OF THE SPECIMEN

Once you have identified three **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 **B**. You can actually estimate **B** to within about 10° from the vector cross product as follows

$$\mathbf{B} = \mathbf{g}_1 \times \mathbf{g}_2 = \begin{bmatrix} \mathbf{i}_1 & \mathbf{i}_2 & \mathbf{i}_3 \\ h_1 & k_1 & \ell_1 \\ h_2 & k_2 & \ell_2 \end{bmatrix} \qquad [18.11]$$

$$= (k_1\ell_2 - k_2\ell_1,\ \ell_1 h_2 - \ell_2 h_1,\ h_1 k_2 - h_2 k_1) \qquad [18.12]$$

For the three-beam case, you can determine **B** with an accuracy of ~ 3°. 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 **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

$$\mathbf{g}_1 \cdot (\mathbf{g}_2 \times \mathbf{g}_3) = \frac{1}{V}\begin{bmatrix} h_1 & k_1 & \ell_1 \\ h_2 & k_2 & \ell_2 \\ h_3 & k_3 & \ell_3 \end{bmatrix} \qquad [18.13]$$

Now we can write a weighted-average expression for **B**

$$\mathbf{B} = \frac{\mathbf{g}_2 \times \mathbf{g}_3}{|\mathbf{g}_1|^2} + \frac{\mathbf{g}_3 \times \mathbf{g}_1}{|\mathbf{g}_2|^2} + \frac{\mathbf{g}_1 \times \mathbf{g}_3}{|\mathbf{g}_3|^2} \qquad [18.14]$$

A 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**.

In Figures 18.17–18.19, 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 18.17C

$$(12\bar{1}) = (110) + (01\bar{1}) \qquad [18.15]$$

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 (1976), as shown in Figure 18.20 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

$$\cos(\phi_{AB}) = \frac{\mathbf{g}_A \cdot \mathbf{g}_B}{|\mathbf{g}_A||\mathbf{g}_B|} \qquad [18.16]$$

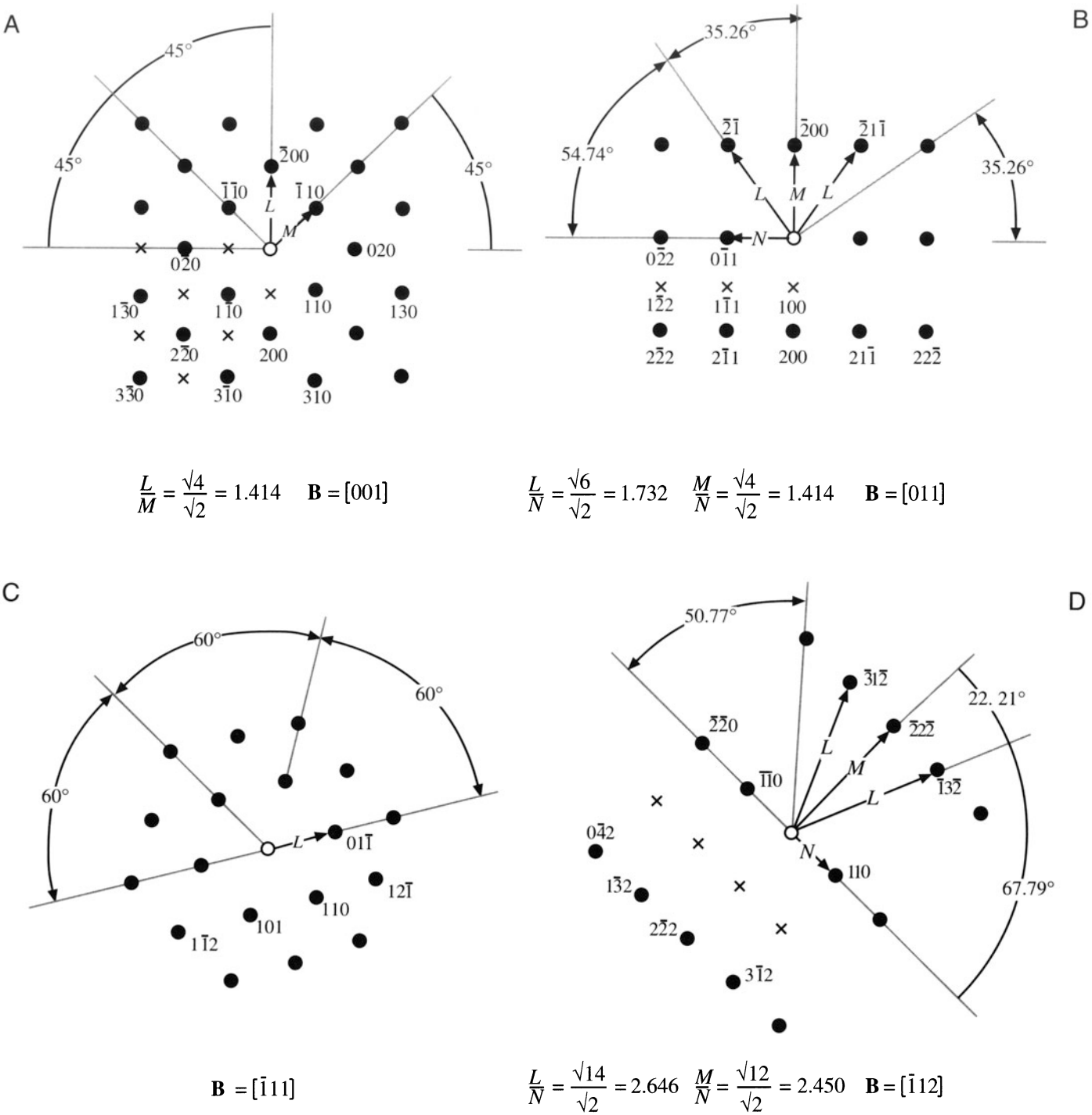
and so on for ϕ_{BC} and ϕ_{CA}. You should immediately recognize that this is the $\pm[12\bar{3}]$ pole, but continue. Now you can plug these indices into equation 18.11 or 18.14 to show that **B** = $[\bar{1}\bar{2}3]$.

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.

- Notice that if you used the stereographic technique with a noncubic material, you would locate a direction not a plane normal.
- You can make the determination of **B** more accurate by making **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).

18.9. ORIENTATION RELATIONSHIPS

Once you've learned how to index a DP and determine **B**, you can determine *orientation relationships*, which are one of the most useful aspects of diffraction in the TEM for the metallurgist. The orientation relationship (OR) between



$\frac{L}{M} = \frac{\sqrt{4}}{\sqrt{2}} = 1.414 \quad \mathbf{B} = [001]$

$\frac{L}{N} = \frac{\sqrt{6}}{\sqrt{2}} = 1.732 \quad \frac{M}{N} = \frac{\sqrt{4}}{\sqrt{2}} = 1.414 \quad \mathbf{B} = [011]$

$\mathbf{B} = [\bar{1}11]$

$\frac{L}{N} = \frac{\sqrt{14}}{\sqrt{2}} = 2.646 \quad \frac{M}{N} = \frac{\sqrt{12}}{\sqrt{2}} = 2.450 \quad \mathbf{B} = [\bar{1}12]$

Figure 18.17. Four standard indexed diffraction patterns 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.

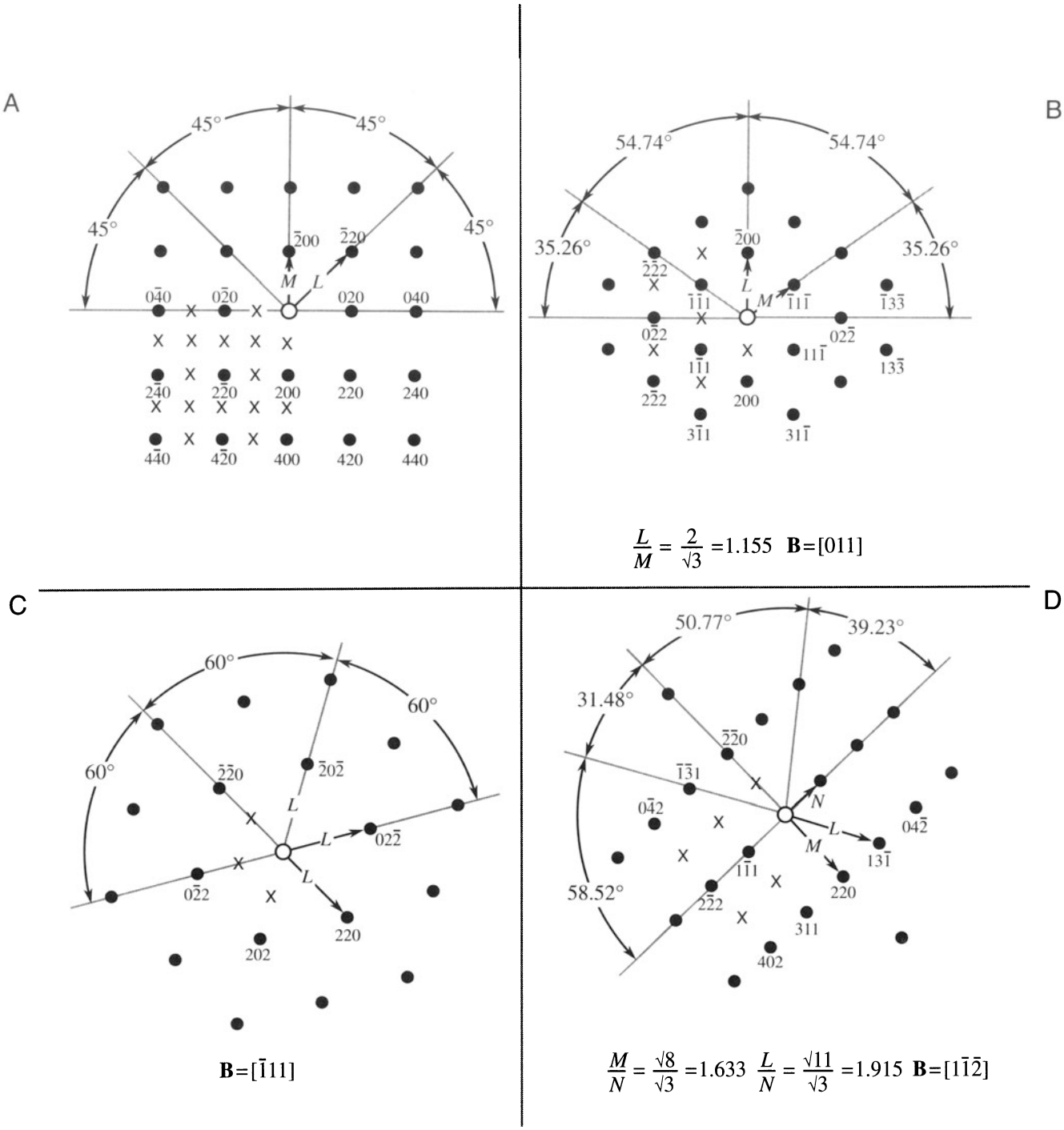


Figure 18.18. Four standard indexed diffraction patterns for fcc crystals in the [001], [011], [$\bar{1}$11], and [$\bar{1}$12] beam directions. Ratios of the principal spot spacings are shown as well as the angles between the principal plane normals. Forbidden reflections are indicated by x.

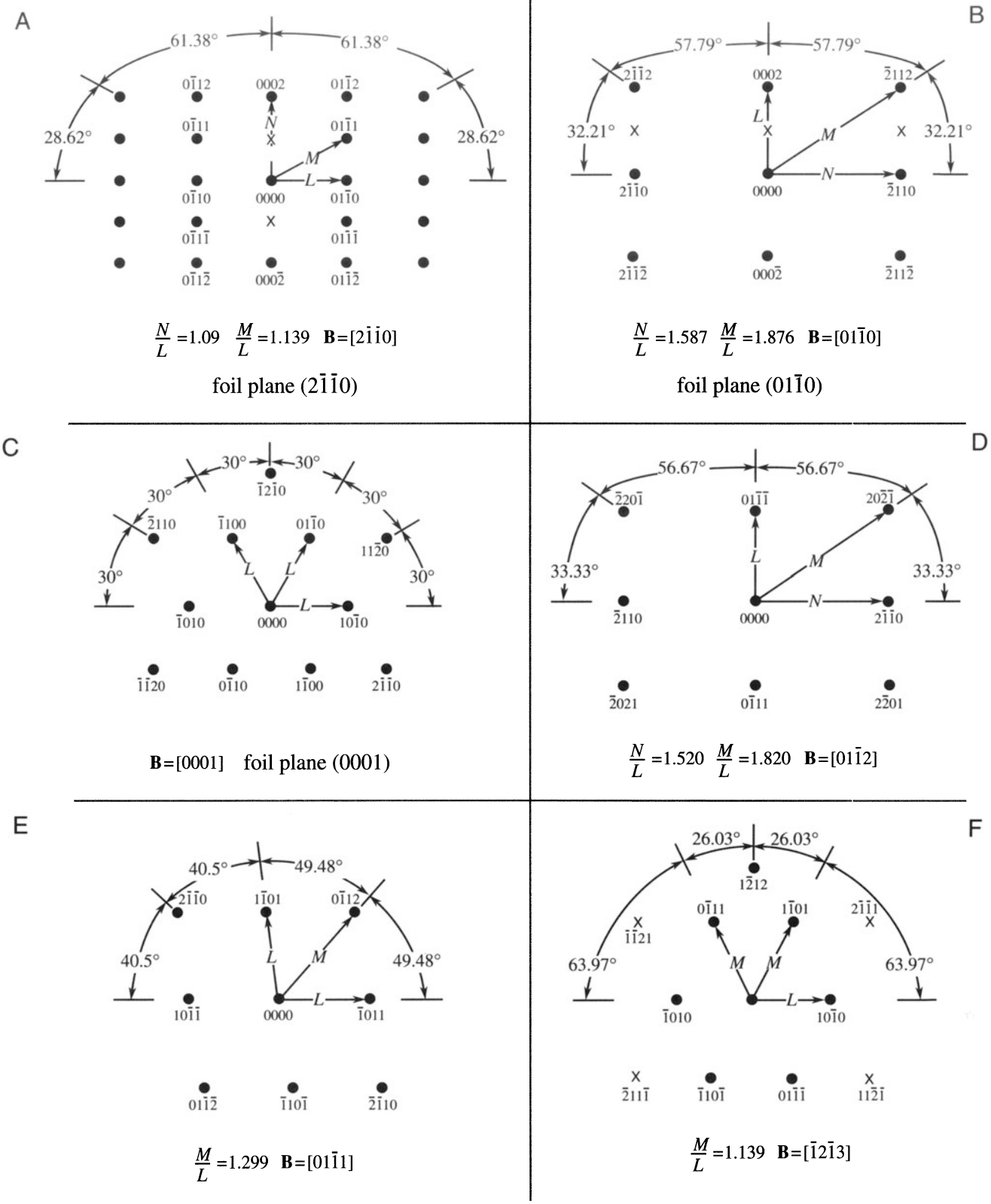


Figure 18.19. Six standard indexed diffraction patterns for hcp crystals in the [2$\bar{1}\bar{1}$0], [01$\bar{1}$0], [0001], [01$\bar{1}$2], [01$\bar{1}$1], and [1$\bar{2}$1$\bar{3}$] beam directions. Ratios of the principal spot spacings are shown as well as the angles between the principal plane normals. Forbidden reflections are indicated by x.

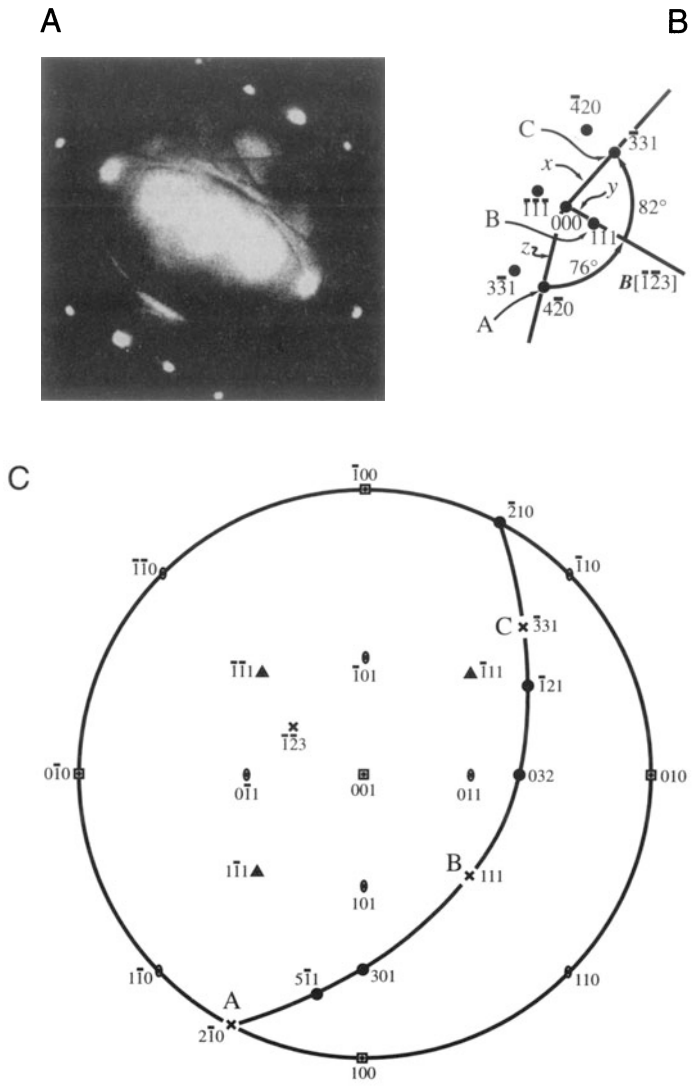


Figure 18.20. (A) An fcc pattern, indexed in (B) with the major indexed poles plotted on a stereographic projection in (C), identifying the pole of the great circle as $\bar{1}\bar{2}3$, which is therefore the beam direction for the pattern in (A).

two different crystals is important in many materials. We often want to know how a precipitate is oriented relative to its surrounding matrix, a fiber to the adjacent matrix, a thin film to the substrate, or two grains are oriented either side of a GB. The OR 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 (β-α) 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 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 or CBED pattern to provide more information. With CBED patterns from very small regions, you'll have to take a pattern in one crystal, translate the specimen, 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 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 from the next chapter 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 $\{hk\ell\}$ 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 this method and some more examples of ORs in Edington (1976).

18.10. 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 software packages to help you with any material, especially if it is not cubic. 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 PDF files, or better still the NIST/Sandia/ICDD electron-diffraction database, until you find a match. Yes, it is a lot of work and you have to remember some rules:

- Measurements made on calibrated SAD patterns 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. An example of such a DP is shown in Figure 18.21. As you can appreciate from the schematic, you must be careful not to confuse such patterns with those showing systematic absences.

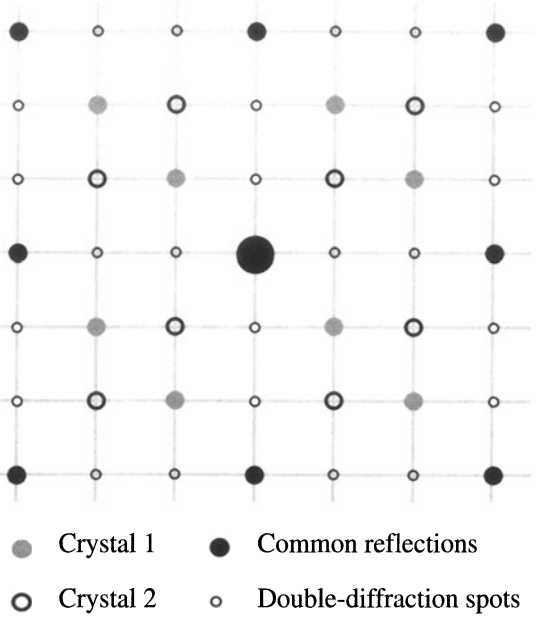


Figure 18.21. Care is needed to recognize diffraction from two similar domains, which appears identical to diffraction from a real structure with a different symmetry. All the spots lie on a square array which may lead to erroneous indexing as a 100 pattern. The DP actually consists of separate patterns from two overlapping crystals, plus double-diffraction spots as indicated.

A strategy for search-and-match procedures has been given by Lyman and Carr (1992). 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 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?
- 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 can also simulate the DPs to confirm that the popular software packages do reproduce what you see.

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) 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 **g** should give you the best value for both d and ϕ, but be absolutely sure that $\mathbf{s} = 0$ for your chosen **g**.
- DPs from polycrystalline and amorphous materials contain a wealth of information. The added value that TEM brings over X-ray diffraction is the spatial resolution and the accompanying images.
- Computer indexing of DPs will be the norm and will be automatic if you know your material. If you understand the principles discussed here you will avoid a few pitfalls. 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!

REFERENCES

General References

Andrews, K.W., Dyson, D.J., and Keown, S.R. (1971) *Interpretation of Electron Diffraction Patterns,* 2nd edition, Plenum Press, New York. An essential resource for anyone using electron diffraction.

Burger, M.J. (1978) *Elementary Crystallography,* MIT Press, Cambridge, Massachusetts. One of the classics.

Cullity, B.D. (1978) *Elements of X-ray Diffraction,* 2nd edition, Addison-Wesley, Reading, Massachusetts. The standard text on X-ray analysis.

Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science,* Van Nostrand Reinhold, New York. Part 2 of the book is full of useful hints and examples.

Giacovazzo, C., Monaco, H.L., Viterbo, D., Scordari, F., Gilli, G., Zanotti, G., and Catti, M. (1992) *Fundamentals of Crystallography,* Oxford University Press and IUCr. Oxford. A comprehensive reference book.

Glazer, A.M. (1987) *The Structure of Crystals,* Adam Hilger, Bristol, United Kingdom. The essentials condensed into a 50-page monograph.

Hammond, C. (1992) *Introduction to Crystallography,* 2nd edition, 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. R.F. Bunshah), Interscience, New York. A very helpful book if you can find a copy.

Kelly, A. and Groves, G.W. (1970) *Crystallography and Crystal Defects,* Addison-Wesley, Reading, MA. The original out-of-print 1970 edition has been reprinted by TechBooks, 4012 Williamsburg Court, Fairfax, Virginia 22032. A classic which all materials scientists should already have on their shelves. This is not only the standard introductory text on this subject but also gives a good review of the stereographic projection.

Klein, C. and Hurlbut, C.S. (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.

Smaill, J.S. (1972) *Metallurgical Stereographic Projections,* Adam Hilger Ltd., London. Chapter 20 is another source for stereographic projections and the Wulff net.

Vainshtein, B.K. (1981) *Modern Crystallography,* **I–IV**, Springer-Verlag, New York.

Specific References

Cockayne, D.J.H., McKenzie, D., and Mueller, D. (1991) *Microscopy, Microanalysis, Microstructure* **2,** 359.

Graczyk, J.F. and Chaudhari, P. (1973) *Phys. Stat. Sol.* **b58,** 163.

Howie, A. (1988) in *High-Resolution Transmission Microscopy and Associated Techniques* (Eds. P. Buseck, J. Cowley, and L. Eyring), p. 607, Oxford University Press, New York.

Lyman, C.E. and Carr, M.J. (1992) in *Electron Diffraction Techniques,* **2** (Ed. J.M. Cowley), p. 373, Oxford University Press, New York.

Randle, V. (1993) *The Measurement of Grain Boundary Geometry,* Institute of Physics, Bristol, United Kingdom.

Rudee, M.L. and Howie, A. (1972) *Phil. Mag.* **25,** 1001.

Tietz, L.A., Carter, C.B., and McKernan, S. (1995) *Ultramicroscopy* **60,** 241.

Vainshtein, B.K., Zuyagin, B.B., and Avilov, A.V. (1992) in *Electron Diffraction Techniques,* **1** (Ed. J.M. Cowley), p. 216, Oxford University Press, New York.

References for Crystallographic Data

Stereographic projections and the Wulff net can be obtained from SPI Supplies, P.O. Box 656, Westchester, Pennsylvania 19831-0656 (1-800 242 4774).

Donnay, J.D.H. and Ondik, H.M. (1972) *Crystal Data: Determinative Tables,* 3rd edition, National Bureau of Standards, US Dept. of Commerce, Washington, DC. This source lists the ratios of lattice parameters for different materials.

Wells, A.F. (1984) *Structural Inorganic Chemistry,* 6th edition, Oxford University Press, New York. The source for crystal-structure data in inorganic materials.

Villars, P. and Calvert, L.D. (1985) *Pearson's Handbook of Crystallographic Data for Intermetallic Phases*, ASM, Metals Park, Ohio is now in many volumes covering an ever-growing number of materials.

ICDD Powder Diffraction File is produced by the International Center for Diffraction Data (Swarthmore, Pennsylvania, 1990). It is available in various formats. Most researchers favor the CD-ROM version.

ICDD Elemental and Lattice Spacing Index is produced by the same Center but is only presently available in printed form. This index used to be known as the ASTM cards (3" by 5" index cards!). Each file gives the *d*-spacings and X-ray diffraction peak intensities. These files should be in a more useful computer-accessible form.

NIST Crystal Data can be purchased as a CD-ROM or on tape. Parts are from the Donnay-Ondik books (see above). A program called NBS*SEARCH will allow you to search this database. These files give not only crystallographic data but also physical data on more than 100,000 organic and inorganic materials. Obtainable from NIST Crystal Data Center, NIST, Gaithersburg, MD 20899.

NIST/Sandia/ICDD Electron Diffraction Database has become available thanks to the tireless efforts of M. Carr, who has also provided methods for searching this database on a PC.

Desktop Microscopist (see Section 1.5). This program can look up crystal data and plot out the diffraction pattern.

Kikuchi Diffraction

19

CHAPTER PREVIEW

In this chapter and the following two, we will discuss two special cases of electron diffraction. In the first we find that inelastically scattered electrons give rise to arrays of lines in DPs known as Kikuchi patterns. In the second, we will form DPs with a convergent rather than a 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." We can then think of these electrons as traveling in different "incident" directions. When the direction is appropriate, these electrons can be scattered again, this time by Bragg diffraction. In the second technique we use a convergent beam intentionally to make the electrons incident on the crystal over a range of different angles. 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 this chapter we will show that these Kikuchi patterns can be used to give us accurate information on the beam direction and can give a direct link in reciprocal space to the stereographic projection. The topics we'll cover are basically experimental. The ideas developed here will carry over to the next two chapters when we discuss HOLZ lines.

Kikuchi Diffraction

19

19.1. THE ORIGIN OF KIKUCHI LINES

The reason we form Kikuchi patterns is that, if the specimen is thick enough, we will generate a large number of scattered electrons which travel in all directions; i.e., they have been *incoherently* scattered but not necessarily *inelastically* scattered. They are sometimes referred to as diffusely scattered electrons. These electrons can then be Bragg diffracted by the planes. The rest of the story is geometry.

We'll discuss a little of the theory in Section 19.5, but for now we'll note the experimental facts.

- Since energy losses are small compared to E_0, the diffusely scattered electrons 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. We learned in Chapter 3 that inelastic scattering is "peaked in the forward direction."
- 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. This is one of the few situations when thinner is not necessarily better.
- Kikuchi (1928) described this phenomenon before the development of the TEM; it can occur in any crystalline specimen.

Diffuse scattering will again be important when we discuss image formation in Chapter 31. 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 Chapter 40). The specimen needs to be thick enough, but if it is too thick then there will be no lines because inelastic scattering then dominates and there is no subsequent Bragg diffraction of these electrons.

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 you see in the DP. We imagine (Figure 19.2A) that electrons have been generated at the point shown and scatter in all directions (but mainly forward). Some of these electrons will travel at an angle θ_B to the $hk\ell$ planes as shown in Figure 19.2B and then be Bragg diffracted by the 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. Construct the cones by considering all the vectors oriented at angle θ_B to the $hk\ell$ plane; these are called *Kossel cones*. There is a pair of Kossel cones for $\pm\mathbf{g}$, another pair for $\pm 2\mathbf{g}$, and so on.

What we see in the DP is the intersection of these two cones with the screen.

Since the screen 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 look like two parallel lines: the pair of Kikuchi lines. We'll sometimes refer to this pair of 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 more complex (see Section 19.6).

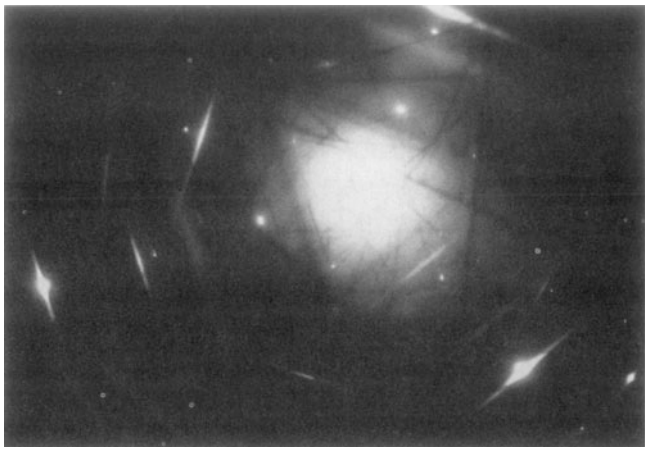


Figure 19.1. An ideal diffraction pattern containing both well-defined spots and clearly visible pairs of bright and dark Kikuchi lines.

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 beam which was initially closest to the optic axis, and therefore the more intense, is further away after being scattered. This beam then gives the excess line and the other the deficient line. We can see that this simple idea really does work in Figure 19.1.

The value of this result is apparent 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 cones shown in Figure 19.2C act as if they are rigidly fixed to the plane $hk\ell$; they are thus fixed to the crystal. We can draw a line halfway between the two Kikuchi lines to represent the trace of the plane $(hk\ell)$. Remember our angles are all small. This simple observation explains why we have a whole chapter on 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. 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 $2\theta_B$.

- When the **g** Kikuchi line passes through the reflection G, $\mathbf{s}_g = 0$ (the Bragg condition is satisfied), and the $\bar{\mathbf{g}}$ Kikuchi line passes through O.

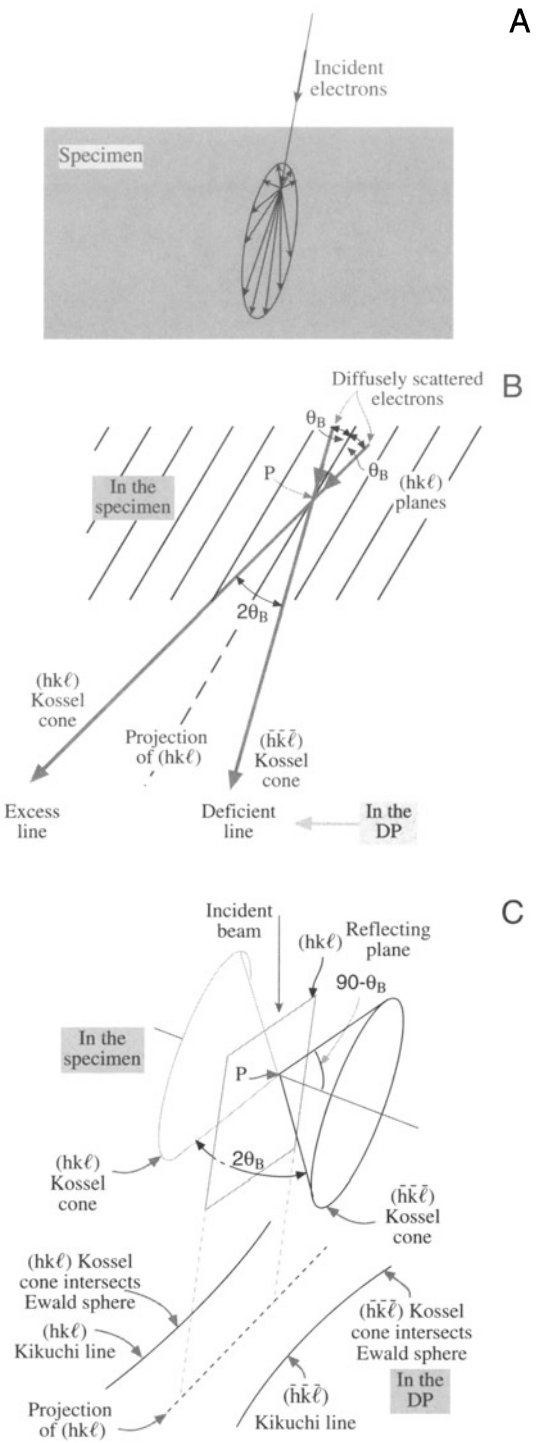


Figure 19.2. (A) Schematic representation of all electron scattering localized at a single point in the specimen. In (B) some of the scattered electrons are diffracted because they travel at the Bragg angles θ_B to certain $hk\ell$ planes. The diffracted electrons form Kossel cones centered at P on the diffracting planes. The lines closest to the incident beam direction are dark (deficient) and the lines furthest away from the beam are bright (excess). In (C) the cones intercept the Ewald sphere, creating parabolas which approximate to straight Kikuchi lines in the diffraction patterns because θ_B is small.

- A corollary: if the direct beam is exactly parallel to the plane $hk\ell$, the **g** and $\bar{\mathbf{g}}$ Kikuchi lines are symmetrically displaced about O with **g** "passing through" **g**/2 and $\bar{\mathbf{g}}$ "passing through" $\bar{\mathbf{g}}$/2.

In this latter case, our simple explanation breaks down because it would predict equal intensity in both excess and deficient Kikuchi lines, and thus they would both be indistinguishable from the diffuse-scattered background. Therefore, no Kikuchi lines should be visible if the beam is exactly down a zone axis, and this is not true. So the full Kikuchi line explanation is more complex and requires Bloch-wave theory.

19.3. CONSTRUCTING KIKUCHI MAPS

The method for constructing Kikuchi maps is illustrated in Figure 19.3. 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 |**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 19.4, keeping the common

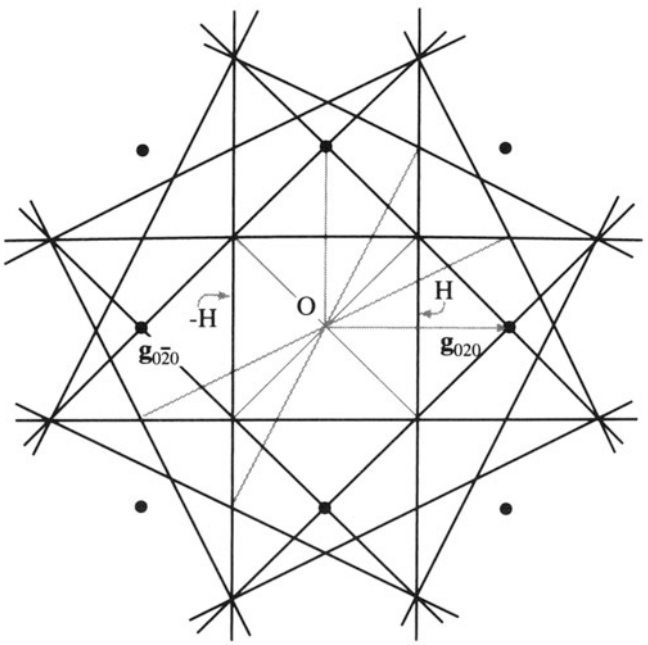


Figure 19.3. To construct a Kikuchi pattern, draw pairs of lines each bisecting the ±**g** vectors. For example, when the [001] fcc pole is on axis, the vector $\mathbf{g}_{020}$ is bisected by the vertical line at H and the companion Kikuchi line is at -H ($0\bar{2}0$). All other Kikuchi line pairs can be constructed for any **g**-vector.

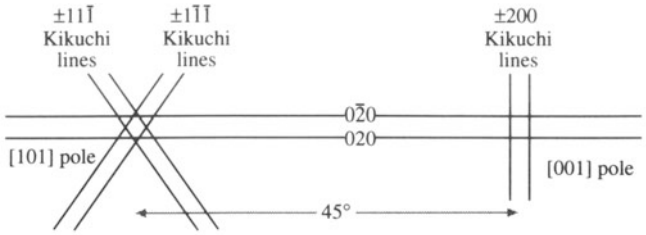


Figure 19.4. From one Kikuchi pattern we can extend the lines to create a second pattern. For example, knowing the [001] pattern we can construct the [101] pattern since a pair of lines is common to both. So we draw the $0\bar{2}0$ and 020 lines from the [001] pole 45° to the [101] pole.

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.

It's a good exercise to construct a Kikuchi map for your material as illustrated for the fcc case 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. Maps can also be downloaded from the WWW using EMS.

You can appreciate the value of Kikuchi maps in noncubic 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: 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 can 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.
- For most noncubic materials and particularly if you are working with monoclinic or triclinic crystals, it's not practical to construct the com-

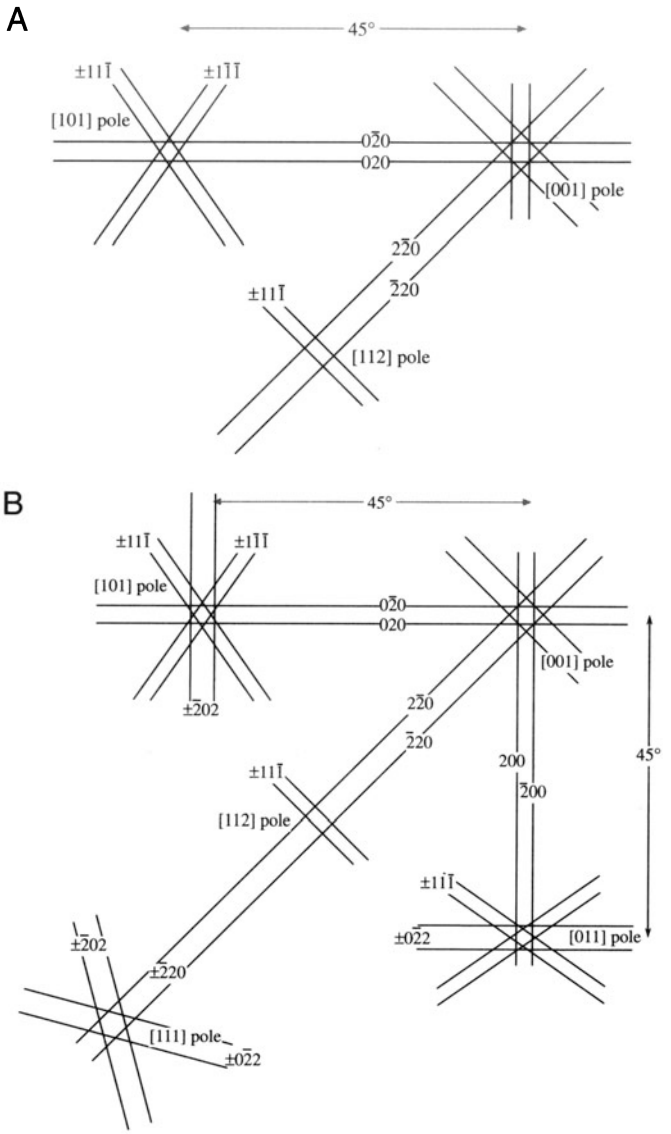


Figure 19.5. (A) Construction of the [112] pattern from the [001] and [101] patterns by extending the Kikuchi lines common to each pair of patterns. The [11$\bar{1}$] pair is common to the [101] and [112] patterns and the $\bar{2}$20 pair is common to [001] and [112] patterns. (B) Other poles can be 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.

plete 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.

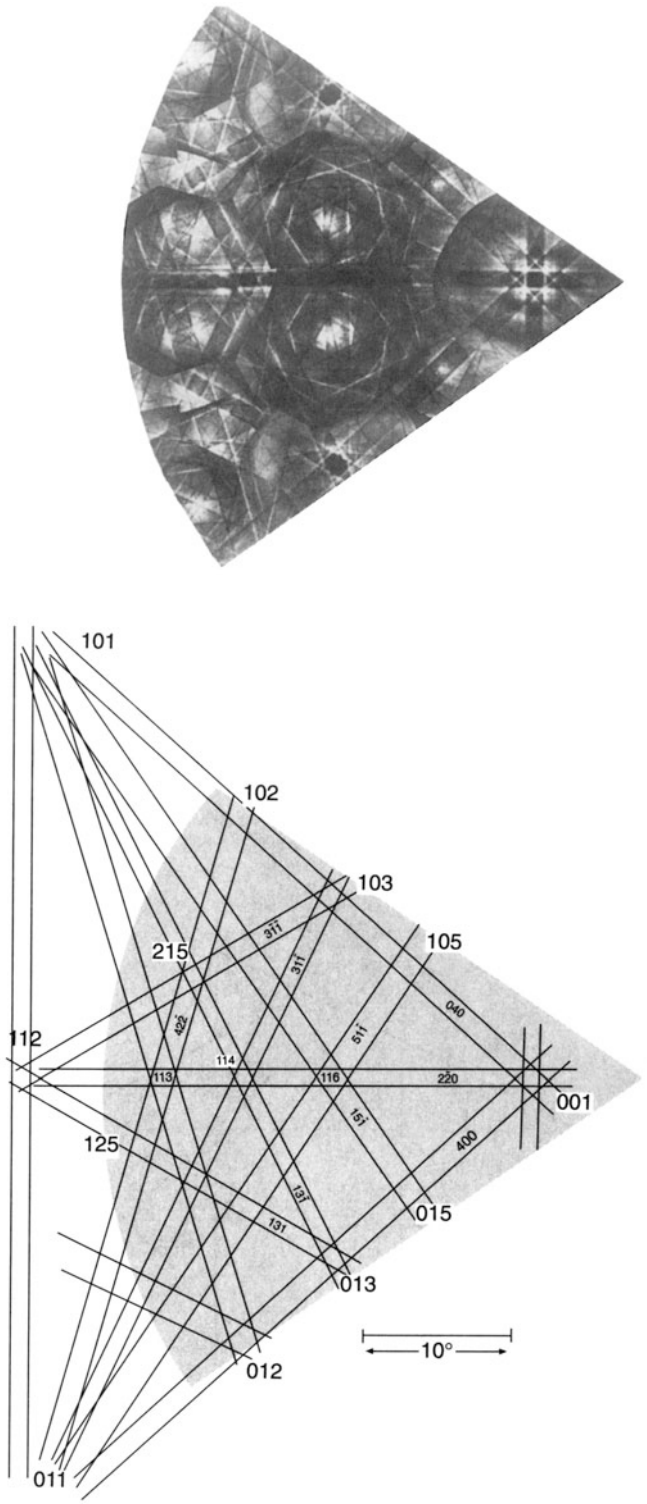


Figure 19.6. (A) Experimental Kikuchi map for fcc crystals and (B) indexed Kikuchi lines in the schematic map.

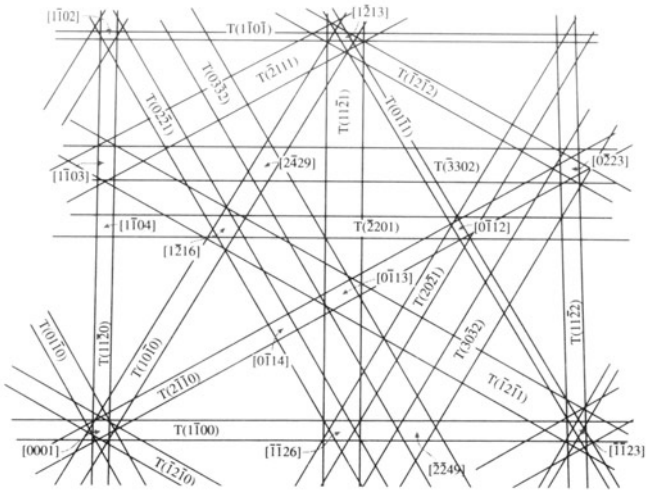


Figure 19.7. Part of the schematic Kikuchi map for hexagonal Ag_2Al with the principal poles and pairs of lines indexed.

- At the TEM, record the Kikuchi pattern for several special low-index poles along with the spot pattern.
- Index the DPs consistently.
- Print DPs for each pole at the scale you used in your line drawing of the Kikuchi map.
- 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.

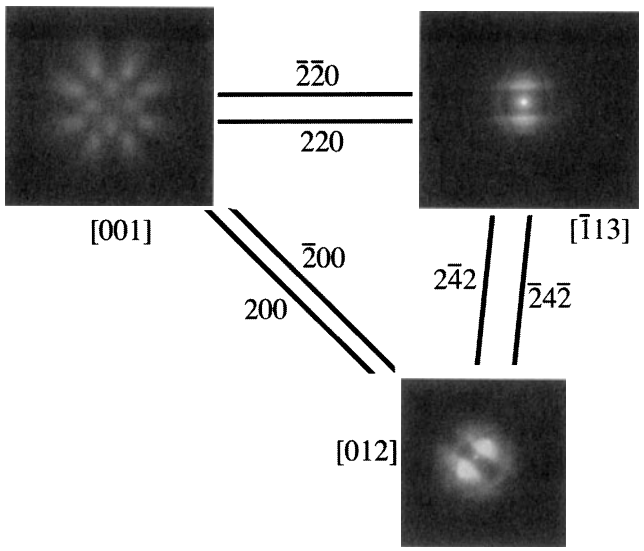


Figure 19.8. Experimental Kikuchi patterns around three principal poles in MgO with the common Kikuchi lines between each pole drawn in. You should compare this figure with the diffraction patterns in Figure 18.7.

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 further you have to go.

By now you will appreciate even more the value of the stereographic projection we introduced in Chapter 18. 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 Minneapolis and London, but use the Kikuchi map to locate the Guthrie Theater and Buckingham Palace.

19.4. CRYSTAL ORIENTATION AND KIKUCHI MAPS

In Chapter 18 we showed how you could estimate the orientation of the beam relative to the crystal with an accuracy of ~3°. Using Kikuchi patterns we can increase this accuracy to ~0.1°.

A routine method for this determination has again been developed by Thomas and co-workers (e.g., Okamoto *et al.* 1967), who pioneered this use of Kikuchi maps. 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 you can check your 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); calculate each angle using equation 18.3 if your material is cubic.

Measure the distances OA, OB, and OC in Figure 19.8 and, using the calibrated camera length, convert these distances into angles, ρ_1, ρ_2, and ρ_3 (which are defined in Figure 19.9B). 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 $(90 - \phi)$.

You can solve these three equations for the three unknowns U, V, and W and hence we have **B**. Finally, always check the sign of **B**, as we described in Section 18.8.

It is possible that the DP you have to work with is not obviously near a zone axis. Normally, while you are at

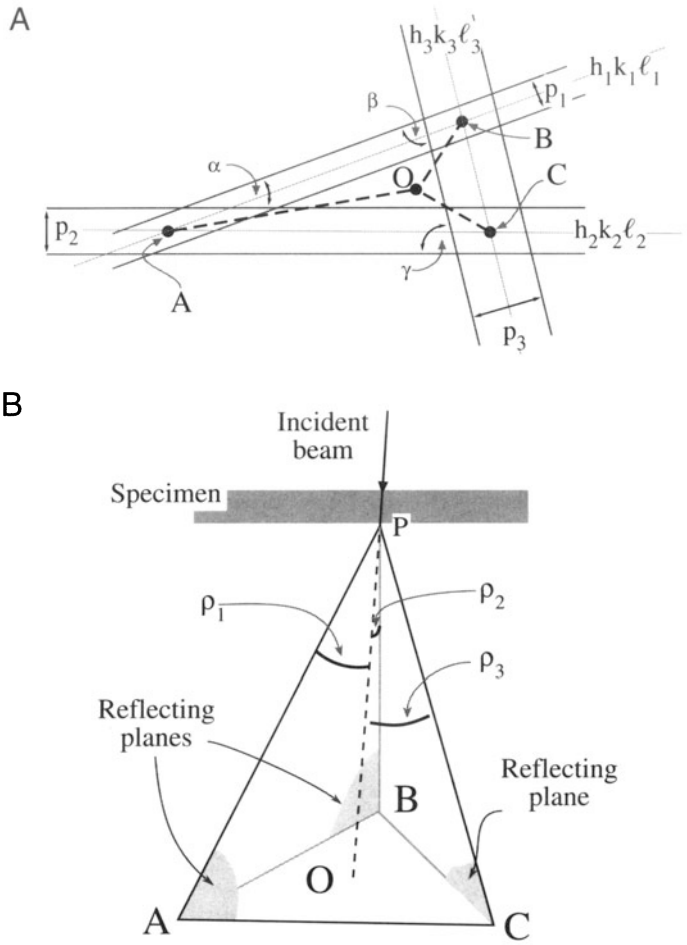


Figure 19.9. In (A) pairs of Kikuchi lines from the reflecting planes also intercept at points A, B, C. The distances from O to the points A, B, C correspond to the angles between the beam direction and the three zone axes while the angles α, β, γ correspond to the angles between pairs of plane normals. The angle α is between the $(h_1k_1\ell_1)$ and $(h_2k_2\ell_2)$ plane normals, etc. (B) Three reflecting planes in the specimen with traces AB $(h_1k_1\ell_1)$, AC $(h_2k_2\ell_2)$, and BC $(h_3k_3\ell_3)$ around the direct beam, O. The traces of pairs of planes intercept at A (AB, AC), B (AB, BC), and C (AC, BC).

the microscope, you will tilt along the different Kikuchi bands until you find the appropriate poles to ease your task later. 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. Now trace these lines in both directions and you will find 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.

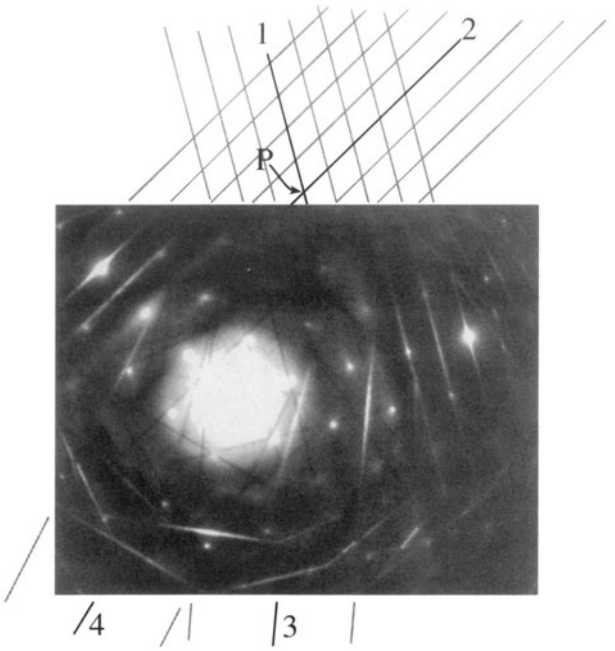


Figure 19.10. To index a diffraction pattern well away from a low-index zone axis, extend the Kikuchi lines. The dark lines 1–4 represent the traces of the diffracting planes which intercept at a pole (P). For Kikuchi lines 1 and 2 the higher-order extensions are also drawn. From the *d*-spacings, index the Kikuchi line pairs. The angles between the beam direction and the poles, P, can then be measured directly.

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.* (1967). When $\mathbf{s}_g$ is negative, the **g** Kikuchi line is on the same side of **g** as O; when $\mathbf{s}_g$ is positive, the line lies on the opposite side of **g**. For high-energy electrons, and knowing the camera length *L*, we can write an expression for the angle η

$$\eta = \frac{x}{L} = \frac{x\lambda}{Rd} \qquad [19.1]$$

where *d* is $|\mathbf{g}|^{-1}$. The distances *x* and *R* are measured on the photographic negative.

The angle ε is given by

$$\varepsilon = \frac{s}{g} \qquad [19.2]$$

Now we can set $\varepsilon = \eta$, to give

$$s = \varepsilon g = \frac{x}{L}g = \frac{x}{Ld} \qquad [19.3]$$

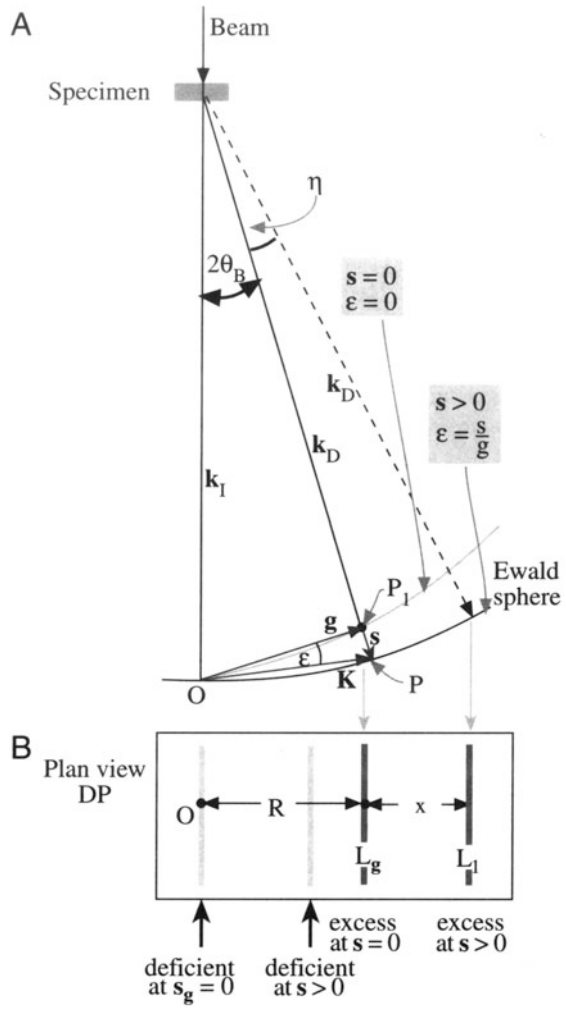


Figure 19.11. (A) The distance between the diffraction spot and its Kikuchi line gives a direct measure of **s**. The angle ε is *s*/*g* and is zero at the exact Bragg condition. (B) Measure x, the spacing between O and the deficient line (or G and the excess line), to determine **s**.

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

$$\frac{R}{L} = 2\theta_B = \frac{\lambda}{d} \qquad [19.4]$$

Hence the expression for s is

$$s = \frac{x}{Ld} = \frac{x}{d} \cdot \frac{\lambda}{Rd} \qquad [19.5]$$

$$s = \frac{x}{R}\frac{\lambda}{d^2} = \frac{x}{R}\lambda g^2 \qquad [19.6]$$

We'll reconsider this equation when we discuss weak-beam microscopy in Chapter 26.

Ryder and Pitsch (1968) 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

$$\mathbf{B} = \alpha_1 |\mathbf{g}_1|^2 (\mathbf{g}_2 \times \mathbf{g}_3) + \alpha_2 |\mathbf{g}_2|^2 (\mathbf{g}_3 \times \mathbf{g}_1) + \alpha_3 |\mathbf{g}_3|^2 (\mathbf{g}_1 \times \mathbf{g}_2) \qquad [19.7]$$

where α_i is given by

$$\alpha_i = \frac{R_i + 2x_i}{R_i} \qquad [19.8]$$

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.* (1971) 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 simply known as electron-backscatter patterns or EBSPs. They were regarded as a curiosity until it was shown (Dingley *et al.* 1992, Randle 1993) that you can rapidly map out the texture of polycrystalline materials using these patterns, without thinning the sample. New detection systems, similar to the YAG- or CCD-based cameras, and some fast computer algorithms have led to the development of "orientation imaging" in the SEM. Similar techniques should be available for TEM Kikuchi maps. They won't be as automated, but TEM can give the interface plane much more accurately so the two techniques will be complementary.
- 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 23 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 should remember that their physical origin is *completely different*; the most im-

portant 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 23) 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 refer you to Reimer (1993) for further reading.
- We mentioned earlier that the contrast between the lines, i.e., the band, is complex. The contrast is actually strongly influenced by anomalous absorption of the Bloch waves which are formed by coherent scattering of the incoherently scattered electrons; 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: 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 of diffraction (including Kikuchi lines and the following CBED patterns) 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. 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 then you must read "Skew thoughts on parallelism" by Christenson and Eades (1988).

CHAPTER SUMMARY

- The Kikuchi lines consist of an excess line and a deficient line. In the DP, the excess line is further from the direct beam than the deficient line.
- The Kikuchi lines are 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 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 $\mathbf{s}_g = 0$).

Pairs of Kikuchi lines define the road. Taken together, the roads make up a map. The rule is different than road maps: in our maps, narrow roads are the most important! What is the relevance of the roadside curbs? They define the roads and tell us when we are standing on them, but we are not too interested in their detailed appearance. We view Kikuchi maps as an invaluable tool for the microscopist.

Kikuchi lines and Kikuchi maps are one of the most important aids we have when orienting, or determining the orientation of, crystalline materials. Knowing the orientation of your specimen is essential for any form of quantitative microscopy, whether you're analyzing dislocation Burgers vectors by diffraction contrast, imaging grain boundaries with lattice resolution, or measuring chemistry variations by EELS or XEDS. They are especially useful when combined with the map of zones and poles (directions and plane normals) on the stereographic projection. Use the computer to check or to assist you in constructing a map for your material.

REFERENCES

General References

Thomas, G. (1978) in *Modern Diffraction and Imaging Techniques in Material Science* (Eds. S. Amelinckx, R. Gevers, and J. Van Landuyt), p. 399, North-Holland, Amsterdam.

Specific References

Christenson, K.K. and Eades, J.A. (1988) *Ultramicroscopy* **26**, 113.

Dingley, D.J., Baba-Kishi, K., and Randle, V. (1992) *An Atlas of Electron Back-Scatter Diffraction Patterns,* Institute of Physics, Bristol, United Kingdom.

Johari, O. and Thomas, G. (1969) *The Stereographic Projection and Its Applications in Techniques of Metals Research* (Ed. R.F. Bunshah), Interscience, New York.

Kikuchi, S. (1928), *Jap. J. Phys.* **5**, 23.

Levine, E., Bell, W.L., and Thomas, G. (1966) *J. Appl. Phys.* **37**, 2141.

Okamoto, P.R., Levine, E., and Thomas, G. (1967) *J. Appl. Phys.* **38**, 289.

Randle, V. (1993) *The Measurement of Grain Boundary Geometry,* Institute of Physics, Bristol, United Kingdom.

Reimer, L. (1993) *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis,* 3rd edition, Springer-Verlag, New York.

Ryder, P.L. and Pitsch, W. (1968) *Phil. Mag.* **18**, 807.

Tan, T.Y., Bell, W.L., and Thomas, G. (1971) *Phil Mag.* **24**, 417.

Obtaining CBED Patterns

20

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 SAD patterns from areas which are less than ~0.5 μm in diameter. This size is large compared to the dimensions of many crystalline features that interest us in materials science (Chapter 16).
- SAD patterns contain only rather imprecise two-dimensional crystallographic information because the Bragg conditions are relaxed for a thin specimen and small grains within the specimen (Chapter 17).

The technique of CBED overcomes both of these limitations and also generates much new diffraction information, which we will use in Chapter 21.

In this chapter we will show you how simple it is to use the versatility of modern (S)TEMs to create a range of CBED patterns containing a variety of contrast effects. We will also take you through the steps required to index the HOLZ spots and lines which occur under certain experimental conditions. In Chapter 21 you will see why these HOLZ features are so useful. They can give us an almost complete crystallographic analysis of the specimen. As is often the case in TEM, our advantage is that we have high spatial resolution. CBED, like many other sophisticated analytical techniques, uses various obscure definitions and acronyms which we will attempt to clarify as we introduce them.

Obtaining CBED Patterns

20

20.1. WHY USE A CONVERGENT BEAM?

Historically, CBED is the oldest TEM diffraction technique. It was originally developed by Kossel and Möllenstedt (1939) well before LePoole (1947) developed SAD. While SAD is the classical 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 traditionally the diameter of the smallest area you can select by SAD is about 0.5 μm with an error of similar dimensions. However, if you have an intermediate voltage HRTEM with a very low C_s you may be able to use SAD to analyze areas ~0.1 μm in diameter. Many crystal defects and second-phase precipitates which influence the properties of materials are much smaller than this. As we've mentioned, one way we can overcome this limitation is to use a convergent beam of electrons (see Figure 9.4) to limit the region of the specimen which contributes to the DP. This region is a function of the beam size and beam–specimen interaction volume, which increases with specimen thickness but can be a lot smaller than in SAD. In fact, several so-called "microdiffraction" techniques have been developed to overcome the spatial-resolution limitations of SAD in a TEM. We'll review these techniques in Section 21.8. CBED is by far the most simple and versatile microdiffraction technique.

In addition to the improved spatial resolution, CBED gives us a wealth of new information not available in SAD, sometimes from a single DP. Such information comprises:

- Specimen thickness.
- Unit cell and precise lattice parameters.
- Crystal system and true 3D crystal symmetry (point group and space group).
- Enantiomorphism, if present.

With such capabilities, CBED has transformed electron diffraction from the "poor relative" of X-ray and neutron diffraction to a precise, and in some senses unique, diffraction technique.

> The big advantage of CBED 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 show you, among other things, how to perform what is known as "electron crystallography." In some materials we can even study scattering from within a unit cell. All these advantages can simultaneously be coupled with XEDS and EELS data, allowing you to achieve a remarkable degree of specimen characterization.

There are two potential drawbacks which you should always keep in mind:

- You may have local contamination which can cause localized stresses.
- The convergent beam may heat or damage the region of the specimen as you examine it.

In early probe-forming TEMs or those containing minilenses, you 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 should not suffer from this problem (see Chapter 8). Small regions of a clean specimen can be studied for minutes or even hours without visible contamination.

> Most specimen contamination is caused by the preparation process.

Beam heating/damage may be a problem in materials with poor thermal conductivity, but this can be mini-

mized by using a thin conductive carbon coating or preferably using a cooling holder. This latter approach has other advantages for CBED, as we'll see.

You may find that it is easier to use CBED rather than conventional SAD procedures. When reading the literature, remember that researchers often use the technique with which they are most familiar rather than the best one available.

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. So a low-background, double-tilt, cooling holder is really useful. Tilt-rotation holders can sometimes be advantageous, but they are not available in a cooling or low-background form.

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 the 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 ~10–100 nm in diameter), as shown in Figure 20.1. We have already seen in Chapters 11 and 16 that parallel illumination means that the SAD pattern consists of an array of sharp maxima in the back focal plane of the objective lens; Figure 20.2A shows such an SAD pattern from pure Si. In contrast, the beam convergence in CBED gives rise to a pattern of disks of intensity; 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, you can certainly see that it contains a wealth of contrast detail not present in the SAD pattern. We'll see that, like SAD, CBED is most useful when the beam is oriented along a zone axis in the crystal, giving a symmetrical zone-axis (diffraction) pattern, commonly called a ZAP.

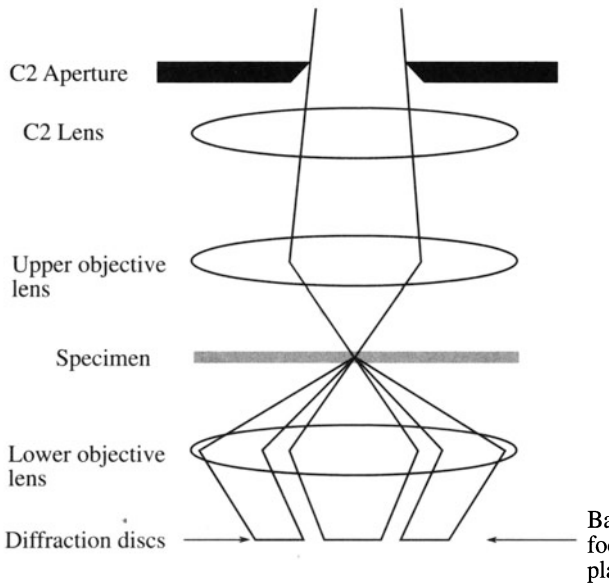


Figure 20.1. Ray diagram showing CBED pattern formation. A convergent beam at the specimen results in the formation of disks in the BFP of the objective lens.

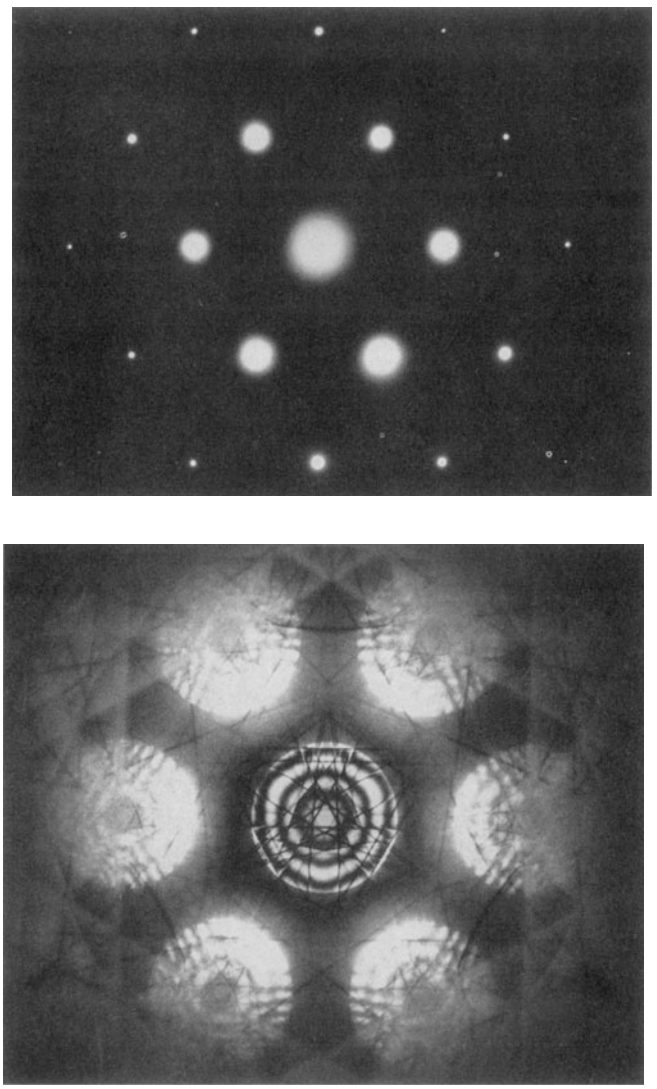


Figure 20.2. (A) SAD pattern from [111] Si showing the first few orders of diffraction spots but no Kikuchi lines. (B) CBED pattern from [111] Si showing dynamical contrast within the disks as well as Kikuchi and other lines.

From Figure 20.1 you can see that we need a strong upper-objective polepiece to create a convergent beam, so any probe-forming instrument, such as a TEM/STEM or dedicated STEM, can generate the patterns. Before the development of STEMs, CBED was possible by the addition of a mini-lens below the C2 lens of a conventional TEM, but then parallel-beam TEM imaging was impossible. We've already seen the detailed 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. We'll start with TEM mode and then describe STEM operation.

20.2.B. CBED in TEM

You can form CBED patterns in any TEM that is capable of creating a small (<< 1 μm) beam with a convergence semiangle (α) >10 mrads. This might not be possible on TEMs made before the late 1970s that do not have condenser-objective lenses, so you should check if your TEM is properly equipped before you spend a lot of time trying to get it to do something which won't be possible.

There are four microscope variables you need to control when forming a CBED pattern:

- The beam convergence semiangle α.
- The camera length (*L*) (i.e., the magnification).
- The focus of the pattern.
- The size of the beam.

When you focus the beam, you probably won't be able to see any useful image information, just a bright spot on the screen, but if the TEM is well aligned 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, and form a focused image on the TEM screen with the area that you want to examine approximately in the middle of the screen.
- Select a large C2 aperture about 100–200 μm in diameter, carefully center it, then adjust the C2 lens to form a focused spot on the area of interest.
- Keep C1 weakly excited to give a relatively large spot, about 100–200 nm FWTM (see Chapter 5), containing sufficient current to give high intensity in the 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 apertures 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.

20.2.C. Choosing the C2 Aperture

Once the CBED pattern is visible you can adjust the convergence semiangle 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.

> The pattern of nonoverlapping disks is a Kossel–Möllenstedt (K–M) pattern.

To get such a pattern you must select a C2 aperture such that $2\alpha < 2\theta_B$ 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; we usually operate in this mode. If 2α 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" may be used (although this can cause confusion with the classical use of the term for geometrically similar X-ray patterns). Figures 20.3A–C show a series of

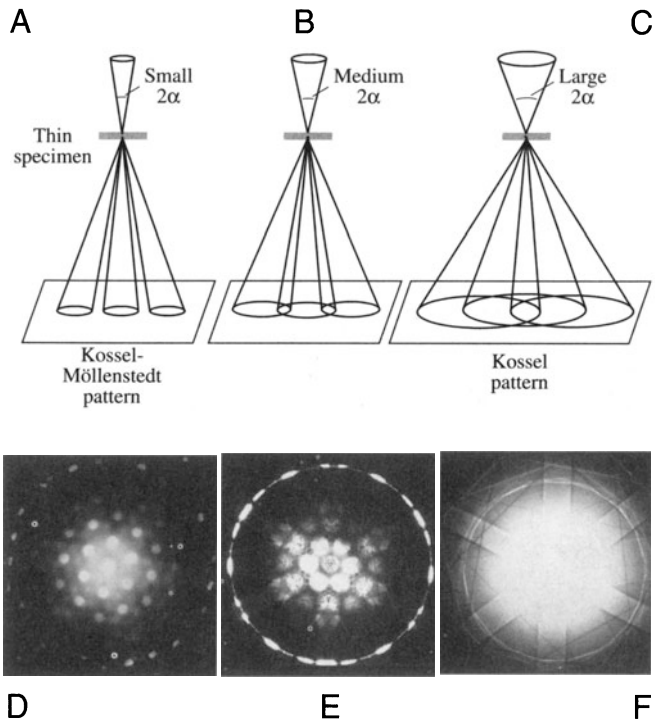


Figure 20.3. (A)–(C) Ray diagrams showing how increasing the C2 aperture size causes the CBED pattern to change from one in which individual disks are resolved to one in which all the disks overlap. In (D)–(F) you can see what happens to experimental patterns on the TEM screen as you select larger C2 apertures.

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 Figures 20.3D–F. The patterns in Figure 20.3 were all taken at a small camera length and you can see rings of intensity which arise from electrons scattered to quite high angles (±10°). We'll return to these HOLZ diffraction effects in Section 20.3.

Kossel patterns are most useful when viewed with a small camera length (L) because they display an enormous area of reciprocal space, and the large 2α gives rise to strong Kikuchi bands. The Kikuchi bands intersect at the zone axes in the center of the pattern, as you can see in Figure 20.3F, and therefore it is very easy for you to tilt to a particular zone axis. So, to form a ZAP, it is best to start at very small L with a large 2α. Later, you can worry about the best choice of C2 aperture and focusing the pattern.

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 of three C2 apertures comprises one of about 200 μm for routine TEM, EELS, and Kossel patterns, a 50–70 μm ultrathick 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, but if you do, the objective lens has to be changed also 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.

> Use the specimen height (z) control to maintain the specimen in the eucentric plane as you tilt. A computer-controlled stage is ideal for this.

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.

20.2.D. Choosing the Camera Length

The choice of L depends on the information that you want to obtain from the pattern. (Remember that L controls the magnification of the DP and a *large* L gives a *high* magnification view, but only spans a *small* angular range.) Typically we choose $L > 1500$–6000 mm to observe detail in the 000 (BF) disk at the highest possible magnification, and $L < 500$ mm to view the low magnification pattern, sometimes called the "whole pattern" (WP), containing electrons scattered to high angles. Figure 20.4 shows a series of CBED patterns obtained over a range of L and you can see that if we start at a high L (Figure 20.4A) we first see only the 000 disk and then the disks that are equivalent to an SAD pattern, but at smaller L the HOLZ effects that we just mentioned become visible (Figure 20.4C). Ultimately, at the smallest L, the angular range of the pattern is limited by the bore of the objective lens polepiece.

20.2.E. Focusing Your Pattern

If you don't focus your patterns you will miss a lot of the fine detail! First, the beam has to be focused in the speci-

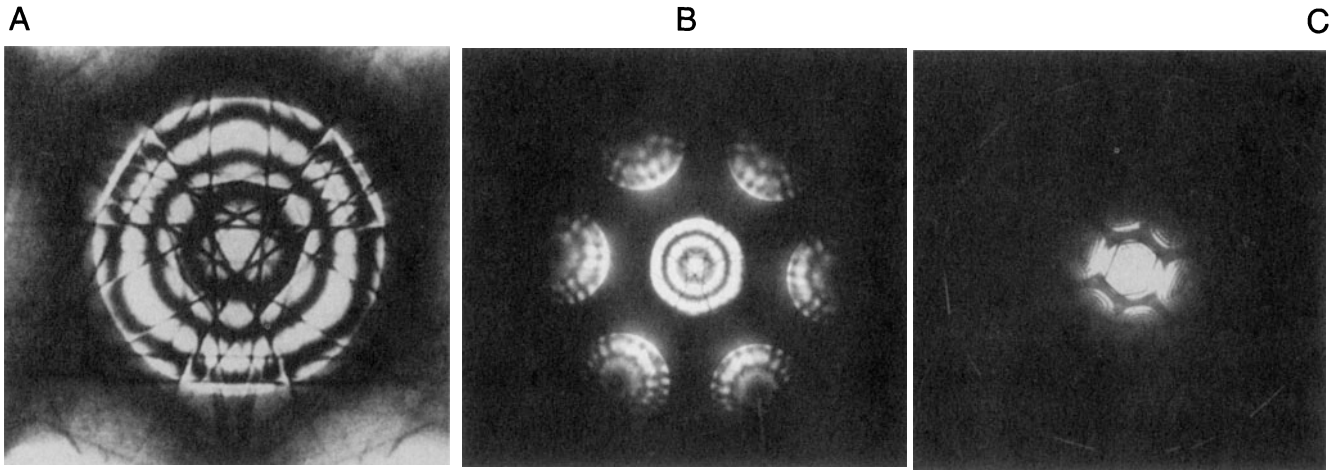


Figure 20.4. Decreasing the camera length expands our view of reciprocal space. (A) Starting at high L with a CBED pattern containing the 000 diffraction disk we then begin to see in (B) the distribution of electrons in the zero-order Laue zone. At the lowest camera lengths (C) the higher-order Laue zone is faintly visible. Typically, we can record electrons scattered over an angular range of ±10°.

men rather than underfocused or overfocused. The easiest way to find focus in TEM mode is to leave the objective lens alone and adjust the C2 lens to form the smallest spot on the TEM screen. Changing C2 thus also changes α, and if you want to maintain a fixed α then you need to adjust the objective lens strength as follows:

- Select K–M conditions and choose a value of L so you can clearly see the 000 disk.
- Deliberately overfocus (strengthen) the objective lens until a BF image is visible in the disk. This is because the beam is spread at the plane of the specimen (see Figure 20.5A and the associated ray diagram back in Figure 6.5A).
- Weaken 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 (Figure 20.5B and Figure 6.5B).
- Underfocus, and again you can see a BF image in the 000 disk, inverted with respect to the overfocused image (Figure 20.5C and Figure 6.5C). As you can see in Figure 20.5B there is nonspatial (i.e., diffraction-contrast) information in the 000 disk when you are at focus. 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.

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 20.5. If you use the second (noneucentric) tilt axis, or move to another region of the specimen, you will probably have to refocus the pattern with the z-control.

The CBED pattern also has to be correctly focused in the back focal plane and you can do this in the conventional manner using the intermediate lens fine focus to sharpen the image of the C2 aperture.

Note that this operation is equivalent to creating multiple DF images with a parallel beam, which we used to calibrate the SAD pattern rotation in Section 9.6.C.

20.2.F. Choice of Beam Size

The last TEM variable is the beam size. We've already mentioned that you should start with a reasonably large beam 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

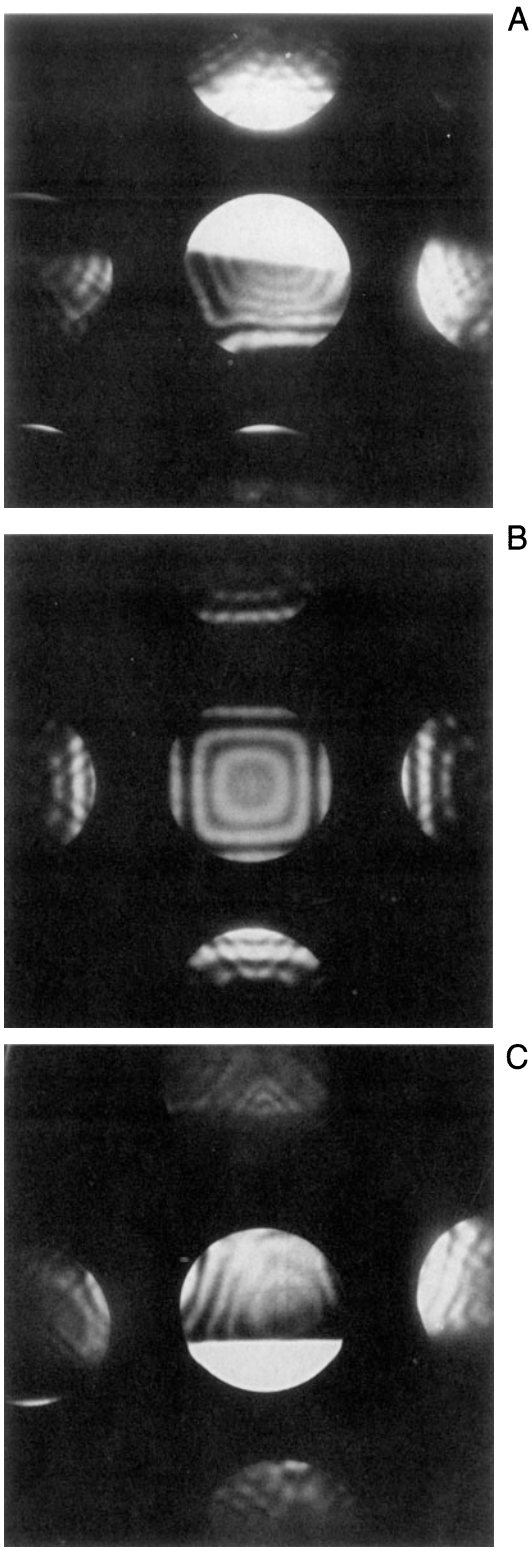


Figure 20.5. The procedure for correctly focusing the CBED pattern by adjusting the strength of the objective or C2 lens. In both overfocus (A) or underfocus (C) conditions you see a BF image in 000 and DF images in the $hk\ell$ disks, but at exact focus (B) the disks contain nonspatial dynamical-diffraction contrast.

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 scatter will spread the beam and degrade the resolution in a manner identical to XEDS (see Chapter 36). Using the thinnest specimens and an FEG, CBED patterns can be obtained from extraordinarily small regions, as we'll see in Section 21.8. Figure 21.15 shows that subnanometer diffraction is possible. However, there is a drawback to using the thinnest specimens because they don't exhibit dynamical-diffraction effects, as you'll now see.

20.2.G. Effect of Specimen Thickness

If your specimen is very thin you may have kinematical-diffraction conditions. Then the diffraction disks are uniformly bright and devoid of contrast, as shown in the ZAP in Figure 20.6A. Moving to a thicker area of the specimen in the same orientation transforms the pattern from an array of uniformly intense disks to a display of strong dynamical contrast (Figure 20.6B), which we'll discuss later. So to get the most out of a CBED pattern the specimen should be thicker than one extinction distance (see Chapter 16). This requirement differs from that of many other TEM techniques, such as HRTEM, XEDS, and EELS, where the best information is obtained from the thinnest specimens. So with CBED, you can almost always get something out of your specimens, even if they are too thick for anything else!

20.2.H. Final Adjustment

Sometimes it is quite difficult to make the ZAP exactly symmetrical as in Figure 20.2B. It often seems as if the last 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 symmetric 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 only be a last resort.

As with SAD patterns, a range of exposure times for all CBED patterns will give you the most information.

20.2.I. 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 then quite simple:

- First stop the beam scanning (i.e., select "spot" mode on the STEM console).
- Position the spot on the STEM screen on the region of interest.

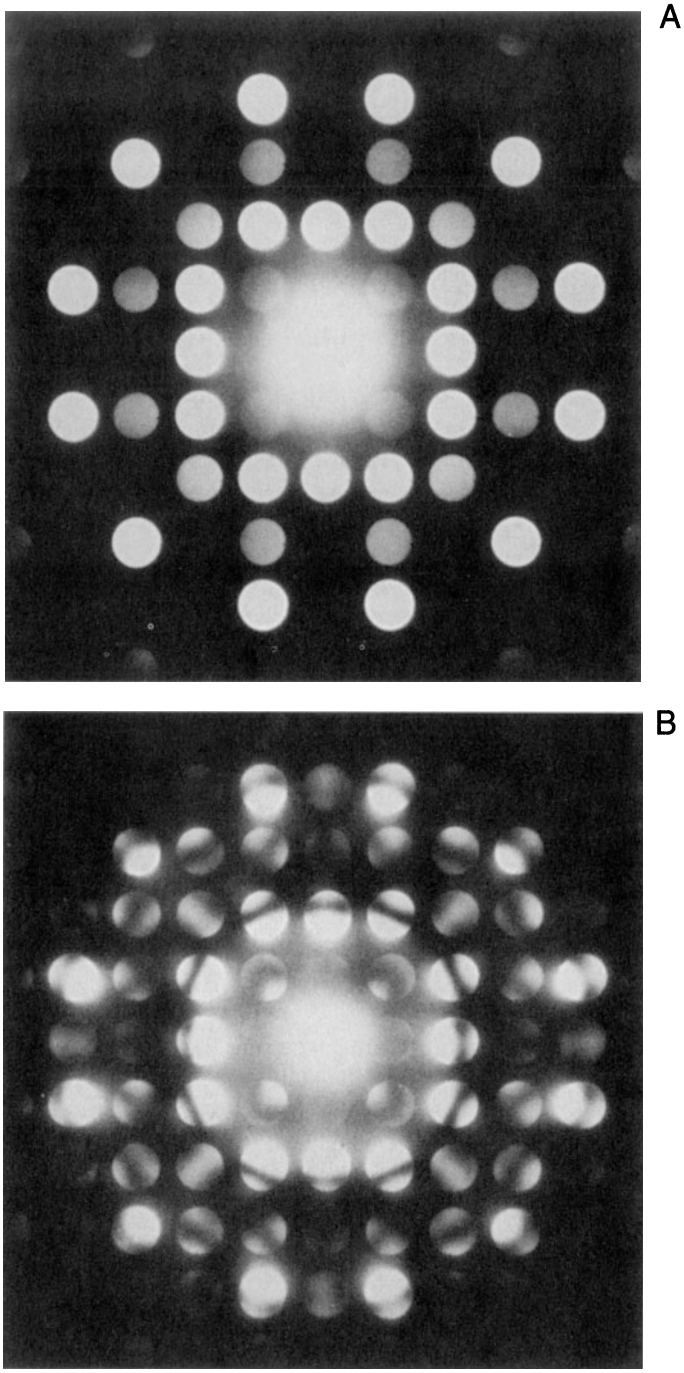


Figure 20.6. (A) [001] CBED pattern from σ phase in 316 stainless steel under kinematical conditions. Such patterns give us no more information than SAD. Their only advantage over SAD is that they come from a smaller region of the foil. (B) CBED pattern from a thicker area than in (A) showing dynamical-contrast phenomena.

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 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.

> 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 looking at the back focal plane of the objective lens, or by scanning the pattern across the BF detector using post-specimen scan coils (see Section 21.7).

The choice of operating mode then is really up to you; TEM and STEM both have their advantages. 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.
- Decrease L to see the full pattern including HOLZ scatter, and tilt to the desired orientation.
- Adjust the convergence semiangle with the C2 aperture.
- Increase the beam size if necessary with the C1 lens to make the pattern brighter.
- Increase L to look at the 000 disk and focus the pattern.

20.3. ZERO-ORDER AND HIGHER-ORDER LAUE-ZONE DIFFRACTION

20.3.A. ZOLZ Patterns

If you increase L above ~800 mm you will only see the first few diffraction maxima, as shown in Figure 20.2B. The CBED pattern consists of disks similar to the array of spots in an SAD pattern, i.e., discrete diffraction maxima surrounding the central 000 disk. Remember that such a pattern is termed a ZOLZ pattern since the permitted $hk\ell$ diffraction maxima must all satisfy the Weiss zone law relationship $hU + kV + \ell W = 0$, where UVW is the beam direction. Remember also that the $hk\ell$ maxima all lie in the reciprocal lattice plane containing the origin 000 of the reciprocal lattice, and this plane is also called the ZOLZ. So in fact SAD patterns 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 and the $hk\ell$ maxima can be indexed and UVW identified, in exactly the same manner as for an SAD pattern. The two options, as we described in Section 18.4, are the method of ratios or a calibration standard to determine L.

Because of the finite size of the diffraction disks you must take care to select equivalent points in each disk when measuring the $hk\ell$ spot spacings. If α is too large, you might not see individual maxima and you should then select a smaller C2 aperture (K–M conditions).

20.3.B. HOLZ Patterns

The central bright portion of the CBED pattern is due to relatively intense low-angle scattering. At higher angles, the ZOLZ intensity drops because the atomic scattering amplitude, $f(\theta)$, has decreased. However, the intensity increases when the Ewald sphere intercepts the HOLZ planes in the reciprocal lattice and a ring of diffracted intensity is observed, as in Figures 20.3D–F and 20.4C.

> Remember that the radial distance from 000 in a DP is related to the angle of scatter; use a smaller L to see higher-angle scattering.

If you've chosen a small convergence semiangle, 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 weak high-angle diffraction from crystal planes that are *not* parallel to the beam. Low temperatures increase the HOLZ scatter and also minimize the thermal diffuse (phonon) scatter that in some materials masks the weak HOLZ intensity. So you'll find a liquid-N_2 cooling holder is often essential.

Consider the intersection of the Ewald sphere with the reciprocal lattice. The HOLZ planes in the reciprocal lattice cross the sphere, unlike the zero layer which is tangential to it, and the intersection of a sphere with a plane creates a ring. The first ring is called the FOLZ, because the possible $hk\ell$ reflections satisfy the relationship $hU + kV + \ell W = 1$, and so on. Where the Ewald sphere intersects these HOLZs, diffracted intensity is expected (taking into account the usual structure-factor effects; see 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 in each HOLZ relrod is sampled, as shown in Figure 20.7A. This range of angles manifests itself in the CBED disk, sampling the intensity distribution along the relrod, as shown in Figure 20.7B. Different interception points on the relrod correspond to different points in the disk, also shown in Figure 20.7B. Figure 20.7C shows a typical experimental distribution of diffraction maxima from the Ewald sphere construction in Figure 20.7A. We've already shown similar experimental patterns, such as Figure 20.4C.

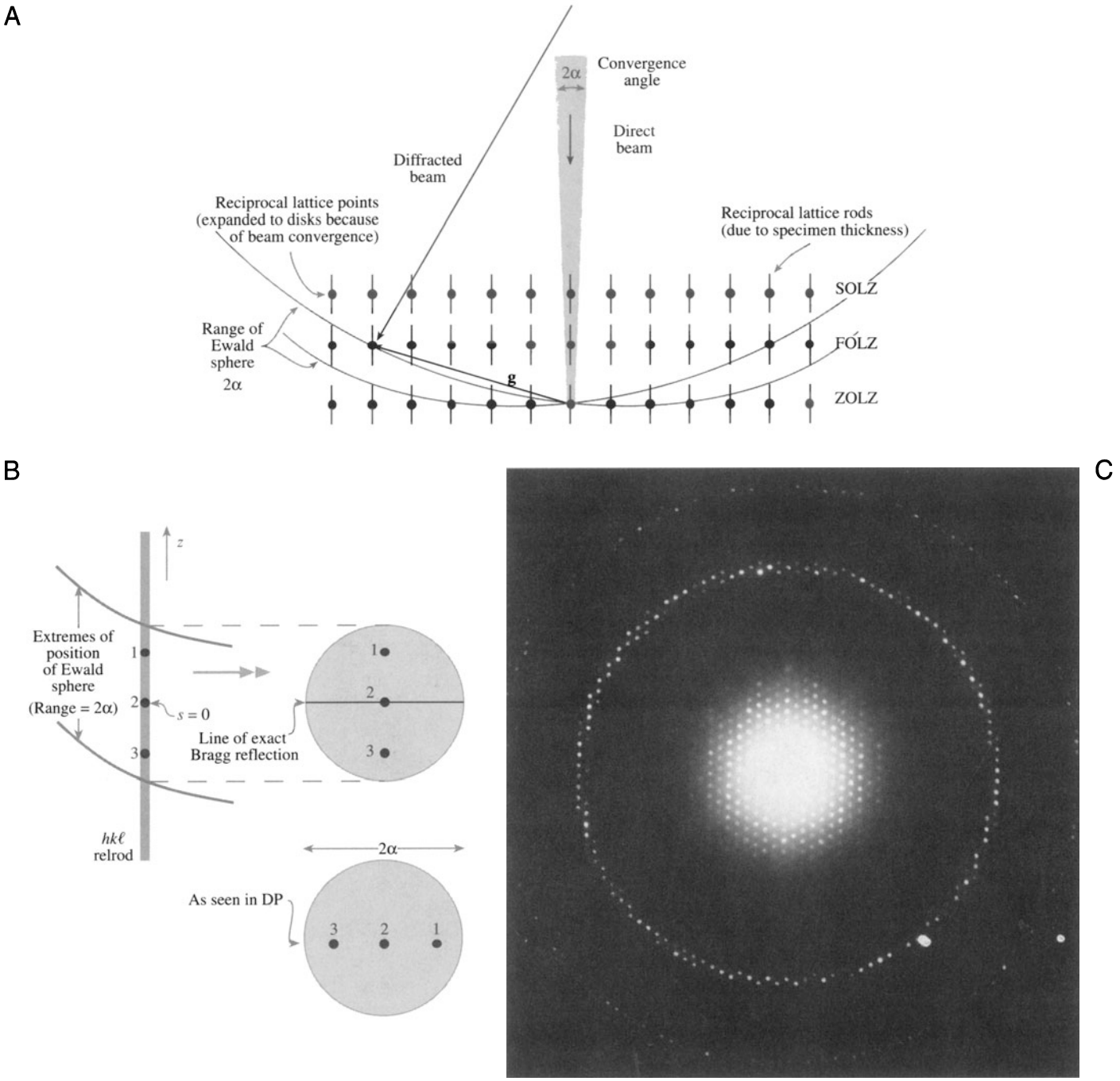


Figure 20.7. (A) The Ewald sphere can intercept reciprocal lattice points from planes not parallel to the electron beam whose **g** vectors are not normal to the beam. The sphere has an effective thickness of 2α because of beam convergence and so intercepts a range of these HOLZ reciprocal lattice points. 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 $hk\ell$ disk. 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 experimentally in (C).

The most important point to remember immediately 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.

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 back focal plane permitted by the imaging system (±10°). As shown schematically in Figure 20.7A, the Ewald sphere only intercepts reciprocal lattice points in HOLZs many orders of diffraction maxima away from the beam. Since the scattering into HOLZ spots is weak, the exposure time to reveal HOLZ maxima is usually long enough to ensure that the ZOLZ is overexposed on the negative (see Figure 20.7C). Therefore, you may have to record two DPs: a relatively short exposure for the ZOLZ pattern containing only two-dimensional 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 here:

- You can use careful photographic processing (Turner and Krishnan 1987) but you must plan this in advance.
- You might be lucky with your thin area; sometimes you can produce reasonable ZOLZ and FOLZ intensity on the same exposure (e.g., look ahead to Figure 20.16).
- A CCD camera will give a greater dynamical range, but perhaps with some minor loss of resolution.
- You can use image processing techniques to combine differently exposed patterns (see Chapter 30).

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 20.8.) If you still can't see a HOLZ ring, then the last thing you can try is increasing λ by lowering the kV.

In the next chapter, we will show you how HOLZ-ring measurements can be used to deduce the lattice-repeat vector of the crystal parallel to the beam direction. You can then determine the unit cell, the crystal system, and also the type of lattice centering.

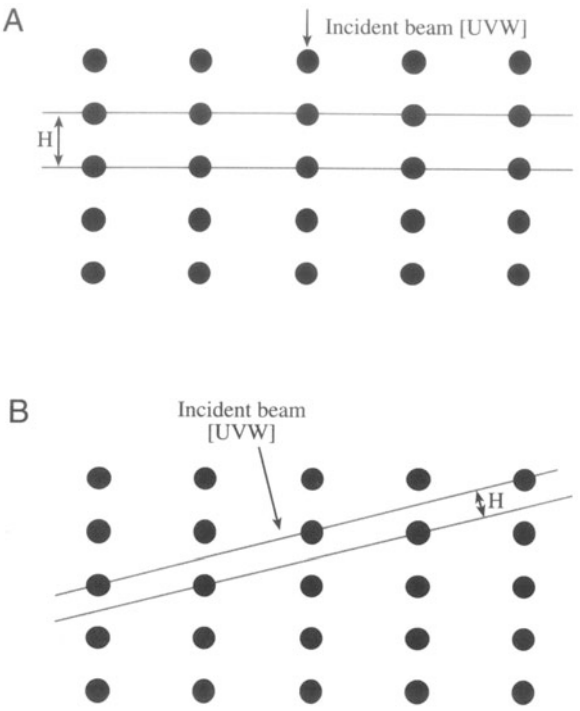


Figure 20.8. (A) The reciprocal lattice spacing (H) is large if the beam is down a major zone axis in the crystal. (B) The spacing is small if the beam is down a low-symmetry direction.

20.3.C. Indexing HOLZ Patterns

If you want to index the individual HOLZ reflections:

- Index the ZOLZ for which $hU + kV + \ell W = 0$ (see Section 18.4).
- Consult a stereographic projection to identify the poles of the principal planes constituting the FOLZ ($hU + kV + \ell W = 1$) and SOLZ ($hU + kV + \ell W = 2$), etc.
- Alternatively, you can just solve the Weiss zone law 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 stereographic projections, see Chapter 18. Remember that the stereographic projection just gives you the major low-index $hk\ell$ planes and ignores any systematic absences.

Examples of indexed ZOLZ, FOLZ, and SOLZ patterns are shown in Figures 20.9A–C for the fcc lattice under (A) [001], (B) [110], and (C) [111] beam directions; Figures 20.10A–C show similar patterns for the bcc lattice.

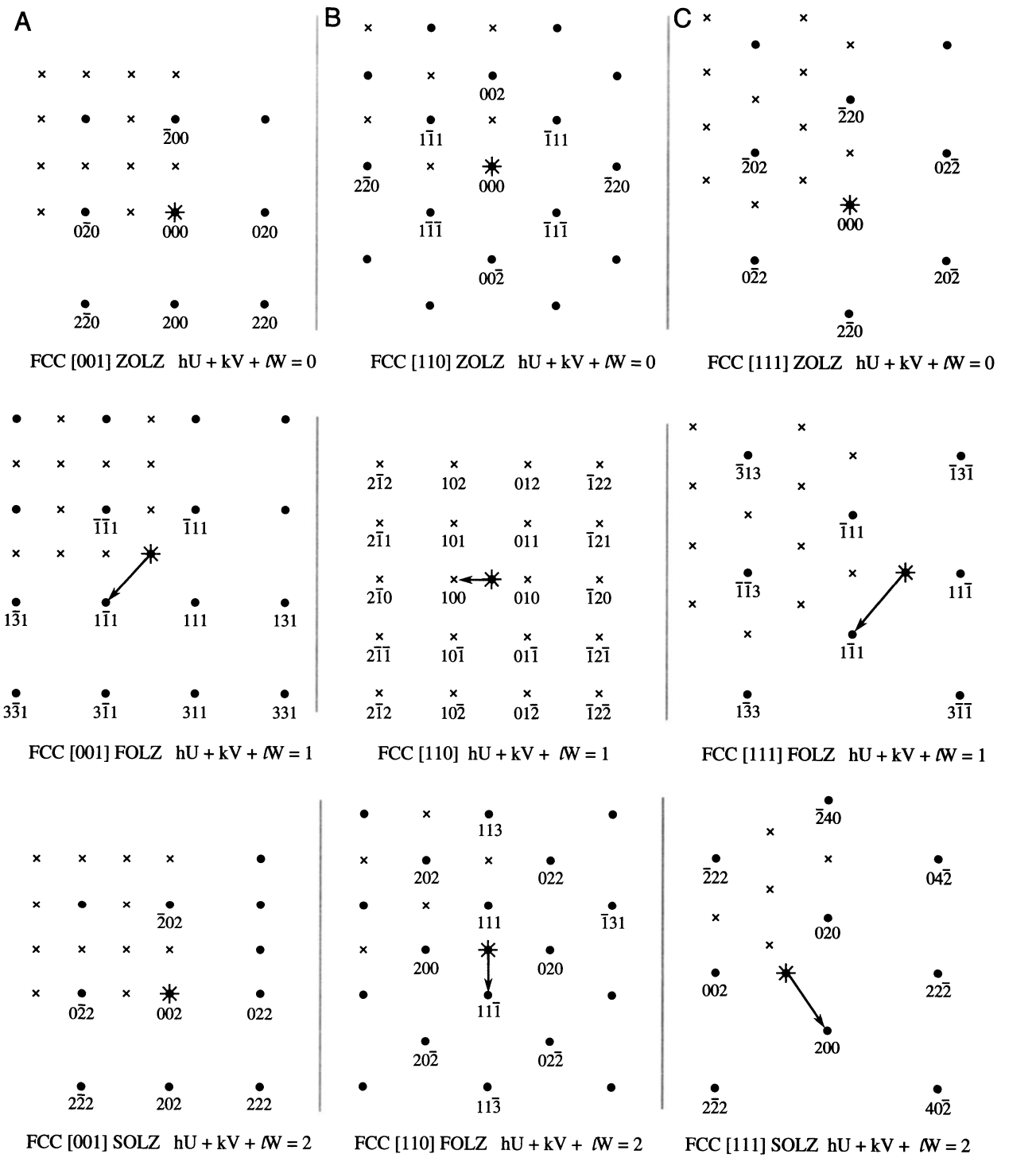


Figure 20.9. The possible ZOLZ, FOLZ, and SOLZ reflections for the three principal zone axes of an fcc specimen. Allowed reflections are shown as dots, forbidden reflections as crosses, the direct-beam direction is a star, and the arrow indicates the displacement vector between the ZOLZ and FOLZ. Note that when all reflections are forbidden for hU + kV + lW = 1 the FOLZ has hU + kV + lW = 2.

Only the first few orders of diffraction maxima are shown in each case. In practice, relatively high orders of diffraction maxima are present in the HOLZs, and the schematic patterns should be extended accordingly to match up with the experimental patterns (see Ayer 1989). From the schematic HOLZ patterns in Figures 20.9 and 20.10 you can see that the symmetry of each *UVW* zone is retained in each HOLZ pattern, but the HOLZ patterns are often shifted by a displacement vector relative to the ZOLZ because there is no allowed reflection on the zone axis. This

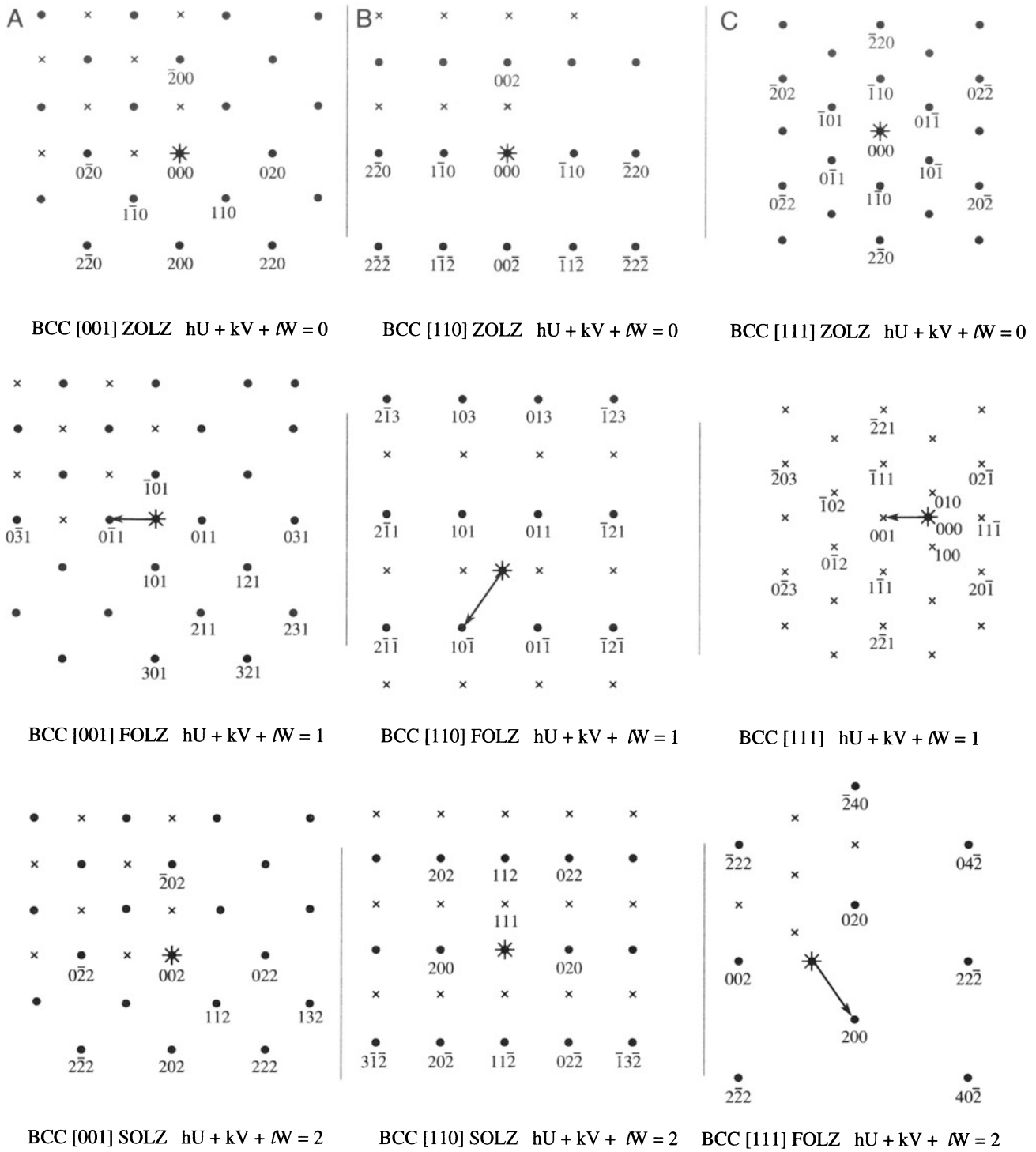


Figure 20.10. The possible ZOLZ, FOLZ, and SOLZ reflections for the three principal zone axes of a bcc specimen. Allowed reflections are shown as dots, forbidden reflections as crosses, the direct-beam direction is a star, and the arrow indicates the displacement vector between the ZOLZ and FOLZ. Note that when all reflections are forbidden for hU + kV + lW = 1 the FOLZ has hU + kV + lW = 2.

displacement can be calculated for any zone axis using equations which we'll discuss in the next chapter.

You should generate similar diagrams for the major zone axes of any specimen that you are going to study by CBED techniques; several computer programs, including EMS and Digital Microscopist, can do this readily (Section 1.5). There is also freeware available on the WWW. You may have to generate less symmetrical patterns (e.g., [233] or [114]) since these give rise to HOLZ rings with smaller radii, in which HOLZ phenomena are easier to see.

Be aware when you are indexing HOLZ rings that structure factors can cause every reflection in a reciprocal-lattice layer to be forbidden.

Under these circumstances the first ring of spots to be observed is from the second layer of the reciprocal lattice *but it is still called the FOLZ.* Well-known examples are the $(110)_{fcc}$ and $(111)_{bcc}$ patterns. You can't predict the total 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 if the [001] beam direction is observed, but in other directions in this system, e.g., [121], [141], and [542], only every third Laue zone is present. A detailed explanation of this is given by Raghavan *et al.* (1984).

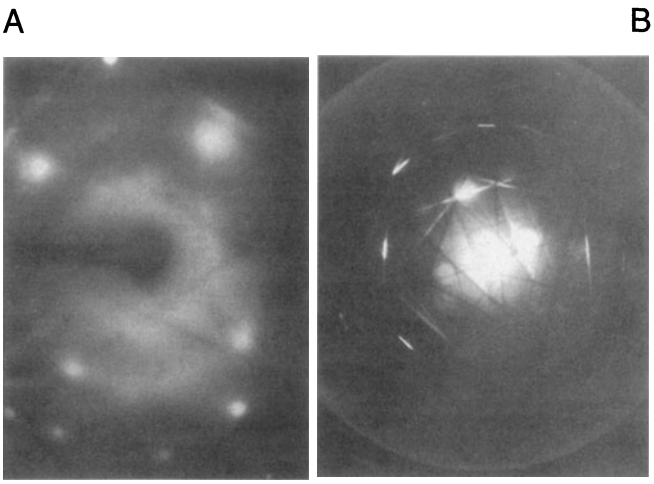


Figure 20.11. (A) Comparison of the poor quality of Kikuchi lines in an SAD pattern and (B) the relatively clear distribution in a CBED pattern from highly deformed copper.

20.4. KIKUCHI LINES IN CBED PATTERNS

In CBED patterns you almost invariably see sharp Kikuchi lines, while in SAD patterns Kikuchi lines are often rather diffuse or absent (see Chapter 19). This difference arises mainly 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. This effect is shown in Figure 20.11A, which is a conventional SAD pattern containing very diffuse Kikuchi lines. This pattern was obtained from a large region of deformed copper. By comparison, Figure 20.11B 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 (Heilman *et al.* 1983).

If the CBED pattern is not a ZAP, as in Figure 20.11B, the Kikuchi lines appear as pairs of excess (bright) and deficient (dark) lines, as in SAD patterns. But when you obtain a ZAP pattern, the ZOLZ Kikuchi lines appear as bright bands. These bands increase in intensity and definition as you increase the convergence semiangle, as shown in Figures 20.3D–F. A similar effect is seen in channeling patterns in the SEM which are generated by rocking a parallel beam around the optic axis. If you need to understand the difference between Kikuchi lines in SAD patterns and

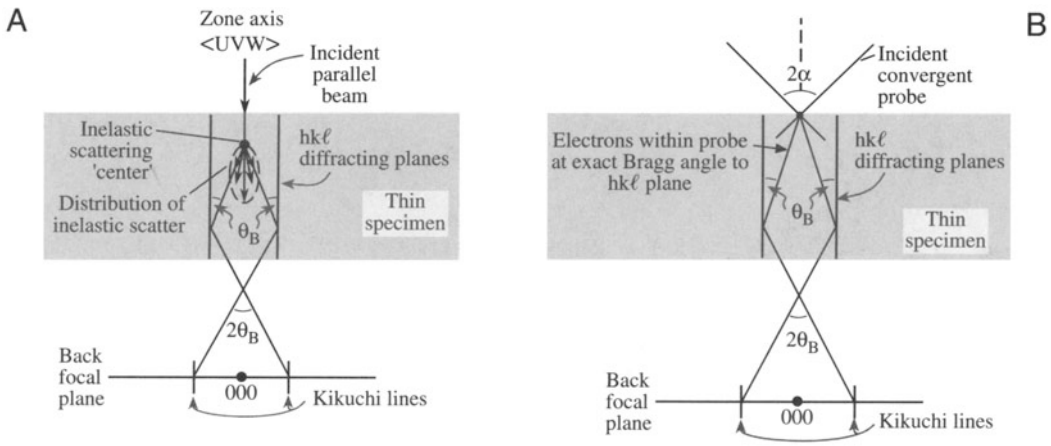


Figure 20.12. Comparison of the generation of Kikuchi lines (A) by inelastic scatter of electrons in a parallel beam and (B) by elastic scatter of electrons in a convergent beam when the convergence semiangle, α, is greater than the Bragg angle, θ_B.

Kikuchi bands in CBED patterns, Reimer (1993) gives a discussion.

The generation of Kikuchi lines in a CBED pattern is marginally more complex than in an SAD pattern. Remember how Kikuchi lines arise in a specimen illuminated by a parallel beam (see Chapter 19). In Figure 20.12 you can see what happens when a convergent beam is used. In this case the incident electrons span an angular range and therefore some electrons in the beam may already be at the Bragg angle to a ZOLZ plane. Thus, there will be an elastic-scattering contribution to the Kikuchi lines where they cross the disks in CBED patterns. If you choose Kossel conditions (i.e., $2\alpha > 2\theta_B$) as in Figure 20.12, 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, you should only use the term "Kikuchi lines" for the situation where inelastic scatter alone is responsible for their formation (i.e., the lines between any $hk\ell$ 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.

20.5. HOLZ LINES

20.5.A. The Relationship between HOLZ Lines and Kikuchi Lines

Kikuchi lines also arise from inelastic scattering by the HOLZ planes and HOLZ Kikuchi lines exist in many CBED patterns. You can see the criss-cross 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 more sensitive to changes in lattice parameter than the ZOLZ lines. Since

$$|\mathbf{g}| = \frac{1}{d}, \quad |\Delta\mathbf{g}| = -\frac{\Delta d}{d^2} \qquad [20.1]$$

So the smaller d, the larger $|\Delta\mathbf{g}|$ 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 part of the line which lies *within* the diffraction disks. By analogy with the production of Kikuchi lines, you can imagine that the lines arise when electrons within the incident beam at the correct Bragg angle for diffraction by a HOLZ plane are scattered out to high angles, creating a bright line through the HOLZ disk and leaving a dark line in the 000 disk. Not surprisingly, the theory for the origin of HOLZ lines is much more complicated than this. When you have time, you should read the paper by Jones *et al.* (1977).

As you see, the HOLZ lines come in pairs, like Kikuchi lines, with the bright (excess) lines within the HOLZ $hk\ell$ disks and the dark (deficient) lines within the 000 disk; an example of deficient lines is shown schematically in Figure 20.13. Because HOLZ lines contain 3D information, they show the true fcc threefold {111} symmetry while the ZOLZ Kikuchi lines and spots show sixfold two-dimensional {111} symmetry. We will make use of such differences when we discuss crystal-symmetry determination in Chapter 21.

20.5.B. Acquiring HOLZ Lines

Steeds (1981) has dealt with the practical problems of recording HOLZ lines in some detail. 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.
- You may have to make changes in the operating voltage or the orientation in order to view the HOLZ rings, especially if the crystal has a small reciprocal-lattice-repeat spacing parallel to the beam so the angular view of the back focal plane is poor.

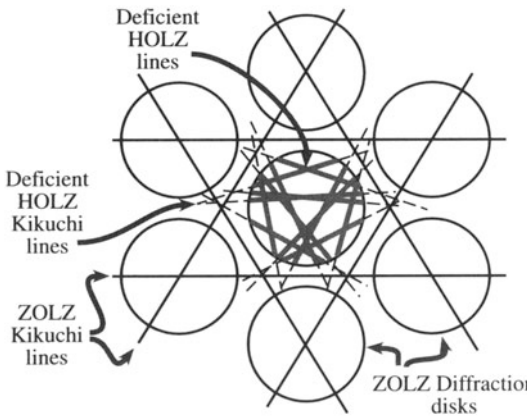


Figure 20.13. The relationship between Kikuchi lines and HOLZ lines is shown in this schematic of a [111] CBED pattern from an fcc crystal. The three principal pairs of $2\bar{2}0$ 110 ZOLZ Kikuchi lines show sixfold symmetry and bisect the **g**-vectors from 000 to the $2\bar{2}0$ ZOLZ disks. The inelastic HOLZ deficient Kikuchi lines are shown in the region between the ZOLZ diffraction disks and the elastic HOLZ deficient lines are present within the 000 disk only. Compare this schematic with the experimental pattern back in Figure 20.2B.

- 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, but will lower the overall intensity of the pattern.
- Planar or point disorder as well as thermal effects can restrict large-angle scattering. In practice, this means that cooling the specimen and reducing the kV can help to increase HOLZ-line visibility; cooling also helps to reduce 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. So continuous kV control is an essential accessory if you're going to do serious CBED work.

The experimental procedure for observing HOLZ lines is quite straightforward, but since the lines themselves can be rather elusive, you should practice with a specimen such as silicon 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 (largest 2α) and go to the smallest L at which you can see the full angular view of the back focal plane (about 500 mm).
- Examine the Kossel pattern, which should reveal Kikuchi line pairs intersecting at many poles, spanning a good fraction of the stereographic triangle as shown in Figure 20.14A.
- Tilt to a suitable zone axis. Remember, the best orientation for seeing the HOLZ lines is *not* a low-index, high-symmetry pole such as <100> or <111>, shown in Figure 20.14B, but a higher-index, lower-symmetry one such as <114>. One such pattern is shown back in Figure 20.3F.
- When you have tilted to a zone-axis orientation, you should see the ring of excess HOLZ lines.
- To see the deficient lines, increase L to look in detail at the 000 region of the pattern, and if necessary put in a smaller C2 aperture, center it, and look for the fine dark lines criss-crossing the bright disk as in Figure 20.14C. Usually you just need the deficient line distribution.

You have to use a range of L (from ~300 mm to 1500 mm) to obtain all this information, and this flexibility is only available on TEMs with more than three imaging lenses. DSTEMs with sufficient post-specimen lenses also offer this versatility.

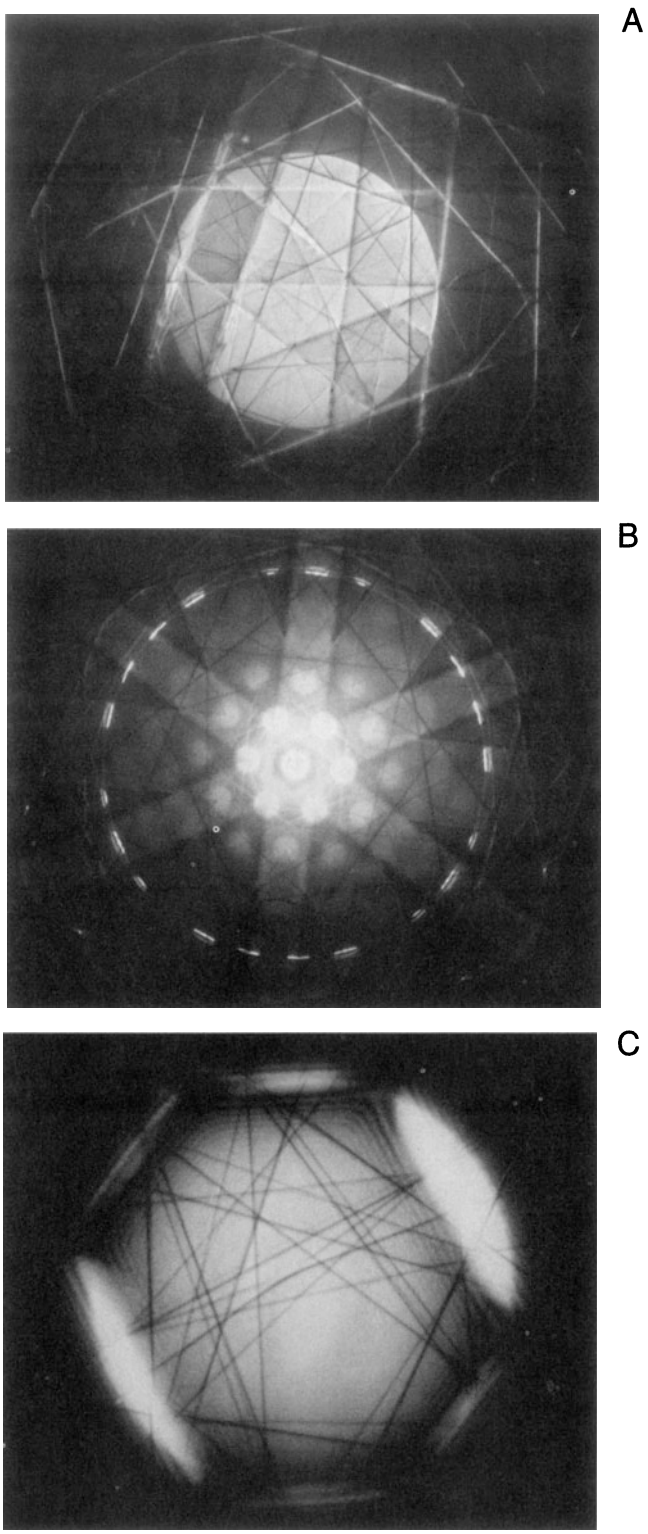


Figure 20.14. (A) Low L, large α CBED pattern showing a wide area of reciprocal space, away from ZAP conditions. (B) When the specimen is tilted to a low-index ZAP, and a smaller C2 aperture is inserted, a ring of excess HOLZ lines appears associated with the HOLZ disks. (C) Deficient HOLZ lines are visible in the central 000 disk of a low-symmetry 114 ZAP taken at high L.

20.5.C. 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 HOLZ disks: the first thing you have to do is index the HOLZ $hk\ell$ maxima in the manner we described in Section 20.4. 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 HOLZ maximum. There should be a parallel HOLZ deficient line in the 000 disk, and this line can be assigned the same indices as the HOLZ maximum. If you repeat this exercise around the HOLZ ring most of the HOLZ lines should be indexed, as shown in Figure 20.15. 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 HOLZ maximum (Ecob *et al.* 1981).

In some cases you may find it difficult to associate a specific HOLZ deficient line in 000 with a specific HOLZ reflection due to strong excitation of two diffraction 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 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. The indexing of HOLZ lines lends itself to computer assistance and several programs have been described in the literature or are available on the WWW (for a review of which software is best, see Eades *et al.* 1993). 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, in the next chapter.

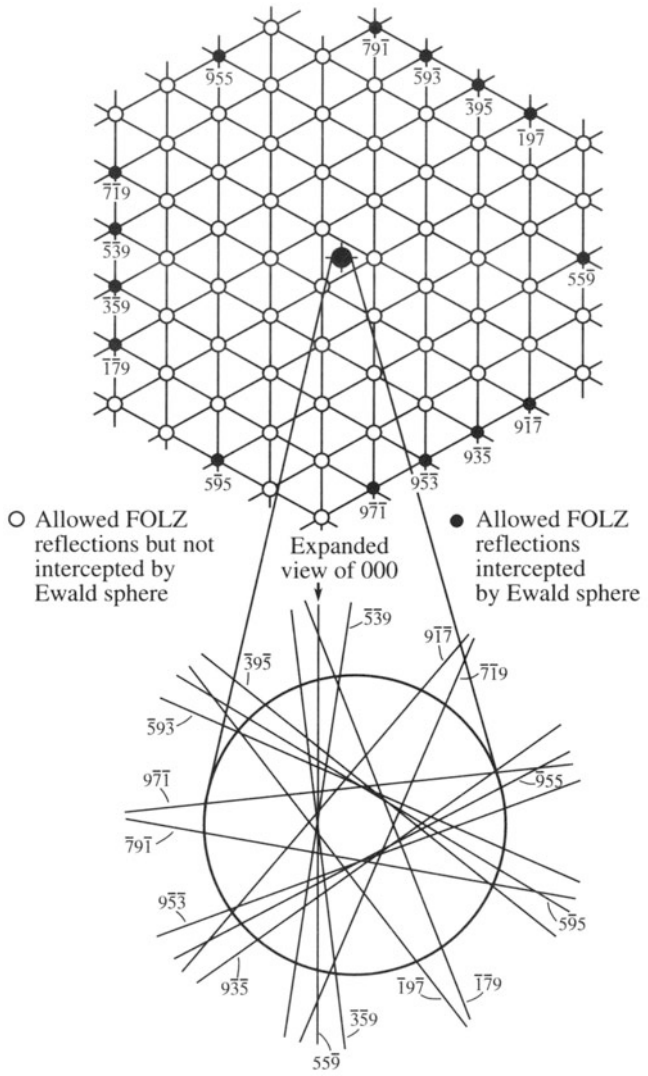


Figure 20.15. How to relate deficient HOLZ lines to the HOLZ maxima; the indexed FOLZ reflections in this [111] pattern are shown as full circles and the open circles are the rest of the FOLZ reciprocal lattice points that don't intercept the Ewald sphere. The **g**-vector from 000 to each $hk\ell$ FOLZ disk is normal to the $hk\ell$ HOLZ line and the lines are shown in the expanded 000 disk below.

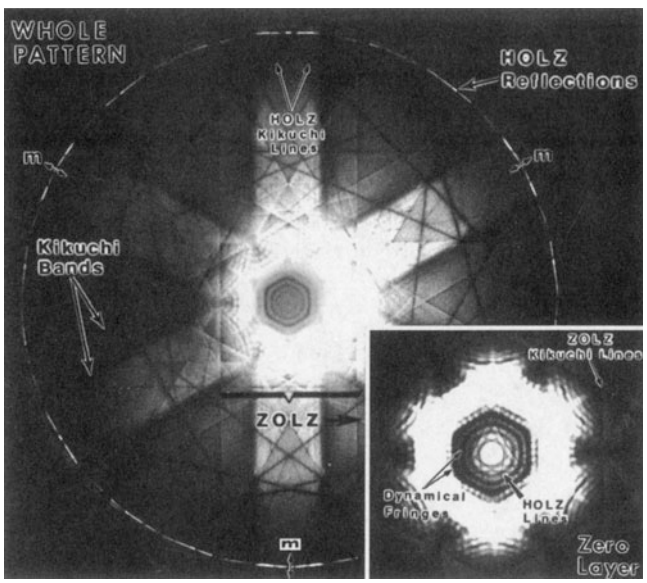


Figure 20.16. CBED pattern from Cu showing all the important characteristics. The ZOLZ pattern is visible, and the ZOLZ Kikuchi line pairs show six fold symmetry. The ring of excess HOLZ-line intensity can be seen. Deficient HOLZ lines in the 000 disk in the inset show only threefold symmetry and this symmetry difference will be exploited in the next chapter.

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, and L, then you can obtain CBED patterns showing many different features, particularly when the beam is down a zone axis.
- It is necessary to record patterns at different camera lengths with different exposure times.
- It is strongly recommended that you use a double-tilt cooling holder.
- You must be able to change the kV by very small steps, if you are studying HOLZ lines.
- Learn the meaning of such terms as ZAP, ZOLZ, FOLZ, HOLZ, and K–M and Kossel patterns.

If you look at Figure 20.16 you'll see many of the essential features of CBED displayed in one highly symmetrical pattern. In the next chapter we'll show you how to use this knowledge along with other aspects of CBED to get the maximum amount of crystallographic information from your specimen.

REFERENCES

General References

Loretto, M.H. (1994) *Electron Beam Analysis of Materials*, 2nd edition, Chapman and Hall, London.

Steeds, J.W. (1979) *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 387, Plenum Press, New York.

Sung, C.M. and Williams, D.B. (1991) *J. Electron Microsc. Tech.* **17**, 95.

Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd edition, p. 117, Philips Electron Optics Publishing Group, Mahwah, New Jersey.

Specific References

Ayer, R. (1989) *J. Electron Microsc. Tech.* **13**, 16.

Eades, J.A., Moore, S., Pfullman, T., and Hangas, J. (1993) *Microscopy Research and Technique*, **24**, 509.

Ecob, R.C., Shaw, M.P., Porter, A.J., and Ralph, B. (1981) *Phil. Mag.* **A44**, 1117.

Heilman, P., Clark, W.A.T., and Rigney, D.A. (1983) *Acta Met.* **31**, 1293.

Jones, P.M., Rackham, G.M., and Steeds, J.W. (1977) *Proc. Roy. Soc.* **A354**, 197.

Kossel, W. and Möllenstedt, G. (1939) *Ann. der Phys.* **36**, 113.

LePoole, J.B. (1947) *Philips Tech. Rundsch.* **9**, 33.

Raghavan, M., Scanlon, J.C., and Steeds, J.W. (1984) *Met. Trans.* **15A**, 1299.

Reimer, L. (1993) *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis*, 3rd edition, Springer-Verlag, New York.

Steeds, J.W. (1981) *Quantitative Microanalysis with High Spatial Resolution* (Eds. G.W. Lorimer, M.H. Jacobs, and P. Doig), p. 210, The Metals Society, London.

Turner, J.H. and Krishnan, K.M. (1987) *J. Electron Microsc. Tech.* **5**, 211.

Using Convergent-Beam Techniques

21

CHAPTER PREVIEW

In the preceding chapter we described how to obtain a variety of CBED patterns under various experimental conditions, but always with a focused beam. In this chapter you will find out why these patterns are so useful: they contain a wealth of quantitative data. First, we'll show how to measure the specimen thickness. Next, we'll describe the steps for a complete crystallographic analysis of your specimen including determination of its unit cell, crystal system, point group, and space group. 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. Other convergent-beam techniques are also available, some of which use a somewhat defocused beam, as well as different microdiffraction methods which we will briefly summarize at the end of the chapter.

We should warn you at this stage that this analysis requires a very good understanding of crystallography. Both the learning and the doing are time-consuming processes. We suggest that first you skim the chapter and review your crystallography. The thickness determination described in Section 21.1 is required reading for anyone doing XEDS in the TEM.

Using Convergent-Beam Techniques

21

21.1. 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 and for determining the attainable X-ray spatial resolution (see Chapters 35 and 36). 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 in Figure 20.2B in the previous chapter, 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 less than one extinction distance thick, then you see no fringes and the 000 disk is uniformly bright, as shown back in Figure 20.6A. Clearly, these fringes contain thickness information. In fact, because the foil thickness can be measured at precisely the point you are doing diffraction and microanalysis, 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 relatively flat and undistorted, 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 $hk\ell$ reflection. If you do this you will see that the CBED disks contain parallel rather than concentric intensity oscillations, as shown in Figure 21.1. If you go to thicker regions of your specimen you'll get many more fringes, and in this case it often helps to energy-filter the pattern (look ahead to Figure 40.15).

> These oscillations are symmetric in the $hk\ell$ disk and asymmetric in the 000 disk.

We'll see in Chapter 23 that these fringes are equivalent to the rocking-curve intensity oscillations which occur across a bend contour in a BF image. 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 21.2A). 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 $hk\ell$ planes (see Figure 21.2B). The procedure to extract the thickness from the fringe pattern was first described by Kelly *et al.* (1975) and developed in detail by Allen (1981).

If you look at the $hk\ell$ disk through a 10× lupe containing a graticule, then it is easy to measure the distance between 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 21.3A, and from these spacings you can obtain a deviation s_i for the ith fringe from the equation

$$s_i = \lambda \frac{\Delta\theta_i}{2\theta_B d^2} \qquad [21.1]$$

where θ_B is the Bragg angle for the diffracting $hk\ell$ plane, d is the $hk\ell$ interplanar spacing, and we'll use the magnitude of $\mathbf{s}$, ignoring its sign. The angle $2\theta_B$ in the CBED pattern is, of course, just the separation of the 000 and $hk\ell$ disks.

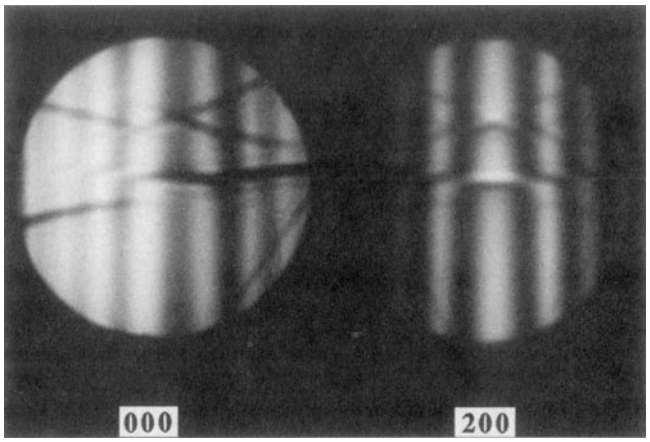


Figure 21.1. Kossel–Möllenstedt fringes in a ZOLZ CBED pattern from pure Al taken under two-beam conditions with (200) strongly excited.

The specimen in Figure 21.1 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

$$\frac{s_i^2}{n_k^2} + \frac{1}{\xi_g^2 n_k^2} = \frac{1}{t^2} \qquad [21.2]$$

where n_k is an integer (k is an integer not related to λ). If you don't know ξ_g, you 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.
- Then 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 21.3B.

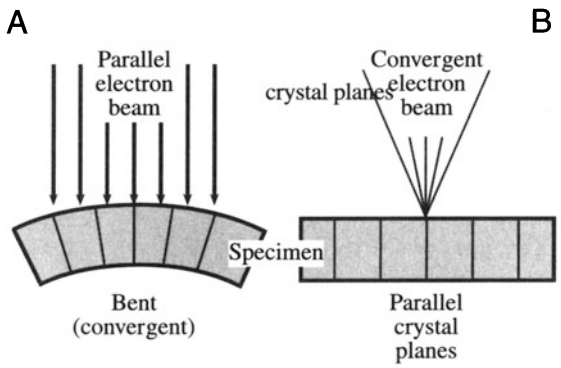


Figure 21.2. The reciprocal relationship between electron ray paths and crystal planes during (A) the formation of bend contours in BF images and (B) the formation of K–M fringes in CBED disks .

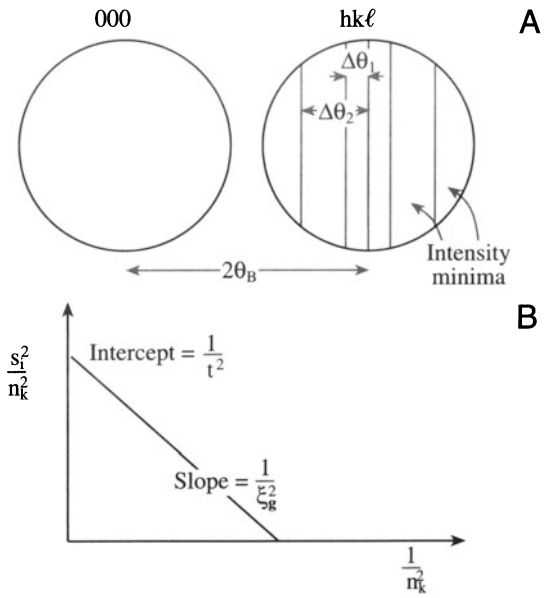


Figure 21.3. (A) The measurements necessary to extract thickness (t) from K–M fringes. From n_i measure spacings of $\Delta\theta_i$, determine the deviation parameters s_i, then (B) plot $(s_i/n_k)^2$ against n_k^2. If the plot is a straight line, extrapolate to the ordinate to find t^{-2}) and hence t.

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.

Example

■ If we apply this method to Figure 21.1, 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 $(s_i/n_k)^2$ in column 3.

These data do not plot as a straight line, since 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

Table 21.1. CBED Data for Thickness Determination

s_i (nm^{-1})	n_i	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_k	s_t^2/n_k^2 (nm^{-2})
$s_1 = 0.84 \times 10^{-2}$	$n_1 = 2$	1.7×10^{-5}
$s_2 = 2.1 \times 10^{-2}$	$n_2 = 3$	4.9×10^{-5}
$s_3 = 3.0 \times 10^{-2}$	$n_3 = 4$	5.6×10^{-5}

second set of values as shown in Table 21.2, and these numbers plot as a straight line, as shown in Figure 21.3B. The intercept of the line with the ordinate is $1/t^2$, and this equals 6.1×10^{-5} nm^{-2}. Therefore, we find that $t^2 = (6.1)^{-1} \times 10^5$ nm^2, and so $t = 128$ nm. This procedure lends itself to computerization. It is possible to digitize the fringes on line by scanning the pattern across the STEM detector or using a CCD camera. Software is available to do this analysis; you may even want to try writing the program yourself or incorporate equation 21.2 into a spreadsheet/graphing program.

21.2. 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 (Ayer, 1989). In fact, such a determination is only possible if you already know the crystal system of the specimen. 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, first of all, and then you can deduce the crystal structure. In this case then you should proceed first to Section 21.3.

We saw in the previous chapter that a CBED pattern at small L often reveals one or more rings of HOLZ intensity and you learned how to index the diffraction disks that make up these rings. If you don't know the structure, then of course it will be rather difficult to index the pattern, since you don't know the appropriate systematic absences. These rings are most useful in themselves, 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].

Hence you should be able to determine all the lattice parameters of the unit cell in a single 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 standard 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 you'll now learn how to utilize a unique aspect of CBED patterns, namely, that from a single two-dimensional pattern you can obtain 3D information about the crystal.

21.2.A. Experimental Considerations

The first thing you have to do is get DPs containing clear ZOLZ and HOLZ maxima. The patterns should have a small L to reveal one or more rings.

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).

Such a pair of patterns is shown in Figures 21.4A 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.

21.2.B. The Importance of the HOLZ-Ring Radius

If you go back and look at Figures 20.7A–C you will see the simple geometrical relationship between H and G:

- H 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 photograph. If the HOLZ ring is split, always measure G_n using the innermost ring.
- If the order of the ring is too large ($\alpha \sim 10°$), then your measurements may suffer from the effects of lens distortion because the scattering semiangle is so large. You must calibrate the distortion in reciprocal space using a known

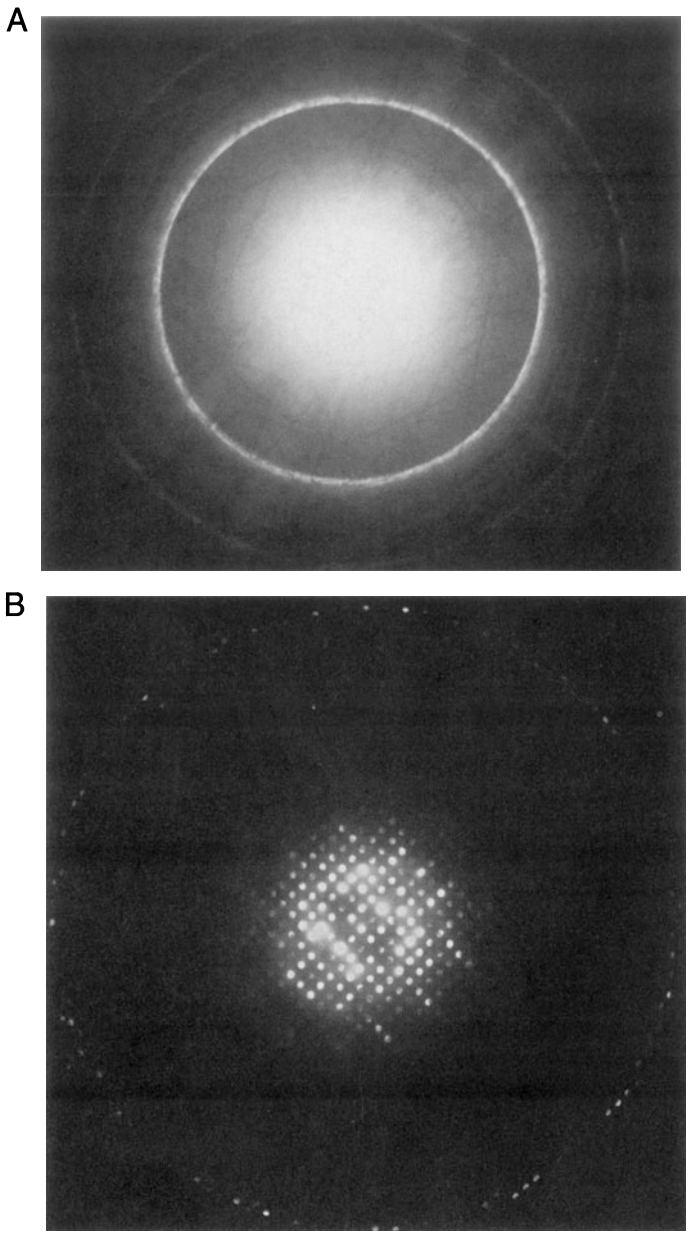


Figure 21.4. (A) CBED Kossel pattern from a carbide particle taken with a 150-μm C2 aperture showing the FOLZ ring of intensity surrounding the overexposed ZOLZ region. In (B) the same pattern, taken with a 20-μm aperture, reveals individual reflections in both the ZOLZ and FOLZ.

specimen from which G_1, G_2, etc. can be calculated and compared with the values obtained experimentally.

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 21.4A. 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 Figure 20.7A and assuming that terms in H^2 are negligible, the radii of the FOLZ and SOLZ rings, G_1, and G_2, are (Steeds, 1979)

$$G_1 = \left(\frac{2H}{\lambda}\right)^{1/2} \qquad [21.3]$$

and

$$G_2 = 2\left(\frac{H}{\lambda}\right)^{1/2} \qquad [21.4]$$

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 space to give, for the FOLZ

$$\frac{1}{H} = \frac{2}{\lambda G_1^{\,2}} \qquad [21.5]$$

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 20.8, you can see that a low-symmetry zone axis with a small H will give rise to a small HOLZ ring of diameter r on the DP, which 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

$$\frac{1}{H} = \left|[UVW]\right| \qquad [21.7]$$

and so this magnitude can be calculated for a specific beam direction $[UVW]$.

Example

For an fcc crystal (Steeds 1979)

$$\frac{1}{H} = \frac{a_0\left(U^2 + V^2 + W^2\right)^{1/2}}{p} \qquad [21.8]$$

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 nonintegral, then your indexing is wrong. A generalized method for determining which Laue zone you should see has been given by Jackson (1990). You can see that, if you have indexed the ZOLZ (i.e., $[UVW]$ is known), r is measured, and λ is known, then H can be determined without the need to index individual spots in the HOLZ ring of intensity.

It is possible to develop more generalized equations for H^{-1} than equation 21.8 (Raghavan *et al.*, 1984, Ayer, 1989).

Other examples

In a crystal system with orthogonal axes (i.e., 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

$$\frac{1}{H} = \left(a^2\left(U^2 + V^2 - UV\right) + c^2W^2\right)^{1/2} \qquad [21.11]$$

and for the four-index system

$$\frac{1}{H} = \left(3\left(U^2 + V^2 + UV\right)^2 + c^2W\right)^{1/2} \qquad [21.12]$$

while for the monoclinic system with a unique b-axis

$$\frac{1}{H} = \left(U^2a^2 + V^2b^2 + W^2c^2 + 2UWac\cos\beta^2\right)^{1/2} \qquad [21.13]$$

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 less symmetrical 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, and so the FOLZ-ring diameter should give the same value of H^{-1} as that determined from 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.

21.2.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 21.4B. 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 the primitive lattice of Figure 21.5, the FOLZ superimposes 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 shown by Hirsch *et al.* (1977) and illustrated schematically in Figure 21.5A. 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 Figures 20.9 and 20.10. It is not so simple in more complex crystals, but Jackson (1987) has developed a generalized method of determining the shift vector $\mathbf{t}$ for all crystal systems and all orientations

$$\mathbf{t} = \mathbf{g} - \mathbf{u}^*\left(\frac{HN_L}{|\mathbf{u}^*|}\right) \qquad [21.14]$$

where $\mathbf{g}$ is the vector for the $hk\ell$ HOLZ reflection, $\mathbf{u}^*$ is the vector normal to the ZOLZ and parallel to H, and N_L is the number of the Laue zone containing $hk\ell$. To determine $\mathbf{t}$, then all you do is look up values of H, $\mathbf{u}^*$, and $H/|\mathbf{u}^*|$, tabulated by Jackson (1987).

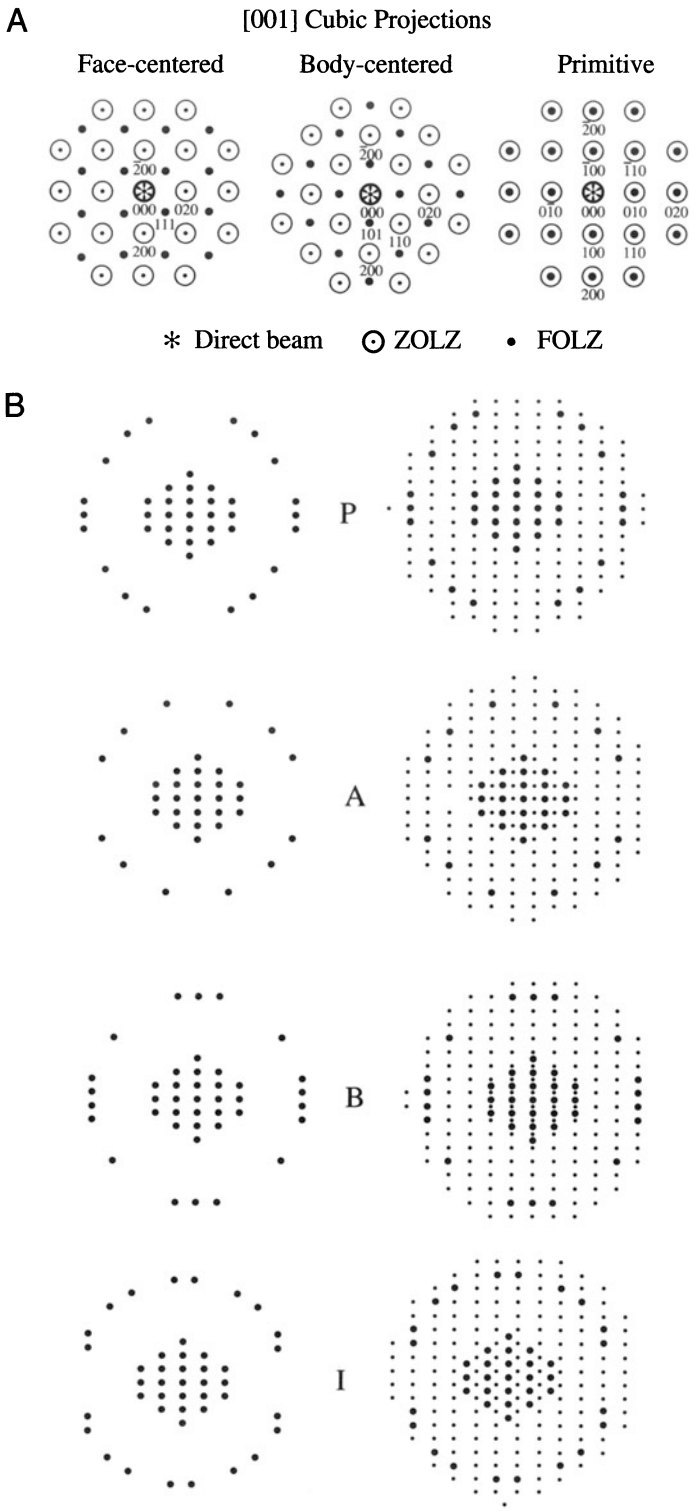


Figure 21.5. (A) The overlap between the ZOLZ and the FOLZ when looking down the [001] axis of cubic crystals. In the fcc pattern, 111 is a FOLZ index, likewise 101 in bcc. In the primitive pattern, only the ZOLZ is indexed. (B) Schematic illustration of the superposition of the FOLZ pattern on the ZOLZ 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 (I) I-centered lattices.

Example

We can illustrate the shift due to lattice type, by looking at Figure 21.5B, which is a series of schematic patterns for an orthorhombic cell (Ayer 1989) 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 chemical analysis by XEDS or EELS.

21.3. SYMMETRY DETERMINATION

21.3.A. Introduction to Symmetry Concepts

Before you study the following two sections 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. 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 21.6 we reproduce the standard point-group table familiar to any student of crystallography; we will refer to this table again. If you aren't familiar with such concepts, then the rest of this section may be incomprehensible and you should go and read one of the crystallography texts which we listed at the end of Chapter 18.

The crystallographic point groups are the sets of crystallographically permissible symmetries which are formed when sets of axes intersect in a common point. The axes correspond to the rotation, inversion, and mirror symmetry elements. We generally choose our lattice points to have a particularly high degree of symmetry. We will ignore translational symmetry elements until we discuss space groups.

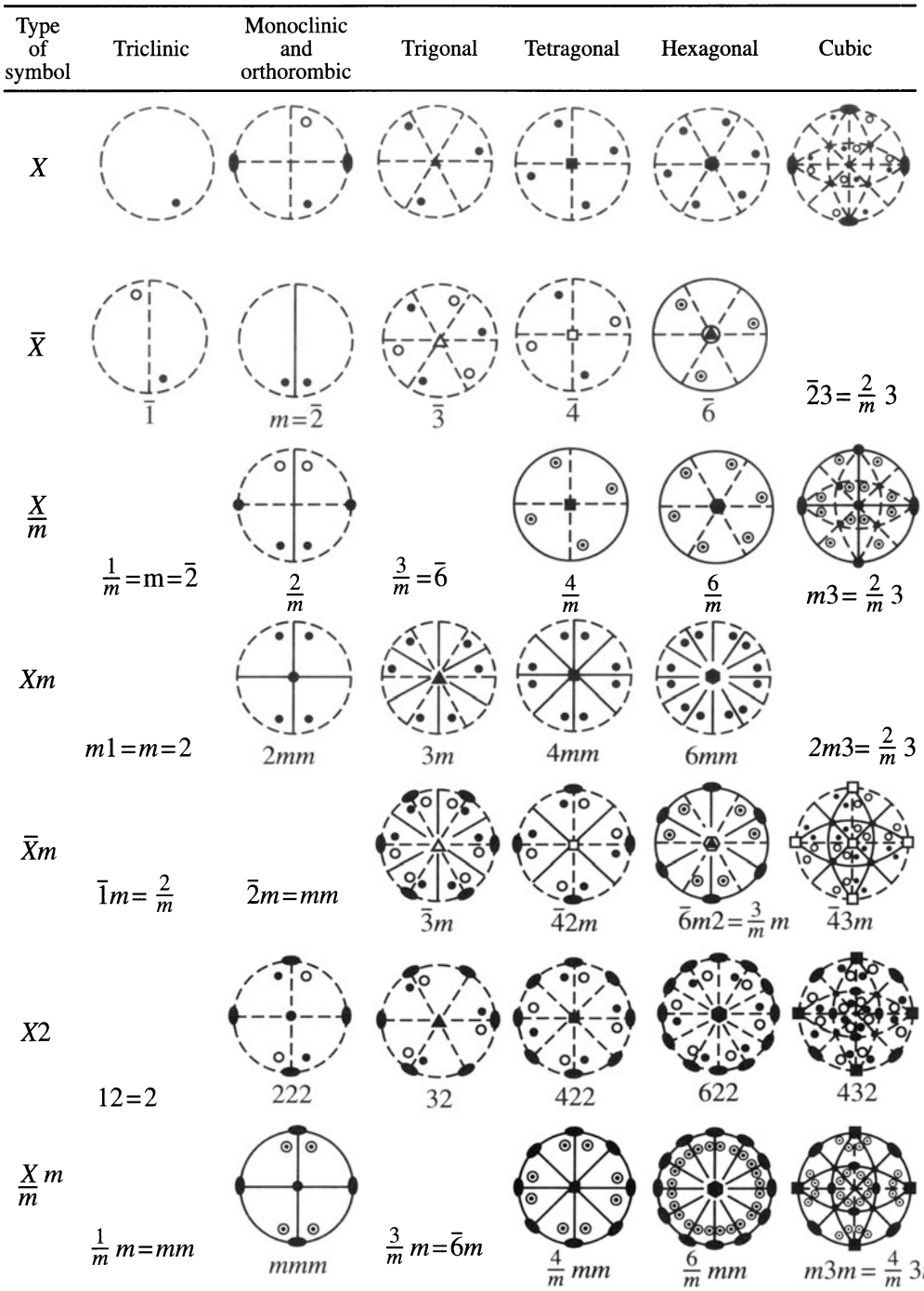


Figure 21.6. The 32 crystal point groups represented by stereograms showing the operation of rotational, mirror, and inversion symmetry elements on a general pole $hk\ell$. The international notation describing the point groups is given under each of the stereograms.

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. With the availability of CBED you can now determine the point group of a thin crystal directly in the TEM, simply by recording two or three low-index ZAPs.

This process has a tremendous advantage over classical 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 nontrivial 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. 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.

21.3.B. Friedel's Law

Symmetry determination in crystals has evolved from the early work of Friedel and von Laue on the kinematical theory of X-ray diffraction. We can summarize a fundamental aspect of Friedel's work as follows.

Friedel's Law: The intensity of some reflection $hk\ell$ in an X-ray DP is equal to the intensity in its opposite reflection $\bar{h}\bar{k}\bar{\ell}$.

If this is so, a center of symmetry exists in the DP. This is the case for most X-ray DPs from single crystals, because most X-ray diffraction occurs under kinematical conditions.

The presence of a mirror plane in a crystal, parallel to axes *a* and *b*, makes the intensity of all reflections of type $hk\ell$ equal to the intensity of the corresponding $\bar{h}\bar{k}\bar{\ell}$ reflection.

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 (see Figure 21.6) *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 $I_{hk\ell} = I_{k\bar{h}\ell} = I_{\bar{h}\bar{k}\ell} = I_{\bar{k}h\ell}$, where $I_{hk\ell}$ denotes the diffracted intensity of a reflection of type $hk\ell$. Under kinematical conditions, these intensities are also the same as $I_{hk\bar{\ell}} = I_{\bar{k}h\bar{\ell}} = I_{\bar{h}\bar{k}\bar{\ell}} = I_{k\bar{h}\bar{\ell}}$.

X-ray diffraction is thus severely limited for point-group determination because of Friedel's law. Since crystals which do not possess true centers of symmetry (noncentrosymmetric 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 21.6 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*3, and *m*3*m*. These are known as the *Laue groups* in X-ray diffraction. Except under "anomalous" scattering conditions, X-ray diffraction can only determine these 11 symmetry groups.

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 $hk\ell$ reflections, you can distinguish centrosymmetric and noncentrosymmetric crystals. That is, the 32 crystal point groups are not reduced to the 11 Laue groups, as occurs in X-ray diffraction.

21.3.C. Looking for Symmetry in Your Patterns

Before we go into details of point-group determination, let's first get some practice at looking at CBED patterns and seeing the symmetry within them. 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 $\bar{X}$ or *X*/*m* are not used. The only combinations we can get are the same as for the ten two-dimensional point groups: 1, 2, *m*, 2*mm*, 3, 3*m*, 4, 4*mm*, 6, or 6*mm*. These symbols refer to the observable symmetry in the pattern, and four examples of different pattern symmetries are shown schematically in Figure 21.7.

Symmetry determination is always carried out using ZAPs:

- 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.

There are two specific kinds of symmetry that you have to look for in CBED patterns:

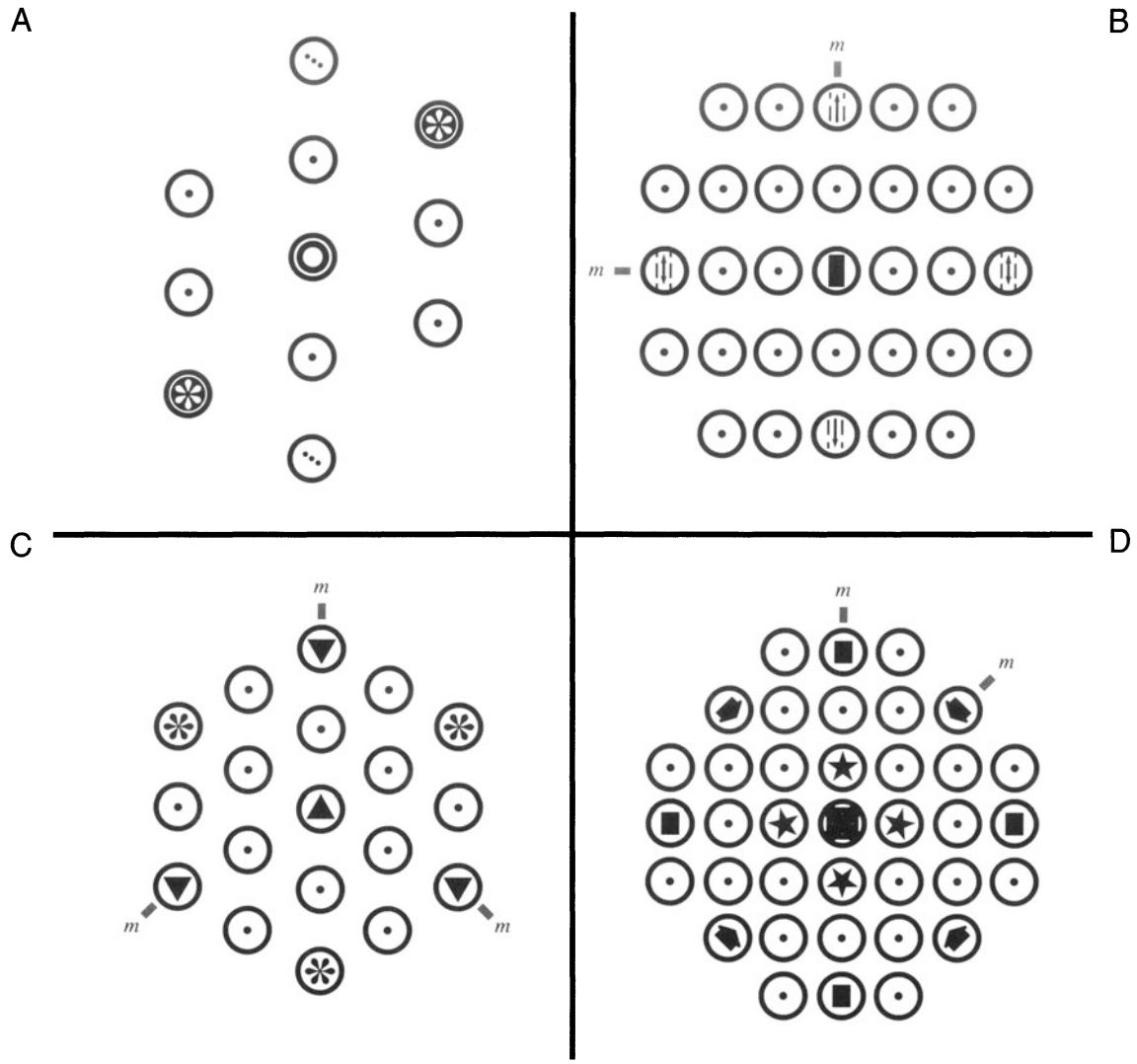


Figure 21.7. Four examples of symmetry in CBED patterns. (A) Symmetry 2 refers to a twofold (diad) rotation axis, i.e., the pattern has symmetry when rotated 180°; (B) 2*mm* is a diad symmetry with two independent mirror planes parallel to the diad; (C) 3*m* indicates threefold rotation (triad) symmetry with one mirror plane, i.e., rotational symmetry every 120° with one mirror plane present at each 120°; (D) 4*mm* indicates a fourfold rotational symmetry (tetrad) with two independent mirror planes parallel to the tetrad.

- Whole-pattern (WP) symmetry.
- Bright-field (BF) symmetry.

The first and most important is the *WP symmetry*. The WP symmetry is just what it says; the symmetry of the complete pattern, including the relative positions of the HOLZ reflections and any HOLZ Kikuchi lines. 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. The pattern can be either a Kossel or a K–M pattern. The WP symmetry at any orientation must belong to one of the ten two-dimensional groups listed above. If you look at the CBED pattern from Cu back in Figure 20.16, the WP symmetry is 3*m* because the HOLZ ring and the array of HOLZ-deficient Kikuchi lines shows threefold symmetry with one mirror plane reproduced every 120°. As a useful exercise look at a few other CBED patterns in this chapter and work out the WP symmetry.

The second kind of symmetry is the *BF symmetry,* which refers to the symmetry of the 000 disk only, *when HOLZ lines are present*. In this case, the BF symmetry also 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 the diffracted maxima. For example, the 000 disk back in Figure 20.2B has an array of deficient HOLZ lines which displays 3*m* 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 ten two-dimensional point groups just listed. We will see that combination of the WP and BF symmetry in three ZAPs is usually sufficient to determine the point group.

On some occasions the projection-diffraction symmetry may be all that is available. This term refers to the symmetry displayed by the intensity in the direct 000 beam plus the $hk\ell$ diffracted beams in the zero layer. It *ignores any contributions from HOLZ layers* such as HOLZ lines and HOLZ reflections, but includes any diffuse intensity within the ZOLZ disks. This is because the diffuse contrast within these disks arises from dynamical interactions within the zero layer of the crystal that give rise to K–M fringes, which we used for thickness determination. The projection-diffraction symmetry corresponds to the projected two-dimensional symmetry of the crystal down the zone axis that you have selected. The projection-diffraction symmetry is simply the symmetry displayed in SAD patterns. Since this symmetry is only two-dimensional, it is not as useful as the WP and BF symmetry. If you go back and examine Figure 20.16, the projection-diffraction symmetry is not clear because the ZOLZ disks overlap, but the ZOLZ Kikuchi bands can also be used and they show a sixfold rotational symmetry with two independent mirror planes, one within the Kikuchi bands and one between them, giving $6mm$ symmetry. Similar symmetry is shown in the SAD pattern from Si in Figure 20.2A.

21.3.D. Point-Group Determination

There are several methods to determine the point group by examining the various symmetry aspects as we have just described, but the easiest method is based on the work of Steeds (1979). This method uses several different ZAPs. For each of these you have to determine the BF and WP symmetries, and ensure that they are consistent with the projection symmetry. To do this we make use of a standard table, which is given here as Table 21.3. This table is a modified version of the original table developed by Buxton *et al.* (1976) in a seminal but rather complex paper.

> The new concept introduced in these tables is the idea of the "*diffraction group*" of a crystal. This term describes the full 3D symmetry of a DP and the "*projection-diffraction group*" describes the full two-dimensional symmetry.

There are 10 two-dimensional projection-diffraction groups which are related to the 10 two-dimensional point group symbols with the addition of the symbol 1_R. 1_R denotes a rotational symmetry present in all diffraction disks around the central point of the $hk\ell$ disk where the Bragg condition is satisfied. The 31 diffraction groups are obtained by combinations of the 10 two-dimensional point group symbols with the rotational symmetry, 1_R, and mirror rotations, m_R. All these groups are listed in column 1 of Table 21.3. You don't have to understand the *derivation* of the nomenclature in order to be able to carry out point group determination, but if you want to pursue the concepts in more depth, the reviews by Steeds (1984) and Tanaka (1989) may help.

The Procedure: From your ZAP, determine the BF and WP symmetry. Go to Table 21.3 and see which of the 31 possible diffraction groups accounts for the observed symmetry. There may be only one, but more often there are two possibilities. For example, if your BF and WP symmetries are $3m$ as we found for the Cu pattern, then possible diffraction groups are $3m$ or $6_R mm_R$. However, if your BF symmetry is $3m$, but the WP symmetry is only 3, then the only possible diffraction group is $3m_R$. If your patterns only contain two-dimensional information, then you may only be able to determine the projection-diffraction group, not the full diffraction group. Under these circumstances, you must go to experimental conditions that give 3D information, i.e., a thicker specimen, a shorter L to see any HOLZ rings, or cool the specimen to enhance HOLZ effects and so on, as we described in Chapter 20.

Repeat the Procedure: Conduct the same type of analysis for up to three orientations in your crystal. Remember that if you can't see the HOLZ contrast in the BF disk, you should choose a lower-symmetry orientation. When you have found the possible diffraction groups for each orientation, the crystal point group can be deduced using Table 21.4, again taken from the pioneering work of Buxton *et al.* (1976). Table 21.4 contains the 31 diffraction groups and the 32 crystal point groups. The diffraction groups, consistent with particular crystal point groups, can be deduced. For each crystal orientation and diffraction group, one or more different crystal point groups are possible. For example, the diffraction group $3m$, which is one of the two possible ones for the Cu pattern, is consistent with point groups $3m$ and $\bar{4}3m$. If you repeat this exercise for three different ZAPs, you will find that only one crystal point group is consistent with all the diffraction groups observed; this is the true point group of your specimen.

Example

Point-group determination in this manner is shown in Figure 21.8, which contains three sets of CBED patterns from three separate zone axes in austenitic stainless steel. The BF and WP symmetries are always identical in this case, although this is not in-

Table 21.3. CBED Pattern Symmetries

Diffraction Group	Bright Field	Whole Pattern	Dark Field		± G		Projection-Diffraction Group
			General	Special	General	Special[a]	
1	1	1	1	none	1	none	1_R
1_R	2	1	2	none	1	none	1_R
2	2	2	1	none	2	none	21_R
2_R	1	1	1	none	2_R	none	21_R
21_R	2	2	2	none	21_R	none	21_R
m_R	m	1	1	m	1	m_R	$m1_R$
m	m	m	1	m	1	m	$m1_R$
$m1_R$	$2mm$	m	2	$2mm$	1	$m1_R$	$m1_R$
$2m_Rm_R$	$2mm$	2	1	m	2	—	$2mm1_R$
$2mm$	$2mm$	$2mm$	1	m	2	—	$2mm1_R$
2_Rmm_R	m	m	1	m	2_R	—	$2mm1_R$
$2mm1_R$	$2mm$	$2mm$	2	$2mm$	21_R	—	$2mm1_R$
4	4	4	1	none	2	none	41_R
4_R	4	2	1	none	2	none	41_R
41_R	4	4	2	none	21_R	none	41_R
$4m_Rm_R$	$4mm$	4	1	m	2	—	$4mm1_R$
$4mm$	$4mm$	$4mm$	1	m	2	—	$4mm1_R$
4_Rmm_R	$4mm$	$2mm$	1	m	2	—	$4mm1_R$
$4mm1_R$	$4mm$	$4mm$	2	$2mm$	21_R	—	$4mm1_R$
3	3	3	1	none	1	none	31_R
31_R	6	3	2	none	1	none	31_R
$3m_R$	$3m$	3	1	m	1	m_R	$3m1_R$
$3m$	$3m$	$3m$	1	m	1	m	$3m1_R$
$3m1_R$	$6mm$	$3m$	2	$2mm$	1	$m1_R$	$3m1_R$
6	6	6	1	none	2	none	61_R
6_R	3	3	1	none	2_R	none	61_R
61_R	6	6	2	none	21_R	none	61_R
$6m_Rm_R$	$6mm$	6	1	m	2	—	$6mm1_R$
$6mm$	$6mm$	$6mm$	1	m	2	—	$6mm1_R$
6_Rmm_R	$3m$	$3m$	1	m	2_R	—	$6mm1_R$
$6mm1_R$	$6mm$	$6mm$	2	$2mm$	21_R	—	$6mm1_R$

[a]Where a dash appears, the special symmetries can be deduced from columns 5 and 6 of this table (or from Table 1 in Buxton *et al.* 1976).

variably so, as you can see from Table 21.3. The symmetries are $3m$ from the [111], $4mm$ from the [100], and $2mm$ from the [110] patterns, respectively. The diffraction group symmetry consistent with each of the three patterns can then be determined from Table 21.3. Table 21.5 lists the possible point groups consistent with the diffraction group symmetry, taken from Table 21.4, and it is immediately apparent that only one point group, $m3m$, is consistent with the symmetry in all three patterns. This conclusion can be checked by reference to Table 21.6, also from Buxton *et al.*, in which it is seen that the symmetries consistent with the $m3m$ point group are correct, i.e., 6_Rmm_R for [111], $4mm1_R$ for [100], and $2mm1_R$ for [110].

So let's summarize this detailed, but nonetheless straightforward process to determine point-group symmetry:

- Obtain at least three low-index ZAPs, and record a small camera-length pattern and large camera-length pattern at each orientation.
- From the small-L pattern determine the WP symmetry.
- From the large-L pattern determine the BF symmetry.
- From Table 21.3 determine the possible diffraction groups consistent with the WP and BF symmetries.
- From Table 21.4 determine the possible point groups and find the one consistent with your pattern symmetries in all orientations.
- If necessary, cross-check with Table 21.6 to see that the diffraction group symmetries at each orientation were consistent with the point group you deduced.

Table 21.4. Relation Between the Diffraction Groups and the Crystal Point Groups

Diffraction Group	1	$\bar{1}$	2	m	$2/m$	222	$mm2$	mmm	4	$\bar{4}$	$4/m$	422	$4mm$	$\bar{4}2m$	$4/mmm$	3	$\bar{3}$	32	$3m$	$\bar{3}m$	6	$\bar{6}$	$6/m$	622	$6mm$	$6m2$	$6/mmm$	23	$m3$	432	$\bar{4}3m$	$m3m$
$6mm1_R$																											X					
$3m1_R$																										X						
$6mm$																									X							
$6m_Rm_R$																								X								
61_R																							X									
31_R																						X										
6																					X											
6_Rmm_R																				X												X
$3m$																			X												X	
$3m_R$																		X												X		
6_R																	X												X			
3																X											X					
$4mm1_R$															X																	X
4_Rmm_R														X																	X	
$4mm$													X																			
$4m_Rm_R$												X																		X		
41_R											X																					
4_R										X																						
4									X																							
$2mm1_R$								X							X												X		X			X
2_Rmm_R					X			X			X				X					X			X				X		X			X
$2mm$							X																			X						
$2m_Rm_R$						X						X		X										X				X		X		
$m1_R$							X						X	X											X	X					X	
m				X			X						X	X					X			X			X	X					X	
m_R			X			X	X		X	X		X	X	X				X			X			X	X	X		X		X	X	
21_R					X															X												
2_R		X			X			X			X				X		X			X			X				X		X			X
2			X															X														
1_R				X															X													
1	X		X	X		X	X		X	X		X	X	X		X		X	X		X	X		X	X	X		X		X	X	

There are other methods to determine the point group symmetry, especially that discussed by Tanaka (1989). If you want to be a dedicated electron crystallographer you should become familiar with all the methods, but we haven't the space here to go into them all. We will now go on briefly to discuss the way in which we can get the space group of the crystal if the point-group identification is not sufficient to identify the unknown sample unambiguously.

21.3.E. Space-Group Determination

For a full crystallographic analysis of your specimen, you should determine the space group.

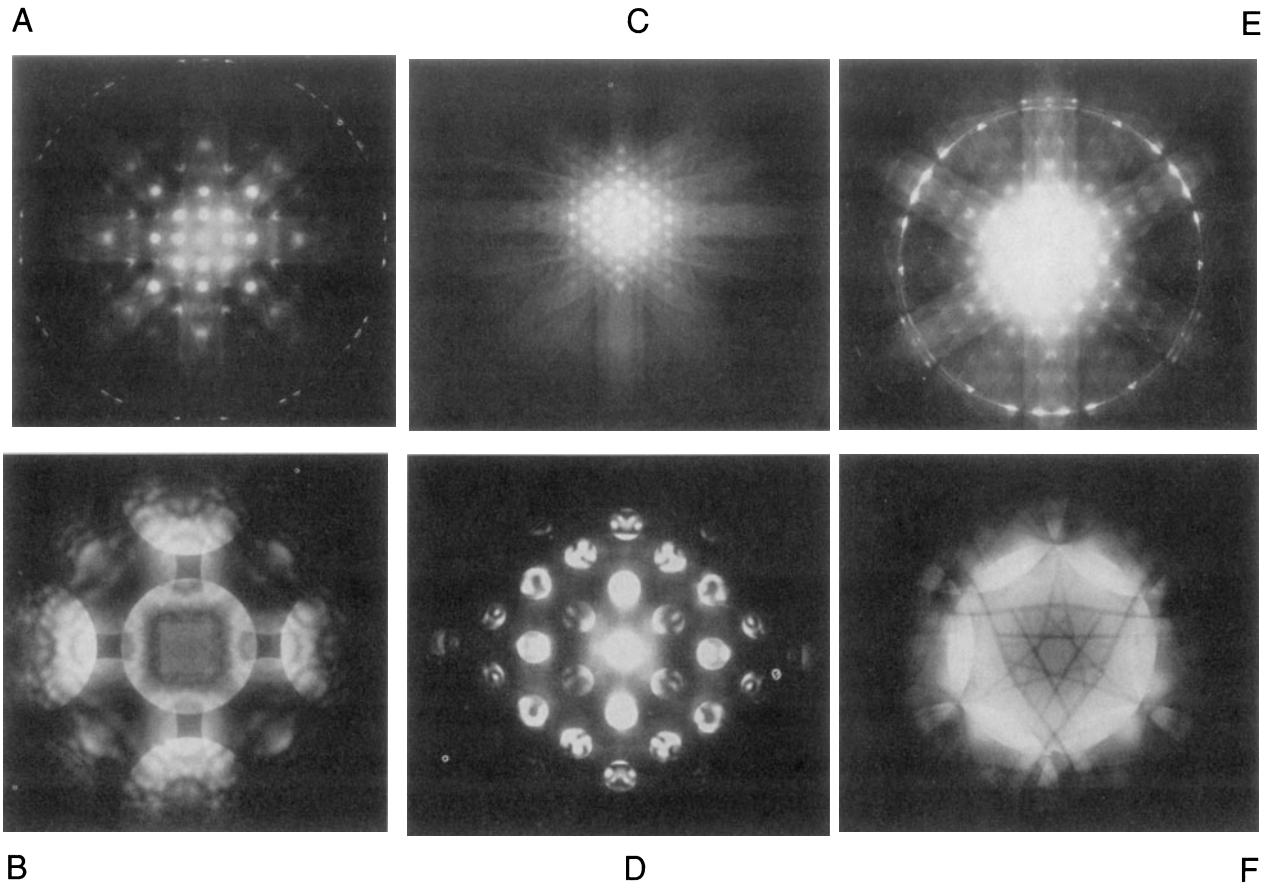


Figure 21.8. (A, B) [100], (C, D) [110], and (E, F) [111] ZAPs from stainless steel used to determine the point group. In (A, C, E) the low-*L* pattern gives the WP symmetry and in (B, D, F) the high-*L* pattern shows the BF symmetry. From the WP and BF symmetries, possible diffraction groups and point groups are determined, as summarized in Table 21.5.

