30

Image Simulation

CHAPTER PREVIEW

When we need to obtain information about the specimen in two directions, we need to align the specimen close to a low-index zone axis. If the HRTEM image information is going to be directly interpretable, the specimen must be oriented with the incident beam exactly aligned with both the optic axis of the TEM and the zone axis of the specimen. Thus, we will have many reflections excited and the simple two-beam analysis of Chapter 27 cannot be used.

A method for modeling the contrast of images obtained under these conditions was developed by Cowley, Moodie, and their co-workers, principally at Melbourne and Arizona State University (ASU), in a series of classic papers beginning with that by Cowley and Moodie in 1957. Fortunately, the growing interest in HRTEM has coincided with the availability of increasingly powerful computers which can handle the extensive calculations.

There are several software packages available commercially, so there is little reason for most users to re-invent the wheel. However, the different packages do not necessarily perform the calculations in the same way: one may be more appropriate for your application than others. Since these packages essentially operate as 'black boxes' it is also reassuring to simulate images from the same structure using different packages (unless they don't give the same answers).

One point to keep in mind as you work through the literature is that this subject already has a lot of history. We will point out some of the things that have been done, and in some cases continue to be done, for historical reasons.

30.1 SIMULATING IMAGES

The idea of simulating HRTEM images arose because of the realization that the loss of phase information when we form an experimental intensity map means that we can't go back from the image to the structure. Instead, we assume a structure (perfect crystal or crystalline material containing defects), simulate the image, see how closely the simulated image resembles the experimental image, modify the structure, and repeat the process. The only difficulty is that the image is sensitive to several factors

- The precise alignment of the beam with respect to both the specimen and the optic axis.
- The thickness of the specimen (as we saw in Chapter 28).
- The defocus of the objective lens.
- Chromatic aberration which becomes more important as t, the thickness, increases.
- Coherence of the beam.
- Other factors: one example would be the intrinsic vibration in the material which we try to take account of through the Debye-Waller factor.

In principle, we could have the same image from two different structures. So obviously, this is the tricky part!

30.2 THE MULTISLICE METHOD

The basic multislice approach used in most of the simulation packages is to section the specimen into many slices, which are normal to the incident beam.

There are different methods for actually performing the multislice calculation. The different approaches have been developed for several reasons. Some try to optimize the use of available hardware. Others were written with the intention of providing a convenient method of simulating DPs using the same program. At least one package was written to make use of a popular

TABLE 30.1. Software

Cerius2	Runs on UNIX; by Accerlys. Look for c2_hrtem www.accelrys.com/products/cerius2/cerius2products/hrtem.html
EMS and jEMS	By Pierre Stadleman. Very widely used and user-friendly. Does Bloch wave and conventional multislice calculations. Is used by CuFour (see Chapter 24). Multi-platform including Mac, Unix, Windows. The 'j' denotes the java version
Kirkland	By Earl Kirkland. Well described in his book on the subject
MacTEMPAS	By Roar Kilaas. Runs on a Mac so it's very user-friendly. NCEMSS was the Unix version that was produced by NCEM
SHRLI81	By Mike O'Keefe. This program made image simulation freely available to everyone and introduced a generation to HRTEM simulation. Unix only, but no longer supported
WinHREM and MacHREM	www.hremresearch.com/Eng/download/documents/HREMcatE.html by HREM Research Inc. (Kazuo Ishizuka)

personal computer with a user-friendly interface (see Table 30.1). The principal methods for performing these calculations are

- The reciprocal-space formalism.
- The FFT formalism.
- The real-space approach.
- The Bloch-wave approach.

We'll go through the special features of each approach. The software packages which are readily available are listed in Section 1.6 and in Table 30.1.

30.3 THE RECIPROCAL-SPACE APPROACH

We project each slice onto a plane somewhere in the slice (usually top, bottom, or middle) giving a projected potential for that slice, and we call this the phase grating. We then calculate the amplitudes and phases for all the beams which will be generated by the incident beam interacting with the first projection plane. We could think of this as being a many-beam image calculation for a single slice. We then allow all these beams to propagate down the microscope in free space until they meet the next phase grating, which need not be the same as the first. The scattering calculation is now repeated for all the beams incident on this plane. This calculation produces a new set of beams which propagate through free space to the next phase grating, and so on. The process is summarized in Figure 30.1.

One point which you must remember: scattering by the phase grating does not just produce Bragg beams. It is crucial to keep track of the scattering in *all* directions. All of these beams will be incident on the next phase grating. So we don't just have Bragg beams, we *sample* all of reciprocal space.

A calculation based on a 128×128 array will impose a limit of ~4096 on the number of 'beams' which can be included in the calculation. This number might appear large, especially when you form a [110] HRTEM image of Si with six Bragg beams (plus the O beam) but, particularly for imperfect crystals, this number will be inadequate.

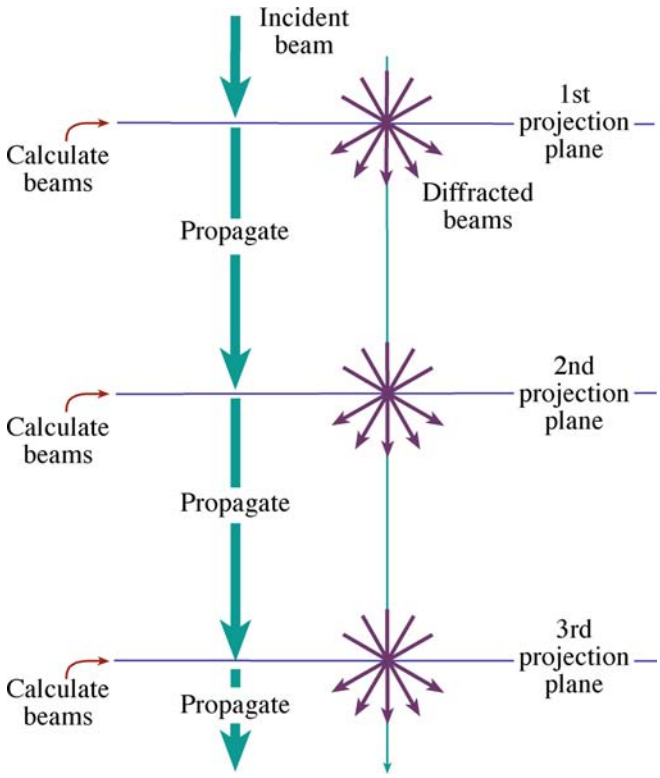


FIGURE 30.1. The potential within a slice is projected onto the first projection plane; this is the phase grating. We calculate the amplitudes and phases for all the beams generated by interacting with this plane and then propagate all the diffracted beams through free space to the next projection plane, and repeat the process.

k-SPACE BETWEEN THE BRAGG BEAMS

Why do we need to consider regions of **k**-space between the Bragg beams? In other words, why do we need to sample all of reciprocal space? The answer is that the Bragg beams contain information about the periodic structure, but all of the information from defects, i.e., non-periodic structure, is contained *between* the Bragg spots, though it will generally be quite close to them.

Essentially the multislice method considers three components

- ψ describes the *electron* wave.
- P is the propagation of the electron wave in free space: the *microscope*.
- Q is the phase grating: the *specimen*.

The process can be described by this equation

$$\psi_{n+1}(\mathbf{k}) = [\psi_n(\mathbf{k})P_{n+1}(\mathbf{k})] \otimes Q_{n+1}(\mathbf{k}) \qquad (30.1)$$

where $\psi_{n+1}(\mathbf{k})$ is the wave function in reciprocal space at the exit of the $n+1$ slice and the symbol $\otimes$ denotes a convolution as before. $P_{n+1}(\mathbf{k})$ is the propagator for the $n+1$ slice. In other words, this is expressing the Fresnel diffraction phenomenon for this one slice because we are making a near-field calculation. (Look back to Section 2.8 for a discussion of near-field versus far-field.) Similarly, $Q_{n+1}(\mathbf{k})$ is the phase-grating function; it is a transmission function, for the $n+1$ slice.

The three functions $\psi(\mathbf{k})$, $P(\mathbf{k})$, and $Q(\mathbf{k})$ are all functions in reciprocal space, so this approach is referred to as the reciprocal-space formulation. Notice that they are all two-dimensional arrays. We can think of the different terms as being diffracted beams within the specimen. We can easily insert a circular objective aperture of radius $\mathbf{r}$; we just require that all values of $\psi(\mathbf{k})$ are zero for $\mathbf{k} > \mathbf{k_r}$.

