Handbook of surface and interface analysis methods for problem-solving

John C. Rivière, Sverre Myhra

Transmission Electron Microscopy: Instrumentation, Imaging Modes, and Analytical Attachments

Publication details
Birgit Hagenhoff
Published online on: 24 Jun 2009

Accessed on: 18 Oct 2023

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
6 Transmission Electron Microscopy: Instrumentation, Imaging Modes, and Analytical Attachments

John M. Titchmarsh

CONTENTS

6.1 Introduction ........................................................................................................................... 140
   6.1.1 List of Symbols (Units), Acronyms, and Abbreviations ........................................... 141
6.2 Electron Microscope Instrumentation .................................................................................. 143
   6.2.1 Conventional TEM .................................................................................................... 143
   6.2.2 Electron Sources ....................................................................................................... 144
   6.2.3 Illumination System .................................................................................................. 145
   6.2.4 Sample Stage ............................................................................................................ .1 4 5
   6.2.5 Objective Lens .......................................................................................................... 146
   6.2.6 Projection and Recording ......................................................................................... 147
6.3 Electron–Sample Interaction ................................................................................................ 148
   6.3.1 Diffraction Pattern .................................................................................................... 148
   6.3.2 Diffraction Contrast Imaging ................................................................................... 152
   6.3.3 Weak-Beam Imaging ................................................................................................ 153
   6.3.4 HREM Image Formation .......................................................................................... 155
      6.3.4.1 Spatial Coherence ....................................................................................... 156
      6.3.4.2 Electron Beam Propagation through the Sample ....................................... 156
      6.3.4.3 Objective Lens Aberrations ......................................................................... 158
      6.3.4.4 Temporal Coherence .................................................................................. 160
      6.3.4.5 Phase Problem ............................................................................................ 162
      6.3.4.6 Implications for HREM Imaging of Interfaces .......................................... 164
6.4 Scanning Transmission Electron Microscopy ...................................................................... 164
   6.4.1 STEM Imaging ......................................................................................................... 165
   6.4.2 STEM Instrumentation ............................................................................................. 165
   6.4.3 STEM Imaging Modes.............................................................................................. 166
6.5 Chemical Analysis Methods ................................................................................................. 170
   6.5.1 Compositional Analysis by EDX .............................................................................. 170
      6.5.1.1 Principles of EDX ...................................................................................... 170
      6.5.1.2 Protocol for EDX Spectrum Acquisition ................................................... 172
      6.5.1.3 EDX Spectrum Processing ......................................................................... 173

© 2009 by Taylor & Francis Group, LLC
6.1 INTRODUCTION

Transmission electron microscopy (TEM) and the related scanning transmission electron microscopy (STEM) have developed over several decades into a huge field with wide ranging areas of application in both the physical and life sciences. Although many of the basic instrumental requirements are common to both areas of application, there are important differences in sample preparation methods, the sensitivity to irradiation by high-energy electrons, the contrast mechanisms, and the dimensions of features of interest. The contents of this chapter and of Chapter 10 describe TEM and STEM methods for the characterization of boundaries and interfaces only in inorganic materials. Even so, the topic is still so vast that, in the space available, only a general overview is possible.

Several comprehensive texts [1–7] covering EM are listed in the references, should the reader wish to delve more deeply into the subject. The topic has been divided into two chapters. This chapter describes instrumentation, basic theory, and techniques. Such aspects underlie all types of analysis and their description is essential before considering specific interfacial applications, examples of which are covered in Chapter 9.

Although this chapter includes current developments and recent results, it also covers the more established characterization methods found on both new and relatively old instruments, because the latter are still widely used in materials science. This is because many areas of materials science still rely on simple diffraction contrast (DC) imaging, selected area diffraction (SAD), and energy dispersive x-ray analysis (EDX) rather than more demanding methods such as high resolution electron microscopy (HREM) imaging, electron energy-loss spectroscopy (EELS), and high-angle annular dark-field (HA-ADF) imaging. The most modern instruments now incorporate aberration correction, monochromators, holography, energy-filtered imaging, and HA-ADF, but these advances are not yet widely available to many materials scientists. However, these methods will become increasingly available in the coming years, and so are included in this chapter in anticipation that the content will remain or become more useful to materials scientists in the future.

An internal interface in a sample can be investigated by TEM provided that an electron beam can be transmitted through the sample and collected with sufficient intensity and spatial resolution by an imaging system. Because they are charged particles, electrons interact strongly with matter compared to other particle beams such as x-rays or neutrons that are also employed for internal structural studies. The disadvantage of this is that the strong interaction causes significant scattering out of the incident electron-beam direction in even the thinnest samples and the scattering increases with atomic number. Penetration by an electron beam is, therefore, much smaller than for an energetic x-ray or neutron beam. TEM will generally be useful when the sample thickness \( t < 150 \) nm. For the highest spatial resolution, much thinner samples are necessary. As the electron energy is increased, penetration improves but the atoms within the sample are then increasingly likely to be displaced from their positions and even ejected from the sample. Such processes interfere with the integrity of both imaging and analytical measurements. Hence, although the accelerating voltage, \( E_o \), of TEMs was increased from <100 keV to 1 MeV, or more, in the 1960s, the vast majority of
Transmission Electron Microscopy

instruments are now designed to operate with $E_o = 200–300 \text{keV}$, because this range is considered to be the optimum compromise between TEM size and the cost of increasing $E_o$, useful penetration, spatial resolution, sample damage, and analytical sensitivity.

Abbreviations and acronyms are used extensively throughout the chapter in the interest of economy of space. Definitions are summarized below as an aid to the reader.

### 6.1.1 List of Symbols (Units), Acronyms, and Abbreviations

- $a_i^a$ partition function for $i$th shell of element $a$
- $A_i^a$ atomic weight of element $a$
- $A(\alpha)$ focused probe aperture function
- $b$ dislocation Burger’s vector (m)
- $B$ electron source brightness (current emitted per unit area per steradian)
- $c_{ab, ...}$ fractional weight concentrations of elements $a$, $b$, ...
- $C_c$ chromatic aberration coefficient (m)
- $C_s$ spherical aberration coefficient (m)
- $d$ crossover diameter (m)
- $d_{hkl}$ interplanar spacing (m)
- $e$ electronic charge (C)
- $E_o$ accelerating voltage/energy of the electron beam (eV)
- $E_i$ temporal coherence envelope function
- $E_B$ continuum x-ray energy (eV)
- $E_x$ characteristic x-ray energy (eV)
- $E_i^a$ energy of $i$th shell x-ray from element $a$ (eV)
- $F(K)$ Fourier transform of crystal potential
- $F_g$ structure factor for diffraction vector $g$ (m)
- $g_{hkl}$ reciprocal lattice diffraction vector (m$^{-1}$)
- $h$ Planck’s constant
- $i, I_o$ probe current (A)
- $I_B, I_D$ bright field, dark field image intensity
- $I_i^a$ number of x-rays from $i$th shell of element $a$
- $I_i(r)$ probe intensity at position $r$
- $k$ Boltzmann’s constant (J/K)
- $k_{ab}$ factor scaling ratio of x-ray counts to weight ratio of elements $a$ and $b$
- $k$ wave vector (m$^{-1}$)
- $K$ wave vector (m$^{-1}$)
- $k_o$ incident beam wave vector (m$^{-1}$)
- $k'$ diffracted beam wave vector (m$^{-1}$)
- $L$ effective camera length of diffraction pattern
- $m$ electron mass (g)
- $n$ integer index
- $N_A$ Avagadro’s number
- $P_o$ probability of electron suffering no inelastic scattering
- $q_o$ limit of illumination angle at specimen (rad)
- $r$ real space position vector (m)
- $r_c$ radius of coherent illumination at specimen (m)
- $s$ deviation vector (m$^{-1}$)
- $S_{i}^A(W, \beta)$ energy loss signal intensity integrated over window width, $W$, and collection angle, $\beta$, above the $i$th ionization edge of element $A$
- $t$ sample thickness (m)
- $T$ temperature (K)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_o$</td>
<td>beam accelerating voltage (V)</td>
</tr>
<tr>
<td>$V_c$</td>
<td>volume of crystal unit cell (m$^3$)</td>
</tr>
<tr>
<td>$z$</td>
<td>coordinate of the optic axis (m)</td>
</tr>
<tr>
<td>$Z$</td>
<td>atomic number</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>semi-angle of probe convergence (rad)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>scattering angle collected by aperture or detector (rad)</td>
</tr>
<tr>
<td>$\chi$</td>
<td>path difference (m): $2\pi\chi = \text{phase angle (rad)}$</td>
</tr>
<tr>
<td>$\gamma(\alpha)$</td>
<td>aberration function</td>
</tr>
<tr>
<td>$\Delta f$</td>
<td>probe defocus distance (m)</td>
</tr>
<tr>
<td>$\Delta E$</td>
<td>energy range (window) in EELS spectrum (eV)</td>
</tr>
<tr>
<td>$\varepsilon(E)$</td>
<td>x-ray detector efficiency for x-ray energy E</td>
</tr>
<tr>
<td>$\phi$</td>
<td>azimuthal angle (rad)</td>
</tr>
<tr>
<td>$\phi_o$</td>
<td>mean inner potential inside sample (eV)</td>
</tr>
<tr>
<td>$\phi(r)$</td>
<td>crystal potential distribution (eV)</td>
</tr>
<tr>
<td>$\phi_p(r)$</td>
<td>projected crystal potential distribution normal to beam (eV)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>electron wavelength (m)</td>
</tr>
<tr>
<td>$\lambda_{\text{total}}$</td>
<td>total inelastic scattering mean free path (m)</td>
</tr>
<tr>
<td>$(\mu/\rho)_{ij}$</td>
<td>mass absorption coefficient of x-ray from element i by atomic mixture j (kg/m$^2$)</td>
</tr>
<tr>
<td>$\Psi(r)$</td>
<td>electron wave function at a point, $r$, in the x–y plane</td>
</tr>
<tr>
<td>$\rho^a$</td>
<td>density of sample of element a (kg/m$^3$)</td>
</tr>
<tr>
<td>$\sigma_K$</td>
<td>K-shell ionisation cross-section (m$^2$)</td>
</tr>
<tr>
<td>$\sigma_i^a$</td>
<td>ionization cross-section for $i$th shell of element a (m$^2$)</td>
</tr>
<tr>
<td>$\tau$</td>
<td>analysis time (s)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>diffraction angle (rad)</td>
</tr>
<tr>
<td>$\theta_{B^i}, \theta_{B^i}$</td>
<td>Bragg diffraction angle for $(hkl)$ planes (rad)</td>
</tr>
<tr>
<td>$\theta_i, \theta_o$</td>
<td>inner and outer angular limits of HA-ADF detector (rad)</td>
</tr>
<tr>
<td>$\omega$</td>
<td>angular frequency (rad)</td>
</tr>
<tr>
<td>$\omega_i^f, \omega_K$</td>
<td>fluorescent yields for $i$th, $K$th shells</td>
</tr>
<tr>
<td>$\xi_g$</td>
<td>extinction distance of diffraction vector $g$ (m)</td>
</tr>
<tr>
<td>$\xi_g^{\text{eff}}$</td>
<td>effective extinction distance of diffraction vector $g$ (m)</td>
</tr>
<tr>
<td>$\psi$</td>
<td>electron wave function</td>
</tr>
<tr>
<td>$\psi_i(r), \psi_e(r)$</td>
<td>incident, exit surface wave functions at position $r$</td>
</tr>
<tr>
<td>$\psi_i(r)$</td>
<td>probe wave functions at position $r$</td>
</tr>
<tr>
<td>$\psi_i, \psi_i^*$</td>
<td>image wave functions and its complex conjugate</td>
</tr>
<tr>
<td>ADF</td>
<td>annular dark field</td>
</tr>
<tr>
<td>BF</td>
<td>bright field (image)</td>
</tr>
<tr>
<td>BFP</td>
<td>back focal plane of the objective lens</td>
</tr>
<tr>
<td>CA</td>
<td>contrast aperture</td>
</tr>
<tr>
<td>CC</td>
<td>camera constant</td>
</tr>
<tr>
<td>CCD</td>
<td>charge-coupled device</td>
</tr>
<tr>
<td>CL</td>
<td>condenser lens</td>
</tr>
<tr>
<td>CTEM</td>
<td>convergent beam TEM</td>
</tr>
<tr>
<td>CTF</td>
<td>contrast transfer function</td>
</tr>
<tr>
<td>DC</td>
<td>diffraction contrast (image)</td>
</tr>
<tr>
<td>DF</td>
<td>dark field (image)</td>
</tr>
<tr>
<td>DOS</td>
<td>density of states</td>
</tr>
<tr>
<td>DP</td>
<td>electron diffraction pattern</td>
</tr>
<tr>
<td>EDX</td>
<td>energy dispersive x-ray analysis</td>
</tr>
<tr>
<td>EELS</td>
<td>electron energy-loss spectrometer</td>
</tr>
<tr>
<td>EFTEM</td>
<td>energy filtered TEM</td>
</tr>
<tr>
<td>EHT</td>
<td>electrostatic high tension</td>
</tr>
<tr>
<td>ELNES</td>
<td>energy-loss near-edge structure</td>
</tr>
</tbody>
</table>

© 2009 by Taylor & Francis Group, LLC
Transmission Electron Microscopy

EPMA electron probe microanalyser
ESI electron spectrum image
FEG field-emission gun
FT Fourier transform
FWHM full-width half-maximum
HA-ADF high-angle annular dark-field detector
HREM high-resolution electron microscopy
IL intermediate lens
IP image plane
MAC mass absorption coefficient
OA objective aperture
OL objective lens
PL projector lens
SAD selected area diffraction
SADA selected area diffraction aperture
SEM scanning electron microscope
SNR signal-to-noise ratio
SSD single scattering distribution
STEM scanning transmission electron microscopy
TDS thermal diffuse scattering
TEM transmission electron microscopy
WB weak beam
WPOA weak-phase object approximation
ZLP zero loss peak

6.2 ELECTRON MICROSCOPE INSTRUMENTATION

6.2.1 CONVENTIONAL TEM

The TEM system is pumped to a low pressure (the 10^{-5} Pa range or lower in the case of FEG [field-emission gun]) to ensure stable electron emission from the source, to minimize scattering of the electron beam between source and detectors, to prevent ion damage in the sample and source, and to minimize contamination deposits on the sample. It follows that great care must always be undertaken during maintenance or repair to use clean and dry gas whenever the system is vented to atmospheric pressure, and any components introduced into the evacuated system should always be degreased, dried, and never handled without gloves. Figure 6.1 shows, schematically, the essential electron-optical components of a TEM instrument used for DC and HREM imaging, which can be found in many materials science laboratories. Different sections of the column are often differentially pumped because the electron gun and sample stage sections demand lower pressure than other regions. The maximum tolerable gun pressure depends on the type of electron emitter (described in the following section). All the lenses are electromagnetic, except for those of the electron gun, which are electrostatic, and have axial symmetry. Deflection coils, necessary for accurate alignment of the electron optical column, and astigmatism correctors are excluded from Figure 6.1.

The electron gun provides a source of electrons at the point S1 (Figure 6.1), which is focused onto the specimen by two or more condenser lenses (CLs) plus an objective mini-lens. The sample is positioned on the optical axis (OA) in the sample stage inside the strong magnetic field of the objective lens (OL). The sample holder, of which there are various designs for different types of experiment, is inserted into the stage through an airlock. Samples can, therefore, be exchanged without venting the column to atmospheric pressure. The OL produces a magnified image of part of the sample and a series of intermediate lens (IL) and projector lens (PL) further magnify the image and transfer it to viewing chamber for assessment and recording.
6.2.2 ELECTRON SOURCES

Various types of electron source are used and all employ electrostatic extraction and acceleration. The source is held at a high negative-potential, $E_o$, with respect to the grounded column and sample. An important characteristic of the source is its brightness, $B$. The thermionic tungsten hairpin filament is still used in older TEMs, but more recent instruments use a heated, pointed, single crystal of LaB$_6$ [8] as the electron emitter because it is brighter and has a much longer lifetime than the W hairpin. Increasingly, the Schottky FEG [9,10] is found in the most modern high-performance instruments. The FEG depends on electron tunneling from a single crystal of oriented W wire, etched to a fine point (∼50 nm radius at the apex) and coated with an appropriate material to lower the potential barrier to electron emission. The Schottky source is brighter than the simple thermal sources but it still needs to be heated to keep the tip clean and to give thermal assistance to the tunneling process. The brightest source is the cold FEG that operates at ambient temperature [11]. However, FEGs require a much lower operational pressure (of order $10^{-9}$ Pa) than simple thermal guns ($10^{-5}$ Pa) and, consequently, FEG-TEMs are considerably more expensive. The electron gun, whatever the type, generates an effective source focused at a crossover, S1, on the optical axis, which forms the object for the condensing lens system.