The space group is a classification of the complete symmetry of a crystal that takes into account translational symmetry elements such as glide planes and screw axes, in addition to the point symmetry elements.

The space-group notation combines the Bravais lattice type, such as primitive (P) or body-centered (I), with the point-group notation (e.g., *mmm* or 4/*mmm*) along with appropriate symbols for glide planes and screw axes, if necessary. The space group is the ultimate classification step and there are 230 possible space groups. So it is possible to be much more precise about grouping your crystal. However, knowing the space group of your crystal doesn't tell you anything more about its properties than you may have learned from knowing its point group. So there is nothing to be gained by pursuing space-group analysis if,

Table 21.5. Possible Diffraction Groups and Point Groups

Zone Axis	BF Symmetry	WP Symmetry	Possible Diffraction Groups	Possible Point Groups									
<111>	3*m*	3*m*	3*m*	3*m*	$\bar{4}3m$								
			$6_R mm_R$			$\bar{3}m$	*m*3*m*						
<100>	4*mm*	4*mm*	4*mm*					4*mm*					
			$4mm1_R$				*m*3*m*		4/*mmm*				
<110>	2*mm*	2*mm*	2*mm*							*mm*2	$\bar{6}m2$		
			$2mm1_R$				*m*3*m*		4/*mmm*			*mmm*	432	6/*mmm*

Table 21.6. Zone-axis Symmetries

point group	<111>	<100>	<110>	<*UV*0>	<*UUW*>	[*UVW*]	
$m3m$	$6_R mm_R$	$4mm1_R$	$2mm1_R$	$2_R mm_R$	$2_R mm_R$	2_R	
$\bar{4}3m$	$3m$	$4_R mm_R$	$m1_R$	m_R	m	1	
432	$3m_R$	$4m_R m_R$	$2m_R m_R$	m_R	m_R	1	
point group	<111>	<100>	<*UV*0>	[*UVW*]			
$m3$	6_R	$2mm1_R$	$2_R mm_R$	2_R			
23	3	$2m_R m_R$	m_R	1			
point group	[0001]	<11$\bar{2}$0>	<1$\bar{1}$00>	[*UV*.0]	[*UU*.*W*]	[*UŪ*.*W*]	[*UV*.*W*]
$6/mmm$	$6mm1_R$	$2mm1_R$	$2mm1_R$	$2_R mm_R$	$2_R mm_R$	$2_R mm_R$	2_R
$\bar{6}m2$	$3m1_R$	$m1_R$	$2mm$	m	m_R	m	1
$6mm$	$6mm$	$m1_R$	$m1_R$	m_R	m	m	1
622	$6m_R m_R$	$2m_R m_R$	$2m_R m_R$	m_R	m_R	m_R	1
point group	[0001]	[*UV*.0]	[*UV*.*W*]				
$6/m$	61_R	$2_R mm_R$	2_R				
$\bar{6}$	31_R	m	1				
6	6	m_R	1				
point group	[0001]	<11$\bar{2}$0>	[*UŪ*.*W*]	[*UV*.*W*]			
$\bar{3}m$	$6_R mm_R$	21_R	$2_R mm_R$	2_R			
$3m$	$3m$	1_R	m	1			
32	$3m_R$	2	m_R	1			
point group	[0001]	[*UV*.*W*]					
$\bar{3}$	6_R	2_R					
3	3	1					
point group	[001]	<100>	<110>	[*U*0*W*]	[*UV*0]	[*UUW*]	[*UVW*]
$4/mmm$	$4mm1_R$	$2mm1_R$	$2mm1_R$	$2_R mm_R$	$2_R mm_R$	$2_R mm_R$	2_R
$\bar{4}2m$	$4_R mm_R$	$2m_R m_R$	$m1_R$	m_R	m_R	m	1
$4mm$	$4mm$	$m1_R$	$m1_R$	m	m_R	m	1
422	$4m_R m_R$	$2m_R m_R$	$2m_R m_R$	m_R	m_R	m_R	1
point group	[001]	[*UV*0]	[*UVW*]				
$4/m$	41_R	$2_R mm_R$	2_R				
$\bar{4}$	4_R	m_R	1				
4	4	m_R	1				
point group	[001]	<100>	[*U*0*W*]	[*UV*0]	[*UVW*]		
mmm	$2mm1_R$	$2mm1_R$	$2_R mm_R$	$2_R mm_R$	2_R		
$mm2$	$2mm$	$m1_R$	m	m_R	1		
222	$2m_R m_R$	$2m_R m_R$	m_R	m_R	1		
point group	[010]	[*U*0*W*]	[*UVW*]				
$2/m$	21_R	$2_R mm_R$	2_R				
m	1_R	m	1				
2	2	m_R	1				
point group	[*UVW*]						
$\bar{1}$	2_R						
1	1						

from the point group and any other data such as XEDS or EELS, you are able to identify the crystal unambiguously. Space-group analysis is only necessary if the point group determination still leaves some uncertainty as to the nature of your crystal.

There may be some uncertainty, because even when you have determined the crystal point group and you know the lattice type, several different space groups are still possible. The next step in CBED analysis is to analyze the reflections which are kinematically forbidden (see Section 16.9) for each possible space group, and thus identify the specific space group of the crystal. Due to the dynamical nature of electron diffraction, reflections which are forbidden in kinematical diffraction often occur in CBED patterns by double diffraction. Kinematically forbidden reflections can occur by double diffraction when they are due to additional symmetry elements such as glide planes or screw axes.

It is these translational symmetry elements which uniquely identify the space group of the crystal.

In CBED we are interested in the kinematically forbidden reflections that occur due to these additional symmetry elements. When two or more equivalent double-diffraction paths exist in a given orientation, the kinematically forbidden reflection that occurs will have a central line of zero intensity passing through the disk. These so-called dynamical absences occur in these reflections because the diffracted beams from two equivalent paths undergo complete destructive interference along the central line of the disk to which the beams are perpendicular, as shown in Figure 21.9. First, you must obtain a ZAP since the dynamical absences don't occur if the crystal is tilted off the zone. The kinematically forbidden reflections, which generally have multiple double-diffraction routes, usually lie along systematic rows of reflections, and occur in *alternate* reflections along a systematic row. You can easily distinguish the lines of absence from other contrast phenomena because they occur for all specimen thicknesses, at all values of kV, and they become narrower as the thickness increases.

The existence of a dynamical absence in a kinematically forbidden reflection indicates that the electron beam is aligned either parallel to a glide plane or perpendicular to a screw axis in the crystal. Steeds and Vincent (1983) established tables, based on the earlier work of Gjønnes and Moodie (1965), which describe the relationship between the dynamical absences and the number of symmetry elements that can be responsible for those absences. These relationships, shown in Table 21.7, are used for interpreting the presence of screw axes and glide planes in space-group determinations using 3D CBED effects. Some rare situations exist where this approach breaks down but, in general, it has withstood the test of time.

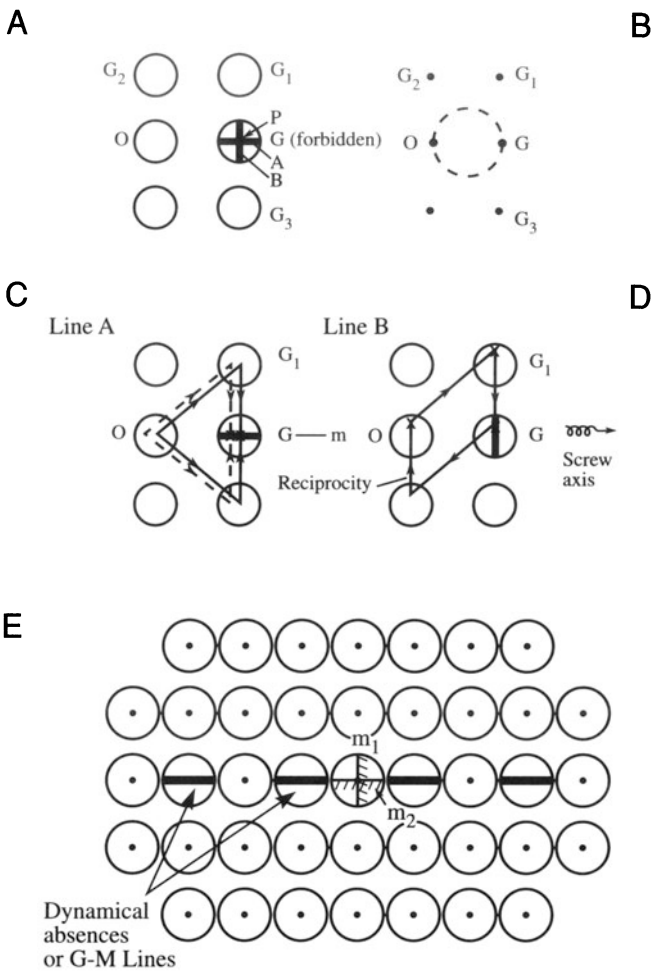


Figure 21.9. The formation of dynamic absences or G–M lines in CBED patterns. In (A) a pair of orthogonal G–M lines form a black cross in a forbidden diffraction disk G. In (B) the position of the Laue circle is shown. In (C) and (D) pairs of diffraction vectors contribute equal and opposite amplitudes, thus causing the lines of dynamic absence. In the ZOLZ of a CBED pattern the G–M lines will occur in alternate reflections, as shown in (E). To determine the translational symmetry elements, examine the orientation of the G–M lines with respect to the BF mirror planes, and consult Table 21.7.

The dynamical absences are referred to as Gjønnes–Moodie or G–M lines.

As you can see from Table 21.7, you have to carry out an analysis of the orientation of the G–M line with respect to the BF mirrors within the 000 disk, in order to ascertain whether a glide plane, screw axis, or both are present. Once the additional glide planes and screw axes in each orientation are known, and you know the point group, the space group can be identified from Volume A of The International Tables for Crystallography (Hahn 1988) in conjunc-

Table 21.7. The Six Different Cases of Dynamic Absences along a Single Systematic Line of Reflections

Symbols used follow Tables 4.1.6 and 4.1.7 of International Tables for X-ray Crystallography (1969)

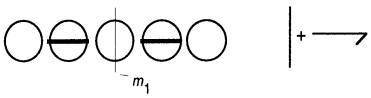
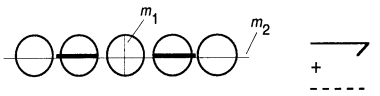
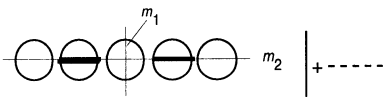
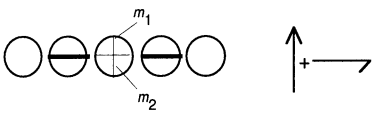
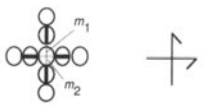
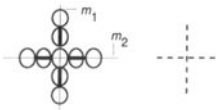
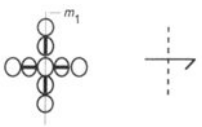
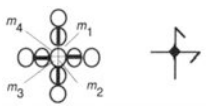
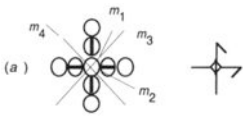
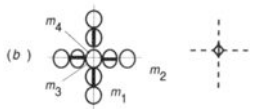
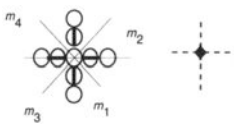
WP	BF	Diffraction Group	Orientation of mirrors with respect to lines of absences in a zone-axis pattern (orthogonal lines are principal axes)	Minimum number of symmetry elements responsible for absence
1	*m*	m_R		Screw axis perpendicular to beam $2_1 4_1 4_3 6_1 6_3$ or 6_5
	m	$2_R mm_R$ (*a*) and (*b*) or	(*a*) 	Screw axis perpendicular to the beam and to a mirror plane; $2_1/m$, $6_3/m$
m		*m* (*b*)	(*b*) 	Glide plane parallel to beam
m	2*mm*	$m\,1_R$		Screw axis (2_1, 6_3) plus parallel glide plane
2*mm*	2*mm*	2*mm* or 2*mm* 1_R		As (*b*) above with an extra mirror plane parallel to the beam and perpendicular to the glide plane
2	2*mm*	$2m_R m_R$		Screw axis ($2_1 4_1 4_3 6_1 6_3$ or 6_5) perpendicular to a 2-fold axis or 2_1 axis perpendicular to a 4-fold axis both perpendicular to the beam

Table 21.7 (continued)
The Seven Cases of Dynamic Absences along Two Orthogonal Lines

WP	BF	Diffraction Group	Orientations of mirrors with respect to lines of absences in zone-axis patterns	Minimum number of symmetry elements responsible for absences
2	2*mm*	$2m_Rm_R$		Orthogonal screw axes, orthogonal to the beam direction. 2_1 perpendicular to 2_1, 4_1, or 4_3
2*mm*	2*mm*	2*mm* or 2*mm* 1_R		Two perpendicular glide planes
m	*m*	2_Rmm_R		2_1 screw axis perpendicular to a glide plane. (*N.B.* although there are 4_1 screw axes perpendicular to glide plane in cubic *F* and *I* centered space groups dynamic absences do not occur. There are no multiple diffraction routes to the forbidden reflections.)
4	4*mm*	$4m_Rm_R$		Orthogonal screw diads normal to a tetrad axis (4, 4_1, or 4_3) which is parallel to the beam direction OR an orthogo-nal set of three 4_1 or 4_3 screw tetrad axes with one axisparallel to the beam
2*mm*	4*mm*	4_Rmm_R		Orthogonal screw diads normal both to an inversion tetrad axis and to the beam direction
				Orthogonal glide planes, parallel to an inversion tetrad axis
4*mm*	4*mm*	4*mm* or 4*mm* 1_R		Orthogonal glide planes, parallel both to a tetrad axis (4 or 4_2) and to the beam direction

tion with the rules for forbidden reflections. We can illustrate the determination of translational symmetry elements by observing Figure 21.10 in detail.

Example

$CaZrO_3$ is orthorhombic, with the *mmm* point group. Within the *mmm* point group the only translational symmetry elements allowed are twofold screw/rotation axes and {100} glide/mirror planes. Determination of which translational symmetry elements are present can be deduced by analyzing CBED patterns along <*UV*0> zone axes. These <*UV*0> zone-axis patterns, such as [210], within the *mmm* point group contain only a single BF/WP mirror. Figure 21.10A shows two orthogonal pairs of G–M lines in alternate reflections along the (001), $1 \neq 2n$, and (*hk*0), $h+k \neq 2n$, systematic rows. Figure 21.10D shows the BF disk which displays single mirror symmetry, with (*hk*0), $h+k \neq 2n$ G–M lines being parallel to the bright-field mirror as a consequence of the (001) *c*-glide plane. On the other hand, in the (001) reflection, $1 \neq 2n$ G–M lines, which are normal to the BF mirror, indicate the presence of a twofold screw axis along [001].

Confirmation that these dark bands are in fact G–M lines is given by the observation of "black crosses" when the Bragg reflection condition is satisfied for the forbidden reflections.

For example, Figure 21.10B shows a black cross in (001) and Figure 21.10C shows a black cross in (120). The *mmm* point group, combined with the presence of the *c*-glide plane and the twofold screw axis, permits the space group of $CaZrO_3$ to be determined as *Pcmn*.

21.4. LATTICE PARAMETER, STRAIN, AND COMPOSITION ANALYSIS

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 which are an order of magnitude more precise (~0.2%) can be obtained by computer simulation (see Section 20.5) of the position of the HOLZ lines using different lattice constants. The values which 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 which may occur for a variety of reasons, but materials scientists are most interested in changes due to composition or lattice strain. It is feasible to use this approach to infer, indirectly, the composition of a specimen by comparing the lattice parameter of an unknown with a standard of known composition; in the case of solid solutions, extrapolating to other compositions assuming that Vegard's law applies. Such a chemical analysis is not limited by elemental considerations, as is the case for XEDS, and does not require ultrathin specimens as for EELS. However, it is indirect and makes several assumptions which may not always apply. In addition to composition measurements we can also make very localized measurements of strain around precipitates or defects, but the data are always averaged through the foil thickness, and in one dimension, and so are rather limited.

For determining precise lattice parameters, we compare the experimental HOLZ-line pattern to the computer-simulated version.

In some programs, the position of the HOLZ lines in the simulation is derived from kinematical-diffraction theory only, but dynamical effects which may be important should really be included (Eades *et al.* 1993). The method is:

- 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.
- 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 Figures 21.11A and B). Theoretically, an accuracy of 0.02% should be achievable but in practice an accuracy of 0.2% is generally obtained.

While this approach has demonstrated reasonable success in measuring lattice-parameter shifts as shown by Randle *et al.* (1989), 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 impossible. Remember, you need to cool the specimen so you can see HOLZ lines in certain materials. There are some necessary precautions you should take:

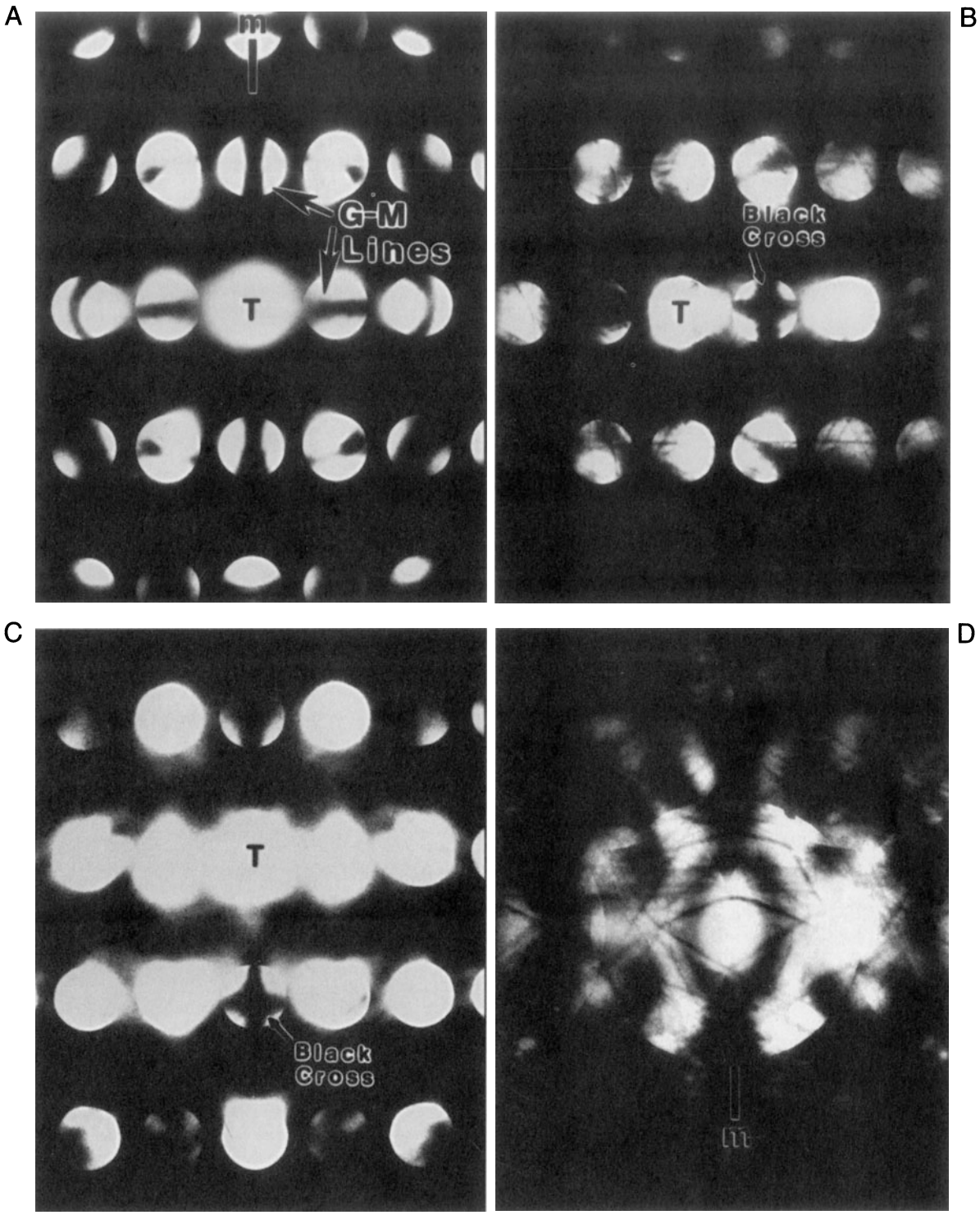


Figure 21.10. The use of G–M lines to determine the space group of $CaZrO_3$ (see text for a full explanation). In (A), orthogonal G–M lines are visible in alternate reflections, while in the (001) reflection (B) and the (120) reflection (C) "black crosses" are visible. In (D) the mirror symmetry is evident, with parallel G–M lines.

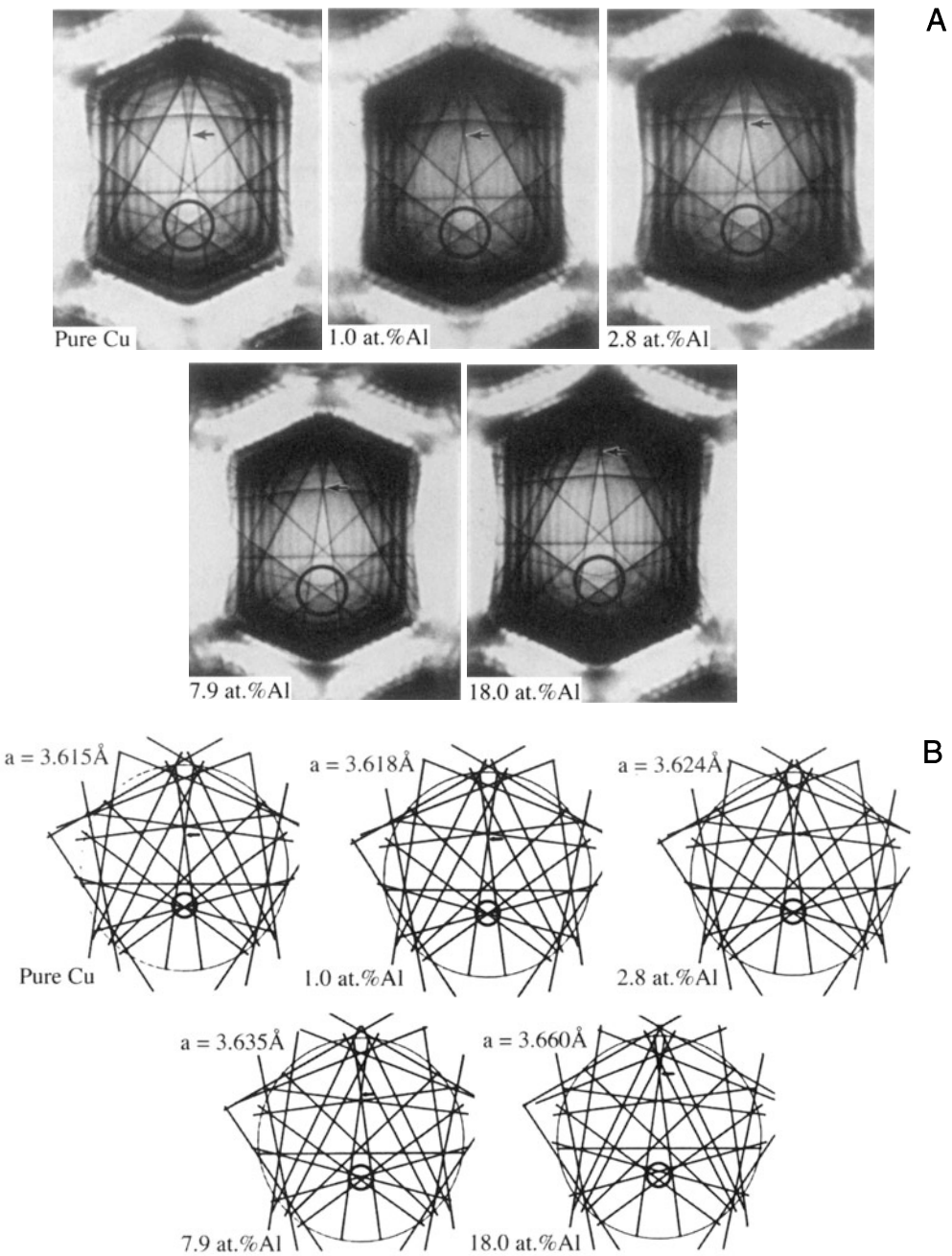


Figure 21.11. (A) HOLZ-line patterns obtained from several different Cu-Al alloys in the [114] orientation at fixed kV. The HOLZ-line shifts are due to changes in lattice parameter. (B) Computer simulation of the HOLZ-line patterns in (A) showing the lattice parameter that corresponds to each experimental pattern.

- 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.

While this is potentially a very precise technique, its accuracy may be less reliable, so it is probably best not to rely on HOLZ-line shifts as the primary method of absolute lattice parameter or strain determination.

21.5. DETERMINATION OF ENANTIOMORPHISM

Crystals belonging to point groups where only rotation axes are allowed can be either right-handed or left-handed. Another unique aspect of CBED patterns is shown in Figure 21.12. We can use CBED patterns to deduce whether a crystal shows right- or left-handedness, since under these circumstances the pattern symmetries vary depending on which way the beam enters the specimen. Thus, the handedness of a crystal can be readily determined using CBED. If a sample is enantiomorphous (i.e., it has no symmetry elements which change "hand," such as inversion centers or mirrors), then the patterns are different when the beam enters the specimen from opposite directions. So you take patterns before and after turning the specimen upside down. In crystals which do not possess a handedness, CBED patterns obtained with illumination from the top or bottom surface of the crystal are identical. However, in right/left-handed crystals (e.g., quartz) the CBED patterns obtained in this way are no longer identical, but are related to each other by a mirror (or a twofold rotation axis). This is clearly the case in Figure 21.12, which is from enantiomorphous quartz (space group $P3_121$ or $P3_221$).

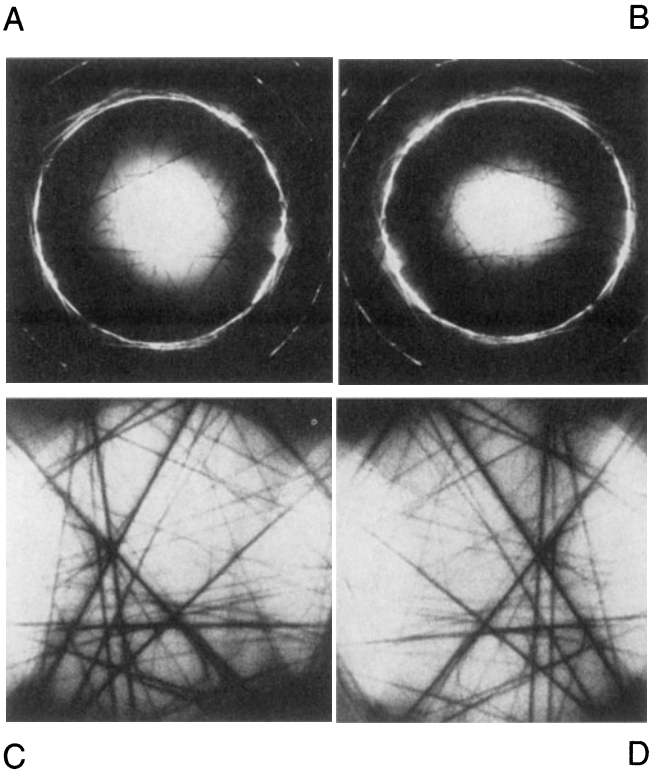


Figure 21.12. (A) and (C) are WP and BF disks, respectively, from a quartz specimen illuminated from the top surface. (B) and (D) are the same patterns, except that the specimen was inverted to illuminate the bottom surface. Note the mirror symmetries between (A) and (B), as well as between (C) and (D).

21.6. CONVERGENT-BEAM IMAGING

You can obtain some direct correlation between CBED symmetry and the image of the specimen. The technique of convergent-beam imaging (CBIM) (Tanaka *et al.* 1988, Humphreys *et al.* 1988) creates a mixture of both image and DP in one. By adjustment of condenser and objective lenses, either the image or the DP can be focused, but not both at the same time.

This technique is analogous to the multiple dark-field imaging in TEM in which BF and DF images are visible in defocused SAD patterns (see Figure 9.25).

CBIM is carried out by focusing the convergent beam either above or below the specimen plane, in which case HOLZ lines are projected, although somewhat broadened, into the normal image plane. So you should use the smallest possible probe, with reasonable defocus values; then the HOLZ-line resolution, governed by the probe size, remains acceptable. This technique has the advantage that your image resolution is that expected from the microscope under normal diffraction-contrast conditions. For example, in Figure 21.13, the change in symmetry across a two-phase interface between NiO and CaO is clearly visible, as is the diffraction contrast associated with the interface. The insets show a TEM image of the interface and an SAD pattern across the interface showing the orientation relationship between the phases.

21.7. SCANNING-BEAM DIFFRACTION

Diffraction phenomena can be imaged in scanning-beam instruments, just as in TEM. For example, the phenomenon of electron channeling is one way to obtain crystallographic information from a bulk crystal in the SEM. This technique is somewhat outside the scope of this book; however, channeling is possible in a STEM or DSTEM and channeling patterns contain similar HOLZ information as CBED patterns. If you're interested you should look up any standard SEM text.

Scanning-diffraction patterns are obtainable in STEMs using either one or two sets of coils both before

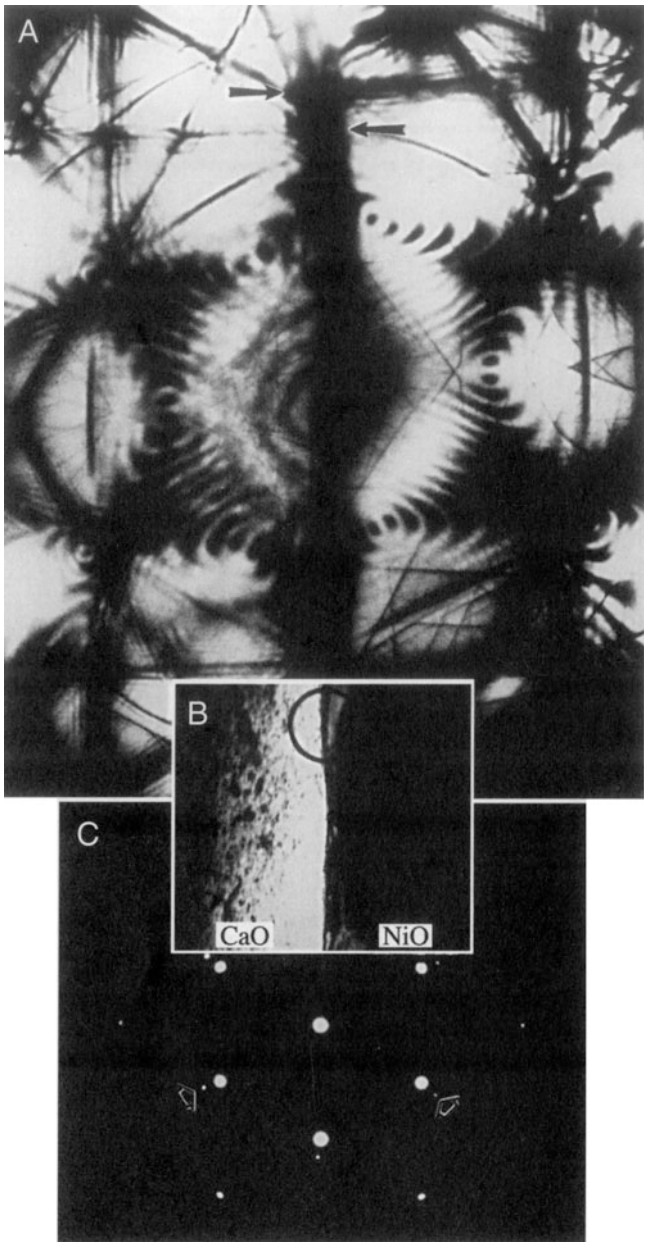


Figure 21.13. (A) CBIM pattern of a NiO-CaO eutectic interface (B). In this directionally solidified eutectic all directions and planes in both phases are parallel to each other, as evident in the SAD pattern (C). The CBIM pattern displays the parallelism of (220) mirror planes which are normal to the planar interface, and the lack of continuity of HOLZ lines and Kikuchi bands indicates the lattice mismatch across the interface.

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 rocking-beam method of microdiffraction (see Section 21.8). Using only one scan coil below the specimen partially "de-rocks" the beam, but two coils

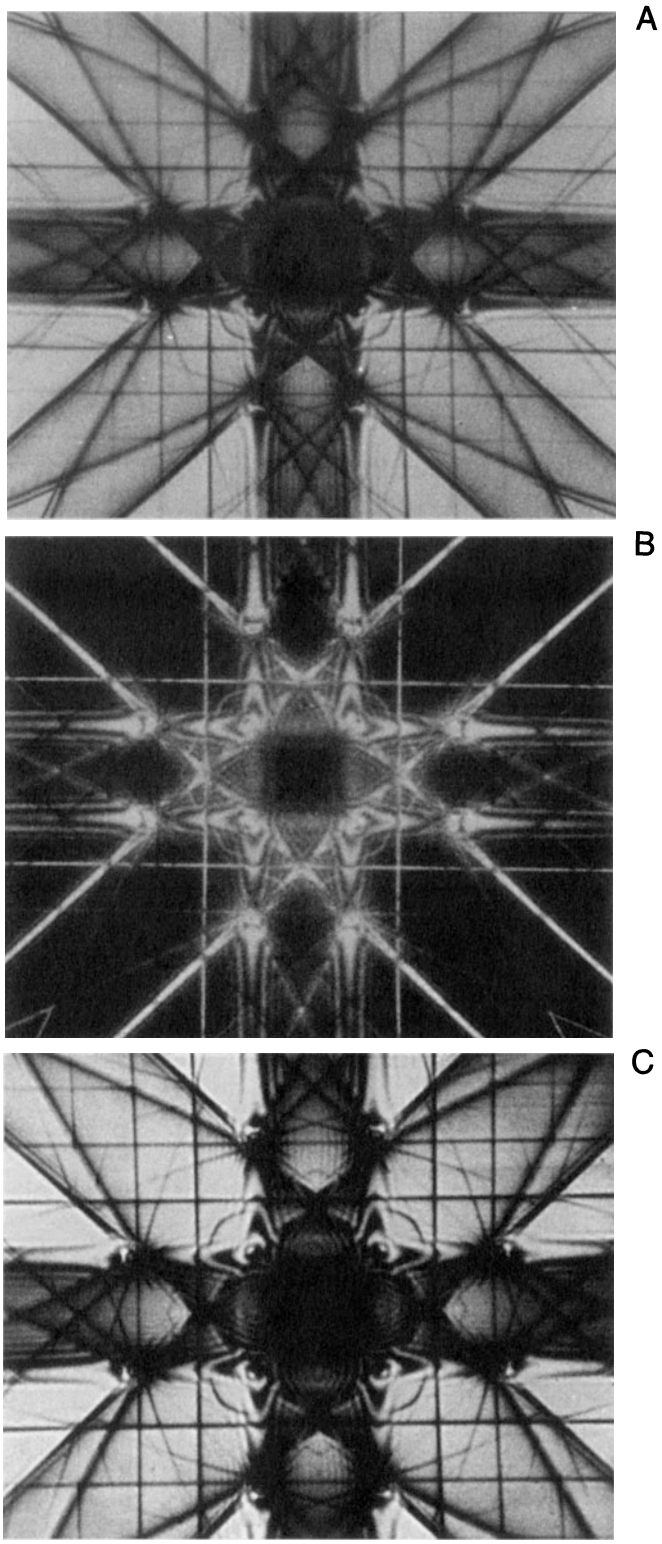


Figure 21.14. A series of Eades double-rocking zone-axis patterns obtained from a thin sample of aluminum in the [001] orientation. (A) is the BF image, (B) is a DF image, and (C) is an energy-filtered version of (A). The removal of energy-loss electrons sharpens the image.

fully "de-rock" the beam such that, instead of having the DP scan across the BF STEM detector, 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 patterns cover several degrees in comparison with the fractions of a degree visible in a normal CBED disk. The resultant patterns are very striking, as you can appreciate from Figure 21.14; such patterns can be viewed in BF or DF. In addition, the images can be sharpened by sending the electrons through an energy filter (Figure 21.4C) prior to displaying the pattern on the CRT (see Chapter 40). These approaches can be used to study the occurrence of forbidden reflections, which are important in crystal-symmetry determinations.

Grigson scanning is a method of obtaining DPs in a DSTEM without any post-specimen lenses. In such an instrument, the DP is only visible if it is scanned across the STEM detector and displayed on the CRT. In a DSTEM with an FEG source but no post-specimen lenses, the pattern is viewed by stopping the scan and positioning the beam on the area of interest. With an FEG, DPs can be obtained from subnanometer-diameter regions if the specimen is thin enough. The Grigson method is highly inefficient and collects only a small fraction of the DP at a time: it's a serial technique. More usually, in a DSTEM the pattern is viewed directly from a TV recording of the back focal plane of the objective lens. Almost always, post-specimen lenses are now included in a DSTEM to allow you to use a range of camera lengths. This method of very high resolution microdiffraction, sometimes termed "nanodiffraction," has been pioneered by Cowley and co-workers (Cowley 1981). Figure 21.15 shows a sequence of DPs obtained across a single carbon nanotube.

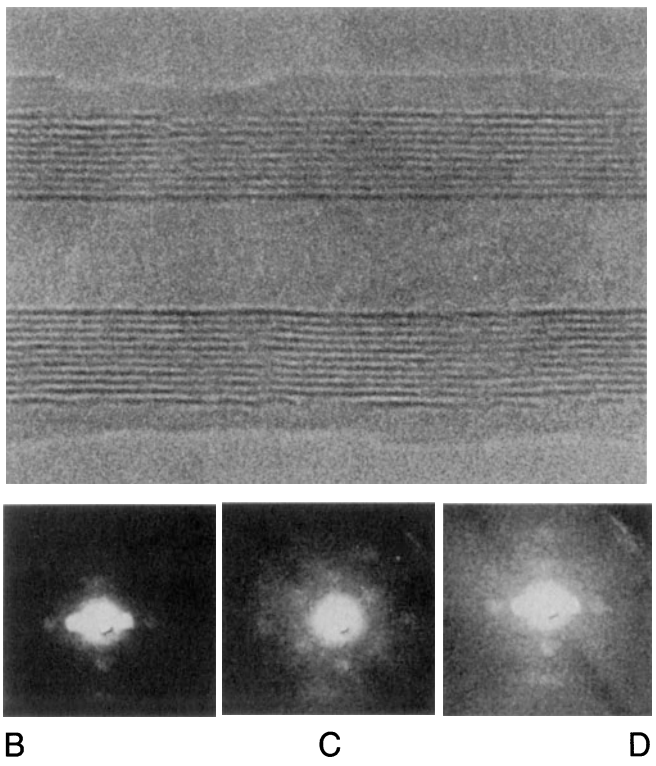


Figure 21.15. (A) A single carbon nanotube of circular cross section and zero helix angle. The 0.34-nm 0002 graphite lattice fringes are clearly resolved. (B–D) Nanodiffraction patterns taken from the upper, center, and lower regions of the nanotube, respectively.

21.8. OTHER METHODS OF MICRODIFFRACTION

It is possible to obtain microdiffraction patterns in TEMs and STEMs by methods other than CBED, although no other technique combines the versatility and ease of CBED. The need for diffraction information from below the SAD limit was originally addressed by Riecke (1962). He developed a strong C3 lens, or mini-lens, which permitted a large demagnification of the C2 aperture onto the specimen. While this approach was similar to SAD, the available demagnification was much larger. However, the state of development of electron optics at that time was such that, to get the best area selection, it was necessary to operate with the specimen raised above the eucentric plane, which meant that few users were prepared to try the technique. This limitation has been overcome with the development of condenser-objective lenses and the Riecke method can still be used to produce patterns from regions of a few tens of nanometers in diameter. In a modern TEM the equivalent method is often called "nanoprobe" mode. You first create a fine probe in TEM mode with a small C2 aperture, defocus the probe to give parallel illumination, then switch to diffraction mode to view the pattern. Because these patterns are obtained with a parallel beam the diffraction spots are points, not disks, and an example is shown in Figure 21.16.

With the development of STEMs, it became possible to obtain diffraction information in a unique manner by using the scan coils to rock the beam through a range of angles above the specimen. This "rocking-beam" method has been refined to the point where patterns from regions below 5 nm are obtainable. The pattern is created by the varying intensity that falls on the STEM BF detector as the beam changes its angle of incidence to the spec-

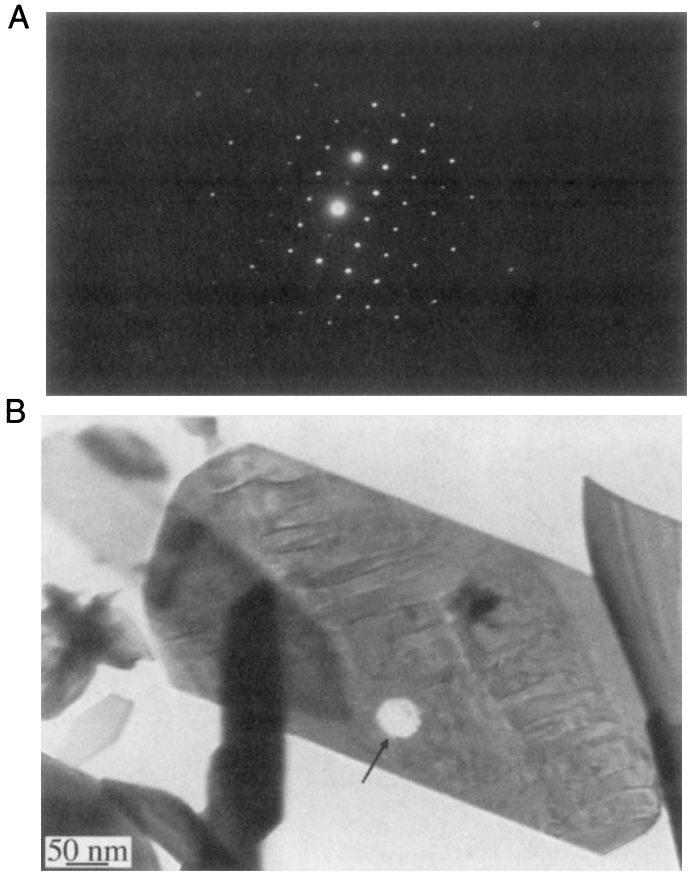


Figure 21.16. (A) A Riecke diffraction pattern from a single crystal of MoO_3. (B) A double exposure of the crystal and the bright area (arrowed) illuminated by the small parallel beam. The pattern is identical to an SAD pattern except that it comes from a region of ~50-nm diameter.

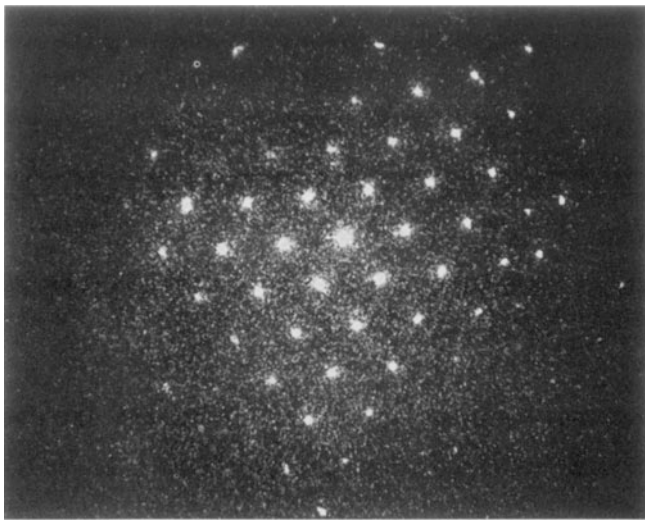


Figure 21.17. Rocking-beam pattern of graphitized carbon from the CRT screen of a DSTEM.

imen, and is recorded on the STEM CRT. In a DSTEM, such patterns are also useful to determine the collection angle of the EELS spectrometer, and an example of a rocking-beam pattern is shown in Figure 21.17. Neither Riecke nor rocking-beam techniques are extensively used. They are historically interesting, but do not compete in any way with CBED, which is by far the best method of obtaining crystallographic information about your specimen.

CHAPTER SUMMARY

CBED patterns contain contrast information which can be used to give the point group, space group, crystal system, and lattice parameter of very small crystals, as well as other information such as thickness and enantiomorphism. To do this requires some 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. If you want to determine the point group, you need to obtain from your CBED patterns:

- WP symmetry.
- BF symmetry.

These symmetries lead you to the diffraction group, which describes the full 3D symmetry of a DP and is directly related to the point group through Buxton's tables.

If you successfully worked your way through these two chapters, then you're ready for the more challenging aspects of CBED, such as measurement of structure factors, Debye–Waller factors, and charge density discussed in the text by Spence and Zuo (1992).

REFERENCES

General References

Eades, J.A. (1988) *Ultramicroscopy* **24**, 143.

Eades, J.A., Ed. (1989) *J. Electron Microsc. Tech.* **13** (I & II). Special issues on CBED.

Loretto, M.H. (1994) *Electron Beam Analysis of Materials,* 2nd edition, Chapman and Hall, New York.

Mansfield, J.F. (1984), *Convergent Beam Electron Diffraction of Alloy Phases,* Adam Hilger, Bristol, United Kingdom.

Spence, J.C.H. and Zuo, J.M. (1992) *Electron Microdiffraction*, Plenum Press, New York. You will want a copy of this if you use CBED extensively. The code for plotting HOLZ lines is included in the appendices along with the Fortran code for two programs, one Bloch wave and one multislice. You may also find a reference to earlier CBED studies on your material here in the selective bibliography organized by material.

Sung, C.M. and Williams, D.B. (1991) *J. Elect. Microsc. Tech.* **17**, 95. A bibliography of CBED papers from 1939–1990.

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.

Specific References

Allen, S.M. (1981) *Phil. Mag.* **A43**, 325.

Ayer, R. (1989) *J. Electron Microsc. Tech.* **13**, 16.

Buxton, B.F., Eades, J.A., Steeds, J.W., and Rackham, G.M. (1976) *Phil. Trans. Roy. Soc.* **281**, 181.

Cowley, J.M. (1981) *Ultramicroscopy* **7**, 19.

Eades, J.A., Moore, S., Pfullman, T., and Hangas, J. (1993) *Microscopy Research and Technique* **24**, 509.

Gjønnes, J. and Moodie, A.F. (1965) *Acta Cryst.* **19**, 65.

Hahn, T., Ed. (1988) *International Tables for Crystallography,* A. *Space-group Symmetry,* Kluwer Academic Publisher, Norwell, Massachusetts.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, p. 122, Krieger, Huntington, New York.

Humphreys, C.J., Maher, D.M., Fraser, H.L., and Eaglesham, D.J. (1988) *Phil. Mag.* **58A**, 787.

Jackson, A.G. (1987) *J. Electron Microsc. Tech.* **5**, 373.

Jackson, A.G. (1990) *Ultramicroscopy* **32**, 181.

Kelly, P.M., Jostons, A., Blake, R.G., and Napier, J.G. (1975) *Phys. stat. sol.* **A31**, 771.

Raghavan, M., Scanlon, J.C., and Steeds, J.W. (1984) *Met. Trans.* **15A**, 1299.

Randle, V., Barker, I., and Ralph, B. (1989) *J. Electron Microsc. Tech.* **13**, 51.

Riecke, W.D. (1962) *Proc. 5th Int. Cong. on Electron Microscopy,* **1** (Ed. S.S. Breese Jr.), p. KK-5, Academic Press, New York.

Steeds, J.W. (1979) *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 387, Plenum Press, New York.

Steeds, J.W. (1984) "Electron Crystallography," in *Quantitative Electron Microscopy,* (Eds. J.N. Chapman and A.J. Craven), p. 49, Scottish Universities Summer School in Physics, Edinburgh.

Steeds, J.W. and Vincent, R. (1983) *J. Appl. Cryst.* **16**, 317.

Tanaka, M. (1989) *J. Electron Microsc. Tech.* **13**, 27.

Imaging

III

Imaging in the TEM

22

CHAPTER PREVIEW

We've already mentioned back 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 these kinds 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 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 BF and DF images and we'll discuss the important differences between the images formed in each of these two modes of operation. We'll then go on to discuss the principles of diffraction contrast, which are sufficiently complex that it takes Chapters 23–26 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 distort the diffracting planes. Therefore, the diffraction contrast from regions close to the defect would 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 27–30. Other forms of TEM imaging and variations on these major types of contrast are gathered in the catch-all Chapter 31.

Imaging in the TEM

22

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 photograph. However, if your image is digitally recorded, you can enhance low contrast electronically to levels at which your eyes can perceive it. We'll return to image processing and contrast enhancement in Chapter 30.

So we see contrast in TEM images as different levels of green light coming from the viewing screen or CRT. On the photograph, 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, 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/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, while if you try and increase the number of electrons falling on the screen, by condensing the beam onto a reduced area of the specimen, you'll usually lower the image contrast. These points are summarized in Figure 22.1, which defines intensity and contrast.

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. PRINCIPLES OF IMAGE CONTRAST

22.2.A. Images and Diffraction Patterns

If you go back and look at Figure 2.1, you'll see that the uniform electron intensity in the incident beam is transformed into a nonuniform 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 or CRT. Now you also know that the DP shows you this nonuniformity because it separates out 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 need to view the DP first, whatever contrast mechanism you want to use, and whatever 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 some of the diffracted beams in the SAD pattern to form BF and

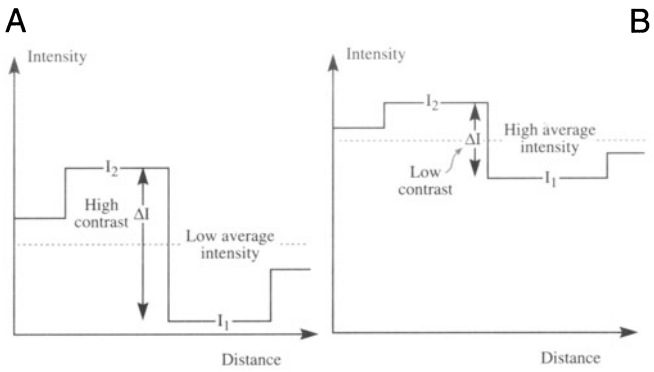


Figure 22.1. Schematic intensity profiles across an image showing (A) different intensity levels (I_1 and I_2) and the difference (ΔI) between them, which defines the contrast. Generally, in a TEM, if the overall intensity is increased (B) the contrast decreases.

DF images, respectively. Note we are justified in using the "beam" terminology, since the electrons have left the specimen. We've already seen back in Section 9.3 that in a TEM we select the direct or a scattered electron beam with the objective aperture. Remember, if you form an image without the aperture, the contrast will be poor because many beams then contribute to the image. Furthermore, aberrations due to the off-axis electrons will make your image impossible to focus. Your choice of the aperture size governs which electrons contribute to the image and thus you control the contrast.

Figure 22.2 shows a DP from a single-crystal Al specimen with the schematic indication of the objective aperture indicated. In this figure, the aperture in position A is selecting 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. Usually

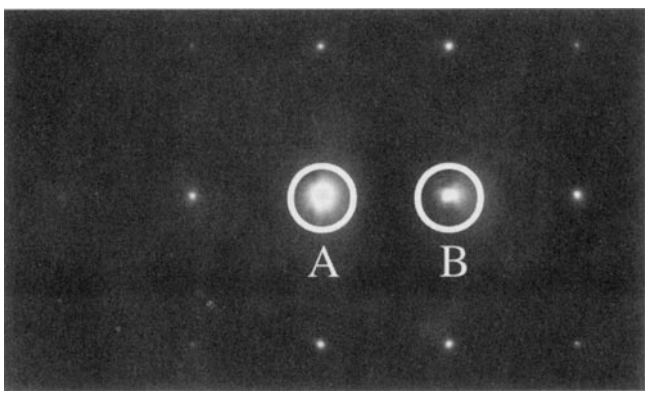


Figure 22.2. The relationship between the objective aperture and the diffraction pattern for forming (A) BF and (B) DF images.

we tilt the incident beam such that the scattered electrons remain on axis, creating 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. *Thus, BF and DF are the two basic ways to form amplitude-contrast images.* We'll see later 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.

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 conjugate plane to the back focal plane. We control which electrons fall on which detector and thus contribute to the image by adjusting the post-specimen (imaging)

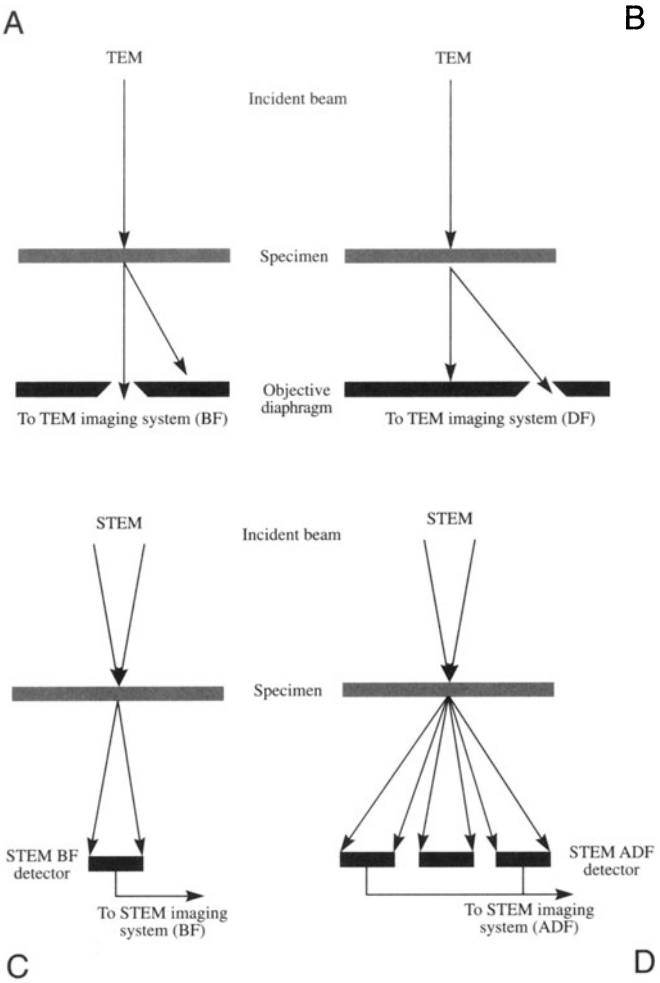


Figure 22.3. Comparison of the use of an objective aperture in TEM to select (A) the direct or (B) the scattered electrons forming BF and DF images, respectively. In STEM we use (C) an on-axis detector or (D) an annular detector to perform equivalent operations.

lenses to change the camera length. Clearly, for DF imaging, the ADF detector gathers 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.

22.3. MASS-THICKNESS CONTRAST

Mass-thickness contrast arises from incoherent (Rutherford) elastic scatter of electrons. As we saw back in Chapter 3, the cross section for Rutherford scatter is a strong function of the atomic number Z, i.e., the mass or the density, ρ, as well as 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 (<~5°), mass-thickness contrast dominates (but it also competes with Bragg-diffraction contrast). However, we'll also see that at high angles (>5°), where coherent scattering is negligible, we can pick up low-intensity, incoherently scattered beams. 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; in a DSTEM, however, we can obtain these images with atomic resolution. 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!

Mass-thickness contrast is most important if you are looking at noncrystalline 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 and so 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 gather many beams and thus minimize any diffraction contrast. As you'll see, you should still use an objective aperture to enhance the mass-thickness contrast, i.e., 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 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.

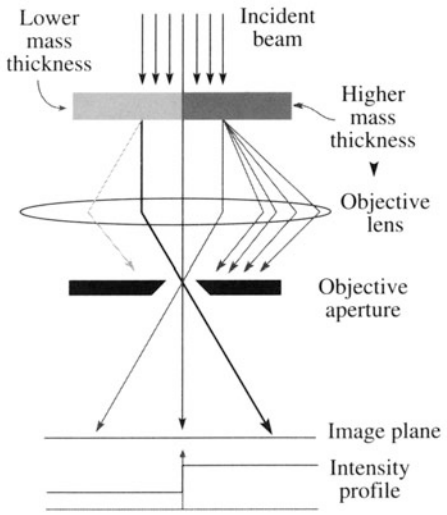


Figure 22.4. Mechanism of mass-thickness contrast in a BF image. Thicker or higher-Z areas of the specimen (darker) will scatter more electrons off axis than thinner or lower-mass (lighter) areas. Thus fewer electrons from the darker region fall on the equivalent area of the image plane (and subsequently the screen), which therefore appears darker in BF images.

This is all you need to know to interpret 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," which 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 27) 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. To tell something about the shape, you need to shadow them, i.e., evaporate a thin heavy metal (Au or Au-Pd) coating at an oblique angle as shown in Figure 22.5B. The shadow shape reveals the true shape of the particles (Watt 1985).

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. There is also an intensity change across the latex spheres reflecting the preferential deposit of Au-Pd on the side of the sphere toward the source of the evaporated metal.

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

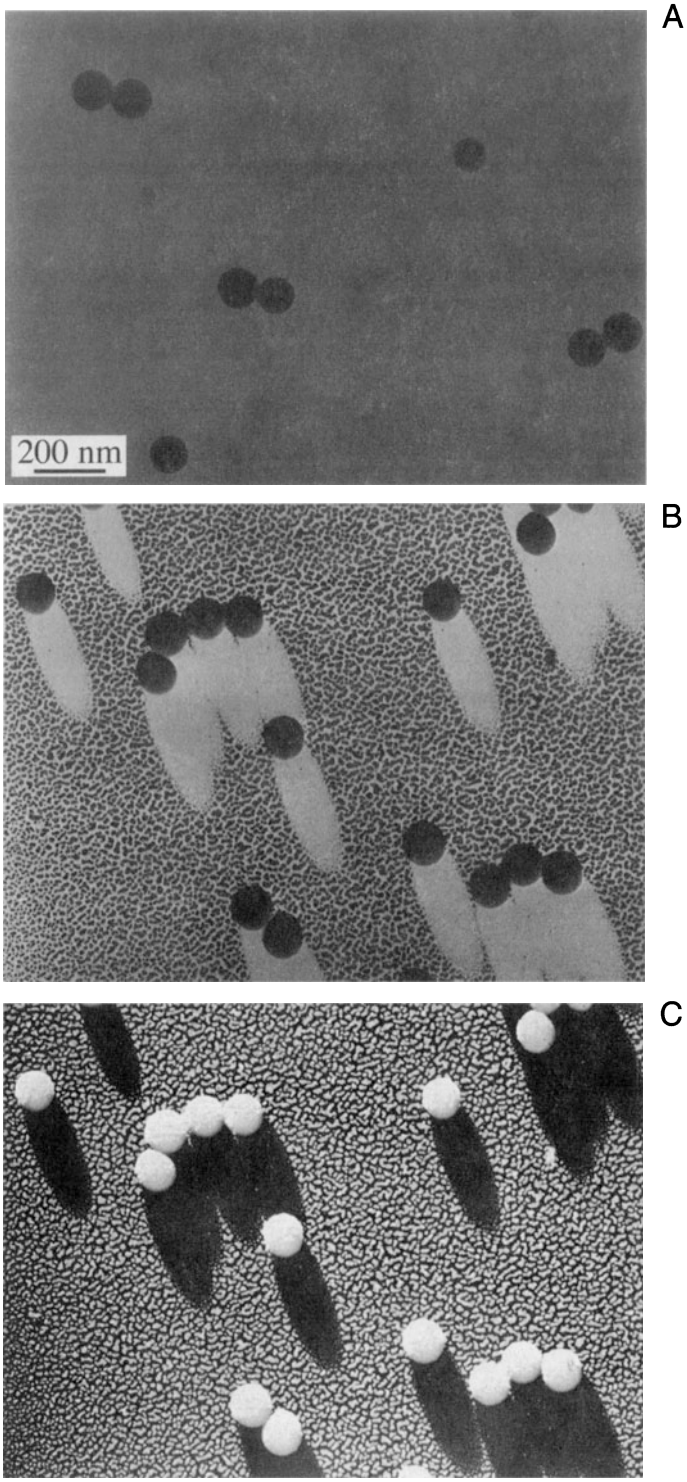


Figure 22.5. (A) TEM BF image of latex particles on a carbon support film showing thickness contrast only. (B) Latex particles on a carbon film shadowed to reveal the shape of the particles through the addition of selective mass contrast to the image. (C) Reverse print of (B) exhibits a 3D appearance.

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 two-dimensional 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 polymers and biological specimens with heavy metals such as Os, Pb, and U (Sawyer and Grubb 1987). 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.

The TEM variables which 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 nonscattering 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. Of course, any decrease in intensity can be offset by increased exposure times until specimen drift becomes a limiting factor.

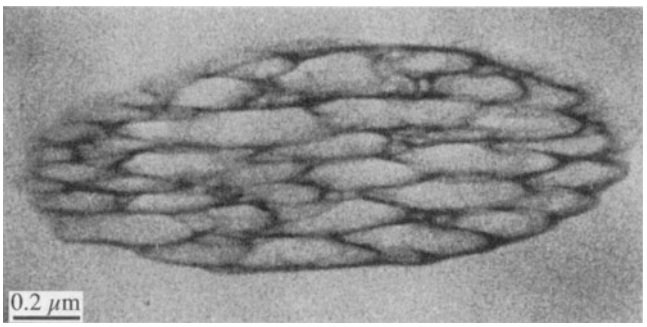


Figure 22.6. BF image of stained two-phase polymer exhibiting mass contrast due to the segregation of the heavy metal atoms to the unsaturated bonds in the darker phase.

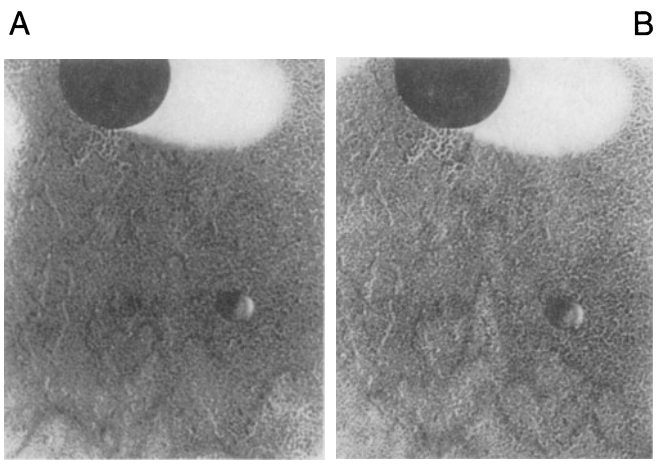


Figure 22.7. The effect of objective aperture size on mass-thickness contrast; the images of the shadowed latex particle were taken with an aperture size of (A) 70 μm and (B) 10 μm. A smaller aperture enhances the contrast, in a similar manner to lowering the kV.

Now for DF images, there isn't much more to be said and the images will generally show complementary contrast to BF, as we saw in Figures 22.5B and C. 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 excellent 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. 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 image generally shows 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 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

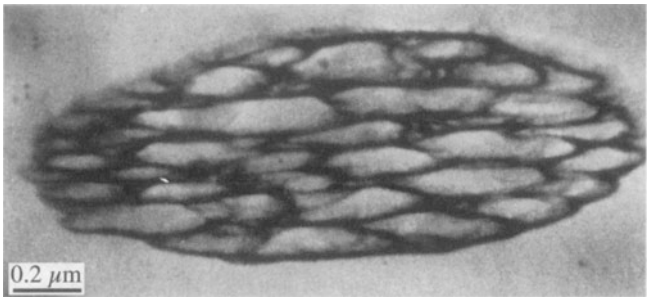


Figure 22.8. STEM BF image of a stained two-phase polymer. Comparison with the TEM image in Figure 22.6 shows that, while the contrast is higher in STEM, the image resolution is poorer.

you record your TEM image using a CCD camera, or digitize the negative, you can enhance the contrast (see Chapter 30). A good way to do this is with differential hysteresis imaging (Section 1.5), as you can see in Figure 22.9C.

> 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 CRT; such options aren't available for analog TEM images.

In a STEM 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.

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, 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 CRT and adjusting L.

> The STEM must be well aligned so the DP expands and contracts on axis.

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. 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.

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 an FEG STEM.

The comparison we've made of TEM and STEM images here is qualitative, but there have been many rigorous comparisons of STEM and TEM contrast, particularly

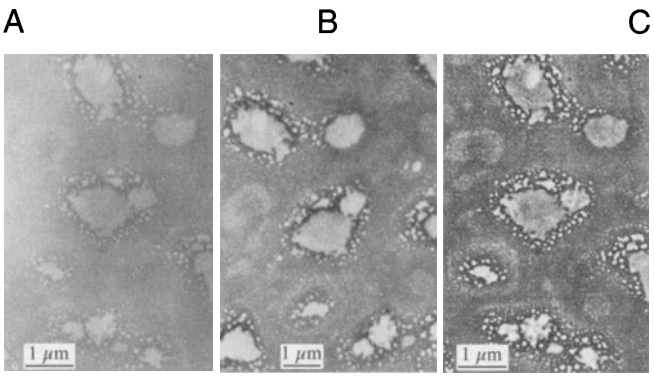


Figure 22.9. Comparison of TEM (A) and STEM (B) images of an amorphous SiO_2 specimen containing Cl-rich bubbles. The low mass contrast in the TEM image can be enhanced in a STEM image through signal processing. (C) A similar effect can be achieved by digitizing the TEM image (A) and applying contrast-enhancement software.

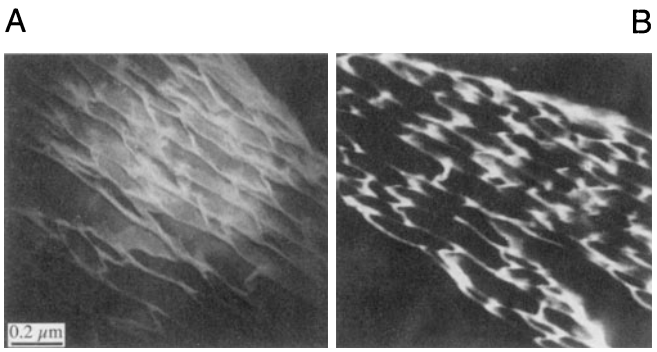


Figure 22.10. Comparison of (A) TEM DF and (B) STEM ADF images of the same two-phase polymer as in Figures 22.6 and 22.8. As in BF the STEM image shows higher contrast but lower resolution. Also, the ADF aperture collects more signal than the TEM objective aperture so the STEM image is less noisy.

for biological specimens (e.g., Cosslett 1979). 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 classical 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 three reasons 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.

22.3.D. Specimens Which Show Mass-Thickness Contrast

Mass-thickness is the primary contrast source in amorphous materials, which is why we've illustrated this portion of the 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 unshadowed (Figure 22.11A) or shadowed (Figure 22.11B). The uneven metal shadowing increases the mass contrast and thus accentuates the topography; see also the latex particles in Figure 22.3. An extraction replica (Figure 22.11C) or particles dispersed on a support film will also show mass-thickness contrast, and shadowing could be useful to reveal the shape of the particles. If the particles are crystalline there will also be a component of diffraction 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 (i.e., the scattering is truly incoherent). This is not the case, since even DPs from amorphous specimens show diffuse rings rather than uniform intensity (Figure 2.5D). 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 (e.g., Heidenreich

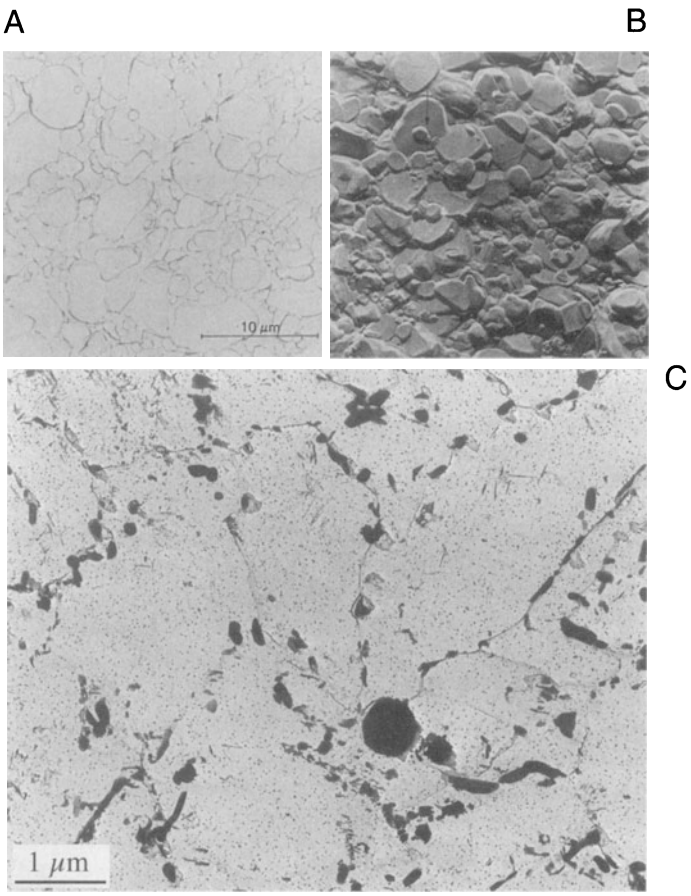


Figure 22.11. More examples of mass-thickness contrast: (A) a carbon replica of a fracture surface doesn't show much of either form of contrast until (B) oblique shadowing enhances the topography. (C) An extraction replica of a range of small precipitate particles in a Cr-Mo steel weld shows both mass and thickness contrast.

1964) that a change in thickness, Δt, at constant atomic number Z creates contrast

$$\frac{\Delta I}{I} = 1 - e^{-Q\Delta t} = Q\Delta t \qquad [22.2]$$

for $Q\Delta t < 1$, where Q is the total elastic scattering cross section. Since the minimum contrast we can see is ~5%, then the minimum Δt that we can see is

$$\Delta t \cong \frac{5}{100\,Q} = \frac{5\,A}{100\,N_0\sigma\rho} \qquad [22.3]$$

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 changes). So, if we want to calculate the contrast, we need to know σ. As we've seen in equation 3.9, 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 3.10

$$f(\theta) = \frac{\left(1 + \frac{E_0}{m_0c^2}\right)}{8\pi^2 a_0}\left(\frac{\lambda}{\sin\frac{\theta}{2}}\right)^2 (Z - f_x) \qquad [22.4]$$

The Z term represents the Rutherford scattering. For unscreened Rutherford scattering (i.e., ignoring the effects of the electron cloud) σ is proportional to Z^2. This unscreened behavior is approximated by electrons scattered through semiangles above ~5° (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 > θ_0, the screening parameter, defined back in equation 3.6.

We can use the unscreened Rutherford scattering expression to determine the probability that an electron will be scattered through greater than a given angle. To do this, we integrate the differential Rutherford cross section from an angle β (defined by the semiangle of collection of the objective aperture) to infinity. Thus

$$\sigma(\beta) = 2\pi\int_{\beta}^{\infty} |f(\theta)|^2\, \theta\, d\theta \qquad [22.5]$$

which can be evaluated (see Reimer 1993) to give

$$\sigma(\beta) = \frac{\left[Z\lambda\left(\frac{a_0}{Z^{0.33}}\right)\left(1 + \frac{E_0}{m_0c^2}\right)\right]^2}{\pi(a_0)^2\left(1 + \left(\frac{\beta}{\theta_0}\right)^2\right)} \qquad [22.6]$$

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 > β. 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

$$\frac{dn}{n} = -N\,\sigma(\beta)\,dx \qquad [22.7]$$

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

$$\ln n = -N\sigma x + \ln n_0 \qquad [22.8]$$

and if we rearrange this expression, we obtain

$$n = n_0\, e^{-N\sigma x} \qquad [22.9]$$

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 somewhat 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 and 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) imaging technique, but rather than discuss it in Chapter 28 under phase-contrast effects we'll talk about it here, because it represents the limit of mass-thickness contrast where detectable scattering arises from single atoms or a 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 (Isaacson *et al.* 1979), 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 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-

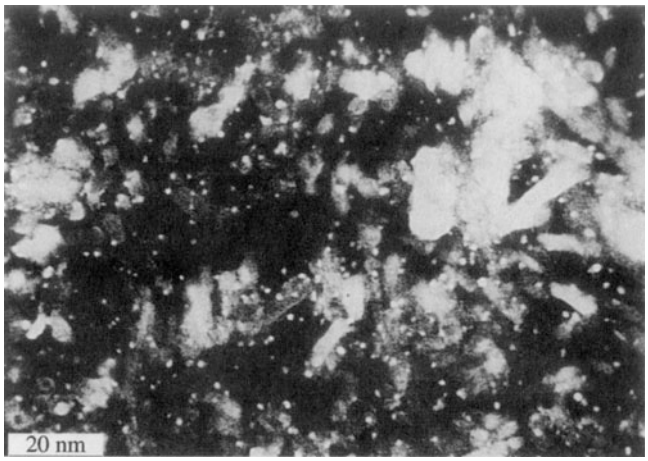


Figure 22.12. Z-contrast ADF image of individual Pt atoms or groups of atoms on a crystalline Al_2O_3 film obtained using an FEG STEM.

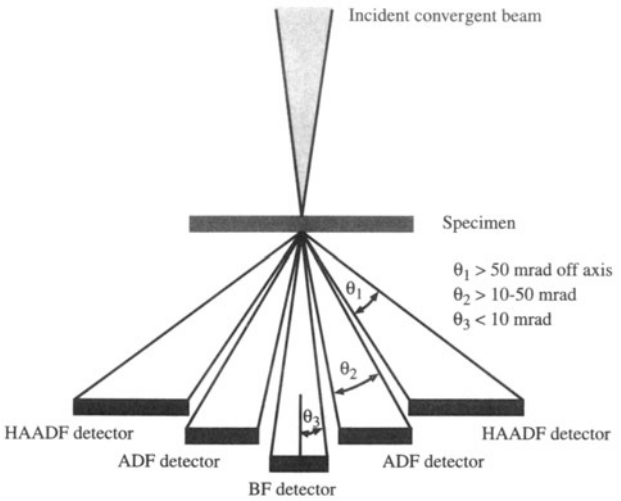


Figure 22.13. Schematic of the HAADF detector set-up for Z-contrast imaging in a STEM. The conventional ADF and BF detectors are also shown along with the range of electron scattering angles gathered by each detector.

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.

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) (Jesson and Pennycook 1995). 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. The image is thus formed only from the very high angle, incoherently scattered electrons.

- The detector is called a high-angle ADF or HAADF detector.
- Sometimes, the term "Howie detector" is used since Howie (1979) first proposed its design.
- *Z*-contrast images are also termed HAADF images.

Bragg effects are avoided if the HAADF detector only gathers electrons scattered through a semiangle 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 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 25) 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 scattering 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. Pennycook (1992), who pioneered this technique, found he could quantify the intensity to an absolute accuracy of ±20%.

In an FEG with probe sizes of < 0.3 nm, *Z*-contrast image resolution of this order 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. Fig-

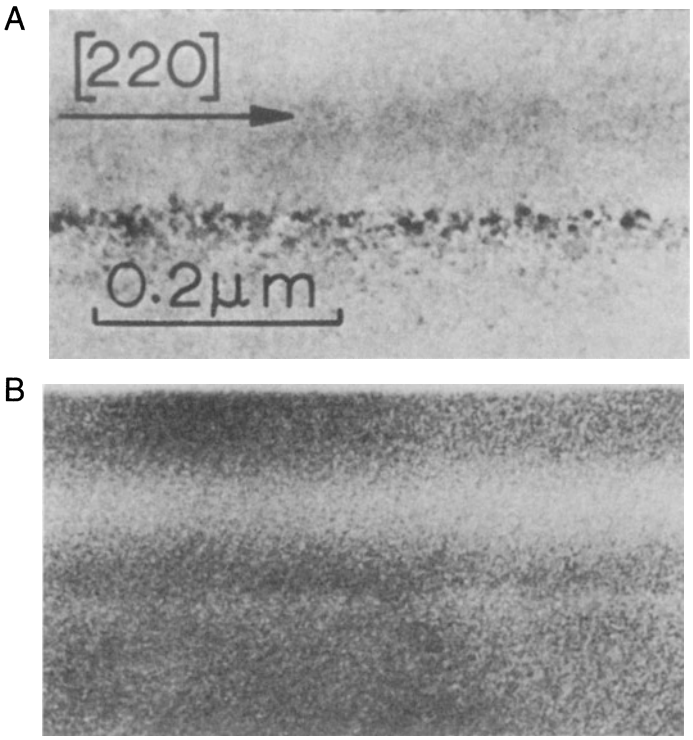


Figure 22.14. (A) Low-resolution TEM BF image showing a row of defects in Bi-implanted Si. In (B), obtained under *Z*-contrast conditions, the defects associated with the implant are invisible but the specimen is bright in the region implanted with Bi.

ure 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 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.

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.

> The $f(\theta)$ map can have atomic level resolution, which XEDS imaging can't provide.

So why do we need a STEM for *Z*-contrast imaging? We are constrained in TEM if we use an analog screen

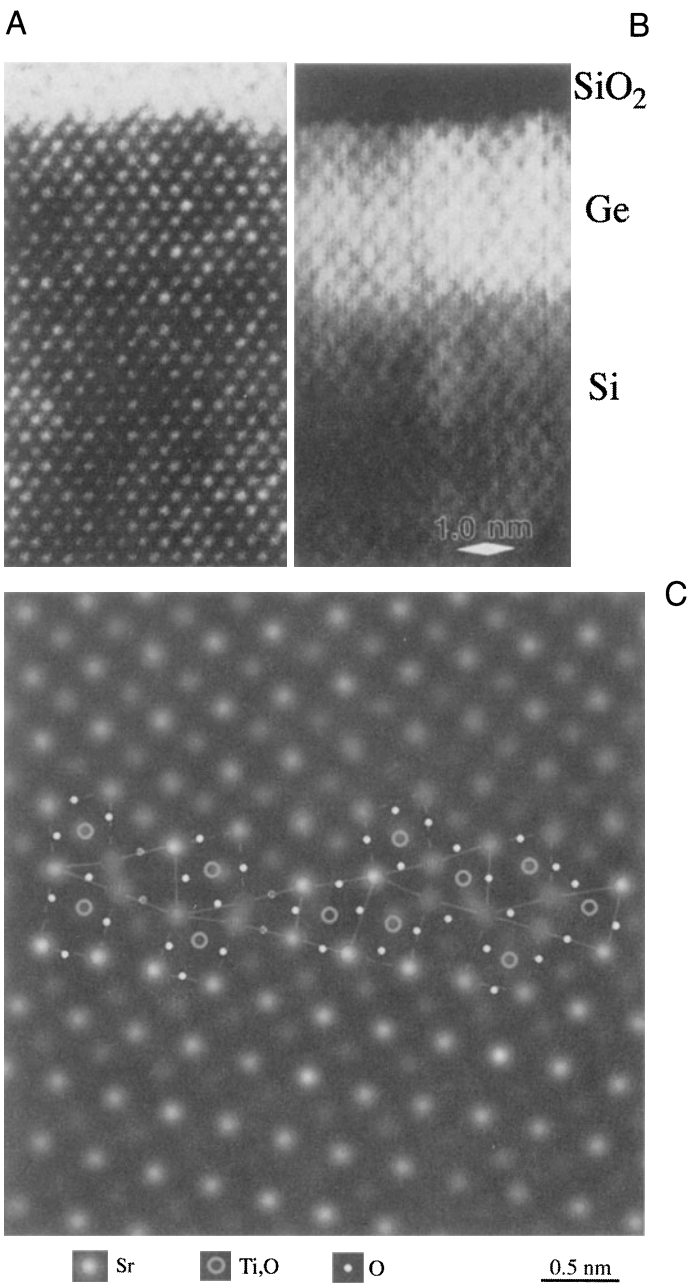


Figure 22.15. (A) High-resolution phase-contrast image of epitaxial Ge on Si with an amorphous SiO_2 surface. The bright array of dots common to the crystalline region represents atomic rows and the Ge and Si regions are indistinguishable. (B) The high-resolution *Z*-contrast STEM image shows the atom 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.

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 reci-

procity" 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.

22.5. TEM DIFFRACTION CONTRAST

Bragg diffraction, as you now know from Part II, 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 because 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 round the direct beam (Figure 22.2A) and DF images come from 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 used *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 $hk\ell$ beam have been diffracted by a *specific* set of $hk\ell$ planes and so the area that appears bright in the DF image is the area where the $hk\ell$ 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 $hk\ell$ 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.

As you can see in Figures 22.16B, and C, 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 essential for viewing crystalline specimens.

> If you're working with crystalline materials, you'll spend a lot of time tilting the specimen to set up different two-beam conditions.

We'll see in the following chapters that two-beam conditions are not only necessary for good contrast, they also greatly simplify interpretation of the images. This is why we emphasized two-beam theory in our discussion of diffraction in Part II.

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 are 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 $\mathbf{s}$ small and positive (the excess $hk\ell$ Kikuchi line just outside the $hk\ell$ spot). This will give you the best possible strong-beam image contrast, as in Figure 22.17B. If you tilt the specimen slightly, so $\mathbf{s}$ increases further as shown in Figure 22.17C, the defect images become narrower but the contrast is reduced.

> Never form strong-beam images with **s** negative; the defects will be difficult to see.

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

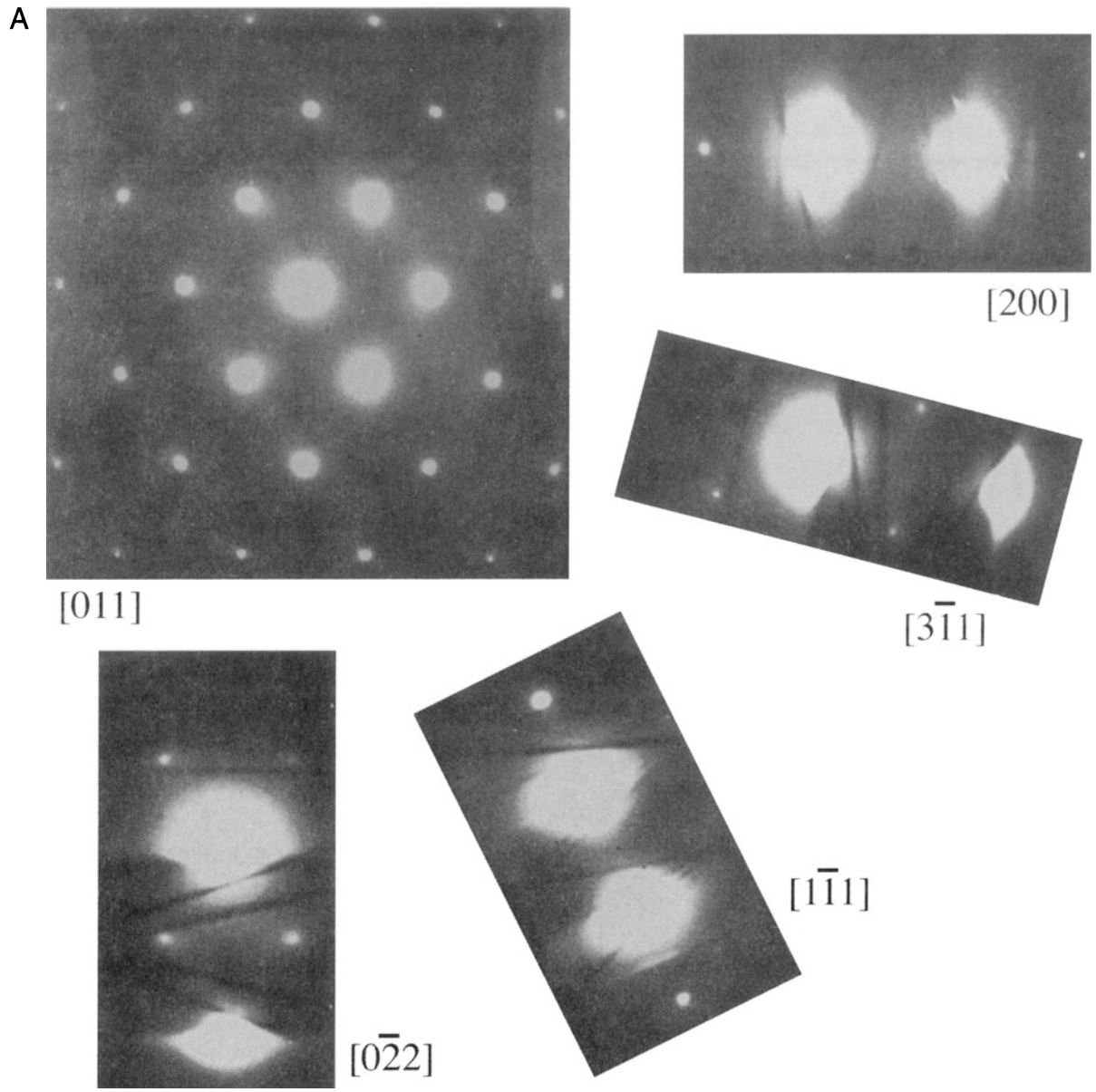


Figure 22.16. (A) The [011] zone-axis diffraction pattern has many planes diffracting with equal strength. In the smaller patterns, the specimen is tilted so there are only two strong beams, the direct 000 on-axis beam and a different one of the $hk\ell$ off-axis diffracted beams. Complementary (B) BF 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.

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 $hk\ell$ reflection moves onto the optic axis. If you do that, you'll find that the $hk\ell$ reflection becomes weaker as you move it onto the axis and the $3h3k3\ell$ 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 26. To set up a *strong-beam* CDF image, tilt in the $\bar{h}\bar{k}\bar{\ell}$ 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.

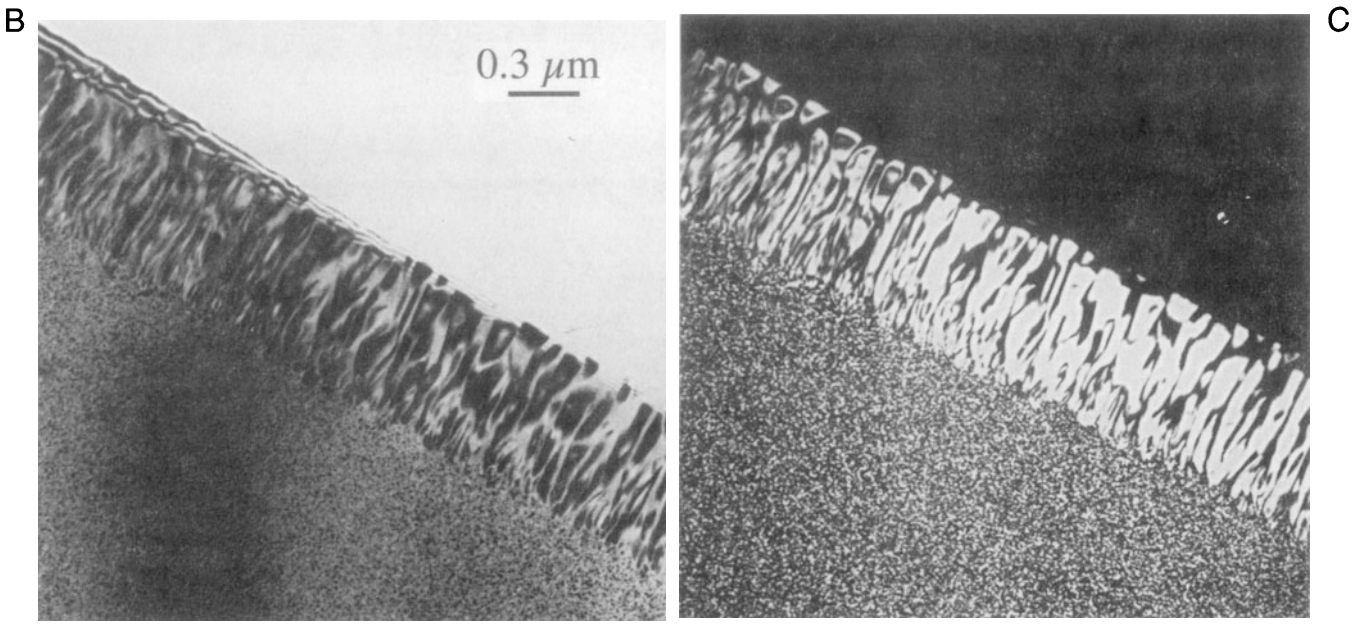


Figure 22.16. *(Continued)*

- Look at the SAD pattern and tilt the specimen until the desired $hk\ell$ reflection is strong. Make sure the incident beam is well underfocused.
- Now tilt the specimen until the $\bar{h}\bar{k}\bar{\ell}$ reflection is strong: $hk\ell$ will now be weak.
- Use the DF tilt controls to move the 000 reflection toward the strong $\bar{h}\bar{k}\bar{\ell}$ reflection. The weak $hk\ell$ reflection will move toward the optic axis and become strong.
- When $hk\ell$ 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 $hk\ell$ and 000 reflections appear in the same position. Make fine adjustments to the DF coils until you can see no shift between 000 and $hk\ell$ 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 $hk\ell$ reflection is too weak (unlikely) or your tilt coils are misaligned (common). In the latter case, realign the coils (see the manufacturer's handbook).

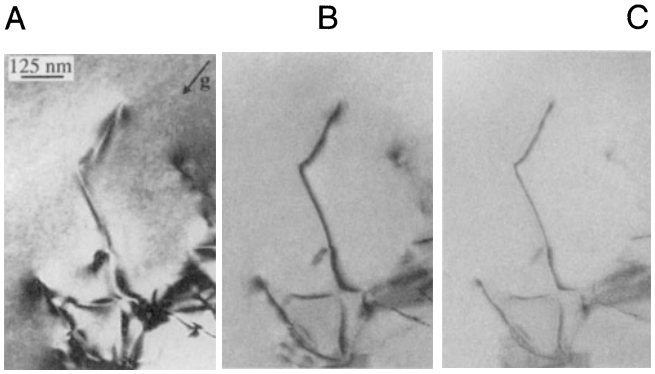


Figure 22.17. Variation in the diffraction contrast when **s** is varied from (A) zero to (B) small and positive and (C) larger and positive.

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.

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. You should usually show the **g** vector in any BF or DF diffraction contrast image after correcting for any rotation between the image and the DP.

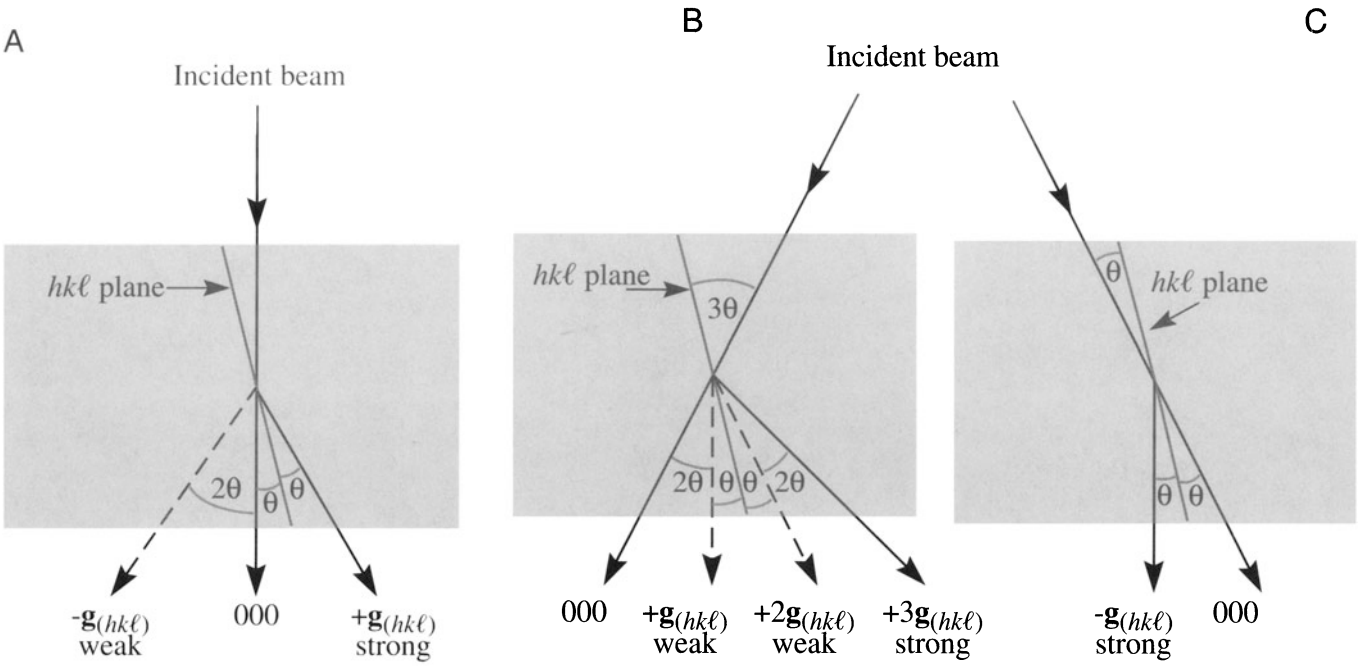


Figure 22.18. (A) Standard two-beam conditions involve the 000 spot and the $hk\ell$ spot bright because one set of $hk\ell$ planes is exactly at the Bragg condition. (B) When the incident beam is tilted through 2θ so that the excited $\mathbf{g}_{hk\ell}$ spot moves onto the optic axis, the $\mathbf{g}_{hk\ell}$ intensity decreases because the $\mathbf{g}_{3h3k3\ell}$ spot becomes strongly excited. (C) To get a strong $\bar{h}\bar{k}\bar{\ell}$ spot on axis for a CDF image, it is necessary to set up a strong $\mathbf{g}_{hk\ell}$ condition first of all, then tilt the initially weak $\mathbf{g}_{\bar{h}\bar{k}\bar{\ell}}$ spot onto the axis.

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 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 $hk\ell$ reflection falls on the BF detector. Either way the CRT 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 semiangles. It's rare in fact that you'll need to do this, but we showed you how to determine the beam-convergence semiangle back in Section 5.5. To calculate the collection semiangle, you need to carry out a similar exercise as we use to determine the EELS spectrometer collection semiangle in Section 37.4.

Remember, there are three conditions that must be fulfilled for strong contrast in your image:

- The incident beam must be parallel, 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 semiangle as α_T and the objective aperture collection semiangle 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

$$\alpha_T = \alpha_S \qquad [22.11a]$$

and

$$\beta_T = \beta_S \qquad [22.11b]$$

Now it should be immediately clear that we can't get such equivalence in a STEM because the convergence angle of

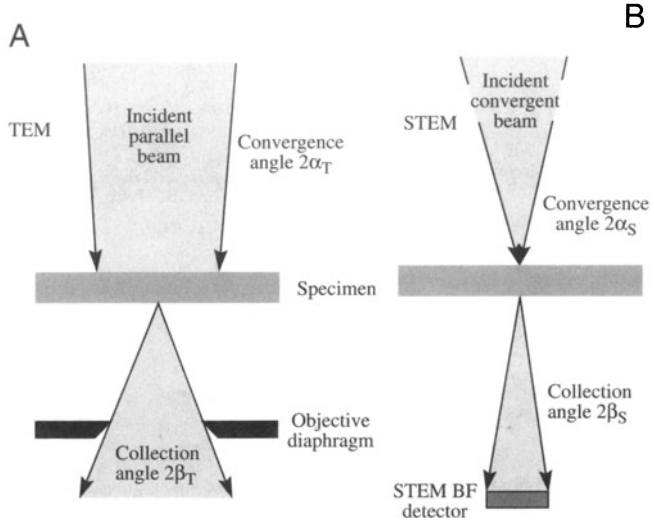


Figure 22.19. Comparison of the important beam-convergence and divergence angles (A) in TEM and (B) in STEM.

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

$$\alpha_S = \beta_T \qquad [22.12a]$$

and

$$\alpha_T = \beta_S \qquad [22.12b]$$

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 semiangle 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 nonparallel (large α_T) destroys the contrast.

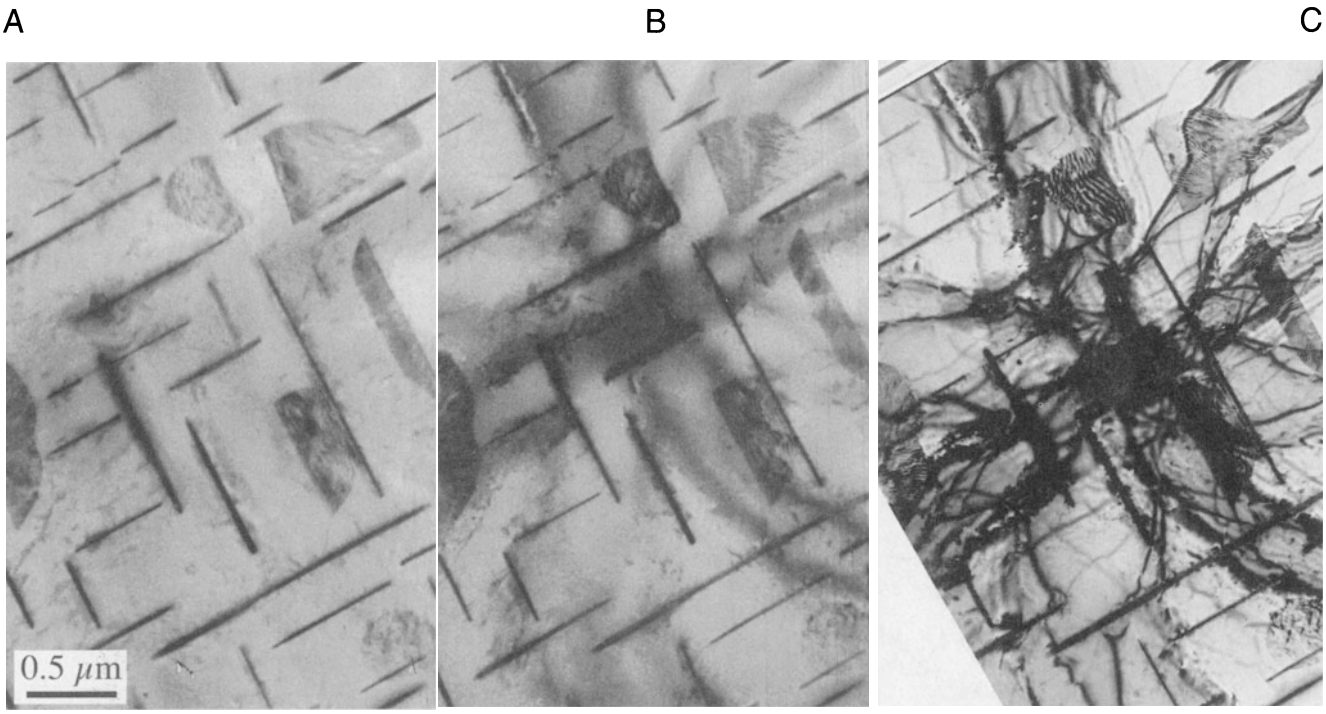


Figure 22.20. (A) BF STEM image of an Al-4 wt.% Cu specimen showing weak diffraction contrast in the form of faint bend contours. As the STEM detector collection angle is lowered (B) the diffraction contrast increases slightly at the expense of increased noise in the image. Even at the smaller collection 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.

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 the next chapter 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 microanalysis, 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 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, and therefore appear dark in BF images and bright in DF images. The contrast can be quantified if necessary.
- TEM mass-thickness contrast images are of 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.

REFERENCES

General References

Grundy, P.J. and Jones, G.A. (1976) *Electron Microscopy in the Study of Materials,* E. Arnold, Ltd., London.

Heidenreich, R.D. (1964) *Fundamentals of Transmission Electron Microscopy,* p. 31, John Wiley & Sons, New York.

Humphreys, C.J. (1979) *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 305, Plenum Press, New York.

Sawyer, L.C. and Grubb, D.T. (1987) *Polymer Microscopy,* Chapman and Hall, New York.

Watt, I.M. (1985) *The Principles and Practice of Electron Microscopy,* Cambridge University Press, New York.

Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd edition, Philips Electron Optics Publishing Group, Mahwah, New Jersey.

Specific References

Cosslett, V.E. (1979) *Phys. stat. sol.* **A55**, 545.

Howie, A. (1979) *J. Microsc.* **177**, 11.

Isaacson, M., Ohtsuki, M., and Utlaut, M. (1979) in *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 343, Plenum Press, New York.

Jesson, D.E. and Pennycook, S.J. (1995) *Proc. Roy. Soc.* (London) **A449**, 273.

Pennycook, S.J. (1992) *Ann. Rev. Mat. Sci.* **22**, 171.

Reimer, L. (1993) *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis,* 3rd edition, p. 194, Springer-Verlag, New York.

Thickness and Bending Effects

23

CHAPTER PREVIEW

We see diffraction contrast in an image of a perfect specimen for two reasons: either the thickness of the specimen varies or the diffraction conditions change across the specimen.

The *thickness* effect: When the thickness of the specimen is not uniform, the coupling 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 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 bend, 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 is rarely constant.
- 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°, since they still have a significant effect on the image contrast. 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.

Thickness and Bending Effects

23

23.1. THE FUNDAMENTAL IDEAS

To understand the origin of thickness fringes and bend contours we limit our 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

$$I_g = \left|\phi_g\right|^2 = \left(\frac{\pi t}{\xi_g}\right)^2 \left(\frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2}\right) = 1 - I_0 \qquad [23.1]$$

where s_{eff} is the effective excitation error

$$s_{\text{eff}} = \sqrt{s^2 + \frac{1}{\xi_g^2}} \qquad [23.2]$$

Although we will concentrate on I_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). 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 23.1 and how you can relate the image to the information contained in the diffraction pattern. 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 unavoidable!

23.2. THICKNESS FRINGES

As a result of the way that we thin TEM specimens, very few of them (only evaporated thin films or ideal ultramicrotomed sections) have a uniform thickness over their entire area. A BF/DF pair of images from the same region of a specimen at 300 kV is shown in Figure 23.1; the thin area is generally in the form of a wedge.

Consider again equation 23.1. 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. 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.

Equation 23.1 tells us that the intensity of both the **0** and the **g** beams oscillates as t varies. Furthermore, these oscillations are complementary for the DF and BF images, as we show schematically in Figure 23.2. 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. In reality, the situation is complicated by the presence of other diffracted beams (we are never truly in a two-beam situation) 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 con-

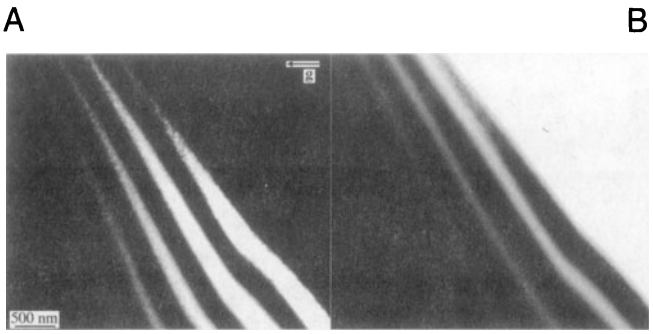


Figure 23.1. (A) DF and (B) BF images from the same region of a wedge-shaped specimen of Si at 300 kV tilted so that **g**(220) is strong. The periodicity and contrast of the fringes is similar and complementary in each image.

tours 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.

> It is important to realize that the image may appear to be black or white depending on the thickness of the specimen.

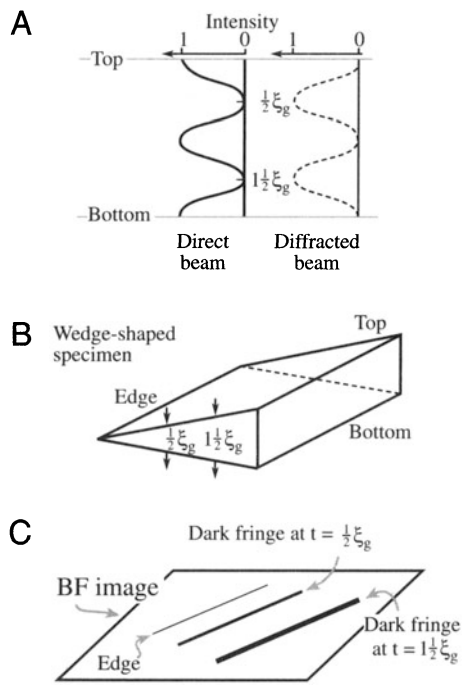


Figure 23.2. (A) At the Bragg condition (**s** = 0), the intensities of the direct and diffracted beams oscillate in a complementary way. (B) For a wedge specimen, the separation of the fringes in the image (C) is determined by the angle of the wedge and the extinction distance, ξ_g.

For example, in BF images, thicker areas are often brighter than thinner areas, which really is counterintuitive.

Several examples of how thickness fringes might appear in your image are shown in Figure 23.3. Although it is often helpful to think of these fringes as thickness con-

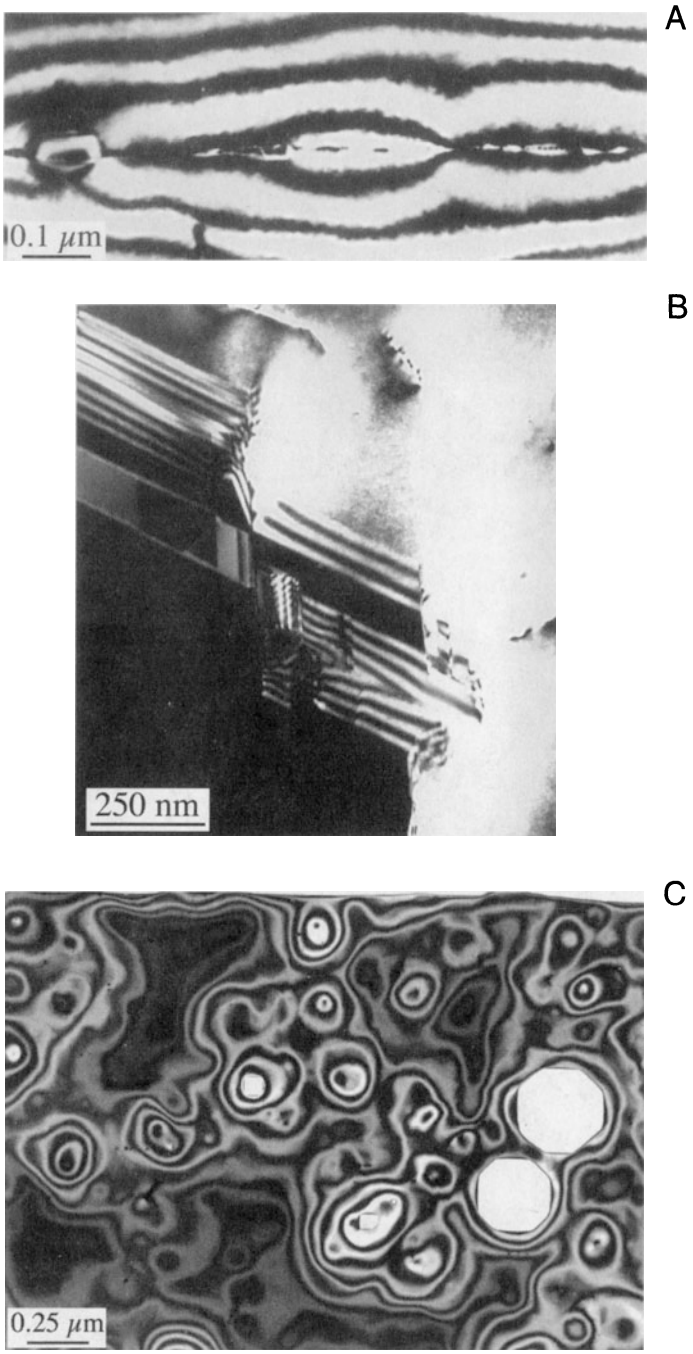


Figure 23.3. Examples of thickness fringes in (A) DF image of a preferentially thinned grain boundary; (B) a strong 220 DF image of microtwinned GaAs taken with only the left-hand grain diffracting, and (C) BF image of a chemically etched thin film of MgO. The white regions in (C) are holes in the specimen.

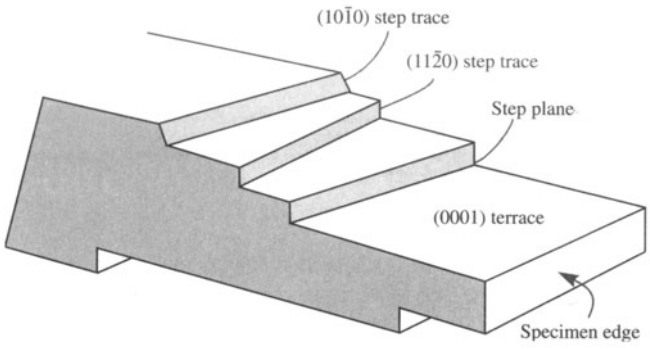


Figure 23.4. Schematic cross-sectional view of a specimen with terraces parallel to the surface and steps connecting terraces.

tours analogous to height or depth contours on a map, with the hole at sea level, remember that there are *two* surfaces to your TEM specimen. A DF image will usually appear to give greater contrast. This is partly because any hole now appears dark, but also because many-beam effects are less important in DF. In Figure 23.3A 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 23.3B, the reflection used to form the 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 23.3C, the specimen is an almost flat, parallel-sided film of MgO with holes formed by chemical etching; 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 23.2 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.

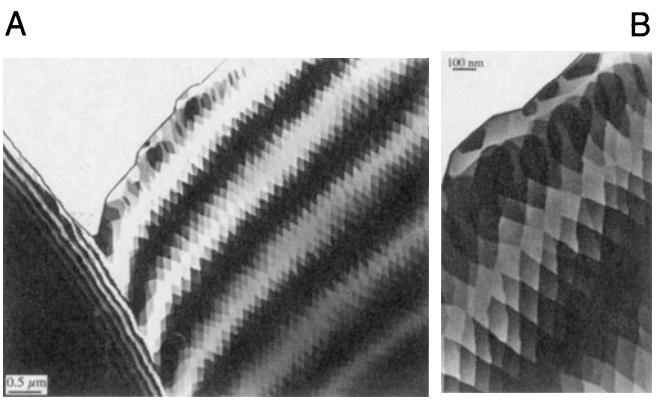


Figure 23.5. Thick fringes from an annealed Al_2O_3 specimen with the geometry shown in Figure 23.4. (A) At low magnification, the fringes are well defined and continuous, even when the wedge angle and wedge axis change. (B) At higher magnification, the contrast is seen to be quantized within a given fringe.

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 23.4. 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 but we never actually inclined the surfaces!

Figure 23.5 shows a striking image of a wedge-shaped specimen of Al_2O_3 which has been heat treated so that the surface has facets 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) but the specimens tend to bend, which obscures these abrupt contrast changes.

23.3. THICKNESS FRINGES AND THE DIFFRACTION PATTERN

A general rule in TEM is that, whenever we see a periodicity in real space (i.e., 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 23.6.

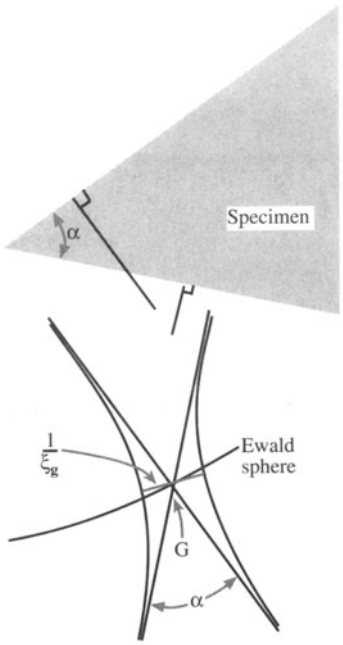


Figure. 23.6. Relrods aligned normal to both surfaces of a wedge-shaped specimen. In practice, the relrods don't cross so there are always two spots in the DP.

The minimum spot spacing in the DP corresponds to the periodicity of the thickness fringes, which at **s** = 0 is given directly by the extinction distance.

This spot spacing is thus related to ξ_g^{-1}. 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 **s** = 0, the Ewald sphere will move up or down (as **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 **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 **s** = 0.

So be wary of trying to make accurate thickness measurements in wedge-shaped crystals.

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, **g**. They actually occur due to interference between two beams, both of which are located close to **g**, so they are really an example of phase contrast although we rarely think of them as such.

23.4. BEND CONTOURS (ANNOYING ARTIFACT, USEFUL TOOL, AND 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 23.7 is aligned so that the $hk\ell$ 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 $hk\ell$ planes are exactly in the Bragg condition at A and the $\bar{h}\bar{k}\bar{\ell}$ 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. Now if we form a BF image we will see two dark lines. Next, we form the DF image using reflection **g**. We see a

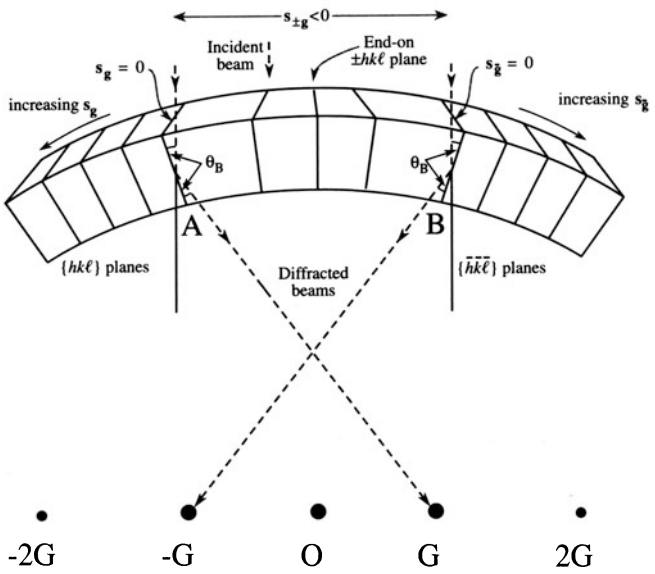


Figure. 23.7. The origin of bend contours shown for a foil symmetrically bent either side of the Bragg conditions. For this geometry, when the $hk\ell$ planes are in the Bragg condition, the reflection G is excited. Notice that G and the diffracting region are on opposite sides of O; if the foil were bent upwards, they would be on the same side.

bright band on the left because that's where **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. Experimental images are shown in Figures 23.8 and 23.9.

> 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 the specimen.

Remembering Bragg's law, the $(2h\ 2k\ 2\ell)$ 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.

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 move as you tilt.

Bend contours are true amplitude contrast, not phase contrast.

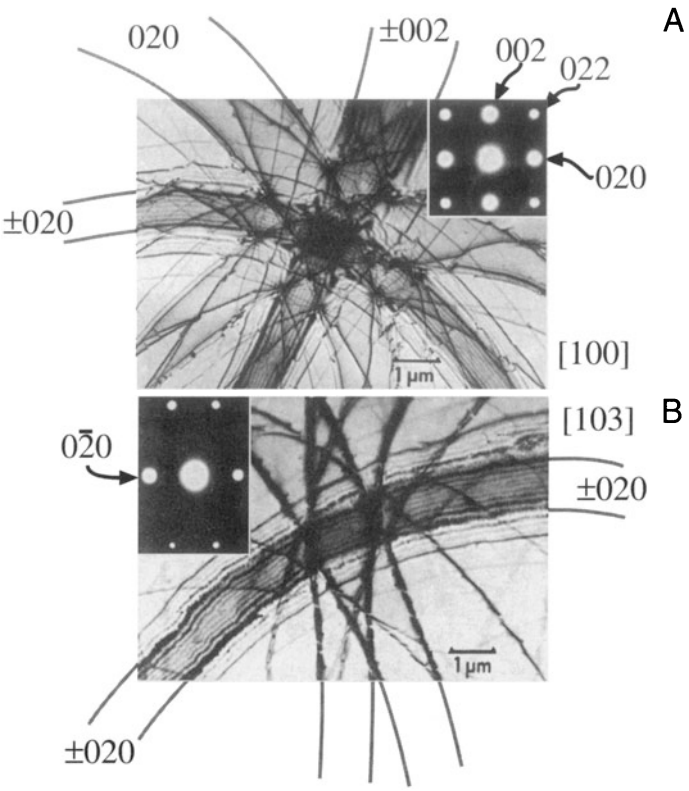


Figure. 23.8. BF images of a bent Al specimen oriented close to the (A) [100] and (B) [103] zone axes. These images are known as (real-space) ZAPs, or zone-axis patterns, and are shown with their respective zone-axis diffraction patterns (insets). Each diffracting plane produces two bend contours, depending 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.

23.5. 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 form a zone-axis pattern, or ZAP. Two examples of these ZAPs are shown in Figure 23.8. 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 (e.g., Rackham and Eades 1977). 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 in large-angle CBED patterns.

In fact, it's the exception that a $\pm\mathbf{g}$ pair of bend contours are straight and parallel. In case you are having a problem visualizing how a pair of contours might diverge, go back to the bent specimen in Figure 23.7, hold the $\bar{h}\bar{k}\bar{\ell}$ 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 $hk\ell$ 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 200 contours in the [100] ZAP are closely spaced, while in the [103] pattern, only one of contours is more closely spaced; the others are more clearly defined and further apart.

> In this case, small **g** in the DP gives a small spacing in the image, contrary to the usual inverse relationship between image and DP.

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 it refers to a DP). You index the contours in the manner described in Section 18.4 but use all the spots in the zone-axis SAD pattern.

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!

23.6. 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.

Figures 23.9A and B show a ZAP in a thin specimen of Al_2O_3 and the associated SAD pattern; the dark bands are $\{3\bar{3}00\}$ bend contours. Figures 23.9C–H show DF images recorded using each of these reflections; the area is identical in each of these images. Using Figures 23.7 you can determine the sense of the bending. You can see directly that the bend contour from one set of $hk\ell$ planes does not necessarily lie parallel to those planes but instead both curves and changes in width.

Remember, for these contrast experiments only, it is important that you move the objective aperture, not the specimen. The resolution of the image will be lower, but this is not critical for this application.

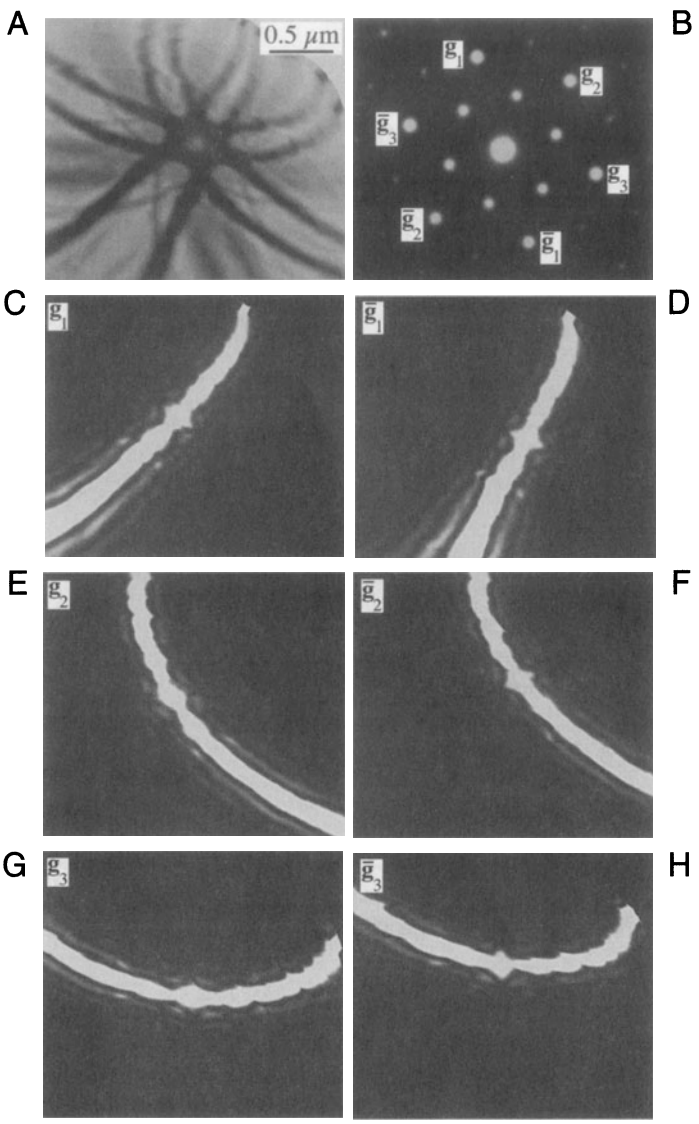


Figure 23.9. An 0001 real space ZAP of Al_2O_3: (A) BF image and (B) corresponding DP. (C–H) Displaced-aperture DF images taken from the spots indicated in (B) identifying the principal dark bend contours in (A); (C,D) $\pm(\bar{3}030)$, (E,F), $\pm(\bar{3}300)$, (G,H) $\pm(03\bar{3}0)$. Note in (A) that the inner $\{11\bar{2}0\}$ spots produce fainter bend contours than the $\{\bar{3}300\}$.

23.7. 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 imply. Much of our thinking about this topic is, however, just as empirical. In fact, it is common to define an imaginary component ξ_g' to the extinction distance, so

$$\xi_g^{abs} = \xi_g \left(\frac{\xi_g'}{\xi_g' + i\xi_g} \right) \qquad [23.3]$$

in the Howie–Whelan equations. The term ξ_g' is found to be approximately $10\xi_g$. The reason for choosing this expression for ξ_g^{abs} is that the $1/\xi_g$ in the Howie–Whelan equation is replaced by $(i/\xi_g' + 1/\xi_g)$. We do the same for ξ_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 IV, 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

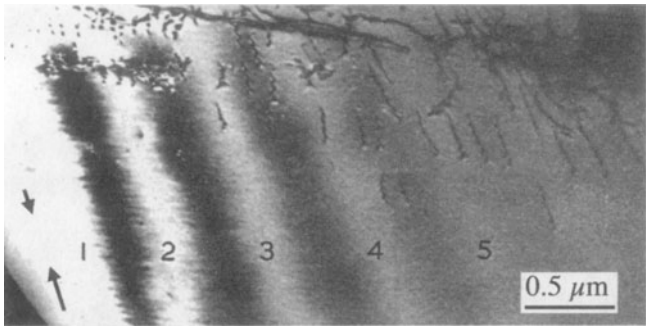


Figure. 23.10. The contrast of thickness fringes in a two-beam BF image decreases when the effect of anomalous absorption is included. Note that the defects are still visible when the fringes have disappeared at a thickness of ~5 ξ_g.

Bloch wave 1. Since thickness fringes result from a beating between the two waves, we will lose the thickness fringes but will still be able to "see through" the specimen, as you can appreciate from Figure 23.10.

Absorption due to the loss of Bloch wave 1 is called anomalous absorption for historical reasons, not because it is unexpected.

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 $\mathbf{s}_g$ is negative the tie line D_1D_2 would be closer to **0** than **g,** and Bloch wave 2 contributes to ϕ_g more strongly. When $\mathbf{s}_g$ 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 $\mathbf{s}_g$ is negative or inside the ±**g** pair of bend contours.

We can summarize this brief description 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 that 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.

23.8. COMPUTER SIMULATION OF THICKNESS FRINGES

The simulation of thickness fringes can be carried out using the Head *et al.* programs or with Comis (see Section 1.5). We'll talk more about these programs in Chapter 25, but be wary: don't use a program as a "black box." Why do we need 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 23.11A. The specimen is a GaAs/Al_xGa_{1-x}As layered composite grown on (001). Since the cleavage surface it {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 nontrivial, especially if you also have to consider the effect shown in Figure 23.6.

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 Figures 23.11B and C which compare the same region of a wedge specimen imaged at 300 kV and 100 kV.

23.9. THICKNESS-FRINGE/ BEND-CONTOUR INTERACTIONS

It's clear from equation 23.1 that both bending and thickness effects can occur together. This combined effect is shown in Figure 23.12, 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} decreases so the thickness contours curve toward the edge of the specimen. This image actually shows the **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 you increase the deviation parameter, **s**. You can use this fact to determine the thickness of an area quite accurately, providing you have a reference value such as zero thick-

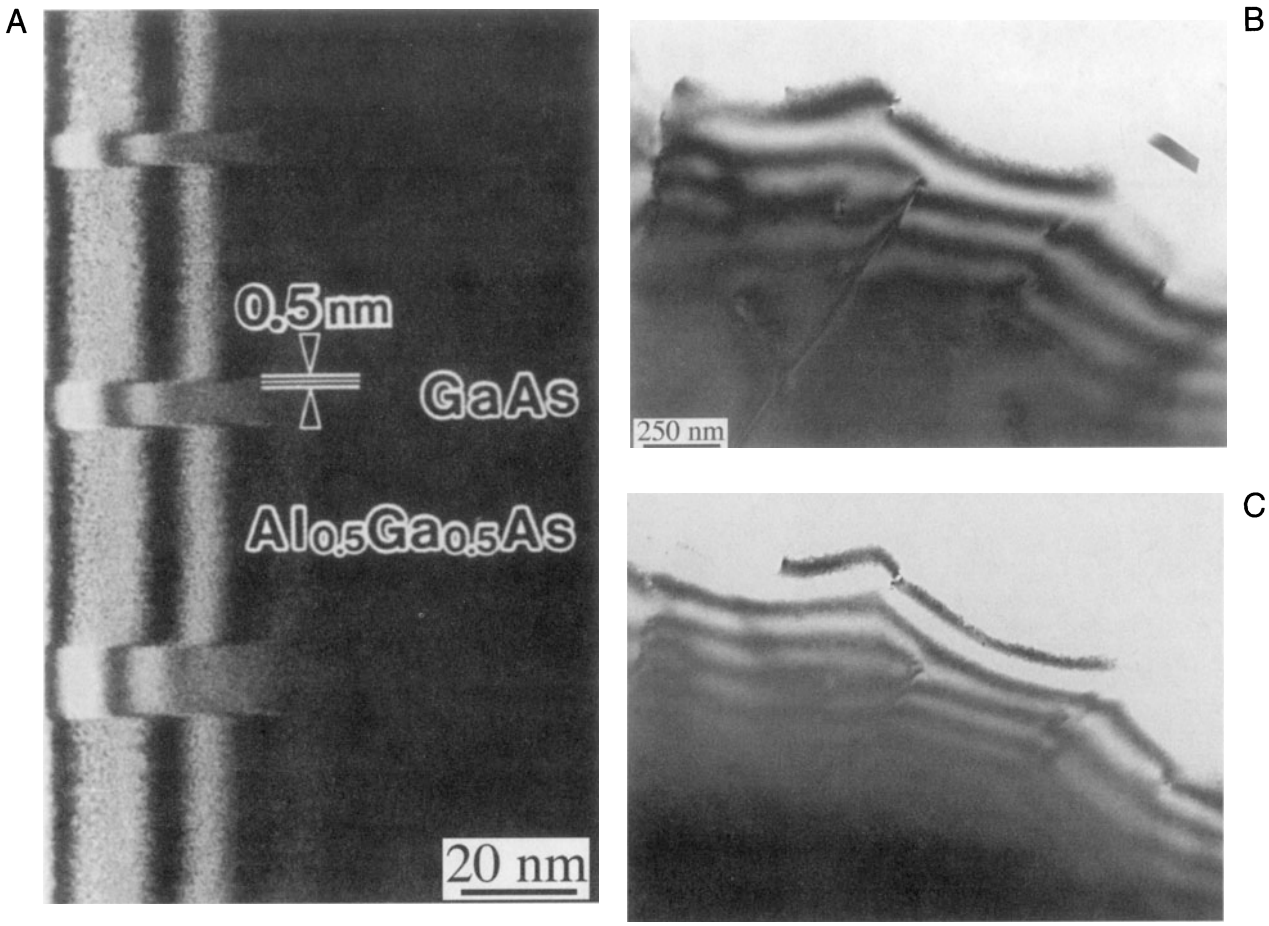


Figure 23.11. (A) Thickness fringes in a 90° wedge of alternating GaAs and AlGaAs. The extinction distance changes in each phase so the fringe spacing changes. Strong beam BF images ($\mathbf{s} = 0$) for (B) 300-kV and (C) 100-kV electrons. The extinction distance increases as the accelerating voltage increases, and you can see through thicker areas; compare with the images in Figure 23.1.

ness 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 you only have 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 ξ_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 26, and where the foil also bends in Section 23.11.

Aside: *You should be careful when using this method for thickness determination 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.*

23.10. 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 by Gibson *et al.* (1985) 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 a (001) (i.e., cube-on-cube) interface plane. One crystal must expand and the other contract normal to this interface. When we prepare a cross-section TEM specimen, we might then imagine it relaxing at the surface as shown in Figure 23.13B. The reason for the relaxation is simply that this allows the one material to expand while the other contracts; the constraint at the surface has been removed during the specimen preparation pro-

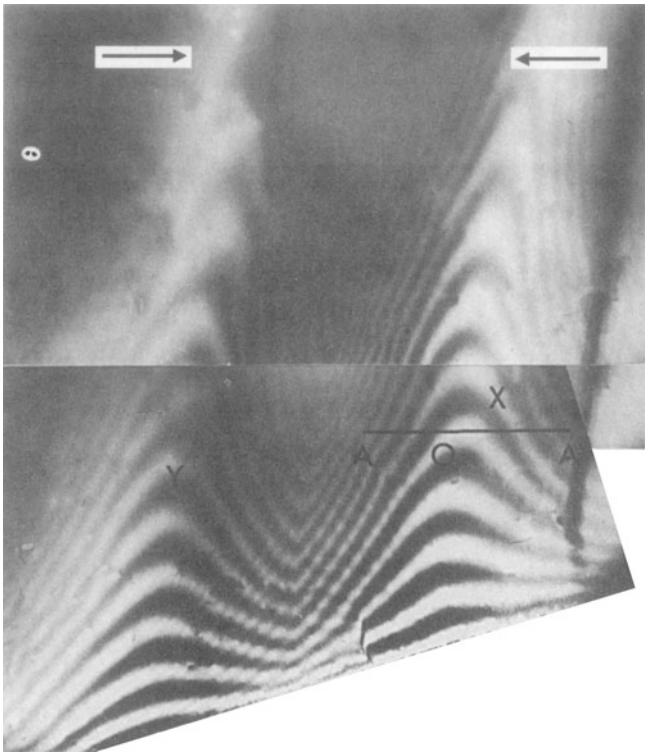


Figure 23.12. Since both thickness fringes and bend contours (X and Y) affect the contrast seen in the image, and both can occur in the same part of the specimen, they can affect, or couple with, one another to give the complex contrast shown in this BF image. Along the line A-A, s changes in sign, being approximately zero at O and negative between the contours.

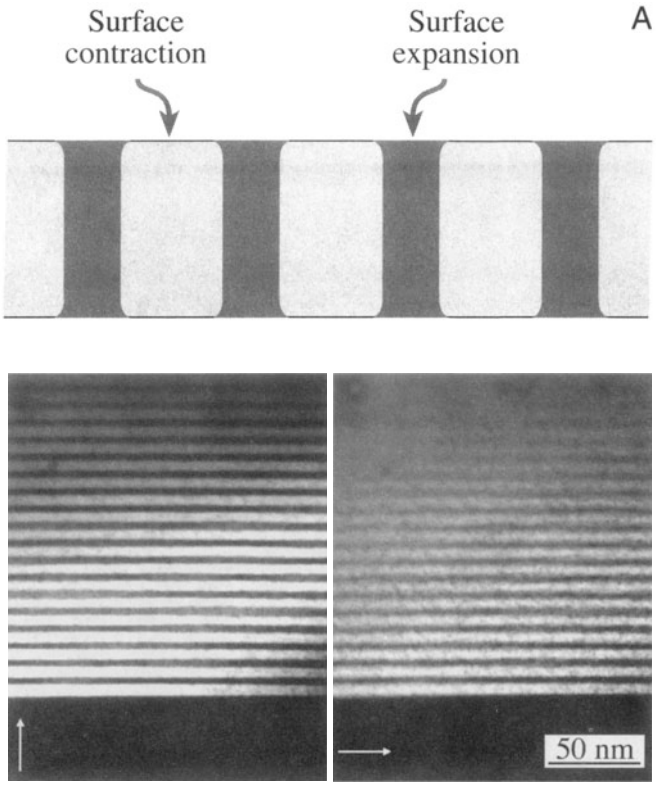


Figure 23.13. (A) A schematic of how interfaces might relax at the surface of a thin specimen. (B,C) DF images of a GaAs/AlGaAs superlattice imaged in two orthogonal reflections, 200 and 020, with the specimen oriented at the 001 pole. (B) The [020] vector is parallel to the interface while (C) the [200] 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.

cess. This argument is admittedly crude, but Figures 23.13B and C show that images recorded with $\mathbf{g} = 020$ normal to this interface appear sharper than images formed when $\mathbf{g} = 200$ is parallel to the interface.

So, no matter whether $\mathbf{g}$ is 020 or $0\bar{2}0$, the Bragg planes are bent closer to $\mathbf{s} = 0$ at one surface or the other.

Here, 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 specimen 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 23.1. 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 $\mathbf{s}$ constant gives thickness fringes.
- Varying $\mathbf{s}$ 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 combine to produce beautiful ZAPs which reflect the true symmetry of the material.

However, if you want to keep defect analysis simple, then you need to avoid bending and work in relatively thin regions of nearly constant thickness. The exception to this rule is that there are quite 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 which 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.

Lastly, be aware that anomalous absorption is not anomalous. It can best be explained by Bloch-wave interactions.

REFERENCES

Specific References

Gibson, J.M., Hull, R., Bean, J.C., and Treacy, M.M.J. (1985) *Appl. Phys. Lett.* **46**, 649.

Rackham, G.M. and Eades, J.A. (1977) *Optik* **47**, 227.

Susnitzky, D.W. and Carter, C.B. (1992) *J. Am. Ceram. Soc.* **75**, 2463. An overview of the surface morphology of heat-treated ceramic thin films studied using TEM.

Planar Defects

24

CHAPTER PREVIEW

Internal interfaces (grain boundaries, phase boundaries, stacking faults) or external interfaces (i.e., surfaces) are perhaps 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. 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 25. 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.

Planar Defects

24

24.1. TRANSLATIONS AND ROTATIONS

An interface is simply a surface which separates any two distinct regions of the microstructure. For most of our discussion, we will assume that the surface is flat and is thus a planar defect. We can sketch a general interface as shown in Figure 24.1.

> 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**.

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, 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.

Now with each of these groups, we can have special examples. We list some of the most common examples in Table 24.1, 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.

For example, 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 ½<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 ½<111> would be a lattice vector and no defect would exist. This particular APB can thus be regarded as an SF. Although we know that {111} is the favored SF plane for fcc metals, SFs in other materials lie on different planes. We will find that the methods used to characterize RBs can often be used to determine **R**(**r**) in other interfaces.

GBs fall into two groups, low-angle and high-angle. Low-angle boundaries necessarily involve a rotation which is usually accommodated by arrays of dislocations; we'll consider these defects in Chapter 25. High-angle boundaries can adopt some special values of **n** and θ such that a large fraction of lattice sites in one grain is shared by the other grain. We characterize the fraction by its inverse, which we call Σ. For example, the common twin boundary in fcc metals is the Σ = 3 grain boundary. 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 coincident-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.

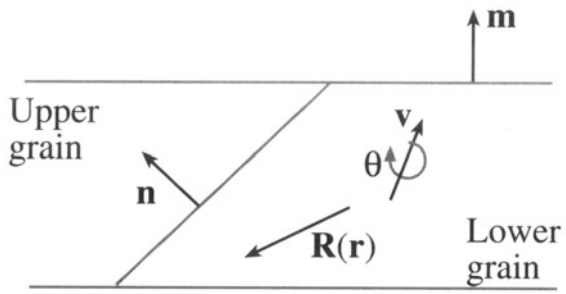


Figure 24.1. A specimen containing a planar defect. The lower grain is translated by a vector **R**(**r**) and rotated through an angle θ about the vector **v**, relative to the upper grain. The defect plane is **n**, the foil normal is **m**.

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). These special interfaces can often be imaged because of the 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 on the other; the $(hk\ell)$ plane on one side is parallel to the $\bar{h}\bar{k}\bar{\ell}$ on the other. We can't distinguish **g** from $\bar{\mathbf{g}}$ unless we use CBED (Taftø and Spence 1982).

PBs are rarely analyzed fully. 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 two surface-sensitive techniques are plan-view and reflection electron microscopy (REM; see Chapter 31).

24.2. 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 Chapter 25. 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 which 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.

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.

Table 24.1. Examples of Internal Planar Defects

Group	Structure	Example	Example
SF	Diamond-cubic, fcc, zinc blende	Cu, Ag, Si, GaAs	**R** = ⅓ [111] or **R** = ⅙ [11$\bar{2}$]
APB/IDB	Zinc blende, wurtzite	GaAs, AlN	inversion
APB	CsCl	NiAl	**R** = ½ [111]
APB/SF	Spinel	$MgAl_2O_4$	**R** = ¼ [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

$$\mathbf{r}'_n = \mathbf{r}_n + \mathbf{R}_n \qquad [24.1]$$

In this expression, $\mathbf{R}_n$ is actually $\mathbf{R}_n(\mathbf{r})$; it can vary throughout the specimen. The term $e^{2\pi i\mathbf{K}\cdot\mathbf{r}}$ in equation 13.3 now becomes $e^{2\pi i\mathbf{K}\cdot\mathbf{r}'}$ so we need to examine the term $\mathbf{K}\cdot\mathbf{r}'$. We know that $\mathbf{K}$ is $\mathbf{g} + \mathbf{s}$, so we can write

$$\begin{aligned}\mathbf{K}\cdot\mathbf{r}'_n &= (\mathbf{g}+\mathbf{s})\cdot(\mathbf{r}_n+\mathbf{R}_n)\\ &= \mathbf{g}\cdot\mathbf{r}_n + \mathbf{g}\cdot\mathbf{R}_n + \mathbf{s}\cdot\mathbf{r}_n + \mathbf{s}\cdot\mathbf{R}_n\end{aligned} \qquad [24.2]$$

Now, since $\mathbf{r}_n$ is a lattice vector, $\mathbf{g}\cdot\mathbf{r}_n$ is an integer as usual. The third term, $\mathbf{s}\cdot\mathbf{r}_n$, gives our usual sz term, so the new terms are $\mathbf{g}\cdot\mathbf{R}_n$ and $\mathbf{s}\cdot\mathbf{R}_n$.

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 $\mathbf{s}\cdot\mathbf{R}_n$. 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 26.
- 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 24.2

$$\frac{d\phi_{\mathbf{g}}}{dz} = \frac{\pi i}{\xi_0}\phi_{\mathbf{g}} + \frac{\pi i}{\xi_{\mathbf{g}}}\phi_0 \exp\left[-2\pi i(sz + \mathbf{g}\cdot\mathbf{R})\right] \qquad [24.3]$$

and

$$\frac{d\phi_0}{dz} = \frac{\pi i}{\xi_0}\phi_0 + \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{\mathbf{g}} \exp\left[+2\pi i(sz + \mathbf{g}\cdot\mathbf{R})\right] \qquad [24.4]$$

Next, we simplify these equations just as we did in Chapter 13 by setting

$$\phi_0(z)_{(\text{sub})} = \phi_0 \exp\left(\frac{-\pi i z}{\xi_0}\right) \qquad [24.5]$$

and

$$\phi_{\mathbf{g}}(z)_{(\text{sub})} = \phi_{\mathbf{g}} \exp\left(2\pi i s z - \frac{\pi i z}{\xi_0}\right) \qquad [24.6]$$

Then the Howie–Whelan equations become

$$\frac{d\phi_{0(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{\mathbf{g}(\text{sub})} \exp\left(2\pi i \mathbf{g}\cdot\mathbf{R}\right) \qquad [24.7]$$

and

$$\frac{d\phi_{\mathbf{g}(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{0(\text{sub})} \exp\left(-2\pi i \mathbf{g}\cdot\mathbf{R}\right) + 2\pi i s \phi_{\mathbf{g}(\text{sub})} \qquad [24.8]$$

These equations are just as before (equations 13.14 and 13.15) but with the addition of the $2\pi i\mathbf{g}\cdot\mathbf{R}$ term. This additional phase is termed α, hence planar defects are seen when $\alpha \neq 0$

$$\alpha = 2\pi \mathbf{g}\cdot\mathbf{R} \qquad [24.9]$$

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 24.2. 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 translation $\mathbf{R}$, but after that they again propagate as if in a 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.

24.3. THE SCATTERING MATRIX

This discussion of the scattering matrix introduces no new concepts. It is just a different way of writing the equations

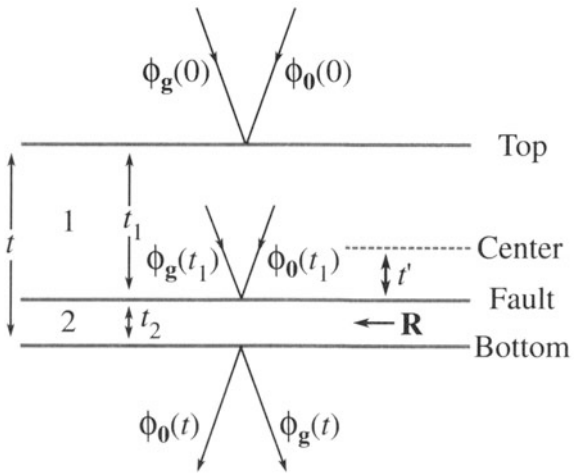


Figure 24.2. A stacking fault lying at depth t_1 in a parallel-sided uniformly thick specimen. The total thickness is t and $t_2 = t - t_1$.

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 equation 13.20, we showed that in the two-beam case, we can write these simple expressions for ϕ_0 and ϕ_g

$$\phi_0 = C_0 e^{2\pi i\gamma z} \qquad [24.10]$$

and

$$\phi_g = C_g e^{2\pi i\gamma z} \qquad [24.11]$$

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

$$\phi_0(z) = C_0^{(1)}\psi^{(1)}\exp\left(2\pi i\gamma^{(1)}z\right) + C_0^{(2)}\psi^{(2)}\exp\left(2\pi i\gamma^{(2)}z\right) \qquad [24.12]$$

$$\phi_g(z) = C_g^{(1)}\psi^{(1)}\exp\left(2\pi i\gamma^{(1)}z\right) + C_g^{(2)}\psi^{(2)}\exp\left(2\pi i\gamma^{(2)}z\right) \qquad [24.13]$$

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 24.12 and 24.13 in a matrix form

$$\begin{pmatrix}\phi_0(z)\\ \phi_g(z)\end{pmatrix} = \begin{pmatrix}C_0^{(1)} & C_0^{(2)}\\ C_g^{(1)} & C_g^{(2)}\end{pmatrix}\begin{pmatrix}\exp\left(2\pi i\gamma^{(1)}z\right) & 0\\ 0 & \exp\left(2\pi i\gamma^{(2)}z\right)\end{pmatrix}\begin{pmatrix}\psi^{(1)}\\ \psi^{(2)}\end{pmatrix} \qquad [24.14]$$

We can express our boundary conditions as

$$C_0^{(1)}\psi^{(1)} + C_0^{(2)}\psi^{(2)} = \phi_0(0) \qquad [24.15]$$

and

$$C_g^{(1)}\psi^{(1)} + C_g^{(2)}\psi^{(2)} = \phi_g(0) \qquad [24.16]$$

which we can now rewrite as

$$\begin{pmatrix}C_0^{(1)} & C_0^{(2)}\\ C_g^{(1)} & C_g^{(2)}\end{pmatrix}\begin{pmatrix}\psi^{(1)}\\ \psi^{(2)}\end{pmatrix} = \begin{pmatrix}\phi_0(0)\\ \phi_g(0)\end{pmatrix} \qquad [24.17]$$

(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.) Now we can use matrix algebra to solve equation 24.17. First rewrite it as

$$C\begin{pmatrix}\psi^{(1)}\\ \psi^{(2)}\end{pmatrix} = \begin{pmatrix}\phi_0(0)\\ \phi_g(0)\end{pmatrix} \qquad [24.18]$$

then rewrite equation 24.18 as

$$\begin{pmatrix}\psi^{(1)}\\ \psi^{(2)}\end{pmatrix} = C^{-1}\begin{pmatrix}\phi_0(0)\\ \phi_g(0)\end{pmatrix} \qquad [24.19]$$

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 24.14 as

$$\begin{pmatrix}\phi_0(z)\\ \phi_g(z)\end{pmatrix} = C\begin{pmatrix}\exp\left(2\pi i\gamma^{(1)}z\right) & 0\\ 0 & \exp\left(2\pi i\gamma^{(2)}z\right)\end{pmatrix}C^{-1}\begin{pmatrix}\phi_0(0)\\ \phi_g(0)\end{pmatrix} \qquad [24.20]$$

Finally, we can define a new matrix $P(z)$ as the scattering matrix for a slice of thickness z

$$P(z) = C\begin{pmatrix}\exp\left(2\pi i\gamma^{(1)}z\right) & 0\\ 0 & \exp\left(2\pi i\gamma^{(2)}z\right)\end{pmatrix}C^{-1} = C\,\Gamma\,C^{-1} \qquad [24.21]$$

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. Notice that z only enters the equation through the Γ matrix in equation 24.21.

24.4. 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 24.2. The idea is that we now have two slices of material of thickness t_1 and t_2. We can easily calculate $\phi_0(t_1)$ and $\phi_g(t_1)$ using equation 24.20. These values for ϕ_0 and ϕ_g then become the incident values for slice 2. The effect of the translation **R** is to multiply the terms in C_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

$$C_2 = \begin{pmatrix}C_0^{(1)} & C_0^{(2)}\\ C_g^{(1)}\exp(-i\alpha) & C_g^{(2)}\exp(-i\alpha)\end{pmatrix} \qquad [24.22]$$

We can write down the expression for $\phi_0(t)$ and $\phi_g(t)$ as

$$\begin{pmatrix} \phi_0(t) \\ \phi_g(t) \end{pmatrix} = C_2 P(t_2) C_2^{-1} C_1 P(t_1) C_1^{-1} \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \quad [24.23]$$

where the subscripts on C_1 and C_2 just identify the slices. Normally, this equation goes straight into the computer. However, we'll go back and consider a few special points:

- Look at equation 24.23 and set $\mathbf{R} = 0$, so that $C_1 = C_2$. 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.
- How do we prove equation 24.22? From equation 14.12 we know that a Bloch wave can be written as

$$b(\mathbf{k}) = \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k}) \exp\left(2\pi i(\mathbf{k}+\mathbf{g})\cdot\mathbf{r}\right) \quad [24.24]$$

If the crystal is displaced by a vector $\mathbf{R}$ then we replace $\mathbf{r}$ with $\mathbf{r} - \mathbf{R}$ (notice the sign, see Section 25.14). (We have just used a "hidden" column approximation.) Equation 24.24 is then written

$$b(\mathbf{k}) = \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k}) \exp\left(2\pi i(\mathbf{k}+\mathbf{g})\cdot(\mathbf{r}-\mathbf{R})\right) \quad [24.25]$$

$$b(\mathbf{k}) = e^{-2\pi i\mathbf{k}\cdot\mathbf{R}} \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k})\, e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\, e^{2\pi i(\mathbf{k}+\mathbf{g})\cdot\mathbf{r}} \quad [24.26]$$

C_0 is not affected by $\mathbf{R}$ since $2\pi\mathbf{0}\cdot\mathbf{r} = 0$, but C_g is multiplied by $e^{-i\alpha}$.

- If you choose the coordinates appropriately (see Chapter 14) 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 24.23 explicitly, as given by Hirsch *et al.* (1977) (omitting a phase factor)

$$\begin{aligned} &\varphi_0(t) = \\ &[\cos(\pi\Delta kt) - i\cos(\beta)\sin(\pi\Delta kt)] \\ &+ \tfrac{1}{2}(e^{i\alpha}-1)\sin^2\beta\cos(\pi\Delta kt) - \tfrac{1}{2}(e^{i\alpha}-1)\sin^2\beta\cos(2\pi\Delta kt') \end{aligned} \quad [24.27]$$

$$\begin{aligned} &\varphi_g(t) = i\sin\beta\sin(\pi\Delta kt) \\ &+ \tfrac{1}{2}\sin\beta\left(1-e^{(-i\alpha)}\right)\left[\cos\beta\cos(\pi\Delta kt) - i\sin(\pi\Delta kt)\right] \\ &- \tfrac{1}{2}\sin\beta\left(1-e^{(-i\alpha)}\right)\left[\cos\beta\cos(2\pi\Delta kt') - i\sin(2\pi\Delta kt')\right] \end{aligned} \quad [24.28]$$

In equations 24.27 and 24.28, 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 24.27 and 24.28 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.

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 platelike 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.

24.5. 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 23.

- Planar defects can, in fact, have a thickness. We'll illustrate this concept using overlapping faults in fcc materials (see also Section 6.6).

Do not assume all faults are the same as in fcc materials!

24.5.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 $\frac{1}{6}\langle\bar{1}\bar{1}2\rangle$ or $\frac{1}{3}\langle111\rangle$. Notice that these definitions differ by the lattice vector, $\frac{1}{2}\langle110\rangle$. (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 is shown in Figure 24.3. If the sample is 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.

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 $\mathbf{R} = \pm\frac{1}{3}[11\bar{1}]$; the phase factor, α, is $2\pi\,\mathbf{g}\cdot\mathbf{R}$. If you form an image with the $\mathbf{g} = (2\bar{2}0)$ reflection strongly excited, then $\mathbf{g}\cdot\mathbf{R} = 0$ and the fault is out of contrast in both BF and DF. If, instead, you use the reflection $\mathbf{g} = (02\bar{2})$, then $\mathbf{g}\cdot\mathbf{R} = \frac{1}{3}$ or $-\frac{1}{3}$ and $\alpha = \frac{8\pi}{3} = \frac{2\pi}{3} = 120°$ or $-\frac{8\pi}{3} = -\frac{2\pi}{3} = -120°$ (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., $\mathbf{g}\cdot\mathbf{R} = 0$ for all values of **g** lying in the fault plane.

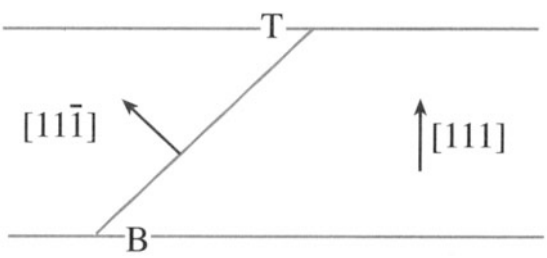


Figure 24.3. A stacking fault in a parallel-sided fcc specimen. The normal to the specimen is [111] and the normal to the SF is [11$\bar{1}$]. T and B indicate the top and bottom of the foil.

Figures 24.4A–D show 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, as summarized in Figures 24.4E and F. The questions which arise are:

- What determines whether a fringe will be gray or white?
- Why are the two images not complementary?

Note in Figures 24.4E and F that Type A reflections are 200, 222, and 440 while Type B are 111, 220, and 400.

24.5.B. Some Rules

There are some experimental rules:

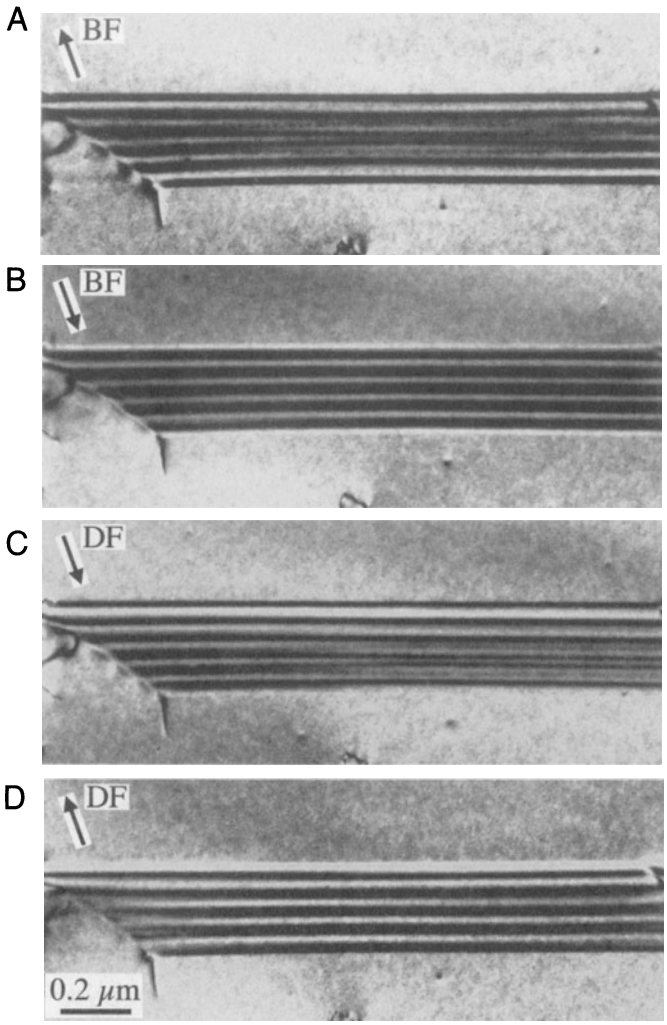
- Be very careful when you record such a pair of images: record the DP for each image. Be sure to note which of the two bright spots corresponds to the direct beam.
- Use the same strong $hk\ell$ reflection for BF and DF imaging. Therefore, to form the CDF image using a strong $hk\ell$ reflection you must first tilt the specimen so that $\bar{h}\bar{k}\bar{\ell}$ is strong, and then use the beam tilts to move $hk\ell$ 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 image, rather than a CDF image.

In a modern IVEM, there is almost no loss of resolution between displaced-aperture DF and CDF.

So to avoid the possibility of confusion about which reflection to use for the DF image (e.g., see Edington 1976, p. 148), just displace the aperture to the strong $hk\ell$ reflection in every case.

This is exactly the opposite of the approach used by Edington, who advocates tilting in $\bar{h}\bar{k}\bar{\ell}$ for the DF image, which reverses the DF contrast in Figures 24.4E and F. Our approach ensures that diffraction from the same $hk\ell$ planes causes the contrast in both BF and DF images.

Then there are some rules for interpreting the contrast:



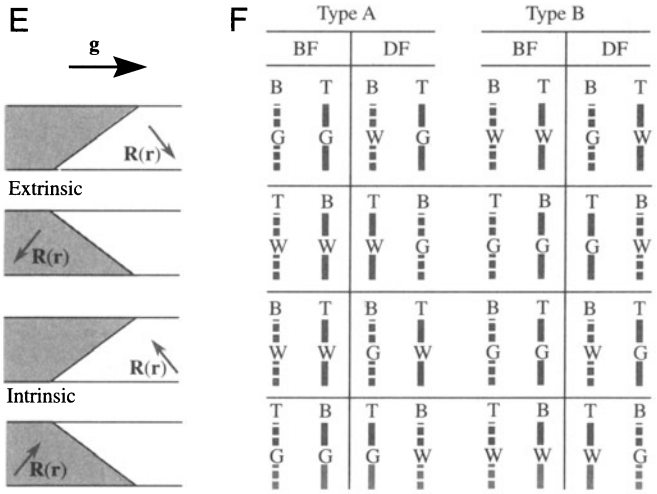


Figure 24.4. (A–D) Four strong-beam images of an SF recorded using ±**g** BF and ±**g** DF. The beam was nearly normal to the surfaces; the SF fringe intensity is similar at the top surface but complementary at the bottom surface. The rules are summarized in (E) and (F) where G and W indicate that the first fringe is gray or white, and (T,B) indicates top/bottom.

- 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 $\mathbf{g}\cdot\mathbf{R}>0$ and black if $\mathbf{g}\cdot\mathbf{R}<0$.
- Using the same strong $hk\ell$ 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 24.10.
- The reason it is important to know the sign of **g** is that you will use this information to determine the sign of **R**.
- For the geometry shown in Figure 24.3, if the origin of the **g**-vector is placed at the center of the SF in the DF image, the vector **g** points away from the bright outer fringe if the fault is extrinsic and toward it if it is intrinsic (200, 222, and 440 reflection); 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 any intensity change due to $\mathbf{g}\cdot\mathbf{R}$ effects is only detectable if $\mathbf{g}\cdot\mathbf{R}>0.02$ (unless you can digitize and process your analog TEM image). 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 the background intensity in the matrix around the defect is gray and this maximizes visibility of lighter and darker fringes.

So just because you can't see a defect doesn't mean it isn't there, or that $\mathbf{g}\cdot\mathbf{R} = 0$.

As we said, these complex rules are summarized in Figures 24.4E and F. 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 $\mathbf{g}\cdot\mathbf{R}$ are given in Table 24.2. As we'll describe in Section 24.11, you should use a computer program to check the contrast.

24.5.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 24.22 and 24.23 that

Table 24.2. Values of g·R for Some Common R/g Combinations

	R	**g**	$\alpha = 2\pi\,\mathbf{g}\cdot\mathbf{R}$ (mod 2π)
SF in fcc	$\frac{1}{3}[111]$	$(11\bar{1})$, (220), $(\bar{1}\bar{1}3)$	$2\pi/3$
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 ξ_g differ slightly	δ

$$\phi_g = \frac{i\pi}{\xi_g}\left\{\int_0^{t_1} e^{-2\pi isz}dz + e^{-i\alpha}\int_{t_1}^{t} e^{-2\pi isz}dz\right\} \qquad [24.29]$$

which gives

$$\phi_g = \frac{i\pi}{s\xi_g} e^{-2\pi ist_1}\{\sin(\pi st_1) + e^{-i\alpha}\sin(\pi s(t-t_1))\} \qquad [24.30]$$

We rearrange equation 24.30 to give an expression for the intensity, I_g (= $\phi_g\phi_g^*$). This rearrangement involves a little manipulation

$$I_g = \frac{1}{(s\xi_g)^2}\left\{\sin^2\left(\pi st_1 + \frac{\alpha}{2}\right) + \sin^2\left(\frac{\alpha}{2}\right) - \sin\left(\frac{\alpha}{2}\right)\sin\left(\pi st + \frac{\alpha}{2}\right)\cos(2\pi st')\right\} \qquad [24.31]$$

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 24.31 becomes

$$I_g \propto \frac{1}{s^2}\{A - B\cos(2\pi st')\} \qquad [24.32]$$

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 24.28 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.

In Chapter 26 we will discuss this SF contrast in terms of phasor diagrams, which give a graphical way to represent these equations.

24.5.D. Overlapping Faults

It is interesting to extend this analysis to the case of overlapping faults. Taking the analytical approach, we can extend equation 24.29 to the case of two overlapping faults, the first at depth t_1 and the second at depth t_2

$$\phi_g = \frac{i\pi}{\xi_g}\left\{\int_0^{t_1} e^{-2\pi isz}dz + e^{-i\alpha}\int_{t_1}^{t_1+t_2} e^{-2\pi isz}dz + e^{-i(\alpha_1+\alpha_2)}\int_{t_1+t_2}^{t} e^{-2\pi isz}dz\right\} \qquad [24.33]$$

An experimental illustration of a somewhat more complex situation, involving several overlapping SFs, is shown in

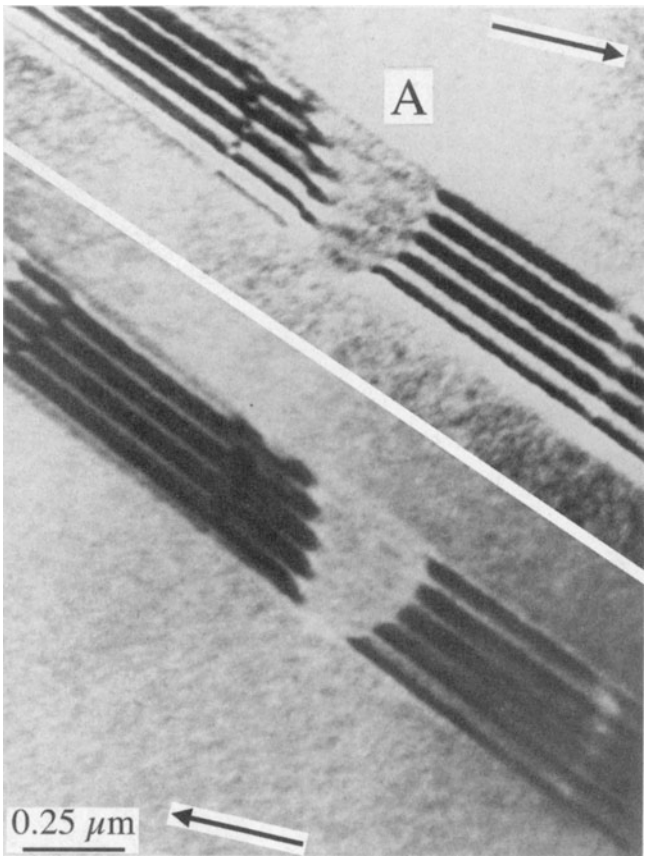


Figure 24.5. Two ±**g** BF images of overlapping SFs in fcc steel, with the direction of **g** indicated. The faults are very close together. When three faults overlap the effective value of **R** is 0 so the contrast disappears.

Figure 24.5. We can see two very interesting features in this image:

- The BF/DF contrast is not complementary.
- 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 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 26.6 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 thickness. We can then analyze the contrast from such interfaces using the overlapping-fault model.

24.6. 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.

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 nonequivalent APBs in this structure.) 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 24.6. These π fringes can give symmetric fringes in DF and BF and complementary BF/DF pairs.

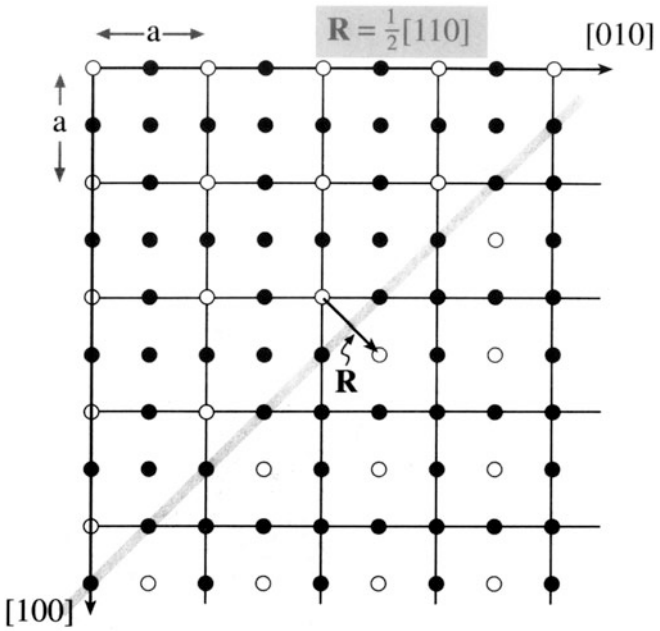


Figure 24.6. Schematic of an interface in the intermetallic Ni_3Al showing how the two structures link coherently. The phase factor at such an interface is π and the fringes seen in the image are called π fringes.

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 Figures 24.7A and B has been called both an SF and an APB in the spinel. These interfaces can show all the features we discussed in Section 24.5 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 24.7C when APBs in TiO_2 overlap as shown schematically in Figure 24.7D (Amelinckx and Van Landuyt, 1978). In Figures 24.7A,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 24.8 is different yet again. This planar defect in GaAs is also known as an IDB (Section 24.1). The fringes you see are caused by a translation, but **R** is not related in a simple way to the structure of the crystal (Rasmussen *et al.* 1991). As shown in the diagram, 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 0.19Å with a statistical uncertainty of ±0.03Å. These fringes are then known as δ fringes (because they are only small transla-

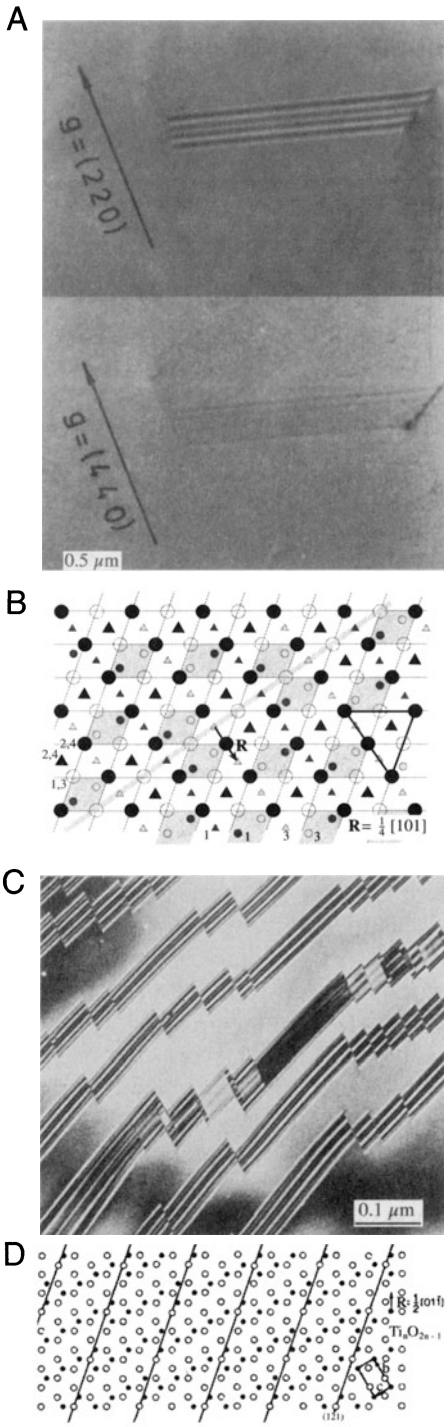


Figure 24.7. (A) Pair of BF images and (B) schematic of an SF in spinel; this interface is also known as an APB since the SF translation vector is a perfect (sub)lattice vector in the underlying fcc oxygen sublattice as shown in (B). Large circles are O anions at different heights (1,3) and (2,4); the small circles and triangles are cations at different heights (1–4 as indicated). (C) 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 (D) shows a series of APBs, each of which is formed by a translation which has little effect on the oxygen sublattice.

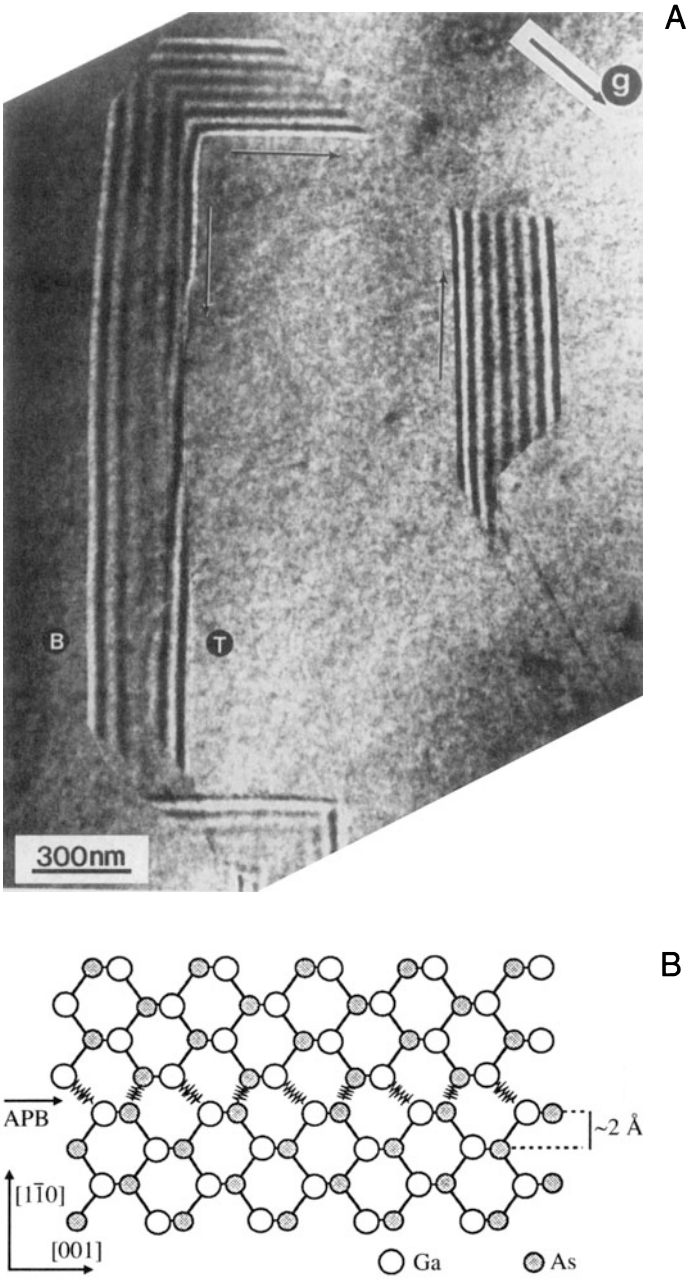


Figure 24.8. (A) A faceted APB (or IDB) in GaAs with (B) a schematic of the (1$\bar{1}$0) facet. The translation is caused by the difference in length of the Ga-Ga and As-As bonds and does not correspond to any length in the GaAs lattice.

tions and are 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 0.025Å), is discussed in Section 24.13.