To give you an idea of the complexities involved, consider what values of $Q(\mathbf{k})$ you must use in the calculation. $Q(\mathbf{k})$ must go out twice as far as $\psi(\mathbf{k})$ or $P(\mathbf{k})$ in reciprocal space. You can understand why by considering Figure 30.2. If you represent the number of

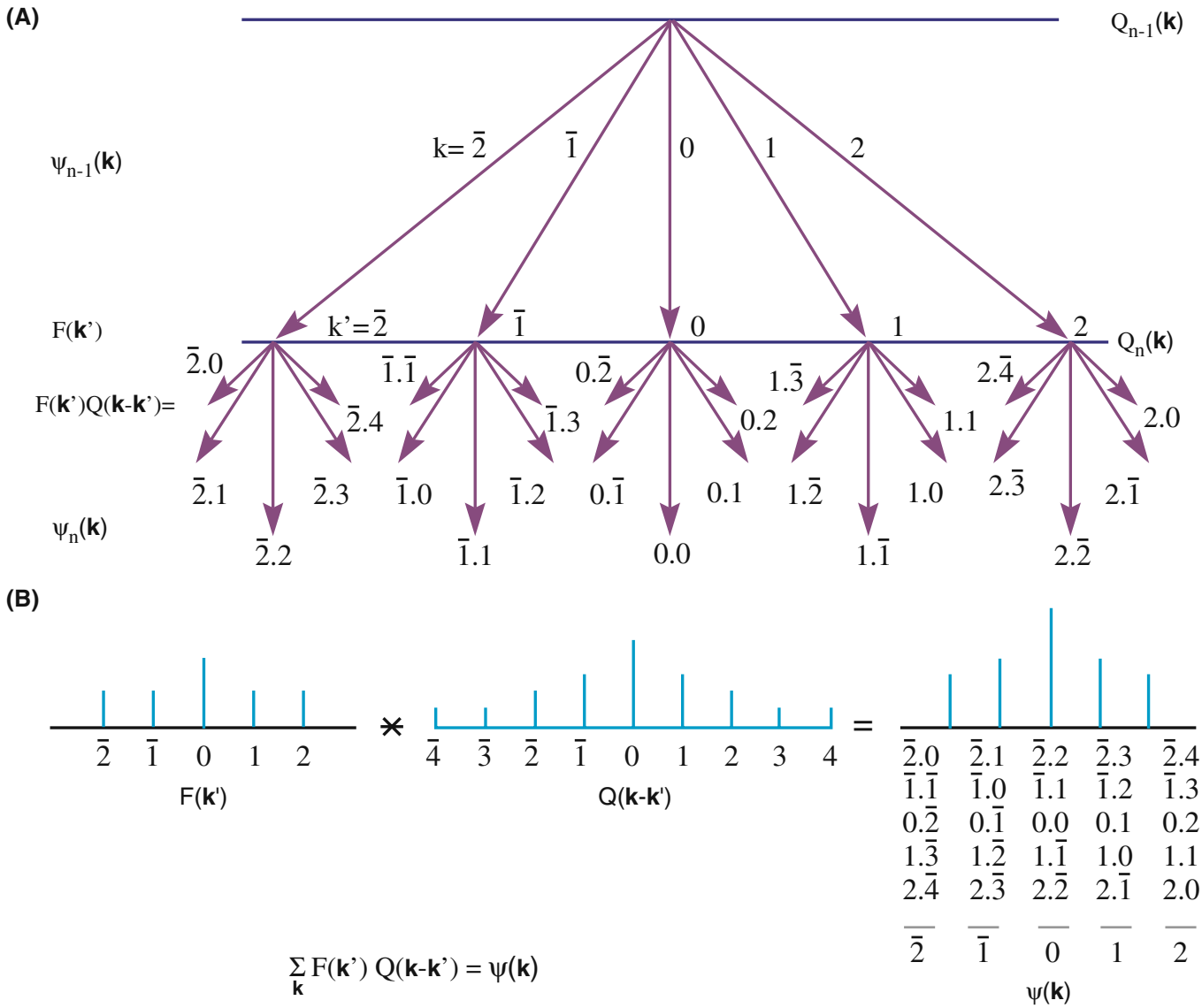


FIGURE 30.2. (A) Schematic used to explain why, in the one-dimensional case, $Q(\mathbf{k})$ must take account of twice as many $\mathbf{k}$ values as $\psi(\mathbf{k})$ or $P(\mathbf{k})$. Consider wave $k = \bar{2}$ from $Q_n(\mathbf{k})$: to produce wave at $+2$ at this point you need to add 4 to $\bar{2}$ and similarly for every possible wave in slice $Q_n(\mathbf{k})$. As summarized in (B) $Q(\mathbf{k}–\mathbf{k}')$ extends from $\bar{4}$ to $+4$ so that $\psi(\mathbf{k})$, which we want, extends from $\bar{2}$ to $+2$, including all possible combinations of $\mathbf{k}'$ and $\mathbf{k}$.

beams from slice $Q_{n-1}(\mathbf{k})$ as $F(\mathbf{k}')$ then $Q(\mathbf{k}–\mathbf{k}')$ must go out to $k = -4$ because when you multiply these two functions to give $\psi(\mathbf{k})$ you can produce $k = -2$ by using $k = -4$ in Q and $k = +2$ in F as in Figure 30.2B. Putting this into an equation we have

$$\sum_{\mathbf{k}'} F(\mathbf{k}')Q(\mathbf{k}-\mathbf{k}') = \psi(\mathbf{k}) \qquad (30.2)$$

where

$$F(\mathbf{k}) = \psi(\mathbf{k})P(\mathbf{k}) \qquad (30.3)$$

The function $Q(\mathbf{k})$ is a 'probability map.' What we are doing here is using the convolution to describe multiple scattering.

We can illustrate the complexity of the calculation by considering a 128 × 128 array for $Q(\mathbf{k})$ using SHRLI81 (see Table 30.1). The maximum value for (k_x,k_y) is only (31,31) but even so, the number of diffracted beams is nearly 4096. Remember, we usually just use the seven inner Bragg beams in, e.g., the Si <110> DP, as we saw in Figure 27.3; most of the beams in our calculation are not Bragg beams. However, you will remember that the information concerning defects in crystals is contained in the regions between the Bragg spots in the DP, so it does make sense. Specific examples of $Q(\mathbf{k})$, including numerical computations of the phase change per slice, are given by John Barry.

30.4 THE FFT APPROACH

We can recast equation 30.3 to maximize the efficiency of the computer in using fast Fourier transform (FFT) routines. In equation 30.4, the notation F and F^{-1} tell us to take the Fourier transform or the inverse transform of the function inside the brackets.

$$\psi_{n+1}(\mathbf{k}) = F\{F^{-1}[\psi_n(\mathbf{k})P_{n+1}(\mathbf{k})]q_{n+1}(\mathbf{r})\} \qquad (30.4)$$

In this equation, $q_{n+1}(\mathbf{r})$ is the real space form of $Q_{n+1}(\mathbf{k})$, i.e., it is the inverse Fourier transform of $Q_{n+1}(\mathbf{k})$. So $q(\mathbf{r})$ is a real-space phase grating. Now we can look at some numbers for the calculation and take $Q(\mathbf{k})$ as a 128 × 128 array to keep the calculation small. The main steps carried out by the computer are

- Multiply $\psi_n(\mathbf{k})$ by $P_{n+1}(\mathbf{k})$: that is a 64 × 64 array times another 64 × 64 array. Remember that we are limited to 64 points, if we have a 128 × 128 array, because the Q array must be twice as large in all directions in $\mathbf{k}$ space.
- Take the inverse Fourier transform of the result.
- Multiply this new result by $q_{n+1}(\mathbf{r})$ which is the 128 × 128 array.
- Fourier transform the final result and set all values outside the inner 64 × 64 array equal to zero so that we can repeat the process for the next slice.

You will notice that this example used a square array. In modern programs, we are not restricted even to using powers of 2 but this helped the original FFT routines. You will see the value of this advance when we examine some defect calculations later. If you are interested in the mechanics of the FFT routine and other aspects of this simulation approach, the article by Mike O'Keefe and Roar Kilaas is required reading.

30.5 THE REAL-SPACE APPROACH

As we noted earlier, image simulation used to be limited by your budget, i.e., by your computer. The real-space approach was developed, in part, to decrease the time needed for the calculations by using our knowledge that $P(\mathbf{r})$ is strongly peaked in the forward direction. In our notation, the method developed by Coene and Van Dyck for calculating $\psi(\mathbf{r})$ can be expressed by the equation

$$\psi_{n+1}(\mathbf{r}) = [\psi_n(\mathbf{r}) \otimes P_{n+1}(\mathbf{r})]q_{n+1}(\mathbf{r}) \qquad (30.5)$$

where $P_{n+1}(\mathbf{r})$ is now the propagator in real space and $q_{n+1}(\mathbf{r})$ is again the real-space phase grating. Once you have written this, it's all computing, which is a substantial task since the size of the multislice calculation is the size of the largest array, i.e., $Q(\mathbf{k})$ or $q(\mathbf{r})$.

