

Part 3

Imaging

Amplitude Contrast

CHAPTER PREVIEW

We've already mentioned in Chapters 2–4 that TEM image contrast arises because of the scattering of the incident beam by the specimen. The electron wave can change both its amplitude and its phase as it traverses the specimen and both types of change can give rise to image contrast. Thus a fundamental distinction we make in the TEM is between *amplitude contrast* and *phase contrast*. In most situations, both types of contrast actually contribute to an image, although we usually select conditions so that one will tend to dominate. In this chapter, we'll discuss only amplitude contrast and we'll see that there are two principal types, namely, *mass-thickness contrast* and *diffraction contrast*. This kind of contrast is observed in both TEM and STEM and in both BF and DF images. We'll discuss the important differences between the images formed in each of these two modes of operation. Then we'll go on to discuss the principles of diffraction contrast, which are sufficiently complex that it takes Chapters 24–27 to show you how this form of contrast is used to identify and distinguish different crystal defects. Diffraction-contrast imaging came into prominence in about 1956 when it was realized that the intensity in a diffracted beam depends strongly on the deviation parameter, s , and that crystal defects rotate the diffracting planes near the defect. Therefore, the diffraction contrast from regions close to the defect will depend on the properties (in particular, the strain field) of the defect. We'll then consider phase contrast and how it can be used to image atomic-level detail in Chapters 23 and 28. Other forms of TEM imaging and variations on these major types of contrast are gathered in the catch-all Chapter 29.

22.1 WHAT IS CONTRAST?

Before we start to describe specific types of contrast it's worth a quick reminder of what exactly we mean by the word 'contrast.' We can define contrast (C) quantitatively in terms of the *difference* in intensity (ΔI) between two adjacent areas

$$C = \frac{(I_2 - I_1)}{I_1} = \frac{\Delta I}{I_1} \quad (22.1)$$

In practice your eyes can't detect intensity changes <5% and even <10% is difficult. So unless the contrast from your specimen exceeds >5–10% you won't see anything on the screen or on the recorded image. However, if your image is digitally recorded, you can enhance low-contrast images electronically to levels at which your eyes can perceive the contrast. We'll return to image processing and contrast enhancement in Chapter 31.

So we see contrast in TEM images as different levels of green light coming from the viewing screen

or the computer display. On the photograph or computer screen, contrast is seen as different gray levels and our eyes can only discern about 16 of these. If we want to quantify the contrast, we need to make direct intensity measurements, e.g., via a microdensitometer if we're using film or directly from the CCD, but usually it's only necessary to see qualitative differences in intensity. Be careful not to confuse *intensity* with *contrast* when you describe your images. We can have strong or weak contrast but not bright or dark contrast. The terms 'bright' and 'dark' refer to density (number/unit area) of electrons hitting the screen or detector, and the subsequent light emission that we see. In fact, you generally get the strongest contrast under illumination conditions that lower the overall intensity. Conversely, you can try to increase the number of electrons falling on the screen by condensing the beam onto a reduced area of the specimen, but then you'll usually lower the image contrast. These points are summarized in Figure 22.1, which defines intensity and contrast.

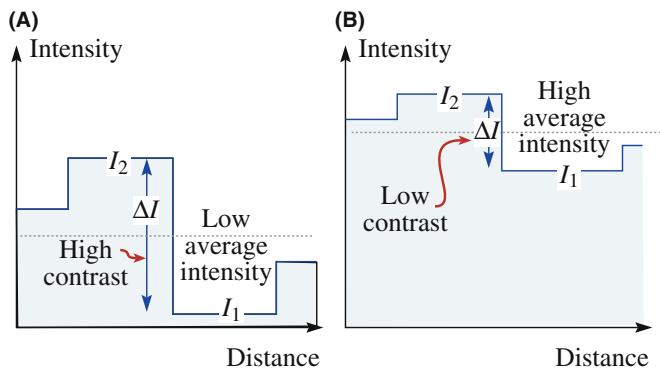


FIGURE 22.1. Schematic intensity profiles across an image showing (A) different intensity levels (I_1 and I_2) and the difference (ΔI) between them which defines the contrast. Generally, in a TEM, if the overall intensity is increased (B) the contrast decreases.

Before we discuss the two forms of amplitude contrast in detail, we need to remind you of the operational principles for creating amplitude contrast in your image. We obtain contrast in our images either by selecting specific electrons or excluding them from the imaging system. We have two choices: we can form either BF or DF images by selecting the direct or scattered electrons, respectively. So this chapter builds on what you learned about electron scattering in Chapters 2–4 and how to operate the TEM, described in Chapter 9.

22.2 AMPLITUDE CONTRAST

Amplitude contrast results from variations in mass or thickness or a combination of the two: the thickness variation can produce contrast because the electron interacts with more material (hence, more mass). Alternatively, diffraction can vary locally because the specimen is not a perfect, uniformly thin sheet.

22.2.A Images and Diffraction Patterns

If you look back at Figure 2.1, you'll see that the uniform electron intensity in the incident beam is transformed into a non-uniform intensity after scattering by the specimen. So a variable electron intensity hits the viewing screen or the electron detector, which translates into contrast on the screen. Now you also know that the DP shows you this non-uniformity because it separates the diffracted and direct beams. Therefore, a fundamental principle of imaging in the TEM is: first, *view the DP*, since this pattern tells you how your specimen is scattering. The relationship between the image and the DP is most critical for crystalline specimens showing diffraction contrast. However, you should view the DP first, no matter which contrast mechanism you want to use, or the specimen you are studying.

22.2.B Use of the Objective Aperture or the STEM Detector: BF and DF Images

In order to translate the electron scatter into interpretable amplitude contrast, we select *either* the direct beam or a diffracted beam in the SADP to form BF and DF images, respectively. (Remember that a small aperture will enhance the contrast but might decrease the resolution.) Note we are justified in using the 'beam' terminology, since the electrons have left the specimen. We've already seen in Section 9.3 that in a TEM we select the direct or a scattered electron beam with the objective aperture. Remember, if we form an image without the aperture, the contrast will be poorer (lower) because many beams then contribute to the image so we lose diffraction contrast. Furthermore, aberrations due to the off-axis electrons will make it impossible to focus your image. Your choice of the aperture size governs which electrons contribute to the image and you thus control the contrast.

Figure 22.2 shows a DP from a single-crystal Al specimen with two possible positions of an objective aperture indicated by white circles. In this figure, if the aperture is in position A it selects the direct beam only and thus a BF image will be formed in the image plane of the lens. This arrangement will produce amplitude contrast whether the specimen is crystalline (as in this case) or amorphous. If the aperture is in position B, it will select only electrons scattered in that specific direction. Thus a DF image will be formed. Traditionally (for 50 years), we have tilted the incident beam such that the scattered electrons remain on axis when forming a DF image. We thus create a centered dark-field (CDF) image, which we described back in Section 9.3. We'll discuss CDF techniques later in Section 22.5 and we'll usually assume CDF is the operational mode in DF imaging. However, you already know that it would be much better if we did not tilt the specimen but instead displaced the aperture (so the BF and DF are formed with the same diffraction conditions). We'll see later

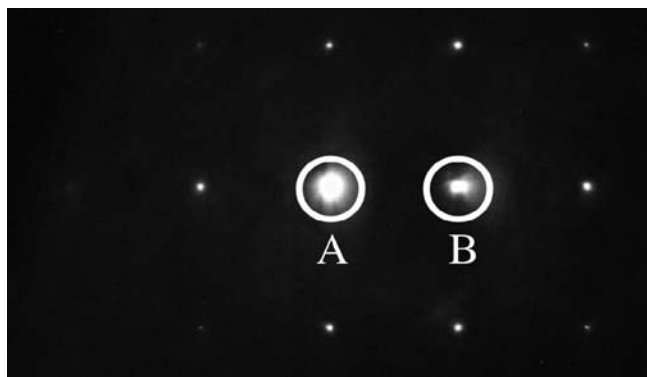


FIGURE 22.2. The relationship between the objective aperture and the diffraction pattern for forming (A) BF and (B) DF images. The circles show the location of the objective aperture.

that if you want to observe phase contrast, you have to use an objective aperture that is large enough to gather more than one beam.

BF AND DF

The two basic ways to form amplitude-contrast images.

In a STEM we select the direct or scattered beams in an equivalent way but use detectors rather than apertures. We compare the two different operational modes in Figure 22.3. Again, we saw back in Section 9.4 that we insert a BF on-axis detector, or an annular DF (ADF) detector, in a plane that's conjugate with the back focal plane. We control which electrons fall on which detector and thus contribute to the image by adjusting the post-specimen (imaging) lenses to change the camera length. Clearly, for DF imaging, the ADF detector will probably gather many more electrons than the objective aperture, which is good for imaging some specimens and bad for imaging others, as we'll see.

So, in summary, we can create BF or DF images with the direct beam or scattered beams, respectively. In

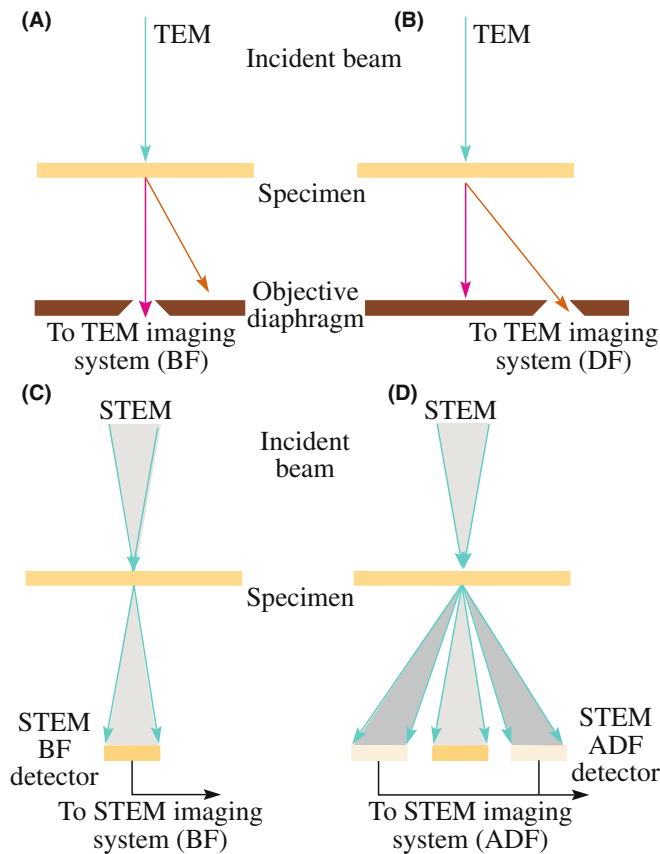


FIGURE 22.3. Comparison of the use of an objective aperture in TEM to select (A) the direct or (B) the scattered electrons forming BF and DF images, respectively. In STEM we use (C) an on-axis detector or (D) an off-axis annular detector to perform equivalent operations.

order to understand and control the contrast in these images you need to know what features of a specimen cause scattering and what aspects of TEM operation affect the contrast.

22.3 MASS-THICKNESS CONTRAST

Mass-thickness contrast arises from incoherent elastic scattering (Rutherford scattering) of electrons. As we saw back in Chapter 3, the cross section for Rutherford scatter is a strong function of the atomic number Z (hence the mass or the density, ρ) and the thickness, t , of the specimen. Rutherford scattering in thin specimens is strongly forward peaked. Therefore, if we form an image with electrons scattered at low angles ($< \sim 5^\circ$), mass-thickness contrast dominates (but it also competes with Bragg-diffraction contrast). However, we'll also see that at high angles ($> 5^\circ$), where Bragg scattering is usually negligible, we can pick up low-intensity, scattered beams. (The former is often referred to as coherent and the latter as incoherent; we'll discuss these terms later.) The intensity of these beams depends on atomic number (Z) only. Thus we can also get so-called Z -contrast, which contains elemental information like that in BSE images in the SEM. We can obtain these images with atomic resolution, particularly in a DSTEM. It is also feasible to form BSE images in a TEM but, because the specimen is thin, the number of BSEs is so small that the images are noisy and of poor quality, so no one does it. You shouldn't waste your money buying a BSE detector (but a secondary-electron detector can be very useful).

Mass-thickness contrast is most important if you are looking at non-crystalline materials, such as polymers and it is *the* critical contrast mechanism for biological scientists. But as we'll see, any variations in mass and thickness will cause contrast. As you learned in Chapter 10, it's almost impossible to thin a bulk sample uniformly (although a FIB can get close) and so nearly all real specimens will show some mass-thickness contrast. In some cases this will be the only contrast you can see.

In this section, we'll assume that there is no contribution to the image from diffraction contrast. This is automatically so if the specimen is amorphous. If the specimen is crystalline, then remove the objective aperture or use the ADF detector to minimize any diffraction contrast. You should still use an objective aperture to enhance the mass-thickness contrast to minimize the effects of lens aberrations. You'll still create BF and DF images of amorphous materials.

22.3.A Mechanism of Mass-Thickness Contrast

The mechanism by which differences in mass and thickness cause contrast is shown in Figure 22.4 and at this

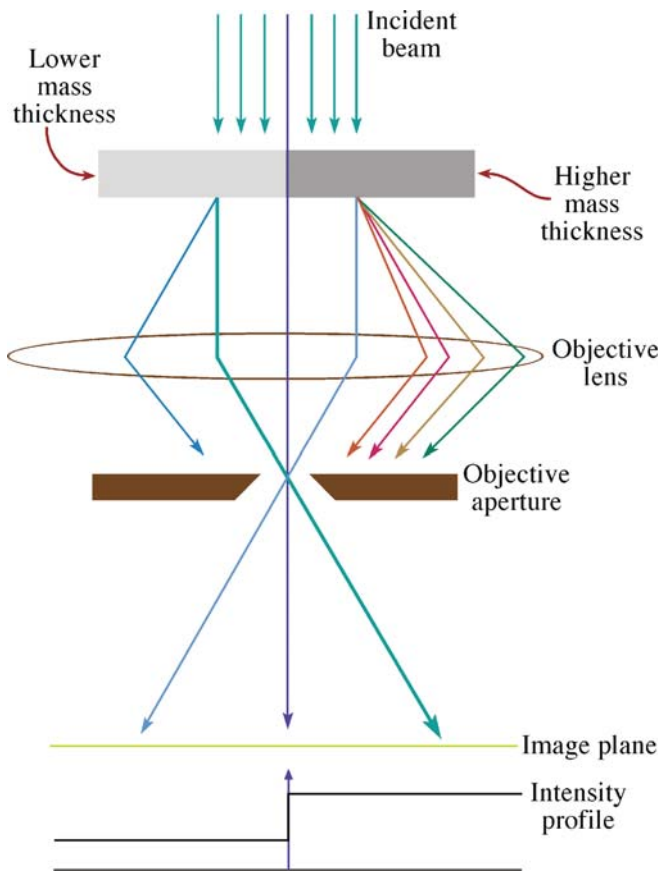


FIGURE 22.4. Mechanism of mass-thickness contrast in a BF image. Thicker or higher- Z areas of the specimen (darker) will scatter more electrons off axis than thinner, lower mass (lighter) areas. Thus fewer electrons from the darker region fall on the equivalent area of the image plane (and subsequently the screen), which therefore appears darker in BF images.

stage we'll talk about the process qualitatively. As electrons go through the specimen, they are scattered off axis by elastic nuclear interactions, i.e., Rutherford scattering. You know two factors from Chapter 3

- The cross section for elastic scattering is a function of Z .
- As the thickness of the specimen increases, there will be more elastic scattering because the mean free path remains fixed.

So, using a very simple, qualitative argument you would expect high- Z (i.e., high-mass) regions of a specimen to scatter more electrons than low- Z regions of the same thickness. Similarly, thicker regions will scatter more electrons than thinner regions of the same average Z , all other factors being constant. Usually, mass-thickness contrast images are interpreted in such a purely qualitative fashion, although we'll see a little later that it is possible to quantify the scattering intensity. So, as you can see from Figure 22.4, for the

case of a BF image, thicker and/or higher-mass areas will appear darker than thinner and/or lower-mass areas. The reverse will be true for a DF image.

This is all you need to know for the simplest interpretation of mass-thickness contrast images. Sometimes mass-thickness contrast is explained in terms of different amounts of electron absorption within the specimen and so you may come across the expression 'absorption contrast.' We think that this term is misleading, because in thin foils the actual amount of electron absorption is small; scattering outside the aperture or the detector, not absorption within the specimen, causes the contrast. For much the same reason, we prefer not to use the term 'structure-factor contrast' that is sometimes used to describe this phenomenon, since this implies a Bragg contribution, which may or may not be present.

However, you should be aware that if there are small crystals of different atoms in a given foil thickness, differences in their structure factor (F) from that of the matrix will cause contrast changes, since $I \propto |F|^2$. For example, you can detect the presence of nanometer-size clusters of Ag atoms in very thin foils of Al alloys in this way. Conversely, an absence of atoms (e.g., a void) will also scatter differently, although Fresnel contrast (see Chapter 23) is a better way to detect voids and bubbles.

Let's first look at a few images showing mass-thickness contrast and see which TEM variables you can control.

22.3.B TEM Images

Figure 22.5A is a TEM BF image of some latex particles on an amorphous-carbon support film. Assuming the latex is predominantly carbon, we have a constant Z and varying t . So the latex particles are darker than the support film since they are thicker. What you are basically seeing is a shadow projection image of the latex particles. Because it is a projection image, you cannot say that the particles are spheres (which in fact they are). They could equally well be disks or cylinders. One way to reveal the shape in a single image is to shadow-coat the particles, i.e., evaporate a thin heavy metal (Au or Au-Pd) coating at an oblique angle as shown in Figure 22.5B. The shape of the shadow reveals the true shape of the particles.

Shadowing introduces some mass contrast to what was just a thickness-contrast image. If we assume the Au-Pd film is very thin compared to the carbon support film, then the contrast across the edge of the shadow is predominantly mass contrast, due to the difference in average Z of the Au-Pd and the carbon film. If the spheres are small, there may also be an intensity change across the latex spheres due to the preferential deposit of Au-Pd on the side of the sphere towards the source of the evaporated metal.

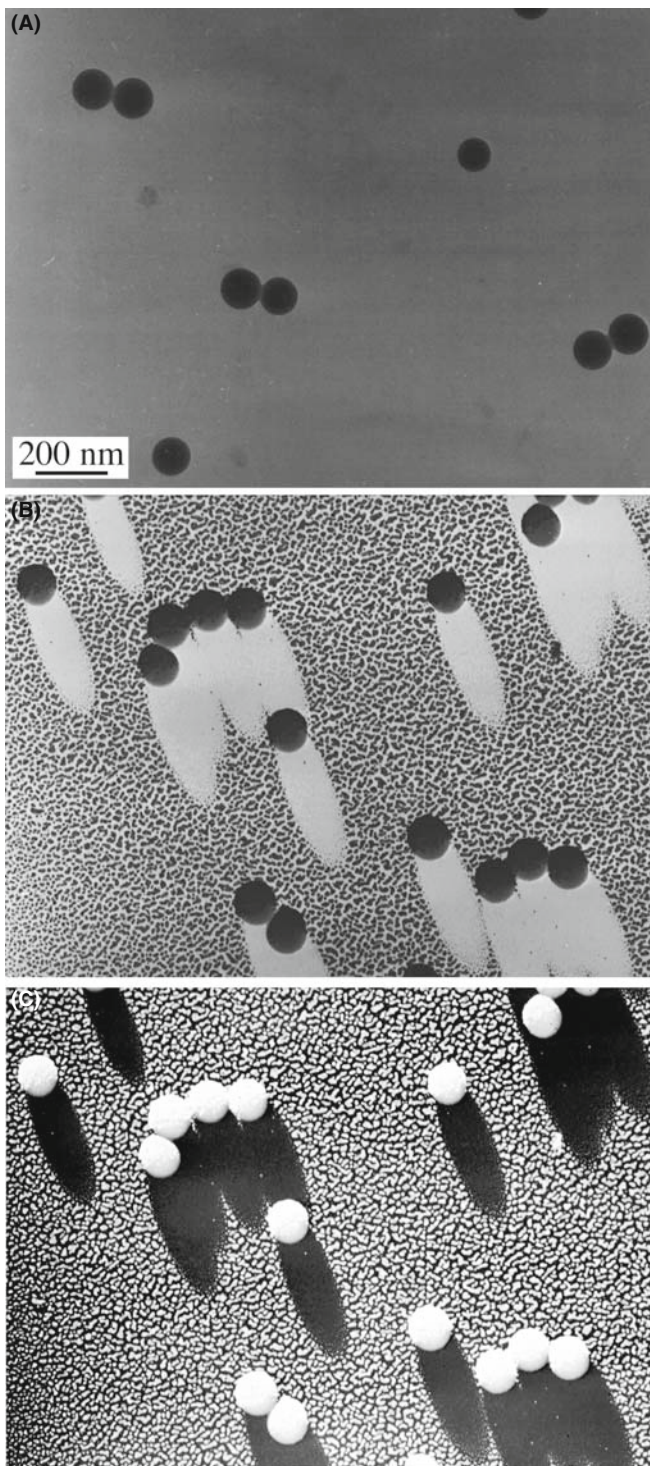


FIGURE 22.5. (A) TEM BF image of latex particles on a carbon support film showing thickness contrast only. (B) Latex particles on a carbon film shadowed to reveal the shape of the particles through the addition of selective mass contrast to the image. (C) Reverse print of (B) exhibits a 3D appearance.

It is an intriguing exercise to print Figure 22.5B in reverse (or take a DF image) as shown in Figure 22.5C. In this image, the latex spheres now appear to stand proud of the surface, even though you're still viewing a

two-dimensional projected image. Because the shadows are now dark your brain interprets the picture as it would a reflected-light image and endows it with a 3D nature. While the interpretation in this case is correct, it may not always be so. Once again we stress that you must be careful when interpreting 2D images of 3D specimens.

In addition to the use of shadowing to enhance mass-thickness contrast, it is common practice to stain different areas of polymer and biological specimens with heavy metals such as Os, Pb, and U. The stain leaves the heavy metal in specific regions of the structure (e.g., at unsaturated C = C bonds in a polymer and cell walls in biological tissue) and therefore these areas appear darker in a BF image. Figure 22.6 shows a BF image of a stained two-phase polymer. Since the specimen is of constant thickness (it was ultramicrotomed) the image shows mass contrast only.

TEM VARIABLES

The TEM variables that affect the mass-thickness contrast for a given specimen are the objective aperture size and the kV.

If you select a larger aperture, you allow more scattered electrons to contribute to the BF image. So the contrast between scattering and non-scattering areas is lowered, although the overall image intensity increases. If you choose a lower kV, both the scattering angle and the cross section increase. Hence, more electrons will be scattered outside a given aperture, hitting the diaphragm, and contrast will increase at the expense of intensity. The decrease in intensity will be worse for TEMs with a thermionic source because the gun brightness decreases as the kV is lowered. Figure 22.7 shows how a smaller aperture size results in improved contrast.

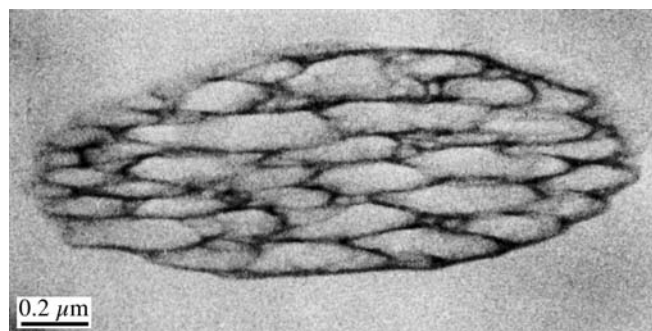


FIGURE 22.6. BF image of stained two-phase polymer exhibiting mass contrast due to the segregation of the heavy metal atoms to the unsaturated bonds in the darker phase.

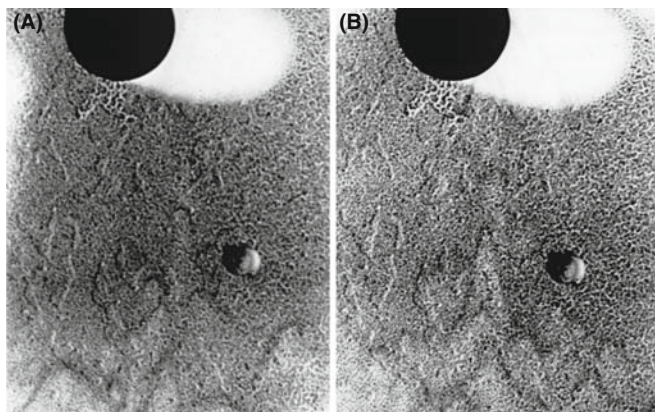


FIGURE 22.7. The effect of objective aperture size on mass-thickness contrast; the images of the shadowed latex particle were taken with an aperture size of (A) 70 μm and (B) 10 μm . A smaller aperture enhances the contrast, in a similar manner to lowering the kV.

Of course, any decrease in intensity can be offset by increased exposure times until specimen drift becomes a limiting factor.

Now for DF images, there isn't much more to be said: the images will generally show regions of contrast that are complementary to those seen in the BF (analogous to the reversed contrast in Figure 22.5B and C) but be prepared for exceptions (in Chapter 24). The overall intensity of the DF image will be much lower than the BF image (hence the relative terms 'dark' and 'bright') because the objective aperture will select only a small fraction of the scattered electrons. It's easy to remember that the BF image of a hole in your specimen will be bright and a DF image will be dark. However, remember that the corollary of low intensity is high contrast and DF images generally show better contrast.

22.3.C STEM Images

In a STEM you have more flexibility than in a TEM because by varying L , you change the collection angle of your detector and create, in effect, a variable objective aperture. (We'll mention the old TEMs later where you could do this and the SAD aperture was a triangle or a square.) So you have more control over which electrons contribute to the image. Even so, STEM BF images offer little more than TEM BF images. Generally, STEM images are noisier than TEM images (unless you've got an FEG STEM). Figure 22.8 shows a noisy STEM BF image of the same two-phase polymer as shown in the TEM image in Figure 22.6. The STEM images have generally shown poorer resolution because, with good thin specimens, the beam size dominates the resolution. To get reasonable intensity in a scanning image in reasonable time we have to use a large beam, as we discussed when we compared scanning and static images back in Chapter 9. Figure 22.9 shows the difference between (A) TEM and (B) STEM BF images from a low-contrast

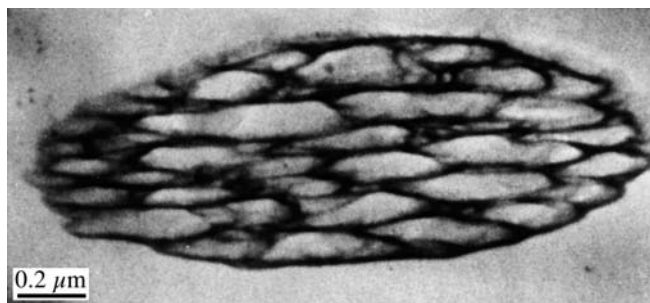


FIGURE 22.8. STEM BF image of a stained two-phase polymer. Comparison with the TEM image in Figure 22.6 shows that while the contrast is higher in STEM, the image quality is poorer.

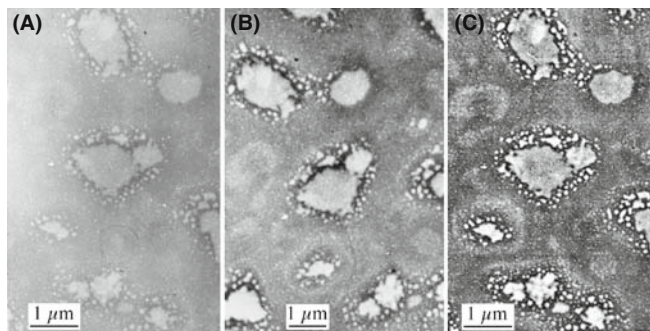


FIGURE 22.9. Comparison of TEM (A) and STEM (B) images of an amorphous SiO_2 specimen containing Cl-rich bubbles. The low mass contrast in the TEM can be enhanced in a STEM image through signal processing. (C) A similar effect can be achieved by digitizing the TEM image (A) and applying contrast-enhancement software.

specimen. The STEM image contrast has been enhanced and is considerably greater than in the TEM image, but the noise in the image is also more visible. However, if you record your TEM image using a CCD camera, or digitize the negative, you can enhance the contrast (see Chapter 31). A good way to do this is with one of the several kinds of image processing software (see Chapter 31) as you can see in Figure 22.9C.

STEM

STEM in a TEM is now a routine high-resolution technique.

In a STEM DF image, the scattered electrons fall onto the ADF detector. This gives rise to a fundamental difference between the TEM and STEM DF modes

- DF TEM images are usually formed by permitting only a fraction of the scattered electrons to enter the objective aperture.
- STEM images are formed by collecting most of the scattered electrons on the ADF detector.

CONTRAST IN STEM

Remember, you can always increase the contrast in the STEM image by adjusting the signal-processing controls, such as the detector gain and black level and the contrast and brightness controls on the computer screen; such options aren't available for analog TEM images.

Therefore, STEM ADF images are less noisy than TEM DF images, as shown in Figure 22.10. Because lenses aren't used to form the STEM image (although they are used to form the probe), the ADF images don't suffer aberrations, as would the equivalent off-axis TEM DF image.

STEM ADF image contrast is greater than TEM DF contrast: in STEM, L can be adjusted to maximize the ratio of the number of scattered electrons hitting the detector to the number of electrons going through the hole in the middle of the detector. You can thus improve the contrast quite easily, just by watching the computer screen and adjusting L .

However, as you can see from Figure 22.10, while the TEM DF image shows poorer contrast and is noisier, it still shows better resolution. Thermionic-source STEM images generally only show better resolution than TEM images when thick specimens are being imaged, because the chromatic aberration effects from thicker specimens do not affect the STEM images. If contrast is more important than resolution, then STEM is more useful. Indeed, in a STEM, you can study unstained polymer specimens which would show negligible contrast in a TEM.

DP AND STEM

The STEM must be well aligned so the DP expands and contracts on axis.

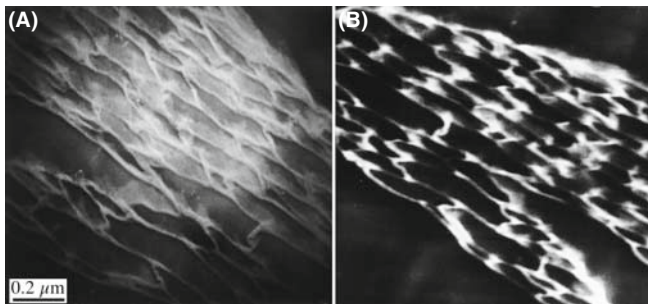


FIGURE 22.10. Comparison of (A) TEM DF and (B) STEM ADF images of the same two-phase polymer as in Figures 22.6 and 22.8. As in BF the STEM image shows higher contrast but lower resolution. Also, the ADF aperture collects more signal than the TEM objective aperture so the STEM image is less noisy.

STEM imaging is also useful if your specimen is beam sensitive, e.g., some polymers. A scanning beam lets you precisely control the irradiated region of the specimen, so it's a form of low-dose microscopy (see Section 4.6). You'll lose some image resolution unless you have access to a FEG STEM.

The comparison we've made of TEM and STEM images here is qualitative, but there have been many quantitative comparisons of STEM and TEM contrast, particularly for biological specimens. When STEMs were first introduced in the 1970s, the absence of chromatic aberration effects led to prophecies that STEM image resolution would invariably be better than TEM; there were even predictions of the end of classic TEM imaging! This hasn't happened because, as we'll see, there is more than just the chromatic aberration factor that governs the image quality, particularly for crystalline specimens. In summary, then, there are occasions when you might want to use STEM mass-thickness contrast images

- The specimen is so thick that chromatic aberration limits the TEM resolution.
- The specimen is beam-sensitive.
- The specimen has inherently low contrast in TEM and you can't digitize your TEM image or negative.
- Your specimen is ideally suited for HRTEM by Z-contrast imaging.

We'll examine the latter in the next section and in Chapter 28, and in much more detail in the companion text.

This said, be aware that STEM in the TEM has greatly improved since many of the comparison studies were carried out. Second, low-dose techniques in TEM will improve in the future. Third, you must be able to digitize your negatives now—negatives will be obsolete in the near future.

22.3.D Specimens Showing Mass-Thickness Contrast

Mass-thickness is the primary contrast source in amorphous materials, which is why we've illustrated this chapter mainly with polymer specimens. Replicas also display thickness contrast (see Figure 22.11A). Remember from Chapter 10 that replicas recreate the specimen topography, e.g., for a fracture surface. The amorphous-carbon replica can be as prepared (without shadowing) (Figure 22.11A) or shadowed (Figure 22.11B). The uneven metal shadowing increases the mass contrast and thus accentuates the topography, as for the latex particles in Figure 22.3. An extraction replica (Figure 22.11C) or a collection of particles dispersed on a support film will also show mass-thickness contrast; shadowing could be useful to reveal the shape of these particles too. If the particles are crystalline there will also be a component of diffraction contrast. Don't shadow if you want to do elemental analysis of the particles.

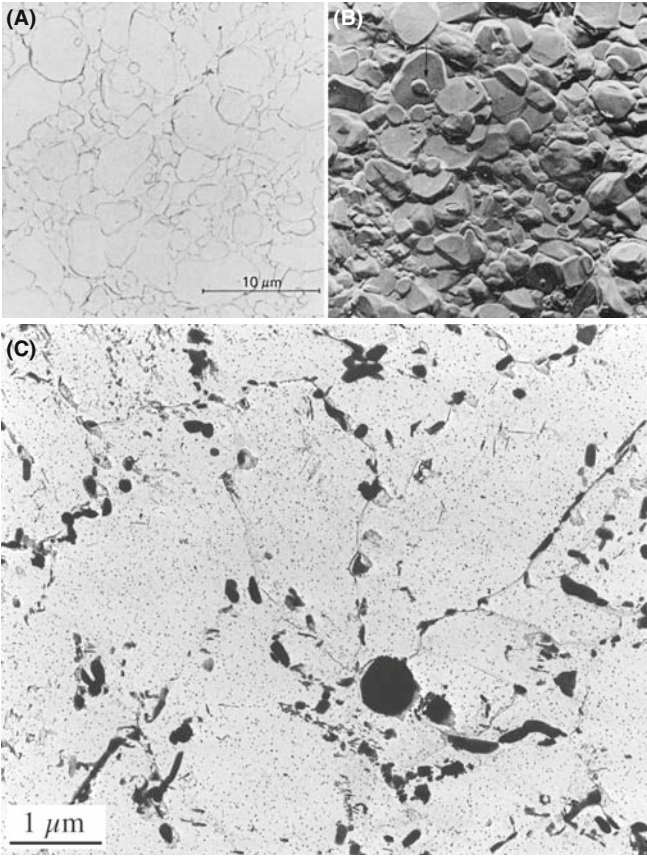


FIGURE 22.11. More examples of mass-thickness contrast: (A) a carbon replica of a fracture surface doesn't show much of either form of contrast until (B) oblique shadowing enhances the topography. (C) An extraction replica of small precipitate particles in a Cr–Mo steel weld shows both mass and thickness contrast.

22.3.E Quantitative Mass-Thickness Contrast

Because mass-thickness contrast is governed by Rutherford scattering, we can use the equations given back in Chapter 3 to predict the effect of Z and t on the scattering angle, θ , and the effect of kV on the cross section. We assume that the atoms scatter independently. (We then say that the scattering is truly incoherent.) This is actually not really the case, since even DPs from amorphous specimens show diffuse rings rather than uniform intensity (Figure 2.13A). Nevertheless, we'll still assume incoherent scattering.

As we stated at the start of this chapter, the contrast C is given by $\Delta I/I$ and it can be shown that a change in thickness, Δt , at constant atomic number Z creates contrast

$$\frac{\Delta I}{I} = 1 - e^{-Q\Delta t} \quad (22.2)$$

Q is the total elastic scattering cross section. So C becomes $Q\Delta t$ for $Q\Delta t < 1$. If 5% is the minimum contrast we can see, then the minimum Δt that we can see is

$$\Delta t \cong \frac{5}{100Q} = \frac{5A}{100N_0\sigma\rho} \quad (22.3)$$

where A is the atomic weight, N_0 is Avogadro's number, σ is the single-atom scattering cross section, and ρ is the density.

A similar argument can be made if there is a ΔZ (in which case σ or ρ changes). So, if we want to calculate the contrast, we need to know σ . As we've seen in equation 3.8, for low-angle scattering, the differential Rutherford cross section is equal to $f(\theta)^2$ where $f(\theta)$ is the atomic scattering factor, given by equation 3.9

$$f(\theta) = \frac{\left(1 + \frac{E_0}{m_0c^2}\right)}{8\pi^2a_0} \left(\frac{\lambda}{\sin\frac{\theta}{2}}\right)^2 (Z - f_x) \quad (22.4)$$

The Z term represents the Rutherford scattering. For unscreened Rutherford scattering (where we ignore the effects of the electron cloud), σ is therefore proportional to Z^2 . This unscreened behavior is quite a good approximation for electrons scattered through angles (remember, we mean semiangles when we talk about scattering) above $\sim 5^\circ$ (e.g., for Cu) although it is dependent on E_0 and Z . At lower angles, scattering becomes increasingly screened (less dependent on Z), and more dominated by inelastic scattering and diffraction. There is no precise angle which we use to define the transition from low- to high-angle scatter but the effect of screening effectively disappears at angles $> \theta_0$, the screening parameter, defined back in equation 3.4.

We can use the atomic scattering factor (equation 22.4) to determine the probability that an electron will be scattered through greater than a given angle. To do this, we integrate $|f(\theta)|^2$ from an angle β (defined by the angle of collection of the objective aperture) to infinity. Thus

$$\sigma(\beta) = 2\pi \int_{\beta}^{\infty} |f(\theta)|^2 \theta d\theta \quad (22.5)$$

which can be evaluated to give

$$\sigma(\beta) = \frac{\left[Z\lambda \left(\frac{a_0}{Z^{0.33}} \right) \left(1 + \frac{E_0}{m_0c^2} \right) \right]^2}{\pi(a_0)^2 \left(1 + \left(\frac{\beta}{\theta_0} \right)^2 \right)} \quad (22.6)$$

where a_0 is the Bohr radius and θ_0 is the characteristic screening angle; all the other terms have their usual meaning (see Chapter 3). So in equation 22.6 you can see directly the effect of Z and kV on electron scatter and hence on contrast. As we've already described, higher- Z specimens scatter more while lowering E_0 increases scattering. The effect of thickness is deduced

from the mean free path for elastic scatter, λ (which is inversely proportional to σ). So, thicker specimens scatter more.

Let's assume that n_0 electrons are incident on the specimen and dn electrons are scattered through an angle $>\beta$. Then, from equation 22.6, ignoring any inelastic scattering (which isn't really reasonable, but we'll do it to simplify matters), the reduction in the number of electrons going through the objective aperture to form the BF image is given by

$$\frac{dn}{n} = -N\sigma(\beta)dx \quad (22.7)$$

where $N = N_0/A$ and N_0 is Avogadro's number, $\sigma(\beta)$ is given by equation 22.6, and $x = \rho t$. So this expression gives the dependence of the contrast on Z and t . If we integrate

$$\ln n = -N\sigma x + \ln n_0 \quad (22.8)$$

and if we rearrange this expression

$$n = n_0 e^{-N\sigma x} \quad (22.9)$$

which describes the exponential decrease in the number of scattered electrons (n) as the specimen mass-thickness ($x = \rho t$) increases.

As you'll have gathered, this equation is something of an approximation but it does give you a feel for the factors that control mass-thickness contrast. For a given specimen, the variables are local changes in Z and t ; within the microscope, the variables are β and E_0 , which you can control to change the contrast, as we saw in Figure 22.7.

In principle, you could use these equations and equation 22.1 to calculate the expected contrast arising from differences in Z or t and see if they were detectable at the 5% contrast level. In practice, however, image-contrast calculations are not carried out for simple mass-thickness contrast in materials specimens.

22.4 Z-CONTRAST

Z-contrast is the name given to a high-resolution (atomic), mass-thickness (Z), imaging technique. We'll talk about it here because it represents the limit of mass-thickness contrast where detectable scattering arises from single atoms or column of atoms.

Back in the 1970s, early FEG STEMs demonstrated the remarkable capability of imaging single heavy atoms (e.g., Pt and U) on low- Z substrates as shown in Figure 22.12. These images were formed by the ADF detector collecting low-angle elastically scattered electrons only. Single atoms scatter incoherently and the image intensity is the sum of the individual atomic scattering

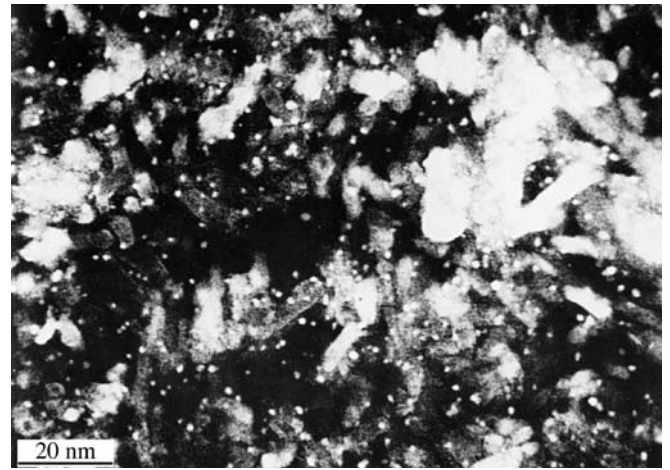


FIGURE 22.12. Z-contrast ADF image of individual Pt atoms or groups of atoms on a crystalline Al_2O_3 film obtained using an FEG STEM.

contributions. There was sometimes a problem with thickness changes in the substrate and contributions to the ADF signal from inelastically scattered electrons. This problem was overcome by dividing the digital ADF signal by the inelastic (energy-loss) signal from the EELS system. A drawback to this technique is that diffraction contrast (e.g., from a crystalline substrate) is preserved in the low-angle EELS signal, which can confuse the image interpretation. In Figure 22.12, the large bright regions arise from the Al_2O_3 substrate diffracting onto the ADF detector, obscuring the scatter from the Pt atoms.

HAADF

The detector is called a high-angle ADF or HAADF detector. Z-contrast images are also termed HAADF images. The outer diameter of the Fischione HAADF detector is 28 mm; the inner diameter is 4 mm.

Because of Bragg scattering, this early approach to Z-contrast was not suited to the study of crystalline specimens. Since the normal ADF detector will always collect some Bragg electrons, it was necessary to design an ADF detector with a very large central aperture. Z-contrast images could then be formed from thin crystals (Figure 22.13). You can decrease the camera length with the post-specimen lenses to ensure that the Bragg electrons (including any HOLZ scatter) don't hit the detector. Thus only the electrons scattered through very high angles contribute to the image.

Bragg effects are avoided if the HAADF detector only gathers electrons scattered through an angle of >50 mrad ($\sim 3^\circ$). Remember that cooling your specimen has the effect of increasing coherent HOLZ scatter, so don't cool it unless you must. Electron channeling effects

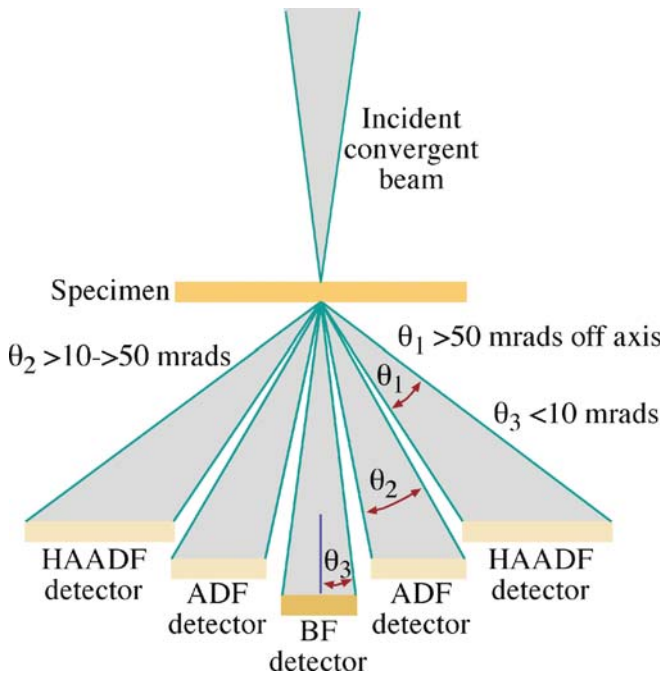


FIGURE 22.13. Schematic of the HAADF detector setup for Z-contrast imaging in a STEM. The conventional ADF and BF detectors are also shown along with the range of electron scattering angles gathered by each detector.

remain at high scattering angles, so imaging away from strong two-beam conditions and closer to zone-axis orientations is wise.

So, what do these Z-contrast images of crystals look like? Figure 22.14 shows a TEM BF image of Bi-implanted Si and below is a Z-contrast image. In the TEM BF image, formed from the direct beam, defects associated with the Bi implant are shown (we'll talk about such diffraction contrast from defects in Chapter 26) but otherwise there is no contrast associated with the Bi. In the Z-contrast image the Bi-implanted area is bright, but note that defect contrast isn't preserved in this image. You can relate the intensity differences in Figure 22.14 to an absolute measure of the Bi concentration. To do this you need to choose a suitable elastic scattering cross section. The contrast is related directly to the cross section for elastic scatter by the matrix (σ_A) and the alloying or dopant element (σ_B)

$$C = \left(\frac{\sigma_A}{\sigma_B} - F_B \right) c_B \quad (22.10)$$

where c_B is the atomic concentration of the alloying element and F_B is the fraction of the alloying element that substitutes for matrix atoms. The intensity can be quantified to an absolute accuracy of better than $\pm 20\%$.

In an FEG with probe sizes of < 0.3 nm, Z-contrast image resolution of close to the probe diameter is possible. Figure 22.15A shows a high-resolution

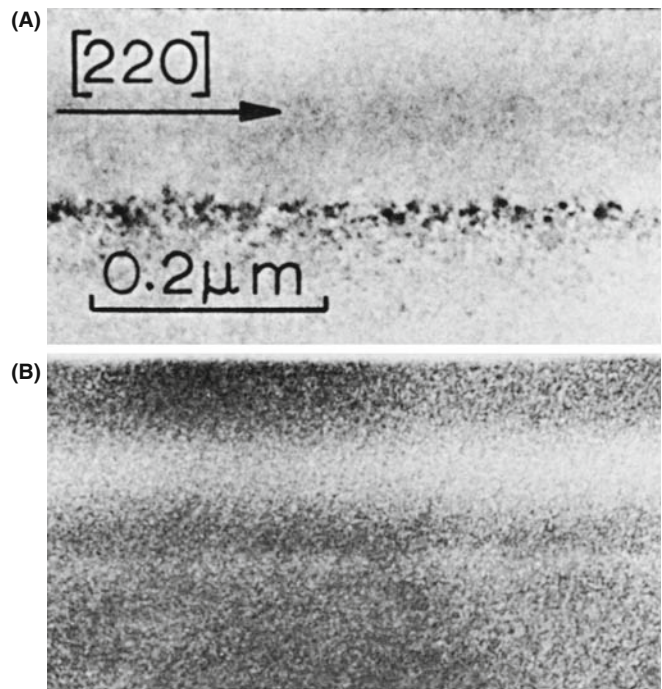


FIGURE 22.14. (A) Low-resolution TEM BF image showing a row of defects in Bi-implanted Si. In (B) obtained under Z-contrast conditions, the defects associated with the implant are invisible but the specimen is bright in the region implanted with Bi.

phase-contrast TEM image of Ge on Si with an amorphous SiO_2 surface layer. The Si and Ge are indistinguishable by phase contrast. In Figure 22.15B, which is a STEM Z-contrast image of the same region, the higher-Z Ge crystal region is clearly visible and the lower-Z SiO_2 layer appears very dark. The atomic structures of the Si and Ge crystals are visible in both phase-contrast and Z-contrast images, although the Z-contrast image is noisier. Phase-contrast TEM images can show similar Z-contrast effects, as we'll detail in Chapter 28. Figure 22.15C shows a model of a grain boundary superimposed on a Z-contrast image which has been refined and processed to reduce the noise via a maximum-entropy approach. You can easily see atomic-level detail.

HAADF

HAADF has the advantage that the contrast is generally unaffected by small changes in objective lens defocus (Δf) and specimen thickness.

We'll see in Chapter 28 that interpretation of atomic-resolution phase-contrast images requires knowledge of t and Δf . Some microscopists claim that Z contrast will become the principal method of high-resolution imaging in the future as more FEG STEMs become available; others strongly disagree!

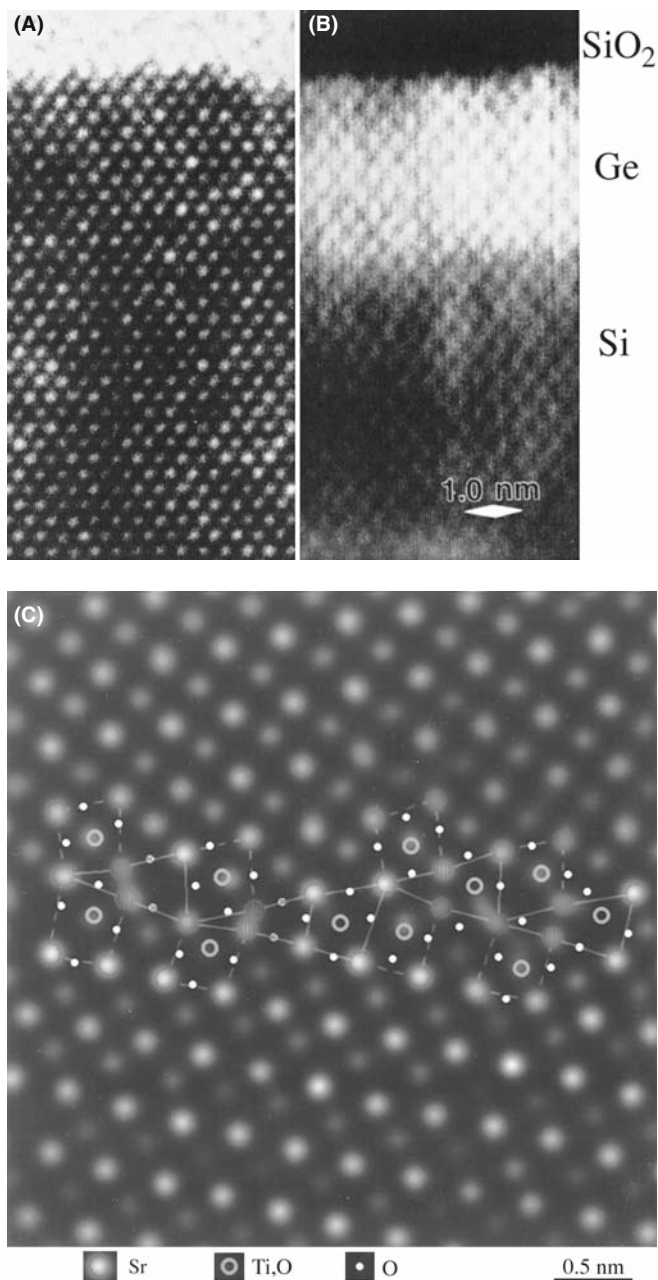


FIGURE 22.15. (A) High-resolution phase-contrast image of epitaxial Ge on Si with an amorphous SiO₂ surface. The bright array of dots common to the crystalline region represents atomic rows and the Ge and Si regions are indistinguishable. (B) The high-resolution Z-contrast STEM image shows the atom rows but with strong contrast at the Si-Ge interface and low intensity in the low-Z oxide. (C) Model structure of a boundary in SrTiO₃ superimposed on a processed Z-contrast image.

We can think of the image in Figure 22.15B as a direct map of the $f(\theta)$ variation in the specimen. In that respect, it is similar to an X-ray map showing the distribution of a certain element.

RESOLUTION

The $f(\theta)$ map can have atomic-level resolution, which XEDS imaging can't provide (yet!).

So why do we need a STEM for Z-contrast imaging? We are constrained in TEM if we use an analog screen rather than a digital detector to form the image. Nevertheless, we can do Z-contrast imaging in a TEM but we have to create electron-optical conditions which are equivalent to those used in STEM. So the beam-convergence angle in TEM must equal the collection angle of the HAADF detector. This is an example of the so-called 'principle of reciprocity' which we'll discuss in more detail in the next section. To converge the TEM beam to the required angular range, we use so-called 'hollow-cone' illumination, which requires an annular C2 aperture. However, the highest incidence angles possible in hollow-cone illumination are typically a few mrad rather than the 50–150 mrad (up to $\sim 9^\circ$) collected by the STEM HAADF detector. So TEM Z-contrast images are not equivalent to STEM and will always contain some diffraction contrast from crystalline specimens. This leads us into the topic of diffraction contrast, which is the other form of amplitude contrast we see in TEM images.

The future (see Section 39.10): Z-contrast images can be collected at the same time and from the same location as the EELS data. There is no other combination of techniques to match this, if it gives the information you need.

22.5 TEM DIFFRACTION CONTRAST

Bragg diffraction, as we discussed in Part 2, is controlled by the crystal structure and orientation of the specimen. We can use this diffraction to create contrast in TEM images. Diffraction contrast is simply a special form of amplitude contrast where the scattering occurs at special (Bragg) angles. We've just seen how incoherent elastic scattering causes mass-thickness contrast. Now we'll see how coherent elastic scattering produces diffraction contrast. As you know, crystalline specimens usually give a single-crystal DP, such as in Figure 22.2. So, as for mass-thickness contrast, we can form BF images by placing the objective aperture around the direct beam (Figure 22.2A) and DF images by selecting any of the diffracted beams (Figure 22.2B). Remember that the incident electrons must be parallel in order to give sharp diffraction spots and strong diffraction contrast. So, if you can, underfocus C2 to spread the beam.

22.5.A Two-Beam Conditions

There is one major difference between forming images to show mass-thickness contrast or diffraction contrast. We can use *any* scattered electrons to form a DF image showing mass-thickness contrast. However, to get good strong diffraction contrast in both BF and DF images

we tilt the specimen to *two-beam conditions*, in which only one diffracted beam is strong. Of course, the direct beam is the other strong spot in the pattern.

Remember: the electrons in the strongly excited *hkl* beam have been diffracted by a *specific* set of *hkl* planes and so the area that appears bright in the DF image is the area where the *hkl* planes are at the Bragg condition. Hence the DF image contains *specific* orientation information, not just general scattering information, as is the case for mass-thickness contrast.

We can tilt the specimen to set up several different two-beam conditions. Figure 22.16A includes a zone-axis DP from a single-crystal specimen in which the beam direction is [011]. The surrounding patterns are a series of two-beam conditions in which the specimen has been tilted slightly so that different *hkl* spots are strongly excited in each pattern. We can form DF images from each strongly diffracted beam after tilting the specimen, and each will give a different image.

TWO-BEAM CONDITIONS

If you're working with crystalline materials, you'll spend a lot of time tilting the specimen to set up different two-beam conditions.

As you can see in Figure 22.16B and C, the BF and DF images show complementary contrast under two-beam conditions. We'll explain the image contrast in detail in Chapter 23. Obviously, to set up a series of two-beam conditions, we need precise tilt control, which explains why a double-tilt eucentric holder is the holder of choice for viewing crystalline specimens.

We'll see in the following chapters that two-beam conditions are not only necessary for good contrast but they also greatly simplify interpretation of the images. This is why we emphasized two-beam theory in our discussion of diffraction in Part 2.

22.5.B Setting the Deviation Parameter, s

Setting up two-beam conditions is very simple. While looking at the DP, tilt around until only one diffracted beam is strong, as in Figure 22.16. As you can see, the other diffracted beams don't disappear because of the relaxation of the Bragg conditions, but they can be made relatively faint. Now if you just do as we've described, the contrast might still not be the best. For reasons we'll discuss in detail in the next chapter, to get the best contrast from defects, your specimen shouldn't be *exactly* at the Bragg condition ($s = 0$) as in Figure 22.17A. Tilt your specimen close to the Bragg condition but make s small and positive. (The excess *hkl* Kikuchi line, just outside the *hkl* spot; go back and check the text for Figures 19.10 and 19.11.) This will give you the best

possible strong-beam image contrast as in Figure 22.17B. If you tilt the specimen slightly, so s increases further as shown in Figure 22.17C, the defect images become narrower but the contrast is reduced.

USE $s > 0$

Never form strong-beam images with s negative; it's easier to see the defects when $s > 0$.

22.5.C Setting Up a Two-Beam CDF Image

We described the basic mechanism of forming BF and DF images back in Chapter 9 (Figure 9.14A). To produce the best BF diffraction contrast, tilt to the desired two-beam condition as in Figure 22.18A, and insert the objective aperture on axis as in Figure 22.2A. A two-beam CDF image is not quite as simple. You might think it involves just tilting the incident beam so the strong *hkl* reflection moves onto the optic axis. If you do that, you'll find that the *hkl* reflection becomes weaker as you move it onto the axis and the $3h3k3l$ reflection becomes strong, as shown in Figure 22.18B. What you've just done is in fact set up a *weak-beam* image condition, which we'll discuss in Chapter 27. To set up a *strong-beam* CDF image, tilt in the $\bar{h}\bar{k}\bar{l}$ reflection which was initially weak, and it becomes strong as it moves on axis, as shown in Figure 22.18C. The CDF technique is absolutely crucial for obtaining and interpreting diffraction-contrast images, so we will take you through it in detail

- Look at the SADP and tilt the specimen until the desired *hkl* reflection is strong. Make sure the incident beam is well underfocused.
- Now tilt the specimen until the $\bar{h}\bar{k}\bar{l}$ reflection is strong: *hkl* will now be weak.
- Use the DF tilt controls to move the 000 reflection towards the strong $\bar{h}\bar{k}\bar{l}$ reflection. The weak *hkl* reflection will move towards the optic axis and become strong.
- When *hkl* is close to the axis, switch off the DF deflectors, insert and carefully center the objective aperture around 000.
- Switch the DF tilt coils on and off while looking through the binoculars. Check that the *hkl* and 000 reflections appear in the same position. Make fine adjustments to the DF coils until you can see no shift between 000 and *hkl* when the deflectors are off and on, respectively.
- Switch to image mode. If necessary, condense the beam slightly with C2 until you can see the CDF image. If you can't see an image, either the *hkl* reflection is too weak (unlikely) or your tilt coils are misaligned (common). In the latter case, realign the coils (see the manufacturer's handbook).

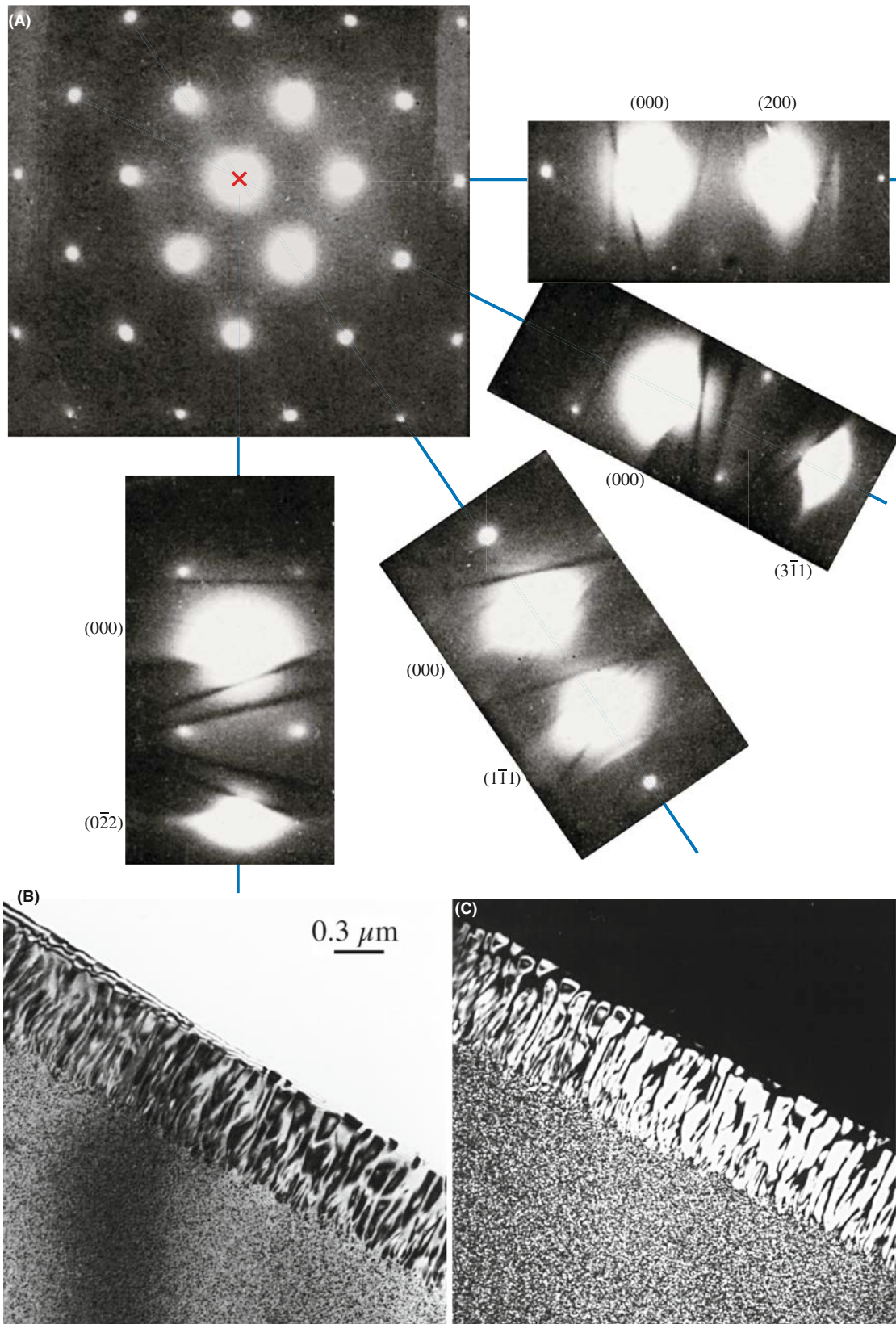


FIGURE 22.16. (A) The $[011]$ zone-axis diffraction pattern has many planes diffracting with equal strength. In the smaller patterns, the specimen is tilted so there are only two strong beams, the direct 000 on-axis beam and a different one of the hkl off-axis diffracted beams. Complementary (B) BF and (C) DF images of Al-3 wt% Li taken under two-beam conditions are shown also. In (B) the Al_3Li precipitate phase (present as tiny spheres in the grain and coarse lamellae at the boundary) is diffracting strongly and appears dark. In (C), imaged with a precipitate spot, only the diffracting precipitates appear bright.

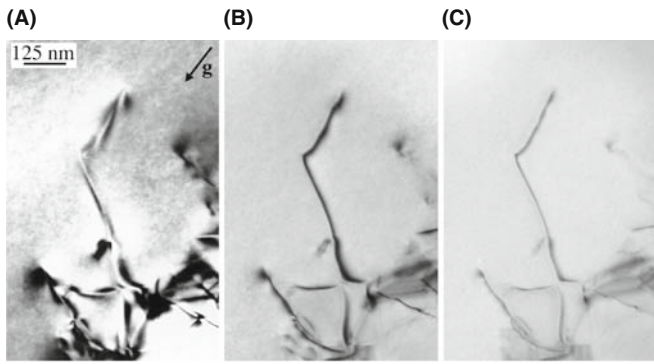


FIGURE 22.17. Variation in the diffraction contrast when s is varied from (A) zero to (B) small and positive and (C) larger and positive.

Now go back and study Figure 9.14C carefully. You'll see that the beam was tilted through an angle $2\theta_B$ to bring the weak beam in Figure 9.14B onto the optic axis.

Notice that \bar{g} is excited in the DF condition; g was excited in the BF condition. If you want to form a CDF image with g excited, you must tilt the beam *and* the specimen. Alternatively, just displace the aperture but be cautious.

22.5.D Relationship Between the Image and the Diffraction Pattern

From what we've just described, there is clearly an important relationship between the DP and a diffraction-

contrast image. If we change the DP in any way, the contrast in the image will change. So it is critical to relate the DP to the image. We need to indicate the direction of the g vector in the image. To relate the two, remember that you may have to calibrate the rotation between the image and the DP if, whenever you change magnification, your image rotates but your DP does not. We described this calibration in Section 9.6.C. You should usually show the g vector in any BF or DF two-beam, diffraction-contrast image after correcting for any rotation between the image and the DP. Remember: you should always check the calibration of any microscope, especially a TEM, and be careful with possible 180° rotations.

ROTATION CALIBRATION

Do it even if only to convince yourself that you don't need to!

We will expand on diffraction contrast in far more detail in the subsequent chapters, making use of the fundamental operational principles we have just described.

22.6 STEM DIFFRACTION CONTRAST

The principle of forming BF and DF images in STEM is just the same as for mass-thickness contrast; i.e., use the BF detector to pick up the direct beam and the ADF

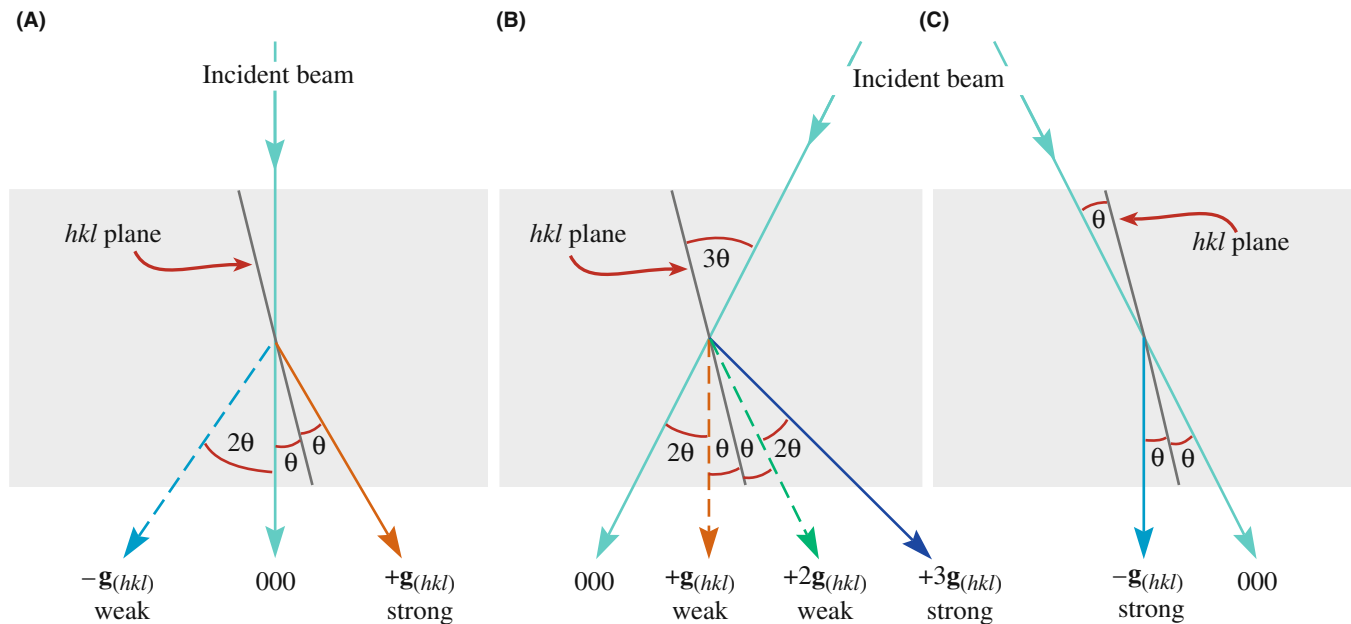


FIGURE 22.18. (A) Standard two-beam conditions involve the 000 spot and the hkl spot bright because one set of hkl planes are exactly at the Bragg condition. (B) When the incident beam is tilted through 2θ so that the excited g_{hkl} spot moves onto the optic axis, the g_{hkl} intensity decreases because the g_{3h3k3l} spot becomes strongly excited. (C) To get a strong hkl spot on axis for a CDF image, it is necessary to set up a strong $-g_{hkl}$ condition first of all, then tilt the initially weak g_{hkl} maximum onto the axis.

detector to pick up the diffracted beams. To preserve two-beam conditions, the ADF detector must only pick up one strong diffracted beam and this can be ensured by inserting the objective aperture and selecting only one diffracted beam. Alternatively, the DP could be displaced so the chosen hkl reflection falls on the BF detector. Either way the computer screen will display a DF image.

However, the diffraction contrast observed in the STEM image will generally be much poorer than TEM contrast; the normal STEM operating conditions are not equivalent to the TEM conditions that ensure strong diffraction contrast. To understand the contrast in STEM images you need to know the beam convergence and detector collection angles. It's rare in fact that you'll need to do this, but we showed you how to determine the beam convergence angle back in Section 5.5. To calculate the collection angle, you need to carry out a similar exercise as we use to determine the EELS spectrometer collection angle in Section 37.4.

Remember, there are three conditions that must be fulfilled for strong contrast in your image

- The incident beam must be coherent, i.e., the convergence angle must be very small.
- The specimen must be tilted to a two-beam condition.
- Only the direct beam or the one strong diffracted beam must be collected by the objective aperture.

This condition is shown schematically in Figure 22.19A. We define the TEM convergence angle as α_T and the objective aperture collection angle as β_T . In a STEM, the equivalent angles are the beam convergence angle α_S and the STEM detector collection angle β_S as shown in Figure 22.19B. Therefore, we have identical operating conditions if

$$\alpha_T = \alpha_S \quad (22.11a)$$

$$\beta_T = \beta_S \quad (22.11b)$$

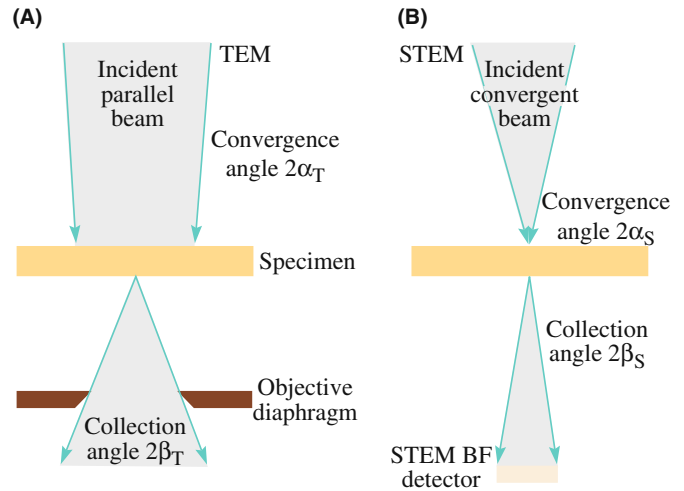


FIGURE 22.19. Comparison of the important beam-convergence and divergence angles (A) in TEM and (B) in STEM. (Note that we show the full convergence and divergence angles, not the semiangles that are usually described in the text.)

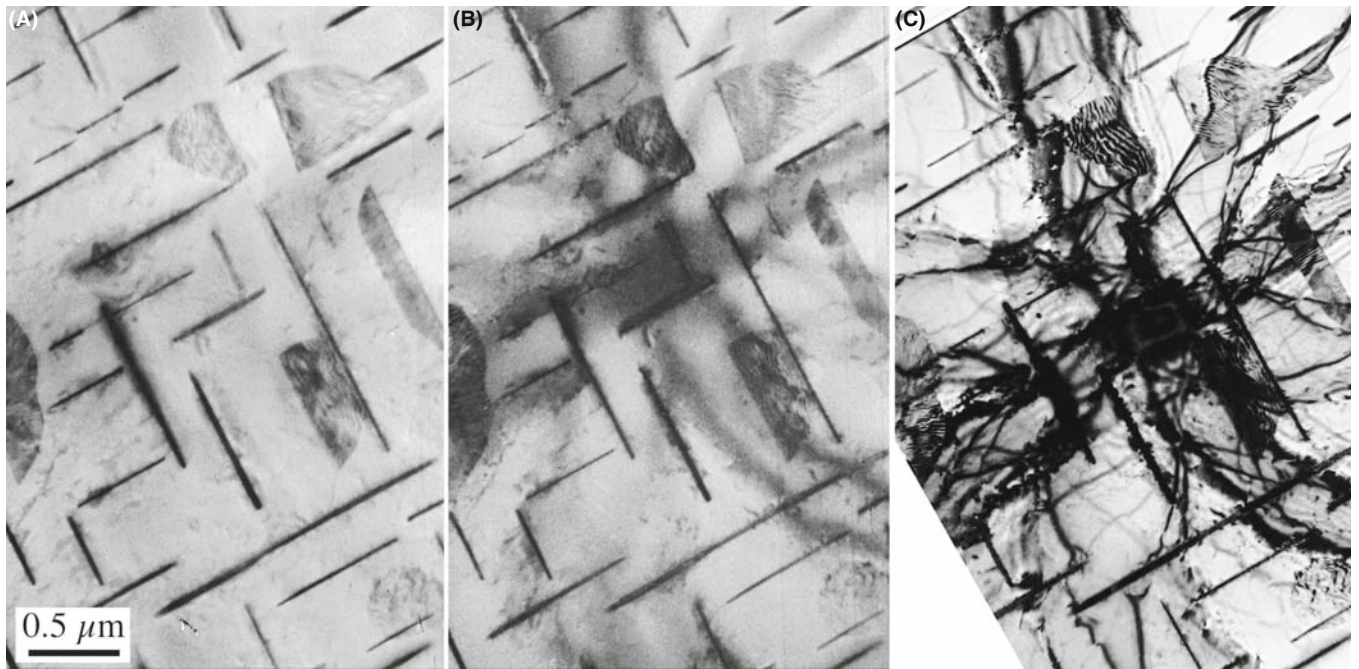


FIGURE 22.20. (A) A BF STEM image of an Al-4 wt% Cu specimen showing weak diffraction contrast in the form of bend contours. As the STEM detector collection angle is lowered (B), the diffraction contrast increases slightly at the expense of increased noise in the image. Even at the smaller collection angle, comparison with the contrast in the TEM image (C) is unfavorable. Note that the Cu-rich θ' precipitates maintain strong mass contrast in all the images.

Now it should be immediately clear that we can't get such equivalence in a STEM because the convergence angle of the beam is very much greater than in a TEM (since in STEM we deliberately create a convergent rather than a parallel beam). However, there is a way around this dilemma and it depends on a theorem that is often used in electron optics called the *principle of reciprocity*. In essence this principle says that so long as the electron ray paths contain equivalent angles (of convergence and collection) at some point in the electron optical system, the image contrast will be identical.

In other words, while the conditions in equations 22.11 can't be fulfilled, we can create conditions such that

$$\alpha_S = \beta_T \quad (22.12a)$$

$$\alpha_T = \beta_S \quad (22.12b)$$

Under these circumstances the electrons in TEM and STEM do see equivalent angular constraints, although not at the equivalent points of convergence and collection.

- Since the objective-aperture collection angle in TEM is about equal to the convergence angle in STEM, the first of this pair of equations is easily satisfied.
- To satisfy the second pair, we have to make a very small STEM collection angle β_S .

We can't simply increase α_T , because we must keep a parallel beam to get good TEM diffraction contrast and making the beam non-parallel (large α_T) destroys the contrast.

There is an obvious drawback to making β_S small. The signal falling on the STEM detector becomes very small and the STEM image becomes noisy. So STEM diffraction-contrast images become noisier as we attempt to increase the amount of diffraction contrast, as in Figure 22.20. (See Chapter 24 for an explanation of the contrast (bend contours) in this figure.) Having an FEG helps to offset this increase in noise, but in general, STEM diffraction-contrast images (in both BF and DF) compare so unfavorably with TEM images (see Figure 22.20C) that, while they may be useful if you're performing analysis, they are rarely used to show diffraction-contrast images of crystal defects. This is solely the domain of TEM, as we'll discuss in detail in the next few chapters.

CHAPTER SUMMARY

Mass-thickness contrast and diffraction contrast are two forms of amplitude contrast. Both arise because the specimen scatters electrons. The operational procedures to produce BF and DF images are identical. Interpretation of mass-thickness contrast is generally simpler than interpretation of diffraction contrast. In fact, the interpretation of diffraction contrast is sufficiently complex that we need to devote several subsequent chapters to the various forms arising in perfect and imperfect crystals.

We can summarize the characteristics of mass-thickness contrast

- Areas of greater Z and/or t scatter electrons more strongly (in total), and therefore appear darker in BF images and brighter in DF images. The contrast can be quantified if necessary.
- TEM mass-thickness contrast images are better quality (lower noise and higher resolution) than STEM images, but digital STEM images can be processed to show higher contrast than analog TEM images.
- STEM mass-thickness contrast images are most useful for thick and/or beam-sensitive specimens.
- Z -contrast (HAADF) images can show atomic-level resolution.

We can summarize the characteristics of diffraction contrast

- Diffraction contrast arises when the electrons are Bragg scattered.
- To form a diffraction-contrast image in TEM, the objective aperture selects one Bragg-scattered beam. Often, the STEM detectors gather several Bragg beams which reduce diffraction contrast.
- Diffraction-contrast images in TEM always show better contrast than in STEM images, which are always noisier and almost never used.

Some of the statements made in the First Edition are not now valid! FEGs are now common. The original DSTEM was discontinued in ~ 1985 ; but DSTEMs are now made by Hitachi, JEOL, and Nion. C_s correctors are becoming more available. Modern TEMs all use digital recording. Z -contrast imaging has greatly improved on commercial machines and rivals CTEM in many applications.

REFERENCES

References on diffraction contrast are included in Chapters 24–27.

GENERAL IMAGING AND SOME HISTORY

The topic of staining is discussed and used much more for biological samples.

- Cosslett, VE 1979 *Penetration and Resolution of STEM and CTEM in Amorphous and Polycrystalline Materials* Phys. Stat. Sol. **A55** 545–548. Early review of using STEM in biological specimens by one of the pioneers of TEM.
- Heidenreich, RD 1964 *Fundamentals of Transmission Electron Microscopy* p 31 John Wiley & Sons New York. More details on the derivation of equation 22.2.
- Humphreys, CJ 1979 *Introduction to Analytical Electron Microscopy* p305 Eds. JJ Hren, JI Goldstein and DC Joy Plenum Press New York.
- Reimer, L 1997 *Transmission Electron Microscopy; Physics of Image Formation and Microanalysis* 4th Ed. Springer-Verlag New York. Includes a more detailed discussion of equation 22.6.
- Sawyer, LC, Grubb, DT and Meyers, DT 2008 *Polymer Microscopy* 3rd Ed. Springer New York. Gives a good introduction to staining for the materials scientist.
- Watt, IM 2003 *The Principles and Practice of Electron Microscopy* 2nd Ed. Cambridge University Press New York. See in particular the discussion on shadowing.

EARLY HAADF

- Brown, LM 1977 *Progress and Prospects for STEM in Materials Science* Inst. Phys. Conf. Ser. No. **36** 141–148. Diffraction effects in low-angle ADF.
- Donald, AM and Craven, AJ 1979 *A Study of Grain Boundary Segregation in Cu-Bi Alloys Using STEM* Phil. Mag. **A39** 1–11.
- Howie, A 1979 *Image Contrast and Localized Signal Selection Techniques* J. Microsc. **117** 11–23.
- Isaacson, M, Ohtsuki, M and Utlaut, M 1979 in *Introduction to Analytical Electron Microscopy* p 343 Eds. JJ Hren, JI Goldstein and DC Joy Plenum Press New York. Includes discussion of early Z-contrast imaging.
- Jesson, DE and Pennycook, SJ 1995 *Incoherent Imaging of Crystals Using Thermally Scattered Electrons* Proc. Roy. Soc. (Lond.) **A449** 273–293.
- Pennycook, SJ 1992 *Z-Contrast Transmission Electron Microscopy: Direct Atomic Imaging of Materials* Annu. Rev. Mat. Sci. **22** 171–195.
- Treacy, MMJ, Howie, A and Pennycook, SJ 1980 *Z Contrast of Supported Catalyst Particles on the STEM* Inst. Phys. Conf. Ser. No. **52** 261–264.
- Treacy, MMJ, Howie, A and Wilson, CJ 1978 *Z Contrast of Platinum and Palladium Catalysts* Phil. Mag. **A38** 569–585. Early paper noting the advantage of higher-angle scattering in Z-contrast imaging.

THE COMPANION TEXT

The chapter on simulating images is particularly relevant but you must understand what you are simulating first. The companion text also includes a full chapter on HAADF.

SELF-ASSESSMENT QUESTIONS

- Q22.1 What do we mean by the term minimum contrast and can we quantify it?
- Q22.2 When considering contrast (either qualitatively or quantitatively), why do we immediately consider the DP?
- Q22.3 Will a TEM be able to resolve two atoms next to each other in the periodic table?
- Q22.4 You are shown a polymer sphere in a TEM image. Why would you be cautious?
- Q22.5 How would you prepare an unstained polymer sample for TEM imaging?
- Q22.6 Why is the resolution in STEM not as good as in TEM on most TEMs?
- Q22.7 Why is resolution on some STEMs as good as, or better than, many TEMs?
- Q22.8 How are diffraction-contrast images different from mass-thickness contrast images?
- Q22.9 No matter how hard a particular student tries to obtain a two-beam condition, he or she still sees the direct beam, a strongly diffracted beam, and several weak spots in the DP. Could the student be at the two-beam condition?
- Q22.10 What is the difference between a phase-contrast image and an amplitude-contrast image?
- Q22.11 What causes mass-thickness contrast?
- Q22.12 What influences the intensity of mass-thickness contrast?
- Q22.13 For what materials is mass-thickness contrast most useful?
- Q22.14 What microscope controls can affect the mass-thickness contrast?
- Q22.15 What features of a STEM give more flexibility in mass-thickness contrast imaging?
- Q22.16 What is diffraction contrast?

- Q22.17 How can you obtain good diffraction contrast in BF and DF images?
Q22.18 How can you obtain the best strong-beam contrast from defects?
Q22.19 How are Z-contrast imaging and HAADF imaging related?
Q22.20 Does a BF STEM image show strong contrast from bend contours? Explain after you've read Chapter 24.

TEXT-SPECIFIC QUESTIONS

- T22.1 Where is the Ewald sphere in Figure 22.2? Give a full explanation.
T22.2 How was the specimen prepared in Figure 22.12 and why is this information important?
T22.3 Redraw Figure 22.13 showing all the angles and distances to scale and with typical values in mrad and degrees. Are the values of θ correct/reasonable/sensible? Explain your reasoning carefully.
T22.4 What do you guess is the geometry of the specimen shown in Figure 22.14?
T22.5 Draw an atomistic diagram to explain the contrast seen in Figure 22.15B.
T22.6 Draw and label direction vectors on Figure 22.15C for both grains.
T22.7 Index all the spots in the diffraction pattern in Figure 22.16.
T22.8 Explain why the bend contours in STEM images are much less pronounced than in TEM diffraction-contrast images.
T22.9 In Figure 22.17, why are some of the defects in C almost out of contrast?
T22.10 Suggest how the contrast in Figure 22.16C is influenced by the fact that the Al_3Li precipitate on the grain boundary is a lamella.

Phase-Contrast Images

CHAPTER PREVIEW

We see phase contrast any time we have more than one beam contributing to the image. In fact, whenever we say “fringes,” we are essentially referring to a phase-contrast phenomenon. Although we often distinguish phase and diffraction contrast, this distinction is generally artificial. For example, in Chapters 24 and 25, we will examine thickness fringes and stacking-fault fringes; both types of contrast result from interference of waves so both are phase-contrast images although we usually think of them as two-beam, diffraction-contrast images.

Phase-contrast imaging is often thought to be synonymous with high-resolution TEM. In fact, phase contrast appears in most TEM images even at relatively low magnifications. We will draw your attention to its role in the formation of moiré patterns and Fresnel contrast at defects. This Fresnel contrast has the same origin as the Fresnel fringes that we used in Chapter 9 to correct the astigmatism of the objective lens.

As with many of the topics we’ve discussed, we can approach the problem at several different levels. One danger is that you may be tempted to use one of the prepackaged simulation programs to predict phase-contrast images, without first learning about the limitations of such packages. The second danger is that with a modern HRTEM you can very easily obtain phase-contrast images that show great detail without you fully knowing/understanding the conditions used to obtain the image. This may impress your advisor but you should know better.

We will begin this chapter by discussing some simple approaches to understanding phase-contrast effects as they relate to lattice-fringe imaging.

23.1 INTRODUCTION

Contrast in TEM images can arise due to the differences in the phase of the electron waves scattered through a thin specimen. This contrast mechanism can be difficult to interpret because it is very sensitive to many factors: the appearance of the image varies with small changes in the thickness, orientation, or scattering factor of the specimen, and variations in the focus or astigmatism of the objective lens. However, its sensitivity is the reason phase contrast can be exploited to image the *atomic structure* of thin specimens. Of course this also requires a TEM with sufficient resolution to detect contrast variations at atomic dimensions and the proper control of instrument parameters that affect the phases of the electrons passing through the specimen and the lenses. If you know what you are doing, the procedures can be straightforward; the level of operator skill that is necessary to obtain such images can be acquired with practice; the challenge is in getting enough practice and knowing when you have!

The most obvious distinction between phase-contrast imaging and other forms of TEM imaging is the number

of beams collected by the objective aperture or an electron detector. As described in the previous chapters, a BF or DF image requires that we select a single beam using the objective aperture. A phase-contrast image requires the selection of *more than one* beam. In general, the more beams collected, the higher the resolution of the image. However, we will see that there are reasons why some beams, which are apparently admitted through the aperture, might not contribute to the image. The details of this process depend on the performance of the electron-optical system. We’ll first examine the theory and then consider the practical aspects.

23.2 THE ORIGIN OF LATTICE FRINGES

We can understand the origin of lattice fringes by extending the analysis of Chapter 13 to allow the two beams, $\mathbf{0}$ and \mathbf{g} , to interfere; i.e., use the objective aperture to select only two beams. We begin by rewriting equation 13.5

$$\psi = \phi_0(z) \exp 2\pi i(\mathbf{k}_I \cdot \mathbf{r}) + \phi_g(z) \exp(2\pi i\mathbf{k}_D \cdot \mathbf{r}) \quad (23.1)$$

where we know

$$\mathbf{k}_D = \mathbf{k}_I + \mathbf{g} + \mathbf{s}_g = \mathbf{k}_I + \mathbf{g}' \quad (23.2)$$

We are thus using a two-beam approximation but allowing \mathbf{s}_g to be non-zero. Now we will make some simple substitutions setting $\phi_0(z) = A$ and take $e^{2\pi i \mathbf{k}_I \cdot \mathbf{r}}$ out as a factor. We will also represent the expression for ϕ_g from equation 13.5 as

$$\phi_g = B \exp i\delta \quad (23.3)$$

where

$$B = \frac{\pi \sin \pi t s_{\text{eff}}}{\xi_g \pi s_{\text{eff}}} \quad (23.4)$$

and

$$\delta = \frac{\pi}{2} - \pi t s_{\text{eff}} \quad (23.5)$$

The $\pi/2$ in the expression for δ takes care of i in equation 13.5 and we'll pretend that the specimen is so thin that we can replace s_{eff} by s . Thus equation 23.1 becomes

$$\psi = \exp(2\pi i \mathbf{k}_I \cdot \mathbf{r}) [A + B \exp i(2\pi \mathbf{g}' \cdot \mathbf{r} + \delta)] \quad (23.6)$$

The intensity can then be expressed as

$$I = A^2 + B^2 + AB[\exp i(2\pi \mathbf{g}' \cdot \mathbf{r} + \delta) + \exp -i(2\pi \mathbf{g}' \cdot \mathbf{r} + \delta)] \quad (23.7)$$

$$I = A^2 + B^2 + 2AB \cos(2\pi \mathbf{g}' \cdot \mathbf{r} + \delta) \quad (23.8)$$

Now \mathbf{g}' is effectively perpendicular to the beam so we'll set it parallel to x and replace δ giving

$$I = A^2 + B^2 - 2AB \sin(2\pi g' x - \pi s t) \quad (23.9)$$

Therefore, the intensity is a sinusoidal oscillation normal to \mathbf{g}' , with a periodicity that depends on s and t . (Note that g and s are not bold in equation 23.9 because they represent the magnitude of the vectors, not the vectors themselves.) We can, with care, relate these fringes to the spacing of the lattice planes normal to \mathbf{g}' . Although we have obtained this equation using a very simple model, it gives us some useful insight, which will also be helpful when we talk about many-beam images in Chapter 28.

The intensity varies sinusoidally with different periodicities for different values of \mathbf{g}' . This model will be equally valid even if the incident beam is tilted slightly off the optic axis.

DELOCALIZATION

This simple analysis shows that the location of a fringe does not necessarily correspond to the location of a lattice plane.

23.3 SOME PRACTICAL ASPECTS OF LATTICE FRINGES

23.3.A If $\mathbf{s} = 0$

If we just have $\mathbf{0}$ and \mathbf{g} in the objective aperture and we then set $\mathbf{s} = 0$ for reflection G (so $\mathbf{g}' = \mathbf{g}$), we will see fringes in the image (Figure 23.1A) which have a periodicity of $1/g$ in the x direction; i.e., the fringe periodicity is the same as the spacing of the planes which give rise to \mathbf{g} . This result holds wherever $\mathbf{s} = 0$ no matter how $\mathbf{0}$ and \mathbf{g} are located relative to the optic axis, even if the diffracting planes are not parallel to the optic axis.

Figure 23.1B shows the ideal geometry for producing images like Figure 23.1A. It is called the 'tilted-beam condition' and it means that the planes of interest lie parallel to the optic axis. If we use the geometry shown in Figure 23.1B we have $\mathbf{s} = 0$ and the planes are parallel to the optic axis but not parallel to the incident beam. Therefore the fringes cannot correspond directly to the individual planes. If we use the on-axis geometry shown in Figure 23.1C, the planes are viewed edge on, but $\mathbf{s} \neq 0$ for reflection G ; so we must also consider reflection $-G$.

CONFUSING

The trouble is that the fringes look so like atomic planes that we can be easily misled into thinking that they are atomic planes.

23.3.B If $\mathbf{s} \neq 0$

If the specimen is not exactly flat, then \mathbf{s} will vary across the image; even if you set $\mathbf{s} = 0$ in the DP, it will not be

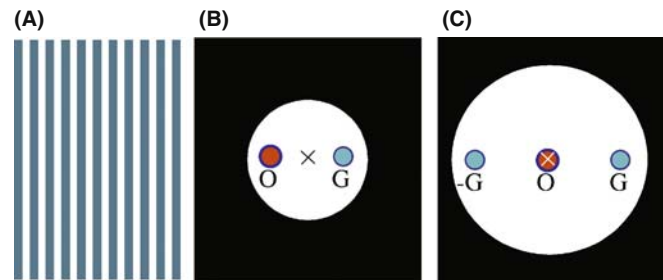


FIGURE 23.1. (A) Schematic tilted-beam 111 lattice fringes in Si formed using the O and G beams symmetrically displaced relative to the optic axis; \mathbf{g} is normal to the fringes. (B) Ideal diffraction geometry to produce tilted-beam fringes. (C) On-axis, three-beam geometry.

zero everywhere. If s is not zero, then the fringes will shift by an amount which depends on both the magnitude of s and the value of t , but the periodicity will not change noticeably. We expect this s dependence to affect the image when the foil bends slightly, as is often the case for thin specimens. We also expect to see thickness variations in many-beam images, since s may be non-zero for all of the beams; s may also vary from beam to beam.

23.4 ON-AXIS LATTICE-FRINGE IMAGING

We've just seen that two beams can interfere to give an image with a periodicity related to $|\Delta\mathbf{g}|^{-1}$. Since one beam is the direct beam, $|\Delta\mathbf{g}|^{-1}$ is just d , the interplanar spacing corresponding to \mathbf{g} . If you align your beam parallel to a low-index zone axis then you'll see fringes running in different directions; these fringes in the image must correspond to an array of spots in the DP. The spacings of the spots may be inversely related to the lattice spacings, as shown in Figure 23.2 which extends Figure 23.1 to the many-beam case. In general, this array of spots bears no direct relationship to the *position* of atoms in the crystal.

We'll see more on this when we discuss image simulation in Chapter 30. In case you are in doubt, compare the beautiful image shown in Figure 23.3A with the projected structure of Si in Figure 23.3B. The Si dumbbells are a pair of atoms which are 1.4 Å apart in this projection of the structure. The aperture used to form the image included 13 reflections, as shown in Figure 23.3C. The difficulty is that in the image, the spots in the dumbbell image are really only 1.3 Å apart but the point resolution of the TEM was only ~ 2.5 Å. You can see from the

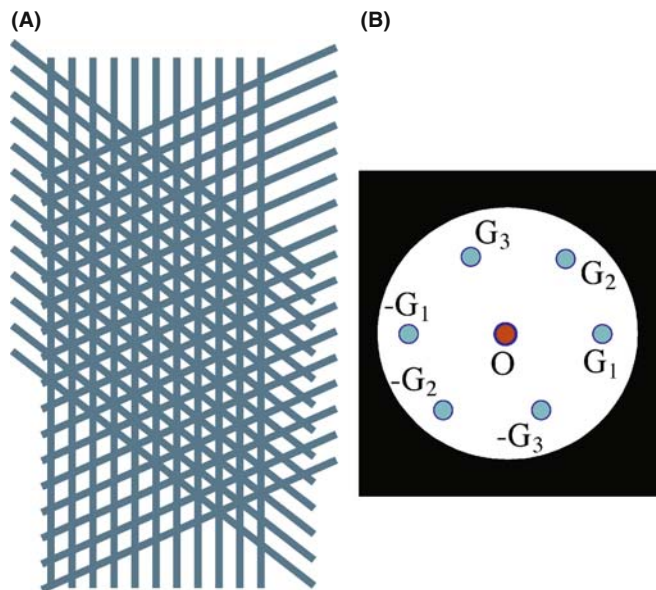


FIGURE 23.2. (A) Schematic many-beam image showing crossing lattice fringes and (B) the diffraction pattern.

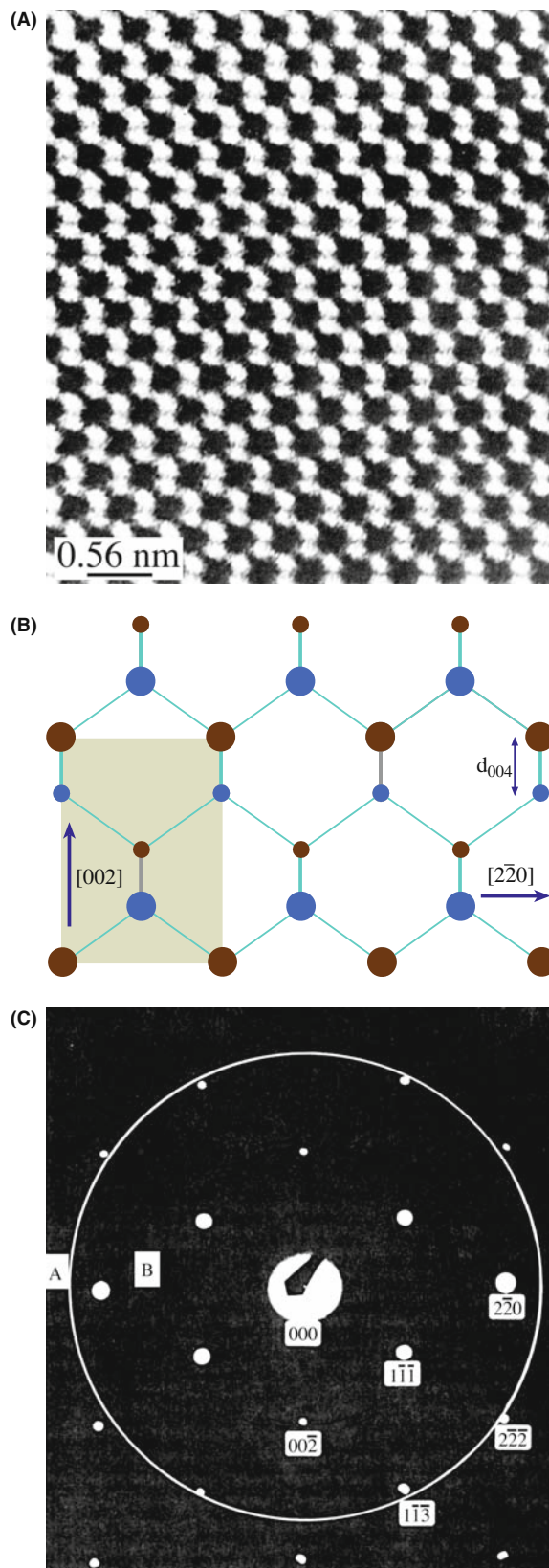


FIGURE 23.3. (A) On-axis image of a perfect Si crystal; (B) the projected structure; (C) the diffraction pattern showing the 13 spots used to form the image inside the aperture (ring). The Si dumbbells do not correspond to the closely spaced pairs of spots in the image.

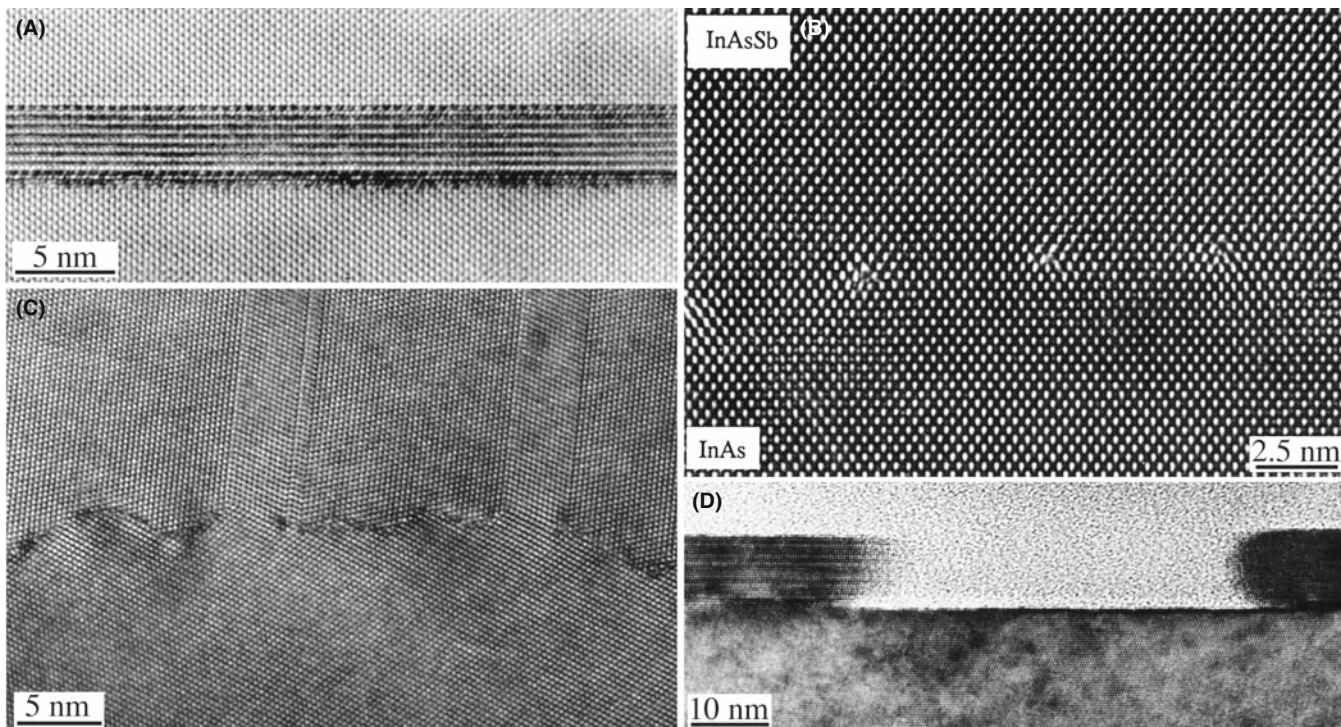


FIGURE 23.4. Illustrations of lattice images which contain easily interpreted information. (A) The spinel/olivine interface; (B) dislocations at a heterojunction between InAsSb and InAs; (C) a grain boundary in Ge faceting on an atomic scale; (D) a profile view of a faceted surface.

structure that the real dumbbell spacing corresponds to the (004) plane spacing, but the 004 reflection was *not* used to form the image. The explanation was given by Krivanek and Rez; the dumbbells in the image are caused by the crossing {113} fringes—so the contrast is an artifact, if we are interested in the real structure.

The lesson is: we only knew the image did not correspond to the structure because we knew the structure! Taking this example as your guide, consider the case where a defect is present in an image where the perfect crystal spots are all in the ‘correct’ position. Could you still be certain that the detail in the image close to the defect gives you a true picture of the location of the atoms close to the defect? The answer is, of course, “no.”

FRINGES NOT PLANES

So lattice fringes are not direct images of the structure, but just give you information on lattice spacing and orientation.

On-axis lattice-fringe images are perhaps best used as a measure of the local crystal structure and orientation. The exception, as we’ll see in the next chapter, is when these images can only be interpreted using extensive computer simulation. Figure 23.4 illustrates some typical applications of the imaging mode, where we can learn a lot about our material by intuitive interpretation

without the need for simulating their images. Our guess is that at least 99% of all lattice-fringe images are interpreted without using image simulation.

Figure 23.4A shows the interface between a spinel particle and an olivine matrix; Figure 23.4B shows how we can locate dislocations at a heterojunction; Figure 23.4C shows the atomic-scale faceting of a grain boundary in Ge; and Figure 23.4D illustrates the faceting of a surface.

23.5 MOIRÉ PATTERNS

Moiré (pronounced mwa-ray) patterns can be formed by interfering two sets of lines that have nearly common periodicities. We can demonstrate two fundamentally different types of interference: the rotational moiré and the translational (often referred to as misfit) moiré. It’s easy to understand moirés if you make three transparent sheets of parallel lines (two with the same spacing and one slightly different): you can generate such sets of lines readily using any computer, choosing the line widths to be similar to the gaps between them. Then try these three exercises. (You should ‘do’ the experiment by making your own ‘specimen’ or using those on the Web site.)

- Take two misfit sets and align them exactly. This gives a set of moiré fringes which are parallel to the lines forming them, as shown in Figure 23.5A.

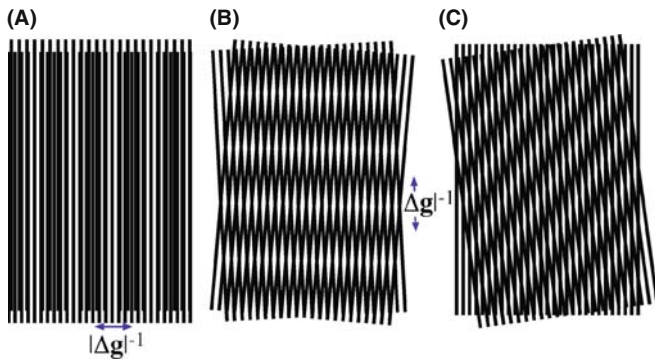


FIGURE 23.5. (A) Translation moiré fringes; (B) rotation moiré fringes; (C) mixed moiré fringes; note the relationship between the fringes and their constituent lattices.

- Take two identical sets of lines and rotate them. Now, we produce a set of moiré fringes which is perpendicular to the average direction of the initial lines (Figure 23.5B).
- Take the first two sets and rotate them so we produce moiré fringes as in Figure 23.5C; but note that their alignment to your reference sets is not obvious.

When the misfit or misorientation is small, the moiré fringe spacing is clearly much coarser than that of the lines themselves. In particular, if the sets of lines in Figure 23.5 are actually lattice planes in a crystal, the moiré fringes may give information about the crystals even if you cannot resolve the lattice planes. The simplest way to analyze the spacings and orientation of the moiré fringes is to consider the diffraction vectors from the two ‘lattices.’ Incidentally, the term ‘moiré’ originated in the textile industry; it’s related to the French word for ‘mohair,’ the silky hair of the Angora goat, hence the watery or wavy pattern seen in silk fabrics—and the lowercase ‘m’!

23.5.A Translational Moiré Fringes

In this case, since the planes are parallel, the \mathbf{g} -vectors will also be parallel. If we write these as \mathbf{g}_1 and \mathbf{g}_2 , we produce a new spacing \mathbf{g}_{tm} given by

$$\mathbf{g}_{tm} = \mathbf{g}_2 - \mathbf{g}_1 \quad (23.10)$$

In Figure 23.6A, we have assigned \mathbf{g}_2 to the smaller ‘lattice’ spacing and tm indicates ‘translational moiré’ fringes. The vector \mathbf{g}_{tm} corresponds to a set of fringes with spacing d_{tm} , as shown by the following simple manipulation

$$d_{tm} = \frac{1}{g_{tm}} = \frac{1}{g_2 - g_1} = \frac{\frac{1}{g_2} \cdot \frac{1}{g_1}}{\frac{1}{g_1} - \frac{1}{g_2}} = \frac{d_2 d_1}{d_1 - d_2} = \frac{d_1}{1 - \frac{d_2}{d_1}} \quad (23.11)$$

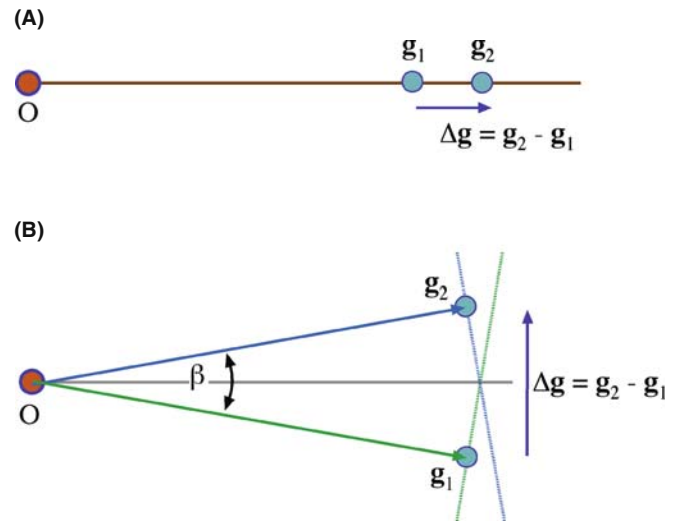


FIGURE 23.6. (A) Relationship between \mathbf{g} -vectors and translational moiré fringes; (B) relationship for rotational moiré fringes.

23.5.B Rotational Moiré Fringes

We follow the same procedure as above, but now the two \mathbf{g} -vectors are identical in length and rotated through an angle β so that the new \mathbf{g} -vector, \mathbf{g}_{rm} , has length $2g \sin \beta/2$, as shown in Figure 23.6B. The fringe spacing is then

$$d_{rm} = \frac{1}{g_{rm}} = \frac{1}{2g \sin \beta/2} = \frac{d}{2 \sin \beta/2} \quad (23.12)$$

23.5.C General Moiré Fringes

If we use the same approach to locate \mathbf{g}_{gm} (gm: ‘general moiré’) we can readily show that, for small misorientation, the spacing d_{gm} of our fringes is given by

$$d_{gm} = \frac{d_1 d_2}{\left((d_1 - d_2)^2 + d_1 d_2 \beta^2 \right)^{1/2}} \quad (23.13)$$

23.6 EXPERIMENTAL OBSERVATIONS OF MOIRÉ FRINGES

Moiré fringes in TEM images were first reported early in the history of the microscope. They were used by Minter to identify a dislocation before lattice imaging was possible. Later, they were regarded as an imaging artifact that obscured the true dislocation structure in twist boundaries. Most recently, there has been renewed interest due to the widespread development of thin films grown on different substrates.

You must be wary of the limitation or pitfall of using moiré fringes to learn about interfaces and defects. Moiré patterns result purely from the interference of two ‘sets of planes.’ Their appearance will be essentially the same even if the two ‘crystals’ are not in contact.

In TEM the moiré patterns correspond to interference between a pair of beams, \mathbf{g}_1 and \mathbf{g}_2 . If \mathbf{g}_1 is generated in the upper crystal and \mathbf{g}_2 in the lower, then each reflection \mathbf{g}_1 in crystal 1 acts as an incident beam for the lower crystal and produces a ‘crystal-2 pattern’ around each \mathbf{g}_1 reflection, as shown in Figure 23.7A. This process is another example of double diffraction, as discussed in Section 18.9. Figure 23.7A is from a pair of

perfectly aligned but misfitting cubic crystals viewed along their common [001] zone axis; the pattern is indexed in Figure 23.7B. When we have many planes diffracting at a zone axis, as in this pattern, we expect to see crossed moiré fringes.

In the following three sections we will discuss examples of the use of moiré fringes.

23.6.A Translational Moiré Patterns

When a continuous film is grown on a thick substrate, one question which is asked is: “Do the lattice parameters of the thin film correspond to the values of the same material in bulk form?” For example, a thin film of a cubic material on an (001) substrate may be tetragonally strained so that the a_{film} lattice parameter is smaller than a_{bulk} , but the c_{film} parameter is larger. If the bulk material has its bulk lattice parameter, then the measurement of d_{tm} , the translational moiré spacing, can give a very accurate value for a_{film} . Furthermore, we can tilt the specimen 45° or 60° and deduce a value for c_{film} to estimate the tetragonal distortion directly.

Tilting the specimen can also give us information about misfitting islands, as illustrated in Figure 23.8. In this case, we see a hexagonal array of fringes when the two pseudo-hexagonal materials are viewed parallel to their common c -axis. The variation in the contrast of the moiré fringes around the edge of the particle occurs because the particle facets on inclined planes, as is confirmed when we tilt the specimen. In this system, when the islands are grown on different substrates, they may still grow as platelets. In Figure 23.8B and C the platelet is thick in the direction of the beam but, when tilted over, we again see moiré fringes. In particular, we can see moiré fringes at the *top* of the platelet.

INTERFERENCE NOT STRUCTURE
 We know that the top of an inclined island is not in contact with the substrate yet it shows fringes; so this reminds us that moiré fringes do not tell us about the interface structure!

23.6.B Rotational Moiré Patterns

We often see rotational moiré fringes at twist boundaries, as illustrated for Si in Figure 23.9. A complicating factor is that the misfit may be accommodated by an array of dislocations having a periodicity which is related to the moiré-fringe spacing. The periodic strain field from the dislocations is, of course, only present if the two materials are in intimate contact. This image is actually recorded under rather special conditions, which we will discuss in more detail in Chapter 27.

23.6.C Dislocations and Moiré Fringes

Since the moiré pattern can often be thought of as a magnified view of the ‘structure’ of the materials, such

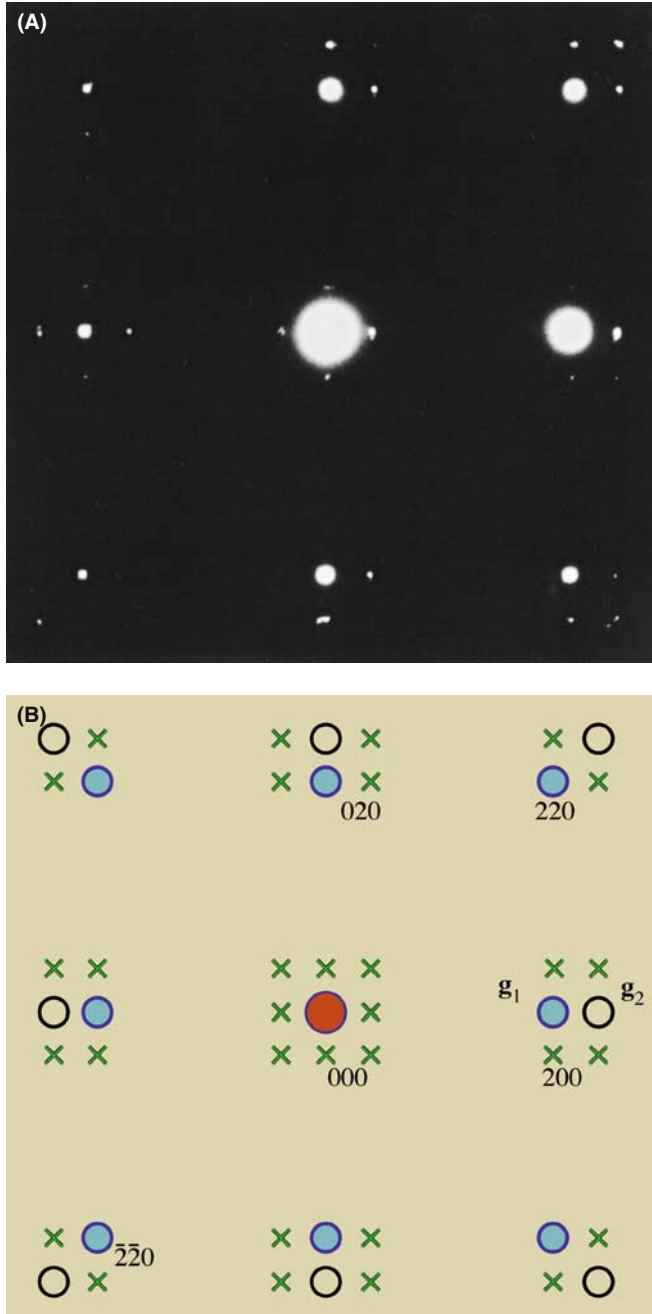


FIGURE 23.7. (A) Experimental diffraction pattern from perfectly aligned Ni and NiO. Brighter spots are from NiO, which has the larger lattice parameter. (B) Schematic which explains translational moiré fringes. Closed circles (\mathbf{g}_1) correspond to crystal 1, open circles (\mathbf{g}_2) to crystal 2 and \times double diffraction of \mathbf{g}_1 beams by crystal 2. Only \times reflections close to \mathbf{g}_1 and \mathbf{g}_2 have appreciable intensity.

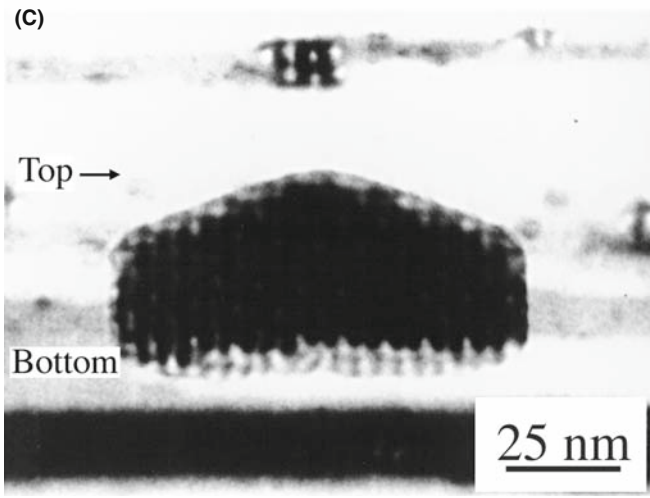
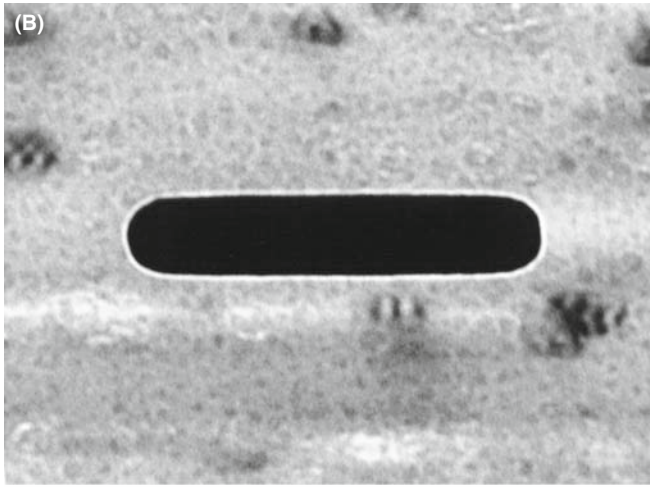
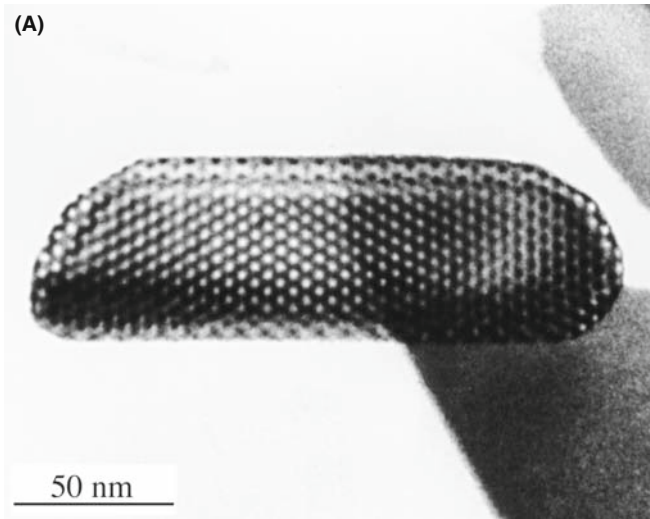


FIGURE 23.8. (A) The appearance of moiré fringes depends on the thickness of the specimen, as you can see where the edges of this island (Fe_2O_3) are inclined relative to the surface of the substrate (Al_2O_3). (B) The particle is too thick to show moiré fringes when edge on. (C) When this thick particle is tilted over, moiré fringes are seen at both the top and the bottom.

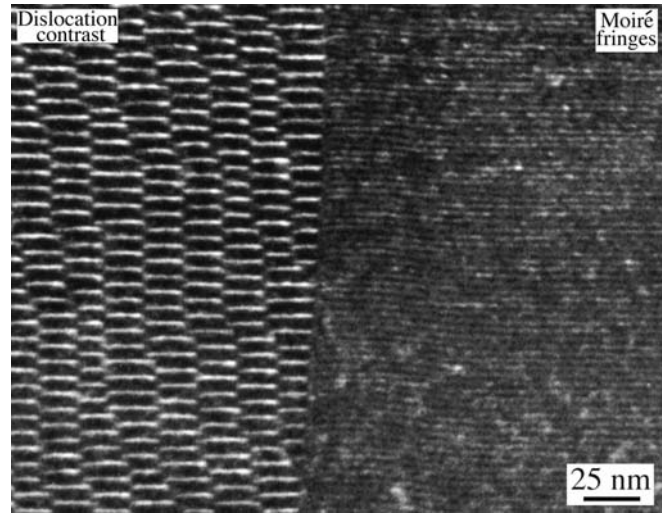


FIGURE 23.9. WBDF image of moiré fringes at a grain boundary showing very different contrast to the region containing dislocations.

patterns can be used to locate and give information on dislocations which are present in one material but not the other (but “which?” is the question). We can form an image which contains information on the dislocation if it is associated with a terminating lattice plane in one material, but we don’t actually ‘see’ the dislocation. This effect is illustrated in Figure 23.10; the image appears as a magnified view of the projection of the dislocation.

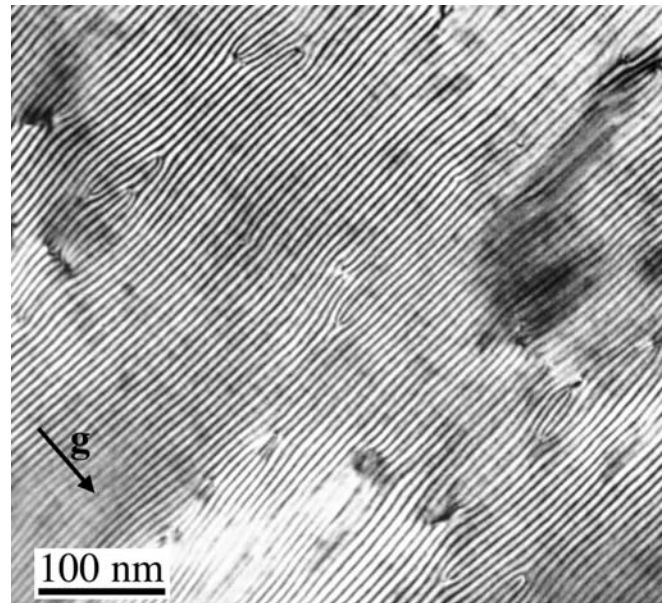


FIGURE 23.10. Moiré fringes reveal the presence of dislocations in a thin film of CoGa grown on a GaAs substrate. The (001) interface lies parallel to the specimen surface. Although the images contain much detail, most of it cannot readily be related to the structure of the defects.

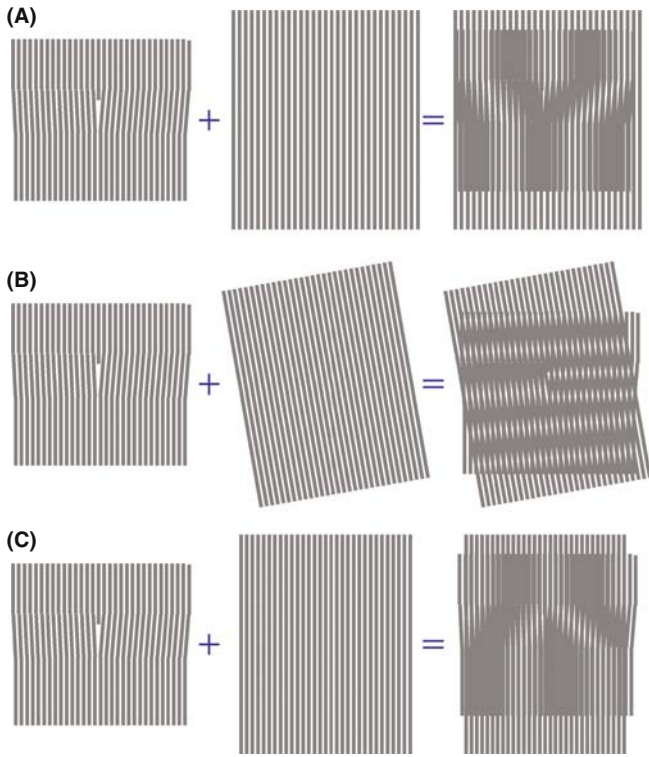


FIGURE 23.11. Schematic diagrams showing why moiré patterns from regions containing dislocations cannot be readily interpreted: (A) a dislocation image formed by interference between a regular lattice and one containing an extra half-plane. (B) In comparison with (A), a small rotation of the lattice of either grain can cause a large rotation of the dislocation fringes. (C) A small spacing change of either lattice can cause the dislocation image to reverse.