The brightness, $B$, is defined as the current emitted per unit area per steradian:

$$ B = \frac{I}{A \theta} $$
Transmission Electron Microscopy

\[ B = \frac{4i}{\pi^{2}d^{2}\alpha^{2}} \]  \hspace{1cm} (6.1)

where

- \( i \) is the probe current
- \( d \) is the effective crossover diameter
- \( \alpha \) is the semi-angle of probe convergence

Provided aberrations are negligible, \( B \) is constant when measured at any focused beam crossover in the TEM column. Table 6.1 lists typical values of operating parameters for the sources discussed above. The FEG is a factor of about \( 10^{3} \) brighter than a thermal source. Compared to a thermal source, the spatial coherence of the FEG beam is much greater at the specimen for a given current density and area of illumination. Coherence is important for HREM imaging and is discussed in Section 3.4. High \( B \) is equally important for analytical work because the current in a very small focused probe is much greater for an FEG. The brightness of any source tends to fall gradually during its operational lifetime.

### 6.2.3 ILLUMINATION SYSTEM

The effective source position (S1 in Figure 6.1) and the sample height are fixed in the column, so the CLs can transfer either a defocused image of S1 (i.e., an approximately parallel beam) to the sample plane, for TEM imaging and diffraction, or a focused probe at the sample, for convergent beam diffraction and spot analysis using EDX or EELS. The three CLs allow, in principle, independent control of the three beam parameters, \( i, d, \) and \( \alpha \). However, the condenser mini-lens is often coupled with CL2, rather than operated in a completely independent manner for TEM imaging, and so independent control of all three parameters is not always readily available to the user. The beam size at the sample is controlled by changing the strength of CL1. However, this requires corresponding adjustment to CL2 to maintain focus at the sample and, from Equation 6.1, \( \alpha \) will increase. In practice, an interchangeable aperture in CL2 is used to adjust \( \alpha \), so that changing CL1 with CL2 results in \( d \) and \( i \) changing while \( \alpha \) remains constant. Instrument manufacturers usually provide a number of fixed settings for CL1. However, when a free lens control facility is available, it is possible to adjust CL settings to allow continuous adjustment of \( \alpha \) and \( d \), which may have advantages for focused beam diffraction experiments.

The CL2 aperture can be regarded as the effective source of electrons and such a source can be considered to be completely incoherent in the sense that individual electrons are seen by the sample as originating at distinct points within the CL2 aperture at different times.

### 6.2.4 SAMPLE STAGE

In older TEMs, the sample holders were of the top-entry type, whereby the sample was inserted into the column through an airlock above the OL and then lowered through the OL upper pole piece. This

---

**TABLE 6.1**

**Comparison of Electron Source Parameters**

<table>
<thead>
<tr>
<th>Type</th>
<th>Operating Temperature (K)</th>
<th>Operating Pressure (Pa)</th>
<th>Maximum Brightness (A/mm²/sr)</th>
<th>Life (h)</th>
<th>Energy Spread (FWHM eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal W</td>
<td>2500–2800</td>
<td>( 10^{-3} )</td>
<td>( 10^{3} )</td>
<td>25</td>
<td>1–2</td>
</tr>
<tr>
<td>Thermal LaB₆</td>
<td>1800</td>
<td>( 10^{-5} )</td>
<td>( 10^{4} )</td>
<td>500</td>
<td>1–2</td>
</tr>
<tr>
<td>Schottky</td>
<td>1000–1500</td>
<td>( 10^{-4} )</td>
<td>( 10^{6} )</td>
<td>10⁴</td>
<td>0.8–1.0</td>
</tr>
<tr>
<td>Cold FEG</td>
<td>300</td>
<td>( 10^{-9} )</td>
<td>( 10^{7} )</td>
<td>10⁴</td>
<td>0.3–0.5</td>
</tr>
</tbody>
</table>

© 2009 by Taylor & Francis Group, LLC
gave improved positional stability for HR imaging compared with the side-entry stages found on lower resolution, analytical TEMs. However, in modern systems, improved stage design for side-entry holders now provides equivalent performance without the mechanical complexity of top-entry. The sample is inserted horizontally, directly between the pole pieces of the OL. Sample tilting is possible about two axes that are approximately orthogonal over a typical range up to ±45° from the horizontal position, which allows considerable flexibility for examining crystal features under a wide range of orientations and diffraction conditions. TEM sample dimensions must usually be no more than 3 mm in diameter and ~0.2 mm thick. It is important to remember that, in any TEM image, the structural features appear as a 2D (two-dimensional) projection of the real 3D structure and that greater insight can be gained by tilting to several orientations. There is currently much interest in the development of tomography by combining series of images acquired with tilt increments [12,13], leading to the development of stages that permit tilting up to ±80°. An HREM will probably have a reduced tilting range, typically ±25°, because the pole-piece gap is lower than for a general purpose TEM, in order to reduce resolution-limiting aberrations of the OL. Special holders can be purchased for in-situ studies of heating, cooling, straining, magnetization, chemical reactions, etc. [6], and these often require extensive changes to the stage design.

6.2.5 OBJECTIVE LENS

The OL is the most important imaging lens in the TEM and is designed to minimize the spherical and chromatic aberration coefficients that limit the spatial resolution. Aberrations are discussed in Section 3.4.3.

Image and diffraction pattern formation in the OL are shown, schematically, in Figure 6.2 using the concept of ray optics. Although the sample is, in practice, immersed in the field of the OL, it is still sufficiently accurate for many purposes to regard the OL as thin, with the object (sample) outside the lens and close to the principal focus of the lens. It is assumed in Figure 6.2 that all rays incident on the sample are parallel to OA, i.e., there is an incident plane wave, and those which are

![Figure 6.2](https://example.com/figure6.2.png)

**FIGURE 6.2** Formation of the image A"B" in the IP and the DP in the diffraction plane (BFP) by the OL.
transmitted through the sample but not deflected, such as $AA'$ and $BB'$, are brought to a focus, $O$, in the back focal plane (BFP) of the OL on OA. All rays that are diffracted through a particular angle, $\theta$, are brought to a common, off-axis point, $P$, in the BFP. The intensity distribution in the BFP is the diffraction pattern (DP). Continuation of the pairs of rays $A'O$ and $A'P$ and $B'O$ and $B'P$ result in intersections at points $A''$ and $B''$, respectively, in the image plane (IP) of the OL. Points $A$ and $B$ in the sample are imaged at $A''$ and $B''$, respectively, in IP. Every point in the sample contributes intensity to every part of the DP and every point in the DP subsequently contributes to the image intensity in the IP. When the sample is very thin and crystalline, the DP intensity is concentrated into an array of spots, with relatively low intensity between the spots. What cannot readily be shown in Figure 6.2 is that there is a rotation about OA between corresponding directions in the image and diffraction planes.

Close to the BFP, the operator can insert and position an objective aperture (contrast aperture [CA]), which is essential for performing DC imaging. Without the insertion of a CA, the final image is a superposition of all the diffracted images from each point in the BFP. When the CA is inserted and positioned on OA to allow only the directly transmitted spot to pass through, the resulting image is called the bright-field (BF) image. (The term “undiffracted” beam is sometimes used as a convenience to describe the directly transmitted beam. However, diffraction also occurs in the directly transmitted beam direction along OA and so it is more accurate to call the undiffracted spot the forward-diffracted spot.) A dark-field (DF) image is formed by positioning the CA around a single off-axis region of the DP. For a crystalline sample, the aperture is usually centered round a single diffraction spot. A tilted DF image is obtained by tilting the incident beam through an angle such that the selected DF spot is moved onto the OA to reduce the blurring caused by OL aberrations. If the diameter of the CA is large relative to the separation of diffraction spots in the BFP, or no CA is inserted, then many DF images, together with the BF image, are superposed in the image plane. An HREM image is formed in this manner. HREM image formation is discussed in Section 3.4.

A retractable SAD aperture (SADA) can be inserted in the IP (see Figure 6.1). By definition, the IP is optically conjugate with the sample plane so that the SADA appears sharply focused in the image. The SADA can be accurately positioned, usually in the center of the image, and any sample region of interest can be moved until it appears within the SADA. The insertion of the SADA, therefore, is equivalent to placing an area-selecting aperture on the sample itself. Figure 6.2 shows how the area of sample, diameter AB, contributing to the DP, is restricted to the area of sample apparently defined by the SADA, diameter $A''B''$. Hence, the SADA can be used to collect diffraction information from, for example, a relatively small particle contained in a matrix, or to determine the local orientation in a bent crystal.

### 6.2.6 Projection and Recording

The excitation of lenses IL and PL (Figure 6.1), which follow the OL, can be changed so that a magnified image of either the BFP or the IP of the OL can be transferred to the fluorescent viewing screen (SC), which is viewed through a glass window. The operator can toggle between the two modes using a single button and the magnification of either image or DP can be separately changed to suit the experiment.

Recording of images was, until recently, performed using photographic emulsion on cellulose plates in a camera located below the retractable SC (Figure 6.1). However, it is now commonplace to use a solid-state charge-coupled device (CCD) camera, the output from which can be read directly into a computer memory, from which real-time viewing and processing of images and DPs can be performed to assess resolution and aberrations [14]. The only limit on the number of recorded images is the capacity of the computer hard drive. This is a great advantage for HREM studies, where large numbers of exposures are often required for through-focal series. CCD cameras with up to $4k \times 4k$ pixel arrays are now becoming affordable, providing a pixel density close to that of photographic emulsion.
Although the CCD camera is better than emulsion, it still lacks the sufficient dynamic range necessary for some measurements. Image plates provide a superior recording medium in this respect [15,16]. The image plate is exposed in the same manner as photographic emulsion such that the electron intensity is stored as a charge distribution across the plate. The plate must be removed from the microscope before the intensity distribution can be read and stored in a computer. Hence, image plates suffer from the same slow pre-pumping and exchange of plate as conventional photographic emulsion. The dynamic range of the image plate, however, is several orders of magnitude greater than that of a CCD camera and this makes image plates suitable for certain specialist applications, for example, for measuring quantitative data from DPs where the intensity at the center of bright spots is often many orders of magnitude greater than the background.

6.3 ELECTRON–SAMPLE INTERACTION

A high-energy electron travels as an isolated wave packet with a wavelength, $\lambda$, much smaller than the typical interatomic distance and is determined by the electron energy:

$$\lambda = \frac{1.226}{E_o^{1/2} (1 + 0.97845 \times 10^{-6} E_o)^{1/2}}$$

(6.2)

where

- $\lambda$ is in nanometers
- $E_o$ in volts

For example, $\lambda = 2.51$ pm for an electron accelerated through 200 keV. A plane wave is represented as

$$\psi = \exp [i(\omega t - 2\pi k \cdot r)]$$

(6.3)

traveling in the direction of $k$, where $r$ is the (vector) distance from a chosen origin and $|k| = 1/\lambda$, although the time varying factor, $\omega t$, where $\omega$ is the angular frequency, is usually omitted for convenience. (It is worth noting that some texts use the definition of $|k| = 2\pi/\lambda$, so Equation 6.2 must then be modified appropriately.) The physical extent, typically tens of nanometers, of the wave packet in the direction, $z$, parallel to OA, is usually much smaller than the average separation of electrons, typically tens of micrometers, in beams used for both HREM and analysis. The manner in which the electron beam interacts with the sample can be described either as the propagation of a continuous plane or as a spherically converging monochromatic wave, even though the electron wave-packets are discrete, or as a stream of energetic particles (e.g., Ref. [17]). The former is more convenient when discussing image and DP formation while the latter is often more useful when considering analytical modes such as EDX and EELS.

6.3.1 DIFFRACTION PATTERN

When a sample is placed in the electron beam, the fast electron wave-field interacts with the electric field of the atomic nuclei and their surrounding electrons. The wave function $\psi_e$, of the fast electron that emerges at the exit surface of the sample, is no longer planar (or spherically converging, depending on the incident illumination condition), and this causes a variation of amplitude with direction of propagation. The angular distribution of the wave function, viewed in a plane at a large distance, $L$, the camera length, from the sample, is called the Fraunhofer electron DP. In the TEM, the DP is observed by magnifying the electron distribution in the BFP of the OL at the detector. In Figure 6.2, all rays emerging from the sample at a common angle to OA are focused by the OL to a single point in the BFP. For an amorphous sample, there will be a continuous distribution of intensity in the BFP. However, the DP is
most conveniently described by considering a section through a simple crystalline sample with just one set of diffracting planes of spacing \( d_{hkl} \), tilted at the Bragg angle, \( \theta_{hkl} \), to the incident beam (Figure 6.3). The phase of \( \psi_e \) then varies periodically in the \( x-y \) plane, and constructive interference occurs in certain directions, \( 2\theta_{hkl} \) from the incident direction, that are derived using the well-known Bragg equation:

\[
\lambda = 2d_{hkl}\sin\theta_{hkl}
\]  

(6.4)

A bright spot is observed in the BFP for every set of lattice planes with a normal lying close to the \( x-y \) plane (i.e., normal to OA) at a location where the path difference between diffracted intensity from adjacent planes is a whole number of wavelengths (shown as \( n\lambda \) in Figure 6.3). The symmetry of the spots in the DP will reflect the symmetry of the projected crystal lattice in the plane perpendicular to OA. An extra constraint on the formation of any diffraction spot is that the structure factor [1] of the sample unit cell must be nonzero, otherwise systematic absences occur. For certain crystals, however, double diffraction by two different sets of planes can lead to additional spots that are kinematically forbidden.

The Ewald sphere construction (Figure 6.4) provides a convenient, geometric method of illustrating, in reciprocal space, the relationship between the incident beam direction, with a wave vector \( k_o \), where \( |k_o| = 1/\lambda \), and a diffracted beam with wave vector \( k' \), also with \( |k'| = 1/\lambda \). The vector \( k_o \) is

![Ewald sphere construction](image-url)
drawn from a point, A, in the direction AO, parallel to the incident electron wave, to the point O. A sphere is drawn with radius AO = 1/λ, centered on A, part of which is shown as a circular arc in the 2D section of reciprocal space in Figure 6.4a. The diffraction vector \( \mathbf{OP} = \mathbf{k} - \mathbf{k}_o \) has a magnitude \( |\mathbf{k}| = 2 |\mathbf{k}_o| \sin \theta \).

The locations of possible Bragg reflections are represented by points, located at the end of vectors, \( \mathbf{g}_{hkl} \), drawn from O and that form a 3D lattice in reciprocal space. As the sample is tilted, the array of spots rotates with respect to the direction of \( \mathbf{k}_o \) so that the surface of the sphere sweeps through the array of reciprocal lattice points. At specific tilt angles of the beam, the sphere will pass exactly through a reciprocal lattice point when \( \mathbf{k}' = k'_o \), as shown in Figure 6.4a for the point \( \mathbf{g}_{hkl} \), where \( |\mathbf{g}_{hkl}| = 1/d_{hkl} \), signifying that the \((hkl)\) planes satisfy the condition for Bragg diffraction, Equation 6.3, such that \( \mathbf{K} = \mathbf{g}_{hkl} \) and the angle between \( \mathbf{k}' \) and \( \mathbf{k}_o \) is \( 2\theta_{hkl} \). Other reciprocal lattice points that do not lie on the sphere are not seen in the DP.

A more accurate construction of the reciprocal lattice requires that the points are elongated into rods, called relrods, in a direction normal to the top surface of the sample (i.e., the entry surface of the electron beam), and by a length that is inversely proportional to the thickness of the sample. Hence, a very thin sample has very long relrods. The implication of this is that when the Ewald sphere intersects a relrod, a spot is generated in the DP, even though there is a deviation from the precise Bragg orientation by the deviation vector, \( \mathbf{s} \), parallel to \( \mathbf{k}_o \), as shown in Figure 6.4b, so that \( \mathbf{k}_g = \mathbf{k}_o + \mathbf{g}_{hkl} + \mathbf{s} \). In Figure 6.4b, \( \mathbf{g}_{hkl} \) lies inside the Ewald sphere and \( \mathbf{s} > 0 \). For the values of \( \lambda \) used in TEM, the Ewald sphere is almost planar and intersects many relrods, so many spots are then visible in a DP from a thin EM sample.

A DP from a thin crystal of Inconel 600, a face-centered cubic metal alloy, oriented with a <110> zone axis almost exactly parallel to the electron beam, is shown in Figure 6.5. Several dozen spots are seen due to the extension of the relrods and the curvature of the Ewald sphere. The spots in the pattern have been indexed using a right-handed axis notation and assuming that the outward direction from the page is [109], i.e., anti-parallel to the beam.