30.6 BLOCH WAVES AND HRTEM SIMULATION

Although we saw in Chapters 14 and 15 that electrons propagate through crystalline specimens as Bloch waves, the multislice method we've described so far is essentially a 'diffracted-beam' approach. In two classic papers Fujimoto (1978) and Kambe (1982) showed that, for the perfect crystal, the HRTEM image may be understood simply in terms of images of Bloch waves. The key point is that, although a large number of diffracted waves are formed, only a small number of Bloch waves determine the appearance of the image, providing the crystal has a sufficiently high symmetry. Following Kambe's 'simple' example, we consider the case where only three Bloch waves i, j, and k are significant. Let's assume that Bloch waves i and j are in phase at a thickness $z = D$. Then we have

$$e^{ik_z^{(i)}z} = e^{ik_z^{(j)}D} \qquad (30.6)$$

Using our expression for ψ, namely

kTH BLOCH WAVE
Don't confuse the kth Bloch wave with the **k**-vector!

$$\psi(\mathbf{r}) = \sum_i C^{(i)}\phi^{(i)}(x,y)e^{ik_z^{(i)}z} \tag{30.7}$$

and the normalization rule

$$\sum_i C^{(i)}\phi^{(i)}(x,y) = 1 \tag{30.8}$$

we can therefore express ψ at $z = D$ in terms of our three Bloch waves.

$$\psi(x,y,D) = \left[C^{(i)}\phi^{(i)} + C^{(j)}\phi^{(j)}\right]e^{ik_z^{(i)}D} + C^{(k)}\phi^{(k)}e^{ik_z^{(k)}D} \tag{30.9}$$

We rearrange this equation so that we can extract the phase factor $e^{ik_z^{(i)}z}\left(= e^{ik_z^{(i)}D}\right)$. We write

$$\psi(x,y,D) = \left[1 - C^{(k)}\phi^{(k)}\right]e^{ik_z^{(i)}D} + C^{(k)}\phi^{(k)}e^{i\left(k_z^{(k)}-k_z^{(i)}\right)D}e^{ik_z^{(k)}D} \tag{30.10}$$

$$\psi(x,y,D) = e^{ik_z^{(i)}D}\left[1 + \beta_{ik}(D)C^{(k)}\phi^{(k)}\right] \tag{30.11}$$

where we've defined a new parameter β given by

$$\beta_{ik}(D) = e^{i\left(k_z^{(k)}-k_z^{(i)}\right)D} - 1 \tag{30.12}$$

These equations tell us that if any two of the Bloch waves (here they are i and j) are in phase then the amplitude of the wave at the exit surface is determined by the third Bloch wave.

If the third Bloch wave is also nearly in phase, we have a relation like equation 30.6 but with i, j, and k all equal. Then we can approximate $\beta_{ik}(D)$ by

$$\beta_{ik}(D) = i\left[\left(k_z^{(k)} - k_z^{(i)}\right)D + 2n\pi\right] = i\gamma_{ik}(D) \tag{30.13}$$

Now we've defined another factor γ_{ik}. If you plug this expression back into equation 30.11, you see we have a pure phase object. All the diffracted beams will be shifted in phase by $\pi/2$.

Now you can test the effects of how we change k. Consider what conditions this will really correspond to using equations 30.11 and 30.13

- If k is such that the phase of Bloch wave k is ahead of i and j (which were equal), then you'll see a 'negative' image of $C^{(k)}\phi^{(k)}$. A 'delayed' k gives us the 'positive' image.
- For the Ge <110> zone axis, HRTEM image at 100 kV, only three Bloch waves are strongly excited.

The relationship to the Bloch-wave contours in Chapter 14 is clear. Using this information and the projected potential shown in Figure 30.3, Kambe calculated the Bloch-wave amplitudes and the two ideal images of the Bloch waves: one positive and the other negative. In the calculation of different images for increasing thickness, several images corresponding to a single Bloch wave can be predicted and identified, as shown in the figure. At other thicknesses, the images form by a combination of Bloch waves. So, what can we learn?

- For a perfect crystal, you may need as few as three Bloch waves to give the essential features of an HRTEM zone-axis image.
- There is a direct connection between the WPOA (go back to the definition if you don't remember) and the propagation of Bloch waves.

THE MULTISLICE APPROACH
We usually use this to simulate HRTEM images. It is actually a very elegant form of brute force.

We saw in Chapter 14 that the electron propagates as Bloch waves inside the crystal. The reason we don't use Bloch waves is that our specimens are not perfect. However, EMS does give you the option of using this approach.

30.7 THE EWALD SPHERE IS CURVED

When you are using the TEM, some other complications arise because the Ewald sphere is curved

- If you align the beam exactly parallel to a zone axis, **s** will be non-zero for every Bragg reflection. In fact, it will also be different for each type of reflection.
- If you do not align the beam exactly parallel to the zone axis, then **s** will also be slightly different for each reflection in that zone.
- If you change the wavelength of the electrons, the radius of the sphere changes.
- If you converge the beam, then you'll add a thickness to the Ewald sphere.

The point is, knowing precisely what the correct values are to put in the program will also require thought and work.

30.8 CHOOSING THE THICKNESS OF THE SLICE

So far, we've just cut the specimen into slices in the computer without considering how thick each slice should be, or even whether they should all be the

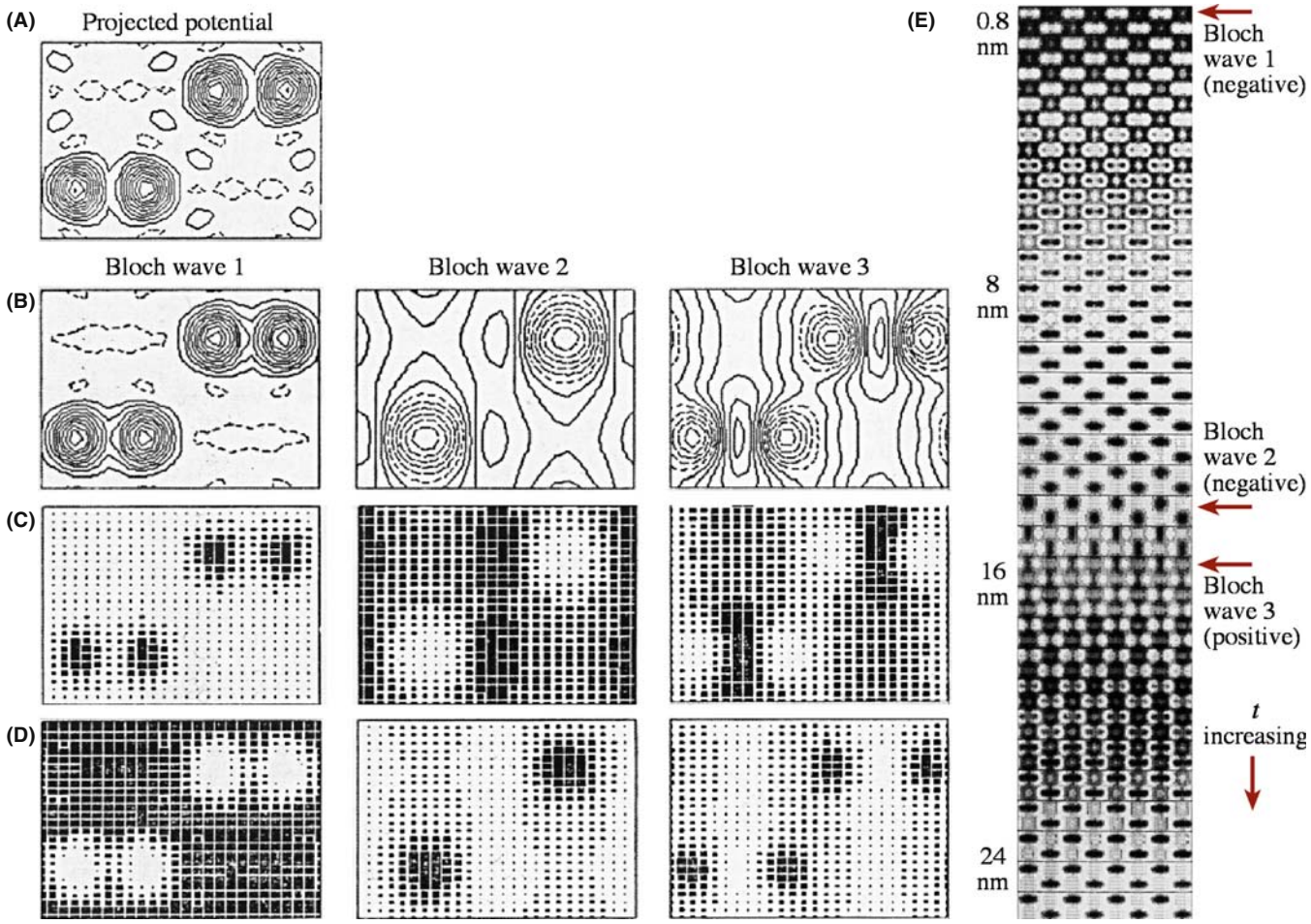


FIGURE 30.3. (A) The projected potential for Ge where the contour lines represent changes in potential of –10 eV, and the dashed lines are positive values. (B–D) are for the three Bloch waves 1, 2, and 3 for 100-keV electrons: (B) the amplitudes, (C) ideal positive images, (D) ideal negative images. (E) Calculated lattice images showing thickness dependence.

same. If all the slices are the same, then there can be no information about the z-direction. Although HOLZ lines are not important for the simulation of HRTEM images, some of the programs we are discussing can now just as readily be used to simulate CBED patterns and HOLZ lines. So, following the philosophy of attacking problems with different techniques, you should be aware of these limitations, since it is easy to overlook the simplifications you made once you see the computed image. You should remember that when you are studying a material with a large unit cell, the reciprocal lattice spacing will be short in the beam direction, so HOLZ effects come into play sooner.