This result can be deceptive, as you can see in Figure 23.11, where we have rotated the perfect grain slightly and changed the dislocation fringe spacing.

The images can always be related directly to the projected Burgers vector of the dislocation, but you must know which planes give rise to the fringes. So make some models and experiment.

This analysis even works if you have two or more terminating fringes, but don't put too much emphasis on the actual location of the fringes. Remember, the dislocation may not be parallel to the beam. Moiré fringes may be related to a dislocation in the plane of the interface, since these locally relax the misfit. One example of such an application comes from the work of Vincent who showed that as Sn islands grew on a thin film of SnTe, the moiré-fringe spacing around the perimeter of the islands gradually increased. Suddenly, the strain at the interface was so large that a dislocation was nucleated to relax the strain and the process began again. The analysis of the changes in moiré-fringe spacing is shown in Figure 23.12.

Since the spacing of moiré fringes essentially gives a magnified view of the misfit between aligned particles

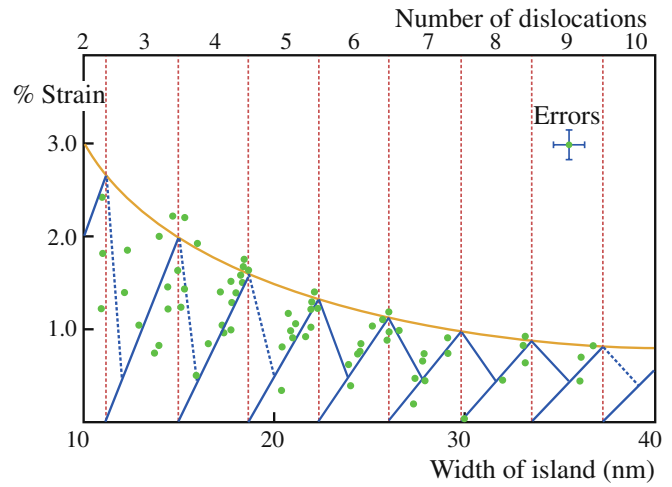


FIGURE 23.12. Moiré-fringe spacings can be used to monitor the change in lattice parameter as small islands of Sn grow in size on a thin film of SnTe. This plot shows how the strain (measured from the moiré-fringe spacing) can be related to the width of the misfitting island and then to the number of dislocations in the interface.

and a substrate, we can use them to measure the strain in such particles. In its simplest form, in one dimension, the strain is given by

$$\varepsilon = \frac{a_1 - a_0}{a_0} \quad (23.14)$$

where a_1 and a_0 are the lattice spacings of the particle and the substrate, respectively. You may need to modify this equation if the alignment is not simple cube-on-cube.

23.6.D Complex Moiré Fringes

Since moiré fringes can occur whenever $\Delta\mathbf{g}$ is small enough to be included in the objective aperture, we can have a situation where the relative rotation is rather large (45° or even 90°) so that \mathbf{g}_1 and \mathbf{g}_2 correspond to different sets of planes. This is illustrated in Figure 23.13 for YBCO grains rotated 45° on an MgO substrate. You can see that, as a bonus, the moiré fringes allow you to locate the 45° boundaries directly. Small rotations of the diffracting planes cause small rotations of \mathbf{g} but large rotations of $\Delta\mathbf{g}$.

Two overlapping lattices produce a pattern of interference fringes that is much coarser than the original pattern and is very sensitive to differences in lattice spacing and relative orientations. We can use this sensitivity to provide an optical method for examining small rotations or lattice-parameter differences in HRTEM images. Make transparencies of the 'distorted' image such as that shown in Figure 23.14A and a reference lattice; the reference lattice could be the perfect-crystal image or a template you have created on the computer. Now overlay the two and rotate/translate them relative

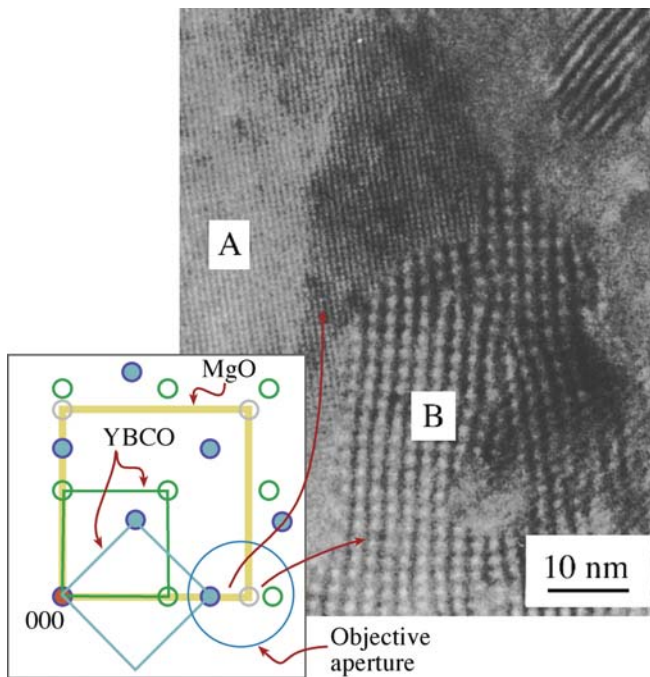


FIGURE 23.13. Moiré fringes formed when grains of YBCO grown on a single crystal of MgO are aligned to the substrate (B) or rotated through 45° (A); the spacing of the fringes is different so the position of the grain boundary can be identified. The circle in the DP shows the spots that cause the fringes. Small rotations of the fringe away from the perfect alignment are exaggerated because the spots are close together.

to one another. You will have created a new artificial moiré image similar to that shown in Figure 23.14B, which was formed for a special grain boundary in Al by Hetherington and Dahmen. This boundary is special because one set of {111} planes in the upper grain is nearly normal to one set in the lower grain. How near is near? Hetherington and Dahmen overlapped their experimental image with a template that was drawn to have two sets of lines normal to one another. Overlaying the two images gave moiré fringes, which were not quite perpendicular to one another. Careful measurements of the rotation and fringe spacing showed that the fringes in the experimental image were actually 89.3° apart, not 90°.

23.7 FRESNEL CONTRAST

We saw in Chapter 9 that we can use Fresnel-contrast images of holes in carbon films to correct the astigmatism of the objective lens. We'll now discuss how we can use this same contrast mechanism to learn more about particular features in the specimen. In the classic demonstration of Fresnel contrast using visible light, bright fringes can appear in the geometric shadow of an opaque mask or dark fringes can appear in the illuminated region. The complication introduced in the TEM version is that the 'mask' is not opaque but simply

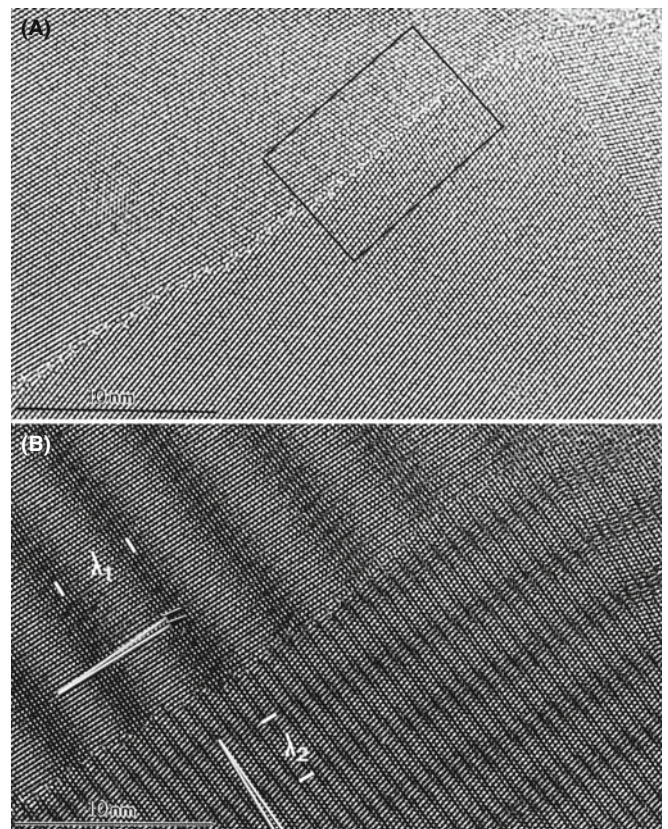


FIGURE 23.14. The use of 'artificial' moiré fringes to analyze a special grain boundary in Al. (A) An experimental image. (B) The same image overlaid with a perfect-crystal lattice transparency, producing moiré fringes of different spacings, λ_1 and λ_2 .

has a different inner potential. Therefore, in any situation where the inner potential changes abruptly, we can produce Fresnel fringes if we image that region out of focus. Since we still focus the objective lens on a plane that is close to the specimen, we are in the near-field or Fresnel regime. Since we often study lines, planes, or platelets by this technique, we'll often see Fresnel fringes.

23.7.A The Fresnel Biprism

We can demonstrate a particularly simple interference phenomenon by placing a wire at a position F on the optic axis, as shown in Figure 23.15A. Since the beam is narrow, the wire should be less than 1 μm in diameter and can be made of a drawn glass fiber coated with Cr or Au. If we apply ~ 10 V to the wire, it will bend the electron beam on either side in opposite directions. The resulting interference fringes can be recorded on photographic film, as shown in Figure 23.15B, or on the CCD. The wire here is acting as a beam splitter; we'll encounter it again when we discuss holography in Section 29.11. The visible-light analog is the prism. Notice how the wire acts to produce two virtual sources s_1 and s_2 , which are D_s apart. Horiuchi gives

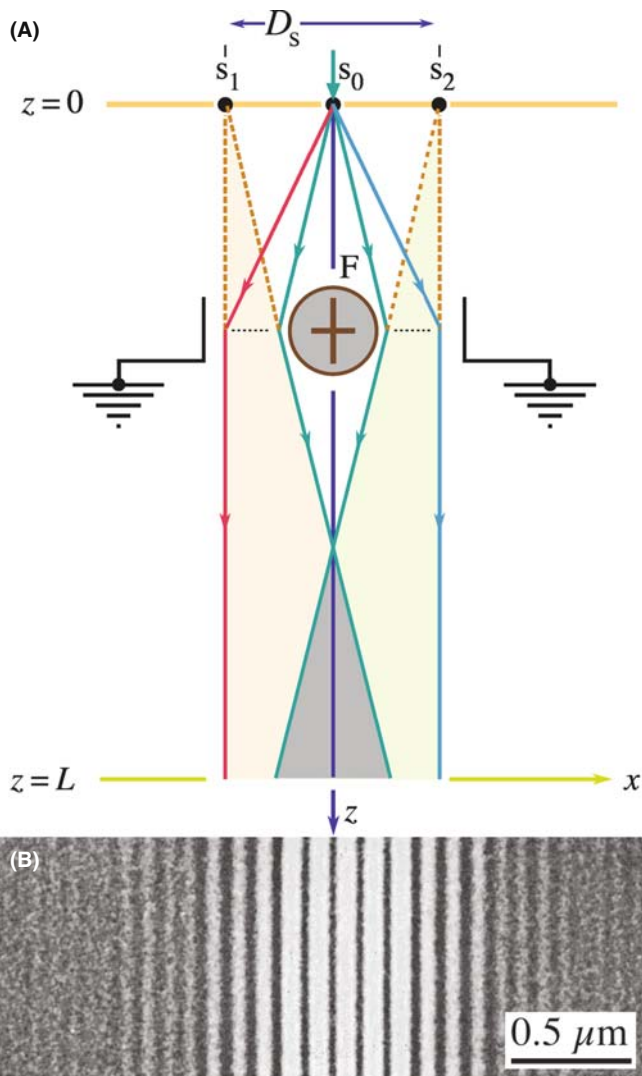


FIGURE 23.15. (A) A Fresnel biprism formed using a charged wire placed in the path of the beam; (B) the resulting interference fringes in the image.

the following equation to define a measure of the degree of spatial coherence, γ , which, as we discussed in Section 5.2 and Figure 5.13, is a function of the source size

$$\gamma = \frac{I_{\text{Max}} - I_{\text{Min}}}{I_{\text{Max}} + I_{\text{Min}}} \quad (23.15)$$

Here I_{Max} is the intensity of the central fringe and I_{Min} is the intensity of the first minimum in Figure 23.15B.

23.7.B Magnetic-Domain Walls

Although we'll discuss imaging magnetic materials in Chapter 29, it is appropriate here to consider briefly the similarity of Lorentz microscopy of magnetic-domain walls to other interference images. We know from our

discussion of magnetic fields in the electron lens in Chapter 6 that the Lorentz force acting on an electron with velocity \mathbf{v} is proportional to $\mathbf{v} \times \mathbf{B}$. If the sign of \mathbf{B} is opposite in two adjacent domains, then the electrons will be deflected in opposite directions as shown in Figure 23.16. The 'converging' domain wall is remarkably similar to the electron interferometer in the previous section. We can indeed produce a series of interference fringes. You should consult the original analysis of Boersch et al.; the basics are given in the book by Hirsch et al. who show that we can calculate the fringe spacing Δx using

$$\Delta x = \frac{\lambda(L + \ell)}{2\ell\beta_m} \quad (23.16)$$

where β_m is the angle of deflection of the beam, λ is the electron wavelength, ℓ is the 'source'-to-specimen distance and L is the specimen-to-'detector' distance. Quotation marks are used to emphasize that these are 'effective' distances like the 'camera length.' The value of Δx can be ~ 20 nm. You'll only see such interference fringes if you form the image using parallel illumination. We'll return to the imaging of magnetic specimens in Chapter 29.

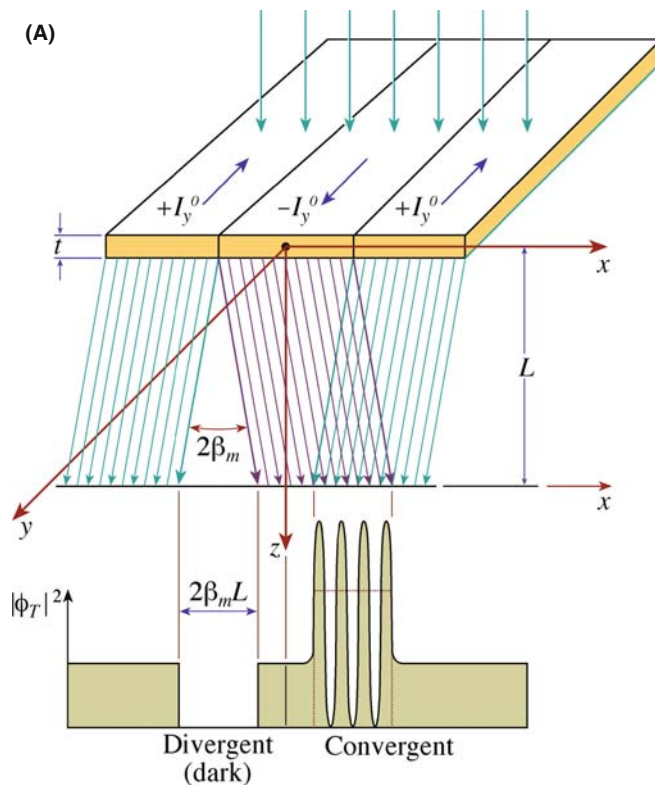


FIGURE 23.16. (A) Deflection of the electron beam by magnetic-domain walls; compare with Figure 23.15A. (B) Experimental interference fringes from one such wall; compare with Figure 23.15B.

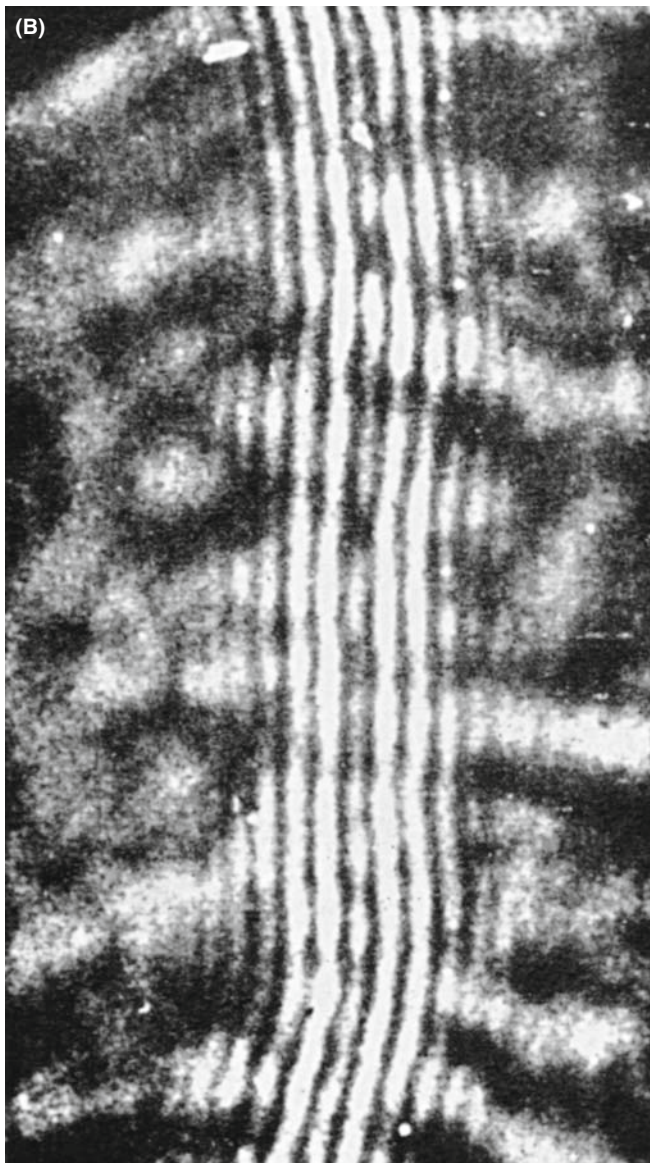


FIGURE 23.16. (Continued).

23.8 FRESNEL CONTRAST FROM VOIDS OR GAS BUBBLES

You might think that it would be difficult to image voids or small gas-filled cavities when there is no associated strain field, because voids or cavities do not scatter electrons. However, we can image holes that are fully enclosed inside the specimen by defocusing the image and observing the special phase contrast; this is the Fresnel contrast that we introduced back in Sections 2.9 and 9.5. In principle, we can apply this technique to holes that contain a liquid or even a solid (i.e., a second phase). In the latter case, however, the Fresnel contrast is likely to be hidden by strain contrast in the specimen.

You can image small voids or gas bubbles in two ways

- By orienting the region of interest so that $\mathbf{s} = 0$; the cavity then reduces the ‘thickness’ of material locally.
- By using Fresnel contrast.

CAUTION

Small particles can give similar contrast to small voids. The Fresnel contrast can easily be misinterpreted as a core-shell structure!

In the Fresnel technique, the image shows contrast whenever the objective lens is not focused on the bottom surface of the specimen.

Wilkins expressed the wave function as

$$\psi(t, \mathbf{r}') = \psi_0(t)[1 + \Delta_r(\mathbf{r}') + i\Delta_i(\mathbf{r}')] \quad (23.17)$$

Here $\psi_0(t)$ is the wave function in the absence of the cavity; Δ_r and Δ_i are real functions which depend on

- The location and dimensions of the cavity.
- The extinction distance and absorption parameter of the matrix (ξ_g and ξ_g').
- The potential difference, ΔV , between the inner potential of the matrix, V_0 , and that of the cavity, V_c (it could be filled or empty).

FRESNEL FOCUS?

Fresnel-contrast images are always out of focus.

In the case of thick foils where z_c , the size of the cavity in the direction of the beam, is $< 0.1\xi_g$, the wave function can be expressed as

$$\psi(t, \mathbf{r}') = \psi(t)[1 + i\Delta_i(\mathbf{r}')] \quad (23.18)$$

where Δ_i (using $w = s\xi_g$) is given by

$$\Delta_i = - \left(2\varepsilon_0 - \frac{1}{\varepsilon_g (1+w)^{1/2}} \right) z_c(\mathbf{r}') p_i(z) \quad (23.19)$$

The difference in inner potential is included in ε_0 , which is defined by the equation

$$\varepsilon_0 = - \frac{\Delta V}{E} k \quad (23.20)$$

Here, k is the magnitude of the wave vector and E is the energy of the electron beam. When the thickness dependence is damped out (the foil is thick), the intensity can be expressed quite simply as

$$|\psi(t, \mathbf{r}')|^2 = |\psi_0(t)|^2(1 + \Delta_i^2) \quad (23.21)$$

We can summarize some results from this analysis

- When the image is in focus, the cavity is invisible so we view it out of focus to observe the Fresnel-fringe contrast.
- The contrast depends on the difference in the inner potential of the matrix and the cavity; we usually see the most contrast if the content of the cavity is vacuum, because then ϵ_0 is greatest.
- The contrast does depend on the wavelength of the electrons through both k and E .
- Cavities as small as 1–2 nm in diameter can be imaged using Δf values of 0.5–1.0 μm .
- In the case where $w = 0$ and $2\epsilon_0 > \xi_{\text{g}}^{-1}$ (so Δ_i is < 0), if $\Delta f < 0$, the image is a bright dot surrounded by a dark fringe; if $\Delta f > 0$ the dot is dark and the fringe is bright.
- This is the same behavior as we saw in Figure 9.21, where we had a dark fringe at underfocus and a bright fringe at overfocus.

The contrast is illustrated in Figure 23.17. You should note that it is not the same as the black-white contrast from small precipitates that result from strain fields as we'll discuss in Chapter 26. You'll find a more detailed analysis in the article by Rühle and Wilkens, but you can already appreciate the relevance for imaging nanoparticles.

23.9 FRESNEL CONTRAST FROM LATTICE DEFECTS

This topic is one that is receiving more attention as computers and simulation programs become, respectively, more powerful and more user-friendly. The reason for this increased attention is clear, as Bursill et al. showed in their pioneering studying of Fresnel fringes from edge-on defects. They demonstrated that if you take great care in determining all the electron-optical

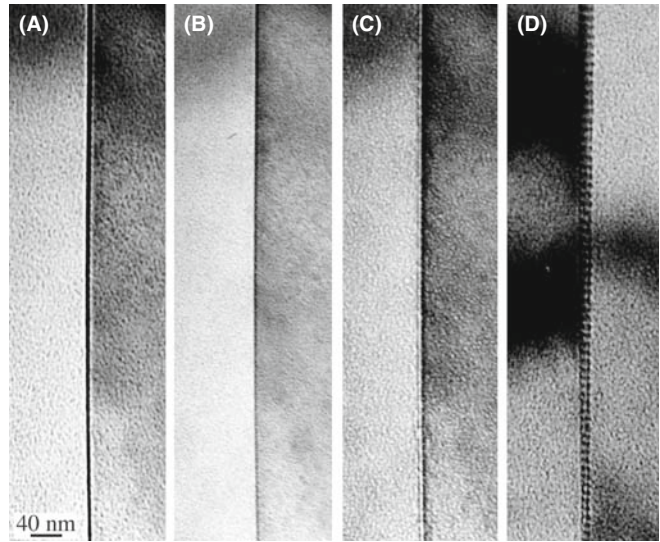


FIGURE 23.18. (A–D) A through-focus series of images from an edge-on GB showing the changes in Fresnel contrast. The image in (D) shows the boundary tilted over to reveal its periodic structure more clearly.

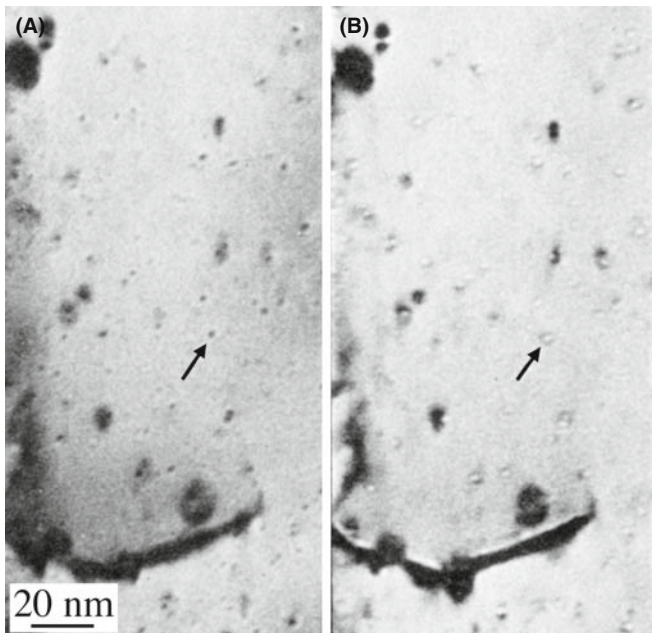


FIGURE 23.17. Fresnel contrast from He bubbles in Au. (A) Overfocus image. (B) Underfocus image.

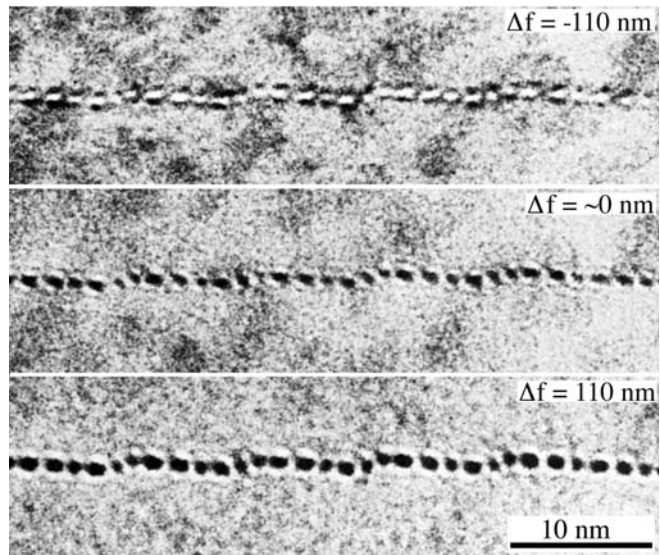


FIGURE 23.19. Series of experimental images recorded at different values of Δf for a low-angle grain boundary in NiO. Each white or black spot corresponds to one end-on dislocation.

parameters (particularly what defocus steps you are using), you can obtain new information on edge-on defects. The defect images are very sensitive to the model used to simulate them, but you will need other information for a full analysis. We'll refer you to their

paper on the {100} platelets in diamond and concentrate on two more widely applicable situations, namely, end-on dislocations and edge-on grain boundaries; in both cases, we now have many techniques, such as XEDS, EELS, and HRTEM, to complement the Fresnel-fringe studies.

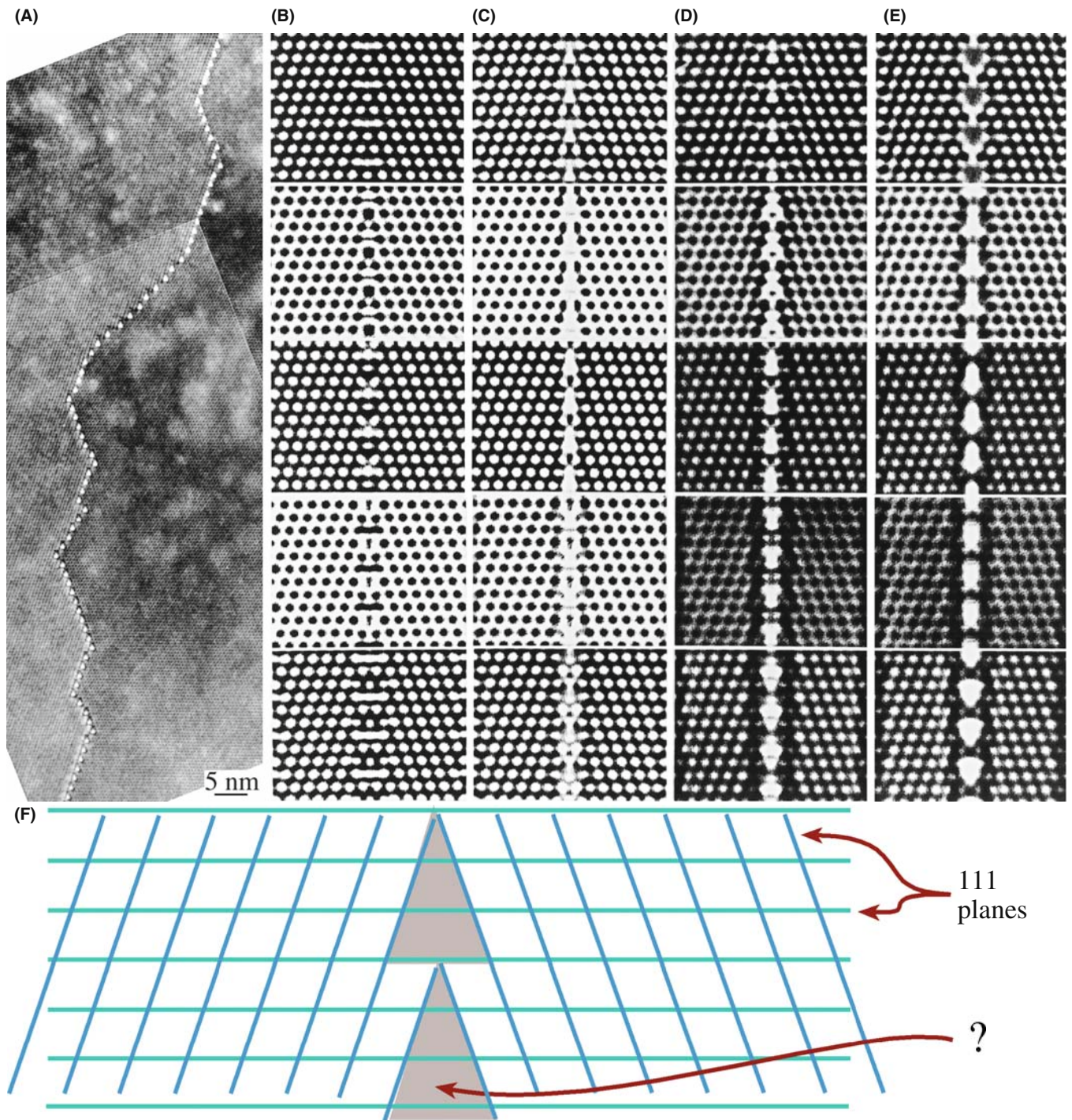


FIGURE 23.20. The structure of the (112) lateral twin boundary in a thin-foil of spinel consists of triangular prisms with a density lower than the bulk crystal as shown in the diagram at the bottom. (A) Fresnel contrast is seen when these prism-like defects are imaged out of focus. (B–E) Simulated images; each column is a different model: (B) no ions removed, (C) ion sites in the prisms half occupied, (D, E) all ions removed. The defocus in each row increases from the top (–10, –70, –130, –160, –210 nm), and the thickness is 5.7 nm. (F) Schematic of the prism ‘tubes.’

23.9.A Grain Boundaries

We might expect almost any grain boundary to show a localized change in the inner potential. However, following the original suggestion by Clarke (1979), Fresnel-contrast imaging has been used most extensively to study those interfaces which are thought to contain a thin layer of glass. Part of the reason for this emphasis is simply that other techniques tend to give ambiguous results for such interfaces.

When you use the Fresnel-fringe technique to study grain boundaries or analyze intergranular films, you must orient the boundary in the edge-on position so that you can probe the potential at the boundary. Later, in Section 30.11, we will consider the actual shape of this ‘potential well.’

In a real TEM specimen, the specimen thickness is likely to change at the grain boundary even if only by a nanometer or so. Since the specimen will be quite thin, this change can give an appreciable contribution to the difference in the ‘effective inner potential’ seen by the electron beam. You can defocus the image to see the Fresnel contrast shown in Figure 23.18.

The Fresnel-contrast technique can equally well be applied to phase boundaries, with perhaps the most thoroughly studied example being the Si/SiO₂ interface. Since the details of the contrast are sensitive to the abruptness of the change in the inner potential, the technique can also produce information on this aspect of the interface. Nevertheless, you must always look for associated changes in the real geometry that can occur even when you’re just forming Fresnel fringes from the edge of the specimen.

23.9.B End-On Dislocations

We’ve just seen that we can detect Fresnel-fringe contrast from edge-on high-angle grain boundaries. We might then ask: “is it possible to detect similar contrast from low-angle grain boundaries, i.e., grain boundaries which consist of arrays of distinct dislocations?” It is indeed possible, as shown in the series of images from a tilt boundary in NiO in Figure 23.19. This through-focus series, and images of other grain boundaries, was analyzed by assuming that

there is a change in $\Delta V(\mathbf{r})$ in the mean inner potential at the core of the dislocation. Two models were proposed for $\Delta V(\mathbf{r})$. In model 1, when $r < r_0$

$$\Delta V(\mathbf{r}) = \Delta V_0 \left\{ 1 - e^{-\frac{(r-r_0)}{a}} \right\} \quad (23.22)$$

but when $r > r_0$

$$\Delta V(\mathbf{r}) = 0 \quad (23.23)$$

The constant a is $\sim 0.1r_0$. In model 2

$$\Delta V(\mathbf{r}) = \Delta V_0 \exp(-r^2/r_0^2) \quad (23.24)$$

In both cases, ΔV_0 is negative. As an example of the quantities involved in these equations, if the Burgers vector of the dislocations is $1/2 [110]$, Rühle and Sass found that $\Delta V_0 = 0.09V_0$ for $r_0 = 3.2 \text{ \AA}$. They could not distinguish between the two models for $\Delta V(\mathbf{r})$ but two clear points come out of this study

- You must know the inclination of your foil surface. If the lower surface is inclined to the horizontal, then thicker parts of the specimen can be much closer to the objective lens than in the thin area; you can do a quick calculation to prove this point.
- The inner potential at a dislocation core is not the same as the bulk value. You should expect the value of ΔV_0 to be influenced by a change in stoichiometry or impurity segregation.

Before leaving this topic, we should point out that the inner potential at the grain boundary may not be uniform, perhaps because the width of the interface varies or the interface facets on a mesoscopic scale. You can still see Fresnel effects which relate to the periodicity in the grain boundary even if this periodicity is not associated with dislocations. A particularly clear example of such a variation is shown in Figure 23.20, where a twin boundary in spinel is essentially constructed of parallel triangular tubes; the inner potential inside the tube is much lower than the matrix value and the tubes are only about 1.2 nm high.

CHAPTER SUMMARY

Phase contrast will occur whenever we have more than one beam contributing to the image. The clue is: if you see fringes (periodic contrast) of any sort, then you are almost certainly observing a phase-contrast image. This conclusion even applies to thickness fringes (Chapter 24) and stacking-fault fringes (Chapter 25) in what are traditionally called two-beam diffraction-contrast images.

Phase-contrast images are widely used in three forms

- Images which relate directly to the structural periodicity of the crystalline specimen
- Moiré-fringes
- Fresnel-contrast

It is even possible for an image to show all three effects at the same time. So you must remember that phase-contrast effects don't just occur when you are forming high-resolution images. You will create Fresnel contrast whenever your specimen is thick or you are working out of focus. We should note that it is difficult, but in principle not impossible, to be quantitative in your analysis of Fresnel fringes.

The usefulness of moiré fringes continues to surprise even experienced users of the TEM. However, you still have to exercise caution when interpreting what they are telling you about defects in your material.

The appearance of the Fresnel image varies with small changes in the thickness, orientation, or scattering factor of your specimen, and variations in the focus or astigmatism of the objective lens.

REFERENCES

We encourage you to delve into the early papers on moiré fringes. Search the literature for those by Pashley, Stowell, et al. on Au islands on mica.

The books by Hirsch et al. (1977), Horiuchi (1994), and De Graef (2003) are useful references. That by Spence (2003) is essential follow-up reading (and described in Chapter 1).

SPECIAL TOPICS

Boersch, H, Hamisch, H, Wohlleben, D and Grohmann, K 1920 *Z. Phys.* **159** 397–404. The original analysis of interference fringes from domain walls.

Bursill, LA, Barry, JC, Hudson, PRW 1978 *Fresnel Diffraction at {100} Platelets in Diamond* *Phil. Mag.* **A37** 789–812.

Heavens, OS and Ditchburn, RW 1991 *Insight into Optics* p 73 John Wiley & Sons New York. Gives a clear treatment of Fresnel fringes.

Hetherington, CJD and Dahmen, U 1992 *Scanning Microscopy Supplement* **6** p 405 Scanning Microscopy International AMF O'Hare IL.

Krivanek, OL and Rez, P 1980 *Proc. 38th Ann. EMSA Meeting* p 170 Ed. GW Bailey Claitors Baton Rouge LA.

MOIRÉ FRINGES

Menter, JW 1956 *The Direct Study by Electron Microscopy of Crystal Lattices and Their Imperfections* *Proc. Roy. Soc. (London)* **A236** 119–135. Classic early work.

Norton, MG and Carter, CB 1995 *Moiré Patterns and Their Application to the Study of the Growth of $YBa_2Cu_3O_{7-\delta}$ Thin Films* *J. Mater. Sci.* **30** 381–389. Islands of YBCO.

Rühle, M and Wilkens, M 1975 *Defocusing Contrast of Cavities; Theory* *Cryst. Lattice Defects* **6** 129–400. Voids—the definitive work.

Vincent, R 1969 *Analysis of Residual Strains in Epitaxial Thin Films* *Phil. Mag.* **19** 1127–1139. Showing the magical size effect using moirés.

Wilkens, M 1975 in *Electron Microscopy in Materials Science II* p 647 Eds. U Valdré and E Ruedl CEC Brussels. A clear explanation of the use of Fresnel fringe theory to analyze voids.

FRESNEL FRINGES, SURFACES, AND LAYERS OF GLASS

Clarke, DR 1979 *On The Detection of Thin Intergranular Films by Electron Microscopy* *Ultramicrosc.* **4** 33–44.

Fukushima, K, Kawakatzu, H and Fukami, A 1974 *J. Phys.* **D7** 257. Just surfaces.

Longworth, S 2006 Ph.D. Thesis, Cambridge University. Concludes that the Fresnel fringe technique is not reliable for measuring amorphous films.

Ness, JN, Stobbs, WM and Page, TF 1986 *A TEM Fresnel Diffraction-Based Method for Characterizing Interfacial Films* *Phil. Mag.* **54** 679–702.

Rasmussen, DR and Carter, CB 1990 *On the Fresnel-Fringe Technique for the Analysis of Interfacial Films* *Ultramicrosc.* **32** 337–348.

Rasmussen, DR, Simpson, YK, Kilaas, R and Carter, CB 1989 *Contrast Effects at Grooved Interfaces* *Ultramicrosc.* **30** 52–55.

Simpson, YK, Carter, CB, Morrissey, KJ, Angelini, P and Bentley, J 1986 *Identification of Thin Amorphous Films at Grain-Boundaries in Al_2O_3* *J. Mater. Sci.* **21** 2689–2696. Discussion of different techniques for characterizing these layers.

FRESNEL FRINGES INTERFACES

- Carter, CB, Elgat, Z and Shaw, TM 1987 *Lateral Twin Boundaries in Spinel* Phil. Mag. **55** 21–38. Tubes along GBs.
- Ross, FM and Stobbs, WM 1991a *Study of the Initial Stages of the Oxidation of Silicon Using the Fresnel Method* Phil. Mag. **A63** 1–36.
- Ross, FM and Stobbs, WM 1991b *Computer Modelling for Fresnel Contrast Analysis* Phil. Mag. **A63** 37–70. The inner potential.
- Rühle, M and Sass, SL 1984 *Detection of the Change in Mean Inner Potential at Dislocations in Grain Boundaries in NiO* Phil. Mag. **A49** 759–782. Individual dislocations with different Burgers vectors show different Fresnel contrast.
- Tafto, J, Jones, RH and Heald, SM 1986 *Transmission Electron Microscopy of Interfaces Utilizing Mean Inner Potential Differences Between Materials* J. Appl. Phys. **60** 4316–4318.

THE COMPANION TEXT

The companion text includes several chapters that build on phase contrast as it applies to HRTEM and holography.

SELF-ASSESSMENT QUESTIONS

- Q23.1 What are the differences between electron diffraction and X-ray diffraction?
- Q23.2 Deduce a simple expression for the intensity when both beams are admitted through the objective aperture in the two-beam geometry.
- Q23.3 What is the key advantage of using the three-beam condition where \mathbf{g} and \mathbf{O} are equally strongly excited and \mathbf{O} stays on the optic axis?
- Q23.4 Might you see a different fringe periodicity if you use the three-beam rather than two-beam condition?
- Q23.5 If you want to resolve the dumbbells in Si, what reflections must you include when forming the image?
- Q23.6 Deduce the spacing for translational moiré fringes by using the DP construction.
- Q23.7 Deduce the spacing for rotational moiré fringes by using the DP construction.
- Q23.8 What do moiré fringes tell you about the structure between the two materials which overlap to cause the fringes?
- Q23.9 The image of a dislocation seen in a moiré pattern may suggest a Burgers vector which is 90° off or even of the wrong sign. What is the most important factor to know when you interpret moiré patterns of dislocations?
- Q23.10 Islands of Sn growing on a thin film of SnTe show a magic number effect. What is this effect?
- Q23.11 Looking at a particular specimen in the TEM we see fringes when we are out of contrast and say this is a Fresnel image. Why?
- Q23.12 Why is the Fresnel biprism so named?
- Q23.13 What is Lorentz microscopy?
- Q23.14 You could image a small gas bubble if the specimen were in focus or you could intentionally make the image out of focus. Why do both techniques work and which approach would you recommend?
- Q23.15 What information are we obtaining when we image grain boundaries using Fresnel contrast?
- Q23.16 Could we use Fresnel contrast to examine twist boundaries? (Be careful.)
- Q23.17 Would you prefer underfocus or overfocus when imaging a grain boundary using the Fresnel method?
- Q23.18 Approximately how large is the defocus value used when imaging end-on dislocations?
- Q23.19 You could use the Cornu spiral in visible-light optics to calculate the location and intensity of fringes from a semi-infinite opaque sheet. Why is the TEM specimen more complex?

TEXT-SPECIFIC QUESTIONS

- T23.1 Consider Figure 23.1B. Why does this not give a true image of the lattice planes? Hence why is 'C' better?
- T23.2 Consider Figure 23.3. Why do the dumbbells in the image not correspond to those in the structure and why are they there? Give a full clear explanation.
- T23.3 Consider Figure 23.4B. Why would Z-contrast imaging be useful here? Explain in detail.
- T23.4 Consider Figure 23.5A and B. (A) If one set of these 111 planes (smaller d) corresponds to Si and the other set is 111 planes from a different fcc material. What is the lattice parameter of the second material. (B) If the fringes are the 111 fringes of Si, show that the rotation is consistent with the moiré fringe spacing.
- T23.5 Consider Figure 23.16. Use reasonable values of the experimental parameters to suggest a spacing for the fringes in B. Justify your choice of values.
- T23.6 Consider Figure 23.16. Estimate the value of Z. Justify your choice of other values.
- T23.7 Consider Figure 23.12. Deduce a value for the dislocation periodicity and compare this value with the expected misfit for this system.

- T23.8 Consider Figure 23.13. Index both DPs and then compare the observed fringe spacing with that which you would expect from the crystal parameters.
- T23.9 Consider Figure 23.11. What is the misfit in A and in C? Does the observed angle of rotation in B really produce the fringes (and their spacing) in B?
- T23.10 Assuming the hematite is fully relaxed, what would be the error in the magnification bar in Figure 23.8A.
- T23.11 Show quantitatively with the aid of a diagram, the image you would obtain using (a) the 220 and (b) the 200 regions of the DP in Figure 23.7.
- T23.12 If the InAsSb film in Figure 23.4B is fully relaxed, what is the As:Sb ratio?
- T23.13 Fully describe the geometry of the grains shown in Figure 23.4C.
- T23.14 Figure 23.4A actually shows two spinel/olivine interfaces. Which material is where and what are the orientations?
- T23.15 In Figure 23.4B, what is the character and Burgers vectors of each dislocation?
- T23.16 Confirm that the moiré fringe spacings in Figure 23.5A and B are as predicted by the formula.
- T23.17 Confirm that the dumbbells in Figure 23.3A are not identical to those in Figure 23.3B and explain how they arise.
- T23.18 Work through the equations leading to equation 23.9. Show quantitatively how you might see different 'lattice fringes' from the same set of lattice planes.
- T23.19 Consider Figure 23.9. Deduce the \mathbf{g} -vector used to form the image, the rotation angle, and explain why the spacing of the short bright lines is not the same as that of the moiré fringes.
- T23.20 Through what angle was the specimen in Figure 23.8 tilted between image B and C?
- T23.21 In Figure 23.10, what is \mathbf{g} and what is the misfit between the two materials?
- T23.22 Consider the terminating fringes near the middle of Figure 23.10. Assume the threading dislocation lies in the thin film and hence deduce its Burgers vector. If it were in the substrate, would the image be the same?
- T23.23 Consider Figure 23.17. Estimate the size of the arrowed void and the difference in focus between the two images.
- T23.24 If the specimen shown in Figure 23.17 were imaged in focus, what would you see in the image? Justify your answer.
- T23.25 Discuss why the contrast in Figure 23.19 doesn't disappear when $\Delta f = 0$.
- T23.26 Consider the simulated images in Figure 23.20. Discuss the change in each set of four images going first across, and then down, the set. Do the changes agree with your expectations?

Thickness and Bending Effects

CHAPTER PREVIEW

We see diffraction contrast in an image for two reasons: either the thickness of the specimen varies or the diffraction conditions change across the specimen: the t effect and the s effect!

The *thickness* effect: when the thickness of the specimen is not uniform, the coupling (interference) of the direct and diffracted beams occurs over different distances, thus producing a thickness effect. Don't confuse diffraction contrast due to thickness changes with mass-thickness contrast discussed in the previous chapter. The effects are very different. The diffraction contrast changes with small changes in tilt, but the mass-thickness contrast doesn't.

The *bending* effect: Whenever the orientation of the diffracting planes changes, i.e., when the diffracting planes tilt relative to the beam, the contrast changes. To interpret changes in image contrast we need to understand how the contrast is related to thickness and bending.

We call these two important contrast phenomena 'thickness fringes' and 'bend contours.'

The present chapter is particularly important for three reasons

- All TEM specimens are thin but their thickness invariably changes.
- Because the specimens are so thin they also bend elastically, i.e., the lattice planes physically rotate.
- The planes also bend when lattice defects are introduced.

We can see the effects of these rotations even when they are $<0.1^\circ$ since they still have a significant effect on the image. Therefore, the bending may arise because the specimen is thin (i.e., giving possible artifacts of the technique) or it may be caused by strains which were present in the bulk material. The result is that, in real specimens, bending and thickness effects often occur together.

24.1 THE FUNDAMENTAL IDEAS

To understand the origin of thickness fringes and bend contours, we limit the discussion to the two-beam situation and recall equations 13.46 and 13.47, which we derived from the Howie-Whelan equations. The intensity of the Bragg-diffracted beam is then given by equations 13.48 and 13.49 and summarized as

$$I_g = |\phi_g|^2 = \left(\frac{\pi t}{\xi_g}\right)^2 \cdot \frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} = 1 - I_0 \quad (24.1)$$

where s_{eff} is the effective excitation error in equation 13.47

$$s_{\text{eff}} = \sqrt{s^2 + \frac{1}{\xi_g^2}} \quad (24.2)$$

Although we will concentrate on I_g (the DF image intensity) for most of this discussion, the direct beam (BF image) behaves in a complementary manner (neglecting, for now, the effect of absorption and other diffracted beams). The diffracted intensity is periodic in the two independent quantities, t and s_{eff} . If we imagine the situation where t remains constant but s (and hence s_{eff}) varies locally, then we produce bend contours. Similarly, if s remains constant while t varies, then thickness fringes will result.

This chapter is simply concerned with the physical understanding of equation 24.1 and how you can relate the image to the information contained in the DP. Although these effects are often a hindrance to systematic analysis of lattice defects, they can, in certain situations, be useful. The most important reason for understanding them is that they are essentially unavoidable!

24.2 THICKNESS FRINGES

As a result of the way that we thin TEM specimens, very few of them (only evaporated thin films or ideal ultramicrotomed, cleaved, or FIBbed sections) have a uniform thickness over their entire area. A BF/DF pair of images from the same region of the specimen is shown in Figure 24.1; the thin area is generally in the form of a wedge.

Consider again equation 24.1. You should remember that, in this calculation, t is not the ‘thickness’ of the foil; it is actually the distance ‘traveled’ by the diffracted beam, but we’ll usually just say ‘thickness.’ If we try to treat the many-beam situation rigorously, then the value of t would, in general, be different for each beam. If you are actually viewing the foil flat-on (i.e., one surface normal to the beam), then t will be close to the geometric thickness of the foil. However, it is more difficult to analyze the image thoroughly when the foil is wedge-shaped and inclined to the beam. We almost invariably make the approximation that t is fixed with the justification being that the Bragg angles are small.

TWO BEAMS

The situation is complicated by absorption and the presence of other diffracted beams: we are never truly in a two-beam situation.

Equation 24.1 tells us that intensities of both the 0 and the g beams oscillate as t varies. Furthermore, these oscillations are complementary for the DF and BF

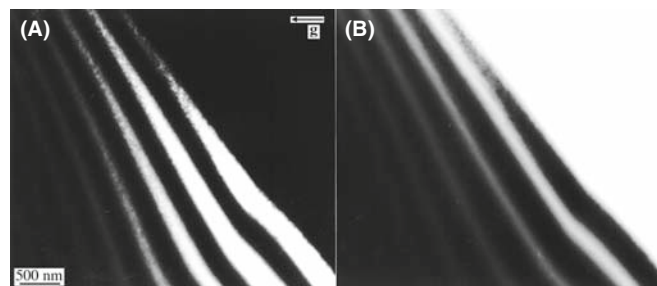


FIGURE 24.1. (A) BF and (B) DF images from the same region of a wedge-shaped specimen of Si at 300 kV tilted so that $g(220)$ is strong. The periodicity and contrast of the fringes are similar and complementary in each image.

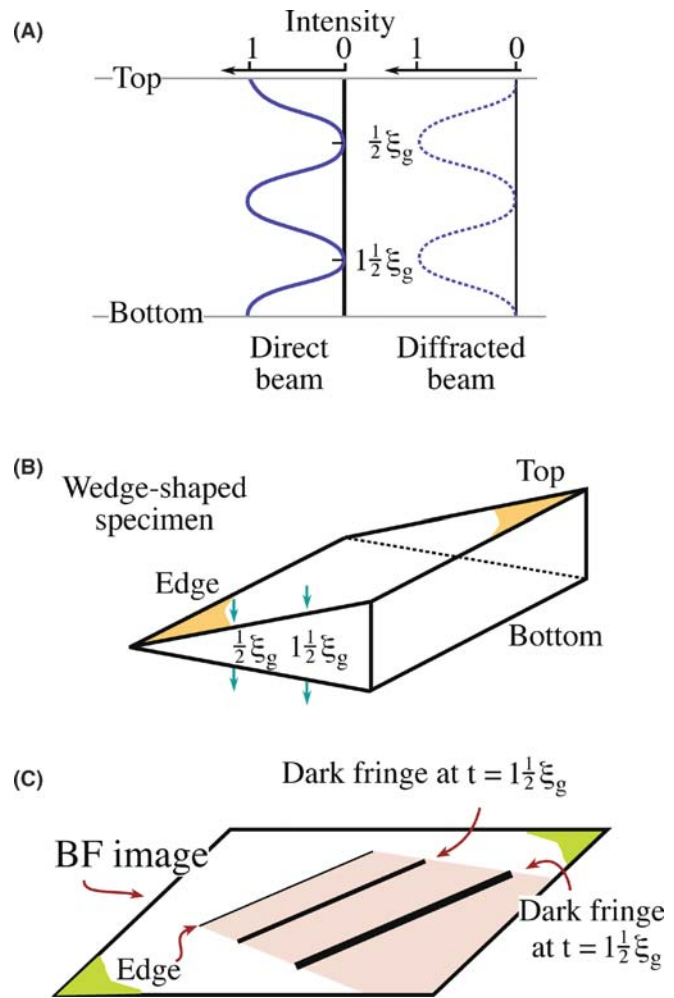


FIGURE 24.2. (A) At the Bragg condition ($s = 0$), the intensities of the direct and diffracted beams oscillate in a complementary way. (B) For a wedge specimen, the separation of the fringes in the image (C) is then determined by the angle of the wedge and the extinction distance, ξ_g .

images, as we show schematically in Figure 24.2. You can, of course, confirm this observation at the microscope by forming the image without using an objective aperture; there is then minimal contrast when you’re in focus. The intensity, I_0 , of the incident beam starts equal to unity and gradually decays, while the intensity of the diffracted beam, I_g , gradually increases until it becomes unity; I_0 is then zero; the process then repeats itself.

THICKNESS, ξ_g , AND ABSORPTION

As a rule of thumb, when other diffracted beams are present the effective extinction distance is reduced. At greater thicknesses, absorption occurs and the contrast is reduced.

These oscillations in I_0 or I_g are known as thickness fringes, though they are often not fringes. We sometimes call them thickness contours, because they denote the

contours where the specimen has constant thickness; you will only see these fringes when the thickness of the specimen varies locally, otherwise the contrast will be a uniform gray. As we'll see, the actual contrast can quickly change if the specimen is tilted through a small angle.

BLACK AND WHITE

It is important to realize that the image may appear to be black or white depending on the thickness of the specimen.

For example, in BF images, thicker areas are often brighter than thinner areas, which really is counter-intuitive.

Several examples of how thickness fringes might appear in your image are shown in Figure 24.3. Although it is often helpful to think of these fringes as thickness contours analogous to height or depth contours on a map, with the hole at sea level, remember that there are *two* surfaces to the TEM specimen. A DF image will usually appear to give greater contrast. This is partly because the hole now appears dark but also because many-beam effects are less important in DF. In Figure 24.3A, the narrow fringe pattern in this DF image is due to the grain boundary region being thinner than the matrix. In the DF image in Figure 24.3B, the reflection used to form this image is only excited in the right grain so the left grain is black; the diffracting grain exhibits strong thickness fringes in the regions where there are microtwins. This image introduces the idea that images of defects can also show thickness effects.

In Figure 24.3C, the specimen is an almost flat, parallel-sided film of MgO with holes formed by preferential chemical etching at defects; after thinning, the surfaces were faceted by heating the specimen at 1400°C. The holes in the image are white, so it is a BF image. The contours, like the holes, are angular because of the faceting, but they are not uniformly spaced because it's not a uniform wedge. Notice that the center of the hole is faceted and that the first fringe is a very narrow dark line. We know that this surface is different because it is curved. We also know from Figure 24.2 that, in a BF image, the first fringe must be bright if the thickness actually decreases to zero. We can therefore conclude from this one image that the specimen is not *tapering* to zero thickness at the center of this hole.

Although we've talked about wedges or specimens with gently curving surfaces so far, the way we actually calculate and analyze the contrast from such wedges is shown in Figure 24.4. We imagine that the specimen has two parallel surfaces which are normal to the electron beam, so that we have a fixed thickness for each calculation. We also assume that the beam is normal to the

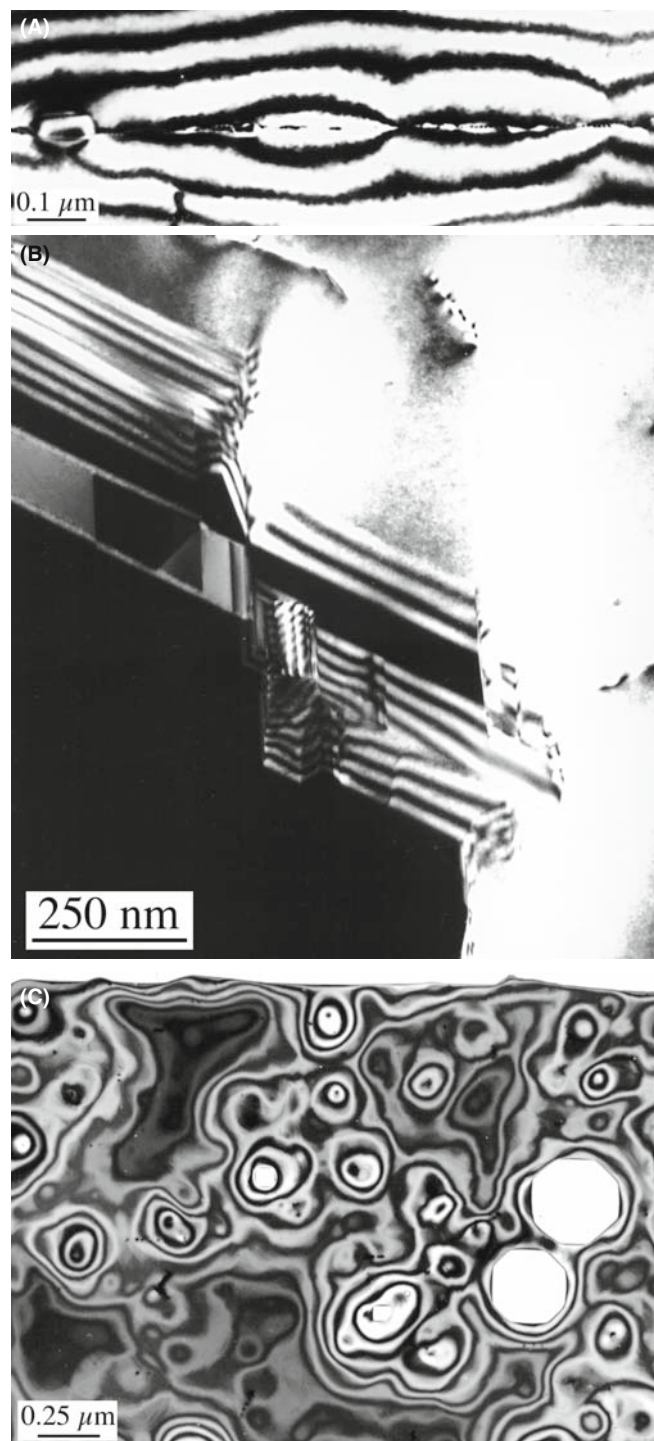


FIGURE 24.3. Examples of thickness fringes in (A) DF image of a preferentially thinned grain boundary; (B) a strong 220 DF image of microtwinned GaAs taken with only the right-hand grain diffracting, and (C) BF image of a chemically etched thin film of MgO. The white regions in (C) are holes in the specimen.

surface. We then change t and recalculate the intensity. Finally, we plot the different values for the intensity against t (as in Figure 24.2A) but we never actually inclined the surfaces!

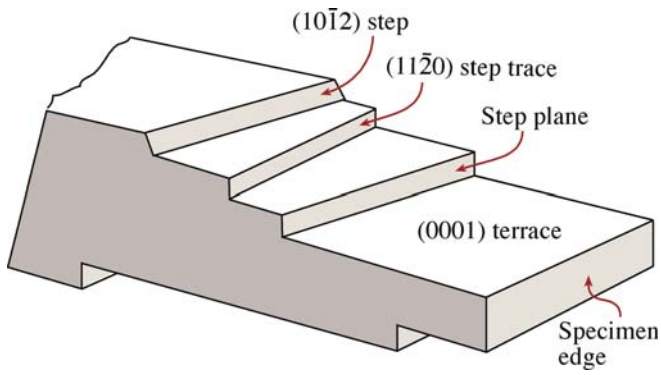


FIGURE 24.4. Schematic cross-sectional view of a specimen with terraces parallel to the surface and steps connecting terraces.

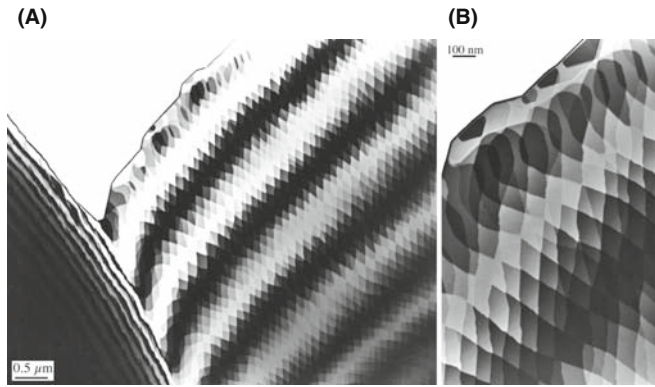


FIGURE 24.5. Thick fringes from an annealed Al_2O_3 specimen with the geometry shown in Figure 24.4. (A) At low magnification, the fringes are well defined and continuous, even when the wedge angle and wedge axis change. (B) At higher magnification, the contrast is seen to be quantized within a given fringe.

Figure 24.5 shows an image of a wedge-shaped specimen of Al_2O_3 which has been heat treated so that the surface has faceted parallel to certain low-index planes. The thickness fringes can then be seen to be discrete regions of different shades of gray; the fringes are, in general, quantized. You can form similar specimens by cleaving layer materials (e.g., graphite or mica) but the specimens tend to bend, which obscures these abrupt contrast changes.

24.3 THICKNESS FRINGES AND THE DP

A general rule in TEM is that, whenever we see a periodicity in real space (i.e., in the image), there must be a corresponding array of spots in reciprocal space; the converse is also true. If we image a specimen with a constant wedge angle, then we will see a uniform spacing of thickness fringes in both the BF and DF two-beam images even when $s_g = 0$. We must therefore have more than one spot ‘at G ’ when $s_g = 0$ otherwise we would not see fringes. We already know that if we increase s or if the wedge angle were larger, then the fringe separation

would decrease and the spacing of these spots must therefore increase.

To understand why there is more than one spot at G , go back to Chapter 17 where we showed that, because the specimen is thin, any spot in the DP will be elongated normal to the surface. When the specimen is wedge shaped, there will be two surfaces and we can imagine the spot being elongated normal to both surfaces, as was shown in Figure 17.4. Actually we have two curved relrods which do not intersect at $s = 0$; we related this curvature to the dispersion surface in Chapter 15. The diffraction geometry close to G is shown in Figure 24.6.

(A)

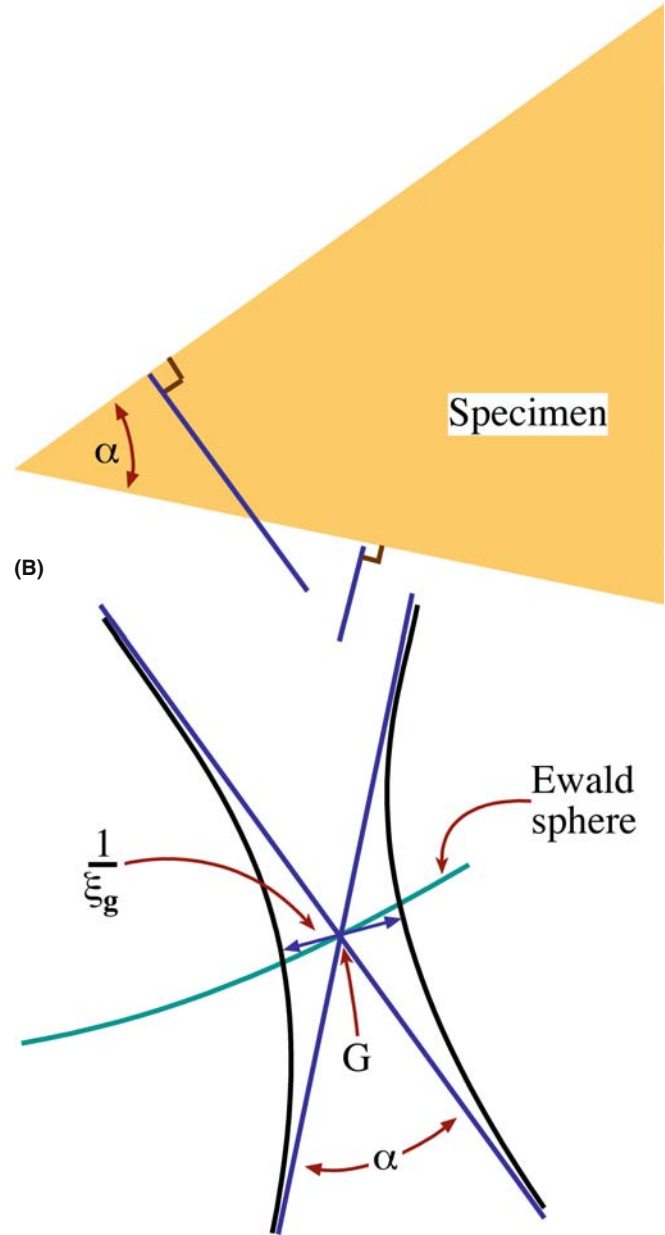


FIGURE 24.6. Relrods in (B) aligned normal to both surfaces of a wedge-shaped specimen in (A). In practice, the relrods (blue lines) don’t actually cross (they are the black curves) so there are always two spots in the DP.

This spot spacing is thus related to ξ_g^{-1} and the wedge angle. While the spacing of the thickness fringes depends on ξ_g , it is not equal to ξ_g . As we referenced back in Chapter 17, Amelinckx's group has shown that we can describe this geometric relationship as shown in Figure 17.4. As we tilt the crystal away from $s = 0$, the Ewald sphere will move up or down (as s becomes negative or positive) to cut the two 'rods.' So, you can see that there will be two spots instead of one at G and their separation will increase as s increases. As the separation increases, the spacing of the fringes decreases; the thickness fringes move closer together because the ξ_{eff} has decreased. The change in the fringe spacing is similar either side of $s = 0$. So be wary of trying to make accurate thickness measurements for wedge-shaped crystals.

FRINGE PERIODICITY

The minimum spot spacing in the DP corresponds to the periodicity of the thickness fringes, which at $s = 0$ is given directly by the extinction distance and the wedge angle.

We refer to thickness fringes as being an example of amplitude contrast because, in the two-beam case, they are associated with a particular reflection, \mathbf{g} . They actually occur due to interference between two beams, both of which are located close to \mathbf{g} , so they are really an example of phase contrast although we rarely think of them as such.

24.4 BEND CONTOURS (ANNOYING ARTIFACT, USEFUL TOOL, INVALUABLE INSIGHT)

This is a particularly satisfying topic, because you can understand it by considering a simple physical picture and yet the concept involved is the basis for understanding most aspects of defect contrast. Bend contours (don't call them extinction contours) occur when a particular set of diffracting planes is not parallel everywhere; the planes rock into, and through, the Bragg condition.

The specimen shown schematically in Figure 24.7 is aligned so that the hkl planes are exactly parallel to the incident beam at the center of the figure and always lie normal to the specimen surface even when it bends. We imagine that the foil bends evenly, so that the hkl planes are exactly in the Bragg condition at A and the $\bar{h}\bar{k}\bar{l}$ planes are exactly in the Bragg condition at B. We can draw the systematic row of reflections as you see below the bent crystal. Notice that $-G$ is now on the left and G on the right (with no lens present). Now if we form a BF

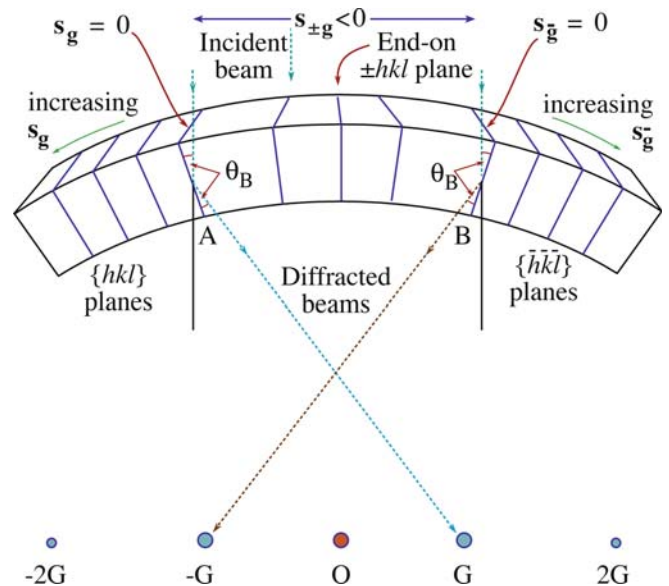


FIGURE 24.7. The origin of bend contours shown for a foil symmetrically bent either side of the Bragg conditions. For this geometry, when the hkl planes are in the Bragg condition, the reflection G is excited. Notice that G and the diffracting region are on opposite sides of O ; if the foil were bent upwards, they would be on the same side.

image we will see two dark lines. Next, we form the DF image using reflection \mathbf{g} . We see a bright band on the left because that's where \mathbf{g} is excited. Now use $\bar{\mathbf{g}}$ to form the image and the bright band is on the right. These bands are referred to as bend contours. A set of experimental images is also shown in Figures 24.8 and 24.9.

Remembering Bragg's law, the $(2h\ 2k\ 2l)$ planes diffract strongly when θ has increased to $\sim 2\theta_B$. So we'll see extra contours because of the higher-order diffraction. As θ increases, the planes rotate through the Bragg condition more quickly (within a small distance Δx) so the bend contours become much narrower for higher-order reflections.

DISPLACE THE SAD APERTURE

In actually doing this imaging experiment, you should translate the objective aperture to form the DF images; you'll lose some resolution but don't move (neither translate nor tilt) the specimen.

Tyro-microscopists occasionally have difficulty in distinguishing higher-order bend contours from real line defects in the crystal. The solution is very simple: tilt your specimen. Bend contours are not fixed to any particular position in the specimen and quickly move as you tilt.

Bend contours are true amplitude contrast, not phase contrast.

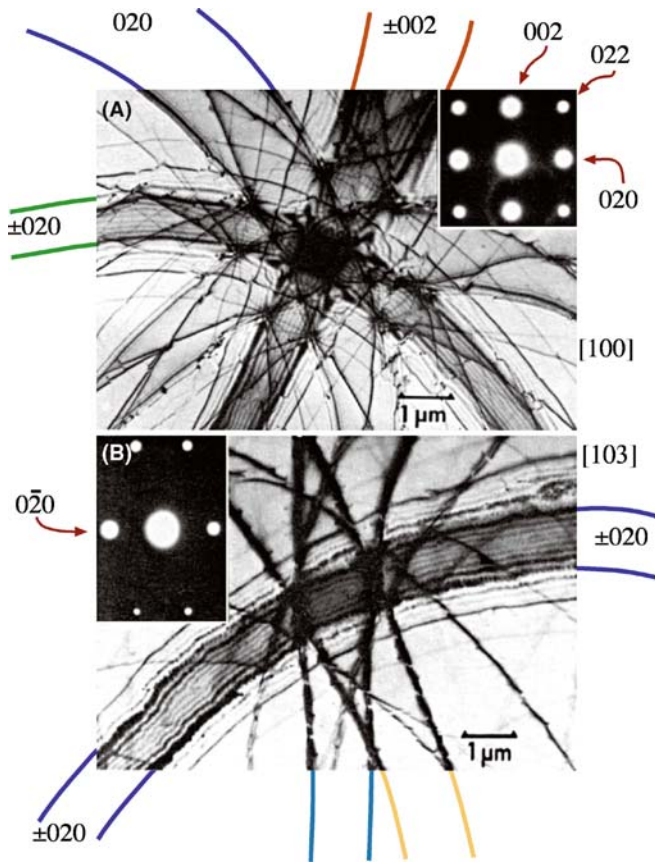


FIGURE 24.8. BF images of a bent Al specimen oriented close to the (A) [100] and (B) [013] zone axes. These images are known as (real-space) zone-axis patterns, or ZAPs, and are shown with their respective zone-axis DPs (insets). Each diffracting plane produces two bend contours, depending on whether θ_B or $-\theta_B$ is satisfied. Note that the separation of the bend contours is not uniform for any particular pair of planes because the curvature of the bending is not, in general, the same.

24.5 ZAPs AND REAL-SPACE CRYSTALLOGRAPHY

In the above discussion, we only considered bending about one axis. In real specimens, the bending will be more complex. This complexity will be important when the bent area is oriented close to a low-index pole, because the bend contours then form a zone-axis pattern or ZAP. Two examples of these ZAPs are shown in Figure 24.8. Although the ZAP is distorted, the symmetry of the zone axis is clear and such patterns have been used as a tool for real-space crystallographic analysis. Each contour is uniquely related to a particular set of diffracting planes, so the ZAP does not automatically introduce the twofold rotation axis that we are used to in SAD patterns. These contours are the real-space analog of the symmetry seen in large-angle CBED patterns.

In fact it's the exception that a $\pm g$ pair of bend contours is straight and parallel. In case you are having

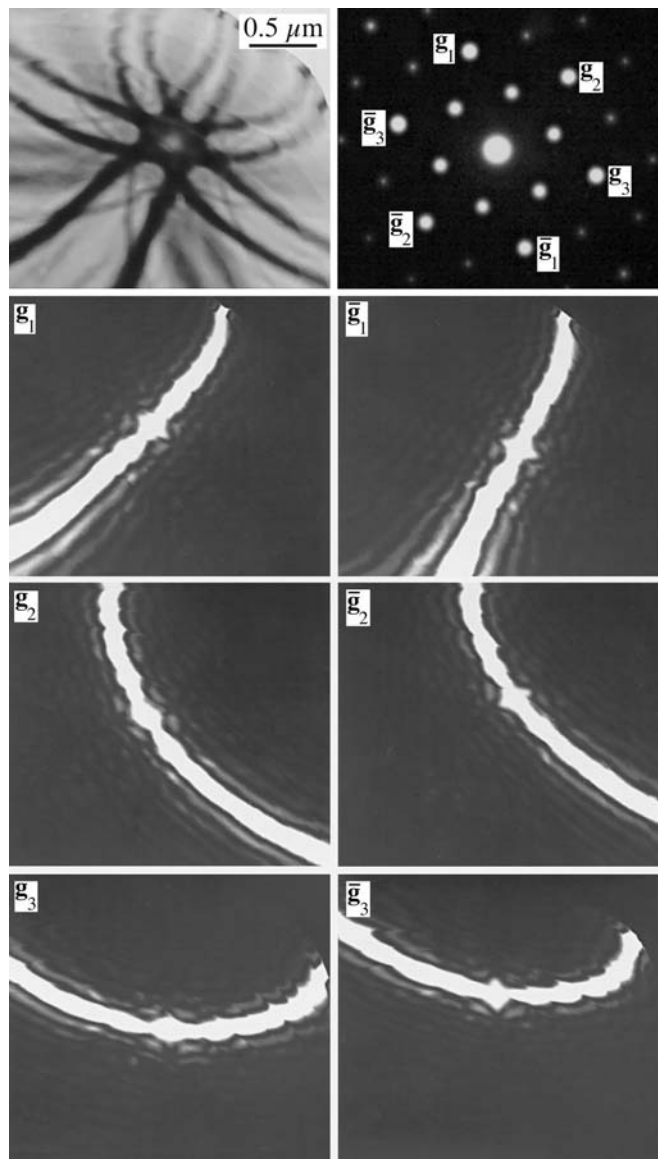


FIGURE 24.9. An 0001 real-space ZAP of Al_2O_3 : (A) BF image and (B) corresponding DP. (C–H) Displaced-aperture DF images taken from the spots indicated in (B) identifying the principal dark bend contours in (A); (C, D) $\pm(3030)$, (E, F) $\pm(3300)$, (G, H) $\pm(0330)$. Note in (A) that the inner $\{1120\}$ spots produce fainter bend contours than the $\{3300\}$.

a problem visualizing how a pair of contours might diverge, go back to the bent specimen in Figure 24.7, hold the $h\bar{k}l$ plane fixed at $x = x_0$, and then as you move along the foil (going into the page) gradually decrease the bend in the foil. The position where the hkl planes are in the Bragg condition gradually moves to the left, so $-x_0$ becomes more negative. Since x_0 is fixed the contours move apart in the image.

Notice how at the zone axis, the main 020 and 002 contours in the [100] ZAP are closely spaced, while in the [103] pattern, only one pair of contours is closely spaced; the others are more clearly defined and farther apart.

When the foil curvature is equal, this effect allows you to recognize a low-index ZAP. Since you can tilt the crystal, you can form different ZAPs from exactly the same area of your specimen just as you can for SAD and CBED (where we also use the term ZAP but then refer to a DP; really a ZADP). You index the contours in the manner described in Section 18.4 but use all the spots in the ZADP.

ZAPs IN REAL SPACE

In this case, a small \mathbf{g} in the DP gives a small spacing in the image, contrary to the usual inverse relationship between image and DP.

If your specimen is buckled, you can tilt it so that a particular bend contour stays at the position you're studying in the image. You're then doing the same operation as we did using Kikuchi lines in Chapter 19. Tilting in image mode is more tricky, but if the specimen is very buckled or too thin, you can't use Kikuchi lines. The ZAP and bend contours let you work in real space. You can even set the value of \mathbf{s} for a particular \mathbf{g} at a particular location on your specimen!

24.6 HILLOCKS, DENTS, OR SADDLES

The simplest use for bend contours is in determining whether an area is a hillock or a dent. This information is useful when analyzing particles grown on a substrate, particularly if the substrate is a thin film.

Figure 24.9A and B shows a ZAP in a thin specimen of Al_2O_3 and the associated SADP; the dark bands are $\{3300\}$ bend contours. Figure 24.9C–H shows DF images recorded using each of these reflections; the area is identical in each of these images. Using Figure 24.7 you can determine the sense of the bending. You can see directly that the bend contour from one set of hkl planes does not necessarily lie parallel to those planes but instead both curves and changes in width.

Remember, for these contrast experiments especially, it is important to move the objective aperture, not the specimen. The resolution of the image will be somewhat lower but this is not critical for this application.

24.7 ABSORPTION EFFECTS

When your specimen is very thick you won't see an image, so we can say that the electrons have then all been *absorbed*. The absorption process is more

important than this obvious statement might suggest. Much of our thinking about this topic is, however, just as empirical. In fact it is common to define an imaginary component ξ'_g to the extinction distance, so

$$\xi_g^{\text{abs}} = \xi_g \left(\frac{\xi'_g}{\xi'_g + i\xi_g} \right) \quad (24.3)$$

Then we use ξ_g^{abs} in the Howie-Whelan equations. The term ξ'_g is found to be approximately $10\xi_g$. The reason for choosing this expression for ξ_g^{abs} is that the $1/\xi_g$ in the Howie-Whelan equations can then be replaced by $(i/\xi'_g + 1/\xi_g)$. We do the same for ξ_0 . The result is that γ in the Howie-Whelan equations has an imaginary component. Consequently, we now have an exponential decay of the diffracted amplitude. It's a completely phenomenological treatment, but you will see reference to it. When we discuss EELS in Part 4, you'll appreciate the difficulties in modeling the effects of inelastic scattering on the image by a single parameter.

We did briefly discuss absorption of Bloch waves in Chapter 14. We showed that Bloch wave 2 (smaller \mathbf{k}) is less strongly absorbed than Bloch wave 1; Bloch wave 1 travels along the atom nuclei while Bloch wave 2 channels between them. As the crystal becomes thicker we lose Bloch wave 1. Since thickness fringes result from a beating between the two beams, we will lose the thickness fringes but will still be able to 'see through' the specimen as you can appreciate from Figure 24.10. Now you can see why equation 24.3 must be phenomenological—there are no beams in the crystal!

Bend contours in thicker parts of the specimen will also show the effect of this anomalous absorption. Looking back at Chapter 15 on the dispersion surface, you'll see that when \mathbf{s}_g is negative, the tie line D_1D_2 would be closer to $\mathbf{0}$ than \mathbf{g} and Bloch wave 2 contributes to ϕ_g more strongly. When \mathbf{s}_g becomes positive, Bloch wave 1 is the more strongly excited. We lose Bloch wave 1 because the specimen is thick. So as we rock through the Bragg condition on the bend contour, we'll lose the thickness fringes faster where Bloch wave 1 was weaker already, i.e., when \mathbf{s}_g is negative or inside the $\pm\mathbf{g}$ pair of bend contours.

ANOMALOUS ABSORPTION

Absorption due to the loss of Bloch wave 1 is called anomalous absorption for historical reasons, not because it is unexpected.

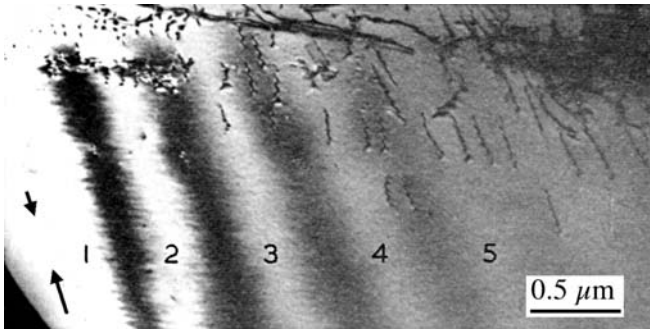


FIGURE 24.10. The contrast of thickness fringes in a two-beam BF image decreases when the effect of anomalous absorption is included. Note that the defects are still visible when the fringes have disappeared at a thickness of $-5 \xi_g$.

We can summarize this discussion of absorption with some conclusions

- We can define a parameter ξ'_g which is usually about $10 \xi_g$ and is really a fudge factor which modifies the Howie-Whelan equations to fit the experimental observations.
- The different Bloch waves are scattered differently. If they don't contribute to the image, we say that they were absorbed. We thus have anomalous absorption which is quite normal!
- Usable thicknesses are limited to about $5 \xi_g$, but you can optimize this if you channel the less-absorbed Bloch wave.

24.8 COMPUTER SIMULATION OF THICKNESS FRINGES

Thickness fringes can be simulated. We'll talk more about the programs in Chapter 25, but be wary: don't use a program as a 'black box.' Why do we need or want to simulate thickness fringes? As an illustration, let's look at a 90° -wedge specimen (see Section 10.6) so that we know how the thickness changes with position. The actual thickness will be *very* sensitive to the orientation of the specimen since the specimen is so thick, as shown in Figure 24.11A. The specimen is a GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ layered composite grown on (001). Since the cleavage surface is {110} it can be mounted at 45° so the beam is nearly parallel to the [100] pole. The value of ξ_g is different for the two materials, so they can be readily distinguished. Clearly, a quantitative simulation of this situation is non-trivial, especially if you have to consider the effect shown in Figure 24.6.

Because the fringe spacing changes as ξ_g changes, it will also change if you vary the accelerating voltage. You can see this effect clearly in Figure 24.11B and C which compares the same region of a wedge specimen imaged at 300 and 100 kV.

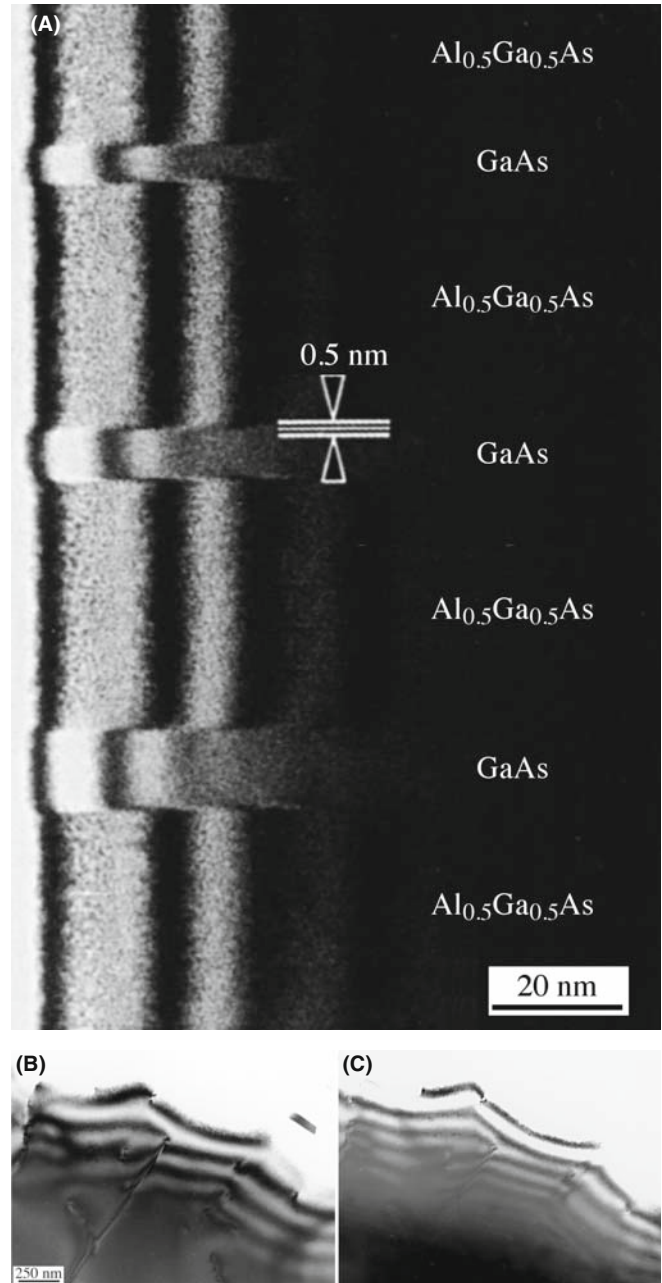


FIGURE 24.11. (A) Thickness fringes in a 90° wedge of alternating GaAs and AlGaAs. The extinction distance changes in each phase so the fringe spacing changes. Strong beam BF images ($s = 0$) for (B) 300-kV and (C) 100-kV electrons. The extinction distance increases as the accelerating voltage increases, and you can see through thicker areas; compare with the images in Figure 24.1.

24.9 THICKNESS-FRINGE/BEND-CONTOUR INTERACTIONS

It's clear from equation 24.1 that both bending and thickness effects can occur together. This combined effect is shown in Figure 24.12, where the axis of bending runs normal to the edge of the wedge specimen. When $s = 0$, the value of ξ_{eff} is largest. As we bend away from the Bragg condition, on either side, ξ_{eff}

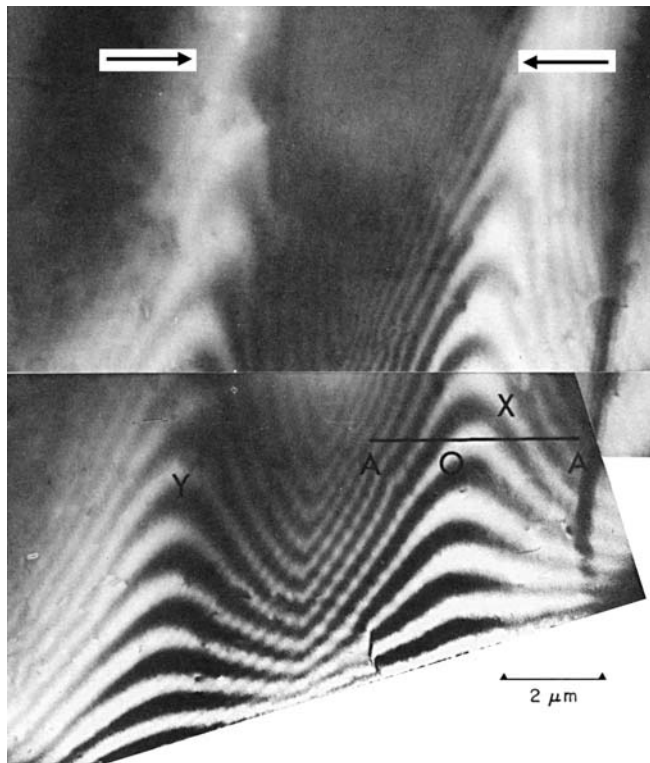


FIGURE 24.12. Since both thickness fringes and bend contours (X and Y) affect the contrast seen in the image, and both can occur in the same part of the specimen, they can affect, or couple with, one another to give the complex contrast shown in this BF image. Along the line $A-A$, s changes in sign, being approximately zero at O and negative between the contours.

decreases so the thickness contours curve towards the edge of the specimen. This image actually shows the $\mathbf{g}(111)$ and $\bar{\mathbf{g}}(\bar{1}\bar{1}\bar{1})$ contours (arrowed). As an exercise you can calculate the value of s at any point between the contours in this image. Assume the wedge angle is constant and $t = 0$ at the edge; then compare the thickness you deduce using ξ_{eff} with the thickness value extrapolated from the regions where $s = 0$.

If a defect causes the specimen to bend, then the contrast from the defect and that from thickness variations will be linked.

Since the effective thickness is s_{eff}^{-1} , it will change as we increase the deviation parameter, s . You can use this fact to determine the thickness of an area quite accurately, provided you have a reference value, such as zero thickness at a hole in the specimen. You initially tilt the specimen so that the chosen reflection is at the Bragg condition. (Remember that this analysis assumes that we have only two beams.) You can determine $\mathbf{s}_{\mathbf{g}}$ quite accurately if you can see the Kikuchi lines. Then you can determine s_{eff}^{-1} at different positions on the specimen. The maximum value of s_{eff}^{-1} is $\xi_{\mathbf{g}}$ and occurs at $\mathbf{s}_{\mathbf{g}} = 0$. As you tilt the specimen to increase $\mathbf{s}_{\mathbf{g}}$ in the positive or negative sense, you'll see the thickness fringes move closer together. We'll examine the situation where $\mathbf{s}_{\mathbf{g}}$ is very large in Chapter 27, and where the foil also bends in Section 24.10.

ON THICKNESS DETERMINATION

Be careful when using fringes for determining t to use in XEDS analysis, since only the thickness of the diffracting (crystalline) material is determined. There may be amorphous material on the surface which has similar or different composition.

24.10 OTHER EFFECTS OF BENDING

In some situations, the bending of the foil may be more subtle. For example, strains in TEM specimens may relax at the surface of the thin specimen. A particularly important example of this effect was found in the study of superlattices in semiconductors.

We'll generalize the situation a little. Imagine that two cubic materials, which normally have slightly different lattice parameters, are grown on one another to form an artificial superlattice with an (001) (i.e., cube-on-cube) interface plane. One crystal must expand and the other contract normal to this interface (the tetragonal distortion). When we prepare a cross section TEM specimen, we might then imagine it relaxing at the surface as shown in Figure 24.13B. The reason for the relaxation is simply that this allows the one material to expand while the other

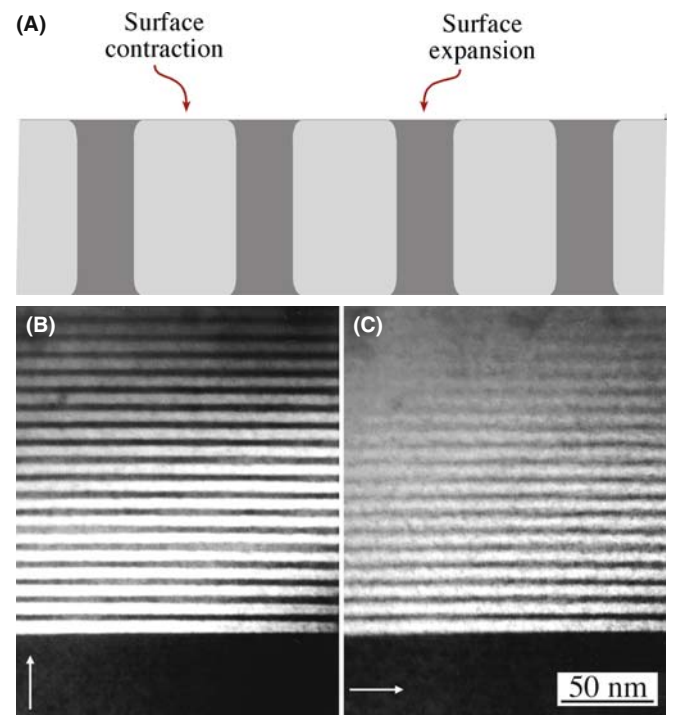


FIGURE 24.13. (A) A schematic of how an interface might relax at the surface of a thin specimen. (B, C) DF images of a GaAs/AlGaAs superlattice imaged in two orthogonal reflections, 200 and 020, with the specimen oriented at the 001 pole. (B) The [200] vector is parallel to the interface while (C) [020] is normal to it. If planes parallel to the interface bend to relax the strain caused by the lattice misfit, then only the 020 image will be affected, giving a more abrupt contrast change.

contracts; the constraint at the surface has been removed during the specimen preparation process. This argument is admittedly simplistic, but Figure 24.13B and C shows that images recorded with $\mathbf{g} = 020$ normal to this interface do indeed appear *sharper* than images formed when $\mathbf{g} = 200$ is parallel to the interface.

So, no matter whether \mathbf{g} is 020 or $0\bar{2}0$, the Bragg planes are bent closer to $\mathbf{s} = 0$ at one surface or the other.

Here the bending only occurs within a short distance of the interface but it significantly affects the appearance of the DF image. The bending is actually making the image appear sharper than it should.

This example is special but emphasizes the point: relaxation at the surface can cause the diffracting planes to bend and this bending will affect the appearance of your image.

CHAPTER SUMMARY

The effects of changes in thickness and specimen bending are both explained by equation 24.1. Although this equation was derived for a two-beam geometry, you'll see similar effects when more strongly excited beams are present but the simple \sin^2 dependence will be lost.

- Varying t while keeping \mathbf{s} constant gives thickness fringes.
- Varying \mathbf{s} while keeping t constant gives bend contours.

Thickness fringes are an interference effect and, with care, can be used to calculate the foil thickness and reveal the topography.

Note that if the two surfaces of the specimen are parallel, then we don't see thickness fringes, even if the specimen is tilted. However, the contrast of that region will depend on the projected thickness.

Bend contours are very useful because they map out the value of \mathbf{s} in the specimen. If your foil is bent around more than one axis, bend contours can combine to produce beautiful ZAPs which reflect the true symmetry of the material.

However, if you want to keep the defect analysis simple, then you need to avoid specimen bending and work in relatively thin regions of nearly constant thickness. There are exceptions to this rule: in a few special cases where you want to do exactly the opposite! So the message is that bending and thickness variations give you extra parameters that you can use in your study as long as you can control these parameters. This control comes from mastering the BF/DF/SAD techniques in Chapter 9.

Last, be aware that anomalous absorption is not anomalous. It can best (only) be explained by Bloch wave interactions.

SOME SPECIFIC READING

Gibson, JM, Hull, R, Bean, JC and Treacy, MMJ 1985 *Elastic Relaxation in Transmission Electron Microscopy of Strained-Layer Superlattices* Appl. Phys. Lett. **46**(7) 649–651. Shows the effect of surface relaxation on the contrast from superlattices.

Rackham, GM and Eades, JA 1977 *Specimen Contamination in the Electron Microscope When Small Probes are Used* Optik **47**(2) 226–232. Example of using CBED for real-space crystallographic analysis.

Susnitzky, DW and Carter, CB 1992 *Surface Morphology of Heat-Treated Ceramic Thin Films* J. Am. Ceram. Soc. **75**(9) 2463–2478. An overview of the surface morphology of heat-treated ceramic thin films studied using TEM.

THE COMPANION TEXT

The simulation programs discussed in the companion text allow you to do the contrast experiments in a computer.

SELF-ASSESSMENT QUESTIONS

- Q24.1 In a BF image, there is a set of dark lines on either side of a particular bend contour. What causes these lines?
- Q24.2 Is it always correct to say “my sample must vary in thickness a lot because I see a lot of thickness fringes”? (Apart from the grammar.)

- Q24.3 In a TEM image, you see a pair of dark lines indicating a bend contour. Is this BF or DF?
- Q24.4 When would a wedge sample give two closely spaced spots in the DP and when would it not?
- Q24.5 Imagine that you can continuously increase the accelerating voltage on the TEM from 200 to 300 kV. How will the thickness fringes change?
- Q24.6 If your specimen is exceptionally thin but bent, how can you find a zone axis?
- Q24.7 A set of bend contours gradually broadens and the contours move apart. What is happening in the specimen?
- Q24.8 If the calculated extinction distance of a sample is 50 nm, what is the maximum thickness of the specimen that you can use to produce reasonable BF images?
- Q24.9 How will thickness fringes curve as the specimen bends away from the Bragg condition?
- Q24.10 Are thickness fringes caused by amplitude contrast or phase contrast?
- Q24.11 Can thickness fringes tell you about surface topography?
- Q24.12 What causes bend contours?
- Q24.13 How do you distinguish between bend contours and dislocations?
- Q24.14 Can a void cause bend contours?
- Q24.15 What is a ZAP and why does it occur?
- Q24.16 Write down the empirical expression for the extinction distance when absorption is included.
- Q24.17 How reliably can you determine the thickness of a specimen using thickness fringes?
- Q24.18 When imaging a GaAs/AlGaAs superlattice grown on (001), which {200} reflection will appear to give the sharpest image and why is this so?
- Q24.19 If you look along [100] at a 90°-wedge specimen of a (001) GaAs quantum well in a AlGaAs matrix, which material will show more thickness fringes in two-beam conditions?
- Q24.20 How thick is the crystal at the center of the third bright fringe in a two-beam DF image (as a multiple of ξ_g)?

TEXT-SPECIFIC QUESTIONS

- T24.1 If the two grains in Figure 24.3A have a common [001] foil normal, describe the shape of the specimen assuming the electron had 120-keV energy.
- T24.2 Is there an error in the caption for Figure 24.3? Estimate the thickness of the sample. Explain all your assumptions.
- T24.3 What is the wedge angle of the two segments of specimen in Figure 24.5A. Explain your assumptions.
- T24.4 In Figure 24.8A, the ± 002 bend contours show that the specimen bends. What is the axis of bending that produces this pair of contours? Deduce the variation in the angle through which the specimen bends.
- T24.5 In Figure 24.8B, index the DP. Estimate the radius of curvature around the $[30\bar{1}]$ axis.
- T24.6 Explain the shape of the specimen in Figure 24.9.
- T24.7 Why does the contrast of the thickness fringes in Figure 24.10 decrease moving left to right?
- T24.8 Why is there a contraction and expansion in Figure 24.13? Does the shading therefore correspond to that in the image?
- T24.9 In Figure 24.11, consider the GaAs region. What voltage was used to form this image? Explain how you would deduce this (and do so) even if you could not look up the original publication.
- T24.10 In equation 24.3, we introduce ξ_g' , the anomalous extinction distance. What is this distance for Cu 220 at 100 and 300 kV? Is this a reasonable value to explain the observations in Figure 24.10 (assuming Cu 220 and 100 kV)? Why is ξ_g' a fudge factor?
- T24.11 A crystal of nickel is prepared in the form of a wedge with its upper surface parallel to (100) and its lower surface to (110). Electrons of energy 100 keV enter through the upper surface and travel in the (001) plane perpendicular to the edge of the wedge, falling on the (020) planes at the exact Bragg angle. If the extinction distance for this reflection is 25 nm, calculate the angular splitting of the reflected beam due to refraction as it leaves through the lower surface. (Courtesy Mike Goringe.)
- T24.12 A specimen of thickness t contains a small inclusion of thickness Δz which scatters electrons differently from the matrix. The crystal is viewed under two-beam dynamical conditions at the Bragg position ($s = 0$) and the effective extinction distances are ξ_{gm} for the matrix and ξ_{gt} for the inclusion, which may be assumed to have the same crystal structure as the matrix. Calculate the visibility of the inclusion as a function of (a) its depth in the specimen and (b) the specimen thickness. How would you expect these results to be modified if the specimen is not precisely at the Bragg position (i.e., $s \neq 0$)? (Courtesy Mike Goringe.)
- T24.13 A bcc sample has been badly bent due to mishandling. You have imaged a bend center and the corresponding ZADP. The zone axis is [011]; draw and index the main diffraction spots you would expect at this zone axis. If the sample is locally bent in the shape of a horse saddle, explain (using a sketch, if necessary) the relationship between the DP and the bent contour lines that form the bend center. How would you confirm the relationship you suggest? (Courtesy Ian Robertson.)

Planar Defects

CHAPTER PREVIEW

Internal interfaces (grain boundaries, phase boundaries, stacking faults) or external interfaces (i.e., surfaces) are surely the most important defects in crystalline engineering materials. Their common feature is that we can usually think of them as all being two-dimensional, or planar, defects (even though they're not really). The main topics of this chapter will be

- Characterizing which type of internal interface we have and determining its main parameters.
- Identifying lattice translations at these interfaces from the appearance of the diffraction-contrast images.

Rotations are usually associated with line defects and they will be discussed in Chapter 26. We can't usually identify the details of the local structure of an interface unless we use HRTEM, so we will return to that topic in Chapter 28. We'll talk a little about simulating images of these defects because it is relatively direct and instructive.

25.1 TRANSLATIONS AND ROTATIONS

This section is a brief summary of some fundamentals of materials science. If the concepts are not immediately familiar to you, then go and study an introductory textbook on defects in materials, several of which are in the reference section.

An interface is simply a surface which separates any two distinct regions of the microstructure. For most of our discussion, we will make the assumption that the surface is flat and is thus a planar (parallel to a plane) defect, though this is rarely the case. We can sketch a general interface as shown in Figure 25.1

With this general definition, we can summarize the different classes of planar defects

- *Translation boundary, RB.* Any translation $\mathbf{R}(\mathbf{r})$ is allowed, θ is zero, and both regions are identical and thus perfectly aligned. Stacking faults (SFs) are a special case. We'll denote the translation boundary as RB so as to avoid confusing it with the twin boundary (TB).
- *Grain boundary, GB.* Any values of $\mathbf{R}(\mathbf{r})$, \mathbf{n} , and θ are allowed (\mathbf{n} is the normal to the defect plane), but the chemistry and structure of the two grains must be the same. The SF is again a special case, but this class also includes TBs.

- *Phase boundary, PB.* As for a GB, but the chemistry and/or structure of the two regions can differ.
- *Surface.* A special case of a PB where one phase is vacuum or gas.

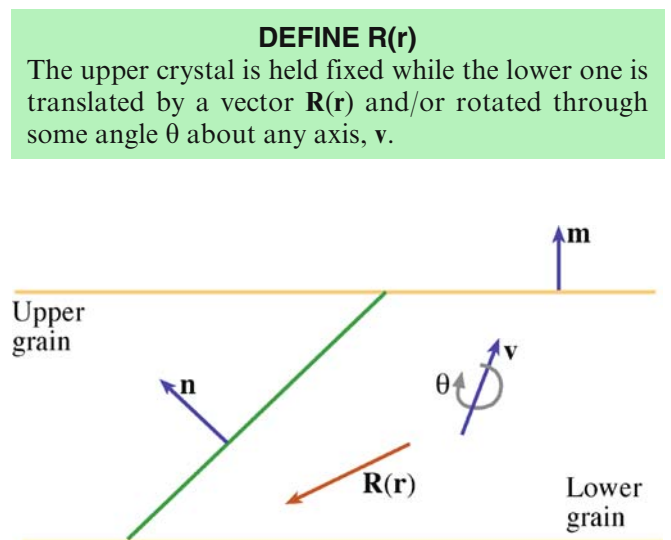


FIGURE 25.1. A specimen containing a planar defect. The lower grain is translated by a vector $\mathbf{R}(\mathbf{r})$, and rotated through an angle θ about the vector \mathbf{v} , relative to the upper grain. The defect plane normal is \mathbf{n} , the foil normal is \mathbf{m} .

TABLE 25.1 Examples of Internal Planar Defects

Planar defect	Structure	Example	Example
SF	Diamond-cubic, fcc, zinc blende	Cu, Ag, Si, GaAs	$R = \frac{1}{3}[111]$ or $R = \frac{1}{6}[112]$
APB/IDB	Zinc blende, wurtzite	GaAs, AlN	Inversion
APB	CsCl	NiAl	$R = \frac{1}{2}[111]$
APB/SF	Spinel	MgAl ₂ O ₄	$R = \frac{1}{4}[110]$
GB	All materials	Often denoted by Σ where Σ^{-1} is the fraction of coincident lattice sites	Rotation plus R
PB	Any two different materials	Sometimes denoted by Σ_1, Σ_2 which are not equal	Rotation plus R plus misfit
RB	Extra translation	{112} Twin boundary in Al	R not related to lattice

Now with each of these groups, we can have special examples. We will list some of the most common examples in Table 25.1, including those that we will consider in this chapter. For a more detailed discussion, we refer you to the general references at the end of the chapter.

RBs include the familiar SFs found in fcc, hcp, diamond-cubic, and layer materials. They have been widely studied because they play an important role in the mechanical properties of the fcc metals, e.g., Cu and stainless steel. They are also found in more complex materials such as spinels, Ni₃Al, Ti₃Al, etc., where the lattice parameters, and therefore the dislocation Burgers vectors, are large.

The anti-phase boundary (APB) in ordered CuAu (which we can describe as two interpenetrating simple-cubic superlattices) is produced by translating one superlattice by $\frac{1}{2} \langle 111 \rangle$ with respect to the other. It is called an APB because one superlattice is out of phase with the other. If the crystal were disordered and the Cu and Au occupied the bcc sites randomly, then $\frac{1}{2} \langle 111 \rangle$ would be a lattice vector and no defect would exist. This particular APB can thus be regarded as an SF. We will find that the methods used to characterize RBs can often be used to determine $R(r)$ in other interfaces.

STACKING FAULT PLANES

Although we know that {111} is the favored SF plane for fcc metals, SFs in other materials lie on different planes.

GBs fall into two groups, low angle and high angle. Low-angle boundaries necessarily involve a rotation through a small angle that is usually accommodated by arrays of dislocations; we'll consider these defects in Chapter 26. High-angle boundaries can adopt some special values of n and θ ; a large fraction of lattice sites in one grain is then shared by (is common to) the other grain. We characterize the fraction by the inverse of this fraction, which we call Σ . For example, the common

twin boundary in fcc metals is the $\Sigma = 3$ grain boundary (the fraction is $\frac{1}{3}$). The reason this is important to our discussion is that if a set of lattice points is common to two grains (as implied by the Σ coincidence-site lattice concept), then certain planes may also be common and may give rise to common reflections. These reflections will remain common even if one grain is translated relative to the other. In that case, we'll have a special type of RB, called the rigid-body translation. Rigid-body translations in grain boundaries behave just like other SFs except R is usually small and is not directly related to the lattice parameters.

There is a second group of APBs where the two grains cannot be related by a translation. These occur in GaAs, ZnO, AlN, and SiC, for example. One lattice can always be related to the other by a rotation of 180° to give the equivalent of an inversion; they are sometimes known as inversion domain boundaries (IDBs). (They are $\Sigma = 1$ GBs.) These special interfaces can often be seen in the image because there is usually a small associated translation. We analyze this translation as if it were a simple RB because all the planes on one side of the IDB are parallel to their counterparts in the other; the (hkl) plane on one side is parallel to the $\bar{h}\bar{k}\bar{l}$ on the other. We can't distinguish g from \bar{g} unless we use CBED or we have a great specimen for HRTEM.

PBs are rarely (if ever) analyzed fully using diffraction contrast. If the orientation, chemistry, and structure can all change on crossing the boundary, then not only will the reflections change, but all extinction distances will change too. Some special examples of such interfaces are hcp-Co/fcc-Co, bcc-Fe/fcc-Fe, NiO/NiFe₂O₄, and GaAs/Al_xGa_{1-x}As. Of course, the number of other such interfaces is countless.

Surface studies using TEM have quite recently become very important although the experimental tools have been available for some time. We will discuss surfaces in this chapter insofar as they are imaged by diffraction contrast. So-called profile imaging will follow in Chapter 28 on HRTEM. The other two surface-sensitive techniques are plan-view and reflection electron microscopy (REM, see Chapter 29).

25.2 WHY DO TRANSLATIONS PRODUCE CONTRAST?

As usual, we will start our analysis considering only two beams, \mathbf{O} and \mathbf{G} . Our approach will use hand-waving arguments, which are not perfect, to justify adapting the Howie-Whelan equations for specimens containing interfaces. We'll use the same approach for other defects in Chapters 26 and 27. Because the Howie-Whelan equations for perfect crystals assume two-beam conditions, we are able to solve them analytically. We'd like to be able to do the same when defects are present, because this gives us a physical understanding of the processes which produce the contrast. There are two important features that we will need to keep in mind

- Diffraction contrast only occurs because we have Bloch waves in the crystal. However, our analysis will initially only consider diffracted beams.
- We make the column approximation so we can solve the equations; we must be wary whenever the specimen or the diffraction conditions change within a distance comparable to the column diameter.

TEM OF SURFACES

The techniques are (i) profile imaging, (ii) plan-view imaging, and (iii) reflection electron microscopy (REM).

A unit cell in a strained crystal will be displaced from its perfect-crystal position so that it is located at position \mathbf{r}'_n instead of \mathbf{r}_n where n is included to remind us that we are considering scattering from an array of unit cells; we'll soon omit the n (and we still have the column approximation)

$$\mathbf{r}'_n = \mathbf{r}_n + \mathbf{R}_n \quad (25.1)$$

In this expression, \mathbf{R}_n is actually $\mathbf{R}_n(\mathbf{r})$; it can vary throughout the specimen. The term $e^{2\pi i \mathbf{K} \cdot \mathbf{r}}$ in equation 13.3 (it's actually $\mathbf{K} \mathbf{r}_n$) now becomes $e^{2\pi i \mathbf{K} \cdot \mathbf{r}'}$ so we need to examine the term $\mathbf{K} \cdot \mathbf{r}'$. We know that \mathbf{K} is $\mathbf{g} + \mathbf{s}$, so we can write

$$\mathbf{K} \cdot \mathbf{r}'_n = (\mathbf{g} + \mathbf{s}) \cdot (\mathbf{r}_n + \mathbf{R}_n) = \mathbf{g} \cdot \mathbf{r}_n + \mathbf{g} \cdot \mathbf{R}_n + \mathbf{s} \cdot \mathbf{r}_n + \mathbf{s} \cdot \mathbf{R}_n \quad (25.2)$$

Now since \mathbf{r}_n is a lattice vector, $\mathbf{g} \cdot \mathbf{r}_n$ is an integer as usual. The third term, $\mathbf{s} \cdot \mathbf{r}_n$ gives our usual sz term so the new terms are $\mathbf{g} \cdot \mathbf{R}_n$ and $\mathbf{s} \cdot \mathbf{R}_n$.

When we discuss strong-beam images we know that \mathbf{s} is very small. Since we are using elasticity theory \mathbf{R}_n must be small. Hence we ignore the term $\mathbf{s} \cdot \mathbf{R}_n$. Remember that we have made a special assumption which may not be valid in two situations

- When \mathbf{s} is large; we'll encounter this when we discuss the weak-beam technique in Chapter 27.
- When the lattice distortion, \mathbf{R} , is large; this occurs close to the cores of some defects.

We now modify equation 13.8 intuitively to include the effect of adding a displacement from equation 25.2

$$\frac{d\phi_{\mathbf{g}}}{dz} = \frac{\pi i}{\xi_0} \phi_{\mathbf{g}} + \frac{\pi i}{\xi_{\mathbf{g}}} \phi_0 \exp[-2\pi i(sz + \mathbf{g} \cdot \mathbf{R})] \quad (25.3)$$

and

$$\frac{d\phi_0}{dz} = \frac{\pi i}{\xi_0} \phi_0 + \frac{\pi i}{\xi_{\mathbf{g}}} \phi_{\mathbf{g}} \exp[+2\pi i(sz + \mathbf{g} \cdot \mathbf{R})] \quad (25.4)$$

Next, we simplify these equations just as we did in Chapter 13 by setting

$$\phi_0(z)_{(\text{sub})} = \phi_0 \exp\left(\frac{-\pi i z}{\xi_0}\right) \quad (25.5)$$

and

$$\phi_{\mathbf{g}}(z)_{(\text{sub})} = \phi_{\mathbf{g}} \exp\left(2\pi i s z - \frac{\pi i z}{\xi_0}\right) \quad (25.6)$$

Then the Howie-Whelan equations become

$$\frac{d\phi_{\mathbf{g}(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}} \phi_{\mathbf{g}(\text{sub})} \exp(2\pi i \mathbf{g} \cdot \mathbf{R}) \quad (25.7)$$

and

$$\frac{d\phi_{\mathbf{g}(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}} \phi_{\mathbf{g}(\text{sub})} \exp(-2\pi i \mathbf{g} \cdot \mathbf{R}) + 2\pi i s \phi_{\mathbf{g}(\text{sub})} \quad (25.8)$$

These equations are just as before (equations 13.14 and 13.15) but with the addition of the $2\pi i \mathbf{g} \cdot \mathbf{R}$ term. This additional phase is termed α , hence planar defects are seen when $\alpha \neq 0$.

$$\alpha = 2\pi \mathbf{g} \cdot \mathbf{R} \quad (25.9)$$

These expressions will be particularly useful in two cases

- When $\mathbf{R} = \text{constant}$.
- Understanding phasor diagrams when defects are present.

We start with a simple stacking fault lying parallel to the surface as shown in Figure 25.2. In this situation, the beams propagate through the upper layer just as if no fault were present. At a depth $z = t_1$, the beams may experience a phase change due to the effect of the

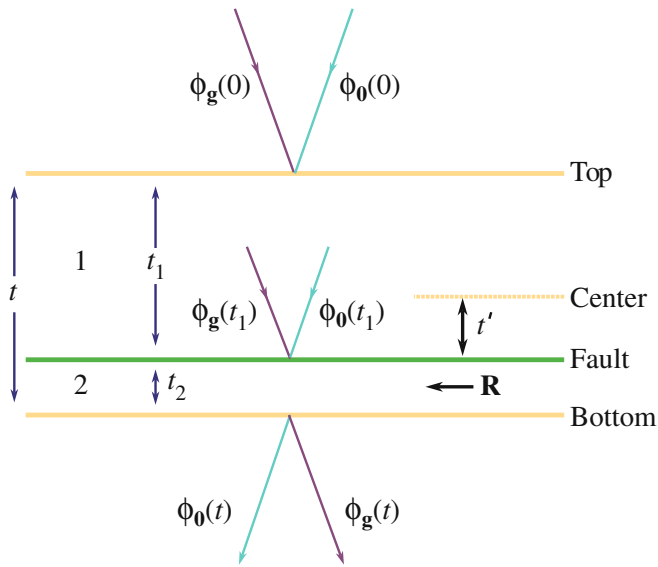


FIGURE 25.2. A stacking fault lying at depth t_1 in a parallel-sided uniformly thick specimen. The total thickness is t and $t_2 = t - t_1$.

translation \mathbf{R} , but after that they again propagate as if in perfect crystal.

In this chapter, we'll see several values of α . A special case occurs when $\alpha = \pm 120^\circ$. This value of α is often encountered since it occurs for fcc SFs. We'll also encounter the case where $\alpha = \pm 180^\circ$; this value arises for some special APBs which are really SFs.

25.3 THE SCATTERING MATRIX

This discussion of the scattering matrix introduces no new concepts. It is just a different way of writing the equations so that if you are calculating the image contrast, you can program the computer more easily, especially when you have complicated arrays of lattice defects. Our reason for delaying the introduction of the scattering matrix until now is that it is much easier to understand when you can apply it to a specific problem.

In equations 13.17–13.20, we showed that in the two-beam case, we can write these simple expressions for ϕ_0 and ϕ_g

$$\phi_0 = C_0 e^{2\pi i \gamma z} \quad (25.10)$$

and

$$\phi_g = C_g e^{2\pi i \gamma z} \quad (25.11)$$

Since there are two values for γ , we can express both the $\mathbf{0}$ and \mathbf{g} beams as the combination of these two contributions to give

$$\phi_0(z) = C_0^{(1)} \psi^{(1)} \exp(2\pi i \gamma^{(1)} z) + C_0^{(2)} \psi^{(2)} \exp(2\pi i \gamma^{(2)} z) \quad (25.12)$$

and

$$\phi_g(z) = C_g^{(1)} \psi^{(1)} \exp(2\pi i \gamma^{(1)} z) + C_g^{(2)} \psi^{(2)} \exp(2\pi i \gamma^{(2)} z) \quad (25.13)$$

where the $\psi^{(i)}$ terms tell us the relative contributions of the $\gamma^{(1)}$ and $\gamma^{(2)}$ terms. (We are really saying that both Bloch waves contribute to both the $\mathbf{0}$ and \mathbf{g} beams.) We can rewrite equations 25.12 and 25.13 in a matrix form. (This is the key step.)

$$\begin{pmatrix} \phi_0(z) \\ \phi_g(z) \end{pmatrix} = \begin{pmatrix} C_0^{(1)} & C_0^{(2)} \\ C_g^{(1)} & C_g^{(2)} \end{pmatrix} \begin{pmatrix} \exp(2\pi i \gamma^{(1)} z) & 0 \\ 0 & \exp(2\pi i \gamma^{(2)} z) \end{pmatrix} \begin{pmatrix} \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} \quad (25.14)$$

We can express our boundary conditions as

$$C_0^{(1)} \psi^{(1)} + C_0^{(2)} \psi^{(2)} = \phi_0(0) \quad (25.15)$$

and

$$C_g^{(1)} \psi^{(1)} + C_g^{(2)} \psi^{(2)} = \phi_g(0) \quad (25.16)$$

which we can now rewrite as

$$\begin{pmatrix} C_0^{(1)} & C_0^{(2)} \\ C_g^{(1)} & C_g^{(2)} \end{pmatrix} \begin{pmatrix} \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} = \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \quad (25.17)$$

(We actually saw in Section 13.9 that $\phi_0(0)$ is 1 and $\phi_g(0)$ is 0 because $z = 0$ is the top surface.) We know from Chapters 13 and 14 that the elements of the matrix C are determined by the two-beam conditions that we set up. The matrix C does not depend on z . Now we can use matrix algebra to solve equation 25.17. First rewrite it as

$$C \begin{pmatrix} \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} = \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \quad (25.18)$$

then rewrite equation 25.18 as

$$\begin{pmatrix} \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} = C^{-1} \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \quad (25.19)$$

where C^{-1} is just the inverse matrix. Remember that the order is important in matrix multiplication and that $C^{-1}C = I$, the unit matrix.

Therefore we can rewrite equation 25.14 as

$$\begin{pmatrix} \phi_0(z) \\ \phi_g(z) \end{pmatrix} = C \begin{pmatrix} \exp(2\pi i \gamma^{(1)} z) & 0 \\ 0 & \exp(2\pi i \gamma^{(2)} z) \end{pmatrix} C^{-1} \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \quad (25.20)$$

Finally, we can define a new matrix $P(z)$ as the scattering matrix for a slice of perfect material with thickness z

$$P(z) = C \begin{pmatrix} \exp(2\pi i \gamma^{(1)} z) & 0 \\ 0 & \exp(2\pi i \gamma^{(2)} z) \end{pmatrix} C^{-1} = C \Gamma C^{-1} \quad (25.21)$$

The matrix $P(z)$ thus gives us the values of the exit wave amplitudes at the bottom of the slice in terms of the incident values. In other words, the matrix $P(z)$ includes all the information to describe the propagation of the beams through the crystal; $P(z)$ is a *propagator* matrix.

P = C Γ C⁻¹ AND VARIES WITH z

Notice that z only enters the ‘ P equation’ through the Γ matrix.

25.4 USING THE SCATTERING MATRIX

Now we illustrate the real strength of the scattering matrix approach by considering the effect of a planar fault lying parallel to the foil surface, as we saw in Figure 25.2. The idea is that we now have two slices of material of thickness t_1 and t_2 . We can easily calculate $\phi_0(t_1)$ and $\phi_g(t_1)$ using equation 25.20. These values for ϕ_0 and ϕ_g then become the incident values for slice 2. The effect of the translation \mathbf{R} is to multiply the terms in C_g in the lower slice by a phase factor $\exp(-i\alpha)$ where $\alpha = 2\pi \mathbf{g} \cdot \mathbf{R}$ as usual. The matrix C for slice 2 is then written as

$$C_2 = \begin{pmatrix} C_0^{(1)} & C_0^{(2)} \\ C_g^{(1)} \exp(i\alpha) & C_g^{(2)} \exp(i\alpha) \end{pmatrix} \quad (25.22)$$

We can write down the expression for $\phi_0(t)$ and $\phi_g(t)$ as

$$\begin{pmatrix} \phi_0(t) \\ \phi_g(t) \end{pmatrix} = C_2 \Gamma(t_2) C_2^{-1} C_1 \Gamma(t_1) C_1^{-1} \begin{pmatrix} \phi_0(0) \\ \phi_g(0) \end{pmatrix} \quad (25.23)$$

Here the subscripts on C_1 and C_2 just identify the slices. Normally, this equation goes straight into the computer, which handles matrices very easily. However, we’ll go back and consider a few special points

- Look at equation 25.22 and set $\mathbf{R} = 0$ so that $C_2 = C_1$. You can see that $P(t) = P(t_1)P(t_2)$. Clearly we could cut the perfect-crystal specimen into

many slices and $P(t)$ would always be the product of the scattering matrices for each slice. (We could call this a multislice approach to scattering from a crystal but we don’t to avoid confusion with HRTEM simulation.)

- How do we prove equation 25.22? From equation 14.12 we know that a Bloch wave can be written as

$$b(\mathbf{k}) = \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k}) \exp(2\pi i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{r}) \quad (25.24)$$

- If the lower crystal is displaced by a vector \mathbf{R} then we replace \mathbf{r} by $\mathbf{r} - \mathbf{R}$ (notice the sign). (We have just used a ‘hidden’ column approximation.) Equation 25.24 is then written as

$$b(\mathbf{k}) = \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k}) \exp(2\pi i(\mathbf{k} + \mathbf{g}) \cdot (\mathbf{r} - \mathbf{R})) \quad (25.25)$$

$$b(\mathbf{k}) = e^{-2\pi i \mathbf{k} \cdot \mathbf{R}} \sum_{\mathbf{g}} C_{\mathbf{g}}(\mathbf{k}) e^{(-2\pi i \mathbf{g} \cdot \mathbf{r})} e^{2\pi i(\mathbf{k} + \mathbf{g}) \cdot \mathbf{r}} \quad (25.26)$$

C_0 is not affected by \mathbf{R} since $2\pi \mathbf{0} \cdot \mathbf{R} = 0$, but C_g is multiplied by $e^{-i\alpha}$.