More rigorous treatment of electron diffraction in standard texts [18] shows that the DP is given by the Fourier transform (FT) of the 3D distribution of the sample potential. Because the potential generally consists of a distribution of spatial-frequency components, \( \phi(r) \), related to the distribution of nuclear and electronic charges in the sample, the DP intensity \( F(K) \) at some point, \( \mathbf{K} \), in the DP is given by

\[
F(K) = C \int [\phi(r)\exp(-2\pi i\mathbf{K} \cdot \mathbf{r})] d\mathbf{r}
\]

where \( C \) is a constant.

The DP can be used to determine \( d_{hkl} \) if \( L \) for the TEM is known. The magnification of the DP and, hence, \( L \) can be changed by altering the setting of the projector lenses, in a manner analogous to changing the image magnification. Typical values of \( L \) are in the range 50–1000 mm. Equations

![Figure 6.5](image-url)
6.6 and 6.7 and Figure 6.6 show how the modulus, \( m \), of \( g \), measured from the recorded DP, is related to \( L \) and to its Bragg angle (in the small angle approximation):

\[
m \approx 2\theta L \quad (6.6)
\]

Combining Equations 6.3 and 6.5 gives

\[
md_{hkl} \approx \lambda L \quad (6.7)
\]

The product \( \lambda L \) is called the camera constant (CC) and, for quantitative measurements, must first be accurately calibrated at each magnification of the DP for the appropriate \( E_0 \) (i.e., \( \lambda \)) using a sample of known crystal structure. The \( d_{hkl} \) spacings of an unknown structure are then easily calculated from measurements of \( m \) and Equation 6.7. In using Equation 6.7, it is important to ensure that the dimensions are correct: if \( L \) and \( m \) are measured in millimeters and \( l \) in nanometers, then the units of \( d_{hkl} \) will be in nanometers.

Unless the sample is very thin, the DP also contains intensity between the spots arising from inelastic phonon scattering (thermal diffuse scattering, TDS) whereby electrons are scattered over all angles with very small energy losses, proportional to \( kT \) (\( \sim 0.025 \text{ eV at ambient temperature} \)). Low intensity is observable in Figure 6.5 between the spots. In thicker samples, when electrons undergo both phonon and diffraction scattering, a pattern of lines, known as the Kikuchi pattern [19], appears in the DP, an example of which is shown in Figure 6.7 from a thicker region of the same sample of Inconel as in Figure 6.5, but at a random incident-beam orientation away from a zone axis. The Kikuchi lines appear in pairs, one brighter (excess, E), and the other darker (deficient, D), than the background intensity, with the angular separation \( 2\theta_{hkl} \) of the particular set of \( (hkl) \) planes giving rise to the contrast and a line direction perpendicular to \( g_{hkl} \). The excess line of any pair is always the further from the central BF spot in the DP. When the incident beam is exactly parallel to a set of planes, then the two lines have equal intensity and form a Kikuchi band with the central spot lying between and equidistant from both lines. As with diffraction spots, additional higher order Kikuchi lines are often seen lying parallel to, and at an angular distance of \( \theta_{hkl} \), outside, the principal lines.

The Kikuchi pattern is extremely useful because, like the DP spots, it displays the same symmetry as the sample. Major zone axes are defined by the intersection of two or more bands. As the sample is tilted, the DP spots remain fixed in position but change in intensity. When large changes in tilt are made, some spots will disappear while others appear at new positions when a new zone axis, with new \( g_{hkl} \) vectors, comes close to OA. However, while the DP is changing, the Kikuchi pattern
intensity stays roughly constant and the lines move across the DP, as the sample is tilted, as though fixed to the sample. The Kikuchi pattern, therefore, can be regarded as a road atlas of the crystal orientation and is used to tilt the sample to specific zone-axis orientations or to select specific diffraction conditions for defect analysis, with an accuracy as high as 0.05 mrad.

The clarity of the Kikuchi pattern is degraded when the sample is bent or contains crystal defects. When such degradation occurs when using a SADA, it is better to remove the SADA and change the illumination to a convergent probe. The area of sample generating the DP is then greatly reduced, as is the range of bending, and it is then often possible to select a defect-free region.

### 6.3.2 Diffraction Contrast Imaging

DC imaging is considered to be a low-resolution imaging mode because the contrast arises from changes on a scale significantly larger than typical interatomic distances. The name HREM imaging is used to denote the imaging of the atomic structure in a crystalline sample and is considered separately in Section 3.4. While DC imaging can be used in thin samples under the so-called kinematic conditions where the undiffracted spot is much brighter than any diffracted beam, the lack of a Kikuchi pattern impedes accurate sample tilting. DC is more often applied in thicker samples than those used for HREM. Dynamic scattering occurs in thicker crystals such that the intensity in the diffracted beams is as strong, or stronger, than in the undiffracted beam, and where the Kikuchi pattern is visible. Scattering then occurs between the various diffracted beams because each becomes an effective illumination direction within the sample.

The most common type of DC image is the two-beam BF image, briefly mentioned in Section 2.5. Figure 6.3 shows how diffraction occurs from a single set of planes oriented exactly at the Bragg condition \((s = 0, \text{Figure 6.4a})\), i.e., when the diffracting planes are inclined at exactly the Bragg angle to the incident beam. At such an orientation and when no other set of planes lies close to its Bragg diffracting orientation, then the DP contains only two strong spots: the undiffracted spot and the spot from the diffracting planes. Figure 6.8a shows such a DP, where only one strong spot, \(g_{200}\), is excited. The sample was too thin in this area to generate visible Kikuchi lines. The positions of the much weaker spots, present because of their extended relrods, still show the symmetry of a \(<110>\) zone axis, tilted a few degrees away from the incident beam direction around the \(<100>\) direction parallel to \(g_{200}\). The BF and DF images in Figure 6.8b and c are formed by placing the CA around the lower...
Transmission Electron Microscopy

153

and upper spots, respectively, in Figure 6.8a. Note that the difference in the direction of \( g_{200} \) arrowed in Figure 6.8a and b is due to the rotation between image and DP mentioned in Section 2.5.

Under two-beam conditions, a perfect lattice of uniform sample thickness, \( t \), will generate a BF or DF image of uniform intensity. If \( t \) varies, however, the intensities in the BF or DF images will increase or decrease in a complementary manner as a function of \( t \). A wedge sample, therefore, will show intensity contours in both images [20,21] as in Figure 6.8. When the crystal is bent, however, the symmetric two-beam condition shown in Figure 6.3 occurs only along a line in the image and the intensity will then vary in an oscillatory manner on each side of this line to generate a bend contour. In practice, the two-beam BF image is often recorded with \( s > 0 \). This reduces the contrast from thickness and bending variations, thereby improving the visibility of microstructural features.

When a lattice defect such as a dislocation is present in the crystal, then some of the lattice planes in a small region around the core are distorted. Under two-beam conditions, a local change of intensity will be seen along the line of the dislocation core, projected onto the image plane, whenever a distorted set of lattice planes is diffracting. The magnitude of the defect contrast can be calculated using numerical calculations, the details of which can be found in standard texts [1,4]. The analysis of the dislocation Burgers vector, \( b \), is performed by systematically recording a series of BF images, using a range of different \( g \)-vectors, to find those in which the dislocation is rendered invisible, i.e., to identify lattice planes that are not distorted by the defect. Invisibility requires that the vector product \( g \cdot b = 0 \) so, by finding three such noncoplanar \( g \)-vectors, a unique set of indices can be derived for \( b \). In practice, it is found that a residual image can still be present when \( g \cdot b = 0 \) and then it is necessary to compare experimental images with calculated images [21]. This is especially so when analyzing the dislocations present in a grain boundary or interface [22] and specific examples of this are described in Chapter 9. An example of dislocation invisibility is shown in Figure 6.9 for a face-centered cubic nickel alloy. In Figure 6.9a, the dislocation line images labeled AA’ have strong contrast when \( g = 220 \), whereas only residual contrast remains in Figure 6.9b for which \( g = 113 \), when \( b = a/2(1,−1,0) \) and \( g \cdot b = 0 \).

6.3.3 Weak-Beam Imaging

The width of a dislocation image in a two-beam BF or DF image is approximately \( \xi_g / 3 \), where the extinction distance \( \xi_g \), Equation 6.8, is a characteristic thickness difference between two adjacent intensity thickness contours in the BF or DF image at the exact Bragg orientation:
\[ \xi_g = \pi V_c \cos \theta_B / (\lambda F_g) \]  

(6.8)

where

- \( V_c \) is the volume of the crystal unit cell
- \( \theta_B \) is the Bragg angle
- \( F_g \) is the structure factor for \( g \)

Calculation of \( \xi_g \) using atomic scattering amplitudes [23] at \( E_0 = 100 \) keV yields values in the range 15–200 nm, increasing as \(|g| \) increases and atomic number, \( Z \), decreases, so that a dislocation will have an image line width of approximately 5–50 nm, severely limiting individual defect discrimination in many materials. However, \( \xi_g \) decreases with increasing \(|s| \), such that the effective value, \( \xi_g^{\text{eff}} \), is

\[ \xi_g^{\text{eff}} = \xi_g / (1 + |s|^2 \xi_g^2)^{1/2} \]  

(6.9)

This is exploited when weak-beam (WB) DF imaging is employed [24,25], the diffraction conditions for which are shown in the Ewald sphere construction in Figure 6.10.

For WB imaging, the sample is tilted about an axis normal to the selected \( g_{hkl} \) so that the Ewald sphere intersects close to a higher order spot, \( n g_{hkl} \), along the systematic row in which \( g \) lies. In Figure 6.10, the \( 5g \) spot is shown intersected by the Ewald sphere, although \( n \) need not be integer, and the WB DF image is then recorded by placing the CA around a spot such as +\( g \) (as shown in Figure 6.10) or −\( g \), for which \( s \) is large (\( \sim 0.2 \) nm\(^{-1} \)). Prior calculation is necessary to determine the value of \( n \) for a given \( \lambda \) and \( g \) suitable for generating a WB image. However, because the image intensity falls inversely with \(|s|^2 \), it becomes increasingly difficult to focus on the image. A dislocation image width, approximately \( \xi_g^{\text{eff}} / 3 \), is then as small as \( \sim 1 \) nm and closely spaced dislocations are resolved much more clearly, as in Figure 6.11 [26]. Because the WB image is always a DF image, the dislocation cores are imaged as thin bright lines on a dark background. It is also important to use tilted illumination so that the DF spot used for imaging lies on OA and aberrations are minimized. This is indicated in Figure 6.10 by the incident electron wave vector, \( k_0 \), inclined to the optical axis AO. The WB image resolution is such that in many materials it is then possible to image the partial dislocations formed when perfect dislocations dissociate, to image the stacking-fault ribbon (labeled SF in Figure 6.11) between the partials, and even to estimate limits for the stacking-fault energy [27].
6.3.4 HREM Image Formation

HREM imaging provides a means for exploring crystal structures and imperfections on the atomic scale when the sample is tilted such that a low-order zone axis is closely parallel to the incident beam, and columns of atoms are viewed end-on by the beam. However, a direct correspondence between atom columns and image features does not generally occur. In particular, the contrast will vary with OL excitation \( (C_1) \), \( t \), and the convergence angle, \( \alpha \), of the illumination. In order to interpret HREM image contrast, a comparison is usually necessary between a series of experimental...
images, recorded as \( C_1 \) is systematically changed by a known increment, with images generated by computer simulation for which the parameters listed above are also systematically varied. Even when an accurate match is obtained, it is commonly observed that the experimental values of image contrast are significantly lower (typically by a factor of roughly 3) than computed values [28].

In this section, the issues relevant to HREM imaging, both experimental and computational, are described. These include sample illumination, the propagation through the sample, the distortion introduced by OL aberrations and defocus, \( C_1 \), and the stability of the microscope power supplies.

### 6.3.4.1 Spatial Coherence

First, consider the nature of the illumination at the sample. It was noted earlier that the sample sees electrons as emanating sequentially in time from different points within the CL2 aperture, of angular radius \( q_o \), and that the electron wave packets are spaced far apart in the \( z \) direction parallel to OA; thus, the probability of electron overlap in the sample is very small when using the nearly parallel illumination required for HREM. Any image contrast, therefore, must arise from self-interference within individual electron wave packets and, across the area of interest, the wave front must therefore be spatially coherent [29].

On the wave front, \( \psi_f \), of any incident wave packet, a small path difference arises between two points in the sample, separation \( r \), which increases with \( |r| \). The incident wave is spatially coherent over \( |r| \) provided that this path difference is negligible compared with \( \lambda \). Clearly, electrons emitted at the periphery of the CA generate the greatest path difference across the sample. Hence, as \( q_o \) increases, the fraction of electrons in the beam that coherently illuminates any small area of the specimen falls. The spatial coherence length, \( r_c \), is a measure of the lateral distance over which the phase of the incident wave packet is coherent at the sample. For a wave packet with incident illumination angle, \( \alpha \), at the sample,

\[
r_c = \frac{\lambda}{2\pi \alpha}
\]  
(6.10)

Even at very high magnification and with \( q_o \sim 0.5 \) mrad, the field of view is usually larger than \( r_c \) because \( q_o > \lambda/2\pi r_c \), and the illumination is then partially coherent because any area of the sample will be illuminated by a mixture of both coherent and incoherent illuminations [30,31].

For atomic imaging by HREM through the interference of diffracted beams, it follows that \( r_c \) needs to be larger than the interatomic spacing for a significant fraction of the incident wave packets, so that the phase of the atomic scattering accurately contributes to the wave function at the exit surface, \( \psi_e \). Atomic resolution is usually required over a field of view typically tens of nanometers in diameter. When the illumination is spread over the sample to become more parallel, \( \alpha \) tends to zero and \( r_c \) increases. However, the intensity of illumination also falls as \( q_o \) and \( \alpha \) are reduced. The need for both spatial coherence and sufficient intensity of illumination for HREM imaging is readily achieved with an FEG source and even with LaB6. (Although partial coherence is possible with a simple heated W source, the intensity of illumination is weak, so HREM imaging becomes more difficult.) Even though \( q_o \) is small (~0.5 mrad or less), computer image simulation still requires that the CL2 illumination aperture area is divided into many small elements, with each element being treated as a parallel beam at a specific incident angle, \( \alpha \), to OA. A separate image calculation is then made for each element, and the contributions from all the elements added incoherently, weighted according to the intensity distribution across the CL2 aperture, to give a single image.

### 6.3.4.2 Electron Beam Propagation through the Sample

Although HREM is primarily concerned with imaging atomic structure, first consider propagation through a homogeneous continuum sample of thickness \( t \), inside which the potential of the fast electron is uniformly raised by a few volts, \( \phi_o \), the mean inner potential, relative to vacuum. Within the
sample, therefore, the electrons travel faster than in vacuum and the refractive index is increased. A plane wave incident on the sample will emerge with a path difference, $\chi$, relative to the same wave traveling through vacuum. The phase difference, $2\pi\chi/\lambda$, can be derived from the dependence of $\lambda$ on $E_o$, Equation 6.2, as

$$2\pi\chi/\lambda = \pi\phi_\text{o} / \lambda E_o = \sigma \phi_\text{o}$$

(6.11)

where $\sigma = 2\pi m e \lambda / h^2$ [1]. Hence, an incident plane wave of unit amplitude, given by Equation 6.3, exits with a wave function, $\psi_e$, of the form

$$\psi_e = \exp(-2\pi i k \cdot r) \cdot \exp(-i\sigma \phi_\text{o} t)$$

(6.12)

i.e., the phase of the wave is modified by the term $\sigma \phi_\text{o} t$. As only the variations of phase introduced by the sample are of interest, Equation 6.12 is usually abbreviated to

$$\psi_e = \exp(-i\sigma \phi_\text{o} t) \approx 1 - i\sigma \phi_\text{o} t$$

(6.13)

provided $|\sigma \phi_\text{o} t| \ll 1$. This is known as the weak-phase object approximation (WPOA) and assumes that the scattered intensity is much smaller than the incident beam intensity.

When a crystalline sample is viewed down a zone axis, then the phase change, $\phi_\text{p}(r)$, at a point, $r$, in the exit surface of the sample will be determined by the integrated path difference, along the $z$ direction in the sample. Thus, along the path AB in Figure 6.12, which passes close to an atomic nucleus represented by the black dots, $\phi_\text{p}(r)$ will increase with depth faster than $\phi_\text{p}(r)$ along the path CD, which passes between nuclei, as the path difference increases. The exit wave function, $\psi_e$, therefore, in the WPOA will vary with position $r$ and Equation 6.12 can be rewritten as

$$\psi_e(r) = \exp[-i\sigma \phi_\text{p}(r)] \approx 1 - i\sigma \phi_\text{p}(r)$$

(6.14)

Equation 6.14 shows that $\psi_e(r)$ has two terms, the first term, i.e., “1,” is the undiffracted beam amplitude, while the scattering term $\sigma \phi_\text{p}(r)$, which is complex and has both an amplitude and phase variation across the sample, contains all the information about the crystal structure. The DP is given by the FT of $\psi_e(r)$.