Table 24.3. Examples of Special Phase Boundaries

Boundary	Example of material	Features
Ferromagnetic domain boundaries	NiO	
Ferroelectric/piezoelectric boundaries	$BaTiO_3$	Small tetragonal distortion
Composition boundary	GaAs/AlGaAs	ξ_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	

24.7. PHASE BOUNDARIES

We list a few special phase boundaries in Table 24.3.

An example of a PB is shown in Figure 24.9. In NiO, which is ferromagnetic, some of the planes rotate when the structure changes from cubic symmetry below the Curie temperature. Now we can also define the cubic structure as rhombohedral with $\alpha = 60°$ in the rhombohedron. Below the Curie temperature, the rhombohedral angle is distorted by only 4.2′ 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 24.9, 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.

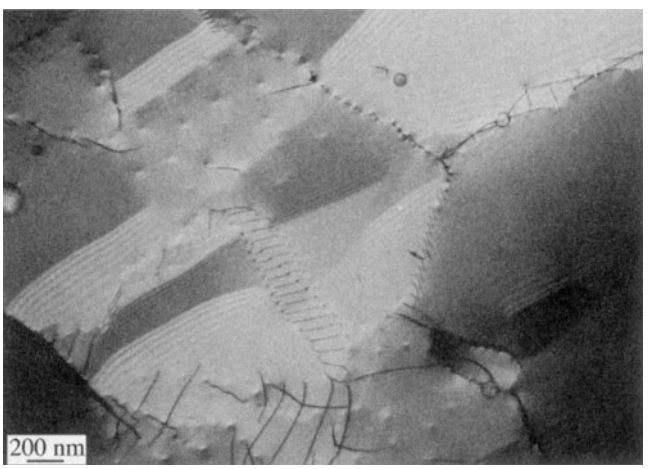


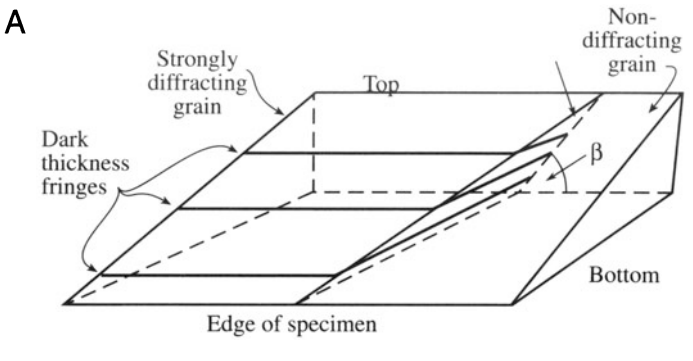
Figure 24.9. The ferromagnetic material NiO undergoes a structural change from cubic to distorted rhombohedral at the Curie temperature. Although the distortion in the rhombohedral structure is very small, it causes a detectable rotation of the lattice planes which results in the δ fringes in the image.

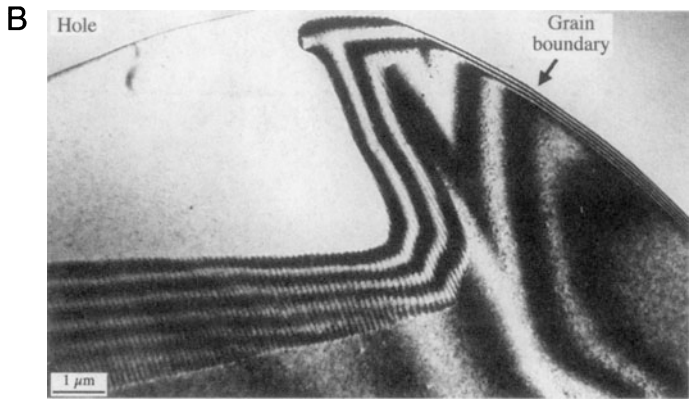
24.8. ROTATION BOUNDARIES

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 interfaces are shown in Figure 24.10.

24.9. 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 in a parallel-sided specimen will produce at least two spots in the diffraction pattern. 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 even more complex. Figure 24.11 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 diffraction pattern. 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 periodicities of the fringes in the image are inversely related to the distances (M_1N, and M_2N) between the spots in the DP.





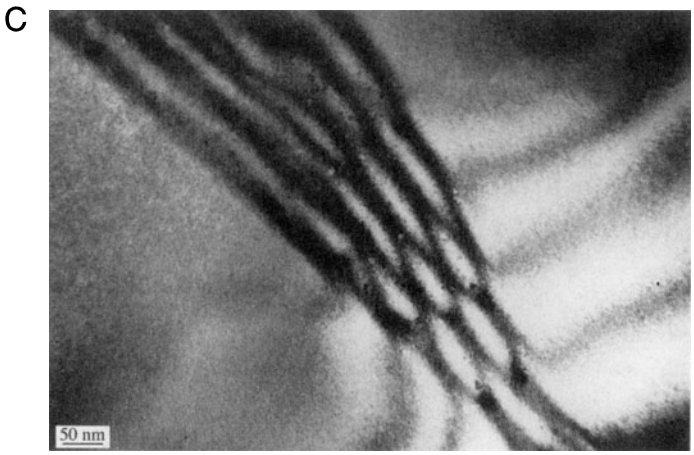


Figure 24.10. (A) If the adjoining grains are rotated so that they do not share a common reflection, images can be formed where only one of the grains diffracts. As shown in (B), the thickness fringes associated with the wedge-shaped foil merge into the thickness fringes associated with the inclined 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.

At **s** = 0, the fringe spacing is determined by $\Delta k = \xi_g^{-1}$; the spacing of the fringes is ξ_g.

When a planar defect is present in the specimen, the two branches of the dispersion surface are not only coupled

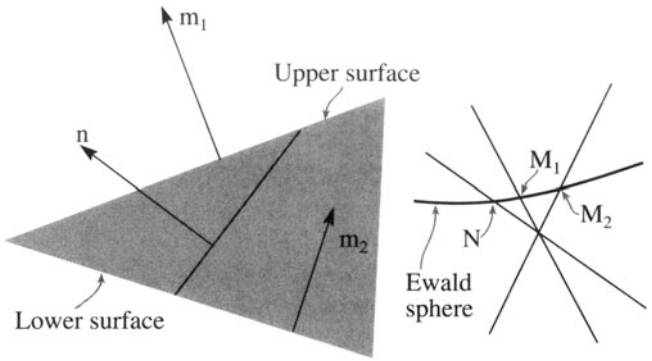


Figure 24.11. In a wedge specimen, the planar defect will, in general, be inclined to both surfaces and the relrod geometry is complex. The fringe spacing in the image is related to the reciprocal of the distances M_1N and M_2N.

along a tie line normal to the surface of the specimen but also along that normal to the planar defect. However, when **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 are a kinematical construction) must actually separate to give the two hyperbolas shown in Figure 24.12, which is why we drew Figures 17.15 and 23.6 as we did.

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 24.11 and 24.12.

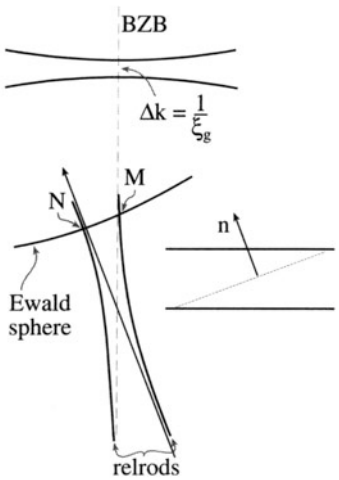


Figure 24.12. The dispersion-surface construction for an inclined planar defect in a parallel-sided specimen. (Compare with Figures 17.15 and 23.6.) For simplicity, we show the hyperbolas due to the defect alone, not the extra effects that would arise in a wedge specimen.

24.10. 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 given by s_{eff}. 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 nonchanneled one, as you can see in Figure 24.13. The effect of the planar defect is simply to couple the Bloch waves; in other words, the defect links the different branches of the dispersion surface. The noncomplementary contrast at SFs in fcc metals is directly explained by this coupling.

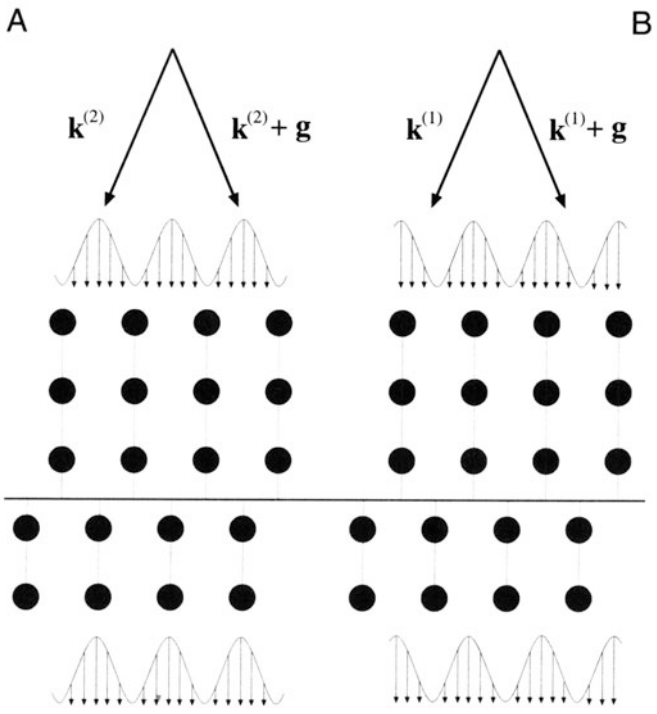


Figure 24.13. Bloch waves 1 and 2 are channeled along and between the atom columns, respectively, until they meet the fault. There the atomic columns are translated so that the channeled Bloch wave may become the nonchanneled one.

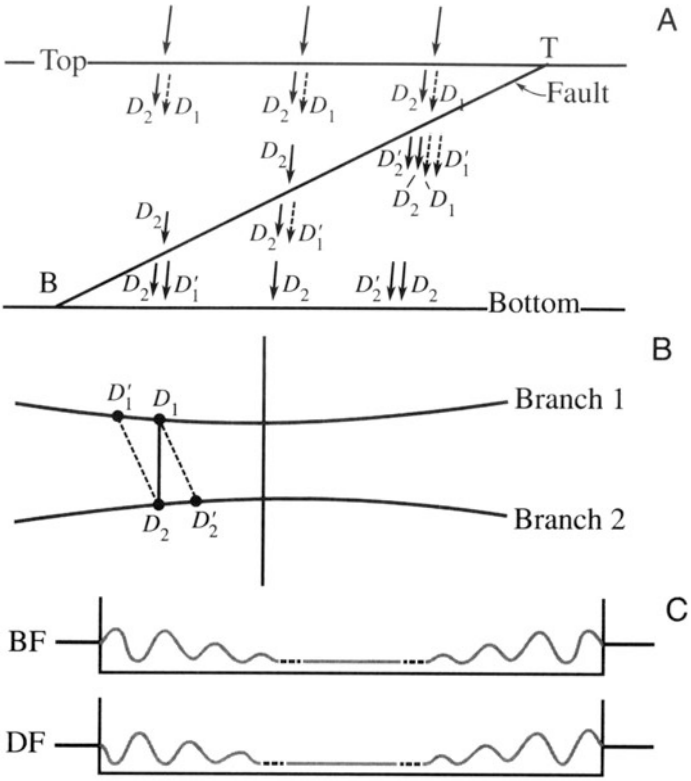


Figure 24.14. The absorption of the branch-1 Bloch wave near the top surface, T, of the specimen and its re-creation when the planar defect is near the bottom, B, determines the contrast we see. (A) shows which Bloch waves are present at the different depths in the specimen, (B) shows how the Bloch waves are coupled along tie lines joining the two branches of the dispersion surface, and (C) shows the resulting contrast.

As soon as the beam enters the specimen we excite Bloch waves 1 and 2. Therefore, in Figure 24.14A, 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 24.14A, along the tie line, D_1D_2' and $D_1'D_2$. We'll analyze the three situations shown in Figure 24.14B, 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), 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, 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, only wave D_2 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 24.14C.

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.* (1962) for further discussion on this topic.

24.11. COMPUTER MODELING

From the discussion in Sections 24.5 and 24.6, 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 first program to attempt the task of simulating two-dimensional images of planar defects is described in the book by Head *et al.* (1973) (see Section 1.5). One of several modern approaches is Comis, a Unix program, which is also available in a Macintosh version. 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 image, too.

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.* (1994) have shown using the Cufour simulation package (Schäublin and Stadelmann 1993) that the rules for fringe contrast given by Gevers *et al.* (1963) 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 24.15.

Image simulation tells us that $\mathbf{g}\cdot\mathbf{R}$ 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 1.5 before writing your own program. The subject is just as relevant to the topics of Chapters 25 and 26, but we include it in this

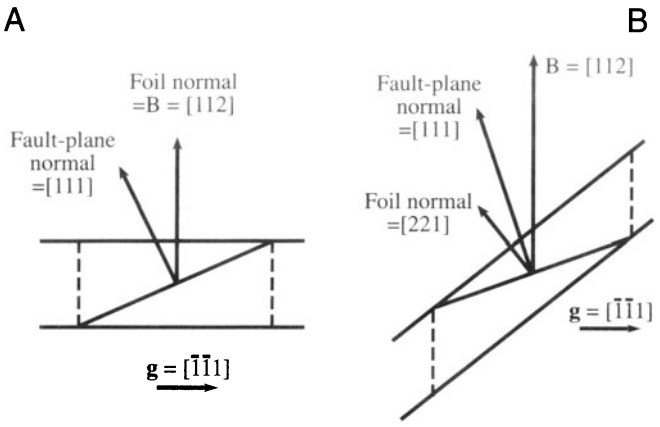


Figure 24.15. Many of the "rules" for predicting the contrast from planar "defects" make certain assumptions about the geometry of the defect relative to the surface of the specimen which may not always hold. Here it is demonstrated that the intersection of the planar defect with the upper specimen surface may be lower than the intersection with the bottom surface. This geometry can cause a reversal in the rules.

chapter mainly because the analysis of planar defects is the most straightforward application.

24.12. THE GENERALIZED CROSS SECTION

Head *et al.* (1973) 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

$$u_{\mathbf{k}}(x, y, z) = u_{\mathbf{k}}(x, 0, z + cy) \qquad [24.34]$$

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.

Examples of comparisons between experimental and simulated images using Comis are shown in Figure 24.16. This package is particularly attractive since it also performs elasticity calculations for simple defect configurations. You'll also find it instructive for simulating the effect of changing different parameters, such as:

- 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 ($z = 0.5t$).

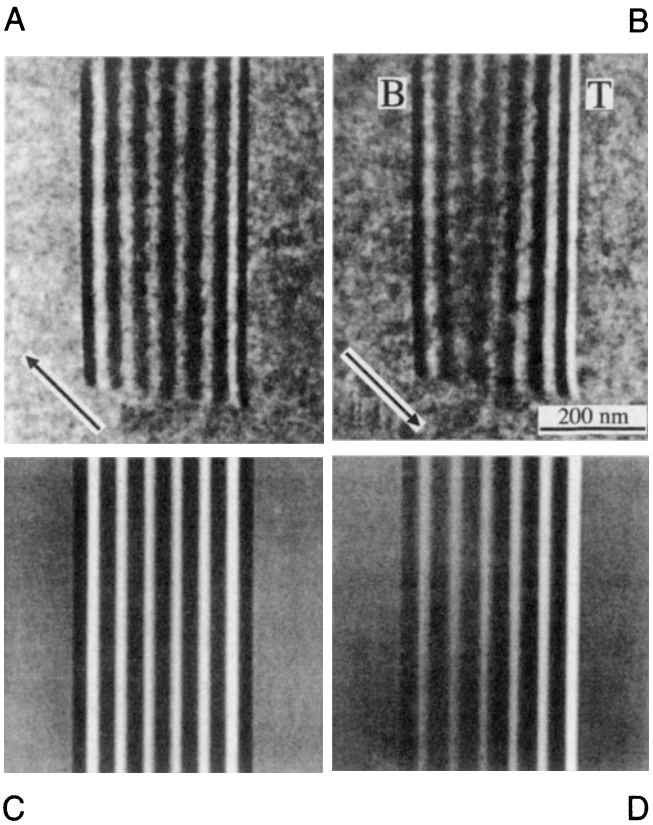


Figure 24.16. (A) Experimental BF image of an APB with **g** = 220. (B) DF image of the same defect, **g** = $\bar{2}20$. (C,D) Corresponding simulated images.

- 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 $\mathbf{s}_g$.

24.13. 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 analyses has been the uncertainty in the background-level intensity caused by inelastic scattering. As energy-filtered images become more widely available (see Chapter 40) this problem will disappear. Direct digital recording of the intensities, using a CCD camera, will also make a quantitative analysis more tractable, eliminating uncertainties associated with the calibration of the response of the emulsion of the photographic film. (See also Section 30.4.)

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 you not only to calculate the image but also give defect-interaction geometries. In simulating images of crystal defects you'll encounter several problems which are almost independent of one another. You must:

- 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 you can employ for different defect geometries. In Chapter 25 we'll consider other types of defects which may be analyzed, methods for defining them, and how you can calculate the displacement field.

24.13.A. Theoretical Basis and Parameters

We'll use the Comis program as an example of the considerations which go into a simulation program. One message which 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.

> 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 is based on the Howie–Whelan dynamical theory of electron diffraction and therefore neglects diffuse scattering. We'll follow the approach given by Howie and Basinski (1968).

- 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 $\mathbf{r} - \mathbf{R}(\mathbf{r})$ in the perfect crystal. The model is good unless $\mathbf{R}(\mathbf{r})$ 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.

> We are now including terms which allow for a variation in x and y: these terms were specifically excluded in Chapter 13 when we made the column approximation.

The equations are now written as

$$\frac{\partial \phi_{\mathbf{g}}(r)}{\partial z} = i\pi \sum_{\mathbf{h}} \left(\frac{1}{\xi_{\mathbf{g}-\mathbf{h}}} + \frac{i}{\xi'_{\mathbf{g}-\mathbf{h}}} \right) \phi_h e^{2\pi i \left(\left(s_{\mathbf{h}} + s_{\mathbf{g}} \right) z - (\mathbf{h}-\mathbf{g}) \cdot \mathbf{R} \right)} \quad [24.35]$$

$$\theta_x \frac{\partial \phi_{\mathbf{g}}}{\partial x} - \theta_y \frac{\partial \phi_{\mathbf{g}}}{\partial y} + \frac{i}{4\pi\chi_z} \left(\frac{\partial^2 \phi_{\mathbf{g}}}{\partial x^2} + \frac{\partial^2 \phi_{\mathbf{g}}}{\partial y^2} \right)$$

As usual, χ 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

$$\theta_x = \frac{(\chi + \mathbf{g})_x}{\chi_z} \quad \text{and} \quad \theta_y = \frac{(\chi + \mathbf{g})_y}{\chi_z} \qquad [24.36]$$

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, nonsystematic 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, $\mathbf{s}_h$; there are many beams and each $\mathbf{s}$ can be different.

Each extinction distance ξ_g is defined as the ratio $\chi_g/|U_g|$, as usual with U_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 (Mott and Massey 1965). 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 (1968).

You'll need to know the unit cell for your material and the Debye–Waller factors. Comis, for example, can then automatically calculate ξ_g using a built-in table of the Doyle–Turner parameters. The Debye–Waller factor (B) is related to the (mean-square) vibrational amplitude of an atom on a lattice site. So it is a temperature-sensitive term. 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.

Equation 24.35 is only valid if your crystal has a center of symmetry, otherwise you have to redefine ξ_g. Simulations involving noncentrosymmetric crystals can be performed, but you have to replace the ξ_g and ξ_g' with complex quantities and then defining all the parameters is really difficult (Gevers *et al.* 1966) (see also equation 14.2).

You can take account of absorption effects in the usual way by adding imaginary Fourier components, U_g'(Yoshioka 1957); the absorption distances, ξ_g', are then defined as we discussed in Section 23.8. There are equations which a program can use to estimate the absorption distances, e.g., the linear relation $|U_g'|/|U_g| = a + b|\mathbf{g}|$, as suggested by Humphreys and Hirsch (1968), or you could specify each individual absorption distance directly.

24.13.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 ξ_g may be replaced by an apparent extinction distance, $\xi_g^a < \xi_g$. This substitution compensates for scattering into beams that are not included in the two-beam calculation. The term ξ_g^a depends on the 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 ξ_g^a. Alternatively, you may determine ξ_g^a by comparing simulated thickness fringes calculated using many-beam and two-beam approximations.

24.13.C. Avoiding the Column Approximation

You 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 24.35. 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 26, there are situations where the column approximation will not be acceptable (Howie and Sworn 1970).

Without the column approximation, equation 24.35 gives us a system of coupled partial differential equations. The boundary conditions (at $z = 0$) can be generally written in the form

$$a_g \phi_g + b_g \frac{\partial \phi_g}{\partial x} = c_g \qquad [24.37]$$

where we're ignoring changes in the y-dimension.

You can use fixed boundary conditions (Howie and Basinski 1968):

- 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 29; we are still using

Howie–Whelan equations. Then, equation 24.35 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 24.37, $a_g = 1$, $b_g = 0$, and $c_g = \phi_g$ at the initial surface. In order to avoid distortion of the image, you must choose the step size, Δz, carefully and be sure that the distance between columns (mesh size) is small enough (Anstis and Cockayne 1979).

24.13.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, for example, certain standard menus are available for special purposes. The user can also build (and save) menus interactively. This allows all the relevant parameters and commands for a particular problem to be present within a single menu. At any time, all the commands are 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 includes a "submit" command which will start a batch job based on your current data and parameters. Thus, the interactive mode may 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 ξ_g^a such that the simulations can be carried out with only two beams (Head *et al.* 1973). 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 $\alpha = 2\pi\mathbf{g}\cdot\mathbf{R}$.
- 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 of certain Bloch waves.

You can now use computer modeling of diffraction-contrast images of planar defects to perform quantitative analysis and image matching.

REFERENCES

General References

Amelinckx, S. and Van Landuyt, J. (1978) in *Diffraction and Imaging Techniques in Material Science,* **1** and **2** (Eds. S. Amelinckx, R. Gevers, and J. Van Landuyt), 2nd edition, p. 107, North-Holland, New York.

Christian, J.W. (1975) *The Theory of Transformations in Metals and Alloys,* Part 1, 2nd edition, Pergamon Press, New York.

Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science,* Van Nostrand Reinhold, New York.

Forwood, C.T. and Clarebrough, L.M. (1991) *Electron Microscopy of In-*

terfaces in Metals and Alloys, Adam Hilger, New York. Invaluable for anyone studying interfaces by TEM.

Head, A.K., Humble, P., Clarebrough, L.M., Morton, A.J., and Forwood, C.T. (1973) *Computed Electron Micrographs and Defect Identification,* North-Holland, New York.

Rasmussen, D.R. and Carter, C.B. (1991) *J. Electron Microsc. Techniques* **18**, 429.

Sutton, A.P. and Balluffi, R.W. (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.

Specific References

Anstis, G.R. and Cockayne, D.J.H. (1979) *Acta Cryst.* **A35**, 511.

Doyle, P.A. and Turner, P.S. (1968) *Acta Cryst.* **A24**, 390.

Gevers, R., Art, A., and Amelinckx, S. (1963) *Phys. Stat. Sol.* **3**, 1563.

Gevers, R., Blank, H., and Amelinckx, S. (1966) *Phys. Stat. Sol.* **13**, 449.

Hashimoto, H., Howie, A., and Whelan, M.J. (1962) *Proc. Roy. Soc. London* **A269**, 80.

Hirsch, P., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, p. 225, Krieger, Huntington, New York.

Howie, A. and Basinski, Z.S. (1968) *Phil. Mag.* **17**, 1039.

Howie, A. and Sworn, H. (1970) *Phil. Mag.* **31**, 861.

Humphreys, C.J. and Hirsch, P.B. (1968) *Phil. Mag.* **18**, 115.

Mott, N.F. and Massey, H.S.W. (1965) *The Theory of Atomic Collisions,* 3rd edition, Clarendon Press, Oxford.

Rasmussen, D.R., McKernan, S., and Carter, C.B. (1991) *Phys. Rev. Lett.* **66**, 2629.

Schäublin, R. and Stadelmann, P. (1993) *Mater. Sci. Engng.* **A164**, 373.

Taftø, J. and Spence, J.C.H. (1982) *J. Cryst.* **15**, 60.

Thölen, A.R. (1970) *Phil. Mag.* **22**, 175.

Thölen, A.R. (1970) *Phys. Stat. Sol. (A)* **2**, 537.

Viguier, B., Hemker, K.J., and Vanderschaeve, G. (1994) *Phil. Mag.* **A69**,19.

Yoshioka, H. (1957) *J. Phys. Soc. Japan* **12**, 618.

Strain Fields

25

CHAPTER PREVIEW

As we discussed in Chapter 23, 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.

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 com-

prehensive 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.

Strain Fields

25

25.1. 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. 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 25.1. 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 23.7 showing bend contours. Regions far from the dislocation are tilted well away from the Bragg condition, while the regions either side of the dislocation core are at the Bragg condition for $\pm\,\mathbf{g}_{hk\ell}$. It is more difficult to "see" the diffracting planes for a screw dislocation (Figure 25.2A) but the effect is 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 $hk\ell$ diffracting planes (Figures 25.1 and 25.2B).
- The line direction (a vector) and therefore the character of the dislocations (edge, screw, or mixed).
- The glide plane.

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 (Section 31.1) can be very helpful, although we will not emphasize that technique. The basic requirement if you do use stereomicroscopy is that you must form all of your images using the same **g**-vector.

25.2. HOWIE–WHELAN EQUATIONS

Let's start with the two-beam phenomenological approach because it worked so well in Chapter 24. 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.

In Chapter 24, we showed that we could modify the Howie–Whelan equations to include a lattice distortion **R**. So for the imperfect crystal

$$\frac{d\phi_{\mathbf{g}}}{dz} = \frac{\pi i}{\xi_{\mathbf{0}}}\phi_{\mathbf{g}} + \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{\mathbf{0}} \exp\left[-2\pi i\left(sz + \mathbf{g}\cdot\mathbf{R}\right)\right] \qquad [25.1]$$

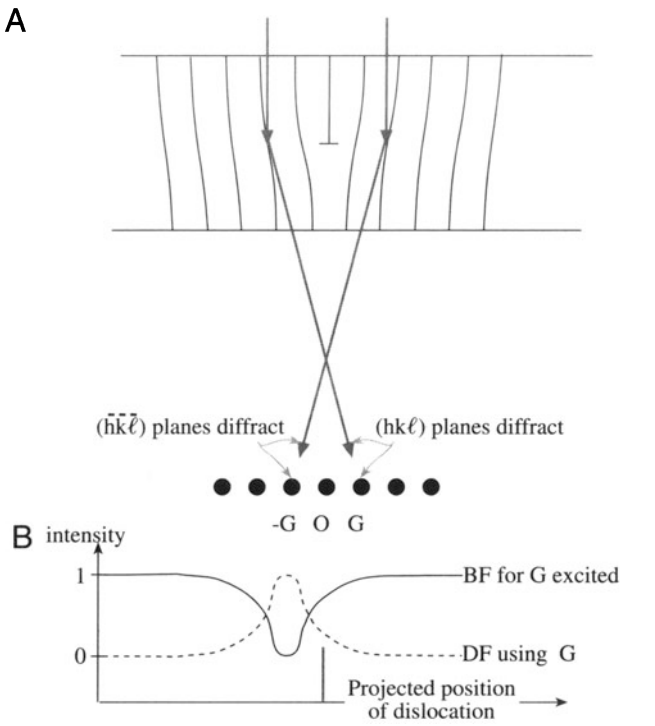


Figure 25.1. (A) The specimen is tilted slightly away from the Bragg condition ($\mathbf{s} \neq 0$). The distorted planes close to the edge dislocation are bent back into the Bragg-diffracting condition ($\mathbf{s} = 0$), diffracting into G and –G as shown. (B) Schematic profiles across the dislocation image showing that the defect contrast is displaced from the projected position of the defect.

Now we make a different substitution of variables (compare with equations 24.5 and 24.6). Set

$$\phi_0(z)_{(\text{sub})} = \phi_0(z) \exp\left(\frac{-\pi i z}{\xi_0}\right) \qquad [25.2]$$

and

$$\phi_{\mathbf{g}}(z)_{(\text{sub})} = \phi_{\mathbf{g}} \exp\left(2\pi i s z - \frac{\pi i z}{\xi_0} + 2\pi i\, \mathbf{g}\cdot\mathbf{R}\right) \qquad [25.3]$$

The justification for this substitution is the same as always. You'll notice that $\phi_0(z)_{(\text{sub})}$ is the same as before, but $\phi_{\mathbf{g}(\text{sub})}$ now includes a $\mathbf{g}\cdot\mathbf{R}$ term. The reason for this substitution is that it will give us a simple expression for $d\phi_{\mathbf{g}}/dz$.

The equations become

$$\frac{d\phi_0(z)_{(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}}\phi_{\mathbf{g}}(z)_{(\text{sub})} \qquad [25.4]$$

and

$$\frac{d\phi_{\mathbf{g}(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}}\phi_0(z)_{(\text{sub})} + \left[2\pi i\left(s + \mathbf{g}\cdot\frac{d\mathbf{R}}{dz}\right)\right]\phi_{\mathbf{g}}(z)_{(\text{sub})} \qquad [25.5]$$

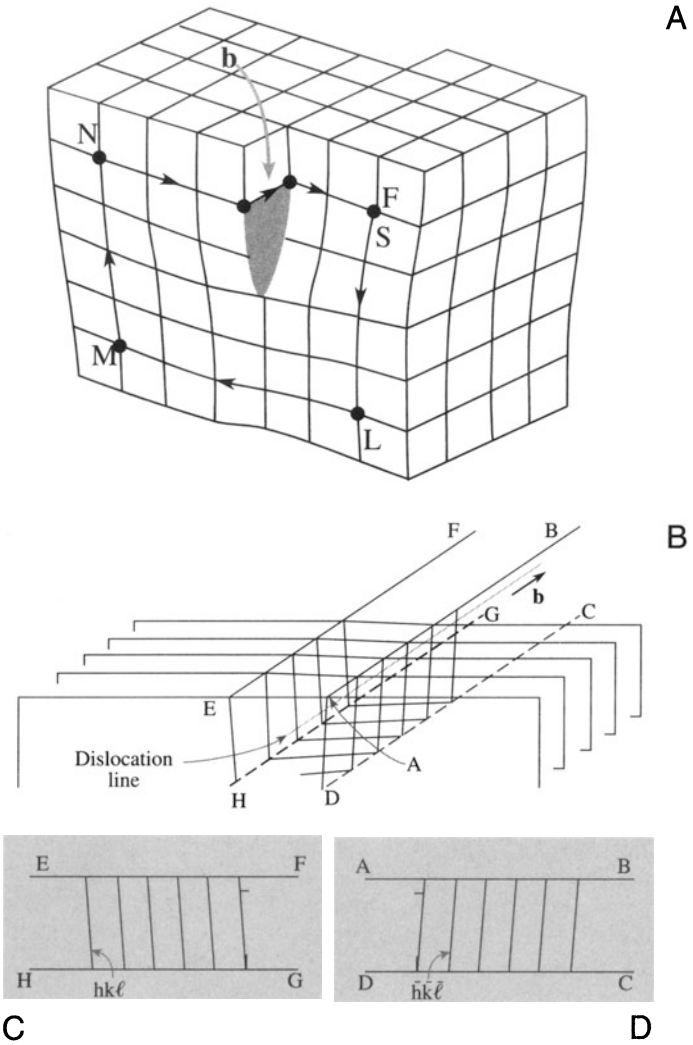


Figure 25.2. (A) Distortion of planes around a screw dislocation. The circuit SLMNF is used to define the Burgers vector, **b** (see Figure 25.5). (B) Schematic showing the rotation of the diffracting planes by a screw dislocation. The planes are rotated in opposite directions on either side (C, D) of the dislocation.

which can be rewritten, while dropping the subscript

$$\frac{d\phi_{\mathbf{g}}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}}\phi_0 + 2\pi i s_{\mathbf{R}}\phi_{\mathbf{g}} \qquad [25.6]$$

This equation looks just like equation 13.14 but with $s_{\mathbf{R}}$ instead of s, where

$$s_{\mathbf{R}} = s + \mathbf{g}\cdot\frac{d\mathbf{R}}{dz} \qquad [25.7]$$

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 ξ_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 $\mathbf{g}\cdot d\mathbf{R}/dz$ and $\mathbf{g}\cdot\mathbf{R}$ terms are used to understand dislocations. Since the equations we have just derived have the same form as those we discussed in Chapters 11 and 24, 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 24, so we have two ways of looking at the defects:

- $\mathbf{g}\cdot\mathbf{R}$ 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 $\mathbf{g}\cdot d\mathbf{R}/dz$.

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 (Hirsch *et al.* 1960). So how does the column approximation relate to the theory? The model relates **R** to the column, as shown in Figure 25.3, and the calculation is for a continuum even though we have atoms. The important point is that for a 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 (Howie and Basinski 1968; Howie and Sworn 1970). 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 **R**(z) (see Section 24.13). Remember that z is actually measured along the column.

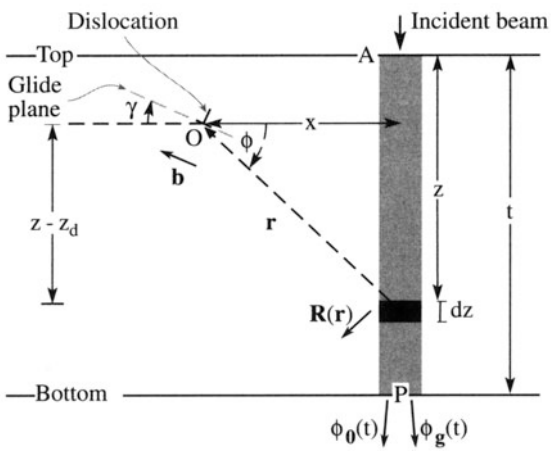


Figure 25.3. The effect of a dislocation with Burgers vector, **b,** at O on a column, distance x away. The effect of the strain field on the electron waves in the column is integrated in increments dz over its total length t, giving amplitudes $\phi_0(t)$ and $\phi_g(t)$ at P.

25.3. 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 (Hirth and Lothe 1982) can be written as

$$\mathbf{R} = \frac{1}{2\pi}\left(\mathbf{b}\phi + \frac{1}{4(1-\nu)}\{\mathbf{b}_e + \mathbf{b}\times\mathbf{u}(2(1-2\nu)\ln r + \cos 2\phi)\}\right) \quad [25.8]$$

For convenience, **R** is given here in polar coordinates (r and ϕ) shown in Figure 25.3; **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 (Steeds 1973) or just feed in displacements calculated from a computer model of the atom structure.

The amplitude of the diffracted beam, ϕ_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 25.8 simplifies to

$$\mathbf{R} = \mathbf{b}\frac{\phi}{2\pi} = \frac{\mathbf{b}}{2\pi}\tan^{-1}\left(\frac{z-z_d}{x}\right) \quad [25.9]$$

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 25.3). The dependence on $(z - z_d)$ emphasizes that the displacement field is present above

Table 25.1. Different Burgers Vectors and Different Reflections Give Different g·b = *n* Values[a]

g \ b	$\frac{1}{6}[11\bar{2}]$	$\frac{1}{6}[1\bar{2}1]$	$\frac{1}{6}[\bar{2}11]$	$\frac{1}{3}[111]$
$\pm(1\bar{1}1)$	± 1/3	± 2/3	± 1/3	± 1/3
$\pm(\bar{1}\bar{1}1)$	± 2/3	± 1/3	± 1/3	± 1/3
$\pm(0\bar{2}2)$	± 1	± 1	0	0
±(200)	± 1/3	± 1/3	± 2/3	± 2/3
$\pm(3\bar{1}1)$	0	± 1	± 1	± 1
$\pm(\bar{3}\bar{1}1)$	± 1	0	± 1	± 1

[a]The dislocations all lie on a (111) plane in an fcc material; the beam direction is [011].

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 25.1.

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 (Hirth and Lothe 1982) as illustrated in Figure 25.4. This buckling can be important because it complicates the analysis of **b** for some dislocations with an edge component, as we'll see below.

- 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**.

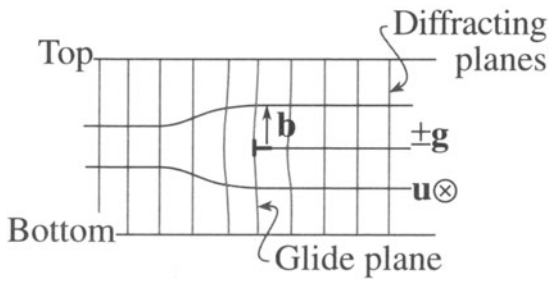


Figure 25.4. Buckling of the glide planes arises because of the term **g·b×u** and is important because it complicates the analysis of **b**.

An experimental point: you usually set *s* to be greater than 0 when imaging a dislocation. 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 25.1.

The + and – signs in Table 25.1 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 position of the dislocation core. If you look carefully at Figure 25.1, 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 (**g·b**)*s* ("g-dot-b-times-s"), as shown in Figure 25.5.

If **g·b** = 0 , then you won't see any contrast because the diffracting planes are then parallel to **R**. This is termed the *invisibility criterion.*

If we identify two reflections, $\mathbf{g}_1$ and $\mathbf{g}_2$, for which **g·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 **g·b** < 1/3; similarly, dislocations need not be invisible even if **g·b** = 0 when **g·b×u** ≠ 0. Further exceptions to the rule are given in Edington (1976).

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 in the experimental image shown in Figure 25.6A.

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 **g·b** or, more completely, on (**g·b**)*s* and **g·b×u**.
- If we repeat this analysis for other values of **g·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.

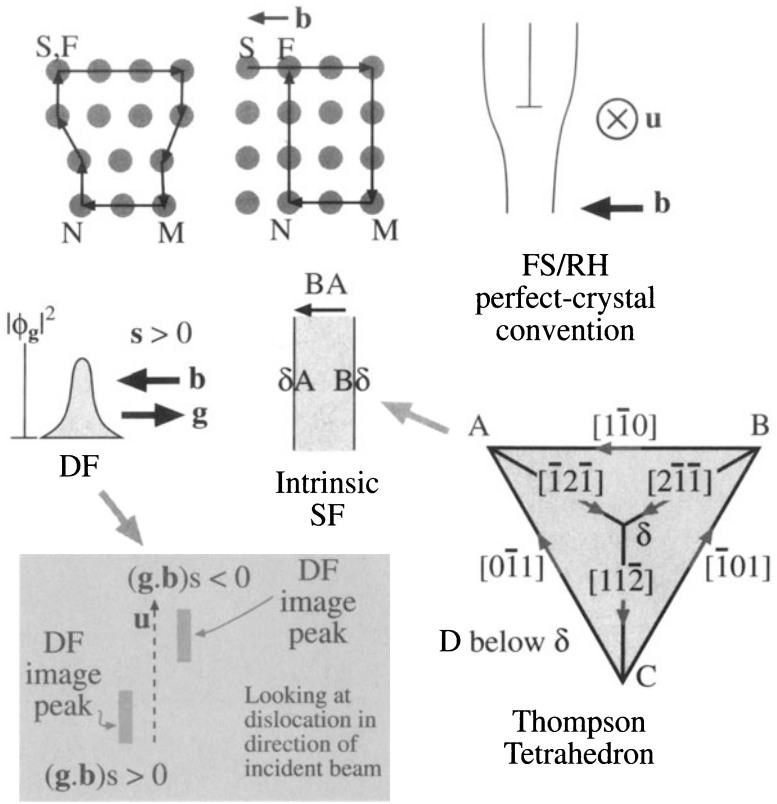


Figure 25.5. A brief summary of dislocations in an fcc crystal: **b** is defined by the finish- (F) to-start (S) vector in a right-hand (RH) circuit that comes to closure around the dislocation but fails to close in the perfect crystal. The location of the diffracted intensity $|\phi_g|^2$ relative to 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 dislocations is given by the Thompson tetrahedron.

A final "rule of thumb" which you may find useful (from computer modeling and early analytical calculations) is

> If **g**·**b** = 0, you can still "see" dislocations when **g**·**b**×**u** ≥ 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 25.6. 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 Figures 25.6A–C. The single line below the arrow in (C) is a dislocation having its Burgers vector parallel to **g** here (so **g**·**b** = 2); you can see two "peaks" in the image, one darker and broader than the other. Notice that one peak has nearly disappeared in (B) and the broad peak is on the other side of the dislocation. A group of three parallel lines is present in (C) but is out of contrast in (B). These are three Shockley partial dislocations all having the same **b** and thus all 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; these faults will never give contrast at the 111 pole since **g**·**R** is then always 0 (or an integer). The 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 (Figures 25.6D 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 25.6F). Dislocations in interphase boundaries can be revealed by imaging with different reflections (Figures 25.6G 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. 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 25.6.

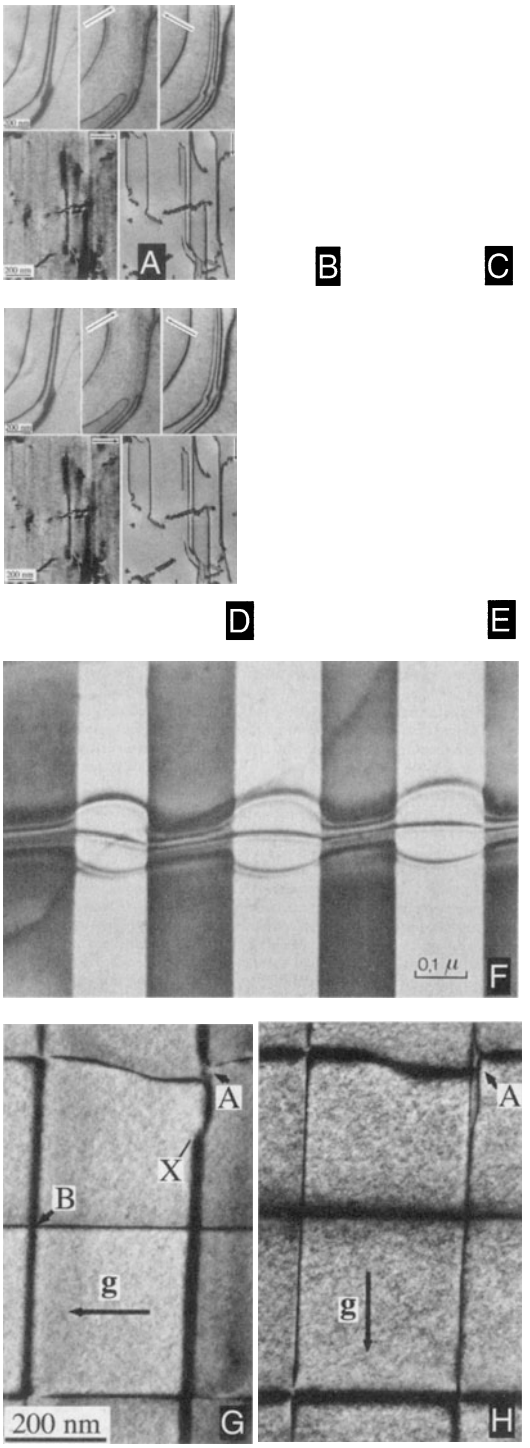


Figure 25.6. (A–C) Three strong-beam BF images from the same area using (A) {11$\bar{1}$} and (B,C) {220} reflections to image dislocations which lie nearly parallel to the (111) foil surface in a Cu alloy which has a low stacking-fault energy. (D,E) Dislocations in Ni_3Al in a (001) foil imaged in two 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 in a (001) interface between two slightly lattice-mismatched III-V compounds.

25.4. DISPLACEMENT FIELDS AND EWALD'S SPHERE

In Section 25.3, we showed that when a displacement $\mathbf{R}(\mathbf{r})$ is present, we can think of s as being replaced by $s_{\mathbf{R}}$ (equation 25.7). Hirsch *et al.* (1977) (see also Goringe 1975) showed that this new s should be written more completely as

$$s_{\mathbf{R}} = s + \mathbf{g} \cdot \frac{\partial \mathbf{R}}{\partial z} + \theta_B \mathbf{g} \cdot \frac{\partial \mathbf{R}}{\partial x} \qquad [25.10]$$

The point is that, as you can see in Figure 25.7, **R** causes the lattice planes to bend through an angle $\delta\phi$. So two other parameters, namely, **g** and s, also change. The diffraction vector is actually lengthened by $\Delta\mathbf{g}$ but, more importantly, **g** is rotated. The result is that s increases by the two com-

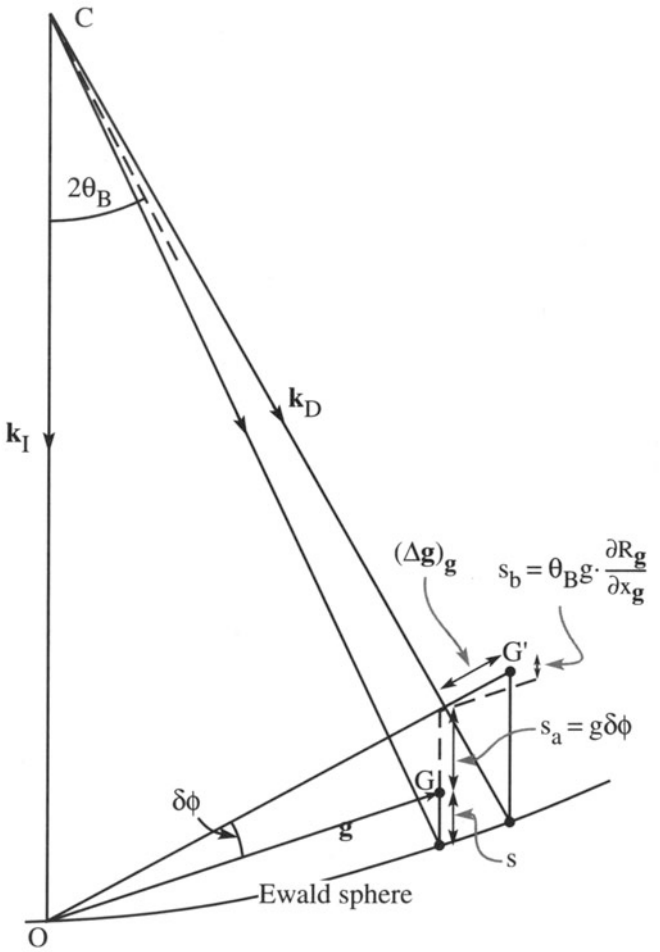


Figure 25.7. The strain field of the dislocation causes the lattice planes to bend through an angle $\delta\phi$. So **g** and s also change. The diffraction vector is lengthened by $\Delta\mathbf{g}$ and **g** is rotated. So s increases by the two components of $s_{\mathbf{R}}$, i.e., s_a and s_b.

ponents, s_a and s_b, shown in the figure, to give s_R. If you manipulate this equation for small angles you produce equation 25.8. We usually neglect the third term because θ_B is small, but it can become important when screw dislocations intersect the surface (Tunstall *et al.* 1964).

An alternative way of looking at this deformation is to think of **g** as changing by $\Delta\mathbf{g}$. We can define this change by the equation

$$\mathbf{g} \cdot \left(\mathbf{r} - \mathbf{R}(\mathbf{r})\right) = \left(\mathbf{g} + \Delta\mathbf{g}\right) \cdot \mathbf{r} \qquad [25.11]$$

so that

$$-\mathbf{g} \cdot \mathbf{R}(\mathbf{r}) = \Delta\mathbf{g} \cdot \mathbf{r} \qquad [25.12]$$

The implication is that the information about the displacement field, $\mathbf{R}(\mathbf{r})$, 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. 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 24.13), we make the assumption that the atom doesn't know it has moved. If $\mathbf{R}(\mathbf{r})$ 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 $\mathbf{R}(\mathbf{r})$, are large, as at dislocation cores.

25.5. 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 25.8 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 25.8A. 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.

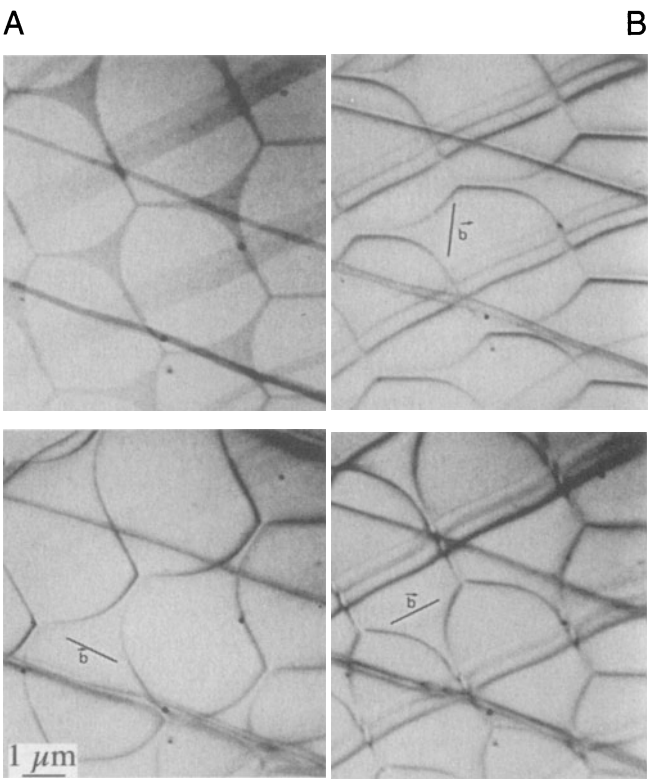


Figure 25.8. Dislocation networks in graphite. In (A) the stacking faults parallel to the foil surface are imaged, in (B), (C), and (D) two of the three dislocations at each node are in contrast but the third is invisible. Knowing **g** for each image, the Burgers vector of the dislocations can be determined as shown.

25.6. 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 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**, as shown in Figure 25.9.
- Loops can be present which show no **g**·**b** contrast.
- Loops can enclose single or multiple stacking faults, and so exhibit SF contrast as shown in Figure 25.10.
- 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!

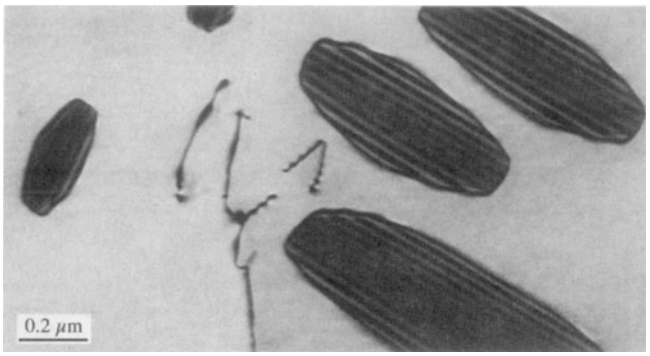


Figure 25.10. Dislocation loops in irradiated Ni showing SF contrast.

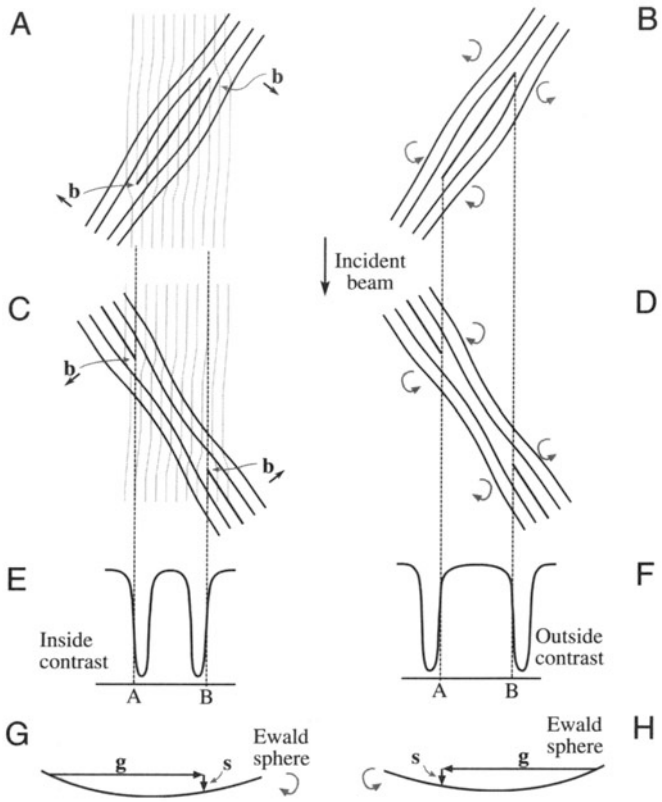


Figure 25.9. (A) Structure of an interstitial loop relative to the diffracting planes (faint lines). (B) Arrows show the rotation of the diffracting planes around the dislocation. (C,D) Vacancy loops. (E,F) Position of the image contrast relative to the projected dislocation position. Inside contrast 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).

Dislocations in Zn provide a particularly nice illustration of **g**·**b**×**u** contrast. If the specimen surface is parallel to the (0001) basal plane, then dislocation loops can readily form by coalescence of vacancies. In Figure 25.11, **b** is normal to **g** so that **g**·**b** = 0. These loops give a clear illustration of how the appearance of the image depends on the line direction, **u**, of a dislocation. Note that you can see the dislocation, 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.

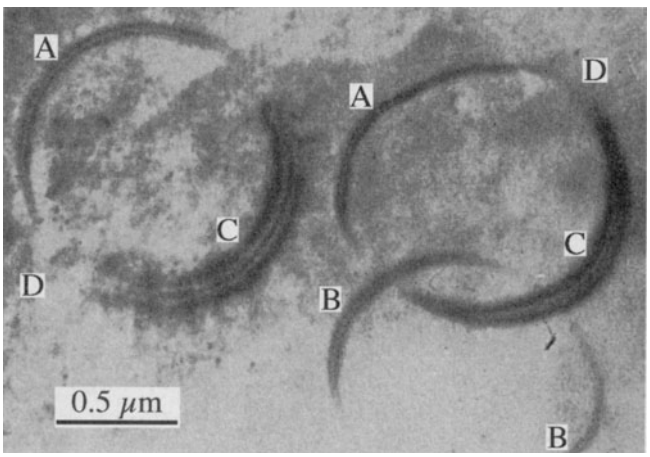


Figure 25.11. Prismatic loops in Zn parallel to the (0001) surface of the specimen with **b** = *c*[0001]. All round the loop, **b** is normal to **g** so that **g**·**b** = 0 and the vector **b**×**u** lies in the plane of the loop. At A, B, and C, **b**×**u** is parallel to **g** so that we see strong contrast. However, at D, **b**×**u** and **g** are mutually perpendicular so that **g**·**b**×**u** = 0 and the loop disappears.

The basic idea is that the appearance of the image is now dependent on the thickness of the specimen.

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 25.12 summarizes the contrast which 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 common terms. Notice that the behavior of the contrast differs in BF and DF images; this effect is similar to that which we discussed in Chapter 24 and is again related to anomalous absorption. A detailed description of this complex contrast behavior is given by Wilkens (1978).

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 vectors, lying on parallel glide planes. As a result, they are

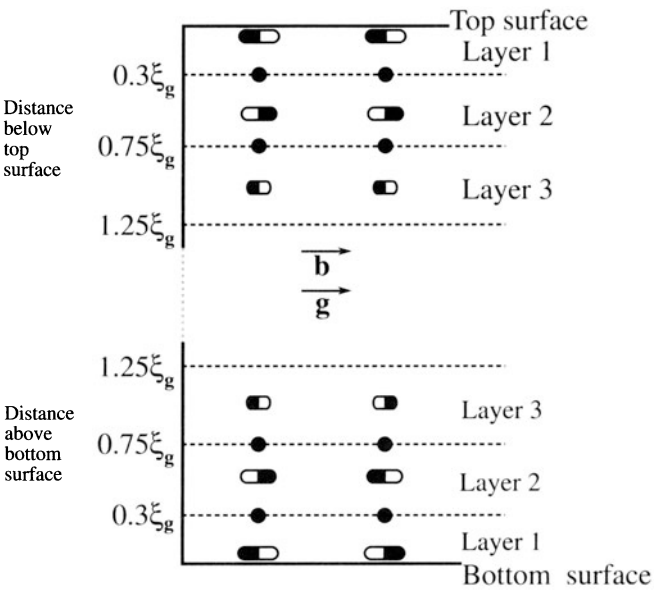


Figure 25.12. Changes in the black–white contrast from small dislocation loops at different layer depths in the specimen. The DF shows the same contrast at the top and bottom while the BF contrast is complementary at the two surfaces.

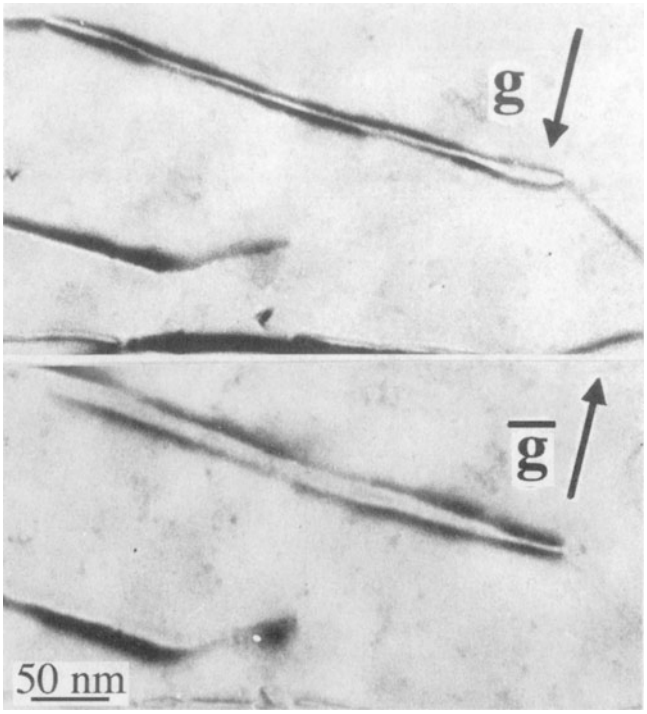


Figure 25.13. Images of dislocation dipoles in Cu showing inside–outside contrast on reversing **g** (±220).

best recognized by their "inside–outside" contrast as illustrated in Figure 25.13. 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 25.9 tells you that one image will lie on one side of the core and the other on the opposite side. The order reverses when you reverse **g**.

25.7. 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 25.8, this is often helpful if you have SFs associated with the dislocations; you can then produce $\mathbf{g}\cdot\mathbf{R}$ contrast for the fault. We will discuss dislocation dissociation more in Chapter 26. If you look back at Figure 25.6, 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

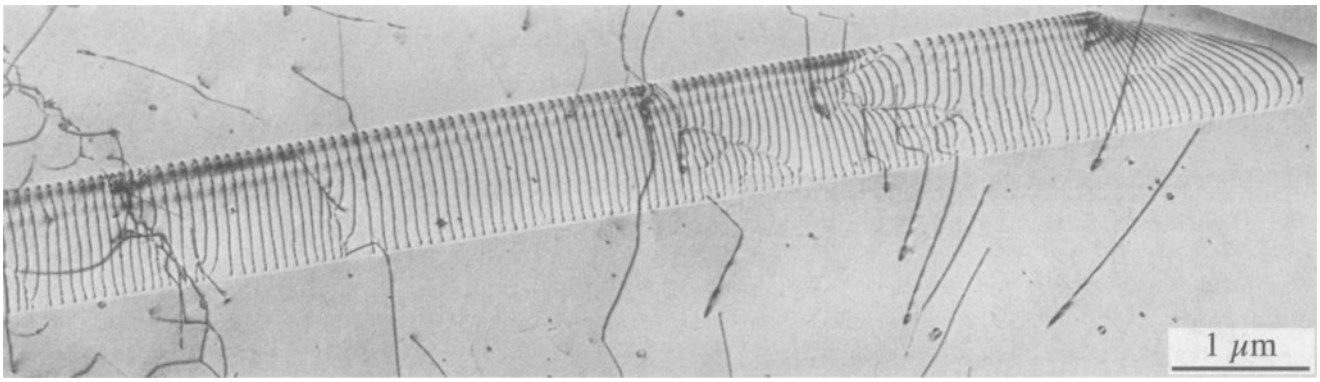


Figure 25.14. Dislocations threading through a very thick specimen in an image recorded using a high-voltage TEM.

$$\frac{1}{2}\left[1\bar{1}0\right] = \frac{1}{6}\left[1\bar{2}1\right] + \frac{1}{6}\left[2\bar{1}\bar{1}\right] \quad \text{on} \left(111\right) \qquad [25.13]$$

If we image this dislocation using the $(2\bar{2}0$ reflection, then $\mathbf{g}\cdot\mathbf{b} = 2$. If, instead, we use the $(20\bar{2})$ reflection, then $\mathbf{g}\cdot\mathbf{b} = 1$. 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 25.14; 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. 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 25.15A. The defects may be very close together in heavily deformed materials, as shown in Figure 25.15B; you then need to make the specimen thin over very large areas to minimize image overlap (e.g., Hughes and Hansen 1995). If the density of defects is too large, the weak-beam technique may be the only way to "look into" the walls (Chapter 26).

25.8. 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 25.16 (Amelinckx 1964).

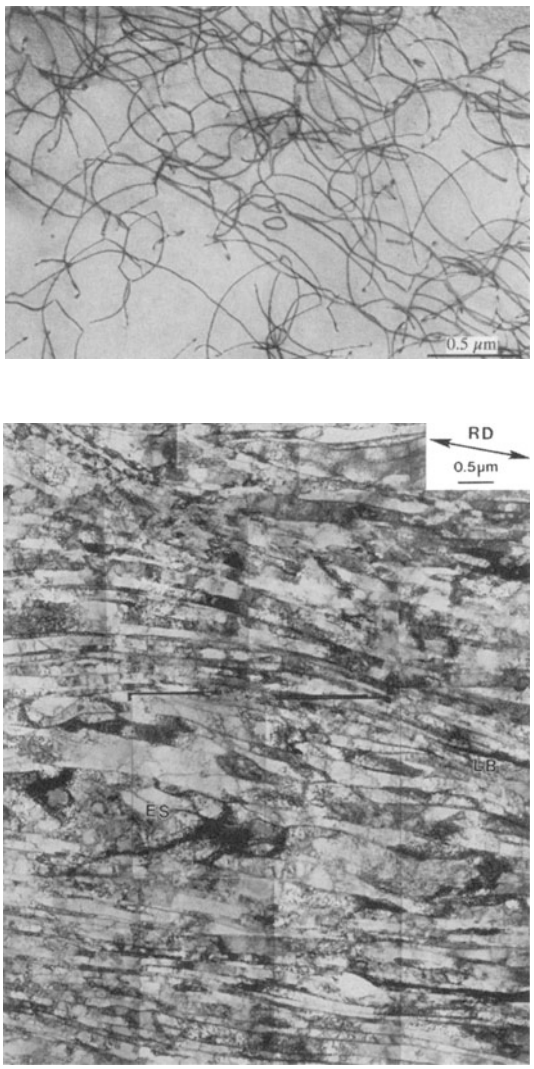


Figure 25.15. (A) Dislocation tangles in an Fe-35% Ni-20% Cr alloy, creep tested at 700°C; the dislocations have moved by glide and climb and do not lie on well-defined planes. (B) Dislocation walls in Al which has been heavily deformed by directional rolling.

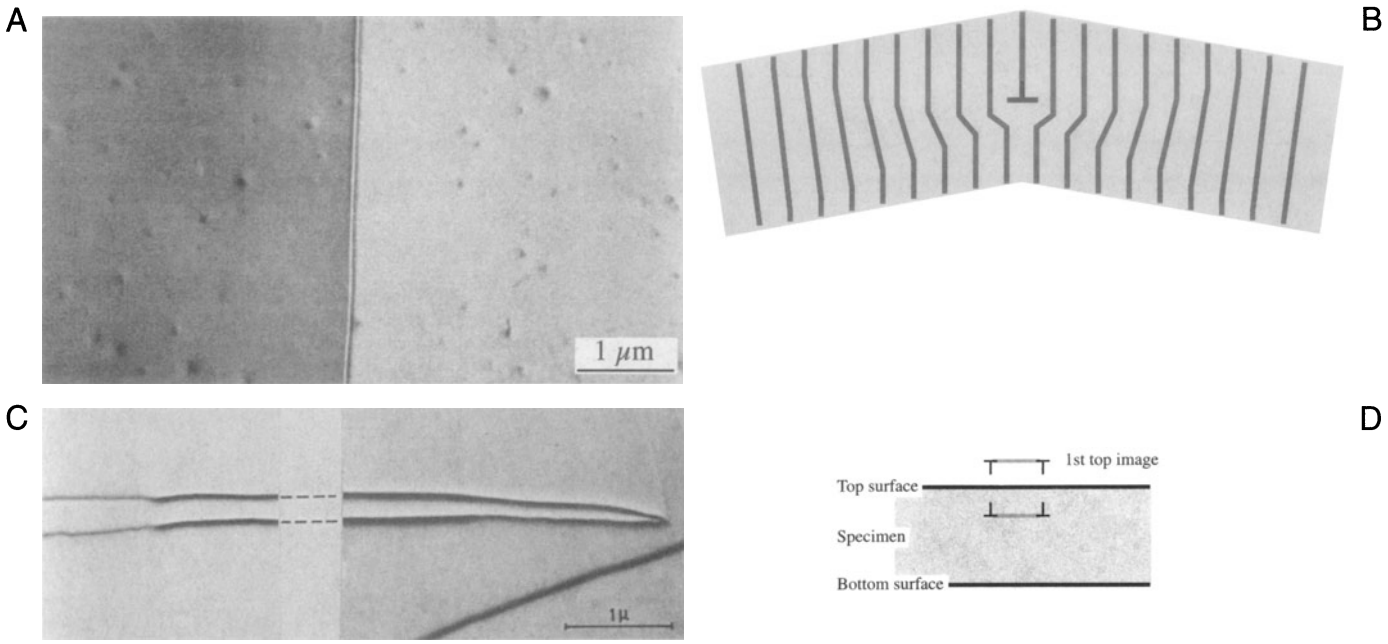


Figure 25.16. (A) A single-edge dislocation lying parallel to the surface of a very thin foil of $SnSe_2$ causes the diffracting planes to bend (B) so we see different intensity in the matrix on either side of the defect. (C,D) If the dislocation is dissociated, the image forces due to the surface cause its width to decrease. The schematic in (D) shows the image dislocations included to represent the effect of the surface.

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* $\theta_D = 0.29^\circ$, *which you can compare to a Bragg angle from sin* $\theta_B = (n\lambda/2d)$ *(0.0037 nm/0.05 nm = 0.0074).* $\theta_B = 0.42^\circ$. *Notice how* θ_D *increases for thinner foils and* θ_B *decreases for increasing voltage (decreasing* λ*).*

Similarly, if the dislocation is dissociated, the proximity of the surface causes its width to decrease. We can model this situation using "image dislocations" as shown in Figures 25.16C and D. 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 geometry, as we'll show in Section 26.8.

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 25.17.

Takayanagi (1988) has shown that 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 (111) Au film is more dense than the rest of the film and this misfit is accommodated by surface dislocations, as shown in Figure 25.18. 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.

Dislocation can be viewed nearly parallel to their line directions, when we still see contrast even for screw dislocations, as you can see in Figure 25.19A (Tunstall *et al.* 1964). Initially, this contrast is surprising since $\mathbf{g}\cdot\mathbf{b}$ and $\mathbf{g}\cdot\mathbf{b}\times\mathbf{u}$ must be zero for any screw dislocation. However, the screw dislocation can relax at the surface, as shown in Figure 25.19B (Amelinckx 1992).

25.9. 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 25.20. This is one topic where image simulation, which we'll discuss in Section 29.12, is invaluable.

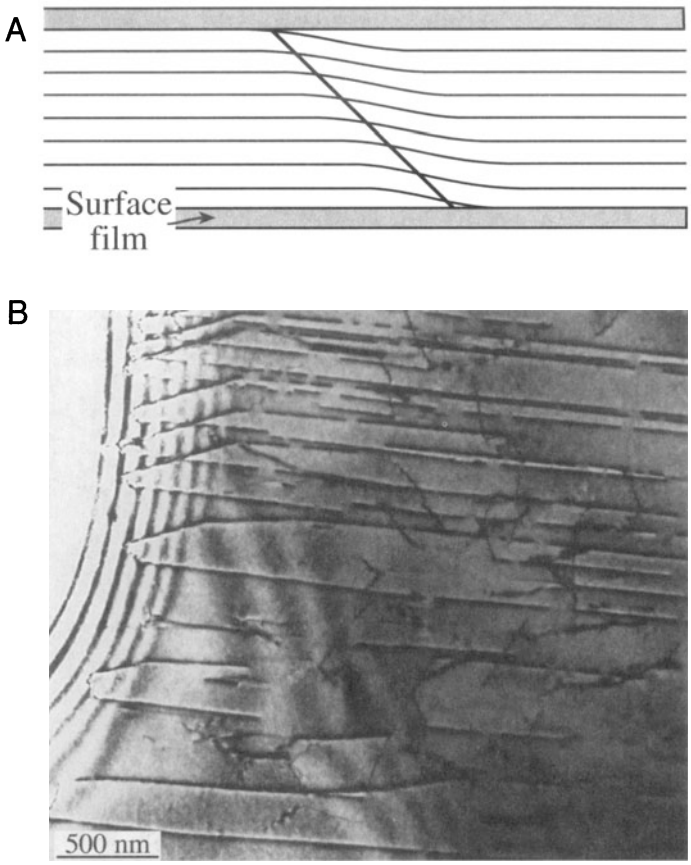


Figure 25.17. (A) Schematic diagram of dislocations pinned at the surface of the specimen by surface films such as oxides. (B) A reduced (i.e., metal) film on NiO pins dislocations. Such films may be introduced during or after thinning to electron transparency.

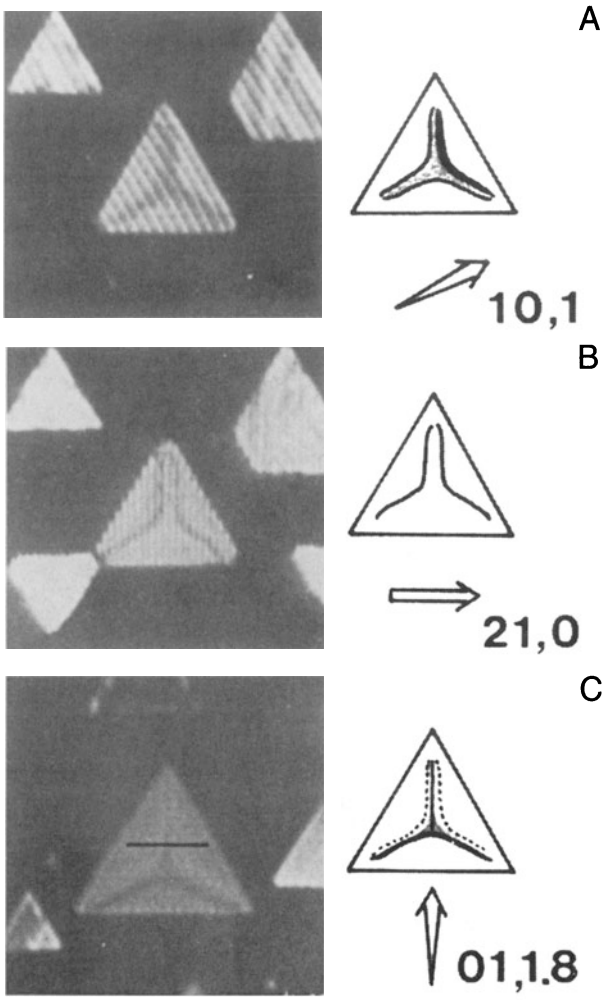


Figure 25.18. Dislocations networks can form at the surface of (111) Au islands because the surface layer relaxes to a "lattice" parameter which is different from that of the bulk material. Different dislocations are visible under different diffracting conditions (A–C).

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 grain boundary model of an interface.

Dislocations can be present at interfaces where the composition, or structure, or both change.

- Misfit dislocations accommodate the difference in lattice parameter between two well-aligned crystalline grains. Surface dislocations (as we saw in Section 25.8) are a special subgroup of misfit dislocations.
- Transformation dislocations are the dislocations which 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. An example of such steps is shown in Figure 25.21. 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 26 and using HRTEM in Chapter 28.

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, grain boundaries appear as arrays of dislocations. In general, the grains are misoriented. There are some special cases, as we saw in Chapter 24.

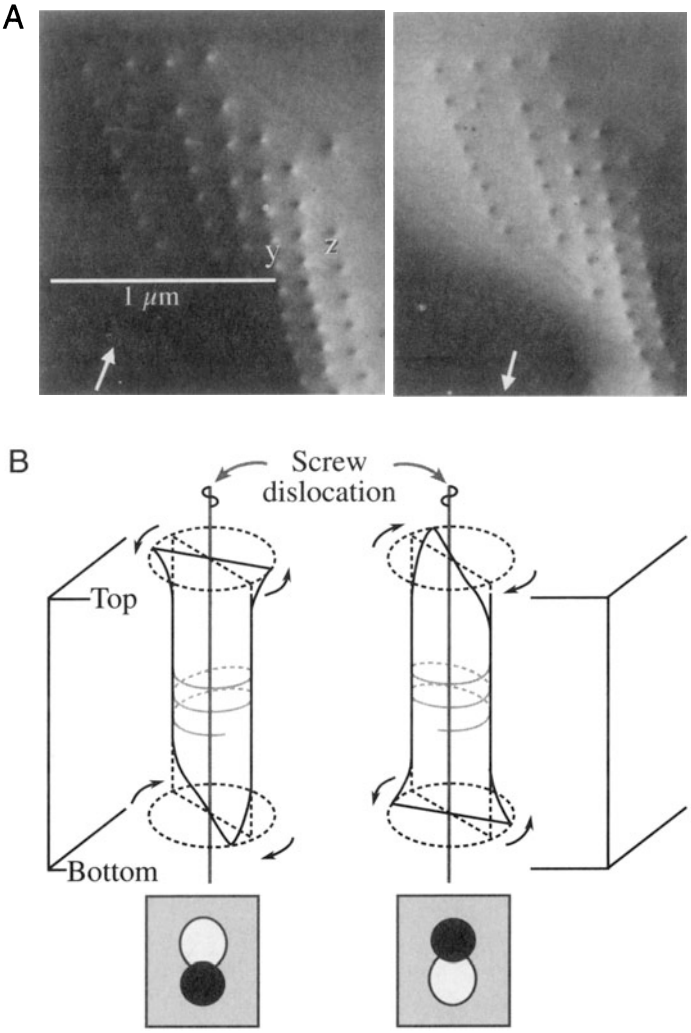


Figure 25.19. (A) Screw dislocations viewed end on (±**g**). The schematics (B) show the diffracting planes rotating in the same direction away from the edge-on orientation at both surfaces.

- Two grains may have a near-common plane and therefore a nearly common, but different, **g**-vector.
- In small-angle grain boundaries, θ is small, so the separation of the dislocations is large ($\sin \theta/2 \cong b/2d$).

The $\Sigma = 3$ twin boundary in fcc materials is an example of an interface where you can use common, but different, **g**-vectors. Here, the $(3\bar{3}\bar{3})$ plane is parallel to the $(\bar{5}11)$ plane, so these two **g**-vectors are identical, as you can see in Figure 25.22. However, this common reflection would not normally be used because **g** is rather large. This coincidence can also occur for other grain boundaries.

In the case of small-angle boundaries, we can pretend that the reflection is common to both grains, as illustrated in Figure 25.23. What we are really doing is treating

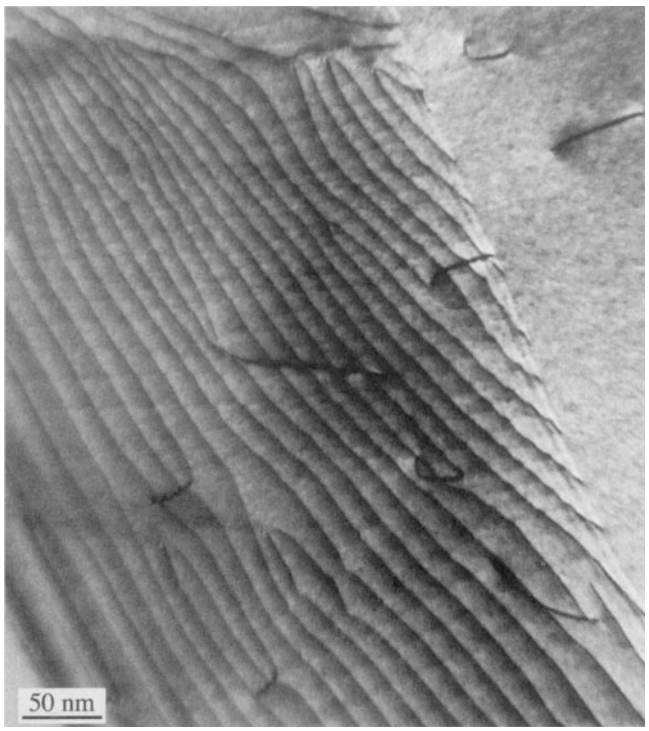


Figure 25.20. Dislocations interacting with a grain boundary; the dislocation contrast changes because its strain field changes when it enters the boundary and becomes part of the dislocation structure.

the dislocations as if they were isolated lattice defects; actually, the **g**-vectors for the two grains will be rotated relative to one another.

Lattice misfit is very important whenever we are studying thin films; dislocations are often present to accommodate the misfit. An example is shown in Figure 25.24, where dislocations are present between spinel and

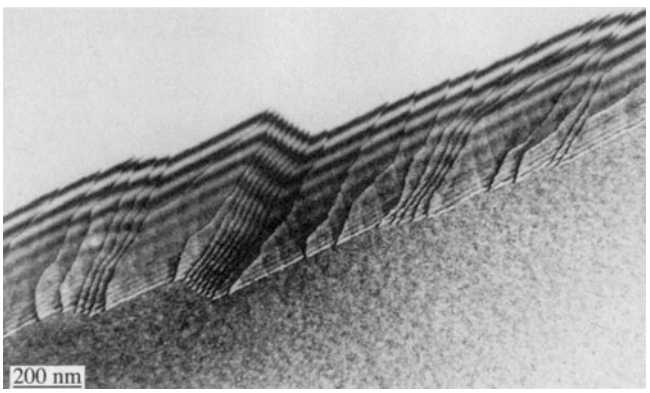


Figure 25.21. Steps at interfaces may also cause diffraction contrast. In this Ge specimen, the steps displace the thickness fringes in the GB. The fringe spacing is different at the top and bottom of the boundary because the diffraction conditions are different in each grain.

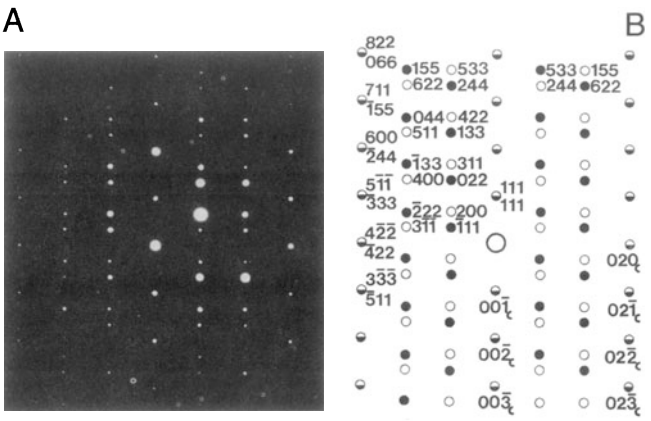


Figure 25.22. (A) DP and (B) its indexed schematic for a $\Sigma = 3$ twin boundary in an fcc material. Notice that many pairs of **g**-vectors exactly overlap but have very different indices.

NiO; these two materials both have the same fcc crystal structure. Although you can easily appreciate the change in lattice parameter, you must remember that there is also a less obvious change in the elastic constants.

> This means that the strain field at phase boundaries is *not* the same as at a grain boundary.

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.

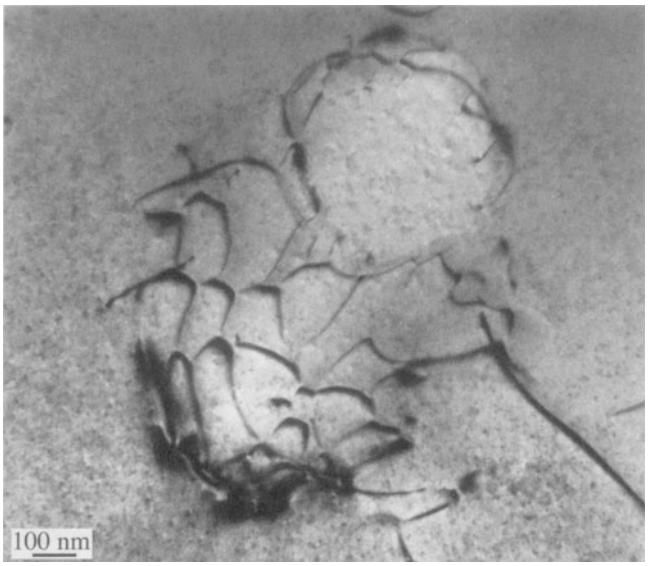


Figure 25.24. An irregular array of misfit dislocations at the interface between a spinel particle and an NiO matrix. The lattice mismatch is very small as you can appreciate from the scale. Although you can "see" a distorted hexagonal array of dislocations, you have to remember that this interface is actually curving within the specimen so that we are only seeing a projection of the structure.

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, so as to physically translate the interface as the transformation proceeds. However, you will find it dif-

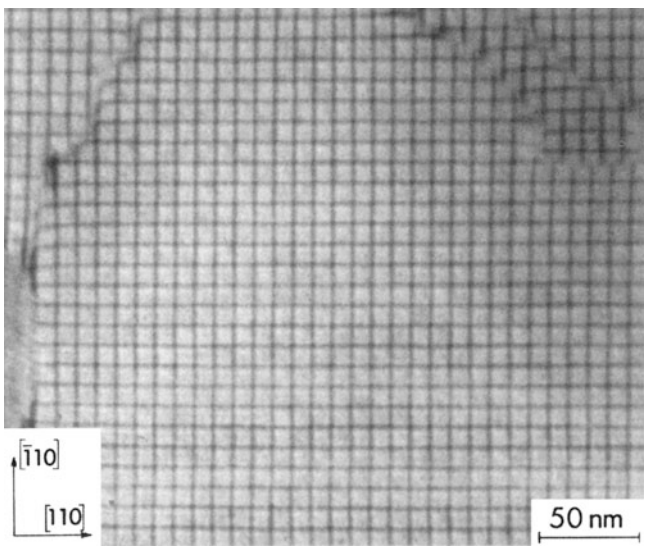


Figure 25.23. A low-angle (001) twist boundary in Si oriented almost exactly parallel to the specimen surface. Two (040) reflections were excited to form this BF image, but for small misorientations these are so close that we treat them as one reflection.

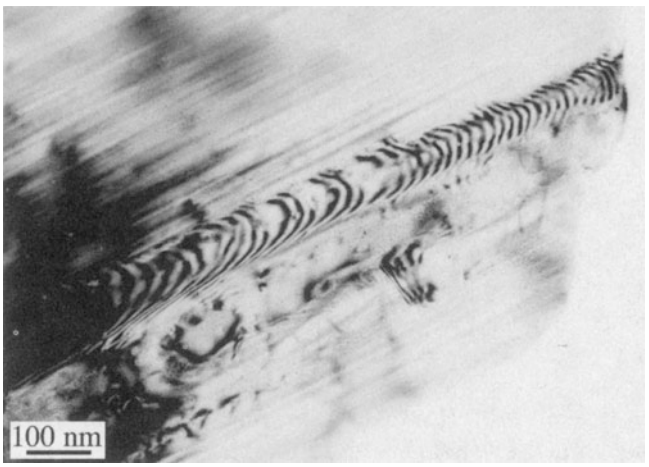


Figure 25.25. Transformation dislocations in the interface between a growing lath of hematite (pseudo-hexagonal alumina structure) in a ferrite (cubic spinel structure) matrix. The dislocations are curved because they were moving while heating the thinned specimen, which is why we know they are transformation dislocations, not simply misfit dislocations.

ficult to model the contrast from such dislocations, especially when you have a thin layer of the new phase enclosed by the matrix, as in the case when a precipitate grows, as illustrated in Figure 25.25.

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'll discuss moiré fringes in Chapter 27). 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 26) and carefully examine the DP.

Humble and Forwood (1975) 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.

25.10. VOLUME DEFECTS AND PARTICLES

When the defects are small, the image may be dominated by the strain-field contrast, and that is the aspect we are considering here. You have to remember, though, that the defects may have a different structure, lattice parameter, and composition. The theory for a spherical particle in a matrix was given over 30 years ago by Ashby and Brown (1963).

> The theory works well for coherent particles but as soon as the first interface dislocation appears, 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 in Figure 25.26. 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

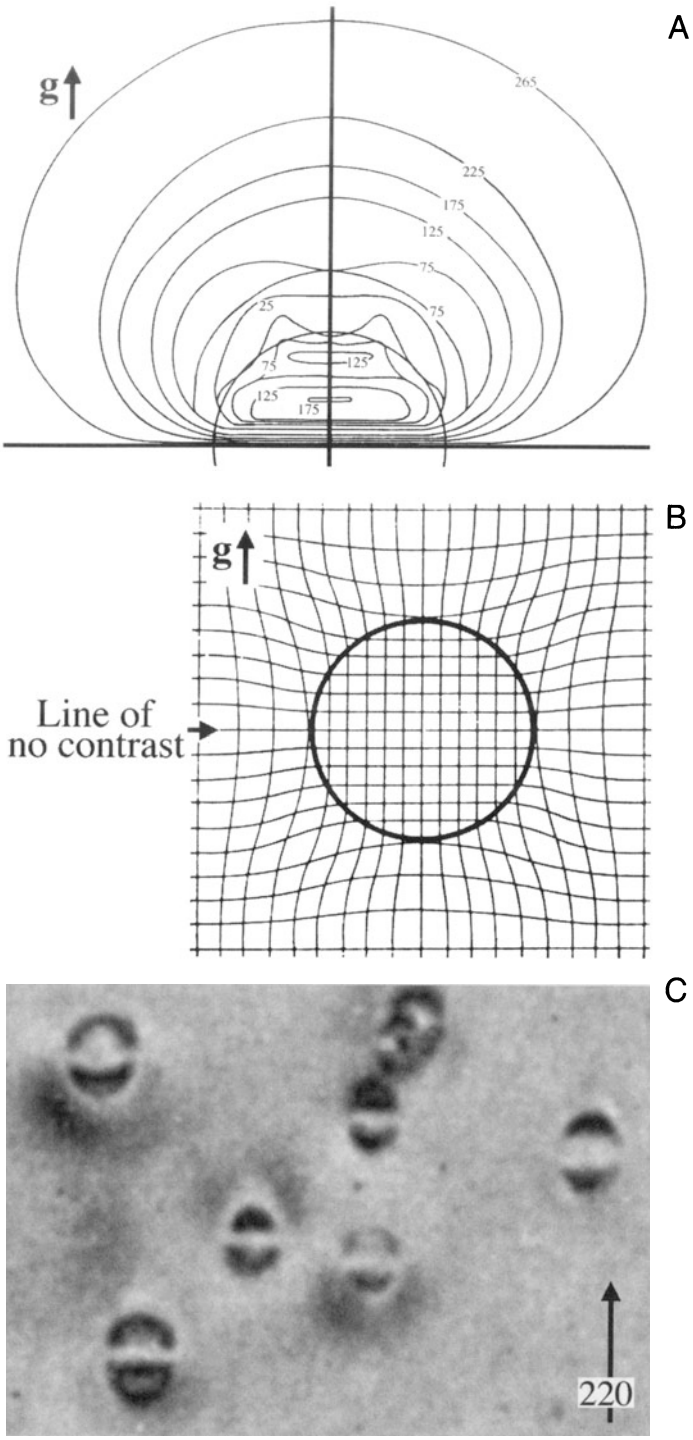


Figure 25.26. (A) Intensity contours from a simulated image of a particle like that shown schematically in (B). Notice the line of no contrast which corresponds to the plane that is not distorted by the strain field of the particle. (C) Experimental image of coherent particles in Cu-Co showing strain contrast.

the lattice strain surrounding a single precipitate, which is really quite remarkable. The process requires some specific experimental conditions and careful calibration of the photographic recording procedure, which there isn't space to describe, 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 25.26B 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

$$\mathbf{R} = C_\varepsilon \mathbf{r} \qquad [25.14]$$

when $\mathbf{r} \leq \mathbf{r}_0$, and

$$\mathbf{R} = C_\varepsilon \frac{r_0^3}{r^3} \mathbf{r} \qquad [25.15]$$

when $\mathbf{r} \geq \mathbf{r}_0$, where C_ε is an expression for the elastic constants, given by

$$C_\varepsilon = \frac{3K\delta}{3K + 2E(1+\nu)} \qquad [25.16]$$

and 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 shows radial symmetry. Thus, when we consider the Howie–Whelan equations, we realize that when $\mathbf{g}\cdot\mathbf{R} = 0$ 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 25.26A. In the image from a specimen of a Cu-Co alloy containing small Co precipitates shown in Figure 25.26C, 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 (e.g., Karth *et al.* 1995).

25.11. 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 24.11 to 24.13.

If you want to make quantitative comparisons with real images recorded on film, you must correct for the nonlinearity of the film (see Chapter 30).

Although the algorithm employed by the Head *et al.* (1973) programs allow 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 nonparallel dislocations, Thölén (1970a) introduced a matrix algorithm. As we saw in Chapter 24, Howie and Basinski (1968) extended the two-beam calculations to include several beams on the systematic row and presented a method for circumventing the column approximation.

25.11.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.* (1973), 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.

25.11.B. Crystal Defects and Calculating the Displacement Field

The program Comis can simulate amplitude contrast from any number of defects consisting of fault planes and straight, infinite dislocations (Rasmussen and Carter 1991). You just need to define the Burgers vector, line di-

rection, and relative position; planar faults are defined by the plane normal, the displacement vector, and the relative position. You can predefine certain standard geometries to ease the process of defining the defect system.

Once you've defined the defect geometry, 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), the program will determine this region and provide it as the default. However, you can always set the image region manually in Ångström units, to obtain a desired magnification.

The displacement field for the dislocations is calculated using linear, anisotropic elasticity theory (Eshelby *et al.* 1953) and is based on the algorithms given by Head *et al.* (1973), so you must 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. You can introduce image dislocations outside the crystal in order to include surface effects.

25.11.C. The Parameters

An example which shows simulated images of an orthogonal network of screw dislocations is given in Figure 25.27. Thölén (1970b) has analyzed this situation in detail. Comis can calculate the equilibrium configuration of certain types of interacting dislocations (Morton and Forwood 1973) using anisotropic elasticity theory, and then directly incorporate the resulting geometry in subsequent image simulations. As you can appreciate from equation 25.5, 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 $|U_g'|/|U_g| = 0.1$.
- The number of beams included in the calculation.

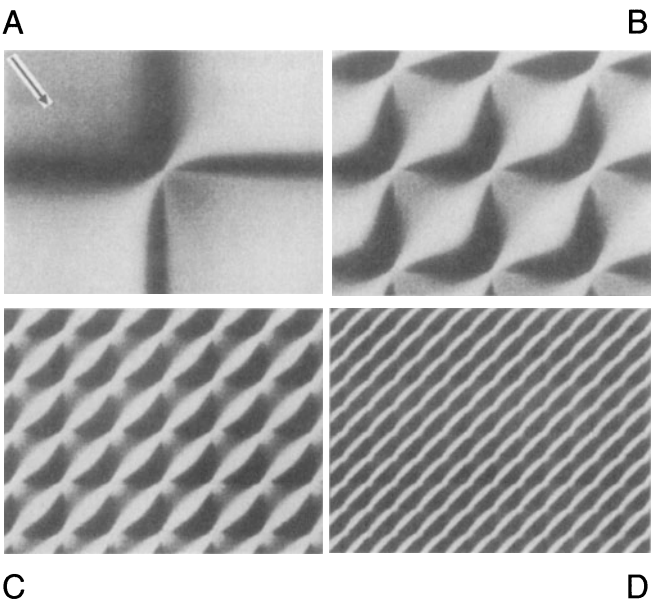


Figure 25.27. Simulated two-beam BF images of networks of screw dislocations, located in the middle of a foil with thickness equal to 4 times the extinction distance, ξ_g; $\mathbf{g}\cdot\mathbf{b} = 1$ for both dislocation types. The separation between the dislocations is: (A) ∞, (B) $1\xi_g$, (C) $0.5\xi_g$, (D) $0.25\xi_g$.

- 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 is then specified by defining the "center" of the Laue zone, giving the coordinates in terms of the **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 **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 **g**/2, i.e., at the Bragg position with the **0** and **g** beams excited. If you change the second coordinate to give, say, (0.5, 0.5), you need to 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 dislocations 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 **g·b×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 $(\mathbf{g}\cdot\mathbf{b})s$.
- 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 $\mathbf{s} > 0$, detail in the image is more localized relative to the defect than if we use the $\mathbf{s} = 0$ 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 (de Graf and Clarke 1993) or the analysis of buckling of thin specimens (Thölén and Taftø 1993). Although these are rather specialized situations, they do illustrate the growing applications of diffraction contrast in the TEM.

REFERENCES

General References

Amelinckx, S. (1964) *The Direct Observation of Dislocations,* Academic Press, New York. A fascinating summary of the early studies by TEM.

Amelinckx, S. (1979) in *Dislocations in Solids,* 2 (Ed. F.R.N. Nabarro), North-Holland, New York. If you're interested in dislocations there are many other volumes in this set.

Amelinckx, S. and Van Dyck, D. (1992) in *Electron Diffraction Techniques,* **2** (Ed. J.M. Cowley), p. 1, Oxford University Press, New York.

Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science,* Van Nostrand Reinhold, New York.

Head, A.K., Humble, P., Clarebrough, L.M., Morton, A.J., and Forwood, C.T. (1976) *Computed Electron Micrographs and Defect Identification,* North-Holland, New York.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, New York.

Hirth, J.P. and Lothe, J. (1982) *Theory of Dislocations,* 2nd edition, John Wiley & Sons, New York. The definitive textbook, but not for the beginner.

Hull, D. and Bacon, D.J. (1984) *Introduction to Dislocations,* 3rd edition, Pergamon Press, New York. A great introductory text.

Matthews, J.W., Ed. (1975) *Epitaxial Growth,* Parts A and B, Academic Press, New York.

Nabarro, F.R.N. (1987) *Theory of Dislocations,* Dover Publications, New York.

Porter, D.A. and Easterling, K.E. (1992) *Phase Transformations in Metals and Alloys,* 2nd edition, Chapman and Hall, New York.

Smallman, R.E. (1985) *Modern Physical Metallurgy,* 4th edition, Butterworth-Heinemann, Boston.

Sutton, A.P. and Balluffi, R.W. (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.

The original series of papers by P.B. Hirsch, A. Howie, M.J. Whelan, and H. Hashimoto in *Proc. Roy. Soc. London* **A252**, 499 (1960), **263**, 217 (1960), **267**, 206 (1962), and **268**, 80 (1962) are strongly recommended.

Specific References

Amelinckx, S. (1992) in *Electron Microscopy in Materials Science* (Eds. P.G. Merli and M.V. Antisari), World Scientific, River Edge, New Jersey.

Ashby, M.F. and Brown, L.M. (1963) *Phil. Mag.* **8**, 1083 and 1649.

de Graf, M. and Clarke, D.R. (1993) *Ultramicroscopy* **49**, 354.

Eshelby, J.D., Read, W.T., and Shockley, W. (1953) *Acta Metall.* **1**, 251.

Goringe, M.J. (1975) in *Electron Microscopy in Materials Science* (Eds. U. Valdrè and E. Ruedl), p. 555, Commission of the European Communities, Luxembourg.

Hirsch, P.B., Howie, A., and Whelan, M.J. (1960) *Phil. Trans. Roy. Soc.* **A252**, 499.

Howie, A. and Basinski, Z.S. (1968) *Phil. Mag.* **17**, 1039.

Howie, A. and Sworn, H. (1970) *Phil. Mag.* **31**, 861.

Hughes, D.A. and Hansen, N. (1995) *Scripta Met. Mater.* **33**, 315.

Humble, P. and Forwood, C.T. (1975) *Phil. Mag.* **31**, 1011 and 1025.

Karth, S., Krumhansl, J.A., Sethna, J.P., and Wickham, L.K. (1995) *Phys. Rev. B* **52**, 803.

Morton, A.J. and Forwood, C.T. (1973) *Cryst. Lattice Defects* **4**, 165.

Rasmussen, D.R. and Carter, C.B. (1991) *J. Electron Microsc. Technique* **18**, 429.

Steeds, J.W. (1973) *Anisotropic Elastic Theory of Dislocations,* Clarendon Press, Oxford, United Kingdom.

Takayanagi, K. (1988) *Surf. Sci.* **205**, 637.

Thölén, A.R. (1970a) *Phil. Mag.* **22**, 175–182.

Thölén, A.R. (1970b) *Phys. Stat. Sol. (a)* **2**, 537.

Thölén, A.R. and Taftø, J. (1993) *Ultramicroscopy* **48**, 27

Tunstall, W.J., Hirsch, P.B., and Steeds, J.W. (1964) *Phil. Mag.* **9**, 99.

Wilkens, M. (1978) in *Diffraction and Imaging Techniques in Material Science,* 2nd edition (Eds. S. Amelinckx, R. Gevers, and J. Van Landuyt), p. 185, North-Holland, New York.

Weak-Beam Dark-Field Microscopy

26

CHAPTER PREVIEW

The term "weak-beam microscopy" refers to the formation of a diffraction-contrast image in either BF or DF. 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 di-

rect 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 26.9 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.

Weak-Beam Dark-Field Microscopy

26

26.1. INTENSITY IN WBDF IMAGES

We showed in Chapter 11 that, in a two-beam situation, the intensity of the diffracted beam **g** in a perfect crystal can be written as

$$\left|\phi_g\right|^2 = \left(\frac{\pi t}{\xi_g}\right)^2 \cdot \frac{\sin^2(\pi t s_{eff})}{(\pi t s_{eff})^2} \qquad [26.1]$$

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 26.9. The important variables in equation 26.1 are the thickness, t, and the effective excitation error, s_{eff}, which is given by

$$s_{eff} = \sqrt{s^2 + \frac{1}{\xi_g^{\,2}}} \qquad [26.2]$$

In the WB technique we increase s to about 0.2 nm^{-1} so as to increase s_{eff}. (In most WB papers you will see this value as 2×10^{-2} $Å^{-1}$.) This large value of s means that s_{eff}, and therefore the intensity, I_g, become independent of ξ_g except as a scaling factor for t (in the prefactor in equation 26.1). 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 11, ξ_g varies with both the reflection used to form the WB image and the energy of the electrons

$$\xi_{eff} = \frac{\xi_g}{\sqrt{w^2 + 1}} \qquad [26.3]$$

A comment on equation 26.2: if $s \gg \xi_g^{-2}$ then $s \approx s_{eff}$, so that equation 26.2 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.

Practical Considerations: As the value of s_{eff} increases, equation 26.1 shows that the intensity of the G beam decreases very rapidly. The result is that the exposure time needed to record the image on a 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. The problem can be 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 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 change in the temperature of the water in the objective lens is 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.

26.2. SETTING s_g USING THE KIKUCHI PATTERN

Since the contrast in the WB image is so dependent on the value of $\mathbf{s}_g$ we need a method for determining $\mathbf{s}_g$. 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 26.1 illustrates this situation. How can we "see" what the value of n is? Of course we can't, since we are looking ap-

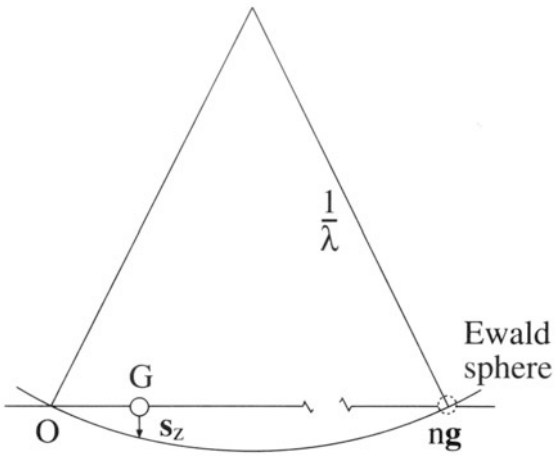


Figure 26.1. The Ewald sphere construction showing the diffraction conditions used to obtain weak-beam images. The sphere cuts the row of systematic reflections at "*n***g**" where *n* is not necessarily an integer.

proximately 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 26.2. 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 26.2, 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 26.3. Therefore, when the Ewald sphere passes through 3.2**g**, the 3**g** Kikuchi line will pass through 3.1**g**; we can express this simple geometric result as

$$n = 2m - N \qquad [26.4]$$

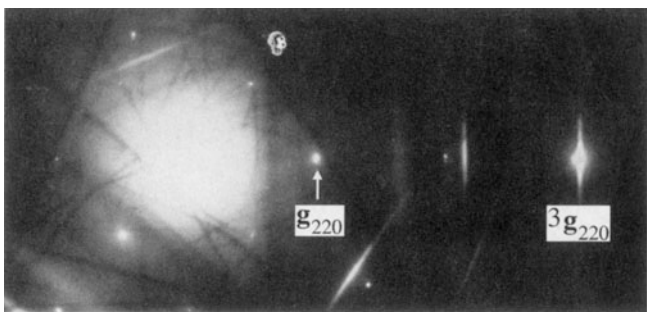


Figure 26.2. A DP obtained when the specimen is tilted to a suitable orientation for WB microscopy. Here **g** is a 220 reflection and 3**g** is strong.

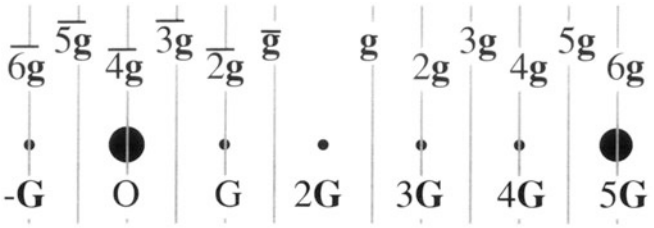


Figure 26.3. A schematic diagram showing the positions of the Kikuchi lines for the systematic row of reflections when 4**g** is excited.

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

$$\mathbf{s} = \frac{1}{2}(n-1)|\mathbf{g}|^2\lambda \qquad [26.5]$$

which you can derive from Figure 26.4 using the intersecting chord theorem (*ab* = *cd*) and the fact that 1/λ 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 |**g**| than on λ, but it depends on *both*.
- The specific nature of the material enters through **g**, the microscope affects **s** through λ. 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 26.1.

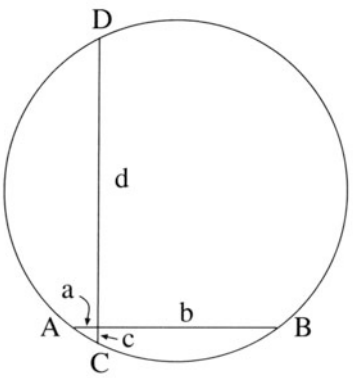


Figure 26.4. The intersecting chord construction used to deduce the value of $\mathbf{s}_g$: we approximate *d* to 2/λ, *c* is **s**, *a* is |**g**|, and *b* is (*n*-1)|**g**|.

Table 26.1. Values of S (Å^{-1}) for Various kV[a]

Si	Cu	Accelerating Voltage			
n_{Si}	n_{Cu}	100 kV	200 kV	300 kV	400 kV
4.9	2.8	0.020	0.014	0.011	0.009
6.9	3.6	0.030	0.020	0.016	0.013
8.5	4.3	0.038	0.025	0.020	0.017
9.9	4.9	0.044	0.030	0.024	0.020

[a]For Cu, a = 0.3607 nm; for Si, a = 0.534 nm. The values of n are accurate to better than 0.1. In both cases the calculation is for **g** = 220.

As an exercise, you can use equations 26.2 and 26.5 to calculate s_{eff} with ξ_g = 42 nm for the 220 reflection in Cu at 100 kV. You can then see when s_{eff} becomes "independent" of ξ_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 ≥ 0.2 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!

26.3. 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 your 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 to 30 seconds using Kodak SO-163 film. 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.

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, we generally refer to it as "**g**(3**g**) with $\mathbf{s}_{3g}$ 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 $\mathbf{s}_g$ slightly positive, to locate the defect and to focus your image. The first two steps are illustrated in Figure 26.5, 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 $\mathbf{s}_g$ is just greater than zero. Make sure that no other reflections are excited.
- 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.

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 SAD pattern.

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 diffraction condition **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

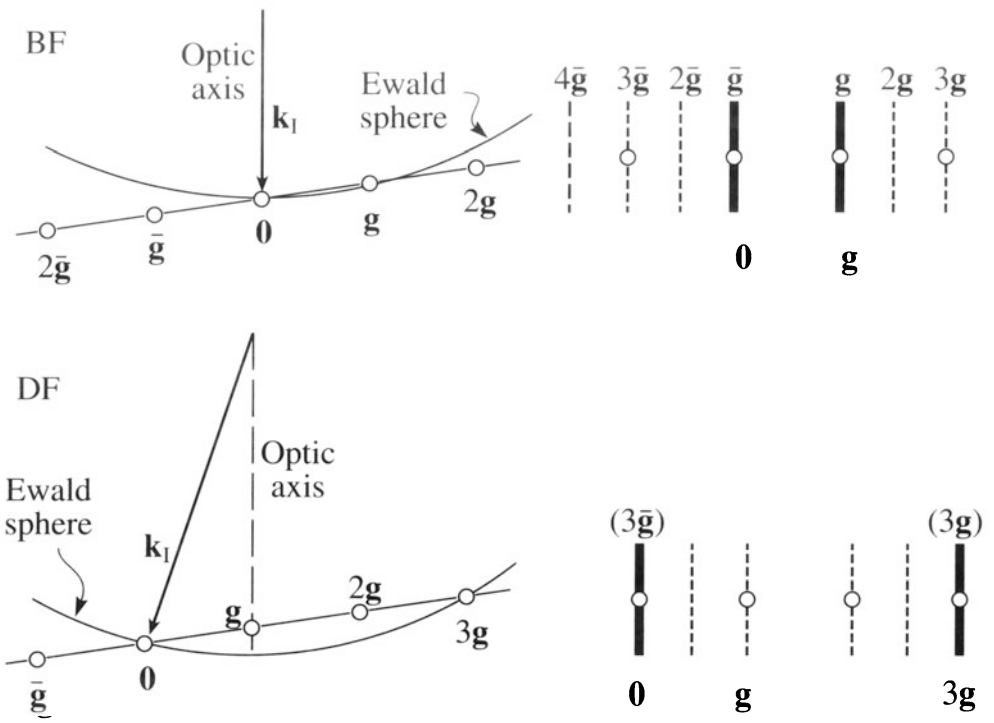


Figure 26.5. Relationship between the orientation of the Ewald sphere and the position of the Kikuchi lines for the **0**(**g**) (upper) and **g**(3**g**) (lower) diffraction conditions. The two pairs of diagrams are related by tilting the beam; the specimen has not tilted so the position of the Kikuchi lines is unchanged.

this condition useful when you need to use a **g**(*n***g**) condition with a large value of *n*.

Weak-beam microscopy becomes much easier when you are comfortable with using the TV 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 for most images. 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.

26.4. 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 26.1 we can see that the intensity minimum occurs at thicknesses of $\mathcal{N}(s_{eff}^{-1})$ with maxima at $(\mathcal{N}+ {}^1/_2)(s_{eff}^{-1})$. The effective extinction distance for s = 0.2 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 26.6.

The thickness effect is illustrated in Figure 26.7. These images were recorded with s = 0.2 nm^{-1}. The MgO specimen has been heat-treated so that there are large regions where the surface is atomically flat on both sides.

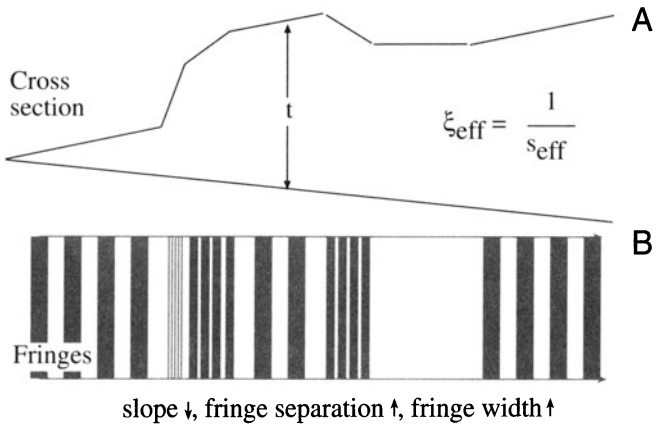


Figure 26.6. (A) In WB imaging, the thickness periodicity depends on the effective extinction distance, ξ_{eff}. (B) The separation of the fringes varies accordingly.

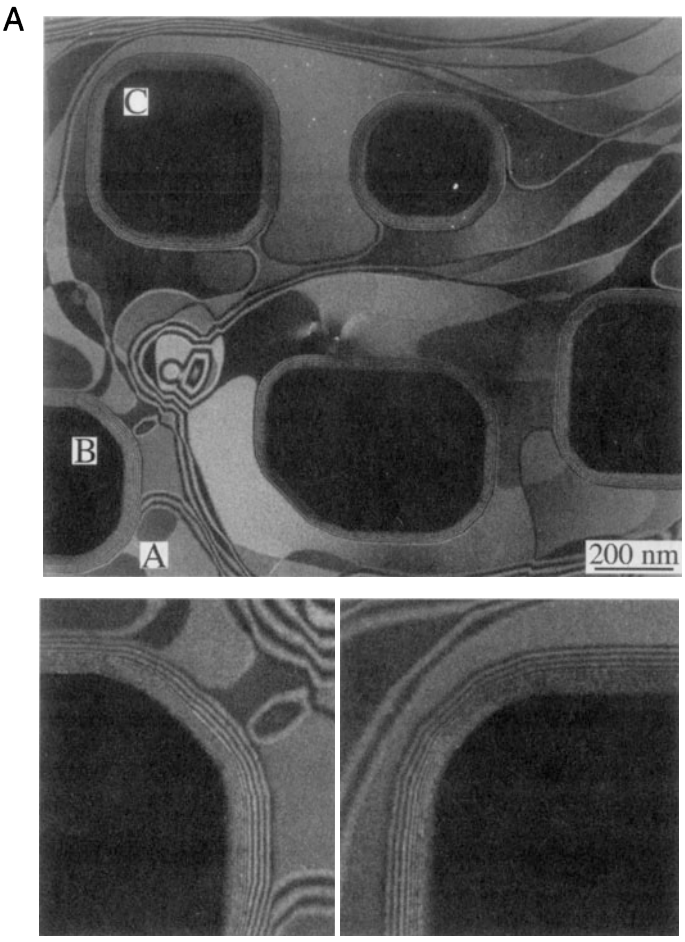


Figure 26.7. (A) WB thickness fringes from annealed MgO. (B,C) Higher magnification of regions B and C. Compare to Figure 23.3.

Before heating, the specimen had been acid etched, which caused the holes seen in this image (they're black because we're in DF). 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 increases at A. Around the holes (Figures 26.7B 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.

26.5. 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 ~2° to change $\mathbf{s}$ locally from 0.2 nm^{-1} to zero.

> 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 26.8. You can see that the WBDF image is much narrower;

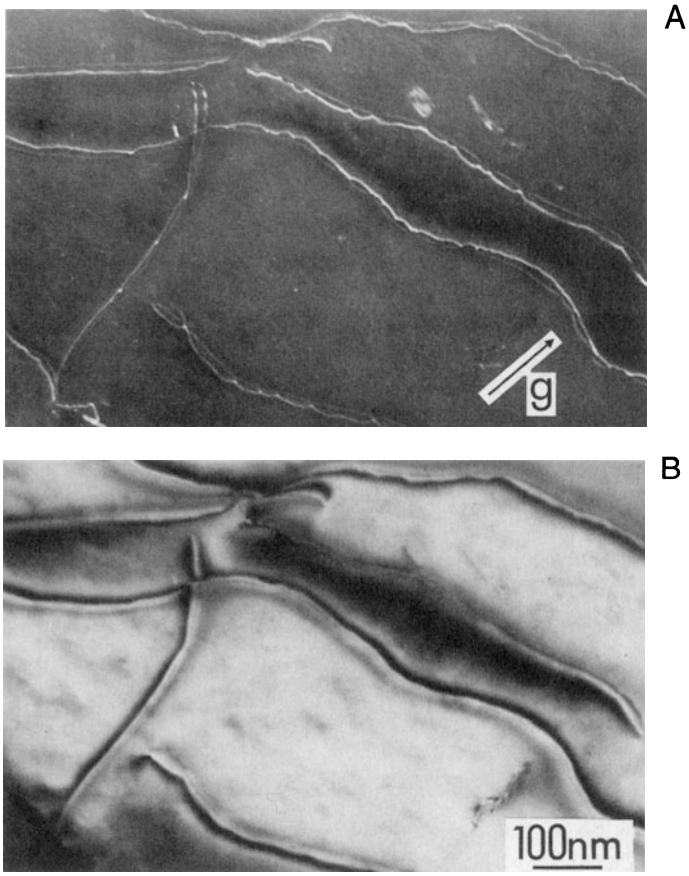


Figure 26.8. A comparison of dislocation images in a Cu alloy formed using (A) WB and (B) strong-beam ($\mathbf{s}_g > 0$) conditions.

you could make the comparison look even better if you make **s** very close to zero in the BF image.

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 26.9, 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 26.9 through 180°), but keep **g** the same, the peak again moves to the other side.

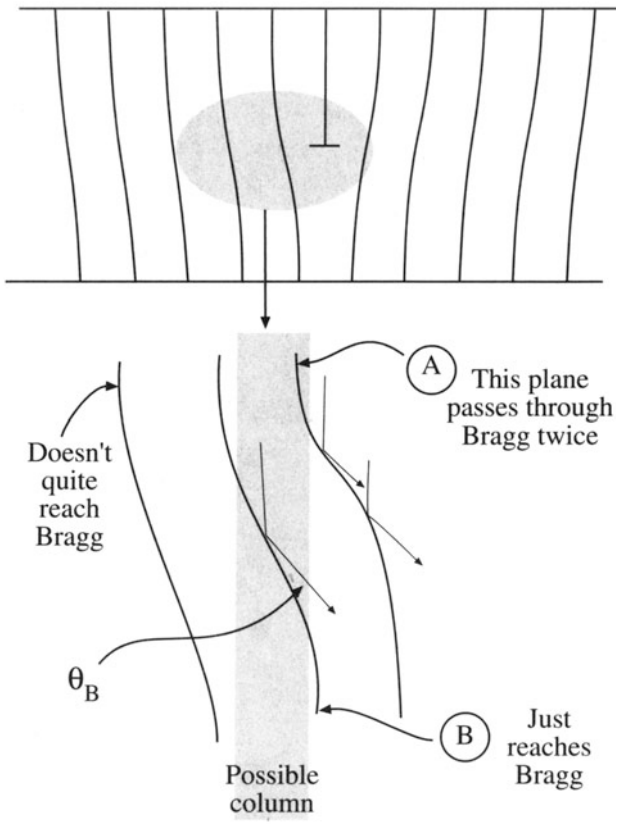


Figure 26.9. WB images of defects show high intensity close to the defect because only there are the diffracting planes bent back into the Bragg condition. This illustration is for an edge dislocation.

- If you increase **s** in the crystal, then the planes must bend more to satisfy the Bragg condition, which means the peak will move closer to the dislocation core.
- 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 strain is not large enough to compensate for the **s** you have chosen. Then you will only see poor contrast in the image.

26.6. 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 ϕ_g in the WB image occurs when s_R, which we derived in equation 25.7, is zero. We can express this result as

$$s_R = s_g + \mathbf{g} \cdot \frac{d\mathbf{R}}{dz} = 0 \qquad [26.6]$$

Equation 26.6 tells us that if the effective value of s (i.e., s_{eff}) is zero, even though s_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 (Hirsch *et al.* 1977). Therefore, the crystal can be oriented so that ϕ_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 (Cockayne *et al.* 1969, Cockayne 1972 and 1981). Therefore, the position of the WB peak should occur when equation 26.6 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 by Cockayne (1972). In the approximation where 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

$$\int_{\text{column}} e^{\left\{-2\pi i\left(s_g z + \mathbf{g}\cdot\mathbf{R}\right)\right\}} dz \qquad [26.7]$$

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 ϕ_g is large, but still kinematical (i.e., $\mathbf{s}$ is large); we want to maximize the kinematical integral in equation 26.7. We can do this using the stationary-phase method described by Stobbs (1975). We write the integral as

$$\int_0^t \exp\left(-2\pi i\left[\frac{z^2}{2}\cdot\frac{d^2}{dz^2}(\mathbf{g}\cdot\mathbf{R}) - \frac{z^3}{3}\cdot\frac{d^3}{dz^3}(\mathbf{g}\cdot\mathbf{R})\right]\right)dz \qquad [26.8]$$

where we have set $s + d/dz\ (\mathbf{g}\cdot\mathbf{R}) = 0$. If we also set $d^2/dz^2(\mathbf{g}\cdot\mathbf{R}) = 0$ (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. 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 variables, such as the convergence of the beam.

In the kinematical approximation, the half-width, Δx, of the image of an undissociated screw dislocation with $|\mathbf{g}\cdot\mathbf{b}| = 2$ is given approximately by the relation derived by Hirsch *et al.* (1960)

$$\Delta x = \frac{1}{\pi s_{\text{eff}}} \simeq \frac{\xi_{\text{eff}}}{3} \qquad [26.9]$$

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 ξ_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: ξ_g is 42 nm, and the width Δx is 14 nm. So, even if equation 26.9 is slightly wrong, the image width is greatly reduced in WB. If we make $s_g = 0.2$ 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_g is used. A series of peak profiles for different values of t is shown in Figure 26.10. Notice that although the intensity of the peak may be only

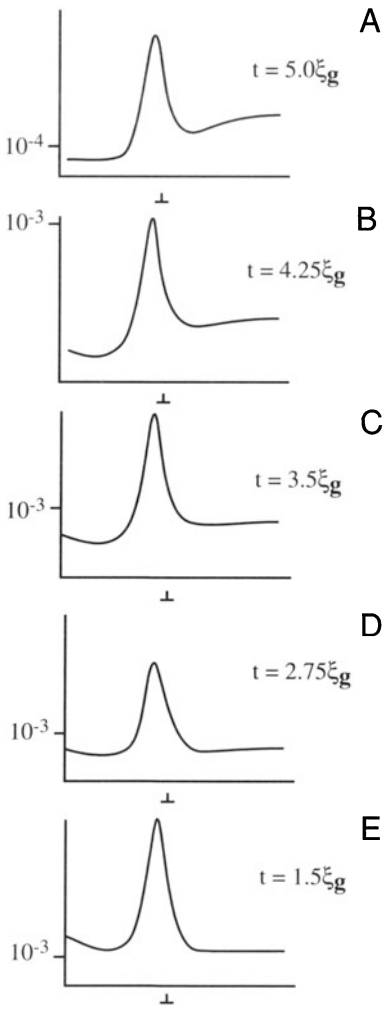


Figure 26.10. Examples of computer-calculated intensity peaks in WB images of an edge dislocation in Cu for different values of t. The intensity is relative to unit incident beam intensity.

about 0.1% of the incident beam, it is still much higher than the background.

For WB microscopy, $s_g = 0.2$ 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 26.9 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:

- $s \geq 2 \times 10^{-2}$ Å^{-1} to give sufficiently narrow peaks for fine detail to be studied.
- $s \leq 3 \times 10^{-2}$ Å^{-1} because the intensity varies as s^{-2} in the kinematical limit.
- $s\xi_g \geq 5$ to give sufficient contrast in the WB image.

If you use the **g**(3**g**) condition for Cu with **g** = 220 and 100-keV electrons, then the value of $\mathbf{s}_g$ will be 0.238 nm^{-1}.

26.7. 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 **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 basic idea is shown in Figure 26.11. 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

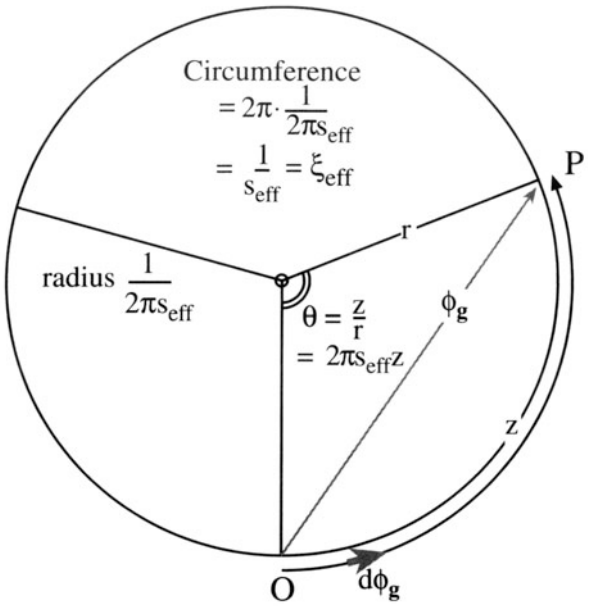


Figure 26.11. A phasor diagram for the WB case. The distance z is the arc OP measured around the circumference and the radius of the circle is $(2\pi s_{eff})^{-1}$.

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 26.12. 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 round the first circle until it meets the

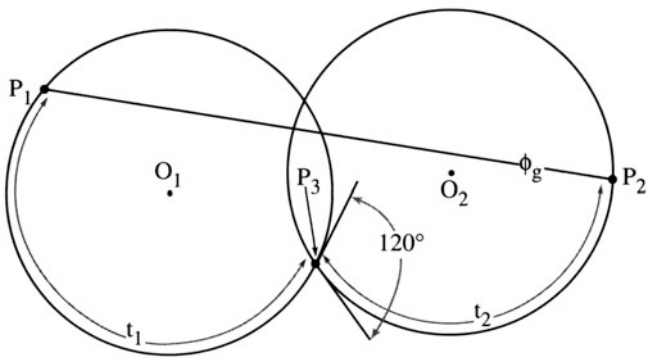


Figure 26.12. A phasor diagram used to explain the WB contrast from a SF at depth t_1 in a foil of thickness $t_1 + t_2$. P_1 is the top and P_2 the bottom of the foil. Here, the phase change at the SF (P_3) is 120°.

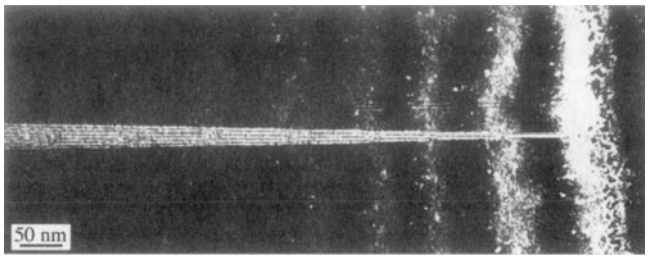


Figure 26.13. Illustration of thickness fringes in an experimental image of two inclined SFs in a wedge specimen of spinel. The thickness fringes from the wedge and from the inclined SF can be clearly seen and counted. The second SF shows little contrast because it has a different **R** value.

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 more difficult to envision if we keep t fixed but vary t_1, but the principle is the same. Images of WB fringes at inclined SFs are illustrated in Figure 26.13. 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**.

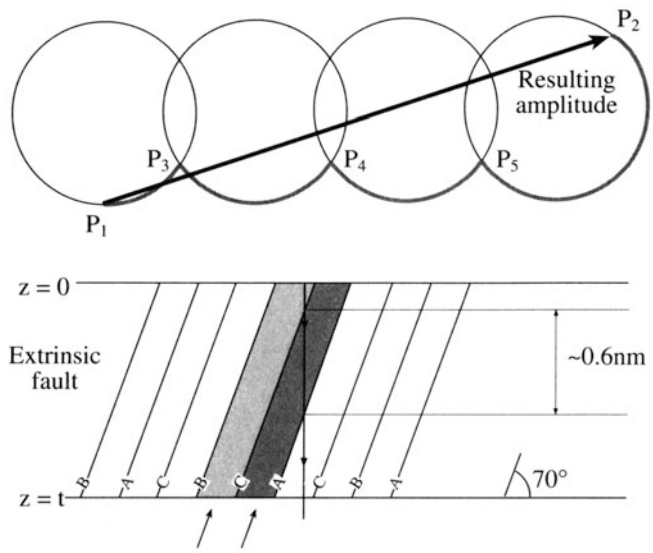


Figure 26.14. Phasor diagram for a series of overlapping planar defects at P_3, P_4, and P_5. For a 111 foil normal, the inclined $11\bar{1}$ planes lie at 70° to the surface so that the spacing between adjacent planes in the direction of the beam is 0.627 nm. Compare with Figure 26.12.

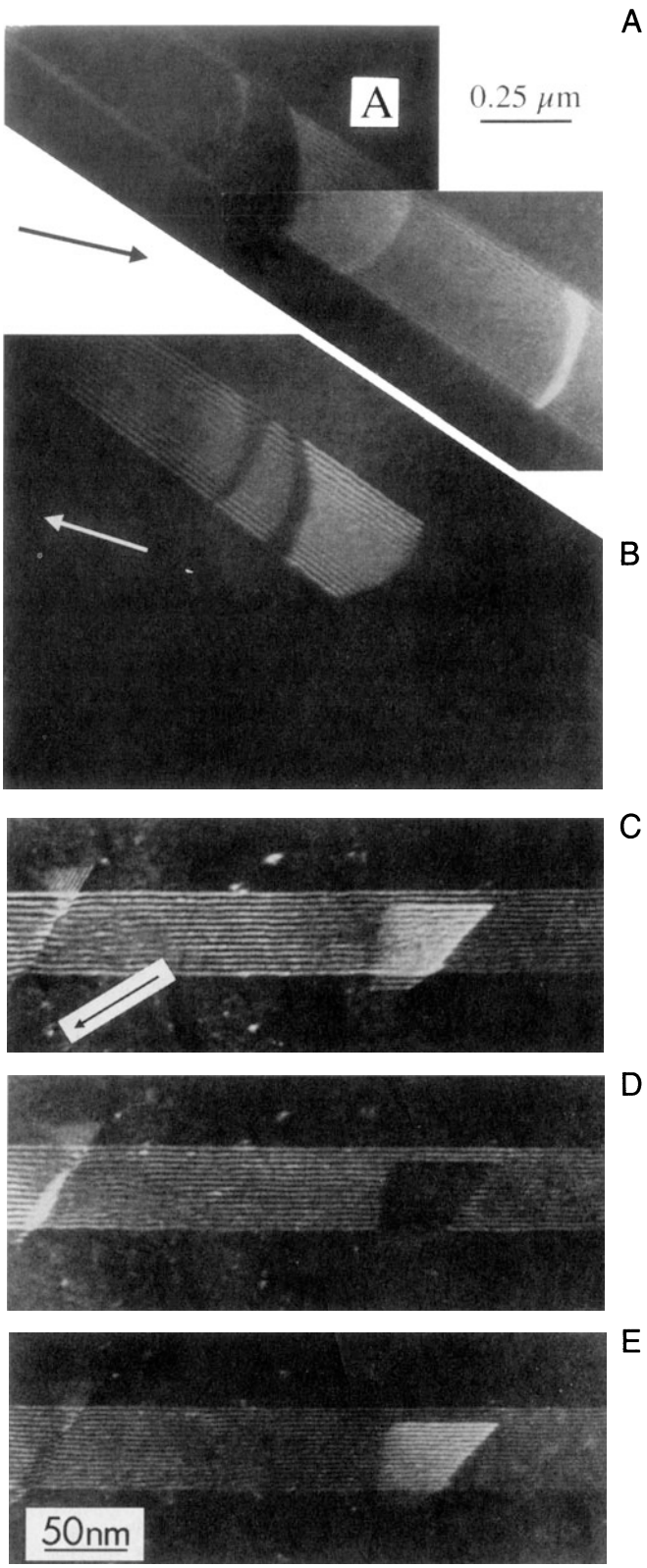


Figure 26.15. (A,B) Overlapping SFs imaged in WBDF conditions for ±**g**. Many more fringes occur than in the BF image in Figure 24.5. Fringes occur at A where the BF image showed no contrast. (C–E) Changes in fringe spacing and intensity in the overlapping region as **s** increases.

You can imagine applying this analysis to the situation where you have several overlapping planar faults, as shown in Figure 26.14. 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 26.15. Here we have several overlapping faults. The bright ones are bright even when ϕ_g reaches its minimum, as you would expect from Figure 26.14. If you compare the WB image with its strong-beam counterpart in Figure 24.5, you will notice that there is much more detail in the WB image; as we saw in Chapter 24, 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 Figures 26.15C–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 (Föll *et al.* 1980, Wilson and Cockayne 1985). This approach can be used to image other planar defects, such as the {112} twin boundary which can have a significant thickness (Carter *et al.* 1995). 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 range of thickness rather than at a particular value. As illustrated in Figure 26.16, the phase can either add or subtract depending on the sign of $\mathbf{g}\cdot\mathbf{R}$. When the phase changes quickly with a change in t, as at the center of Figure 26.16, it means that we are strongly coupling the incident and diffracted beams.

We'll summarize our discussion of phasor diagrams with 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, especially when crystal defects are present.

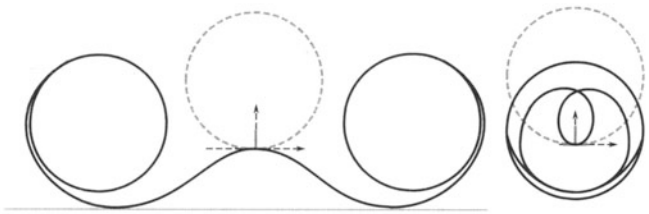


Figure 26.16. Phasor diagrams for a dislocation for ±**g**. The phase change is not abrupt but rather occurs over an extended distance along the column. In the left diagram the phase change is the direction shown in Figure 26.12, but in the right diagram the phase change has the opposite sign.

26.8. 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 26.17. We gave some general references on the theory of dislocations in Chapter 25.

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 $|\mathbf{g}\cdot\mathbf{b}_T| = 2$ images of dissociated dislocations. Hence we can estimate the stacking-fault energy (SFE) of semiconductors and several fcc metals. (See Carter 1984 for a review of dissociated dislocations and Geerthsen and Carter 1993 for a comparison of WB and HRTEM.) In order to interpret WB images of extended dislocation configurations, we 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*

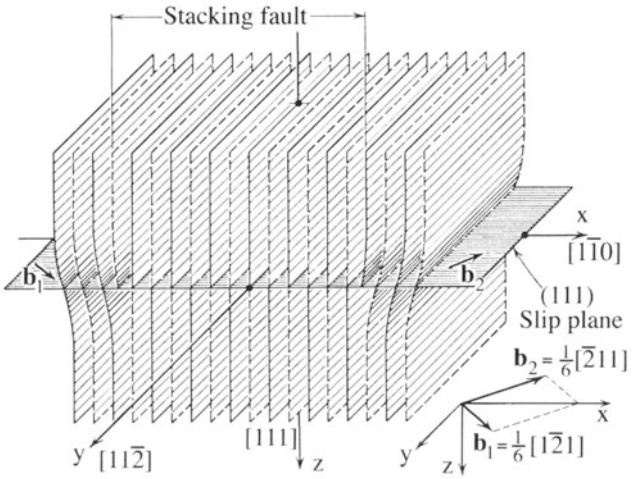


Figure 26.17. The geometry of a perfect dislocation in Cu. The perfect dislocation separates into two Shockley partial dislocations with Burgers vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ separated by a SF on a {111} plane.

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 $|\mathbf{g}\cdot\mathbf{b}_T| = 2$, so that each partial dislocation gives rise to an intensity peak having a half-width of 1 to 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 26.8 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 25, these lines are not related to the detailed structure of the dislocation but rather to n in the equation $\mathbf{g}\cdot\mathbf{b} = n$.

Ideally, for quantitative analysis, you should choose long, nearly straight dislocations. As shown in Figure 26.17, 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 $|\mathbf{g}\cdot\mathbf{b}_p| = 1$, and either no peak or a diffuse one (arising from the anisotropy of the lattice) if the partial dislocation has $|\mathbf{g}\cdot\mathbf{b}_p| = 0$. When $|\mathbf{g}\cdot\mathbf{b}_T| = 2$, 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 $|\mathbf{g}\cdot\mathbf{b}_p| = 1$).

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 $\mathbf{g}(3\mathbf{g})$.

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 $|\mathbf{g}\cdot\mathbf{b}_T| = 1$ images, when $|\mathbf{g}\cdot\mathbf{b}_p| = 1$ for one partial and $|\mathbf{g}\cdot\mathbf{b}_p| = 0$ for the other, and it can be used to identify the reflection for which $|\mathbf{g}\cdot\mathbf{b}| = 2$. Confirmation of the Burgers vector is always obtained using the BF mode, observing characteristic $|\mathbf{g}\cdot\mathbf{b}_T| = 2$ or $|\mathbf{g}\cdot\mathbf{b}_T| = 0$ images.

In a WB image with $|\mathbf{g}\cdot\mathbf{b}_T| = 2$, each of the partial dislocations will generally give rise to a peak in the image which is close to the dislocation core (Cockayne *et al.* 1969). You can calculate the approximate positions of these peaks using the criterion from equation 26.6. 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

$$-s_g = \frac{|\mathbf{g}|}{2\pi}\left[\left(|\mathbf{b}_1| + \frac{|\mathbf{b}_{1e}|}{2(1-\nu)}\right)\frac{1}{x} + \left(|\mathbf{b}_2| + \frac{|\mathbf{b}_{2e}|}{2(1-\nu)}\right)\frac{1}{x-\Delta}\right] \qquad [26.10]$$

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 $\mathbf{g}\cdot\mathbf{b}\times\mathbf{u}$ is zero. Using the notation

$$a = -s_g\left[\frac{|\mathbf{g}|}{2\pi}\left(|\mathbf{b}_1| + \frac{|\mathbf{b}_{1e}|}{2(1-\nu)}\right)\right]^{-1} \qquad [26.11]$$

and

$$b = -s_g\left[\frac{|\mathbf{g}|}{2\pi}\left(|\mathbf{b}_2| + \frac{|\mathbf{b}_{2e}|}{2(1-\nu)}\right)\right]^{-1} \qquad [26.12]$$

equation 26.10 reduces to

$$1 = \frac{1}{ax} + \frac{1}{b(x-\Delta)} \qquad [26.13]$$

which has two solutions, x_+ and x_-, given by

$$x_\pm = \frac{ab\Delta + a + b \pm \left[(ab\Delta + a + b)^2 - 4ab^2\Delta\right]^{\frac{1}{2}}}{2ab} \qquad [26.14]$$

These values of x define the positions of the peaks in the image. The separation between these peaks is then given by

$$\Delta_{obs} = \left[\Delta^2 + \frac{(a+b)^2}{a^2b^2} + \frac{2(a+b)\Delta}{ab} - \frac{4\Delta}{a}\right]^{\frac{1}{2}} \qquad [26.15]$$

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.

$$\Delta_{obs} = \left[\left(\Delta + \frac{1}{b} - \frac{1}{a}\right)^2 + \frac{4}{ab}\right]^{\frac{1}{2}} \qquad [26.16]$$

Computed images confirm that this relation is accurate for $\Delta_{obs} > 2.5$ nm, to within ±0.7 nm (Cockayne 1972). 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 (1971) have found, using anisotropic elasticity theory, that the relation (equation 26.16) subject to the ±0.7 nm uncertainty is still a good approximation.

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 26.18 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 26.19A tells you very quickly that the two nodes are different; if we form images using other $\mathbf{g}$ vectors, (B–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; $\mathbf{g}\cdot\mathbf{R}$ is zero in this image for the stacking fault. The other node is constricted, within the detectability of the WB technique. Comparison with the BF images in Figure 25.8 is instructive.

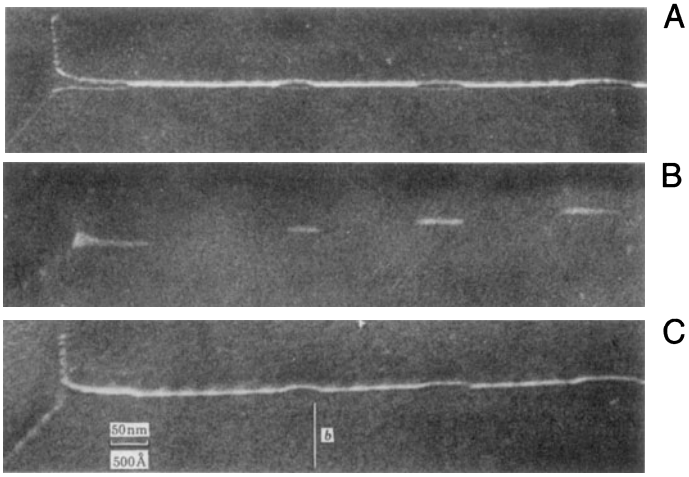


Figure 26.18. WB image of a dislocation in Si which has both dissociated and constricted segments: (A) $\mathbf{g}\cdot\mathbf{b} = 2$; both partial dislocations are visible. (B) $\mathbf{g}\cdot\mathbf{b}_T = 0$ showing SF contrast. (C) $\mathbf{g}\cdot\mathbf{b} = 1$; only one partial dislocation is visible.

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 dipole as you can see in Figure 26.20. 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 $\mathbf{g}\cdot\mathbf{b}$ 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 (just more "equiaxed"). The images shown in Figures 26.20B,C illustrate the dramatic change in contrast which you can see on reversing $\mathbf{g}$. Some of the dipoles completely disappear in (A–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 the lattice is thus only distorted over very small distances. When we use the WB technique, we are probing the structure on these very small dimensions and the contrast can be high. Again, compare with the BF image in Figure 25.13.

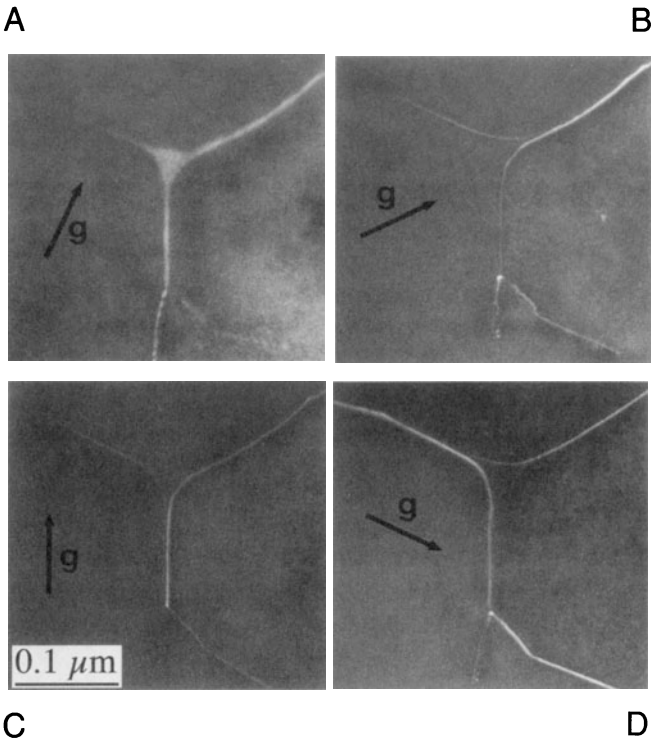


Figure 26.19. WB images of a pair of dislocation nodes formed by the dissociation of interacting dislocations lying on a {111} plane in a Cu alloy. The SF is imaged in (A) and, imaging with different reflections (B–D), the partial dislocations are out of contrast when $\mathbf{g}\cdot\mathbf{b} = 0$. Compare with the BF image in Figure 25.8 and note the difference in magnification.

Examples 4 and 5. The WB technique allows us to see features which would be hidden if we used strong-beam imaging. Figure 26.21 gives an 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 this WB image. Seeing small particles close to dislocations is difficult in strong-beam imaging. Although not easy in WB, Figure 26.22 does illustrate that it can be done. These images show, for example, that the behavior of the dissociated dislocation is different on each side of the particle (see also Figure 26.23), and again stress the advantage of WB over BF imaging.

Example 6. We noted in Section 25.8 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, surface effects can be even more important. Figure 26.23 shows an example where this effect is particularly clear. Dislocations which were uniformly separated in the bulk material now appear wedge-shaped: the effects of the two surfaces are different in this case (Hazzledine *et al.* 1975).

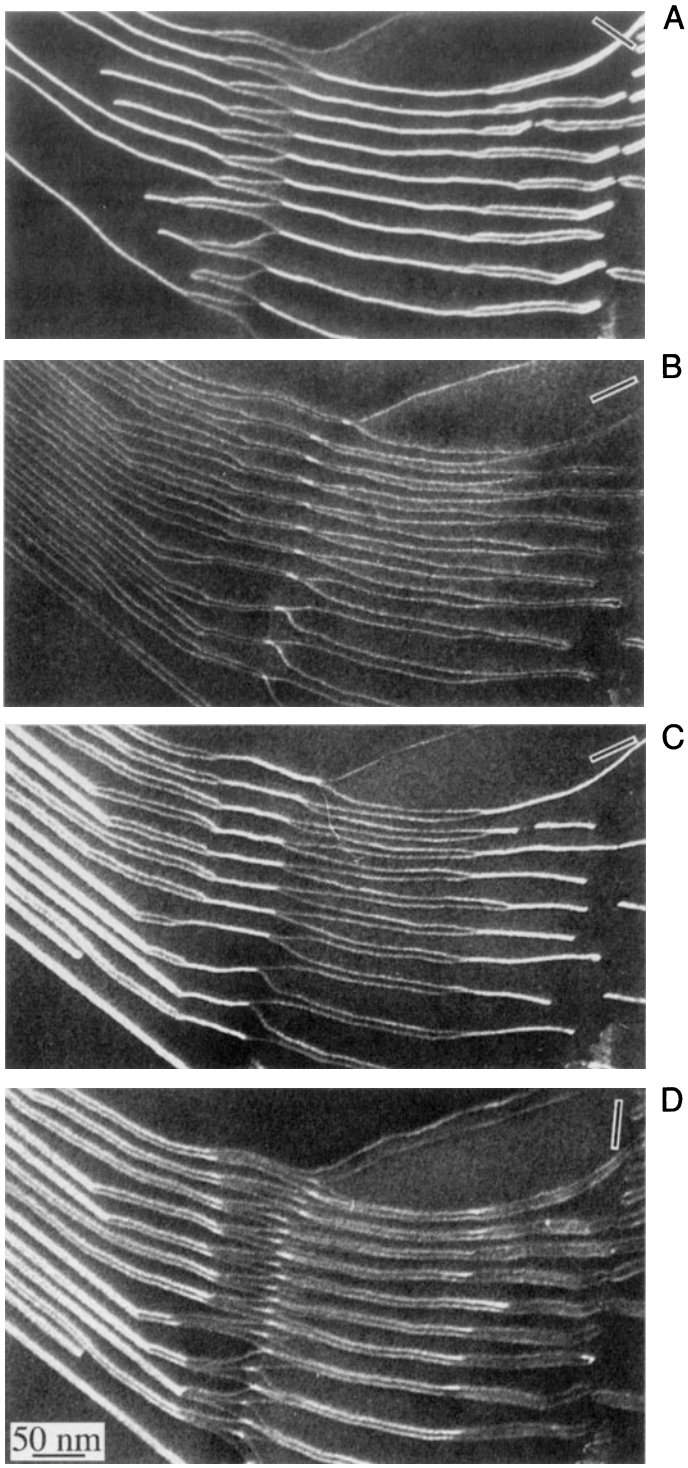


Figure 26.20. Four WBDF images showing an array of dislocation dipoles in a Cu alloy having a low stacking-fault energy. The reflections are all 220-type and the dislocations all lie on (111) planes which are nearly parallel to the surface of the specimen. All the dislocations are dissociated. All the 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.

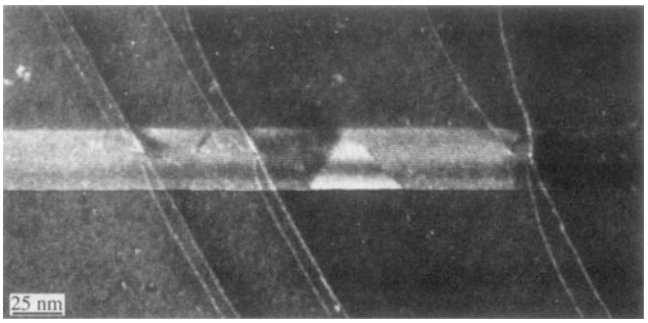


Figure 26.21. WB image of an inclined SF cutting through a series of dissociated dislocations lying parallel to the surface of the Cu-alloy specimen.

26.9. OTHER THOUGHTS

26.9.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 26.24. 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, we will achieve 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.

26.9.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 usually interested in defects. However, we can make some basic comments. For the reflection $\mathbf{g}$ to give a WB image, $|\phi_g|$ must be much smaller than unity in the regions of perfect crystal but, in strained regions, a change

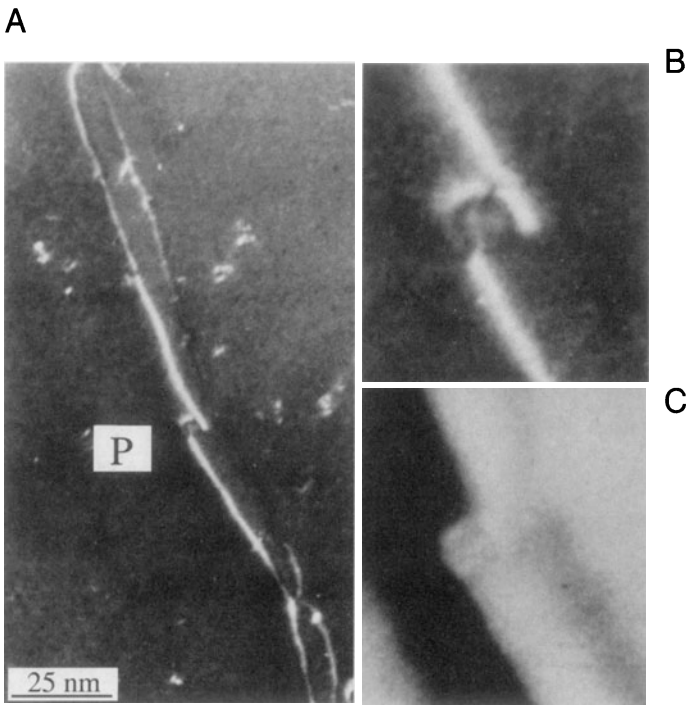


Figure 26.22. (A) WB image of a dissociated dislocation interacting with a particle (P) in a Cu alloy. (B) and (C) are enlargements of the WB and corresponding BF images, respectively.

$\Delta\psi^{(j)}$ in the amplitude of the Bloch wave j can give rise to a change $\Delta\phi_g$. 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 $C_g^{(j)}$, i.e., from the Bloch wave with the largest amplitude to the one which is most strongly excited. The dispersion surface for the $\mathbf{g}(3\mathbf{g})$ diffraction geometry is shown in Figure 26.25. 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., $\mathbf{0}(2\mathbf{g})$.

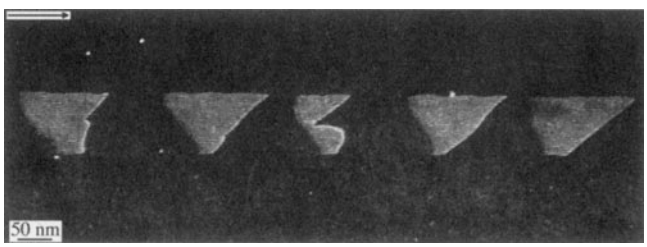


Figure 26.23. WB image showing the dissociation of a group of dislocations which are inclined to the foil surface to give wedge-shaped SFs. The shape of the SFs is caused by surface stresses.

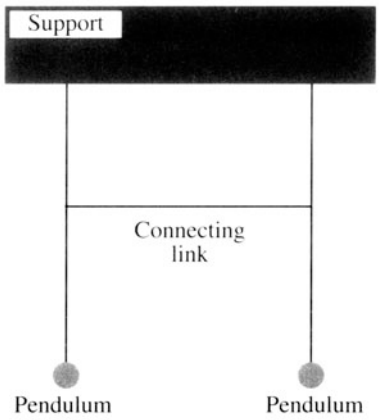


Figure 26.24. The coupled pendulum.

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 ($= s\xi_g$) 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 nm^{-1}. For example, for a $\{2\bar{2}0\}$ reflection in copper with 100-keV electrons, w is automatically greater than 8 since ξ_g is 42 nm.

The **g**(3.1**g**) condition may also be preferable to **g**($\bar{\mathbf{g}}$) 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.

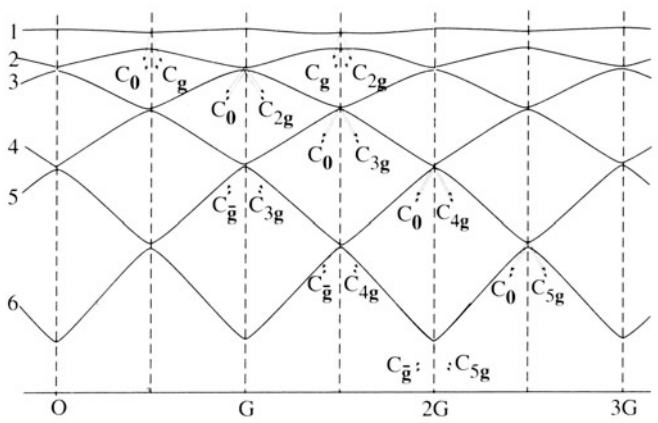


Figure 26.25. Dispersion surface construction which is used to describe the **g**(3**g**) geometry. The BZB is at 1.5G and tells you which reflections are strongly coupled; see Figure 15.9.

26.9.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 11. When you are using WB conditions, you must be even more careful. Consider the **g**,3**g** geometry shown in Figure 26.5. 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 13. The coupling of beams **g** and **h** is determined by $(s_g - s_h)^{-1}$ and has an extinction distance given by ξ_{g-h}. If s_g and s_h are equal, then the coupling between these beams will be strong. Furthermore, the characteristic length for the coupling in this example will be ξ_{2g-g} or ξ_g, which is what you would have guessed from Figure 26.25.

26.9.D. The Future

Several new developments will change 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.
- An FEG and energy-filtered imaging will allow us either to minimize the effect of variations in the energy or to form WB images using particular sections of the energy-loss spectrum. Then, we will need to extend the theory.
- Image processing and frame averaging should allow us to reduce the noise and again aid quantification.

Image simulation, as we described in Chapter 25, will allow us to be more quantitative in our interpretation of WB images.

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 the lattice parameter of the crystal and the wavelength of the electrons. You will see the "magic number" **s** = 0.2 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 24.13 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.

REFERENCES

Specific References

Carter, C.B. (1984) in *Dislocations–1984* (Eds. P. Veyssière, L. Kubin, and J. Castaing), p. 227, Editions du CNRS, Paris.

Carter, C.B., Mills, M.J., Medlin, D.L., and Angelo, J.E. (1995) 7th International Conference on Intergranular and Interphase Boundaries in Materials, Lisbon, Portugal.

Cockayne, D.J.H. (1972) *Z. für Naturforschung.* **27a**, 452.

Cockayne, D.J.H. (1981) *Ann. Rev. Mater. Sci.* **11**, 75.

Cockayne, D.J.H., Ray, I.L.F., and Whelan, M.J. (1969) *Phil. Mag.* **20**, 1265.

Föll, H., Carter, C.B., and Wilkens, M. (1980) *Phys. stat. sol. (A)* **58**, 393.

Geerthsen, D. and Carter, C.B. (1993) *Phys. stat. sol. (A)* **136**, 29.

Hazzledine, P.M., Karnthaler, H.P., and Wintner, E. (1975) *Phil. Mag.* **32**, 81.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, p. 164, Krieger, Huntington, New York.

Hirsch, P.B., Howie, A., and Whelan, M.J. (1960) *Proc. Roy. Soc. London* **A252**, 499.

Stobbs, W.M. (1975) in *Electron Microscopy in Materials Science,* Vol. **II** (Eds. U. Valdrè and E. Ruedl), p. 593, CEC, Brussels.

Stobbs, W.M. and Sworn, C.H. (1971) *Phil. Mag.* **24**, 1365.

Wilson, A.R. and Cockayne, D.J.H. (1985) *Phil. Mag.* **A51**, 341.

Phase-Contrast Images

27

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, as we saw in Chapters 23 and 24, even thickness fringes and stacking-fault fringes 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 that which 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 considering the limitations of such packages. We will begin this chapter by discussing some simple approaches to understanding phase-contrast effects as they relate to lattice-fringe imaging.

Phase-Contrast Images

27

27.1. 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 necessary to obtain such images can be acquired with practice by most TEM users.

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.

27.2. 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

$$\psi = \varphi_0(z) \exp 2\pi i(\mathbf{k}_I \cdot \mathbf{r}) + \varphi_g(z) \exp(2\pi i \mathbf{k}_D \cdot \mathbf{r}) \quad [27.1]$$

where we know

$$\mathbf{k}_D = \mathbf{k}_I + \mathbf{g} + \mathbf{s}_g = \mathbf{k}_I + \mathbf{g}' \quad [27.2]$$

We are thus using a two-beam approximation but allowing $\mathbf{s}_g$ to be nonzero. Now we will make some simple substitutions, setting $\varphi_0(z) = A$ and take $e^{2\pi i \mathbf{k}_I \cdot \mathbf{r}}$ out as a factor. We will also represent the expression for φ_g from equation 13.5 as

$$\varphi_g = B \exp i\,\delta \quad [27.3]$$

where

$$B = \frac{\pi}{\xi_g} \frac{\sin \pi t s_{\text{eff}}}{\pi s_{\text{eff}}} \quad [27.4]$$

and

$$\delta = \frac{\pi}{2} - \pi t s_{\text{eff}} \quad [27.5]$$

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} with s. Thus equation 27.1 becomes

$$\psi = \exp(2\pi i \mathbf{k}_I \cdot \mathbf{r})[A + B \exp i(2\pi \mathbf{g}' \cdot \mathbf{r} + \delta)] \quad [27.6]$$

The intensity can then be expressed as

$$I = A^2 + B^2 + AB[\exp i(2\pi \mathbf{g}' \cdot \mathbf{r} + \delta) + \exp -i(2\pi \mathbf{g}' \cdot \mathbf{r} + \delta)] \quad [27.7]$$

$$I = A^2 + B^2 + 2AB\cos\left(2\pi\mathbf{g}'\cdot\mathbf{r} + \delta\right) \qquad [27.8]$$

Now $\mathbf{g}'$ is effectively perpendicular to the beam so we'll set it parallel to x and replace δ, giving

$$I = A^2 + B^2 - 2AB\sin\left(2\pi\mathbf{g}'x - \pi st\right) \qquad [27.9]$$

Therefore, the intensity is a sinusoidal oscillation normal to $\mathbf{g}'$, with a periodicity that depends on s and t (just like thickness fringes). (Note that g and s are not bold in equation 27.9 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 g'.

This model will be equally valid even if the incident beam is tilted slightly off the optic axis.

27.3. SOME PRACTICAL ASPECTS OF LATTICE FRINGES

27.3.A. If s = 0

If we just have **0** and **g** in the objective aperture and we then set **s** = 0 for reflection G (so $\mathbf{g}' = \mathbf{g}$), we will see fringes in the image (Figure 27.1A) which have a periodicity in the x direction of $1/g$, i.e., the spacing of the planes which give rise to **g**. This result holds wherever **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.

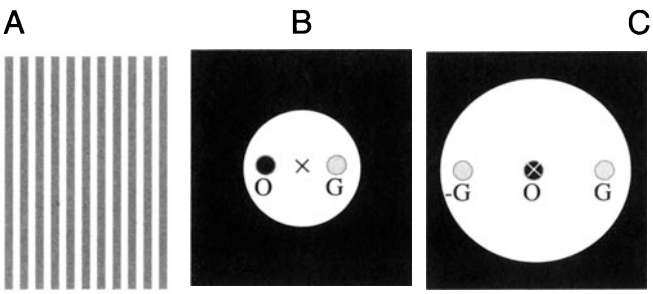


Figure 27.1. (A) Schematic tilted-beam 111 lattice fringes in Si formed using the O and G beams symmetrically displaced relative to the optic axis; **g** is normal to the fringes. (B) Ideal diffraction geometry to produce tilted-beam fringes. (C) On-axis three-beam geometry.

Figure 27.1B shows the ideal geometry for producing images like Figure 27.1A. 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 27.1B we have **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 27.1C, the planes are viewed edge on, but **s** ≠ 0 for reflection G; so we must also consider reflection –G.

27.3.B. If s ≠ 0

If the specimen is not exactly flat, then **s** will vary across the image; even if you set **s** = 0 in the DP, it will not be 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 ideally **s** is not zero for any of the beams; **s** may also vary from beam to beam.

27.4. 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 27.2, which extends Figure 27.1 to the many-beam case. In general, this array of spots bears no direct relationship to the *position* of atoms in the crystal.

The trouble is that the fringes look so like atomic planes that you can be easily misled into thinking that they are atomic planes.

We'll see more on this when we discuss image simulation in Chapter 29. In case you are in doubt, compare the beautiful image shown in Figure 27.3A with the projected structure of Si in Figure 27.3B. 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 27.3C. The diffi-

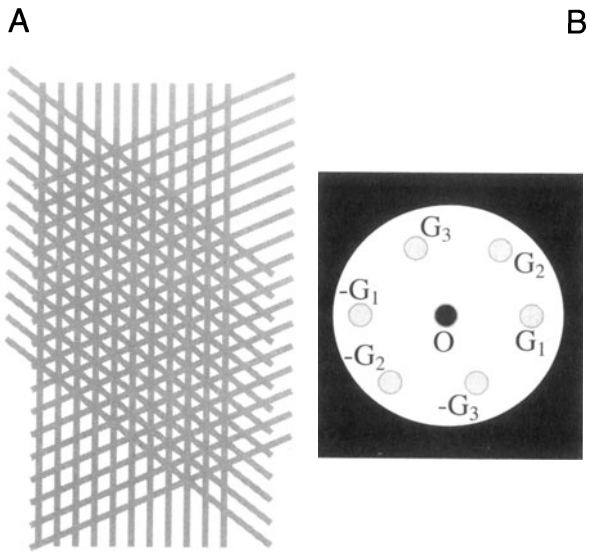


Figure 27.2. (A) Schematic many-beam image showing crossing lattice fringes and (B) the diffraction pattern.

culty is that, in the image, the spots in the dumbbell are really just 1.3 Å apart but the point resolution of the TEM was only ~2.5 Å. You can see from the 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 (1980); the dumbbells in the image are caused by the crossing {113} fringes.

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 details in the image close to the defect give you a true picture of the location of the atoms close to the defect? The answer is of course "no."

So lattice fringes are *not* direct images of the structure, but just give you lattice spacing information.

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 27.4 illustrates some typical applications of the phase-contrast imaging mode, where we can learn a lot about our material by intuitive interpretation without the need for simulating their images.

Figure 27.4A shows interfaces between a spinel particle and an olivine matrix; Figure 27.4B shows how we

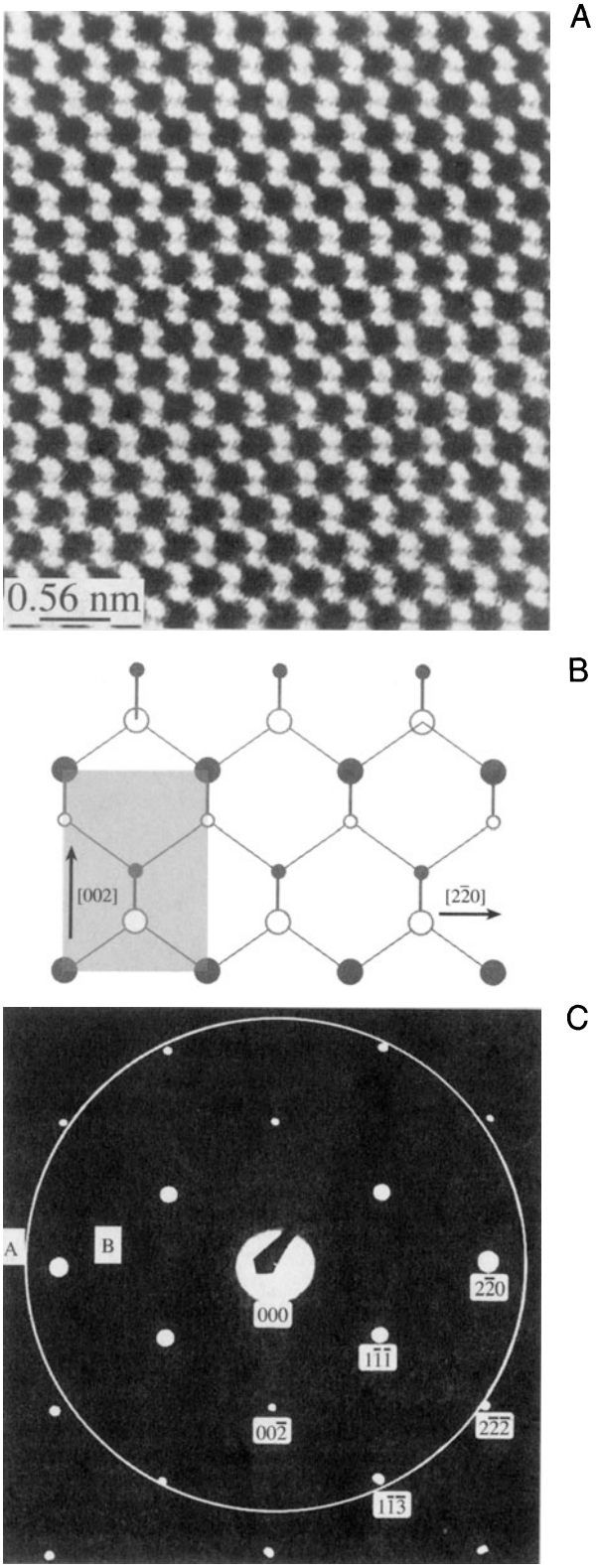


Figure 27.3. (A) On-axis image of a perfect Si crystal; (B) the projected structure; (C) the diffraction pattern showing the 13 spots used to form the image inside the aperture (ring). The Si dumbbells do not correspond to the closely spaced pairs of spots in the image.

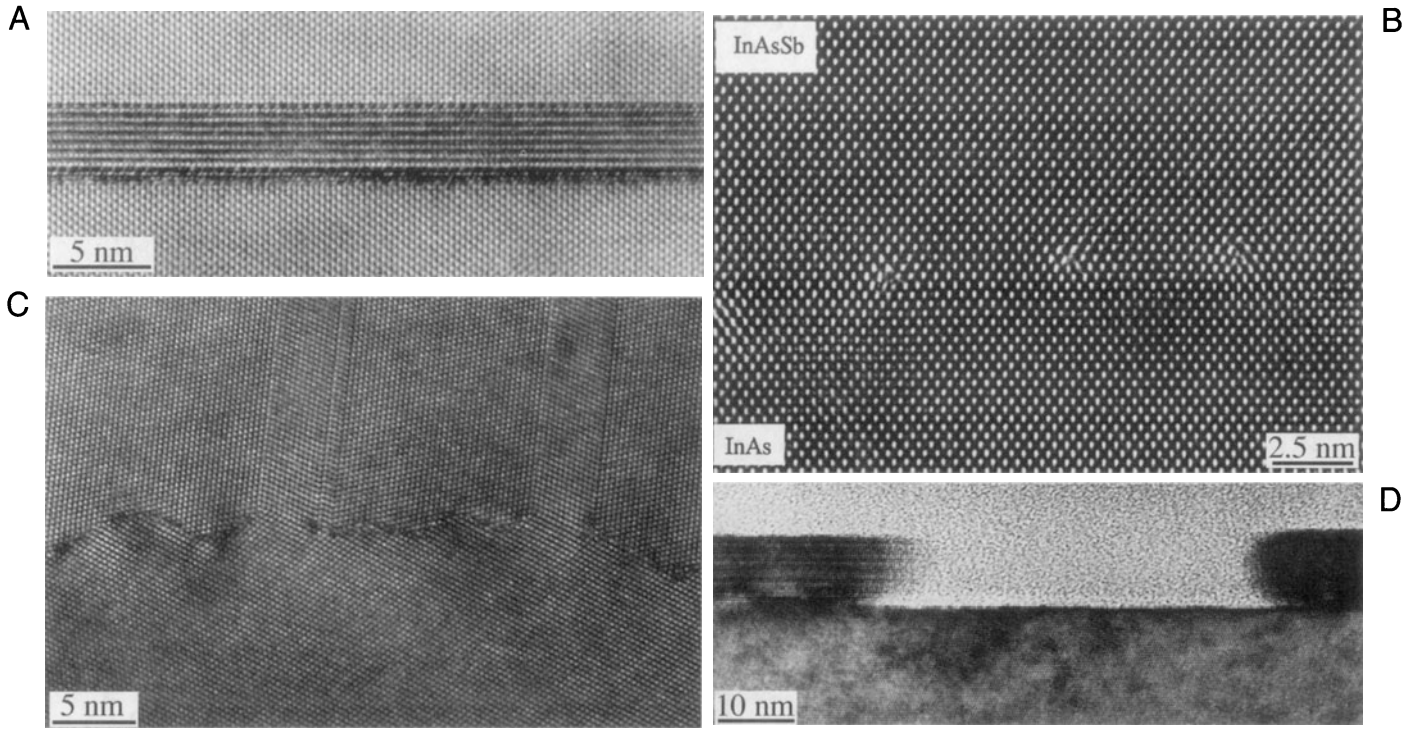


Figure 27.4. Illustrations of lattice images which contain easily interpreted information. (A) The spinel/olivine interface; (B) dislocations at a heterojunction between InAsSb and InAs; (C) a grain boundary in Ge faceting on an atomic scale; (D) a profile view of a faceted surface.

can locate dislocations at a heterojunction; Figure 27.4C shows the atomic-scale faceting of a grain boundary in Ge, and Figure 27.4D illustrates the faceting of a surface.

