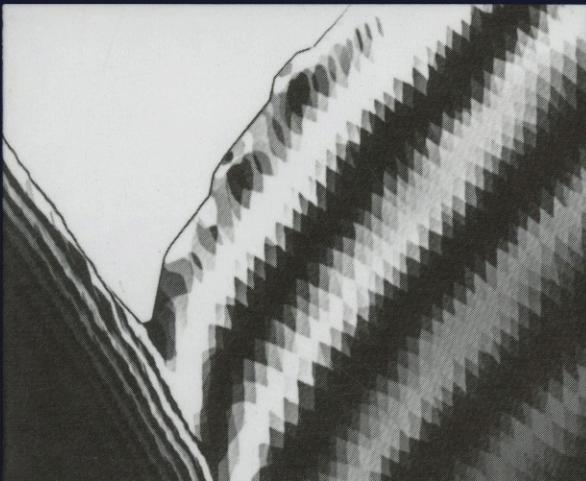
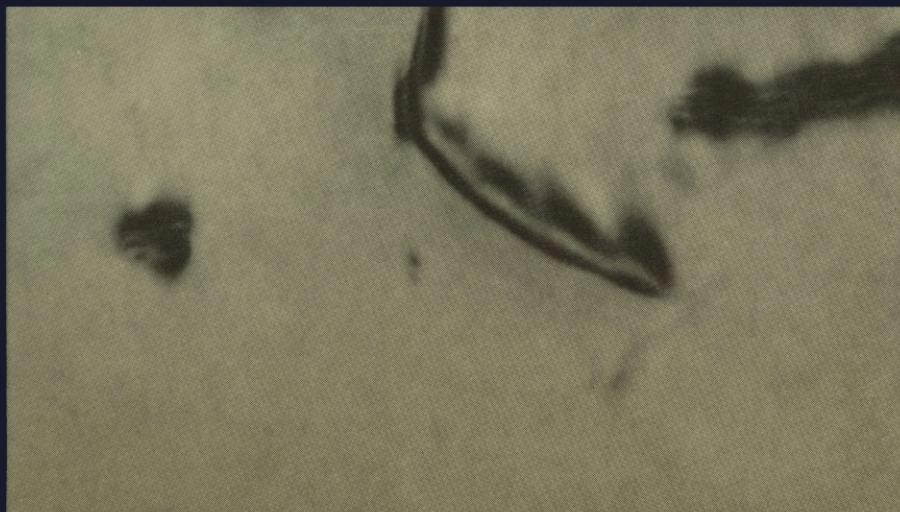
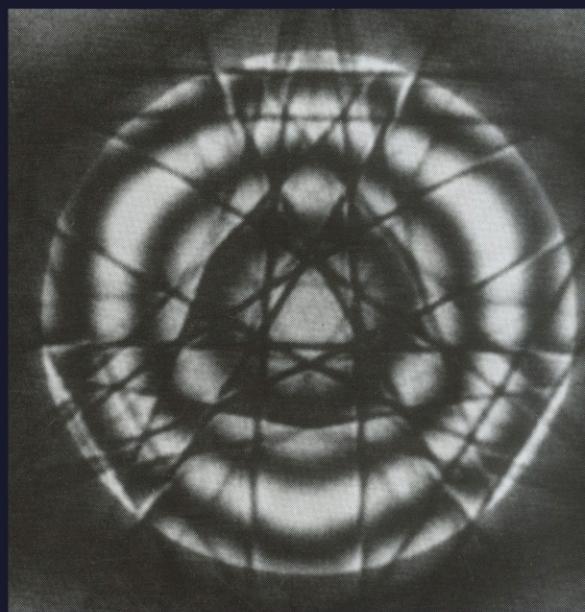


# TRANSMISSION ELECTRON MICROSCOPY

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

502'.8'25—dc20

96-28435

CIP

---

ISBN 978-0-306-45324-3

© 1996 Springer Science+Business Media New York  
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

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher

*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@cem.s.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 Veyssiére, 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	CL	cathodoluminescence
A/D	analog to digital (converter)	CRT	cathode-ray tube
ADF	annular dark field	CS	crystallographic shear
AEM	analytical electron microscope/microscopy	CSL	coincident-site lattice
AES	Auger electron spectrometer/spectroscopy	DF	dark field
AFF	aberration-free focus	DOS	density of states
ALCHEMI	atom location by channeling-enhanced micro-analysis	DP	diffraction pattern
ANL	Argonne National Laboratory	DQE	detection quantum efficiency
APB	anti-phase domain boundary	DSTEM	dedicated scanning transmission electron microscope/microscopy
ASU	Arizona State University	DTSA	desktop spectrum analyzer
ATW	atmospheric thin window	EBIC	electron beam-induced current/conductivity
BF	bright field	EELS	electron energy-loss spectrometry
BFP	back focal plane	EFI	energy-filtered imaging
BSE	backscattered electron	ELNES	energy-loss near-edge structure
BSED	backscattered-electron diffraction	ELP	energy-loss program (Gatan)
BZB	Brillouin-zone boundary	EMMA	electron microscope microanalyzer
C(1, 2, etc.)	condenser (1, 2, etc.) lens	EMS	electron microscopy image simulation
CB	coherent bremsstrahlung	EPMA	electron probe microanalyzer
CBED	convergent-beam electron diffraction	ESCA	electron spectroscopy for chemical analysis
CBIM	convergent-beam imaging	ESI	electron spectroscopic imaging
CCD	charge-coupled device	EXAFS	extended X-ray absorption fine structure
CCF	cross-correlation function	EXELFS	extended energy-loss fine structure
CCM	charge-collection microscopy	FCF	fluorescence correction factor
CDF	centered dark field	FEG	field-emission gun
CF	coherent Fresnel/Foucault	FET	field-effect transistor
CFEG	cold field-emission gun	FFT	fast Fourier transform

FOLZ	first-order Laue zone	QHRTEM	quantitative high-resolution transmission electron microscopy
FSE	fast secondary electron	RB	translation boundary (yes, it does!)
FTP	file transfer protocol	RCP	rocking-beam channeling patterns
FWHM	full width at half maximum	RDF	radial distribution function
FWTM	full width at tenth maximum	REM	reflection electron microscope/microscopy
GB	grain boundary	RHEED	reflection high-energy electron diffraction
GCS	generalized cross section	RHF	relativistic Hartree–Fock
GIF	Gatan image filter	RHFS	relativistic Hartree–Fock–Slater
GOS	generalized oscillator strength	SAD	selected-area diffraction
HAADF	high-angle annular dark field	SE	secondary electron
HOLZ	higher-order Laue zone	SEELS	serial electron energy-loss spectrometer/spec-trometry
HPGe	high-purity germanium	SEM	scanning electron microscope/microscopy
HRTEM	high-resolution transmission electron micro-scope/microscopy	SF	stacking fault
HV	high vacuum	SHRLI	simulated high-resolution lattice images
HVEM	high voltage electron microscope/microscopy	SIMS	secondary ion mass spectrometry
IDB	inversion domain boundary	S/N	signal-to-noise ratio
IEEE	International Electronics and Electrical Engineering	SOLZ	second-order Laue zone
IG	intrinsic Ge	SRM	standard reference material
IVEM	intermediate voltage electron microscope/micro-scopy	STEM	scanning transmission electron microscope/micro-scopy
K–M	Kossel–Möllenstedt	STM	scanning tunneling microscope/microscopy
LEED	low-energy electron diffraction	TB	twin boundary
LLS	linear least-squares	TEM	transmission electron microscope/microscopy
LUT	look-up table	TMBA	too many bloody acronyms
MC	minimum contrast	UHV	ultrahigh vacuum
MCA	multichannel analyzer	UTW	ultrathin window
MDM	minimum detectable mass	V/F	voltage to frequency (converter)
MLS	multiple least-squares	VLM	visible-light microscope/microscopy
MMF	minimum mass fraction	WB	weak beam
MSDS	material safety data sheets	WBDF	weak-beam dark field
NCEMSS	National Center for Electron Microscopy sim-ulation system	WDS	wavelength-dispersive spectrometer/spectrometry
NIH	National Institutes of Health	WP	whole pattern
NIST	National Institute of Standards and Technology	WPOA	weak-phase object approximation
OR	orientation relationship	WWW	World Wide Web
OTEDP	oblique-textured electron diffraction pattern	XANES	X-ray absorption near-edge structure
PB	phase boundary	XEDS	X-ray energy-dispersive spectrometer/spectrometry
P/B	peak-to-background ratio	XRD	X-ray diffraction
PEELS	parallel electron energy-loss spectrometer/spec-trometry	YBCO	yttrium–barium–copper oxide
PIMS	Precision Ion-Milling System™	YAG	yttrium–aluminum garnet
PIPS	Precision Ion-Polishing System™	ZAF	atomic number, absorption, fluorescence (correc-tion)
PM	photomultiplier	ZAP	zone-axis pattern
POA	phase-object approximation	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	$C_s$	spherical aberration coefficient
$a_0$	Bohr radius	$C_x$	fraction of X atoms on specific sites
$\mathbf{a}, \mathbf{b}, \mathbf{c}$	lattice vectors	$C_0$	amplitude of direct beam
$\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$	reciprocal lattice vectors	$C_e$	combination of the elastic constants
$A$	amplitude of scattered beam	$(C_s\lambda)^{1/2}$	scherzer
$\mathbf{A}$	ampères	$(C_s\lambda^3)^{1/4}$	glaser
$A$	absorption correction factor	$d$	beam (probe) diameter
$A$	active area of the detector	$d$	diameter of spectrometer entrance aperture
$A$	Richardson's constant	$d$	spacing of moiré fringes
$A$	atomic weight	$d$	interplanar spacing
$\text{\AA}$	Ångstrom	$d_c$	effective source size
$\mathcal{A}$	Bloch wave amplitude	$d_d$	diffraction-limited beam diameter
$A(\mathbf{U})$	aperture function	$d_g$	Gaussian beam diameter
$b$	beam-broadening parameter	$d_{hkl}$	$hkl$ interplanar spacing
$\mathbf{b}_e$	edge component of the Burgers vector	$d_{im}$	smallest resolvable image distance
$\mathbf{b}_p$	Burgers vector of partial dislocation	$d_{ob}$	smallest resolvable object distance
$\mathbf{b}_T$	Burgers vector of total dislocation	$d_s$	spherical-aberration limited beam diameter
$\mathbf{B}$	beam direction	$d_{eff}$	effective entrance aperture diameter at recording plane
$\mathbf{B}$	magnetic field strength	$dz$	thickness of a diffracting slice
$B$	background intensity	$d\sigma/d\Omega$	differential cross section of one atom
$B(\mathbf{U})$	aberration function	$D$	change in focus
$c$	velocity of light	$D$	distance from projector crossover to recording plane
$C$	contrast	$D$	electron dose
$C$	composition	$D_{im}$	depth of focus
$C_a$	astigmatism aberration coefficient	$D_{ob}$	depth of field
$C_c$	chromatic aberration coefficient	$D_1, D_2$	tie-line points on dispersion surfaces in presence of defect
$C_g$	$\mathbf{g}$ component of Bloch wave		

$e$	charge on the electron	$H$	spacing of the reciprocal-lattice planes parallel to the electron beam
$E$	energy	$H(\mathbf{u})$	Fourier transform of $h(\mathbf{r})$
$\mathbf{E}$	electric field	$I$	intensity
$E$	Young's modulus	$I$	intrinsic line width of the detector
$E_a$	spatial coherence envelope	$i_e$	emission current
$E_c$	chromatic aberration envelope	$i_f$	filament heating current
$E_c$	critical ionization energy	$I_g$	intensity in the diffracted beam
$E_d$	displacement energy	$I_K$	K-shell intensity above background
$E_K$	ionization energy for K-shell electron	$I(\mathbf{k})$	kinematical intensity
$E_L$	ionization energy for L-shell electron	$I_p$	intensity in the first plasmon peak
$E$	total energy	$I_T$	total transmitted intensity
$E$	energy loss	$I_0$	intensity in the zero-loss peak
$E_m$	average energy loss	$I_0$	intensity in the direct beam
$E_p$	plasmon energy loss	$I(\ell)$	low-loss spectrum intensity
$E_p$	plasmon energy	$J$	current density
$E_t$	threshold energy	$k$	magnitude of the wave vector
$E_0$	beam energy	$k$	Boltzmann's constant
$E(\mathbf{U})$	envelope function	$k$	kilo
$E_c(\mathbf{u})$	envelope function for chromatic aberration	$\mathbf{k}_I$	$\mathbf{k}$ -vector of the incident wave
$E_d(\mathbf{u})$	envelope function for specimen drift	$\mathbf{k}_D$	$\mathbf{k}$ -vector of the diffracted wave
$E_D(\mathbf{u})$	envelope function for the detector	$k_{AB}$	Cliff–Lorimer factor
$E_s(\mathbf{u})$	envelope function for the source	$K$	bulk modulus
$E_v(\mathbf{u})$	envelope function for specimen vibration	$K$	Kelvin
$f$	focal length	$K$	Kramers constant
$f(\mathbf{r})$	strength of object at point $x,y$	$K$	sensitivity factor
$f(\theta)$	atomic scattering factor/amplitude	$K$	inner core shell/characteristic X-ray line/ionization edge
$f_x$	scattering factor for X-rays	$\mathbf{K}$	change in $\mathbf{k}$ due to diffraction
$f_i(x)$	residual of least-squares fit	$K_B$	magnitude of $\mathbf{K}$ at the Bragg angle
$F$	Fano factor	$K_o$	kernel
$F$	fluorescence correction factor	$L$	camera length
$\mathbf{F}$	Lorentz force	$m$	number of focal increments
$F_B$	fraction of B alloying element	$m_0$	rest mass of the electron
$F^g$	special value of $F(\theta)$ when $\theta$ is the Bragg angle	$M$	magnification
$F(P)$	Fourier transform of plasmon intensity	$M$	mega
$F(\mathbf{u})$	Fourier transform of $f(\mathbf{r})$	$M_A$	angular magnification
$F(0)$	Fourier transform of elastic intensity	$M_T$	transverse magnification
$F(1)$	Fourier transform of single-scattering intensity	$M_1, M_2$	tie-line points on dispersion surfaces
$F(\theta)$	structure factor	$n$	integer
$\mathbf{g}$	diffraction vector (magnitude of $\mathbf{K}$ at the Bragg angle)	$n$	free-electron density
$g(\mathbf{r})$	intensity of image at point $(x,y)$	$n$	nano
$G$	Bragg reflection	$\mathbf{n}$	vector normal to the surface
$G$	radius of a HOLZ ring	$n_s$	number of electrons in the ionized subshell
$G$	giga	$N$	noise
$G(\mathbf{u})$	Fourier transform of $g(\mathbf{r})$	$N$	number of counts
$h$	Planck's constant	$N$	number of atoms per unit area
$h$	distance from specimen to the aperture	$N$	$h + k + \ell$
$h(r)$	contrast transfer function		
$(hkl)$	Miller indices of a crystal plane		
$hkl$	indices of diffraction spots from $hkl$ plane		

$N(E)$	number of bremsstrahlung photons of energy $E$	$v$	image distance
$N_0$	Avogadro's number	$v$	velocity of an electron
$O$	direct beam	$V$	accelerating voltage
$p$	pico	$\mathcal{V}$	potential energy
$\mathbf{p}$	momentum	$V_c$	the volume of the unit cell
$p$	integer	$V_{\text{c}}$	inner potential of cavity
$P$	peak intensity	$V_t$	projected potential through the thickness of the specimen
$P$	FWHM of a randomized electronic pulse generator	$V(\mathbf{r})$	crystal inner potential
$P_K$	probability of K-shell ionization	$T$	absolute temperature
$P(z)$	scattering matrix for a slice of thickness $z$	$T$	Tesla
$Q$	number of scattering events per unit distance	$T_C$	period of rotation
$Q$	cross section	$(UVW)$	indices of a crystal direction
$r$	radius	$UVW$	indices of beam direction
$r$	distance a wave propagates	$w$	$s\xi_g$ (excitation error $\times$ extinction distance)
$r$	power term to fit background in EELS spectrum	$\times$	times
$r_M$	image translation distance	$x$	distance
$\mathbf{r}_n$	lattice vector	$\times$	times (magnification)
$\mathbf{r}^*$	reciprocal lattice vector	$x, y, z$	atom coordinates
$r_{\text{ast}}$	astigmatism disk radius	$X$	FWHM due to detector
$r_{\text{chr}}$	chromatic-aberration disk radius	$y$	parallax shift in the image
$r_{\text{sph}}$	spherical-aberration disk radius	$y$	displacement at the specimen
$r_{\min}$	minimum disk radius	$z$	specimen height (distance along the optic axis)
$r_{\text{th}}$	theoretical disk radius	$Z$	atomic number/atomic number correction factor
$\mathbf{r}'_n$	lattice vector in strained crystal		Greek symbols
$r_0$	maximum radius of DP in focal plane of spectrometer	$\alpha$	phase shift due to defect
$\mathbf{R}$	crystal lattice vector	$\alpha$	semiangle of incidence/convergence
$R$	count rate	$\alpha$	X-ray take-off angle
$R$	resolution of XEDS detector	$\alpha_{\text{opt}}$	optimum convergence semiangle
$R$	distance on screen between diffraction spots	$\beta$	brightness
$\mathbf{R}_n$	lattice displacement vector	$\beta$	ratio of electron velocity to light velocity
$s$	excitation error or deviation parameter	$\beta$	semiangle of collection
$s_R$	excitation error due to defect	$\beta_{\text{opt}}$	optimum collection semiangle
$s_z(s_g)$	excitation error in the $z$ direction	$\gamma$	degree of spatial coherence
$s_{\text{eff}}$	effective excitation error	$\gamma$	phase of direct beam
$S$	distance from the specimen to detector	$\Delta$	change/difference
$S$	signal	$\Delta$	width of energy window
$S$	standard deviation for $n$ measurements	$\Delta\phi$	phase difference
$\text{sr}$	steradians	$\Delta\theta_i$	angles between Kossel–Möllenstedt fringes
$\mathbf{t}$	shift vector between the ZOLZ and the HOLZ	$\Delta_{AB}$	difference in mass-absorption coefficients
$t'$	absorption path length	$\Delta E$	energy spread
$T(\mathbf{u})$	objective-lens transfer function	$\Delta E_p$	plasmon line width
$T_{\text{eff}}(\mathbf{u})$	effective transfer function	$\Delta f$	maximum difference in focus
$u$	object distance	$\Delta f_{\text{AFF}}$	aberration-free (de)focus
$\mathbf{u}$	unit vector along the dislocation line	$\Delta f_{\text{MC}}$	minimum contrast defocus
$\mathbf{u}^*$	vector normal to the ZOLZ	$\Delta f_{\text{opt}}$	optimum defocus
$u_k$	displacement field		
$U$	overvoltage		
$U_g$	Fourier component of the perfect-crystal potential		

$\Delta f_{\text{sch}}$	Scherzer defocus	$\xi_g$	extinction distance for diffracted beam
$\Delta h$	relative depth in specimen	$\xi_g'$	absorption parameter
$\Delta I$	change in intensity	$\xi_0$	extinction distance for direct beam
$\Delta p$	parallax shift	$\xi_g^{\text{abs}}$	absorption-modified $\xi_g$
$\Delta V$	change in the inner potential	$\lambda$	mean-free path
$\Delta x$	path difference	$\lambda$	wavelength
$\Delta x$	half-width of image of undissociated screw dislocation	$\lambda_c$	coherence length
$\Delta x_{\text{res}}$	resolution at Scherzer defocus	$\lambda_p$	plasmon mean free path
$\Delta z$	change in height	$\lambda^{-1}$	radius of Ewald sphere
$\delta$	angle between detector normal and line from detector to specimen	$\mu$	micro
$\delta$	diameter of disk image	$\mu$	refractive index
$\delta$	diffuseness of interface	$\mu/\rho$	mass absorption coefficient
$\delta$	fluorescence enhancement ratio	$\mu^{(j)}(\mathbf{r})$	Bloch function
$\delta$	precipitate/matrix misfit	$\nu$	frequency
$\epsilon$	angle of deflection	$\psi$	amplitude of a wave
$\epsilon$	detector efficiency	$\psi^T$	total wave function
$\epsilon$	energy to create an electron–hole pair	$\psi^{\text{tot}}$	total wave function
$\epsilon$	strain	$\rho$	angle between directions
$\epsilon_0$	permittivity of free space (dielectric constant)	$\rho$	density
$\eta(\theta)$	phase of the atomic scattering factor	$\rho_c$	information limit due to chromatic aberration
$\eta$	phase change	$\rho(\mathbf{r})$	radial distribution function
$\nu$	Poisson's ratio	$\rho t$	mass thickness
$\Phi$	work function	$\rho_i^2$	area of a pixel
$\Phi_A^{\Delta p}$	X-ray emission from element A in an isolated thin film	$\sigma$	scattering cross section of one atom
$\phi$	angle between Kikuchi line and diffraction spot	$\sigma$	standard deviation
$\phi$	angle between two Kikuchi line pairs	$\sigma$	stress
$\phi$	angle between two planes	$\sigma_K$	ionization cross section for K-shell electron
$\phi$	angle of tilt between stereo images	$\sigma_T$	total scattering cross section
$\phi$	phase of a wave	$\sigma_K(\beta\Delta)$	partial ionization cross section
$\phi^*$	complex conjugate of $\phi$	$\theta$	scattering semiangle
$\phi_g$	amplitude of the diffracted beam	$\theta_B$	Bragg angle
$\phi_0$	amplitude of the direct beam	$\theta_C$	cut-off semiangle
$\phi_x$	angle of deflection of the beam	$\theta_E$	characteristic scattering semiangle
$\phi(\rho t)$	depth distribution of X-ray production	$\theta_0$	screening parameter
$\chi$	wave vector outside the specimen	$\tau$	detector time constant
$\chi_G$	wave vector which terminates on the point G in reciprocal space	$\tau$	dwell time
$\chi_O$	wave vector which terminates on the point O in reciprocal space	$\omega$	fluorescence yield
$\chi(\mathbf{u})$	phase-distortion function	$\omega_c$	cyclotron frequency
$\kappa$	thermal conductivity	$\omega_p$	plasmon frequency
		$\Omega$	solid angle of collection
		$\otimes$	convolution (multiply and integrate)

# Contents

---

## I ■ Basics

### 1 ■ The Transmission Electron Microscope

Chapter Preview .....	3
1.1. Why Use Electrons? .....	5
1.1.A. An Extremely Brief History .....	5
1.1.B. Microscopy and the Concept of Resolution .....	6
1.1.C. Interaction of Electrons with Matter .....	7
1.1.D. Depth of Field .....	8
1.1.E. Diffraction .....	9
1.2. Limitations of the TEM .....	9
1.2.A. Sampling .....	9
1.2.B. Interpreting Transmission Images .....	10
1.2.C. Electron Beam Damage and Safety .....	10
1.2.D. Specimen Preparation .....	11
1.3. Different Kinds of TEMs .....	11
1.4. Some Fundamental Properties of Electrons .....	11
1.5. Microscopy on the Internet/World Wide Web .....	13
1.5.A. Microscopy and Microanalysis-Related WWW Sites .....	14
1.5.B. Microscopy and Microanalysis Software .....	14
Chapter Summary .....	16
References .....	16

### 2 ■ Scattering and Diffraction

Chapter Preview .....	19
2.1. Why Are We Interested in Electron Scattering? .....	21
2.2. Terminology of Scattering .....	22
2.3. The Characteristics of Electron Scattering .....	23
2.4. The Interaction Cross Section .....	24
2.5. The Mean Free Path .....	25
2.6. The Differential Cross Section .....	25

## 2.7. Other Factors Affecting Scattering .....

2.8. Comparison to X-ray Diffraction .....	26
2.9. Fraunhofer and Fresnel Diffraction .....	27
2.10. Diffraction of Light from Slits .....	27
2.11. Coherent Interference .....	28
2.12. A Word about Angles .....	30
2.13. Electron Diffraction Patterns .....	30
Chapter Summary .....	31
References .....	33

## 3 ■ Elastic Scattering

Chapter Preview .....	35
3.1. Particles and Waves .....	37
3.2. Mechanisms of Elastic Scattering .....	37
3.3. Scatter from Isolated Atoms .....	38
3.4. The Rutherford Cross Section .....	39
3.5. Modifications to the Rutherford Cross Section .....	39
3.6. Coherency of the Rutherford-Scattered Electrons .....	40
3.7. The Atomic Scattering Factor .....	41
3.8. The Origin of $f(\theta)$ .....	42
3.9. The Structure Factor $F(\theta)$ .....	43
3.10. Simple Diffraction Concepts .....	44
3.10.A. Interference of Electron Waves .....	44
3.10.B. Diffraction Equations .....	45
Chapter Summary .....	46
References .....	47

---

4 ■ Inelastic Scattering and Beam Damage	49
Chapter Preview .....	49
4.1. Which Inelastic Processes Occur in the TEM? .....	51
4.2. X-ray Emission .....	51
4.2.A. Characteristic X-rays .....	52
4.2.B. Bremsstrahlung X-rays .....	56

4.3. Secondary Electron Emission .....	57	6.3. Electron Lenses .....	91
4.3.A. Slow Secondary Electrons .....	57	6.3.A. Polepieces and Coils .....	91
4.3.B. Fast Secondary Electrons .....	58	6.3.B. Different Kinds of Lenses .....	92
4.3.C. Auger Electrons .....	58	6.3.C. Electron Ray Paths through Magnetic Fields .....	93
4.4. Electron–Hole Pairs and Cathodoluminescence (CL) .....	59	6.3.D. Image Rotation and the Eucentric Plane .....	95
4.5. Plasmons and Phonons .....	60	6.3.E. Deflecting the Beam .....	95
4.6. Beam Damage .....	61	6.4. Apertures and Diaphragms .....	95
4.6.A. Electron Dose .....	62	6.5. Real Lenses and Their Problems .....	97
4.6.B. Specimen Heating .....	62	6.5.A. Spherical Aberration .....	97
4.6.C. Beam Damage in Polymers .....	62	6.5.B. Chromatic Aberration .....	98
4.6.D. Beam Damage in Covalent and Ionic Crystals .....	63	6.5.C. Astigmatism .....	98
4.6.E. Beam Damage in Metals .....	63	6.6. The Resolution of the Electron Lens (and Ultimately of the TEM) .....	99
4.6.F. Sputtering .....	65	6.6.A. Theoretical Resolution .....	99
Chapter Summary .....	65	6.6.B. Spherical Aberration-Limited Resolution (The Practical Resolution) ...	100
References .....	65	6.6.C. Confusion in the Definitions of Resolution .....	101
<b>5 ■ Electron Sources</b>		<b>6.7. Depth of Focus and Depth of Field .....</b>	102
Chapter Preview .....	67	Chapter Summary .....	103
5.1. The Physics of Different Electron Sources .....	69	References .....	104
5.1.A. Thermionic Emission .....	69	<b>7 ■ How to “See” Electrons</b>	
5.1.B. Field Emission .....	70	Chapter Preview .....	105
5.2. The Characteristics of the Electron Beam .....	70	7.1. Electron Detection and Display .....	107
5.2.A. Brightness .....	70	7.2. Viewing Screens .....	108
5.2.B. Temporal Coherency and Energy Spread .....	71	7.3. Electron Detectors .....	108
5.2.C. Spatial Coherency and Source Size .....	72	7.3.A. Semiconductor Detectors .....	108
5.2.D. Stability .....	72	7.3.B. Scintillator–Photomultiplier Detectors ...	110
5.3. Electron Guns .....	72	7.3.C. TV Cameras and Charge-Coupled Devices .....	111
5.3.A. Thermionic Guns .....	72	7.3.D. Faraday Cup .....	112
5.3.B. Field-Emission Guns (FEGs) .....	76	7.4. Which Detector Do We Use for Which Signal? .....	112
5.4. Comparison of Guns .....	77	7.5. Image Recording .....	114
5.5. Measuring Your Gun Characteristics .....	78	7.5.A. Photographic Emulsions .....	114
5.5.A. Beam Current .....	78	7.5.B. Other Image Recording Methods .....	114
5.5.B. Convergence Angle .....	78	7.6. Comparison of Scanning Images and Static TEM Images .....	115
5.5.C. Calculating the Beam Diameter .....	79	Chapter Summary .....	115
5.5.D. Measuring the Beam Diameter .....	79	References .....	115
5.5.E. Energy Spread .....	82	<b>8 ■ Pumps and Holders</b>	
5.5.F. Spatial Coherency .....	82	Chapter Preview .....	117
5.6. What kV Should You Use? .....	82	8.1. Vacuums .....	119
Chapter Summary .....	82	8.2. Roughing Pumps .....	119
References .....	83	8.3. High/Ultra-High Vacuum Pumps .....	120
<b>6 ■ Lenses, Apertures, and Resolution</b>		8.3.A. Diffusion Pumps .....	120
Chapter Preview .....	85	8.3.B. Turbomolecular Pumps .....	120
6.1. Why Learn about Lenses? .....	87	8.3.C. Ion Pumps .....	121
6.2. Light Optics and Electron Optics .....	87	8.3.D. Cryogenic (Adsorption) Pumps .....	121
6.2.A. How to Draw a Ray Diagram .....	88		
6.2.B. The Principal Optical Elements .....	89		
6.2.C. The Lens Equation .....	90		
6.2.D. Magnification, Demagnification, and Focus .....	90		

8.4. The Whole System .....	122	10.3.A. Creation of a Thin Slice from the Bulk Sample .....	159
8.5. Leak Detection .....	123	10.3.B. Cutting the Disk .....	159
8.6. Contamination: Hydrocarbons and Water Vapor .....	123	10.3.C. Prethinning the Disk .....	160
8.7. Specimen Holders and Stages .....	124	10.4. Final Thinning of the Disk .....	161
8.8. Side-Entry Holders .....	124	10.4.A. Electropolishing .....	161
8.9. Top-Entry Holders .....	125	10.4.B. Ion Milling .....	162
8.10. Different Types of Holders .....	125	10.5. Cross-Section Specimens .....	165
Chapter Summary .....	128	10.6. Specimens on Grids/Washers .....	166
References .....	129	10.6.A. Electropolishing—The Window Method for Metals and Alloys .....	166
<b>9 ■ The Instrument</b>		10.6.B. Ultramicrotomy .....	167
Chapter Preview .....	131	10.6.C. Grinding and Crushing .....	167
9.1. The Illumination System .....	133	10.6.D. Replication and Extraction .....	168
9.1.A. TEM Operation Using a Parallel Beam ..	133	10.6.E. Cleaving .....	168
9.1.B. Convergent-Beam (S)TEM Mode ..	133	10.6.F. The 90° Wedge .....	169
9.1.C. Translating and Tilting the Beam ..	135	10.6.G. Lithography .....	169
9.1.D. Alignment .....	136	10.6.H. Preferential Chemical Etching .....	169
9.1.E. Condenser Lens Defects .....	137	10.7. Storing Specimens .....	170
9.1.F. Calibration .....	138	Chapter Summary .....	170
9.2. The Objective Lens and Stage .....	139	References .....	172
9.3. Forming Diffraction Patterns and Images:			
The TEM Imaging System .....	140	<b>II ■ Diffraction</b>	
9.3.A. Selected-Area Diffraction .....	140		
9.3.B. Bright-Field and Dark-Field Imaging ..	142	<b>11 ■ Diffraction Patterns</b>	
9.3.C. Centered DF Operation .....	143	Chapter Preview .....	177
9.4. Forming Diffraction Patterns and Images:		11.1. Why Use Diffraction in the TEM? .....	179
The STEM Imaging System .....	144	11.2. The TEM, Diffraction Cameras, and the TV ..	179
9.4.A. Bright-Field STEM Images .....	144	11.3. Scattering from a Plane of Atoms .....	181
9.4.B. Dark-Field STEM Images .....	145	11.4. Scattering from a Crystal .....	182
9.4.C. Annular DF Images .....	145	11.5. Meaning of $n$ in Bragg's Law .....	184
9.4.D. Magnification in STEM .....	146	11.6. A Pictorial Introduction to Dynamical Effects .....	185
9.5. Alignment and Stigmation .....	147	11.7. Use of Indices in Diffraction Patterns .....	185
9.5.A. Lens Rotation Centers .....	147	11.8. Practical Aspects of Diffraction-Pattern Formation .....	185
9.5.B. Correction of Astigmatism in the Imaging Lenses .....	147	11.9. More on Selected-Area Diffraction Patterns ..	186
9.6. Calibration of the Imaging System .....	148	Chapter Summary .....	189
9.6.A. Magnification Calibration .....	148	References .....	189
9.6.B. Camera-Length Calibration .....	150		
9.6.C. Rotation of the Image Relative to the Diffraction Pattern .....	151		
9.6.D. Analysis of TEM Images and Diffraction Patterns .....	151	<b>12 ■ Thinking in Reciprocal Space</b>	
9.7. Other Calibrations .....	151	Chapter Preview .....	191
Chapter Summary .....	153	12.1. Why Introduce Another Lattice? .....	193
References .....	153	12.2. Mathematical Definition of the Reciprocal Lattice .....	193
<b>10 ■ Specimen Preparation</b>		12.3. The Vector $\mathbf{g}$ .....	194
Chapter Preview .....	155	12.4. The Laue Equations and Their Relation to Bragg's Law .....	195
10.1. Safety .....	157	12.5. The Ewald Sphere of Reflection .....	196
10.2. Self-Supporting Disk or Use a Grid? .....	158	12.6. The Excitation Error .....	197
10.3. Preparing a Self-Supporting Disk for Final Thinning .....	159	12.7. Thin-Foil Effect and the Effect of Accelerating Voltage .....	199

Chapter Summary .....	200	16.3. Some Important Structures: bcc, fcc, and hcp .....	241
References .....	200	16.4. Extending fcc and hcp to Include a Basis .....	242
<b>13 ■ Diffracted Beams</b>		16.5. Applying the bcc and fcc Analysis to Simple Cubic .....	243
Chapter Preview .....	201	16.6. Extending hcp to Ti <sub>3</sub> Al .....	244
13.1. Why Calculate Intensities? .....	203	16.7. Superlattice Reflections and Imaging .....	244
13.2. The Approach .....	203	16.8. Diffraction from Long-Period Superlattices .....	246
13.3. The Amplitude of a Diffracted Beam .....	204	16.9. Forbidden Reflections .....	247
13.4. The Characteristic Length $\xi_g$ .....	205	16.10. Using the International Tables .....	247
13.5. The Howie–Whelan Equations .....	206	Chapter Summary .....	249
13.6. Reformulating the Howie–Whelan Equations .....	207	References .....	249
13.7. Solving the Howie–Whelan Equations .....	207	<b>17 ■ Diffraction from Small Volumes</b>	
13.8. The Importance of $\gamma^{(1)}$ and $\gamma^{(2)}$ .....	208	Chapter Preview .....	251
13.9. The Total Wave Amplitude .....	209	17.1. Introduction .....	253
13.10. The Effective Excitation Error .....	210	17.1.A. The Summation Approach .....	253
13.11. The Column Approximation .....	210	17.1.B. The Integration Approach .....	254
13.12. The Approximations and Simplifications .....	211	17.2. The Thin-Foil Effect .....	255
13.13. The Coupled Harmonic Oscillator Analog .....	212	17.3. Diffraction from Wedge-Shaped Specimens .....	256
Chapter Summary .....	212	17.4. Diffraction from Planar Defects .....	256
References .....	213	17.5. Diffraction from Particles .....	258
<b>14 ■ Bloch Waves</b>		17.6. Diffraction from Dislocations, Individually and Collectively .....	259
Chapter Preview .....	215	17.7. Diffraction and the Dispersion Surface .....	261
14.1. Wave Equation in TEM .....	217	Chapter Summary .....	262
14.2. The Crystal .....	217	References .....	263
14.3. Bloch Functions .....	219	<b>18 ■ Indexing Diffraction Patterns</b>	
14.4. Schrödinger's Equation for Bloch Waves .....	219	Chapter Preview .....	265
14.5. The Plane-Wave Amplitudes .....	221	18.1. Choosing Your Technique .....	267
14.6. Absorption of Bloch Waves .....	223	18.2. Experimental Techniques .....	267
Chapter Summary .....	223	18.3. The Stereographic Projection .....	269
References .....	224	18.4. Indexing Single-Crystal Diffraction Patterns .....	271
<b>15 ■ Dispersion Surfaces</b>		18.5. Ring Patterns from Polycrystalline Materials .....	273
Chapter Preview .....	225	18.6. Ring Patterns from Amorphous Materials .....	274
15.1. Introduction .....	227	18.7. Double Diffraction .....	278
15.2. The Dispersion Diagram when $U_g = 0$ .....	228	18.8. Orientation of the Specimen .....	280
15.3. The Dispersion Diagram when $U_g \neq 0$ .....	228	18.9. Orientation Relationships .....	280
15.4. Relating Dispersion Surfaces and Diffraction Patterns .....	229	18.10. Computer Analysis .....	285
15.5. The Relation between $U_g$ , $\xi_g$ , and $s_g$ .....	232	Chapter Summary .....	286
15.6. The Amplitudes of Bloch Waves .....	233	References .....	286
15.7. Extending to More Beams .....	234		
15.8. Dispersion Surfaces and Defects .....	235		
Chapter Summary .....	235		
References .....	236		
<b>16 ■ Diffraction from Crystals</b>			
Chapter Preview .....	237		
16.1. Review of Diffraction from a Primitive Lattice .....	239		
16.2. Structure Factors: The Idea .....	240		
		<b>19 ■ Kikuchi Diffraction</b>	
		Chapter Preview .....	289
		19.1. The Origin of Kikuchi Lines .....	291
		19.2. Kikuchi Lines and Bragg Scattering .....	291
		19.3. Constructing Kikuchi Maps .....	293

CONTENTS	xxiii		
19.4. Crystal Orientation and Kikuchi Maps .....	295	21.7. Scanning-Beam Diffraction .....	341
19.5. Setting the Value of $s_g$ .....	296	21.8. Other Methods of Microdiffraction .....	343
19.6. Intensities .....	297	Chapter Summary .....	344
Chapter Summary .....	298	References .....	345
References .....	298		
<b>20 ■ Obtaining CBED Patterns</b>		<b>III ■ Imaging</b>	
Chapter Preview .....	301		
20.1. Why Use a Convergent Beam? .....	303	<b>22 ■ Imaging in the TEM</b>	
20.2. Obtaining CBED Patterns .....	304	Chapter Preview .....	349
20.2.A. Comparing SAD and CBED .....	304	22.1. What Is Contrast? .....	351
20.2.B. CBED in TEM .....	305	22.2. Principles of Image Contrast .....	351
20.2.C. Choosing the C2 Aperture .....	305	22.2.A. Images and Diffraction Patterns .....	351
20.2.D. Choosing the Camera Length .....	306	22.2.B. Use of the Objective Aperture	
20.2.E. Focusing Your Pattern .....	306	or the STEM Detector: BF and DF	
20.2.F. Choice of Beam Size .....	307	Images .....	351
20.2.G. Effect of Specimen Thickness .....	308	22.3. Mass-Thickness Contrast .....	353
20.2.H. Final Adjustment .....	308	22.3.A. Mechanism of Mass-Thickness	
20.2.I. CBED in STEM Mode .....	308	Contrast .....	353
20.3. Zero-Order and Higher-Order Laue-Zone		22.3.B. TEM Images .....	354
Diffraction .....	309	22.3.C. STEM Images .....	355
20.3.A. ZOLZ Patterns .....	309	22.3.D. Specimens Which Show	
20.3.B. HOLZ Patterns .....	309	Mass-Thickness Contrast .....	357
20.3.C. Indexing HOLZ Patterns .....	311	22.3.E. Quantitative Mass-Thickness	
20.4. Kikuchi Lines in CBED Patterns .....	314	Contrast .....	357
20.5. HOLZ Lines .....	315	22.4. Z Contrast .....	358
20.5.A. The Relationship between		22.5. TEM Diffraction Contrast .....	361
HOLZ Lines and Kikuchi Lines .....	315	22.5.A. Two-Beam Conditions .....	361
20.5.B. Acquiring HOLZ Lines .....	315	22.5.B. Setting the Deviation	
20.5.C. Indexing HOLZ Lines .....	317	Parameter, $s$ .....	361
Chapter Summary .....	318	22.5.C. Setting Up a Two-Beam CDF	
References .....	318	Image .....	361
<b>21 ■ Using Convergent-Beam Techniques</b>		22.5.D. Relationship between the Image	
Chapter Preview .....	319	and the Diffraction Pattern .....	363
21.1. Thickness Determination .....	321	22.6. STEM Diffraction Contrast .....	364
21.2. Unit-Cell Determination .....	323	Chapter Summary .....	366
21.2.A. Experimental Considerations .....	323	References .....	366
21.2.B. The Importance of the HOLZ-Ring			
Radius .....	323		
21.2.C. Determining the Lattice Centering .....	325		
21.3. Symmetry Determination .....	326		
21.3.A. Introduction to Symmetry Concepts ..	326		
21.3.B. Friedel's Law .....	328		
21.3.C. Looking for Symmetry in Your			
Patterns .....	328		
21.3.D. Point-Group Determination .....	330		
21.3.E. Space-Group Determination .....	332		
21.4. Lattice Parameter, Strain,			
and Composition Analysis .....	338		
21.5. Determination of Enantiomorphism .....	341		
21.6. Convergent-Beam Imaging .....	341		
		<b>23 ■ Thickness and Bending Effects</b>	
		Chapter Preview .....	367
		23.1. The Fundamental Ideas .....	369
		23.2. Thickness Fringes .....	369
		23.3. Thickness Fringes and the Diffraction	
		Pattern .....	371
		23.4. Bend Contours (Annoying Artifact,	
		Useful Tool, and Valuable Insight) .....	372
		23.5. ZAPs and Real-Space Crystallography .....	373
		23.6. Hillocks, Dents, or Saddles .....	374
		23.7. Absorption Effects .....	374
		23.8. Computer Simulation of Thickness Fringes .....	375
		23.9. Thickness-Fringe/Bend-Contour	
		Interactions .....	375

23.10. Other Effects of Bending .....	376	<b>26 ■ Weak-Beam Dark-Field Microscopy</b>	
Chapter Summary .....	377	Chapter Preview .....	421
References .....	378	26.1. Intensity in WBDF Images .....	423
<b>24 ■ Planar Defects</b>		26.2. Setting $s_g$ Using the Kikuchi Pattern .....	423
Chapter Preview .....	379	26.3. How to Do WBDF .....	425
24.1. Translations and Rotations .....	381	26.4. Thickness Fringes in Weak-Beam Images .....	426
24.2. Why Do Translations Produce Contrast? .....	382	26.5. Imaging Strain Fields .....	427
24.3. The Scattering Matrix .....	383	26.6. Predicting Dislocation Peak Positions .....	428
24.4. Using the Scattering Matrix .....	384	26.7. Phasor Diagrams .....	430
24.5. Stacking Faults in fcc Materials .....	385	26.8. Weak-Beam Images of Dissociated	
24.5.A. Why fcc Materials? .....	386	Dislocations .....	432
24.5.B. Some Rules .....	386	26.9. Other Thoughts .....	436
24.5.C. Intensity Calculations .....	387	26.9.A. Thinking of Weak-Beam	
24.5.D. Overlapping Faults .....	388	Diffraction as a Coupled Pendulum .....	436
24.6. Other Translations: $\pi$ and $\delta$ Fringes .....	389	26.9.B. Bloch Waves .....	436
24.7. Phase Boundaries .....	391	26.9.C. If Other Reflections Are Present .....	437
24.8. Rotation Boundaries .....	391	26.9.D. The Future .....	437
24.9. Diffraction Patterns and Dispersion Surfaces .....	391	Chapter Summary .....	438
24.10. Bloch Waves and BF/DF Image Pairs .....	393	References .....	438
24.11. Computer Modeling .....	394	<b>27 ■ Phase-Contrast Images</b>	
24.12. The Generalized Cross Section .....	395	Chapter Preview .....	439
24.13. Quantitative Imaging .....	396	27.1. Introduction .....	441
24.13.A. Theoretical Basis and Parameters .....	396	27.2. The Origin of Lattice Fringes .....	441
24.13.B. Apparent Extinction Distance .....	397	27.3. Some Practical Aspects of Lattice Fringes .....	442
24.13.C. Avoiding the Column Approximation .....	397	27.3.A. If $s = 0$ .....	442
24.13.D. The User Interface .....	398	27.3.B. If $s \neq 0$ .....	442
Chapter Summary .....	398	27.4. On-Axis Lattice-Fringe Imaging .....	442
References .....	398	27.5. Moiré Patterns .....	444
<b>25 ■ Strain Fields</b>		27.5.A. Translational Moiré Fringes .....	445
Chapter Preview .....	401	27.5.B. Rotational Moiré Fringes .....	445
25.1. Why Image Strain Fields? .....	403	27.5.C. General Moiré Fringes .....	445
25.2. Howie–Whelan Equations .....	403	27.6. Experimental Observations of Moiré	
25.3. Contrast from a Single Dislocation .....	405	Fringes .....	445
25.4. Displacement Fields and Ewald’s Sphere .....	408	27.6.A. Translational Moiré Patterns .....	446
25.5. Dislocation Nodes and Networks .....	409	27.6.B. Rotational Moiré Fringes .....	447
25.6. Dislocation Loops and Dipoles .....	409	27.6.C. Dislocations and Moiré Fringes .....	447
25.7. Dislocation Pairs, Arrays, and Tangles .....	411	27.6.D. Complex Moiré Fringes .....	448
25.8. Surface Effects .....	412	27.7. Fresnel Contrast .....	450
25.9. Dislocations and Interfaces .....	413	27.7.A. The Fresnel Biprism .....	450
25.10. Volume Defects and Particles .....	417	27.7.B. Magnetic-Domain Walls .....	451
25.11. Simulating Images .....	418	27.8. Fresnel Contrast from Voids or Gas Bubbles .....	451
25.11.A. The Defect Geometry .....	418	27.9. Fresnel Contrast from Lattice Defects .....	452
25.11.B. Crystal Defects and Calculating the Displacement Field .....	418	27.9.A. Grain Boundaries .....	453
25.11.C. The Parameters .....	419	27.9.B. End-On Dislocations .....	453
Chapter Summary .....	419	Chapter Summary .....	455
References .....	420	References .....	455
<b>28 ■ High-Resolution TEM</b>		<b>28.1. The Role of an Optical System</b> .....	457
Chapter Preview .....		28.2. The Radio Analogy .....	459

CONTENTS	xxv		
28.3. The Specimen .....	461	30.5.D. Kernels .....	507
28.4. Applying the WPOA to the TEM .....	462	30.6. Applications .....	507
28.5. The Transfer Function .....	462	30.6.A. Beam-Sensitive Materials .....	507
28.6. More on $\chi(\mathbf{u})$ , $\sin \chi(\mathbf{u})$ , and $\cos \chi(\mathbf{u})$ .....	463	30.6.B. Periodic Images .....	508
28.7. Scherzer Defocus .....	465	30.6.C. Correcting Drift .....	508
28.8. Envelope Damping Functions .....	466	30.6.D. Reconstructing the Phase .....	508
28.9. Imaging Using Passbands .....	467	30.6.E. Diffraction Patterns .....	508
28.10. Experimental Considerations .....	468	30.6.F. Tilted-Beam Series .....	510
28.11. The Future for HRTEM .....	469	30.7. Automated Alignment .....	511
28.12. The TEM as a Linear System .....	470	30.8. Quantitative Methods of Image Analysis .....	512
28.13. FEGTEMs and the Information Limit .....	470	30.9. Pattern Recognition in HRTEM .....	513
28.14. Some Difficulties in Using an FEG .....	473	30.10. Parameterizing the Image Using QUANTITEM .....	514
28.15. Selectively Imaging Sublattices .....	475	30.10.A. The Example of a Specimen with Uniform Composition .....	514
28.16. Interfaces and Surfaces .....	476	30.10.B. Calibrating the Path of $R$ .....	516
28.17. Incommensurate Structures .....	478	30.10.C. Noise Analysis .....	516
28.18. Quasicrystals .....	479	30.11. Quantitative Chemical Lattice Imaging .....	518
28.19. Single Atoms .....	480	30.12. Methods of Measuring Fit .....	518
Chapter Summary .....	481	30.13. Quantitative Comparison of Simulated and Experimental HRTEM Images .....	521
References .....	481	30.14. A Fourier Technique for Quantitative Analysis .....	523
<b>29 ■ Image Simulation</b>		30.15. Real or Reciprocal Space? .....	523
Chapter Preview .....	483	30.16. The Optical Bench .....	524
29.1. Simulating Images .....	485	Chapter Summary .....	526
29.2. The Multislice Method .....	485	References .....	526
29.3. The Reciprocal-Space Approach .....	485		
29.4. The FFT Approach .....	487		
29.5. The Real-Space Approach .....	488		
29.6. Bloch Waves and HRTEM Simulation .....	488		
29.7. The Ewald Sphere Is Curved .....	489		
29.8. Choosing the Thickness of the Slice .....	490		
29.9. Beam Convergence .....	490		
29.10. Modeling the Structure .....	491		
29.11. Surface Grooves and Simulating Fresnel Contrast .....	492		
29.12. Calculating Images of Defects .....	493		
29.13. Simulating Quasicrystals .....	494		
29.14. Bonding in Crystals .....	496		
Chapter Summary .....	497		
References .....	497		
<b>30 ■ Quantifying and Processing HRTEM Images</b>			
Chapter Preview .....	499		
30.1. What Is Image Processing? .....	501		
30.2. Processing and Quantifying Images .....	501		
30.3. A Cautionary Note .....	502		
30.4. Image Input .....	502		
30.5. Processing Techniques .....	502		
30.5.A. Fourier Filtering and Reconstruction .....	502		
30.5.B. Analyzing Diffractograms .....	504		
30.5.C. Averaging Images and Other Techniques .....	506		
<b>31 ■ Other Imaging Techniques</b>			
Chapter Preview .....	529		
31.1. Stereo Microscopy .....	531		
31.2. 2½D Microscopy .....	532		
31.3. Magnetic Specimens .....	534		
31.3.A. The Magnetic Correction .....	534		
31.3.B. Lorentz Microscopy .....	535		
31.4. Chemically Sensitive Images .....	538		
31.5. Imaging with Diffusely Scattered Electrons .....	538		
31.6. Surface Imaging .....	540		
31.6.A. Reflection Electron Microscopy .....	540		
31.6.B. Topographic Contrast .....	540		
31.7. High-Order BF Imaging .....	541		
31.8. Secondary-Electron Imaging .....	541		
31.9. Backscattered-Electron Imaging .....	542		
31.10. Charge-Collection Microscopy and Cathodoluminescence .....	543		
31.11. Electron Holography .....	543		
31.12. <i>In situ</i> TEM: Dynamic Experiments .....	546		
31.13. Other Variations Possible in a STEM .....	547		
Chapter Summary .....	548		
References .....	548		

<b>IV ■ Spectrometry</b>	
<b>32 ■ X-ray Spectrometry</b>	
Chapter Preview .....	553
32.1. X-ray Analysis: Why Bother? .....	555
32.2. Basic Operational Mode .....	555
32.3. The Energy-Dispersive Spectrometer .....	558
32.4. Semiconductor Detectors .....	559
32.4.A. How Does XEDS Work? .....	559
32.4.B. Different Kinds of Windows .....	560
32.4.C. Intrinsic Germanium Detectors .....	562
32.5. Pulse Processing and Dead Time .....	563
32.6. Resolution of the Detector .....	564
32.7. What You Should Know about Your XEDS ...	565
32.7.A. Detector Variables .....	565
32.7.B. Processing Variables .....	566
32.7.C. Artifacts Common to XEDS Systems .....	568
32.8. Wavelength-Dispersive Spectrometers .....	570
Chapter Summary .....	571
References .....	572
<b>33 ■ The XEDS-TEM Interface</b>	
Chapter Preview .....	573
33.1. The Requirements .....	575
33.2. The Collimator .....	575
33.2.A. Collection Angle .....	575
33.2.B. Take-Off Angle .....	576
33.2.C. Orientation of the Detector to the Specimen .....	577
33.3. Protecting the Detector from Intense Radiation .....	578
33.4. System X-rays and Spurious X-rays .....	578
33.4.A. Pre-Specimen Effects .....	578
33.4.B. Post-Specimen Scatter .....	580
33.4.C. Coherent Bremsstrahlung .....	583
33.5. Peak-to-Background Ratio .....	583
Chapter Summary .....	584
References .....	585
<b>34 ■ Qualitative X-ray Analysis</b>	
Chapter Preview .....	587
34.1. Microscope Variables .....	589
34.2. Acquiring a Spectrum .....	589
34.3. Peak Identification .....	590
34.4. Peak Visibility .....	593
Chapter Summary .....	595
References .....	595
<b>35 ■ Quantitative X-ray Microanalysis</b>	
Chapter Preview .....	597
35.1. Historical Perspective .....	599
35.2. The Cliff-Lorimer Ratio Technique .....	600
35.3. Practical Steps for Quantitative Microanalysis .....	600
35.3.A. Background Subtraction .....	601
35.3.B. Peak Integration .....	603
35.4. Determining <i>k</i> Factors .....	605
35.4.A. Experimental Determination of <i>k</i> <sub>AB</sub> ..	605
35.4.B. Errors in Quantification .....	606
35.4.C. Calculating <i>k</i> <sub>AB</sub> .....	609
35.5. Absorption Correction .....	612
35.6. Extrapolation Techniques for Absorption Correction .....	614
35.7. The Fluorescence Correction .....	615
35.8. ALCHEMI .....	616
35.9. Examples: Profiles and Maps .....	617
Chapter Summary .....	619
References .....	619
<b>36 ■ Spatial Resolution and Minimum Detectability</b>	
Chapter Preview .....	621
36.1. Why Is Spatial Resolution Important? .....	623
36.2. Definition of Spatial Resolution .....	623
36.3. Beam Spreading .....	624
36.4. The Spatial Resolution Equation .....	625
36.5. Measurement of Spatial Resolution .....	626
36.6. Thickness Measurement .....	628
36.6.A. TEM Methods .....	628
36.6.B. Contamination-Spot Separation Method .....	629
36.6.C. X-ray Spectrometry Methods .....	630
36.6.D. Electron Energy-Loss Spectrometry Methods .....	631
36.6.E. Convergent-Beam Diffraction Method .....	631
36.7. Minimum Detectability .....	631
36.7.A. Experimental Factors Affecting the MMF .....	632
36.7.B. Statistical Criterion for the MMF ..	632
36.7.C. Comparison with Other Definitions ..	633
36.7.D. Minimum Detectable Mass .....	634
Chapter Summary .....	634
References .....	634
<b>37 ■ Electron Energy-Loss Spectrometers</b>	
Chapter Preview .....	637
37.1. Why Do Electron Energy-Loss Spectrometry? .....	639
37.2. The Magnetic Prism: A Spectrometer and a Lens .....	640
37.2.A. Focusing the Spectrometer .....	641
37.2.B. Calibrating the Spectrometer .....	642
37.3. Acquiring a Spectrum .....	642
37.3.A. Serial Collection .....	642

37.3.B. Parallel Collection .....	644	39.4.C. Edge Integration .....	676	
37.3.C. Spectrometer Dispersion .....	645	39.4.D. The Zero-Loss Integral .....	676	
37.3.D. Spectrometer Resolution .....	645	39.4.E. The Partial Ionization Cross Section .....	676	
37.3.E. Point-Spread Function .....	646	39.5. Measuring Thickness from the Energy-Loss Spectrum .....	678	
37.4. Image and Diffraction Modes .....	646	39.6. Deconvolution .....	680	
37.4.A. Spectrometer Collection Angle .....	647	39.7. Correction for Convergence of the Incident Beam .....	682	
37.4.B. Spatial Selection .....	649	39.8. The Effect of the Specimen Orientation .....	682	
37.5. What You Need to Know about Your PEELS .....	649	39.9. Spatial Resolution .....	682	
37.6. Imaging Spectrometers .....	650	39.10. Detectability Limits .....	683	
Chapter Summary .....	651	Chapter Summary .....	684	
References .....	651	References .....	684	
<b>38 ■ The Energy-Loss Spectrum</b>				
Chapter Preview .....	653	<b>40 ■ Everything Else in the Spectrum</b>		
38.1. A Few Basic Concepts .....	655	Chapter Preview .....	687	
38.2. The Zero-Loss Peak .....	656	40.1. Fine Structure in the Ionization Edges .....	689	
38.3. The Low-Loss Spectrum .....	656	40.1.A. ELNES .....	689	
38.3.A. Plasmons .....	656	40.1.B. EXELFS .....	694	
38.3.B. Inter- and Intra-Band Transitions .....	658	40.2. The Low-Loss Spectrum .....	696	
38.4. The High-Loss Spectrum .....	658	40.2.A. Plasmon Losses .....	696	
38.4.A. Inner-Shell Ionization .....	658	40.2.B. Dielectric-Constant Determination .....	697	
38.4.B. Ionization-Edge Characteristics .....	662	40.2.C. Band-Gap and Inter-Band Transitions .....	698	
38.5. Artifacts in the Spectrum .....	664	40.2.D. Angle-Resolved EELS .....	698	
Chapter Summary .....	665	40.3. Energy-Filtered and Spectrum Imaging .....	699	
References .....	666	40.3.A. STEM Digital Imaging .....	700	
<b>39 ■ Microanalysis with Ionization-Loss Electrons</b>				
Chapter Preview .....	667	40.3.B. TEM Analog Imaging .....	700	
39.1. Choice of Operating Parameters .....	669	Chapter Summary .....	703	
39.2. What Should Your Spectrum Look Like? .....	670	References .....	703	
39.3. Qualitative Microanalysis .....	671	<b>Index</b> .....	705	
39.4. Quantitative Microanalysis .....	672	<b>Acknowledgements for Figures</b> .....		
39.4.A. Derivation of the Equations for Quantification .....	673	717		
39.4.B. Background Subtraction .....	674			

# **Transmission Electron Microscopy**

**A Textbook for  
Materials Science**

# Basics



---

# The Transmission Electron Microscope

---

1.1. Why Use Electrons? .....	5
1.1.A. An Extremely Brief History .....	5
1.1.B. Microscopy and the Concept of Resolution .....	6
1.1.C. Interaction of Electrons with Matter .....	7
1.1.D. Depth of Field .....	8
1.1.E. Diffraction .....	9
1.2. Limitations of the TEM .....	9
1.2.A. Sampling .....	9
1.2.B. Interpreting Transmission Images .....	10
1.2.C. Electron Beam Damage and Safety .....	10
1.2.D. Specimen Preparation .....	11
1.3. Different Kinds of TEMs .....	11
1.4. Some Fundamental Properties of Electrons .....	11
1.5. Microscopy on the Internet/World Wide Web .....	13
1.5.A. Microscopy and Microanalysis-Related WWW Sites .....	14
1.5.B. Microscopy and Microanalysis Software .....	14

## 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. 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 Festschriften 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,  $\delta$ , is given approximately by

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

In equation 1.1,  $\lambda$  is the wavelength of the radiation,  $\mu$  the refractive index of the viewing medium, and  $\beta$  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,  $\lambda$  is about 550 nm (5500 Å), and so the resolution of a good light microscope is about 300 nm. In TEMs we can approximate

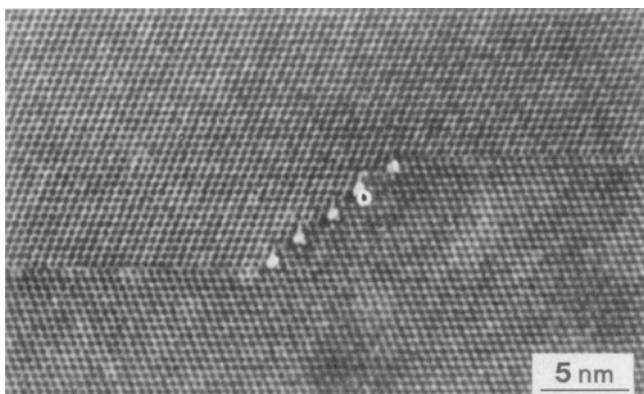
We’ll try to use nanometers throughout this book, but you’ll find that many microscopists still insist on using Ångstroms rather than the SI units. However, the Ångstrom is close to the atomic diameter and so is a more convenient unit because it saves us using convoluted phrases like “three tenths of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

the resolution in equation 1.1 to  $0.61\lambda/\beta$  which, as we’ll see later, is very small.

Now although 300 nm is a small dimension to us it corresponds to about 1000 atom diameters, and therefore many of the features that control the properties of materials are on a scale well below the resolution of the 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

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



**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  $\lambda$  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  $\lambda \sim 4 \text{ pm}$  (0.004 nm), which is much smaller than the diameter of an atom.

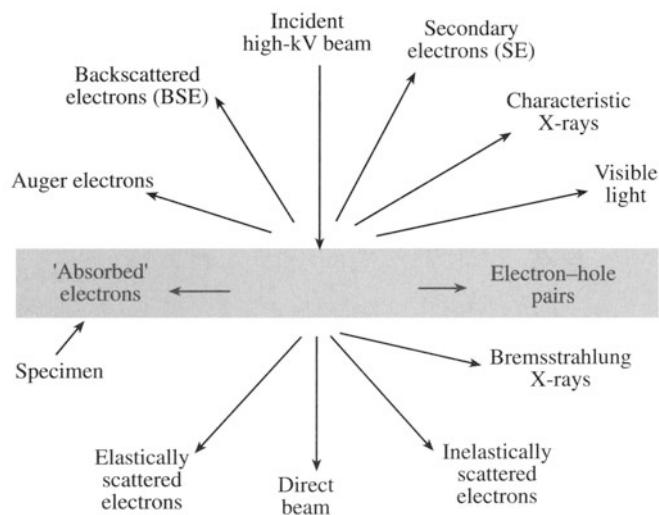
We'll see later that we are nowhere near building TEMs that approach this wavelength limit of resolution, because we can't make perfect electron lenses (see Chapter 6). But progress was rapid after Ruska's early work on lenses and, since the mid-1970s, many commercial TEMs have been capable of resolving 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

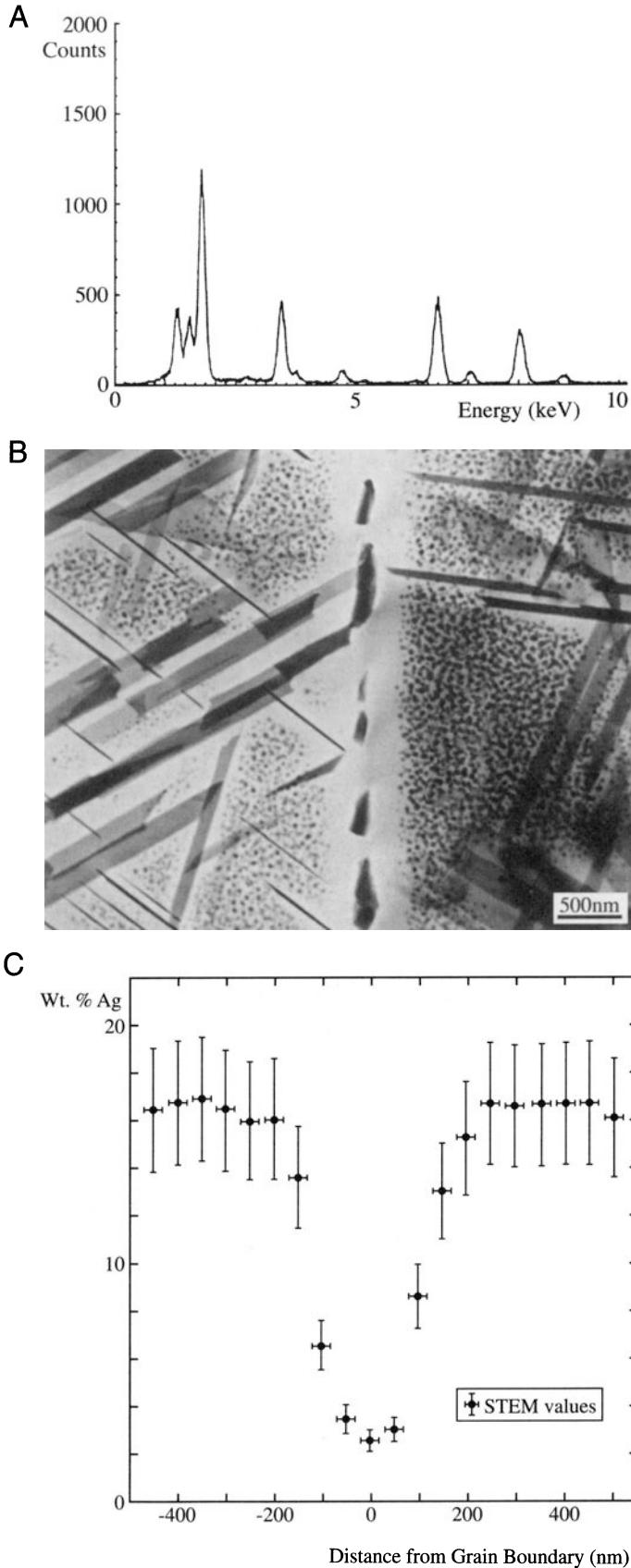
Electrons are one type of "ionizing radiation," which is the general term given to radiation that is capable of removing one of the tightly bound inner-shell electrons from the attractive field of the nucleus.

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.



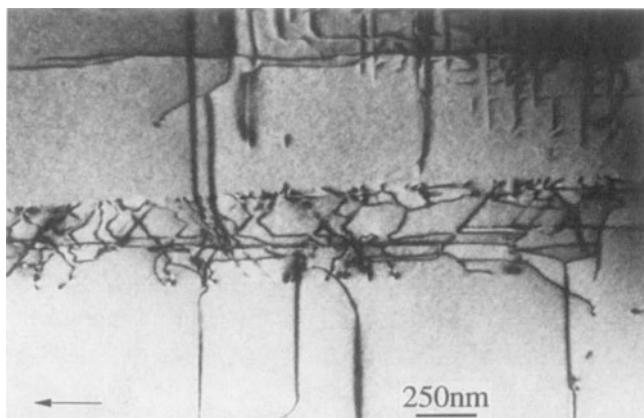
signal-generating instruments. To localize these signals we need to get our TEM to form a very fine electron beam, typically  $<10$  nm and at best  $<1$  nm in diameter. We accomplish this by combining TEM and scanning electron microscope (SEM) technology to create a scanning transmission electron microscope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

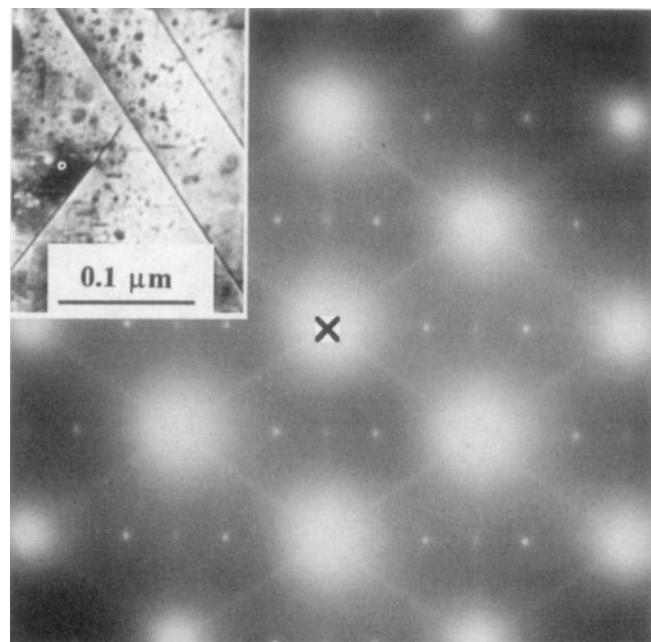
The depth of field of a microscope is a measure of how much of the *object* we are looking at remains “in focus” at the same time. Like the resolution, this property is governed by the lenses in the microscope. The best electron lens is not a very good one, as we’ve already mentioned, and has been compared 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.

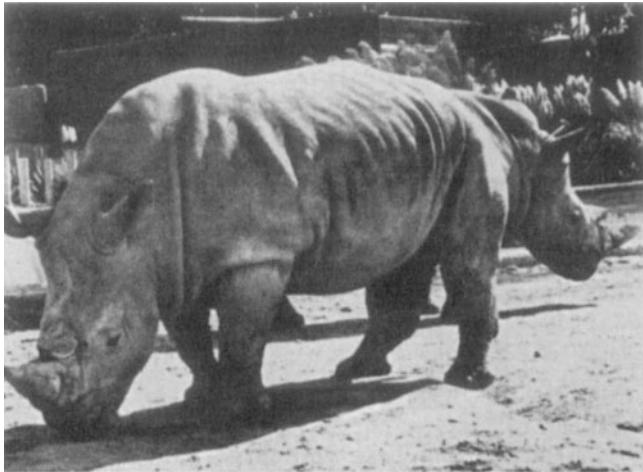
So an electron microscope can produce atomic level images, can generate a variety of signals telling you

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  $0.3 \text{ mm}^3$  of material! Extending that calculation to the present time at best doubles the volume to  $0.6 \text{ mm}^3$ . 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

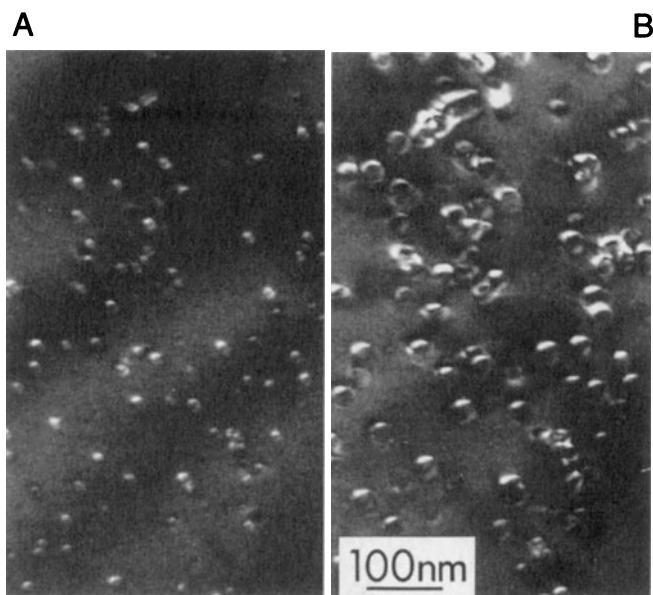
Another problem is that the TEM presents us with 2D images of 3D specimens, viewed in transmission. Our eyes and brain routinely understand reflected light images but are ill-equipped to interpret TEM images, and so we must be cautious. Hayes (1980) illustrates this problem well by showing a picture 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 back-scattering, 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  $\mu\text{m}$  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.

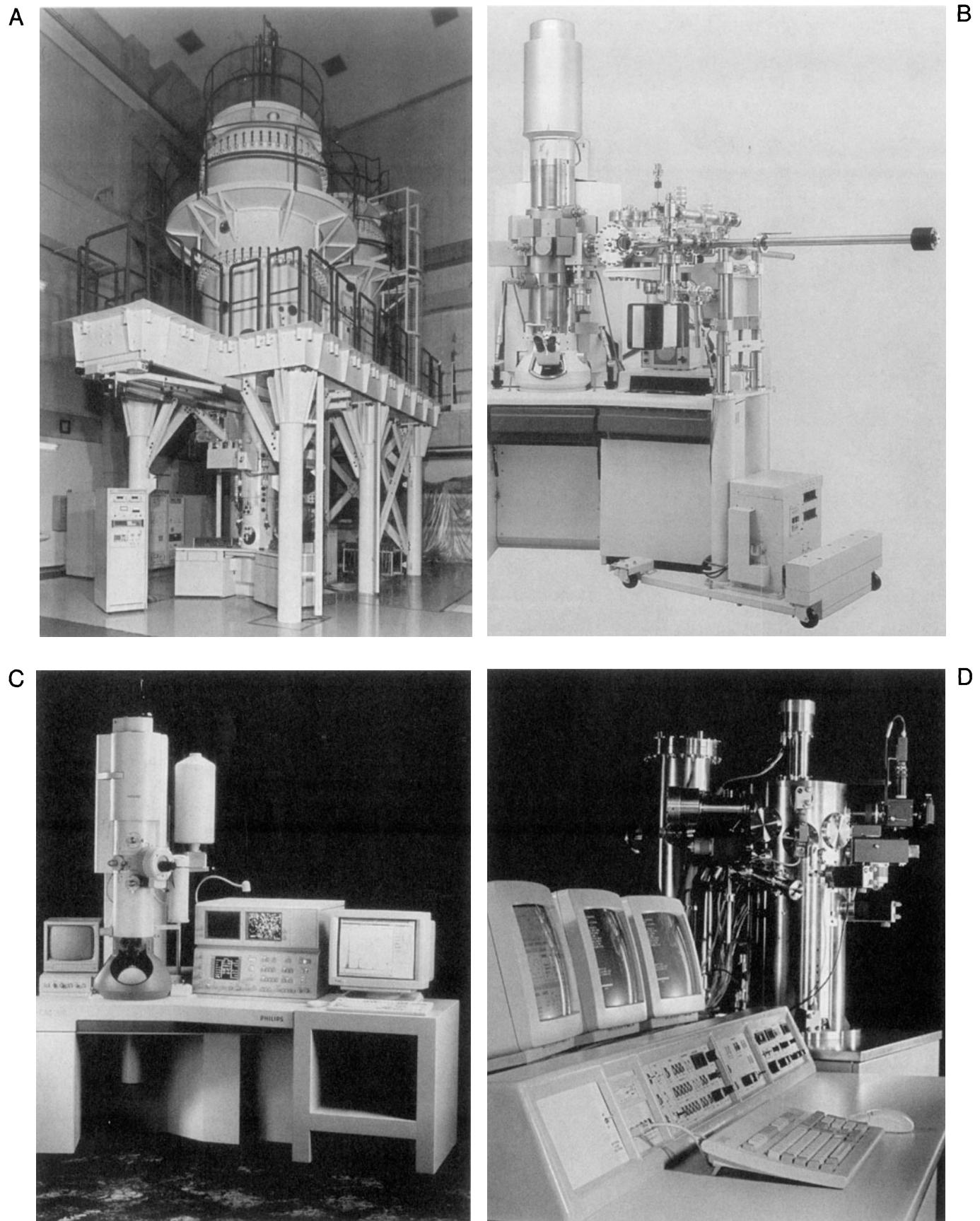
## 1.4. SOME FUNDAMENTAL PROPERTIES OF ELECTRONS

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  $\mu\text{A}$ , which corresponds to about  $10^{12}$  electrons passing through the specimen plane. But as we'll see below, with 100-keV energy, these electrons travel at about  $0.5c$  ( $1.6 \times 10^8 \text{ m/s}$ ), 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  $\lambda$  through Planck's constant; thus

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

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<sub>2</sub> 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 ( $e$ )	$(-)\ 1.602 \times 10^{-19}\text{ C}$
1 eV	$1.602 \times 10^{-19}\text{ J}$
Rest mass ( $m_0$ )	$9.109 \times 10^{-31}\text{ kg}$
Rest energy ( $m_0 c^2$ )	$511\text{ keV}$
Kinetic energy (charge $\times$ voltage)	$1.602 \times 10^{-19}\text{ N m}$ (for 1 volt potential)
Planck's constant ( $h$ )	$6.626 \times 10^{-34}\text{ N m s}$
1 ampere	$1\text{ C/sec}$
Speed of light in vacuum ( $c$ )	$2.998 \times 10^8\text{ m/sec}$

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

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

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

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

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

This expression is identical to equation 1.2. This relationship between  $\lambda$  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  $2.998 \times 10^8\text{ m/s.}$ ) So to be exact we must modify equation 1.6 to give

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

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 ( $\times m_0$ )	Velocity ( $\times 10^8\text{ m/s}$ )
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  $(hkl)$  distances, structure factors, and extinction distances, draw the microscope contrast transfer function, with  $(hkl)$  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/news-group.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.uhc.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](mailto: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  $63 \times 63$ . Available from the Internet by anonymous ftp from zippy.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 lo-

cally. 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  $3 \text{ nm} \times 3 \text{ nm}$  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.

Davison, 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

---

2.1. Why Are We Interested in Electron Scattering?	21
2.2. Terminology of Scattering	22
2.3. The Characteristics of Electron Scattering	23
2.4. The Interaction Cross Section	24
2.5. The Mean Free Path	25
2.6. The Differential Cross Section	25
2.7. Other Factors Affecting Scattering	26
2.8. Comparison to X-ray Diffraction	26
2.9. Fraunhofer and Fresnel Diffraction	27
2.10. Diffraction of Light from Slits	27
2.11. Coherent Interference	28
2.12. A Word about Angles	30
2.13. Electron Diffraction Patterns	30

## 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.

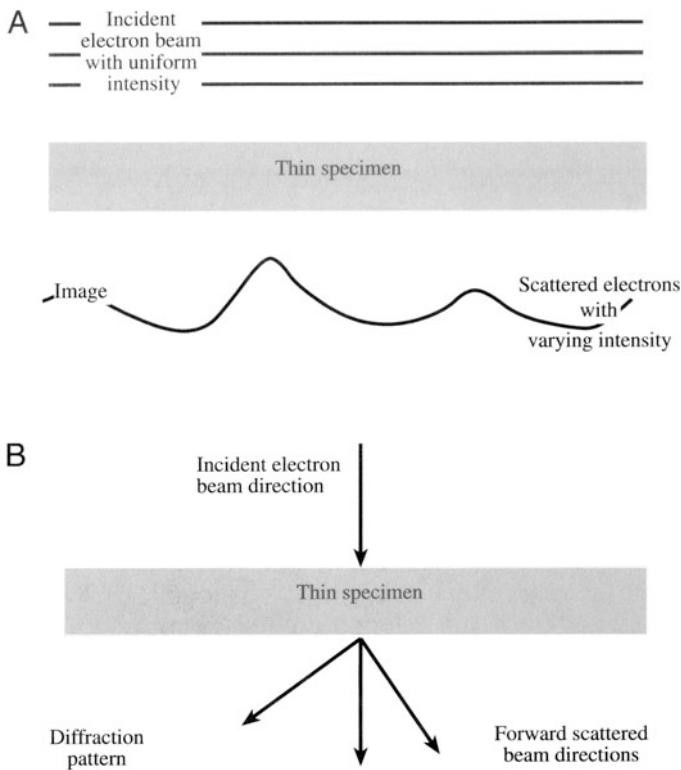
To give you some feel for the importance of electron scattering, it is worth illustrating at this stage the basic principles of the TEM. You will see in due course that in a TEM we illuminate a thin specimen with electrons in which the electron intensity is uniform over the illuminated area.

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.

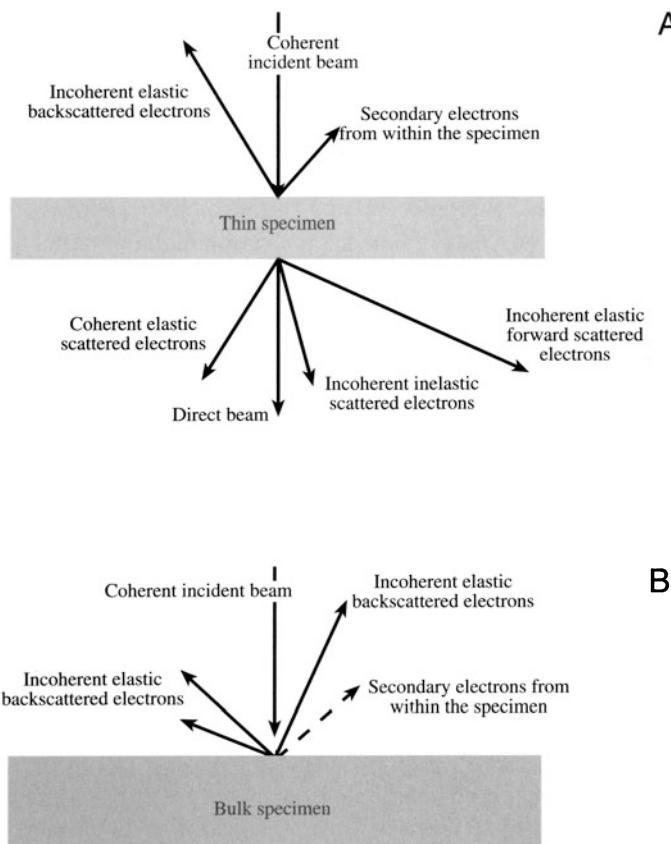
## 2.2. TERMINOLOGY OF SCATTERING

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  $< 90^\circ$  then it is forward scattered, and if  $> 90^\circ$  it is backscattered. These various terms are related by the following general principles, summarized in Figure 2.2.

- Elastic scattering is usually coherent, if the specimen is thin and crystalline.



**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 ( $1\text{--}10^\circ$ ), i.e., in the forward direction.
- At higher angles ( $>\sim 10^\circ$ ) elastic scattering becomes more incoherent.
- Inelastic scattering is almost always incoherent and relatively low angle ( $<1^\circ$ ) 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.

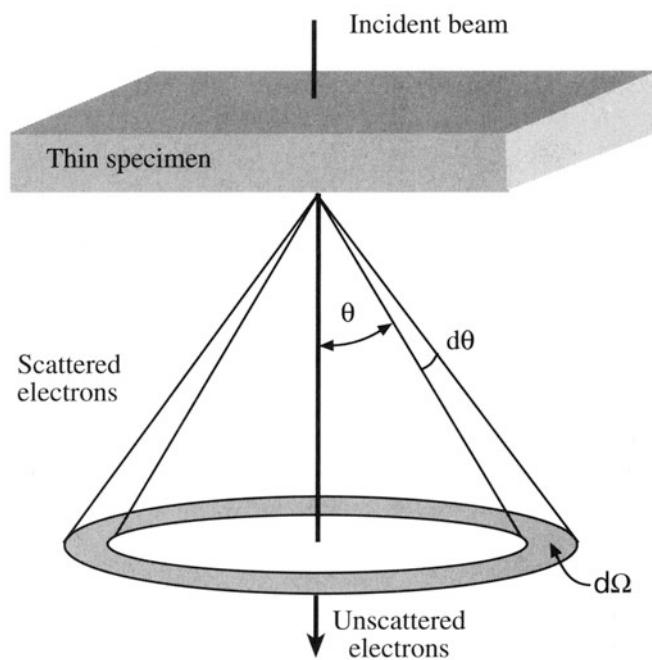
Forward scattering causes most of the signals used in the TEM.

Forward scattering includes elastic scattering, Bragg scattering, the events called diffraction, refraction, and in-

elastic 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  $\theta$  (radians) into some solid angle  $\Omega$  measured in steradians (sr). Often we assume that  $\theta$  is small enough such that  $\sin \theta \approx \tan \theta \approx \theta$ . When  $\theta$  is this small, it is often convenient to use milliradians or mrad; 1 mrad is  $0.0573^\circ$ , 10 mrad is  $\sim 0.5^\circ$ .



**Figure 2.3.** Electron scattering by a single isolated atom. The electrons are scattered through a semiangle  $\theta$  and the total solid angle of scattering is  $\Omega$ . 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 mrad.

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  $\sigma$ ) 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  $10^{-24} \text{ cm}^2$  [that's  $(10^4 \text{ \AA})^2$ ] 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 ( $\text{cm}^2$ ) which persists in the literature, although we can easily define the barn as  $10^{-28} \text{ m}^2$ . 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,  $\sigma_T$ , which is simply the sum of all elastic and inelastic scattering cross sections such that the total  $\sigma_T$  is

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

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

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

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

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

where  $V$  is the potential of the incoming electron, charge  $e$ , which is scattered through an angle greater than  $\theta$  by atoms of atomic number  $Z$ . At first sight this equation seems dimensionally incorrect, since  $Z$  and  $\theta$  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  $\text{cm}^{-1}$  or  $\text{m}^{-1}$ ) as

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

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  $\rho$  (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

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

The product of  $\rho$  and  $t$  is called the "mass-thickness" of the specimen (e.g., doubling  $\rho$  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  $r_{\text{elastic}}$  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  $\sigma$  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  $\sim 10^{-22} \text{ m}^2$  ( $\sim 10^{-18} \text{ cm}^2$ ). 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  $\sim 10^{-22} \text{ m}^2$  down to  $10^{-26} \text{ m}^2$  (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,  $\lambda$ . 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  $(\text{length})^{-1}$  there is a simple expression for the mean free path  $\lambda$  which has units of length

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

Typical values of  $\lambda$  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  $\lambda$  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  $t$

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

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  $\theta$  into a solid angle  $\Omega$  and there is a simple geometrical relationship between the  $\theta$  and  $\Omega$

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

and therefore

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

So the differential scattering cross section can be written as

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

Now, we can calculate  $\sigma$  for scattering into all angles which are greater than  $\theta$  by integrating equation 2.10. This yields

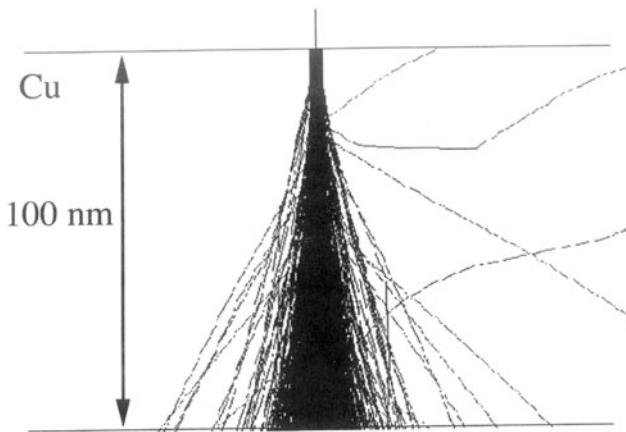
$$\sigma_\theta = \int_0^\pi d\sigma = 2\pi \int_0^\pi \frac{d\sigma}{d\Omega} \sin \theta \, d\theta \quad [2.11]$$

The limits of the integration are governed by the fact that the values of  $\theta$  can vary from 0 to  $\pi$ , depending on the specific type of scattering. If we work out the integral we find that  $\sigma$  decreases as  $\theta$  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  $\sigma$  for an atom in the specimen:  $\sigma$  for all values of  $\theta$  is simply the integral from 0 to  $\pi$ . 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.