**FIGURE 6.12** Wavelength variation and path difference as a function of position at a crystal zone-axis orientation. AB and CD denote electron paths.
As \( t \) increases, the WPOA breaks down because the scattered electron intensity in the diffracted beams is no longer small compared with the incident wave intensity and multiple elastic scattering occurs between the diffracted beams. In practice, the WPOA breaks down even when the sample has a low \( Z \) and is only a few nanometers thick. Computer simulation must then be used to model the wave propagation through a known crystal structure to calculate \( \psi_e(r) \) [32–36]. Such calculations must be performed at discrete thicknesses within a likely range of \( t \) when comparing with experimental images.

### 6.3.4.3 Objective Lens Aberrations

It has been known for many years that round electromagnetic lenses suffer from aberrations that cannot be eliminated. When a plane wave is transmitted through an ideal lens, it is transformed into a spherical wave, represented by the solid curves in Figure 6.13 that converge to an exact focus, \( P \), on OA. Hence, an ideal OL changes the phase along each ray path in an incident plane wave such that all the rays have identical phase at the point of interference, \( P \). In a real lens, the converging wave front is no longer spherical but becomes increasingly distorted by various lens aberrations as \( \alpha \) increases, indicated by the dashed curves in Figure 6.13. Different parts of the wave front converge to different points along OA. Additional, azimuthal, distortion arises through astigmatism, when the focusing field departs from the ideal cylindrical shape, and through coma, when the incident illumination is inclined to OA. The effect of the various aberrations (the wave aberration function or path difference), \( \chi(\alpha) \), can be represented as a power series in \( \alpha \) and the azimuthal angle, \( \phi \) [37], i.e.,

\[
\chi(\alpha) = A_1 \alpha \cos(\phi - \phi_1) + 1/2A_2 \alpha^2 \cos 2(\phi - \phi_2) + 1/2C_1 \alpha^2 \\
+ 1/3A_2 \alpha^3 \cos 3(\phi - \phi_3) + 1/3B_2 \alpha^3 \cos(\phi - \phi_3) \\
+ 1/4A_4 \alpha^4 \cos 4(\phi - \phi_4) + 1/4S_3 \alpha^4 \cos 2(\phi - \phi_4) + 1/4C_3 \alpha^4 + \cdots
\]  

(6.15)

The coefficients corresponding to the different aberrations are defined as follows: \( A_1 \), first-order (twofold) astigmatism; \( C_1 \), defocus; \( A_2 \), threefold astigmatism; \( B_2 \), coma; \( A_3 \), fourfold astigmatism; \( S_3 \), star distortion; \( C_3 \), threefold spherical aberration, etc. \( \phi \) is the azimuthal angle and the various angles \( \phi_{ij} \) refer to the azimuthal values of reference planes.

![FIGURE 6.13 Distortion of a converging spherical wave front by objective lens aberrations. P is the focal point for undistorted spherical waves. Waves distorted by aberration come to a focus at other points on OA.](image)
All microscopes are fitted with correctors to minimize first-order astigmatism, $A_1$, while coma, $B_2$, can be effectively eliminated by careful alignment of the illumination in a modern HREM. However, the defocus setting, $C_1$, of the OL is a user-controlled parameter that is optimally non-zero (see below) and, hence, is an important aberration. The other major aberration that affects $\chi$ is spherical aberration, $C_3$. Note that, although the IL and PL further magnify the image by several orders of magnitude, the aberrations in these lenses are not significant because electrons travel through the OL at much greater angles to OA than in the IL and PL. Consequently, the pole-piece bore and separation in the IL and PL can be much larger than in the OL and their aberrations still be insignificant.

Each spatial frequency of the crystal potential, $\phi_p(r)$, generates intensity in the DP at the corresponding reciprocal lattice position, $K$ (Figure 6.4), which is transferred through the OL at its unique angle, $\alpha$, to OA. Hence, the aberrations of the OL change the phase of each diffracted beam by an amount related to $\alpha$ and to the azimuthal angle of propagation, $\phi$, before the beams interfere (Figure 6.13), thereby distorting the structural information in the HREM image. Because the phase varies with $C_1$, the image intensity at any point varies with OL focus. When computing images, it is necessary to know the value of $C_3$ and to vary systematically the value of $C_1$ for every value of $\alpha$ and $t$.

The phase distortion, $2\pi\chi/\lambda$, of a diffracted beam from the ideal spherical wave, described in terms of its scattering vector $K$, is given by [5]

$$2\pi\chi(K)/\lambda = \pi C_1 \lambda K^2 + \pi C_3 \lambda^3 K^4/2$$  \hspace{1cm} (6.16)

Constructive interference between the central, undiffracted beam and a scattering direction, $K$, at some value of $C_1$ will enhance the image contrast of that spatial frequency. Other spatial frequencies, however, may be destroyed by destructive interference in that image. Under the WPOA and the linear imaging approximation [38], whereby scattering between diffracted beams is assumed to be negligible, the contrast transfer function (CTF) is related to the gradient of the phase function given in Equation 6.16 by

$$\text{CTF} = \sin \left[ \frac{2\pi}{\lambda} \cdot \frac{d\chi(K)}{dK} \right]$$  \hspace{1cm} (6.17)

The effect of partial spatial coherence (Section 3.4.1) resulting from focused illumination with maximum semi-angle $q_o$, requires multiplication of the CTF in Equation 6.17 by an exponential function that attenuates the CTF at high $K$. The form of the CTF is then given by

$$\text{CTF} = \exp \left[ -\pi^2 q_o^2 (C_1 \lambda K + C_3 \lambda^3 K^3)^2 \right] \sin \left[ \frac{2\pi}{\lambda} \cdot \frac{d\chi(K)}{dK} \right]$$  \hspace{1cm} (6.18)

The CTF describes how the amplitude of the transmitted spatial frequency, $K$, depends on the defocus $C_1$, $\lambda$, and $C_3$. CTFs are shown in Figure 6.14 for a 200 keV TEM with $C_3$ of 0.5 mm, $q_o = 1$ mrad, and at two defocus values, $C_1 = -35$ and $-48$ nm.

Astigmatism is assumed to have been corrected and all other higher order aberrations ignored. At very small $K$, CTF $\approx 0$ but it then falls to $-1$ over a range, $\Delta K$, before oscillating between $+1$ and $-1$. This first wide band, $\Delta K$, in the curves in Figure 6.14, where the CTF $\approx -1$, implies that $\chi$ is faithfully transferred by the OL over $\Delta K$, and that direct interpretation of the image intensity is possible for this range of $K$ up to the point where the CTF first crosses the $K$-axis yields. This intersection point of $K$ defines the point resolution, $d$ ($d = 1/K$), given by [39]

$$d = 0.707 \ (C_1 \lambda^3)^{1/4}$$  \hspace{1cm} (6.19)

The corresponding value of $C_1$ is called the Scherzer defocus value and is calculated as $C_1 = -(C_3 \lambda)^{1/2}$ [39].
FIGURE 6.14 CTFs for two values of defocus, −35 and −48 nm, for $q_o = 1$ mrad, $C_3 = 0.5$ mm, and $E_o = 200$ keV.

In Figure 6.14, the curve for $C_1 = -35$ nm is close to Schertzer defocus and gives a point resolution of $-0.47$ nm$^{-1}$ (or $-0.21$ nm). At higher $K$, however, contrast reversal occurs such that alternate passbands of $K$ are transferred with reversed phase. Comparison of the curves in Figure 6.14 shows that the positions of these passbands have a sensitive dependency on $C_1$, implying that contrast reversal of image features are induced by small changes in $C_1$, preventing direct image interpretation at higher $K$. Equation 6.19 suggests that the resolution can be improved by reducing $\lambda$ (i.e., by increasing $E_o$) or $C_3$ through OL design or $C_3$ aberration correction. A detailed discussion of the passbands can be found in Ref. [5].

6.3.4.4 Temporal Coherence

In addition to the inherent aberrations included in Equation 6.15, the image is further degraded by chromatic aberration and limited temporal coherence. Temporal coherence is determined by the mechanical and electrical instabilities in the microscope system that cause changes in $C_1$ during the image recording time. Such instabilities are always present and manufacturers are constantly trying to minimize fluctuations in power supplies and vibrational resonances. (Protection from stray external acoustic and electromagnetic sources is of prime concern in the selection of a site for a TEM.) Assuming that these instabilities occur independently, their effect in blurring the image is combined using the central limit theorem to give a defocus distribution $F(C_1)$ [40]:

$$F(C_1) = \frac{1}{\sqrt{2\pi\Delta}} \exp \left( -\frac{C_1^2}{2\Delta} \right)$$

(6.20)

where,

$$\Delta = 2^{1/2} C_c \left[ \frac{\sigma^2(V_o)}{V_o^2} + 4 \frac{\sigma^2(I_o)}{I_o^2} + \frac{\sigma^2(E_o)}{E_o^2} \right]^{1/2}$$

(6.21)

There are two voltage-dependent terms in Equation 6.21: $\sigma^2(V_o)$ is the variance of the electronic fluctuations of the EHT power supply and $\sigma^2(E_o)$ is the variance of the natural spread of energy of the electrons from the source (given in Table 6.1) that are always present even if the EHT were perfectly stable; $\sigma^2(I_o)$ is the variance of the OL current; and $C_c$ is the chromatic aberration coefficient of the OL. The factor 4 occurs because the OL focal length is inversely proportional to the square of
the OL current. The effect of temporal incoherence is to attenuate the transfer of higher spatial frequencies, described mathematically by another exponential envelope function, \( E_t \), given by [41,42]

\[
E_t = \exp\left[-\pi^2 \Delta^2 (0.5 \lambda K^2)^2\right]
\]

(6.22)

The effect on the CTF curves of Figure 6.14, following multiplication by \( E_t \), is shown in Figure 6.15. The temporal incoherence leads to much greater attenuation of the CTF at high \( K \) compared with spatial incoherence under normal HREM imaging conditions. The value of \( K \) at which the envelope function falls from unity to \( 1/e \) defines the information limit, an important parameter in HREM [5]. Even if the point resolution, \( d \), Equation 6.19, could be made very small, the information limit will usually determine the minimum \( K \) that can be usefully recorded by HREM. In Figure 6.15, a value of \( \Delta = 5 \) eV has been assumed, consistent with \( C_c = 1 \) mm, and instabilities in power supplies of 1 part in \( 10^6 \). Great efforts are being made continually to reduce instabilities further (e.g., Ref. [43]).

Following careful alignment of the TEM column, correction of astigmatism, and orientation of the sample to a selected zone axis, the HREM experimental procedure is to record a series (10–20) of images while systematically varying \( C_1 \) over a range of values around Scherzer defocus. The incremental changes in \( C_1 \) are, typically, a few nanometers. The experimental images are then compared in detail with an appropriate series of computed images to look for correspondence. Figure 6.16 illustrates the need for careful \( C_1 \) adjustment. The four images of the same sample area are from a more extensive through-focal series from the complex oxide Nb16W18O94 projected along the [010] direction, recorded using a 300 keV FEG TEM (JEOL 3000F) [44]. Fine detail is observed to vary rapidly for focus changes of only a few nanometers; focus values are marked in the top left of each image (negative implies underfocus). Because there is clear, periodic structure in all images, every image contains structural information and it is impossible to state that any particular image is the one which is correctly focused. (Note that the sample is covered by a very thin amorphous film of contamination, which can be seen in each image at the edge of the sample.)

The expressions for the envelope functions of the CTF imposed by spatial and temporal coherence limits are applicable only in the WPOA and the linear imaging approximation. No scattering between one diffracted beam and another is permitted in this approximation. As \( t \) increases, scattering between diffracted beams increases and nonlinear imaging becomes increasingly important. The
envelope functions, Equations 6.18 and 6.22, are modified by nonlinear imaging such that the point resolution and information limit change [45].

6.3.4.5 Phase Problem

$\psi_i$ is a complex function, Equation 6.14, and the phases of the $K$ components, containing the crystal structural information, are distorted by the OL before the image, $\psi_i$, is formed. Moreover, the detector records only the wave intensity, $\psi_i\psi_i^*$, in the image plane, in which case all the phase information is lost. This is the phase problem of HREM.

It can be seen from Figure 6.15 that the point resolution, which defines the limit of the zero-order passband, $\Delta K$, within which all spatial frequencies are transferred to the image with roughly the same phase and efficiency, varies with $C_1$, as do the higher passbands of $K$. One possibility for imaging a known crystal structure is to adjust $C_1$ so that all the spatial frequencies present at a selected zone axis lie in a single passband above the point resolution, or lie in different passbands that all have the same polarity in the CTF. Then, all $K$ at that zone axis are correctly combined into the image, even though the point resolution is exceeded. The problem is that small focus fluctuations can easily reverse the phase of some of the higher frequency passbands because the passbands are relatively narrow. The method is, therefore, unsuitable for unknown structures and the information limit still imposes a fundamental limit that cannot be overcome by this approach.
Schiske [46] has suggested that it should be possible to recover both the phase and amplitude of \( y \) by recording and subsequently analyzing a series of HREM images at appropriate \( C_1 \) values, such that the whole range of \( K \) up to the information limit would be transferred strongly in at least one of the images. Other groups have developed algorithms to achieve this [47–49], and it is now routinely performed in a number of laboratories. Further advantages, using series of images for which both \( C_1 \) and \( \alpha \) [50] are systematically changed, have also been demonstrated [51].

Equation 6.19 indicates that increasing \( E_o \) and reducing \( C_3 \) increases the point resolution. However, there is a limit to the smallest \( C_3 \) that can be produced in an OL. An alternative approach has been to develop correctors, based on non-round lenses, to correct and eliminate aberrations up to third order, including \( C_3 \) [52–57]. The through-focal series reconstruction methods can still be used. The point resolution can be moved much closer to the information limit in a \( C_3 \)-corrected TEM, as illustrated in Figure 6.17, in which the CTFs for \( C_3 \) values of 0.5 mm (\( C_1 = -35 \) nm) for a conventional OL, and 5 \( \mu \)m (\( C_1 = -3.5 \) nm) for a \( C_3 \)-corrected OL, are compared for \( \Delta = 5 \) eV, \( q_o = 1 \) mrad, and 200 keV. For the latter, the spatial coherence has been improved to the extent that the CTF oscillations have been completely suppressed, and the structure can be interpreted directly, out to the information limit \( K \approx 7.5 \) nm\(^{-1}\).

The development of monochromation [58–62] for FEG guns provides a means of pushing the information limit to higher spatial frequencies by reducing \( \sigma^2(E_o) \) in Equation 6.21. However, the development of chromatic aberration correctors [63,64] offers the most dramatic means of improving the information limit by reducing \( C_c \), again in Equation 6.21. In combination with \( C_3 \) correctors, a point and information limit close to 0.05 nm should ultimately be obtainable, with acceptable image intensity, when \( C_1 \) can be corrected. Figure 6.17 also shows a CTF for \( \Delta = 1 \) eV to simulate such a microscope. At present, \( C_c \) correctors are still under development.

It is desirable to use a TEM that has both point resolution and information limits that allow direct interpretation of an HREM image, and that is particularly true when characterizing an interface on the atomic scale. The information limit of a TEM can be measured using the Young’s fringe method [65] with an amorphous sample of metal, typically Ge or Au, evaporated on to a holey C support film. Two images of the same area are sequentially recorded with a small lateral displacement. The images are then added and the power spectrum is formed by computing the FT (readily performed online when a CCD detector and computer image storage are available). An example is

**FIGURE 6.17** Comparison of CTFs. Full line: \( C_1 = -35 \) nm, \( C_3 = 0.5 \) mm, \( \Delta = 5 \) eV; dotted line: \( C_1 = 3.5 \) nm, \( C_3 = 0.5 \) mm, \( \Delta = 5 \) eV; and dashed line: \( C_1 = 3.5 \) nm; \( C_3 = 0.5 \) mm, \( \Delta = 1 \) eV. \( E_o = 200 \) keV and \( q_o = 1 \) mrad.
shown in Figure 6.18 for an amorphous C sample using an aberration-corrected TEM (JEOL 2200MCO) at 200 keV with very stable EHT and lens power supplies. The Young’s fringes have uniform intensity close to the center (the OA) but with gradually fading intensity as the distance from the OA increases. The superimposed circle marks the position of the 0.07 nm information limit, to which the fringes extend. The shape of the temporal coherence envelope is determined from the intensity attenuation with radius (proportional to $K$) in this power spectrum.