Consider the different methods for making the slice

- You could calculate the projected potential for a thick slice and then do n calculations with slices which are $1/n$ times this thickness.
- A better approach would be to subdivide the cell into layers of atoms, create a different grating for each of these layers, and then run the program with the sequence.

For example, if the beam is aligned along the [111] direction of an fcc crystal, then you would have three identical gratings displaced relative to one another, giving the ABC stacking of close-packed planes. This approach would allow you to test for the effect of a real error in the stacking sequence normal to the beam. Even this point can be a bit difficult. In general, you orient the beam to be parallel to a particular zone axis $[UVW]$ so that the planes in that zone are parallel to the beam (so our projection works). If the material is not cubic, we will not generally have a low-index plane normal to the beam to make this slice.

30.9 BEAM CONVERGENCE

When you are recording HRTEM images, you need to keep exposure times short. So, if you don't use parallel illumination, you have to take account of the beam convergence when simulating the images. O'Keefe and Kilaas (see also the paper by Self and O'Keefe) have developed one approach to address this problem. If

the beam actually has some convergence, then the diffraction spots will be disks, as illustrated in Figure 30.4, so you need to simulate disks in the DP. Experimentally, the large objective aperture admits many disks, so in the simulation routine you should sample each disk at many points. This means the program needs to calculate the image at each of these convergence angles and average all the resulting images. Of course, the objective aperture is easily applied in the computer. If you choose 49 points, you can make the sampling interval in reciprocal space $\leq 0.1\,\mathrm{nm}^{-1}$. It is instructive to examine just how much work is necessary to sample the 49 points.

We can start by writing the usual expression for the χ, the phase change due to the objective lens

$$\chi = \pi \Delta f \lambda u^2 + \pi C_s \lambda^3 \left(\frac{u^4}{2}\right) \tag{30.14}$$

Then differentiate this with respect to the variable u

$$\frac{d\chi}{du} = 2\pi\left(\lambda u \Delta f + C_s \lambda^3 u^3\right) \tag{30.15}$$

This equation tells us that if u changes by δu, then χ changes by

$$\delta\chi = 2\pi\lambda\left(u\Delta f + C_s\lambda^2 u^3\right)\delta u \tag{30.16}$$

Now we choose $\delta\chi$ so that

$$\delta\chi < \frac{2\pi}{n} \tag{30.17}$$

where n will allow us to determine the maximum change in χ between two points in the disk. For example, if $n = 12$, then the maximum value of $\delta\chi$ is 30°. Combining equations 30.15 and 30.17, we can write

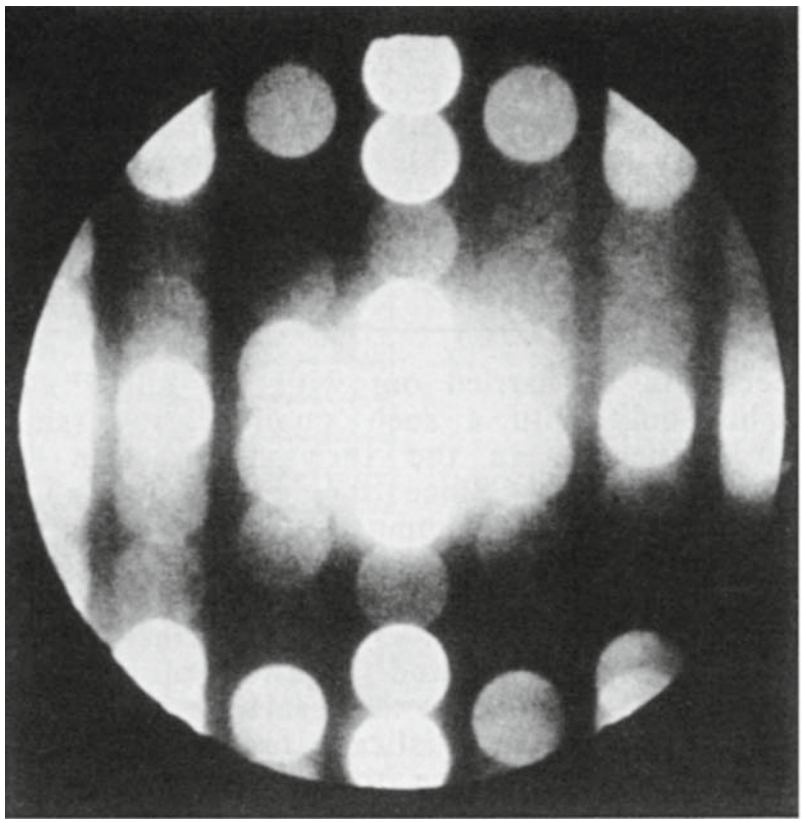
$$\delta u = \left[n\lambda u\left(\Delta f + C_s\lambda^2 u^2\right)\right]^{-1} \tag{30.18}$$

If we plot χ versus u (or play with equation 30.15 and its derivative), then we find a minimum in χ at

$$\Delta f = -C_s\lambda^2 u^2 \tag{30.19}$$

and an inflection at

$$\Delta f = -3C_s\lambda^2 u^2 \tag{30.20}$$



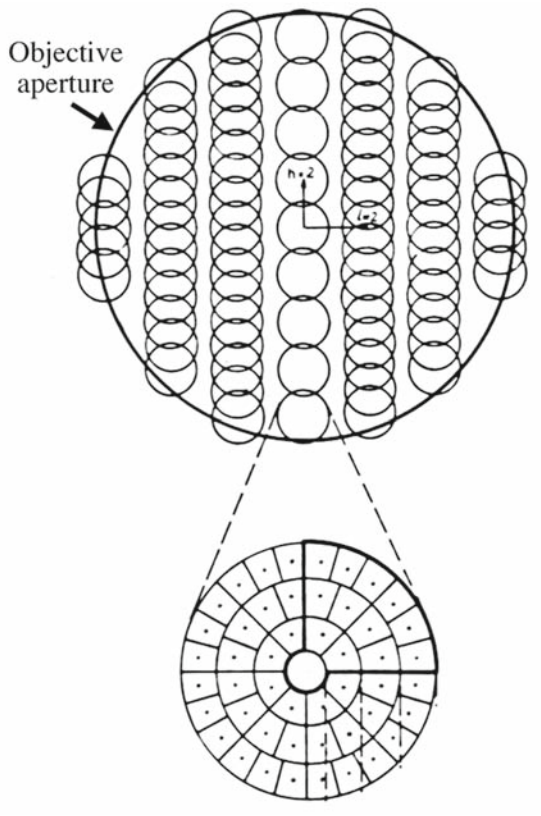


FIGURE 30.4. Disks in the DP from a crystal of $Nb_{12}O_{29}$. The computer simulation can divide each disk into many sectors and simulate the image for each sector, as shown in the schematic, excluding sectors which are intersected by the objective aperture.

So the simulation program can check to find the smallest δu at an inflection point, which equations 30.18 and 30.20 tell us is

$$\delta u = -\left[\frac{27C_s}{(\Delta f)^3}\right]^{1/2}\left(\frac{1}{2n}\right) \qquad (30.21)$$

The value of δu therefore depends on both C_s and Δf.

THE BLACK BOX
Remember where all these calculations takes place!

You can also appreciate the relevance of this type of approach if your disks actually intersect the objective aperture, as shown in Figure 30.4. Put another way, you can learn two lessons from this analysis

- Always try to minimize the convergence of the beam when recording HRTEM images.
- Use an aperture which does *not* cut through the diffraction disks.

30.10 MODELING THE STRUCTURE

To simulate any HRTEM image, you need a unit cell. If you are concerned only with perfect crystals, then your program should have all the space groups already included so that you only need to add the lattice parameters (lengths and angles) and the occupied sites for your material. If you are interested in simulating images from defects, then you have to create a new unit cell which must be sufficiently large that it will not add effects due to the edges. There are many ways to create this defect unit cell. You can input from other programs, such as those performing atomistic modeling of defects, or create your own starting structure. In either case, you will need to move atoms, either manually or following a rule you've selected for image matching, to optimize the match between your experimental series of through-focus images and the simulated images.

At some stage, you will find it useful to combine different slices, as when simulating GBs with or without a surface groove, or modeling large complex unit cells using a multilayer approach. We'll now go through some specific features of this task and return to modeling in Chapter 31 when we discuss quantitative HRTEM.

30.11 SURFACE GROOVES AND SIMULATING FRESNEL CONTRAST

The analysis of interfaces by the Fresnel-fringe technique, which we introduced in Chapter 23, illustrates the importance of image simulation and emphasizes that it is not just for HRTEM. The calculation is complicated for several reasons, as shown in Figure 30.5A

- The potential change at the interface is probably not abrupt.
- The potential depends on the detailed structure of the interface.
- During preparation, TEM specimens may be preferentially damaged at GBs, giving rise to surface grooves.

If you use a thicker specimen, you'll reduce the effect of surface grooves on any Fresnel fringes, but in practice, your foil thickness is usually limited (~20 nm), since you need to view the boundary exactly edge-on.