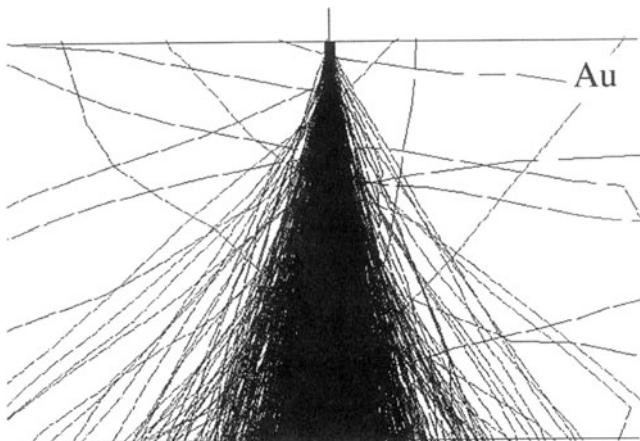
Our knowledge of the values of  $\sigma$  and  $\lambda$  is very sketchy, particularly at the 100–400 keV beam energies used in TEMs. Cross sections and mean free paths for particular scattering events may only be known within a factor of two, but we can often measure  $\theta$  very precisely in the 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!

A



B



**Figure 2.4.** Monte Carlo simulation of the paths followed by 10<sup>3</sup> 100-kV 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  $\theta$ ), you are changing the effective scattering cross

section ( $\sigma_\theta$ ), 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  $\pm 5^\circ$  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

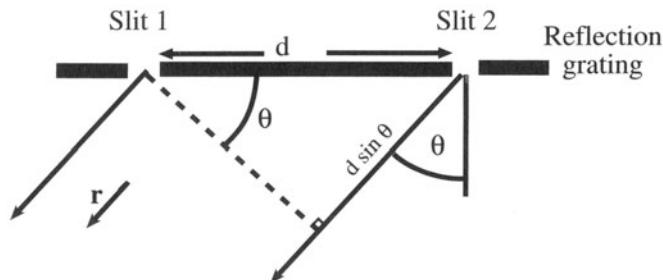
acts as a new source for a spherical wavelet. The wavelets interfere with one another to give the new wavefront and the process is repeated.

## 2.10. DIFFRACTION OF LIGHT FROM SLITS

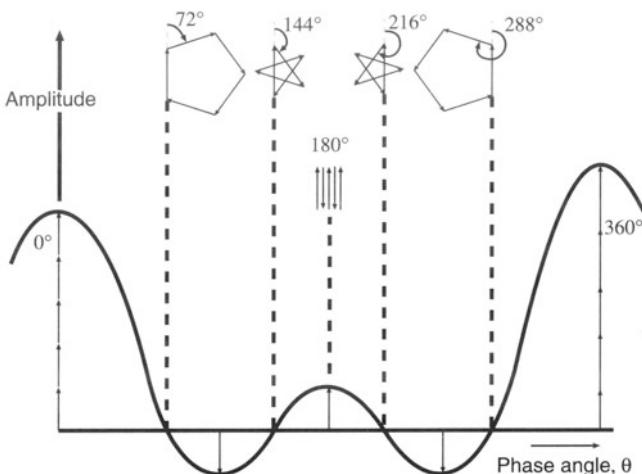
When we place a pair of very narrow slits in front of a wavefront, we select just two of the Huygens wavelets; these wavelets then must have the same phase at the slits. As they propagate past the slits, their phases differ, depending on the position of the detector. The important term is the path difference  $L = d \sin \theta$  as shown in Figure 2.5. The two wavelets propagating in direction  $\mathbf{r}$  are out of phase by  $2\pi L/\lambda$ . If  $d$  and  $\lambda$  are such that this phase difference is actually a multiple of  $2\pi$ , then the rays are again in phase. Therefore, there is an inverse relationship between  $d$  and  $\theta$  for a given  $d$ ; as  $d$  decreases,  $\sin \theta$  increases.

The inverse relationship between  $d$  and  $\theta$  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  $\theta$  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  $\theta$  increases the rays become out of phase, but the phasors can still add to give a large resultant vector. When  $\theta$  is exactly  $72^\circ$  ( $360^\circ/5$  for five slits), the phasor diagram is a closed pentagon and the resultant amplitude is zero. This process repeats at  $144^\circ$  ( $2 \times 360^\circ/5$ ) and  $216^\circ$  ( $3 \times 360^\circ/5$ ). In between these values at  $108^\circ$  ( $1.5 \times 360^\circ/5$ ) we produce a local maximum in amplitude which is repeated at  $180^\circ$  ( $2.5 \times 360^\circ/5$ ). If we plot the amplitude as a function of  $\theta$ , 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  $\phi$  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  $\theta$  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  $\lambda$  (increasing the accelerating voltage) will improve resolution: as  $\lambda$  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  $\lambda/w$  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

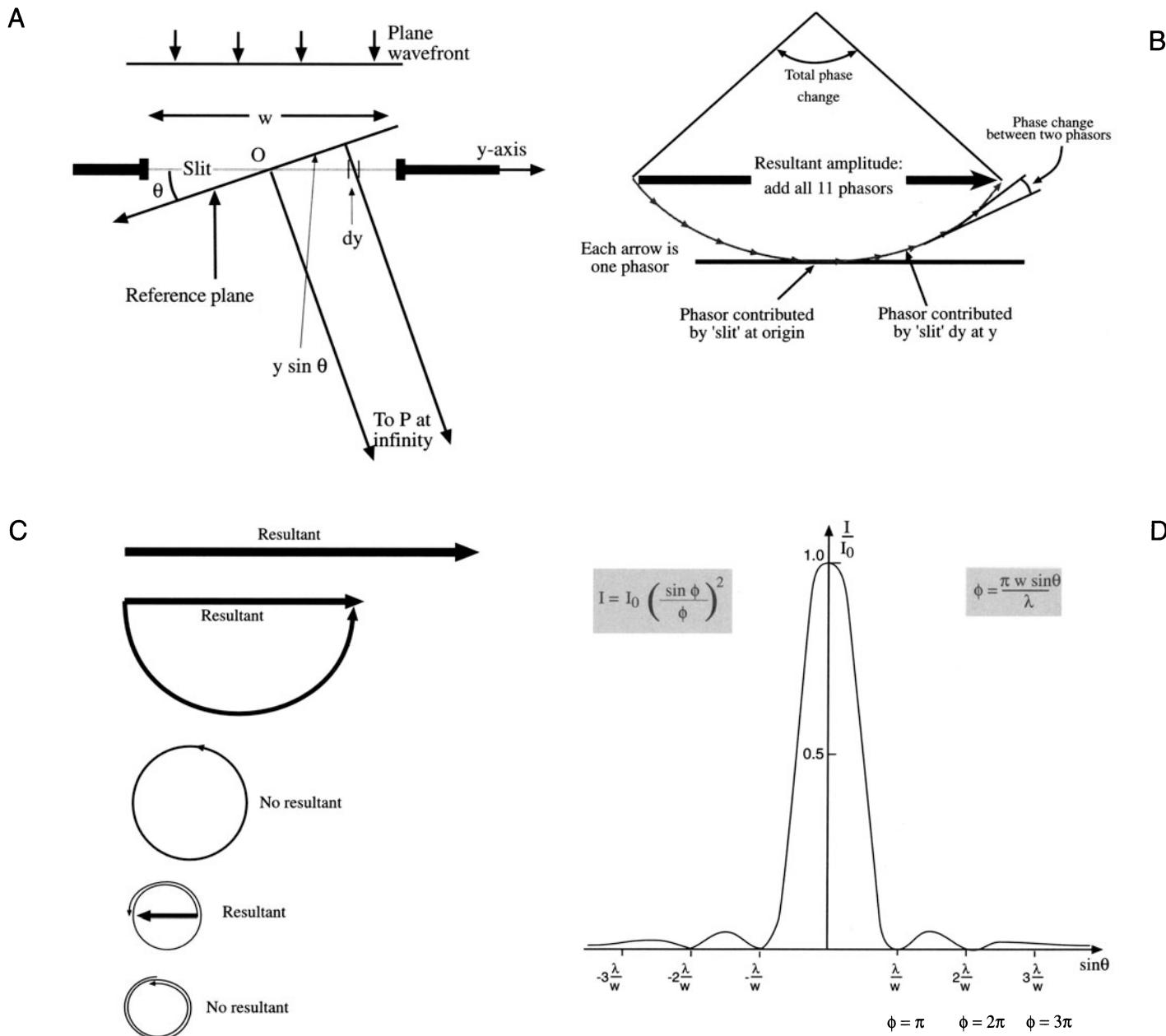
To expand on this point, consider an infinite plane wave described by the usual characteristics of amplitude and phase. We can describe the wave function for this wave by the expression

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

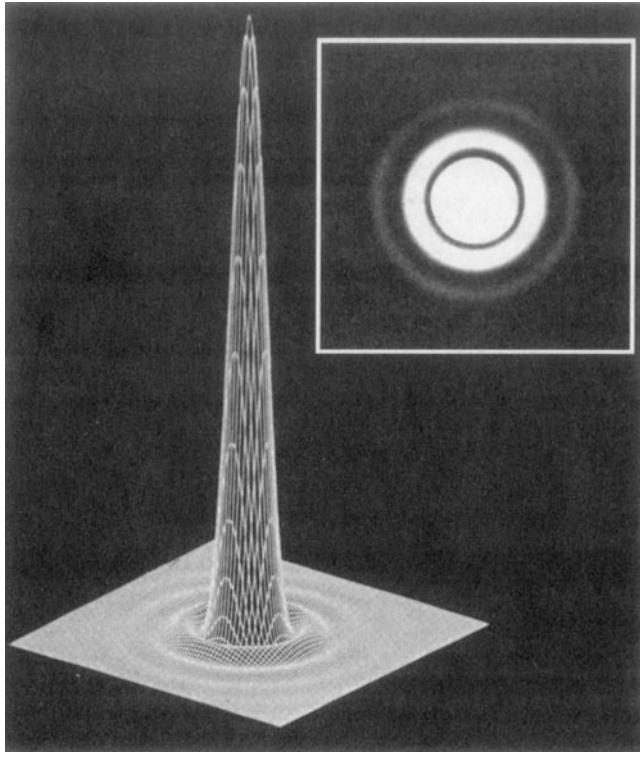
where  $\psi_0$  is the amplitude and  $\phi$  the phase of the wave. The phase depends on position  $x$ , such that if  $x$  changes by one wavelength  $\lambda$ , the phase difference is  $2\pi$ . Stated another way, the phase difference  $\Delta\phi$  between any two monochromatic (same wavelength) waves is related to the path difference  $\Delta x$  they must travel in going from source to detector. The relationship is

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

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  $2\pi$ . 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  $\theta$  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).



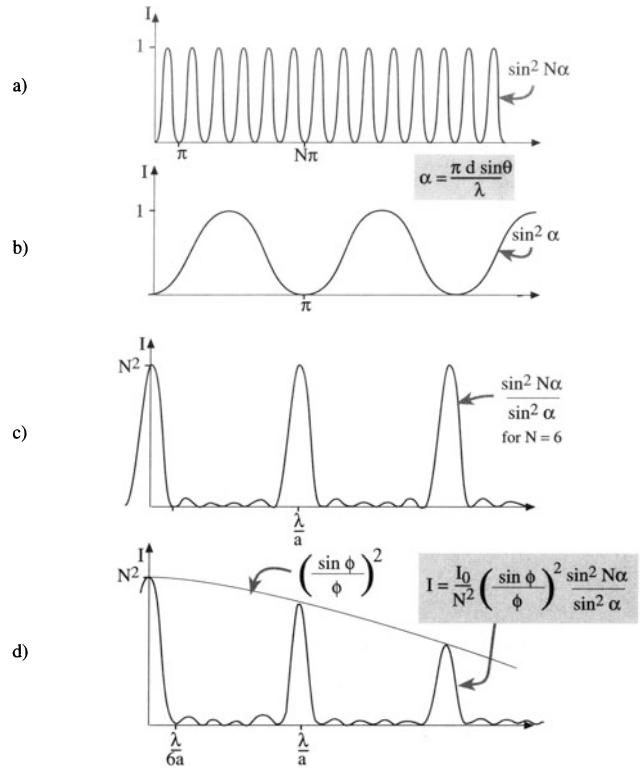
**Figure 2.8.** The visible-light intensity produced by a 0.5-mm-diameter circular aperture and the observed Airy rings (inset).

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  $\alpha$ , 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  $\beta$ .
- We will define all scattering semiangles controlled by the specimen as  $\theta$ . This may be a specific angle, such as twice the Bragg angle,

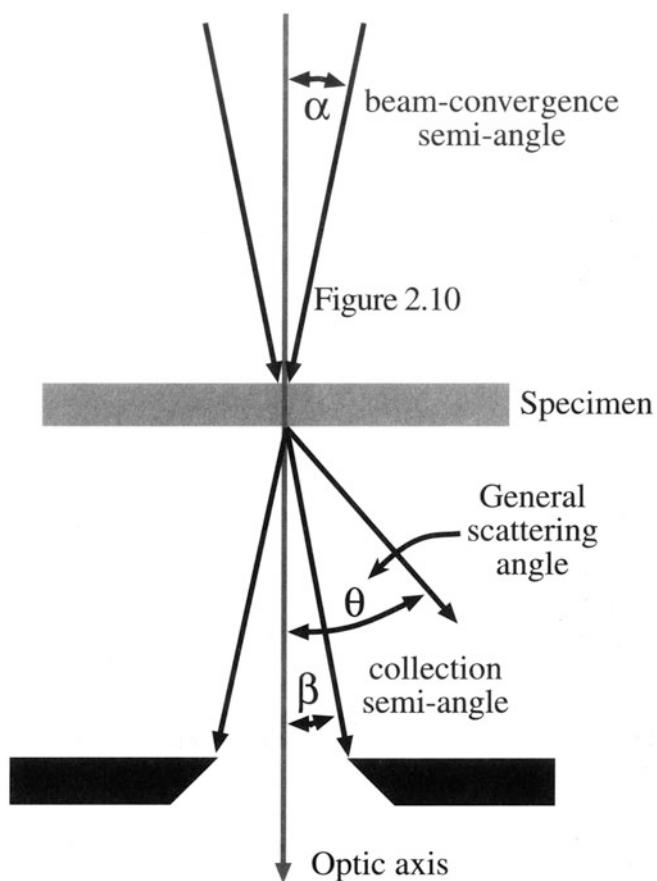


**Figure 2.9.** The scattered intensity from  $N$  slits (shown here for  $N = 6$ ) where each slit would give the intensity shown in (B). (C) is the curve in (A) divided by curve in (B) and (D) is the curve in (C) multiplied by the curve in (Figure 2.7D). The distance  $d$ , the separation of the slits,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

where  $\theta = 2\theta_B$  (see Section 11.4), or a general scattering semiangle  $\theta$ .

## 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  $\alpha$ ; any collection semiangle is  $\beta$  and general scattering semiangles are  $\theta$ .

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  $\theta$  (increasing distance from the direct beam), which reflects the decrease in the scattering cross section with  $\theta$ . Third, the scattering intensity varies strongly with the structure of the specimen. You'll see much more of this in Part II.

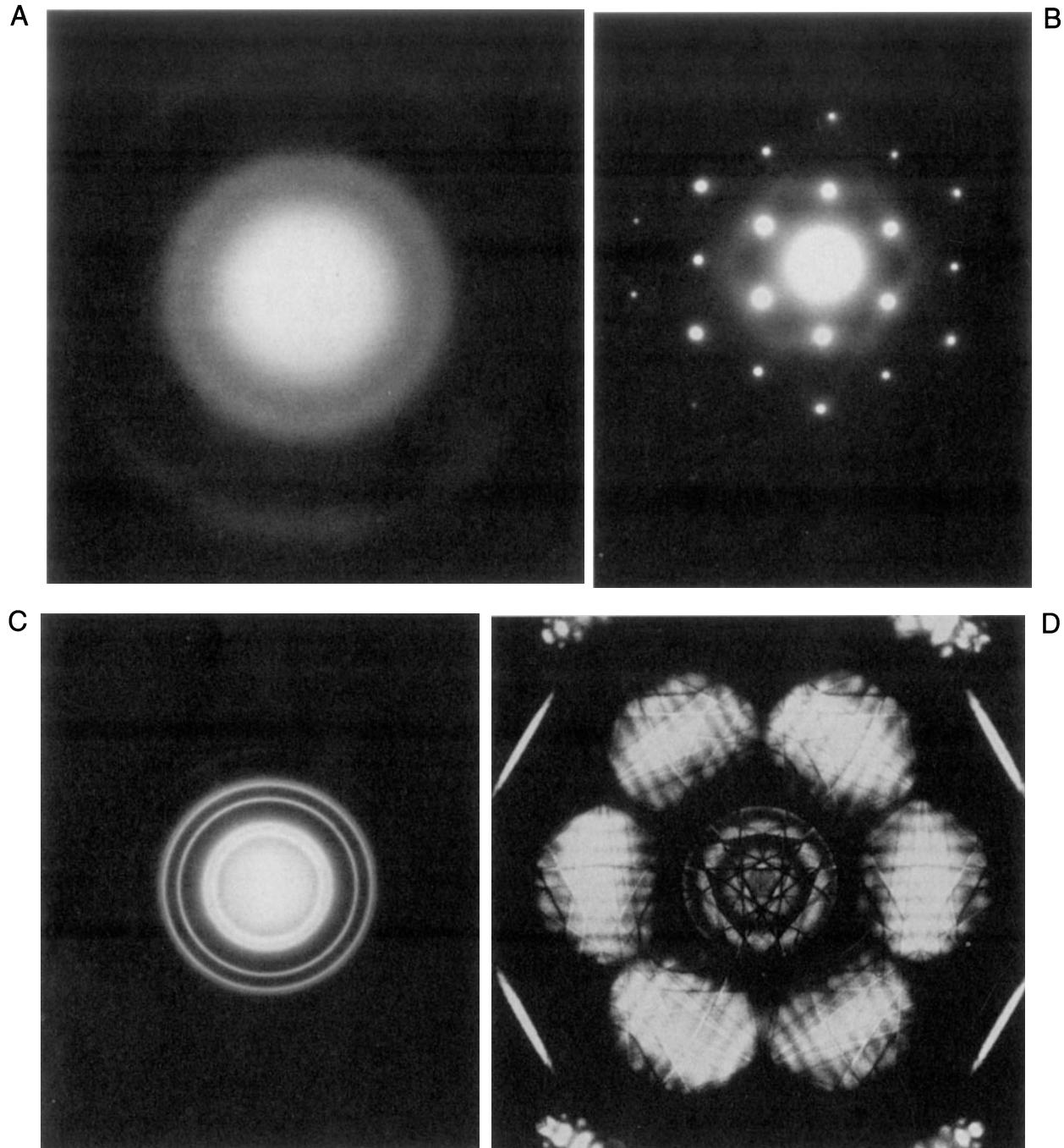
So far, in fact, we've only considered the amplitude of the electron wave and we've neglected the phase. When a wave is scattered, it will change its phase with respect to the incident wave. This is because a wave cannot change direction and remain in step with a wave that is unscattered. The phase of 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:

$\sigma$	the scattering cross section of one atom
$\lambda$	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.

# 3

# Elastic Scattering

---

3.1. Particles and Waves .....	37
3.2. Mechanisms of Elastic Scattering .....	37
3.3. Scatter from Isolated Atoms .....	38
3.4. The Rutherford Cross Section .....	39
3.5. Modifications to the Rutherford Cross Section .....	39
3.6. Coherency of the Rutherford-Scattered Electrons .....	40
3.7. The Atomic Scattering Factor .....	41
3.8. The Origin of $f(\theta)$ .....	42
3.9. The Structure Factor $F(\theta)$ .....	43
3.10. Simple Diffraction Concepts .....	44
3.10.A. Interference of Electron Waves .....	44
3.10.B. Diffraction Equations .....	45

## 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  $180^\circ$ . 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 ( $\theta$ ). 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  $\sim 3^\circ$  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

## 3.1. 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  $\alpha$  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):

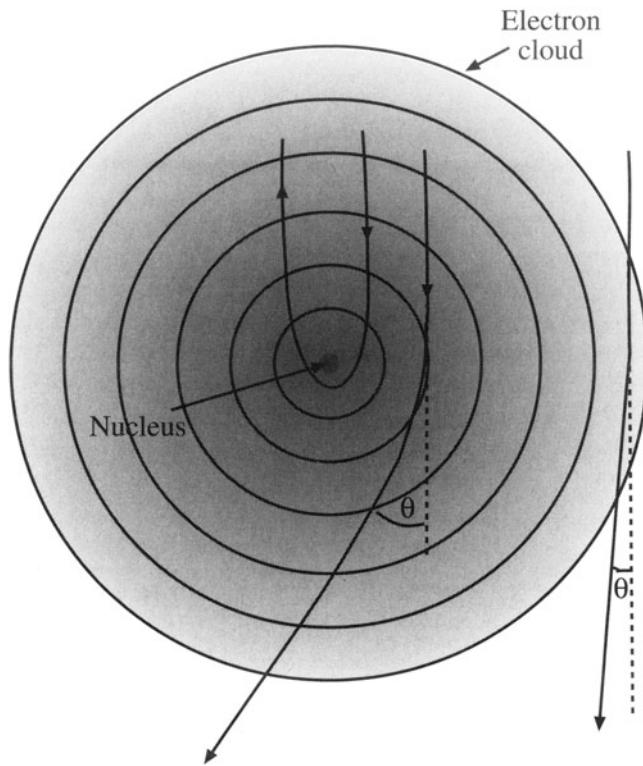
*An interaction between a wave of any kind and an object of any kind.*

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

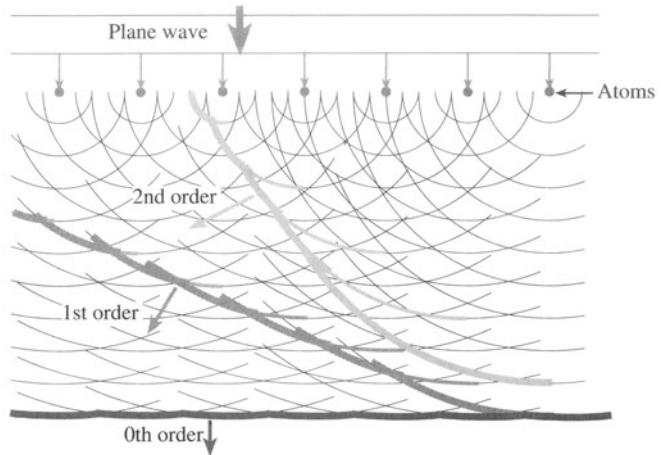


**Figure 3.1.** Two mechanisms by which a high-energy electron is scattered by an isolated atom. Coulombic interaction within the electron cloud results in low-angle ( $\theta$ ) scatter while Coulombic attraction by the nucleus causes high  $\theta$  scatter and perhaps complete backscatter. The potential 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  $180^\circ$  (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  $\theta$ .

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)

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

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

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  $\theta$  will be large. We'll see in Chapter 22 that this dependence on  $\theta$  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  $\alpha$  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

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

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  $\pi$  to obtain the total elastic nuclear cross section (in scattering events per electron per atom per m<sup>2</sup>) in a more accurate form than that given in equation 3.2

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

Again we see that the beam energy ( $E_0$ ), the angle of scattering ( $\theta$ ), 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  $t$

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

Notice that we still have the mass-thickness dependence,  $\rho t$ , 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  $< 5^\circ$ ) 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  $\theta_0$  is called the screening parameter given by

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

(Here  $E_0$  is in keV.) What we are saying is that the screening parameter can be described by a particular scattering angle,  $\theta_0$ . When the scattering angle is greater than  $\theta_0$ , we can neglect electron-electron interactions and the nuclear interaction is dominant. The value of  $\theta_0$  at 100 keV is only  $\sim 2^\circ$  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  $\lambda_R$ , the relativistically corrected wavelength (see equation 1.7), and  $a_0$ , the Bohr radius of the scattering atom

$$a_0 = \frac{\hbar^2 \epsilon_0}{\pi m_0 e^2} \quad [3.7]$$

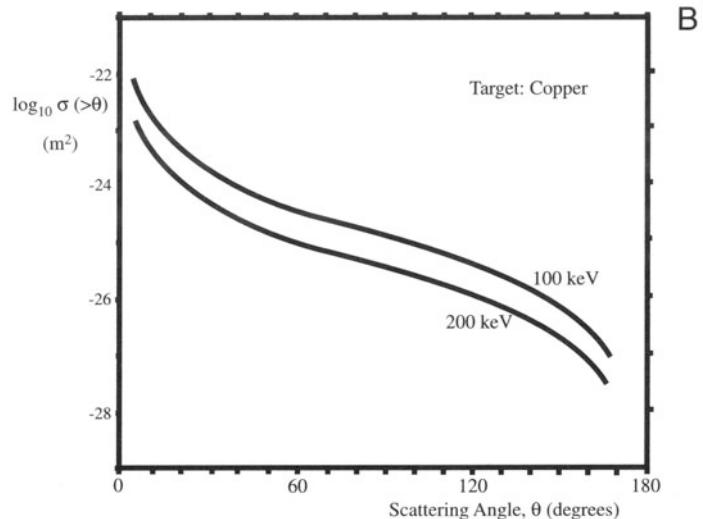
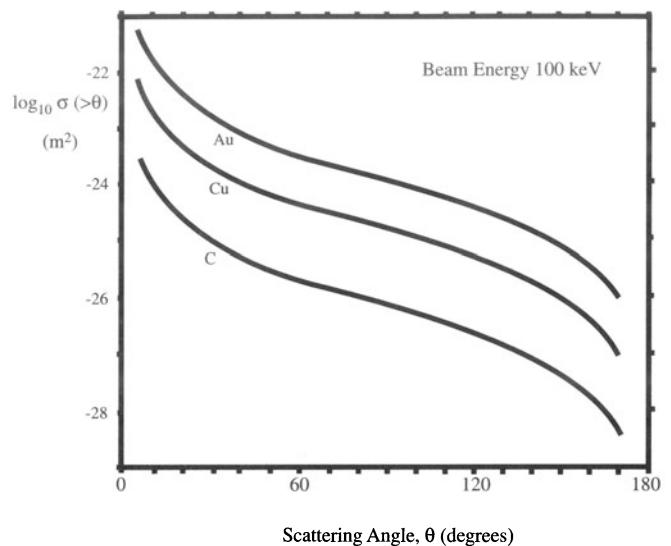
where  $\epsilon_0$  is the dielectric constant. Using the other constants listed in Table 1.1 we find  $a_0$  is 0.0529 nm (remember this as 0.5 Å). The result is

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

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  $\sim 10^6$  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 ( $\theta$  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 scat-

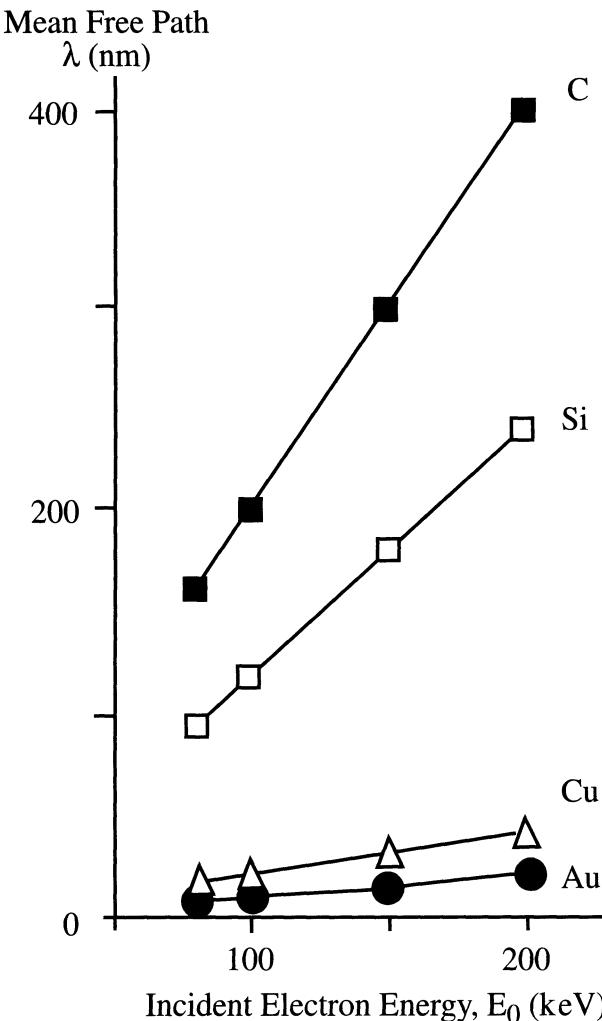


**Figure 3.3.** The variation of the logarithm of the screened relativistic Rutherford cross section with scattering angle: (A) for different elements at 100 keV and (B) for Cu at different TEM voltages.

tering. 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.

### 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 spec-

imen 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

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

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

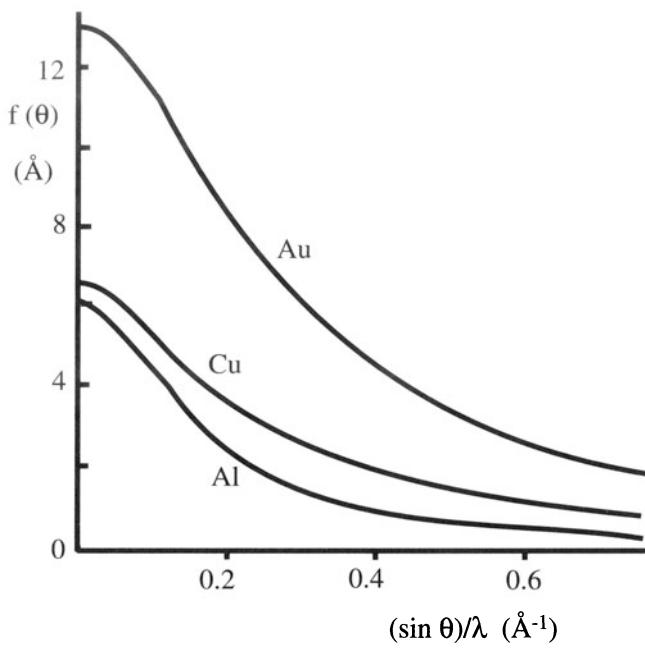
- $f(\theta)$  is a measure of the amplitude of an electron wave scattered from an isolated atom.
- $|f(\theta)|^2$  is proportional to the scattered intensity.

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

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

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  $\lambda$  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  $\theta$  (calculated from equation 3.10) showing that elastic scattering decreases with angle away from the incident beam direction ( $\theta = 0^\circ$ ) 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  $\lambda$ ,  $\theta$ , 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  $\theta$  increases ( $\theta = 0^\circ$  for the incident beam direction).
- It decreases as  $\lambda$  decreases (i.e., as the accelerating voltage increases).
- It increases with  $Z$  for any value of  $\theta$ .

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  $\theta$ .

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 $f(\theta)$

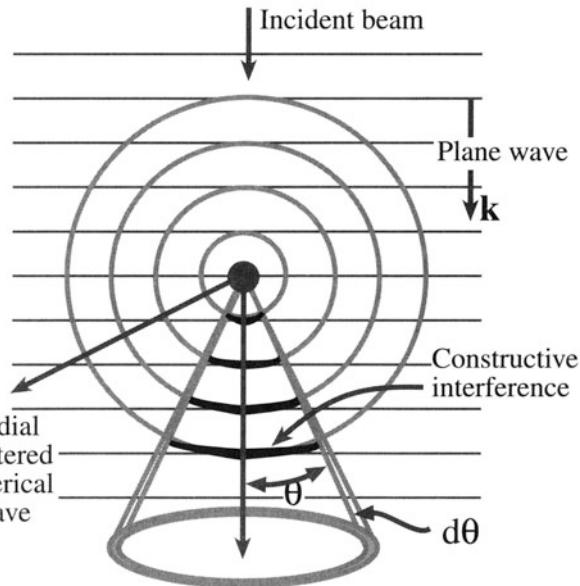
Since  $f(\theta)$  relates to the amplitude of a scattered wave, we will briefly consider how it arises. The following analysis is not intended to be completely rigorous, but only to give the fundamental ideas behind the meaning of  $f(\theta)$  and its relation to the differential scattering cross 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 be described as a wave of amplitude  $\psi$  and phase  $2\pi kr$

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

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  $\lambda$ ). The in-phase constructive interference between the plane and spherical waves is shown by the dark arcs. The angles  $\theta$  and  $d\theta$  are the same as in Figure 2.3.

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

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

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

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

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

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

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

**First aside:** In writing equation 3.13, we have introduced two wave propagation parameters: the vector  $\mathbf{k}_I$  for the incident plane wave and the scalar  $k$  for the spherical scattered wavelet. By writing the  $2\pi$  factor separately as part of the phase term, we have implicitly defined  $k$  to be  $1/\lambda$ . Many physics textbooks include the  $2\pi$  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  $\psi_0 \sin(2\pi kr)$ , then after it has passed through the specimen it will be  $\psi_{tot}$ ; after scattering, the phase is increased by  $\phi$ . We can express the new  $\psi_{tot}$  as

$$\begin{aligned} \psi_{tot} &= \psi_0 \sin(2\pi kz + \phi) = \psi_0 \sin(2\pi kz) \cos \phi \\ &\quad + \psi_0 \cos(2\pi kz) \sin \phi \end{aligned} \quad [3.15]$$

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

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

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

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

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

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  $(hkl)$ . The scattering angle  $\theta$  is the angle between the incident and scattered electron beams. So we can write

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

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  $(hkl)$  that make up the crystal structure. None of this is very surprising, but it turns out that this equation predicts that in certain circumstances the amplitude of 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.

### 3.9. 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

### 3.10. SIMPLE DIFFRACTION CONCEPTS

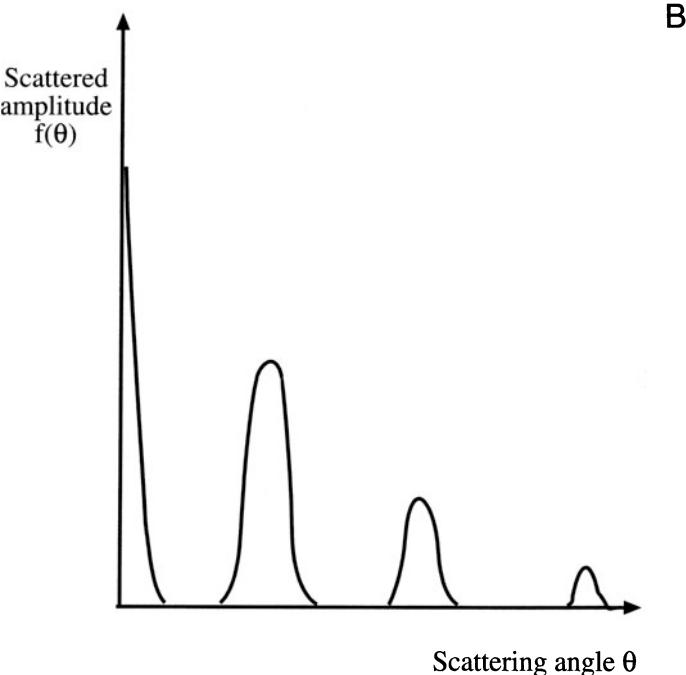
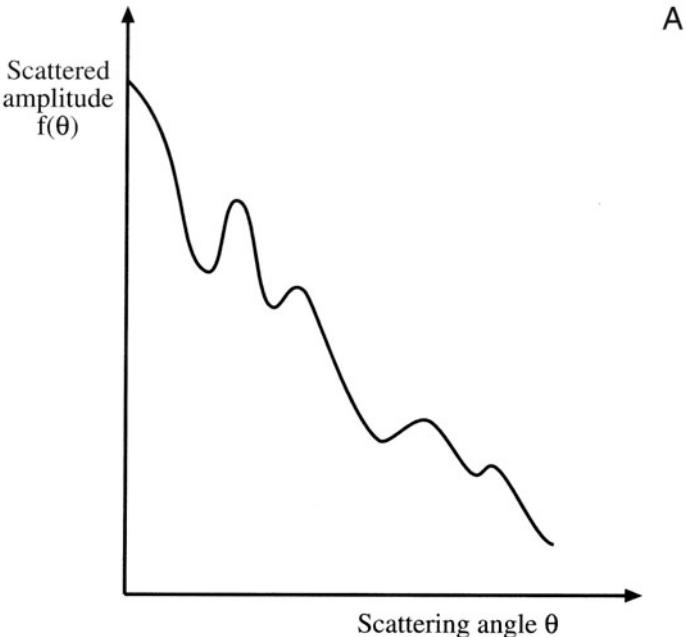
As we mentioned earlier, electron diffraction is by far the most important scattering phenomenon in the TEM. The importance arises because, as we'll show you in Chapters 11 and 12, we can use diffraction to determine the spacing of planes in crystals. The interplanar spacings in different crystal structures 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  $\theta$  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.

#### 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



**Figure 3.7.** Change in  $f(\theta)$  with  $\theta$  (A) for an amorphous specimen and (B) for a crystalline specimen. The amplitude (and therefore the intensity) of scatter generally decreases with increasing  $\theta$ , 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).

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  $\lambda$ ) is advancing toward the centers. Each slit acts as a new source of a wave of the same  $\lambda$ . 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.

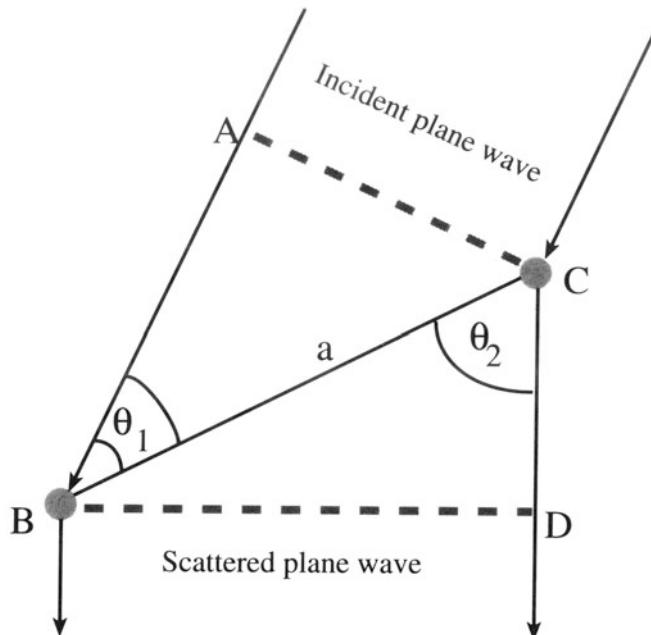
### 3.10.B. Diffraction Equations

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<sub>4</sub> 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  $\lambda$ ) makes an angle  $\theta_1$  with the line connecting the scattering centers, and is diffracted at an angle  $\theta_2$ , then the path difference (AB – CD) is then

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

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



**Figure 3.8.** The approach used by von Laue to calculate the path difference for a wave (wavelength  $\lambda$ ). In this one-dimensional figure the wave is incident at angle  $\theta_1$  and scattered at angle  $\theta_2$  from two atoms (B and C) spaced distance  $a$  apart. The path difference between scattered waves is AB – CD.

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

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

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  $hkl$  are the indices of the diffracted beam and are equivalent to the Miller indices ( $hkl$ ) of a crystal plane, or some multiple thereof.

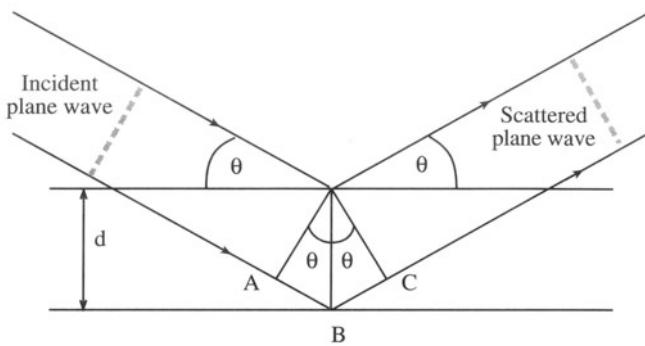
Usually in TEM, we use a simpler approach to describe diffraction. Von Laue's approach was simplified by Bragg (the elder) in England who argued that the waves behaved as if they were reflected off atomic planes as shown in Figure 3.9. Bragg showed that waves reflected off adjacent scattering centers 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"  $hkl$  planes are spaced a distance  $d$  apart and the wave is incident and reflected at an angle  $\theta_B$ , 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

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

We'll reserve  $\theta_B$  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  $\lambda$  for the inci-



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

dent electron and we can measure  $\theta$  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.

## 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  $\sigma$ , there is an angle involved in any  $\sigma$ .

- 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,  $\theta$ .
- $F(\theta)$  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. A* **24**, 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.

# 4

# Inelastic Scattering and Beam Damage

---

4.1. Which Inelastic Processes Occur in the TEM .....	51
4.2. X-ray Emission .....	51
4.2.A. Characteristic X-rays .....	52
4.2.B. Bremsstrahlung X-rays .....	56
4.3. Secondary Electron Emission .....	57
4.3.A. Slow Secondary Electrons .....	57
4.3.B. Fast Secondary Electrons .....	58
4.3.C. Auger Electrons .....	58
4.4. Electron–Hole Pairs and Cathodoluminescence (CL) .....	59
4.5. Plasmons and Phonons .....	60
4.6. Beam Damage .....	61
4.6.A. Electron Dose .....	62
4.6.B. Specimen Heating .....	62
4.6.C. Beam Damage in Polymers .....	62
4.6.D. Beam Damage in Covalent and Ionic Crystals .....	63
4.6.E. Beam Damage in Metals .....	63
4.6.F. Sputtering .....	65

## 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

## 4.1. 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.

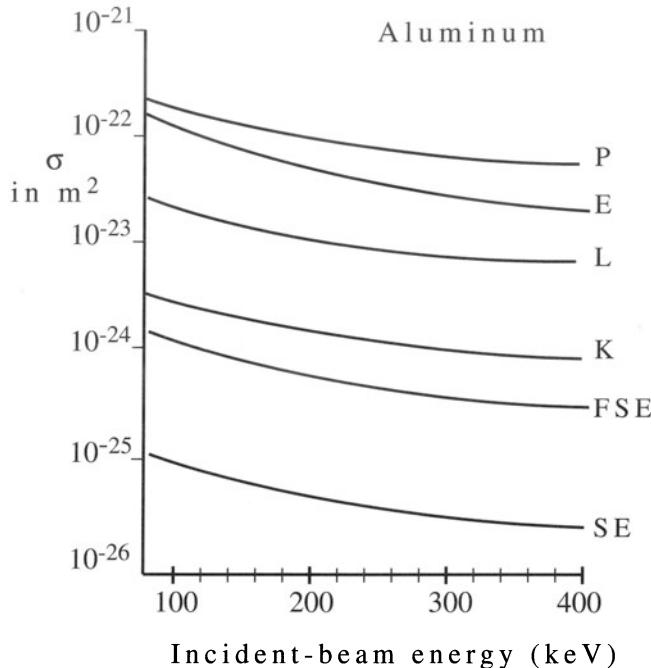
This range of inelastic scattering produces a range of scattering angles, but there is no simple relationship between the energy lost and the scattering angle. We'll separate the inelastic processes into three components:

- 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 ( $\theta \sim 0^\circ$ ); 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

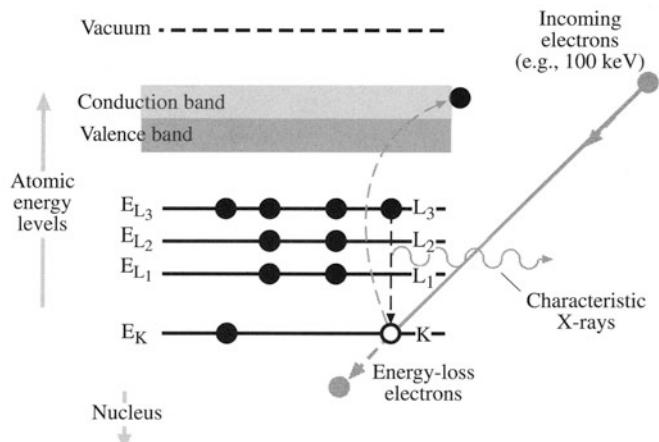
How do we produce characteristic X-rays and of what are they “characteristic”? First of all, a high-energy beam electron must penetrate through the outer electron shells and interact with the inner-shell (or core) electrons. If more than a critical amount of energy is transferred to an inner-shell electron, 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  $K_\alpha$  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  $K_{\alpha}$  X-ray, but if we fill it from the M shell we get a  $K_{\beta}$  X-ray. If the hole is in the L shell and we fill it from the M shell we get an  $L_{\alpha}$  X-ray, and if we fill it from the N shell we get an  $L_{\beta}$  X-ray. The notation is in fact more complex because we differentiate the  $\alpha$  X-rays in terms of  $\alpha_1$  and  $\alpha_2$ , depending from which subshell of the outer shell the electron falls to fill the hole. The  $\alpha_1$  X-ray is from the outermost subshell (e.g., the  $L_{III}$  or  $M_V$ ), the  $\alpha_2$  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  $\alpha$  and  $\beta$  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 in-

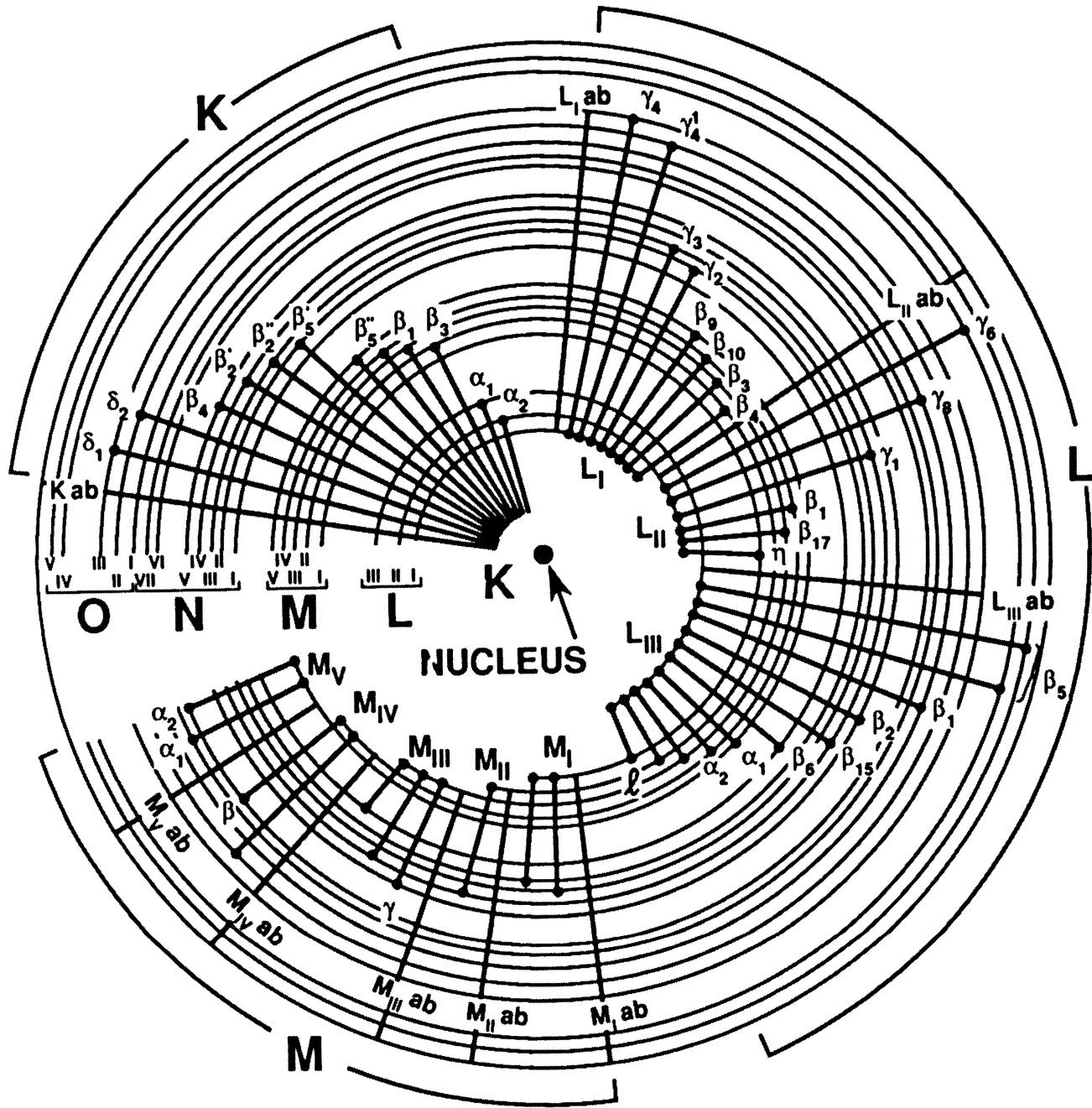
tense lines, usually the  $\alpha$  lines (or, if we can't resolve them, we use the  $\alpha$  and  $\beta$  lines). This will become more obvious when you've learned about X-ray qualitative analysis in Chapter 34.

**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 ( $\sigma$ ) 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  $\geq 100$  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

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

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 ( $\sim 10$  mrad).
  - The resultant characteristic X-ray is emitted uniformly over  $4\pi$  sr.

As with the Rutherford cross section, the simple Bethe expression needs to be corrected for the effect of rel-

activity at TEM electron energies, and this means substituting the term  $m_0 v^2/2$  for the beam energy and introducing a standard relativistic factor,  $\beta (=v/c)$

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

**Table 4.1. Weights of Lines (Approximate)<sup>a</sup>**

$K_\alpha$	(1)	$K_\beta$	(0.1)		
$L_{\alpha 1,2}$	(1)	$L_{\beta 1}$	(0.7)	$L_{\beta 2}$	(0.2)
$L_{\gamma 1}$	(0.08)	$L_{\gamma \beta}$	(0.03)	$L_\epsilon$	(0.04)
$M_\eta$	(0.01)				
$M_\alpha$	(1)	$M_\beta$	(0.6)	$M_\zeta$	(0.06)
$M_\gamma$	(0.05)	$M_{II}N_{IV}$	(0.01)		

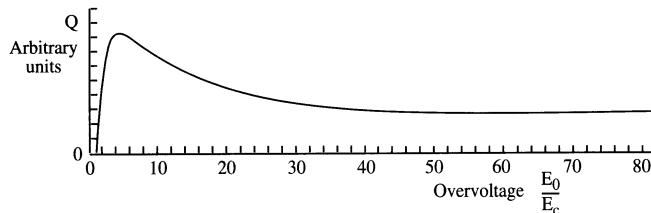
<sup>a</sup>The weights are given in parentheses.

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  $\lambda$ . 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  $\nu$  is the frequency, and in



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

order to express this energy in electron volts we equate it to  $E$ , where  $E$  is the X-ray energy.

Thus

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

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

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

where  $\lambda$  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

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

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.

**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 $\alpha$  X-ray (8.048 keV), then an L $\alpha$  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,  $\omega$ , 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  $\omega$  gives

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

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 ( $Z = 6$ ),  $\omega$  is  $\sim 10^{-3}$  and, for Ge ( $Z = 32$ ),  $\omega$  is  $\sim 0.5$ . This means you have to ionize 1000 carbon atoms before you get a single C K $\alpha$  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).

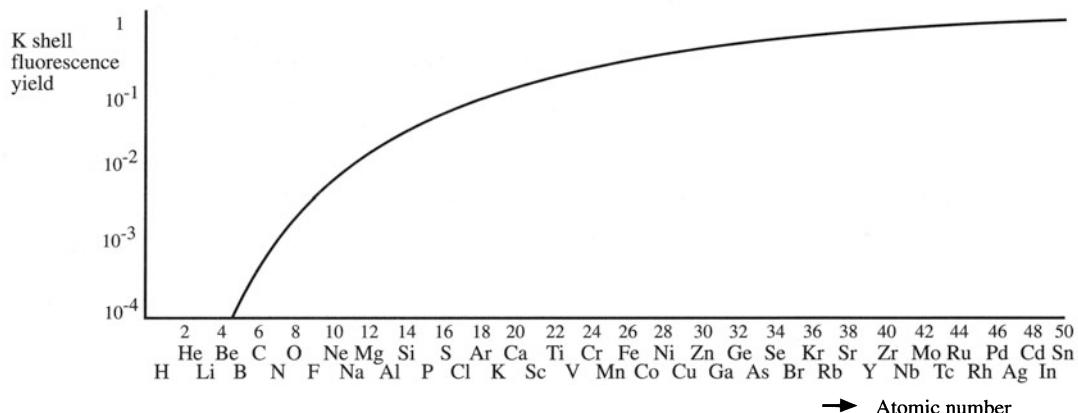
## 4.2.B. Bremsstrahlung X-rays

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."

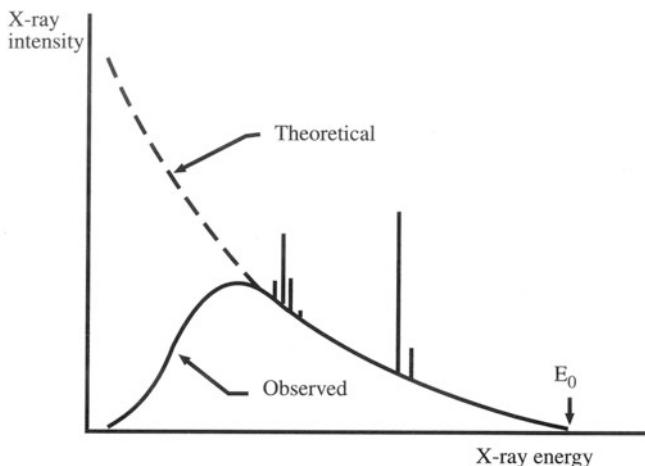
The likelihood of bremsstrahlung creation is usually described by the cross section derived by Kramers (1923). This expression is often used for thin TEM specimens, although it was originally derived for bulk specimens. It is common to use the Kramers cross section to predict the bremsstrahlung production rather than the probability of interaction. Thus

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

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  $\sim 50$  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

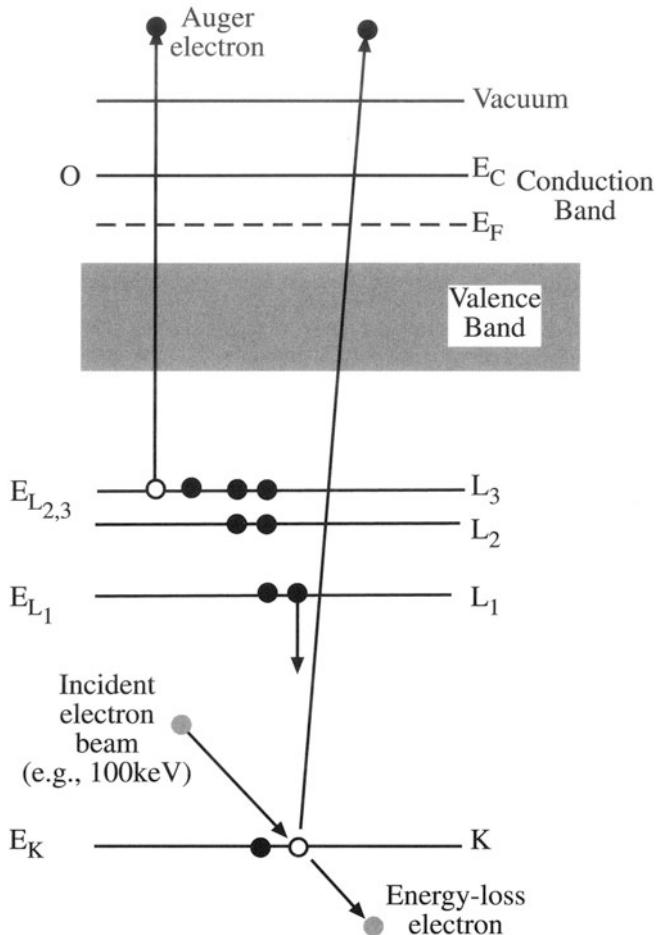
Fast secondary electrons (FSEs) are high-energy electrons which are generated in the specimen; they are high-energy because they receive a large fraction of the beam energy (Joy 1984). From the cross-section data in Figure 4.1 you can see that they should be an order of magnitude more probable than slow 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  $\sim 50$ – $200$  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.

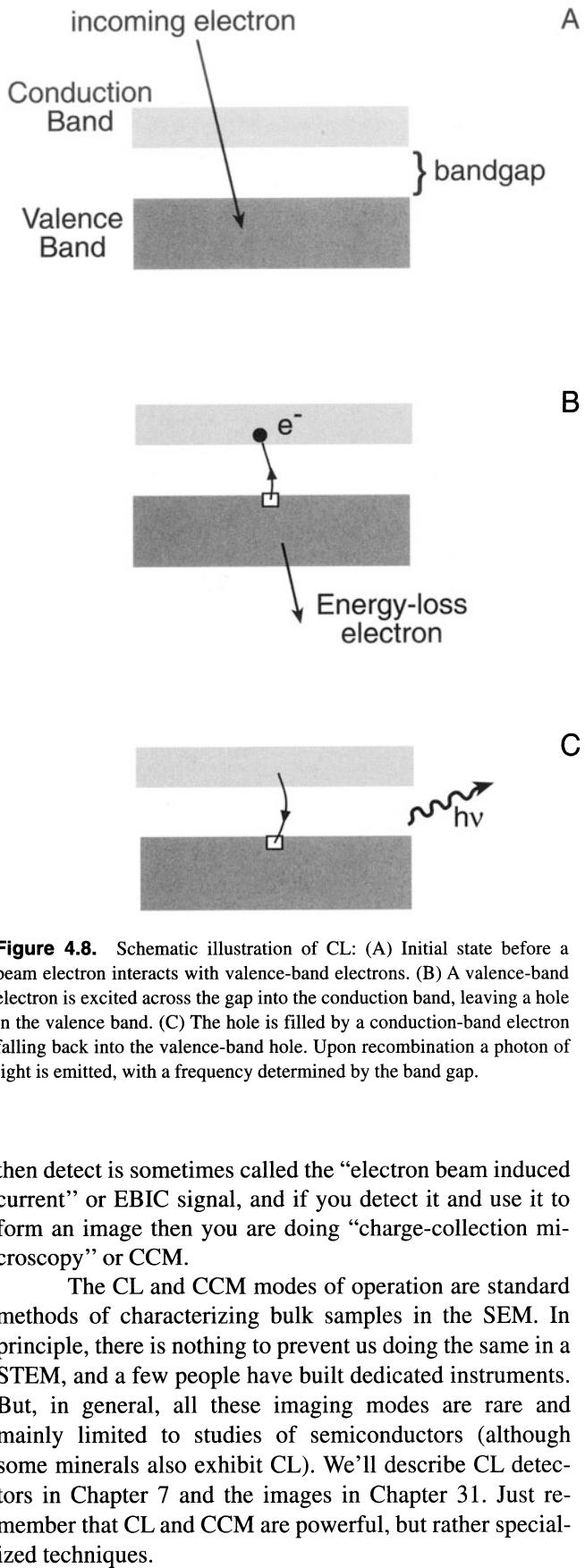
#### 4.4. ELECTRON–HOLE PAIRS AND CATHODOLUMINESCENCE (CL)

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



**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.

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.

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.”

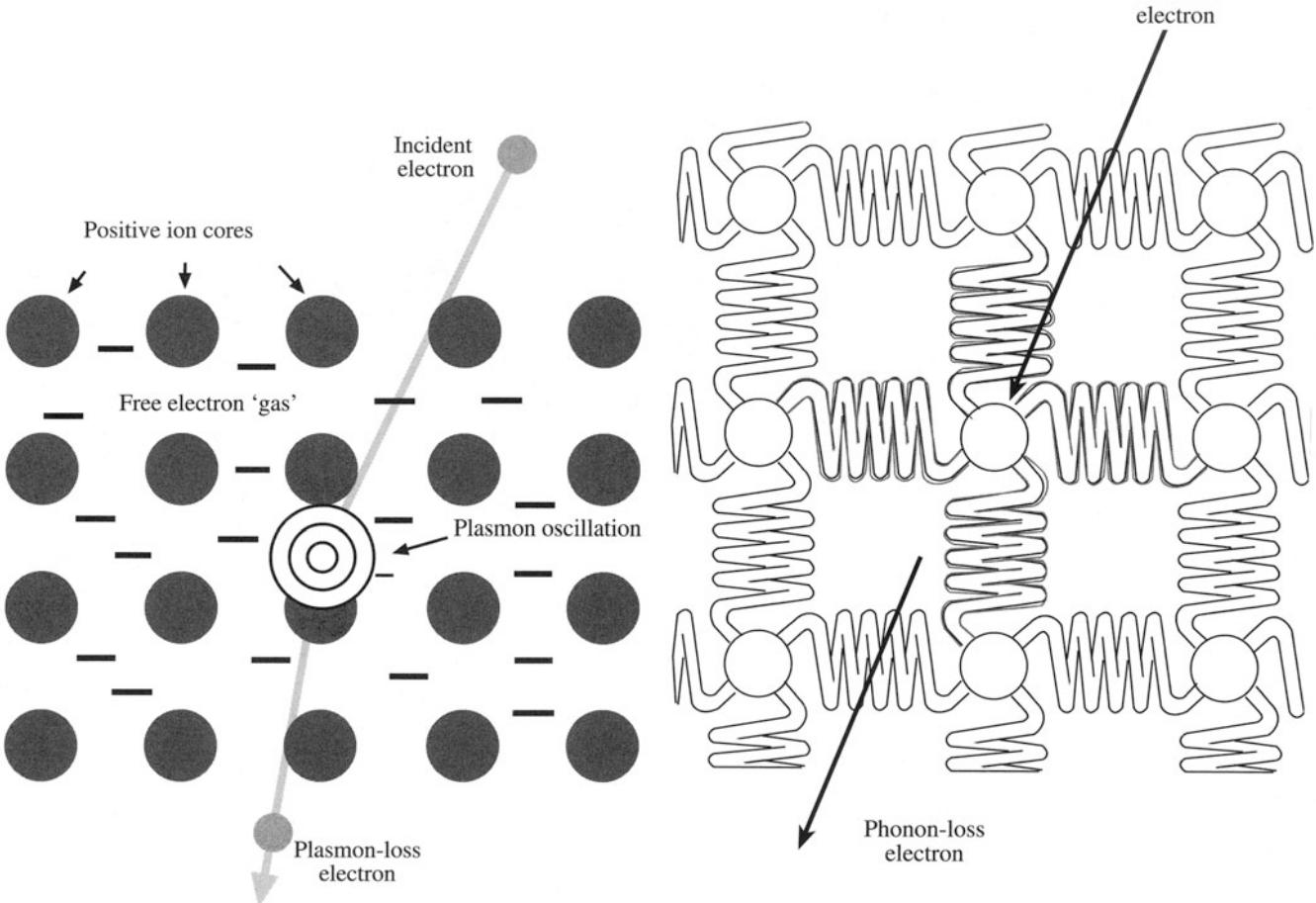
Plasmons are collective oscillations of free electrons that occur when the beam electron passes through the free electron “gas.”

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 pre-

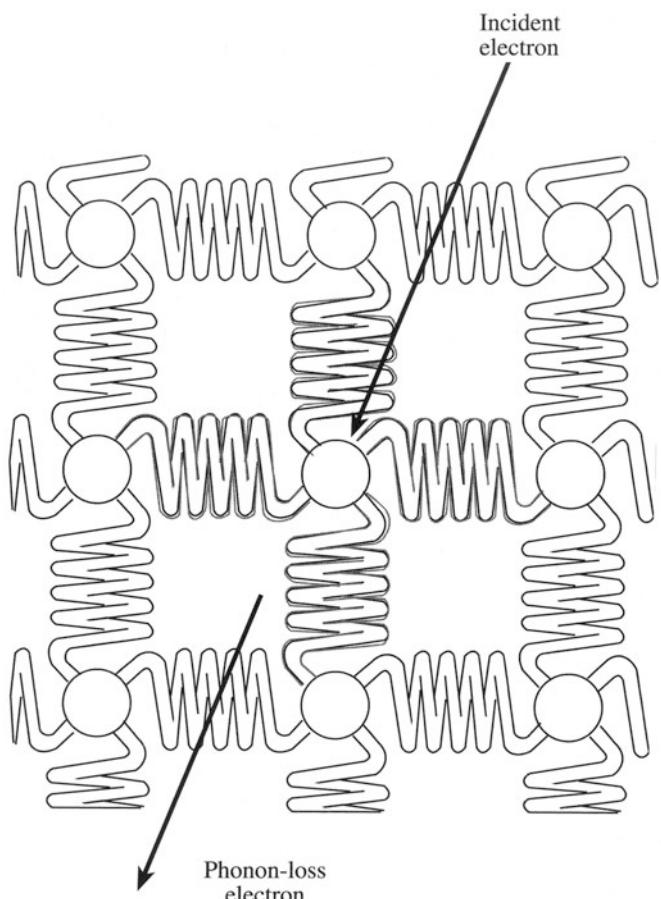
dominantly in metals, particularly ones like aluminum which have a large Fermi surface and thus a high free-electron density. The plasmon oscillation is quantized and the mean free path for plasmon excitation is of the order of 100 nm. As we’ll see in Chapter 39, this makes the number of plasmon excitations 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)

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



**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,  $\theta$  is the scattering angle, and  $\theta_E$  is the so-called “characteristic scattering angle” given by  $E_p/2E_0$ . Since  $E_p$ , the plasmon energy, is almost fixed ( $\sim 15\text{--}25$  eV), the cross section is a strong function of  $\theta$ , dropping rapidly to zero at values much above 10 mrad, 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 mrad), 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  $\sim 2\text{--}3\times$ .

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:

- **Radiotherapy:** 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 *radiotherapy*; 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.

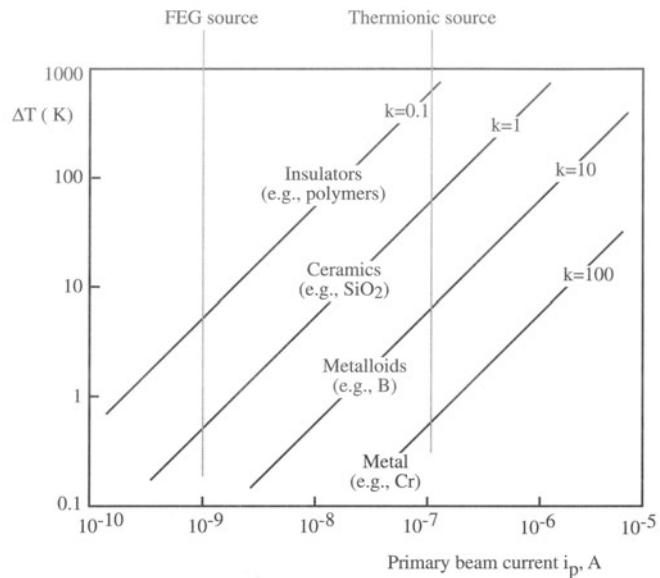
#### 4.6.A. Electron Dose

In the TEM we define the electron dose as the charge density ( $\text{Cm}^{-2}$ ) hitting the specimen. It is easy to convert this to the number of electrons/unit area (usually  $e/\text{nm}^2$ ) knowing that  $e = 1.6 \times 10^{-19} \text{ C}$ . 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  $\text{W m}^{-1}\text{K}^{-1}$  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  $\sim 1700^\circ\text{C}$ .

#### 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<sub>2</sub> 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.

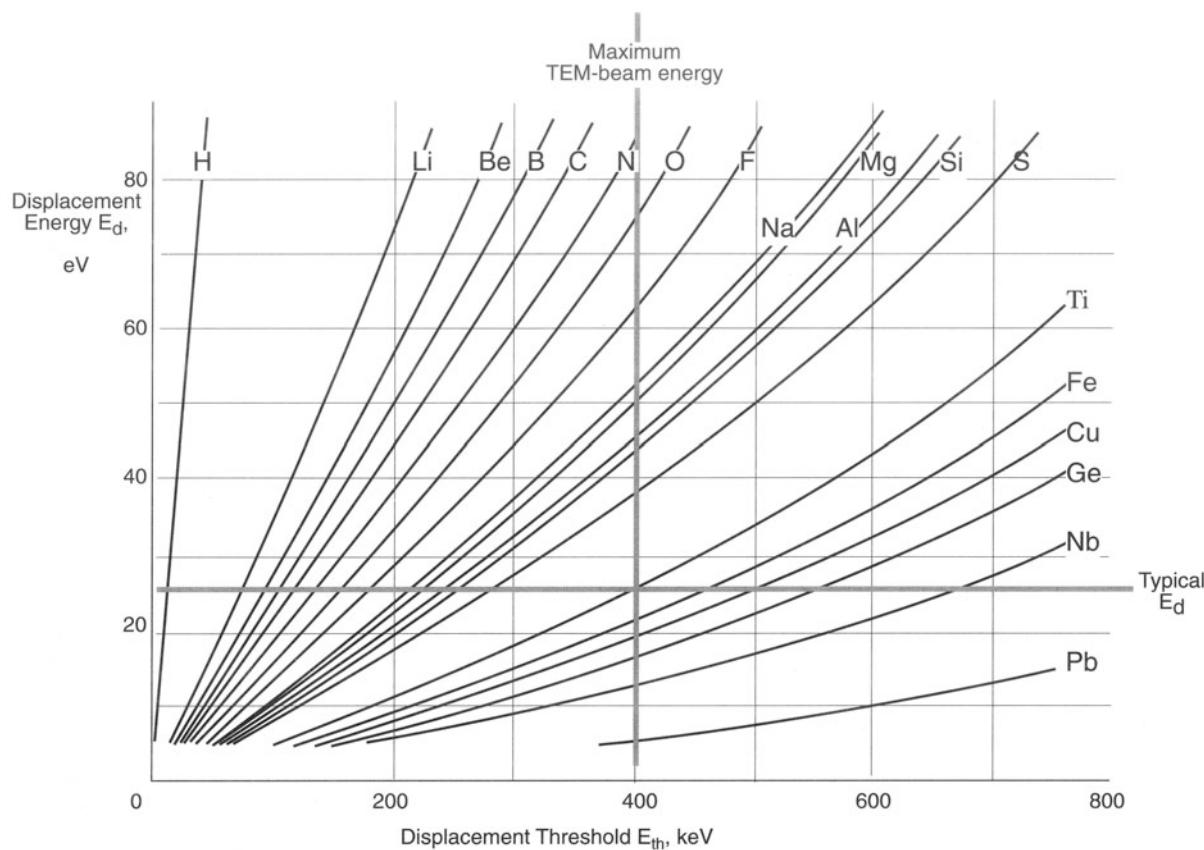
How strongly the atoms are bonded to their neighbors will also be a factor. A simple expression given by Hobbs (1979) for the displacement energy  $E_d$  allows us to determine the threshold energy ( $E_t$ ) for displacement of atoms of atomic weight  $A$

$$E_t = \frac{\left(\frac{100 + AE_d}{5}\right)^{1/2} - 10}{20} \quad [4.9]$$

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  $\sim 15\text{--}25$  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)**

Element	$T$ (eV)				$E_d$ (eV)	$E_s$ (eV)
	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 in-

terpretation 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

---

<b>5.1. The Physics of Different Electron Sources .....</b>	<b>69</b>
<b>5.1.A. Thermionic Emission .....</b>	<b>69</b>
<b>5.1.B. Field Emission .....</b>	<b>70</b>
<b>5.2. The Characteristics of the Electron Beam .....</b>	<b>70</b>
<b>5.2.A. Brightness .....</b>	<b>70</b>
<b>5.2.B. Temporal Coherency and Energy Spread .....</b>	<b>71</b>
<b>5.2.C. Spatial Coherency and Source Size .....</b>	<b>72</b>
<b>5.2.D. Stability .....</b>	<b>72</b>
<b>5.3. Electron Guns .....</b>	<b>72</b>
<b>5.3.A. Thermionic Guns .....</b>	<b>72</b>
<b>5.3.B. Field-Emission Guns (FEGs) .....</b>	<b>76</b>
<b>5.4. Comparison of Guns .....</b>	<b>77</b>
<b>5.5. Measuring Your Gun Characteristics .....</b>	<b>78</b>
<b>5.5.A. Beam Current .....</b>	<b>78</b>
<b>5.5.B. Convergence Angle .....</b>	<b>78</b>
<b>5.5.C. Calculating the Beam Diameter .....</b>	<b>79</b>
<b>5.5.D. Measuring the Beam Diameter .....</b>	<b>79</b>
<b>5.5.E. Energy Spread .....</b>	<b>82</b>
<b>5.5.F. Spatial Coherency .....</b>	<b>82</b>
<b>5.6. What kV Should You Use? .....</b>	<b>82</b>

## 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 ( $\text{LaB}_6$ ) 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” ( $\Phi$ ) 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

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

where  $k$  is Boltzmann’s constant ( $8.6 \times 10^{-5}$  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  $\Phi$  is given to the electrons; then they will escape from the source and be available to form an electron beam. Unfortunately, when we put a few eV of thermal energy into most materials they either melt or vaporize. So the only viable thermionic sources are either refractory (high melting point) materials or those with an exceptionally low 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  $\Phi$  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  
 $\text{LaB}_6$  and other low- $\Phi$  materials, e.g.,  $\text{CeB}_6$

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

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

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  $<0.1\text{ }\mu\text{m}$ . If we apply a 1-kV potential to this tip, then  $E$  is  $10^{10}\text{ V/m}$  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  $\text{LaB}_6$ , depends on the crystallography of the tungsten tip; the  $\langle 310 \rangle$  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 ( $<10^{-11}\text{ Torr}$ ), 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  $\text{ZrO}_2$  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  $\alpha_0$ .

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  $\alpha_0$  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,  $\beta$ , as

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

This equation is an important one which you should remember. The units of  $\beta$  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  $\beta$  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  $\beta$ , 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<sup>2</sup>. 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  $\lambda_c$  is

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

where  $v$  is the electron velocity,  $\Delta E$  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  $\Delta E$ , 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  $\Delta E$  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  $\Delta E$  is still small. From these values of  $\Delta E$ , 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

$$d_c \ll \frac{\lambda}{2\alpha} \quad [5.5]$$

where  $\lambda$  is the electron wavelength and  $\alpha$  is the angle subtended by the source at the specimen. We can control  $\alpha$  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  $\alpha$ .
- If your source size is large (e.g., a W hairpin) decrease the accelerating voltage and thus increase  $\lambda$ .

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  $\beta$  and better spatial coherency, but less stability.

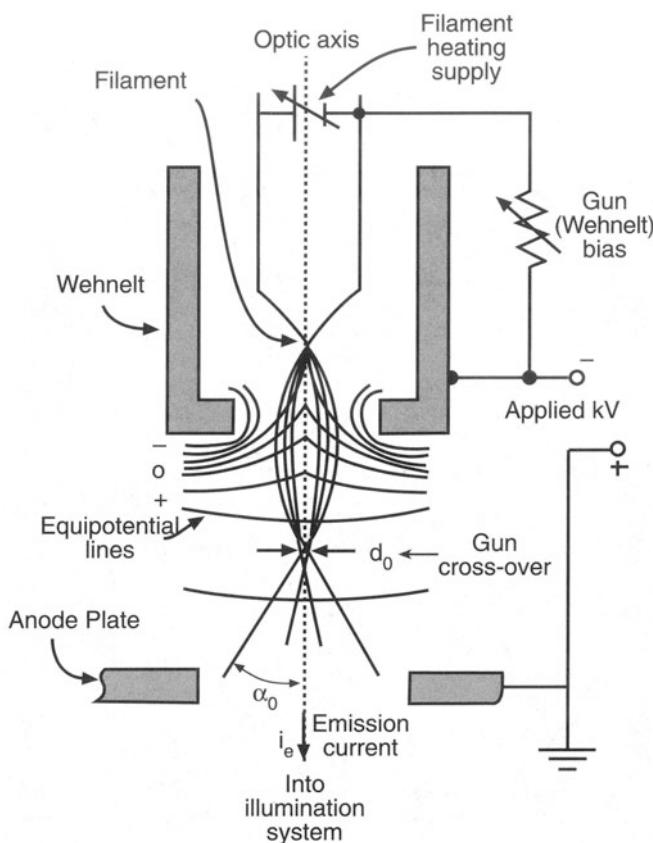
Now that we know the critical characteristics required of electron sources, let's examine those used in commercial TEMs.

## 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<sub>6</sub> 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  $\alpha_0$ .

voltage power supply. This cable also connects to the tungsten filament to supply a current to heat the filament resistively to the operating temperature.  $\text{LaB}_6$  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.

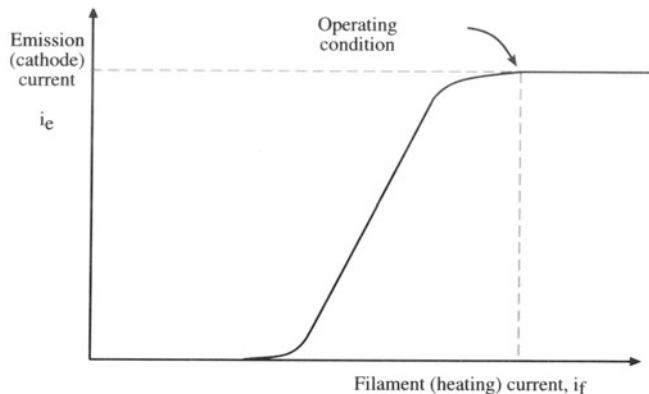
When the electrons leave the cathode they have a negative potential of 100 kV with respect to the earthed anode, so they accelerate through this potential difference ac-

quiring an energy of 100 keV, and a velocity of greater than half the speed of light.

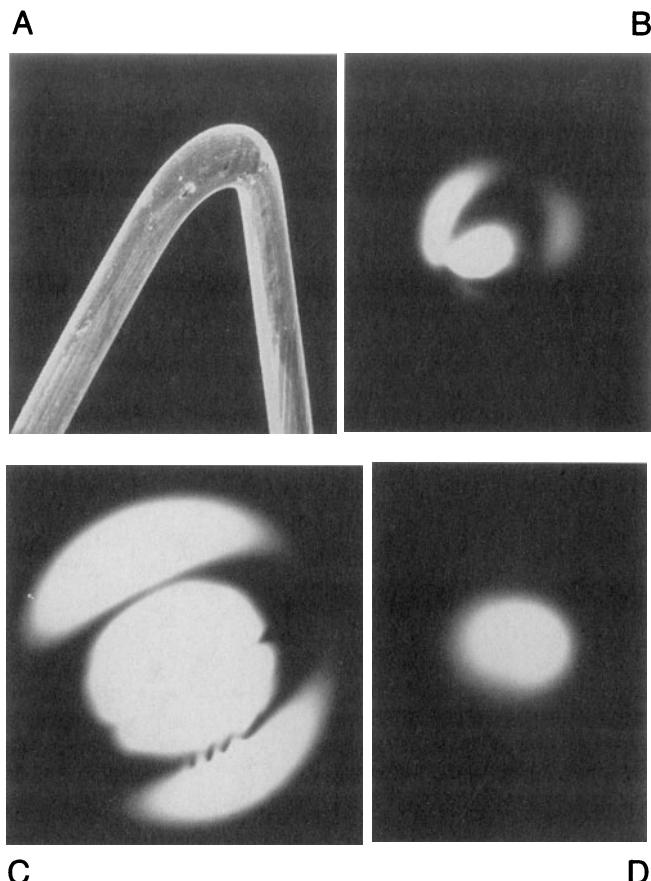
Now to get a controllable beam of electrons through the hole in the anode and into the microscope itself, we apply a small negative bias to the Wehnelt cylinder. The electrons coming off the cathode see the negative field and are converged to a point called a crossover between the Wehnelt and the anode 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.

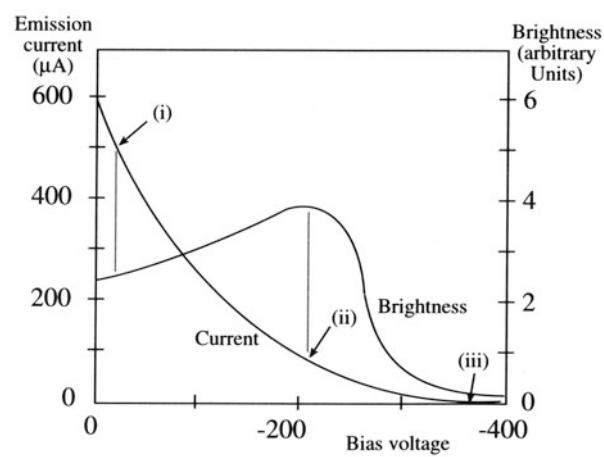
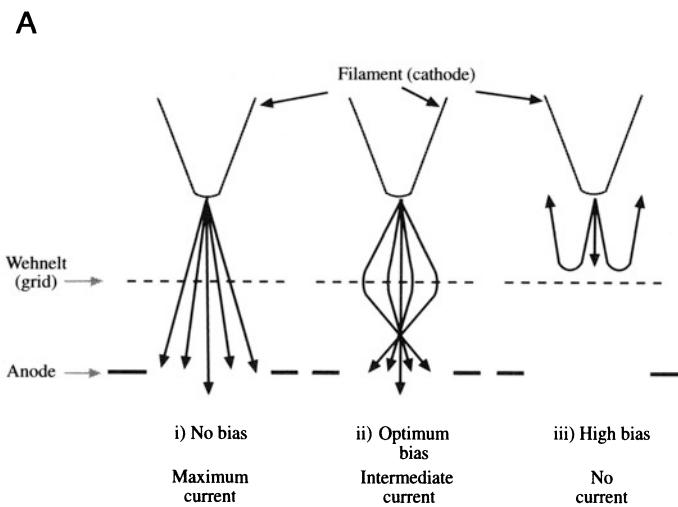
The Wehnelt acts as a simple electrostatic lens: the first lens in the microscope.

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

C

D

**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.



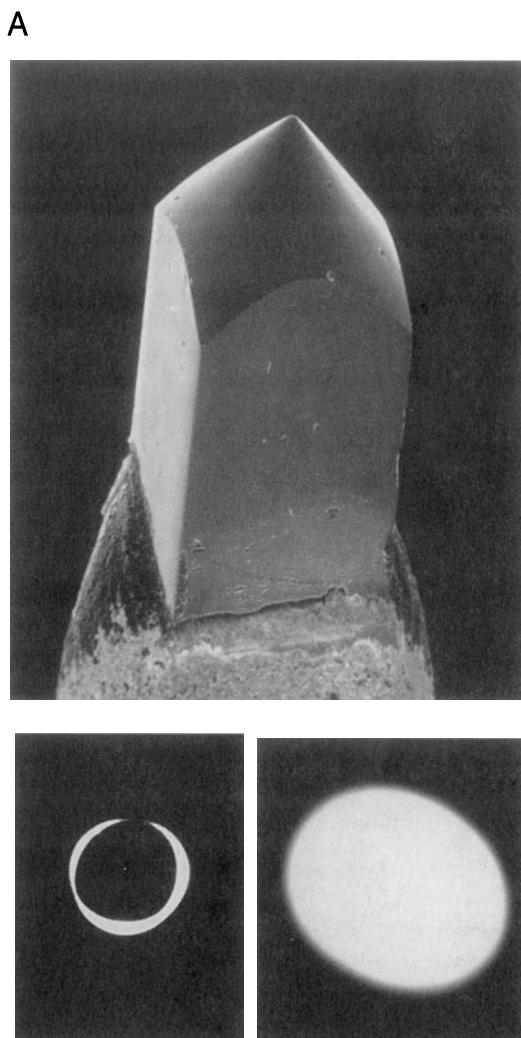
**Figure 5.4.** (A) The effect of increasing Wehnelt bias (i–iii) on the distribution of electrons coming through the anode. (B) The relationship between the bias and the emission current/gun brightness. Maximum brightness is achieved at an intermediate Wehnelt bias, and an intermediate emission current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

back in the brightness equation (equation 5.3) and the convergence/divergence angle at the crossover is  $\alpha_0$  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  $\beta$  would be low. The optimum  $\beta$  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  $\text{LaB}_6$  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  $\text{LaB}_6$  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.  $\text{LaB}_6$  crystals are more susceptible to thermal shock than tungsten, and so you should take care when heating *and cooling* an  $\text{LaB}_6$  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  $\text{LaB}_6$  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  $\text{LaB}_6$  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  $\beta$  is critical in any operations that require a fine beam ( $<0.1 \mu\text{m}$ ).

In an SEM, which always requires a small probe, the gun is carefully adjusted by the manufacturer to produce optimum  $\beta$  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  $\beta$ , 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  $\beta$ . This is not important if you’re operating with a broad beam, but if you want to operate at maximum  $\beta$  with a focused beam, as is the case for AEM, then you need to be able to measure  $\beta$ ; we’ll show you how to do that in Section 5.5.