- If you choose the coordinates appropriately then C is a unitary matrix. In this case, you can find C^{-1} just by reflecting across the diagonal and taking the complex conjugate of each term. This *trick* will allow you to express equation 25.23 explicitly, as given by Hirsch et al. (omitting a phase factor).

$$\begin{aligned} \phi_0(t) = & [\cos(\pi \Delta k t) - i \cos(\beta) \sin(\pi \Delta k t)] \\ & + \frac{1}{2} (e^{i\alpha} - 1) \sin^2 \beta \cos(\pi \Delta k t) \\ & - \frac{1}{2} (e^{i\alpha} - 1) \sin^2 \beta \cos(\pi \Delta k t') \end{aligned} \quad (25.27)$$

$$\begin{aligned} \phi_g(t) = & i \sin(\beta) \sin(\pi \Delta k t) \\ & + \frac{1}{2} \sin \beta (1 - e^{(-i\alpha)}) [\cos \beta \cos(\pi \Delta k t) - i \sin(\pi \Delta k t)] \\ & - \frac{1}{2} \sin \beta (1 - e^{(-i\alpha)}) [\cos \beta \cos(\pi \Delta k t') - i \sin(\pi \Delta k t')] \end{aligned} \quad (25.28)$$

In equations 25.27 and 25.28, t' is the distance of the fault below the center of the slice, i.e., we define $t' = t_1 - t/2$ where t_1 lies between 0 and t . (It’s a good, but tedious, exercise to derive these equations for yourself.) The right-hand side of equations 25.27 and 25.28 each contains three terms.

- The first term is just what we found in Chapter 13 where the phase factor $\alpha = 0$, i.e., it’s just like the perfect crystal.
- The second term is independent of the position of the planar fault because it doesn’t depend on t' .

- The third term depends on t' such that both ϕ_0 and ϕ_g change with a periodicity in t' given by Δk^{-1} . So these amplitudes show the same dependence on ξ_g^{eff} . They will both show thickness variations.

THE GEOMETRY

You should keep in mind that we derived these equations for a planar defect lying parallel to the parallel surfaces of our specimen and normal to the beam.

We can now take these ideas and apply them to planar defects which are inclined to the surface, by calculating the contrast for all values of t' between 0 and t . The important points to remember are

- The model used in the calculation was a flat interface parallel to the surface of a plate-like specimen. You'll see fault fringes when t' varies across the fault but you don't usually have to consider the fact that either the surface or the fault may be inclined to the beam.
- The concept of the scattering matrix allows you to identify very clearly the effect of the defect on ϕ_0 and ϕ_g .

25.5 STACKING FAULTS IN FCC MATERIALS

We'll begin our discussion of actual examples with the SF in fcc materials. Before we discuss the details of contrast from SFs in fcc materials, we'll summarize the important results which hold for all planar defects

- The appearance of the image depends on the specimen thickness.
- Pairs of BF/DF SF images are not generally complementary even though we are using a two-beam approximation. Compare to the complementary behavior of the thickness fringes discussed in Chapter 24.
- Planar defects always really have a thickness. We'll illustrate this concept using overlapping faults in fcc materials (see also Section 26.6).

NOT ALL MATERIALS ARE FCC

Do not assume all faults are the same as in fcc materials!

25.5.A Why fcc Materials?

There are several reasons for emphasizing the analysis of stacking faults in fcc crystals

- Many important materials are fcc, including the metals Cu, Ag, Au, and austenitic stainless steel, and the semiconductors Si, Ge, and GaAs.
- Most of the analysis of SFs derives from the study of fcc materials.
- The translations are well known and directly related to the lattice parameter: \mathbf{R} is either $1/6\langle\bar{1}\bar{1}2\rangle$ or $1/3\langle 111\rangle$. Notice that these definitions differ by the lattice vector, $1/2\langle 110\rangle$. (Actually there may be small deviations from these ideal values but we'll ignore them for now.)

We want to learn how to extend this analysis to other fault vectors and avoid making unfounded assumptions when we do extend it. The geometry often encountered when studying defects in fcc materials is shown in Figure 25.3.

You should note that $(11\bar{1})$ is one of three possible planes for an inclined SF. In this case, the translation at the stacking fault will be $\mathbf{R} = \pm 1/3[11\bar{1}]$; the phase factor, α , is $2\pi \mathbf{g} \cdot \mathbf{R}$. If we form an image with the $\mathbf{g} = (2\bar{2}0)$ reflection strongly excited then $\mathbf{g} \cdot \mathbf{R} = 0$ and the fault is out of contrast in both BF and DF. If instead we use the reflection $\mathbf{g} = (02\bar{2})$ then $\mathbf{g} \cdot \mathbf{R} = 4/3$ or $-4/3$ and $\alpha = 8\pi/3 = 2\pi/3 = 120^\circ$ or $-8\pi/3 = -2\pi/3 = -120^\circ$ (modulo 2π in each case). Notice that if the stacking fault lies parallel to the surface of this (111) -oriented specimen, you must tilt the specimen to see *any* contrast from the SF, i.e., $\mathbf{g} \cdot \mathbf{R} = 0$ for all values of \mathbf{g} lying in the fault plane.

INVISIBILITY CRITERION

When $\mathbf{g} \cdot \mathbf{R} = 0$ the defect is 'out of contrast'; it's invisible.

Figure 25.4A and B shows two typical BF/DF pairs of $\pm \mathbf{g}$ strong-beam images from the same SF. In the BF images, the outer fringes are the same on both sides of the fault, both gray or both white while in the DF images one outer fringe is white but the other is gray. The questions which arise are

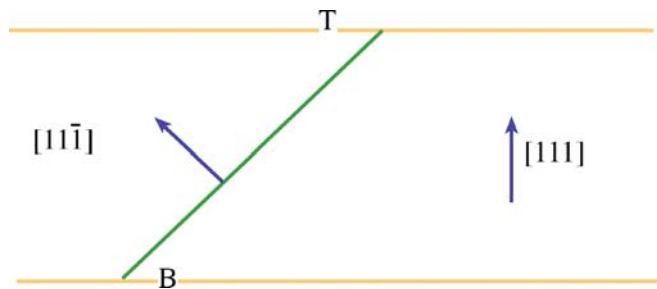


FIGURE 25.3. A stacking fault in a parallel-sided fcc specimen. The normal to the specimen is $[111]$ and the normal to the SF is $[11\bar{1}]$. T and B indicate the top and bottom of the foil.

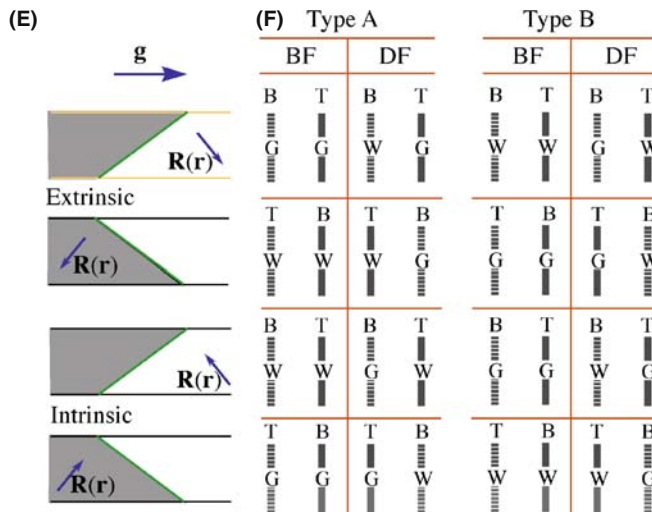
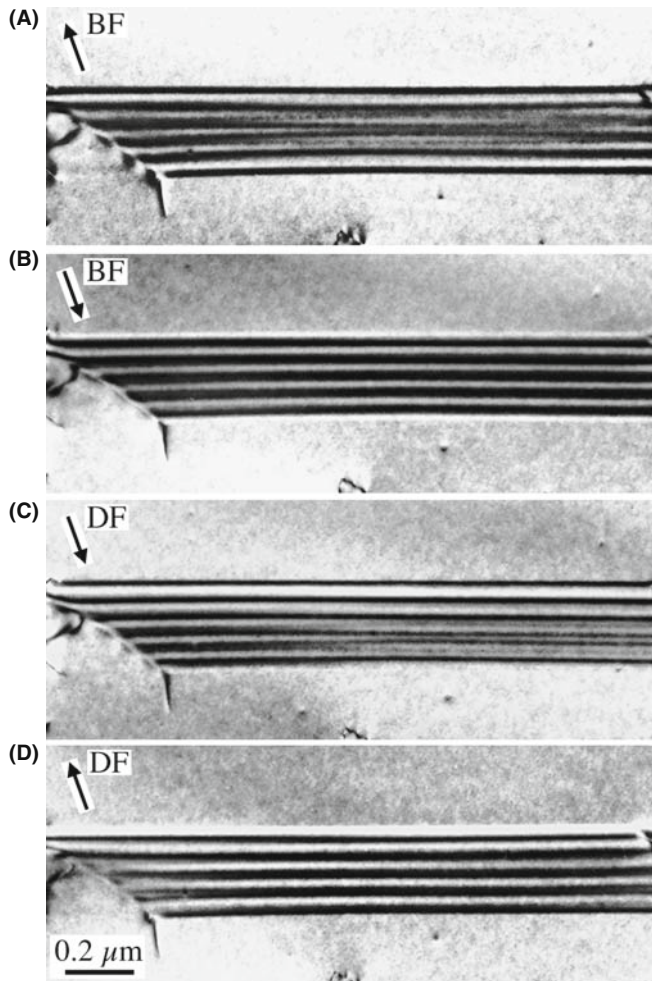


FIGURE 25.4. (A–D) Four strong-beam images of an SF recorded using $\pm\mathbf{g}$ BF and $\pm\mathbf{g}$ DF. The beam was nearly normal to the surfaces; the SF-fringe intensity is similar at the top surface but complementary at the bottom surface. The rules are summarized in (E) and (F) where G and W indicate that the first fringe is gray or white; (T, B) indicates top/bottom.

- What determines whether a fringe will be gray or white?
- Why are the two images not complementary?

Note that in the summary schematic shown in Figure 25.4E, 200, 222, and 440 are type A reflections while 111, 220, and 400 are type B reflections. Also we'll call the first fringe gray (G), not black (i.e., B), to reduce confusion!

25.5.B Some Rules

There are some experimental rules

- Be very careful when you record such a pair of images: record the DP for each image. Be sure to note which of the two bright spots corresponds to the direct beam.
- Use the same strong hkl reflection for BF and DF imaging. Therefore, to form the CDF image using a strong hkl reflection, you must first tilt the specimen so that $\bar{h}\bar{k}\bar{l}$ is strong and then use the beam tilts to move hkl onto the optic axis where it will become strong (see Section 22.5). This is confusing, so we recommend that you sacrifice a little image resolution and compare the BF image with a displaced-aperture DF (DADF) image, rather than a CDF image.

This is exactly the opposite of the approach used by Edington, who advocates tilting in $\bar{h}\bar{k}\bar{l}$ for the DF image, which reverses the DF contrast in Figure 25.4E and F. Our approach ensures that diffraction from the same hkl planes causes the contrast in both DF and CDF images but DADF beats both! (TEMs were not as good in 1975 as they are now.)

Then there are some rules for interpreting the contrast

- In the image, as seen on the screen or on a print, the fringe corresponding to the top surface (T) is white in BF if $\mathbf{g} \cdot \mathbf{R}$ is > 0 and black if $\mathbf{g} \cdot \mathbf{R} < 0$.
- Using the same strong hkl reflection for BF and DF imaging, the fringe from the bottom (B) of the fault will be complementary whereas the fringe from the top (T) will be the same in both the BF and DF images.
- The central fringes fade away as the thickness increases. If this seems anomalous, the explanation is in Section 25.10.
- The reason it is important to know the sign of \mathbf{g} is that you will use this information to determine the sign of \mathbf{R} .
- For the geometry shown in Figure 25.3, if the origin of the \mathbf{g} vector is placed at the center of the SF in the DF image, the vector \mathbf{g} points away from the bright outer fringe if the fault is extrinsic and toward it if it is intrinsic (200, 222, and 440 reflections); if the reflection is a 400, 111, or 220 the reverse is the case.

- Don't forget that, as we said at the start of Chapter 22, any contrast must be $> \sim 5\text{--}10\%$ to be visible to the eye, so we traditionally say an intensity change due to $\mathbf{g}\cdot\mathbf{R}$ effects is only detectable if $\mathbf{g}\cdot\mathbf{R} > 0.02$; of course, now you can digitally record the TEM image or digitize your analog version and then process this so the limit could be lower. With experience, you'll find there is an optimum thickness to view defect contrast, before absorption effects make it difficult. You must also carefully select \mathbf{s} so that the background intensity in the matrix around the defect is gray and this maximizes visibility of lighter and darker fringes.

DISPLACE THE APERTURE

To avoid the possibility of confusion about which reflection to use for the DF image, just displace the aperture to the strong hkl reflection in every case. In a modern IVEM, there is almost no loss of resolution between DADF and CDF. DADF is always preferred but CDF is used to give much better images.

As we said, these complex rules are summarized in Figure 25.4C and D. Although they are very useful, in practice you should remember that they were derived for a very special combination of \mathbf{R} and \mathbf{g} in fcc materials. Some important examples of $\mathbf{g}\cdot\mathbf{R}$ are given in Table 25.2. As we'll describe in Section 25.11, if this is your research, you should use a computer program to check the contrast.

INVISIBILITY

Just because you can't see a defect doesn't mean it isn't there or that $\mathbf{g}\cdot\mathbf{R} = 0$.

25.5.C Intensity Calculations

Now let's consider intensity calculations using the column approximation, which we briefly discussed in Section 13.11. If the fault cuts the column at a depth t_1 we can deduce from equations 25.22 and 25.23 that

$$\phi_{\mathbf{g}} = \frac{i\pi}{s\xi_{\mathbf{g}}} \left\{ \int_0^{t_1} e^{-2\pi i s z} dz + e^{-i\alpha} \int_{t_1}^t e^{-2\pi i s z} dz \right\} \quad (25.29)$$

TABLE 25.2. Values of $\mathbf{g}\cdot\mathbf{R}$ for Some Common \mathbf{R} Plus \mathbf{g} Combinations

	\mathbf{R}	\mathbf{g}	$\alpha = 2\pi \mathbf{g}\cdot\mathbf{R}$ (mod 2π)
SF in fcc	$\frac{1}{3}[111]$	(111), (220), (113)	$2\pi/3$
SF in fcc	$\frac{1}{3}[111]$	(113)	$4\pi/3$
Translation at APB in Fe_3Al	$\frac{1}{2}[110]$	(100)	π
Small \mathbf{R} , e.g., NiO	Any	\mathbf{g} or \mathbf{s} or $\xi_{\mathbf{g}}$ differ slightly	δ

which gives

$$\phi_{\mathbf{g}} = \frac{i\pi}{s\xi_{\mathbf{g}}} e^{-2\pi i s t_1} \left\{ \sin(\pi s t_1) + e^{-i\alpha} \sin(\pi s(t - t_1)) \right\} \quad (25.30)$$

We rearrange equation 25.30 to give an expression for the intensity, $I_{\mathbf{g}}$ ($= \phi_{\mathbf{g}} \cdot \phi_{\mathbf{g}}^*$). This rearrangement involves a little manipulation.

$$I_{\mathbf{g}} = \frac{1}{(s\xi_{\mathbf{g}})^2} \left\{ \sin^2\left(\pi s t_1 + \frac{\alpha}{2}\right) + \sin^2\left(\frac{\alpha}{2}\right) - \sin\left(\frac{\alpha}{2}\right) \sin\left(\pi s t + \frac{\alpha}{2}\right) \cos(2\pi s t') \right\} \quad (25.31)$$

where $t' = t_1 - t/2$ as before. So the contrast depends on both the thickness and the depth. Note that $t/2$ is the center of the foil. Since α is fixed for a particular defect, let's fix t . Then equation 25.31 becomes

$$I_{\mathbf{g}} \propto \frac{1}{s^2} \{A - B \cos(2\pi s t')\} \quad (25.32)$$

Now we have cosine depth fringes or defect thickness fringes, just as we did for the perfect crystal

- The thickness periodicity depends on s^{-1} .
- The intensity varies as s^{-2} .

We could have derived this equation from equation 25.28 with more work. However, the value of the scattering matrix approach is that we don't derive the analytical expression but just run the computer program.

In Chapter 27, we will discuss this SF contrast in terms of phasor diagrams which give a graphical way to represent these equations.

25.5.D Overlapping Faults

It is interesting to extend this analysis to the case of overlapping faults. Taking the analytical approach, we can extend equation 25.29 to the case of two overlapping faults, the first at depth t_1 and the second at depth t_2 .

$$\phi_{\mathbf{g}} = \frac{i\pi}{s\xi_{\mathbf{g}}} \left\{ \int_0^{t_1} e^{-2\pi i s z} dz + e^{-i\alpha} \int_{t_1}^{t_1+t_2} e^{-2\pi i s z} dz + e^{-i(\alpha_1+\alpha_2)} \int_{t_1+t_2}^t e^{-2\pi i s z} dz \right\} \quad (25.33)$$

An experimental illustration of a somewhat more complex situation, involving several overlapping SFs, is shown in Figure 25.5

- It sometimes appears that there is no contrast, even when we know that there are overlapping SFs. This can happen if, e.g., three SFs overlap on adjacent (or nearly adjacent) planes; then the effective R can

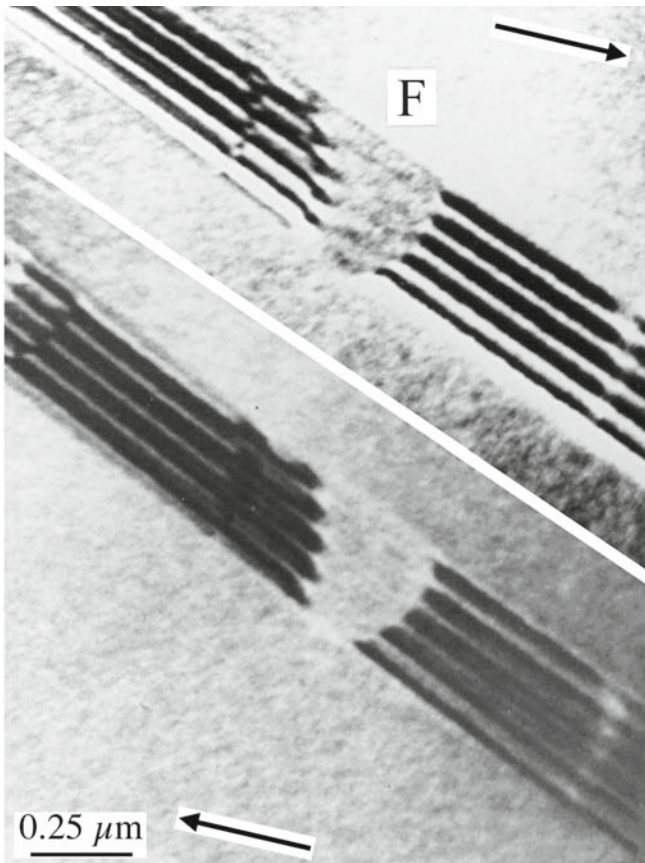


FIGURE 25.5. Two BF images of overlapping SFs in fcc steel with the direction \mathbf{g} indicated. The faults are very close together. When three faults overlap at F the effective value of \mathbf{R} is 0 so the contrast disappears.

$\mathbf{be} = 3 \times \frac{1}{3} [11\bar{1}]$, which is a perfect lattice vector and can therefore appear to give $2\pi\mathbf{g} \cdot \mathbf{R} = 0$.

We will return to this topic in Section 27.8, where we'll show that some planar defects, such as the extrinsic SF in Si, or the dissociated $\{112\}$ twin boundary in some fcc metals, really have a significant thickness. We can then analyze the contrast from such interfaces using the overlapping-fault model.

25.6 OTHER TRANSLATIONS: π AND δ FRINGES

We discussed the $L1_0$ structure of NiAl in Section 16.5. This intermetallic is an example of a large group of materials which can contain a different type of RB. If the Ni atoms sit at the corners of the cell in one crystal region but the Al atoms sit at the corners in another part of the crystal, then the two crystal regions are related by a translation of $\frac{1}{2}[111]$. The two crystals would otherwise still be perfectly aligned but are separated by this RB, which we call an APB (but which could be called an SF or an IDB).

Similarly in the $L1_2$ structure of the intermetallic Ni_3Al , we could have the Al atoms on the corners of the unit cell in one part of the crystal and displaced by $\mathbf{R} = \frac{1}{2}[110]$ in the adjacent region. (We can actually form six non-equivalent APBs in this structure and these are *not* IDBs.) The crystal structure looks like fcc but the Al atoms are at the corners of the unit cell (forming the simple-cubic superlattice) with the Ni at the face-centered positions. The easy way to appreciate this RB is to think what would happen if the alloy were completely disordered: there would be no planar defect. This RB can be imaged using the (100) reflection. Notice again that for a disordered structure, the $\{100\}$ reflections would be absent if the alloy were disordered; the $\{100\}$ planes are said to give rise to superlattice reflections; these reflections would be forbidden if the material were disordered. For this case we can readily show that the phase factor $\alpha = \pi$, so the fringes we see are called π fringes. The structure of this interface is shown schematically in Figure 25.6. These π fringes can give symmetric fringes in DF and BF and complementary DF/DF pairs.

SUPERLATTICES

We usually use the term *superlattice* in the sense that it is the *real* lattice, because we have ordering of the components. If the same material was disordered (but with the atoms on the same sites), the lattice parameters, in real space, would be smaller.

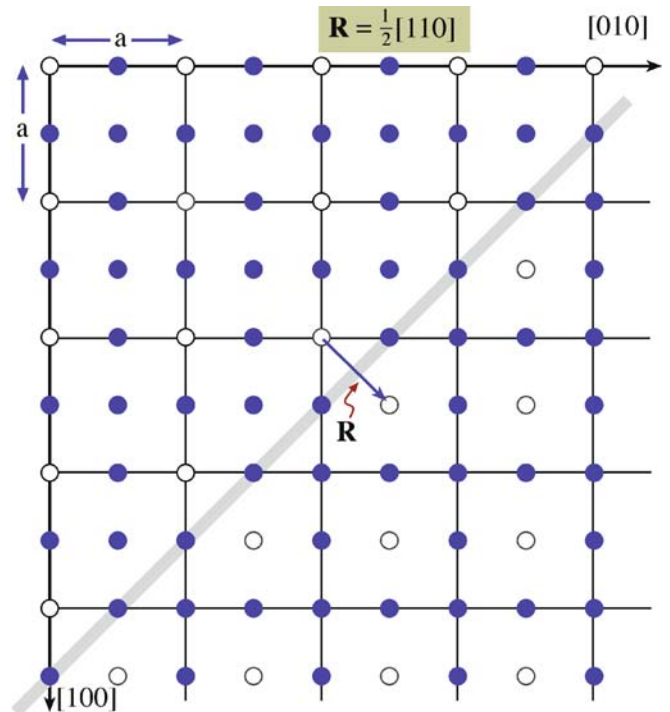


FIGURE 25.6. Schematic of an interface in the intermetallic Ni_3Al showing how the two structures link coherently but are displaced by the vector \mathbf{R} . The phase factor at such an interface is π and the fringes seen in the image are called π fringes.

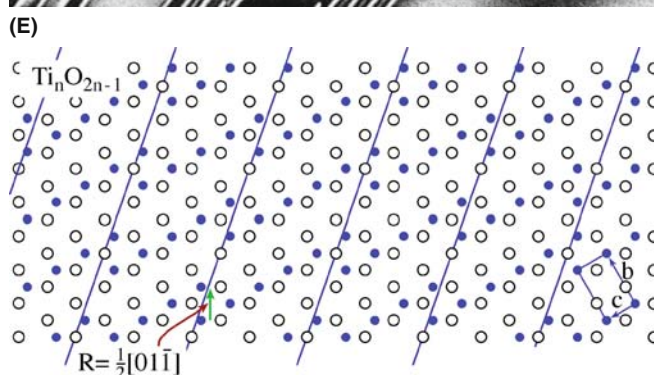
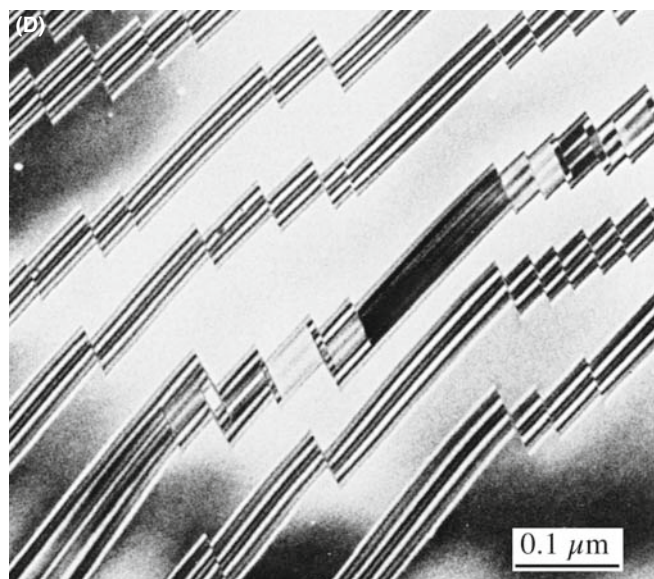
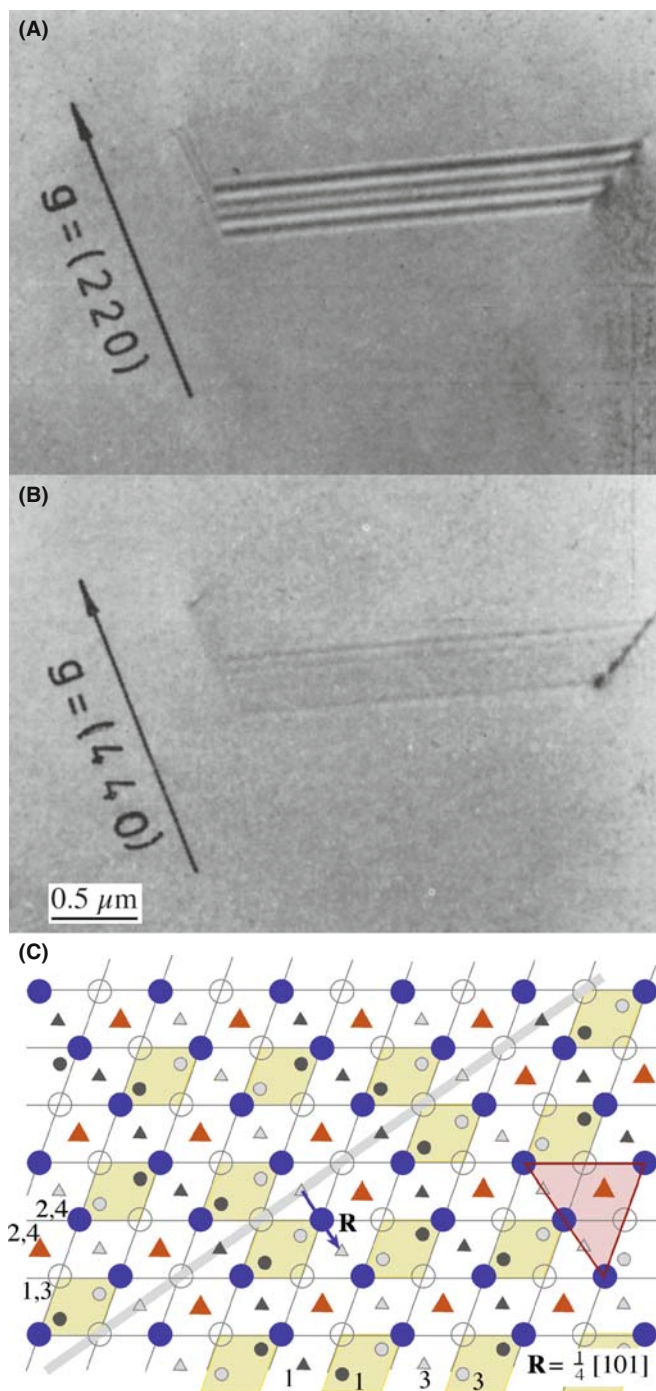


FIGURE 25.7. (Continued).

Similar RBs are very common in oxides because the unit cell is often quite large, giving more opportunities to form such interfaces. The interface shown in Figure 25.7A and B has been called both an SF and an APB in spinel. These interfaces can show all the features we discussed in Section 25.5 for SFs in fcc materials, and those we've just discussed depend on which reflection you use. You can again see a change in contrast in Figure 25.7D when APBs in TiO_2 overlap as shown schematically in Figure 25.7E (Amelinckx and Van Landuyt 1978). In Figure 25.7A and B, if you image the fault using the 220 reflection $2\pi\mathbf{g}\cdot\mathbf{R} = \pi$ and so you'll see SF fringes. If, however, you image using 440, $2\pi\mathbf{g}\cdot\mathbf{R} = 2\pi$, so you'll only see residual contrast (because \mathbf{R} is not exactly $\frac{1}{4}[101]$).

The APB shown in Figure 25.8A is different yet again. This planar defect in GaAs is also known as an IDB (Section 25.1). The fringes you see are caused by a translation, but \mathbf{R} is not related in a simple way to the structure of the crystal or to the inversion symmetry. Rasmussen et al. showed that the translation is present because there is a small relaxation of the Ga–Ga and As–As bonds at this $\{110\}$ interface. The value of \mathbf{R} was determined to be 0.19 \AA with a statistical uncertainty of $\pm 0.03 \text{ \AA}$, so it's a translation that's almost parallel to the

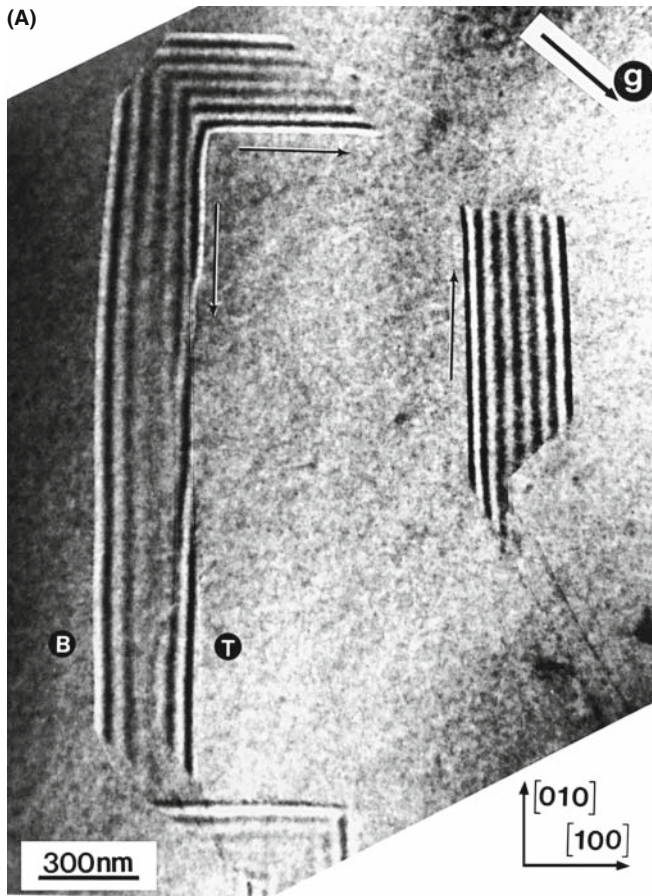


FIGURE 25.8. (A) A faceted APB (or IDB) in GaAs with (B) a schematic of the (110) facet. The translation is caused by the difference in length of the Ga–Ga and As–As bonds and is much smaller than a lattice vector in the GaAs lattice.

interface, as shown schematically in Figure 25.8B. These fringes are then known as δ fringes (because there is only a small translation and α is not related in a simple way to 2π). The use of image-simulation programs, which are necessary to determine \mathbf{R} (remember that the wavelength of the 200-kV electrons used for the measurement is itself 0.025 Å or 2.5 pm), is discussed in Section 25.13.

25.7 PHASE BOUNDARIES

We'll list a few special phase boundaries in Table 25.3.

TABLE 25.3 Examples of Special Phase Boundaries		
Boundary	Example of material	Features
Ferromagnetic domain boundaries	NiO	
Ferroelectric and piezoelectric boundaries	BaTiO ₃	Small tetragonal distortion
Composition boundary	GaAs/AlGaAs	$\xi_{\mathbf{g}}$ is different on two sides of boundary, even for perfect lattice matching
Structure boundaries	α -SiC/ β -SiC hcp-Co/fcc-Co	
Composition/structure	NB/Al ₂ O ₃ Al/Cu α -Fe/Fe ₃ C	

An example of a PB is shown in Figure 25.9. In NiO, which is ferromagnetic, some of the planes rotate slightly when the structure changes from cubic symmetry below the Curie temperature. Now we can also define the cubic structure as rhombohedral with $\alpha = 60^\circ$ in the rhombohedron. Below the Curie temperature, the rhombohedral angle is distorted by only 4.2' (yes, minutes of arc) from the true 60° . Therefore, most \mathbf{g} -vectors will rotate through a very small angle and hence produce a change in the value of s . However, as you can see in Figure 25.9, this small rotation can readily be detected by the change in contrast and the faint fringes at the phase boundary.

We can have overlapping PBs, so the warning is the same: be very wary and use tilting experiments while you are at the TEM.

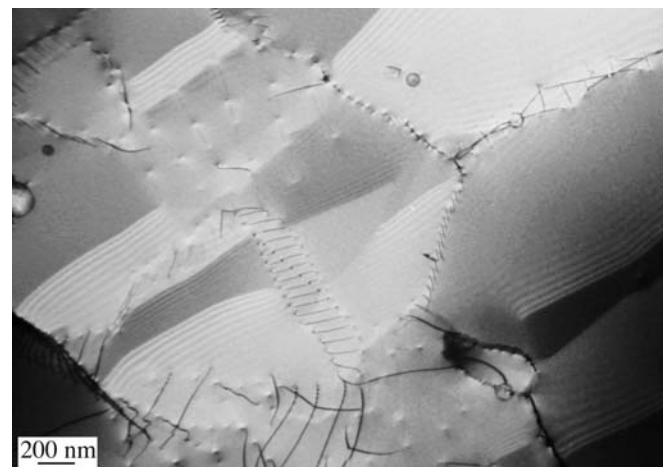


FIGURE 25.9. The ferroelectric material NiO undergoes a structural change from cubic to distorted rhombohedral at the Curie temperature. Although the distortion in the rhombohedral structure is very small, it does cause a detectable rotation of the lattice planes that results in the δ fringes in the image.

25.8 ROTATION BOUNDARIES

What can we learn about rotation boundaries when the rotation angle is greater than about 0.1° ? Unfortunately, the answer is “not a lot,” unless we have defects which accommodate the rotation. Then we are into the subject of diffraction contrast of line defects in interfaces. However, with care you may be able to excite \mathbf{g} in one grain or in both by tilting the

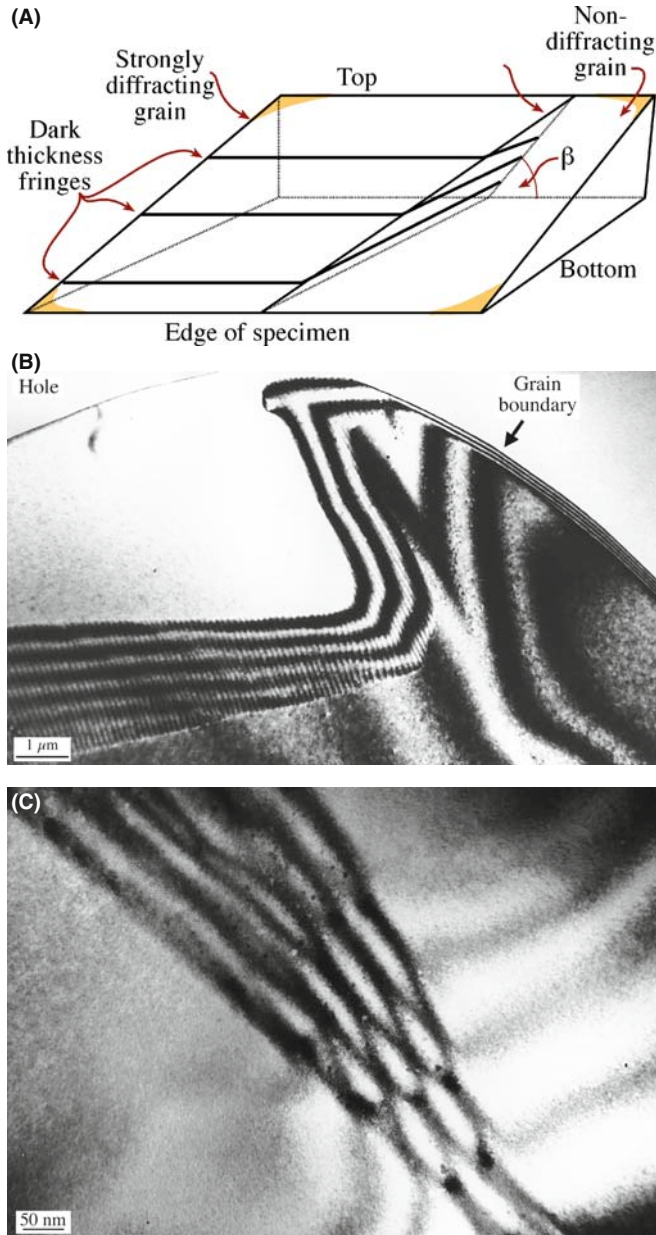


FIGURE 25.10. (A) If the adjoining grains are rotated so that they do not share a common reflection, images can be formed where only one of the grains diffracts. As shown in (B), the thickness fringes associated with the wedge merge into the thickness fringes associated with the inclined interface. (C) If the foil is tilted so that the same (though not coincident) reflection is excited in both grains, the number of fringes in the interface increases with each incremental increase in the wedge thickness.

specimen. The difficulty, of course, is that \mathbf{s}_g is likely to be different in each material. Complications will also arise if other defects are present, since you may or may not see those defects. Examples of such inclined, rotation interfaces are shown schematically in Figure 25.10A and experimentally in Figure 25.10B and C.

25.9 DIFFRACTION PATTERNS AND DISPERSION SURFACES

You read in Chapter 17, that what you see in an image must be related to what happens in the DP which in turn is determined by how the Ewald sphere intersects the reciprocal lattice. Figure 17.5 showed that a planar defect which is inclined to the surface of a parallel-sided specimen will give rise to relrods. Therefore, a planar defect that is inclined to the surface of a parallel-sided specimen will produce at least two spots in the DP. Since most specimens are wedges (see Figure 17.4), and the planar defect will, in general, be inclined to both surfaces, the relrod geometry is actually even more complex. Figure 25.11 shows lines normal to each interface and their associated relrods. You can appreciate that when the Ewald sphere cuts these relrods, several spots may appear in the DP. Now we need to relate these relrods to the fringes we see in the image. This model would predict that we would not produce fringes when $s = 0$, so we should modify what we did in Figure 17.15. The periodicity of the fringes in the image is inversely related to the distances (M_1N and M_2N) between the spots in the DP.

FRINGE SPACING

At $s = 0$, the thickness fringe spacing is determined by $\Delta k = \xi_g^{-1}$ and the wedge angle: the spacing of the fringes is proportional to ξ_g .

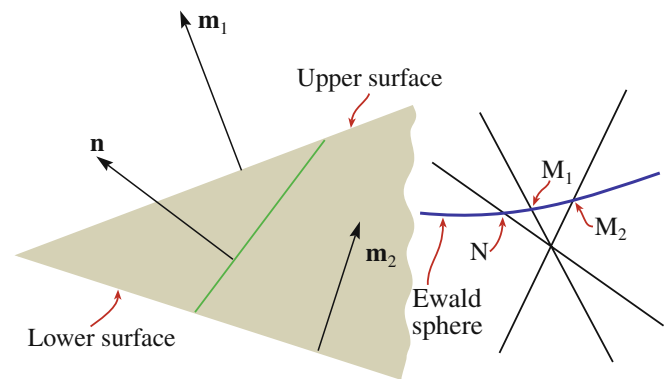


FIGURE 25.11. In a wedge specimen, the planar defect will, in general, be inclined to both surfaces and the relrod geometry is complex. The fringe spacing in the image is related to the reciprocal of the distances M_1N and M_2N .

When a planar defect is present in the specimen, the two branches of the dispersion surface are not only coupled along a tie-line normal to the surface of the specimen but also along the normal to the planar defect. However, when $s = 0$, the thickness periodicity in the image corresponds to the extinction distance. When we relate this to the region G in the reciprocal lattice, the two relrods (which result from a kinematical construction) must actually separate to give the two hyperbolas shown in Figure 25.12 which is why we drew Figures 17.15 and 23.6 as we did.

FRINGES IMPLY SPOTS

If you see fringes in the image, spots will be present in the DP.

The spots in the DP are associated with points M and N in Figures 25.11 and 25.12.

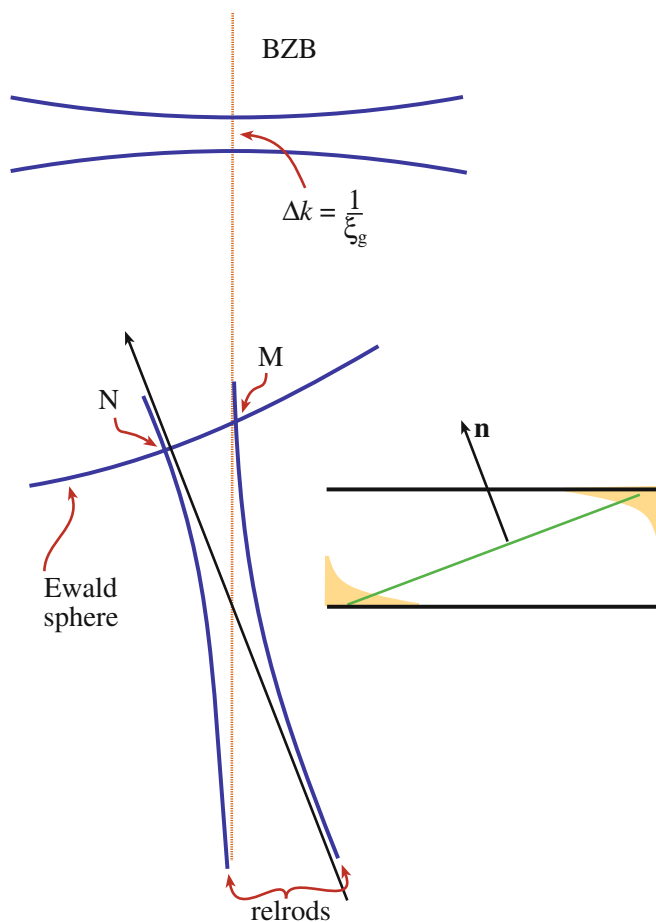


FIGURE 25.12. The dispersion-surface construction for an inclined planar defect in a parallel-sided specimen. (Compare Figures 17.15 and 24.6.) For simplicity, we show the hyperbolas due to the defect alone, not the extra effects that would arise in a wedge specimen.

25.10 BLOCH WAVES AND BF/DF IMAGE PAIRS

In Chapter 14 we saw that, in a crystal, the electron must propagate as Bloch waves, and yet we have not mentioned Bloch waves in our discussion of thickness and bending so far. Most of the analysis of this topic is beyond the scope of this text, but it is important to understand the basic ideas, particularly since they will also apply to scattering from defects in the crystals. Remember that ξ_g is a direct consequence of having two Bloch waves. The important message here is: don't let the words overawe you.

The idea is quite simple. Since we have two Bragg beams excited, then we must have two Bloch waves in the crystal. The propagation vectors of these two waves are \mathbf{k}_1 and \mathbf{k}_2 with the difference $|\Delta\mathbf{k}|$ being related to s_{eff} (and $\gamma_1 - \gamma_2$ in Section 13.10). We see a thickness dependence in the image because the two waves are interfering. The only two waves which are really present in the crystal are the two Bloch waves. It's the beating of these two waves which gives rise to thickness effects.

In the two-beam case, the Bloch waves, 1 and 2, are channeled along and between the atom columns (see Figure 14.2). A fault may change the channeled wave into the non-channeled one, as you can see in Figure 25.13. So, the effect of the planar defect is to couple the

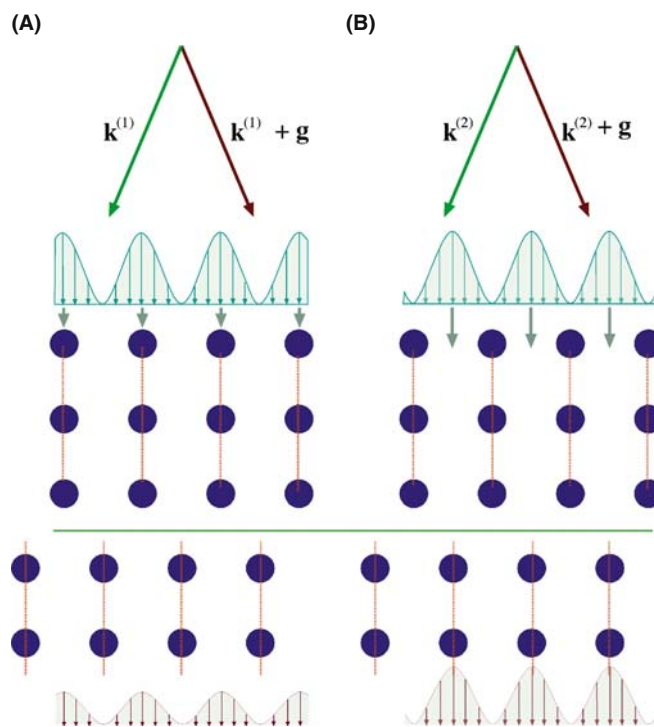


FIGURE 25.13. Bloch waves 1 and 2 are channeled (A) along and (B) between the atom columns, respectively, until they meet the fault. There the atomic columns are translated so that the channeled Bloch wave may become the non-channeled one and vice versa.

Bloch waves; in other words, the defect links (or connects or ties) the different branches of the dispersion surface (along the tie line). The non-complementary contrast at SFs in fcc metals is directly explained by this coupling.

As soon as the beam enters the specimen we excite Bloch waves 1 and 2. Therefore, in Figure 25.14A, the two Bloch waves 1 and 2 are shown everywhere at the top surface of the foil. The planar defect links points D_1 and D_2 on the two branches of the dispersion surface, as shown in Figure 25.14B, along the tie line, D_1D_2' and $D_1'D_2$. We'll analyze the three situations shown in Figure 25.14A, which correspond to the planar defect being close to the top, the middle, and the bottom of the specimen. The key feature is that, as we saw in Section 14.6, Bloch wave 1, which has the larger \mathbf{k} -vector, will be preferentially absorbed. It is actually totally absorbed in thicker specimens.

- When the planar defect is close to the top surface (as occurs near T in Figure 25.14A), waves 1 and 2 are both coupled (or scattered) to the other branch of the dispersion surface so we form four Bloch waves (but with only two \mathbf{k} -vectors). Both Bloch waves which are associated with the upper branch of the dispersion surface (wave vector \mathbf{k}_1) are preferentially

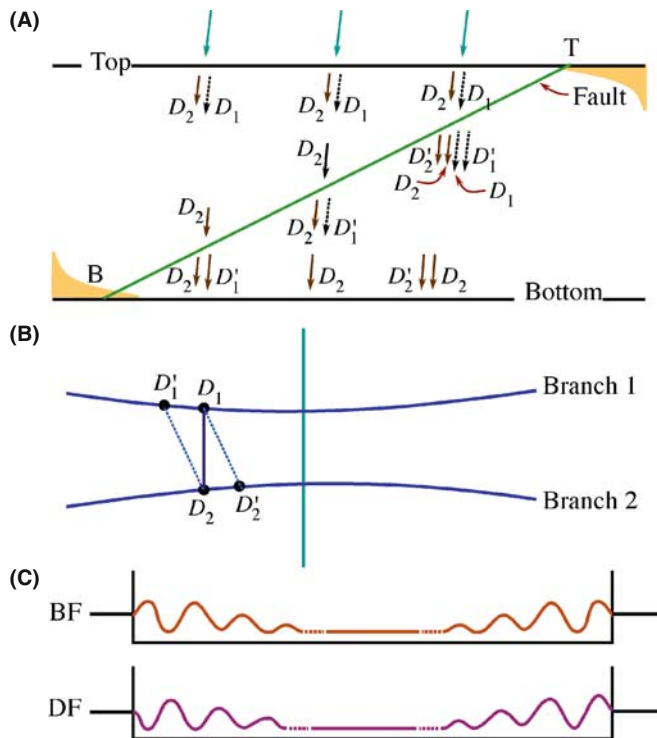


FIGURE 25.14. The absorption of the branch-1 Bloch wave near the top surface, T, of the specimen and its re-creation when the planar defect is near the bottom, B, determines the contrast we see. (A) shows which Bloch waves are present at the different depths in the specimen, (B) shows how the Bloch waves are coupled along tie lines joining the two branches of the dispersion surface, and (C) shows the resulting contrast profiles.

absorbed, but the waves D_2 and D_2' both reach the lower surface. There, they interfere to give the thickness fringes even though they are both associated with the lower branch of the dispersion surface; D_2' retains a 'memory' of D_1 .

- When the fault is close to the middle of the specimen in Figure 25.14A, the branch-1 Bloch wave is absorbed before it reaches the planar defect but a new Bloch wave D_1' is formed at the defect. However, while traversing the other half of the foil, this wave is also absorbed so that only wave D_2 reaches the lower surface. Thus the electrons can propagate through the specimen (we can see through it) but there are no thickness fringes because only one Bloch wave survives. However, we can still image defects in these thicker areas, as you'll see if you look back at Figure 23.10.
- At the lower surface, B in Figure 25.14A, only wave D_1 survives to reach the planar defect but it now produces a new wave D_1' which can reach the lower surface, recombine with Bloch wave D_2 , and produce thickness fringes. The resulting contrast is summarized in Figure 25.14C.

Bloch-wave absorption is a critical factor in explaining the appearance of contrast from planar defects. The part of this argument which is not intuitive is the fact that D_2' retains a memory of D_1 ; this memory allows it to interfere with D_2 to produce the thickness fringes near the top of the specimen, even though no Bloch wave from branch 1 reaches the bottom of the specimen. We'll refer you to the article by Hashimoto et al. for further discussion on this topic.

25.11 COMPUTER MODELING

From the discussion in Sections 25.5 and 25.6, you will realize that α and π fringes are usually understandable as long as you know what the defect is and as long as it's not actually a set of overlapping defects. The contrast from δ fringes is much more complex and combinations of α , π , and δ fringes are difficult! The situation will become even more complicated if you want to understand the contrast occurring when other defects interact with these planar faults.

A computer program is then really the only way to analyze the contrast from these defects. The problem here is that many programs have been written for computers that now only exist in museums, by researchers who became professional programmers and never saw a TEM image again.

The first program to attempt the task of simulating two-dimensional images (rather than line profiles) of planar defects is described in the book by Head et al. Two current software packages are CuFour and

TEMACI. CuFour is the subject of an entire chapter in the companion text; TEMACI incorporated the ideas from Comis that we'll refer to here. We will mention some of the features of these programs to help you select one, but leave the detailed descriptions to the appropriate manuals. The most important reason for using any program must be your desire to understand the contrast and thus characterize the defect.

These programs are tools to assist you toward the goal of quantitative analysis of diffraction contrast, but you always need the fully quantitative experimental data, too. Few TEM users have ever collected such data.

You need an accurate simulation of the image you see in the microscope. The fact that the image varies with depth, thickness, \mathbf{g} -vector, etc., is actually to your advantage, since you then have many variables, all of which you must be able to measure to achieve a good match with your experimental image. From the point of view of quantitative analysis, a one-dimensional line (intensity) profile is as valid as a two-dimensional image. Of course, if you can compare the contrast in an image and a simulation point by point, then you can have much greater confidence in the matching. The two-dimensional simulated image is also more viewer-friendly!

A great advantage of more powerful computers is that you can also test the effect of specimen geometry more readily. Thus, for example, Viguiet et al. showed, using the CuFour simulation package, that the rules for fringe contrast given by Gevers et al. will not work if the specimen is tilted such that it intersects the bottom of the foil above the point where it intersects the top! You can understand this situation more easily by looking at the specimen geometry shown in Figure 25.15.

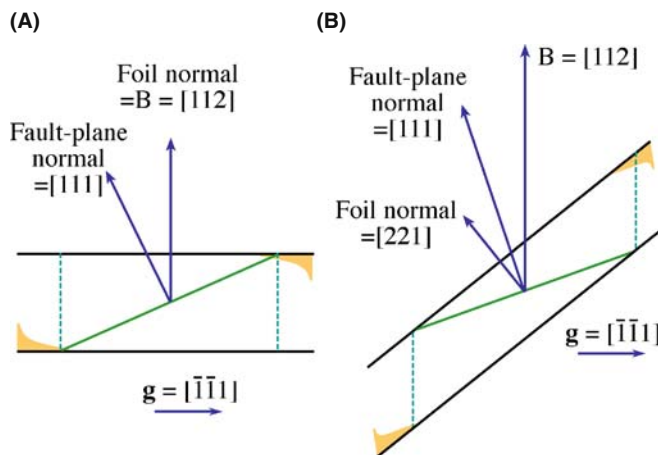


FIGURE 25.15. Many of the 'rules' for predicting the contrast from planar 'defects' make certain assumptions about the geometry of the defect relative to the surface of the specimen that may not always hold. The 'usual' situation is shown in (A) but the intersection of the planar defect with the upper specimen surface may be lower than the intersection with the bottom surface as in (B). This geometry can cause a reversal in the rules.

Image simulation tells us that $\mathbf{g} \cdot \mathbf{R}$ must be > 0.02 to produce visible fringes, and you don't need to know the local structure at the planar defect when determining this condition. You could, in principle, detect smaller values of \mathbf{R} by using larger \mathbf{g} -vectors but in practice it then becomes more difficult to set up a well-defined diffraction geometry.

The next two sections are rather specialized and you may wish to leave them until much later, especially if you don't have access to a suitable program or until you are prepared to write your own. Do consult the key references and list of available programs in Section 1.6 before writing your own program. The subject is just as relevant to the topics of Chapters 26 and 27, but we include it in this chapter mainly because the analysis of planar defects is the most straightforward application.

25.12 THE GENERALIZED CROSS SECTION

Head et al. presented a method and a computer program for the computation of BF and DF images of line and planar defects. The source code is given in their book and is available from the WWW. You should note several important features of this program

- It uses the two-beam theory of electron diffraction.
- It uses the column approximation.
- The simulated image can be displayed as a halftone image rather than as intensity profiles.

This program was so successful, in part, because Head et al. were able to calculate the images quite quickly in spite of the fact that the computers available to microscopists were often not particularly powerful in ~1970. The calculations used a concept which they called the generalized cross section (GCS). (The GCS is not a scattering cross section, it is actually a slice through the specimen.) The GCS can be used when the displacement field, $u_{\mathbf{k}}$, satisfies the requirement that

$$u_{\mathbf{k}}(x, y, z) = u_{\mathbf{k}}(x, 0, z + cy) \quad (25.34)$$

Here c is a constant and the foil is imagined to be laterally infinite. When this requirement is satisfied, the calculation of $u_{\mathbf{k}}$ is greatly simplified. One such situation is the important case where several dislocations and their associated fault planes are all parallel to one another. Then you only need to calculate the many-beam, Howie-Whelan equations on the plane $y = 0$. The displacement field for two columns y_1 and y_2 will only differ by a translation along the column, i.e., the z -direction. You don't want to repeat calculations you've already done; just calculate the image on a mesh in the x - y plane.

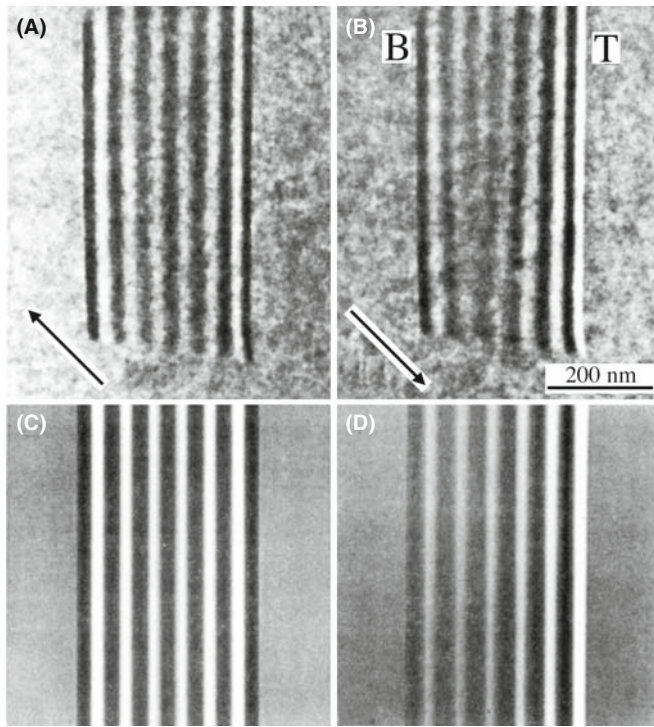


FIGURE 25.16. (A) Experimental BF image of an APB with $g = 220$. (B) DF image of the same defect, $g = \bar{2}20$. (C, D) Corresponding simulated images.

Examples of experimental and simulated images are shown in Figure 25.16. The simulation package used here (Comis again) was particularly attractive since it also performed elasticity calculations for simple defect configurations. It could simulate the effect of changing different parameters, and hence

- Change the accelerating voltage to see the extinction distances change.
- Change the absorption parameters to see the loss of SF-fringe contrast near the middle of the foil ($z = 0.5t$).
- Change the number of beams contributing to the image; how good is your two-beam assumption?
- Look at how reversing g changes the geometry of the image.
- Compare BF and DF as you vary the value of s_g .

The Comis program ran on a Convex mini-super-computer with an intelligent terminal, then on a Mac with OS6. Zhou used the concepts in the new and advanced package, TEMACI.

25.13 QUANTITATIVE IMAGING

One of the important applications of diffraction-contrast images is the detailed characterization of defects. With the improvements in the TEM, particularly in resolution

and drift, we are now able to pay more attention to the fine structure of defects and this requires *quantitative* image analysis. In particular, we need to use the actual intensity levels in the image. One obstacle for quantitative analysis has been the uncertainty in the background-level intensity caused by inelastic scattering. As energy-filtered images become more widely available (see Chapter 37), we hope this problem will disappear. Direct digital recordings of the intensities, using a CCD camera are making quantitative analysis more tractable, eliminating uncertainties associated with the calibration of the response of the emulsion of the photographic film.

With these new applications of diffraction-contrast images in mind, improved simulation programs have become essential. An ideal program will be versatile, but user-friendly; it will allow us not only to calculate the image but also to give defect interaction geometries. In simulating images of crystal defects, you'll encounter several problems that are almost independent of one another. You must be able to do the following

- Define the geometry of the defects and of the specimen (the diffracting conditions).
- Calculate the displacement field associated with the defects.
- Propagate and scatter the electron beams throughout the foil (i.e., solve the Howie-Whelan equations).

We've already discussed the theoretical basis of the diffraction process, so we'll now illustrate some of the numerical methods which we can employ for different defect geometries. In Chapter 26, we'll consider other types of defects which may be analyzed, methods for defining them, and how we can calculate the displacement field.

25.13.A Theoretical Basis and Parameters

We'll use the Comis program as an example of the considerations that go into a simulation program. One message that you should certainly understand from this discussion is that you must be very cautious when using any program to simulate images. All such programs make assumptions and simplifications.

BLACK BOX

As always, when using the computer to simulate TEM images: beware of the black box. Don't automatically believe everything that comes out of it.

Comis was based on the Howie-Whelan dynamical theory of electron diffraction and thus neglects diffuse scattering but it could simulate without the column approximation. The basis of the approach used was given by Howie and Basinski.

- Use the deformable-ion approximation to describe how the crystal is influenced by the displacement field, \mathbf{R} . In this model, the potential at \mathbf{r} in the deformed crystal is assumed to be equal to the potential at the point $\mathbf{r} - \mathbf{R}(\mathbf{r})$ in the perfect crystal. The model is good unless $\mathbf{R}(\mathbf{r})$ varies too rapidly.
- Extend the Howie-Whelan approach to many beams and avoid the column approximation.

The resulting equations are basically the same as those we derived in Chapter 13, so don't be put off by their appearance.

In treatments avoiding the column approximation, we include terms which allow for a variation in x and y : these terms were specifically excluded in Chapter 13.

The equations are now written as

$$\frac{\partial \phi_{\mathbf{g}}(r)}{\partial z} = i\pi \sum_{\mathbf{h}} \left(\frac{1}{\xi_{\mathbf{g}-\mathbf{h}}} + \frac{i}{\xi'_{\mathbf{g}-\mathbf{h}}} \right) \phi_{\mathbf{h}} e^{2\pi i((s_{\mathbf{h}}-s_{\mathbf{g}})z - (\mathbf{h}-\mathbf{g}) \cdot \mathbf{R})} - \theta_x \frac{\partial \phi_{\mathbf{g}}}{\partial x} - \theta_y \frac{\partial \phi_{\mathbf{g}}}{\partial y} + \frac{i}{4\pi\chi_z} \left(\frac{\partial^2 \phi_{\mathbf{g}}}{\partial x^2} + \frac{\partial^2 \phi_{\mathbf{g}}}{\partial y^2} \right) \quad (25.35)$$

As usual, χ is the incident-beam wave vector in vacuum, \mathbf{g} is a particular diffraction vector, and \mathbf{h} represents all the other possible diffraction vectors; you might like to compare this equation with equation 13.8! We have defined two new parameters to take account of the direction of the beam

$$\theta_x = \frac{(\chi + \mathbf{g})_x}{\chi_z} \quad \text{and} \quad \theta_y = \frac{(\chi + \mathbf{g})_y}{\chi_z} \quad (25.36)$$

The x - y plane in the reciprocal lattice contains the dominant reflections, and z is almost parallel to the incident beam. The number of beams you can include in the calculation is limited only by the capacity of your computer. The standard default for a program would be to select only beams on the systematic row. However, non-systematic beams can substantially influence your image, so it is useful if you can include sets of beams which are coplanar with the systematic row. We define the deviation of the crystal orientation from the exact Bragg condition by specifying the wave-vector components χ_x and χ_y ; the latter applies when reflections outside the systematic row are included. Now we can calculate *all* the deviation parameters, $s_{\mathbf{h}}$; there are many beams and each \mathbf{s} can be different.

Each extinction distance $\xi_{\mathbf{g}}$ is defined as the ratio $\chi_{\mathbf{g}}/|U_{\mathbf{g}}|$, as usual with $U_{\mathbf{g}}$ being a Fourier component of the perfect-crystal potential. The Fourier components can be calculated from X-ray scattering factors, using the Mott expression. For most situations the scattering angle is small enough, so the X-ray scattering factors

may, in turn, be calculated using the nine-parameter Gaussian fit given by Doyle and Turner.

You'll need to know the unit cell for your material and the Debye-Waller factor, B . Comis could then automatically calculate $\xi_{\mathbf{g}}$ using a built-in table of the Doyle-Turner parameters. You need to know B if you want to convert X-ray structure factors to electron structure factors (or vice versa) at a given temperature, or if you want to compare structure-factor measurements taken at different temperatures. When you calculate extinction distances, the Debye-Waller factors are essential to determine the effect of temperature.

DEBYE-WALLER FACTOR

B is related to the (mean-square) vibrational amplitude of an atom on a lattice site. It is a temperature-sensitive term.

Equation 25.35 is only valid if your crystal has a center of symmetry, otherwise we have to redefine $\xi_{\mathbf{g}}$. Simulations involving non-centrosymmetric crystals can be performed, but we have to replace the $\xi_{\mathbf{g}}$ and $\xi'_{\mathbf{g}}$ with complex quantities and then defining all the parameters is really difficult (see equations 14.2–14.8).

You can take account of absorption effects in the usual way by adding imaginary Fourier components, $U'_{\mathbf{g}}$ the absorption distances, $\xi'_{\mathbf{g}}$ are then defined as we discussed in Section 24.7. There are equations which a program can use to estimate the absorption distances, e.g., the linear relation $|U'_{\mathbf{g}}|/|U_{\mathbf{g}}| = a + b|\mathbf{g}|$, as suggested by Humphreys and Hirsch, or you could specify each individual absorption distance directly.

25.13.B Apparent Extinction Distance

The program developed by Head et al. was based on the two-beam approximation. The success of such calculations relies on the fact that the $\xi_{\mathbf{g}}$ may be replaced by an apparent extinction distance, $\xi_{\mathbf{g}}^a < \xi_{\mathbf{g}}$. This substitution compensates for scattering into beams that are not included in the two-beam calculation. The term $\xi_{\mathbf{g}}^a$ depends on t and must be estimated in each individual situation, e.g., by fitting the simulated image to your experimental image. For a quantitative image analysis it is important that you should have as few adjustable parameters as possible; using the many-beam program eliminates the need to use the parameter $\xi_{\mathbf{g}}^a$. Alternatively, you may determine $\xi_{\mathbf{g}}^a$ by comparing simulated thickness fringes calculated using many-beam and two-beam approximations.

25.13.C Avoiding the Column Approximation

Now you know we can perform simulations with or without the column approximation. With the column

approximation, you only keep the first term on the right-hand side of equation 25.35. The equations are reduced to a system of ordinary differential equations which the program must solve at each image point (x, y) . In practice, the equation is solved on the nodes of a mesh (columns) using a fifth-order Runge-Kutta integration routine (which you, or the program, can look up when you need it). You need to choose the size and ‘resolution’ of the mesh. As we’ll see in Chapter 27, there are situations where the column approximation will not be acceptable.

Without the column approximation, equation 25.35 gives us a system of coupled partial differential equations. The boundary conditions (at $z = 0$) can be generally written in the form

$$a_{\mathbf{g}}\phi_{\mathbf{g}} + b_{\mathbf{g}}\frac{\partial\phi_{\mathbf{g}}}{\partial x} = c_{\mathbf{g}} \quad (25.37)$$

where we’re ignoring changes in the y -dimension.

We can use fixed boundary conditions again following Howie and Basinski

- The foil is divided into thin slices of thickness Δz . You should not confuse this with the multislice method for lattice-image simulation which we’ll see in Chapter 30; we are still using Howie-Wheelan equations. Then, equation 25.35 is integrated, using the column approximation, through the first slice, i.e., from $z = 0$ to $z = \Delta z$, at all the mesh-points.
- The corrections to the column approximation, i.e., the terms containing derivatives with respect to x and y , are then evaluated by interpolation and included.
- The procedure is repeated until the exit surface of the foil is reached.

With this procedure, you are actually applying the column approximation to the outer boundary of the mesh. So, in equation 25.37 $a_{\mathbf{g}} = 1$, $b_{\mathbf{g}} = 0$, and $c_{\mathbf{g}} = \phi_{\mathbf{g}}$ at the initial surface. In order to avoid distortion of the image, we must choose the step size, Δz , carefully and be sure that the distance between columns (mesh size) is small enough (see Anstis and Cockayne).

25.13.D The User Interface

You’ll want to run your program interactively so it should include commands which allow you to change parameters easily. Ideally, it will allow you to access each command through the keyboard using a menu. In Comis certain standard menus were available for special purposes. The user could also build (and save) menus interactively. This allowed all the relevant parameters and commands for a particular problem to be present within a single menu. At any time, all the commands were available through the keyboard.

Although typical simulations may be performed in a matter of seconds, many-beam calculations including several dislocations may require more CPU time. For this situation, Comis included a ‘submit’ command that would start a batch job based on your current data and parameters. Thus, the interactive mode could be used as a convenient way of submitting several jobs with varying parameter values.

For many problems, a purely visual comparison of experimental and simulated images is sufficient to allow you to interpret your image. In these situations, you can often find a $\xi_{\mathbf{g}}^a$ such that the simulations can be carried out with only two beams (à la Head et al.). However, since many parameters are involved in the image-matching process, it is best to eliminate as many unknown variables as possible. Many-beam calculations are even more important for quantitative analyses.

CHAPTER SUMMARY

The key points discussed in this chapter are

- We see contrast from planar defects because the translation, \mathbf{R} , causes a phase shift $\alpha = 2\pi\mathbf{g}\cdot\mathbf{R}$.
- In the two-beam case, we can derive analytical expressions to describe the contrast.
- We can use the scattering matrix method in the two-beam case and can readily extend it to more complicated multibeam situations.

Many different types of planar defect can be studied. You should be careful not to assume that all defects behave the same as SFs in fcc materials.

There is a direct relationship between the information in the images and that in the DPs, which you can understand using the concept of the relrod.

You need to understand how Bloch waves behave to explain why BF/DF pairs of images are not complementary and why the contrast from planar defects can disappear in the ‘middle’ of the image. The latter is a result of preferential absorption (really scattering) of certain Bloch waves.

We can now use computer modeling of diffraction-contrast images of planar defects to perform quantitative analysis and image matching.

REFERENCES

We suggest a few books or chapters of books for background reading. You'll realize that careers have been built on this topic. As usual, we also recommend that you read some of the original papers.

INTERFACES

- Christian, JW 1975 *The Theory of Transformations in Metals and Alloys*, Part 1, 2nd edition, Pergamon Press New York.
- Carter, CB and Norton, MG 2007 *Ceramic Materials* Springer New York. Chapters illustrating interfaces in ceramics.
- Forwood, CT and Clarebrough, LM 1991 *Electron Microscopy of Interfaces in Metals and Alloys*, Adam Hilger New York. Invaluable for anyone studying interfaces by TEM.
- Howe, JM 1997 *Interfaces in Materials: Atomic Structure, Thermodynamics and Kinetics of Solid-Vapor, Solid-Liquid and Solid-Solid Interfaces* Wiley New York.
- Matthews, John Wauchope 1975, *Epitaxial Growth* Academic Press/Elsevier New York. Unfortunate use of 'epitaxial' instead of epitactic by a pioneer in the subject. (Yes, the father of Dave Matthews.)
- Sutton, AP and Balluffi, RW 1995 *Interfaces in Crystalline Materials*, Oxford University Press New York.
- Wolf, D and Yip, S, Eds. 1992 *Materials Interfaces, Atomic-level Structure and Properties*. Chapman and Hall New York.

THINKING ABOUT CONTRAST

The book by Head et al. (see Section 1.5) is the starting text for simulating diffraction-contrast images.

- Amelinckx, S and Van Landuyt, J 1978 in *Diffraction and Imaging Techniques in Materials Science*, **1** and **2**, (Eds., S Amelinckx, R Gevers and J Van Landuyt), 2nd Ed., p. 107, North-Holland New York.
- Anstis GR and Cockayne DJH 1979 *The Calculation and Interpretation of High-Resolution Electron Microscope Images of Lattice Defects* Acta Cryst. **A35**, 511–524.
- Edington, JW 1976 *Practical Electron Microscopy in Materials Science*, Van Nostrand Reinhold New York. A classic practical guide but sometimes not suitable for modern TEMs.
- Gevers R, Art, A and Amelinckx S 1963 *Electron Microscopic Images of Single and Intersecting Stacking Faults in Thick Foils – 1. Single faults* Phys. Stat. Sol. **3**, 1563–93.
- Gevers R, Blank H and Amelinckx S 1966 *Extension of the Howie-Whelan Equations for Electron Diffraction to Non-Centro Symmetrical Crystals* Phys. Stat. Sol. **13**, 449–465. See this paper for discussion of imaginary extinction distances, ξ'_g .

THE COLUMN APPROXIMATION

- Howie A and Basinski ZS 1968 *Approximations of the Dynamical Theory of Diffraction Contrast* Phil. Mag. **17**, 1039–63. The non-column approximation paper.
- Howie A and Sworn H 1970 *Column Approximation Effects in High Resolution Electron Microscopy Using Weak Diffracted Beams* Phil. Mag. **22**, 861–4. Example of a situation where the column approximation gives the wrong answer.

EXTINCTION DISTANCES

- Doyle PA and Turner PS 1968 *Relativistic Hartree-Fock X-ray and Electron Scattering Factors* Acta Cryst. **A24**, 390–7.
- Hashimoto, H, Howie, A and Whelan, MJ 1962 *Anomalous Electron Absorption Effects in Metal Foils: Theory and Comparison with Experiment* Proc. Roy. Soc. London A **269**, 80–103. Bloch waves and planar defects.
- Hirsch PB, Howie A, Nicholson RB, Pashley DW and Whelan MJ 1977 *Electron Microscopy of Thin Crystals*, 2nd edition, p. 225, Krieger, Huntington, New York. Deduces equation 25.27.
- Humphreys CJ and Hirsch PB 1968 *Absorption Parameters in Electron Diffraction Theory* Phil. Mag. **18**, 115–122.
- Mott NF and Massey HSW 1965 *The Theory of Atomic Collisions*, 3rd Ed., Clarendon Press Oxford. Source of the Mott expression for X-ray scattering factors.
- Taftø, J and Spence, JCH 1982 *A Simple Method for the Determination of Structure-Factor Phase Relationships and Crystal Polarity Using Electron Diffraction* J. Appl. Cryst. **15**, 60–4. Shows how to distinguish g and \bar{g} using CBED.
- Yoshioka, H 1957 *Effect of Inelastic Waves on Electron Diffraction* J. Phys. Soc. Japan **12**, 618–628. See this paper for discussion of imaginary Fourier components, U'_g .

SIMULATION

The original citation is the text by Head et al. (Section 1.5). It is still useful reading but needs a trip to the library.

- Rasmussen, DR and Carter, CB 1991 *A Computer Program for Many-Beam Image Simulation of Amplitude-Contrast Images* J. Electron Microsc. Techniques **18**, 429–36. An idea of what we need in a user-friendly (Comis) program.
- Rasmussen, DR, McKernan, S and Carter, CB 1991 *Rigid-Body Translation and Bonding Across {110} Antiphase Boundaries in GaAs* Phys. Rev. Lett. **66**, 2629–32.
- Schaublin, R and Stadelmann, P 1993 *Method for Simulating Electron Microscope Dislocation Images* Mater. Sci. Engng. **A164**, 373–8. The paper describing CuFour.
- Thölen AR 1970 *A Rapid Method for Obtaining Electron Microscope Contrast Maps of Various Lattice Defects* Phil. Mag. **22**, 175–182.
- Thölen AR 1970 *On the Ambiguity Between Moiré Fringes and the Electron Diffraction Contrast from Closely Spaced Dislocations* Phys. Stat. Sol. (A) **2**, 537–550.
- Viguier, B, Hemker, KJ and Vanderschaeve, G 1994 *Factors Affecting Stacking Fault Contrast in Transmission Electron Microscopy: Comparisons with Image Simulations* Phil. Mag. **A69**, 19–32.
- Zhou, Z 2005 *Electron Microscopy and Elastic Diffuse Scattering of Nanostructures* D.Phil. Thesis, Oxford University. Zhongfu developed TEMACI.

THE COMPANION TEXT

Chapters on CuFour and more on TEMACI.

SELF-ASSESSMENT QUESTIONS

- Q25.1 Describe the distinguishing features of the following interfaces: (a) phase boundary; (b) anti-phase boundary; (c) stacking fault.
- Q25.2 What is α and for what values of α will planar defects not be seen?
- Q25.3 Deduce the value of α for an fcc stacking fault.
- Q25.4 A series of overlapping stacking faults show no contrast in the BF image. Explain how this could occur.
- Q25.5 Which SF fringe is the same in both BF and DF images?
- Q25.6 What interface gives complementary BF/DF fringes?
- Q25.7 You can calculate the extinction distance. Can you confirm this value experimentally from an image?
- Q25.8 What is the difference between a low-angle GB and a high-angle GB?
- Q25.9 What does the concept of the scattering matrix allow to describe?
- Q25.10 Why are BF/DF SF images usually not complementary?
- Q25.11 Is the column approximation relevant to the study of planar defects?
- Q25.12 Why might we see no contrast from the SF in the middle of the specimen even when we see fringes from the part close to each surface?
- Q25.13 How does the planar defect affect the Bloch wave when it is close to the lower surface?
- Q25.14 We describe SF contrast in terms of the inclination of the planar defect in the foil. When does this description clearly fail?
- Q25.15 What does the term ‘deformable-ion approximation’ mean?
- Q25.16 In deducing the propagation of Bloch waves in the presence of a SF, we use the tie line concept. What is the physical reason for using this concept in this situation?
- Q25.17 Why do we recommend that you use the displaced-aperture DF approach for comparing BF and DF images?
- Q25.18 Why is the displaced-aperture DF approach more useful today than in 1960?

TEXT-SPECIFIC QUESTIONS

- T25.1 Consider Figure 25.5. At the far left of the lower image, the fringes are shifted. Explain this shifting using words and equations.
- T25.2 Consider Figure 25.7 and if possible, the original paper. Confirm that the translation vector given in the text is correct. (Draw a larger diagram and deduce the direction and magnitude of \mathbf{R} .)
- T25.3 Consider Figure 25.8. Is this DF or BF and what is g ? Explain your answer.
- T25.4 In Figure 25.9, the contrast changes abruptly between regions that we might think of as ‘grains.’ Is this change real and why is it occurring?
- T25.5 By considering Figure 25.10A and B, explain how the image will differ if the angle β is 30° or 150° .
- T25.6 Is the shape of the sample in Figure 25.10c a perfect wedge? Discuss the reasons for your answer and the possible implications.
- T25.7 Suggest a geometry for the anti-phase domain in Figure 25.8 using information from the fringes. Be as complete as possible.
- T25.8 Explain why π fringes in Ni_3Al are so named.

- T25.9 Extra challenge. By referring to the literature or otherwise, discuss the origin of the δ fringes in Figure 25.9. Are these fringes SF-type fringes? Is it coincidental that they end on arrays of dislocations? How would the fringes in CoO differ?
- T25.10 Extra challenge. Explain the contrast in Figure 25.4 and relate them to one of the geometries. One difficulty you will encounter is identifying the first W fringe. Discuss why this is so.
- T25.11 Sketch the possible SFs in Si for which the \mathbf{g} vector is 131 and the beam direction is [2, 3, 11]. The planes on which the faults lie have been determined and are (111), (111), (111), and (111) for faults A, B, C, and D, respectively. How would you determine the nature of the SFs. Comment on why any intensity is observed for the faults on the $(\bar{1}\bar{1}\bar{1})$ and $(\bar{1}\bar{1}\bar{1})$ planes. (Adapted from Ian Robertson.)
- T25.12 Consider a DP from two thin crystals and the corresponding moiré pattern. Both crystals are fcc structures. If you know that the magnification of the moiré pattern is 10^6 times and that the moiré pattern and the DP are correctly oriented with respect to each other, show how you would calculate the corresponding lattice parameters (a_1 and a_2) of the two fcc crystals. (Courtesy Matt Halvarson.)

Imaging Strain Fields

CHAPTER PREVIEW

As we discussed in Chapter 24, bending of the lattice planes causes a change in the diffraction conditions and therefore a change in the contrast of the image. The presence of a lattice defect in the specimen causes the planes to bend close to the defect. The special feature here is that the bending varies not just laterally, but also through the specimen. Since the details of the bending generally depend on the characteristics of the defect, we can learn about the defect by studying the contrast in the TEM image. This simple principle has led to one of the main applications of TEM, namely, the study of defects in crystalline materials. We can claim that our understanding of the whole field of dislocations and interfaces, for example, has advanced because of TEM. We have even discovered new defects using TEM—like the stacking-fault tetrahedron, the faulted dipole, and the multipole.

Usually we want to learn two things about these defects: we want to know where they are and then understand what they are. So the idea underlying this chapter is the same as for bend contours: we use different reflections corresponding to different sets of lattice planes. We see how the defects affect the image contrast from those different lattice planes and thus characterize the defects. In case you are worried, we would like to emphasize that this is *not* a chapter about defects; it is concerned with understanding contrast in the TEM. We will introduce the necessary terminology and notation concerning defects, but we won't try to give you a comprehensive discussion of them. You should consult the standard references on dislocations at the end of the chapter if you need more details. However, we will show lots of pictures because now we are concerned with the appearance of images. We finish with some discussion of image simulation.

26.1 WHY IMAGE STRAIN FIELDS?

First, we should review our terminology. When we displace the atom at position \mathbf{r} a distance $\mathbf{R}(\mathbf{r})$ from its site in the perfect crystal, we say the crystal is under a strain (ϵ). If the crystal is strained, then it must be subject to a stress which we'll call σ . (Metallurgists traditionally use these symbols and although σ means 'cross section' to a microscopist, we'll stick with it.) Since $\mathbf{R}(\mathbf{r})$ varies with position in the crystal, ϵ and σ will in general also vary with \mathbf{r} . We will assume that ϵ and σ can each be defined at a point. Then we will refer to these quantities as the displacement field, $\mathbf{R}(\mathbf{r})$, the strain field, $\epsilon(\mathbf{r})$, and the stress field, $\sigma(\mathbf{r})$. You will notice that these terms are used interchangeably in the literature although there is a clear causal sequence. What we image is the effect of the $\mathbf{R}(\mathbf{r})$.

To have an intuitive feel for why we see contrast from dislocations, consider the geometry shown in Figure 26.1. The diffraction geometry has been set up so that

the specimen is slightly tilted away from the Bragg condition. The distortion due to the dislocation will then bend the near-diffracting planes back into the Bragg-diffracting condition. We have relrods so there will still be some intensity in the electron beam even when we are not at the exact Bragg condition. The figure shows planes bending at a dislocation; compare this to Figure 24.7 showing bend contours. Regions far from the dislocation are tilted well away from the Bragg condition, while the regions on either side of the dislocation core are at the Bragg condition for $\pm\mathbf{g}_{hkl}$. It is more difficult to recognize the diffracting planes for a screw dislocation (Figure 26.2) but the planes are bending just the same.

When studying a particular dislocation (edge or screw), we want to determine the following parameters

- The direction and magnitude of the Burgers vector, \mathbf{b} , which is normal to the hkl diffracting planes (Figures 26.1 and 26.2B).

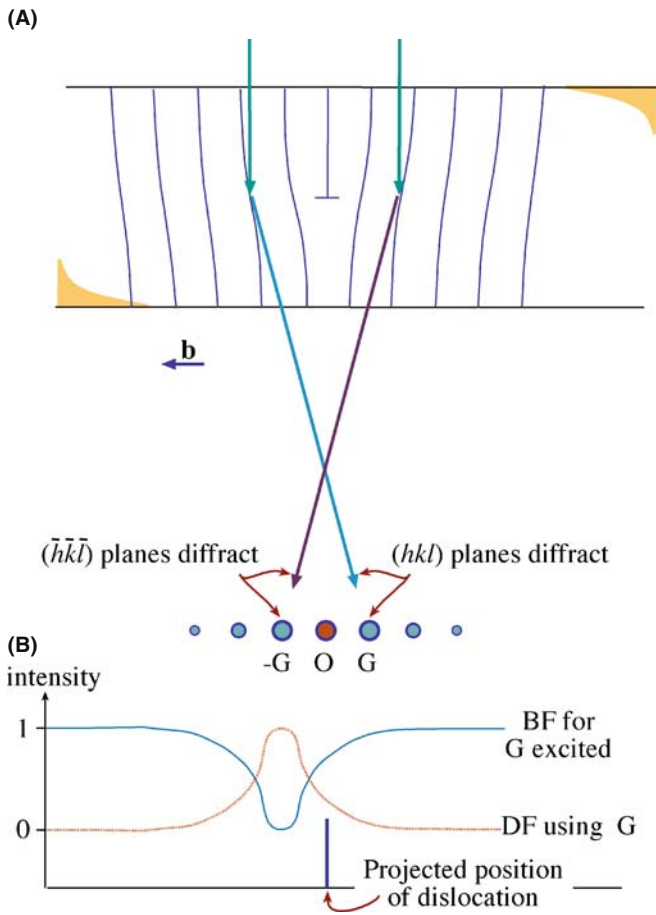


FIGURE 26.1. (A) The specimen is tilted slightly away from the Bragg condition ($s \neq 0$). The distorted planes close to the edge dislocation are bent back into the Bragg-diffracting condition ($s = 0$), diffracting into G and $-G$ as shown. (B) Schematic profiles across the dislocation image showing that the defect contrast is displaced from the projected position of the defect. (As usual for an edge dislocation, \mathbf{u} points into the paper.)

- The line direction, \mathbf{u} (a vector), and therefore, the character of the dislocations (edge, screw, or mixed).
- The glide plane: the plane that contains both \mathbf{b} and \mathbf{u} .

There are other questions we want to answer

- Is the dislocation interacting with other dislocations, or with other lattice defects?
- Is the dislocation jogged, kinked, or straight?
- What is the density of dislocations in that region of the specimen (and what was it before we prepared the specimen)?
- Has the dislocation adopted some special configuration, such as a helix?

In many of these questions, you may find that stereomicroscopy or tomography (Section 29.1) can be very helpful although we will not emphasize those techniques here. The basic requirement if you do use stereomicroscopy is that you must form all of your images using the same \mathbf{g} vector.

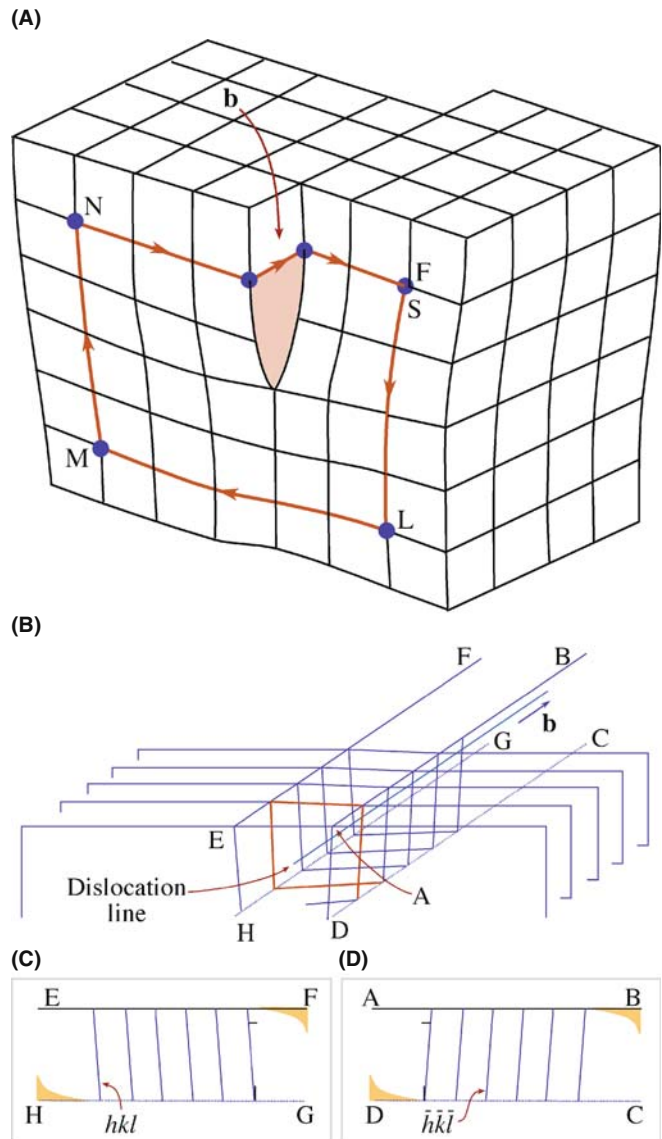


FIGURE 26.2. (A) Distortion of planes around a screw dislocation. The circuit SLMNF is used to define the Burgers vector, \mathbf{b} (see Figure 26.5). (B) Schematic showing the rotation of the diffracting planes by a screw dislocation. The planes are rotated in opposite directions on either side of the dislocation. (C) and (D) show sections (ABCD and EFGH) through these diffracting planes.

26.2 HOWIE-WHELAN EQUATIONS

Let's start with the two-beam phenomenological approach because it worked so well in Chapter 25. An important assumption is that we have linear elasticity. What this means is that if we have \mathbf{R}_1 due to one defect and \mathbf{R}_2 due to a second defect, then at any point in the specimen we can just add these two values to determine the total displacement field, \mathbf{R} . We will not consider anisotropic elasticity although this can readily be included in calculations if you're using a computer.

In Chapter 25, we showed that we could modify the Howie-Whelan equations to include a lattice distortion \mathbf{R} . So for the imperfect crystal

$$\frac{d\phi_{\mathbf{g}}}{dz} = \frac{\pi i}{\xi_0} \phi_{\mathbf{g}} + \frac{\pi i}{\xi_{\mathbf{g}}} \phi_0 \exp[-2\pi i(sz + \mathbf{g} \cdot \mathbf{R})] \quad (26.1)$$

Now we make a different substitution of variables (compare with equations 25.5 and 25.6). Set

$$\phi_0(z)_{(\text{sub})} = \phi_0(z) \exp\left(\frac{-\pi iz}{\xi_0}\right) \quad (26.2)$$

and

$$\phi_{\mathbf{g}}(z)_{(\text{sub})} = \phi_{\mathbf{g}} \exp\left(2\pi isz - \frac{\pi iz}{\xi_0} + 2\pi i\mathbf{g} \cdot \mathbf{R}\right) \quad (26.3)$$

The justification for this substitution is the same as always. You'll notice that $\phi_0(z)_{(\text{sub})}$ is the same as before, but $\phi_{\mathbf{g}}(z)_{(\text{sub})}$ now includes a $\mathbf{g} \cdot \mathbf{R}$ term. The reason for this substitution is that it will give us a simple expression for $d\phi_{\mathbf{g}}/dz$.

The equations become

$$\frac{d\phi_0(z)_{(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}} \phi_{\mathbf{g}}(z)_{(\text{sub})} \quad (26.4)$$

and

$$\frac{d\phi_{\mathbf{g}}(z)_{(\text{sub})}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}} \phi_0(z)_{(\text{sub})} + \left[2\pi i\left(s + \mathbf{g} \cdot \frac{d\mathbf{R}}{dz}\right)\right] \phi_{\mathbf{g}}(z)_{(\text{sub})} \quad (26.5)$$

which can be rewritten, while dropping the subscript

$$\frac{d\phi_{\mathbf{g}}}{dz} = \frac{\pi i}{\xi_{\mathbf{g}}} \phi_0 + 2\pi is_{\mathbf{R}} \phi_{\mathbf{g}} \quad (26.6)$$

This equation looks just like equation 13.14 but with $s_{\mathbf{R}}$ instead of s , where

$$s_{\mathbf{R}} = s + \mathbf{g} \cdot \frac{d\mathbf{R}}{dz} \quad (26.7)$$

The concept of $s_{\mathbf{R}}$ is new.

The importance of this result is that although we have a new 's,' we have the same equation so we can use the rest of the analysis of Chapter 13 and obtain the same results with a modified value of s , i.e., $s_{\mathbf{R}}$. Therefore, we'll have the same thickness dependence so that the contrast of the defects will depend on both s and $\xi_{\mathbf{g}}$. The big change is that we can now treat the case where \mathbf{R} is a continuous function of z .

We will examine how the $\mathbf{g} \cdot d\mathbf{R}/dz$ and $\mathbf{g} \cdot \mathbf{R}$ terms are used to understand dislocations. Since the equations we have just derived have the same form as those we discussed in Chapters 13 and 25, we can expect many of the same properties in the images. In particular, the images of

defects will show the same sort of thickness dependence. We can also use the equations we derived in Chapter 25, so we have two ways of looking at the defects

- $\mathbf{g} \cdot \mathbf{R}$ contrast is used when \mathbf{R} has a single value,
- $s_{\mathbf{R}}$ contrast is used when \mathbf{R} is a continuously varying function of z , which in turn is associated with $\mathbf{g} \cdot d\mathbf{R}/dz$.

Now let's consider the principles of this analysis. Remember, we are not trying to be quantitative or totally rigorous. We will generalize the two-beam treatment for the imperfect crystal. Note that we still have beams, it's a dynamical situation, and we assume that the column approximation is valid. So how does the column approximation relate to the theory? The model relates \mathbf{R} to the column as shown in Figure 26.3 and the calculation is for a continuum even though we have atoms. The important point is that the displacement field, \mathbf{R} , varies with position, \mathbf{r} ; we can define the origin as the core of the defect. We'll go through the calculation for a dislocation parallel to the foil surface.

As we saw in Section 13.11, the column approximation is equivalent to assuming that the crystal can be divided into narrow columns. We then calculate the amplitudes of the beams in any such column as if the whole crystal consisted of an infinite number of identical columns. The approximation is valid when we don't need to see image detail below 2–3 nm. The actual diameter of the column depends on the diffracting conditions. We can include the effect of distortions due to strains from lattice defects by imagining that the column consists of slabs of perfect crystal, each displaced by an amount $\mathbf{R}(z)$ (like a mini-SF that's different on every plane). Remember that z is actually measured along the column.

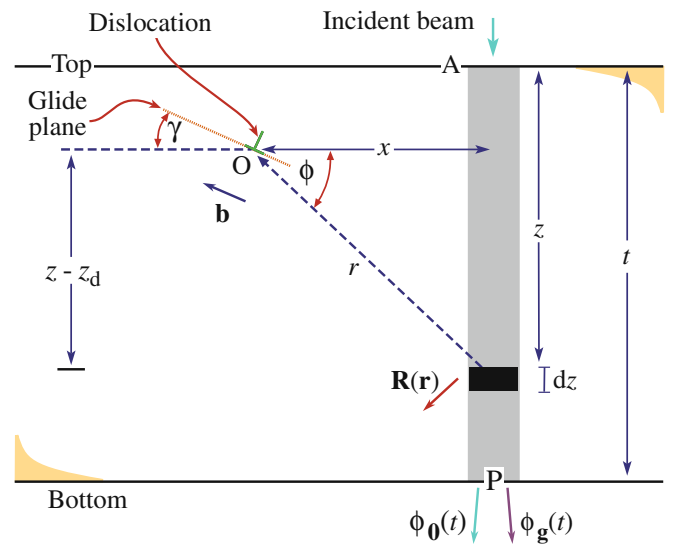


FIGURE 26.3. The effect of a dislocation with Burgers vector, \mathbf{b} , at O on a column, distance x away. The effect of the strain field on the electron waves in the column is integrated in increments dz over its total length t , giving amplitude $\phi_0(t)$ and $\phi_{\mathbf{g}}(t)$ at P.

26.3 CONTRAST FROM A SINGLE DISLOCATION

When we study dislocations, we usually want to know how many there are (the density) and whether they are edge, screw, or mixed in character. The displacement field in an isotropic solid for the general, or mixed, case can be written as

$$\mathbf{R} = \frac{1}{2\pi} \left(\mathbf{b}\phi + \frac{1}{4(1-\nu)} \{ \mathbf{b}_e + \mathbf{b} \times \mathbf{u} (2(1-2\nu)\ln r + \cos 2\phi) \} \right) \quad (26.8)$$

For convenience, \mathbf{R} is given here in polar coordinates (r and ϕ) shown in Figure 26.3; \mathbf{b} is the Burgers vector, \mathbf{b}_e is the edge component of the Burgers vector, \mathbf{u} is a unit vector along the dislocation line (the line direction), and ν is Poisson's ratio.

It was particularly important to be able to write down this expression when we did the calculations by hand. However, when we have a computer available, it's quite straightforward to use anisotropic elasticity or just feed in displacements calculated from a computer model of the atom structure.

The amplitude of the diffracted beam, ϕ_g , is directly influenced by the value of \mathbf{R} . We can consider two particular cases, namely, the screw and edge dislocations. For the screw dislocation, $\mathbf{b}_e = 0$ and \mathbf{b} is parallel to \mathbf{u} so that $\mathbf{b} \times \mathbf{u} = 0$. Then the expression for \mathbf{R} in equation 26.8 simplifies to

$$\mathbf{R} = \mathbf{b} \frac{\phi}{2\pi} = \frac{\mathbf{b}}{2\pi} \tan^{-1} \left(\frac{z - z_d}{x} \right) \quad (26.9)$$

Here, z is the distance traveled down the column and z_d is the distance of the dislocation core below the top surface (again, refer to Figure 26.3). The dependence on $(z - z_d)$ emphasizes that the displacement field is present above and below the dislocation; it affects the whole column. From these two equations we see that $\mathbf{g} \cdot \mathbf{R}$ is proportional to $\mathbf{g} \cdot \mathbf{b}$. For this reason, we often discuss images of dislocations in terms of $\mathbf{g} \cdot \mathbf{b}$ (g-dot-b) contrast. Examples of $\mathbf{g} \cdot \mathbf{b}$ values for some dislocations lying on a (111) plane in an fcc material with a [011] beam direction are given in Table 26.1.

THE SPECIMEN

If the sample is a single crystal, then you need to prepare a specimen with a [111] foil normal, so that you can image long segments of the dislocations lying in the plane of the foil on their (111) glide plane.

The second special case arises when the dislocation is pure edge in character. Then $\mathbf{b} = \mathbf{b}_e$ and $\mathbf{g} \cdot \mathbf{R}$

TABLE 26.1. Different Burgers Vectors and Different Reflections Give Different $\mathbf{g} \cdot \mathbf{b} = n$ Values

$\mathbf{g} \setminus \mathbf{b}$	$\frac{1}{6}[11\bar{2}]$	$\frac{1}{6}[1\bar{2}1]$	$\frac{1}{6}[\bar{2}11]$	$\frac{1}{3}[111]$
$\pm(1\bar{1}1)$	$\pm\frac{1}{3}$	$\pm\frac{2}{3}$	$\pm\frac{1}{3}$	$\pm\frac{1}{3}$
$(\bar{1}\bar{1}1)$	$\pm\frac{2}{3}$	$\pm\frac{1}{3}$	$\pm\frac{1}{3}$	$\pm\frac{1}{3}$
$(0\bar{2}2)$	± 1	± 1	0	0
(200)	$\pm\frac{1}{3}$	$\pm\frac{1}{3}$	$\pm\frac{2}{3}$	$\pm\frac{2}{3}$
$(3\bar{1}1)$	0	± 1	± 1	± 1
$(\bar{3}\bar{1}1)$	± 1	0	± 1	± 1

The dislocations all lie on a (111) plane in an fcc material; the beam direction is [011].

involves two terms $\mathbf{g} \cdot \mathbf{b}$ and $\mathbf{g} \cdot \mathbf{b} \times \mathbf{u}$. (The latter term is read as 'g-dot-b-cross-u.') The displacement field causes the Bragg-diffracting planes associated with \mathbf{g} to bend. Incidentally, the origin of $\mathbf{g} \cdot \mathbf{b} \times \mathbf{u}$ is interesting; it arises because the glide plane is buckled by the presence of an edge dislocation as illustrated in Figure 26.4. This buckling can be important because it complicates the analysis of \mathbf{b} for some dislocations with an edge component, as we'll see below.

- Always remember: $\mathbf{g} \cdot \mathbf{R}$ causes the contrast and for a dislocation, \mathbf{R} changes with z .
- We say that $\mathbf{g} \cdot \mathbf{b} = n$. If we know \mathbf{g} and we determine n , then we know \mathbf{b} .

An experimental point: you usually set s to be greater than 0 for \mathbf{g} when imaging a dislocation in two-beam conditions. Then the dislocation can appear dark against a bright background in a BF image. Of course, you still need to think about $s_{\mathbf{R}}$ and $d\mathbf{R}/dz$ since these will vary with z , as we saw in Figure 26.1.

The + and - signs in Table 26.1 are very important. If the sign of \mathbf{R} , and hence $\mathbf{g} \cdot \mathbf{R}$ or $\mathbf{g} \cdot \mathbf{b}$, reverses, then the image of the dislocation will move to the other side of the projected

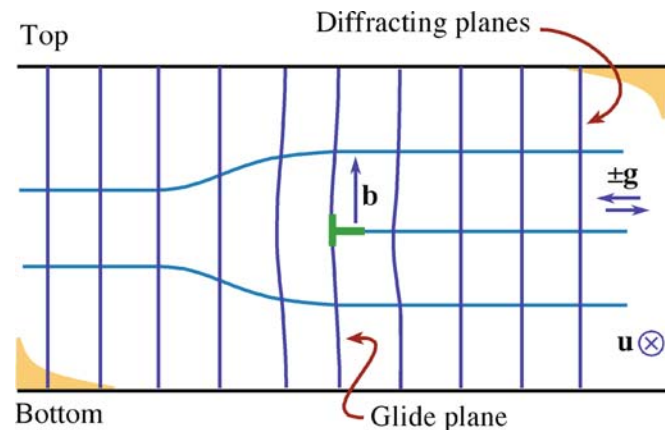


FIGURE 26.4. Buckling of the glide planes arises because of the term $\mathbf{g} \cdot \mathbf{b} \times \mathbf{u}$ and is important because it complicates the analysis of \mathbf{b} .

position of the dislocation core. If you look carefully at Figure 26.1, you can appreciate that reversing the sign of s produces the same effect as reversing the sign of \mathbf{g} . We can summarize these two ideas in terms of the quantity $(\mathbf{g}\cdot\mathbf{b})s$ ($\mathbf{g}\cdot\mathbf{b}$ -times- s), as shown in Figure 26.5.

If we identify two reflections \mathbf{g}_1 and \mathbf{g}_2 for which $\mathbf{g}\cdot\mathbf{b} = 0$, then $\mathbf{g}_1 \times \mathbf{g}_2$ is parallel to \mathbf{b} . This identification of \mathbf{b} is actually a little more complicated because dislocations

INVISIBILITY CRITERION
 If $\mathbf{g}\cdot\mathbf{b} = 0$, then you won't see any contrast because the diffracting planes are then parallel to \mathbf{R} .

appear out of contrast when $\mathbf{g}\cdot\mathbf{b} < 1/3$; similarly, dislocations need not be invisible even if $\mathbf{g}\cdot\mathbf{b} = 0$ when $\mathbf{g}\cdot\mathbf{b} \times \mathbf{u} \neq 0$.

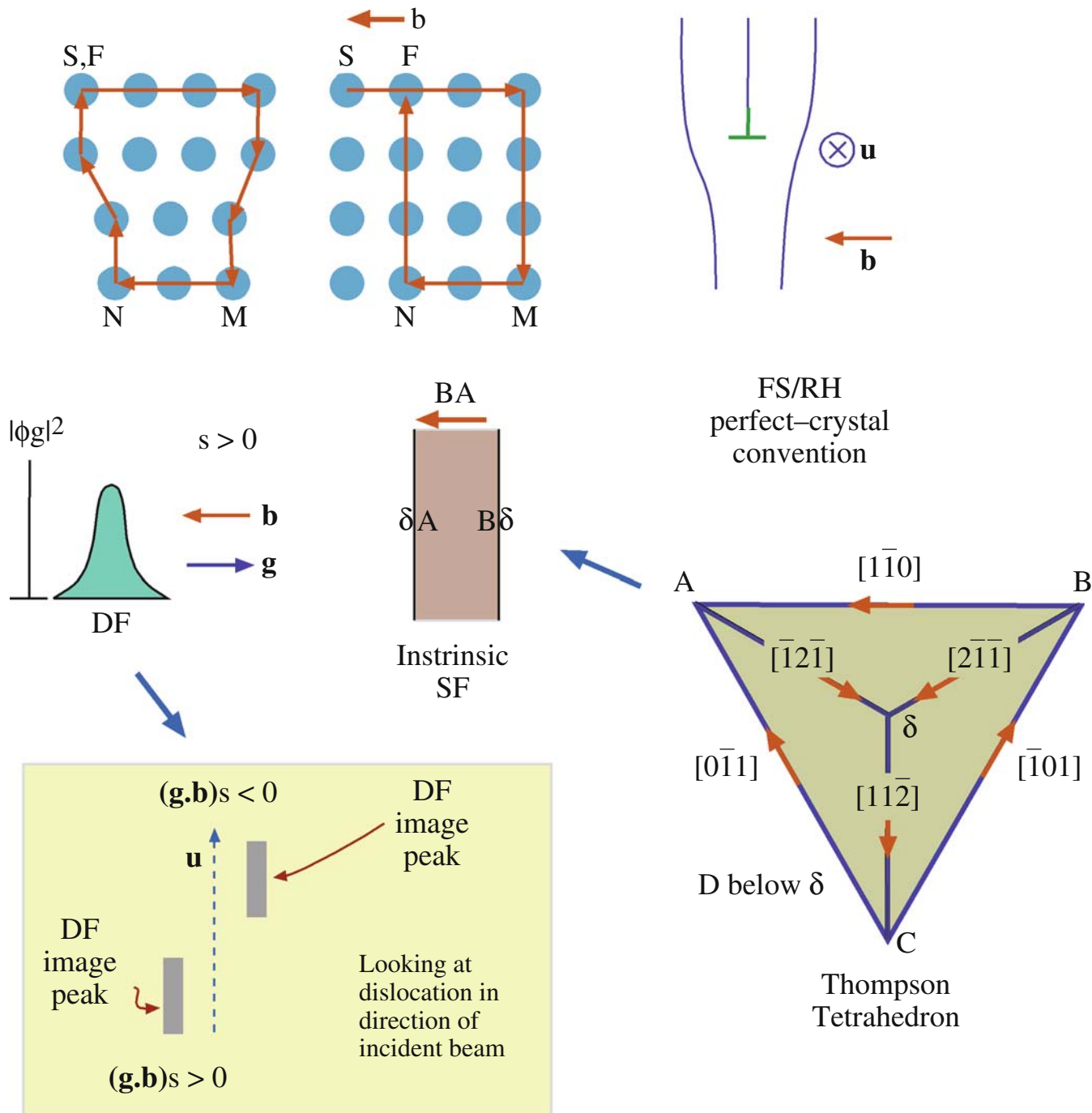
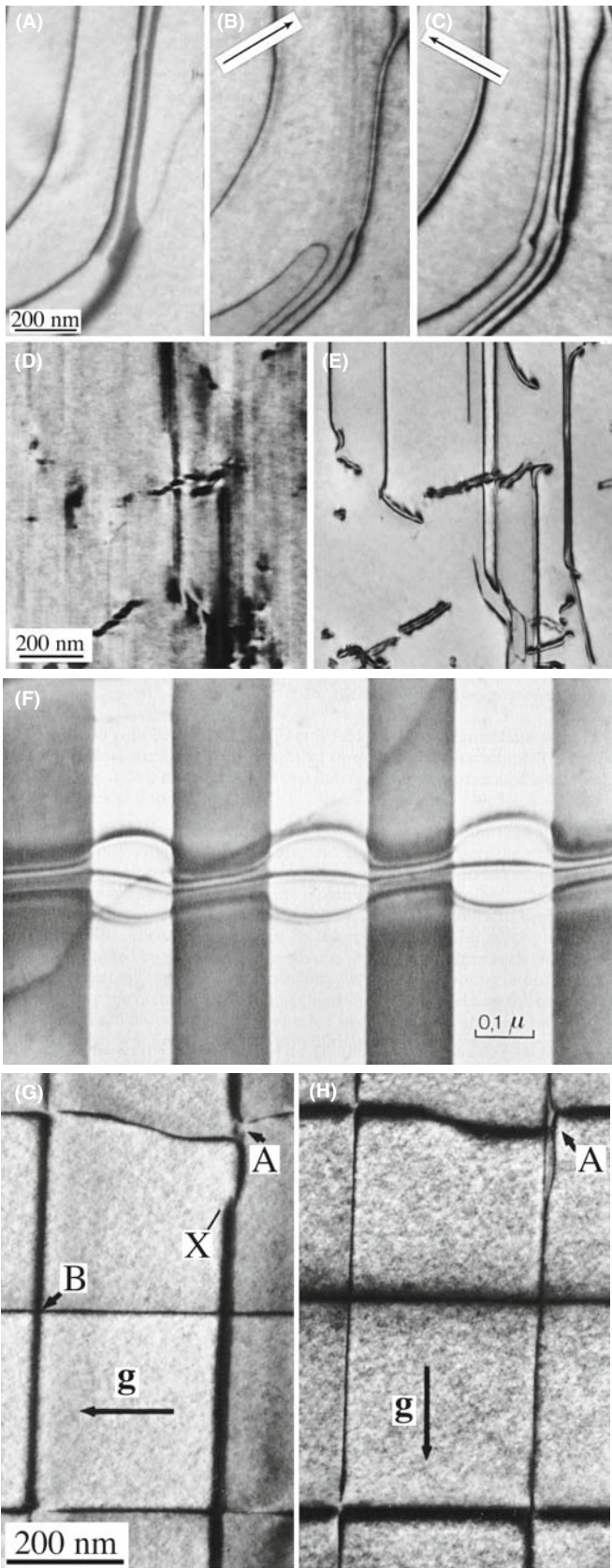


FIGURE 26.5. A brief summary of dislocations in an fcc crystal: \mathbf{b} is defined by the finish- (F) to-start (S) vector in a right-hand (RH) circuit that comes to closure around the dislocation but fails to close in the perfect crystal. The location of the diffracted intensity $|\phi_g|^2$ relative to the core depends on the sign of \mathbf{b} , \mathbf{g} , and s for the FSRH convention. If any sign is reversed, the contrast shifts across the core. When a perfect dislocation splits into Shockley partial dislocations, the order of the partial dislocation is given by the Thompson tetrahedron.



If we compare the contrast from a dislocation with that from a SF, the difference is that now α is a continuously varying function of z . The image of the dislocation itself shows thickness fringes but it may be ‘out of contrast’ at some depths or thicknesses, as you can see (look carefully) in the experimental image shown in Figure 26.6E.

Some points to remember from this discussion are

- The sign of s affects the image.
- The sign of x affects the image; the image is asymmetric.
- The magnitude of s affects the image.
- The depth of the dislocation and the thickness of the specimen affect the image.
- The appearance of the image depends on $\mathbf{g}\cdot\mathbf{b}$ or, more completely, on $(\mathbf{g}\cdot\mathbf{b})s$ and $\mathbf{g}\cdot\mathbf{b} \times \mathbf{u}$.
- If we repeat this analysis for other values of $\mathbf{g}\cdot\mathbf{b}$ ($=n$) and plot intensities, we would find that the image width becomes broader as n increases.
- Note where the dislocation image ‘comes from’: the position of the line in the image only rarely corresponds to the projected position of the dislocation; it is usually displaced to one side of the core.
- As a complication, remember that the dislocations will probably be found in wedge specimens, not ideal parallel-sided ones.

A final ‘rule of thumb’ which you may find useful (from computer modeling and early analytical calculations) is that if $\mathbf{g}\cdot\mathbf{b} = 0$, you can still ‘see’ dislocations when $\mathbf{g}\cdot\mathbf{b} \times \mathbf{u} \geq 0.64$. For fcc materials, this rule can be useful when the foil is not parallel to a $\{111\}$ plane.

Other examples of dislocation images are illustrated in Figure 26.6. Remember that partial dislocations are not only present in fcc metals; they also occur in many fcc semiconductors and many layer materials. Such materials may have a very low stacking-fault energy allowing the partial dislocations to separate, forming wide ribbon-like defects, as shown in Figure 26.6A–C. The single ‘line’ (that is actually a dark line and a less-dark line) below the arrow in (C) is a dislocation having its Burgers vector parallel to \mathbf{g} (so $\mathbf{g}\cdot\mathbf{b} = 2$); you can see two ‘peaks’ in the image, one darker and broader than



FIGURE 26.6. (A–C) Three strong-beam BF images from the same area using (A) $\{111\}$ and (B, C) $\{220\}$ reflections to image dislocations which lie nearly parallel to the (111) foil surface in a Cu alloy which has a low stacking-fault energy. (D, E) Dislocations in Ni_3Al in a (001) foil imaged in two orthogonal $\{220\}$ reflections. Most of the dislocations are out of contrast in (D). (F) A complex dislocation crossing a (rotational) domain boundary; the character of the dislocation changes and thus its dissociation width changes. (G, H) Dislocations of a (001) interface between two slightly lattice-mismatched III–V compounds.

the other. Notice that one of those dark lines in (C) has nearly disappeared in (B) and the dark peak is on the other side of the dislocation. A group of three parallel lines is present in (C) but is almost out of contrast in (B). These are Shockley partial dislocations all having the same \mathbf{b} and thus giving $\mathbf{g}\cdot\mathbf{b} = 0$ in (B) (the three lines actually form by the dissociation of a perfect dislocation with Burgers vector $\frac{1}{2}\langle 112 \rangle$). The $11\bar{1}$ image (A) is formed by tilting the specimen to a 112 pole ($\sim 20^\circ$ from the 111 pole) and shows contrast from the stacking faults themselves (imagine them filling the area between the dislocations); these faults will never give contrast at the 111 pole since $\mathbf{g}\cdot\mathbf{R}$ is then always 0 (or an integer). Intermetallics tend to have large unit cells so that the superlattice dislocations dissociate into partial dislocations, which would have been perfect dislocations in the disordered crystal (Figure 26.6D and E). These super-partial dislocations can dissociate further as they might have in the disordered lattice, or they can separate differently in different ordered domains (Figure 26.6F). Dislocations in interphase boundaries can be revealed by imaging with different reflections (Figure 26.6G and H). Since the dislocations are present to accommodate the mismatch, they must lie at, or close to, the (001) phase boundary. It can be difficult to analyze their Burgers vectors unambiguously, because the adjoining materials have different extinction distances, etc. The topic is ‘image interpretation’: now you know why. One of the extra challenges is determining the plane on which this dissociation occurs. We’ll illustrate how we can see $\mathbf{g}\cdot\mathbf{b} \times \mathbf{u}$ contrast when we examine dislocation loops in Zn in Section 26.6.

A QUESTION

How many dislocations can you really see in Figure 26.6H? Answer: zero.

26.4 DISPLACEMENT FIELDS AND EWALD’S SPHERE

In Section 26.2, we showed that when a displacement $\mathbf{R}(\mathbf{r})$ is present, we can think of s as being replaced by $s_{\mathbf{R}}$ (equation 26.7). This new s should be written more completely as

$$s_{\mathbf{R}} = s + \mathbf{g} \cdot \frac{\partial \mathbf{R}}{\partial z} + \theta_B \mathbf{g} \cdot \frac{\partial \mathbf{R}}{\partial x} \quad (26.10)$$

The point is that, as you can see in Figure 26.7, \mathbf{R} causes the lattice planes to bend through an angle $\delta\phi$. So two other parameters, namely \mathbf{g} and s , also change. The diffraction vector is actually lengthened by $\Delta\mathbf{g}$ but, more

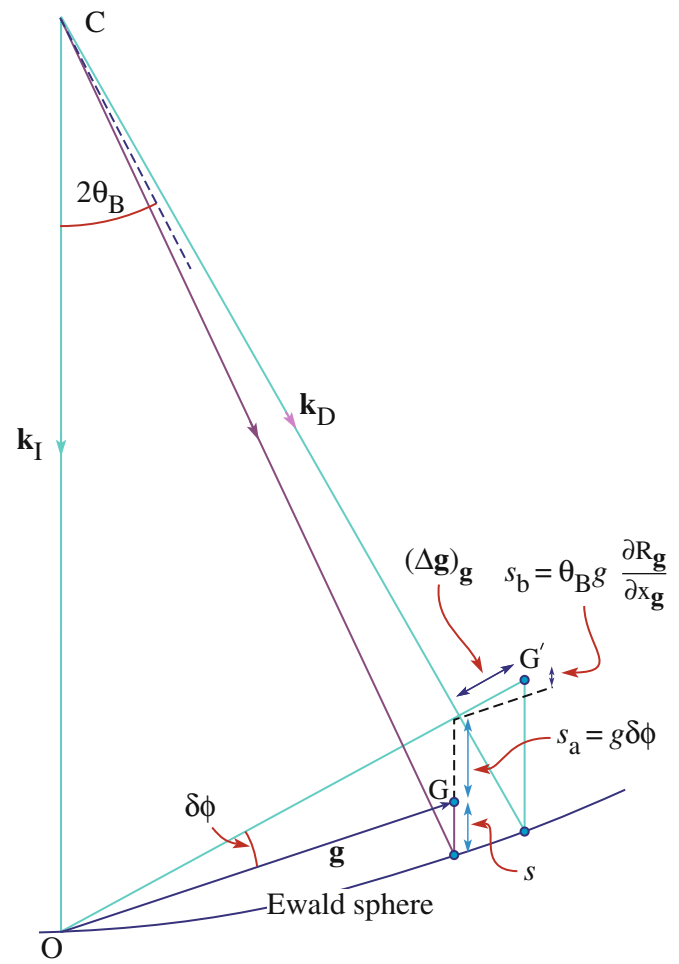


FIGURE 26.7. The strain field of the dislocation causes the lattice planes to bend through an angle $\delta\phi$. So \mathbf{g} and s also change. The diffraction vector is lengthened by $\Delta\mathbf{g}$ and \mathbf{g} is rotated. So s increases by the two components of $s_{\mathbf{R}}$, i.e., s_a and s_b .

importantly, \mathbf{g} is rotated. The result is that s increases by the two components s_a and s_b shown in the figure, to give $s_{\mathbf{R}}$. If you manipulate this equation for small angles you can produce equation 26.7. We usually neglect the third term because θ_B is small but it can become important when screw dislocations intersect the surface.

An alternative way of looking at this deformation is to think of \mathbf{g} as changing by $\Delta\mathbf{g}$. We can define this change by the equation

$$\mathbf{g} \cdot (\mathbf{r} - \mathbf{R}(\mathbf{r})) = (\mathbf{g} + \Delta\mathbf{g}) \cdot \mathbf{r} \quad (26.11)$$

so that

$$-\mathbf{g} \cdot \mathbf{R}(\mathbf{r}) = \Delta\mathbf{g} \cdot \mathbf{r} \quad (26.12)$$

The implication is that the information about the displacement field, $\mathbf{R}(\mathbf{r})$, is present in the region around \mathbf{g} but not actually at \mathbf{g} . Remember that the reflection \mathbf{g} is present because we have a perfect crystal. It is difficult to image this type of scattering (but it becomes easier using

EFTEM). If you displace the objective aperture, you will still see the dislocation, but other inelastic scattering will complicate image interpretation. We saw that scattering does indeed occur between Bragg reflections in Section 17.6. An analogy for scattering from dislocations is the scattering of light from a single slit which we discussed in Chapter 2.

In the deformable-ion approximation (Section 25.13), we make the assumption that the atom doesn't know it has moved. If $\mathbf{R}(\mathbf{r})$ varies rapidly, as it does near the core of a dislocation, the approximation must fail. You can draw the same conclusion whenever the density of the material changes rapidly. So what we should do is use a better model for the atomic potential, one that also takes account of what happens to the valence electrons at such a defect. Of course, linear elasticity theory also fails when the strains, and hence $\mathbf{R}(\mathbf{r})$, are large, as at dislocation cores.

26.5 DISLOCATION NODES AND NETWORKS

You can analyze the Burgers vectors of dislocations which form networks directly and easily, if all the dislocations lie in a plane parallel to the surface of the specimen, as illustrated in Figure 26.8 for the case of

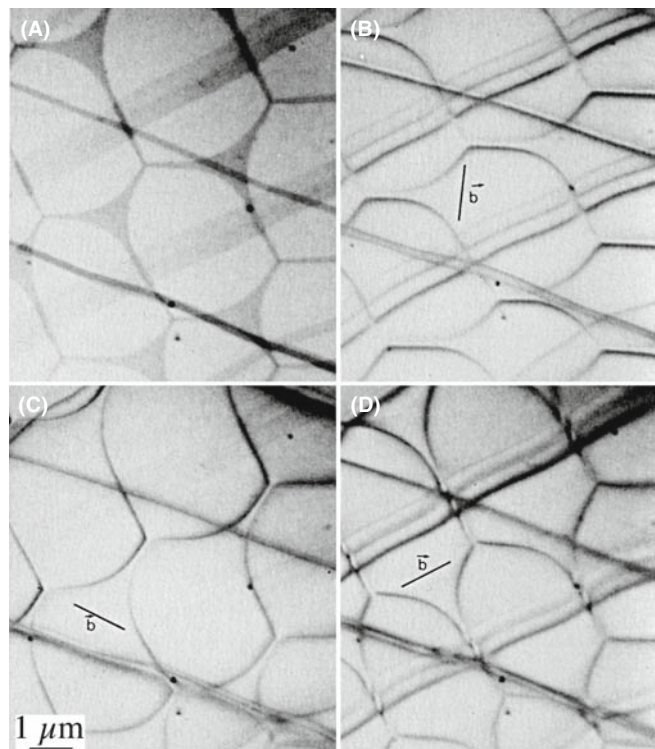


FIGURE 26.8. Dislocation networks in graphite. In (A) the stacking faults parallel to the foil surface are imaged by using a \mathbf{g} that is inclined to the fault (so $\mathbf{g}\cdot\mathbf{R}\neq 0$); in (B–D) two of the three dislocations at each node are in contrast but the third is invisible. Knowing \mathbf{g} for each image, the Burgers vector of the dislocations can be determined as shown in (B–D) ($\mathbf{g}\cdot\mathbf{b} = 0$ for the *missing* dislocations).

graphite. The idea is simple: you form a series of images using different \mathbf{g} -vectors. Don't forget that you can tilt to other poles; in fact, you'll often need to tilt the specimen just to image SFs which lie parallel to the foil surface, as in Figure 26.8A (but you won't see SF fringes then). When the diffraction vector is parallel to the plane of the SF you'll only see contrast from the dislocations as in Figure 26.8B–D. Such tilting experiments are essential if you're examining networks of misfit dislocations since the dislocations will then often have a component of their Burgers vector out of the plane of the network.

26.6 DISLOCATION LOOPS AND DIPOLES

Loops have been studied extensively because they can form when point defects coalesce. There are probably thousands of papers describing TEM studies of radiation damage and the formation of dislocation loops. In fact, many HVEMs were built in the 1960s just to study this problem. Questions which were answered led to a greatly improved understanding of irradiation processes (but failed to justify the construction of more nuclear power stations). We found that

- The loops can form by coalescence of interstitials or vacancies.
- The rate of growth, critical size, and nucleation time for different loops can be measured.
- Some of the loops are found to be faulted (containing a SF) while others are not faulted. The faulting should be related to the size of the loop and the stacking-fault energy of the material.