### 6.3.4.6 Implications for HREM Imaging of Interfaces

In order to apply HREM imaging to the study of the atomic displacements at a grain boundary, it is necessary (1) to orient the sample such that the axis about which one grain is rotated into the orientation of the other is parallel to OA and (2) that both grains are viewed along a low-index zone axis parallel to the rotation axis. For many general grain boundaries and interphase interfaces in metals, it is unlikely that the common rotation axis will coincide with a low-index direction. HREM imaging is limited to the study of simple tilt boundaries, a few special boundaries, and the interfaces between coherent and semi-coherent epitaxial layers. Fortunately, there are many important, nonmetallic, structures that contain such interfaces, including semiconductor devices, giant magnetoresistant structures, ceramic alloys, metal alloys, incommensurate minerals, etc.

### 6.4 SCANNING TRANSMISSION ELECTRON MICROSCOPY

In both DC and HREM conventional imaging modes described in Section 6.3, all parts of the sample are simultaneously illuminated by the electron beam, and the intensity in the image is integrated in parallel by the detector at each point (or pixel). STEM provides an alternative imaging method, which has both advantages and disadvantages compared with convergent beam TEM (CTEM), and many TEMs now offer a choice of both CTEM and STEM modes. There are also dedicated STEM instruments that allow imaging only in the scanning mode. The so-called HA-ADF STEM mode is an incoherent high-resolution image mode that is particularly powerful for imaging atomic
Transmission Electron Microscopy

structures as an alternative to conventional HREM. The same requirements and restrictions on zone-axis orientation and range of interfaces apply to both imaging modes.

6.4.1 STEM IMAGING

An STEM image is formed in a manner analogous to that from a solid sample in a scanning electron microscope (SEM). The incident illumination has the form of a cone of semi-angle of convergence \( q_o \), defined by the diameter of the aperture of the probe-forming OL. The probe is focused to a fine crossover at the entrance surface of the thin foil sample, and then scanned as a raster over a selected area of the sample. In a modern STEM, the scanning is performed digitally such that the probe is held for a selected dwell time sequentially at each pixel. A detector on the exit side of the sample intercepts part of the transmitted intensity (cf. SEM) over the selected angular range determined by the detector geometry, and the electron current is integrated and stored at each pixel, sequentially in time, as the probe is scanned over the image area. Ideally, STEM requires an FEG to ensure sufficient intensity in a very small probe, Equation 6.1, for atomic resolution. When STEM facilities are interfaced with CTEMs with LaB\(_6\) sources, resolution is limited to a few nanometers.

6.4.2 STEM INSTRUMENTATION

A typical imaging configuration for a dedicated STEM is shown in Figure 6.19. The electron probe is focused on, and transmitted through, the thin sample. A retractable fluorescent screen is inserted to view the DP and a retractable annular detector is positioned symmetrically around the OA to record the annular dark-field (ADF) image: the ADF detector collects electron intensity scattered over a range defined by its inner (\( \theta_i \)) and outer (\( \theta_o \)) cutoff semi-angles. Usually, there are no lenses between the sample and the ADF detector in a dedicated STEM, so the physical size and ADF detector position determine \( \theta_i \) and \( \theta_o \). (When an STEM facility is incorporated into a CTEM, the ADF detector is located within the intermediate/projector lens system such that lenses between the sample and the detector can be used to change the effective values of \( \theta_i \) and \( \theta_o \) at a detector of fixed physical dimensions.) A separate, on-axis, disk detector is used to collect the BF and low-order diffracted beams up to a maximum semi-angle of \( \beta (\leq \theta_i) \), determined by the collector aperture. When an EELS spectrometer is installed, the undiffracted beam passes through the spectrometer (described later in Section 5.2) before reaching the BF detector, as shown in Figure 6.19. The difference between ADF and HA-ADF imaging lies simply with the range of \( \theta_i \) and outer \( \theta_o \), as discussed in more detail in the following section.

![FIGURE 6.19 Dedicated STEM: the configuration of the probe, sample, detectors, EDX, and EELS.](image)
STEM offers significant, and increasingly exploited, advantages over CTEM. These include the following:

- Image of higher resolution can be formed using incoherent HA-ADF than can be formed using coherent HREM under equivalent electron-optical conditions ($E_0$, OL aberrations, optimum focus, etc.) [66–70].
- More than one detector can be used to record simultaneous multiple images. Thus, BF and ADF electron images can be acquired together with chemical compositional maps, for example.
- Information using EELS, EDX, and diffraction can be acquired with a spatial resolution similar to the probe diameter using a stationary probe.
- No changes in lens excitation are required when changing from a scanning (imaging) mode to an analytic (stationary probe) mode. Although, in principle, in a CTEM, it should be possible to switch from an imaging mode to a static-focused probe at a selected point, in practice, it is always necessary to make minor adjustments to stigmators, $C_1$, and probe position due to hysteresis in the power supplies. Such adjustments can be time consuming and increase damage and contamination.

Potential disadvantages of STEM compared with HREM include the following:

- Very slow image recording times because pixels are recorded sequentially.
- Image distortions and drift over extended acquisition times limit the accuracy of atomic column positions, for example, at interfaces and around crystal defects.
- Greater difficulty in performing DC imaging and working with DPs in STEM.

### 6.4.3 STEM Imaging Modes

The formation of the CTEM image and DP was shown in Figure 6.2 using simple ray optics. In the STEM, the incident illumination is a cone (Figure 6.20) and every diffracted beam emerges as a cone. The axes of the cones lie at $2\theta_B$ with respect to the axis, OA, of the forward scattered beam. Figure 6.20a shows just one diffracted beam such that $q_o < \theta_B$ and the diffraction disks do not overlap. However, when $q_o > \theta_B$ (Figure 6.20b), the disks overlap. A ray from point E in the incident cone can reach point Q in the disk overlap region, without diffraction, passing through the point P in the overlap range.

**FIGURE 6.20** DP disk geometry when (a) $q_o < \theta_B$ and disks do not overlap and (b) $q_o > \theta_B$, where disk overlap and coherent STEM BF imaging is possible.
sample, while a ray from point F in the incident cone can be diffracted through $2\theta_B$ at P and then travel to Q. If the source disk is coherent, then the two rays traveling along PQ will interfere. As the probe position, P, is translated across columns of atoms in a thin sample, to generate a scanned image, the phase of the diffracted ray at Q will change, depending on the probe position within the unit cell of the crystal. However, this would not be observed if $\beta$ were made very large (say $\pi/2$) such that the BF detector always collected all the transmitted intensity.

In practice, the DP from point P in a thin sample will contain many disks, with overlap between adjacent disks. Also, $\beta$ is 10–20 mrad and $q_o$ is $\sim$ 10 mrad, so the BF detector integrates the intensity over relatively few regions of disk overlap, together with the areas within disks where there is no overlap. As the probe is scanned, the intensity in the coherent overlap regions integrated by the BF detector changes with the periodicity of the projected crystal potential, and an HR image is generated, i.e., the fraction of signal lost to the BF detector varies as the probe is scanned. The principle of reciprocity [71,72] shows that a CTEM HR image will be identical to the corresponding STEM BF image provided (1) the beam convergence angle in CTEM is equal to the STEM BF detector angle, $\beta$, and (2) the STEM probe convergence angle, $q_o$, is equal to the angular diameter of the OL CA used to form the CTEM BF image. This is true both for elastic scattering and for inelastic scattering with small energy losses, such as TDS. In practice, HREM imaging in CTEM uses an illumination angle $q_o$ of $\sim$ 0.5 mrad and a CA that allows many diffracted beams to contribute to the image. When an STEM BF $\beta$ of only 0.5 mrad is used, the image is extremely noisy and, consequently, STEM BF is rarely used for HREM. It is important to stress that, for application of the principle of reciprocity, the STEM BF HREM image must have the same spatial and temporal coherence as the corresponding HTEM image and, therefore, depends on defocus, aberrations, and power supply instabilities as described in Section 3.4.4. Contrast reversals in the image will still occur as the $C_1$ changes and the limitations of the WPOA will apply (Section 3.4.2).

It follows that a DF image, $I^{DF}(r)$, formed by collecting the signal lost to the BF detector with an ADF detector (Figure 6.19) that extends from $\beta$ to $\pi/2$ will be complementary to the BF image, $I^{BF}(r)$, i.e.,

$$I^{DF}(r) = 1 - I^{BF}(r) \tag{6.23}$$

Contrast will, therefore, still arise by coherent interference between adjacent diffraction cones. Coherent STEM ADF lattice fringes were first demonstrated by Cowley [73]. However, the intensities of diffracted beams fall with increasing angle and, above a few tens of milliradians, the scattered signal arises largely through quasi-elastic TDS, manifest in the Kikuchi pattern in the CTEM DP (Figure 6.6). Such scattering is believed to be largely incoherent. The principle of reciprocity requires that a large STEM collection angle corresponds to a large illumination angle in CTEM. In Section 3.4.1, it was noted that the coherence of the CTEM illumination falls as the CL2 aperture angle, $q_o$, increases and, for very large angles, the spatial coherence distance, Equation 6.10, will eventually become too small for coherent HREM. Hence, for an ADF detector with collection limits of $\theta_t$ and $\theta_o$, the DF image will become increasingly incoherent as $\theta_t$ increases. Typically, when $\theta_t > 40 mrad$, corresponding to the TDS vibrational amplitude [74], it is possible to form an ADF image that is essentially incoherent and this is known as an HA-ADF image. Such images have low intensity, even when $\theta_t$ is very large ($\sim 200 mrad$).

When a cold FEG is used, an OL aperture of semi-angle 10 mrad can be filled with coherent illumination, and the probe intensity distribution at the sample can simultaneously be made very small by appropriate demagnification of the source. The probe wave function, $\psi_p(r)$, and intensity, $I_p(r) = |\psi_p(r)|^2$, are mathematically defined by the FT of the aperture disk function, $A(\alpha)$, convoluted by the OL phase function, $\chi(\alpha)$, i.e.,

$$\psi_p(r) = \text{FT}[A(\alpha) * \chi(\alpha)] \quad \text{and} \quad I_p(r) = |\psi_p(r)|^2 \tag{6.24}$$
A(α) = 1 if α < q_o, and A(α) = 0 if α > q_o. Three theoretical probe current distributions are shown in Figure 6.21 for 200 keV electrons, and these take the approximate shape of a Bessel function, with a large central peak and weaker oscillations as r increases. The two curves for which C_s = 0.5 mm and q_o = 8 mrad demonstrate the way in which the tail in the distribution changes as the defocus changes from 20 to 50 nm, while the third curve is for an aberration-corrected STEM where C_s and Δf are zero and q_o = 15 mrad. The benefits of C_s-correction are obvious; even though q_o is almost doubled, giving an increase of almost four times the current in the probe, the FWHM (full-width half-maximum) of the central peak is almost halved. The optimal probe shape approaches a Gaussian distribution for C_1 = −0.75(C_3l)^{1/2} and α = 1.27(λ/C_3)^{1/4} [75]. It must be stressed, however, that such probe current distributions are modified immediately when the probe enters the sample, leading to channeling of intensity along atom columns at zone-axis orientations. Although the HA-ADF signal intensity is often described as having a power law dependence on atomic number, Z, of the form Z^n, where 1.7 < n < 2 [67], in thicker alloy samples the simple relationship between intensity and Z no longer applies for atomic resolution. Channeling and scattering between the atom columns occur, leading to more complex dependence on both location and crystal thickness [76,77]. Currently, there is much interest in atomic imaging by combining STEM with EELS to use chemically specific signals generated by inner shell ionization. Such experiments will be complicated because recent calculations show that it is possible, when the probe is positioned on a specific atom column, to cause ionization in adjacent atom columns that might contain completely different atoms [78–81].

The incoherent image is defined by the convolution of I_p(r) with the object function, which can be represented as an array of scattering objects at the positions of the atom columns (when the sample is oriented with a zone axis parallel to OA). A detailed analysis [5,82] suggests that the form of the object function depends on θ_i. The image is then an array of bright spots close to the positions of the atomic columns, provided the probe-forming aperture is filled with coherent illumination. However, unless the image is completely incoherent (i.e., θ_i > 3q_o), the apparent separation of adjacent atom columns can be inaccurate. Kirkland [35] has written a computer code to calculate the STEM ADF image using an alternative, multi-slice approach to the Bloch wave analysis of Ref. [82].

The first atomic images in STEM were reported by Pennycook and Boatner [83]. An example of an HA-ADF image is shown in Figure 6.22 for a GaAs sample at a <110> zone axis [84]. The great advantage of the incoherent HA-ADF image is that there are no contrast reversals with focus; the image can be interpreted directly with confidence, unlike the CTEM HR image, i.e., the optical transfer function does not have the reversals shown in Figures 6.15 and 6.17 but tends to decay out to the information limit, and the point resolution is extended to higher K compared to the HREM image obtained using the same E_o, C_3, and C_c.
The first images of single atoms in the electron microscope were obtained using a STEM with \( E_o = 30 \text{ keV} \) by displaying the image as the ratio of the inelastic to elastic signals [85,86]. The atoms were those of U, i.e., with high \( Z \), on a low-\( Z \) supporting film. Such work stimulated much of the subsequent development of STEM and FEG. The ratio method of Crewe is not now commonly used because it is relatively straightforward to generate HA-ADF images of atom columns in many crystal structures. However, the imaging of low-\( Z \) atoms remains a challenge, as is the imaging of a single substitutional atom in a lattice of another element. The imaging of Sb atom complexes and even single Sb atoms in Si, Figure 6.23 [87], has demonstrated the power and potential of HA-ADF imaging. The application of HA-ADF imaging to interfaces is described in Chapter 9.

![FIGURE 6.23](image_url)

**FIGURE 6.23** HA-ADF STEM image showing the locations of atomic columns in a \(<110>\) cross section through an undoped Si substrate, to the right of the dashed line, and a Sb-doped epilayer. The sample thickness increases from left to right. The atom column intensity is almost constant in the substrate but, in the doped layer, some columns appear much brighter than neighboring columns, consistent with the presence of one or more heavier Sb-dopant atoms located in such columns. (From Voyles, P.M., Grazul, J.L., and Muller, D.A., *Ultramicroscopy*, 96, 251, 2003. With permission.)
6.5 CHEMICAL ANALYSIS METHODS

6.5.1 COMPOSITIONAL ANALYSIS BY EDX

EDX [88,89] has been used to determine chemical composition in the TEM and STEM for at least three decades. Although EELS [3] is a more sensitive method and can give more detailed information at higher spatial resolution than EDX, the relative simplicity of EDX still ensures that it is the method of choice for simple compositional measurement for many materials scientists using TEM. In this section, the principles of the EDX method are outlined, together with its advantages, limitations, and potential pitfalls. Examples of its detailed application to the measurement of compositional changes at interfaces and grain boundaries are described in Chapter 9.

6.5.1.1 Principles of EDX

When a fast electron passes through the cloud of bound electrons around an atom in a TEM sample, it can lose energy and generate two kinds of x-ray. If the atom is ionized, then it can return to an equilibrium state by emitting either a characteristic x-ray or an Auger electron. (The analysis of surface composition using Auger electrons is described in Sections 3.2.2 and 3.3.3.) The probability of X-ray emission is called the fluorescent yield, \( w (0 < w < 1) \), the value of which varies with \( Z \) and ionized shell. For example, \( w \) is greater for L-shells than for K-shells and increases with \( Z \) in a nonlinear manner. These x-rays have precise energies, \( E_X \), that are characteristic of the atom within which they were generated and are emitted isotropically from the sample. Alternatively, the fast electron can transfer energy to the atom without causing ionization, yet the energy is still emitted in the form of an x-ray. Such a continuum, or Bremsstrahlung, x-ray has energy, \( E_B \), between 0 and \( E_o \). Continuum x-ray emission is anisotropic, being strongly peaked in the forward direction.

In electron probe microanalysis (EPMA) [89,90] of a bulk sample, \( E_X \) is measured using x-ray diffraction by a large single crystal. Spectral accumulation over a range of \( E_X \) is sequential in time, although it provides greater sensitivity and much better energy resolution than EDX. The EDX method is based on the detection and energy measurement of individual x-rays, whatever the order of detection, and was developed from the same technology used by the nuclear industry to measure gamma radiation. A full description of the EDX method can be found in reference texts [88–90] so only the broadest outline is given here. Currently, the only type of x-ray detector interfaced to (S)TEM instruments is the Li-drifted Si type; a few Ge crystal detectors have been supplied and the same operational principles and analytical procedures apply. The detector is cooled by liquid nitrogen to reduce artifacts from electronic noise.