### 5.3.B. Field-Emission Guns (FEGs)

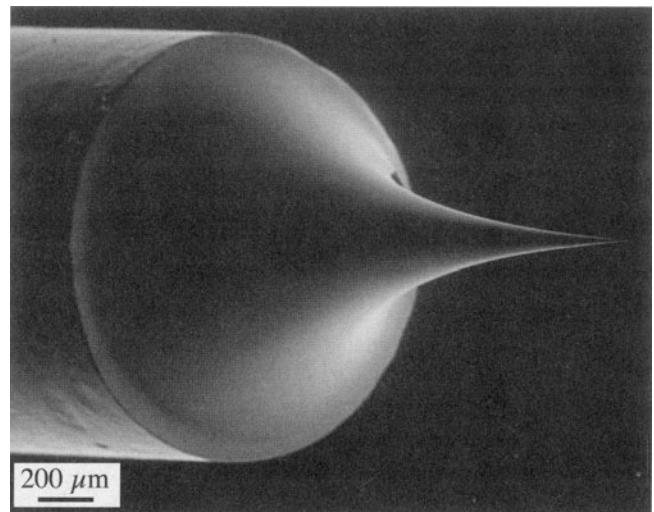
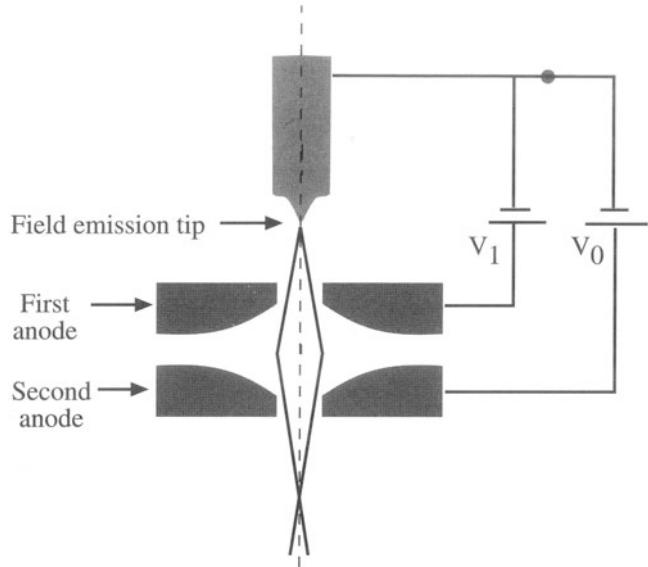
In many ways, FEGs are much simpler than thermionic guns. In order to get an FEG to work we make it the cathode with respect to *two* anodes. The first anode is positively charged by several kV with respect to the tip. This is called the “extraction voltage” since it generates the intense electric field-extracting 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  $\beta$ . 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.

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



**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.

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  $\sim 5000$  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.

$\text{LaB}_6$  is a more useful source for several reasons. While it is not as refractory as tungsten,  $\text{LaB}_6$  has a much lower value of  $\Phi$ , and since  $\Phi$  appears in the exponential in the Richardson equation, its effect on the current density is dominant.  $\text{LaB}_6$  crystals can be produced with a fine tip about 1  $\mu\text{m}$  in radius, which accounts for the smaller crossover size. As a result  $\text{LaB}_6$  current densities are considerably higher than for tungsten. The brightness is typically 10 times that of tungsten, even though  $\text{LaB}_6$  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  $\text{LaB}_6$  is purely economic.  $\text{LaB}_6$  sources cost several hundred dollars each while tungsten filaments are so cheap that the manufacturer often provides them free. Because  $\text{LaB}_6$  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,  $\text{LaB}_6$  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  $\text{LaB}_6$  sources

gently and you will be well rewarded. If users are not careful, your TEM supervisor may try to extend the life of the  $\text{LaB}_6$  to the point where it behaves no better than a W filament.  $\text{LaB}_6$  sources don't die, they fade away.

In FEGs, the current density is enormous and  $\beta$  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  $\text{LaB}_6$  sources,  $\beta$  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  $\beta$  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 ( $\approx 1\text{--}10 \mu\text{m}$ ) 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	$\text{LaB}_6$	Field Emission
Work function, $\Phi$	eV	4.5	2.4	4.5
Richardson's constant	$\text{A/m}^2\text{K}^2$	$6 \times 10^5$	$4 \times 10^5$	
Operating temperature	K	2700	1700	300
Current density	$\text{A/m}^2$	$5 \times 10^4$	$10^6$	$10^{10}$
Crossover size	$\mu\text{m}$	50	10	<0.01
Brightness	$\text{A/m}^2\text{sr}$	$10^9$	$5 \times 10^{10}$	$10^{13}$
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  $\beta$ , 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  $\beta$ , then  $\Delta E$ , 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  $\beta$ . However, while we can easily get a measure of the emission current at the gun, it is more difficult to measure  $d_0$  and  $\alpha_0$  there. So we make the approximation that, if we neglect lens aberrations,  $\beta$  is constant throughout the electron optical system so it doesn't matter where it is measured. It is easiest, practically, to determine  $\beta$  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  $\mu\text{m}$ ) 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  $\alpha$ .

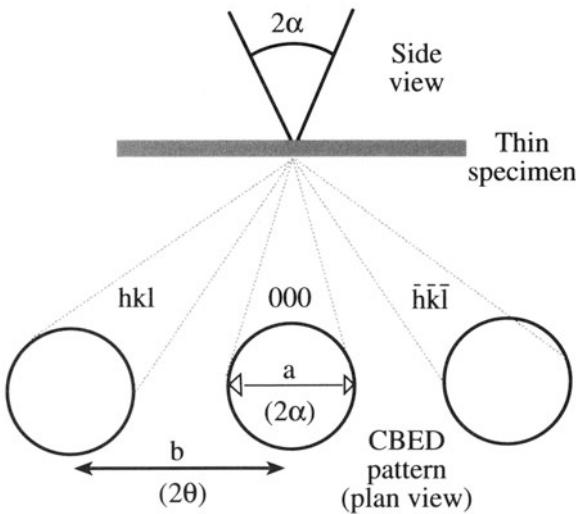
- The beam current is usually in the range from nanoamps to picoamps.
- The emission current is typically several microamps.

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  $\alpha$  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  $2\alpha$  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  $h\bar{k}\bar{l}$  disk. Thus

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



**Figure 5.8.** The distances on a convergent-beam diffraction pattern from which you can measure the beam-convergence semiangle,  $\alpha$ , 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  $\alpha$  is useful in many aspects of TEM. The value of  $\alpha$  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  $\alpha$ , 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.

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,  $d_t$

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

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  $\beta$ , and a value of  $\beta$  has to be assumed for the purposes of calculation. The expression for  $d_g$  is

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

We have already defined  $i$ ,  $\beta$ , and  $\alpha$ .

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

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

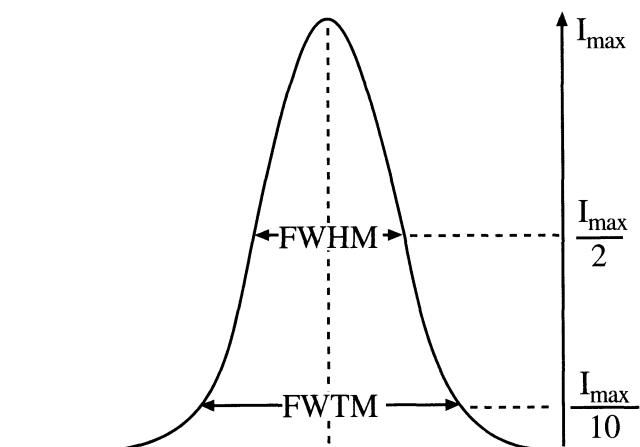
where  $C_s$  is the spherical aberration coefficient, which we discuss in detail in Chapter 6. This is the full diameter containing 100% of the beam current. Clearly, this term is not Gaussian unless the beam is correctly apertured which, as we just discussed, is not always possible. The calculated diameter due to diffraction is

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

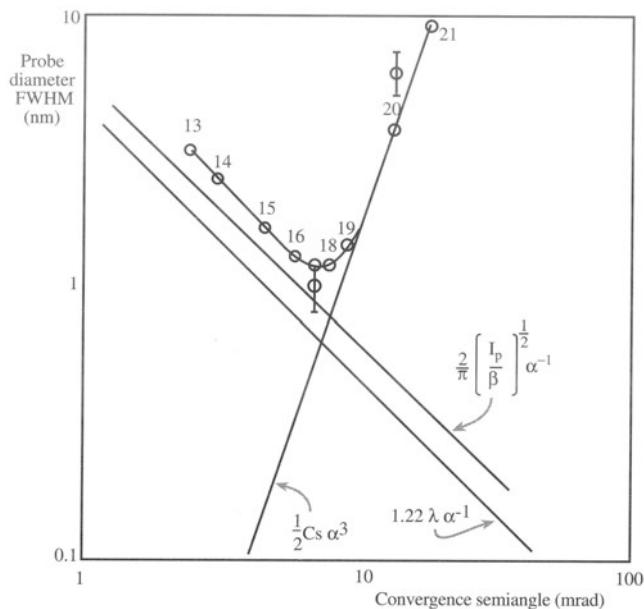
which is the Rayleigh criterion which we discussed in Chapter 1. Although all these definitions do not define the same diameter of the electron distribution, they are all combined to give a first approximation of the FWHM of the beam. Clearly, it is more reliable, but more time-consuming, to measure  $d$  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

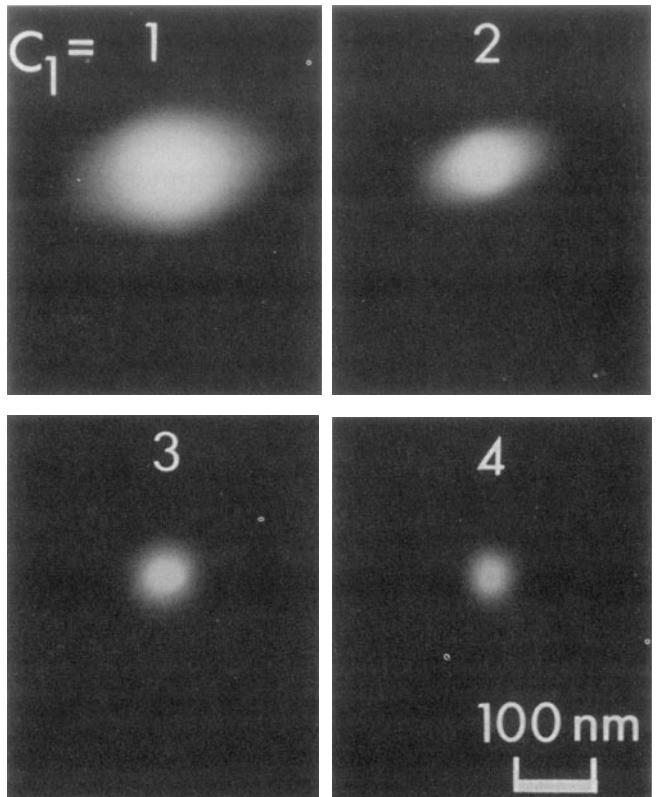
To measure the beam size in a TEM/STEM, you must form an image of the beam on the TEM viewing screen under conditions where you know, or can calibrate, the magnification. This is a nontrivial exercise and you may need to



**Figure 5.9.** 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.



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

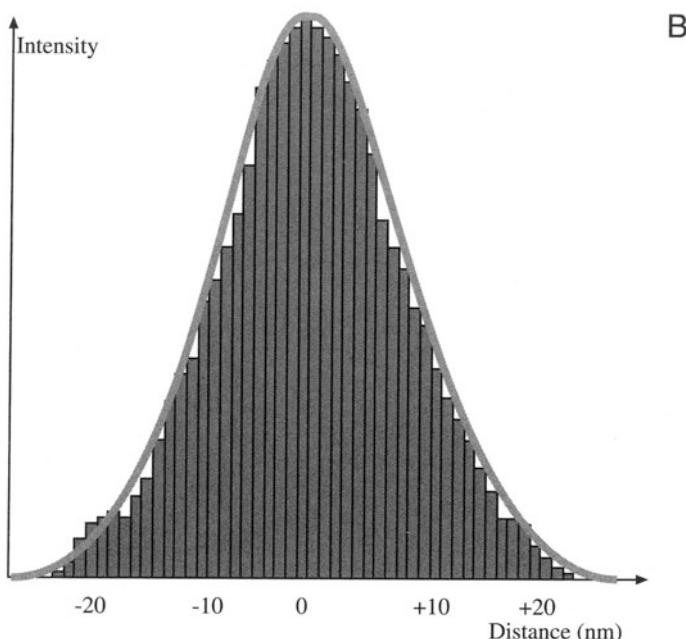


consult the manufacturer to be sure that you are doing it correctly. You can then photograph the beam and determine the intensity distribution from a microdensitometer trace across the image, as shown in Figure 5.11.

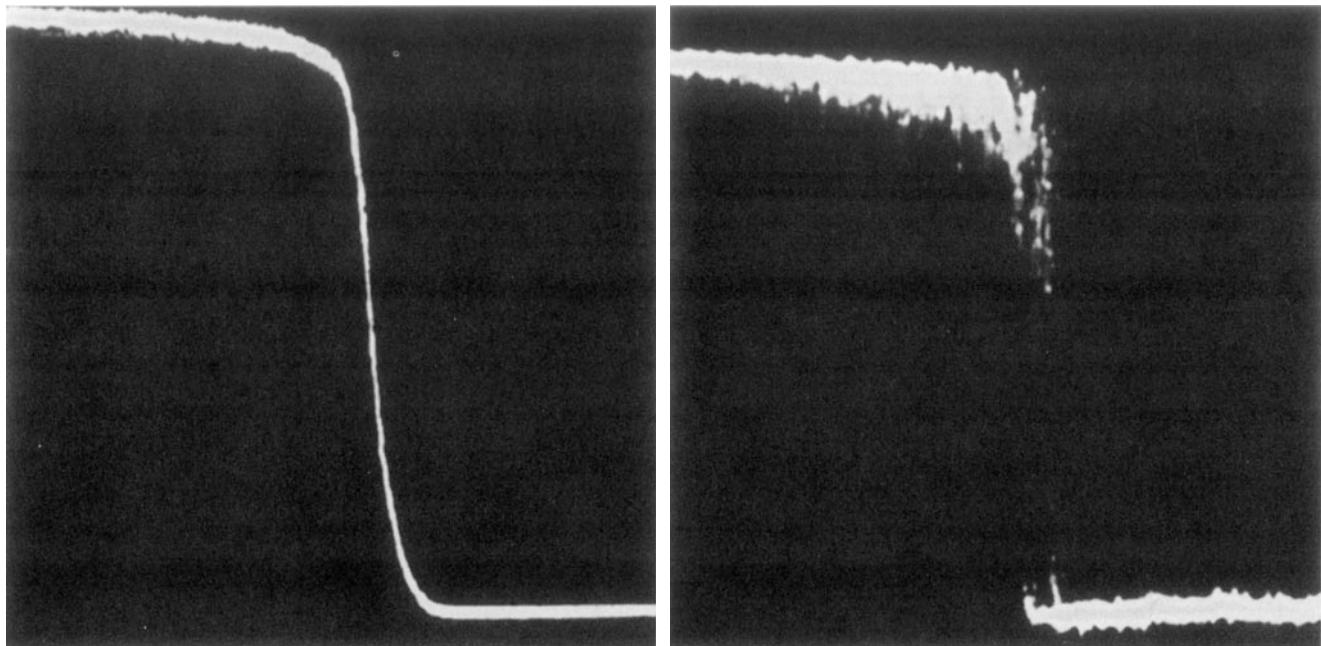
- The FWHM contains 50% of the integrated intensity. It is the value used by the manufacturers when they report beam sizes. It is also the important dimension when considering the effect of  $d$  on the STEM image resolution.
- The full width at tenth maximum (FWTM) contains 90% of the integrated intensity. 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  $1.82 \times$  FWHM and this is also shown in Figure 5.9. You should note, therefore, that you overestimate  $\beta$  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.

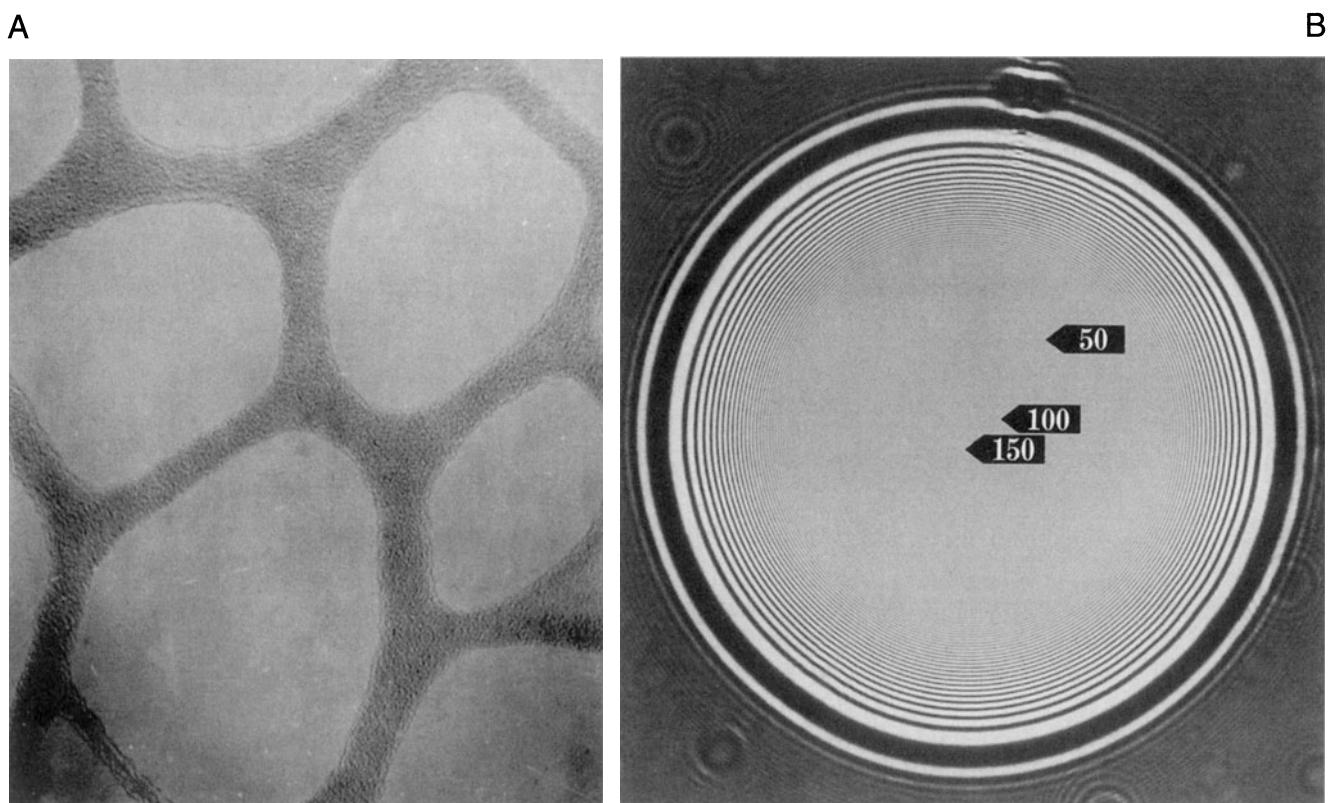
In a dedicated STEM you can't image the beam directly, since there are no post-specimen lenses to magnify



**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  $1 \times 10^6$ ) and in (right) 1.8 nm (magnification  $11 \times 10^6$ ). 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 ( $\Delta E$ ) 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  $\Delta E$  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  $\Delta E$  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  $\text{LaB}_6$  source and run at the highest kV. Take care when heating and cooling the  $\text{LaB}_6$  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  $\beta$ , 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

---

6.1. Why Learn about Lenses? . . . . .	87
6.2. Light Optics and Electron Optics . . . . .	87
6.2.A. How to Draw a Ray Diagram . . . . .	88
6.2.B. The Principal Optical Elements . . . . .	89
6.2.C. The Lens Equation . . . . .	90
6.2.D. Magnification, Demagnification, and Focus . . . . .	90
6.3. Electron Lenses . . . . .	91
6.3.A. Polepieces and Coils . . . . .	91
6.3.B. Different Kinds of Lenses . . . . .	92
6.3.C. Electron Ray Paths through Magnetic Fields . . . . .	93
6.3.D. Image Rotation and the Eucentric Plane . . . . .	95
6.3.E. Deflecting the Beam . . . . .	95
6.4. Apertures and Diaphragms . . . . .	95
6.5. Real Lenses and Their Problems . . . . .	97
6.5.A. Spherical Aberration . . . . .	97
6.5.B. Chromatic Aberration . . . . .	98
6.5.C. Astigmatism . . . . .	98
6.6. The Resolution of the Electron Lens (and Ultimately of the TEM) . . . . .	99
6.6.A. Theoretical Resolution . . . . .	99
6.6.B. Spherical Aberration-Limited Resolution (The Practical Resolution) . . . . .	100
6.6.C. Confusion in the Definitions of Resolution . . . . .	101
6.7. Depth of Focus and Depth of Field . . . . .	102

## 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

## 6.1. 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.

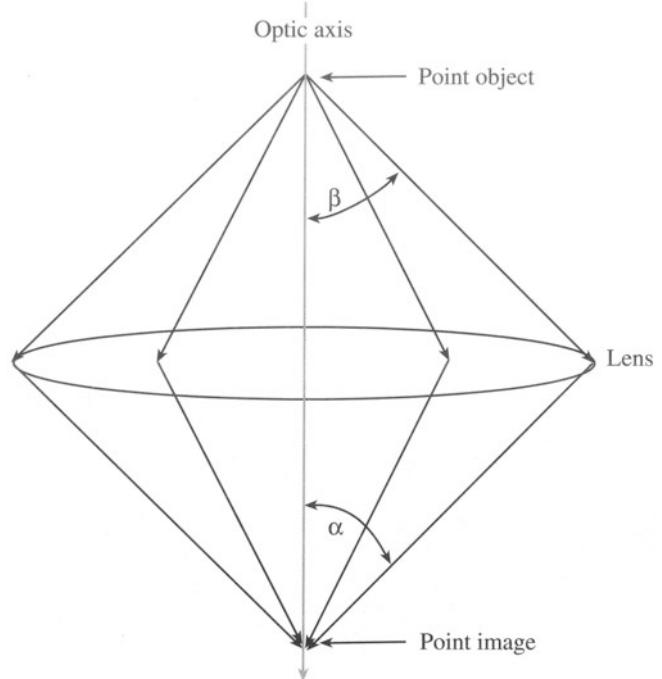
The first thing we need to do is to have a base line on which to draw our diagrams; this line is called the optic axis.

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  $\beta$  in Figure 6.1. Ultimately, as you can see,  $\beta$  is governed by the size of the lens, but we often choose to limit  $\beta$  by inserting an aperture, as we'll discuss later in this chapter. You'll often see the semiangle of collection defined as  $\alpha$ , but we will reserve  $\alpha$  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.

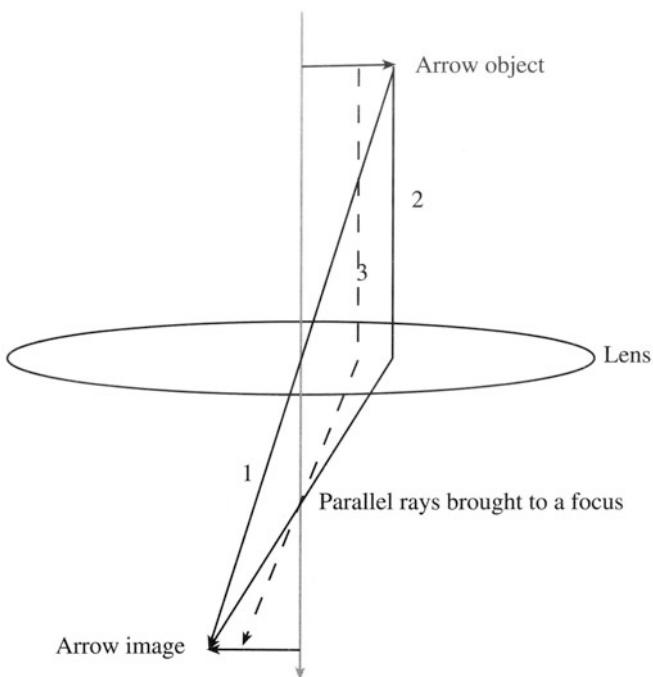
As you know from Chapters 2–4, most electrons are strongly forward scattered, so we can in practice gather a high fraction of the scattered electrons.



**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 ( $\beta$ ) or the image ( $\alpha$ ).

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  $\beta$  is maybe a few tens of milliradians (10 mrad =  $0.57^\circ$ ) 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  $180^\circ$ . 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 lens.



**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.

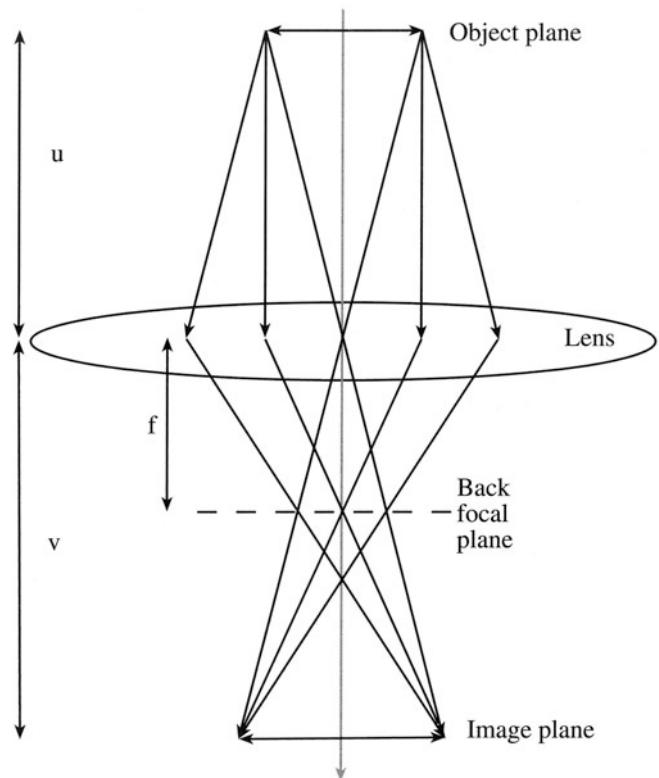
The image formed after each lens is rotated by 180° with respect to the object.

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

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

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

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

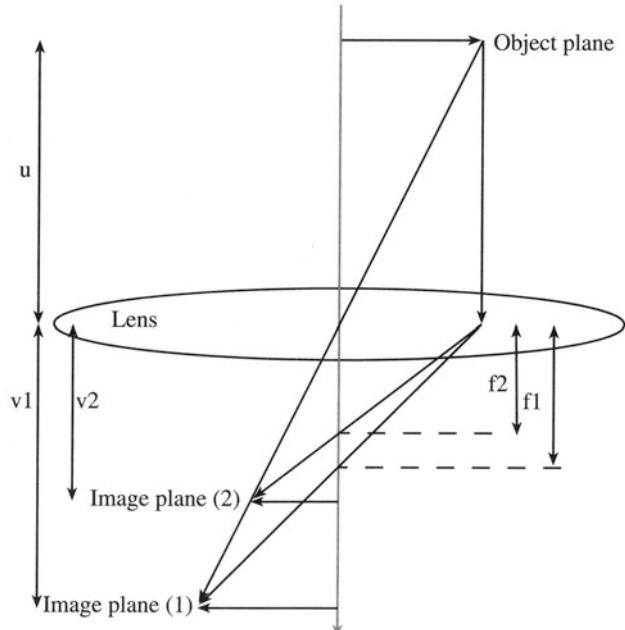
$M$  is also approximately equal to the ratio of the collection semiangles of the lens subtended at the object ( $\beta$ ) and at the image ( $\alpha$ ) as shown in Figure 6.1, assuming that these angles are small, as they invariably are in a TEM. In this example the magnification is unity.

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 ( $f_1$ ) produces a higher magnification of the object than a stronger lens ( $f_2$ ) since the image distance  $v$  increases, but the object distance is unchanged.

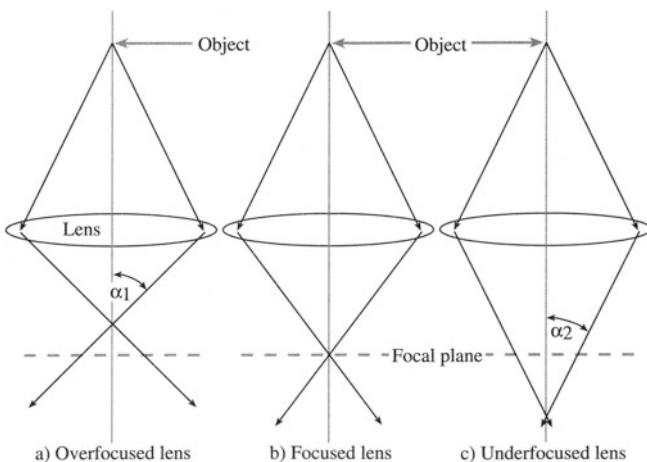
pound optical microscope. (We'll discuss the details of the lens combinations in the illumination and imaging systems of the TEM in Chapter 9.) In these circumstances, the image plane of the upper lens acts as the object plane for the lower lens, assuming we want the lower lens to further magnify the 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 ( $\alpha_2 < \alpha_1$ ).

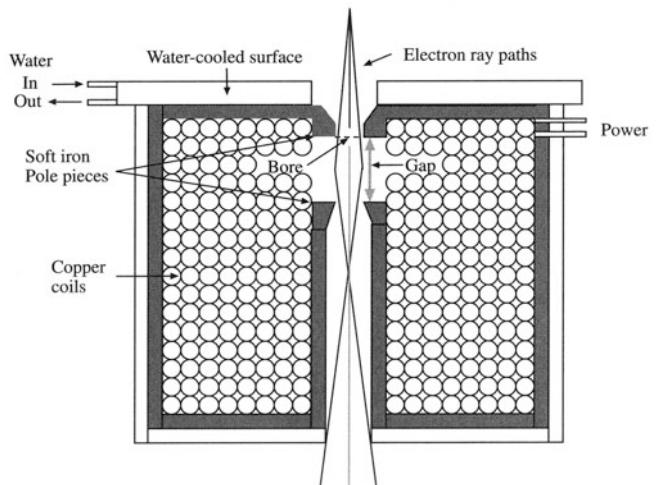
## 6.3. ELECTRON LENSES

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.

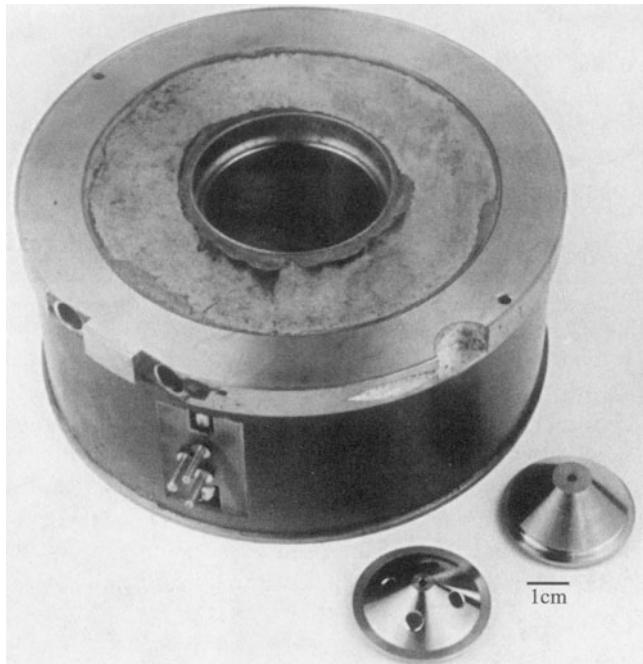
### 6.3.B. Different Kinds of Lenses

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.

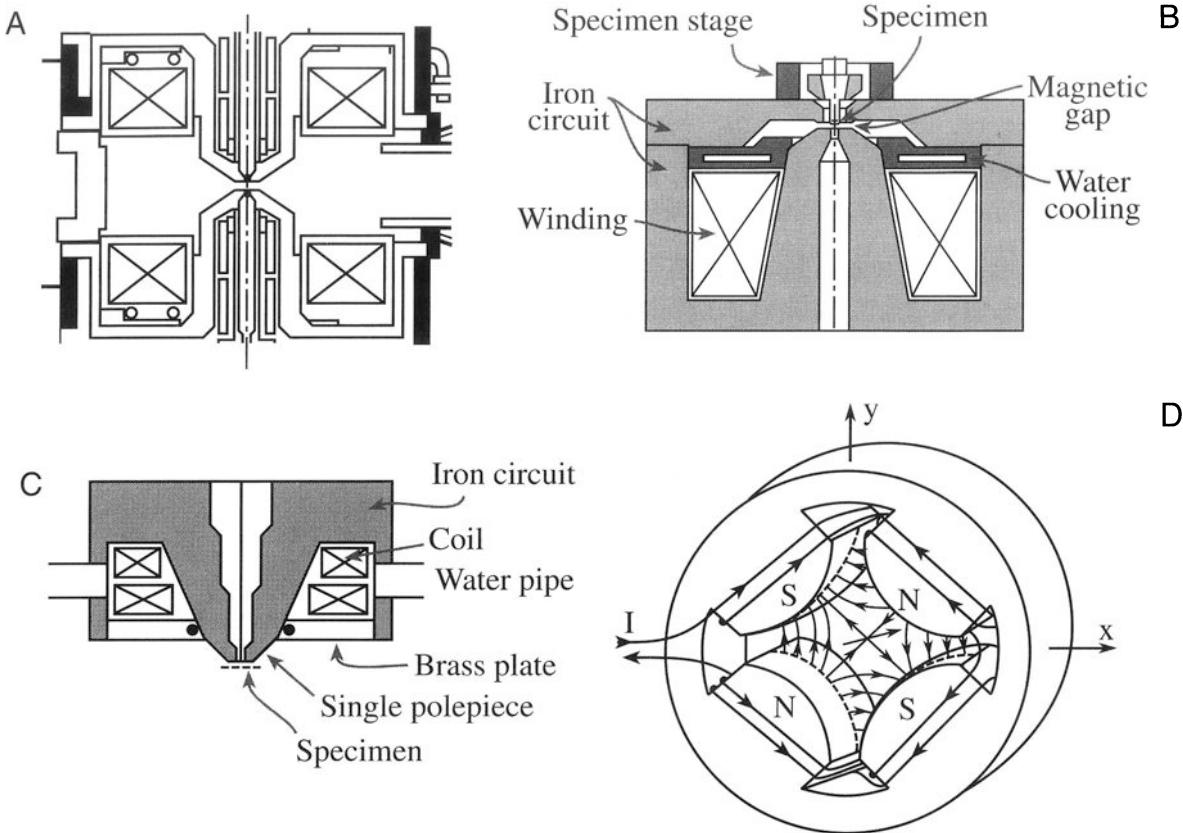
With split polepieces it is possible to make the upper polepiece behave differently than the lower polepiece.



**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  $\mathbf{B}$  (Tesla) and an electric field of strength  $\mathbf{E}$ , it experiences a force  $\mathbf{F}$ , known as the Lorentz force, which depends on the velocity of the electron,  $\mathbf{v}$ . All these factors are related through the equation

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

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  $\mathbf{F}$  is a vector normal to  $\mathbf{v}$  and  $\mathbf{B}$ , which are inclined to one another at an angle  $\theta$ . You can easily work out the relative directions of  $\mathbf{E}$ ,  $\mathbf{v}$ ,  $\mathbf{B}$ , and  $\mathbf{F}$  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  $90^\circ$  to  $\mathbf{B}$  is

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

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

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

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

$$r = \left[ \frac{2m_0 E \left( 1 + \frac{E}{2E_0} \right)}{eB} \right]^{1/2} \quad [6.6A]$$

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)

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

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  $\theta$  does equal  $90^\circ$ , the electron is traveling straight across the optic axis and is not focused! It is the deviation from  $\theta = 90^\circ$  that gives the lens effect. The next step therefore is to separate the electron velocity  $\mathbf{v}$  in a magnetic field into two components,  $\mathbf{v}_1$  perpendicular to, and  $\mathbf{v}_2$  parallel to, the magnetic field direction  $\mathbf{B}$ , as shown

in Figure 6.9, where  $v_1 = v \sin \theta$  and  $v_2 = v \cos \theta$ . The parallel component,  $\mathbf{v}_2$ , results in motion parallel to the optic axis in the  $z$  direction, with  $z = v_2 t$ , while the perpendicular component produces circular motion with radius given by equation 6.5. Hence we see the very important result:

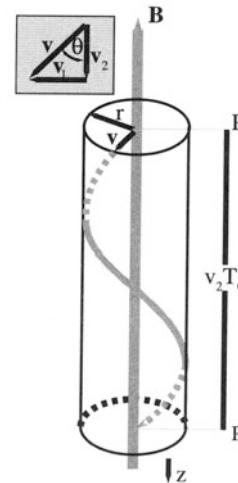
The electron spirals through the lens field with a helical trajectory.

The period of rotation ( $T_c$ ) through the field gives rise to a so-called “cyclotron frequency”  $\omega$

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

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 ( $\theta$ ) 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”

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



**Figure 6.9.** Electron trajectories in a homogeneous magnetic field, strength  $\mathbf{B}$ . The electrons have velocity components parallel and perpendicular to the field, so long as they are not traveling at  $90^\circ$  to the direction of  $\mathbf{B}$ . 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,  $r$ .

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

where  $V$  is the accelerating voltage of the microscope and  $\eta$  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  $\epsilon$  is (for small  $\epsilon$ )

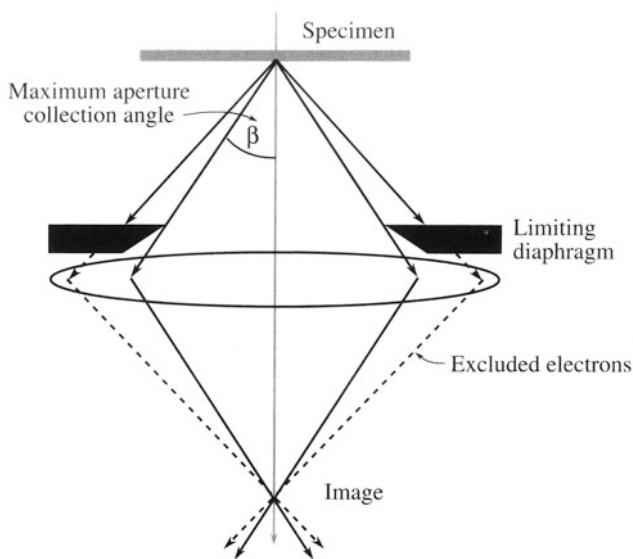
$$\epsilon = \frac{eLB}{mv} \quad [6.10]$$

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.

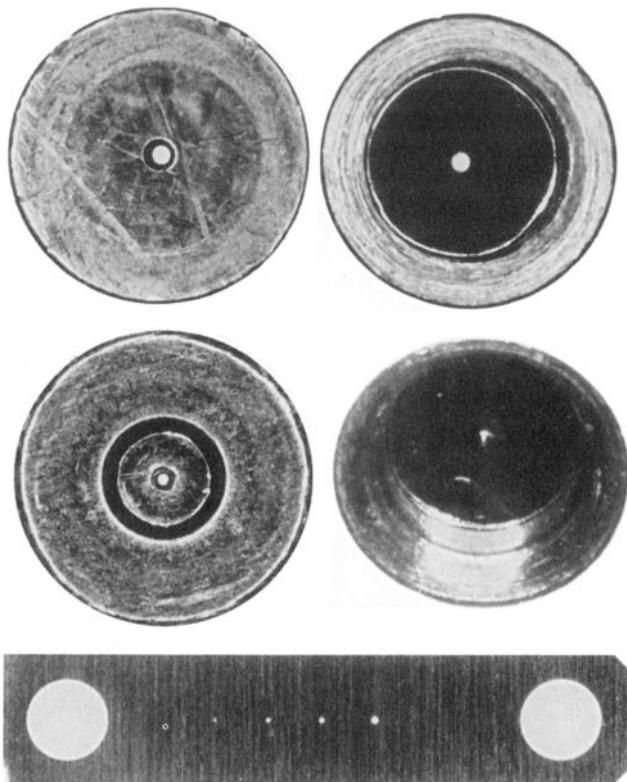
## 6.4. APERTURES AND DIAPHRAGMS

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

A



B



**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  $\beta$  subtended by the aperture at the object are allowed through the lens (full ray paths). Electrons from the object scattered at angles  $>\beta$  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  $\sim 3$  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  $10\text{ }\mu\text{m}$ , which is about the smallest circular aperture we can make consistently, up to  $\sim 0.3\text{ mm}$  ( $300\text{ }\mu\text{m}$ ). Usually the individual diaphragms or the strips are about  $25\text{--}50\text{ }\mu\text{m}$  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.

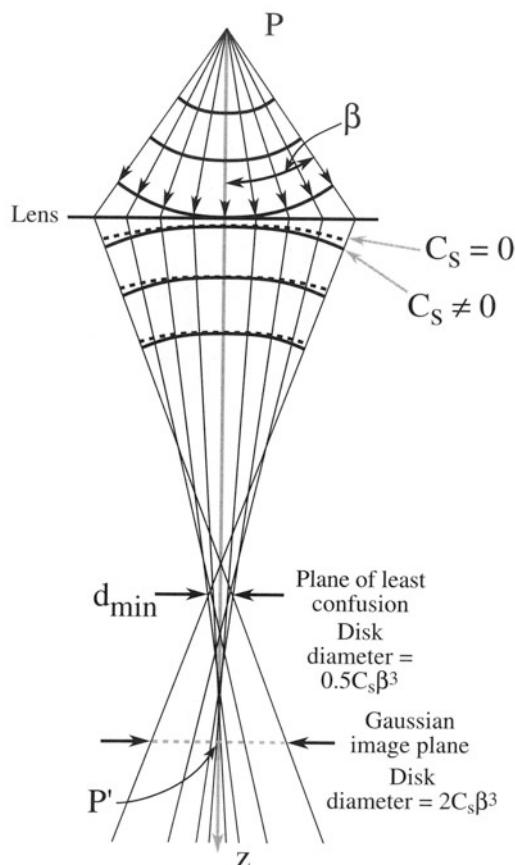
### 6.5.A. Spherical Aberration

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)

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

In this relationship  $L_0 = PP'_0$ , where  $P'_0$  is the Gaussian image of the point P for very small  $\theta$  (i.e., paraxial conditions). As  $\theta$  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,  $\delta$ , 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  $\theta$  in equation 6.14 is replaced by the maximum semiangle of collection of the objective lens aperture,  $\beta$ . We'll use  $\beta$  to define the objective lens semiangle of collection in the forthcoming discussion of resolution. This contrasts with our use of  $\alpha$  when discussing beam size in Chapter 5, since  $\alpha$  defined a semiangle of beam convergence. Most other TEM texts use  $\alpha$  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  $r_{\text{sph}}$  in the Gaussian plane under nonparaxial (i.e., realistic) conditions, given by

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

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

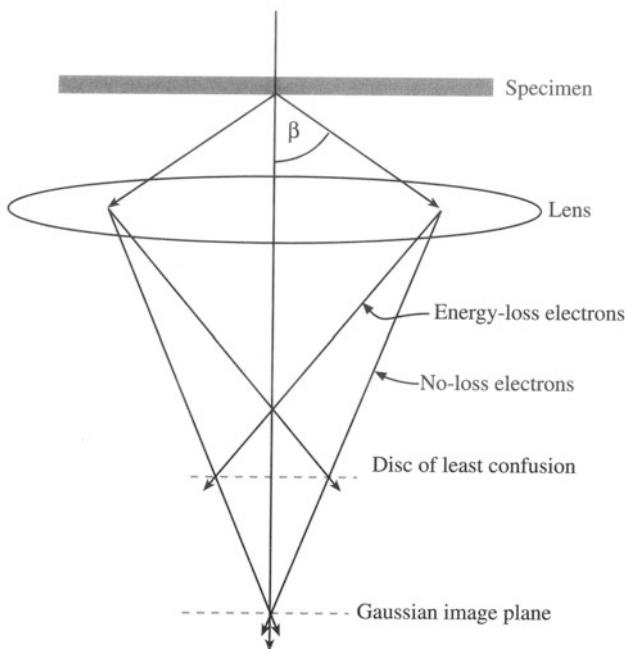
This term is related to the "color" (i.e., wavelength and hence energy) of the electrons. We've assumed the electrons are monochromatic, but they aren't. However, we can now make very good high-tension supplies and the variation of the electron energy is usually smaller than one part in  $10^6$ , which 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  $r_{\text{chr}}$  of this disk is given by

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

where  $C_c$  is the chromatic aberration coefficient of the lens (like  $C_s$  it is a length, approximately equal to the focal length),  $\Delta E$  is the energy loss of the electrons,  $E_0$  is the initial beam energy, and  $\beta$  is the semiangle of collection of the lens. While  $\Delta E$  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  $r_{\text{ast}}$  where

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

and  $\Delta f$  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  $r_{\text{sph}}$ ,  $r_{\text{chr}}$ ,  $r_{\text{ast}}$ ) 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

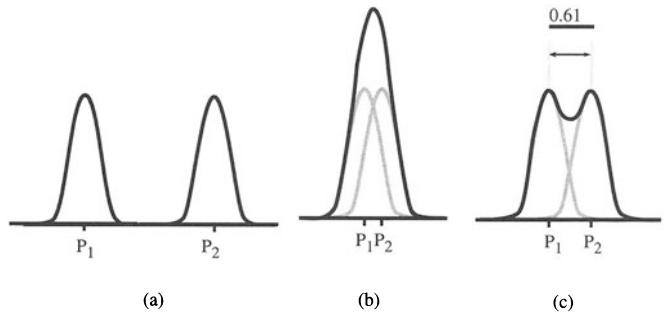
If there are *no* aberrations at all, the resolution of *any* lens (glass, electromagnetic) is customarily defined in terms of

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  $r_{\text{th}}$  and is given by the radius of the Airy disk

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

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  $\lambda$  or increase  $\beta$ . 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  $\lambda$  accounts for much of the drive toward intermediate and high voltage microscopes since  $\lambda$  decreases with keV, as we saw back in Chapter 1. But the obvious question is why don't we just increase  $\beta$  (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  $\beta$ .



**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  $r_{\text{sph}}$  limits the resolution. Now if you go back and look at equation 6.15 you'll see that  $r_{\text{sph}}$  increases with the cube of  $\beta$ , 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

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

We can thus find how  $r$  varies with  $\beta$

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

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

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

From this equation the optimum (compromise) value of  $\beta$  is given by

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

(Sometimes, this compromise value is determined approximately by simply equating the equations for  $r_{\text{th}}$  and  $r_{\text{sph}}$ . A quick calculation for 100-keV electrons ( $\lambda = 0.0037$  nm) for an instrument with  $C_s = 3$  mm gives a  $\beta_{\text{opt}}$  value of 4.5 mrad.)

If this expression for  $\beta_{\text{opt}}$  is substituted into equation 6.18, we get a minimum value of  $r(\beta)$

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

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

Typically, the value for  $r_{\text{min}}$  is  $\sim 0.25$ – $0.3$  nm, but the high-resolution instruments have  $r_{\text{min}} \sim 0.15$  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  $r_{\text{min}}$  (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  $\sim 0.2$  mm, then the maximum useful magnification of the best high-resolution TEM is  $\sim 10^6$ . 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  $r_{\text{th}}$  and  $r_{\text{sph}}$  is another possible way to determine  $r_{\text{min}}$  (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  $\beta_{\text{opt}}$  and 6.23 for  $r_{\text{min}}$ . 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  $\beta_{\text{opt}}$ , from equation 6.16 the value of  $r_{\text{chr}}$  is  $\sim 2$  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  $\sim 30$  nm, while at 300 keV you can probably get away with  $\sim 50$  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  $r_{\text{sph}}$ , 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  $\beta$  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  $r_{\text{sph}}$  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  $B \sim 1$  (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 ( $r_{\text{sph}}$ ) 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.,  $r_{\text{sph}} = C_s \beta^3$ .

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  $r_{\text{sph}}$  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  $r_{\text{sph}}$ , 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  $\beta_{\text{opt}}$  is extracted,  $A$  only varies by about  $\pm 15\%$ . 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.

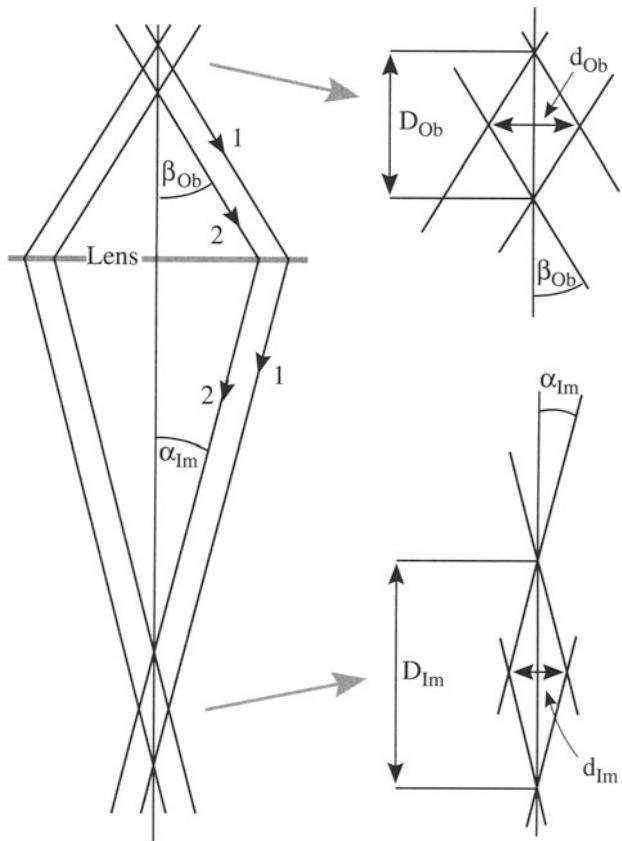
Basically, we are trying to find out how much of the object (the specimen) is in focus at the same time and over what range the image is in focus. (This latter question is irrelevant in SEM and STEMs where we don't use lenses to form the image, so both terms are equivalent.) In TEM both terms are important.

The depth of field,  $D_{\text{ob}}$ , 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,  $D_{\text{im}}$ , 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  $d_{\text{ob}}$  and  $d_{\text{im}}$  correspond to the smallest distances which we can resolve in the object or image, respectively. Now we can show that angles  $\alpha_{\text{im}}$  and  $\beta_{\text{ob}}$ , which are both small, are given by

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

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  $\pm D_{\text{ob}}/2$  either side of the specimen. The same rays define the depth of field over which the image is in focus  $\pm D_{\text{im}}/2$  either side of the image plane. The resolution in the object is  $d_{\text{ob}}$  and that in the image is  $d_{\text{im}}$ .

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

The angular magnification is thus

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

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

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

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

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

and the depth of field is

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

So we get a much increased depth of focus and field by using small apertures which reduce  $\beta$ . Notice that for a cor-

rect calculation of either  $D_{\text{ob}}$  or  $D_{\text{im}}$  you must be careful to select the right value of  $\beta$ . Under different circumstances the limiting semiangle is defined by the illumination aperture  $\alpha$  (in the C2 lens) or the objective aperture  $\beta_o$  (in the objective lens). In thin specimens, which scatter weakly, most electrons emerge from the specimen in a cone closer to that defined by  $\alpha_{\text{im}}$ , which is often small ( $\sim 10^{-4}$  rad). In a thicker, more strongly scattering specimen, the objective aperture defines the angle and  $\beta_o$  is usually  $10^{-2}$  rad.

For a collection semiangle,  $\beta_{\text{ob}}$ , of 10 mrad and a  $d_{\text{ob}}$  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 2 Å 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.,  $r_{\text{sph}} = C_s \beta^3$ .*

*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, New York.  
 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

---

7.1. Electron Detection and Display .....	107
7.2. Viewing Screens .....	108
7.3. Electron Detectors .....	108
7.3.A. Semiconductor Detectors .....	108
7.3.B. Scintillator–Photomultiplier Detectors .....	110
7.3.C. TV Cameras and Charge-Coupled Devices .....	111
7.3.D. Faraday Cup .....	112
7.4. Which Detector Do We Use for Which Signal? .....	112
7.5. Image Recording .....	114
7.5.A. Photographic Emulsions .....	114
7.5.B. Other Image Recording Methods .....	114
7.6. Comparison of Scanning Images and Static TEM Images .....	115

## 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

$$DQE = \frac{\left(\frac{S_{\text{out}}}{N_{\text{out}}}\right)^2}{\left(\frac{S_{\text{in}}}{N_{\text{in}}}\right)^2} \quad [7.1]$$

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  $<100 \mu\text{m}$  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  $\mu\text{m}$ , although they may be as small as 10  $\mu\text{m}$  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.

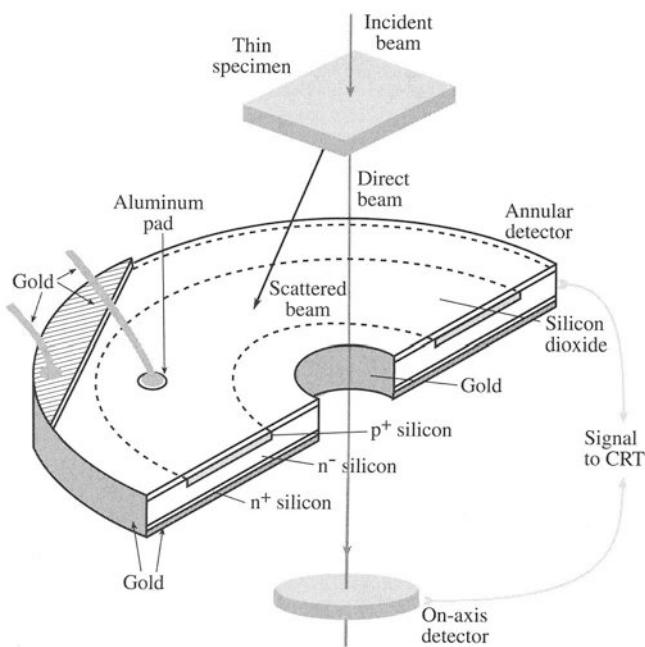
## 7.3. ELECTRON DETECTORS

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  $1 \mu\text{m}^2$ . 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  $3 \times 10^4$ , 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  $2 \times 10^4$ .

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:

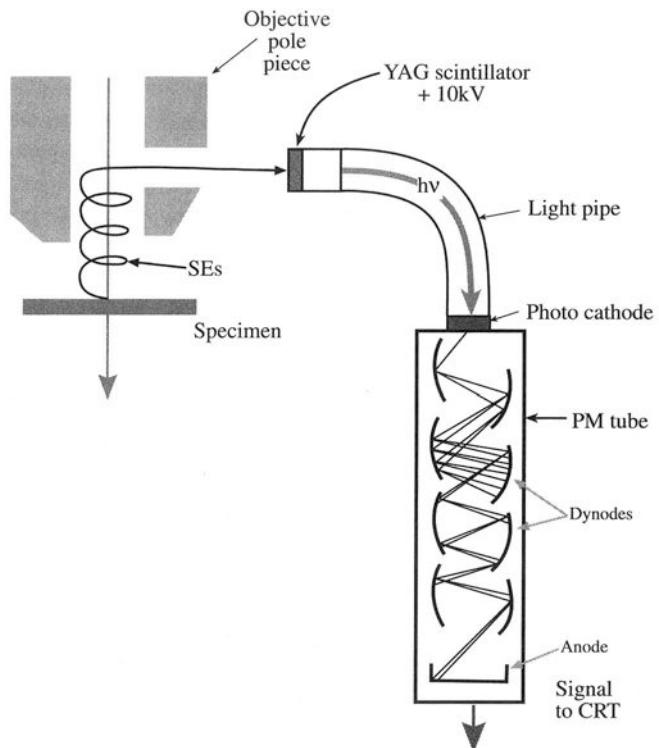
- The gain of the system is very high. The gain for the total detector system is of the order of  $10^n$ , depending on the number ( $n$ ) of dynodes in

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.

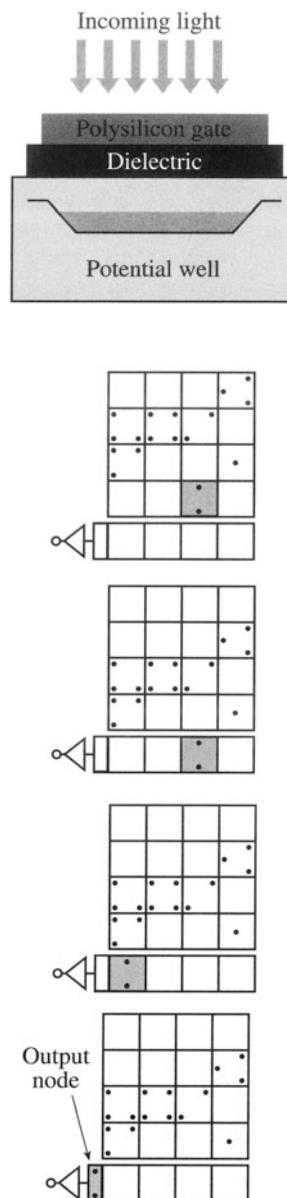
- ation damage, particularly after long-time exposure to the beam.
- The scintillator-PM combination is also substantially more expensive and bulky compared to semiconductor detectors and therefore it does not fit well within the TEM stage, nor is it easily manufactured into multidetector configurations. 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  $\mu\text{m}$ . 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.

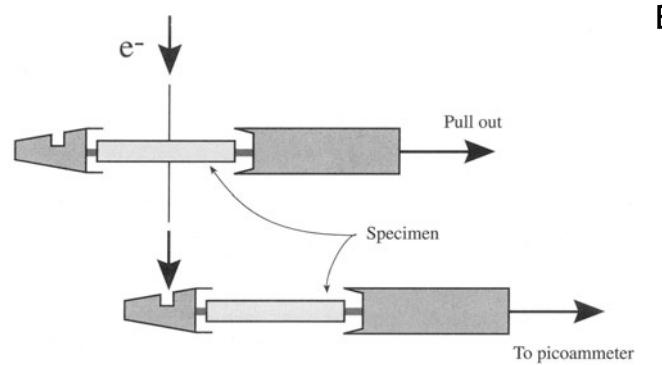
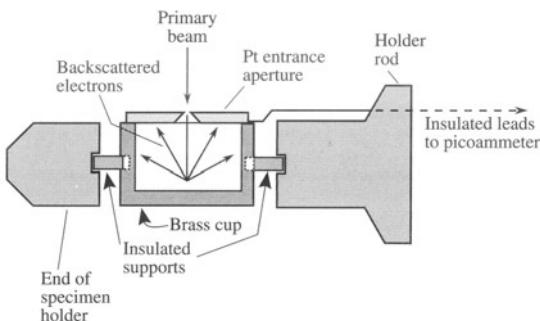
The major disadvantages are the speed and the expense. These devices can also be used as two-dimensional arrays for parallel-collection electron energy-loss spectrometers rather than the more conventional one-dimensional silicon diode arrays described in Chapter 37.

### 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.

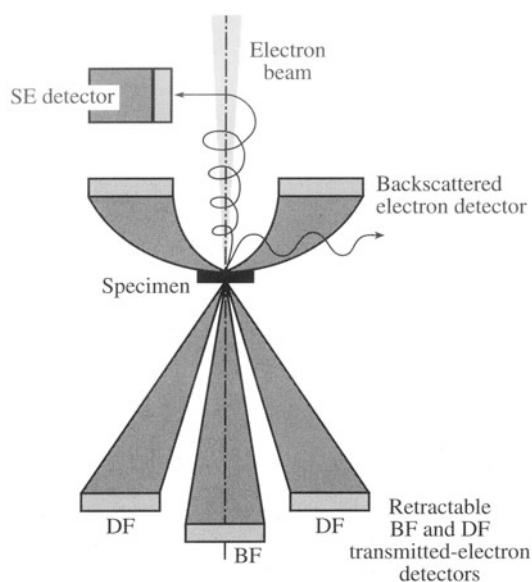


**Figure 7.4.** (A) Schematic diagram of a Faraday cup in the end of a side-entry specimen holder. The entrance aperture has to be found using SEs or BSEs. In (B) the holder is retracted slightly so the electrons fall into a cup on the tip of the holder. The electron current is measured as it goes to 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.



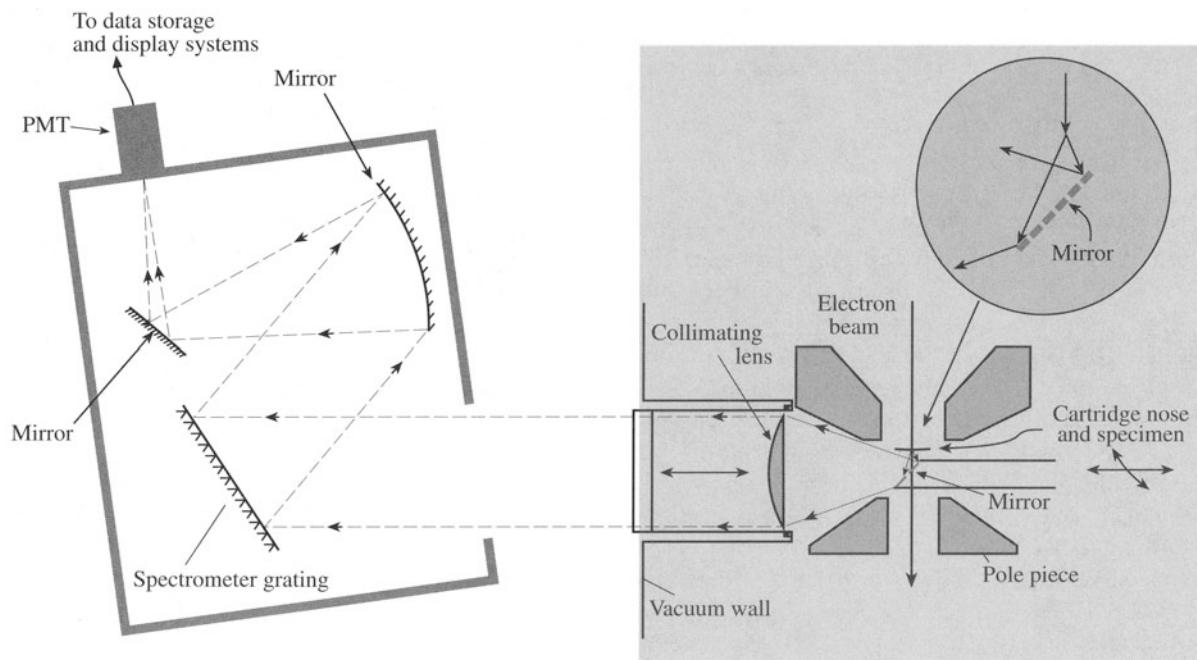
**Figure 7.5.** The various electron detectors in a STEM. Scintillator–PM detectors are invariably used for SE detection and semiconductor detectors for the BSE. The on-axis and annular forward-scattered detectors may be either type, depending on the microscope.

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 semi-

conductor 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  $\mu\text{m}$  compared to about 4  $\mu\text{m}$  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  $\mu\text{m}$ . Despite this degradation we still have more than  $10^7$  picture elements or pixels available to store information in a 10 cm  $\times$  10 cm image (Kodachrome film has the silver halide equivalent of  $1.8 \times 10^7$  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  $> 2.5 \times 10^7$  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  $2 \times 10^3$  lines, giving a total of  $2 \times 10^6$  pixels in the  $10\text{ cm} \times 10\text{ cm}$  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  $10\text{ cm} \times 10\text{ cm}$  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  $\ll 1\text{ ms}$  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

---

8.1. Vacuums .....	119
8.2. Roughing Pumps .....	119
8.3. High/Ultra-High Vacuum Pumps .....	120
8.3.A. Diffusion Pumps .....	120
8.3.B. Turbomolecular Pumps .....	120
8.3.C. Ion Pumps .....	121
8.3.D. Cryogenic (Adsorption) Pumps .....	121
8.4. The Whole System .....	122
8.5. Leak Detection .....	123
8.6. Contamination: Hydrocarbons and Water Vapor .....	123
8.7. Specimen Holders and Stages .....	124
8.8. Side-Entry Holders .....	124
8.9. Top-Entry Holders .....	125
8.10. Different Types of Holders .....	125

## 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  $\sim 10^{-5}$  Pa, which compares with atmospheric pressure of  $\sim 10^5$  Pa. It is quite remarkable that we can transfer a specimen into the TEM, reducing the ambient pressure at its surface by 10 orders of magnitude in a matter of a few seconds. This rapid transfer is a testament to the skills of TEM designers, and particularly the construction of the specimen holder and the airlock system. Specimen holders are the physical contact between you and your specimen across this extraordinary vacuum range. 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

## 8.1. 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  $\sim 10^5$  Pa and so

One Torr is  $\sim 130$  Pa, 1 Pa is  $7.5 \times 10^{-3}$  Torr.

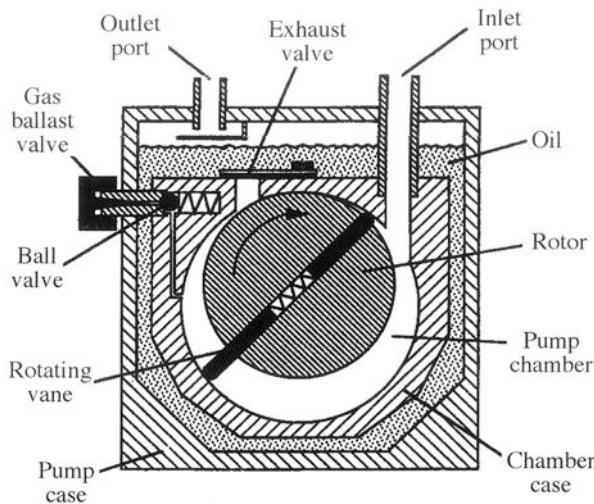
We'll mainly use Pa, but since Torr is still very common terminology, we'll occasionally put highly approxi-

mate 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 ( $\sim 1$  and  $10^{-3}$  Torr) are rough vacuums,  $0.1\text{--}10^{-4}$  Pa ( $\sim 10^{-3}\text{--}10^{-6}$  Torr) are low,  $10^{-4}\text{--}10^{-7}$  Pa ( $\sim 10^{-6}\text{--}10^{-9}$  Torr) are high vacuums (HV). If the pressure is  $<10^{-7}$  Pa ( $\sim 10^{-9}$  Torr) you have an ultrahigh vacuum (UHV). These are approximate, not standardized definitions. A typical modern TEM has a pressure inside the column of  $\sim 10^{-7}$  Torr ( $1.3 \times 10^{-5}$  Pa), which is in the HV range. UHV TEMs operate below  $10^{-7}$  Pa and the gun region of an FEG TEM operates at  $\sim 10^{-9}$  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  $<\sim 0.1$  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



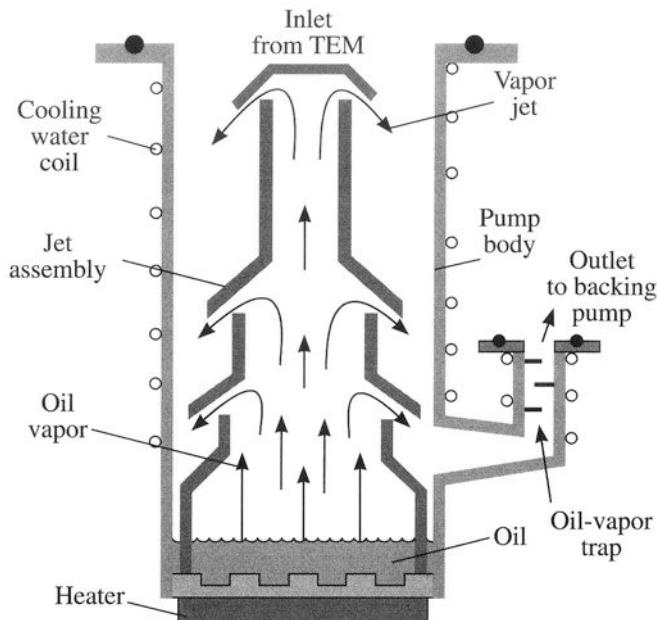
**Figure 8.1.** A mechanical pump for roughing vacuums. The eccentric motion of the pump creates a vacuum in the RH side when it rotates and the vacuum sucks air into the inlet valve. As the cylinder rotates further, it cuts off the inlet and forces the air through the outlet on the LH side, creating 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.

chamber and expels it through an exit valve, as shown in Figure 8.1. Such pumps are very reliable, relatively inexpensive, noisy and dirty, and only lower the pressure to  $\sim 10^{-1}$  Pa ( $\sim 10^{-3}$  Torr). Mechanical pumps should be housed outside your TEM room, and connected to the column through a line that doesn't transmit their vibration. These pumps use a hydrocarbon oil as a medium. If you have such a pump, the line from the pump to the vacuum should contain a foreline trap to condense out oil vapor before it is deposited in the column. Also, the exhaust line from the pump must 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



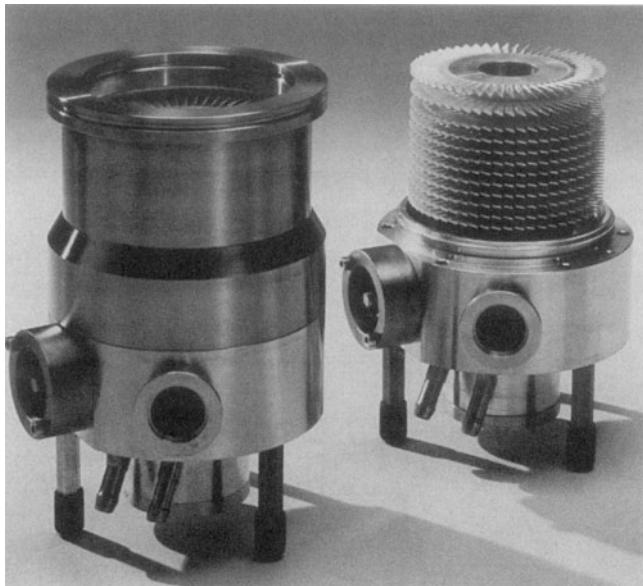
**Figure 8.2.** Principles of diffusion pump operation. A heater plate at the base of the pump boils synthetic oil. The expansion of the oil vapor on boiling creates a pressure, which forces the vapor up the central column and out of several holes. The stream of oil vapor pulls gas molecules out of the 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.

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<sub>2</sub> 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<sub>2</sub> to maintain a clean system.

Diffusion pumps are capable of very efficient pumping from  $\sim 10^{-1}$  to  $\sim 10^{-9}$  Pa ( $10^{-11}$  Torr) and, if properly trapped, will provide a clean UHV system that is very reliable. The VG series UHV DSTEMs use only oil diffusion pumps to attain UHV conditions.

#### 8.3.B. Turbomolecular Pumps

Turbomolecular pumps, or turbopumps, as the name implies, use a turbine to force gases from the microscope.



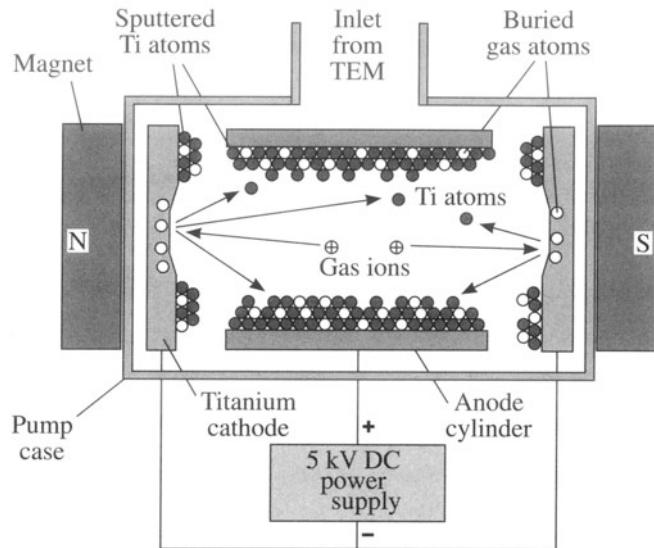
**Figure 8.3.** A turbopump (with and without its casing) which is nothing more than a small turbine that rotates at high speed. Like a jet turbine it pulls air in at the front end and forces it out of the back. The blades are designed like airfoils to enhance the flow of gas through the system.

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 cryo-transfer 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



**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.

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<sub>2</sub> 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.

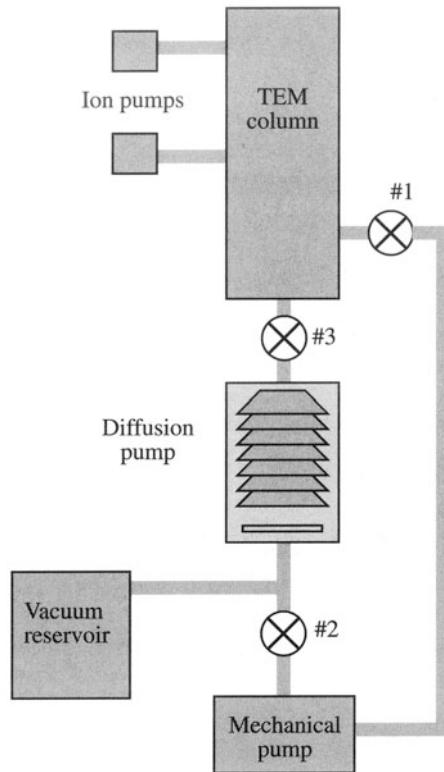
## 8.4. THE WHOLE SYSTEM

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



**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.

heater can be safely switched on without oxidizing. So close #1, open #2, and close #3.

When the diffusion pump is warmed up, you rough out the column: open #1, close #2 and #3, until the column is at a low enough pressure that the diffusion pump can be used.

At this point, close #1, open #2 and then #3, so the diffusion pump is open to the TEM and 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  $>100^{\circ}\text{C}$  (or  $>150\text{--}200^{\circ}\text{C}$  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  $>100^{\circ}\text{C}$  in a heating holder or with a halogen lamp in the prepump chamber, or cooling the specimen to liquid-N<sub>2</sub> 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<sub>2</sub>O 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<sub>2</sub> 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.

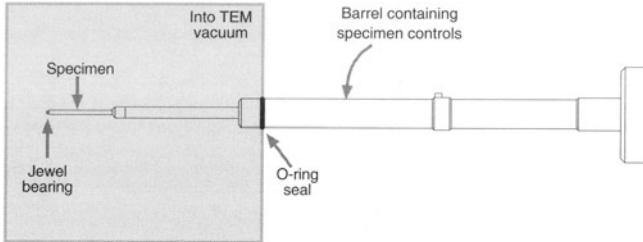
The actual cup that holds your specimen is either 2.3 mm or 3.05 mm in diameter, so the specimen disk or support grid has to be the same dimension, as we'll see in Chapter 10. The reasons for these dimensions are again partly historical. In the top-entry holder the specimen and part of the holder fit 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 specimen is accomplished from outside the column via controls within the rod.

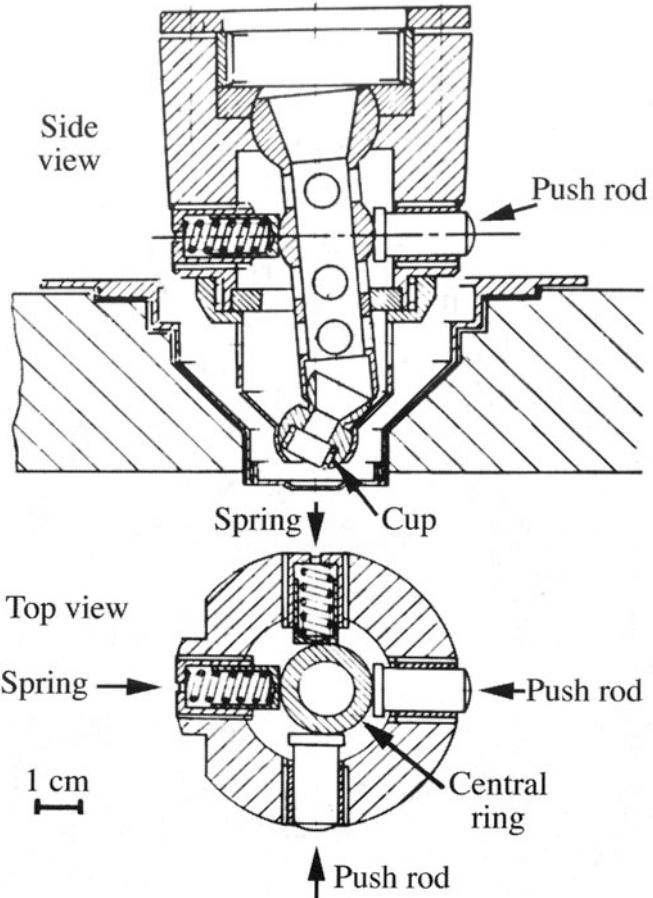
ation of X-rays that would interfere with microanalysis.

- *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.

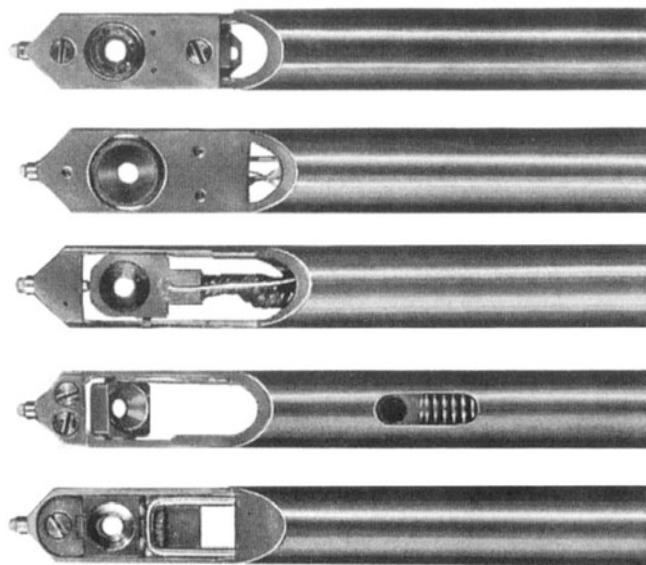


**Figure 8.7.** Cross section of a top-entry holder. The cartridge has a cone shape which fits into the tapered bore of the objective lens pole-piece. The specimen sits in a cup at the base of a column through the cone down which the incident beam travels. Simple manipulations such as tilting or rotating require complex micromechanical design, since the specimen is at the base of the cartridge and completely surrounded by the pole-piece. 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

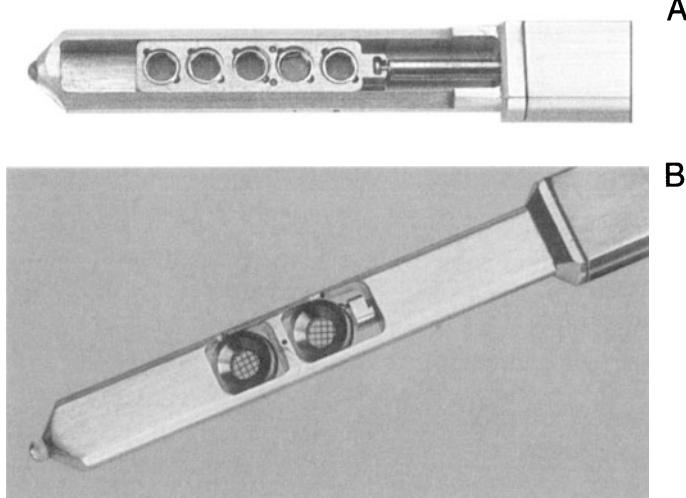
One feature of TEMs which may surprise you if you are a new user is the wide variety of holders which is available. Figure 8.8 shows illustrations of some different designs for the side-entry holder:

- *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

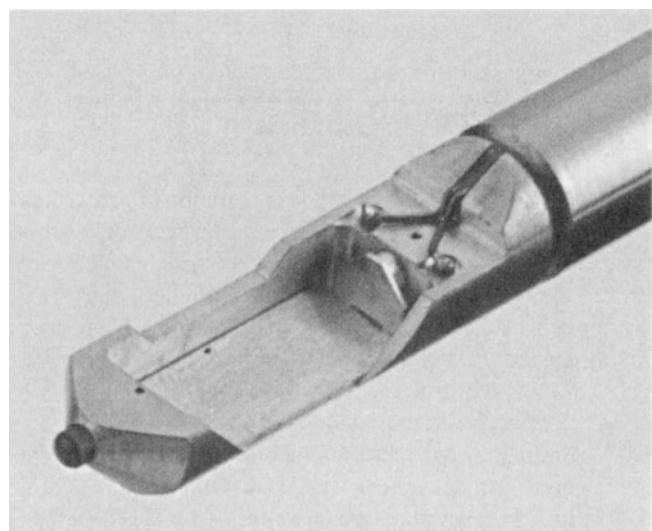


**Figure 8.9.** Multiple-specimen holders. (A) Five-specimen single-tilt and (B) two-specimen double-tilt.

will allow you to sample more of your material at one time (Figure 8.10).

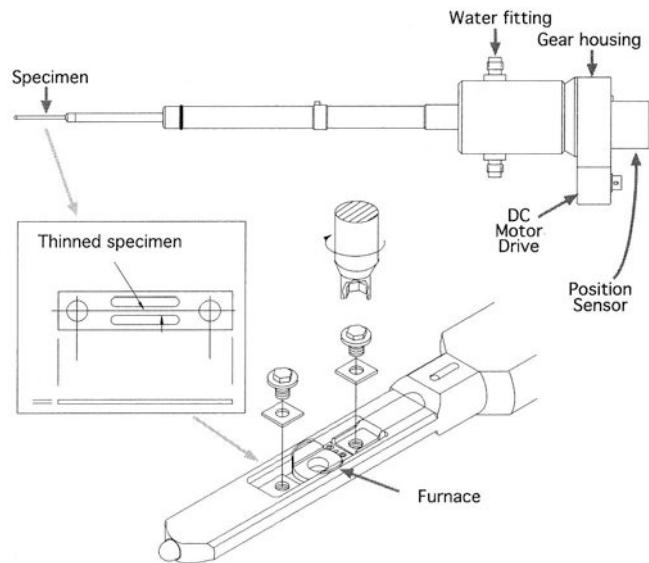
So don't always think that you are limited to 3-mm specimens!

- **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

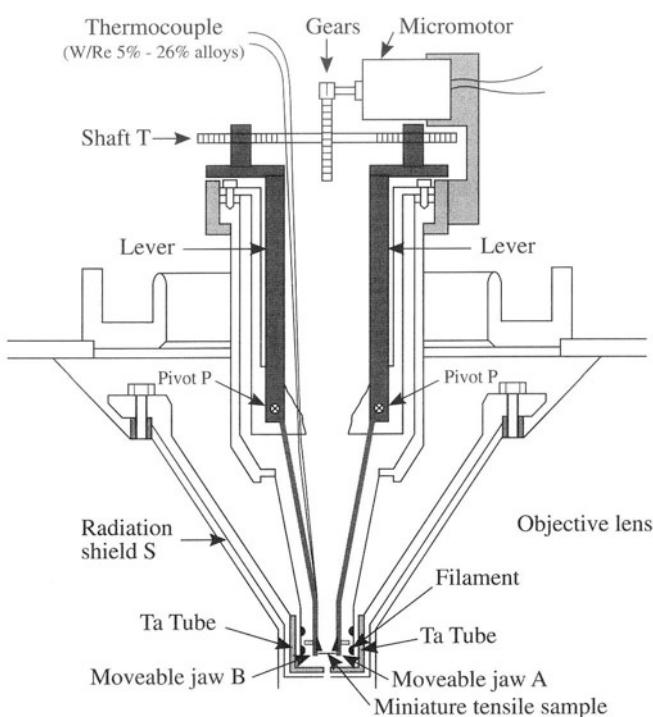


**Figure 8.10.** A bulk holder for large specimens.

- 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  $\sim 1300^{\circ}\text{C}$ , 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<sub>2</sub> 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.

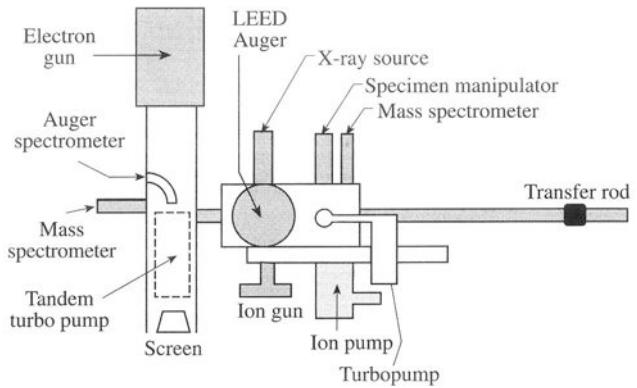


**Figure 8.12.** A top-entry, heating-straining holder which can be used at temperatures up to 2300 K in a 3-MV HVEM.

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



**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.

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

---

9.1. The Illumination System . . . . .	133
9.1.A. TEM Operation Using a Parallel Beam . . . . .	133
9.1.B. Convergent-Beam (S)TEM Mode . . . . .	133
9.1.C. Translating and Tilting the Beam . . . . .	135
9.1.D. Alignment . . . . .	136
9.1.E. Condenser Lens Defects . . . . .	137
9.1.F. Calibration . . . . .	138
9.2. The Objective Lens and Stage . . . . .	139
9.3. Forming Diffraction Patterns and Images: The TEM Imaging System . . . . .	140
9.3.A. Selected-Area Diffraction . . . . .	140
9.3.B. Bright-Field and Dark-Field Imaging . . . . .	142
9.3.C. Centered DF Operation . . . . .	143
9.4. Forming Diffraction Patterns and Images: The STEM Imaging System . . . . .	144
9.4.A. Bright-Field STEM Images . . . . .	144
9.4.B. Dark-Field STEM Images . . . . .	145
9.4.C. Annular DF Images . . . . .	145
9.4.D. Magnification in STEM . . . . .	146
9.5. Alignment and Stigmation . . . . .	147
9.5.A. Lens Rotation Centers . . . . .	147
9.5.B. Correction of Astigmatism in the Imaging Lenses . . . . .	147
9.6. Calibration of the Imaging System . . . . .	148
9.6.A. Magnification Calibration . . . . .	148
9.6.B. Camera-Length Calibration . . . . .	150
9.6.C. Rotation of the Image Relative to the Diffraction Pattern . . . . .	151
9.6.D. Analysis of TEM Images and Diffraction Patterns . . . . .	151
9.7. Other Calibrations . . . . .	151

## 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

## 9.1. 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 ( $20,000\times$ – $100,000\times$ ). 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 ( $\alpha$ ) 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,  $\alpha$  under these conditions is  $<10^{-4}$  rads ( $0.0057^\circ$ ), which is effectively a parallel beam.