These studies were particularly instructive illustrations of the value of diffraction contrast.

- Dislocation loops can have either positive or negative \mathbf{b} , and can be inclined to the beam on different planes as shown in Figure 26.9. Many person-years have been spent analyzing the inside-outside contrast from such defects to determine if they were caused by vacancies or interstitials.
- Loops can be present that show no $\mathbf{g}\cdot\mathbf{b}$ contrast.
- Loops can enclose single or multiple stacking faults, and so exhibit SF contrast as shown in Figure 26.10.
- The dislocation dipole is a special case and gives an important example of interacting dislocations. TEM is the best way to image dipoles because they have no long-range strain fields; the Burgers vector of the complete dipole is zero! (More on this in Chapter 27.)

Dislocations in Zn provide a particularly nice illustration of $\mathbf{g}\cdot\mathbf{b}\times\mathbf{u}$ contrast. If the specimen surface is parallel

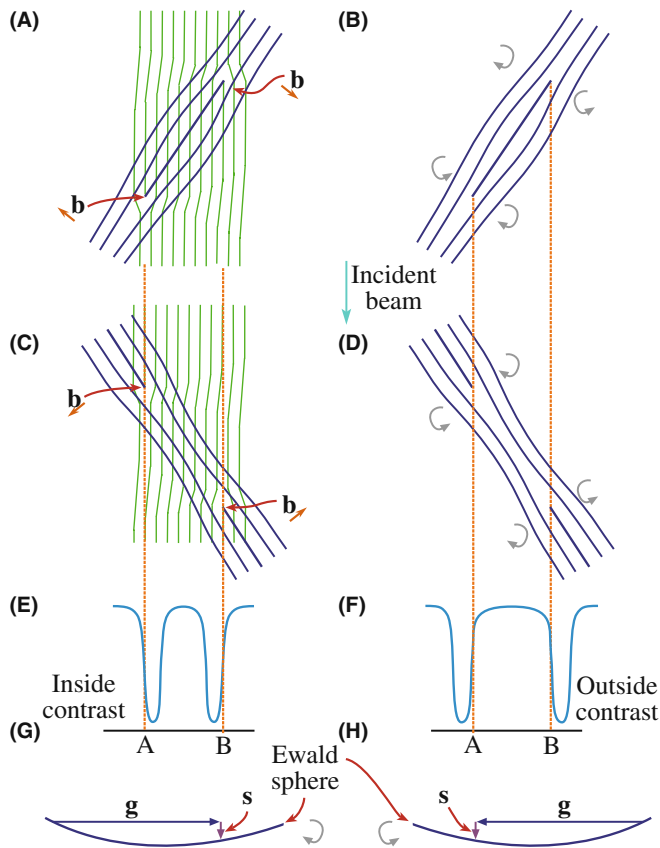


FIGURE 26.9. (A) Structure of an interstitial loop relative to the diffracting plane (faint lines). (B) Arrows show the rotation of the diffracting planes around the dislocation. (C, D) Vacancy loops. (E, F) Position of the image contrast relative to the projected dislocation position. Inside contrast occurs when clockwise rotation of the diffracting planes brings them into the Bragg condition. Outside contrast occurs for the counter-clockwise case. (G, H) The relationship between g , s , and the sense of rotation. Everything is reversed if the loops are tilted in the opposite direction relative to the beam (i.e., reflect this figure in a mirror).

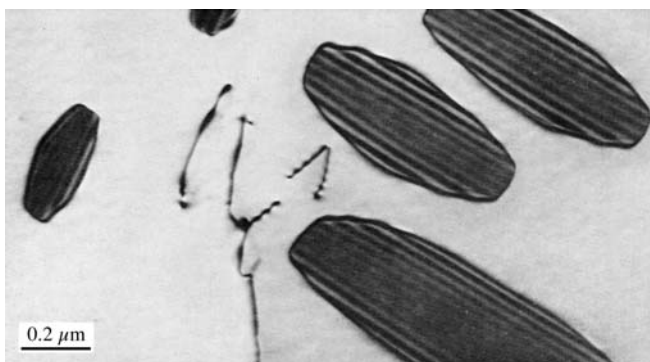


FIGURE 26.10. Dislocation loops in irradiated Ni showing SF contrast.

to the (0001) basal plane, then dislocation loops can readily form by coalescence of vacancies. In Figure 26.11, b is normal to g so that $g \cdot b = 0$. These loops give a clear illustration of how the appearance of the image

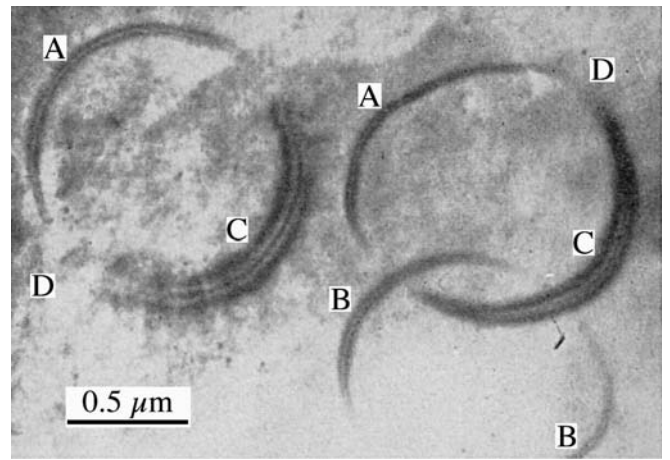


FIGURE 26.11. Prismatic loops in Zn parallel to the (0001) surface of the specimen with $b = c[0001]$. All round the loop, b is normal to g so that $g \cdot b = 0$ and the vector $b \times u$ lies in the plane of the loop. At A, B, and C, $b \times u$ is parallel to g so that we see strong contrast. However, at D, $b \times u$ and g are mutually perpendicular so that $g \cdot b \times u = 0$ and the loop disappears.

depends on the line direction, u , of a dislocation. Note that you can see the dislocation contrast, even though $g \cdot b$ is zero, so this is not an absolute criterion for invisibility.

The above discussion is fine if the loops are large, but a problem arises when they are small. You must then consider the details of the contrast mechanism.

This was, of course, true for all these images, but now the size of the defect is small compared to the extinction distance. The schematic shown in Figure 26.12 summarizes the contrast that arises from small vacancy loops; if the loops were interstitial in nature, the contrast would be reversed. Not only does the black/white contrast change as the position of the defect changes in the specimen, but its size also appears to change. When the nature of the loops becomes more complex, the appearance of the image may also become more difficult to interpret with ‘butterflies,’ ‘lozenges,’ and ‘peanuts’ being seen. Notice that the behavior of the contrast differs in the BF and DF images; this effect is similar to that which we discussed in Chapter 25 and is again related to anomalous absorption.

IMAGES AND THICKNESS

The basic idea is that the appearance of the image is now dependent on the thickness of the specimen.

Dislocation dipoles can be present in great numbers in heavily deformed metals, but can also be important in the degradation of some semiconductor devices. Dipoles can be thought of as loops which are so elongated that they look like a pair of single dislocations of opposite Burgers vector, lying on parallel glide planes. As a result, they are best recognized by their

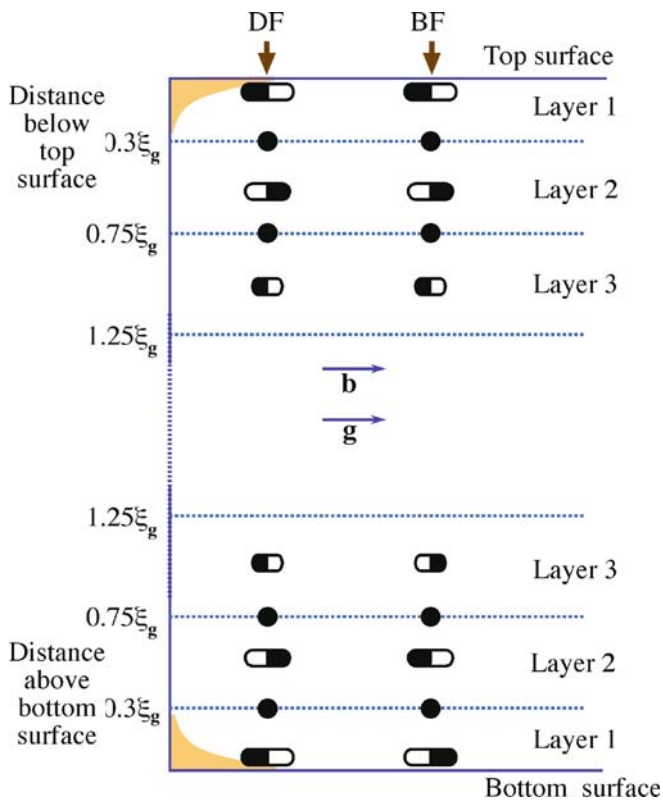


FIGURE 26.12. Changes in the black-white contrast from small dislocation loops at different depths in the specimen. The DF shows the same contrast at top and bottom while the BF contrast is complementary at the two surfaces.

‘inside-outside’ contrast, as illustrated in Figure 26.13. You can appreciate the origin of the term by looking at the projection of the images of the two dislocations when you reverse the sign of \mathbf{g} : since the two dislocations have opposite Burgers vectors, Figure 26.9 tells you that one image will lie on one side of the core and the other on the opposite side. The order reverses when we reverse \mathbf{g} .

26.7 DISLOCATION PAIRS, ARRAYS, AND TANGLES

Remember, you are not limited to \mathbf{g} -vectors which are parallel to the foil surface; hence you can tilt the specimen to see SF contrast. As we saw in Figure 26.8A, this is often helpful if you have SFs associated with the dislocations; you can then produce $\mathbf{g}\cdot\mathbf{R}$ contrast for the fault. We will discuss dislocation dissociation more in Chapter 27. If you look back at Figure 26.6 or 26.8, you will see the benefit of being able to see the SF. This figure also illustrates the effect of n on the dislocation contrast.

Consider a dislocation in an fcc metal which can dissociate into two Shockley partial dislocations on the (111) plane. We can write down the dislocation reaction as

$$\frac{1}{2}[1\bar{1}0] = \frac{1}{6}[1\bar{2}1] + \frac{1}{6}[21\bar{1}] \text{ on } (111) \quad (26.13)$$

If we image this dislocation using the $(2\bar{2}0)$ reflection, then $\mathbf{g}\cdot\mathbf{b} = 2$. If, instead, we use the $(20\bar{2})$ reflection, then $\mathbf{g}\cdot\mathbf{b} = 1$. The appearance of the image is very different even if we cannot see the individual partial dislocations.

The advantage of using high voltages to study arrays of dislocations is illustrated in Figure 26.14; everything we said in Chapter 11 applies when we study dislocations. We see thickness fringes at the surface, but these disappear in the central region of the foil—the contrast becomes constant. When the foils are this thick, you may find stereomicroscopy helpful in giving a 3D view of the defect arrangement; you can imagine its value in interpreting an image such as that shown in Figure 26.15A. The defects may be very close together in heavily deformed materials, as shown in Figure 26.15B; of course, the interpretation of such an image depends on the thickness of the specimen so you also need to know the voltage used to form the image. Ideally, make the specimen thin over very large areas to minimize image overlap. If the density of defects is too large, the weak-beam technique may be the only way to ‘look into’ the walls (Chapter 27) but then the specimen needs to be even thinner.

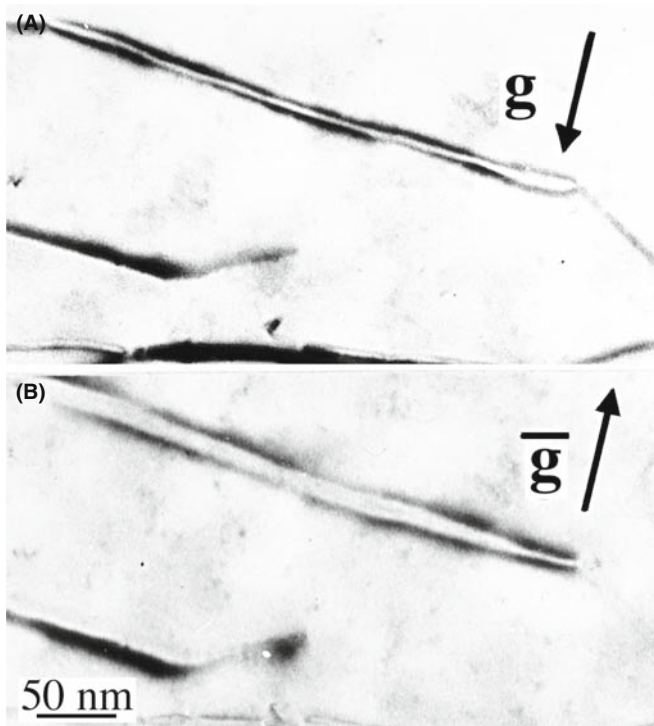


FIGURE 26.13. Images of dislocation dipoles in Cu showing inside-outside contrast on reversing $\mathbf{g} (\pm 220)$. (A) is inside; (B) is outside.

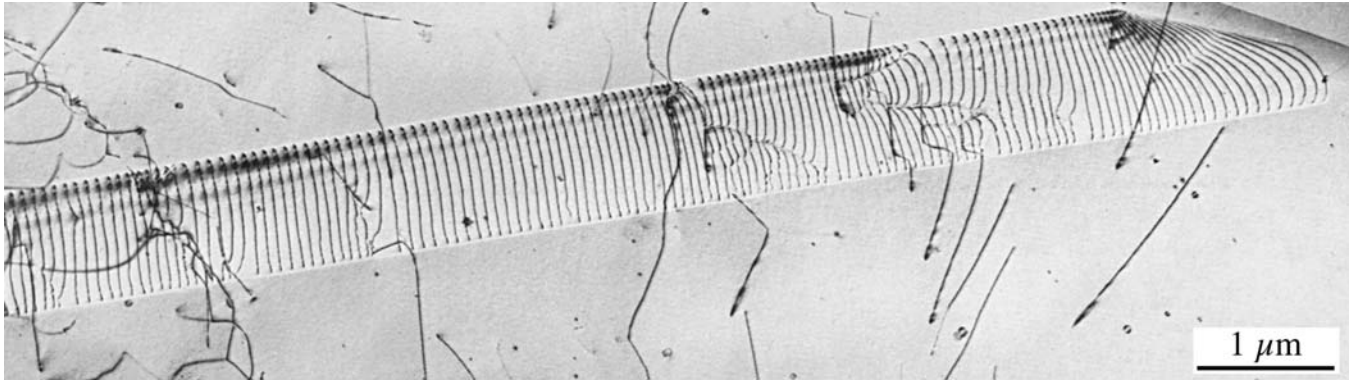


FIGURE 26.14. Dislocations threading through a very thick specimen in an image recorded using a (very) high-voltage (3 MeV) TEM.

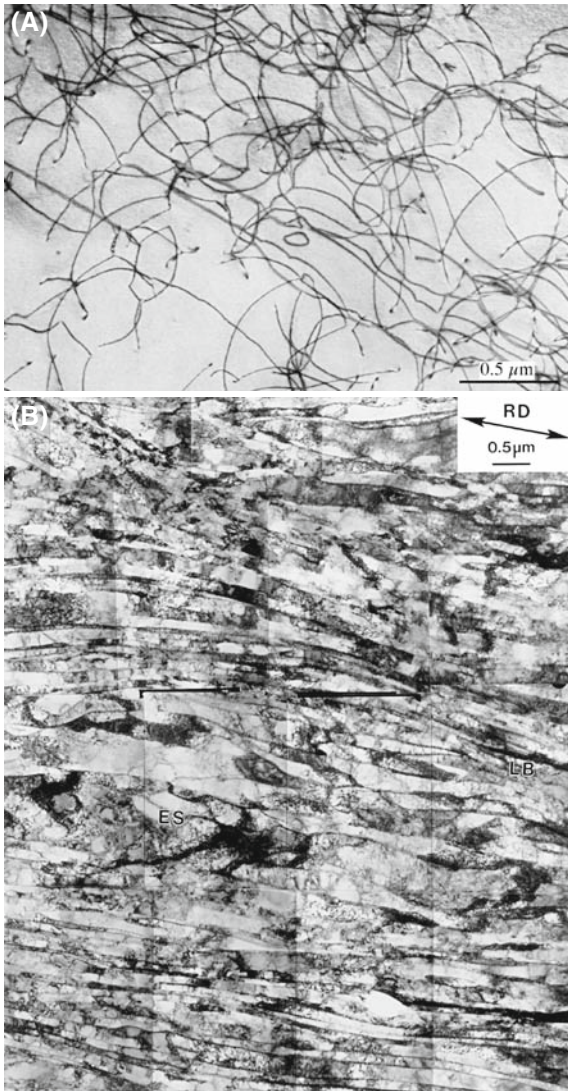


FIGURE 26.15. (A) Dislocation tangles in an Fe-35% Ni-20% Cr alloy, creep tested at 700°C; the dislocations have moved by glide and climb and do not lie on well-defined planes. (B) Dislocation walls in Al which has been heavily deformed by directional rolling. (200-keV electrons but super specimen.)

A CALCULATION

$\sin \theta_D/2 = b/2t.$

If $b \sim 0.25$ nm and $t = 50$ nm $\sin \theta_D/2 = 0.0025$ and $\theta_D = 0.29^\circ$; compare to a Bragg angle from $\sin \theta_B = (n\lambda/2d)$ (0.0037 nm/ 0.05 nm= 0.0074). $\theta_B = 0.42^\circ$. Notice how θ_D increases for thinner foils and θ_B (and λ) decreases as voltage increases.

26.8 SURFACE EFFECTS

In TEM, we always have thin foils. Dislocation strain fields are long range, but we often assign them a cut-off radius of ~ 50 nm. However, the thickness of the specimen might only be 50 nm or less, so we can expect the surface to affect the strain field of the dislocation, and vice versa.

When an edge dislocation lies parallel to the surface of a very thin specimen, it causes the specimen to bend. The effect is not large, but large enough compared to the Bragg angle, as illustrated schematically and with an example in Figure 26.16.

Similarly, if the dislocation is dissociated, the proximity of the surface causes its width to decrease (as in Figure 26.16C). We can model this situation using ‘image dislocations’ as shown in Figure 26.16D. The main point is that we can think of these image dislocations as forcing the partial dislocations closer together; the proximity of the surface can really change the structure of the defect, not just its contrast. A similar effect can occur when the dislocation is inclined to the surface and can result in a V-shaped dissociated dislocations, as we’ll show in Section 27.8.

A special interaction between dislocations and surfaces occurs when a dislocation tries to glide out of the material but can’t penetrate a surface layer (which might even be amorphous, as in the case of oxide films on metals), as shown in Figure 26.17.

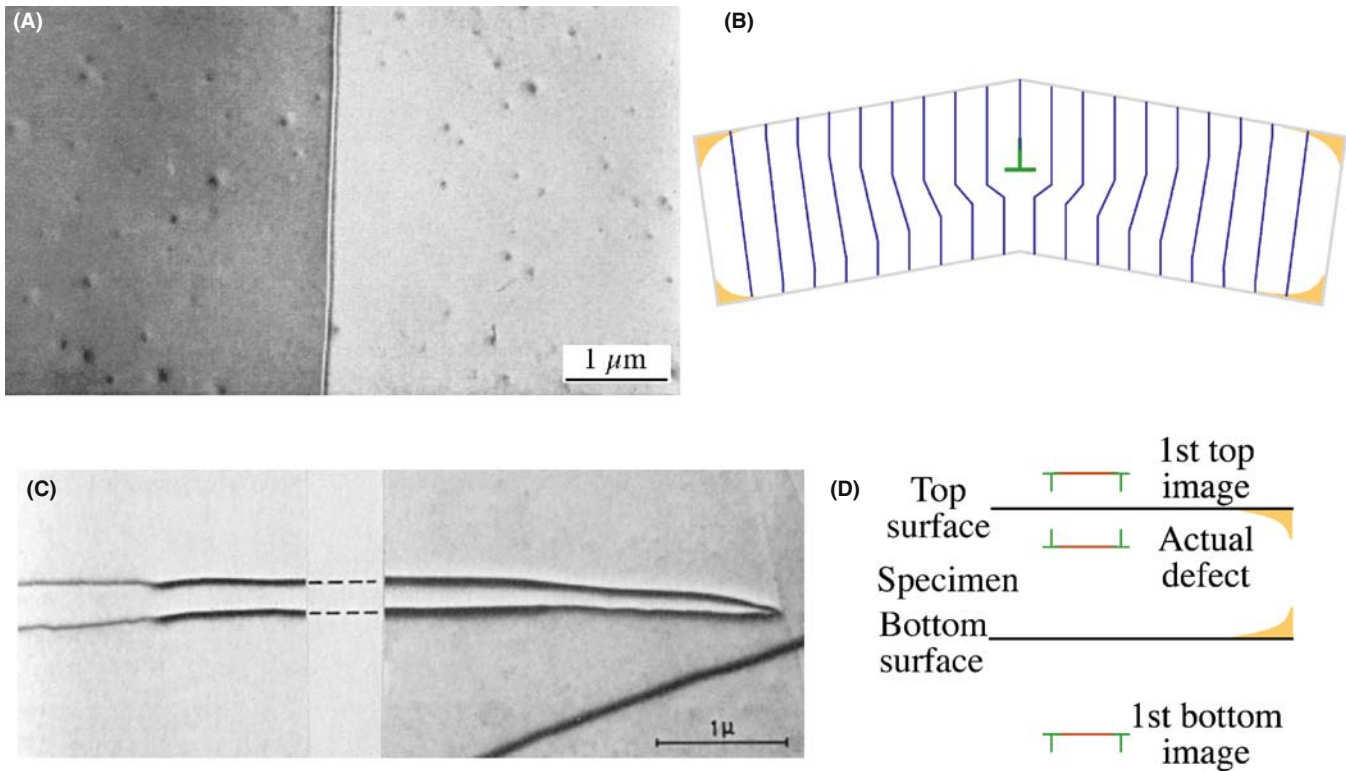


FIGURE 26.16. (A) BF image of a single-edge dislocation lying parallel to the surface of a very thin foil of SnSe, causes the diffracting planes to bend as shown schematically in (B) so we see different contrast intensity in the matrix on either side of the defect. (C, D) If the dislocation is dissociated, the image forces due to the surface cause its width to decrease as it approaches the surface. The schematic in (D) shows two image dislocations included to represent the effect of the nearest surface.

We can have dislocations at the surface just because the structure of the surface layer is different from that of the bulk material. The surface of materials can actually reconstruct. The surface of a very clean (111) Au film is actually more dense than the rest of the film and this misfit is accommodated by surface dislocations, as shown in Figure 26.18. We see the contrast because the strain field extends into the bulk layer. The identification of these dislocations has been confirmed using STM, which also gives more information on the detailed surface structure. However, they were observed first by TEM. The difficulty in TEM studies is that the surface contaminates unless you operate under UHV conditions. The same comment is true for STM of metals but we are used to having a poorer vacuum in TEM and faster/easier specimen exchanges.

GOLD

The dislocation in Figure 26.18 is one atomic plane below the surface but $\mathbf{g} \cdot \mathbf{b}$ still works.

Dislocations can be viewed nearly parallel to their line directions, when we still see contrast even for screw dislocations, as you can see in Figure 26.19A. Initially, this contrast is surprising since $\mathbf{g} \cdot \mathbf{b}$ and $\mathbf{g} \cdot \mathbf{b} \times \mathbf{u}$ must be

zero for any dislocation. However, the screw dislocation can relax at the surface, as shown in Figure 26.19B; it gives a twist. You see the screw dislocation even though $\mathbf{g} \cdot \mathbf{b} = 0$ because of this twist.

26.9 DISLOCATIONS AND INTERFACES

Interfaces are, of course, important in all polycrystalline materials. In metals, semiconductors, and thin films on substrates, the interaction between dislocations and interfaces is critical. So now we'll briefly examine the special features we see when combining line and planar defects, as illustrated in Figure 26.20. This is one topic where image simulation, which we'll discuss in Section 26.12, is invaluable. The location of the GB is known because we can see (just) the thickness fringes and the GB dislocations. Notice that there are also 'extrinsic' (lattice) dislocations running into the GB. This is a great figure for student questions!

When we have an array of dislocations, the strain fields overlap so that the value of $\mathbf{R}(\mathbf{r})$ for each dislocation tends to be reduced. This is the GB model of an interface.

Dislocations can be present at interfaces where the composition, or structure, or both change

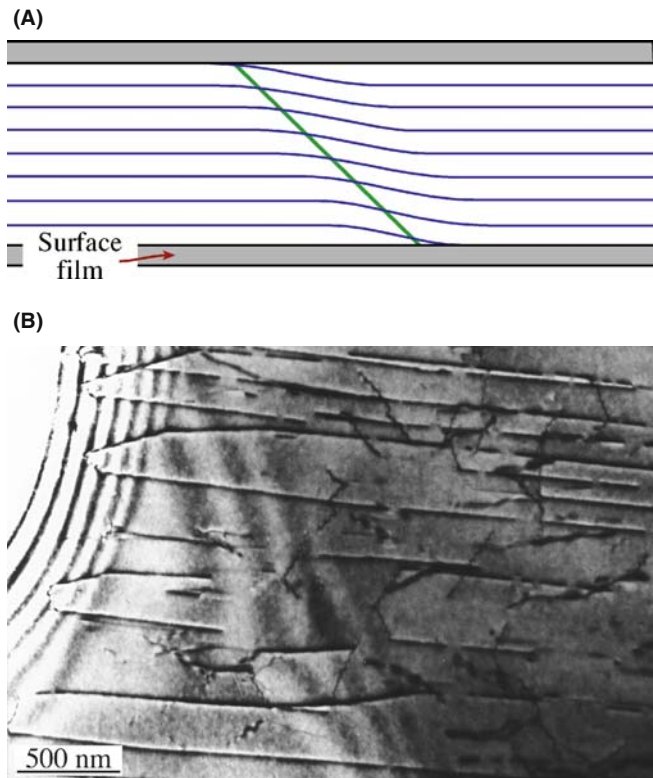


FIGURE 26.17. (A) Schematic diagram of dislocations pinned at the surface of the specimen by surface films such as oxides. (B) A reduced (i.e., metal) film on NiO pins dislocations. Such films may be introduced during or after thinning to electron transparency.

- Misfit dislocations accommodate the difference in lattice parameter between two well-aligned crystalline grains. Surface dislocations (as we saw in Section 26.8) are a special subgroup of misfit dislocations.
- Transformation dislocations are the dislocations that move to create a change in orientation or phase. The $\frac{1}{6}\langle 112 \rangle$ dislocations in twin boundaries in fcc materials are an example of transformation dislocations (twinning dislocations).

A complication in the analysis of images of interfacial dislocations is that they are often associated with steps in the interface. There are many examples of such steps in Figure 26.21, including an example of ‘step bunching.’ (Notice also that s is different for the two grains!) Sometimes, as is the case for the $\frac{1}{6}\langle 112 \rangle$ twinning dislocations, the dislocations must introduce a step. In other situations, steps are present but there is no dislocation. The difficulty is that we often encounter all three of these situations at the same time. We will also examine these defects using weak-beam conditions in Chapter 27 and using HRTEM in Chapter 28.

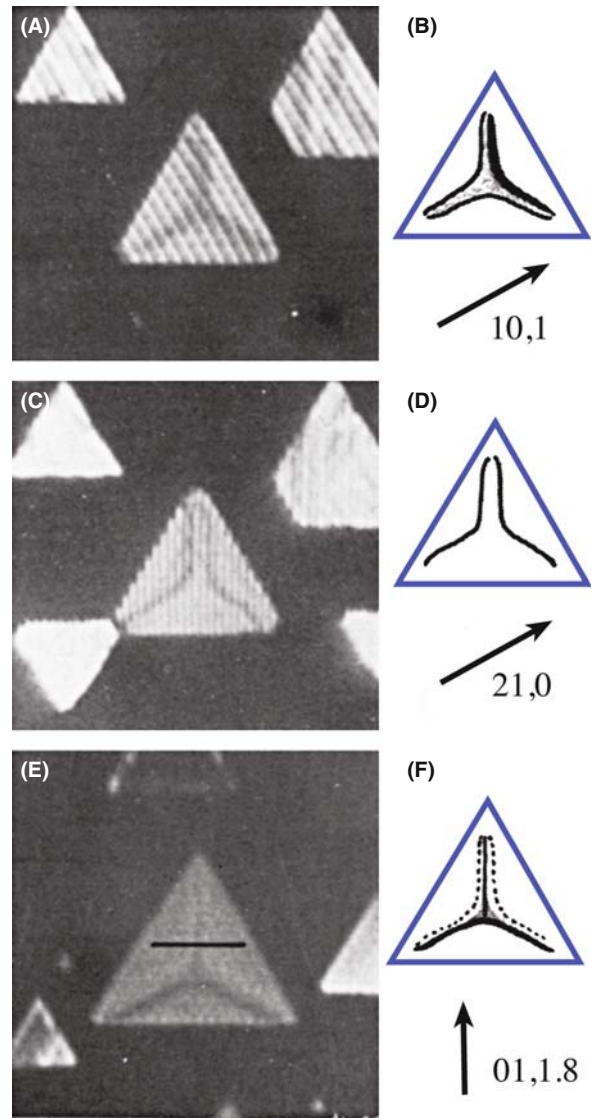


FIGURE 26.18. Dislocation networks can form at the surface of (111) Au islands because the surface layer relaxes to a ‘lattice’ parameter that is different from that of the bulk material. Different dislocations are visible under different diffracting conditions. The edge misfit dislocations form nodes in the triangular islands and show the usual $\mathbf{g}\cdot\mathbf{b}$ contrast behavior. (A), (C), and (E) are experimental images; (B), (D), and (F) are schematics showing \mathbf{g} and summarizing the contrast.

We will discuss the images first and then, remembering that information must also be present in the DP, we will relate the two.

In many cases that interest us, GBs appear as arrays of dislocations. In general, the grains are mis-oriented. There are some special cases as we saw in Chapter 24.

- Two grains may have a near-common plane and therefore a nearly common, but different, \mathbf{g} -vector.
- In small-angle GBs, θ is small, so the separation of the dislocations is large ($\sin \theta/2 = \mathbf{b}/2d$).

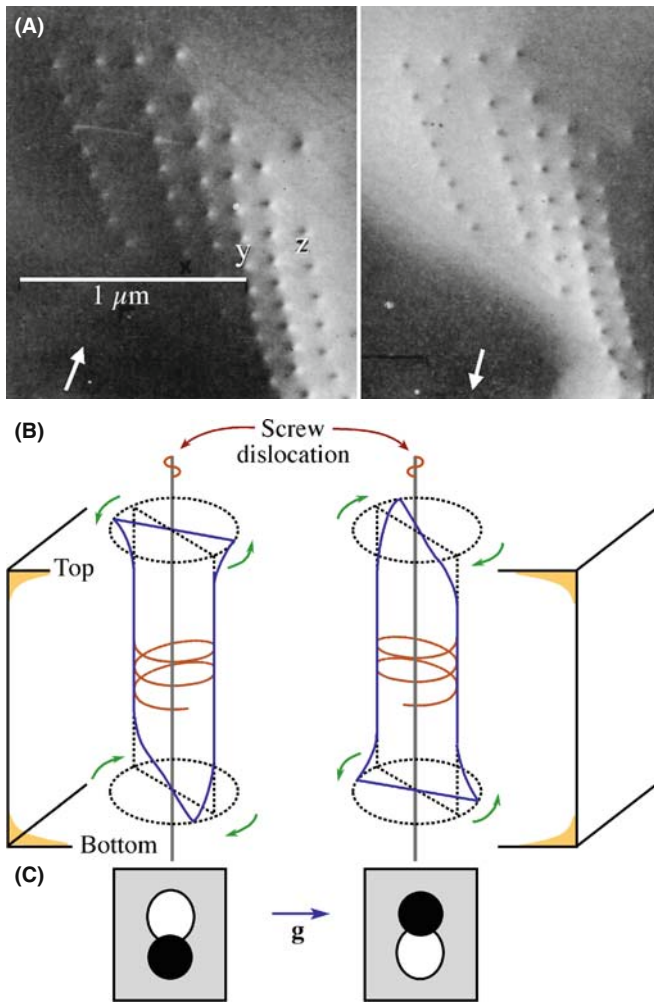


FIGURE 26.19. Screw dislocations viewed end on $\pm\mathbf{g}$. (A) Two images in $\pm\mathbf{g}$. (B) The twist relaxation at the surface. (C) Representations of the resulting contrast. The schematics in (B) show the diffracting planes rotating in the same direction away from the edge-on orientation at both surfaces.

The $\Sigma = 3$ twin boundary in fcc materials is an example of an interface where you can use common, but different, \mathbf{g} vectors. Here, e.g., the (333) plane in one grain is parallel to the $(\bar{5}11)$ plane in the other grain (because these two reflections superimpose), and the lengths of these two \mathbf{g} vectors are identical, as you can see in Figure 26.22. However, this common reflection would not normally be used to form an image because \mathbf{g} is rather large. This type of coincidence of dissimilar planes can also occur for other GBs.

In the case of small-angle boundaries, we can pretend that the reflection is common to both grains. The image of the twist GB in Figure 26.23 was formed using such a ‘reflection’ (i.e., two nearby reflections). (Look back at Figure 17.7 to get an idea of what the DP would be like.) What we are really doing is treating the dislocations as if they were isolated lattice defects; actually, the \mathbf{g} vectors for the two grains will be rotated relative to one another.

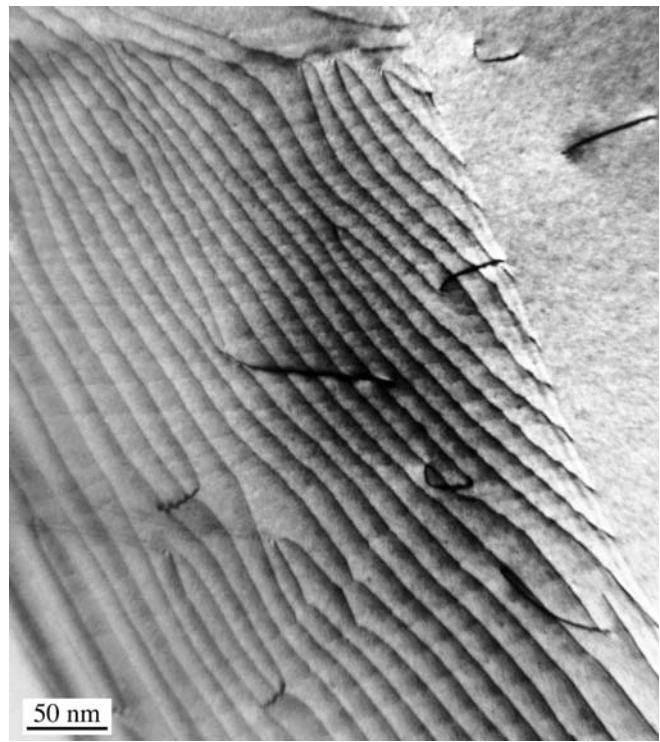


FIGURE 26.20. Dislocations interacting with a GB; the dislocation contrast changes because its strain field changes when it enters the boundary and becomes part of the dislocation structure.

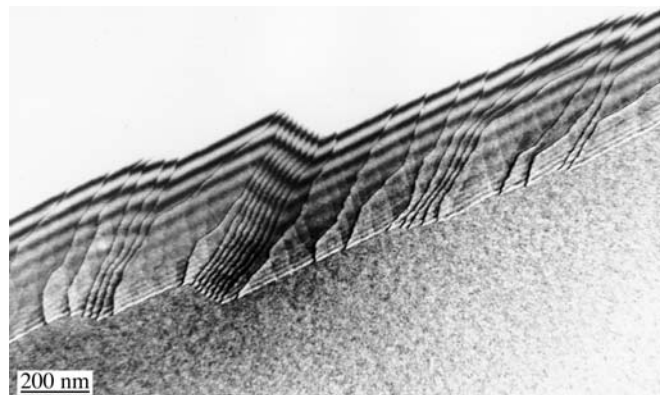
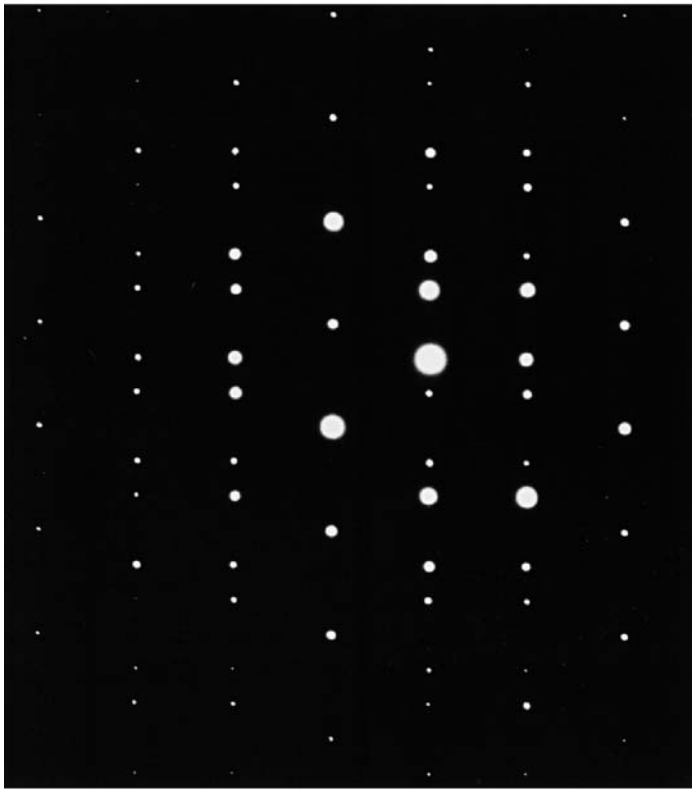


FIGURE 26.21. Steps at interfaces may also cause diffraction contrast when associated with strain. In this Ge specimen, the steps displace the thickness fringes in the GB so they are readily visible. The fringe spacing is different at the top and bottom of the boundary because the diffraction conditions are different at each grain.

Lattice misfit is very important whenever we are studying thin films; dislocations are often present to accommodate misfit. An example is shown in Figure 26.24, where dislocations are present between spinel and NiO; these two materials each have an fcc crystal structure. Although you can easily appreciate the change in lattice parameter, there is also a less obvious change in the elastic constants.

(A)



(B)

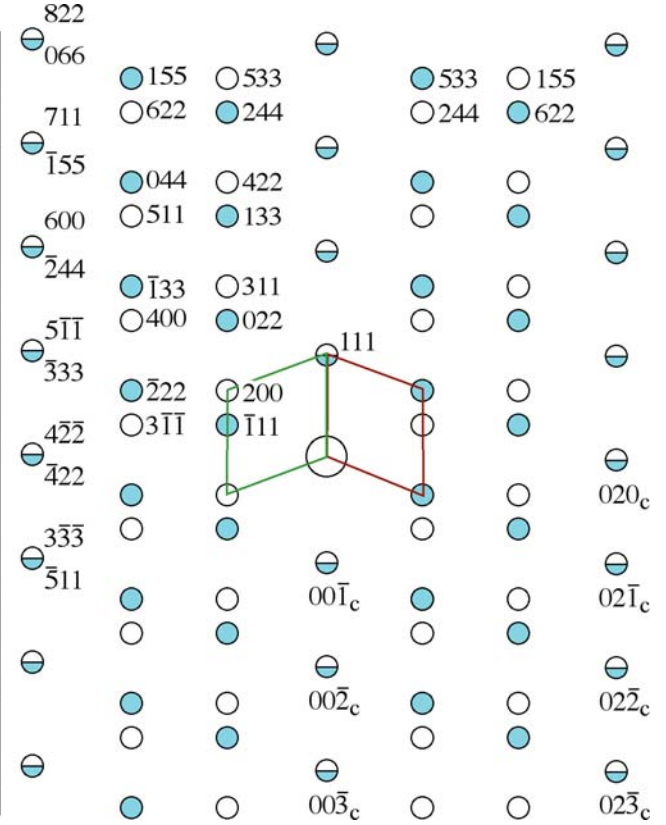


FIGURE 26.22. (A) DP and (B) its indexed schematic for a $\Sigma = 3$ twin boundary in an fcc material. Notice that many pairs of g -vectors exactly overlap but have very different indices.

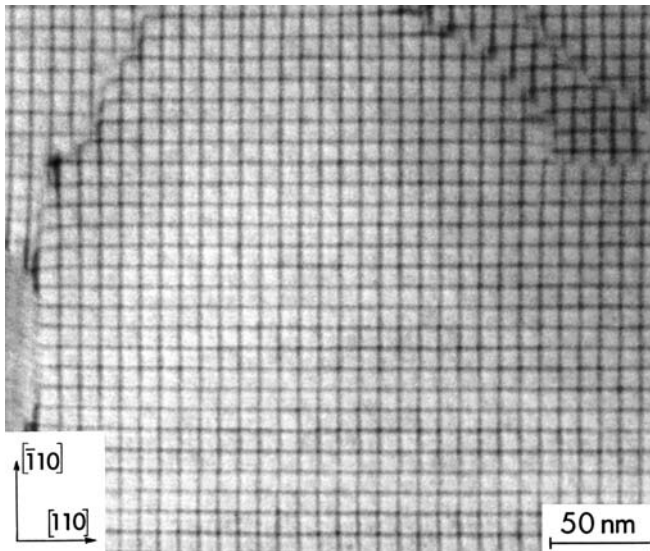


FIGURE 26.23. A low-angle (001) twist boundary in Si oriented almost exactly parallel to the specimen surface. Two (040) reflections were excited to form this BF image, but for small misorientations these are so close that we treat them as one reflection.

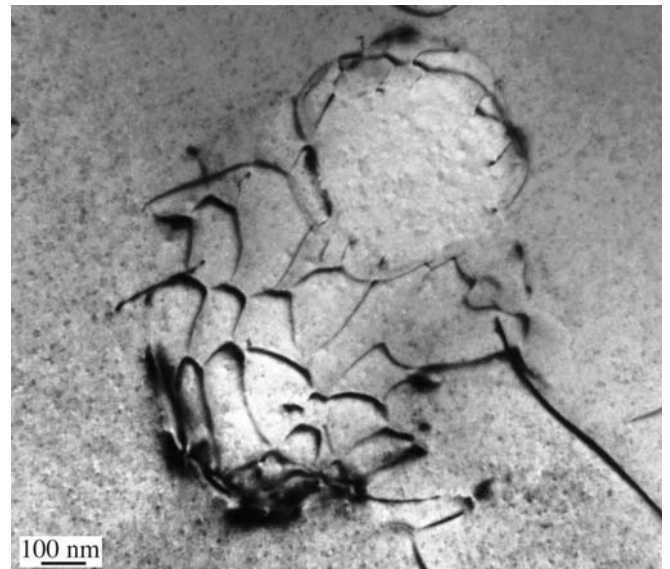


FIGURE 26.24. An irregular array of misfit dislocations at the interface between a spinel particle and a NiO matrix. The lattice mismatch is very small as you can appreciate from the scale. Although you can ‘see’ a distorted hexagonal array of dislocations, you have to remember that this interface is actually curving within the specimen so that we are only seeing a projection of the structure.

The TEM beam ‘sees’ yet another change: the extinction distance is different. The result is that, if the crystal is inclined to the electron beam, you will see thickness fringes associated with the interface. Not much work has been done on this, but you may find that it is more difficult to use the $\mathbf{g}\cdot\mathbf{b}$ criterion for determining Burgers vectors, especially when the misfit is large.

GB VERSUS PB

This means that the strain field at phase boundaries is *not* the same as at a GB.

Phase transformations often involve the movement of dislocations, generally at semicoherent interfaces. All the conditions discussed above may hold; however, now the dislocations will certainly be associated with a step on the interface to allow the transformation to proceed, so as to physically translate the interface. However, you’ll find it difficult to model the contrast from such dislocations, especially when you have a thin layer of the new phase enclosed by the matrix, as is the case when a precipitate grows, as illustrated in Figure 26.25.

The main effect of steps on such interfaces is that they cause a shift in the thickness fringes. It is often difficult to tell if there is also a dislocation present.

We’ll summarize some features you should remember when studying dislocations in interfaces

- If the orientation of the grains is different, the distribution of strain from the dislocation may be different in the two grains; the diffraction contrast is determined by this strain field.

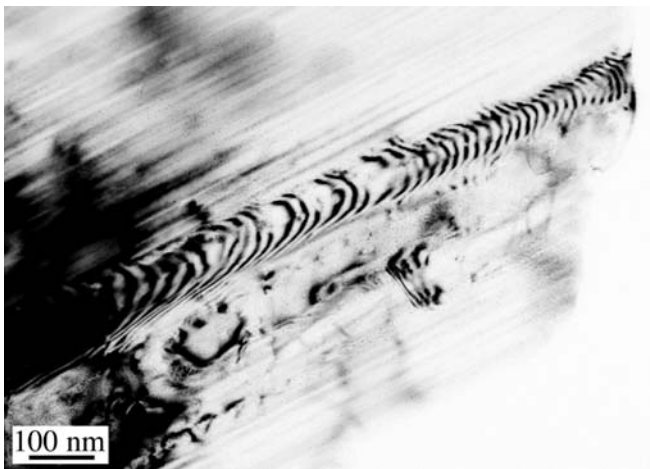


FIGURE 26.25. Transformation dislocations in the interface between a growing lath of hematite (pseudo-hexagonal alumina structure) in a ferrite (cubic spinel structure) matrix. The dislocations are curved because they were moving while heating the thinned specimen, which is why we know they are transformation dislocations, not simply misfit dislocations.

- If the chemistry of the two grains is different or if you use different but equal \mathbf{g} vectors, the extinction distances will be different and the image of the dislocations must therefore be affected.
- Be careful not to confuse moiré fringes with dislocations (we discussed moiré fringes in Chapter 23). The guide is that the dark and light moiré fringes have approximately equal widths; if there is any ambiguity, you should use weak-beam imaging (Chapter 27) and carefully examine the DP.

Humble and Forwood have shown using computer simulation of dislocations in interfaces that it is best to use diffraction conditions where a reflection is satisfied in both grains, otherwise the dislocation images tend to be rather featureless relative to the interface thickness fringes.

26.10 VOLUME DEFECTS AND PARTICLES

When the defects are small, the image may be dominated by the strain-field contrast; that is the aspect we are considering here. You have to remember, though, that these defects may have a different structure, lattice parameter, and composition. The theory for a spherical particle in a matrix was given over 40 years ago and is known as Ashby-Brown contrast. The theory works well for coherent particles but as soon as the first interface dislocation forms, analysis becomes much more difficult.

Lattice-strain effects around spherical precipitates appear as lobes of low intensity with a line of no contrast perpendicular to \mathbf{g} , as shown schematically in Figure 26.26. If you measure the size of the precipitates from a DF image and the size of the strain-contrast lobes in BF, you can get a direct measure of the lattice strain surrounding a single precipitate, which is really quite remarkable. The process requires some specific experimental conditions and careful calibration of the image recording procedure. It is quite a specialized study, so you should read the original references for details. If your precipitates aren’t spherical, intuitive interpretation of the images is unreliable and you have to resort to computer simulation.

Figure 26.26B shows how a spherical particle might strain the lattice. Notice that in this case, all the planes continue across the particle so it is coherent and there are no misfit dislocations. The figure here assumes that all the strain occurs in the matrix, which is only true for a hard particle in a soft matrix. The displacement field used to model this situation is

$$\mathbf{R} = C_{\epsilon}\mathbf{r} \quad (26.14)$$

when $\mathbf{r} \leq \mathbf{r}_0$, and

$$\mathbf{R} = C_{\epsilon}\frac{r_0^3}{r^3}\mathbf{r} \quad (26.15)$$

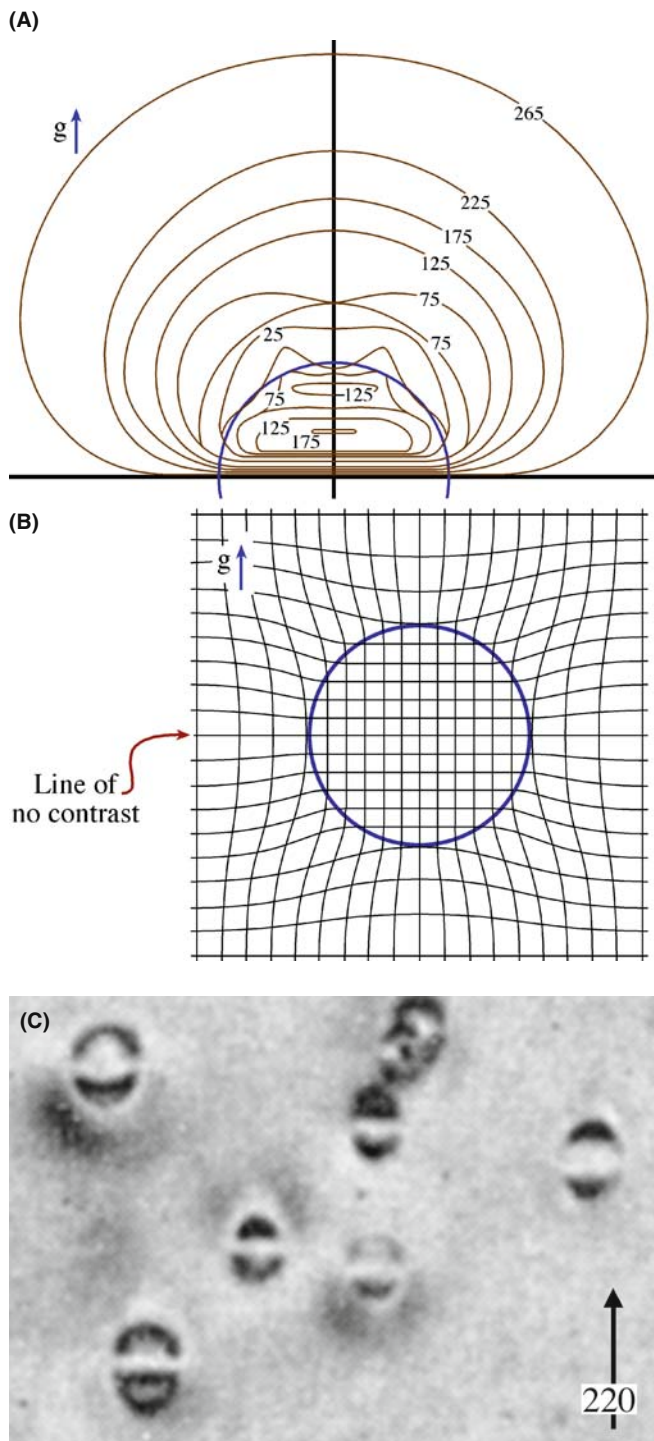


FIGURE 26.26. (A) Intensity contours from a simulated image of a particle like that shown schematically in (B). Notice the line of no contrast which corresponds to the plane which is not distorted by the strain field of the particle. (C) Experimental image of coherent particles in Cu-Co showing strain contrast and a line of no contrast as predicted.

when $r \geq r_0$. C_ϵ is an expression for the elastic constants, given by

$$C_\epsilon = \frac{3K\delta}{3K + 2E(1 + \nu)} \quad (26.16)$$

K is the bulk modulus of the precipitate; E and ν are Young's modulus and Poisson's ratio, respectively, for the matrix. The important feature is that \mathbf{R} always has radial symmetry. Thus, when we consider the Howie-Whelan equations, we realize that when $\mathbf{g} \cdot \mathbf{R} = 0$ we will see no contrast. So, there will be a 'line of no contrast' normal to \mathbf{g} .

The strain can be plotted using the equations given by Ashby and Brown and the image simulated (see below), as shown in Figure 26.26A. In the image from a specimen of a Cu-Co alloy containing small Co precipitates shown in Figure 26.26C, we can see that the images of the particles resemble butterflies or coffee beans. With the improvement in computers, the image contrast expected from much more complex particle geometries can now be calculated and can even consider statistical structural fluctuations.

26.11 SIMULATING IMAGES

It is important that you understand the origins of diffraction contrast from strain fields before you try to simulate this contrast using a computer. Having said that, few students would want to calculate image intensities by hand. The Howie-Whelan equations can be used to simulate images of dislocations, which is especially important when the dislocations are close together. The principal approaches used to simulate diffraction-contrast images were discussed in Sections 25.11–25.13.

Although the algorithm employed by the Head et al. programs allows very fast computation of the image, it does so by restricting the geometry of the defects. To cope with more general geometries, e.g., the strain field from end-on screw dislocations or non-parallel dislocations, we can use Thölén's matrix algorithm. As we saw in Chapter 25, the Howie-Basinski treatment extends the two-beam calculations to include several beams on the systematic row and provides a method for circumventing the column approximation.

PHOTOGRAPHIC FILM

If you want to make quantitative comparisons with real images on film, you must correct for the non-linearity of the film (see Chapters 30 and 31).

26.11.A The Defect Geometry

When choosing the optimal simulation method, depending on the defect geometry, the problem of calculating the image belongs to one of three categories

- *Two-dimensional problem*: including the most general geometries where integration of the full two-dimensional (x,y) grid is necessary.
- *One-dimensional problem*: geometries where the image depends only on either x or y and can be represented by a profile, e.g., problems involving a dislocation parallel to the foil surfaces.
- *GCS problem*: geometries where the method of generalized cross sections (GCS), developed by Head et al., can be applied. Situations where the dislocations and fault planes are parallel to each other, but inclined to the foil surface, are included in this group.

Choosing the best method can speed up the simulations considerably, as we'll show later. The Head et al. program automatically determines the category and selects the appropriate calculation method.

26.11.B Crystal Defects and Calculating the Displacement Field

The program Comis could simulate amplitude contrast from any number of defects consisting of fault planes and straight, infinite dislocations. You just needed to define the Burgers vector, line direction, and relative position; planar faults were defined by the plane normal, the displacement vector, and the relative position. You could then predefine certain standard geometries to ease the process of defining the defect system.

Once you've defined the defect geometry, you must consider the region of the crystal you want to simulate. In situations where the 'interesting' region is well defined (as in the case of inclined dislocations or intersecting dislocations), Comis determined this region and provided it as the default. However, you could always set the image region manually in Angstrom units, to obtain a desired magnification.

The displacement field for the dislocations was calculated in Comis using linear, anisotropic elasticity theory and was based on the algorithms of Head et al., so you had to specify the elastic constants of your crystal. The displacement field then corresponds to straight, infinite dislocations in an infinite medium with no account taken of surface relaxations. In any simulation, you should be able to introduce image dislocations outside the crystal in order to include surface effects.

26.11.C The Parameters

An example which shows simulated images of an orthogonal network of screw dislocations is given in Figure 26.27. Comis could calculate the equilibrium

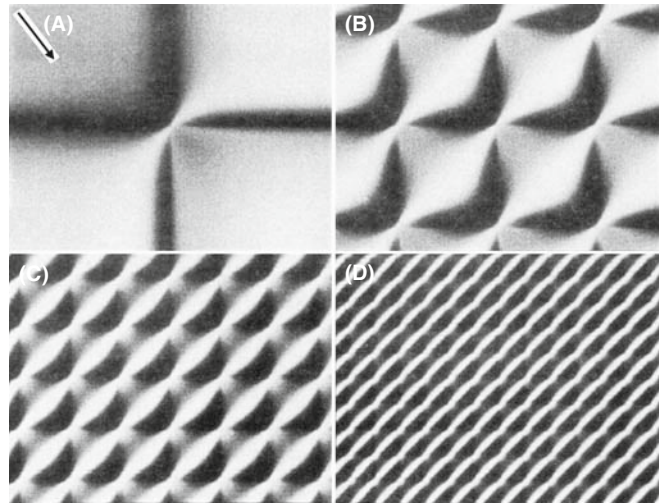


FIGURE 26.27. Comis simulations of two-beam BF images of networks of screw dislocations, located in the middle of a foil with thickness equal to four times the extinction distance, $\xi_{\mathbf{g}}$; $\mathbf{g} \cdot \mathbf{b} = 1$ for both dislocation types. The separation between the dislocations is (A) ∞ , (B) $1\xi_{\mathbf{g}}$, (C) $0.5\xi_{\mathbf{g}}$, (D) $0.25\xi_{\mathbf{g}}$.

configuration of certain types of interacting dislocations using anisotropic elasticity theory, and then directly incorporates the resulting geometry in subsequent image simulations. As you can appreciate from equation 26.5, in such simulation studies, you will need *all* the parameters for the defects, the specimen, and the diffraction conditions

- The foil thickness.
- The stacking-fault energy.
- The absorption parameters, usually using $|U'_{\mathbf{g}}|/|U_{\mathbf{g}}| = 0.1$.
- The number of beams included in the calculation.
- The zone-axis and the diffracting vectors.
- Also required are the electron energy, the elastic constants, the normal to the foil surface, the Burgers vectors, and the line direction of the dislocations.

The exact beam direction can then be specified by defining the 'center' of the Laue zone, giving the coordinates in terms of the \mathbf{g} vector and \mathbf{g}_z . Here \mathbf{g}_z is a specially defined vector in reciprocal space, which is automatically set to lie in the ZOLZ and to be perpendicular to \mathbf{g} . Thus, if you place the center of the ZOLZ at $(0,0)$, the specimen is oriented on the zone-axis; if you place the center of the ZOLZ at $(0.5,0)$, it corresponds to being at $\mathbf{g}/2$, i.e., at the Bragg position with the $\mathbf{0}$ and \mathbf{g} beams excited. If you change the second coordinate to be, say, $(0.5,0.5)$, you would include beams from off the systematic row.

CHAPTER SUMMARY

The central idea of this chapter is that the strain field moves atoms off their perfect-crystal positions. We've concentrated on dislocations because the edge dislocation gives the clearest illustration of how the deformation produces the contrast and its structure can be understood with a two-dimensional projection. We can summarize the topics of the chapter as follows

- There is a new feature to the column approximation. The displacement moves atoms out of the column and brings others into the column.
- The basis of the $\mathbf{g}\cdot\mathbf{b}$ analysis of a dislocation is simply that the contrast is determined by $\mathbf{g}\cdot\mathbf{R}(\mathbf{r})$ and that $\mathbf{R}(\mathbf{r})$ is linearly related to \mathbf{b} . For the screw dislocation, $\mathbf{R}(\mathbf{r})$ is directly proportional to \mathbf{b} . For the edge dislocation, the image can also be affected by a $\mathbf{g}\cdot\mathbf{b} \times \mathbf{u}$ component which is caused by the buckling of the dislocation glide plane.
- Dislocation images are usually asymmetric. The contrast depends on the sign of $(\mathbf{g}\cdot\mathbf{b})s$.
- As a practical rule, we usually set s to be >0 . Then the distortion due to the defect will bend the near-diffracting planes back into the Bragg-diffracting condition to give strong contrast. When $s >0$, detail in the image is more localized relative to the defect than if we use the $s = 0$ condition.

There are many other situations which are closely related to the topics we've discussed in this chapter. For example, we have not discussed strain contrast associated with crack tips or the analysis of buckling of thin specimens. Although these are rather specialized situations, they do illustrate the growing applications of diffraction contrast in the TEM.

REFERENCES

The text by Hirsch et al. summarizes the early work by the Cambridge group including the derivation of equation 26.10.

BACKGROUND TO DISLOCATIONS AND INTERFACES

- Amelinckx, S 1964 *The Direct Observation of Dislocations*, Academic Press New York. A fascinating summary of the early studies by TEM by the man who influenced many. Source of Figure 26.16.
- Amelinckx, S 1979 in *Dislocations in Solids 2*, (Ed. FRN Nabarro), North-Holland New York. If you're interested in dislocations, there are many other volumes in this set.
- Eshelby, JD, Read WT and Shockley W 1953 *Anisotropic Elasticity with Applications to Dislocation Theory* Acta Metall. **1**, 251–9. An early paper on anisotropic elasticity theory (used by Comis).
- Hirth, JP and Lothe, J 1982 *Theory of Dislocations*, 2nd Ed., John Wiley & Sons New York. The definitive textbook, but not for the beginner. Source for equation 26.8, buckling of the glide plane, etc. Not a TEM book.
- Hull, D. and Bacon, DJ 2001 *Introduction to Dislocations*, 4th Ed., Pergamon Press New York. A great introductory text.
- Matthews, JW, Ed. 1975 *Epitaxial Growth*, Parts A and B, Academic Press New York. Our hero. Unfortunate grammar in the title.
- Nabarro, FRN 1987 *Theory of Dislocations*, Dover Publications New York.
- Porter, DA and Easterling, KE 1992 *Phase Transformations in Metals and Alloys*, 2nd Ed., Chapman and Hall New York.
- Smallman, RE 1985 *Modern Physical Metallurgy*, 4th Ed., Butterworth-Heinemann Boston.
- Steeds, JW 1973 *Anisotropic Elastic Theory of Dislocations*, Clarendon Press Oxford, UK. As readable as this subject can be: written by a microscopist. See also Eshelby et al.
- Sutton, AP and Balluffi, RW 1995 *Interfaces in Crystalline Materials*, Oxford University Press New York.
- Wolf, D and Yip, S, Eds. 1992 *Materials Interfaces, Atomic-level Structure and Properties*. Chapman and Hall New York. A collection of review articles.

IMAGE SIMULATION

- Head, AK, Humble P, Clarebrough LM, Morton AJ and Forwood, CT 1973 *Computed Electron Micrographs and Defect Identification* North-Holland New York.
- Morton, AJ and Forwood CT 1973 *Equilibria of Extended Dislocations* Cryst. Lattice Defects **4** 165–177. TEM of arrays of interacting dislocations (Section 26.11.C).

- Humble, P and Forwood, CT 1975 *Identification of Grain Boundary Dislocations I and II* Phil. Mag. **31**, 1011–23 and 1025–48. Simulating images of interfaces.
- Rasmussen, DR and Carter, CB 1991 *A Computer Program for Many-Beam Image Simulation of Amplitude-Contrast Images* J. Electron Microsc. Techniques **18**, 429. Description of Comis, a great program that died with the advances in operating systems.
- Thölén, AR 1970a *Rapid Method for Obtaining Electron Microscope Contrast Maps of Various Lattice Defects* Phil. Mag. **22** 175–182. Matrix algorithm for more complex geometries.
- Thölén, AR 1970b *On the Ambiguity between Moiré Fringes and the Electron Diffraction Contrast from Closely Spaced dislocations* Phys. stat. sol. (a) **2** 537–550. Applying the algorithm to a network of orthogonal dislocations (Section 26.11.C).
- Thölén, AR and Taftø, J 1993 *Periodic Buckling of the Lattice Planes in the Thin Regions of Wedge-Shaped Crystals* Ultramicrosc. **48** 27–35. TEM of buckled specimens: challenging exercise.

CONTRAST THEORY

- Amelinckx, S 1992 in *Electron Microscopy in Materials Science*, (Eds PG Merli and MV Antisari), World Scientific River Edge NJ. Includes discussion of dislocation strain field relaxing at the surface.
- Amelinckx, S and Van Dyck, D 1992 in *Electron Diffraction Techniques 2* (Ed. J.M. Cowley), p. 1, Oxford University Press New York.
- Ashby, MF and Brown, LM 1963 *Diffraction Contrast from Spherically Symmetrical Coherency Strain* Phil. Mag. **8** 1083–1103 and *On Diffraction Contrast from Inclusions* Phil. Mag. **8** 1649–1676. Ashby-Brown contrast.
- de Graef, M and Clarke, DR 1993 *Strain Contrast at Crack Tips for in-situ Transmission Electron Microscopy Straining Experiments* Ultramicrosc. **49**, 354–365. TEM of crack tips: challenging exercise.
- Goringe, MJ 1975 in *Electron Microscopy in Materials Science* (Eds. U Valdré and E Ruedl), p. 555, Commission of the European Communities Luxembourg. Further discussion of equation 26.10 and much more.
- Hirsch, PB, Howie, A and Whelan, MJ 1960 *Kinematical Theory of Diffraction Contrast of Electron Transmission Microscope Images of Dislocations and Other Defects* Phil Trans Roy Soc. **A252** 499–529. Early paper you should read, with more on the column approximation.
- The original series of papers by H Hashimoto, PB Hirsch, A Howie, MJ Whelan in Proc. Roy. Soc. London A **252** 499 (1960), **263** 217 (1960), **267** 206 (1962), and **268** 80 (1962) are strongly recommended.

ON THE NON-COLUMN APPROXIMATION

- Howie, A and Basinski ZS 1968 *Approximations of the Dynamical Theory of Diffraction Contrast* Phil. Mag. **17** 1039–1063.
- Howie, A and Sworn H 1970 *Column Approximation Effects in High Resolution Electron Microscopy using Weak Diffracted Beams* Phil. Mag. **31** 861–864.

EXAMPLES

- Hughes, DA and Hansen, N 1995 *High Angle Boundaries and Orientation Distributions at Large Strains* Scripta Met. Mater. **33** 315–321. Particularly clear illustration of the value of producing a large specimen area.
- Karth, S, Krumhansl, JA, Sethna, JP and Wickham, LK 1995 *Disorder-Driven Pretransitional Tweed Pattern in Martensitic Transformations* Phys. Rev. B **52** 803–822. Early work considering statistical structural fluctuations on image contrast (in Section 26.10).
- Takayanagi, K, Tanishiro, Y, Yagi, K, Kobayashi, K and Honjo, G 1988 *UHV-TEM Study on the Reconstructed Surface of Au(111): Metastable $p'' \times p''$ and Stable $p \times 1$ Surface Structure* Surf. Sci. **205** 637–651. They showed surface dislocations by TEM before they were discovered by STM.
- Tunstall, WJ, Hirsch, PB and Steeds, JW 1964 *Effects of Surface Stress Relaxation on Electron Microscope Images of Dislocations Normal to Thin Metal Foils* Phil. Mag. **9** 99–119. Classic paper on the contrast seen when a screw dislocation intersects a surface.
- Wilkens, M 1978 in *Diffraction and Imaging Techniques in Materials Science*, 2nd Ed. (Eds S Amelinckx, R Gevers and J Van Landuyt) p.185 North-Holland New York. A detailed analysis of butterflies, lozenges and peanuts (Figure 26.12).

THE COMPANION TEXT

The programs TEMACI and Cufours are available for simulating these defects. TEMACI was designed for the study of small defects.

SELF-ASSESSMENT QUESTIONS

- Q26.1 Give the physical reason for the change in contrast near a dislocation.
- Q26.2 We are given an assignment to perform a computer simulation of contrast due to defects. The assignment includes areas of a hypothetical specimen with dimensions of about 1 nm. Is the column approximation valid?
- Q26.3 A screw dislocation in your specimen appears to be out-of-contrast. Does this mean $\mathbf{g}\cdot\mathbf{R}=0$?
- Q26.4 An edge dislocation has $\mathbf{g}\cdot\mathbf{b} = 0$ but it still shows strong contrast. Why?
- Q26.5 Is the image of a dislocation located directly at the projected position of the dislocation?
- Q26.6 Is the information in the image of dislocation core located coming from the reflection \mathbf{g} ?
- Q26.7 You are looking at a region in the image that is about 1 μm in diameter. Could this be the strain field from a defect?
- Q26.8 You see abrupt steps in the thickness fringes at different locations in the specimen. What might cause this?
- Q26.9 How can you 'see' a dislocation and why do we use quotes in this question?
- Q26.10 What parameters are needed to describe a particular dislocation?
- Q26.11 How do we determine the Burgers vector of a dislocation?
- Q26.12 How is the Burgers vector determined for dislocation nodes or networks?
- Q26.13 What complications arise for small dislocation loops?
- Q26.14 Why do dislocation dipoles show 'inside-outside' contrast?
- Q26.15 What happens to $\mathbf{g}\cdot\mathbf{b}$ dislocation contrast when \mathbf{g} is reversed?
- Q26.16 How does using a higher kV help in 'viewing' dislocations?
- Q26.17 How does using a lower kV help in 'viewing' dislocations?
- Q26.18 Give two ways that the surface can affect the image of a dislocation.
- Q26.19 Why do we talk about $\mathbf{g}\cdot\mathbf{b}$ rather than $\mathbf{g}\cdot\mathbf{R}$?
- Q26.20 How does the term $\mathbf{g}\cdot\mathbf{b}\times\mathbf{u}$ relate to $\mathbf{g}\cdot\mathbf{R}$?

TEXT-SPECIFIC QUESTIONS

- T26.1 Consider Figure 26.1A. How wide do you expect the dislocation image (the peak in 'B') to be? Just give your thought process without doing displacement-field calculations, etc.
- T26.2 Consider Figure 26.3. Why do we always like to have \mathbf{b} lying in the plane of the specimen? (Give at least three different valid reasons.)
- T26.3 Figure 26.5 suggests that if we keep $\mathbf{g}\cdot\mathbf{b}$ the same but reverse the sign of s , the image of the dislocation will arise from the other side of the dislocation core. Use your analysis of the Howie-Whelan equations to explain this result.
- T26.4 Give a physical explanation of equation 26.12 and discuss how it differs from using \mathbf{R} .
- T26.5 Consider Figure 26.6A. Give a consistent set of \mathbf{b} vectors for all the dislocations you see.
- T26.6 Consider Figure 26.6C. By consulting the original source or otherwise, explain the variation in the separation of the dislocation images.
- T26.7 Consider the lower half of Figure 26.6G,H. Explain why all the dislocations are visible in the two \mathbf{g} vectors.
- T26.8 Consider Figure 26.8. Suggest a set of consistent \mathbf{g} vectors for the four images.
- T26.9 Consider Figure 26.10. Why are the fringes on one loop running in a different direction? Suggest what \mathbf{g} might be and include it on the diagram.
- T26.10 Consider Figure 26.11. Explain how $\mathbf{g}\cdot\mathbf{b}\times\mathbf{u}$ can lie in the plane of the loop and thus confirm the interpretation of the contrast in this image.
- T26.11 Consider Figure 26.16A and B. Estimate the angle through which the specimens bend.
- T26.12 How can we increase the information on dislocations parallel to the specimen surface if dislocations with values of $\mathbf{g}\cdot\mathbf{b}$ which are <0.3 can't be distinguished from those with $\mathbf{g}\cdot\mathbf{b} = 0$?

Weak-Beam Dark-Field Microscopy

CHAPTER PREVIEW

The term ‘weak-beam microscopy’ refers to the formation of a diffraction-contrast image in either BF or DF where the useful information is transferred by weakly excited beams. The DF approach has been more widely used, in part because it can be understood using quite simple physical models. It also gives stronger contrast; we see white lines on a dark gray background. This chapter will be concerned only with the DF approach. Historically, the weak-beam dark-field (WBDF often abbreviated to WB) method became important because under certain special diffraction conditions, dislocations can be imaged as narrow lines which are approximately 1.5 nm wide. Equally important is the fact that the positions of these lines are well defined with respect to the dislocation cores; they are also relatively insensitive to both the foil thickness and the position of the dislocations in the specimen. The technique is particularly useful if you are studying dissociated dislocations where pairs of partial dislocations may only be ~4 nm apart and yet this separation greatly affects the properties of the material.

We first choose a particular \mathbf{g} and bring this onto the optic axis as if intending to form a regular on-axis DF image. We then tilt the specimen to make \mathbf{s}_g large and examine the DF image using reflection \mathbf{g} . If a defect is present, the diffracting planes may be bent locally back into the Bragg-diffracting orientation to give more intensity in the DF image. The problem is that, as we increase \mathbf{s}_g , the average intensity decreases as $1/s^2$; in the DP the beam appears as a weak spot, hence the name. When \mathbf{s}_g is large, the coupling between \mathbf{g} and the direct beam becomes small and the diffracted beam is said to be ‘kinematically diffracted.’ So, this chapter is where we will discuss the ‘kinematical approximation’.

- You will sometimes see reference to the $\mathbf{g}(3\mathbf{g})$ WB condition. Beware! Sometimes you don’t need to be this weak; sometimes this is not weak enough.
- It is often not the fact that \mathbf{s} is large that is important; what is important is that ξ_{eff} is small.

This chapter is unusual in that it deals with a special imaging technique, rather than a concept or theory. Also, WBDF is only really useful when the specimen is not perfect, i.e., when you are interested in defects in the specimen or small changes in thickness. Therefore, you can skip this chapter if crystal-lattice defects are not relevant to your microscopy study. If you are interested in defects, you will find that this chapter really covers much more than WB microscopy. For example, we will use concepts developed for diffracted beams and carefully set the excitation error, \mathbf{s}_g , by referring to the Kikuchi-line pattern. In Section 27.9, we will discuss some of the ways that new developments in TEM design are changing the way we do WB microscopy and how we interpret the images.

27.1 INTENSITY IN WBDF IMAGES

$$|\phi_g|^2 = \left(\frac{\pi t}{\xi_g}\right)^2 \cdot \frac{\sin^2(\pi t s_{\text{eff}})}{(\pi t s_{\text{eff}})^2} \tag{27.1}$$

We showed in Chapter 13 that, in a two-beam situation, the intensity of the diffracted beam \mathbf{g} in a perfect crystal can be written as

Remember that when we derived this expression we assumed that only two beams, O and G, are important.

We will consider complications which arise when more beams are present in Section 27.9. The important variables in equation 27.1 are the thickness, t , and the effective excitation error, s_{eff} , which was given by equation 13.47

$$s_{\text{eff}} = \sqrt{s^2 + \frac{1}{\xi_g^2}} \quad (27.2)$$

In the WB technique we increase s to about 0.2 nm^{-1} so as to increase s_{eff} . (In most WB papers you will see this value as $2 \times 10^{-2} \text{ \AA}^{-1}$; remember 50 \AA is 5 nm .) This large value of s means that s_{eff} , and therefore the intensity, I_g , becomes independent of ξ_g except as a scaling factor for t (in the prefactor in equation 27.1). The actual value of s can be set by carefully positioning the Kikuchi lines for the systematic row of reflections, which includes \mathbf{g} . You can best appreciate this effect by calculating a range of values for s . Remember when doing this that you must specify \mathbf{g} and the kV because, as we saw in Chapter 13, ξ_g varies with both the reflection used to form the WB image and the energy of the electrons, as implied in equation 13.47

$$\xi_{\text{eff}} = \frac{\xi_g}{\sqrt{w^2 + 1}} \quad (27.3)$$

Practical Considerations: As the value of s_{eff} increases, equation 27.1 shows that the intensity of the \mathbf{G} beam decreases very rapidly. The result is that the exposure time needed to record the image on the CCD or on photographic film also rapidly increases and has, in the past, been the factor which limited the usefulness of the technique. Although manufacturers may guarantee a drift rate of less than 0.5 nm per minute for new machines, values of six times this rate are common on many older instruments. In early WB studies, this problem was partly overcome by using photographic film with a fast emulsion or by modifying photographic processing conditions. In either case, the grainy appearance of the photographic emulsion would be increased. The problem of drift can, in principle, be overcome by using a video system to record the image and capturing frames from the video. You could then, in principle, reduce the noise by frame averaging, particularly if you can take account of any drift. The causes of drift (specimen and thermal effects) and their correction or minimization are discussed in Chapter 8, but it is worth reminding you that changes in the temperature of the water in the objective lens are a major cause of drift. Although WBDF imaging may aim for 0.5 nm resolution rather than 0.2 nm in HRTEM, exposure times are often 10 times greater for WBDF than HRTEM, so drift may be even more important.

ON EQUATION 27.2

If $s \gg \xi_g^{-2}$ then $s \approx s_{\text{eff}}$ so that equation 27.2 becomes what is known as the ‘kinematical equation’; the kinematical equation cannot be applied for small s unless the thickness, t , is also very small.

27.2 SETTING s_g USING THE KIKUCHI PATTERN

Since the contrast in the WB image is so dependent on the value of s_g we need a method for determining s_g . We draw a line through the \mathbf{g} -systematic row and let the Ewald sphere cut through this line at $n\mathbf{g}$ where n is not an integer. Figure 27.1 illustrates this situation. How can we ‘see’ what the value of n is? Of course we can’t, since we are looking approximately normal to the ZOLZ. We can’t just judge n by looking at the intensities of the spots except in special circumstances.

The solution to the problem can be appreciated by looking at the Kikuchi pattern shown in Figure 27.2. When \mathbf{g} is exactly at the Bragg condition, the \mathbf{g} Kikuchi line passes through the \mathbf{g} reflection; when $3\mathbf{g}$ is exactly satisfied, the $3\mathbf{g}$ Kikuchi line passes through the $3\mathbf{g}$ reflection. We might guess that n is ~ 3.2 in Figure 27.2, but we don’t have a $3.2\mathbf{g}$ Kikuchi line; we have to deduce this value of n from the position of the $3\mathbf{g}$ Kikuchi line. Remember (from Chapter 19) that when the $3\mathbf{g}$ Kikuchi line passes through $3.5\mathbf{g}$ on the \mathbf{g} -systematic row, the $4\mathbf{g}$ Kikuchi line and the Ewald sphere pass through the $4\mathbf{g}$ reflection as shown in Figure 27.3. Therefore, when the Ewald sphere passes through $3.2\mathbf{g}$,

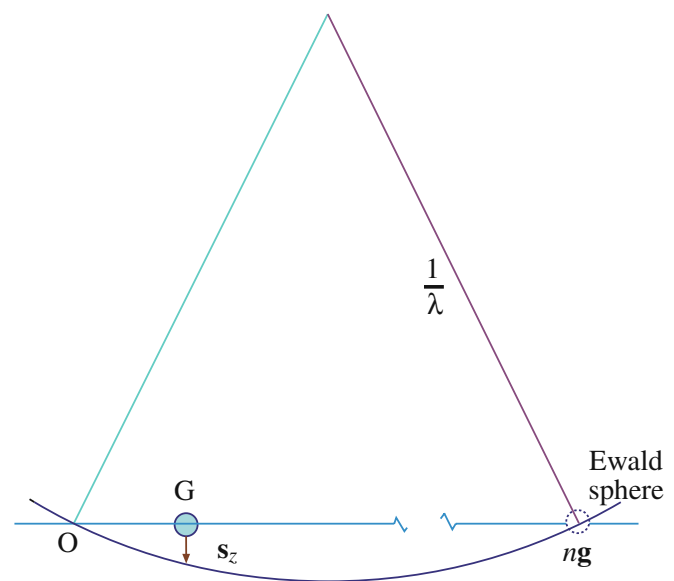


FIGURE 27.1. The Ewald-sphere construction showing the diffraction conditions used to obtain weak-beam images. The sphere cuts the row of systematic reflections at ‘ $n\mathbf{g}$ ’ where n is not necessarily an integer.

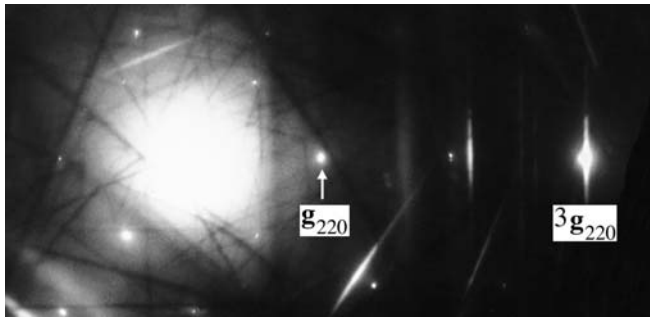


FIGURE 27.2. DP obtained when the specimen is tilted to a suitable orientation for WB microscopy. Here g is a 220 reflection and $3g$ is strong.

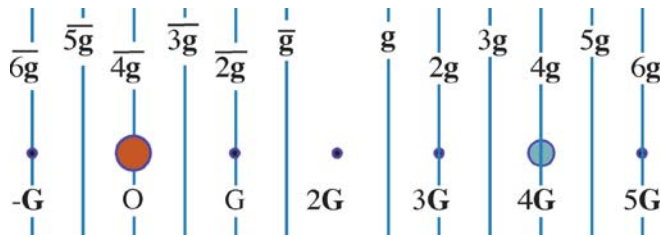


FIGURE 27.3. A schematic diagram showing the positions of the Kikuchi lines for the systematic row of reflections when $4g$ is excited.

the $3g$ Kikuchi line will pass through $3.1g$; we can express this simple geometric result as

$$n = 2m - N \quad (27.4)$$

where Ng refers to the Kikuchi line closest to ng (N is an integer) and mg is the location of the Kikuchi line as we measure it. In the example above, we can choose N to be 3 so that if m is 3.1 then n is 3.2; if, instead, we choose N to be 4, then m is 3.6 (because we measure the position of the $4g$ Kikuchi line) and n is still 3.2. Having determined n we need to estimate s . This we do using the expression

$$s = \frac{1}{2}(n - 1)|g|^2\lambda \quad (27.5)$$

which you can derive from Figure 27.4 using the intersection chord theorem ($ab = cd$) and the fact that $1/\lambda$ is much larger than s .

You can immediately appreciate some important results from this expression

- Setting $n = -1$ gives the same value of s as for $n = 3$ but the sign is reversed.
- The magnitude of s is more strongly dependent on $|g|$ than on λ but it depends on *both*.
- The specific nature of the material enters through g , the microscope affects s through λ .

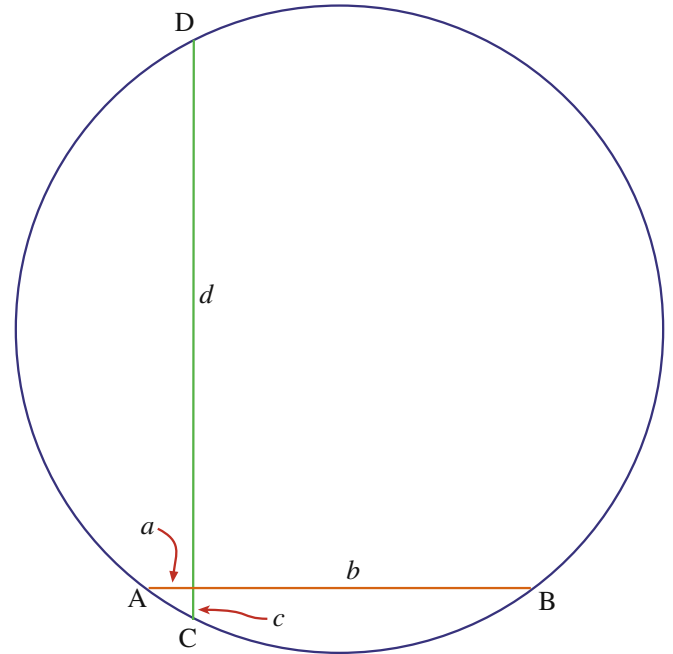


FIGURE 27.4. The intersection chord construction used to deduce the value of s_g : we approximate d to $2/\lambda$, c is s , a is $|g|$, and b is $(n-1)|g|$.

Here we recommend that you use a spreadsheet to calculate different values of s as you vary g or λ . A selection of these for Cu and Si is given in Table 27.1.

As an exercise, you can use equations 27.2 and 27.5 to calculate s_{eff} with $3g = 42 \text{ nm}$ for the 220 reflection in Cu at 100 kV. You can then see when s_{eff} becomes ‘independent’ of ξ_g . Next, repeat the exercise with other values of λ or for other reflections and materials.

One point you should bear in mind is that none of the above discussion requires a particular value for s and yet you will often read that s must be $\geq 0.2 \text{ nm}^{-1}$ for a WB image. This value of s is recommended when you are studying defects quantitatively, because computer calculations show that the position of the image can then be directly related to the position of the defect. You will find that smaller values of s will often give you WB images which contain the information you want and you can more easily see and record the image!

TABLE 27.1. Values of s (nm^{-1}) for Various kV					
Si	Cu	Accelerating voltage			
		100 kV	200 kV	300 kV	400 kV
n_{Si}	n_{Cu}				
4.9	2.8	0.20	0.14	0.11	0.09
6.9	3.6	0.30	0.20	0.16	0.13
8.5	4.3	0.38	0.25	0.20	0.17
9.9	4.9	0.44	0.30	0.24	0.20

For Cu, $a = 0.3607 \text{ nm}$; for Si, $a = 0.534 \text{ nm}$. The values of n are accurate to better than 0.1. In both cases the calculation is for $g = 220$.

27.3 HOW TO DO WBDF

The nature of the WB image imposes a restriction on the maximum specimen thickness which you can use because the visibility of such images decreases as the thickness increases (due to a corresponding increase in inelastic scattering). However, the orientation of the specimen is accurately set by reference to the Kikuchi lines observed in the DP (Chapter 19), and these are not visible in specimens which are too thin. Therefore, your specimen thickness must be greater than a certain minimum value. Also, if the observations you made on certain defects, in particular, dislocation ribbons and nodes, are to be interpreted as representative of the bulk and not influenced by the surface of the foil, then again the foil must not be too thin. You can generally satisfy these requirements by selecting defects for detailed study which, in the case of Cu and its alloys, lie in areas which are about 70 nm thick.

Due to the very low intensity of the WB images, exposure times required are typically on the order of 4–30 seconds using Kodak SO-163 film (CCD cameras are better!). The main factor limiting the exposure time is the inherent instability of any specimen stage. To minimize the exposure time, you can usually use a highly convergent (or divergent) beam, contrary to the assumption implicit in the simple theory. The effect of this convergence is that oscillations in image intensity and position, which result from variations in depth and thickness parameters, are diminished.

WB STEP-BY-STEP

We'll go through how to set up the $\mathbf{g}(3.1\mathbf{g})$ diffraction mode condition since this is widely used in practice. Actually, it is generally referred to as ' $\mathbf{g}(3\mathbf{g})$ with $\mathbf{s}_{3\mathbf{g}}$ positive' because we guess the value of n by estimating m . This condition ensures that the $3\mathbf{g}$ reflection is not satisfied and also that you can use the BF $\mathbf{0}(\mathbf{g})$ image, with $\mathbf{s}_{\mathbf{g}}$ slightly positive, to locate the defect and to focus your image.

The first two steps in setting up the WB conditions are illustrated in Figure 27.5, relating what happens in the Ewald-sphere model with what you see happening to the Kikuchi lines.

- Orient the specimen in BF so that \mathbf{g} is excited and $\mathbf{s}_{\mathbf{g}}$ is just greater than zero. Make sure that no other reflections are excited (see upper diagram in Figure 27.5).
- Use the DF beam-deflecting coils to bring the reflection \mathbf{g} on to the optic axis. Use the binoculars because \mathbf{g} becomes very weak; underfocus the beam before you use the high-resolution screen.

- Insert the objective aperture. In BF, check that the aperture is centered, then switch to DF and check that the spot \mathbf{G} is centered in the aperture.
- Fine-tune your conditions looking at the DP with \mathbf{G} centered.
- Go to imaging mode; you now have a WB image with the required $\mathbf{g}(3.1\mathbf{g})$ condition (see lower diagram in Figure 27.5).

Since you have inserted a small objective aperture, you should now check the objective astigmatism. We use a small objective aperture so as to remove inelastic scattering; remember that this aperture will then limit our potential resolution. If you focus the beam, you may change the position of the beam and probably the astigmatism! Remove the objective aperture and check that no other reflections are strongly excited when you are in DF. Then repeat the process starting at the third step (insert the objective aperture).

After finely focusing the image, record it together with its SADP.

If you're not sure why this 'trick' for setting the $\mathbf{g}(3.1\mathbf{g})$ condition works, go back to Chapter 19 and draw the systematic row and the corresponding Kikuchi lines. Then move the spots, while keeping the crystal, and thus the Kikuchi lines, fixed. You may see the conditions $\mathbf{g}(\bar{\mathbf{g}})$ used. This was the original condition suggested by Cockayne et al. (1969); it does give you the same values of s as the $\mathbf{g}(3.1\mathbf{g})$ condition, but the interband scattering processes are different and it is not so convenient to change from BF to WBDF. However, you may find variations on this condition useful when you need to use a $\mathbf{g}(n\mathbf{g})$ condition with a large value of n .

WB USING CDF

Note that we are using CDF imaging—we are not comparing contrast in BF and DF images.

Weak-beam microscopy becomes much easier when you are comfortable with using a CCD camera. Just being able to see whether the image is moving due to specimen drift can save boxes of photographic film. As we mentioned earlier, you could use frame averaging to reduce the noise. However, you will realize that the extra magnification from the video system tends to limit the area which can be viewed, so that film is still preferred by many WB practitioners but this is changing as higher resolution CCD cameras become affordable. You will find that $30\text{ k}\times$ magnification on the plate is a good compromise; without video, you will use $50\text{--}60\text{ k}\times$ to allow you to focus with the binoculars.

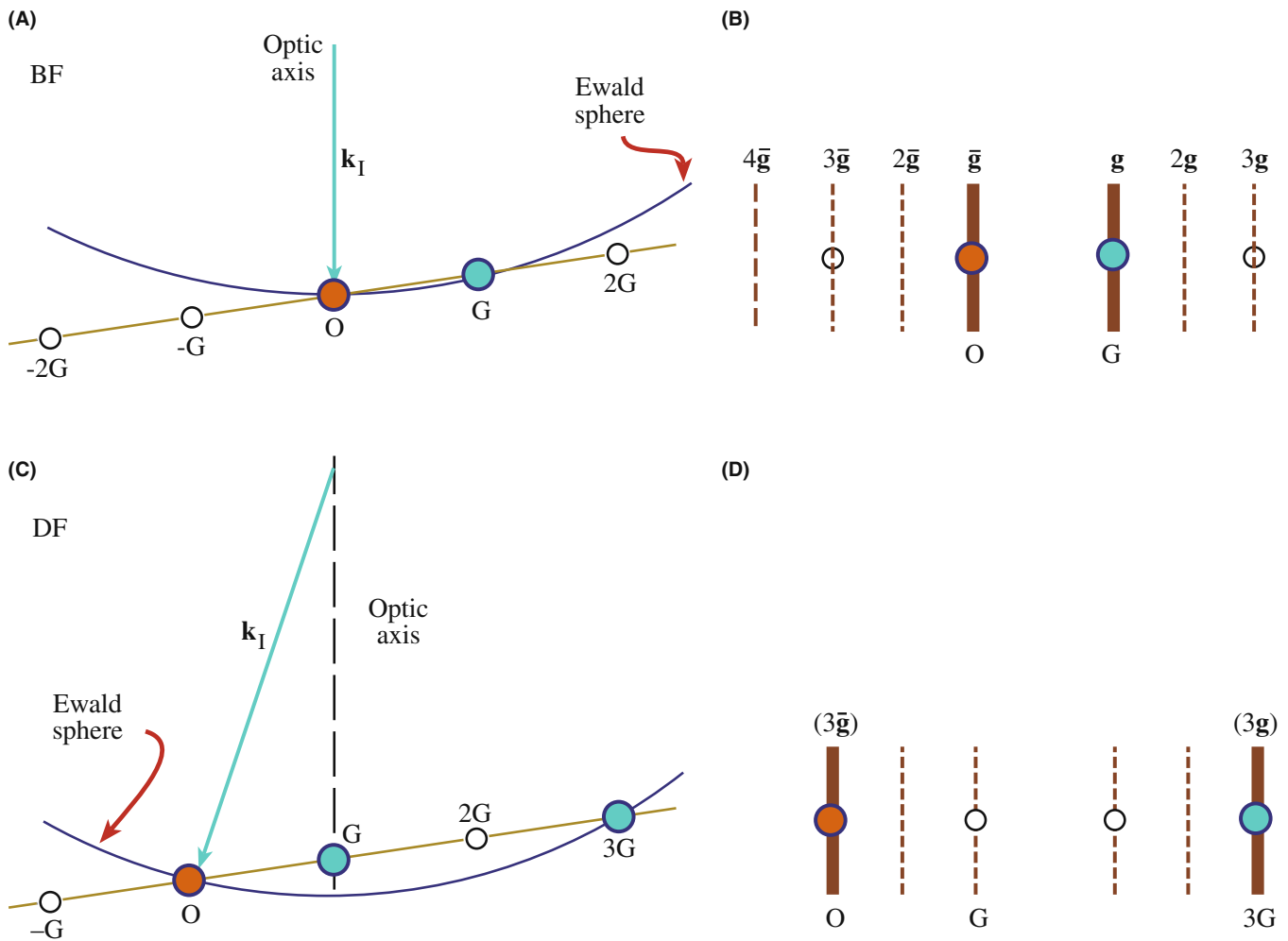


FIGURE 27.5. Relationship between the orientation of the Ewald sphere and the position of the Kikuchi lines for the $0(g)$ (A and B) and $g(3g)$ (C and D) diffraction conditions. The two pairs of diagrams are related by tilting the beam; the specimen has not tilted so the position of the Kikuchi lines is unchanged.

27.4 THICKNESS FRINGES IN WEAK-BEAM IMAGES

Thickness fringes in WB images are just like thickness fringes in strong-beam images but the effective extinction distance, ξ_{eff} , is much smaller. From equation 27.1 we can see that the intensity minimum occurs at thicknesses of $\mathcal{N}(s_{\text{eff}}^{-1})$ with maxima at $(\mathcal{N} + 1/2)\mathcal{N}(s_{\text{eff}}^{-1})$. The effective extinction distance for $s = 0.2 \text{ nm}^{-1}$ is 5 nm; this value is rather sensitive to the precise value of s so that the fringes will change if the foil bends. Using WB images we can form a rather detailed contour map of the specimen, but you must remember that both surfaces may be inclined to the beam, as shown in Figure 27.6.

The thickness effect is illustrated in Figure 27.7. These images were recorded with $s = 0.2 \text{ nm}^{-1}$. The MgO specimen has been heat-treated so that there are

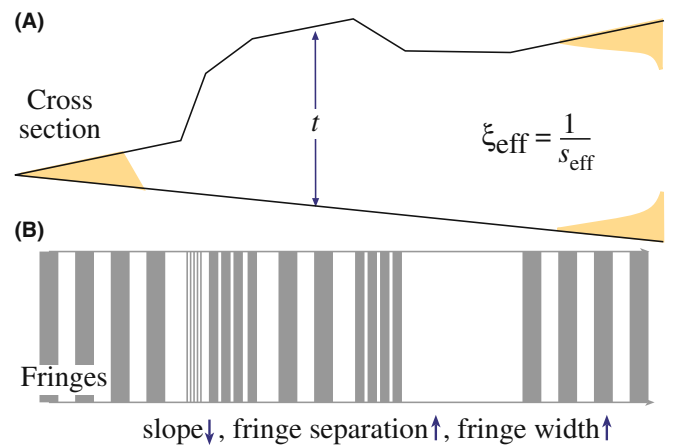


FIGURE 27.6. (A) In WB imaging, the thickness periodicity depends on the effective extinction distance, ξ_{eff} . (B) The separation of the fringes varies accordingly.

27.5 IMAGING STRAIN FIELDS

The principle of the technique is very simple. When the area of your specimen in which the defect of interest lies is oriented away from the Bragg position, the reflecting planes may be bent back into the reflecting position close to the defect. The region over which this occurs is very small because the strain has to be quite large to cause this bending. For the (220) planes in Cu (which fixes the plane spacing, d), the planes must rotate through an angle of $\sim 2^\circ$ to change s locally from 0.2 nm^{-1} to zero.

WHAT IS WEAK?

The intensity of the reflection that we see in the DP is still small even though a relatively intense peak may occur in the image close to the defect core because the DP averages over a large area.

When you look at dislocations in the WB image, you see bright lines on a dark background. Let's compare the WB image to a BF image of the same defect in Figure 27.8. You can see that the WBDF image is much narrower; you could make the comparison look even better if you make s very close to zero in the BF image.

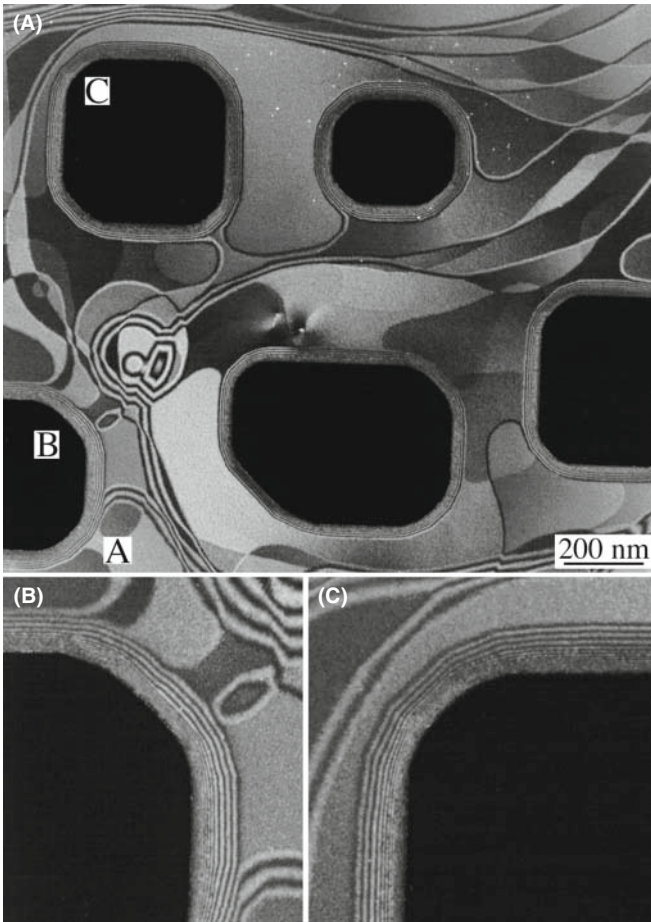


FIGURE 27.7. (A) WB thickness fringes from annealed MgO. (B, C) Higher magnifications of regions B and C. Compare to Figure 23.3.

large regions where the surface is atomically flat on both sides. Before heating, the specimen had been acid etched, which caused the holes seen in this image (they're black because we're in DF); you can see the defects causing the etch pits in some pits. The specimen shows inclined steps which curve across the surface. Where we see wide, uniform gray regions the surface is atomically flat. At A there is a large inclined step which runs into the hole B. Notice how the number of fringes around the hole B increases at A. Around the holes (Figure 27.7B and C) we see much more closely spaced fringes because the thickness changes more quickly here. Now if we look at the edge of any hole such as C, we see that the spacing of the fringes has one value away from the hole but another, smaller value close to the hole. What we find is that the inclined surface facets on different planes, with each facet becoming steeper closer to the hole. This topology is a result of the way the specimen has been prepared and would not normally be found, say, in electrochemically polished specimens, but it does illustrate the possibility of 'profiling' using thickness fringes.

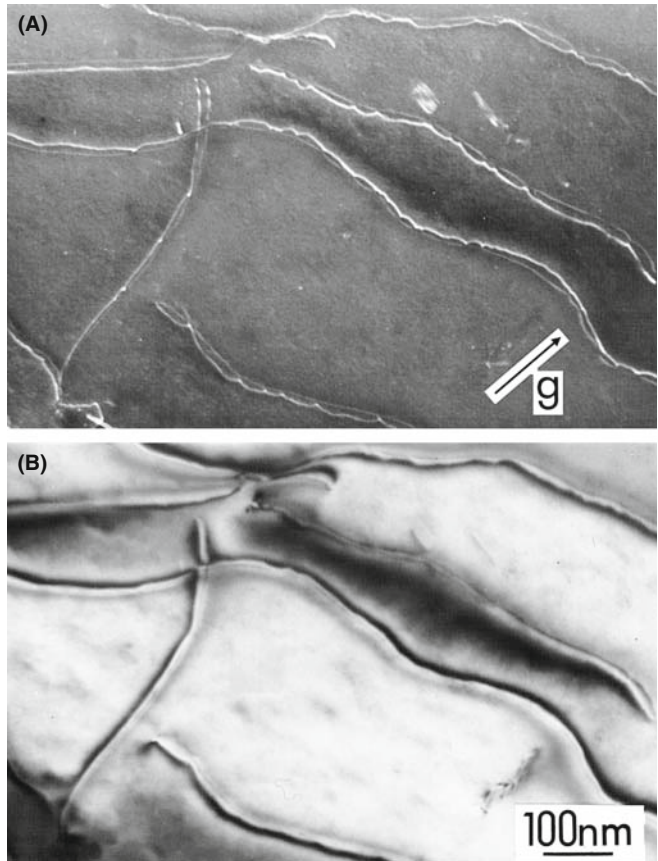


FIGURE 27.8. A comparison of dislocation images in a Cu alloy formed using (A) WB and (B) strong-beam ($s_g > 0$) conditions.

We'll keep our discussion of dislocations brief but draw your attention to a few particular points

- In the WB technique, most of the specimen is tilted so that s is large; the lattice planes in most of the specimen are then rotated away from the Bragg condition. However, as you can see in Figure 27.9, near the core of the dislocation the planes are locally bent back into the Bragg condition.
- This bending is only large close to the core of the dislocation (i.e., at the same depth from the surface).
- The peak you see in the WB image is always displaced to one side of the dislocation core. If you reverse the sign of \mathbf{g} , the peak moves to the other side of the core. If you reverse the Burgers vector, \mathbf{b} (rotate the diagram in Figure 27.9 through 180°), but keep \mathbf{g} the same, the peak again moves to the other side.
- If you increase s in the crystal, then the planes must bend more to satisfy the Bragg condition, which means the observed peak will move closer to the dislocation core.

- When we say 'position of the peak,' we are always talking about a projected position where the projection is along \mathbf{k}_D .
- There will be some situations where the local strain at the defect is not large enough to compensate for the s you have chosen. Then you will only see poor contrast in the image.

27.6 PREDICTING DISLOCATION PEAK POSITIONS

There are three ways to calculate the contrast in a WB image. Since each teaches us something new, we'll go through them in turn.

Method 1: The WB Criterion states that the largest value of $\phi_{\mathbf{g}}$ in the WB image occurs when $s_{\mathbf{R}}$, which we derived in equation 26.7, is zero. We can express this result as

$$s_{\mathbf{R}} = s_{\mathbf{g}} + \mathbf{g} \cdot \frac{d\mathbf{R}}{dz} = 0 \quad (27.6)$$

Equation 27.6 tells us that if the effective value of s (i.e., s_{eff}) is zero, even though $s_{\mathbf{g}}$ is not zero, then the direct beam and the diffracted beam, \mathbf{g} , are strongly coupled. In this situation the strain field effectively rotates the lattice planes into the Bragg-reflecting position. Therefore, the crystal can be oriented so that $\phi_{\mathbf{g}}$ is small for all columns except those near the dislocation core where it can attain a considerable magnitude due to the strong coupling with the transmitted beam as it passes through the region close to the core of the dislocation where s_{eff} is zero. This increased amplitude is then retained below the core when the coupling between the two beams is decreased again. The intensity is expected to be largest for that column where s_{eff} remains closest to zero over the longest length, and this occurs for the column where there is an inflection in the curve of \mathbf{R} versus z . Therefore, the position of the WB peak should occur when equation 27.6 is satisfied at a turning point of $\mathbf{g} \cdot d\mathbf{R}/dz$.

Method 2: The Kinematical Integral. An alternative criterion for defining the position of the WB peak was derived with the approximation that only two beams are considered and s is sufficiently large. Cockayne showed that the maximum scattering from the transmitted to the diffracted beam occurs where the kinematical integral, defined as

$$\int_{\text{column}} e^{-2\pi i(s_{\mathbf{g}}z + \mathbf{g} \cdot \mathbf{R})} dz \quad (27.7)$$

is maximized. This maximum, in general, occurs for a column which is closer to the dislocation core than predicted by *Method 1*. The reason for this difference

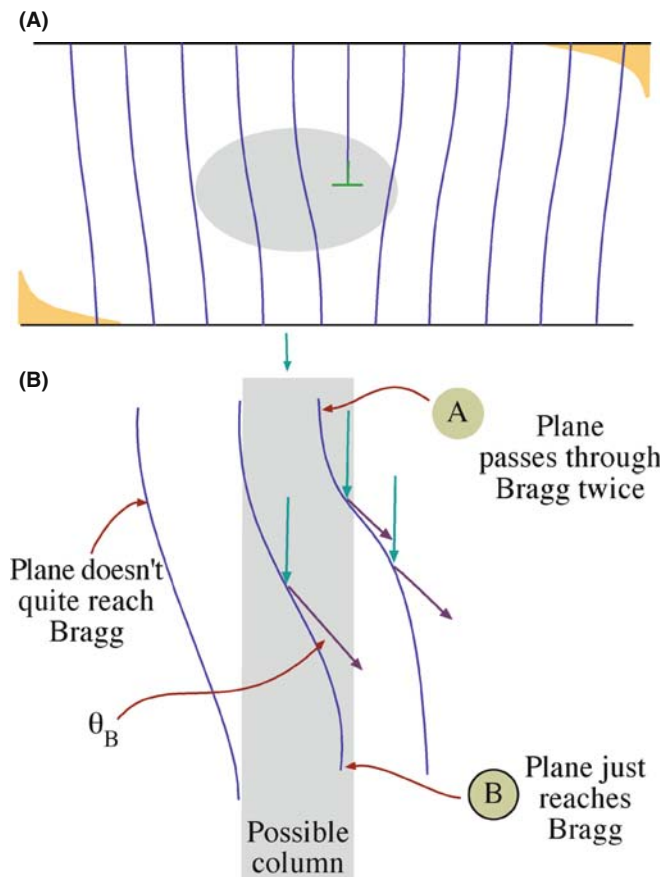


FIGURE 27.9. WB images from strained regions of the specimen show high intensity only from those regions where the diffracting planes are bent back into the Bragg condition. This illustration is for an edge dislocation. (B) shows the core region selected in (A).

is interesting: because the planes are bent, the reciprocal lattice point is, on average, nearer to the Ewald sphere. Therefore, the integral has a larger value over the length of the column.

Without doing all the math, we can illustrate how these two approaches are related. What we want to do is determine when $\phi_{\mathbf{g}}$ is large, but still kinematical (i.e., \mathbf{s} is large); we want to maximize the kinematical integral in equation 27.7. We can do this using the stationary-phase method described by Stobbs. We write the integral as

$$\int_0^t \exp\left(-2\pi i \left[\frac{z^2}{2} \cdot \frac{d^2}{dz^2}(\mathbf{g} \cdot \mathbf{R}) - \frac{z^3}{3} \cdot \frac{d^3}{dz^3}(\mathbf{g} \cdot \mathbf{R}) \right]\right) dz \quad (27.8)$$

where we have set $s + d/dz(\mathbf{g} \cdot \mathbf{R}) = 0$. If we also set $d^2/dz^2(\mathbf{g} \cdot \mathbf{R}) = 0$ (at the inflection), we go further to ensuring that the term in the square brackets is zero. This condition is what we guessed for the first method of defining the WB criterion.

Method 3: Compute the Contrast. Now that personal computers are widely available, we can calculate the position of the WB peak and graph the results. What we then find is that the WB peak actually lies between the two values predicted by the two criteria deduced using the first two methods. We also find, using the computer, that the position and width of the image peak are affected by any strongly excited diffracted beams so these *must* be avoided. Some apparently weak beams can also affect the image. A practical point is that the computer sometimes gives a rather pessimistic view of the variability in the peak position, so we have to weight the results carefully. Remember that the important advantages of this approach are that we can include the effects of the other diffracted beams which are always present, and we can take account of other effects, such as the convergence of the beam.

In the kinematical approximation, the half-width, Δx , of the image of an undissociated screw dislocation with $|\mathbf{g} \cdot \mathbf{b}| = 2$ is given approximately by the relation derived by Hirsch et al. (1960)

$$\Delta x = \frac{1}{\pi s_{\mathbf{g}}} \cdot \frac{\xi_{\text{eff}}}{3} \quad (27.9)$$

This expression is a very useful rule of thumb. You will realize that this WB image width is special for three reasons, which arise because it doesn't depend on $\xi_{\mathbf{g}}$. So, once \mathbf{s} is fixed in a WB image, we can make several surprising statements regarding the width of the dislocation peak

- It does not depend on the material.
- It does not depend on the reflection.
- It does not depend on the kV.

Take the example of the 220 reflection in Cu at 100 kV: $\xi_{\mathbf{g}}$ is 42 nm, and the width Δx is 14 nm. So, even if equation 27.9 is slightly wrong, the image width is greatly reduced in WB. If we make $s_{\mathbf{g}} = 0.2 \text{ nm}^{-1}$, then the half-width is 1.7 nm.

Computed many-beam images confirm that dislocations in other orientations give rise to similar narrow peaks when this value of $s_{\mathbf{g}}$ is used. A series of peak profiles for different values of t is shown in Figure 27.10. Notice that although the intensity of the peak may be only about 0.1% of the incident beam, it is still much higher than the background.

THE WB 'CRITERION'

$s_{\mathbf{g}} = 0.2 \text{ nm}^{-1}$ is always a useful guide; it satisfies the requirements that the image should have a narrow width and show a high contrast between the defect and background regions.

Equation 27.9 indicates that as we increase the value of s , the half-width of the image peak decreases. However, a maximum is imposed on s by the fact that the intensity of the diffracted beam varies as s^{-2} . If we make s much larger, the contrast of the image therefore becomes too small to be of practical use.

The basic requirements governing the value of s which you must use for quantitative imaging are

- $s \geq 2 \times 10^{-2} \text{ \AA}^{-1}$ to give sufficiently narrow peaks for fine detail to be studied.
- $s \leq 3 \times 10^{-2} \text{ \AA}^{-1}$ because the intensity varies as s^{-2} in the kinematical limit.
- $s\xi_{\mathbf{g}} \geq 5$ to give sufficient contrast in the WB image.

If you use the $\mathbf{g}(3\mathbf{g})$ condition for Cu with $\mathbf{g} = 220$ and 100-keV electrons, then the value of $s_{\mathbf{g}}$ will be 0.238 nm^{-1} .

27.7 PHASOR DIAGRAMS

We sometimes find it useful to demonstrate the depth dependence of the contrast in the WB image using phasor, or amplitude-phase, diagrams which we introduced in Chapter 2. You can generally use such diagrams whenever the kinematical approximation holds; they are equivalent to a graphical integration of the two-beam equations for the case where \mathbf{s} is large. In fact, many of the early calculations of defect contrast were made using this approach before computers became widely available. We recommend that you glance at the original paper by Hirsch et al. (1960): the contrast profiles were calculated by plotting the phasor diagram.

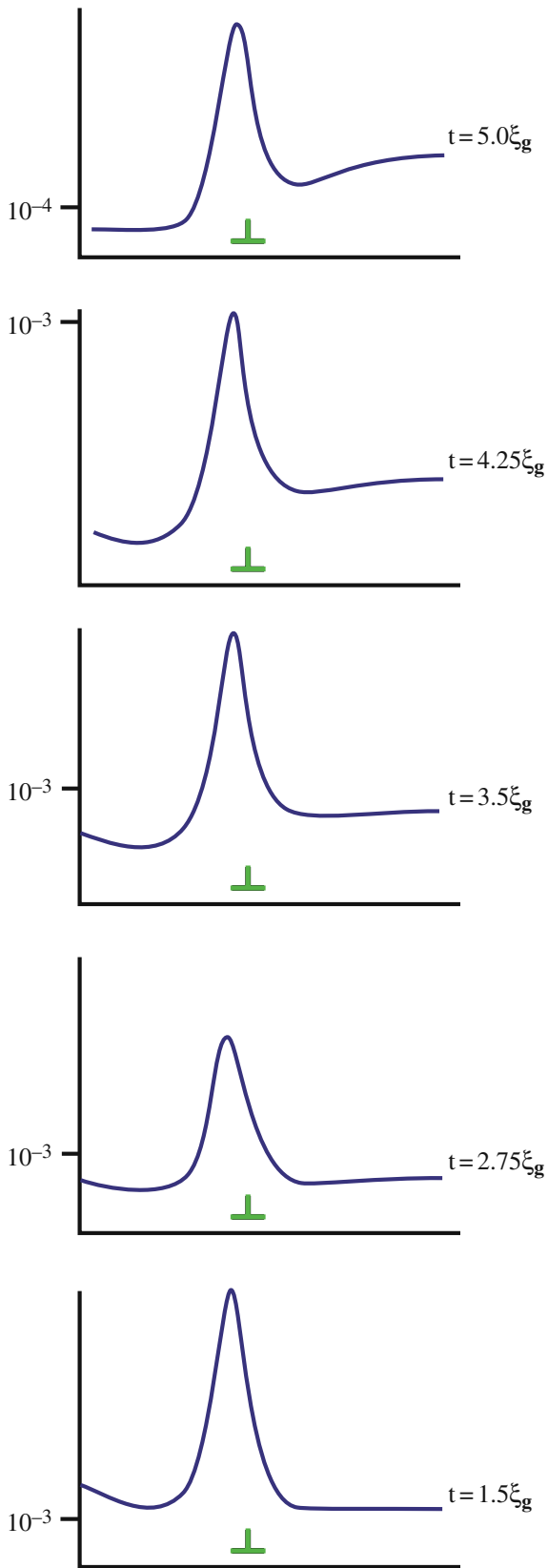


FIGURE 27.10. Examples of computer-calculated intensity peaks in WB images of an edge dislocation in Cu for different values of t . The intensity is relative to the unit incident-beam intensity. Note that the dislocation position and the peak intensity never coincide.

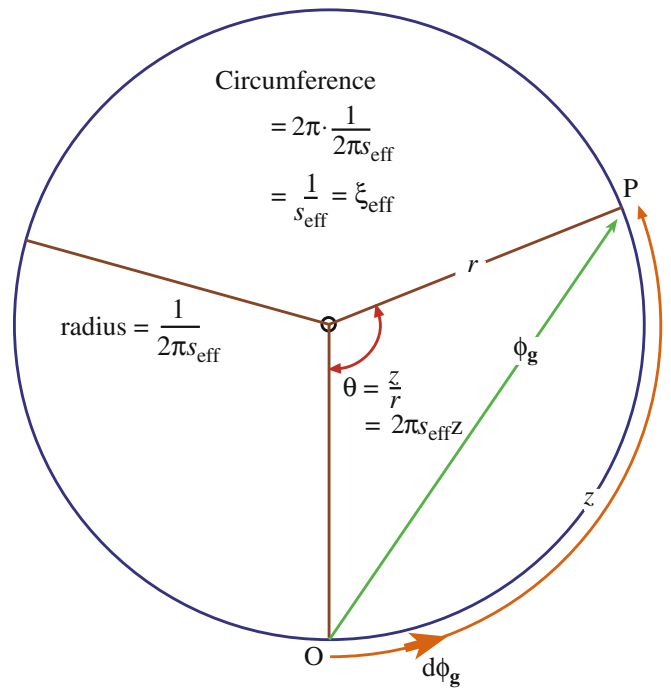


FIGURE 27.11. A phasor diagram for the WB case. The distance z is the arc OP measured around the circumference and the radius of the circle is $(2\pi s_{\text{eff}})^{-1}$. An increase in θ of 2π means the thickness increased by the ξ_{eff} . The amplitude in the WB is $\phi_{\mathbf{g}}$.

The basic idea is shown in Figure 27.11. We simply add all the $d\phi_{\mathbf{g}}$ increments to $\phi_{\mathbf{g}}$. In doing so, we take account of the phase changes which occur as the beam passes through the crystal. Remember that in this approximation, no electrons leave the \mathbf{g} beam! If the crystal is perfect and our increments are sufficiently small, we will produce a smooth circle.

The circumference of this circle is ξ_{eff} as we require for the depth periodicity, and the radius is $\xi_{\text{eff}}/2\pi$ or $(2\pi s_{\text{eff}})^{-1}$. Notice that as we increase s , we decrease ξ_{eff} and the circle becomes smaller. Thus we move around the circle more quickly if s is large. In other words, our effective extinction distance is reduced, as we knew from Chapter 13.

If the diffracted beam passes through a stacking fault, it will experience an extra phase shift given by $2\pi\mathbf{g}\cdot\mathbf{R}$. Using the familiar example for an fcc crystal, we take an example where $\mathbf{R} = \frac{1}{3}[11\bar{1}]$ and $\mathbf{g} = (20\bar{2})$ which gives $\alpha = 2\pi/3 = 120^\circ$ (modulo 2π). The abrupt phase change is shown at P_3 , in Figure 27.12. Now, the value of $\phi_{\mathbf{g}}$ (P_1P_2) can be much larger than in a perfect crystal. The locus of $\phi_{\mathbf{g}}$ still travels around the first circle until it meets the planar fault at $z = n\xi_{\text{eff}} + t_1$, where z and t_1 are measured parallel to \mathbf{k}_D . It then moves onto the second circle until it reaches $z = t$. You can readily see that if we keep the depth of the fault, t_1 , fixed, we then see depth fringes which vary with periodicity ξ_{eff} as we vary the total thickness, t . (The value of α is still 120° .) The situation is a little

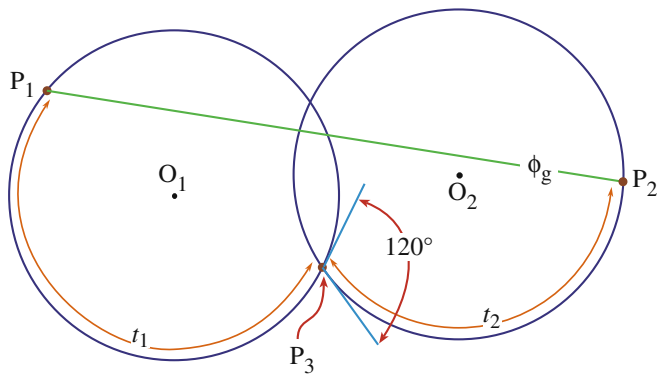


FIGURE 27.12. A phasor diagram used to explain the WB contrast from a SF at depth t_1 in a foil of thickness $t_1 + t_2$. P_1 is the top of the foil (on a circle centered on O_1); P_2 is the bottom of the foil (on a circle centered on O_2). Here, the phase change at the SF (P_3) is 120° . The total amplitude in the WB is ϕ_g .

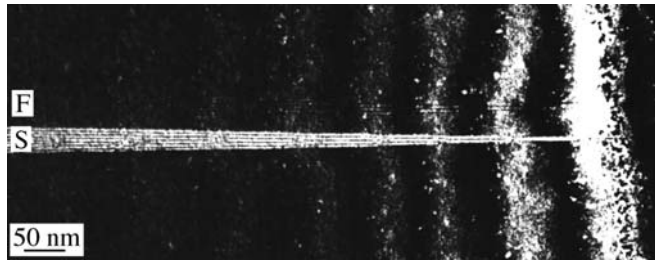


FIGURE 27.13. Illustration of thickness fringes in an experimental image of two inclined translation boundaries (S and F) in a wedge specimen of spinel. The thickness fringes from the wedge and from the inclined boundaries can be clearly seen and counted. Boundary ‘F’ shows very little contrast because it has a small R value.

more difficult to envision if we keep t fixed but vary t_1 , but the principle is the same. Images of WB fringes from inclined SFs are illustrated in Figure 27.13. The thickness fringes of the wedge specimen and those from the inclined stacking fault can both be clearly seen and counted. Notice that the number of bright fringes on the planar defect really does increase by one for every increment ξ_{eff} in the thickness of the wedge and you don’t need to know R .

You can imagine applying this analysis to the situation where you have several overlapping planar faults, as shown in Figure 27.14. $\phi_g (P_1P_2)$ can become very large. (Does our approximation of unit incident intensity still hold?) This situation does occur in practice, as shown in Figure 27.15. Here we have several overlapping faults. The bright ones are bright even when ϕ_g reaches its minimum, as you would expect from Figure 27.14. If you compare the WB image with its strong-beam counterpart in Figure 25.5, you will notice that there is much more detail in the WB image; as we saw in Chapter 25, overlapping faults on nearly adjacent

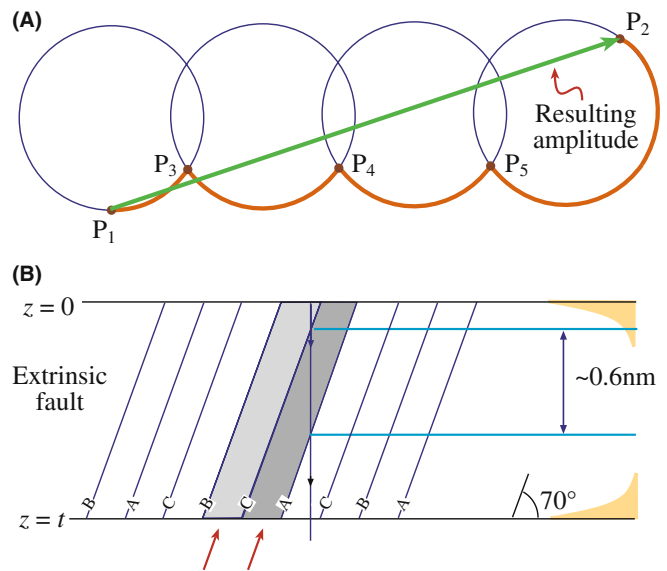


FIGURE 27.14. Phasor diagram (A) for WB encountering a series of overlapping planar defects (B) at P_3 , P_4 , and P_5 . For a $11\bar{1}$ foil normal, the inclined $11\bar{1}$ planes lie at 70° to the surface so that the spacing between adjacent planes in the direction of the beam is 0.627 nm . Compare with Figure 27.12.

planes can essentially give no contrast in the BF image. You can easily check such effects in WB by adjusting s as in Figure 27.15C–E. It’s interesting to realize that this effect can occur even when two intrinsic SFs lie on adjacent planes to give the extrinsic SF. This approach can be used to image other planar defects such as the $\{112\}$ twin boundary which can have a significant thickness. The key factor is that, under WB conditions, ξ_{eff} can become comparable to the distance that the beam travels between encountering successive planes of atoms, particularly when the interface is quite steeply inclined relative to the beam.

We can also use a phasor diagram to describe the contrast from a dislocation, but now the phase change occurs over a length of the column rather than at a particular depth in the sample. As illustrated in Figure 27.16, the phase can either add or subtract depending on the sign of $\mathbf{g} \cdot \mathbf{R}$. When the phase changes quickly with a change in t , as at the center of Figure 27.16B, it means that we are strongly coupling the incident and diffracted beams.