The detected x-ray undergoes complete absorption in a Si single-crystal detector by the photoelectron process. The emitted photoelectron has a limited range in the Si and soon loses energy through further ionization events involving successively less tightly bound electrons, finally resulting in the generation of electron–hole (e–h) pairs. On average, one e–h pair is produced for approximately every 3.8 eV of \( E_X \) or \( E_B \). An electric field of about 200 V/mm is applied across the Si crystal to separate the e–h pairs, and the resulting charge pulse is amplified, shaped, and measured by sophisticated electronics. This whole process must be concluded in time to be repeated when the next x-ray arrives and hence the time constant of the amplifier determines the maximum measurement rate. If a second x-ray is absorbed too soon, then the system will detect it and prevent it from interfering with the measurement of the first. However, the second x-ray will not be measured, and its absorption delays the return of the amplifier to its quiescent state in readiness for the measurement of the x-ray after that. The greater the frequency of overlap then the larger the dead time the system spends in a state of unreadiness. The microscopist can vary the x-ray generation rate by changing the electron probe current or, possibly, the sample thickness. A dead time \(<25\%\) is recommended. The total time for analysis, \( \tau \), is selected to allow accumulation of enough counts in the
spectrum for analysis with the required level of precision. The user must base this on previous experience and appropriate calculation. The accuracy of the analysis will depend on the care taken with the calibration of the system and measurement of a number of parameters that are described below.

A typical spectrum is shown in Figure 6.24 from a sample of Inconel 600 containing Fe, Cr, and Ni as the major alloying elements, with smaller concentrations of Si and Mo. The range of the energy scale is from 0 to \( \sim 20 \) keV, divided into 1024 channels of 20 eV/channel. Characteristic peaks from the major elements have a peak height much larger than the continuum intensity and are readily visible. The peak from Si is much smaller and noisier, as seen in the enlarged inset section between 0 and 10 keV. The continuum, also noisy, has a maximum intensity typically between 1 and 2 keV and then falls gradually as the energy increases. The width of the characteristic peaks is a measure of the detector resolution. Although a specific characteristic x-ray has a natural linewidth less than 1 eV, the measured spectrum peak-width is much larger, reflecting the stochastic processes that cause different numbers of e–h pairs to be generated by x-rays of the same energy. Hence the peak width increases with x-ray energy because the variance of the numbers of e–h pairs increases with \( E_X \) or \( E_B \). The capacitance and electronic noise of the detector and amplifiers also contribute to the peak width but, as this cannot be known before fabrication, the resolution and price of similar detectors can vary considerably. A good-quality detector will provide an FWHM resolution of \( \sim 60 \) eV at the C K peak at \( \sim 0.28 \) keV and an FWHM value about 130 eV at the Mn K\( \alpha \) peak at about 5.9 keV.

The shape of the continuum is the first feature that the microscopist should always examine when performing EDX. The actual generation of the continuum x-ray intensity is predicted to rise steeply as the x-ray energy falls [91,92] and is roughly proportional to \( Z^2 \). However, the detector has an energy-dependent detection efficiency factor, \( \varepsilon(E_X) \). Low energy x-rays are always strongly absorbed within the sample, in the protective window in front of the detector, and in the Au contact on the front surface of the detector, so the intensity always falls to zero at \( E_X = 0 \) eV. Within the energy range 3–10 keV, almost all x-rays are absorbed within the Si crystal, so \( \varepsilon(E_X) \approx 1.0 \). As the energy increases further, there is an increasing probability that the x-ray will pass right through
the detector crystal and be lost to the analysis, so $\varepsilon(E_X)$ begins to fall. Table 6.2 lists the values of $\varepsilon(E_X)$ measured for a windowless detector interfaced to an FEG-STEM at the $K_{\alpha}$ and $L_{\alpha}$ x-ray energies of several elements, to illustrate the magnitude of the variation [94].

There should, therefore, be a maximum in the measured continuum intensity, the position of which depends on (1) the type of detector window and (2) the extent of self-absorption of x-rays within the sample itself. The peak is usually between 1 and 2 keV and moves to higher energies for Be windows compared to ultrathin polymeric windows or windowless detectors, as the self-absorption (i.e., sample thickness) increases. However, the peak will also move to higher energies when there is a strong contribution from stray x-rays generated by thick material in the vicinity of the sample, or when there is not a complete line of sight between the electron probe position and the whole of the detector window.

### 6.5.1.2 Protocol for EDX Spectrum Acquisition

Although it is apparently very straightforward to acquire an EDX spectrum, high accuracy in quantitative analysis is difficult to achieve. The following protocol should alert the potential analyst to some of the many potential pitfalls:

- Detector should be reconditioned within the recommended period prior to use, otherwise preferential absorption of low energy x-rays can occur and $k$-factors used for quantification can be erroneous.
- Liquid nitrogen dewar should contain sufficient coolant to last longer than the duration of the experiment.
- Detector and amplifier should be under power for several hours before use in order to ensure stability of the pulse-processing electronics, amplifiers, and the energy scale in the spectrum.
- Low-background holder and sample-retaining screw made from Be should be used to minimize spectral contributions from x-rays generated outside the sample by stray scattered electrons.
- Ensure that the sample and holder are both clean, ideally by plasma cleaning, before insertion into the TEM. Wherever possible, ensure that the thin regions of the sample to be analyzed are close to the center of the holder and remote from the supporting cup or grid bars, which can generate x-rays by stray scattered electrons.
- Ensure that the probe size is small enough to provide the desired spatial resolution. (In a TEM, it is possible to image the probe directly. This is more difficult to ensure in a STEM.)

### Table 6.2

<table>
<thead>
<tr>
<th>L-line</th>
<th>Energy (keV)</th>
<th>Efficiency, $\varepsilon$</th>
<th>K-Line</th>
<th>Energy (keV)</th>
<th>Efficiency, $\varepsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>0.70</td>
<td>0.53</td>
<td>Fe</td>
<td>6.40</td>
<td>1.00</td>
</tr>
<tr>
<td>Ni</td>
<td>0.85</td>
<td>0.66</td>
<td>Ni</td>
<td>7.48</td>
<td>1.00</td>
</tr>
<tr>
<td>Ge</td>
<td>1.19</td>
<td>0.82</td>
<td>Ge</td>
<td>9.88</td>
<td>1.00</td>
</tr>
<tr>
<td>Mo</td>
<td>2.29</td>
<td>0.97</td>
<td>Mo</td>
<td>17.48</td>
<td>0.81</td>
</tr>
<tr>
<td>Sn</td>
<td>3.44</td>
<td>1.00</td>
<td>Sn</td>
<td>25.25</td>
<td>0.45</td>
</tr>
</tbody>
</table>

• Select the analysis position and first acquire a spectrum from a nearby location.
• Check that the dead time is $\leq 25\%$ and, if too high, reduce the beam current or the probe size.
• Confirm that the sample is not significantly damaged or contaminated during the trial analysis. If it is, then try using a smaller beam current or spread the beam over a larger area (if the desired spatial resolution permits this). If contamination persists, then it might be reduced by flooding the analysis area with a high intensity of electrons for several minutes. To do this, retract the detector from the TEM column, select the largest probe size and largest CA and defocus the illumination to spread the beam over a diameter of 10–20 $\mu$m. Illuminate for $\sim 5$ min before returning the TEM to the normal analysis conditions.
• Examine the shape of the continuum to ensure that no gross absorption is present. If it is, then tilt the sample holder further toward the detector and repeat until it seems that a clear line of sight is present between the point of analysis and the detector. If the sample is supported on a grid, then ensure that the analysis region is not shadowed by a grid bar. When analyzing an interface, this might be difficult because an additional geometrical constraint must be also met, i.e., to keep the interfacial plane parallel to the optic axis.
• Commercial analysis packages assume that the sample is homogeneous and amorphous. Hence, if the sample is crystalline at the point of analysis, it is important to ensure that the OA is not parallel to a low-index zone axis and is not strongly exciting any diffracted beams, i.e., that a kinematic diffraction condition operates, rather than dynamic scattering. Strong electron channeling operates when dynamic scattering is present and this can result in large variations in acquired counts [95,96]. In addition, false, coherent, Bremsstrahlung peaks can arise in the spectrum, which cause inaccuracies in the analysis of a number of elements, if present in small concentrations [97,98].
• Ensure that the characteristic peaks are accurately located at the correct positions on the energy scale. (Some systems require the collection of additional spectra using a range of counting rates to calibrate or check this.)
• If possible, collect a spectrum with the probe positioned in the center of a large hole near the analysis point using the same probe configuration to be used for the actual analysis. If the hole-count spectrum contains appreciable counts, then check that the CA and SADA are fully retracted and the hard x-ray aperture, if fitted on the TEM, is inserted and accurately centered. A badly positioned hard x-ray aperture can generate large stray x-ray signals instead of removing them.
• Collect spectra as the probe is moved from a hole and into the sample, noting the distance over which the counts increase from zero to a maximum. When this distance is much larger than the probe diameter, then there is significant intensity in an extended tail of the probe that will affect accuracy when high spatial resolution is required. A possible solution is to reduce the probe convergence angle by changing the CA diameter.
• Collect and store the required spectra.
• Perform experiments to measure the sample thickness and tilt with respect to the detector for use in correction for self-absorption of x-rays in the sample.
• Collect spectra from other samples to be used as standards for quantification.

### 6.5.1.3 EDX Spectrum Processing

The calculation of elemental concentrations from an EDX spectrum has been described in several references [99]. The number, $N_i^a(\text{ion})$, of ionization events in shell $i$, in a sample composed of one pure element, $a$, can be calculated from

$$N_i^a(\text{ion}) = (i/e) \tau \rho^s N_H \sigma_i^a / \lambda^a$$

(6.25)
where

- $i$ is the probe current
- $e$ is the electron charge
- $\tau$ is the analysis time
- $t$ is the time
- $\rho$ is the sample density
- $N_A$ is Avogadro’s number
- $\sigma_i^a$ is the total ionization cross section of shell $i$ in element $a$ at the given electron accelerating voltage
- $A^a$ is the atomic weight

Only a fraction of these events, $\omega_i^a a_i$, generate characteristic x-rays, where $\omega_i^a$ is the fluorescent yield and $a_i$ is the partition function [100–103], i.e., the fraction of the x-rays that arise in the specific sub-shell, $\alpha$, $\beta$, etc., chosen for analysis. If all the lines in a sub-shell are combined for the analysis, then $a_i = 1$. Only those x-rays emitted into the solid angle subtended by the detector, $\Omega$, have a chance of measurement, so the fraction of such x-rays is $\Omega/4\pi$. Finally, the number actually measured is further reduced by the relevant detector efficiency factor, $e(E_i^a)$, for the specific x-ray energy. Hence, the expression for the number of counts actually detected, $I_i^a$, becomes

$$I_i^a = (i/e)\tau\rho N_A \sigma_i^a a_i^a \Omega e(E_i^a)/(A^a 4\pi)$$

(6.26)

In a binary alloy containing elements $a$ and $b$, with weight fractions $c_a$ and $c_b$, the number of x-rays from element $a$ is given by the expression in Equation 6.26 multiplied by $c_a$. An equivalent expression gives the number of x-rays collected from shell $j$ of element $b$:

$$I_j^b = c_b(i/e)\tau\rho N_A \sigma_j^b a_j^b \Omega e(E_j^b)/(A^b 4\pi)$$

(6.27)

The two expressions finally yield the relationship between the weight fractions and the collected counts:

$$c_a/c_b = (k_{ab} I_a^a) / I_j^b$$

where

$$k_{ab} = \left[\sigma_j^b a_j^b \varepsilon(E_j^b) / A_j^b\right] / \left[\sigma_i^a a_i^a \varepsilon(E_i^a) / A_i^a\right]$$

(6.28)

Knowledge of the $k$-factors, $k_{ab}$, also known as the Cliff–Lorimer factors [104,105] permits quantification of spectra.

Spectra can be processed to yield a chemical composition using either a standardless procedure or using standard samples to determine the $k$-factors, $k_{ab}$. For both methods, it is necessary (1) to identify the characteristic peaks present and (2) to remove the underlying continuum to extract the number of counts, $I_i^a$, $I_j^b$, etc., in the peaks. Automated routines are provided by system manufacturers to remove the background, which can be problematic when peak overlap occurs. Digital spectrum filtering or fitting of theoretical or experimentally recorded standard peaks to the spectrum are usually employed.

The use of $k$-factors measured from standard samples of known composition provides the most accurate results provided suitable standards are available. Standards must be homogeneous on the nanoscale, be stable, and not damaged or oxidized. Moreover, appropriate correction may be required for self-absorption (see below) that can be tedious to compute. Unlike EPMA, where many bulk standards can be stored inside the microscope column and readily accessed and analyzed at precisely known geometry, sample exchange in the (S)TEM is slow and analysis is tedious. Hence, standard spectra acquired on the same or a similar (S)TEM at an earlier time are sometimes used. Standardless analysis requires calculation of the $k$-factors from first principles using values for $\sigma$, $\omega$, $\alpha$, $\varepsilon$, and $A$, the last of which presents no problem. Values for $\omega$ and $a$ are available in the literature.
[100,101], and either parametric expressions (National Bureau of Standards [NBS] special technical publication] or numerically computed values [106,107] are used for $\sigma$. The detector efficiency, $e$, is perhaps the most difficult parameter to calculate, especially for low x-ray energies where it varies quickly and is sensitive to the precise thickness of detector components. The absorbing layers in the detector that influence the efficiency are shown in Figure 6.25 and include (1) the window, (2) possible ice and contamination, (3) the Au contact in the front of the detector, (4) an inactive (dead) layer of Si just below the Au contact, and (5) the active Si crystal.

The thickness, density, and mass absorption coefficient (MAC) (a function of absorbing material $Z$ and x-ray energy) must be known for all five layers of Figure 6.25. The detector efficiency is then computed as

$$\epsilon(E_x) = I/I_o = \left\{1 - \exp\left[-\frac{(\mu/\rho) t}{y}\right]\right\} \cdot \prod \exp\left[-\frac{(\mu/\rho) t}{y}\right]$$

Here, the multiplication operation, denoted by $\Pi$, is for the factors of layers 1–4, and the parameter $(\mu/\rho)$ is the MACs for the x-ray energy in the absorbing material in layer $i$. Parameterized expressions for MACs for the whole periodic table can be found in the literature [108]. Detector manufacturers can provide data on construction materials and thicknesses of the layers (apart from layer 2), which, hopefully, can be eliminated by periodic conditioning of the detector. In practice, the system software should include appropriate expressions for the detector efficiency for use with standardless analysis.

Whether measured or standardless $k$-factors are used, the EDX system software derives the composition of the sample at the probe position, along with the standard error of each element. If the sample is thick or the angle between the plane of the sample and the direction to the center of the detector crystal (the takeoff angle) is small, then a correction may be necessary for the absorption of x-rays within the sample. This is especially important when one element in the sample, present in a significant concentration, strongly absorbs x-rays of another element. The expression in Equation 6.26 for the detected x-ray intensity, $I^a$, must be scaled by the factor

$$\left[1 - \exp(-x_i)\right]/x_i$$

where $x_i = (\mu/\rho)\rho y$ and $y = t/(\sin \theta + \cos \theta \cos \phi \tan \beta)$ [109]. The term $(\mu/\rho)$ is the effective MAC of the x-ray by the sample itself. The geometric parameters $\theta$, $\phi$, $\beta$, and the distance $y$ are shown in Figure 6.26. The upper surface of the wedge sample is inclined at an angle $\beta$ from the horizontal plane down its slope of steepest descent. The probe travels along TB, the local sample thickness, $t$, within the sample, generating x-rays uniformly along the trajectory. Depending on the generation depth, the x-rays have to travel through different distances before emerging from the sample. Figure 6.26 shows the distance, $y (=BE)$, to be used in Equation 6.30; it equals the distance traveled by x-rays generated at the point B. The direction of the x-ray detector is at an angle $\theta$ above the horizontal and at an azimuthal angle, $\phi$, to the plane containing TB and the normal to the plane of steepest descent.
As the mass absorption \((\mu/\rho)\) depends on the sample composition, whose measurement is required, an initial calculation of composition is performed without correction for self-absorption. This first estimate is used to generate MACs for the relevant x-rays in the specimen, for example, \((\mu/\rho)_{\text{spec}}^a\) for element \(a\), defined in Equation 6.31, allowing a first-absorption correction to be made:

\[
(\mu/\rho)_{\text{spec}}^a = c_a (\mu/\rho)_a^a + c_b (\mu/\rho)_a^b + c_c (\mu/\rho)_a^c + \cdots
\]  

(6.31)

where

\(c_i\) is the estimate of the concentration of a
\((\mu/\rho)^j_i\) is the MAC of the x-ray from element \(i\) by element \(j\)

In practice, the analyst is required to measure values for \(t\), \(q\), \(f\), and \(b\) at each analysis point, and to store details of the detector structure and geometry in the computer, which also stores values for MACs in look-up tables.