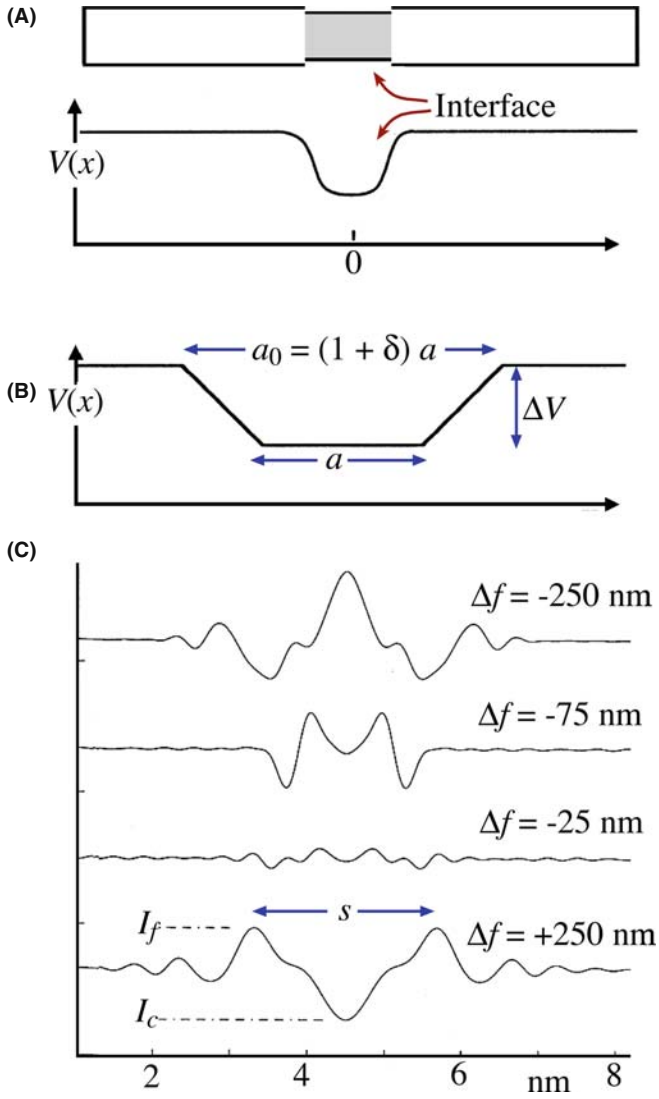


FIGURE 30.5. (A) Schematic of a GB containing a layer of material with a different inner potential; (B) one model used to represent such a GB giving variable parameters a, a_0, and δ; (C) a typical set of simulated Fresnel-fringe intensity profiles at increasing Δf: s is the distance between the first two fringes, I_c and I_f are the intensities of the central and first fringe, respectively.

Even for foils this thick, surface grooves can influence the projected potential considerably. If we assume that the bulk has a mean inner potential $V = 20$ V, and take a typical potential drop for an intergranular film to be $\sim$1 V, then the total projected potential drop for a 20 nm thick foil would be the same as that caused by a pair of grooves at the top and bottom surfaces which are only 0.5 nm deep. Although the surface groove may be partly filled with a second phase, the effect on the Fresnel fringes can still be substantial.

We can examine Fresnel fringes using different methods. In all of them, we describe the potential at the interface in terms of the projected potential drop $\Delta V_p = t\Delta V$, an inner width a, and an outer width a_0, and a 'diffuseness,' δ, defined by

$$a_0 = (1 + \delta)a \tag{30.22}$$

These parameters are shown in Figure 30.5B. Then we construct models of a foil with a surface groove at the edge-on interface by combining such potentials.

The models: Values of $\delta = 0.5$ and $\delta = 0.2$ represent shallow and steep surface grooves, respectively. The total projected potential drop can be due to a real change in V or a change in t. A groove without a film implies $a = 0$. If $a = 1$ nm and $a_0 = 1.5$ nm, the model could correspond to two different situations

- If the atoms at the interface relax, then the atomic density at the interface will usually be reduced. This occurs at both structured interfaces and those where a layer of glass is present.
- Surface grooves at the interface.

What image simulation shows is that the relative shapes and sizes of these models are more important than the actual dimensions. Therefore, we can give most of the following analysis in terms of dimensionless quantities. Inner potentials are typically 5–10 eV. Except for very small defocus values ($\Delta f < \approx t$), we find that the distribution of the potential through the foil is not important. Usually, the projected potential at the interface is lower than that in the bulk. However, the opposite situation can occur, e.g., when a Bi_2O_3-rich phase is present at interfaces in ZnO. When we discuss the calculated profiles, the term 'interface width' will be used for the parameters a and a_0, whether they actually correspond to an intergranular film, a surface groove, or otherwise.

SPACING OF FRESNEL FRINGES

The distance from the edge to the first fringe is proportional to $(\lambda\Delta f)^{1/2}$. The fringe spacing, s_f, can be extrapolated to zero defocus to obtain the interface width, based on the relation $(s_f - a) \propto \sqrt{(\lambda\Delta f)}$.

This relation was first described by Clarke and only holds when a is large and Δf is relatively small; then the fringes from each 'edge' at the interface are independent. We observe the minimum fringe spacing at small values of defocus and this spacing can be used to provide a measure of the interface width. For more details on the simulation of Fresnel fringes, you should check the original articles.

In practice, the analysis of Fresnel fringes is impaired, not only by specimen artifacts such as surface grooves, but also by various sources of noise, which all add to the uncertainty of measurements, especially at small values of Δf. For diffuse interfaces, the contrast decreases rapidly as Δf approaches zero (Figure 30.5C), and measurements of the fringe spacing for $\alpha > 0.7$ are increasingly susceptible to noise and artifacts. You can always use larger defocus values and thus obtain higher contrast. However, without prior knowledge about the shape of the potential drop (its 'diffuseness'), we can't reliably determine the interface width by measuring the fringe spacing alone. Since the fringe spacing is dominated by the outer width, a_0, we may easily overestimate the interface width. The atomic density in a region close to the boundary is also often reduced, even if the boundary is structured, so you can easily misinterpret the image as showing the presence of an intergranular film when actually there is no amorphous film.

The region of defocus, where the central fringe shows little contrast, provides complementary information to the fringe spacings, so it is more sensitive to the inner width.

From this discussion, you'll appreciate that, before you can completely understand the effect of any intergranular films, you must estimate the extent to which surface grooves are present in your specimen. Shadowing (e.g., using platinum or gold) may provide evidence for surface grooves, but in the case where the surface groove is already filled (e.g., if your specimen was coated with carbon or contaminated without your knowing), this technique won't work.

CHARACTERIZING A POTENTIAL

The conclusion is that you must use all the information in the image to characterize the shape of a potential well and you need to know what is in it!

To summarize, this discussion gives us a method for analyzing Fresnel fringes from a GB. We can draw some conclusions

- To interpret the contrast from Fresnel fringes at GBs, you must simulate images of many different interface models. In particular, it is essential that you

consider the possibility of artifacts such as surface grooving. Even a rather 'flat' or diffuse surface groove may influence the fringes in some range of defocus values.

- Both the fringe spacing and the central fringe intensity depend on the shape of the potential well and are sensitive to surface grooving.
- The interface width, which you can infer from the fringe spacing, is dominated by the outer width of a diffuse interface.
- A direct match with the $s_f - a$ curve (or with similar simulated curves when the assumptions employed here fail) leads to a better estimate of the average interface width, but cannot give you much information on the shape of the potential well.
- Determining when the central fringe is weak (the range Δf) gives complementary information on the interface width which, in combination with the estimate based on the fringe spacing, you can use to evaluate the diffuseness of the potential well.

30.12 CALCULATING IMAGES OF DEFECTS

When we simulate HRTEM images of perfect crystals, we only need to input the unit cell and the program generates the rest of the specimen. If we want to calculate the image of a defect, we have to use the same approach: we set up a unit cell to contain the defect and the program treats it like any other unit cell. This is known as the *periodic continuation* method for defect calculation. What we've actually done is shown in Figure 30.6: there is an array of defects throughout our specimen in all directions. We need to know two things

- To what extent does this ordered array introduce artifacts in the image?
- Have we created interfaces where the 'cells' join which may influence our image?

An example of a supercell for a GB is shown in Figure 30.6. This figure illustrates clearly how we can create a cell which is more suitable for this periodic continuation by including two defects in a single supercell. As shown in this figure, the periodic continuation then not only creates many other GBs but also makes them very long. If we don't match the crystals exactly at the edges of the supercell we create a different 'ghost' boundary.

SIMULATING DEFECTS
The periodic continuation is most widely used. Real-space patching avoids the artificial cell.