We’ll summarize our discussion of phasor diagrams by two points

- They should only be used when the kinematical approximation holds.
- They then give us a graphical method for understanding the variation of ϕ_g with thickness; the method is especially useful when crystal defects are present.

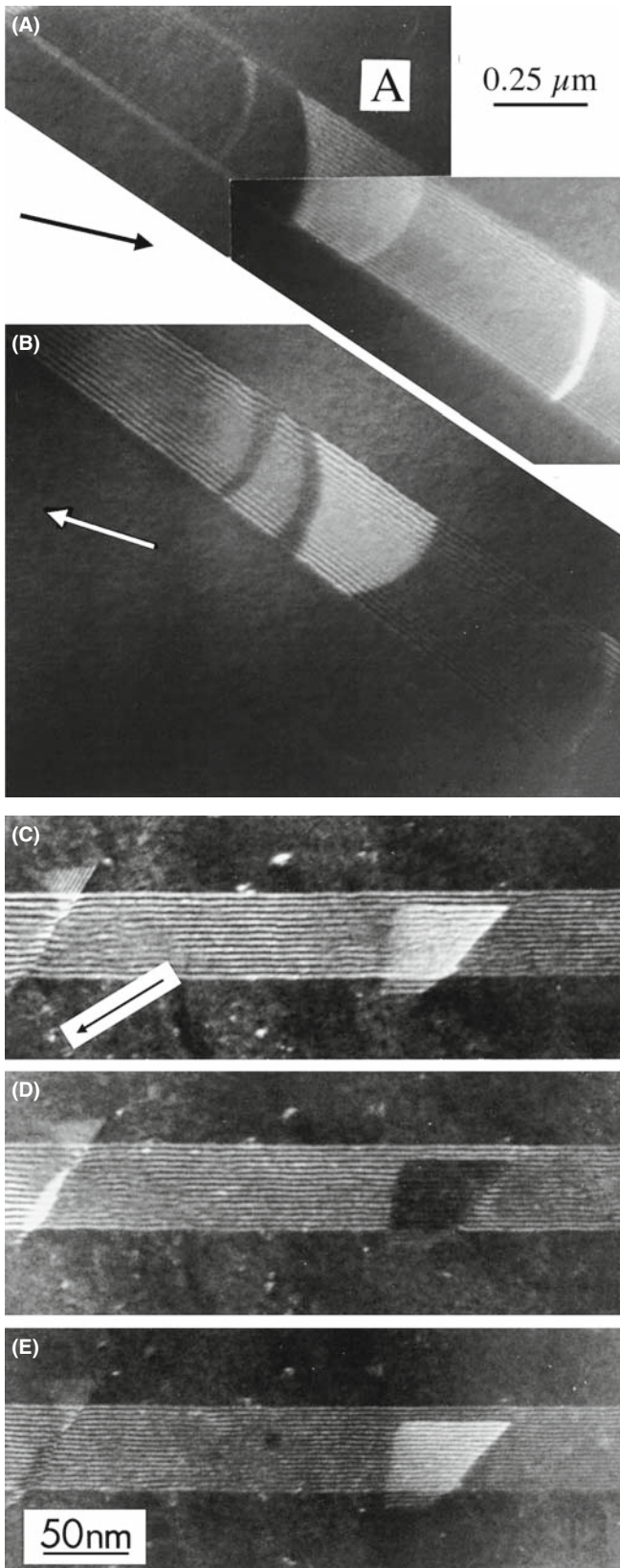


FIGURE 27.15. (A, B) Overlapping SFs imaged in WBDF using two different conditions ($\pm\mathbf{g}$). Many more fringes occur than in the corresponding BF image in Figure 24.5. Fringes are present at A where the BF image showed no contrast. (C–E) Changes in fringe spacing and intensity when one SF overlaps another; s increases from C to E. (Note the decrease in the fringe spacing.)

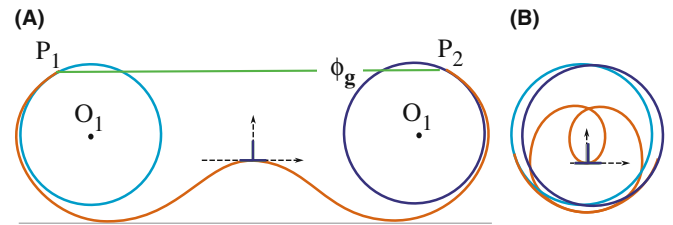


FIGURE 27.16. Phasor diagrams for a dislocation for $\pm\mathbf{g}$. The phase change is not abrupt but instead occurs over an extended distance along the column. (A) The phase change causes an increase in the amplitude scattered: $\phi_{\mathbf{g}}$ is larger than in the perfect crystal (compare to Figure 27.12). (B) When \mathbf{g} is reversed the phase change has the opposite sign and the resulting $\phi_{\mathbf{g}}$ is much smaller (so it's not shown in this confusing figure).

27.8 WEAK-BEAM IMAGES OF DISSOCIATED DISLOCATIONS

Although the study of dislocations is a very specialized topic, it beautifully illustrates the potential of the WB technique. Dissociated dislocations are common in face-centered-cubic (fcc) materials (including Si) and ordered intermetallics such as Ni_3Al . The geometry of a dissociated dislocation in Cu is summarized in Figure 27.17. We gave some general references on the theory of dislocations in Chapter 26.

Since the computed many-beam images show that the position of the dislocation image lies close to the position predicted by the WB criterion, this criterion is used in practice because it allows us to deduce an equation for the position of the image. We can thus directly relate the separation of Shockley partials, for example, to the measured separation of the two peaks observed in $|\mathbf{g} \cdot \mathbf{b}_T| = 2$ images of dissociated dislocations. Hence we can estimate the stacking-fault energy (SFE) of

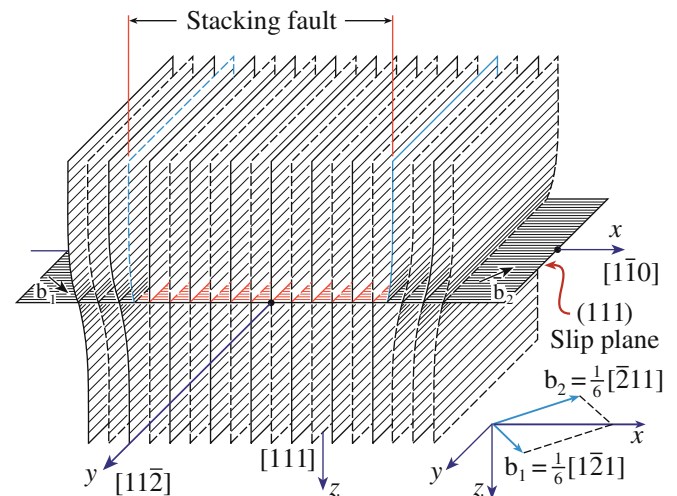


FIGURE 27.17. The geometry of a perfect dislocation in Cu. The perfect dislocation separates into two Shockley partial dislocations with Burgers vectors \mathbf{b}_1 and \mathbf{b}_2 separated by a SF on a $\{111\}$ plane. Although it might appear that partial dislocations are edge type, they are actually 60° dislocations (shown in the figure) with an edge component.

semiconductors and several fcc metals. In order to interpret WB images of extended dislocation configurations, we may need to know how the position of the image peak is related to the position of the dislocation core. This information is essential whenever we use the WB technique to collect *quantitative* data. We will now answer two questions for quantitative WB analysis

- What factors determine the position of the WB peak?
- What methods are used to determine the Burgers vectors of the dislocations?

When the WB conditions are satisfied, dissociated dislocations can be imaged in a $\{220\}$ reflection for which $|\mathbf{g} \cdot \mathbf{b}_T| = 2$, so that each partial dislocation gives rise to an intensity peak having a half-width of 1–1.5 nm. The separation of the partials can be deduced to within ± 0.7 nm providing that the peak separation is greater than ~ 2.5 nm. For many materials, we need to use anisotropic elasticity theory to relate the atomic displacements to the image, so the computer becomes essential.

Why not just use HRTEM since, as we'll see in Chapter 28, HRTEM can give detail down to below 0.2 nm, whereas WB is often limited to ~ 1 nm? If you want to interpret an HRTEM image the defect must be *absolutely straight*, parallel to the beam, and located in a very thin region of the specimen. The segment of dislocation studied by HRTEM will thus be no longer than 20 nm and less than 10 nm for the highest resolution. In a WB image the defect can be micrometers long and, in a relatively thick foil, it can even change direction. If you look back at Figure 27.8 you'll see pairs of lines which correspond to partial dislocations in the WB image. You can see other features such as constrictions. The strong-beam image may also show two or more lines for a particular dislocation but as we saw in Chapter 26, these lines are generally not related to the detailed structure of the dislocation but rather to n in the equation $\mathbf{g} \cdot \mathbf{b} = n$.

Ideally, for quantitative analysis, you should choose long, nearly straight dislocations. As shown in Figure 27.17, the two Shockley partial dislocations lie in the (111) plane of the foil. The Burgers vectors of the total and partial dislocations can be determined by imaging the dislocation in the WB mode using the $\{2\bar{2}0\}$ reflections. A sharp peak is found for a partial dislocation with $|\mathbf{g} \cdot \mathbf{b}_p| = 1$, and either no peak or a diffuse one (arising from the anisotropy of the lattice) if the partial dislocation has $|\mathbf{g} \cdot \mathbf{b}_p| = 0$. When $|\mathbf{g} \cdot \mathbf{b}_T| = 2$, the diffraction vector \mathbf{g} and the Burgers vector \mathbf{b}_T are parallel and, in the image of a dissociated dislocation, two sharp peaks are formed, one corresponding to each of the partial dislocations (both now have $|\mathbf{g} \cdot \mathbf{b}_p| = 1$).

We can explain this difference in intensity if one peak (the weaker) arises from the region between the partial dislocations and the other from outside the dissociated dislocation. This effect cannot occur in $|\mathbf{g} \cdot \mathbf{b}_T| = 1$ images, when $|\mathbf{g} \cdot \mathbf{b}_p| = 1$ for one partial and $|\mathbf{g} \cdot \mathbf{b}_p| = 0$ for the other, and it can be used to identify the reflection for which $|\mathbf{g} \cdot \mathbf{b}| = 2$. Confirmation of the Burgers vector is always obtained using the BF mode, observing characteristic $|\mathbf{g} \cdot \mathbf{b}_T| = 2$ or $|\mathbf{g} \cdot \mathbf{b}_T| = 0$ images.

DISSOCIATED DISLOCATIONS

One of the peaks is on average more intense than the other; the order reverses when $\bar{\mathbf{g}}(3\bar{\mathbf{g}})$ is used instead of $\mathbf{g}(3\mathbf{g})$.

In a WB image with $|\mathbf{g} \cdot \mathbf{b}_T| = 2$, each of the partial dislocations will generally give rise to a single peak in the image which is close to the dislocation core. You can calculate the approximate positions of these peaks using the criterion from equation 27.6. Then, you can relate the separation of the peaks in the image to the separation Δ of the partial dislocations.

We can write the displacement, using isotropic elasticity theory, as the sum of the displacements due to the individual partial dislocations. If the Burgers vector of a straight, mixed dislocation lies in the (111) plane parallel to the surface of a foil, then at a distance x from the dislocation core

$$-s_g = \frac{|\mathbf{g}|}{2\pi} \left[\left(|\mathbf{b}_1| + \frac{|\mathbf{b}_{1e}|}{2(1-\nu)} \right) \frac{1}{x} + \left(|\mathbf{b}_2| + \frac{|\mathbf{b}_{2e}|}{2(1-\nu)} \right) \frac{1}{x-\Delta} \right] \quad (27.10)$$

Here x defines an axis perpendicular to both the dislocation line and the beam direction, and e refers to the edge component of the Burgers vectors of the partial dislocations, 1 and 2. This relation is particularly simple because, for the geometry we have chosen, the term $\mathbf{g} \cdot \mathbf{b} \times \mathbf{u}$ is zero. Using the notation

$$a = -s_g \left[\frac{|\mathbf{g}|}{2\pi} \cdot \left(|\mathbf{b}_1| + \frac{|\mathbf{b}_{1e}|}{2(1-\nu)} \right) \right]^{-1} \quad (27.11)$$

and

$$b = -s_g \left[\frac{|\mathbf{g}|}{2\pi} \cdot \left(|\mathbf{b}_2| + \frac{|\mathbf{b}_{2e}|}{2(1-\nu)} \right) \right]^{-1} \quad (27.12)$$

equation 27.10 reduces to

$$1 = \frac{1}{ax} + \frac{1}{b(x-\Delta)} \quad (27.13)$$

which has two solutions, x_+ and x_- , given by

$$x_{\pm} = \frac{ab\Delta + a + b \pm \left[(ab\Delta + a + b)^2 - 4a^2b^2\Delta \right]^{1/2}}{2ab} \quad (27.14)$$

These values of x define the positions of the peaks in the image. The separation between these peaks is then given by

$$\Delta_{\text{obs}} = \left[\Delta^2 + \frac{(a+b)^2}{a^2b^2} + \frac{2(a+b)\Delta}{ab} - \frac{4\Delta}{a} \right]^{1/2} \quad (27.15)$$

We can rearrange this equation to make it appear more symmetric in a and b . Of course, it will not be symmetric because the peak is always located on one side of the dislocation.

$$\Delta_{\text{obs}} = \left[\left(\Delta + \frac{1}{b} - \frac{1}{a} \right)^2 + \frac{4}{ab} \right]^{1/2} \quad (27.16)$$

Computed images confirm that this relation is accurate for $\Delta_{\text{obs}} > 2.5$ nm, to within ± 0.7 nm. This uncertainty is due to the variation of the peak position with the depth of the dislocation in the foil and the foil thickness. A small uncertainty arises when you have not determined the actual direction of \mathbf{b}_T , i.e., whether it is in the direction of \mathbf{g} or $\bar{\mathbf{g}}$. Stobbs and Sworn found that the relation (equation 27.16) subject to the ± 0.7 nm uncertainty is still a good approximation when anisotropic elasticity theory is used.

As a simple exercise, consider the WB images of a dissociated screw dislocation and a dissociated edge dislocation. You should pay particular attention to ‘ a ’ and ‘ b ’ because in one case, \mathbf{b}_{1e} and \mathbf{b}_{2e} have the same sign, while in the other the sign is opposite. Does the image always have the same width when you reverse \mathbf{g} ?

Example 1. Even if you never want to calculate the actual separation of two dislocations from observations of two peaks, you can learn new ideas about dislocations from such images. Figure 27.18 is a famous set of images showing a dislocation in Si which is constricted along part of its length and dissociated along the rest. Even if you don’t know the precise details of the dislocation structure, you know that it can adopt two variants; the rest of the task is modeling the defect.

Example 2. The WB image of the node pair in Figure 27.19A, tells you very quickly that the two nodes are different; if we form images using other \mathbf{g} vectors (Figure 27.19B–D), we find that one of the partial dislocations is out of contrast in the image. The extended node contains the same type of intrinsic stacking fault that is present in the dissociated dislocation; $\mathbf{g} \cdot \mathbf{R}$ is zero in this image for the stacking fault. The other node is constricted,

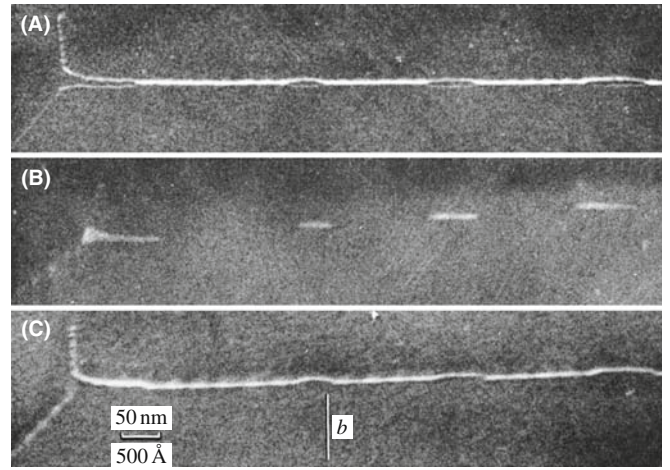


FIGURE 27.18. WB image of a dislocation in Si which has both dissociated and constricted segments: (A) $\mathbf{g} \cdot \mathbf{b} = 2$; both partial dislocations are visible. (B) $\mathbf{g} \cdot \mathbf{b}_T = 0$ showing SF contrast. (Note: no fringes.) (C) $\mathbf{g} \cdot \mathbf{b} = 1$; only one partial dislocation is visible.

within the detection limit of the WB technique. Comparison with the BF images in Figure 26.8 is instructive. Note that in Figure 27.19A the SF in the node does not show fringes because the SF plane is parallel to the surface of the specimen; we just tilted the specimen.

Example 3. We mentioned that the peak moves to the other side of the dislocation if we reverse \mathbf{b} . This is exactly what happens for a dislocation

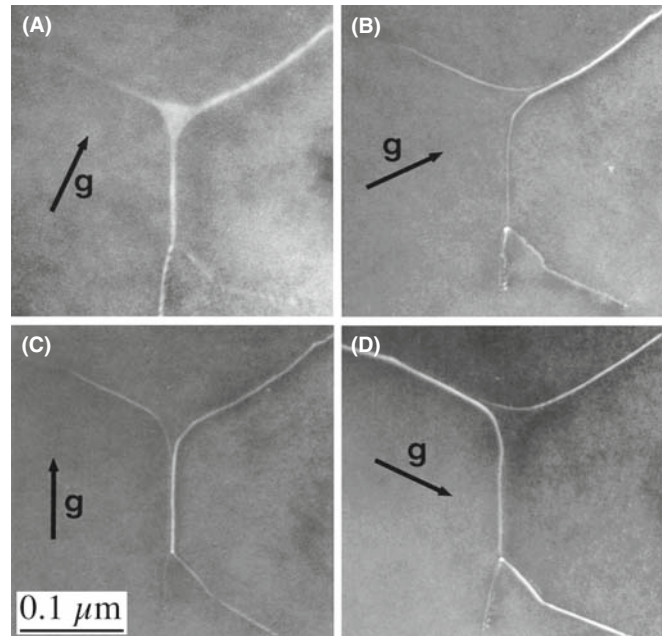


FIGURE 27.19. WB images of a pair of dislocation nodes formed by the dissociation of interacting dislocations lying on a $\{111\}$ plane in a Cu alloy. The SF is imaged in (A), and imaging with different reflections (B–D) the partial dislocations are out of contrast when $\mathbf{g} \cdot \mathbf{b} = 0$. Compare with the BF image in Figure 26.8 and note the difference in magnification.

dipole as you can see in Figure 27.20. This is a complicated figure except that you can use it not only to see the inside-outside contrast in WB

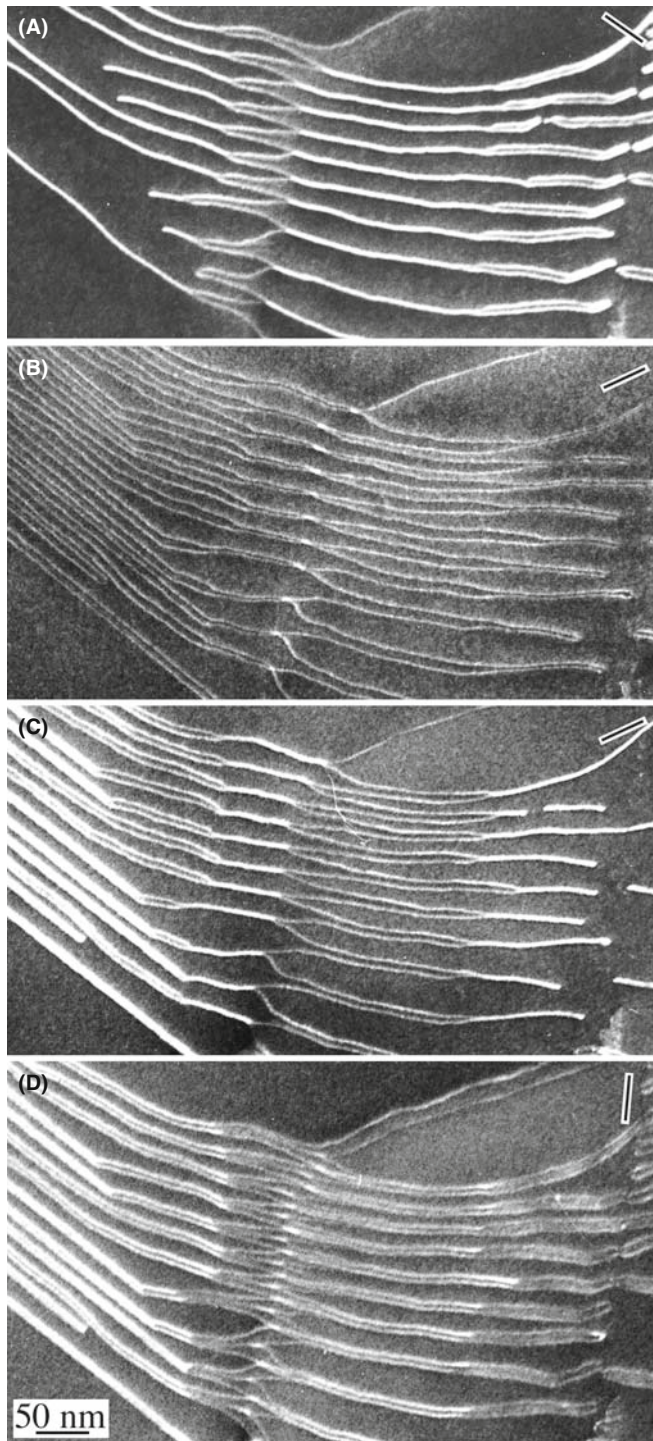


FIGURE 27.20. Four WB images showing an array of dislocation dipoles in a Cu alloy having a low stacking-fault energy. The reflections are all 220-type and the dislocations all lie on (111) planes which are nearly parallel to the surface of the specimen. All the dislocations are dissociated. All the Shockley partial dislocations are in contrast in (D) while half are out of contrast in (A–C). Notice that the narrower images are brighter than the wider ones; the strain is large in between the dislocations but decreases rapidly outside the dipole since the total Burgers vector of a dipole is zero.

images of dislocation dipoles but also as an exercise in $\mathbf{g} \cdot \mathbf{b}$ analysis. A dislocation dipole is a pair of dislocations identical in every way, apart from the sign of the Burgers vector. If we now reverse \mathbf{g} , then both peaks move to the other side of their respective dislocations. This change in contrast is referred to as inside-outside contrast and is commonly seen on dislocation loops, which are themselves closely related to these dipoles. Loops are just more ‘equiaxed’ than dipoles and always consist of one dislocation (which may be dissociated). The images shown in Figure 27.20B and C illustrate the dramatic change in contrast which you can see on reversing \mathbf{g} . Some of the dipoles completely disappear in Figure 27.20A–C because they are a special form of defect known as a faulted dipole. Such dipoles usually give very low contrast in strong-beam BF images because the dislocations are always very close together so that their strain fields overlap and cancel, and the lattice is thus only distorted over very small distances. All the dislocations are in contrast in Figure 27.20B: it’s the $\mathbf{g} \cdot \mathbf{b} = 2$ image. When we use the WB techniques, we are probing the structure on these very small dimensions and the contrast can be high. Again compare with the BF image in Figure 26.13.

Examples 4 and 5. The WB technique allows us to see features which would be hidden if we used strong-beam imaging. Figure 27.21 gives a simple example where an inclined SF is cutting through several dissociated dislocations. The interaction of the two defects would be masked by the SF fringes in the strong-beam image but is clearly visible in the WB image. Seeing small particles close to dislocations is difficult in strong-beam imaging. Although not easy in WB, Figure 27.22 does illustrate that it can be done. This image shows, for example, that the behavior of the dissociated dislocation is different on each side of the particle.

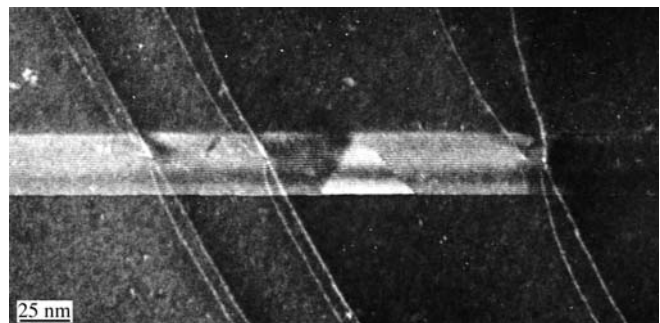


FIGURE 27.21. WB image of an inclined SF cutting through a series of dissociated dislocations lying parallel to the surface of the Cu-alloy specimen.

27.9 OTHER THOUGHTS

27.9.A Thinking of Weak-Beam Diffraction as a Coupled Pendulum

We can illustrate the principle which underlies the increase in intensity in the WB image close to a dislocation using the mechanical analog of a coupled pendulum. A diagram is shown in Figure 27.24. The two pendula are connected (coupled) by a third string. If we start the left pendulum swinging but hold the connecting string, the right-hand pendulum remains stationary. Now release the connecting string. You will see that the right-hand pendulum now begins to swing. If we let the process continue, eventually the right-hand pendulum is swinging as much as the original one did, but the original one is stationary: this is the strong-beam analog! All the kinetic energy has been transferred from one pendulum to the other. Given more time, the pendulum will return to the original condition. Now repeat the exercise but hold the connecting string again after the right-hand pendulum has begun to swing; you will notice that both pendula continue to swing, each with a constant amplitude. The role of the connecting string is to couple the two pendula (beams) so that we transfer energy from one beam to the other. In WB TEM, the defect acts as the connecting string. The two beams are only coupled over a short length as they travel past the defect. We can plot this amplitude (or intensity); try this as an exercise.

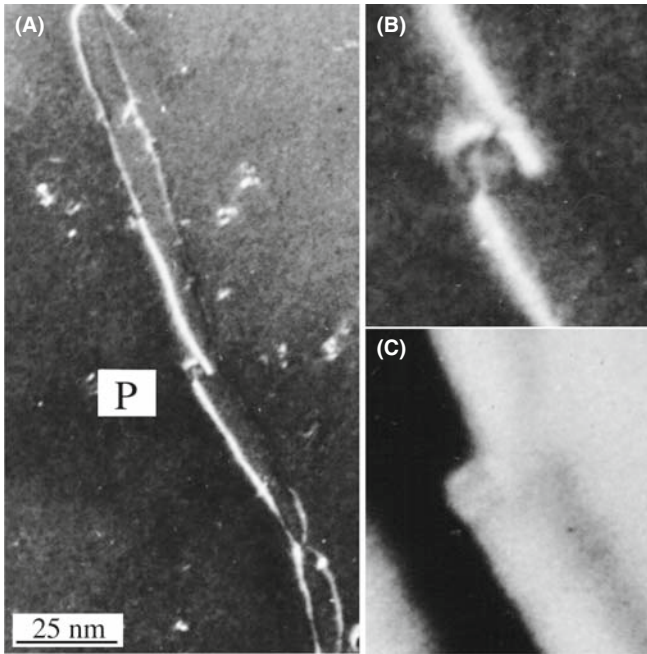


FIGURE 27.22. (A) WB image of a dissociated dislocation interacting with a particle (P) in a Cu alloy. (B) and (C) are enlargements of the WB and corresponding BF images, respectively. Note the difference in detail.



FIGURE 27.23. WB image showing the dissociation of a group of dislocations which are inclined to the foil surface and give wedge-shaped SFs. The shape of the SFs is caused by surface stresses.

Figure 27.22B and C compares magnified WB and BF images from the region close to the particle, and again stresses the advantage of WB over BF imaging.

Example 6. We noted in Section 26.8 that the surface of the specimen can affect the geometry of the defects we are examining. In general, the specimen needs to be thinner for WB imaging than for strong-beam imaging. Therefore, the surface can affect not only the appearance of the image but also the actual geometry of the defect. Figure 27.23 shows an example where this effect is particularly clear. Dislocations that were uniformly separated in the bulk material now appear wedge-shaped: the effects of the two surfaces are different in this case.

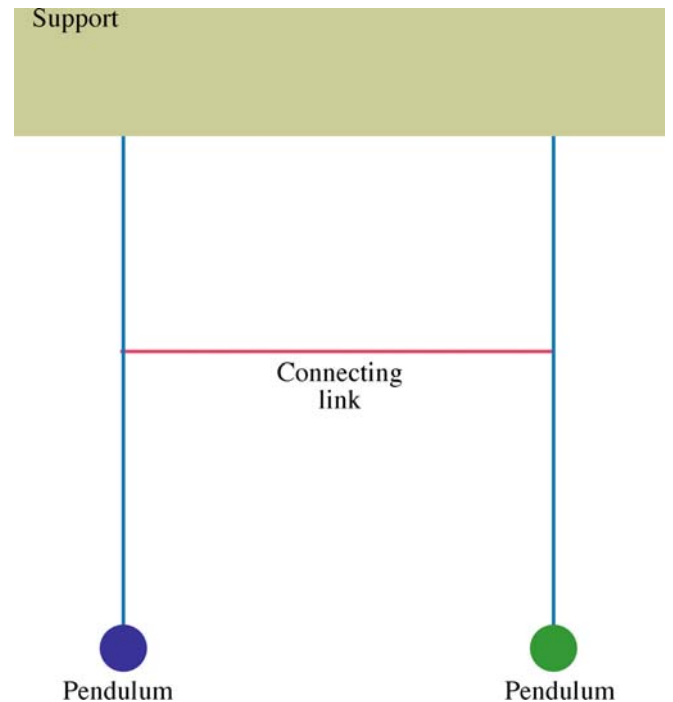


FIGURE 27.24. The coupled pendulum: a thought experiment.

27.9.B Bloch Waves

We discussed Bloch waves in Chapter 14. The difficulty in applying Bloch-wave analysis to the WB situation is that we are now usually interested in defects while Bloch waves are properties of the perfect crystal. However, we can offer some basic comments. For the reflection \mathbf{g} to give a WB image, $|\phi_{\mathbf{g}}|$ must be much smaller than unity in the regions of perfect crystal but, in strained regions, a change $\Delta\psi^{(j)}$ in the amplitude of the Bloch wave j can give rise to a change $\Delta\phi_{\mathbf{g}}$. Cockayne has shown that the appreciable contrast which can then be present in the WB image is due, in the two-beam approximation, to the interband scattering from Bloch wave 1 to Bloch wave 2. In the general case, the scattering is from the branch of the dispersion surface with the largest $\psi^{(j)}$ to those branches with the largest $C_{\mathbf{g}}^{(j)}$, i.e., from the Bloch wave with the largest amplitude to the one which is most strongly excited. The dispersion surface for the $\mathbf{g}(3\mathbf{g})$ diffraction geometry is shown in Figure 27.25. It's an instructive exercise to reread this paragraph thinking how each statement relates to this figure and to consider other diffraction geometries, e.g., $\mathbf{0}(2\mathbf{g})$.

The Bloch-wave analysis of the problem leads to two further points which simplify the interpretation of WB images

- The diffraction conditions should be such that only one interband scattering process is important.
- In the two-beam approximation, in order for the image peaks to show sufficient contrast, it is generally found that $w (= s\xi_{\mathbf{g}})$ is greater than ~ 5 .

You can satisfy the first requirement by ensuring that no reflections are strongly excited. The second condition is usually already satisfied because of the more stringent requirement that s should be greater than 0.2 nm^{-1} . For example, for a $\{2\bar{2}0\}$ reflection in copper with 100-keV electrons, w is automatically greater than 8 since $\xi_{\mathbf{g}}$ is 42 nm.

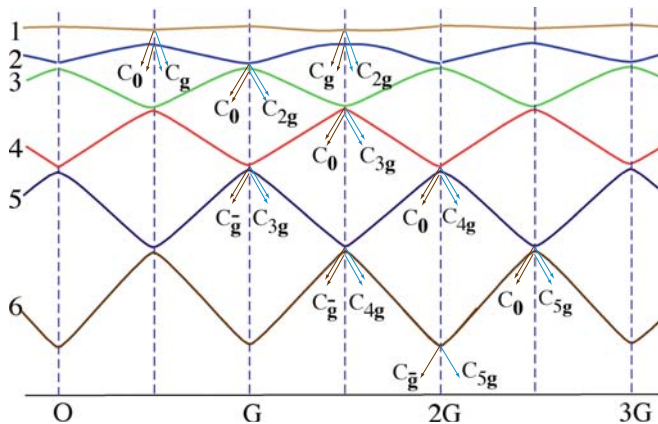


FIGURE 27.25. Dispersion-surface construction which is used to discuss the $\mathbf{g}(3\mathbf{g})$ geometry. The BZB is then at $1.5\mathbf{G}$ and tells you which reflections are strongly coupled; see Figure 15.9.

The $\mathbf{g}(3.1\mathbf{g})$ condition may also be preferable to $\mathbf{g}(-1.1\mathbf{g})$ on theoretical grounds if images are to be compared with computed profiles made using the column approximation; i.e., it simplifies your interpretation. The basis for this suggestion is that the region of the dispersion surface from which the scattering occurs is flatter for the $\mathbf{g}(3.1\mathbf{g})$ diffraction geometry than for the $\mathbf{g}(\bar{\mathbf{g}})$ case.

27.9.C If Other Reflections are Present

Several times in the previous discussion, we have said that no reflections should be strongly excited. Much of our thinking has been based on the two-beam approximation we introduced in Chapter 13. When you are using WB conditions, you must be even more careful. Consider the $\mathbf{g}, 3\mathbf{g}$ geometry shown in Figure 27.5. We form the WB image using reflection \mathbf{g} so electrons are weakly scattered from the O beam into the G beam. However, once in the G beam, they can be strongly scattered into the $2\mathbf{G}$ beam. We can picture this process by drawing the new Ewald sphere for the 'new' incident beam, G; this sphere passes through $2\mathbf{G}$!

For the mathematically inclined, you can go back to the many-beam equations, which we introduced briefly in Chapter 14. The coupling of beams \mathbf{g} and \mathbf{h} is determined by $(s_{\mathbf{g}} - s_{\mathbf{h}})$ and has an extinction distance given by $(\xi_{\mathbf{g}-\mathbf{h}})$. If $s_{\mathbf{g}}$ and $s_{\mathbf{h}}$ are equal, then the coupling between these beams will be strong. Furthermore, the characteristic length for the coupling in this example will be $\xi_{2\mathbf{g}-\mathbf{g}}$ or $\xi_{\mathbf{g}}$ which is what you would have guessed by thinking about Figure 27.5A and C.

27.9.D The Future Is Now

Several new developments are changing how we practice the WB technique. The main point here is that you should remember the principles because they will not change.

- Slow-scan CCD cameras give a very linear response and therefore make quantitative analysis of WB images possible. For this to happen, computer modeling of the defect and simulation of the image will be needed. We will return to this topic in Chapter 30 but mainly for HRTEM.
- An FEG and energy filtering allow us either to minimize the effect of variations in the energy (decrease diffuse scattering) or to form WB images using particular sections of the energy-loss spectrum. To make use of this extra information, we need to extend the theory.
- Image processing and frame averaging should allow us to reduce the noise, aid quantification, and use WB microscopy for in-situ studies.

Image simulation, as we described in Section 26.11, will allow us to be more quantitative in our interpretation of WB images when we actually have quantitative recording.

CHAPTER SUMMARY

The basic idea of the WB technique is very simple: using a large value of s gives a small ξ_{eff} and hence a narrow image of most defects, since the width of a dislocation is related to $\xi_{\text{eff}}/3$. What you should remember is that the value of s for a particular diffraction condition $\mathbf{g}(n\mathbf{g})$ depends not only on n and \mathbf{g} , but also on the lattice parameter of the crystal and the wavelength of the electrons. You will see the ‘magic number’ $s = 0.2 \text{ nm}^{-1}$ quoted often.

- Remember that this number gives a rule of thumb if you want to do quantitative analysis. It does not usually correspond to $\mathbf{g}(3\mathbf{g})$.
- Don’t use the $\mathbf{g}(3\mathbf{g})$ condition without calculating the value of $\mathbf{s}_{\mathbf{g}}$.
- The term $\mathbf{s}\cdot\mathbf{R}$ has been neglected in this analysis; we usually assume that the deformable model from Section 25.13 is valid.

You can often get all the information you need with less effort using a somewhat smaller value of s . As always, the longer you take to perfect the image, the more likely you are to alter your specimen, especially the defect structure.

Finally, remember that the diffracted beam travels parallel to \mathbf{k}_D . Therefore, the image of any defect is also projected in this direction. Even though the Bragg angle is small, this means, for example, that the apparent separation of defects in the image may not be equal to their horizontal separation relative to their glide planes if the defects are located at different heights in the specimen. This projection error can vary depending on the \mathbf{g} and s used to form the image and the orientation of the specimen.

THE TECHNIQUE

Cockayne, DJH 1972 *A Theoretical Analysis of the Weak-Beam Method of Electron Microscopy [Defectoscopy]* Z. Naturf. **27a**, 452–460. Derives equation 27.2.

Cockayne, DJH 1981 *Weak-Beam Electron Microscopy* Ann. Rev. Mater. Sci. **11**, 75–95. A review of WB.

Cockayne, DJH, Ray, ILF and Whelan, MJ 1969 *Investigations of Dislocation Strain Fields Using Weak Beams* Phil. Mag. **20**, 1265–1270. The original paper; derives equations 27.10 and 27.15.

Hirsch, PB, Howie, A and Whelan, MJ 1960 *A Kinematical Theory of Diffraction Contrast of the Electron Transmission Microscope Images of Dislocations and Other Defects* Proc. Roy. Soc. London A **252**, 499–529. Always worth another look.

Stobbs, WM 1975 *The Weak Beam Technique in Electron Microscopy in Materials Science*, vol. **II** (Eds. U Valdrè and E Ruedl), p. 591–646, CEC Brussels. A masterful review which also compared WB and a Stradivarius.

Stobbs, WM and Sworn, C 1971 *The Weak Beam Technique as Applied to the Determination of the Stacking-Fault Energy of Copper* Phil. Mag. **24**, 1365–1381. WB calculations using anisotropic elasticity.

SOME DEFECTS

Carter, CB 1984 *What’s New in Dislocation Dissociation?* in *Dislocations –1984* (Eds. P Veysseyre, L Kubin, and J Castaing), p. 227, Editions du CNRS Paris. A 50th birthday review of dissociated dislocations.

Carter, CB and Holmes, SM, 1975 *The Study of Faulted Dipoles in Copper Using Weak-Beam Electron Microscopy* Phil. Mag. **32** (3), 599–614.

Carter, CB, Mills, MJ, Medlin, DL and Angelo, JE 1995 *The 112 Lateral Twin Boundaries in FCC Metals in 7th International Conference Intergranular and Interphase Boundaries in Materials*, Lisbon, Portugal. The ‘thickness’ of a plane is important in WB! (Section 27.7.)

Föll, H, Carter, CB and Wilkens, M 1980 *Weak-Beam Contrast of Stacking Faults in Transmission Electron Microscopy* Phys. stat. sol. (A) **58**, 393–407. Discusses the anomalous WB contrast from SFs.

Gerthsen, D and Carter, CB 1993 *Stacking-Fault Energies of GaAs* Phys. stat. sol. (A) **136**, 29–43. An experimentalist’s comparison of WB and HRTEM.

Hazzledine, PM, Karnthaler, HP and Wintner, E 1975 *Non-parallel Dissociation of Dislocations in Thin Foils* Phil. Mag. **32**, 81–97. Used WB to show the unambiguous effect of surfaces on dislocation core splitting: a paper with broad implications for the microscopists.

Wilson, AR and Cockayne, DJH 1985 *Calculated Asymmetry for Weak Beam Intrinsic Stacking Fault Images* Phil. Mag. **A51**, 341–354. More on the WB contrast from SFs.

SELF-ASSESSMENT QUESTIONS

- Q27.1 Write down the equation relating s_{eff} to s .
- Q27.2 Write down the equation relating ξ_{eff} to $s_{\mathbf{g}}$.
- Q27.3 Write down the equation relating s to n , \mathbf{g} , and λ .
- Q27.4 If the $3\mathbf{g}$ reflection is excited, where are the \mathbf{g} and $3\bar{\mathbf{g}}$ Kikuchi lines?
- Q27.5 What is the optimum value of s for WB and why is this a compromise value?
- Q27.6 If an Si specimen is set to $\mathbf{g}, 3\mathbf{g}$ conditions at 100 kV, how wide will the image of a dislocation be?
- Q27.7 How will the image differ if we use 300 kV rather than 100 kV for $\mathbf{g}, 3\mathbf{g}$ conditions?
- Q27.8 Name the principal methods used for calculating WB images.
- Q27.9 What is the radius of the phasor diagram of a perfect crystal in the WB condition?
- Q27.10 Why is the separation of the two lines in a $\mathbf{g}, \mathbf{b}=2$ WB image not equal to the separation of the two partial dislocations which cause these lines?
- Q27.11 A WB image can show two lines when $\mathbf{g}, \mathbf{b}=2$ for a non-dissociated dislocation. How can this happen?
- Q27.12 Why is the circumference of the circle in a phasor diagram of a perfect crystal imaged under WB conditions interesting/instructive?
- Q27.13 Why might the contrast from an extrinsic stacking fault differ from that from an intrinsic SF under WB conditions?
- Q27.14 Sketch the phasor diagrams which show the difference behavior of a dislocation image in $\pm\mathbf{g}$.
- Q27.15 What extinction distance do you expect to have for the coupling of \mathbf{g} and $2\mathbf{g}$ when $3\mathbf{g}$ is excited?
- Q27.16 In WB images, we usually set $w \geq 5$. Why?
- Q27.17 Why do we avoid exciting other reflections when using the WB technique? Is it more important than when forming BF images?
- Q27.18 What is the classic value for $s_{\mathbf{g}}$ in WB images and why was this value originally chosen?
- Q27.19 Why is the condition with s positive preferred over that with s negative even though both conditions may have the same value of s ?
- Q27.20 Why is WB microscopy more difficult as we go to higher kV or to lower kV? (Higher and lower indicate relative to 100 kV.)

TEXT-SPECIFIC QUESTIONS

- T27.1 Starting from Figures 27.1 and 27.4, deduce equation 27.5.
- T27.2 Check Table 27.1 (any errors?) and add a column for 120-keV electrons. (Why?)
- T27.3 Consider the lowest part of Figure 27.7C (a line at and parallel to the bottom of the figure). Draw an accurate sketch of what the cross section of the sample might be and justify your answer.
- T27.4 If the sample in Figure 27.8 is close to the (111) plane, deduce the character of the dislocations in this image which lie in the (111) plane. Explain your answer very carefully.
- T27.5 If we have $\mathbf{g}(3\mathbf{g})$ conditions, through what angle must the planes near the dislocation bend in order to be in the Bragg condition in Figure 27.9?
- T27.6 Using equation 27.9, deduce the width of the dislocation image for $\mathbf{g}(3\mathbf{g})$ and $\mathbf{g}(4\mathbf{g})$ conditions if \mathbf{g} is $2\bar{2}0$ and the sample is parallel to the (111) plane.
- T27.7 Consider Figure 27.15. Deduce how many $\{111\}$ planes apart the continuous and dark (in D) SFs are. Explain all the assumptions and steps in your answers.
- T27.8 Consider Figure 27.19. Deduce the character of all of the perfect and partial dislocations in this image assuming that they all lie in the (111) plane.
- T27.9 Consider Figure 27.23. If the sample is (111) and the $\mathbf{g}(3\mathbf{g})$ conditions were used to form the image, deduce the thickness of the sample. Explain all your assumptions and arguments.
- T27.10 Consider Figure 27.21. What and where (indices and direction) is the most likely \mathbf{g} used to form this image? Explain your reasoning.
- T27.11 (An extra challenge question) Consider Figure 27.16. When will you see the left phasor and when will you see the one on the right, assuming you are looking at the same dislocation.
- T27.12 Determine the value of s for 220 taking 100 kV and Cu (Figure 27.2).
- T27.13 Figure 27.8. Estimate the value for the SFE. Assume a Cu fcc alloy, $\mathbf{g}, 3\mathbf{g}$ conditions.
- T27.14 Estimate the maximum thickness of the sample in Figure 27.13. Spinel with 220, $\mathbf{g}, 3\mathbf{g}$, 100 kV.
- T27.15 Figure 27.14. Phasor diagram-100 kV. For Cu, $\mathbf{g}=220$ and $\mathbf{g}, 3\mathbf{g}$ conditions. How thick is the sample?
- T27.16 Challenging. Derive equation 27.8 explaining all your assumptions.
- T27.17 Challenging: Figure 27.9 greatly exaggerates the curvature of the dislocations. Estimate the length of the column that contributes to \mathbf{g} for $\mathbf{g}, 3\mathbf{g}$ and 100-kV electrons.
- T27.18 What value of n would you use to produce WB images of dislocation for these reflections assuming 100-, 200-, and 300-kV electrons? (a) $(11\bar{2}0)$ in Al_2O_3 (alumina); (b) (220) in MgAl_2O_4 (spinel); (c) (111) in MgAl_2O_3 ; (d) (222) in YAG; (e) (220) in Si; (f) (111) in Si; (g) (110) in W; (h) (200) in Au.

- T27.19 If the specimen is oriented for $\mathbf{g},3\mathbf{g}$ (Cu, 100 kV) throughout, what angle must you tilt to set up $\mathbf{g},4\mathbf{g}$ conditions? (Answer in degrees and radians.)
- T27.20 Imagine four overlapping SFs in a crystal. What is the maximum intensity you could produce in the image? Give your answer as a multiple of the intensity possible for a perfect crystal. Use Cu 100 kV and $\mathbf{g},5\mathbf{g}$ for $\mathbf{g} = 111$.
- T27.21 Discuss how you might use weak-beam imaging to profile the topography of a thin sample even though the thickness can be changed by undulations on either surface. Hint: this question is related to electron tomography.
- T27.22 Discuss how weak-beam images will be improved using a new energy-filtered TEM equipped with a drift-compensated stage. Will you still use the same value of s ?
- T27.23 Professor J wants to use weak-beam imaging to obtain information on the core of dislocations. What factors will ultimately limit the information she can obtain? How will these limitations depend on the nature of the sample?

High-Resolution TEM

CHAPTER PREVIEW

We will now rethink what we mean by a TEM, in a way that is more suitable for HRTEM, where the purpose is to maximize the useful detail in the image. (Note the word *useful* here.) You should think of the microscope as an optical device that transfers information from the specimen to the image. The optics consists of a series of lenses and apertures aligned along the optic (symmetry) axis. What we would like to do is to transfer *all* the information from the specimen to the image, a process known as mapping. There are two problems to overcome and we can never be completely successful in transferring *all* the information. First, as you know from Chapter 6, the lens system is not perfect so the image is distorted and you lose some data because the lens has a finite size (Abbe’s theory). The second problem is we have to interpret the image using an atomistic model for the material. Ideally, this model will include a full description of the atomic potential and the bonding of the atoms, but we don’t know that either. We will also need to know exactly how many atoms the electron encountered on its way through the specimen. So most of our task will be concerned with finding the best compromise and producing models for the real situation. To conclude our discussion of the theory, we will introduce the language of *information theory*, which is increasingly used in HRTEM.

Perhaps the biggest challenge for HRTEM now is interpretation; getting an image that shows detail finer than 0.2 nm is now routine: correctly interpreting it may not be routine. This caution is particularly critical when you’re examining nanomaterials. We close the chapter with a review of the experimental applications of HRTEM to include periodic and non-periodic materials, mixtures of the two, or just single atoms.

A few words of caution: HRTEM is one of the most important aspects of TEM and is usually an essential consideration in obtaining funds for a new TEM. This chapter introduces some standard concepts and some ideas that are just beginning to be explored. Some topics are straightforward; others are exceedingly difficult.

28.1 THE ROLE OF AN OPTICAL SYSTEM

What the microscope does is to transform each point on the specimen into an extended region (at best, a circular disk) in the final image. Since each point on the specimen may be different, we describe the specimen by a specimen function, $f(x,y)$. The extended region in the image which corresponds to the point (x,y) in the specimen is then described as $g(x,y)$ as shown schematically in Figure 28.1; note that both f and g are functions of x and y .

If you consider two nearby points, A and B, they will produce two overlapping images g_A and g_B as shown in Figure 28.2. If we extend this argument, we can see that each point in the image has contributions from many points in the specimen. We express this result mathematically by

$$g(\mathbf{r}) = \int f(\mathbf{r}')h(\mathbf{r} - \mathbf{r}')d\mathbf{r}' \tag{28.1}$$

$$= f(\mathbf{r}) \otimes h(\mathbf{r} - \mathbf{r}') \tag{28.2}$$

POINT-SPREAD FUNCTION
 Since $h(\mathbf{r})$ describes how a point spreads into a disk, it is known as the point-spread function or smearing function, and $g(\mathbf{r})$ is called the convolution of $f(\mathbf{r})$ with $h(\mathbf{r})$.

Here $h(\mathbf{r}-\mathbf{r}')$ is a weighting term telling us how much each point in the specimen contributes to each point in the image.

Spence calls $h(\mathbf{r})$ the impulse response function, and notes that it can only apply to small patches of specimen

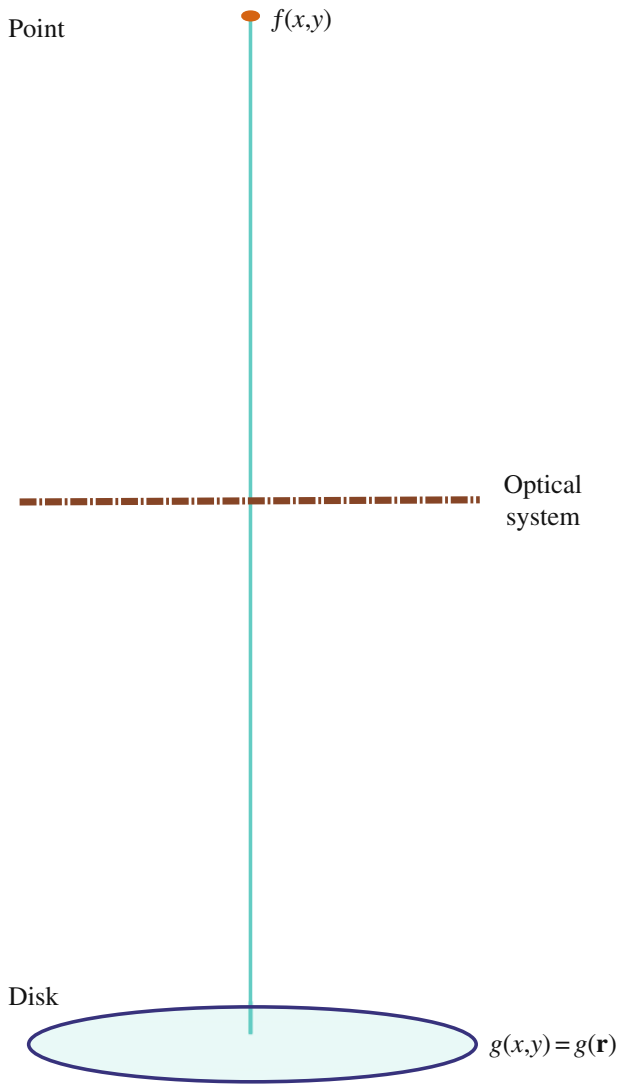


FIGURE 28.1. An optical system transforms a point in the specimen (described by $f(x,y)$) into a disk in the image described by $g(x,y)$. The intensity in the image at point (x,y) can be described by the function $g(x,y)$ or $g(\mathbf{r})$. It has a unique value for each value of (x,y) , so we say it is a representation of the image.

which lie in the same plane and are close to the optic axis. The symbol \otimes indicates that the two functions, f and h , are ‘folded together’ (multiplied and integrated) or ‘convoluted with one another.’

28.2 THE RADIO ANALOGY

We can compare this imaging process with the task of recording the sound of an orchestra on a record/tape/CD or even transmitting to the brain directly or via a radio. We want to hear the loud drum and quiet flute (large amplitude and small amplitude); we want to hear the high note on the violin and the low note on the double bass (high frequency and low frequency). Our audio amplifier has limits on both the low and high

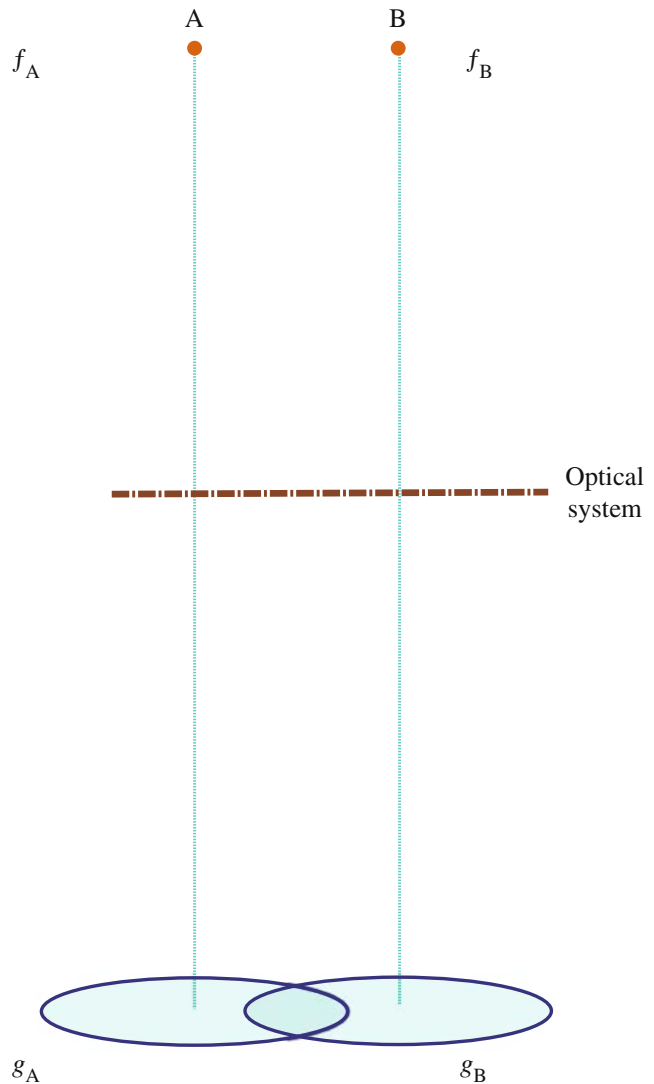


FIGURE 28.2. Two points, f_A and f_B , in the specimen produce two disks, g_A and g_B , in the image.

frequencies, so we won’t achieve perfect reproduction. The importance of amplitude is obvious as discussed in Chapter 22, but how do we define frequency in a TEM image? High frequency in audio is related to $1/t$; frequencies in lattice images are related to $1/x$. So the high spatial frequencies simply correspond to small distances. What we are looking for in high-resolution work are the high spatial frequencies. Notice our use of high/low and large/small.

SPATIAL FREQUENCY
High resolution requires high spatial frequencies.

Figure 28.2 shows two points A and B in the specimen and their disk images on the screen. We see disks (see our discussion of the Rayleigh disk in Chapter 6)

because the lens system is not perfect. We can also write $g(x,y)$, the intensity of an image at point (x,y) , as $g(\mathbf{r})$, and in the simplest case, these disks have uniform intensity. We can always represent any function in two dimensions as a sum of sine waves.

$$g(x,y) = \sum_{u_x, u_y} G(u_x - u_y) \exp(2\pi i(xu_x + yu_y)) \quad (28.3)$$

$$g(x,y) = \sum_{\mathbf{u}} G(\mathbf{u}) \exp(2\pi i \mathbf{u} \cdot \mathbf{r}) \quad (28.4)$$

Here \mathbf{u} is a reciprocal-lattice vector, the spatial frequency for a particular direction. We have expressed $g(\mathbf{r})$ in terms of a combination of the possible values of $G(\mathbf{u})$, where $G(\mathbf{u})$ is known as the Fourier transform of $g(\mathbf{r})$. We can now define two other Fourier transforms

$F(\mathbf{u})$ is the Fourier transform of $f(\mathbf{r})$,

and

$H(\mathbf{u})$ is the Fourier transform of $h(\mathbf{r})$.

FOURIER TRANSFORM

The Fourier transform of a function is an expression of that function as a ‘sum’ of frequencies; it is the frequency-domain representation of the function.

Since $h(\mathbf{r})$ tells us how information in real space is transferred from the specimen to the image, $H(\mathbf{u})$ tells us how information (or contrast) in \mathbf{u} space is transferred to the image.

$H(\mathbf{u})$ is the contrast transfer function.

Now these three Fourier transforms are related by

$$G(\mathbf{u}) = H(\mathbf{u})F(\mathbf{u}) \quad (28.5)$$

So a convolution in real space (equation 28.1) gives multiplication in reciprocal space (equation 28.5).

The factors contributing to $H(\mathbf{u})$ include

Apertures → The aperture function $A(\mathbf{u})$
 Attenuation of the wave → The envelope function $E(\mathbf{u})$
 Aberration of the lens → The aberration function $B(\mathbf{u})$

We write $H(\mathbf{u})$ as the product of these three terms

$$H(\mathbf{u}) = A(\mathbf{u})E(\mathbf{u})B(\mathbf{u}) \quad (28.6)$$

The aperture function says that the objective diaphragm cuts off all values of \mathbf{u} (spatial frequencies) greater than (higher than) some selected value governed by the radius of the aperture. The envelope function has the same effect but is a property of the lens itself, and so may be either more or less restricting than $A(\mathbf{u})$. $B(\mathbf{u})$ is usually expressed as

$$B(\mathbf{u}) = \exp(i\chi(\mathbf{u})) \quad (28.7)$$

The term $\chi(\mathbf{u})$ can be written as

$$\chi(\mathbf{u}) = \pi\Delta f\lambda u^2 + \frac{1}{2}\pi C_s\lambda^3 u^4 \quad (28.8)$$

We will give a simple (simplified) derivation of this equation in Section 28.6. It builds on the concepts we discussed in Chapter 6 when we examined the origin of C_s .

OVERFOCUS

$\Delta f > 0$ is known as overfocus. It means we have focused the objective lens on a plane above the specimen. (By above, we mean before the electrons reach the specimen; the story is the same if the microscope is upside down!)

Summarizing so far: High spatial frequencies correspond to large distances from the optic axis in the DP. The rays which pass through the lens at these large distances are bent through a larger angle by the objective lens. They are not focused at the same point by the lens, because of spherical aberration, and thus cause a spreading of the point in the image. The result is that the objective lens magnifies the image but confuses the fine detail. The resolution we require in HRTEM is limited by this ‘confusion’

- Each point in the specimen plane is transformed into an extended region (or disk) in the final image.
- Each point in the final image has contributions from many points in the specimen.

LINEAR

What we need for our analysis to ‘work’ is a ‘linear relationship between the image and the weak specimen potential.’

We now have to go back and look at how we can represent the specimen. That is, what is $f(\mathbf{r})$ in equation 28.1? (We’ll use the coordinates \mathbf{r} and x,y interchangeably in this discussion; the former is more compact but we can extend the latter notation to emphasize the possibility of a z component.)

28.3 THE SPECIMEN

Since we are using a TEM, we call the specimen function, $f(\mathbf{r})$, the specimen transmission function. Here you have to be very careful to remember that we are going to

use a model to represent the specimen and the model will make certain assumptions. A general model would describe $f(\mathbf{r})$ as

$$f(x, y) = A(x, y) \exp(-i\phi_t(x, y)) \quad (28.9)$$

where $A(x, y)$ is the amplitude (not the aperture function) and $\phi_t(x, y)$ is the phase which depends on the thickness of the specimen.

For our application to HRTEM, we simplify our model further by setting $A(x, y) = 1$; i.e., we set the incident wave amplitude to be unity. We can show that the phase change only depends on the potential $V(x, y, z)$ which the electron sees as it passes through the specimen (by following Van Dyck's argument). We will assume that the specimen is so thin that we can write down a projected potential $V_t(x, y)$ with t being the thickness of the specimen, as usual.

$$V_t(x, y) = \int_0^t V(x, y, z) dz \quad (28.10)$$

What we are doing is creating a two-dimensional projection of the crystal structure; this approach is critical to much of our interpretation of HRTEM images.

We can relate the wavelength, λ , of the electrons in vacuum to the energy. (Ideally, λ should have its relativistic value, but the principle is correct.)

$$\lambda = \frac{h}{\sqrt{2meE}} \quad (28.11)$$

(We'll give the analysis in a simple non-relativistic form for simplicity.) When the electrons are in the crystal, λ is changed to λ'

$$\lambda' = \frac{h}{\sqrt{2me(E + V(x, y, z))}} \quad (28.12)$$

so we can say that, when passing through a slice of material of thickness dz , the electrons experience a phase change given by

$$d\phi = 2\pi \frac{dz}{\lambda'} - 2\pi \frac{dz}{\lambda} \quad (28.13)$$

$$d\phi = 2\pi \frac{dz}{\lambda} \left(\frac{\sqrt{E + V(x, y, z)}}{\sqrt{E}} - 1 \right) \quad (28.14)$$

$$d\phi = 2\pi \frac{dz}{\lambda} \left(\left(1 + \frac{V(x, y, z)}{E} \right)^{\frac{1}{2}} - 1 \right) \quad (28.15)$$

$$d\phi = 2\pi \frac{dz}{\lambda} \frac{1}{2} \frac{V(x, y, z)}{E} \quad (28.16)$$

$$d\phi = \frac{\pi}{\lambda E} V(x, y, z) dz \quad (28.17)$$

$$d\phi = \sigma V(x, y, z) dz \quad (28.18)$$

So the total phase shift is dependent only on $V(x, y, z)$ since

$$d\phi = \sigma \int V(x, y, z) dz = \sigma V_t(x, y) \quad (28.19)$$

where $V_t(x, y)$ is the potential projected in the z -direction.

We call σ the interaction constant. (See more discussion of σ in Chapter 3 but be careful—all σ s are not the same.) This σ tends to a constant value as V increases, since the energy of the electron is proportional to E or λ^{-1} (i.e., changes in the two variables, λ and E , tend to compensate for one another).

THE INTERACTION CONSTANT

This σ is not the stress or the scattering cross section. It is another elastic interaction.

Now, we can take account of absorption by including a function $\mu(x, y)$ so that our specimen transfer function $f(x, y)$ is now given by

$$f(x, y) = \exp[-i\sigma V_t(x, y) - \mu(x, y)] \quad (28.20)$$

The effect of this model is that, apart from $\mu(x, y)$, we have represented the specimen as a 'phase object.' This is known as the phase-object approximation or POA. We are actually lucky because the absorption will usually be small in the regime where the rest of the approximation holds.

POA

In general, the phase-object approximation only holds for thin specimens.

We can simplify the model further if the specimen is *very* thin, so that $V_t(x, y)$ is $\ll 1$. Then we expand the exponential function, neglecting μ and higher-order terms, so that $f(x, y)$ becomes

$$f(x, y) = 1 - i\sigma V_t(x, y) \quad (28.21)$$

Now we have reached the weak phase-object approximation or the WPOA. We see that the WPOA essentially says that, for a very thin specimen, the amplitude of a transmitted wave function will be linearly related to

the projected potential of the specimen. Note that in this model the projected potential is taking account of variations in the z -direction, and is thus very different for an electron passing through the center of an atom compared to one passing through its outer regions.

Fortunately, there are software packages that allow us to calculate what an image will look like for a particular specimen geometry. However, you must always remember that a model has been used to represent the specimen and have a clear understanding of its limits. To emphasize this last point, bear in mind that the WPOA fails for an electron wave passing through the center of a single uranium atom! An atomic layer of U would be too thick for the WPOA. As a second example, Fejes has shown that, for the complex oxide $Ti_2Nb_{10}O_{27}$, the WPOA is only valid if the specimen thickness is <0.6 nm! The good news is that the approach appears to be more widely applicable than these particular estimates would suggest.

28.4 APPLYING THE WPOA TO THE TEM

So far our treatment has been quite general, but now we use our WPOA model. If we use the expression for $f(\mathbf{r})$ given by equation 28.21, then equation 28.2 tells us that the wave function as seen in the image is given by

$$\psi(x, y) = [-i\sigma V_t(x, y)] \otimes h(x, y) \quad (28.22)$$

If we represent $h(x, y)$ as $\cos(x, y) + i \sin(x, y)$, then $\psi(x, y)$ becomes

$$\psi(x, y) = 1 + \sigma V_t(x, y) \otimes \sin(x, y) - i\sigma V_t(x, y) \otimes \cos(x, y) \quad (28.23)$$

As usual, the intensity is given by

$$I = \psi\psi^* = |\psi|^2 \quad (28.24)$$

Multiplying this out and neglecting terms in σ^2 (because σ is small), we find that

$$I = 1 + 2\sigma V_t(x, y) \otimes \sin(x, y) \quad (28.25)$$

Knowing this result we can say that, in the WPOA, only the imaginary part of $B(\mathbf{u})$ in equation 28.7 contributes to the intensity in equation 28.24 (because it gives the imaginary part of $h(x, y)$). Thus, we can set $B(\mathbf{u}) = 2 \sin \chi(\mathbf{u})$ rather than $\exp(i\chi(\mathbf{u}))$. (Notice the 2!)

We can now define a new quantity, $T(\mathbf{u})$, which we could call the intensity transfer function to distinguish it from $H(\mathbf{u})$. It's given by

$$T(\mathbf{u}) = A(\mathbf{u})E(\mathbf{u})2 \sin \chi(\mathbf{u}) \quad (28.26)$$

Note that $T(\mathbf{u})$ is not identical to $H(\mathbf{u})$, which we defined in equation 28.6. The '2' in equation 28.26 is the '2' in

equation 28.25 and arises because we are interested in the intensity in the beam, and therefore we multiplied ψ by its complex conjugate in equation 28.24. You may also see authors use a negative sign in equation 28.26 (in particular, in Reimer's text). This has the effect of inverting the graph of $B(\mathbf{u})$ versus \mathbf{u} and making $B(\mathbf{u}) > 0$ for positive phase contrast.

A note on terminology. You will often see $T(\mathbf{u})$ rather than $H(\mathbf{u})$ called the contrast transfer function in the HRTEM literature. This terminology comes from the analysis of the imaging process for incoherent light in visible-light optics. With coherent illumination, $T(\mathbf{u})$ and $H(\mathbf{u})$ are identical. The smearing function (point-spread function) for that case is the Fourier transform of the CTF. The equation describing $T(\mathbf{u})$ was derived for the situation where we have *coherent* imaging. For incoherent light the smearing function would be

$$\cos^2(x, y) + \sin^2(x, y) \quad (28.27)$$

which is just unity.

So the CTF in HRTEM would be different from $T(\mathbf{u})$, and therefore we will call $T(\mathbf{u})$ the *objective lens transfer function*.

28.5 THE TRANSFER FUNCTION

You must note two things here. First, as we just said, the transfer function, $T(\mathbf{u})$, formulation applies to any specimen, and second, $T(\mathbf{u})$ is *not* the CTF of HRTEM. The problem with this formulation is that the image wave function is not an observable quantity! What we observe in an image is contrast, or the equivalent in optical density, current readout, etc., and this is not linearly related to the object wave function. Fortunately, there is a linear relation involving observable quantities under the special circumstances, where the specimen acts as a WPO.

TRANSFER FUNCTION

When $T(\mathbf{u})$ is negative, positive phase contrast results, meaning that atoms would appear dark against a bright background. When $T(\mathbf{u})$ is positive, negative phase contrast results, meaning that atoms would appear bright against a dark background. When $T(\mathbf{u}) = 0$, there is no detail in the image for this value of \mathbf{u} . (Note that we assume here that $C_s > 0$.)

If the specimen acts as a weak-phase object, then the transfer function $T(\mathbf{u})$ is sometimes called the CTF, because there is no amplitude contribution, and the output of the transmission system is an observable quantity (image contrast). The transfer function appropriate for this image formation process has the form

which we derived above (equation 28.26) and if we ignore $E(\mathbf{u})$,

$$T(\mathbf{u}) = 2A(\mathbf{u}) \sin \chi(\mathbf{u}) \quad (28.28)$$

where we know that $A(\mathbf{u})$ is the aperture function and might call $\chi(\mathbf{u})$ the phase-distortion function.

$\chi(\mathbf{u})$

In other words, the phase-distortion function has the form of a phase shift expressed as $2\pi/\lambda$ times the path difference traveled by those waves affected by spherical aberration (C_s), defocus (Δz), and astigmatism (C_a).

Assuming that astigmatism can be properly corrected, the phase-distortion function is the sum of two terms. If the CTF is now compared to the phase-distortion function, a number of observations can be made. Note that the CTF is oscillatory; there are ‘bands’ of good transmission separated by ‘gaps’ (zeros) where no transmission occurs.

The CTF shows maxima (meaning maximum transfer of contrast) whenever the phase-distortion function assumes multiple odd values of $\pm\pi/2$. Zero contrast occurs for $\chi(\mathbf{u}) = \text{multiple } \pm\pi$.

The reason that negative $T(\mathbf{u})$ gives positive phase contrast is that there is a phase shift of $-\pi/2$ due to diffraction. If a diffracted beam is further phase shifted by $-\pi/2$, it subtracts amplitude from the forward scattered beam, causing atoms to appear dark (positive contrast). If the same beam is instead phase shifted by $+\pi/2$, it adds amplitude to the forward scattered beam (they are ‘in phase’), causing atoms to appear bright (negative contrast).

28.6 MORE ON $\chi(\mathbf{u})$, $\sin \chi(\mathbf{u})$, AND $\cos \chi(\mathbf{u})$

The ideal form of $T(\mathbf{u})$ would be a constant value as \mathbf{u} increases, as shown in Figure 28.3; $T(\mathbf{u})$ must be zero at $\mathbf{u} = 0$ but, since small values of \mathbf{u} correspond to very large values of x (i.e., long distances in the specimen), this is not a problem. If $T(\mathbf{u})$ is large, it means that information with a periodicity or spatial frequency corresponding to that value of \mathbf{u} will be strongly transmitted, i.e., it will appear in the image. What we then need is that the different values of \mathbf{u} give the same contrast. Then all the atoms in a crystal appear as black spots, say, rather than some as black spots and others as white spots; if the latter occurs, interpretation will be difficult!

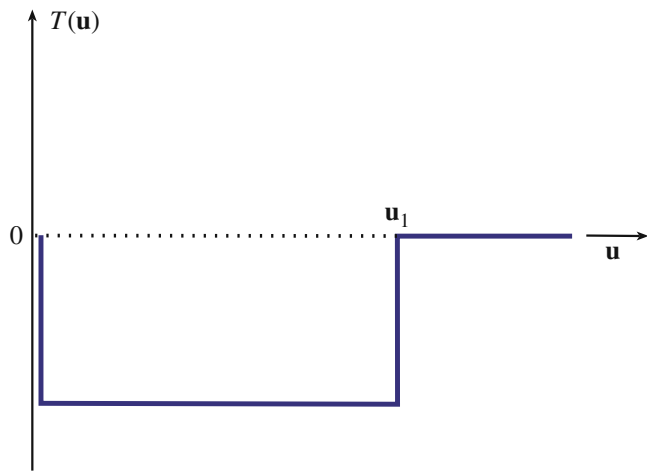


FIGURE 28.3. The ideal form of the transfer function, $T(\mathbf{u})$. In this example, $T(\mathbf{u})$ is large and negative between $\mathbf{u} \neq 0$ and $\mathbf{u} = \mathbf{u}_1$.

USEFUL \mathbf{u}

$T(\mathbf{u})$ becomes zero again at $\mathbf{u} = \mathbf{u}_1$; what we would like is for \mathbf{u}_1 to be as large as possible. If $T(\mathbf{u})$ crosses the \mathbf{u} -axis the sign of the transfer function reverses. This means that \mathbf{u}_1 defines the limit at which our image may be quite directly interpreted; it is a very important parameter.

We will now go through a simple exercise to produce an expression for $\chi(\mathbf{u})$. If we combine the effects of the spherical aberration (equation 6.14) and the defocus (equations 11.18) of the objective lens, we find that a point at the specimen will actually be imaged as a disk with radius $\delta(\theta)$.

$$\delta(\theta) = C_s \theta^3 + \Delta f \theta \quad (28.29)$$

Due to the spherical aberration of the objective lens and the finite value of Δf , the rays which pass through the objective lens at angle θ are not focused on the Gaussian image plane. If we only had one value of θ , we would still be all right! Of course, we have a range of values, so we average (integrate) these with respect to θ to give

$$D(\theta) = \int_0^\theta \delta(\theta) d\theta = \frac{C_s \theta^4}{4} + \Delta f \frac{\theta^2}{2} \quad (28.30)$$

Now, Bragg’s law tells us that

$$2d \sin \theta_B = n\lambda \quad (28.31)$$

or since θ_B is small

$$2\theta_B \cong \lambda g \quad (28.32)$$

So, we can replace θ in equation 28.30 by λu where \mathbf{u} is a general reciprocal-lattice vector. (Remember that the scattering angle is $2\theta_B$, not θ_B .)

We are interested in the phase $\chi(\mathbf{u})$, so we write

$$\chi(\mathbf{u}) = \text{phase} = \frac{2\pi}{\lambda} D(\mathbf{u}) = \frac{2\pi}{\lambda} \left(C_s \frac{\lambda^4 u^4}{4} + \Delta f \frac{\lambda^2 u^2}{2} \right) \quad (28.33)$$

and we have

$$\chi = \pi \Delta f \lambda u^2 + \frac{1}{2} \pi C_s \lambda^3 u^4 \quad (28.34)$$

which we quoted as equation 28.8. Clearly $\sin \chi(\mathbf{u})$ will be a complicated curve which will depend on the values of C_s (the lens quality), λ (the accelerating voltage), Δf (the defocus value you choose to form the image), and \mathbf{u} (the spatial frequency). This topic is addressed in a novel way in the companion text; see also Section 3.3 of John Spence's book; most of us just start with equation 28.34 and this reasonable justification.

The best way to appreciate the importance of χ is to use one of the simulation packages discussed in Chapter 30 and vary each of the parameters one by one. The plot of $T(\mathbf{u}) (= 2 \sin \chi)$ versus \mathbf{u} , shown in Figure 28.4, illustrates the main features. The curve has been drawn for $C_s = 1 \text{ mm}$, $E_0 = 200 \text{ kV}$, and a defocus value of -58 nm .

The important features of this curve are shown in Figures 28.4–28.6

- $\sin \chi$ starts at 0 and decreases. When \mathbf{u} is small, the Δf term dominates.
- $\sin \chi$ first crosses the \mathbf{u} -axis at \mathbf{u}_1 and then repeatedly crosses the \mathbf{u} -axis as \mathbf{u} increases.
- χ can continue forever but, in practice, it is modified by other functions which we discuss in Section 28.8.

Once you've selected your microscope and its objective lens, you have fixed C_s (unless you have a TEM with a C_s corrector); C_s does depend to some extent on the λ you choose. The curve of $T(\mathbf{u})$ versus \mathbf{u} does not depend on your specimen. Figure 28.5 shows a series of $\sin \chi$ curves

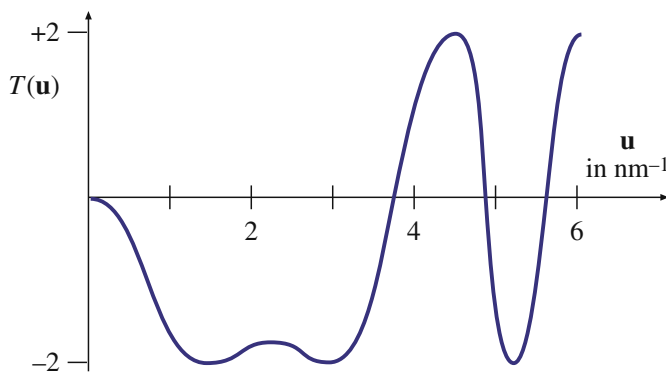


FIGURE 28.4. A plot of $T(\mathbf{u})$ versus \mathbf{u} ($C_s = 1 \text{ mm}$, $E_0 = 200 \text{ kV}$, $\Delta f = -58 \text{ nm}$).

for an imaginary 200-kV microscope where C_s has been changed. In each case, the 'best' curve (we'll discuss this in a moment) has been chosen. You can appreciate that the smaller C_s values give the larger \mathbf{u}_1 values; so a small C_s means we can achieve a higher spatial resolution.

C_s , Δf , AND β

High spatial frequencies \Rightarrow large diffraction angles \Rightarrow larger effect of objective lens (C_s).

So for a large objective aperture semi-angle β , the β^4 term wins, i.e., C_s wins. We can all vary Δf ; some can vary C_s too.

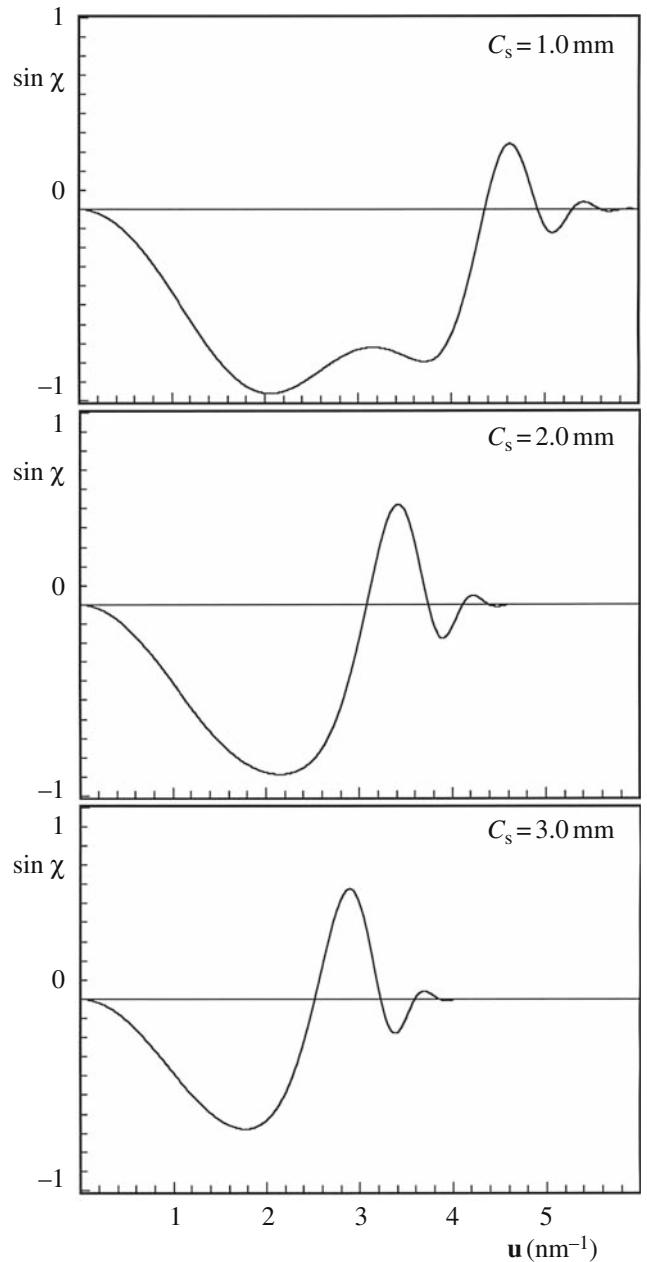


FIGURE 28.5. A series of $\sin \chi$ curves calculated for different values of C_s . Remember $2 \sin \chi = T(\mathbf{u})$. ($E_0 = 200 \text{ kV}$, $\Delta f = -60 \text{ nm}$).

28.7 SCHERZER DEFOCUS

The presence of zeros in the CTF means that we have gaps in the output spectrum which do not contribute to the output signal: it's as if these frequencies were filtered out. Obviously, the best transfer function is the one with the fewest zeros, which would be the case for a perfect lens, for example. What Scherzer did back in 1949 was to notice that the CTF could be optimized by balancing the effect of spherical aberration against a particular negative value of Δf . This value has come to be known as 'Scherzer defocus,' Δf_{Sch} which occurs at

$$\Delta f_{\text{Sch}} = -1.2(C_s \lambda)^{1/2} \quad (28.35)$$

At this defocus (which we'll derive below) all the beams will have nearly constant phase out to the 'first crossover' of the zero axis. This crossover point is defined as the instrumental resolution limit. This is the best performance that can be expected from a microscope unless we use image processing schemes to extract more information. In other words, this is not the information limit but it is the limit where we can use nearly intuitive arguments to interpret what we see. Again, as we discussed in Chapter 6 when we defined image resolution, you will see other authors give different values for the constant rather than the 1.2 given in equation 28.34; remember that this number is a calculated value, so it does depend on the details of your approximations.

This definition of resolution has new implications. The Rayleigh criterion which we used in Chapter 6, was only concerned with the ability to distinguish closely spaced point objects by eye. Our new definition requires a flat response in the object spectrum, and the goal is to have as many beams as possible being transferred through the optical system with identical phase, i.e., within the flat response regime. This is the underlying principle governing phase-contrast imaging in HRTEM.

DETAIL AND INFORMATION

A TEM image with detail of 0.66 Å was demonstrated in 1970 when the *interpretable* resolution was about 3.3 Å. So just because you can see detail in the image does not mean that you can gain useful information about your specimen.

The closest we can get to the ideal curve in Figure 28.6 occurs when $\chi(\mathbf{u})$ is close to -120° ; then $\sin \chi$ will be near -1 when χ is between -120° and -60° . We know that when $\chi = \pi$, $\sin \chi = 0$ so we want $\sin \chi$ to be as large as possible over a large range of \mathbf{u} ; $\sin \chi$ will be a nearly flat function if $d\chi/du$ is zero. So we look for the value of

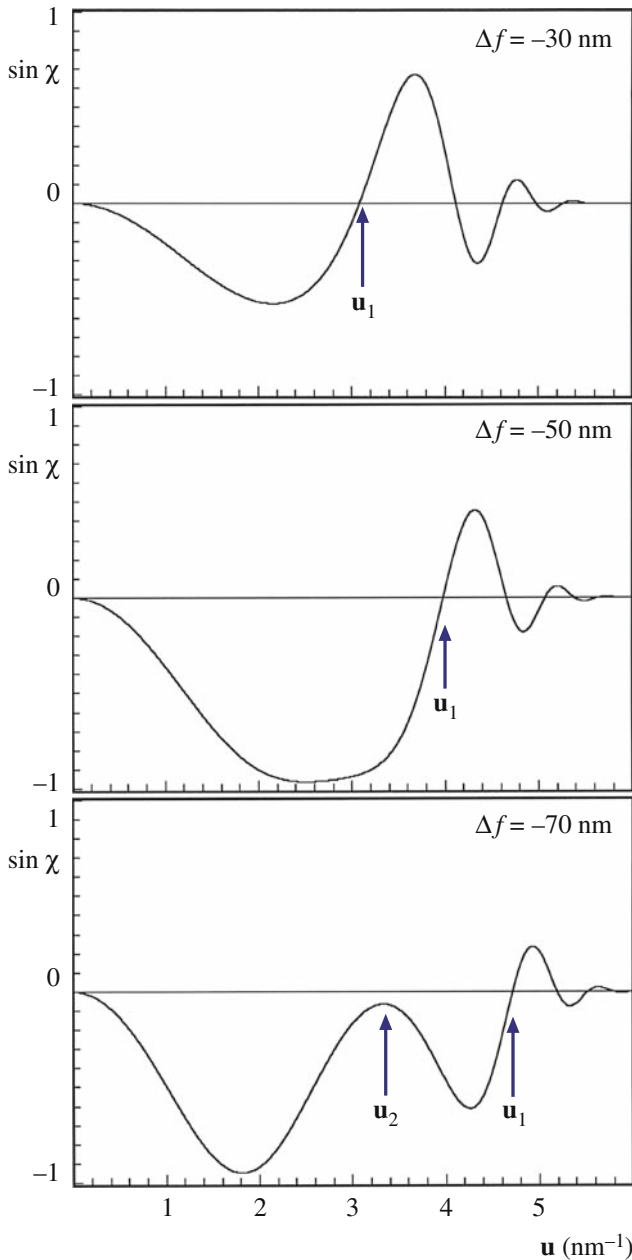


FIGURE 28.6. A series of $\sin \chi$ curves calculated for different values of Δf . ($E_0 = 200$ kV; $C_s = 1.0$ mm.)

If instead we fix C_s and again plot the best curves, but this time varying λ , you can see that the smallest value of λ allows us to achieve a higher spatial resolution. The result is not surprising; we want a small C_s and a small λ or a high voltage. So we choose the microscope to optimize C_s and λ . Now we only have Δf to vary. The set of curves shown in Figure 28.6 illustrates the effect of varying Δf . Notice how the bump in the curve at u_2 will eventually increase as Δf increases until it crosses the \mathbf{u} -axis so that u_1 is suddenly much smaller. If we just make Δf smaller, then u_1 steadily decreases. In the next section we will discuss the optimum value for Δf .

Δf when $d\chi/du$ is zero and χ is -120° (you should consider why we choose this value of χ). Differentiating equation 28.29 gives

$$\frac{d\chi}{du} = 2\pi\Delta f\lambda u + 2\pi C_s\lambda^3 u^3 \quad (28.36)$$

Set the left-hand term equal to 0

$$0 = \Delta f + C_s\lambda^2 u^2 \quad (28.37)$$

When $\chi = -120^\circ$, equation 28.29 becomes

$$-\frac{2\pi}{3} = \pi\Delta f\lambda u^2 + \frac{1}{2}\pi C_s\lambda^3 u^4 \quad (28.38)$$

Combining equations 28.37 and 28.38 gives a special value for Δf

$$\Delta f_{\text{Sch}} = -\left(\frac{4}{3}C_s\lambda\right)^{\frac{1}{2}} \quad (28.39)$$

The subscript denotes the Scherzer defocus value. Since $(1.33)^{1/2} = 1.155$ (~ 1.2), we have deduced equation 28.35. At this value of Δf (in equation 28.37) we find by substituting Δf_{Sch} into equation 28.37 (and using $(1.155)^{1/2} = 1.51$) that we next cross the axis at

$$u_{\text{Sch}} = 1.51 C_s^{-\frac{1}{4}} \lambda^{-\frac{3}{4}} \quad (28.40)$$

The resolution at the Scherzer defocus can then be defined as the reciprocal of u_{Sch} .

$$r_{\text{Sch}} = \frac{1}{1.51} C_s^{\frac{1}{4}} \lambda^{\frac{3}{4}} = 0.66 C_s^{\frac{1}{4}} \lambda^{\frac{3}{4}} \quad (28.41)$$

You will often see this expression with different values for the constant (0.66) for reasons discussed back in Section 6.6.B (here we are essentially summing the effects of Δf and C_s). The value of the constant can be increased, thus lowering r_{Sch} (i.e., giving higher resolution) if we are less restrictive about the value we choose for χ .

The quantities $(C_s\lambda)^{\frac{1}{2}}$ and $(C_s\lambda^3)^{\frac{1}{4}}$ seen in equations 28.39 and 28.41 are so important in HRTEM that Hawkes has designated them to be the units 1 Sch and 1 Gl (the scherzer and the glaser) in honor of two of the most noted pioneers of HRTEM. Notice that these *units* vary depending on the microscope you're using!

You'll find it interesting to plot the phase shift due to varying of Δf and C_s using EMS (Section 1.6). An excellent, though advanced, discussion of such diagrams is given by Thon, who describes how they can be used to design phase plates for the TEM. Spence shows how you can use a plot of nu^{-2} versus u^2 to help you determine experimental values of Δf and C_s ; see Figure 31.6A.

28.8 ENVELOPE DAMPING FUNCTIONS

The plots of $\chi(\mathbf{u})$ as a function of \mathbf{u} could extend out as far as you want to plot them. In practice, they don't because of the envelope damping function. In other words, the $\chi(\mathbf{u})$ plot stops where it does because the microscope is incapable of imaging the finest detail due to reasons other than the simple transfer characteristics of a linear system.

We know from Chapters 5 and 6 that resolution is also limited by the spatial coherence of the source and by chromatic effects. We can include these effects in our analysis of images by imposing an envelope function on the transfer function. The result is that higher spatial frequencies that might normally pass through higher-order windows are in fact damped out, as shown in the plot in Figure 28.7B.

The exact mathematical form of these envelope functions is complex. In general, the result is described by multiplying the (objective lens) transfer function $T(\mathbf{u})$ by both the chromatic aberration envelope E_c and the spatial coherence envelope E_a to yield an effective transfer function $T_{\text{eff}}(\mathbf{u})$.

$$T_{\text{eff}}(\mathbf{u}) = T(\mathbf{u})E_cE_a \quad (28.42)$$

The effect of the envelope functions is to impose a virtual aperture in the back focal plane of the objective

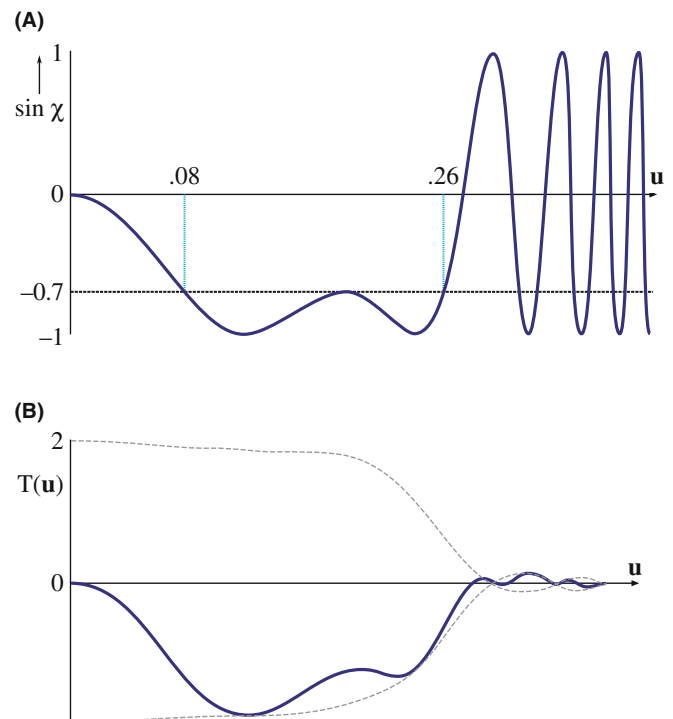


FIGURE 28.7. (A) $\sin \chi(\mathbf{u})$ versus \mathbf{u} without damping of the higher spatial frequencies. (B) $T(\mathbf{u})$ versus \mathbf{u} modified by the damping envelope (dashed line); $\Delta f = -100$ nm, $C_s = 2.2$ nm.

lens, *regardless* of the setting of focus. If we are going to use a physical aperture to remove unwanted noise, we should make it no larger than the ‘virtual aperture’ present due to this envelope. The presence of this virtual aperture means that higher-order passbands are simply not accessible. This cut-off thus imposes a new resolution limit on the microscope. This is what we earlier called the ‘information retrieval limit’ or simply the ‘information limit.’

If we keep these restrictions in mind then we can say that, up to the instrumental resolution limit, phase-contrast images are directly (i.e., intuitively) interpretable; this limit is set by the crossover at Scherzer defocus or the envelope function, i.e., whichever equals zero first. If the information limit is beyond the Scherzer resolution limit, we need to use image-simulation software (see Chapter 30) to interpret any detail beyond the Scherzer limit.

So you can image columns of atoms along the incident-beam direction and their positions are faithfully rendered with respect to one another up to Scherzer resolution. If the microscope is operated at different defocus values, the crossovers in the transfer function make image interpretation more indirect and you have to resort to using computer simulation.

28.9 IMAGING USING PASSBANDS

Because of the focus dependence of the CTF, you, the microscope operator, have control over its overall form. For example, the worst case of contrast transfer is where all contrast is minimized. This minimum contrast (MC) defocus condition (Δf_{MC}) is also known as the dark-field focus condition in STEM imaging and occurs for a special value of χ .

$$\sin \chi(\mathbf{u}) = 0.3 \quad (28.43)$$

or

$$\Delta f_{MC} = -0.44(C_s \lambda)^{1/2} \quad (28.44)$$

The importance of this focus setting is that when you are actually working on the TEM, you can recognize this focus setting visually on the TEM screen, since it occurs when you can’t see anything! If you adjust the focus to this condition visually, you then have a reference point from which you can change to the Scherzer defocus. The procedure is actually quite simple (after lots of practice), since you can minimize the contrast easily, providing you have correctly aligned the microscope and corrected the astigmatism.

Some other special settings of the CTF may also be useful. The idea is to make use of *passbands* or large ‘windows’ in the CTF to allow higher spatial frequencies to contribute to the image. As you see in Figure 28.8, what this requires is that χ is constant, or $d\chi/du$ small,

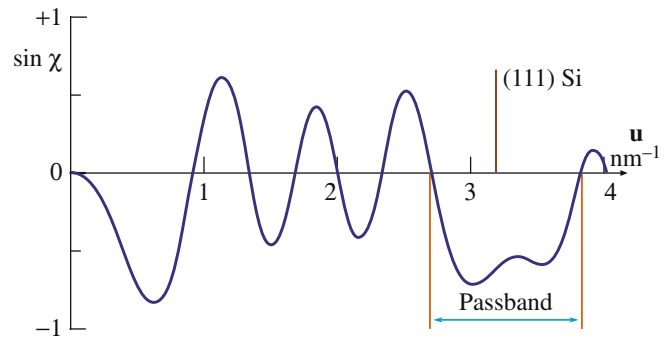


FIGURE 28.8. Special settings of the CTF to make use of passbands or ‘windows’ in the transfer function, here optimized to image Si(111).

over a range of \mathbf{u} which includes the reflection of interest. These passbands occur periodically with underfocus at values set by

$$\Delta f_p^n = -\{([8n + 3]/2)(C_s \lambda)\}^{1/2} \quad (28.45)$$

This formula is not an exact relationship but it gives us a good guide; its derivation is given by Spence. The $n = 0$ passband is, in fact, equivalent to the Scherzer defocus setting. This technique gives us access to higher spatial frequencies and thus finer detail in real space. The price we pay is that there are now zeros in the transfer function at lower spatial frequencies. For some applications, the presence of these zeros may be a problem, but for others, useful information can be obtained in these higher passband settings. For a microscope like the JEM 200 CX, these passband settings are -66 nm (Scherzer, or $n = 1$), -129 nm ($n = 2$), -169 nm ($n = 3$), -202 nm ($n = 4$), etc. Note that all are negative values of focus.

Hashimoto and Endoh defined an ‘aberration-free focus’ (AFF) condition for any specific crystal. The idea is to set the transfer function so that the gaps will only occur between Bragg reflections. All Bragg reflections would then see a window in the CTF out to very high order. This aberration-free focus setting is defined by

$$\Delta f_{AFF} \cong \{2(4m \pm 0.23)C_s \lambda^3 / d^4\} (d^2 / 2\lambda) \quad (28.46)$$

where $m = 0, 1, 2, 3$, etc., and d is the fundamental lattice spacing of the first-order Bragg beam to be resolved. Applying this equation to a Au crystal in the [001] orientation where $d(020) = 0.2035$ nm, using a 100-kV microscope with $C_s = 0.75$ mm, gives $\Delta f_{AFF} = -53.3$ nm. At this setting of focus, the transfer function peaked at -2 for beams 020, 220, 040, 420, 440, and 060.

There is, of course, a catch. We can only use this technique when we know which spatial frequencies we are interested in. In other words, it is great for perfect crystals since we are only concerned with Bragg peaks. If defects are present we will then lose much of the information about the defect, since defects scatter between

the Bragg peaks. Any information falling in the gap of the CTF is lost to the image: in effect, the defect will be invisible!

You should therefore be very cautious in using higher passband settings. You may obtain a pretty picture which does not give a true image of your specimen. If you do use higher-order passbands, you must realize that you are imaging the specimen beyond the instrumental resolution limit so you can't use the intuitive approach for image interpretation. You must know exactly where the zeros are in the CTF. You can only know that by very careful evaluation of your images using diffractograms, computer simulation, and image processing.

28.10 EXPERIMENTAL CONSIDERATIONS

Whenever you are using HRTEM imaging, you must first ask what information you are hoping to obtain. Lattice-fringe images which show lots of straight lines but tell you nothing of where the atoms are located may be all that you need. These fringes are giving you information about the crystal orientation on a very fine scale. Another situation is illustrated by early studies of spinel. You would like to obtain information at, say, 0.23 nm (the spacing of the oxygen 111 planes), but your point-to-point resolution is 0.27 nm. You could still learn a lot about the spinel from the 46 nm spinel (111) planes, so you might even use an aperture to remove information which only adds uninterpretable detail below 0.46 nm. The difficulty comes when you want to relate your HRTEM image to the atomic structure of your specimen. Then you must remember that all of the above treatment is based upon the TEM specimen behaving as a WPO. Most specimens of interest do *not* satisfy this criterion.

MULTIPLE AND PLURAL

Note that the HRTEM community uses the term 'multiple scattering' to denote >1 scattering event. This terminology differs from that used by analytical microscopists, who define 'multiple' as >20 scattering events and reserve 'plural' for 2–20 events. In HRTEM you never hear of 'plural scattering.'

If you look at a typical HRTEM specimen, there will be a wedge-shaped region near the thinnest edge, and thickness extinction contours will be visible. As soon as the first contour is visible, the specimen is already much too thick to behave as a WPO! Multiple scattering limits most phase-contrast imaging conditions for crystalline materials.

Thicker specimens also are susceptible to Fresnel effects associated with spreading of the wave front as it is transmitted through more of the specimen along the beam direction. Inelastic scattering effects, etc., will also

become important as the thickness increases. These effects are not easy to simulate in the computer, although the techniques we will discuss in Chapter 30 are very helpful.

To be really sure that you have correctly interpreted the image, the match between experimental and simulated images should be good over a range of thicknesses and defocus values, as we'll see more clearly in Chapter 30.

We can now summarize the ten steps you need to take to obtain a phase-contrast image with atomic resolution

- Choose an instrument of low C_s and small λ .
- Align it well; it will take time for the electronics and moving parts to become stable.
- Work with an undersaturated LaB₆ filament and a small condenser aperture (unless you have an FEG: see later).
- Perform current and voltage centering of the objective lens routinely and frequently at high magnification.
- Work in thin, flat, and clean regions of the specimen.
- Orient the specimen using small SAD apertures or bend contours in the image, so the beam is aligned along a zone axis.
- Correct the astigmatism, using optical diffractograms if necessary, but ideally on-line (Chapter 31).
- Find the minimum-contrast focus setting and record a through-focus series.
- Record the DP at the same setting of the condenser; calculate α , the convergence angle, and remember that angle means semiangle!
- Simulate and/or process the images using available computer codes (Chapter 30).

A comment on alignment: You'll find that it's relatively straightforward to align the electron beam with the current center or voltage center. The result is an image which does not shift as the current changes in the objective lens or the accelerating voltage fluctuates. As you'll appreciate more from Chapter 31, for the highest resolution, it is also critical that the incident beam is precisely parallel to the optic axis of the microscope. If the incident beam is not exactly aligned with the optic axis we can see the coma aberration, which is only important at the highest resolution. (A discussion of other aberrations is given in the companion text.) We refer to the process of aligning the beam with the optic axis as 'coma-free alignment.' The process involves alternately applying equal and opposite beam tilts to the incident beam; you choose the magnitude of the tilt to match the periodicity in the image. If there is a residual beam tilt of the incident beam away from the optic axis, then one image will look more distorted than the other. Adjust the beam tilt controls until both tilted images look equally distorted. Repeat this procedure for the orthogonal direction. You will need a lot of practice to do this successfully (see Section 30.5).

COMA

If a point object on the optic axis is imaged as a point, a similar point object off the axis may appear distorted. This distortion is known as coma or comatic aberration. In a telescope, the point (a star) would look like a comet—a cometary coma. (Coma: Latin for hair.)

Some final remarks on experimental techniques: Always remember that specimen orientation is very critical for HRTEM. Always be aware of contamination and damage caused by the electron beam; the specimen will have changed long before you can see the change by eye. Although HRTEM is now so much easier because high-quality CCD video cameras are available to give you an image at TV rates, don't spend any longer than you have to with the beam on the area of interest in your specimen. Get used to using the CCD and the computer. If you are going to do quantitative HRTEM, you'll have to be comfortable with both. Of course, if you use a remote microscope you aren't even sitting near the screen.

28.11 THE FUTURE FOR HRTEM

The historical approach to HRTEM was: be pleased if you recorded what you saw. Now machines are sufficiently stable that we can reliably record images at different values of Δf . Certainly as important is the availability of computers, as we will discuss in Chapters 30 and 31, since we can 'predict' the image for model structures and quantify the contrast of the image. We are thus able to do quantitative HRTEM (QHRTEM or HRQTEM!).

Another approach to improve resolution is provided by the FEG TEM. The beam in an FEG TEM is now more highly coherent, so the envelope function shown in Figure 28.7 extends to greater values of \mathbf{u} . The computer becomes indispensable because we have to interpret images which have contrast reversals beyond Scherzer defocus. If a set of carefully designed multipole lenses is inserted into an HRTEM (by the manufacturer), it is possible to correct C_s or even make it variable like Δf .

If you use such a C_s -corrected TEM, you have to rethink your approach to HRTEM. Think what will happen to the scherzer and the glaser. What will the Gl/Sch be? The corrector proposed by Rose was shown schematically in Figure 6.12B. The actual corrector, incorporated into a 200-kV JEOL TEM, is shown in Figure 28.9 (with the covers removed); it significantly increased the height of the column. It is a combination of round lenses and hexapoles, all of which are magnetic elements. The hexapoles don't affect the paraxial path of the rays and only need to be stabilized to an accuracy

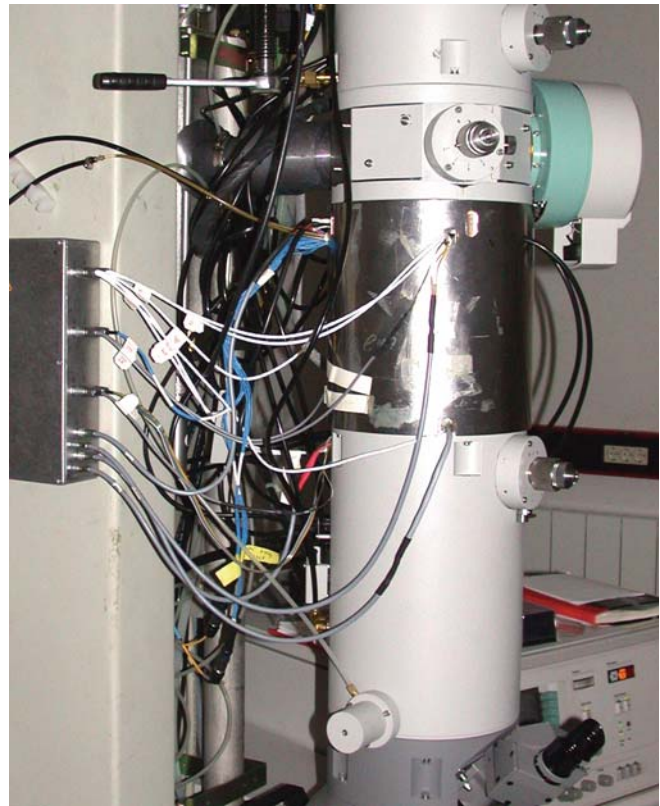


FIGURE 28.9. The post-objective lens corrector system as incorporated into a 200-kV JEOL TEM; it looks sleeker in the finished product. (See also Figure 6.12B.)

of 1 in 10^4 to give atomic resolution. When C_s is zero, the specimen resolution limit will be determined by C_c

$$d_{C_s=0} \approx \left[\left(\frac{\Delta E}{E} \right) \lambda C_c \right]^{\frac{1}{2}} \quad (28.47)$$

If $C_c = 2$ mm and $\Delta E \sim 0.3$ eV, a 200-kV FEG TEM could achieve a resolution of 0.8 \AA . If C_c is also corrected, which we'll see in Chapter 40, it is possible that the resolution will become limited by the fifth-order spherical aberration constant. In practice, it will be important to correct C_c in the lens design first. In Rose's original C_s -correcting lens, $C_s = 3$ mm and the resolution limit is 0.28 \AA for this 200-kV FEG TEM. Other lens defects will limit this to $\sim 0.5 \text{ \AA}$ but with a price of \$12 M for a 1.25-MeV machine which changes your specimen in seconds, the Rose corrector will be quite important!

28.12 THE TEM AS A LINEAR SYSTEM

The discussion we went through above is an example of a much larger topic known as information theory. The concept of a 'phase-contrast transfer function' is central to this field. So you can understand the practice of

phase-contrast imaging at high resolution, we will briefly discuss the way an information specialist might view this process. We will define the transfer function in elementary terms, and make detailed reference to phase-contrast imaging in the TEM.

Remember, the purpose of the TEM is to transmit information about the specimen to the image. We can thus consider the microscope to be an 'information channel' and use the concepts of information theory.

- The input signal comes from the specimen.
- The output signal is the image.

If we neglect the effects of noise, there is a unique relation between the input signal and the output signal, determined by the optical system of the microscope.

Most information theory treats linear systems. A linear system is one which is characterized by the property that if

$$S_0(r_0) \rightarrow \text{Transmission System} \rightarrow S_1(r_1),$$

and if

$$S'_0(r_0) \rightarrow \text{Transmission System} \rightarrow S'_1(r_1),$$

(the prime here denotes the derivative) then the system is linear if

$$a(S_0) + b(S'_0) \rightarrow \text{Transmission System} \rightarrow a(S_1) + b(S'_1)$$

for any values of a and b . (It's like linear elasticity where we can simply add stresses and strains.)

The linear relation between input and output signals can be described by the concept of the CTF. Overall, the transfer function relates an input spectrum to an output spectrum, and it operates only in the frequency domain.

LINEAR SYSTEM

Schrödinger's wave equation is linear. Therefore, the amplitudes of an electron wave in the specimen are linearly related to the amplitudes of an electron wave in the image.

In general, for a linear system, if we know the CTF, then the relation between S_0 and S_1 is uniquely defined. On the other hand, if the relation between S_0 and S_1 could be empirically determined, then we can deduce the CTF.

One of the best examples of a linear system is an electrical transmission cable. The transfer of electrical signals through transmission lines can be made linear enough for the above theory to apply. Conversely, the transfer of mass-thickness information from a specimen

to the optical density of a developed photographic negative is far from linear, and the above theory does not apply. Then why bother to discuss this in HRTEM? The answer lies in finding an appropriate linear relation between the object and the image.

28.13 FEG TEMs AND THE INFORMATION LIMIT

We've mentioned that an FEG reduces the instrumental contribution to chromatic aberrations and extends the envelope function to larger values of \mathbf{u} . This means that information with higher spatial frequencies is transferred to the image. We've just analyzed the Scherzer defocus problem, so now we'll consider the information limit. The reason for emphasizing the FEG here is that it really makes a difference and we're just beginning to learn how to use this information: the contrast reversals mean that any image interpretation is not intuitive. The topic has been laid out in two papers by Van Dyck and de Jong but be warned—this topic is exceedingly tricky!

Since the information limit is determined by the envelope function, this is split into its separate terms. The total envelope function, $E_T(\mathbf{u})$, is the product of all of these (like equation 28.5).

$$E_T(\mathbf{u}) = E_c(\mathbf{u})E_s(\mathbf{u})E_d(\mathbf{u})E_v(\mathbf{u})E_D(\mathbf{u}) \quad (28.48)$$

The individual envelope functions in equation 28.48 are

- $E_c(\mathbf{u})$: for chromatic aberration.
- $E_s(\mathbf{u})$: for the source dependence due to the small spread of angles from the probe.
- $E_d(\mathbf{u})$: for specimen drift.
- $E_v(\mathbf{u})$: for specimen vibration.
- $E_D(\mathbf{u})$: for the detector.

As you can see, some of these envelope functions are new, some are old. We won't discuss all these functions; we'll only mention a couple of the key points.

The chromatic aberration is well known, and its envelope function $E_c(\mathbf{u})$ can be expressed by the equation

$$E_c(\mathbf{u}) = \exp\left[-\frac{1}{2}(\pi\lambda\delta)^2 u^4\right] \quad (28.49)$$

where c reminds us that this is a chromatic aberration and δ is the defocus spread due to this aberration

$$\delta = C_c \left[4 \left(\frac{\Delta I_{\text{obj}}}{I_{\text{obj}}} \right)^2 + \left(\frac{\Delta E}{V_{\text{acc}}} \right)^2 + \left(\frac{\Delta V_{\text{acc}}}{V_{\text{acc}}} \right)^2 \right]^{\frac{1}{2}} \quad (28.50)$$

The terms $\Delta V_{\text{acc}}/V_{\text{acc}}$ and $\Delta I_{\text{obj}}/I_{\text{obj}}$ represent the instabilities in the high-voltage (accelerator) supply

and the objective lens current. $\Delta E/V_{\text{acc}}$ is the intrinsic energy spread in the electron gun. Notice that ΔE and ΔV are different: ΔV depends on how well we can control the voltage supply whereas ΔE depends on our choice of electron source (see Chapter 5). If we neglect any other contributions to the envelope function, then we can define an information limit due to instrument chromatic aberrations by ρ_c

$$\rho_c = \left(\frac{\pi\lambda\delta}{\sqrt{2\ln(s)}} \right)^{\frac{1}{2}} \quad (28.51)$$

where e^{-s} is the cut-off value for the envelope. If we take $\ln_e s$ to be 2, (s is not the deviation parameter) then

$$\rho_c = \left(\frac{\pi\lambda\delta}{2} \right)^{\frac{1}{2}} \quad (28.52)$$

The source-dependent envelope function is new because, until we have a, FEG, we don't usually consider the 'source of a probe.' If we imagine that the source has a Gaussian distribution, we have an envelope function $E_s(\mathbf{u})$ given by

$$\begin{aligned} E_s(\mathbf{u}) &= \exp \left[\left(\frac{\alpha}{2\lambda} \right)^2 \left(\frac{\partial\chi(\mathbf{u})}{\partial u} \right)^2 \right] \\ &= \exp \left[- \left(\frac{\pi\alpha}{\lambda} \right)^2 (C_s\lambda^3 u^3 + \lambda u)^2 \right] \end{aligned} \quad (28.53)$$

Here α is the semi-angle characterizing the Gaussian distribution. What this equation tells us is that if α is too large (≤ 1 mrad) it can limit the information limit. If we say that u must lie between u and some maximum value u_{max} , we can maximize the argument of the exponential in equation 28.53 to give an optimum focus

$$\Delta f_{\text{opt}} = -\frac{3}{4} C_s \lambda^2 u_{\text{max}}^2 = -\frac{3}{4} \frac{C_s \lambda^2}{\rho_i^2} \quad (28.54)$$

In this equation ρ_i is the information limit of the microscope (because of how we chose u_{max}). This defocus value will be important later when we discuss holography in a FEG TEM. The two curves shown in Figure 28.10 illustrate graphically how this envelope function varies within Δf . It can also be optimized by decreasing the angle α . With a little more manipulation, de Jong and Van Dyck show that the information limit due to the limited coherence of the source is given by

$$\rho_\alpha = \left(\frac{6\pi\alpha a}{\lambda\sqrt{\ln(s)}} \rho_s^4 \right)^{\frac{1}{3}} \quad (28.55)$$

The envelope functions for the drift/translation and vibration of the specimen represent a new method for

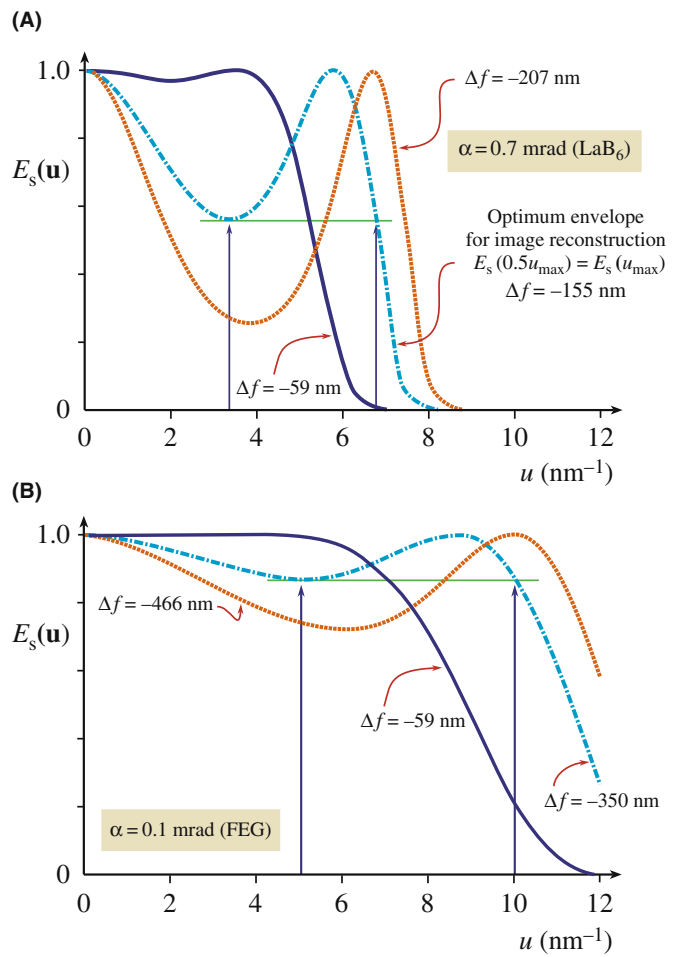


FIGURE 28.10. Variations in the envelope function, $E_s(\mathbf{u})$, for different objective lens defocus: (A) LaB₆ source, (B) FEG.

taking account of these two unavoidable quantities. We'll just quote the results for the two 'information limits' which are the crossover values for the two envelope functions $E_d(\mathbf{u})$ and $E_v(\mathbf{u})$

$$\rho_d = \frac{\pi d}{\sqrt{6\ln(s)}} \quad (28.56)$$

and

$$\rho_v = \frac{\pi v}{\sqrt{\ln(s)}} \quad (28.57)$$

In these equations, d is the total specimen drift during the exposure, t_{exp} , so $d = v_d t_{\text{exp}}$ for a drift velocity v_d ; v is the amplitude of the vibration.

The detector envelope function, $E_D(\mathbf{u})$, is something we never worried about with film, but CCD cameras have a limited number of pixels, i.e., we only have a limited number of resolved image points. This envelope function results from two effects

- Delocalization of the information in the image.
- The finite pixel size.

DELOCALIZATION

Image delocalization depends on \mathbf{u} . It is large when $\partial\chi(\mathbf{u})/\partial u$ oscillates rapidly as it does for large \mathbf{u} , i.e., where we are placing the information limit.

The idea is simple but the math is more difficult. Let's assume that the image which is actually captured by the camera is circular. We can then say that if R is less than R_w , the radius of the window, then we capture the information; if it's greater than R_w , we don't. So the CCD detector is acting like an aperture! Now de Jong and Van Dyck show how u_{\max} is related to R_w

$$\alpha C_s \lambda^3 u_{\max}^3 = R_w \quad (28.58)$$

The important result is that the delocalization of the information in the image must be less than the half-width of the CCD detector array.

The value of R_w is related to the number of pixels, N , and their size, D

$$R_w = \frac{1}{2} ND \quad (28.59)$$

The information limit due to the detector (i.e., the cross-over values of the detector envelope functions $E_D(\mathbf{u})$ is

$$\rho_D = \left(\frac{12\sqrt{2}\pi a}{N\sqrt{\ln(s)}} \right)^{\frac{1}{4}} \rho_s \quad (28.60)$$

Clearly we can decrease ρ_D by increasing N , but not quickly. With this analysis in mind, we can summarize the conditions necessary for ρ_i to be limited by chromatic aberration.

$$\alpha \leq \frac{\lambda}{6\pi a \rho_s} \left(\frac{\rho_c}{\rho_s} \right)^3 \quad (28.61)$$

$$N \geq 12\sqrt{2}\pi a \left(\frac{\rho_s}{\rho_c} \right)^4 \quad (28.62)$$

$$d \leq \frac{\sqrt{6}}{\pi} \rho_c; \quad 0.8\rho_c \quad (28.63)$$

$$u \leq \frac{1}{\pi} \rho_c; \quad 0.3\rho_c \quad (28.64)$$

Table 28.1 gives some numerical examples of what these equations mean.

To see whether we will ever reach the information limit, we have to consider the effect of the noise. We know that the signal-to-noise ratio is proportional

TABLE 28.1. Maximum Convergence Angle α and Minimum Number of Unusable Image Points N for Different Values of the Point Resolution to (Chromatic) Aberration Limit Ratio (ρ_s/ρ_c)

ρ_s/ρ_c	α (mrad)		N (pix)	
	ϵ_0	ϵ_{opt}	ϵ_0	ϵ_{opt}
1	0.58	2.3	53	13
1.5	0.17	0.69	270	67
2	0.07	0.30	853	213
2.5	0.04	0.15	2082	521
3	0.02	0.09	4320	1080

ϵ_0 : Gaussian focus, ϵ_{opt} : optimum focus; $\lambda = 0.011 \rho_s$ (de Jong and Van Dyck 1993).

to $\beta^{1/2}$, where β is the brightness of the electron gun. The smallest image element we need to examine has an area ρ_i^2 . Then the background signal I_0 is given by

$$I_0 = D\rho_i^2 = \beta\pi\alpha^2 t\rho_i^2 \quad (28.65)$$

where D is the electron dose, α is the angle of convergence, and t is the time.

For white noise, the noise in an element will be related to $I_0^{1/2}$. The total contrast in our small pixel can be written as $DEF\rho_i^2$ where D is the dose, E is the envelope function, F is the structure factor, and ρ_i^2 is the area of the pixel. Now we can say that the minimum detectable signal-to-noise ratio is k , which gives

$$DEF\rho_i^2 = kI_0^{1/2} = k\rho_i D^{1/2} \quad (28.66)$$

Therefore for $\alpha = 1$ mrad and $t = 1$ second, we can express the signal-to-noise ratio as

$$s_0 = 443\rho_i \frac{F}{k} \beta^{1/2} \quad (28.67)$$

Now you can use some real numbers: take $k = 2$ (think what this means for the minimum contrast), and assume that β is 10^{10} A/m² sr for a LaB₆ gun and 10^{13} A/m² sr for a Schottky FEG. You can show that for $\rho_i = 0.15$ nm (a LaB₆ gun) $\ln_e s_0$ is 1.2–2.2 whereas for a FEG, $\rho_i = 0.1$ nm and $\ln_e s_0$ is 4.5–5.2. You can also appreciate why s_0 depends on your material: low atomic numbers mean weak scattering. For our last two equations we'll again quote de Jong and Van Dyck. We can deduce optimum values for both the angle of convergence α and the exposure t by differentiating the envelope equations.

$$\alpha_{\text{opt}} = \frac{1}{k_s \sqrt{2}} \left(\frac{\rho_i}{\rho_c} \right)^3 = \frac{1}{6\pi a \sqrt{2}} \frac{\lambda}{\rho_s} \left(\frac{\rho_i}{\rho_s} \right)^3 \quad (28.68)$$

$$t_{\text{opt}} = \frac{1}{2k_d} \left(\frac{\rho_i}{\rho_c} \right) \simeq 0.39 \left(\frac{\rho_i}{v_d} \right) \quad (28.69)$$

Notice that α_{opt} depends not only on ρ_s and ρ_i , but also on λ (of course ρ_s and ρ_i also depend on λ) and that t_{opt} only depends on the drift rate; fortunately v_d will never be zero!

We can now summarize these new concepts

- Microscopy is much more complex when you try to use the information limit rather than the Scherzer limit!
- If you want to use a computer, the size of the CCD camera will also affect the actual information limit; this is the effect of $E_D(\mathbf{u})$.
- Drift and vibrations must be minimized or they will determine your resolution; these contributions were described by $E_d(\mathbf{u})$ and $E_v(\mathbf{u})$.
- When everything else is perfect, your resolution will be controlled by the signal-to-noise ratio of the detector and the coherence functions, $E_D(\mathbf{u})$ and $E_c(\mathbf{u})$.
- An FEG improves the information limit because of the large increase in the brightness, β . This increase allows us to decrease α , increase the dose, and increase the signal-to-noise ratio.

This section has been quite long and not what you traditionally find in a chapter on HRTEM. This is quite deliberate. The TEM produces a great deal of information but many researchers want numbers, quantities, not data encoded in images. So we need to convert our data into numbers; we have to be quantitative.

MORE DELOCALIZATION

To fully interpret an HRTEM image, you must understand the implications of this concept. See the papers by (i) Coene and Janssen and (ii) Lichte.

28.14 SOME DIFFICULTIES IN USING AN FEG

We've discussed the advantages of using an FEG for HRTEM, but there are some practical difficulties which have been analyzed by Otten and Coene at FEI. A cold FEG (CFEG) allows us to extract a very high current per unit area, but the total area of the emitting region is very small so that the extraction current is <5 nA. This current can be increased if we thermally assist the field emission by heating the Schottky emitter to $\sim 1500^\circ\text{C}$. It gives the same high brightness, but a larger maximum current because of the larger emitting area. So what are the difficulties?

- The emitter area may be so small that we have to 'fan' the beam in order to illuminate the area used in TEM.

This fanning may actually increase the effect of coma aberration (a radial aberration as noted in Section 28.10). If a CFEG has a source size of ~ 3 nm, we can study ~ 15 nm with a $5\times$ magnification. The Schottky source has a source diameter $\sim 10\times$ greater, and the price you pay for this is a decrease in spatial coherence, and a larger energy spread.