27.5. MOIRÉ PATTERNS

Moiré (pronounced "mwa-ray") patterns can be formed by interfering two sets of lines which 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:

- Take two misfit sets and align them exactly. This gives you a set of moiré fringes which are parallel to the lines forming them, as shown in Figure 27.5A.
- Take two identical sets of lines and rotate them. Now, you produce a set of moiré fringes which is perpendicular to the average direction of the initial lines (Figure 27.5B).
- Take the first two sets and rotate them so you produce moiré fringes, as in Figure 27.5C; 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

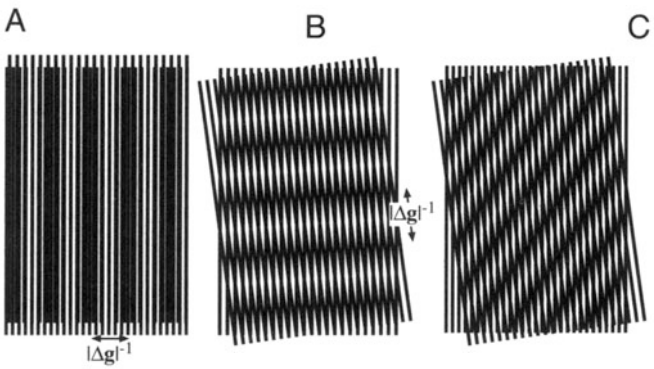


Figure 27.5. (A) Translational moiré fringes; (B) rotational moiré fringes; (C) mixed moiré fringes; note the relationship between the fringes and their constituent lattices.

lines themselves. In particular, if the sets of lines in Figure 27.5 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.

27.5.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

$$\mathbf{g}_{tm} = \mathbf{g}_2 - \mathbf{g}_1 \qquad [27.10]$$

As shown in Figure 27.6A (which is from fcc Ni on fcc NiO), 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

$$d_{tm} = \frac{1}{g_{tm}} = \frac{1}{g_2 - g_1} = \frac{\frac{1}{g_2}\cdot\frac{1}{g_1}}{\frac{1}{g_1} - \frac{1}{g_2}} = \frac{d_2 d_1}{d_1 - d_2} = \frac{d_2}{1 - \frac{d_2}{d_1}} \qquad [27.11]$$

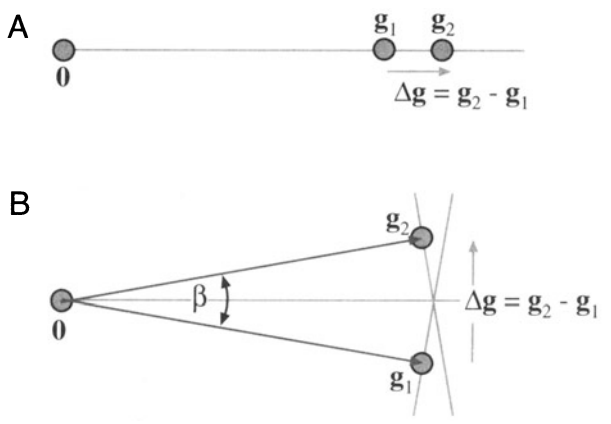


Figure 27.6. (A) Relationship between **g**-vectors for translational moiré fringes; (B) relationship for rotational moiré fringes.

27.5.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 $2g \sin \beta/2$, as shown in Figure 27.6B. The fringe spacing is then

$$d_{rm} = \frac{1}{g_{rm}} = \frac{1}{2g \sin \beta/2} = \frac{d}{2\sin \beta/2} \qquad [27.12]$$

27.5.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

$$d_{gm} = \frac{d_1 d_2}{\left((d_1 - d_2)^2 + d_1 d_2 \beta^2\right)^{\frac{1}{2}}} \qquad [27.13]$$

27.6. 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 (1956) to identify a dislocation before lattice imaging was possible. Later, they were regarded as an imaging artifact which 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 27.7A. This process is another example of double diffraction as discussed in Section 18.7. Figure 27.7A is from a pair of perfectly aligned but misfitting cubic crystals viewed along their common [001] zone axis; the pattern is indexed in Figure 27.7B. 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.

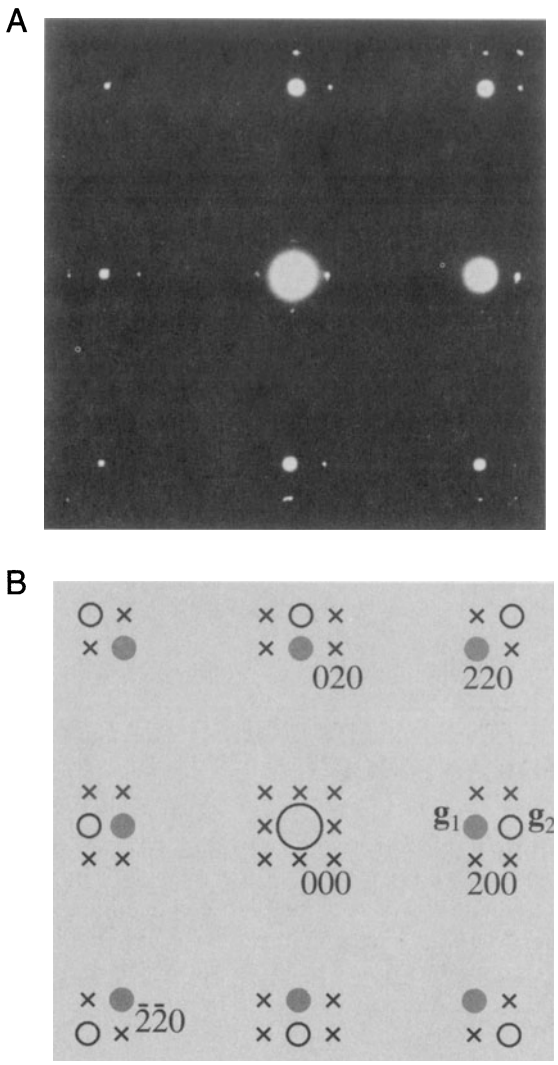


Figure 27.7. (A) Experimental diffraction pattern from perfectly aligned Ni and NiO. Brighter spots are from NiO, which has the larger lattice parameter. (B) Schematic which explains translational moiré fringes. Closed circles ($\mathbf{g}_1$) correspond to crystal 1, open circles ($\mathbf{g}_2$) to crystal 2 and × to double diffraction of $\mathbf{g}_1$ beams by crystal 2. Only × reflections close to $\mathbf{g}_1$ and $\mathbf{g}_2$ have appreciable intensity.

27.6.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 a (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.

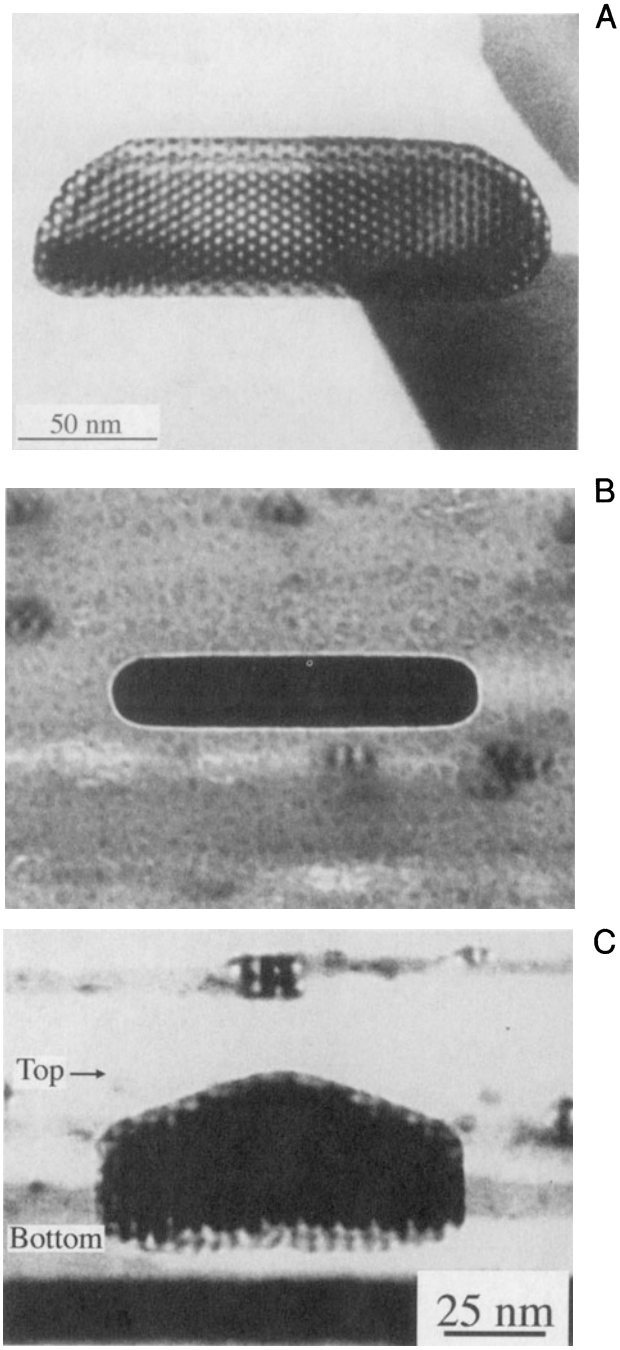


Figure 27.8. (A) The appearance of moiré fringes depends on the thickness of the specimen, as you can see where the edges of this island are inclined relative to the surface of the substrate. (B) The particle is too thick to show moiré fringes when edge on. (C) When this thick particle is tilted over, moiré fringes are seen at both the top and bottom.

Tilting the specimen can also give us information about misfitting islands, as illustrated in Figure 27.8. 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 Figures 27.8B 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.

We know that this region is not in contact with the substrate, which reminds us that these fringes do not tell us about the interface structure!

27.6.B. Rotational Moiré Fringes

We often see rotational moiré fringes at twist boundaries, as illustrated for Si in Figure 27.9. 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.

27.6.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 patterns can be used to locate and give information on dislocations which are present in one material but not the other. We can form an image which contains information about the dislocation if it is associated with a terminating lattice plane in one material, but we don't actually "see" the dislocation.

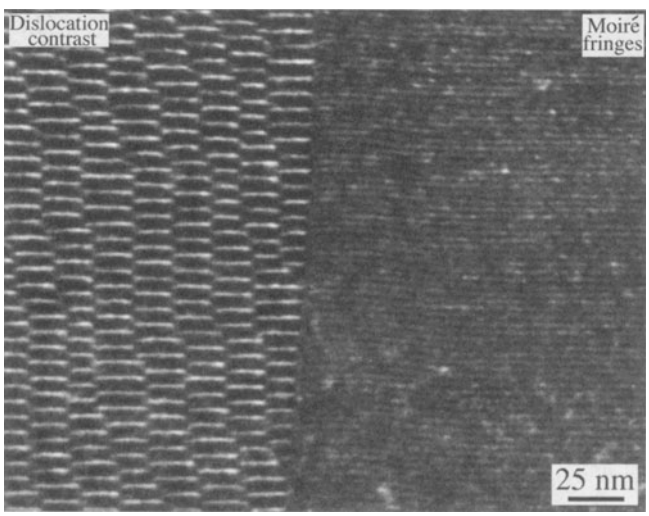


Figure 27.9. WBDF image of moiré fringes at a grain boundary showing very different contrast than the region containing dislocations.

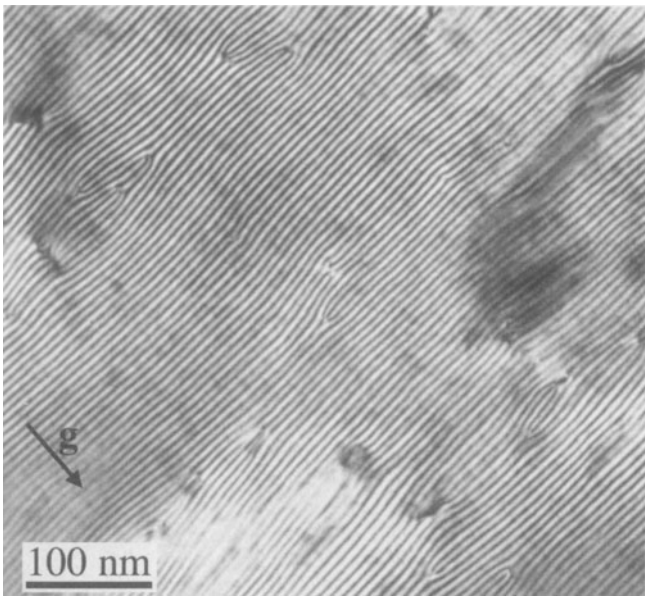


Figure 27.10. Moiré fringes reveal the presence of dislocations in a thin film of CoGa grown on a GaAs substrate. The (001) interface lies parallel to the specimen surface. Although the images contain much detail, most of it cannot readily be related to the structure of the defects.

This effect is illustrated in Figure 27.10; the image appears as a magnified view of the projection of the dislocation. This result can be deceptive, as you can see in Figure 27.11, where we have rotated the perfect grain slightly and changed its 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 (1969), 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 27.12.

Since the spacing of moiré fringes essentially gives a magnified view of the misfit between aligned particles and a substrate, we can use them to measure the strain in

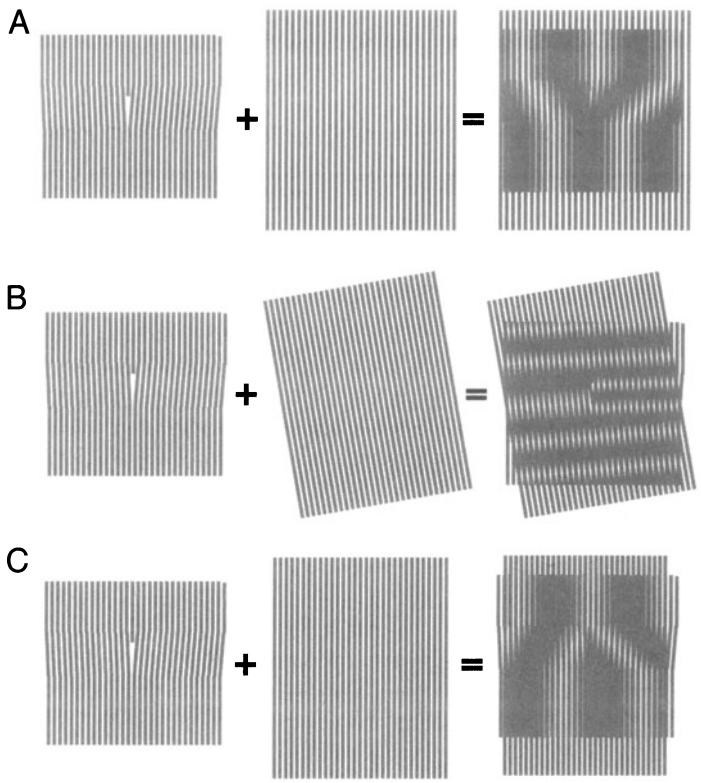


Figure 27.11. Schematic diagrams showing why moiré patterns from regions containing dislocations cannot be readily interpreted: (A) a dislocation image formed by interference between a regular lattice and one containing an extra half plane. (B) In comparison with (A), small rotation of the lattice of 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.

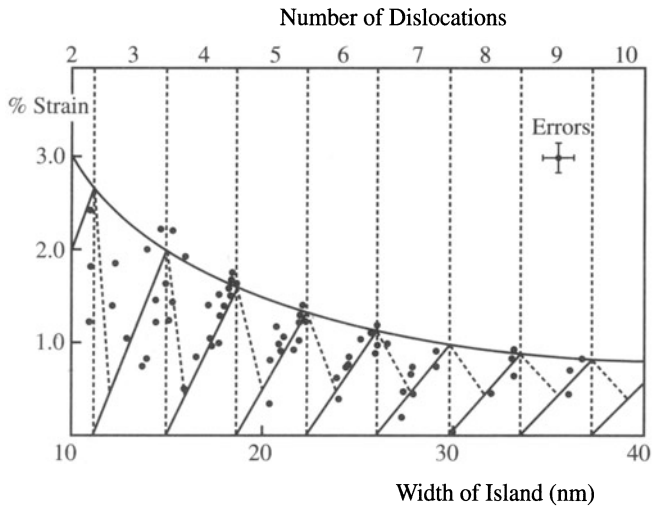


Figure 27.12. Moiré-fringe spacings can be used to monitor the change in lattice parameter as small islands of Sn grow in size on a thin film of SnTe. This plot shows how the strain (measured from the moiré-fringe spacing) can be related to the width of the misfitting island and then to the number of dislocations in the interface.

such particles. In its simplest form, in one dimension, the strain is given by

$$\varepsilon = \frac{a_1 - a_0}{a_0} \qquad [27.14]$$

where a_1 and a_0 are the lattice spacings of the particle and substrate, respectively. You may need to modify this equation if the alignment is not simple cube-on-cube.

27.6.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 27.13 for YBCO grains rotated 45° on a MgO substrate (Norton and Carter 1995). 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 which is much coarser than the original pattern and is very sensitive to differences in lattice spacing

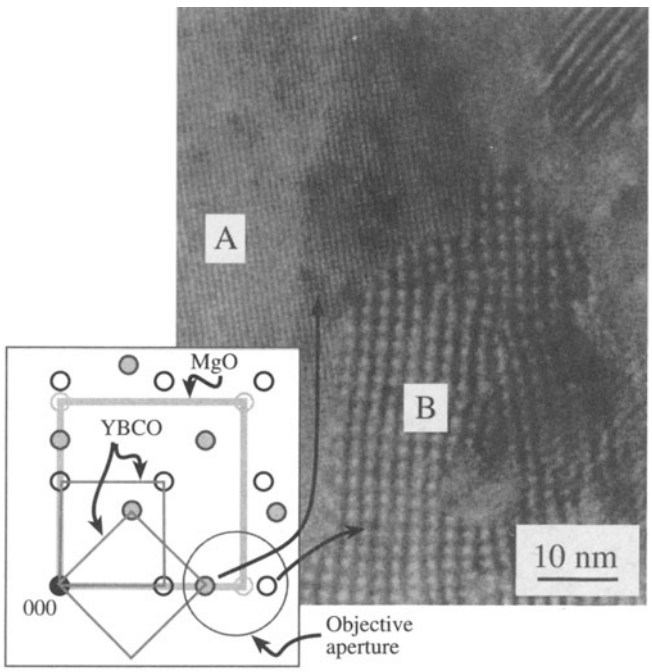


Figure 27.13. Moiré fringes formed when grains of YBCO grown on a single crystal of MgO are aligned to the substrate (B) or rotated through 45°(A); the spacing of the fringes is different so the position of the grain boundary can be identified. The circle in the DP shows the spots that cause the fringes. Small rotations of the fringes away from perfect alignment are exaggerated because the spots are close together.

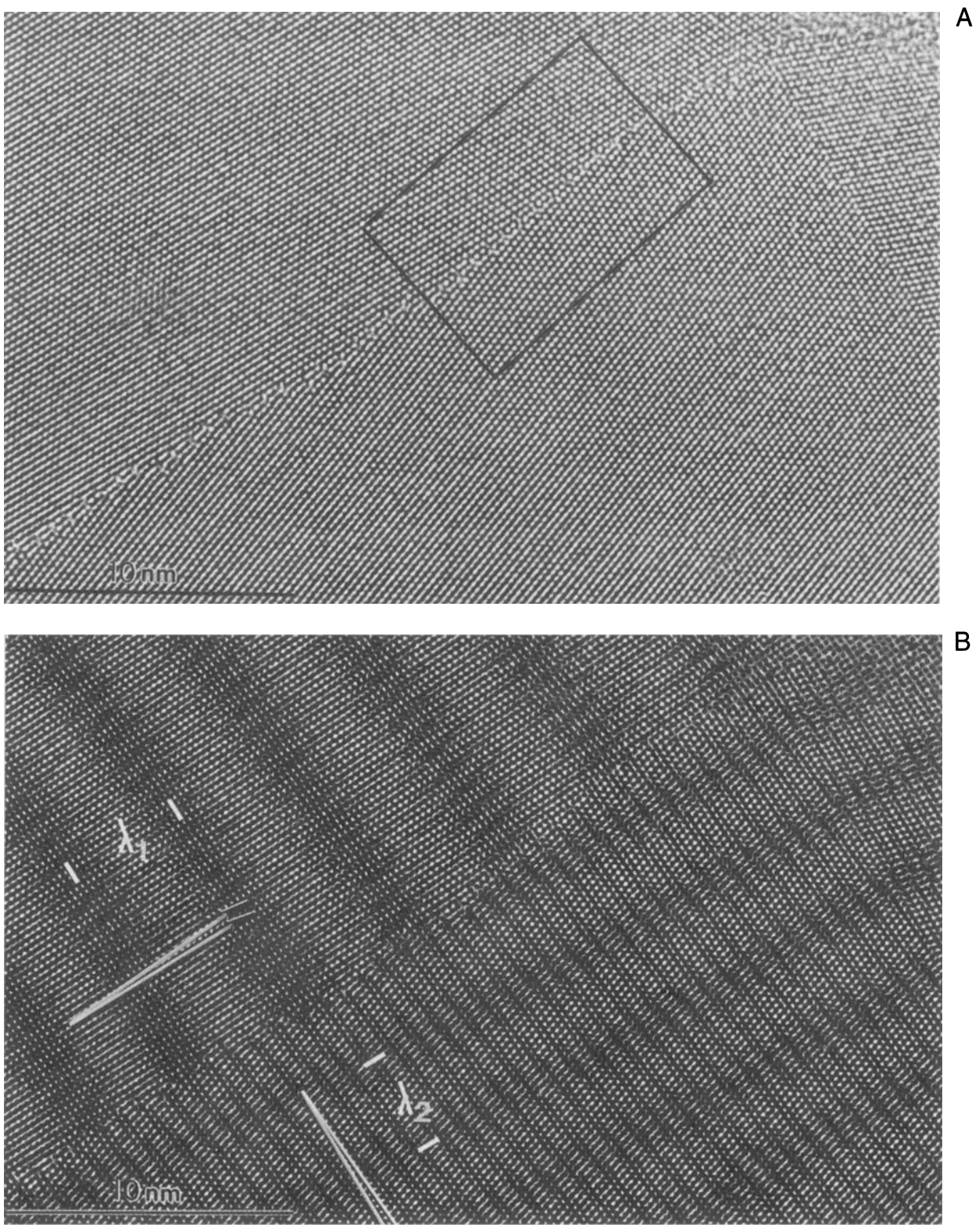


Figure 27.14. The use of "artificial" moiré fringes to analyze a special grain boundary in Al. (A) An experimental image. (B) The same image overlaid with a perfect-crystal lattice transparency, producing moiré fringes of different spacings, λ_1 and λ_2.

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 27.14A 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 to one another. You will have created a new artificial moiré image similar to that shown in Figure 27.14B, which was formed for a special grain boundary in Al by Hetherington and Dahmen (1992). 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 which 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°.

27.7. 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 classical 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 (e.g., Heavens and Ditchburn 1991). 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 on a plane which is close to the specimen, we are in the near-field or Fresnel regime. Now we extend this concept and say that

> Whenever we observe contrast only because we are forming an out-of-focus image, we are forming a Fresnel image.

Since we often study lines, planes, or platelets by this technique, we'll often see Fresnel fringes.

27.7.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 27.15A. 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 a photographic film, as shown in Figure 27.15B. The wire here is acting as a beam splitter; we'll encounter it again when we discuss holography in Chapter 31. 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 (1994) gives 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. Horuchi shows that

$$\gamma = \frac{I_{Max} - I_{Min}}{I_{Max} + I_{Min}} \qquad [27.15]$$

where I_{Max} is the intensity of the central fringe and I_{Min} is the intensity of the first minimum in Figure 27.15B.

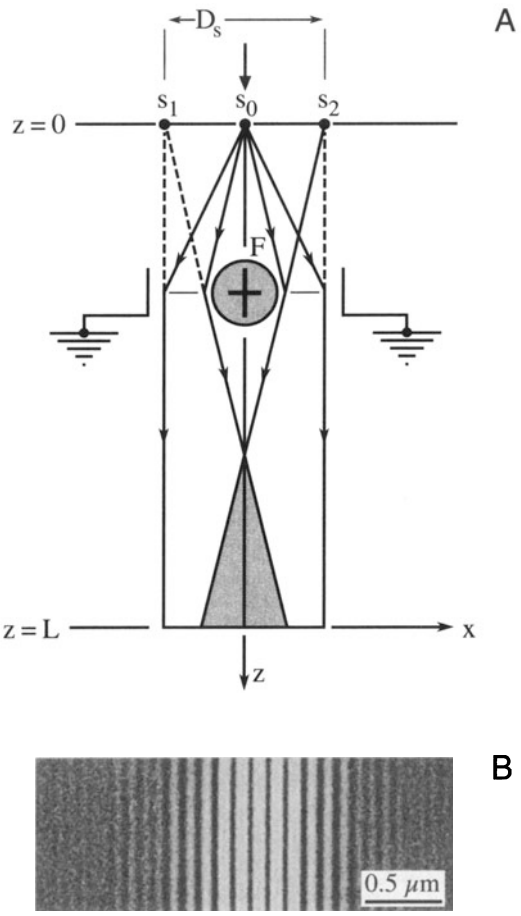


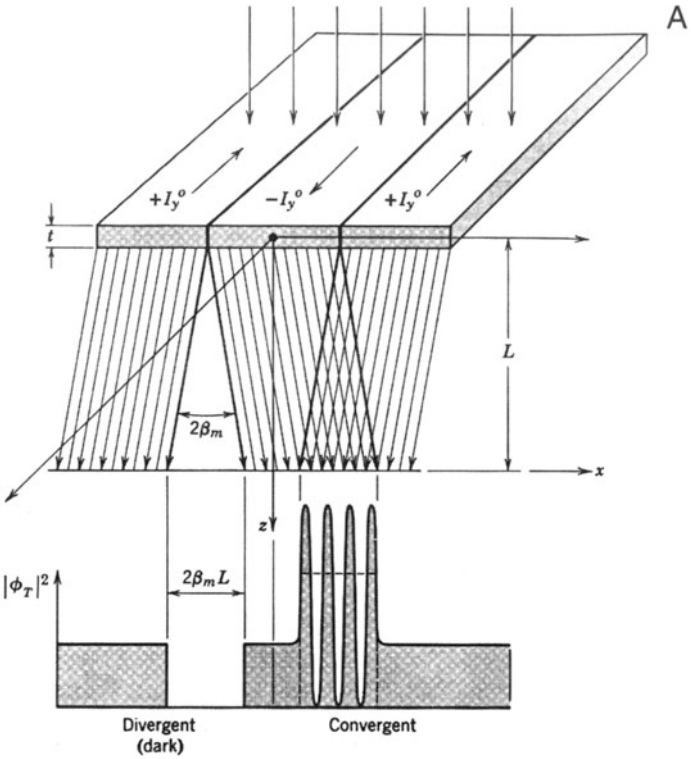
Figure 27.15. (A) A Fresnel biprism formed using a charged wire placed in the path of the beam; (B) the resulting interference fringes in the image.

27.7.B. Magnetic-Domain Walls

Although we'll discuss imaging magnetic materials in Chapter 31, 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 on an electron with velocity **v** is proportional to **v**×**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 27.16. The "converging" domain wall is remarkably similar to the electron interferometer. We can indeed produce a series of interference fringes. You should consult the original analysis of Boersch *et al.* (1960), but the basics are given by Hirsch *et al.* (1977) who show that the fringe spacing Δx is given by

$$\Delta x = \frac{\lambda(L + \ell)}{2\ell\beta_m} \qquad [27.16]$$

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 magnetic imaging in Chapter 31.



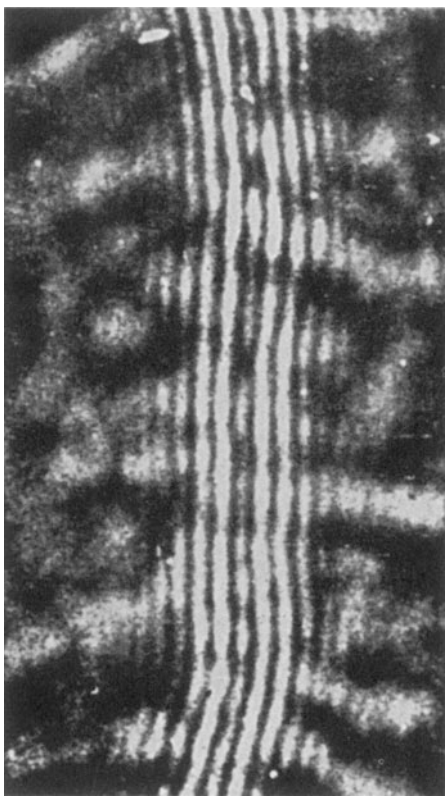


Figure 27.16. (A) Deflection of the electron beam by magnetic-domain walls; compare with Figure 27.15A. (B) Interference fringes from one such wall; compare with Figure 27.15B.

27.8. 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 which are fully enclosed inside the specimen by defocusing the image and observing the special form of phase contrast termed Fresnel contrast, which we introduced back in Sections 2.9 and 9.5. In principle, we can apply this technique to holes which 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 **s** = 0; the cavity then reduces the "thickness" of material locally.
- By using Fresnel contrast, as explained by Wilkens (1975).

In the Fresnel technique, the image shows contrast whenever the objective lens is not focused on the bottom surface of the specimen.

Fresnel-contrast images are always out of focus.

Wilkens expresses the wave function as

$$\psi(t, \mathbf{r}') = \psi_0(t)[1 + \Delta_r(\mathbf{r}') + i\Delta_i(\mathbf{r}')] \quad [27.17]$$

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).

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

$$\psi(t, \mathbf{r}') = \psi(t)[1 + i\Delta_i(\mathbf{r}')] \quad [27.18]$$

where Δ_i (using $w = s\xi_g$) is given by

$$\Delta_i = -\left(2\varepsilon_0 - \frac{1}{\xi_g}\frac{1}{(1+w)^{1/2}}\right) z_c(\mathbf{r}')p_i(z) \quad [27.19]$$

The difference in inner potential is included in ε_0, which is defined by the equation

$$\varepsilon_0 = -\frac{\Delta V}{E}k \quad [27.20]$$

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

$$|\psi(t, \mathbf{r}')|^2 = |\psi_0(t)|^2\left(1 + \Delta_i^2\right) \quad [27.21]$$

We can summarize some results from this study:

- When the image is in focus, the cavity is invisible so you have to 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.

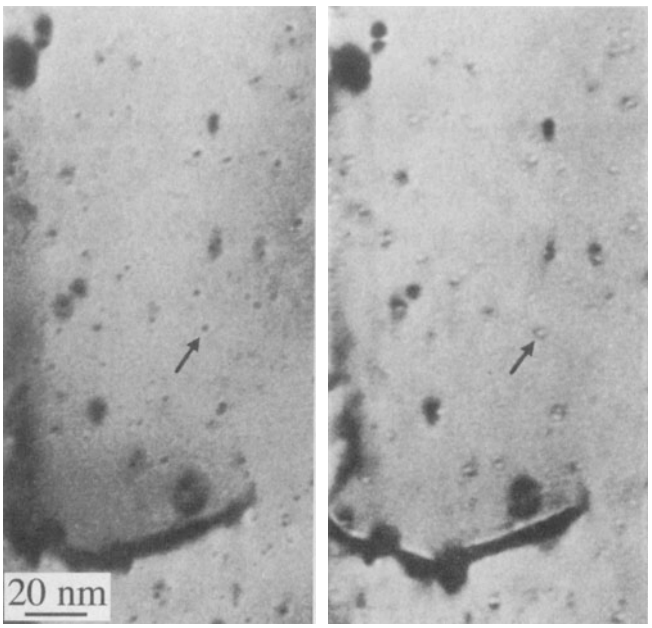


Figure 27.17. Fresnel contrast from He bubbles in Au. (A) Overfocus image. (B) Underfocus image.

- 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 to 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 9.20, where we had a dark fringe at underfocus and a bright fringe at overfocus.

The contrast is illustrated in Figure 27.17. You should note that it is not the same as the black–white contrast from small precipitates discussed in Chapter 25. You'll find a more detailed analysis in the article by Rühle and Wilkens (1975).

27.9. FRESNEL CONTRAST FROM LATTICE DEFECTS

This topic is one which is receiving more attention as the computer and simulation programs become, respectively, more powerful and more user-friendly. The reason for this increased attention is clear, as Bursill *et al.* (1978) 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 parameters (particu-

larly 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.

27.9.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 (Clarke 1979, Ness *et al.* 1986, Rasmussen *et al.* 1989, Rasmussen and Carter 1990). Part of the reason for this emphasis is simply that other techniques tend to give ambiguous results for such interfaces (Simpson *et al.* 1986).

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 29.11, we will consider the actual shape of this "potential well."

In a real TEM specimen, the grain boundary is likely to change in thickness 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 27.18.

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 (Taftø *et al.* 1986, Ross and Stobbs 1991a). 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 (Ross and Stobbs 1991b). Nevertheless, you must always look for associated changes in the real geometry which can even occur when you're just forming Fresnel fringes from the edge of the specimen (Fukushima *et al.* 1974).

27.9.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

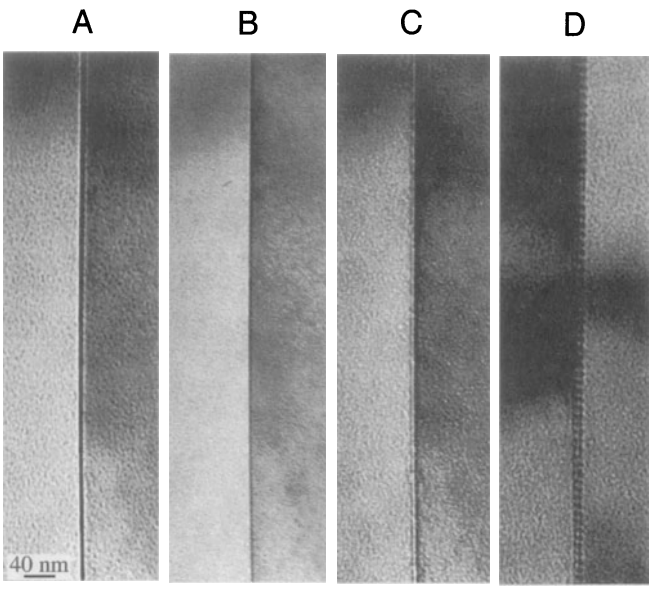


Figure 27.18. (A–D) A through-focus series of images from an edge-on GB showing the changes in Fresnel contrast. The image in (D) shows the boundary tilted over to reveal its periodic structure more clearly.

in Figure 27.19. Rühle and Sass (1984) analyzed this through-focus series, and images of other grain boundaries, by assuming that there is a change in $\Delta V(\mathbf{r})$ in the mean inner potential at the core of the dislocation. They proposed two models for $\Delta V(\mathbf{r})$. In model 1, when $r < r_0$

$$\Delta V(\mathbf{r}) = \Delta V_0 \left\{ 1 - e^{\frac{-(r - r_0)}{a}} \right\} \qquad [27.22]$$

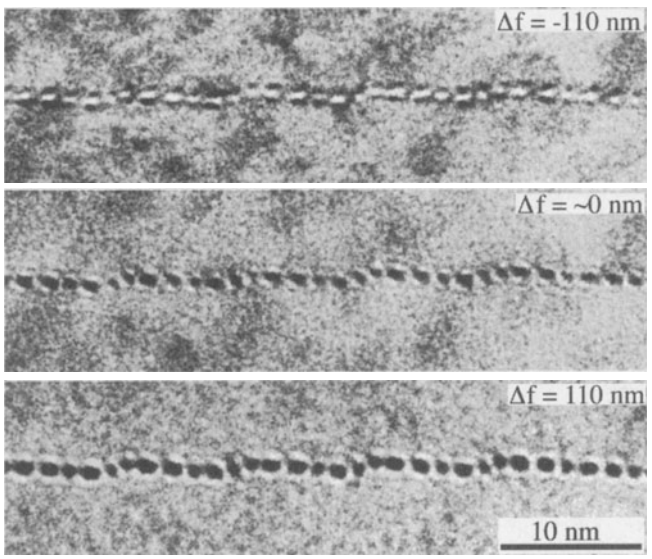


Figure 27.19. Series of experimental images recorded at different values of Δf for a low-angle grain boundary in NiO. Each white or black spot corresponds to one end-on dislocation.

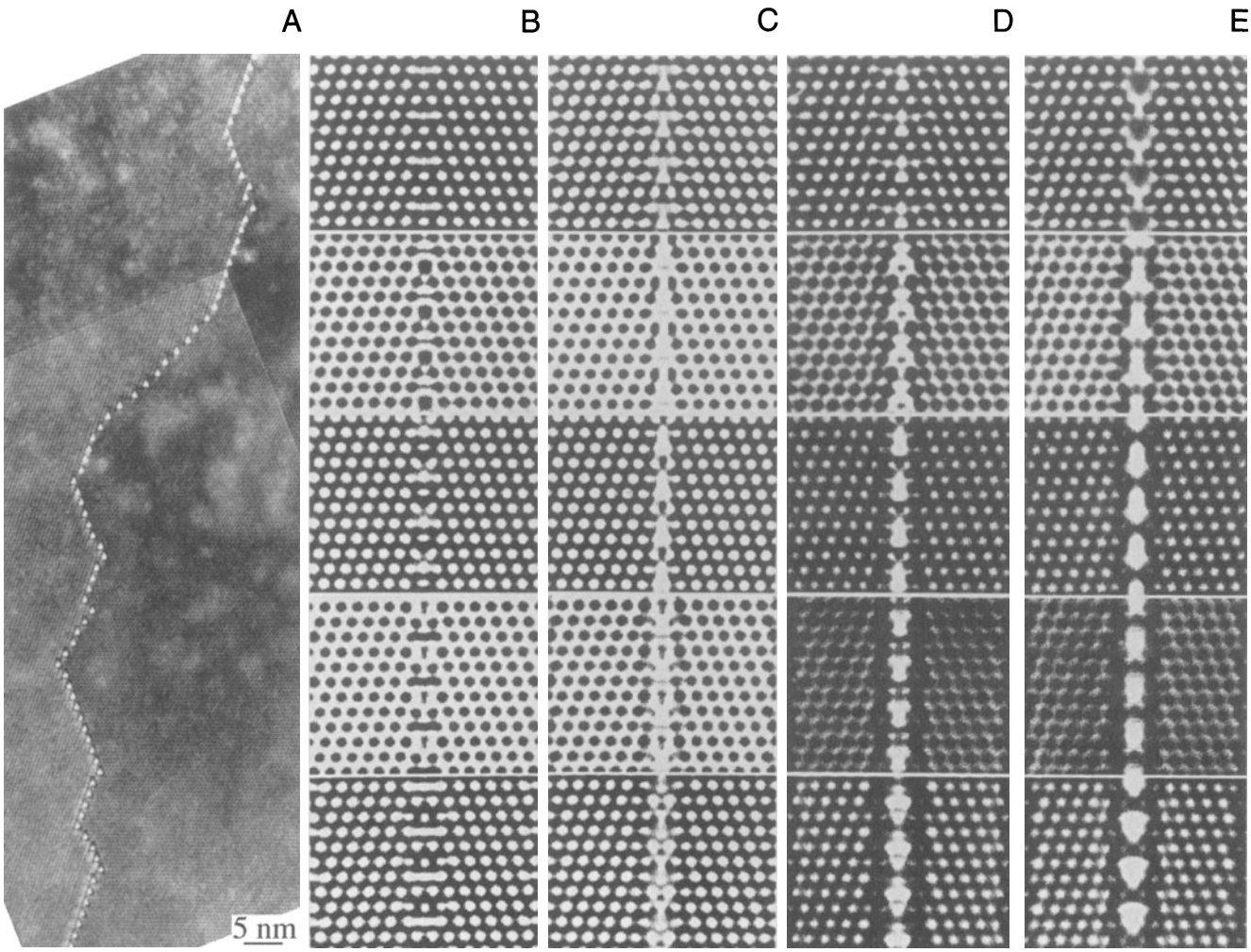


Figure 27.20. The structure of the (112) lateral twin boundary in a thin-foil of spinel consists of triangular prisms with a density lower than the bulk crystal. (A) Fresnel contrast is seen when these prism-like defects are imaged out of focus. (B–E) Simulated images; each column is a different model: (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 nm, –70 nm, –130 nm, –160 nm, –210 nm), and the thickness is 5.7 nm.

but when $r > r_0$

$$\Delta V(\mathbf{r}) = 0 \qquad [27.23]$$

The constant a is ~ $0.1r_0$. In model 2

$$\Delta V(\mathbf{r}) = \Delta V_0 \exp\left(-r^2/r_0^2\right) \qquad [27.24]$$

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 [110], Rühle and Sass found that $\Delta V_0 = 0.09V_0$ 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 different than 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. Even then, 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 27.20, 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 (Carter *et al.* 1987).

CHAPTER SUMMARY

Phase contrast will occur whenever we have more than one beam contributing to the image. The clue is: if you see fringes of any sort, then you are almost certainly observing a phase-contrast image. This conclusion even applies to stacking-fault fringes (Chapter 24) and thickness fringes (Chapter 23) 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-fringe images.
- Fresnel-contrast images.

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. You should note that it is difficult, but 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

Specific References

Boersch, H., Hamisch, H., Wohlleben, D., and Grohmann, K. (1960) *Z. Phys.* **167**, 72.

Bursill, L.A., Barry, J.C., and Hudson, P.R.W. (1978) *Phil. Mag.* **A37**, 789.

Carter, C.B., Elgat, Z., and Shaw, T.M. (1987) *Phil. Mag.* **55**, 1.

Clarke, D.R. (1979) *Ultramicroscopy* **4**, 33.

Fukushima, K., Kawakatzu, H., and Fukami, A. (1974) *J. Phys.* **D7**, 257.

Heavens, O.S. and Ditchburn, R.W. (1991) *Insight into Optics,* p. 73, John Wiley & Sons, New York.

Hetherington, C.J.D. and Dahmen, U. (1992) *Scanning Microscopy Supplement* **6**, p. 405, Scanning Microscopy International, AMF O'Hare, Illinois.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977) *Electron Microscopy of Thin Crystals,* 2nd edition, Krieger, Huntington, New York.

Horiuchi, S. (1994) *Fundamentals of High-Resolution Transmission Electron Microscopy,* North-Holland, Amsterdam.

Krivanek, O.L. and Rez, P. (1980) Proc. 38th Ann. EMSA Meeting (Ed. G.W. Bailey), p. 170, Claitors, Baton Rouge, Louisiana.

Minter, J.W. (1956) *Proc. Roy. Soc.* (London) **A236**, 119.

Ness, J.N., Stobbs, W.M., and Page, T.F. (1986) *Phil. Mag.* **54**, 679.

Norton, M.G. and Carter, C.B. (1995) *J. Mater. Sci.* **30**, 381.

Rasmussen, D.R. and Carter, C.B. (1990) *Ultramicroscopy* **32**, 337.

Rasmussen, D.R., Simpson, Y.K., Kilaas, R., and Carter, C.B. (1989) *Ultramicroscopy* **30**, 52.

Ross, F.M. and Stobbs, W.M. (1991a) *Phil. Mag.* **A63**, 1.

Ross, F.M. and Stobbs, W.M. (1991b) *Phil. Mag.* **A63**, 37.

Rühle, M. and Sass, S.L. (1984) *Phil. Mag.* **A49**, 759.

Rühle, M. and Wilkens, M. (1975) *Crystal Lattice Defects* **6**, 129.

Simpson, Y.K., Carter, C.B., Morrissey, K.J., Angelini, P., and Bentley, J. (1986) *J. Mater. Sci.* **21**, 2689.

Taftø, J., Jones, R.H., and Heald, S.M. (1986) *J. Appl. Phys.* **60**, 4316.

Vincent, R. (1969) *Phil. Mag.* **19**, 1127.

Wilkens, M. (1975) in *Electron Microscopy in Materials Science,* **II** (Eds. U. Valdrè and E. Ruedl), p. 647, CEC, Brussels.

High-Resolution TEM

28

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 which transfers information from the specimen to the image. The optics consist 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. As you know from Chapter 6, the lens system is not perfect so the image is distorted and you lose some data (Abbe's theory). The second problem is that 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. We close the chapter with a review of the experimental applications of HRTEM to include periodic and nonperiodic materials, mixtures of the two, or just single atoms.

High-Resolution TEM

28

28.1. THE ROLE OF AN OPTICAL SYSTEM

What the microscope does is to transform each point on the specimen into an extended region 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 we consider two nearby points, A and B, they will produce two overlapping images, g_A and g_B. 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

$$g(\mathbf{r}) = \int f(\mathbf{r}')\, h(\mathbf{r} - \mathbf{r}')\, dr' \qquad [28.1]$$

$$= f(\mathbf{r}) \otimes h(\mathbf{r}) \qquad [28.2]$$

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.

Since $h(\mathbf{r})$ 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 $h(\mathbf{r})$.

Spence (1988) calls $h(\mathbf{r})$ the impulse response function, and notes that it can only apply to small patches of specimen 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 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 (more of this later), 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.

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

$$g(x,y) = \sum_{u_x,\, u_y} G\left(u_x, u_y\right) \exp\left(2\pi i\left(x u_x + y u_y\right)\right) \qquad [28.3]$$

$$g(x,y) = \sum_{\mathbf{u}} G(\mathbf{u}) \exp\left(2\pi i\, \mathbf{u} \cdot \mathbf{r}\right) \qquad [28.4]$$

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 $G(\mathbf{u})$,

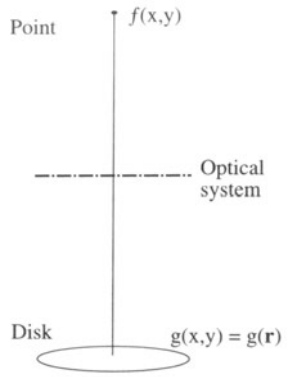


Figure 28.1. An optical system transforms a point in the specimen (described by $f(x,y)$) into a disk in the image described by $g(x,y)$. The intensity in the image at point (x,y) can be described by the function $g(x,y)$ or $g(\mathbf{r})$. It has a unique value for each value of (x,y) so we say that $g(\mathbf{r})$ is a representation of the image.

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 $h(\mathbf{r})$.

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.

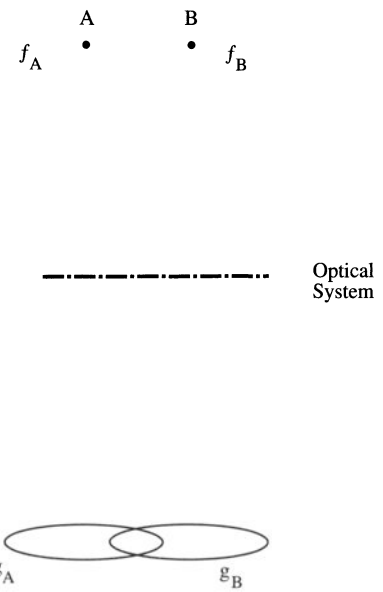


Figure 28.2. Two points, f_A and f_B, in the specimen produce two disks, g_A and g_B, in the image.

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

$$B(\mathbf{u}) = \exp\left(-i\chi(\mathbf{u})\right) \qquad [28.7]$$

The term $\chi(\mathbf{u})$ can be written as

$$\chi(\mathbf{u}) = \pi\, \Delta f\, \lambda\, u^2 + \tfrac{1}{2}\pi\, C_s\, \lambda^3\, u^4 \qquad [28.8]$$

We will give a crude 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.

$\Delta f > 0$ 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.

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

$$f(x,y) = A(x,y)\exp\left(i\phi_t(x,y)\right) \qquad [28.9]$$

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 the following argument (Van Dyck 1992). To do this we 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

$$V_t(x,y) = \int_0^t V(x,y,z)dz \qquad [28.10]$$

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.)

$$\lambda = \frac{h}{\sqrt{2meE}} \qquad [28.11]$$

(We'll give the analysis in a simple nonrelativistic form for simplicity.) When the electrons are in the crystal, λ is changed to λ'

$$\lambda' = \frac{h}{\sqrt{2me\left(E + V(x,y,z)\right)}} \qquad [28.12]$$

so we can say that, when passing through a slice of material of thickness dz, the electrons experience a phase change given by

$$d\phi = 2\pi\frac{dz}{\lambda'} - 2\pi\frac{dz}{\lambda} \qquad [28.13]$$

$$d\phi = 2\pi\frac{dz}{\lambda}\left(\frac{\sqrt{E + V(x,y,z)}}{\sqrt{E}} - 1\right) \qquad [28.14]$$

$$d\phi = 2\pi\frac{dz}{\lambda}\left(\left(1 + \frac{V(x,y,z)}{E}\right)^{\frac{1}{2}} - 1\right) \qquad [28.15]$$

$$d\phi \simeq 2\pi\frac{dz}{\lambda}\frac{1}{2}\frac{V(x,y,z)}{E} \qquad [28.16]$$

$$d\phi \simeq \frac{\pi}{\lambda E}V(x,y,z)dz \qquad [28.17]$$

$$d\phi \simeq \sigma\, V(x,y,z)dz \qquad [28.18]$$

So the total phase shift is indeed dependent only on $V(x,y,z)$ since

$$d\phi \simeq \sigma\int V(x,y,z)dz = \sigma\, V_t(x,y) \qquad [28.19]$$

where $V_t(x,y)$ is the potential projected in the z-direction.

We call σ the interaction constant. It is not a scattering cross section, but is another expression for the elastic interaction we discussed in Chapter 3. It 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).

Now, we can take account of absorption by including a function $\mu(x,y)$ (Cowley 1992) so that our specimen transfer function $f(x,y)$ is given by

$$f(x,y) = \exp\left[-i\,\sigma\, V_t(x,y) - \mu(x,y)\right] \qquad [28.20]$$

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.

> 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 <<1. Then we expand the exponential function, neglecting μ and higher-order terms, so that $f(x,y)$ becomes

$$f(x,y) = 1 + i\,\sigma\,V_t(x,y) \qquad [28.21]$$

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 many 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! As a second example, Fejes (1977) has shown that, for the complex oxide $Ti_2Nb_{10}O_{27}$, the WPOA is only valid if the specimen thickness is < 6 Å! 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 will 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

$$\psi(x,y) = \left[1 - i\sigma V_t(x,y)\right] \otimes h(x,y) \qquad [28.22]$$

If we represent $h(x,y)$ as cos (x,y) + i sin (x,y), then $\psi(x,y)$ becomes

$$\psi(x,y) = 1 + \sigma\,V_t(x,y) \otimes \sin(x,y) - i\,\sigma\,V_t(x,y) \otimes \cos(x,y) \qquad [28.23]$$

and 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 $h(x,y)$). Therefore, we can set $B(\mathbf{u}) = 2 \sin \chi(\mathbf{u})$ rather than exp $(i\chi(\mathbf{u}))$.

We can now define a new quantity, $T(\mathbf{u})$, which we'll call the 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 but is closely related 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 28.25. You may also see authors use a negative sign in equation 28.26 (in particular, in Reimer 1993). 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. The terminology comes from the analysis of the imaging process for incoherent light in visible-light optics. With incoherent 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 contrast transfer function (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 "contrast transfer function" 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 "weak-phase object."

If the specimen acts as a weak-phase object, then the transfer function $T(\mathbf{u})$ is sometimes called the "contrast transfer function," 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.

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 contrast transfer function is now compared to the phase-distortion function, a number of observations can be made. Note that the contrast transfer function is oscillatory; there are "bands" of good transmission separated by "gaps" (zeros) where no transmission occurs.

The contrast transfer function 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 of $\pm\pi$.

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 $\mathbf{u}$.

The reason for this behavior is that the phase shift due to diffraction is $-\pi/2$. 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 χ(u), SIN χ(u), AND COS χ(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 occurred, interpretation would be difficult!

$T(\mathbf{u})$ 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 6.15) and the defocus (equation 11.19) of the objective lens, we find that a point at the specimen will actually be imaged as a disk with radius $\delta(\theta)$

$$\delta(\theta) = C_s\theta^3 + \Delta f\theta \qquad [28.29]$$

The rays which pass through the objective lens at angle θ are not focused to the same point due to the spherical aberration of the objective lens and the finite value of Δf. 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

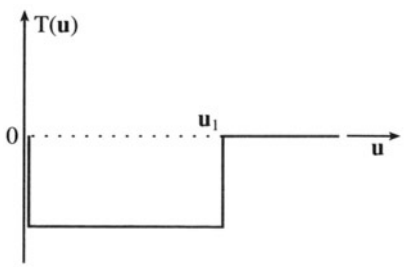


Figure 28.3. The ideal form of the transfer function, $T(\mathbf{u})$. In this example $T(\mathbf{u})$ is large and negative between $\mathbf{u} = 0$ and $\mathbf{u} = \mathbf{u}_1$.

$$D(\theta) = \int_0^{\theta} \delta(\theta) d\theta = \frac{C_s \theta^4}{4} + \Delta f \frac{\theta^2}{2} \qquad [28.30]$$

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 with 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

$$\chi(\mathbf{u}) = \text{phase} = \frac{2\pi}{\lambda} D(\mathbf{u}) = \frac{2\pi}{\lambda}\left(C_s \frac{\lambda^4 u^4}{4} + \Delta f \frac{\lambda^2 u^2}{2}\right) \qquad [28.33]$$

and we have

$$\chi = \pi \Delta f \lambda u^2 + \frac{1}{2}\pi C_s \lambda^3 u^4 \qquad [28.34]$$

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). (When you are ready for a more rigorous derivation of equation 28.34, see Spence, 1988, Section 3.3; 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 29 and vary each of the parameters one by one. The plot of $T(\mathbf{u})$ (= 2 sin χ) versus **u,** shown in Figure 28.4, illustrates the main features. The curve has been drawn for C_s = 1 mm, E_0 = 200 keV, and a defocus value of –58 nm.

The important features of this curve are shown in Figures 28.4–28.6:

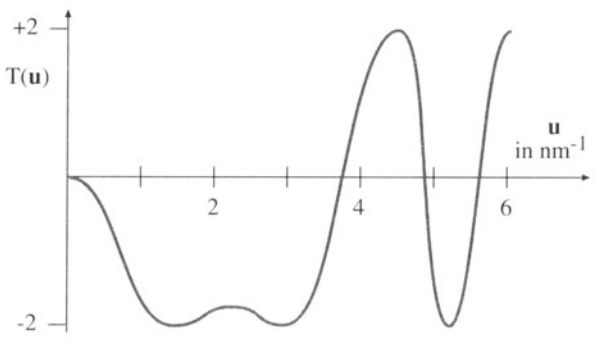


Figure 28.4. A plot of $T(\mathbf{u})$ versus **u** (C_s = 1 mm, E_0 = 200 keV, Δf = –58 nm).

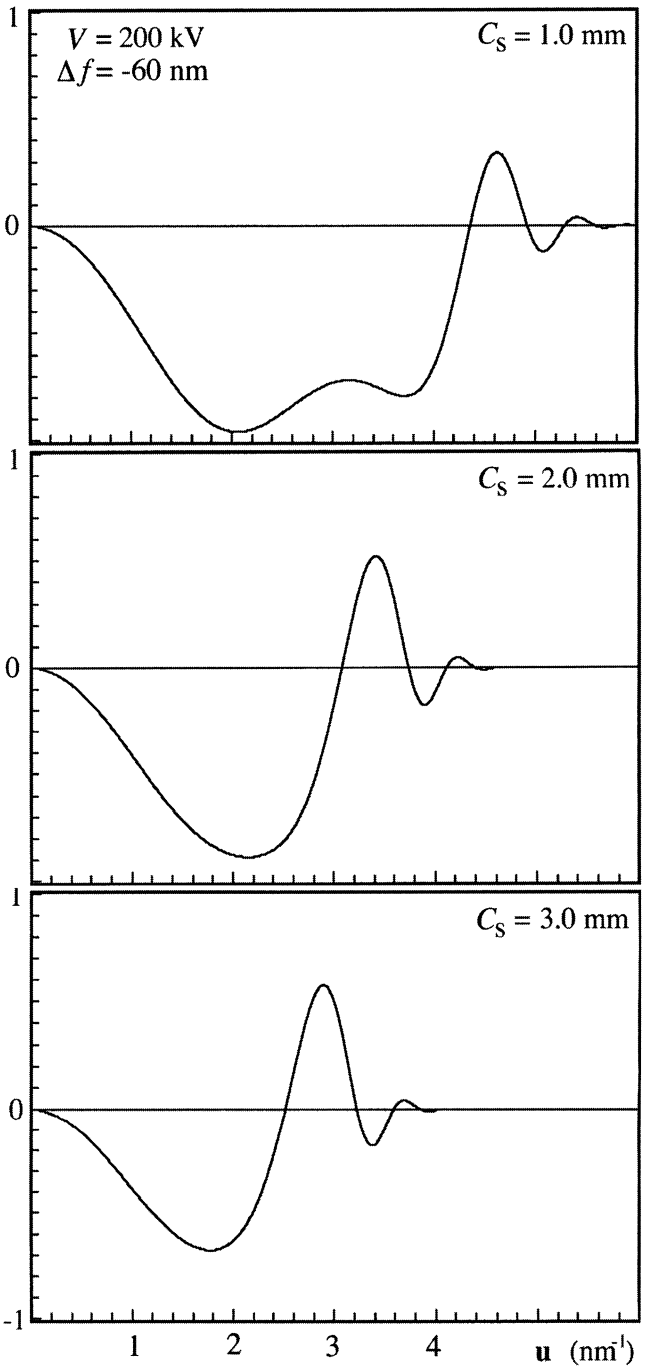


Figure 28.5. A series of sin χ curves calculated for different values of C_s. Remember 2 sin $\chi = T(\mathbf{u})$.

- sin χ starts at 0 and decreases. When **u** is small, the Δf term dominates.
- sin χ 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.

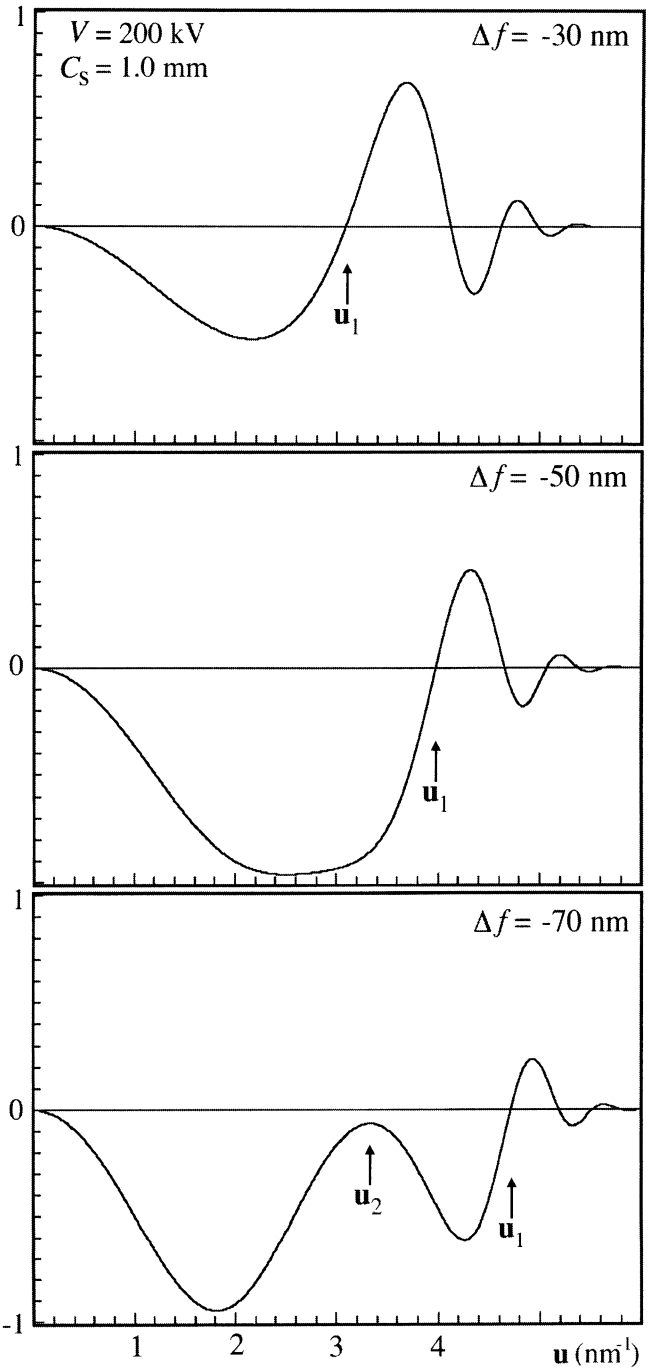


Figure 28.6. A series of sin χ curves calculated for different values of Δf.

Once you've selected your microscope and its objective lens, you have fixed C_s although C_s does depend to some extent on the λ you choose. The curve of $T(\mathbf{u})$ versus $\mathbf{u}$ does not depend on your specimen. Figure 28.5 shows a series of sin χ 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.

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 that the bump in the curve at $\mathbf{u}_2$ will eventually increase as Δf increases until it crosses the $\mathbf{u}$-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.

High spatial frequencies ⇒ large diffraction angles ⇒ larger effect of objective lens (C_s).
So for a large objective aperture semiangle β, the β^4 term wins, i.e., C_s wins, but you can vary Δf.

28.7. SCHERZER DEFOCUS

The presence of zeros in the contrast transfer function 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 transfer function 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

$$\Delta f_{Sch} = -1.2\,(C_s\lambda)^{\frac{1}{2}} \qquad [28.35]$$

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 sophisticated 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 28.35; 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 our ability to distinguish closely spaced object points. 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.

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 $\chi(\mathbf{u})$ is close to –120°; then sin χ will be near –1 when χ is between –120° and –60°. We know that when $\chi = \pi$, sin $\chi = 0$, so we want sin χ to be as large as possible over a large range of $\mathbf{u}$. 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 28.34 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 $\chi = -120°$, equation 28.34 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 $(1.33)^{1/2} = 1.155$, we have deduced equation 28.35. At this value of Δf we find 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 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 $(C_s\lambda)^{1/2}$ and $(C_s\lambda^3)^{1/4}$ seen in equations 28.39 and 28.41 are so important in HRTEM that Hawkes (1980) has designated them to be the units 1 Sch and 1 Gl (the scherzer and the glaser) in honor of the two 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 the variation of Δf and C_s using EMS (Section 1.5). An excellent, though advanced, discussion of such diagrams is given by Thon (1975), who describes how they can be used to design phase plates for the TEM. Spence (1988) 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 30.6A.

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 28.7.

The exact mathematical form of these envelope functions is complex. In general, the result is described by multiplying the 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 $T_{eff}(\mathbf{u})$

$$T_{eff}(\mathbf{u}) = T(\mathbf{u})\, E_c\, E_a \qquad [28.42]$$

The effect of the envelope functions is to impose a virtual aperture in the back focal plane of the objective 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 micro-

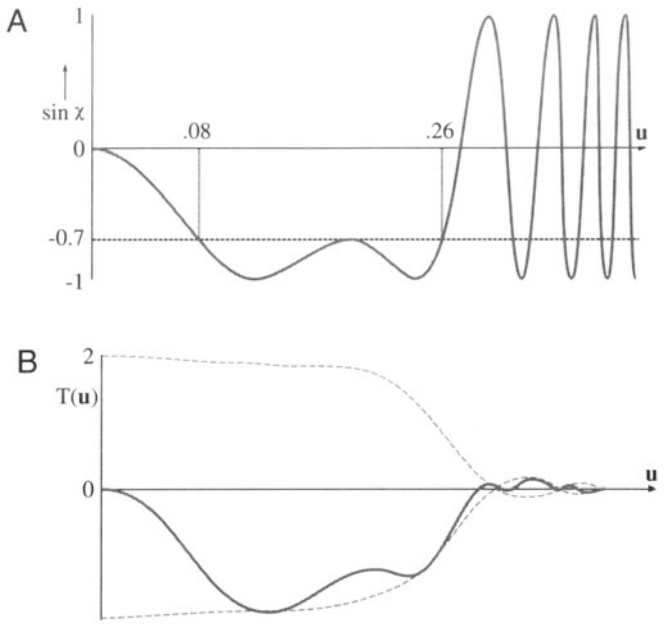


Figure 28.7. (A) sinχ(**u**) versus **u** without damping of the higher spatial frequencies. (B) T(**u**) versus **u** modified by the damping envelope (dashed line); $\Delta f = -100$ nm, $C_s = 2.2$ mm.

scope. 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 the next chapter) 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 simulations.

28.9. IMAGING USING PASSBANDS

Because of the focus dependence of the contrast-transfer function, 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 when

$$\sin \chi(\mathbf{u}) \simeq 0.3 \qquad [28.43]$$

or

$$\Delta f_{MC} = -0.44 \left(C_s \lambda\right)^{\frac{1}{2}} \qquad [28.44]$$

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, since you can minimize the contrast easily, providing you have correctly aligned the microscope and corrected the astigmatism.

Some other special settings of the transfer function may also be useful. The idea is to make use of *passbands* or large "windows" in the transfer function 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, over a range of u which includes the reflection of interest. These passbands occur periodically with underfocus at values set by

$$\Delta f_p^n = \left\{ \left(\frac{8n+3}{2} \right) \left(C_s \lambda\right) \right\}^{\frac{1}{2}} \qquad [28.45]$$

This formula is not an exact relationship but it gives us a good guide; its derivation is given by Spence (1988). 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 JEOL 200 CX, these pass-

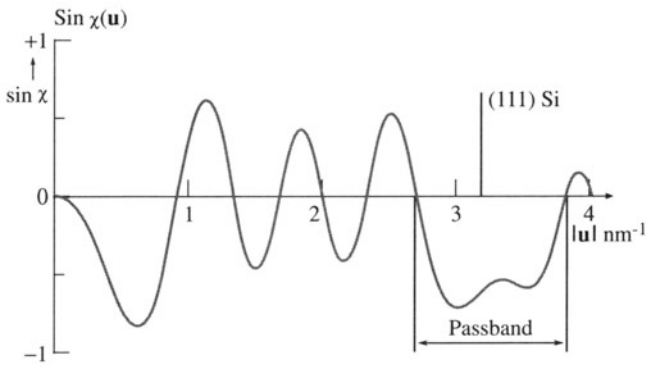


Figure 28.8. Special setting of the transfer function to make use of passbands or "windows" in the transfer function, here optimized to image Si (111).

band 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.

We can define an "aberration-free focus" (AFF) condition for any specific crystal (Hashimoto and Endoh 1978). 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 transfer function out to very high order. This aberration-free focus setting is defined by

$$\Delta f_{\mathrm{AFF}} \lessgtr \left\{ 2(4\,m \pm 0.23) + C_s\lambda^3/d^4 \right\} \left(d^2/2\lambda \right) \qquad [28.46]$$

where $m = 0, 1, 2, 3$, etc. and d is the fundamental lattice spacing of the first-order Bragg beam to be resolved. In the application to a Au crystal in [001] orientation where $d(020) = 0.2035$ nm, using a 100-kV microscope with $C_s = 0.75$ mm, Δf_{AFF} is –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 lose all the information about the defect, since defects scatter between the Bragg peaks. Any information falling in the gap of the transfer function 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 transfer function. You can only know that by very careful evaluation of your negatives 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 just what 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 (Carter *et al.* 1986). You would like to obtain information at, say, 2.3 Å (the spacing of the oxygen 111 planes), but your point-to-point resolution is 2.7 Å. You could still learn a lot about the spinel from the 4.6 Å spinel (111) planes, so you might use an aperture to remove information which only adds uninterpretable detail below 4.6 Å. 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 weak-phase object. Most specimens of interest do *not* satisfy this criterion.

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 weak-phase object! Multiple scattering limits most phase-contrast imaging conditions for crystalline materials.

> 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."

Thicker specimens also are susceptible to Fresnel effects associated with spreading of the wave front as it is transmitted through more 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 29 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 electronic 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 30).
- 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 semiangle).
- Simulate and/or process the images using available computer codes (Chapter 29).

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 30, 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. (See Hall (1953) for a discussion of other aberrations.) 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).

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 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 specimen. You must get used to using the TV and the computer. If you are going to do quantitative HRTEM, you'll have to be comfortable with both.

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 29 and 30, 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 FEGTEM. Such a beam is now highly coherent, so the envelope function shown in Figure 28.7 extends to greater values of **u**. The computer now becomes indispensable because we have to interpret images which have contrast reversals beyond Scherzer defocus. If a carefully designed multipole lens is inserted into an HRTEM, Rose (1990, 1991) has shown that it is, in principle, possible to correct C_s!

When this happens we will have to rethink our approach to HRTEM. Think what will happen to the scherzer

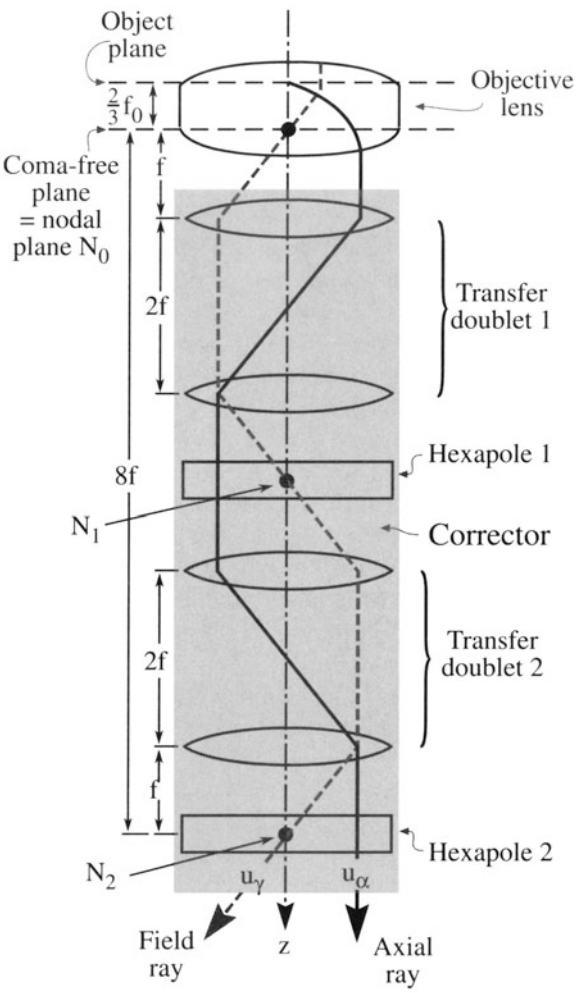


Figure 28.9. The post-objective lens corrector system proposed by Rose to correct spherical aberrations in the objective lens.

and the glaser. What will the Gl/Sch be? The corrector proposed by Rose is shown schematically in Figure 28.9. 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 of 1 in 10^4 to give atomic resolution. When C_s is zero, the specimen resolution limit will be determined by C_c

$$d_{C_s=0} \simeq \left[\left(\frac{\Delta E}{E}\right)\lambda C_c\right]^{\frac{1}{2}} \qquad [28.47]$$

If C_c = 2 mm and ΔE ~ 0.3 eV, a 200-kV FEGTEM could achieve a resolution of 0.8 Å. If C_c is also corrected, which we'll see how to do 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 proposed lens, C_s = 3 mm and the resolution limit is 0.28 Å for this 200-kV FEGTEM. Other lens defects will limit this to ~0.5 Å, but with a price of $12M for a 1.25-MeV machine which damages your specimen in seconds, the Rose corrector could 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 (Shannon and Weaver 1964, Van Dyck 1992). 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.

The linear relation between input and output signals can be described by the concept of the transfer function. Overall, the transfer function relates an input spectrum to an output spectrum, and it operates only in the frequency domain.

In general, for a linear system, if we know the transfer function, 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 transfer function.

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 should we bother to discuss this in HRTEM? The answer lies in finding an appropriate linear relation between the object and the image.

> 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.

28.13. FEGTEMs 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 (Van Dyck and de Jong 1992, de Jong and Van Dyck 1993).

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

$$E_T(\mathbf{u}) = E_c(\mathbf{u})E_s(\mathbf{u})E_d(\mathbf{u})E_v(\mathbf{u})E_D(\mathbf{u}) \qquad [28.48]$$

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 the 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

$$E_c(\mathbf{u}) = \exp\left[-\tfrac{1}{2}(\pi\lambda\delta)^2 u^4\right] \qquad [28.49]$$

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 $\Delta I_{obj}/I_{obj}$ are the instabilities in the high-voltage 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

$$\rho_c = \left(\frac{\pi\lambda\delta}{\sqrt{2\ln(s)}}\right)^{\frac{1}{2}} \qquad [28.51]$$

where e^{-s} is the cut-off value for the envelope. If we take $\ln_e s$ to be 2, then

$$\rho_c = \left(\frac{\pi\lambda\delta}{2}\right)^{\frac{1}{2}} \qquad [28.52]$$

The source-dependent envelope function is new because, until we have an 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

$$E_s(\mathbf{u}) = \exp\left[\left(\frac{\alpha}{2\lambda}\right)^2\left(\frac{\partial\chi(\mathbf{u})}{\partial u}\right)^2\right]$$
$$= \exp\left[-\left(\frac{\pi\alpha}{\lambda}\right)^2\left(C_s\lambda^3u^3 + \Delta f\lambda u\right)^2\right] \qquad [28.53]$$

Here, α is the semiangle characterizing the Gaussian distribution. What this equation tells us is that if α is too large ($\geq$ 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

$$\Delta f_{opt} = -\frac{3}{4}C_s\lambda^2u_{max}^2 = -\frac{3}{4}\frac{C_s\lambda^2}{\rho_i^2} \qquad [28.54]$$

In this equation, ρ_i is the information limit of the microscope (which is how we chose u_{max}). This defocus value will be important later when we discuss holography in an FEGTEM. The two curves shown in Figure 28.10 illustrate how this envelope function varies within Δf. It can also be optimized by decreasing the semiangle α. With a little more manipulation, de Jong and Van Dyck (1993) show that the information limit due to the limited coherence of the source is given by

$$\rho_\alpha = \left(\frac{6\pi\alpha a}{\lambda\sqrt{\ln(s)}}\rho_s^4\right)^{\frac{1}{3}} \qquad [28.55]$$

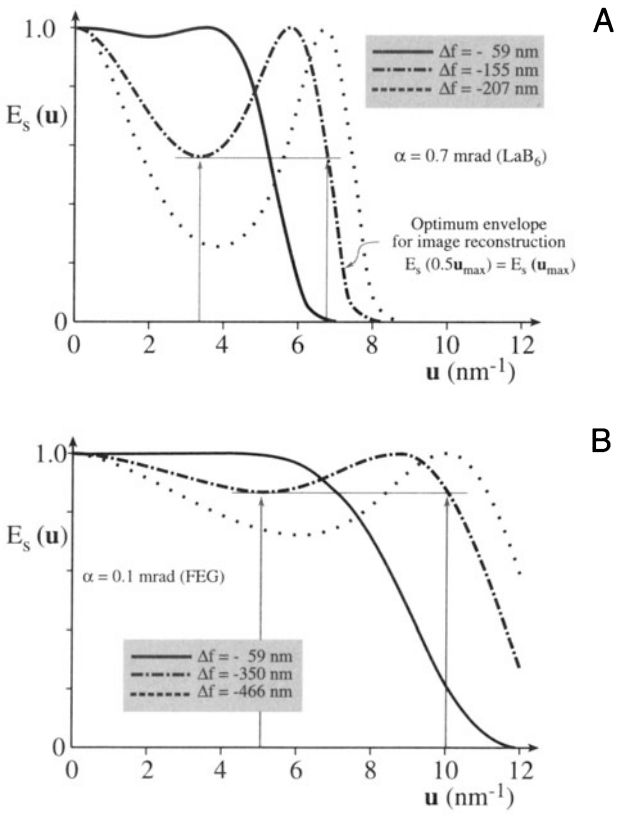


Figure 28.10. Variations in the envelope function, $E_s(\mathbf{u})$, for different objective lens defocus: (A) LaB_6 source, (B) FEG.

The envelope functions for the drift and vibration represent a new method for 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 $E_v(\mathbf{u})$

$$\rho_d = \frac{\pi d}{\sqrt{6 \ln (s)}} \quad [28.56]$$

and

$$\rho_v = \frac{\pi v}{\sqrt{\ln (s)}} \quad [28.57]$$

In these equations, d is the total drift during the exposure time, 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.

The idea is simple but the math is more difficult. If the image which is actually captured by the camera is a circle, we can 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

$$\alpha C_s \lambda^3 u_{max}^3 = R_w \quad [28.58]$$

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.

> 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 value of R_w is related to the number of pixels, N, and their size, D

$$R_w = \frac{1}{2} N D \quad [28.59]$$

The information limit due to the detector (i.e., the crossover values of the detector envelope functions $E_D(\mathbf{u})$) is

$$\rho_D = \left(\frac{12\sqrt{2}\pi a}{N\sqrt{\ln (s)}} \right)^{\frac{1}{4}} \rho_s \quad [28.60]$$

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

$$\alpha \le \frac{\lambda}{6\pi a \rho_s} \left(\frac{\rho_c}{\rho_s} \right)^3 \quad [28.61]$$

$$N \ge 12\sqrt{2}\pi a \left(\frac{\rho_s}{\rho_c} \right)^4 \quad [28.62]$$

$$d \le \frac{\sqrt{6}}{\pi} \rho_c \simeq 0.8 \rho_c \quad [28.63]$$

$$u \le \frac{1}{\pi} \rho_c \simeq 0.3 \rho_c \quad [28.64]$$

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. If the smallest image element we need to examine has an area ρ_i^2, then the background signal I_0 is given by

$$I_0 = D\rho_i^2 = \beta\pi\alpha^2 t \rho_i^2 \quad [28.65]$$

where D is the electron dose, α is the semiangle 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

$$DEF\rho_i^2 = kI_0^{\frac{1}{2}} = k\rho_i D^{\frac{1}{2}} \quad [28.66]$$

Table 28.1. Maximum Convergence Semiangle 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).[a]

	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

[a] ε_0: Gaussian focus, ε_{opt}: optimum focus; $\lambda = 0.011\ \rho_s$ (de Jong and Van Dyck 1993).

Therefore, the signal-to-noise ratio for $\alpha = 1$ mrad and $t = 1$ second can be written as

$$s_0 = 443\,\rho_i \frac{F}{k}\beta^{\frac{1}{2}} \qquad [28.67]$$

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} $Am^{-2}sr^{-1}$ for a LaB_6 gun and 10^{13} $Am^{-2}sr^{-1}$ 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 to 2.2, whereas for an FEG, $\rho_i = 0.1$ nm and $\ln_e s_0$ is 4.5 to 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

$$\alpha_{opt} = \frac{1}{k_s\sqrt{2}}\left(\frac{\rho_i}{\rho_c}\right)^3 = \frac{1}{6\pi a\sqrt{2}}\frac{\lambda}{\rho_s}\left(\frac{\rho_i}{\rho_s}\right)^3 \qquad [28.68]$$

$$t_{opt} = \frac{1}{2k_d}\left(\frac{\rho_i}{\rho_c}\right) \simeq 0.39\left(\frac{\rho_i}{v_d}\right) \qquad [28.69]$$

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 FEGTEM 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.

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 (1993). 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 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 you can appreciate from Figure 28.11, if the image is astigmatic you'll see this 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 define Δf. If you try to use the wobbler to do coma-free alignment (Otten and Coene 1993), 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 30) or converge the beam! The latter way deteriorates the spatial coherence and you've made your $1.5M FEG machine behave like an old $200k LaB_6 machine.
- Through-focus series 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 equation 28.36, we can express the delocalization as

$$\Delta R = \lambda u\left(\Delta f + C_s\lambda^2 u^2\right) \qquad [28.70]$$

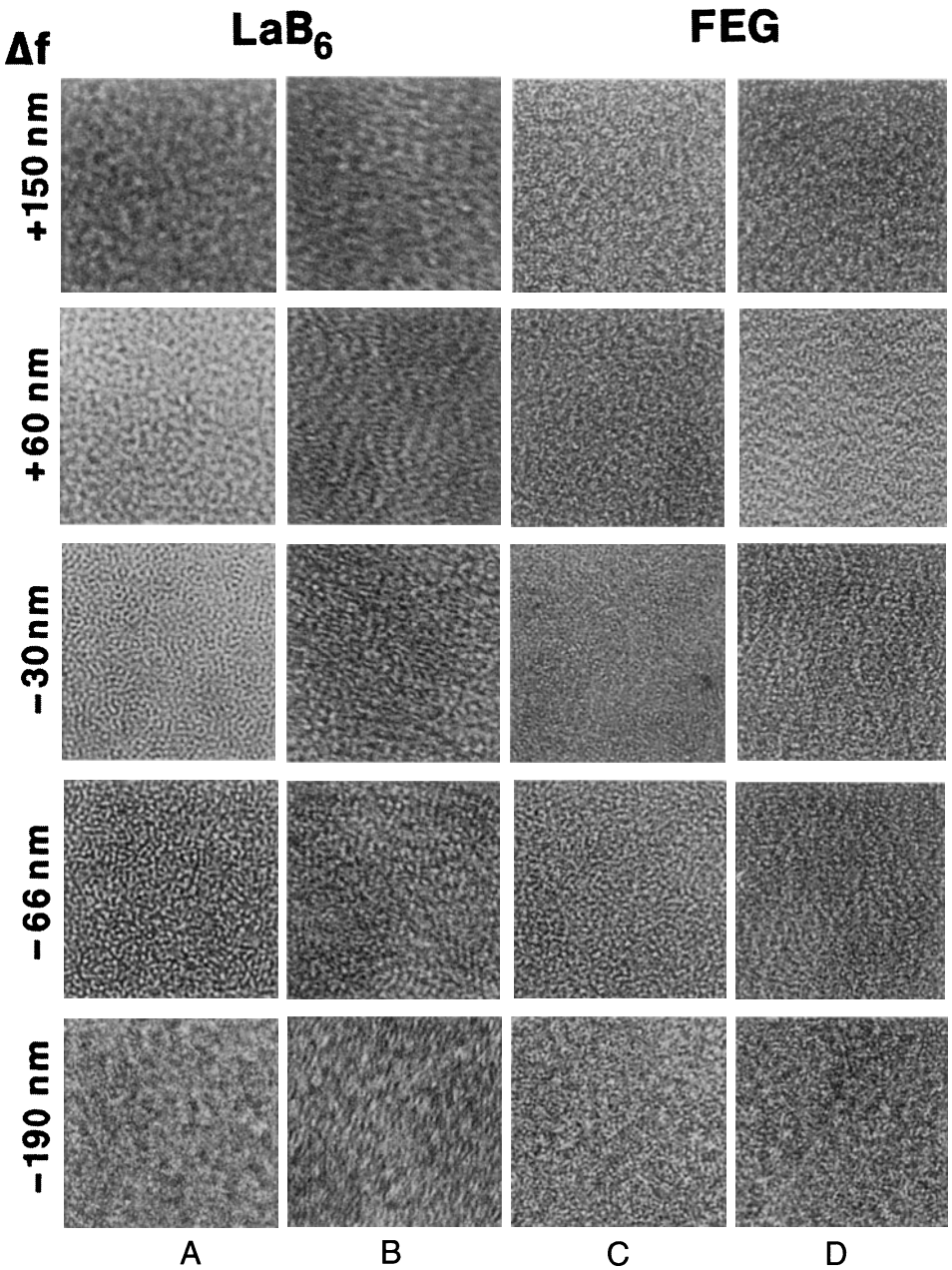


Figure 28.11. A tableau of images from an amorphous film. (A,B) LaB_6 source; (C,D) FEG. (A,C) Without astigmatism, (B,D) with astigmatism. With LaB_6 you can easily see the astigmatism while with an FEG, you can't.

You may notice a similarity between this equation and that for the SAD error (Chapter 11). Two values have been proposed for Δf_{opt}, the optimum defocus setting to minimize delocalization (Coene and Janssen 1992, Lichte 1991). They give an optimum value for the defocus of

$$\Delta f_{opt} = -MC_s\lambda^2 u_{max}^2 \qquad [28.71]$$

where M is a factor between 0.75 and 1. A value for ΔR_{min} is close to

$$\Delta R_{min} = \tfrac{1}{4}C_s\lambda^3 u_{max}^3 \qquad [28.72]$$

The actual value of M is determined by where you define the cut-off value for **u**. There are three conclusions on delocalization:

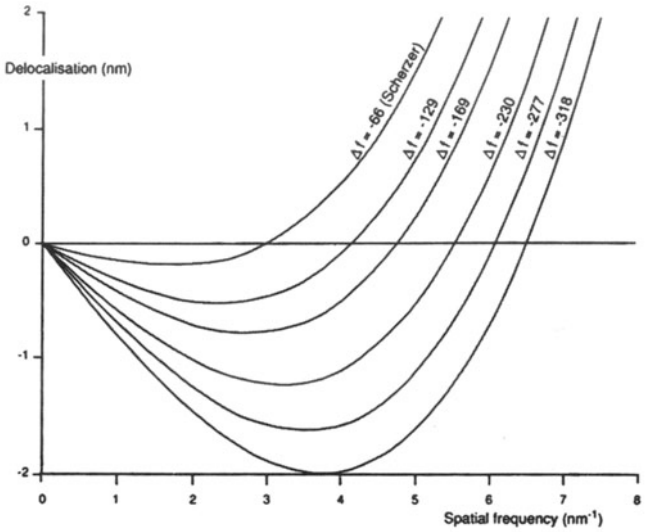


Figure 28.12. Image delocalization plotted as Δf is changed for a Philips CM20 FEG with C_s = 1.2 mm.

- As C_s decreases, delocalization decreases.
- As λ decreases (accelerating voltage increases), delocalization decreases.
- Delocalization cannot be avoided in an FEG!

28.15. SELECTIVELY IMAGING SUBLATTICES

In Chapter 16, where we discussed the 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 (Amelinckx *et al.* 1993). Two domains are present in the combined pattern. Both patterns have 4-fold 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 with a small Σ. This approach has been used to

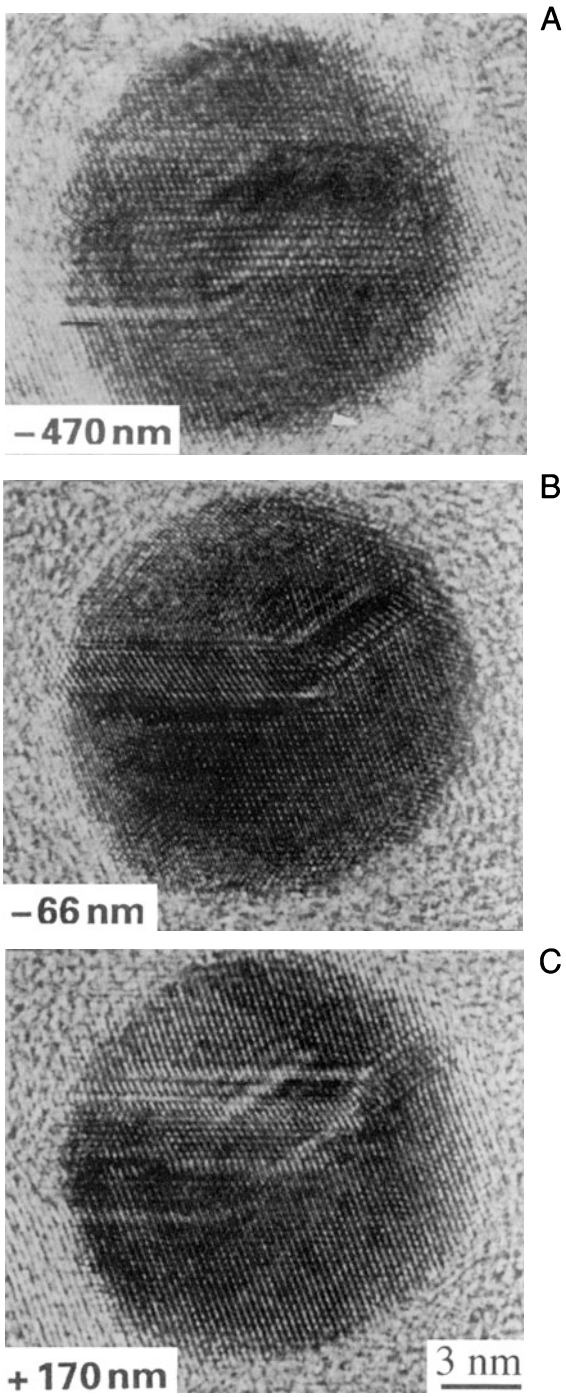


Figure 28.13. Experimental images showing delocalization in HRTEM images of a Au particle: (A) underfocus, (B) Scherzer defocus, (C) overfocus.

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 (Rasmussen *et al.* 1995). The lattice parameter of the spinel is twice that of the NiO but the latter is generally above and below the particle. This ap-

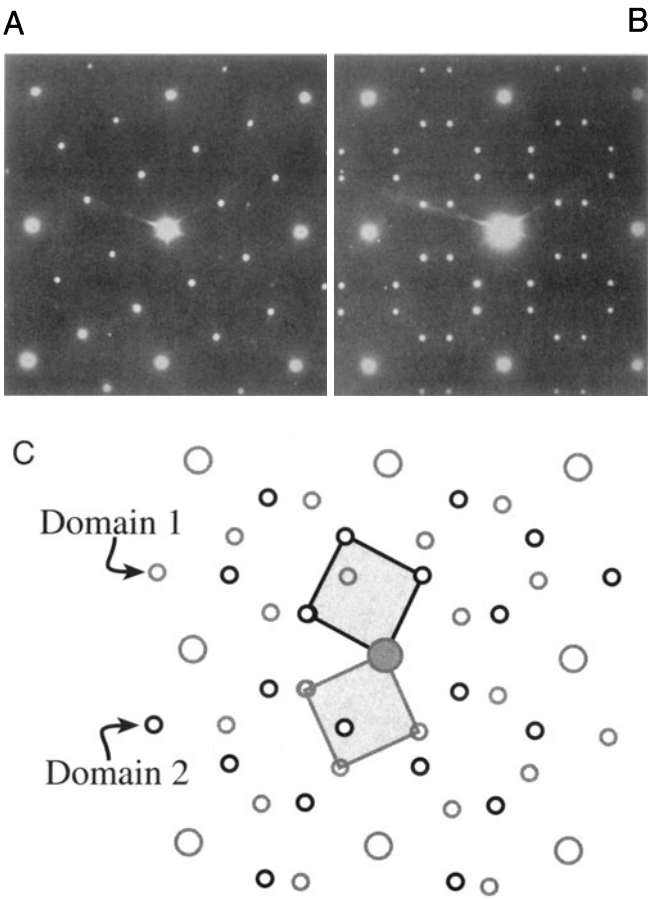


Figure 28.14. Two [001] DPs from an ordered alloy of Au_4Mn. (A) One domain. (B) Two symmetry-related domains. (C) Schematic diagram showing how (B) arises from the relative rotation of the two domains.

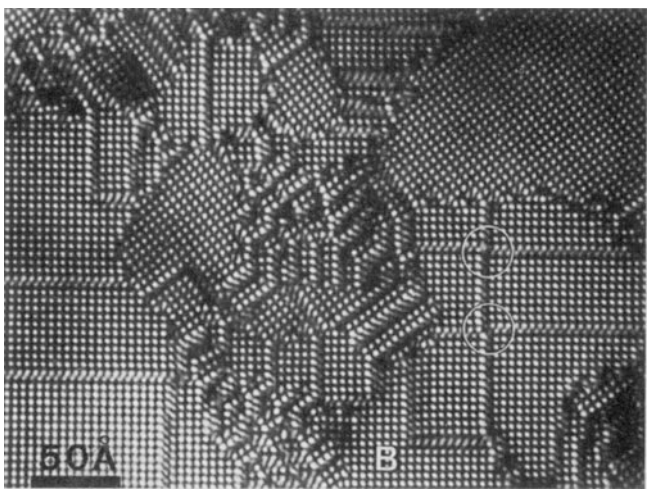


Figure 28.15. DF lattice image of Au_4Mn using an objective diaphragm to exclude all the fcc reflections. The two differently oriented domains correspond to the two orientations in Figure 28.14C.

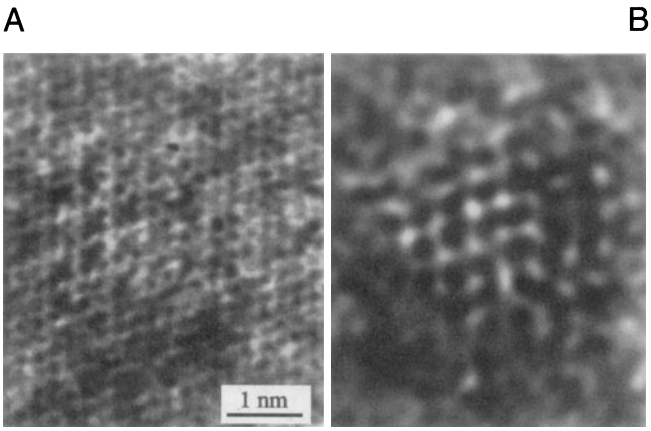


Figure 28.16. (A) The experimental image of a small spinel particle in a NiO matrix. The NiO is thicker and dominates the image. (B) After filtering out the NiO contribution (its lattice parameter is twice that of the spinel) we can see the spinel particle and estimate its size. See also Figures 30.2 and 30.3.

proach 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 29.

28.16. INTERFACES AND SURFACES

Interfaces of all kinds have been extensively studied by HRTEM, in part because we need the near-atomic resolution, but sometimes it's because they make ideal subjects for study! Point defects require extensive image processing and simulation, dislocations 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.

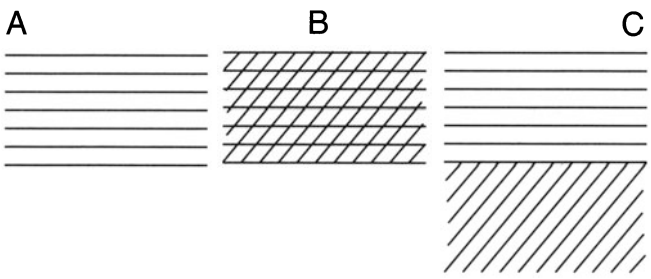


Figure 28.17. Schematic HRTEM images of grain boundaries showing (A) one set of fringes in one grain; (B) specimen tilted to give crossing fringes in one grain; (C) one set of fringes in each grain. In each case the boundary remains parallel to the beam.

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 you can rarely be sure that this is the case, but because you are looking at a very thin specimen, the projected width of even a tilted boundary is small. 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 29.
- 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 chemically abrupt 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? Clearly this question is related to the first.

We will address some of these problems in Chapter 29 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 thickness across the interface. If you do not *know* that this is so, then your interpretation may 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.

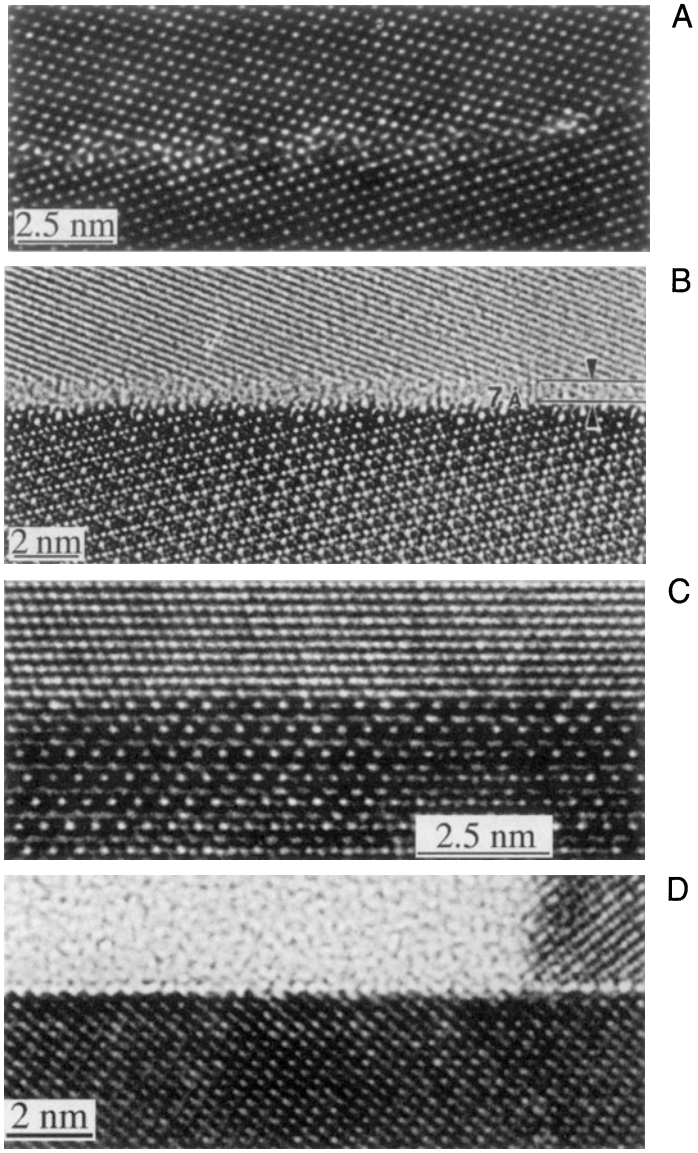


Figure 28.18. Examples of HRTEM images of planar interfaces. (A) Grain boundary in Ge; (B) grain boundary in Si_3N_4 with a layer of glass along the interface; (C) phase boundary separating NiO and $NiAl_2O_4$; (D) profile image of the (0001) surface of Fe_2O_3.

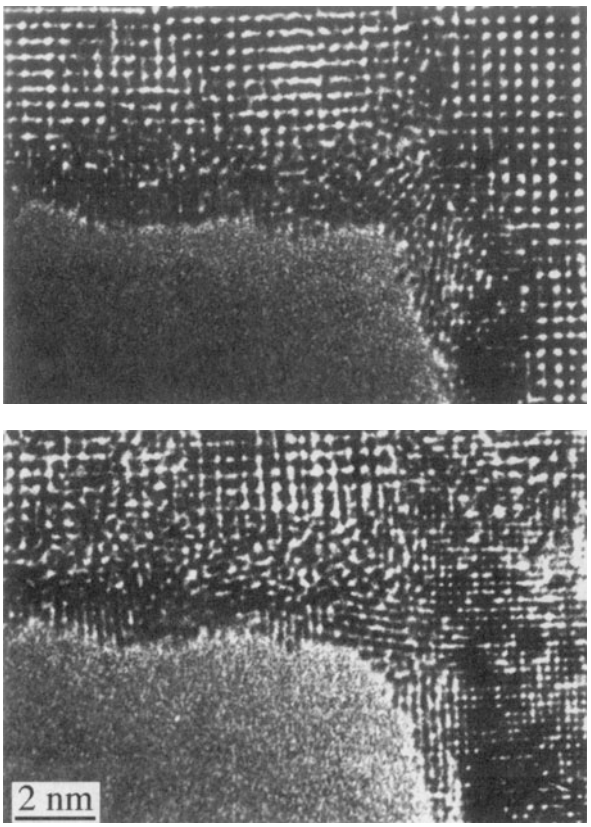


Figure 28.19. Reduction of Nb oxide to the metal by beam-induced loss of oxygen during observation of the edge of the foil. The reduction increases with time (lower image).

- 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 a good vacuum!

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.* (1991) 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.
- 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

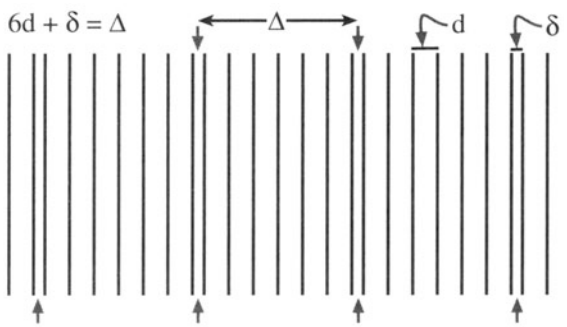


Figure 28.20. An incommensurate structure formed by inserting a planar defect after every seventh layer to expand the lattice by δ every seventh plane.

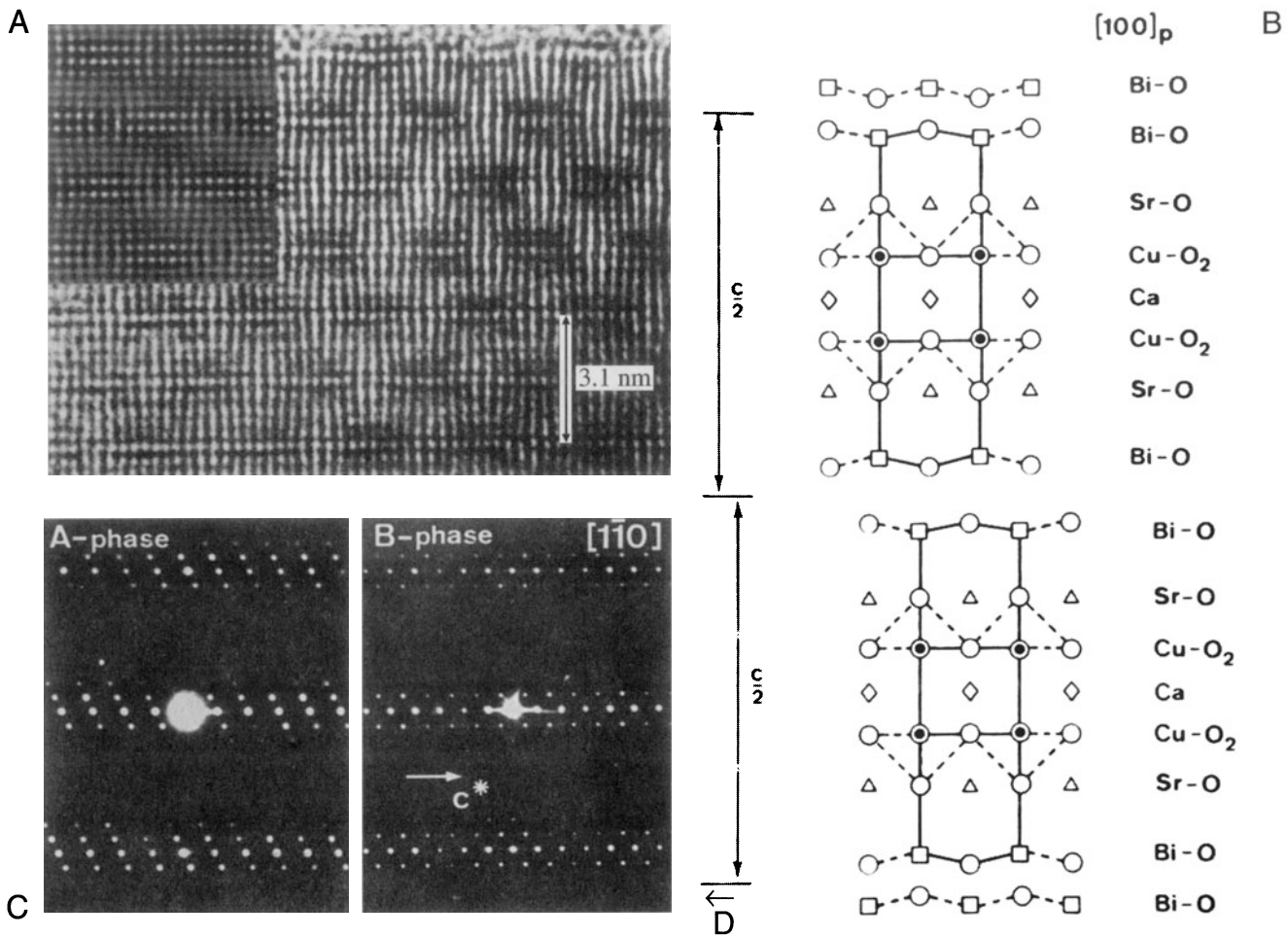


Figure 28.21. (A) HRTEM image of the superconductor $Bi_2Sr_2Ca_nCu_{n+1}O_{2n+\delta}$; the simulation (inset) assumes that the lattice relaxation occurs in the Bi-O layer. (B) For $n = 1$, the structure is built up from blocks which are shifted relative to one another. The DPs are from (C) the n = o phase and (D) the $n = 1.2$ phase. Notice that the spacings of the spots (the satellite sequence) are different.

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 you 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 (Butler and Thomas, 1970). Many ceramics are described as polytypes, e.g., SiC, or polytypoids, which are just polytypes with larger composition fluctuations from layer to layer (e.g. SiAlONs) (Bailey 1977). 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 image the individual modulations, but 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 transla-

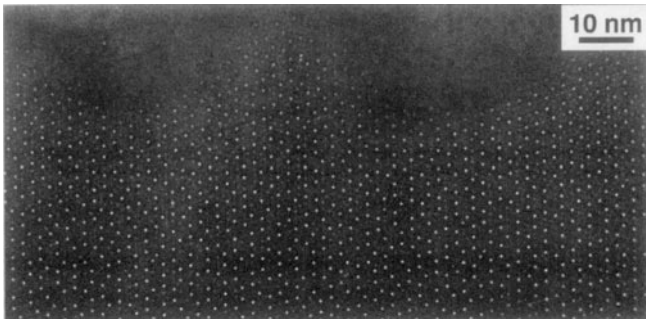


Figure 28.22. Tenfold symmetry in a decagonal Al-Mn-Pd quasicrystal.

tional symmetry which we associate with crystals. However, they are strongly ordered, as you can appreciate from Figure 28.22 (Nissen and Beeli 1991). The HRTEM image shows many sharp white spots from a stable decagonal quasicrystal of Al-Mn-Pd. The DP from another specimen also showed very strong, clear, well-defined spots. In our 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.
- 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! Reimer (1993) gives over 20 references to this topic. The techniques used include phase contrast and amplitude contrast in a conventional TEM, and (see Section 22.4) a dedicated STEM. Parsons *et al.* (1973) used mellitic acid molecules stained with uranyl ions from uranyl acetate so the atoms that were imaged are heavy. Parsons *et al.* 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. The challenge is recognizing that the contrast from the individual uranium atoms reverses as you change defocus, just as we've seen for columns of atoms. You can see this effect in Figure 28.24.

Some points to notice:

- This is a case where we really do have "white atoms or black atoms"!

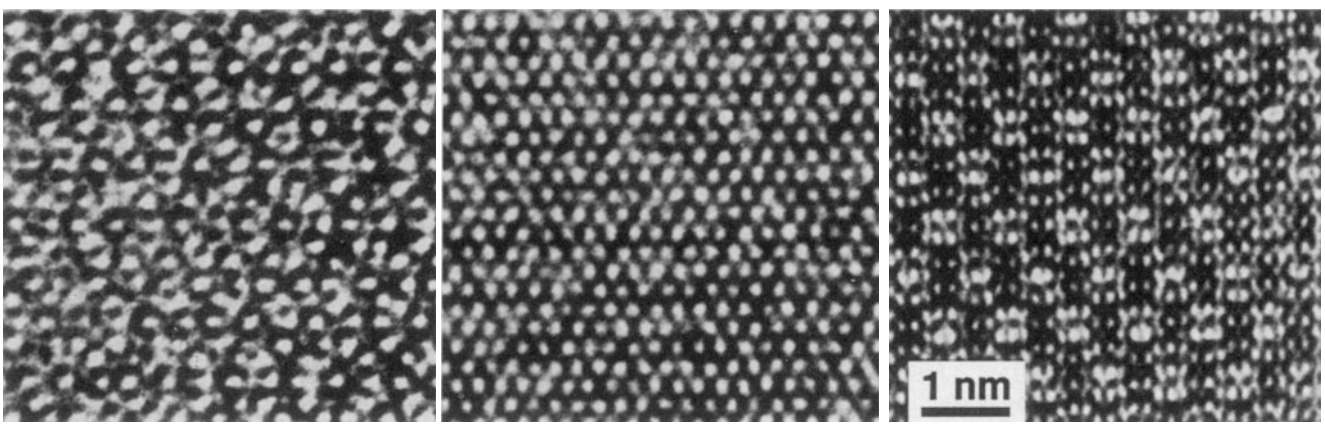


Figure 28.23. Images of (A) fivefold, (B) threefold and (C) twofold projections of an Al-Li-Cu quasicrystal; $\Delta f = -27$ nm.

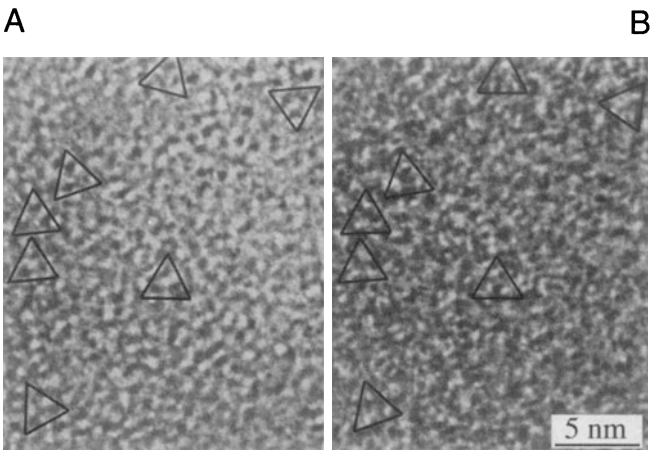


Figure 28.24. Images from triangular arrangements of uranium atoms at different values of defocus.

- Parsons *et al.* (1973) 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 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 were seen to move on the surface, agglomerate, etc. The imaging mode was essentially high-angle dark-field so that the heavy atoms are by far the strongest scatterers and appear bright, as we discussed in Chapter 22. The difficulty with this approach 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 27 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).
- 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 works by Cowley (1981) and Spence (1988), but 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 effect of the specimen.
- Finally, we combine the two models.

We will take the next two chapters to achieve these three tasks. Once you pass those chapters you are ready to attack John Spence's text.

REFERENCES

General References

Buseck, P.R., Cowley, J.M., and Eyring, L. (Eds.) (1988) *High-Resolution Electron Microscopy and Associated Techniques,* Oxford University Press, New York.

Hawkes, P.W. (1980) *Ultramicroscopy* **5**, 67. An enjoyable and informative diversion.

Horiuchi, S. (1994) *Fundamentals of High-Resolution Transmission Electron Microscopy,* North-Holland, Amsterdam.

Spence, J.C.H. (1988) *Experimental High-Resolution Electron Microscopy,* 2nd edition, Oxford University Press, Oxford, United Kingdom.

Vladár, A.V., Postek, M.T., and Davilla, C.D. (1995) *Scanning* **17**, 287. Looks at the SEM as a hi-fi instrument.

Specific References

Amelinckx, S., Milat, O., and Van Tendeloo, G. (1993) *Ultramicroscopy* **51**, 90.

Bailey, S.J. (1977) International Union of Crystallography, *Acta Cryst.* **A33**, 681.

Butler, E.P. and Thomas, G. (1970) *Acta Met.* **18**, 347.

Carter, C.B., Elgat, Z., and Shaw, T.M. (1986) *Phil. Mag.* **A55**, 1.

Coene, W. and Janssen, A.J.E.M. (1992) in *Signal and Image Processing in Microscopy and Microanalysis,* Scanning Microscopy Supplement **6** (Ed. P.W. Hawkes) p. 379, SEM Inc. AMF, O'Hare, Illinois.

Cowley, J.M. (1981) *Diffraction Physics,* 2nd edition, North-Holland, Amsterdam.

Cowley, J.M. (1992) in *Electron Diffraction Techniques.* **1** (Ed. J.M. Cowley) p. 1, I.U.Cr., Oxford Science Publication.

de Jong, A.F. and Van Dyck, D. (1993) *Ultramicroscopy* **49**, 66.

Fejes, P.L. (1977) *Acta Cryst.* **A33**, 109.

Hall, C.E. (1953) *Introduction to Electron Microscopy,* McGraw-Hill, New York.

Hashimoto, H. and Endoh, H. (1978) *Electron Diffraction 1927–1977* (Eds. P.J. Dobson, J.B. Pendry, and C.J. Humphreys), p. 188, The Institute of Physics, Bristol and London.

Lichte, H. (1991) *Ultramicroscopy* **38**, 13.

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 Electronmikroskopie, Halle/Saale, Germany.

Otten, M.T. and Coene, W.M.J. (1993) *Ultramicroscopy* **48**, 77.

Parsons, J.R., Johnson, H.M., Hoelke, C.W., and Hosbons, R.R. (1973) *Phil. Mag.* **29**, 1359.

Rasmussen, D.R., Summerfelt, S.R., McKernan, S.R., and Carter, C.B. (1995) *J. Microscopy* **179**, 77.

Reimer, L. (1993) *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis,* 3rd edition, Springer-Verlag, New York.

Rose, H. (1990) *Optik* **85**, 19.

Rose, H. (1991) in *High Resolution Electron Microscopy: Fundamentals and Applications* (Eds. J. Heydenreich and W. Neumann), p. 6, Institut für Festkörperphysik und Electronmikroskopie, Halle/Saale, Germany.

Shannon, C.E. and Weaver, W. (1964) *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, Illinois.

Thon, F. (1975) in *Electron Microscopy in Materials Science* (Ed. U. Valdrè), p. 570, Academic Press, New York.

Van Dyck, D. (1992) in *Electron Microscopy in Materials Science,* p. 193, World Scientific, River Edge, New Jersey.

Van Dyck, D. and de Jong, A.F. (1992) *Ultramicroscopy* **47**, 266.

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 Electronmikroskopie, Halle/Saale, Germany.

Image Simulation 29

CHAPTER PREVIEW

When we need to obtain information about the specimen in two directions, we need to align the specimen so the beam is 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 TEM's optic axis and the specimen's zone axis. 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 (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.

Image Simulation

29

29.1. 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 27).
- 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 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!

29.2. 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 personal computer with a user-friendly interface. 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 1.5.

29.3. THE RECIPROCAL-SPACE APPROACH

We project each slice onto a plane somewhere in the slice (usually the 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 plane. 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 plane, and so on. The process is summarized in Figure 29.1.

One point which you must remember: scattering by the phase grating does not just produce Bragg beams. It is

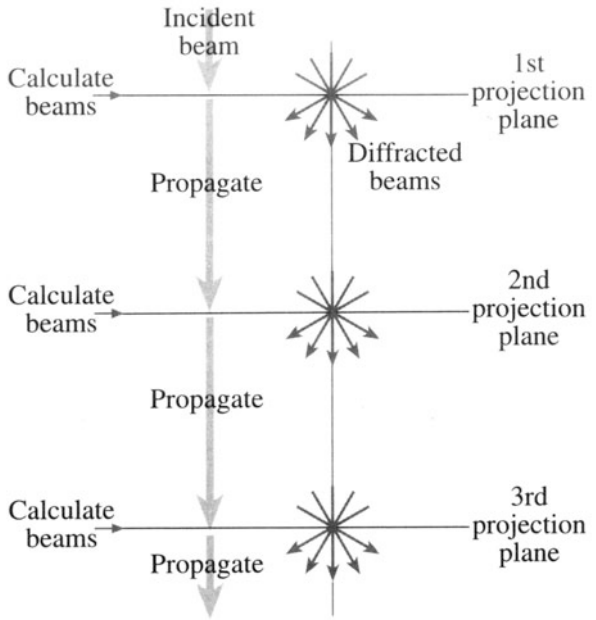


Figure 29.1. The potential within a slice is projected onto the first projection plane; this is the phase grating. We calculate the amplitudes and phases for all the beams generated by interacting with this plane and then propagate all the diffracted beams through free space to the next projection plane, and repeat the process.

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.

Aside: 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., nonperiodic 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 propagator of the electron wave in free space: the *microscope.*
- Q is the phase grating: the *specimen.*

The process can be described by this equation:

$$\psi_{n+1}(\mathbf{k}) = [\psi_n(\mathbf{k}).P_{n+1}(\mathbf{k})] \otimes Q_{n+1}(\mathbf{k}) \qquad [29.1]$$

where $\psi_{n+1}(\mathbf{k})$ is the wave function in reciprocal space at the exit of the n+1 slice and the symbol ⊗ denotes a convolution; $P_{n+1}(\mathbf{k})$ 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 Chapter 2 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 the functions 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 **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 29.2. If you represent the number of 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 29.2B. Putting this into an equation we have

$$\sum_{\mathbf{k}'} F(\mathbf{k}')Q(\mathbf{k} - \mathbf{k}') = \psi(\mathbf{k}) \qquad [29.2]$$

where

$$F(\mathbf{k}) = \psi(\mathbf{k})\, P(\mathbf{k}) \qquad [29.3]$$

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 Section 1.5). 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 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 diffraction pattern, so it does make sense. Specific examples of $Q(\mathbf{k})$, including nu-

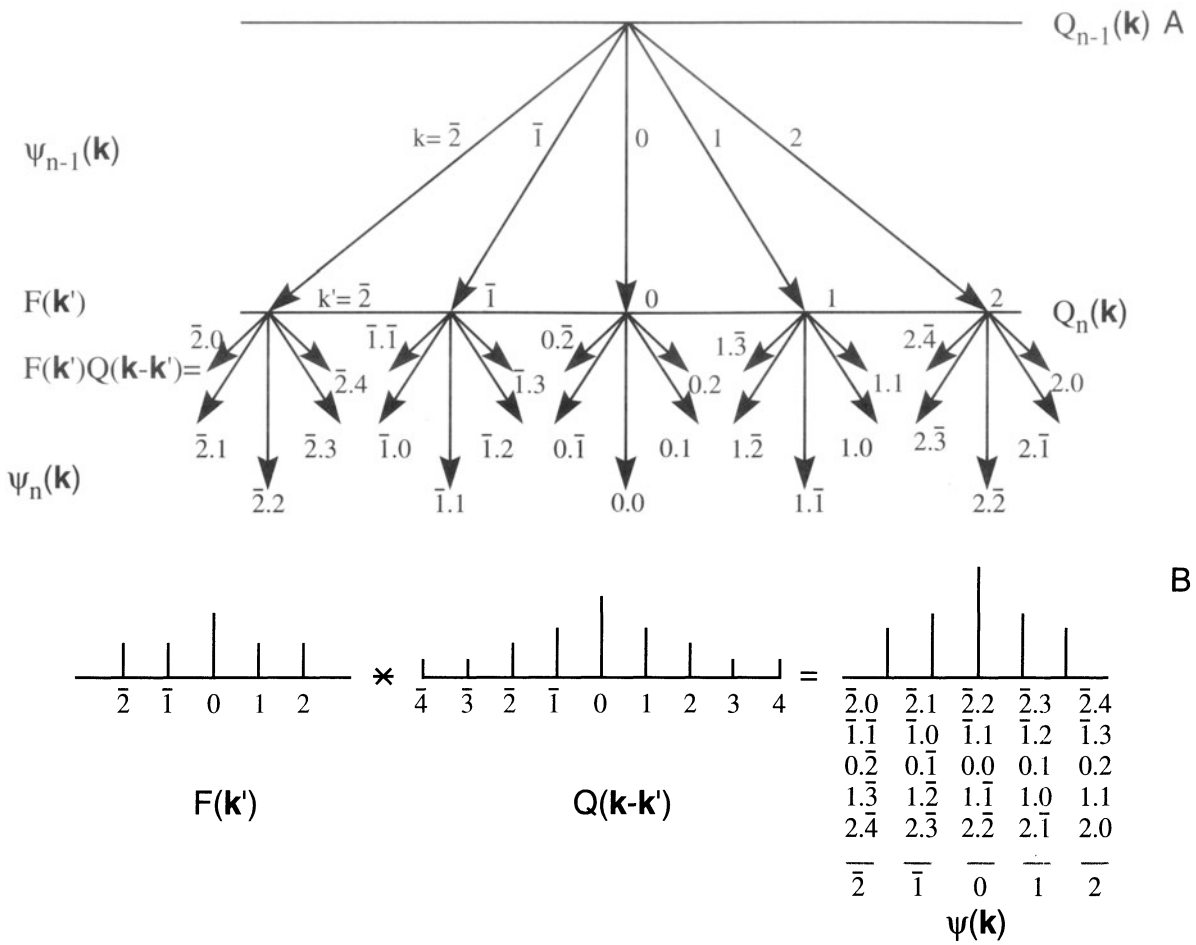


Figure 29.2. (A) Schematic used to explain why, in the one-dimensional case, $Q(\mathbf{k})$ must take account of twice as many **k** values as $\psi(\mathbf{k})$ or $P(\mathbf{k})$. Consider wave $k = \bar{2}$ from $Q_n(\mathbf{k})$: to produce wave +2 at this point you need to add 4 to $\bar{2}$ and 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 posssible combinations of $\mathbf{k}'$ and **k.**

merical computations of the phase change per slice, are given by Barry (1992).

29.4. THE FFT APPROACH

We can recast equation 29.3 to maximize the efficiency of the computer in using fast Fourier transform (FFT) routines. In this equation, F and F^{-1} tell us to take the Fourier transform or the inverse transform of the function inside the brackets

$$\psi_{n+1}(\mathbf{k}) = F\left\{F^{-1}\left[\psi_n(\mathbf{k})\, P_{n+1}(\mathbf{k})\right] q_{n+1}(\mathbf{r})\right\} \qquad [29.4]$$

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 you are limited to 64 points because the Q array must be twice as large in all directions in **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 you 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 pow-

ers 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 O'Keefe and Kilaas (1988) is required reading.

29.5. 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 Coene and Van Dyck (1984a,b) method for calculating $\psi(\mathbf{x})$ can be expressed by the equation

$$\psi_{n+1}(\mathbf{r}) = [\psi_n(\mathbf{r}) \otimes P_{n+1}(\mathbf{r})]\, q_{n+1}(\mathbf{r}) \qquad [29.5]$$

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 $q(\mathbf{x})$.

29.6. 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 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

$$e^{ik_z^{(i)}z} = e^{ik_z^{(j)}D} \qquad [29.6]$$

(Don't confuse the kth Bloch wave with the **k**-vector!)

Using our expression for ψ, namely

$$\psi(\mathbf{r}) = \sum_i C^{(i)}\phi^{(i)}(x, y)e^{ik_z^{(i)}z} \qquad [29.7]$$

and the normalization rule

$$\sum_i C^{(i)}\phi^{(i)}(x, y) = 1 \qquad [29.8]$$

we can therefore express ψ at $z = D$ in terms of our three Bloch waves

$$\psi(x, y, D) = \left[C^{(i)}\phi^{(i)} + C^{(j)}\phi^{(j)}\right]e^{ik_z^{(i)}D} + C^{(k)}\phi^{(k)}e^{ik_z^{(k)}D} \qquad [29.9]$$

We rearrange this equation so that we can extract the phase factor $e^{ik_z^{(i)}z}$ (= $e^{ik_z^{(j)}D}$). We write

$$\psi(x, y, D) = \left[1 - C^{(k)}\phi^{(k)}\right]e^{ik_z^{(i)}D} + C^{(k)}\phi^{(k)}e^{i\left(k_z^{(k)} - k_z^{(i)}\right)D}e^{ik_z^{(i)}D} \qquad [29.10]$$

$$\psi(x, y, D) = e^{ik_z^{(i)}D}\left[1 + \beta_{ik}(D)C^{(k)}\phi^{(k)}\right] \qquad [29.11]$$

where we've defined a new parameter β given by

$$\beta_{ik}(D) = e^{i\left(k_z^{(k)} - k_z^{(i)}\right)D} - 1 \qquad [29.12]$$

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 29.6 but with i, j, and k all equal. Then we can approximate $\beta_{ik}(D)$ by

$$\beta_{ik}(D) \simeq i\left[\left(k_z^{(k)} - k_z^{(i)}\right)D + 2n\pi\right] = i\gamma_{ik}(D) \qquad [29.13]$$

Now we've defined another factor γ_{ik}. If you plug this expression back into equation 29.11, 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 29.11 and 29.13.

- 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 29.3, 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 and the propagation of Bloch waves.

We saw in Chapter 14 that the electron propagates as Bloch waves inside the crystal. The multislice approach, which we usually use to simulate HRTEM images, is actually a very elegant form of brute force. The reason we don't use Bloch waves is that our specimens are not perfect. However, EMS (see Section 1.5) does give you the option of using this approach.

29.7. 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 nonzero for every Bragg reflection. In fact, it will also be different for each type of reflection.

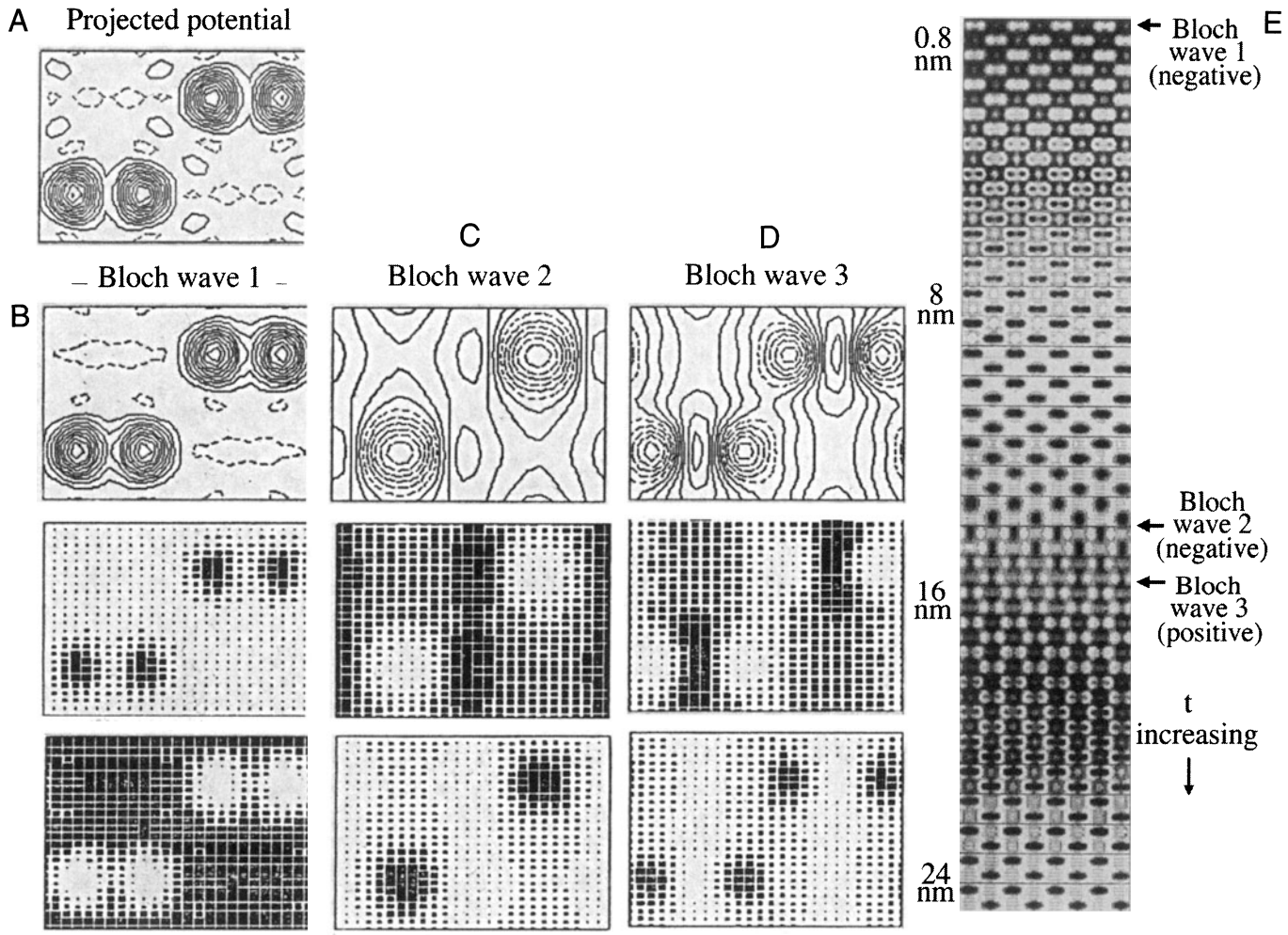


Figure 29.3. (A) The projected potential for Ge where the contour lines represent changes in potential of -10 eV, and the dashed lines are positive values; (B) the amplitudes for Bloch waves 1, 2, and 3 for 100-keV electrons; (C) ideal positive images of the Bloch waves; (D) ideal negative images of the Bloch waves; (E) the thickness dependence of the lattice image.

- 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.

29.8. 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 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, you will not generally have a low-index plane normal to the beam to make this slice.

29.9. 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 (1988) (see also Self and O'Keefe 1988) 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 29.4, 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 ≤ 0.1 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 χ, the phase change due to the objective lens

$$\chi = \pi\Delta f\lambda u^2 + \pi C_s\lambda^3\left(\frac{u^4}{2}\right) \qquad [29.14]$$

Then differentiate this with respect to the variable u

$$\frac{d\chi}{du} = 2\pi\left(\lambda u\Delta f + C_s\lambda^3 u^3\right) \qquad [29.15]$$

This equation tells us that if u changes by δu, then χ changes by

$$\delta\chi = 2\pi\lambda\left(u\Delta f + C_s\lambda^2 u^3\right)\delta u \qquad [29.16]$$

Now we choose $\delta\chi$ so that

$$\delta\chi < \frac{2\pi}{n} \qquad [29.17]$$

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 29.15 and 29.17, we can write

$$\delta u = \left[n\lambda u\left(\Delta f + C_s\lambda^2 u^2\right)\right]^{-1} \qquad [29.18]$$

If we plot χ versus u (or play with equation 29.15 and its derivative), then we find a minimum at

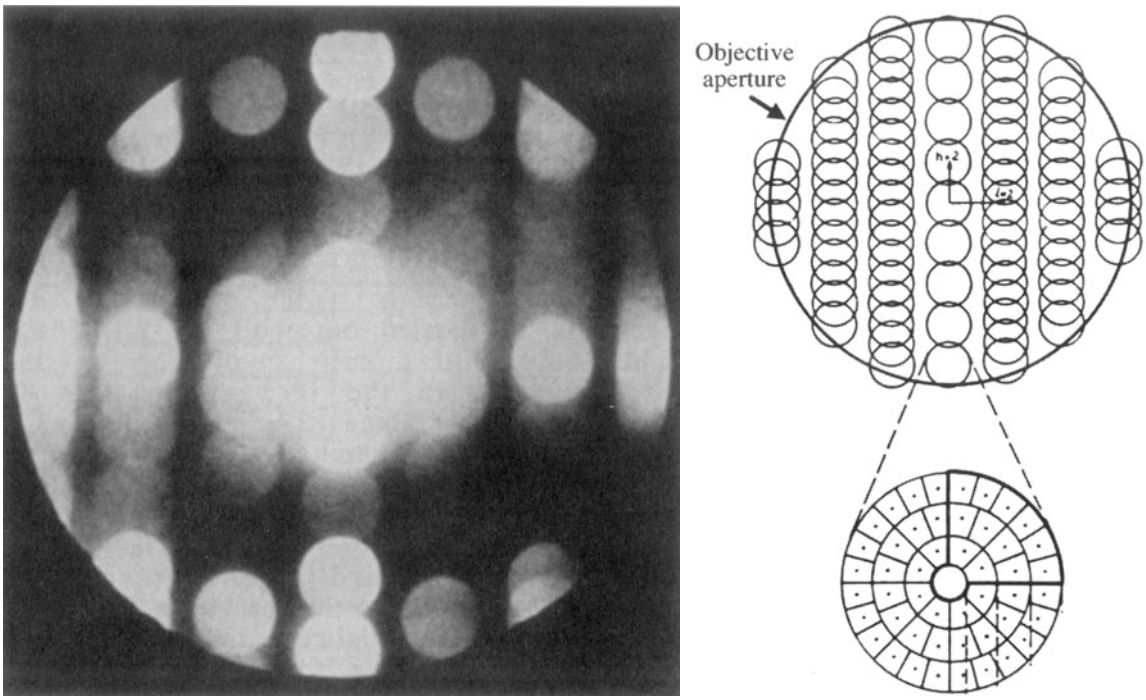


Figure 29.4. Disks in the DP from a crystal of $Nb_{12}O_{29}$. The computer simulation can divide each disk into many sectors and simulate the image for each sector, as shown in the schematic, excluding sectors which are intersected by the objective aperture.

$$\Delta f = -C_s\lambda^2 u^2 \qquad [29.19]$$

and an inflection at

$$\Delta f = -3C_s\lambda^2 u^2 \qquad [29.20]$$

So the simulation program can check to find the smallest δu at an inflection point, which equations 29.18 and 29.20 tell us is

$$\delta u = -\left[\frac{27C_s}{(\Delta f)^3}\right]^{\frac{1}{2}}\left(\frac{1}{2n}\right) \qquad [29.21]$$

The value of δu therefore depends on both C_s and Δf.

Remember that all this calculation takes place in that black box!

You can also appreciate the relevance of this type of approach if your disks actually intersect the objective aperture, as shown in Figure 29.4. 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.

29.10. MODELING THE STRUCTURE

To simulate any HRTEM image, you need a unit cell. If you are only concerned 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 grain boundaries 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 30 when we discuss quantitative HRTEM.

29.11. SURFACE GROOVES AND SIMULATING FRESNEL CONTRAST

The analysis of interfaces by the Fresnel-fringe technique, which we introduced in Chapter 27, 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 29.5A:

- 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 grain boundaries, 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. 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, an outer width a_0, and a "diffuseness," δ, defined by

$$a_0 = (1 + \delta)a \qquad [29.22]$$

These parameters are shown in Figure 29.5B. Then we construct models of a foil with a surface groove at the edge-on interface by combining such potentials.

The models: Values of δ = 0.5 and δ = 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

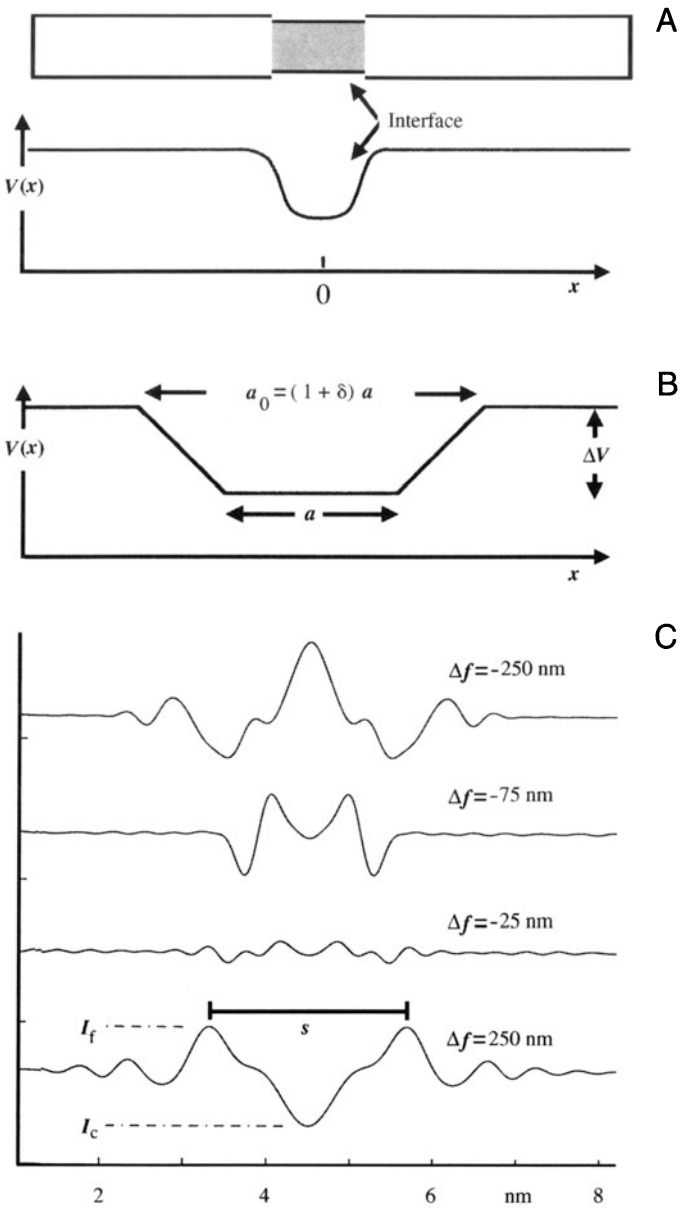


Figure 29.5. (A) Schematic of a grain boundary containing a layer of material with a different inner potential; (B) one model used to represent such a grain boundary giving variable parameters a, a_0, and δ; (C) a typical set of simulated Fresnel-fringe intensity profiles at increasing Δf: s is the distance between the first two fringes, I_c and I_f are the intensities of the central and first fringe, respectively.

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.
- The 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, i.e., $\Delta f \sim t$, 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.

In the case of Fresnel fringes from an edge, 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 (Clarke 1979) 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, we refer you to the original articles and the papers by Taftø *et al.* (1986), Rasmussen and Carter (1990), Stobbs and Ross (1991), and Zhu *et al.* (1995).

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 low defocus. For diffuse interfaces, the contrast decreases rapidly as Δf approaches zero (Figure 29.5C), and measurements of the fringe spacing 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, e.g., the diffuseness, you 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, you may easily overestimate the interface width. The atomic density in a region close to the boundary is often reduced, even if the boundary is structured, so you can easily misinterpret the image as showing the presence of an intergranular film when it is actually film-free (Simpson *et al.* 1986).

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.

The conclusion is that you must use all the information in the image to characterize the shape of the potential well.

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), this technique won't work.

To summarize, this discussion gives us a method for analyzing Fresnel fringes from a grain boundary. We can draw some conclusions:

- To interpret the contrast from Fresnel fringes at grain boundaries, 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.

29.12. CALCULATING IMAGES OF DEFECTS

When we simulate HRTEM images of perfect crystals, we only need 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 29.6: 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 grain boundary is shown in Figure 29.6. 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 grain boundaries 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.

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 (1982).

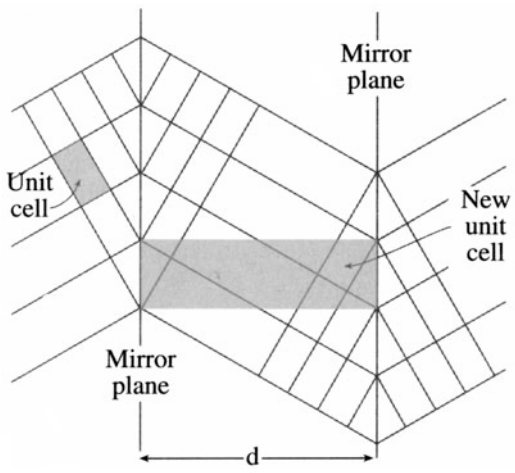


Figure 29.6. The periodic continuation technique illustrating how an artificial unit cell can be constructed to contain two grain boundaries, thus allowing the HRTEM image to be simulated. The distance (d) between the two interfaces can be varied to check for overlap artifacts.

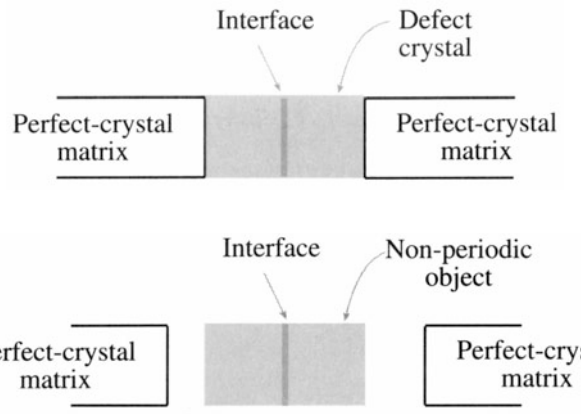


Figure 29.7. In the real-space patching method, the defect crystal (in this case the interface and several adjacent layers) is a nonperiodic object which is surrounded by a perfect-crystal matrix.

An alternative approach to the periodic continuation approach has been developed by Coene *et al.* (1985) 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 29.7; the image from each patch is 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 that 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. As explained by Coene *et al.* (1985), the defect does not now "see" its own image, it only sees the perfect matrix on all sides.

29.13. SIMULATING QUASICRYSTALS

There are several problems in simulating HRTEM images of quasicrystals, not least of which concerns which model you should use. Several models have been reviewed by Shoemaker (1993) and the possibilities are illustrated by the work of Beeli and Horiuchi (1994), who used a combination of 10 layers in the multislice calculation. The layers are made up from the planar structures shown in Figure 29.8. The final structure (shown in Figure 29.8A) is made up of two sets of five layers. The first set of layers

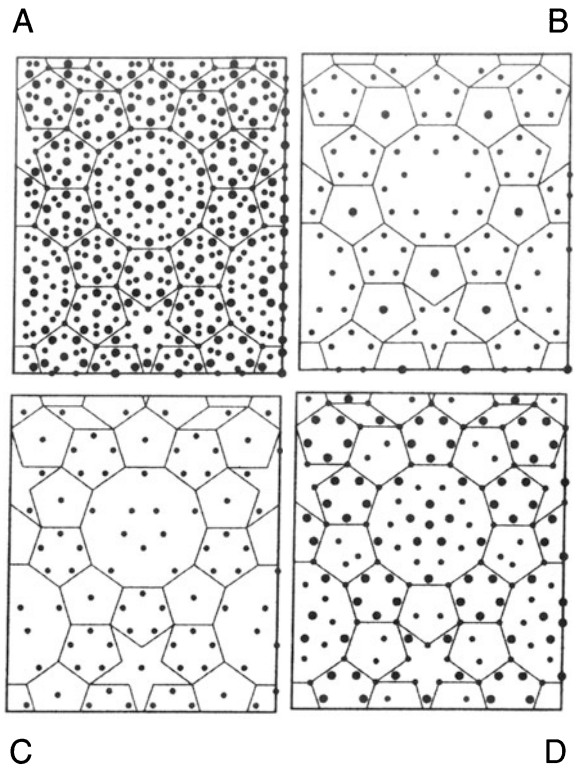


Figure 29.8. Projections used to simulate images of $Al_{70}Mn_{17}Pd_{13}$ quasicrystals. (A) Combination of all the layers; the (B, C, D) layers are used to contribute (repeatedly) to (A). The edges of all the tiles are 0.482 nm. The large circles denote Al atoms.

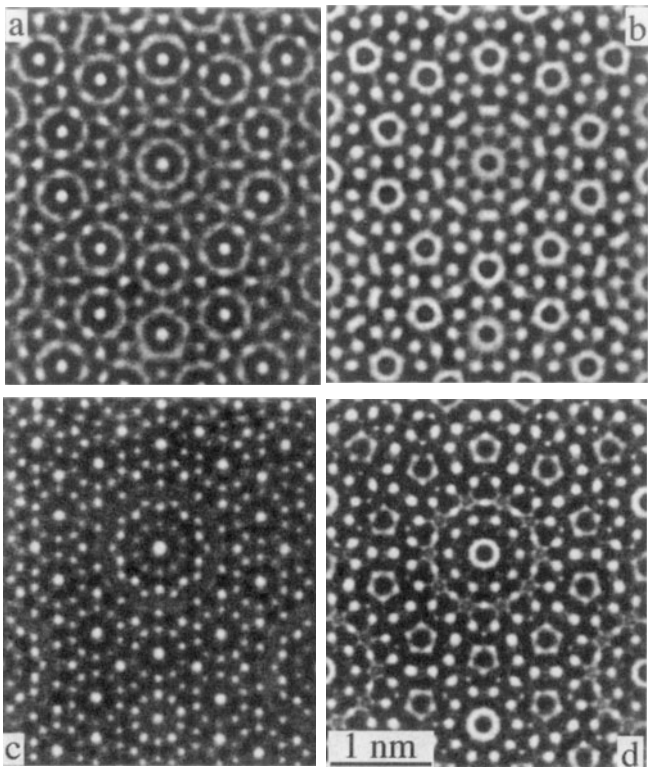


Figure 29.9. Four simulated images of the model constructed from the layers shown in Figure 29.8 using only Al and Mn atoms. The thickness is 3.77 nm, which corresponds to three periods in the beam direction. The values of Δf are (a) 0 nm, (b) 46 nm, (c) 88 nm, and (d) 124 nm.

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 that is 2.04 nm in diameter and the center part of a pentagon tile. As you realize, the problems in such an image calculation are increased because the quasicrystal does not have translational symmetry, but you must impose such a symmetry to do the calculation. The calculation was then carried out for thicknesses up to 10 nm.

The results of such calculations with only Al and Mn atoms are illustrated in Figure 29.9. The edges of the cells are essentially artificial because, as we just noted, the structure used in the calculation is a "unit cell." 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 29.10.

Another illustration of the success of HRTEM comes from the work of Jiang *et al.* (1995) 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 29.11 with the B′ layer being a 45°-rotated B layer, i.e., the B and B′ layers are again related by a screw axis, but 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.

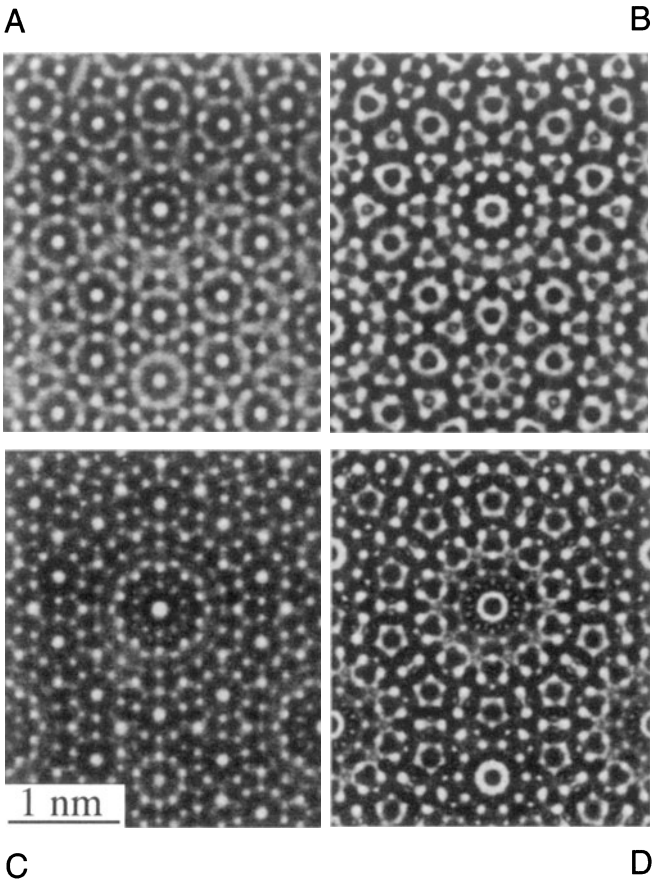


Figure 29.10. Examples of simulated images of the quasicrystal shown in Figure 29.8 but substituting Pd atoms for Mn atoms. The values of Δf are (A) 0 nm, (B) 48 nm, (C) 88 nm, and (D) 128 nm.

29.14. BONDING IN CRYSTALS

We mentioned early on that one problem we have with image 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 (1968) and Doyle and Cowley (1974). 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.* (1970) give tabulated results while Tang and Dorignac (1994) have made detailed comparisons for HRTEM imaging.

O'Keefe and Spence (1994) have re-examined the meaning of the mean inner potential. One of the reasons that you 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.

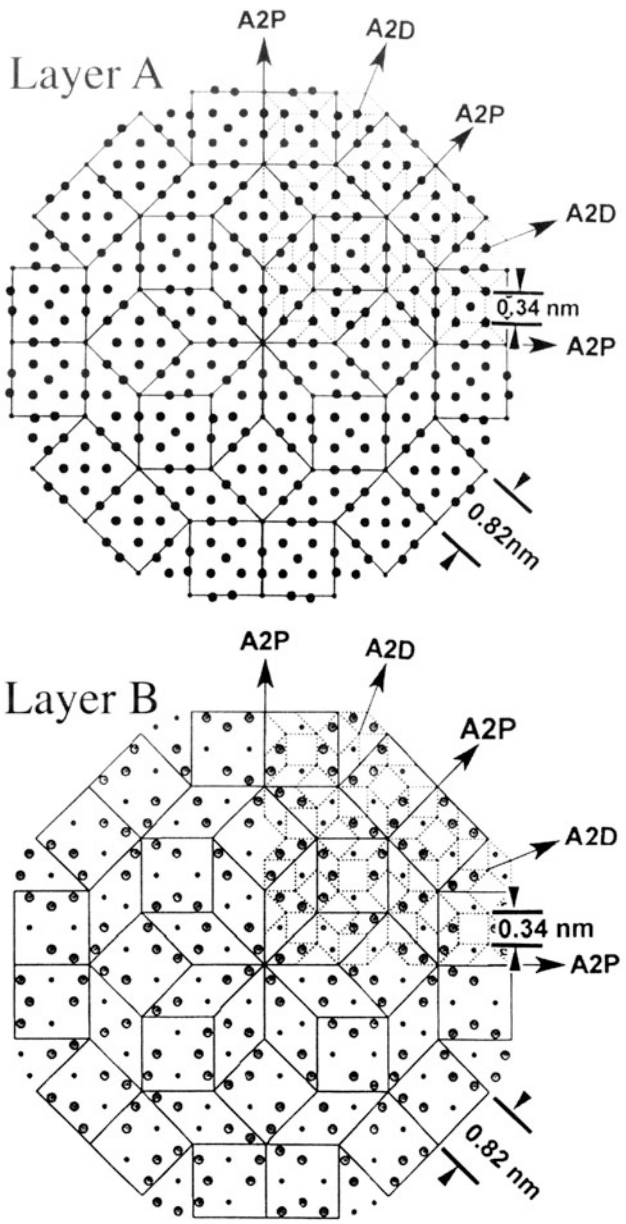


Figure 29.11. The model used to simulate quasicrystals with eightfold symmetry. The structure for the simulation was constructed as a four-layer sequence ABAB' where the B and B' layers are related by an 8_4 screw axis.

While this is an evolving study, some important results have been obtained:

- The inner potential is very sensitive to bonding effects. O'Keefe 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'Keefe and Spence is highly recommended reading for those who have a strong physics background but think TEM is a "known" subject!

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, as we'll see in Chapter 30. Most materials scientists using TEM will want to use one of the established software packages we listed in Section 1.5. 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 grain boundaries. 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.
- Make sure that you accurately align your TEM before you record any images.
- If possible, use more than one program to simulate the images.
- 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.

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.

REFERENCES

General References

Buseck, P.R., Cowley, J.M., 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, Pennsylvania. A collection of defocused review articles.

O'Keefe, M.A. and Kilaas, R. (1988) in *Advances in High-Resolution Image Simulation in Image and Signal Processing in Electron Microscopy,* Scanning Microscopy Supplement 2. (Eds. P.W. Hawkes, F.P. Ottensmeyer, W.O. Saxton, and A. Rosenfeld), p. 225. SEM Inc., AMF O'Hare, Illinois. This article is another excellent introduction to the subject.

Spence, J.C.H. (1988) *Experimental High-Resolution Electron Microscopy,* 2nd edition, Oxford University Press, New York. This is *the* text for users of the HRTEM.

Specific References

Barry, J. (1992) in *Electron Diffraction Techniques* **1**, (Ed. J.M. Cowley), p. 170, I.U.Cr., Oxford Science Publication, New York.

Beeli, C. and Horiuchi, S. (1994) *Phil. Mag.* **B70**, 215.

Carlson, T.A., Lu, T.T., Tucker, T.C., Nestor, C.W., and Malik, F.B. (1970) Report ORNL-4614, ORNL, Oak Ridge, Tennessee.

Clarke, D.R. (1979) *Ultramicroscopy* **4**, 33.

Coene, W. and Van Dyck, D. (1984) *Ultramicroscopy* **15**, 41.

Coene, W. and Van Dyck, D. (1984) *Ultramicroscopy* **15**, 29.

Coene, W., Van Dyck, D., Van Tendeloo, G., and Van Landuyt, J. (1985) *Phil. Mag.* **52**, 127.

Cowley, J.M. and Moodie, A.F. (1957) *Acta Cryst.* **10**, 609.

Doyle, P.A. and Cowley, J.M. (1974) in *International Tables for X-ray Crystallography,* Vol. **IV**, pp. 152–173, Kluwer Academic Publ., Dordrecht, the Netherlands.

Doyle, P.A. and Turner, P.S. (1968) *Acta. Cryst.* **A24**, 309.
Fujimoto, F. (1978) *Phys. stat. sol. (a)* **45**, 99.
Jiang, J.-C., Hovmöller, S., and Zou, X.-D. (1995) *Phil. Mag.* **71**, 123.
Kambe, K. (1982) *Ultramicroscopy* **10**, 223.
Kouh, Y.M., Carter, C.B., Morrissey, K.J., Angelini, P., and Bentley, J. (1986) *J. Mat. Sci.* **21**, 2689.
O'Keefe, M. and Spence, J.C.H. (1994) Acta. Cryst. **A50**, 33.
Rasmussen, D.R. and Carter, C.B. (1990) *Ultramicroscopy,* **32**, 337.
Ross, F.M. and Stobbs, W.M. (1991) *Phil. Mag.* **A63**, 1.
Self, P.G. and O'Keefe, M.A. (1988) in *High-Resolution Transmission Electron Microscopy and Associated Techniques,* (Eds. P.R. Buseck, J.M. Cowley, and L. Eyring), p. 244, Oxford University Press, New York.
Shoemaker, C.B. (1993) *Phil. Mag.* **B67**, 869.
Simpson, Y.K., Carter, C.B., Morrissey, K.J., Angelini, P., and Bentley, J. (1986) *J. Mater. Sci.* **21**, 2689.
Taftø, J., Jones, R.H., Heald, S.M. (1986) *J. Appl. Phys.* **60**, 4316.
Tang, D. and Dorignac, D. (1994) *Acta. Cryst.* **A50**, 45.
Wilson, A.R. and Spargo, A.E.C. (1982) *Phil. Mag.* **A46**, 435.
Zhu, Y., Taftø, J., Lewis, L.H., and Welch, D.O. (1995) *Phil Mag. Lett.* **71**, 297.

Quantifying and Processing HRTEM Images

30

CHAPTER PREVIEW

In this chapter we will equate our title with the use of the computer. We will simply use image processing to extract more information from a TEM image than we can obtain by eye. 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 our 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, and how to display the result!

Quantifying and Processing HRTEM Images

30

30.1. 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.

> The basic idea of image processing is changing images into numbers and manipulating the numbers.

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 back 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 between 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.

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.

30.2. 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.
- We may want to quantify the information contained in the image.

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.

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 freeware 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 be unprocessed.

30.3. 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 29, 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. This reporting is particularly critical when the raw data (the "original" image) are 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.

30.4. 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 CRT, 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 video tape.
- Record the image on film, then print it and use a flat-bed 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 to CCD cameras is that high-quality CCD chips are very expensive for 1k × 1k arrays and astronomically expensive for 2k × 2k arrays. Although such cameras may become routine add-ons for all TEMs in the near future, even when you do have such a camera you will probably also want to use film or video. With film, you can record a larger area than you can using a CCD; you should use a video-recorder for *in situ* studies when using a heating or straining holder.

We can transfer the image from a video tape 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 with a resolution of 1k × 1k cost about the same as a video 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.

30.5. PROCESSING TECHNIQUES

30.5.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 an 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. This is not possible in a modern TEM with normal fixed-diameter objective apertures. A single

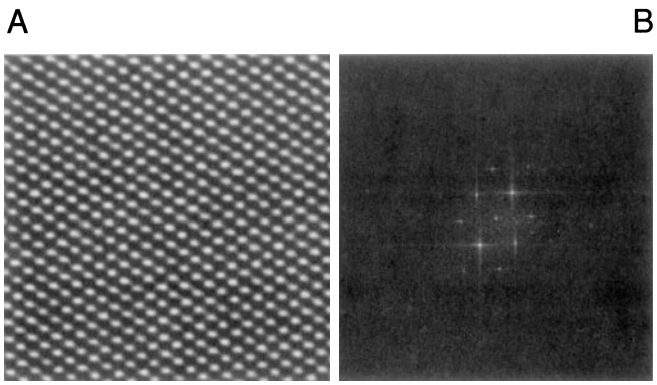


Figure 30.1. A square mask has been used to select the area shown in (A) from a much larger print of the image. The Fourier transform of this region is shown in (B) where you can see not only the spots in the 110 DP but also long streaks which run normal to the edges of the mask.

variable SAD aperture was used on some early TEMs, but it was triangular in shape and used three movable blades, so it was not much use as an objective aperture. You can best understand the procedure by an example. A square mask was used to select the region in Figure 30.1A from a much larger region of the HRTEM image, and its Fourier

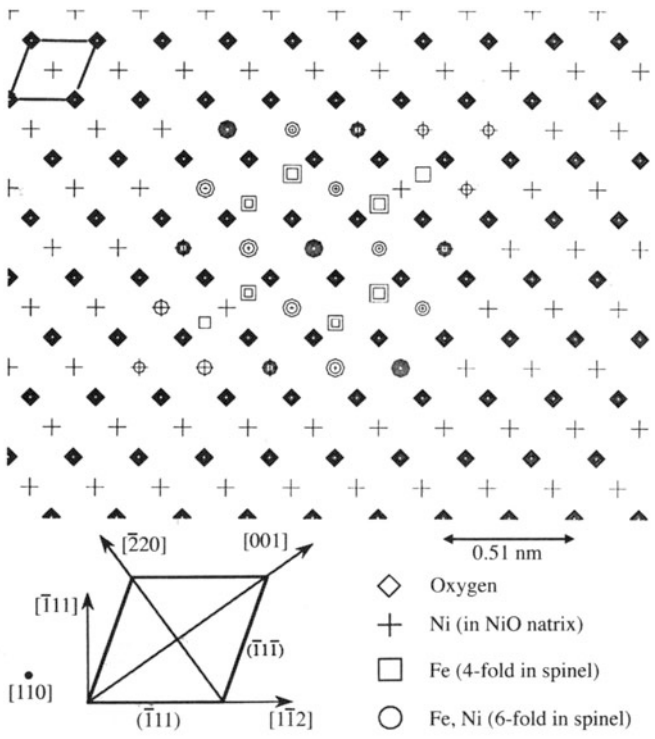


Figure 30.2. A model of an octahedron of spinel which is fully enclosed within a matrix of NiO. The rest of the specimen could then be modeled by adding extra layers of NiO above or below the defect layer.

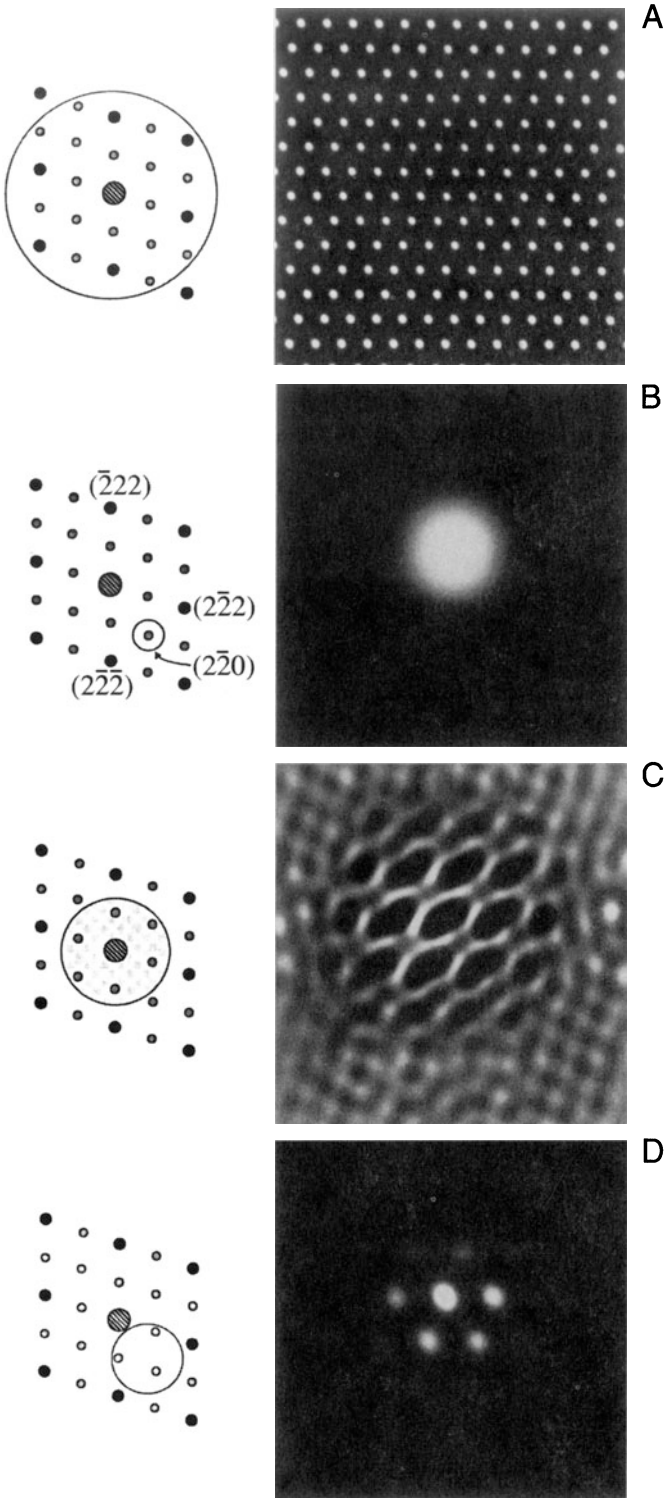


Figure 30.3. The DP from a specimen containing a particle such as that shown in Figure 30.2 is shown schematically in (A) together with the resulting lattice image. The other three pairs of diagrams illustrate how we can use the computer to produce different masks and thus generate different images, such as the DF image in (B). The image in (C) corresponds to the image shown in Figure 28.16, while the DF lattice image shown in (D) is analogous to that discussed in Figures 28.14 and 28.15.

transform (i.e., DP effectively from a few nanometers) is shown in Figure 30.1B.

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 as in the "real" TEM because the information about anything other than the perfect lattice is carried between the reciprocal lattice points. Figure 30.2 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 30.3. This was done using the Digital Micrograph package (Section 1.5).

30.5.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 you had 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.

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.

We thus record the image at high resolution, preferably directly using a slow-scan CCD camera, although digitizing the negative is fine. Then, by comparing your experimental plot of I versus **u** with those calculated for different values of Δf and C_s you 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 30.4. Notice that as the defocus of the objective lens increases, the number of rings increases but the rings become narrower. The contrast transfer gradually extends to larger values of **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 30.5, even a small amount of astigmatism can be detected by eye. The computer can readily measure and correct such a pattern, as we will see shortly. This set of diffractograms also shows that the computer can distinguish astigmatism and drift in the image while your 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.

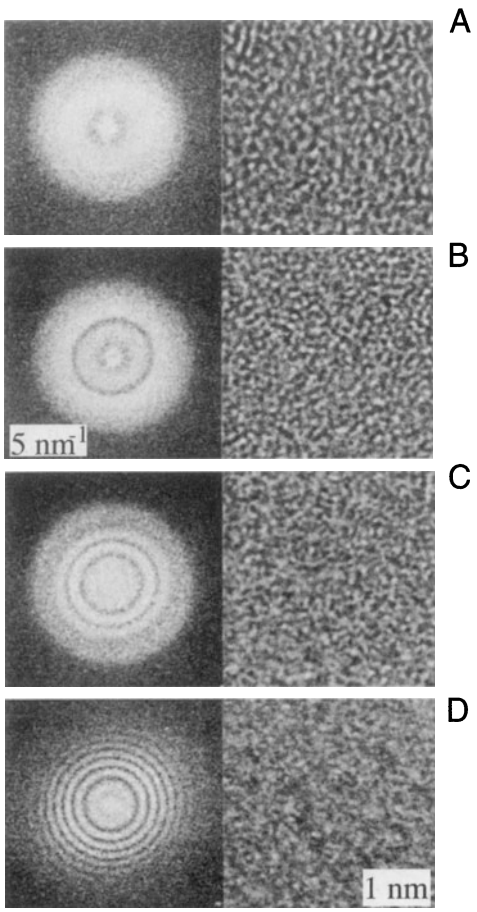


Figure 30.4. Four images of an amorphous Ge film and their corresponding diffractograms. Δf has the following values: (A) 1 sch; (B) 1.87 sch; (C) 2.35 sch; (D) 3.87 sch. 1 sch = $-(C_s\lambda)^{1/2}$.

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 $\sin \chi(\mathbf{u}) = 0$.

$$\sin \chi(\mathbf{u}) = 1 \quad \text{when} \chi(\mathbf{u}) = \frac{n\pi}{2} \text{ and } n \text{ is odd} \quad [30.1]$$

$$\sin \chi(\mathbf{u}) = 0 \quad \text{when} \chi(\mathbf{u}) = \frac{n\pi}{2} \text{ and } n \text{ is even} \quad [30.2]$$

Since C_s will also influence the location of the rings, you need at least two rings. Krivanek (1976) has given a simple

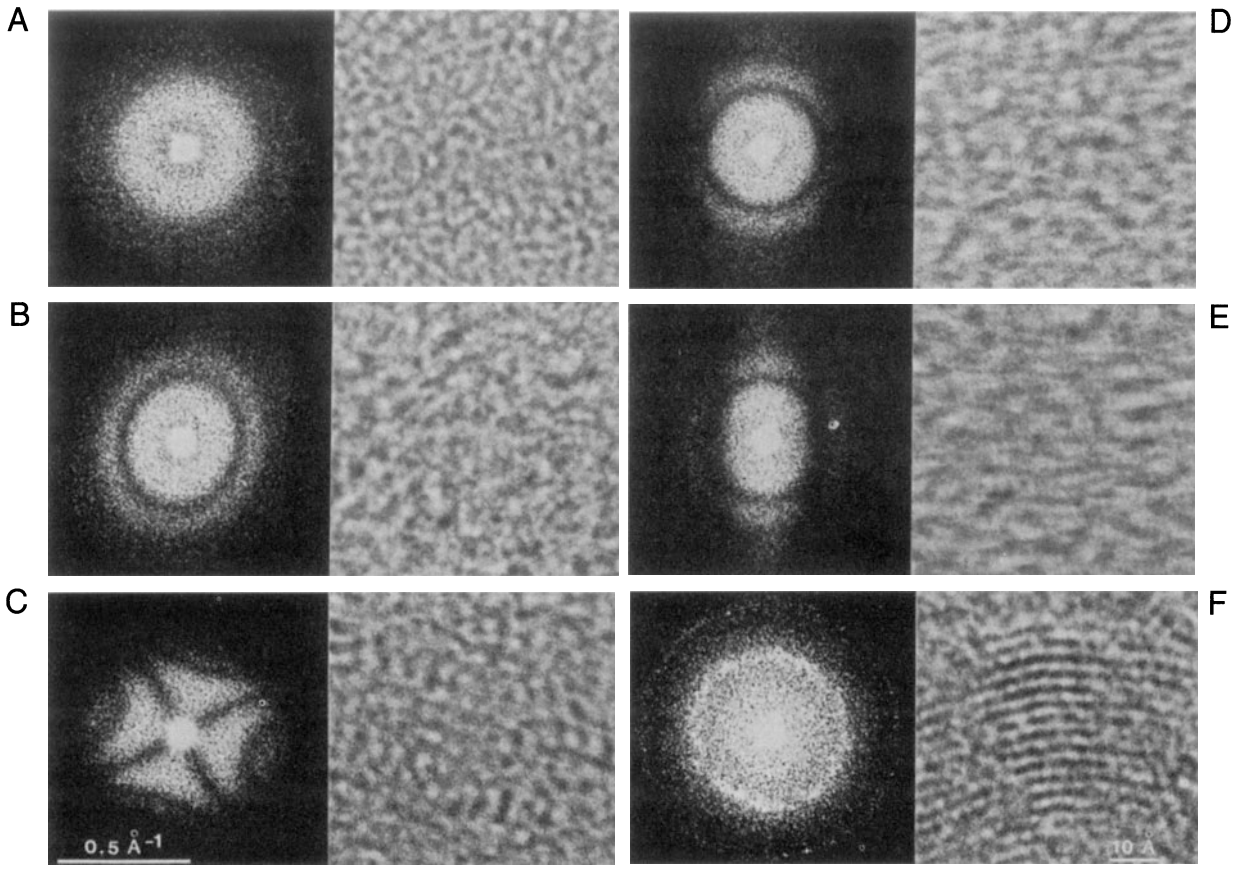


Figure 30.5. Six images of an amorphous carbon film and their corresponding diffractograms illustrating different misalignments of a 300-kV HRTEM: (A) well aligned and no drift; (B) some astigmatism (C_a = 14 nm); (C) more astigmatism (C_a = 80 nm); (D) no astigmatism but drifted 0.3 nm; (E) no astigmatism but drifted 0.5 nm; (F) well aligned and no drift showing graphite calibration fringes of 0.344-nm spacing. (B,D,E) $\Delta f = -2.24$ sch; (C) $\Delta f = 0$.

procedure for finding both C_s and Δf. If we start with our definition of χ

$$\chi(\mathbf{u}) = \pi \Delta f \lambda u^2 + \frac{1}{2}\pi C_s \lambda^3 u^4 \qquad [30.3]$$

then, inserting the values given in equations 30.1 and 30.2 leads to

$$\frac{n}{u^2} = C_s \lambda^3 u^2 + 2\Delta f \lambda \qquad [30.4]$$

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. You should find that your value of C_s will be 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 30.6A. You can use these hyperbolas to determine C_s for any microscope and Δf for any diffractogram (Krivanek 1976).