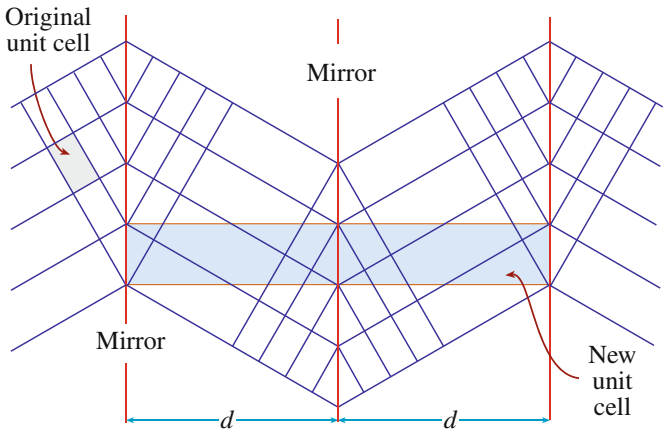


FIGURE 30.6. The periodic continuation technique illustrating how an artificial unit cell can be constructed to contain two GBs, thus allowing the HRTEM image to be simulated. The distance (d) between the two interfaces can be varied to check for overlap artifacts.

You can see that this really can be a problem by considering the DP which our new cell would produce. We are calculating the image of a small part of a periodic array of interfaces. Periodic arrays in real space produce rows of extra spots in reciprocal space. If we include these spots in forming the image, we should change the image. The solution for image simulation is quite simple, make the supercell wider and wider until the change in the image detail is less than some specified limit. However, don't try to interpret the data in the calculated DP without consulting the paper by Wilson and Spargo.

An alternative approach to the periodic continuation approach has been developed by Coene et al. and is called the real-space patching method. This method uses the 'real-space' image simulation approach to perform the calculation. The structure you want to simulate can be divided into a number of different 'patches' as illustrated in Figure 30.7; the image from each patch is

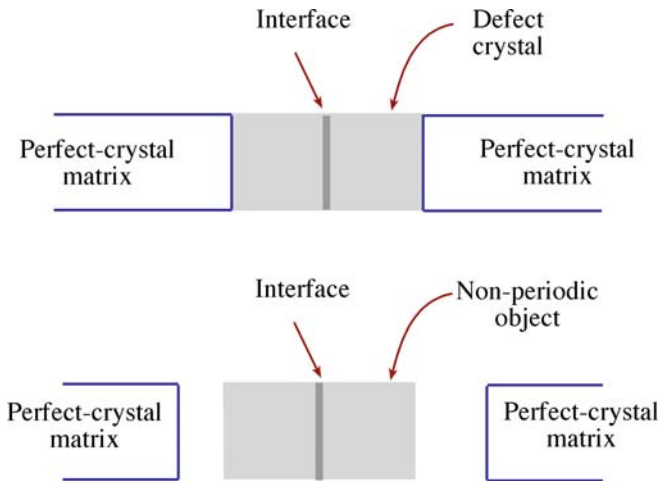


FIGURE 30.7. In the real-space patching method, the defect crystal (in this case the interface and several adjacent layers) is a non-periodic object that is surrounded by a perfect crystal matrix.

calculated for a slice and then the patches are joined together. The key, of course, is that you must correctly take account of what happens at the edge of each patch. This means each patch needs some information about the neighboring patches. Assuming (correctly) that this can be done, you can appreciate the nice feature of this approach: we avoid the artificial interference effects due to the array of defects that would be produced by the periodic continuation technique. The defect does not now 'see' its own image; it only sees the perfect matrix on all sides.

30.13 SIMULATING QUASICRYSTALS

There are several problems in simulating HRTEM images of quasicrystals, not least of which concerns which model you should use. (Remember that quasicrystals are not periodic objects anyway!) Several models have been reviewed by Shoemaker, and the possibilities are illustrated by the work of Beeli and Horiuchi, who used a combination of 10 layers in the multislice calculation. The layers are made up from the planar structures shown in Figure 30.8. The final structure (shown in Figure 30.8A) is made up of two sets of five layers. The first set of layers is B-C-D-C-B in this figure. The second set of five layers is constructed from the first by using the screw symmetry of the structure; the screw axis has 10_5 screw symmetry. The supercell used was 3.882 nm by 3.303 nm which was chosen to contain a complete decagonal cluster which is 2.04 nm in diameter and the center part of a pentagon tile. The calculation was then carried out for thicknesses up to 10 nm.

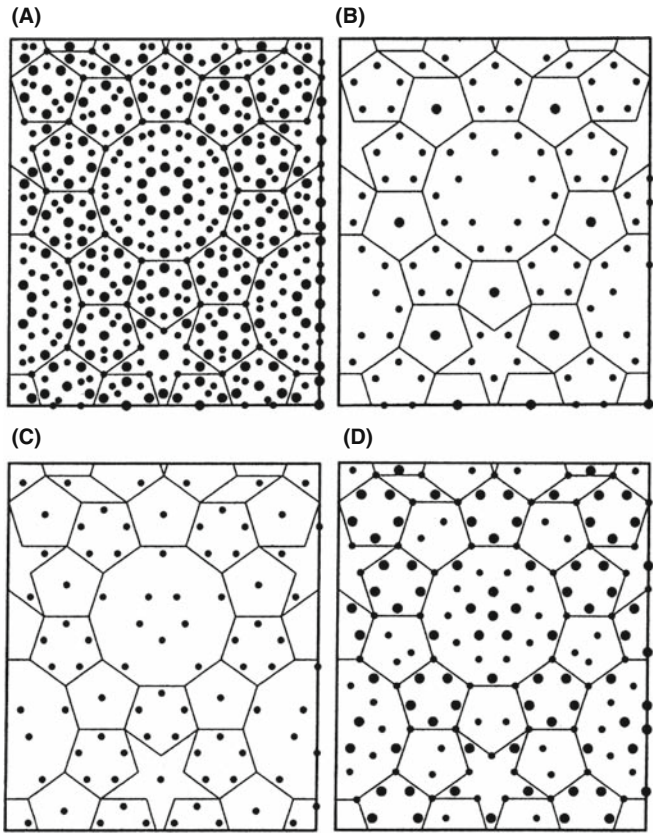


FIGURE 30.8. Projections used to simulate images of $Al_{70}Mn_{17}Pd_{13}$ quasicrystals. (A) Combination of all the layers; (B–D) layers are used to contribute (repeatedly) to (A). The edges of all the tiles are 0.482 nm. The large circles denote Al atoms.

Z-CONTRAST

Since the atoms lie along columns for phase-contrast imaging we can also use *Z*-contrast imaging.

The results of such calculations with only Al and Mn atoms are illustrated in Figure 30.9. The edges of the cells are essentially artificial because, as we just noted, the structure used in the calculation is a 'unit cell' and a quasicrystal doesn't have one. In spite of these difficulties Beeli and Horiuchi could conclude that the image match was much improved when Pd atoms were included to replace some of the Mn atoms in the D layer and Al atoms in the B-C layers with the results shown in Figure 30.10.

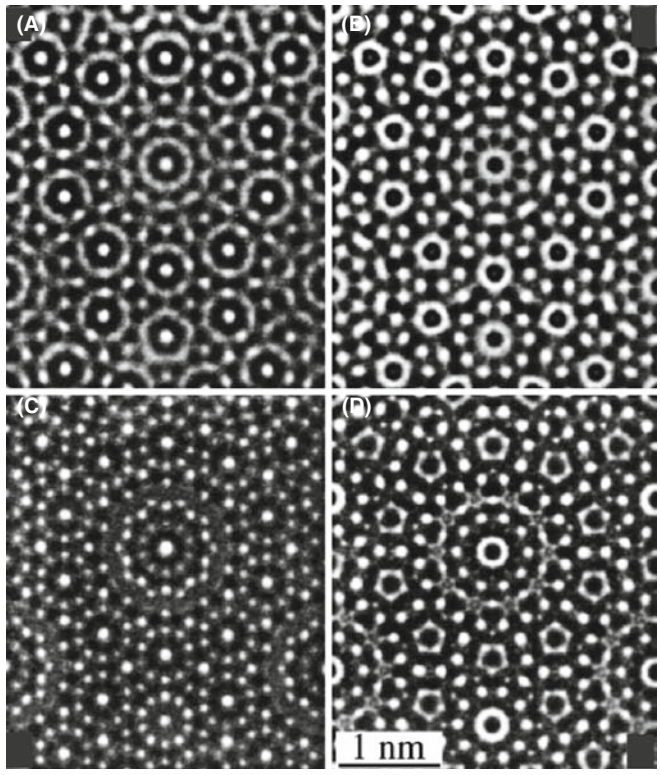


FIGURE 30.9. Four simulated images of the model constructed from the layers shown in Figure 30.8 using only Al and Mn atoms. The thickness is 3.77 nm, which corresponds to three periods in the beam direction. The values of Δf are (A) 0 nm, (B) 46 nm, (C) 88 nm, and (D) 124 nm.