- Correcting astigmatism is very tricky with an FEG. As shown in Figure 28.11, if the image is astigmatic you'll see that at all defocus settings with an LaB_6 source. In an FEG, when astigmatism is present, all the images look similar and you can't use the technique of finding the minimum-contrast defocus (at ~ 0.4 sch) to determine Δf . If you try to use the wobbler to do coma-free alignment, that fails too, because you can't interpret the focus difference between two FEG images for the two wobbler directions. There is a solution to finding Δf , fortunately; either use on-line processing (Chapter 31) or converge the beam! The latter way deteriorates the spatial coherence and you've made your \$2 M FEG behave like an old \$200 K LaB_6 machine.
- Focal series of images are a challenge, because you can now use a very large range of Δf values, and it becomes a major task just to determine your value of Δf .
- Image delocalization occurs when detail in the image is displaced relative to its 'true' location in the specimen. The effect is emphasized by the graph shown in Figure 28.12 and becomes worse as you go away from Scherzer defocus. The effect is illustrated in Figure 28.13, where fringes from the gold particles can appear outside the particle. If we rewrite equations 28.36, we can express the delocalization as

$$\Delta R = \lambda u(\Delta f + C_s \lambda^2 u^2) \quad (28.70)$$

You may notice a similarity between this equation and that for the SAD error (Chapter 11): there is a good reason for this similarity. Two values have been proposed for Δf_{opt} , the optimum defocus setting to minimize delocalization. They give an optimum value for the defocus of

$$\Delta f_{\text{opt}} = -MC_s \lambda^2 u_{\text{max}}^2 \quad (28.71)$$

where M is a factor between 0.75 and 1. The value for ΔR_{min} is then close to

$$\Delta R_{\text{min}} = \frac{1}{4} C_s \lambda^3 u_{\text{max}}^3 \quad (28.72)$$

The actual value of M is determined by where you define the cut-off value for \mathbf{u} . There are three important conclusions on delocalization

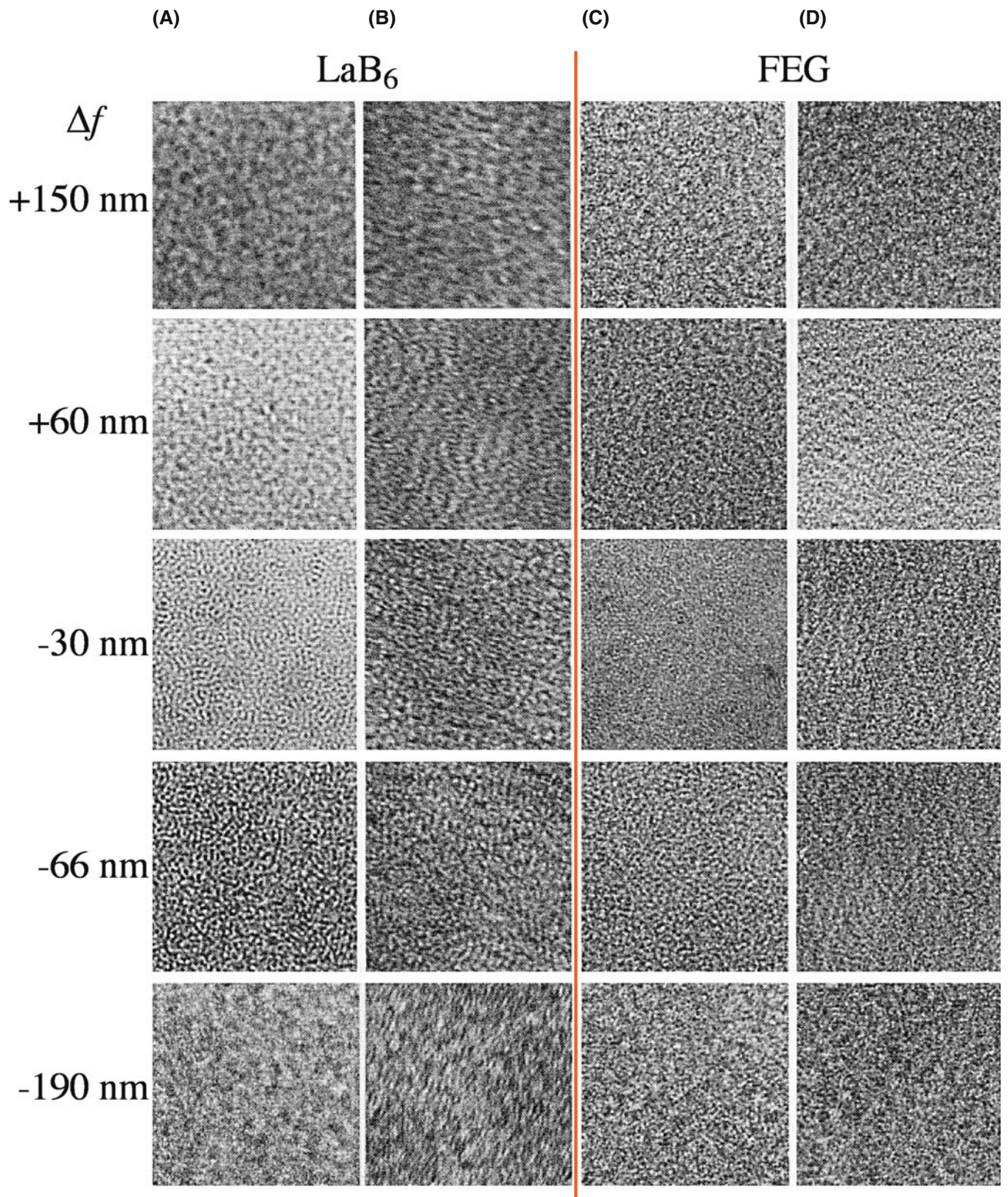


FIGURE 28.11. A tableau of images from an amorphous film at varying defocus. (columns A and B) LaB₆; (columns C and D) FEG. (columns A and C) Without astigmatism; (columns B and D) with astigmatism. With LaB₆ you can easily see the astigmatism, while with an FEG, you can't.

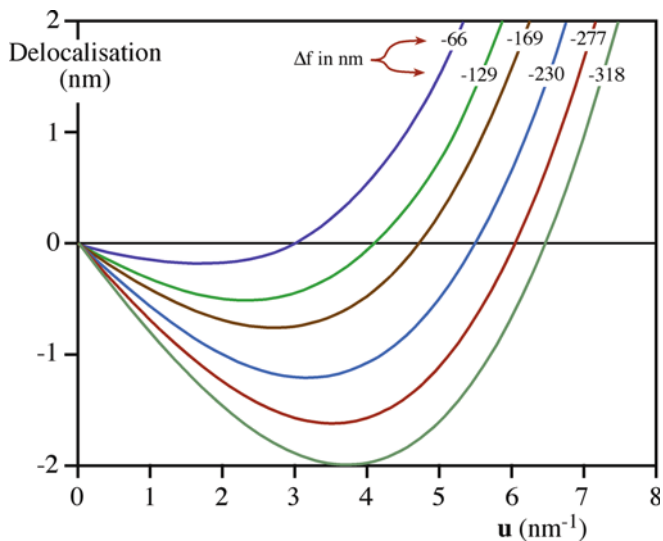


FIGURE 28.12. Image delocalization plotted against u as Δf is changed for a Philips CM20 FEG with $C_s = 1.2$ nm.

- As C_s decreases, delocalization decreases.
- As λ decreases (accelerating voltage increases), delocalization decreases.
- Delocalization cannot be avoided in an FEG, except by greatly reducing C_s !

28.15 SELECTIVELY IMAGING SUBLATTICES

In Chapter 16, where we discussed ordered intermetallic alloys, we saw that many materials with a large-unit cell are closely related to a material with a smaller unit cell. If the two structures don't have the same symmetry, then the two unit cells can show several different orientation relationships, as was the case for vanadium carbide.

We can use this information to form different high-resolution images instead of different DF images. Two [001] DPs from an ordered alloy of Au_4Mn are shown in Figure 28.14, together with a schematic of one pattern. Two domains are present in the combined pattern. Both patterns have fourfold symmetry, but they are rotated relative to one another. If we use the DF lattice-imaging mode and exclude all the fcc reflections using the objective aperture, we form an image like that shown in Figure 28.15. The two variants are not only easily recognized, but we know where they are with an accuracy of atomic dimensions. If you're used to grain-boundary theory, the original cell has become the coincident-site lattice (CSL) in reciprocal space and the two sublattices are like grains related by a small Σ . This approach has been used to estimate the size of very small particles of NiFe_2O_4 spinel which are completely contained in a matrix of NiO , as illustrated in Figure 28.16. The lattice

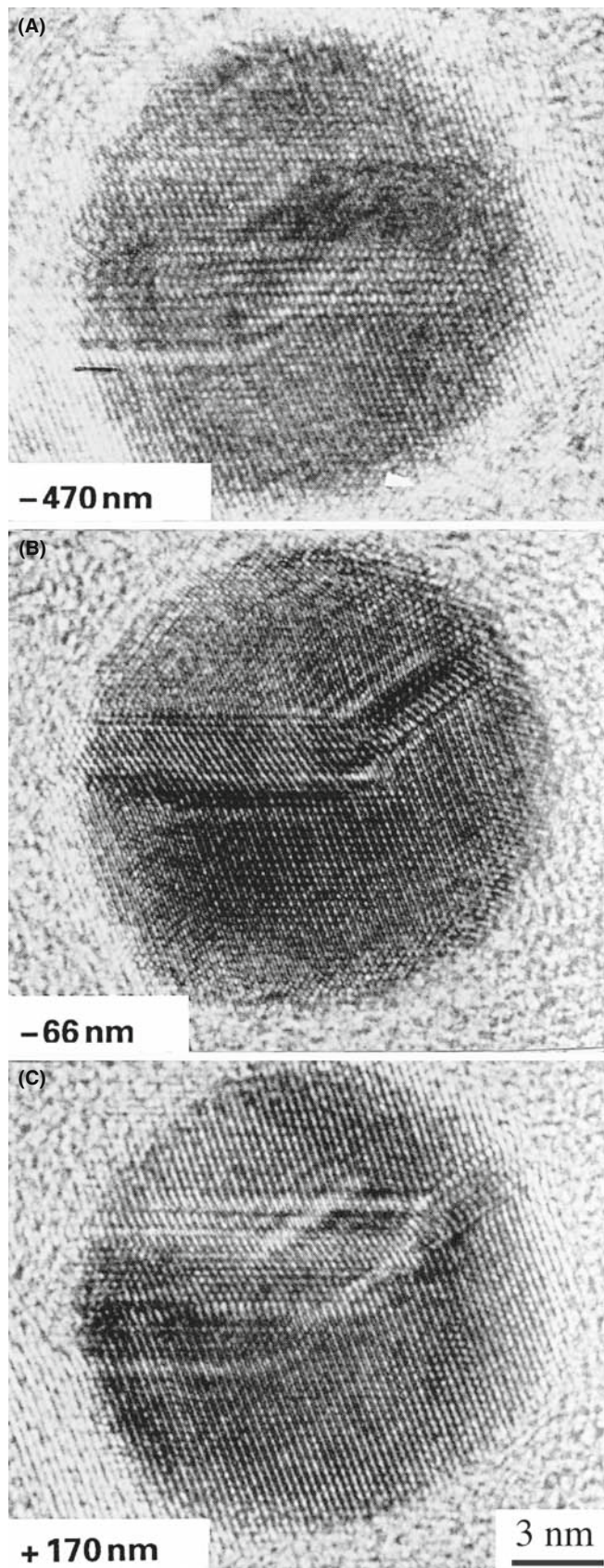


FIGURE 28.13. Experimental images showing delocalization in HRTEM images of an Au particle: (A) underfocus, (B) Scherzer focus, (C) overfocus.

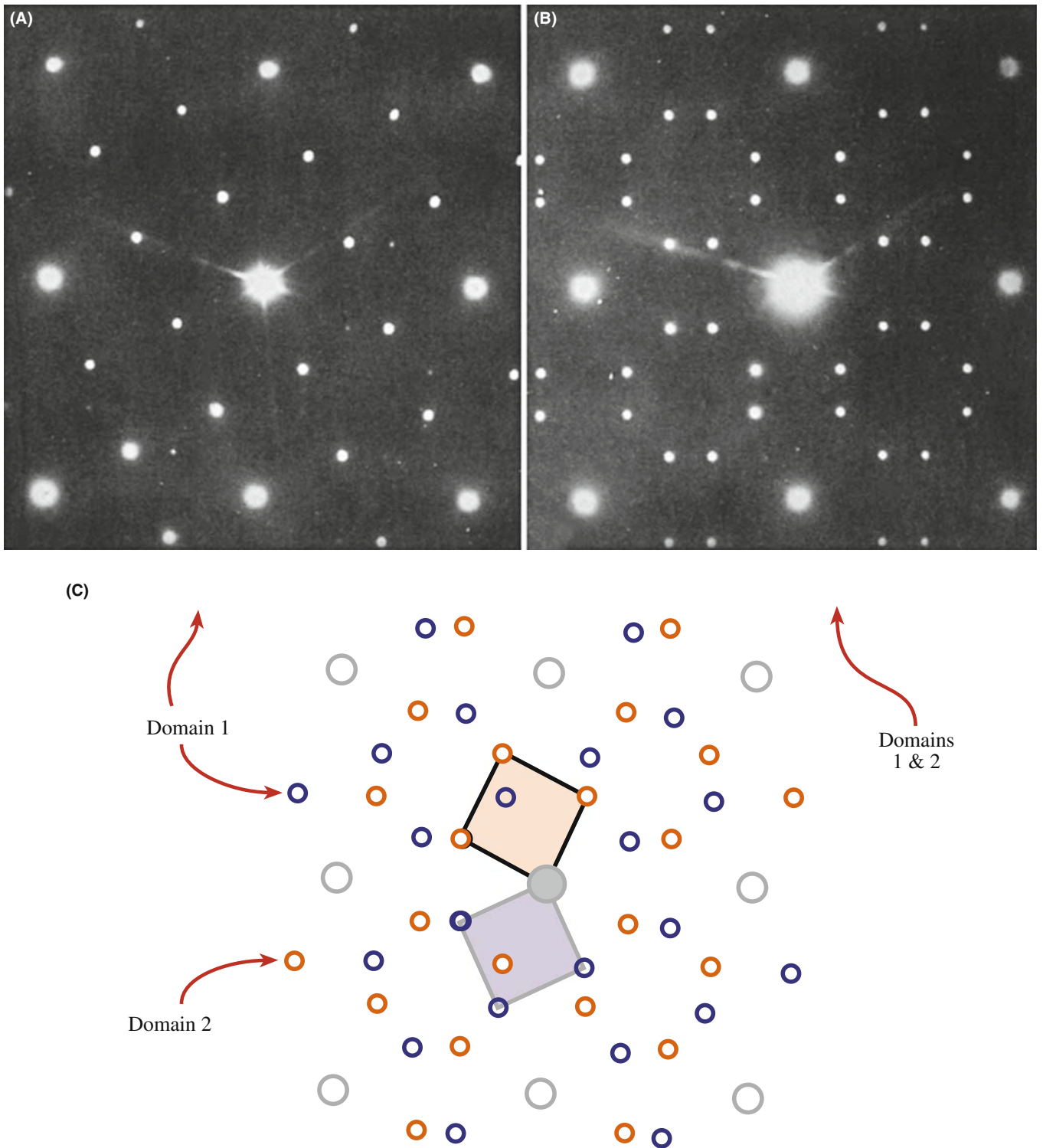


FIGURE 28.14. Two [001] DPs from an ordered alloy of Au_4Mn with schematics. (A) One domain. (B) Two symmetry-related domains. (C) Schematic diagram showing how (B) arises from the relative rotation of the two domains.

parameter of the spinel is twice that of the NiO but the latter is generally above and below the particle. This approach is therefore quite difficult, especially if, as in

the analysis of Figure 28.16, the shape of the particle is important. Then you need to resort to simulation and processing, as we'll discuss in Chapter 30.

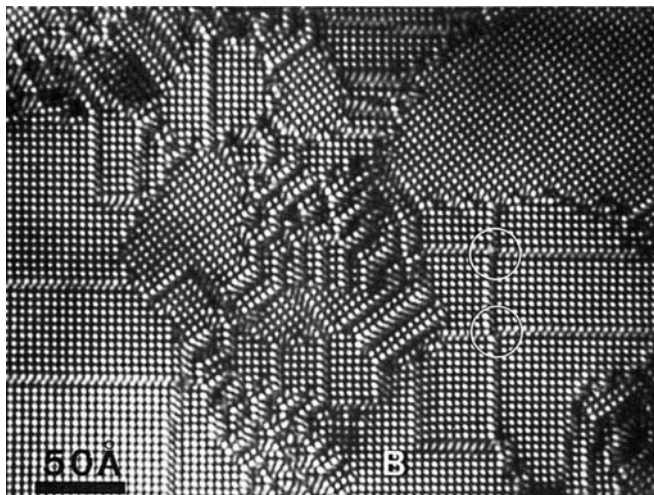


FIGURE 28.15. DF lattice image of Au_4Mn using an objective diaphragm to exclude all the fcc reflections. The two differently oriented domains correspond to the two orientations in Figure 28.14C.

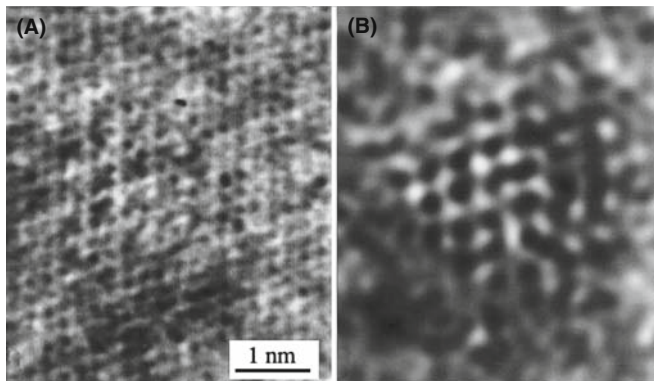


FIGURE 28.16. (A) The experimental image of a small spinel particle in a NiO matrix. The NiO is thicker and dominates the image. (B) After filtering out the NiO contribution (its lattice parameter is twice that of the spinel), we can see the spinel particle and estimate its size. See also Figures 31.2 and 31.3.

28.16 INTERFACES AND SURFACES

Interfaces of all kinds have been extensively studied by HRTEM. Of course, we want the near-atomic resolution. Sometimes it's because they make ideal subjects for study! Point defects require extensive image processing and simulation, dislocations tend to move, but interfaces seem to remain stationary if you're careful. However, we are always limited as to which interfaces we can study.

The fundamental requirement is that the interface plane must be parallel to the electron beam.

If a low-index plane in one grain (but preferably in both grains) is parallel to the interface, you're in business. The problem is that we are rarely sure that this is the case, but because you are looking at a very thin specimen, the projected width of even a slightly tilted boundary is small.

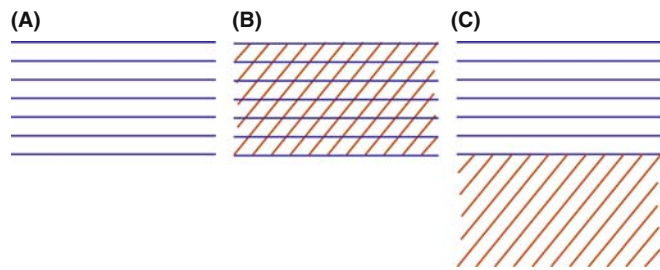


FIGURE 28.17. Schematic HRTEM images of grain boundaries showing (A) one set of fringes in one grain; (B) specimen tilted to give crossing fringes in one grain; (C) one set of fringes in each grain boundary remains parallel to the beam.

So you can tilt the specimen to look down a pole in one grain or to make the beam parallel to a low-index plane in the second grain as shown in Figure 28.17.

If you are lucky (and we often are because we only study tilt boundaries by HRTEM), you can produce crossed fringes in both grains. A selection of images is shown in Figure 28.18. Here you can see structured boundaries, boundaries with an amorphous layer between the grains, interfaces between two different materials, and a surface profile image. We can make some general comments about these images

- Even a 'low'-resolution, lattice-fringe image gives you information on the local topology of your interface.
- If the layer of amorphous material in the boundary is quite thick (>5 nm), you can see it directly.
- You can quite easily see detail like five-membered rings in grain boundaries but you should be wary of interpretation until you've covered Chapter 30.
- You can see abrupt interfaces at near-atomic dimensions.

Now we can also list our concerns

- Has grooving at the interface affected the appearance of the image? The answer is "maybe, but does it affect what you wanted to learn?"
- Is the phase boundary as abrupt chemically as it is structurally? It is very difficult to answer this. The appearance of the image changes at the interface in Figure 28.18C, mainly because the total number and location of the cations (Fe^{3+} and Ni^{2+}) changes, not because there is a 2:1 ratio of Fe:Ni.
- Are all of the black spots in Figure 28.18D complete columns? The next question is "Complete columns of what?"

We will address some of these problems in Chapter 30 but we can make some comments now.

- The quality of your imaging data will be governed by how well you prepare your specimen. Nearly all subsequent analysis will assume that it has a uniform

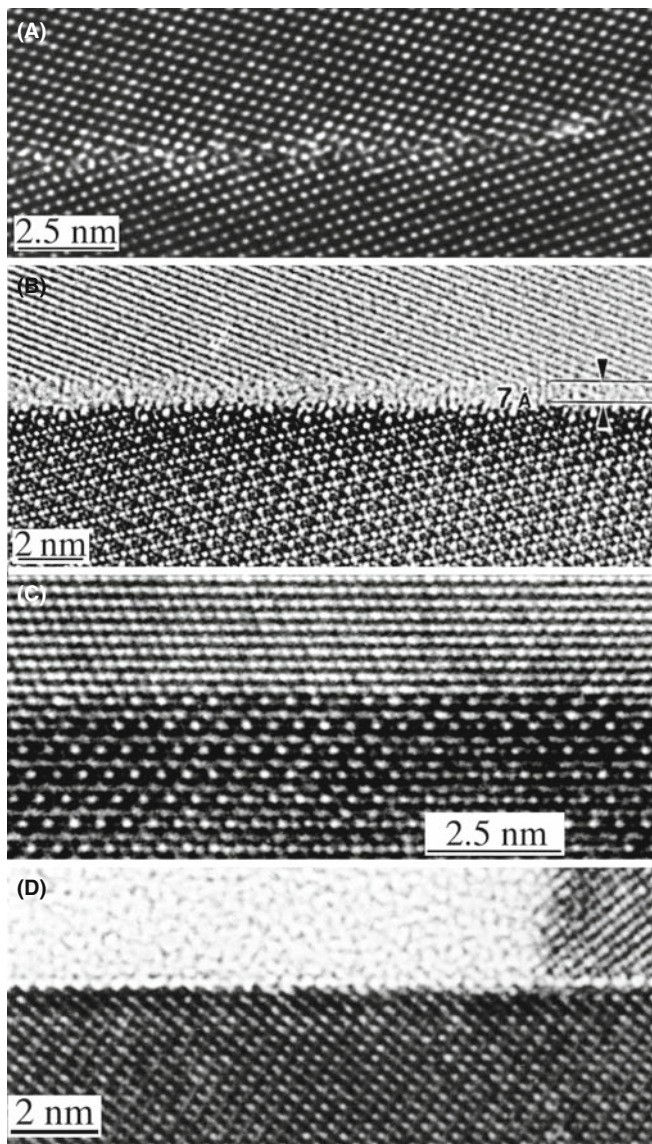


FIGURE 28.18. Examples of HRTEM images of planar interfaces. (A) Grain boundary in Ge; (B) grain boundary in Si_3N_4 with a layer of glass along the interface; (C) phase boundary separating NiO and NiAl_2O_4 ; (D) profile images of the (0001) surface of Fe_2O_3 .

thickness across the interface. If you do not *know* that this is so, then your interpretation may and should be questioned.

- Crystalline grains also thin at different rates if they have different orientations or different structures or different chemistries. The grain-boundary layer, whether crystalline or amorphous, will also thin at a different rate. Why? Because the bonding and density are different. So careful specimen preparation is absolutely critical.
- You can learn a lot about your interface using HRTEM without trying to use atomic-resolution imaging.
- The longer you look at your specimen, the more it will differ from what you started with. Use at least a pseudo-

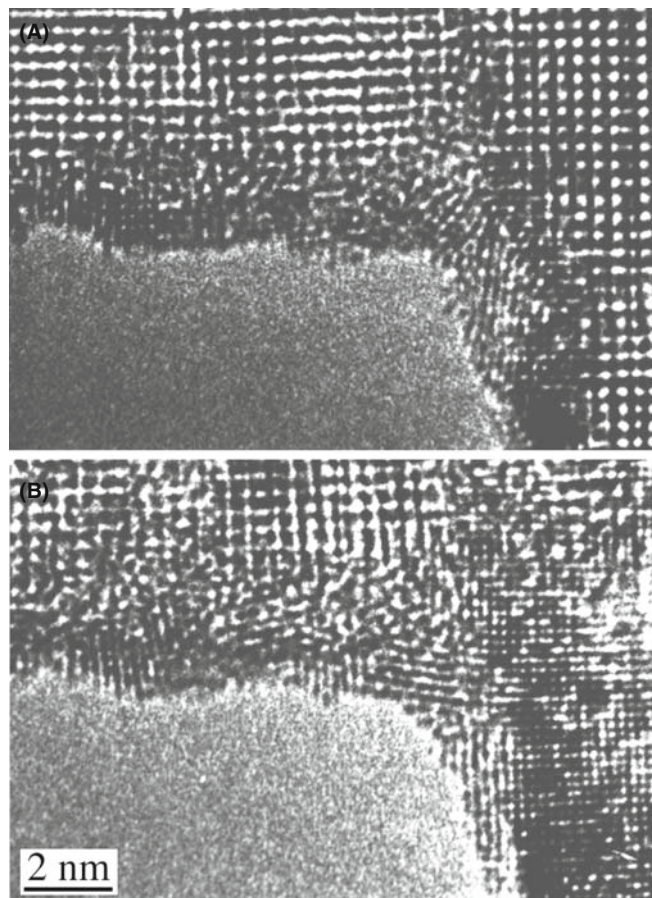


FIGURE 28.19. Reduction of Nb oxide to the metal by beam-induced loss of oxygen during observation of the edge of the foil. The reduction increases with time.

low-dose approach, if possible. Figure 28.19 illustrates an extreme example. Here, the oxide has been completely reduced to the metal at the edge of the foil. Of course, this now provides a method for studying the reduction of oxides under the electron beam in the presence of hydrocarbons and in a good vacuum!

28.17 INCOMMENSURATE STRUCTURES

We'll illustrate this topic by considering several types of incommensurate (modulated) structures. In each case, the structure consists of a 'parent' structure to which we then add a periodic modulation by means of an internal planar defect. Van Landuyt et al. have characterized three different types of incommensurate structures.

- Periodic modules of the parent separated by interfaces. The interface may be a stacking fault (SF), twin boundary (TB), anti-phase boundary (APB), inversion domain boundary (IDB), crystallographic shear (CS) plane, or discommensuration wall.
- A parent structure with a superimposed periodic deformation wave with a larger periodicity.

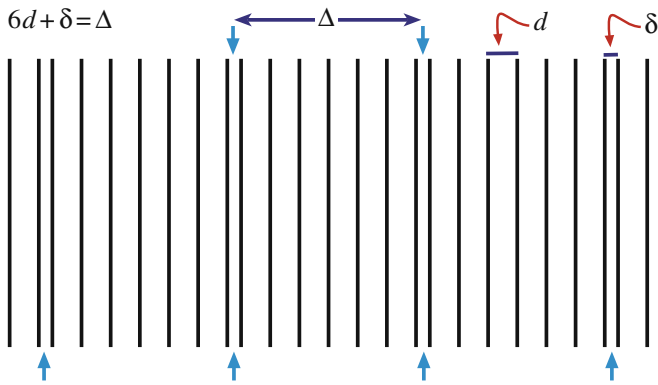


FIGURE 28.20. An incommensurate structure formed by inserting a planar defect after every seventh layer to expand the lattice by δ every seventh plane.

- A parent structure where the composition or site occupancy changes periodically.

The next complication is that we can find commensurate and incommensurate structures, and also structures where the modulation is variable. To understand how a structure can be incommensurate, consider Figure 28.20, where we've placed a planar defect after every seventh layer so that it expands the lattice by δ every seventh plane. The parent lattice will show a spot spacing in the DP proportional to d^{-1} but the 'superlattice' will have a periodicity of Δ^{-1} , which means that we need not have a simple relationship between the two arrays of spots.

These different kinds of modulation can be combined! We'll illustrate this type of specimen with two examples. The Bi-Sr-Ca-Cu-O superconductor provides a good illustration of this type of structure. The parent is a perovskite-like cube; we may have two, three, or four layers of the perovskite with each group separated from the next by a bismuth oxide layer. The formula can be written as $\text{Bi}_2\text{Sr}_2\text{Ca}_n\text{Cu}_{n+1}\text{O}_{2n+\delta}$ so for $n = 1$ we have a sequence of layers (planes) described as BiO-SrO-CuO₂-Ca-CuO₂-SrO-BiO, as shown in Figure 28.21. The DPs depend on the particular value of n in the chemical formula and show rows of satellite reflections due to the modulation of the basic structure. When we form the HRTEM image, the lattice planes appear wavy although we can recognize an orthorhombic pattern. The wavy modulation in the image is probably due to excess oxygen in the BiO layers: the BiO layers don't fit very well to the perovskite block but the misfit stresses can be relaxed by introducing excess oxygen. There are several clear lessons from this example

- HRTEM is essential if we are to understand such structures.
- You need supplementary information, such as the chemistry of the specimen.

- Images of this structure produced with the beam along the orthogonal direction would be difficult to explain.

Modulated structures are not confined to the superconductors. In fact, such structures are ubiquitous in the materials world. Many useful engineering alloys exhibit spinodal decomposition and the spinodal wavelength can be directly measured from the extra spots in the DP. Many ceramics are described as polytypes, e.g., SiC or polytypoids, which are just polytypes with the composition fluctuating from layer to layer (e.g., SiAlONs). These structures consist of random or locally ordered stacking of specific atomic layers, which often give predictable effects in the DP. All these materials are particularly amenable to HRTEM analysis, because you can imagine the individual modulations, and still characterize the overall structure with conventional amplitude contrast.

28.18 QUASICRYSTALS

The study of quasicrystals continues to be a challenge for HRTEM, since these materials do not have the translational symmetry that we associate with crystals. However, they are strongly ordered, as you can appreciate from Figure 28.22. The HRTEM image shows many sharp white spots from a stable decagonal quasicrystal of Al-Mn-Pd. The DP from another specimen also shows very strong, clear, well-defined spots. In our earlier discussion of DPs, we associated each spot with a single set of planes, which were present throughout the specimen. Although the quasicrystals do not contain such planes, there is clearly far more order than in an amorphous material. You can indeed see that the spots in the HRTEM image are aligned in certain well-defined directions, but the spacing is difficult to identify. We have a growing understanding of these materials and it appears that the spots in the HRTEM image are this sharp because, at least in decagonal quasicrystals (but not in icosahedral ones), they really do correspond to columns; we don't need translational periodicity along the column. In fact we can rotate the quasicrystal (they can be grown as large as 1 mm) to reveal twofold and threefold axes, as illustrated in Figure 28.23. You can see how this might arise by looking along the rows of spots in Figure 28.22. We can draw some interesting lessons from the use of TEM to study quasicrystals

- HRTEM excels when materials are ordered on a local scale.
- For HRTEM, we need the atoms to align in columns because this is a 'projection technique,' but the distribution along the column is not so critical, and we

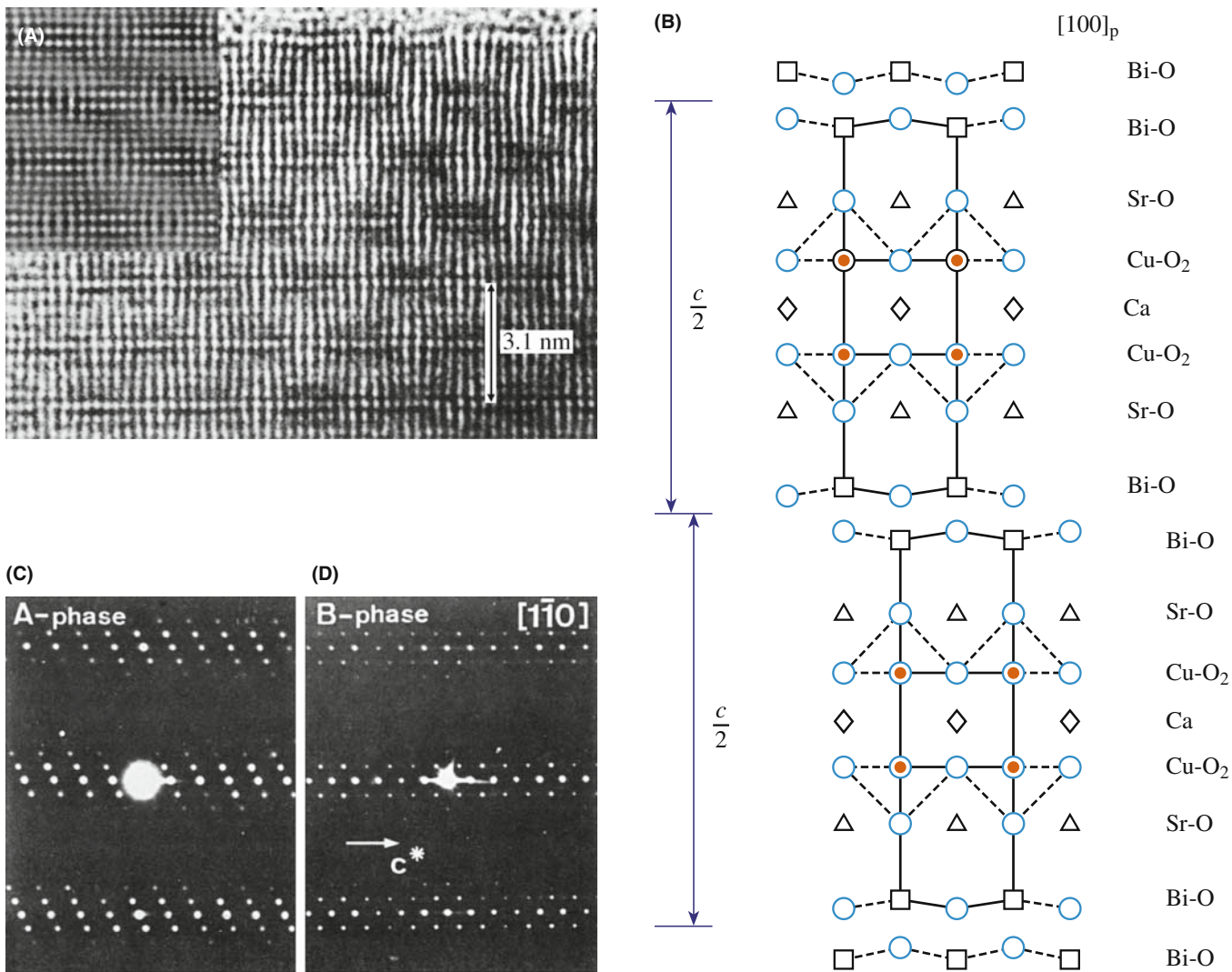


FIGURE 28.21. (A) HRTEM image of the superconductor $\text{Bi}_2\text{Sr}_2\text{Ca}_n\text{Cu}_{n+1}\text{O}_{2n+8}$; this simulation (inset) assumes that the lattice relaxation occurs in the Bi-O layer. (B) For $n = 1$, the structure is built up from blocks which are shifted relative to one another. The DPs are from (C) the $n = 0$ phase and (D) the $n = 1.2$ phase. Notice that the spacings of the spots (the satellite sequence) are different.

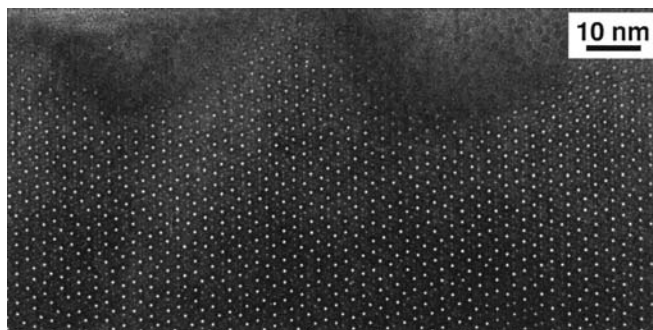


FIGURE 28.22. Tenfold symmetry in a decagonal Al-Mn-Pd quasicrystal.

can't determine it without tilting to another projection in the perfect crystal.

- SAD and HRTEM should be used in a complementary fashion.

28.19 SINGLE ATOMS

You have read that it has become possible to study materials at atomic resolution in the TEM quite recently. So you may be surprised to find that many groups have been reporting studies of individual atoms since about 1970! The techniques used include phase contrast and amplitude contrast in a conventional TEM, and (see Section 22.4) a dedicated STEM. Parsons et al. used mellitic acid molecules stained with uranyl ions from uranyl acetate so the atoms that were imaged are heavy. Parsons et al. then knew that the uranium atoms would be 1 nm apart at each apex of an equilateral triangle and they knew that there were 10^{13} of these per cm^2 supported on a thin (0.8 nm) film of evaporated carbon. One challenge is recognizing that the contrast from the individual uranium atoms reverses as you change defocus, just as we've seen for columns of

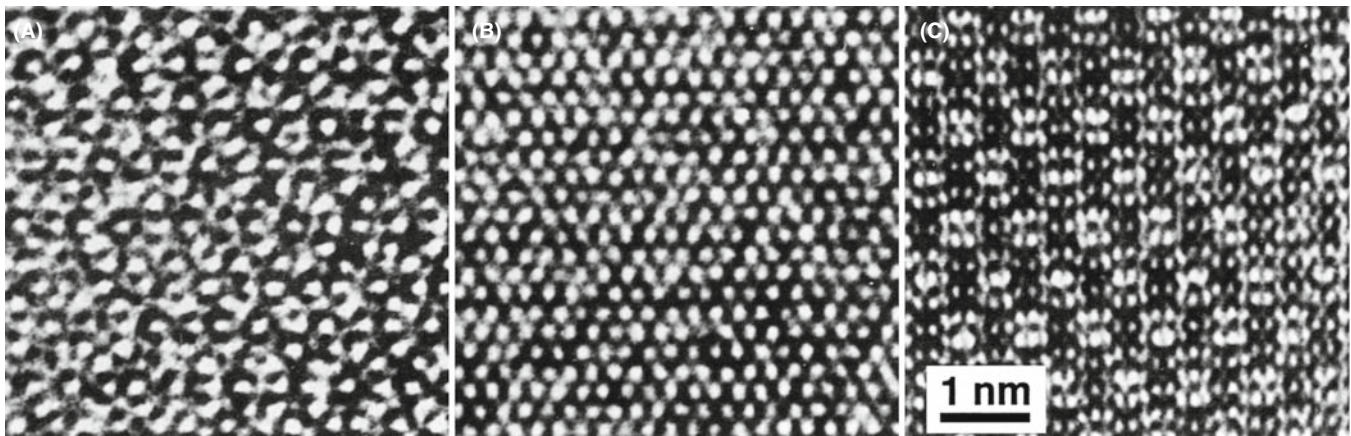


FIGURE 28.23. Images of (A) fivefold, (B) threefold, and (C) twofold projections of an Al-Cu-Li quasicrystal; $\Delta f=27$ nm.

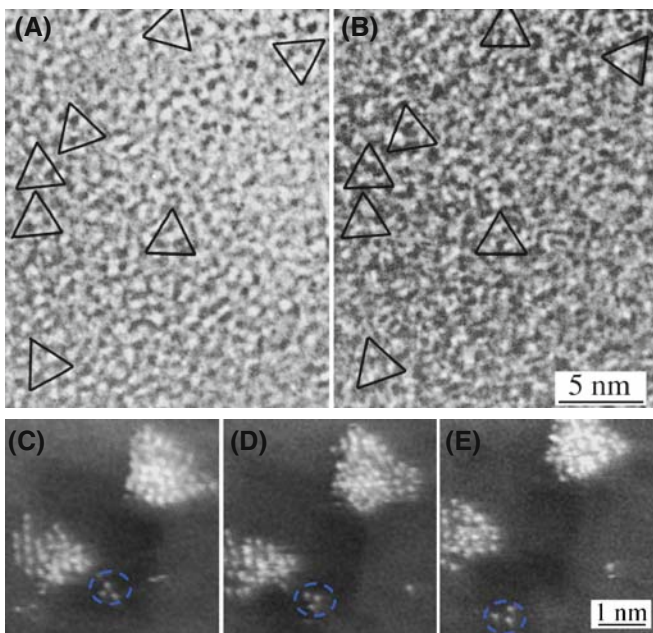


FIGURE 28.24. Images of atoms. (A, B) Triangular arrangements of uranium atoms at different values of defocus in conventional TEM. (C–E) STEM images of Pt/Fe catalyst supported on MgO; time between images is 10–15 seconds. These images have been processed to reduce noise and optimize visibility of the atoms.

atoms (and voids). You can see this effect in Figure 28.24A and B; for comparison, a series of Z-contrast STEM images is shown in Figure 28.24C–E. In this case the contrast is much higher and the movement of the cluster is clear.

Some points to notice

- This is a case where we really do have ‘white atoms or black atoms!’
- Parsons et al. used a Siemens 101 TEM operating at 100 kV with a point-to-point resolution of about 0.33 nm; this is not today’s state-of-the-art machine!
- The specimen in the TEM study was so stable that they could do ‘through-focus’ imaging.

Z-contrast imaging heralded the arrival of the STEM as a real research tool: atoms are now seen to move on the surface, agglomerate, etc. The imaging mode is essentially a high-angle dark-field technique so that the heavy atoms are by far the strongest scatterers and appear bright, as we discussed in Chapter 22. The difficulty with Z-contrast imaging in the STEM is that it requires an FEG, whereas almost any TEM operating today can produce images like those first demonstrated by Parsons et al.

CHAPTER SUMMARY

The major problem that separates this chapter from Chapter 23 is the language. In HRTEM, the language is that of physics or electrical engineering; you can get so involved with the language and the equations that you miss the point. Having said that, you must know the following terms and understand what they mean or don’t mean

- Point-spread function.
- Contrast transfer function (CTF) (and transfer function).
- Weak-phase object approximation (WPOA).

With this understanding of the restrictions that our model involves, we can now consider the simulation of high-resolution images. If you want to delve further into the theory, we

recommend starting with the companion text, then going to the books by Cowley and Spence. You will need to have a strong background in math and physics to appreciate fully the further subtleties of more complex models. Always keep in mind that all of the above discussion was concerned with arriving at models or approximations

- We *model* the effect of the lens.
- Then we *model* the specimen.
- Finally, we combine the two models.

We will take the next two chapters to achieve these three tasks. Although by 2008 ~25 TEMs with C_s correctors had been installed, we leave the discussion of this exciting topic to the companion text since it is certainly not an instrument for someone learning TEM.

REFERENCES

One of the pioneers in the interpretation of HRTEM images was the late John Cowley. Much of our analysis of the specimen transfer function, $f(x,y)$, follows directly from his teaching. When pronouncing names, don't confuse Lord Rayleigh (born John William Strutt) with Walter Raleigh. Otto Scherzer was professor in Darmstadt and actually built an aberration corrector for his TEM. He was succeeded at Darmstadt by Harald Rose who with his former student, Max Haidar, made aberration correction work for the rest of us. Ondrej Krivanek and Nicolas Delby did the same for STEMs. Articles by Shannon and Weaver 1964, Van Dyck 1992 on information theory will start you on this topic; for HRTEM, you must then have access to John Spence's book.

ESSENTIAL FURTHER READING

Buseck, PR, Cowley, JM and Eyring, L Eds. 1988 *High-Resolution Electron Microscopy and Associated Techniques*, Oxford University Press New York.
Horiuchi, S 1994 *Fundamentals of High-Resolution Transmission Electron Microscopy*, North-Holland Amsterdam.
Spence, JCH 2003 *High-Resolution Electron Microscopy*, 3rd Ed., Oxford University Press New York.

NEW INSIGHTS

Haider, M, Müller, H, Uhlemann, S., Zach, J, Loebau, U and Hoeschen, R 2008 *Prerequisites for a C_c/C_s -Corrected Ultrahigh-Resolution TEM* *Ultramicrosc.* **108** 167–178.
Hawkes, PW 1980 *Units and Conventions in Electron Microscopy, for Use in Ultramicroscopy* *Ultramicrosc.* **5**, 67–70. An enjoyable and informative diversion. Includes definitions of the Glaeser and the Scherzer units (very non-SI).
Vladár, AE, Postek, MT, and Davilla, SD 1995 *Is Your Scanning Electron Microscope Hi-Fi?* *Scanning* **17**, 287–295. Looks at the SEM as a hi-fi instrument.

DELOCALIZATION

Coene, W and Janssen, AJEM 1992 in *Signal and Image Processing in Microscopy and Microanalysis*, *Scanning Microscopy Supplement 6* (Ed. PW Hawkes), p. 379, SEM inc. O'Hare IL.
Lichte, H 1991 *Optimum Focus for Taking Electron Holograms* *Ultramicrosc.* **38**, 13–22.

INFORMATION LIMIT AND INFORMATION THEORY

Van Dyck, D 1992 in *Electron Microscopy in Materials Science*, p 193, World Scientific, River Edge New Jersey.
Gives more detail on the derivation of equations 28.19 and 28.55 and on information theory for the TEM.
Van Dyck, D and De Jong, AF 1992 *Ultimate Resolution and Information in Electron Microscopy: General Principles* *Ultramicrosc.*, **47** 266–281. On the information limit—part 1.
de Jong, AF and Van Dyck, D 1993 *Ultimate Resolution and Information in Electron Microscopy. II. The Information Limit of Transmission Electron Microscopes* *Ultramicrosc.* **49**, 66–80. On the information limit—part 2.
Shannon, CE and Weaver, W 1964 *The Mathematical Theory of Communication*, University of Illinois Press Urbana IL. Early text on information theory.

MORE THEORY

Cowley, JM 1992 in *Electron Diffraction Techniques 1*, (Ed. JM Cowley), p. 1, IUCr. Oxford Science Publication. More extensive discussion of equations 28.20 and 28.21.
Cowley, JM 1995 *Diffraction Physics*, 3rd edition, North-Holland, Amsterdam.

- Fejes, PL 1977 *Approximations for the Calculation of High-Resolution Electron-Microscope Images of Thin Films* Acta Cryst. **A33**, 109–113.
- Hall, CE 1983 *Introduction to Electron Microscopy*, Krieger New York. An early clear discussion of other lens aberrations.
- Hashimoto H and Endoh H 1978 in *Electron Diffraction 1927–1977*, (Eds. PJ Dobson, JB Pendry and CJ Humphreys), p. 188, IoP, Bristol. The aberration-free focus paper.
- Otten, MT and Coene, WMJ 1993 High-Resolution Imaging on a Field Emission TEM Ultramicrosc. **48**, 77–91. Some practical difficulties, including the use of the wobbler, when using FEGs.
- Rose, H 1990 *Outline of a Spherically Corrected Semiaplanatic Medium-Voltage Transmission Electron Microscope* Optik **85**, 19–24. One of two early articles by the father of today's aberration correctors.
- Rose, H 1991 in *High Resolution Electron Microscopy: Fundamentals and Applications*. (Eds. J. Heydenreich and W Neumann), p. 6, Institut für Festkörperphysik und Elektronmikroskopie, Halle/Salle, Germany.
- Thon, F 1975 in *Electron Microscopy in Materials Science*, (Ed. U Valdrè) p. 570, Academic Press New York.

MATERIALS APPLICATIONS

- Amelinckx, S, Milat, O and Van Tendeloo, G 1993 *Selective Imaging of Sublattices in Complex Structures* Ultramicrosc. **51**, 90–108.
- Bailey, SJ 1977 *Report of the International Mineralogical Association (IMA)-International Union of Crystallography (IUCr) Joint Committee on Nomenclature* Acta Cryst. **A33**, 681–684. Definitions of polytypes, polytypoids, etc.
- Butler, EP and Thomas, G 1970 *Structure and Properties of Spinodally Decomposed Cu-Ni-Fe Alloys* Acta Met. **18**, 347–365. Classic study of DPs from spinodal decomposition.
- Carter, CB, Elgat, Z and Shaw, TM 1986 *Twin Boundaries Parallel to the Common-{111} Plane in Spinel* Phil. Mag. **A55**, 1–19. Early study of spinel showing that you don't always need the highest resolution. See also p21–38.
- Nissen H-U, and Beeli, C 1991 in *High Resolution Electron Microscopy: Fundamentals and Applications*, (Eds. J Heydenreich and W Neumann) p. 272, Institut für Festkörperphysik und Elektronmikroskopie Halle/Salle Germany. HRTEM from quasicrystals (Figure 28.22).
- Parsons, JR, Johnson, HM, Hoelke, CW and Hosbons, RR 1973 *Imaging of Uranium Atoms with the Electron Microscope by Phase Contrast* Phil. Mag. **29**, 1359–1368. Individual Pt atoms in 1973!
- Rasmussen, DR, Summerfelt, SR, McKernan, S and Carter, CB 1995 *Imaging Small Spinel Particles in an NiO Matrix* J. Microsc. **179**, 77–89.
- Van Landuyt, J, Van Tendeloo, G and Amelinckx, S 1991 in *High Resolution Electron Microscopy: Fundamentals and Applications*, (Eds. J Heydenreich and W Neumann), p. 254, Institut für Festkörperphysik und Elektronmikroskopie Halle/Salle Germany.

THE COMPANION TEXT

The companion text includes a complete chapter by Van Aert and Van Dyck on direct imaging methods for HRTEM. Aberrations are discussed by Kruit (on lenses) and Haider (on aberration correction). Since HRTEM is changing rapidly, the companion text is our way of keeping more current.

SELF-ASSESSMENT QUESTIONS

- Q28.1 If $f(\mathbf{r})$ describes the specimen function and $g(\mathbf{r})$ describes the image, write down the relationship between the two.
- Q28.2 The function $h(\mathbf{r})$ is given two names in the text, state the two names and explain why each is used.
- Q28.3 Describe in 30 words or less how a TEM is like a radio.
- Q28.4 $H(\mathbf{u})$ is described as a product of three or more terms. What are they and why do we use the product?
- Q28.5 Write down an expression for $\chi(\mathbf{u})$.
- Q28.6 Justify the definition of the term 'overfocus.'
- Q28.7 In the weak-phase object approximation, the electrons experience a phase change in passing through a sample thickness dz . Write down an expression for the phase shift.
- Q28.8 Write down Cowley's expression for the transfer function of a weak-phase object with absorption.
- Q28.9 Give a numerical value for the maximum thickness for $\text{Ti}_2\text{Nb}_{10}\text{O}_{27}$ in the WPOA.
- Q28.10 Write down an expression defining $T(\mathbf{u})$, the objective lens transfer function.
- Q28.11 What is the problem with using the term 'contrast transfer function' in TEM?
- Q28.12 A point in the specimen is imaged as a disk of radius $\delta(\theta)$ in the image. Write down an expression for $\delta(\theta)$ and use it to deduce an expression for $\chi(\mathbf{u})$.
- Q28.13 Sketch the graph of $T(\mathbf{u})$ versus \mathbf{u} with and without C_c .
- Q28.14 Write down an equation for Scherzer defocus and explain the sign of Δf .

- Q28.15 Write down an expression for u_{Sch} and hence write an equation for r_{Sch} .
- Q28.16 What is coma-free alignment?
- Q28.17 What is the special feature of the Rose corrector?
- Q28.18 What is image delocalization and why is it so important in an FEG TEM?
- Q28.19 Atoms were seen in ~ 1972 . How can this be true if the resolution of the TEM used was only 0.33 nm?
- Q28.20 Can we image incommensurate structures by HRTEM? Explain and justify your answer.

TEXT-RELATED QUESTIONS

- T28.1 Explain fully why the WPOA is only valid if the specimen thickness is < 0.6 nm for $\text{Ti}_2\text{Nb}_{10}\text{O}_{27}$.
- T28.2 Explain fully the connection between equations 28.29 and 28.34.
- T28.3 In Section 28.7, we say “and χ is -120° .” Why do we say this?
- T28.4 Explain why it is important that the TEM is a linear system.
- T28.5 Choose two TEMs, one 100 kV from 1970 and the other 300 kV from 2008. Plot χ for each using equation 28.34 and Excel (or similar).
- T28.6 Consider a group of TEMs with $C_s = 3, 2.8, 2.6, \dots, 0.6$ mm (or just one with variable C_s) operating at 200 kV. Plot the values for the resolution at Scherzer defocus.
- T28.7 By delving into the literature, explain how equations 28.43 and 28.44 were deduced.
- T28.8 Using equation 28.47 construct a table for d , varying ΔE from 0.1 to 1.0 eV in increments of 0.1 eV for TEMs operating at 100, 200, and 300 kV.
- T28.9 Discuss (using the literature) why we write $E_s(\mathbf{u})$ as shown in equation 28.53.

Other Imaging Techniques

CHAPTER PREVIEW

Much of what we've discussed in the preceding imaging chapters is what we might call 'classical' TEM imaging. It began with BF and DF techniques and quickly expanded to include many beams. Diffraction contrast, phase contrast, and to a lesser extent, mass-thickness contrast are the mechanisms we use to characterize our specimens. We control the contrast by inserting the objective aperture, or a STEM detector, and excluding or collecting electrons that have been scattered by the different processes. However, there are variations to the standard ways in which we can extract more information from a TEM image; in this chapter, we'll present a brief overview of some of them. Most of these operational modes that we'll discuss here are somewhat esoteric and have rather specialized applications. Nevertheless, you should know that they exist because they may be just what you need to solve your particular problem. There's no importance to the order in which we go through the various modes, but we'll cover modifications to conventional parallel-beam TEM imaging as well as those techniques that require STEM and use some of the electron detectors we discussed in Chapter 7. It turns out, however, that the various procedures are often feasible in either TEM or STEM mode.

This is a bit of a potpourri of a chapter but the techniques are not, to our knowledge, gathered together in any other text. The descriptions will, of necessity, be brief but we'll reference suitable source material so you can follow up if you really want to try the technique for yourself.

29.1 STEREO MICROSCOPY AND TOMOGRAPHY

This section is an excellent example of why we've included this catch-all chapter. In the first edition of this textbook, stereo microscopy was an old technique only used by biologists and tomography was used in an MRI. Now tomography fills an entire chapter in the companion text. The principle is, of course, the same as stereo microscopy: record images from more than one direction to give a 3D view of the sample.

You've have realized by now that any TEM or STEM image is a two-dimensional projection of a 3D specimen and this is a fundamental limitation. Sometimes we can discern differences in diffraction-contrast images of certain defects, depending on whether the defect is intersecting the top or the bottom of the foil, but generally we lose the depth dimension. To regain this depth information we use stereo microscopy, but only for features showing mass-thickness or diffraction contrast. We cannot use stereo for phase-contrast imaging because the essential experimental step, tilting the specimen, changes the phase contrast and the

projected potential of the specimen. So any stereo effect in the image is lost. You may need stereo microscopy if, for example, you want to know whether precipitates have formed on your specimen surface rather than in the interior, or if you want to see how dislocations are interacting with each other.

Stereo imaging works because your brain gauges depth by simultaneously interpreting signals from both your eyes, which view the same scene from slightly different angles (about 5°), giving a parallax shift. So in the TEM, if you take two pictures of the same area but tilted a few degrees relative to each other, then present the two images simultaneously to your brain using a stereo viewer, you'll see a single image in which the different depths of the features are apparent. In fact, some people are able to see the stereo effect without the aid of a viewer and some people are incapable of discerning the stereo effect at all.

To see in stereo, the two images should be separated by ~60 mm, but in practice it's often sufficient just to move the pictures relative to each other until your eye and brain seize on the effect.

STEREO RULES

You have to make sure that the field of view, the contrast, and the magnification in each of the two images are the same.

2D TOMOGRAPHY

The stereo technique of the future: the data is processed by a computer, not by your brain.

A couple of points are worth noting before we describe the method. First, if the features you want to observe show diffraction contrast, then the only way to maintain contrast is to tilt along a Kikuchi band, keeping both \mathbf{g} and \mathbf{s} fixed; so tilt while looking at the DP. This procedure almost invariably requires a double-tilt stage and may be difficult, or impossible, if your specimen is heavily deformed. If you just want to measure the foil thickness, any tilt is sufficient and contrast does not have to be maintained. Second, if you want to be pedantic, there is a right way and a wrong way to view the stereo images. You have to present the images in the same way that your eyes would see a scene, that is, the two images have to be correctly positioned, otherwise the brain will interpret depth the wrong way round. If you are trying to perceive the true surface topography (such as with SE images) using SEM or STEM images, then the choice of which image goes into the left eye and which into the right is crucial. (See any SEM text, such as that by Goldstein et al., for more details on stereo viewing.) Of course, for TEM images this difference is irrelevant. TEM applications are reviewed by Hudson and a whole set of related papers appears in that same issue of *Journal of Microscopy*.

So if you want to take a stereo pair, follow these steps

- Select the region of interest, making sure that the specimen is eucentric.
- Record an image (BF, DF, or WBDF, it doesn't matter, although usually the BF image is used).
- Tilt the specimen by at least 5° (much higher tilts give a larger parallax shift but it's more difficult to keep the focus and the diffraction contrast constant).
- Ensure that the whole field of view didn't move while you tilted. If it did, translate it back to its original position (using the beam stop as a point of reference if you wish). All features in the image will be shifted slightly relative to each other: it is this parallax shift which your brain interprets in stereo.
- If the area is now out of focus (it will be if you had to use the second, non-eucentric, tilt axis), refocus *using*

the specimen-height (z) control; otherwise you'll change your image magnification. Obviously, computerized stages will help in this respect.

- Record another image.
- Develop or display the images and observe them under a stereo viewer.

Figure 29.1 shows a pair of BF images showing precipitates. If the images are correctly spaced and you look through a stereo viewer then you should be able to see the relative depth of the precipitates. You can purchase cheap cardboard stereo viewers through any EM supplier. Although a proper stereo viewer is an expensive optical tool, you can use it to calculate the relative depth (Δh) of a feature in a stereo pair, since

$$\Delta h = \frac{\Delta p}{2 M \sin \frac{\phi}{2}} \quad (29.1)$$

where Δp is the parallax shift between the same feature in the two images tilted by ϕ at a magnification M . Be careful how you define ϕ because some microscopists define the tilt angle as $\pm\phi$, in which case $\sin \phi/2$ becomes $\sin \phi$. For true depth determination you need to deposit some recognizable feature on the surface, such as gold islands, but that is not usually important in TEM images and relative depth is often sufficient. If you need quantitative stereo measurements, you can either learn about the field of stereology, which is an old established discipline or you can learn the tomography technique.

We will discuss electron tomography in the companion text. This is a very large topic and an emerging activity in TEM. It has been limited until recently by the need for fast computers and large amounts of memory, and specimen holders that allow you to tilt the specimen through nearly the full 360° . Obviously, disk specimens are not ideal for such tilting experiments. The principle involved is just as we've described for stereo, but we use about 100 times as many images and a computer is essential to analyze and view the results.

29.2 2½D MICROSCOPY

This is an example of imaginative terminology in the TEM field. If there are diffraction spots in the SADP which are too close together to give separate DF images, center the objective aperture around them all. Then, if you view two DF images taken at different focus settings through a stereo viewer, you see features at different apparent depths. However, you're using a pseudo-stereo technique because the 'depth' difference is due to a difference

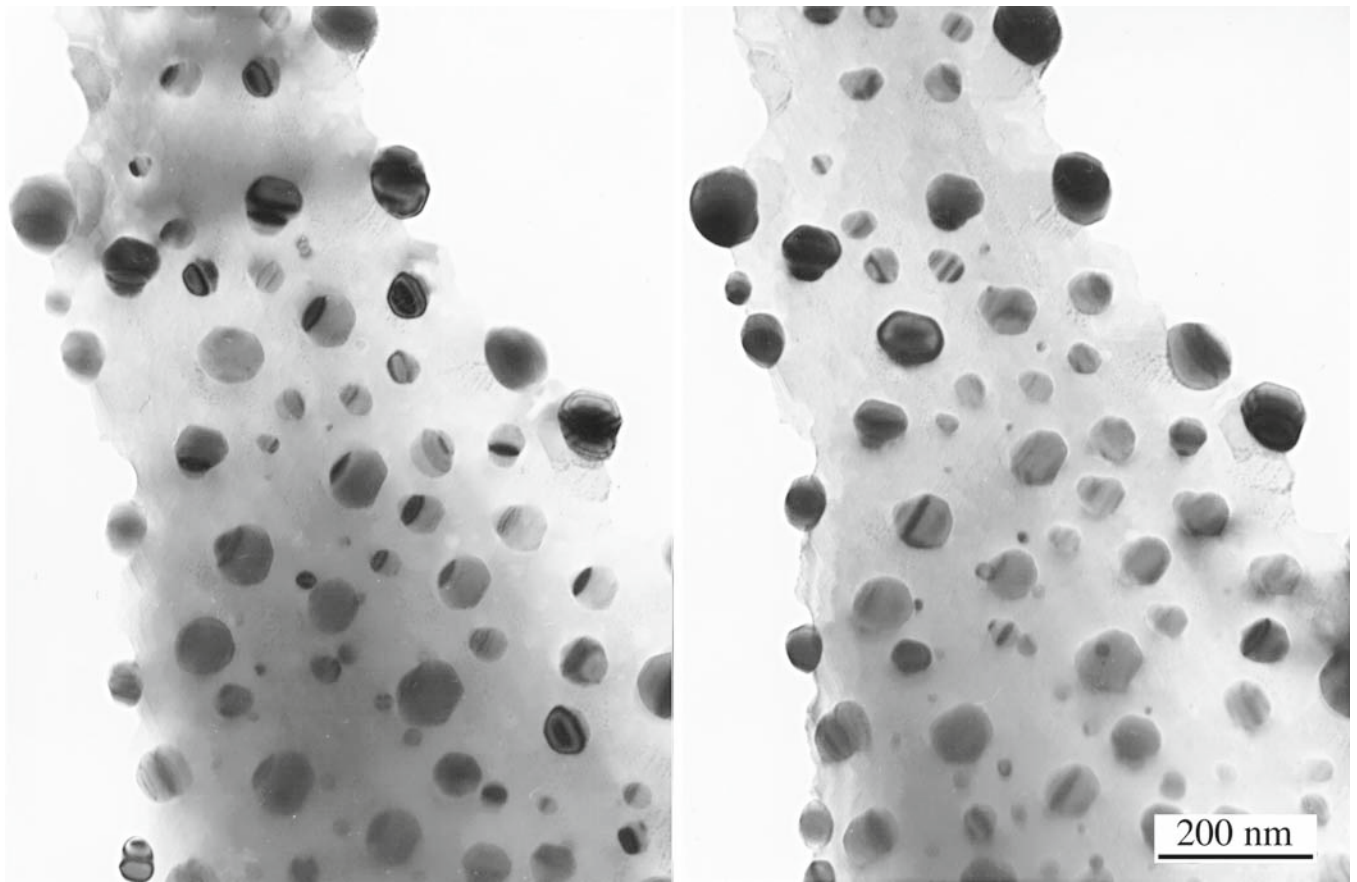


FIGURE 29.1. A stereo pair of spinel precipitates in (and on) a NiO-Cr₂O₃ specimen which, when combined, show the relative depth of the precipitates. You can see the small parallax shift between the two images. Note also the mixture of mass-thickness and diffraction contrast.

in \mathbf{g} , not a true depth difference at all: hence the term $2\frac{1}{2}D$ or ‘not quite 3D.’ The technique is also known by the rather more staid name of ‘through-focus DF.’

Bell developed a simple theory explaining how a change in focus Δf introduces a parallax shift in the image, y .

$$y = M\Delta f \lambda \mathbf{g} \quad (29.2)$$

$$\Delta y_{12} = \Delta f \lambda \mathbf{g}_{12} \quad (29.3)$$

where M is the magnification, λ is the electron wavelength, and \mathbf{g} is the diffraction vector. So, as you change focus, the parts of the image coming from different diffracted beams shift relative to each other by an amount, $M\Delta y_{12}$.

The term \mathbf{g}_{12} ($= \mathbf{g}_1 - \mathbf{g}_2$) is the vector between the spots arising from features 1 and 2. It is this parallax shift which introduces a stereo effect in the same manner as one introduced through tilting.

When would you need to use such a technique? Well, if you’re dealing with a multiphase specimen in which many of the phases have similar structures and/or lattice parameters, then there would be many closely spaced diffraction spots. In these circumstances, if you succeed in separating the spots with a small objective aperture,

the selected spot will undoubtedly be very close to the aperture and the image will display serious aberrations. It’s more likely you will be unable to select only one spot in the aperture.

You must start by setting up a CDF imaging configuration, but one in which several diffraction maxima are contained within the objective aperture. In the normal CDF image, therefore, all the parts of the specimen diffracting into the several spots will appear bright. However, if you take two images at different defocus conditions, the features diffracting into different spots will shift by slightly different amounts given by equation 29.3. This relative shift between the two images is perceived by your brain as a relative depth difference when you view the images simultaneously through a stereo viewer. You might argue that your images are going to be out of focus, and strictly speaking you’re right, but remember that in the TEM there’s a large depth of field (see Section 6.7) and it’s easy to keep both images reasonably in focus over quite a range of objective lens excitation, so long as the magnification isn’t too high. The greater the change in the objective lens current, the better the ability to resolve similarly diffracting features, so this process works better at lower magnifications where the depth of field is greater.

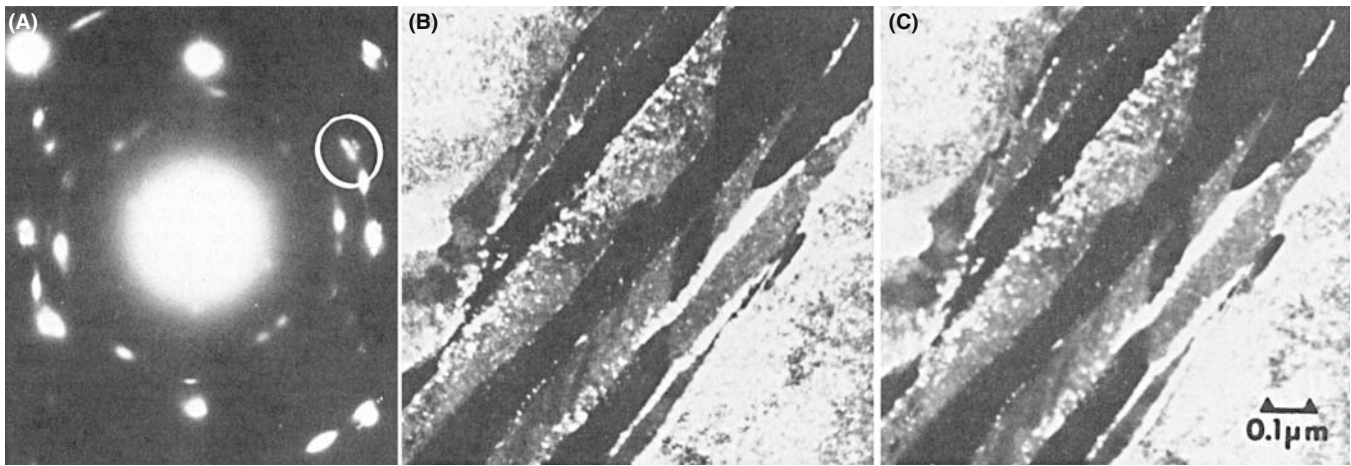


FIGURE 29.2. (A) SAD pattern of retained austenite and carbide precipitates in steel. (B, C) Stereo pair showing $2\frac{1}{2}$ D effect in which the relative depth of austenite and carbide is related to the position of their diffraction spots in the objective aperture.

So the experimental steps are as follows

- Select the area of interest making sure the specimen is eucentric.
- Tilt the beam so that the several diffraction spots cluster around the optic axis; center the objective aperture around this group of spots.
- Return to image mode and *underfocus* the objective lens until you see the clarity of the image begin to degrade; then *overfocus* one click back to maintain focus.
- Record an image.
- *Overfocus* the objective lens until the image begins to lose clarity again, then *underfocus* back one click.
- Record another image.
- Develop or display the images and view them in a stereo viewer.

Figure 29.2 shows a pair of CDF images taken under these conditions. The DP, as you can see, has many closely spaced spots. This example and many more are given in a well-illustrated review by Sinclair et al. If you look at the images through a stereo viewer, different bright regions will appear at different ‘heights.’ If you read the review, you’ll find that it is quite easy to determine which features in the image are responsible for which diffraction spot.

29.3 MAGNETIC SPECIMENS

If you happen to be looking at a magnetic material, then this can cause major problems for both you and the TEM and you might want to switch to studying aluminum. However, if you’re patient and want a challenging task, then you can learn to correct for the magnetic disturbance introduced by the specimen. You can also get more information by making use of the interaction

of the electron beam with the magnetic field of the specimen. The use of TEM to study magnetic effects has seen a resurgence since the discovery of high- T_c superconductors, and the interest in magnetic recording media capable of storing a high density of information.

MAGNETIC DISKS

You should not use self-supporting magnetic disk specimens if you can avoid doing so.

First, we’ll look at how to get the best images from your magnetic specimen and then we’ll describe a couple of specialized imaging techniques that allow you to see either the domains or the domain walls, which are relevant to both ferromagnetic and ferroelectric materials. This topic will be extended in the companion text.

29.3.A The Magnetic Correction

If your specimen is magnetic, its magnetic field will deviate the electron beam as it passes through and then the electrons that you use to form images will not be on the optic axis. All your images will be severely aberrated and shift when you try to focus them. You can minimize these effects as we’ll now describe.

The most important step is to make the magnetic part of your specimen as thin and small as possible to reduce its total magnetic field strength.

However, small flake specimens are more easily pulled out from between support grids when you put the specimen into the objective lens. So, make sure your foil is well clamped into the specimen holder and use oyster grids for thin flakes with conducting glue if possible. Insert the holder into the lens with the objective lens strength as low as possible; look at the current running through the lens and minimize it. Ideally, you will use a TEM that has been optimized for the task.

Such a TEM may, for example, have a modified objective lens. Notice that none of the images shown in this section are HRTEM.

MAGNETIC MATERIAL

If you lose any specimen, or part of a specimen, in the TEM it is essential that you stop work and get help; if the specimen is magnetic this is doubly essential—and you can be identified if you don't get help!

You'll find it very difficult to set a magnetic specimen to the eucentric height because, as you tilt it through the 0° position, the specimen tends to rotate, or it is pulled out of the holder (if you're unlucky) or it shifts slightly in height and position. If you lose the specimen, the column must be split and the lost specimen retrieved, otherwise you may introduce a fixed, non-correctible, astigmatism into the microscope. In the extreme, you may end up with your specimen welded to the polepiece, which gets to be quite expensive. It is better to incur the wrath of your technical support staff when the objective polepiece is still undamaged.

So when you have found a general area of interest, set up the eucentric height by tilting the specimen, taking care *not* to tilt through 0° ; keep the tilt range to one side of zero. When the eucentricity is reasonable (and it rarely gets perfect for magnetic materials), focus and center a recognizable feature on the screen, remove the objective diaphragm, then carry out the following steps

- Underfocus C2 and the objective lens by one coarse step. The feature should move off center. If you lose illumination you haven't underfocused C2 enough.
- Return the feature to the screen center using the DF beam tilt potentiometers. This brings the imaging beam onto the optic axis.
- Refocus the image, recondense the beam, and if the feature shifts, recenter it with the stage translators.
- Repeat the procedure for objective lens overfocus, until the feature stays centered as the objective lens is moved through focus

If you then go to the DP, you'll see that the 000 beam should now be on axis and you can put the objective diaphragm back in again and check the astigmatism. However (and here's the catch), if you tilt or traverse the specimen, then you will have to redo the whole correction since the magnetic field will change as you move the specimen relative to the beam. You'll rapidly get accomplished at this procedure if you practice for a while. Now, if you want to do CDF imaging at the same time as doing the magnetic correction, then you need an extra set of external DF controls, which may be an option or may be built into your microscope. Then you center the image

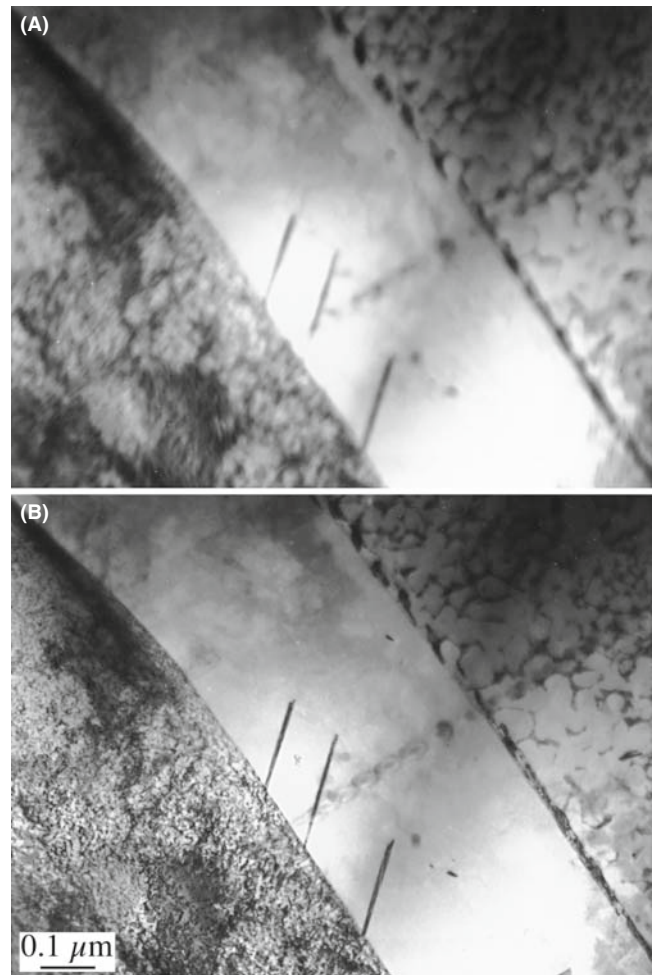


FIGURE 29.3. The effect of the magnetic correction on the quality of the image of an Fe-Ni meteorite: (A) before magnetic correction, (B) after magnetic correction.

feature with one set of potentiometers and tilt in the required hkl maximum with the other set. Figure 29.3 shows the image of a magnetic specimen before and after the magnetic correction. It is more difficult to image magnetic specimens in STEM mode, because the scanning beam may interact differently with different areas of the specimen and so the image quality will be variable.

29.3.B Lorentz Microscopy

When you've corrected for the magnetic field introduced by the specimen, you can image the magnetic domains if they're on the right scale to see in the microscope. This is a form of phase-contrast microscopy which we saw back in Chapter 23. The general term for this kind of imaging is called Lorentz microscopy and it comes with two options, and a special lens.

THE LENSES

The objective lens produces a magnetic field around the sample. A Lorentz lens replaces the objective lens: it gives a lower magnification and is farther from the sample.

Lorentz? Because we use the Lorentz force.

Foucault Images: If there are several domains in the illuminated area, the electron beam will be deviated in different ways by different domains. This will result in a splitting of the diffraction spot, as shown in Figure 29.4A. It is useful to know immediately that the

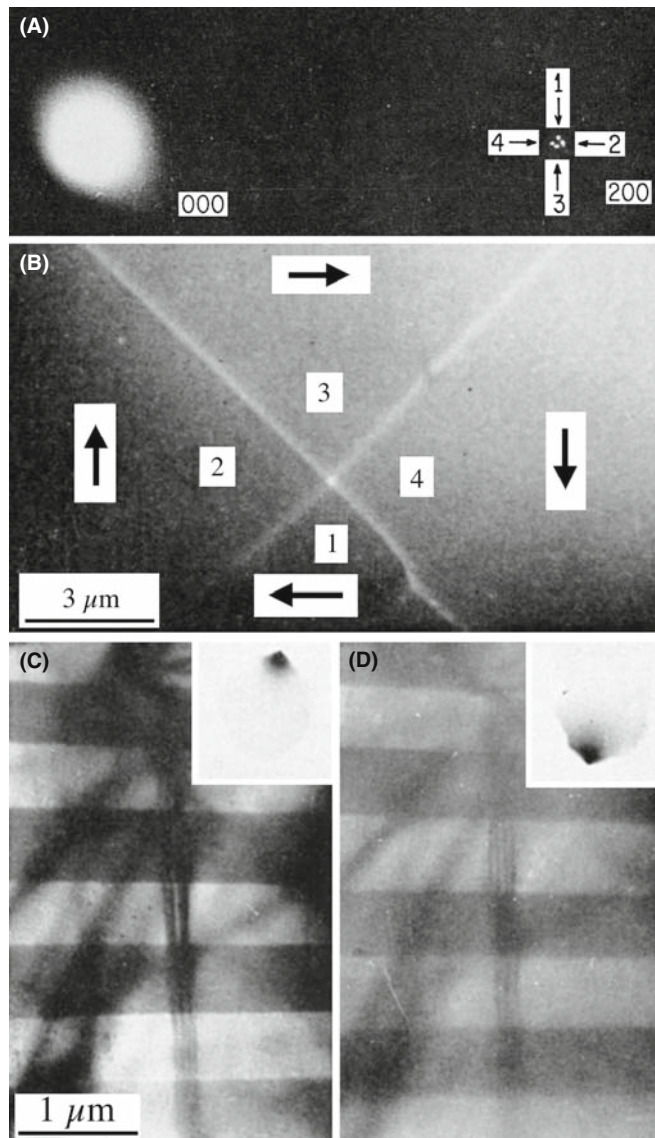


FIGURE 29.4. (A) Splitting of the 200 spot from Ni_3Mn due to the presence of magnetic domains. (B) Image taken from all of the split spots in (A) showing the four domains which scattered electrons into the various spots. (C, D) Foucault images of domains in Co formed by displacing the objective aperture to select one of the two split spots as shown in the insets.

direction of magnetization is normal to the direction of spot splitting (the effect of the Lorentz force). You can then take a BF image using either of the split 000 reflections. If you use all the spots, you can just see the domain walls (Figure 29.4B). The domains bending electrons into the chosen spot will then appear bright and the other domains will appear darker, as in Figure 29.4C. These images of the like domains are called Foucault images (perhaps after the French inventor of the pendulum used to demonstrate the earth's rotation) for reasons unknown to the authors. It is possible to see an analogous effect in STEM images if you use a detector that is cut into electrically isolated segments so different portions (usually halves or quadrants) pick up electrons coming through different domains. In STEM you can also add, subtract, or divide the various signals, as shown in Figure 29.5, because you are picking up each signal digitally. However, the intensity in Foucault images cannot be related quantitatively to the magnetic induction, so their only use is to give a rapid estimate of the domain size.

Fresnel Images: This option allows you to see the domain walls rather than the domains. As we discussed in Chapter 23, Fresnel imaging is named for another famous Frenchman whose TEM images would never be in focus. If you over- or underfocus the objective lens, then the electrons coming through different domains will produce images in which the walls appear as bright or

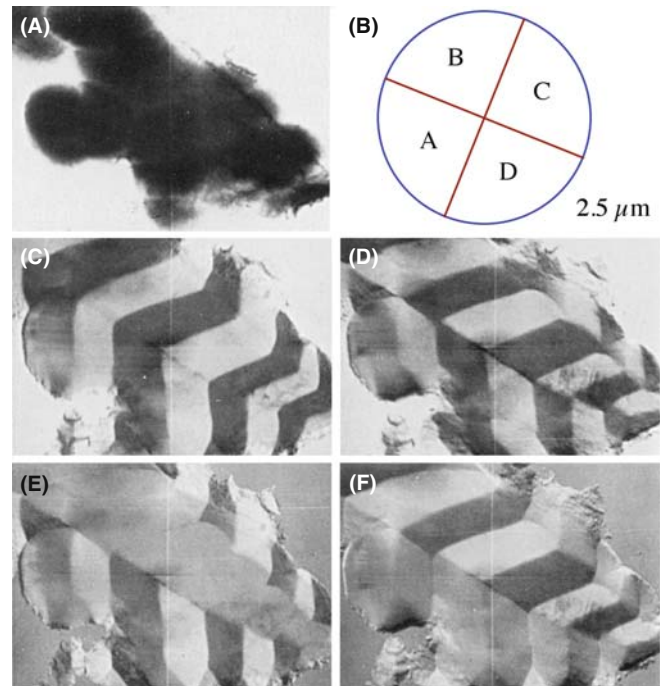


FIGURE 29.5. The use of a quadrant detector in STEM to differentiate regions of different magnetic induction in pure Fe. (A) The BF image from all four quadrants is shown schematically in (B). (C-F) Images formed by different quadrant combinations. Regions showing the same intensity are regions of like magnetic induction.

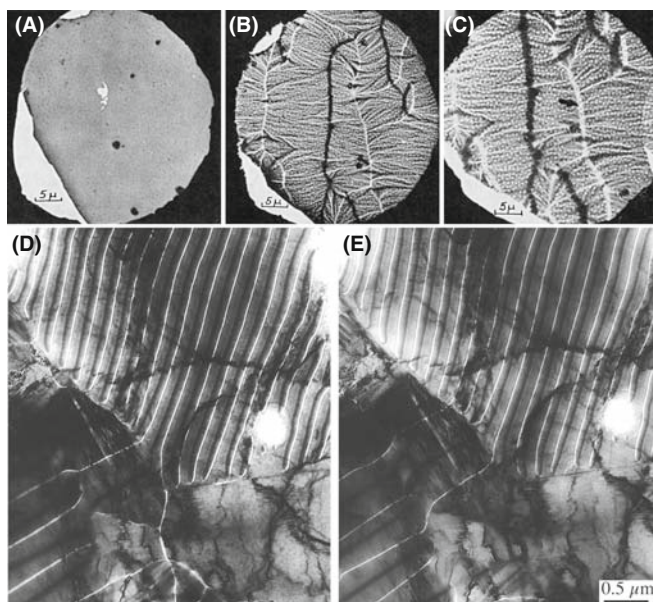


FIGURE 29.6. (A) In-focus image showing no magnetic contrast. (B, C) Fresnel-defocus images showing the magnetic domain walls which image as bright and dark lines and increase in width with defocus. (D) Under-focus and (E) overfocus images showing reversal of domain-wall contrast.

dark lines, which reverse contrast as you go through focus, as in Figure 29.6. The contrast depends on whether the electrons going through the domains on either side of the wall were deflected toward or away from each other, as shown schematically in Figure 23.16A. From such images you can work out whether you're looking at Bloch walls, Néel walls, or cross-tie walls. Figure 29.7 explains why this technique is so important.

In an FEG TEM, the highly coherent source means that coherent Fresnel and Foucault (CF) imaging is possible. These techniques give quantitative measurements of the magnetic induction.

If you want to do Lorentz microscopy, you have to decrease the field strength of your objective lens, which will otherwise dominate the internal field and thus control the domain size in the specimen. So, you either switch off the objective lens and use the intermediate lens for focusing, or use a specially designed low-field lens. TEM manufacturers offer appropriate objective lenses for Lorentz microscopy and it can also be done in STEM.

In addition to imaging the domains and the walls, you can also image the flux lines in the specimen if you evaporate Fe on the surface, just as you use Fe filings to delineate flux lines around a bar magnet. If you heat the specimen *in situ*, then the spot splitting decreases linearly to become zero at the Curie temperature.

29.4 CHEMICALLY SENSITIVE IMAGES

In Section 16.7, we showed that in many materials the structure factor, F , for some reflections was sensitive to the difference between the atomic-scattering amplitudes

of the constituents. If we form a DF image using such a reflection, it will, in principle, be sensitive to changes in the composition of the material. The material systems which have been studied most extensively are related to GaAs and the other III–V compound semiconductors. There is great interest in partially replacing either the group III element or the group V element locally to produce superlattices and quantum wells, as illustrated in Figure 29.8. The contrast in the DF images will be brighter in thin specimens when the difference in the two atomic-scattering amplitudes is large. Therefore, layers of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ appear brighter than the surrounding GaAs matrix.

This imaging technique can, in principle, be applied to many different materials; it is the same, in principle, as the use of superlattice reflections to image ordered regions (see Figures 16.5 and 16.6). We will use the same information when discussing quantitative chemical lattice imaging in Section 31.11; we also use this effect when studying site location by ALCHEMI (see Section 35.9). In practice, it is often important to know how abruptly the composition changes, since this interface affects the properties of the material, so you may be interested in the change in contrast which occurs exactly at the interface. However, you should keep in mind our discussion (Section 24.10) of surface relaxations in the thin specimen which can influence any diffraction-contrast images.

29.5 IMAGING WITH DIFFUSELY SCATTERED ELECTRONS

This is just a variation on the theme of normal DF imaging. If your specimen contains non-crystalline regions which scatter electrons into a different region of reciprocal space away from the diffracted beams, then the non-crystalline regions can be seen in strong contrast. To do this, you perform a DF operation with the objective aperture away from any strong diffraction spots but at a position to intercept a fraction of the diffuse scatter. For example, in silicate glasses the diffuse scatter peaks radially at $3\text{--}4\text{ nm}^{-1}$. As shown in Figure 29.9, a DF image reveals the amorphous regions with high contrast at grain boundary in a ceramic bicrystal. However, you must be very careful when interpreting such images, as shown by Kouh et al.

Diffuse scatter can also arise from short-range ordering in the specimen due either to microdomains of ordered nuclei or local regions of increased order analogous to spinoidal decomposition. These regions produce diffuse intensity maxima at positions that will eventually correspond to a superlattice spot when the short-range order has developed to long-range order (Cowley 1973a,b). A DF image from the diffuse scatter will reveal the short-range ordered regions as diffuse

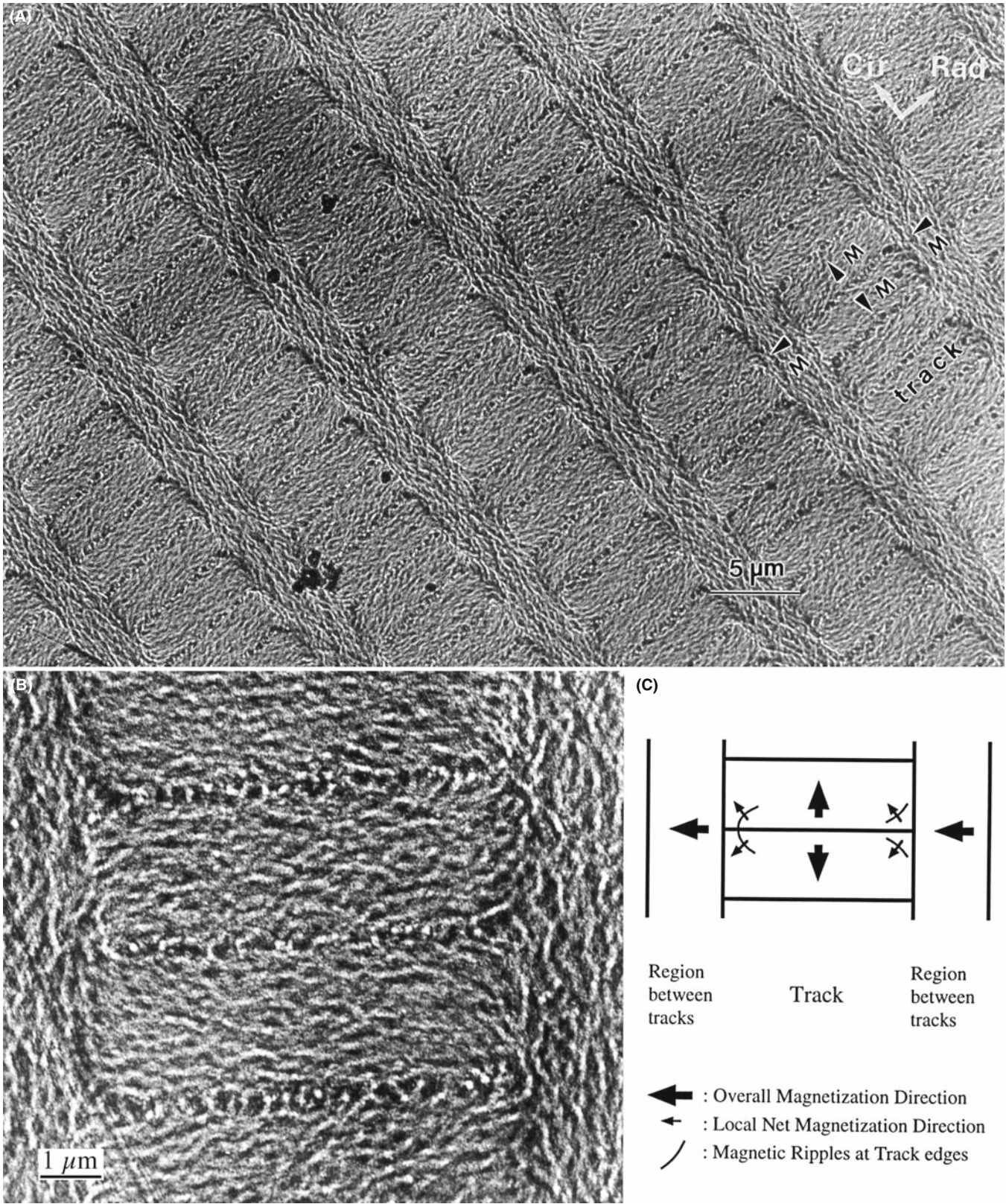


FIGURE 29.7. (A) Fresnel image of $\text{Co}_{84}\text{Cr}_{10}\text{Ta}_6/\text{Cr}$ film on a smooth NiP/Al substrate used for magnetic data storage, imaged in the bits-written magnetic state. (B) Fresnel image at higher magnification. (C) Schematic of magnetic ripples at the track edges. The bits were in alternating direction of magnetization along the tracks in the circumferential direction of the hard disk, while the inter-track regions had remnant magnetization in the radial direction, perpendicular to that within the bits.

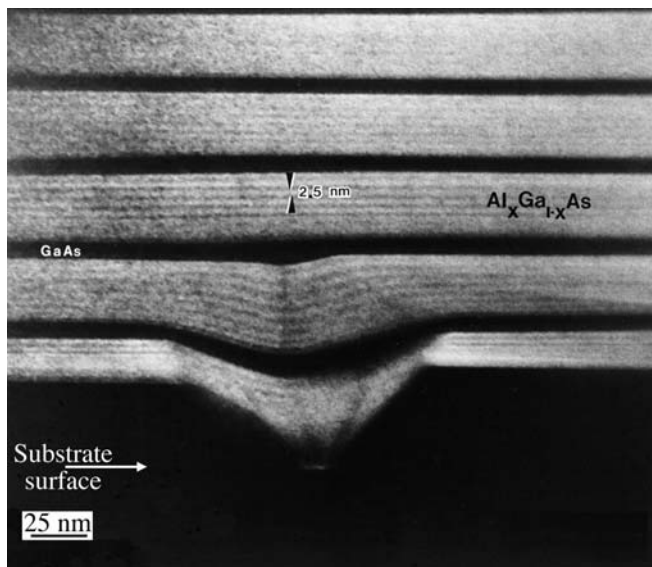


FIGURE 29.8. A chemically sensitive image of a GaAs-AlGaAs quantum-well structure. The composition of the AlGaAs is non-uniform because of growth fluctuations and the substrate surface was imperfectly covered by the first GaAs layer.



FIGURE 29.9. Diffuse-scatter DF image of a glassy material at a GB in a bicrystal of Al_2O_3 .

intensity maxima. Remember that we saw similar contrast effects, *but in the DPs*, of short-range ordered materials in Figure 17.11.

The practical suggestion in either of these cases for a modern TEM is to use a displaced aperture approach rather than translating the beam.

29.6 SURFACE IMAGING

We can obtain surface information in the TEM in a variety of ways. We can do reflection electron microscopy (REM), or use a technique called ‘topographic contrast.’ We can also form an SE image in an SEM or STEM, as we describe later in the chapter. We’ve already mentioned profile imaging by HRTEM in Section 28.16.

29.6.A Reflection Electron Microscopy

Reflection electron microscopy (REM) of surfaces requires that you mount your specimen in the holder

so the beam hits at a glancing angle, as shown in Figure 29.10A. Since the electrons are scattered from the surface, your specimen doesn’t need to be thinned. The image is foreshortened by an amount which depends on the reflection used. Different parts of the specimen will be focused at different positions behind the lens. Once you have chosen an image plane, you can move the specimen in the z -direction to focus different regions of the surface. A reflection high-energy electron diffraction (RHEED) pattern is generated by the surface layers of the specimen, as shown in Figure 29.10B. This diffraction geometry is exactly that which we used to derive Bragg’s law.

REM

Notice that the surface is actually tilted away from the optic axis so that the diffracted beam will be on the optic axis: The electron source can’t ‘see’ the diffracting surface!

Once you’ve formed the DP, the experimental procedure is essentially the same as for conventional diffraction-contrast DF imaging. You insert an objective aperture to select a Bragg-reflected beam and form an REM image, as in Figure 29.10C and D. The images are strongly foreshortened in the beam direction but maintain the usual TEM image resolution in the plane normal to the beam where the image is in focus. This can make interpretation difficult; you’ll see two different magnification markers in the two orthogonal directions. Note how different the two rotated images appear.

Ideally, you would like to have a very clean surface to simplify interpretation of the contrast. Of course, if the surface is not very flat, or is covered with a thick contamination or oxide layer, you won’t learn much from the image. However, with care, you can use REM to study the surface of many different materials. All that you’ve learned about diffraction-contrast images will apply to REM images. For example, you can also detect chemically sensitive contrast, as shown in Figure 29.10E. Here the dark bands are layers of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ in a GaAs matrix; the contrast is sensitive to the actual value of x . Notice that this contrast is the reverse of what we saw in TEM using chemically sensitive reflections.

Since you can do REM using any regular TEM holder, you can, in principle, easily heat or cool your bulk sample. In many ways, this is easier than using a transmission specimen since the sample is much more robust. In situ REM studies of Si provided a leap in our understanding of the reconstruction of Si surfaces. Yagi et al. showed the surface reconstruction taking place in situ. This technique has many potential applications but requires practice.

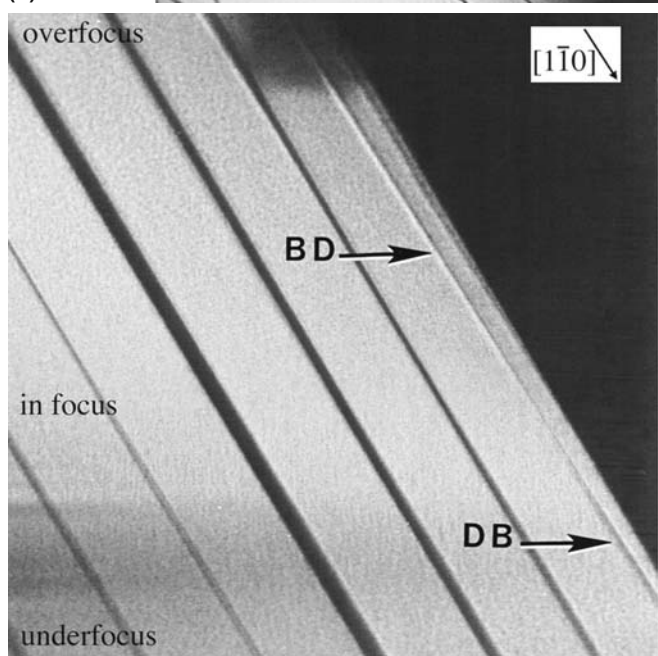
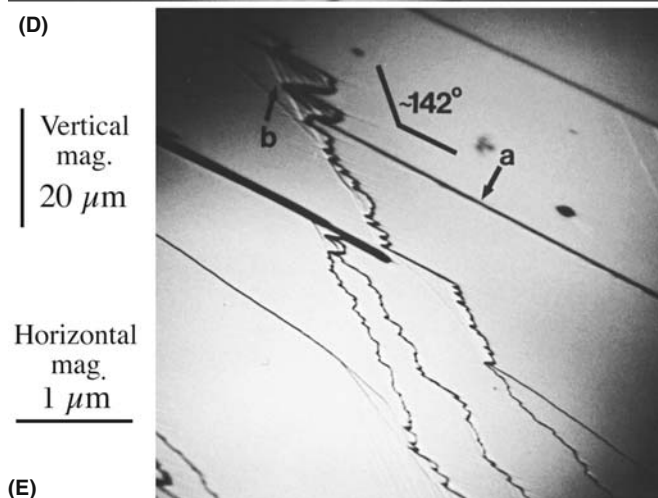
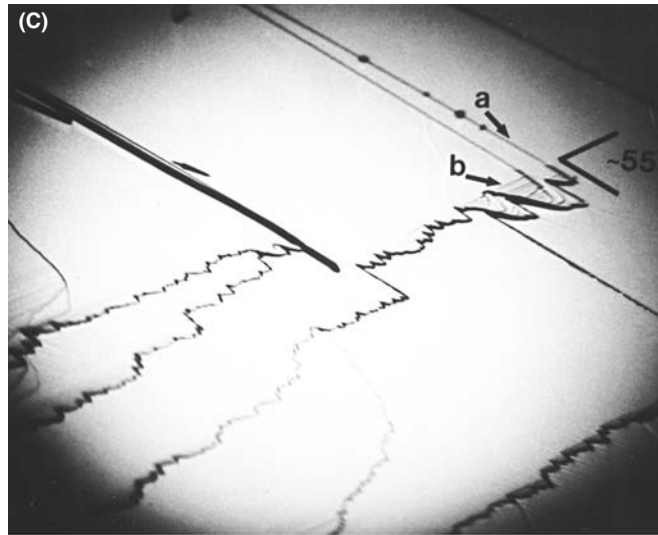
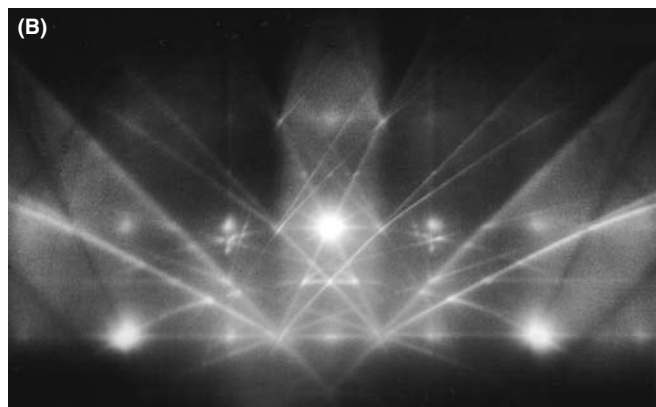
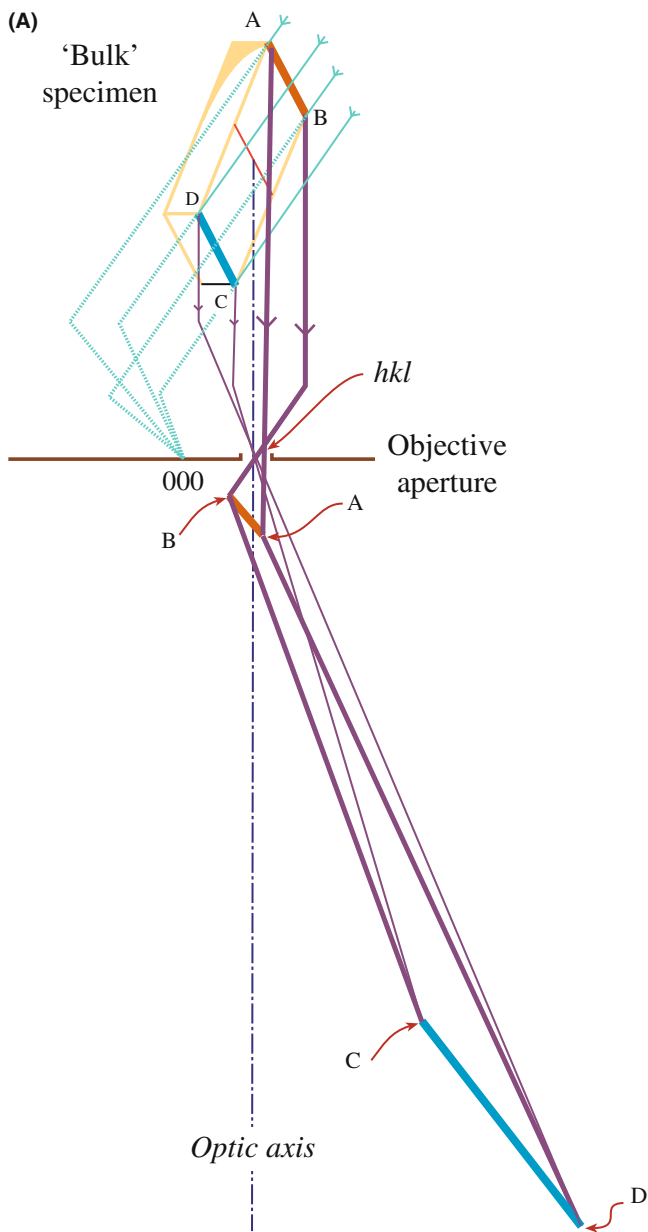


FIGURE 29.10. (A) Schematic diagram of the formation of an image using REM. (B) A 100 REM DP from a Si 001 surface. You can see the square array of spots but the 000 reflection has been blocked by the specimen itself. (C, D) REM images showing surface steps on a cleaved single crystal of GaAs. The two images are rotated with respect to one another and show the effect of the foreshortening. This emphasizes the care needed when interpreting foreshortened REM images. (E) REM image showing chemically sensitive contrast from quantum wells in GaAs/AlGaAs.

29.6.B Topographic Contrast

You can get a sense for the topography of your specimen by a neat technique which simply involves displacing the objective aperture until its shadow is visible across the region of the image that you're looking at. Around this area of the specimen you will see contrast which arises from an electron refraction effect but can be simply interpreted in terms of thickness changes in the specimen. In Figure 29.11, you can easily see that the Fe_3O_4 particles are supported on top of the carbon film and you can also see that the carbon film is not flat. Although the displacement of the aperture introduces astigmatism into the image, it does not limit resolution at the relatively low magnification at which this image was taken. To carry out an equivalent operation in STEM, you just displace the BF detector or shift the 000 diffraction disk so it falls half on and half off the detector. Again, the technique is not widely used, partly because refraction is a difficult topic in TEM and partly because the images don't look so 'nice.'



FIGURE 29.11. Topographic contrast from Fe_3O_4 particles on a carbon film. Ripples in the film are clearly visible. The shadow of the displaced objective diaphragm crosses the field of view.

29.7 HIGH-ORDER BF IMAGING

As we've taken some pains to point out, you get the strongest diffraction contrast in your images when s is small and positive. Under these circumstances, it's

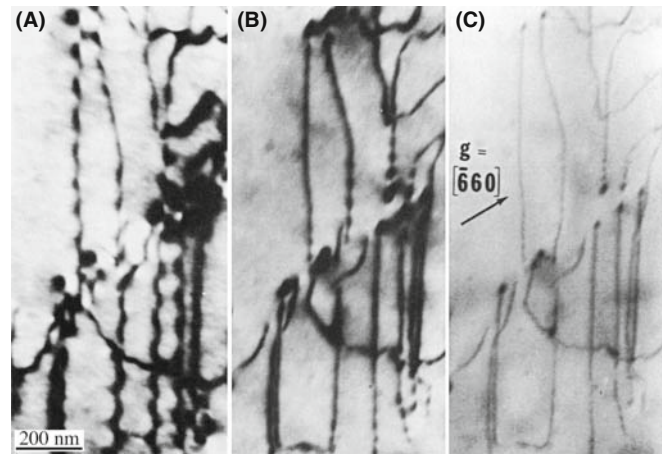


FIGURE 29.12. BF images of dislocations in GaAs taken with (A) g , (B) $2g$, and (C) $3g$. In the higher-order BF images the width of the defects decreases.

easiest to see crystal defects and interpret their contrast. This strong dynamical contrast can mask the details of the defect and sometimes it is more important to lower the contrast. You can do this by operating under kinematical conditions. Rather than having one strong low-index spot in the DP, if you tilt to make a high-order spot bright then the overall contrast is less and the background is not dominated by bend contours or thickness fringes, and the effects are shown in Figure 29.12. You can get a similar effect called 'kinematical diffraction imaging' if you form a BF image when no hkl spot is strong.

We get a similar effect in STEM imaging because, as we saw back in Section 22.6, dynamical-contrast effects are reduced when the reciprocity conditions between TEM and STEM are not fulfilled and all STEM images are more kinematical in nature.

Notice that Figure 29.12 was recorded at 650 kV. This tells you that high-order BF imaging can be a very useful technique when you're using an IVEM. You'll see there are subtle contrast effects that can occur at higher voltages. For example, in addition to improving the spatial localization of defect images, there are times when the $2g$ reflection gives better contrast.

PLAN YOUR EXPERIMENT

The point is, as usual, be prepared to experiment when you are sitting at the microscope.

Many of the early texts assume that you will be working at 100 kV, but only because that was all that was routinely available when they were written. The world is now different.

29.8 SECONDARY-ELECTRON IMAGING

SE images reveal the surface topography, which isn't much if you've polished your sample well, but is very important if you're looking at particulate specimens such as catalysts. You have to use a STEM if you're going to form SE images. If you look back to Figure 7.2, you'll see how the SE signal is detected by a scintillator-PM detector situated in the upper objective polepiece of a STEM. SEs generated in the top few nanometers of the specimen surface are confined by the strong magnetic field of the upper polepiece and spiral upward until they see the high voltage (~ 10 kV) on the aluminized surface of the scintillator. This design differs from the conventional SEM, in which the SE detector is situated under the final polepiece. In the STEM we get SE images of superior resolution and quality, which is why this design has been adopted for HRSEM. There are several reasons for this improvement in resolution

- A conventional SE image in an SEM invariably has noise contributions from BSEs which can enter the scintillator directly. In the STEM there is no line of sight to the SE detector for the BSE, and so the SE signal in STEM lacks the BSE noise that exists in an SEM image.
- The brightness of a thermionic source in a STEM is higher than in an SEM, because of the higher kV, and so the SE signal will be correspondingly stronger. An FEG will be brighter still.
- The C_s of STEM objective lenses is usually a lot smaller than for conventional SEMs. Therefore, SE images in a STEM are invariably of better quality than in an SEM because these two factors increase the S/N ratio.

The fourth advantage depends on a knowledge of the different types of SE which we'll now discuss.

The presence of remote SE signals in an SEM also decreases the S/N ratio. As shown in Figure 29.13, the SE detector in an SEM can pick up four different types of secondary electrons labeled SE I–SE IV generated by the primary electron (PE) beam

- The SE I signal is the only signal we want, since it emanates from the region around the probe and contains high-resolution topographic information from that region only.
- The SE II signal comprises SEs generated by BSEs that emerge from the specimen some distance away from the beam. The only way to reduce SE II is to reduce the BSE fraction and we can do this by using thin specimens, as in STEM.
- The SE III component arises from parts of the microscope stage that are struck by BSE from the specimen. So, reducing the BSE fraction by using thin specimens also reduces SE III.

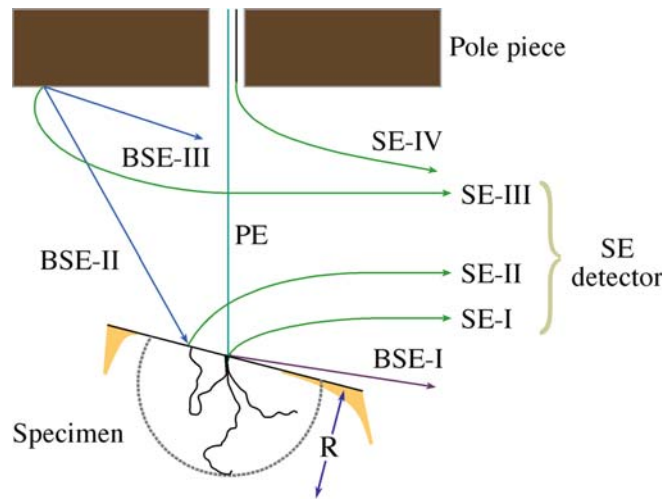


FIGURE 29.13. The four possible sources of SEs that enter the detector in a conventional SEM. The SE I signal is the only desirable one since it comes from the probe region, but SE II from BSE electrons, SE III from the stage of the microscope, and SE IV from the final aperture all combine to reduce the S/N ratio in an SE image. In a STEM, the relative contributions are different.

- The SE IV signal comes from SEs generated at the edges of the final probe-forming aperture, which in an SEM, often sits in the final lens. In a STEM, however, the C2 aperture performs this function and it is well away from the stage and the SE detector.

So, in a STEM with a thin specimen, the ratio of desirable SE I to undesirable SE II–IV is higher than in a conventional SEM. If you look at a bulk sample in the STEM, then the SE II signal will be the same as in an SEM since the BSE yield does not vary with beam energy. Furthermore, the SE III signal may be slightly worse than in an SEM because of the smaller STEM stage. Experimentally, the net result of all these differences is that the STEM still provides higher-resolution SE images of the same bulk sample than a conventional SEM, as shown in Figure 29.14. Now, the highest resolution FEG SEMs incorporate many of the design aspects of the STEM stage and can produce SE image resolutions < 1 nm at 30 kV; some SEMs even encourage you to use thin specimens and give a simple STEM image.

A thermionic source STEM cannot better than SEM resolution; however, an FEG STEM at 100 kV or higher should, in theory, offer the best possible SE images and may even show atomic-level topographic information, especially with an aberration-corrected probe.

Despite these obvious advantages and the ready availability of STEMs with SE detectors, very few high-resolution studies of the surface topography of specimens have been carried out with a STEM. It appears that one of the major uses of the SE image is simply to find the hole in the TEM disk! Why this is so remains unclear, but we strongly encourage you to use

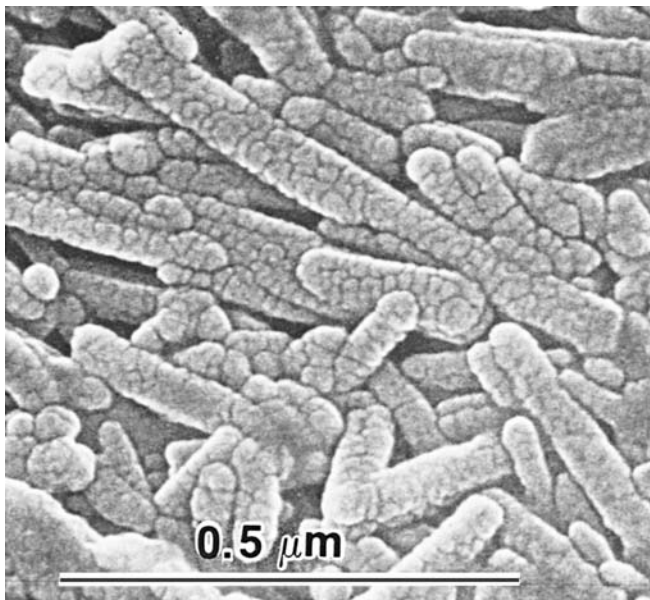


FIGURE 29.14. High-resolution SE image of coated magnetic tape in a TEM/STEM with an LaB₆ source at 100 kV showing ~2 nm spatial resolution.

SE images if they are available in your STEM. Of course, your specimen surface must be prepared carefully and kept clean, otherwise you'll image contamination, oxide, or some other artifact of your preparation process. If you have to coat your specimen because it's an insulator, then you should use a modern high-resolution, high-vacuum coater to generate a continuous thin film of a refractory metal such as Cr. Don't use the more conventional Au-Pd coatings, which merely mask the fine detail of the specimen surface.

29.9 BACKSCATTERED-ELECTRON IMAGING

Remember that the BSE detector in a STEM is situated directly under the upper objective polepiece. In an SEM, the detector is in the same position and we get equal collection efficiency. Higher kV gives a brighter source, but specimen thickness dominates the yield and it is abysmal in thin specimens (0.4% in 100 nm of Au). (Look at Figures 2.4, 36.1, and 36.2 and check the number of BSE in the Monte-Carlo simulations.) So there is a very low total signal and the *S/N* is poor. Nevertheless, in a thin specimen we can still form BSE images in the STEM that show resolution approaching the probe size, but only if a high-contrast specimen is used, e.g., Au islands on a C film, as shown in Figure 29.15. So in general, the BSE signal in a STEM offers no obvious advantages over a modern SEM with a high-efficiency scintillator or semiconductor detector. In fact,

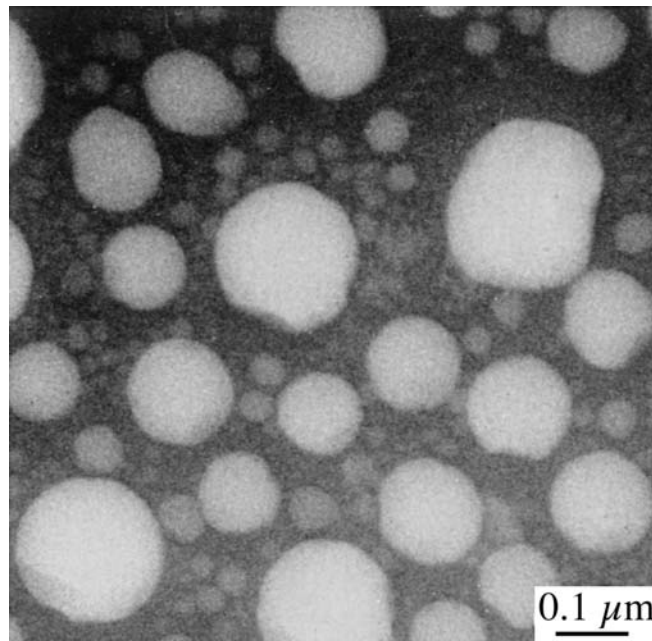


FIGURE 29.15. High-resolution STEM BSE image of Au islands on a carbon support film obtained in a TEM/STEM at 100 kV with an LaB₆ source. Resolution of ~7 nm is obtained, although the image quality is not very good because of the poor signal level.

the manufacturers of dedicated STEMs do not even offer a BSE detector as an option, and it is doubtful if it is worth installing one on a TEM/STEM.

For both SE and BSE imaging in STEM, you can look at both electron transparent or bulk samples of the sort normally examined in a conventional SEM. A major drawback to the STEM compared with the SEM is the relative volume of the microscope stage. In an SEM stage you can insert samples up to several centimeters in diameter and several centimeters thick, which means that sample preparation from the original object is often minimal. The confined-stage region of a STEM means that, even with a specially designed holder for bulk samples, the largest specimen will be about 10 mm × 5 mm and not more than a couple of millimeters deep.

29.10 CHARGE-COLLECTION MICROSCOPY AND CATHODOLUMINESCENCE

Charge-collection microscopy (CCM), otherwise known as electron-beam induced conductivity (EBIC), and the related phenomenon of cathodoluminescence (CL) are common techniques for the characterization of semiconductors in the SEM. It is possible to do the same things in a STEM, but few have been brave enough to try. You learned back in Section 7.1 about semiconductor electron detectors in which the incident beam generates electron-hole pairs which are swept apart by the internal field of the p-n junction and not allowed to

recombine. So if your *specimen* is a semiconductor, electron-hole pairs will be formed during the normal imaging process. You have to separate out the pairs by applying an external voltage through ohmic contacts evaporated on the surface of your foil. Then you can use the electron charge pulse to generate a signal on the STEM screen. The signal is strong wherever the pairs are separated, and weak at recombination centers such as dislocations and stacking faults. You can also measure the minority carrier diffusion length. Now in a STEM you can easily see the recombination centers by standard imaging techniques, so CCM is not really any great advantage, but it is essential in an SEM because the defects are subsurface.

If we don't separate the electron-hole pairs, they recombine and give off visible light. This light is an extremely weak signal, but it can be detected and dispersed by mirrors and spectrometers shown back in Figure 7.6. Again, recombination centers appear dark in CL images because most of the recombination is deep-level and non-radiative. The advantage of CL over CCM is that you don't have to coat your specimen to produce ohmic contacts, and you get a spectrum of light which contains information about doping levels and band-gap changes. However, you have to dedicate your STEM primarily to this imaging mode and it is a difficult and tedious technique. Early work in this area was pioneered by Petroff et al. and more recently by Batstone. It is not obvious, however, that doing the work in transmission offers major advantages over studies of bulk samples in the SEM (see the textbooks by Yacobi and Holt and by Newbury et al.) and this is reflected in the limited TEM applications of CL in the literature.

29.11 ELECTRON HOLOGRAPHY

Although the technique of electron holography became widely known in the early 1990s, Gabor had originally proposed the technique in 1948 as a way to improve the resolution of the TEM. The delay in its wide implementation was due to the lack of affordable, reliable FEG TEMs. An FEG is required so that the source will be sufficiently coherent; the FEG is then the electron equivalent of the laser. The topic is broader than you might have guessed, unless you've read the 1992 article by John Cowley entitled 'Twenty forms of electron holography' and the text by Völkl et al. We'll discuss the general principles of the technique and refer you to the companion text and the references for the details and more ideas.

The key feature is that, unlike conventional TEM imaging techniques, both the amplitude and the phase of the beam can be recorded. We can use this feature in two ways

- The effects of C_s can be partially corrected; thus we can improve the resolution of the TEM.
- We can examine other phase-dependent phenomena, such as those associated with magnetism.

Several different forms of holography are possible; the three best known are all based on the approach described in Chapter 23

- In-line holography.
- Single-sideband holography.
- Off-axis holography.

The approach which is mainly used in the TEM is the off-axis variation, as we'll discuss later. Figure 29.16 shows a hologram of a wedge-shaped crystal of Si oriented close to the [110] pole. The boxed

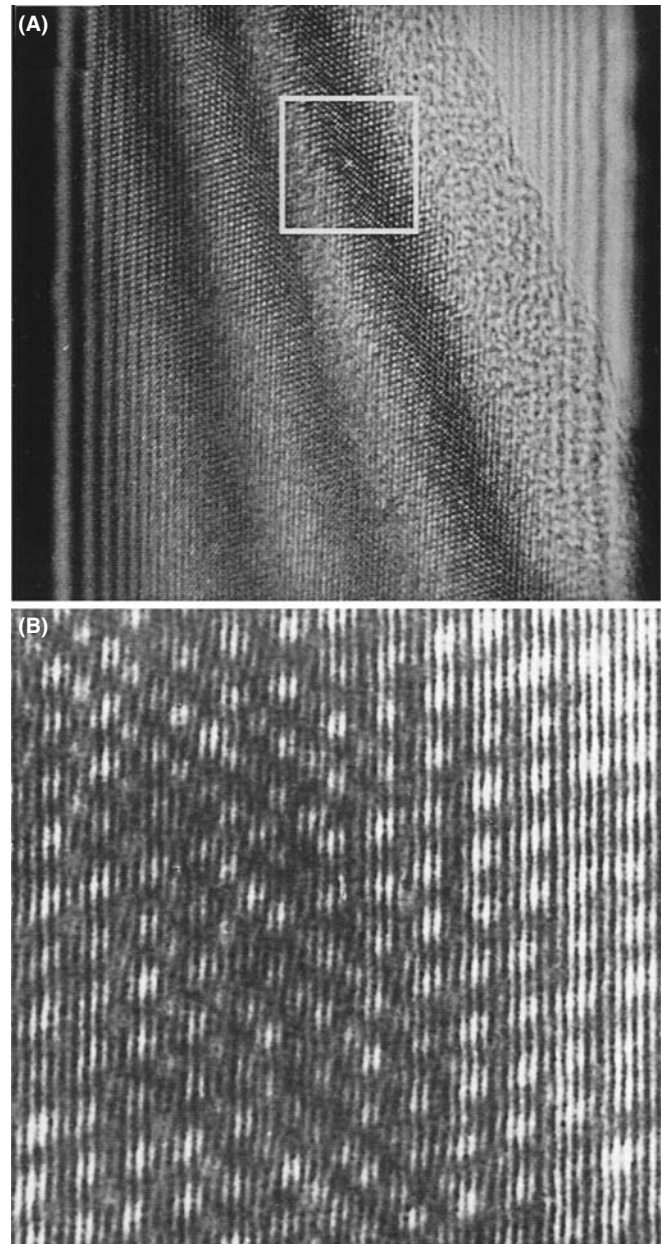


FIGURE 29.16. (A) Hologram of a [110]-oriented Si specimen; the broad bands are thickness fringes where the specimen is thickest at the lower right corner. (B) An enlargement of the boxed region in (A) showing fringes with a 70 pm spacing.

region of the figure shows a set of fringes which are only 70 pm apart. These are not lattice fringes but do contain phase information relating to the diffracted beams. In order to analyze such images, you'll need the appropriate software although much of the original research used optical processing techniques; this is such a specialized branch of image processing and image reconstruction that we will not cover it in Chapter 31, but it is discussed in the holography chapter in the companion text.

The principle of holography is shown by the schematic in Figure 29.17. Here $\chi(\mathbf{u})$ is the function we used in Chapter 28 to describe the effect of the objective-lens aberrations and defocus. In conventional TEM, we choose conditions so that the specimen acts as a pure phase object and the imaginary part of $e^{i\chi(\mathbf{u})}$, i.e., the sine term, converts this phase information into an amplitude which we record as the image. With holography, we can use the real part of the exponential too. A nice way to think of the process is that, for a real specimen, $\chi(\mathbf{u})$ mixes the amplitude, A , and phase, ω , from the specimen to give amplitude, A , and phase, Ω , in the image. In conventional imaging we record A^2 and lose all the information on Ω . With holography, we don't lose any information but we have to work hard to recover it. The original proposal by Gabor, to use holography to improve the resolution of the TEM, is being actively pursued. Although holography has still not reached its full potential in this field, you can see the future by looking again at Figure 29.16.

REFERENCE BEAM

The essential feature of the experiment is that a reference beam from the same coherent source passes outside your specimen and is then deflected by the biprism so that it interferes with the beam that passed through the specimen.