### 6.5.2 EELS: Principles and Instrumentation

EELS is a very powerful and versatile technique when coupled to either a TEM or STEM [3]. The following kinds of information can be found using EELS:

- Chemical composition: point, line profile, or map
- Valence and electronic bonding
- Local foil thickness
- Dielectric function
- Mapping of oxidation states and variations in bonding
- Atomic site location of impurities in ordered crystals (ALCHEMI) [110]
- Measurement of the bandgap of semiconductor materials [111]

In this section, the basic instrumentation and implementation of EELS techniques are presented. Detailed examples of applications to the characterization of interfaces appear in Chapter 9. The spatial and energy resolution limits in EELS have recently been reassessed by Egerton [112].

#### 6.5.2.1 EELS: Principles and Instrumentation

EELS provides information from the distribution of energy loss, \(\Delta E\), by the electrons in the beam during interaction with the atoms in the thin foil sample. There are two types of spectrometer: the
Transmission Electron Microscopy

in-column $\Omega$-filter and the post-column sector-magnet filter. Both types of spectrometer generate EELS spectra for $\Delta E$ up to $\sim 3$ keV, above which the intensity falls to very low levels over the angular range over which the signal can be collected. At the present time, the post-column sector-magnet systems made by Gatan Inc. are more common [113,114]. The sector-magnet system can be fitted to both the TEM and the dedicated FEG-STEM, and the basic design is illustrated schematically in Figure 6.27.

The incident electrons are scattered and undergo energy losses, $\Delta E$, in passing through the sample. An aperture limits the angular range of the scattered intensity that passes into the dispersing magnetic field with a direction normal to the plane of this page. All electrons are deflected along a curved trajectory within a drift tube located inside the sector magnet. Figure 6.27 shows how the electrons with energy $E_o$ (solid lines), and electrons with energy-loss $\Delta E$ (dashed lines), are focused by the sector magnet. An energy-selecting slit is inserted to limit the transmitted range of energy loss when forming an energy-filtered TEM image (EFTEM). The slit is removed and the magnifying lenses adjusted when the spectrum is viewed. The CCD array detector, of size 1 K $\times$ 1 K pixels or larger, is located inside the vacuum and is bombarded directly by the high-energy electron beam. The integrated charge at the CCD is periodically read into a computer memory for display as a spectrum or image, and successive readouts are added to achieve the required statistical accuracy, typically for times of up to tens of seconds. A spectrum can be shifted along the direction of dispersion by application of a voltage to the drift tube, in order to choose a range of the spectrum for investigation. Electromagnetic lenses control energy dispersion at the detector. Second-order aberrations in the magnetic field are reduced by shaping the entrance and exit surfaces of the sector magnet and further alignment and astigmatism correction is achieved by incorporating additional compensating coils that are not shown in Figure 6.27.

The simple spectrometer in Figure 6.27 is positioned below the standard viewing chamber in a TEM, so that the viewing screen can be used to select the area for analysis and then retracted to allow passage of the beam into the spectrometer. In a dedicated FEG-STEM, where there are no imaging lenses, the spectrometer is placed after the ADF detector and post-sample deflection coils (Figure 6.19). BF imaging is achieved using all, or part of, the forward-diffracted intensity that has passed through the spectrometer. The spectrum is acquired with the beam focused on a selected point in the STEM (or TEM operated in a STEM mode) or with a more parallel beam in a TEM.

© 2009 by Taylor & Francis Group, LLC
A typical EELS spectrum is shown in Figure 6.28. A large zero-loss peak (ZLP) marks the origin of the energy-loss scale, and the width of the ZLP is a measure of the energy resolution of the combined spectrometer/microscope system. The highest probability of any electron losing energy is by phonon excitation, which, at ambient temperature, requires ~0.025 eV energy transfer to the lattice and this is too small, currently, to be detected by spectrometers interfaced to EMs. Hence, electrons contributing to the ZLP include electrons that have suffered both zero loss and phonon losses. Apart from the ZLP, most of the spectral intensity is concentrated in the first few tens of electron volts, often peaking in the range 15–30 eV. This is the plasmon-loss region where the energy losses arise through cooperative excitations of loosely bound electrons, and ionization of valence electrons. As the sample thickness increases, the ZLP intensity falls, while the chance of multiple plasmon losses increases, leading to overlapping plasmon-loss distributions that decay into a fairly smooth intensity distribution as \( \Delta E \) increases further. In a few materials such as Al alloys, the plasmon peaks are much narrower (in the range 1–2 eV FWHM).

Superimposed on the decaying background are ionization edges at values of \( \Delta E \) corresponding to characteristic electron ionization energies of the elements that compose the sample. At high-energy resolution, many edges are observed to contain fine structure, and the precise value of \( \Delta E \) at the onset of the edge can shift over a few electron volts, depending on the valence state of the excited atom, and the nature and configuration of the surrounding atoms.

Many studies have been reported in which experimental fine structure has been compared with modeling predictions in attempts to understand the local atomic environment and interatomic bonding, and this is becoming increasingly important for studying atomic arrangements at interfaces. For a fast electron to contribute to an ionization edge, it must impart energy to a bound electron in the sample, such that the bound electron is transferred to an empty state above the Fermi level. Hence, the fine structure at an edge is proportional to the product of the density of vacant states and the probability of an electron transition from the occupied inner-shells to empty states, initiated by the fast electron. This product must then be convoluted by the spectrometer/detector response function and the energy spread of the incident electron beam to generate the theoretical spectral details.

In theory, the energy resolution in the spectrum is determined by a number of factors. Clearly, one limit is imposed by the physical separation of the diodes in the array detector and the dispersion. However, the dispersion can be made as large as necessary to overcome this constraint. The low-order aberrations in the spectrometer have now been largely overcome by improvements in design so that, for a small angular range of electrons close to the OA, it is the energy spread in the incident electron beam (see Table 6.1) that then limits energy resolution in many instruments. The advent of monochromators [58–60] overcomes even this constraint, so that the fundamental limit imposed by
the natural line width of the electron transitions during ionization is now achievable with a few sophisticated EMs. Currently, a ZLP resolution of ∼0.15 eV is possible with adequate intensity for spectroscopy.

In practice, however, many EELS experiments are still performed using an LaB₆ or FEG source without monochromation, and in spectrometers in which higher order aberrations are still present. Also, it is not always possible to eliminate the degrading effects of stray fields and vibrations, which can change on a daily basis due to factors beyond the control of the analyst. An energy resolution around 1 eV is typically achievable and acceptable for some of the purposes listed at the beginning of this section. In addition, there is usually a compromise to be made between energy resolution and SNR (signal-to-noise ratio). Because the electrons undergo momentum change during energy loss, they are scattered over a range of angles from the OA. As the collection angle at the entrance to the spectrometer increases, the signal intensity will increase. However, the resolution will degrade due to distortion by higher order aberrations, which increase with angle.

6.5.2.2 Protocol for Collection of EELS Spectrum

It is assumed that the microscope site is adequately protected from stray fields, mechanical vibration, and earth loops. During spectrum acquisition, no movement of metallic objects, such as chairs or noise generation, e.g., talking, should occur.

- Ensure the sample is clean because EELS spectra are very sensitive to C contamination. Select a region of the sample for use during alignment close to the region to be analyzed.
- Decide which coupling mode is to be used: the image mode requires a DP on the viewing screen and the diffraction mode an image on the screen [115]. The image mode is more accurate for compositional analysis. The diffraction mode might be more convenient for analyzing a very small particle using a convergent beam. The projector lens crossover located at the exit aperture of the final PL is close to the object point for the sector magnet (Figure 6.27). In image mode, there is a very low magnification image close to the crossover and in diffraction mode, there is a very low magnification DP close to the crossover (hence the names of the operating modes).
- Angular scattering range at the sample, β, collected by the spectrometer is selected using one of the interchangeable spectrometer entrance apertures located below the final viewing screen, together with adjustment of the DP camera length. An alternative method is to use the CA in the OL to limit β.
- β should be measured for later calculation of partial differential scattering cross sections. When the OL CA is used, β can be measured directly from a DP. When the spectrometer aperture defines β, then an estimate is obtained by dividing the aperture radius by the distance from the aperture to the crossover point in the final PL.
- Chosen spectrometer entrance aperture must provide the desired energy resolution in the spectrum. A sector magnet is equivalent to a lens. A larger aperture implies a larger angular range of electrons passing through the sector magnet and a greater influence of aberrations on the spectrum resolution. The best energy resolution, therefore, is achieved only when using the smallest available spectrometer entrance aperture (unless the image of the OL CA diameter is smaller at the spectrometer entrance aperture), but the signal intensity is reduced.
- Sample orientation should ideally be set to avoid strongly diffracting conditions when measuring composition, otherwise ordered alloys can generate orientation-dependent compositional data. When strongly diffracting conditions are unavoidable or chosen deliberately, as for a zone-axis orientation, then the angle-defining aperture should be symmetrically located with respect to the spots in the DP [116].
• Incident illumination convergence angle, $\alpha$, at the sample should be recorded. The theory underlying the conversion of an ionization edge intensity counts ratio into an atomic concentration ratio assumes that $\alpha$ is zero (i.e., parallel illumination). When $\alpha \approx \beta$, then a significant correction factor might be necessary to ensure compositional accuracy [3].

• Align the spectrometer. Nearly all sector-magnet systems have been provided by Gatan Inc. and aligned during initial installation to ensure that the OA of the microscope closely matches that of the spectrometer. However, further small adjustments are essential at the beginning of each session, and then should be checked periodically throughout an experiment. The procedure should be described in the user manual.

• Fine alignment is initially performed using the ZLP formed through a hole in the sample. Stigmators and compensators are adjusted to optimize the sharpness of the ZLP such that the ZLP does not move along the energy scale as the focus control is changed over a wide range about the optimum setting. The ZLP position on the CCD array is selected by deflection coils and major offsets are made by adjustment of the voltage on the drift tube. The sample is then moved under the beam.

• Ensure that the ZLP does not saturate the CCD during the data acquisition period by reducing the acquisition period, or the beam current, or both.

• Trial spectrum should be examined to ensure that there are no unexplained artifacts present, such as regions where the slowly changing background intensity rises with increasing $\Delta E$, rather than falling. The cause of any such artifact should be found, which might be difficult and even demand cessation of any further analysis until it can be rectified.

• Energy scale must be accurately calibrated. The energy scale is determined by the dispersion of the sector magnet and subsequent magnification (in $\mu$m/eV), selected by the user, before projection on the CCD array. The channel energy width in the spectrum is determined by dividing the physical spacing of the pixels by the dispersion. Calibration is performed automatically by the system software following the identification of two known edges, or the ZLP and one known edge, in a spectrum. (Note: Calibration can be performed post-acquisition on stored spectra.)

• Choice must be made about which $\Delta E$ is to be focused optimally. Although the initial alignment checks are most easily performed on the ZLP because it is so intense, ionization edges will then be progressively more blurred as $\Delta E$ increases. When energy-loss near-edge structure (ELNES) information of a particular edge is important, then the spectrometer should be focused on the selected edge rather than on the ZLP. The ZLP will then be blurred.

• Because the intensity of most ionization edges is so much lower than that of the ZLP, it is common practice to offset the energy scale of the spectrum so that the ZLP is not recorded. Long acquisition times are then feasible for ionization edges without saturation of any spectral channels. The analyst must decide on the extent of the offset voltage applied to the drift tube. Some further experimentation with acquisition times might then be necessary in order to optimize data recording at the edges of interest.

• Sample is then moved to enable recording of spectra from the required areas. Sufficient time must be allowed for any sample drift to subside and, if the CCD has recently been exposed to a ZLP, for any afterglow in the CCD to decay.

• With an older spectrometer, it might be necessary to record a dark current spectrum, with the beam deflected away from the spectrometer, to measure the background noise in the spectrum. Modern systems should automatically record such data with every spectrum acquisition.

• Spectrum containing the unsaturated ZLP and the adjacent low-loss region of the spectrum should be separately recorded. This spectrum can later be used for deconvolution of the multiple scattering contribution in the spectra containing the ionization edges, using the Fourier-ratio method [117,118].

© 2009 by Taylor & Francis Group, LLC
Each individual pixel of the CCD array has an individual gain efficiency. Ideally, a calibration of the CCD pixel response should be made for correction of spectra. If such a calibration is not performed automatically, then the user manual should provide information on how to measure and apply this scaling correction.

### 6.5.2.3 Sample Thickness Measurement Using EELS

The sample thickness can be determined from the low-loss region of the EELS spectrum. The single scattering distribution (SSD) is the energy-loss distribution arising from an infinitely thin sample such that no electron undergoes more than one inelastic event. Almost all the SSD intensity occurs within ~100 eV of the ZLP because plasmon, single electron excitations, and ionization of weakly bound core electrons with energy losses of only a few tens of electron volts have much larger cross-section (i.e., probabilities) than for ionization of the more tightly bound core electrons. The total inelastic scattering mean free path, \( \lambda_{\text{inel}} \), is the average distance traveled by a fast electron in the sample between any two inelastic scattering events. The low-loss region of the EELS spectrum from a very thin sample displays the SSD. As the sample thickness, \( t \), gradually increases from zero, intensity in the ZLP is gradually transferred to the SSD. However, the probability of multiple scattering increases with \( t \), and the low-loss region then includes contributions from multiple convolutions of the SSD. The probability of a fast electron undergoing multiple scattering is given by the Poisson distribution, from which it can be shown that the probability of an electron suffering no inelastic scattering, \( P_o \), is given by

\[
P_o = I_o / I_t = \exp (-t/\lambda_{\text{inel}})
\]

where \( I_o \) and \( I_t \) are, respectively, the intensity in the ZLP and in the total EELS spectrum, including the ZLP. Even when multiple scattering is present, almost all the total spectrum intensity lies within the first 100 eV, so that integration of counts in the ZLP and up to 100 eV provides experimental values of \( I_o \) and \( I_t \). \( t/\lambda_{\text{inel}} \) can be therefore be calculated rapidly. Figure 6.29 illustrates, schematically, the integration ranges for the ZLP and the low-loss region. Provided \( \lambda_{\text{inel}} \) is known, \( t \) can be derived.

The simple theory assumes that all the emergent electron intensity is collected in the spectrum. In practice, collection is limited over an angular range, \( \beta \), as already described earlier, so that a high proportion can be lost by high-angle elastic scattering (diffraction). It is reasonable to assume that the lost intensity is divided equally between the ZLP and the inelastic intensity, so that Equation 6.32 remains valid. However, \( \lambda_{\text{inel}} \) then becomes a function of \( \beta \), as well as \( E_o \) and atomic number \( Z \). An empirical formula for \( \lambda_{\text{inel}} \) as a function of \( \beta, E_o \) and \( Z \) [119] has been shown to generate values of \( \lambda_{\text{inel}} \) with an accuracy of ~10% over a wide range of the three parameters, and can be used with Equation 6.32 to yield values of \( t \) that are sufficiently accurate for many purposes.

**FIGURE 6.29** Schematic illustration of the energy ranges of \( I_o \) (ZLP) and \( I_t \) (low-loss region) for calculation of \( t \).
It is recommended (cf. EELS protocol) that measurements of \( t/\lambda \) be made whenever EELS experiments are performed. When \( t/\lambda \geq 0.5 \), the results will contain multiple scattering contributions that can distort ELNES peaks and lower the accuracy of compositional measurements unless they are removed by deconvolution [117,118].

### 6.5.2.4 Measurement of Chemical Composition by EELS

EELS provides an alternative to EDX for compositional measurement. The combinations of elements present in the sample often determine which method is preferable. EDX is generally suitable for quantitative analysis of all elements with \( Z \geq 10 \) (x-ray energy \( \geq 1 \) keV) even when peaks overlap; most of the \( k \)-factors lie within a factor of 3, and thicker samples are more readily analyzed. EELS, however, has a huge advantage for low-Z analysis. Although elements such as B, C, N, and O have large \( s_k \), their very low x-ray fluorescent yields, \( \omega_k \), high absorption in both sample and detector window, and the small solid angle of collection lead to very low EDX detection rates. EELS is not affected by \( \omega_k \) or absorption in the sample, and a high fraction of events is detected, so that these low-Z elements, and even He, Li, and Be are very easily detected and quantified. Count rates for many edges are orders of magnitude larger than for peaks in EDX so that precision and detection limits are generally much better in EELS [3].

Ideally, a single spectrum should be acquired containing all peaks of interest. However, when the ionization edges are too widely separated, it might be necessary to acquire consecutive spectra over different energy ranges, with consequently increased chance of changes due to drift, damage, and contamination, or to changes in beam current and electron-optical parameters. Ideally, any multiple scattering should be deconvoluted to yield the SSD; however, for \( t < \lambda \), failure to deconvolute is unlikely to degrade accuracy by more than a few percent.