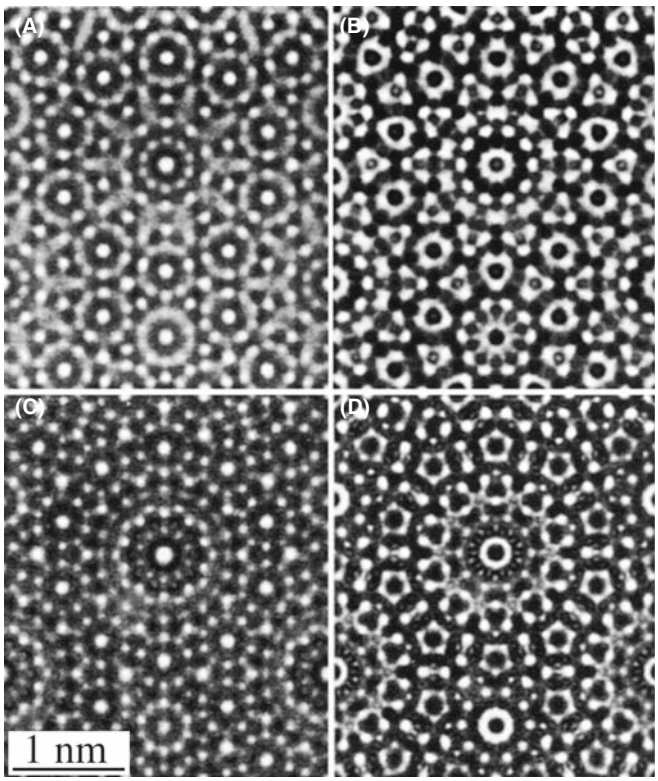


FIGURE 30.10. Examples of simulated images of the quasicrystal shown in Figure 30.8 but substituting Pd atoms for Mn atoms. The values of Δf are (A) 0 nm, (B) 48 nm, (C) 88 nm, and (D) 128 nm.

SIMULATING IMAGES OF A QUASICRYSTAL

As you realize, the problems in such an image calculation are increased because the quasicrystal does not have translational symmetry but we must impose such a symmetry to do the calculation.

Another illustration of the success of HRTEM comes from the work of Jiang et al. on quasicrystals with eightfold symmetry. Here the multislice calculation could again be made using a relatively simple sequence of four layers ABAB′, where the layers are at $z = 0$, 0.25, 0.5, and 0.75. The structures of the A and B layers are shown in Figure 30.11 with the B′ layer being a 45°-rotated B layer; i.e., the B and B′ layers are again related by a screw axis, this time, it's an 8_4 screw axis.

- In each of these examples, it is possible to view the same structure parallel to an orthogonal axis.
- Quasicrystals do not have translational symmetry, but we pretend they do for thickness calculations and for the periodic continuation of the unit cell.

Our reason for showing so much detail on these rather esoteric materials is that they show what can be done using image simulation. Furthermore, they emphasize the important fact that although we can construct the crystal using different layers and different sequences of layers, we always use a projection of the structure, to compare with the experimental image.

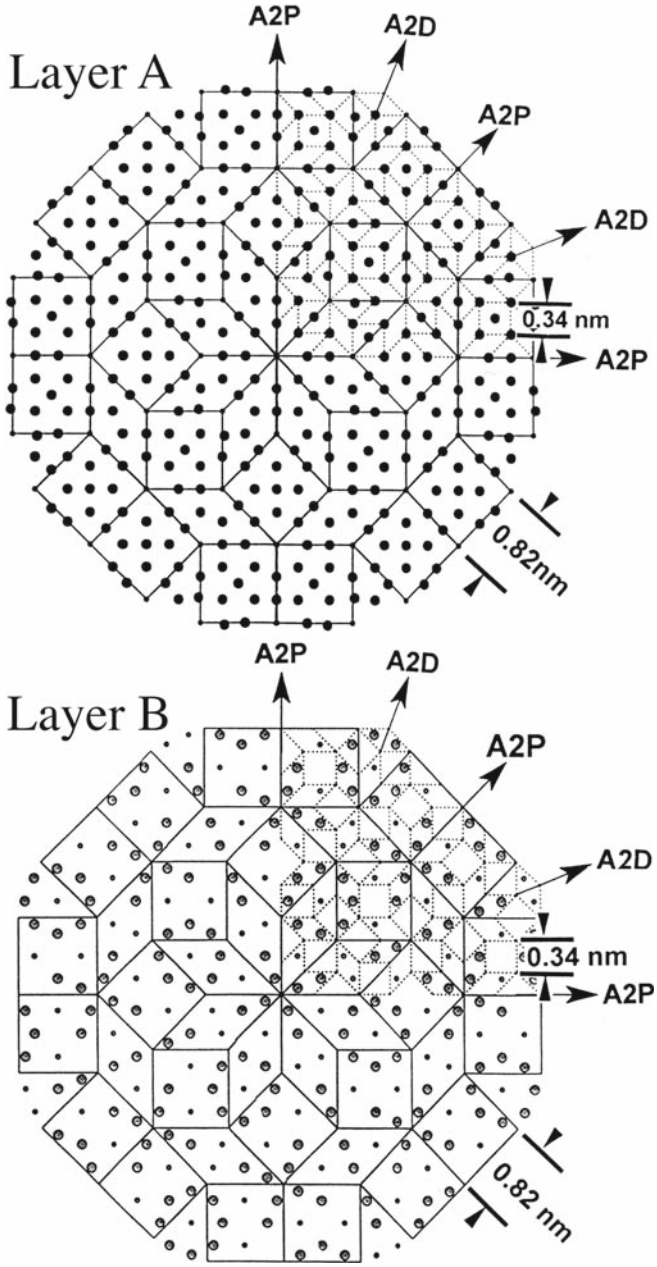


FIGURE 30.11. The model used to simulate quasicrystals with eightfold symmetry. The structure for the simulation was constructed as a four-layer sequence ABAB′ where the B and B′ layers are related by an 8_4 screw axis.

30.14 BONDING IN CRYSTALS

We mentioned early on that one problem we have with simulation concerns the fact that atoms are bonded in different ways in different materials. The standard

approach has been to use values for structure factors tabulated by Doyle and Turner and by Doyle and Cowley. These values were calculated using a relativistic Hartree-Fock (RHF) model for the atomic potential. An alternative approach is to relate the scattering factor for electrons (f_e) to that for X-rays (f_x) using the Mott equation, or to use a more sophisticated atomic potential known as the relativistic Hartree-Fock-Slater (RHFS) model. Carlson et al. give tabulated results while Tang and Dorignac have made detailed comparisons for HRTEM imaging.

O'Keeffe and Spence have re-examined the meaning of the mean inner potential. One of the reasons that we need to understand this concept is that we often link data from X-ray diffraction and data from electron diffraction. As usual, computers are making it possible to do more elaborate calculations using other potentials.

While this is an evolving study, some important results have been obtained

- The inner potential is very sensitive to bonding effects. O'Keeffe and Spence discuss this result for MgO (large ionic component), Si (covalent), and Al (metallic).
- We are still not able to take full account of bonding effects, which could be important for HRTEM images.

This paper by O'Keeffe and Spence is a highly recommended reading for those who have a strong physics background but think TEM is a 'known' subject! For a very different approach, see the paper by Zuo and Spence which uses the DP to deduce information about bonding.

30.15 SIMULATING *Z*-CONTRAST

This is a topic where you can make an impact! The clues are in Chapters 2 and 3 and in the companion text.

30.16 SOFTWARE FOR PHASE-CONTRAST HRTEM

This is a short section because it is summarized in Table 30.1. HRTEMs are priced at up to $4 M as of 2009. Image simulation is essential for HRTEM. Very few students can, or should, write their own programs to simulate HRTEM images—it has already been done many times. The catch is that no manufacturer supports any software package for simulating HRTEM images. Your lab must have at least two programs for you to use. In future we will see much more use of Mathematica and Matlab in this field.

CHAPTER SUMMARY

If you are going to do HRTEM imaging, you must be prepared to use image simulation to assist you in interpreting your images. If you want to do quantitative imaging, simulation is an essential component of the process. Most materials scientists using TEM will want to use one of the established software packages in Table 30.1. There are several important conclusions contained in this chapter

- Make sure that you know all you can about your specimen. We illustrated the dangers with our discussion of grooved GBs. You can waste too much time looking at artifacts caused by specimen preparation.
- Make sure that you know all you can about your TEM. You now have some idea of how many parameters are required by the simulation routines. Beware of the parameters which you did not measure for your machine. The program will need to use some value: *you* should make sure it is appropriate.
- Make sure that you accurately align your TEM before you record any images.
- If possible, use more than one program to simulate the images. At least try this once.
- Record a through-focus series and check for changes in Δf by repeating the first image.
- The fact that the thickness of your specimen varies can be a great asset provided you can determine that thickness; i.e., it gives you another variable.

If you can afford a C_s-corrected HRTEM, then C_s becomes another variable.

The traditional method of using simulated images has often involved looking at a series of simulated images for different values of Δf and t and finding the best match with your experimental image. Clearly this is not the ideal approach! Remember that the interpretation of HRTEM images may not be straightforward or unique. We must next compare the simulated images with those generated experimentally. This is the subject of the next chapter and is the basis of quantitative HRTEM.

GENERAL

Buseck, PR, Cowley, JM and Eyring, L, Eds. 1988 *High-Resolution Transmission Electron Microscopy and Associated Techniques* Oxford University Press New York. An excellent resource covering more than HRTEM.

Horiuchi, S 1994 *Fundamentals of High-Resolution Electron Microscopy* North-Holland, New York. An ideal complement to Spence's book.

Kihlborg, L, Ed. 1979 *Direct Imaging of Atoms in Crystals and Molecules* Nobel Symposium 47 The Royal Swedish Academy of Sciences Stockholm. A classic collection of papers on HRTEM.