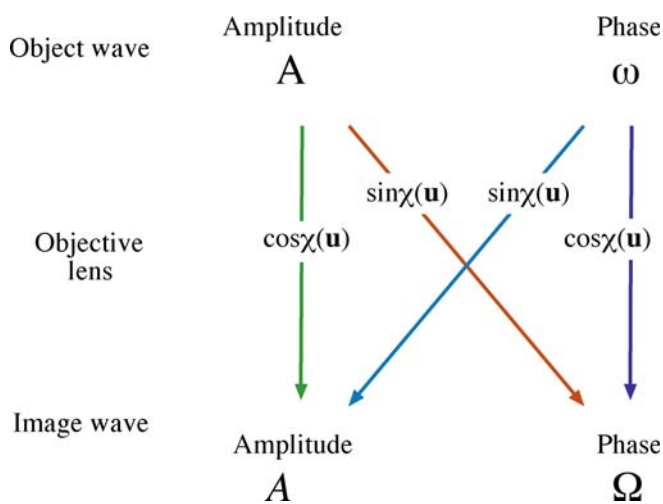


FIGURE 29.17. A schematic of how the objective lens mixes the amplitude and phase components of the specimen to form the amplitude and phase components of the image.

In practice, electron holography is carried out using an FEG TEM which has been fitted with a beam splitter; not every FEG TEM is so equipped. The beam splitter is made by coating a thin glass fiber with metal to prevent it charging, and assembled to give the biprism; i.e., it's the biprism we discussed in Section 23.7 and can be $<0.5 \mu\text{m}$ in diameter. Part of the beam passes through the specimen while the other part forms the reference beam, as shown in Figure 29.18. You must be able to rotate either the biprism or the specimen. For holography, the biprism is placed below the objective lens, e.g., at the position usually occupied by the SAD aperture rod.

HOLOGRAM

A 3D image of an object; the image is made by holography.

HOLOGRAPHY

A method of creating a 3D image of an object. The process requires recording both the intensity and the phase information of the waves emitted or scattered by the object.

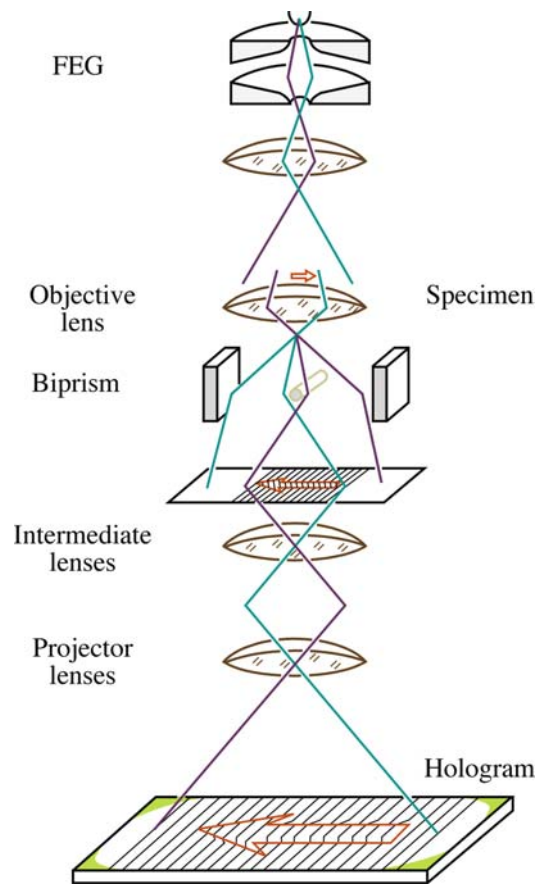


FIGURE 29.18. The operation of the biprism in electron holography.

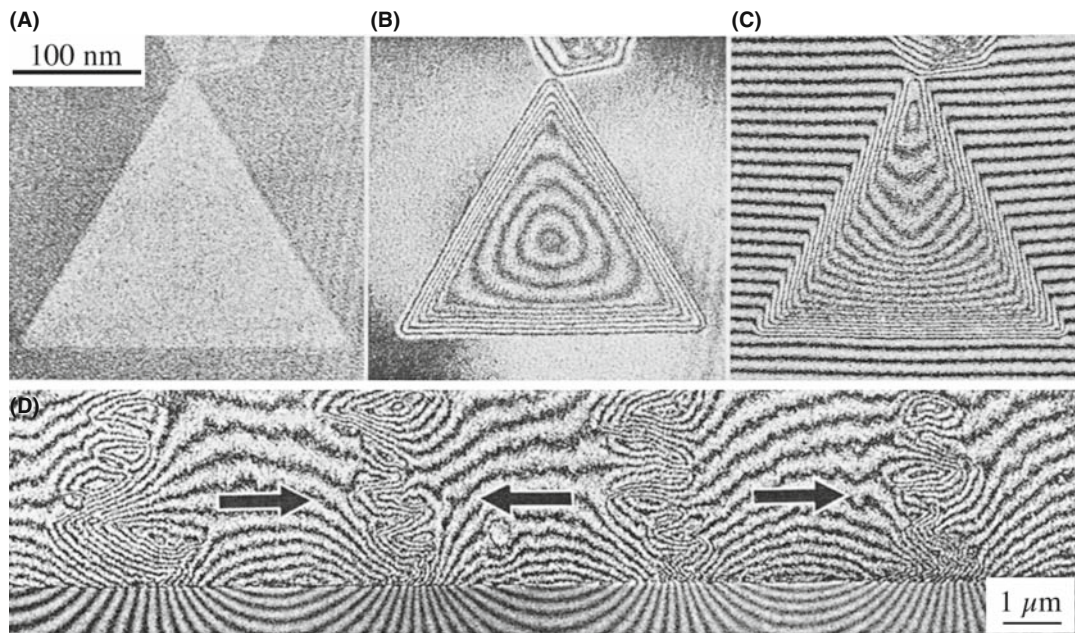


FIGURE 29.19. Holographic images of a magnetic Co particle showing (A) the reconstructed image, (B) the magnetic lines of force, and (C) the interferogram. (D) Line of force in a magnetic medium.

All that you then have to do is interpret the interference pattern, i.e., the electron hologram, but as you'll appreciate after reading Chapter 31, this process is not trivial. However, the technique does provide the possibility for 'coherent processing' of the electron wave. In principle, all the required data can be recovered from one hologram but the technique is unlikely to replace conventional HRTEM in most applications.

We'll conclude this section by illustrating one of the unique applications of the technique, namely, the imaging of magnetization characteristics and flux lines which has been developed by Tonomura. Figure 29.19 shows a series of images illustrating how magnetization patterns can be imaged both for single particles and for an actual recording medium. Tonomura's classic studies of the quantization of flux lines in superconducting materials is illustrated in Figure 29.20A. As the ring is gradually cooled, the phase shift due to the magnetization changes until it becomes superconducting: the phase shift inside the ring then becomes exactly π . The other images in Figure 29.20B and C show flux lines constricting as they enter the superconductor. Notice that you can also clearly see these flux lines outside the specimen: the magnetic field outside the specimen influences the electron beam.

29.12 IN-SITU TEM: DYNAMIC EXPERIMENTS

We've mentioned several times that we can use the TEM for in-situ experiments. In Chapter 8, we described holders you can use to heat the specimen, e.g., causing phase transformations, or to strain it to change the defect

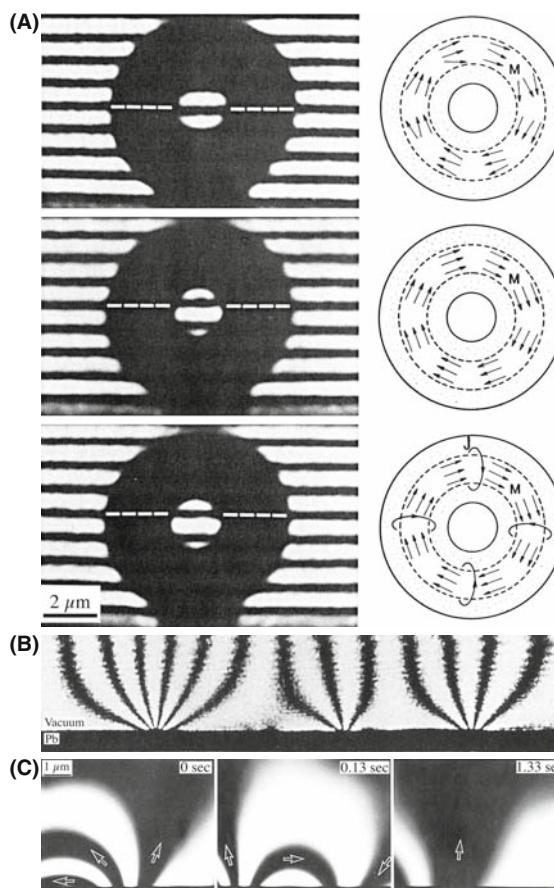


FIGURE 29.20. (A) Gradually cooling a ring (torus) of a superconducting material from 300 to 15 to 5 K to demonstrate the quantization of flux. (B) Magnetic flux lines penetrating a superconducting Pb film. (C) Interference of fluxons trapped in superconducting Pb as time increases from 0 to 0.13 to 1.33 seconds.

structure. In Section 7.3.C, we discussed CCD video cameras, which are the best way to record dynamic changes in the microstructure, and in this chapter, we've discussed imaging moving flux lines in magnetic materials. In situ experiments remove the doubts that exist whenever you observe materials after heat treatment or after deformation and then try to infer what actually happened at temperature or during deformation. It is implicitly assumed in most TEM investigations that cooling a heat-treated specimen to room temperature or removing the applied stress does not change the microstructure. Having made this assumption, we draw conclusions about what happened during our experiment. However, this assumption is clearly not valid for many situations. Nevertheless, we generally view our specimen at ambient temperatures and not under load.

In fact, there are good reasons why we rarely do in-situ experiments. The main reason is that such studies are difficult to perform on thin specimens. As we indicated in Section 26.8, when the surface properties dominate the bulk properties, as is often the case in thin specimens, TEM images and their analysis can be misleading. Surface diffusion is much more rapid than bulk diffusion and defects are subject to different stress states in thin foils where surface relaxation effect may dominate. The best

way to overcome this limitation is to use much thicker specimens and this requires higher voltages. In-situ experimentation was widely used in the 1960s and 1970s when 1–3 MV instruments were first constructed and used to look through foils which were $>1\ \mu\text{m}$ thick (see the book by Butler and Hale and Figures 26.14 and 26.15). So in-situ experiments can be expensive: currently there is no longer a TEM operating at $>400\ \text{kV}$ in the USA.

However, the advent of 300–400 kV IVEMs, the construction of megaVolt instruments in Stuttgart and Japan and the extra space (because of the larger pole-piece gap) that is potentially available in C_s -corrected TEMs have brought a resurgence of interest in using in situ experimentation. Developments in electron optics, stage design, and recording media since the 1960s mean that combined HRTEM and in-situ experimentation is now possible. This combination is a powerful tool, permitting the observation of reactions at the atomic level, such as the motion of individual ledges at interfaces, as in Figure 29.21, although lower-resolution images are no less impressive.

You must bear in mind that the experiments you perform *in situ* are taking place under conditions that still do not approach the bulk conditions experienced by many engineering materials. In particular, the high-keV

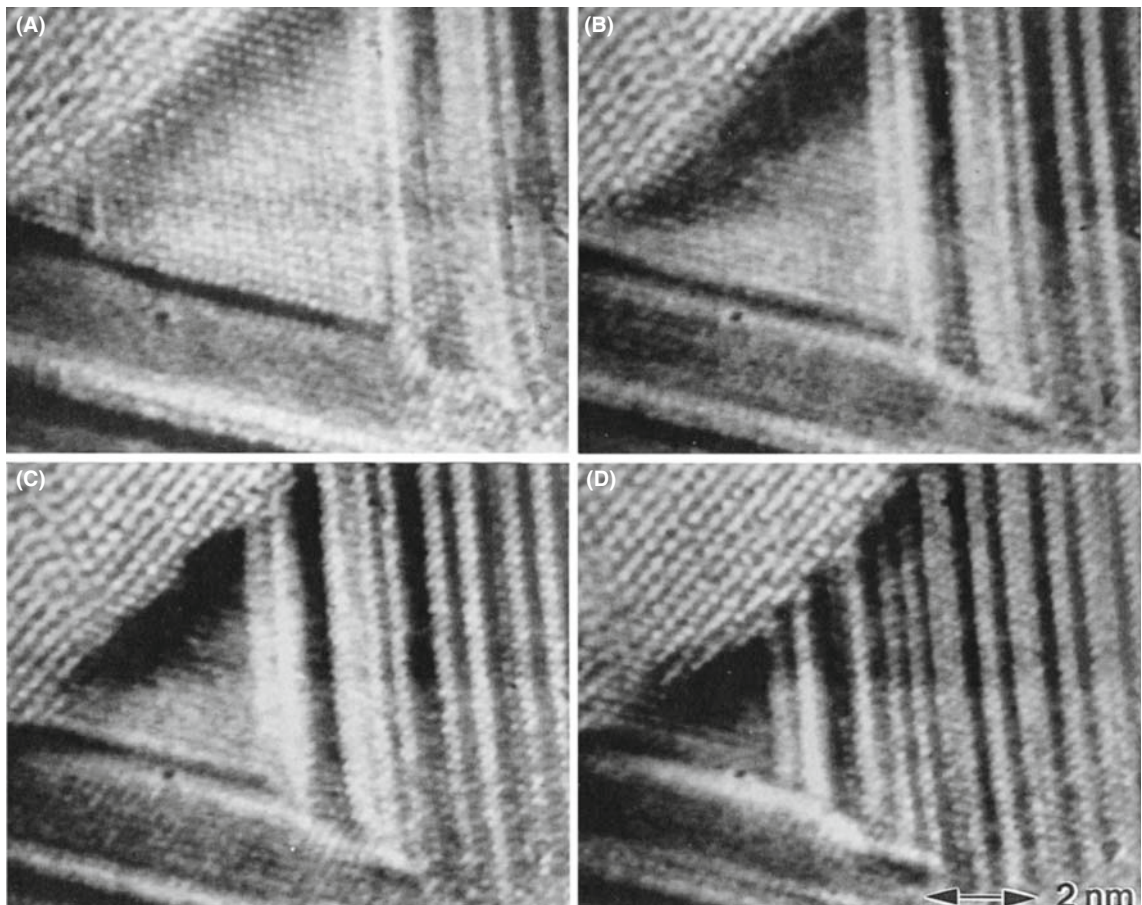


FIGURE 29.21. (A–D) Four images taken from a video showing the reaction front of a Ge/Ag/Ge trilayer while heating the specimen *in situ* at 250°C. The time interval between the images was 8 seconds. The Ge crystal grows from the upper left without moving the Ag lattice.

electron flux brings an uncertainty to all in-situ experiments, particularly if you are trying to infer kinetic data from measurements taken of reactions. So while in-situ experiments can give you powerful demonstrations of real-time changes in materials, be wary of direct interpretation. You have to ensure you are controlling all the variables, and you must cross-check kinetic data with calculations to verify that, e.g., diffusion fluxes are consistent with the known temperature and not influenced by surface or vacancy effects, having said that, it is nonetheless very useful to have access to an IVEM for this kind of experiment.

29.13 FLUCTUATION MICROSCOPY

Fluctuation electron microscopy (FEM) is a technique invented by Treacy and Gibson that has come into prominence in the last decade to study medium-range order ($\sim 0.5\text{--}2$ nm) in amorphous materials, such as glasses, amorphous silicon and carbon. We saw back in Section 18.7 how to use different kinds of diffraction and imaging to study these increasingly important materials and FEM is a hybrid imaging and diffraction technique. The ‘fluctuation’ refers to changes in the local structure and orientation of the glass where ‘local’ means not quite down at the atomic level.

You can do FEM experiments in two equivalent ways. Either you can form a hollow-cone DF image (go back and check Figure 18.14) or you can do STEM diffraction with a fine probe, scanning the pattern across the STEM detector. The normalized variance of the intensity distribution in the microdiffraction patterns gives equivalent information to the hollow-cone image. In both modes, you insert a small objective aperture which, in either mode, limits the resolution of the fluctuation image to $\sim 1\text{--}2$ nm. This resolution matches the dimensional scale of local structural fluctuations, which appear as speckles in the image (see Figure 29.22A and B). Depending on the degree of local coherence in the random network structure, the speckles (both bright and dark regions) vary in intensity. More random structures give gray regions; more ordered structures give locally brighter or darker regions depending on whether the local electron scatter is on axis or off axis (intercepted by the diaphragm), respectively (Figure 29.22C).

29.14 OTHER VARIATIONS POSSIBLE IN A STEM

We can form images with the characteristic X-rays and the energy-loss electrons that we detect with the appropriate spectrometers. Rather than discuss these here, we’ll save the topics for the chapters devoted to XEDS and EELS in Part 4.

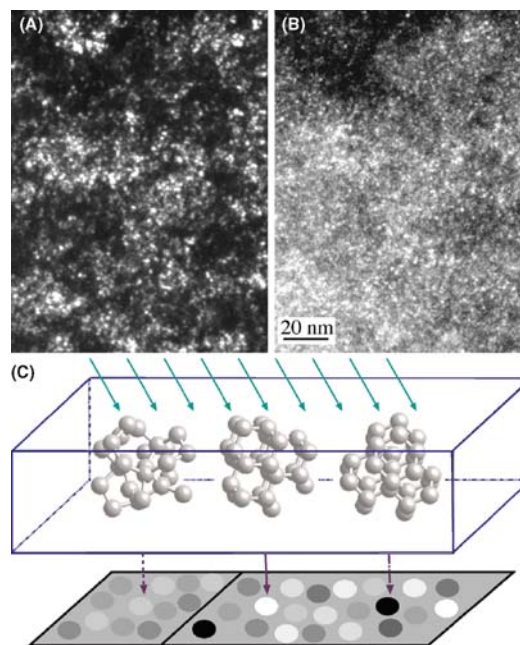


FIGURE 29.22. Two images of the same region of shungite. (A) The speckle corresponding to fluctuations on the graphite 002 reflection (3.0 nm^{-1}). (B) Ditto but for 7.1 nm^{-1} . Some of this speckle corresponds to fullerenes (and some to noise). The presence of fullerenes is inferred from the fact that there is a peak (such as that shown in the plot) due to the C–C bond distance, which is ordinarily forbidden. Structure noise from truly amorphous sample will not give a significant peak corresponding to ~ 1.5 nm. (C) Schematic diagram showing how the local coherence influences the intensity in the image.

In a STEM then, we can, in theory, detect and measure all of the signals generated in a thin specimen and shown way back in Figure 1.3. We can reproduce all the conventional TEM imaging methods and some unconventional ones too, such as Z-contrast and most of the specialized techniques in this chapter. Remember that for all the STEM signals we use a detector of one form or another. We can make detectors of different shapes and sizes, as we already described for looking at directional scattering from magnetic specimens. Remember also that the detectors permit us to digitize the signal so we can process it, manipulate it, and present it for viewing in ways that are impossible with analog TEM images. Most STEM systems come with relatively basic image processing, such as black level, gain, contrast, brightness, gamma, Y modulation, and signal addition and subtraction. The computer system for X-ray and EELS analysis often has a lot of image analysis software thrown in. So there’s a lot you can do with the STEM images that really fall into the category of digital image processing which you may study in Russ’s specialist texts, after you’ve read Chapter 31.

CHAPTER SUMMARY

We've described only a few of the many ways that we can manipulate the electron beam to produce different images and different contrast phenomena, and it is more than likely that there are still unknown methods awaiting discovery. For example, the introduction of a continuously variable detector collection angle in STEM is equivalent to using a continuously variable iris-type objective aperture in TEM and offers new imaging possibilities. Similarly, there is much to be learned from mutually complementary techniques in SEM and TEM.

We scarcely mentioned the use of annular apertures, conical illumination, pre- and post-specimen scanning, or rocking. All of these are possible, so you should be prepared to try new things in the microscope and see what the effects are. Many advances of the sort we have described only came about by accident, but the microscopist was wise enough to see the effect and try and understand it, rather than dismiss it as unimportant.

REFERENCES PHILOSOPHY

Dennis Gabor (Gábor Dénes; 1900–1979) worked in Rugby, England with the British Thomson-Houston company. This company was already merged with the Metropolitan-Vickers company and in 1960 became part of the holding company, Associated Electrical Industries (AEI). The early TEMs, including the HVEMs, in England were manufactured by AEI. He was awarded the Nobel Prize for physics in 1971 for discovering holography.

REM

De Cooman, BC, Kuesters, K-H and Carter, CB 1984 *Reflection Electron Microscopy of Epilayers Grown by Molecular Beam Epitaxy* Phil. Mag. A **50** 849–56. Using chemically sensitive reflections.
Hsu, T, Ed. 1992 *J. Electron Microsc. Tech.* **20**, part 4. A special issue using REM for imaging surfaces.
Wang, ZL 1996 *Reflection Electron Microscopy and Spectroscopy for Surface Analysis* Cambridge University Press New York. The latest book—reflecting that not much has happened recently.
Yagi, K, Ogawa, S and Tanishiro, Y 1987 *Reflection High-Energy Electron Diffraction and Reflection Electron Imaging of Surfaces* (Eds. PK Larsen and PJ Dobson), p. 285, Plenum Press, New York.

HOLOGRAPHY

Electron holography: The following key articles will give you a sense of the possible applications but you also have the monograph by Tonomura. An electron holography plug-in is available for digital micrograph.

Gabor, D 1949 *Microscopy by Reconstructed Wave-Fronts: I* Proc. Roy. Soc. London **197A**, 454–487; (1951) *Microscopy by Reconstructed Wave Fronts: II* Proc. Phys. Soc. **B64**, 449–69. Not easy to find.
Harscher, A, Lang, G and Lichte, H 1995 *Interpretable Resolution of 0.2 nm at 100 kV Using Electron Holography* Ultramicroscopy **58**, 79. Pursuing higher resolution in TEM using a FEG.
Lehmann, M and Lichte, H 2005 *Electron Holographic Material Analysis at Atomic Dimensions* Crystal Research and Technology **40** 149–160.
Tonomura, A 1987 *Application of Electron Holography* Rev. Mod. Phys. **59**, 639–669; 1992 *Electron-Holographic Interference Microscopy* Adv. Phys. **41**, 59–103. Overview of his work on imaging flux lines.
Tonomura, A 1999 *Electron Holography* 2nd Ed Springer Heidelberg. A focused text on the subject.
Völkl, E, Allard, LF, Joy, DC Eds. 1999 *Introduction to Electron Holography* Springer NY. A comprehensive text on the subject

MAGNETIC MATERIALS

Chapman, JN, Johnston, AB, Heyderman, LJ, McVitie, S and Nicholson, WAP 1994 *Coherent Magnetic Imaging by TEM* IEEE Trans. Magn. **30** 4479–4484. Qualitative measurement of magnetic induction.
Petford-Long, AK and Chapman JM 2005 *Lorentz Microscopy in Magnetic Microscopy of Nanostructures* (Eds. H Hopster and HP Oepen) pp. 67–86 Springer New York. A recent review in a broad-ranging review text.

CL ETC.

Newbury, DE, Joy, DC, Echlin, P, Fiori, CE and Goldstein, JI 1986 *Advanced Scanning Electron Microscopy and X-ray Microanalysis*, p. 45 Plenum Press New York. Review of CL imaging. Compare to Section 29.10.
Yacobi, G and Holt, DB 1990 *Cathodoluminescence Microscopy of Inorganic Solids* Plenum Press New York.

SEM IN THE TEM

Goldstein, JI, Newbury, DE, Echlin, P, Joy, DC, Lyman, CE, Lifshin, E, Sawyer, LC and Michael, JR 2003 *Scanning Electron Microscopy and X-ray Microanalysis*, 3rd Ed., Springer New York. Whenever you think of SEM, this is *the* text.

IN SITU TEM

Butler P, Hale KF 1981 *In Situ Studies of Gas–Solid Reactions in Practical Methods in Electron Microscopy*, Ed. AM Glauert 9 North-Holland New York. A collection of impressive images and techniques.

Gai, PL Ed. 1997 *In-Situ Microscopy in Materials Research* Springer New York. Comprehensive set of review articles.

Rühle, M, Philipp, F, Seeger, A and Heydenreich, J, Eds. 1994 *Ultramicroscopy* 56 (1-3). An issue on in-situ TEM.

Sinclair, R and Konno, TJ 1994 *In-Situ HREM: Application to Metal-Mediated Crystallization* Ultramicroscopy 56 225–232. HRTEM in situ studies.

Wang, ZL, Poncharal, P and de Heer, WA 2000 *Measuring Physical and Mechanical Properties of Individual Carbon Nanotubes by In-Situ TEM* J. Phys. Chem. Sol. 61 1025–1030. The combination of in situ and nano.

Special collections of papers:

J. Mater. Res. 20 2005 *In-Situ TEM* Eds I Robertson, J Yang, R Hull, M Kirk and U Messerschmidt.

J. Mater. Sci. 41 2006 *Characterization of Real Materials and Real Processing by Transmission Electron Microscopy* Ed. H Saka.

FLUCTUATION MICROSCOPY

Treacy, MMJ and Gibson, JM 1996 *Variable Coherence Microscopy: a Rich Source of Structural Information from Disordered Systems* Acta Cryst A52 212–220. The first description of fluctuation microscopy.

Treacy, MMJ, Gibson, JM, Fan, L, Paterson, DJ and McNulty, I 2005 *Fluctuation Microscopy: a Probe of Medium Range Order* Rep. Prog. Phys. 68 2899–2944. The latest in-depth review.

COMBINED TECHNIQUES

Williams, DB and Newbury, DE 1984 *Recent Advances in the Electron Microscopy of Materials in Advances in Electronics and Electron Physics* (Ed. PW Hawkes) 62 161–288 Academic Press New York. Lots of techniques, some of which are still relevant.

GENERAL

Bell, WL 1976 *2½D Electron Microscopy: Through-Focus Dark-Field Image Shifts* J. Appl. Phys. 47 1676–82. The paper on this topic.

Bell, WL and Thomas, G 1972 *Applications and Recent Developments in Transmission Electron Microscopy* Electron Microscopy and Structure of Materials (Ed. G Thomas), p. 23–59 University of California Press Berkeley CA. High-order BF imaging; not widely used but interesting.

Clarke, DR 1979 *On the Detection of Thin Intergranular Films by Electron Microscopy* Ultramicroscopy 4 33–44. The original paper on using diffuse DF to study intergranular films.

Cowley, JM 1973a *High-Resolution Dark-Field Electron Microscopy. I. Useful Approximations* Acta Cryst. A29 529–536.

Cowley, JM 1973b *High-Resolution Dark-Field Electron Microscopy. II. Short-Range Order in Crystals* Acta Cryst. A29 537–540.

Cowley, JM 1992 *Twenty Forms of Electron Holography* Ultramicroscopy 41 33–348.

Franke, F-J, Hermann, K-H and Lichte, H 1988 in *Image and Signal Processing for Electron Microscopy* (Eds. PW Hawkes, FP Ottensmeyer, A Rosenfeld and WO Saxton), p. 59 Scanning Microscopy Supplement 2 AMF O'Hare IL.

Hudson, B 1973 *Application of Stereo-Techniques to Electron Micrograph* J. Microsc. 98 396–401. A review of stereomicroscopy in the TEM.

Imeson D 1987 *Studies of Supported Metal Catalysts Using High Resolution Secondary Electron Imaging in a STEM* J. Microsc. 147 65–74. More on SE imaging in the STEM.

Joy, DC, Maher, DM and Cullis, AG 1976 *The Nature of Defocus Fringes in Scanning-Transmission Electron Microscope Images* J. Microsc. 108 185–193. Topographic contrast in the TEM (Section 29.6.B).

Russ, J 1990 *Computer-Assisted Microscopy: The Analysis and Measurement of Images* Plenum Press New York.

Simpson, YK, Carter, CB, Morrissey, KJ, Angelini, P and Bentley, J 1986 *The Identification of Thin Amorphous Films at Grain-Boundaries in Al₂O₃* J. Mater. Sci. 21 2689–96. An early paper comparing different techniques for studying intergranular films.

Spence, JCH 1999 *The Future of Atomic-Resolution Electron Microscopy for Materials Science* Mat. Sci. Eng. R26 1–49.

COMPANION TEXT

Four complete chapters in the companion text are devoted to topics described in this chapter: electron holography (Lichte and Lehmann), magnetic specimens (Petford-Long and de Graef), and in situ TEM (CBC). Chemically sensitive imaging is becoming much more commonplace now but using more AEM combined with imaging. A fourth chapter is dedicated to tomography.

SELF-ASSESSMENT QUESTIONS

- Q29.1 Make a list of all the imaging techniques in this chapter and determine the electron-scattering mechanism(s) operating in each case.
- Q29.2 Make a list of all the imaging techniques in this chapter and determine the contrast mechanism(s) operating in each case.
- Q29.3 Make a list of all the imaging techniques in this chapter and determine the TEM operating mode(s) operating in each case.
- Q29.4 List the pros and cons of stereo versus tomographic approaches to the TEM image-projection problem. Why is stereo more difficult for diffraction-contrast experiments than mass-thickness contrast? Why is a phase-contrast, atomic-resolution stereo experiment effectively impossible?
- Q29.5 Why is SE imaging much more useful than BSE imaging in STEM while both are equally useful in SEM?
- Q29.6 Search the Web and determine how many different kinds of experiments can be performed *in situ* in a TEM.
- Q29.7 List the pros and cons of a typical in-situ experiment.
- Q29.8 Why is $2\frac{1}{2}$ D imaging not called 2D?
- Q29.9 If your specimen is magnetic, how does this affect your choice of preparation method?
- Q29.10 If your specimen is magnetic, how does this affect your choice of TEM versus STEM mode?
- Q29.11 If your specimen is magnetic, how does this affect your choice of DF imaging method?
- Q29.12 Magnetism is becoming increasingly important in nanotechnology. Why is nanomagnetism easier to study in a (S)TEM than a conventional ferromagnetic steel specimen?
- Q29.13 Who were Lorentz, Fresnel, and Foucault (yes; go and study some history of science) and why are their names given to specific forms of TEM of magnetic specimens?
- Q29.14 Discuss the factors limiting the resolution in Foucault and Fresnel images. Why is this resolution important in nanomagnetism studies?
- Q29.15 Distinguish Bloch, Néel walls, and cross-tie walls and how best to image them in a TEM.
- Q29.16 Why would you use chemically sensitive imaging techniques when XEDS or EELS spectrum imaging (see Part 4) can reveal the distribution of every element in your specimen down to the atomic level?
- Q29.17 Find out every way possible to explore the structure and chemistry of glasses that is discussed in this book. List the pros and cons of each approach.
- Q29.18 Why is DADF preferable to CDF when forming images with diffusely scattered electrons?
- Q29.19 Why do REM in your TEM when you can do SEM in your STEM?
- Q29.20 How can we image atomic-level detail in REM images when the TEM is not being operated in phase-contrast HRTEM mode?
- Q29.21 Compare and contrast HOBf and HODf imaging. What is the more common name for HODf?

TEXT-SPECIFIC QUESTIONS

- T29.1 Why would you be surprised to find any examples in the literature of BSE imaging in a STEM?
- T29.2 If you want to watch nanotube electronic circuits operating *in situ* in a TEM what technique(s) would you use? Could you do this in a SEM?
- T29.3 Why would Denis Gabor be delighted if he were still alive? Why was he a 'microscopist before his time' and what experimental development helped most to realize his dream?
- T29.4 Explain how a biprism works and what you give up if you put one into your TEM column.
- T29.5 Suggest why fluctuation microscopy is not common in the Far East.
- T29.6 Why in fluctuation microscopy is the *variance* of the intensity distribution used to form images rather than the intensity distribution itself?
- T29.7 Why is fluctuation microscopy not simply described as a subset of diffuse-scatter imaging?
- T29.8 Search the Web to find at least one new TEM imaging or diffraction technique *not* described in this chapter or anywhere else in the text. What does your search tell you about TEM as an experimental materials-characterization tool?
- T29.9 Look at Figure 29.1 and try and cross your eyes until the left and right images overlap. You then may be able to see the stereo effect without using a stereo viewer. To see the effect, you may have to tilt your head to get exact image overlap (why?) and you will still see de-focused image of the two originals (why?).
- T29.10 Look at Figure 29.2. Use this to explain why $2\frac{1}{2}$ D microscopy effects can better be achieved with more recent developments in high spatial resolution diffraction.

- T29.11 List all the information about magnetic specimens that is discernible in Figures 29.3–29.7. What can you not find out about magnetism in TEM images?
- T29.12 Why do the dislocation images get (a) sharper and (b) fainter across the sequence shown in Figure 29.12?
- T29.13 Explain why you can see the structure inside the amorphous carbon support film in Figure 29.11.
- T29.14 What are the lines in Figure 29.10B? Explain how they arise.
- T29.15 Draw a similar diagram to Figure 29.13 for a thin specimen and explain how the different SE contributions vary between SEM and STEM SE images.
- T29.16 Find examples in the literature of SE imaging in a STEM for small particles on a substrate and explain the advantages over conventional TEM imaging.
- T29.17 Some scientists believe strongly that Tonomura’s work in Figure 29.20 is worthy of the Nobel Prize. Justify this claim.
- T29.18 Explain why Figure 29.21 is an extraordinary sequence of images.

Image Simulation

CHAPTER PREVIEW

When we need to obtain information about the specimen in two directions, we need to align the specimen close to a low-index zone axis. If the HRTEM image information is going to be directly interpretable, the specimen must be oriented with the incident beam exactly aligned with both the optic axis of the TEM and the zone axis of the specimen. Thus, we will have many reflections excited and the simple two-beam analysis of Chapter 27 cannot be used.

A method for modeling the contrast of images obtained under these conditions was developed by Cowley, Moodie, and their co-workers, principally at Melbourne and Arizona State University (ASU), in a series of classic papers beginning with that by Cowley and Moodie in 1957. Fortunately, the growing interest in HRTEM has coincided with the availability of increasingly powerful computers which can handle the extensive calculations.

There are several software packages available commercially, so there is little reason for most users to re-invent the wheel. However, the different packages do not necessarily perform the calculations in the same way: one may be more appropriate for your application than others. Since these packages essentially operate as ‘black boxes’ it is also reassuring to simulate images from the same structure using different packages (unless they don’t give the same answers).

One point to keep in mind as you work through the literature is that this subject already has a lot of history. We will point out some of the things that have been done, and in some cases continue to be done, for historical reasons.

30.1 SIMULATING IMAGES

The idea of simulating HRTEM images arose because of the realization that the loss of phase information when we form an experimental intensity map means that we can’t go back from the image to the structure. Instead, we assume a structure (perfect crystal or crystalline material containing defects), simulate the image, see how closely the simulated image resembles the experimental image, modify the structure, and repeat the process. The only difficulty is that the image is sensitive to several factors

- The precise alignment of the beam with respect to both the specimen and the optic axis.
- The thickness of the specimen (as we saw in Chapter 28).
- The defocus of the objective lens.
- Chromatic aberration which becomes more important as t , the thickness, increases.
- Coherence of the beam.

- Other factors: one example would be the intrinsic vibration in the material which we try to take account of through the Debye-Waller factor.

In principle, we could have the same image from two different structures. So obviously, this is the tricky part!

30.2 THE MULTISLICE METHOD

The basic multislice approach used in most of the simulation packages is to section the specimen into many slices, which are normal to the incident beam.

There are different methods for actually performing the multislice calculation. The different approaches have been developed for several reasons. Some try to optimize the use of available hardware. Others were written with the intention of providing a convenient method of simulating DPs using the same program. At least one package was written to make use of a popular

TABLE 30.1. Software

Cerius2	Runs on UNIX; by Accelrys. Look for c2_hrtem www.accelrys.com/products/cerius2/cerius2products/hrtem.html
EMS and jEMS	By Pierre Stadleman. Very widely used and user-friendly. Does Bloch wave and conventional multislice calculations. Is used by CuFour (see Chapter 24). Multi-platform including Mac, Unix, Windows. The 'j' denotes the java version
Kirkland	By Earl Kirkland. Well described in his book on the subject
MacTEMPAS	By Roar Kilaas. Runs on a Mac so it's very user-friendly. NCEMSS was the Unix version that was produced by NCEM
SHRLI81	By Mike O'Keefe. This program made image simulation freely available to everyone and introduced a generation to HRTEM simulation. Unix only, but no longer supported
WinHREM and MacHREM	www.hremresearch.com/Eng/download/documents/HREMcatE.html by HREM Research Inc. (Kazuo Ishizuka)

personal computer with a user-friendly interface (see Table 30.1). The principal methods for performing these calculations are

- The reciprocal-space formalism.
- The FFT formalism.
- The real-space approach.
- The Bloch-wave approach.

We'll go through the special features of each approach. The software packages which are readily available are listed in Section 1.6 and in Table 30.1.

30.3 THE RECIPROCAL-SPACE APPROACH

We project each slice onto a plane somewhere in the slice (usually top, bottom, or middle) giving a projected potential for that slice, and we call this the phase grating. We then calculate the amplitudes and phases for all the beams which will be generated by the incident beam interacting with the first projection plane. We could think of this as being a many-beam image calculation for a single slice. We then allow all these beams to propagate down the microscope in free space until they meet the next phase grating, which need not be the same as the first. The scattering calculation is now repeated for all the beams incident on this plane. This calculation produces a new set of beams which propagate through free space to the next phase grating, and so on. The process is summarized in Figure 30.1.

One point which you must remember: scattering by the phase grating does not just produce Bragg beams. It is crucial to keep track of the scattering in *all* directions. All of these beams will be incident on the next phase grating. So we don't just have Bragg beams, we *sample* all of reciprocal space.

A calculation based on a 128×128 array will impose a limit of ~ 4096 on the number of 'beams' which can be included in the calculation. This number might appear large, especially when you form a [110] HRTEM image of Si with six Bragg beams (plus the O beam) but, particularly for imperfect crystals, this number will be inadequate.

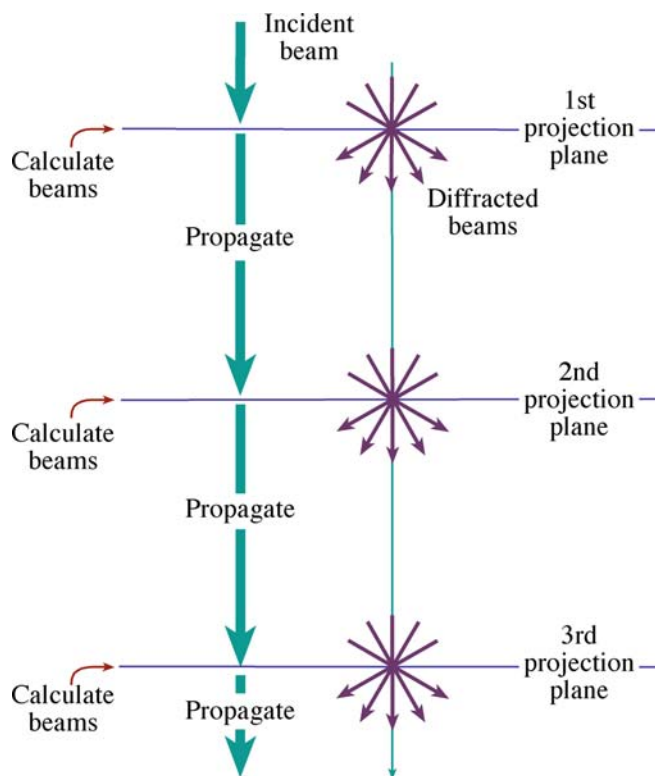


FIGURE 30.1. The potential within a slice is projected onto the first projection plane; this is the phase grating. We calculate the amplitudes and phases for all the beams generated by interacting with this plane and then propagate all the diffracted beams through free space to the next projection plane, and repeat the process.

k-SPACE BETWEEN THE BRAGG BEAMS

Why do we need to consider regions of *k*-space between the Bragg beams? In other words, why do we need to sample all of reciprocal space? The answer is that the Bragg beams contain information about the periodic structure, but all of the information from defects, i.e., non-periodic structure, is contained *between* the Bragg spots, though it will generally be quite close to them.

Essentially the multislice method considers three components

- ψ describes the *electron wave*.
- P is the propagation of the electron wave in free space: the *microscope*.
- Q is the phase grating: the *specimen*.

The process can be described by this equation

$$\psi_{n+1}(\mathbf{k}) = [\psi_n(\mathbf{k})P_{n+1}(\mathbf{k})] \otimes Q_{n+1}(\mathbf{k}) \quad (30.1)$$

where $\psi_{n+1}(\mathbf{k})$ is the wave function in reciprocal space at the exit of the $n+1$ slice and the symbol \otimes denotes a convolution as before. $P_{n+1}(\mathbf{k})$ is the propagator for the $n+1$ slice. In other words, this is expressing the Fresnel diffraction phenomenon for this one slice because we are

making a near-field calculation. (Look back to Section 2.8 for a discussion of near-field versus far-field.) Similarly, $Q_{n+1}(\mathbf{k})$ is the phase-grating function; it is a transmission function, for the $n+1$ slice.

The three functions $\psi(\mathbf{k})$, $P(\mathbf{k})$, and $Q(\mathbf{k})$ are all functions in reciprocal space, so this approach is referred to as the reciprocal-space formulation. Notice that they are all two-dimensional arrays. We can think of the different terms as being diffracted beams within the specimen. We can easily insert a circular objective aperture of radius \mathbf{r} ; we just require that all values of $\psi(\mathbf{k})$ are zero for $\mathbf{k} > \mathbf{k}_r$.

To give you an idea of the complexities involved, consider what values of $Q(\mathbf{k})$ you must use in the calculation. $Q(\mathbf{k})$ must go out twice as far as $\psi(\mathbf{k})$ or $P(\mathbf{k})$ in reciprocal space. You can understand why by considering Figure 30.2. If you represent the number of

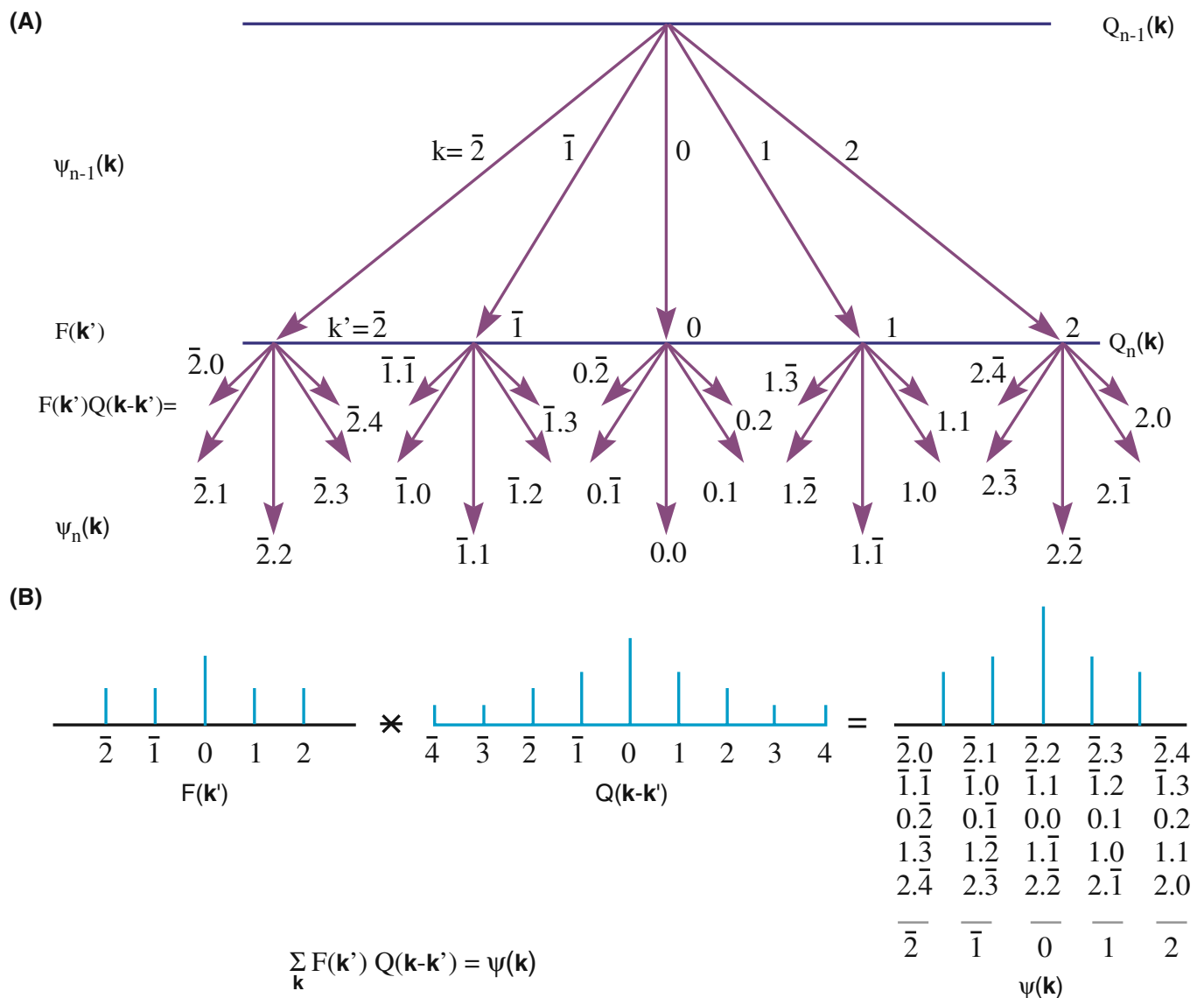


FIGURE 30.2. (A) Schematic used to explain why, in the one-dimensional case, $Q(\mathbf{k})$ must take account of twice as many \mathbf{k} values as $\psi(\mathbf{k})$ or $P(\mathbf{k})$. Consider wave $k = \bar{2}$ from $Q_n(\mathbf{k})$: to produce wave at $+2$ at this point you need to add 4 to $\bar{2}$ and similarly for every possible wave in slice $Q_n(\mathbf{k})$. As summarized in (B) $Q(\mathbf{k}-\mathbf{k}')$ extends from 4 to $+4$ so that $\psi(\mathbf{k})$, which we want, extends from $\bar{2}$ to $+2$, including all possible combinations of \mathbf{k}' and \mathbf{k} .

beams from slice $Q_{n-1}(\mathbf{k})$ as $F(\mathbf{k}')$ then $Q(\mathbf{k}-\mathbf{k}')$ must go out to $k = -4$ because when you multiply these two functions to give $\psi(\mathbf{k})$ you can produce $k = -2$ by using $k = -4$ in Q and $k = +2$ in F as in Figure 30.2B. Putting this into an equation we have

$$\sum_{\mathbf{k}'} F(\mathbf{k}') Q(\mathbf{k} - \mathbf{k}') = \psi(\mathbf{k}) \quad (30.2)$$

where

$$F(\mathbf{k}) = \psi(\mathbf{k}) P(\mathbf{k}) \quad (30.3)$$

The function $Q(\mathbf{k})$ is a ‘probability map.’ What we are doing here is using the convolution to describe multiple scattering.

We can illustrate the complexity of the calculation by considering a 128×128 array for $Q(\mathbf{k})$ using SHRLI81 (see Table 30.1). The maximum value for (k_x, k_y) is only (31,31) but even so, the number of diffracted beams is nearly 4096. Remember, we usually just use the seven inner Bragg beams in, e.g., the Si $\langle 110 \rangle$ DP, as we saw in Figure 27.3; most of the beams in our calculation are not Bragg beams. However, you will remember that the information concerning defects in crystals is contained in the regions between the Bragg spots in the DP, so it does make sense. Specific examples of $Q(\mathbf{k})$, including numerical computations of the phase change per slice, are given by John Barry.

30.4 THE FFT APPROACH

We can recast equation 30.3 to maximize the efficiency of the computer in using fast Fourier transform (FFT) routines. In equation 30.4, the notation F and F^{-1} tell us to take the Fourier transform or the inverse transform of the function inside the brackets.

$$\psi_{n+1}(\mathbf{k}) = F\{F^{-1}[\psi_n(\mathbf{k}) P_{n+1}(\mathbf{k})] q_{n+1}(\mathbf{r})\} \quad (30.4)$$

In this equation, $q_{n+1}(\mathbf{r})$ is the real space form of $Q_{n+1}(\mathbf{k})$, i.e., it is the inverse Fourier transform of $Q_{n+1}(\mathbf{k})$. So $q(\mathbf{r})$ is a real-space phase grating. Now we can look at some numbers for the calculation and take $Q(\mathbf{k})$ as a 128×128 array to keep the calculation small. The main steps carried out by the computer are

- Multiply $\psi_n(\mathbf{k})$ by $P_{n+1}(\mathbf{k})$: that is a 64×64 array times another 64×64 array. Remember that we are limited to 64 points, if we have a 128×128 array, because the Q array must be twice as large in all directions in \mathbf{k} space.
- Take the inverse Fourier transform of the result.
- Multiply this new result by $q_{n+1}(\mathbf{r})$ which is the 128×128 array.

- Fourier transform the final result and set all values outside the inner 64×64 array equal to zero so that we can repeat the process for the next slice.

You will notice that this example used a square array. In modern programs, we are not restricted even to using powers of 2 but this helped the original FFT routines. You will see the value of this advance when we examine some defect calculations later. If you are interested in the mechanics of the FFT routine and other aspects of this simulation approach, the article by Mike O’Keefe and Roar Kilaas is required reading.

30.5 THE REAL-SPACE APPROACH

As we noted earlier, image simulation used to be limited by your budget, i.e., by your computer. The real-space approach was developed, in part, to decrease the time needed for the calculations by using our knowledge that $P(\mathbf{r})$ is strongly peaked in the forward direction. In our notation, the method developed by Coene and Van Dyck for calculating $\psi(\mathbf{r})$ can be expressed by the equation

$$\psi_{n+1}(\mathbf{r}) = [\psi_n(\mathbf{r}) \otimes P_{n+1}(\mathbf{r})] q_{n+1}(\mathbf{r}) \quad (30.5)$$

where $P_{n+1}(\mathbf{r})$ is now the propagator in real space and $q_{n+1}(\mathbf{r})$ is again the real-space phase grating. Once you have written this, it’s all computing, which is a substantial task since the size of the multislice calculation is the size of the largest array, i.e., $Q(\mathbf{k})$ or $q(\mathbf{r})$.

30.6 BLOCH WAVES AND HRTEM SIMULATION

Although we saw in Chapters 14 and 15 that electrons propagate through crystalline specimens as Bloch waves, the multislice method we’ve described so far is essentially a ‘diffracted-beam’ approach. In two classic papers Fujimoto (1978) and Kambe (1982) showed that, for the perfect crystal, the HRTEM image may be understood simply in terms of images of Bloch waves. The key point is that, although a large number of diffracted waves are formed, only a small number of Bloch waves determine the appearance of the image, providing the crystal has a sufficiently high symmetry. Following Kambe’s ‘simple’ example, we consider the case where only three Bloch waves i, j , and k are significant. Let’s assume that Bloch waves i and j are in phase at a thickness $z = D$. Then we have

$$e^{ik_z^{(i)}z} = e^{ik_z^{(j)}D} \quad (30.6)$$

Using our expression for ψ , namely

KTH BLOCH WAVE

Don't confuse the k th Bloch wave with the \mathbf{k} -vector!

$$\psi(\mathbf{r}) = \sum_i C^{(i)} \phi^{(i)}(x, y) e^{ik_z^{(i)} z} \quad (30.7)$$

and the normalization rule

$$\sum_i C^{(i)} \phi^{(i)}(x, y) = 1 \quad (30.8)$$

we can therefore express ψ at $z = D$ in terms of our three Bloch waves.

$$\psi(x, y, D) = \left[C^{(i)} \phi^{(i)} + C^{(j)} \phi^{(j)} \right] e^{ik_z^{(i)} D} + C^{(k)} \phi^{(k)} e^{ik_z^{(k)} D} \quad (30.9)$$

We rearrange this equation so that we can extract the phase factor $e^{ik_z^{(i)} z}$ ($= e^{ik_z^{(i)} D}$). We write

$$\psi(x, y, D) = \left[1 - C^{(k)} \phi^{(k)} \right] e^{ik_z^{(i)} D} + C^{(k)} \phi^{(k)} e^{i(k_z^{(k)} - k_z^{(i)}) D} e^{ik_z^{(i)} D} \quad (30.10)$$

$$\psi(x, y, D) = e^{ik_z^{(i)} D} \left[1 + \beta_{ik}(D) C^{(k)} \phi^{(k)} \right] \quad (30.11)$$

where we've defined a new parameter β given by

$$\beta_{ik}(D) = e^{i(k_z^{(k)} - k_z^{(i)}) D} - 1 \quad (30.12)$$

These equations tell us that if any two of the Bloch waves (here they are i and j) are in phase then the amplitude of the wave at the exit surface is determined by the third Bloch wave.

If the third Bloch wave is also nearly in phase, we have a relation like equation 30.6 but with i, j , and k all equal. Then we can approximate $\beta_{ik}(D)$ by

$$\beta_{ik}(D) = i \left[(k_z^{(k)} - k_z^{(i)}) D + 2n\pi \right] = i\gamma_{ik}(D) \quad (30.13)$$

Now we've defined another factor γ_{ik} . If you plug this expression back into equation 30.11, you see we have a pure phase object. All the diffracted beams will be shifted in phase by $\pi/2$.

Now you can test the effects of how we change k . Consider what conditions this will really correspond to using equations 30.11 and 30.13

- If k is such that the phase of Bloch wave k is ahead of i and j (which were equal), then you'll see a 'negative' image of $C^{(k)} \phi^{(k)}$. A 'delayed' k gives us the 'positive' image.
- For the Ge $\langle 110 \rangle$ zone axis, HRTEM image at 100 kV, only three Bloch waves are strongly excited.

The relationship to the Bloch-wave contours in Chapter 14 is clear. Using this information and the projected potential shown in Figure 30.3, Kambe calculated the Bloch-wave amplitudes and the two ideal images of the Bloch waves: one positive and the other negative. In the calculation of different images for increasing thickness, several images corresponding to a single Bloch wave can be predicted and identified, as shown in the figure. At other thicknesses, the images form by a combination of Bloch waves. So, what can we learn?

- For a perfect crystal, you may need as few as three Bloch waves to give the essential features of an HRTEM zone-axis image.
- There is a direct connection between the WPOA (go back to the definition if you don't remember) and the propagation of Bloch waves.

THE MULTISLICE APPROACH

We usually use this to simulate HRTEM images. It is actually a very elegant form of brute force.

We saw in Chapter 14 that the electron propagates as Bloch waves inside the crystal. The reason we don't use Bloch waves is that our specimens are not perfect. However, EMS does give you the option of using this approach.

30.7 THE EWALD SPHERE IS CURVED

When you are using the TEM, some other complications arise because the Ewald sphere is curved

- If you align the beam exactly parallel to a zone axis, \mathbf{s} will be non-zero for every Bragg reflection. In fact, it will also be different for each type of reflection.
- If you do not align the beam exactly parallel to the zone axis, then \mathbf{s} will also be slightly different for each reflection in that zone.
- If you change the wavelength of the electrons, the radius of the sphere changes.
- If you converge the beam, then you'll add a thickness to the Ewald sphere.

The point is, knowing precisely what the correct values are to put in the program will also require thought and work.

30.8 CHOOSING THE THICKNESS OF THE SLICE

So far, we've just cut the specimen into slices in the computer without considering how thick each slice should be, or even whether they should all be the

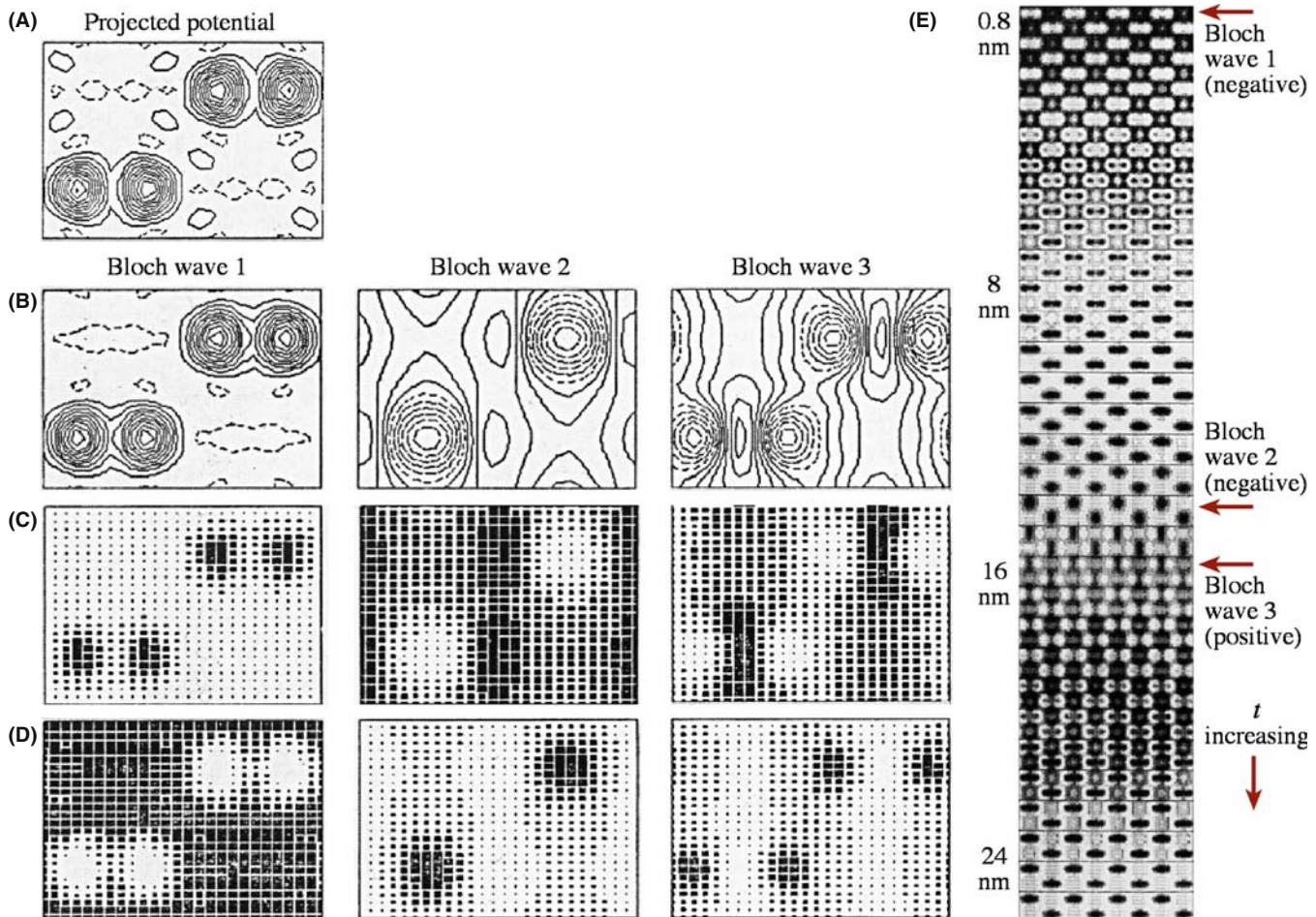


FIGURE 30.3. (A) The projected potential for Ge where the contour lines represent changes in potential of -10 eV , and the dashed lines are positive values. (B–D) are for the three Bloch waves 1, 2, and 3 for 100-keV electrons: (B) the amplitudes, (C) ideal positive images, (D) ideal negative images. (E) Calculated lattice images showing thickness dependence.

same. If all the slices are the same, then there can be no information about the z -direction. Although HOLZ lines are not important for the simulation of HRTEM images, some of the programs we are discussing can now just as readily be used to simulate CBED patterns and HOLZ lines. So, following the philosophy of attacking problems with different techniques, you should be aware of these limitations, since it is easy to overlook the simplifications you made once you see the computed image. You should remember that when you are studying a material with a large unit cell, the reciprocal lattice spacing will be short in the beam direction, so HOLZ effects come into play sooner.

Consider the different methods for making the slice

- You could calculate the projected potential for a thick slice and then do n calculations with slices which are $1/n$ times this thickness.
- A better approach would be to subdivide the cell into layers of atoms, create a different grating for each of these layers, and then run the program with the sequence.

For example, if the beam is aligned along the $[111]$ direction of an fcc crystal, then you would have three identical gratings displaced relative to one another, giving the ABC stacking of close-packed planes. This approach would allow you to test for the effect of a real error in the stacking sequence normal to the beam. Even this point can be a bit difficult. In general, you orient the beam to be parallel to a particular zone axis $[UVW]$ so that the planes in that zone are parallel to the beam (so our projection works). If the material is not cubic, we will not generally have a low-index plane normal to the beam to make this slice.

30.9 BEAM CONVERGENCE

When you are recording HRTEM images, you need to keep exposure times short. So, if you don't use parallel illumination, you have to take account of the beam convergence when simulating the images. O'Keefe and Kilaas (see also the paper by Self and O'Keefe) have developed one approach to address this problem. If

the beam actually has some convergence, then the diffraction spots will be disks, as illustrated in Figure 30.4, so you need to simulate disks in the DP. Experimentally, the large objective aperture admits many disks, so in the simulation routine you should sample each disk at many points. This means the program needs to calculate the image at each of these convergence angles and average all the resulting images. Of course, the objective aperture is easily applied in the computer. If you choose 49 points, you can make the sampling interval in reciprocal space $\leq 0.1 \text{ nm}^{-1}$. It is instructive to examine just how much work is necessary to sample the 49 points.

We can start by writing the usual expression for the χ , the phase change due to the objective lens

$$\chi = \pi\Delta f\lambda u^2 + \pi C_s\lambda^3\left(\frac{u^4}{2}\right) \quad (30.14)$$

Then differentiate this with respect to the variable u

$$\frac{d\chi}{du} = 2\pi(\lambda u\Delta f + C_s\lambda^3 u^3) \quad (30.15)$$

This equation tells us that if u changes by δu , then χ changes by

$$\delta\chi = 2\pi\lambda(u\Delta f + C_s\lambda^2 u^3)\delta u \quad (30.16)$$

Now we choose $\delta\chi$ so that

$$\delta\chi < \frac{2\pi}{n} \quad (30.17)$$

where n will allow us to determine the maximum change in χ between two points in the disk. For example, if $n = 12$, then the maximum value of $\delta\chi$ is 30° . Combining equations 30.15 and 30.17, we can write

$$\delta u = [n\lambda u(\Delta f + C_s\lambda^2 u^2)]^{-1} \quad (30.18)$$

If we plot χ versus u (or play with equation 30.15 and its derivative), then we find a minimum in χ at

$$\Delta f = -C_s\lambda^2 u^2 \quad (30.19)$$

and an inflection at

$$\Delta f = -3C_s\lambda^2 u^2 \quad (30.20)$$

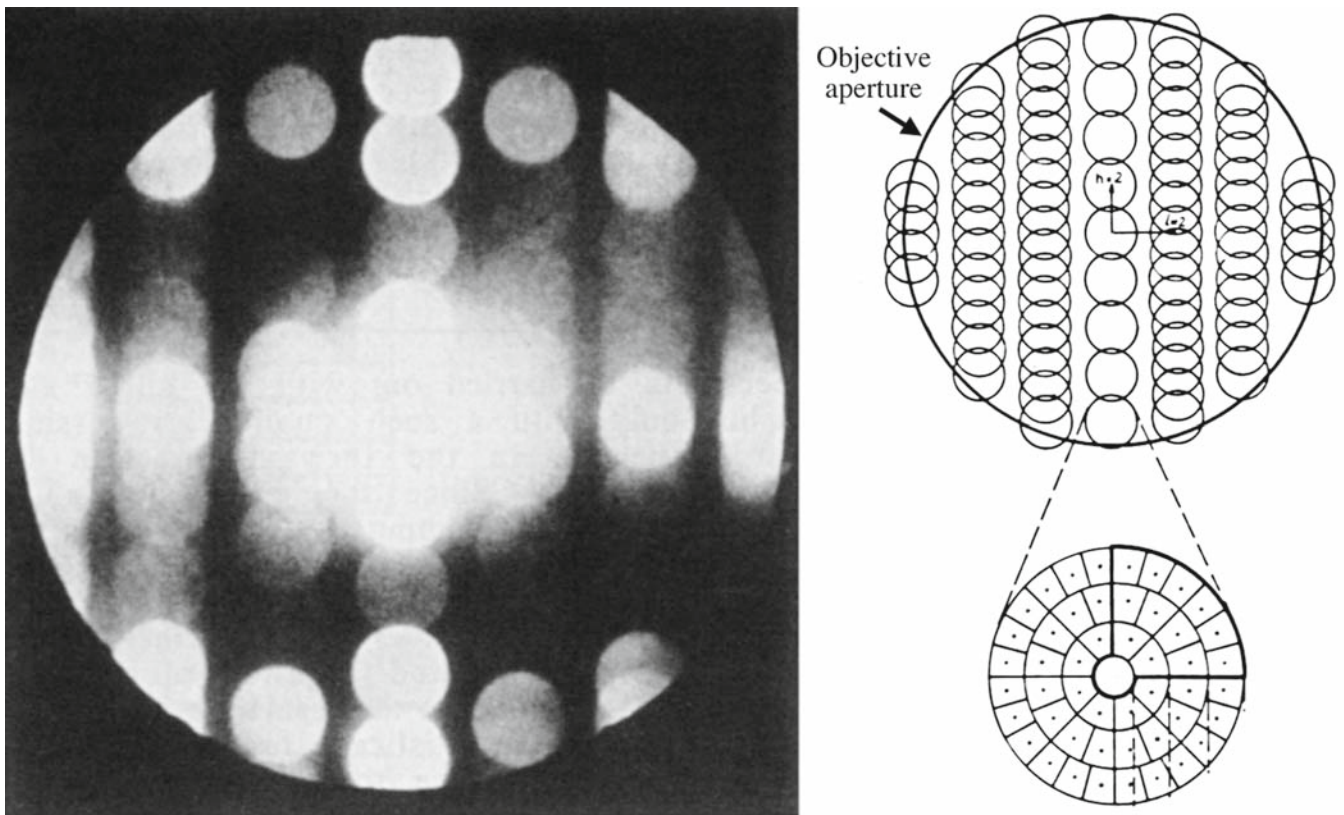


FIGURE 30.4. Disks in the DP from a crystal of $\text{Nb}_{12}\text{O}_{29}$. The computer simulation can divide each disk into many sectors and simulate the image for each sector, as shown in the schematic, excluding sectors which are intersected by the objective aperture.

So the simulation program can check to find the smallest δu at an inflection point, which equations 30.18 and 30.20 tell us is

$$\delta u = - \left[\frac{27C_s}{(\Delta f)^3} \right]^{1/2} \left(\frac{1}{2n} \right) \quad (30.21)$$

The value of δu therefore depends on both C_s and Δf .

THE BLACK BOX

Remember where all these calculations takes place!

You can also appreciate the relevance of this type of approach if your disks actually intersect the objective aperture, as shown in Figure 30.4. Put another way, you can learn two lessons from this analysis

- Always try to minimize the convergence of the beam when recording HRTEM images.
- Use an aperture which does *not* cut through the diffraction disks.

30.10 MODELING THE STRUCTURE

To simulate any HRTEM image, you need a unit cell. If you are concerned only with perfect crystals, then your program should have all the space groups already included so that you only need to add the lattice parameters (lengths and angles) and the occupied sites for your material. If you are interested in simulating images from defects, then you have to create a new unit cell which must be sufficiently large that it will not add effects due to the edges. There are many ways to create this defect unit cell. You can input from other programs, such as those performing atomistic modeling of defects, or create your own starting structure. In either case, you will need to move atoms, either manually or following a rule you've selected for image matching, to optimize the match between your experimental series of through-focus images and the simulated images.

At some stage, you will find it useful to combine different slices, as when simulating GBs with or without a surface groove, or modeling large complex unit cells using a multilayer approach. We'll now go through some specific features of this task and return to modeling in Chapter 31 when we discuss quantitative HRTEM.

30.11 SURFACE GROOVES AND SIMULATING FRESNEL CONTRAST

The analysis of interfaces by the Fresnel-fringe technique, which we introduced in Chapter 23, illustrates the importance of image simulation and emphasizes

that it is not just for HRTEM. The calculation is complicated for several reasons, as shown in Figure 30.5A

- The potential change at the interface is probably not abrupt.
- The potential depends on the detailed structure of the interface.
- During preparation, TEM specimens may be preferentially damaged at GBs, giving rise to surface grooves.

If you use a thicker specimen, you'll reduce the effect of surface grooves on any Fresnel fringes, but in practice, your foil thickness is usually limited (~ 20 nm), since you need to view the boundary exactly edge-on.

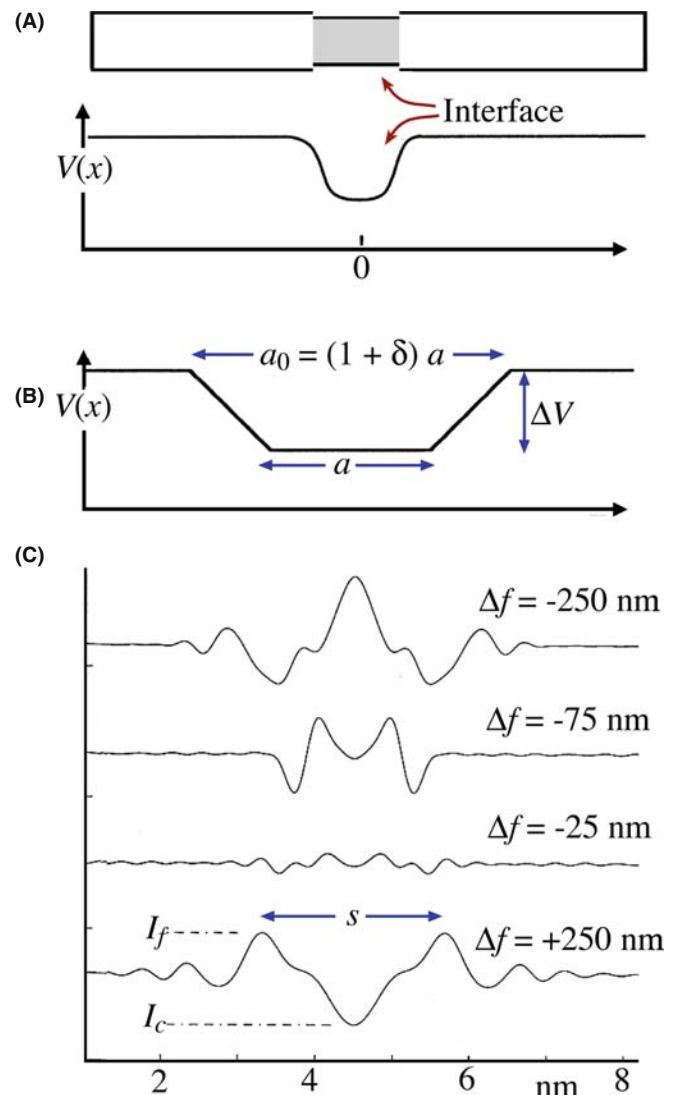


FIGURE 30.5. (A) Schematic of a GB containing a layer of material with a different inner potential; (B) one model used to represent such a GB giving variable parameters a , a_0 , and δ ; (C) a typical set of simulated Fresnel-fringe intensity profiles at increasing Δf : s is the distance between the first two fringes, I_c and I_f are the intensities of the central and first fringe, respectively.

Even for foils this thick, surface grooves can influence the projected potential considerably. If we assume that the bulk has a mean inner potential $V = 20$ V, and take a typical potential drop for an intergranular film to be ~ 1 V, then the total projected potential drop for a 20 nm thick foil would be the same as that caused by a pair of grooves at the top and bottom surfaces which are only 0.5 nm deep. Although the surface groove may be partly filled with a second phase, the effect on the Fresnel fringes can still be substantial.

We can examine Fresnel fringes using different methods. In all of them, we describe the potential at the interface in terms of the projected potential drop $\Delta V_p = t\Delta V$, an inner width a , and an outer width a_0 , and a ‘diffuseness,’ δ , defined by

$$a_0 = (1 + \delta)a \quad (30.22)$$

These parameters are shown in Figure 30.5B. Then we construct models of a foil with a surface groove at the edge-on interface by combining such potentials.

The models: Values of $\delta = 0.5$ and $\delta = 0.2$ represent shallow and steep surface grooves, respectively. The total projected potential drop can be due to a real change in V or a change in t . A groove without a film implies $a = 0$. If $a = 1$ nm and $a_0 = 1.5$ nm, the model could correspond to two different situations

- If the atoms at the interface relax, then the atomic density at the interface will usually be reduced. This occurs at both structured interfaces and those where a layer of glass is present.
- Surface grooves at the interface.

What image simulation shows is that the relative shapes and sizes of these models are more important than the actual dimensions. Therefore, we can give most of the following analysis in terms of dimensionless quantities. Inner potentials are typically 5–10 eV. Except for very small defocus values ($\Delta f < \approx t$), we find that the distribution of the potential through the foil is not important. Usually, the projected potential at the interface is lower than that in the bulk. However, the opposite situation can occur, e.g., when a Bi_2O_3 -rich phase is present at interfaces in ZnO. When we discuss the calculated profiles, the term ‘interface width’ will be used for the parameters a and a_0 , whether they actually correspond to an intergranular film, a surface groove, or otherwise.

SPACING OF FRESNEL FRINGES

The distance from the edge to the first fringe is proportional to $(\lambda\Delta f)^{1/2}$. The fringe spacing, s_f , can be extrapolated to zero defocus to obtain the interface width, based on the relation $(s_f - a) \propto \sqrt{(\lambda\Delta f)}$.

This relation was first described by Clarke and only holds when a is large and Δf is relatively small; then the fringes from each ‘edge’ at the interface are independent. We observe the minimum fringe spacing at small values of defocus and this spacing can be used to provide a measure of the interface width. For more details on the simulation of Fresnel fringes, you should check the original articles.

In practice, the analysis of Fresnel fringes is impaired, not only by specimen artifacts such as surface grooves, but also by various sources of noise, which all add to the uncertainty of measurements, especially at small values of Δf . For diffuse interfaces, the contrast decreases rapidly as Δf approaches zero (Figure 30.5C), and measurements of the fringe spacing for $\alpha > 0.7$ are increasingly susceptible to noise and artifacts. You can always use larger defocus values and thus obtain higher contrast. However, without prior knowledge about the shape of the potential drop (its ‘diffuseness’), we can’t reliably determine the interface width by measuring the fringe spacing alone. Since the fringe spacing is dominated by the outer width, a_0 , we may easily overestimate the interface width. The atomic density in a region close to the boundary is also often reduced, even if the boundary is structured, so you can easily misinterpret the image as showing the presence of an intergranular film when actually there is no amorphous film.

The region of defocus, where the central fringe shows little contrast, provides complementary information to the fringe spacings, so it is more sensitive to the inner width.

From this discussion, you’ll appreciate that, before you can completely understand the effect of any intergranular films, you must estimate the extent to which surface grooves are present in your specimen. Shadowing (e.g., using platinum or gold) may provide evidence for surface grooves, but in the case where the surface groove is already filled (e.g., if your specimen was coated with carbon or contaminated without your knowing), this technique won’t work.

CHARACTERIZING A POTENTIAL

The conclusion is that you must use all the information in the image to characterize the shape of a potential well and you need to know what is in it!

To summarize, this discussion gives us a method for analyzing Fresnel fringes from a GB. We can draw some conclusions

- To interpret the contrast from Fresnel fringes at GBs, you must simulate images of many different interface models. In particular, it is essential that you

consider the possibility of artifacts such as surface grooving. Even a rather ‘flat’ or diffuse surface groove may influence the fringes in some range of defocus values.

- Both the fringe spacing and the central fringe intensity depend on the shape of the potential well and are sensitive to surface grooving.
- The interface width, which you can infer from the fringe spacing, is dominated by the outer width of a diffuse interface.
- A direct match with the $s_f - a$ curve (or with similar simulated curves when the assumptions employed here fail) leads to a better estimate of the average interface width, but cannot give you much information on the shape of the potential well.
- Determining when the central fringe is weak (the range Δf) gives complementary information on the interface width which, in combination with the estimate based on the fringe spacing, you can use to evaluate the diffuseness of the potential well.

30.12 CALCULATING IMAGES OF DEFECTS

When we simulate HRTEM images of perfect crystals, we only need to input the unit cell and the program generates the rest of the specimen. If we want to calculate the image of a defect, we have to use the same approach: we set up a unit cell to contain the defect and the program treats it like any other unit cell. This is known as the *periodic continuation* method for defect calculation. What we’ve actually done is shown in Figure 30.6: there is an array of defects throughout our specimen in all directions. We need to know two things

- To what extent does this ordered array introduce artifacts in the image?
- Have we created interfaces where the ‘cells’ join which may influence our image?

An example of a supercell for a GB is shown in Figure 30.6. This figure illustrates clearly how we can create a cell which is more suitable for this periodic continuation by including two defects in a single supercell. As shown in this figure, the periodic continuation then not only creates many other GBs but also makes them very long. If we don’t match the crystals exactly at the edges of the supercell we create a different ‘ghost’ boundary.

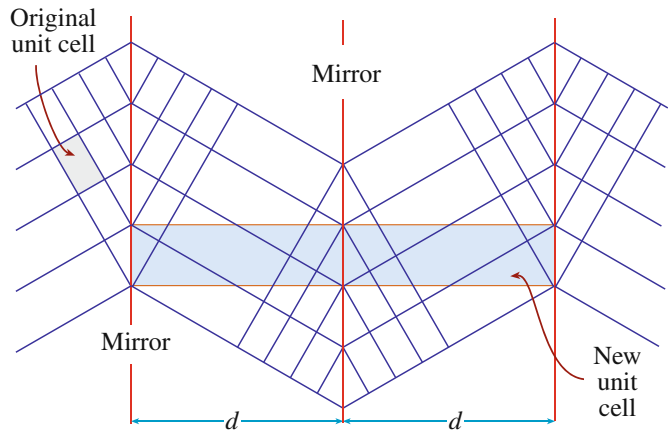


FIGURE 30.6. The periodic continuation technique illustrating how an artificial unit cell can be constructed to contain two GBs, thus allowing the HRTEM image to be simulated. The distance (d) between the two interfaces can be varied to check for overlap artifacts.

You can see that this really can be a problem by considering the DP which our new cell would produce. We are calculating the image of a small part of a periodic array of interfaces. Periodic arrays in real space produce rows of extra spots in reciprocal space. If we include these spots in forming the image, we should change the image. The solution for image simulation is quite simple, make the supercell wider and wider until the change in the image detail is less than some specified limit. However, don’t try to interpret the data in the calculated DP without consulting the paper by Wilson and Spargo.

An alternative approach to the periodic continuation approach has been developed by Coene et al. and is called the real-space patching method. This method uses the ‘real-space’ image simulation approach to perform the calculation. The structure you want to simulate can be divided into a number of different ‘patches’ as illustrated in Figure 30.7; the image from each patch is

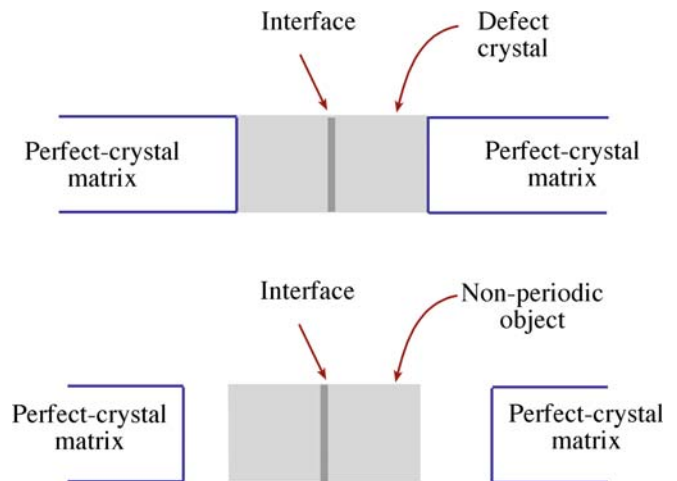


FIGURE 30.7. In the real-space patching method, the defect crystal (in this case the interface and several adjacent layers) is a non-periodic object that is surrounded by a perfect crystal matrix.

SIMULATING DEFECTS

The periodic continuation is most widely used. Real-space patching avoids the artificial cell.

calculated for a slice and then the patches are joined together. The key, of course, is that you must correctly take account of what happens at the edge of each patch. This means each patch needs some information about the neighboring patches. Assuming (correctly) that this can be done, you can appreciate the nice feature of this approach: we avoid the artificial interference effects due to the array of defects that would be produced by the periodic continuation technique. The defect does not now 'see' its own image; it only sees the perfect matrix on all sides.

30.13 SIMULATING QUASICRYSTALS

There are several problems in simulating HRTEM images of quasicrystals, not least of which concerns which model you should use. (Remember that quasicrystals are not periodic objects anyway!) Several models have been reviewed by Shoemaker, and the possibilities are illustrated by the work of Beeli and Horiuchi, who used a combination of 10 layers in the multislice calculation. The layers are made up from the planar structures shown in Figure 30.8. The final structure (shown in Figure 30.8A) is made up of two

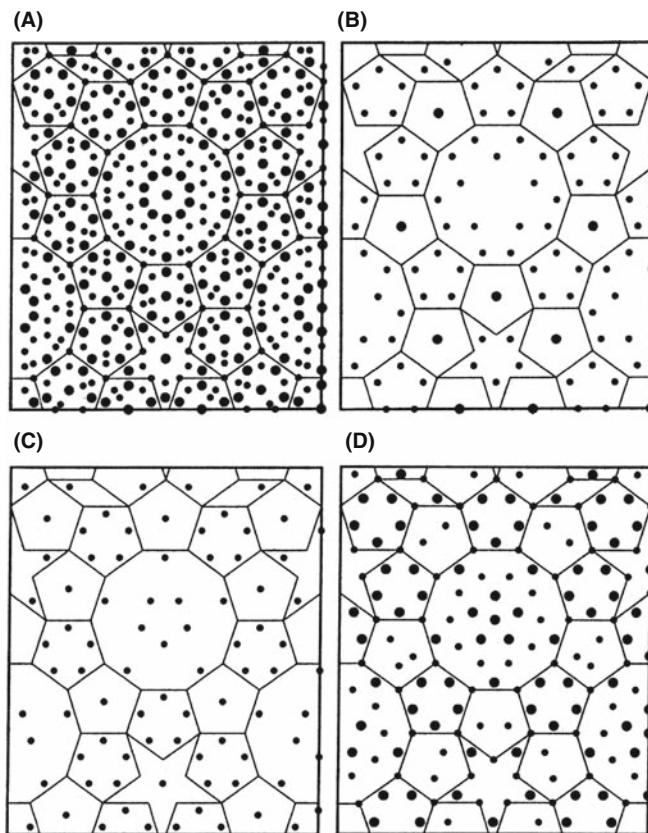


FIGURE 30.8. Projections used to simulate images of $\text{Al}_{70}\text{Mn}_{17}\text{Pd}_{13}$ quasicrystals. (A) Combination of all the layers; (B–D) layers are used to contribute (repeatedly) to (A). The edges of all the tiles are 0.482 nm. The large circles denote Al atoms.

sets of five layers. The first set of layers is B-C-D-C-B in this figure. The second set of five layers is constructed from the first by using the screw symmetry of the structure; the screw axis has 10_5 screw symmetry. The supercell used was 3.882 nm by 3.303 nm which was chosen to contain a complete decagonal cluster which is 2.04 nm in diameter and the center part of a pentagon tile. The calculation was then carried out for thicknesses up to 10 nm.

Z-CONTRAST

Since the atoms lie along columns for phase-contrast imaging we can also use Z-contrast imaging.

The results of such calculations with only Al and Mn atoms are illustrated in Figure 30.9. The edges of the cells are essentially artificial because, as we just noted, the structure used in the calculation is a 'unit cell' and a quasicrystal doesn't have one. In spite of these difficulties Beeli and Horiuchi could conclude that the image match was much improved when Pd atoms were included to replace some of the Mn atoms in the D layer and Al atoms in the B-C layers with the results shown in Figure 30.10.

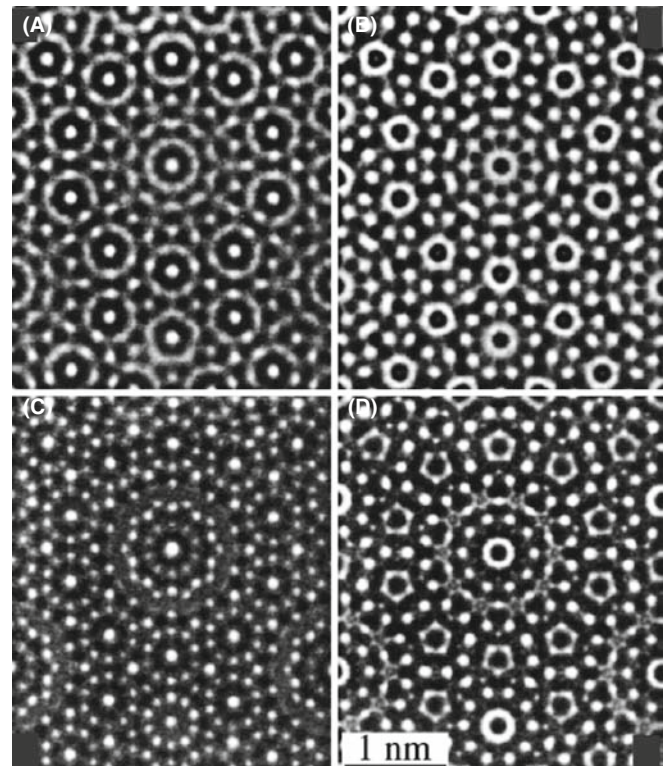


FIGURE 30.9. Four simulated images of the model constructed from the layers shown in Figure 30.8 using only Al and Mn atoms. The thickness is 3.77 nm, which corresponds to three periods in the beam direction. The values of Δf are (A) 0 nm, (B) 46 nm, (C) 88 nm, and (D) 124 nm.

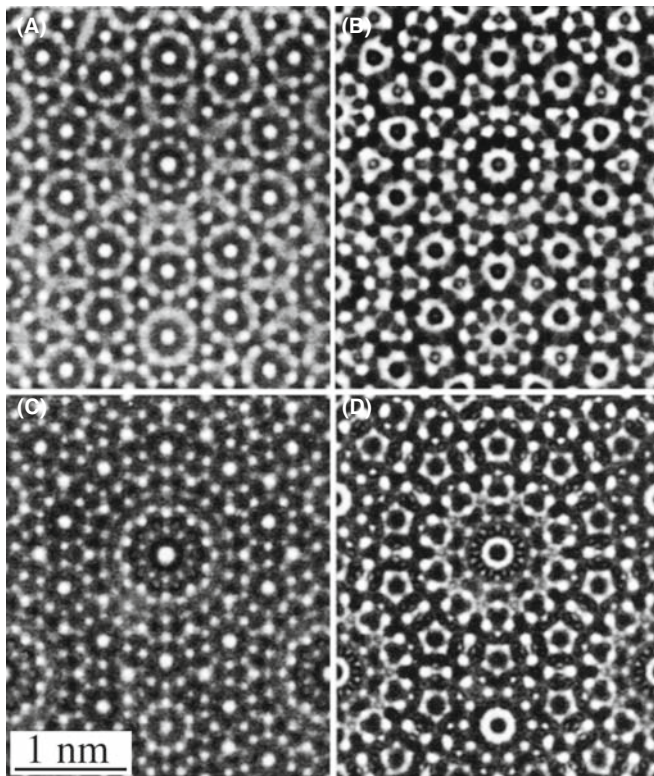


FIGURE 30.10. Examples of simulated images of the quasicrystal shown in Figure 30.8 but substituting Pd atoms for Mn atoms. The values of Δf are (A) 0 nm, (B) 48 nm, (C) 88 nm, and (D) 128 nm.

SIMULATING IMAGES OF A QUASICRYSTAL

As you realize, the problems in such an image calculation are increased because the quasicrystal does not have translational symmetry but we must impose such a symmetry to do the calculation.

Another illustration of the success of HRTEM comes from the work of Jiang et al. on quasicrystals with eightfold symmetry. Here the multislice calculation could again be made using a relatively simple sequence of four layers ABAB', where the layers are at $z = 0, 0.25, 0.5,$ and 0.75 . The structures of the A and B layers are shown in Figure 30.11 with the B' layer being a 45° -rotated B layer; i.e., the B and B' layers are again related by a screw axis, this time, it's an 8_4 screw axis.

- In each of these examples, it is possible to view the same structure parallel to an orthogonal axis.
- Quasicrystals do not have translational symmetry, but we pretend they do for thickness calculations and for the periodic continuation of the unit cell.

Our reason for showing so much detail on these rather esoteric materials is that they show what can be done using image simulation. Furthermore, they

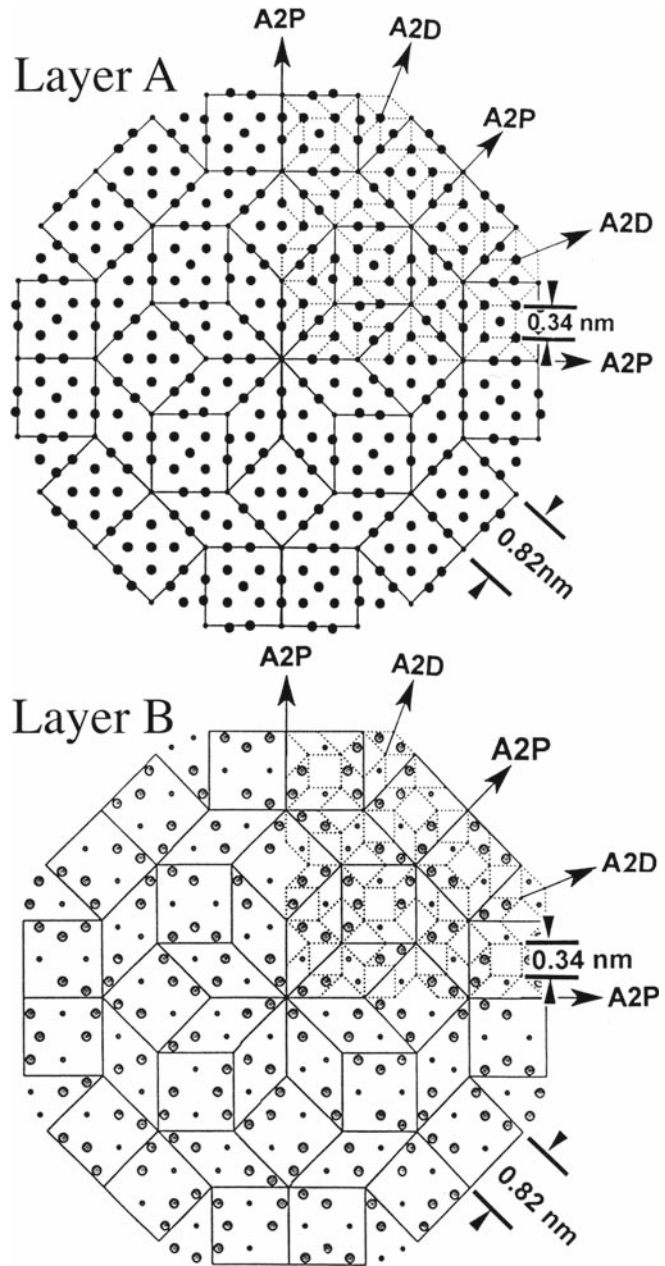


FIGURE 30.11. The model used to simulate quasicrystals with eightfold symmetry. The structure for the simulation was constructed as a four-layer sequence ABAB' where the B and B' layers are related by an 8_4 screw axis.

emphasize the important fact that although we can construct the crystal using different layers and different sequences of layers, we always use a projection of the structure, to compare with the experimental image.

30.14 BONDING IN CRYSTALS

We mentioned early on that one problem we have with simulation concerns the fact that atoms are bonded in different ways in different materials. The standard

approach has been to use values for structure factors tabulated by Doyle and Turner and by Doyle and Cowley. These values were calculated using a relativistic Hartree-Fock (RHF) model for the atomic potential. An alternative approach is to relate the scattering factor for electrons (f_e) to that for X-rays (f_x) using the Mott equation, or to use a more sophisticated atomic potential known as the relativistic Hartree-Fock-Slater (RHFS) model. Carlson et al. give tabulated results while Tang and Dornignac have made detailed comparisons for HRTEM imaging.

O’Keeffe and Spence have re-examined the meaning of the mean inner potential. One of the reasons that we need to understand this concept is that we often link data from X-ray diffraction and data from electron diffraction. As usual, computers are making it possible to do more elaborate calculations using other potentials.

While this is an evolving study, some important results have been obtained

- The inner potential is very sensitive to bonding effects. O’Keeffe and Spence discuss this result for MgO (large ionic component), Si (covalent), and Al (metallic).
- We are still not able to take full account of bonding effects, which could be important for HRTEM images.

This paper by O’Keeffe and Spence is a highly recommended reading for those who have a strong physics background but think TEM is a ‘known’ subject! For a very different approach, see the paper by Zuo and Spence which uses the DP to deduce information about bonding.

30.15 SIMULATING Z-CONTRAST

This is a topic where you can make an impact! The clues are in Chapters 2 and 3 and in the companion text.

30.16 SOFTWARE FOR PHASE-CONTRAST HRTEM

This is a short section because it is summarized in Table 30.1. HRTEMs are priced at up to \$4 M as of 2009. Image simulation is essential for HRTEM. Very few students can, or should, write their own programs to simulate HRTEM images—it has already been done many times. The catch is that no manufacturer supports any software package for simulating HRTEM images. Your lab must have at least two programs for you to use. In future we will see much more use of Mathematica and Matlab in this field.

CHAPTER SUMMARY

If you are going to do HRTEM imaging, you must be prepared to use image simulation to assist you in interpreting your images. If you want to do quantitative imaging, simulation is an essential component of the process. Most materials scientists using TEM will want to use one of the established software packages in Table 30.1. There are several important conclusions contained in this chapter

- Make sure that you know all you can about your specimen. We illustrated the dangers with our discussion of grooved GBs. You can waste too much time looking at artifacts caused by specimen preparation.
- Make sure that you know all you can about your TEM. You now have some idea of how many parameters are required by the simulation routines. Beware of the parameters which you did not measure for your machine. The program will need to use some value: *you* should make sure it is appropriate.
- Make sure that you accurately align your TEM before you record any images.
- If possible, use more than one program to simulate the images. At least try this once.
- Record a through-focus series and check for changes in Δf by repeating the first image.
- The fact that the thickness of your specimen varies can be a great asset provided you can determine that thickness; i.e., it gives you another variable.

If you can afford a C_s -corrected HRTEM, then C_s becomes another variable.

The traditional method of using simulated images has often involved looking at a series of simulated images for different values of Δf and t and finding the best match with your experimental image. Clearly this is not the ideal approach! Remember that the interpretation of HRTEM images may not be straightforward or unique. We must next compare the simulated images with those generated experimentally. This is the subject of the next chapter and is the basis of quantitative HRTEM.

GENERAL

- Buseck, PR, Cowley, JM and Eyring, L, Eds. 1988 *High-Resolution Transmission Electron Microscopy and Associated Techniques* Oxford University Press New York. An excellent resource covering more than HRTEM.
- Horiuchi, S 1994 *Fundamentals of High-Resolution Electron Microscopy* North-Holland, New York. An ideal complement to Spence's book.
- Kihlborg, L, Ed. 1979 *Direct Imaging of Atoms in Crystals and Molecules* Nobel Symposium 47 The Royal Swedish Academy of Sciences Stockholm. A classic collection of papers on HRTEM.
- Krakow, W and O'Keefe, M, Eds. 1989 *Computer Simulation of Electron Microscope Diffraction and Images* TMS, Warrendale, PA. A collection of focused review articles.
- O'Keefe, MA and Kilaas, R 1988 in *Advances in High-Resolution Image Simulation in Image and Signal Processing in Electron Microscopy*, Scanning Microscopy Supplement 2. (Eds. PW Hawkes, FP Ottensmeyer, WO Saxton and A Rosenfeld) p. 225 SEM Inc. AMF O'Hare IL. This article is another excellent introduction to the subject.
- Spence, JCH 2003 *High-Resolution Electron Microscopy* 3rd edition Oxford University Press New York. This is *the* text for users of the HRTEM.
- Zuo, JM and Spence, JCH 1992 *Electron Microdiffraction* Springer, NY

SCATTERING FACTORS

- Carlson, TA, Lu TT, Tucker, TC, Nestor, CW and Malik, FB 1970 *Report ORNL-4614*, ORNL Oak Ridge TN. Tabulated values of scattering factors.
- Doyle, PA and Turner, PS 1968 *Relativistic Hartree-Fock X-ray and Electron Scattering Factors* Acta Cryst. **A24** 390–7.
- Tang, D. and Dorignac, D 1994 *The Calculation of Scattering Factors in HREM Image Simulation* Acta Cryst. **A50** 45–52. Detailed comparisons of scattering factors.

QUASICRYSTALS

- Beeli, C and Horiuchi, S 1994 *The Structure and its Reconstruction in the Decagonal $Al_{70}Mn_{17}Pd_{13}$ Quasicrystal* Phil. Mag. **B70** 215–40. Calculations of HRTEM images.
- Shoemaker, CB 1993 *On the Relationship between μ - $MnAl_{4,12}$ and the Decagonal Mn-Al Phase* Phil. Mag. **B67** 869–81. A review of several models.

MODELS

- Clarke, DR 1979 *On The Detection of Thin Intergranular Films by Electron Microscopy* Ultramicrosc. **4** 33–44.
- Coene, W and Van Dyck, D 1984 *The Real Space Method for Dynamical Electron Diffraction Calculations in High Resolution Electron Microscopy: II. Critical Analysis of the Dependency on the Input Parameters* Ultramicrosc. **15** 41–50.
- Coene, W and Van Dyck, D 1984 *The Real Space Method for Dynamical Electron Diffraction Calculations in High Resolution Electron Microscopy: I. Principles of the Method* Ultramicrosc. **15** 29–40.
- Rasmussen, DR and Carter, CB 1990 *On the Fresnel-fringe Technique for the Analysis of Interfacial Films* Ultramicrosc. **32** 337–348.
- Ross, FM and Stobbs, WM 1991 *A Study of the Initial Stages of the Oxidation of Silicon Using the Fresnel Method* Phil. Mag. **A63** 1–36.
- Ross, FM and Stobbs, WM 1991 *Computer Modelling for Fresnel Contrast Analysis* Phil. Mag. **A63** 37–70.
- Simpson, YK, Carter, CB, Morissey, KJ, Angelini, P and Bentley, J 1986 *The Identification of Thin Amorphous Films at Grain-Boundaries in Al_2O_3* J. Mater. Sci. **21** 2689–96.

IMAGE SIMULATION

- Barry, J 1992 in *Electron Diffraction Techniques 1* (Ed. JM Cowley) p. 170 I.U.Cr. Oxford Science Publication New York.
- Coene, W, Van Dyck, D, Van Tendeloo, G and Van Landuyt, J 1985 *Computer Simulation of High-Energy Electron Scattering by Non-Periodic Objects. The Real Space Patching Method as an Alternative to the Periodic Continuation Technique* Phil. Mag. **52** 127–143. Real-space patching for Section 30.12.
- Cowley, JM and Moodie, AF 1957 *The Scattering of Electrons by Atoms and Crystals. I. A New Theoretical Approach* Acta Cryst. **10** 609–19.
- Kambe, K 1982 *Visualization of Bloch Waves of High Energy Electrons in High Resolution Electron Microscopy* Ultramicrosc. **10** 223–227.
- Kirkland EJ 1998 *Advanced Computing in Electron Microscopy* Springer, NY

Stadelmann, PA 1979. *EMS - a Software Package for Electron Diffraction Analysis and HREM Image Simulation in Materials Science Ultramicrosc.* **21** 131–145. The original description of EMS.

SOME PHYSICS

- Doyle, PA and Cowley, JM 1974 in *International Tables for X-ray Crystallography*, Vol. **IV**, pp. 152–173, Kluwer Academic Publ. Dordrecht Netherlands.
- Fujimoto, F 1978 *Periodicity of Crystal Structure Images in Electron Microscopy with Crystal Thickness* Phys. stat. sol. (a) **45** 99–106.
- Jiang, J-C, Hovmöller, S and Zou, X-D 1995 *A Three-dimensional Structure Model of Eight-Fold Quasicrystals Obtained by High-Resolution Electron Microscopy* Phil. Mag. Lett. **71** 123–129.
- O’Keefe, M and Spence, JCH 1994 *On the Average Coulomb Potential (Φ_0) and Constraints on the Electron Density in Crystals*. Acta Cryst. **A50**(1) 33–45.
- Self, PG and O’Keefe, MA 1988 in *High-Resolution Transmission Electron Microscopy and Associated Techniques* (Eds. PR Buseck, JM Cowley and L Eyring) p. 244 Oxford University Press New York.
- Taftø, J, Jones, RH and Heald, SM 1986 *Transmission Electron Microscopy of Interfaces Utilizing Mean Inner Potential Differences between Materials* J Appl. Phys. **60** 4316–8.
- Wilson, AR and Spargo, AEC 1982 *Calculation of the Scattering from Defects Using Periodic Continuation Methods* Phil. Mag. **A46** 435–49.
- Zhu, Y, Taftø, J, Lewis, LH and Welch, DO 1995 *Electron Microscopy of Grain Boundaries. An Application to RE-Fe-B (RE = Pr or Nd) Magnetic Materials* Phil. Mag. Lett. **71** 297–306.

THE COMPANION TEXT

The companion text extends the topic of simulation in chapters on EMS (jEMS is the java version that Pierre Stadleman supports) and on simulating diffraction contrast images. One idea which you might examine immediately is the use of Mathematica or similar packages to graph the equations.

SELF-ASSESSMENT QUESTIONS

- Q30.1 List the four principal methods of calculating multiple images.
- Q30.2 What is a phase grating?
- Q30.3 Why do we sample all of reciprocal space, not just the Bragg beams?
- Q30.4 Write down an equation relating ψ_{n+1} and ψ using a propagator and a phase grating.
- Q30.5 Why are we using Fresnel diffraction in this multislice treatment rather than Fraunhofer?
- Q30.6 We use convolution to describe what sort of scattering and why?
- Q30.7 Why must $Q(\mathbf{k})$ take account of nearly twice as many k values as $\psi(\mathbf{k})$ or $P(\mathbf{k})$?
- Q30.8 Write down an equation for $\psi_{n+1}(\mathbf{k})$ used in the FFT method.
- Q30.9 Extra challenge. Why is it FFT not just FT?
- Q30.10 Write down an equation for $\psi_{n+1}(\mathbf{r})$ in the Coene-Van Dyck method.
- Q30.11 Why do we use ‘ ’ when describing Kambé’s ‘simple’ example?
- Q30.12 In Kambé’s ‘simple’ example, we consider the case where only three Bloch waves are significant. Why choose just three, not more (or fewer)?
- Q30.13 We mention the direct connection between WPOA and Bloch wave propagation. Why is this?
- Q30.14 Why is it relevant that λ is finite (not zero)?
- Q30.15 Why does beam convergence alter a HRTEM image according to the Ewald construction?
- Q30.16 Why will GB grooving influence HRTEM images?
- Q30.17 What is the periodic continuation method?
- Q30.18 What is the problem in imaging crystals with 5- and/or 10-fold symmetry?
- Q30.19 How do most multislice images simulation programs take care of ionic and covalent bonding?
- Q30.20 Why do many researchers prefer a through-focus series to just a focal series of images?

TEXT-SPECIFIC QUESTIONS

- T30.1 Assuming that the first Fresnel fringe is $\delta = (\lambda\Delta f)^{1/2}$ away from the edge of the sample, calculate Δf for Figure 9.21B.
- T30.2 How well does the relationship $\delta = (\lambda\Delta f)^{1/2}$ for the position of the first Fresnel fringe predict the images in Figure 30.5C assuming the structure given in Figure 30.5A?
- T30.3 Construct a unit cell to use in HRTEM simulation of the $\Sigma = 3$ 111 twin boundary in Si.
- T30.4 Construct a unit cell to use in HRTEM simulation of the $\Sigma = 3$ 112 twin boundary in GaP. Discuss the information you would obtain for an HRTEM giving point (structural) resolution of 0.3, 0.2, and 0.1 nm.
- T30.5 Show the smallest reproducible unit for Figure 30.11 given the symmetry shown in the figure.
- T30.6 Plot equation 30.14 for different values of Δf . How does this plot change as we change the kV of the microscope?

- T30.7 By playing with equation 30.15 deduce equations 30.19 and 30.20.
- T30.8 Deduce equation 30.21 using equations 30.18 and 30.20.
- T30.9 Explain why the 'images' from Bloch wave 1 differs from those in Bloch waves 2 and 3, as shown in Figure 30.3.
- T30.10 Index the DP in Figure 30.4 and thus determine the radius (in nm^{-1}) of the objective aperture.

Processing and Quantifying Images

CHAPTER PREVIEW

In this chapter we will equate processing with the use of the computer to analyze our data. We will simply use image processing to extract more information from the data than we can obtain by eye. The data will generally be an HRTEM image but could be other images or DPs. We'll quantify spectra after we describe them in Part 4 of this text. In the past, the optical bench was also used for this purpose, but the number of optical benches is negligible compared to the number of computers now found in every TEM lab. Optical benches did allow us to form DPs which we could then modify to produce a processed image. This analog approach has now largely been replaced by its digital counterpart. The computer can be much cheaper than the optical bench and is far more flexible. The number of software packages which are designed for, or can easily be adapted to, TEM is also growing.

We can use image processing to produce a clearer view of the image, for example, by subtracting unwanted background detail, correcting for noise or drift, or removing artifacts. The big warning, though, is that, when removing one artifact, you must be very careful not to introduce others.

Although it's nice to see information more clearly, the unique feature of the computer approach is that we can *quantify* the data in any image and then normalize these data. Now we can directly compare the quantified experimental image with computer-simulated images. Although throughout this chapter we will be concerned with HRTEM images, most of what we say can be transferred directly to the analysis of diffraction-contrast images.

The other general point is that the ideas we'll discuss are also applicable to images derived from different sources. Once the data are in the computer, i.e., in digital form, the source becomes unimportant as far as processing possibilities are concerned. Examples of 'images' which might be obtained from the TEM include X-ray or EELS maps, STEM images, TEM images, and CBED or BSE patterns.

Most of the discussion will concern the use of computers. All you need to know is how best to get the data into the computer, how to process it, what to do with the data, how to display the result, and how to report what you've done. Most importantly, when we have the image in a numerical form we can start to consider statistics, which means much more than adding 'error' bars.

31.1 WHAT IS IMAGE PROCESSING?

Image processing is essentially manipulating images. The topic arises in many fields so we need to understand the words/jargon; we'll discuss the language of image processing as it is applied to TEM.

Image processing is not only becoming more common, but is also finding new applications in many fields. Faster, more powerful computers and increased memory storage are making tasks possible which could not previously have been considered. As a result of this increasing user base, there are now

many software packages available which can be used in microscopy; we listed some in Section 1.5. These range from programs used widely in desktop publishing to those which have been custom designed for EM. The goals of image processing include that of quantitative microscopy. You must choose among the different packages, commercial and freeware, and match them to the computer available in your lab. One point to remember is that some very simple optical methods which don't rely on a computer can often be very helpful. The other point is that the eye is hard to beat.

THE BASIC IDEA

Change images into numbers and manipulate the numbers.

There are many specialized books on this topic for the beginner or the expert; a selection is given in the references. The purpose of this chapter is to give a generalized overview. One problem in discussing this topic is that it is a very rapidly changing field. We will try to avoid specifics concerning particular programs but will mention these programs at the end of this chapter.

31.2 PROCESSING AND QUANTIFYING IMAGES

We process images primarily for two reasons

- We may want to improve the appearance of an image, make it look sharper, more even in contrast, higher contrast, etc. This processing can be unwise.
- We may want to quantify the information contained in the image. This processing may be essential: many physicists can read equations but not images.

Processing for improving the appearance of TEM images has been practiced for many years using such photographic techniques as ‘dodging,’ using ‘filters,’ selecting different emulsions or varying the developer, etc. It is only recently that relatively powerful personal computers have become widely available, but the term ‘image processing’ almost automatically implies the use of computers. Computer image processing will be the emphasis of our discussion. We have three requirements.

- We must be able to create a digital form of the image in the computer.
- We need appropriate software for processing the image.
- We need a computer which can perform the processing in an acceptable period of time with the required resolution.

THE DIFFICULTY

When image processing, the difficulty may be describing what you have done.

Many comments here are similar to those we made in discussing the microscope itself. For example, you may have to work with the built-in system or the system that’s already available in the lab. The difference is that some of the free programs are extremely powerful, so that all you need is the desktop computer. Many programs designed for desktop publishing are relatively inexpensive. Thus, you can almost always find a way to extend your processing capabilities.

The motivation is that we need to obtain more information from an image than we can get by just looking at it. This principle applies to more than HRTEM; we are discussing it here because HRTEM is where at present it is most needed/used in TEM. However, any TEM, X-ray map, or energy-filtered image or DP may benefit from processing and quantification. We need to quantify the TEM parameters, in particular C_s . One unique aspect of image processing in the TEM is that we have a choice between on-line and off-line processing. In fact, we often use on-line processing (frame averaging and background subtraction on the video image) to see the image even though the image we record may not be processed. However, it may indeed already be processed—so you should know and report this.

31.3 A CAUTIONARY NOTE

For most of our discussion we will consider only processing techniques using computers. To a large extent we can simulate a TEM using the computer. As we saw in Chapter 30, we can model a crystal, insert apertures, define the electron beam, including its broadening in the specimen, and then calculate the image. What we do in image processing is start with the image, add apertures and special filters, and then create a new image, the processed image. This image is a *real* image. What we must be careful about is explaining just what processing procedures we have used, since these may affect the interpretation of the data. Reporting the processing procedure is particularly critical when the raw data (the ‘original’ image) is not being reported at the same time.

ALWAYS

Report how you have treated your image, so that the reader can compare your data with related data that may have been processed differently or not at all. Be very careful. One honest (or dishonest) mistake may change your career!

31.4 IMAGE INPUT

There are several methods you can use to put the TEM image on the computer. The choice depends in part on how much detail you want in your digitized image, but also depends on how much work you’re prepared to do. In this discussion, we’ll only consider images which you have looked at on a video monitor, a computer display, or the fluorescent screen. Your basic choices are

- Transfer the image directly from the TEM to the computer.
- Record the image on film, then digitize it using a microdensitometer.

- Record the image on videotape.
- Record the image on film, then print it and use a flat-bed (or other) scanner.

There are many methods for creating a digital form of an image in your computer. The simplest is to use a slow-scan CCD camera, which we discussed in Chapter 7. The drawback of CCD cameras is that high-quality CCD chips are very expensive for $2\text{ k} \times 2\text{ k}$ arrays and astronomically expensive for $4\text{ k} \times 4\text{ k}$ arrays (but they are getting cheaper). Such cameras may become routine additions for all TEMs in the future. You will probably also want to use video. With film, you can record a larger area than you can using a CCD unless you have the $4\text{ k} \times 4\text{ k}$ version. You should use a digital video-recorder for in situ studies when using a heating or straining holder.

We can transfer the image from a videotape or a video camera to the computer using a frame grabber. Frame-grabber boards are readily available for most computers. You can use a high-resolution scanner for photographs or negatives. At this time, scanners cost about the same as a digital camera with comparable resolution. The purist's approach is to use a microdensitometer to measure the intensity of the film point by point and read this directly into the computer. The advantage of the microdensitometer is that it is very precise and can achieve the highest resolution for a very large area. The main problem is that it is slow, being a serial-collection technique. If you use it to its best advantage, your image will require a large amount of computer memory, which in itself is not a problem, but manipulating such images will still be slow.

31.5 PROCESSING TECHNIQUES

31.5.A Fourier Filtering and Reconstruction

The principle involved in filtering is that a mask is used to remove some information from an image in order to enhance or emphasize other information. As an extra complication we can process the image, e.g., Fourier transform a HRTEM image, then apply a mask, and then reverse the processing.

We can vary the size of the apertures and the sharpness of their edges, something you can't do in a TEM with normal fixed-diameter objective apertures. A single variable SAD aperture was used on some early TEMs; it was triangular or square in shape and used three or two movable blades. You can best understand the procedure by an example. A square mask was used to select the region in Figure 31.1A from a much larger region of the HRTEM image, and its Fourier transform (i.e., the DP effectively from a few nanometers) is shown in Figure 31.1B.

THE SUPPORT FILM IS IMPORTANT

You will often see similar diffractograms obtained using a film of amorphous carbon. While such films are easier to make, they give little diffracted intensity for the range of u values between 6 and 8.5 nm^{-1} , which is important in HRTEM.

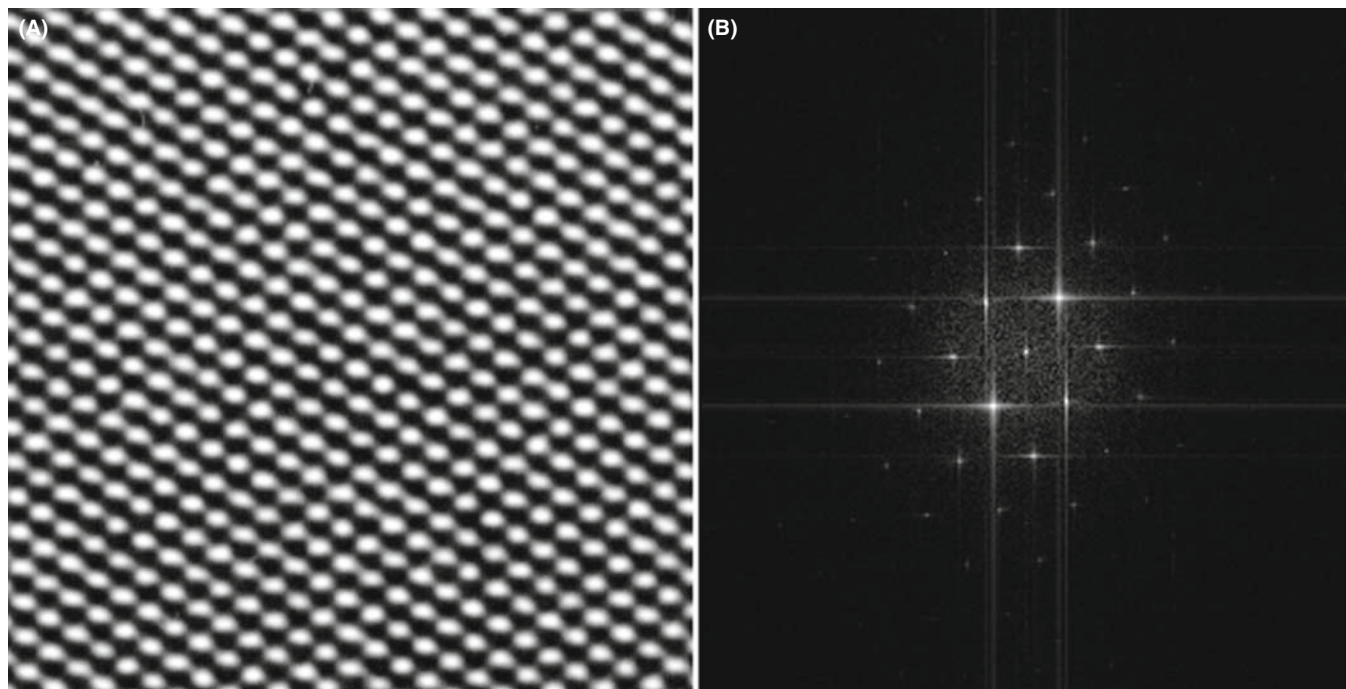


FIGURE 31.1. A square mask has been used to select the area shown in (A) from a much larger print of the image. The Fourier transform of this region is shown in (B) where you can see not only the spots in the 110 DP but also long streaks (artifacts of the processing) that run normal to the edges of the mask.

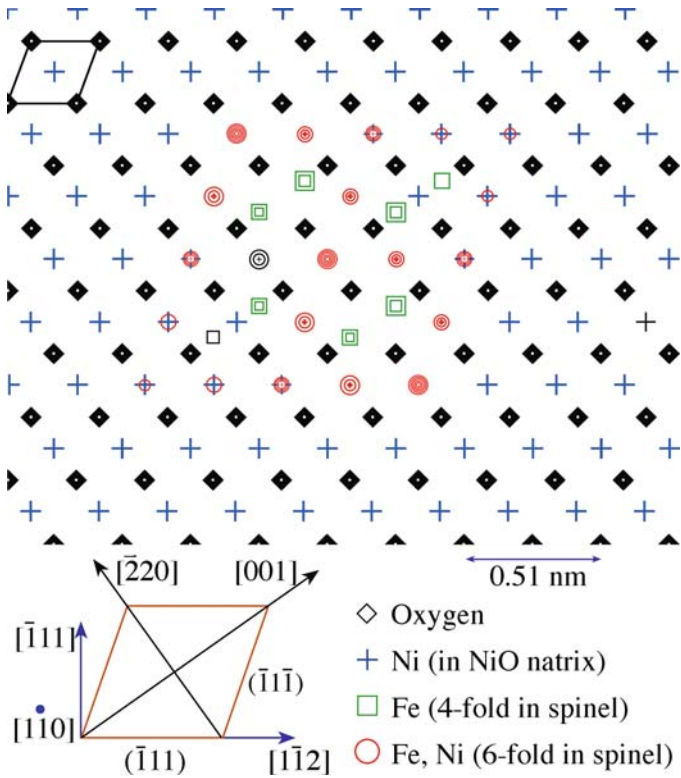


FIGURE 31.2. A model of an octahedron of spinel fully enclosed within a matrix of NiO. The rest of the specimen could then be modeled by adding extra layers of NiO above or below the defect layer.

What this technique does is to allow you to do microscopy in the computer. Your image becomes the specimen. You form the DP, then you can use apertures to select one or more beams to form the image; these apertures are the computer version of the objective aperture. Small apertures limit resolution just as in the 'real' TEM because the information about anything other than the perfect lattice is carried between the reciprocal lattice points. Figure 31.2 gives an illustration of how a model can be constructed of a particle in a matrix, which can be useful when simulating HRTEM and conventional BF/DF images in the computer, as shown in Figure 31.3. This model was made using the Digital Micrograph package (see Section 1.6).

31.5.B Analyzing Diffractograms

In Chapter 28 we showed that the transfer function could be plotted out as shown schematically in, e.g., Figure 28.4. Another way of thinking about this plot is to imagine what would happen if we have a specimen which generated equally every possible value of \mathbf{u} , i.e., every possible spatial frequency.

An amorphous film of Ge can provide just such a plot, but it is difficult to record the result because the scattered intensity is low.

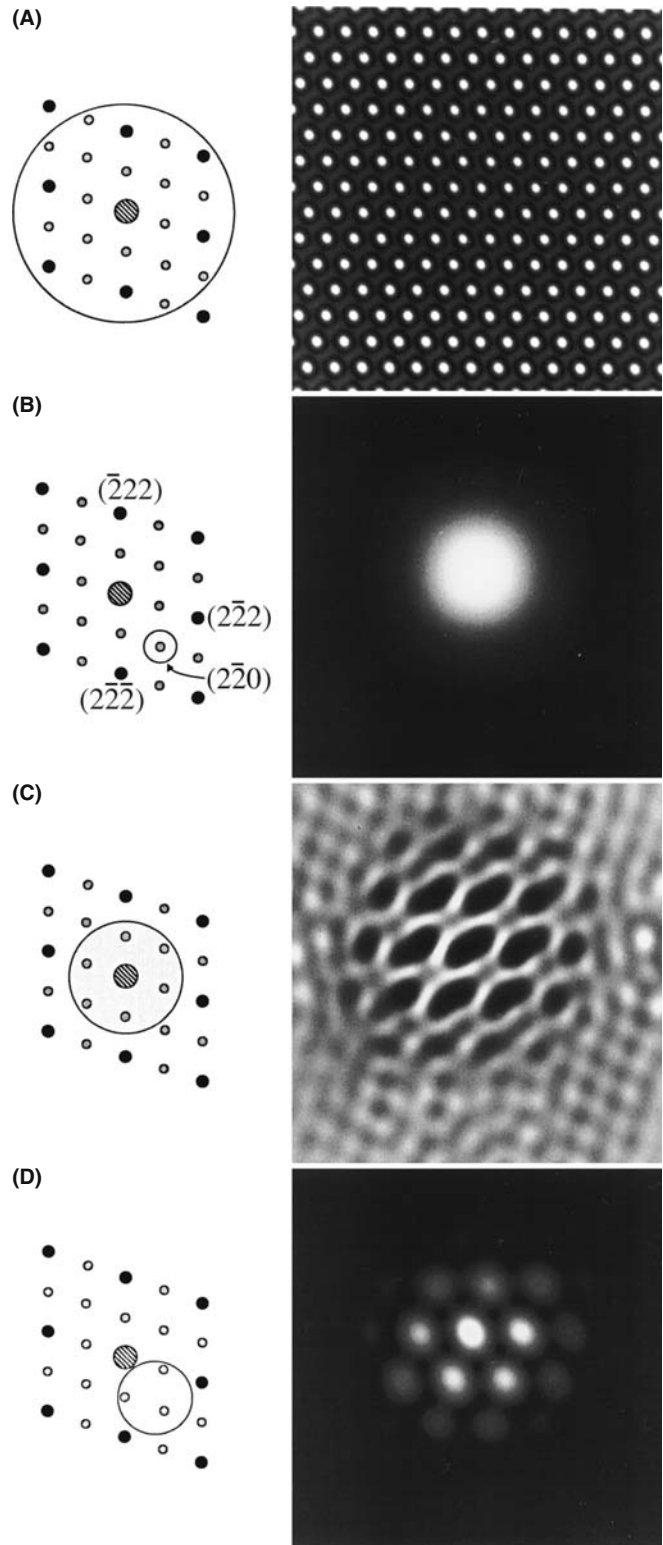


FIGURE 31.3. The DP from a model specimen containing a particle such as that shown in Figure 31.2 is shown schematically in (A) together with the resulting lattice image. The other three pairs of diagrams illustrate how we can use the computer to produce different masks and thus generate different images, such as the DF image in (B). The image in (C) corresponds to the image shown in Figure 28.26, while the DF lattice image shown in (D) is analogous to that discussed in Figures 28.14 and 28.15.