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 30.6B 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 ± θ° tilt will only look the same (though rotated) if the beam had zero tilt at θ = 0°. 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.

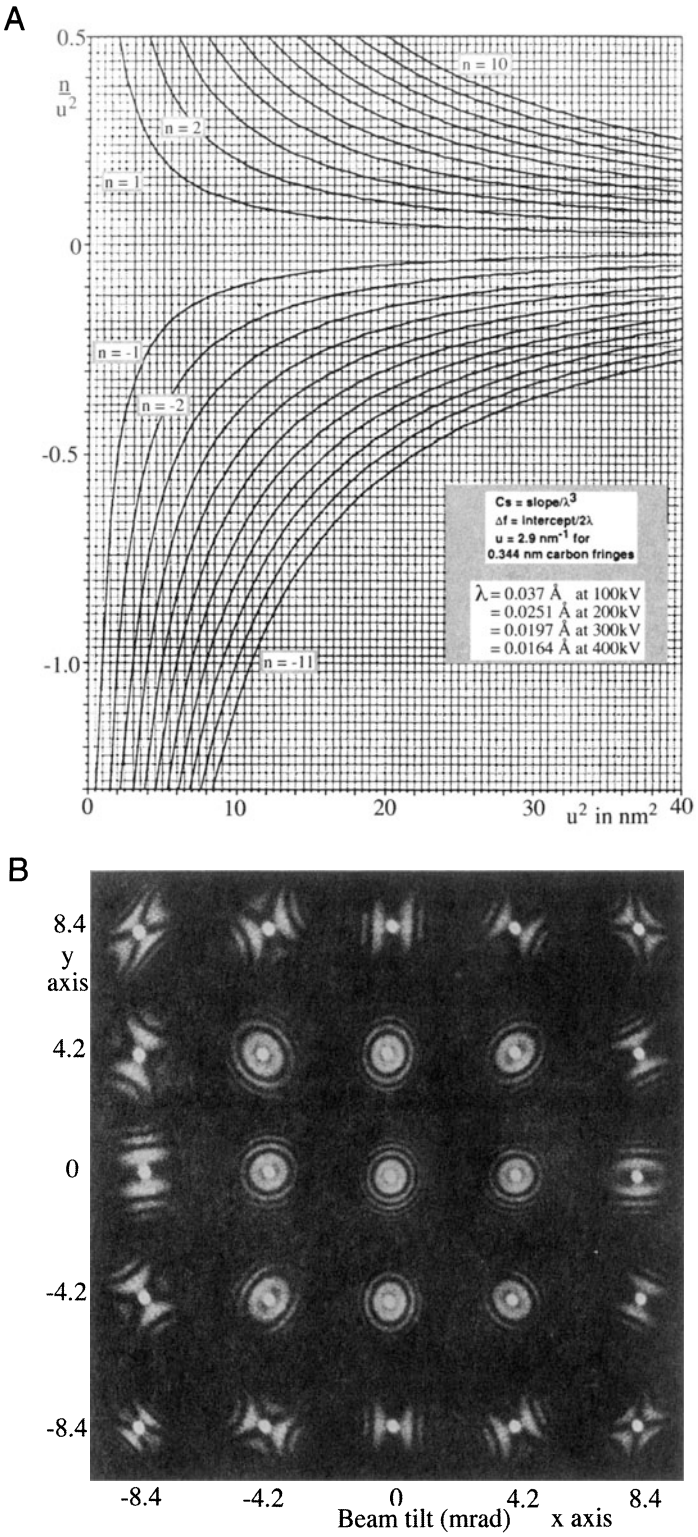


Figure 30.6. (A) Plot of nu^{-2} versus u^2. The rings in any diffractogram correspond to a series of n values which allow you to draw straight lines on this figure and thus determine the slope and the ordinate intercept, giving C_s and Δf respectively. (B) Set of diffractograms showing the effect of incident beam tilt.

30.5.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 30.7. 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.* (1992) 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.

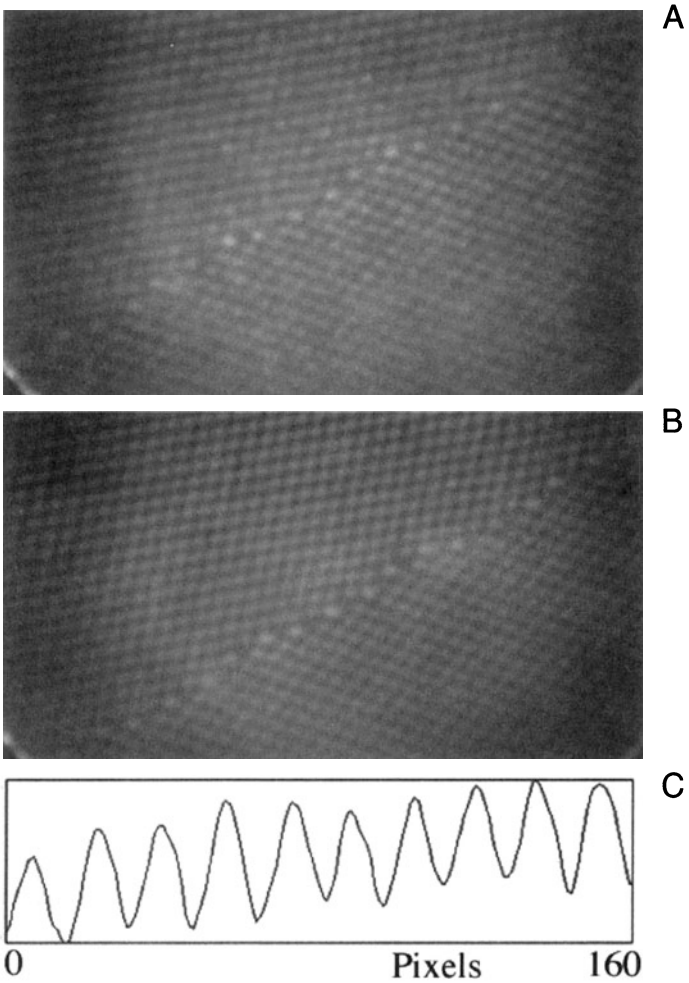


Figure 30.7. An example of the benefit of frame averaging to improve information from a video recording: (A) one frame, (B) 16 frames, (C) intensity profile along a (111) plane in (B).

If you use a TV-rate video, you'll almost certainly use background subtraction as a matter of routine. 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 (pseudo-color) to your TEM images. Although it is often assumed that this is just done to make the images even more appealing to the nonmicroscopist (or nonscientist), 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. In the TEM, all of our apertures have relatively sharp edges. In the computer, you have the possibility of using multiple apertures, apertures with different shapes, and apertures with diffuse edges. Apertures with diffuse edges will help eliminate the streaking which will otherwise be present. 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 and 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 (Russ 1990, 1995).

30.5.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, eg., 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.

30.6. 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.

30.6.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, e.g., Erickson and Klug 1971). 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 a pre-

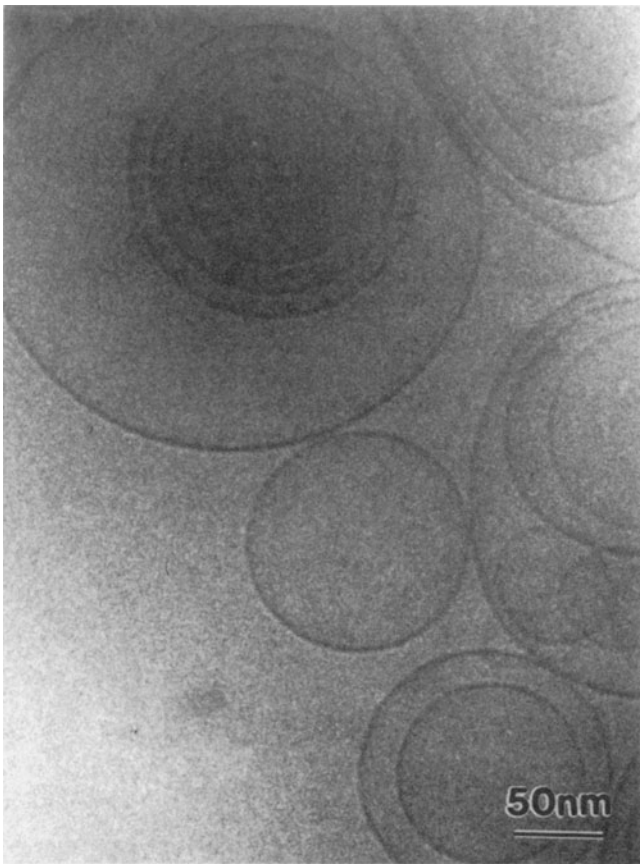


Figure 30.8. The image is from a highly beam-sensitive solution of surfactants in water. The solution has been frozen by plunging a film into liquid ethane and then transferring it to the TEM. The large circles are the surfactants that have aggregated to form vesicles; the concentration of the surfactants in the solution is just right for them to form lamellar structures. Texture starts to appear in the image as soon as the beam interacts with the specimen.

determined 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. The image shown in Figure 30.8 illustrates the possibilities. If you read the review by van Heel *et al.* (1992), you will get some idea of how far you can already go in this field.

30.6.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 (Downing 1992, Dorset 1995) and even correcting for distortions in the specimen (Saxton 1992).

30.6.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 will be particularly valuable for frame averaging using a video camera. Then there will also be many applications for diffraction-contrast imaging, as well as for microanalysis.

30.6.D. Reconstructing the Phase

Although we are studying phase contrast, the image intensity doesn't directly give us phase information. Kirkland *et al.* (1982) have shown that the phase can be reconstructed by processing a defocus series. In their approach they use an iterative nonlinear image-processing technique to reconstruct the complex 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 Figures 30.9a–e, together with the reconstructed transmission function plotted both as a real and imaginary part and then as an amplitude and phase part Figures 30.9f–i. The projected structure of the known unit cell is also shown Figure 30.9j. 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 few examples of full phase reconstructions published. You must record such a series of defocus images if you want to do quantitative HRTEM.

30.6.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 SAD pattern to carry out electron crystallography in the same way as in classical X-ray crystallography. As you can appreciate from Figure 30.10, particularly if the unit cell is large and the specimen examined is thin, there is a great deal of information in the SAD pattern but

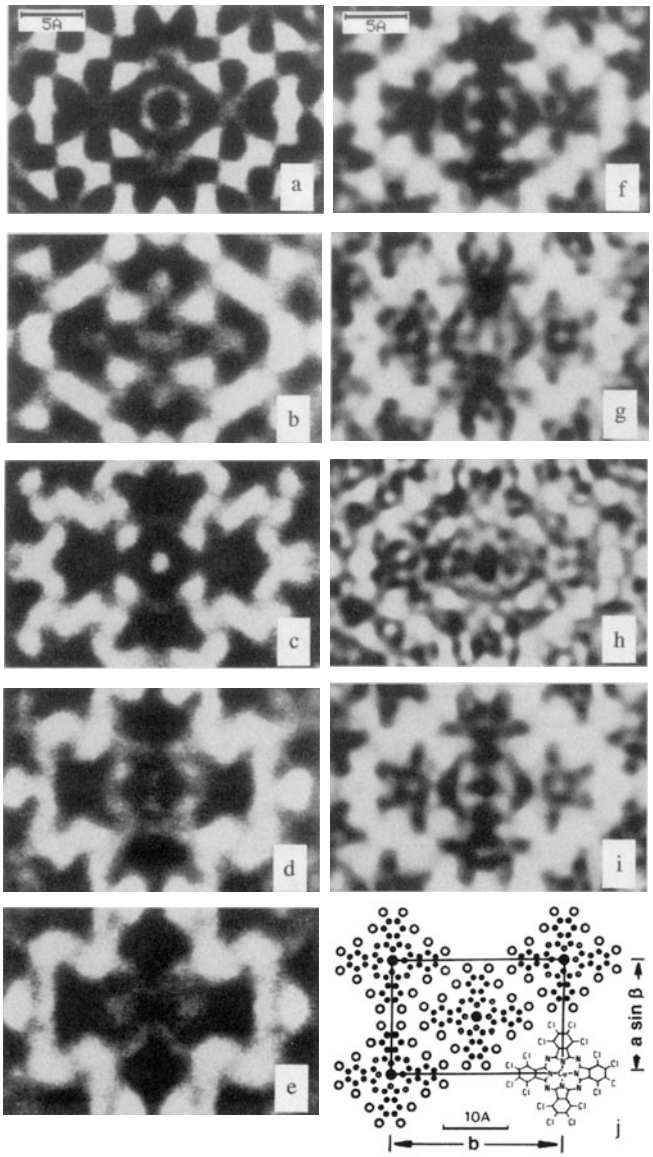


Figure 30.9. (a–e) Images from an experimental defocus series of $CuCl_{16}PC$; the reconstructed transmission function plotted as both a real and imaginary part (f,g) and then as an amplitude and phase part (h,i); (j) the projected structure of the crystal.

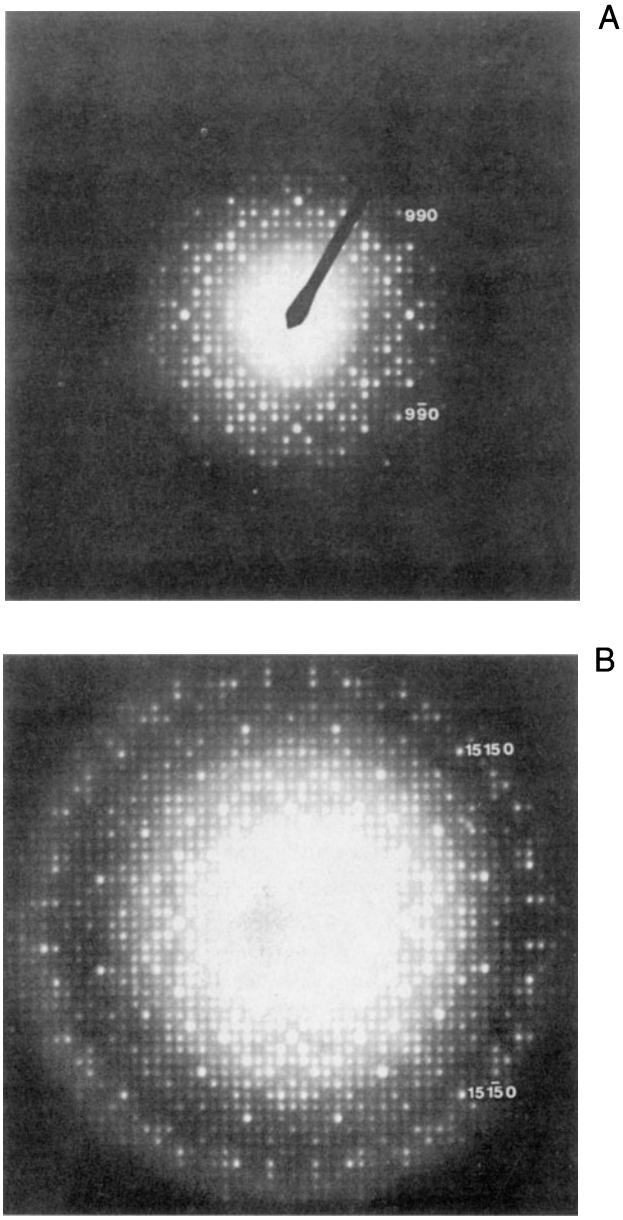


Figure 30.10. SAD patterns from $K_2O.7Nb_2O_5$ recorded using two different exposures. More than one exposure is needed to get all the information in the pattern. The space group in P4bm with $a = b = 2.75$ nm. The (15, 15, 0) reflections correspond to spacing of 0.13 nm.

you can't get it all in one exposure. Hovmöller's group (see, e.g., Zou 1995 and the ELD program in Section 1.5) have 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.

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 SAD patterns. 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).

> The structure deduced from the HRTEM approach should generate the experimental SAD pattern, 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.* 1995). 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.

30.6.F. Tilted-Beam Series

Having gone to great trouble to remove any beam tilt, we will now 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 27.2. Now you use a computer to combine information in different tilted-beam images. The method proposed by Kirkland *et al.* (1995) 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 30.11A; you also need the on-axis image, as shown in the tableau in Figure 30.11B. Since it is important that

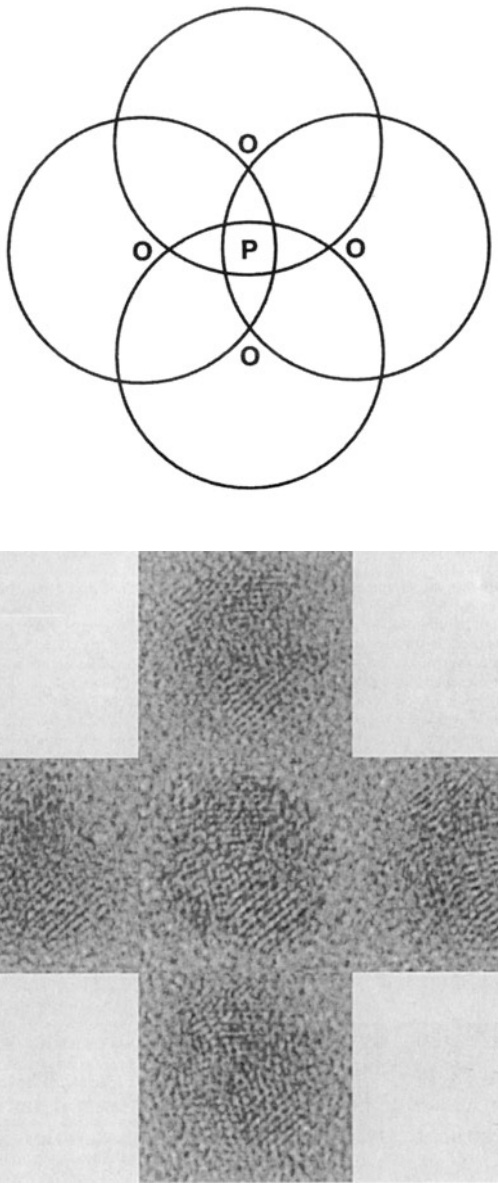


Figure 30.11. A method for extending the resolution of your TEM. Set the beam tilt to zero, then tilt the beam through different angles. (A) The four regions of Fourier space are shown by the four circles; O is each position of the tilted beam, P is the optic axis, and PO corresponds to the angle of 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.

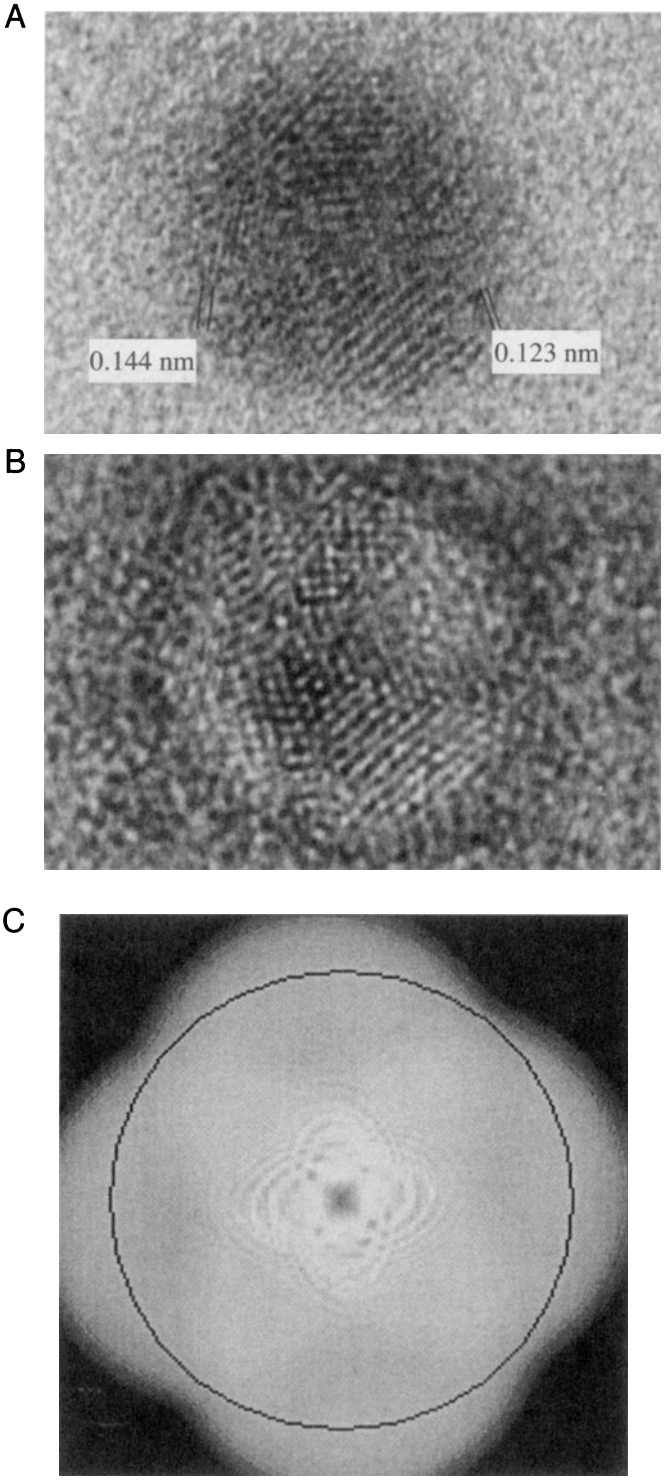


Figure 30.12. The restored image of a gold particle on amorphous Ge: (A) the amplitude (modulus) image showing 0.123-nm fringes, (B) the corrsponding phase image, (C) the transfer function after restoration, plotted in two dimensions. The circle corresponds to 0.125-nm detail successfully transferred to the image. For such thin specimens, atom-dense positions have a reduced modulus [black in (A)] and an increased phase [white in (B)].

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 30.12 demonstrates the potential of the technique: detail is present in the image at a resolution of 0.123 nm using a 400-kV microscope.

30.7. 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 big advantage in remote control will not be just that you can sit in Bethlehem and operate a microscope in California but that you can locate the microscope in its own 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. Also, if your specimen is not ideal or the microscope breaks down, you won't have to go for a walk on the beach in California but can continue word processing in Pennsylvania.

The procedure has been extensively discussed and implemented by Saxton and Koch (1982), Krivanek and Mooney (1993), Koster *et al.* (1988), Koster and de Ruijter (1992), and others. Your role is to select a suitable region of the specimen close to the area of interest; the area of interest should ideally only be examined at low magnification. At present, you will make the initial alignment manually and then turn the process over to the computer. The computer will then adjust the astigmatism and correct the beam tilt independently and quickly.

Figure 30.13 shows how well and quickly this procedure can now be done. The different diffractograms in each tableau correspond to incremental changes in the

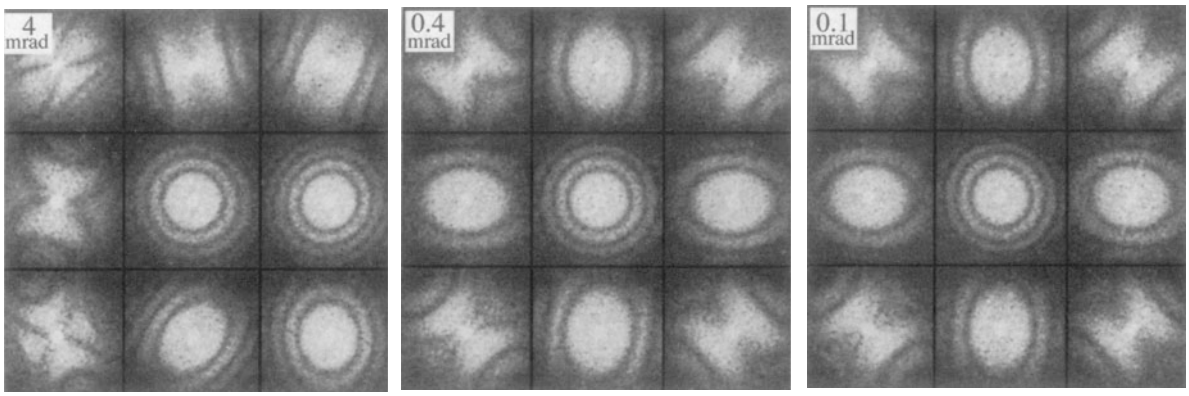


Figure 30.13. Using the computer to correct the beam tilt: The different diffractograms in each tableau correspond to incremental changes in the beam tilt of 6 mrad in the x and y directions away from the initial beam tilt in the central diffractogram. The computer initially determined a misalignment of 4 mrad, then corrected this to 0.4 mrad and finally to 0.1 mrad. Note the central diffractogram is almost unchanged, emphasizing the need for computer-controlled tilting to give correct alignment.

beam tilt of 6 mrad in the x and y directions. The computer showed that the initial tilt error was 4 mrad, which was reduced to 0.4 mrad after one pass and < 0.1 mrad after the second pass. Each pass only took 28 s! The astigmatism shown in Figure 30.14 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 an 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 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.* (1983) 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.

30.8. QUANTITATIVE METHODS OF IMAGE ANALYSIS

In the next six sections we will go through several particular illustrations of image processing in HRTEM. 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 recom-

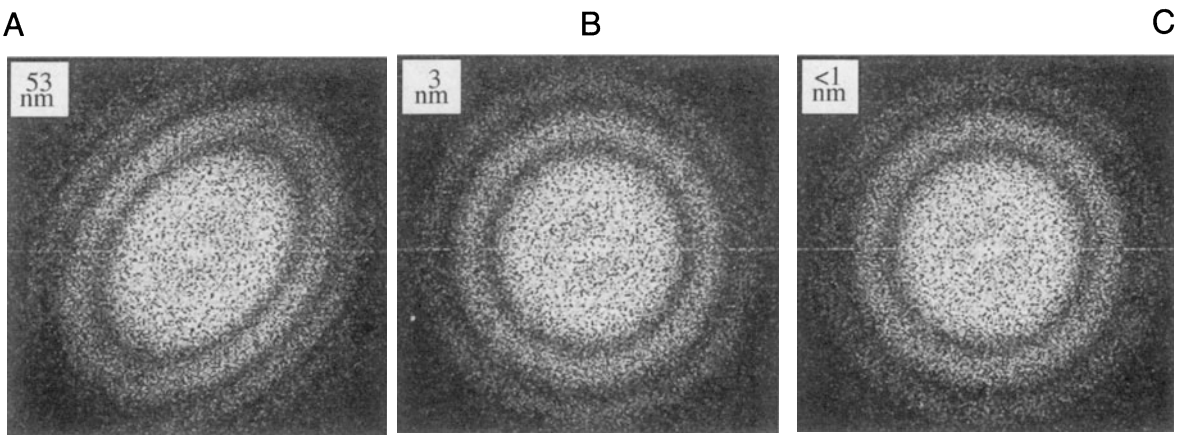


Figure 30.14. Diffractograms showing the astigmatism corrections made by the computer following a similar procedure to that shown in Figure 30.13. The final diffractogram shows that the HRTEM is now very well stigmated.

mended. At this time we can summarize the situation as follows:

- Quantitative analysis is difficult, often tedious, and invariably time-consuming.
- You have to understand the basic ideas of image theory.
- Your analysis is only as good as your image and your image is only as good as your specimen.

We gave the necessary information on how to obtain the software in Section 1.5.

30.9. PATTERN RECOGNITION IN HRTEM

The most obvious feature of the majority of 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.

> The principle of pattern recognition is to take or make a template, move it across your image, and measure how closely the image resembles your template.

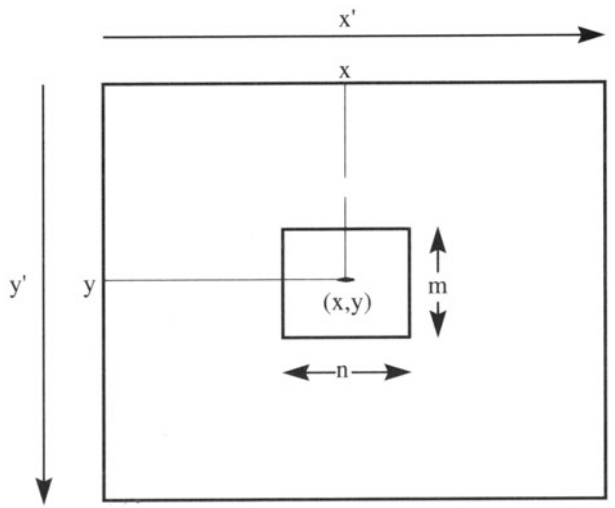


Figure 30.15. The large rectangle represents the digitized image, size $(x' \times y')$; the small rectangle, $(m \times n)$ pixels, represents the template used in the cross-correlation calculation. The small rectangle is moved to different (x,y) positions during the process.

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 as shown in Figure 30.15 following Paciornik *et al.* (1996). The large rectangle represents your digitized image and could be 1k × 1k; 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 already got the

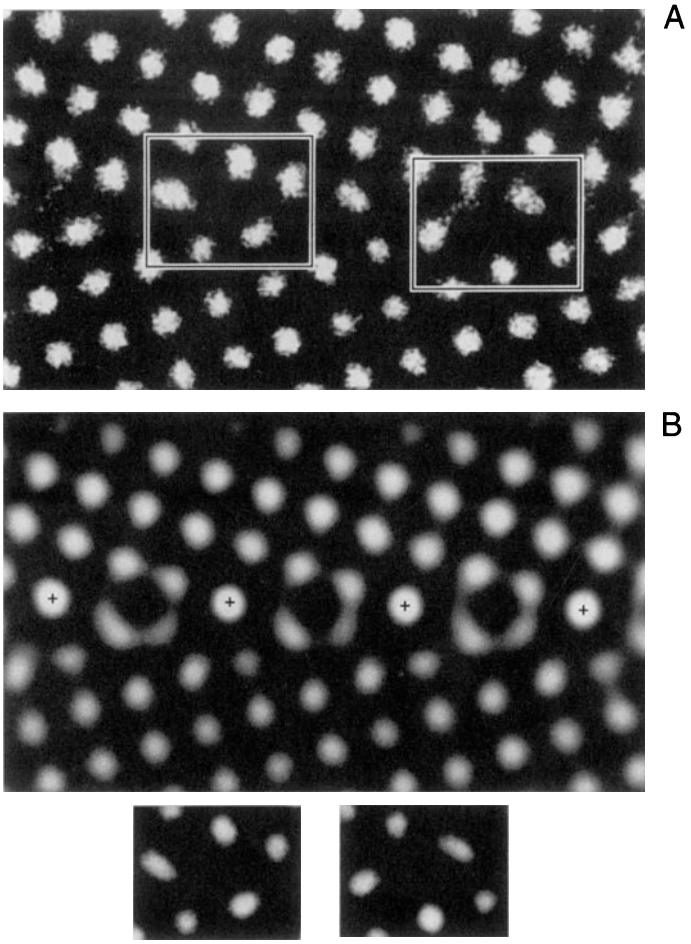


Figure 30.16. Analysis of a small region of a $\Sigma = 5$ tilt boundary in TiO_2. The two small boxed regions in (A) are only present at the boundary; these are used as the templates for the cross-correlation method. (B) The cross-correlated image. The small rectangles at the bottom of the figure are the low-noise averaged images of the GB templates.

right magnification and rotation. If not, you have to set these first and we will return to this problem shortly.

This is a real-space approach.

The process is best understood by an example. Figure 30.16A 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 30.16B. Having found all the regions which match the template, we could then take the average of these to produce low-noise images of the grain boundary templates. The final step is the comparison of these templates with models of the grain boundary structure. There are two important points to remember:

- When you average images, you implicitly assume that they are all the same.
- Don't forget our discussion in Chapter 27 of interface grooving and the problems associated with interfacial segregation.

30.10. 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 you must derive them from image simulations. This approach has been described for two special cases by Kisielowski *et al.* (1995) and Ourmazd *et al.* (1990).

30.10.A. The Example of a Specimen with Uniform Composition

In the QUANTITEM approach developed by Ourmazd, the results of Chapter 29 are summarized by a general equation linking the intensity and all the imaging (S_i) and materials (P) parameters,

$$I(x, y) = F(P(x, y), S_i) \qquad [30.5]$$

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

$$I(x, y) = F\left(P(x, y), S_i^0\right) = F^0(P(x, y)) \qquad [30.6]$$

The basis of this approach is quite straightforward:

- 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. Ourmazd gives a helpful simple analogy for this process, as illustrated in Figures 30.17A and B. The function F^0 describes the path of a swinging pendulum as it varies with time. 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. 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 Mbyte 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 digitize these to give many templates, which are n pixels by m pixels as shown in Figures 30.17C–E. 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 the path of these N-dimensional vectors as the projected potential changes.

The next step is to define a reference frame for these vectors. Three basis vectors are derived from the experimental image. Ourmazd *et al.* (1990) argue that three basis vectors will be sufficient, as we can show in the 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.

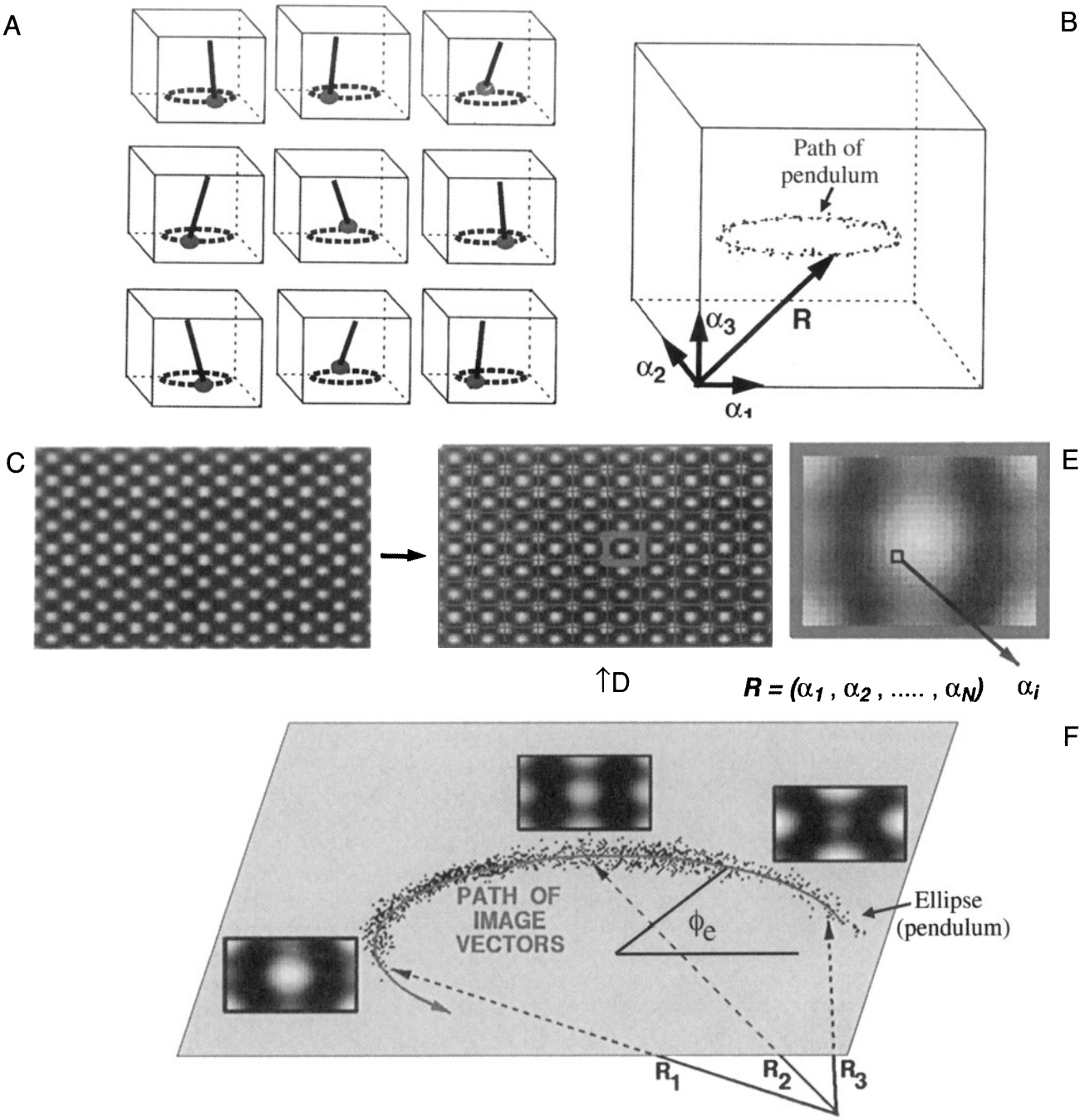


Figure 30.17. The principle of vector parameterization used in QUANTITEM. (A) shows the different pendulum positions and (B) shows the path of the pendulum. Each HRTEM image is represented by a single vector R which has *N* dimensions; (C–E) The image is separated into unit cells and digitized to give ($n \times m$) pixel templates; (F) Three vector-parameterized image ($\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{R}_3$) of a wedge-shaped specimen of Si at different thicknesses.

- 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

$$\mathbf{R}^G = a_G\mathbf{R}^B + b_G\mathbf{R}^S + c_G\mathbf{R}^D \quad [30.7]$$

Each of the basis vectors (images) can be expressed in the same manner

$$\mathbf{R}_i^T = a_i\mathbf{R}^B + b_i\mathbf{R}^S + c_i\mathbf{R}^D \quad [30.8]$$

giving three vectors for $i = 1$, 2, and 3.

We can turn these equations around to define any vector $\mathbf{R}^G$ in terms of the basis vectors

$$\mathbf{R}^G = \alpha_G \mathbf{R}_1^T + \beta_G \mathbf{R}_2^T + \gamma_G \mathbf{R}_3^T \qquad [30.9]$$

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 in Figure 30.17F.

30.10.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 a series of simulated images as shown in Figure 30.17F. Each point on the curve corresponds to a unit-cell image, 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 29, 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 in Figure 30.17F. Thinking back to the pendulum analogy, the path parameters are the image version of the coordinates for the harmonic oscillator.

Now the ϕ_e curve parameters can be obtained from a series of images. We can change 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 Chapter 29. In materials such as YBCO, this is not the case, because too many Bloch waves are important and the curve is not a straight line.

30.10.C. Noise Analysis

If noise moves the vector off the ellipse, we can analyze the noise. If it 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 method should reduce the noise by a factor of $\sqrt{N}$, which for a 10 pixel × 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

$$I = B + S + D \qquad [30.10]$$

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 Figures 30.18A and B. In the first example, 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, as in Figure 30.18B.

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 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:

- 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 dif-

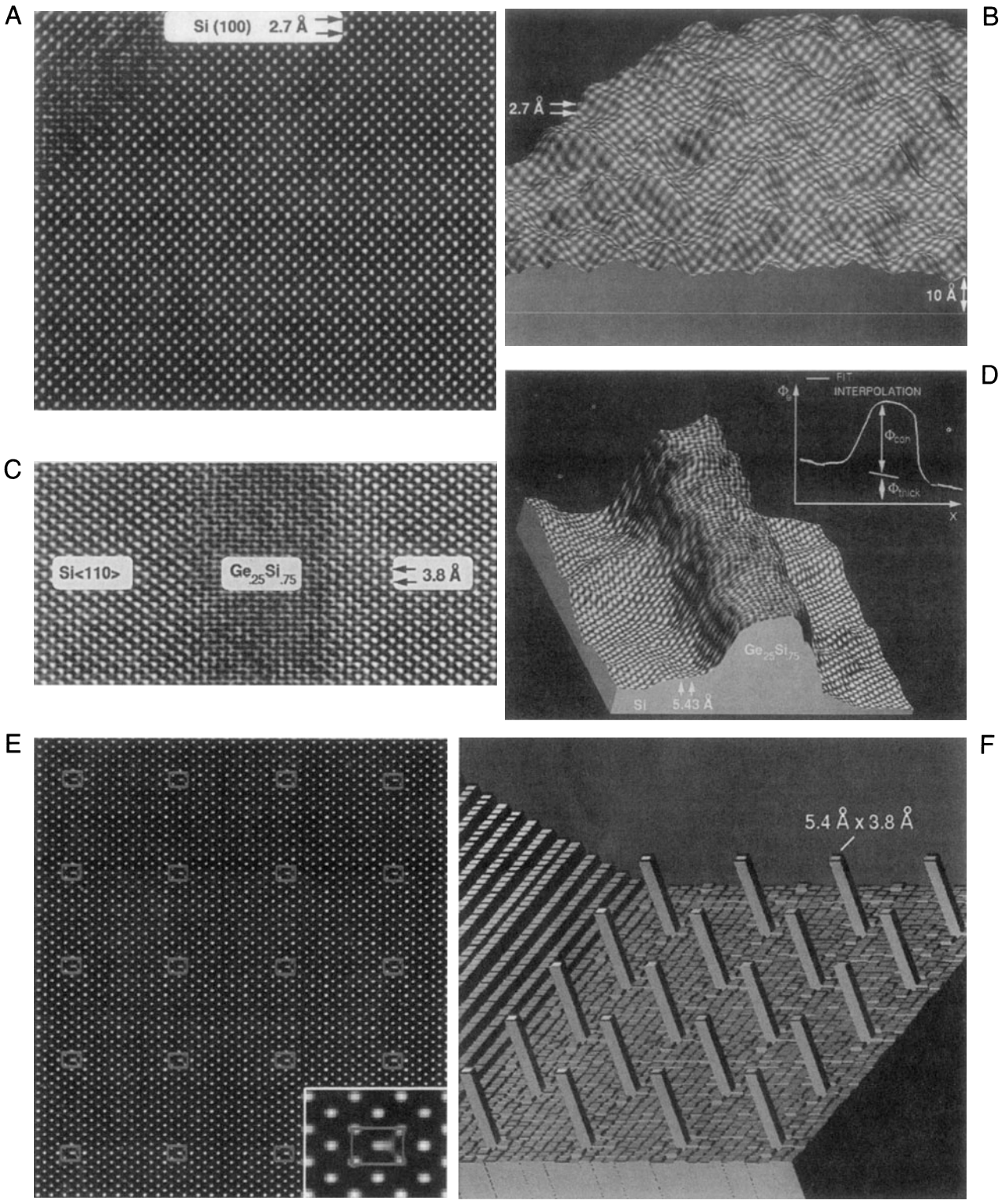


Figure 30.18. Examples of the application of QUANTITEM: on the left is the conventional HRTEM image, on the right is the QUANTITEM image. (A,B) Mapping of the roughness of the Si surface covered by SiO_2; (C,D) a layer of Ge_xSi_{1-x} in a matrix of Si, the inset shows the plot of ϕ_e versus x; (E,F) analyzing a simulated image of columns of Ge (a δ-function in concentration) in Si.

ferent compositions, you have determined the local composition.

The example shown in Figures 30.18C 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.

You can test the potential resolution of the technique, its sensitivity to the alignment of the beam, bending of the specimen, and beam divergence in the usual way by creating model structures, simulating the images, and then analyzing them. Figures 30.18E and F show that the potential resolution is superb, but beam tilt can cause 10% errors in thickness measurement. The conclusion is clear: as always, you will only get the best results if the specimen is ideal and both the microscope and the specimen were 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).

30.11. QUANTITATIVE CHEMICAL LATTICE IMAGING

This technique uses the approach described in Section 30.10, but can only be applied to materials where we have chemically sensitive reflections, which we discussed in Chapter 17. We will use these reflections in Chapter 31 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 Figures 30.19A and B 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. Figure 30.19C shows the sort of image we can analyze with this approach. 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 characterized by the two vectors, $\mathbf{R}_{GaAs}$ and $\mathbf{R}_{AlGaAs}$, following the approach described in Section 30.10. This was done in this case by simulating the cells, dividing them into 30 × 30 pixel arrays (so $N = 900$), and then plotting **R**. 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 by Figure 30.19D. The three known simulated templates each produce 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 flow chart shown in Figure 30.20 helpful.

- In Figure 30.20, the experimental image (A) is 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 (B).
- The pair of templates shown (C) 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 30.19D also).

The maximum chemical difference determines how far apart the two principal distributions can be (E). Since the image is now fully parameterized, we can do the statistics and finally invert the angular data to give the compositions (K).

This technique has enormous potential, but you must also remember that it is susceptible to all the drawbacks inherent in HRTEM. The big 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 back in Figure 30.18E. 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.

30.12. METHODS OF MEASURING FIT

There are two methods presently used to obtain a measurement of the goodness of fit, namely:

- Cross-correlation.
- Least-squares refinement (Section 30.13).

In this section, we'll use the cross-correlation method to compare an $n \times m$ pixel template (see Section 30.9) with every possible $n \times m$ rectangle in the image. The computer

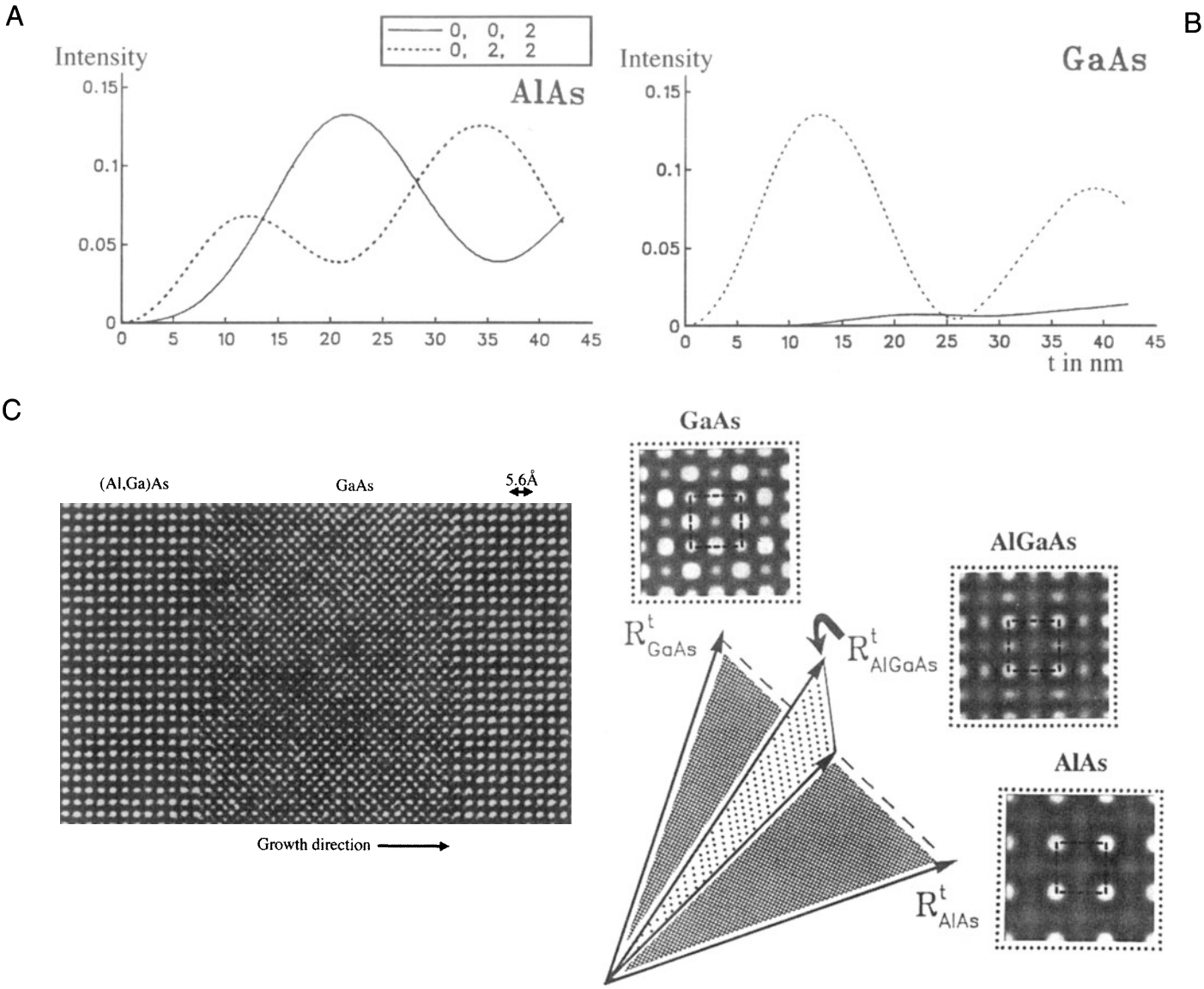


Figure 30.19. (A,B) Variation in intensity of the (002) and (022) beams along [100] in AlAs and GaAs (400 keV); (C) chemical lattice image of a layer of GaAs between two layers of AlxGa$_{1-x}$As (x = 0.4); (D) templates simulated for different values of x each produce a vector $\mathbf{R}^t$.

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

$$CCF(x, y) = \frac{\sum_{x'}\sum_{y'}\left[i(x', y') - \langle i(x', y')\rangle\right]\cdot\left[t(x' - x, y' - y) - \langle t\rangle\right]}{\sqrt{\left\{\sum_{x'}\sum_{y'}\left[i(x', y') - \langle i(x, y)\rangle\right]^2 \sum_{x'}\sum_{y'}\left[t(x' - x, y' - y) - \langle t\rangle\right]^2\right\}}} \quad [30.11]$$

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.
- $\langle t \rangle$ is the average value of the pixels in $t(x', y')$; it is computed just once.
- $\langle i(x, y) \rangle$ 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 equa-

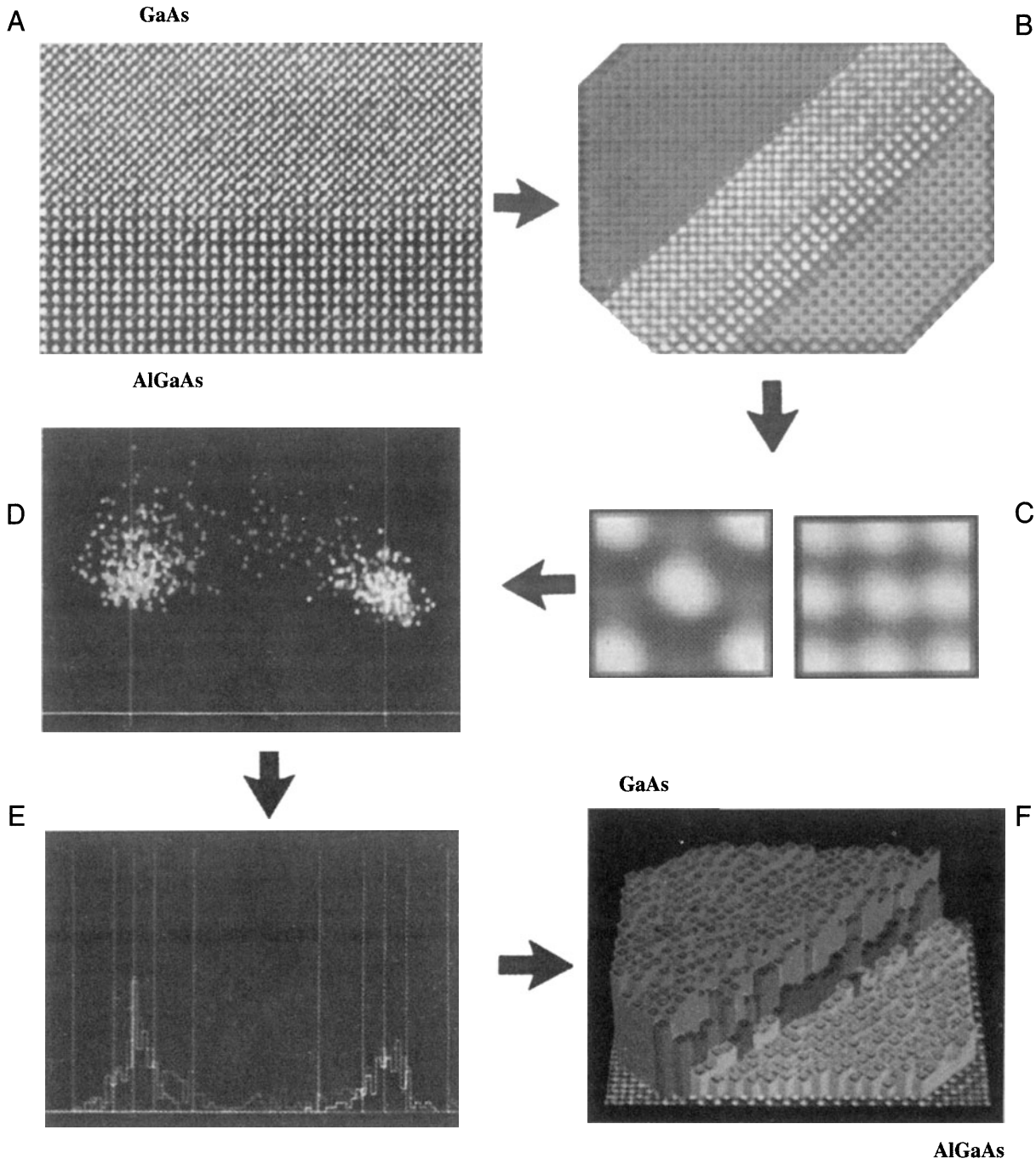


Figure 30.20. A flow chart summarizing the pattern recognition procedure.

tion 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 30.11 as the dot product of two vectors **t** and **i** which gives us the $n \times m$ component of the template

$$CCF(x, y) = \cos(\theta) = \frac{\mathbf{t} \cdot \mathbf{i}}{|\mathbf{t}||\mathbf{i}|} \qquad [30.12]$$

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

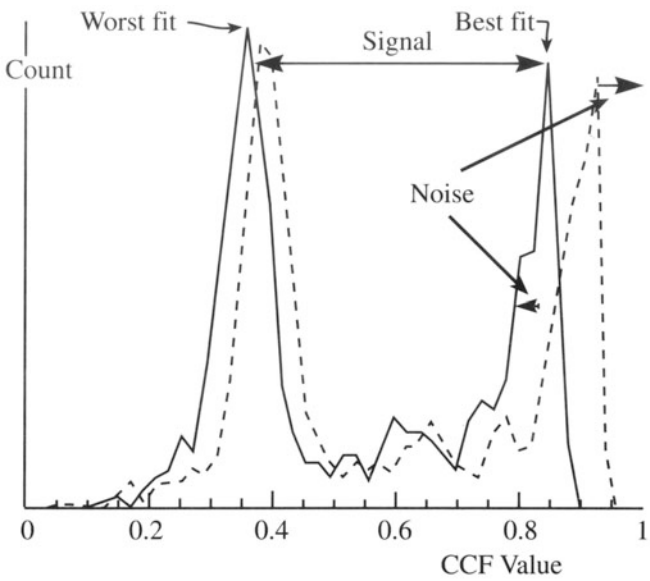


Figure 30.21. A plot of how often a particular CCF value occurs. The two peaks in the continuous curve are the best fit and the worst fit: their separation gives a measure of the discrimination signal; the width of the peaks gives a measure of the noise and hence a signal-to-noise ratio. The plot can be redrawn after repeating the process (dashed line) to estimate the improvement in signal-to-noise ratio.

out the number of times each particular CCF value occurs, as illustrated in Figure 30.21. 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 1.5). Your 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 (1980).

30.13. 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 of 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 your lattice image such that you correct for variations in intensity across the field of view and the nonlinearity of the response from the photographic film. Figure 30.22 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.

> 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.

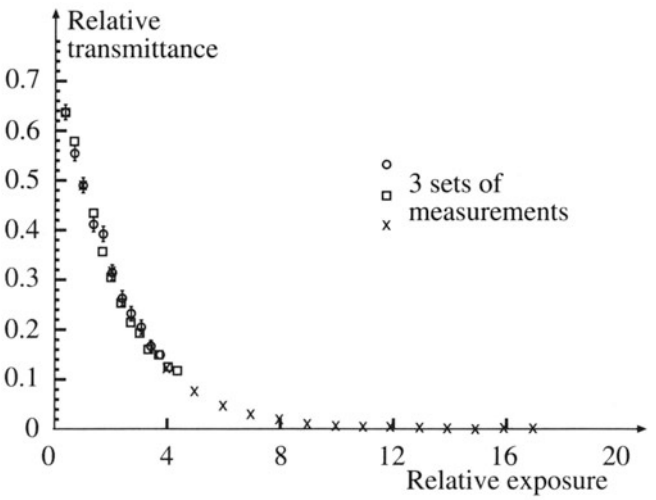


Figure 30.22. Plot of transmittance versus relative exposure measured using a CCD camera to digitize images from Kodak SO-163 film (the symbols indicate three different microscopes).

When you analyze the image, you'll find out if the area you photographed was correctly aligned. Since your image only takes two seconds or so to record, you 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

$$f_i(x) = \frac{\left[f_i^{obs} - f_i^{calc}(x)\right]}{W_i} \qquad [30.13]$$

and your task is to minimize $f_i(x)$. The difference between the intensity in the experimentally observed ith pixel and its calculated value would be zero if everything had been scaled correctly, the imaging conditions (Δf, C_s, etc.) were correct, and you have the right structure.

Let's say W_i is the image which represents the error bars for pixel i. Then we can write that

$$W_i = \min\left[\sum_{i=1}^{N} f_i(x)^2\right] \qquad [30.14]$$

This equation defines the nonlinear 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 (Moré 1977, Moré *et al.* 1980)].

In their demonstrations of this approach to analyze a [001] tilt grain boundary in Nb, King and Campbell (1993, 1994) 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 1.5), the optimization took 20 iterations and 80 multislice calculations.
- Next, they had to optimize the structure of the grain boundary. 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. Secondly, they tell you that this is a computer-intensive process; that you 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 grain boundary. 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 intensity value of the ith pixel in the experimental image and f_i^{calc} as the corresponding value in the simulated image. We then calculate the residual $f_i(x)$ as follows

$$f_i(x) = \frac{\left(f_i^{obs}(x) - \left(f_i^{calc}(x) + b^{fit}\right)\right)}{W_i} \qquad [30.15]$$

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

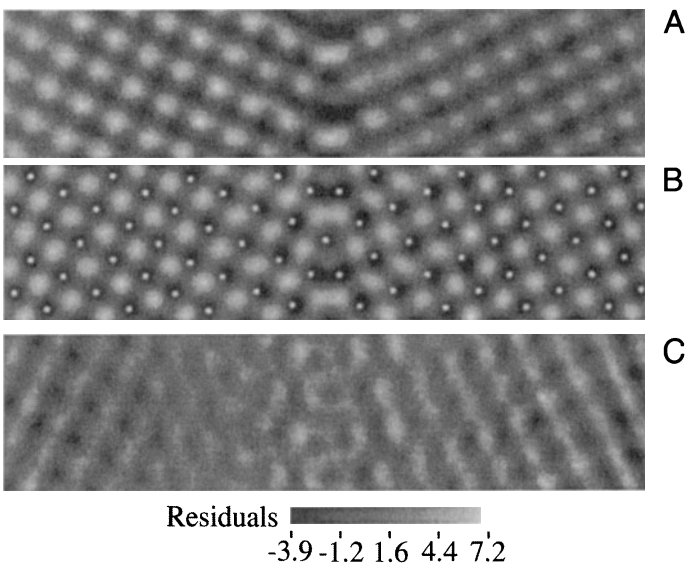


Figure 30.23. (A) Experimental image, (B) best-fit simulation, and (C) normalized residuals of a Σ = 5 symmetric tilt boundary in Nb.

$$W_i = \sigma_i^{obs} + 0.05 f_i^{obs} \quad [30.16]$$

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 30.23 for images from a $\Sigma = 5$, (310), [001] GB in Nb. The values for the thickness and Δf show how consistent this technique can be, especially since the images in Figures 30.23A and B were from *opposite sides of the grain boundary*; (C) was like (B), but for a different defocus value.

30.14. A FOURIER TECHNIQUE FOR QUANTITATIVE ANALYSIS

Möbus *et al.* (1993) have 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 most promising for analyzing regions which contain defects.

> An adaptive filter is one where the shape of the filter, or mask, is adapted to fit the shape of the "image" it's filtering.

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 $\Sigma = 5$ grain boundary 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 30.24A. The power spectrum (computer-generated DP) of the micrograph is shown in Figure 30.24B. The adaptive filter and the filtered image are shown in Figure 30.24C and D. The important feature of the adaptive filter is that it was created as such because the computer detected the doubling of the periodicity, which is *only* present in the grain boundary. Secondly, the mask consists of elongated openings which we know we need when analyzing the grain boundary because of the shape effect.

30.15. REAL OR RECIPROCAL SPACE?

In principle we could equally well compare two images in reciprocal space rather than real space. However, while 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.

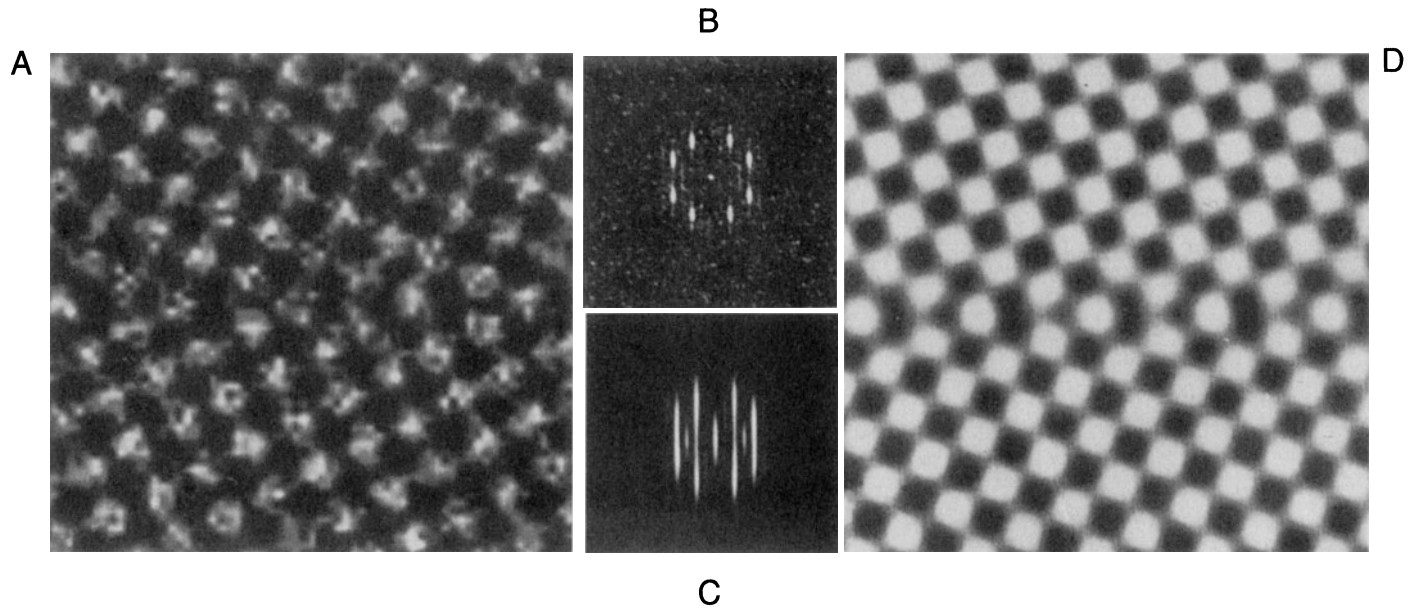


Figure 30.24. (A) White noise added to a calculated image of a $\Sigma = 5$ grain boundary; (B) The power spectrum of (A); (C) the adaptive filter; (D) the filtered image.

- 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.

30.16. THE OPTICAL BENCH

Although not widely used now, the optical bench is still a useful instructional tool. A typical experimental set-up is shown in Figure 30.25. 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 our objective apertures. 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 30.26. So Figure 30.26D is effectively a BF image: you lose a lot of information in such images!

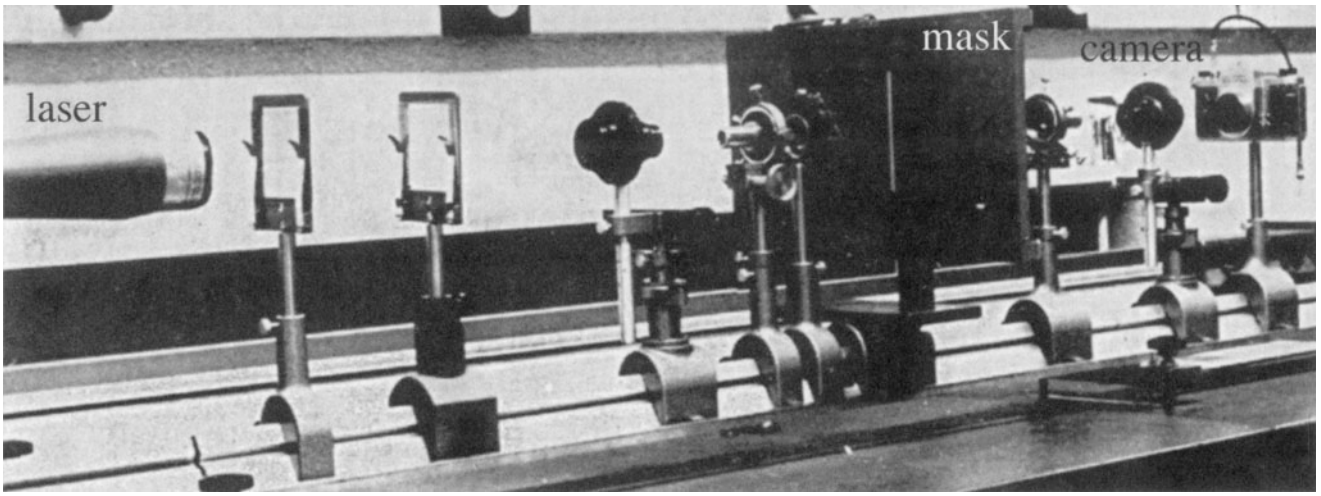


Figure 30.25. A typical experimental set-up for an optical bench, with the mask in the back focal plane.

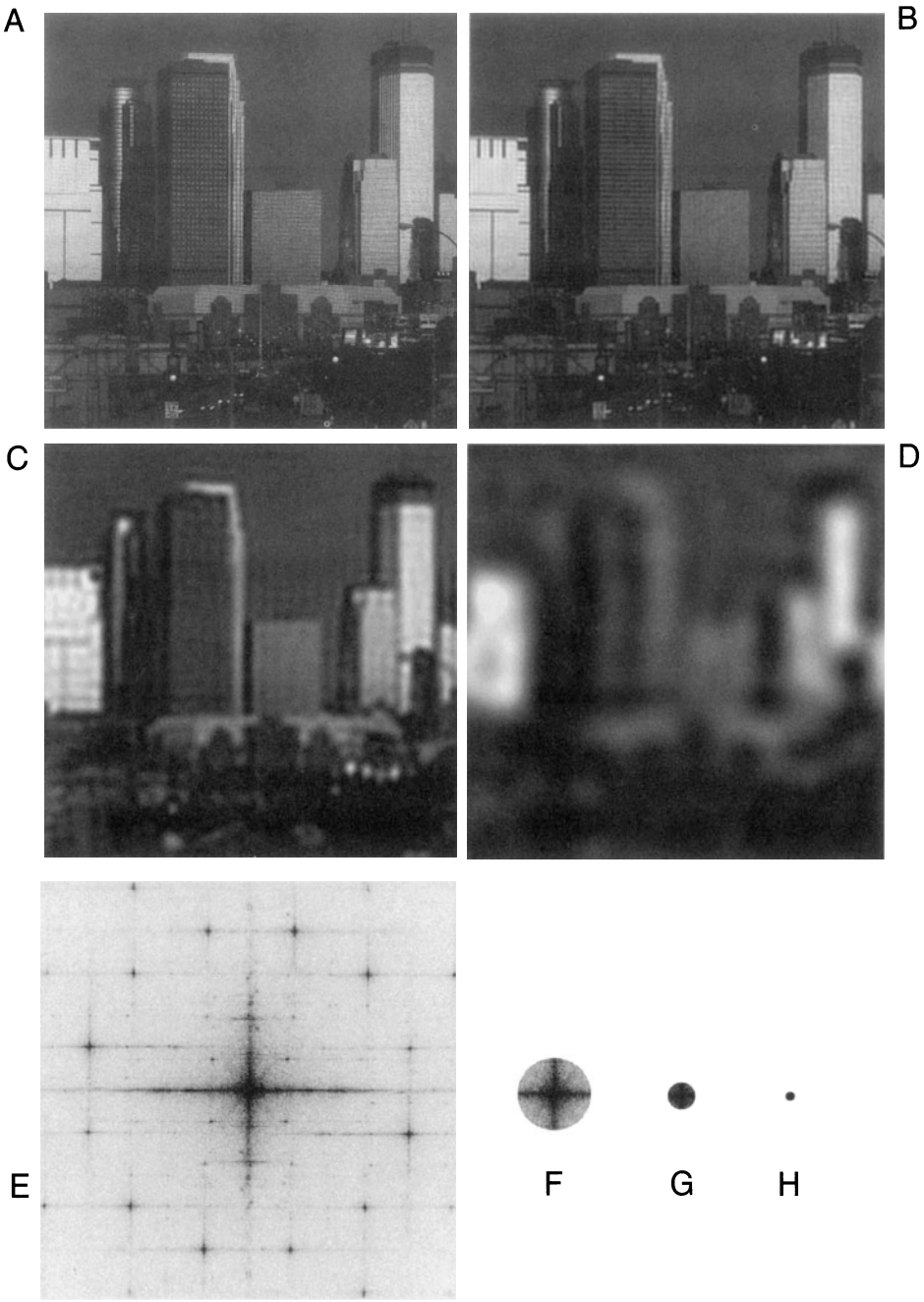


Figure 30.26. The effect of mask (aperture) size on a nonperiodic image of the Minneapolis skyline viewed from near the Guthrie Theater. (A–D) Reducing the aperture size, as indicated in the corresponding optical transforms (diffractograms) (E–H) reduces the image detail. The streaks in (E) arise from 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 with a special enlarger. However, we have done very little quantitative imaging in materials science. 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 your experimental parameters.
- If you are going to use reciprocal-space techniques for quantitative analysis, you should let the computer design the optimum mask as part of this process; usually, it will not be a circular mask, especially if you are studying interfaces!
- The potential for image restoration is not limited by the signal mixing due to C_s and Δf. You can unscramble those effects. Ultimately, the limit is set by the signal-to-noise ratio in your image.
- 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.

There is always the possibility of removing information which is important. For example, Fresnel fringes often should be there! Beware of making reality match your simulation, rather than the reverse. In the same vein, we draw your attention to the conclusion of Hÿtch and Stobbs (1994), 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.

In this chapter we have discussed several different techniques used for processing TEM images. Several software packages are widely used by the TEM community and have been listed in Section 1.5. 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.

REFERENCES

General References

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, P.W., Ed. (1980) *Computer Processing of Electron Microscope Images,* Springer-Verlag, New York. This collection of articles is for advanced students.

Hawkes, P.W., Ed. (1992) *Signal and Image Processing in Microscopy and Microanalysis,* Scanning Microscopy Supplement **6**, SEM, Inc. AMF, O'Hare, Illinois. This volume and its 1988 and 1996 companions are required reading for this subject.

Hawkes, P.W., Ottensmeyer, F.P., Saxton, W.O., and Rosenfeld, A. Eds. (1988) *Image and Signal Processing for Electron Microscopy,* Scanning Microscopy Supplement **2**, SEM, Inc. AMF, O'Hare, Illinois.

Hawkes, P.W., Saxton, W.O., and Frank, J., Eds. (1996) *Image Processing,* Scanning Microscopy Supplement, SEM, Inc. AMF, O'Hare, Illinois.

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, J.C. (1995) *The Image Processing Handbook,* 2nd edition, CRC Press, Boca Raton, Florida. A beautiful, comprehensive, and an essential component of any EM lab (or home).

Specific References

Dorset, D.L. (1995) *Structural Electron Crystallography,* Plenum Press, New York.

Downing, K.H. (1992) Scanning Microscopy Supplement **6**, ibid., p. 405.

Erickson, H.P. and Klug, A. (1971) *Phil. Trans. Roy. Soc. Lond.* **B261**, 105.

Frank, J. (1980) in *Computer Processing of Electron Microscope Images,* p. 187.

Hÿtch, M.J. and Stobbs, W.M. (1994) *Ultramicroscopy* **53**, 191.

King, W.E. and Campbell, G.H. (1993) *Ultramicroscopy* **51**, 128.

King, W.E. and Campbell, G.H. (1994) *Ultramicroscopy* **56**, 46.

Kirkland, E.J., Siegel, B.M., Uyeda, N., and Fujiyoshi, Y. (1982) *Ultramicroscopy* **9**, 65.

Kirkland, A.I., Saxton, W.O., Chau, K.-L., Tsuno, K., and Kawasaki, M. (1995) *Ultramicroscopy* **57**, 355.

Kisielowski, C., Schwander, P., Baumann, F.H., Seibt, M., Kim, Y., and Ourmazd, A. (1995) *Ultramicroscopy* **58**, 131.

Koster, A.J., van den Bos, A., and van der Mast, K.D. (1988) Scanning Microscopy Supplement **2**, p. 83.

Koster, A.J. and de Ruijter, W.J. (1992) *Ultramicroscopy* **40**, 89.

Krivanek, O.L. (1976) *Optik* **45**, 97.

Krivanek, O.L. and Mooney, P.E. (1993) *Ultramicroscopy* **49**, 95.

Möbus, G., Necker, G., and Rühle, M. (1993) *Ultramicroscopy* **49**, 46.

Moré, J.J. (1977) in *Lectures Notes in Mathematics* (Ed. G.A. Watson), p. 630, Springer, Berlin.

Moré, J.J., Garbow, B.S., and Hillstrom, K.E. (1980) User Guide for MINIPACK-1.

Ourmazd, A., Baumann, F.H., Bode, M., and Kim, Y. (1990) *Ultramicroscopy* **34**, 237.

Paciornik, S., Kilaas, R., Turner, J., and Dahmen, U. (1996) *Ultramicroscopy* **62**, 15.

Saxton, W.O. (1992) Scanning Microscopy Supplement **6**, p. 405.

Saxton, W.O. and Koch, T.L. (1982) *J. Microsc.* **127**, 69.

Saxton, W.O., Smith, D.J., and Erasmus, S.J.J. (1983) *J. Microsc.* **130**,187.

Tang, D., Jansen, J., Zandbergen, H.W., and Schenk, H. (1995) *Acta Cryst.* **A51**, 188.

Trus, B.L., Unser, M., Pun, T., and Stevens, A.C. (1992) *Scanning Microscopy Supplement* **6**, p. 441.

van Heel, M., Winkler, H., Orlora, E., and Schatz, M. (1992) *Scanning Microscopy Supplement* **6**, p. 23.

Zou, X. (1995) *Electron Crystallography of Inorganic Structures,* Chemical Communications, Stockholm University, Stockholm.

Other Imaging Techniques

31

CHAPTER PREVIEW

What we've discussed in the preceding nine chapters comprises "classical" TEM imaging, based on BF or DF techniques. 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 and 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 about them 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 de-

tectors 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.

Other Imaging Techniques

31

31.1. STEREO MICROSCOPY

You should 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 effect at all.

> You have to make sure that the field of view, the contrast, and the magnification in each of the two images are the same.

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.

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 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 Goldstein *et al.* (1992), for more details on stereo viewing.) Of course, for TEM images this difference is irrelevant. TEM applications are reviewed by Hudson (1973) 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 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 should be shifted slightly relative to each other and 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, noneucentric, tilt axis), refocus *using the specimen-height (z) control;* otherwise you'll change your image magnification. Obviously, computerized stages will help in this respect.
- Take another photograph.
- Develop the images and observe them under a stereo viewer.

Figure 31.1 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

$$\Delta h = \frac{\Delta p}{2M \sin \frac{\phi}{2}} \qquad [31.1]$$

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. For quantitative stereo measurements you have to enter the field of stereology, which is a discipline on its own and something that we can't cover at all, so you need to go and read an appropriate text, such as Russ (1990).

31.2. 2½D MICROSCOPY

This is one of the few examples of imaginative terminology in the TEM field, and is due to Bell (1976). If there are diffraction spots in the SAD pattern 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 seeing a pseudo-stereo technique because the "depth" difference is due to a difference in **g**, not a true depth difference at all: hence the term "2½D" or "not quite

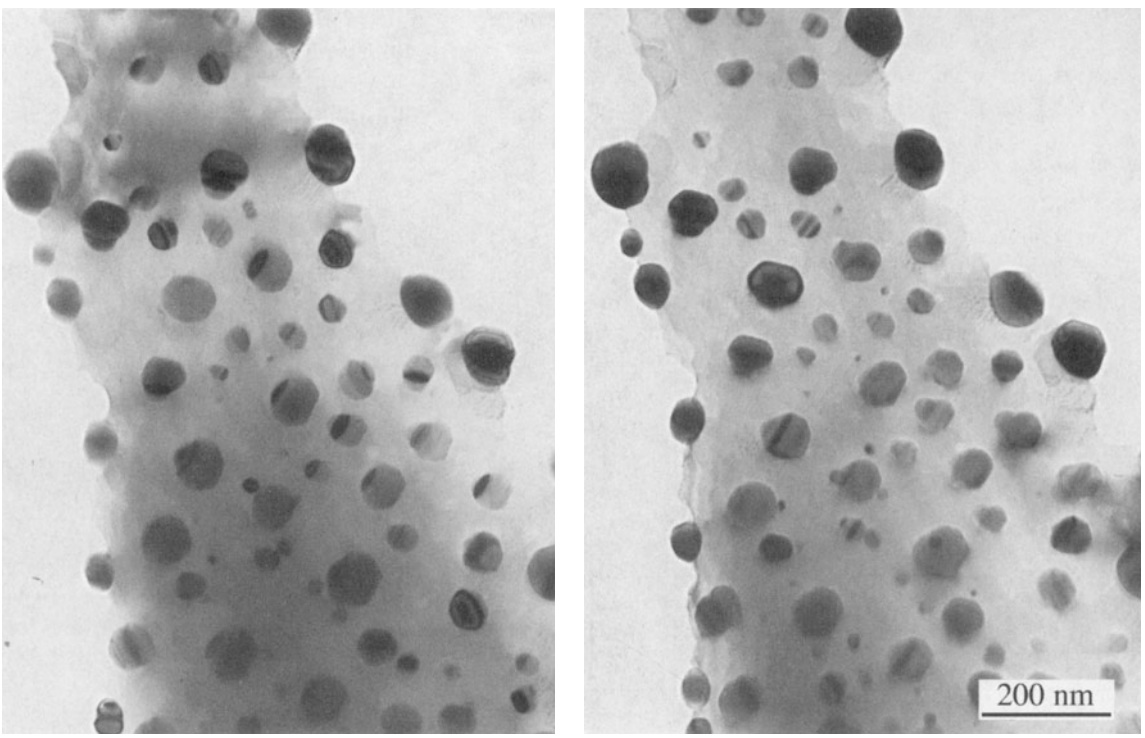


Figure 31.1. A stereo pair of spinel precipitates in a NiO-Cr_2O_3 specimen which shows the relative depth of the precipitates. You can see the small parallax shift between the two images.

3D." The technique is also called by the rather more staid term "through-focus DF."

Bell developed a simple theory explaining how a change in focus Δf introduces a parallax shift in the image, y

$$y = M\Delta f\lambda\mathbf{g} \qquad [31.2]$$

$$\Delta y_{12} = \Delta f\lambda\mathbf{g}_{12} \qquad [31.3]$$

where M is the magnification, λ is the electron wavelength, and $\mathbf{g}$ 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 that 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 31.3. 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 your ability to resolve similarly diffracting features, so this process works better at lower magnifications where the depth of field is greater.

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 the images and view them in a stereo viewer.

Figure 31.2 shows a pair of CDF images taken under these conditions. The DP, as you can see, has many closely

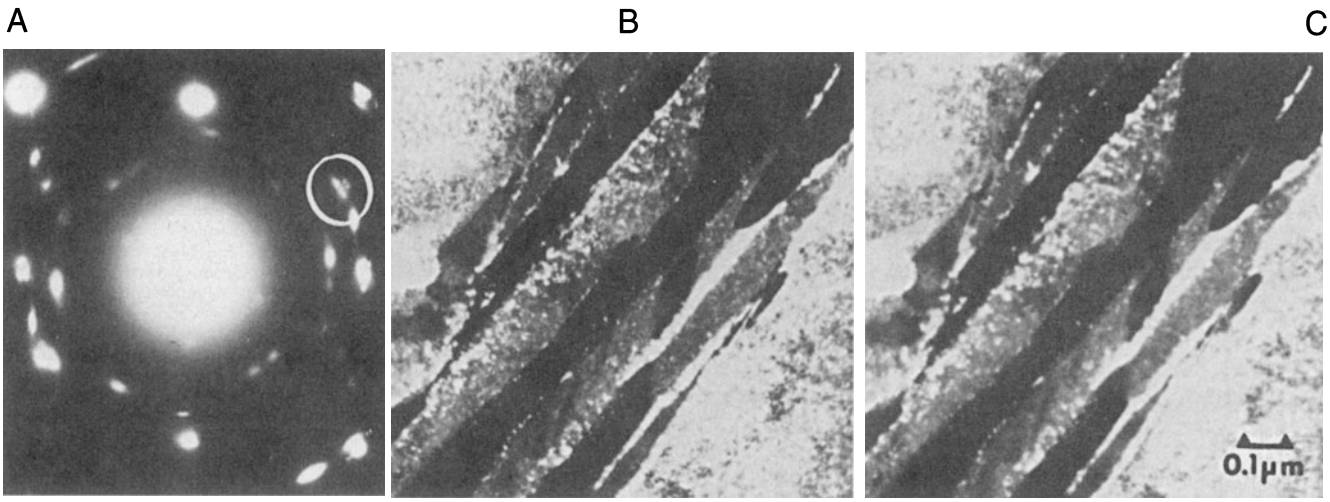


Figure 31.2. (A) SAD pattern of retained austenite and carbide precipitates in steel. (B,C) Stereo pair showing $2\frac{1}{2}$D effect in which the relative depths of the austenite and carbide are related to the position of their diffraction spots in the objective aperture in (A).

spaced spots. This example and many more are given in a well-illustrated review by Sinclair *et al.* (1981). 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 then which features in the image are responsible for which diffraction spot.

31.3. 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. TEM study of magnetic effects has seen somewhat of a resurgence since the discovery of high-T_c superconductors, and the growth of magnetic recording media.

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.

31.3.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 your whole specimen as thin and small as possible to reduce its total magnetic field strength.

> You should not use self-supporting magnetic disk specimens unless your material is brittle and likely to break in the microscope.

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. 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.

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 try and rotate and either gets pulled out of the holder (if you're unlucky) or shifts slightly in height and position.

> If you lose your specimen it is paramount that you stop, switch off the microscope, and get help.

The column must be split and the lost specimen retrieved, otherwise you may introduce a fixed astigmatism into the microscope and, 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 aperture, 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 traverses.
- 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 aperture 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 re-do 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 $hk\ell$ maximum with the other set. Figure 31.3 shows the image of a mag-

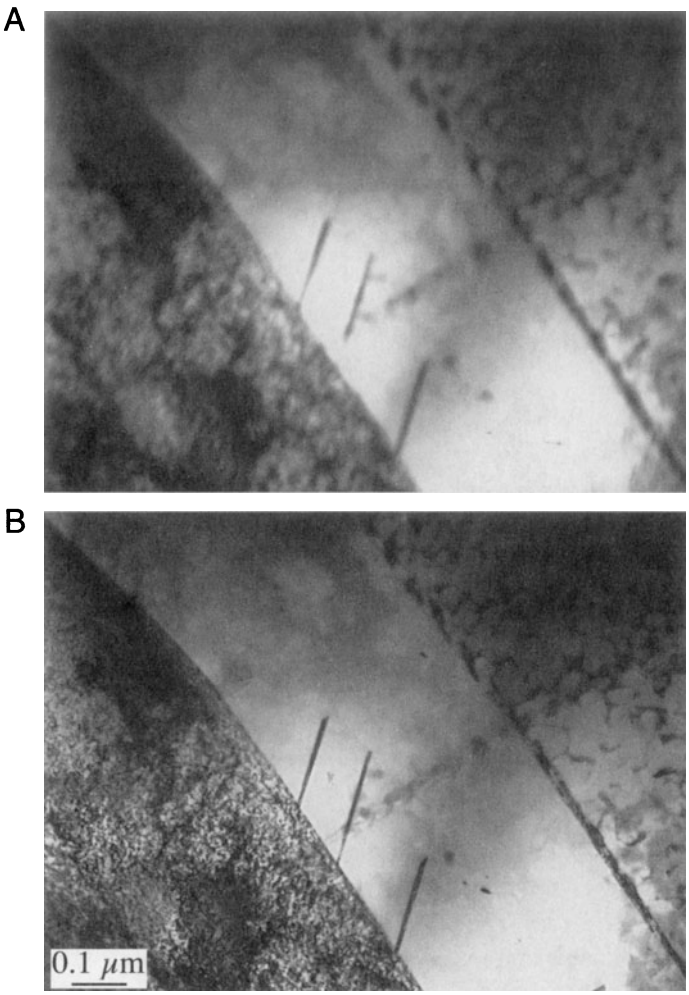


Figure 31.3. The effect of the magnetic correction on the quality of the image of an Fe-Ni meteorite: (A) before magnetic correction, (B) after correction.

netic 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.

31.3.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 mentioned back in Chapter 27. The general term for this kind of imaging is called Lorentz microscopy (Chapman 1984) and it comes with two options.

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 31.4A. It is useful to know immediately that the direction of magnetization is normal to the direction of spot splitting. You can then take a BF image using either all or one of the split 000 reflections. If you use all the spots, you can just see the domain walls (Figure 31.4B). But the domains bending electrons into the chosen spot will then appear bright and the other domains will appear darker, as in Figure 31.4C.

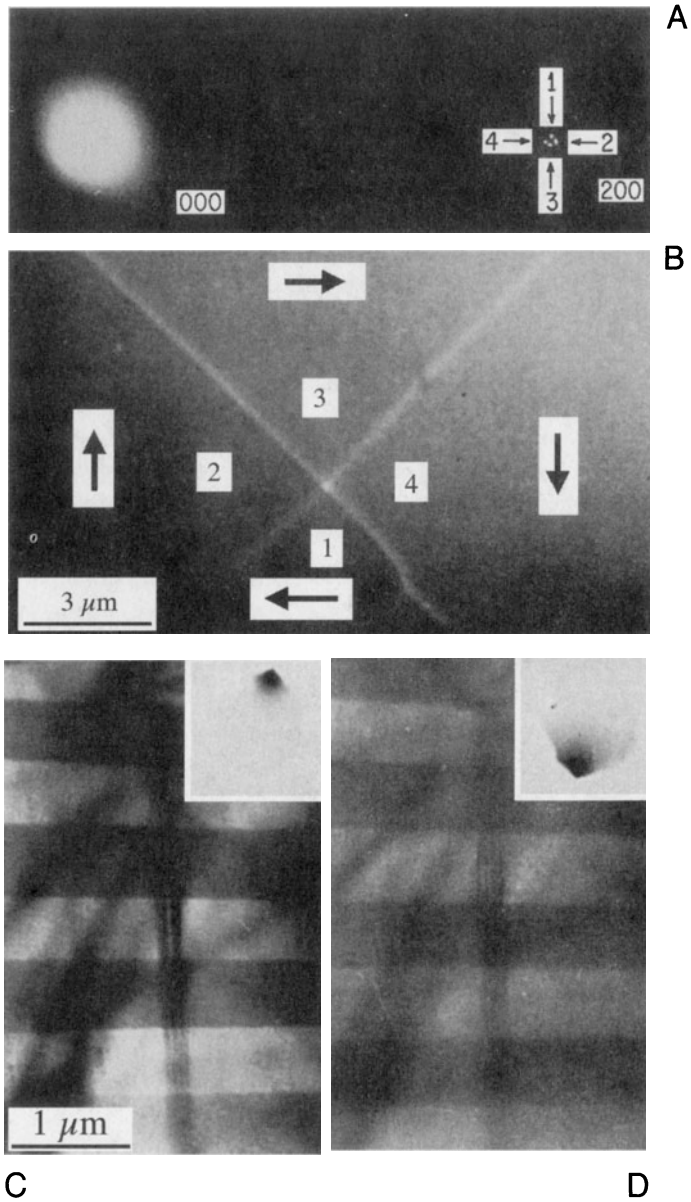


Figure 31.4. (A) Splitting of the 200 spot from Ni_3Mn due to the presence of magnetic domains. (B) Image taken from all of the split spots in (A) showing the four domains which scattered electrons into the various spots. (C,D) Foucault images of domains in Co formed by displacing the objective aperture to select one of two split spots as shown in the insets.

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 31.5, because you are picking up each signal digitally (Craven and Colliex 1977). 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 27, Fresnel imaging is named for another famous Frenchman whose micrographs were never 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 dark lines, which reverse contrast as you go through focus as in Figure 31.6. The contrast depends on whether the electrons going through the domains either side of the wall were deflected

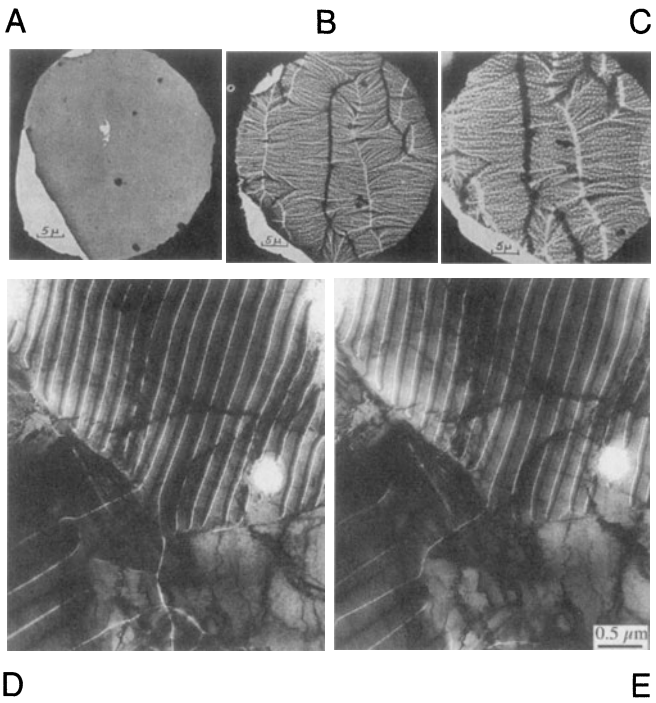


Figure 31.6. (A) In-focus image showing no magnetic contrast. (B,C) Fresnel-defocus images showing the magnetic domain walls which image as bright and dark lines and increase in width with defocus. (D) Underfocus and (E) overfocus images showing reversal of domain-wall contrast.

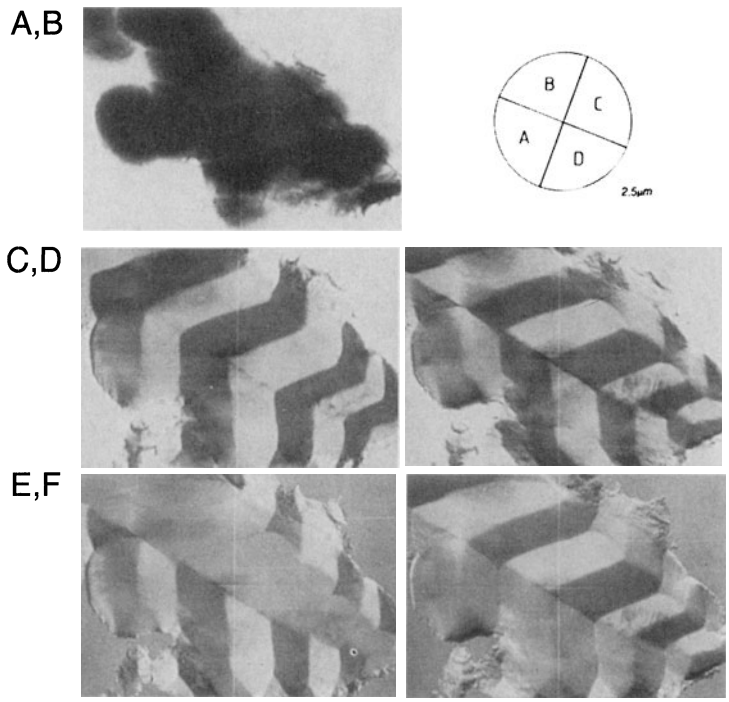


Figure 31.5. The use of a quadrant detector in STEM to differentiate regions of different magnetic induction in pure Fe. (A) The BF image from all four quandrants shown schematically in (B). (C–F) Images formed by different quadrant combinations. Regions showing the same intensity are regions of like induction.

toward or away from each other, as shown schematically in the diagram (Figure 27.16A). From such images you can work out whether you're looking at Bloch walls, Néel walls, or cross-tie walls. Figure 31.7 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 (Chapman *et al.* 1994).

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 zero at the Curie temperature.

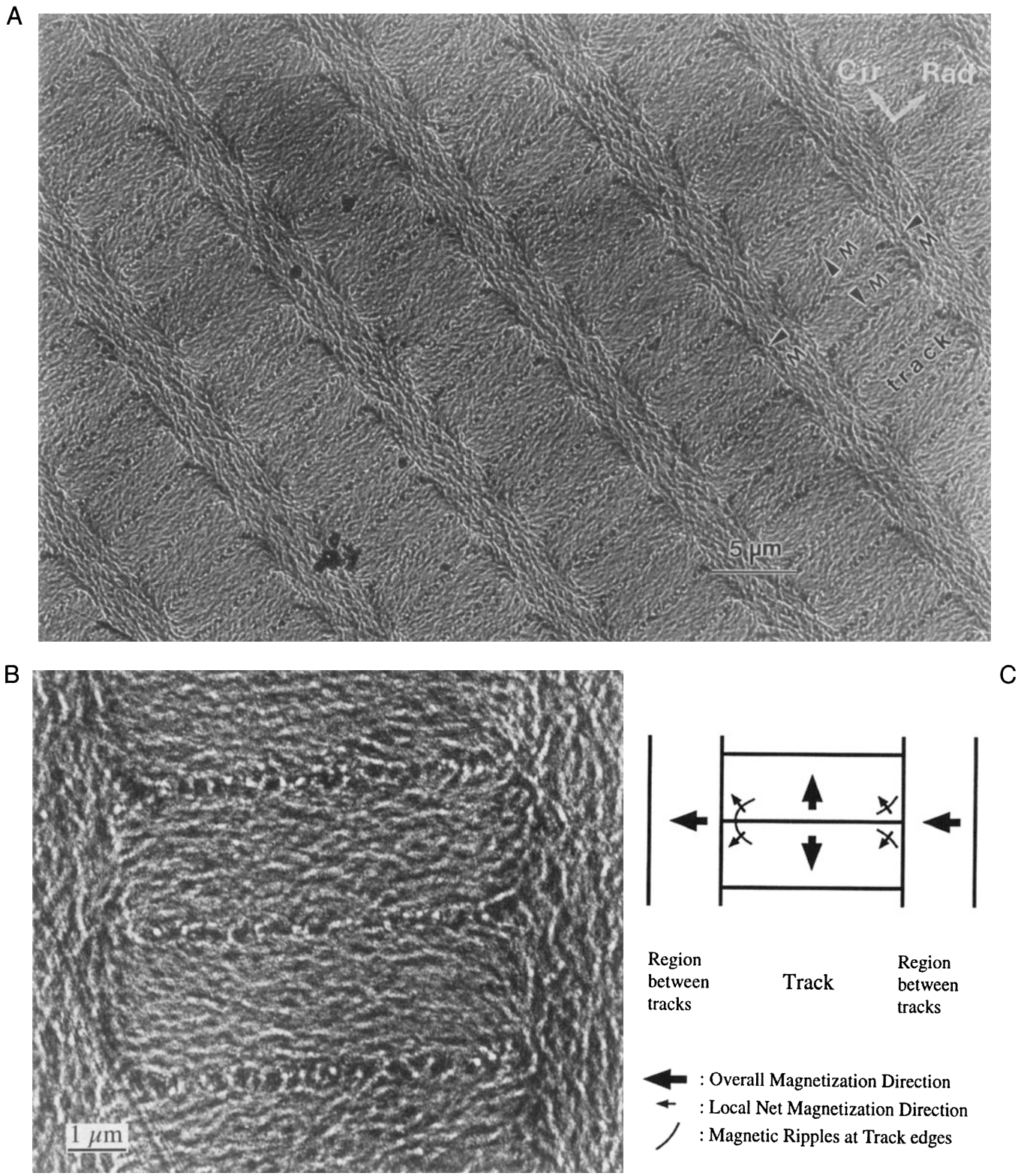


Figure 31.7. (A) Fresnel image of $Co_{84}Cr_{10}Ta_6$/Cr film on a smooth NiP/Al substrate used for magnetic data storage, imaged in the bits-written magnetic state. (B) Fresnel image at higher magnification. (C) Schematic of magnetic ripples at the track edges. The bits were in alternating direction of magnetization along the tracks in the circumferential direction of the hard disk, while the inter-track regions had remanent magnetization in the radial direction, perpendicular to that within the bits.

31.4. CHEMICALLY SENSITIVE IMAGES

In Section 16.4 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 31.8. 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 used the same information when discussing quantitative chemical lattice imaging in Section 30.11; we also use this effect when studying site location by ALCHEMI (see Section 35.8). 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 23.10) of surface relaxations in the thin specimen which can influence any diffraction-contrast images.

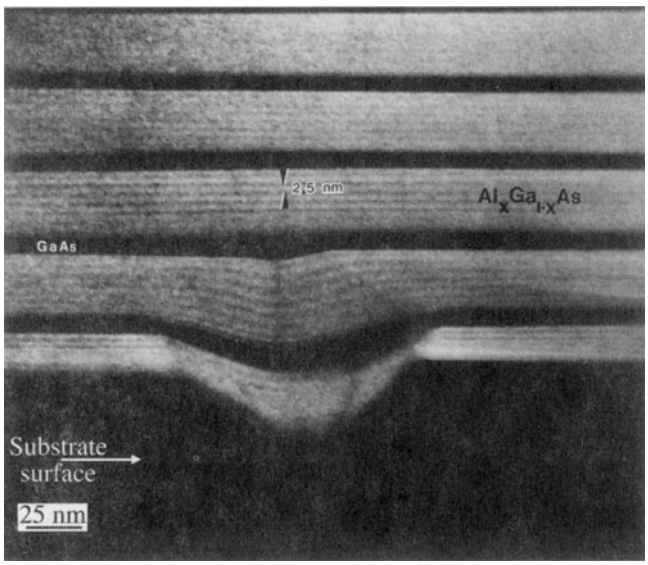


Figure 31.8. A chemically sensitive image of a GaAs-AlGaAs quantum-well structure. The composition of the AlGaAs is nonuniform because of growth fluctuations and the substrate surface was imperfectly covered by the first GaAs layer.

31.5. IMAGING WITH DIFFUSELY SCATTERED ELECTRONS

This is just a variation on the theme of normal DF imaging. If your specimen contains noncrystalline regions which scatter electrons weakly compared to the diffraction spots, then the noncrystalline regions can be seen in strong contrast. To do this, you perform a CDF operation with the objective aperture centered 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} (Clarke 1979). As shown in Figure 31.9, a DF image reveals the amorphous regions with high contrast at the grain boundary in a ceramic bicrystal. However, you must be very careful when interpreting such images, as shown by Kouh *et al.* (1986).

Diffuse scatter can also arise from short-range ordering in the specimen due to either microdomains of ordered nuclei or local regions of increased order analogous to spinodal 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 intensity maxima. Remember that we saw similar contrast effects, *but in the DPs* of short-range ordered materials in Figure 17.11.

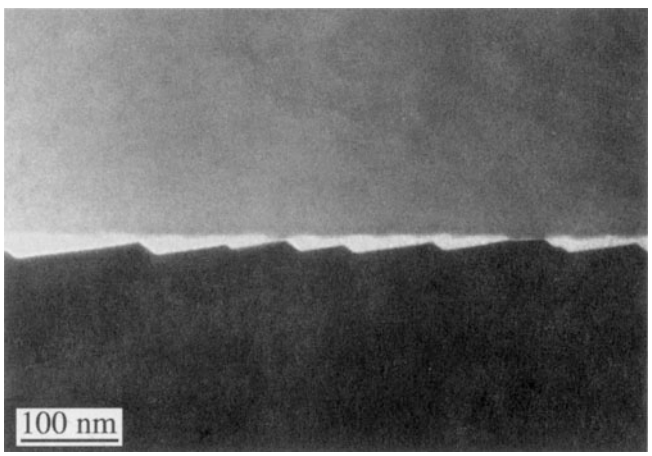


Figure 31.9. Diffuse-scatter DF image of a glassy material at a GB in a bi-crystal of Al_2O_3.

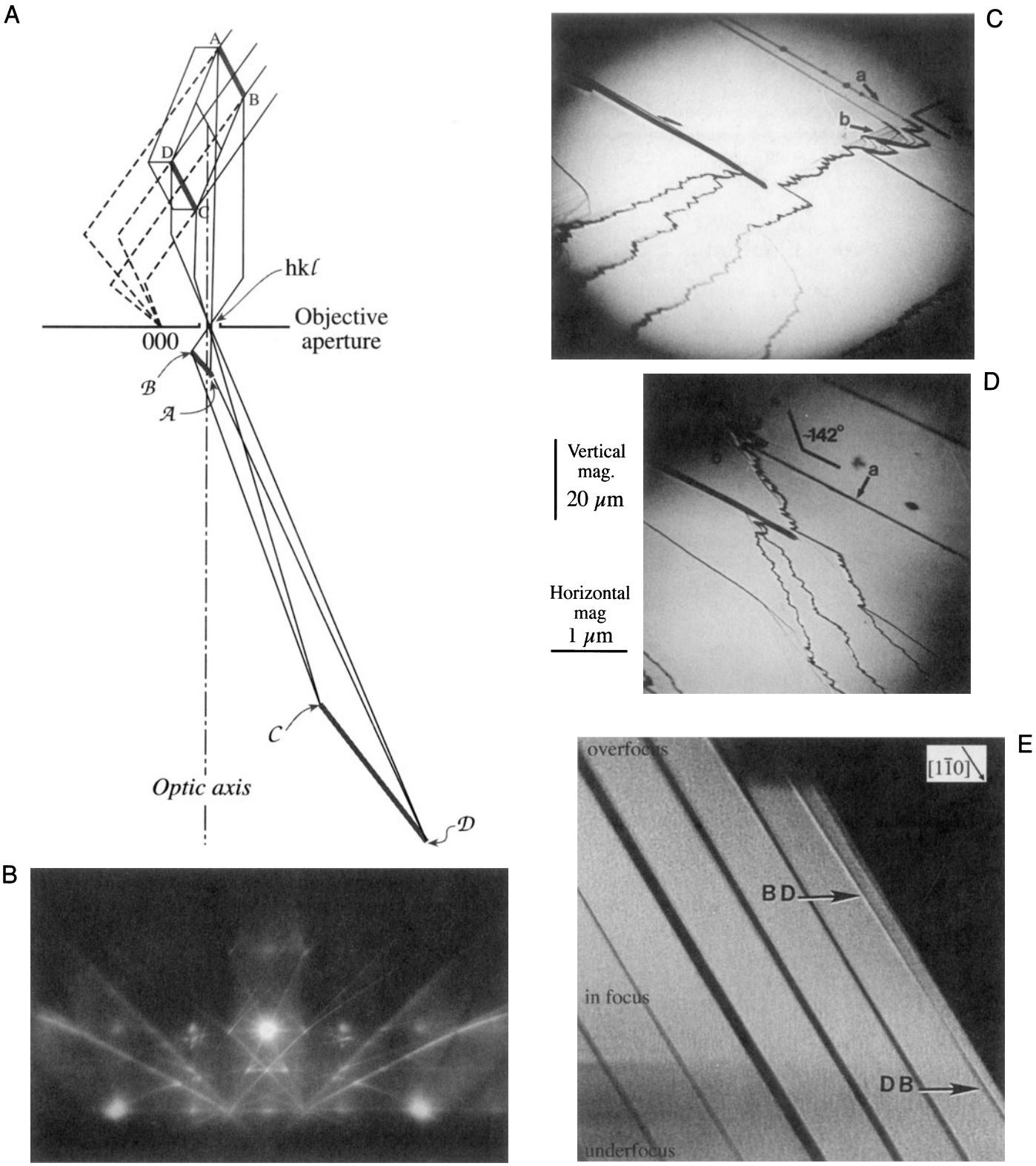


Figure 31.10. (A) Schematic diagram of the formation of an image using REM. (B) A 100 REM DP from a Si 001 surface. You can see the square array of spots but the 000 reflection has been blocked by the specimen itself. (C,D) REM images showing surface steps on a cleaved single crystal of GaAs. The two 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.

31.6. SURFACE IMAGING

We can get surface information in the TEM in a variety of ways. We can do reflection electron microscopy (REM) (Hsu 1992), 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.

31.6.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 31.10A. Since the electron is 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 31.10B. This diffraction geometry is exactly that which we used to derive Bragg's Law.

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 Figures 31.10C 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 which 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 31.10E. 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 (De Cooman *et al.* 1984).

Since you can do REM using any regular TEM holder, you can easily heat or cool your bulk sample. In many ways, this is easier than using a thin 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.* (1987) showed the surface reconstruction taking place *in situ*. You'll find more details and examples in the books and journal issues listed in the general references.

31.6.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 (Joy *et al.* 1976) but can be simply interpreted in terms of thickness changes in the specimen. In Figure 31.11 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

Figure 31.11. Topographic contrast from Fe_3O_4 particles on a carbon film. Ripples in the film are clearly visible. The shadow of the displaced objective crosses the field of view.

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.

31.7. 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 31.12 (Bell and Thomas 1972). You can get a similar effect called "kinematical diffraction imaging" if you form a BF image when no $hk\ell$ spot is strong.

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 31.12 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 than the usual **g** reflection. The point is, as usual, be prepared to experiment when you are sitting at the microscope. Many standard texts and many standard microscopists assume that you will be working at 100 kV, but only because that was all that was routinely available twenty years ago. The world is now different.

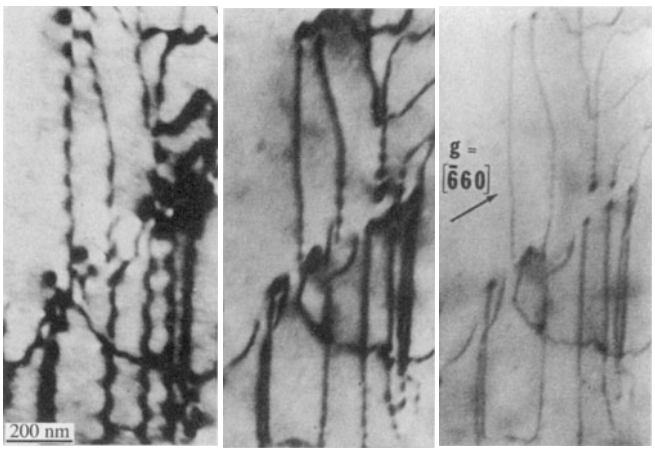


Figure 31.12. BF images of dislocations in GaAs taken with **g, 2g, 3g.** In the higher-order BF images the width of the defects decreases.

31.8. 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 is different than the SEM, in which the SE detector is situated under the final polepiece, and in the STEM we get SE images of superior resolution and quality. There are several reasons for this:

- An SE image in an SEM invariably has noise contributions from BSEs which can enter the scintillator directly. But 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.
- 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 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 31.13, the SE detector in an SEM can pick up four different types of secondary electrons, labeled SE I - SE IV (Peters 1984):

- The SE I signal is the only signal we want, since it emanates from the region around the

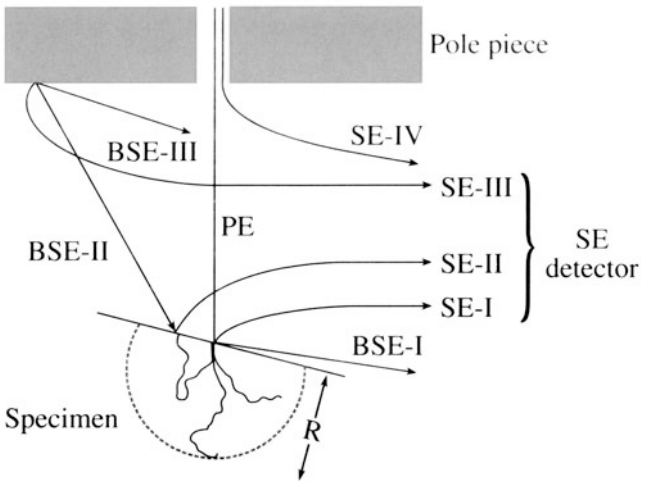


Figure 31.13. The four possible sources of secondary electrons that enter the detector in a conventional SEM. The SE I signal is the only desirable one since it comes from the probe region, but SE II from BSE electrons, SE III from the stage of the microscope and SE IV from the final aperture all combine to reduce the *S/N* ratio in an SE image.

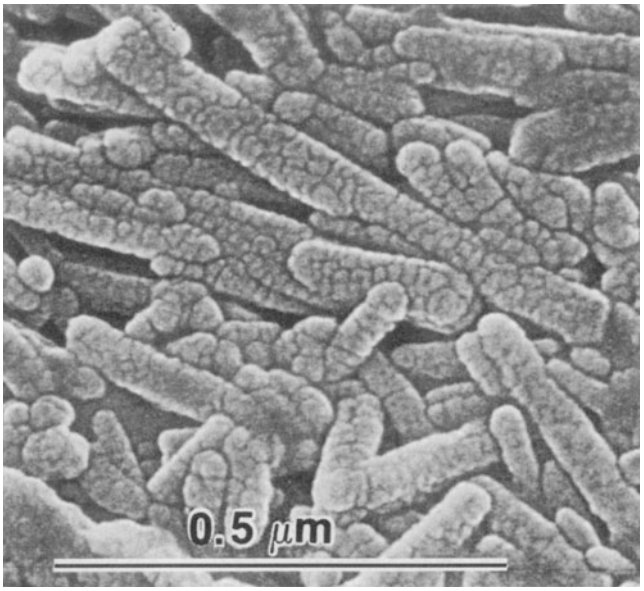


Figure 31.14. High-resolution SE image of coated magnetic tape in a TEM/STEM with an LaB_6 source at 100 kV showing ~2 nm spatial resolution.

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.
- 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 a bulk sample than a conventional SEM, as shown in Figure 31.14. The highest-resolution FEG SEMs incorporate many of the design aspects of the STEM stage and can now produce SE image resolutions < 1 nm at 30 kV. A thermionic source STEM cannot better that performance, however an FEG STEM at 100 kV or higher should, in theory, offer the best possible SE images (Imeson 1987) and may even show atomic-level topographic information.

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 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.

31.9. 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). 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 resolutions 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 31.15. 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 manufacturer of dedicated STEMs does 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.

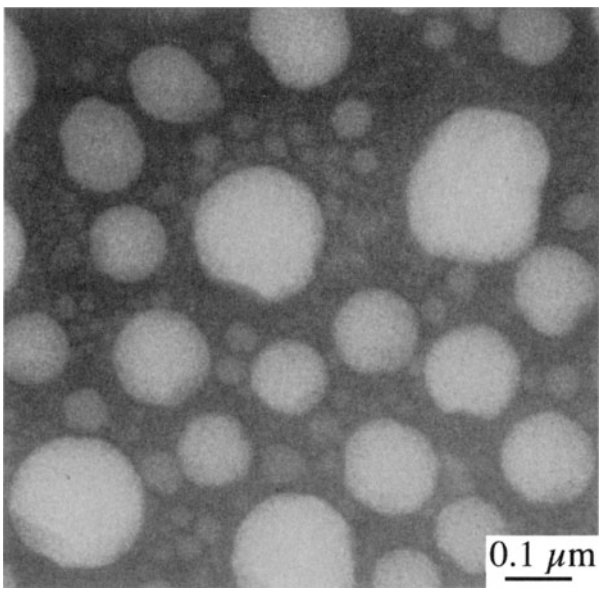


Figure 31.15. High-resolution STEM BSE image of Au islands on a carbon support film obtained in a TEM/STEM at 100 kV with an LaB_6 source. Resolution of ~7 nm is obtained, although the image quality is not very good because of the poor signal level.

31.10. 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 (Newbury *et al.* 1986). 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 you 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 nonradiative. 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.* (1978) and more recently by Batstone (1989). It is not obvious, however, that doing the work in transmission offers major advantages over studies of bulk samples in the SEM.

31.11. ELECTRON HOLOGRAPHY

Although the technique of electron holography achieved prominence 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 article by Cowley (1992) entitled "Twenty forms of electron holography." We'll discuss the general principles of the technique and refer you to the articles given in the general 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 (see Cowley 1992, for more).

- 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 31.16 shows a hologram of a wedge-shaped crystal of Si oriented close to the [110] pole. The boxed region of the figure shows a set of fringes which are only 0.7 Å apart (Lichte 1992). 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 did not cover it in Chapter 30.

The principle of the technique is shown by the schematic in Figure 31.17. Here, $\chi(\mathbf{u})$ is the function we used in Chapter 29 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, $\mathcal{A}$, 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

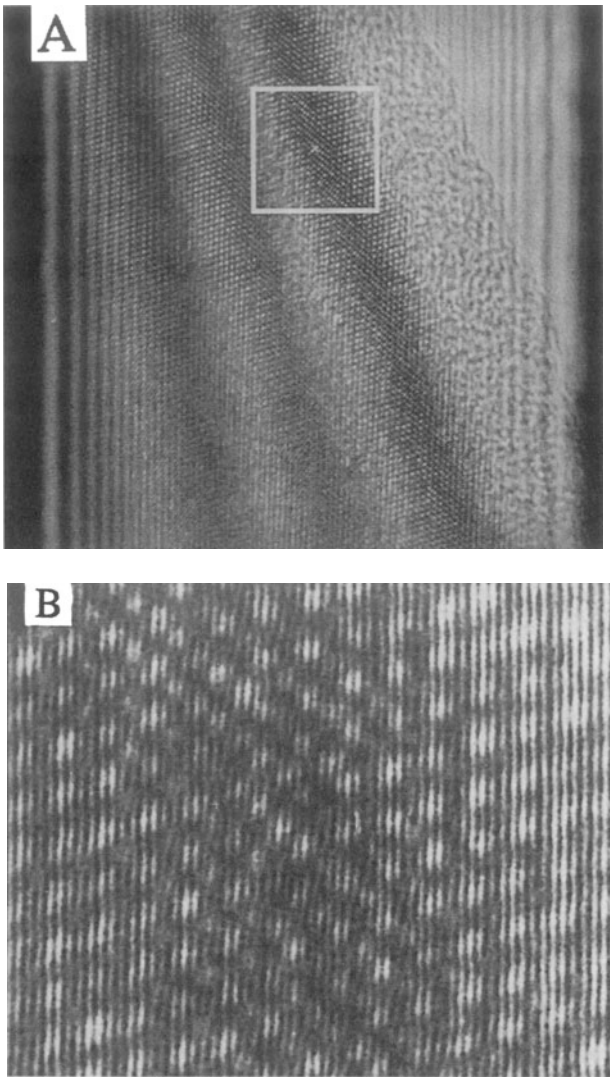


Figure 31.16. (A) Hologram of a [110]-oriented Si specimen; the broad bands are thickness fringes where the specimen is thickest at the lower left corner. (B) An enlargement of the boxed region in (A) showing with a 0.07 nm spacing.

to improve the resolution of the TEM, is being actively pursued wherever there is an FEG TEM (e.g., Harscher *et al.* 1995). Although holography has still not reached its full potential in this field, you can see the future by looking above at Figure 31.16.

In practice, electron holography is carried out using an FEG TEM which has been fitted with a beam splitter. 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 27.7 and can be < 0.5 μm in diameter. Part of the beam passes through the specimen while the other part forms the reference beam, as shown in Figure 31.18. You must be able to rotate either the biprism or the specimen. For holography,

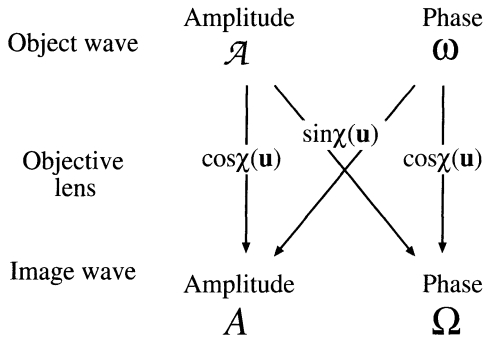


Figure 31.17. A schematic of how the objective lens mixes the amplitude and phase components of the specimen to form the amplitude and phase components of the image.

it's placed below the objective lens, e.g., at the position usually occupied by the SAD aperture rod.

The essential feature of the experiment is that a reference beam passes outside your specimen and is then deflected by the biprism so that it interferes with the beam that passed through the specimen.

All that you then have to do is interpret the interference pattern, i.e., the electron hologram, but as you'd guess after reading Chapter 30, this process is not trivial. However, the technique does provide the possibility for "coherent processing" of the electron wave (Lichte 1992). In principle, all the required data can be recovered from one hologram

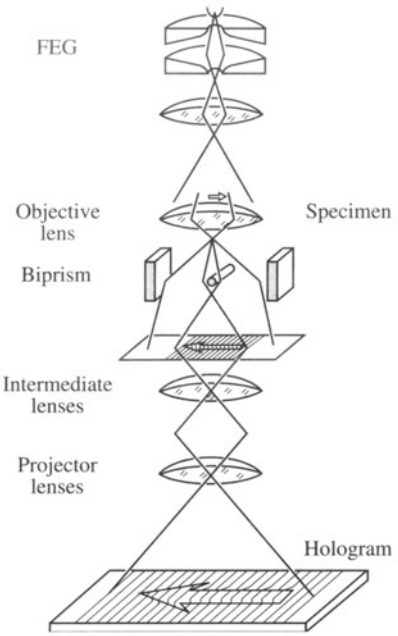


Figure 31.18. The operation of the biprism in electron holography.

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 (1987). Figure 31.19 shows a series of images illustrating how magnetization

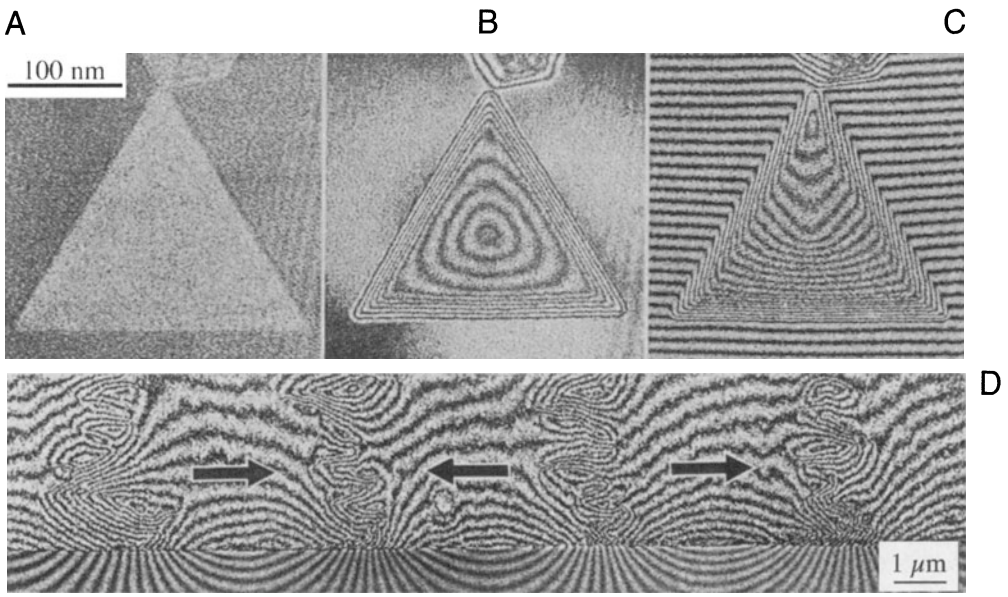


Figure 31.19. Holographic images of a magnetic Co particle showing (A) the reconstructed image, (B) the magnetic lines of force, and (C) the interferogram. (D) Lines of force in a magnetic recording medium.

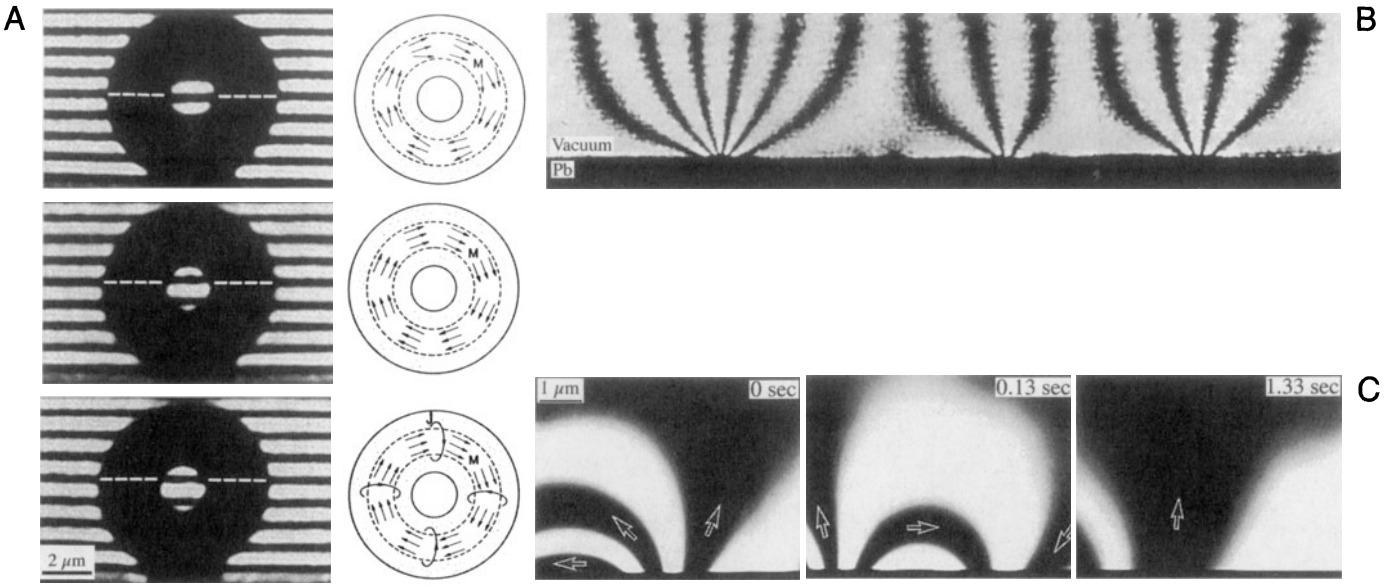


Figure 31.20. (A) Gradually cooling a ring (torus) of a superconducting material from 300K to 15K to 5K to demonstrate the quantization of flux. (B) Magnetic flux lines penetrating a superconducting Pb film. (C) Interference of fluxons trapped in superconducting Pb as time increases from 0 s to 0.13 s to 1.3 s.

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 31.20A. As the ring is gradually cooled, the phase shift due to the magnetization changes until it becomes superconducting, when the phase shift inside the ring becomes exactly π. The other images in Figures 31.20B,C show flux lines constricting as they enter the superconductor, but notice that you can clearly see these flux lines outside the specimen: the magnetic field outside the specimen also influences the electron beam.

31.12. *IN SITU* TEM: DYNAMIC EXPERIMENTS

We've mentioned several times that it is possible to do *in situ* experiments inside the TEM. 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 Chapter 30 we talked about 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 implicit 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 25.5, when the surface properties dominate the bulk, as is often the case in thin specimens, TEM images and analyses can be misleading. Surface diffusion is much more rapid than bulk diffusion and defects are subject to different stress states. 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 70s when 1–3 MV instruments were first constructed and used to look through foils which were > 1 μm thick (see the text by Butler and Hale 1981). So *in situ* experiments were expensive!

However, the advent of 300–400 kV IVEMs and the construction of new 1.25-MV instruments in Stuttgart and Tsukuba has brought a resurgence of interest in *in situ* experimentation (see Rühle *et al.* 1994). Developments in electron optics, stage design, and recording media since the 1960s mean that combined HRTEM and *in situ* experimentation is now possible at intermediate voltages, as demonstrated in the beautiful images of Sinclair and Konno

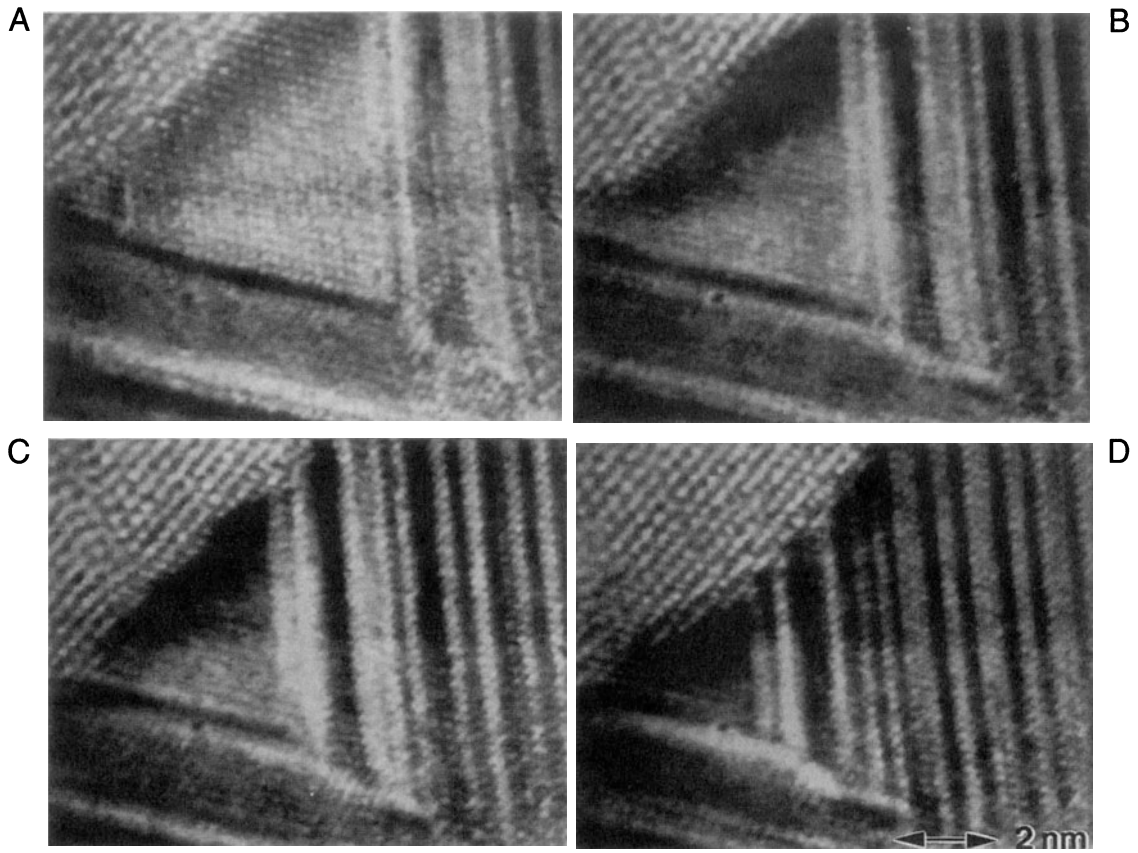


Figure 31.21. (A–D) Four images taken from a video showing the reaction front of a Ge/Ag/Ge trilayer while heating the specimen *in situ* at 250°C. The time interval between the images was 8 s. The Ge crystal grows from the upper left without moving the Ag lattice.

(1994). 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 31.21, although lower-resolution images are no less impressive.

However, 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 electron flux brings an uncertainty to all *in situ* experiments, particularly if you want to try and 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.

31.13. 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 (Chapter 35) and EELS (Chapter 40).

In a STEM then, we can pick up, in theory, 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 as well as 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 falls into the category of digital image processing which you may study in specialist texts such as Russ (1990), after you've read Chapter 30.

CHAPTER SUMMARY

We've described only some 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 (Williams and Newbury 1984).

We haven't discussed the use of annular apertures, conical illumination, pre- and post-specimen beam scanning, or rocking. All of these are possible, so you should always 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

General References

Hsu, T., Ed. (1992) *J. Electron Microsc. Tech.* **20**, part 4. A special issue devoted to REM.

Larsen, P.K and Dobson, P.J., Eds. (1987) *Reflection High-Energy Electron Diffraction and Reflection Electron Imaging of Surfaces,* Plenum Press, New York.

Electron holography: The following key articles will give you a sense of the possible applications:

Gabor, D. (1949) *Proc. Roy. Soc. London* **197A**, 454; (1951) *Proc. Roy. Soc. London* **64**, 450. These are parts 1 and 2 of a pair of articles entitled "Microscopy by Reconstructed Wave Fronts."

Hansen, K.-J. (1982) *Adv. in Electronics and Electron Physics* **59**, 1.

Tonomura, A. (1987) *Rev. Mod. Phys.* **59**, 639; (1992) *Adv. Phys.* **41**, 59.

Ru, Q., Endo, J., Tanji, T., and Tonamura, A. (1991) *Appl. Phys. Lett.* **59**, 2372.

Spence, J.C.H. (1992) *Optik* **92**, 52.

Franke, F.-J., Hermann, K.-H., and Lichte, H. (1988) in *Image and Signal Processing for Electronic Microscopy,* (Eds. P.W. Hawkes, F.P. Ottensmeyer, A. Rosenfeld, and W.O. Saxton), p. 59, Scanning Microscopy Supplement **2**, AMF, O'Hare, Illinois.

Völkl, E., Allard, L.F., Datye, A., and Frost, B. (1995) *Ultramicroscopy* **58**, 97.

Steeds, J.W., Vincent, R., Vine, W.J., Spellward, P., and Cherns, D. (1992) *Acta Microsc.* **1.2**, 1. So that you don't forget the value of using an FEG for diffraction.

Specific References

Batstone, J. (1989) *Microbeam Analysis-1989* (Ed. P.E. Russell), San Francisco Press, San Francisco, California.

Bell, W.L. (1976) *J. Appl. Phys.* **47**, 1676.

Bell, W.L. and Thomas, G. (1972) *Electron Microscopy and Structure of Materials* (Ed. G. Thomas), p. 23, University of California Press, Berkeley, California.

Butler, E.P. and Hale, K.F. (1981) in *Practical Methods in Electron Microscopy* **9**, (Ed. A.M. Glauert), North-Holland, New York.

Chapman, J.N. (1984) *J. Phys. D* **17**, 623.

Chapman, J.N., Johnston, A.B., Heyderman, L.J., McVitie, S., and Nicholson, W.A.P. (1994) *IEEE Trans. Magn.* **30**, 4479.

Clarke, D.R. (1979) *Ultramicroscopy* **4**, 33.

Cowley, J.M. (1973a) *Acta Cryst.* **A29**, 529.

Cowley, J.M. (1973b) *Acta Cryst.* **A29**, 537.

Cowley, J.M. (1992) *Ultramicroscopy* **41**, 335.

Craven, A.J. and Colliex, C. (1977) *J. Microsc. Spectrosc. Electr.* **2**, 511.

De Cooman, B.C., Kuesters, K.-H., and Carter, C.B. (1984) *Phil. Mag. A* **50**, 849.

Goldstein, J.I., Newbury, D.E., Echlin, P., Joy, D.C., Romig, A.D. Jr., Lyman, C.E., Fiori, C., and Lifshin, E. (1992) *Scanning Electron Microscopy and X-ray Microanalysis,* 2nd edition, Plenum Press, New York.

Harscher, A., Lang, G., and Lichte, H. (1995) *Ultramicroscopy* **58**, 79.

Hudson, B. (1973) *J. Microsc.* **98**, 396.

Imeson D. (1987) *J. Microsc.* **147**, 65.

Joy, D.C., Maher, D.M., and Cullis, A.G. (1976) *J. Microsc.* **108**, 185.

Kouh, Y.M., Carter, C.B., Morrissey, K.J., Angelini, P., and Bentley, J. (1986) *J. Mat. Sci.* **21**, 2689.

Lichte, H. (1992) *Signal and Image Processing in Microscopy and Microanalysis,* Scanning Microscopy Supplement, **6**, p. 433, SEM Inc., AMF O'Hare, Illinois.

Newbury D.E., Joy. D.C., Echlin, P., Fiori, C.E., and Goldstein, J.I. (1986) *Advanced Scanning Electron Microscopy and X-ray Microanalysis,* p. 45, Plenum Press, New York.

Peters, K.R. (1984) *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography,* (Eds. D.F. Kyser, H. Niedrig, D.E. Newbury, and R. Shimizu), p. 363, SEM Inc., AMF, O'Hare, Illinois.

Petroff, P., Lang, D.V., Strudel, J.L., and Logan, R.A. (1978) *Scanning Electron Microscopy* **1** (Ed. O. Johari), p. 325, SEM Inc., AMF, O'Hare, Illinois.

Rühle, M., Phillipp, F., Seeger, A., and Heydenreich, J., Eds. (1994) *Ultramicroscopy* **56**, (1-3).

Russ, J. (1990) *Computer-Assisted Microscopy: The Analysis and Measurement of Images,* Plenum Press, New York.

Sinclair, R. and Konno, T.J. (1994) *Ultramicroscopy* **56**, 225.

Sinclair, R., Michal, G.M., and Yamashita, T. (1981) *Met. Trans.* **12A**, 1503.

Williams, D.B. and Newbury, D.E. (1984) in *Advances in Electronics and Electron Physics,* (Ed. P.W. Hawkes), Academic Press, New York.

Yagi, K. Ogawa, S., and Tanishiro, Y. (1987) *Reflection High-Energy Electron Diffraction and Reflection Electron Imaging of Surfaces,* (Eds. P.K. Larsen and P.J. Dobson), p. 285, Plenum Press, New York.

Spectrometry

IV

X-ray Spectrometry

32

CHAPTER PREVIEW

To use the X-rays generated when the electron beam strikes a TEM specimen, we have to detect them first and then identify them as coming from a particular element. This is accomplished by X-ray spectrometry, which is one way we can transform a TEM into a far more powerful instrument, called an analytical electron microscope (AEM). Currently, the only kind of X-ray spectrometer that we use in an AEM is an X-ray energy-dispersive spectrometer (XEDS), which comprises a detector interfaced to signal-processing electronics and a computer-controlled multichannel analyzer (MCA) display. The XEDS is a complex and rather sophisticated piece of instrumentation which takes advantage of modern semiconductor technology. The principal component of the XEDS is a semiconductor detector which has the benefit of being compact enough to fit within 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 start with the basic physics you need to understand how the detector works and give you a very brief overview of the processing electronics. We then describe a few simple tests you can perform to confirm that your XEDS is working correctly. It is really most important from a practical point of view that you know the limitations of your XEDS system. Therefore, we will describe these limitations in some detail, especially the unavoidable artifacts. Finally, we briefly mention the wavelength-dispersive spectrometer (WDS), which is used in bulk X-ray microanalysis. The WDS is old technology which might see a renaissance in the AEM in the future.

X-ray Spectrometry

32

32.1. X-RAY ANALYSIS: WHY BOTHER?

When characterizing a specimen in the TEM, the limitations of only using 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 III, the TEM gives us two-dimensional projected images of 3D transparent specimens and you, the operator, need substantial experience in order to interpret the images correctly. For example, Figure 32.1 shows six images, taken with both light and electron microscopes (can you distinguish which images are from which kind of microscope?). The scale of the microstructures varies from nanometers to millimeters and yet the images appear very similar. Without any prior knowledge it would be almost impossible, even for an experienced microscopist, to identify the nature of these specimens simply by looking at the image.

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. We will be discussing such spectra in detail later in this and subsequent chapters, but even with no knowledge of XEDS, you can easily see that each specimen gives a different spectrum.

> Different spectra mean that each specimen must have a different elemental composition and it is possible to obtain this information in a matter of minutes.

Armed with the elemental make-up of your specimen, any subsequent image and diffraction analysis is greatly facilitated. For your interest, the identity of each specimen is given in the caption to Figure 32.2. While Figures 32.2A–E are all 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. 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 other similar microstructures.

> The main message you should get from this illustration is that the *combination* of imaging and spectrometry is most powerful and this combination transforms a TEM to an AEM.

Within a very short time, you can get a qualitative elemental analysis of most features in a complex microstructure, and the important features can be isolated for full quantitative analysis, which we will address in Chapter 35. In the chapters before this you will first learn something about how the XEDS detector works and the problems that arise when the detector is inserted into the column of an AEM.

32.2. BASIC OPERATIONAL MODE

To produce spectra such as those in Figure 32.2, all that you have to do is obtain a TEM or STEM image of the area you wish to analyze. In TEM mode, you then have to condense the beam down to an appropriate size for analysis. This may mean exciting the C1 lens more strongly and changing the C2 aperture and C2 lens strength. These steps may misalign the illumination system. For this reason it is recommended that you operate in STEM mode. Create your STEM image using the appropriate C1 lens setting and C2 aperture to give the desired probe dimension. It is then a simple matter to stop the scanning probe and position it on the feature you wish to analyze. Furthermore, critical software routines that you can use to check for specimen drift during your analysis can only work via a digital image of the analysis region.

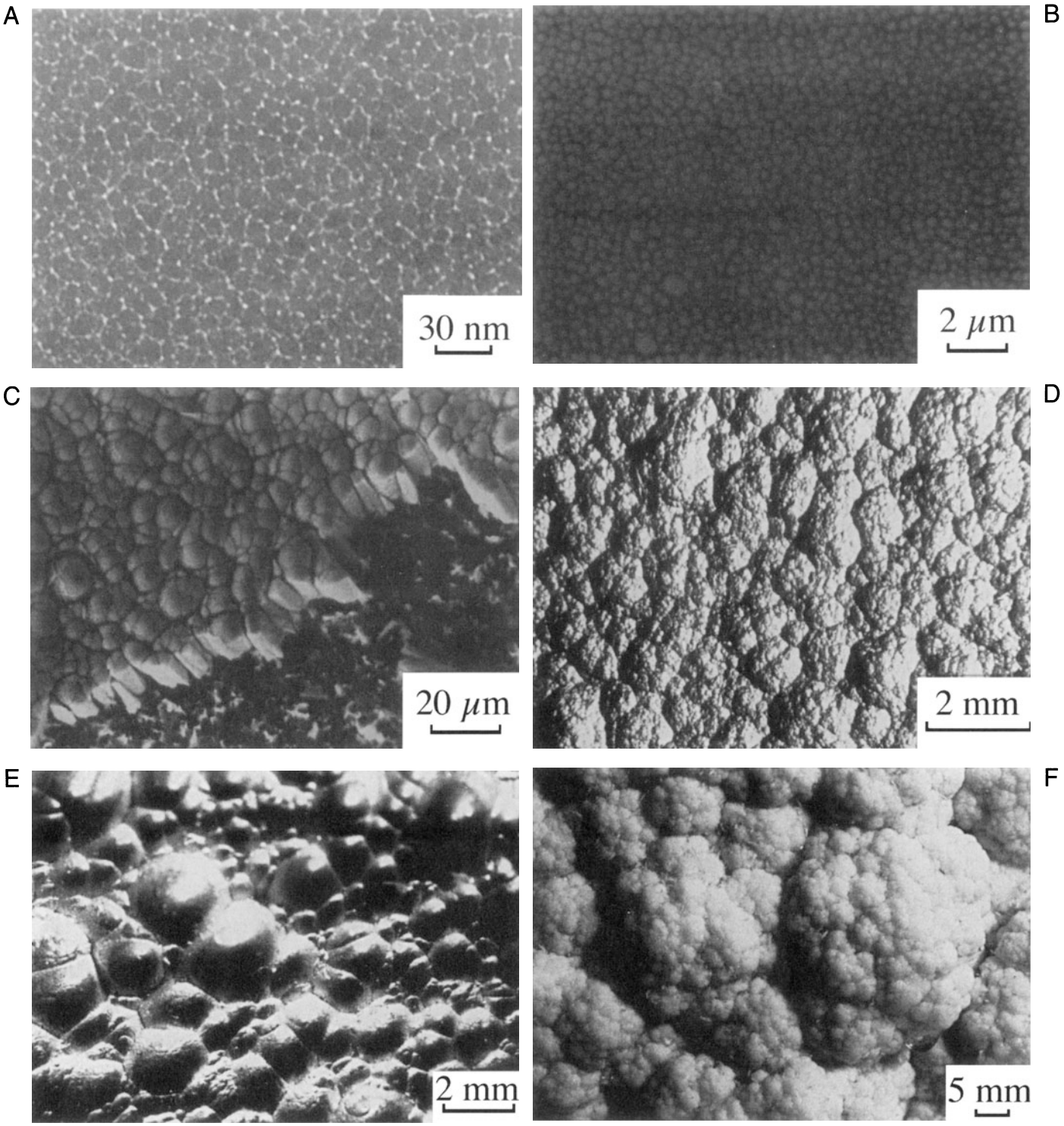


Figure 32.1. Six images of various microstructures, spanning the dimensional range from nanometers to millimeters. The images were taken with TEMs, SEMs, and light microscopes, but the characteristic structures are very similar, and it is not possible, without prior knowledge, to identify the samples.

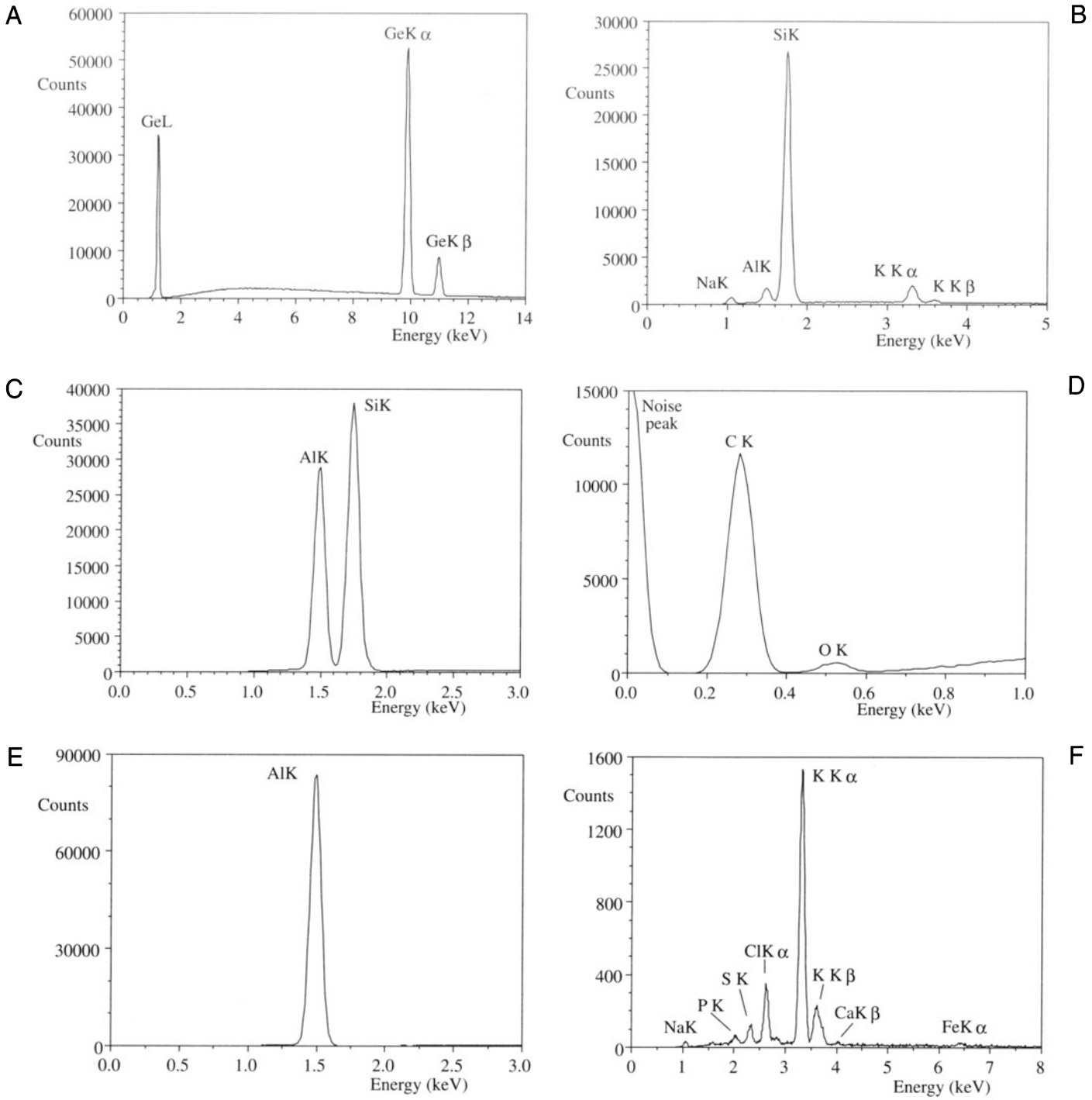


Figure 32.2. XEDS spectra from the six materials in Figure 32.1. Each spectrum is clearly different from the others, and helps to identify the samples as (A) pure Ge, (B) silica glass, (C) Al evaporated on a Si substrate, (D) pyrolitic graphite, (E) pure Al, and (F) a cauliflower.

Use STEM images to select your analysis region. This makes it easier to move between image mode and analysis mode.

Microanalysis should *always* be performed with your specimen in a low-background (Be) holder (see Chapter 8). The holder should be capable of being cooled to liquid-N_2 temperature to minimize contamination, and a double-tilt version is recommended so diffraction and imaging can be carried out simultaneously.

32.3. THE ENERGY-DISPERSIVE SPECTROMETER

The XEDS produces spectra which are plots of X-ray *counts* (imprecisely termed "intensity") versus X-ray *energy*. Before we get into details, recall from back in Chapter 4 that electrons generate two kinds of X-rays. When electrons ionize an atom, the emitted characteristic X-ray *energy* is unique to the ionized atom. When electrons are slowed by interaction with the nucleus, they produce a continuum of bremsstrahlung X-rays. The result is that, as we have seen in Figures 1.4A, 4.6, and 32.2, the characteristic X-rays appear as Gaussian-shaped peaks superimposed on a background of bremsstrahlung X-rays, most clearly visible in Figure 32.2A. Many more spectra will appear throughout this and subsequent chapters.

The XEDS was developed in the late 1960s and, by the mid-1970s, it was available as an option on many TEMs and even more widespread on other electron beam instruments, such as SEMs. This 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 and these are listed in the general reference section. Figure 32.3A shows a schematic diagram of an XEDS system and we'll deal with each of the major components as we go through this chapter.

The three main parts are the detector, the processing electronics, and the MCA display.

A 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 at a time so the detector is switched off when an X-ray signal is detected. Second, the computer controls the processing electronics, setting the time required to analyze the X-ray signal and assigning

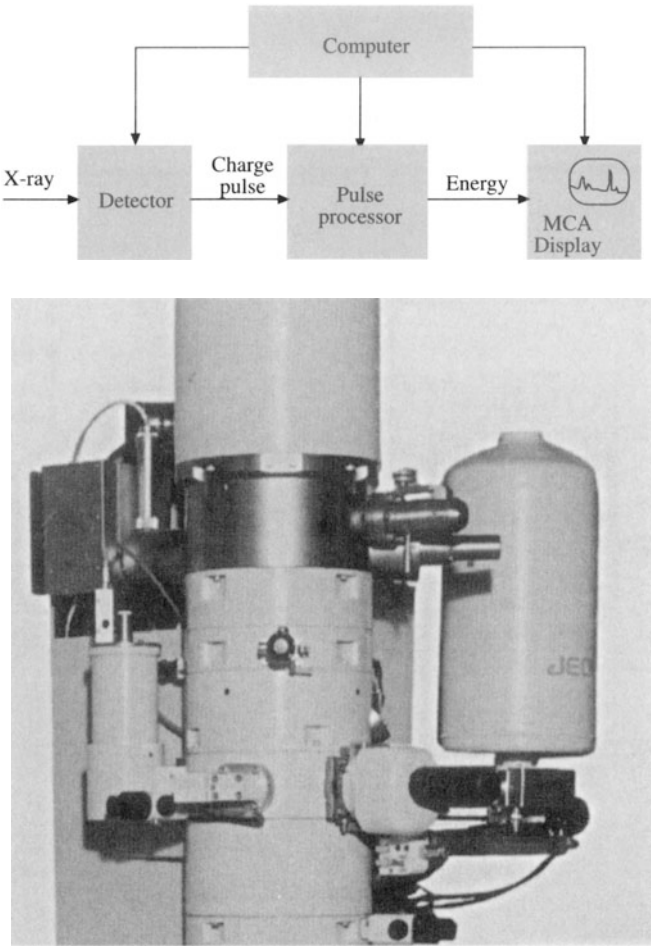


Figure 32.3. (A) Diagram of the XEDS system showing how the computer controls the detector, the processing electronics, and the display. (B) An XEDS system interfaced to the stage of an AEM. All that is visible is the large liquid-N_2 dewar attached to the side of the column.

the signal to the correct channel in the MCA. Third, the computer software governs both the calibration of the spectrum readout on the MCA screen and all the alphanumerics which tell you the conditions under which you acquired the spectrum. Any data processing is also carried out using 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.
- Then the signal is amplified through a field effect transistor (FET), isolated from other pulses, further amplified, then identified elec-

tronically as resulting from an X-ray of specific energy.
- Finally, a digitized signal is stored in a channel assigned to that energy in the MCA.

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 a detector interfaced to an AEM. In fact, you can't see the processing electronics or the MCA display, nor the detector itself because it sits close to the specimen within the microscope column. The most prominent feature that you can see is the liquid-N_2 dewar, which cools the detector.

32.4. SEMICONDUCTOR DETECTORS

The detector in an XEDS is a reverse-biased p-i-n diode. Almost all AEMs use silicon–lithium [Si(Li)] semiconductor detectors and so we will take these as our model. Later, in Section 32.4.C, we'll discuss the role of intrinsic Ge (IG) detectors, which can be useful on intermediate voltage AEMs.

32.4.A. How Does XEDS Work?

While you don't need to understand 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 interact with a semiconductor, the primary method of energy deposition is the transfer of electrons from the valence band to the conduction band, creating an electron–hole pair. High-energy electrons lose energy in Si in a similar way as we saw in Section 4.4. The energy required for this transfer in Si is ~3.8 eV at the liquid-N_2 operating temperature. (This quantity is a statistical value so don't try to link it directly to the band gap.) Since characteristic X-rays typically have energies well in excess of 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 incoming X-ray. 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. The way this is achieved is summarized in Figure 32.4, which is a schematic diagram of the Si(Li) detector.

The design of the Si(Li) detector is very similar to the semiconductor electron detectors discussed in Chapter 7.

> The electron detectors separate the electrons and holes by the internal reverse bias of a very narrow p-n junction; since X-rays penetrate matter much more easily than electrons, we need a much thicker region for the X-rays to generate electron–hole pairs and lose all their energy.

In practice, we need to have an intrinsic region between p- and n-type regions which is about 3 mm thick. So the Si should have low conductivity, with no impurity atoms to contribute electrons or holes to the charge pulse, and no defects to act as recombination sites for the electron–hole pairs. However, we still cannot make intrinsic Si on a commercial basis. It usually contains acceptor impurities and so acts as a p-type semiconductor. We compensate for the impurity effects by "filling" any recombination sites with Li, thus creating a region of intrinsic Si, hence the term Si(Li), popularly pronounced "silly." Without the Li, commercial-purity Si would suffer electrical breakdown when the bias was applied to separate the electrons and holes. The Li is introduced either by diffusion under an applied voltage (hence the term "Li-drifted" detector) or, in a more controlled fashion, by ion implantation followed by a diffusion anneal.

While many electrons and holes are generated by an X-ray, they still constitute a very small charge pulse (about 10^{-16} C), and so a negative bias of ~0.5–1 keV is ap-

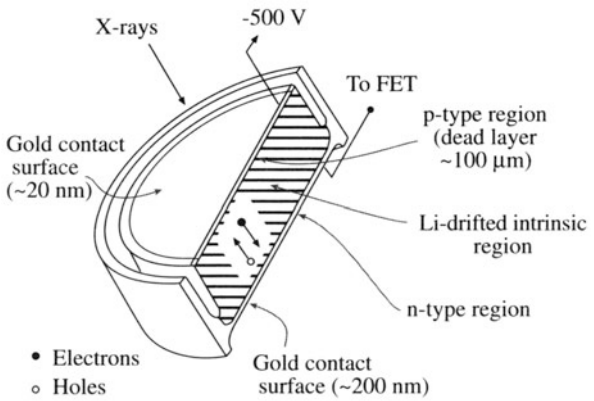


Figure 32.4. Cross section of a Si(Li) detector. The incoming X-rays generate electron–hole pairs in the intrinsic Si which are separated by an applied bias. A positive bias attracts the electrons to the rear ohmic contact after which the signal is amplified by an FET.

plied across the Si to ensure collection of most of the signal. We apply the bias between ohmic contacts, which are evaporated metal films such as Au or Ni, ~10–20 nm thick for the front face and ~200 nm at the back. This metal film also produces a p-type region at the front of the crystal; the back of the crystal is doped to produce n-type Si.

So the whole crystal is now a p-i-n device, with relatively shallow junctions less than 200 nm deep at either side of the central 3-mm intrinsic region.

When a reverse bias is applied to the crystal (i.e., a negative charge is placed on the p-type region at the front of the detector and a positive charge on the rear), the electrons and holes are separated and a charge pulse of electrons can be measured at the rear ohmic contact. Remember that the magnitude of this pulse is proportional to the energy of the X-ray that generated the electron–hole pairs. (We could equally well measure the whole pulse, but available low-noise FETs are n-channel devices, requiring electron collection.)

The p and n regions, at either end of the detector, are usually termed "dead layers." The traditional argument for use of this term is that the Li compensation is not completely effective so most of the electron–hole pairs generated in these end regions recombine, and contribute nothing to the charge pulse. Recently, however, Joy (1995) has shown that the dead layer can be explained if the diffusion length of the charge-carrying electrons exceeds the distance they travel under the drift field, in which case they are not "dead" but they will not be gathered at the surface electrode and contribute to the charge pulse. However, we'll still use the inaccurate term "dead layer" because it is so common. In practice it is the layer at the entrance surface of the detector that is most important since the X-rays must traverse it to be detected, and we will refer to this as *the* dead layer. This dead layer affects the spectrum, particularly when you are studying peaks from the low-Z elements (McCarthy 1995).

The p and n regions are called "dead layers" and the intrinsic region in between is referred to as the "active layer."

Why do we have to cool the detector with liquid N_2? 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 applied bias, destroying the intrinsic properties of the detector.
- The noise level in the FET would mask signals from low-energy X-rays.

For these reasons we cool the detector and the FET with liquid N_2, necessitating the characteristic dewar mentioned above (see Figure 32.3B). The FET gets to a temperature of about 140 K and the detector surface is at about 90 K.

Cooling the detector and the FET brings some undesirable consequences, which we have to accommodate. The minor irritations are that we have to regularly monitor and fill up the liquid-N_2 dewar. The more severe consequence of the cooling is that both hydrocarbons and ice from the microscope environment can condense on the cold detector surface, causing absorption of lower-energy X-rays. There are two obvious solutions to this problem. Either we can isolate the detector from the microscope vacuum, or we can remove hydrocarbons and water vapor from the microscope. The latter is the more desirable solution but the former is far easier. So, detectors are sealed in a prepumped tube with a "window" to allow X-rays through into the detector.

You have a choice of three different kinds of detector: those with a Be window, those with an ultrathin window, and those without a protective window.

Let's examine the pros and cons of each detector window; a good review has been given by Lund (1995).

32.4.B. Different Kinds of Windows

Beryllium window detectors use a thin sheet (nominally 7 μm) of beryllium which is transparent to most X-rays, and can withstand atmospheric pressure when the stage is vented to air. (In fact, 7 μm Be is expensive ($3M/pound!), rare, and slightly porous, so thicker sheet (> ~ 12 μm) is more commonly used.) Production of such thin Be sheet is a remarkable metallurgical achievement, but the window is still too thick to permit passage of all characteristic X-rays; any that have energy less than ~1 keV are strongly absorbed. Therefore, we cannot detect K_α X-rays from elements below about Na ($Z = 11$) in the periodic table. The Be window prevents microanalysis of the lighter elements such as B, C, N, and O, which are important in materials science (and also in other disciplines that use the AEM, such as the biological and geological sciences). Other factors, such as the low fluorescence yield and absorption within the specimen, make light-element X-ray microanalysis somewhat of a challenge, and EELS is often preferable.

Ultrathin window (UTW) detectors have windows that are less absorbent than Be; usually these are made from very thin (<100 nm) films of polymer, diamond, boron nitride, or silicon nitride, all of which are capable of withstanding atmospheric pressure while still transmitting 192-eV boron K_α X-rays. Early UTWs were very thin polymer membranes, such as parylene. Unfortunately, these were unable to withstand atmospheric pressure; they had to be withdrawn behind a vacuum isolation valve whenever a specimen was exchanged or the column was vented to air. Newer, composite Al/polymer UTWs or very thin diamond or BN windows, sometimes termed "atmospheric thin windows" (ATWs), are really the only sensible option. You should remember that different window materials absorb the light-element X-rays differently, so you need to know the characteristics of the window in the particular system you are using. For example, carbon-containing windows absorb N K_α X-rays very strongly.

Windowless detectors were first tried in the early 1970s, but microscope vacuums were relatively poor, resulting in rapid hydrocarbon and/or ice contamination of the detector surface. You should only use windowless detectors in a UHV AEM. Take great care to eliminate hydrocarbons from your specimen and keep the partial pressure of water vapor below ~10^{-8} Pa by efficient pumping. The best performance by a windowless system is the detection of Be (110 eV) K_α X-rays as shown in Figure 32.5, which is a remarkable feat of electronics technology.

You may recall that 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 !

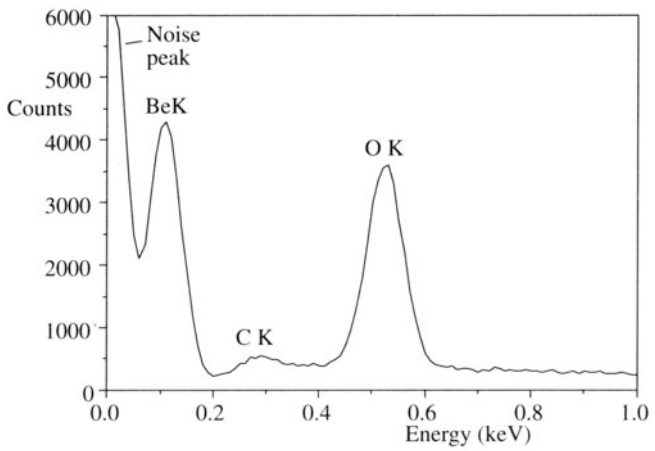


Figure 32.5. XEDS spectrum showing the detection of Be in an oxidized Be foil in an SEM at 10 keV. The Be K_α line is not quite resolved from the noise peak.

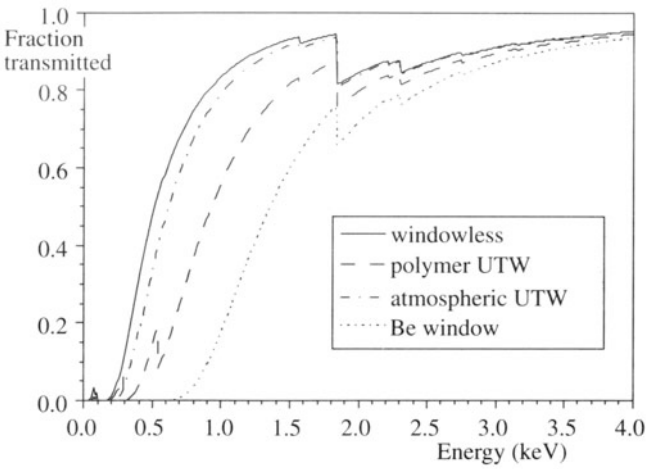


Figure 32.6. Low-energy efficiency calculated for a windowless detector, UTW detector (1 μm Mylar coated with 20 nm of Al), an ATW detector and a 13-μm Be window detector. Note that the efficiency is measured in terms of the fraction of X-rays transmitted by the window.

The relative performance of the various types of Si(Li) detector windows is summarized in Figure 32.6. Here we plot the detector efficiency as a function of the energy of the incoming X-ray. You can clearly see the rapid drop in efficiency at the low-energy end and the increased efficiency of UTW and windowless detectors. In fact, the Si(Li) detector absorbs X-rays with almost 100% efficiency over the energy range from about 2 to 20 keV, as shown in Figure 32.7. Within this energy range you will find X-rays from all the elements in the periodic table

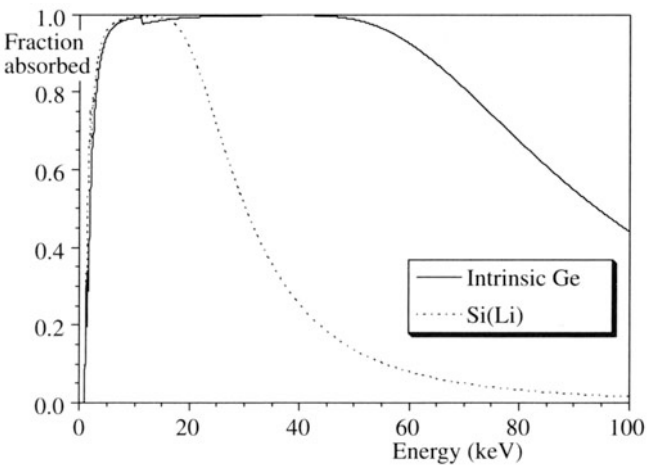


Figure 32.7. High-energy efficiency up to 100-keV X-ray energy calculated for Si(Li) and IG detectors, assuming a detector thickness of 3 mm in each case. Note the large effect of the Ge absorption edge at about 11 keV. In contrast to Figure 32.6, the efficiency in this case is measured by the fraction of X-rays absorbed within the detector.

above phosphorus. This uniform high efficiency is a major advantage of the XEDS detector.

32.4.C. Intrinsic Germanium Detectors

You can see in Figure 32.7 that Si(Li) detectors show a drop in efficiency above ~20 keV. This is because X-rays with such high energy can pass through the detector without depositing their energy by creating electron–hole pairs. This effect limits the use of Si(Li) detectors in intermediate voltage AEMs because at 300–400 keV we can generate K_α X-rays from all the high-atomic-number elements; Pb K_α X-rays at 75 keV are easily excited by 300-keV electrons. As we'll see in Chapter 35, there are certain advantages to using the K lines rather than the lower-energy L or M lines for quantification; with a Si(Li) detector the K lines from elements above silver (Z = 47) are barely detectable. The answer to this problem is to use a Ge detector, which more strongly absorbs high-energy X-rays (Sareen 1995).

We can manufacture Ge of higher purity than Si, and therefore Li compensation is not needed to produce a large intrinsic region; clearly this is a major advantage. Like the Si(Li) detector, the intrinsic Ge (IG) or high-purity Ge (HPGe) detector can have a Be window, a UTW/ATW, or no window; in any form it has some advantages over Si(Li). The detector is more robust and it can be warmed up repeatedly, which, as we'll see, sometimes solves certain problems.

> 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 mm thick, which results in 100% efficient detection of Pb K_α X-rays at ~ 75 keV. Figure 32.7 compares the efficiency of Si(Li) and IG detectors up to 100 keV and Figure 32.8 shows detection of Pb $K_{\alpha1}$ and $K_{\alpha2}$ lines generated from a lead glass specimen at 200 kV.

There is an even more fundamental advantage to IG detectors. Since it takes only ~2.9 eV of energy to create an electron–hole pair in Ge, compared with 3.8 eV in Si, a given X-ray produces more electron–hole pairs, and so the energy resolution and signal to noise are better. However, as you may have guessed, there are some difficulties in using IG detectors. The high-energy K lines, for which these detectors are ideally suited, have very small ionization cross sections when using 300–400 keV electrons, and so the spectral intensities are rather low, as you can see in Figure 32.8. A minor drawback is that IG detectors have to be

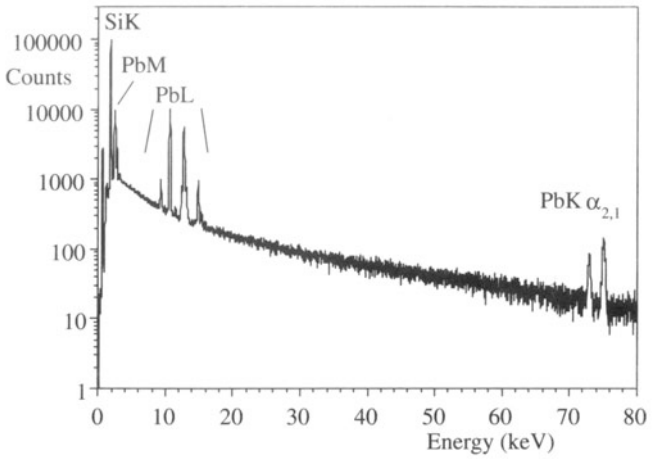


Figure 32.8. High-energy spectrum from lead silicate glass analyzed in a 200-kV AEM with an IG detector. Note the logarithmic scale for the counts which masks somewhat the very low intensity of the K lines compared with the L and M families. Note also that the $K_{\alpha1}$ and $K_{\alpha2}$ lines are clearly resolved.

cooled 25 K lower than Si(Li) to give the same leakage current, and they invariably need an Al-coated UTW since they are more sensitive to infrared radiation than Si(Li) detectors. This UTW reduces the maximum collection angle of IG detectors (see Section 33.2). Si(Li) detectors are easier to manufacture and are more reliable. They have a long history of dependable operation and a large number are already in use. IG detectors should eventually become more widely accepted as users install new systems.

IG detectors have been used since ~1970 by nuclear physicists to detect MeV radiation, but the first one was not installed on an AEM until 1986. One reason for this slow transfer of technology was that the high-energy K lines were very inefficiently excited in the lower voltage AEMs available at the time. Also, the low-energy performance of IG detectors was very poor in early detectors due mainly to a thick dead layer, which resulted in very non-Gaussian peak shapes. This problem has now been overcome and Gaussian characteristic peaks can be generated across the full energy range of the spectrum.

> In fact, UTW IG detectors are capable of detecting X-rays from boron to uranium, although the low-energy spectrum is still a little better in a Si(Li) system.

It is arguable that all intermediate voltage AEMs should be equipped with two detectors, an IG and a UTW Si(Li), to give the most efficient X-ray detection across the widest possible elemental range. Look ahead to Table 32.1 for a comparison of the two kinds of detector.

32.5. PULSE PROCESSING AND DEAD TIME

The electronic components attached to the detector convert the charge pulse created by the incoming X-ray into a voltage pulse, which can be stored in the appropriate energy channel of the MCA. The pulse-processing electronics must maintain good energy resolution across the spectrum without peak shift or distortion, even at high counting rates. To accomplish this, all the electronic components beyond the detector crystal must have low-noise characteristics and must employ some means of handling pulses that arrive in rapid succession. Currently, this whole process relies on analog pulse processing, but it is likely that, in the near future, many of the problems we'll now describe will be solved by digital techniques (Mott and Friel 1995).

Let's consider first of all what happens if a single isolated X-ray enters the detector and creates a pulse of electrons at the back of the Si(Li) crystal.

- The charge pulse enters the FET, which acts as a preamplifier and converts the charge into a voltage pulse.
- This voltage pulse is amplified several thousand times by a pulse processor, and shaped so that an analog-to-digital converter can recognize the pulse as coming from an X-ray of specific energy. (The XEDS doesn't do a very good job of accurate energy assessment.)
- The computer assigns it to the appropriate channel in the MCA display.

The accumulation of pulses or counts entering each energy channel at various rates produces a histogram of counts versus energy that is a digital representation of the X-ray spectrum. The MCA display offers multiples of 1024 channels in which to display the spectrum, and various energy ranges can be assigned to these channels. For example, 10, 20, or 40 keV full horizontal scales 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) detector is 20 keV, and in 2048 channels this gives you a display resolution of 10 eV per channel.

You should keep the display resolution at about 10 eV per channel. A smaller value ties up a lot of memory and often you can't display the whole spectrum at once. A larger value gives you only a few channels for each characteristic peak. For an IG detector, more channels (at least 4096) are needed to display the complete spectrum up to 80 or 100 keV, but the resolution of the MCA display is usually poorer, ~20 eV/channel.