Krakow, W and O'Keefe, M, Eds. 1989 *Computer Simulation of Electron Microscope Diffraction and Images* TMS, Warrendale, PA. A collection of focused review articles.

O'Keefe, MA and Kilaas, R 1988 in *Advances in High-Resolution Image Simulation in Image and Signal Processing in Electron Microscopy*, Scanning Microscopy Supplement 2. (Eds. PW Hawkes, FP Ottensmeyer, WO Saxton and A Rosenfeld) p. 225 SEM Inc. AMF O'Hare Il. This article is another excellent introduction to the subject.

Spence, JCH 2003 *High-Resolution Electron Microscopy* 3rd edition Oxford University Press New York. This is *the* text for users of the HRTEM.

Zuo, JM and Spence, JCH 1992 *Electron Microdiffraction* Springer, NY

SCATTERING FACTORS

Carlson, TA, Lu TT, Tucker, TC, Nestor, CW and Malik, FB 1970 *Report ORNL-4614*, ORNL Oak Ridge TN. Tabulated values of scattering factors.

Doyle, PA and Turner, PS 1968 *Relativistic Hartree-Fock X-ray and Electron Scattering Factors* Acta. Cryst. **A24** 390–7.

Tang, D. and Dorignac, D 1994 *The Calculation of Scattering Factors in HREM Image Simulation* Acta. Cryst. **A50** 45–52. Detailed comparisons of scattering factors.

QUASICRYSTALS

Beeli, C and Horiuchi, S 1994 *The Structure and its Reconstruction in the Decagonal $Al_{70}Mn_{17}Pd_{13}$ Quasicrystal* Phil. Mag. **B70** 215–40. Calculations of HRTEM images.

Shoemaker, CB 1993 *On the Relationship between* μ-*$MnAl_{4.12}$ and the Decagonal Mn-Al Phase* Phil. Mag. **B67** 869–81. A review of several models.

MODELS

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

Coene, W and Van Dyck, D 1984 *The Real Space Method for Dynamical Electron Diffraction Calculations in High Resolution Electron Microscopy: II. Critical Analysis of the Dependency on the Input Parameters* Ultramicrosc. **15** 41–50.

Coene, W and Van Dyck, D 1984 *The Real Space Method for Dynamical Electron Diffraction Calculations in High Resolution Electron Microscopy: I. Principles of the Method* Ultramicrosc. **15** 29–40.

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

Ross, FM and Stobbs, WM 1991 *A Study of the Initial Stages of the Oxidation of Silicon Using the Fresnel Method* Phil. Mag. **A63** 1–36.

Ross, FM and Stobbs, WM 1991 *Computer Modelling for Fresnel Contrast Analysis* Phil. Mag. **A63** 37–70.

Simpson, YK, Carter, CB, Morissey, KJ, Angelini, P and Bentley, J 1986 *The Identification of Thin Amorphous Films at Grain-Boundaries in Al_2O_3* J. Mater. Sci. **21** 2689–96.

IMAGE SIMULATION

Barry, J 1992 in *Electron Diffraction Techniques* **1** (Ed. JM Cowley) p. 170 I.U.Cr. Oxford Science Publication New York.

Coene, W, Van Dyck, D, Van Tendeloo, G and Van Landuyt, J 1985 *Computer Simulation of High-Energy Electron Scattering by Non-Periodic Objects. The Real Space Patching Method as an Alternative to the Periodic Continuation Technique* Phil. Mag. **52** 127–143. Real-space patching for Section 30.12.

Cowley, JM and Moodie, AF 1957 *The Scattering of Electrons by Atoms and Crystals. I. A New Theoretical Approach* Acta Cryst. **10** 609–19.

Kambe, K 1982 *Visualization of Bloch Waves of High Energy Electrons in High Resolution Electron Microscopy* Ultramicrosc. **10** 223–227.

Kirkland EJ 1998 *Advanced Computing in Electron Microscopy* Springer, NY

Stadelmann, PA 1979. *EMS - a Software Package for Electron Diffraction Analysis and HREM Image Simulation in Materials Science* Ultramicrosc. **21** 131–145. The original description of EMS.

SOME PHYSICS

Doyle, PA and Cowley, JM 1974 in International Tables for X-ray Crystallography, Vol. **IV**, pp. 152–173, Kluwer Academic Publ. Dordrecht Netherlands.

Fujimoto, F 1978 *Periodicity of Crystal Structure Images in Electron Microscopy with Crystal Thickness* Phys. stat. sol. (a) **45** 99–106.

Jiang, J-C, Hovmöller, S and Zou, X-D 1995 *A Three-dimensional Structure Model of Eight-Fold Quasi-crystals Obtained by High-Resolution Electron Microscopy* Phil. Mag. Lett. **71** 123–129.

O'Keeffe, M and Spence, JCH 1994 *On the Average Coulomb Potential (Φ_0) and Constraints on the Electron Density in Crystals*. Acta Cryst. **A50**(1) 33–45.

Self, PG and O'Keefe, MA 1988 in *High-Resolution Transmission Electron Microscopy and Associated Techniques* (Eds. PR Buseck, JM Cowley and L Eyring) p. 244 Oxford University Press New York.

Taftø, J, Jones, RH and Heald, SM 1986 *Transmission Electron Microscopy of Interfaces Utilizing Mean Inner Potential Differences between Materials* J Appl. Phys. **60** 4316–8.

Wilson, AR and Spargo, AEC 1982 *Calculation of the Scattering from Defects Using Periodic Continuation Methods* Phil. Mag. **A46** 435–49.

Zhu, Y, Taftø, J, Lewis, LH and Welch, DO 1995 *Electron Microscopy of Grain Boundaries. An Application to RE-Fe-B (RE = Pr or Nd) Magnetic Materials* Phil. Mag. Lett. **71** 297–306.

THE COMPANION TEXT

The companion text extends the topic of simulation in chapters on EMS (jEMS is the java version that Pierre Stadleman supports) and on simulating diffraction contrast images. One idea which you might examine immediately is the use of Mathematica or similar packages to graph the equations.

SELF-ASSESSMENT QUESTIONS

Q30.1 List the four principal methods of calculating multiple images.
Q30.2 What is a phase grating?
Q30.3 Why do we sample all of reciprocal space, not just the Bragg beams?
Q30.4 Write down an equation relating ψ_{n+1} and ψ using a propagator and a phase grating.
Q30.5 Why are we using Fresnel diffraction in this multislice treatment rather than Fraunhofer?
Q30.6 We use convolution to describe what sort of scattering and why?
Q30.7 Why must $Q(\mathbf{k})$ take account of nearly twice as many k values as $\psi(\mathbf{k})$ or $P(\mathbf{k})$?
Q30.8 Write down an equation for $\psi_{n+1}(\mathbf{k})$ used in the FFT method.
Q30.9 Extra challenge. Why is it FFT not just FT?
Q30.10 Write down an equation for $\psi_{n+1}(\mathbf{r})$ in the Coene-Van Dyck method.
Q30.11 Why do we use ' ' when describing Kambé's 'simple' example?
Q30.12 In Kambé's 'simple' example, we consider the case where only three Bloch waves are significant. Why choose just three, not more (or fewer)?
Q30.13 We mention the direct connection between WPOA and Bloch wave propagation. Why is this?
Q30.14 Why is it relevant that λ is finite (not zero)?
Q30.15 Why does beam convergence alter a HRTEM image according to the Ewald construction?
Q30.16 Why will GB grooving influence HRTEM images?
Q30.17 What is the periodic continuation method?
Q30.18 What is the problem in imaging crystals with 5- and/or 10-fold symmetry?
Q30.19 How do most multislice images simulation programs take care of ionic and covalent bonding?
Q30.20 Why do many researchers prefer a through-focus series to just a focal series of images?

TEXT-SPECIFIC QUESTIONS

T30.1 Assuming that the first Fresnel fringe is $\delta = (\lambda \Delta f)^{1/2}$ away from the edge of the sample, calculate Δf for Figure 9.21B.
T30.2 How well does the relationship $\delta = (\lambda \Delta f)^{1/2}$ for the position of the first Fresnel fringe predict the images in Figure 30.5C assuming the structure given in Figure 30.5A?
T30.3 Construct a unit cell to use in HRTEM simulation of the $\Sigma = 3$ 111 twin boundary in Si.
T30.4 Construct a unit cell to use in HRTEM simulation of the $\Sigma = 3$ 112 twin boundary in GaP. Discuss the information you would obtain for an HRTEM giving point (structural) resolution of 0.3, 0.2, and 0.1 nm.
T30.5 Show the smallest reproducible unit for Figure 30.11 given the symmetry shown in the figure.
T30.6 Plot equation 30.14 for different values of Δf. How does this plot change as we change the kV of the microscope?

T30.7 By playing with equation 30.15 deduce equations 30.19 and 30.20.
T30.8 Deduce equation 30.21 using equations 30.18 and 30.20.
T30.9 Explain why the 'images' from Bloch wave 1 differs from those in Bloch waves 2 and 3, as shown in Figure 30.3.
T30.10 Index the DP in Figure 30.4 and thus determine the radius (in nm^{-1}) of the objective aperture.