We therefore record the image at high resolution, preferably directly, using a slow-scan CCD camera, although digitizing the negative is fine. Then by comparing the experimental plot of I versus \mathbf{u} with those calculated for different values of Δf and C_s we can determine the astigmatism, the defocus, Δf , and the value of C_s (as we'll see below). It helps if you have a few particles of Au on the Ge film since the Au spots then give an internal calibration. Such a set of images and their corresponding diffractograms is shown in Figure 31.4. Notice that as the defocus of the objective lens increases the number of rings increase but they become narrower. The contrast transfer gradually extends to larger values of \mathbf{u} .

Determining astigmatism. you can use such diffractograms to correct the astigmatism, since a perfectly stigmatized image will give a DP with circular symmetry. As you can see in Figure 31.5, even a small amount of astigmatism can be detected by eye. The computer can readily measure and provide feedback to the TEM to correct the lens defect, as we will see shortly. This set of diffractograms also shows that the computer can distinguish astigmatism and drift in the image while the eye can easily mistake one for the other. Drift produces a circular pattern but the higher spatial frequencies are lost in the direction of drift.

Determining Δf and C_s . You can determine Δf for any image by measuring the radii of the bright and dark rings in the diffractogram, since bright rings correspond to $\sin \chi(\mathbf{u}) = 1$ and dark rings correspond to $\sin \chi(\mathbf{u}) = 0$.

$$\sin \chi(\mathbf{u}) = 1 \quad \text{when } \chi(\mathbf{u}) = \frac{n\pi}{2} \text{ and } n \text{ is odd} \quad (31.1)$$

$$\sin \chi(\mathbf{u}) = 0 \quad \text{when } \chi(\mathbf{u}) = \frac{n\pi}{2} \text{ and } n \text{ is even} \quad (31.2)$$

Since C_s will also influence the location of the rings, you need at least two rings. Krivanek has given a simple procedure for finding both C_s and Δf . If we start with our definition of χ

$$\chi(\mathbf{u}) = \pi\Delta f\lambda u^2 + \frac{1}{2}\pi C_s\lambda^3 u^4 \quad (31.3)$$

then, inserting the values given in equations 31.1 and 31.2 leads to

$$\frac{n}{u^2} = C_s\lambda^3 u^2 + 2\Delta f\lambda \quad (31.4)$$

All we now have to do is plot nu^{-2} versus u^2 to obtain a straight line with slope $C_s\lambda^3$ and with an intercept on the nu^{-2} axis of $2\Delta f\lambda$. Assign $n=1$ to the intensity maximum of the central bright ring, $n=2$ to the first dark ring, etc. The analysis can be trickier if you have used an underfocus condition or if you are very close to Scherzer

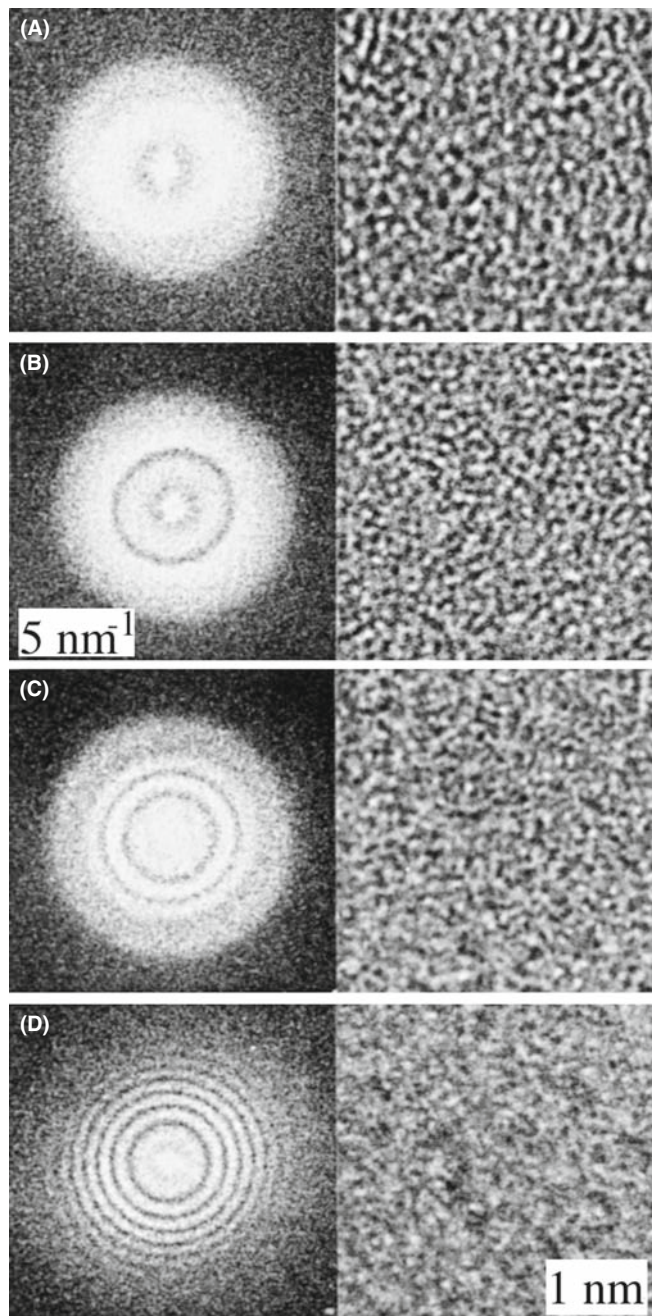


FIGURE 31.4. Four images of an amorphous Ge film and their corresponding diffractograms. Δf has the following values: (A) 1 sch; (B) 1.87 sch; (C) 2.35 sch; (D) 3.87 sch. Remember that 1 sch = $-(C_s\lambda)^{1/2}$.

defocus, but you will know when you have not found a straight line! Hopefully, you'll find that your value of C_s is close to that given by the manufacturer! A rather neat result is that if you plot nu^{-2} versus u^2 for different diffractograms (i.e., different values of Δf), then the points corresponding to each particular value of n will lie on a hyperbola, as shown in Figure 31.6A. You can use these hyperbola to determine C_s for any microscope and Δf for any diffractogram.

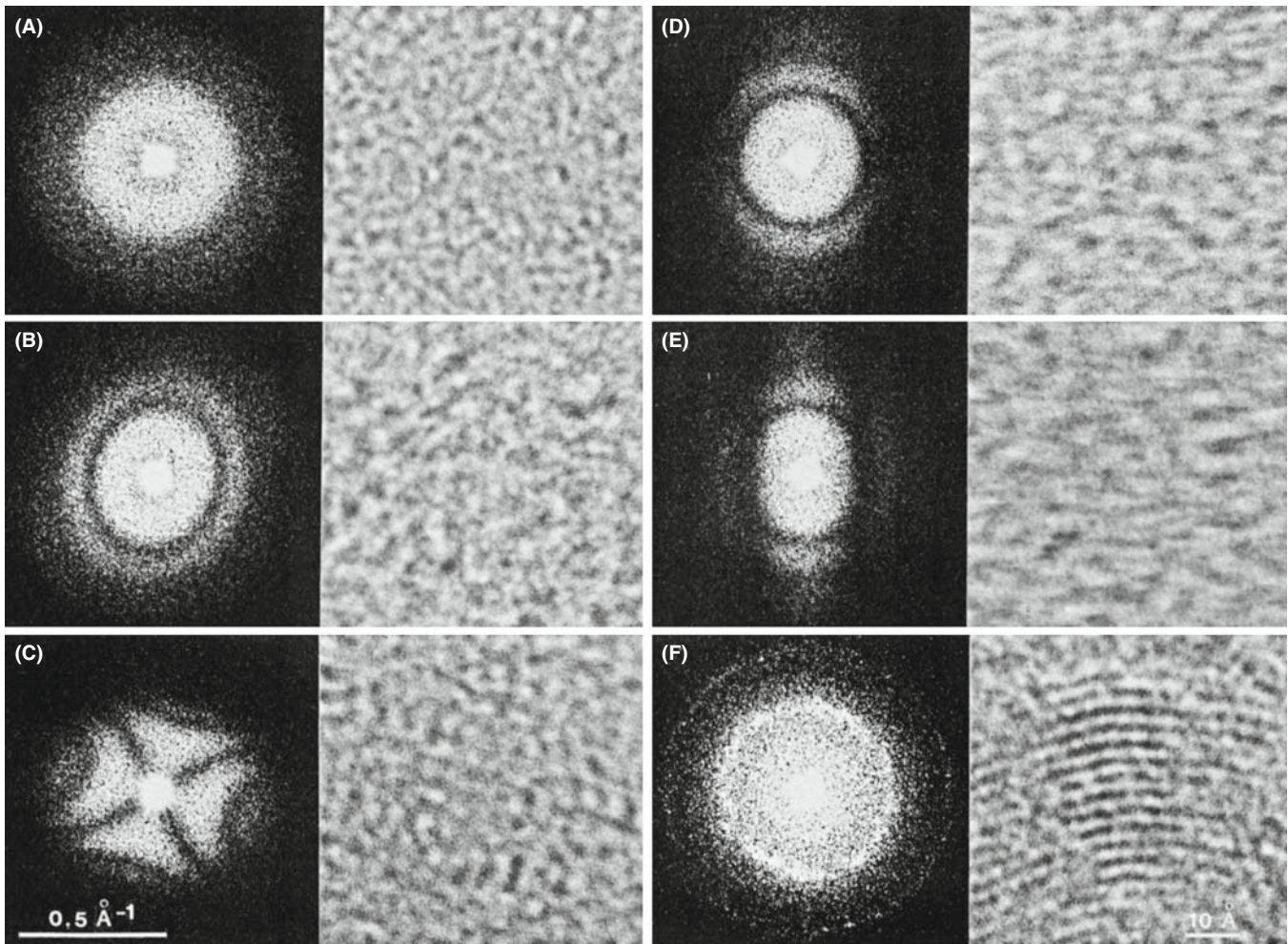


FIGURE 31.5. Six images of an amorphous carbon film and their corresponding diffractograms illustrating different misalignments of a 300-kV HRTEM. (A) Well aligned and no drift; (B) some astigmatism ($C_a = 14$ nm); (C) more astigmatism ($C_a = 80$ nm); (D) no astigmatism but drifted 0.3 nm; (E) no astigmatism but drifted 0.5 nm; (F) well aligned and no drift showing graphite calibration fringes of 0.344 nm spacing. (B,D,E) $\Delta f = 2.24$ sch; (F) $\Delta f = 0$.

THE ZEMLIN TABLEAU

This is a display of diffractograms recorded by systematically changing a parameter, such as tilt or defocus, as in Figure 31.5.

Diffractograms and beam tilt. Beam tilt is very difficult to correct by eye; even worse, it can cause the diffractogram to look astigmatic, so you correct the astigmatism instead. In the image, as we saw earlier, beam tilt can improve the appearance but confuse the interpretation! The set of diffractograms shown in Figure 31.6B shows you how to overcome the problem. You have to compare diffractograms taken at different beam tilts to determine the zero-tilt condition. A pair of diffractograms taken at $\pm\theta^\circ$ tilt will only look the same (though rotated) if the beam had zero tilt at $\theta=0^\circ$. In the example shown, the diffractograms above and below the horizontal line are similar, so θ_y was very close to zero for the central condition.

However, the pairs of diffractograms on opposite sides of the vertical axis differ slightly, so the alignment of θ_x was not perfect.

31.5.C Averaging Images and Other Techniques

If you have recorded a series of images using a video camera, for example, you can average them over several frames as your eye does automatically. The result of such a process is illustrated in Figure 31.7. Different methods can be used to average the images. The easiest approach appears to be as good as any and simply involves taking the unweighted average of your best images, i.e., in the video example, just average over a series of frames. If you know that the object you're studying has a certain symmetry, you can use that information to improve the image further. The article by Trus et al. will give you a start on this process. If you want to remove the blur due to motion of the image, then you will really need to delve much more into this subject.

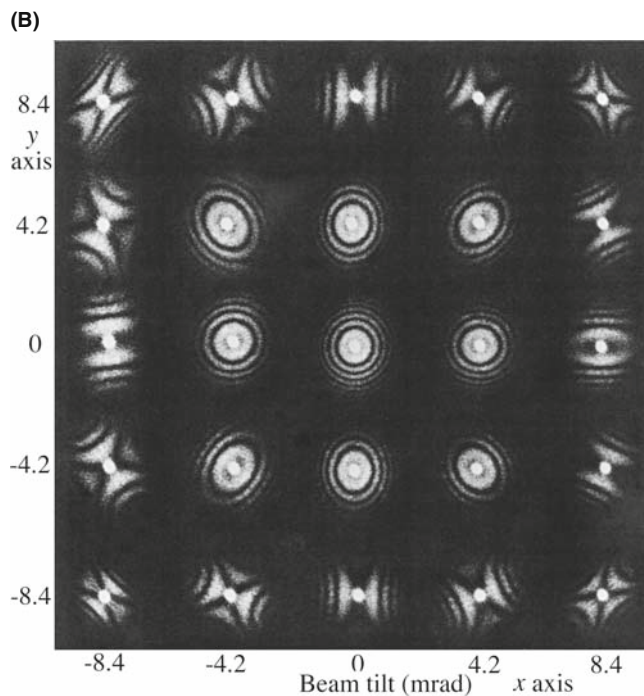
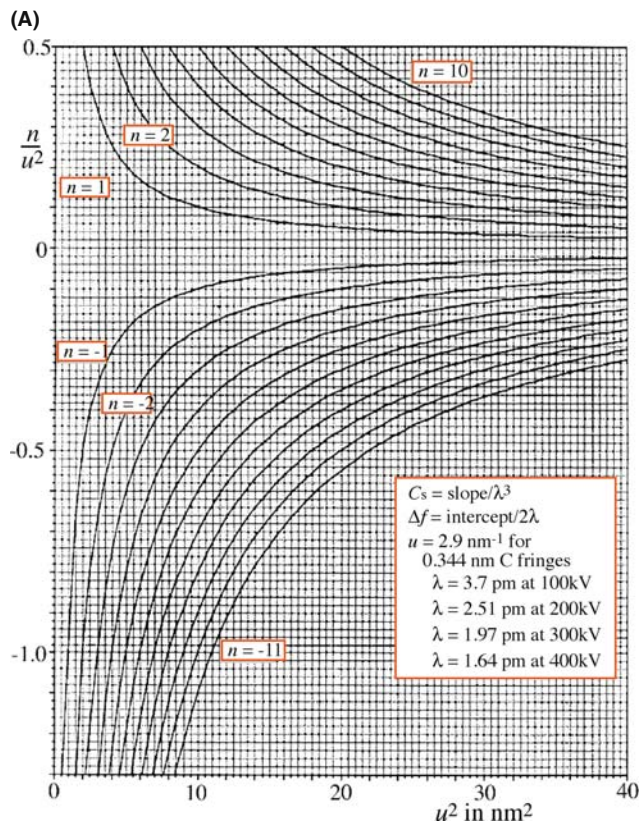


FIGURE 31.6. (A) Plot of nu^{-2} versus u^2 . The rings in any diffractogram correspond to a series of n values which allow you to draw straight lines on this figure and thus determine the slope and the ordinate intercept, giving C_s and Δf , respectively. (B) Set of diffractograms showing the effect of the incident beam tilt.

If you use a TV-rate video, you'll almost certainly use background subtraction routinely. For example, you can record an image of the honeycomb

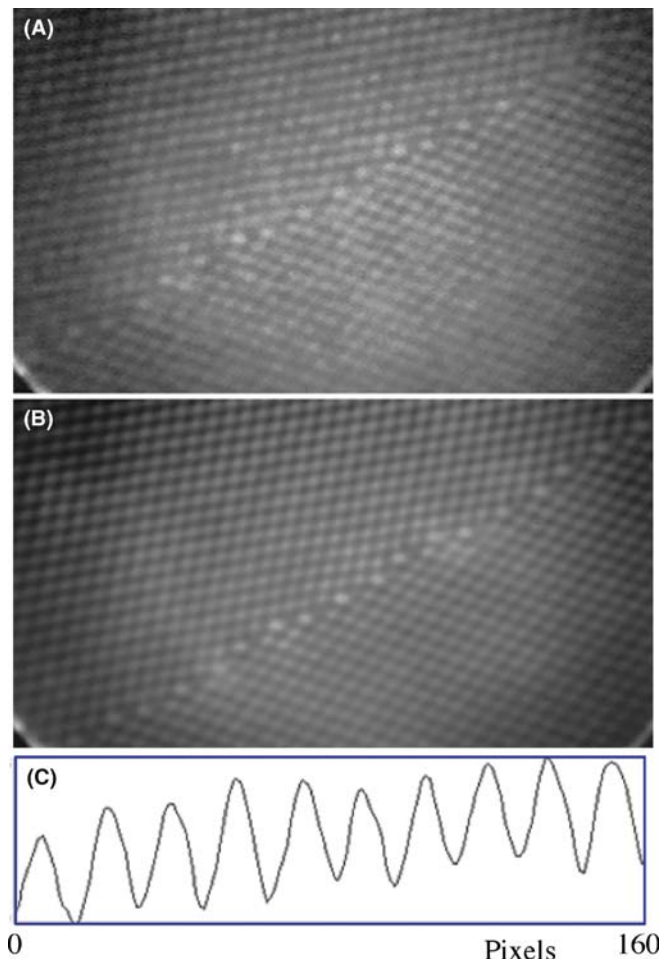


FIGURE 31.7. An example of the benefit of frame averaging to improve information from a video recording: (A) one frame; (B) 16 frames; (C) intensity profile along a (111) plane in (B).

pattern of the YAG detector, store it, and then automatically subtract it from all subsequent images in real time.

You may find it useful to add artificial color (pseudo-color) to your TEM images, as illustrated in Figure 31.8. Although it is often assumed that this is done just to make the images even more appealing to the non-microscopist (or non-scientist), there is actually a valid reason for the practice. Our eyes are much more sensitive to small variations in color than they are to small variations in gray level. You might therefore find color useful if you have a wide range of gray levels and want the viewer to be able to 'see' some subtle variations. Similarly, you can use color to emphasize a particular gray level in an image. However, you have to be very careful in your choice of look-up table (LUT), the table which relates each gray level to a particular color. To get a feel for the dangers, play with Photoshop™ and your favorite TEM image.

As we've already noted, in TEM, all of our apertures have relatively sharp edges but in the computer, you have the possibility of using multiple apertures, apertures with different shapes and apertures with diffuse

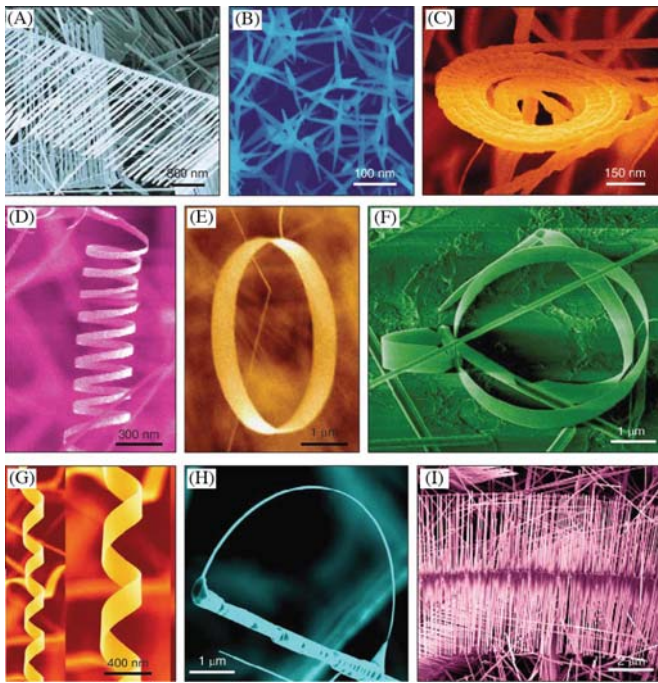


FIGURE 31.8. Illustrating the effectiveness of using false color.

edges. Apertures with diffuse edges will help eliminate the streaking which will otherwise be present (go back and look at Figure 31.1). You can also use the computer to do ‘unsharp masking,’ which is not the same as simply using a diffuse mask. The technique comes from the photographic process whereby we first print an image out of focus onto film, thus making a complementary image except where there is fine detail present in the original image; in digital processing this is called Laplacian filtering. Many more examples are given in Russ’s two books.

31.5.D Kernels

A kernel is simply an array of numbers which we can use to perform operations on a digital image. If we have the 3×3 kernel, K (we can have 5×5 , 7×7 , etc., but the computation time becomes too long, especially for real-time situations)

$$K = \begin{matrix} -1 & -1 & -1 \\ -1 & +8 & -1 \\ -1 & -1 & -1 \end{matrix} \quad K_o = \begin{matrix} A & B & C \\ D & E & F \\ G & H & I \end{matrix}$$

we can apply it to every 3×3 group of pixels in our image, e.g., K_o and put the result in a new digital image. If we call our new 3×3 image, K_i then

$$K_i = \begin{matrix} A' & B' & C' \\ D' & E' & F' \\ G' & H' & I' \end{matrix}$$

The new image will have, for example, $E' = 8E - A - D - G - B - H - C - F - I$. This kernel then gives us a digital Laplacian (an approximation to the second linear derivative, ∇^2). What this kernel is doing is subtracting the brightness value of each neighboring pixel from the center pixel. If the area is a uniform gray it will become white so changes in contrast will be exaggerated. We can design a wide range of kernel operators. For example, the edge enhancer kernel has the effect of digitally differentiating the image. (We’ll see a related digital-processing procedure applied to spectra in Chapters 35 and 39.) The Sobel and Kirsch operators are examples of such edge detectors; each can be thought of as the sum of several kernel operators. We can also use binary morphological operators which make binary features become larger or smaller. All of these operations can be carried out in any standard image-processing package. In general, you should be very careful when using such techniques in TEM; their value is in displaying data which might otherwise be missed, rather than helping you quantify an image.

REPORT IT

We’ve said it before. Check HREM papers in the literature. Have the images been processed? Are you told how? Can you reproduce the ‘experiment’? The last test is the essential component of science.

31.6 APPLICATIONS

This section will give you a taste of how image processing is being used now. It is just part of a rapidly growing list, so we are not going to be detailed or inclusive. We can separate the applications into two groups

- Noise reduction or improving the signal/noise ratio.
- Quantifying images.

Of course, the first topic is included in the second.

31.6.A Beam-Sensitive Materials

Low-dose microscopy necessarily implies that the signal-to-noise ratio will not be large; if it is large, the dose could have been smaller. This problem has been extensively addressed in biological EM and led to Klug’s Nobel prize for ‘Development of crystallographic electron microscopy and the structural elucidation of biologically important nucleic acid–protein complexes’ in 1982 (see the 1971 paper by Erickson and Klug). In materials science, we have tended to accept ‘beam damage’ as a fact of life, but this attitude will not be acceptable for future quantitative HRTEM. Most modern microscopes will allow you to perform all your

alignments on one area and then translate the beam to a predetermined distance in a predetermined direction before recording the image of a pristine area. Clearly, the CCD camera will not only let you see your image without waiting to develop the plates, but you can take a series of images for noise reduction purposes and/or assess whether the imaging conditions were what you had intended. (Sometimes, by taking a chronological sequence of images you can get a sense of the 'undamaged' image by extrapolating information back to zero time). The image shown in Figure 31.9 illustrates the possibilities. If you read the review by van Heel et al., you will get some idea of how far you can already go in this field.

BEAM DAMAGE

The electron beam will invariably change your specimen. Historically, we have worried about beam damage. Beam healing is just as bad for your image interpretation!

31.6.B Periodic Images

In discussing quantitative analysis, we have already noted how we can use the computer to identify similar features and combine them in order to reduce the noise. This technique has many possible variations. Again biological applications are leading the way with 3D crystallographic reconstruction, as illustrated by Downing and by Dorset and even correcting for distortions in the specimen, as shown by Saxton et al.

31.6.C Correcting Drift

Although drift is not as limiting on new machines, many older TEMs are still in use. Drift can be corrected now if the rate and direction of movement are constant. The computer can calculate the relative translation of two images and change the current in the image translation coils appropriately (which avoids moving the specimen). The difficulty is that the drift may not be linear. When implemented, such routines are already particularly valuable for frame averaging using a video camera. There will also be many applications for diffraction-contrast imaging as well as for X-ray or EELS analysis.

31.6.D Reconstructing the Phase

Although we are studying phase contrast, the image intensity doesn't directly give us phase information. Kirkland et al. showed as early as 1982 that the phase can be reconstructed by processing a defocus series. In their approach they use an iterative non-linear image-processing technique to reconstruct the complex

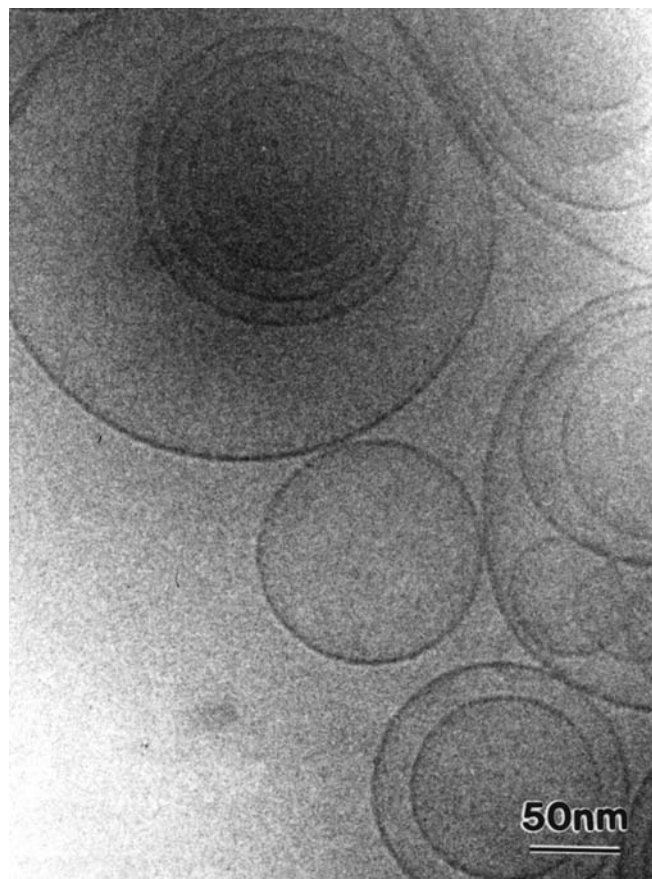


FIGURE 31.9. The image is from a highly beam-sensitive solution of surfactants in water. The solution has been frozen by plunging a film into liquid ethane and then transferring it to the TEM. The large circles are the surfactants that have aggregated to form vesicles; the concentration of the surfactants in the solution is just right for them to form lamellar structures (the circles seen flat on). Texture starts to appear in the image as soon as the beam interacts with the specimen due to crystalline ice formation and/or beam damage.

electron-transmission function. The technique was demonstrated using images of $\text{CuCl}_{16}\text{PC}$ (hexadeca-chlorophthalocyanine copper).

Five images from the experimental defocus series are shown in Figure 31.10A–E, together with the reconstructed transmission function plotted both as a real and imaginary part and then as an amplitude and phase part in Figure 31.10F–I. The projected structure of the known unit cell is also shown in Figure 31.10J. The phase image contains most of the structural information: it corresponds to the projected potential while the amplitude image contains features due to inelastic scattering. Notice in particular that we can now identify the benzene ring. This is one of the earliest published examples of the full-phase reconstructions. You must record such a series of defocus images if you want to do quantitative HRTEM. The companion text includes an extensive chapter giving details and an overview of how the process is practiced today.

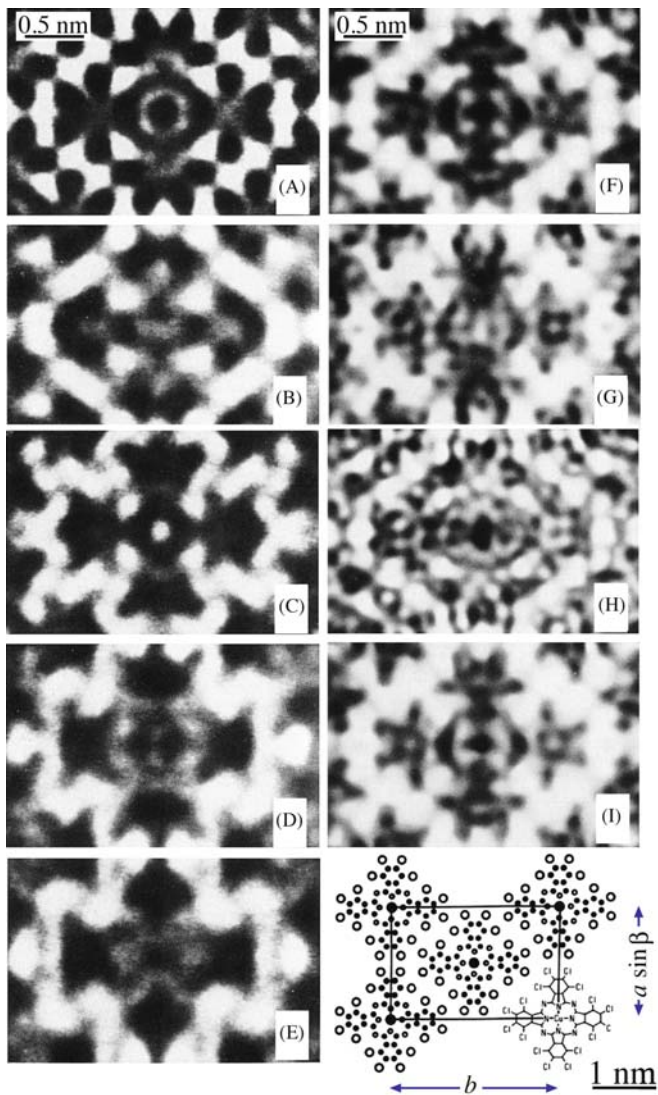


FIGURE 31.10. (A–E) Images from an experimental defocus series of CuCl_6PC ; the reconstructed transmission function plotted as both a real and imaginary part (F,G) and then as an amplitude and phase part (H,I); (J) the projected structure of the crystal.

31.6.E Diffraction Patterns

We have generally ignored the intensities in DPs because they are so strongly influenced by dynamical scattering. However, if the specimen is very thin, we can use the intensity in the SADP to carry out electron crystallography in the same way as in classic X-ray crystallography. As you can appreciate from Figure 31.11, particularly if the unit cell is large and the specimen examined is thin there is a great deal of information in the SADP but you can't get it all in one exposure. Hovmöller's group (see, e.g., Zou and the ELD program in Section 1.5) has provided a routine for analyzing such patterns and getting structure-factor information. For example

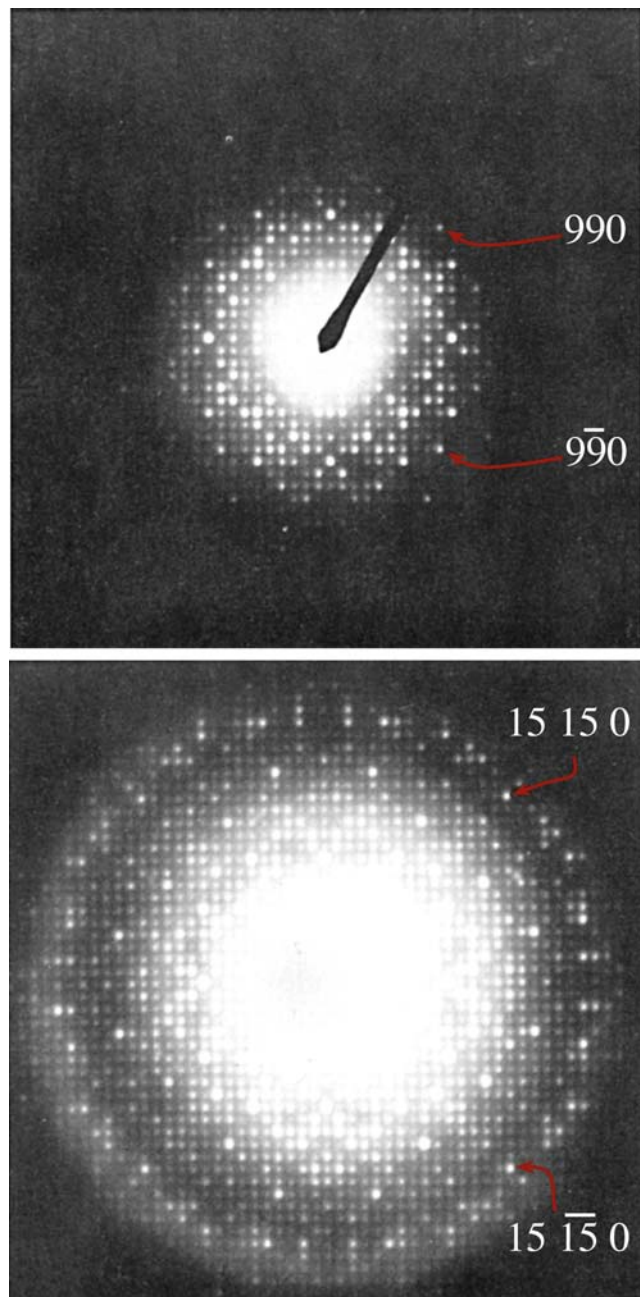


FIGURE 31.11. SADPs from $\text{K}_2\text{O} \cdot 7\text{Nb}_2\text{O}_5$ recorded using two different exposures. More than one exposure is needed to get all the information in the DP. The space group is P4bm and $a = b = 2.75$ nm. The (15, 15, 0) reflections correspond to spacing of 0.13 nm.

- Several patterns are recorded using exposure times of 0.5 s to >15 s.
- The patterns are digitized directly from the negatives using a CCD camera and a light-box for backlighting.
- The intensity of each film is calibrated using a calibration strip with 20 equal exposure steps.
- The intensity is measured for all the spots and the processing begins.

Thus far it is just like the classic analysis of XRD intensities. This digitization process is particularly

demanding, because each reflection typically covers an area of <0.5 mm diameter on the photographic film. You will need to be able to index three strong, but clear, reflections. The computer can then perform a series of functions

- Optimize the location of these points using a center-of-gravity approach, locate the origin, and index the rest of the pattern.
- Extract the intensities of each peak, taking care not to be misled by any shape effects of the specimen.
- Use reflections which are present on two successive negatives (since the intensities are now in digital form) to calibrate films recorded with different exposures and thus develop a very large dynamic range.

A cooled slow-scan CCD camera will give you a large dynamic range and better linearity than its room-temperature counterpart and should simplify this type of analysis. There are other complications in using electrons rather than X-rays for this kind of crystallography. While the Ewald sphere is still curved, as with X-rays, electrons can easily damage your specimen. However, the technique clearly has potential! Like all TEM techniques it can be applied to much smaller regions of the specimen than is possible for X-ray beams. We can also use the symmetry present in the SADPs. Because the specimen is very thin, this technique could be described as ‘kinematical’ crystallography and complements the ‘dynamical’ electron crystallography that we described for CBED patterns from thicker specimens in Chapter 21. The process of extracting intensities from DPs can now be carried out using the ELD software package (Section 1.5).

USING THE DP

The structure deduced from the HRTEM approach should generate the experimental SADP so it should be possible to use these diffraction data to further refine the structure.

If we can use the quantitative information available in DPs, we could combine this information with our experimental and simulated HRTEM images. The quantitative analysis of the DP is known as structure-factor-modulus restoration or reconstruction (Tang et al.). One limitation of this approach is that the specimen be sufficiently thin that diffraction is kinematical. Of course, this requirement is necessarily similar to the HRTEM requirement of the WPOA.

31.6.F Tilted-Beam Series

Having gone to great trouble to remove any beam tilt, we will mention how beam tilt can be used to extend the resolution of your microscope! The basic idea goes back to the tilted-beam lattice-fringe imaging we discussed in Section 23.3. Now you use a computer to combine information in different tilted-beam images. The method proposed by Kirkland et al. assumes that you know when the beam tilt is zero. You tilt the beam through different angles in well-defined directions so that you transfer information in overlapping regions of reciprocal space, as shown in Figure 31.12A; you also need the on-axis image, as shown in the tableau in

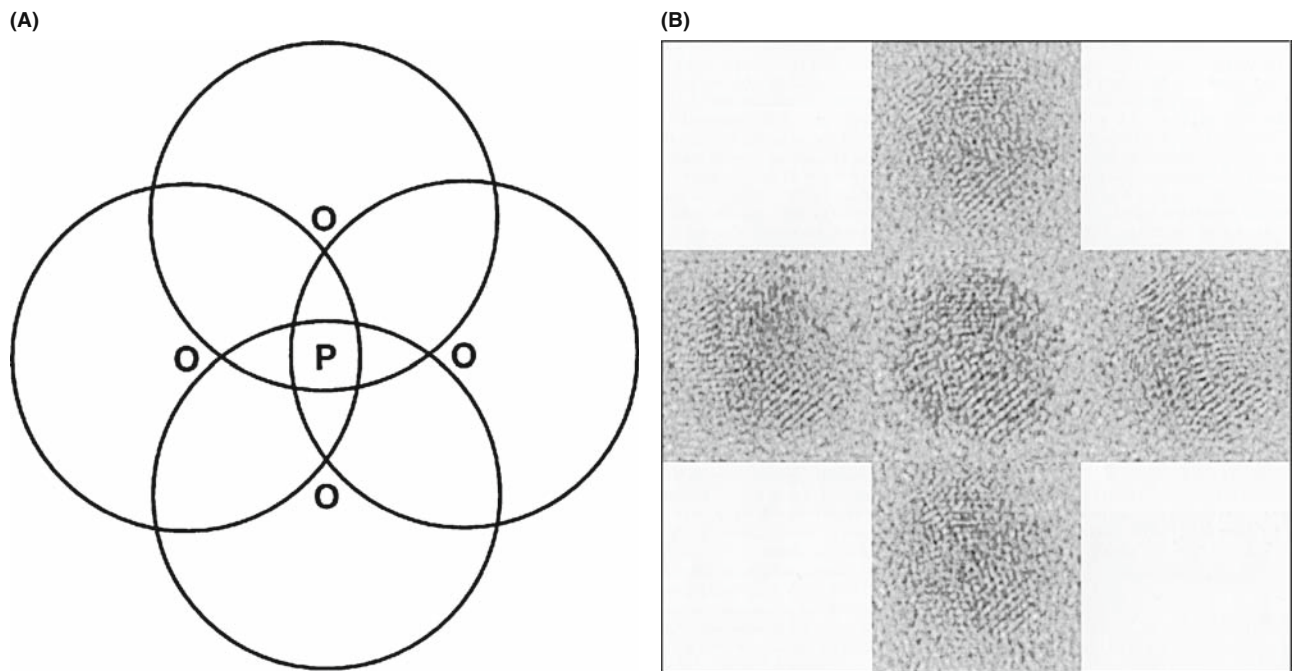


FIGURE 31.12. A method for extending the resolution of your TEM. Set the beam tilt to zero, then tilt the beam through different angles. (A) The four regions of Fourier space are shown by the four circles; O is each position of the tilted beam, P is the optic axis, and PO corresponds to the angle of tilt. (B) The five images used in the restoration arranged according to the beam tilt used in (A) with the on-axis image at the center.

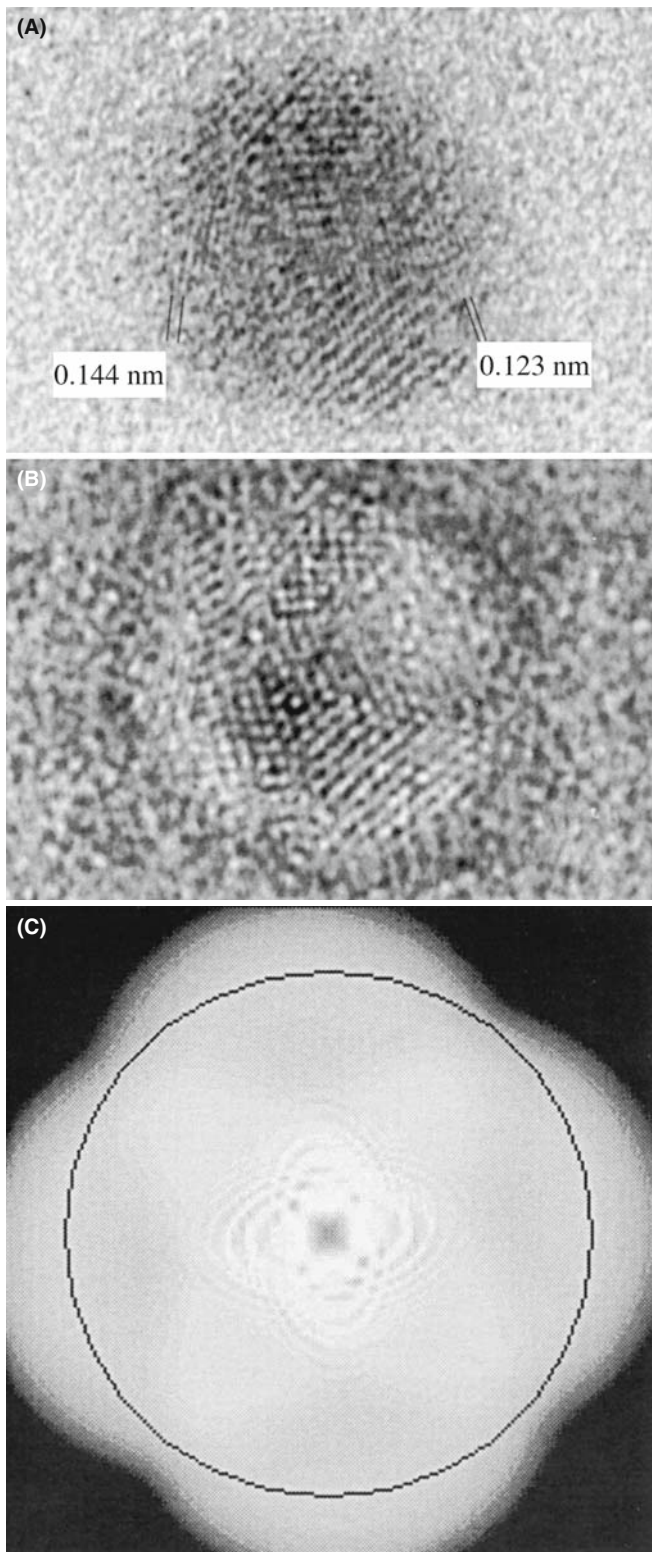


FIGURE 31.13. The restored image of a gold particle on amorphous Ge: (A) the amplitude (modulus) image showing 0.123-nm fringes, (B) the corresponding phase image, (C) the transfer function after restoration, plotted in two dimensions. The circle corresponds to 0.125-nm detail successfully transferred to the image. For such thin specimens, atom-dense positions have a reduced modulus [black in (A)] and an increased phase [white in (B)].

Figure 31.12B. Since it is important that the same area is imaged, a sixth (on-axis) image is recorded and correlated with the first (on-axis) image to check for drift and specimen degradation. You now need to restore the modulus and phase to create a higher-resolution image.

Kirkland's paper is a beautiful demonstration of the care needed in image processing. Even aligning the images is not trivial. However, the resulting restoration shown in Figure 31.13, demonstrates the potential of the technique: detail is present in the image at a resolution of 0.123 nm using a 400-kV microscope.

31.7 AUTOMATED ALIGNMENT

In the not-too-distant future, all TEMs will have automatic beam alignment, astigmatism correction, and readout for Δf . What makes this possible is the diffractogram analysis, a slow-scan CCD camera to digitize the image, and computer control of all the microscope functions. By microscope functions, we mean all lens currents, deflector currents, specimen drive, and aperture drives. The slow-scan camera is needed because the computer needs to make measurements on more than one ring in the diffractogram. So you will not actually have to sit in front of the microscope once you have loaded the specimen.

THE ADVANTAGE OF REMOTE TEM

The big advantage in remote control will not be that you can sit in Huntsville and operate a microscope in California but that you can locate the microscope in its controlled environment with no one opening the door to check if it is working (it was) or entering the room and thus changing the heat-load. If your specimen is not ideal or the TEM breaks down, you won't have to go for a walk on the beach in California but can continue word processing in Alabama.

Autoalignment is now available to everyone with a digital camera and the plug-in for Digital Micrograph (see Section 1.6). The early developments are summarized in the references. Your role is to select a suitable region of the specimen close to the area of interest; the area that you are interested in should ideally only be examined at low magnification. You will make the initial alignment manually and then ideally turn the process over to the computer. The computer will then adjust the astigmatism and correct the beam tilt independently and quickly.

Figure 31.14 shows how well and quickly this procedure can now be done. The different diffractograms in each tableau correspond to incremental changes in the beam tilt of 6 mrad in the x and y directions. The computer showed that the initial tilt error was 4 mrad, which

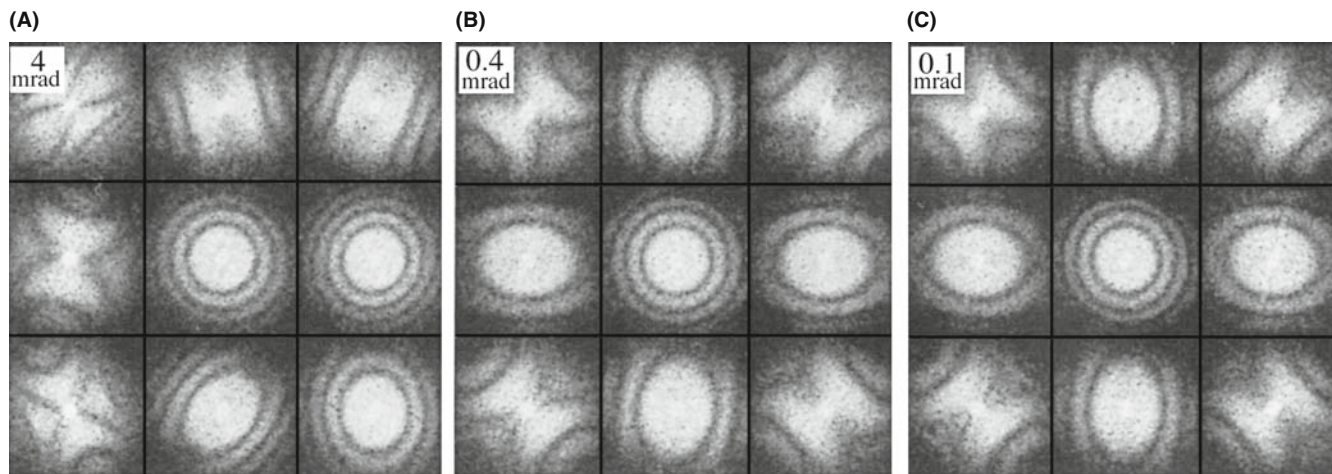


FIGURE 31.14. Using the computer to correct the beam tilt. (A) is the initial setting and (C) is the final. The incremental changes in the beam tilt are 6 mrad in the x and y directions away from the initial beam tilt in the (B). Note the central diffractogram is almost unchanged, emphasizing the need for computer-controlled tilting to give correct alignment. The misalignment in each setting is given in the upper left corner. (A) is the initial setting and (C) is the final.

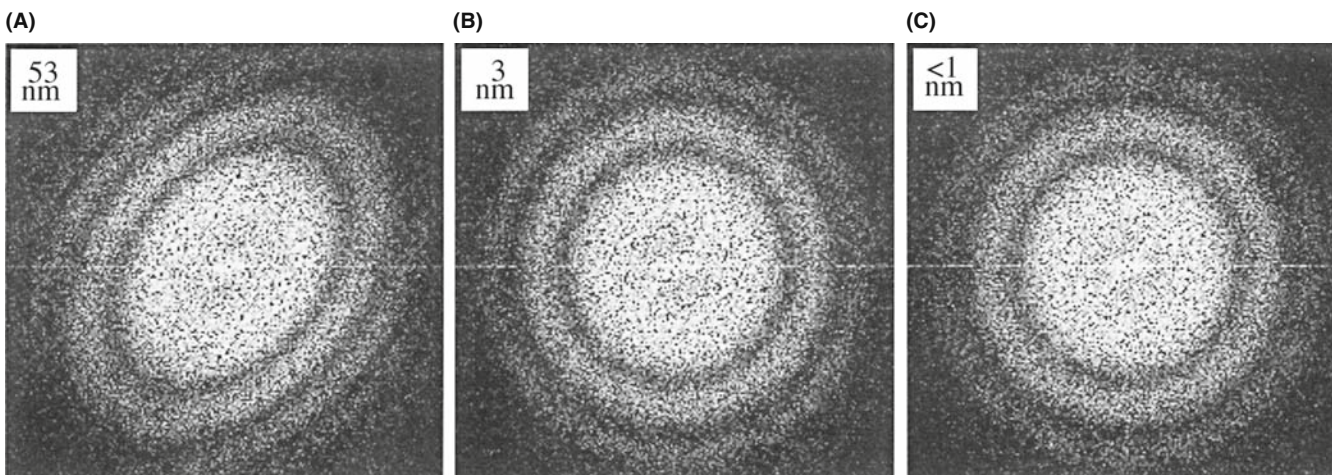


FIGURE 31.15. Diffraction patterns showing the astigmatism corrections made by the computer following a similar procedure to that shown in Figure 31.13. (A) is the initial setting and (C) is the final (and is not the same as (B)). (C) shows that the HRTEM is now well stigmatized. The misalignment in each setting is given in the upper left corner.

was reduced to 0.4 mrad after one pass and <0.1 mrad after the second pass. Each pass took only 28 s! The astigmatism shown in Figure 31.15 was initially 53 nm. It was reduced to 3 nm after one pass and to <1 nm after the second pass. For this correction, each pass took only 8 s. Even the most experienced operator can't match this speed or accuracy for either correction, and both corrections are now *quantitative*.

The defocus value is then found by calibrating the image with minimum contrast occurring at Δf_{MC} . The value of Δf_{Sch} can then be found when the image contrast is a maximum. Although the method described here uses the diffractogram, a corresponding approach can be followed by analyzing variations in the contrast of the image. This technique has been described by Saxton et al. and uses a method of cross-correlating

pairs of images recorded at each focus setting of the microscope. The reason for cross-correlating images is to remove the effects of electron shot noise; variations due to the photographic emulsion are avoided by using the slow-scan camera.

31.8 QUANTITATIVE METHODS OF IMAGE ANALYSIS

In the next six sections, we will go through several special topics and illustrations of image processing in HRTEM.

- Pattern recognition
- Parameterizing the image
- Chemical information from images

- Measuring the fit
- Quantitative comparison of simulation and experiment
- Fourier techniques

Our discussion will draw heavily on the work of a few pioneers in this field; we will also emphasize that, although this subject is still in its infancy, it is developing rapidly. The main cause for the delay in its application in materials science has been the lack of affordable fast computers and the feeling that everyone must write their own image-processing program; the latter is not true and is certainly not recommended. At this time, we can summarize the situation as follows

- Quantitative analysis is difficult, often tedious, and invariably time consuming.
- You must understand the basic ideas of image theory before you start processing images.
- Your analysis is only as good as your image and your image is only as good as your specimen.

We gave some information on software in Section 1.6 and expand on this in Section 31.16.

31.9 PATTERN RECOGNITION IN HRTEM

The most obvious feature of most HRTEM images is that we see patterns of white, gray, and black dots or other shapes. If the pattern is perfect everywhere, your specimen is probably a single crystal with no defects, no thickness variations, no variation in atomic composition, and no use. If it is not perfect, then we can use pattern recognition to quantify the variations.

PATTERN RECOGNITION

The idea is to take/make a template, move it across your image, and measure how closely the image resembles your template.

Clearly you need a computer for this! Your template needs to match the magnification and rotation of the pattern you are examining. Then you need a method to say how close your match is, i.e., you need to know your 'goodness of fit.' We will go through some basics here, but strongly recommend that you consult the list of original papers given at the end of this chapter when you are ready to apply this technique.

We can illustrate the approach following Paciornik et al. The large rectangle shown in Figure 31.16 represents your digitized image and could be $1\text{ k} \times 1\text{ k}$; remember the numbers indicate pixels. The small rectangle represents your template. This template might be a small area of the pattern or a simulated image, in which case it might be a 128×128 pixel template. If the template is taken from your image, then you have

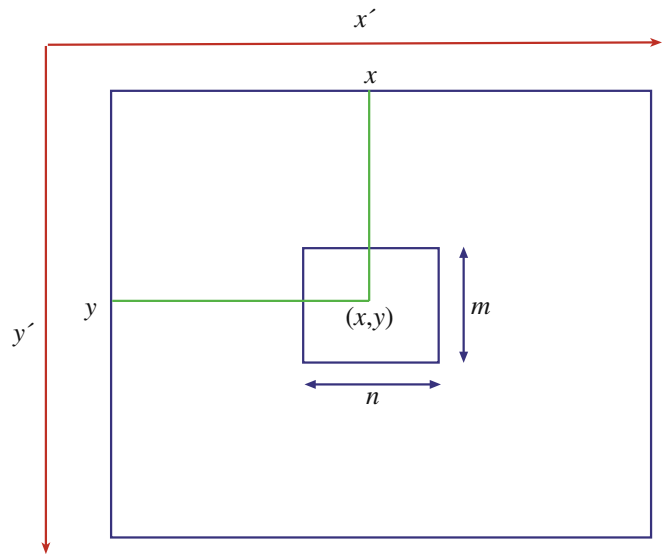


FIGURE 31.16. The large rectangle represents the digitized image, size $(x' \times y')$; the small rectangle, $(m \times n)$ pixels, represents the template used in the cross-correlation calculation. The small rectangle is moved to different (x, y) positions during the process.

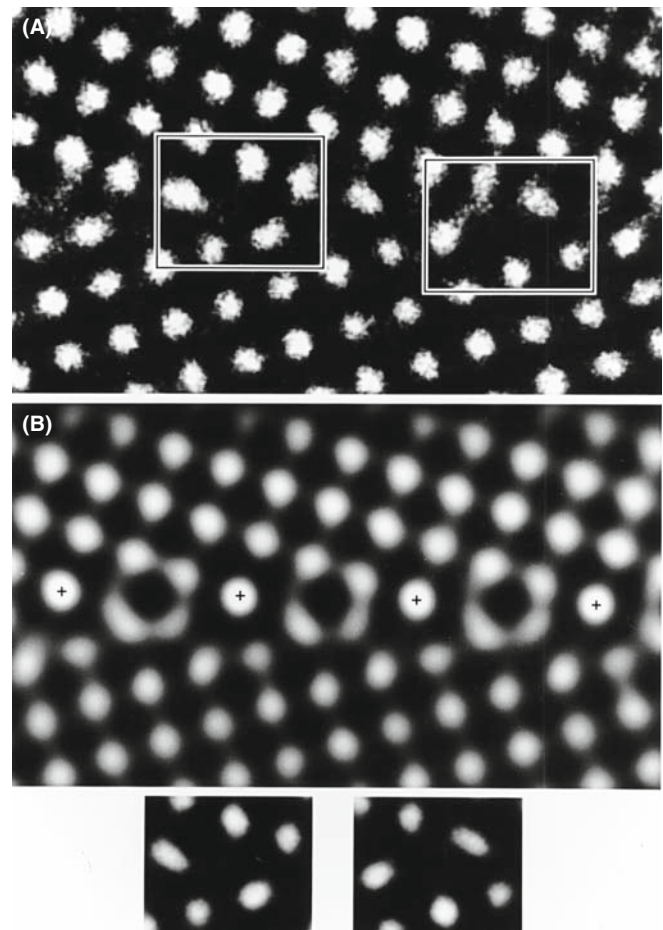


FIGURE 31.17. Analysis of a small region of a $\Sigma = 5$ tilt boundary in TiO_2 . The two small boxed regions in (A) are only present at the boundary; these are used as the templates for the cross-correlation method. (B) The cross-correlated image. The small rectangles at the bottom of the figure are the low-noise-averaged images of the GB templates.

already got the right magnification and rotation. If not, you have to set these first, but remember that the lens will still cause distortions. We will return to this problem shortly. Klug actually used pattern recognition in the work that led to his Nobel Prize.

REAL-SPACE APPROACH

We only look at the image.

This is a real-space approach. The process is best understood by an example. Figure 31.17A shows an HRTEM image of a small region of a $\Sigma = 5$ tilt boundary in TiO_2 . The two small boxed regions appear only at the boundary and are selected as templates. The matching process has then been carried out and the new image is shown in Figure 31.17B. Having found all the regions which match the template, we could then take the average of these to produce low-noise images of the GB templates. The final step is the comparison of these templates with models of the GB structure. There are two important points to remember

- When you average images, you implicitly assume that all the images are the same except for random noise.
- Don't forget our discussion in Chapter 30 of interface grooving and the problems associated with interfacial segregation.

31.10 PARAMETERIZING THE IMAGE USING QUANTITEM

In general, the thickness or chemistry will vary as you cross the specimen, i.e., the projected potential varies across the specimen. This means that one template will only match a small area, so you have to use many templates. These templates could, in principle, be totally empirical, but to be quantitative we must derive them from image simulations. This approach has been described for two special cases in papers by Kisielowski et al. and Ourmazd et al.

31.10.A The Example of a Specimen with Uniform Composition

In QUANTITEM, the results of Chapter 30 are summarized by a general equation linking the intensity and all the imaging (S_i) and materials (P) parameters

$$I(x, y) = F(P(x, y), S_i) \quad (31.5)$$

This equation just tells us that the intensity depends on the imaging conditions and on the specimen. For a particular set of imaging conditions, S_i will be known (more or less) and we'll call it S_i^0 . Then we can write that

$$I(x, y) = F(P(x, y), S_i^0) = F^0(P(x, y)) \quad (31.6)$$

The basis of this approach is quite straightforward, potentially powerful but almost never used

- Define the function F^0 for each image that you may obtain.
- Then construct a set of templates for your matching process.

Providing you stay within one extinction band, F^0 will be directly related to the projected potential of the specimen. A helpful simple analogy for this process is illustrated in Figure 31.18. The function F^0 describes the path of a swinging pendulum as it varies with time (Figure 31.18A). Each value of F^0 corresponds to a snapshot of the pendulum, so if you plot F^0 you can 'see' the path of the pendulum (Figure 31.18B). The velocity of the pendulum is related to the density of points along the path. So it should be possible to plot out the function F^0 from a single lattice image even if you don't know the microscope parameters used to form the image.

Yes, there are limitations and conditions and we'll discuss them later. All we need now is a method for representing each image by a snapshot of the pendulum: we have to *parameterize* the image. This process is the key to the technique. Manipulating and quantifying, in principle, thousands of images, each requiring 4 M byte of memory is not a fast process, even if you do have that much memory. If we could characterize each image by a few numbers (a vector or parameter) the comparison process could be much faster.

We separate the image into unit cells and digitized these to give many templates which are n pixels by m pixels, as shown here for one cell (Figure 31.18C–F). If we define N to be $n \times m$, then we have N numbers for the N pixels, where each number represents a gray level. Now the N numbers are regarded as the N components of an N -dimensional vector. (The math is not complicated but don't try to visualize this vector.) So now all the information in each unit cell is represented by a vector in N -dimensional space. The function F^0 describes how these N -dimensional vectors (the path) change as the projected potential changes.

PARAMETERIZING IMAGES

This may remind you of using the Hough transform to quantify EBSD patterns from the SEM.

The next step is to define a reference frame for these vectors. Three basis vectors are derived from the experimental image. Ourmazd et al. argue that three basis vectors will be sufficient, as we can show in the

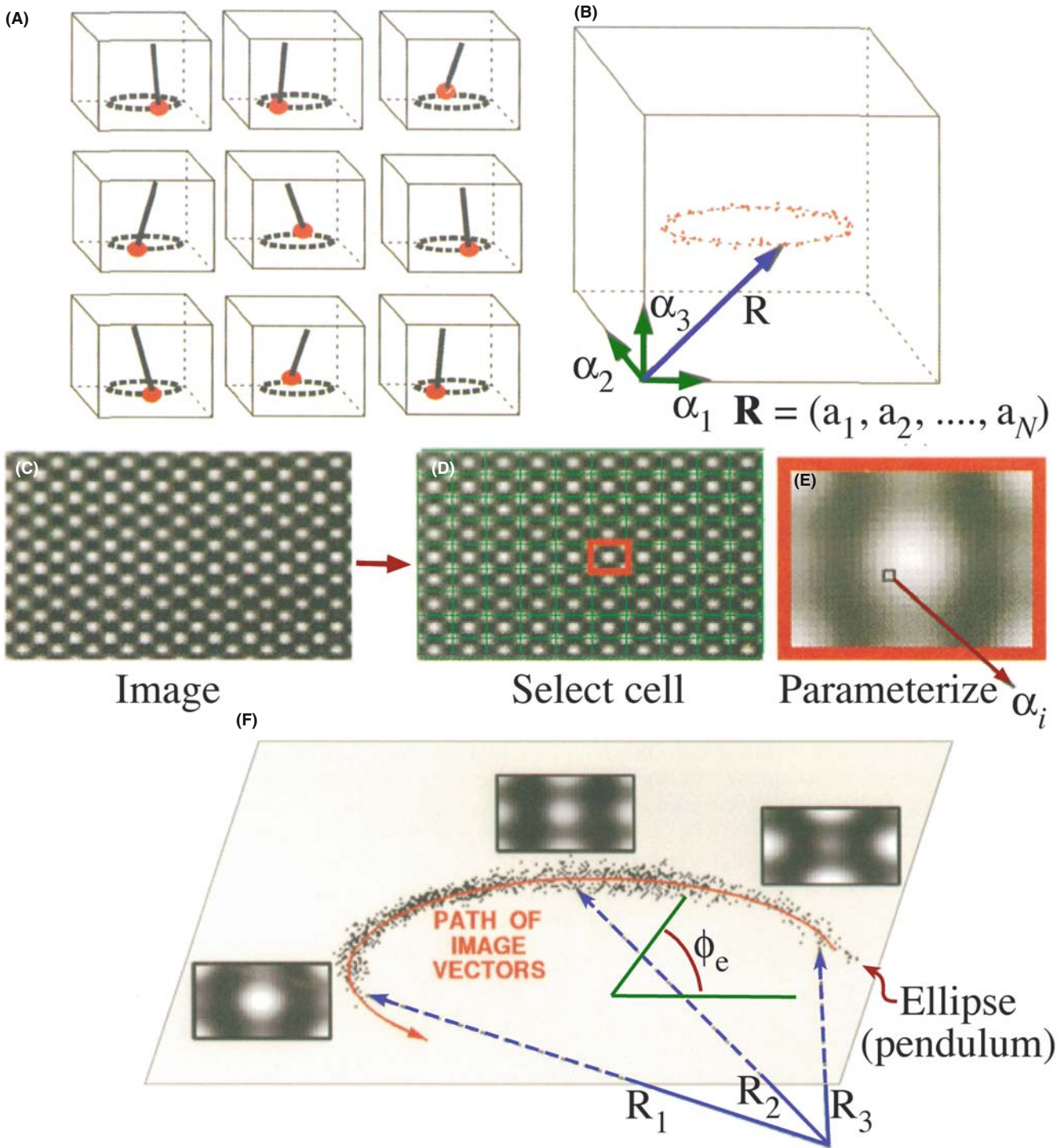


FIGURE 31.18. The principle of vector parameterization used in QUANTITEM. Each HRTEM image is represented by a single vector \mathbf{R} which has N dimensions. (A) and (B) show the ‘swinging pendulum’; (B) shows its ‘path’; (C–E) the image is separated into unit cells and digitized to give $(n \times m)$ pixel templates; (F) three vector-parameterized images ($\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$) of a wedge-shaped specimen of Si at different thicknesses.

following way. We will be using a low-index zone axis for any HRTEM analysis. Then we have three types of images

- The background, \mathbf{R}^B , due to the direct beam, O .
- A single-period image, \mathbf{R}^S , due to the interference between O and the strongest reflections, G_i .

- A double-period image, \mathbf{R}^D , due to the interference between these strong G_i reflections.

Each of these \mathbf{R} terms is a vector which represents an image. Any image we can form must be a combination of these three types of image, so a general image, G , can be written as

$$\mathbf{R}^G = a_G \mathbf{R}^B + b_G \mathbf{R}^S + c_G \mathbf{R}^D \quad (31.7)$$

Each of the basis vectors (images) can be expressed in the same manner

$$\mathbf{R}_i^T = a_i \mathbf{R}^B + b_i \mathbf{R}^S + c_i \mathbf{R}^D \quad (31.8)$$

giving three vectors for $i = 1, 2, \text{ and } 3$.

We can, in principle, turn these equations around (invert them) to define any vector \mathbf{R}^G in terms of the basis vectors

$$\mathbf{R}^G = \alpha_G \mathbf{R}_1^T + \beta_G \mathbf{R}_2^T + \gamma_G \mathbf{R}_3^T \quad (31.9)$$

which is what we wanted to show.

Ourmazd et al. point out that this treatment gives three important results

- The vector notation allows us to parameterize the lattice image.
- Projecting the vectors onto planes and/or paths aids noise reduction.
- Any noise which remains can be quantified.

The result of vector-parameterizing an experimental image of a wedge-shaped specimen of Si is shown by the three images corresponding to \mathbf{R}_1 , \mathbf{R}_2 , and \mathbf{R}_3 in Figure 31.18F.

31.10.B Calibrating the Path of R

In order to relate any image to the projected potential, we have to calibrate the curve showing the path of \mathbf{R}^G . This is where the image simulation comes in. We start with the vector-parameterized analysis of the series of simulated images from the Si wedge. Each point on the curve corresponds to an image of the unit cell, and thus to a vector \mathbf{R}^G . The ellipse has been fitted empirically, and the thickness of the cell has been increased by 0.38 nm for successive calculations. The points are closer together in some parts of the plot because, as we saw in Chapter 30, some characteristic images appear for a wider range of thicknesses. Now we have a way to quantify this ‘experimental’ observation. What the ellipse does is to allow us to parameterize the path in terms of the phase angle of the ellipse ϕ_e , shown by the ellipse in Figure 31.18F. Thinking back to the pendulum analogy, the path parameters are the image version of the coordinates for the harmonic oscillator.

Now the parameters of the curve (ϕ_e) can be obtained from a series of images. We can vary the material and in each case examine three other variables

- The orientation of the specimen (i.e., the zone axis).
- The defocus, Δf , of the objective lens.
- The specimen thickness.

The remarkable result is that when we plot ϕ_e versus the thickness, normalized by the extinction distance, we obtain a straight line. The explanation for this result is related to the fact that only a small number of Bloch waves usually contribute to the image, as we saw in Section 30.6. In materials such as YBCO, this is not the case, and the curve is not a straight line.

This analysis gives more detail than most students will ever need, but it introduces a concept that could (should) be more widely recognized: the power of parameterizing the image.

31.10.C Noise Analysis

Noise in the image moves the vector off the ellipse; so, we have a method for analyzing the noise. If the noise moves the vector exactly along the ellipse, we can’t analyze the noise, but that is quite unlikely since the noise would then be accurately mimicking a change in projected potential. So this parameterization gives a way to reduce the noise by a factor of \sqrt{N} which for a 10 pixel \times 10 pixel cell is a factor of 10!

The analysis given by Ourmazd et al. then shows that, in the case where only two Bloch waves are excited, the image intensity, I , can be expressed as

$$I = B + S + D \quad (31.10)$$

where B , S , and D are the contributions from the background, single interaction, and double interaction as we defined them above. The point (B, S, D) does indeed describe an ellipse which lies on a plane independent of Δf .

The value of this approach can be appreciated if you look at the examples shown in Figure 31.19. In the first example (Figure 31.19A and B), the technique has been used to provide a map of the roughness of the Si surface. The experimental image looks really uniform until you analyze it using this method, when you can discern the roughness at the 0.5-nm level; remember that there are two surfaces.

As you know from earlier discussions, changes in chemistry produce effects which are similar to changes in thickness, because they change the projected potential. In terms of the present analysis, the effects are different: composition changes cause changes in the ellipse and in ξ (notice that there is no subscript, since this ξ is a many-beam value).

The method is more limited, in this respect, to change in composition but can be used if the thickness and roughness are known, i.e., if you can measure the roughness elsewhere on your specimen (using a known reference cell) and infer it for the area you want to analyze. (Warning lights should be flashing.) The approach is as follows

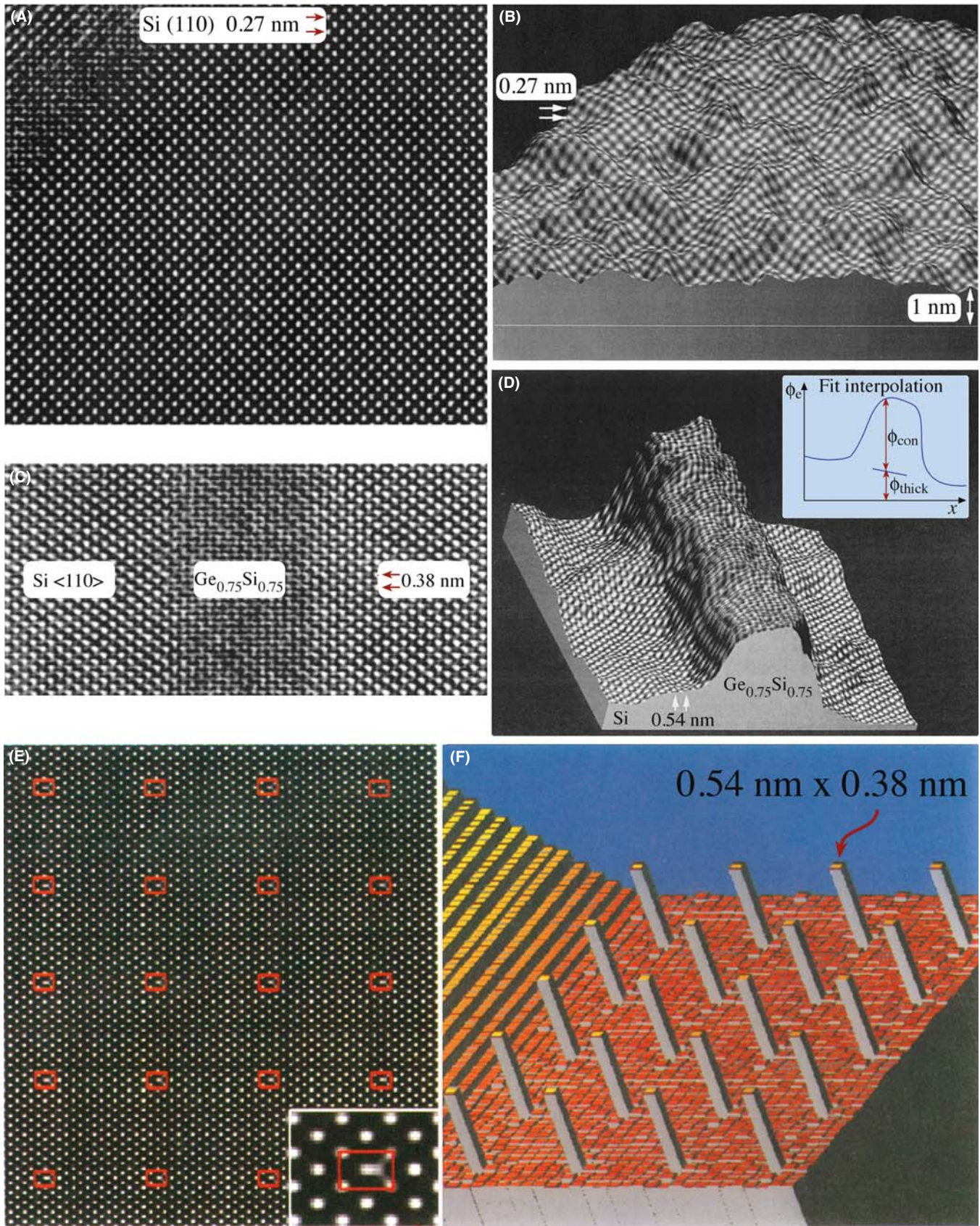


FIGURE 31.19. Examples of the application of QUANTITEM: on the left is the image, on the right is the QUANTITEM image. (A, B) Mapping of the roughness of the Si surface covered by SiO_2 ; (C, D) a layer of $\text{Ge}_x\text{Si}_{1-x}$ in a matrix of Si, the inset shows the plot of ϕ_c versus x ; (E, F) analyzing a simulated image of columns of Ge (a δ -function in concentration) in Si.

- Use QUANTITEM to measure the advance in ϕ_e at your target cell relative to your reference cell.
- Subtract $\Delta\phi_e$, which is due to a thickness change.
- Then the rest of the change in ϕ_e must be due to changes in ξ . If you know how ξ varies for different compositions, you have determined the local composition.

The second example shown in Figure 31.19C and D is a near-perfect application for the technique, since the elements in the alloy $\text{Ge}_x\text{Si}_{1-x}$ are randomly located on the lattice sites. The slope tells us how abruptly the composition varies.

NOISE

You can test the potential resolution of the technique, its sensitivity to the alignment of the beam, bending of specimen, and beam divergence in the usual way by creating model structures, simulating the images, and then analyzing them.

The third example in Figure 31.19E and F shows that the potential resolution is superb (we can see the quantum dots) but beam tilt can cause 10% errors in thickness measurement. The conclusion is clear: as always, you only get the best results if your specimen is ideal and both the microscope and the specimen are perfectly aligned. Note, however, that the technique has not yet been successfully applied to a wide range of materials, but it is complementary in many ways to STEM Z-contrast (see Figure 22.15).

31.11 QUANTITATIVE CHEMICAL LATTICE IMAGING

This technique is an instructive tool but is perhaps less necessary now that STEM is commonly available. The technique has not been widely used since the software was not readily available.

This technique uses the approach described in Section 30.4, but can only be applied to materials where we have chemically sensitive reflections, which we discussed in Section 16.4. We used these reflections in Section 29.4 to produce chemically sensitive DF images. In HRTEM, the chemically sensitive reflections not only contribute to the overall image but they will generally have a different dependence on thickness, too.

This effect is shown in Figure 31.20A–D for AlAs and GaAs, which have identical structures: the 002 reflection is allowed for both, but is stronger for AlAs since F , the structure factor, is proportional to $f_{\text{III}} - f_{\text{V}}$. You can see that, under the conditions chosen for this comparison, the intensity of the 022 reflections is also

very different for the different thicknesses. The sort of image we can analyze with this approach is illustrated here by the quantum wire. We want to know how abruptly the composition changes at the interface. In this example, the ideal GaAs and $\text{Al}_{0.4}\text{Ga}_{0.6}\text{As}$ unit-cell images are first characterized by the two vectors, \mathbf{R}_{GaAs} and $\mathbf{R}_{\text{AlGaAs}}$, following the approach described in Section 31.10. In this case the cells were simulated, divided into 30×30 pixel arrays (so $N=900$), and then \mathbf{R} was plotted. The information content is contained in θ_C . As before, we can directly assess the noise in such an image. So how is the direction of \mathbf{R} dependent on composition?

The technique is explained using the three known simulated templates (GaAs, AlGaAs, and AlAs); each one produces a vector \mathbf{R}^t . Although the vector for the intermediate composition does not lie in the plane, it can be projected onto this plane to give a unique vector for certain ranges of thickness. Since this is a complex procedure, you'll find the 'image flow chart' shown in Figure 31.21 helpful.

- The experimental image is first digitized; the image contained approximately 25×25 unit cells and used a 514×480 frame buffer.
- Next, the image must be separated into individual cells.
- The pair of templates shown is then used to calculate the angular positions of the \mathbf{R} vectors for all the unit cells. Such templates can be calculated or taken from known areas of the specimen.
- These \mathbf{R} vectors are characterized in terms of where they cut through a plane (D) (see Figure 31.21D also).

The maximum chemical difference determines how far apart the two principal distributions can be (see Figure 31.21E). Since the image is now fully parameterized we can do the statistics and finally invert the angular data to give the compositions.

This technique has enormous potential, but you must also remember that it is susceptible to all the drawbacks inherent in HRTEM. The advance is that now you can put numbers on those effects. The technique is material-specific, but if you know your material, you can combine image simulation and this processing method to examine what will be the limiting factors for your material. You can construct a test image like that shown in Figure 31.19E. If your specimen is ideal, you could, in principle, easily detect a column of Al in a mainly GaAs matrix without any 'spreading' due to the electron beam and in this respect alone makes QUANTITEM better than atomic-resolution X-ray mapping and equivalent to atomic-column EELS (see Part 4). This technique could be applied to other materials.

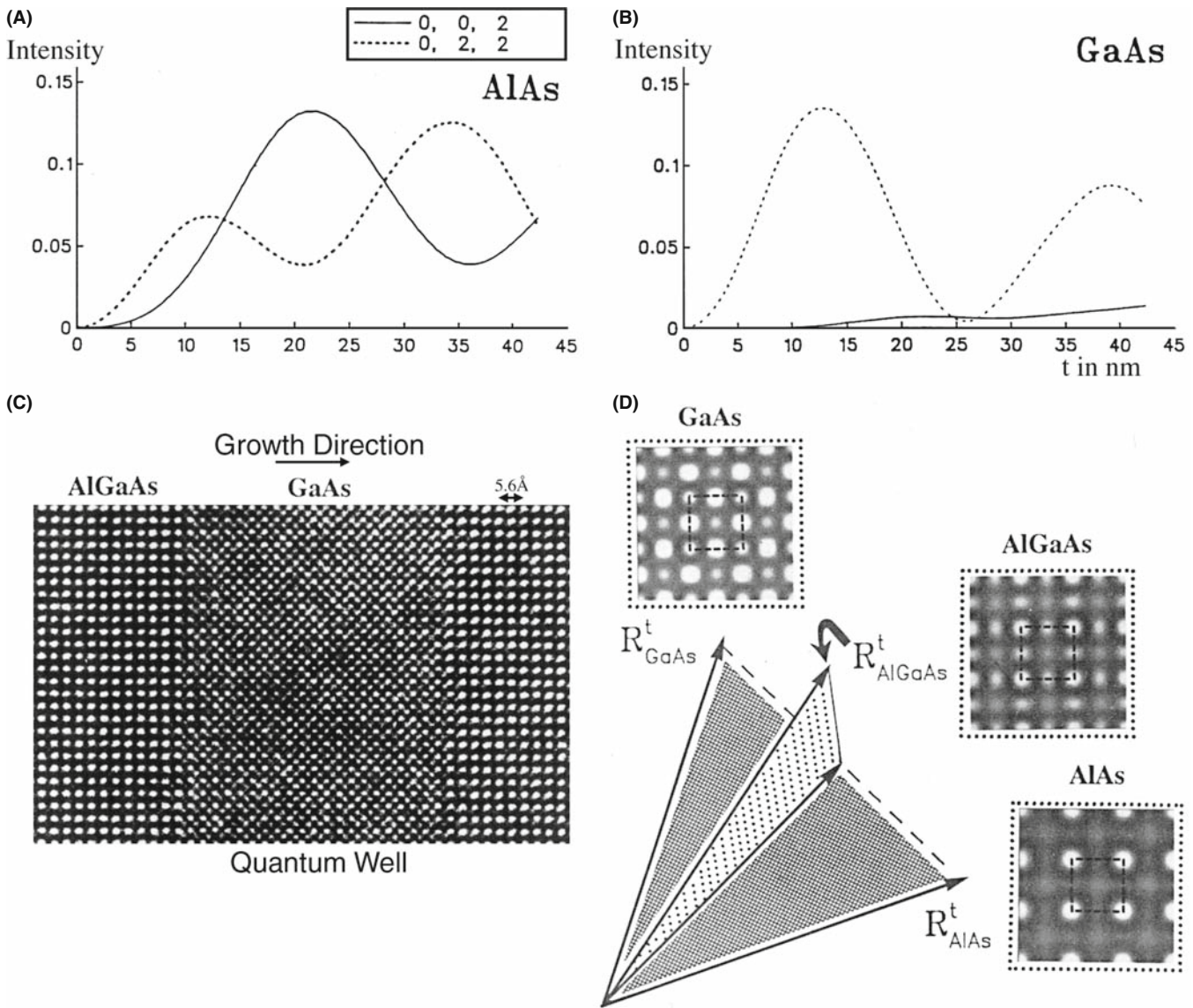


FIGURE 31.20. (A,B) Variation in intensity of the (002) and (022) beams along [100] in AlAs and GaAs (400 keV); (C) chemical lattice image of a layer of GaAs between two layers of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ($x = 0.4$); (D) templates simulated for different values of x each produce a vector \mathbf{R}^t .

31.12 METHODS OF MEASURING FIT

There are two methods presently used to obtain a measurement of how well the data and the model match. In statistics, we determine this ‘goodness of fit’ using cross correlation or a least-squares refinement (Section 31.13).

CROSS CORRELATION

The cross correlation gives a measure of how similar two images (or signals) are. The math process looks very much like a convolution of two functions but we don’t take the Fourier transform. What’s special is that the computer does it easily as long as our images are numbers!

LEAST-SQUARES

A method of fitting a curve to a series of values.

In this section we’ll use the cross-correlation method to compare an $n \times m$ pixel template (see Section 31.9) with every possible $n \times m$ rectangle in the image. The computer moves the template across the image one pixel column at a time, then shifts down one pixel row, and repeats the exercise. The cross-correlation function (CCF) gives the goodness of fit or a ‘measure of similarity’ between the template and each $n \times m$ image

$$CCF(x, y) = \frac{\sum_{x'} \sum_{y'} [i(x', y') - \langle i(x', y') \rangle] \cdot [t(x' - x, y' - y) - \langle t \rangle]}{\sqrt{\left\{ \sum_{x'} \sum_{y'} [i(x', y') - \langle i(x, y) \rangle]^2 \sum_{x'} \sum_{y'} [t(x' - x, y' - y) - \langle t \rangle]^2 \right\}}}$$

(31.11)

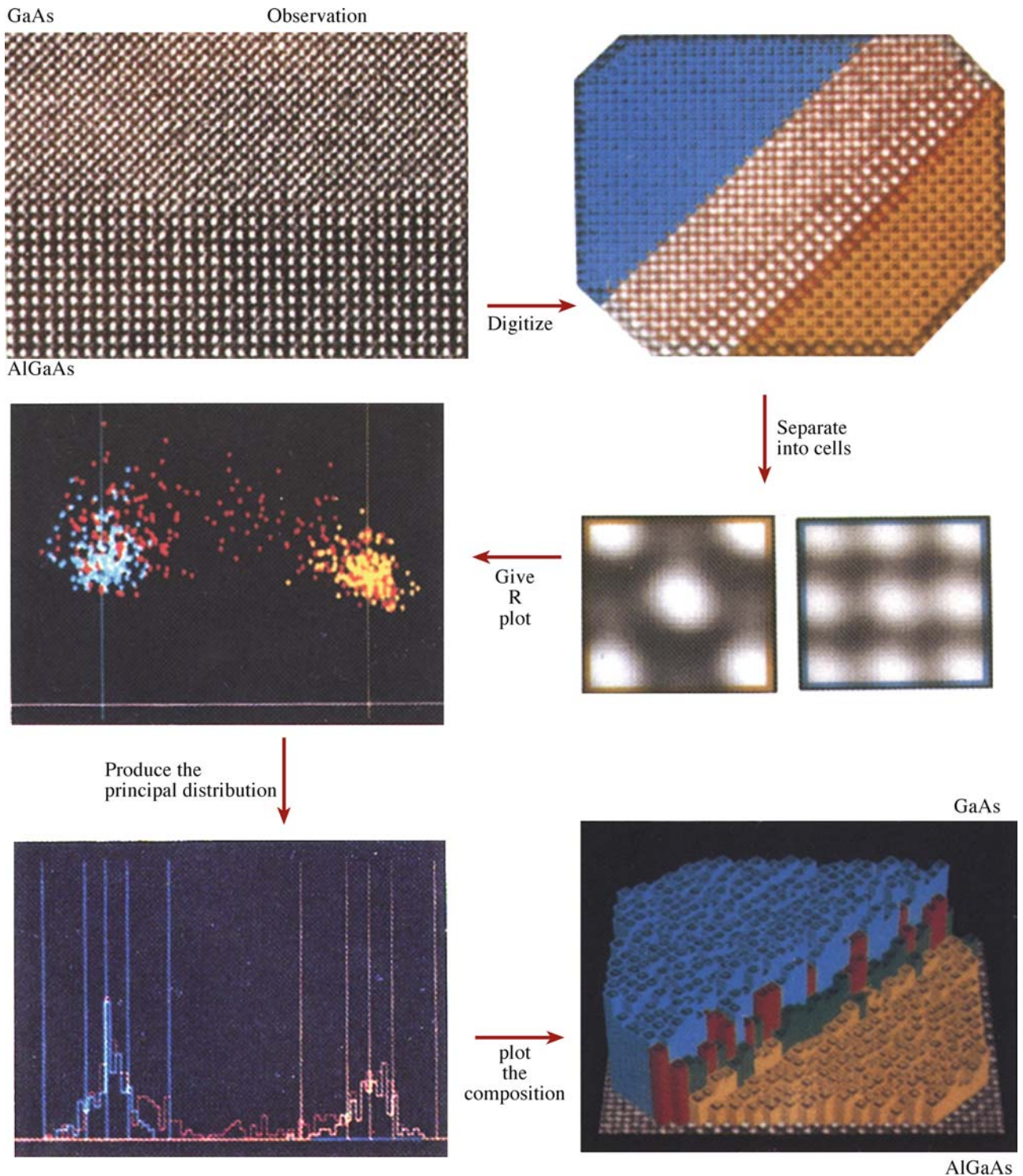


FIGURE 31.21. A flow chart summarizing the pattern-recognition procedure.

In this equation x varies from 0 to x_{\max} , y varies from 0 to y_{\max} .

- $i(x', y')$ represents the image.
- $t(x', y')$ represents the template.
- $\langle t \rangle$ is the average value of the pixels in $t(x', y')$; it is computed just once.

- $\langle i(x, y) \rangle$ is the average of $i(x', y')$ in the region coincident with the current location of t .

The summations are taken over the coordinates common to both i and t . The origin of the image is at its top left corner and the origin of the template is at its center. In this equation, the denominator is a

normalization factor, so that the CCF will not depend on any difference in the intensity scale between the template and the image.

We can rewrite equation 31.11 as the dot product of two vectors \mathbf{t} and \mathbf{i} , which gives us the $n \times m$ component of the template

$$\text{CCF}(x, y) = \cos(\theta) \frac{\mathbf{t} \cdot \mathbf{i}}{|\mathbf{t}| |\mathbf{i}|} \quad (31.12)$$

Now we can plot the CCF as a map of our image and then examine it to deduce where there is a particularly good match. Since the CCF value varies from 0 to 1, we can plot out the number of times each particular CCF value occurs, as illustrated in Figure 31.22. The two peaks in this curve correspond to the best fit and the worst fit, so the distance between them gives a measure of the 'discrimination signal.' From the width of the peaks we have a measure of the noise and hence a signal-to-noise ratio. The regions of good fit can be combined to produce a better template, and the process repeated, giving the dashed line. A second measure of the noise is then given by how far the good peak differs from unity. A particularly nice feature of this approach is that the procedure is available as a plug-in module for Digital Micrograph (see Section 1.6). The template could alternatively be a simulated image and the process repeated for a series of different thicknesses and/or defocus values. When you want to learn more about correlation techniques, see the article by Frank.

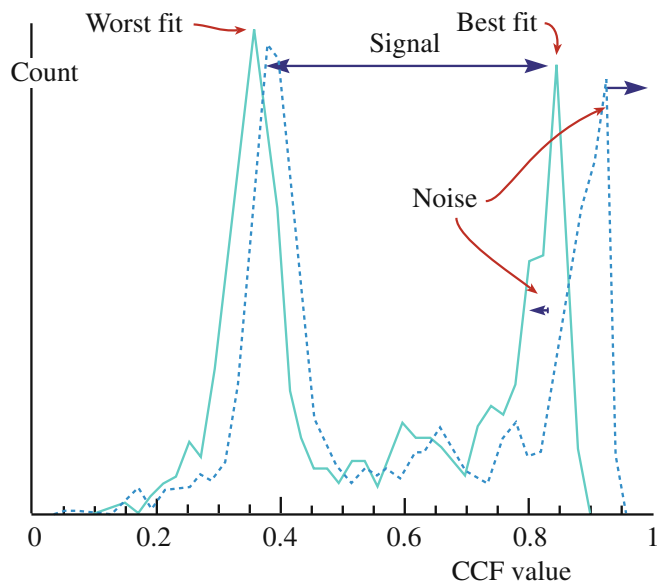


FIGURE 31.22. A plot of how often a particular CCF value occurs. The two peaks in the continuous curve are the best fit and the worst-fit: their separation gives a measure of the discrimination signal; the width of the peaks gives a measure of the noise and hence a signal-to-noise ratio. The plot can be redrawn after repeating the process (dashed line) to estimate the improvement in signal-to-noise ratio.

31.13 QUANTITATIVE COMPARISON OF SIMULATED AND EXPERIMENTAL HRTEM IMAGES

If we want to compare simulated and experiment images quantitatively, we really should modify our usual approaches to both simulation and experiment (King and Campbell 1993 and 1994). When doing the simulation, most programs automatically adjust the gray scale for each image so that darkest is 1 and brightest is 0 (or vice versa). This means that two simulated images might appear similar even though you would hardly see the pattern in one if both appeared on the same negative. In a similar way, we usually print an image to be as clear as possible using the full range contrast of the photographic paper.

We need methods for normalizing these procedures if we want to make quantitative comparisons. The solution for the simulation is simple. For the experimentalist, it means recording extra data while you're at the microscope. After recording the image, you record another image with the specimen removed. You then use this image to scale the lattice image such that you correct for variations in intensity across the field of view and the non-linearity of the response from the photographic film. Figure 31.23 illustrates the experimental transmittance for Kodak SO-163 film, 400-keV electrons, plotted against the digital value on a CCD array. Of course, you must process both images at the same time. This is called the 'flat-field' correction; a slow-scan CCD camera would simplify this procedure at the cost of reducing the area you examine.

When you analyze the image, you'll find out if the area you photographed was correctly aligned. Since your image takes only 2 seconds or so to record, you

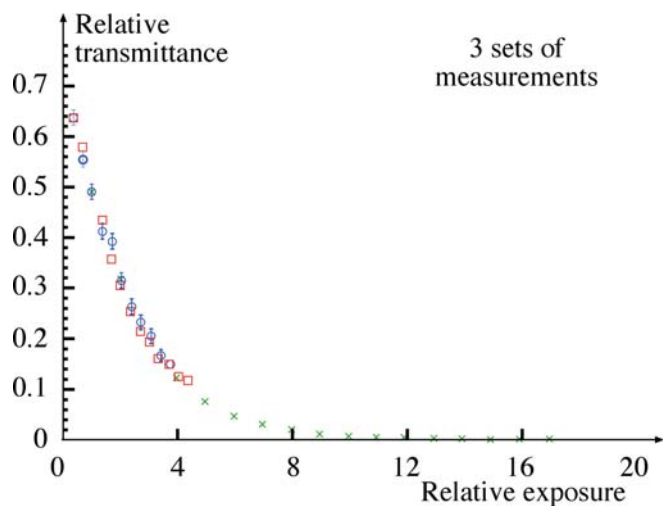


FIGURE 31.23. Plot of transmittance versus relative exposure measured using a CCD camera to digitize images from Kodak SO-163 film. (The symbols indicate three different microscopes.)

may risk several exposures using this technique. You are now comparing numbers, so you can use a least-squares fit where the residual $f_i(x)$ is defined as

$$f_i(x) = \frac{[f_i^{\text{obs}} - f_i^{\text{calc}}(x)]}{W_i} \quad (31.13)$$

and your task is to minimize $f_i(x)$. The difference between the intensity in the experimentally observed i th pixel and its calculated value will be zero if everything has been scaled correctly, the imaging conditions (Δf , C_s , etc.) are correct, and you have the right structure.

HRTEM AND LOW DOSES

Don't forget that since HRTEM uses higher voltages, the perfect image will only be recorded from an area of the specimen that has only seen the beam while you recorded the image! So you should always use low-dose techniques for quantitative imaging (but others don't).

Let's say W_i is the image which represents the error bars for pixel i . Then we can write that

$$W_i = \min \left[\sum_{i=1}^N f_i(x)^2 \right] \quad (31.14)$$

This equation defines the non-linear least-squares problem. We use x to summarize a set of parameters (Δf , C_s , the model, etc.); N is the number of pixels in the image. Fortunately, this analysis is now routine statistics. You'll need a computer program to tell you how good the first guess was, make an improvement and continue until it meets our specific criterion for matching. [King and Campbell used MINIPACK-1. (See articles by Moré et al.)]

In their demonstrations of this approach to analyze a [001] tilt GB in Nb, King and Campbell varied four parameters: thickness, defocus, x -tilt, and y -tilt. The steps were as follows

- They first optimized the electron-optical parameters using a 64×64 pixel image, giving $N = 4096$ and an image computational cell of 3.303 nm by 3.303 nm. Using the EMS program (Section 1.6), the optimization took 20 iterations and 80 multislice calculations.
- Next they had to optimize the structure of the GB. This process required defining 84 atomic positions in a unit cell of 4.16 nm \times 1.04 nm and a 512×128 (= 65,536) pixel image. Now the optimization required 16 iterations and 1300 multislice calculations.

These numbers are instructive. First they tell you that this computation can be done, which wasn't obvious. Second they tell you that this is a computer-intensive process; that we could have guessed!

You'll need to take enormous care in this type of analysis

- Align the simulated cell with the experimental cell and measure the unit cell in pixels.
- Choose a number of cells and relate them by the translation vector parallel to the rows of the image array.
- Calculate the standard-deviation images.
- Rotate the unit cell and repeat the exercise several times.

The orientation which gives the smallest standard deviation is your alignment. You must now adjust the magnification of the experimental image to fit the simulation, in a similar way to what you did for rotation. Next, you have to match the origins of both cells; the procedure is the same as we just described but translating the unit cell not rotating it. For a bicrystal, you now repeat this exercise for the other grain and then for the GB. You can improve the fit further if you take account of a constant background contribution, which probably arises due to the amorphous layer on both surfaces. Comparing experimental and calculated images quantitatively, we define f_i^{obs} as the corresponding value in the simulated image. We then calculate the residual $f_i(x)$ as follows

$$f_i(x) = \frac{(f_i^{\text{obs}} - (f_i^{\text{calc}}(x) + b^{\text{fit}}))}{W_i} \quad (31.15)$$

where b^{fit} is included as a free parameter in the optimization procedure. King and Campbell's calculations showed that W_i could be expressed as

$$W_i = \sigma_i^{\text{obs}} + 0.05f_i^{\text{obs}} \quad (31.16)$$

where σ_i^{obs} is the standard deviation of the i th pixel. Examples of the experimental, best fit, and normalized residuals are shown in Figure 31.24 for images from a $\Sigma = 5$, (310), [001] GB in Nb.

31.14 A FOURIER TECHNIQUE FOR QUANTITATIVE ANALYSIS

Möbus et al. proposed using what is referred to as an adaptive Fourier-filtering technique. The HRTEM image is digitized in the usual manner and then a special spatial-frequency filter is applied. This type of mask is designed for analyzing regions which contain defects.

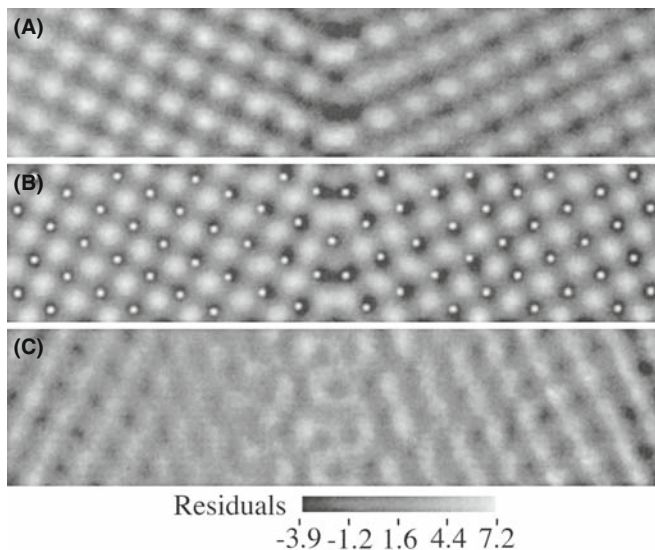


FIGURE 31.24. (A) Experimental image, (B) best fit simulation, and (C) normalized residuals of a $\Sigma = 5$ symmetric tilt boundary in Nb.

So the idea is that the computer automatically optimizes the mask to maximize the separation of the signal and the noise. This approach has not been widely practiced in TEM but clearly holds enormous promise. By varying the mask, this approach can prevent the analysis of a defect layer being dominated by the bulk information. Since the approach is quite straightforward signal processing, we will just illustrate an example found in the analysis of a simulated $\Sigma = 5$ GB with an extra period along the boundary. To test the analysis, white noise was added to a calculated image to give the image shown in Figure 31.25A. The power spectrum (the computer-generated DP) of the micrograph is shown in Figure 31.25B. The adaptive filter and the filtered

image are shown in Figure 31.25C and D. The important feature of the adaptive filter in this example is that it was created as such because the computer detected the doubling of the periodicity which is *only* present in the GB. Second the mask consists of elongated openings which we know we need when analyzing the GB because of the shape effect that we discussed when analyzing DPs (see Chapter 17).

AN ADAPTIVE FILTER
 One where the shape of the filter, or mask, is adapted to fit the shape of the 'image' it's filtering.

31.15 REAL OR RECIPROCAL SPACE?

In principle, we could equally well compare two images in reciprocal space rather than real space. However, because the Fourier transforms can generally be carried out much faster, the real-space approach has several advantages

- Fourier analysis separates local information into sine and cosine functions which are delocalized. When we reassemble the real-space image, higher parts of the frequency spectrum will be lost which will degrade the resolution.
- We want to maintain information on the absolute value of the intensities.
- Real-space methods are visually more intuitive for most of us. We can easily see what we have removed in the process.
- The real-space approach allows us to choose any values of n and m in defining our templates. Fourier space prefers aspect ratios given by 2^n .

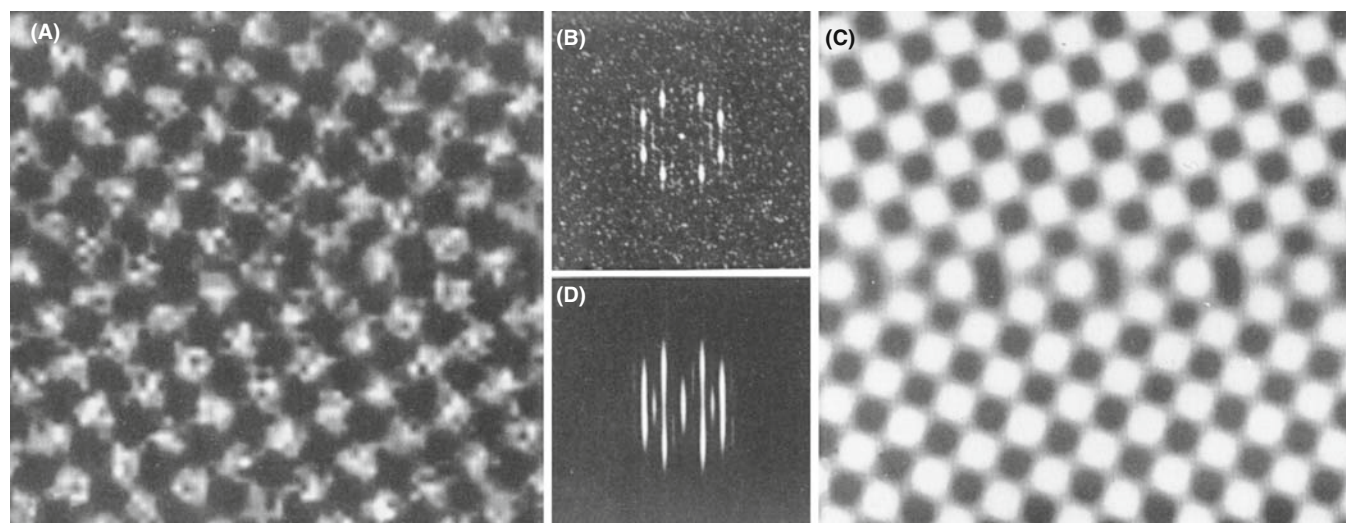


FIGURE 31.25. (A) White noise added to a calculated image of a $\Sigma = 5$ GB; (B) the power spectrum of (A); (C) the adaptive filter; (D) the filtered image.

31.16 SOFTWARE

Several software packages are widely used by the TEM community and have been listed in Section 1.6. We summarize the ones we know in Table 31.1. In its earliest application, image processing in TEM was almost exclusively applied to HRTEM images. This is no longer the case. Remember: always start with the best possible data. You can't always obtain a perfect image because your specimen might be beam sensitive, or coated with oxide, and you need to be aware of these limitations when processing or quantifying the image. This chapter has given you a hint of what is possible and where the subject is developing. We recommend that you obtain the software and start experimenting.

In this chapter we have discussed several different techniques used for processing TEM images. Different software is suitable for processing different types of image. The problem is that there may not be sufficient financial benefit unless a package supports another product; a similar problem was found for simulation software in Chapter 30. Many programs have been written

to run on particular computer systems and support can disappear even though the TEM has a lifetime of 15+ years.

31.17 THE OPTICAL BENCH—A LITTLE HISTORY

Although not widely used now, the optical bench is a useful instructional tool but you need to be old fashioned and record the image on film! A typical experimental setup is shown in Figure 31.26. The laser provides a coherent source of illumination representing the electron beam. The negative acts as the specimen. If it contains a set of lattice fringes, these act as a diffraction grating and give rise to a row of spots on the screen placed at the back focal plane of the 'objective' lens. The lens is thus performing an optical Fourier transform of the photograph. If you 'move the screen' to the image plane, the fringes reappear. You can make different masks and place them at the back focal plane or even create an 'adaptive filter' by exposing a photographic film and using this as the template for your mask. These masks correspond to the objective aperture in TEM. Students will find it instructive to transform their instructor or another suitable photograph, examine the frequency spectrum, and investigate the resulting spatial effect of different masks. The detail in the image is quickly lost as you remove the high spatial frequencies. This corresponds to inserting a smaller aperture in the back focal plane of the objective lens, as illustrated in Figure 31.27. So Figure 31.27D is effectively a BF image: clearly, you discard a lot of information in such images!

TABLE 31.1. Software for Image Processing

Package	Supplier	Cost
Digital Micrograph	Gatan	Commercial
NIH Image	NIH	Freeware

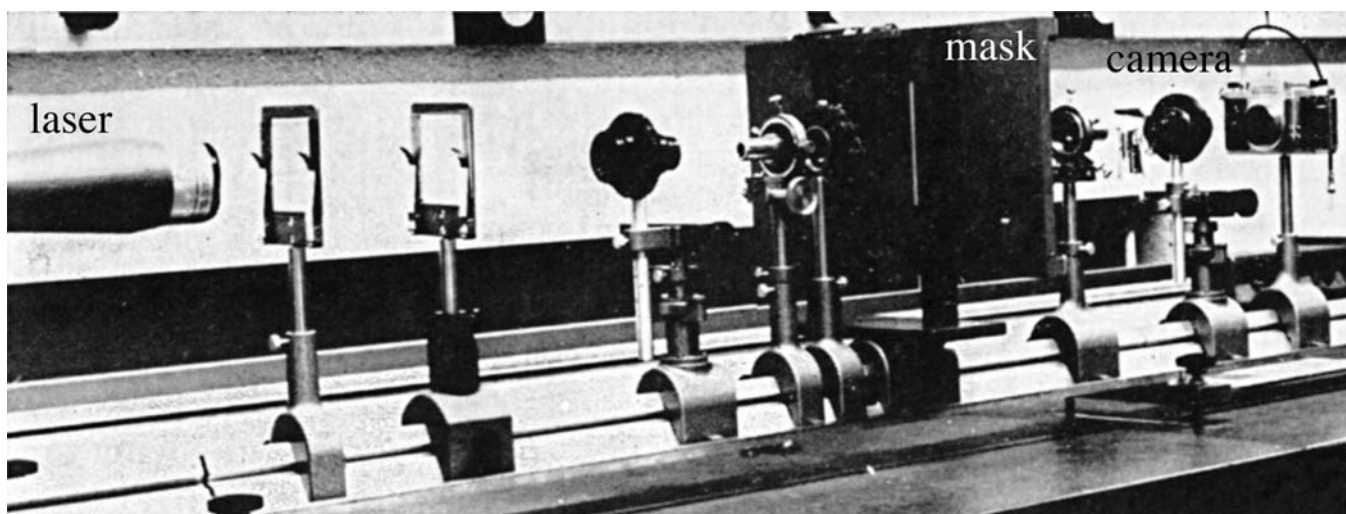


FIGURE 31.26. A typical experimental setup for an optical bench with the mask in the back focal plane.

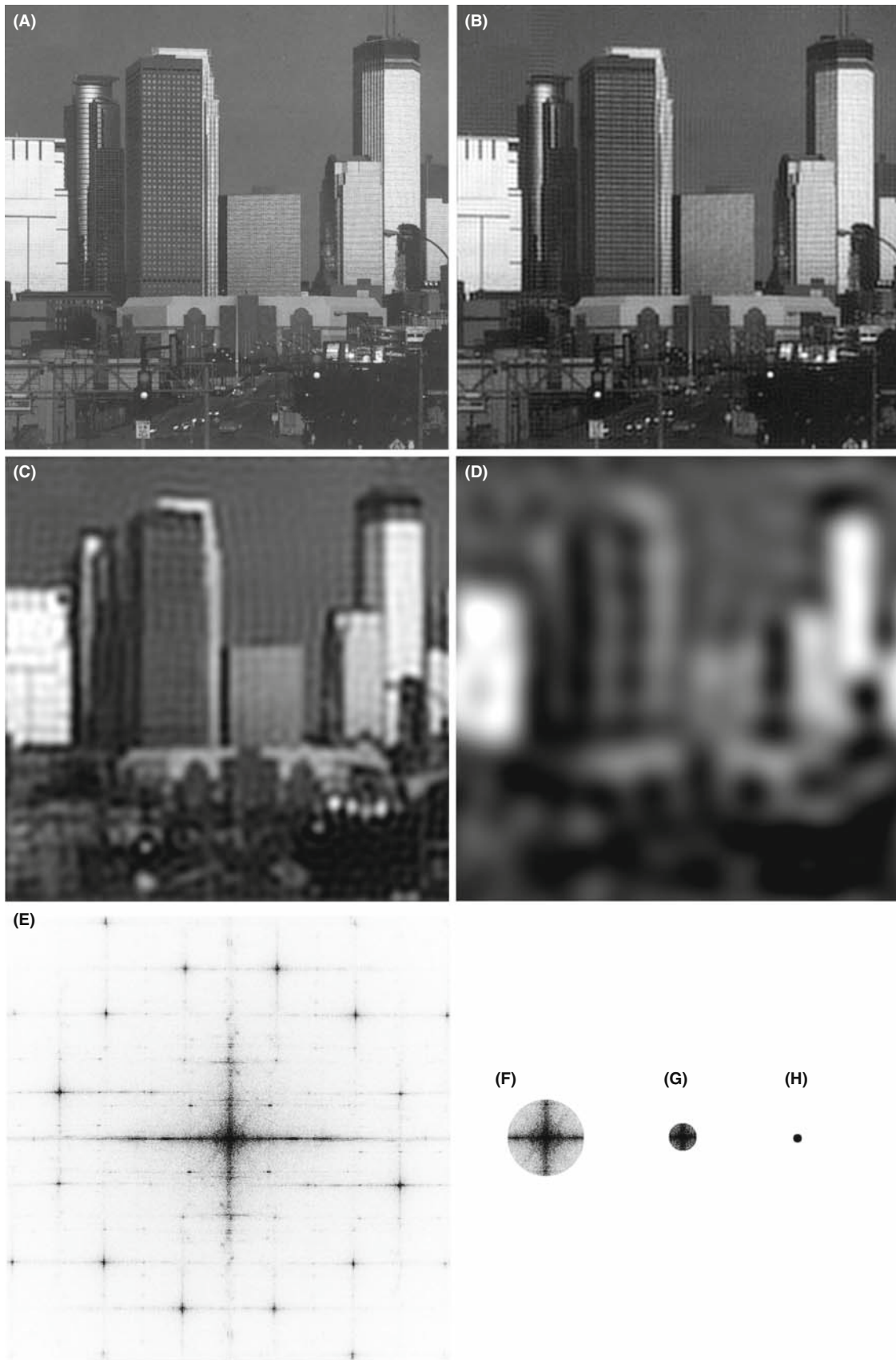


FIGURE 31.27. The effect of a mask (aperture) size on a non-periodic image of the Minneapolis skyline. (A–D) Reducing the aperture size, as indicated in the corresponding optical transform (diffractograms) (E–H) reduces the image detail. The streaks in (E) arise from the edges of the photographs.

CHAPTER SUMMARY

We have been doing image processing for many years; it's called 'dodging' in the photographic darkroom. You can even do this automatically with a special commercial enlarger. However, we have done very little quantitative imaging in materials science and even less in nanotechnology. The points you should remember when starting in the field are the following

- Quantitative comparison of simulated and experimental images depends on both the simulation program and the experimental parameters.
- If you are going to use reciprocal-space techniques for quantitative analysis, you should consider the shape of the optimum mask; it may not be a circular mask especially if you are studying interfaces! The process can, in principle, be assisted by the computer.
- The potential for image restoration is no longer limited by how the signals are mixed due to C_s and Δf . You can unscramble those effects if you have the right equipment.
- You will notice the repeated use of the word 'potential' where we don't mean $V(\mathbf{r})$! In many ways this chapter is a guide to the future of HRTEM and TEM in general. Some of the features won't be commonly available or optimized until the manufacturers realize their importance to the user and the user realizes that stability (for example) costs money.

Ultimately, the resolution limit is set by the signal-to-noise ratio in your image.

When you are processing images, there is always the possibility of removing information which is important. For example, Fresnel fringes often *should* be there! Beware of making reality match simulation, rather than the reverse. In the same vein, we draw your attention to the conclusion of Hÿtch and Stobbs, who found that they could only match their experimental and simulated images if they used a value for the specimen thickness which they knew was wrong! Their study emphasizes that, wherever possible, you should obtain independent measurements of the characteristics of your specimen and your machine. Remember the double-headed rhino in Figure 1.7; don't publish artifacts, even well-processed ones. Even if it's just for entertainment, images can reappear in unexpected places. Remember that an early image of Si dumbbells was later interpreted correctly because the authors gave the necessary data on how the image was formed when they published it. You should *always* do likewise.

PROCESSING IMAGES

Harrel, B *et al.* 1995 *Using Photoshop for Macintosh* Que Indianapolis. A detailed description of the possibilities using Photoshop; this volume should also serve as a warning for all microscopists. Ask the question: has the image you are examining been processed, and if so, how?

Hawkes, PW, Ed. 1980 *Computer Processing of Electron Microscope Images*, Springer-Verlag New York. Despite its age, this collection of articles is for advanced students.

ANALYZING C_s , Δf , ETC.

Krivanek, OL 1976 *Method for Determining the Coefficient of Spherical Aberration from a Single Electron Micrograph* *Optik* **45**, 97–101. An early paper giving a method for analyzing C_s and Δf using a diffractogram.

Zemlin, F, Weiss, K, Schiske, P, Kunath, W and Herrmann, K-H 1978 *Come-Free Alignment of High Resolution Electron Microscopes with the Aid of Optical Diffractograms* *Ultramicrosc.* **3** 49. The Zemlin tableau paper.

FOCAL SERIES AND TILT SERIES

Kirkland, EJ, and Siegel, BM, Uyeda, N and Fujiyoshi, Y 1982 *Nonlinear high resolution image processing of conventional transmission electron microscopy: II. Experiment* *Ultramicrosc.* **9**, 65–74.

Kirkland, AI, Saxton, WO, Chau, K-L, Tsuno, K and Kawasaki, M 1995 *Super-resolution by aperture synthesis: tilt series reconstruction in CTEM* *Ultramicrosc.* **57**, 355–374.

Kirkland, AI, Meyer, RR and Chang L-YS 2006 *Local Measurement and Computational Refinement of Aberrations for HR TEM* *Microsc. Microanal.* **12** 461–468.

Kirkland, AI and Meyer, RR 2004. *'Indirect' High-Resolution Transmission Electron Microscopy: Aberration Measurement and Wavefunction Reconstruction* Microsc. Microanal. **10** 401–413.

ADVANCED TEM PROCESSING

- Dorset, DL 1995 *Structural Electron Crystallography* Plenum New York. An early example of 3D reconstruction.
- Downing, KH 1992 Scanning Microscopy Supplement **6** AMF O'Hare IL p. 405.
- Hawkes, PW, Ed. 1992 *Signal and Image Processing in Microscopy and Microanalysis* Scanning Microscopy Supplement **6**, AMF O'Hare IL. This volume and its 1988 and 1996 companions are required reading for this subject.
- Hawkes, PW, Ottensmeyer, FP, Rosenfeld, A and Saxton, WO (Eds) 1988 *Image and Signal Processing for Electron Microscopy* Scanning Microscopy Supplement **2** AMF O'Hare IL.
- Hawkes, PW, Saxton, WO and Frank, J Eds. 1996 *Image Processing* Scanning Microscopy Supplement X AMF O'Hare IL.
- Russ, J.C. (1990) *Computer-Assisted Microscopy* Plenum Press New York. Chapter 3 is particularly relevant. Subsequent chapters give detailed analyses of the topics introduced here.
- Russ, JC 1995 *The Image Processing Handbook* 2nd edition CRC Press Boca Raton. A beautiful, comprehensive, and an essential component of any EM lab (or home).
- Saxton, WO 1992 Scanning Microscopy Supplement **6**, AMF O'Hare IL p. 405. Correcting for distortion in the specimen (Section 31.6B).
- Trus, BL, Unser, M, Pun, T and Stevens, AC 1992 Scanning Microscopy Supplement **6**, AMF O'Hare IL p. 441.

CHEMISTRY FROM IMAGES (SECTION 31.10)

- Kisielowski, C, Schwander, P, Baumann, FH, Seibt, M, Kim, Y, and Ourmazd, A 1995 *An Approach to Quantitative High-Resolution Transmission Electron Microscopy of Crystalline Materials* Ultramicrosc. **58**, 131–155.
- Ourmazd, A Baumann, FH, Bode, M and Kim, 1990 *Quantitative Chemical Lattice Imaging: Theory and Practice* Ultramicrosc. **34**, 237–255.

PATTERN RECOGNITION AND FILTERING ON INTERFACES

- King, WE and Campbell, GH 1993 *Determination of Thickness and Defocus by Quantitative Comparison of Experimental and Simulated High-Resolution Images* Ultramicrosc. **51**, 128–135.
- King, WE and Campbell, GH 1994 *Quantitative HREM Using Non-Linear Least-Squares Methods* Ultramicrosc. **56**, 46–53.
- Möbus, G, Necker, G and Rühle, M 1993 *Adaptive Fourier-Filtering Technique for Quantitative Evaluation of High-Resolution Electron Micrographs of Interfaces* Ultramicrosc. **49**, 46–65. The adaptive Fourier-filtering technique.
- Paciornik, S, Kilaas, R, Turner, J, and Dahmen, U 1995 *A Pattern Recognition Technique for the Analysis of Grain Boundary Structure by HREM* Ultramicrosc. **62**, 15–27. Pattern recognition at interfaces.

SPECIFICS

- Erickson, HP and Klug, A 1971 *Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs* Phil. Trans. Roy. Soc. London B. **261**, 105–118. An example from the Nobel Prize winner.
- Frank, J 1980 in *Computer Processing of Electron Microscope Images*, Springer-Verlag New York. p. 187. Correlation techniques (Section 31.12).
- Hýtch, MJ and Stobbs, WM 1994 *Quantitative Comparison of High Resolution TEM images with image simulations* Ultramicrosc. **53** 191–203. A very interesting conclusion on the thickness of the specimen.
- Koster, AJ, van den Bos, A and van der Mast, KD 1988 Scanning Microscopy Supplement **2** AMF O'Hare IL p. 83. Autoalignment of the TEM.
- Koster, AJ and de Juijter, WJ 1992 *Practical Autoalignment of Transmission Electron Microscopes* Ultramicrosc. **40** 89–107. Autoalignment of the TEM.
- Krivanek, OL and Mooney, PE 1993 *Applications of Slow-Scan CCD Cameras in Transmission Electron Microscopy* Ultramicrosc. **49** 95–108. Autoalignment of the TEM.
- Moré, JJ 1977, in *Lectures Notes in Mathematics*, Ed. G.A. Watson p. 630 Springer Berlin.
- Moré, JJ, Garbow, BS and Hillstrom, KE 1980 *User Guide for MINIPACK-1*.
- Saxton, WO and Koch, TL 1982 *Interactive Image Processing with an Off-Line Minicomputer: Organization, Performance and Applications* J. Microsc. **127** 69–83. Autoalignment of the TEM.
- Saxton, WO, Smith, DJ and Erasmus, SJ 1983 *Procedures for Focusing, Stigmating and Alignment in High Resolution Electron Microscopy* J. Microsc. **130** 187–201. Cross-correlating pairs of images.

- Tang, D, Jansen, J, Zandbergen, HW and Schenk, H 1995 *The Estimation of Crystal Thickness and the Restoration of Structure-Factor Modulus from Electron Diffraction: A Kinematical Approach* Acta Cryst. **A51** 188–197. (31.6E) Quantitative analysis of DPs.
- van Heel, M, Winkler, H, Orlora, E and Schatz, M 1992 Scanning Microscopy Supplement **6**, AMF O'Hare IL p. 23. Review of TEM of beam-sensitive materials.
- Zou, XD 1995 *Electron Crystallography of Inorganic Structures*, Chemical Communications Stockholm University, Stockholm, Sweden.

THE COMPANION TEXT

The topic of image processing is often associated only with HRTEM images. This aspect is the subject of an entire chapter on focal-series reconstruction in the companion text.

One point you must always keep in mind is that the data are what you recorded originally. You will usually modify these data in some way before it is reported to the scientific community. The question you must always keep in mind is how much detail on the processing must you report so that other scientists will be able to check and repeat your analysis.

SELF-ASSESSMENT QUESTIONS

- Q31.1 Define the term 'quantifying HRTEM images.'
- Q31.2 Why is Fourier filtering so named?
- Q31.3 Sketch a pair of diffractograms from an amorphous Ge film with $\Delta f = 1$ sch and 4 sch, respectively.
- Q31.4 Why is amorphous Ge used rather than amorphous C to support Au nanoparticles in testing the resolution of an HRTEM?
- Q31.5 If we plot $\frac{n}{u^2}$ against u^2 , what is the slope of the curve and what is n ?
- Q31.6 What is unsharp masking?
- Q31.7 How would you obtain the best TEM still image from a TEM video?
- Q31.8 What is a kernel in image processing?
- Q31.9 Give an example of a kernel for enhancing the contrast in image processing.
- Q31.10 Why do we say that the signal-to-noise ratio will always be small in low-dose TEM?
- Q31.11 Why is it helpful to record a focal series of images if you want to do quantitative HRTEM?
- Q31.12 Will Kikuchi diffraction influence electron crystallography using DPs?
- Q31.13 What is the reason for using a tilted-beam series to enhance resolution in HRTEM?
- Q31.14 What is the most important use of remote microscopy?
- Q31.15 We use a diffractogram tableau to bring the beam alignment to within 0.1 mrad of the optic axis. What aberration are we correcting?
- Q31.16 What is the use of pattern recognition in HRTEM and why would you like to have a new 4 k × 4 k camera for this purpose?
- Q31.17 What is the basic principle behind QUANTITEM?
- Q31.18 There are two methods for estimating the 'goodness of fit'? What are they and how do they differ?
- Q31.19 For the best quantitative data, you should use a focal series and low-dose imaging. What fundamental principle of quantum physics does this statement remind you of?
- Q31.20 What is the adaptive Fourier-filtering technique?

TEXT-SPECIFIC QUESTIONS

Some of the questions for this chapter require that you have access to image-processing software.

- T31.1 Identify another particle/matrix pair that might be studied like that in Figures 31.2 and 31.3. Describe how you would use image processing in your analysis to show it is a spherical particle, not a cylinder.
- T31.2 Describe how you would extract the best image from a 30-second video clip of a stable grain boundary. (Hint: use several sections in this chapter.)
- T31.3 Use your scanner and Digital Micrograph to check the correlation between the images and diffractograms in Figures 31.4 and 31.5. You can consider the original publications. Comment on your results and the scientific process.
- T31.4 By considering the original paper or just from scratch, reproduce the graphs shown in Figure 31.6A.
- T31.5 Explain how you would actually use the kernel for a digital Laplacian in analyzing a TEM image. What type of image might this be useful for? Find (in the literature) and describe two other kernels.
- T31.6 Low-dose microscopy should be used much more routinely. Describe in detail the routines that are actually used by two different cryo-TEM manufacturers.
- T31.7 Consider Figure 30.10. Summarize how you would carry out this type of study in your lab. Pay special attention to the important features of the specimen (e.g., is it applicable to all specimens).

- T31.8 By delving into the literature, write a description of how you would use a program like QUANTITEM. Identify other groups who have used similar programs. Give a critical review of the applications of the concept of parameterization in TEM (with special attention to the range of materials that can be studied).
- T31.9 Use a laser pointer and a small-mesh TEM grid to produce a diffraction pattern of the grid on a nearby wall. (See Figure 31.27.) Relate the spacing of the spots to the distances (e.g., mesh spacing) that you used.
- T31.10 Use Digital Micrograph, or a similar product, to reproduce the images shown in Figure 31.27 using your home town or a face as the object. Comment on the reduction of resolution as you decrease the aperture diameter.