The background intensity, \( B \), must first be determined and subtracted from under each ionization edge. The simplest method of background removal is to fit a power law of the form

\[
B = A(\Delta E)^r
\]

where \( A \) and \( r \) are constants. This is applicable only over regions where \( B \) is varying smoothly, so that the low-energy region of the spectrum (\( \Delta E \leq 100 \) eV) and regions on the tails of preceding ionization edges must be avoided. Even then, in practice, \( r \) increases as \( \Delta E \) increases, so that the fitting function is only approximate. Figure 6.30 shows a section of an EELS spectrum from a stainless steel sample displaying the L_{23} edges of Cr, Fe, and Ni. The variation of intensity in the energy window A has been used to extrapolate a background under the Fe and Ni edges, which has then been subtracted to generate the lower spectrum. The intensity in a window such as B is then used to quantify the Fe content of the sample.

More complex methods of background fitting to improve accuracy have been assessed in Ref. [120]. Subtraction of \( B \) reveals the ionization edges that extend over the range \( I_i^A < \Delta E < E_o \), where \( I_i^A \) is the \( i \)th shell ionization energy for element \( A \). For quantification, the edge intensity, \( S_i^A(W, \beta) \), is found by integration over a selected window of width, \( W \). The same procedure is repeated for all edges of interest. The edge intensity, measured in this way, is proportional to the probability that the fast electron is scattered through an angle no larger than \( \beta \) while losing energy in the range \( I_i^A \) to \( I_i^A + W \), namely the partial differential cross section, \( \sigma_i^A(\beta, W) \). As with EDX, it is usual to measure ratios of elemental concentration, so that the atomic ratio \( c^A/c^B \) of two elements, A and B, is given by

\[
c^A/c^B = S^A \sigma_j^B (\beta, W) / [S^B \sigma_i^A (\beta, W)]
\]

The validity of this formula has been discussed in Ref. [3] and some of the assumptions relating to crystalline sample and probe convergence were covered in Section 6.5.2.2.
6.5.2.5 Ionization Edge Shape and ELNES

Spectrometers typically transfer electrons that have undergone losses of no more than 3 keV. Within the energy range 0–3 keV, most elements have ionization edges that can be detected by EELS [121], although the intensities fall rapidly as $\Delta E$ increases. Edge shapes can be grouped broadly into different types:

- **Type 1.** Hydrogenic-like, with a large step at the onset, followed by a gradual decay (K-edges for all elements)
- **Type 2.** A sharp peak, or peaks, at the onset, superposed on a hydrogenic-like edge (L$_{2,3}$-edges for K–Cu and Rb+)
- **Type 3.** A delayed edge showing a gradual rise from the onset and then a gradual decay (L$_{2,3}$-edges for P–Ar, Zn–Br, and most M- and N-edges)

An energy resolution of $\sim$1 eV in EELS spectra is often sufficient to reveal small shifts in the onset energy, and changes in the fine structure related to differences in the unoccupied density of states (DOS) with coordination and bonding. Good experimental agreement has been obtained between XANES, using synchrotron irradiation, and EELS edges, for crystals with known structure. Figure 6.31 illustrates how the shape and fine structure of the O–K edge varies within iso-structural transition metal chromites [122]. Changes in the ratio of the L$_2$/L$_3$ peak intensities are sensitive to the valence state of 3d-transition elements [123], and the ELNES in many materials varies with orientation of the beam with the crystal sample [124].

Theoretical ELNES structure can be predicted for periodic crystal structures by self-consistent calculations using density functional methods [125] to derive the DOS, both occupied and unoccupied. The Wien2K code [126] is suitable for large crystal structures as it is based on a band-structure approach. However, this requires super-cells to be constructed in the computer from large numbers of atoms, which is computationally demanding. An alternative approach, also using self-consistent density functional theory, but requiring much less computing power, is to use a multiple scattering code such as FEFF8 [127,128] to derive the equilibrium configuration of an atomic cluster and its associated DOS. As the cluster size is increased, good agreement with experiment can be achieved [129].

![Figure 6.30](image-url)  
**FIGURE 6.30** Background extrapolation and subtraction under the Fe-L$_{2,3}$ edge in a spectrum from a stainless steel sample.
the DOS has been calculated, the theoretical EELS spectrum is generated from the product of the probability of the excitation of occupied states below the fermi level to a specific unoccupied state and the square of the appropriate transfer matrix elements. Convolution is then required with an appropriate function to account for the spectrometer resolution and the spread of energy in the electron beam.

Figure 6.32 illustrates the experimental ELNES Mg-L$_{23}$ ionization edge in MgO with corresponding calculated fine structure. There is excellent qualitative agreement between the two with respect to the presence and positions of all the major spectral peaks [130], in which all the observed fluctuations in the ELNES are reproduced with high accuracy by the calculations. Self-consistent approaches using density functional theory are essential to predict accurately the ELNES close to the edge onset. A number of less rigorous calculations based on non-self-consistent calculations have been reported over a number of years, but the level of agreement with experiment has generally been poorer.

As computing power increases, it is becoming feasible to model the ELNES measured from the disordered region of an interface. In addition to generating the DOS, such calculations will also yield the most energetically favorable positions of the atoms. For special interfaces, such as tilt boundaries and coherent interphase interfaces, the atomic structural information derived from HREM and HA-ADF images can be directly used as input into the calculations.

### 6.5.2.6 Energy Filtered Imaging in the TEM

So far, the description of EELS has concentrated on spectroscopy using the sector-magnet spectrometer. The analyzer system has been developed further by Gatan into the Gatan imaging filter [113] by the inclusion of additional lenses between the spectrum plane and the detector, as shown schematically in Figure 6.33. The object plane of the spectrometer is shown as a DP plane, where the spots contain electrons with a range of energy losses determined by the sample thickness and composition. The direction of the field in Figure 6.33 is out of the plane of the page. Inside the sector magnet the electrons are dispersed, the deflection increasing with the energy loss. The width and position of the energy-selecting slit determines the range of energies used to form the final image.
FIGURE 6.32 Comparison of ELNES experiment and theory for the Mg-L_{2,3} edge in MgO. (From Mizoguchi, T., Tatsumi, K., and Tanaka, I., Ultramicroscopy, 106, 1120, 2006. With permission.)

FIGURE 6.33 Schematic of an imaging filter illustrating how the paths of electrons with two specific energies differ in the magnetic field (the direction of which is out of the plane of the page) and how the slit controls the energy range of an EFTEM image.
The dashed lines represent the paths of electrons with energies lower than those of electrons with ray paths represented by the solid lines.

In practice, the slit position is fixed and \( E_0 \) is changed to form images with selected energy loss. The EELS computer can be programmed to record an image for each selected incremental value of \( E_0 \), thereby generating a series of energy-loss images. For a specific image from a selected ionization edge, three energy-loss images are usually recorded, two with energies below the identified ionization edge and one with energy just above the edge. For each corresponding pixel in the images, the intensities in the two images below the edge are used to calculate and subtract the background, \( B \), from the pixel intensity in the image above the edge [3]. When the sample thickness is constant, the intensity at that pixel is then proportional to the atomic concentration at that pixel, so that an elemental map is generated. A thickness map can also be made from two further energy-loss images recorded from (1) the ZLP and (2) the low-loss region, including the ZLP, from which \( \rho/\lambda_{\text{inel}} \) is calculated at each pixel using Equation 6.32. The elemental maps can then be normalized to remove the effects of any \( t \) variation.

Figure 6.34 shows an EFTEM example from an aged Fe–Cu alloy in which small spherical Cu-rich precipitates are formed in the Fe matrix. Several Cu precipitates with diameters of only a few nanometers are imaged in both the Cu-L\(_{2,3}\) and Fe-L\(_{2,3}\) maps [131]. Precipitates appear as bright objects in Figure 6.34a because they are Cu-rich, and as dark objects in Figure 6.34b because they are Fe-deficient. Intensity profiles along the diameter of an object were shown to correspond to a sphere, and the diameter of the sphere, calculated from the profile, was equal within experimental error to the diameter measured independently, from the total Fe-deficient signal of the whole object for radii \( \gtrsim 3 \) nm. Below this size signal broadening prevented accurate measurement from line profiles. However, blurred images of particles with radii as small as 1 nm, measured from the total Fe-depletion, were still detected [131].

The energy slit width, \( W \), and image acquisition time, \( \tau \), must be adjusted to collect enough counts to generate images with suitable SNR. Typically, \( \tau \) is of order 10 s, even when the illumination at the sample is focused to a relatively small area. An LaB\(_6\) source generates images with higher SNR than an FEG except at the very highest magnifications. Although it is possible to generate EFTEM images with \( W \) about 1 eV in order to select specific features in the ELNES [132–134], the SNR is very low and the smaller energy spread in the FEG beam cannot readily be exploited. When a series of many images is acquired, a 3D EFTEM data-cube can be recorded \((x, y, \Delta E)\). An alternative way of generating such data is to use an FEG-STEM with a conventional sector magnet so that a full spectrum is recorded at every pixel with both the highest possible spatial and energy resolutions.

**FIGURE 6.34** EFTEM images of spherical Cu-rich precipitates in Fe. (a) Cu-L\(_{2,3}\) map and (b) Fe-L\(_{2,3}\) map, in which Cu particles show reduced intensity. (From Lozano-Perez, S., Titchmarsh, J.M., and Jenkins, M.L., Ultramicroscopy, 106, 75, 2006. With permission.)
Transmission Electron Microscopy

available. This approach is called electron spectroscopic imaging (ESI) [135]. Systematic background subtraction is then performed for all peaks at every pixel in order to generate energy-filtered images, as for EFTEM. The potential for imaging with specific ELNES features using ESI is much greater than with EFTEM, although it can take tens of minutes to record full data sets and significant image distortion due to drift is likely.

6.5.2.7 In-Column Filters

An alternative to the sector magnet for EELS and EFTEM is the in-column filter, first introduced by Zeiss into a TEM column [136,137]. The in-column filter, located between the intermediate and the projector lenses, as shown in Figure 6.35, is often known as the omega filter because of its shape. Four sector magnets, M1–M4, are located symmetrically about the mirror plane AB. Figure 6.35a shows the ray paths in the x–z plane for two electrons entering along the same trajectory but with an energy difference of ΔE. The electrons emerge from the spectrometer at different angles, i.e., a spectrum is generated. A slit located below the filter allows a small energy range of the spectrum to be used for filtering the image or for DP. Figure 6.35b shows ray paths, also in the x–z plane, of three electrons with the same energy diverging from a point in the plane C1. The C1 plane corresponds to the image plane of the IL, so that either an image or a DP can be selected for filtering. In practice, a circular aperture is inserted between C1 and M1 in order to limit the angular range of electrons entering the filter. The rays in Figure 6.35b are parallel as they cross the plane AB but are subsequently focused to a point in the plane C2, located symmetrically with respect to C1.

In Figure 6.35c, three rays in the y–z plane through the filter are shown diverging from the same point in C1 as depicted in Figure 6.35b. (Note that the sector magnets are still shown in this plane as they would appear in the x–z plane to allow positional correspondence to be made.) These rays pass through a focus in the plane AB and are again focused at the same point in C2 as in Figure 6.35b. When the emerging rays in Figure 6.35a are linearly projected back into the filter, they intersect at a point where the height defines the achromatic image plane of the filter in M4, labeled AP in Figure 6.35a. Hence, in the plane C2, a spectrum is formed, apparently emanating from a point in AP.

![FIGURE 6.35 Schematic of an in-column filter showing electron paths for (a) two electrons entering along the same trajectory but with an energy difference, ΔE, (b) three electrons of equal energy but different angles of entry in the x–z plane, and (c) three electrons of equal energy but different angles of entry in the y–z plane.](image-url)
The spectrum is in the form of a thin line, very bright at one end where the high-intensity ZLP is located and then smoothly fading to almost zero intensity at the other end when the range of $\Delta E$ is large. The whole spectrum is imaged by the following PL and recorded using a CCD detector. Alternatively, an energy-selecting slit can be inserted across the spectrum line in the plane of C2, as indicated in Figure 6.35a, to transmit an EFTEM image. In practice, as with the post-column filter, the slit is permanently positioned on the optic axis and $E_0$ adjusted to select the desired $\Delta E$.

The symmetric construction of the filter ensures that low-order aberrations generated in the first half are canceled exactly in the second half (provided the mechanical construction and alignment is perfect). Because the image formed by the IL in the plane C1 is transferred to plane C2 with unit magnification, the filter is essentially invisible to the TEM operator in standard imaging and diffraction modes. An example of an EFTEM image obtained with an in-column filter of the same sample material shown in Figure 6.34, but recorded with a post-column spectrometer, is shown in Figure 6.36 [138].

### 6.6 CURRENT INSTRUMENTAL DEVELOPMENTS

Currently, there are several major developments in TEM instrumentation in progress, the most notable concerning aberration correction and monochromation.

The effect of aberrations in HREM imaging has been described in Section 3.4.3. Aberrations in the OL adversely affect the direct interpretation of HREM images in CTEM (Equation 6.15) and enlarge and distort the focused STEM probe (Equation 6.24). Chromatic effects also limit resolution (Equation 6.21). Early in the history of EM, it was recognized that the aberrations of round electromagnetic lenses could be corrected by combinations of multipole lenses [39]. Developments in manufacturing precision, computer control of multiple power supplies, and in real-time analysis of aberrations have recently allowed correctors to be fitted by all major EM instrument manufacturers. The principle of aberration correction is straightforward but difficult to realize. A multipole lens, such as a hexapole, introduces aberrations into an image, but the relative components of the aberration function, Equation 6.15, will be different from those of a round lens. When two identical multipole lenses are combined symmetrically with a round OL and excited in a complementary manner, then many of the aberrations introduced by the multipoles cancel exactly while others remain and are doubled. The magnitude of these remaining aberrations can be adjusted selectively and the polarity controlled to reduce, eliminate,
or even reverse the polarity of an aberration such as $C_3$ in a round OL. Other remaining aberrations introduced by the multipoles must themselves be canceled by the introduction of further multipoles. Clearly, as the lower order coefficients in Equation 6.15 are successively minimized or eliminated, then the relative importance of the higher order coefficients becomes more important and will determine the point resolution and Scherzer defocus value. Correctors of increasing complexity will be necessary to reduce simultaneously higher and higher orders of aberration.

Two types of aberration correctors have been developed for control of $C_3$. One of these [51–54] uses a combination of hexapoles and dipoles and is now increasingly incorporated into the OL of TEMs. A similar design of corrector has also been incorporated into the probe-forming optics of a CTEM, in addition to the OL, to generate probes in the range <0.1 nm [139]. A second type of corrector, using quadrupoles and octupoles, was designed for retrofitting to the dedicated HB501 STEM sold by Vacuum Generators for several decades [55,56], also to generate sub-0.1 nm probes. A point resolution below 0.1 nm is now achievable in both CTEM and STEM instruments. The possibility of using core-loss signals to form energy-filtered atomic images [78] is currently being attempted in several laboratories.

Unfortunately, the introduction of $C_s$ correction increases the chromatic aberration, $C_c$, significantly, so that the information limit is made worse than in the corresponding non-corrected instrument. Equations 6.20 and 6.21 show that the information limit can be extended by (1) reducing the energy spread in the incident beam, (2) improving the stability of power supplies, and (3) correcting $C_c$. The first of these can be achieved by including a monochromator in the electron gun and this has recently been realized [59,60]. A monochromator disperses the electron beam before it is accelerated to high energy, in such a way that an aperture can be used to permit only part of the energy distribution to proceed down the column. The design and operating principles of a monochromator are similar to those of an EELS system. The disadvantages of monochromation include a significant reduction in beam current and an increase in the minimum focused probe diameter at the sample. However, a huge advantage is that the achievable resolution in an EELS spectrum can be reduced to $\sim0.15$ eV, leading to significant advantages for ELNES applications. The second in the above list is a never-ending aim of the instrument manufacturers, and instabilities are now typically at a level of less than one part in $2 \times 10^6$. The third is currently one of the aims of a major instrument design program based in the United States [140] and should be realized very soon.

6.7 SUMMARY

In this chapter, a description has been presented of many of the features of EM as currently applied to a wide range of materials science issues. The spectroscopic methods for compositional and bonding characterization on the nanoscale are becoming increasingly important, and these are driving instrumental development just as hard as the desire for improved image resolution. Space limitation has prevented any description of some specialized but important areas for interfacial characterization, and the reader is referred to the specialist literature for such information. In particular, electron holography is developing as an alternative means to HREM image series deconvolution for the recovery of the phase of the exit-surface wave [141,142]. The study of magnetic materials and domain boundaries [143] requires specific methods and instrument modification, and so has not been directly included. Examples of the application of the methods described in this chapter to the study of grain boundaries and interfaces are described in Chapter 9.

REFERENCES