In TEMs used for generating the very small electron beams we need in STEM and AEM, the upper pole-piece, 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?

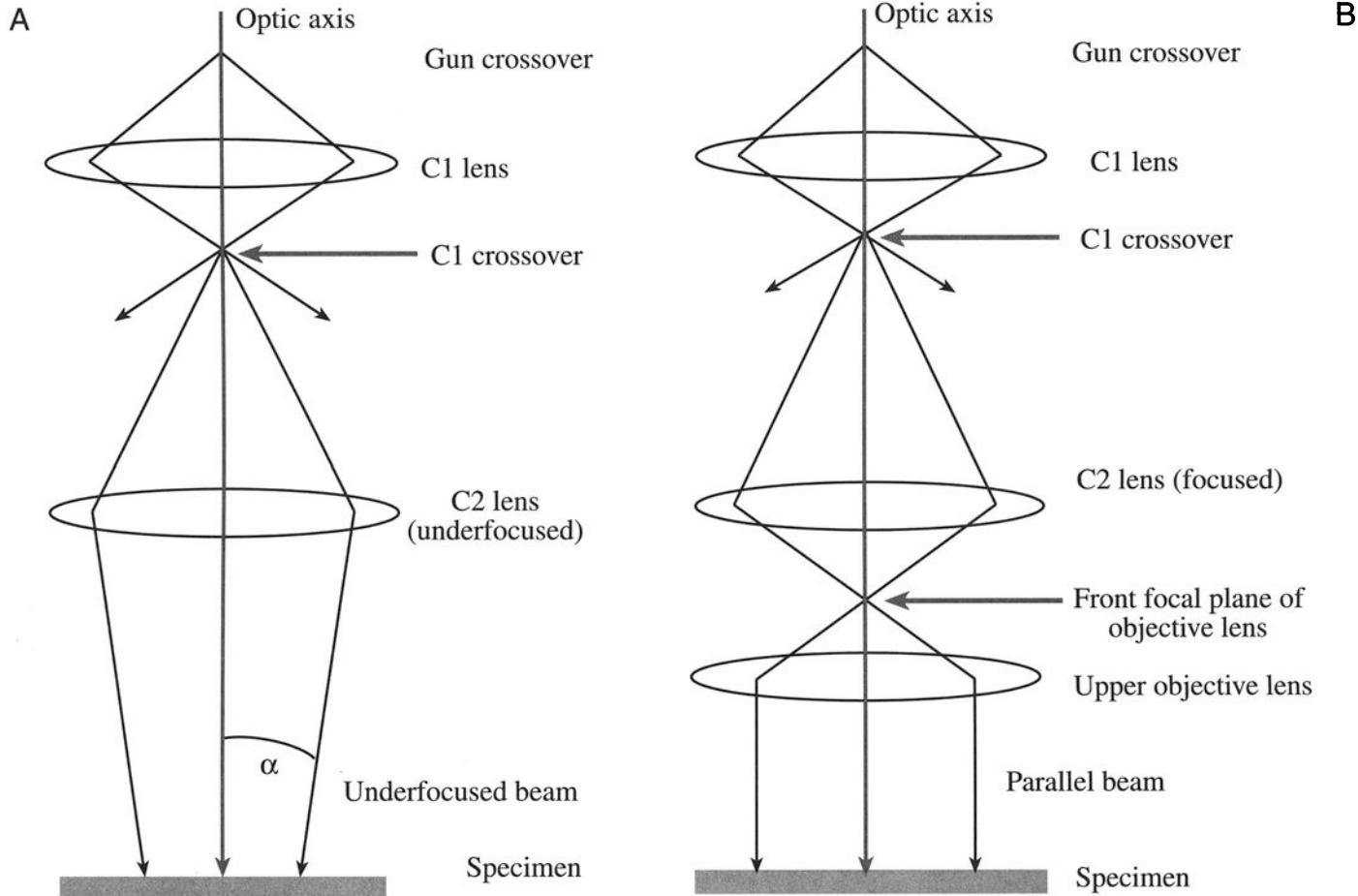
When the beam is parallel, it is as coherent as possible. We'll see in Chapter 18 that parallel illumination is essential to get the sharpest diffraction patterns as well as the best image contrast. (It is also usually assumed in the interpretation of our images that the beam is parallel.) Usually you 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.

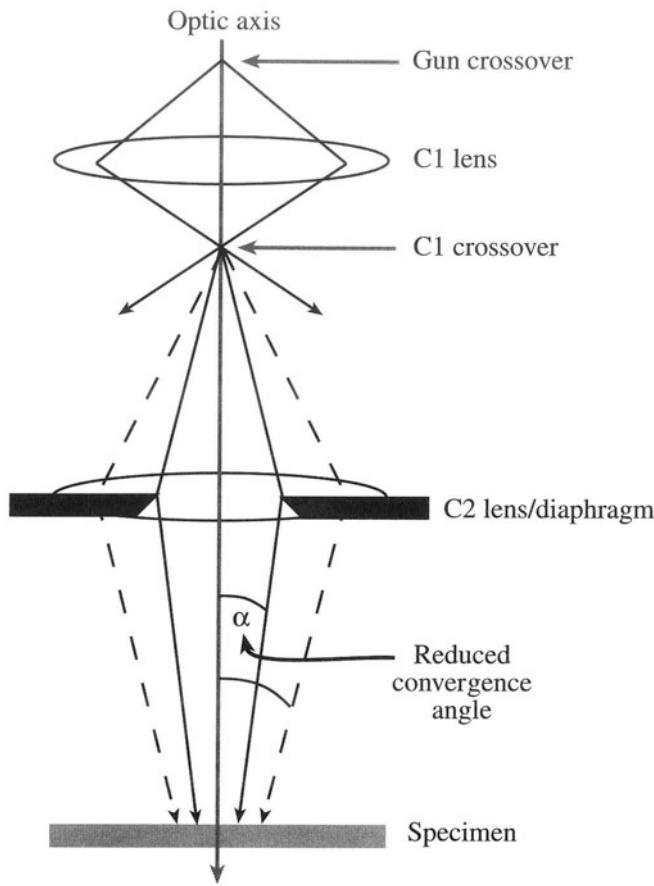
focused rather than underfocused, and you form an image of the C1 crossover at the specimen, as shown in Figure 9.3. This is the condition under which you can view the source image to adjust its saturation or to measure the dimensions of the beam. When C2 is focused like this, the beam is at its least 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.

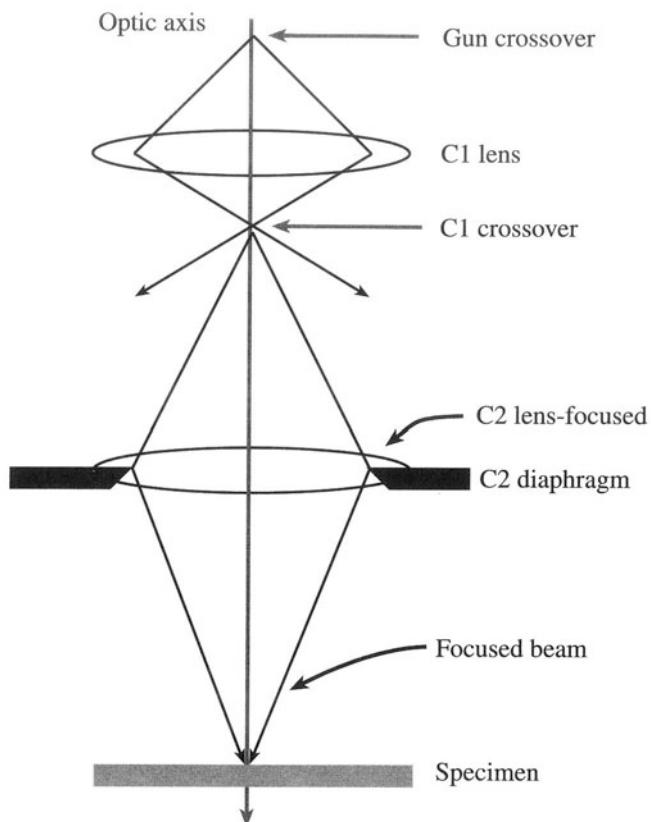
Now, unless you have an FEG, it isn't possible to use just the C1 and C2 lenses as in Figure 9.3 to converge the beam to as small a probe as you would like (<10 nm). This is because the C1 and C2 lenses can't demagnify the gun crossover sufficiently. So the usual solution is to con-



**Figure 9.2.** Effect of the C2 aperture on the parallel nature of the beam: a smaller aperture creates a more parallel beam.

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 ( $\alpha$ ) of the beam on the specimen. As was the case for parallel-beam mode, a smaller C2 aperture gives a smaller  $\alpha$ . 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

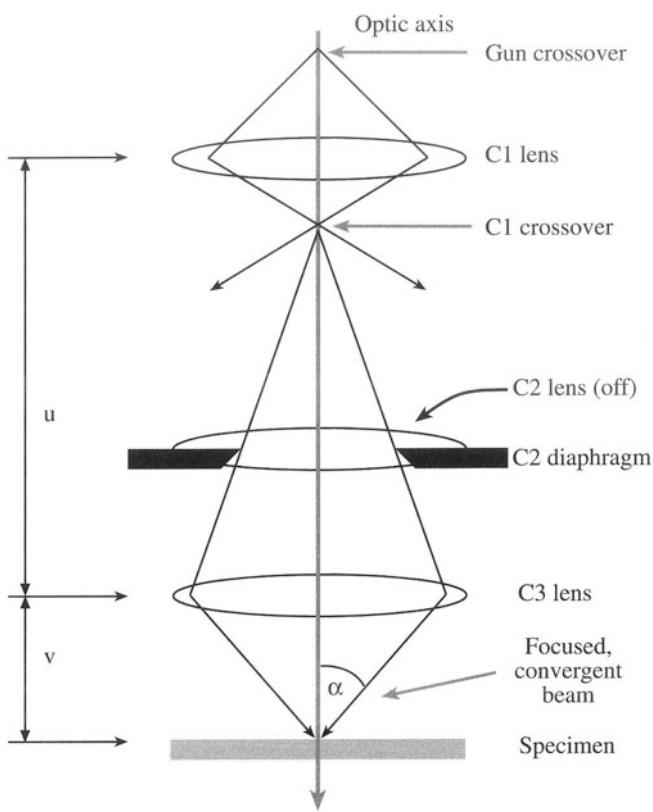


**Figure 9.3.** A focused C2 lens illuminates a small area of the specimen with a nonparallel beam.

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



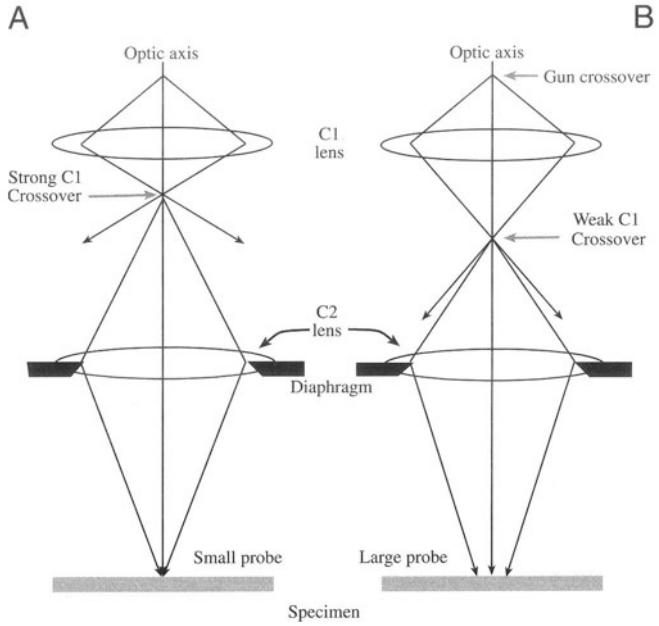
**Figure 9.4.** Use of the objective polepiece as a third condenser lens (also called a condenser-objective, or C3, lens) gives the smallest possible probe and large convergence angles. The large  $u/v$  ratio gives the maximum demagnification of the image of the gun crossover.

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.

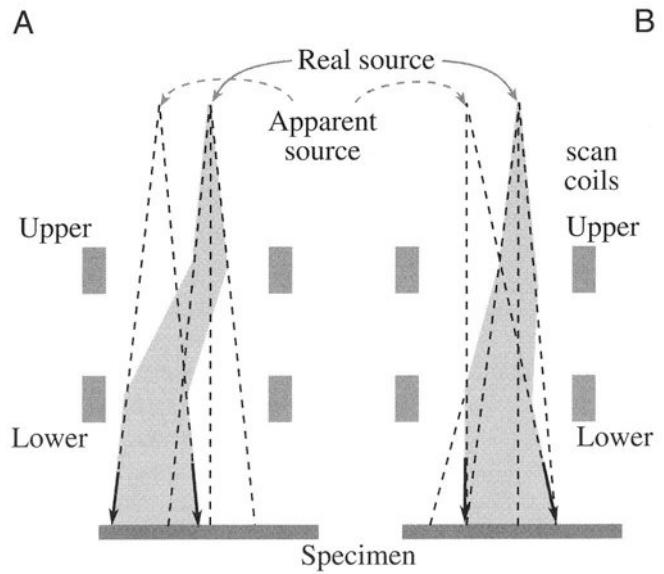
When we create a scanning beam for STEM imaging, the beam must always move parallel to the optic axis. Such scanning is accomplished by tilting the beam twice with two sets of scan coils, one above the other, to ensure that the beam crosses the optic axis at the front focal plane of C3. Then, wherever 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 9.5.** 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.



**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.

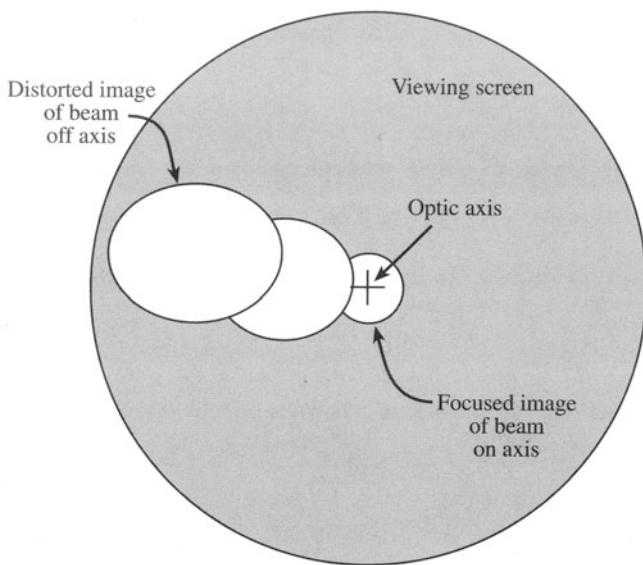
Instructions for alignment vary for different TEMs so we'll simply describe the principles. Even if you won't be doing the alignment you will want to check that it is correctly aligned; you can recognize if the wheels are not aligned on your car and you know it is important to balance them for best performance 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

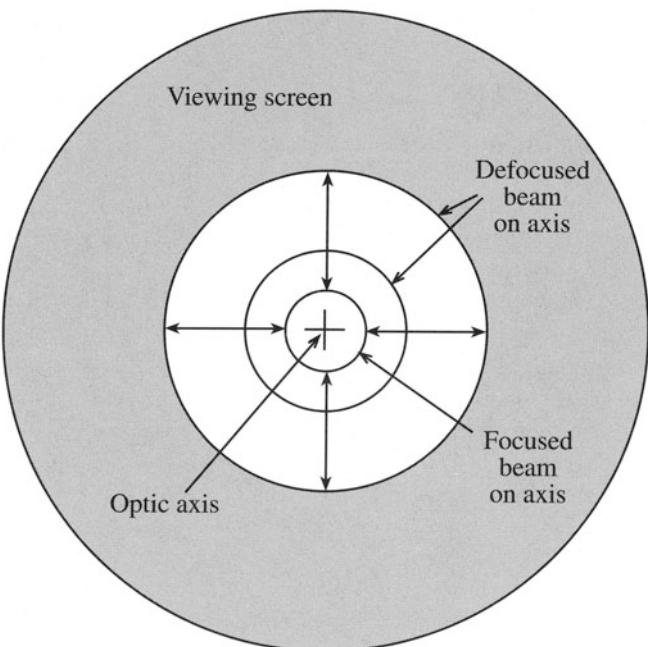
You must have the C2 aperture accurately centered on the optic axis of the TEM. If the aperture is misaligned the image of the beam on the screen moves off axis and distorts as you underfocus or overfocus C2, as shown in Figure 9.7. To align the aperture on axis, you need first to overfocus C2 so the 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  $\alpha_{\text{opt}} = 0.77 \lambda^{1/4}/C_s^{1/4}$ . 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.

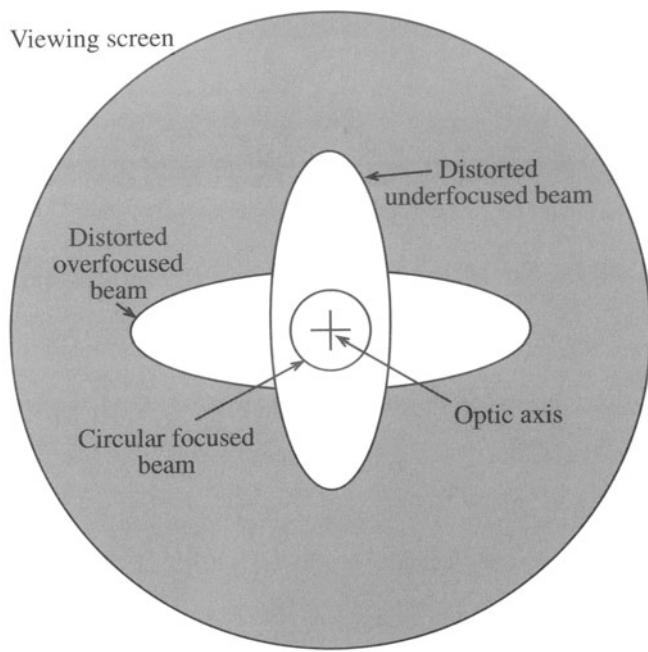
## Chromatic Aberration

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



**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.

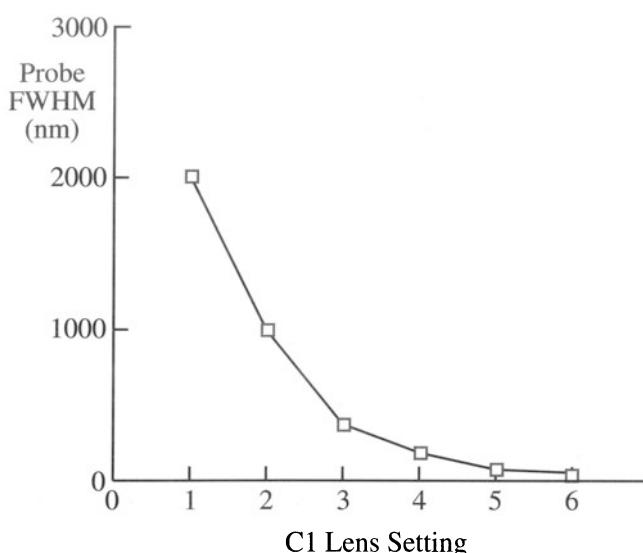
the astigmatism (i.e., the beam appears elliptical). Then adjust the stigmators so the image appears circular. Now underfocus the beam and repeat the correction. Repeat the whole over/underfocus procedure iteratively until the image of the beam remains circular as you expand and contract it on the screen with the C2 lens.

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

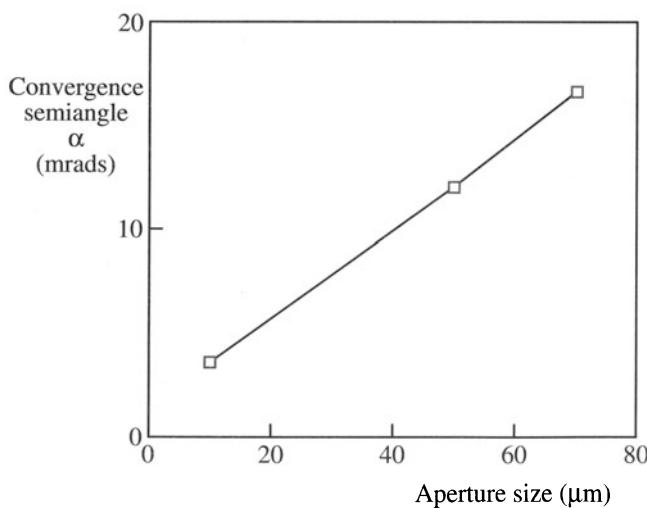


**Figure 9.10.** Calibration of the illumination system requires determining the variation of the probe size with C1 lens strength.

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  $\alpha$ , as we also discussed in Chapter 5 when we were determining the gun brightness.

- We measure the total convergence angle  $2\alpha$  from a CBED pattern (Figure 5.8).
- We increase  $2\alpha$  by increasing the C2 aperture size (Figure 9.11).



**Figure 9.11.** Variation of the beam-convergence semiangle,  $\alpha$ , 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  $2\alpha$ . (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  $\alpha$  and then you have to calibrate  $\alpha$  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  $\pm 30^\circ$  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.

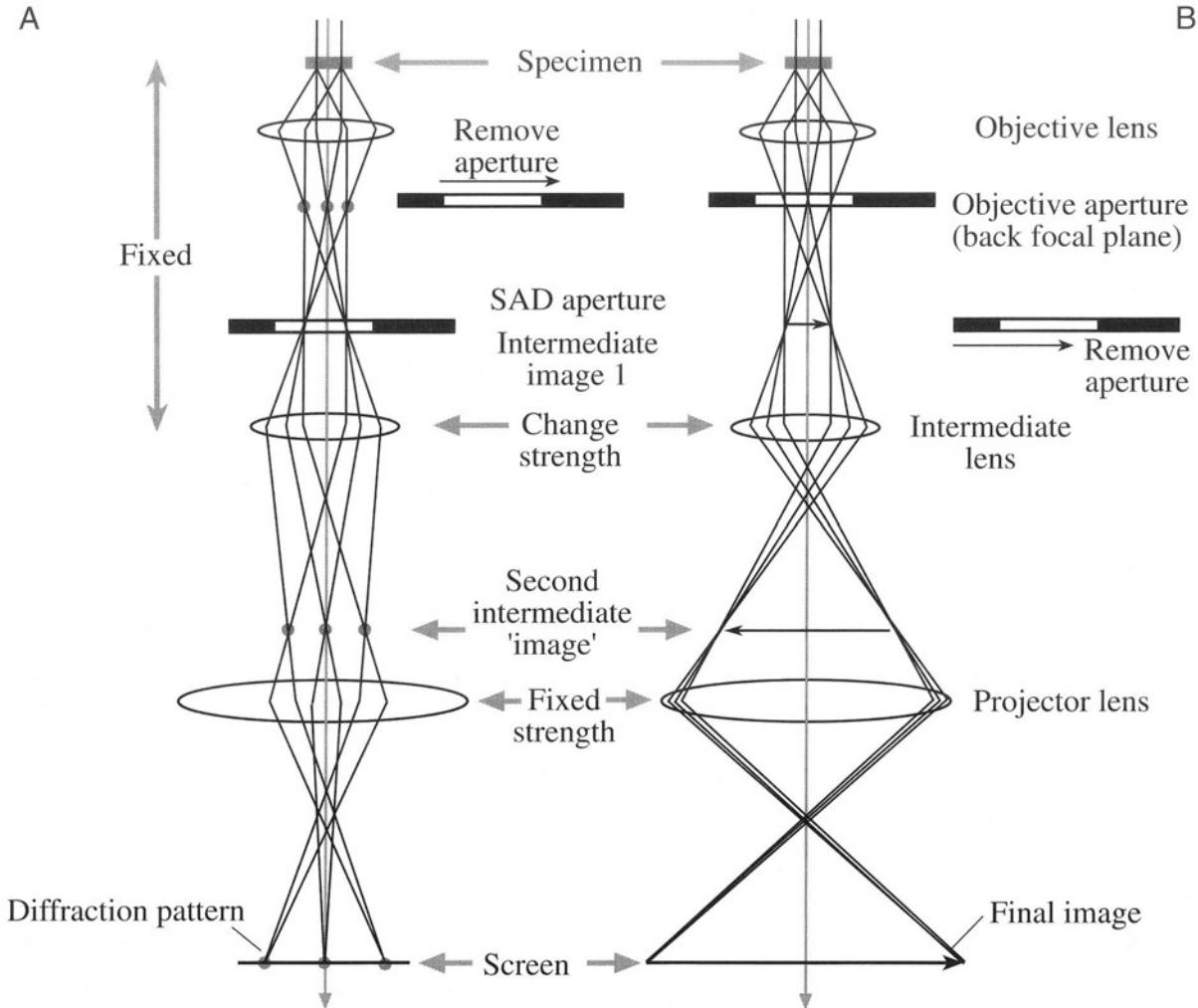
Let's look now at the details of these two fundamental operations from the point of view of the instrument. In subsequent chapters we will discuss how to understand the images and why we form them in the ways we do.

#### 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-



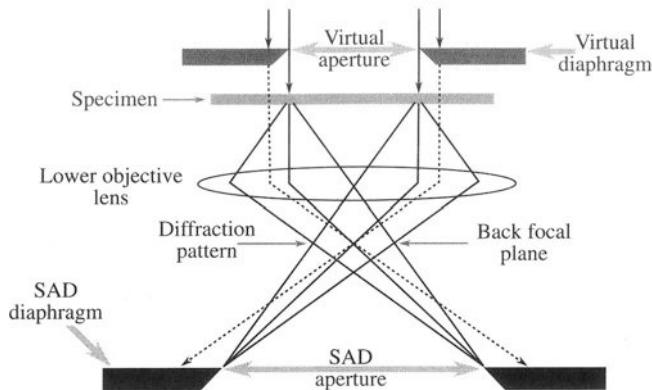
**Figure 9.12.** The two basic operations of the TEM imaging system involve (A) projecting the diffraction pattern on the viewing screen and (B) projecting the image onto the screen. In each case the intermediate lens selects either the back focal plane or the image plane of the objective lens as its object.

lected-area diffraction, or SAD, and was invented by Le Poole (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  $\mu\text{m}$ , and the demagnification back to the plane of the specimen is only about 25 $\times$ , which gives a minimum selected area of  $\sim 0.4 \mu\text{m}$ —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).

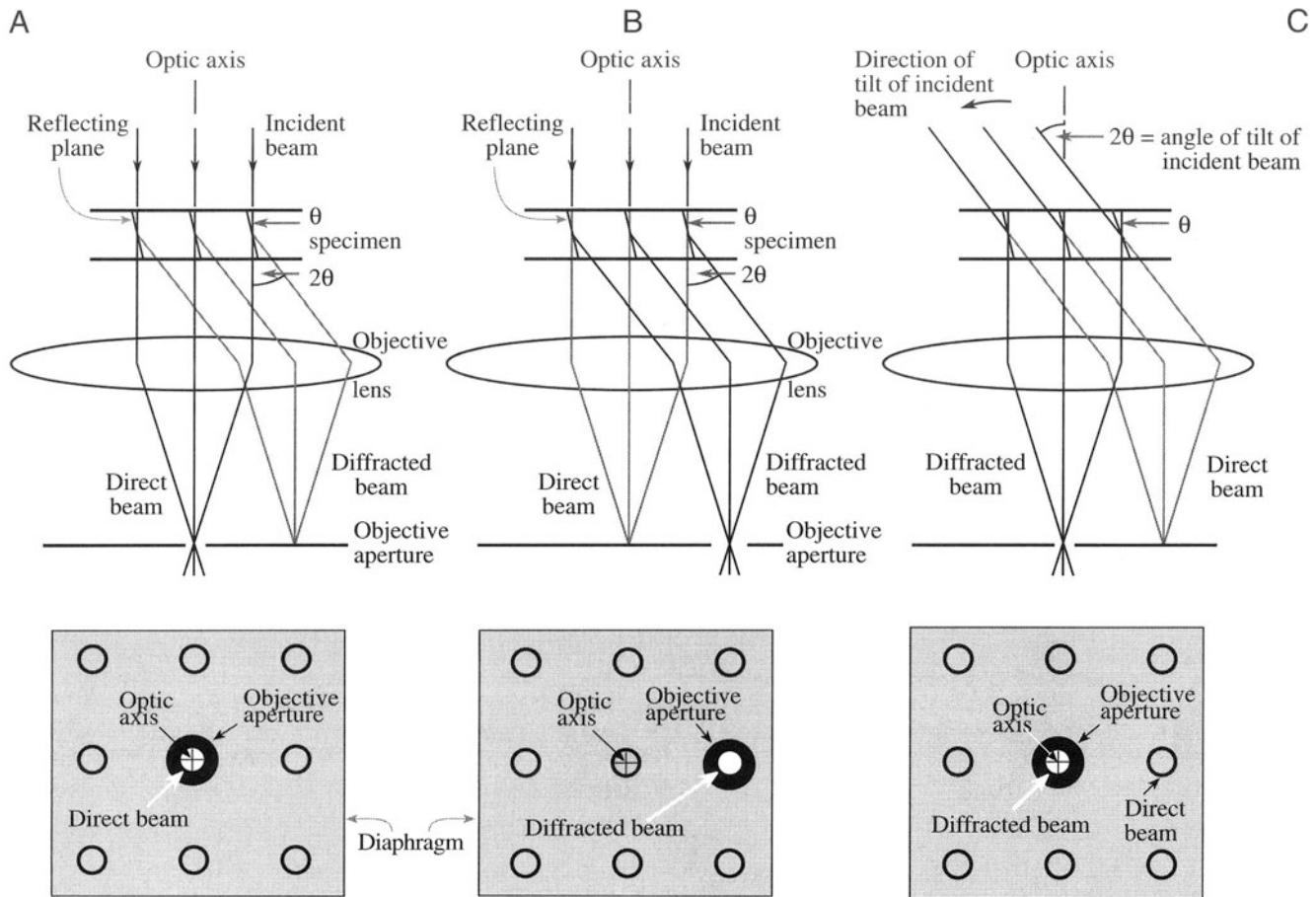
On all other occasions, when the diffraction maxima provide the most important information in the pattern, then we strongly recommend that you use CBED.

### 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 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  $25,000\times$ – $100,000\times$ .

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 ( $\beta$ ) 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.

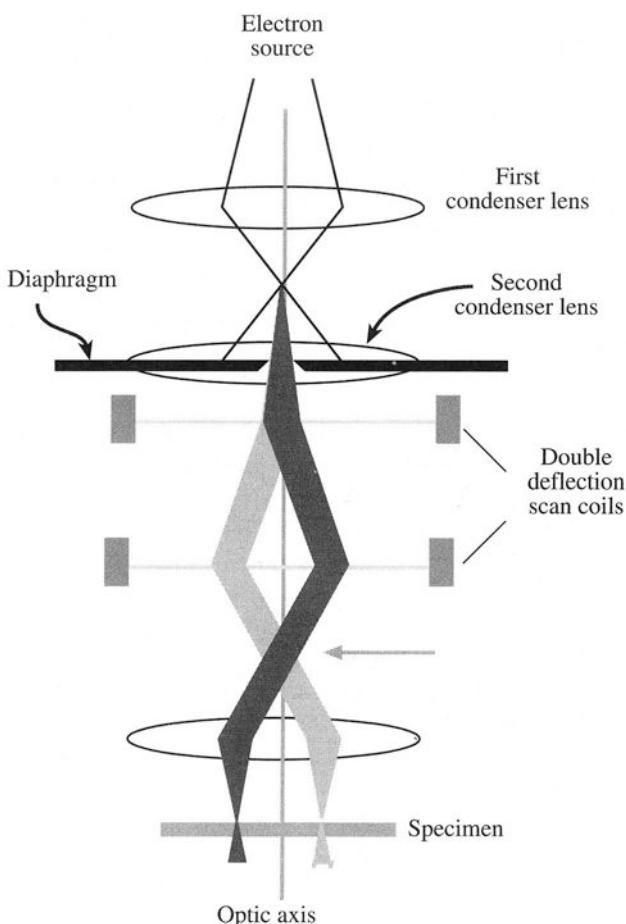
We'll return to BF, CDF, and SAD operations when we discuss specific contrast mechanisms that occur in TEM images in Chapter 22.

## 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



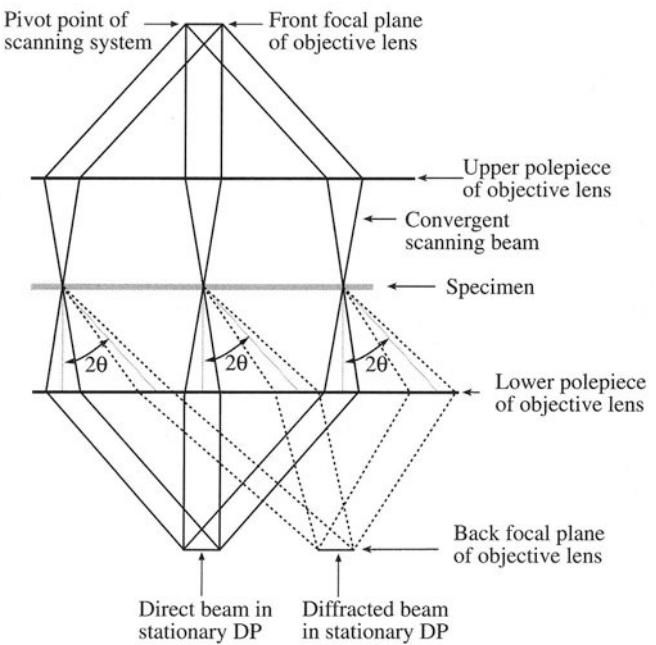
**Figure 9.15.** Scanning the convergent probe for STEM image formation using two pairs of scan coils between the C2 lens (usually switched off) and the upper objective polepiece. The probe remains parallel to the optic axis as it scans.

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.

### 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  $2\theta$  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 speci-

men. 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.

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

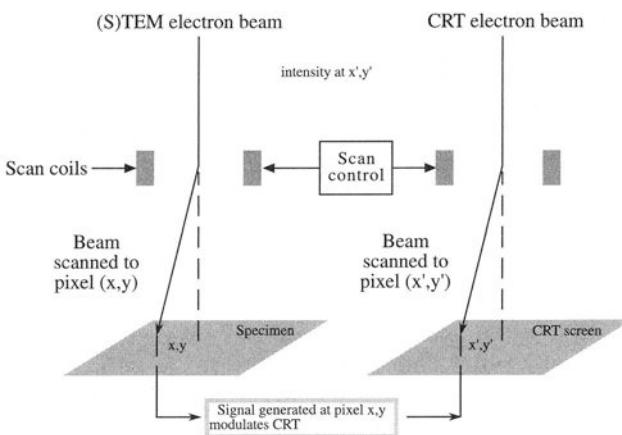
The approach is analogous to that of TEM. We form a DF image by selecting any of the scattered electrons, rather than the direct electrons. Remember, in a TEM we tilt the incident beam so the scattered electrons that we want to form the image travel down the optic axis and are selected by the objective 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.

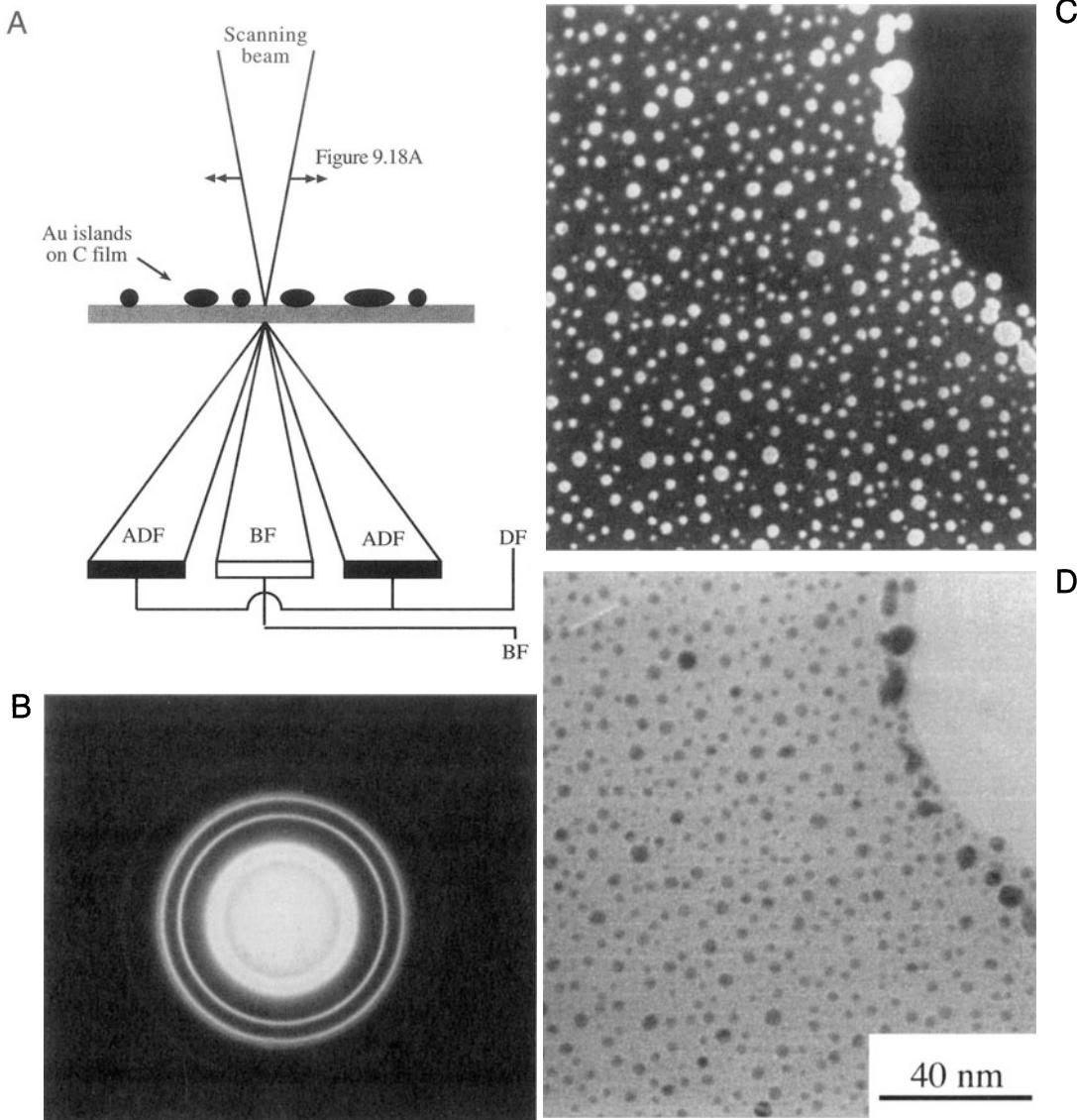
It's simple to do this with the diffraction pattern centering controls, or you could also displace the C2 aperture. The former is to be preferred, since doing the latter misaligns the illumination system.

#### 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 9.17.** 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.



**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  $1\text{ cm} \times 1\text{ cm}$ , and the resultant image is displayed on a CRT with an area  $10\text{ cm} \times 10\text{ cm}$  (which is the standard size of the *record* CRT, though rarely the size of the *viewing* CRT screen) then the magnification is  $10\times$ . If the scan dimension is reduced to  $1\text{ mm}$ , the magnification is  $100\times$ , and so on, up to magnifications in excess of  $10^6\times$ , 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

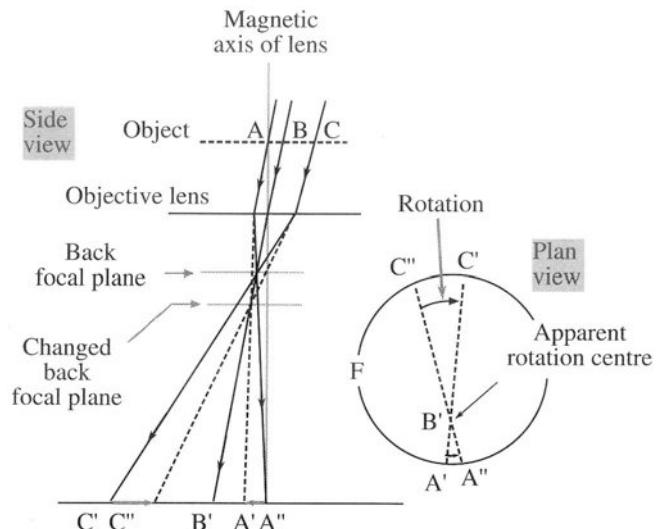
You only need to perform two alignments to ensure that the imaging system is operating correctly. By far the most important is the alignment of the objective lens center of rotation, and the second is the alignment of the diffraction pattern on the optic axis. To get the best out of your TEM, you must 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  $10,000\times$ ), 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  $\sim 100,000\times$  the wobbler may introduce too large a rotation, so you may have to defocus the objective lens manually. The actual steps to do this are instrument-dependent, so consult the manufacturer's handbook. This process is also called "current centering."

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.

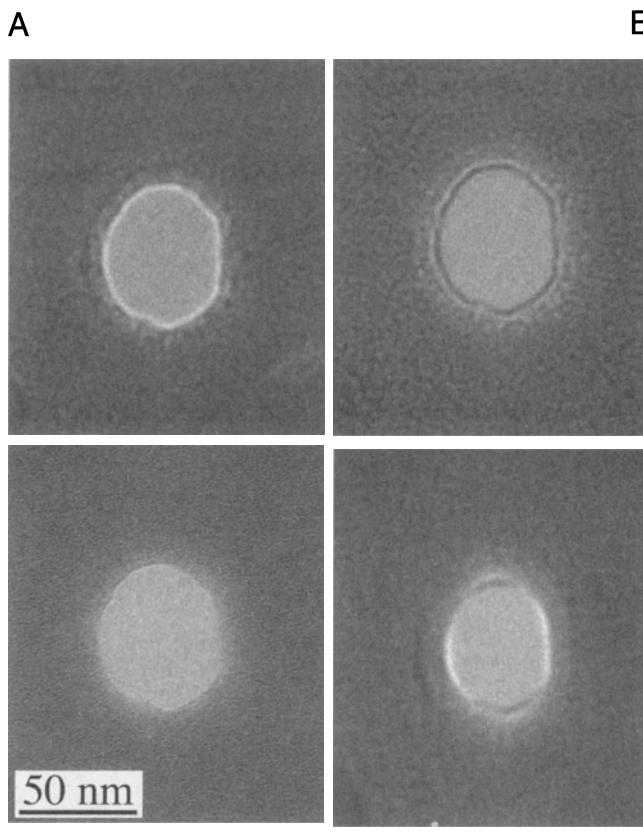
Objective lens astigmatism occurs mainly if the objective aperture is misaligned, so you must carefully center the aperture on the optic axis, symmetrically around the electron beam used to form the BF or DF image. Despite careful centering, however, residual contamination may cause astigmatism and then 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  $90^\circ$  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



**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.

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).

This method of correcting the astigmatism is reasonable at magnifications up to several hundred thousand times. For high-resolution imaging at magnifications of  $>300,000\times$ , 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.

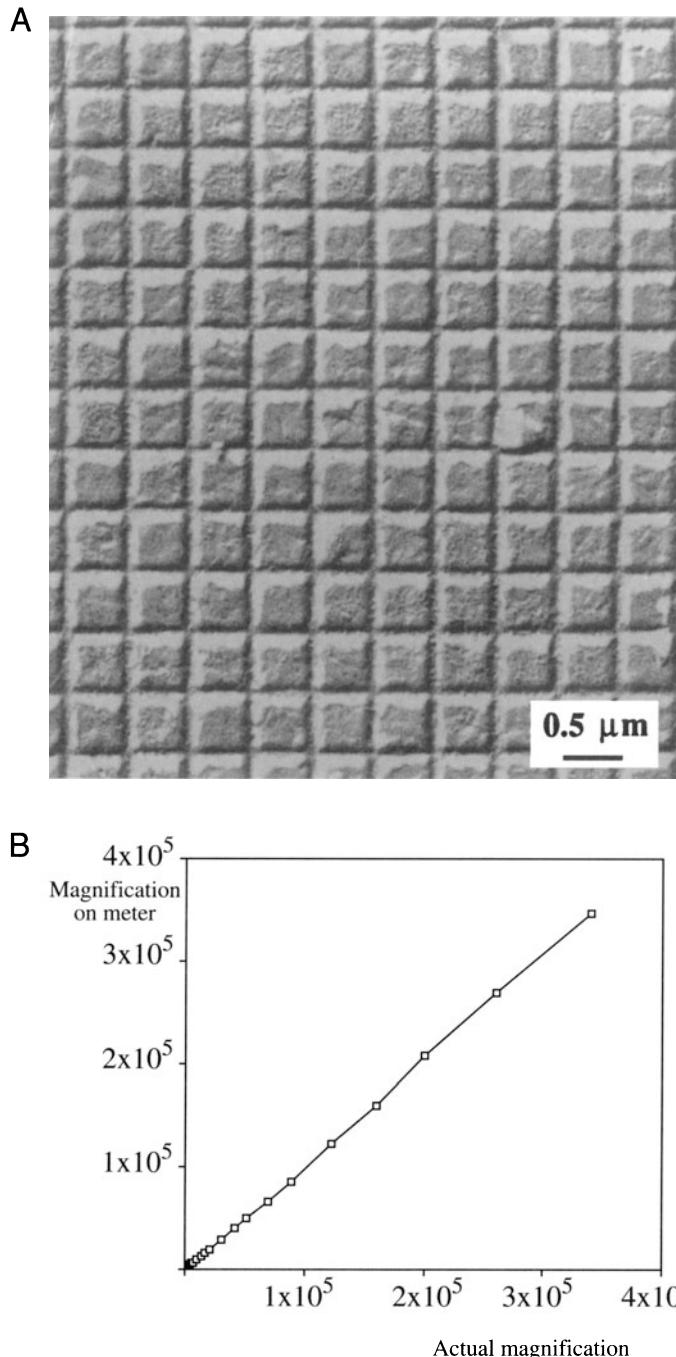
## 9.6. CALIBRATION OF THE IMAGING SYSTEM

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



**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  $0.463 \mu\text{m}$ ), which enables calibration up to magnifications of about  $200,000\times$ . Above this magnification, individual grating spacings are wider than the film. So we then use small latex spheres ( $50\text{--}100 \text{ nm diameter}$ ) 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 ( $0.344 \text{ nm}$ ), 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 ( $<5000\times$ ). 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 ( $\pm 5\%$ ), so long as you note the precautions we've just described. Without a calibration the digital readout is probably no better than  $\pm 10\%$  accurate, and so it is unwise to state magnifications to better than  $\pm 10\%$ . You should be suspicious of any micrographs that you see in the literature with a magnification that is more precise than this (e.g.,  $52,700\times$ ). 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.

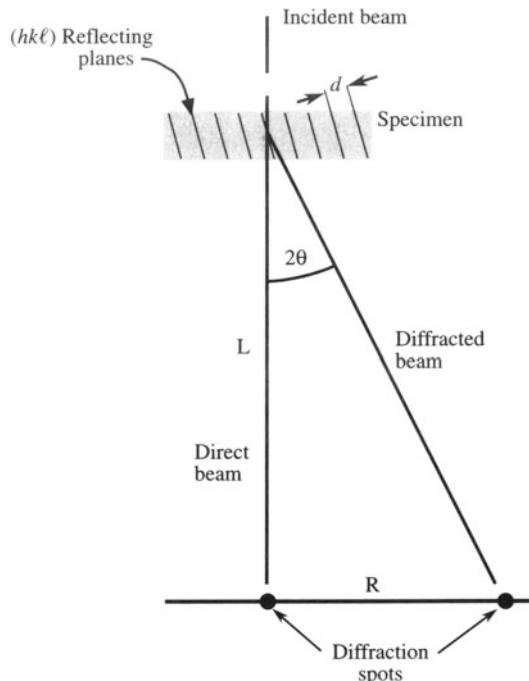
## 9.6.B. Camera-Length Calibration

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  $2\theta$  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

$$\frac{R}{L} = \tan 2\theta \sim 2\theta \quad [9.1]$$



**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

$$Rd = \lambda L \quad [9.2]$$

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  $\lambda$  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 ( $\lambda = 0.0335\text{\AA}$ )**

Camera length Setting	Digital readout (mm)	Measured camera length, $L$ (mm)	Camera constant $\lambda L$ (mm $\text{\AA}$ )
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  $\Phi$  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  $\alpha$ -MoO<sub>3</sub>, 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  $\Phi$  as shown in Figure 9.23B. You can do the same if necessary for different values of  $L$ , which introduce a systematic change in  $\Phi$ . 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

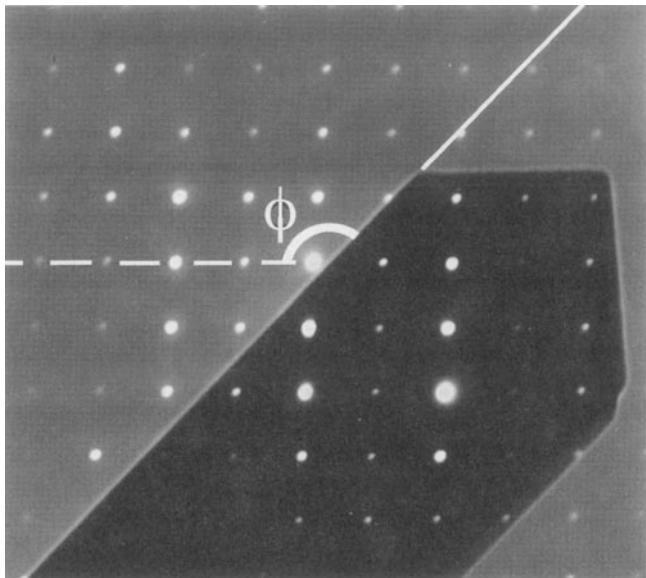
If you don't use a double exposure when comparing images and diffraction patterns (or indeed when comparing directions in any two films), you need a fixed reference line. This line must be independent of slight variations that may arise depending on the film size, how you loaded it, etc. The best reference 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.

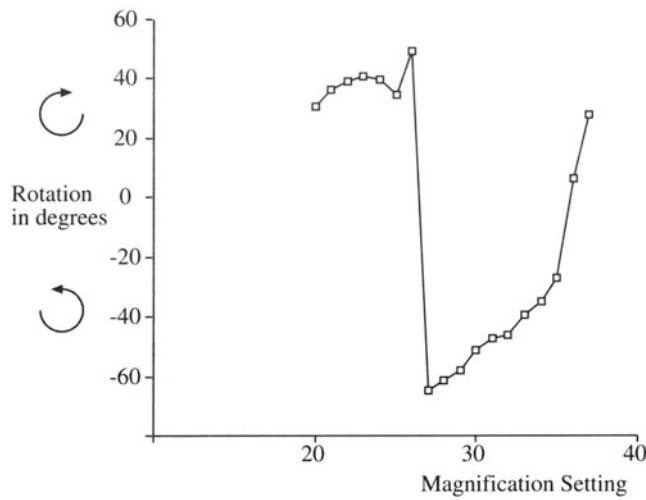
## 9.7. OTHER CALIBRATIONS

**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

A



B



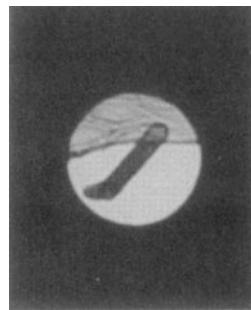
**Figure 9.23.** (A) A double exposure showing the superposition of an image of a  $\text{MoO}_3$  crystal on a diffraction pattern from the same crystal, defining the rotation angle  $\Phi$ . (B) The rotation calibration gives the angle  $\Phi$  between equivalent directions in the image and the diffraction pattern as the magnification is varied. The calibration assumes a constant camera length.

measure the electron wavelength  $\lambda$  by measuring the angle  $\phi$  between two Kikuchi line pairs (see Chapter 19) which intersect at a distance  $R$  from the direct beam

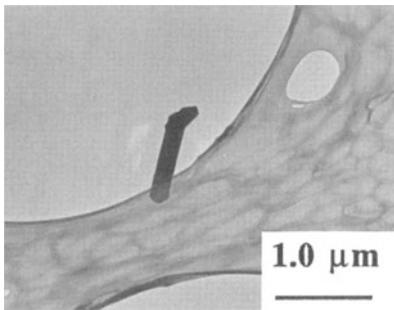
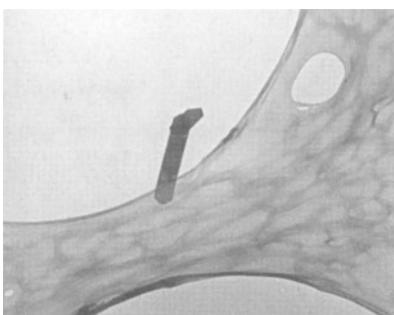
$$\tan \phi = \frac{R}{L} = \frac{\lambda}{d} \quad [9.3]$$

Alternatively, you can match simulations of higher-order Laue-zone lines to experimental lines in CBED patterns (see Chapter 21) and determine which  $\lambda$  gives the best match. Finally, if you can get your X-ray computer system

A



C



D

**Figure 9.24.** Defocused direct beam in a diffraction pattern from  $\text{MoO}_3$  compared with a BF image, showing how to determine if a  $180^\circ$  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  $180^\circ$  inversion is required. In (A) and (B), no rotation occurs between the DP and BF image.

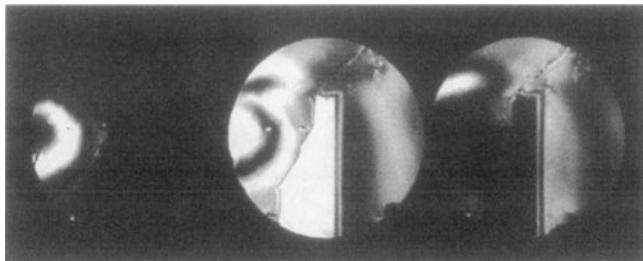
to display the X-ray spectrum (Chapter 34) out to  $E_0$ , the beam energy, the bremsstrahlung intensity vanishes to zero at the exact beam energy (this is called the Duane-Hunt limit).

*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 ( $\Delta f$ ). The two images will be separated by a distance  $\Delta x$  which is related to  $\Delta f$  by

$$\Delta f = \frac{\Delta x}{2Mm\theta} \quad [9.4]$$

where  $M$  is the magnification,  $m$  is the number of focal increments, and  $\theta$  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

---

10.1. Safety .....	157
10.2. Self-Supporting Disk or Use a Grid? .....	158
10.3. Preparing a Self-Supporting Disk for Final Thinning .....	159
10.3.A. Creation of a Thin Slice from the Bulk Sample .....	159
10.3.B. Cutting the Disk .....	159
10.3.C. Prethinning the Disk .....	160
10.4. Final Thinning of the Disk .....	161
10.4.A. Electropolishing .....	161
10.4.B. Ion Milling .....	162
10.5. Cross-Section Specimens .....	165
10.6. Specimens on Grids/Washers .....	166
10.6.A. Electropolishing—The Window Method for Metals and Alloys .....	166
10.6.B. Ultramicrotomy .....	167
10.6.C. Grinding and Crushing .....	167
10.6.D. Replication and Extraction .....	168
10.6.E. Cleaving .....	168
10.6.F. The 90° Wedge .....	169
10.6.G. Lithography .....	169
10.6.H. Preferential Chemical Etching .....	169
10.7. Storing Specimens .....	170

## 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  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ , 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.

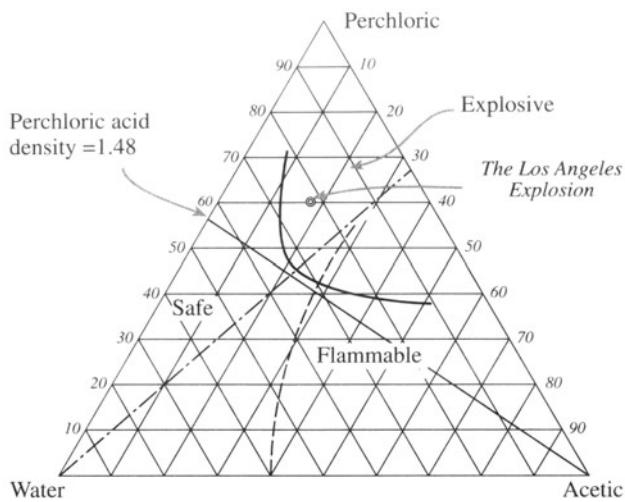
Since these four acids can be so dangerous, we'll mention them specifically, but remember—always seek advice before chemically preparing specimens.

*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<sub>2</sub> 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)

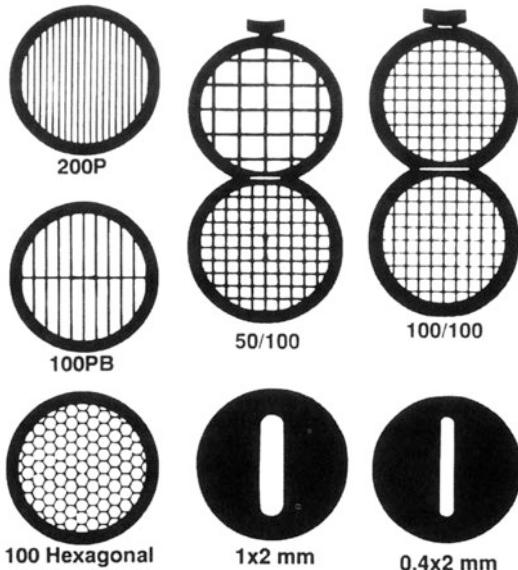
poised skin; HF rapidly penetrates the flesh and dissolves bone.

## 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. How-



**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.

ever, 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  $\mu\text{m}$  and 200  $\mu\text{m}$  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.

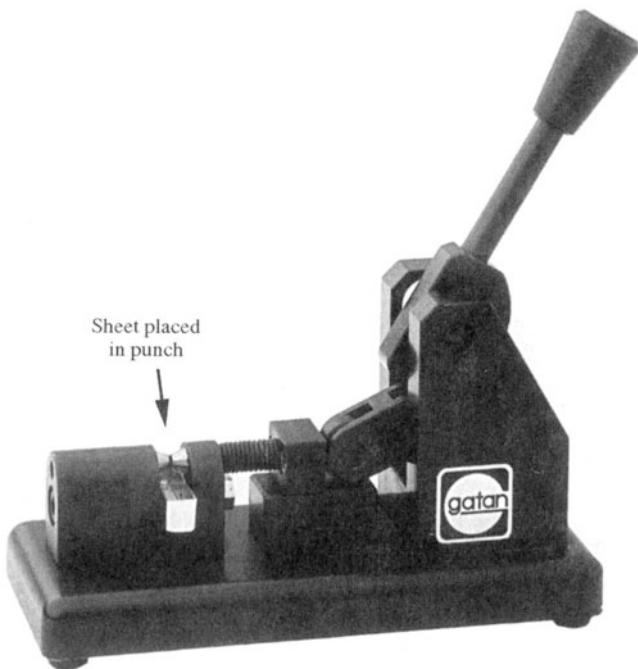
(a) *Ductile materials such as metals.* Usually you don't want to introduce mechanical damage. For example, you want to study the defect structure or the density of defects in processed materials. The ideal method is to use a chemical wire/string saw, a wafering saw (not diamond—the soft metal will dull the blade), or spark erosion (electro-discharge machining) to get a thin slice  $< 200 \mu\text{m}$ . (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  $\text{H}_2\text{O}$ ) 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  $3\times$  their grit size. So a 1- $\mu\text{m}$  abrasive will cause damage to 3  $\mu\text{m}$  below the surface of each side of the specimen. Hence the final disk must be thicker than  $2\times$  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

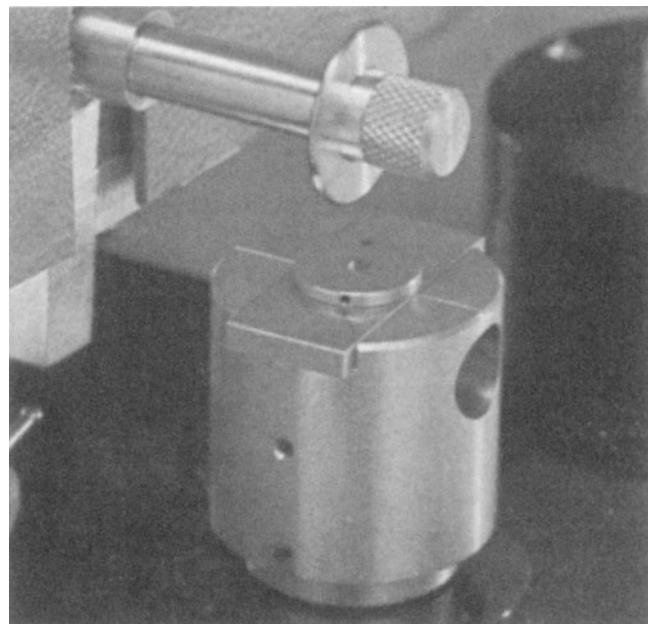
load, precisely determine the thickness of removed material (the depth of the dimple), quickly change the polishing tool, and interrupt the process to remove the sample for closer examination before continuing. The investment is well justified for materials laboratories. One alternative that has been used successfully is a (recycled) dentist’s drill and some imagination. Typically, dimpling can be carried out to produce regions  $\sim 10\ \mu\text{m}$  thick, although in principle, precision dimpling with microprocessor control can sometimes produce electron transparent specimens which are  $<1\ \mu\text{m}$  thick.

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 thickness is  $>2\times$  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  $\text{HNO}_3$  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  $\text{HNO}_3$  oxidizes the Si and the HF removes the  $\text{SiO}_2$ . 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  $\mu\text{m}$ . 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.

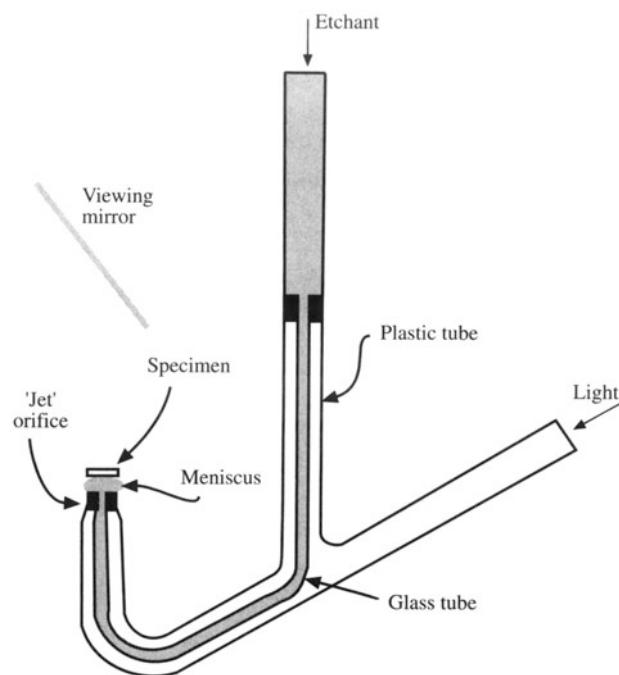
For some materials, such as Si, you can use this polisher to thin the specimen to electron transparency.



**Figure 10.4.** A dimpling apparatus showing grinding tool, and specimen support block.

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.

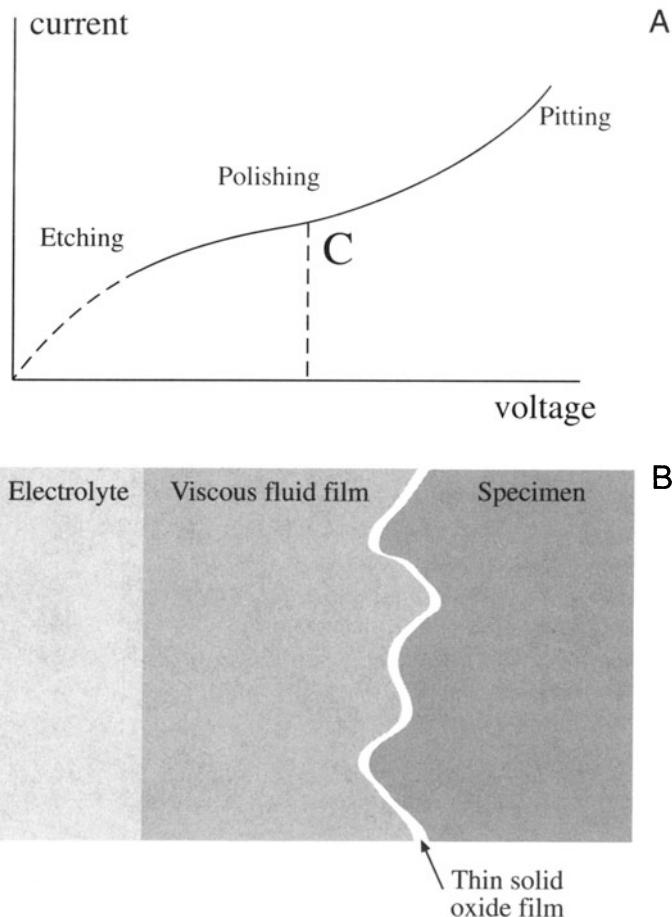
With practice, you can dramatically reduce the time required for the final thinning step. This tool has had a major impact on making TEM a quality control instrument, particularly in the semiconductor industry.

## 10.4. FINAL THINNING OF THE DISK

### 10.4.A. Electropolishing

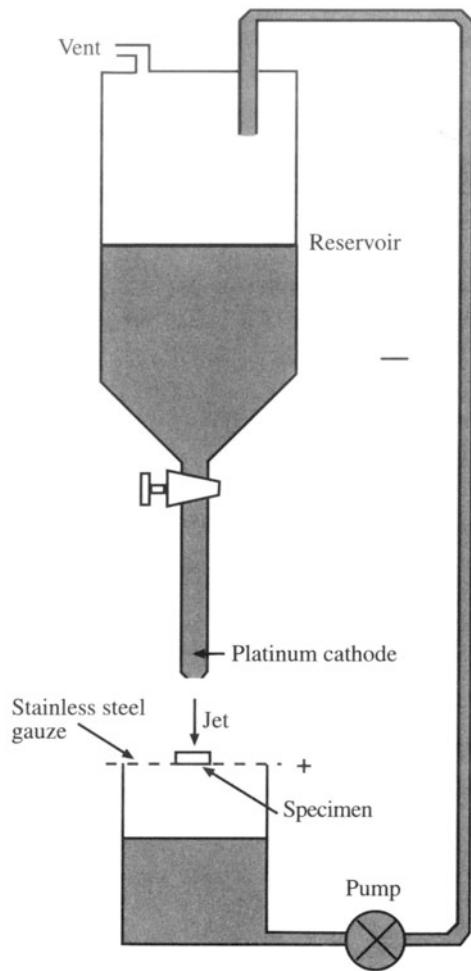
Electropolishing can only be used for electrically conducting samples such as metals and alloys. The method can be relatively quick (a few minutes to an hour or so) and it can produce foils with no mechanical damage. But it can change the surface chemistry of the specimen and it can be hazardous to your 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

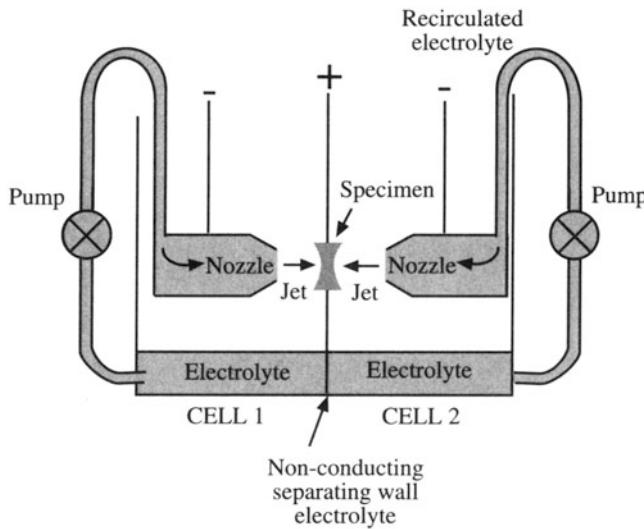


**Figure 10.6.** (A) Electropolishing curve showing the increase in current between the anode and the cathode as the applied voltage is increased. Polishing occurs on the plateau, etching at low voltages, and pitting at high voltages. (B) The ideal conditions for obtaining a polished surface require the formation of a viscous film between the electrolyte and the specimen surface.

A



B



**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.

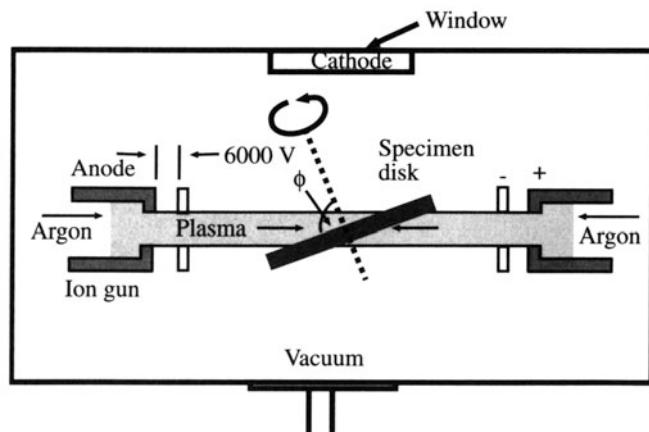
reservoir constant, the jet falls under constant pressure. The voltage is applied between the tip of the pipette and the specimen. A twin-jet apparatus can be used to pump a jet of electrolyte onto both sides of the dimpled disk, as shown schematically in Figure 10.7B. A laser beam or light sensor detects 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<sub>2</sub>)], 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<sub>2</sub> 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  $\leq 5^\circ$  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 ( $< 5^\circ$ ), 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)

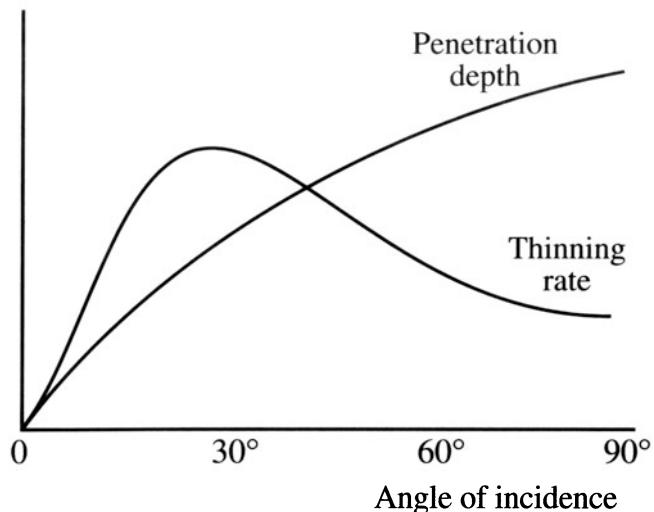
One thing you must remember is that ion thinning is closely related to ion-beam deposition. One manufacturer uses a similar arrangement to coat samples for SEM. The result is that material removed from one part of the sample can easily be redeposited elsewhere on the sample.

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.

$\text{Ar}$  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.

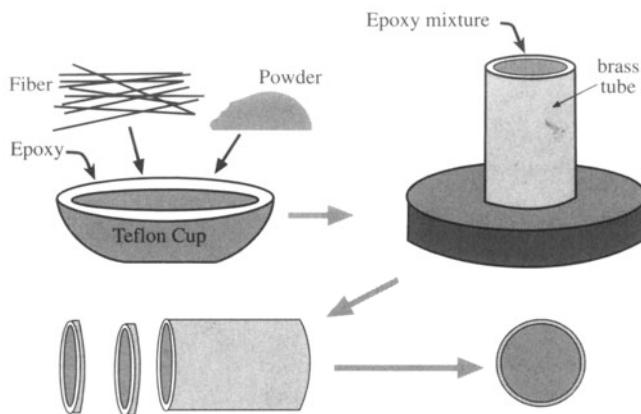


**Figure 10.9.** Variation in ion penetration depth and thinning rate with the angle of incidence. High incidence angles promote implantation, which is undesirable. The rate of thinning reaches a maximum at  $\sim 20^\circ$  incidence, after which the beam penetrates rather than sputters the sample surface. Initial thinning should start at 20–30°, reducing to  $< 5^\circ$  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  $\text{Ar}$  bubbles form in silicon at a depth of  $\sim 10$  nm after 5-keV thinning. Chemical analysis (EDS) of some  $\beta$ -aluminas which had the correct structure by HRTEM (composition  $\text{K}_2\text{O}\cdot 11\text{Al}_2\text{O}_5$ ) 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  $>200^{\circ}\text{C}$  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

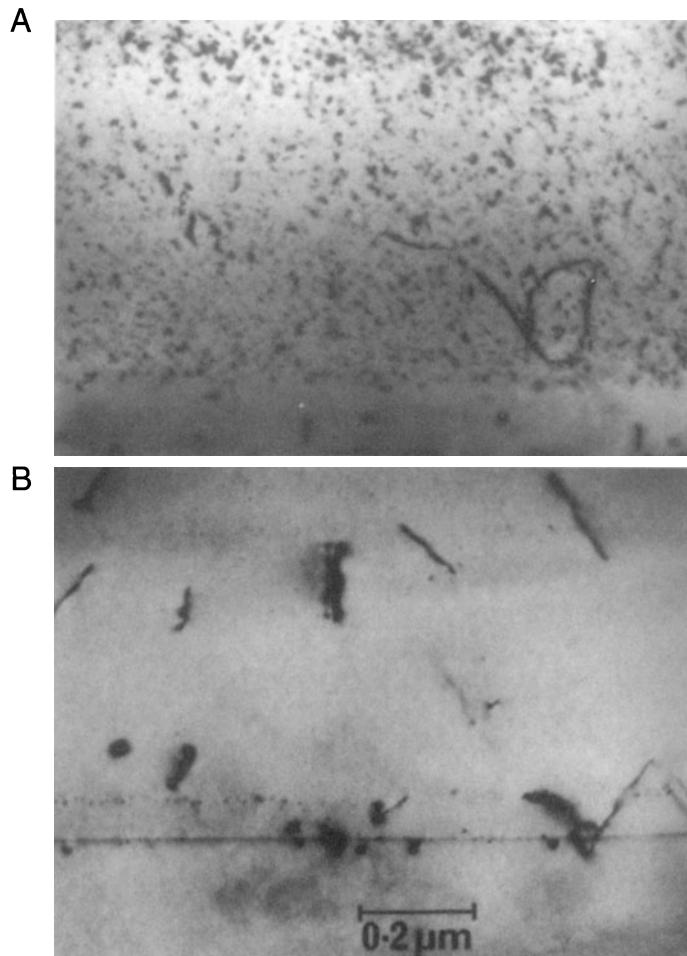
This depends on your ion miller, but if you're choosing a new machine there may be an advantage in tilting the gun rather than the specimen. If the specimen is inclined, then you need a clamping ring and you may sputter this when you thin the specimen. This has led to the development of ion polishing 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  $4\text{--}5^{\circ}$ .

### 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  $10^{-2}\text{--}10^{-3}$  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 ( $\sim 1 \mu\text{m} \times 1 \mu\text{m}$ ) 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 ( $4^{\circ}$ ) 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.

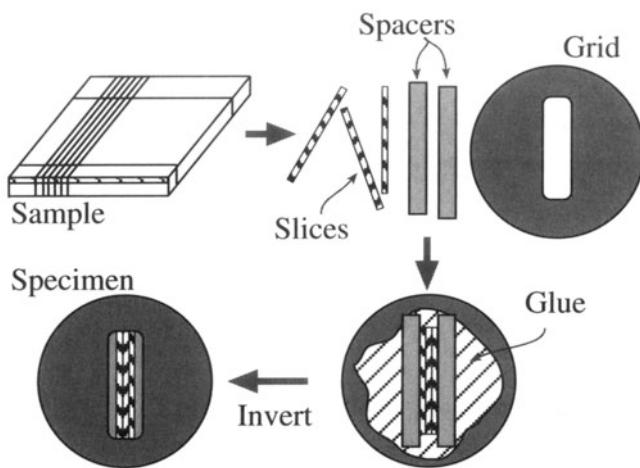
- Ion milling will form a layer on the surface which will probably be a combination of amorphous, highly damaged, and implanted material! The chemistry of the layer will be different than the rest of the specimen. The thickness of crystalline material will thus be less than the total thickness.

## 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.

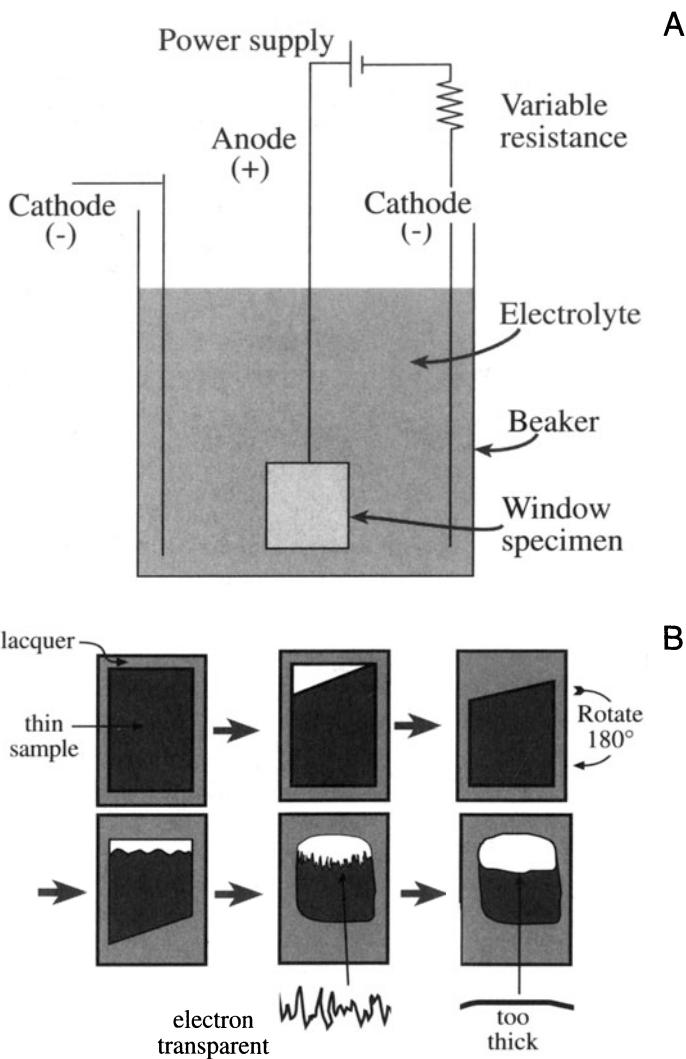
The thin supporting film should have a uniform thickness; the idea is that you are not actually interested in this material and therefore want to minimize its effect on the image of the material you are interested in.

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  $\sim 1\text{ cm} \times 1\text{ cm}$  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  $180^\circ$  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  $\sim 1\text{ cm}^2$  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  $180^\circ$  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.

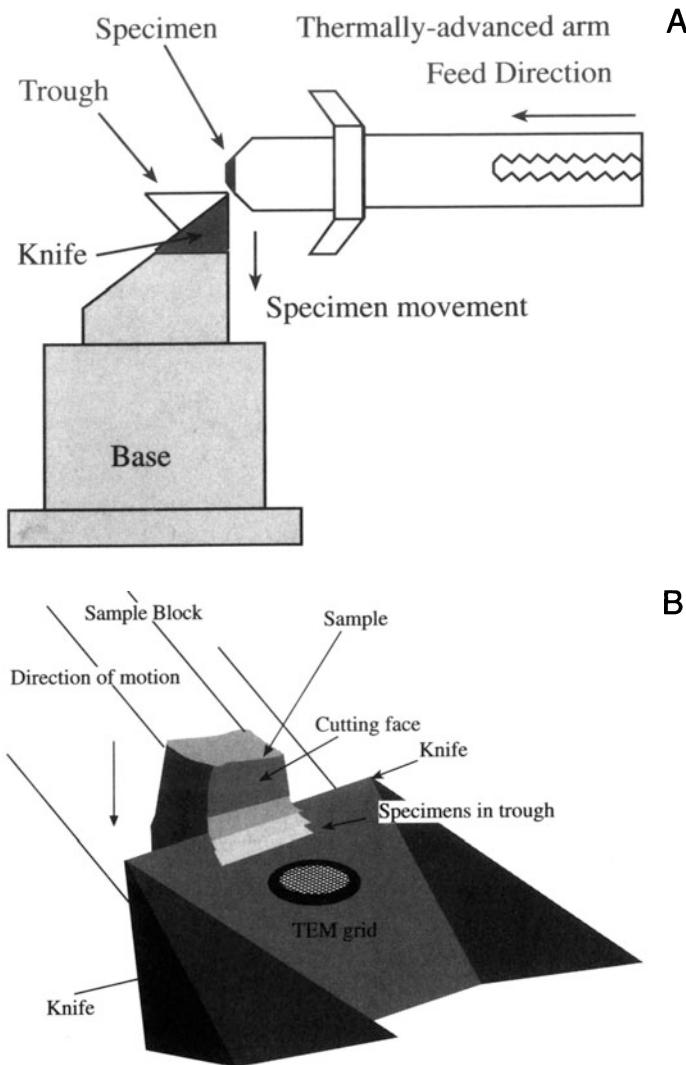
the sheet will finally thin in the center. If final thinning occurs too near the top of the sheet, the edge of the perforation is smooth and relatively thick. After perforation, remove the sheet and cut off slivers of material from around the perforation using a scalpel under an inert solvent such as 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



**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.

epoxy into the pores. After curing, you can ultramicrotome the sample in the usual way.

### 10.6.C. Grinding and Crushing

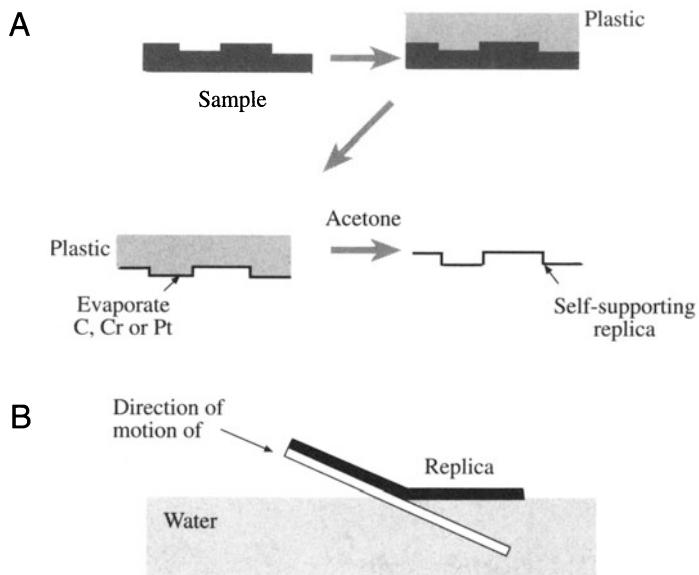
Many brittle materials such as ceramics and minerals are most easily prepared by crushing in a clean pestle and mortar (preferably in an inert liquid). The liquid containing the particles can then be ultrasonically stirred and allowed to settle. The thin particles are too small to be seen and the supernatant 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.

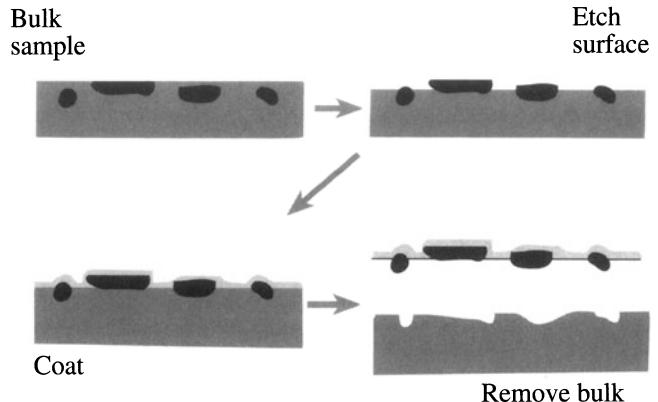
Dust particles found in airborne pollution can be collected simply by leaving the grid and a support film out for a period of time.

### 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 repli-



**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.

cate 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.

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

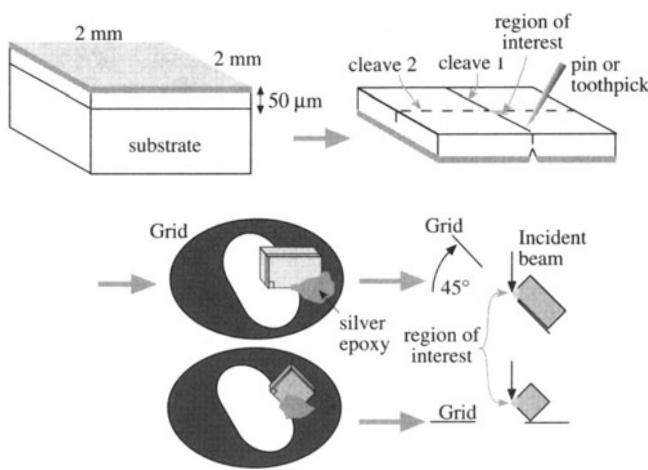
This is one of the oldest techniques and has been used to make thin specimens of graphite, mica, and other layer ma-

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 ( $\text{MoS}_2$ ) 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 (1̄10) 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.