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 the time (~10–50 μs) allowed for the pulse processor to evaluate the magnitude of the pulse. If you select a longer τ, the system is better able to assign an energy to the incoming pulse, but fewer counts can be processed in a given analysis time. You have a choice of τ given by the manufacturer:

- The shortest τ (typically a few μs) will allow you to process more counts per second (cps) but with a greater error in the assignment of a specific energy to the pulse, and so the energy resolution (see Section 32.6 below) will be poorer.
- A longer τ (up to about 50 μs) will give better resolution but the count rate will be lower.

> 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 (1995).

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. The details of the electronics can be found, e.g., in Goldstein *et al.* (1992). When the electronic circuitry detects the arrival of a pulse, it takes less than a microsecond before the detector is effectively switched off for the period of time called the *dead time* while the pulse processor analyzes that pulse. The dead time is clearly closely related to τ, which is so small that you should expect your detector system to process up to 10,000 cps quite easily. The dead time will increase as more X-rays try to enter the detector, which closes down more often. The dead time can be defined in several ways. Take the ratio of the output count rate (R_{out}) to the input count rate (R_{in}), which you can usually measure. Then we can say

$$\text{Dead time in } \% = \left(1 - \frac{R_{out}}{R_{in}}\right) \times 100\% \qquad [32.1a]$$

An alternative definition is

$$\text{Dead time in \%} = \frac{(\text{clock time} - \text{live time})}{\text{live time}} \times 100\% \quad [32.1b]$$

Put another way, this equation says that if you ask the computer to collect a spectrum for a "live time" of 100 s, this means that the detector must be "live" and receiving X-rays for this amount of time. If it is actually dead for 20 s while it is processing the X-rays, then the dead time will be 20%, and it will take 120 s of "clock time" to accumulate a spectrum. 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 old systems) mean that the detector is being swamped with X-rays and collection becomes increasingly inefficient, and it is better to turn down the beam current or move to a thinner area of the specimen to lower the count rate.

32.6. RESOLUTION OF THE DETECTOR

We can define the energy resolution R of the detector as follows

$$R^2 = P^2 + I^2 + X^2 \quad [32.2]$$

The term P is a measure of the quality of the associated electronics. It is 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 below). 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

$$I = 2.35\left(F\varepsilon E\right)^{\frac{1}{2}} \quad [32.3]$$

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 specific analysis conditions.

The IEEE standard for R is the FWHM of the Mn K_α peak, generated (off the microscope) by an Fe^{55} source which produces 10^3 cps with an 8-μs pulse-processor time constant.

Rather than using radioactive Fe^{55}, we recommend measuring the detector resolution on the AEM column! Now, since Mn is not a common sample to have lying around, you will find it useful to keep a thin Cr-film specimen to check the resolution when the detector is on the column. An evaporated Cr film about 100 nm thick supported on a carbon film and a Cu grid is ideal. Because Cr is next to Mn in the periodic table, the resolution of the K_α peak will be just a few eV less than that of Mn. The Cr is also stable with a resilient thin oxide film; it doesn't degrade in the electron beam. As we'll see later, these Cr films are very useful for other calibration checks and performance criteria (Zemyan and Williams 1994).

Many XEDS computer systems have an internal software routine which measures the resolution. Alternatively, you can gather a Mn or Cr peak, and select a window encompassing the peaks from the channels on both sides of the peak that contain half the maximum counts in the central channel, as shown in Figure 32.9.

Typically, Si(Li) detectors have a resolution of ~140 eV at Mn K_α with the best being 127 eV. The best reported IG resolution is 114 eV.

Because the value of ε is lower for Ge (2.9 eV) than for Si (3.8 eV), IG detectors have higher resolution than Si(Li). The resolution is also a function of the area of the detector, and the values given relate to the performance of 10-mm^2 detectors. The 30-mm^2 detectors which are usually installed on AEMs have resolutions about 5 eV worse than the figures just mentioned. However, you should also be aware that when R is measured on the microscope, there

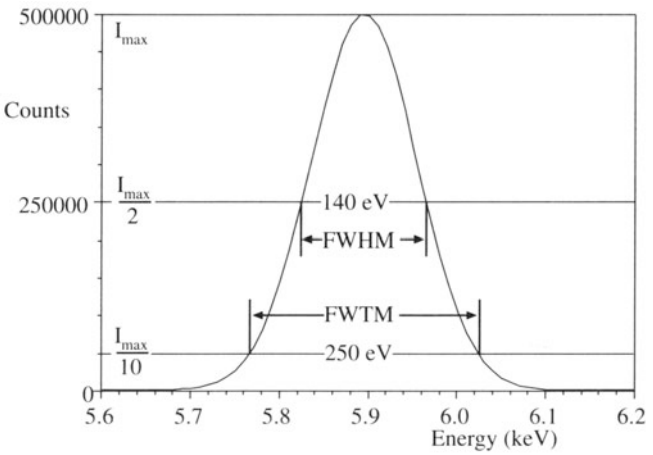


Figure 32.9. Measurement of the energy resolution of an XEDS detector by determining the number of channels that encompass the FWHM of the Mn K_α peak. The number of channels multiplied by the eV per channel gives the resolution, which typically should be about 130–140 eV. We can measure the FWTM also to give an indication of the degree of the incomplete charge collection which distorts the low-energy side of the peak. The FWTM should be ~1.82 times the FWHM.

may be a further degradation in resolution. It is rare to find a 30-mm^2 Si(Li) detector delivering a resolution much better than 140 eV on the AEM column, even though the quoted values are typically ~10 eV lower.

> Remember also that there is always a trade-off between resolution and count rate, unless digital pulse processing is used.

How close are XEDS detectors to their theoretical resolution limit? If we assume that there is no leakage and the electronics produced 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 XEDS detectors won't approach that of crystal spectrometers, which is 5 to 10 eV, although, because of the dependence of I on the X-ray energy, light-element K lines have widths <<100 eV.

32.7. WHAT YOU SHOULD KNOW ABOUT YOUR XEDS

There are several fundamental parameters of both Si(Li) and IG XEDS systems which you can specify, measure, and monitor to ensure that your system is performing acceptably. Many of these tests are standard procedures (e.g., see the XEDS laboratories in Lyman *et al.* 1990) and have been summarized by Zemyan and Williams (1995). In an SEM, which is relatively well behaved, Si(Li) detectors have been known to last ten years or more before requiring service or replacement. In contrast, an AEM (particularly the higher-voltage variety) is a hostile environment and the life of a detector is often less than three years. For this reason, most detectors are equipped with protective shutters (see Section 33.3).

> It is particularly important to monitor the detector performance on your AEM, in order that quantitative analyses you make at very different times may be compared in a valid manner.

You need to know both the operating specifications for your own system and how to measure them. We can break these specifications down into detector variables and signal-processing variables.

32.7.A. Detector Variables

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.

You can help to minimize the ice build-up by filtering the liquid N_2 before putting it into the dewar. Never re-cycle liquid N_2 into the dewar; use it elsewhere. If the nitrogen in the dewar is bubbling, you should consider warming up the detector, but do it *after consultation with the manufacturer, and without the applied bias*. (Think what happens to the Li otherwise.) Emptying out the dewar, filling it with hot water, and then drying it with a hair dryer will often solve the problem. However, after several such cycles, you may find that the detector resolution doesn't return to acceptable levels (the window seal may develop a leak due to the repeated thermal oscillations). When this happens, it is necessary to return the detector to the manufacturer to have it repaired.

Incomplete Charge Collection (ICC). Because of the inevitable presence of the dead layer, the X-ray peak will not be represented by a perfect Gaussian shape when displayed after processing. Usually, the peak will have a low-energy tail, because some energy will be deposited in the dead layer and will not create detectable electron–hole pairs. 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 32.9.

> The ideal value for FWTM/FWHM is 1.82 (Mn K_α or Cr K_α), but this value will be larger for X-ray peaks from the lighter elements, which are more strongly absorbed by the detector.

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 a detector has a large number of recombination sites arising, for example, through damage from a high flux of backscattered electrons. The crystal defects that act as recombination sites may be annealed out by warming the detector, as we just described. IG detectors used to show worse ICC effects than Si(Li) detectors, but this is no longer the case. Now, an IG detector should meet the same FWTM/FWHM ratio criterion as a Si(Li) detector. If the ratio is higher than 2 for the Cr peak, there is something seriously wrong with the detector and you should have it replaced.

Detector Contamination. Over a period of time, ice and/or hydrocarbons will eventually build up on the cold detector surface or on the window. If ice or hydrocarbon contamination does occur, it will reduce the efficiency with which we detect low-energy X-rays. While this is most

likely to happen for a windowless detector, Be and UTW systems also suffer the same problem 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 may develop over many months and you will not notice the degradation of the spectrum until differences in light-element quantification are apparent from the same specimen analyzed at different times. Therefore, you should regularly monitor the quality of the low-energy spectrum. The ratio of the NiK_α/NiL_α has been used (Michael 1995), but you can also use the CrK_α/CrL_α intensity ratio on the same evaporated film of pure Cr used for resolution, as long as there is no significant oxide film, since the O K_α line overlaps the Cr L_α line.

The NiK_α/NiL_α ratio will rise with time if contamination or ice is building up on the detector and selectively absorbing the lower-energy L line.

The K/L ratio will differ for different detector dead layers, for different UTWs or ATWs, and for different specimen thicknesses. So we can't define an accepted figure of merit. The best you can do is to measure the ratio immediately after installing a new (or repaired) detector and be aware that as the ratio increases, then any quantification involving similar low-energy X-ray lines will become increasingly unreliable. When the ratio increases to what you deem to be an unacceptably high value, then you must remove the ice/contamination. Automatic *in situ* heating devices which raise the detector temperature sufficiently to sublime the ice make this process routine. If your detector doesn't have such a device, then you should warm up the detector, as we described above.

In summary, you should measure and continually monitor changes in:

- *Your detector resolution* at the Mn or CrK_α line [typically 140–150 eV for Si(Li) and 120–130 eV for IG].
- *The ICC* defined by the FWTM/FWHM ratio of the Cr K_α line (ideally 1.82).
- *The ice build-up* reflected in the Ni (or Cr) K_α/L_α ratio.

If any of these figures of merit get significantly larger than the accepted values, then you should have your detector serviced by the manufacturer. Warming up the detector may cure some or all of the detector problems, but it could be an expensive procedure if you get it wrong. So only do it with the bias off and after consultation with the manufacturer.

So, 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* warm up the detector with the bias applied and without consulting the manufacturer.
- *Do not* use unfiltered or re-cycled liquid N_2.

32.7.B. Processing Variables

Other ways to monitor the performance of the XEDS system relate to the processing of the detector output signal. There are three things you have to 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 calibration should not change significantly from day to day, unless you change the range or the time constant. Electronic circuit stability has improved to a level where these three checks need only be done a few times a year (or if the detector has been repaired or a new detector installed).

Calibration of the energy display range. This process is quite simple; collect a spectrum from a material which generates a pair of K X-ray lines separated by about the width of the display range (e.g., Al-Cu for 0–10 keV). Some systems use an internal electronic strobe to define zero, and in this case you only need a specimen with one K line at high energy. Having gathered a spectrum, see if the computer markers are correctly positioned at the peak centroid (e.g., Al K_α at 1.54 keV and Cu K_α at 8.04 keV). If the peak and marker are more than 1 channel (10 eV) apart, then you should re-calibrate your display using the software routine supplied by the manufacturer.

Checking the dead-time correction circuit. If the dead-time correction circuitry is working properly, the pulse processor will give a linear increase in output counts as the input counts increase, for a fixed live time.

- Choose a specimen of a pure element, say our favorite Cr foil, which we know will give a strong K_α peak.
- Choose a live time, say 50 s, and a beam current to give a dead-time readout of about 10%.

- Measure the total Cr K_α counts that accumulate in 50 s.
- Then repeat the experiment with higher input count rates.

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 will remain at the chosen value. 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 32.10. But you will see that it will take increasingly longer clock times to attain the preset live time. 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.

Determination of the maximum output count rate. Again the procedure is simple:

- Gather a spectrum for a fixed clock time, say 10 s, with a given dead time, say 10%.
- Increase the dead time by increasing the beam current, C2 aperture size, or specimen thickness.
- See how many counts accumulate in the Cr K_α peak.

The number of counts should rise to a maximum and then drop off, because beyond a certain dead time, which depends on the system electronics, the detector will be closed more than it is open and so the counts in a given clock time will decrease. In Figure 32.11, this maximum is at about 60%, typical of modern systems, although in older XEDS units this peak can occur at as little as 30% dead time. You

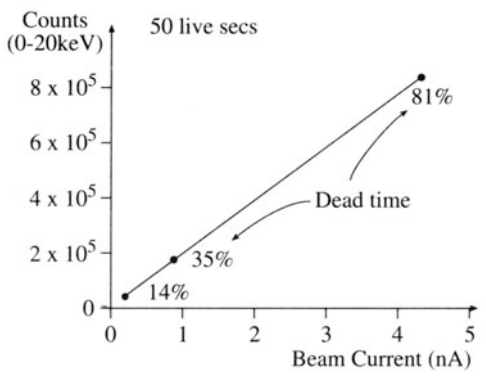


Figure 32.10. A plot of the output counts in a fixed live time as a function of increasing beam current showing good linear behavior over a range of dead times, implying that the dead-time correction circuitry is operating correctly.

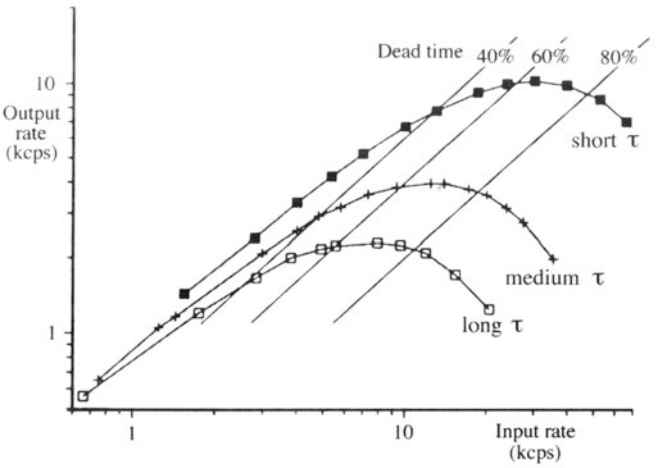


Figure 32.11. The output count rate in a given clock time as a function of dead time. The maximum processing efficiency is reached at about 60% dead time. It is very inefficient to use the system above the maximum output rate. Increasing the time constant results in fewer counts being processed and a drop in the output count rate.

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 32.11. 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 information from your specimen. As we've already said, it is generally better to have more counts than to have the best energy resolution, so select the shortest τ unless you have a peak overlap problem.

While it is rare that a good thin foil produces enough X-ray counts to overload modern detector electronics, there are situations (e.g., maximizing analytical sensitivity) when it's desirable to generate as many counts as possible. Under these circumstances, use of thicker specimens and high beam currents may produce too many counts for conventional analog processing systems. As shown in Figure 32.12A, digital processing permits a higher throughput over a continuous range of energy resolution than the fixed ranges available from specific (in this case, six) time constants. Even more dramatic, as shown in Figure 32.12B, megahertz-rate beam blanking, which deflects the beam off the specimen as soon as the XEDS detects an incoming pulse, permits a remarkable increase in throughput (Lyman *et al.* 1994).

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. In this case, just use a thicker specimen.

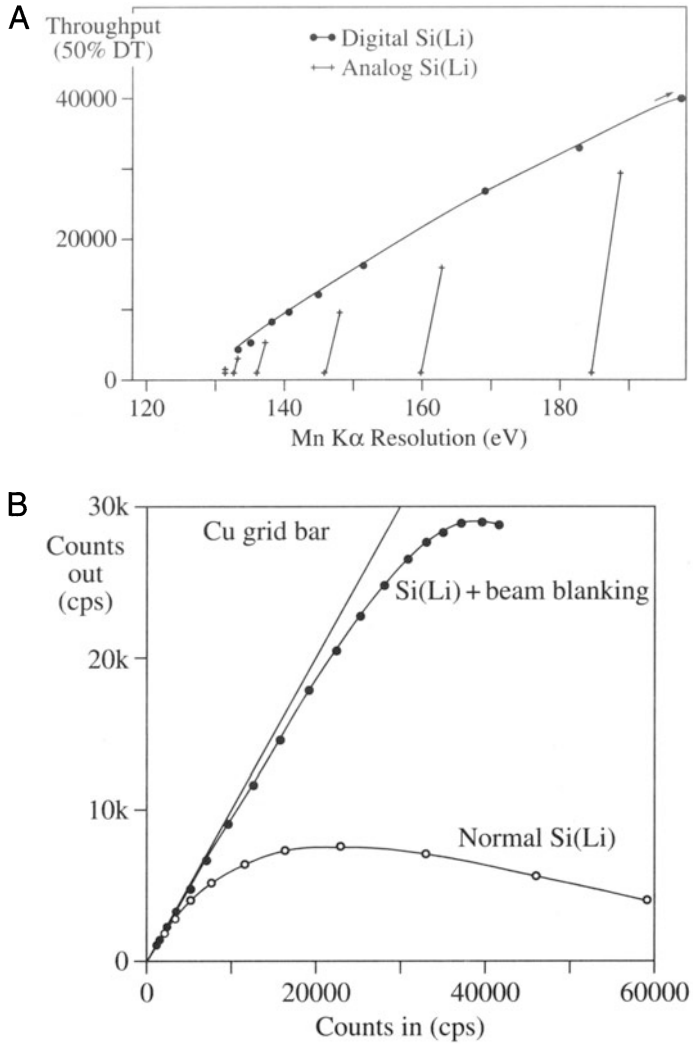


Figure 32.12. (A) Digital pulse processing gives a continuous range of X-ray throughput at 50% dead time, compared with a set of fixed throughput ranges for specific analog processing time constants. (B) Megahertz beam blanking results in a four times improvement in X-ray throughput compared to processing without beam blanking.

In summary you should occasionally:

- Check the energy calibration of the MCA display.
- Check the dead-time circuitry by the linearity of the output count rate versus beam current.
- Check the counts in a fixed clock time as a function of beam current to determine the maximum output count rate.

32.7.C. Artifacts Common to 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 (1995). We can separate the artifacts into two groups:

- Signal-detection artifacts. Examples are "escape" peaks and "internal fluorescence" peaks.
- Signal-processing artifacts. One example is "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 energy of (E-1.74) keV. An example is shown in Figure 32.13.

> 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) characteristic X-rays in the detector. Each of these can cause corresponding escape peaks, but the L_α has much less chance of escaping.

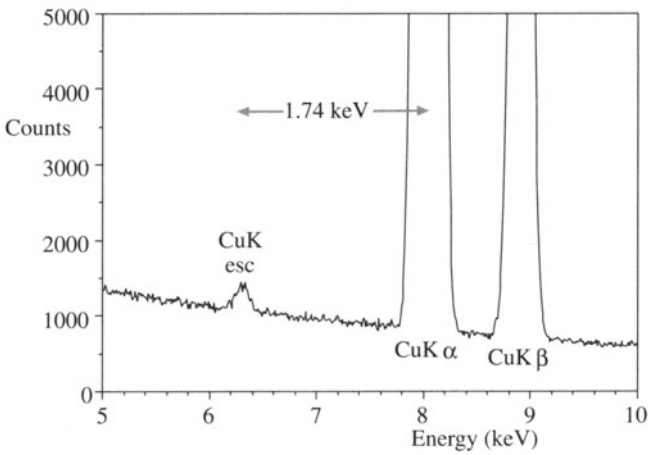


Figure 32.13. The escape peak in a spectrum from pure Cu, 1.74 keV below the Cu K_α peak. The intense K_α peak is truncated because it is ~50–100 times more intense than the escape peak.

The quantitative analysis software should be able to recognize any escape peak in the spectrum, 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.

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 registers a small peak in the spectrum. As detector design has improved and dead layers have decreased in thickness, the internal-fluorescence peak artifact has shrunk. However, it has not yet disappeared entirely.

A small Si K_α peak will occur in all spectra from Si(Li) detectors for long counting times.

Obviously you must be wary when looking for small amounts of Si in a specimen, because you'll always find it! Depending on the detector, particularly the dead-layer thickness, the Si signal has an intensity corresponding to about 0.1% to 1% of the composition of the specimen (see Figure 32.14), so again it is not a major problem. Similar effects are observed in IG detectors also. The Au absorption edge from the ohmic contact layer at the front end of the detector is sometimes detectable as a small disturbance in the bremsstrahlung intensity, around 1 keV, but the effect on microanalysis is negligible.

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 electronics are not fast enough. We can identify the conditions where this is likely to occur:

- The input count rate is high.
- The dead times are in excess of about 60%.
- There are major characteristic peaks in the spectrum.

The system simply cannot be perfect. Occasionally two photons will enter the detector at almost exactly the same time. The analyzer then registers an energy corresponding to the sum of the two photons. Since this event is most likely for the X-ray giving the major peak, a sum peak (sometimes called a coincidence or double peak) appears at twice the energy of the major peak, as shown in Figure 32.15.

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%.

In an AEM, you can't usually generate such high count rates unless your specimen is very thick. As always, 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. In the past the sum peak has led some researchers to report argon being present in aluminum specimens when it wasn't, and others to ignore argon which actually was present in ion-milled specimens! As detector electronics have improved, the sum peak has ceased to cause significant problems, except for intense low-energy peaks below ~1.25 keV, e.g., Mg K_α, where the residual noise in the electronic circuitry interferes with the pile-up rejection. So

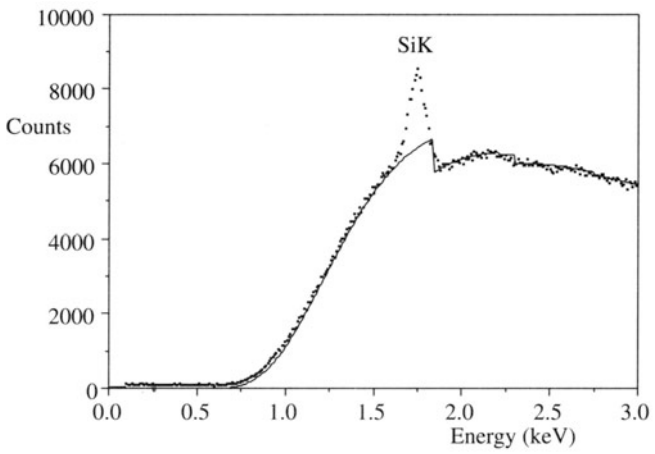


Figure 32.14. The Si internal fluorescence peak in a spectrum from pure C obtained with a Si(Li) detector. The ideal spectrum is fitted as a continuous line which exhibits the Si K absorption edge only.

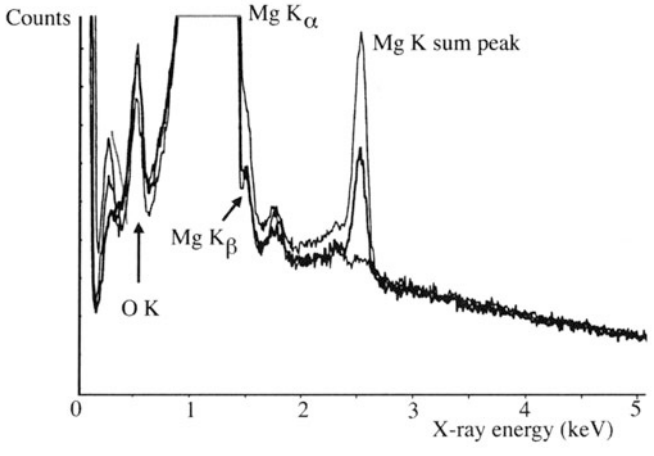


Figure 32.15. The Mg sum (coincidence) peak at various dead times; upper trace 70% dead time, middle trace 47%, lower trace 14%. The artifact is absent at 14% dead time.

if you're analyzing lighter elements 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 32.15.

Much of what we have just discussed 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 purchase the simulation software Desktop Spectrum Analyzer (DTSA) from NIST, which is listed in the recommended software in Section 1.5. This software permits realistic simulation of XEDS spectra in TEMs (and SEMs) and introduces you to all the aspects of spectral processing, artifacts, modeling, etc., that are discussed in the next four chapters.

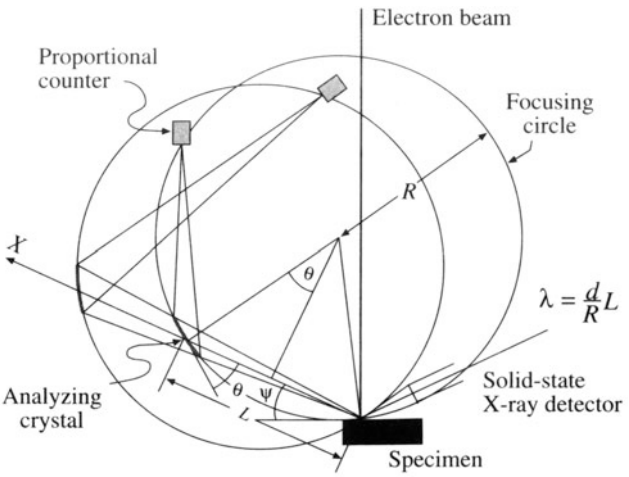


Figure 32.16. Schematic diagram of a WDS system showing how the specimen, crystal, and detector are constrained to move on the focusing circle, radius *R*, such that the specimen–crystal distance *L* is directly proportional to the X-ray wavelength.

32.8. WAVELENGTH-DISPERSIVE SPECTROMETERS

Before the invention of the XEDS, the wavelength-dispersive spectrometer (WDS), or crystal spectrometer, was widely used. The WDS uses one or more diffracting crystals of known interplanar spacing, as devised by W. H. Bragg in 1913. Bragg's Law ($n\lambda = 2d\sin\theta$), which we've already come across in Part II, also describes the dispersion of X-rays of a given wavelength λ through different scattering angles, 2θ. We accomplish this dispersion by placing a single crystal of known interplanar spacing (d) at the center of a focusing circle which has the X-ray source (the specimen) and the X-ray detector on the circumference, as shown in Figure 32.16. The detector is usually a gas-flow proportional counter, but there's no reason why it couldn't be a Si(Li) or Ge semiconductor detector. In fact, these detectors may see more use as the need for better vacuums increases.

The mechanical motions of the crystal and detector are coupled such that the detector always makes an angle θ with the crystal surface while it moves an angular amount 2θ as the crystal rotates through θ. By scanning the spectrometer, a limited range of X-ray wavelengths of about the same dimension as the d-value of the analyzing crystal can be detected. For example, diffraction from the (200) planes of a LiF crystal covers an energy range of 3.5–12.5 keV (0.35–0.1 nm) for a scanning range of θ = 15–65°. To detect X-rays outside this energy range, another crystal of different d-value must be employed.

As we'll discuss in Chapter 35, the forerunner of the AEM was the electron microscope microanalyzer (EMMA), which used a WDS. However, the WDS was a large and inefficient addition to the microscope, and never attained general acceptance by transmission microscopists.

There are two major drawbacks to the WDS compared with the XEDS:

- The crystal has to be moved to a precise angle where it collects only a tiny fraction of the total number of X-rays coming from the specimen, whereas the XEDS detector can be placed almost anywhere in the TEM stage above the specimen and subtends a relatively large solid angle at the specimen.
- The WDS collects a single wavelength at a given time while the XEDS detects X-rays of a large range of energies. WDS is a serial collector; XEDS is effectively a parallel collector.

The geometrical advantage in the collection efficiency of XEDS, combined with the ability to detect X-rays simultaneously over a wide energy range without the mechanical motion required of the WDS, accounts for the present dominance of XEDS systems in all types of electron microscopes. However, there are several advantages of WDS over XEDS:

- Better energy resolution (5–10 eV) to unravel the peak overlaps that plague XEDS (see Section 34.3).
- Better peak-to-background capability to detect smaller amounts of elements.
- Better detection of light elements (minimum Z = 4, Be) by careful choice of crystal, rather than solely through a dependence upon electronics as in the XEDS.

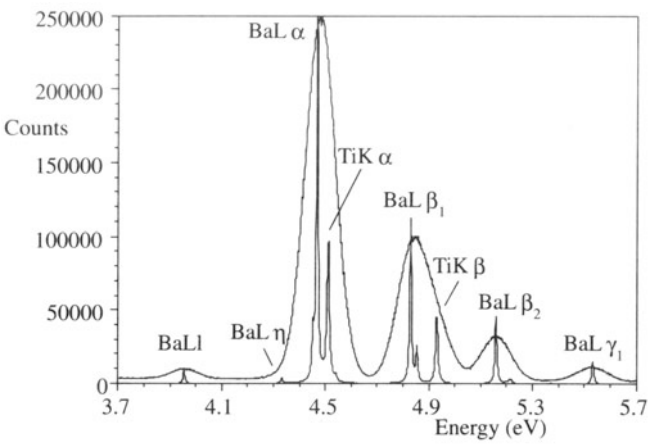


Figure 32.17. A WDS spectrum from $BaTiO_3$, but plotted against energy rather than wavelength. WDS easily resolves the Ba L_α/Ti K_α overlap, which is impossible with an EDS as shown in the overlapping spectrum. The improved resolution of WDS (~8 eV) is obvious.

- 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.

A typical WDS spectrum from $BaTiO_3$ is shown in Figure 32.17. For comparison, an XEDS spectrum is shown superimposed; the improved resolution and *P/B* for WDS are obvious. Because of these advantages, the WDS is often the spectrometer of choice in the X-ray fluorescence spectrometer (XRF), which has a spatial resolution of a few millimeters, and in the electron probe microanalyzer (EPMA) with a spatial resolution of a few micrometers. The advantages of the WDS may make it an attractive complement to the XEDS in future AEMs. However, WDS systems have not been applied to AEMs because of their low X-ray collection efficiency compared to the XEDS, and we discussed the drawbacks of WDS at the start of the chapter. Only when a WDS is designed that is compact enough to be placed inside an AEM specimen stage will it be possible to realize these advantages, and even then the output count rate from thin specimens might be too low unless the AEM has an FEG (Goldstein *et al.* 1989). Spence and Lund (1991) show preliminary results from a WDS in an AEM giving 40-eV resolution in a study of coherent bremsstrahlung (see Section 33.4.C).

CHAPTER SUMMARY

The XEDS is the only X-ray spectrometer currently used in TEMs. It is remarkably compact, efficient, and sensitive. A combination of Si(Li) and Ge detectors can detect K_α lines from all the elements, from Be to U. However, the XEDS has limits in terms of its need for cooling, its poor energy resolution, and the many artifacts that appear in the spectra. The XEDS is simple to run and maintain if you 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 weekly at the Mn or Cr K_α line (typically, 130–140 eV for Si(Li) and 120–130 eV for IG).
- Measure the ICC defined by the FWTM/FWHM ratio of the Cr K_α line (ideally 1.82) on a monthly basis.
- Monitor any ice build-up via the Ni (or Cr) K_α/L_α ratio on a weekly basis.
- Check the calibration of the energy range of your MCA display every few months.
- Check the dead-time correction circuitry by the linearity of the output count rate versus beam current, every six months.
- Check the counts in a fixed clock time as a function of beam current, to determine the maximum output count rate, every six months.
- Be aware of artifacts in *all* your spectra.

For the sake of completeness, Table 32.1 below shows you the relative merits of the various detectors that we have discussed in this chapter.

Table 32.1. Comparison of X-ray Spectrometers

Characteristic energy resolution	Intrinsic Ge	Lithium-drifted Si	WDS
Typical value	140 eV	150 eV	10 eV
Best value	114 eV	127 eV	5 eV
Energy required to form electron–hole pairs (77 K)	2.9 eV	3.8 eV	n.a.
Band-gap energy (indirect)	0.67 eV	1.1 eV	n.a.
Cooling required	LN_2 or thermoelectric	LN_2 or thermoelectric	none
Typical detector active area	10–30 mm^2	10–30 mm^2	n.a.
Typical output counting rates	5–10,000 cps	5–10,000 cps	50,000 cps
Time to collect full spectrum	1 min	1 min	30 min
Collection angle	0.03–0.20 sr	0.03–0.30 sr	10^{-4}–10^{-3} sr
Take-off angle	0°/20°/72°	0°/20°/72°	40°–60°
Artifacts	Escape peaks Sum peaks Ge K/L internal fluorescence peaks	Escape peaks Sum peaks Si K internal fluorescence peaks	High-order lines

REFERENCES

General References

Goldstein, J.I., Newbury, D.E., Echlin, P., Joy, D.C., Romig, A.D. Jr., Lyman, C.E., Fiori, C.E., and Lifshin, E. (1992) *Scanning Electron Microscopy and X-ray Microanalysis,* 2nd edition, Plenum Press, New York.

Heinrich, K.F.J., Newbury, D.E., Myklebust, R.L., and Fiori, C.E., Eds. (1981) *Energy Dispersive X-ray Spectrometry,* NBS Special Publication 604, U.S. Department of Commerce, Washington, DC.

Russ, J.C. (1984) *Fundamentals of Energy Dispersive X-ray Analysis,* Butterworths, Boston, Massachusetts.

Williams, D.B., Goldstein, J.I., and Newbury, D.E., Eds. (1995) *X-Ray Spectrometry in Electron Beam Instruments,* Plenum Press, New York.

Specific References

Goldstein, J.I., Lyman, C.E., and Williams, D.B. (1989) *Ultramicroscopy* **28**, 162.

Joy, D.C. (1995) in *X-Ray Spectrometry in Electron Beam Instruments* (Eds. D.B. Williams, J.I. Goldstein, and D.E. Newbury), p. 53, Plenum Press, New York.

Lund, M.W. (1995) *ibid.,* p. 21.

Lyman, C.E., Newbury, D.E., Goldstein, J.I., Williams, D.B., Romig, A.D. Jr., Armstrong, J.T., Echlin, P.E., Fiori, C.E., Joy, D.C., Lifshin, E., and Peters, K.R. (1990) *Scanning Electron Microscopy, X-Ray Microanalysis and Analytical Electron Microscopy; A Laboratory Workbook,* Plenum Press, New York.

Lyman, C.E., Goldstein, J.I., Williams, D.B., Ackland, D.W., von Harrach, S., Nicholls, A.W., and Statham, P.J. (1994) *J. Microsc.* **176**, 85.

McCarthy, J.J. (1995) in *X-Ray Spectrometry in Electron Beam Instruments* (Eds. D.B. Williams, J.I. Goldstein, and D.E. Newbury) p. 67, Plenum Press, New York.

Michael, J.R. (1995) *ibid.,* p. 83.

Mott, R.B. and Friel, J.J. (1995) *ibid.,* p. 127.

Newbury, D.E. (1995) *ibid.,* p. 167.

Sareen, R.A. (1995) *ibid.,* p. 33.

Spence, J.C.H. and Lund, M. (1991) *Phys. Rev.* **B44**, 7054.

Statham, P.J. (1995) in *X-Ray Spectrometry in Electron Beam Instruments* (Eds. D.B. Williams, J.I. Goldstein, and D.E. Newbury), p. 101, Plenum Press, New York.

Zemyan, S.M. and Williams, D.B. (1994) *J. Microsc.* **174**, 1.

Zemyan, S.M. and Williams, D.B. (1995) *X-Ray Spectrometry in Electron Beam Instruments* (Eds. D.B. Williams, J.I. Goldstein, and D.E. Newbury), p. 203, Plenum Press, New York.

The XEDS–TEM Interface

33

CHAPTER PREVIEW

In principle, all you have to do to create an AEM is to hang an XEDS detector on the side of a TEM. However, in practice it isn't always that simple because the TEM is designed primarily as an imaging tool, and microanalysis requires different design criteria. The AEM illumination system and specimen stage are rich sources of 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 record comes from the area you chose and can ultimately be converted to quantitative elemental information. You therefore need to understand the problems associated with the XEDS–TEM interface and find ways to maximize the useful data. We describe several tests you should perform to ensure that the XEDS–TEM interface is optimized.

The XEDS–TEM Interface

33

33.1. THE REQUIREMENTS

The interfacing of the XEDS to the TEM is not something over which you have too much control. It has already been carried out by the manufacturers, so you purchase an AEM system of TEM and XEDS. Often you won't be able to change anything about the system, but nevertheless you should be aware of the important factors that characterize the interface between the XEDS and the TEM column and what effect these factors will have on your microanalysis experiments. Knowing these factors may help you to select the best AEM to use.

The stage of a TEM is a harsh environment. An intense beam of high-energy electrons bombards a specimen which interacts with and scatters the electrons. The specimen *and any other part of the microscope* that is hit by electrons emit both characteristic and bremsstrahlung X-rays which have energies up to that of the electron beam. X-rays of such energy can penetrate long distances into 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 33.1, it is not possible to prevent radiation from the microscope stage and other areas of the specimen from entering the detector. The X-rays from the microscope itself we will call "system X-rays." All the X-rays arising from regions of the specimen other than that chosen for analysis, we term "spurious X-rays." Your job, as an analyst, is to learn how to identify the presence of these undesirable X-rays and to minimize their effect on your microanalysis.

33.2. THE COLLIMATOR

As you can see from Figure 33.1, the XEDS has a collimator in front of the detector crystal. This collimator is the front line of defense against the entry of undesired radiation from the stage region of the microscope.

> The collimator also defines the (desired) collection angle of the detector (see below) and the average take-off angle of X-rays entering the detector.

Ideally, the collimator should be 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 will minimize the production of X-rays from any backscattered electrons that happen to spiral into the collimator and the high-Z material will absorb any high-energy bremsstrahlung radiation. The inside of the collimator should also have baffles to prevent any backscattered electrons from generating X-rays that then penetrate the detector. Such a design is shown in Figure 33.2 (Nicholson *et al.* 1982), and aspects of this design are available in some commercial systems. They are strongly recommended.

33.2.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 33.1 and is defined as

$$\Omega = \frac{A \cos \delta}{S^2} \qquad [33.1]$$

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.

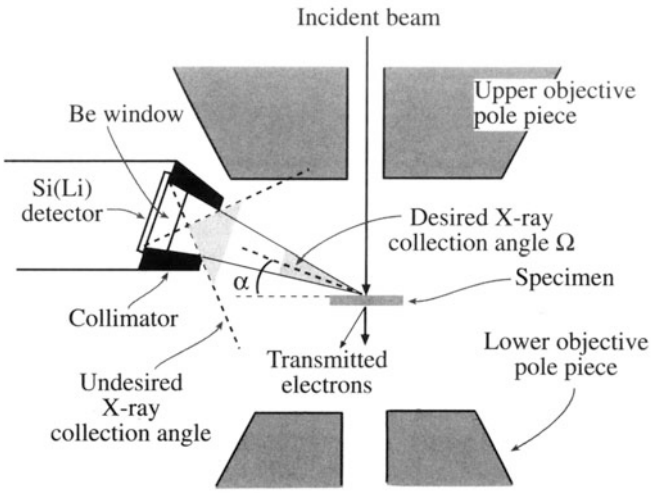


Figure 33.1. The interface between the XEDS and the AEM stage, showing how the detector can "see" undesired X-rays from regions other than the beam–specimen interaction volume. The desired collection angle Ω and take-off angle α are also shown.

The value of Ω is the most important parameter in determining the quality of your X-ray microanalysis.

In most cases, particularly with thermionic-source instruments, it is the low X-ray counts that limit the accuracy of the experiment. Now in commercial AEMs, S varies from about 10–30 mm, and as a result values of Ω 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 (usually etched Si) which reduces the collection angle by ~20%. So, at best, Ω is a small fraction of the total solid angle of characteristic X-ray generation which is, of course, 4π sr. These values of Ω are calculated from the dimensions of the stage and the collimator. Unfortunately, there is no way 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 Cr film, and a known beam current. A figure of merit for this parameter is given in terms of the X-ray counts per second detected from the standard, when a given beam current is used with 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 an energy of 100 keV, it is about 13,000 (Zemyan and Williams 1994). 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. So a compromise has to be made in the design of the stage of the AEM to ensure both adequate current in the beam and the best possible collection angle. If we move the detector too close to the specimen, it will eventually suffer direct bombardment by backscattered electrons. The other alternative we have, looking at equation 33.1, is to increase A. However, increasing the detector area results in a small decrease in energy resolution. As we noted already, there are certainly situations where increased count rate is to be preferred at the expense of a small decrease in resolution; while 50-mm^2 detectors are available, they are rarely used.

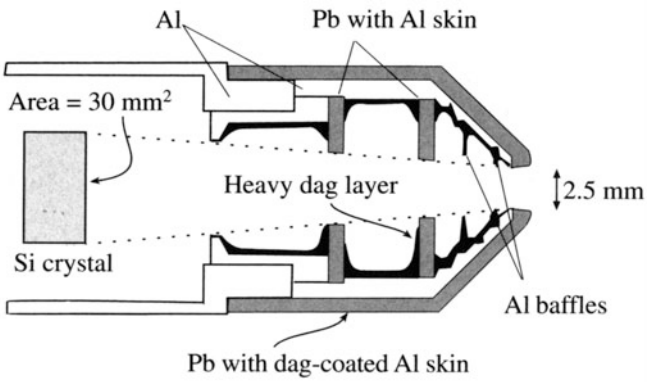


Figure 33.2. Combination high-Z (Pb) and low-Z (Al/carbon paint) collimator design to prevent high-energy bremsstrahlung from penetrating the collimator walls. Baffles are incorporated to minimize BSE entry into the detector.

33.2.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 33.1. Sometimes, it is also defined as the angle between the transmitted beam and the line to the detector, which is simply (90° + α). Traditionally in the EPMA, the value of α is kept high to minimize the absorption of X-rays as they travel through bulk specimens. Unfortunately, if we maximize α the price we pay is lowering Ω. Because the detector then has to be positioned above the upper objective polepiece, it will "look" through a hole in the polepiece. Therefore, it will be much further 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.

We would like to optimize the take-off angle and maximize the count rate.

In AEMs where the detector "looks" through the objective polepiece giving 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 value 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 bulk EPMA is that absorption can usually be neglected. However, if absorption is a problem in your specimen then you can reduce the path length of X-rays traveling through your specimen by tilting it toward the detector, thus increasing α (see Section 35.5). Tilting may increase spurious effects, which we'll discuss later, and also generally lowers the P/B (peak-to-background) ratio in the spectrum; so tilting is always a compromise.

33.2.C. Orientation of the Detector to the Specimen

(a) Is the detector pointing on axis? The detector is inserted into a position 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. We have to assume that the XEDS and the TEM manufacturers have collaborated closely in the design of the collimator and stage. To find out if your system is well aligned, you can make a low-magnification X-ray map from a homogeneous specimen such as a thin Cr film (Nicholson and Craven 1993). If the detector is not pointing on axis, the map will show an asymmetric intensity distribution. Alternatively, if you cannot map at low enough magnification, simply see how the Cr K_α intensity varies from area to area on the foil with the specimen 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 up or down away from the eucentric plane using the *z*-control. Again, the maximum intensity should be recorded at the eucentric plane. If the intensity is asymmetric, then the detector or the collimator is not well aligned and some of your precious X-ray intensity is being shadowed from the detector, probably by the collimator; so you need to consult the manufacturer.

(b) Where is the detector with respect to the image? When you look at a TEM or STEM image to position the beam on an area for microanalysis, it is best if the detector is "looking" toward the thin region of the specimen rather than toward the thicker region, as shown in Figure 33.3A. This position minimizes the X-ray path length through the specimen and helps to ensure that any absorption is minimized. In TEM mode, the detector orientation with respect to the BF image on the screen will vary with magnification if the BF image rotates when changing magnification. In a STEM BF image there is no rotation, so the relative orientation of the detector to the image will be 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. Under these circumstances, if you push in the end of the holder while the specimen is in the column, the image will move in the +*x*-direction. Then you can determine by simple geometry 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 is sometimes rotated 180° with respect to

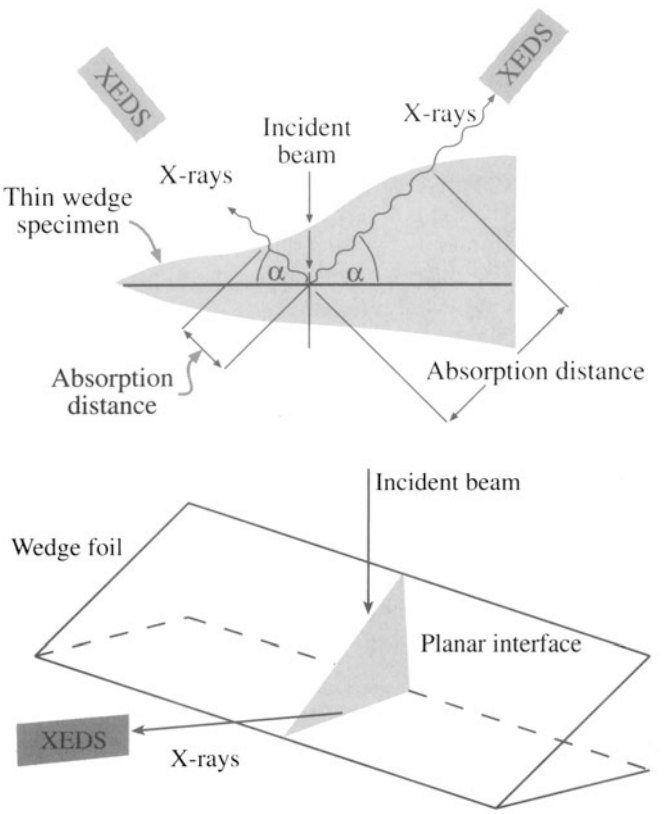


Figure 33.3. (A) The position of the XEDS detector relative to a wedge-shaped thin foil results in different X-ray path lengths. The shortest path length with the detector "looking" at the thinnest region of the foil is best. (B) The preferred orientation of the XEDS detector when analyzing a planar defect: the interface plane is parallel to the detector axis and the incident beam direction.

the TEM image, so you have to take this into account (just check TEM and STEM images of a recognizable area of your specimen).

If you're doing microanalysis across a planar interface, which is a common AEM application, then you will also need to orient your specimen such that the interface is parallel to the detector axis and the beam. A tilt-rotation holder would be ideal for this, but a low-background version is not available, so you may need to reposition your foil manually until the interface is in the right orientation (see Figure 33.3B).

> The XEDS detector must be "looking" at the thin edge of your specimen and aligned with any planar interface you are studying.

33.3. 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 microscopes. These situations usually occur when you place bulk material under the beam. This can happen if you leave the objective diaphragm inserted, if you go to low magnification and expose the bulk regions of a disk to the beam, or if you are traversing around a thin specimen supported on a grid and a grid bar is hit by the beam. To avoid these problems there are various kinds of shutter systems built into XEDS detectors which automatically protect the detector crystal if the instrument is switched to low magnification or if the pulse processor detects too high a flux of radiation.

> 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.

If you don't have a shutter, then you can physically retract the detector, which lowers Ω (if it is retracted along a line of sight to the specimen) or removes 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 and also you may reposition your detector slightly differently each time, unless the system is designed so that you can push the detector up to a fixed stop, thus insuring a constant collection and take-off angle. A shutter is highly recommended!

33.4. SYSTEM X-RAYS AND SPURIOUS X-RAYS

In an ideal AEM, all spectra would be characteristic only of the chosen region of your specimen. The analysis of bulk specimens in the EPMA approaches this ideal, but in the AEM several factors combine to introduce false information which can introduce serious errors into both qualitative and quantitative microanalysis unless you are aware of the dangers and take appropriate precautions to identify and minimize the problems. The factors unique to the AEM that are responsible for these problems are:

- The high accelerating voltages which generate intense doses of stray X-rays and electrons in the illumination system.
- The scattering of high-energy electrons and X-rays by the thin specimen in the limited confines of the TEM stage.

Most AEMs are now designed to minimize some of these problems. Nevertheless, when identification and quantification of small (<~5%) elemental amounts are required, you have to be wary of system and spurious X-rays, which we will now discuss in some detail. These artifacts, which are in addition to the XEDS system artifacts, can be responsible for large errors in quantification or, in the extreme, may make your microanalysis impossible.

33.4.A. Pre-Specimen Effects

Ideally, the electron beam should be the sole source of radiation incident on your specimen, and the X-rays then originate in a well-defined interaction volume. In practice, the illumination system can produce high-energy bremsstrahlung X-rays and uncollimated electrons, which may strike the specimen anywhere, producing spurious X-rays indistinguishable from those generated in the region of interest. 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 several review papers (e.g., Williams and Goldstein 1981, Allard and Blake 1982) which describe in detail how to identify and minimize these artifacts from the illumination system, so we will just describe the precautions necessary to ensure 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 pole-pieces, you must take extra care when using intermediate-voltage instruments.

The standard way to detect stray radiation from the illumination system is to position the focused electron beam down a hole in your specimen and see if you can detect an X-ray spectrum characteristic of the specimen.

> Such a spectrum, sometimes termed a "hole count," is *invariably* obtained in all AEMs if you count for long enough.

If the hole count contains more than a few percent of the characteristic intensity obtained from a thin area of your specimen under similar conditions, then we say the illumination system is not "clean."

You can easily determine whether stray electrons or X-rays are the problem, as illustrated in Figure 33.4 (Goldstein and Williams 1978). Almost invariably, the problem is caused by stray X-rays penetrating the C2 diaphragm.

> 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.

These diaphragms should be a standard fixture in all AEMs (check with your manufacturer), but they are expensive, and you cannot flame-clean them in the usual way. When the thick diaphragms do contaminate, you should discard them, otherwise the contamination itself will become a source of X-rays and also deviate beam electrons by charging. 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. Other AEMs use virtual beam-defining apertures, keeping the diaphragm itself well away from the specimen, and this is ideal. Another good way to minimize the effects of the bremsstrahlung is to use an evaporated film or window-polished flake on a Be grid, rather than a self-supporting disk. If the specimen is thinner than the path length for fluorescence, spurious X-rays will not be generated. Of course, such thin specimens may not be possible to prepare, or 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.

For a quantitative, reproducible measure of the hole count, you should use a uniform thin specimen such as the Cr 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 molybdenum or gold washer 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.

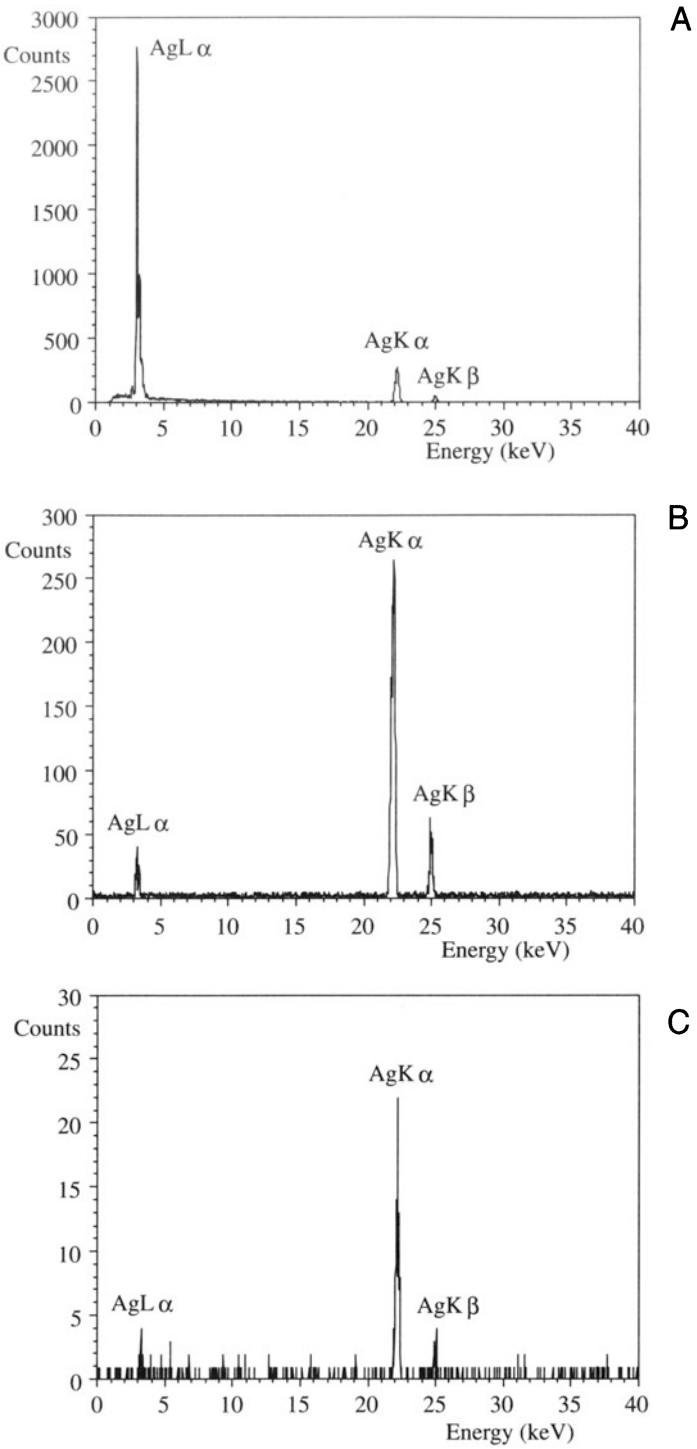
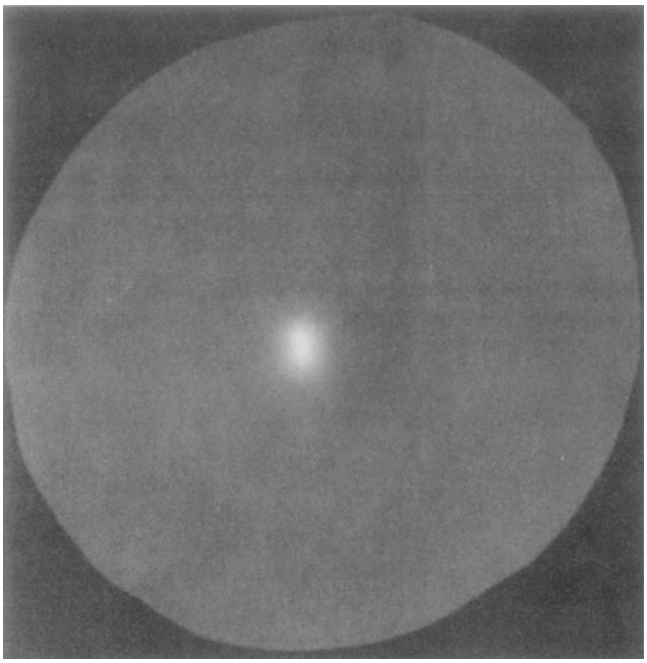


Figure 33.4. The "hole count." (A) A Ag self-supporting disk produces an electron-characteristic (high L/K ratio) spectrum when struck by the primary beam. (B) Without a thick C2 diaphragm, an intense Ag spectrum is also detected when the beam is placed down a hole in the specimen. This spectrum has a low L/K ratio, which indicates high-energy bremsstrahlung fluorescence of the K lines. Approximately 50% of the K_α line in (A) was due to these spurious X-rays. (C) Use of a thick Pt C2 diaphragm reduces the intensity of the hole count substantially. The K_α intensity in (C) is about 30 times less than in (B). (Note the scale change.)

A



B

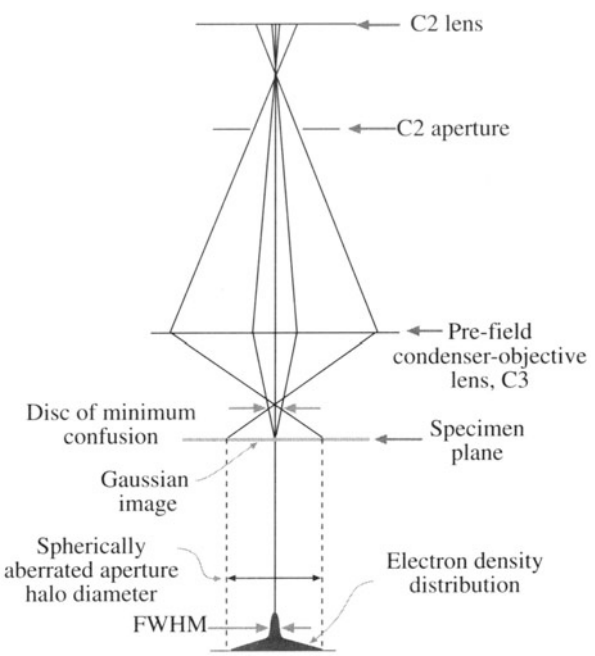


Figure 33.5. (A) The shadow of the diaphragm defines the extent of the halo which excites X-rays remote from the chosen microanalysis region. (B) Ray diagram showing how the STEM probe obtained with a large C2 aperture has a broad halo of electrons surrounding the intense Gaussian central portion. Such a halo is the major source of uncollimated electrons and arises due to spherical aberration in the C3 lens.

As a rule of thumb, the ratio of Mo K_α or Au L_α intensity detected (when the beam is down the hole) to the Cr K_α intensity obtained with the beam on the specimen should be less than 1%.

Under these conditions 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 detail on this test see Lyman and Ackland (1991). 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 from your specimen and subtract it from your experimental spectrum.

In addition to stray X-rays, it is possible that all the electrons are not confined to the beam. If your microscope has a non-beam-defining spray aperture below the C2 aperture, 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 C3 lens, as shown in Figure 33.5 from Cliff and Kenway (1982). The best way to minimize this effect is to image the beam on the TEM screen under the conditions that you will use for microanalysis and select the best C2 aperture, as we discussed earlier in Chapter 6. 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.

In summary:

- Always operate with clean, top-hat C2 diaphragms.
- Use very thin flake specimens, if possible.
- Always image the electron beam on the TEM screen prior to microanalysis, to ensure that it is well collimated by the C2 aperture.

Under these circumstances, the primary X-ray source will be the region where you put the beam.

33.4.B. Post-Specimen Scatter

After the electrons interact with the specimen, they are scattered elastically or inelastically. It is fortunate for us that the intensity of elastic and inelastic scatter from a thin specimen is greatest in the forward direction. Most of the forward-scattered electrons are gathered by the field of the lower objective polepiece and proceed into the imaging system of the microscope. Unfortunately, some electrons are scattered through high enough angles that they strike some part of the specimen holder or the objective lens polepiece or other material in the stage of the microscope.

This effect will be severely exacerbated if the objective diaphragm is not removed during microanalysis.

It is instructive to try this experiment (just once!) to see the enormous increase in spurious and system X-rays that result. Usually, the X-ray flux is so great that the pulse-processing electronics are saturated and the dead time reaches 100%, and the automatic shutter will activate. Even when you remove the objective diaphragm, scattered electrons may create X-rays characteristic of the materials used to construct the holder, the polepiece (mainly copper and iron), and the collimator, and these X-rays could be picked up by the XEDS detector. Furthermore, the scattered electrons may travel directly into the XEDS detector, generating electron–hole pairs, or they may hit your specimen at some point remote from the area of interest and produce specimen-characteristic spurious X-rays. 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 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 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 Cr film) on a copper grid. When you place 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 copper peak arising from the grid as a result of interactions with electrons or X-rays scattered by the specimen. An example of this effect is shown in Figure 33.7. You can remove the presence of the copper peak by using a beryllium grid, since Be K_α X-rays are not routinely detectable. However, using Be grids merely removes the observable effect, not the cause. Therefore, 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.

Remember that 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!

To minimize the effects of the scattered radiation, you should keep your specimen close to zero tilt (i.e., normal to the beam). Experimentally, it seems that if you tilt less than about 10°, 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.

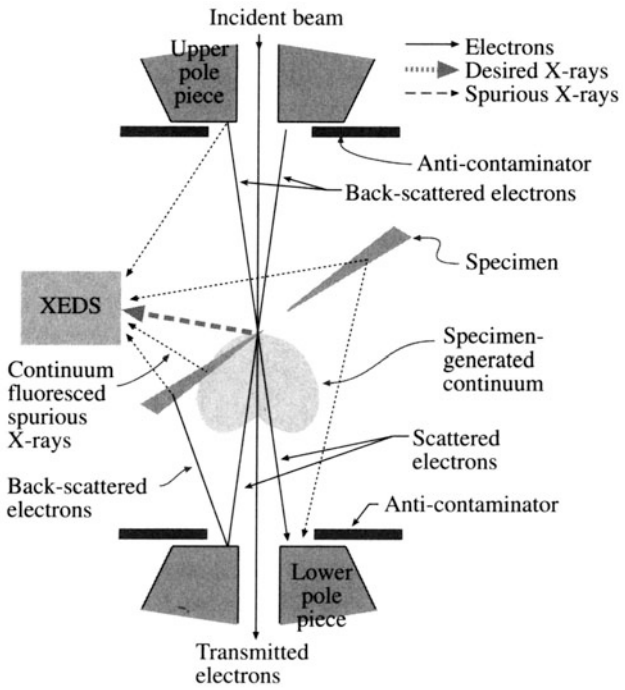


Figure 33.6. Sources of system and spurious X-rays generated when the primary beam is scattered by a tilted, wedge specimen. Note the BSEs which excite X-rays in the stage and elsewhere in the specimen and the specimen-generated bremsstrahlung which fluoresces X-rays from the specimen itself, but well away from the region chosen for analysis.

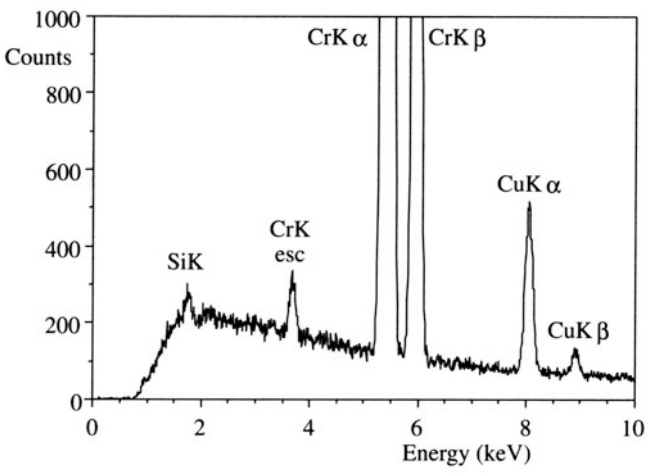


Figure 33.7. Cu peaks in a spectrum from a thin Cr film on a Cu grid. Although the beam is many micrometers from the grid, Cu X-rays are excited by electron scatter and bremsstrahlung from within the specimen and their intensity generally increases with tilt. The Cr escape peak and the Si internal fluorescence peak are also visible.

Both of these phenomena have only a small horizontal component of intensity. The effects of the specimen interacting with X-rays which it has generated will be further reduced if you use thin foils, 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. We do not know 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.

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-atomic-number material. Use of low-Z materials will also remove from your 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 in addition to Be support grids are essential for X-ray microanalysis. 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. The narrow polepiece gap, required to produce high probe currents, and the cold finger, used to reduce hydrocarbon contamination of the specimen, 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 (Nicholson *et al.* 1982), 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 microanalysis.

> If you are seeking small amounts (<~1–2%) of an 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* conclusively determine the presence of element A in the specific region of the specimen.

A small peak from A will *invariably* be present in all spectra, just as surely as the Si or Ge internal fluorescence peak 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 fraction of time (say 10–20 minutes), then in addition to the C or B peak, if your XEDS can detect it, the various instrumental 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 microscope 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.

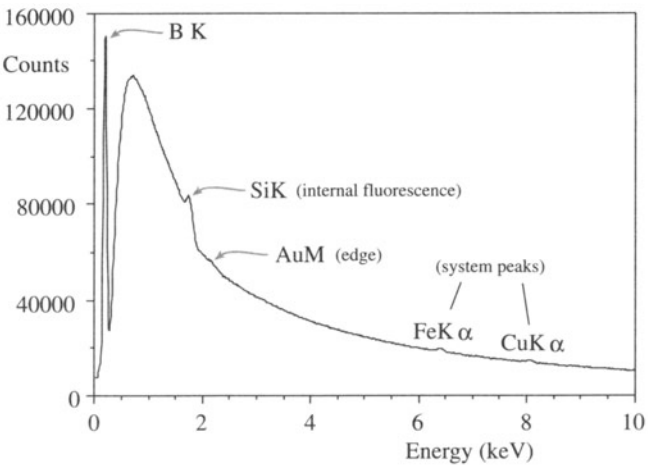


Figure 33.8. An XEDS spectrum from high-purity boron, showing system peaks. The Si K_α peak and the Au M absorption edges are detector artifacts, but the small peaks at 6.4 keV and 8 keV are Fe and Cu system peaks, respectively.

33.4.C. Coherent Bremsstrahlung

As we noted back in Chapter 4, 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 materials by electrons with energies <~30 keV, such as in an SEM. However, in thin monocrystalline 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.* (1984). Figure 33.9A shows a portion of an X-ray spectrum 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 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 with the regularly spaced nuclei. As the beam electron proceeds through the crystal 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 (= $1/H$ in a zone-axis orientation), 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. Unfortunately, you can't remove the CB effects entirely by operating far from a major zone axis, since some residual peaks are invariably detectable.

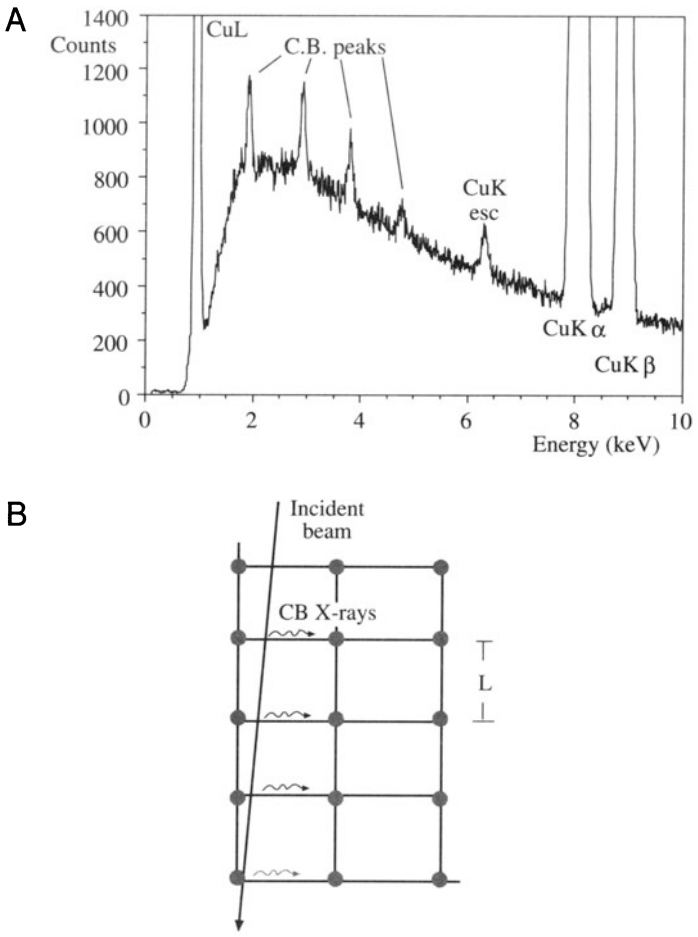


Figure 33.9. (A) CB peaks in a spectrum from pure Cu and (B) the regular generation of CB when the beam passes close to a row of atoms in the specimen.

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.

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 the specimen. While the CB peaks are a nuisance, it may be possible to use them to advantage. There is some evidence that the true bremsstrahlung intensity is low in the regions between the CB peaks. Therefore, if you are seeking to detect a very small amount of segregant, e.g., S segregated to grain boundaries in Cu, then it is possible to "tune up" the CB peaks by careful choice of kV and orientation to ensure that the S K_{α} line will appear between two CB peaks and not be masked by them.

33.5. PEAK-TO-BACKGROUND RATIO

The best test of how well your XEDS is interfaced to your TEM is to measure the peak-to-background (P/B) ratio in a standard specimen (the 100-nm Cr film). There are several

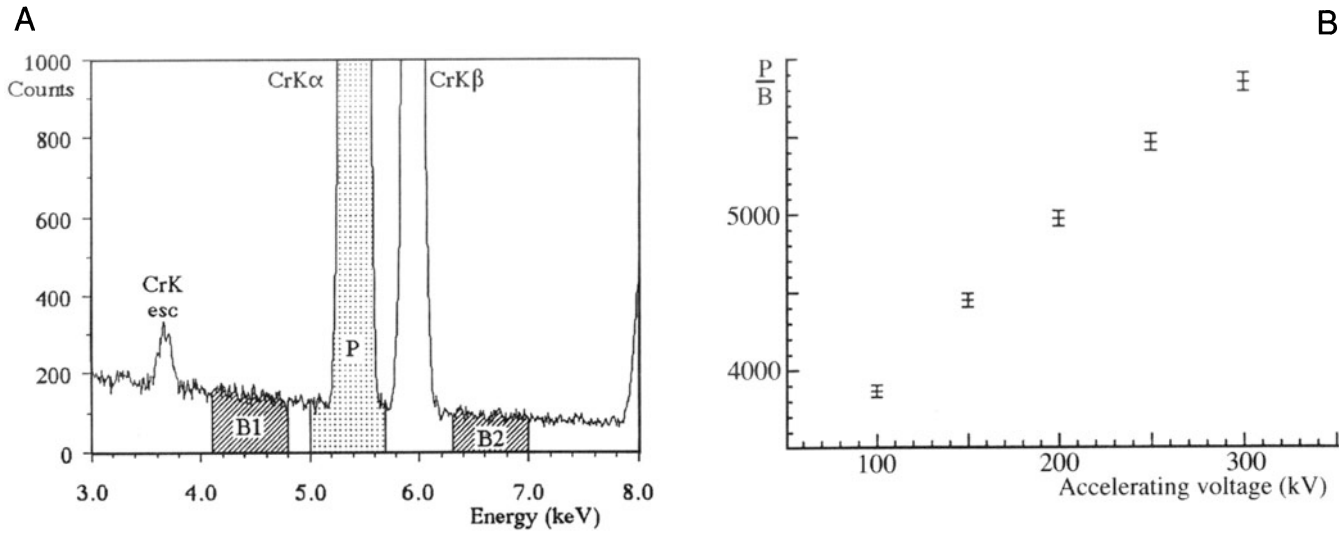


Figure 33.10. (A) The Fiori definition of the peak to background ratio for Cr. The total CrK_α peak intensity *P* is integrated from 5.0 to 5.7 keV. The background windows B1 and B2 are integrated over seventy 10-eV channels from 4.1 to 4.8 keV and from 6.3 to 7.0 keV, respectively. The average of the two windows [(B1 + B2)/2] is B(avg). This B(avg) is divided by 70 to give the background in a single 10-eV channel [B(10 eV)]. The Fiori definition is given by P/B = [P – B(avg)]/[B(10 eV)]. (B) The increase of the Fiori P/B with accelerating voltage in a well-behaved 300-keV IVEM.

definitions of P/B ratio, but the best one (termed the "Fiori" definition, Fiori *et al.* 1982) is shown in Figure 33.10A. For the Cr K_α peak, you should integrate the peak intensity from 5.0 keV to 5.7 keV and divide this by the average background intensity in a 10-eV channel, as shown in Figure 33.10A. In a well-behaved AEM, the P/B ratio will increase with keV. Recommended P/B values, as shown in Figure 33.10B (Zemyan and Williams 1994), should be close to 4000 at 100 keV, rising to almost 6000 at 300 keV.

CHAPTER SUMMARY

The (S)TEM is not well designed for unambiguous X-ray analysis because X-rays are generated and detected from many sources 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 spectrum is primarily from your specimen and your 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.

You must always:

- Ensure the XEDS is pointing toward the thin edge of any wedge specimen.
- Have the shutter closed until you know the area you want to analyze.
- Operate with clean, top-hat C2 diaphragms.
- Use thin foils, flakes, or films rather than self-supporting disks.
- Image the electron beam on the TEM screen to ensure that it is Gaussian.
- Remove the objective diaphragm.
- Operate as close to zero tilt as possible.

Check that:

- The hole count is <1%.
- You know your system peaks and other artifacts.

REFERENCES

General References

Goldstein, J.I. (1979) in *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 83, Plenum Press, New York.

Goldstein, J.I., Williams, D.B., and Fiori, C.E. (1986) in *Principles of Analytical Electron Microscopy* (Eds. D.C. Joy, A.D. Romig Jr., and J.I. Goldstein), p. 123, Plenum Press, New York.

Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd Edition. Philips Electron Optics Publishing Group, Mahwah, New Jersey.

Zaluzec, N.J. (1979) in *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren J.I. Goldstein, and D.C. Joy), p. 121, Plenum Press, New York.

Specific References

Allard, L.F. and Blake, D.F. (1982) in *Microbeam Analysis-1982* (Ed. K.F.J. Heinrich), p. 8, San Francisco Press, San Francisco.

Cliff, G. and Kenway, P.B., *ibid.,* p, 107.

Fiori, C.E., Swyt, C.R., and Ellis, J.R., *ibid.,* p. 57.

Goldstein, J.I. and Williams, D.B. (1978) in *SEM-1978* **1**, (Ed. O. Johari), p. 427, SEM Inc., AMF O'Hare, Illinois.

Lyman, C.E., and Ackland, D.W. (1991) in *Proceedings of 49th EMSA Meeting* (Ed. G.W. Bailey and E.L. Hall), p. 720, San Francisco Press, San Francisco.

Nicholson, W.A.P. and Craven, A.J. (1993) *J. Microsc.* **168**, 289.

Nicholson, W.A.P., Gray, C.C., Chapman, J.N., and Robertson, B.W. (1982) *J. Microsc.* **125**, 25.

Reese, G.M., Spence, J.C.H., and Yamamoto, N. (1984) *Phil. Mag.* **A49**, 697.

Williams, D.B. and Goldstein, J.I. (1981) in *Energy Dispersive X-ray Spectrometry* (Eds. K.F.J. Heinrich, D.E. Newbury, R.L. Myklebust, and C.E. Fiori), p. 341, NBS Special Publication 604, U.S. Department of Commerce/NBS, Washington, DC.

Zemyan, S.M. and Williams, D.B. (1994) *J. Microsc.* **174**, 1.

Qualitative X-ray Analysis

34

CHAPTER PREVIEW

It is a waste of time to proceed with *quantitative* microanalysis from your XEDS spectrum without first carrying out *qualitative* analysis. First we will show you how to choose the best operating conditions, for both the microscope and the XEDS system. Then we will explain the best way to obtain a spectrum for qualitative analysis. Qualitative analysis requires that every peak in your spectrum be identified unambiguously, and with statistical certainty, otherwise it should be ignored. So you have to acquire a spectrum with sufficient X-ray counts to allow you to draw the right conclusions with confidence. There are a couple of simple rules to follow which allow you to do this.

> 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 this approach. First, you may be able to solve the problem at hand without the necessity of a full quantification routine. Second, when quantification is carried out, you will not spend an inordinate amount of time analyzing an artifact or a statistically insignificant peak, and you can be confident that your results are valid.

Qualitative X-ray Analysis

34

34.1. MICROSCOPE VARIABLES

When you first acquire a spectrum, the operating conditions should maximize the X-ray count rate to give you the most intensity in the spectrum, in the shortest time, with the minimum of artifacts. In addition, you want to get a good idea of which elements in your specimen are detectable. The best conditions for qualitative analysis require that you obtain the spectrum from a large area of the specimen, using a large probe, and so spatial resolution will be poor. Having carried out qualitative analysis of a relatively large region, you may then wish to do further analysis of smaller areas, under conditions that optimize spatial resolution, which we discuss in Chapter 36.

To get the most X-ray counts in your spectrum, use the highest operating voltage, since this gives the highest brightness. Notice that in Chapter 33 we said that you get a higher count rate 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. While the two effects counter each other, the added advantage of increased peak-to-background ratio with increased kV tips the scales in favor of using the maximum kV. Only choose a lower voltage if knock-on damage is a problem, as might be the case, for example, in a 200-kV to 400-kV instrument. Pick a portion of your specimen that is single phase in the area of interest and is well away from strong diffraction conditions, so as to minimize crystallographic effects and coherent bremsstrahlung. You will need a probe current of several tens of nanoamps. The necessary combination of probe size and final aperture obviously depends on the type of source in your microscope. To get several tens of nanoamps, from a thermionic source, you will have to select a relatively large probe size, say several tens of nanometers, and a large C2 aperture. As we shall see later, these are just the opposite of the requirements for microanalysis at high spatial resolution. In contrast to the limitations of a thermionic source, under most operating conditions an FEG source will give sufficient current for both initial qualitative analysis and subsequent quantitative analysis for high spatial resolution. However, the lower current from an FEG may mean that you have to count X-rays for a longer time compared to a thermionic source.

You can always gather a more intense spectrum by choosing a thicker region of the specimen. There is nothing wrong with doing this when you are carrying out *qualitative* microanalysis. The only danger is that, if you have a few weight percent of a light element present in the sample, the X-rays may be absorbed within the specimen and so may not be detected. Also, a thick specimen degrades the spatial resolution, but we've already agreed to compromise that aspect of the microanalysis during the initial qualitative analysis.

So, good qualitative analysis requires a large number of X-ray counts in the spectrum. These counts take a long time to generate, so you run the danger of damaging or changing the chemistry of beam-sensitive specimens. You may also contaminate the chosen area. To minimize these effects 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. Use a liquid-N_2-cooled low-background holder if contamination is still a problem.

34.2. ACQUIRING A SPECTRUM

The first and most important step in qualitative analysis is to acquire a spectrum across the complete X-ray energy range. Microanalysis can often be accomplished using X-rays with energies from ~1–10 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 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.

> The first thing to do is to adjust your MCA system to display the widest possible energy range. For a Si(Li) detector, 0–40 keV is sufficient, and for an IG detector, 0–80 keV may be useful.

Of course, if you know the specimen you are analyzing, such a step may not seem essential, but it is still a wise precaution since unanticipated contaminants or trace impurities may be present. Collect a spectrum for several hundred seconds and ascertain the actual energy range over which all the characteristic peaks occur. Then, if all the peaks are present in an energy range <40 keV, regather the spectrum over that reduced range, thereby improving the resolution of your MCA display by lowering the number of eV per channel. The spectrum that you finally gather for qualitative analysis should be displayed with no more than 10 eV per channel resolution on the MCA, and a display range of 0–20 keV should be possible under these conditions (i.e., 2048 channels in total).

You can also increase the intensity of the spectrum by lowering the detector time constant to maximize the throughput of counts. This step also degrades the energy resolution of the XEDS but, for many qualitative analyses, this is not important.

> You should use the shortest time constant while maintaining adequate resolution to discriminate the characteristic peaks in the spectrum.

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 5000–10,000 cps is about the best that can be handled by most analog detector electronics under these conditions.

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 — detectability depending on P/B (see Chapter 36).
- Energy resolution — identifying elements by distinguishing peaks; different eV (this chapter).

When you've got a good intense spectrum over a suitable energy range, there is a well-defined sequence of steps that should be followed to ensure that you correctly identify each peak in the spectrum and disregard those peaks that are not statistically significant. 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, if your spectrum contains many peaks, and particularly if peak overlap is occurring, then misidentification may occur during such an "autosearch" routine, especially if there is no operator intervention. In addition, small peaks may sometimes be missed and phenomena such as coherent bremsstrahlung may not be taken into account. Under these circumstances there is a well-established manual sequence, developed for analysis of spectra from the SEM by Goldstein *et al.* (1992), and we will describe a modified form of this procedure in the next section.

34.3. PEAK IDENTIFICATION

First of all, we assume that you know what system peaks, if any, occur in your AEM, and what artifacts are likely to occur in your XEDS system. Now, ensure that the MCA display is calibrated to be accurate at the display resolution over the energy range you selected. So if your spectrum is displayed at 10 eV per channel, the peak centroids must be within 10 eV of their true position on the energy scale.

> The key to good qualitative analysis is to be suspicious and to not just seek the peaks you expect, but to be prepared also to find peaks that you don't expect.

Our peak analysis will always include three steps:

- Look at the most intense peak and work on down through its family; this is just bookkeeping.
- Go to the next most intense peak not included in the previous step and repeat the search. Then repeat this exercise until all peaks are identified.
- Think about pathological overlaps; look for spurious peaks, system peaks, and artifact peaks.

The bookkeeping. Starting from the high-energy end of the spectrum, select the most intense peak and determine the possible K, L, or M lines that could be present at that energy, either by using the computer-generated X-ray line markers on the MCA screen or by consulting an appropriate source, such as the "slide rules" offered by most commercial manufacturers. Good "bookkeeping" is essential during the sequence we will now describe, particularly if the spectrum contains many peaks. You must take care to label each peak when you have decided on its source. Proceed as follows:

- If a K_α line matches the peak, look for the K_β line which has ~10–15% of the K_α intensity. The K_β line *must* be present at X-ray energies above ~2.3 keV (S K_α), but below this energy the detector may not be able to resolve the two lines.
- If a K_α and K_β pair fit 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 a UTW detector, the L_α lines from Cl and above (>0.2 keV) may be detectable. Ni L_α = 849 eV, Cl L_α = 200 eV.
- If a K_α line does not fit, check for an L_α or M_α line fit.
- 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. 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 intensity 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 L_ℓ (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-line family is extraordinarily intense, but this is rare.
- If the L lines fit, there *must* be a higher-energy K_α/K_β pair, assuming the beam energy is sufficient to generate the K lines and the MCA energy range is wide enough to display them.
- The M lines are only usually 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. La M_α = 833 eV, Nb M_α = 200 eV.
- The M_α/M_β line overlap is difficult to resolve because all the M lines are below 4 keV. If an M_α/M_β line fits, look for three very small M_ζ (0.06), M_γ (0.05), and $M_{II}N_{IV}$ (0.01) lines .
- If the M_α line fits, there *must* be a higher-energy L-line family and possibly the very high energy K lines may exist; again, this depends on the detector, MCA display, and the accelerating voltage.

Figure 34.1 shows the families of lines expected in the display range from 0 to 10 keV, giving you some idea of the distribution of families of elemental lines that you should expect when you follow the procedure outlined above. For example, you can see for which elements you should expect to see 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.

> The idea is that you are looking for families of peaks. If a family member is missing your identification may be wrong.

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.* Finally, look for the escape peak(s) and sum peak associated with all *major* characteristic peaks that you have conclusively identified. Remember that these artifacts and any CB peaks will be very small, and before you worry about them you should make sure that the peaks are statistically significant; we discuss how to do this for all minor peaks in Section 34.4 below. If you have a Si(Li) detector, 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 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 (much less intense) 1.19 keV for the Ge L_α escape). If you suspect a sum peak at twice the energy of any major peaks, then reacquire the spectrum at a much lower dead time (<20%) and see if the suspected sum peak disappears. If you suspect a CB peak, then reacquire 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 samples that cannot be resolved. These are called "pathological overlaps" and include (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; and finally (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 systems.

> *Pathological overlap:* When it is impossible to separate two peaks even when you know they are both there.

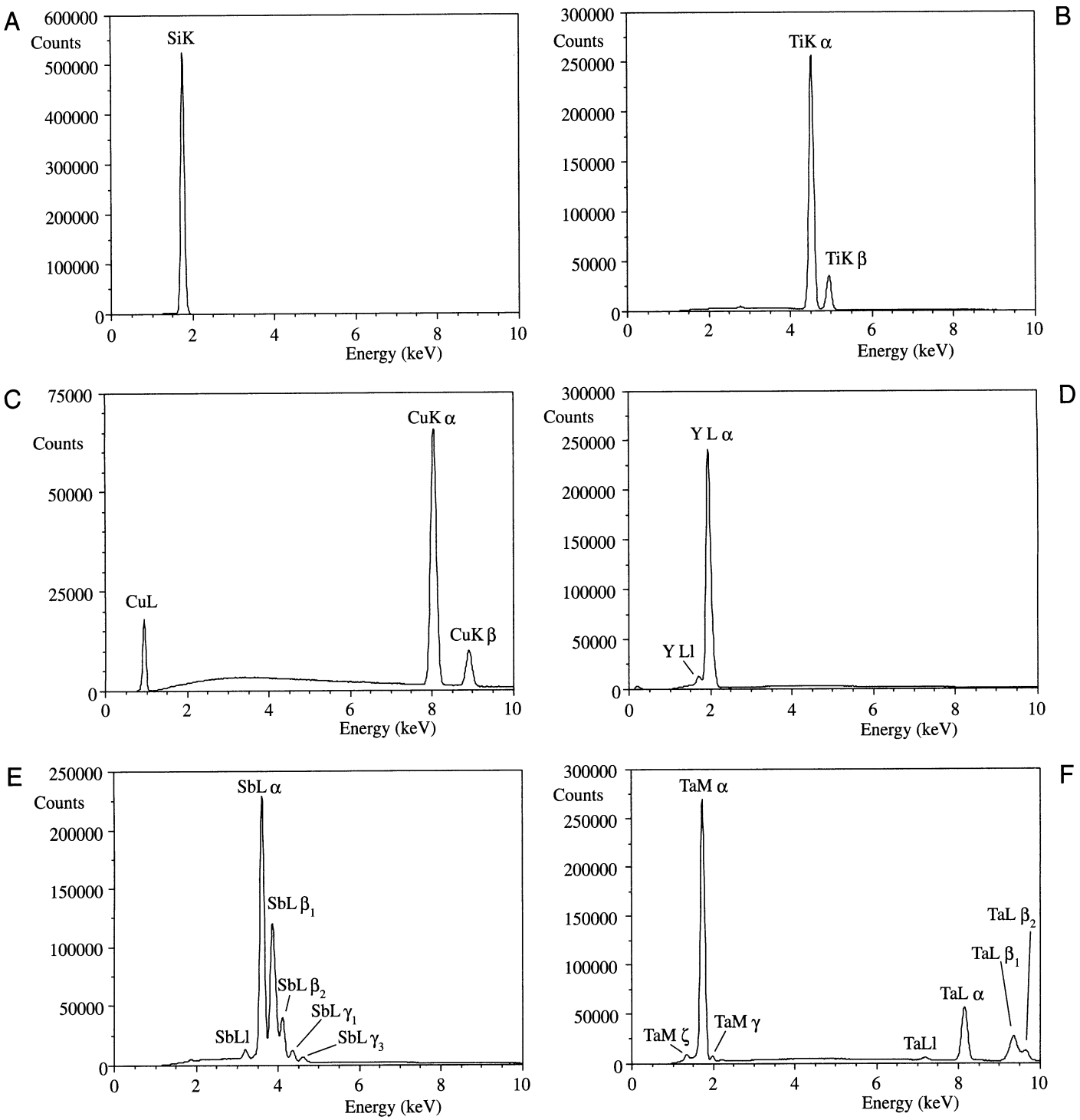


Figure 34.1. X-ray spectra from six elements spanning much of the periodic table, showing the families of characteristic lines. Starting with a single K_α line at low Z and low X-ray energy, the series progresses through the appearance of the L- and M-line families. Note the increasing separation of the lines in a given family as Z and keV increase.

These problems can sometimes be approached by careful choice of the MCA display range. For example, if you are only observing from 0–10 keV, the S K/Mo L line overlap would be clarified by the presence or absence of the Mo K lines around 18 keV. If you suspect that these or any other pathological peak overlap is occurring in your spectrum, then the first thing to do is to regather the spectrum under conditions that maximize the energy resolution of the detector system (i.e., longest time constant and count rate below 5000 cps), and also maximize the display resolution to at least 5 eV per channel. If the overlap is still not resolvable, then you should run a peak deconvolution routine that should be present in the available computer programs (e.g., Schamber 1981). Such routines are capable of detecting and resolving many of the classical materials science overlaps such as Mo L_α and S K_α, and a schematic deconvolution is shown in Figure 34.2.

A historical aside: peak deconvolution has been of great concern to microanalysts since the earliest days of EPMA, using pulse-height analyzers, and the first primitive attempt to deconvolute overlapping peaks was developed by Dolby (1959). With hindsight, it is not difficult to make the connection between the problems of extracting peak information from a low-resolution spectrum and the problem of extracting clear sound signals from noisy recording tape. Dolby, however, saw the potential before anyone else and went on to commercialize his ideas with resounding success, prompting his Ph.D. supervisor Ellis Cosslett to remark that Dolby was the only graduate student he had known to become a millionaire from his Ph.D. research!

This procedure should permit you to identify all the major peaks in the spectrum, but there may still be minor peaks which may or may not be significant and you have to decide whether you are going to identify or ignore these peaks. We'll tell you how to make this decision in the next section.

34.4. PEAK VISIBILITY

Small-intensity fluctuations that you cannot clearly identify as peaks are often present in your spectrum. In this case, there is a simple statistical criterion (Liebhafsky *et al.* 1972) 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 relatively smooth and any peaks are clearly visible, as summarized in Figure 34.3.

- 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.

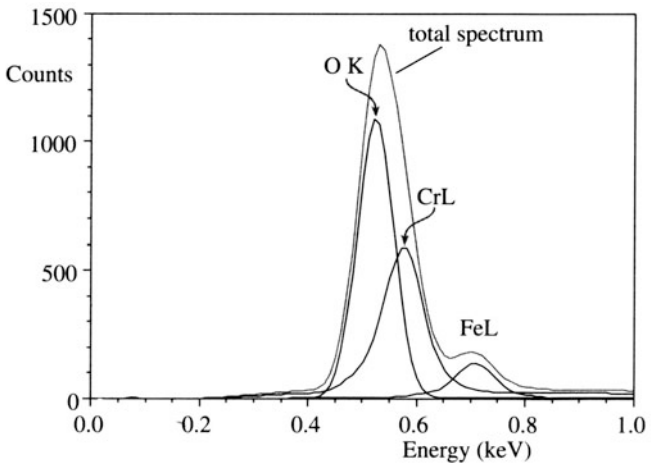


Figure 34.2. The total spectrum arises from the overlap of three Gaussian spectral peaks (the L_α lines of Fe and Cr and the O K_α line) from a mixed Fe-Cr oxide. Deconvolution is essential to determine the intensities in the three constitutent peaks, prior to any attempt at quantification.

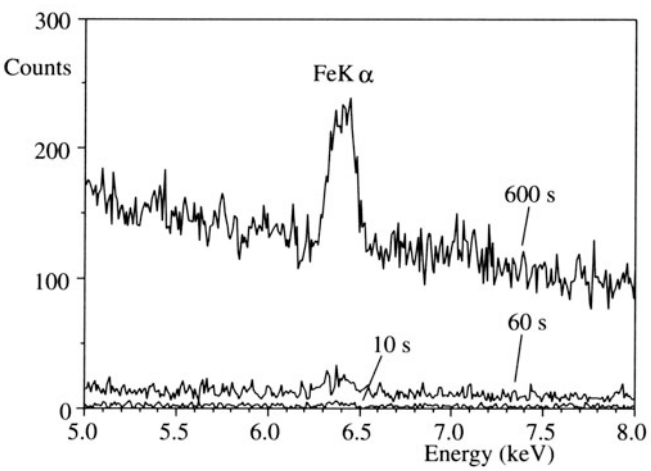


Figure 34.3. With increasing counting time a clear characteristic Fe K_α peak develops above background in a spectrum from Si-0.2% Fe. This demonstrates the need to acquire statistically significant counts before deciding if a small peak is present or absent.

If $I_A > 3\sqrt{I_A^b}$, then the peak is statistically significant at the 99% confidence limit and must be identified. You will make an erroneous peak identification in less than 1% of analyses using this criterion.
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 sample, 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, as long as doing so does not change or contaminate 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. Be aware that when you count for long periods of time to search for characteristic peaks of low intensity, you will also begin to detect the small peaks that arise from the various spurious effects that were discussed in detail above, e.g., CB peaks, Si or Ge internal-fluorescence peaks, and system peaks such as Fe and Cu. Also, you increase the possibility of contamination and beam damage to your specimen. So, 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 detectability limits in Chapter 36. However, you may be able to identify the phase that is being analyzed without any further work. For example, in the material that you are investigating, there may be 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 shall see in the next chapter, one of the marvelous advantages of thin-foil microanalysis 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:

An example of qualitative microanalysis is shown in the spectrum in Figure 34.4. 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 0–10 keV and the first peak to be examined was the most intense high-energy peak, labeled #1 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). 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 K_β) and line #3 was consistent with the Si K_α line at 1.74 keV (the K_α/K_β pair cannot be resolved at this energy). 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. No escape or sum peaks were detectable, 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 we cannot conclude that there is any Cu within the specimen.

The absence of the Cu L_α line at 0.93 keV is evidence that the thick Cu grid is responsible, since the low-energy L X-rays will be absorbed in the grid itself before they can be detected.

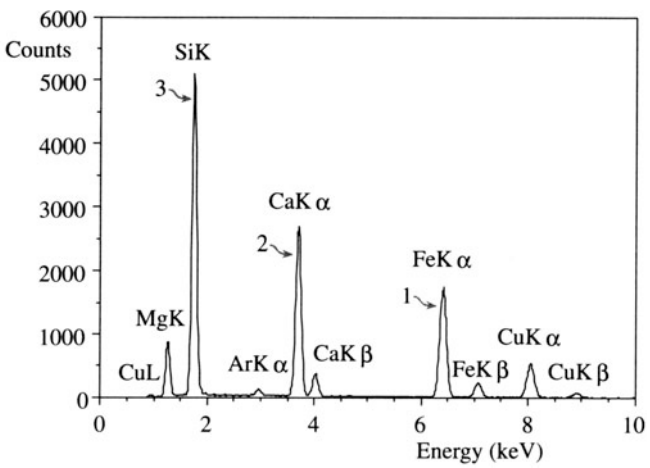


Figure 34.4. Energy-dispersive spectrum obtained at 300 kV from a thin oxide glass film on a Cu grid, with the characteristic peaks identified through the procedure outlined in the text.

Another example is shown in Figure 34.5 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. To the average metallurgist, this sample 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 more information is required, such as the specific grade of stainless steel, then it is necessary to make some measurements of the relative peak intensities, and this is the first step in the quantification procedure. In fact, we will see that the quantification equation, to a first approximation, predicts that the amount of each element is directly proportional to the peak height, and so measuring the relative heights of the K_α peaks in Figure 34.5 with a ruler will give an estimate of the composition as ~Fe-20% Cr-10% Ni. A full quantification using the procedures described in the next chapter gives a very similar result, but with much greater confidence in the true composition.

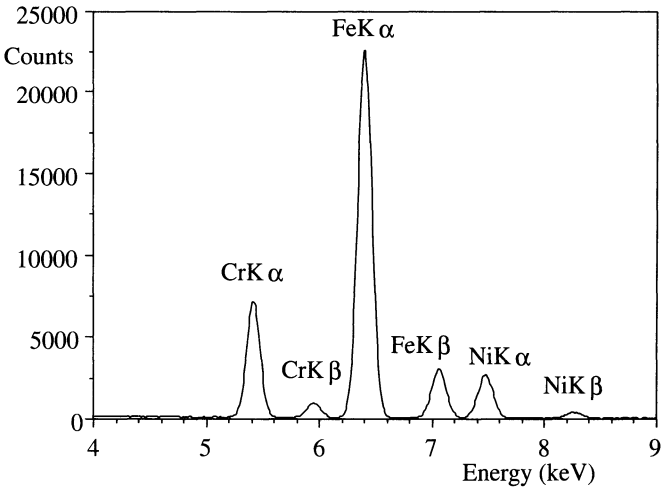


Figure 34.5. Spectrum from a stainless-steel foil. From such a spectrum, in which the peaks are resolved and close together in energy, a first-approximation quantification is possible simply by measuring the relative heights of the K_α peaks.

CHAPTER SUMMARY

One last time: doing the qualitative analysis first is not an option. It is essential.

- 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.

REFERENCES

General References

Goldstein, J.I., Newbury, D.E., Echlin, P., Joy, D.C., Romig, A.D. Jr., Lyman, C.E., Fiori, C.E., and Lifshin, E. (1992) *Scanning Electron Microscopy and X-ray Microanalysis,* 2nd edition, p. 341, Plenum Press, New York.

Russ, J.C. (1984) *Fundamentals of Energy Dispersive X-Ray Analysis,* Butterworths, Boston.

Specific References

Dolby, R.M. (1959) *Proc. Phys. Soc. London* **73**, 81.

Liebhafsky, H.A., Pfeiffer, H.G., Winslow, E.H., and Zemany, P.D. (1972) *X-rays, Electrons and Analytical Chemistry,* p. 349, John Wiley and Sons, New York.

Schamber, F.H. (1981) in *Energy Dispersive X-ray Spectrometry* (Eds. K.F.J. Heinrich, D.E. Newbury, R.L. Myklebust, and C.E. Fiori), p. 193, NBS Special Publication 604, U.S. Department of Commerce/NBS, Washington, DC.

Quantitative X-ray Microanalysis

35

CHAPTER PREVIEW

You've now got an idea of how to acquire XEDS data from thin foils. You understand what factors may limit the information in them and what false and misleading effects may arise. Also, you know how to be sure that a certain peak is due to the presence of a certain element and the occasions when you may not be so sure. Having obtained a spectrum that is qualitatively interpretable, it turns out to be a remarkably simple procedure to convert that spectrum into quantitative data about the elements in your specimen, and this is what we describe in this chapter.

This chapter is rather long. You will 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 the analysis on the microscope.

Quantitative X-ray Microanalysis

35

35.1. HISTORICAL PERSPECTIVE

Quantitative X-ray analysis in the AEM is a most straightforward technique. What is surprising is that, given its simplicity, relatively few users take the trouble to extract quantitative data from their spectra, 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 microanalysis, because this will emphasize the advantages of thin-foil microanalysis over analysis of bulk specimens.

Historically, X-ray microanalysis in electron-beam instruments started with the study of bulk specimens in which the electron beam is 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 (1944), and the necessary instrumentation was built several years later by Castaing (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 $I_i/I_{(i)}$, where I_i is the measured intensity emerging from (*not* generated within) the *specimen* and $I_{(i)}$ is the measured intensity emerging from the *standard*.

Castaing then proposed that, to a reasonable approximation

$$C_i / C_{(i)} = [K]\, I_i / I_{(i)} \qquad [35.1]$$

where K is a sensitivity factor 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 microanalysis 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, Heinrich and Newbury 1991.)

It was soon realized that if a thin electron-transparent specimen was used rather than a bulk specimen, then the correction procedure could be greatly simplified because, to a first approximation, the A and F factors could be ignored and only the Z correction would be necessary. In addition, if thin specimens were used, the analyzed 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 microanalysis led to the development of the so-called electron microscope microanalyzer (EMMA), pioneered by Duncumb in England in the 1960s. Unfortunately the EMMA was ahead of its time, mainly because the WDS was the only X-ray detector system available. As we have seen, the WDS is handicapped by its poor collection efficiency, rela-

tively 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 intensity for quantification, 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) 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. We've seen that the XEDS detector was developed in the late 1960s, and about the same time the development of commercial TEM/STEM systems was beginning. However, before the demise of the EMMAs, they were to play a critical role in the development of the thin-foil microanalysis procedures that we use today. The EMMA at the University of Manchester, operated by Cliff and Lorimer, was refitted 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 microanalysis.

35.2. THE CLIFF–LORIMER RATIO TECHNIQUE

The basis for this technique is to rewrite equation 35.1 for 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 intensity from a standard.
- 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 be related to the measured intensities by the so-called Cliff–Lorimer equation

$$\frac{C_A}{C_B} = k_{AB}\frac{I_A}{I_B} \qquad [35.2]$$

The term k_{AB} is often termed the Cliff–Lorimer factor. It is actually *not* a constant, so don't be fooled by the use of "*k*." It varies according to the TEM/XEDS system and the kV, as we will see later. Because we are ignoring the effects of absorption and fluorescence, k_{AB} is related to the atomic-number correction factor (Z) in Castaing's original ratio equation. (We will derive this relationship rigorously in Section 35.4.C.) 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

$$C_A + C_B = 100\% \qquad [35.3]$$

We can easily extend these equations to ternary and higher-order systems by writing extra equations of the form

$$\frac{C_B}{C_C} = k_{BC}\frac{I_B}{I_C} \qquad [35.4]$$

and

$$C_A + C_B + C_C = 100\% \qquad [35.5]$$

You should also note that the *k* factors for different pairs of elements AB, BC, etc., are related, thus

$$k_{AB} = \frac{k_{AC}}{k_{BC}} \qquad [35.6]$$

It is a convention that we define the units of composition as wt.%.

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. Thus the Cliff–Lorimer equation is the basis for quantitative microanalysis on the AEM. Let's see how we use it in practice.

35.3. PRACTICAL STEPS FOR QUANTITATIVE MICROANALYSIS

First of all, you should try to use only the K_α lines for the measured intensity. (The K_β is combined with K_α if it 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 for your Si(Li) detector.

We can break the process down into four accumulation steps:

- 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 this number in a reasonable time before specimen drift, damage, or contamination limit 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.
- Keep your specimen as close to 0° tilt as possible to minimize spurious effects.
- Orient your specimen so the thin portion of the wedge faces the detector to minimize X-ray absorption (see Section 35.5).
- If the area of the specimen is close to a strong two-beam dynamical-diffraction condition, you should tilt the specimen slightly to kinematical conditions.

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. We will see in Section 35.8 that under certain conditions there are 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 intensity from the spectrum and then you integrate the peak intensity. In a modern computerized MCA system, both these steps are accomplished by one of several available software routines. 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

We are not very precise in the terminology that we use in the discussion of the X-ray background intensity, so it can be confusing. The "background" refers to the intensity under the characteristic peaks in the spectrum displayed on the MCA screen. Now, as we saw back in Chapter 4, 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 these bremsstrahlung X-rays decreases continuously as the X-ray energy increases, reaching zero at the beam energy. Thus the energy distribution can be described as a "continuum," although, as we've seen, the phenomenon of coherent bremsstrahlung disturbs this continuum.

> Now 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 2 keV by absorption within the detector and the specimen, 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 whether the region of interest in your spectrum is in this low-energy regime, and if the characteristic peaks you want to measure are close together or isolated.

Window methods. In the most simple case of isolated characteristic peaks superimposed on a slowly varying background, you can easily remove the background intensity 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

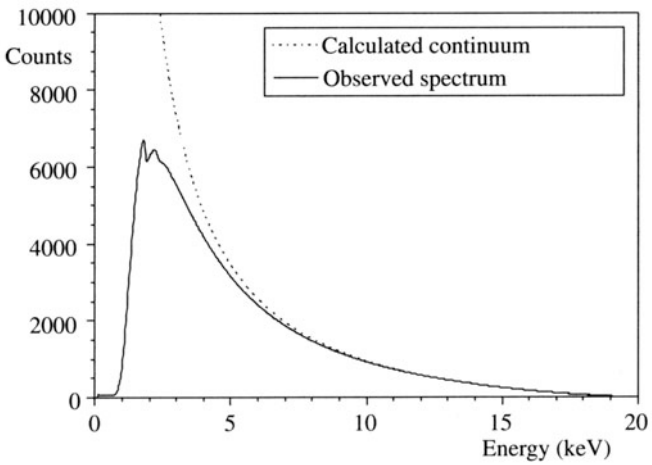


Figure 35.1. The theoretically generated and experimentally observed bremsstrahlung intensity distribution as a function of energy. Both curves are similar until below about 2 keV, when absorption within the specimen and the XEDS system reduces the experimental intensity. Background removal depends on where in the spectrum your characteristic peaks are present.

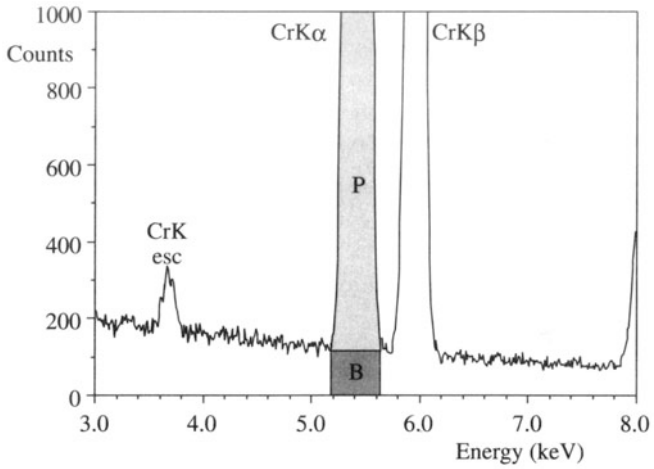


Figure 35.2. The simplest method of estimating the background contribution (B) to the intensity in the characteristic peak (P); a straight line drawn beneath the Cr K_α peak provides a good estimate if the counting statistics are good, and the bremsstrahlung intensity approximates to a slowly varying linear function of energy. There should be no overlap with any other characteristic peaks, and the peaks should be well above ~2 keV.

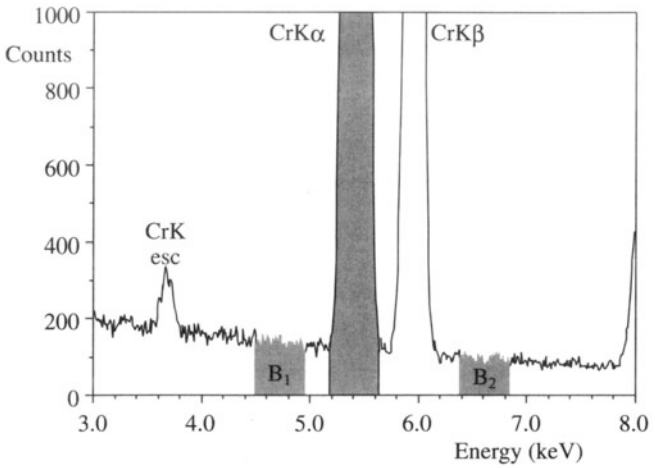


Figure 35.3. Background subtraction can be achieved by averaging the bremsstrahlung intensity in two identical windows (B_1, B_2) either side of the characteristic (Cr K_α and K_β) peaks. There should be no overlap with any other characteristic peaks, and the peaks should be well above ~2 keV.

of the peak, and then draw the line between the background intensities in the channels just outside the window. As with all spectral manipulations, this method gives better results with greater intensity levels 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. Furthermore, the background intensity variation better approximates to a straight line.

Another similarly primitive approach involves averaging the bremsstrahlung intensity above and below the characteristic peak by integrating the intensity in two identical windows either side of the peak, as shown in Figure 35.3. You then assume that the average of the two intensities equals the background intensity under the peak, so you subtract this average from the total peak intensity. 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, which would cause a discrete change in intensity under the peak. When you use this approach it is essential to 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 used to determine the k factor (see Section 35.4).

A typical choice of window width is FWHM, but this throws away a substantial amount of the intensity in the peak. FWTM gives better statistics, but incorporates more bremsstrahlung; 1.2 FWHM is the optimum window.

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 intensity 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:

$$N_E = KZ\frac{(E_0 - E)}{E} \qquad [35.7]$$

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.

All these terms have to be factored into the computer calculation when you use this method of background modeling. In Figure 35.4, the bremsstrahlung is modeled using the Kramers–Small cross section in the DTSA software (see Section 1.5).

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.

A more satisfactory approach, from a scientific standpoint, is to use the modified Bethe–Heitler formula, as discussed by Chapman *et al.* (1984). This formula is explicitly derived for thin foils and high-keV electrons. This model yields an expression for the bremsstrahlung cross section as a function of the X-ray energy (E) and the atomic number (Z) of the specimen. The slow variation of the cross section with Z leads to the possibility of fitting a simple quadratic expression of the form

$$N_E = \left(\frac{a_0}{E} + a_1 + a_2 E\right)\varepsilon \qquad [35.8]$$

where a_i are simply fitting parameters and ε is the detector efficiency (plotted back in Figure 32.7). The term is only important when modeling the background below ~1.5 keV, and we discuss it in detail later on in Section 35.4.

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 intensity by a window method. Figure 35.4 shows an example of a spectrum containing three adjacent low-energy peaks, with the background intensity 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 intensity as a function of energy (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.

After the top-hat filter, the background with small dI/dE is transformed to a linear 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 steps required for the filtering process and Figures 35.5B and C show an example of a spectrum before and after digital filtering.

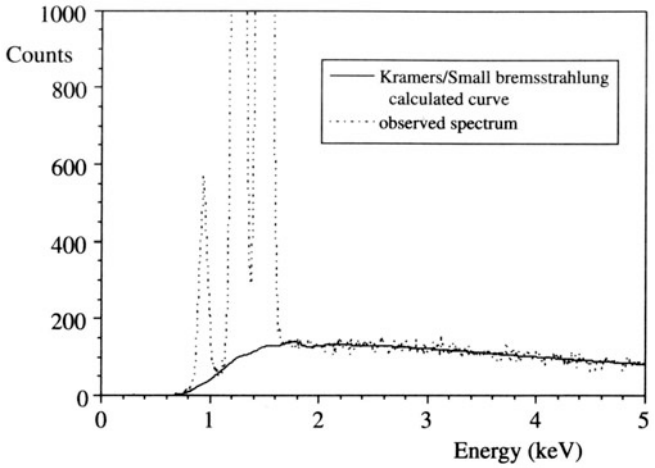


Figure 35.4. The bremsstrahlung intensity modeled using a modified Kramers' Law, which includes the effects of absorption of low-energy X-rays in the specimen and the detector. This method is useful when the spectrum contains overlapping peaks, particularly in the low-energy range, such as the Cu L_α and the Mg and Al K_α lines shown in this spectrum.

In summary, you can remove the background by selecting appropriate windows to estimate the intensity under 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 and you must always take care to apply the same process to both the standard and the unknown. After removing the background, the next thing you have to do is 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 intensity is obtained simply by subtracting the estimated background intensity from the total intensity in the chosen window. Therefore, if you drew a line under the peak as in Figure 35.2, then the peak intensity is that above the line.

- If you chose a window of FWHM and averaged the background on either side of the peak, then

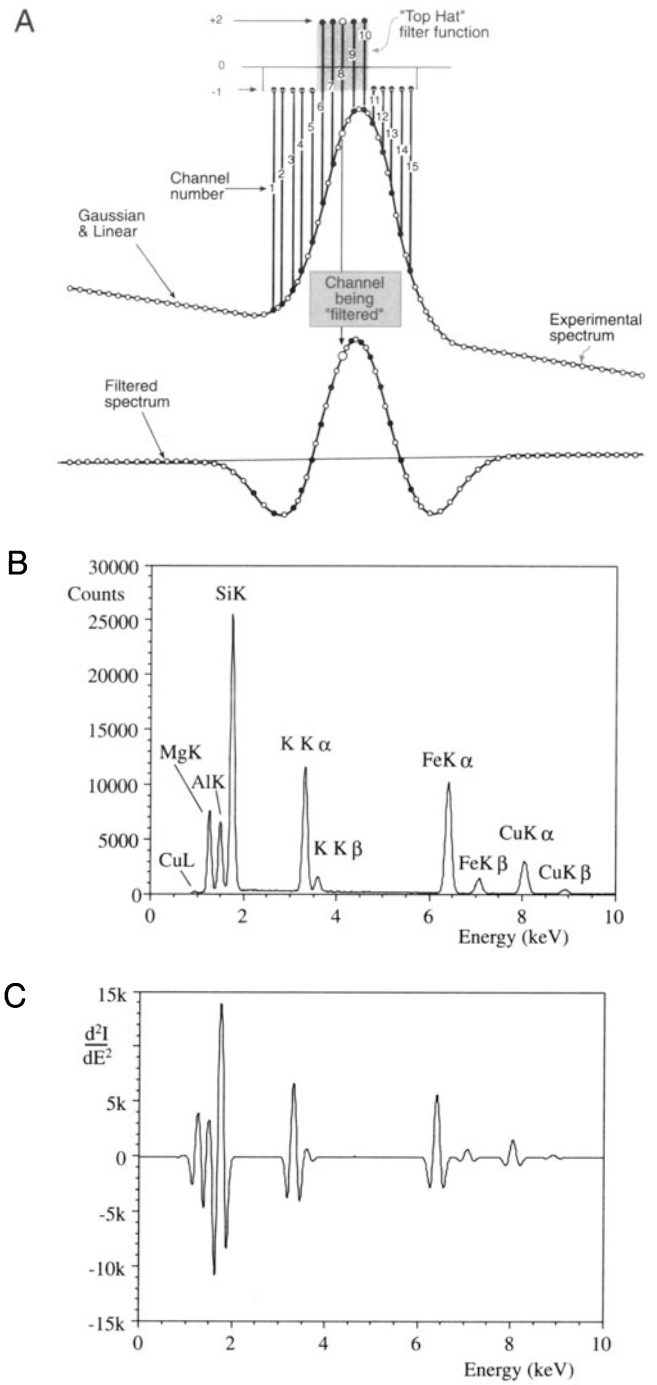


Figure 35.5. (A) Digital filtering involves convolution of a top-hat filter function with the acquired spectrum. To obtain the filtered spectrum, each channel has the top-hat filter applied to it. The channels either side of that being filtered (#8 in this case) are multiplied by the appropriate number 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 assigned to zero at all places, and the characteristic peaks remain effectively unchanged.

the average value must be subtracted from the total intensity in the FWHM window.

- 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 intensity determined through calculation of the fitting parameters.

Each of these curve-matching processes is rapid. Each can be used to deconvolute overlapping peaks, and each uses all the counts in the peak. The Kramers' 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 "tailing" due to incomplete charge collection can vary.
- There may be a low-energy background "shelf" and an absorption edge if the specimen is too thick.

The digital-filter approach requires comparison of experimental peaks with library standards, and this means that you have to create a library of stored spectra under conditions that match those liable to be encountered during microanalysis (particularly, similar count rates and dead times). This 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 chi-squared value is given which has no absolute significance, but is a most useful diagnostic tool. Typically, the chi-squared 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 chi-squared 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 hid-

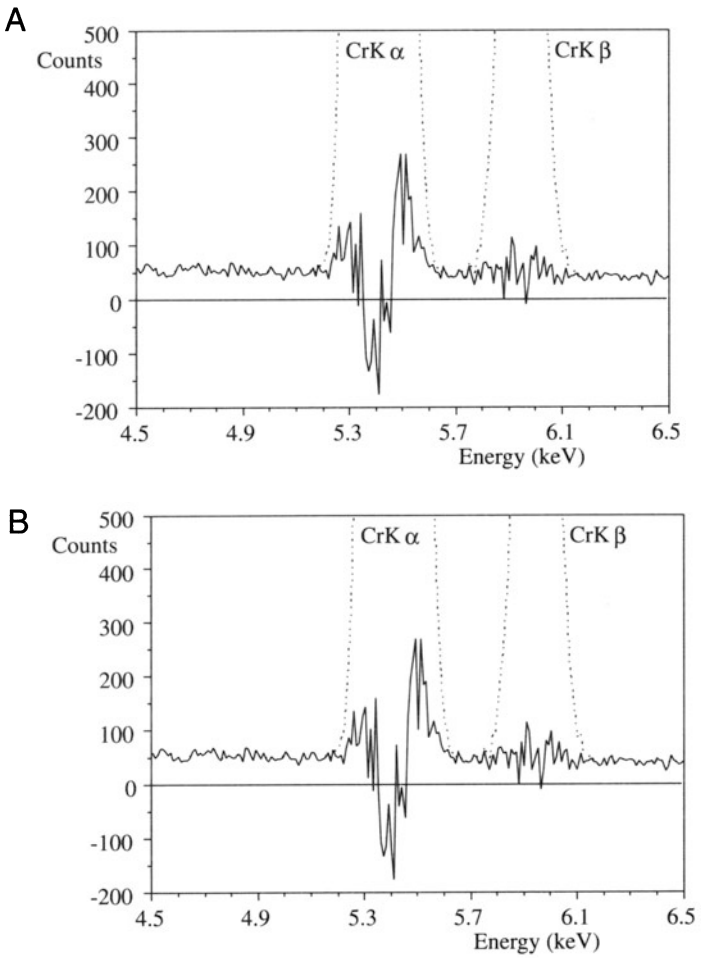


Figure 35.6. (A) A filtered Cr spectrum showing the residual background intensity after the peaks have been removed for integration. The approximately linear residual intensity distribution indicates that the peak matched well with the library standard stored in the computer. (B) A similar filtered spectrum showing the distorted residual counts characteristic of a poor fit with the library standard.

den 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 intensity 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).

The point we are making is that any of the above methods is valid for obtaining values of the peak intensity. 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.

The next step is to insert the values of the peak intensities in the Cliff–Lorimer ratio 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 microanalysis 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.

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 (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 bulk.
- You must be sure that the thin foil is stable under the electron beam at the voltage you intend to use for microanalysis.

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, Fe, and O (SRM #2063). 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. In fact, there are no generally accepted standards that meet all the above criteria for ideal k-factor determination. It is best to use your own judgment in this respect, and also make use of the knowledge gained in previous k-factor studies. The original work of Cliff and Lorimer (1975) was based on crushed mineral standards. Their approach has two advantages:

- Crushing does not affect the chemistry; the stoichiometry is well known.
- Minerals contain Si, thus permitting the creation of a whole series of k_{ASi} factors.

The drawbacks are that the mineral samples often contain more than one phase, or may be naturally nonstoichiometric. Clearly, some prior knowledge of the mineralogy of the sample 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.

Several alternative approaches that attempt to avoid the problems with k_{ASi} have been proposed:

- Wood *et al.* (1984) generated a series of k_{AFe} factors to overcome the Si absorption and the beam sensitivity problems.
- Graham and Steeds (1984) used crystallized microdroplets.
- Sheridan (1989) demonstrated the value of the NIST multi-element glasses.

In all cases the bulk chemistry has to be determined by some acceptable technique, 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 submicron 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. A typical k-factor determination therefore involves taking many spectra from different parts of the thin-foil standard to check both the homogeneity and the stability of the specimen. 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 thin foil that removes the problems of absorption and fluorescence usually results in 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. The end result is that poor counting statistics are the primary source of error in the quantification. 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. In any case you should be prepared to count for a long time, assuming that specimen drift and/or contamination don't compromise the data.

> 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.

The rest of this section is purely statistics. If you know it, then jump ahead.

Given that our characteristic peak is Gaussian, then the standard deviation σ is obtained from

$$\sigma = N^{\frac{1}{2}} \qquad [35.9]$$

where N is the number of counts in the peak above the background. For a single measurement there is a 67% chance that the 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

$$\text{Relative Error} = \frac{3N^{\frac{1}{2}}}{N}\,100\% \qquad [35.10]$$

Clearly the error decreases as N increases, and hence the emphasis throughout this chapter is 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 and, strictly speaking, we should add the relative standard deviations in quadrature using the expression

$$\left(\frac{\sigma_C}{C_A/C_B}\right)^2 = \left(\frac{\sigma_{k_{AB}}}{k_{AB}}\right)^2 + \left(\frac{\sigma_{I_A}}{I_A}\right)^2 + \left(\frac{\sigma_{I_B}}{I_B}\right)^2 \qquad [35.11]$$

So we can determine the error for each datum point in this manner. If we are determining the composition of a single phase, for example during the determination of 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 "t" distribution. For example, in this approach the error is given by

$$\text{Absolute Error} = \frac{(t_{95})^{n-1} S}{n^{\frac{1}{2}}} \qquad [35.12]$$

where $t_{95}{}^{n-1}$ is the student "t" value at the 95% confidence limit for n measurements of k_{AB} (see any statistics text for a list of student "t" values, e.g., Owen 1962). Obviously, you could choose a lower or higher confidence level. Here, S is the standard deviation for n measurements of the intensity, N_i, which on average contain $\bar{N}_i$ counts.

$$S = \left(\sum_{i=1}^{n} \frac{(N_i - \bar{N}_i)^2}{n-1}\right)^{1/2} \qquad [35.13]$$

Hence 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 ±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 35.9 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] × 100%, which is ~3%. A similar value for I_B gives a total error in C_A/C_B of ~±4.5%, using equation 35.11. If you take the time to accumulate 100,000 counts for I_A and I_B, the total error is reduced to ~±1.7%, which represents about the best accuracy that can be expected for quantitative analysis in the AEM. It is appropriate here to go through an illustration of a k_{AB} determination using actual experimental data.

Before deciding that a particular specimen is suitable, it should be checked for its level of homogeneity, and 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 sample 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 microanalysis is such that there is no need to be concerned about them.

Example

A homogenized thin foil of Cu-Mn solid solution was used to determine k_{CuMn}. The sample 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 35.13). 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

$$\frac{96.64}{3.36} = k_{CuMn}\frac{271{,}500}{10{,}800}$$

$$k_{CuMn} = 1.14$$

To determine an error on this value of the k factor, equation 35.12 must be used. The student "t" analysis of the k factors from the other 29 spectra gives an error of ±0.01 for a 95% confidence limit. This error of about ±1% relative is about the best that can be achieved using the experimental approach to k-factor determination.

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 the tables are very similar, the differences cannot be 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, but sensitivity factors.

Table 35.1. Experimentally Determined k_{ASi} and k_{AFe} Factors for K_α X-ray Lines[a]

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_α X-ray Lines[a]

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

[a]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).

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, and *same* peak-integration and background-subtraction routines. Even then there will be differences if one or more of the measured X-ray lines is 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 or accuracy might not be critical, but you need a quick analysis. Then you can calculate an approximate k factor. The programs necessary to calculate k_{AB} are stored in the computer and will give a value of k in a fraction of a second. The calculated value should be accurate to within ±20% relative. Often, this level of accuracy is all you need to draw a sensible conclusion concerning the problem.

Calculating k factors is the recommended approach when a quick answer is required and accuracy is not essential.

We will derive the expression for calculating the k factor from first principles, starting in a manner similar to the development of the expressions for the analysis of bulk samples in the EPMA. The derivation gives a good illustration of the relationship between bulk and thin-film microanalysis, and provides insight into the details of X-ray interactions with solids. In addition, the equations will provide us with the necessary grounding to pursue the problems of absorption and fluorescence in thin foils, when they occur. A full discussion of this derivation is given in the paper by Williams and Goldstein (1991). If you don't need to know the details of this derivation, you may wish to move on to the final expression given in equation 35.23.

The intensity of the *generated* X-ray emission from element A in the specimen, I_A^{Spec*}, is

$$I_A^{Spec*} = \Phi_A^{\Delta\rho t} \int_0^{\infty} \varphi(\rho t)\, e^{-\chi\rho t} \left(1 + \delta_A\right) d(\rho t) \qquad [35.14]$$

- The term $\Phi_A^{\Delta\rho t}$ is the X-ray emission (in cps) generated from element A in an isolated thin film of the specimen with mass thickness $\Delta\rho t$; the thickness of this isolated film is Δ and its mass thickness is ρt (it is *not* the change in ρt).
- The term $\varphi(\rho t)$ is the depth distribution of X-ray production. We define it as the ratio of the X-ray emission from a layer of element A of thickness $\Delta\rho t$ at a depth t in the specimen to the X-ray emission from an identical, but isolated, film.
- The expression $e^{-\chi\rho t}$ accounts for X-ray absorption in the specimen, where χ is defined as

$$\chi = \left.\frac{\mu}{\rho}\right]_{Spec}^{A} \operatorname{cosec} \alpha \qquad [35.15]$$

- The term $\mu/\rho]_{Spec}^{A}$ is the mass absorption coefficient for X-rays from element A in the specimen and α is the X-ray take-off angle.

X-rays from element A may also be fluoresced by other characteristic X-rays emerging from the specimen. The fluorescence contribution to the generated intensity is $(1+\delta_A)$. We can write an expression for the intensity of X-rays from an isolated thin film within a specimen as

$$\Phi_A^{\Delta\rho t} = N\left(\frac{Q\omega a}{A}\right)_A C_A \Delta\rho t \qquad [35.16]$$

where N is Avogadro's number. The subscript A denotes the element A in each case, Q_A is the ionization cross section, ω_A is the fluorescence yield for the characteristic X-rays, A_A is the atomic weight and C_A is the weight fraction of the element.

The use of the weight fraction rather than the atomic fraction is an anomaly which has persisted from the earliest days of microanalysis, when it was thought by Castaing that an atomic number correction was not required.

The remaining term "a" is the relative transition probability. This term takes account of the fact that when a K-shell electron is ionized, the atom will return to ground state through the emission of either a K_α or K_β X-ray. The term "a" in equation 35.14 in this case would be given by

$$a = \frac{I(K_\alpha)}{I\left(K_\alpha + K_\beta\right)} \qquad [35.17]$$

You may remember that we listed the relative "weights" of the X-ray lines in Table 4.1. We can easily apply equation 35.14 to thin-film specimens. For such specimens we can then make some simplifications:

- Assume that the electrons only lose a small fraction of their energy in traversing the specimen. Therefore, Q_A is taken as a constant, and evaluated for the incident beam energy E_0.

- Limit the integral in equation 35.12 to the foil thickness t.

Thus if we substitute equation 35.16 into equation 35.14, we find

$$I_A^{Spec} = N\left(\frac{Q\omega a}{A}\right)_A C_A \Delta\rho t \int_0^t \varphi_A(\rho t)\, e^{-\chi\rho t}(1+\delta_A)\, d(\rho t) \quad [35.18]$$

The Cliff–Lorimer equation assumes that we can measure two characteristic X-ray intensities simultaneously and so we can ratio two equations like equation 35.18, cancel N and $\Delta(\rho t)$, and rewrite them thus

$$\frac{I_A^{Spec*}}{I_B^{Spec*}} = \frac{C_A \frac{Q_A \omega_A a_A}{A_A} \int_0^t \varphi_A(\rho t)\, e^{-\chi\rho t}(1+\delta_A)\, d(\rho t)}{C_B \frac{Q_B \omega_B a_B}{A_B} \int_0^t \varphi_B(\rho t)\, e^{-\chi\rho t}(1+\delta_B)\, d(\rho t)} \quad [35.19]$$

This equation can be conveniently shortened to

$$\frac{I_A}{I_B} = \frac{C_A}{C_B}(ZAF) \quad [35.20]$$

where Z, A, and F stand for the atomic-number, absorption, and fluorescence corrections, respectively. Now remember that the Cliff–Lorimer equation (equation 35.2) assumes that A and F are negligible in a thin foil. We therefore rearrange equation 35.20 to look like equation 35.2

$$\frac{C_A}{C_B} = \frac{1}{Z}\frac{I_A}{I_B} \quad [35.21]$$

By comparison of the two equations (35.2 and 35.19) we can write an expression for k_{AB}

$$k_{AB} = \frac{1}{Z} = \frac{(Q\omega a)_B A_A}{(Q\omega a)_A A_B} \quad [35.22]$$

Thus, 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 microanalysis. From equation 35.22 we can easily see which experimental factors determine the value of k.

- Obviously, 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 35.20 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 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 the 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 35.22, in the following manner

$$k_{AB} = \frac{1}{Z} = \frac{(Q\omega a)_A}{(Q\omega a)_B}\frac{A_B}{A_A}\frac{\varepsilon_A}{\varepsilon_B} \quad [35.23]$$

The symbol ε represents simply a detector-efficiency term (plotted back in Figure 32.7) that we can write as follows

$$\varepsilon_A = \exp\left(-\left[\frac{\mu}{\rho}\right]_{Be}^{A} \rho_{Be} t_{Be}\right) \exp\left(-\left[\frac{\mu}{\rho}\right]_{Au}^{A} \rho_{Au} t_{Au}\right) \cdot \exp\left(-\left[\frac{\mu}{\rho}\right]_{Si}^{A} \rho_{Si} t_{Si}\right)\left\{1 - \exp\left(-\left[\frac{\mu}{\rho}\right]_{Si}^{A} \rho_{Si} t'_{Si}\right)\right\} \quad [35.24]$$

Here, the mass absorption coefficient for the X-rays from element A are required for the Be window, the Au (or other) contact layer, the Si dead layer, and the Si intrinsic region (thickness t'). We can also write a similar expression for an IG detector. These various detector parameters each have density ρ (available from standard elemental density tables) and thickness t. Typical values of t for each part of the detector were discussed in Chapter 32. The first three terms thus account for absorption of weak X-rays passing through the Be window, the Au contact layer, and the Si dead layer before entering the detector. The last term adjusts the k factor for X-rays that *do not* deposit their energy in the active region of the detector which has density ρ and thickness t'.

While equations 35.23 and 35.24 look 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 1976).
- Interpolate values of Q to give the best fit to experimental k factors (Williams *et al.* 1984).

Table 35.3a. Calculated k_{AFe} Factors for K Lines Using Different Theoretical Cross Sections[a]

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

The other major variable in equation 35.24 is the Be window thickness, which is nominally 7.5 μm but in practice may be substantially thicker. Tables 35.3a and b list calculated k factors obtained using various expressions for Q. As you can see, the value of k may easily vary by >±10%, particularly for the lighter elements and the heavier elements. This variation is due to the uncertainties in the detector-efficiency terms in equation 35.22. 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. This point again emphasizes the advantages of K-line analysis where possible. When the heavy elements are being studied, 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 35.3b. Calculated k_{AFe} Factors for L Lines Using Different Theoretical Cross Sections[a]

Element	k_{MM}	k_{P}	k_{HP}	k_{SW}	k_{Z}
Sr[a]	1.73	1.33	1.32	1.64	1.39
Zr[a]	1.62	1.26	1.24	1.51	1.33
Nb[a]	1.54	1.21	1.18	1.43	1.28
Ag[a]	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.o5	3.34	4.97

[a]k factors use the L intensity from the L_{α} and L_{β} lines. MM = Mott–Massey; GC = Green–Cosslett; P = Powell; BP = Brown–Powell; SW = Schreiber–Wims; Z = Zaluzec.

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 will have predetermined values of all the terms in equations 35.21 and 35.22 stored in its memory. You don't usually have control over which particular parameters are being used. However, you should at least ask the manufacturer to list the sources of the values of Q, ω, and a in the computer. You should then carry out a cross-check calculation with a known specimen to ensure that the calculated k factor gives the correct answer.

If you replace or service a detector, which is not an unusual occurrence on an AEM, then the new detector parameters must be inserted into the software.

We cannot recommend a "best" set of values for Q, ω and a, but the values of Q given by Powell (1976), ω by Bambynek *et al.* (1972), and a by Schreiber and Wims (1982) have been used in the past. Also, we can't give you specific detector parameters, so you should obtain an estimate from the manufacturer. The values of μ/ρ which we recommend are those determined by Heinrich (1986), although there is still considerable uncertainty in the mass absorption coefficients for the low-energy X-rays from the light elements. If

you use the DTSA program from NIST (see Section 1.5), you may find that it predicts a worse value.

> Remember that the problem is that all software packages use preset values in their calculations and these may vary from package to package.

Figures 35.7A and B show 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 35.21. 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 nonexistent, but data for the K lines from the heavier elements will become more common if IG detectors are more widely used.

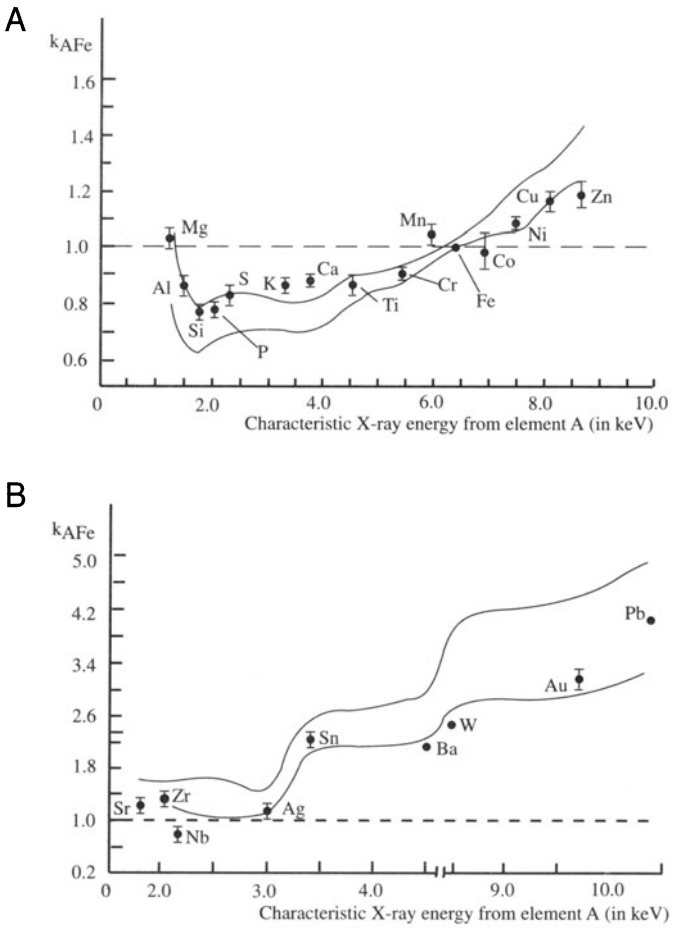


Figure 35.7. (A) Experimental k_{AFe} factors for the K_α X-rays from a range of elements A with respect to Fe. The solid lines represent the spread of calculated k factors using different values for the ionization cross section. (B) Similar data for L_α lines from relatively high-Z elements. The errors in the calculated values of k are large, reflecting the uncertainties in L-line ionization cross sections.

We can summarize the k-factor approach to microanalysis 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.

The point at which the simple Cliff–Lorimer approach breaks down is when the thin-foil criterion is invalid. Absorption is far more common than fluorescence in thin foils. You must be wary of absorption when you have X-ray lines in your spectrum that differ in energy by >5–10 keV, particularly if any are light-element X-rays. To understand why this is so, we must investigate the absorption correction factor.

35.5. ABSORPTION CORRECTION

Preferential absorption of the X-rays from one of the elements in your specimen means that the detected X-ray intensity will be less than the generated intensity and so C_A is no longer simply proportional to I_A. So you have to modify your k factor to take into account the reduction in I_A. This is a problem if your specimen is too thick, or if one or more of the characteristic X-rays has an energy less than ~1 keV (i.e., light element analysis). 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

$$k_{AB}^* = k_{AB}(ACF) \qquad [35.25]$$

The absorption correction factor (ACF) is the A term in equation 35.20 and this can be written out fully, incorporating the expression for χ from equation 35.15, to give

$$\mathrm{ACF} = \frac{\int_0^{\rho t} \left\{ \varphi_B(\rho t)\, e^{-\left(\frac{\mu}{\rho}\right]_{Spec}^{B} \rho t \operatorname{cosec}\alpha} \right\} d(\rho t)}{\int_0^{\rho t} \left\{ \varphi_A(\rho t)\, e^{-\left(\frac{\mu}{\rho}\right]_{Spec}^{A} \rho t \operatorname{cosec}\alpha} \right\} d(\rho t)} \qquad [35.26]$$

In this expression $\mu/\rho]^A_{Spec}$ is the mass absorption coefficient of X-rays from element *A* in the specimen, α is the detector take-off angle, ρ is the density of the specimen, and *t* is the thickness. Since the units of μ/ρ are usually cm^2/gm, be sure to use ρ in gm/cm^3 and t in cm, rather than SI units. Obviously, the value of the ACF is unity when no absorption occurs. Typically, if the ACF is >10% then the absorption is significant, since 10% accuracy is routinely attainable in quantitative microanalysis using experimental *k* factors. However, accuracy better than 10% can be obtained if you decide what constitutes a "significant" level of absorption for the problem at hand and the accuracy required of the data. 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 (1986). 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

$$\left.\frac{\mu}{\rho}\right]^A_{Spec} = \sum_i \left(\left.\frac{C_i\mu}{\rho}\right]^A_i \right) \qquad [35.27]$$

where C_i is the fractional concentration of element *i* in the specimen such that

$$\sum_i C_i = 1 \qquad [35.28]$$

The absorption of X-rays from element *A* by all elements *i* in the specimen is summed, including self absorption by element *A* itself, absorption by elements that may not be of interest in the experiment, and even by materials whose X-rays might not be detectable.

An 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 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 (Bender *et al.* 1980).

In equation 35.26, the depth distribution of X-ray production $\varphi(\rho t)$ is assumed to be 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*. Depending on the thickness of the foil, it is possible that this assumption may be the limiting factor in the accuracy of the absorption correction, but for most thin foils, particularly if *Z* is < 30, variations in $\varphi(\rho t)$ can be ignored. If $\varphi(\rho t)$ does affect the absorption correction, then it will result in a slight overcompensation for the effects of absorption, which will get worse as the thickness increases.

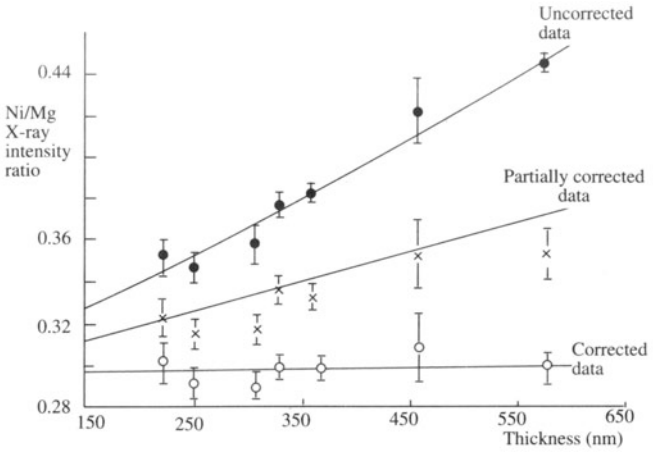


Figure 35.8. The upper curve shows the raw Ni K_α/Mg K_α intensity ratio as a function of thickness in a homogeneous sample of NiO-MgO. The slope indicates strong absorption of Mg K_α X-rays. The middle curve shows the effect of correcting for absorption of the Mg K_α line by Ni and the bottom line shows the effect of a further correction for absorption of the Mg K_α line by O to give the expected horizontal line.

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 about 5% in foil thicknesses of <300 nm. Therefore, the assumption that $\varphi(\rho t)$ equals unity does indeed appear reasonable. 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 35.26 to give

$$ACF = \left(\frac{\left.\frac{\mu}{\rho}\right]^A_{Spec}}{\left.\frac{\mu}{\rho}\right]^B_{Spec}} \right) \left(\frac{1 - e^{-\left(\left.\frac{\mu}{\rho}\right]^B_{Spec} \rho t \operatorname{cosec} \alpha \right)}}{1 - e^{-\left(\left.\frac{\mu}{\rho}\right]^A_{Spec} \rho t \operatorname{cosec} \alpha \right)}} \right) \qquad [35.29]$$

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 convergent-beam electron diffraction

$$\rho = \frac{n A}{V N} \qquad [35.30]$$

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

$$t' = t \operatorname{cosec} \alpha \qquad [35.31]$$

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 36.6. No method is universally applicable, and few are either easy or accurate. The value of α with the specimen at 0° tilt is fixed by the design geometry of the stage and the only way to vary α is by tilting the specimen. As we have seen, there are good reasons not to tilt the specimen beyond about 10°, because of the increase in spurious X-rays, but if there is a severe absorption problem, then decreasing t' by tilting the specimen toward the detector is a sensible first step toward minimizing the problem.

In some AEMs it is necessary to tilt the specimen toward the detector before any X-rays can be detected. Matters get even more complicated if your detector axis is not orthogonal to the tilt axis. Such a design is very poor from an analytical standpoint but, even under these conditions, the geometry is relatively straightforward and Zaluzec *et al.* (1981) have listed all the necessary equations.

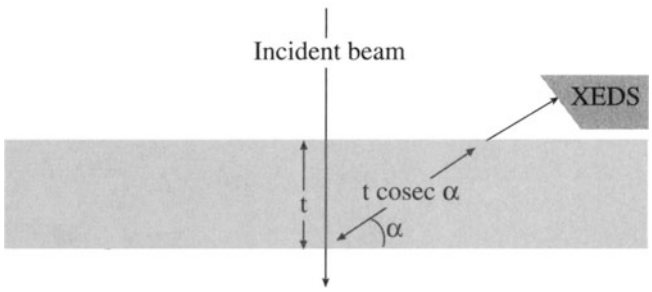


Figure 35.9. Relationship between the specimen thickness t and the absorption path length t cosecα for a take-off angle α.

Fluorescence is usually a minor effect and often occurs for X-rays that are not of interest.

So far we've assumed that the specimen is parallel-sided, but this is uncommon. Most 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 mentioned in Figure 33.3. 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, a method has been developed to correct for absorption without measuring t, as we discuss in the next section.

Because the sample density, ρ, and the values of μ/ρ vary with the composition of the specimen (see equations 35.22 and 35.23), 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, you perform a first iteration calculation of μ/ρ and ρ, and generate modified values of C_A and C_B, and so on. 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 nm to 80 nm. This change is still quite small and within the limits of all but the most accurate microanalyses. 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.

It should always be remembered that large errors will only occur in strongly absorbing systems and/or very thick specimens.

35.6. EXTRAPOLATION TECHNIQUES FOR ABSORPTION CORRECTION

A different approach to the absorption problem has been developed by Horita *et al.* (1987) and Van Cappellen (1990) which neatly avoids the problems of measuring the thickness at each analysis point, but does require that you measure the beam current. This is the way you should proceed with the absorption correction if it is at all possible.

You still need to know μ/ρ, ρ, and α, but not t. This approach uses a simplified correction factor

$$ACF' \simeq e^{-\left(\left[\frac{\mu}{\rho}\right]^{B}_{Spec} - \left[\frac{\mu}{\rho}\right]^{A}_{Spec} \frac{\rho t}{2} \operatorname{cosec} \alpha\right)} \quad [35.32]$$

which assumes that all X-rays are generated at $t/2$, ignoring $\varphi(\rho t)$ effects, and requires that the X-rays from one of the two elements are not absorbed. Applying the ACF′ to the measured intensity ratios, we can show that, if we measure k_{AB} over a range of thicknesses, we can extrapolate to $(k_{AB})_0$ at $t = 0$ to give

$$\log_{10}(k_{AB}) = \log_{10}(k_{AB})_0 + \frac{\Delta_{AB}}{\varphi} I_x \quad [35.33]$$

where Δ_{AB} is related to the difference in μ/ρ for X-rays from elements A and B:

$$\Delta_{AB} = 0.217\left(\left[\frac{\mu}{\rho}\right]^{A}_{Spec} - \left[\frac{\mu}{\rho}\right]^{B}_{Spec}\right)\rho \operatorname{cosec}(\alpha) \quad [35.34]$$

and

$$\varphi = C_A\left(\frac{Q\,\omega\,a}{A}\right)_A i_A \quad [35.35]$$

for element A, where all the terms are described in equation 35.16, except for the electron probe current i_A, which is assumed constant.

So to apply this method, you need to keep the beam current and X-ray acquisition time constant, and the specimen must contain one X-ray that shows negligible absorption, as shown in Figure 35.10.

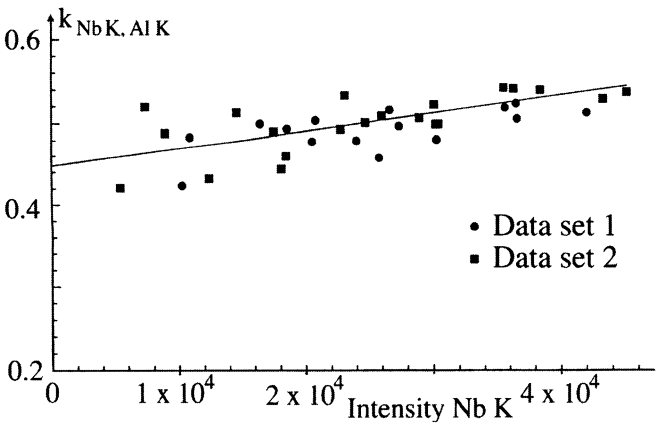


Figure 35.10. A plot of two independent sets of k-factor data for a Nb-Al alloy at 300 kV, showing the variation of the effective k factor with thickness as indicated by the Nb K_α X-ray intensity. The Al K_α X-rays which are absorbed give increasing effective k factors with thickness. X-rays for which absorption is insignificant would give a constant k factor with thickness.

The method can be extended to the microanalysis of unknown specimens by using the extrapolation method to determine the absorption-free intensity ratio at zero thickness, and using this ratio in combination with the k_{AB} factor at zero thickness to give a value of C_A/C_B. This k factor is then applied to the calculation of the composition using the intensity ratio measured at the same thickness.

Obviously, it is very time-consuming to do the full absorption correction as accurately as we would like, because of all the uncertainties. You need an absorption correction if you are dealing with light elements, with K lines <1 keV. Under these circumstances, the extrapolation technique of Horita *et al.* is the best approach; a detailed example of the determination of light-element k factors using this method has been given by Westwood *et al.* (1992). Further refinements of Horita's method have been proposed by Eibl (1993). In the specific case of ionic compounds in which electroneutrality must be maintained (i.e., the sum of all anions and cations, times their valence states, must balance), it is even possible to devise an absorption correction with no estimate of t (Van Cappellen and Doukhan 1994).

35.7. THE FLUORESCENCE CORRECTION

X-ray absorption and fluorescence are intimately related because the 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 microanalysis anyhow, since the higher-energy Ni K X-rays are not absorbed or fluoresced.

Fluorescence is usually a minor effect and often occurs for X-rays that are not of interest.

However, in the rare case that fluorescence occurs to a degree that limits the accuracy of microanalysis, the equation used for the fluorescence correction factor (FCF) is that developed by Nockolds *et al.* (1980); a detailed discussion is given by Anderson *et al.* (1995). Practical examples of the fluorescence correction are hard to come by and the classic case is Cr in stainless steels, where the Cr K_α line is fluoresced by the major peak, the Fe K_α line, giving rise to an increase in apparent Cr content as the foil gets thicker. You can also avoid the problem of thickness measurement for fluorescence corrections, just as for absorption, using a similar parameterless correction (Van Cappellen 1990).

35.8. ALCHEMI

We told you early on in this chapter to take your X-ray spectra away from strong diffraction conditions. This is because of the "Borrmann effect." As we saw back in Sections 13.8, 13.9, and 14.6, 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, as shown in Figure 35.11A. Now 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 many crystals. The technique was first developed for the TEM by Spence and Taftø (1983), who coined the acronym. The derivation of the quantitative expressions that we give below follows that paper. Channeling is widely used for atom site location in other analysis techniques (e.g., see Chu *et al.* 1978).

The way to do ALCHEMI experimentally is to acquire a spectrum under strong channeling conditions, such that the Bloch wave is interacting strongly with a particular systematic row of atoms. This channeling orientation should be chosen so that the planes interacting strongly with the electron beam also contain the candidate impurity atom sites, so you must 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 intensity from the atoms in that

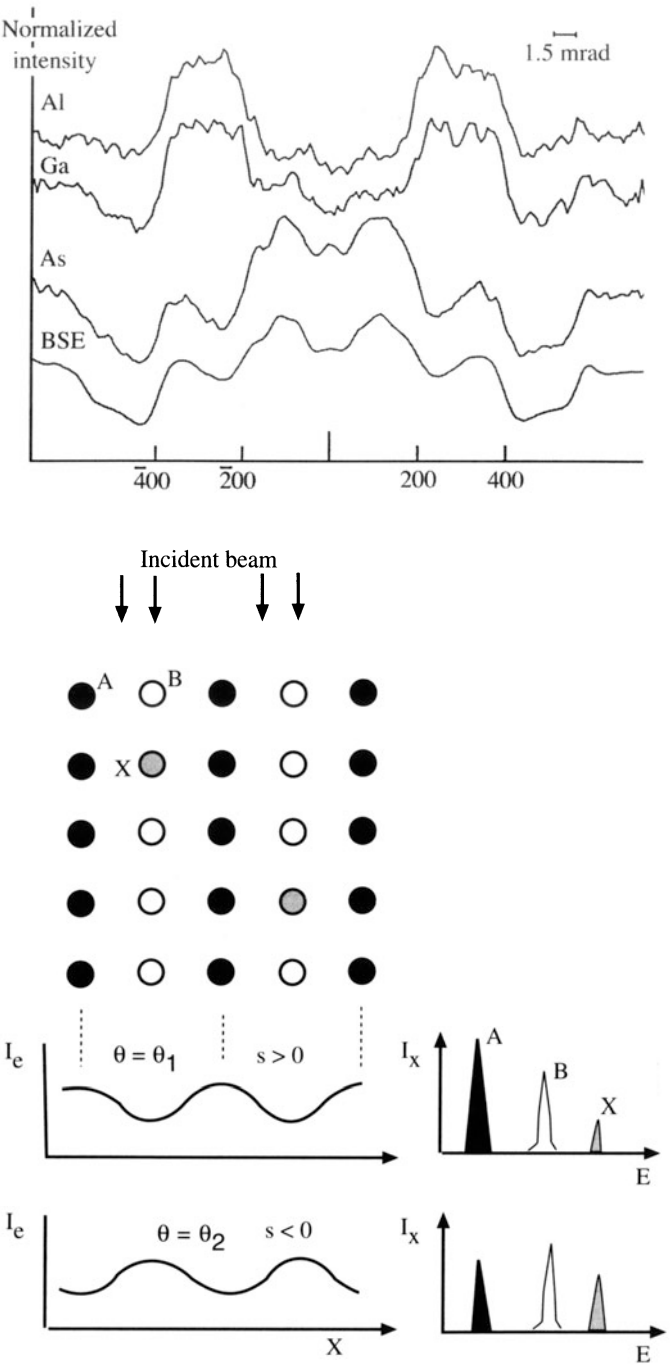


Figure 35.11. (A) The Borrmann effect: the variation in characteristic X-ray emission close to strong two-beam conditions as the beam is rocked across the 400 planes of GaAlAs. The X-rays from Al, which occupies Ga sites, follow the Ga X-ray emission while the As varies in an approximately complementary fashion. The backscattered electron signal (BSE) is inversely proportional to the amount of electron channeling, so the As signal is strongest where the channeling is weakest. (B) ALCHEMI allows the determination of the site occupancy of atom *X* in columns of atoms *A* and *B*. By tilting to $\mathbf{s} > 0$ and $\mathbf{s} < 0$, the Bloch waves interact strongly with row *A* and 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.

plane will be highest. Start by finding the orientations that give the most pronounced channeling effects for the atoms *A* and *B*, as shown schematically in Figure 35.11B. Usually a very small tilt is all that is necessary to get a different spectrum.

If you are looking at two elements, *A* and *B*, and a substitutional element *X*, follow this procedure:

- Measure X-ray intensities from each element in orientations 1 and 2.
- Then find a nonchanneling orientation (3) where electron intensity is uniform for both planes.

In this orientation we define the ratio *k* as

$$k = \frac{I_B}{I_A} \qquad [35.36]$$

where I_B is the number of X-ray counts from the element *B* in the nonchanneling orientation. For the two channeling orientations 1 and 2, we define two parameters β and γ such that

$$\beta = \frac{I_B^{(1)}}{k\, I_A^{(1)}} \qquad [35.37]$$

and

$$\gamma = \frac{I_B^{(2)}}{k\, I_A^{(2)}} \qquad [35.38]$$

Now assuming we know that the element *X* sits on specific sites, say it substitutes for atom *B*, then we define an intensity ratio term *R* such that

$$R = \frac{I_A^{(1)}\, I_X^{(2)}}{I_X^{(1)}\, I_A^{(2)}} \qquad [35.39]$$

Hence the fraction of atom *X* on *B* sites is given by

$$C_X = \frac{R - 1}{R - 1 + \gamma - \beta R} \qquad [35.40]$$

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. This makes it 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.

35.9. EXAMPLES; PROFILES AND MAPS

The best way to appreciate the value of quantitative analysis is to go and study some applications. In Figure 35.12, composition data from a complex three-component Ni-Cr-Mo high-temperature superalloy are plotted to reveal a section of the ternary phase diagram (Raghavan *et al.* 1984).

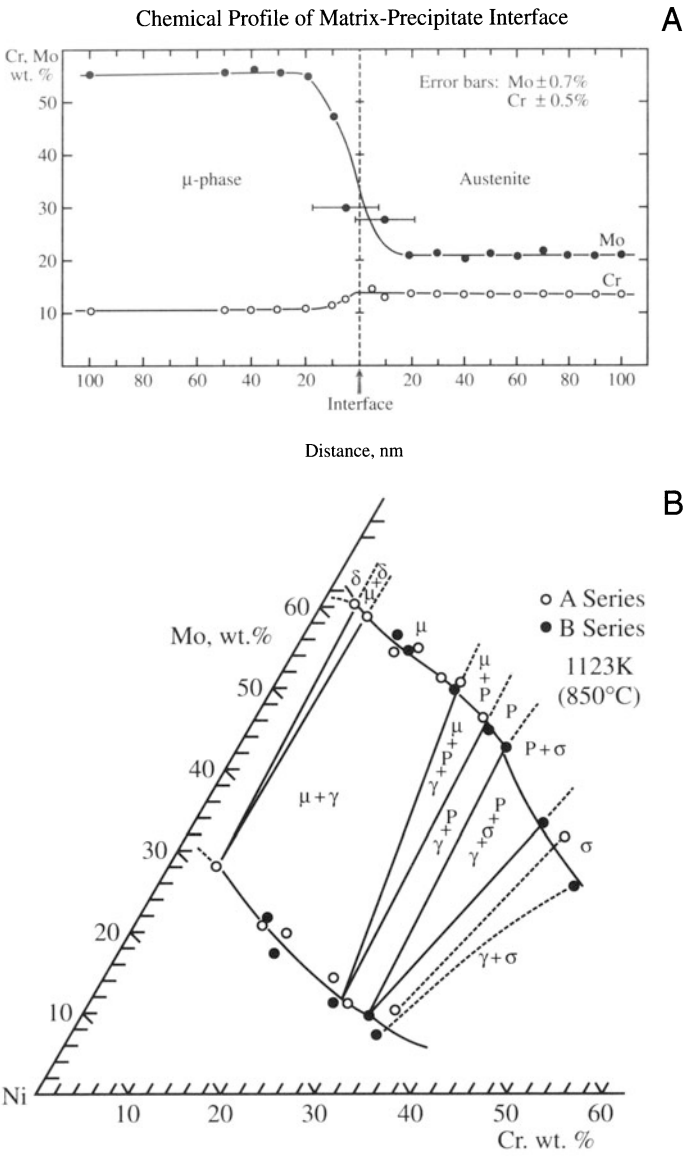


Figure 35.12. (A) Cr and Mo composition profiles across a two-phase μ–γ interface in a Ni-10Cr-30Mo alloy which has been aged 1000 hr at 1123 K. The profiles show composition changes that define tie lines in the ternary phase diagram. (B) A corner of the Ni-Cr-Mo ternary phase diagram determined by XEDS microanalysis of thin foils of heat-treated specimens containing up to three phases. The limits of the undesirable σ–phase regions are the important phase boundaries in this material.

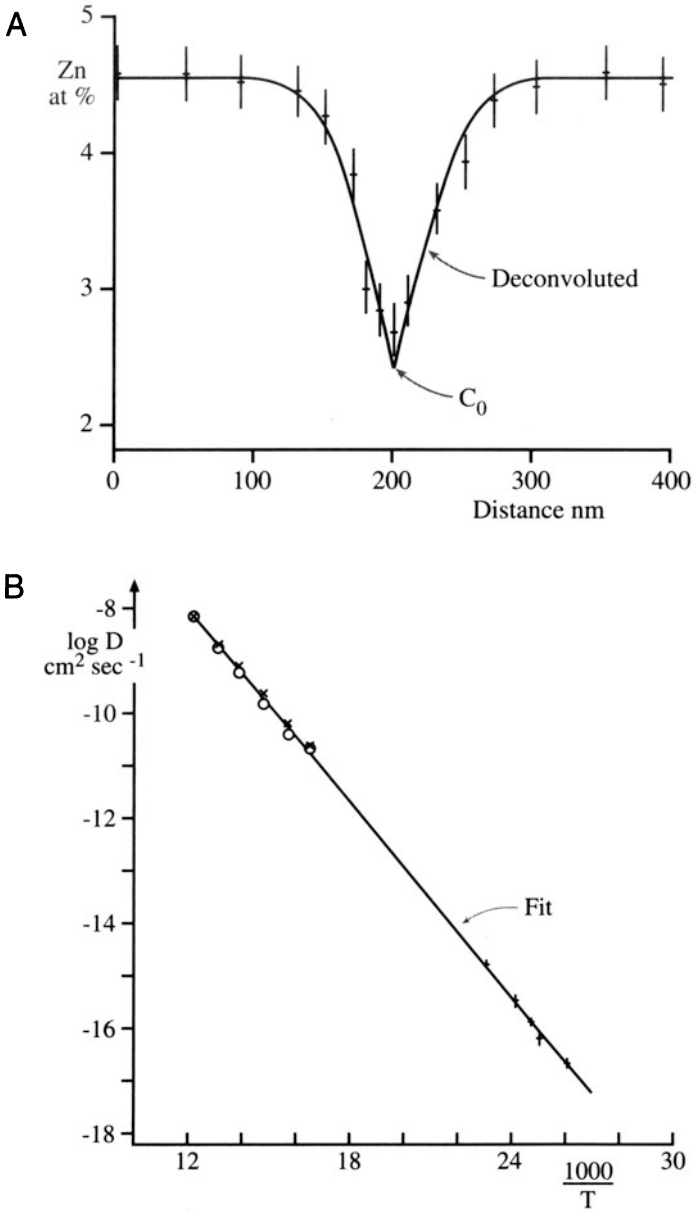


Figure 35.13. (A) Zn concentration profile across a grain boundary in an Al-4.5 at.% Zn alloy aged at 125°C to produce a solute-depleted region due to equilibrium grain boundary precipitation. Grube analysis of the profiles give a measure of D. The low aging temperatures produce such small profiles that AEM is the only way to measure them. (B) Arrhenius plot of the diffusivity of Zn in Al, as a function of temperature, derived from measurements of Zn composition profiles in (A). Extrapolation of high-temperature diffusion data match up well with the AEM results.

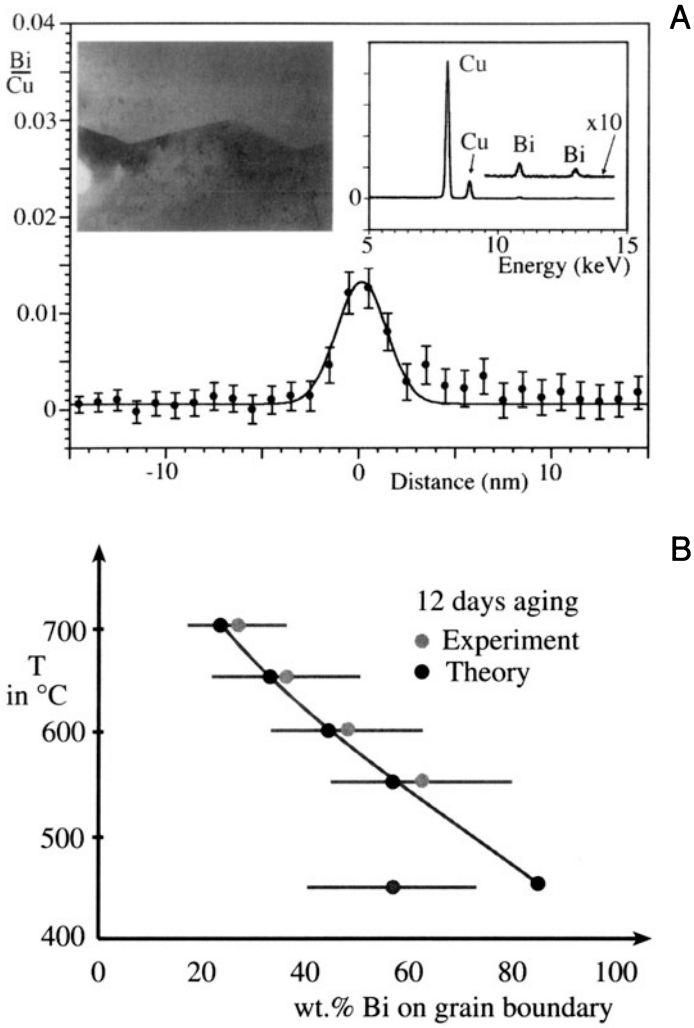


Figure 35.14. (A) XEDS measurements of the distribution of Bi segregated to a grain boundary in Cu. A faceted Cu GB, typical of those to which Bi is segregated, is shown in the upper-left inset and an XEDS spectrum from the GB region in the upper right reveals small Bi peaks. (B) Quantification of similar Bi spectra to that in (A) shows an inverse relationship with temperature, consistent with classical McLean adsorption isotherm predictions, shown as the fitted line.

Assuming interface equilibrium, sections such as the three-phase triangle can be measured from a single thin foil. In this study the undesirable σ–phase boundaries were sought, to avoid embrittlement of the alloy in service. In Figure 35.13, solute profiles measured across grain boundaries in Al-Zn (Nicholls and Jones 1983) permitted measurement of the diffusion coefficient of Zn in Al to much lower temperatures than previously attained with traditional EPMA methods. In Figure 35.14, the detection of Bi equilibrium segregation to grain boundaries in Cu is modeled by a simple McLean-type adsorption isotherm (Michael and Williams 1984). Previous studies of such Gibbsian segregation required *in situ* fracture of embrittled Cu inside Auger spectrometer systems.

We should also note that individual point analysis, or profiles across an interface, are not the only way to display X-ray data. It is possible to produce X-ray images or maps in which the intensity of the signal in the map is directly related to the X-ray intensity I_A. In a quantitative

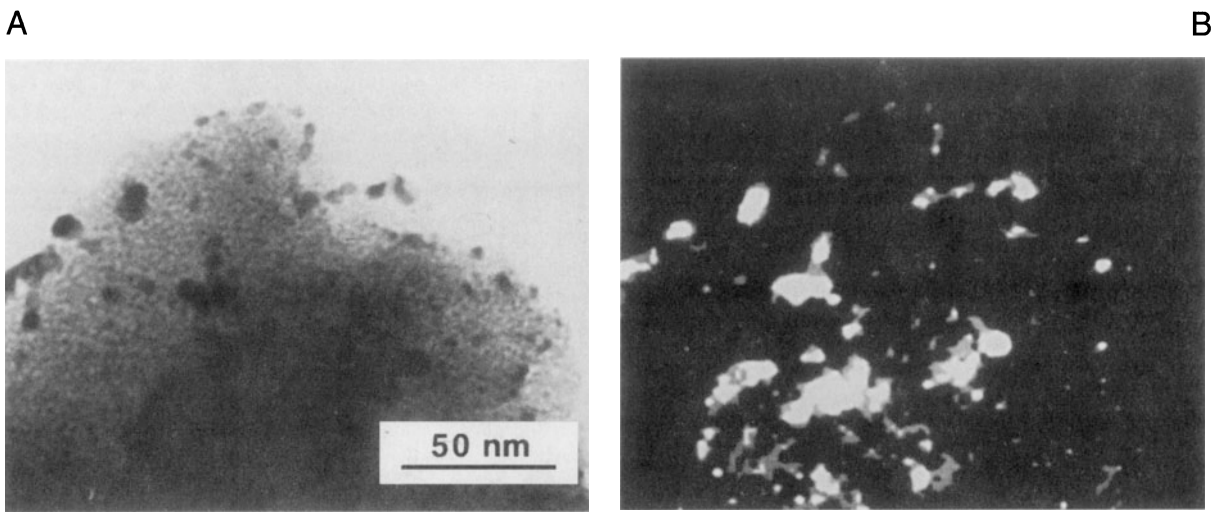


Figure 35.15. (A) BF STEM image of PD catalyst particles on a carbon support film. (B) Qualitative Pd L_α X-ray image of the distribution of Pd.

map, the X-ray signal is proportional to the concentration C_A. While there are obvious advantages to comparing quantitative maps of elemental distributions with other TEM images, this process is limited by the relatively poor statistics of X-ray acquisition. Remember that good quantification requires ~10,000 counts for I_A. Even in an efficient AEM, this intensity may easily take one minute to acquire. At this acquisition rate, even a 56 × 56 pixel image will take 50 hours to gather, so it is impractical. We need to increase the efficiency of X-ray acquisition markedly, or we just have to make do with qualitative, noisy maps, as shown in Figure 35.15, even when using an FEG-AEM.

CHAPTER SUMMARY

Quantitative microanalysis of spectra from thin foils is straightforward in most cases, so long as you take care to determine the *k* factors with sufficient accuracy. The software to handle the more difficult problem of absorption is well known and commercially available. Perhaps the greatest difficulty remains the need to know the specimen thickness in order to compensate for X-ray absorption, and extrapolation techniques are invaluable in avoiding this. We can minimize absorption by making the thinnest possible specimens, but then the possibility arises that the number of X-ray counts may be so small that errors in the quantification are large. The use of FEG sources and improved TEM-EDS configurations to maximize the collection angle will help in this situation.

REFERENCES

General References

Goldstein, J.I. (1979) in *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 83, Plenum Press, New York.

Goldstein, J.I., Williams, D.B., and Cliff, G. (1986) in *Principles of Analytical Electron Microscopy* (Eds. D.C. Joy, A.D. Romig, and J.I. Goldstein), p. 155, Plenum Press, New York.

Romig A.D. Jr. (1986) *Analytical Transmission Electron Microscopy,* Metals Handbook, 9th edition, **10**, p. 429, American Society for Metals, Metals Park, Ohio.

Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd edition, Philips Electron Optics Publishing Group, Mahwah, New Jersey.

Williams D.B. and Goldstein J.I. (1991) in *Electron Probe Quantitation* (Eds. K.F.J. Heinrich and D.E. Newbury), p. 371, Plenum Press, New York.

Zaluzec, N.J. (1979) in *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 121, Plenum Press, New York.

Specific References

Anderson, I.M., Bentley, J., and Carter, C.B. (1995) *J. Microsc.* **178**, 226.

Bambynek, W., Crasemann, B., Fink, R.W., Freund, H.U., Mark, H., Swift, C.D., Price, R.E., and Rao, P.V. (1972) *Rev. Mod. Phys.* **44**, 716.

Bender, B.A., Williams, D.B., and Notis, M.R. (1980) *J. Am. Ceram. Soc.* **63**, 149.

Castaing, R. (1951) Thesis, University of Paris, ONERA Publication, #55.

Chapman, J.N., Nicholson, W.A.P., and Crozier, P.A. (1984), *J. Microsc.* **136**, 179.

Chu, W.-K., Mayer, J.M., and Nicolet, M.-A. (1978) *Backscattering Spectrometry,* Academic Press, Orlando, Florida.

Cliff, G. and Lorimer, G.W. (1975) *J. Microsc.* **103**, 203.

Eibl, O. (1993) *Ultramicrosopy* **50**, 179.

Goldstein, J.I., Costley, J.L., Lorimer, G.W., and Reed, S.J.B. (1977) *SEM 1977,* **1** (Ed. O. Johari), p. 315, IITRI, Chicago, Illinois.

Graham, R.J. and Steeds, J.W. (1984) *J. Microsc.* **133**, 275.

Heinrich, K.F.J. (1986) in *Proc. ICXOM-11* (Eds. J. Brown and R. Packwood), p. 67, University of Western Ontario, Canada.

Heinrich, K.F.J. and Newbury, D.E., Eds. (1991) *Electron Microprobe Quantitation,* Plenum Press, New York.

Hillier, J. and Baker, R.F. (1944) *J. Appl. Phys.* **15**, 663.

Horita, Z., Sano, T., and Nemoto, M. (1987) *Ultramicroscopy* **21,** 271.

Kramers, H.A. (1923) *Phil. Mag.* **46**, 836.

Lorimer, G.W., Al-Salman, S.A., and Cliff, G. (1977) in *Developments in Electron Microscopy and Analysis* (Ed. D.L. Misell), p. 369, The Institute of Physics, Bristol and London.

McGill, R.H. and Hubbard, F.H. (1981) in *Quantitative Microanalysis with High Spatial Resolutions* (Eds. G.W. Lorimer, M.H. Jacobs, and P. Doig), p. 30, The Metals Society, London.

Michael, J.R. and Williams, D.B. (1984) *Met. Trans.* **15A**, 99.

Nicholls, A.W. and Jones, I.P. (1983) *J. Phys. Chem. Solids* **44**, 671.

Nockolds, C., Nasir, M.J., Cliff, G., and Lorimer, G.W. (1980) in *Electron Microscopy and Analysis-1979* (Ed. T. Mulvey), p. 417, The Institute of Physics, Bristol and London.

Owen, D.B. (1962) *Handbook of Statistical Tables,* Addison-Wesley, Reading, Massachusetts.

Powell, C.J. (1976) in *Use of Monte Carlo Calculations in Electron Probe Microanalysis and Scanning Electron Microscopy* (Eds. K.F.J. Heinrich, D.E. Newbury, and H. Yakowitz), p. 61, U.S. Department of Commerce/NBS, Washington, DC.

Raghavan, M., Mueller, R.R., Vaughn, G.A., and Floreen, S. (1984) *Met. Trans.* **15A**, 783.

Schreiber, T.P. and Wims, A.M. (1981) *Ultramicroscopy* **6**, 323.

Schreiber, T.P. and Wims, A.M. (1982) *X-ray Spectrometry* **11**, 42.

Sheridan, P.J. (1989) *J. Electr. Microsc. Tech.* **11**, 41.

Spence, J.C.H. and Taftø, J. (1983) *J. Microsc.* **130**, 147

Sprys, J.W. and Short, M.A. (1976) *Proc. 34th EMSA Meeting* (Ed. G.W. Bailey), Claitors, Baton Rouge, Louisiana.

Van Cappellen, E. (1990) *Microsc. Microanal. Microstruct.* **1**, 1.

Van Cappellen, E. and Doukhan, J.C. (1994) *Ultramicroscopy* **53**, 343.

Westwood, A.D., Michael, J.R., and Notis, M.R. (1992) *J. Microsc.* **167**, 287.

Williams, D.B. and Goldstein, J.I. (1991) in *Electron Probe Quantitation* (Eds. K.F.J. Heinrich and D.E. Newbury), p. 371, Plenum Press, New York.

Williams, D.B., Newbury, D.E., Goldstein, J.I., and Fiori, C.E. (1984) *J. Microsc.* **136**, 209.

Wood, J.E., Williams, D.B., and Goldstein, J.I. (1981) in *Quantitative Microanalysis with High Spatial Resolutions* (Eds. G.W. Lorimer, M.H. Jacobs, and P. Doig), p. 24, The Metals Society, London.

Wood, J.E., Williams, D.B., and Goldstein, J.I. (1984) *J. Microsc.* **133**, 255.

Zaluzec, N.J., Maher, D.M., and Mochel, P.E. (1981) in *Analytical Electron Microscopy-1981* (Ed. R.H. Geiss), p. 25, San Francisco Press, San Francisco, California.

Spatial Resolution and Minimum Detectability

36

CHAPTER PREVIEW

Often, when you do X-ray microanalysis of thin foils, you are seeking information that is close to the limits of spatial resolution. Before you carry out any such microanalysis you need to understand the various controlling factors, which we explain in this chapter. 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.

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 we find it very difficult to detect the presence of trace constituents in thin foils. Consequently, the minimum mass fraction (MMF) in AEM is not very small compared with other analytical instruments which have poorer spatial resolution. This trade-off is true for any microanalysis technique, and so it is only sensible to discuss the ideas of spatial resolution in conjunction with analytical detectability limits. We'll make this connection in the latter part of the chapter. Despite the relatively poor MMF we can detect the presence of just a few atoms of one particular element if the analyzed volume is small enough, and so the AEM actually exhibits excellent minimum detectable mass (MDM) characteristics.

Spatial Resolution and Minimum Detectability

36

36.1. WHY IS SPATIAL RESOLUTION IMPORTANT?

As we described in the introduction to Chapter 35, perhaps the major driving force for the development of X-ray microanalysis in the AEM was the improvement in spatial resolution compared with the EPMA. This improvement arises for two reasons:

- We use thin specimens, so less electron scatter occurs as the beam traverses the specimen.
- The higher electron energy (>100–400 keV in the AEM compared with 5–30 keV in the EPMA) further reduces scatter.

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, as you can see in the Monte Carlo simulations in Figure 36.1A. Conversely, with bulk samples, *increasing* the voltage *increases* the interaction volume, and spatial resolution rarely improves below ~0.5 μm, as shown in Figure 36.1B. So no one is very interested in the theory of spatial resolution of microanalysis for bulk samples and little effort, beyond lowering the kV, is routinely made to optimize this parameter in practice. By contrast, much theoretical and experimental work has been carried out to both define and measure the spatial resolution of XEDS in the AEM, and we'll introduce some of the major ideas here.

36.2. DEFINITION OF SPATIAL RESOLUTION

We can define the spatial resolution of X-ray microanalysis as the smallest distance (R) between two volumes from which independent X-ray microanalyses can be obtained. The definition of R has evolved as AEMs have improved and smaller analysis volumes have become attainable. It has long been recognized that the analysis volume, and hence R, is governed by the beam–specimen interaction volume, since the XEDS can detect X-rays generated anywhere within that volume. The interaction volume is a function of the incident beam diameter (d) and the beam spreading (b) caused by elastic scatter of the beam within the specimen. Therefore, the measured spatial resolution 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.

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 customarily defined as the FWTM of the Gaussian electron intensity. You can measure d directly from the TEM viewing screen, 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 an approximation. Remember that the intensity distribution in the incident beam is Gaussian only if you are careful in your choice and alignment of the C2 aperture. It is a little more diffi-

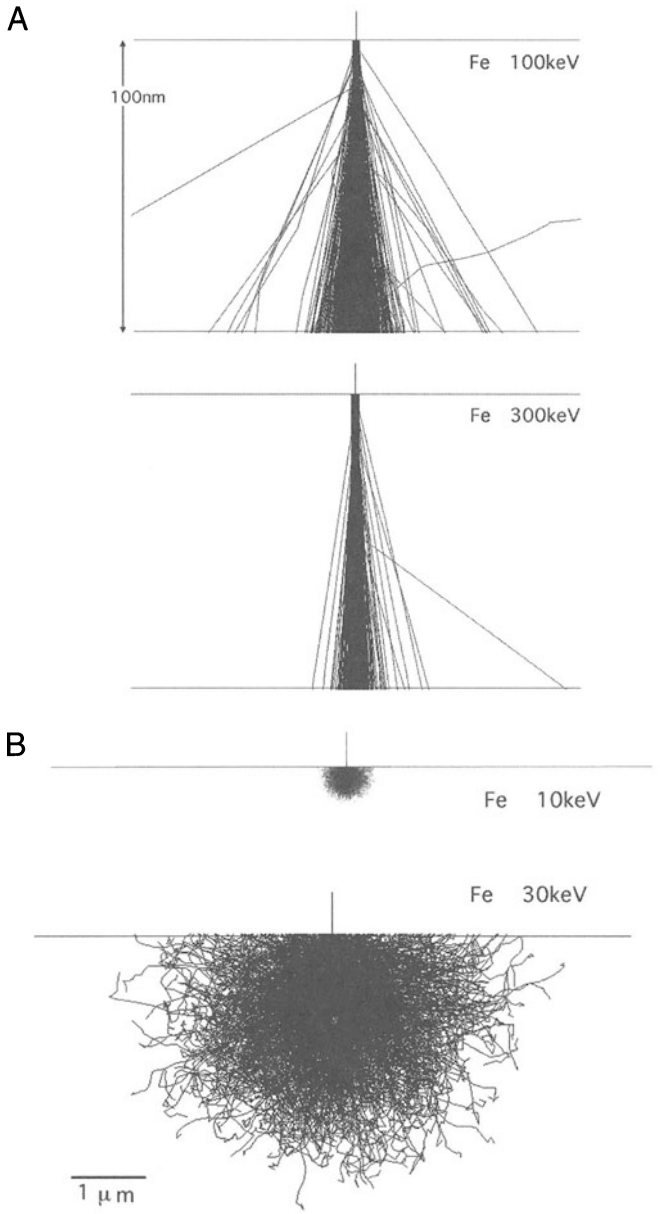


Figure 36.1. (A) Monte Carlo simulations of 10^3 electron trajectories through a 100-nm thin foil of Fe at 100 kV and 300 kV. Note the *improved* spatial resolution at higher kV. (B) Conversely, in a bulk sample the interaction volume at 30 kV is significantly more than at 10 kV, giving *poorer* X-ray spatial resolution at higher kV.

cult to define and measure b, so this needs more explanation.

36.3. BEAM SPREADING

The amount that the beam spreads on its way through the specimen (b) 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 density (ρ). It turns out that the most simple theory for b usually gives a good approximation under most microanalysis 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.* (1977), and refined by Reed (1982). This single-scattering model states

$$b = 7.21 \times 10^5 \frac{Z}{E_0}\left(\frac{\rho}{A}\right)^{\frac{1}{2}} t^{\frac{3}{2}} \qquad [36.1]$$

This well-known expression is *not* in SI units because b and t are in given in cm, ρ is in g/cm^3, and E_0 is in eV.

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 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.

> We recommend that you keep this equation stored in your calculator or in the AEM computer system, so you can quickly estimate the expected beam spreading in your planned experiment.

You should, of course, estimate b *prior* to spending an inordinate amount of time trying to do an experiment which is impossible for lack of spatial resolution.

When we can't apply equation 36.1 (for example, if the specimen geometry or microstructure is complex) the best alternative is the Monte Carlo computer simulation (see Figure 36.1), which we introduced in Section 2.6 as a way of modeling electron scatter. Remember that the Monte Carlo technique uses a random number generator (hence the name) to simulate elastic and inelastic electron–specimen interactions and generate a feasible set of electron paths through a defined specimen. A relatively small number of paths (typically 10^3–10^6) can give a very good measure of the behavior of the very large number of electrons in a typical beam (remember, a 1-nA probe current implies ~10^{10} electrons/second entering the specimen). A full description is beyond the scope of this text and complete books exist on the topic (Heinrich *et al.* 1975, Joy 1995). After simulating 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 elec-

trons. This definition of b is consistent with that described at the start and is the dimension of b given by equation 36.1. In Joy's book, you'll find a code listing for a Monte Carlo simulation program which can be run on a PC. These simulations are now extremely rapid, and in a few minutes on a PC they can provide much of the information you need to estimate the beam spreading in heterogeneous microstructures that are not amenable to simple modeling with the single-scattering approach. Parallel supercomputers have even been used to simulate millions of trajectories in more complex specimens (Michael *et al.* 1993).

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 X-ray source size. Of course this is closely related to the electron distribution, but we can only relate the two directly using Monte Carlo simulations. In these simulations you can easily get the computer to calculate the distribution of X-ray photons generated throughout the specimen, and factor it into any calculations of the composition of the analyzed volume. Monte Carlo simulations are useful for estimating the X-ray spatial resolution because they:

- Incorporate the effects of different kVs and beam diameters.
- Handle difficult specimen geometries and multiphase 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 its parameters, ρ, Z, A, 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 inverse relationship between b and E_0. If you're interested in the details of the various theories you'll find a discussion in Goldstein *et al.* (1986).

36.4. THE SPATIAL RESOLUTION EQUATION

Now that we've defined d and b, all we have to do is combine them to come up with a definition of R. Reed (1982) argued that if the incident beam was Gaussian, and if the beam emerging from the specimen retains a Gaussian intensity distribution, then b and d should be added in quadrature to give a value for R

$$R = \left(b^2 + d^2\right)^{\frac{1}{2}} \qquad [36.2]$$

This equation remained the standard definition of R for almost a decade, despite the fact that no set of experiments ever investigated the effects of all the variables affecting b in equation 36.1. About the same time as this definition of R was proposed, Gaussian beam-broadening models were introduced which were based on equation 36.1 but permitted 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.* (1990) 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 36.2

$$R = \frac{d + R_{\max}}{2} \qquad [36.3]$$

where $R_{\max}$ is given by equation 36.2.

This equation is the formal definition of the X-ray spatial resolution.

Like all definitions of spatial resolution, there is no fundamental justification for the choice of various factors

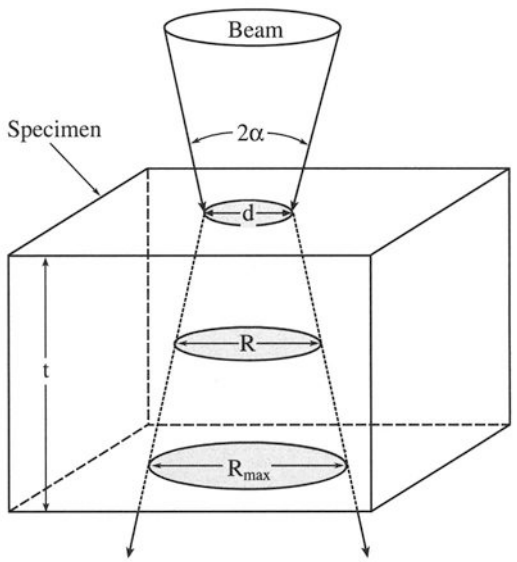


Figure 36.2. Definition of spatial resolution: schematic diagram of how the combination of incident beam size d and beam spreading through the foil combine to define the spatial resolution R of X-ray microanalysis in a thin foil.

such as the FWTM diameter, and the selection of the midplane of the foil at which to define R. Similarly, this approach ignores any contribution of electron diffraction in crystalline specimens, beam tailing beyond the 90% limit, and the effects of fast secondary electrons, which, in some circumstances, can be important. Nevertheless, the definition has been shown to be consistent with experimental results and sophisticated Monte Carlo simulations (Williams *et al.* 1992). Finally, this definition retains the advantage of the original single-scattering model, i.e., it has a simple form, easily amenable to calculation.

36.5. MEASUREMENT OF SPATIAL RESOLUTION

Any theory of spatial resolution must be tested against practical measurements in the TEM if it is to be relevant. Experimental measurements of the spatial resolution appeared slightly before the first theoretical treatments. Composition profiles measured across atomically sharp interphase interfaces were first presented by Lorimer *et al.* (1976). Since then, several other kinds of specimens have been proposed, such as spherical particles in a foil of known thickness, artificial specimens of Au lines deposited on a Si foil, grain boundary films, and quantum well structures, among others.

We believe the first method, using interphase interfaces, retains its validity since there are fewer unknowns than for the other specimens.

If thermodynamic equilibrium exists either side of the interphase interface, the solute content of each phase is well defined. Also, interphase interfaces are common to many engineering materials, as shown back in Figure 35.12.

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 36.3. 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:

$$R = 1.414L \qquad [36.4]$$

Assuming this relationship holds, we just measure the distance L between the 2% and 98% points on the profile, as shown in Figure 36.3. This spread contains 90% of the beam electrons, consistent with our assumption of a 90%

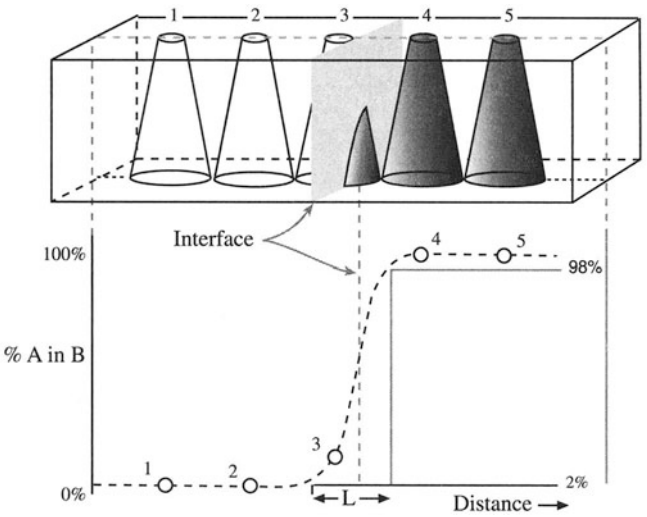


Figure 36.3. Schematic diagram showing the measured composition profile obtained across a planar interface at which an atomically discrete composition change occurs. The spatial resolution can be related to the extent (L) of the measured profile between the 2% and 98% points.

(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.

Note that this definition of R, like the definitions of b and d that we have used, is arbitrary.

Nevertheless, it is easy to remember, relatively easy to measure, consistent with the definitions of b and d, and, most importantly, gives a number that is close to the experimentally measured degradation of the discrete composition change introduced by the beam and the specimen.

Typical measurements of the spatial resolution in two different AEMs are shown in Figures 36.4 and 36.5. Two composition profiles are shown. Each was taken from an Fe-Ni-Cr foil aged to give large Cr composition changes between Cr-rich α-Cr precipitates and the Cr-poor matrix. The specimen was aged sufficiently that the precipitate and matrix are in thermodynamic equilibrium, so that a discrete (atomic level) composition change occurs at the interface. The smooth profiles that appear in the figure are then due to the effects of b and d. In Figure 36.4, the data were obtained from an FEG AEM in which the accelerating voltage was kept constant at 100 kV and the specimen thickness

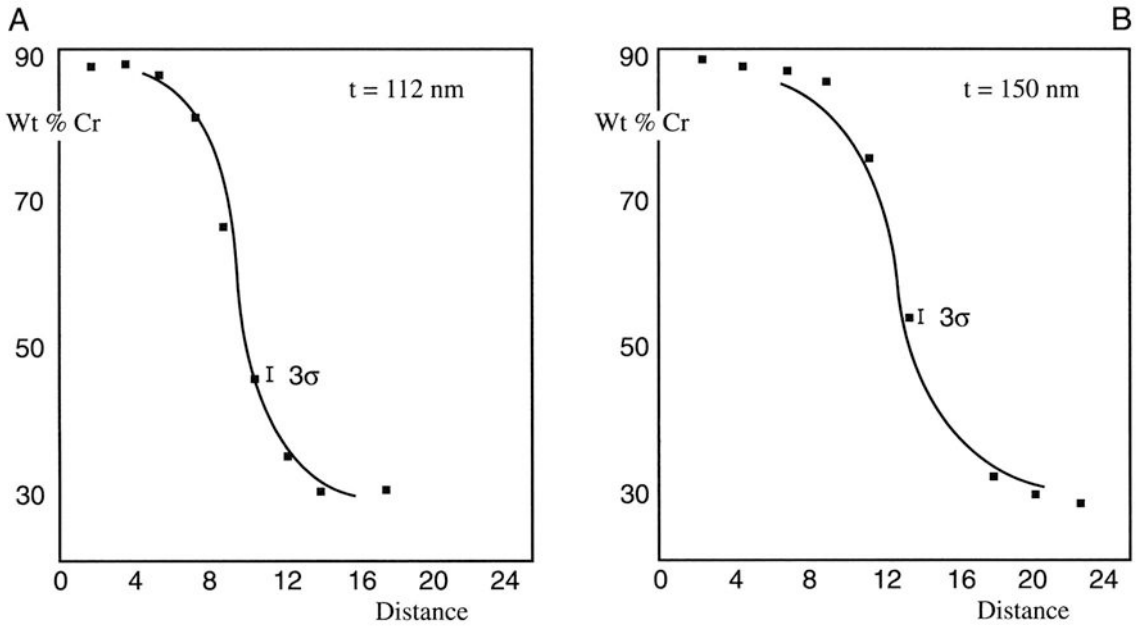


Figure 36.4. Measured Cr composition profiles across the same interface in an Fe-Ni-Cr alloy at two different thicknesses, (A) 112 nm and (B) 150 nm. The solid lines are the fits to the experimental data obtained using a Gaussian convolution model. The profiles were obtained in an FEG AEM at 100 kV and show that the $t^{3/2}$ relationship, assumed in the Gaussian model, applies over the range of thicknesses studied.

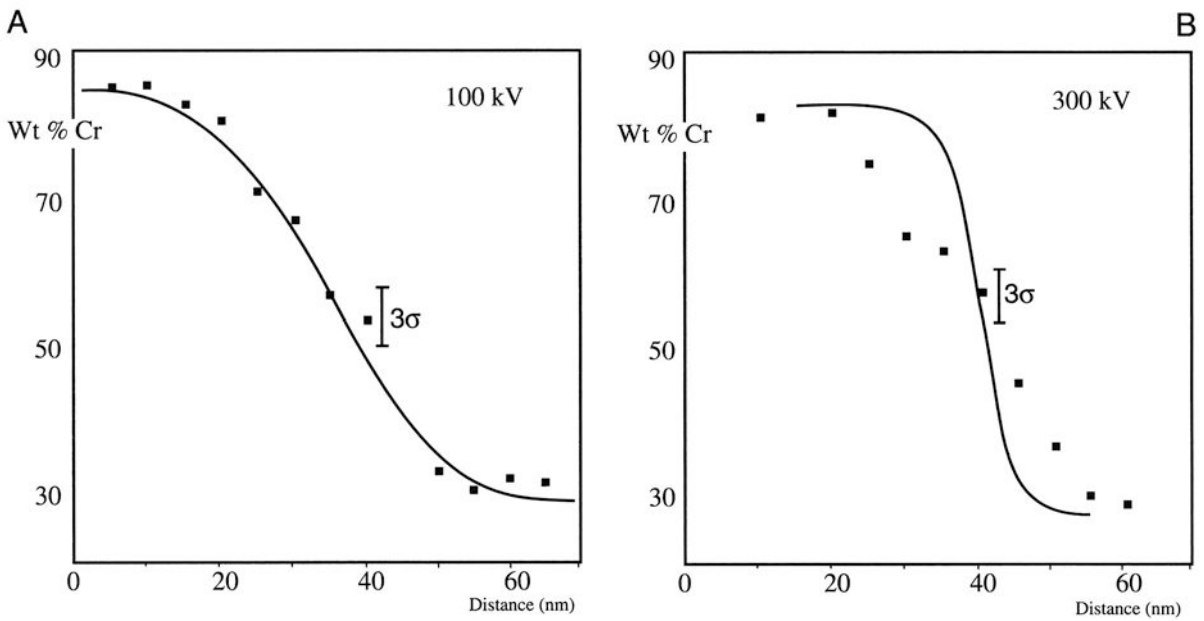


Figure 36.5. Composition profiles from an Fe-Ni-Cr foil, 112 nm thick, obtained with a thermionic source AEM at (A) 100 kV and (B) 300 kV. The solid lines are the fits obtained with a Gaussian convolution model and demonstrate poor spatial resolution, dominated by the large beam size. The bad fit at 300 kV indicates possible specimen drift.

was varied from 112 nm to 150 nm. In Figure 36.5, the data were obtained on an intermediate voltage AEM in which the foil thickness was a constant 120 nm but the voltage was varied between 100 kV and 300 kV. The line drawn through the experimental measurements in each case is derived from the Gaussian model.

It is obvious from equation 36.2 that if we want to improve spatial resolution, then both d and b must be minimized. But if we minimize d, we reduce the input beam current and, for thermionic sources, if d <10 nm, count rates are unacceptably low. However, with an FEG, sufficient current (~1 nA) can be generated in a 1-nm beam to permit quantitative analysis. Comparison of the 100-kV data from the thermionic source instrument (Figure 36.5A) and the FEG source (Figure 36.4A) shows the improvement gained by the use of very small beams in the FEG instrument.

So if you have a thermionic source AEM:

- Your specimen has to be thick enough that sufficient counts are generated for quantification and the net result may be that b is the main contributor to R.
- Alternatively, you may have to increase the beam size such that d dominates rather than b, as in Figure 36.5A (beam diameter 56 nm).

Such a large beam was needed in that example in order to generate sufficient beam current to get a reasonable X-ray count rate at 100 kV. This is one reason why there's been a lot of effort put into developing 300–400 kV AEMs and, more recently, 200–300 kV FEG AEMs.

There are some practical factors which can also limit your experimental spatial resolution:

- Specimen drift and carbon contamination are real problems with side-entry goniometer stages.
- Drift is often exacerbated by the liquid-N_2 cooling required to minimize carbon contamination.
- Changing the kV in intermediate voltage instruments subjects the objective lens cooling coil to large changes in thermal load, which causes drift.

Improvements in image analysis software mean that online drift correction is now available. If you're planning to carry out microanalysis 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. Perhaps one unfortunate side effect of higher voltages is that analysis can be performed in thicker areas than at 100 kV and spatial resolution degrades.

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 of the beam spreading, so thin specimens are essential for the best resolution. FEG sources give sufficient beam current to generate reasonable counts even from very thin specimens and invariably give the best spatial resolution.

36.6. 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 and fluorescence of characteristic X-rays, as we saw in the previous chapter. Furthermore, you should remember that knowledge of t is important in high-resolution phase-contrast imaging and CBED. You'll see in Chapter 39 that in EELS, minimizing t is again critical to obtaining the best results. In almost all TEM techniques your specimen has to be as thin as possible to get the best results; CBED studies are a notable exception 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 remember is that the thickness we are interested in is t, the thickness through which the beam penetrates. This value 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

$$t = \frac{t_0}{\cos \gamma} \qquad [36.5]$$

If your specimen is wedge-shaped, then t and t_0 will vary in an arbitrary fashion depending on the foil shape.

36.6.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 Chapter 23. 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

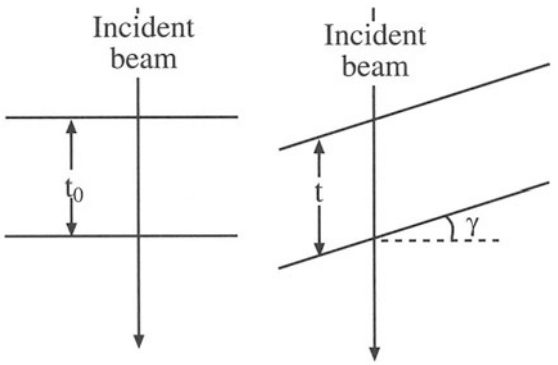


Figure 36.6. The difference between the specimen thickness, t_0, and the distance traveled by the beam, t, that determines beam spreading in a specimen tilted through an angle γ.

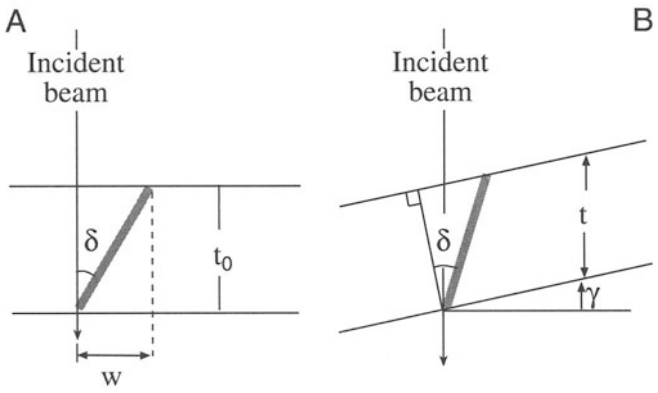


Figure 36.7. The parameters required to measure foil thickness t from a planar defect (projected width w), inclined to the incident beam by angle δ; comparison of (A) an untilted specimen normal to the beam and (B) a specimen tilted through an angle γ.

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 **g**-vector has to be specified. You can calculate ξ_g from the expression

$$\xi_g = \frac{\pi \Omega \cos\theta}{\lambda f(\theta)} \qquad [36.6]$$

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

$$t_0 = w \cot \delta \qquad [36.7]$$

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

$$t = w(\cot \delta - \tan \gamma) \qquad [36.8]$$

Of course both of these methods are inapplicable to noncrystalline materials, and it is not always possible to find a suitable inclined defect next to the analysis region. Furthermore, two-beam conditions are not recommended for microanalysis because of the dangers of anomalous X-ray emission (see Section 35.8). More insidious is the fact that oxidation, during or after specimen preparation, 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. Make all the intensity measurements on your specimen under the same diffraction conditions and incident beam current but with no objective aperture. By calibrating the intensity falling on the screen with a Faraday cup, 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. 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.

Finally, an old method for thickness determination was to deposit small latex spheres on either side of your specimen and measure the thickness by noting parallax shifts between balls on the top and bottom sides as you tilt. This is not recommended, because the latex solution will contribute to specimen contamination and there are alternative and better methods. However, there are cases where you'll have particles or other markers already present on both surfaces, so you might use these for the parallax method.

36.6.B. Contamination-Spot Separation Method

This method, unique to a probe-forming (S)TEM, relies on the propensity of such instruments to generate carbon contamination on both the top and bottom surfaces of the specimen at the point of analysis. If you tilt your specimen by a

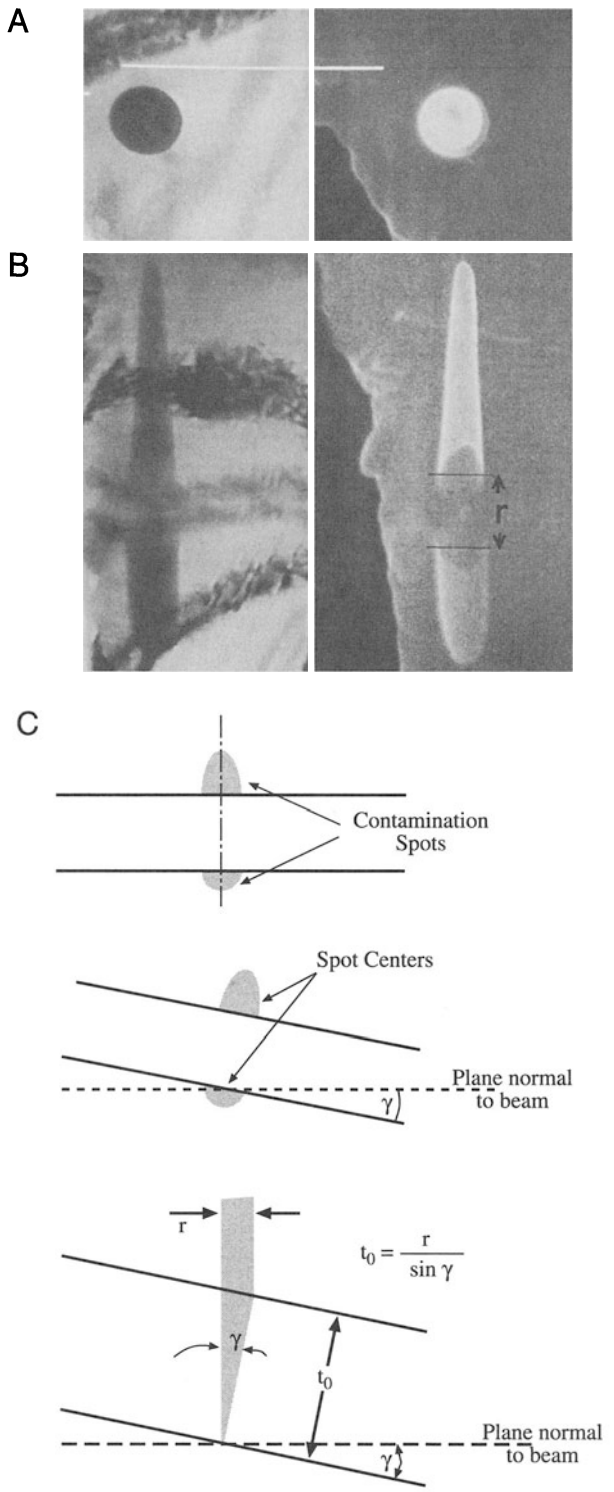


Figure 36.8. The contamination-spot separation method for thickness determination; (A) the contamination is deposited on both surfaces of the specimen and the separation (r) is only visible in (B) when the specimen is tilted sufficiently. The images show the contamination at zero tilt and high tilt angle γ. The two images on the left are obtained in STEM BF and on the right in SE mode. SE mode gives the best image contrast. (C) Geometry required to determine thickness t_0 from the projected spacing r of the contamination spots.

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

$$t_0 = \frac{r}{M \sin \gamma} \qquad [36.9]$$

If the specimen itself is tilted by an angle ε when the contamination is deposited, then

$$t_0 = \frac{r \cos \varepsilon}{M \sin \gamma} \qquad [36.10]$$

Although this method is straightforward, it relies on highly undesirable contamination, which obscures the 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. Having said that, and despite ample evidence that the spot separation method overestimates the thickness by as much as 100%, it is often used because it is quick (and dirty!); it measures t exactly at the analysis point and the shape of the spots can indicate if the beam or the specimen has drifted during microanalysis. So if you can't avoid it, use it with caution.

36.6.C. X-ray Spectrometry Methods

Your X-ray spectrum intensity is a measure of the specimen thickness. Indeed, the standard method of quantification in biological materials uses the bremsstrahlung intensity as a measure of the mass thickness of the specimen, although this isn't used by materials scientists. If an element has two characteristic X-ray lines visible in the spectrum, e.g., the L and K lines, then the relative intensity of these lines will change as the specimen thickness increases because the lower-energy line will be more strongly absorbed. Knowing the necessary absorption parameters, such as μ/ρ, it is then possible to deduce ρt by an iterative process, which is essentially the same as the absorption correction discussed in Section 35.5. If ρ is known, then t can be determined. A similar method involves recording spectra containing an X-ray line that is strongly absorbed at two different tilts, or using two detectors with different take-off angles. In such cases, the two spectra will show different characteristic peak intensities because of the different absorption path lengths in each case. Again, an iterative process is required to extract t.

Porter and Westengen (1981) proposed such a method using standards, which still relies on an iterative procedure based on the absorption correction. In this method, the mass thickness is determined from the following equation

$$\rho t = \frac{\cos\gamma I_0(A)}{e_A m_A} \qquad [36.11]$$

where γ is the tilt angle, m_A is the mass fraction of element A in the specimen, e_A is the number of counts from element A detected per unit incident electron, per unit mass thickness of A in the absence of absorption, and $I_0(A)$ is the absorption-corrected intensity of A from the specimen. The calibration constant e_A is obtained from a standard of known mass thickness, such as a pure element foil. Since $I_0(A)$ is unknown, it is obtained from the absorption correction equation given by

$$I_0(A) = -\frac{\ln\left[1 - u_A I(A)\right]}{u_A} \qquad [36.12]$$

where $I(A)$ is the observed intensity of X-rays from element A and u_A is given by

$$u_A = \frac{\left(\frac{\mu}{\rho}\right)_{\text{Spec}}^{A} \operatorname{cosec}\alpha \cos\theta}{e_A m_A} \qquad [36.13]$$

You can write a similar equation for any element in your specimen. Thus, once your standard has been calibrated, the mass thickness can easily be calculated. However, as in all the absorption correction methods, an iterative process is required. If the foils are bent, or change thickness rapidly, then these methods all become very difficult to carry out and rather inaccurate. Furthermore, if e_A is to be a reliable calibration factor, the electron beam current must be stable and easily measurable; this isn't usually the case in the AEM.

Another closely-related method, described in Section 35.3 when we were discussing absorption, is the extrapolation technique in which the X-ray intensity is related directly to the mass thickness. So if you know the specimen composition you can obtain a value for t (Horita *et al.* 1989).

36.6.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 your 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 is discussed in detail in Section 39.5.

We can apply the EELS method to any specimen, amorphous or crystalline.

> EELS is applicable over a wide range of thicknesses, and with parallel-collection spectrometers it is so rapid that you can even produce thickness "maps" of thin foils. So this approach is highly recommended.

36.6.E. 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 21.1 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. We can envisage on-line thickness determination by digitizing the CBED pattern, scanning it across the STEM BF detector, and measuring the fringe spacing from the Y-modulation output on the STEM CRT. For clean crystalline specimens, this is *the* way to determine t.

In summary, there are many ways you can determine t, but no one method is convenient, accurate, and universally applicable. The various methods also measure different thicknesses, e.g., the crystalline thickness, neglecting surface films, or the thickness including surface films, or the mass thickness. The EELS, CBED, and X-ray absorption methods all have the possibility of widespread on-line use, and we recommend these methods, in order of preference. Detailed reviews of the methods of determining t have also been given by Berriman *et al.* (1984) and Scott and Love (1987).

36.7. MINIMUM DETECTABILITY

Minimum detectability is a measure of the smallest amount of a particular element that can be detected with a defined statistical certainty. Minimum detectability and spatial resolution are intimately related.

> It is a feature of any microanalysis technique that an improvement in spatial resolution is balanced by a worsening of the detectability limit (all other factors being equal).

At higher spatial resolution 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. Accordingly, in the AEM, the price that you pay for improved spatial resolution is a relatively poor minimum detectability. By way of comparison, Figure 36.9 compares the size of the analyzed volume in an EPMA, a TEM/STEM with a thermionic source, and a dedicated STEM with an FEG. The enormous reduction in the beam–specimen interaction volume explains the small signal levels that we obtain in the AEM. It also explains why we have spent so much time emphasizing the need to optimize the beam current through use of higher brightness sources and modifying the specimen-detector configuration to maximize the collection angle, while minimizing the various sources of spurious radiation.

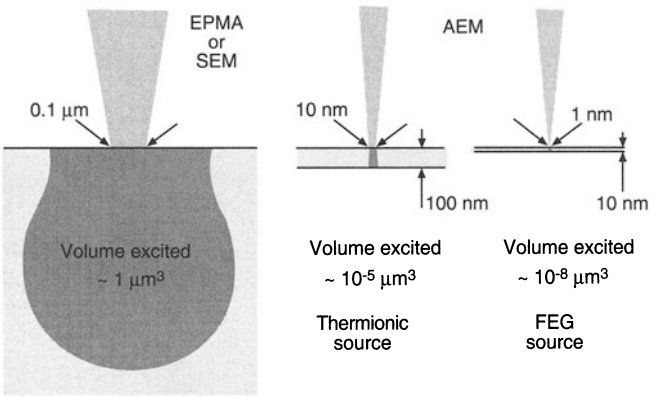


Figure 36.9. Comparison of the relative size of the beam–specimen interaction volumes in an EPMA, a thermionic source AEM, and an FEG-AEM with a bulk, thin, and ultrathin specimen, respectively.

We'll define the minimum detectability in terms of the minimum mass fraction (MMF), which represents the smallest concentration of an element (e.g., in wt.% or ppm) that can be measured in the analysis volume.

Alternatively, the minimum detectable mass (MDM) is sometimes used; the MDM describes the smallest amount of material (e.g., in mg) we can detect. We'll use the MMF, since materials scientists are more used to thinking of composition in terms of wt.% or at.%.

36.7.A. Experimental Factors Affecting the MMF

We can relate the MMF to the practical aspects of microanalysis through the expression of Ziebold (1967)

$$\mathrm{MMF} \propto \frac{1}{\sqrt{P\frac{P}{B}n\tau}} \qquad [36.14]$$

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 electron beam and/or increase the thickness (t) of the specimen. 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 (Zemyan and Williams 1994). Improvements in AEM instrument design, such as using a high-brightness and/or an intermediate voltage source, 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 deterioration and contamination. Improved AEM 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, however, that the Fiori definition of P/B is not the one used in Ziebold's equation (36.14) if you actually want to calculate the MMF.

The other variables in equation 36.14 are the time and number of analyses, which are entirely within your control as operator. Usually, both n and τ are a direct function of your patience and a 5–10 min coffee break is usually the maximum time for any one analysis. With computer control of the analysis procedure, however, there should 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. In the future, a period of several hours or overnight will not be considered unreasonable. 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.

36.7.B. Statistical Criterion for the MMF

We can also define the MMF by a purely statistical criterion. We discussed back in Chapter 34 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 we can come up with a definition of the detectability limit which, when combined with the Cliff–Lorimer equation (assuming Gaussian statistics), gives the MMF (in wt.%) of element B in element A as

$$C_B(\text{MMF}) = \frac{3\left(2\,I_B^b\right)^{\frac{1}{2}} C_A}{k_{AB}\left(I_A - I_A^b\right)} \qquad [36.15]$$

where I_A^b 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} is the Cliff–Lorimer factor. However, if we express the Cliff–Lorimer equation as

$$\frac{C_A}{k_{AB}\left(I_A - I_A^b\right)} = \frac{C_B}{\left(I_B - I_B^b\right)} \qquad [36.16]$$

and substitute it into equation 36.15, the MMF is

$$C_B(\text{MMF}) = \frac{3\left(2\,I_B^b\right)^{\frac{1}{2}} C_B}{I_B - I_B^b} \qquad [36.17]$$

Experimentally, low count-rates from thin specimens mean that typical values of MMF are in the range 0.1% to 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 to 400 kV) and thin specimens to minimize beam broadening. The loss of X-ray intensity, 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 to 2 nm can still be maintained. In summary, optimum MMF and spatial resolution can be obtained by using a high-brightness, intermediate voltage source with thin foils, perhaps of the order of $t \sim 10$ nm. Under these circumstances, MMF values <0.1 wt.% will become routine. Figures 36.9 and 36.10 summarize the classic compromise between resolution and detectability (Lyman 1987).

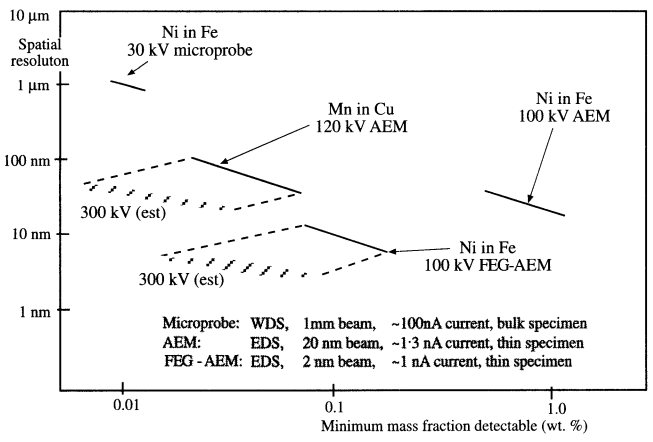


Figure 36.10. Calculation of the relationship between MMF and spatial resolution for the EPMA and a range of AEMs. The inverse relationship between the MMF and resolution is clear, although it is also apparent that the high-brightness sources and high-kV electron beams in the AEM can compensate for the decreased interaction volume in a thin foil.

36.7.C. Comparison with Other Definitions

The MMF definition is not the only way we can measure detectability limits. Currie (1968) has noted at least eight definitions in the analytical chemistry literature. Currie defined three specific limits:

- The decision limit (L_c): Do the results of your analysis indicate detection or not?
- The detection limit (L_d): Can you rely on a specific analysis procedure to lead to detection?
- The determination limit (L_q): Is a specific analysis procedure precise enough to yield a satisfactory quantification?

For I_B counts from element B in a specific peak window and I_B^b in the background, it can be shown that

$$L_C = 2.33\sqrt{I_B^b} \qquad [36.18]$$

$$L_d = 2.71 + 4.65\sqrt{I_B^b} \qquad [36.19]$$

$$L_q = 50\left\{1 + \left(1 + \frac{I_B^b}{12.5}\right)^{\frac{1}{2}}\right\} \qquad [36.20]$$

If there are sufficient counts in the background

$$L_d = 4.65\sqrt{I_B^b} \qquad \text{when } I_B^b > 69 \qquad [36.21]$$

$$L_q = 14.1\sqrt{I_B^b} \qquad \text{when } I_B^b > 2500 \qquad [36.22]$$

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 (~3×) of the element in your specimen (Zemyan 1995). 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. 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.

36.7.D. Minimum Detectable Mass

The MMF values may seem poor compared with other analytical techniques which report ppm or ppb detectability limits. However, it's a different matter if you calculate what the MMF translates to in terms of the minimum detectable mass (MDM).

> The MDM is the minimum number of atoms detectable in the analyzed volume.

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 (1987) 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 and the MDM was less than 10^{-19} g. If the counting time is increased by a factor of 10 and 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 standards. Therefore in ~10-nm-thick specimens, with a spatial resolution approaching the beam diameter d, of 1 to 2 nm, we will be able to detect the presence of 10 to 100 atoms in the analysis volume (10^{-8} μm^3); preliminary data reporting <10 atoms have been published (Lyman *et al.* 1994). For comparison, in the EPMA with 1 μm^3 excitation volume and a 0.01 wt.% MMF, ~10^7 atoms are detected in the analysis volume.

CHAPTER SUMMARY

You cannot optimize spatial resolution and minimum detectability in the same experiment. 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.
- To measure the specimen thickness, use the parameterized EELS approach, otherwise choose between any of the several X-ray intensity methods, or CBED for a crystalline foil.
- To get the best MMF, use the brightest electron source, or the largest beam and thickest specimen, and count for as long as possible.
- If you want the best resolution *and* MMF, an FEG is essential, along with a clean specimen and computer-controlled drift correction; patience is also desirable.

REFERENCES

General References

Berriman, J., Bryan, R.K., Freeman, R., and Leonard, K.R. (1984) *Ultramicroscopy* **13**, 351.

Goldstein, J. I., Williams, D.B., and Cliff, G. (1986) in *Principles of Analytical Electron Microscopy* (Eds. D.C. Joy, A.D. Romig Jr., and J.I. Goldstein), p. 155, Plenum Press, New York.

Scott, V.D. and Love, G. (1987) *Mat. Sci. Tech.* **3**, 600.

Specific References

Currie, L.A. (1968) *Anal. Chem.* **40**, 586.

Goldstein J.I., Costley, J.L., Lorimer, G.W., and Reed, S.J.B. (1977) *Scanning Electron Microscopy*, **1** (Ed. O. Johari), p. 315, IITRI, Chicago, Illinois.

Heinrich, K.F.J., Newbury, D.E., and Yakowitz, H., Eds. (1975) NBS Special Publication 460, U.S. Dept. of Commerce, Washington, D.C.

Horita, Z., Ichitani, K., Sano, T., and Nemoto, M. (1989) *Phil. Mag.* **A59**, 939.

Joy, D.C. (1995) *Monte Carlo Modeling for Electron Microscopy and Microanalysis,* Oxford University Press, New York.

Lorimer, G.W., Cliff, G., and Clark, J.N. (1976) in *Developments in Electron Microscopy and Analysis* (Ed. J.A. Venables), p. 153, Academic Press, London.

Lyman, C.E. (1987) in *Physical Aspects of Microscopic Characterization of Materials* (Eds. J. Kirschner, K. Murata, and J.A. Venables), p. 123, *Scanning Microscopy International,* AMF O'Hare, Illinois.

Lyman, C.E. and Michael, J.R. (1987) in *Analytical Electron Microscopy-1987* (Ed. D.C. Joy), p. 231, San Francisco Press, San Francisco, California.

Lyman, C.E., Goldstein, J.I., Williams, D.B., Ackland, D.W., Von Harrach, S., Nicholls, A.W., and Statham, P.J. (1994) J. *Microsc.* **176**, 85.
Michael, J.R., Williams, D.B., Klein, C.F., and Ayer, R. (1990) *J. Microsc.* **160**, 41.
Porter, D.A. and Westengen, H. (1981) in *Quantitative Microanalysis with High Spatial Resolution* (Eds. M.H. Jacobs, G.W. Lorimer, and P. Doig), p. 94, The Metals Society, London.
Reed, S.J.B. (1982) *Ultramicroscopy* 7, 405.
Williams, D.B., Michael, J.R., Goldstein, J.I., and Romig, A.D. Jr. (1992) *Ultramicroscopy* **47**, 121.
Zemyan, S.M. (1995) Ph.D. dissertation, Lehigh University.
Zemyan, S.M. and Williams, D.B. (1994) *J. Microsc.* **174**, 1.
Ziebold, T.O. (1967) *Anal. Chem.* **39**, 858.

Electron Energy-Loss Spectrometers

37

CHAPTER PREVIEW

Electron energy-loss spectrometry (EELS) is the analysis of the energy distribution of electrons that have interacted inelastically with the specimen. These inelastic collisions tell us a tremendous amount about the electronic structure of the specimen atoms, which in turn reveals details of the nature of these atoms, their bonding and nearest-neighbor distributions, and their dielectric response. In order to examine the spectrum of electron energies we almost invariably use a magnetic prism spectrometer which, when interfaced to a TEM, creates another form of AEM. The magnetic prism is a simple, but highly sensitive, device with resolving power of approximately 1 eV even when the energy of the incident electron beam is up to 400 keV. Despite its simplicity the magnetic prism is operator-intensive and there is not yet the degree of software control to which we are accustomed with XEDS. In this chapter we'll describe the operational principles, how to focus and calibrate the spectrometer, and how to determine the collection semiangle (β). This angle is a most important parameter for interpreting your experimental data. In subsequent chapters we'll go on to look at the spectra, the information they contain, and how we extract quantitative data and images from them. As with XEDS there are standard tests to determine that the spectrometer is working correctly, and we'll describe these also.

As a word of encouragement, or warning, you may get the impression from reviewing the older literature that EELS is the study of small blips which can only be seen by the "trained" eye. While the blips are still often small, we can now be very confident of our interpretation of these spectra. The EELS technique has come to be an excellent complement to the more widely used X-ray spectrometry, since it is well suited to the detection of light elements which are difficult to analyze with XEDS.

Electron Energy-Loss Spectrometers

37

37.1. WHY DO ELECTRON ENERGY-LOSS SPECTROMETRY?

When the electron beam traverses a thin specimen, it loses energy by a variety of processes that we first discussed way back in Chapter 4. The reason we do EELS is so we can separate these inelastically scattered electrons and quantify the information they contain. We've already seen contrasting aspects of inelastic scattering in the TEM:

- Kikuchi lines occur in DPs; these electrons in Kikuchi lines are diffracted at precisely the Bragg angle, and give us much more accurate crystallographic information than the SAD pattern. In thick specimens, many of the electrons in Kikuchi lines are inelastically scattered.
- Chromatic aberration due to energy-loss electrons following different paths through the lenses limits the TEM image resolution, although you can avoid this by using very thin specimens, or STEM imaging. We'll see also, in Chapter 40, that EELS can filter out the chromatic aberration effect in TEM images.
- Specimen damage, which is of course undesirable, is often caused by inelastic interactions.

After reading the next three chapters you should agree that EELS is useful too. If you have an energy-loss spectrometer, it may be that inelastic scatter, in general, is something you would like to happen in your specimens.

The technique of EELS predates X-ray spectrometry; if you want to read a brief history of the technique see the book by Egerton (1996), which we will refer to on many occasions. 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. In contrast to X-ray analysis, EELS has been very slow to develop and still remains firmly in research laboratories rather than applications laboratories. Since probe-forming AEMs became widespread, EELS has become more popular, primarily because it complements XEDS through better detection of the light elements. But as you'll see, we can extract a lot more from the spectra than merely elemental identification. When you have finished this set of chapters you will be ready to read Egerton's text and you will find the book edited by Disko *et al.* (1992) to be another excellent review. Two special issues of *Microscopy, Microanalysis, Microstructure* (Krivanek 1991, 1995a) and *Ultramicroscopy* (Krivanek 1995b) contain papers from EELS workshops.

Currently, there are only two major manufacturers of electron spectrometers for TEMs and they produce radically different instruments, designed for different purposes. We'll describe these two types of electron spectrometers in some detail.

> Two types of commercial spectrometers are presently manufactured: the magnetic prism spectrometer (Gatan) and the omega filter [Zeiss (now LEO)].

The magnetic prism is designed with energy spectrometry as its primary function; this application will constitute the bulk of this chapter. The omega filter is used mainly for energy-filtered imaging although spectra can be obtained; it is a specialized technique because this spectrometer has to be built into the microscope column rather than an optional addition (see later in Section 37.6). However, the Gatan Image Filter (GIF) may soon change this situation, particularly in the materials sciences, since the GIF combines both spectral and imaging capabilities. Magnetic electron spectrometers, along with electrostatic or combined electrostatic and magnetic systems, have been the subject of reviews by Metherell (1971) and Egerton (1996). If you're

an instrument enthusiast you should read these articles. Perhaps because there are so few manufacturers of the spectrometers, the competitively driven progress in user-friendly instrumentation and software control that has pushed X-ray spectrometry forward over the last two decades has been slow to occur in EELS; this lack of user friendliness, in part, accounts for the relatively small number of users of the technique.

> You should know also that there is another area of EELS research which uses electron spectrometers to measure exceedingly small (millivolt) energy losses in low-energy electron beams reflected from the surfaces of samples in UHV surface-chemistry instrumentation such as ESCA and Auger systems. We will ignore this type of EELS completely and deal only with transmission EELS studies of high-voltage electron beams.

There are two fundamentally different ways of detecting the spectrum generated by the magnetic prism spectrometer: either serially or, more efficiently, in parallel (see Section 37.3). Also, you can operate your TEM either in image mode or diffraction mode (see Section 37.4) and this choice has a major effect on the information that can be gathered. Before we look at these options, however, we need to look at the magnetic prism spectrometer itself.

37.2. THE MAGNETIC PRISM; A SPECTROMETER AND A LENS

We use a magnetic prism rather than one of the other kinds of spectrometer (e.g., electrostatic) for several reasons:

- It is compact, and therefore easily interfaced to the TEM. (Remember the WDS problem.)
- It offers sufficient energy resolution to distinguish all the elements in the periodic table and so is ideal for microanalysis.
- 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.

Schematic diagrams of the spectrometer optics are shown in Figures 37.1A and B. A picture of a Gatan spectrometer, which has to be installed beneath the camera system of a TEM or after the ADF detector in a DSTEM, is shown in Figure 37.2. Because these spectrometers are so widespread, many of the numerical values in this chapter are taken from the Gatan literature. For the details of operation you should, of course, read the instruction manual.

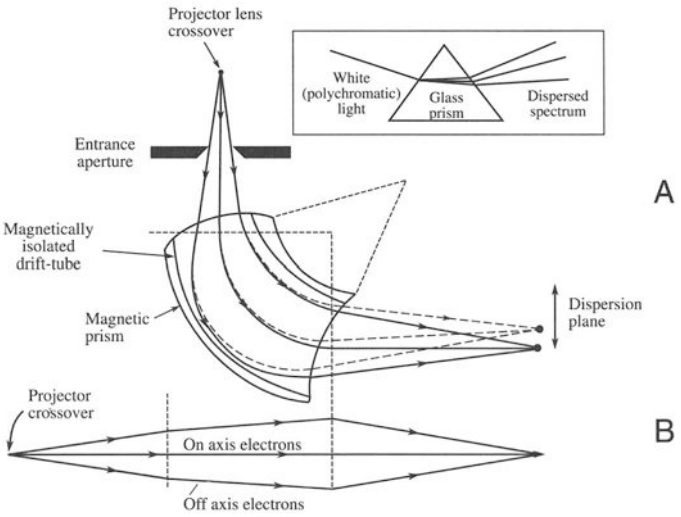


Figure 37.1. Ray paths through a magnetic prism spectrometer showing (A) dispersion and focusing of the electrons in the plane of the spectrometer and (B) the lens-focusing action in the plane normal to the spectrometer; compare the nonfocusing action of a glass prism on visible light (inset).

From Figure 37.1A we can see that electrons are selected by a variable entrance aperture (diameters: 1, 2, 3, or 5 mm in the Gatan system). The electrons travel down a "drift tube" through the spectrometer and are deflected through $\geq 90°$ by the surrounding magnetic field. Electrons with greater energy loss (dashed line) are deflected further than those suffering zero loss (full line). A spectrum is thus formed in the dispersion plane which consists of a distribution of electron counts (again incorrectly referred to as intensity) (I) versus energy loss ($\mathcal{E}$). This process is exactly analogous to the dispersion of white light by a glass prism.

> 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). As a compromise we will use $\mathcal{E}$ (note the different font), but remember it really means a change in E.

Now if you look at Figure 37.1B, 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 plane of the spectrometer,

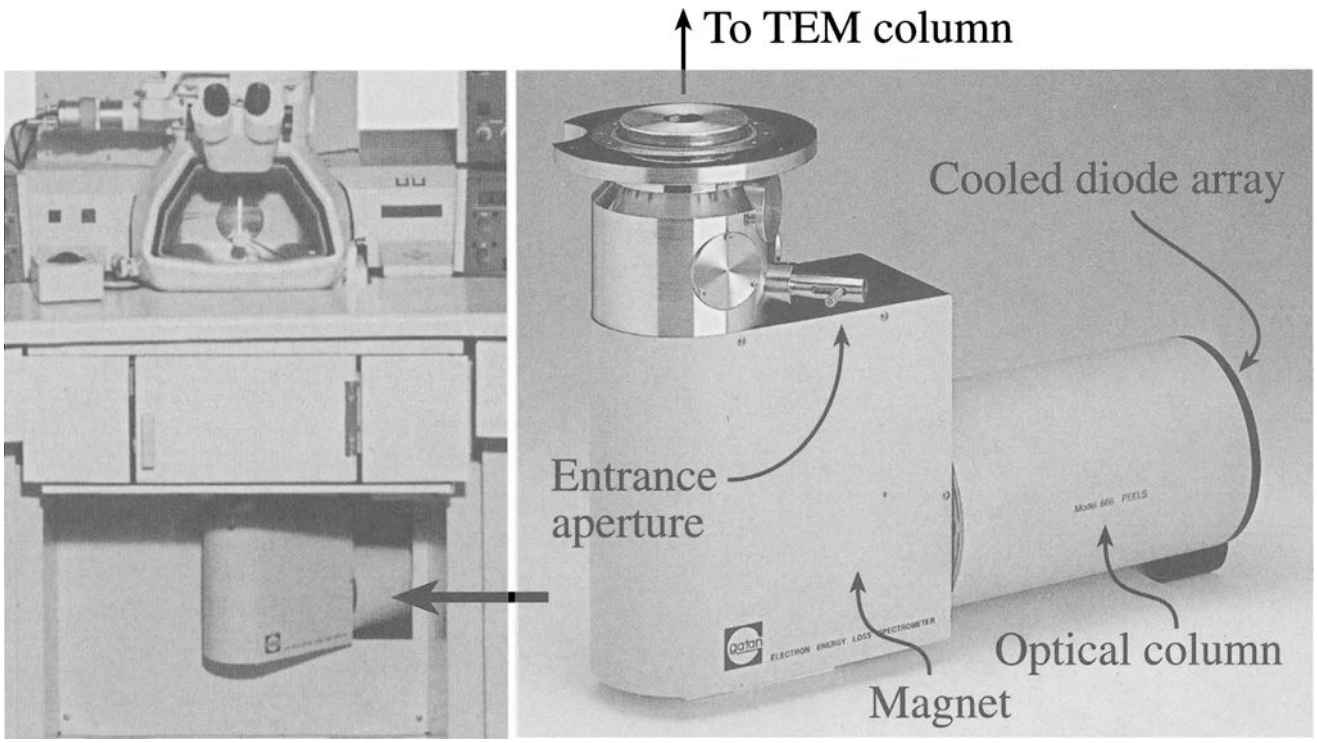


Figure 37.2. A Gatan parallel-collection magnetic prism spectrometer interfaced below the viewing screen of an AEM showing the shielded container for the magnet and the optical column and diode-array detection system.

which thus acts as a magnetic lens. The object plane of the spectrometer is usually set at the projector lens back focal plane, coincident with the differential pumping aperture. This focusing action is not seen in the otherwise analogous glass prism. Many examples of spectra will be given in the next two chapters.

37.2.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. Correction of second-order aberrations and astigmatism are minor steps which we will not describe, but focusing is rather important.

The spectrometer has to focus the electrons because off-axis electrons experience a different magnetic field than 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 37.1A. These nonnormal faces also act to ensure that electrons traveling out of the plane of the paper in Figure 37.1A are also focused in the dispersion plane, as shown in Figure 37.1B. Such a spectrometer is described as "double focusing." The faces are also curved to minimize aberrations.

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 depends on the TEM you are using.

> In a TEM/STEM, or a DSTEM with post-specimen lenses, the object plane is the back focal plane of the projector lens.
> In DSTEMs with no post-specimen lenses, the object plane is the plane of the specimen.

In the TEM the projector lens setting is usually fixed, and the manufacturer has already set this plane to coincide with the differential pumping aperture separating

the column from the viewing chamber. In some dedicated DSTEMs there are no post-specimen lenses, so the object plane of the spectrometer must be the plane of the specimen. In this case it is essential that you keep the specimen height constant.

Now in practice, the back focal plane of the projector does move a little 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 which we call the "zero-loss peak"; we'll talk about this more in the next chapter. You can see the zero-loss peak on the CRT or computer display of the EELS system. You focus the peak by adjusting a pair of pre-spectrometer quadrupoles until it has a minimum width and maximum height. To correct second-order effects, a pair of sextupoles is also available. The actual method of focusing depends on the type of spectrometer. In a parallel-collection system, focusing is controlled directly by the quadrupoles. In a serial-collection system (see below), there is a slit in the dispersion plane. There are some rules to guide you when adjusting the slits:

- If the slit is too wide, then the zero-loss peak has a flat top and is very broad.
- If the slit is too narrow, you lose spectral intensity.
- If your spectrometer is not focused on the slit, the peak is also broad, but more Gaussian-shaped, as in Figure 37.3.

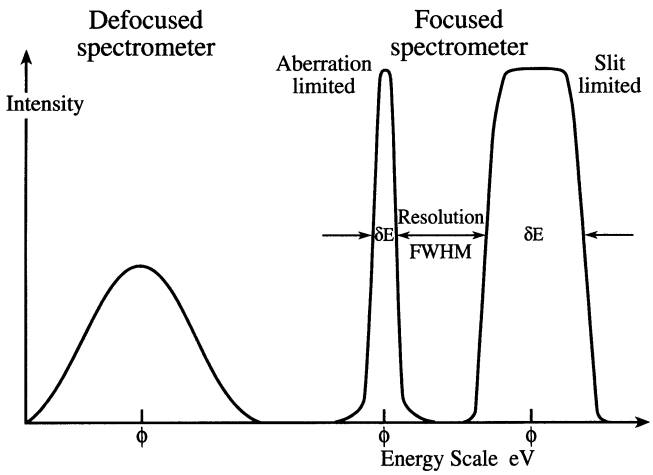


Figure 37.3. The zero-loss peak in a spectrum showing the effect of a defocused spectrometer, and the slit-limited condition, i.e., slits too wide (for serial collection only). In the usual focused condition, the resolution of the spectrometer is defined as the FWHM of the peak.

37.2.B. Calibrating the Spectrometer

We calibrate the spectrometer by placing an accurately known voltage on the drift tube. You will see this voltage displace the spectrum by a fixed amount. Alternatively, as in XEDS, you can look for features in a spectrum from a known specimen that occur at specific energies spanning the spectral display range, such as the zero-loss peak (at 0 eV) and the Ni L edge at 855 eV. (See Figure 37.4 and the next chapter for more details on the spectrum itself.) Modern electronics are reasonably stable and the calibration doesn't shift substantially, but you should check it regularly throughout an operating session since shifts of a few eV do occur and these are of the same order as the energy resolution of the spectrometer.

37.3. ACQUIRING A SPECTRUM

Figure 37.4A shows an EELS spectrum, which we'll describe in detail in the next chapter. For the time being note:

- The zero-loss peak is very intense.
- The intensity range is enormous; this graph uses a logarithmic scale.

There are two ways in which we acquire the spectrum. We either build up the spectrum one channel at a time, which is known as serial-acquisition EELS, or SEELS, or we acquire all the channels simultaneously, which is parallel-acquisition EELS, or PEELS (the word "acquisition" is often omitted when discussing the subject). The two modes are shown in the schematic spectra in Figure 37.4B and C . The intensity changes shown correspond to ionization "edges" and each element exhibits characteristic edges at specific values of $\mathcal{E}$. For example, the carbon K edge onset at 284 eV corresponds to the critical energy E_C required to eject the carbon K shell electron; more about this in Chapter 38.

37.3.A. Serial Collection

SEELS is the original, but least efficient, method of acquisition. No serial systems are now manufactured, but many remain in TEM labs. As shown in Figure 37.5A, the spectrometer system scans or "ramps" the spectrum across a slit in the dispersion plane, leaving the spectrum for a fixed "dwell time" (τ) at each energy loss. The collection and measurement is quite straightforward.

- Ramping is achieved magnetically by changing the current through a set of coils placed after the spectrometer magnet.

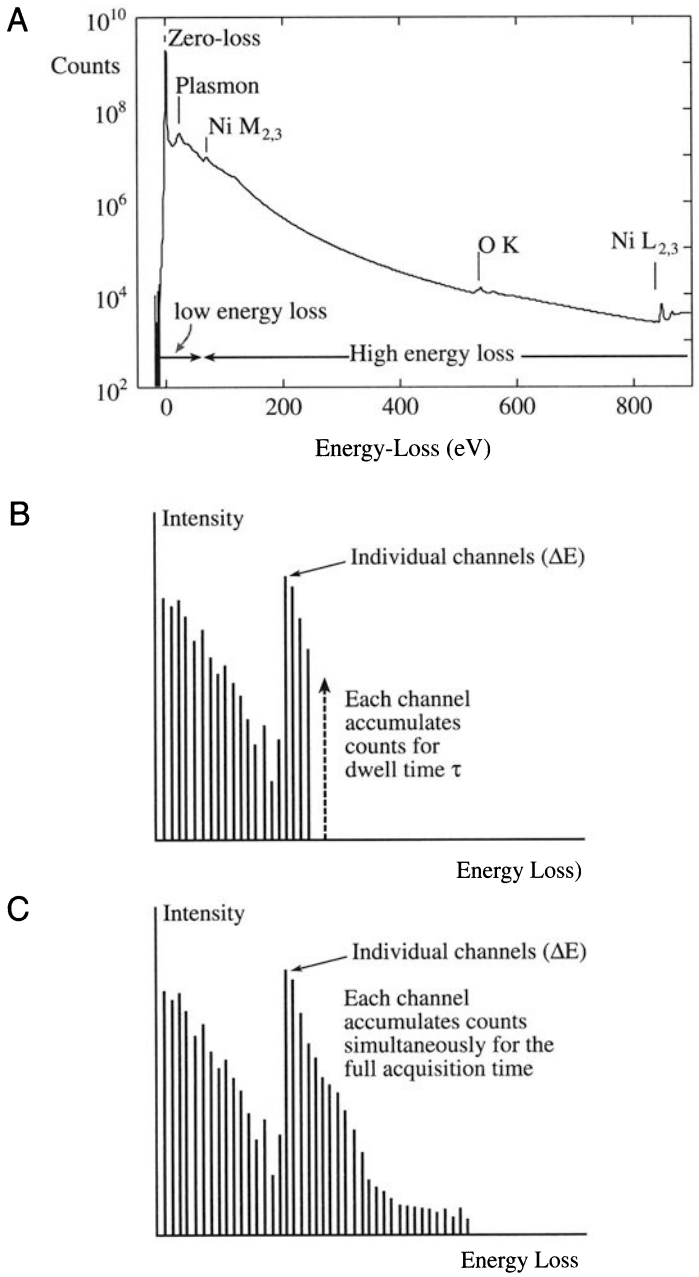


Figure 37.4. (A) An EELS spectrum displayed in logarithmic intensity mode. The zero loss is an order of magnitude more intense than the low-energy-loss portion, which is itself many orders of magnitude more intense than the small ionization edges, identified in the high-energy-loss range. The spectra can be acquired (B) serially or (C) in parallel. In serial collection, each of the energy channels accumulates counts for a given dwell time τ (typically 100 ms) before the next channel (energy range) is selected. In parallel collection, the complete spectrum is gathered simultaneously in all the available channels.

- A plastic scintillator located directly behind the slit receives an electron flux in time τ.
- The integrated electron current is converted to photons and amplified by a photomultiplier (PM) tube.
- After each dwell time, the total signal from the PM is then assigned to a channel corresponding to a specific energy loss, $\mathcal{E}$, in a multichannel analyzer (MCA).

We've already discussed scintillator-PM systems of the sort used for SE detection in the SEM or STEM, or direct beam (BF) detectors in the STEM (see Chapter 7).

- They have a tremendous gain and therefore can handle a large intensity range.
- They show a rapid response to intensity changes and exhibit both low noise and a high detective quantum efficiency, or DQE.

The DQE is close to 1 under ideal conditions, although the absolute DQE of SEELS is very low (≈0.001) since most of the electrons are wasted at any single acquisition time. These factors are particularly important in EELS, because the spectrum intensity can vary by many orders of magnitude. However, the susceptibility of scintillators to beam damage also means that the intense low-energy part of the EELS spectrum can be a health danger to the scintillator. The Gatan systems give an unmistakable audible warning when the electron flux is too great. Even with careful operation the plastic scintillator in your SEELS will become damaged and, as a wise precaution, you should replace it every few months.

The spectrum display is built up in a serial manner, as we showed in Figure 37.4B. Typically, the MCA will have 1024 or 2048 channels. The display resolution can be selected from < 0.1 eV/channel to about 10 eV/channel, depending on how much of the spectrum you want to gather. For example, if you set the display resolution to 1 eV per channel, the entire EELS signal out to 1024 eV energy loss will be recorded. But even if you choose a short τ, e.g., 100 ms, it will still take you about 100 s to accumulate a full spectrum out to 1024 eV, and in 100 ms the total intensity in each channel is going to be small. In particular, the limited energy range that may be of interest to you will only have been sampled for a few seconds at best. So when you know what portion of the spectrum is of interest, you should restrict collection to fewer channels for a longer τ, or accumulate several spectra and add the intensities together in order to get satisfactory counting statistics across the full spectrum. We'll see later that you can influence the total intensity by your choice of operating mode.

During acquisition of a spectrum, the intensity changes by several orders of magnitude (see Figure 37.4A), and usually we are interested primarily in the low-intensity (high-energy-loss) part of the spectrum. Therefore, we have

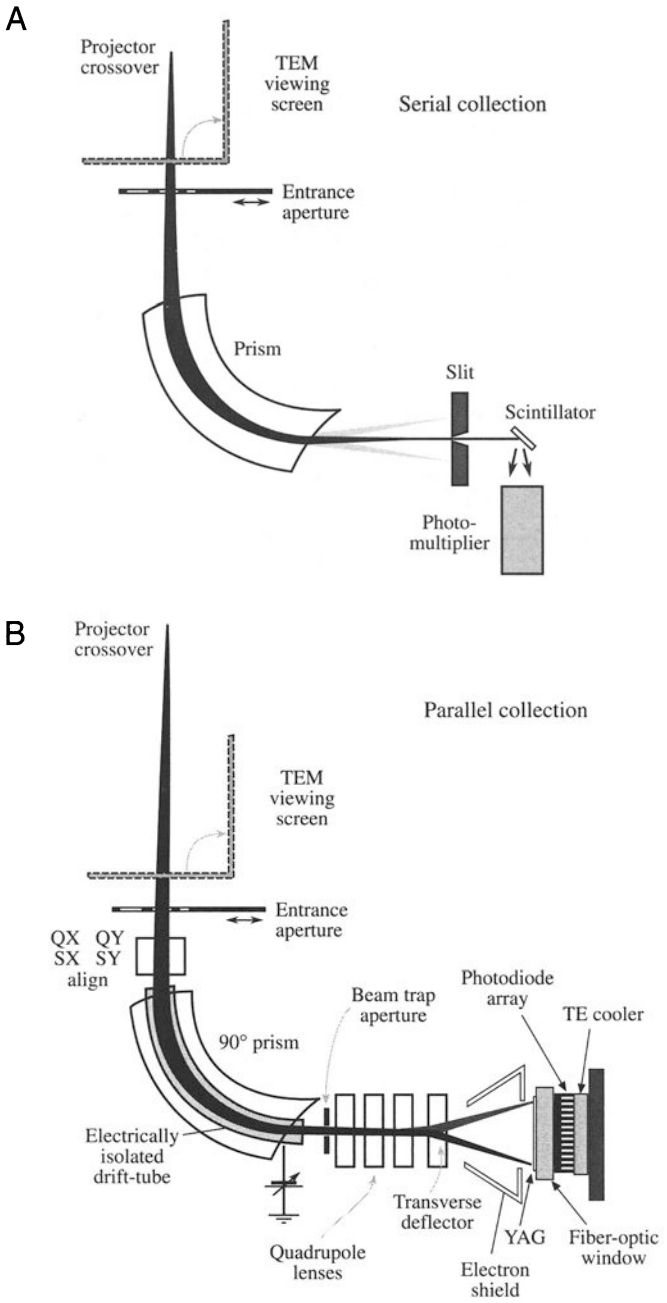


Figure 37.5. Comparison of EELS acquisition modes. (A) Serial collection of the spectrum through a slit onto a scintillator-PM system. (B) Parallel collection onto a YAG scintillator fiber-optically coupled to a thermoelectrically cooled (TE) semiconductor diode array. (Q = quadrupole, S = sextupole)

to change the display scale during SEELS spectrum collection in order to present a visible display of the low-intensity part of the spectrum. We can do this in one of two ways: either we increase τ by a factor of 10–100, or we electronically increase the gain of the scintillator-PM system by some orders of magnitude and operate it in two different modes, as explained below.

For very high electron fluxes encountered in the zero-loss and low-energy-loss portion of the spectrum, we use "analog collection." The total voltage generated at the exit of the PM can be related directly to the total electron current incident on the scintillator. Using a voltage-to-frequency (V/F) converter, this voltage is converted to a pulse that can be fed directly to the computer display. By changing the gain of the V/F converter we can handle electron fluxes in excess of 10^{10}/s (about 1 nA) with no problem. At this kind of incident current, the individual pulses overlap and are all integrated to give a continuous output. At the high-energy-loss end of the spectrum, where electron intensity is very low, the collection system can be changed to count the photon bursts generated by each electron, and this is termed "single-electron counting." You throw a switch on the SEELS electronics control panel to change the detection mode at a given point (energy loss) in the serial collection, or you can change modes under software control, usually at the same energy in the spectrum where the display gain change occurs.

We will see that it is sometimes essential to record the spectrum over a wide range, say from zero to several hundred eV loss. In this case, the intense low-loss region must be gathered and displayed in the same spectrum as the less-intense high-loss region. A potential problem in this situation is that, even if the intense low-loss region does not physically damage the detector, it may cause it to glow and the glow persists for some time after the electrons hit the scintillator. So it is possible that while you are gathering the low-intensity part of the spectrum, the afterglow will contribute unwanted noise.

> To avoid the afterglow, you should acquire all SEELS spectra in reverse-scan fashion. Start at the high loss, then go to the low-loss region, and finish at zero.

If the zero-loss peak is not required, then it is good practice to cease acquisition at about 5–10 eV to save the scintillator from the accumulated effects of high electron fluxes.

37.3.B. Parallel Collection

PEELS gathers the whole energy spectrum simultaneously and is much more efficient than SEELS. PEELS comprises a YAG scintillator coupled via fiber optics to a semiconductor photodiode array in the dispersion plane of the spectrometer, as shown in Figure 37.5B. The array consists of

1024 electrically isolated and thermoelectrically cooled silicon diodes, each about 25 μm across. These diode arrays show varying responses and exhibit specific artifacts, which we'll discuss in the next chapter.

The resultant spectrum accumulates across the whole energy range simultaneously, as we showed schematically back in Figure 37.4C. Rather than having a dwell time as in SEELS, we now have an integration time which can vary from a few msec to several hundred seconds. After integration, the whole spectrum is read out via an amplifier through an A/D converter and into an MCA system. Reasonable spectra can be acquired in a fraction of a second, making PEELS imaging a practical reality. We'll see more about this in Section 40.3.

The advantage of PEELS is that all regions of interest are gathered for the whole integration time, and not just some fraction of the acquisition time as in SEELS. Thus PEELS is much more efficient than SEELS, and its DQE is ~0.5.

A quick warning: you can damage the YAG scintillator, particularly in intermediate-voltage microscopes. Ways to avoid this problem, especially for the intense zero-loss beam, are still being developed. Currently, the zero-loss beam intensity is attenuated or deflected off the scintillator if it is not required, or if the beam current exceeds 0.5 nA. While the SEELS system can handle intense signals, the PEELS diode array saturates at signal intensities of about 16,000 counts.

You must select an integration time that won't saturate the detector and then collect as many consecutive integrations as you need.

One other advantage of the PEELS system is that the scintillator shield is designed to act as a Faraday cup, and so you can use it to measure the total beam current. Usually, the beam is moved onto the shield whenever acquisition ceases, so a constant record of the beam current is available.

To summarize:

- SEELS detects one channel at a time; the detector is easy to optimize and simple to operate.
- PEELS detects the whole spectrum at one time, but the diode array is hard to optimize.
- PEELS exhibits artifacts, and has more complex electron optics, but is 2–3 orders of magnitude more efficient than SEELS with a relatively high DQE.

37.3.C. Spectrometer Dispersion

We define the dispersion as the distance in the spectrum (dx) between the positions of electrons differing by energy 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. In the commercial serial spectrometers, the radius of curvature (R) of electrons traveling on axis is about 200 mm, and for 100-keV electrons dx/dE is about 2 μm/eV. This dispersion, while small, is sufficiently large so as not to limit the energy resolution (see below). A serial-detection system can process the spectrum without any post-spectrometer magnifying lenses. For parallel collection 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 ~15× before the spectrum can be detected with resolution closer to 1eV. This magnification requires post-spectrometer lenses; 4 quadrupoles are used in the Gatan system. The dispersion should be linear across the 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 the diode array.

You may wonder why we don't record the spectrum on film rather than collect it electronically. In fact, this was the first method used to detect electron spectra, but it is an analog method and photographic film does not have a linear response over the usual range of spectral intensities. Furthermore, the grain size of the photographic emulsion (10–20 μm) would limit energy resolution to about 5 eV unless the dispersion were increased, so photographic recording is no longer used.

37.3.D. Spectrometer Resolution

We define the energy resolution of the spectrometer as the FWHM of the zero-loss peak (see back in Figure 37.3). If you don't focus your spectrometer as we just described, then you won't get the best resolution. The best resolution you can get is determined by the type of electron source. As we discussed back in Chapter 5 (see Table 5.1), at ~100 keV a W source has the worst energy resolution (2.5 eV), and a LaB_6 is slightly better than W at 1.5 eV while a cold FEG gives the best value (0.3 eV). Because of the high emission current from thermionic sources, the energy resolution is in fact limited by electrostatic interactions between electrons at the filament crossover. This electron–electron interaction is called the Boersch effect. You can partially overcome this by undersaturating the filament and using only the electrons in the halo. Under these circumstances a LaB_6 source can attain a resolution below

1 eV, but at the expense of a considerable loss of signal, for which you can compensate by increasing the beam size and/or the C2 aperture.

> The energy resolution decreases slightly as the energy loss increases, but it should be no worse than ~1.5 times the zero-loss peak width up to 1000-eV energy loss.
> If you operate at higher voltage, you should also expect a degradation of energy resolution as the kV increases, approximately tripling from 100 kV to 400 kV.

Because the magnetic prism is so extremely sensitive, external magnetic fields in the microscope room may limit the resolution. You may see a disturbance to the spectrum directly if you sit in a metal chair and move around, or if you open metal doors into the TEM room. Remember, for comparison, that the energy resolution of XEDS is >100 eV.

For different EELS operations, different factors affect the resolution. In a SEELS system the resolution is most easily changed by adjusting the slit width. A larger slit width results in poorer energy resolution, but has the advantage of increasing the total current into the detection system. In a PEELS system optimum resolution requires a small projector crossover and a small (1 mm or 2 mm) entrance aperture. The resolution is degraded by choosing a larger entrance aperture because of off-axis beams; i.e., degradation is caused by a C_s effect for the lenses in the energy analyzing spectrometer. Similarly, the resolution may change as you deflect the zero-loss peak onto different regions of the photodiode, although this should not happen if the spectrometer optics are properly aligned.

37.3.E. Point-Spread Function

In a PEELS, you can reduce the magnification of your spectrum so that the zero-loss peak occupies only a single photodiode channel. Any intensity registered outside that single channel is an artifact of the detector system array and is called the point-spread function. This function acts to degrade the inherent resolution of the magnetic spectrometer. The zero-loss peak may spread on its way through the YAG scintillator and the fiber optics before hitting the photodiode. Figure 37.6 shows the point-spread function of a PEELS and clearly there is intensity well outside a single channel. This is important because this spreading broadens features in your spectrum, such as fine structure in ionization edges, and you need to remove it by deconvolution (see Section 39.6). The concept is essentially the same as the point-spread function we discussed for HRTEM.

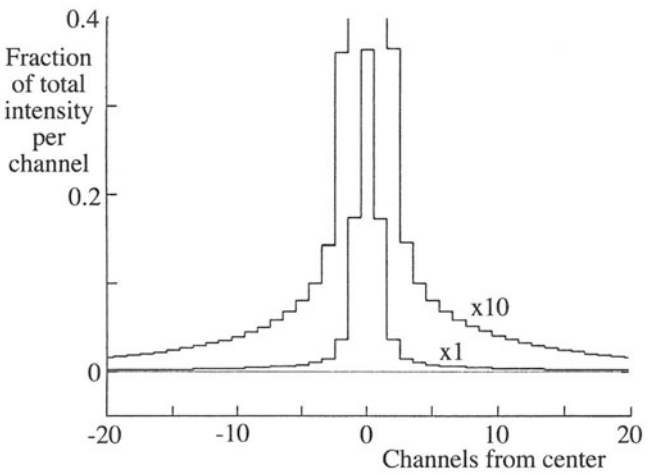


Figure 37.6. The point-spread function, showing the degradation of the intense well-defined zero-loss peak through spreading of the signal as it is transferred from the scintillator via the fiber-optic coupling to the photodiodes. The peak should occupy a single channel but is spread across several channels.

37.4. IMAGE AND DIFFRACTION MODES

When performing EELS in a TEM/STEM, you can operate in either of two modes, and the terminology for this is confusing. If you 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, and the spectrometer uses this pattern as its object. From the spectroscopists' viewpoint, therefore, this is termed "diffraction mode" or "diffraction coupling," but from the microscopists' viewpoint it is more natural to call this "image mode" since you 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 spectrometer object plane contains an image, and the terminology is reversed.

> The spectroscopist uses the term "image mode" or "image coupling" and the microscopist says "diffraction mode."

Both sets of terms appear in the literature, often without precise definition, so it can be rather confusing.

> In this text "image mode" means an image is present on the TEM screen; i.e., we use the microscopists' terminology.

So your first step is to ensure that a focused image or DP is present on your TEM screen, and then the spectrum can be focused onto the dispersion plane.

37.4.A. Spectrometer Collection Angle

The collection semiangle of the spectrometer (β) is the most important variable in quantification, so you should know β for all your standard operating situations. If you do gather spectra with different β, it is difficult to make sensible comparisons without considerable post-acquisition processing. The detailed intensity variations in the spectrum depend on the range of electron scattering angles which are gathered by the spectrometer. Under certain circumstances, the effective value of β can be modified by the beam-convergence semiangle, α, but we'll discuss that when we talk about quantification in Chapter 39.

> β is the semiangle subtended at the specimen by the entrance aperture to the spectrometer.

This definition is illustrated in Figure 37.7. The value of β is affected by the mode of microscope operation, and so we will describe how to measure β under different conditions that may be encountered.

Dedicated STEMs. In a basic DSTEM the situation is straightforward if there are no post-specimen lenses because, as shown in Figure 37.7, the collection angle 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

$$\beta \approx \frac{d}{2h} \qquad [37.1]$$

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 available 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 discussed below.

TEM-image mode. Remember that in image mode, a magnified image of the specimen is present on the viewing screen and the spectrometer object plane contains a DP. 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 you can control the angular distribution of electrons contributing to any TEM image by the size of the objective aperture in the DP in the back focal plane of the objective lens. If you don't use an objective aperture, then the collection semiangle is very

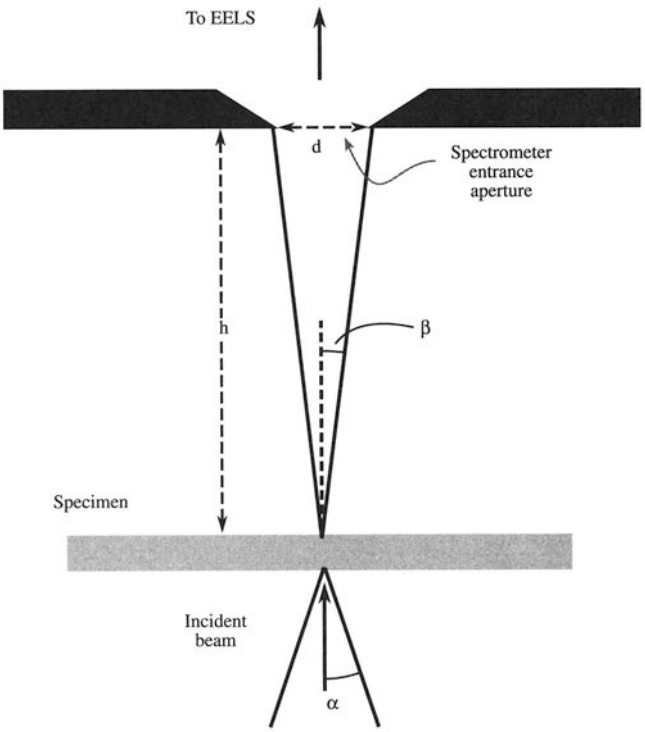


Figure 37.7. Schematic diagram showing the definition of β in a DSTEM in which no lenses exist between the specimen and the spectrometer entrance aperture.

large (>~100 mrads) and need not be calculated accurately, because we'll see that small differences in a large β value 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). This magnification may be described in terms of the camera length L of the DP, and this is given by

$$L \approx \frac{D}{M} \qquad [37.2]$$

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 that

$$\beta \approx \frac{r_0}{L} \qquad [37.3]$$

where r_0 is the maximum radius of the DP in the focal plane of the spectrometer. Typically, r_0 is approximately 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 mrad any calculation or quantification you do will be 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 above, β is the objective aperture diameter divided by twice the focal length of the objective lens, as shown in Figure 37.8. For example, with a focal length of 3 mm and a 30-μm aperture, β is about 5 mrad.

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 (with some considerable error) to the area of the image that sits directly above the spectrometer entrance aperture. We will return to this point in more detail in Section 37.4 when we discuss the spatial resolution of microanalysis. Remember also that with the objective aperture in, you cannot do XEDS. Therefore, 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. Remember, the object plane of the spectrometer (the projector lens BFP) contains a low-magnification image of the specimen; so you see a DP on the screen and the same DP is in the plane of the spectrometer entrance aperture. Under these circumstances we control β by our choice of the spectrometer entrance aperture, as shown in Figure 37.9.

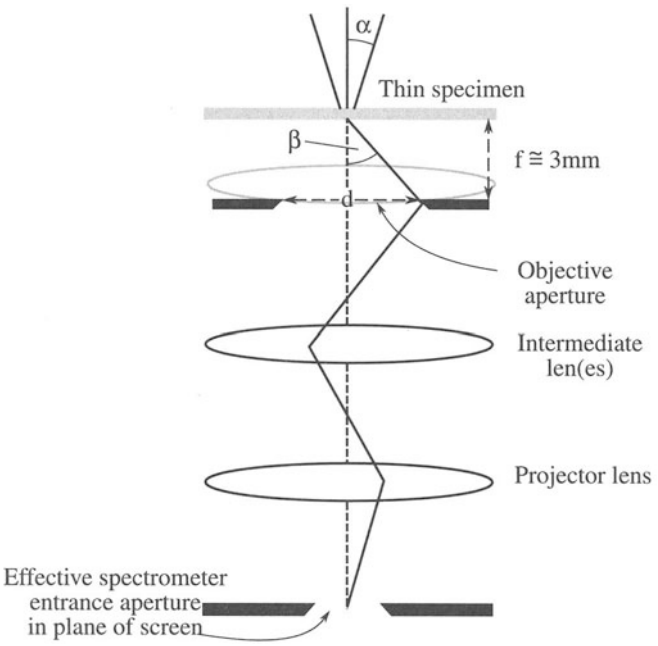


Figure 37.8. The value of β in TEM-image mode is governed by the dimensions of the objective aperture.

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 also shown in Figure 37.9. Knowing the size of the spectrometer entrance aperture, the value of β can be calibrated by twice the Bragg angle, $2\theta_B$, that separates the 000 spot and a known *hkℓ*, disk. If the effective aperture diameter in the recording plane is d_{eff} and the distance b is related to the angle $2\theta_B$, as shown in Figure 5.8, so

$$\beta \approx \frac{d_{eff}}{2}\frac{2\theta_B}{b} \qquad [37.4]$$

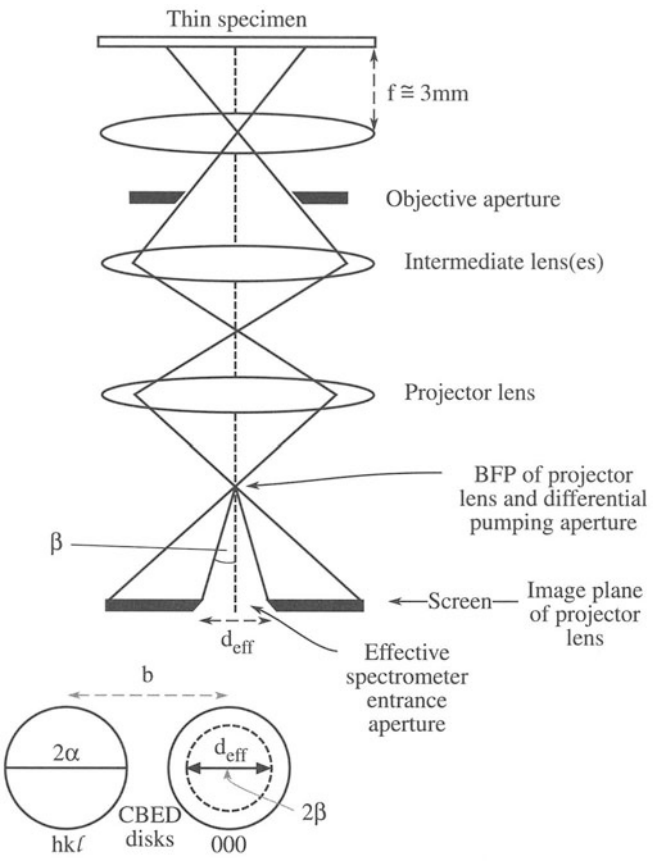


Figure 37.9. The value of β in TEM/STEM diffraction mode is determined by the effective diameter of the spectrometer entrance aperture (d_{eff}), projected into the plane of the diffraction pattern. The value of the d_{eff} can be calibrated by reference to a known diffraction pattern (below) in which d_{eff} can be related to $2\theta_B$.

The effective entrance aperture diameter d_{eff} at the recording plane is related to the actual diameter d by

$$d_{\text{eff}} = \frac{d\,D}{D_A} \qquad [37.5]$$

where D is the distance from the projector crossover to the recording plane (remember, the film is not at the same height as the screen); D_A is the distance between the crossover and the actual entrance aperture. Alternatively, β can be determined directly if the camera length on the recording plane (L) is known, since

$$\beta = \frac{D}{D_A}\frac{d}{L} \qquad [37.6]$$

D_A is typically 610 mm for most Gatan PEELS systems, but D varies from microscope to microscope; you have control over d and L. For example, if D is 500 mm and L is 800 mm, then for the 5-mm entrance aperture, β is ~5 mrads.

If you 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, you have moved the specimen to the object plane of the spectrometer. This special value of L is equal to the D, which you should know for your own microscope. Then β is simply the entrance aperture diameter divided by D_A (610 mm). Life will be much easier when all these calculations are incorporated in ELP (see Section 1.5).

In summary, the collection angle is a crucial factor in EELS. Large collection angles will give high intensity in the spectrum. If you collect your spectrum in image mode without an objective aperture, then you don't compromise your energy resolution. If you're in diffraction mode and you control β with the entrance aperture, then a large aperture (high intensity, large β) will degrade the energy resolution.

37.4.B. 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, you position the area to be analyzed on the optic axis, above the spectrometer 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 microanalysis 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

$$d = \theta\,\Delta f \qquad [37.7]$$

where θ is the angle of scatter, typically <10 mrads, and Δf is the defocus error due to chromatic aberration given by

$$\Delta f = C_c\frac{E}{E_0} \qquad [37.8]$$

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 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 large compared to the value of 10 nm which we calculated without considering chromatic aberration effects.

> While TEM-image mode is good for gathering spectra with a large β and high-energy resolution, the price you pay is poorer spatial resolution.

In TEM diffraction mode, you select the area of the specimen contributing to the DP in the usual way. You can either use 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 microanalysis; just as for XEDS microanalysis, STEM operating mode is recommended for EELS microanalysis.

- Form an image in STEM mode.
- Stop the probe from scanning.
- Position it on the area to be analyzed.
- Switch on the EELS; in a TEM/STEM you have to lift up the TEM viewing screen also!

37.5. WHAT YOU NEED TO KNOW ABOUT YOUR PEELS

As with XEDS, where there are several standard tests you need to perform to determine that all is well with the detector and the electronics, there are similar tests for the PEELS diode array and electronics. Some of these are described in the Gatan handbook and others have been proposed by Egerton *et al.* (1993). We'll discuss specific artifacts visible in the spectrum in the next chapter.

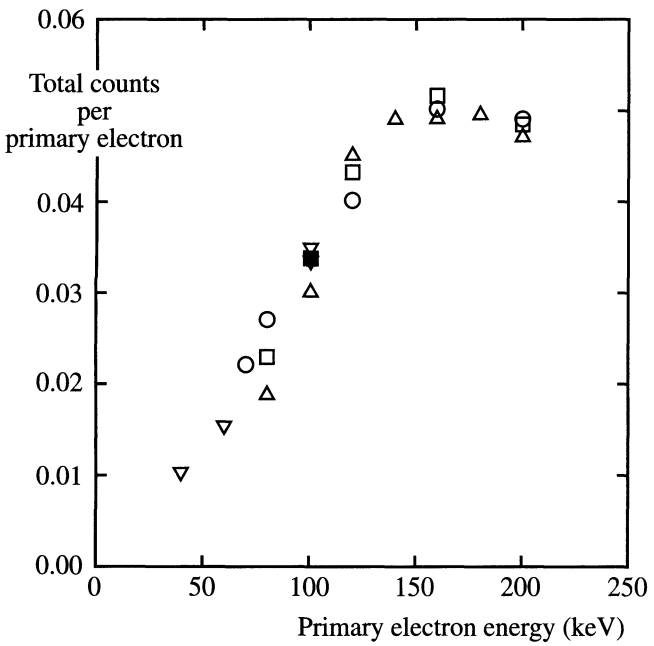


Figure 37.10. Nonlinear response of the diode array as a function of the beam energy. The response saturates at ~150 keV. Different symbols represent different dispersion settings.

Since increasing the keV means more electrons are generated in the scintillator, the sensitivity of the diode array should be linearly related to the electron energy. Egerton *et al.* (1993) have shown (see Figure 37.10) that, in fact, the Gatan diode-array response saturates at ~150 keV because of electron penetration. This nonlinearity doesn't affect quantification, since we typically make measurements over a very small energy range (<1 keV), but it means that there is no gain in count rate by operating >~150 keV. More important is the need for the YAG to respond linearly to different intensities incident on it; you should check that this is so 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, you have to subtract the dark current (see Section 38.5). Obviously, the ratio of these two intensities should be unity for all levels of signal falling on the YAG. If it is not so, then you should consult the manufacturer.

37.6. IMAGING SPECTROMETERS

Two types of electron spectrometers are designed for energy-filtered imaging:

- In-column spectrometers on Zeiss 902 and LEO 912 series TEMs for "electron-spectroscopic imaging" (ESI).
- The Gatan Imaging Filter (GIF), which is a variation of the magnetic prism spectrometer.

Zeiss first used a mirror-prism system originally devised by Castaing and Henry (1962) and described by Zanchi *et al.* (1982). The drawback to the mirror-prism is the need to split the high-tension supply and raise the mirror to the same voltage as the gun. So LEO now use a magnetic omega (Ω) filter (Lanio *et al.* 1986). The Ω filter disperses the electrons in the column, as shown in Figure 37.11A. The spectrometer is placed in the TEM column between the intermediate and the projector lenses. Usually, you project an image into the prism, which is focused on a DP in the back focal plane of the intermediate lens. Therefore, the entrance aperture to the spectrometer selects an area of the specimen and the angle of collection is governed by the objective aperture (i.e., the same as image mode for the magnetic prism spectrometer). Electrons following a particular path through the spectrometer 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 screen. ESI has several advantages over conventional TEM images, as we'll see in Section 40.3. We will also see then that the magnetic prism, which is primarily used for spectrometry, can also be used in a STEM to form energy-filtered images.

You can also change the microscope optics and project a DP into the prism, thus producing an energy-filtered DP on the TEM screen. Then, if you 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.

The GIF (Krivanek *et al.* 1992) shown in Figure 37.11B is basically a PEELS with an energy-selecting slit after the magnet and a two-dimensional slow scan CCD array detector rather than a single line of diodes. There are also more quadrupoles and sextupoles in the optics of the GIF. The first two quadrupoles before the slit increase the dispersion of the spectrometer onto the slit and the quadrupoles after the slit have two functions. Either they project an image of the spectrum at the slit onto the CCD, or they compensate for the energy dispersion of the magnet and project a magnified image of the specimen onto the CCD (which has advantages over the diode array in a conventional PEELS). 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. 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. We'll describe energy filtering with the GIF in Section 40.3.

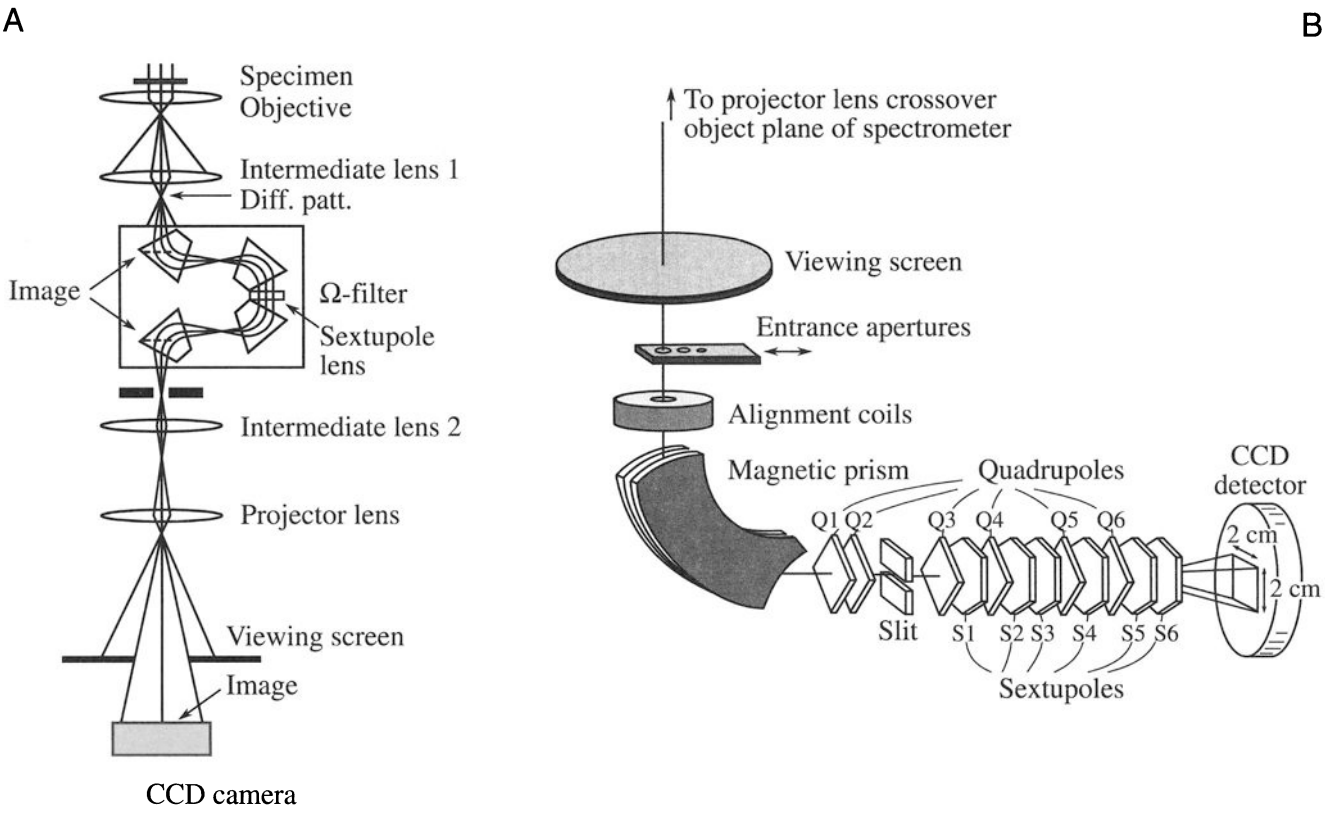


Figure 37.11. (A) Ray paths through the Ω filter system inserted in the imaging lens system of the LEO TEM. (B) The Gatan Imaging Filter attached to the TEM column after the imaging lenses, in the same position as a PEELS.

CHAPTER SUMMARY

We generally use a magnetic prism spectrometer for EELS. It is a simple device and very sensitive, but requires careful operation and an understanding of how it functions in combination with different TEM modes. PEELS is the preferred type of spectrometer, and it is best operated with the TEM in diffraction or STEM mode, or with a DSTEM. You have to know how to focus and calibrate it and how to determine the collection semi-angle, β. Once you can do this you're in a position to analyze energy-loss 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 $\mathcal{E}$.

REFERENCES

General References

Disko, M.M., Ahn, C.C., and Fultz, B., Eds. (1992) *Transmission Electron Energy Loss Spectrometry in Materials Science*, TMS, Warrendale, Pennsylvania.

Egerton, R.F. (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope,* 2nd edition, Plenum Press, New York.

Krivanek, O.L., Ed. (1991) *Microsc. Microanal. Microstruct.* **2** (2,3).

Krivanek, O.L., Ed. (1995a) *Microsc. Microanal. Microstruct.* **6** (1).

Krivanek, O.L., Ed. (1995b) *Ultramicroscopy* **59** (1–4).

Specific References

Castaing, R. and Henry, L. (1962) *C. R. Acad. Sci. Paris* **B255**, 76.

Egerton, R.F., Yang, Y.Y., and Cheng, S.Y. (1993) *Ultramicroscopy* **48**, 239.

Hillier, J. and Baker, R.F. (1944) *J. Appl. Phys.* **15**, 663.

Krivanek, O.L., Gubbens, A.J., Dellby, N., and Meyer, C.E. (1992) *Microsc. Microanal. Microstruct.* **3**, 187.

Lanio, S., Rose, H., and Krahl, D. (1986) *Optic* **73**, 56.

Metherell, A.J.F. (1971) *in Advances in Optical and Electron Microscopy,* **4** (Eds. R. Barer and V.E. Cosslett), p. 263, Academic Press, New York.

Zanchi, G., Kihn, Y., and Sevely, J. (1982) *Optik* **60**, 427.

The Energy-Loss Spectrum

38

CHAPTER PREVIEW

The term "energy-loss" spectrometry implies that we are only interested in inelastic interactions, but the spectrum will also contain electrons which have not lost any energy so we need to consider elastic scattering as well. We'll deal with three principal regions of the energy-loss spectrum:

- The zero-loss peak, which consists primarily of elastic forward-scattered electrons, but also contains electrons that have suffered minor (unresolvable) energy losses.
- The low-loss region up to an energy loss of ~50 eV contains electrons which have interacted with the weakly bound outer-shell electrons of the atoms in the specimen.
- Electrons in the high-loss region have interacted with the more tightly bound inner-shell or "core" electrons.

These different regimes of energy losses can give us different information about the specimen. The terminology is a bit vague but is generally accepted. The zero-loss peak defines the energy resolution and is essential in calibrating your spectrum. The electrons in the low-loss region have only interacted weakly with the atoms via their outer-shell electrons, so they contain information about the electronic properties of the specimen. The electrons in the high-loss region have "probed" the inner electron shells and therefore contain information characteristic of the atoms in the specimen.

We can also obtain information about how the atoms are bonded to one another, and even how the neighboring atoms are distributed around a specific atom. In principle, the energy-loss spectrum is far more useful than an XEDS spectrum. However, it is also far more complex. To understand its content you need a greater understanding of the physics of beam–specimen interactions. The spectrum also contains artifacts which we need to identify and minimize.

In this chapter we will discuss the different features of electron energy-loss spectra and go on to use these spectra in Chapters 39 and 40.

The Energy-Loss Spectrum

38

38.1. A FEW BASIC CONCEPTS

Back in Chapters 2–4 we talked about the difference between elastic and inelastic beam–specimen interactions and introduced the ideas of scattering cross sections and the associated mean-free path. It would be a good idea to remind yourself of 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 is usually manifest 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. Therefore, 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 semiangle of the spectrometer is so important.

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. 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 scattering (>1 scattering event) and multiple scattering (>20 scattering events) imply that the electron has undergone a combination of interactions.

We'll see that the energy-loss spectrum is most understandable when it represents single scattering. This ideal is approached when we have very thin specimens. 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. If multiple scattering occurs, the specimen is too thick for EELS and for much of TEM in general.

The principal inelastic interactions in order of increasing importance (and energy loss) are phonon excitations, inter- and intra-band transitions, plasmon excitations, and inner-shell ionizations. We've already introduced these processes back in Chapter 4 and we will emphasize inner-shell ionizations almost exclusively from here on. The two major characteristics of any inelastic scattering are the energy loss $\mathcal{E}$ and the scattering semiangle θ, and we summarize typical values in Table 38.1.

It's a little difficult to be specific about the values of the scattering angle because the angle varies with energy. In fact there are different definitions of scattering angle which you may come across, and these can be confusing. You can find derivations of the equations governing scattering in Egerton (1996).

> The symbol θ in all cases refers to the scattering semiangle.

We will always assume that the scattering is symmetrical around the direct beam. The most important angle is θ_E, the so-called characteristic or most-probable scattering semiangle for an energy loss, $\mathcal{E}$. This angle is given by

$$\theta_E \approx \frac{\mathcal{E}}{2E_0} \qquad [38.1]$$

This equation is an approximation and it ignores relativistic effects, so you should only use it for rough calculations at and above 100 keV. We can be more precise and define θ_E as

Table 38.1. Characteristics of the Principal Energy-Loss Processes

Process	Energy loss (eV)	θ_E (mrads)
Phonons	~0.02	5–15
Inter/intra-band transitions	5–25	5–10
Plasmons	~5–25	<~0.1
Inner-shell ionization	~10–1000	1–5

$$\theta_E \approx \frac{\mathcal{E}}{(\gamma m_0 v^2)} \qquad [38.2]$$

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

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \qquad [38.3]$$

The electron velocity is v and c is the velocity of light. One other useful angle, θ_C, is the cut-off angle above which the scattered intensity is zero, and this is given by

$$\theta_C = (2\theta_E)^{\frac{1}{2}} \qquad [38.4]$$

In Table 38.1 we have given some typical values of θ_E. This is the scattering angle that we'll usually refer to from now on. Let's now move on to the energy-loss spectrum. We'll start at the low-energy end and proceed to higher-energy losses.

38.2. THE ZERO-LOSS PEAK

If your specimen is thin, the predominant feature in the energy-loss spectrum will be the zero-loss peak. As the name implies, this peak consists mainly of electrons that have completely retained the 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 we were to tilt the incident beam so a diffracted beam entered the spectrometer, then it too would give a zero-loss peak. The scattering angles for diffraction ($2\theta_B$) are relatively large (~20 mrad) compared to the smaller collection angles in EELS, and so the diffracted beams rarely enter the spectrometer. Actually, we can also measure the intensity and energy of electrons as a function of their angular distribution, and we'll discuss this aspect briefly in Chapter 40.

Now the term "zero-loss peak" is really a misnomer for two reasons. First, our spectrometers have a finite energy resolution (at best ~0.3 eV) so the zero-loss peak will also contain electrons that have lost very small amounts of energy, mainly those that excited phonons. So in EELS in the TEM we never resolve phonon losses. This is not a "great loss" since phonon-loss electrons don't carry any useful information anyway; they only cause the specimen to heat up. However, it does explain why we shouldn't really call this the zero-loss peak. Second, we can't produce a beam of monochromatic electrons; the beam has a finite energy range about the nominal value E_0. Despite this imprecision, we will continue to use the zero-loss terminology.

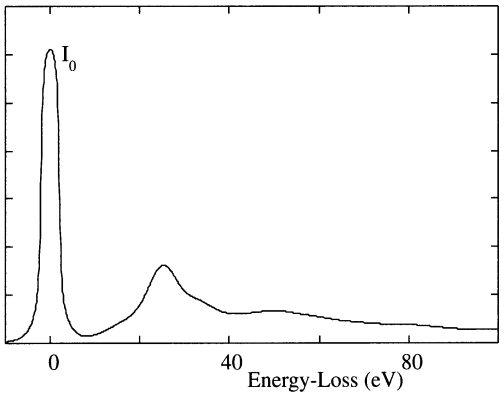


Figure 38.1. The intense zero-loss peak I_0 in a spectrum from stainless steel. The rest of the spectrum comprises energy-loss electrons which constitute a relatively small fraction of the total intensity in the spectrum.

The zero-loss peak is usually a problem rather than a useful feature in the spectrum, because it is so intense that it can damage the scintillator or saturate the photodiode array. We don't collect it except under certain circumstances. Figure 38.1 shows the intense zero-loss peak in a spectrum. To the right of the peak is a relatively small peak, which is part of the low-loss spectrum. This small peak is where we start to get useful information, but you can also see immediately that the useful part of the spectrum is very much less intense than the somewhat useless zero-loss peak, and this is one of several fundamental problems in EELS.

38.3. THE LOW-LOSS SPECTRUM

We use the term "low-loss" to describe energy-loss electrons in the range up to about 50 eV. In this part of the spectrum we come across electrons that have set up plasmon oscillations or have generated inter- or intra-band transitions. Plasmons are by far the most important, so we'll look at these first.

38.3.A. Plasmons

Plasmons are longitudinal wave-like oscillations of weakly bound electrons. The oscillations are rapidly damped, typi-

Table 38.2. Plasmon Loss Data for 100-keV Electrons for Several Elements

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

cally 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 zero-loss peak. The small peak beside the zero-loss peak in Figure 38.1 is a plasmon peak.

The energy $\mathcal{E}_P$ lost by the beam electron when it generates a plasmon of frequency ω_p is given by

$$\mathcal{E}_P = \frac{h}{2\pi}\omega_P = \frac{h}{2\pi}\left(\frac{ne^2}{\varepsilon_0 m}\right)^{\frac{1}{2}} \qquad [38.5]$$

where h is Planck's constant, e and m are the electron charge and mass, ε_0 is the permittivity of free space, and n is the free-electron density. Typical values of $\mathcal{E}_P$ are in the range 5–25 eV and a summary is given in Table 38.2.

Plasmon losses dominate in materials with free-electron structures, such as Li, Na, Mg, and Al, but occur to a greater or lesser extent in all materials.

We even see a plasmon-like peak in spectra from materials with no free electrons (such as polymers) for reasons that are not well understood. From equation 38.5 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. So in principle, measurement of the plasmon energy loss can give indirect microanalytical information, as we'll see later in Section 40.2. Plasmon-loss electrons also carry contrast information and therefore are important because they limit image resolution through chromatic aberration. We can remove them from the image by energy filtering, as we'll also describe in Section 40.3.

Because of the low values of λ_P, the characteristic scattering angles θ_E are very small, being typically <0.1 mrad (as listed in Table 38.2). So, plasmon-loss electrons are strongly forward-scattered. Their cut-off angle $\theta_C \sim 100\ \theta_E$. Hence if you use a β of only 10 mrad, you will gather virtually all the plasmon-loss electrons. Also, their line width $\Delta\mathcal{E}_P$ is at most a few eV.

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 the specimen. Figure 38.2 shows the plasmon-loss spectra from thin and thick foils of pure Al. 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 a most important phenomenon because it eventually limits the interpretation of part of the spectrum containing chemical information from ionization losses in which we are really interested (see Section 38.4). The well-known properties of plasmon

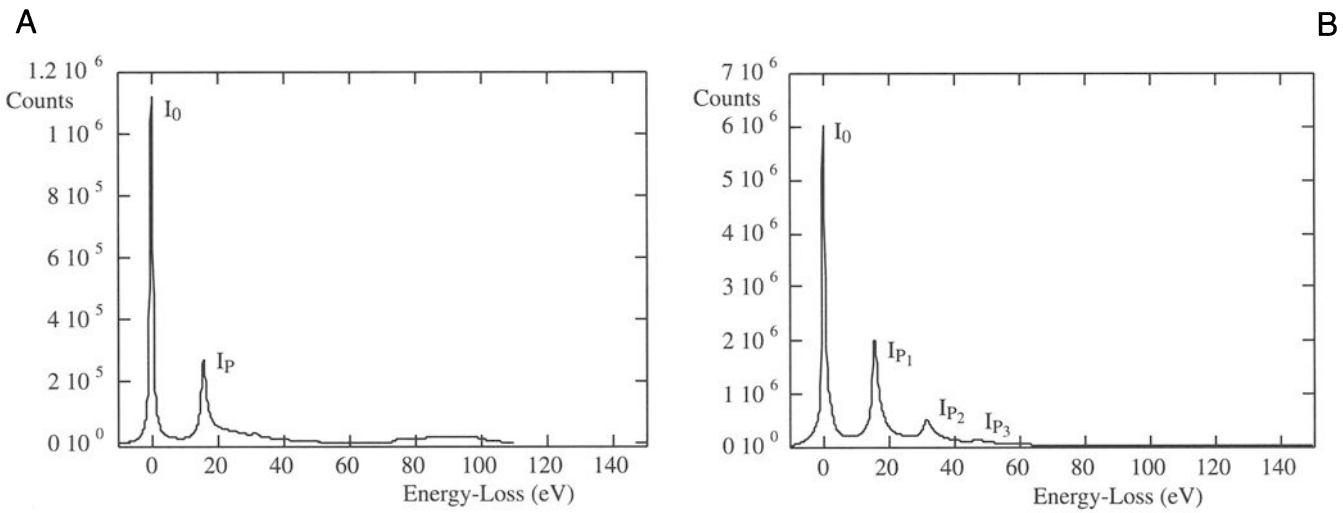


Figure 38.2. (A) The low-loss spectrum from a very thin sample of pure Al showing the intense zero-loss peak (I_0) and a small plasmon peak (I_P) at about 15 eV. (B) The low-loss spectrum from a thicker specimen of pure Al showing several plasmon peaks.

loss electrons from several elements are summarized in Table 38.2.

The plasmon losses which we've just described all arise from interactions with the electrons in the interior of the specimen, but the incident electrons can also set up plasmon oscillations on the surface of the specimen. We can envisage these surface plasmons as transverse charge waves. Surface plasmons have about half the energy of bulk plasmons. Generally, however, the surface plasmon peak is much less intense than the volume plasmon peaks, even in the thinnest specimens.

38.3.B. Inter- and Intra-Band Transitions

An electron in the beam may transfer sufficient energy to a core electron to cause it to change its orbital state, for example, to a Bohr orbit of higher quantum number. We call these events "single electron interactions" and they result in energy losses of up to ~25 eV. Interactions with molecular orbitals such as the π orbitals produce characteristic peaks in this low-energy region of the spectrum, and it is possible sometimes to use the intensity variation in this part of the spectrum to identify a particular specimen. However, the details of the spectrum intensity variations due to single electron interactions are not well understood and cannot yet be predicted *a priori.*

> Use of the low-loss spectrum for phase identification is only possible through a "fingerprinting" process by which the low-loss spectra of known specimens are stored in a library in the computer.

Spectra from unknown specimens may then be compared with the stored library standards. Figure 38.3 shows the low-loss spectra of Al and Al-containing compounds exhibiting differences in the detailed intensity variation. A collection of low-loss spectra from all the elements has been compiled in the EELS Atlas (Ahn and Krivanek 1983) and this can help with "fingerprinting" unknown specimens.

If the beam electron gives a weakly bound valence-band electron sufficient energy to escape the attractive field of the nucleus, then we've created a secondary electron (SE), of the sort used to give topographic images in the SEM and STEM. Typically, we give <20 eV to a SE and therefore the electrons causing SE emission appear in the same low-energy region of the spectrum as the inter- and intra-band transitions.

The weakly bound outer-shell electrons control the reaction of an atom to an external field and thus control the dielectric response of the material. We'll see in Chapter 40 that it is possible to get a measure of the dielectric constant by careful processing of the very low loss portion (<~10 eV) of the spectrum.

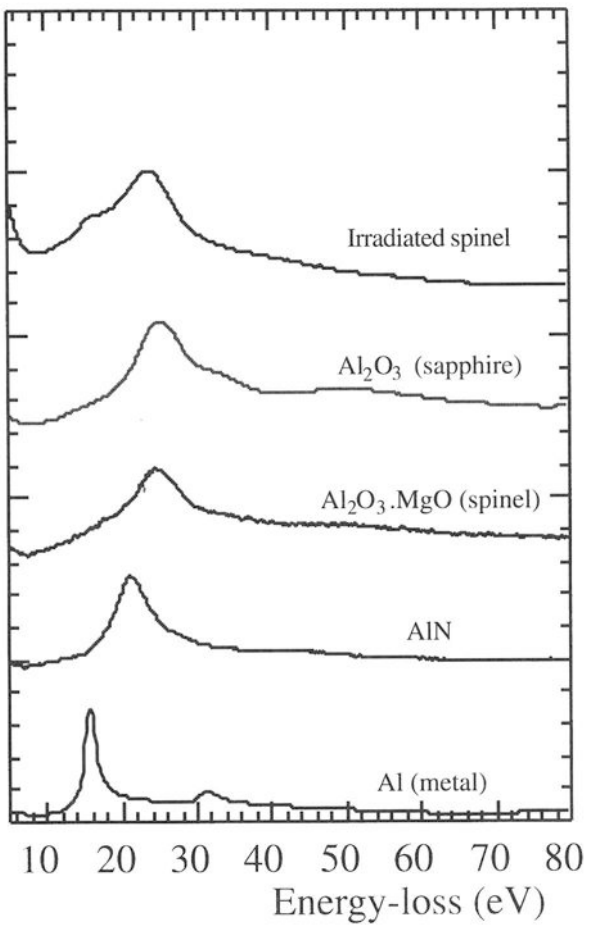


Figure 38.3. The low-loss spectrum from specimens of Al and Al-containing compounds, showing differences in intensity that arise from differences between the bonding in the different materials. The spectra are displaced vertically for ease of comparison.

38.4. THE HIGH-LOSS SPECTRUM

The high-loss portion of the spectrum above about 50 eV contains information from inelastic interactions with the inner or core shells.

38.4.A. Inner-Shell Ionization

When a beam electron transfers sufficient energy to a K, L, M, N, or O shell electron to move it outside the attractive field of the nucleus, as shown back in Figure 4.2, the atom is said to be ionized. As you know from the earlier chapters on X-ray analysis, the decay of the ionized atom back to its ground state may produce a characteristic X-ray, or an Auger electron. So the processes of inner-shell ionization-

loss EELS and XEDS are different aspects of the same phenomenon. 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 elemental information, just like the characteristic X-ray. We call the ionization-loss signal an "edge" for reasons we'll describe shortly.

You should appreciate that detection of the beam electron that ionized the atom is independent of whether the atom emits an X-ray or an Auger electron. EELS is not affected by the fluorescence-yield limitation that restricts light-element X-ray analysis. This difference explains, in part, the complementary nature of XEDS and EELS.

Inner-shell ionization is generally a 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 ~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, as in XEDS, we tend to look for other lower-energy ionizations, such as the L and M edges, when dealing with high-Z atoms. Typically, we start to use the L edges when the K-shell energy exceeds ~1 keV (Na) and M edges when the L shell exceeds ~1 keV (Zn).

It's worth a short mention here about the nomenclature used for EELS edges. Just like in 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 greater energy resolution of the EELS spectrometer means that it is much easier to detect differences in spectra that arise from the presence of different energy states in the shell. For example:

- The K-shell electron is in the 1s 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, then we get an L_1 edge, and a 2p electron causes either an L_2 or L_3 edge.

The L_2 and L_3 edges may not be resolvable at lower ionization energies (e.g., they aren't in Al but they are in Ti), and sometimes we call this edge the $L_{2,3}$. The full range of possible edges is shown schematically in Figure 38.4, and you can see that other "dual" edges exist, such as the $M_{4,5}$. There will be more about this in Chapter 40.

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 very much smaller than the plasmon peak, and becomes even smaller as the energy loss increases (look back to Figure 37.4A). 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 with a plasmon loss is by no means uncommon, and this phenomenon distorts the resultant spectrum.

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 chances of ionization occurring become less with increasing energy above E_C, because the value of the cross section decreases with increasing energy. 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."

This idealized triangular or saw-tooth 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.

You'll notice that this edge, shown in Figure 38.5A, has almost the same intensity profile as the "absorption edges" in X-ray spectroscopy. In reality, because we aren't dealing with isolated atoms but atoms integrated into a crystal lattice or amorphous structure, the spectra become more complex. The ionization edges are superimposed on a rapidly decreasing background intensity from electrons that have undergone random, plural inelastic scattering events (Figure 38.5B). The edge shape may also contain fine structure around E_C (Figure 38.5C) which is due to bonding effects, and is termed energy-loss near-edge structure (ELNES). More than ~50 eV after the edge, small intensity oscillations may be detectable (Figure 38.5D) due to diffraction effects from the atoms surrounding the ion-

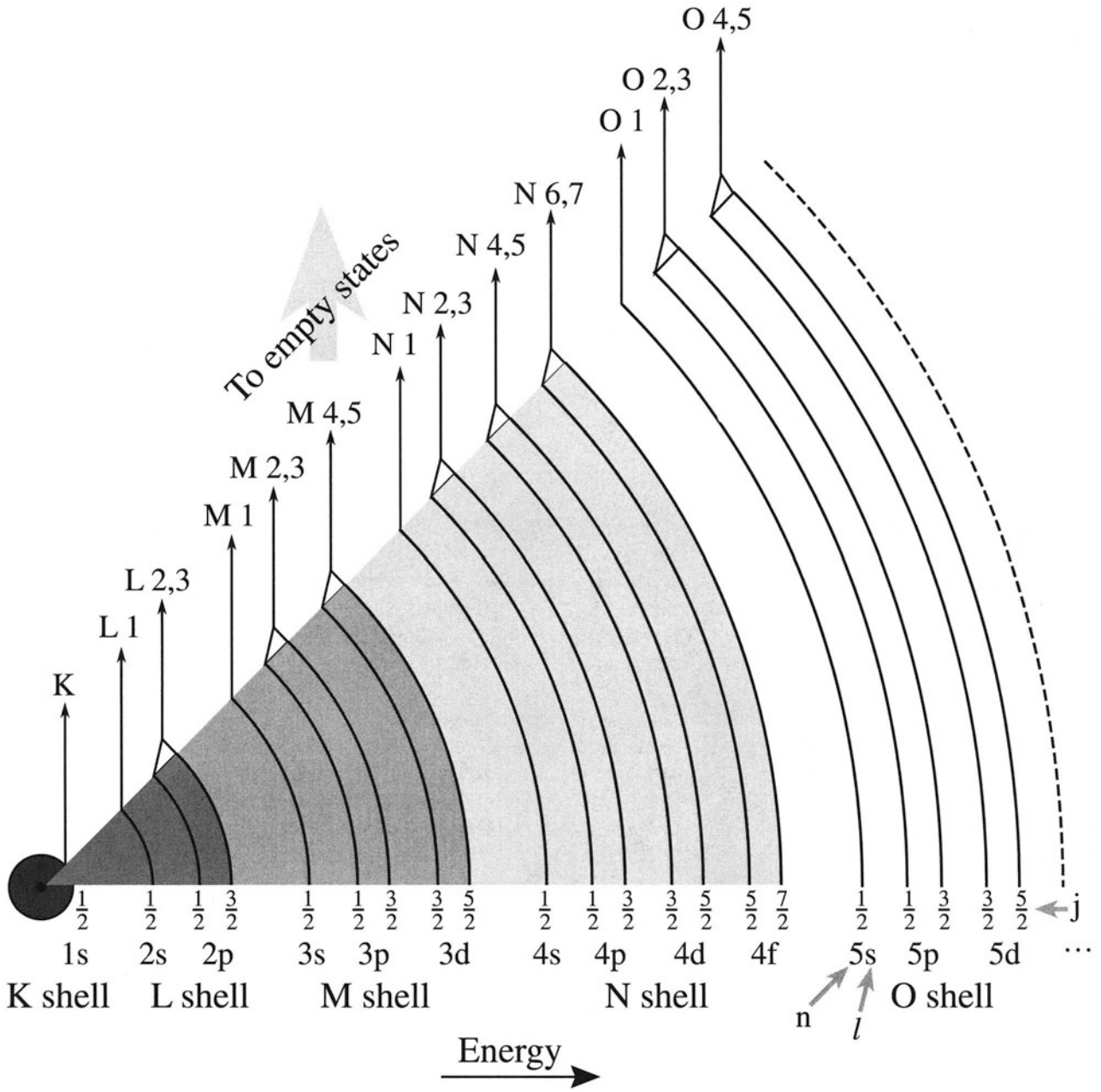


Figure 38.4. The full range of possible edges due to inner-shell ionization, and their associated nomenclature.

ized atom, and these oscillations 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.

- Fine structure before or around the peak 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 further low-loss interactions. They may create plasmons, in which case the ionization edge contains plural scattering intensity ~15–25 eV above E_C, as shown in Figure 38.5E. So the resultant ionization edge is far more complicated than the simple Gaussian peak seen in an XEDS spectrum. Clearly, the edge details contain far more information about the specimen than a characteristic X-ray peak. From an X-ray spectrum you only get *elemental* identification rather than *chemical* in-

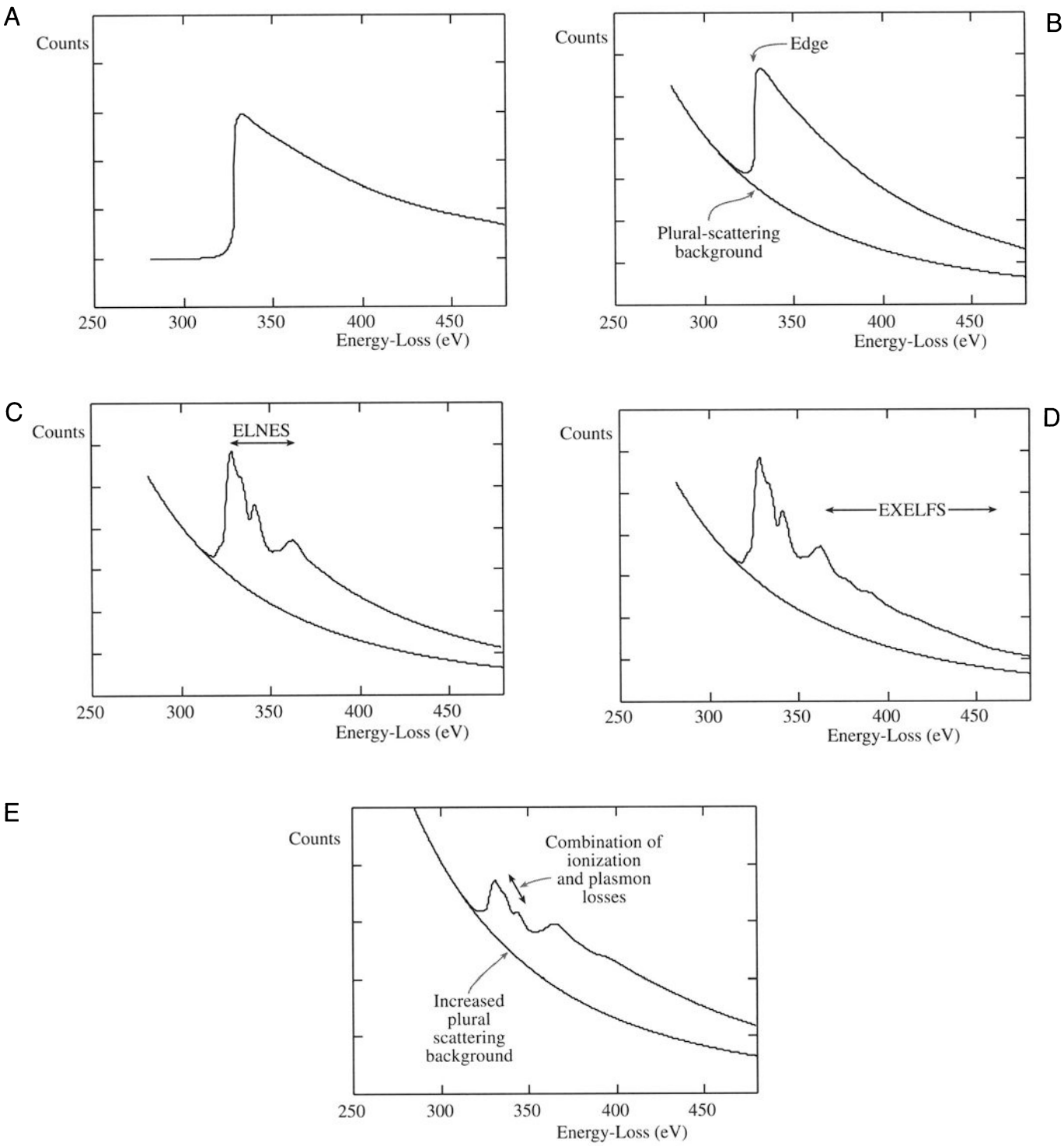


Figure 38.5. The characteristic features of an inner-shell ionization edge: (A) the idealized saw-tooth (hydrogenic) edge, (B) the edge superimposed on the background arising from plural inelastic scattering, (C) the presence of ELNES, (D) the EXELFS. (E) Plural scattering in a thick specimen, such as the combination of ionization and plasmon losses, distorts the post-edge structure and give an increase in the background level.

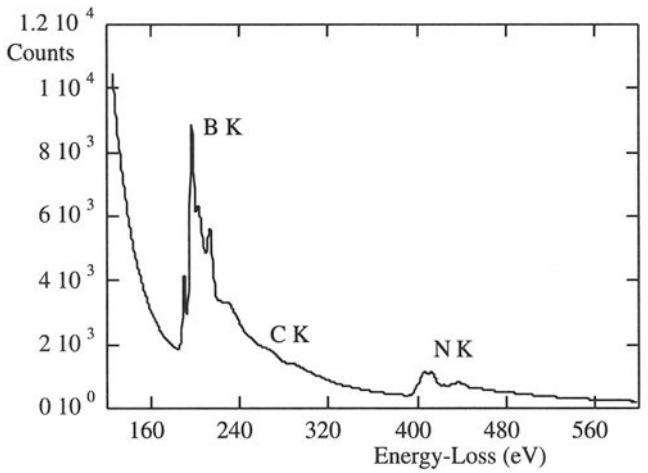


Figure 38.6. High-energy-loss spectrum from a particle of BN over a hole in a C film showing the B and N K-shell ionization edges superimposed on a rapidly decreasing background. A faint C K edge is also visible at ~280 eV.

formation, such as bonding, which is contained in the ELNES. Figure 38.6 shows a spectrum from BN on a C film. The various ionization edges show some of the features drawn schematically in Figure 38.5; we'll discuss these "fine structure" effects more in Section 40.1.

38.4.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 we have relatively large values of E_C compared to $\mathcal{E}_P$:

- θ_E ~5 mrad for ionization-loss electrons at E_C = 1000 eV, for a beam energy of 100 keV.
- The average plasmon-loss scattering was broadened to ~10–15 mrad.

The characteristic scattering angles for both plasmon and inner-shell ionization are still much lower than the characteristic scattering angles for phonon and elastic scattering. The angular distribution varies depending on the energy loss, and because of the extended energy range of ionization-loss electrons above E_C, this can be quite complicated. For $\mathcal{E} \sim E_C$ the intensity drops rapidly to zero over about 10 mrad at θ_C, 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. Since this effect is irrelevant for EELS studies, we'll ignore it, but you can find more information in Egerton (1996).

So, in the region immediately following the ionization edge the angular distribution of the electrons is generally confined to a semiangle of <10–15 mrad and drops to zero beyond this. In other words, like the plasmon-loss electrons, the ionization-loss electrons are very strongly forward-scattered. Consequently, efficient collection of the major inelastically scattered electrons is a straightforward matter, since a spectrometer entrance aperture semiangle (β) of <20 mrad will collect the great majority of these electrons. As a result, collection efficiencies in the range 50–100% are not unreasonable, which contrasts with the situation in XEDS, where the isotropic generation of characteristic X-rays results in very inefficient collection. Figure 38.7 compares the collection of X-rays and energy-loss electrons in the AEM. Figure 38.8 shows the variation in collection efficiency for ionization-loss electrons as a function of both β and energy.

While the K edges in Figure 38.6 show reasonably 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 specific edge types and specific shapes. The situation is further complicated by the fact that the edge shapes change significantly depending on whether or not

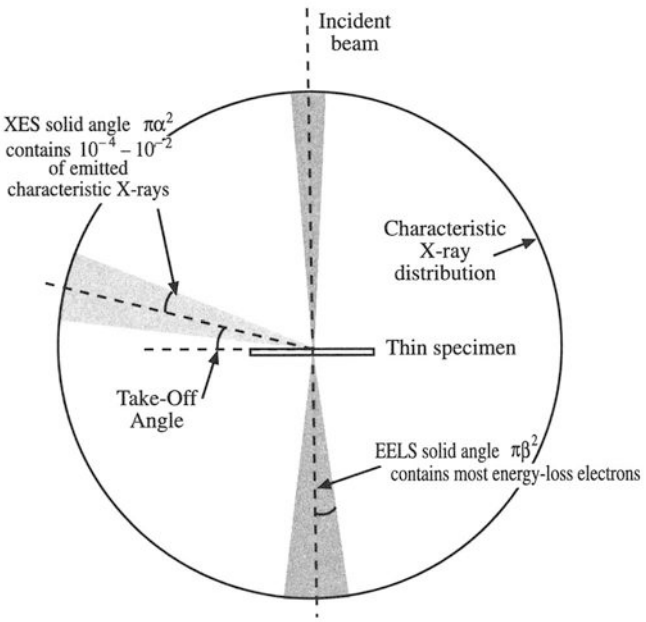


Figure 38.7. Comparison of the relative efficiencies of collection of EELS and XEDS. The forward-scattered energy-loss electrons are more efficiently collected than the uniformly emitted characteristic X-rays.

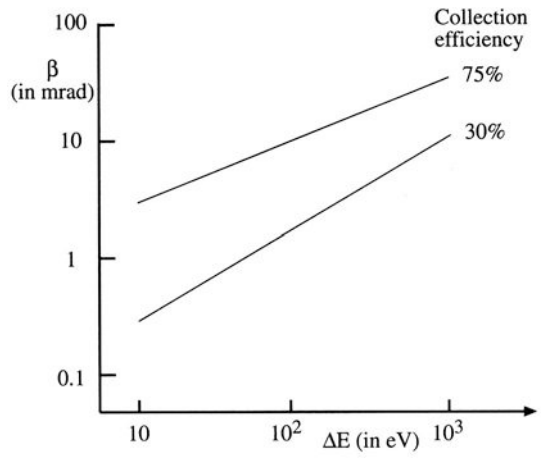


Figure 38.8. Variation in the collection efficiency of ionization-loss electrons as a function of their energy and the spectrometer collection semiangle, β.

certain energy states are filled or unfilled. For example, if you look below at Figure 38.9, 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, which are present in Ni. In contrast, in Cu, in which 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 details of the fine structure and edge shapes are also affected by bonding. For example, the Ni edge in NiO in Figure 38.9 is different from the Ni edge in pure Ni. To sort all this out it's best if you consult the EELS Atlas (Ahn and Krivanek 1983), which contains representative edges from all the elements and many oxides.

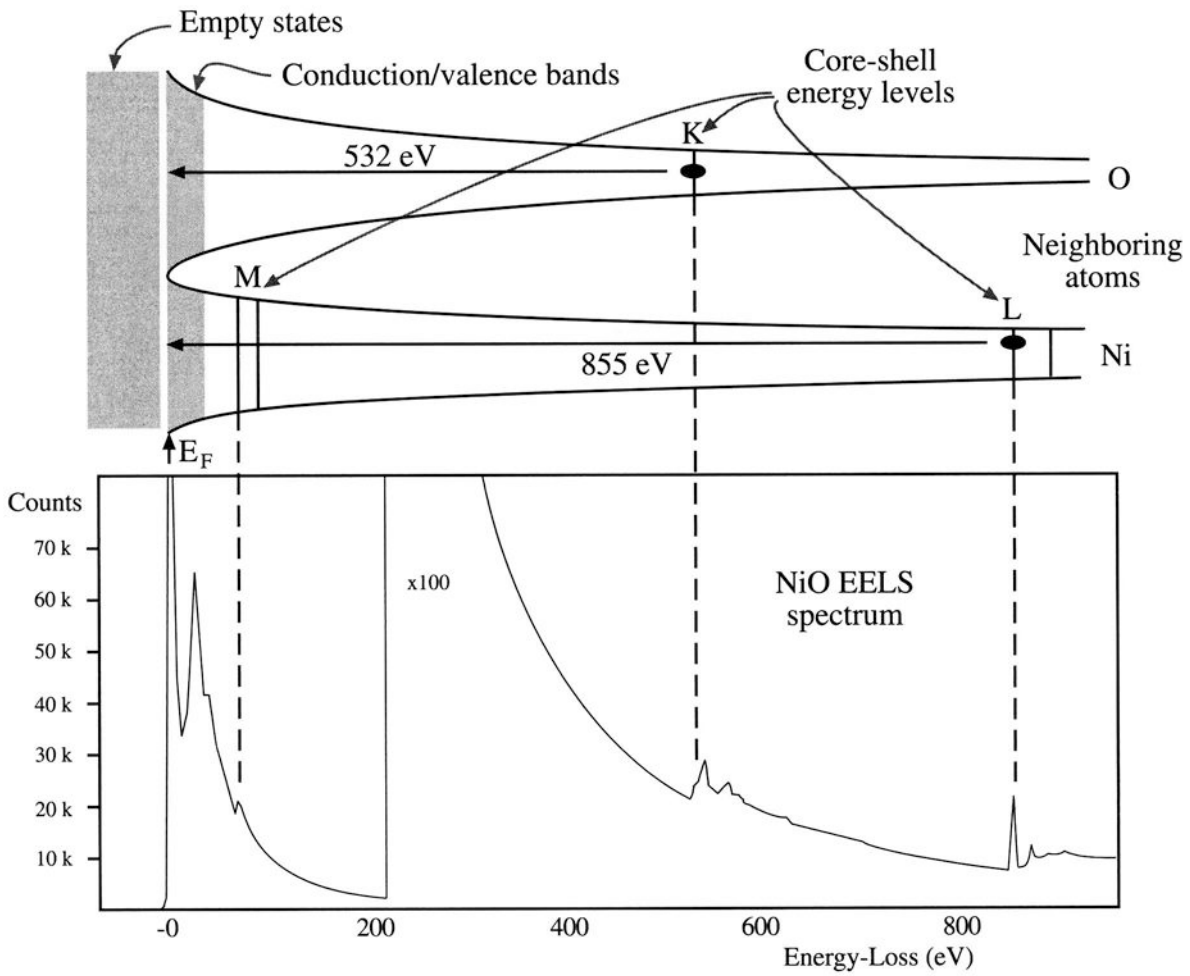


Figure 38.9. The correspondence between the energy levels of electrons surrounding adjacent Ni and O atoms and the energy-loss spectrum: the zero-loss peak is above the Fermi energy E_F, the plasmon peak is at the energy level of the conduction/valence bands, and the critical ionization energy required to eject specific K-, L-, and M-shell electrons is shown.

We can summarize the characteristics of the energy-loss spectrum by showing a complete spectrum from NiO containing both low- and high-loss electrons, as shown in Figure 38.9. In this figure we also compare the spectrum to the energy-level diagram for NiO. You can see that:

- The plasmon peak corresponds to the energy of the valence electron band just below the Fermi level (E_F).
- The relative energy levels of the ionized atom (K, L, or M) control the position of the ionization edge in the spectrum.
- The different density of states in the valence (3d) band of the Ni atom is indicated by shading at the top of the potential wells and is reflected in the characteristic, intense, near-edge, fine structure at the Ni L edge.

The electrons could also be given sufficient energy to travel into the conduction band well above E_F; as we just mentioned, in this case we see extended fine structure 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, which are the major signal for elemental analysis, show relatively low intensity, have an extended energy range above the ionization energy, and ride on a rapidly varying, relatively high background. All these factors, as we shall see, combine to make quantitative microanalysis using EELS a difficult and less accurate technique when compared with XEDS. However, for the light 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 simple, but below oxygen in the periodic table, EELS has shown better performance than XEDS and, for elements below boron, there is no sensible alternative to EELS for microanalysis at high spatial resolution.

38.5. ARTIFACTS IN THE SPECTRUM

The SEELS spectrum contains no artifacts of any consequence, unless it is grossly misaligned, in which case the beam may scatter through the slits or off the drift tube, giving distorted background intensities. These effects are easy to spot and correct.

Unfortunately, the highly efficient PEELS system generates more artifacts which you have to recognize and remove before analyzing the spectrum. Details are available in the Gatan manual, but here we'll summarize the major problems (which are in addition to the point-spread function that we talked about in the previous chapter).

> All 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. This will be characteristic of each individual diode array.

One way to determine the gain variation is to spread the beam uniformly over the array using at least the 3-mm entrance aperture and looking at the diode readouts, as shown in Figure 38.10. This is difficult with an FEG system because the probe is too small, and then it is necessary to scan the beam across the array, although this is not very satisfactory. Then you have to divide your experimental spectrum by this response spectrum to remove the gain variation. Alternatively, and this is recommended, you can gather two or more spectra with slight energy shifts (~1–2 eV) or spatial shifts between them and superimpose them electronically. The gain variation then disappears, as you can see if you look at Figure 38.12. Using a two-dimensional array, as in the GIF, removes this problem also.

Gathering many spectra and superimposing them can bring another problem, namely, that of readout noise. There are two kinds of readout noise, random and fixed. The random readout noise, or shot noise, arises from the electronics chain from the diode to the display, and is minimized by taking as few readouts as possible, and also by

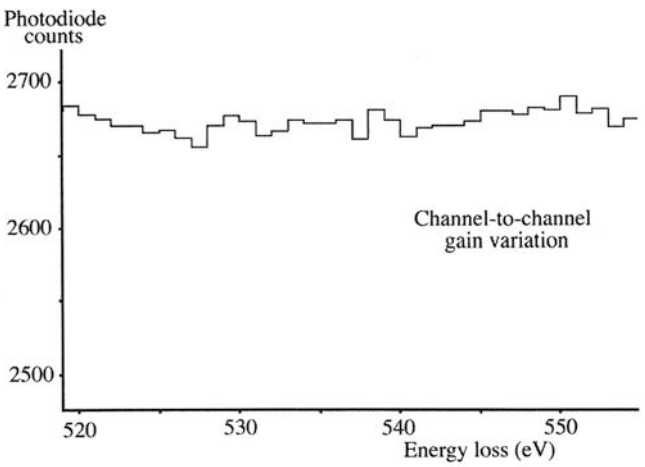


Figure 38.10. The variation in the response of individual diodes in the PEELS detection system to a constant incident electron intensity. A channel-to-channel gain variation is clear and each detector array has its own characteristic response function.

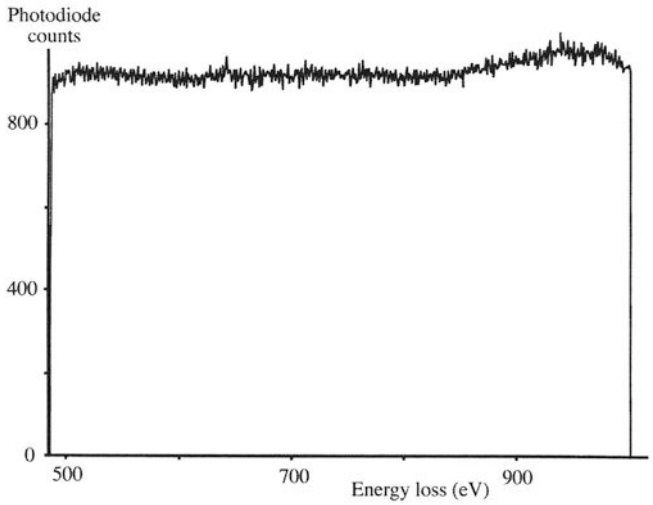


Figure 38.11. The intensity of the dark current which flows from the diode array when no electron beam is present.

cooling the diode array. Individual diodes may have high leakage currents which give a spike on the display. The fixed pattern readout 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 38.11). The dark current is small unless you have a bad diode array, and it is only a problem when there are very few counts in your spectrum or you have added together 10 or more spectra. Figure 38.12 shows some of these effects and how to remove them.

Finally, there is the problem of incomplete readout of the display. When the diodes are cooled, only ~95% of the signal is read out in the first integration, ~4.5% in the second, ~0.25% in the third, and so on. This is only a problem if you have saturated the diodes with an intense signal like the zero-loss peak. This 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 should never confuse a ghost peak with a genuine edge.

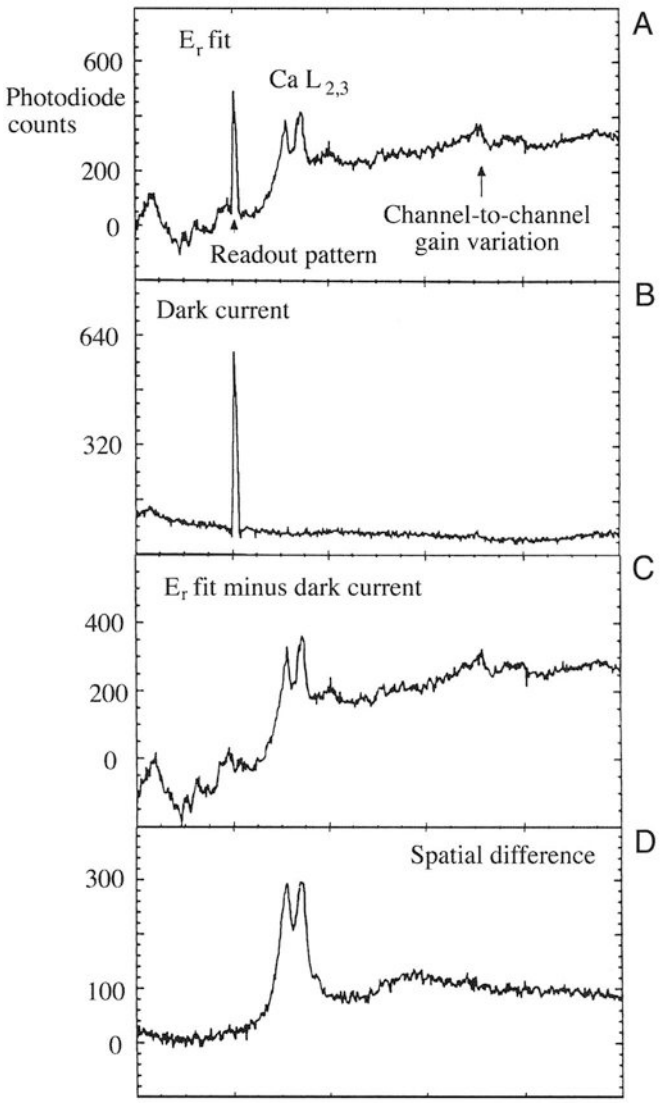


Figure 38.12. How to remove artifacts from a specimen: (A) A Ca $L_{2,3}$ edge spectrum showing both channel-to-channel gain variation and a faulty diode with a high leakage current which appears as a spike in the spectrum. The spike is referred to as the readout pattern and is present in every recorded spectrum. Subtracting the dark current (B) removes the spike (C) and a difference spectrum (D) removes the gain variation, leaving the desired edge spectrum.

CHAPTER SUMMARY

The EELS spectrum varies in intensity over several orders of magnitude.

- The least useful signal (the zero-loss peak) is the most intense, and the most useful signals (the ionization edges) are among the least intense signals.
- The low-loss spectrum reflects beam interactions with loosely bound conduction and valence-band electrons.

Table 38.3. PEELS Artifacts and How to Eliminate Them

Noise name	Source	Elimination
Leakage current	Different diodes	Subtract dark current
Internal scanning noise	Electronics readout	Adjust the electronics and subtract the dark count
Nonuniform sensitivity	Diode sensitivities differ	Determine the response characteristic by sweeping the beam along the array and divide the real spectrum by this result, i.e., normalize the diodes

- The high-loss spectrum contains small ionization edges riding on a strong plural-scattered background.
- Differences in the energy onset of the ionization edges distinguish different elements in the specimen.
- Differences in the fine structure of the edges reflect chemical (bonding) effects and structural (atomic arrangement) effects.
- Artifacts can complicate spectrum interpretation, but they are well understood and easily removed.

We summarize the different sources of noise and how we eliminate this noise in Table 38.3.

REFERENCES

Specific References

Egerton, R.F. (1996) *Electron Energy Loss Spectroscopy in the Electron Microscope,* 2nd edition, Plenum Press, New York.

Ahn, C.C. and Krivanek, O.L. (1983) *EELS Atlas,* Gatan Inc., 780 Commonwealth Drive, Warrendale, Pennsylvania.

Microanalysis with Ionization-Loss Electrons

39

CHAPTER PREVIEW

In the previous two chapters we've described how to acquire an energy-loss spectrum and have also given you some idea of the information in such spectra. Most importantly, there are elemental composition data which can be extracted primarily from the high-loss ionization edges. In this chapter we'll examine how to get this information and quantify it. As we've already indicated, the prime use for these kind of data is light-element microanalysis, where 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 for microanalysis. Next, we'll discuss the various quantification routines which, in principle, are just as straightforward as those for XEDS but in practice require a rather more sophisticated level of knowledge to carry them out successfully. Finally, we'll say a bit about spatial resolution and minimum detectability, although these topics aren't as important in EELS as they are in XEDS.

Microanalysis with Ionization-Loss Electrons

39

39.1. CHOICE OF OPERATING PARAMETERS

Perhaps a major reason why EELS is not as widespread as XEDS is the relative complexity of the experimental procedure and the number of variables which you have to define before you can get started. EELS is not yet a "turn-key" operation so you cannot simply place your specimen under the beam, switch on the spectrometer and the computer, and acquire a spectrum. This is in marked contrast to the situation in XEDS, where the high degree of software control means that the XEDS system is almost invariably ready to go when you push the "acquire" button. Furthermore, as you'll see, little useful information is present in the acquired EELS spectrum unless your specimen is very thin. Disko (1986) succinctly summarized the important experimental variables. We've already told you back in Chapter 37 how to control most of these factors. In this chapter, we'll go through all the parameters and indicate reasonable values for each.

- *Beam energy E_0:* It's probably best to use the highest E_0, unless doing so causes displacement damage or surface sputtering. A higher E_0 does reduce the scattering cross section and so you get reduced edge intensity. However, as E_0 increases, the plural-scattering background intensity falls faster than the edge intensity and so the ionization-edge signal-to-background increases and this is useful. The increase in signal-to-background varies with the particular edge but it is never a strong variation; so while we make a lukewarm recommendation to use the highest kV, it's not a good reason to justify purchasing an IVEM.
- *Convergence semiangle α:* You know how to control α with the C2 aperture and/or the C2 lens, but α is only important in the quantification process 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 any effects of α; otherwise, use the correction factor we give 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 limiting the spatial resolution in STEM mode, and the beam current controls the signal intensity. You have to make the same compromise between improved spatial resolution and loss of signal intensity, or vice versa, as we discussed at some length in Chapter 36 for XEDS.
- *Specimen thickness:* The specimen must be thin because then the plural-scattering contributions to the spectrum are minimized and quantification is most straightforward.

> 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 how to determine the thickness from your spectrum and how to decide if you need to deconvolute the spectrum.

- *Collection semiangle 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 limited spatial resolution, use TEM-image mode with no objective aperture ($\beta \sim 100$ mrad). A small spectrometer entrance aperture would provide better 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 a small spectrometer entrance aper-

ture for good energy resolution. In the STEM case you also get good spatial resolution.

> Remember that a 5-mm entrance aperture gives β ~5 mrad at a camera length of ~800 mm.

Generally, for microanalysis β ~1–10 mrad is fine, so long as it's less than the Bragg angle for your particular specimen and orientation; but for EELS imaging, which we discuss in Section 40.3, 100 mrad may be necessary.

- *Energy resolution:* ΔE is limited by your electron source, assuming you've focused the spectrum. In a SEELS, the slit width can control ΔE. Microanalysis and imaging do not require the best ΔE and ~5 eV will generally suffice. You really only need the best ΔE for ELNES, and plasmon-shift studies, both of which are somewhat esoteric pursuits. Use an FEG source and a PEELS if you want to do this kind of thing.
- *Energy-loss range and spectrum dispersion:* The full spectrum extends out to the beam energy E_0, but the useful portion only extends to about 1 keV. Above this energy loss, the intensity is very low, and microanalysis by XEDS is both easier and more accurate, although arguably a little less sensitive. So you rarely need to collect a spectrum above about 1 keV and therefore, with a minimum of 1024 channels in the MCA display, 1 eV/channel is always a good starting point. 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 $\Delta E < 1$ eV.
- *Signal processing:* In SEELS, remember that your two choices (in which the spectral intensity is determined by the total current on the scintillator) are analog processing or single electron counting. You should collect the high-intensity, low-loss portion of the spectrum in analog mode and the lower-intensity, high-loss region in single-electron mode. The change in counting mode is most conveniently made at the same point in the collection process as the gain change. Set this point around 50–100 eV, well above the plasmon range but at lower $\mathcal{E}$ than most ionization edges. There is no equivalent of this variable in PEELS.
- *Dwell time:* In SEELS, typical dwell times are in the range from 10 ms to 1 s per channel, depending on the number of channels in the spectrum and the intensity necessary to extract the analytical result. Because the magnetic prism is not very stable, it is unwise to collect spectra for periods longer than a few minutes. If more counts are required you should sum several spectra (see below), each recorded over a limited time range, with intermediate checks on the calibration. In PEELS, you set the dwell time (or integration time) such that the maximum intensity in the spectrum doesn't saturate the photodiode array, 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 sweeps:* It is better in both SEELS and PEELS to sum many spectra rather than gather one SEELS spectrum for several minutes, or saturate the PEELS detector. Remember also that each sweep in SEELS should be a reverse scan from high to low energy. Furthermore, if the intense zero loss has to be recorded, several minutes should elapse between each scan to ensure that the scintillator after-glow has subsided to below the normal dark-current output of the detector. In PEELS, multiple acquisitions can give rise to artifacts, as we discussed in Section 38.5.

So now you can see why EELS is not a straightforward turn-key operation. You must have a very good understanding of your TEM and spectrometer optics; be aware that the system is not very stable, needs recalibrating regularly and, PEELS, particularly, is prone to artifacts.

39.2. WHAT SHOULD YOUR SPECTRUM LOOK LIKE?

Before you analyze a particular spectrum, you should check three things:

- Display the zero-loss peak to ensure that the spectrometer is giving you the necessary ΔE, if this is important.
- Look at the low-loss 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.

The first of these tasks is not critical, as we noted earlier. Regarding the second task, you'll see in Section 39.5 that, to a first approximation, if the plasmon peak intensity is less than about one-tenth the zero-loss peak, then the specimen is thin enough for microanalysis. Otherwise, you'll

probably have to deconvolute plural-scattering effects from your experimental spectrum. For the third task, you should ideally see a discrete edge on a smoothly varying background, but you need to see at least a change in slope in the background intensity at any expected edge energies. The Gatan ELP program can identify and quantify such hidden peaks (see Section 1.5). If the background intensity is noisy, it will make quantification more difficult.

> An important parameter in determining the quality of your spectrum is the signal-to-background ratio which, in EELS, we call the jump ratio.

This 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 39.1.

If the jump ratio is above ~5, for the carbon K edge at 284 eV from a standard (< 50 nm thick) carbon film at 100 kV, then your TEM-EELS system is operating satisfactorily. Figure 39.1 is a well-defined edge from a film of amorphous carbon. If you can't get such a jump ratio, then perhaps you need to realign the spectrometer, or find a thinner specimen. The jump ratio should increase with increasing kV.

39.3. QUALITATIVE MICROANALYSIS

As with XEDS, you should always carry out qualitative microanalysis to ensure that you have identified all the features in your spectrum. Then you can decide which edges to use for microanalysis.

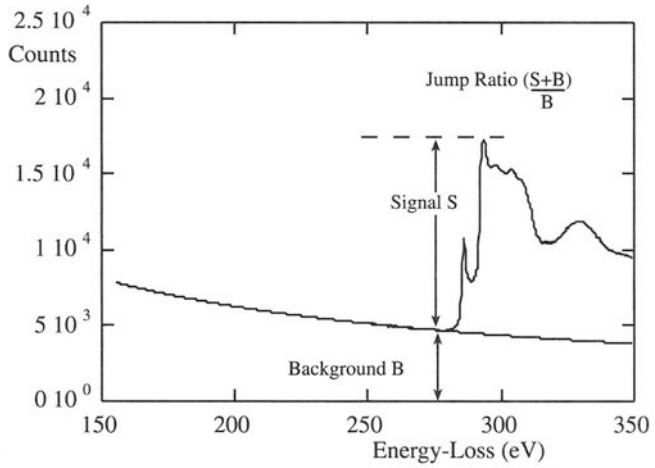


Figure 39.1. Definition of the jump ratio of an ionization edge which should be about 5–10 for the carbon K edge if the EELS is well aligned.

Qualitative microanalysis 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 so-called ghost edge from diode saturation in PEELS spectra (see Section 38.5). So long as you calibrate the spectrum to within 1–2 eV you can unambiguously identify the edge energy.

> 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 half-way 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 38.6, is usually sufficient to let you draw a definite conclusion about the identity of the specimen, which is BN on a C support film. In addition, it is wise to compare your spectrum with reference spectra from one of several EELS libraries that are available (Zaluzec 1981, Ahn and Krivanek 1983, Colliex 1984).

Remember that there are families of edges (K, L, M, etc.) just as there are families of peaks in X-ray spectra. 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. This is not always the case.

> L edges for Z = 19–28 and 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, which we first saw back in Figure 38.9, are so named because of their appearance in early, photographically recorded energy-loss spectra; more details are 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 only possible with standards, as we'll see shortly.

The energy-loss spectrum clearly does not lend itself to a quick "semiquantitative" analysis, as we can do with XEDS. For example, the spectrum in Figure 38.6 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).

Example

Sometimes, qualitative analysis is all that you need to do. Figure 39.2A and Figure 39.2B show 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 39.2A and Figure 39.2B, 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 39.2 would be difficult using windowless XEDS, because the energy resolution is close to the separation of the Ti L (452 eV) and the N K (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 MICROANALYSIS

To quantify the spectrum, you have to extract intensity in the ionization edge(s) by removing the plural-scattering background and integrating the intensity (I) in the edge.

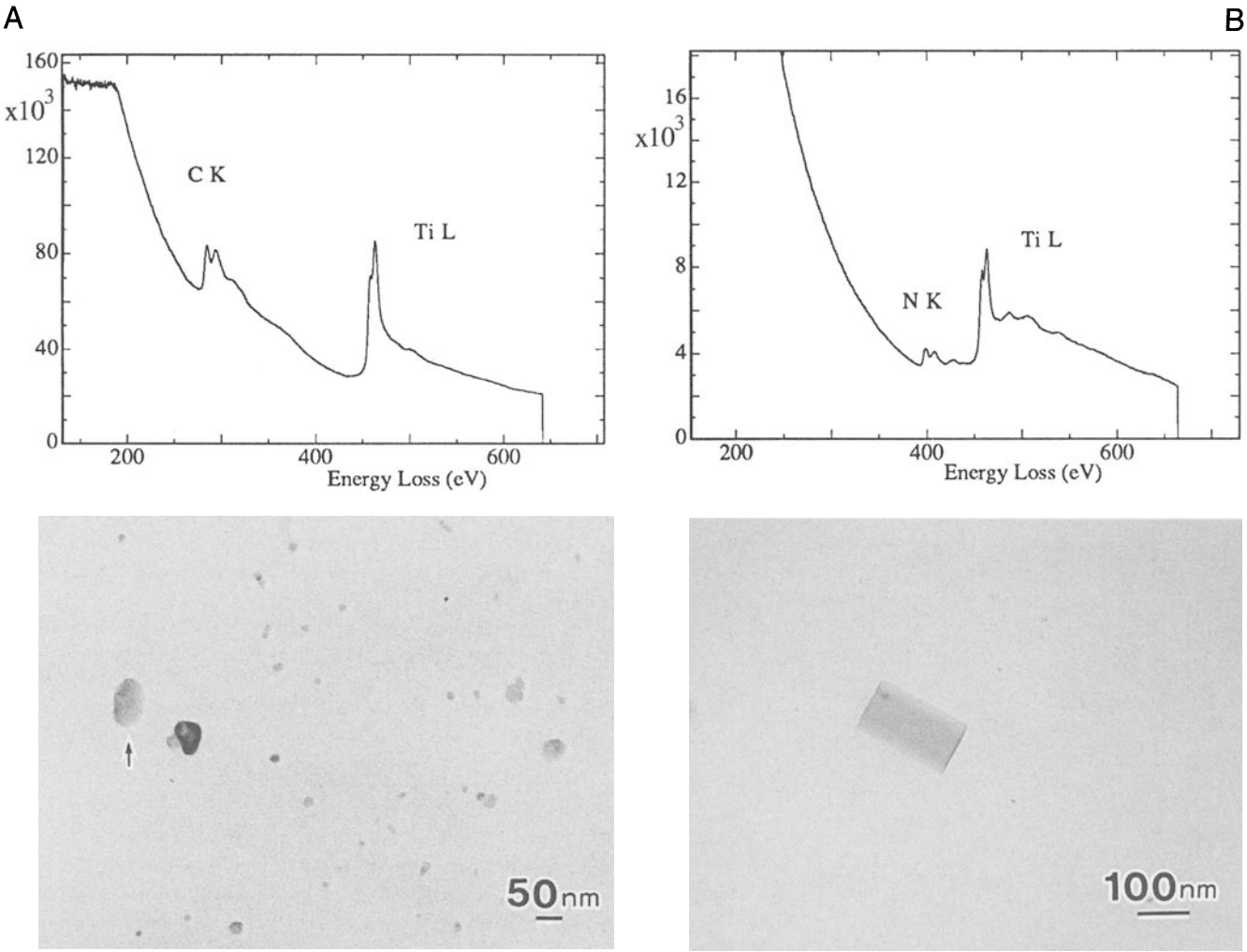


Figure 39.2. Images of small precipitates in a stainless steel specimen, and the corresponding ionization edges showing qualitatively the presence of Ti, C, and N. Thus the precipitates can be identified as (A) TiC and (B) TiN, respectively.

Then you have to determine the sensitivity factor, that is, you need to determine the number of atoms N responsible for I. This sensitivity factor is called the "partial ionization cross section." We'll see that it plays a similar role to the k_{AB} factor in X-ray microanalysis. If you go back and look at Figure 38.5, 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 38.5A, which contains the single-scattering "hydrogenic" edge intensity.

39.4.A. Derivation of the Equations for Quantification

The equations we use for quantitative analysis have been derived, refined, and applied by Egerton and co-workers. The following derivation is a summary of the full treatment by Egerton (1996), which itself draws on work carried out over the preceding two decades.

We'll assume that we are quantifying a K edge, although the basic approach can be used for all edges. The K-shell intensity above background, I_K, is related to the probability of ionization, P_K, and the total transmitted intensity, I_T

$$I_K = P_K I_T \qquad [39.1]$$

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 for 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

$$P_K = N\sigma_K \exp\left(\frac{t}{\lambda_K}\right) \qquad [39.2]$$

where N is the number of atoms *per unit area* of the specimen (thickness t) that contribute to the K edge. The assumption of a single K-shell ionization event with cross-section σ_K is reasonable, given the large mean free path (λ_K) for ionization losses; but it explains why you have to make thin specimens. It also means that the exponential term is very close to unity, and so

$$I_K \approx N\sigma_K I_T \qquad [39.3]$$

and therefore

$$N = \frac{I_K}{\sigma_K I_T} \qquad [39.4]$$

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 and the ionization cross section. We can easily extend this expression to a spectrum containing two edges from elements A and B, in which case the total intensity drops out and we can write

$$\frac{N_A}{N_B} = \frac{I_K^A \sigma_K^B}{I_K^B \sigma_K^A} \qquad [39.5]$$

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 zero-loss peak, and this makes life much easier for the spectrometer scintillator or diode array.

In both equations 39.4 and 39.5 we assumed that we could accurately subtract the background under the ionization edge and that we know σ. Unfortunately, as you'll see, both background subtraction and determination of σ are nontrivial and limit the accuracy of quantification. We will discuss these points later, but initially 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 energy-loss spectrum out to the beam energy, E_0, because above 1–2 keV the intensity decreases to a level close to the system noise. Furthermore, 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 ionization threshold, E_C. In addition, the background extrapolation process becomes increasingly inaccurate beyond about 100 eV, and so it is imperative to restrict the integration of spectral intensities to some window, Δ, usually in the range of 20–100 eV. So we modify equation 39.4 to give

$$I_K(\Delta) = N\sigma_K(\Delta) I_T(\Delta) \qquad [39.6]$$

The term $I_T(\Delta)$ is more correctly written as $I_\ell(\Delta)$, 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. As a result, the technique is inherently far more efficient than XEDS. However, because we cannot physically collect the spectrum over 4π sr, but are limited by our

choice of collection semiangle β, we must further modify the equation and write

$$I_K(\beta\Delta) = N\sigma_K(\beta\Delta)I_\ell(\beta\Delta) \qquad [39.7]$$

The factor $\sigma_K(\beta\Delta)$ is termed the "partial ionization cross section," from this equation, therefore, the absolute quantification for N is given by

$$N = \frac{I_K(\beta\Delta)}{I_\ell(\beta\Delta)\sigma_K(\beta\Delta)} \qquad [39.8]$$

For a ratio of two elements A and B, the low-loss intensity drops out

$$\frac{N_A}{N_B} = \frac{I_K^A(\beta\Delta)\sigma_K^B(\beta\Delta)}{I_K^B(\beta\Delta)\sigma_K^A(\beta\Delta)} \qquad [39.9]$$

We can draw a direct analogy between this equation and the Cliff–Lorimer expression (equation 35.2) used in thin-foil 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 electron spectrometry 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 plural scattering, although in very thin specimens the approximation remains valid, if errors of ±10–20% are acceptable. If plural scattering is significant then the spectrum must be deconvoluted, and we will discuss ways to do this in Section 39.6 when we describe the limitations of specimen thickness. You should also note when using the ratio equation that 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 then, these equations give us an absolute value of the atomic content of the specimen or a ratio of the amounts of two elements. You have to carry out two essential practical steps:

- The background subtraction to obtain I_K.
- The determination of the partial ionization cross section $\sigma_K(\beta\Delta)$.

So now you can see why it is important to know β.

39.4.B. Background Subtraction

The background intensity comes from plural-scattering events which are usually associated with outer-shell interactions. In the spectrum the background appears as a rapidly changing continuum decreasing from a maximum just after the plasmon peak at about 15–25 eV, down to a minimum at which it is indistinguishable from the instrumental noise, typically when $\mathcal{E} \sim$ 1–2 keV. In addition to the true plural scattering, there is also the possibility of single-scattering contributions to the background from the tails of preceding ionization edges. Because of the complexity of the many combinations of plural-scattering processes, it has not proven possible to model the background from first principles to the same degree that is possible in XEDS using variations on Kramers' Law. There are two ways commonly used to remove the background:

- Curve fitting.
- Using difference spectra.

We'll now describe these in some detail.

Curve Fitting: You select a window δ in the background before the edge onset and fit a curve to the channels. Then you extrapolate the curve over the desired energy window Δ under the edge. This process is shown schematically in Figure 39.3, and experimentally in Figure 39.4.

We assume that the energy dependence of the background has the form

$$I = A\,\mathcal{E}^{-r} \qquad [39.10]$$

where I is the intensity in the channel of energy loss $\mathcal{E}$, and A and r are constants for a particular curve fit. The fitting parameters are only valid over a limited energy range because they depend on $\mathcal{E}$. The exponent r is typically in the

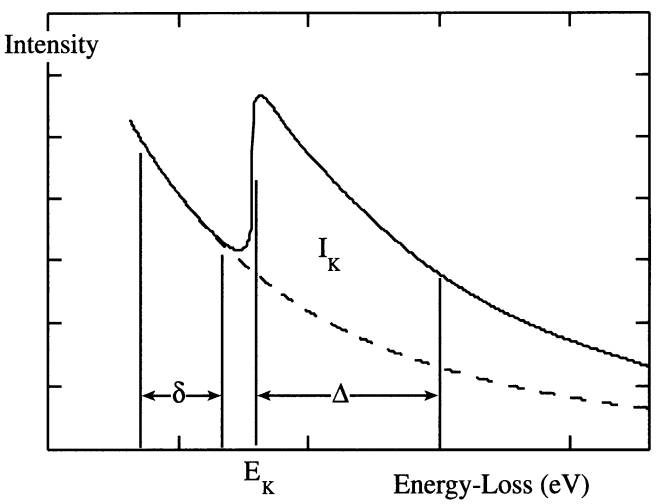


Figure 39.3. The parameters required for background extrapolation and subtraction under an ionization edge. The pre-edge fitting window δ is extrapolated over a post-edge window Δ then subtracted to give the edge intensity I_K.

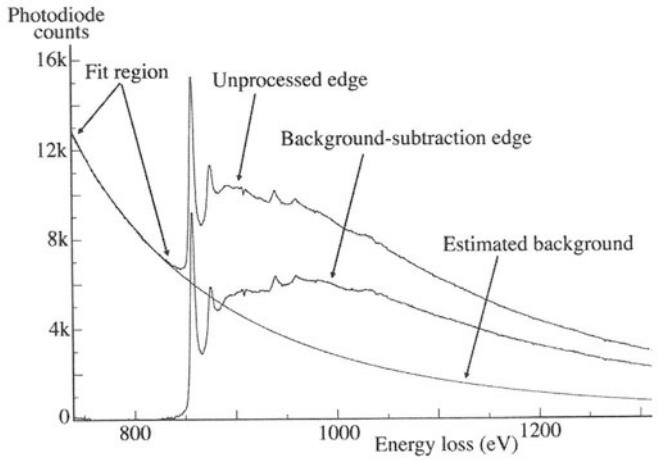


Figure 39.4. A Ni $L_{2,3}$ edge before and after background subtraction. The fit region before the unprocessed edge is extrapolated to give the estimated background, which is then removed, leaving the background-subtracted edge.

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 semiangle, β, increases.
- The electron energy loss, $\mathcal{E}$, increases.

The fit to the tail of a preceding edge also shows a similar power-law dependence on the plural-scattering background, and may be fitted in a similar manner, i.e., $I = B\mathcal{E}^{-s}$. The energy range δ over which you fit the background should not be <~10 channels, and at most not >~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.

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. In Figure 39.4, the extrapolation window (~450 eV) is a little larger than ideal. So, Δ is smaller for lower edge energies. Using larger windows, although improving the statistics of the total edge integration, eventually reduces the accuracy of the final quantification because the fitting parameters A and r are only valid over ~100 eV. If there's a lot of near-edge structure, 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 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.

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, chi-squared, 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 $y = \ln_e I$. The squared term in the denominator ensures suitable weighting of the channels close to 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 39.5, the difference process results in the slowly varying intensity (i.e., background) being reduced to zero and the rapidly varying intensity features (i.e., ionization edges) showing up as classical difference-peaks, similar to what you may have seen in an Auger spectrum. 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 sup-

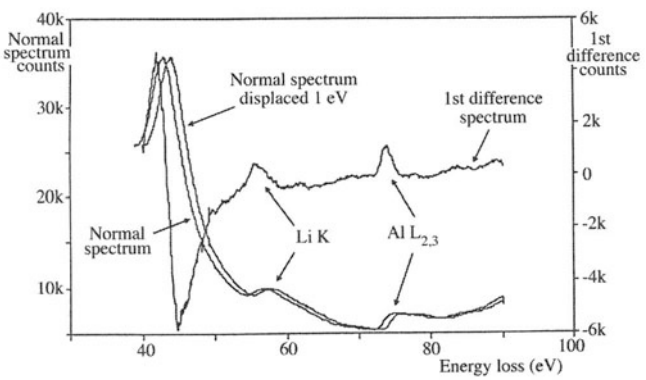


Figure 39.5. First-difference method of background subtraction, showing two PEELS spectra from a specimen of Al-Li displaced by 1 eV and subtracted to give a spectrum in which the background intensity falls to zero and the small Li K and Al $L_{2,3}$ edges are clearly revealed.

presses spectral artifacts common to PEELS, particularly the channel-to-channel gain variation.

Another kind of difference method involves convoluting the experimental spectrum with a filter function. A top-hat filter function, similar to the one we described in Section 35.2 for background subtraction in XEDS, gives a second-difference spectrum which also removes the background but exacerbates some artifacts.

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 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 fine structure effects (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 Zero-Loss Integral

Remember from equation 39.8 that if you want *absolute* quantification of N, then you have to integrate the low-loss spectrum I_ℓ out to about 50 eV. In a SEELS you should always do this using a reverse scan to avoid any problems with after-glow of the scintillator, and in PEELS you must be careful to integrate for a short enough time so you don't saturate the diode array. If you are doing a ratio, then I_ℓ is not needed (equation 39.9).

39.4.E. The Partial Ionization Cross Section

There are several ways we can determine the partial ionization cross section, $\sigma(\beta\Delta)$, which is the sensitivity factor relating intensity (I) to the number of atoms (N). 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 (1979, 1981), 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 and are available in Gatan's ELP software, but the code is also given in 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.

While at first sight this is an absurd approximation, the approach is 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. Because this treatment neglects the outer-shell electrons, it is best suited to K-shell electrons, and Figure 39.6 shows a comparison between the measured N K-shell 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 very good fit, on average, to the experimental edge. Figure 39.7 compares the Cr L edge with the SIGMAL model. The L-shell fit is almost as good as the K fit, but the white lines are imperfectly modeled. These programs are very widely used since they are simple to understand and easy and quick to apply.

There is another theoretical approach which uses empirical parameterized equations to calculate the terms that modify σ for the effects of β and Δ. Both Joy (1986b) and Egerton (1989) have given relatively simple expressions, amenable to evaluation on a hand-held calculator, which you can look up if you wish. Joy's parameterization approach and the SIGMAK/SIGMAL models give good agreement, as shown in Figure 39.8. There are more com-

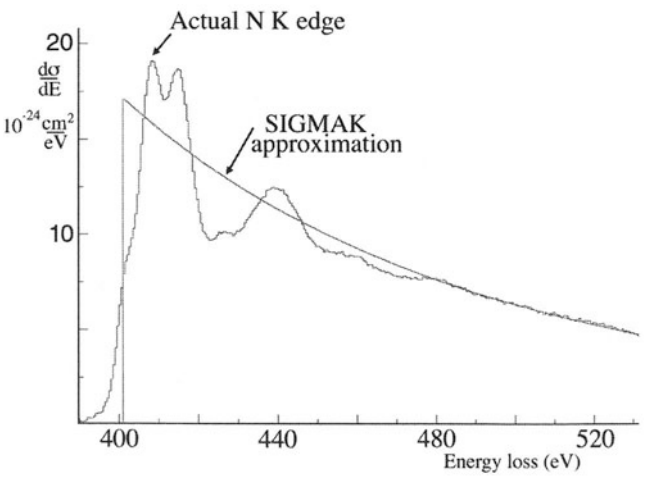


Figure 39.6. Comparison of an experimental N K edge and the hydrogenic fit to the edge using the SIGMAK program.

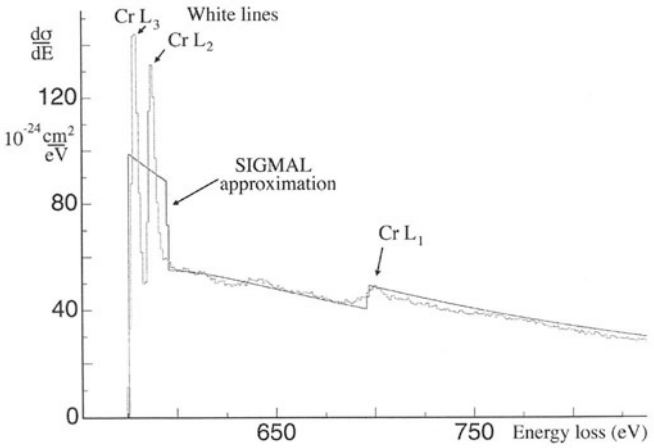


Figure 39.7. Comparison between an experimental Cr $L_{2,3}$ edge and a modified hydrogenic approximation to the edge obtained using the SIGMAL program. The fit makes no attempt to model the intense white lines, but only makes a rough estimate of their average intensity.

plex methods available which calculate the cross section in a more realistic way than the hydrogenic model, e.g., using Hartree–Slater models or atomic-physics approaches, which are better for the more complex L and M edges (Rez 1989). Egerton (1993) has compared experimental and theoretical cross sections, and the M-shell data (which are the worst case) are shown in Figure 39.9. The data are actually plotted in terms of the oscillator strength f (which is a measure of the response of the atom to the incident electron). This term is the integral of the generalized oscillator strength, which is proportional to the differential cross section, so just think of f as proportional to σ. There is still relatively poor agreement between experiment and theory for the M shell, as well as between the atomic and hydrogenic theoretical models. Similar data in Egerton's paper show

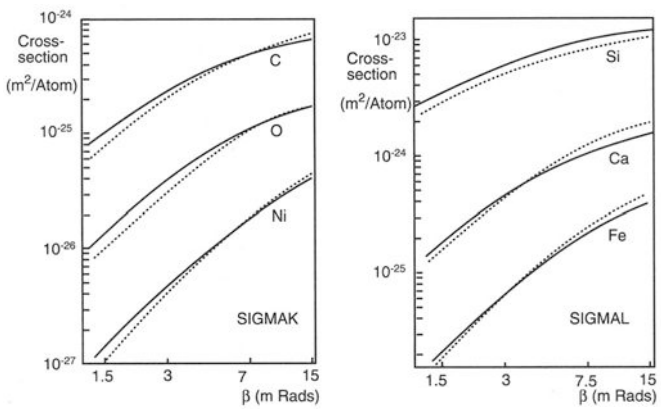


Figure 39.8. Comparison of the SIGMAK and SIGMAL hydrogenic models (full lines) for the ionization cross section with the parametric model (dotted lines).

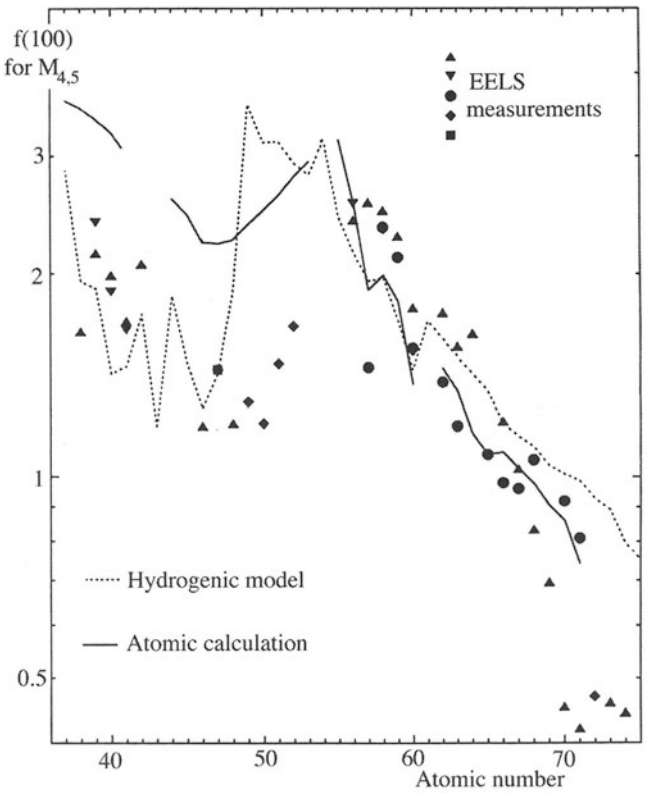


Figure 39.9. Comparison of the experimental and theoretical approaches to determination of the M-shell ionization cross section shown in terms of the variation in the dipole oscillator strength (f) as a function of atomic number. The data points are different experimental measurements, the solid line is a fundamental atomic calculation, and the dotted line is a hydrogenic calculation.

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 microanalysis, you rarely need to go to such lengths to obtain a better value of $\sigma(\beta\Delta)$, and you should generally stick with the SIGMAK/L approach 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 classical 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 and 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.

Again, it is the problem of thickness measurement that appears to be the main limitation to improving the accuracy of microanalysis.

Despite these limitations, data are available comparing cross sections of various elements, relative to oxygen (Hofer *et al.* 1988), just as *k* factors are determined relative to Si or Fe. Malis and co-workers (Malis *et al.* 1987, Malis and Titchmarsh 1988) have also produced a large number of experimental cross sections for light-element compounds. It is intriguing to note that the experimental approach appears increasingly popular in EELS, while in XEDS the reverse trend, toward more theoretical modeling of *k* factors, seems to be the case!

The SIGMAK/L programs may introduce large errors when quantifying the lightest metallic elements, Li and Be, so for the most accurate quantification of these elements the standards approach is still the best.

Example

In a study of Al-Li alloys (Liu and Williams 1989), a homogeneous sample of Al-12.7 at.% Li was used as a standard; *t* was determined from the relative intensities of the first plasmon peak and the zero-loss peak. The integrated intensity ratio for the Li K/Al L_{23} edges was determined after background subtraction to be 0.106±0.006. This number was the average of six separate spectra, and the errors were based on a student *t* analysis at the 95% confidence limit. From equation 39.9, the Li/Al partial-cross-section ratio was calculated to be 1.37±0.07. An Al-Li specimen containing an intermetallic of unknown Li content was then examined and 13 spectra obtained which gave an average Li K/Al L_{23} intensity ratio of 0.188±0.009. Combining this ratio with the partial ionization cross section and substituting back into equation 39.9 gives the composition of the intermetallic as Al-20.5±1 at.%Li. This result and others are given in Figure 39.10, which shows the low-Li portion of the Al-Li phase diagram, determined through direct Li composition measurements. For comparison, the partial ionization cross section was also determined from the SIGMAK/L programs, and the ratio was 0.969, ~30% less than that obtained using the standard. While this is a large difference compared to most SIGMAK/L calculations, it still sounds a note of caution against unquestioning use of the calculated cross sections.

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, tedious experimental data are still the best option. Of course, for such elements it is probably better to revert to XEDS analysis anyhow.

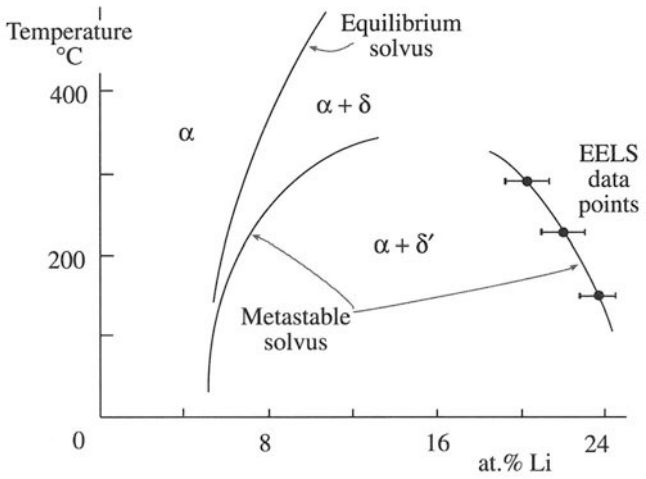


Figure 39.10. The Al-Li phase diagram determined by EELS, showing a variation in the Li content of the metastable Al_3Li (δ′) phase as the temperature is raised. The equilibrium phases are α (Al-Li) solid solution and δ (Al-Li) intermetallic.

So now we're in a position where we have all the data needed to solve the quantification equations. However, our assumption all along has been that the spectra were the result of single scattering and we neglected plural scattering. Now in practice there will *always* be some plural-scattering contribution to the ionization edges.

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 38.5F and, if you look ahead, in Figure 39.15. So how do we go about correcting for this? We can either make our specimens so thin that plural scattering is negligible, or we can deconvolute the spectra. The former approach is possible, but you have to be lucky or exceptionally skilled at specimen thinning. The latter approach is mathematically simple, but can be misleading if not done properly, so we will need to examine deconvolution in more detail; but let's look at how we determine *t* because EELS offers us a simple method for this.

39.5. MEASURING THICKNESS FROM THE ENERGY-LOSS SPECTRUM

There is thickness information in the energy-loss spectrum since the amount of all inelastic scatter increases with

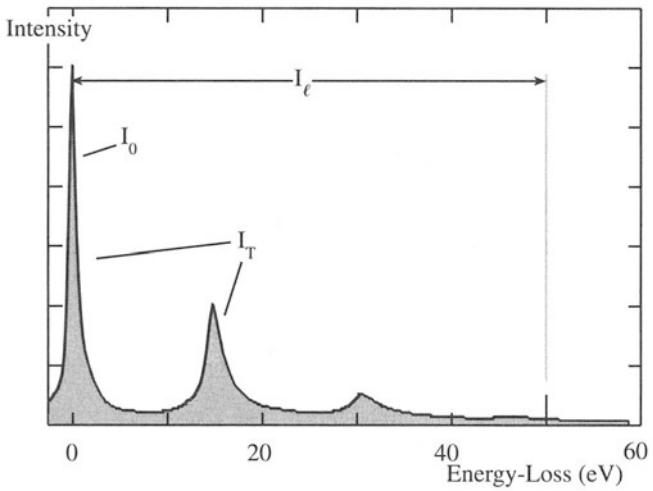


Figure 39.11. Definition of the zero-loss intensity I_0, the total intensity I_T, and the low-loss (I_ℓ) intensity required for thickness determination.

specimen thickness. In principle we have to measure the intensity under the zero-loss peak (I_0) and ratio this to the total intensity in the spectrum (I_T), as defined in Figure 39.11. But in practice the intensity in the EELS spectrum falls so rapidly with increasing energy loss that we can reasonably approximate I_T to the intensity in the low-loss portion of the spectrum I_ℓ, out to about 50 eV. The relative intensity of the zero loss and the total intensity is governed by the average mean free path (λ) for energy losses up to 50 eV, and thus

$$t = \lambda \ln\left(\frac{I_\ell}{I_0}\right) \qquad [39.12]$$

All we need is to determine λ for the specimen, and we get this from a parameterization based on many experimental measurements (see Malis *et al.* 1988). The expression is

$$\lambda = \frac{106\, F E_0}{\left\{\mathcal{E}_m \ln\left(\frac{2\beta E_0}{\mathcal{E}_m}\right)\right\}} \qquad [39.13]$$

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

$$\mathcal{E}_m = 7.6\, Z^{0.36} \qquad [39.14]$$

The relativistic factor (F) is given by

$$F = \frac{\left\{1 + \frac{E_0}{1022}\right\}}{\left\{1 + \left(\frac{E_0}{511}\right)^2\right\}} \qquad [39.15]$$

You can easily store these equations in the TEM computer or in your calculator and they give an accuracy for t of better than ±20%.

If indeed your specimen is so thin that only single scattering occurs, then you can use a similar expression but assume that the only significant scatter was a single plasmon event. Thus

$$t = \lambda_P \frac{I_P}{I_0} \qquad [39.16]$$

where λ_P is the plasmon mean-free path (see Table 38.2), I_P is the intensity in the first (and only) plasmon peak, and I_0 is the intensity in the zero-loss peak.

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.

If plural scattering is significant, then your quantification results become unreliable.

A typical ball-park figure is that, if the intensity in the first plasmon peak is greater than one-tenth the zero-loss intensity, then your specimen is too thick.

Another way of saying this is that, if $t > 0.1\lambda_P$, then errors >~10% are expected, as shown in Figure 39.12. Of course,

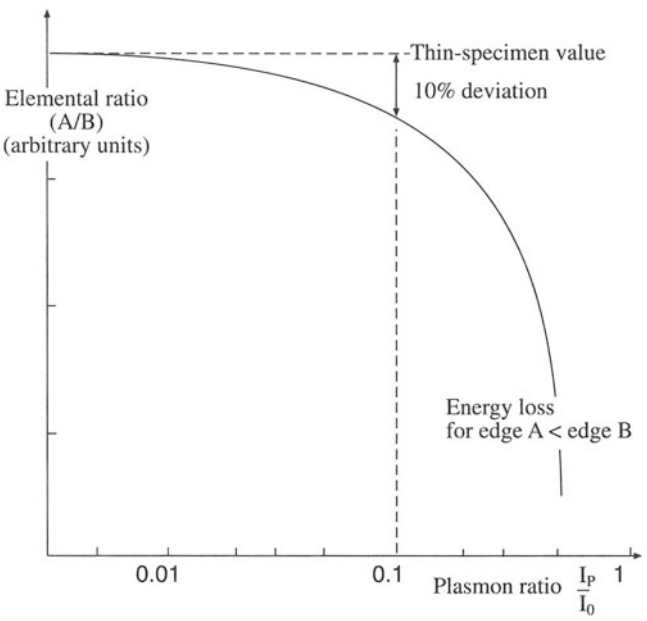


Figure 39.12. The intensity ratio of two ionization edges (A/B) as a function of specimen thickness. The thickness is plotted in terms of the ratio of the plasmon to the zero-loss intensity (I_P/I_0). The intensity ratio is affected significantly when I_P is above about 0.1 I_0.

one way round 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 to make the single-scattering assumption valid.

39.6. DECONVOLUTION

We saw back in Figure 38.5 that the effect of plural scattering is to add intensity to the ionization edge, mainly as a result of combined inner- and outer-shell losses.

> We can represent the experimental ionization edge as a true single-scattering (hydrogenic) edge convoluted with the plasmon, or low-loss, spectrum.

The aim of the deconvolution process therefore, as shown schematically in Figure 39.13, is to extract the single-scattering intensity distribution. We'll describe two methods, the Fourier-Log and the Fourier-Ratio, which are both based on the work of Egerton *et al.* (1985), and both methods are incorporated in the Gatan ELP proprietary software. Strictly speaking, the deconvolution should be carried out in both the energy dimension and the angular dimension, but in practice all the routines ignore the angular dimension; this simplification introduces a small systematic error into any deconvolution. The error is usually <10% up to typical energy losses below about 1 keV, so we can usually ignore it. A smaller β increases the deconvolution error, since the plural-scattered electrons have a wider 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 junction we described in Section 37.3. The Fourier transform of the whole spectrum (F) is then given by

$$F = F(0)\exp\left(\frac{F(1)}{I_0}\right) \qquad [39.17]$$

where $F(0)$ is the transform of the elastic contribution, $F(1)$ 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 $F(1)$, 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.

The danger of this approach is that you may introduce artifacts into the single-scattering spectrum. In particular, any gain change in a SEELS spectrum must be removed or not incorporated in the original spectrum at all. Despite the 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 39.14.

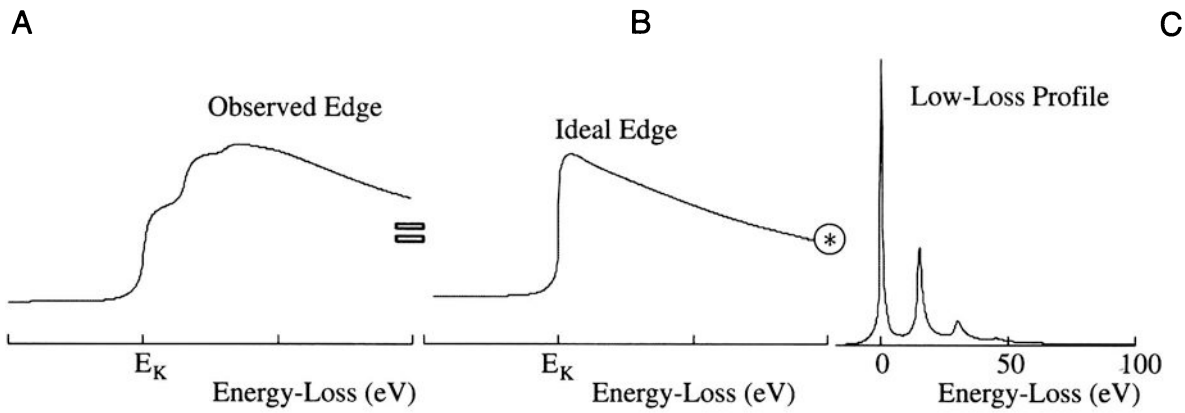


Figure 39.13. The contribution of plural scattering to the experimentally observed ionization edge intensity profile (A) is determined by the convolution of the ideal single-scattering ionization edge (B) with the low-loss plasmon region (C).

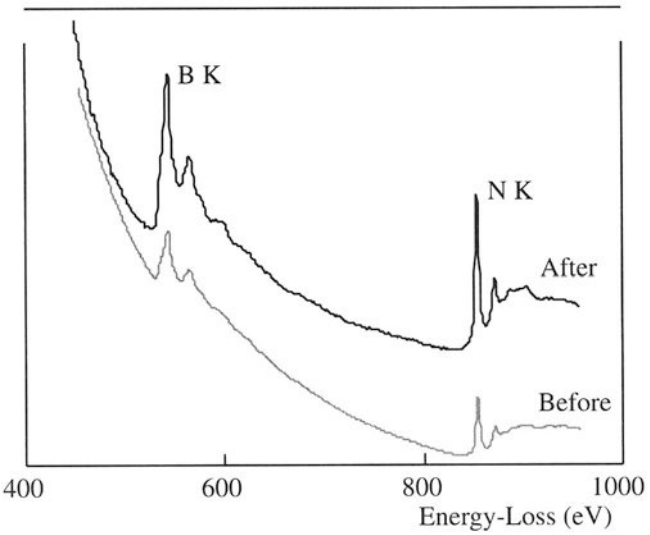


Figure 39.14. A spectrum from a thick crystal of BN before and after Fourier-Log deconvolution. The jump ratio is increased in the deconvoluted spectrum, which is displaced vertically for clarity.

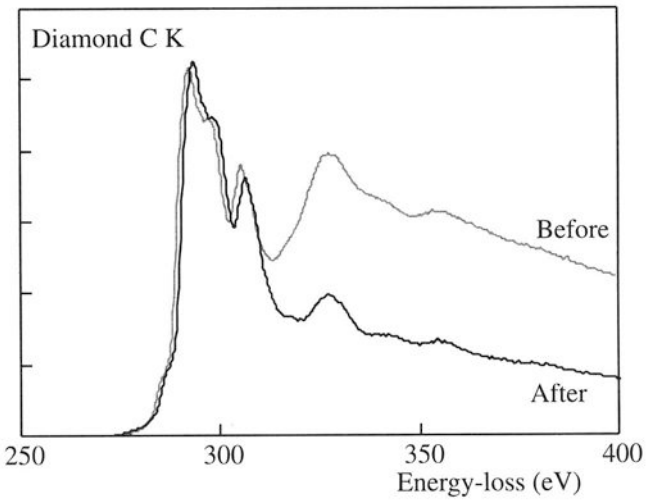


Figure 39.15. A carbon K edge from a thick specimen before and after Fourier ratio deconvolution. The plural scattering plasmon contribution to the post-edge structure is removed.

The *Fourier-Ratio* technique approximates the experimental spectrum to the ideal single-scattering spectrum, convoluted with the low-loss spectrum. We define the low-loss portion of the spectrum as the region up to ~50 eV, including the zero-loss peak, but before the appearance of any ionization edges. So we can now write

$$F' = F(1).F(P) \qquad [39.18]$$

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 zero-loss peak. If we rearrange equation 39.18 to give a ratio (hence the name of the technique)

$$F(1) = \frac{F'}{F(P)} \qquad [39.19]$$

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 you deconvolute. Again, to avoid the problem of increased noise, it is necessary to multiply equation 39.19 by the transform of the zero-loss peak. Figure 39.15 shows a carbon K edge before and 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 reference spectra (Leapman 1992). A single-

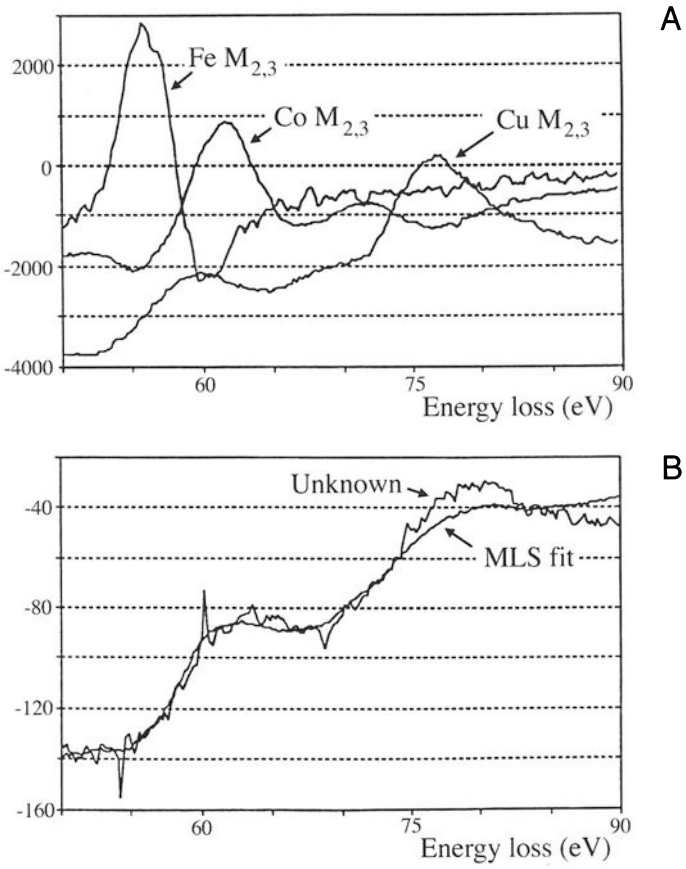


Figure 39.16. (A) Three first-difference M edge reference spectra from Fe, Co, and Cu. (B) MLS fit of the reference spectra superimposed on a low-energy portion of an experimental spectrum from an intermetallic particle in a Cu alloy showing the good fit that can be obtained.

scattering reference spectrum $R_0(\mathcal{E})$ 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 $R_1(\mathcal{E}) = P^*R_0(\mathcal{E})$ is used to generate several reference spectra ($R_2(\mathcal{E}) = P^*R_1(\mathcal{E})$, 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 39.16A and the actual fit to part of the experimental spectrum from an intermetallic in a Cu-Be-Co alloy is shown in Figure 39.16B.

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 all spectra should be deconvoluted 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.

> Always check the validity of the deconvolution routine by applying it to spectra from a known specimen obtained over a range of thickness.

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, 2α, may introduce an error into your quantification. When 2α is equal to or greater than 2β, convergence effects can limit the accuracy because the experimental angular distribution of scattered electrons will be wider than expected. 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 $\sigma(\beta\Delta)$ when α is greater than β

$$R = \frac{\left[\ln\left(1 + \frac{\alpha^2}{\theta_{E^2}}\right)\beta^2\right]}{\left[\ln\left(1 + \frac{\beta^2}{\theta_{E^2}}\right)\alpha^2\right]} \qquad [39.20]$$

where θ_E is the characteristic scattering angle. 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.

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. 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; the energy-loss processes behave similarly. This phenomenon, known as the Borrmann effect in XEDS (see Section 35.8) is not important for low-energy edges, but intensity changes of a factor of two have been reported for Al and Mg K edges (Taftø and Krivanek 1982). The use of large α minimizes the problem in XEDS, but beam-convergence effects are themselves a problem in EELS. 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. SPATIAL RESOLUTION

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

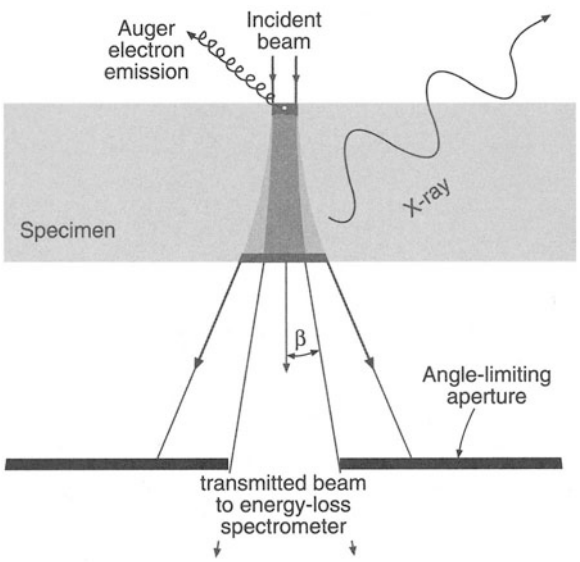


Figure 39.17. The effect of the spectrometer collection angle is to limit the contribution to the spectrum from high-angle scattered electrons, thus ensuring a high spatial resolution signal.

electrons emanating from the specimen in a narrow cone, as shown in Figure 39.17. 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 mode on a TEM, is mainly the size of the probe; we can easily get data with probe sizes <10 nm, and <1 nm with more difficulty.
- 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.

In TEM mode, chromatic aberration usually limits the spatial resolution, as we showed back in Section 37.4.B. We know that the EELS signal isn't affected much by beam spreading and we can easily limit the source of the signal to a few nanometers with an FEG. So, there have been correspondingly fewer studies of the limits of spatial resolution. Most work on defining the spatial resolution has been pursued in France by Colliex and co-workers, e.g., Colliex (1985). Because the primary factor when operating in STEM mode (especially with an FEG) is the incident-probe diameter, we have to be concerned about the problems of spherical aberration broadening the probe (Colliex and Mory 1984). So you must be careful in your selection of the beam-defining aperture. We discussed this topic in detail in the section on the spatial resolution of XEDS.

One factor that we often consider in EELS, but ignore in XEDS (although it occurs in X-ray generation also), is the phenomenon of delocalization.

> Delocalization is the ejection of an inner-shell electron by the passage of a high-energy electron some distance from the atom.

If you are physics oriented and want to read more about this, see Muller and Silcox (1995). The scale of this wave-mechanical effect is small, in the range 2–5 nm, and it is inversely proportional to the energy loss. So it appears that, except in rare cases, delocalization will not limit the spatial resolution; the practical factors such as probe aberrations, signal-to-background in the EELS signal, and damage are much more important. So we can conclude that spatial resolution for EELS will be somewhat better than for XEDS and experiments seem to indicate that this indeed is the case (Colliex 1985). In fact, in certain zone-axis misorientations, it appears that the (FEG) electron beam, if it is <1 nm, can be localized to individual rows of atoms, producing atomic-level spatial resolution (see Figure 40.5C and Browning *et al.* 1993).

39.10. DETECTABILITY LIMITS

The detectability limits for ionization-loss spectrometry are governed by the same factors as we discussed for XEDS. Therefore, the inverse relationship with spatial resolution also applies. Clearly we have to optimize several factors:

- The edge intensity.
- The signal-to-background ratio (jump ratio).
- The efficiency of signal detection.
- The time of microanalysis.

EELS has an inherently higher efficiency than XEDS, but a correspondingly poorer signal-to-background because of the higher background in the spectrum. Joy (1986a) has attempted to compare the two techniques in some detail and calculations based on a thermionic source. He concluded that the MMF for EELS would be of the order of 1–10% in a Si foil 50 nm thick; this value is somewhat worse than the experimental data for XEDS in similar specimens. Leapman and Hunt (1991) have argued that in most situations, PEELS is more sensitive to the presence of small amounts of material than XEDS.

The time of collection, which strongly influences the detectability limit, is particularly dependent on whether serial or parallel collection is used. Colliex (1985) reckons

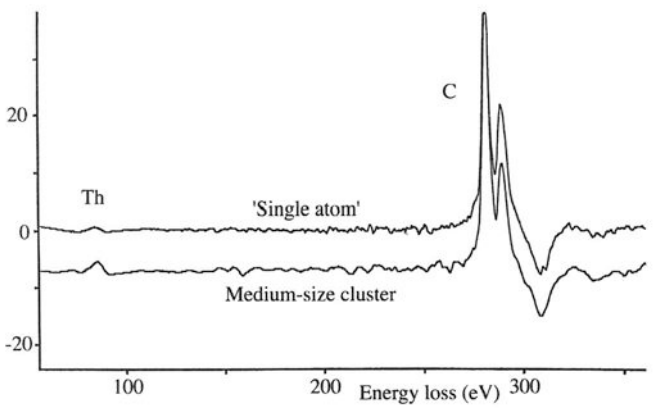


Figure 39.18. First-difference spectra showing the detection of a small cluster and a single atom of Th on a carbon support film.

that a tenfold improvement in all EELS performance criteria is to be expected if parallel collection is used. The best results, combining sensitivity and spatial resolution, will be obtained with an FEG. Krivanek *et al.* (1991) used an FEG DSTEM, parallel detection, and sophisticated data processing to detect the presence of single atoms of Th on thin carbon films, as shown in Figure 39.18. While this is a most favorable analysis situation because of high Z of the atoms and the low average Z of the support film, the result still shows clearly the superiority of the best possible EELS microanalysis over XEDS, which cannot yet detect single atoms.

In conclusion, microanalysis using ionization edges, while considerably more difficult to perform than XEDS, appears to offer both improved spatial resolution and analytical sensitivity. Parallel collection is significantly better than serial collection in both aspects. As was the case for XEDS, an FEG source is required for the best performance.

CHAPTER SUMMARY

The ionization edges can be used to give quantitative elemental analyses from all the elements in the periodic table using a ratio equation. Beware, however, of the many experimental variables you have to define for your TEM, the PEELS, and the specimen. Compared to XEDS there have been very few quantitative analyses or composition profiles measured using EELS.

To use Egerton's ratio equation:

- You have to subtract the background using a power law or MLS approach. The former is easier.
- Integrate the edge intensity. That's straightforward.
- Then you have to determine the partial ionization cross section $\sigma_K(\beta\Delta)$. This is more difficult.
- Calculate $\sigma_K(\beta\Delta)$ with SIGMAK and SIGMAL for most K and L edges.
- For M edges and for the lightest elements (e.g., Li), use known standards.

The difficulty with using standards is that the specimen thickness has to be the same as the unknown and the standard also has to have the same bonding type as the "unknown." This is often impossible, although specimen thicknesses can be deduced directly from the low-loss spectrum intensity. The biggest limitation to quantification is that, ideally, your specimens have to be less than one mean-free path in thickness (typically < 50 nm) otherwise deconvolution routines are needed, which can introduce artifacts on their own.

Spatial resolution and minimum detectability are better than XEDS. Single-atom detection has been demonstrated.

REFERENCES

General References

Egerton, R.F. (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope,* 2nd edition, Plenum Press, New York.

Joy, D.C. (1986a) in *Principles of Analytical Electron Microscopy* (Eds. D.C. Joy, A.D. Romig Jr., and J.I. Goldstein), p. 249, Plenum Press, New York.

Joy, D.C. (1986b) in *Principles of Analytical Electron Microscopy* (Eds. D.C. Joy, A.D. Romig Jr., and J.I. Goldstein), p. 277, Plenum Press, New York.

Maher, D.M. (1979) in *Introduction to Analytical Electron Microscopy* (Eds. J.J. Hren, J.I. Goldstein, and D.C. Joy), p. 259, Plenum Press, New York.

Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science,* 2nd edition, Philips Electron Optics Publishing Group, Mahwah, New Jersey.

Specific References

Ahn, C.C. and Krivanek, O.L. (1983) *EELS Atlas,* Gatan, Inc., 780 Commonwealth Drive, Warrendale, Pennsylvania 15086.

Browning, N.D., Chisholm, M.F., and Pennycook, S.J. (1993) *Nature* **366**, 143.

Colliex, C. (1984) *Advances In Optical and Electron Microscopy* **9** (Eds. R. Barer and V. E. Cosslett), p. 65, Academic Press, New York.

Colliex, C. (1985) *Ultramicroscopy* **18**, 131.

Colliex, C. and Mory, C. (1984) *Quantitative Electron Microscopy* (Eds. J.N. Chapman and A.J. Craven), p. 149, SUSSP Publications, Edinburgh, Scotland.

Disko, M.M. (1986) in *Microbeam Analysis-1986* (Eds. A.D. Romig Jr. and W.F. Chambers), p. 429, San Francisco Press, San Francisco, California.

Egerton, R.F. (1979) *Ultramicroscopy* **4**, 169.

Egerton, R.F. (1981) in *Proc. 39th EMSA Meeting* (Ed. G.W. Bailey), p. 198, Claitors, Baton Rouge, Louisiana.
Egerton, R.F. (1989) *Ultramicroscopy* **28**, 215.
Egerton, R.F. (1993) *Ultramicroscopy* **50**, 13.
Egerton, R.F., Williams, B. G., and Sparrow, T.G. (1985) *Proc. Roy. Soc.* **A398**, 395.
Hofer, F., Golob, P., and Brunegger, A. (1988) *Ultramicroscopy* **25**, 181.
Krivanek, O.L., Mory, C., Tence, M., and Colliex, C. (1991) *Microsc. Microanal. Microstruct.* **2**, 257.
Leapman, R.D. (1992) in *Transmission Electron Energy Loss Spectrometry in Materials Science* (Eds. M.M. Disko, C.C. Ahn, and B. Fultz), p. 47, TMS, Warrendale, Pennsylvania.
Leapman, R.D. and Hunt, J.A. (1991) *Microsc. Microanal. Microstruct.* **2**, 231.
Liu, D.R. and Williams, D.B. (1989) Proc. Roy. Soc. London **A425**, 91.
Malis, T., Rajan, K., and Titchmarsh, J.M. (1987) in *Intermediate Voltage Electron Microscopy* (Ed. K. Rajan), p. 78, Philips Electron Optics Publishing Group, Mahwah, New Jersey
Malis, T. and Titchmarsh, J.M. (1988) in *Electron Microscopy and Analysis–1985* (Ed. G.J. Tatlock), p. 181, Adam Hilger Ltd., Bristol and Boston, Massachusetts.
Malis, T., Cheng, S., and Egerton, R.F. (1988) *J. Electron Microsc. Tech.* **8**, 193.
Muller, D.A. and Silcox, J. (1995) *Ultramicroscopy* **59**, 195.
Rez, P. (1989) *Ultramicroscopy* **28**, 16.
Taftø, J. and Krivanek, O.L. (1982) *Phys. Rev. Lett.* **48**, 560.
Zaluzec, N.J. (1981) in *Analytical Electron Microscopy-1981* (Ed. R.H. Geiss), p. 193, San Francisco Press, San Francisco, California.

Everything Else in the Spectrum

40

CHAPTER PREVIEW

The energy resolution of the magnetic prism spectrometer is very good, which means that the energy-loss spectrum contains a wealth of information about the specimen in addition to its basic elemental chemistry. In the previous chapter, we mentioned how we can learn about chemistry using ionization edges. Much of this chemical information is contained in fine-detail intensity variations at the ionization edges in the core-loss spectra termed *energy-loss near-edge structure* (ELNES) and *extended energy-loss fine structure* (EXELFS). From this fine structure, we can obtain information on how the ionized atom is bonded, the coordination of the atom, and its density of states. Furthermore, we can probe the distribution of other atoms around the ionized atom, i.e., the radial distribution function (RDF). Understanding these phenomena requires that we use certain concepts from atomic and quantum physics. The nonphysicist 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.

If high spatial resolution is important, you can't obtain this additional information by any other spectroscopic technique.

In addition to the extra information around the ionization edges, we can extract useful data from the low-loss region (<50 eV) of the spectrum. The predominant features in this part of the spectrum are the plasmon

peaks, which represent the response of the weakly bound valence and conduction electrons to the high-energy incident electron. The plasmon response contains direct information about the free-electron density. In some binary free-electron alloys, plasmon-peak shifts reflect the composition of the specimen. Within the low-loss region, but separate from the intense plasmon peak, we can find intensity that is related to the dielectric constant of the specimen. Furthermore, we can discern certain inter/intraband transitions, especially in polymers, and we can measure directly the band gap of semiconductors and insulators. We also introduce briefly the effect of the angle of scatter of the energy-loss electrons, which can be studied using the DP.

We note how the intense nature of the EELS spectrum, due mainly to the very high collection efficiency of the spectrometer, permits EELS imaging. Energy-loss (or energy-filtered) images and DPs can be formed in two ways: slowly, quantitatively, and digitally, or rapidly and qualitatively in an analog fashion. The primary advantage of EELS imaging is that *all* the information available in the spectrum can be imaged and related to all the other diffraction and imaging techniques that come from the TEM.

Everything Else in the Spectrum

40

40.1. FINE STRUCTURE IN THE IONIZATION EDGES

We saw in Section 38.4 that the ionization edges have intensity variations both within about 30 eV of the onset of the edge (ELNES) and extending 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.

> Both ELNES and EXELFS arise because the ionization process can impart more than the critical ionization energy (E_c) needed by the core electron to escape the attraction of the nucleus.

Any excess energy ($>E_c$) that the core electron possesses can be imagined as a wave emanating from the ionized atom. So again, we have 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 II. If this wave has only a few eV of excess energy, it undergoes plural elastic scattering from the surrounding atoms, as shown schematically in Figure 40.1A; this scattering is responsible for the ELNES, as we'll show. If the wave has even more excess energy, then it is less likely to be scattered several times and we can approximate the cause of the EXELFS 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 ELNES arises from a more complex process than EXELFS, it is more widely used, because the ELNES is more intense, and so we'll discuss it first.

40.1.A. ELNES

The Physics: A core electron may receive enough energy from the beam electron to be ejected, but not enough to escape to the vacuum level. So it is still not free of all specific nuclear attraction. 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 classical energy level diagram of an atom in Figure 40.2. In an insulator or a semiconductor, E_F is between the valence band (which has all filled states) and the conduction band (which has no filled states).

The EELS: The excited electron can reside in any of the unfilled states, but not with equal probability. 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 electron energy levels is termed the density of states (DOS) and this is also shown in the right diagram in Figure 40.2. Because of the greater probability of electrons filling certain unoccupied states above E_F, the intensity in the ionization edge is greater at the corresponding energy losses above the critical ionization energy E_C (which is equivalent to E_F), as shown in Figure 40.3.

> This variation in intensity, extending several tens of eV above E_C, is the ELNES and is effectively a probe of the DOS above E_F.

The Application: 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 and diamond and the Cu L ELNES changes when Cu is oxidized to CuO. On an even more detailed level, we can deduce the coordination of the ionized atom from the shape of the ELNES.

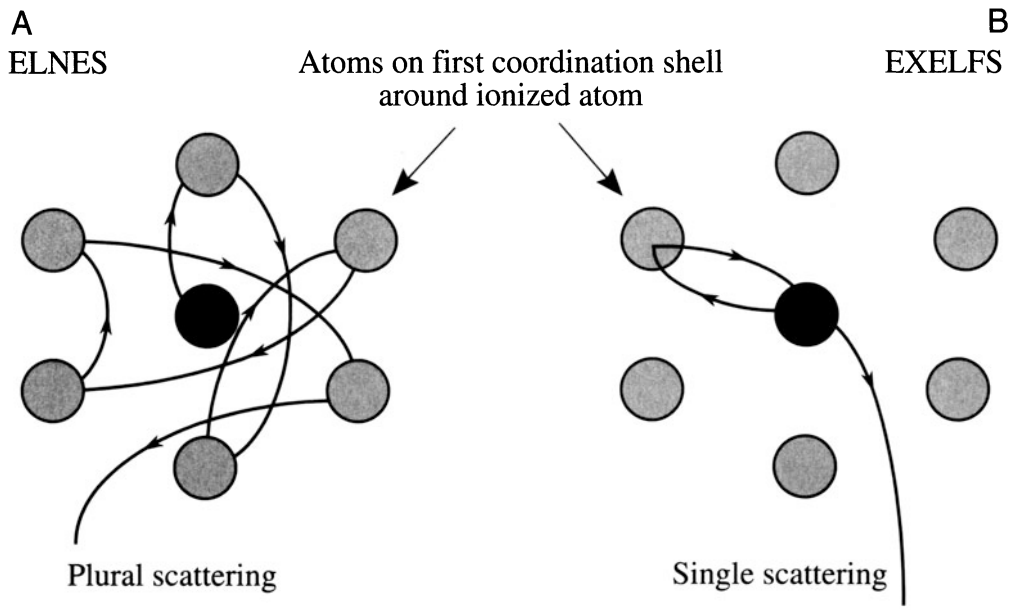


Figure 40.1. Schematic diagram showing the source of (A) ELNES and (B) EXELFS. The excess energy above the ionization threshold creates a wave radiating from the ionized atom which is scattered by surrounding atoms. The low-energy ELNES arises from multiple scatter and is affected by the bonding between the atoms. The higher-energy EXELFS approximates to single scatter and is affected by the local atomic arrangement.

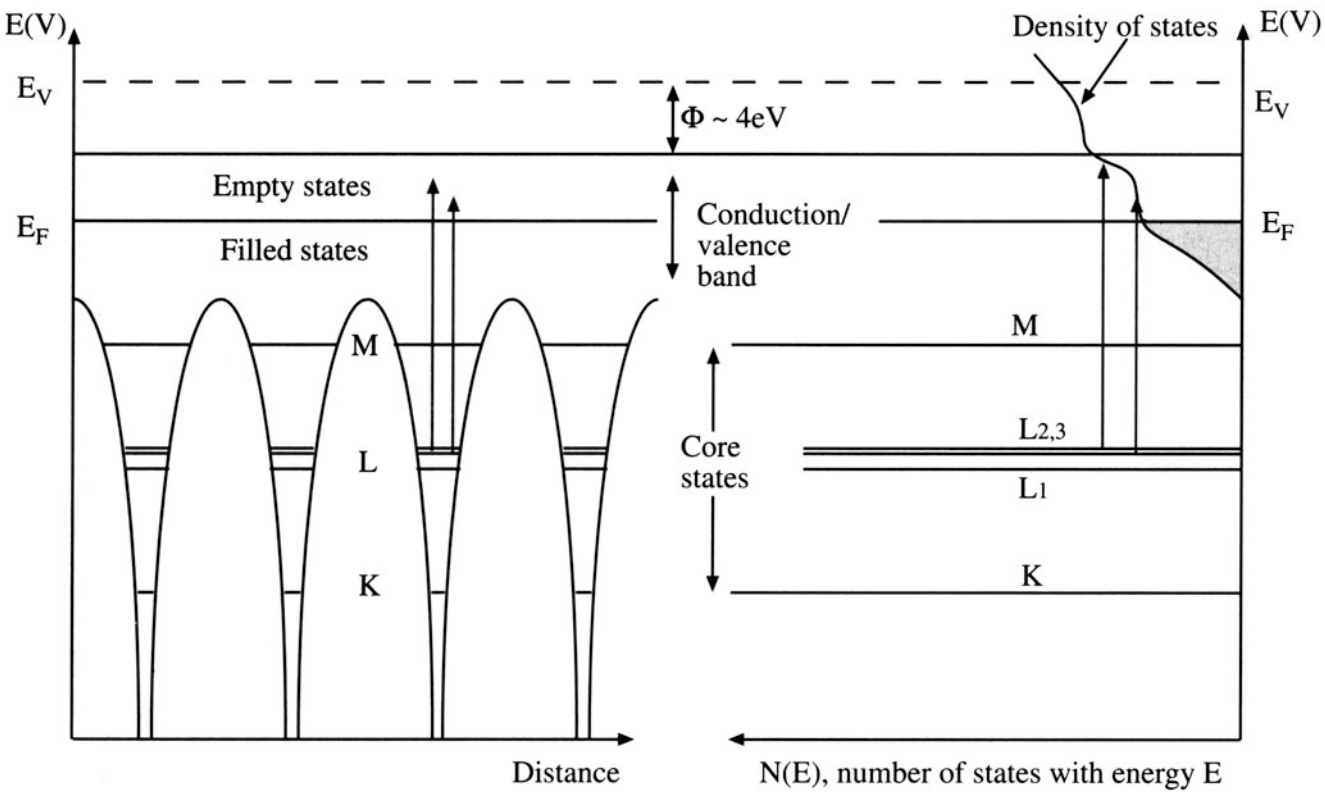


Figure 40.2. Relationship between the classical energy diagram of a metal atom (left) and the density of filled (shaded) and empty (unshaded) states (DOS) in the conduction band (right). The DOS is approximately a quadratic function on which small variations are superimposed. Ionization results in electrons ejected from the core states into empty states above the Fermi level (E_F).

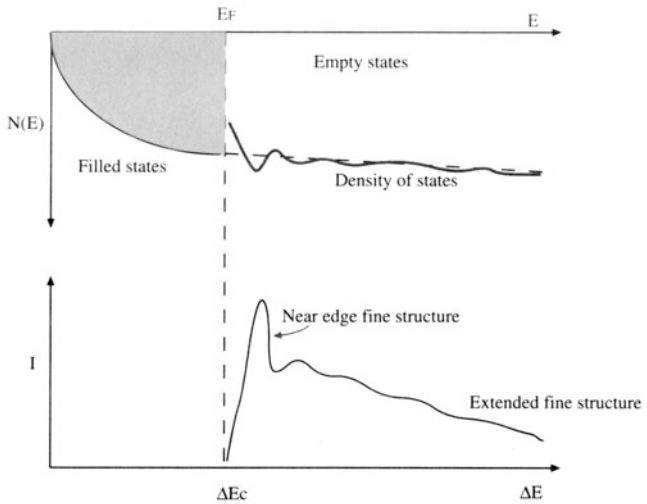


Figure 40.3. Relationship between the empty DOS and the ELNES intensity. Note the equivalence between the Fermi energy E_F and the ionization edge onset E_C. Electrons ejected from the inner shells reside preferentially in regions of the DOS with the greatest density of empty states. The filled states below E_F are shown as a quadratic function, but this is an approximation.

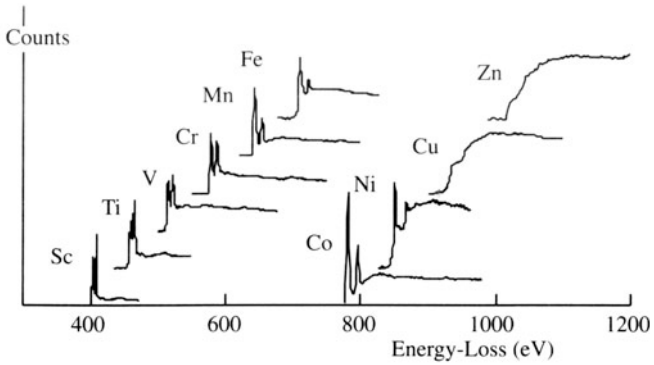


Figure 40.4. The L_3 and L_2 white lines in spectra from the transition metals show a slow variation in intensities until Cu, which has no white lines because the d shell is full.

Even if you don't understand the intricacies of the DOS and Fermi surfaces, you can still deduce bonding information simply by comparing your experimental ELNES with that from standard specimens of known valence state or coordination.

The EELS Atlas, which we've already referred to several times, contains many oxide spectra as well as elemental ones.

Perhaps the most startling example of ELNES is the presence of the "white lines," which we introduced in Section 38.4. They are intense sharp peaks on certain ionization edges; the L edges of the transition metals show such lines. Reminder: white lines were first seen on photographic plates in early X-ray absorption spectroscopy experiments. The white lines in the transition metal L edges are the L_3 and L_2 edges, respectively, as shown in Figure 40.4A. 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 not discuss this somewhat arcane argument.

More Physics: First, go back and look at Figure 38.4 to remind yourself 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 have no counterpart in the EELS spectra we obtain.

As we noted in Section 38.4, 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 equal 1/2, 3/2, 5/2, etc. The spin quantum number, s (not to be confused with the s state), can only equal ±1/2. Taking all this into account along with other quantum number restrictions, it turns out that in the higher-energy (more tightly bound) L_2 shell we can have 2 p electrons with j = ±1/2 while in the L_3 shell we can have 4 p electrons with j = ±1/2, ±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 of 2. While this rule is approximately obeyed in the Fe spectrum only, 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.

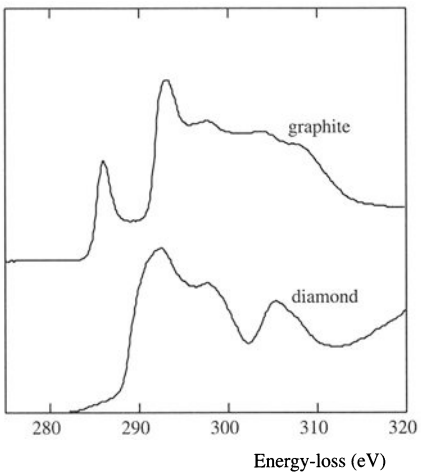
The change $\Delta\ell$ in the angular momentum quantum number between the initial and final states must equal ±1. This constraint is called the *dipole selection rule.*

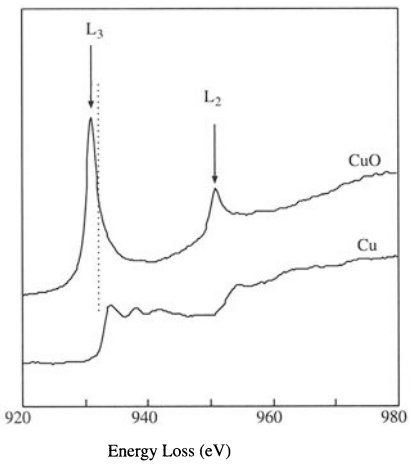
So for the p state ($\ell = 1$) the only permitted final states are either an s state ($\ell = 0$) or a d state ($\ell = 2$). Consequently, the electrons go up primarily into the unoccupied d states, since there are very few unfilled s states in the conduction band.

It is because of the dipole selection rules 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 ($\ell = 0$) so they can only be excited to a p state ($\ell = 1$), but not to a d state ($\ell = 2$), 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 be invisible in the $L_{2,3}$ post-edge structure.

In fact, 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 ΔE than the L_3 line, which is much sharper.

Back to Applications: So let's see how all of this can be useful. If you look at Figure 40.5A 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 sp^2 bonds in the basal plane with van der Waals bonding between the planes. The diamond structure, in contrast, has four directional hybridized sp^3 covalent bonds. In diamond, atoms are tetrahedrally coordinated rather than arranged in graphite sheets. The strong peak on the rising portion of the K edge identifies the empty π^* states into which the K-shell electrons are transferred in graphite, while the diamond K edge has no such peak. This kind of information is extremely useful in the study of thin diamond and diamond-like carbon films, which are of tremendous current interest to both semiconductor manufacturers and the coatings industry. 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 sp^3 (diamond) and sp^2 (graphite) bonding from the K-edge ELNES (Bruley *et al.* 1995). 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 classical example, since Cu metal has all its 3d states filled so 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





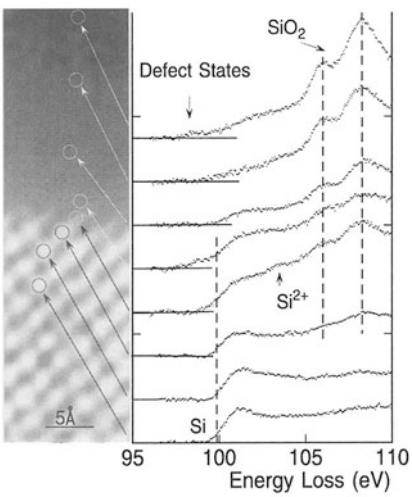


Figure 40.5. (A) Differences between the ELNES of the carbon K edge from graphite and diamond, (B) change in the Cu L edge as Cu metal is oxidized, (C) change in ELNES of the Si $L_{2,3}$ edge in a series of spectra gathered from individual atom rows across the interface between crystalline Si and amorphous SiO_2.

different from that of the metal, because this electron transfer changes the value of E_C.

This phenomenon is called a chemical shift and also helps to fingerprint the specimen.

Finally, in Figure 40.5C the Si L 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.

The combination of *Z*-contrast imaging (see Section 22.4) and PEELS is arguably the most powerful analytical technique for atomic characterization (see, e.g., Batson 1995, Browning and Pennycook 1995).

ELNES Calculations: Many attempts have been made to compare the ELNES with calculations of the density of states in simple materials such as metals and oxides. While the experimental and calculated spectra show reasonable agreement in terms of the energy of the various spectral features, there are still some discrepancies in the measured and calculated intensities. 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. There are two approaches to calculating the ELNES:

- Calculate the band structure directly in reciprocal space.
- Calculate the effect of multiple scattering of the electron wave in real space using the model shown in Figure 40.1A.

It can be shown that in fact these two approaches are mathematically equivalent. We'll emphasize the latter method since it is more commonly used.

Various approximations are made in ELNES modeling. The most critical approximation arises from the choice of the atomic potential. A common choice is the so-called "muffin-tin potential" in which a constant potential is assumed in the regions between atoms that don't touch. The potential within the atom must be spherically symmetrical. This model modifies the classical energy diagram, as shown in Figure 40.6. Apparently, this energy profile approximates to a cross section of a tin used by physicists to bake muffins. The potential profile across dissimilar adjacent atoms is asymmetrical, as also shown in Figure 40.6.

Having chosen a potential, the ELNES is determined by calculating all possible inter- and intra-shell scattering events suffered by the electron after it emerges above the Fermi level. A similar calculation is made in X-ray absorption near-edge structure (XANES) studies, and the two phenomena are equivalent. The wavelength is governed by the excess electron energy ($\mathcal{E}$-E_F). If the wavelength is long (i.e., low energy), the electron is scattered many times by the first few shells of atoms surrounding the ionized atom. One of the problems that confuses the issue is that the ion-

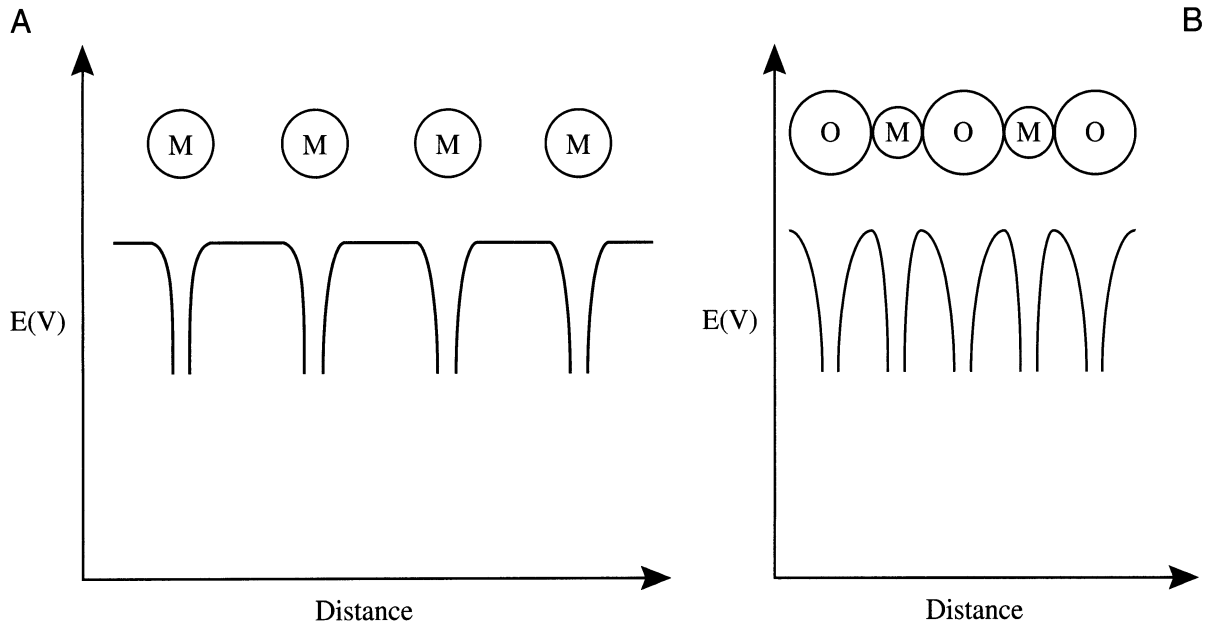


Figure 40.6. The muffin-tin potential energy diagram for (A) a non-closed-packed metal and (B) a metal oxide. Note the symmetry of the potential wells for the metal and the asymmetry for the oxide.

ization event results in a hole in the core shell, thus changing the atomic potential. This is called the *core-hole effect*, and it is accounted for by approximating the ionized atom to one with a nuclear charge of $Z+1$, since the missing electron lowers the shielding effect of the core electrons.

In ceramics and semiconductors the ionized electron remains localized to the ionized atom and may interact with the hole creating an electron-core hole bound state, termed an exciton. Creation of an exciton may influence the ELNES, although this remains a matter of some debate.

It can be shown that the multiple-scattering calculations predict modulations to the intensity of the ionization edge that 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.

Figure 40.7 shows a comparison of the calculated and theoretical ELNES for the Al L edge in tetrahedral and octahedral coordination in spinels. The difference due to different coordination 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 seminal paper in the field of ELNES experiments on transition metals and oxides is by Leapman *et al.* (1982), and a concise summary was given by Brydson (1991).

40.1.B. EXELFS

If the ejected electron does not fill an empty state, then its excess energy can also be interpreted as an electron wave which can be diffracted by the surrounding atoms in the structure, giving rise to EXELFS. Because the electron has higher energy than those which gave rise to ELNES, the diffraction is assumed to be single scattering, as shown in Figure 40.1B. As with any diffraction event, there is information about atomic positions in the EXELFS.

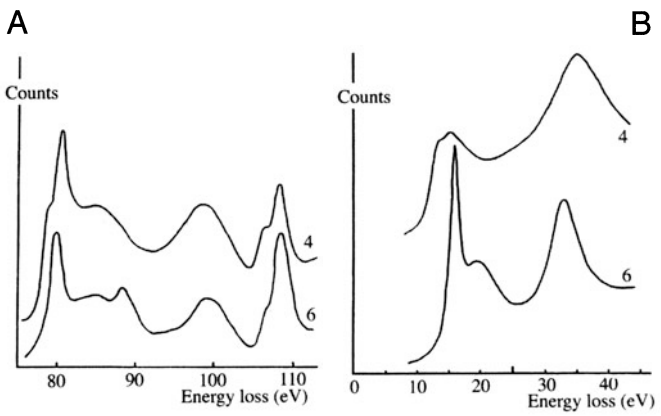


Figure 40.7. Comparison of (A) experimental spectra and (B) theoretical ELNES calculations for the Al $L_{2,3}$ edge in tetrahedrally coordinated (CN-4) and octahedrally coordinated minerals (CN-6). The calculated energy axis in (B) refers to eV above the L edge onset of ~75 eV.

So ELNES is multiple scattering and EXELFS is single scattering, although the two phenomena overlap since the L_1 ELNES peak is often far enough past the edge onset to be included in the EXELFS.

The EXELFS modulations are each 20–50 eV wide (just visible in Figure 40.8A), and continue for several hundred eV. EXELFS is exactly analogous to the oscillations seen in the extended X-ray absorption edge fine structure (EXAFS) in synchrotron X-ray spectra. However, EXAFS results from complete photoabsorption of the incident X-ray while EXELFS involves absorption of only a small fraction of the energy of the beam electron.

Experimentally, it's not easy to see the EXELFS modulations because they are only about 5% of the edge intensity, and so you need good counting statistics. With SEELS you may have to gather the spectrum for many minutes or even hours, so PEELS is the only realistic way to pursue EXELFS. A thermionic source is probably best because it can deliver more current than an FEG; for this application, energy resolution is often less important. TEM diffraction mode will also increase your total signal intensity. Either way, 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, an FEG and STEM mode is best.

We're interested in EXELFS because of the structural information contained in the intensity oscillations. To extract this information, you can use the Gatan ELP software (see Section 1.5), but you first have to ensure that the spectrum contains single-scattering information only, otherwise the plural-scattering intensity may mask the small EXELFS peaks.

Deconvolution is always the first step if the specimen isn't thin enough, i.e., if the plasmon peak is greater than 10% of the zero-loss peak.

Next, you have to remove the background if it wasn't done prior to deconvolution. Then the EXELFS intensity modulations are fitted to a smooth curve, and the intensity either side of the curve is plotted in **k** space (reciprocal space) (Figure 40.8B)

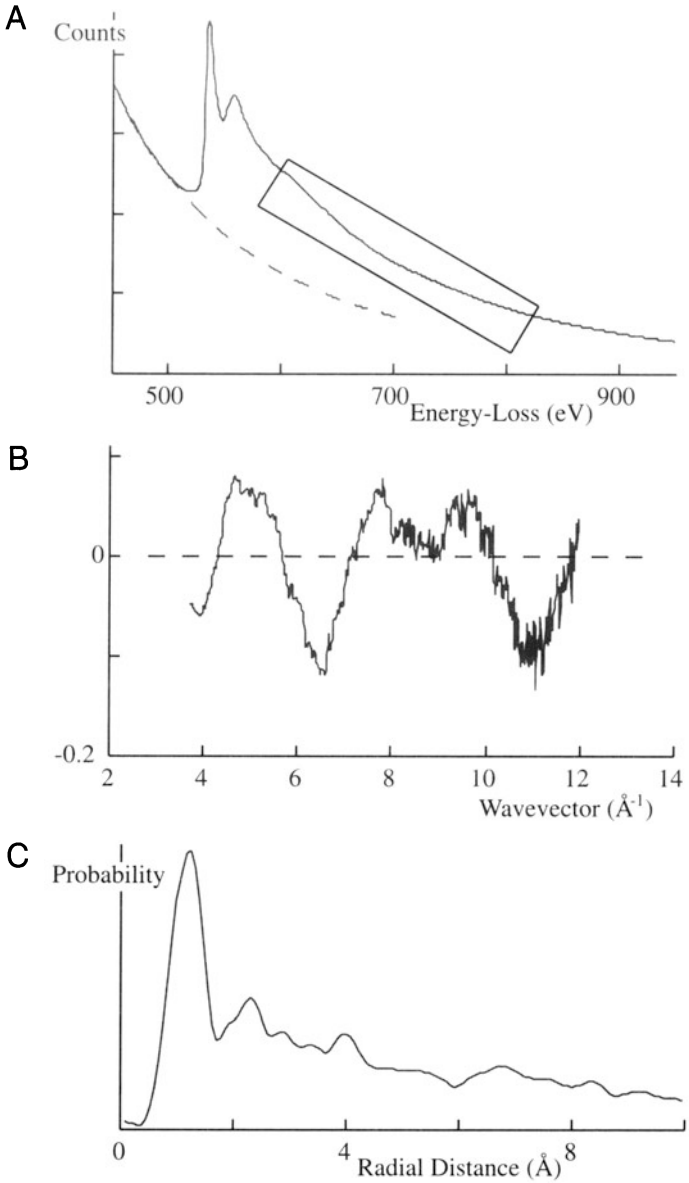


Figure 40.8. (A) EXELFS modulations are barely detectable in the selected post-edge region of an ionization edge. (B) The oscillations either side of a curve fitted to the post-edge spectrum are plotted in k space before (C) Fourier transforming the data to produce a radial distribution function.

$$k = \frac{2\pi}{\lambda} = \frac{[2m_0(E - E_K)]^{\frac{1}{2}}}{h} \qquad [40.1]$$

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, there are periodic maxima occurring for n = 1, 2, etc., and for different interatomic spacings. Consequently, you should be able to determine the local atomic environment, if the various interferences can be discriminated. The atomic spacing is obtained by a Fourier transform of the **k**-space modulations to give a radial distribution function, originating at the ionized atom (Figure 40.8C). Peaks in the RDF indicate the probability of an atom occurring a certain distance from the origin.

With EXELFS we can determine the partial 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, and quasicrystalline structures. The high spatial resolution is obviously advantageous and all the data can be compared with diffraction patterns and images of the analyzed area. However, like all EELS techniques, we can't get good EXELFS unless the specimen is very thin. Despite these advantages, RDF work continues to be dominated by synchrotron sources because of the intensity of the signal, but EXELFS studies are increasing, e.g., Sklad *et al.* (1992), Qian *et al.* (1995).

RDF data acquired through EXELFS complement another TEM method of acquiring RDF information. This involves energy filtering of SAD patterns by scanning the pattern across the entrance aperture to the PEELS using post-specimen scan coils (Cockayne *et al.* 1991; see also Sections 18.6 and 40.3). 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 zero-loss intensity as a function of scattering angle constitutes a line profile across a filtered diffraction pattern from which the RDF can be extracted; you can see a related example if you look ahead to Figure 40.15. This process does not have the spatial resolution of EXELFS, since typical SAD patterns are integrated over ~0.2–1 μm^2, but the signal is much stronger than EXELFS. Accuracies of ±0.001 nm in nearest-neighbor distances can be obtained, and the process is rapid enough to be performed on-line.

The techniques of ELNES and EXELFS are really quite remarkable demonstrations of quantum theory and the wave-particle duality. Consider that within the spectrum we are only gathering beam electrons that have been scattered by the specimen atoms, yet we are able to deduce information about what happened to those atoms *after* the beam–specimen interaction and where the atoms are in the structure!

An approximate 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 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 also to 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 core 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 know the possible final state of the core electron.

40.2. THE LOW-LOSS SPECTRUM

40.2.A. Plasmon Losses

The low-energy plasmon-loss region of the spectrum also contains chemical information, because the composition of the specimen may affect the free-electron density, n, which in turn changes the plasmon-energy peak position, since the two are related, as we described back in equation 38.6. Historically, this technique was the first aspect of EELS to produce quantitative microanalysis data, and it has been 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. For a review see Williams and Edington (1976).

The principle of plasmon-loss microanalysis 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

$$\mathcal{E}_P(C) = \mathcal{E}_P(0) \pm C\left(\frac{d\mathcal{E}_P}{dC}\right) \qquad [40.3]$$

where $\mathcal{E}_P(0)$ is the plasmon energy 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. Table 40.1 summarizes the available plasmon-loss data for Al alloys, gathered in this manner.

Table 40.1. Alloys in Which the Variation of Plasmon Energy Loss $\mathcal{E}_p$ Has Been Measured as a Function of Composition

Alloy (at. %)	Range	$\mathcal{E}_p$ (eV) variation with fractional concentration C
Al-Mg	0–100	$\mathcal{E}_p = 15.3 - 5.0\ C_{Mg}$
Al-Mg	0–8	$\mathcal{E}_p = 15.3 - 4.4\ C_{Mg}$
Mg-Al	0–9	$\mathcal{E}_p = 10.61 + 5.9\ C_{Al}$
Al-Cu	0–2	Nonlinear
Al-Cu	0–2	$\mathcal{E}_p = 15.3^a - 10\ C_{Cu}$
Al-Cu	0–17.3	$\mathcal{E}_p = 15.3 + 4.0\ C_{Cu}$
Al-Zn	0–30	$\mathcal{E}_p = 15.3 - 0.2\ C_{Zn}$
Al-Ag	0–6	$\mathcal{E}_p = 15.3^a + 1.6\ C_{Ag}$
Al-Li	0–25	$\mathcal{E}_p = 15.3^a - 4.0\ C_{Li}$
Al-Ge	0–10	$\mathcal{E}_p = 15.3 + 0.1\ C_{Ge}$
Al-Zn-Mg	0–4	$\mathcal{E}_p = 15.3 - 4.7\ C_{Mg}$

[a]Normalized to 15.3 eV energy loss for pure Al.

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 access to FEGs and so the resolution of the thermionic source was a limiting factor. The poor resolution was compensated for to some extent by utilizing a high-dispersion Wien Filter or a Möllenstedt electrostatic spectrometer and recording the spectra photographically. More recently, similar results have been achieved using the relatively low-dispersion magnetic-prism spectrometer and electronic recording, but with an FEG. While the shift in the position of the plasmon peak may be as small as ~0.1 eV, the position of the peak centroid can still be measured to an accuracy of ~0.05 eV by computerized peak-fitting (Hunt 1995). Figure 40.9 illustrates some early plasmon-loss concentration data and the visible peak shifts that occur.

Plasmon-loss spectrometry has high 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 about 10 nm, since the plasmon disturbance is rapidly damped in the free-electron gas. Your specimen thickness only affects the number and intensity of the plasmon peaks, not their position, as we described back in Figure 38.2. 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. The plasmon signal is intense and is the dominant loss feature in the spectrum. 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.

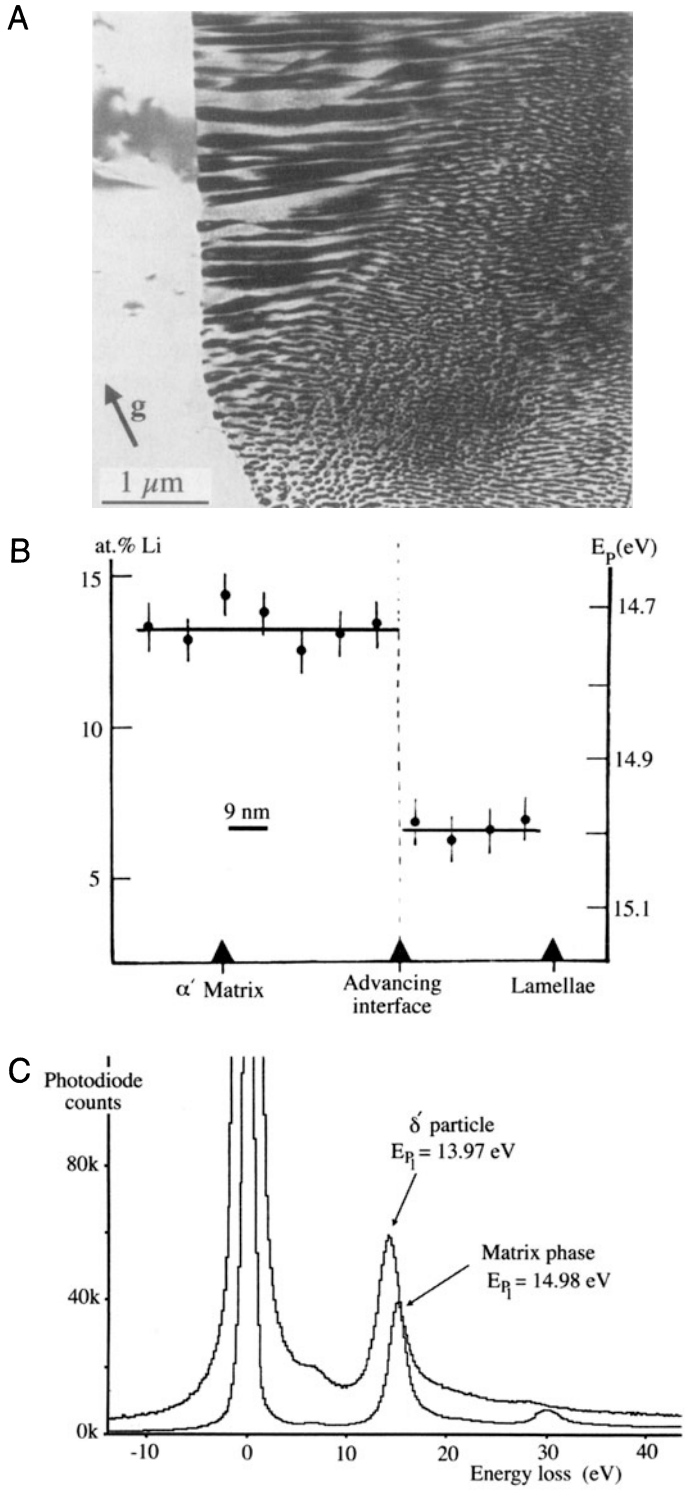


Figure 40.9. (A) Discontinuous transformation interface in Al-11 at.% Li. (B) Plasmon-loss variation and related Li composition change across the interface in (A). (C) Comparison of spectra from the matrix (5 at.% Li) and the precipitate (25 at.% Li) reveals the shift in the plasmon peak.

We are limited to specimens showing well-defined peaks, and only binary specimens can be sensibly analyzed.

In addition, the alloying element must produce a detectable change in $\mathcal{E}_p$ and this is by no means 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 acquiring and analyzing plasmon-loss spectra. However, it is not clear that they will permit the technique to be expanded significantly past the limited range of materials to which it has already been successfully applied.

While quantitative plasmon-loss microanalysis is limited, you can still use the plasmon part of the spectrum to identify unknown phases by the technique of "fingerprinting," as we showed in Figure 38.3. The low-loss portion of the spectrum is often sufficiently distinctive for different compounds that, since suitable libraries of known spectra exist that we've already referenced, you can use these libraries to cross-check the spectra from unknowns. In fact the plasmon-loss spectrum is more robust than the ionization-loss spectrum, since it will not change significantly as you change such experimental variables as α, β, kV, and it is insensitive to the data-processing variables that plague ionization-loss spectra. For example, direct examination of the low-loss spectrum is sufficient to distinguish between free-electron metals and transition metals, as shown in Figures 40.10A,B. Similarly, the low-loss spectra of the different oxides are equally distinctive (Figure 40.10C).

40.2.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, your energy-loss spectrum contains information about the dielectric constant or permittivity (ε). The single-scattering spectrum intensity $I(\ell)$ is related to ε by the expression (Egerton 1996)

$$I(\ell) = I_0 \frac{t}{k} \mathrm{Im}\left(-\frac{1}{\varepsilon}\right) \ln\left|1 + \left(\frac{\beta}{\theta_E}\right)^2\right| \qquad [40.4]$$

where I_0 is the intensity in the zero-loss peak, t is the specimen thickness, and k is a constant incorporating the electron momentum and the Bohr radius. 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 (Im) in equation 40.4, and details

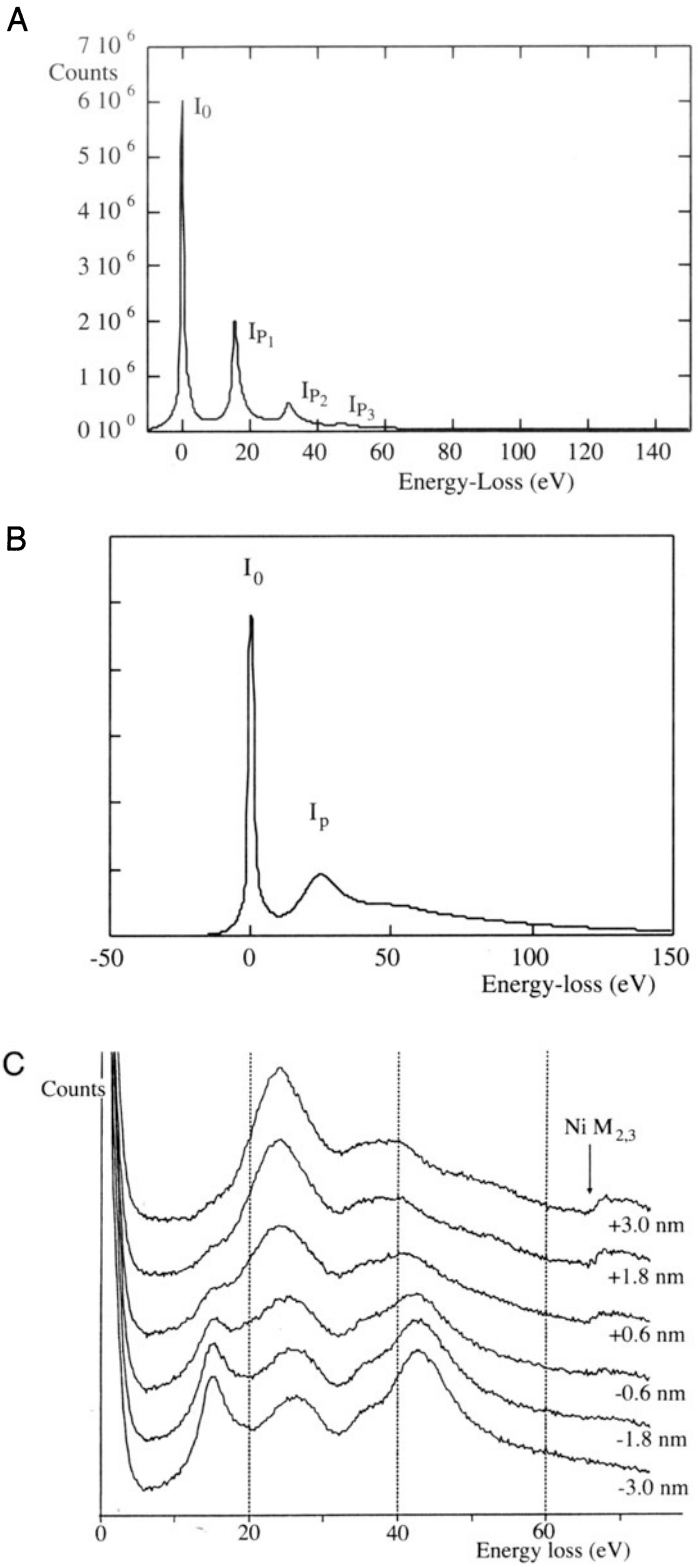


Figure 40.10. (A) Multiple plasmon peaks from Al, which is a free-electron metal, compared with (B) the single weak plasmon from a transition metal, Fe. (C) Six low-loss spectra taken across a NiO (top)–ZrO_2 (bottom) interface showing characteristic differences in the plasmon intensities which occurred within ±3 nm of the interface.

are given in Egerton (1996). Since you need a single-scattering spectrum, deconvolution is again the first step.

> The Kramers–Kronig analysis gives the energy dependence of the dielectric constant and other information which we usually obtain by optical spectroscopy.

The advantage of EELS for this kind of work is the improvement in spatial resolution over electromagnetic radiation techniques. Also, the frequency range which is available is more extended. The low-energy plasmon part of the energy-loss spectrum out to about 20 eV is of most interest to us, and corresponds to optical analysis of the dielectric response from the visible through the ultraviolet frequency range. So in a single EELS experiment you can, in theory, substitute for a whole battery of optical spectroscopy instrumentation. Physicists are most interested in the low-frequency range around 1 eV, since this is less accessible through optical spectroscopy. For this you need an FEG and a high-resolution spectrometer, and you need to deconvolute out the tail of the zero-loss peak so it does not mask the low-energy intensity. An example of the correspondence between EELS and optical dielectric constant spectra is shown in Figure 40.11.

40.2.C. Band-Gap and Interband Transitions

In the region of the spectrum immediately after the zero-loss peak, and before the rise in intensity preceding the plasmon peak, you can see a region of low intensity. If this intensity approaches the dark noise of the detector, then there are no electron–electron energy transfers occurring. This effect implies that there is a forbidden transition region, which is simply the band gap between the valence and conduction bands in semiconductors and insulators. Figure 40.12A illustrates the variable band gap from specimens of Si, SiO_2, and Si_3N_4. In this region there are sometimes small peaks that correspond to interband transition which require energy losses of <10 eV, and surface plasmons may occur. An example of an interband transition is given in the spectra from two polymers shown in Figure 40.12B.

40.2.D. Angle-Resolved EELS

Most of the time we've been talking about locating the beam at different positions on the specimen and gathering a spectrum by sending the direct beam into the spectrometer. This is often called "spatially resolved" EELS since spectra came from different spatial locations on the speci-

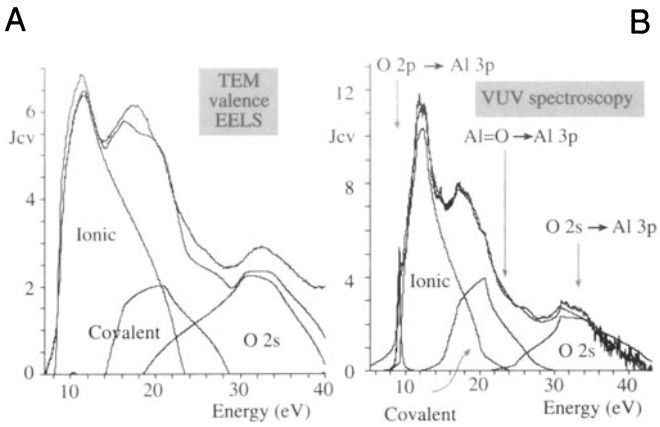


Figure 40.11. Comparison of thin-specimen EELS (A) and bulk-specimen optical dielectric constant data (B) for α-Al_2O_3. J_{cv} is the interband transition strength and the various transitions are labeled: transitions from the filled O 2p level represent ionic bonding, transitions from the hybridized Al=O level represent covalent bonding, interband transitions from O 2s–Al 3p are also detected. Individual contributions to the spectra have been obtained via a critical point model.

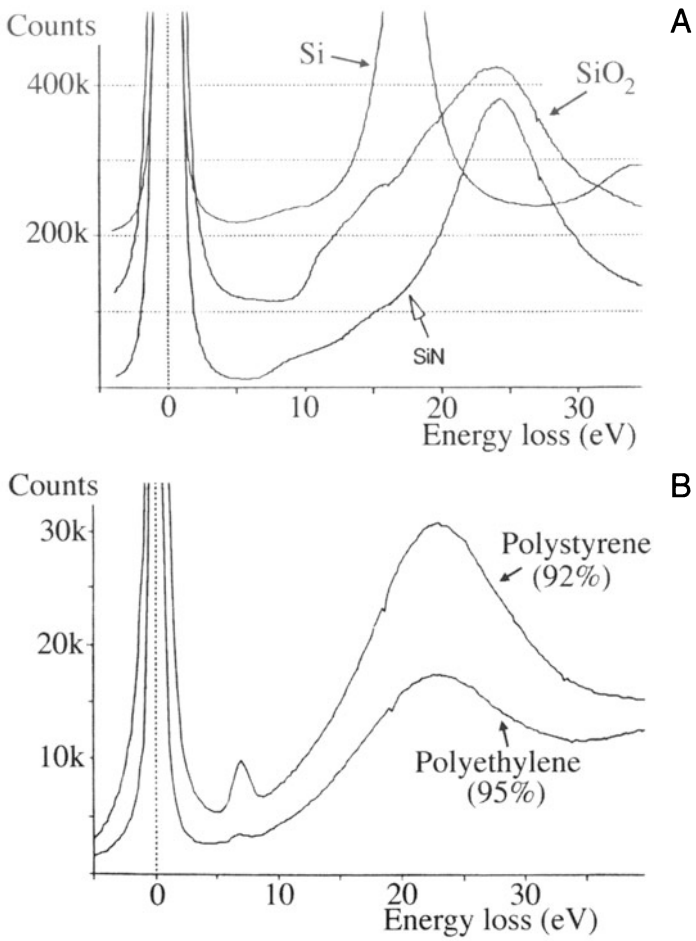


Figure 40.12. (A) Band-gap differences evident in the low-loss spectra of a Si semiconductor and SiO_2 and Si_3N_4 ceramic insulators. (B) The interband transition characteristic of polystyrene, clearly visible on the rise of the plasmon peak, compared with the absence of such a transition in polyethylene.

men. However, we have occasionally mentioned that the *angle* of scatter of 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 (or the SEELS slit) and gather spectra at different angles, as for 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 (e.g., Leapman and Silcox 1979), and 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.* 1995).

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 (about 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 (Schattschneider and Exner 1995).

You can appreciate now that there is a wealth of detail in the energy-loss spectrum beyond the basic chemistry of the specimen. To extract this information you need a single-scattering (deconvoluted) spectrum and sophisticated mathematical analysis. Often, our interpretation of the data is limited by lack of knowledge of the physics of the electron–specimen interaction. However, considerable research is going on into these aspects of EELS and these fine structure studies are the future of the technique.

40.3. ENERGY-FILTERED AND SPECTRUM IMAGING

We can select the intensity in any part of the EELS spectrum and use it to form an image, either in a digital manner by modulating the signal to the STEM screen or in an analog manner in the energy-selecting TEM. A variety of images can be formed in this way and they have several advantages over conventional TEM and STEM images. We will describe the experimental procedures first and then discuss the different types of images.

40.3.A. STEM Digital Imaging

In SEELS, the ramp voltage to the magnetic prism must be held at a constant setting, so only those electrons in the energy range accepted by the slit pass through the spectrometer. In PEELS you select the output from specific diodes. In either case you've used the spectrometer to select electrons of a fixed energy range. If these electrons are then allowed to hit a detector, the signal can be used to form energy-filtered images. In a TEM/STEM you use the signal from the EELS scintillator to modulate the STEM CRT, while in a dedicated STEM the BF detector sits beyond the EELS and so all your BF images are energy-filtered. To avoid image shifts due to scanning of the beam on the specimen, you must descan the beam using a set of post-specimen coils, which are usually present in the TEM or STEM column as a matter of course. In SEELS, if the spectrometer slit width is too large, your image will suffer from chromatic aberration because electrons of different energy are focused at different planes. In PEELS, the intensity is controlled by the total spectral-acquisition time per pixel. Here, an FEG is best if high-resolution images with reasonable pixel numbers are to be acquired.

With PEELS, it is possible to collect a spectrum in a sufficiently short time (< 50 ms) that you can create images in which a complete spectrum is stored at each pixel and all data processing is carried out after the acquisition.

Then we have what is known as a *spectrum image* (Jeanguillaume and Colliex 1989). Such images contain immense amounts of data and you need sophisticated hardware and software routines to handle the data. For example, a 512×512 pixel image with a full 1024 channel spectrum at each pixel contains more than 25 Mb of data. Because of the relatively long time to acquire spectrum images, drift correction and other PEELS corrections are necessary (Hunt and Williams 1991). Within such an image you have a *complete* record of the electron–specimen interaction and from such you can create multiple images, as shown in Figure 40.13. This figure is from an Al-Li alloy and shows the distribution of the component elements. There are three strong advantages to this approach:

- You can analyze the "specimen" at a later time, without putting it back in the microscope, and look for elements that were not initially thought to be present or to be important.
- You can process the data in several different ways to compare quantification schemes and the possibility of discerning unexpected correlations between elemental distributions.
- All the information in the EELS spectrum can be mapped discretely, creating, for example, not just elemental images, but dielectric-constant images, valence-state images, thickness images, etc.

40.3.B. TEM Analog Imaging

For analog imaging in a TEM, an Ω filter spectrometer sits between the first and second pairs of projector lenses, as shown back in Figure 37.11 for the LEO EM912. To select

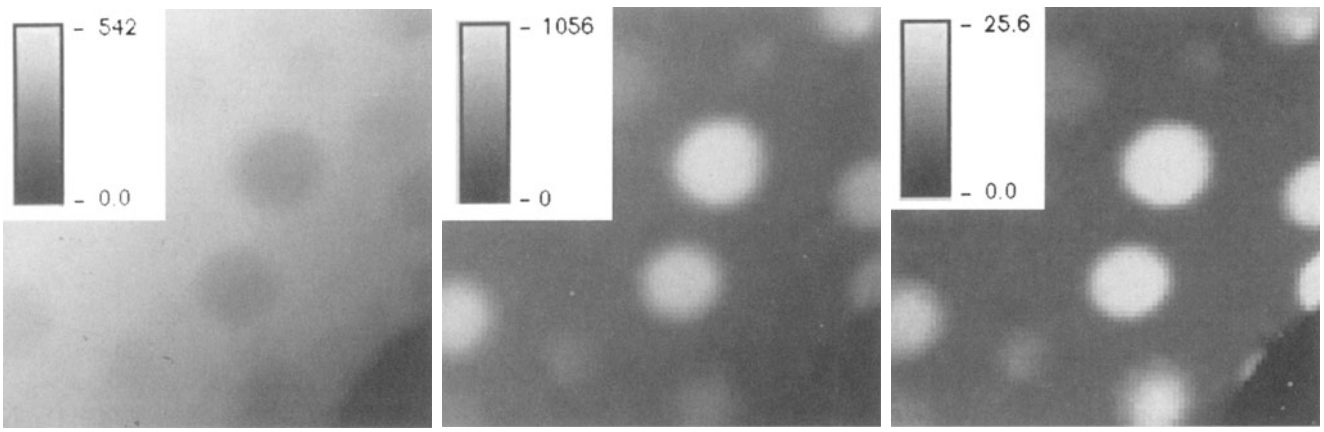


Figure 40.13. Three processed spectrum images of an Al-Li alloy aged to give a dispersion of δ′ precipitates. The left image shows the absolute concentration of Al (atoms/nm^2) obtained from quantification of the Al $L_{2,3}$ edge, the middle image is the absolute Li content (atoms/nm^2) from the Li K edge, and the right image is the Li content (at. %) obtained from the shift of the first plasmon peak. The inserts show the correlation between image intensity and the range of composition imaged.

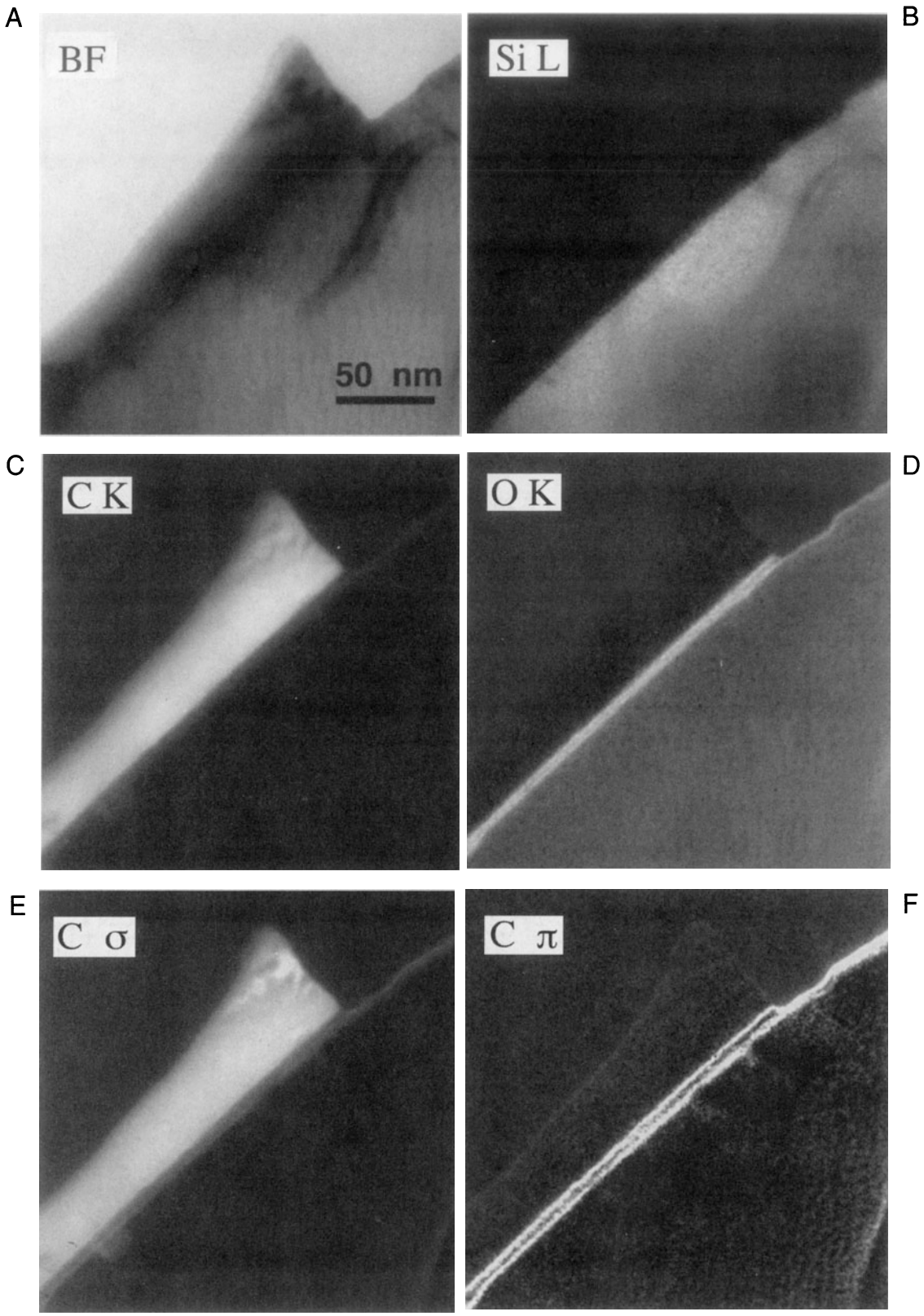


Figure 40.14. (A) TEM BF and (B–F) a series of electron spectroscopic images revealing the Si, C, and O elemental distributions and the carbon bonding maps at the interface between a diamond-like carbon film and a Si substrate. In the oxygen-rich amorphous layer at the interface the carbon atoms exhibit a double layer of π bonds while the carbon film itself is predominantly σ bonded, indicating a high degree of diamond-like character.

the electrons for this imaging (ESI), you shift the spectrum relative to the slit that is positioned after the filter but before the final projector lens. In the LEO instrument, you make the shift by increasing the accelerating voltage of the microscope by $+E$ in order to keep the energy-loss electrons of interest ($-\mathcal{E}$) on the optic axis; this shift correction is prealigned for the chosen kV. You can filter either an image or DP simply by changing the strength of the intermediate lens preceding the Ω filter. In addition, you can also see the energy-loss spectrum on the TEM screen. This method of spectral display has the advantage that the angular distribution of the energy-loss electrons is spatially resolved, although the absolute intensity has to be determined by digitizing the spectrum or using a microdensitometer. However, this is not the major mode of operation of the instrument, which is optimized for electron spectroscopic imaging (ESI), and many examples are given in a special issue of the *Journal of Microscopy* (Knowles 1994).

For ESI, you adjust the energy window by varying the slit width. With a 20-eV window you can obtain images with a chromatic-aberration limit of ~2.5 nm, which compares well with normal TEM C_c-limited resolution. Resolution may be as good as 0.5 nm under ideal conditions. You can select the area to image via a selected-area aperture, or by using Kohler illumination conditions, in which a small parallel beam of electrons is created.

A drawback to this ESI process is that, while background-subtracted core-loss images are easily obtainable, they are not quantitative if significant changes in specimen thickness occur.

However, the images can be acquired in a few seconds, rather than many minutes or hours for a digital image, and a range of filtered images are compared with a conventional TEM BF image in Figure 40.14A–F. ESI is equally applicable to diffraction patterns; energy-filtered CBED is a very powerful technique for extracting more data from conventional CBED patterns, as shown in Figure 40.15. Deininger *et al.* (1994) have demonstrated how energy-filtered CBED patterns can be used to determine structure factors, lattice strains, and the accelerating voltage of the TEM. Removing the inelastic electrons removes much of the diffuse scattering from your diffraction patterns, making comparison of experimental and simulated patterns much easier. In addition to energy-filtered CBED patterns, as shown in Figure 40.15, SAD patterns can be similarly sharpened up, and used for RDF determination, as we already mentioned in Section 40.1.B

In summary, you can perform EELS imaging in two very different ways in a TEM and STEM. You can obtain a variety of images, depending on which portion of your spectrum is selected and the nature of your specimen. The strong forward-scattered EELS signal, combined with close to 100% detection efficiency, means that EELS imaging is much more statistically viable than thin-film X-ray mapping. This fact, combined with the enormous number

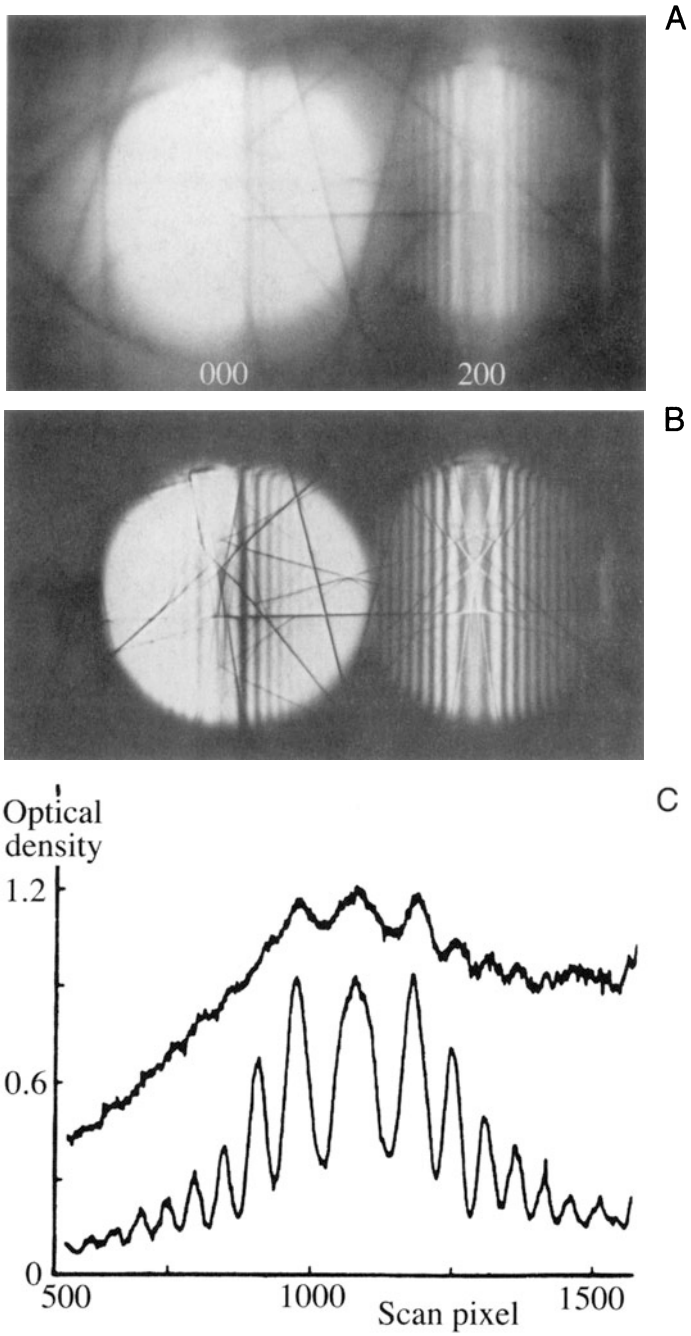


Figure 40.15. (A) Experimental CBED two-beam pattern (000 and 220) from a Si specimen, 270 nm thick. (B) The same pattern energy-filtered using the Zeiss Ω filter with an 8-eV window revealing the K-M fringes useful for thickness determination. (C) Densitometer traces across the 220 diffraction disk, unfiltered (above) and filtered (below).

of signals available in the EELS spectrum, make EELS imaging an extremely attractive technique.

Combination of energy filtering with conventional TEM imaging and diffraction will increase as the development of digital TEM technology continues. It is likely that all images will routinely be filtered to remove chromatic aberration effects, which will be a tremendous aid to the materials scientist struggling to make thin specimens from complex multiphase materials. As digital storage becomes cheaper, the ability to save complete spectrum images of all your specimens will become the norm, thus enhancing the claim that the TEM is *the* most versatile instrument for the characterization of materials.

As a final word, never forget to combine techniques wherever possible to characterize your material. If you are creative, you can even do simultaneous experiments, e.g., by constructing an STM in a TEM (Spence *et al.* 1990). We encourage you to experiment with the microscope at *all* opportunities. Don't think there is nothing new to discover; there is still ample room for you to exercise your imagination and innovation, and the TEM is a fascinating place in which to do just that.

CHAPTER SUMMARY

Both the experimental techniques and the theoretical understanding of EELS are still developing. We have introduced several specialized topics:

- Energy-loss near-edge structure (ELNES).
- Extended energy-loss fine structure (EXELFS).
- Low-loss fine structure.
- Angle-resolved (momentum transfer) EELS.
- Electron spectroscopic imaging (ESI) and spectrum imaging.

However, we have only given you a suspicion of the potential of these topics. If EELS becomes a technique you use in your research, or if you have the time, we recommend watching developments of the technique in the journals referenced at the end of this chapter. EELS, particularly fine structure and imaging, is one of the most dynamic areas of TEM development.

REFERENCES

General References

Egerton, R.F. (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope,* 2nd edition. Plenum Press, New York.

Raether, H. (1965) *Electron Energy-Loss Spectroscopy, Springer Tracts in Modern Physics,* Springer-Verlag, New York.

Teo, B.K. and Joy, D.C. (1981) *EXAFS Spectroscopy; Techniques and Applications,* Plenum Press, New York.

Specific References

Batson, P.E. (1995) *Ultramicroscopy* **59**, 63.

Browning, N.D. and Pennycook, S.J. (1995) *J. Microsc.* **180**, 230.

Bruley, J., Williams, D.B., Cuomo, J.J., and Pappas, D.P. (1995) *J. Microsc.* **180**, 22.

Brydson, R. (1991) *EMSA Bulletin* **21**, 57.

Cockayne, D.J.H., McKenzie, D., and Muller, D. (1991) *Microsc. Microanal. Microstruct.* **2**, 359.

Deininger, C., Necker, G., and Mayer, J. (1994) *Ultramicroscopy* **54**, 15.

Hunt, J.A. (1995) in *Microbeam Analysis-1995* (Ed. E.S. Etz), p. 215, VCH Publishers, New York.

Hunt, J.A. and Williams, D.B. (1991) *Ultramicroscopy* **38**, 47.

Jeanguillaume, C. and Colliex, C. (1989) *Ultramicroscopy* **28**, 252.

Knowles, K.M., Ed. (1994) *J. Microsc.* **174**, 131.

Leapman, R.D. and Silcox, J. (1979) *Phys. Rev. Lett.* **42**, 1362.

Leapman, R.D., Grunes, L.A., and Fejes, P.L. (1982) *Phys. Rev.* **B26**, 614.

Qian, M., Sarikaya, M., and Stern, E.A. (1995) *Ultramicroscopy* **59**, 137.

Schattschneider, P. and Exner, A. (1995) *Ultramicroscopy* **59**, 241.

Sklad, P., Angelini, P., and Sevely, J. (1992) *Phil Mag.* **A65**, 1445.

Spence, J.C.H., Lo, W., and Kuwabara, M. (1990) *Ultramicroscopy* **33**, 69.

Wang, Y.Y., Cheng, S.C., Dravid, V.P., and Zhang, F.C. (1995) *Ultramicroscopy* **59**, 109.

Williams, D.B. and Edington, J.W. (1976) *J. Microsc.* **108**, 113.

Index

Note: all materials examples are indexed under "Materials examples in text."

Acknowledgments for Reproduction of Figures

TEM is a visual science, and any TEM text is heavily dependent on figures and halftones 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 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.

Figures 1.4A: Courtesy of S.M. Zemyan.

Figures 1.4B,C: Courtesy of S.M. Merchant.

Figure 1.6: Courtesy K.S. Vecchio.

Figure 1.7: Courtesy T. Hayes, from Hayes, T. (1980) in Johari, O. (Ed.) SEM—1980 **1**, 1, Fig. 8, reproduced by permission of Scanning Microscopy International.

Figure 1.9A: Courtesy of JEOL USA Inc.

Figure 1.9B: Courtesy of NSA Hitachi Scientific Instruments, Ltd.

Figure 1.9C: Courtesy of Philips Electronic Instruments, Inc.

Figure 1.9D: Courtesy of V.G. Scientific.

Chapter 2

Figure 2.4: Courtesy of J. Bruley and V.J. Keast.

Figure 2.8: Modified from Hecht, E. (1988) *Optics* Fig. 10.21, Addison-Wesley.

Figure 2.11A,D: Courtesy of K.S. Vecchio.

Figure 2.11C: Courtesy of D.W. Ackland.

Chapter 3

Figure 3.3: Courtesy of D.E. Newbury, from Newbury, D.E. (1986) in Joy, D.C., *et al.* (Eds.) *Principles of Analytical Electron Microscopy*, p. 6, Fig. 2, reproduced by permission of Plenum Press.

Figure 3.4: Courtesy of D.E. Newbury, from data in Newbury D.E. (1986) in Joy, D.C., *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 D.E. Newbury, from Newbury, D.E. (1986) in Joy, D.C., *et al.* (Eds.) *Principles of Analytical Electron Microscopy*, p. 20, Fig. 4, reproduced by permission of Plenum Press.

Figure 4.3: From Woldseth, R. (1973) *X-ray Energy Spectrometry*, Fig. 3, reproduced by permission of Kevex Instruments.

Figure 4.4: From Williams, D.B. (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 L.W. Hobbs, from Hobbs, L.W. (1979) in Hren, J.J., *et al.* (Eds.) *Introduction to Analytical Electron Microscopy*, Fig. 17.2, reproduced by permission of Plenum Press.

Figure 4.12: Courtesy of L.W. Hobbs, from Hobbs, L.W. (1979) in Hren, J.J., *et al.* (Eds.) *Introduction to Analytical Electron Microscopy*, Fig. 17.4, reproduced by permission of Plenum Press.

Table 4.1: Courtesy of J.I. Goldstein, from Goldstein, J.I., *et al.* (1992) *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd edition, Table 3.11, reproduced by permission of Plenum Press.

Table 4.2: Courtesy of N.J. Zaluzec and J.F. Mansfield, from Zaluzec, N.J. and Mansfield, J.F. in Rajan, K. (Ed.) *Intermediate Voltage Electron Microscopy and Its Ap-*

plication to Materials Science (1987), p. 29, Table 1, reproduced by permission of Philips Electron Optics.

Chapter 5

Figure 5.1: Modified from Hall, C.E. (1966) *Introduction to Electron Microscopy*, Fig. 7.8, McGraw-Hill.

Figure 5.4B: Courtesy of J.I. Goldstein, from Goldstein, J.I., *et al.* (1992) *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd edition, Fig. 2.7, reproduced by permission of Plenum Press.

Figure 5.5: Courtesy of D.W. Ackland.

Figure 5.6: Courtesy of D.W. Ackland.

Figure 5.7A: Modified from Crewe, A.V., *et al.* (1969) *Rev. Sci. Instrum.* **40**, 241, Fig. 2.

Figure 5.7B: Courtesy of D.W. Ackland.

Figure 5.10: Courtesy of J.R. Michael, from Michael, J.R. and Williams, D.B. (1987) *J. Microsc.* **147**, 289, Fig. 3, reproduced by permission of the Royal Microscopical Society.

Figure 5.11A: Modified from Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 2.12b, Philips Electron Optics.

Figure 5.12: Courtesy of J.R. Michael, from Michael, J.R. and Williams, D.B. (1987) *J. Microsc.* **147**, 289, Fig. 2, reproduced by permission of the Royal Microscopical Society.

Figure 5.13A: Courtesy of D.W. Ackland.

Figure 5.13B: Courtesy of NSA Hitachi Scientific Instruments Ltd.

Chapter 6

Figure 6.7: Courtesy of D.W. Ackland.

Figure 6.8A: Courtesy of Philips Electronic Instruments Inc.

Figure 6.8B: Courtesy of Kratos Ltd.

Figure 6.8C: From Mulvey, T. (1974) Electron Microscopy—1974, p. 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 A.O. Benscoter.

Figure 6.11: Modified from Reimer, L. (1993) *Transmission Electron Microscopy*, 3rd edition, Fig. 2.13, Springer Verlag.

Figure 6.14: 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, D.B: (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 1.2, Philips Electron Optics.

Figure 7.6: From Berger, S.D., *et al.* (1985) *Electron Microscopy and Analysis*, p. 137, Fig.1, reproduced by permission of The Institute of Physics Publishing.

Chapter 8

Figure 8.1: Courtesy of W.C. Bigelow, from Bigelow, W.C. (1994) *Vacuum Methods in Electron Microscopy*, Fig. 4.1, reproduced by permission of Portland Press, Ltd.

Figure 8.2: Courtesy of W.C. Bigelow, from Bigelow, W.C. (1994) *Vacuum Methods in Electron Microscopy*, Fig. 5.1, reproduced by permission of Portland Press, Ltd.

Figure 8.3: Courtesy of Leybold Vacuum Products Inc.

Figure 8.4: Courtesy of W.C. Bigelow, from Bigelow, W.C. (1994) *Vacuum Methods in Electron Microscopy*, Fig. 7.1, reproduced by permission of Portland Press, Ltd.

Figure 8.6: Courtesy of Gatan Inc.

Figure 8.7: From Valdrè, U., and Goringe, M.J. (1971) in Valdrè, U. (Ed.) *Electron Microscopy in Materials Science*, p. 217, Fig. 6, reproduced by permission of Academic Press Inc.

Figure 8.8: Courtesy of NSA Hitachi Scientific Instruments Ltd.

Figure 8.9A,B: Courtesy of Gatan Inc.

Figure 8.10A,B: Courtesy of Gatan Inc.

Figure 8.11: Courtesy of Gatan Inc.

Figure 8.12: From Komatsu, M., *et al.* (1994) *Journal of the American Ceramic Society* **77**, 839, Fig.1, reproduced by permission of The American Ceramic Society.

Figure 8.13: Courtesy of NSA Hitachi Scientific Instruments Ltd.

Chapter 9

Figure 9.6: Modified from Reimer, L. (1993) *Transmission Electron Microscopy*, 3rd edition, Fig. 4.14a, Springer Verlag.

Figure 9.16: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 1.7, reproduced by permission of Philips Electron Optics.

Figure 9.18B,C,D: Courtesy of D.W.Ackland.

Figure 9.19: From Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Fig. 1.5, reproduced by permission of Philips Electron Optics.

Figure 9.20: Courtesy of S. Ramamurthy.
Figure 9.21: Courtesy of D.W. Ackland.
Figure 9.23: Courtesy of D.W. Ackland.
Figure 9.24: Courtesy of D.W. Ackland.
Figure 9.25: Courtesy of S. Ramamurthy.
Table 9.1: From Williams, D.B. (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, D.B. (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 10.2: Courtesy of SPI Inc.
Figure 10.3: Courtesy of Gatan Inc.
Figure 10.4: Courtesy of VCR Inc.
Figure 10.7A: From Thompson-Russell, K.C. and Edington, J.W. (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 9, reproduced by permission of Philips Electron Optics.
Fig. 10.7b: From Thompson-Russell, K.C. and Edington, J.W. (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 7, reproduced by permission of Philips Electron Optics.
Figure 10.8: From Thompson-Russell, K.C. and Edington, J.W. (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 12, reproduced by permission of Philips Electron Optics.
Figure 10.9: Modified from Thompson-Russell, K.C. and Edington, J.W. (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 11, Philips Electron Optics.
Figure 10.10: Courtesy of R. Alani, Gatan Inc.
Figure 10.11: Courtesy of A.G. Cullis, from Cullis, A.G., *et al.* (1985) Ultramicroscopy **17**, 203, Figs. 1a, 3, reproduced by permission of Elsevier Science B.V.
Figure 10.12: From van Hellemont, J., *et al.* (1988) in Bravman, J., *et al.* (Eds.) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Mat. Res. Soc. Symp. **115**, 247, Fig. 1, reproduced by permission of MRS.
Figure 10.15: From Thompson-Russell, K.C. and Edington, J.W. (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Figs. 20, 21, reproduced by permission of Philips Electron Optics.
Figure 10.16: From Thompson-Russell, K.C. and Edington, J.W. (1977) *Electron Microscope Specimen Preparation Techniques in Materials Science*, Fig. 25, reproduced by permission of Philips Electron Optics.
Figure 10.17: From Hetherington, C.J.D. (1988) in Bravman, J., *et al.* (Eds.) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Mat. Res. Soc. Symp. **115**, 143, Fig. 1, reproduced by permission of MRS.
Figure 10.18: Modified from Dobisz, E.A., *et al.* (1986) *J. Vac. Sci. Technol. B* **4**, 850, Fig. 1, reproduced by permission of MRS.
Figure 10.19: From Fernandez, A. (1988) in Bravman, J., *et al.* (Eds.) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Mat. Res. Soc. Symp. **115**, 119, Fig. 1, reproduced by permission of MRS.
Figure 10.20: Courtesy of P. Goodhew, from Goodhew, P.J. (1988) in Bravman, J., *et al.* (Eds.) *Specimen Preparation for Transmission Electron Microscopy of Materials*, Mat. Res. Soc. Symp. **115**, 52, reproduced by permission of MRS.
Table 10.1: Courtesy of T. Malis.

Chapter 11

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

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 M.L. Jenkins, from Jenkins, M.L., *et al.* (1976) *Phil. Mag.* **34**, 1141, Fig. 2, reproduced by permission of Taylor and Francis.
Figure 16.6: Courtesy of B.C. 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 B.C. De Cooman.
Figure 16.9: From Gajdardziska-Josifovska, M., *et al.* (1995) *Ultramicroscopy* **58**, 65, Fig. 1, reproduced by permission of Elsevier Science B.V.

Figure 16.10: Courtesy of S. McKernan.

Figure 16.11: From Hahn, T. (Ed.) *International Tables for Crystallography A*, pp. 538–539, No. 164, reproduced by permission of The International Union of Crystallography.

Table 16.1: Modified from Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Appendix 8, Van Nostrand Reinhold.

Chapter 17

Figure 17.2: From Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Fig. 2.16, reproduced by permission of Philips Electron Optics.

Figure 17.9: Modified from Hirsch, P.B., *et al.* (1977) *Electron Microscopy of Thin Crystals*, 2nd edition, Fig. 4.11, Krieger.

Figure 17.7: From Carter, C.B., *et al.* (1981) *Phil. Mag.* **A43**, 441, Fig. 5c, reproduced by permission of Taylor and Francis.

Figure 17.10: From Driver, J.H., *et al.* (1972) *Phil Mag.* **26**, 1227, Fig. 3, reproduced by permission of Taylor and Francis.

Figure 17.11A–C: From Lewis, M.H., and Billingham, J. (1972) *JEOL News* **10e**(1), 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, C.B., *et al.* (1981) *Phil. Mag.* **A43**, 441, Figs. 5a, b.

Figure 17.13: Modified from Carter, C.B., *et al.* (1980) *J. Electron Microsc.* **63**, 623, Fig. 8.

Figure 17.14: Modified from Carter, C.B. (1984) *Phil. Mag.* **A50**, 133, Figs. 1–3.

Chapter 18

Figure 18.2: From Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Fig. A1.7, 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,E: From Vainshtein, B.K., *et al.* (1992) in Cowley, J.M. (Ed.) *Electron Diffraction Techniques* **1**, Fig. 6.13, reproduced by permission of Oxford University Press.

Figure 18.11: Modified from James, R.W. (1965) in Bragg, L. (Ed.) *The Optical Principles of the Diffraction of X-ray*, Vol. II of the Crystalline State, Figs. 170, 184, Cornell University Press.

Figure 18.12: Courtesy of D.J.H. Cockayne, from Sproul, A., *et al.* (1986) *Phil. Mag.* **B54**, 113, Fig. 1, reproduced by permisson of Taylor and Francis.

Figure 18.13: From Graczyk, J.F. and Chaudhari, P. (1973) *Phys. stat. sol.* (b), **58**, 163, Fig. 10a, reproduced by permission of Akademie Verlag GmbH.

Figure 18.14: Courtesy of A. Howie, from Howie, A. (1988) in Buseck, P.R., *et al.* (Eds.) *High-Resolution Transmission Microscopy and Associated Techniques*, p. 607, Fig. 14.12, reproduced by permission of Oxford University Press.

Figure 18.15: From Tietz, L.A., *et al.* (1995) *Ultramicroscopy* **60**, 241, Figs. 2, 3, 4, reproduced by permision of Elsevier Science B.V.

Figure 18.16: From Tietz, L.A., *et al.* (1995) *Ultramicroscopy* **60**, 241, Fig. 5, reproduced by permision of Elsevier Science B.V.

Figure 18.17: From Andrews, K.W., *et al.* (1971) *Interpretation of Electron Diffraction Patterns*, 2nd edition, Fig. 41, reproduced by permission of Plenum Press.

Figure 18.18: From Andrews, K.W., *et al.* (1971) *Interpretation of Electron Diffraction Patterns*, 2nd edition, Fig. 41, reproduced by permission of Plenum Press.

Figure 18.19: From Andrews, K.W., *et al.* (1971) *Interpretation of Electron Diffraction Patterns*, 2nd edition, Fig. 41, reproduced by permission of Plenum Press.

Figure 18.20: From Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Fig. 2.20, reproduced by permission of Philips Electron Optics.

Chapter 19

Figure 19.6A: Courtesy of G. Thomas, from Levine, E., *et al.* (1966) *J. Appl. Phys.* **37**, 2141, Fig. 1a, reproduced by permission of the American Institute of Physics.

Figure 19.7: Modified from Okamoto, P.R., *et al.* (1967) *J. Appl. Phys.* **38**, 289, Fig. 5.

Figure 19.8: Courtesy of S. Ramamurthy.

Figure 19.9A: Modified from Thomas, G. and Goringe, M.J. (1979) *Transmission Electron Microscopy of Metals*, Fig. 2.30, John Wiley & Sons Inc.

Figure 19.9B: Modified from Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Fig. 2.27, Van Nostrand Reinhold.

Figure 19.11: Modified from Thomas, G. and Goringe, M.J. (1979) *Transmission Electron Microscopy of Metals*, Fig. 2.29, John Wiley & Sons Inc.

Chapter 20

Figure 20.2A: Courtesy of K.S. Vecchio, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.5, reproduced by permission of Oxford University Press.

Figure 20.2B: Courtesy of K.S. Vecchio, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.17, reproduced by permission of Oxford University Press.

Figure 20.3: Modified from Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 6.6, Philips Electron Optics.

Figure 20.5: From Lyman, C.E., *et al.* (Eds.) (1990) *Scanning Electron Microscopy, X-ray Microanalysis and Analytical Electron Microscopy—a Laboratory Workbook*, Fig. A27.2, reproduced by permission of Plenum Press.

Figure 20.6: Courtesy of J.F. Mansfield, from Mansfield, J.F. (1984) *Convergent Beam Diffraction of Alloy Phases*, Fig. 5.3, reproduced by permission of Institute of Physics Publishing.

Figure 20.7C: Courtesy of R. Ayer.

Figure 20.8A: Modified from Ayer, R. (1989) *J. Electron Microscopy Tech.* **13**, 3, Fig. 3.

Figure 20.9: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 6.13, reproduced by permission of Philips Electron Optics.

Figure 20.10: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 6.14, reproduced by permission of Philips Electron Optics.

Figure 20.11: Courtesy of W.A.T. Clark, from Heilman, P., *et al.* (1983) *Acta Met.* **31**, 1293, Fig. 4, reproduced by permision of Elsevier Science B.V.

Figure 20.12: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 6.9, reproduced by permission of Philips Electron Optics.

Figure 20.13: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 6.16, reproduced by permission of Philips Electron Optics.

Figure 20.14: Courtesy of C.M. Sung.

Figure 20.15: Courtesy of B. Ralph, from Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 6.18, reproduced by permission of Philips Electron Optics.

Figure 20.16: Courtesy of K.S. Vecchio, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.14, reproduced by permission of Oxford University Press.

Chapter 21

Figure 21.1: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 4.29a, reproduced by permission of Philips Electron Optics.

Figure 21.3: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Figs. 4.29b, c, reproduced by permission of Philips Electron Optics.

Figure 21.4: Courtesy of R. Ayer, from Raghavan, M., *et al.* (1984) *Met. Trans.* **15A**, 783, Fig. 6, reproduced by permission of ASM International.

Figure 21.5A: Courtesy of K.S. Vecchio, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.23, reproduced by permission of Oxford University Press.

Figure 21.5B: Courtesy of R. Ayer, from Ayer, R. (1989) *J. Electron Microsc. Tech.* **13**, 3, Fig. 7, reproduced by permission of John Wiley & Sons Inc.

Figure 21.7: Courtesy of K.S. Vecchio, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.19, reproduced by permission of Oxford University Press.

Figure 21.8: Courtesy of V.P. Dravid.

Figure 21.9A–D: Courtesy of J.W. Steeds, from Chapman J.N. and Craven, A.J. (Eds.) (1984) *Quantitative Electron Microscopy*, p. 68, Fig. 8, reproduced by permission of the Scottish Universities Summer School in Physics.

Figure 21.9E: Courtesy of K.S. Vecchio, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.25, reproduced by permission of Oxford University Press.

Figure 21.10: Courtesy of K.S. Vecchio, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.24, reproduced by permission of Oxford University Press.

Figure 21.11: Courtesy of K.S. Vecchio and V.P. Dravid, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Figs. 6.27, 6.28, reproduced by permission of Oxford University Press.

Figure 21.12: Courtesy of K.S. Vecchio, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.29, reproduced by permission of Oxford University Press.

Figure 21.13: Courtesy of K.S. Vecchio and V.P. Dravid, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.30, reproduced by permission of Oxford University Press.

Figure 21.14: Courtesy of R. McConville, from Williams, D.B., *et al.* (Eds.) (1992) *Images of Materials*, Fig. 6.33, reproduced by permission of Oxford University Press.

Figure 21.15: Courtesy of J.M. Cowley, from Liu, M. and Cowley, J.M. (1994) *Ultramicroscopy* **53**, 333, Figs. 1, 2, reproduced by permission of Elsevier Science B.V.

Figure 21.16: Courtesy of W.D. Riecke, from Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 6.2, reproduced by permission of Philips Electron Optics.

Figure 21.17: Courtesy of J.R. Michael.

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

Table 21.3: Courtesy of B.F. Buxton, from Buxton, B.F., *et al.* (1976) *Phil. Trans. Roy. Soc.* (London) **281**, 181, Table 2, reproduced by permission of The Royal Society.

Table 21.4: Courtesy of B.F. Buxton, from Buxton, B.F., *et al.* (1976) *Phil. Trans. Roy. Soc.* (London) **281**, 181, Table 3, reproduced by permission of The Royal Society.

Table 21.6: Courtesy of B.F. Buxton, from Buxton, B.F., *et al.* (1976) *Phil. Trans. Roy. Soc.* (London) **281**, 181, Table 4, reproduced by permission of The Royal Society.

Table 21.7: Courtesy of J.W. Steeds, from Steeds, J.W., *et al.* (1983) *J. Appl. Cryst.* **16**, 317, Tables 5, 6, reproduced by permission of the International Union of Crystallography.

Chapter 22

Figure 22.5: Courtesy of K.A. Repa.

Figure 22.6: From Williams, D.B. (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 K.B. Reuter.

Figure 22.8: From Williams, D.B. (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 H. Tsai, from Williams, D.B. (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, D.B. (1983) in Krakow, W., *et al.* (Eds.) *Electron Microscopy of Materials*, Mat. Res. Soc. Symp. **31**, 11, Figs. 3a, b, MRS.

Figure 22.11A,B: Courtesy of I.M. Watt, from Watt, I.M. (1996) *The Principles and Practice of Electron Microscopy*, 2nd edition, Figs. 5.5a, b, reproduced by permission of Cambridge University Press.

Figure 22.12: Courtesy of M.M.J. Treacy, from Williams, D.B. (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 S.J. Pennycook, from Pennycook, S.J., *et al.* (1986) *J. Microsc.* **144**, 229, Fig. 8, reproduced by permission of the Royal Microscopical Society.

Figure 22.15A,B: Courtesy of S.J. Pennycook, from Lyman, C.E. (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 S.J. Pennycook, from Browning, *et al.* (1995) *Interface Science* **2**, 397, Fig. 4d, reproduced by permission of Kluwer.

Figure 22.16A: From Edington, J.W. (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 23.1: Courtesy of S. Ramamurthy.

Figure 23.2: From Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Fig. 3.2a, reproduced by permission of Philips Electron Optics.

Figure 23.3B: Courtesy of D. Cohen.

Figure 23.3C: Courtesy of S. King.

Figure 23.5: Courtesy of D. Susnitzky.

Figure 23.7: Modified from Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Fig. 3.3, Van Nostrand Reinhold.

Figure 23.8: Modified from Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Figs. 3.4b, d, Van Nostrand Reinhold.

Figure 23.9: Courtesy of S. Ramamurthy.

Figure 23.10: From Hashimoto H., *et al.* (1962) *Proc. Roy. Soc.* (London) **A269**, 80, Fig. 11, reproduced by permission of The Royal Society.

Figure 23.11A: Courtesy of NSA Hitachi Scientific Instruments Ltd.

Figure 23.11B,C: Courtesy of D. Cohen.

Figure 23.12: From Edington, J.W. (1976) *Practical Electron Microscopy in Materials Science*, Fig. 3.3d, reproduced by permission of Philips Electron Optics.

Figure 23.13B,C: From De Cooman, B.C., *et al.* (1987) in J.D. Dow and I.K. Schuller (Eds.) *Interfaces, Superlattices, and Thin Films*, Mat. Res. Soc. Symp. **77**, 187, Fig. 1, reproduced by permission of MRS.

Chapter 24

Figure 24.4A–D: Courtesy of D. Cohen.

Figure 24.4 E,F: Modified from Gevers, R., *et al.* (1963) *Phys. stat. sol.* **3**, 1563, Table 3.

Figure 24.5: From Föll, H., *et al.* (1980) *Phys. stat. sol.* (a) **58**, 393, Figs. 6a, c, reproduced by permission of Akademie Verlag GmbH.

Figure 24.7A,B: From Lewis, M.H. (1966) *Phil. Mag.* **14**, 1003, Fig. 9, reproduced by permission of Taylor and Francis.

Figure 24.7C,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**, p. 107, Figs. 3, 18, North-Holland.

Figure 24.8: From Rasmussen, D.R., *et al.* (1991) *Phys. Rev. Lett.* **66**, (20), 262, Fig. 2, reproduced by permission of The American Physical Society.

Figure 24.9: Courtesy of S. Summerfelt.

Figure 24.13: Modified from Metherell, A.J.F. (1975) in Valdrè, U. and Ruedl, E. (Eds.) *Electron Microscopy in Materials Science* **II**, 397, Fig. 13, Commission of the European Communities.

Figure 24.14: From Hashimoto, H., *et al.* (1962) *Proc. Roy. Soc.* (London) **A269**, 80, Fig. 15, reproduced by permission of The Royal Society.

Figure 24.16: Modified from Rasmussen, R., *et al.* (1991) *Phil. Mag.* **63**, 1299, Fig. 4.

Chapter 25

Figure 25.2B: Modified from Amelinckx, S. (1964) *Solid State Physics Suppl.* **6**, Fig. 76.

Figure 25.6A–C: Modified from Carter, C.B. (1980) *Phys. stat. sol.* (a) **62**, 139, Fig. 4.

Figure 25.6f: From Van Landuyt, J., *et al.* (1970) *Phys. stat. sol.* **41**, 271, Fig. 19, reproduced by permission of Akademie Verlag GmbH.

Figure 25.6G–H: Courtesy of B.C. De Cooman.

Figure 25.7: Modified from Hirsch, P.B., *et al.* (1977) *Electron Microscopy of Thin Crystals*, 2nd edition, Fig. 7.8, Krieger.

Figure 25.8: From Delavignette, P. and Amelinckx, S. (1962) *J. Nucl. Mat.* **5**, 17, Fig. 7, reproduced by permission of Elsevier Science B.V.

Figure 25.10: From Urban, K. (1971) in Koda, S. (Ed.) *The World Through the Electron Microscope*, Metallurgy **V**, p. 26, reproduced by permission of JEOL USA Inc.

Figure 25.11: Courtesy of A. Howie, from Howie, A. and Whelan, M.J. (1962) *Proc. Roy. Soc.* (London) **A267**, 206, Fig. 14, reproduced by permission of The Royal Society.

Figure 25.12: Modified from M. Wilkens (1978) in Amelinckx, S., *et al.* (Eds.) *Diffraction and Imaging Techniques in Material Science* **I**, p. 185, Fig. 4, North-Holland.

Figure 25.14: From Dupouy G. and Perrier, F. (1971) in Koda, S. (Ed.) *The World Through the Electron Microscope*, Metallurgy **V**, p. 100, reproduced by permission of JEOL USA Inc.

Figure 25.15A: From Modeer, B. and Lagneborg, R. (1971) in Koda, S. (Ed.) *The World Through the Electron Microscope*, Metallurgy **V**, p. 44, reproduced by permission of JEOL USA Inc.

Figure 25.15B: Courtesy of D.A. Hughes, from Hansen, N. and Hughes, D.A. (1995) *Phys. stat. sol.* (a) **149**, 155, Fig. 5, reproduced by permission of Akademie Verlag GmbH.

Figure 25.16A: From Siems, F., *et al.* (1962) *Phys. stat. sol.* **2**, 421, Fig. 5a, reproduced by permission of Akademie Verlag GmbH.

Figure 25.16C: From Siems, F., *et al.* (1962) *Phys. stat. sol.* **2**, 421, Fig. 15a, reproduced by permission of Akademie Verlag GmbH.

Figure 25.17A: Modified from Whelan, M.J. (1958–59) *J. Inst. Met.* **87**, 392, Fig. 25a.

Figure 25.17B: Courtesy of K. Ostyn.

Figure 25.18: From Takayanagi, L. (1988) *Surface Science* **205**, 637, Fig. 5, reproduced by permission of Elsevier Science B.V.

Figure 25.19A: From Tunstall, W.J., *et al.* (1964) *Phil. Mag.* **9**, 99, Fig. 9, reproduced by permission of Taylor and Francis.

Figure 25.19B: From Amelinckx, S. in Merli, P.G. and Antisari, V.M. (Eds.) *Electron Microscopy in Materials Science*, p. 128, Fig. 45, reproduced by permission of World Scientific.

Figure 25.20: Courtesy of W. Skrotski.

Figure 25.21: Courtesy of W. Skrotski.

Figure 25.22: From Carter, C.B., *et al.* (1986) *Phil. Mag.* **A55**, 21, Fig. 2, reproduced by permission of Taylor and Francis.

Figure 25.23: From Carter, C.B., *et al.* (1981) *Phil. Mag.* **A43**, 441, Fig. 3, reproduced by permission of Taylor and Francis.

Figure 25.24: Courtesy of K. Ostyn.

Figure 25.25: Courtesy of L. Tietz.

Figure 25.26A: Courtesy of L.M. Brown, from Ashby, M.F. and Brown, L.M. (1963) *Phil. Mag.* **8**, 1083, Fig. 10, reproduced by permission of Taylor and Francis.

Figure 25.26B: Modified from Whelan, M.J. (1978) in Amelinckx, S., *et al.* (Eds.) *Diffraction and Imaging Techniques in Material Science* **I**, p. 43, Fig. 36, North-Holland.

Figure 25.26C: Courtesy of L.M. Brown, from Ashby, M.F. and Brown, L.M. (1963) *Phil. Mag.* **8**, 1083, Fig. 12, reproduced by permission of Taylor and Francis.

Figure 25.27: From Rasmussen, D.R. and Carter, C.B. (1991) *J. Electron Microsc. Technique* **18**, 429, Fig. 2, reproduced by permission of John Wiley & Sons Inc.

Figure 25.27: From Rasmussen, D.R. and Carter, C.B. (1991) *J. Electron Microsc. Technique* **18**, 429, Fig. 2, reproduced by permission of John Wiley & Sons Inc.

Chapter 26

Figure 26.7: Courtesy of S. King.

Figure 26.10: Courtesy of D.J.H. Cockayne, from Cockayne,

D.J.H. (1972) *Z. Naturforschung* **27a**, 452, Fig. 6c, reproduced by permission of Verlag der Zeitschrift für Naturforschung, Tübingen.

Figure 26.13: Modified from Carter, C.B., *et al.* (1986) *Phil. Mag.* **A55**, 1, Fig. 9.

Figure 26.15: Modified from Föll, H., *et al.* (1980) *Phys. stat. sol.* (a) **58**, 393, Figs. 6b, c.

Figure 26.17: From Heidenreich, R.D. (1964) *Fundamentals of Transmission Electron Microscopy*, Fig. 9.20, reproduced by permission of John Wiley & Sons Inc.

Figure 26.18: Courtesy of D.J.H. Cockayne, from Ray, I.L.F. and Cockayne, D.J.H. (1971) *Proc. Roy. Soc.* (London) **A325**, 543, Fig. 10, reproduced by permission of The Royal Society.

Figure 26.23: Modified from Carter, C.B. (1979) J. Phys. (A) **54** (1) 395 Fig. 8a.

Chapter 27

Figure 27.3A: From Izui, K.J., *et al.* (1977) *J. Electron Microsc.* **26**, 129, Fig. 1. reproduced by permission of the Japanese Society of Electron Microscopy.

Figure 27.3C: Courtesy of J.C.H. Spence, from Spence, J.C.H. *Experimental High-Resolution Electron Microscopy*, Fig. 5.15, reproduced by permission of Oxford University Press.

Figure 27.4B: Courtesy of J.L. Hutchison, from Hutchison, J.L., *et al.* (1991) in Heydenreich, J. and Neumann, W. (Eds.) *High-Resolution Electron Microscopy—Fundamentals and Applications*, p. 205, Fig. 3, reproduced by permission of Halle/Saale.

Figure 27.4C: Courtesy of S. McKernan.

Figure 27.4D: From Carter, C.B., *et al.* (1989) *Phil. Mag.* **A63**, 279, Fig. 3, reproduced by permission of Taylor and Francis.

Figure 27.8: From Tietz, L.A., *et al.* (1992) *Phil. Mag.* **A65**, 439, Figs. 3a, 12a, c, reproduced by permission of Taylor and Francis.

Figure 27.10: Courtesy of J. Zhu.

Figure 27.12: Modified from Vincent, R. (1969) *Phil. Mag.* **19**, 1127, Fig. 4.

Figure 27.13: Modified from Norton, M.G. and Carter, C.B. (1995) *J. Mat. Sci.* **30**, Fig. 6.

Figure 27.14: Courtesy of U. Dahmen, from Hetherington, C.J.D. and Dahmen, U. (1992) in Hawkes, P.W. (Ed.) Signal and Image Processing in Microscopy and Microanalysis, *Scanning Microscopy* Supplement **6**, 405, Fig. 9, reproduced by permission of Scanning Microscopy International.

Figure 27.15: From Heidenreich, R.D. (1964) *Fundamentals of Transmission Electron Microscopy*, Figs. 5.4, 5.6, reproduced by permission of John Wiley & Sons Inc.

Figure 27.16A: From Heidenreich, R.D. (1964) *Fundamentals of Transmission Electron Microscopy*, Fig. 11.2, reproduced by permission of John Wiley & Sons Inc.

Figure 27.16B: From Boersch, H., *et al.* (1962) *Z. Phys.* **167**, 72, Fig. 4, reproduced by permission of Springer-Verlag.

Figure 27.17: Courtesy of M. Rühle.

Figure 27.18: Modified from Kouh, Y.M., *et al.* (1986) *J. Mat. Sci.* **21**, 2689, Fig. 9.

Figure 27.19: Courtesy of M. Rühle, from Rühle, M. and Sass, S.L. (1984) *Phil. Mag.* **A49**, 759, Fig. 2, reproduced by permission of Taylor and Francis.

Figure 27.20B–E: From Carter, C.B., *et al.* (1986), *Phil. Mag.* **A55**, 21, Fig. 11, reproduced by permission of Taylor and Francis.

Chapter 28

Figure 28.4: Courtesy of R. Gronsky, from Gronsky, R. (1992) in Williams, D.B., *et al.* (Eds.) *Images of Materials*, Fig. 7.6, reproduced 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, J.M. (1988) in Buseck, P.R., *et al.* (Eds.) *High-Resolution Electron Microscopy and Associated Techniques*, Fig. 1.9, Oxford University Press.

Figure 28.8: Courtesy of J.C.H. Spence, from Spence, J.C.H. (1988) *Experimental High-Resolution Electron Microscopy*, 2nd Ed., Fig. 4.3, reproduced by permission of Oxford University Press.

Figure 28.9: Modified from Rose, H. (1991) in Heydenreich, J. and Neumann, W. (Eds.) *High-Resolution Electron Microscopy—Fundamentals and Applications*, p. 6, Fig. 3, Halle/Saale.

Figure 28.10: From de Jong, A.F. and Van Dyck, D. (1993) *Ultramicroscopy* **49**, 66, Fig. 1, reproduced by permission of Elsevier Science B.V.

Figure 28.11: Courtesy of M.T. Otten, from Otten, M.T. and Coene, W.M.J. (1993) *Ultramicroscopy* **48**, 77, Fig. 8, reproduced by permission of Elsevier Science B.V.

Figure 28.12: Courtesy of M.T. Otten, from Otten, M.T. and Coene, W.M.J. (1993) *Ultramicroscopy* **48**, 77, Fig. 11, reproduced by permission of Elsevier Science B.V.

Figure 28.13: Courtesy of M.T. Otten, from Otten, M.T. and Coene, W.M.J. (1993) *Ultramicroscopy* **48**, 77, Fig. 10, reproduced by permission of Elsevier Science B.V.

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

Figure 28.15: From Amelinckx, S., *et al.* (1993) *Ultramicros-*

copy **51**, 90, Fig. 3, reproduced by permission of Elsevier Science B.V.

Figure 28.16: From Rasmussen, D.R., *et al.* (1995) *J. Microsc.* **179**, 77, Figs. 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) Ultramicroscopy **55**, 101, Fig. 4b, reproduced by permission of Elsevier Science B.V.

Figure 28.18C: Courtesy of S. Summerfelt.

Figure 28.18D: Courtesy of S. McKernan.

Figure 28.19: Courtesy of D.J. Smith.

Figure 28.21B: From Van Landuyt, J., *et al.* (1991) in Heydenreich, J. and Neumann, W. (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 Heydenreich, J. and Neumann, W. (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 Heydenreich, J. and Neumann, W. (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 Heydenreich, J. and Neumann, W. (Eds.) *High-Resolution Electron Microscopy—Fundamentals and Applications*, p. 272, Fig. 2, reproduced by permission of Halle/Saale.

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

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

Chapter 29

Figure 29.2A,B: Courtesy of M.A. O'Keefe, from O'Keefe, M.A. and Kilaas, R. (1988) in Hawkes, P.W., *et al.* (Eds.) Image and Signal Processing in Electron Microscopy, *Scanning Microscopy* Supplement **2**, p. 225, Fig. 1, reproduced by permission of Scanning Microscopy International.

Figure 29.3: From Kambe, K. (1982) *Ultramicroscopy* **10**, 223, Figs. 1a–d, reproduced by permission of Elsevier Science B.V.

Figure 29.4: Courtesy of M.A. O'Keefe, from O'Keefe, M.A. and Kilaas, R. (1988) in Hawkes, P.W., *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 29.5: Modified from Rasmussen, D.R. and Carter, C.B. (1990) *Ultramicroscopy* **32**, 337, Figs. 1 and 2.

Figure 29.8: From Beeli, C. and Horiuchi, S. (1994) *Phil. Mag.* **B70**, 215, Figs. 6a–d, reproduced by permission of Taylor and Francis.

Figure 29.9: From Beeli, C. and Horiuchi, S. (1994) *Phil. Mag.* **B70**, 215, Figs. 7a–d, reproduced by permission of Taylor and Francis.

Figure 29.10: From Beeli, C. and Horiuchi, S. (1994) *Phil. Mag.* **B70**, 215, Fig. 8, reproduced by permission of Taylor and Francis.

Figure 29.11: From Jiang, J., *et al.* (1995) *Phil. Mag. Lett.* **71**, 123, Fig. 4, reproduced by permission of Taylor and Francis.

Chapter 30

Figure 30.1: Courtesy of J. Heffelfinger.

Figure 30.2: From Rasmussen, D.R., *et al.* (1995) *J. Microsc.* **179**, 77, Fig. 1b, reproduced by permission of The Royal Microscopical Society.

Figure 30.3: From Rasmussen, D.R., *et al.* (1995), *J. Microsc.* **179**, 77, Fig. 5, reproduced by permission of The Royal Microscopical Society.

Figure 30.4: Courtesy of O.L. Krivanek, from Krivanek, O.L. (1988) in Buseck, P.R., *et al.* (Eds.) *High-Resolution Electron Microscopy and Associated Techniques*, Fig. 12.6, reproduced by permission of Oxford University Press.

Figure 30.5: Courtesy of O.L Krivanek, from Krivanek, O.L. (1988) in Buseck, P.R., *et al.* (Eds.) *High-Resolution Electron Microscopy and Associated Techniques*, Fig. 12.7, reproduced by permision of Oxford University Press.

Figure 30.6A: Courtesy of J.C.H. Spence, from Spence, J.C.H. and Zuo, J.M. (1992) *Electron Microdiffraction*, Fig. A1.3, reproduced by permission of Plenum Press.

Figure 30.6B: Courtesy of O.L. Krivanek, from Krivanek, O.L. (1988) in Buseck, P.R., *et al.* (Eds.) *High-Resolution Electron Microscopy and Associated Techniques*, Fig. 12.8, reproduced by permission of Oxford University Press.

Figure 30.7: Courtesy of S. McKernan.

Figure 30.8: Courtesy of Z.C. Lin, from Lin, Z.C. (1993) Ph.D. dissertation, Fig. 4.15, University of Minnesota.

Figure 30.9: Courtesy of O. Saxton, from Kirkland, A.I. (1992) in Hawkes, P.W. (Ed.) Signal and Image Processing in Microscopy and Microanalysis, *Scanning Microscopy* Supplement **6**, 139, Figs. 1, 2, 3, reproduced by permission of Scanning Microscopy International.

Figure 30.10: From Zou, X.D. and Hovmöller, S. (1993) *Ultramicroscopy* **49**, 147, Fig. 1, reproduced by permission of Elsevier Science B.V.

Figure 30.11A: From Kirkland, A.I., *et al.* (1995) *Ultramicroscopy* **57**, 355, Fig. 1, reproduced by permission of Elsevier Science B.V.

Figure 30.11B: From Kirkland, A.I., *et al.* (1995) *Ultramicroscopy* **57**, 355, Fig. 3, reproduced by permission of Elsevier Science B.V.

Figure 30.12A–C: From Kirkland, A.I., *et al.* (1995) *Ultramicroscopy* **57**, 355, Fig. 8, reproduced by permission of Elsevier Science B.V.

Figure 30.13: Courtesy of O.L. Krivanek, from Krivanek, O.L. and Fan, G.Y. (1992) in Hawkes, P.W. (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 30.14: Courtesy of O.L. Krivanek, from Krivanek, O.L.and Fan, G.Y. (1992) in Hawkes, P.W. (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 30.15: Courtesy of U. Dahmen, from Paciornik, S., *et al.* (1996) *Ultramicroscopy* **62**, 15, Fig. 1, reproduced by permission of Elsevier Science B.V.

Figure 30.16: Courtesy of U. Dahmen, from Paciornik, S., *et al.* (1996) *Ultramicroscopy* **62**, 15, Fig. 5, reproduced by permission of Elsevier Science B.V.

Figure 30.17: Courtesy of A. Ourmazd, from Kisielowski, C., *et al.* (1995) *Ultramicroscopy* **58**, 131, Figs. 2–4, reproduced by permission of Elsevier Science B.V.

Figure 30.18: Courtesy of A. Ourmazd, from Kisielowski, C., *et al.* (1995) *Ultramicroscopy* **58**, 131, Figs. 8, 10, 12, reproduced by permission of Elsevier Science B.V.

Figure 30.19A–D: Courtesy of A. Ourmazd, from Ourmazd, A., *et al.* (1990) *Ultramicroscopy* **34**, 237, Figs. 1, 1, 2, 5, reproduced by permission of Elsevier Science B.V.

Figure 30.20A–F: Courtesy of A. Ourmazd, from Kisielowski, C., *et al.* (1995) *Ultramicroscopy* **58**, 131, Figs. 15a–f, reproduced by permission of Elsevier Science B.V.

Figure 30.21: Courtesy of U. Dahmen, from Paciornik, S., *et al.* (1996) *Ultramicroscopy*, in press, Fig. 2, reproduced by permission of Elsevier Science B.V.

Figure 30.22: From King, W.E. and Campbell, G.H. (1994) *Ultramicroscopy* **56**, 46, Fig. 1, reproduced by permission of Elsevier Science B.V.

Figure 30.23: From King, W.E. and Campbell, G.H. (1994) *Ultramicroscopy* **56**, 46, Fig. 6, reproduced by permission of Elsevier Science B.V.

Figure 30.24: Courtesy of M. Rühle, from Möbus, G., *et al.* (1993) *Ultramicroscopy* **49**, 46, Fig. 6, reproduced by permission of Elsevier Science B.V.

Figure 30.25: From Thon, F. (1970) in Valdrè, U. (Ed.) *Electron Microscopy in Materials Science*, p. 571, Fig. 36, reproduced by permission of Academic Press.

Figure 30.26: Courtesy of J. Heffelfinger.

Chapter 31

Figure 31.2: Courtesy of R. Sinclair, from Sinclair, R., *et al.* (1981) *Met. Trans.* **12A**, 1503, Figs. 13, 14, reproduced by permission of ASM International.

Figure 31.4A,B: From Marcinkowksi, M.J. and Poliak, R.M. (1963) *Phil. Mag.* **8**, 1023, Figs. 15a, b, reproduced by permission of Taylor and Francis.

Figure 31.4C,D: Courtesy of J. Silcox, from Silcox, J. (1963) *Phil. Mag.* **8**, 7, Fig. 7, reproduced by permission of Taylor and Francis.

Figure 31.5: Courtesy of A.J. Craven, from Buggy, T.W., *et al.* (1981) *Analytical Electron Microscopy—1981*, p. 231, Fig. 5, reproduced by permission of San Francisco Press.

Figure 31.6D,E: Courtesy of NSA Hitachi Scientific Instruments Ltd. and S. McKernan.

Figure 31.7: Courtesy of R. Sinclair.

Figure 31.8: From Kuesters, K.-H., *et al.* (1985) *J. Cryst. Growth* **71**, 514, Fig. 4, reproduced by permission of Elsevier Science, B.V.

Figure 31.9: Courtesy of of M. Mallamaci.

Figure 31.10A: From De Cooman, B.C., *et al.* (1985) *J. Electron Microsc. Tech.* **2**, 533, Fig. 1, reproduced by permission of John Wiley & Sons Inc.

Figure 31.10B: Courtesy of S.M. Zemyan.

Figure 31.10C–E: Courtesy of B.C. De Cooman.

Figure 31.12: Courtesy of G. Thomas, from Bell, W.L. and Thomas, G. (1972) in Thomas G., *et al.* (Eds.) *Electron Microscopy and Structure of Materials*, p. 53, Fig. 28, reproduced by permission of University of California Press.

Figure 31.13: Courtesy of K.-R. Peters, from Peters, K.-R. (1984) in Kyser, D.F., *et al.* (Eds.) *Electron Beam Interactions with Solids for Microscopy, Microanalysis and Lithography*, p. 363, Fig. 1, reproduced by permission of Scanning Microscopy International.

Figure 31.14: Courtesy of R. McConville, from Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 3.11, reproduced by permission of Philips Electron Optics.

Figure 31.15: Courtesy of Philips Electronic Instruments, from Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig.

3.10, reproduced by permission of Philips Electron Optics.

Figure 31.16: Courtesy of H. Lichte, from Lichte, H. (1992) *Scanning Microscopy*, p. 433, Fig. 1, reproduced by permission of Scanning Microscopy International.

Figure 31.17: Modified from Lichte, H. (1992) *Ultramicroscopy* **47**, 223, Fig. 1.

Figure 31.18: Modified from Tonomura, A.: Courtesy of NSA Hitachi Scientific Instruments Ltd.

Figure 31.19A–C: From Tonomura, A. (1992) *Adv. Phys.* **41**, 59, Fig. 29, reproduced by permission of Taylor and Francis.

Figure 31.19D: From Tonomura, A. (1987) *Rev. Mod. Phys.* **59**, 639, Fig. 41, reproduced by permission of The American Physical Society.

Figure 31.20A: From Tonomura, A. (1992) *Adv. Phys.* **41**, 59, Fig. 38, reproduced by permission of Taylor and Francis.

Figure 31.20B: From Tonomura, A. (1992) *Adv. Phys.* **41**, 59, Fig. 42, reproduced by permission of Taylor and Francis.

Figure 31.20C: From Tonomura, A. (1992) *Adv. Phys.* **41**, 59, Fig. 44, reproduced by permission of Taylor and Francis.

Figure 31.21: Courtesy of R. Sinclair, from Sinclair, R., *et al.* (1994) *Ultramicroscopy* **56**, 225, Fig. 5, reproduced by permission of Elsevier Science B.V.

Chapter 32

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

Figure 32.2: Courtesy of S.M. Zemyan.

Figure 32.3B: Courtesy of JEOL USA Inc.

Figure 32.4: Modified from Woldseth, R. (1973) *X-ray Energy Spectrometry*, Kevex Instruments.

Figures 32.5–8: Courtesy of S.M. Zemyan.

Figure 32.10: Courtesy of S. M. Zemyan, from Zemyan, S. and Williams, D.B. (1995) in Williams, D.B., *et al.* (Eds.) *X-ray Spectrometry in Electron Beam Instruments*, Fig. 12.9, reproduced by permission of Plenum Press.

Figure 32.11: Courtesy of S. M. Zemyan, from Zemyan, S. and Williams, D.B (1995) in Williams, D.B., *et al.* (Eds.) *X-ray Spectrometry in Electron Beam Instruments*, Fig. 12.10, reproduced by permission of Plenum Press.

Figure 32.12A: Courtesy of J.J. Friel, from Mott, R.B. and Friel, J.J. (1995) in Williams, D.B., *et al.* (Eds.) *X-ray Spectrometry in Electron Beam Instruments*, Fig. 9.8, reproduced by permission of Plenum Press.

Figure 32.12B: Courtesy of C.E. Lyman, from Lyman C.E., *et al.* (1994) *J. Microsc.* **176**, 85, Fig. 9, reproduced by permission of the Royal Microscopical Society.

Figure 32.13: Courtesy of S.M. Zemyan.

Figure 32.14: Courtesy of S.M. Zemyan.

Figure 32.15: Courtesy of D.E. Newbury, from Newbury, D.E. (1995) in Williams, D.B., *et al.* (Eds.) *X-ray Spectrometry in Electron Beam Instruments*, Fig. 11.18, reproduced by permission of Plenum Press.

Figure 32.16: Courtesy of J.I. Goldstein, from Goldstein, J.I., *et al.* (1992) *Scanning Electron Microscopy and X-ray Microanalysis*, 2nd edition, Fig. 5.3, reproduced by permission of Plenum Press.

Figure 32.17A,B: Courtesy of S.M. Zemyan.

Chapter 33

Figure 33.1: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 4.5a, reproduced by permission of Philips Electron Optics.

Figure 33.2: Courtesy of W.A.P. Nicholson, from Nicholson, W.A.P., *et al.* (1982), *J. Microsc.* **125**, 25, Fig. 4, reproduced by permission of the Royal Microscopical Society.

Figure 33.3: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 4.30, reproduced by permission of Philips Electron Optics.

Figure 33.4: Courtesy of S.M. Zemyan.

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

Figure 33.6: Modified from Williams D.B. and Goldstein, J.I. (1981) in Heinrich, K.F.J., *et al.* (Eds.) *Energy-Dispersive X-ray Spectrometry*, p. 346, Fig. 7a, NBS.

Figure 33.7: Courtesy of S.M. Zemyan.

Figure 33.8: Courtesy of S.M. Zemyan.

Figure 33.9A: Courtesy of S.M. Zemyan.

Figure 33.9B: Courtesy of K.S. Vecchio, from Vecchio, K.S. and Williams, D.B. (1987) *J. Microsc.* **147**, 15, Fig. 1, reproduced by permission of the Royal Microscopical Society.

Figure 33.10A: Courtesy of S.M. Zemyan.

Figure 33.10B: Courtesy of S.M. Zemyan, from Zemyan, S. and Williams, D.B. (1994) *J. Microsc.* **174**, 1, Fig. 6, reproduced by permission of the Royal Microscopical Society.

Chapter 34

Figures 34.1–5: Courtesy of S.M. Zemyan.

Chapter 35

Figures 35.1–4: Courtesy of S.M. Zemyan.

Figure 35.5A: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 4.20, reproduced by permission of Philips Electron Optics.

Figure 35.5B,C: Courtesy of S.M. Zemyan.

Figure 35.6: Courtesy of S.M. Zemyan.

Figure 35.7: From Wood, J.E., *et al.* (1984) *J. Microsc.* **133**, 255, Figs. 2, 8, reproduced by permission of the Royal Microscopical Society.

Figure 35.8: From Bender, B.A., *et al.* (1980) *J. Amer. Ceram. Soc.* **63**, 149, Fig. 1, reproduced by permission of the American Ceramic Society.

Figure 35.10: Courtesy of S. Vivekenand and K. Barmak.

Figure 35.11A: Courtesy of J.A. Eades, from Christenson, K.K. and Eades, J.A. (1986) *Proc. 44th EMSA Meeting*, p. 622, Fig. 2, reproduced by permission of MSA.

Figure 35.12: Courtesy of R. Ayer, from M. Raghavan, *et al.* (1984) *Met. Trans.* **15A**, 783, Figs. 4, 11, reproduced by permission of ASM International.

Figure 35.13: Courtesy of A.W. Nicholls, from Nicholls, A.W. and Jones, I.P. (1983) *J. Chem. Phys.* **44**, 671, Figs. 3, 6a, reproduced by permission of The American Physical Society.

Figure 35.14A: Courtesy of V.J. Keast.

Figure 35.14B: Courtesy of J.R. Michael, from Michael J.R. and Williams, D.B. (1987) *Met. Trans.* **15A**, Fig. 7, Reproduced by permission of ASM International.

Figure 35.15: Courtesy of C.E. Lyman, from Lyman C.E. (1986) *Ultramicroscopy* **20**, 119, Figs. 1b, c, reproduced by permission of Elsevier Science B.V.

Tables 35.1, 35.2: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Tables 4.2a, b, reproduced by permission of Philips Electron Optics.

Tables 35.3A,B: From Wood, J.E., *et al.* (1984) *J. Microsc.* **133**, 255, Tables 9,11, reproduced by permission of the Royal Microscopical Society.

Chapter 36

Figure 36.1A,B: Courtesy of V.J. Keast.

Figure 36.2: Courtesy of J.R. Michael, from Williams D.B., *et al.* (1992) *Ultramicroscopy* **47**, 121, Fig. 1, reproduced by permission of Elsevier Science B.V.

Figure 36.3: Courtesy of J.R. Michael, from Williams D.B., *et al.* (1992) *Ultramicroscopy* **47**, 121, Fig. 2, reproduced by permission of Elsevier Science B.V.

Figure 36.4: Courtesy of R. Ayer, from Michael, J.R., *et al.* (1989) *J. Microsc.* **160**, 41, Fig. 2, reproduced by permission of the Royal Microscopical Society.

Figure 36.5.: Courtesy of R. Ayer, from Michael, J.R., *et al.* (1989) *J. Microsc.* **160**, 41, Figs. 3, 4, reproduced by permission of the Royal Microscopical Society.

Figure 35.8A,B: From Williams, D.B. (1987) *Practical Analytical Electron Microscopy in Materials Science*, 2nd Edition, Fig. 4.27, reproduced by permission of Philips Electron Optics.

Figure 36.9: Courtesy of C.E. Lyman, modified from Lyman C.E. (1987) in Kirschner, J., *et al.* (Eds.) *Physical Aspects of Microscopic Characterization of Materials*, p. 123, Fig. 1, Scanning Microscopy International.

Figure 36.10: Courtesy of C.E. Lyman, modified from Lyman C.E. (1987) in Kirschner, J., *et al.* (Eds.) *Physical Aspects of Microscopic Characterization of Materials* p. 123, Fig. 7, Scanning Microscopy International.

Chapter 37

Figure 37.1: Courtesy of R.F. Egerton, modified from Egerton, R.F. (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd edition, Fig. 2.2, Plenum Press.

Figure 37.2: Courtesy of Gatan Inc.

Figure 37.3: Courtesy of D.C. Joy, from Joy, D.C. in Joy, D.C., *et al.* (Eds.) (1986) *Principles of Analytical Electron Microscopy*, Fig. 5, p. 259, reproduced by permission of Plenum Press.

Figure 37.4: Courtesy of J. Bruley.

Figure 37.5: Courtesy of Gatan Inc.

Figure 37.6: Courtesy of J.A. Hunt, from Hunt, J.A. and Williams, D.B. (1994) *Acta Microsc.* **3**, 1, Fig. 7, reproduced by permission of the Venezuelan Society for Electron Microscopy.

Figure 37.10: Courtesy of R.F. Egerton, from Egerton, R.F., *et al.* (1993) *Ultramicroscopy* **48**, 239, Fig. 2, reproduced by permission of Elsevier Science B.V.

Figure 37.11: Courtesy of O.L. Krivanek, modified from Krivanek, O.L, *et al.* (1991) *Microsc. Microanal. Microstruct.* **2**, p. 315, Fig. 8.

Chapter 38

Figures 38.1–3: Courtesy of J. Bruley.

Figure 38.4: Courtesy of O.L. Krivanek, modified from Ahn, C.C. and Krivanek, O.L (1983) *EELS Atlas* p. iv, Gatan Inc.

Figure 38.5: Courtesy of J. Bruley, modified from Joy, D.C. (1986) in Joy D.C., *et al.* (Eds.) *Principles of Analytical Electron Microscopy*, p. 249, Fig. 8, Plenum Press.

Figure 38.6: Courtesy of J. Bruley.

Figure 38.7: Courtesy of C.E. Lyman, from Lyman C.E. (1987) in Kirschner, J., *et al.* (Eds.) *Physical Aspects of Microscopic Characterization of Materials*, p. 123, Fig.

2, reproduced by permission of Scanning Microscopy International.

Figure 38.8: Courtesy of D.C. Joy, modified from Joy, D.C. in Hren, J.J., *et al.* (Eds.) (1979) *Introduction to Analytical Electron Microscopy*, p. 235, Fig. 7.6, Plenum Press.

Figure 38.9: Courtesy of M. Kundmann.

Figure 38.10: Courtesy of J.A. Hunt, from Hunt, J.A. and Williams, D.B. (1994) *Acta Microsc.* **3**, 1, Fig. 5, reproduced by permission of the Venezuelan Society for Electron Microscopy.

Figure 38.11: Courtesy of J.A. Hunt, from Hunt, J.A. and Williams, D.B. (1994) *Acta Microsc.* **3**, 1, Fig. 4, reproduced by permission of the Venezuelan Society for Electron Microscopy.

Figure 38.12: Courtesy of J. Bruley, from Hunt, J.A. and Williams, D.B. (1994) *Acta Microsc.* **3**, 1, Fig. 6, reproduced by permission of the Venezuelan Society for Electron Microscopy.

Table 38.2: Courtesy of R.F. Egerton, from Egerton R.F. (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd edition, p. 157, Table 3.2, reproduced by permission of Plenum Press.

Chapter 39

Figure 39.1: Courtesy of J. Bruley.

Figure 39.2: Courtesy of K. Sato and Y. Ishiguro, modified from Sato K. and Ishiguro Y. (1996) *Materials Transactions Japan Institute of Metals* **37**, 643, Figs. 1 and 7.

Figure 39.3: Courtesy of J. Bruley.

Figure 39.4: Courtesy of J.A. Hunt, from Hunt, J.A. and Williams, D.B. (1994) *Acta Microsc.* **3**, 1, Fig. 14, reproduced by permission of the Venezuelan Society for Electron Microscopy.

Figure 39.5: Courtesy of J.A. Hunt, from Williams, D.B. and Goldstein, J.I. (1992) *Microbeam Anlaysis* **1**, 29, Fig. 11, reproduced by permission of VCH.

Figure 39.6: Courtesy of J.A. Hunt, from Hunt, J.A. and Williams, D.B. (1994) *Acta Microsc.* **3**, 1, Fig. 17a, reproduced by permission of the Venezuelan Society for Electron Microscopy.

Figure 39.7: Courtesy of J.A. Hunt, from Hunt, J.A. and Williams, D.B. (1994) *Acta Microsc.* **3**, 1, Fig. 17b, reproduced by permission of the Venezuelan Society for Electron Microscopy.

Figure 39.8: Courtesy of D.C. Joy, from Joy, D.C. in Joy, D.C., *et al.* (Eds.) (1986) *Principles of Analytical Electron Microscopy*, p. 288, Figs. 6,7, reproduced by permission of Plenum Press.

Figure 39.9: Courtesy of R.F. Egerton, modified from Egerton R.F. (1993) *Ultramicroscopy* **50**, Fig. 6.

Figure 39.10: From Liu D.R. and Williams, D.B. (1989) *Proc. Roy. Soc. (London)* **A425**, Fig. 7, reproduced by permission of The Royal Society.

Figure 39.11: Courtesy of J. Bruley.

Figure 39.12: Courtesy of D.C. Joy, from Joy, D.C. in Joy, D.C., *et al.* (Eds.) (1986) *Principles of Analytical Electron Microscopy*, p. 293, Fig. 10, reproduced by permission of Plenum Press.

Figures 39.13, 14, 15: Courtesy of J. Bruley.

Figure 39.16: Courtesy of J.A. Hunt, from Hunt, J.A. and Williams, D.B. (1994) *Acta Microsc.* **3**, 1, Fig. 16, reproduced by permission of the Venezuelan Society for Electron Microscopy.

Figure 39.17: Courtesy of R.F. Egerton, from Egerton, R.F. (1996) *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd edition, Fig. 1.11, reproduced by permission of Plenum Press.

Figure 39.18: Courtesy of O.L. Krivanek, modified from Krivanek, O.L., *et al.* (1991), *Microsc. Microanal. Microstruc.* **2**, 257, Fig. 5.

Chapter 40

Figure 40.4: Courtesy of N.J. Zaluzec, modified from Zaluzec, N.J. (1982) *Ultramicroscopy* **9**, 319, Fig. 3.

Figure 40.5A,B: Courtesy of J. Bruley.

Figure 40.5C: Courtesy of P.E. Batson, from Batson, P.E. (1993) *Nature* **366**, 727, Fig. 1, reproduced by permission of Macmillan Journals Ltd.

Figure 40.7: Courtesy of R. Brydson, modified from Hansen, P.L., *et al.* (1994) *Microsc. Microanal. Microstruc.* **5**, 173, Figs. 1, 2.

Figure 40.8: Courtesy of J. Bruley.

Figure 40.9: From Williams D.B. and Edington, J.W. (1976) *Acta Met.* **24**, 323, Fig. 7, reproduced by permission of Elsevier Science B.V.

Figure 40.9C: Courtesy of A.J. Strutt.

Figure 40.10A,B: Courtesy of J. Bruley.

Figure 40.10C: Courtesy of J.A. Hunt from Hunt J.A. and Williams, D.B. (1991) *Ultramicroscopy* **38**, 47, Fig. 11, reproduced by permission of Elsevier Science B.V.

Figure 40.11: Courtesy of J. Bruley, from Müllejans H., *et al.* (1993) *Electron Microscopy and Analysis—1993*, p. 62, Fig. 4, reproduced by permission of the Institute of Physics Publishing.

Figures 40.12, 13: Courtesy of J.A. Hunt.

Figure 40.14: Courtesy of J. Bruley and J. Mayer.

Figure 40.15: Courtesy of J.C.H. Spence and J. Mayer, from Mayer, *et al.* (1991) *Proc. 49th EMSA Meeting*, p. 787, Figs. 2, 3, San Francisco Press, reproduced by permission of MSA.

Table 40.1: From Williams D.B. and Edington, J.W. (1976) *J. Microsc.* **108**, 113, Table 1, reproduced by permission of The Royal Microscopical Society.