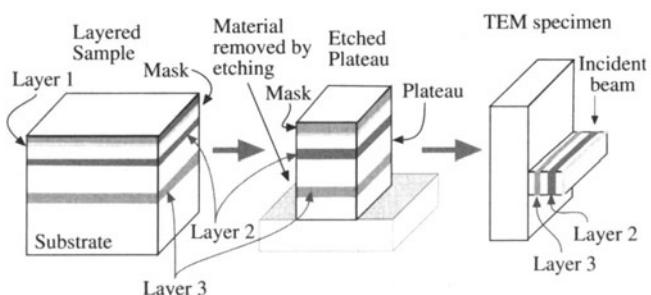
**Figure 10.17.** The 90°-wedge specimen: prethin to create a 2-mm square of the multilayers on a Si substrate; scribe the Si through the surface layers, turn over, and cleave; inspect to make sure the cleavage is clean, giving a sharp 90° edge; reject if not; mount the 90° corner over the edge of a 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

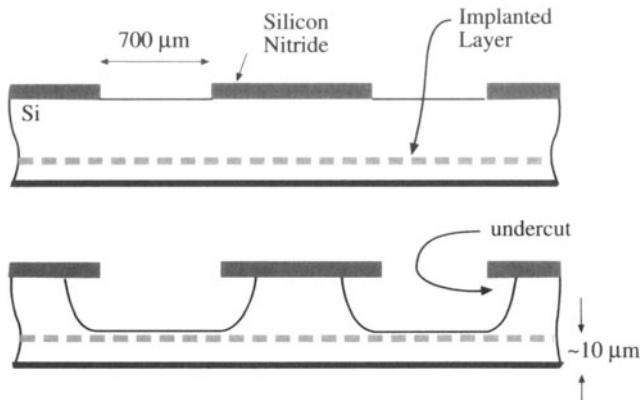
Using a technique developed for advanced engineering applications, lithography is used in the microelectronics industry to define fine lines of width down to 100 nm. An illustration of how lithography can be used specifically to prepare TEM specimens (as opposed to generating a structure which might 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  $\text{Al}_{1-x}\text{Ga}_x\text{As}$  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 multi-layer specimen: in the top diagram, the unthinned sample is shown with a grid of  $\text{Si}_3\text{N}_4$  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  $\sim 10 \mu\text{m}$  thick. Further thinning with a different solution produces large areas of uniformly thin material (not shown) supported by the  $\text{Si}_3\text{N}_4$  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.

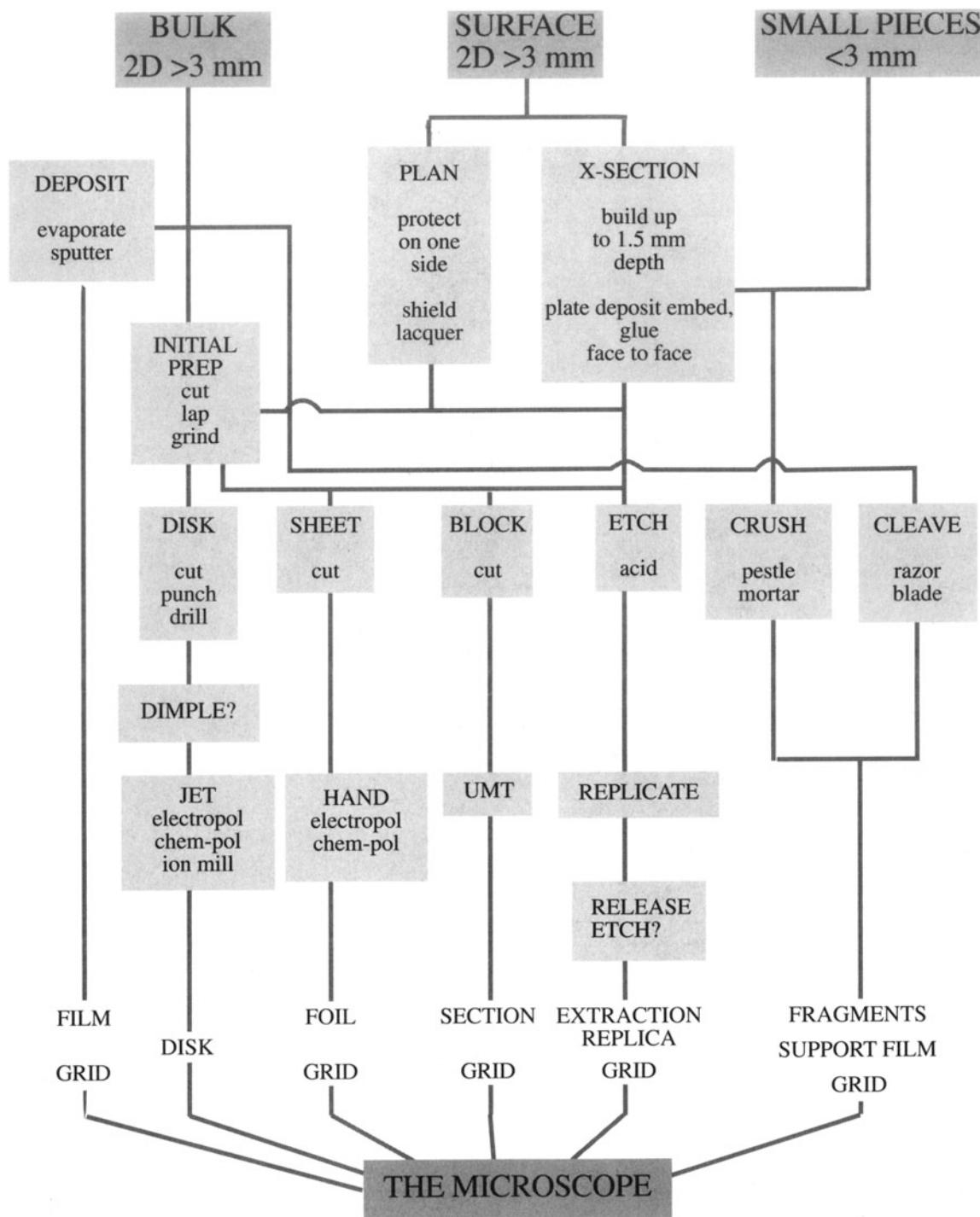


Figure 10.20. Summary flow chart for specimen preparation.

**Table 10.1. Artifacts Produced during Specimen Preparation<sup>a</sup>**

Artifact/Problem	Consequence
<i>Variable thickness</i>	
	<ul style="list-style-type: none"> <li>■ limited local area for chemical mapping (EP, IT, C, CD)</li> <li>■ very limited area for EELS</li> <li>■ somewhat limited area for absorption-free XEDS</li> <li>■ omission of low density defects</li> <li>■ distorted defect densities (EP, IT, TP)</li> </ul>
<i>Uniform thickness</i>	
	<ul style="list-style-type: none"> <li>■ limited diffraction information (UM)</li> <li>■ limited microstructure information (UM)</li> <li>■ handling difficulties (UM)</li> </ul>
<i>Surface films</i>	
	<ul style="list-style-type: none"> <li>■ bath residue, spec. dissolution and/or redeposition (EP)</li> <li>■ enhanced surface oxide (EP)</li> <li>■ extremely irregular topographies (IT)</li> <li>■ faster contamination buildup under beam (EP, R)</li> <li>■ retention of matrix on extracted particle</li> <li>■ C-redeposition (UM—embedded, UM, C, R—support films)</li> <li>■ Cu<sub>2</sub>O formation from Cu grids upon heating (R, UM, C)</li> <li>■ ion amorphization, diffusion-pump oil, redeposition (IT)</li> </ul>
<i>Differential thinning</i>	
	<ul style="list-style-type: none"> <li>■ different phases thin at different rates (EP, IT)</li> <li>■ different orientations thin at different rates (IT)</li> <li>■ grain/phase boundary grooving (EP, IT)</li> <li>■ anodic attack of matrix/particle (UM)</li> </ul>
<i>"Selectivity"</i>	
	<ul style="list-style-type: none"> <li>■ perforation influenced by local defect structure (EP, IT)</li> <li>■ very limited or no microstructure information (C, R)</li> <li>■ weak local regions debond and fall out (all)</li> </ul>
<i>"False" defects</i>	
	<ul style="list-style-type: none"> <li>■ microstructure obscured by high defect density (UM, CD)</li> <li>■ deformation-induced defects (EP, TP)</li> <li>■ ion-induced loops, voids (IT)</li> <li>■ heat-altered defects (EP, IT)</li> </ul>

<sup>a</sup>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. II: 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.

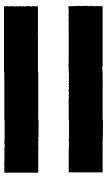
- Kestel, B.J. (1986) "Polishing Methods for Metallic and Ceramic Transmission Electron Microscopy Studies," in ANL-80-120 Rev. 1 (from Argonne National Laboratory). 66 pages of recipes and directions which have all been extensively used by Kestel.
- Marcus, R.B. and Sheng, T.T. (1983) *Transmission 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 Gorringe, 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 Sarlarius, R. (1949) *Rev. Mat.* **46**, 549.

# Diffraction



---

# Diffraction Patterns

# 11

---

11.1. Why Use Diffraction in the TEM? .....	179
11.2. The TEM, Diffraction Cameras, and the TV .....	179
11.3. Scattering from a Plane of Atoms .....	181
11.4. Scattering from a Crystal .....	182
11.5. Meaning of $n$ in Bragg's Law .....	184
11.6. A Pictorial Introduction to Dynamical Effects .....	185
11.7. Use of Indices in Diffraction Patterns .....	185
11.8. Practical Aspects of Diffraction-Pattern Formation .....	185
11.9. More on Selected-Area Diffraction Patterns .....	186

## 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?

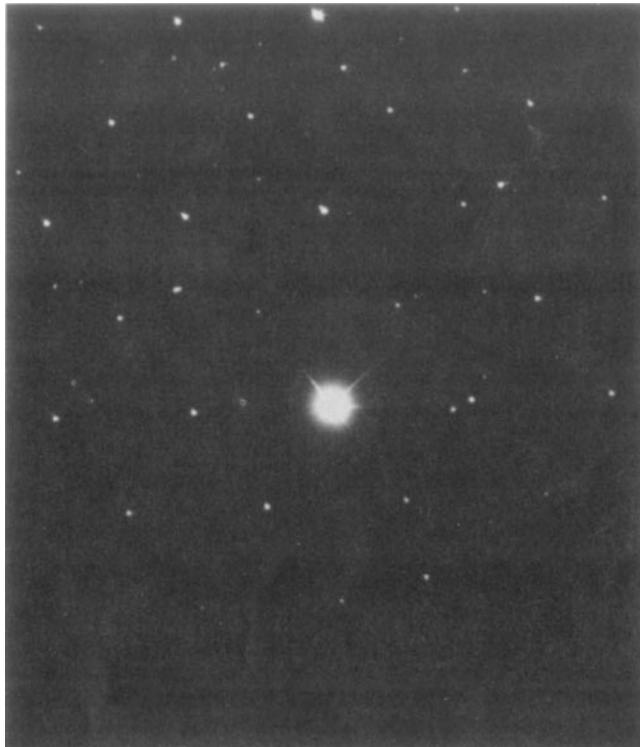
In general, if we see spots then the specimen is at least partly crystalline. (We'll discuss quasicrystals later.) The ability to determine crystallographic orientations locally (down to the nm level) gives TEM its great advantage over SEM and visible-light microscopes. Later on we can make this determination even more precise (to an accuracy of  $\sim 0.001^\circ$ ) using convergent-beam patterns, as we'll see in Chapter 21.

In this chapter we will restrict the discussion to the geometry of the spot patterns. These are necessarily associated with crystalline materials. We'll see that spot patterns provide a great deal of information themselves; they also provide the basis for understanding other DPs. We will find that standard DPs which are common to a group of materials allow us quickly to recognize 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 ( $\geq 100$  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.

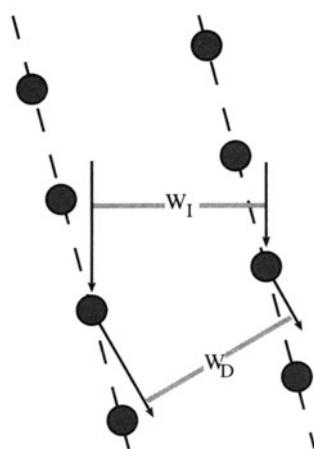
### 11.3. SCATTERING FROM A PLANE OF ATOMS

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.

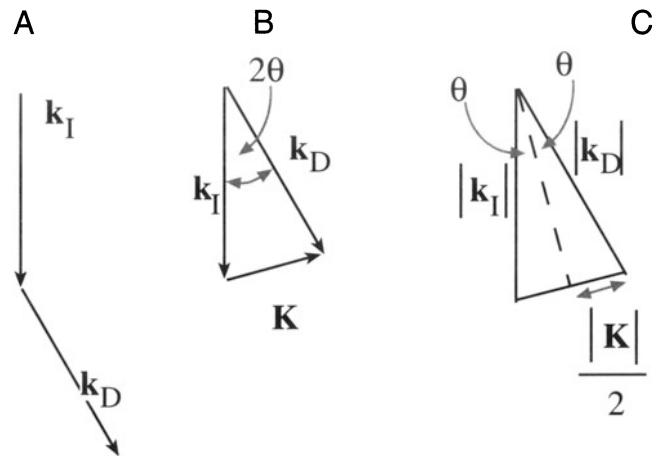
The simple diagram in Figure 11.2 shows an *initial* wavefront,  $W_I$ , being scattered by two planes of atoms to produce a *diffracted* wavefront,  $W_D$ . Whether or not  $W_D$  corresponds to a diffracted beam will depend on whether the atoms are scattering in phase, which itself is determined by 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  $\mathbf{k}$  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  $\mathbf{k}$  is normal to this wavefront. The diagram in Figure 11.3a,b defines vectors  $\mathbf{k}_I$ ,  $\mathbf{k}_D$ , and  $\mathbf{K}$  and gives us the following important equation (which is just vector addition)



**Figure 11.2.** Scattering from two planes of atoms.  $W_I$  and  $W_D$  are the incident and diffracted wavefronts, respectively.



**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)  $\mathbf{K}$  is the difference vector ( $= \mathbf{k}_D - \mathbf{k}_I$ ); (c)  $\sin \theta$  is defined as  $K/2k_I$ .

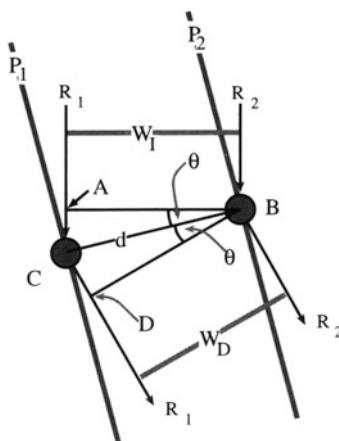
$$\mathbf{K} = \mathbf{k}_D - \mathbf{k}_I \quad [11.1]$$

where  $\mathbf{k}_I$  and  $\mathbf{k}_D$  are the  $\mathbf{k}$  vectors of the incident and diffracted waves, respectively. The vector  $\mathbf{K}$  is thus the change in  $\mathbf{k}$  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  $\mathbf{K}$ ; the angle  $\theta$  shown here need not be a Bragg angle.

Following our discussion in Section 3.10.B, we can always write that

$$|\mathbf{k}_I| = |\mathbf{k}_D| = \frac{1}{\lambda} = |\mathbf{k}| \quad [11.2]$$

providing the energy of the electron is unchanged during diffraction, i.e., the scattering process is elastic. From Fig-



**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  $\theta$  using simple trigonometry

$$\sin \theta = \frac{|\mathbf{K}|/2}{|\mathbf{k}_i|} \quad [11.3]$$

or

$$|\mathbf{K}| = \frac{2 \sin \theta}{\lambda} \quad [11.4]$$

$|\mathbf{K}|$ , like  $|\mathbf{k}_i|$ , has units  $\text{\AA}^{-1}$  if  $\lambda$  is measured in  $\text{\AA}$ .  $\mathbf{K}$  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  $\mathbf{k}$ -vectors are all appropriate to the electrons inside the crystal (rather than in the vacuum).

Equation 11.4 is very important; whenever you see the term  $(\sin \theta)/\lambda$  remember that it is just  $\mathbf{K}/2$  and is thus related to a change in wave vector.

If we now extend this argument to consider the interference between waves scattered from two points (which you can visualize as being atom sites) then we have the situation sketched in Figure 11.4. This figure should remind you of the idea of constructive and destructive interference, which we discussed back in Section 3.10. You will recognize that the geometry of Figure 11.4 is essentially a cross section of the two slits used by Young to demonstrate the wave nature of light (see also Section 2.10). We can then define two planes,  $P_1$  and  $P_2$ , to be normal to the vector  $\mathbf{CB}$ , 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

$$AC + CD = 2d \sin \theta \quad [11.5]$$

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  $\theta$  equals the Bragg angle,  $\theta_B$ , equation 11.4 becomes

$$|\mathbf{K}| = \frac{2 \sin \theta_B}{\lambda} \quad [11.6]$$

When  $\theta$  is  $\theta_B$ , the path difference in equation 11.5 is  $n\lambda$ , where  $n$  is any integer, and the equation becomes

$$n\lambda = 2d \sin \theta_B \quad [11.7]$$

which is Bragg's Law (equation 3.22). If  $n$  is 1

$$2 \sin \theta_B = \frac{\lambda}{d} \quad [11.8]$$

but we already know from equation 11.6 that, at the Bragg angle,

$$2 \sin \theta_B = \lambda |\mathbf{K}| \quad [11.9]$$

so when we are at the Bragg angle, the magnitude of the vector  $\mathbf{K}$  has a special value,  $K_B$ ,

$$|\mathbf{K}_B| = \frac{1}{d} \quad [11.10]$$

and we define this vector,  $\mathbf{K}_B$ , to be  $\mathbf{g}$  so that

$$\mathbf{K}_B = \mathbf{g} \quad [11.11]$$

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  $\mathbf{g}$  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;  $\lambda$  is 3.35 pm (0.00335 nm or 0.0335  $\text{\AA}$ ) for 120-kV electrons;

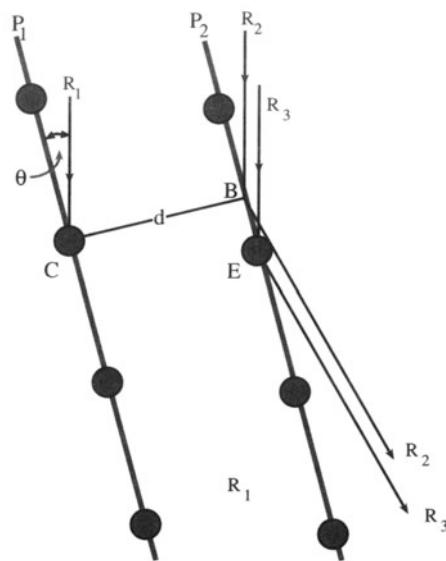
equation 11.8 then gives  $\theta = 7.97$  mrad ( $0.46^\circ$ ) for  $n = 1$ . As a rule of thumb, the Bragg angles of interest are usually no more than  $1^\circ$  when we are forming images, although important information may be present in DPs at much larger ( $10^\circ$  to  $20^\circ$ ) angles; you will find it useful to remember the order of magnitude of these numbers.

Remember that 10 mrad is  $0.573^\circ$ , i.e., about  $0.5^\circ$ .

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.

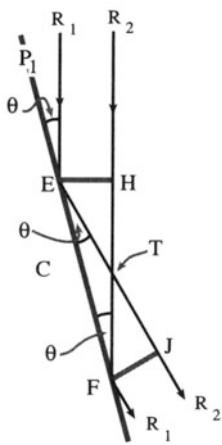
It does not matter how the atoms (scattering centers) are distributed on these two planes; the scattering from any two points on planes  $P_1$  and  $P_2$  will produce the same path difference  $2d \sin \theta$ .



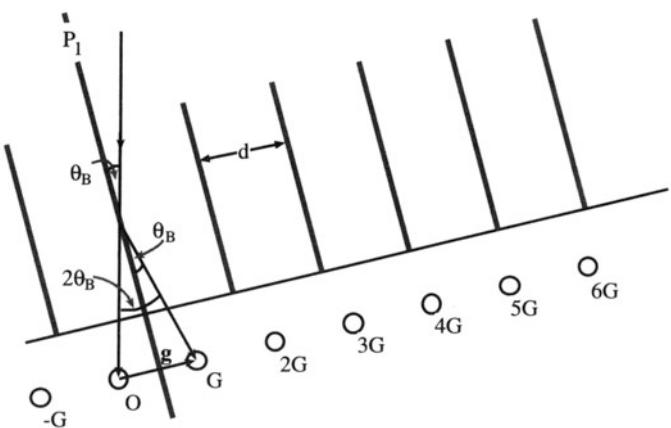
**Figure 11.6.** Scattering from three points on two planes. The path difference for scattering from points B and C is  $2d \sin \theta$ , so the path difference for scattering from points C and E is also  $2d \sin \theta$ . Hence scattering in the direction of the diffracted beam from all points shown 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 11.5.** 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 11.7.** Diffraction from a set of planes a distance  $d$  apart. The planes have been oriented to be in the Bragg diffracting condition ( $\theta_B$  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  $\mathbf{g}$  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,  $\mathbf{g}$ , 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  $\mathbf{K}$  has the value  $\mathbf{g}$ .

## 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,  $\bar{\mathbf{g}}$ ,  $\mathbf{0}$ ,  $\mathbf{g}$ ,  $2\mathbf{g}$ ,  $3\mathbf{g}$ , etc.

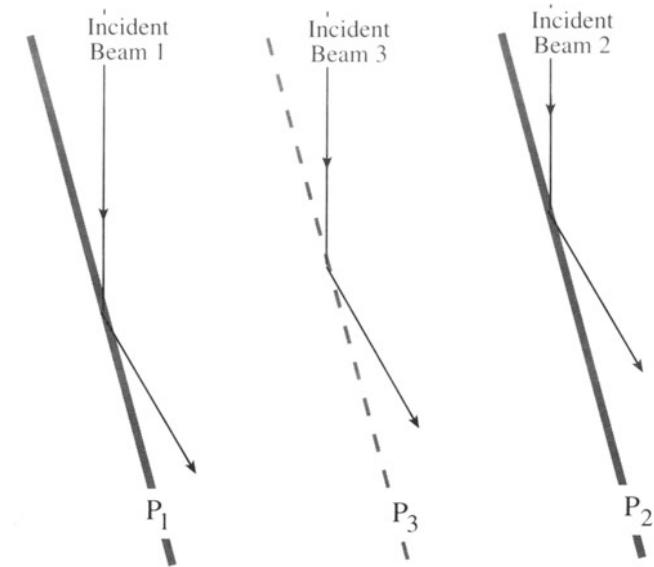
**Notation:** When discussing beams in diffraction patterns, the letter O will refer to the “direct” beam which is present even when there is no specimen, the letter G (not bold—it’s not a vector) will refer to any single diffracted beam; the number  $\mathbf{0}$  (bold) will refer to the diffraction vector for beam O (it is a vector of zero length), and the letter  $\mathbf{g}$  (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  $\mathbf{g}$  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 ( $n\mathbf{g}$ , where  $n \neq 1$ ), called higher-order reflections, are particularly important in TEM. Pictorially, you can imagine them as arising from the interference 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

$$2\left(\frac{d}{2}\right) \sin \theta = \lambda \quad [11.12]$$



**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

$$|\mathbf{g}_2| = \frac{2}{d} \quad [11.13]$$

i.e., when

$$|\mathbf{g}_2| = 2 |\mathbf{g}| \quad [11.14]$$

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

$$2\left(\frac{d}{n}\right) \sin \theta = \lambda \quad [11.15]$$

or rewrite this as

$$2d \sin \theta = n\lambda \quad [11.16]$$

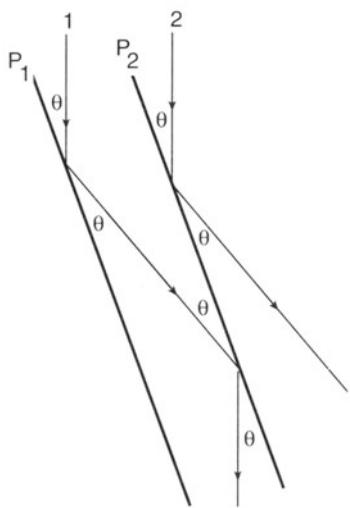
which gives a physical explanation for the  $n$  in equation 11.7.

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  $(hkl)$  and a set of such planes is  $\{hkl\}$ . We define the direct beam as the  $000$  reflection and each diffracted beam as a reflection with different  $hkl$  indices. It is a crystallographic convention to refer to the diffraction spot from a specific  $(hkl)$  plane as  $hkl$ , i.e., without the parentheses. If we assign  $hkl$  to  $\mathbf{g}$ , then the second-order  $(2g)$  spot is  $2h 2k 2l$ , the  $3g$  spot is  $3h 3k 3l$ , etc. Similarly, the  $\bar{g}$  reflection is  $\bar{h}\bar{k}\bar{l}$ . We'll discuss these points further in Section 12.3.

Now we can explain why we see so many spots in the DP. If we look along a zone axis in a crystal, we will see sets of planes in the edge-on orientation. Remember that a zone axis is the direction along the intersection of two or more planes.

*Notation:* The zone axis,  $[UVW]$ , is a direction which is common to all the planes of the zone. So  $[UVW]$  is perpendicular to the normal to the plane  $(hkl)$  if the plane is in the  $[UVW]$  zone. Later, we will see that  $[UVW]$  is defined as the incident beam direction. This result applies to all crystal systems and gives the Weiss zone law:  $hU + kV + \ell W = 0$ .

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.

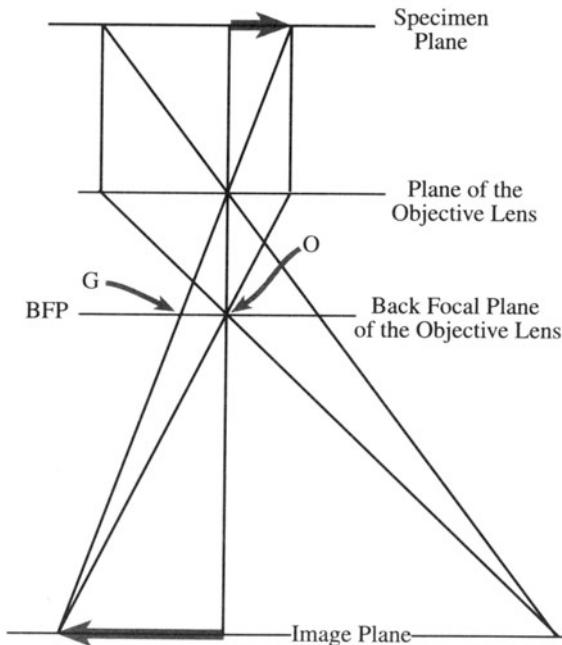
SAD patterns are sharply focused spot patterns, which we use to select reflections for all imaging modes. We can easily associate the sharp spots with our diffraction vectors,  $\mathbf{g}$ .

CBED patterns are arrays of disks. We can associate a  $\mathbf{g}$  vector with each disk but the location of  $\mathbf{g}$  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:

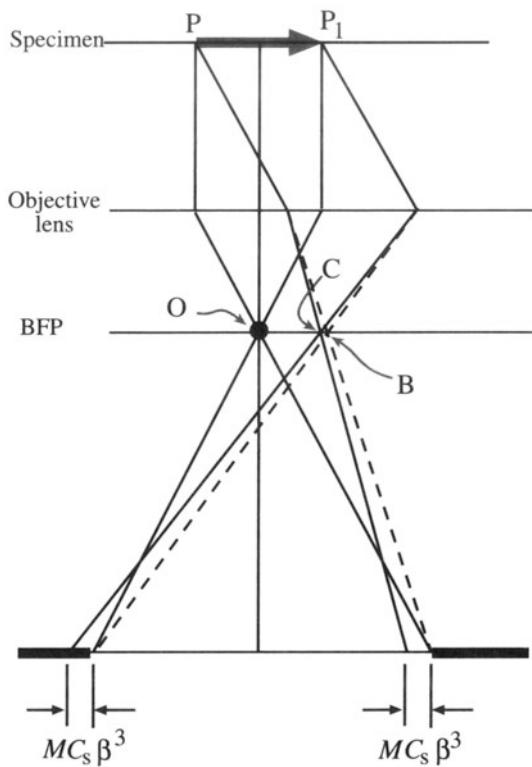
- Be sure that you are at the eucentric focus position, with an image of the area of interest focused on the screen.
- Insert the SAD aperture.
- Remove the objective aperture.
- Focus the SAD aperture.
- Switch to diffraction mode.
- Spread the beam using C2, within the limits imposed by your specimen.
- 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  $25\times$ . Thus a 50- $\mu\text{m}$  aperture will select a 2- $\mu\text{m}$  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  $\beta$  to the optic axis, the image formed at magnification  $M$ , is translated a distance  $r_M$  given by

$$r_M = M C_s \beta^3 \quad [11.17]$$

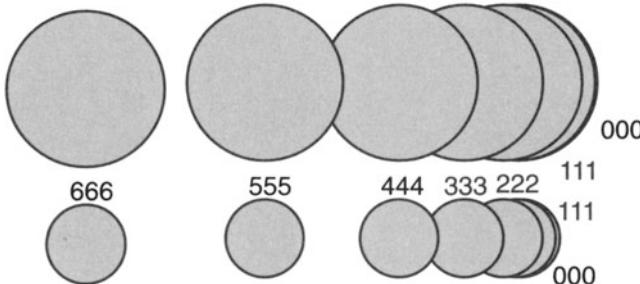
So the area we select using the SAD aperture corresponds to the area  $PP_1$  in the object plane *only* for the direct beam. The error increases as  $\beta$  increases, so that it's larger for a larger Bragg angle or for a larger  $\mathbf{g}$ . The result is illustrated schematically in Figure 11.12 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- $\mu\text{m}$  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 fo-



**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 ( $a_0 = 4.04 \text{ \AA}$ ) assuming 100-keV electrons and  $C_s = 3 \text{ mm}$ . The 000 and 111 disks almost exactly overlap (the translation is 13 nm). The diameter of each disk in the top row is 1  $\mu\text{m}$ , and the diameter of each disk in the bottom row is 0.5  $\mu\text{m}$ .

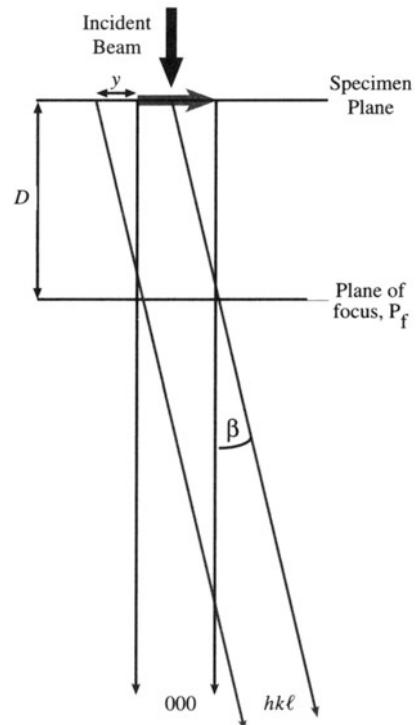
**Table 11.1. The Displacement Distance of the Image Formed by "reflection G" Due to Spherical Aberration**

Reflection in SAD pattern	$C_s \beta^3 (\text{nm})$ old TEM	$C_s \beta^3 (\text{nm})$ modern IREM
111	13	1.2
222	100	9.1
333	350	31.9
444	760	69.3
555	1620	150
666	2800	250

cused 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

$$y = D \beta \quad [11.18]$$

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  $\sim 3 \mu\text{m}$  (see Figure 6.14). You will still find on many TEMs that the aperture in the



**Figure 11.13.** If the lens is not focused on the SAD plane, images associated with the different  $\mathbf{g}$  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  $\lambda$  (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  $\mu\text{m}$  is about the smallest that can be manufactured.

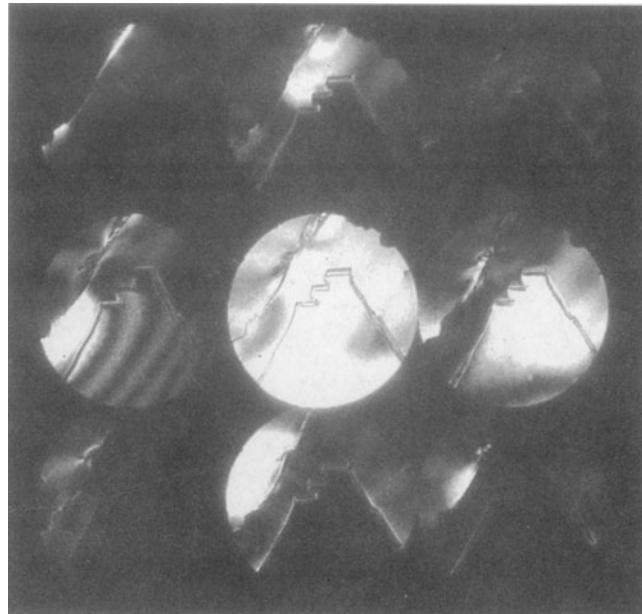
One question which is often asked is: if the SAD aperture is placed at the first image plane, how can it affect the DP which is formed above it? The relationship between the SAD pattern and the image(s) can be illustrated by forming a multiple dark-field image of the type illustrated in Figure 11.14A. 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.

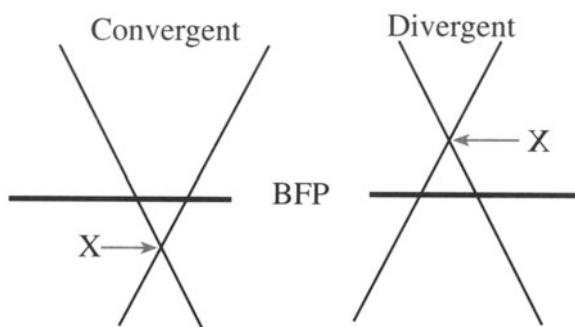
This uncertainty in the area of selection of SAD patterns is one reason that CBED patterns can have some advantage whenever you want to get crystallographic information about specific regions of your specimen.

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.



A



B

**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  $\mathbf{g}$  associated with each Bragg reflection and associate each  $\mathbf{g}$  with a crystal plane  $hkl$ . The diffracting planes are all in a specific zone axis  $UVW$ , which we can define as parallel to the incident beam direction.

## 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

---

12.1. Why Introduce Another Lattice? .....	193
12.2. Mathematical Definition of the Reciprocal Lattice .....	193
12.3. The Vector $\mathbf{g}$ .....	194
12.4. The Laue Equations and Their Relation to Bragg's Law .....	195
12.5. The Ewald Sphere of Reflection .....	196
12.6. The Excitation Error .....	197
12.7. Thin-Foil Effect and the Effect of Accelerating Voltage .....	199

## CHAPTER PREVIEW

In the previous chapter, you've already encountered vectors  $\mathbf{k}$  and  $\mathbf{g}$  and seen that they have lengths with units  $\text{\AA}^{-1}$  or  $\text{nm}^{-1}$ . 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  $\mathbf{K}$  is equal to  $\mathbf{g}$ . The reciprocal lattice concept allows us to define a lattice where all the lattice points correspond to the possible  $\mathbf{g}$  vectors.

In the reciprocal lattice, sets of parallel  $(hkl)$  atomic planes are represented by a single point located a distance  $1/d_{hkl}$  from the lattice origin.

To understand why we use the reciprocal lattice, remember that we can always write Bragg's Law (equations 11.2 and 11.3) as

$$\frac{2 \sin \theta_B}{\lambda} = \frac{n}{d} = |\mathbf{K}| \quad [12.1]$$

Thus the vector  $\mathbf{K}$  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,  $\mathbf{g}$ . 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

$$\mathbf{r}_n = n_1 \mathbf{a} + n_2 \mathbf{b} + n_3 \mathbf{c} \quad [12.2]$$

where the vectors  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$  are the unit-cell translations in real space while  $n_1$ ,  $n_2$ , and  $n_3$  are all integers.

Any reciprocal lattice vector,  $\mathbf{r}^*$ , can be defined in a similar manner

$$\mathbf{r}^* = m_1 \mathbf{a}^* + m_2 \mathbf{b}^* + m_3 \mathbf{c}^* \quad [12.3]$$

where  $\mathbf{a}^*$ ,  $\mathbf{b}^*$ , and  $\mathbf{c}^*$  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,  $\mathbf{a}^*$  is normal to both  $\mathbf{b}$  and  $\mathbf{c}$ , 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  $\mathbf{a}^*$  is parallel to  $\mathbf{a}$  (think about this!). The direction of  $\mathbf{a}^*$  is actually completely defined by equation 12.4. It is perpendicular to both  $\mathbf{b}$  and  $\mathbf{c}$  and must therefore be the normal to the plane containing  $\mathbf{b}$  and  $\mathbf{c}$ .

The vector  $\mathbf{a}^*$  is always perpendicular to the plane (100) even when  $\mathbf{a}$  is not.

Equation 12.5 then uniquely defines the length of the vector  $\mathbf{a}^*$  in terms of the length of the vector  $\mathbf{a}$ . Therefore, this equation gives the scale or dimension of the reciprocal lattice. The product of the projection of  $\mathbf{a}^*$  on the vector  $\mathbf{a}$  multiplied by the length of  $\mathbf{a}$  is unity. We can see that if  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$  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  $\mathbf{a}^*$  as

$$\mathbf{a}^* = \frac{\mathbf{b} \wedge \mathbf{c}}{V_c} \quad [12.6]$$

This definition emphasizes that the vector  $\mathbf{a}^*$  is orthogonal to the vectors  $\mathbf{b}$  and  $\mathbf{c}$ . However, just as  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{c}$  need not be normal to one another,  $\mathbf{a}^*$ ,  $\mathbf{b}^*$ , and  $\mathbf{c}^*$  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 $\mathbf{g}$

We can generalize our definition of  $\mathbf{g}$  a little more. Any vector in reciprocal space can be defined as a combination of the vectors  $\mathbf{a}^*$ ,  $\mathbf{b}^*$ , and  $\mathbf{c}^*$ . In particular, we can write  $\mathbf{K}$  in this form for use later

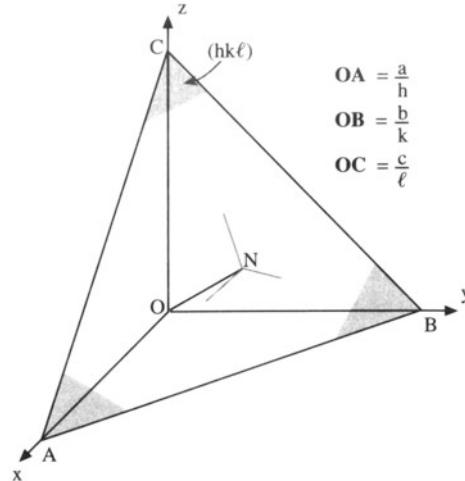
$$\mathbf{K} = \xi \mathbf{a}^* + \eta \mathbf{b}^* + \zeta \mathbf{c}^* \quad [12.7]$$

A particularly important reciprocal lattice vector is the vector  $\mathbf{g}_{hkl}$ , which is defined as

$$\mathbf{g}_{hkl} = h \mathbf{a}^* + k \mathbf{b}^* + \ell \mathbf{c}^* \quad [12.8]$$

where  $h$ ,  $k$ , and  $\ell$  are all integers and together define the plane  $(hkl)$ .

The definition of the plane  $(hkl)$  is that it cuts the  $a$ ,  $b$ , and  $c$  axes at  $1/h$ ,  $1/k$ , and  $1/\ell$ , respectively. If you look at Figure 12.1, you'll see that the vector  $\mathbf{AB}$  can be written



**Figure 12.1.** The plane ABC has Miller indices  $(hkl)$ . The vectors  $\mathbf{OA}$ ,  $\mathbf{OB}$ , and  $\mathbf{OC}$  have lengths  $a/h$ ,  $b/k$ , and  $c/\ell$ . The vector  $\mathbf{ON}$ , which may be written as  $\mathbf{n}$ , is normal to the plane  $(hkl)$ . In the text we see that the reflection,  $\mathbf{g}$ , which is associated with diffraction from the  $(hkl)$  planes, is parallel to  $\mathbf{n}$  and normal to all vectors in  $(hkl)$ .

as  $b/k - a/h$ . This vector, and all vectors in the  $(hkl)$  plane, are normal to the vector  $\mathbf{g}_{hkl}$  defined in equation 12.8. You can prove this by taking the dot product of  $\mathbf{AB}$  and  $\mathbf{g}$  and using equations 12.4 and 12.5. Therefore, the vector  $\mathbf{g}_{hkl}$  must be *normal* to the plane  $(hkl)$ .

$$\left( \frac{\mathbf{b}}{k} - \frac{\mathbf{a}}{h} \right) \cdot \left( h\mathbf{a}^* + k\mathbf{b}^* + \ell\mathbf{c}^* \right) = 0 \quad [12.9]$$

The vectors  $\mathbf{AB}$ ,  $\mathbf{BC}$ , and  $\mathbf{CA}$  all lie in the plane  $(hkl)$  and each is normal to  $\mathbf{g}_{hkl}$ . All that we now have to prove is that the length of the vector  $|\mathbf{g}_{hkl}|$  is given by  $(d_{hkl})^{-1}$ . To show this relationship, consider a unit vector,  $\mathbf{n}$ , normal to the plane (i.e., parallel to  $\mathbf{g}_{hkl}$ ) and take the dot product with any unit vector inclined to this plane (e.g.,  $\mathbf{a}/h$  or  $\mathbf{b}/k$ ).

The unit vector,  $\mathbf{n}$ , parallel to  $\mathbf{g}$  is simply  $\mathbf{g}/|\mathbf{g}|$ . Therefore, the shortest distance from the origin O to the plane is the dot product of  $\mathbf{n}$  with vector  $\mathbf{OB}$  (or  $\mathbf{OC}$ , etc.)

$$\mathbf{n} \cdot \frac{\mathbf{a}}{h} = \frac{\mathbf{g}}{|\mathbf{g}|} \cdot \frac{\mathbf{a}}{h} = \frac{(h\mathbf{a}^* + k\mathbf{b}^* + \ell\mathbf{c}^*)}{|\mathbf{g}|} \cdot \frac{\mathbf{a}}{h} = \frac{1}{h} = \frac{1}{|\mathbf{g}|} \quad [12.10]$$

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  $(hkl)$  planes, so that

$$d_{hkl} = \frac{1}{|\mathbf{g}|} \quad [12.11]$$

as we required.

- The definition of the  $hkl$  indices is  $OA = a/h$ ;  $OB = b/k$ ;  $OC = c/\ell$ .
- The plane ABC can then be represented as  $(hkl)$ .

We should emphasize a few points before moving on:

- Remember: the reciprocal lattice is so called because all lengths are in reciprocal units.
- If you are familiar with the derivation of band-gap concepts in elementary solid-state physics, you will have already used these ideas. The difference is that the energies of the electrons being produced in the microscope are  $\geq 100$  keV, whereas those in solids are  $\sim 1$  eV. This will affect the magnitudes of  $\mathbf{k}$  but the  $a^*$ , etc., will not change with kV.
- Reciprocal-space notation. We introduced the use of brackets in Section 11.7. Now we'll extend this notation to the reciprocal lattice:  $(hkl)$  is shorthand notation for a particular vector in reciprocal space because it is normal to the  $(hkl)$  plane in real space;  $\{hkl\}$  is then the general form for these reciprocal lattice vectors.  $[UVW]$  is a particular plane in reciprocal space, e.g., it may contain many  $\{hkl\}$  points so that in real space it would be a direction—the zone axis for the  $\{hkl\}$  real-space planes (see Table 12.1). 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]

**TABLE 12.1. Notation for Planes, Directions, and Reflections**

Real space	Reciprocal space	Indices
Particular direction	Particular plane	$[UVW]$
General direction	General plane	$\langle UVW \rangle$
Particular plane	Particular direction	$(hkl)$
General plane	General direction	$\{hkl\}$
Diffracting plane	Indexed reflection	$hkl$

zone axis in an orthorhombic crystal such as olivine, the beam will *not* be normal to the (123) plane.

## 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

$$\mathbf{K} = \mathbf{g} \quad [12.12]$$

From Figure 12.2 we can see that the magnitude of  $\mathbf{K}$  is always  $2 \sin \theta / \lambda$ . At the Bragg condition it is also equal to the magnitude of  $\mathbf{g}$ , i.e.,  $1/d$ . Therefore, at the Bragg condition we can write

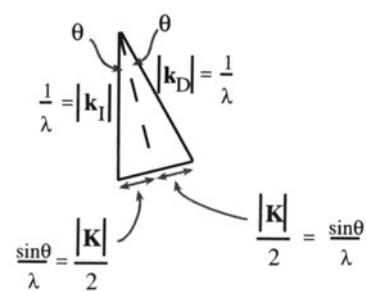
$$\frac{2 \sin \theta}{\lambda} = \frac{1}{d_{hkl}} \quad [12.13]$$

i.e.

$$\lambda = 2d \sin \theta \quad [12.14]$$

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  $\mathbf{k}_I$ ,  $\mathbf{k}_D$ ,  $\mathbf{K}$ ,  $\theta$ , and  $\lambda$ .

$$\mathbf{K} \cdot \mathbf{r}_n = N \quad [12.15]$$

This equation tells us that we must satisfy certain conditions on  $\mathbf{K}$  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

$$\mathbf{K} \cdot \mathbf{r}_n = N \text{ when } \xi, \eta, \text{ and } \zeta \text{ are the integers } h, k, \text{ and } \ell.$$

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

$$\mathbf{K} \cdot \mathbf{a} = h \quad [12.16]$$

$$\mathbf{K} \cdot \mathbf{b} = k \quad [12.17]$$

$$\mathbf{K} \cdot \mathbf{c} = \ell \quad [12.18]$$

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

$$n\lambda = 2d \sin \theta \quad [12.19]$$

We also discussed the physical reason for  $n$ . We can now treat the same situation mathematically. If the integers  $h, k$ , and  $\ell$  have a common factor then we can write

$$n d_{nh, nk, n\ell} = d_{hk\ell} \quad [12.20]$$

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.

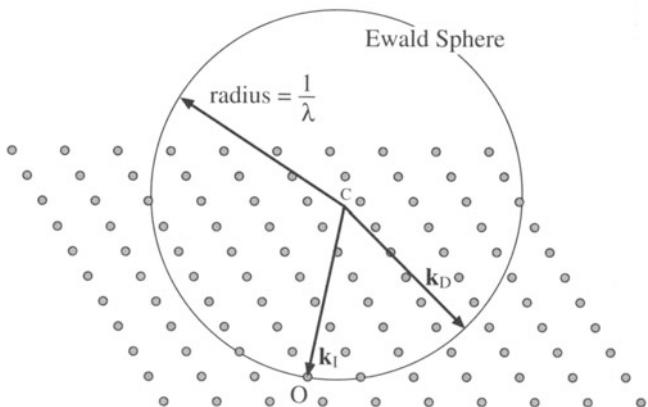
We now construct a sphere of radius  $1/\lambda$ . The sphere is known as the sphere of reflection or generally, and more simply, the "Ewald sphere," in honor of its inventor P.P. Ewald. Due to Ewald's German origins, Ewald is pronounced "A. Valt" rather than "E. Walled." Ewald's paper which first described 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  $\mathbf{CO}$  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  $\lambda$  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  $\mathbf{g}$  which can represent the quantity  $d^{-1}$ . The vector has a length and a direction. We choose the obvious length for  $\mathbf{g}$  to be  $d^{-1}$  and make  $\mathbf{g}$  the only unique vector for the plane  $(hkl)$ , 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,  $\mathbf{g}$ , 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  $\mathbf{k}_I$  and has length  $1/\lambda$ ; this defines where C is located, i.e., we start with O and measure back to C.

Now you can appreciate that it is only when the incident beam lies in our chosen plane that the vector CO will lie in that plane. For example, we may choose the plane to be parallel to the optic axis of the microscope but tilt the incident beam off this axis; in such cases we will still often be interested in the plane containing both the optic axis and the incident beam. Also notice that  $\mathbf{k}_D$  could be any vector which begins at C and ends on the sphere.

Consider the relative dimensions of  $d_{hkl}$  and  $\lambda$ . We can see that for X-rays where  $\lambda$  is  $\sim 0.2$  nm and  $1/\lambda$  is  $\sim 5$  nm $^{-1}$ , the Ewald sphere can only intersect a small number of relrods because  $1/d$  is only  $\sim 3$  nm $^{-1}$ . This explains why it is necessary in X-ray diffraction to use white radiation (giving a wide range of  $\lambda$ ) or to oscillate, rotate, or powder the specimen

(thus producing many variations of  $d$  and  $\theta$ ) in order to produce enough diffraction spots to analyze the structure. For 100-keV electrons, however,  $\lambda$  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.

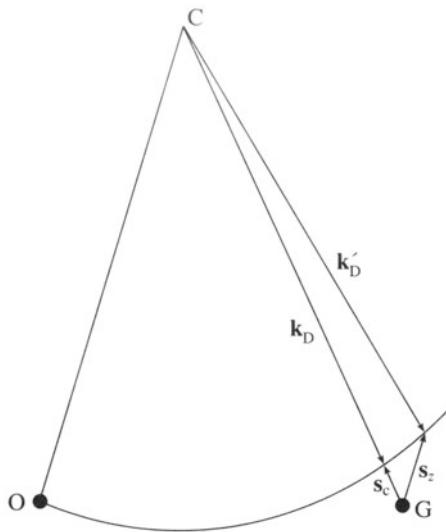
## 12.6. THE EXCITATION ERROR

We'll now introduce a new quantity,  $\mathbf{s}$ , 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,  $\mathbf{s}$ , in reciprocal space such that

$$\mathbf{K} = \mathbf{g} + \mathbf{s} \quad [12.21]$$

This vector,  $\mathbf{s}$ , 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  $\mathbf{g}$  is well defined,  $\mathbf{K}$  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  $\mathbf{s}$  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  $s = s_z$  and  $\mathbf{k}_D$  becomes  $\mathbf{k}'_D$ . In each case,  $\mathbf{k}_D$  ends on the Ewald sphere.

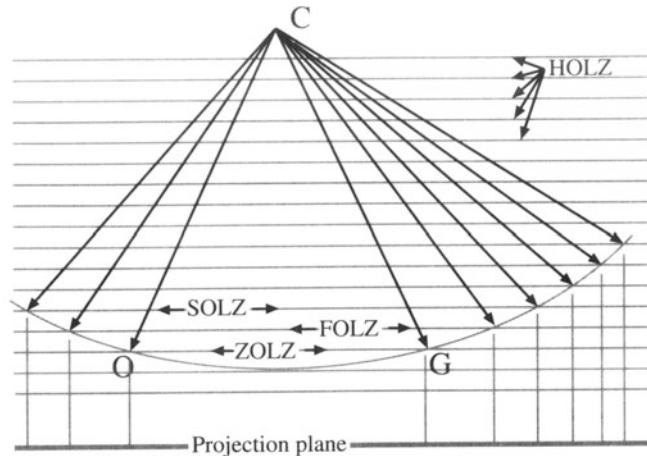
to define  $\mathbf{s}_m$  as being perpendicular to the surface of the specimen, but we don't know where that is. Actually, we will often assume that  $\mathbf{s}_m$  is perpendicular to  $OG$ , but this need not be the case. We will refer to  $s$  in several ways:  $s_g$  will emphasize that  $s$  is defined for a particular  $\mathbf{g}$  while  $s_z$  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,  $\mathbf{g}$ , from the origin to the point. In Figure 12.4, the row of reciprocal lattice points (only  $G$  is shown) is essentially at  $90^\circ$  to the incident beam. If we take all such rows, we define a plane of points which are all at  $90^\circ$  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, and 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.

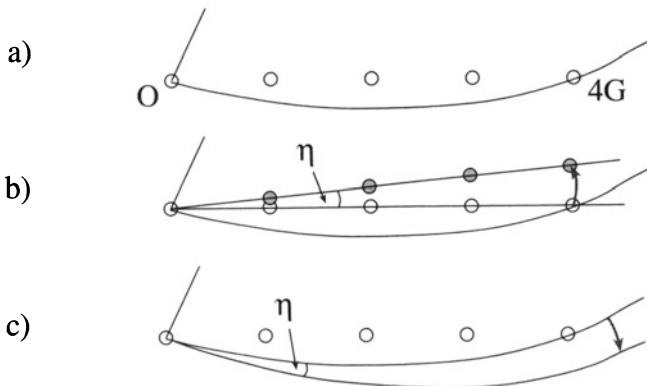


**Figure 12.5.** The Ewald sphere intercepts points in higher-order Laue zones (HOLZ) at large angles to the incident-beam direction. If the radius of the sphere increases (higher kV beam) then the sphere flattens and the HOLZ interception is at still larger angles.

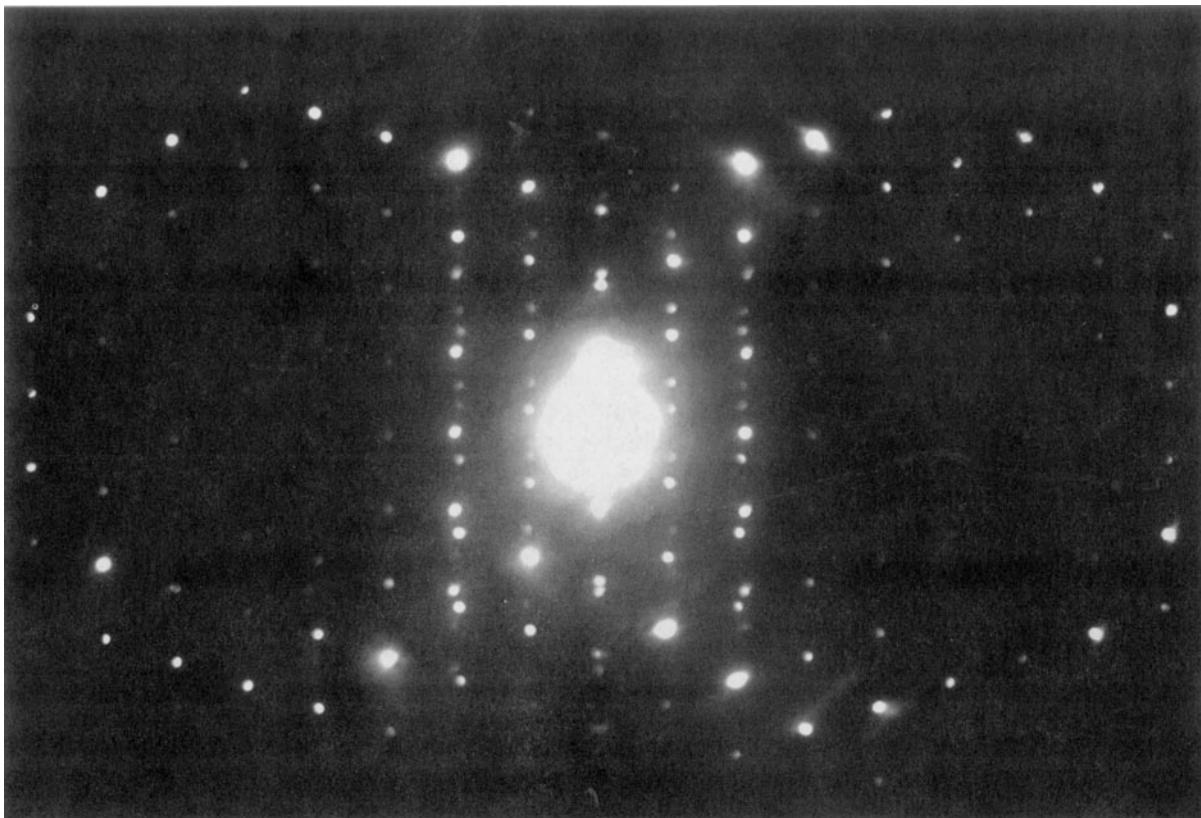
- Second, if we tilt the beam above the specimen, the Ewald sphere moves, because  $\mathbf{k}_i$  tilts, because  $C$  moves!

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 than first meets the eye, as we'll see in Chapter 19.



**Figure 12.6.** In (a)  $s_z = 0$  for  $4G$ . We can change  $s_z$  in two ways: (b) if we tilt the specimen through angle  $\eta$ , the row of spots moves inside the sphere; (c) if we tilt the beam through  $\eta$  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  $\text{MgAl}_2\text{O}_4$  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  $\lambda$ . Since  $d$  does not change but  $\lambda$  does, then Bragg's Law tells us that  $\theta$  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  $\lambda$  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  $\lambda$  is small, the radius of the Ewald sphere,  $\lambda^{-1}$ , 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  $\lambda$  and  $\lambda^{-1}$  as a Function of Accelerating Voltage**

$E$	$\lambda$ (Å)	Radius, $\lambda^{-1}$ (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 1:  $m_0 = 9.109 \times 10^{-31}$  kg,  $c = 2.998 \times 10^8$  ms $^{-1}$ ,  $h = 6.626 \times 10^{-34}$  Nms, and 1eV =  $1.602 \times 10^{-19}$  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  $\mathbf{s}$  to quantify this excitation error. In other words,  $\mathbf{s}$  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

---

13.1. Why Calculate Intensities? .....	203
13.2. The Approach .....	203
13.3. The Amplitude of a Diffracted Beam .....	204
13.4. The Characteristic Length $\xi_g$ .....	205
13.5. The Howie–Whelan Equations .....	206
13.6. Reformulating the Howie–Whelan Equations .....	207
13.7. Solving the Howie–Whelan Equations .....	207
13.8. The Importance of $\gamma^{(1)}$ and $\gamma^{(2)}$ .....	208
13.9. The Total Wave Amplitude .....	209
13.10. The Effective Excitation Error .....	210
13.11. The Column Approximation .....	210
13.12. The Approximations and Simplifications .....	211
13.13. The Coupled Harmonic Oscillator Analog .....	212

## 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  $>1 \mu\text{m}$  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.

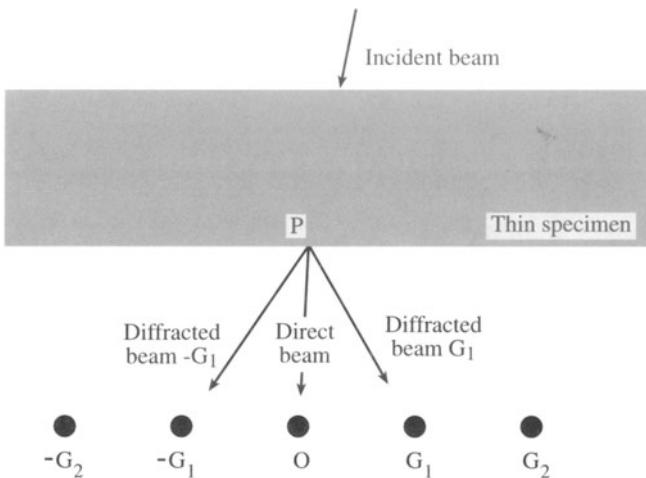
## 13.2. THE APPROACH

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,  $G_i$  ( $i=1, 2, \dots$ ). When discussing images we will often refer to  $\mathbf{g}_i$ , the diffraction vector for the beam  $G_i$ . Then colloquially we will call  $\mathbf{g}$  the “reflection  $\mathbf{g}$ ”; 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

$$A_{\text{cell}} = \frac{e^{2\pi i \mathbf{k} \cdot \mathbf{r}}}{r} \sum_i f_i(\theta) e^{2\pi i \mathbf{K} \cdot \mathbf{r}_i} \quad [13.1]$$

where the summation is over all  $i$  atoms in the unit cell and  $\theta$  is the angle at which the diffracted beam is traveling relative to the incident beam. We have added the term outside

**Table 13.1. Terminology and Notation**

$\Psi^T$ at P	The <i>total</i> 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 $\Psi^T$ but $\phi_g$ and $\phi_0$ .
$\phi_g$	The amplitude of the <i>diffracted</i> beam for reflection $G$ . The intensity is $ \phi_g ^2$ .
$\phi_0$	The amplitude of the <i>direct</i> beam. Don't use the term “transmitted” beam; all the beams we are studying are transmitted. Don't call it the “forward-scattered” beam; diffracted beams can also be forward scattered. $\phi_0$ is a special value of $\phi_g$ for the case where $\mathbf{g} = \mathbf{0}$ .
$\theta$	The angle between a particular set of lattice planes and the direction of the beam scattered constructively by those planes.
$\theta_B$	The Bragg angle; a specific value of $\theta$ when $\mathbf{s} = \mathbf{0}$ .
$d_z$	The thickness of a diffracting slice. This thickness can be as small as we wish to make it; it is not limited to atomic planes.
$\xi_g$	A characteristic length for reflection $g$ ; it is called the <i>extinction distance</i> .
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.)
$\chi$	the electron wave vector in vacuum
$\mathbf{k}$	the electron wave vector in the specimen

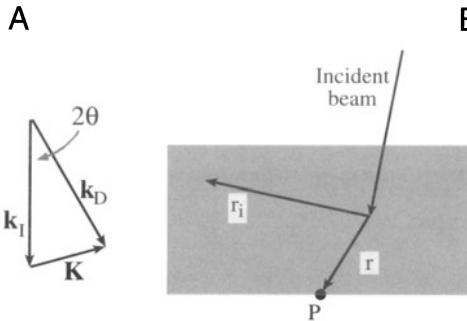
the summation because of how the wave propagates; the  $r^{-1}$  term is present because we have a constant flux of electrons traveling through an expanding spherical surface, radius  $r$ . The quantities  $\mathbf{k}$ ,  $\mathbf{K}$ , and  $\mathbf{r}$  were defined in Chapter 11 and  $f(\theta)$  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.

Figure 13.2 reminds us that  $\mathbf{K} = \mathbf{k}_D - \mathbf{k}_I$ . The vectors  $\mathbf{r}$  and  $\mathbf{r}_i$  are different:  $\mathbf{r}$  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  $\mathbf{K}$  and hence  $\theta$ ).

Therefore, equation 13.1 can be rewritten as

$$A_{\text{cell}} = \frac{e^{2\pi i \mathbf{k} \cdot \mathbf{r}}}{r} F(\theta) \quad [13.2]$$

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  $i$ th 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  $\theta$ ) is denoted as  $\phi_g$  and is given by

$$\phi_g = \frac{\pi a i}{\zeta_g} \sum_n e^{-2\pi i \mathbf{K} \cdot \mathbf{r}_n} e^{2\pi i \mathbf{k}_D \cdot \mathbf{r}} \quad [13.3]$$

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  $\xi_g$  means in equation 13.3; it is a length because  $\phi_g$ , the scattering amplitude, is dimensionless ( $\xi$  (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 $\xi_g$

At this stage in our analysis it is best to think of the quantity  $\xi_g$  as a "characteristic length" for the diffraction vector  $\mathbf{g}$  so as not to have any preconceived ideas of what it represents. A detailed analysis shows that the magnitude of  $\xi_g$  can be expressed as

$$\xi_g = \frac{\pi V_c \cos \theta_B}{\lambda F_g} \quad [13.4]$$

**Table 13.2. Examples of Extinction Distances (in nm)\***

Material	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  $\mathbf{g}$  (i.e.,  $F_g$  is a special value of  $F(\theta)$  when  $\theta$  is the Bragg angle  $\theta_B$ ). The volume of a unit cell,  $V_c$ , is simply  $a/n$ .

$\xi_g$  is the characteristic length for the diffraction vector  $\mathbf{g}$ . We call it the *extinction distance*. The quantity  $\xi_g$  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  $\xi_g$  is a scalar quantity.

From equation 13.4, you can see that the magnitude of  $\xi_g$  is related to  $F_g$  (and through  $V_c$  to the lattice parameter) and the wavelength of the electrons,  $\lambda$ . If the structure factor ( $F_g$ ) is large,  $\xi_g$  will be small. Therefore,  $\xi_g$  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,  $\xi_g$ , 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  $\xi_g$  is illustrated nicely by comparing values of  $\xi_{111}$  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  $\xi_g$  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.

$\xi_g$  depends on the lattice parameters (through  $V_c$ ), the atomic number (through  $F_g$ ), and the kV used (through  $\lambda$ ).

### 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  $\phi_0$  (bold 0 to emphasize that the diffraction vector has zero length) and the amplitudes of the diffracted beams can be written as  $\phi_{g_1}$ ,  $\phi_{g_2}$ , etc. Each beam has an appropriate phase factor. We write  $\psi^T$ , the total wave function, as a series

$$\begin{aligned}\psi^T = & \phi_0 e^{2\pi i \chi_O \cdot r} + \phi_{g_1} e^{2\pi i \chi_{G_1} \cdot r} \\ & + \phi_{g_2} e^{2\pi i \chi_{G_2} \cdot r} + \dots\end{aligned}\quad [13.5]$$

with wave vectors  $\chi_O$  and  $\chi_D$  ( $\chi$  (chi) is pronounced “kai” and rhymes with “sky”);  $\chi_O$  is often written simply as  $\chi$ . We use  $\chi_O$  here to emphasize that it is a vector which terminates on the point O in reciprocal space;  $\chi_{G_1}$  terminates on the “point”  $G_1$ , etc. At this stage, we are using wave vectors  $\chi_O$  and  $\chi_D$  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  $\chi$  as  $\mathbf{k}$ , but there are occasions when the difference is important so we start with  $\chi$  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  $s = 0$ ). All other diffracted beams are weak ( $s >$  or  $< 0$ ), and we ignore their contribution to  $\phi_g$ . Then if the amplitude  $\phi_g$  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  $\phi_g$  and  $\phi_0$  by using the concept introduced in equation 13.3 but replacing  $a$  by the short distance  $dz$

$$d\phi_g = \left\{ \frac{\pi i}{\xi_g} \phi_0 e^{2\pi i (\chi_O - \chi_D) \cdot r} + \frac{\pi i}{\xi_0} \phi_g \right\} dz \quad [13.6]$$

and

$$d\phi_0 = \left\{ \frac{\pi i}{\xi_0} \phi_0 + \frac{\pi i}{\xi_g} \phi_g e^{2\pi i (\chi_D - \chi_O) \cdot r} \right\} dz \quad [13.7]$$

Here  $\chi_O - \chi_D$  is the change in wave vector as the  $\phi_g$  beam scatters into the  $\phi_0$  beam. Similarly,  $\chi_D - \chi_O$  is the change in wave vector as the  $\phi_0$  beam scatters into the  $\phi_g$  beam. Now the *difference*  $\chi_O - \chi_D$  is identical to  $\mathbf{k}_D - \mathbf{k}_O$  although

the individual terms are not equal. Then remember that  $\mathbf{k}_D - \mathbf{k}_O$  ( $= \mathbf{K}$ ) is  $\mathbf{g} + \mathbf{s}$  for the perfect crystal.

You might wonder why we have introduced the wave vector  $\chi$  when it appears to be the same as the  $\mathbf{k}$  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  $\chi$ ) and one in a crystal (wave vector  $\mathbf{k}$ ). Incidentally, the excitation error,  $s$ , should really be written as  $\mathbf{s}_g$ , since it refers to a particular  $\mathbf{g}$  vector. You can think of the parameter  $\xi_0$  as the characteristic length for forward scattering, i.e., scattering from any beam into itself, whereas  $\xi_g$  corresponds to scattering through an angle corresponding to a change of diffraction vector  $\mathbf{g}$ .

The change in  $\phi_g$  depends on the magnitude of both  $\phi_g$  and  $\phi_0$ .

These two equations (13.6 and 13.7) can then be rearranged to give a pair of coupled differential equations. We say that  $\phi_0$  and  $\phi_g$  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

$$e^{-2\pi i s \cdot r} = e^{-2\pi i s z} \quad [13.10]$$

In doing so, we are making the approximation that  $\mathbf{s}$  and  $\mathbf{r}$  are both parallel to  $z$ , i.e., at this time, we ignore components of  $\mathbf{s}$  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  $\phi_0$  is due to the magnitude of  $\phi_0$  itself gives rise to the term *forward scattering*, remember the origin of scattering from Section 2.2. Note that scattering from  $\phi_g$  to  $\phi_g$  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,  $\xi_0$ ; 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)

$$\phi_{0(\text{sub})} = \phi_0 e^{-\frac{\pi i z}{\xi_0}} \quad [13.12]$$

and

$$\phi_{g(\text{sub})} = \phi_g e^{2\pi i s z - \frac{\pi i z}{\xi_0}} \quad [13.13]$$

Then equations 13.8 and 13.9 become

$$\frac{d\phi_{g(\text{sub})}}{dz} = \frac{\pi i}{\xi_g} \phi_{0(\text{sub})} + 2\pi i s \phi_{g(\text{sub})} \quad [13.14]$$

and

$$\frac{d\phi_{0(\text{sub})}}{dz} = \frac{\pi i}{\xi_0} \phi_{g(\text{sub})} \quad [13.15]$$

Since  $\phi_0$  and  $\phi_{0(\text{sub})}$  only differ by a phase factor, we will ignore the difference in calculating intensities since only the amplitude is then important; similarly for  $\phi_g$  and  $\phi_{g(\text{sub})}$ . The result of our substitution is that we have removed the phase factor involving  $\xi_0$ , 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  $\phi_0$

$$\frac{d^2\phi_0}{dz^2} - 2\pi i s \frac{d\phi_0}{dz} + \frac{\pi^2}{\xi_g^2} \phi_0 = 0 \quad [13.16]$$

We can obtain a similar equation for  $\phi_g$  and then obtain solutions for these reformulated expressions.

Note that the only other quantities appearing in this equation for  $\phi_0$  are  $z$ ,  $s$ , and  $\xi_g$ :  $z$  and  $s$  are geometric parameters; the nature of the material only enters through  $\xi_g$ .

## 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,  $\phi_0$ ) must have the form

$$\phi_0 = C_0 e^{2\pi i \gamma z} \quad [13.17a]$$

So we can write that

$$\frac{d\phi_0}{dz} = 2\pi i \gamma C_0 e^{2\pi i \gamma z} \quad [13.17b]$$

and

$$\frac{d^2\phi_0}{dz^2} = -4\pi^2 \gamma^2 C_0 e^{2\pi i \gamma z} \quad [13.17c]$$

What we need to determine is the phase  $\gamma$  and the amplitude  $C_0$ . Note that since  $z$  is a distance in real space, then  $\gamma$  must be a distance in reciprocal space. Substituting this expression into equation 13.16 shows that  $\gamma$  must be a solution to the algebraic equation

$$\gamma^2 - s\gamma - \frac{\xi_g^{-2}}{4} = 0 \quad [13.18]$$

Now  $\phi_g$  is related to  $\phi_0$  through equation 13.15. By substituting equation 13.17 into 13.15 we find that, for each  $\phi_0$ , we also have a  $\phi_g$  given by

$$\phi_g = 2\xi_g \gamma C_0 e^{2\pi i \gamma z} \quad [13.19]$$

To emphasize the similarity to equation 13.17 we can define

$$\phi_g = C_g e^{2\pi i \gamma z} \quad [13.20]$$

Then we can see directly that

$$\frac{C_g}{C_0} = 2\xi_g \gamma \quad [13.21]$$

We've actually got this far without solving any equation! There are two solutions to the quadratic equation (13.18). Use the standard formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad [13.22]$$

to give

$$\gamma^{(1)} = \frac{\left(s - \sqrt{s^2 + \frac{1}{\xi_g^2}}\right)}{2} \quad [13.23a]$$

and

$$\gamma^{(2)} = \frac{\left(s + \sqrt{s^2 + \frac{1}{\xi_g^2}}\right)}{2} \quad [13.23b]$$

We have now found two solutions to the Howie–Whelan equations.

There are two different values for  $\phi_0$  and two corresponding values for  $\phi_g$ .

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

$$\gamma^{(1)} + \gamma^{(2)} = s \quad [13.24]$$

which is a purely geometric quantity, and

$$\gamma^{(1)} \times \gamma^{(2)} = -\frac{1}{4\xi_g^2} \quad [13.25]$$

which is a property of the material. Remember that  $\gamma$  is a length in reciprocal space.

In order to make the equations easier to work with, it is useful to define another quantity,  $w$ , which is *dimensionless* but has the same sign as  $s$ .

$$w = s\xi_g \quad [13.26]$$

In practical situations  $w$  may vary from 0 to  $\pm 10$ . We can then express the two forms of equation 13.21 (because there are *two* values of  $\gamma$ ) in terms of  $\gamma$  or, more conveniently, in terms of  $w$

$$\frac{C_g^{(1)}}{C_0^{(1)}} = 2\xi_g \gamma^{(1)} = w - \sqrt{w^2 + 1} \quad [13.27]$$

and

$$\frac{C_g^{(2)}}{C_0^{(2)}} = 2\xi_g \gamma^{(2)} = w + \sqrt{w^2 + 1} \quad [13.28]$$

(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  $\beta$  by

$$w = \cot \beta \quad [13.29]$$

Now we can impose a restriction on the absolute magnitudes of  $\phi_0$  and  $\phi_g$  so that they satisfy the relations

$$C_0^{(1)2} + C_g^{(1)2} = 1 = C_0^{(2)2} + C_g^{(2)2} \quad [13.30]$$

By normalizing these values for  $C$  separately for each value of  $\gamma$ , 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

$$\begin{aligned} C_0^{(1)} &= \cos \frac{\beta}{2} & C_g^{(1)} &= -\sin \frac{\beta}{2} \\ C_0^{(2)} &= \sin \frac{\beta}{2} & C_g^{(2)} &= \cos \frac{\beta}{2} \end{aligned} \quad [13.31]$$

Now you can understand why we introduced  $\beta$  in equation 13.29. The two independent solutions to the reformulated Howie–Whelan equation for  $\phi_0$  (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  $\phi_g$ .

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  $\gamma$ , 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  $\phi_0$  and  $\phi_g$ , both of which depend on  $\gamma$  (equation 13.17), so equation 13.5 can then be written in terms of both values of  $\gamma$  (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  $\psi^T$ , the total wave function. Alternatively, the total wave function could be some combination of them, i.e., part  $b^{(1)}$  plus part  $b^{(2)}$ . Both of these wave functions are dependent on  $\mathbf{r}$  and have their own values of  $\mathbf{k}$  which we identify as  $\mathbf{k}^{(j)}$ .

Each value of  $\gamma$  gives a different value of  $\mathbf{k}$  which we call  $\mathbf{k}^{(j)}$ .

Thus we can write expressions for  $b^{(1)}$  and  $b^{(2)}$

$$b^{(1)}(\mathbf{k}^{(1)}, \mathbf{r}) = C_0^{(1)} e^{2\pi i \mathbf{k}^{(1)} \cdot \mathbf{r}} + C_g^{(1)} e^{2\pi i (\mathbf{k}^{(1)} + \mathbf{g}) \cdot \mathbf{r}} \quad [13.32]$$

and

$$b^{(2)}(\mathbf{k}^{(2)}, \mathbf{r}) = C_0^{(2)} e^{2\pi i \mathbf{k}^{(2)} \cdot \mathbf{r}} + C_g^{(2)} e^{2\pi i (\mathbf{k}^{(2)} + \mathbf{g}) \cdot \mathbf{r}} \quad [13.33]$$

*Remember:* each of these Bloch-wave functions could be a wave in the crystal—each one depends on only one of the  $\mathbf{k}$  values. In general, the total wave function will be a combination of these two waves. We'll return to the important relationship between  $\mathbf{k}$  and  $\gamma$  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 deter-

mine what  $\phi_0$  and  $\phi_g$  are. The total wave vector,  $\psi^T$ , 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  $\mathbf{g}$ , so this must be the  $\phi_g$  term. We know that at the top of the specimen ( $\mathbf{r} = 0$ ),  $\phi_0$  is unity and  $\phi_g$  is zero (independent of  $\gamma$ )—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  $\phi_0$  and  $\phi_g$ , 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  $\chi_D = \chi_0 + \mathbf{g} + s$  (or  $\mathbf{k}_D = \mathbf{k}_0 + \mathbf{g} + s$ ), where  $\mathbf{k}_0$  is written as  $\mathbf{k}$  and D is  $G_1$  in equation 13.5; then you'll

see that the term containing  $s$  in equation 13.13 drops out.) The  $\phi_0$  and  $\phi_g$  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.

$$(\mathbf{k}^{(2)} - \mathbf{K})_z = \gamma^{(2)} \quad \text{and} \quad (\mathbf{k}^{(1)} - \mathbf{K})_z = \gamma^{(1)} \quad [13.41]$$

What we are interested in is the magnitude of  $\gamma^{(1)}$  and  $\gamma^{(2)}$ . We have also shown directly that  $\phi_0$  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  $\Delta k$  is simply  $|\mathbf{k}^{(2)} - \mathbf{k}^{(1)}|$ . Leaving the term  $e^{\pi i s z}$  in these equations does not affect the amplitudes of  $\phi_0$  and  $\phi_g$ , 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  $\Delta k$  and  $w$ . The term  $\Delta k$  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  $\Delta k$  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,  $|\phi_g|^2 = \phi_g \phi_g^*$ , 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,  $s_{\text{eff}}$ , where

$$s_{\text{eff}} = \sqrt{s^2 + \frac{1}{\xi_g^2}} = \frac{\sqrt{w^2 + 1}}{\xi_g} \quad [13.47]$$

Now the equation becomes

$$|\phi_g|^2 = \left( \frac{\pi t}{\xi_g} \right)^2 \frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} \quad [13.48]$$

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,  $s_{\text{eff}}$ , 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  $I_g$  is proportional to  $\cos^2(\pi t \Delta k)$ .  $I_g$  and  $I_0$  are both periodic in both  $t$  and  $s_{\text{eff}}$ . As  $\phi_g$  increases and decreases,  $\phi_0$  behaves in a complementary manner so that

$$I_0 = 1 - I_g \quad [13.49]$$

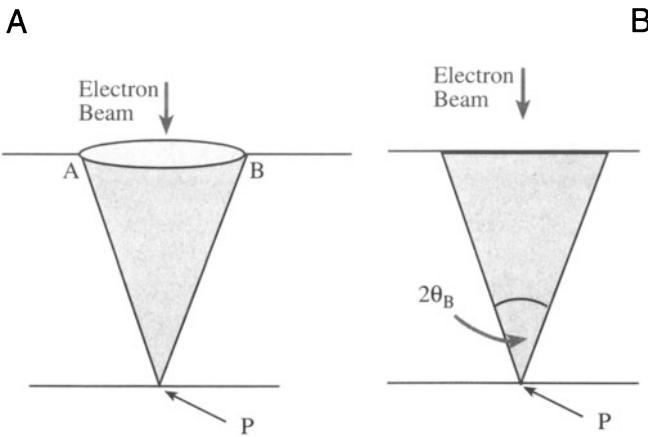
Remember when testing this formula that  $I = \phi \phi^*$  ( $\phi^*$  is the complex conjugate of  $\phi$ ).

The effective excitation error,  $s_{\text{eff}}$ , is a very important quantity. We can summarize some important properties:

- The quantity  $s_{\text{eff}}$  is never zero.
- When  $s$  is zero,  $s_{\text{eff}}$  is  $\xi_g^{-1}$ .
- When  $s$  is very large, then  $s_{\text{eff}}$  becomes essentially the same as  $s$ .

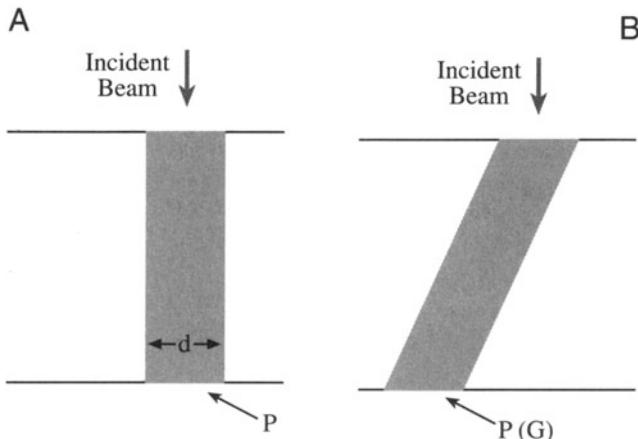
### 13.11. THE COLUMN APPROXIMATION

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  $\phi_0$  and  $\phi_g$  at this point to construct our image. Where do the electrons come from in order to contribute to  $\phi_0$  and  $\phi_g$ ? 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  $\lambda$ . 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  $\sim 2$  nm.

Let's consider some actual numbers: at 100 kV,  $\lambda = 0.0037$  nm,  $\theta_B \sim 0.01$  radians or  $\sim 0.5^\circ$ . So if the thickness ( $t$ ) of the specimen is 100 nm, then AB is  $\sim 2$  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  $\phi_0$  and  $\phi_g$ .

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  $\xi_g$ . 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  $\xi_g$  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  $\mathbf{k}$  and  $\mathbf{r}$

$$e^{2\pi i \mathbf{k} \cdot \mathbf{r}} \quad \text{or} \quad e^{-2\pi i \mathbf{k} \cdot \mathbf{r}} \quad [13.50]$$

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

$$\frac{\hbar^2}{8\pi^2 m} \nabla^2 \psi = -i \frac{\hbar}{2\pi} \frac{d\psi}{dt} \quad [13.51]$$

with the full solution being

$$\psi(\mathbf{r}, t) = A e^{+i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \quad [13.52]$$

- 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

$$|\phi_g|^2 = \sin^2\left(\frac{\pi t}{\xi_g}\right) \quad [13.53]$$

and similarly

$$|\phi_0|^2 = 1 - \sin^2\left(\frac{\pi t}{\xi_g}\right) \quad [13.54]$$

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  $\xi_g$  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  $s = 0$ .

## 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  $\xi_g$ , and explained why this parameter is called the extinction distance. This length was defined in equation 13.4, which shows that  $\xi_g$  depends on the *material*, the *reflection*, and the *wavelength of the electrons*. Two particular points to remember are:

- If the voltage increases then  $\lambda$  decreases and  $\xi_g$  increases.
- The contribution of each Bloch wave is determined by  $\mathbf{s}$ .

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. A* **263**, 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

---

14.1. Wave Equation in TEM .....	217
14.2. The Crystal .....	217
14.3. Bloch Functions .....	219
14.4. Schrödinger's Equation for Bloch Waves .....	219
14.5. The Plane-Wave Amplitudes .....	221
14.6. Absorption of Bloch Waves .....	223

## 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  $\mathbf{g}$  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*,  $\xi_g$ , 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  $\xi_g$ . 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.

---

## 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

$$\left[ -\frac{\hbar^2}{8\pi^2 m} \nabla^2 + V(\mathbf{r}) \right] \Psi(\mathbf{r}) = E \Psi(\mathbf{r}) \quad [14.1]$$

The first term (in  $\nabla^2$ ) represents the kinetic energy, and the second term the potential energy;  $E$  represents the total energy. In TEM we usually talk in terms of the accelerating voltage and the crystal potential and, therefore, we rearrange this equation in terms of voltages. In doing so, we have to be careful about signs, since the charge on the electron is negative and the applied electric field (associated with the accelerating voltage) points toward the gun! All that equation 14.1 says is that the electron has a kinetic energy due to the acceleration it is given in the gun. Initially, this is the total energy of the electron. When the electron passes through the crystal, it will have a potential energy due to the periodic potential associated with the atoms in the crystal.

*The signs:*

- The charge on the electron,  $q$ , is a negative number,  $-e$ , where  $e$  is a positive number.
- The accelerating voltage,  $-E$  (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,  $E$  (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,  $V$ , of the electron outside the crystal is zero; it decreases when the electron is inside the crystal ( $V$  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

$$\nabla^2 \Psi(\mathbf{r}) + \frac{8\pi^2 me}{\hbar^2} [E + V(\mathbf{r})] \Psi(\mathbf{r}) = 0 \quad [14.2]$$

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.

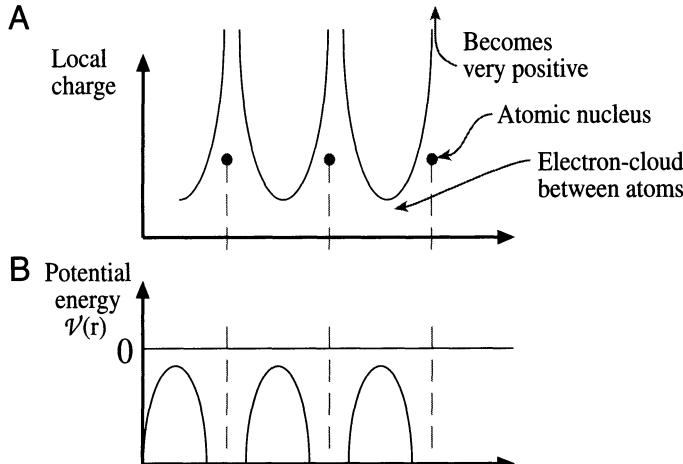
---

## 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

$$V(\mathbf{r}) = V(\mathbf{r} + \mathbf{R}) \quad [14.3]$$

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  $\psi^{\text{tot}}$  (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

$$V(\mathbf{r}) = V^*(\mathbf{r}) \quad [14.4a]$$

Now to make the treatment simple, we consider the case of crystals with a *center of symmetry*

$$V(\mathbf{r}) = V(-\mathbf{r}) \quad [14.4b]$$

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 ex-

press it as a Fourier series in which we sum over all the lattice points in reciprocal space

$$V(\mathbf{r}) = \sum_{\mathbf{g}} V_{\mathbf{g}} e^{2\pi i \mathbf{g} \cdot \mathbf{r}} \quad [14.5]$$

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

$$V_{\mathbf{g}} = \frac{\hbar^2}{2me} U_{\mathbf{g}} \quad [14.6]$$

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

$$V(\mathbf{r}) = \frac{\hbar^2}{2me} \sum_{\mathbf{g}} U_{\mathbf{g}} e^{2\pi i \mathbf{g} \cdot \mathbf{r}} \quad [14.7]$$

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

$$U_{\mathbf{g}} = U_{\mathbf{g}}^* = U_{-\mathbf{g}} \quad [14.8]$$

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  $\xi_0$  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.: $T = 293\text{K}$ )	0.025
Band gap of Si	1.1
Inner potential energy for Si	~11
Energy of electrons in TEM	$\geq 100,000$

**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

$$\psi^{(j)}(\mathbf{r}) = b(\mathbf{k}^{(j)}, \mathbf{r}) \quad [14.9]$$

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

$$\begin{aligned} b^{(j)}(\mathbf{r}) &= b(\mathbf{k}^{(j)}, \mathbf{r}) = \mu(\mathbf{k}^{(j)}, \mathbf{r}) e^{2\pi i \mathbf{k}^{(j)} \cdot \mathbf{r}} \\ &= \mu^{(j)}(\mathbf{r}) e^{2\pi i \mathbf{k}^{(j)} \cdot \mathbf{r}} \end{aligned} \quad [14.10]$$

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  $\mathbf{r}$

$$\mu^{(j)}(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)}(\mathbf{k}^{(j)}) e^{2\pi i \mathbf{g} \cdot \mathbf{r}} \quad [14.11]$$

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

$$b^{(j)}(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i (\mathbf{k}^{(j)} + \mathbf{g}) \cdot \mathbf{r}} \quad [14.12]$$

Using our notation, the superscript on  $C$  indicates that  $C^{(j)}$  depends on  $\mathbf{k}^{(j)}$ . We can now write the expanded expres-

sion for  $b^{(j)}(\mathbf{r})$ , which is a solution to the Schrödinger equation

$$b^{(j)}(\mathbf{r}) = C_0^{(j)} e^{2\pi i \mathbf{k}^{(j)} \cdot \mathbf{r}} + C_{\mathbf{g}}^{(j)} e^{2\pi i (\mathbf{k}^{(j)} + \mathbf{g}) \cdot \mathbf{r}} + \dots \quad [14.13]$$

The first term in this series is  $C_0$ ; 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  $2\pi$  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  $\psi^{\text{tot}}$  using equation 14.9 to give

$$\psi^{\text{tot}} = \sum_{j=1}^n \mathcal{A}^{(j)} \psi^{(j)} = \sum_{j=1}^n \mathcal{A}^{(j)} b(\mathbf{k}^{(j)}, \mathbf{r}) \quad [14.14]$$

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  $\mathbf{g}$ , 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,  $\mathbf{g}$ . The  $U_0$  term does not depend on  $\mathbf{g}$ . 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  $\chi^2$

$$\chi^2 = \frac{2meE}{h^2} \quad [14.20]$$

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,  $\mathbf{p}$  the momentum, and  $\mathbf{k}$  a wave vector. Thus  $\chi$  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  $\chi$ . 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 ( $\lambda$  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  $\geq 1$ ;  $c$  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  $\chi$ , 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  $\mathbf{k}$  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  $b^{(j)}(\mathbf{r})$ .

$$\nabla^2 \psi(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} \nabla^2 \left( e^{2\pi i (\mathbf{k}^{(j)} + \mathbf{g}) \cdot \mathbf{r}} \right) \quad [14.22]$$

$C_{\mathbf{g}}^{(j)}$  does not depend on  $\mathbf{r}$ , 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  $\mathbf{g}$  in the summation in equation 14.19 by  $\mathbf{h}$  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  $\mathbf{g}$

$$\begin{aligned} & \sum_{\mathbf{g}} \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}}^{(j)} \cdot e^{2\pi i (\mathbf{k}^{(j)} + \mathbf{g} + \mathbf{h}) \cdot \mathbf{r}} \\ & = \sum_{\mathbf{g}-\mathbf{h}} \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}-\mathbf{h}}^{(j)} \cdot e^{2\pi i (\mathbf{k}^{(j)} + \mathbf{g}) \cdot \mathbf{r}} \end{aligned} \quad [14.25]$$

Now the sum over all  $\mathbf{g}-\mathbf{h}$  vectors is the same as the sum over all  $\mathbf{g}$  vectors, so we can replace  $\mathbf{g}-\mathbf{h}$  with  $\mathbf{g}$ . (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

$$\sum_{\mathbf{g}} \left( \left\{ -|\mathbf{k}^{(j)} + \mathbf{g}|^2 + \mathcal{K}^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}-\mathbf{h}}^{(j)} \right) e^{2\pi i(\mathbf{k}^{(j)} + \mathbf{g}) \cdot \mathbf{r}} = 0 \quad [14.26]$$

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  $\mathbf{g}$ )

$$\left\{ -|\mathbf{k}^{(j)} + \mathbf{g}|^2 + \mathcal{K}^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq 0} U_{\mathbf{h}} C_{\mathbf{g}-\mathbf{h}}^{(j)} = 0 \quad [14.27]$$

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  $\mathbf{g}$  in equation 14.27. The reason for excluding  $\mathbf{h} = 0$  from the sum is that we have already included it in the first term.

## 14.5. THE PLANE-WAVE AMPLITUDES

We can rewrite and reorder equation 14.27 by, yet again, renaming the variable  $\mathbf{h}$  as  $\mathbf{g}-\mathbf{h}$ . When we do this, we must exclude  $\mathbf{h} = \mathbf{g}$  in the sum

$$\left\{ \mathcal{K}^2 - |\mathbf{k}^{(j)} + \mathbf{g}|^2 \right\} C_{\mathbf{g}}^{(j)} + \sum_{\mathbf{h} \neq \mathbf{g}} U_{\mathbf{g}-\mathbf{h}} C_{\mathbf{h}}^{(j)} = 0 \quad [14.28]$$

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  $\mathbf{g}$  and  $\mathbf{h}$  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  $\mathbf{g} = \mathbf{0}$  in equation 14.27 gives

$$(\mathcal{K}^2 - |\mathbf{k}^{(j)}|^2) C_{\mathbf{0}}^{(j)} + U_{-\mathbf{p}} C_{\mathbf{p}}^{(j)} = 0 \quad [14.29]$$

In deriving this and the following equation, we consider all the possible values of  $\mathbf{h}$  which would give us  $C_{\mathbf{0}}^{(j)}$  or  $C_{\mathbf{p}}^{(j)}$ .

Next, if we let  $\mathbf{g} = \mathbf{p}$  in equation 14.28 and reverse the order of terms to emphasize that we have two equations in  $C_{\mathbf{0}}$  and  $C_{\mathbf{p}}$

$$U_{\mathbf{p}} C_{\mathbf{0}}^{(j)} + (\mathcal{K}^2 - |\mathbf{k}^{(j)} + \mathbf{p}|^2) C_{\mathbf{p}}^{(j)} = 0 \quad [14.30]$$

There are no other possible equations, so to solve these two equations we set the determinant of the coefficients equal to zero

$$\begin{vmatrix} \mathcal{K}^2 - |\mathbf{k}^{(j)}|^2 & U_{-\mathbf{p}} \\ U_{\mathbf{p}} & \mathcal{K}^2 - |\mathbf{k}^{(j)} + \mathbf{p}|^2 \end{vmatrix} = (\mathcal{K}^2 - |\mathbf{k}^{(j)}|^2)(\mathcal{K}^2 - |\mathbf{k}^{(j)} + \mathbf{p}|^2) - U_{\mathbf{p}} U_{-\mathbf{p}} = 0 \quad [14.31]$$

The inner potential of the crystal is usually  $\leq 20$  V while the energy of the electrons is  $\geq 100,000$  eV. Because  $|\mathbf{k}^{(j)} + \mathbf{p}|$  and  $|\mathbf{k}^{(j)}|$  are both very close to  $\mathcal{K}$ , 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!

$$\begin{vmatrix} \mathcal{K}^2 - |\mathbf{k}^{(j)}|^2 & U_{-\mathbf{g}} \\ U_{\mathbf{g}} & \mathcal{K}^2 - |\mathbf{k}^{(j)} + \mathbf{g}|^2 \end{vmatrix} = (\mathcal{K}^2 - |\mathbf{k}^{(j)}|^2)(\mathcal{K}^2 - |\mathbf{k}^{(j)} + \mathbf{g}|^2) - U_{\mathbf{g}} U_{-\mathbf{g}} = 0 \quad [14.32]$$

Now we can use the simple algebraic relation

$$x^2 - y^2 = (x - y)(x + y) \quad [14.33]$$

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

$$(|\mathbf{k}^{(j)}| - \mathcal{K})(|\mathbf{k}^{(j)} + \mathbf{g}| - \mathcal{K}) = \frac{U_g U_{-g}}{4\mathcal{K}^2} = \frac{|U_g|^2}{4\mathcal{K}^2} \quad [14.34]$$

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  $\chi$  can point in any direction.

From equation 14.29 (renaming  $\mathbf{p}$  as  $\mathbf{g}$ ), we have

$$\frac{C_g^{(j)}}{C_0^{(j)}} = \frac{|\mathbf{k}^{(j)}|^2 - \mathcal{K}^2}{U_{-g}} \quad [14.35]$$

which we can rewrite as

$$\frac{C_g^{(j)}}{C_0^{(j)}} = \frac{(|\mathbf{k}^{(j)}| - \mathcal{K})(|\mathbf{k}^{(j)}| + \mathcal{K})}{U_{-g}} \approx \frac{2\mathcal{K}(|\mathbf{k}^{(j)}| - \mathcal{K})}{U_{-g}} \quad [14.36]$$

Thus we can, in principle, say how  $C_0^{(j)}$  and  $C_g^{(j)}$  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

$$\mathbf{A}^{(j)} = \{C_g^{(j)}\} = 0 \quad [14.37]$$

where  $\{C_g^{(j)}\}$  now denotes a column vector with elements  $C_g^{(j)} \mathbf{A}^{(j)}$  is a matrix defined by

$$a_{gg} = K^2 - |\mathbf{k}^{(j)} + \mathbf{g}|^2 \quad [14.38]$$

with the off-diagonal elements given by the Fourier coefficients of the crystal potential

$$a_{gh} = U_{g-h} \quad [14.39]$$

Here,  $g$  refers to rows and  $h$  to columns in the  $\mathbf{A}$  matrix. Except in special cases, such as the two-beam case in equa-

tion 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  $\mathbf{g}_0$ ,  $\mathbf{g}_1$ ,  $2\mathbf{g}_1$ , and  $3\mathbf{g}_1$  beams. The  $5 \times 5$  matrix can be written out (using  $g$  and  $h$  rather than  $\mathbf{g}$  and  $\mathbf{h}$ ) as

$$\mathbf{A} = \begin{pmatrix} a_{-g-g} & U_{-g-0} & U_{-g-g} & U_{-g-2g} & U_{-g-3g} \\ U_{0-(-g)} & a_{00} & U_{0-g} & U_{0-2g} & U_{0-3g} \\ U_{g-(-g)} & U_{g-0} & a_{gg} & U_{g-2g} & U_{g-3g} \\ U_{2g-(-g)} & U_{2g-0} & U_{2g-g} & a_{2g2g} & U_{2g-3g} \\ U_{3g-(-g)} & U_{3g-0} & U_{3g-g} & U_{3g-2g} & a_{3g3g} \end{pmatrix} \quad [14.40]$$

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

$$\mathbf{A} = \begin{pmatrix} a_{-g} & U_{-g} & U_{-2g} & U_{-3g} & U_{-4g} \\ U_g & a_0 & U_{-g} & U_{-2g} & U_{-3g} \\ U_{2g} & U_g & a_g & U_{-g} & U_{-2g} \\ U_{3g} & U_{2g} & U_g & a_{2g} & U_{-g} \\ U_{4g} & U_{3g} & U_{2g} & U_g & a_{3g} \end{pmatrix} \quad [14.41]$$

Some points to notice are:

- The terms  $U_g$ ,  $C_g^{(j)}$ , 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  $C_g^{(j)}$  terms, but we can find the ratios  $C_g^{(j)}/C_0^{(j)}$ .

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  $\{C_g^{(j)}\}$  appears as the eigenvectors and the wave vectors  $\mathbf{k}^{(j)}$  appear as the eigenvalues. He expresses this equation as

$$\mathcal{M}\{C_g^{(j)}\} = \gamma^{(j)}\{C_g^{(j)}\} \quad [14.42]$$

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,  $\xi_{g-h}$ . Remember that  $h$  is the column and notice that the subscript here is  $\mathbf{g}-\mathbf{h}$ ; this extinction distance is related to the interference between the  $\mathbf{g}$  beam and the  $\mathbf{h}$  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  $C_g^{(j)}$  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  $\xi_0$  and  $\xi_g$  in the two-beam case, namely,

$$\xi_0 = \frac{2 \mathcal{K} \cos \theta_B}{U_0} \quad [14.43]$$

and

$$\xi_g = \frac{2 \mathcal{K} \cos \theta_B}{U_g} = \frac{1}{|\Delta \mathbf{k}|} \quad [14.44]$$

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

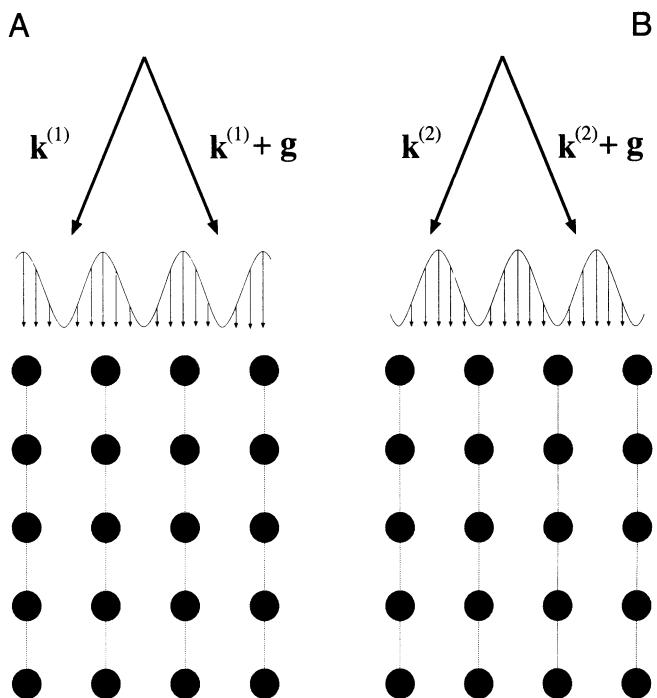
When we have just two beams excited, O and G, we showed in Section 13.9 that we can express the wave function  $\psi$  as

$$\psi(\mathbf{r}) = \mathcal{A}^{(1)} b^{(1)}(\mathbf{r}) + \mathcal{A}^{(2)} b^{(2)}(\mathbf{r}) \quad [14.45]$$

where

$$\mathcal{A}^{(1)} = \cos \frac{\beta}{2}; \quad \mathcal{A}^{(2)} = \sin \frac{\beta}{2} \quad [14.46]$$

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



**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.

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.

## 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  $\phi_0$  and  $\phi_g$  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  $\phi_0$  and  $\phi_g$ ) 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  $\mathbf{k}^{(i)}$  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

---

15.1. Introduction .....	227
15.2. The Dispersion Diagram When $U_g = 0$ .....	228
15.3. The Dispersion Diagram When $U_g \neq 0$ .....	228
15.4. Relating Dispersion Surfaces and Diffraction Patterns .....	229
15.5. The Relation between $U_g$ , $\xi_g$ , and $s_g$ .....	232
15.6. The Amplitudes of Bloch Waves .....	233
15.7. Extending to More Beams .....	234
15.8. Dispersion Surfaces and Defects .....	235

## 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  $\geq 100$  keV.

In this chapter we will see the real origin of the extinction distance  $\xi_g$ , 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  $\mathbf{k}$  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  $U_g$ . (See Section 14.2 for the definition of  $U_g$ .) 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

$$\frac{C_{\mathbf{g}}^{(j)}}{C_{\mathbf{0}}^{(j)}} = \frac{(k^{(j)})^2 - \mathcal{K}^2}{U_{-\mathbf{g}}} = \frac{U_g}{(k^{(j)} + g)^2 - \mathcal{K}^2} \quad [15.1]$$

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

$$b^{(j)}(\mathbf{r}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{(j)} e^{2\pi i (\mathbf{k}^{(j)} + \mathbf{g}) \cdot \mathbf{r}} \quad [15.2]$$

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

$$A_{2n}(\mathbf{k}^{(j)})^{2n} + A_{2n-1}(\mathbf{k}^{(j)})^{2n-1} + \dots = 0 \quad [15.3]$$

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 $U_g = 0$

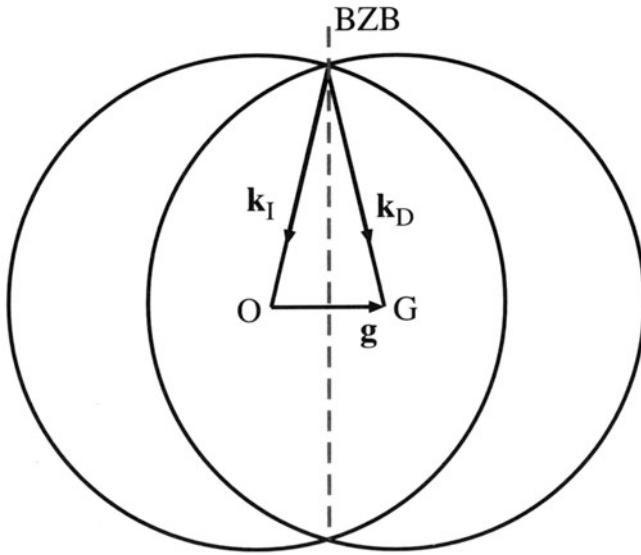
When the electrons are in the vacuum, i.e., outside the specimen, the Fourier coefficients,  $U_g$ , are 0. We start with equation 14.34, namely,

$$(|\mathbf{k}^{(j)}| - \mathcal{K})(|\mathbf{k}^{(j)} + \mathbf{g}| - \mathcal{K}) = \frac{|U_g|^2}{4\mathcal{K}^2} \quad [15.4]$$

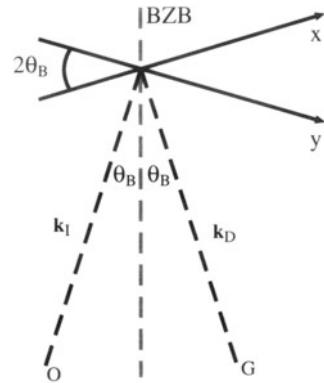
Remember that this equation was derived for the two-beam case. When  $U_g = 0$ , the left side of this equation is zero and the equation has two solutions.

$$\mathcal{K} = |\mathbf{k}^{(j)}| \quad \text{or} \quad \mathcal{K} = |\mathbf{k}^{(j)} + \mathbf{g}| \quad [15.5]$$

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

$$E = \frac{\mathbf{p}^2}{2m} = \frac{\hbar^2 \chi^2}{2m} \quad [15.6]$$

where  $\mathbf{p}$ , the momentum, is related to the wave vector in a vacuum,  $\chi$ , by  $\mathbf{p} = \hbar\chi$ . Here,  $\chi$  is the  $\mathbf{k}$  when the electron is in a vacuum.

Rearranging, we have

$$\chi = \left\{ \frac{2m}{\hbar^2} E \right\}^{\frac{1}{2}} \quad [15.7]$$

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  $\lambda$  is very small.

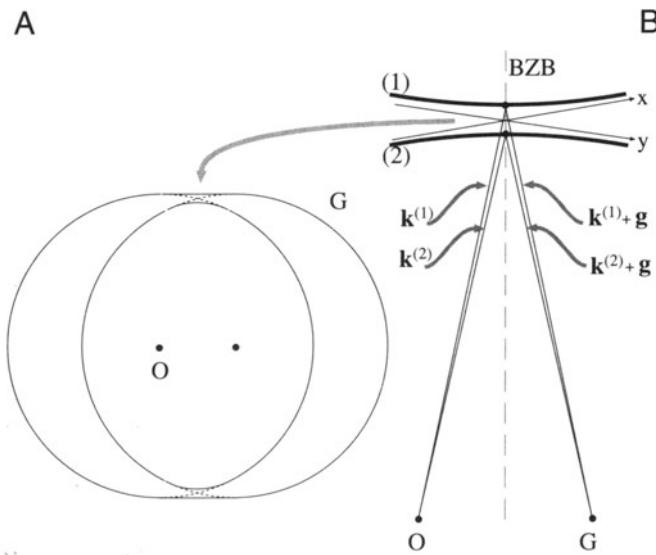
## 15.3. THE DISPERSION DIAGRAM WHEN $U_g \neq 0$

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  $\mathbf{k}$ , 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  $\chi$  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  $C_{\mathbf{g}}^{(2)}$ .

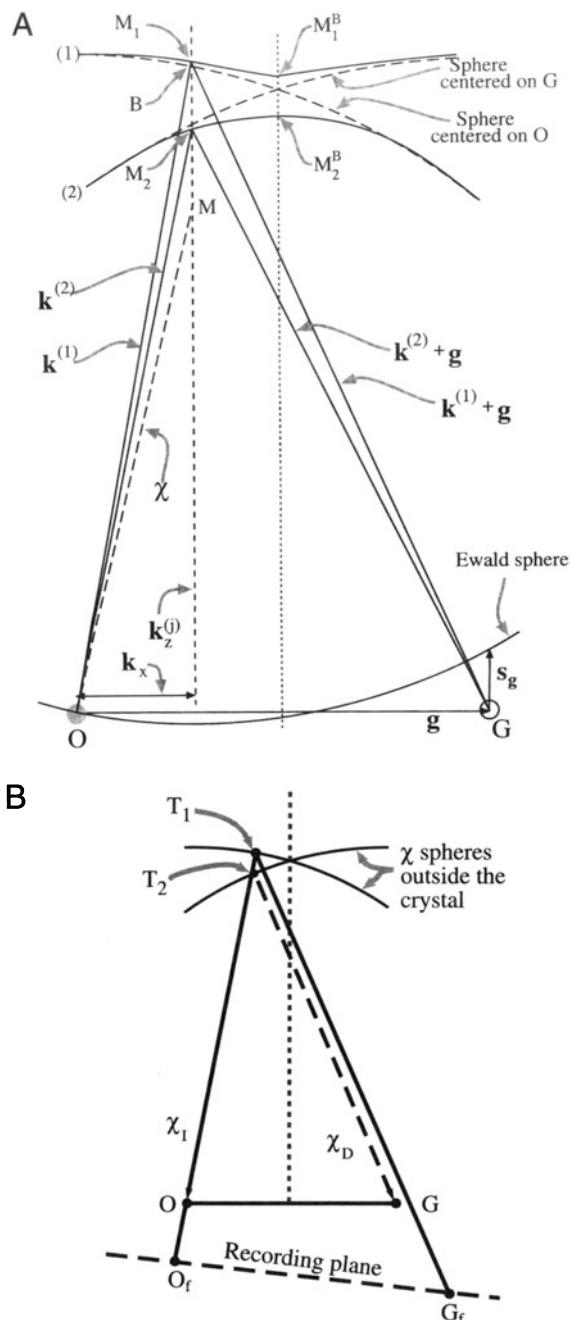
First, we need to know which points on the dispersion surface will actually correspond to the diffraction condition we have chosen. Next, we need to know the orientation of the specimen relative to the beam and the orientation of the Bragg planes.

We begin by considering the case where the surface of the specimen is parallel to  $\mathbf{g}$ ; we will explain why we are so specific on this point in a moment.

Now we have fixed the specimen and  $\mathbf{g}$ . 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  $\Delta k^{-1}$  for  $s = 0$ , as in Section 13.10. If we now tilt the incident beam so that  $\chi$  moves closer to the vertical (keeping the specimen fixed), then the excited points become  $M_1$  and  $M_2$  and, as we see in Figure 15.4,  $s$  becomes negative.

We define the lines  $M_1^B M_2^B$  and  $M_1 M_2$  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  $\mathbf{g}$ .

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  $\mathbf{g}$ , 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  $\chi$  (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  $\chi$  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  $\mathbf{k}$  vectors has an associated wave amplitude  $C_g^{(j)}$  associated with it.

The diagrams of the dispersion surface in Figures 15.4 and 15.5 contain lots of reminders:

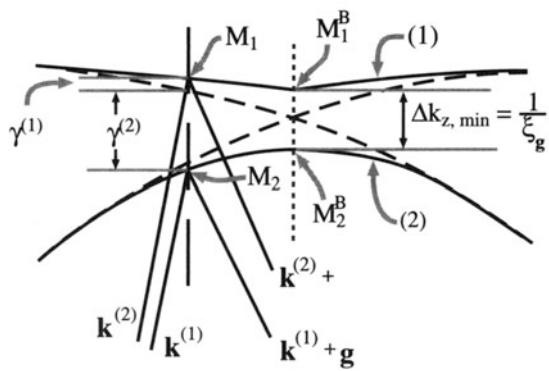
- For this orientation,  $\mathbf{k}_x$  is the same for all  $\mathbf{k}$  vectors ending on O.
- You can recognize  $\gamma^{(1)}$  and  $\gamma^{(2)}$  from Section 13.7.
- The vacuum wave vector  $\chi$  is always shorter than  $\mathcal{K}$  or  $\mathbf{k}$ .

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  $\mathbf{g}$ . Here, the normal to the surface,  $\mathbf{n}$ , is not parallel to the BZB. However, the tie line is always parallel to  $\mathbf{n}$  so the tie line is no longer parallel to the BZB. Remember: this construction matches the components of the  $\mathbf{k}$  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  $\Delta k_z$ .

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  $\mathbf{g}$ . 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  $\chi$  and that  $\chi$  defines a pair of spheres centered on O and G. So we extend the  $\mathbf{n}_2$  tie lines until they reach the  $\chi$  spheres. Now we have excited four points, as we see graphically in Figure 15.6A. The points on the O circle are labeled  $O_1$  and  $O_2$ ; those on the G circle are  $D_1$  and  $D_2$ . We have labeled the subscripts this way because they correspond to the plane waves  $\chi_0^{(1)}$ ,  $\chi_0^{(2)}$ , etc., as also shown in Figure 15.6A.

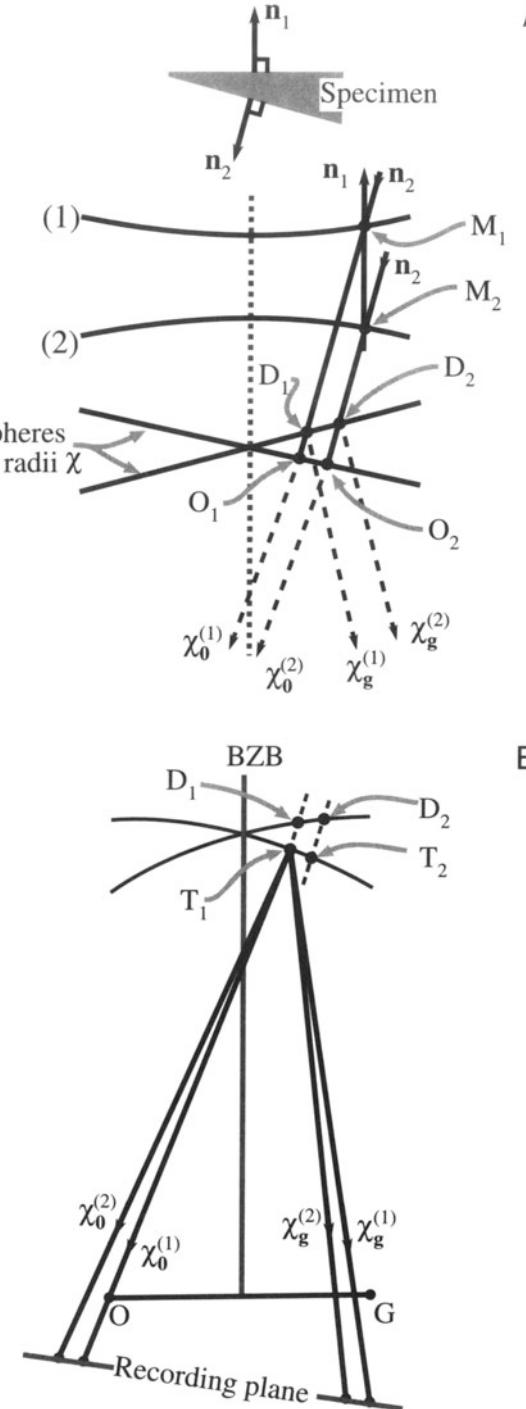
Now we have reached the final step: we have to relate these beams to the DP. Yes, they are real beams because we are now outside the specimen and in a vacuum. We show this in Figure 15.6B. All of the  $\chi$  beams have been related to point  $O_1$  because  $\chi_0^{(1)}$  is the direct beam. Remember:  $\chi_0^{(1)}$  is not vertical because  $\mathbf{g}$  is horizontal. The vectors  $\chi_0^{(1)}$  and  $\chi_0^{(2)}$  are not parallel because they are both radii of the circle  $\chi_0$  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.

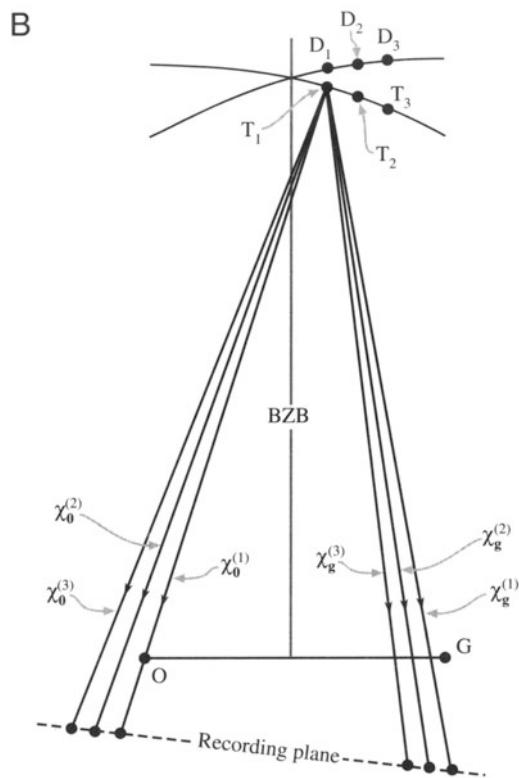
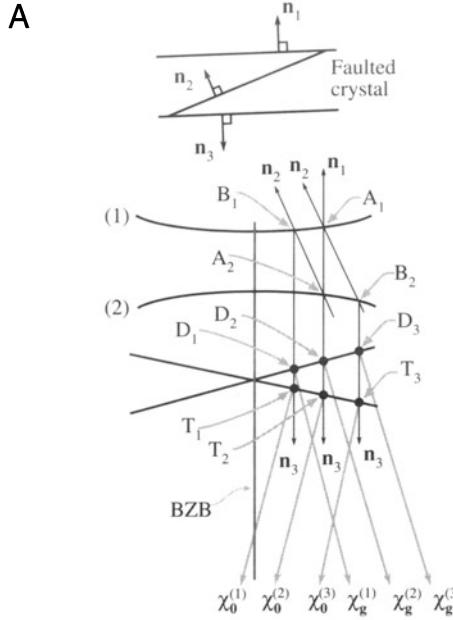
It can be useful to extend the wedge case to the double wedge. For example, imagine an inclined planar defect in a parallel-sided slab with  $\mathbf{g}$  parallel to the slab surface, as shown in Figure 15.7. Everything is as before at the top surface. At the inclined interface then, tie lines do create 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  $\chi$  spheres and find that now we have three  $\chi_0$  vectors and three  $\chi_D$  vectors. Translating these  $\chi$  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  $\mathbf{g}$  (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  $\mathbf{g}$  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  $\chi$  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 $U_g$ , $\xi_g$ , 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  $\chi$  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  $\mathbf{k}$  vectors in the crystal are  $\mathbf{k}^{(1)}$  and  $\mathbf{k}^{(2)}$ . One  $\mathbf{k}$  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  $\mathbf{k}$  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  $\mathbf{k}_g^{(2)}$  and  $\mathbf{k}_g^{(1)}$  by adding  $\mathbf{g}$ . Now, how does  $\mathbf{k}_g^{(1)}$ , say, relate to  $\mathbf{K}$ ? The point  $K$  is also determined by the tie line through  $\chi$ , and lies on the circle centered on  $O$ . Most importantly, neither  $\mathbf{k}_1$  nor  $\mathbf{k}_2$  is equal to  $\mathbf{K}$ . If you look back at equation 13.41 you can see that

$$\mathbf{k}_z^{(i)} - \mathbf{K}_z = \gamma^{(i)} \quad [15.8]$$

So  $\gamma^{(i)}$  is simply the distance of the point  $M_i$  from the  $\mathcal{K}$  sphere centered on  $O$ . We can write this relationship explicitly

$$\mathbf{k}_z^{(i)} = \mathbf{k}_z^{(1)} + \mathbf{k}_x^{(i)} \quad [15.9]$$

$$= (K_z + \gamma^{(i)}) \mathbf{u}_z + k_x \mathbf{u}_x \quad [15.10]$$

Notice that the last term here is independent of  $i$ . Look again at Figure 15.4. You can see that  $\Delta k_z$  is a minimum when  $M_1$  and  $M_2$  lie on the BZB. In that situation

$$\Delta k_{z_{\min}} = \gamma^{(1)} - \gamma^{(2)} \quad [15.11]$$

Simply by looking at the diagram, and as expected from Chapter 13, you also know that

$$\gamma^{(1)} - \gamma^{(2)} = \frac{U_g}{k} = \frac{1}{\xi_g} \quad [15.12]$$

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  $\Delta k_z$  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  $\Delta k_z$ .
- Hence we have a finite extinction distance (i.e.,  $\xi_g$  is not infinite).

An aside: think how  $s_{\text{eff}}$  and  $s$  would be related if  $\xi_g$  were infinite. (Go back to equation 13.47.)

If the tie line  $M_1M_2$  does not lie on the BZB, then when we draw the Ewald sphere centered just below  $M_1$  (with radius of length  $1/\lambda$  or  $|K|$ ) we see that  $s_g$  is nonzero. We can easily see from the equations in Section 13.10 that, in general,  $\Delta k_z$  is given by

$$\Delta k_z = s_{\text{eff}} = \frac{1}{\xi_{\text{eff}}} \quad [15.14]$$

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  $\Delta k$  increases.

Some questions raised here are:

- What is the physical reason that  $\Delta k_z$  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

$$\Psi(\mathbf{r}) = \mathcal{A}^{(1)} b^{(1)} + \mathcal{A}^{(2)} b^{(2)} \quad [15.15]$$

The relative contributions of the two Bloch waves,  $\mathcal{A}^{(1)}$  and  $\mathcal{A}^{(2)}$ , were shown to be  $\cos \beta/2$  and  $\sin \beta/2$ , respectively, and  $w = \cot \beta = s\xi_g$ .

We also showed in Section 13.8 that

$$b^{(1)}(\mathbf{k}^{(1)}, \mathbf{r}) = C_0^{(1)} e^{2\pi i \mathbf{k}^{(1)} \cdot \mathbf{r}} + C_g^{(1)} e^{2\pi i (\mathbf{k}^{(1)} + \mathbf{g}) \cdot \mathbf{r}} \quad [15.16]$$

and

$$b^{(2)}(\mathbf{k}^{(2)}, \mathbf{r}) = C_0^{(2)} e^{2\pi i \mathbf{k}^{(2)} \cdot \mathbf{r}} + C_g^{(2)} e^{2\pi i (\mathbf{k}^{(2)} + \mathbf{g}) \cdot \mathbf{r}} \quad [15.17]$$

The Bloch-wave coefficients were given by equation set 13.31

$C_0^{(1)}$	$C_0^{(2)}$	$C_g^{(1)}$	$C_g^{(2)}$
$\cos \beta/2$	$\sin \beta/2$	$-\sin \beta/2$	$\cos \beta/2$

Now we can consider some special cases and examine the actual values for  $C_0^{(1)}$ ,  $\mathcal{A}^{(1)}$ , 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  $\cos(\beta/2) > \sin(\beta/2)$  so that  $\mathcal{A}^{(1)}$  is greater than  $\mathcal{A}^{(2)}$ . If we reverse the sign of  $s$ ,  $\cos(\beta/2) < \sin(\beta/2)$  and  $\mathcal{A}^{(1)}$  is less than  $\mathcal{A}^{(2)}$ .

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**

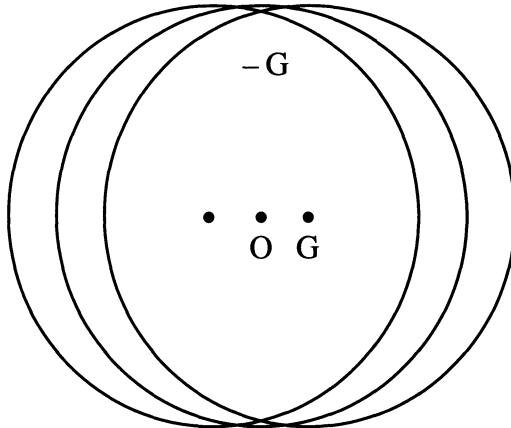
$s$	$w$	$\beta$	$\beta/2$	$\cos(\beta/2)$	$\sin(\beta/2)$
0	0	$\pi/2$	$\pi/4$	$1/\sqrt{2}$	$1/\sqrt{2}$
+0.01	$+\Delta$	$\pi/2 - \delta$	$\pi/4 - \delta/2$	$1/\sqrt{2} + \epsilon$	$1/\sqrt{2} - \epsilon$
-0.01	$-\Delta$	$\pi/2 + \delta$	$\pi/4 - \delta/4$	$1/\sqrt{2} - \epsilon$	$1/\sqrt{2} + \epsilon$

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

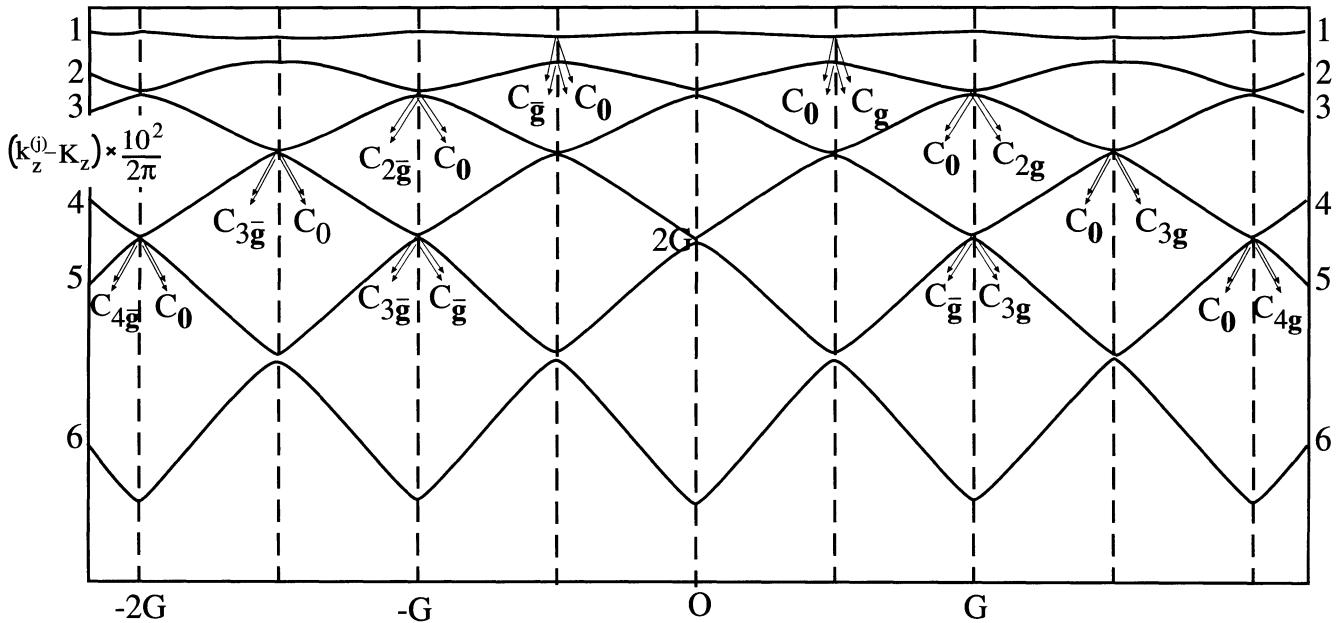
If we allow more beams to contribute to the image, we can picture the dispersion surface for the case where  $U_g = 0$  by constructing more spheres, shown in Figure 15.8. If we have  $n$  beams, then we will have  $n$  spheres. Note that each sphere is centered on its corresponding reciprocal lattice 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  $\mathbf{g}$  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  $3\mathbf{g}$  is excited. We will actually consider the two-beam condition, where  $\mathbf{0}$  and  $3\mathbf{g}$  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  $O, \pm G$ , and  $\pm 2G$ , 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  $\mathbf{0}$ ,  $\mathbf{g}$ , 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  $\mathbf{0}$  and  $\mathbf{g}$ ; they intersect on the BZB which passes through  $\mathbf{g}/2$ , so the  $C_0$ ,  $C_g$  are labeled as shown.

Similarly, the spheres centered on  $\mathbf{0}$  and  $3\mathbf{g}$  intersect on the BZB which passes through  $3\mathbf{g}/2$ , 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.,  $\mathbf{0}$  and  $n\mathbf{g}$ , and will be located on the  $n\mathbf{g}/2$  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  $\mathbf{g}$  is excited is that they do not intersect the  $\mathbf{0}$  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  $C_0^{(1)}$ ,  $C_0^{(2)}$ ,  $C_{0g}^{(1)}$ ,  $C_{2g}^{(1)}$ , and  $C_{2g}^{(2)}$  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  $\mathbf{s}$  values for  $\bar{\mathbf{g}}$  and  $3\mathbf{g}$  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  $\mathbf{g}$  and  $3\mathbf{g}$  when  $2\mathbf{g}$  is strongly excited is  $\xi_{4g}$  ( $\xi_{3g - (-g)}$ ). 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  $\mathbf{k}$  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  $\mathbf{k}$  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  $\Delta k_z$ ,  $s_{\text{eff}}$ , and  $\xi_{\text{eff}}$ . Notice that  $\Delta k_z$  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.  $\Delta k$  is only nonzero because we have a crystal.  $\Delta k$  gives rise to thickness fringes and all thickness effects. Thus we see that thickness variations are due to the interference, or beating, of pairs of Bloch waves. As we increase  $n$ ,  $\xi_g$  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  $\omega$  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 Mermin, 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

---

16.1. Review of Diffraction from a Primitive Lattice .....	239
16.2. Structure Factors: The Idea .....	240
16.3. Some Important Structures: bcc, fcc, and hcp .....	241
16.4. Extending fcc and hcp to Include a Basis .....	242
16.5. Applying the bcc and fcc Analysis to Simple Cubic .....	243
16.6. Extending hcp to Ti <sub>3</sub> Al .....	244
16.7. Superlattice Reflections and Imaging .....	244
16.8. Diffraction from Long-Period Superlattices .....	246
16.9. Forbidden Reflections .....	247
16.10. Using the International Tables .....	247

## 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

$$\phi_g = \frac{ai\lambda F_g}{V_c \cos \Theta} \sum_n e^{-2\pi i \mathbf{k} \cdot \mathbf{r}_n} e^{2\pi i \mathbf{k}_n \cdot \mathbf{r}} \quad [16.1]$$

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  $\theta$  (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  $\theta$  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)

$$|f(\theta)|^2 = \frac{d\sigma(\theta)}{d\Omega} \quad [16.2]$$

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  $\theta = \theta_B$**

Element	$f(\theta)(\text{\AA})$	Element	$f(\theta)(\text{\AA})$
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

<sup>a</sup>These are values given by Edington (1976) using a self-consistent field theory ( $\sin \theta/\lambda = 0.2 \text{\AA}^{-1}$ ) 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  $\mathbf{g}$  can give a reflection in the DP. Each reciprocal lattice point will then correspond to a possible beam. The next step will be to add the basis (i.e., the group of atoms associated with each lattice point) to the primitive lattice. Since we still have the primitive lattice, all of these points will still exist in the reciprocal lattice but the reflections will be weighted. You will find that there are three different ways to look at the situation, which in fact are all equivalent:

- *Selection rules:* This is perhaps closest to physics. The structure of the crystal imposes certain selection rules which determine which beams are allowed.
- *Weights (or weighting factors):* We can assign a weight (which may be zero) to each of the points in the reciprocal lattice. This is the terminology used by Ewald. The nice feature about weighting factors is that they are analogous to scattering factors.
- *Structure factors ( $F$ ):* These are the unit-cell equivalents of the atomic scattering amplitude,  $f(\theta)$ ; they can be thought of as unit-cell scatter-

ing 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<sub>3</sub>Al, the 100 reflection is weak but in Ni it is absent.
- Some materials have a special lattice in real space, for example, fcc or bcc lattices. In these cases, we can describe a corresponding special lattice in reciprocal space. What this means is that certain reflections are always forbidden for these particular structures; these are known as "kinematically forbidden" reflections. (We'll see, however, that they can be present due to dynamical scattering events, and structure factors do not take any account of dynamical scattering.) The reciprocal lattice (of allowed reflections) of an fcc crystal is bcc, and vice versa.

In equation 13.1 we described the scattering from the unit cell by the expression

$$A_{\text{cell}} = \frac{e^{2\pi i k r}}{r} \sum_i f_i(\theta) e^{2\pi i \mathbf{K} \cdot \mathbf{r}_i} \quad [16.3]$$

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

$$\mathbf{r}_i = x_i \mathbf{a} + y_i \mathbf{b} + z_i \mathbf{c} \quad [16.4]$$

We'll start by considering only the case where  $\mathbf{K} = \mathbf{g}$  since this is an infinite, perfect crystal

$$\mathbf{K} = h \mathbf{a}^* + k \mathbf{b}^* + \ell \mathbf{c}^* \quad [16.5]$$

So we can write

$$F_{hkl} = \sum_i f_i e^{2\pi i (hx_i + ky_i + \ell z_i)} \quad [16.6]$$

This is our key equation; it is completely general.

This equation applies whether there is one atom or one hundred atoms in the unit cell, no matter where they are located, and it applies to all crystal lattices. What we do

next is simply insert the atomic coordinates into equation 16.6 and calculate  $F_{hkl}$ .

### 16.3. SOME IMPORTANT STRUCTURES: bcc, fcc, AND hcp

We will now calculate the structure factor for bcc and fcc crystals, because they illustrate the points we just made in Section 16.2 and because, as a materials scientist, you must know these results. You can regard the reciprocal lattice in two ways:

- The reciprocal 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

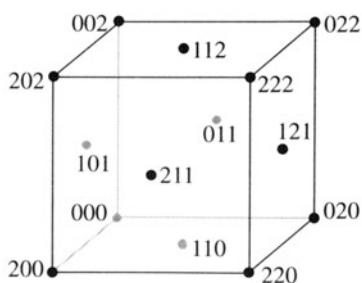
$$F = f \left\{ 1 + e^{\pi i(h+k+\ell)} \right\} \quad [16.7]$$

Now, since  $h, k, \ell$  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.

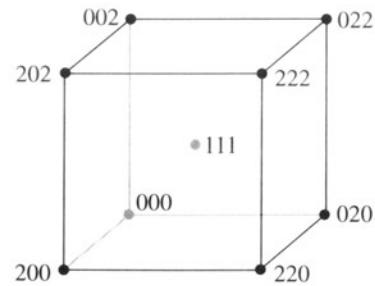
Thus, we can say that:

- $F = 2f$  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

$$(\mathbf{x}, \mathbf{y}, \mathbf{z}) = (0, 0, 0), (\frac{1}{2}, \frac{1}{2}, 0), (\frac{1}{2}, 0, \frac{1}{2}), (0, \frac{1}{2}, \frac{1}{2}) \quad [16.8]$$

Substituting these values for  $\mathbf{r}_i$  into equation 16.6 gives

$$F = f \left\{ 1 + e^{\pi i(h+k+\ell)} + e^{\pi i(h+\ell)} + e^{\pi i(k+\ell)} \right\} \quad [16.9]$$

Again, we consider the possible values of the integers  $h, k, \ell$ . If all three are either odd or even, then all of the exponential terms are  $e^{2\pi i}$ . Therefore, all the phases of the diffracted waves are multiples of  $2\pi$  and are in phase. However, if one of  $h, k$ , or  $\ell$  is odd but the other two are even, or vice versa, then two of the three phase factors will be odd multiples of  $\pi$ , giving two terms of -1 in equation 16.9. Therefore:

- $F = 4f$  if  $h, k, \ell$  are all even or all odd,
- $F = 0$  if  $h, k, \ell$  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

$$(x, y, z) = (0, 0, 0), \left(\frac{1}{3}, \frac{2}{3}, \frac{1}{2}\right) \quad [16.10]$$

Substituting these values for  $\mathbf{r}_i$  into equation 16.6 gives

$$F = f \left\{ 1 + e^{2\pi i \left( \frac{h}{3} + \frac{2k}{3} + \frac{\ell}{2} \right)} \right\} \quad [16.11]$$

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

$$|F|^2 = f^2 (1 + e^{2\pi i X}) (1 + e^{-2\pi i X}) = f^2 (2 + e^{2\pi i X} + e^{-2\pi i X}) \quad [16.12]$$

$$|F|^2 = f^2 (2 + 2 \cos 2\pi X) = f^2 (4 \cos^2 \pi X) \quad [16.13]$$

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  $\ell$  is odd,
- $|F|^2 = 4f^2$  if  $h + 2k = 3m$  and  $\ell$  is even,
- $|F|^2 = 3f^2$  if  $h + 2k = 3m + 1$  and  $\ell$  is odd,
- $|F|^2 = f^2$  if  $h + 2k = 3m + 1$  and  $\ell$  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  $(\bar{1}\bar{2}0)$  are crystallographically equivalent.

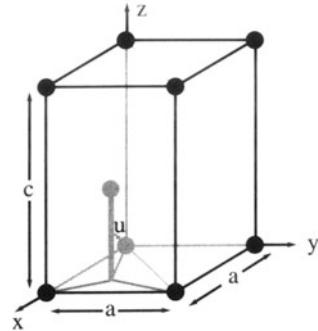
You need to know a few other expressions for this system. If you are working with hcp materials, you *must* have a copy of Frank's 1965 paper on indexing this system.

If the direction  $[uvw]$  lies in the plane  $(hki\ell)$ , then we can show that

$$uh + vk + ti + w\ell = 0 \quad [16.14]$$

The normal to the plane  $(h, k, i, \ell)$  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

$$[u, v, t, \lambda w] \cdot [h, k, i, \ell/\lambda] = 0 \quad [16.15]$$



**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  $[hkl]$  is always normal to the plane  $(hkl)$ , but this is not the case for hcp crystals. You can show using some simple geometry that

$$\lambda^2 = \left(\frac{2}{3}\right) \left(\frac{c}{a}\right)^2 \quad [16.16]$$

Thus the Cartesian vector  $[HKIL]$ , which is normal to the plane  $(hki\ell)$ , is the vector

$$\left(h, k, i, \frac{3}{2} \left(\frac{a}{c}\right)^2 \ell\right) \quad [16.17]$$

So,  $[11\bar{2}0]$  is normal to the  $(11\bar{2}0)$  plane because  $\ell$  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,  $\phi$ , 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

$$\cos \phi = \frac{hd + ke + \frac{1}{2}(he + kd) + \frac{3}{4}\ell g \left(\frac{a}{c}\right)^2}{\sqrt{h^2 + k^2 + hk + \frac{3}{4}\ell^2 \left(\frac{a}{c}\right)^2} \sqrt{d^2 + e^2 + de + \frac{3}{4}g^2 \left(\frac{a}{c}\right)^2}} \quad [16.18]$$

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

$$F = \{f_{\text{Na}} + f_{\text{Cl}} e^{\pi i(h+k+\ell)}\} \{1 + e^{\pi i(h+k)} + e^{\pi i(h+\ell)} + e^{\pi i(k+\ell)}\} \quad [16.19]$$

This again gives rise to some rules:

- $F = 4(f_{\text{Na}} + f_{\text{Cl}})$  if  $h, k, \ell$  are all even,
- $F = 4(f_{\text{Na}} - f_{\text{Cl}})$  if  $h, k, \ell$  are all odd,
- $F = 0$  if  $h, k, \ell$  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  $f_{\text{Cl}}$  is zero. Whether the sign in  $(f_{\text{Na}} \pm f_{\text{Cl}})$  is positive or negative is the new feature. What this means in practice is that reflections with  $h, k, \ell$  all even will appear much more intense in the DP than those with  $h, k, \ell$  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  $4(f_{\text{Na}} - f_{\text{Cl}})$  will be different in each case. Reflections with  $h, k, \ell$  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  $F_{\text{fcc}}$ )

$$F = \left\{ f_{\text{Ga}} + f_{\text{As}} e^{\frac{\pi i}{2}(h+k+\ell)} \right\} F_{\text{fcc}} \quad [16.20]$$

So the rules are slightly more complicated:

- $F = 0$  if  $h, k, \ell$  are mixed as always for fcc,
- $F = 4(f_{\text{Ga}} \pm i f_{\text{As}})$  if  $h, k, \ell$  are all odd,
- $F = 4(f_{\text{Ga}} - f_{\text{As}})$  if  $h, k, \ell$  are all even and  $h + k + \ell = 2N$  where  $N$  is odd (e.g., the 200 reflection),
- $F = 4(f_{\text{Ga}} + f_{\text{As}})$  if  $h, k, \ell$  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  $16(f_{\text{Ga}}^2 + f_{\text{As}}^2)$  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  $f_{\text{Ga}}$  and  $f_{\text{As}}$  in our results with  $f_{\text{Si}}$ . 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

$$F = \{f_{\text{Ni}} + f_{\text{Al}} e^{\pi i(h+k+\ell)}\} \quad [16.21]$$

This leads to two values for  $F$ , neither of which is zero:

- $F = f_{\text{Ni}} + f_{\text{Al}}$  if  $h+k+\ell$  is even,
- $F = f_{\text{Ni}} - f_{\text{Al}}$  if  $h+k+\ell$  is odd.

This would, of course, be the bcc result if  $f_{\text{Ni}}$  and  $f_{\text{Al}}$  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<sub>3</sub>Au (LI<sub>2</sub>) structure:* There are many important ordered intermetallics with this structure such as Al<sub>3</sub>Li and Fe<sub>3</sub>Al. The most important is Ni<sub>3</sub>Al (because of its role in Ni-base superalloys.) We can treat Ni<sub>3</sub>Al 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

$$F = f_{\text{Al}} + f_{\text{Ni}} \{e^{\pi i(h+k)} + e^{\pi i(h+\ell)} + e^{\pi i(k+\ell)}\} \quad [16.22]$$

The rules for Ni<sub>3</sub>Al are:

- $F = (f_{\text{Al}} + 3f_{\text{Ni}})$  if  $h, k, \ell$  are all even or all odd,
- $F = (f_{\text{Al}} - f_{\text{Ni}})$  if  $h, k, \ell$  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  $hkl$  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  $hkl$  will be zero. For this reason, reflections with mixed  $hkl$  are referred to as superlattice reflections (see Section 16.7).

## 16.6. EXTENDING hcp TO TiAl

The TiAl structure is not as well known as the previous two cases, but illustrates a similar class of materials. We noted

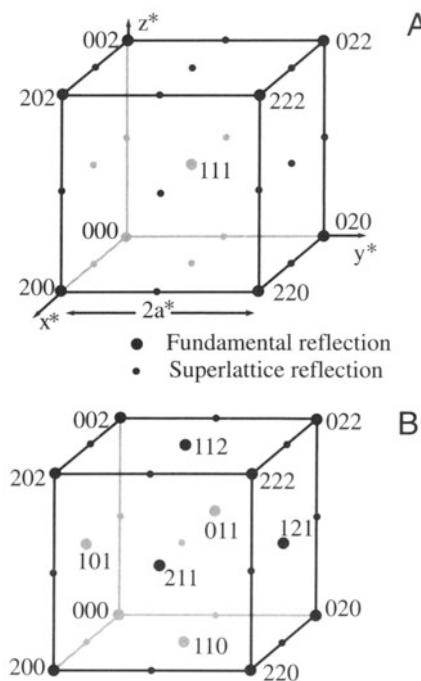
in Section 16.4 that the two atoms in the hcp structure are not equivalent. In TiAl, we actually make them chemically distinct, too. This means that the rules for hcp will be modified again. Using equation 16.11, we find that

$$F = f_{\text{Ti}} + f_{\text{Al}} e^{2\pi i(\frac{h}{3} + \frac{2k}{3} + \frac{\ell}{2})} \quad [16.23]$$

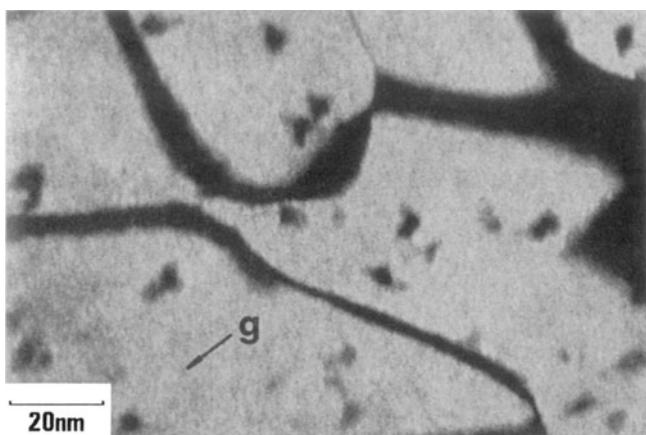
The most important result is that the  $(0001)$  reflection is now allowed since  $F = f_{\text{Ti}} - f_{\text{Al}}$ . TiAl really does have a primitive hexagonal unit cell.

## 16.7. SUPERLATTICE REFLECTIONS AND IMAGING

The reciprocal lattices for Ni<sub>3</sub>Al 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<sub>3</sub>Al and (B) the NiAl structures. In (A) Ni<sub>3</sub>Al is fcc, so the fcc forbidden reflections ( $h, k, \ell$  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.

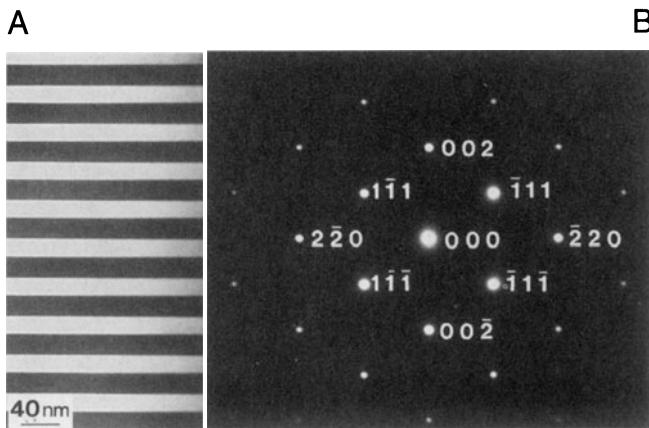


**Figure 16.5.** DF image from a chemically sensitive 110 reflection showing bright ordered domains in  $\text{Cu}_3\text{Au}$ . The dark areas in the bright domains are regions of local disorder induced by ion beam damage.

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.

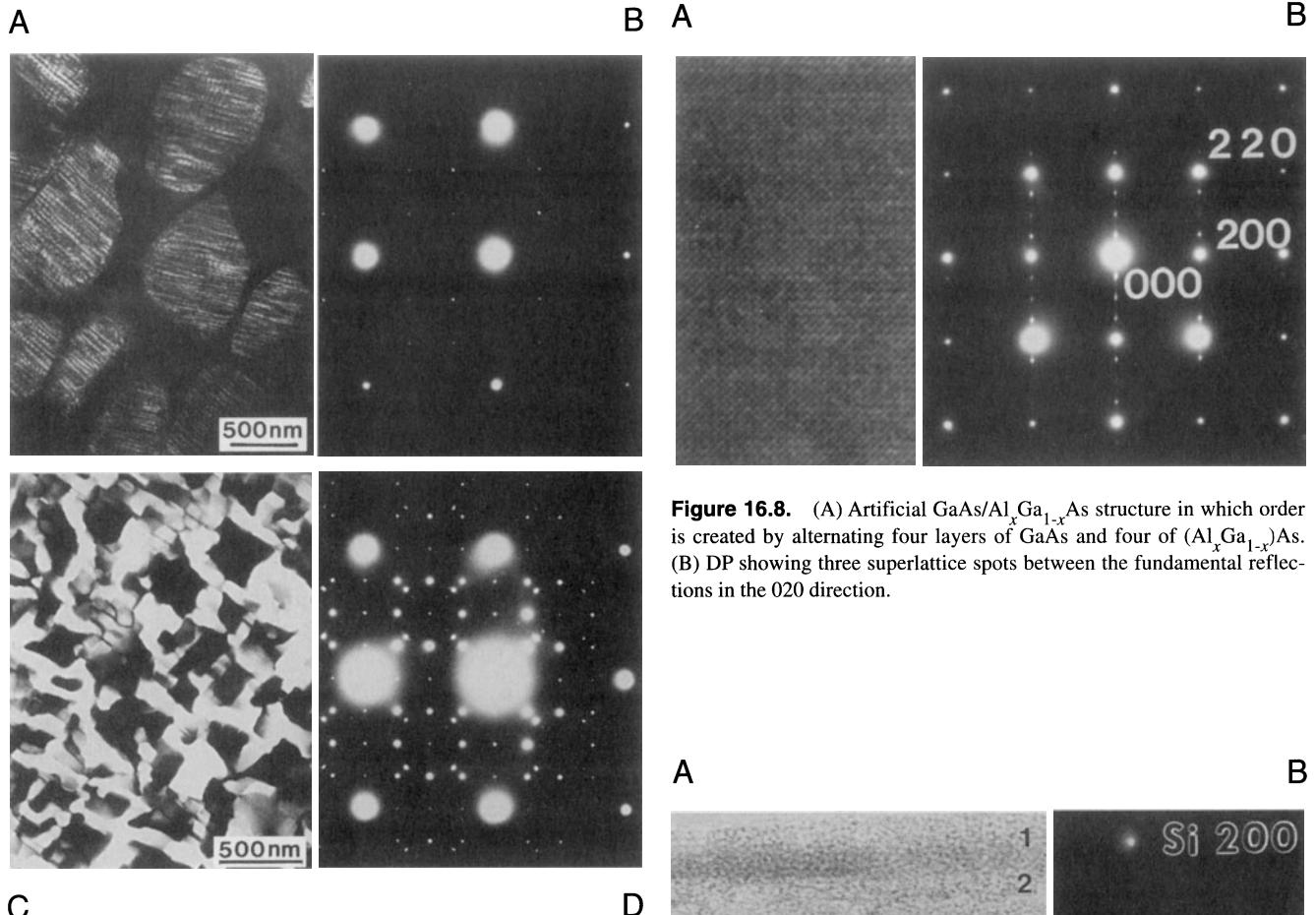


**Figure 16.6.** (A) DF image from a 002 superlattice reflection in GaAs. The  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  is the lighter region because Al has replaced Ga in the GaAs (darker regions). (B) Diffraction pattern showing the less intense 002 and other superlattice reflections.

Figure 16.5 shows an image from  $\text{Cu}_3\text{Au}$ , the archetypal  $\text{A}_3\text{B}$  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  $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$  quantum well structure. The  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  appears lighter than the GaAs because the 002 reflection is a superlattice reflection; remember, it would be forbidden for GaAs if  $f_{\text{Ga}}$  and  $f_{\text{As}}$  were equal. The reason the  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  appears lighter is that we have replaced a fraction  $x$  of the Ga atoms with the lighter Al atoms, thus increasing the difference  $f_{\text{III}} - f_{\text{V}}$ . 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  $\text{V}_x\text{C}_y$ , where  $x > y$ . The two images and DPs shown in Figure 16.7 were taken from well-ordered  $\text{V}_6\text{C}_5$  and  $\text{V}_8\text{C}_7$ , 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).



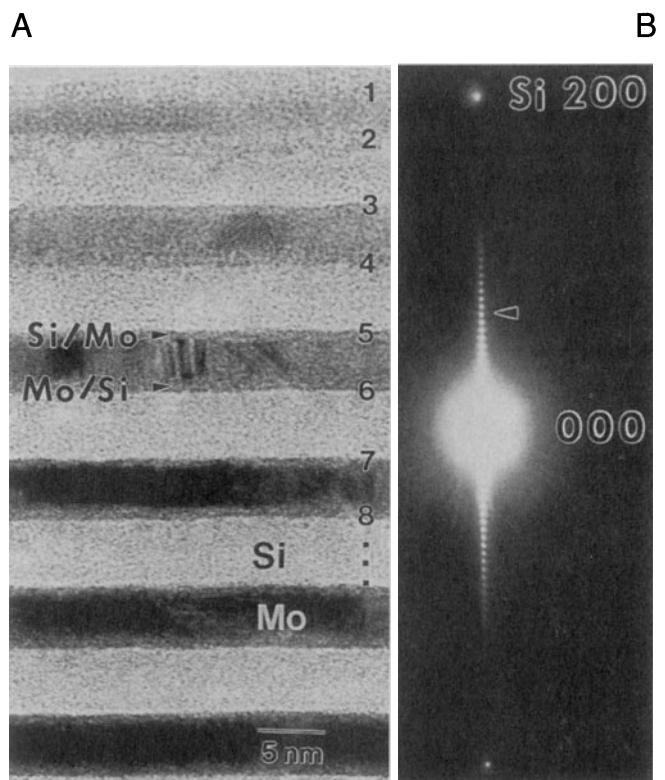
**Figure 16.7.** (A) DF image of ordered  $V_6C_5$  and (B) accompanying diffraction pattern. (C) DF image of  $V_8C_7$  and (D) diffraction pattern. In both carbides the ordering is due to vacancies on the C sublattice.

## 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 ( $\sim 10$  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.



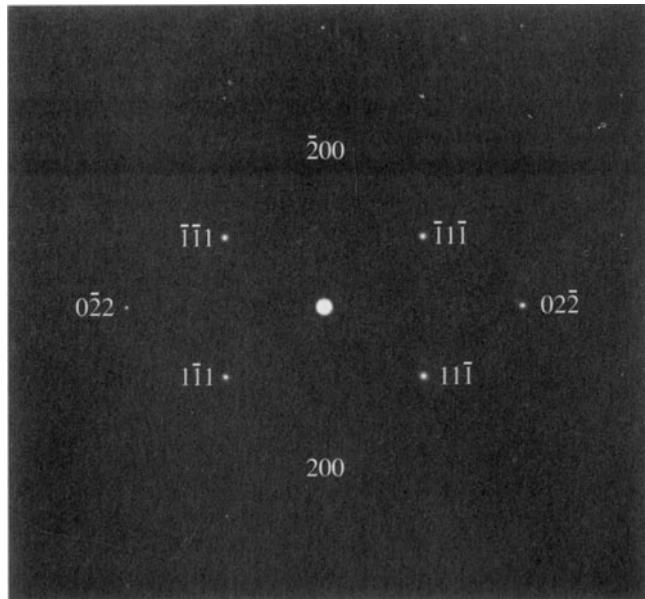
**Figure 16.9.** (A) Artificial superlattice of Si and Mo layers  $\sim 5$  nm thick (B) Expanded DP around  $000$  showing many superlattice spots (arrowed). The large spacing of the superlattice in real space results in very small spacing of the superlattice reflections in the DP in reciprocal space. 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  $\pm 10\%$  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

$$(11\bar{1}) + (\bar{1}\bar{1}1) = (200) \quad [16.24]$$

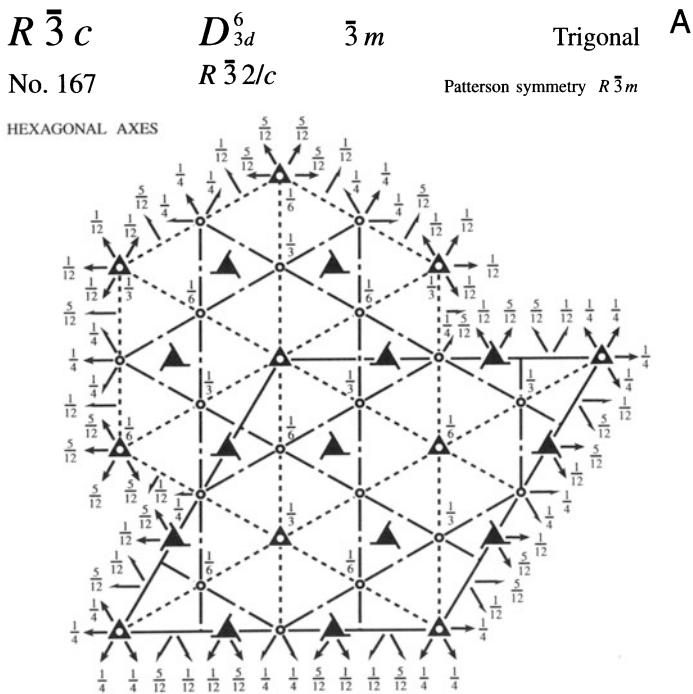


**Figure 16.10.** The [011] diffraction pattern from Si. The 200 reflection is forbidden, but it is present because the allowed  $11\bar{1}$  diffracted beam acts like a new incident beam and is rediffracted by the  $(1\bar{1}1)$  plane. The sum of the two allowed reflections,  $(11\bar{1}) + (\bar{1}\bar{1}1)$ , results in a 200 reflection, which is so weak you may not see it.

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  $\alpha\text{-Al}_2\text{O}_3$ , 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 16.11.** (A) Symmetry information, as given in the International Tables for trigonal  $\alpha\text{-Al}_2\text{O}_3$ , 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)+	( $\frac{2}{3}, \frac{1}{3}, \frac{1}{3}$ )+	( $\frac{1}{3}, \frac{2}{3}, \frac{2}{3}$ )+				
36 <i>f</i> 1	(1) $x, y, z$ (4) $y, x, \bar{z} + \frac{1}{2}$ (7) $\bar{x}, \bar{y}, \bar{z}$ (10) $\bar{y}, \bar{x}, z + \frac{1}{2}$	(2) $\bar{y}, x-y, z$ (5) $x-y, \bar{y}, \bar{z} + \frac{1}{2}$ (8) $y, \bar{x}+y, \bar{z}$ (11) $\bar{x}+y, y, z + \frac{1}{2}$	(3) $\bar{x}+y, \bar{x}, z$ (6) $\bar{x}, \bar{x}+y, \bar{z} + \frac{1}{2}$ (9) $x-y, x, \bar{z}$ (12) $x, x-y, z + \frac{1}{2}$				General: $hkil : -h+k+l = 3n$ $hki0 : -h+k = 3n$ $h\bar{h}2\bar{h}l : l = 3n$ $h\bar{h}0l : h+l = 3n, l = 2n$ $000l : l = 6n$ $h\bar{h}00 : h = 3n$
18 <i>e</i> .2	$x, 0, \frac{1}{4}$	$0, x, \frac{1}{4}$	$\bar{x}, \bar{x}, \frac{1}{4}$	$\bar{x}, 0, \frac{3}{4}$	$0, \bar{x}, \frac{3}{4}$	$x, x, \frac{3}{4}$	Special: as above, plus no extra conditions
18 <i>d</i> $\bar{1}$	$\frac{1}{2}, 0, 0$	$0, \frac{1}{2}, 0$	$\frac{1}{2}, \frac{1}{2}, 0$	$0, \frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, 0, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$	$hkil : l = 2n$
12 <i>c</i> 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 <i>b</i> $\bar{3}$ .	$0, 0, 0$	$0, 0, \frac{1}{2}$					$hkil : l = 2n$
6 <i>a</i> 32	$0, 0, \frac{1}{4}$	$0, 0, \frac{3}{4}$					$hkil : l = 2n$

## Positions

Multiplicity, Wyckoff letter, Site symmetry	Coordinates						Reflection conditions
36 <i>f</i> 1	(1) $x, y, z$ (4) $\bar{y} + \frac{1}{2}, \bar{x} + \frac{1}{2}, \bar{z} + \frac{1}{2}$ (7) $\bar{x}, \bar{y}, \bar{z}$ (10) $y + \frac{1}{2}, x + \frac{1}{2}, z + \frac{1}{2}$	(2) $z, x, y$ (5) $\bar{x} + \frac{1}{2}, \bar{z} + \frac{1}{2}, \bar{y} + \frac{1}{2}$ (8) $\bar{z}, \bar{x}, \bar{y}$ (11) $x + \frac{1}{2}, z + \frac{1}{2}, y + \frac{1}{2}$	(3) $y, z, x$ (6) $\bar{z} + \frac{1}{2}, \bar{y} + \frac{1}{2}, \bar{x} + \frac{1}{2}$ (9) $\bar{y}, \bar{z}, \bar{x}$ (12) $z + \frac{1}{2}, y + \frac{1}{2}, x + \frac{1}{2}$				General: $hh\bar{l} : l = 2n$ $h\bar{h}h : h = 2n$
6 <i>e</i> .2	$x, \bar{x} + \frac{1}{2}, \frac{1}{4}$ $\bar{x}, x + \frac{1}{2}, \frac{3}{4}$	$\frac{1}{4}, x, \bar{x} + \frac{1}{2}$ $\frac{3}{4}, \bar{x}, x + \frac{1}{2}$	$\bar{x} + \frac{1}{2}, \frac{1}{4}, x$ $x + \frac{1}{2}, \frac{3}{4}, \bar{x}$				Special: as above, plus no extra conditions
6 <i>d</i> $\bar{1}$	$\frac{1}{2}, 0, 0$	$0, \frac{1}{2}, 0$	$0, 0, \frac{1}{2}$	$\frac{1}{2}, 0, \frac{1}{2}$	$0, \frac{1}{2}, \frac{1}{2}$	$\frac{1}{2}, \frac{1}{2}, 0$	$hkl : h+k+l = 2n$
4 <i>c</i> 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 <i>b</i> $\bar{3}$ .	$0, 0, 0$	$\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$					$hkl : h+k+l = 2n$
2 <i>a</i> 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 Lagerloef (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 $h, k, \ell$	$f$	1
Body centered cubic	$(h + k + \ell) = 2n$	$2f$	2
Face centered cubic including GaAs and NaCl	$h, k, \text{ and } \ell$ 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 $\ell$ all odd or all even	$2f$	2
Hexagonal close-packed	$h + 2k = 3n$ with $\ell$ odd $h + 2k = 3n$ with $\ell$ even $h + 2k = 3n \pm 1$ with $\ell$ odd $h + 2k = 3n \pm 1$ with $\ell$ even	0 $2f$ $f\sqrt{3}$ $f$	Example 0001 0002 0111 0110

## 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 Lagerloef, K.P.D. (1985) *J. Electron Microsc. Tech.* **2**, 247 (errata in).

# Diffraction from Small Volumes

# 17

---

17.1. Introduction .....	253
17.1.A. The Summation Approach .....	253
17.1.B. The Integration Approach .....	254
17.2. The Thin-Foil Effect .....	255
17.3. Diffraction from Wedge-Shaped Specimens .....	256
17.4. Diffraction from Planar Defects .....	256
17.5. Diffraction from Particles .....	258
17.6. Diffraction from Dislocations, Individually and Collectively .....	259
17.7. Diffraction and the Dispersion Surface .....	261

## 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 ( $\sim 3$  mm) relative to the unit-cell dimensions in the plane of the specimen, but very thin ( $< 0.5 \mu\text{m}$ ) 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  $\mathbf{K}$  is not exactly equal to  $\mathbf{g}$ , or when

$$\mathbf{K} = \mathbf{g} + \mathbf{s} \quad [17.1]$$

Then we can write, from equation 13.48

$$|\phi_{\mathbf{g}}|^2 = \left( \frac{\pi t}{\xi_{\mathbf{g}}} \right)^2 \frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} \quad [17.2]$$

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  $s \neq 0$ . 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

$$A = F \sum_{n_x} e^{i2\pi n_x \mathbf{K} \cdot \mathbf{a}} \sum_{n_y} e^{i2\pi n_y \mathbf{K} \cdot \mathbf{b}} \sum_{n_z} e^{i2\pi n_z \mathbf{K} \cdot \mathbf{c}} \quad [17.3]$$

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  $\mathbf{r}_n$

$$\mathbf{r}_n = n_x \mathbf{a} + n_y \mathbf{b} + n_z \mathbf{c} \quad [17.4]$$

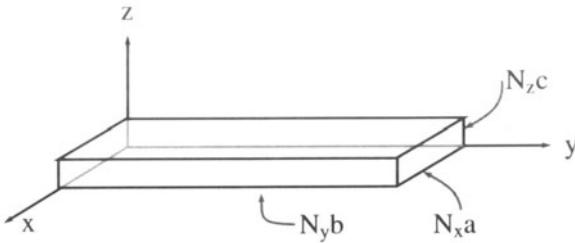
To simplify the first summation we set  $X$  equal to  $e^{i2\pi \mathbf{K} \cdot \mathbf{a}}$ , then each separate summation term is a geometric series, so we can sum the  $n_x$  terms as

$$S = \sum_{n_x=0}^{n_x=N-1} X^n = X^0 + X^1 + \dots + X^{N-1} = \frac{1 - X^N}{1 - X} \quad [17.5]$$

(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

$$\sum_{n_x=0}^{n_x=N-1} e^{i2\pi n_x \mathbf{K} \cdot \mathbf{a}} = \frac{1 - e^{i2\pi n_x \mathbf{K} \cdot \mathbf{a}}}{1 - e^{i2\pi \mathbf{K} \cdot \mathbf{a}}} \quad [17.6]$$



**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

$$\begin{aligned} & (1 - e^{-i\alpha})(1 - e^{i\alpha}) \\ &= (1 - \cos \alpha + i \sin \alpha)(1 - \cos \alpha - i \sin \alpha) \\ &= (1 - 2 \cos \alpha + \cos^2 \alpha) + \sin^2 \alpha \\ &= 2(1 - \cos \alpha) = 4 \sin^2 \frac{\alpha}{2} \end{aligned} \quad [17.7]$$

The intensity is then related to

$$\left| \sum_{n_x=0}^{N_x-1} e^{i2\pi n_x \mathbf{K} \cdot \mathbf{a}} \right|^2 = \frac{1 - \cos (2\pi N_x \mathbf{K} \cdot \mathbf{a})}{1 - \cos (2\pi \mathbf{K} \cdot \mathbf{a})} \quad [17.8]$$

Then we can write

$$\begin{aligned} I &= |A|^2 \\ &= |F|^2 \left( \frac{\sin^2(\pi N_x \mathbf{K} \cdot \mathbf{a})}{\sin^2(\pi \mathbf{K} \cdot \mathbf{a})} \right) \left( \frac{\sin^2(\pi N_y \mathbf{K} \cdot \mathbf{b})}{\sin^2(\pi \mathbf{K} \cdot \mathbf{b})} \right) \left( \frac{\sin^2(\pi N_z \mathbf{K} \cdot \mathbf{c})}{\sin^2(\pi \mathbf{K} \cdot \mathbf{c})} \right) \end{aligned} \quad [17.9]$$

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

$$\pi N_x \mathbf{K} \cdot \mathbf{a} = \frac{\pi}{2} C \quad [17.10]$$

where  $C = \text{an integer}$ . Reordering this equation, we have

$$\mathbf{K} \cdot \mathbf{a} = \frac{C}{2 N_x} \quad [17.11]$$

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

$$\phi_g = \frac{e^{2\pi i \mathbf{k} \cdot \mathbf{r}}}{r} \sum_n F_n e^{(-2\pi i \mathbf{K} \cdot \mathbf{r}_n)} \quad [17.12]$$

Since we have defined  $\mathbf{K}$  to be  $\mathbf{g} + \mathbf{s}$ , we can rewrite this equation as

$$\phi_g = \frac{e^{2\pi i \mathbf{k} \cdot \mathbf{r}}}{r} \sum_n F_g e^{(-2\pi i (\mathbf{g} + \mathbf{s}_g) \cdot \mathbf{r}_n)} \quad [17.13]$$

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  $\mathbf{s}_g$  as  $\mathbf{s}$ . Hence we can write equation 17.13 as

$$\phi_g = \frac{e^{2\pi i \mathbf{k} \cdot \mathbf{r}}}{r} \sum_n F_g e^{(-2\pi i \mathbf{s} \cdot \mathbf{r}_n)} \quad [17.14]$$

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

$$\phi_g = \frac{e^{2\pi i \mathbf{k} \cdot \mathbf{r}}}{r V_c} F_g \int_{\text{crystal}} e^{(-2\pi i \mathbf{s} \cdot \mathbf{r}_n)} dV \quad [17.15]$$

This is where the present treatment differs from the first. If we now express  $\mathbf{s}$  and  $\mathbf{r}_n$  as the vectors

$$\mathbf{s} = u \mathbf{a}^* + v \mathbf{b}^* + w \mathbf{c}^* \quad [17.16]$$

and

$$\mathbf{r}_n = h \mathbf{a} + k \mathbf{b} + \ell \mathbf{c} \quad [17.17]$$

then we can write

$$\phi_g = \frac{e^{2\pi i \mathbf{k} \cdot \mathbf{r}}}{r V_c} F_g \int_0^C \int_0^B \int_0^A e^{(-2\pi i (ux + vy + wz))} dx dy dz \quad [17.18]$$

where  $A = N_x a$ , etc. This integral is straightforward

$$\begin{aligned} \int_0^A e^{-2\pi i ux} &= \frac{e^{-2\pi i uA} - 1}{-2\pi i u} = \left( \frac{e^{-\pi i uA}}{\pi u} \right) \left( \frac{e^{\pi i uA} - e^{-\pi i uA}}{2i} \right) \\ &= \frac{e^{-\pi i uA}}{\pi u} \sin(\pi uA) \end{aligned} \quad [17.19]$$

$$\phi_g = \frac{e^{2\pi i \mathbf{k} \cdot \mathbf{r}}}{r V_c} F_g \frac{(\sin \pi Au)}{(\pi u)} \frac{(\sin \pi Bv)}{(\pi v)} \frac{(\sin \pi Cw)}{(\pi w)} e^{iD} \quad [17.20]$$

( $D$  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

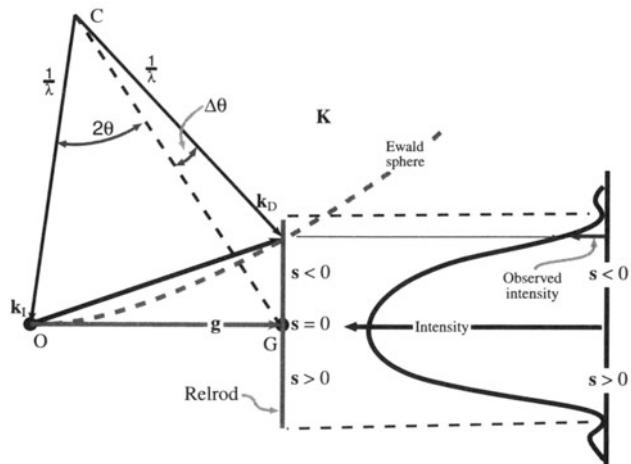
Equation 17.9 is very important for TEM. It tells us why the relrods we introduced in Chapter 12 have a finite length if we measure them to the first minimum. It also tells us that the diffracted intensity does depend on the value of  $s$ ; it is not a constant for any position along the rod.

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.

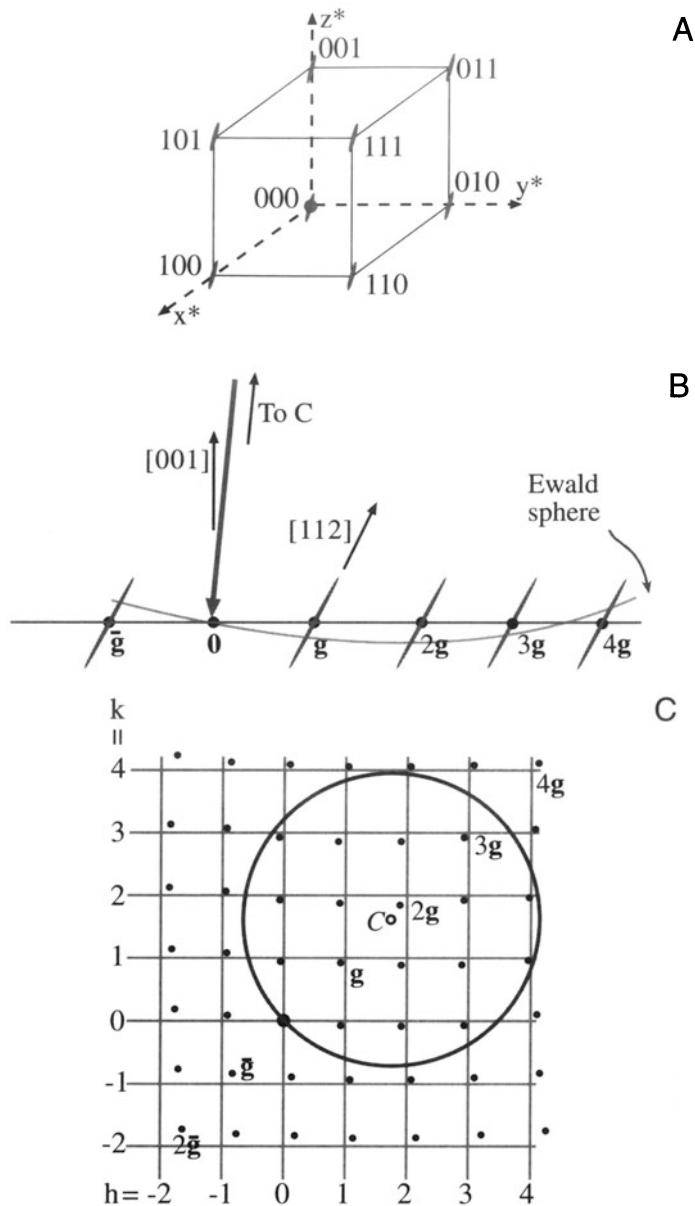
Just remember that when we said intensity in the last sentence, we meant:

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



**Figure 17.2.** The relrod at  $\mathbf{g}_{hkl}$  when the beam is  $\Delta\theta$  away from the exact Bragg condition. The Ewald sphere intercepts the relrod at a negative value of  $s$  which defines the vector  $\mathbf{K} = \mathbf{g} + \mathbf{s}$ . The intensity of the diffracted beam as a function of where the Ewald sphere cuts the relrod is shown on the right of the diagram. In this case the intensity has fallen 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.

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-

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.

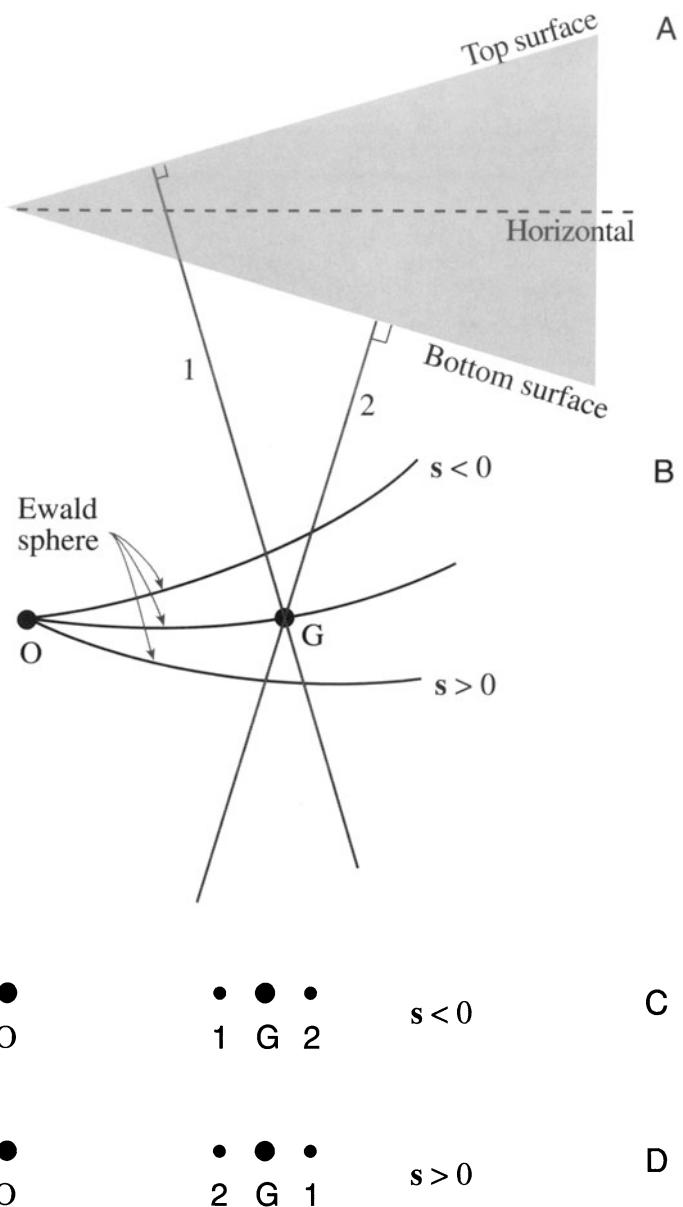
We calculated a "structure factor" for the whole volume which contributes to  $\phi_g$ : we call this calculated factor the shape factor.

We should then use this shape factor rather than the structure factor (since  $F$  is included in equation 17.9) in our dynamical calculations of  $\phi_g$ . 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

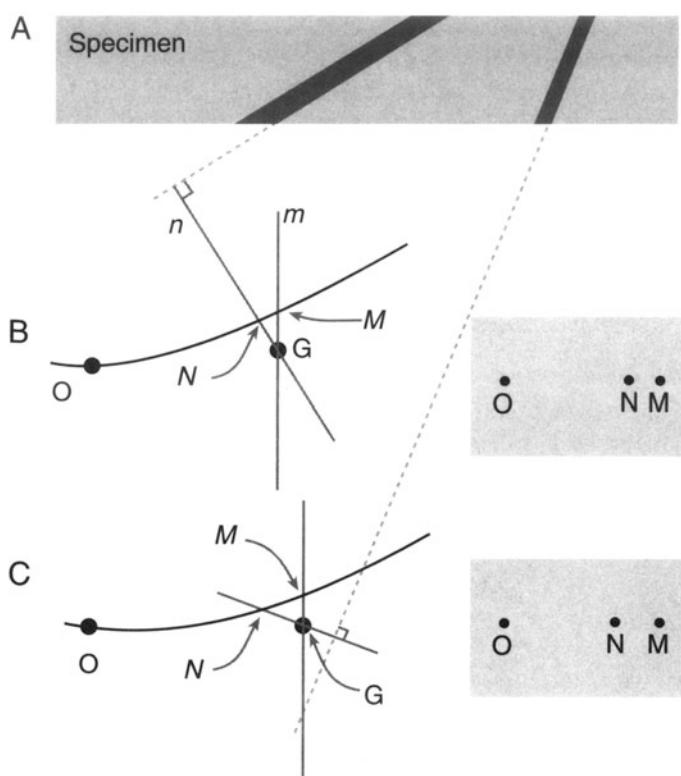
Most TEM specimens do not have parallel surfaces but are wedge-shaped. In drawing the relrods for such a wedge-shaped specimen, we extend the results of Section 17.2 by saying that the relrod will always be normal to the surface. So, for a wedge-shaped specimen (Figure 17.4A) we must have two relrods, 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  $s = 0$ . 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  $s < 0$ , relrod 1 is on the left of relrod 2 but the order reverses when  $s > 0$ . 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

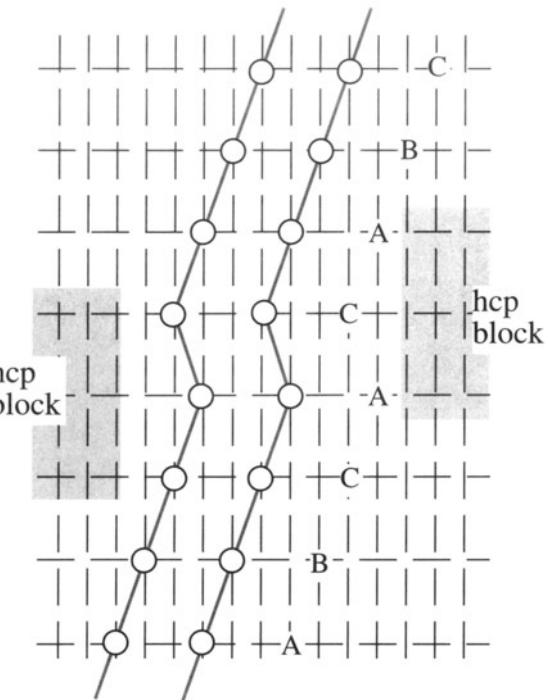


**Figure 17.5** The effect of a thin inclined plate in a thin specimen. (A) Two plates are shown to illustrate the effect of changing the inclination of the plate relative to the foil surface. When  $s \neq 0$  we see two spots in the DP because there are two relrods for the two different planar-defect 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  $90^\circ$ ; 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



**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.

17.6; so it really is a platelet with perfect lattice matching parallel to its surface.

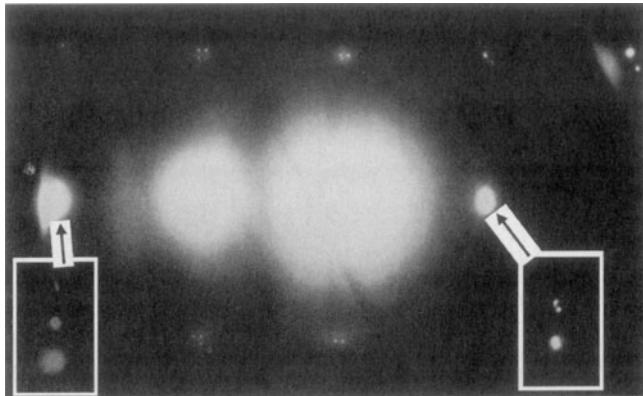
We can understand diffraction effects from other planar interfaces by considering two cases:

- If the grains on either side of the interface contain a common reflection, then the diffraction effects can be modeled by the thin platelet.
- In the case where a reflection is not common to the two grains, then for that reflection the diffracting crystal behaves like a wedge specimen with one surface parallel to the planar defect. We can ignore the crystal that is not diffracting.

The two DPs in Figure 17.7 show that you really do see pairs of spots for these two types of boundary. As 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  $\mathbf{0}$ , 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.

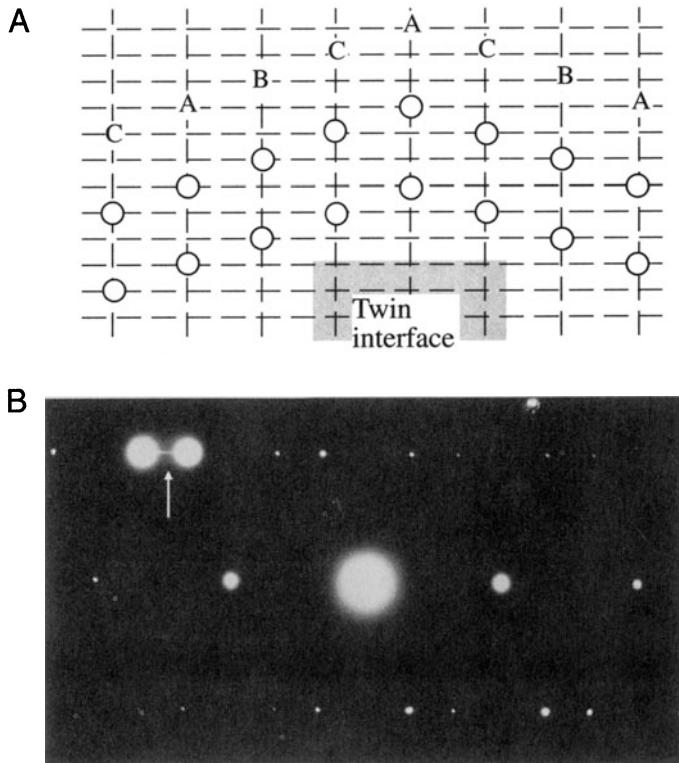
## 17.5. DIFFRACTION FROM PARTICLES

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.

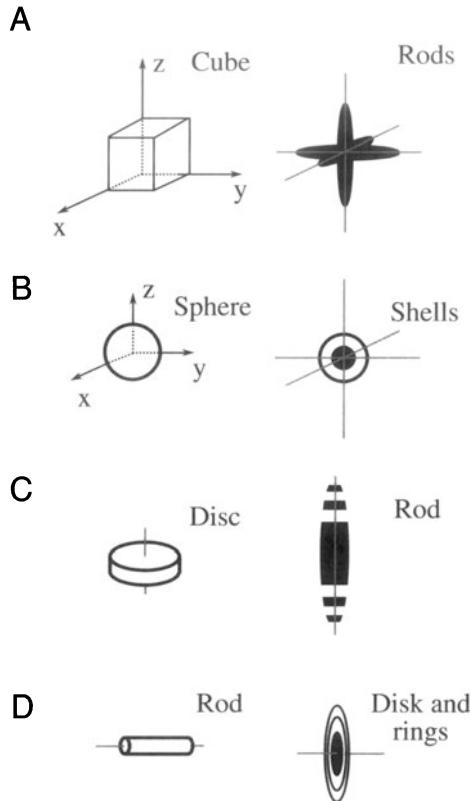
One example, which is common, is the platelets shown in Figure 17.10; these can occur as GP zones or other thin disk-shaped precipitates. When the platelets are oriented parallel to the beam, we see streaks in the DP, just as we saw them in Figure 17.8B. The difference in this figure is that the platelets 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  $\langle 001 \rangle$  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

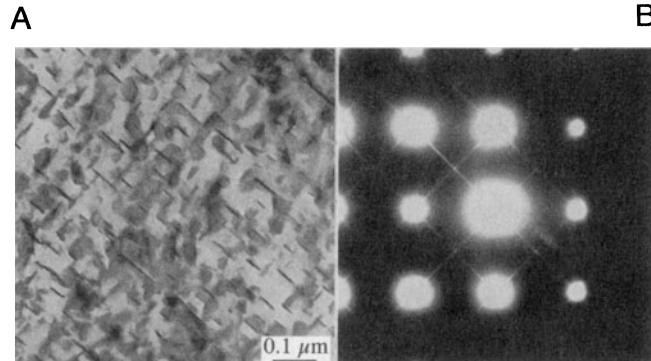


**Figure 17.8.** (A) Schematic of twin and (B) DP with a streak (arrowed) normal to the twin plane. Note that  $s = 0$  for the two bright diffracted spots.

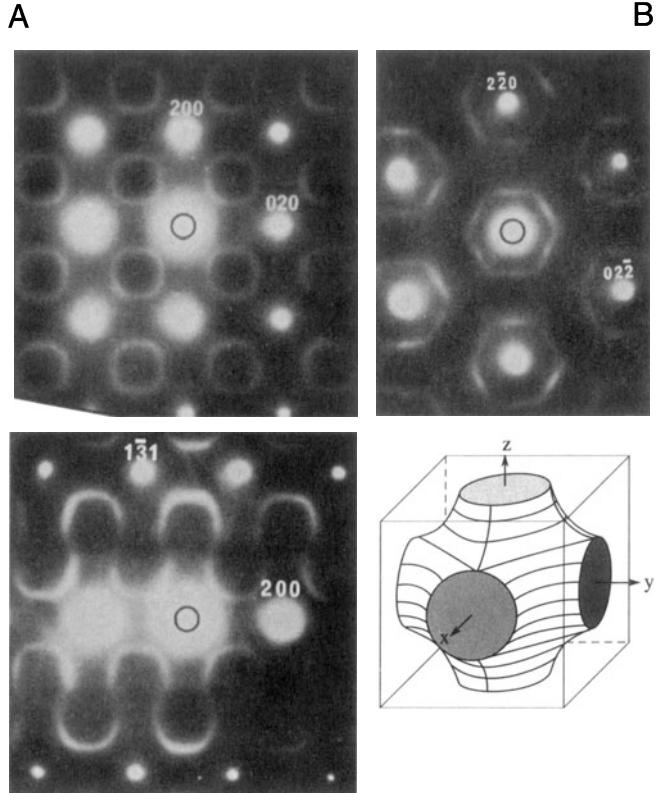


**Figure 17.9.** Examples of how spots in reciprocal space have different shapes, depending on the shape of the particles which are diffracting.

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

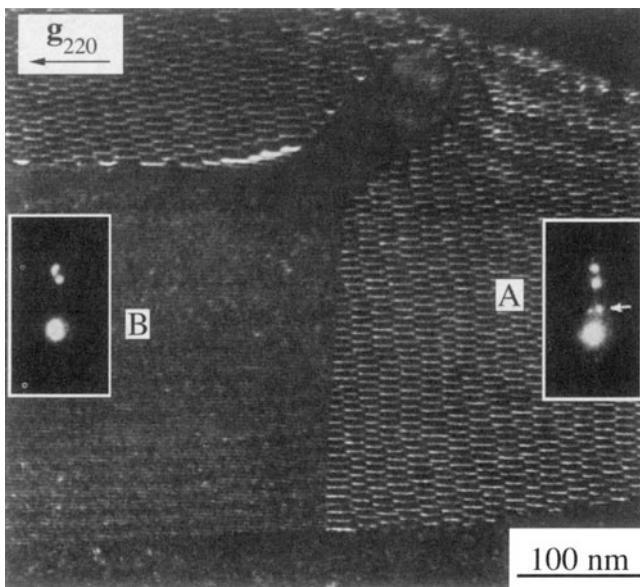
In Chapter 25 we will discuss images of dislocations. A dislocation is a line defect which is characterized by its line direction and its Burgers vector. The crystal around the defect is distorted or strained.

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<sup>-1</sup> to perhaps 0.5 Å<sup>-1</sup> 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

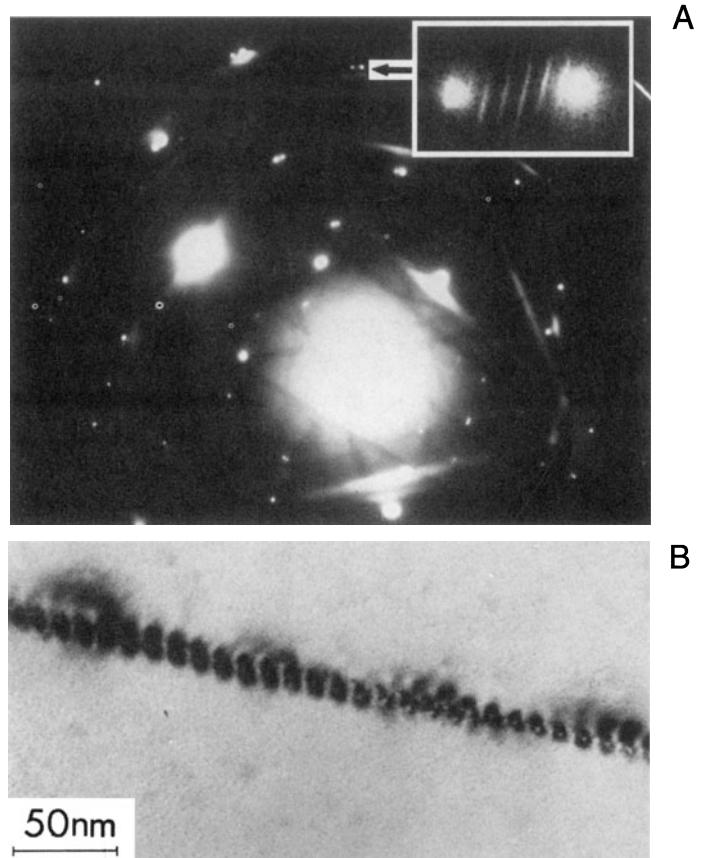


**Figure 17.12.** Diffraction from an ordered array of dislocations. Dislocations are present in region A, but not in B. The insets show a small part of the DP from the two regions. The extra spots arrowed in A are caused by the visible array of dislocations; these spots are a doublet because there is 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.

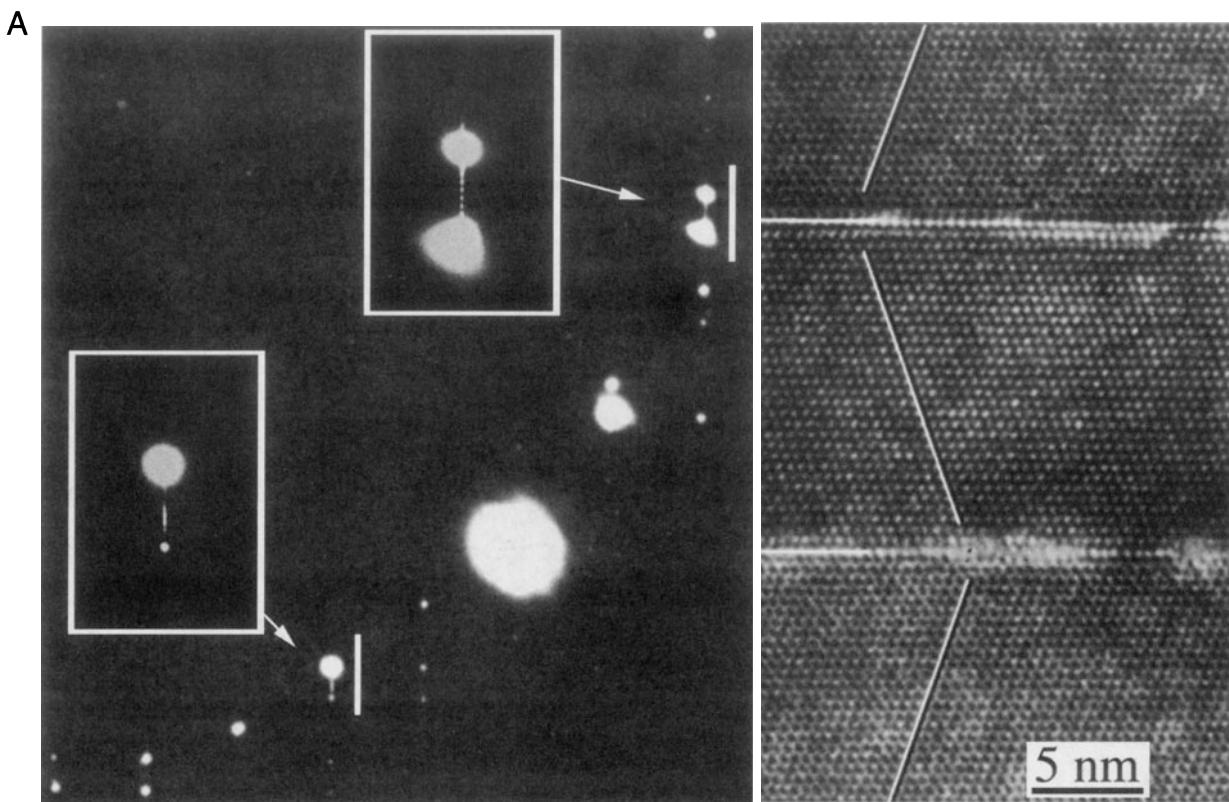
from an ordered array of dislocations, as shown in Figure 17.12. The specimen used to form this image was rather special. Dislocations are present in region A, but not in region B. The array actually forms a structured grain boundary in A, but a layer of glass is present in B. The insets show the same 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<sub>8</sub>C<sub>7</sub> 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 17.13.** (A) The set of streaks from an array of dislocations in Al<sub>2</sub>O<sub>3</sub> 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.)

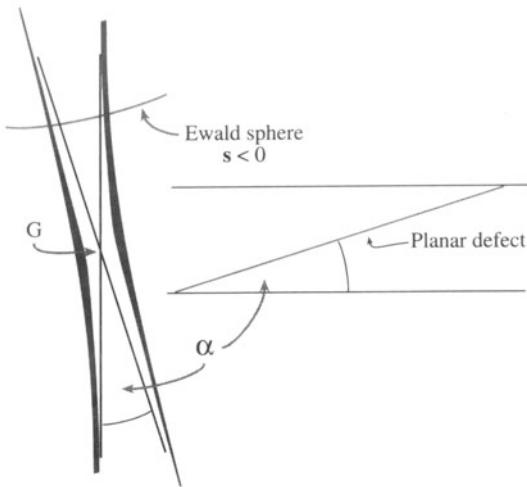
The general rule is:

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  $0.067 \text{ nm}^{-1}$ , 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  $s = 0$ ,” 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  $\alpha$  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  $\xi_g^{-1}$ .

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 remark-

ably 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  $\xi_g$ , the extinction distance for reflection  $g$ .

You can understand why this is so in the following pictorial way. What you see in the image will be determined by the DP. What you see in the DP is determined by which relrods, or surfaces, the Ewald sphere intersects. All the information about extinction distances and coupling of diffracted beams is fundamentally contained in the dispersion surface ( $\xi_g$  is just  $\Delta k^{-1}$  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  $90^\circ$ . 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

18.1. Choosing Your Technique .....	267
18.2. Experimental Techniques .....	267
18.3. The Stereographic Projection .....	269
18.4. Indexing Single-Crystal Diffraction Patterns .....	271
18.5. Ring Patterns from Polycrystalline Materials .....	273
18.6. Ring Patterns from Amorphous Materials .....	274
18.7. Double Diffraction .....	278
18.8. Orientation of the Specimen .....	280
18.9. Orientation Relationships .....	280
18.10. Computer Analysis .....	285

## 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  $hkl$  reflections are characteristic of the crystal system. Indexing associates each spot in the DP with a plane,  $(hkl)$ , in the crystal. From the indexing of the spots, you can deduce the orientation of the crystal in terms of the zone axis  $[UVW]$  in which the indexed planes lie. 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,  $\leq 10$  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  $\sim 100$  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  $\lambda$  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  $\sim 2$   $\mu\text{m}$ . 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  $\Delta f$ , 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  $> 2$   $\mu\text{m}$ . 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 ( $> 2$   $\mu\text{m}$ , even better if they are  $> 5$   $\mu\text{m}$ ) 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.

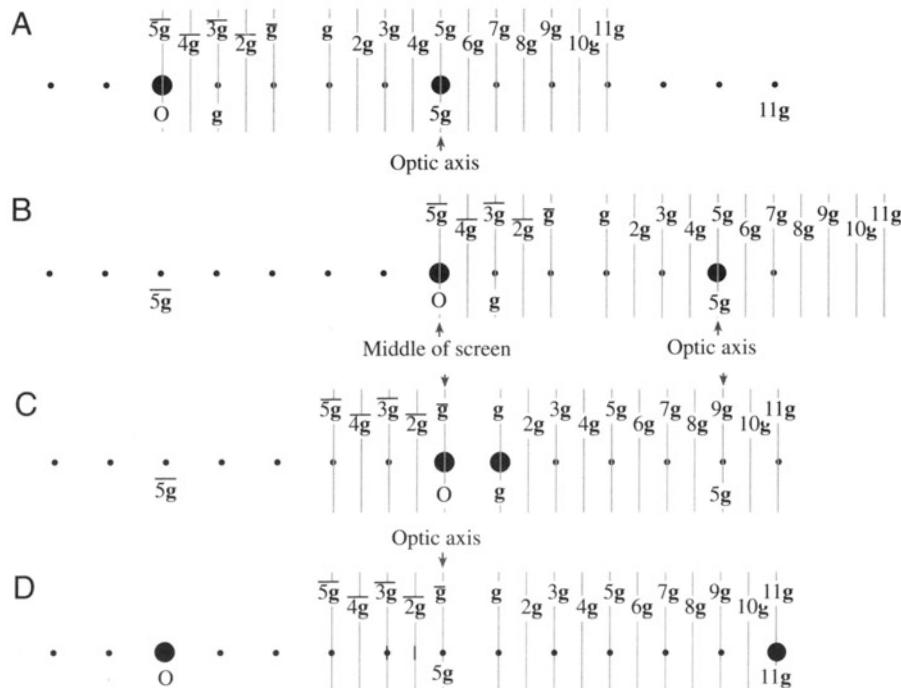
*Tilting and rotating the specimen.* One of the biggest assets of the TEM is that you can monitor the DP as you tilt or rotate your specimen. Rotating the pattern requires a rotation holder, which is ideal if you want to align a particular reflection parallel to the tilt axis, especially for a side-entry 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  $11g$  but you can't see  $11g$ , use the following approach, shown in Figure 18.1:

- Tilt the beam using the dark-field deflection coils so that  $5g$  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  $5g$  is back on the optic axis;  $11g$  should now be excited (Figure 18.1D).

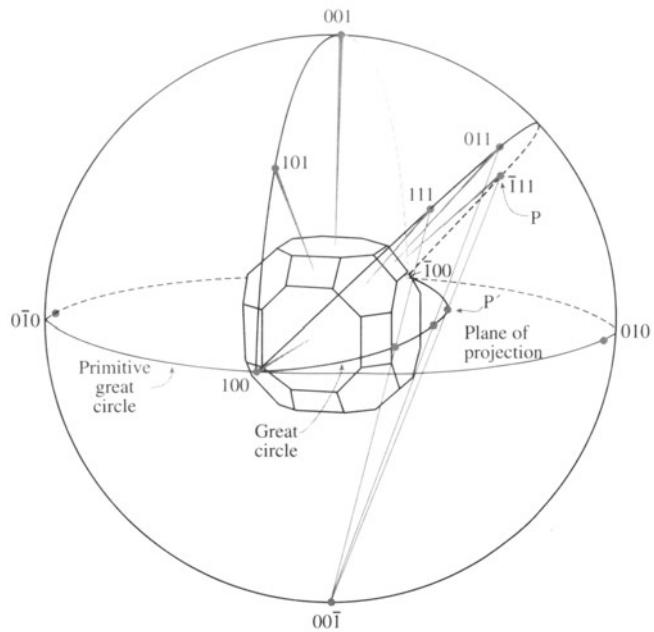
We'll develop other variations of this technique in Chapter 19, and we'll see in Chapter 26 that the situation in Figure 18.1 does arise in weak-beam microscopy at higher voltages.

### 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

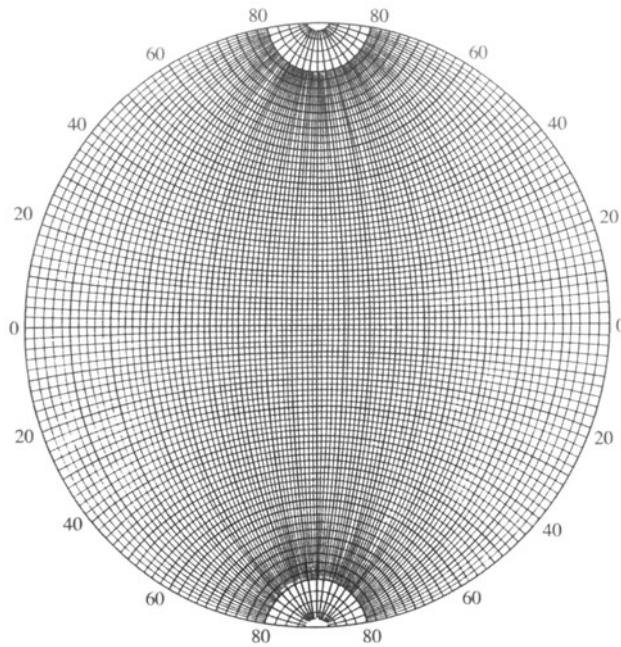


**Figure 18.2.** The stereographic projection. The crystal is at the center of the sphere. Normals to the crystal planes are projected until they intercept the sphere at P, then projected back to the south pole ( $00\bar{1}$ ) of the sphere. Where this projected line crosses the equatorial plane at  $P'$  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.

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  $(\bar{1}\bar{1}1)$  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  $2^\circ$  apart so the net covers  $180^\circ$ . 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  $90^\circ$  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  $(hkl)$  is parallel to the direction  $[hkl]$  only for cubic materials.

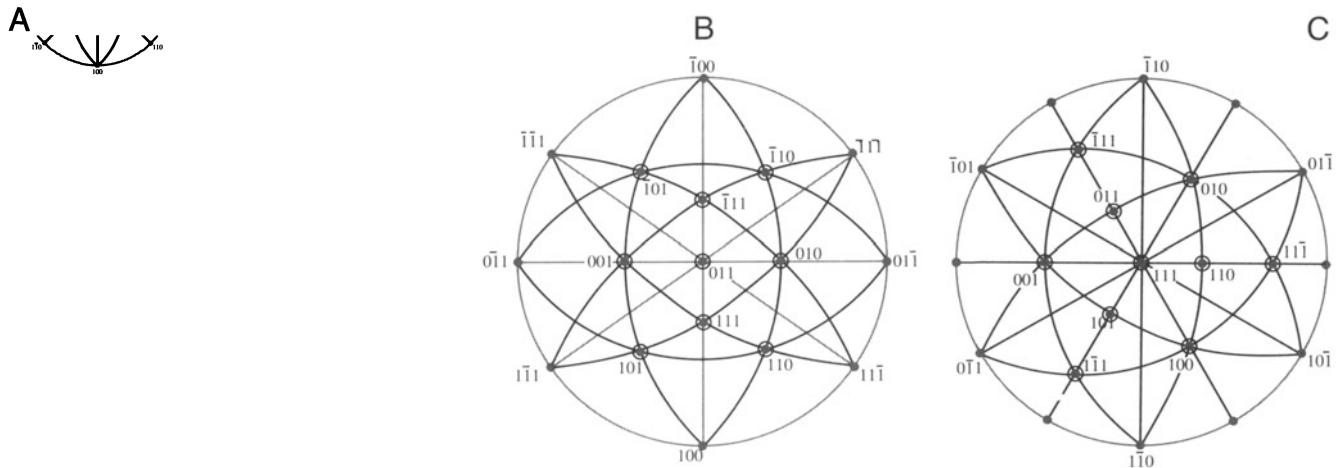
- The zone axis is always  $90^\circ$  away from any plane normal in its zone. All the plane normals in a particular zone,  $[UVW]$ , will lie on a single

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  $hkl$  reflections will be around the circumference of the projection (the primitive great circle).

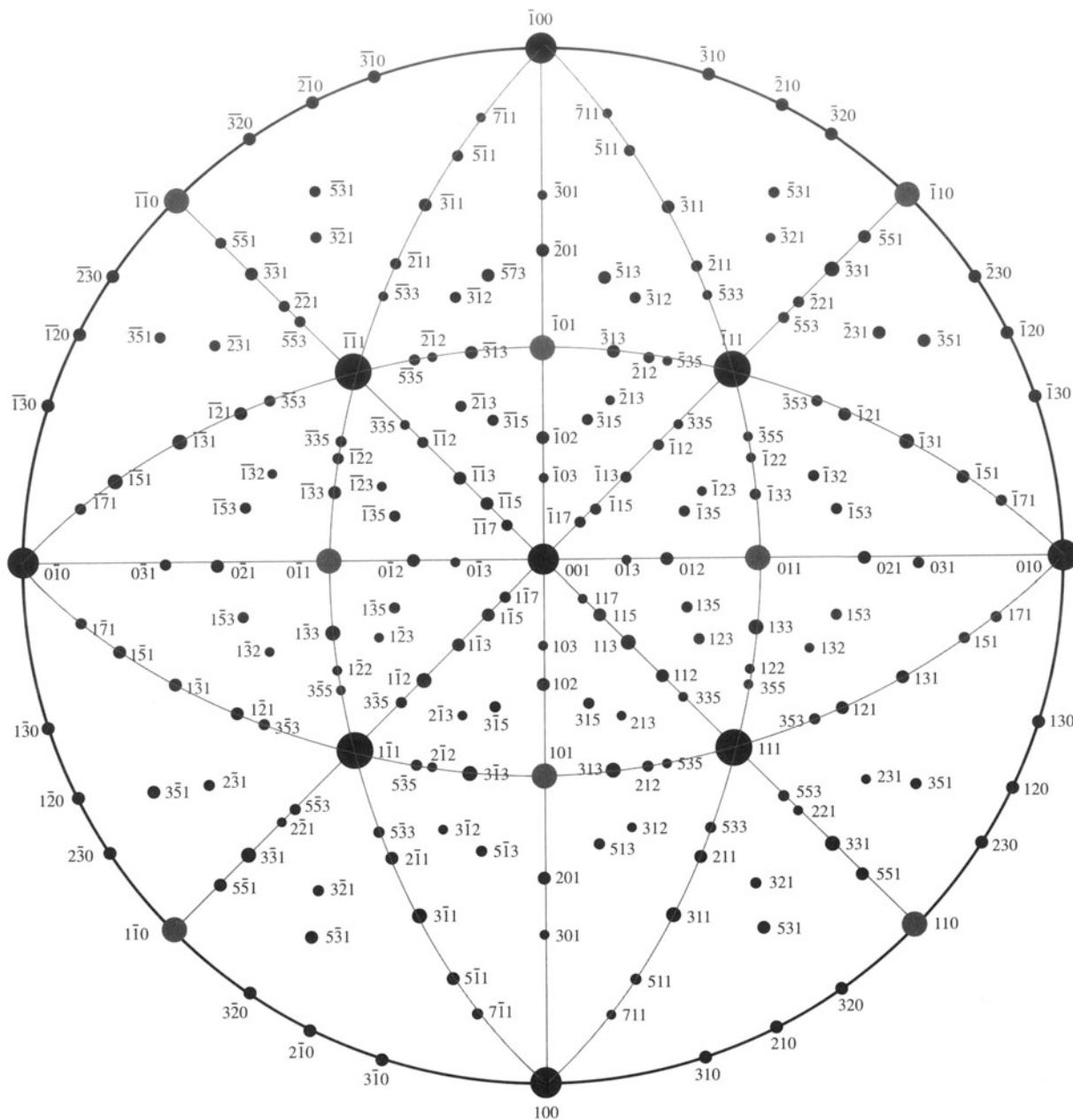
- The angle between any two planes is the angle between their plane normals, measured along a great circle using the Wulff net.
  - We can use the same construction to summarize all the symmetry elements of any particular crystal system.

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  $90^\circ$  away from the  $\bar{[001]}$  direction. How large is the angle between  $(0\bar{1}1)$  and  $(011)$ ? How would this angle change if the material were forced to be tetragonal with  $c/a > 1$ ? What happens in this case to the  $(1\bar{1}1)$  pole or the  $(1\bar{1}0)$  pole? Now consider the more extensive plot shown in Figure 18.5. If the specimen is cubic with the  $[001]$  foil normal, what pole would you tilt to if you wanted to form an image with the  $0\bar{2}2$  reflection? (One answer is the  $[011]$  zone axis, but why?) For the same specimen, if you want to excite the  $\bar{1}11$  reflection, 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  $\bar{0}\bar{2}2$  reflection, you need to tilt the specimen so the  $\bar{0}\bar{1}1$  pole rotates until it is on the primitive, i.e., it is  $90^\circ$  from the beam direction. To do this you need to tilt about an axis that is  $90^\circ$  from the  $\bar{0}\bar{2}2$  reflection, such as the [100], [111], [311], zone axes.

plot the points, or download appropriate free-  
ware from the WWW.

$$Rd = \lambda L \quad [18.1]$$

Any distance,  $R$ , which we measure on the DP is related to a specific spacing in the crystal,  $d$ . Since  $\lambda L$  is a constant, we can measure several values of  $R$  and know that

$$R_1 d_1 = R_2 d_2 = R_3 d_3 = R_4 d_4 = \dots \quad [18.2]$$

## 18.4. INDEXING SINGLE-CRYSTAL DIFFRACTION PATTERNS

Remember the fundamental relationship in a DP (refer back to Chapter 9)

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	
$h^2 + k^2 + \ell^2$	$hkl$	$h^2 + k^2 + \ell^2$	$hkl$	$h^2 + k^2 + \ell^2$	$hkl$
2	110		3	111	3
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  $hkl$  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  $\phi$ ; the angle between directions  $[U_1V_1W_1]$  and  $[U_2V_2W_2]$  is  $\rho$ . 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

$$\cos \phi = \frac{h_1h_2 + k_1k_2 + \ell_1\ell_2}{(h_1^2 + k_1^2 + \ell_1^2)^{\frac{1}{2}}(h_2^2 + k_2^2 + \ell_2^2)^{\frac{1}{2}}} \quad [18.3]$$

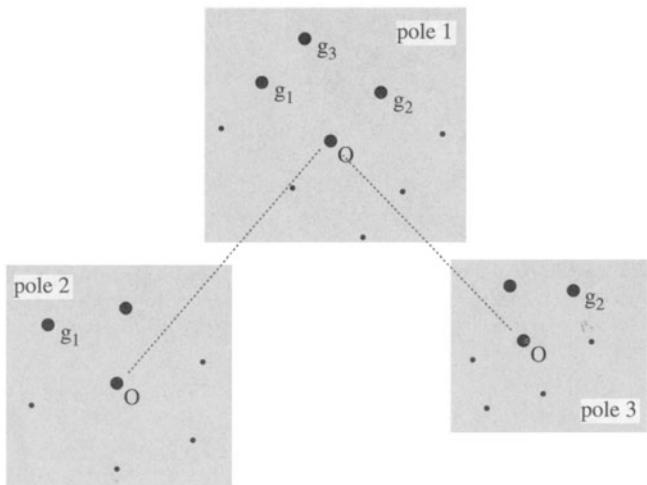
$$\cos \rho = \frac{U_1U_2 + V_1V_2 + W_1W_2}{(U_1^2 + V_1^2 + W_1^2)^{\frac{1}{2}}(U_2^2 + V_2^2 + W_2^2)^{\frac{1}{2}}} \quad [18.4]$$

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.

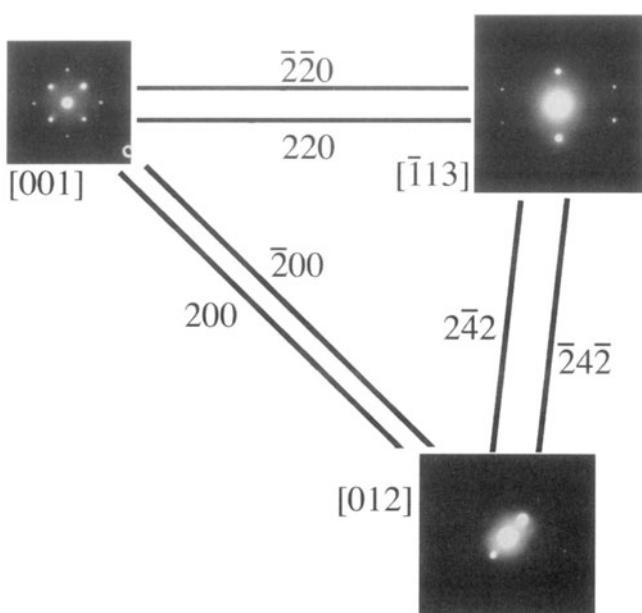
In principle, if we don't know the crystal structure, we can still work out the  $d$ -spacings of the diffracting planes using equation 18.1. However, you should remember that SAD is not the most accurate method for determining the spacing of lattice planes,  $d_{hkl}$ , or the angles between them,  $\phi$ . SAD is generally very good at distinguishing patterns, but it completely fails when the difference between the two patterns is a  $180^\circ$  rotation as occurs in some patterns of polar material, like GaAs.

To summarize, tilt your specimen to a low-index pole, set  $s = 0$  for the innermost reflections, and record the SAD pattern. Repeat the exercise using higher-order reflections after tilting the specimen to set  $s = 0$ . These measurements will be more accurate, but only if you make sure that  $s = 0$ . The discussion on relrods in Chapter 16 told you that both  $d$  and  $\phi$  could be seriously in error if reflections are not set to have  $s = 0$ , 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 recip-

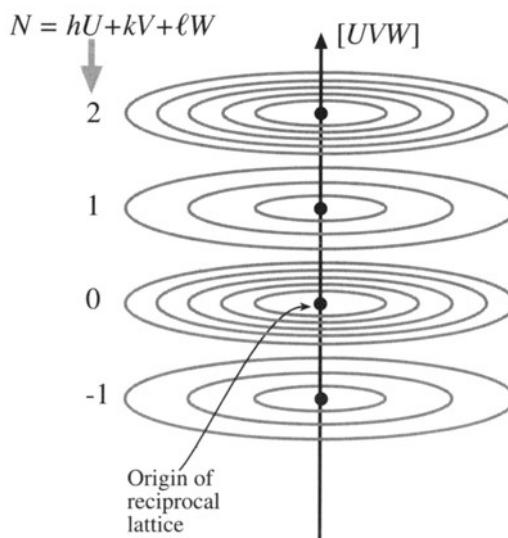
rocal lattice about all axes and produce a set of nested spheres. When we intersect these spheres with the Ewald sphere (which, in the TEM, approximates to a plane) we will see the rings that are recorded in powder patterns.

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  $\mathbf{g}_{hkl}$  will be parallel to the direction  $[hkl]$  in real space. Otherwise, this will not generally be the case.

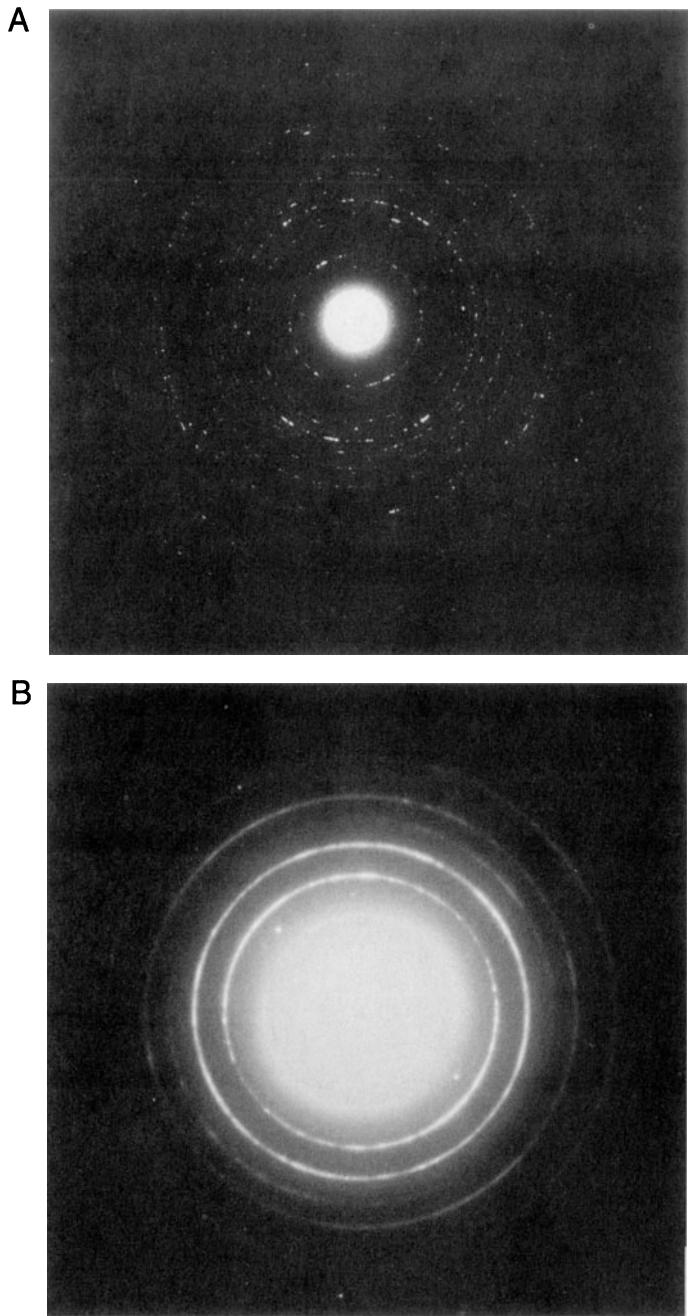
Since the grains are small, all the reciprocal lattice points will be broadened by the shape effect; so will the sphere or circles for the polycrystals.

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 ( $N = 1, 2, \text{ etc.}$ ) produces a set of concentric circles for each allowed reflection in each zone.



**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  $\alpha\text{-Ag}_2\text{Se}$ , 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  $\ell$ , 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  $d_{hkl}$  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.

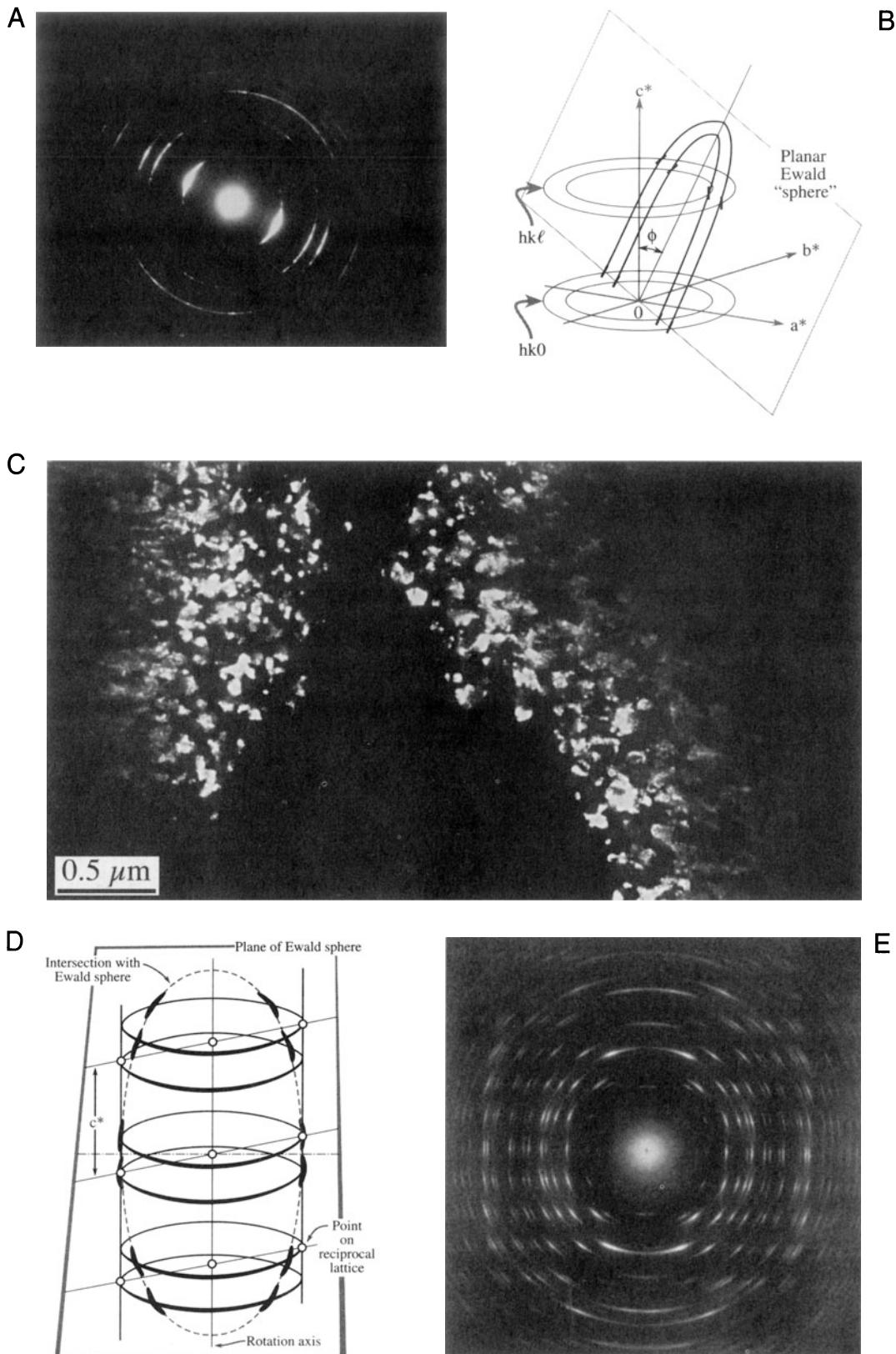
## 18.6. RING PATTERNS FROM AMORPHOUS MATERIALS

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  $\leq 15\text{-}\text{\AA}$  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).



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

$$I(\mathbf{k}) = |f(\mathbf{k})|^2 \sum_{i,j} e^{i2\pi\mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \quad [18.5]$$

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

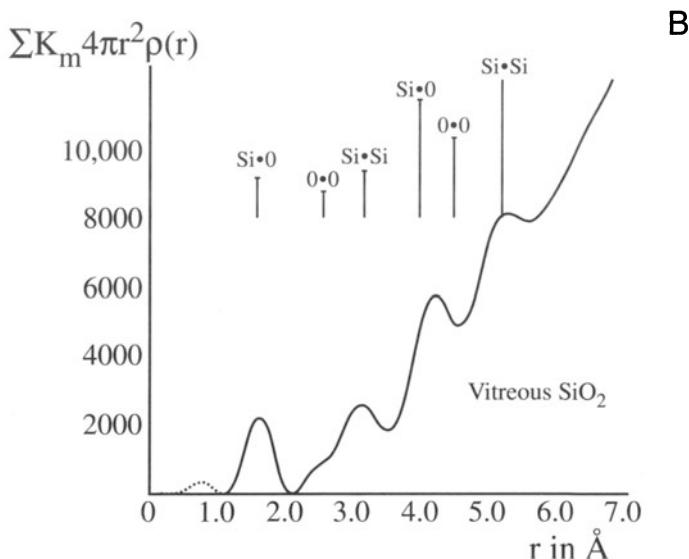
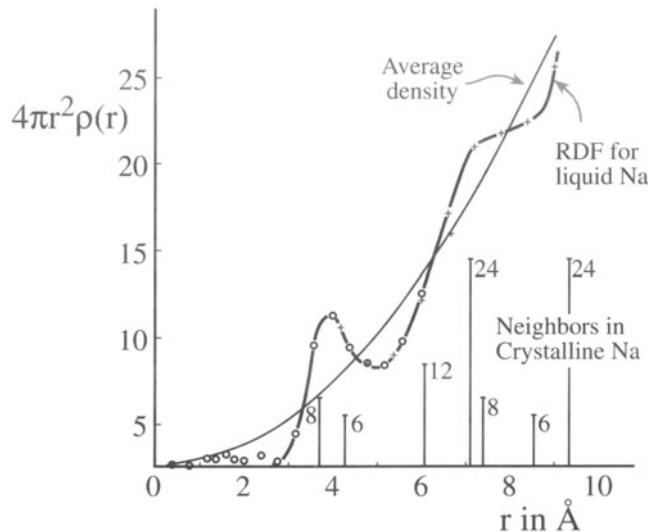
$$I(\mathbf{k}) = N |f(\mathbf{k})|^2 \left( 1 + \frac{F(\mathbf{k})}{k} \right) \quad [18.6]$$

where

$$F(\mathbf{k}) = \sum_{i \neq j} e^{i2\pi\mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \quad [18.7]$$

$$F(\mathbf{k}) = k \int \rho(r) e^{i2\pi\mathbf{k} \cdot \mathbf{r}} dV \quad [18.8]$$

$$F(\mathbf{k}) = 4\pi \int_0^\infty \rho(r) \sin(2\pi kr) r dr \quad [18.9]$$



**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  $\text{SiO}_2$  is peaked at distances that represent various spacings between atoms.

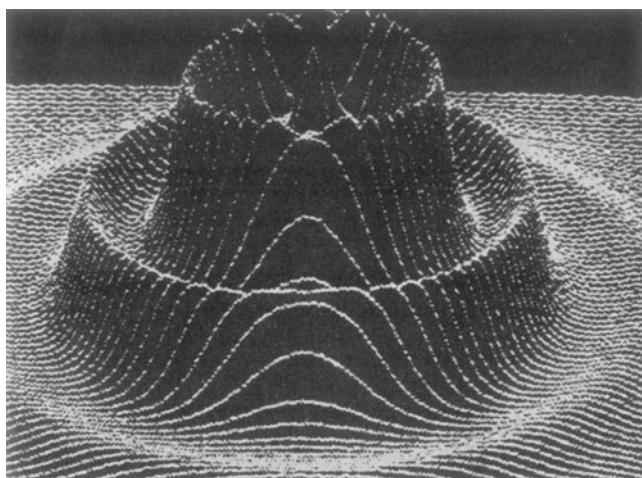
The term  $\rho(r)$  is the RDF. Equation 18.9 can be inverted to give an expression for  $\rho$

$$\rho(r) - \rho_0 = \frac{1}{r} \int_0^\infty F(\mathbf{k}) \sin 2\pi kr dk \quad [18.10]$$

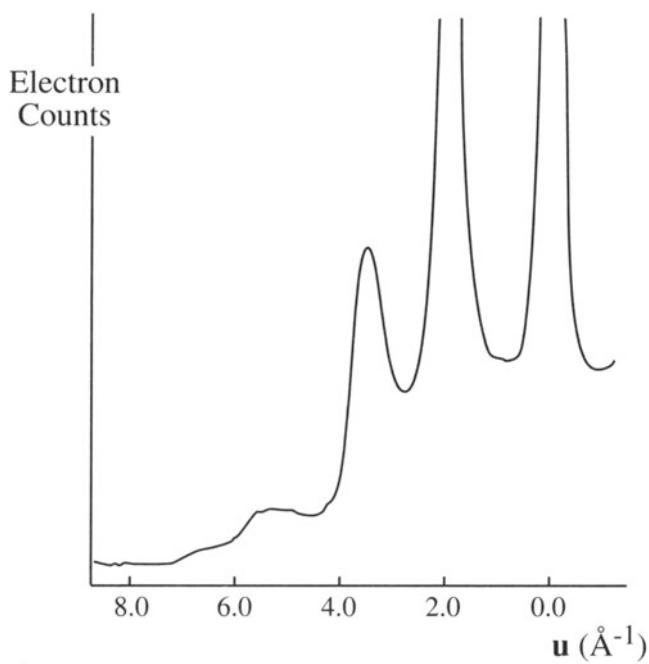
The RDF can be obtained directly from DPs, and this process is enhanced if the patterns are energy-filtered to remove inelastic contributions (see Chapter 40 and the work of Cockayne *et al.* 1991). Software to determine the RDF

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



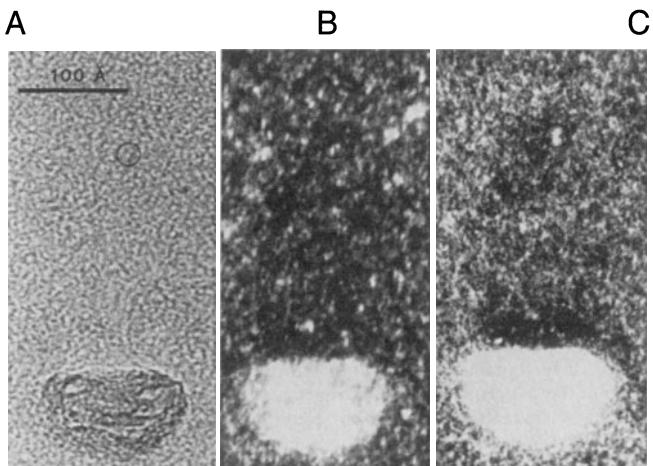
**Figure 18.13.** A computer plot of the diffracted-intensity distribution from an amorphous structure, showing diffuse rings of intensity. The direct-beam intensity is off scale.



**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.

you may well want to study small particles (e.g., catalysts) supported by an amorphous film. In such a case, you need to know what the image of the support film looks like before you add a new component.

*Glass 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 18.14.** (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

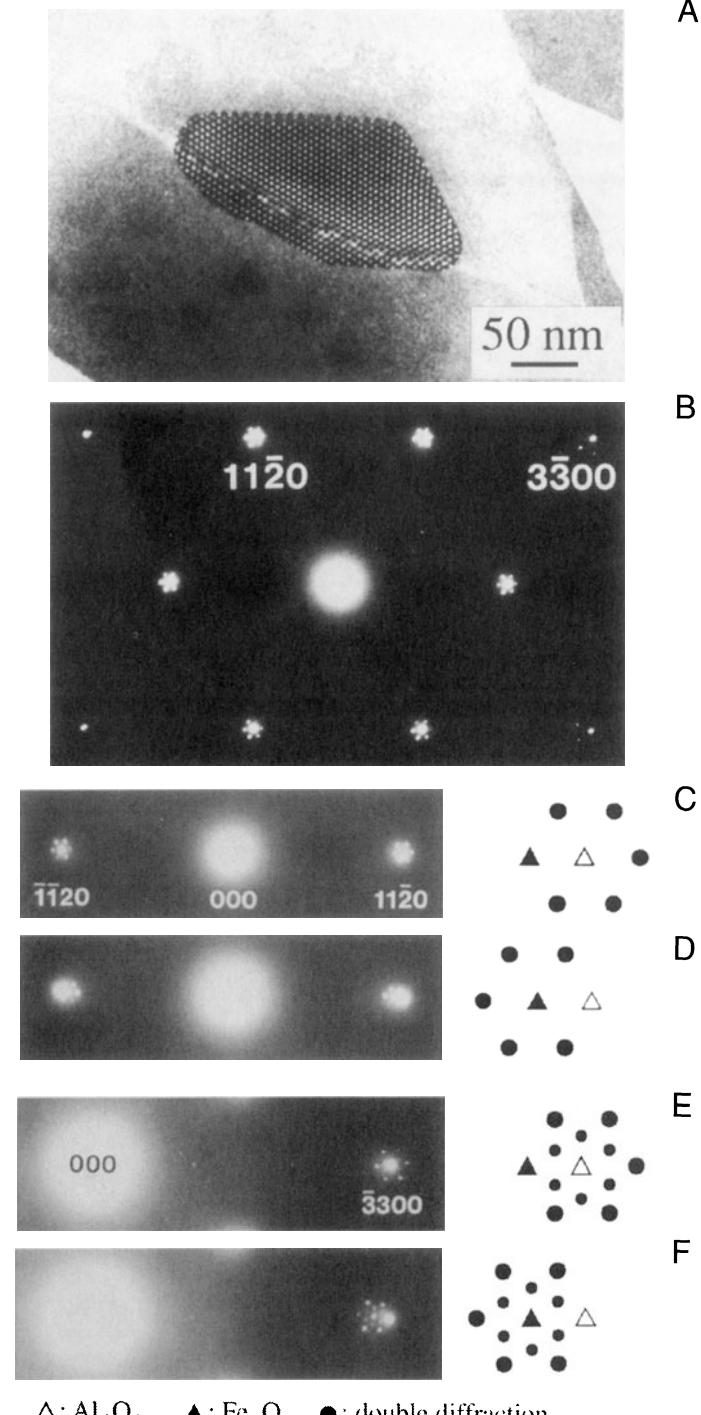
Double diffraction occurs when a diffracted beam traveling through a crystal is rediffracted either within the same crystal or when it passes into a second crystal. If the initial diffraction vector of the beam is  $\mathbf{g}_1$  and it is rediffracted by reflection  $\bar{\mathbf{g}}_2$ , then the resultant diffraction vector of the double-diffracted beam is  $(\mathbf{g}_1 - \bar{\mathbf{g}}_2)$ . If  $\bar{\mathbf{g}}_2$  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  $\alpha\text{-Fe}_2\text{O}_3$  (hematite) islands grown on a single-crystal  $\alpha\text{-Al}_2\text{O}_3$  (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  $\alpha\text{-Fe}_2\text{O}_3$  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  $\text{Al}_2\text{O}_3$  reflection while in (D) the double-diffraction spots surround the  $\text{Fe}_2\text{O}_3$  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  $\alpha\text{-Fe}_2\text{O}_3$  on  $\alpha\text{-Al}_2\text{O}_3$ . (B) [0001] SAD pattern from  $\alpha\text{-Fe}_2\text{O}_3$  showing double-diffraction spots around the  $\{11\bar{2}0\}$  and  $\{\bar{3}300\}$  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  $\text{Al}_2\text{O}_3$ , 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:

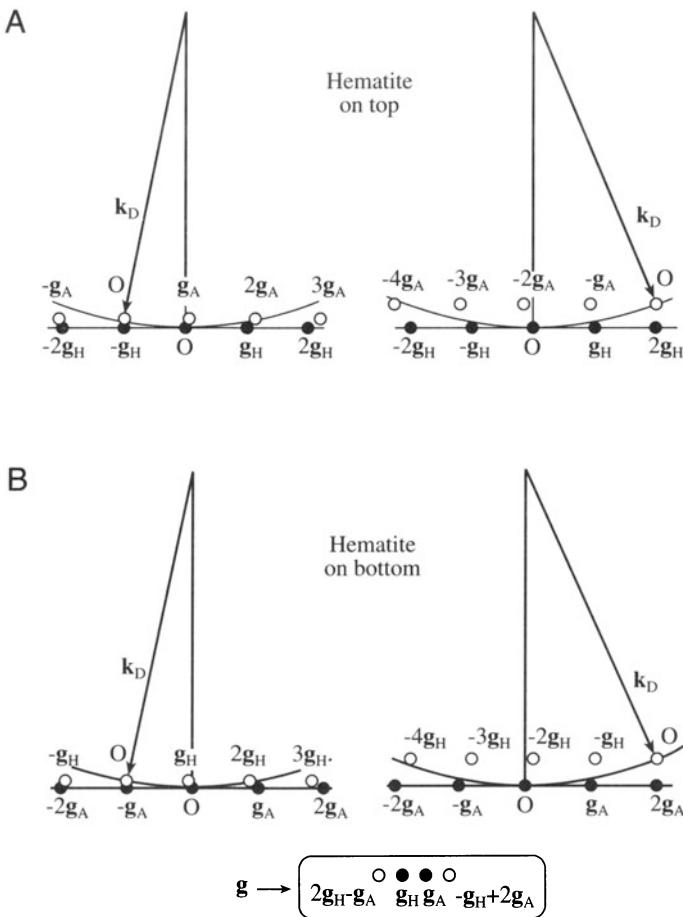
- Draw the reciprocal lattice with the origin of the Ewald sphere at  $2\mathbf{g}_H$  for the first case and on  $\bar{\mathbf{g}}_H$  in the second.
- Keep the radius of the Ewald sphere unchanged since only elastic interactions are considered.
- The incident beam for the lower crystal is in the  $2\mathbf{g}$  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.

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



**Figure 18.16.** Top-bottom effect in double diffraction. The pattern depends on which of the two crystals is on top. In (A)  $\alpha\text{-Fe}_2\text{O}_3$  particles are on top of the  $\text{Al}_2\text{O}_3$ ; in (B)  $\text{Fe}_2\text{O}_3$  particles are below the  $\text{Al}_2\text{O}_3$ . Two non-equivalent paths for double diffraction are shown.

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  $\mathbf{g}$  vectors  $\mathbf{g}_1$ ,  $\mathbf{g}_2$ , and  $\mathbf{g}_3$  in a single-crystal DP, you can calculate the direction of the beam  $\mathbf{B}$ . You can actually estimate  $\mathbf{B}$  to within about  $10^\circ$  from the vector cross product as follows

$$\mathbf{B} = \mathbf{g}_1 \times \mathbf{g}_2 = \begin{bmatrix} \mathbf{i}_1 & \mathbf{i}_2 & \mathbf{i}_3 \\ h_1 & k_1 & \ell_1 \\ h_2 & k_2 & \ell_2 \end{bmatrix} \quad [18.11]$$

$$= (k_1\ell_2 - k_2\ell_1, \ell_1h_2 - \ell_2h_1, h_1k_2 - h_2k_1) \quad [18.12]$$

For the three-beam case, you can determine  $\mathbf{B}$  with an accuracy of  $\sim 3^\circ$ . You first need to make sure that the three vectors are taken in the correct order. Draw a circle through these three reflections: if O is inside the circle, then the  $\mathbf{g}$  vectors should be numbered counterclockwise; if O is outside, number them clockwise. Check your labeling; the determinant of the matrix in equation 18.13 should be positive

$$\mathbf{g}_1 \cdot (\mathbf{g}_2 \times \mathbf{g}_3) = \frac{1}{V} \begin{bmatrix} h_1 & k_1 & \ell_1 \\ h_2 & k_2 & \ell_2 \\ h_3 & k_3 & \ell_3 \end{bmatrix} \quad [18.13]$$

Now we can write a weighted-average expression for  $\mathbf{B}$

$$\mathbf{B} = \frac{\mathbf{g}_2 \times \mathbf{g}_3}{|\mathbf{g}_1|^2} + \frac{\mathbf{g}_3 \times \mathbf{g}_1}{|\mathbf{g}_2|^2} + \frac{\mathbf{g}_1 \times \mathbf{g}_2}{|\mathbf{g}_3|^2} \quad [18.14]$$

*A convention:* The vector  $\mathbf{B}$  points up the column. It is normal to the emulsion side of a photographic negative. The electron beam travels along the direction  $-\mathbf{B}$ .

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

$$(12\bar{1}) = (110) + (01\bar{1}) \quad [18.15]$$

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  $\rightarrow$  fcc reciprocal space.
- fcc real space  $\rightarrow$  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 = ( $\bar{3}\bar{3}1$ ); check that the angles are correct using

$$\cos(\phi_{AB}) = \frac{\mathbf{g}_A \cdot \mathbf{g}_B}{|\mathbf{g}_A| |\mathbf{g}_B|} \quad [18.16]$$

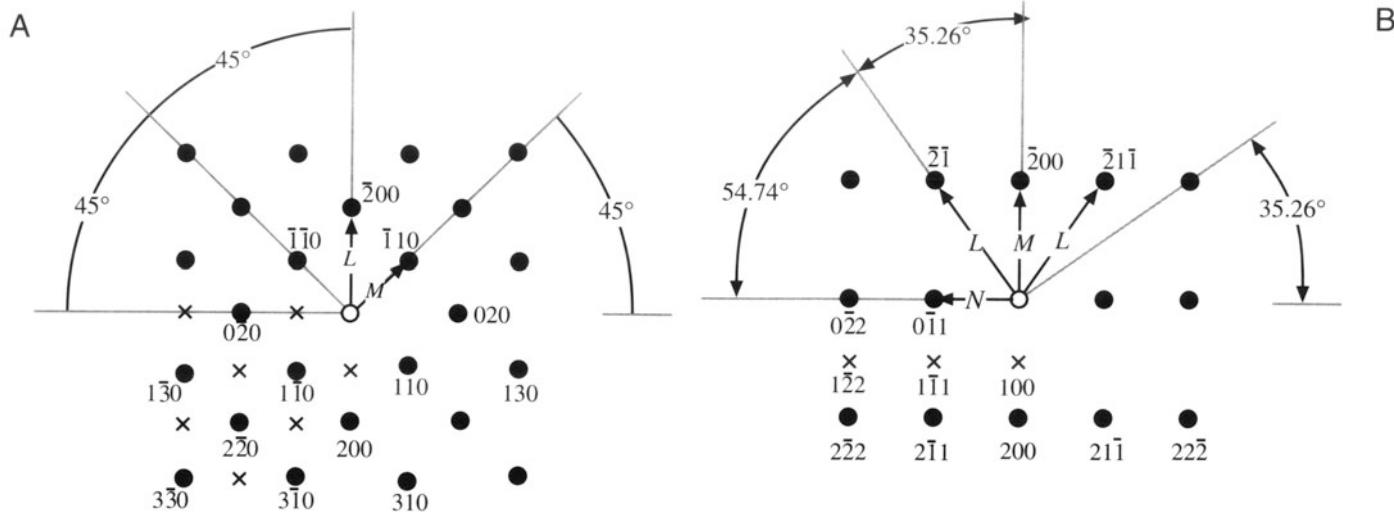
and so on for  $\phi_{BC}$  and  $\phi_{CA}$ . You should immediately recognize that this is the  $\pm[12\bar{3}]$  pole, but continue. Now you can plug these indices into equation 18.11 or 18.14 to show that  $\mathbf{B} = [\bar{1}\bar{2}3]$ .

Finally, use the [001] stereographic projection. Draw a great circle that passes through the (111), (2 $\bar{1}0$ ), and ( $\bar{3}\bar{3}1$ ) 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  $90^\circ$  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  $\mathbf{B}$  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).

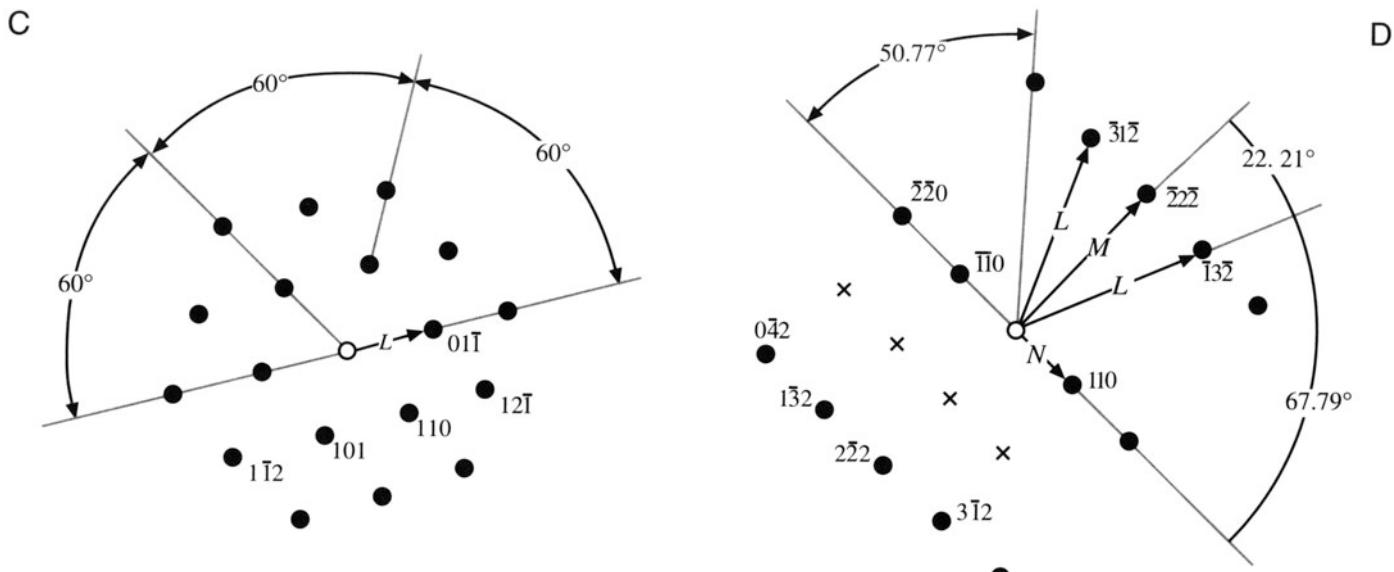
## 18.9. ORIENTATION RELATIONSHIPS

Once you've learned how to index a DP and determine  $\mathbf{B}$ , 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



$$\frac{L}{M} = \frac{\sqrt{4}}{\sqrt{2}} = 1.414 \quad \mathbf{B} = [001]$$

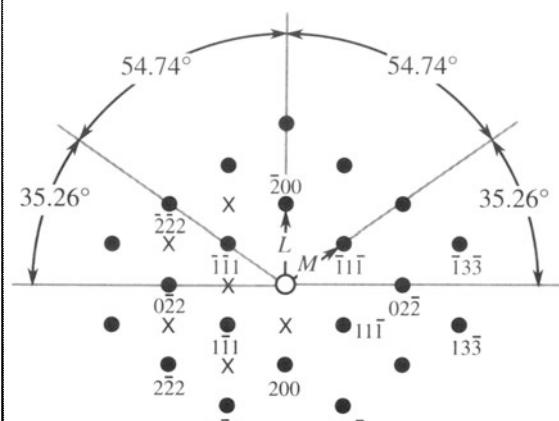
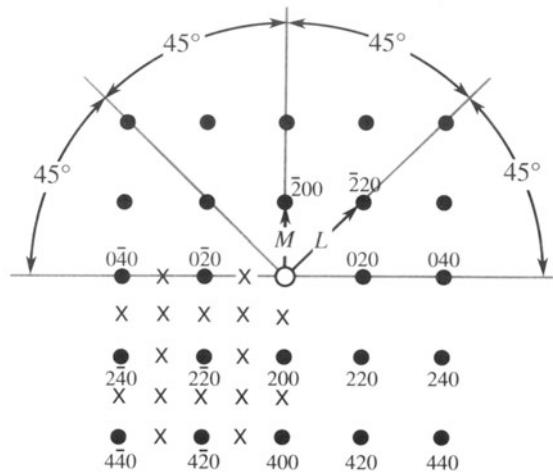
$$\frac{L}{N} = \frac{\sqrt{6}}{\sqrt{2}} = 1.732 \quad \frac{M}{N} = \frac{\sqrt{4}}{\sqrt{2}} = 1.414 \quad \mathbf{B} = [011]$$



$$\mathbf{B} = [\bar{1}11]$$

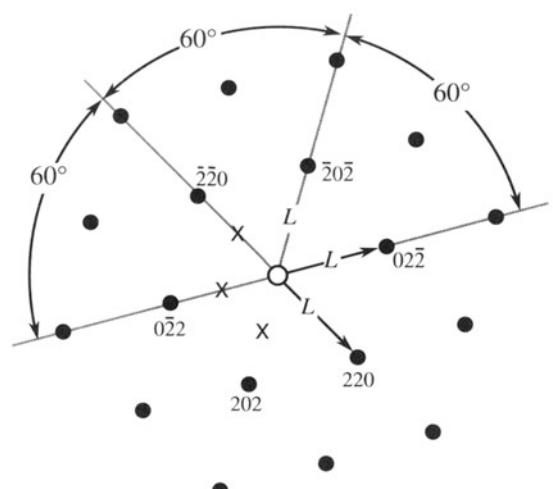
$$\frac{L}{N} = \frac{\sqrt{14}}{\sqrt{2}} = 2.646 \quad \frac{M}{N} = \frac{\sqrt{12}}{\sqrt{2}} = 2.450 \quad \mathbf{B} = [\bar{1}12]$$

**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.

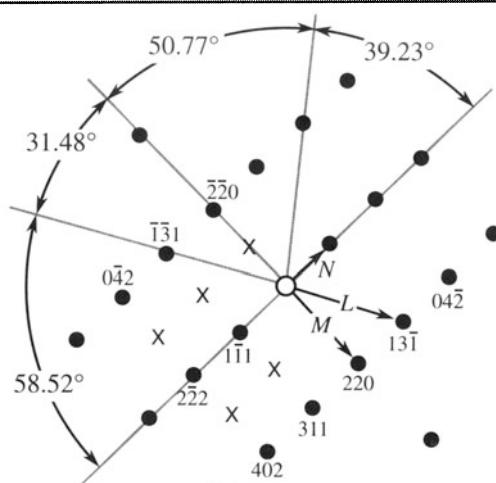


$$\frac{L}{M} = \frac{2}{\sqrt{3}} = 1.155 \quad \mathbf{B} = [011]$$

C

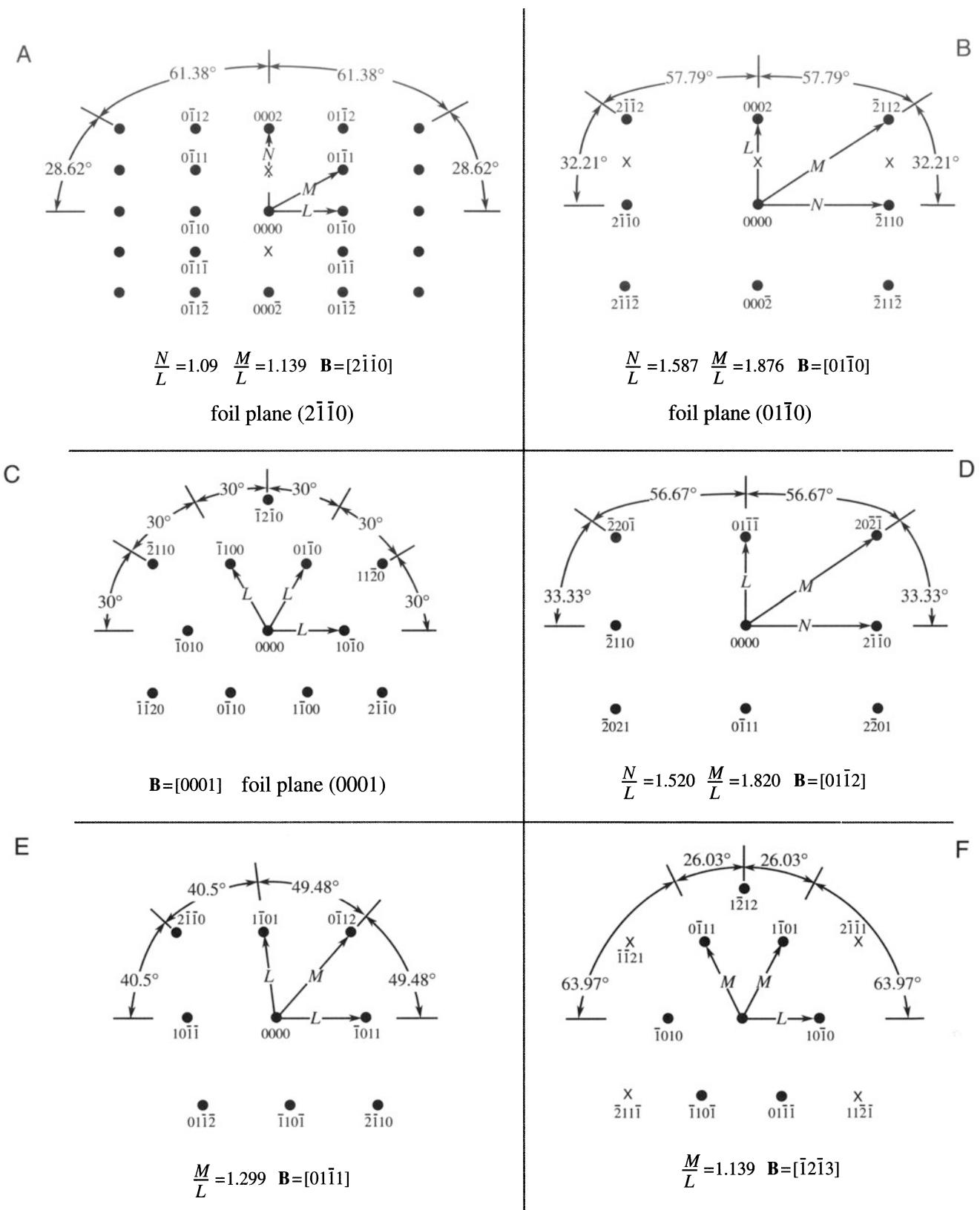


$$\mathbf{B} = [\bar{1} \ 1 \ 1]$$

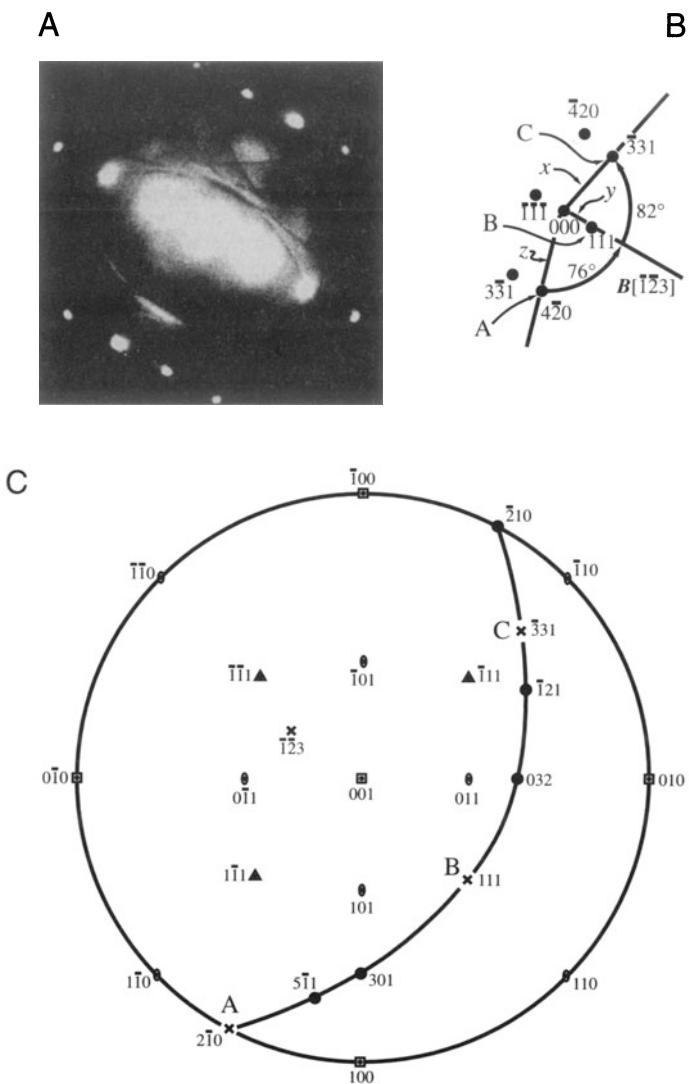


$$\frac{M}{N} = \frac{\sqrt{8}}{\sqrt{3}} = 1.633 \quad \frac{L}{N} = \frac{\sqrt{11}}{\sqrt{3}} = 1.915 \quad \mathbf{B} = [1\bar{1}\bar{2}]$$

**Figure 18.18.** Four standard indexed diffraction patterns for fcc crystals in the [001], [011], [ $\bar{1}11$ ], and [ $\bar{1}\bar{1}2$ ] 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  $[1\bar{2}\bar{1}3]$  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 ( $\beta$ - $\alpha$ ) orientation relationships where the crystal systems may be different.
  - The two crystals have a common direction (axis) so that one crystal can be rotated through some angle 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  $\alpha$  and  $\beta$ :

- Tilt to zone-axis pattern (ZAP) 1 in phase  $\alpha$ , the matrix phase. Record and index it to determine  $\mathbf{B}_1(\alpha)$ .
  - Translate the precipitate,  $\beta$ , 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  $\alpha$  and  $\beta$  on a stereogram and construct the poles of the important planes that are normal to each  $\mathbf{B}$ . 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  $\mathbf{B}$ . 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<sub>3</sub>Li

$(\delta')$  in an Al-Li ( $\alpha$ ) 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  $\langle UVW \rangle$  directions or  $\{hkl\}$  planes in the cubic system would be parallel. It's just convention to choose the lowest-index planes or directions to define the OR. When the lowest-index planes and directions align, the surface energy between the phases tends to be lowest, so this configuration is thermodynamically favored.

- The Kurdjumov-Sachs OR is often found relating fcc and bcc crystalline grains. The close-packed planes (or closest packed in bcc) and close-packed directions are parallel, but these are not now identical.

$(111)_{\text{fcc}}$  is parallel to  $(011)_{\text{bcc}}$  (the closest-packed planes),

$[10\bar{1}]_{\text{fcc}}$  is parallel to  $[11\bar{1}]_{\text{bcc}}$  (the close-packed directions),

$(\bar{1}2\bar{1})_{\text{fcc}}$  is parallel to  $(\bar{2}\bar{1}\bar{1})_{\text{bcc}}$ .

- The Nishiyama-Wassermann OR is related to the Kurdjumov-Sachs OR:

$[011]_{\text{fcc}}$  is parallel to  $[001]_{\text{bcc}}$ ,

$(111)_{\text{fcc}}$  is parallel to  $(\bar{1}\bar{1}0)_{\text{bcc}}$  (the closest-packed planes),

$(211)_{\text{fcc}}$  is parallel to  $(110)_{\text{bcc}}$ .

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:

$(111)_{\text{fcc}}$  is parallel to  $(0001)_{\text{hcp}}$  (the closest-packed planes),

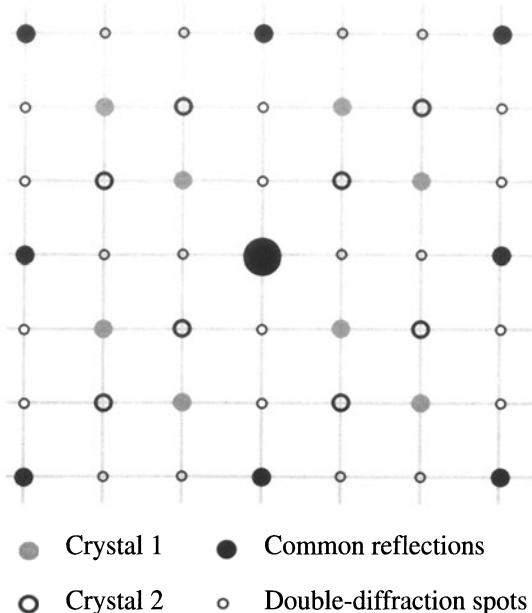
$[1\bar{1}0]_{\text{fcc}}$  is parallel to  $[12\bar{1}0]_{\text{hcp}}$  (the close-packed directions).

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



**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.

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.

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  $\mathbf{g}$  should give you the best value for both  $d$  and  $\phi$ , but be absolutely sure that  $\mathbf{s} = \mathbf{0}$  for your chosen  $\mathbf{g}$ .
- 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  $180^\circ$  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.
- 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.

## 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).

# Kikuchi Diffraction

# 19

---

19.1. The Origin of Kikuchi Lines .....	291
19.2. Kikuchi Lines and Bragg Scattering .....	291
19.3. Constructing Kikuchi Maps .....	293
19.4. Crystal Orientation and Kikuchi Maps .....	295
19.5. Setting the Value of $s_g$ .....	296
19.6. Intensities .....	297

## 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  $\lambda$  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

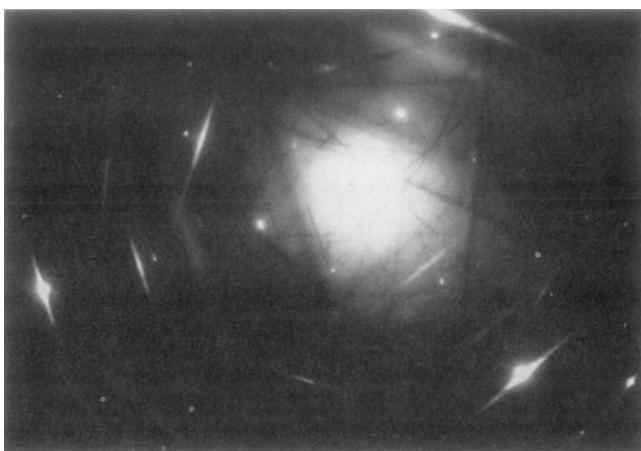
thick enough, but if it is too thick then there will be no lines because inelastic scattering then dominates and there is no subsequent Bragg diffraction of these electrons.

## 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  $\theta_B$  to the  $hkl$  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  $\mathbf{k}$ -vectors rather than a single  $\mathbf{k}$ -vector. Construct the cones by considering all the vectors oriented at angle  $\theta_B$  to the  $hkl$  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  $\theta_B$  and the other to  $-\theta_B$ ; one is the  $\mathbf{g}$  Kikuchi line and the other the  $\bar{\mathbf{g}}$  Kikuchi line. Neither of them is the  $\mathbf{0}$  Kikuchi line.

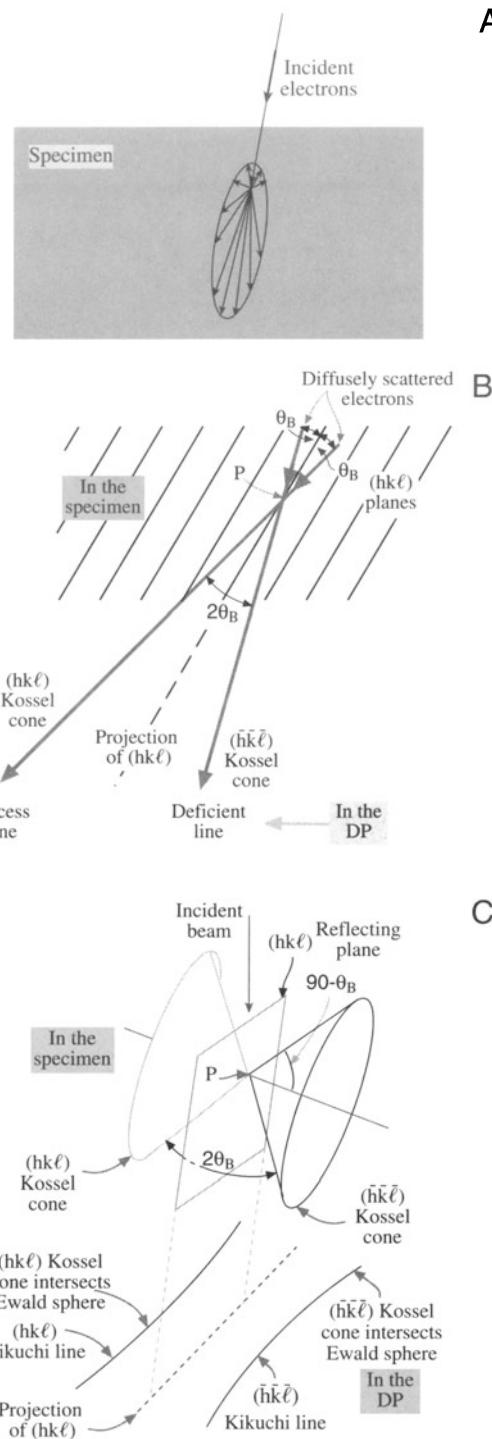
We can make another important observation on the intensity of these lines by considering Figure 19.2 again. In Figure 19.2B you can see that the beam which was initially closest to the optic axis, and therefore the more intense, is further away after being scattered. This beam then gives the excess line 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  $hkl$ ; 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  $(hkl)$ . 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  $\mathbf{g}$  Kikuchi lines is  $\mathbf{g}$  (not  $2\mathbf{g}$ ) because the angle between the two Kossel cones is  $2\theta_B$ .

- When the  $\mathbf{g}$  Kikuchi line passes through the reflection  $G$ ,  $s_g = 0$  (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  $\theta_B$  to certain  $hkl$  planes. The diffracted electrons form Kossel cones centered at P on the diffracting planes. The lines closest to the incident beam direction are dark (deficient) and the lines 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  $\theta_B$  is small.

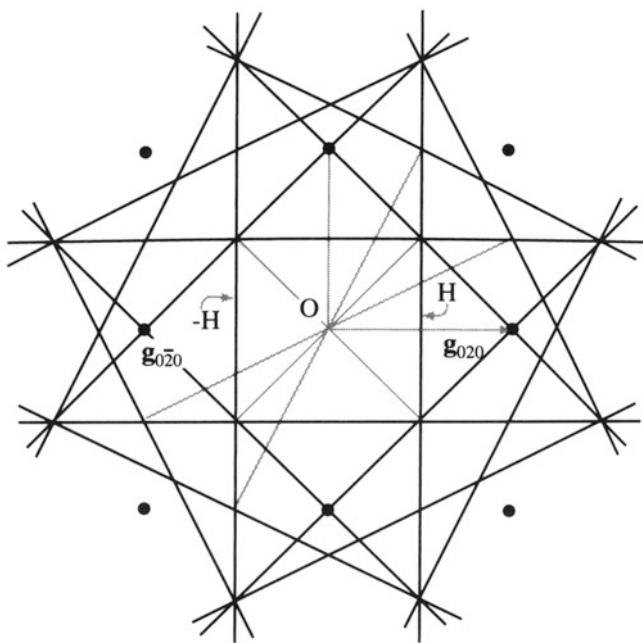
- A corollary: if the direct beam is exactly parallel to the plane  $hkl$ , the  $\mathbf{g}$  and  $\bar{\mathbf{g}}$  Kikuchi lines are symmetrically displaced about O with  $\mathbf{g}$  “passing through”  $\mathbf{g}/2$  and  $\bar{\mathbf{g}}$  “passing through”  $\bar{\mathbf{g}}/2$ .

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.

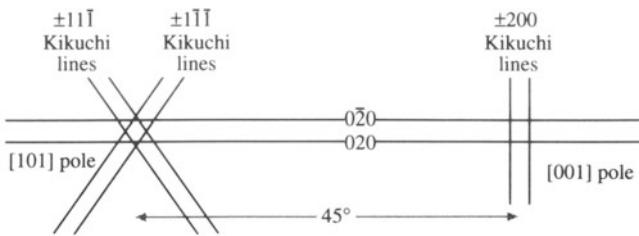
### 19.3. CONSTRUCTING KIKUCHI MAPS

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  $\mathbf{g}$ -vector you can find in the ZOLZ. The distance between each pair of lines is then automatically  $|\mathbf{g}|$ . We can then give each line a unique label  $\mathbf{g}$ .

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  $\pm\mathbf{g}$  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  $\mathbf{g}$ -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  $45^\circ$  to the [101] pole.

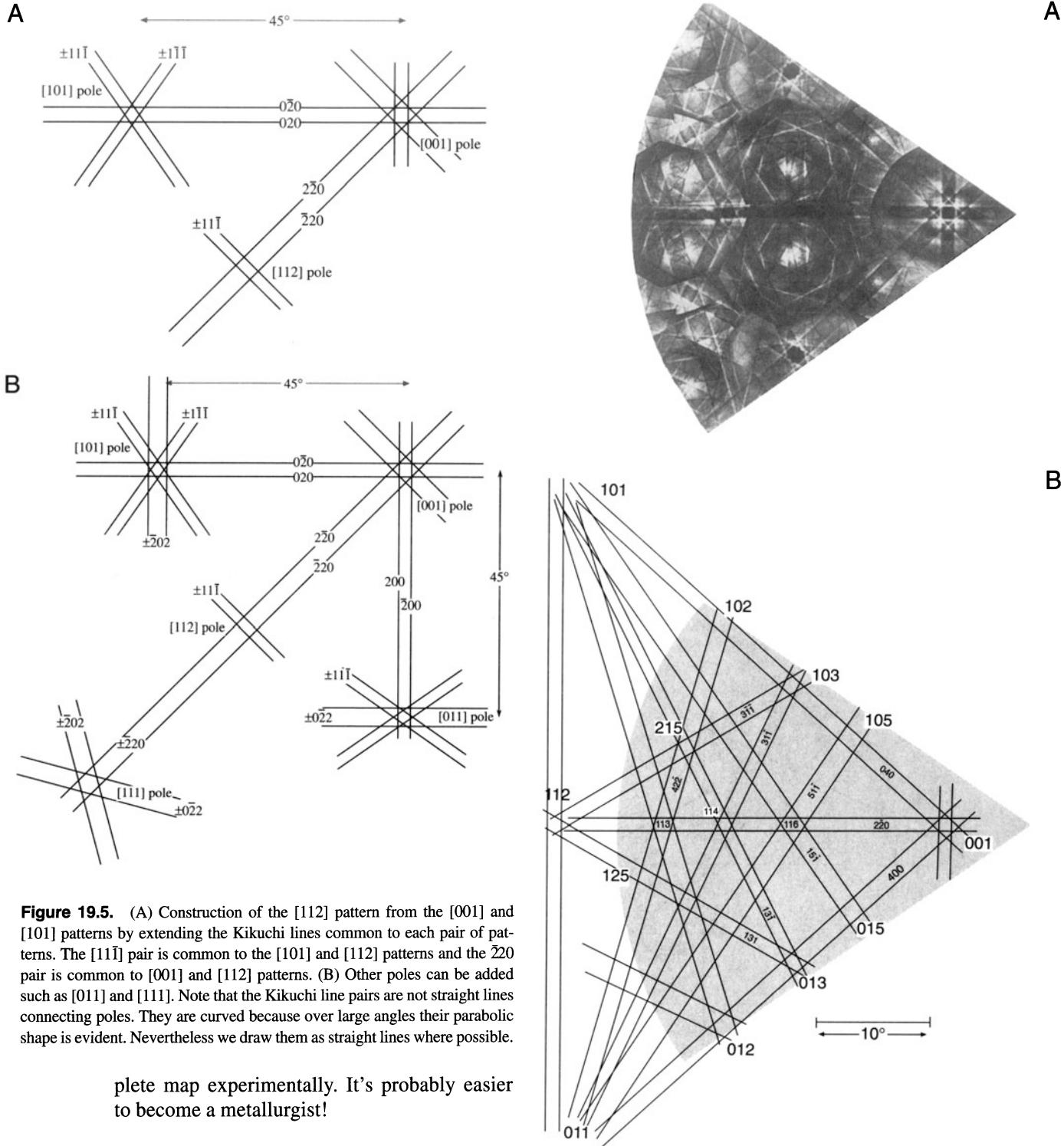
020  $\mathbf{g}$ -vector pointing in the same direction. So, the 020 and  $0\bar{2}0$  Kikuchi lines are common to the two patterns. Although the angle between the [001] and [101] poles is  $45^\circ$ , we draw the 020 lines as parallel and straight because we are always looking at a small segment of 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.

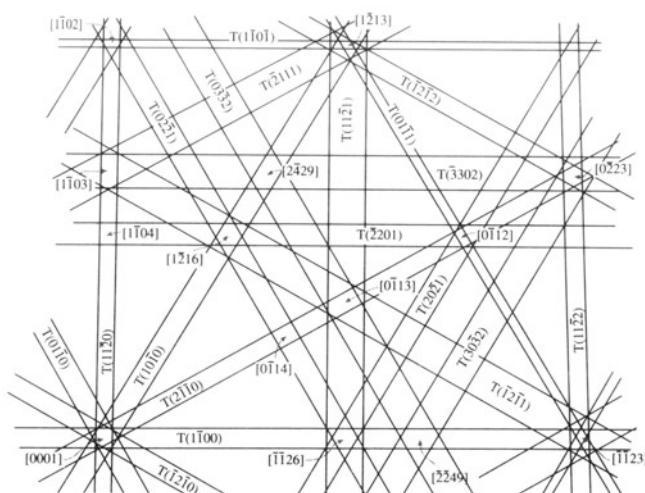
It's a good exercise to construct a Kikuchi map for your material as illustrated for the fcc case in Figure 19.6. The maps are available in the literature for fcc, bcc, diamond-cubic, and some hcp materials. Such maps are mainly from Thomas and co-workers (Levine *et al.* 1966, Okamoto *et al.* 1967, 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  $\text{Ag}_2\text{Al}$ , 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-

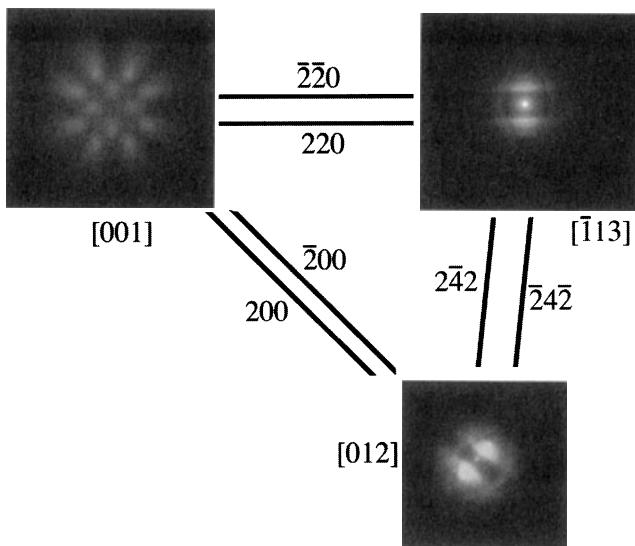


**Figure 19.6.** (A) Experimental Kikuchi map for fcc crystals and (B) indexed Kikuchi lines in the schematic map.



**Figure 19.7.** Part of the schematic Kikuchi map for hexagonal  $\text{Ag}_2\text{Al}$  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  $\sim 3^\circ$ . Using Kikuchi patterns we can increase this accuracy to  $\sim 0.1^\circ$ .

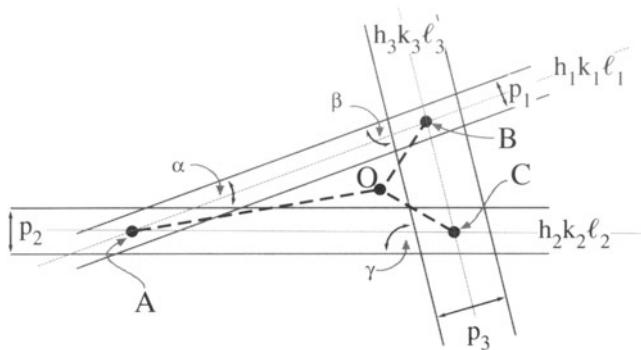
A routine method for this determination has again been developed by Thomas and co-workers (e.g., Okamoto *et al.* 1967), who pioneered this use of Kikuchi maps. The beam direction,  $[UVW]$ , lies along the optic axis O in Figure 19.9. A, B, and C are major poles (i.e., zone axes), which we can determine 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  $\alpha$ ,  $\beta$ , and  $\gamma$  between the traces of the planes in Figure 19.9A (which equals the angle,  $\phi$ , 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,  $\rho_1$ ,  $\rho_2$ , and  $\rho_3$  (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  $\rho_1$ ,  $\rho_2$ , and  $\rho_3$ . Notice we are distinguishing between  $\rho$  and  $\phi$  (see Section 18.4). The angles  $\alpha$ ,  $\beta$ , and  $\gamma$  in Figure 19.9A are slightly distorted values of  $(90 - \phi)$ .

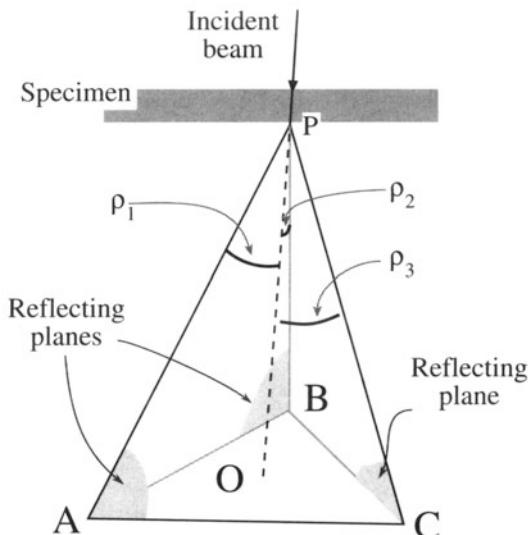
You can solve these three equations for the three unknowns  $U$ ,  $V$ , and  $W$  and hence we have  $\mathbf{B}$ . Finally, always check the sign of  $\mathbf{B}$ , as we described in Section 18.8.

It is possible that the DP you have to work with is not obviously near a zone axis. Normally, while you are at

A

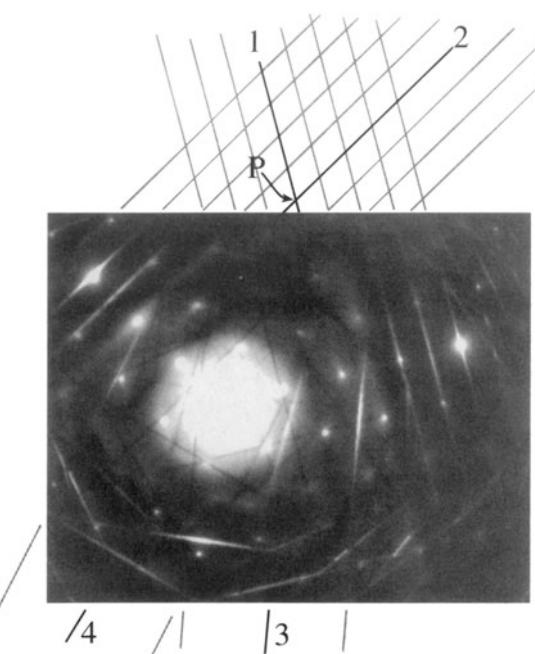


B



**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  $\alpha$ ,  $\beta$ ,  $\gamma$  correspond to the angles between pairs of plane normals. The angle  $\alpha$  is between the  $(h_1 k_1 \ell_1)$  and  $(h_2 k_2 \ell_2)$  plane normals, etc. (B) Three reflecting planes in the specimen with traces AB ( $h_1 k_1 \ell_1$ ), AC ( $h_2 k_2 \ell_2$ ), and BC ( $h_3 k_3 \ell_3$ ) 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  $\mathbf{B}$ , 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  $s_g$ . The diffraction geometry is shown in Figure 19.11 following Okamoto *et al.* (1967). When  $s_g$  is negative, the  $\mathbf{g}$  Kikuchi line is on the same side of  $\mathbf{g}$  as O; when  $s_g$  is positive, the line lies on the opposite side of  $\mathbf{g}$ . For high-energy electrons, and knowing the camera length  $L$ , we can write an expression for the angle  $\eta$

$$\eta = \frac{x}{L} = \frac{x\lambda}{Rd} \quad [19.1]$$

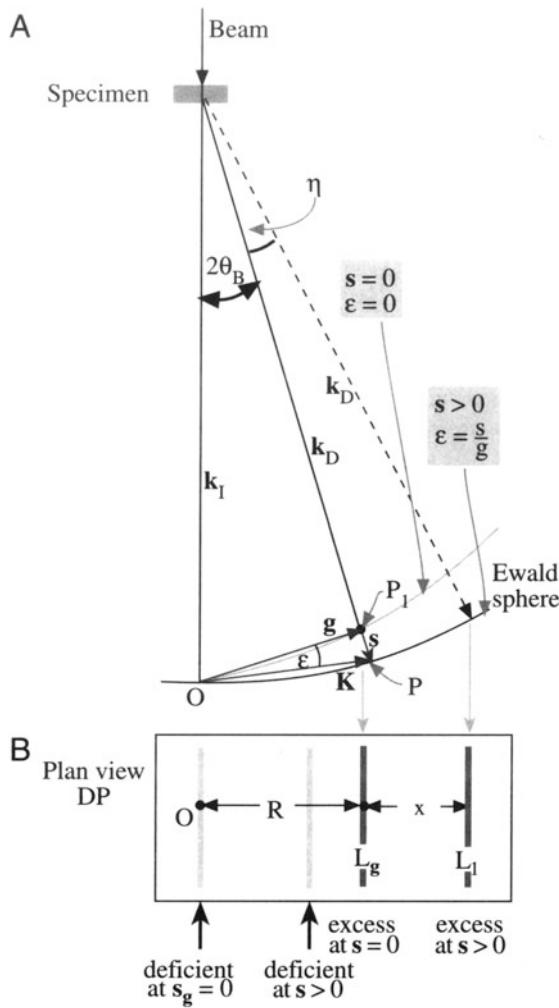
where  $d$  is  $|\mathbf{g}|^{-1}$ . The distances  $x$  and  $R$  are measured on the photographic negative.

The angle  $\epsilon$  is given by

$$\epsilon = \frac{s}{g} \quad [19.2]$$

Now we can set  $\epsilon = \eta$ , to give

$$s = \epsilon g = \frac{x}{L} g = \frac{x}{Ld} \quad [19.3]$$



**Figure 19.11.** (A) The distance between the diffraction spot and its Kikuchi line gives a direct measure of  $s$ . The angle  $\epsilon$  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

$$\frac{R}{L} = 2\theta_B = \frac{\lambda}{d} \quad [19.4]$$

Hence the expression for  $s$  is

$$s = \frac{x}{Ld} = \frac{x}{d} \cdot \frac{\lambda}{Rd} \quad [19.5]$$

$$s = \frac{x}{Rd^2} = \frac{x}{R} \lambda g^2 \quad [19.6]$$

We'll reconsider this equation when we discuss weak-beam microscopy in Chapter 26.

Ryder and Pitsch (1968) have given a method for determining  $\mathbf{B}$  using the approach we described in Section 19.4 with the accuracy given by equation 19.6. Their expression for  $\mathbf{B}$  is

$$\begin{aligned} \mathbf{B} = & \alpha_1 |\mathbf{g}_1|^2 (\mathbf{g}_2 \times \mathbf{g}_3) + \alpha_2 |\mathbf{g}_2|^2 (\mathbf{g}_3 \times \mathbf{g}_1) \\ & + \alpha_3 |\mathbf{g}_3|^2 (\mathbf{g}_1 \times \mathbf{g}_2) \end{aligned} \quad [19.7]$$

where  $\alpha_i$  is given by

$$\alpha_i = \frac{R_i + 2x_i}{R_i} \quad [19.8]$$

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  $\mathbf{g}$  Kikuchi line and the  $\mathbf{G}$  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

---

20.1. Why Use a Convergent Beam? .....	303
20.2. Obtaining CBED Patterns .....	304
20.2.A. Comparing SAD and CBED .....	304
20.2.B. CBED in TEM .....	305
20.2.C. Choosing the C2 Aperture .....	305
20.2.D. Choosing the Camera Length .....	306
20.2.E. Focusing Your Pattern .....	306
20.2.F. Choice of Beam Size .....	307
20.2.G. Effect of Specimen Thickness .....	308
20.2.H. Final Adjustment .....	308
20.2.I. CBED in STEM Mode .....	308
20.3. Zero-Order and Higher-Order Laue-Zone Diffraction .....	309
20.3.A. ZOLZ Patterns .....	309
20.3.B. HOLZ Patterns .....	309
20.3.C. Indexing HOLZ Patterns .....	311
20.4. Kikuchi Lines in CBED Patterns .....	314
20.5. HOLZ Lines .....	315
20.5.A. The Relationship between HOLZ Lines and Kikuchi Lines .....	315
20.5.B. Acquiring HOLZ Lines .....	315
20.5.C. Indexing HOLZ Lines .....	317

## 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  $\sim 0.5 \mu\text{m}$  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  $\mu\text{m}$  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  $\sim 0.1 \mu\text{m}$  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 mini-lenses, 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.

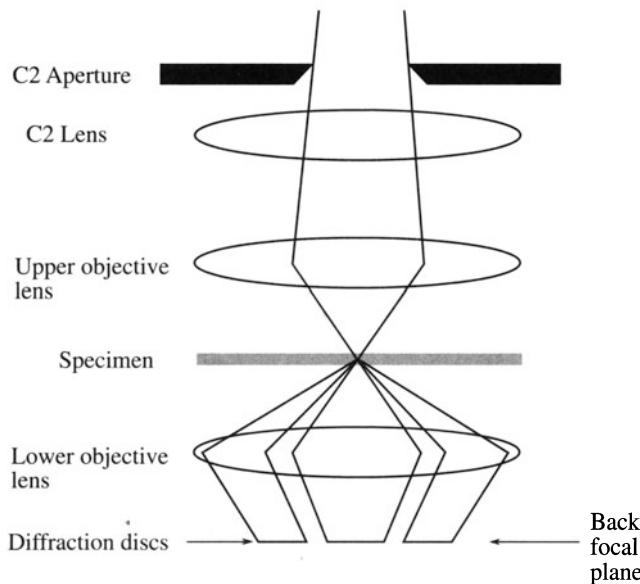
You may find that it is easier to use CBED rather than conventional SAD procedures. When reading the literature, remember that researchers often use the technique with which they are most familiar rather than the best one available.

## 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<sub>2</sub> 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.

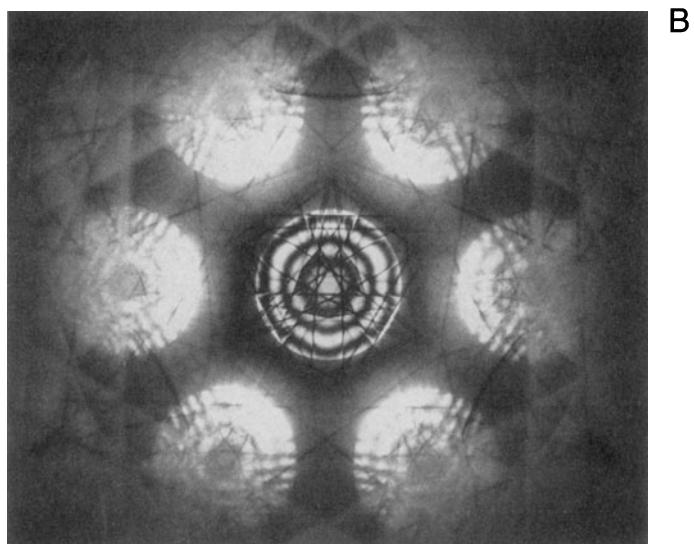
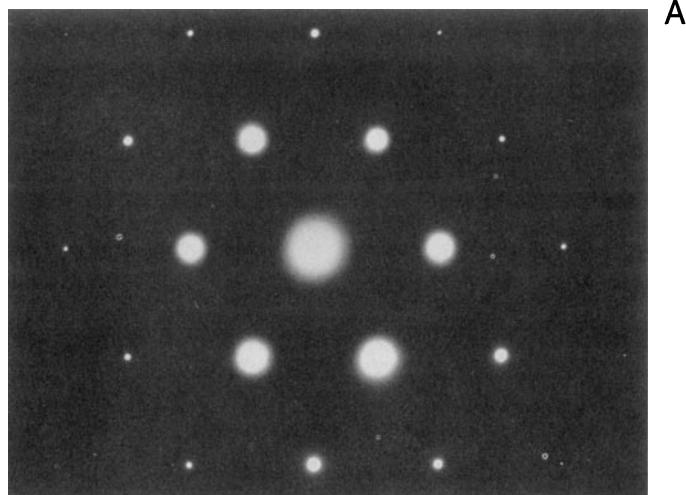
### 20.2.A. Comparing SAD and CBED

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  $\mathbf{k}$ ) and relatively large (usually ~1–10  $\mu\text{m}$  in diameter). In CBED, the beam is convergent (range of  $\mathbf{k}$ -vectors) and relatively small (usually ~10–100 nm in diameter), as shown in Fig-



**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.

ure 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 20.2.** (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 ( $<< 1 \mu\text{m}$ ) beam with a convergence semi-angle ( $\alpha$ )  $> 10$  mrad. 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  $\alpha$ .
- 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  $100\text{--}200 \mu\text{m}$  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  $100\text{--}200 \text{ nm FWTM}$  (see Chapter 5), containing sufficient current to give high intensity in the pattern.
- Select a small camera length,  $< 500 \text{ mm}$ , 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.

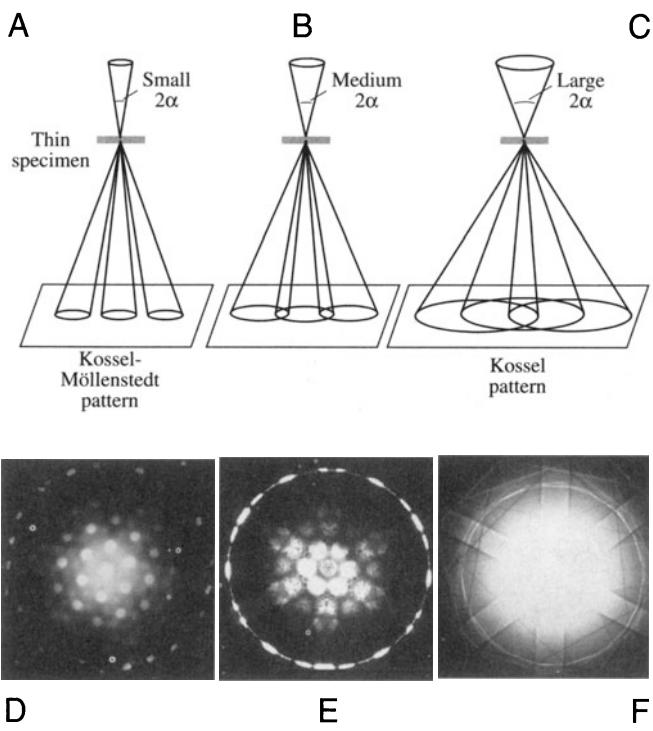
Remember that you control the minimum illuminated area on the screen (i.e., the beam diameter at the specimen) by the strength of the C1 lens.

### 20.2.C. Choosing the C2 Aperture

Once the CBED pattern is visible you can adjust the convergence semiangle  $\alpha$  by changing the C2 aperture, making sure to center the aperture you finally choose. The size of the diffraction disks depends on  $\alpha$ , 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  $10\text{--}50 \mu\text{m}$  range will usually ensure that you have satisfied the K–M conditions; we usually operate in this mode. If  $2\alpha$  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  $2\alpha$ . 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 ( $\pm 10^\circ$ ). 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  $2\alpha$  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  $2\alpha$ . Later, you can worry about the best choice of C2 aperture and focusing the pattern.

Because we need to be able to vary  $\alpha$ , a range of C2 apertures from about  $10 \mu\text{m}$  up to  $200 \mu\text{m}$  is desirable, consistent with the needs of other techniques. A reasonable choice of three C2 apertures comprises one of about  $200 \mu\text{m}$  for routine TEM, EELS, and Kossel patterns, a  $50\text{--}70 \mu\text{m}$  ultrathick aperture for XEDS (which can also be used for STEM imaging and some K-M patterns), and a  $10\text{--}20 \mu\text{m}$  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  $\alpha$ , 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  $\alpha$  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.

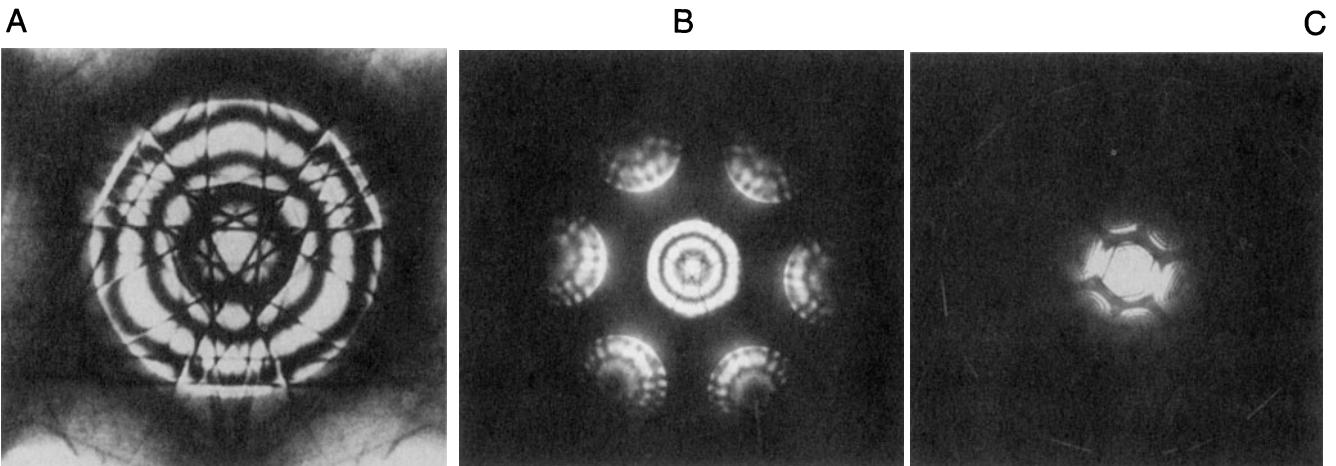
If you need to know the value of  $\alpha$ , you should use a known crystal to calibrate its variation with C2 aperture size for typical C2 lens excitations, as we described back in Section 9.1.

## 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  $L > 1500\text{--}6000 \text{ mm}$  to observe detail in the 000 (BF) disk at the highest possible magnification, and  $L < 500 \text{ mm}$  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  $\pm 10^\circ$ .

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  $\alpha$ , and if you want to maintain a fixed  $\alpha$  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 cross-over 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.

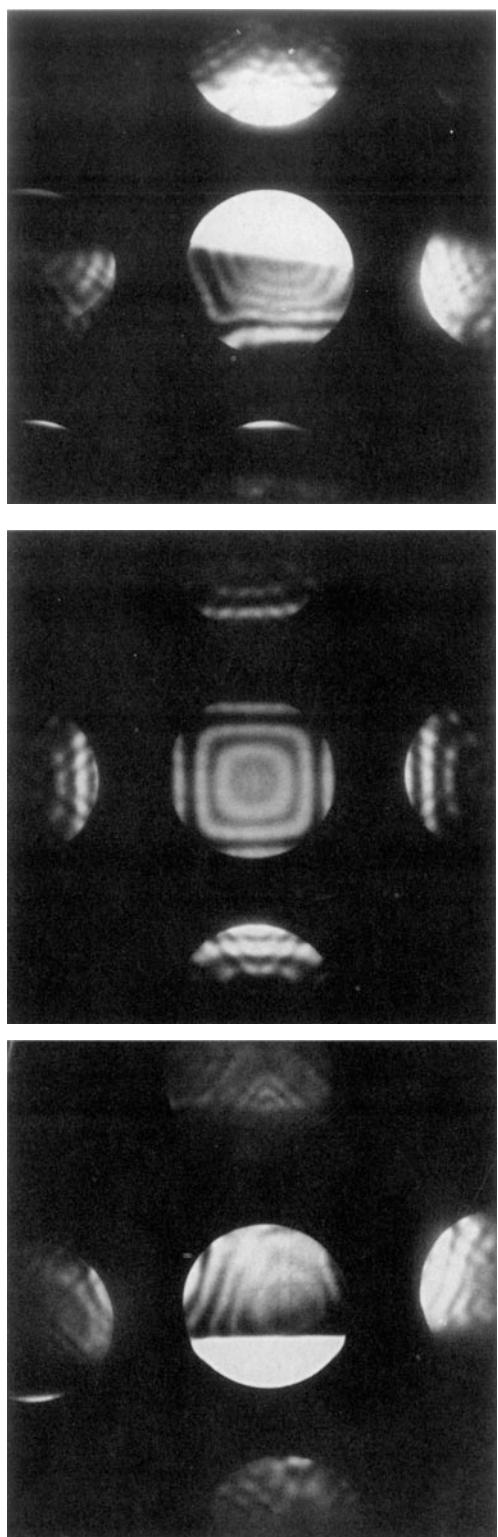
If you leave the objective lens current fixed and focus the beam on the specimen by adjusting C2, you'll see a similar effect to that shown in Figure 20.5. If you use the second (noneucentric) tilt axis, or move to another region of the specimen, you will probably have to refocus the pattern with the  $z$ -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  $hkl$  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.

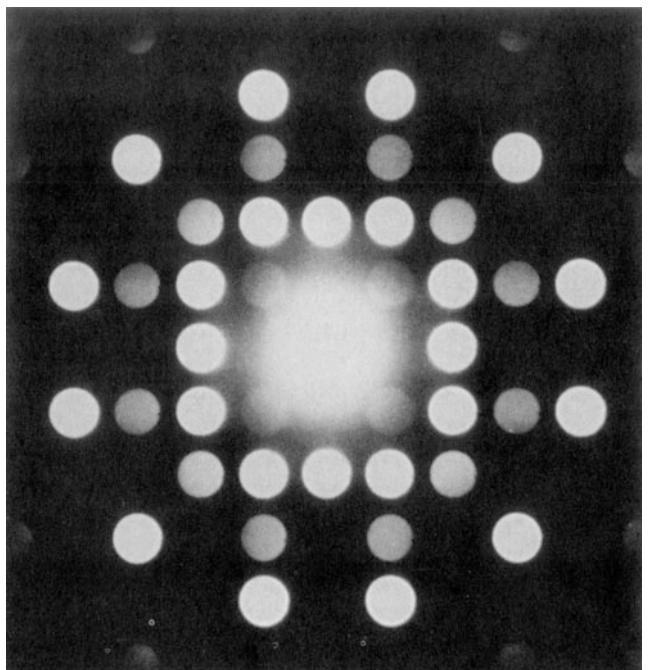
As with SAD patterns, a range of exposure times for all CBED patterns will give you the most information.

## 20.2.I. CBED in STEM Mode

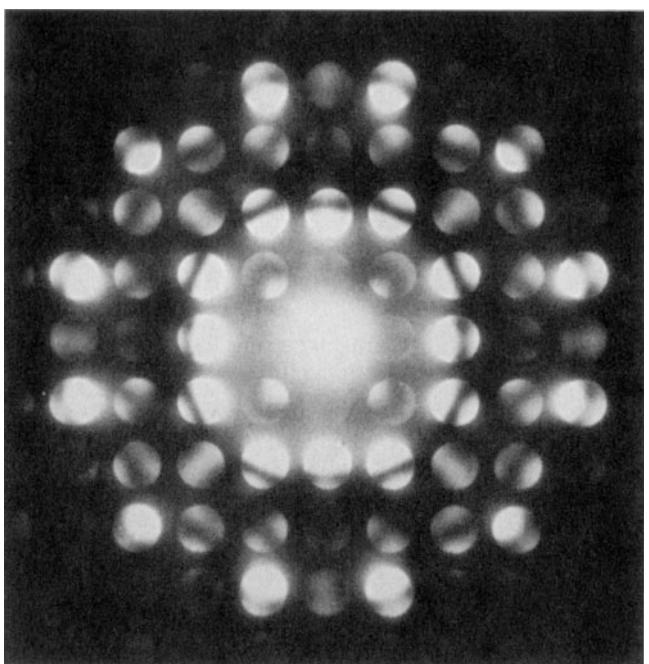
You should first get a focused STEM image of the specimen, as we described in Section 9.4.

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.



A



B

**Figure 20.6.** (A) [001] CBED pattern from  $\sigma$  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  $\alpha$  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  $h\ell$  dif-

fraction maxima must all satisfy the Weiss zone law relationship  $hU + kV + \ellW = 0$ , where  $UVW$  is the beam direction. Remember also that the  $h\ell$  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  $h\ell$  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  $h\ell$  spot spacings. If  $\alpha$  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.

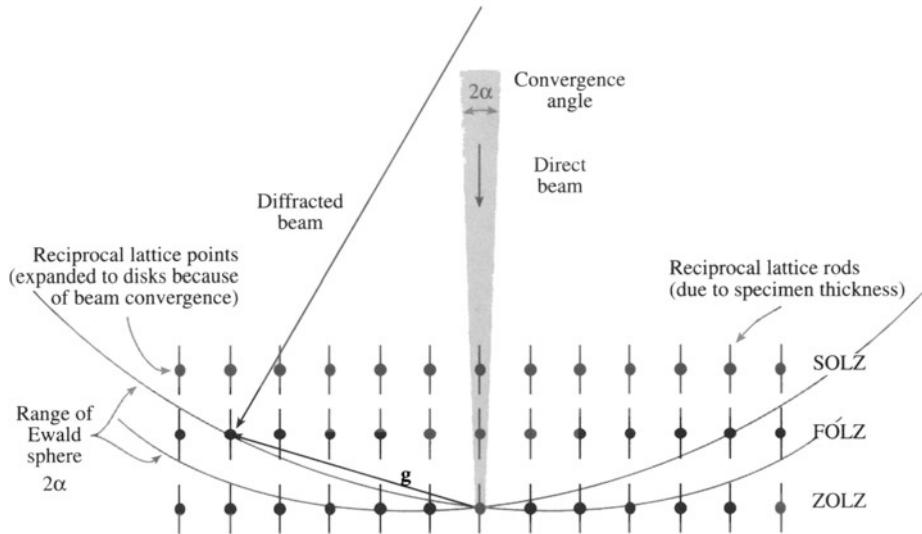
If you've chosen a small convergence semiangle, you'll see a ring of discrete HOLZ spots as in Figure 20.3D while a large C2 aperture gives a HOLZ ring of intersecting lines as in Figure 20.3F. The HOLZ intensity arises from weak high-angle diffraction from crystal planes that are *not* parallel to the 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<sub>2</sub> 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  $h\ell$  reflections satisfy the relationship  $hU + kV + \ellW = 1$ , 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  $2\alpha$ , the Ewald sphere is effectively rotated  $2\alpha$

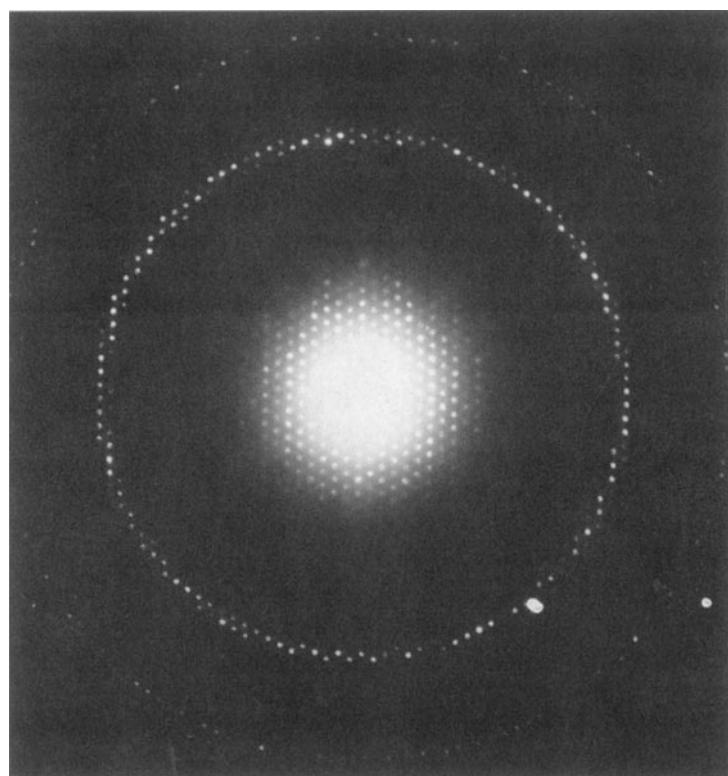
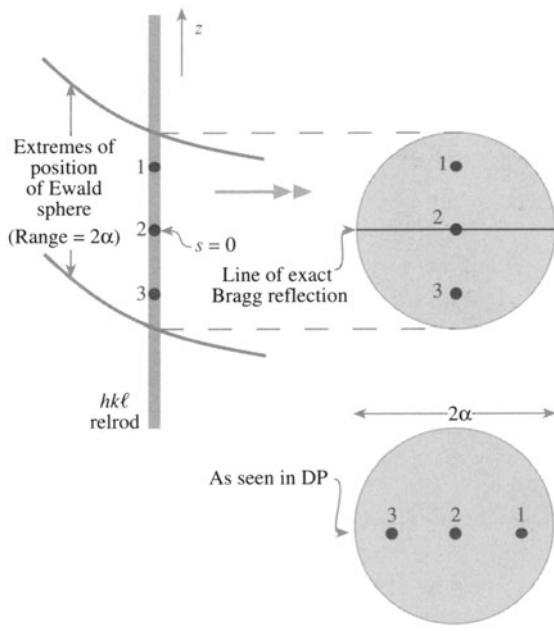
about the origin, and thus a range of angles in each HOLZ relrod is sampled, as shown in Figure 20.7A. This range of angles manifests itself in the CBED disk, sampling the intensity distribution along the relrod, as shown in Figure 20.7B. Different interception points on the relrod correspond to different points in the disk, also shown in Figure 20.7C.

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.

A



B



**Figure 20.7.** (A) The Ewald sphere can intercept reciprocal lattice points from planes not parallel to the electron beam whose  $\mathbf{g}$  vectors are not normal to the beam. The sphere has an effective thickness of  $2\alpha$  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  $h\bar{k}\ell$  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.

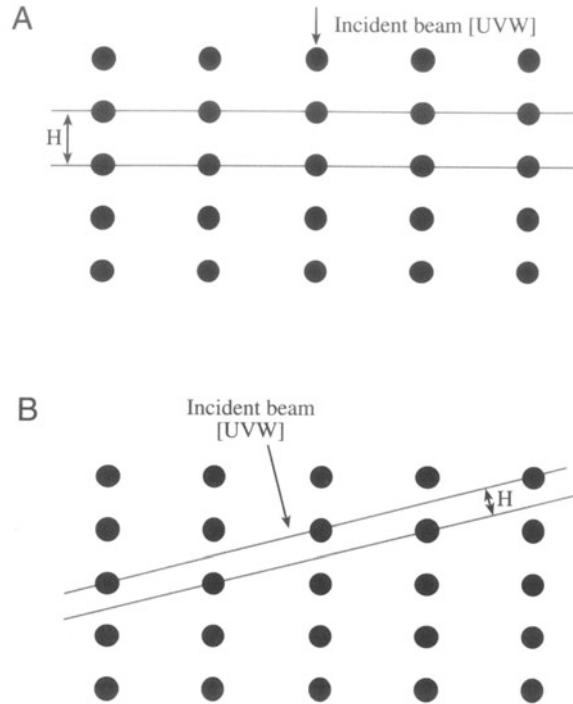
To observe HOLZ rings in addition to the ZOLZ pattern, choose a very small  $L$  ( $<500$  mm) so that you can see the full angular range of the back focal plane permitted by the imaging system ( $\pm 10^\circ$ ). As shown schematically in Figure 20.7A, the Ewald sphere only intercepts reciprocal lattice 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 reflection rods 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.,  $\langle 114 \rangle$ ) since this gives you a better chance of observing the FOLZ than a high-symmetry zone axis, such as  $\langle 001 \rangle$ . (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  $\lambda$  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



**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.

then determine the unit cell, the crystal system, and also the type of lattice centering.

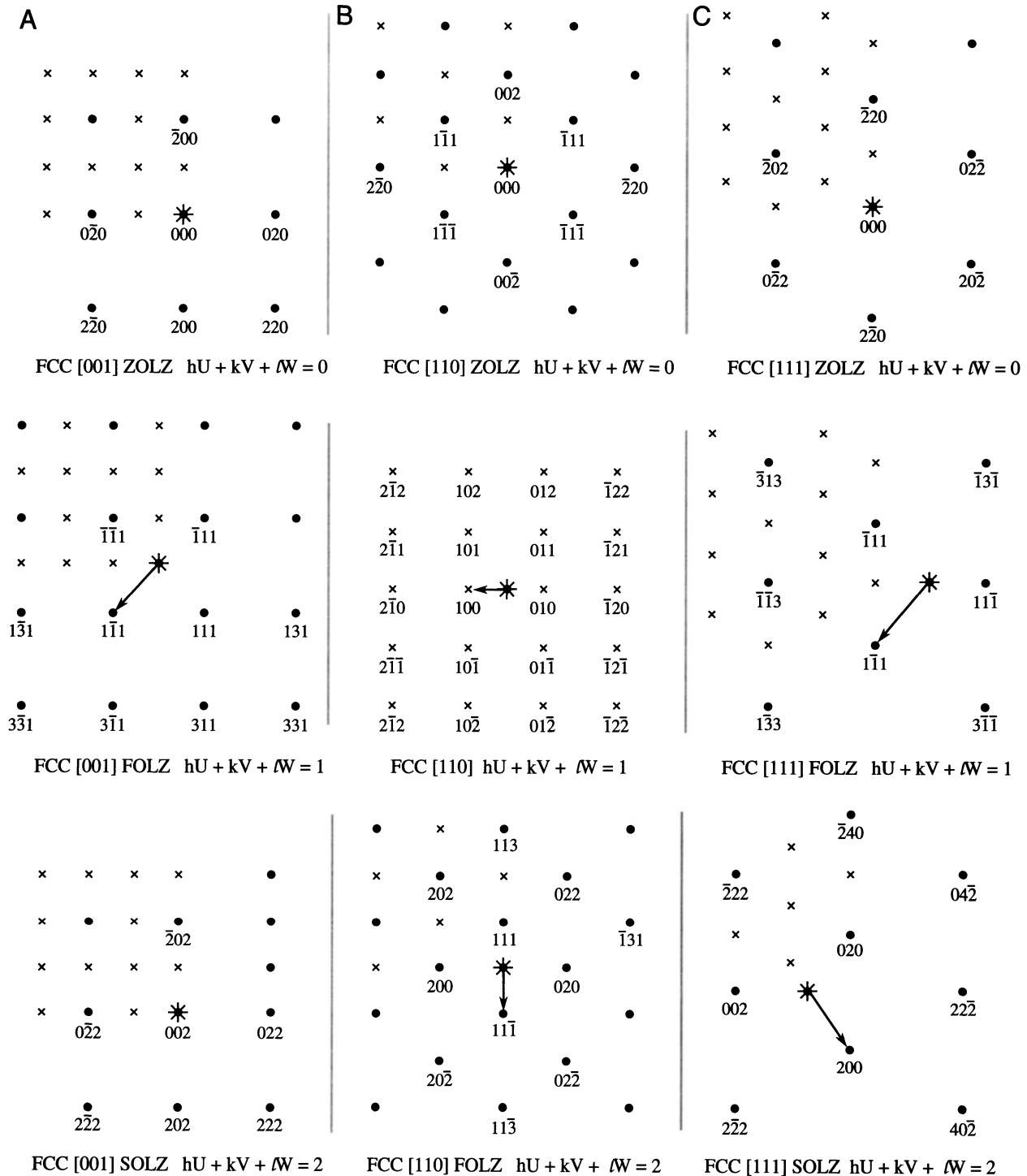
### 20.3.C. Indexing HOLZ Patterns

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  $hkl$  planes and ignores any systematic absences.

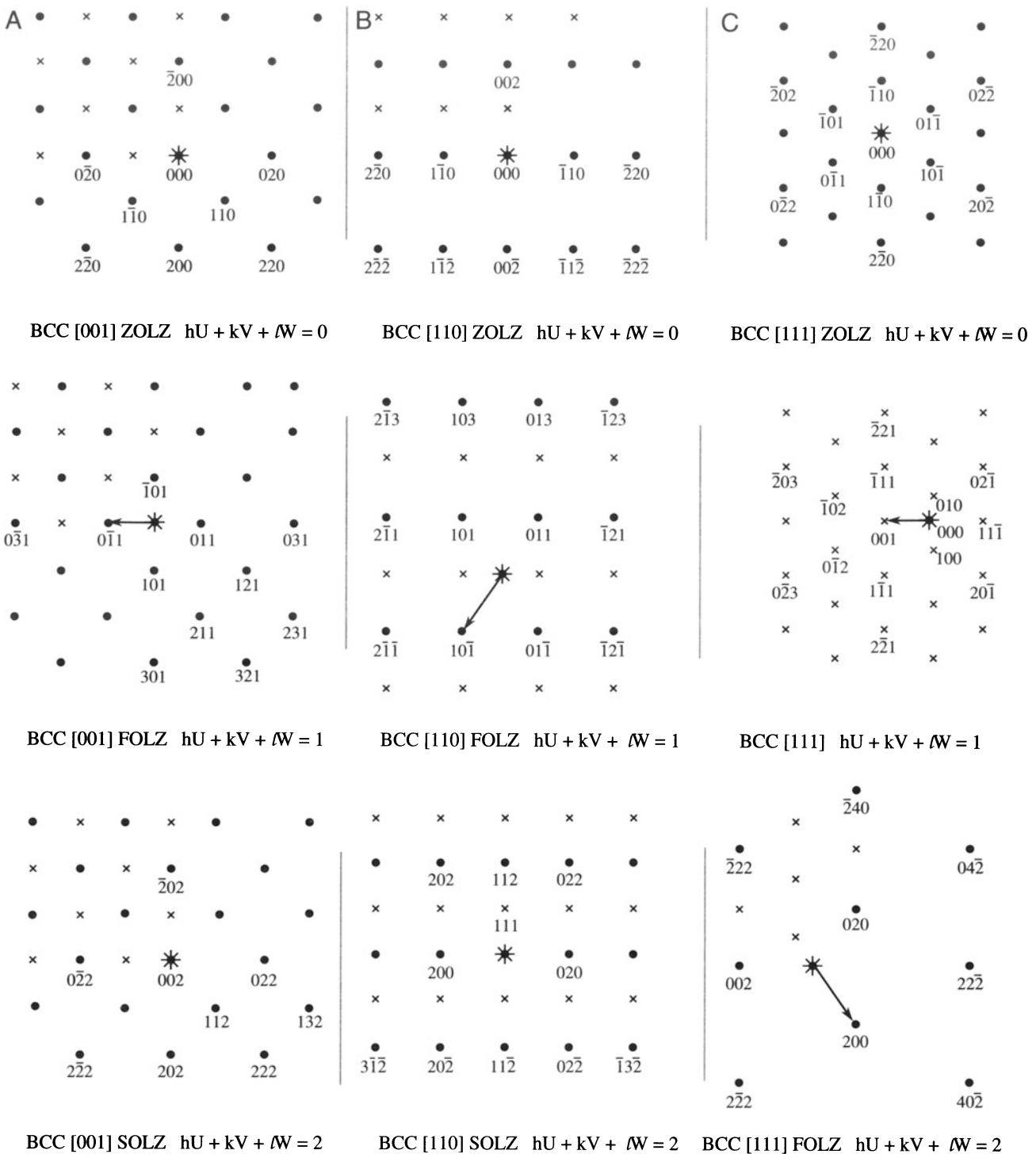
Examples of indexed ZOLZ, FOLZ, and SOLZ patterns are shown in Figures 20.9A–C for the fcc lattice under (A)  $[001]$ , (B)  $[110]$ , and (C)  $[111]$  beam directions; Figures 20.10A–C show similar patterns for the bcc lattice.



**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

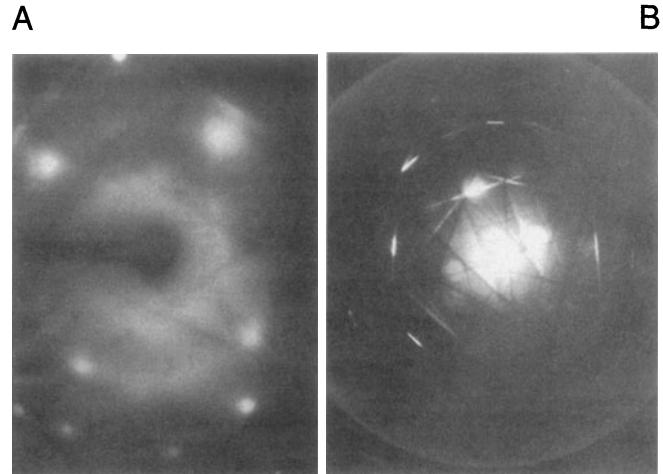
EMS and Digital Microscopist, can do this readily (Section 1.5). There is also freeware available on the WWW. You may have to generate less symmetrical patterns (e.g., [233] or [114]) since these give rise to HOLZ rings with smaller radii, in which HOLZ phenomena are easier to see.

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  $(110)_{\text{fcc}}$  and  $(111)_{\text{bcc}}$  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  $\alpha\text{-Al}_2\text{O}_3$ , 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).

## 20.4. KIKUCHI LINES IN CBED PATTERNS

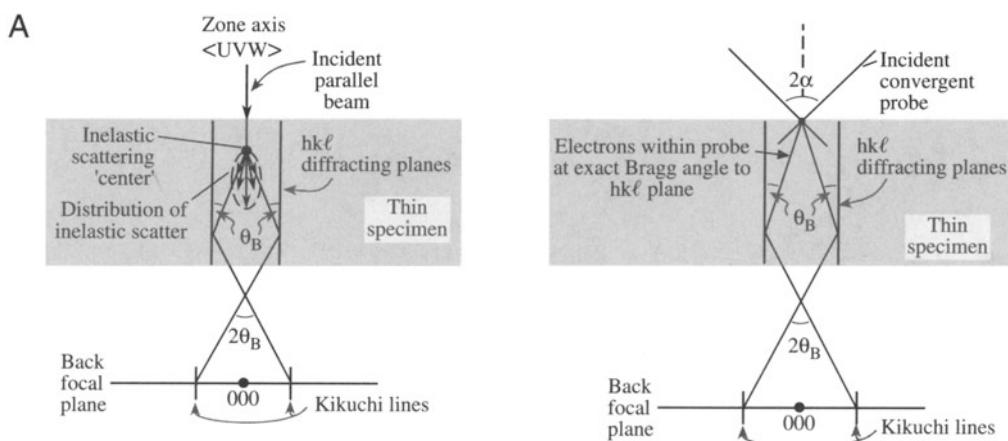
In CBED patterns you almost invariably see sharp Kikuchi lines, while in SAD patterns Kikuchi lines are often rather diffuse or absent (see Chapter 19). This difference arises mainly because the convergent beam samples a much smaller region of the specimen than that selected by the SAD aperture. So in 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 show-



**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.

ing 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,  $\alpha$ , is greater than the Bragg angle,  $\theta_B$ .

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 Kosel 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  $hkl$  disks). However, the term is used rather loosely in the literature to describe the ZOLZ intensity bands, despite the elastic contribution to the scattered intensity.

## 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  $\mathbf{g}$ -vectors), so they are more sensitive to changes in lattice parameter than the ZOLZ lines. Since

$$|\mathbf{g}| = \frac{1}{d}, \quad |\Delta\mathbf{g}| = -\frac{\Delta d}{d^2} \quad [20.1]$$

So the smaller  $d$ , the larger  $|\Delta\mathbf{g}|$  at the same  $\Delta d$ . 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

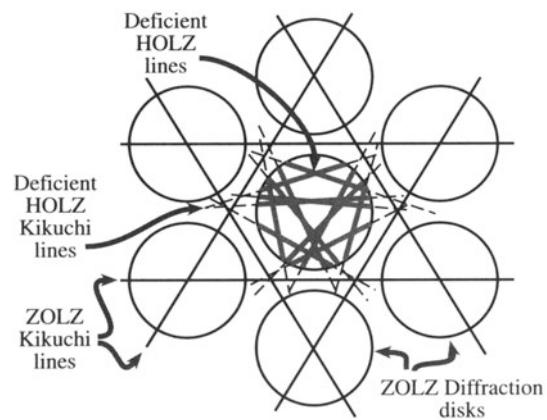
through the HOLZ disk and leaving a dark line in the 000 disk. Not surprisingly, the theory for the origin of HOLZ lines is much more complicated than this. When you have time, you should read the paper by Jones *et al.* (1977).

As you see, the HOLZ lines come in pairs, like Kikuchi lines, with the bright (excess) lines within the HOLZ  $hkl$  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.



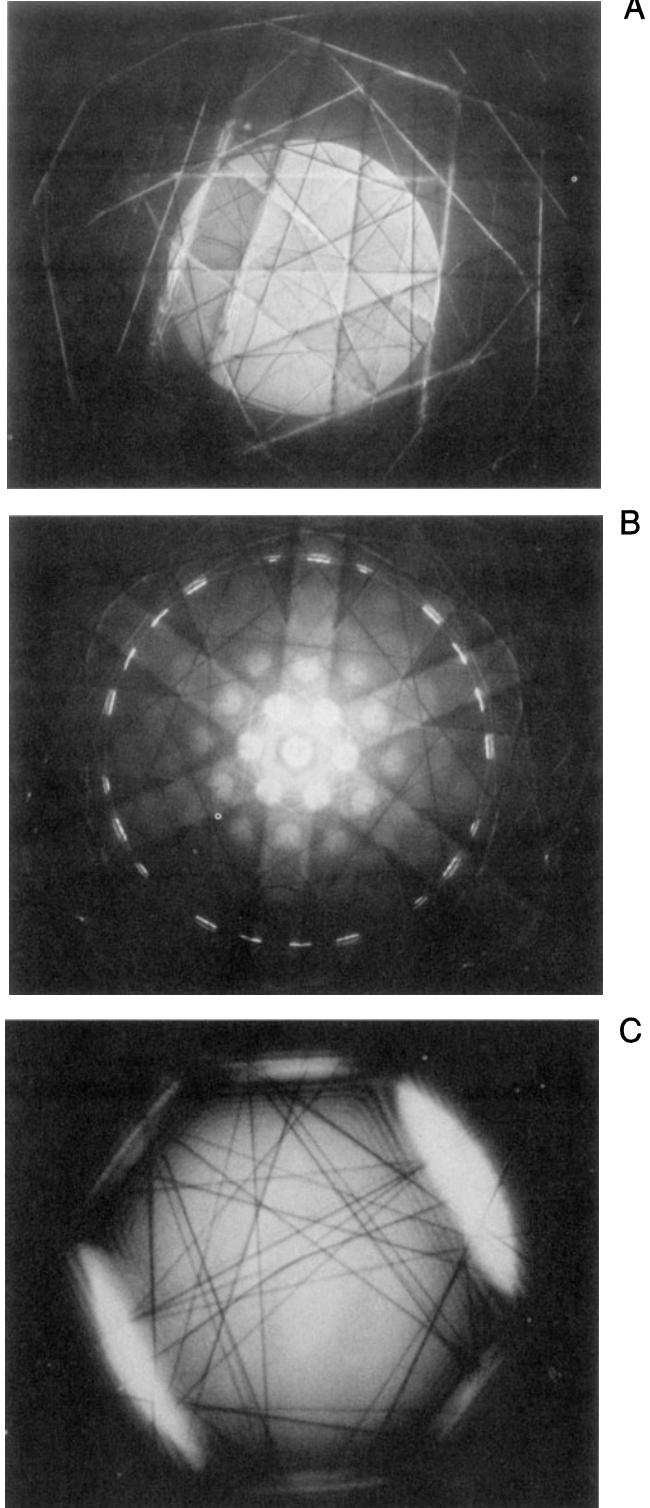
**Figure 20.13.** The relationship between Kikuchi lines and HOLZ lines is shown in this schematic of a [111] CBED pattern from an fcc crystal. The three principal pairs of  $2\bar{2}0$   $1\bar{1}0$  ZOLZ Kikuchi lines show sixfold symmetry and bisect the  $\mathbf{g}$ -vectors from 000 to the  $2\bar{2}0$  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  $2\alpha$ ) 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.



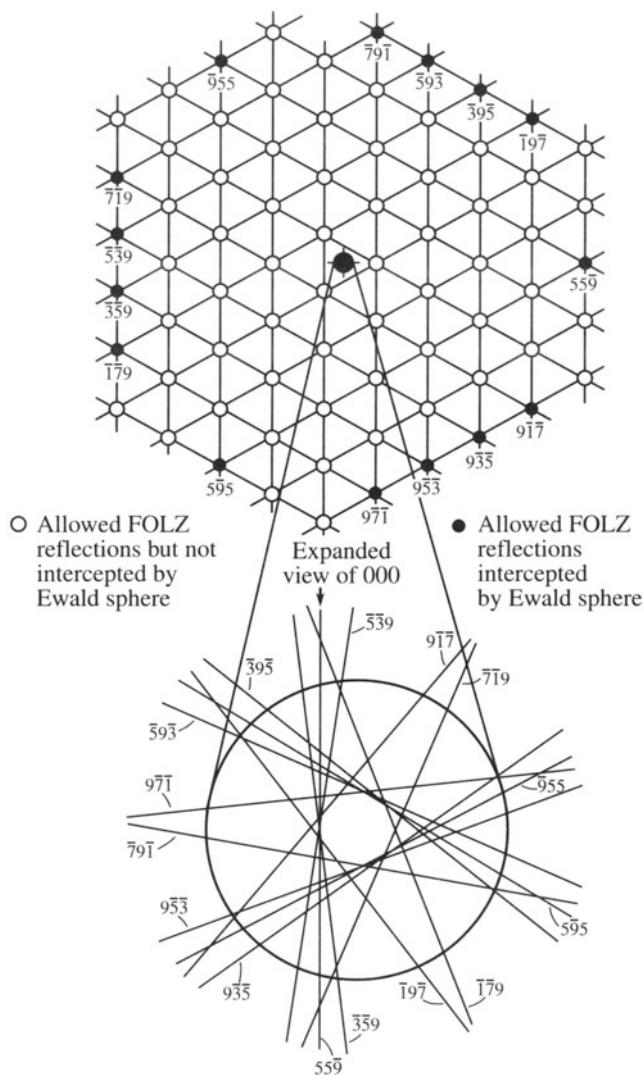
**Figure 20.14.** (A) Low  $L$ , large  $\alpha$  CBED pattern showing a wide area of reciprocal space, away from ZAP conditions. (B) When the specimen is tilted to a low-index ZAP, and a smaller C2 aperture is inserted, a ring of excess HOLZ lines appears associated with the HOLZ disks. (C) Deficient 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

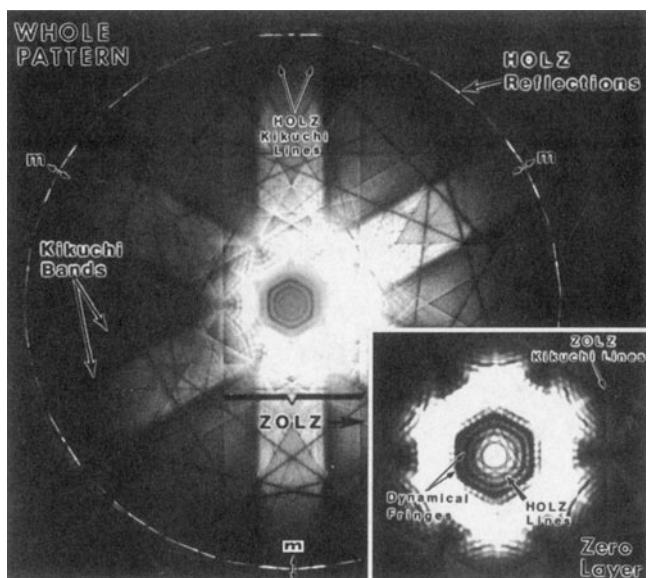
When you're recording the pattern containing the deficient HOLZ lines, you must also record another pattern with a small  $L$  and a small  $\alpha$  to show the HOLZ disks: the first thing you have to do is index the HOLZ  $hkl$  maxima in the manner we described in Section 20.4. Then observe which 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 ex-

ercise 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  $hkl$  FOLZ disk is normal to the  $hkl$  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,  $\alpha$ , 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.
- Ecob, R.C., Shaw, M.P., Porter, A.J., and Ralph, B. (1981) *Phil. Mag. A* **44**, 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. A* **354**, 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.

### 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.

# Using Convergent-Beam Techniques

# 21

---

21.1. Thickness Determination .....	321
21.2. Unit-Cell Determination .....	323
21.2.A. Experimental Considerations .....	323
21.2.B. The Importance of the HOLZ-Ring Radius .....	323
21.2.C. Determining the Lattice Centering .....	325
21.3. Symmetry Determination .....	326
21.3.A. Introduction to Symmetry Concepts .....	326
21.3.B. Friedel's Law .....	328
21.3.C. Looking for Symmetry in Your Patterns .....	328
21.3.D. Point-Group Determination .....	330
21.3.E. Space-Group Determination .....	332
21.4. Lattice Parameter, Strain, and Composition Analysis .....	338
21.5. Determination of Enantiomorphism .....	341
21.6. Convergent-Beam Imaging .....	341
21.7. Scanning-Beam Diffraction .....	341
21.8. Other Methods of Microdiffraction .....	343

## 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,  $\xi_g$ ; 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  $hkl$  reflection. If you do this you will see that the CBED disks contain parallel rather than concentric

intensity oscillations, as shown in Figure 21.1. 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).

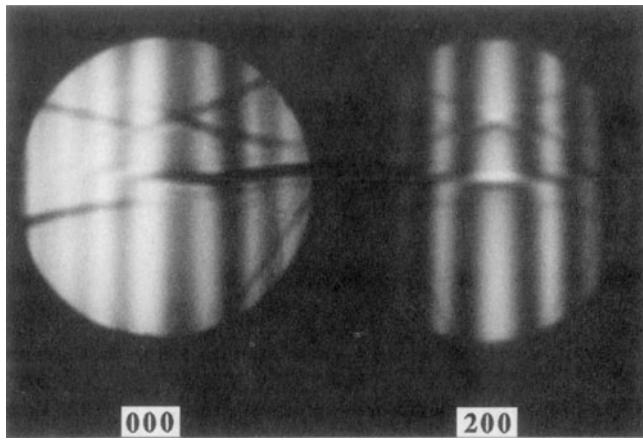
These oscillations are symmetric in the  $hkl$  disk and asymmetric in the 000 disk.

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  $hkl$  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  $hkl$  disk through a  $10\times$  loupe 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  $\pm 0.1$  mm. The central bright fringe is at the exact Bragg condition where  $s = 0$ . 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  $i$ th fringe from the equation

$$s_i = \lambda \frac{\Delta\theta_i}{2\theta_B d^2} \quad [21.1]$$

where  $\theta_B$  is the Bragg angle for the diffracting  $hkl$  plane,  $d$  is the  $hkl$  interplanar spacing, and we'll use the magnitude of  $s$ , ignoring its sign. The angle  $2\theta_B$  in the CBED pattern is, of course, just the separation of the 000 and  $hkl$  disks.



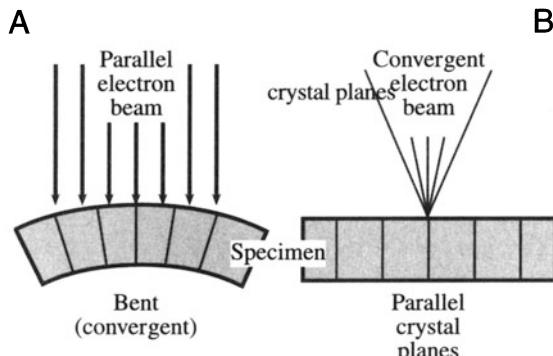
**Figure 21.1.** Kossel-Möllenstedt fringes in a ZOLZ CBED pattern from pure Al taken under two-beam conditions with (200) strongly excited.

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  $\xi_g$  is known, then you can determine the foil thickness  $t$  since

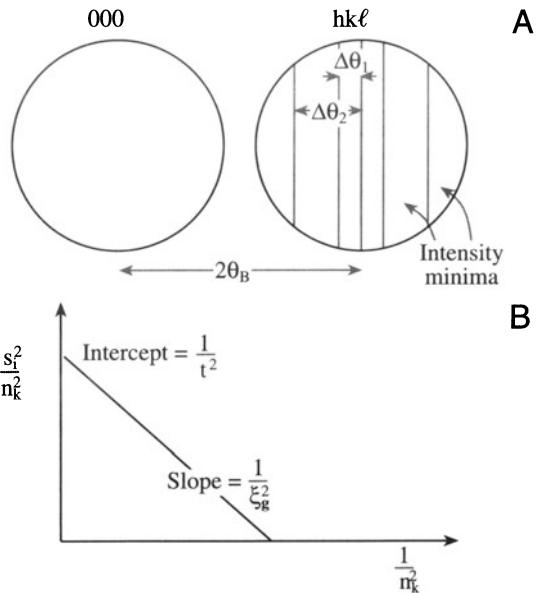
$$\frac{s_i^2}{n_k^2} + \frac{1}{\xi_g^2 n_k^2} = \frac{1}{t^2} \quad [21.2]$$

where  $n_k$  is an integer ( $k$  is an integer not related to  $\lambda$ ). If you don't know  $\xi_g$ , 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  $k$



**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  $t$ .

is given by  $k = i + j$ , where  $j$  is the largest integer  $\langle t/\xi_g \rangle$ .

- 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.

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 <sup>-1</sup> )	$n_i$	$s_i^2/n_i^2$ (nm <sup>-2</sup> )
$s_1 = 0.84 \times 10^{-2}$	$n_1 = 1$	$0.7 \times 10^{-4}$
$s_2 = 2.1 \times 10^{-2}$	$n_2 = 2$	$1.1 \times 10^{-4}$
$s_3 = 3.0 \times 10^{-2}$	$n_3 = 3$	$1.0 \times 10^{-4}$

**Table 21.2. Alternative CBED Data for Thickness Determination**

$s_i$ (nm <sup>-1</sup> )	$n_k$	$s_i^2/n_k^2$ (nm <sup>-2</sup> )
$s_1 = 0.84 \times 10^{-2}$	$n_1 = 2$	$1.7 \times 10^{-5}$
$s_2 = 2.1 \times 10^{-2}$	$n_2 = 3$	$4.9 \times 10^{-5}$
$s_3 = 3.0 \times 10^{-2}$	$n_3 = 4$	$5.6 \times 10^{-5}$

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  $6.1 \times 10^{-5}$  nm<sup>-2</sup>. Therefore, we find that  $t^2 = (6.1)^{-1} \times 10^5$  nm<sup>2</sup>, 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.

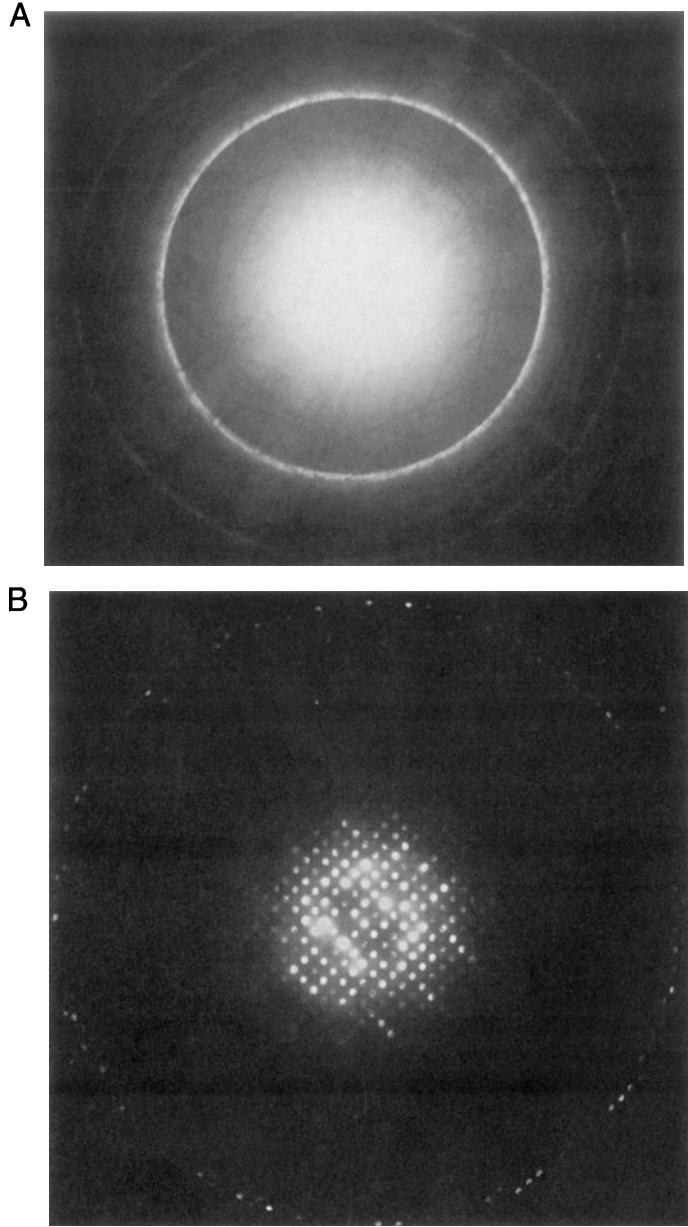
It is good practice to record two patterns, one with a large C2 aperture (therefore large  $\alpha$ , Kossel conditions) and one with a small C2 aperture to show the individual disks (K-M conditions).

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 ( $\alpha \sim 10^\circ$ ), 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- $\mu\text{m}$  C2 aperture showing the FOLZ ring of intensity surrounding the overexposed ZOLZ region. In (B) the same pattern, taken with a 20- $\mu\text{m}$  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  $\lambda$ , 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)

$$G_1 = \left( \frac{2H}{\lambda} \right)^{1/2} \quad [21.3]$$

and

$$G_2 = 2 \left( \frac{H}{\lambda} \right)^{1/2} \quad [21.4]$$

where both  $G$  and  $H$  are in reciprocal-space units ( $\text{nm}^{-1}$  or  $\text{\AA}^{-1}$ ). 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

$$\frac{1}{H} = \frac{2}{\lambda G_1} \quad [21.5]$$

The value of  $H^{-1}$  can be expressed in real-space units (nm) through the measured radius  $r$  (mm) and the camera constant  $\lambda L$  (nm mm)

$$\frac{1}{H} = \left( \frac{2}{\lambda} \right) \left( \frac{\lambda L}{r} \right)^2 \quad [21.6]$$

You must take the time to measure  $\lambda L$  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

$$\frac{1}{H} = \left| [UVW] \right| \quad [21.7]$$

and so this magnitude can be calculated for a specific beam direction  $[UVW]$ .

*Example*

For an fcc crystal (Steeds 1979)

$$\frac{1}{H} = \frac{a_0(U^2 + V^2 + W^2)^{1/2}}{p} \quad [21.8]$$

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

$$\frac{1}{H_c} = n \left( \frac{1}{H_m} \right) \quad [21.9]$$

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  $\lambda$  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  $UVW$

$$\frac{1}{H} = \left( a^2 U^2 + b^2 V^2 + c^2 W^2 \right)^{1/2} \quad [21.10]$$

Similarly, for hexagonal or rhombohedral systems using a three-index system

$$\frac{1}{H} = \left( a^2 (U^2 + V^2 - UV) + c^2 W^2 \right)^{1/2} \quad [21.11]$$

and for the four-index system

$$\frac{1}{H} = \left( 3(U^2 + V^2 + UV)^2 + c^2 W^2 \right)^{1/2} \quad [21.12]$$

while for the monoclinic system with a unique  $b$ -axis

$$\frac{1}{H} = \left( U^2 a^2 + V^2 b^2 + W^2 c^2 + 2UWac \cos \beta^2 \right)^{1/2} \quad [21.13]$$

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.

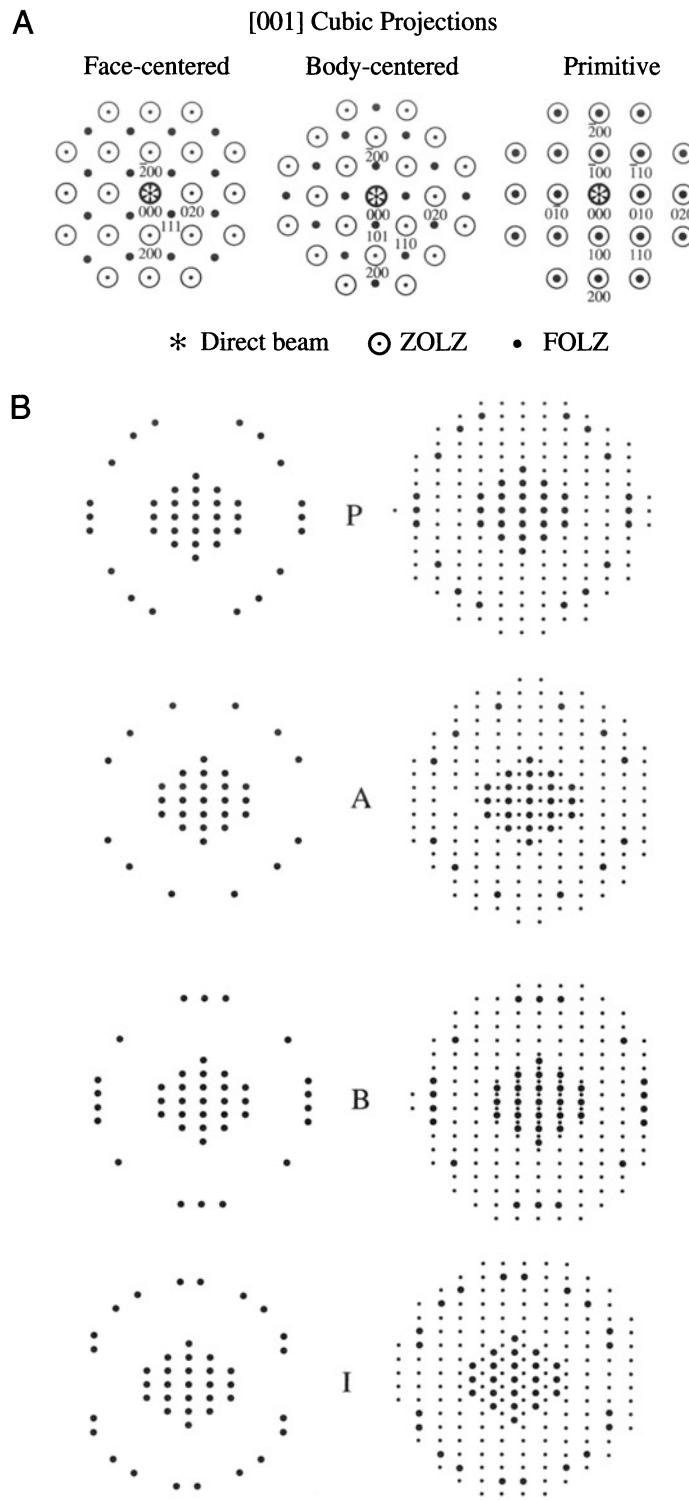
## 21.2.C. Determining the Lattice Centering

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

$$\mathbf{t} = \mathbf{g} - \mathbf{u}^* \left( \frac{HN_L}{|\mathbf{u}^*|} \right) \quad [21.14]$$

where  $\mathbf{g}$  is the vector for the  $hkl$  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  $hkl$ . 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.

Type of symbol	Triclinic	Monoclinic and orthorombic	Trigonal	Tetragonal	Hexagonal	Cubic
X						
$\bar{X}$						$\bar{2}3 = \frac{2}{m} 3$
$\frac{X}{m}$						$m3 = \frac{2}{m} 3$
$Xm$						$2m3 = \frac{2}{m} 3$
$\bar{X}m$						$\bar{1}m = \frac{2}{m}$ $\bar{2}m = mm$ $\bar{3}m$ $\bar{4}2m$ $\bar{6}m2 = \frac{3}{m} m$ $\bar{4}3m$
$X2$						
$\frac{X}{m}m$						$\frac{1}{m} m = mm$ $\frac{3}{m} m = \bar{6}m$ $\frac{4}{m} mm$ $\frac{6}{m} mm$ $m3m = \frac{4}{m} 3m$

**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: