28

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 $h(\boldsymbol{r})$ 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

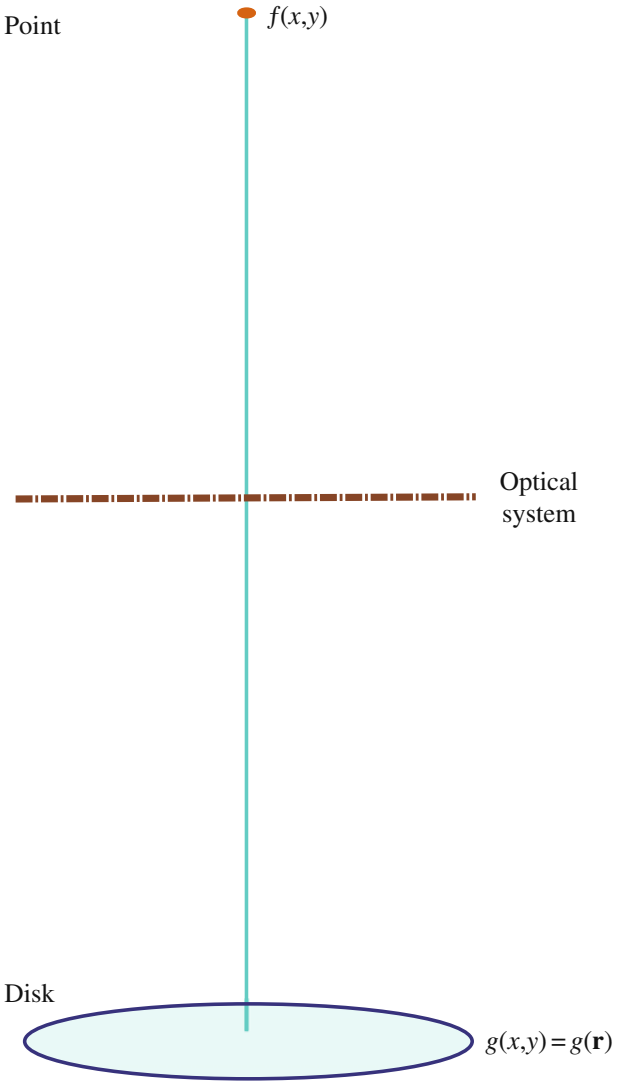

FIGURE 28.1. An optical system transforms a point in the specimen (described by $f(x,y)$) into a disk in the image described by $g(x,y)$. The intensity in the image at point (x,y) can be described by the function $g(x,y)$ or $g(\mathbf{r})$. It has a unique value for each value of (x,y), so we say it is a representation of the image.

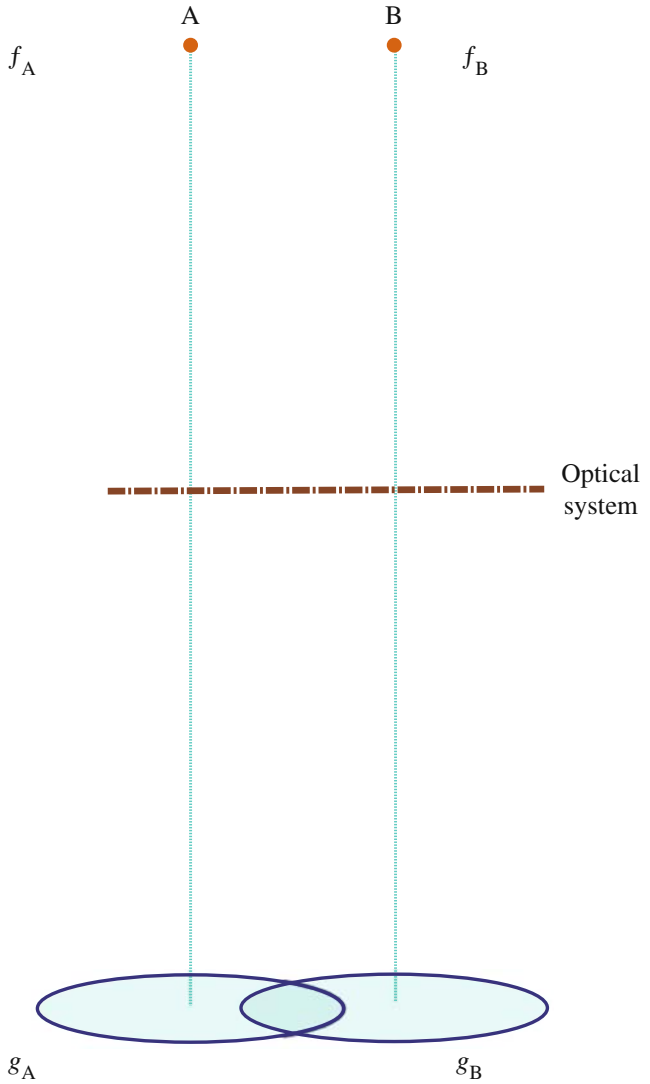

FIGURE 28.2. Two points, f_A and f_B, in the specimen produce two disks, g_A and g_B, in the image.

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 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.

$$g(x,y) = \sum_{u_x,u_y} G(u_x - u_y)\exp\left(2\pi i\left(xu_x + yu_y\right)\right) \qquad (28.3)$$

$$g(x,y) = \sum_{\mathbf{u}} G(\mathbf{u})\exp(2\pi i\ \mathbf{u}\,.\,\mathbf{r}) \qquad (28.4)$$

Here $\mathbf{u}$ is a reciprocal-lattice vector, the spatial frequency for a particular direction. We have expressed $g(\mathbf{r})$ in terms of a combination of the possible values of $G(\mathbf{u})$, 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

$$G(\mathbf{u}) = H(\mathbf{u})F(\mathbf{u}) \qquad (28.5)$$

So a convolution in real space (equation 28.1) gives multiplication in reciprocal space (equation 28.5).

The factors contributing to $H(\mathbf{u})$ include

Apertures → The aperture function $A(\mathbf{u})$
Attenuation of the wave → The envelope function $E(\mathbf{u})$
Aberration of the lens → The aberration function $B(\mathbf{u})$

We write $H(\mathbf{u})$ as the product of these three terms

$$H(\mathbf{u}) = A(\mathbf{u})E(\mathbf{u})B(\mathbf{u}) \qquad (28.6)$$

The aperture function says that the objective diaphragm cuts off all values of $\mathbf{u}$ (spatial frequencies) greater than (higher than) some selected value governed by the radius of the aperture. The envelope function has the same effect but is a property of the lens itself, and so may be either more or less restricting than $A(\mathbf{u})$. $B(\mathbf{u})$ is usually expressed as

$$B(\mathbf{u}) = \exp(i\chi(\mathbf{u})) \qquad (28.7)$$

The term $\chi(\mathbf{u})$ can be written as

$$\chi(\mathbf{u}) = \pi\Delta f\lambda u^2 + \frac{1}{2}\pi C_s\lambda^3 u^4 \qquad (28.8)$$

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

$$f(x,y) = A(x,y)\exp(-i\phi_t(x,y)) \tag{28.9}$$

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.

$$V_t(x,y) = \int_0^t V(x,y,z)dz \tag{28.10}$$

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.)

$$\lambda = \frac{h}{\sqrt{2meE}} \tag{28.11}$$

(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'

$$\lambda' = \frac{h}{\sqrt{2me(E + V(x,y,z))}} \tag{28.12}$$

so we can say that, when passing through a slice of material of thickness dz, the electrons experience a phase change given by

$$d\phi = 2\pi\frac{dz}{\lambda'} - 2\pi\frac{dz}{\lambda} \tag{28.13}$$

$$d\phi = 2\pi\frac{dz}{\lambda}\left(\frac{\sqrt{E + V(x,y,z)}}{\sqrt{E}} - 1\right) \tag{28.14}$$

$$d\phi = 2\pi\frac{dz}{\lambda}\left(\left(1 + \frac{V(x,y,z)}{E}\right)^{\frac{1}{2}} - 1\right) \tag{28.15}$$

$$d\phi = 2\pi\frac{dz}{\lambda}\frac{1}{2}\frac{V(x,y,z)}{E} \tag{28.16}$$

$$d\phi = \frac{\pi}{\lambda E}V(x,y,z)dz \tag{28.17}$$

$$d\phi = \sigma\, V(x,y,z)dz \tag{28.18}$$

So the total phase shift is dependent only on $V(x,y,z)$ since

$$d\phi = \sigma\int V(x,y,z)dz = \sigma V_t(x,y) \tag{28.19}$$

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 σs 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

$$f(x,y) = \exp[-i\sigma V_t(x,y) - \mu(x,y)] \tag{28.20}$$

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

$$f(x,y) = 1 - i\sigma V_t(x,y) \tag{28.21}$$

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

$$\psi(x,y) = [-i\sigma V_t(x,y)] \otimes h(x,y) \qquad (28.22)$$

If we represent $h(x,y)$ as $\cos(x,y) + i\sin(x,y)$, then $\psi(x,y)$ becomes

$$\psi(x,y) = 1 + \sigma V_t(x,y) \otimes \sin(x,y) - i\sigma V_t(x,y) \otimes \cos(x,y) \qquad (28.23)$$

As usual, the intensity is given by

$$I = \psi\psi^* = |\psi|^2 \qquad (28.24)$$

Multiplying this out and neglecting terms in σ^2 (because σ is small), we find that

$$I = 1 + 2\sigma V_t(x,y) \otimes \sin(x,y) \qquad (28.25)$$

Knowing this result we can say that, in the WPOA, only the imaginary part of $B(\mathbf{u})$ in equation 28.7 contributes to the intensity in equation 28.24 (because it gives the imaginary part of $h(x,y)$). 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

$$T(\mathbf{u}) = A(\mathbf{u})E(\mathbf{u})2\sin\chi(\mathbf{u}) \qquad (28.26)$$

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

$$\cos^2(x,y) + \sin^2(x,y) \qquad (28.27)$$

which is just unity.

So the CTF in HRTEM would be different from $T(\mathbf{u})$, and therefore we will call $T(\mathbf{u})$ the *objective lens transfer function*.

28.5 THE TRANSFER FUNCTION

You must note two things here. First, as we just said, the transfer function, $T(\mathbf{u})$, formulation applies to any specimen, and second, $T(\mathbf{u})$ is *not* the 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 $E(\mathbf{u})$,

$$T(\mathbf{u}) = 2A(\mathbf{u}) \sin \chi(\mathbf{u}) \qquad (28.28)$$

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 $\chi(\mathbf{u})$ = 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!

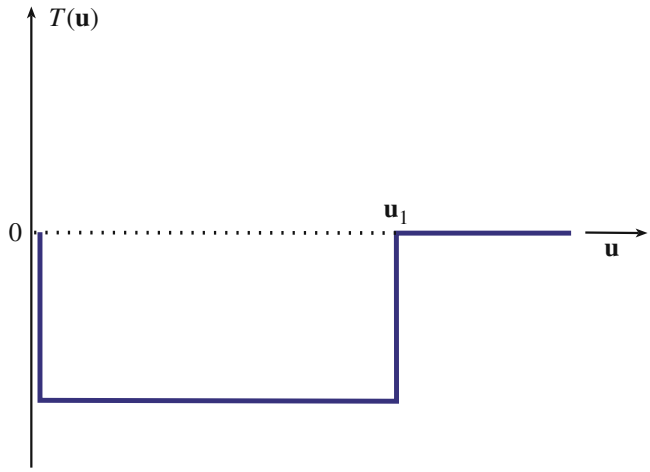

FIGURE 28.3. The ideal form of the transfer function, $T(\mathbf{u})$. In this example, $T(\mathbf{u})$ is large and negative between $\mathbf{u} \neq 0$ and $\mathbf{u} = \mathbf{u}_1$.

USEFUL 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 $\delta(\theta)$.

$$\delta(\theta) = C_s\theta^3 + \Delta f\theta \qquad (28.29)$$

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

$$D(\theta) = \int_0^\theta \delta(\theta) d\theta = \frac{C_s\theta^4}{4} + \Delta f\frac{\theta^2}{2} \qquad (28.30)$$

Now, Bragg's law tells us that

$$2d \sin \theta_B = n\lambda \qquad (28.31)$$

or since θ_B is small

$$2\theta_B \cong \lambda g \qquad (28.32)$$

So, we can replace θ in equation 28.30 by λu where **u** is a general reciprocal-lattice vector. (Remember that the scattering angle is $2\theta_B$, not θ_B.)

We are interested in the phase $\chi(\mathbf{u})$, so we write

$$\chi(\mathbf{u}) = \text{phase} = \frac{2\pi}{\lambda} D(\mathbf{u}) = \frac{2\pi}{\lambda}\left(C_s \frac{\lambda^4 u^4}{4} + \Delta f \frac{\lambda^2 u^2}{2}\right) \quad (28.33)$$

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 **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 **u**, shown in Figure 28.4, illustrates the main features. The curve has been drawn for $C_s = 1$ mm, $E_0 = 200$ 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 **u** is small, the Δf term dominates.
- $\sin\chi$ first crosses the **u**-axis at $\mathbf{u}_1$ and then repeatedly crosses the **u**-axis as **u** increases.
- χ can continue forever but, in practice, it is modified by other functions which we discuss in Section 28.8.

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 **u** does not depend on your specimen. Figure 28.5 shows a series of $\sin\chi$ curves for an imaginary 200-kV microscope where C_s has been changed. In each case, the 'best' curve (we'll discuss this in a moment) has been chosen. You can appreciate that the smaller C_s values give the larger $\mathbf{u}_1$ values; so a small C_s means we can achieve a higher spatial resolution.

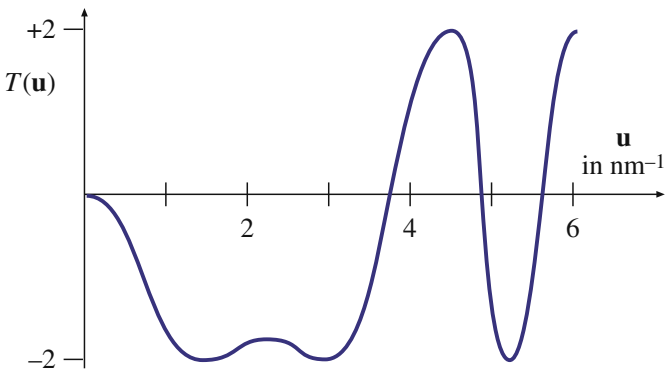

FIGURE 28.4. A plot of $T(\mathbf{u})$ versus **u** (C_s = 1 mm, E_0 = 200 kV, Δf = –58 nm).

C_s, Δf, AND β

High spatial frequencies ⇒ large diffraction angles ⇒ 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.

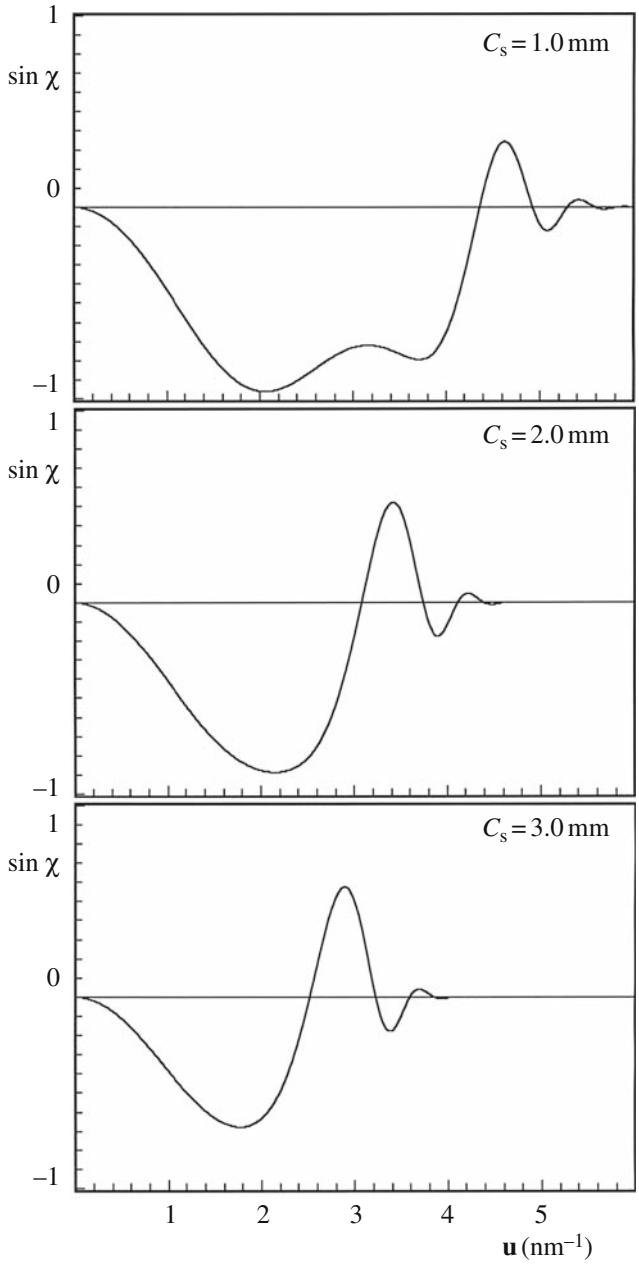

FIGURE 28.5. A series of $\sin\chi$ curves calculated for different values of C_s. Remember $2\sin\chi = T(\mathbf{u})$. ($E_0$ = 200 kV, Δf = –60 nm).

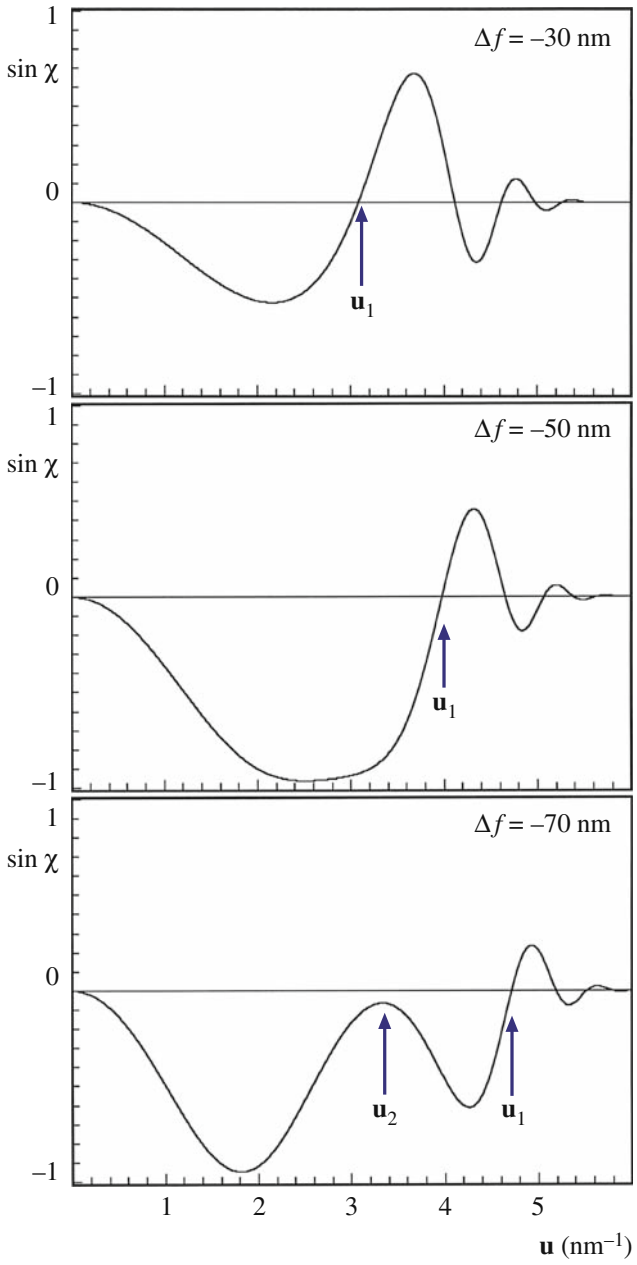

FIGURE 28.6. A series of sin χ curves calculated for different values of Δf. (E_0 = 200 kV; C_s = 1.0 mm.)

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 **u**-axis so that $\mathbf{u}_1$ is suddenly much smaller. If we just make Δf smaller, then $\mathbf{u}_1$ steadily decreases. In the next section we will discuss the optimum value for Δf.

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

$$\Delta f_{Sch} = -1.2(C_s\lambda)^{1/2} \qquad (28.35)$$

At this defocus (which we'll derive below) all the beams will have nearly constant phase out to the 'first crossover' of the zero axis. This crossover point is defined as the instrumental resolution limit. This is the best performance that can be expected from a microscope unless we use 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 χ(**u**) is close to –120°; then sin χ will be near –1 when χ is between –120° and –60°. We know that when χ = π, sin χ = 0 so we want sin χ to be as large as possible over a large range of **u**; sin χ will be a nearly flat function if $d\chi/du$ is zero. So we look for the value of

Δf when $d\chi/du$ is zero and χ is -120° (you should consider why we choose this value of χ). Differentiating equation 28.29 gives

$$\frac{d\chi}{du} = 2\pi\Delta f\lambda u + 2\pi C_s\lambda^3 u^3 \qquad (28.36)$$

Set the left-hand term equal to 0

$$0 = \Delta f + C_s\lambda^2 u^2 \qquad (28.37)$$

When $\chi = -120^\circ$, equation 28.29 becomes

$$-\frac{2\pi}{3} = \pi\Delta f\,\lambda u^2 + \frac{1}{2}\pi C_s\lambda^3 u^4 \qquad (28.38)$$

Combining equations 28.37 and 28.38 gives a special value for Δf

$$\Delta f_{Sch} = -\left(\frac{4}{3}C_s\lambda\right)^{\frac{1}{2}} \qquad (28.39)$$

The subscript denotes the Scherzer defocus value. Since $(1.33)^{1/2} = 1.155$ ($\sim$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

$$u_{Sch} = 1.51 C_s^{-\frac{1}{4}}\,\lambda^{-\frac{3}{4}} \qquad (28.40)$$

The resolution at the Scherzer defocus can then be defined as the reciprocal of u_{Sch}.

$$r_{Sch} = \frac{1}{1.51}C_s^{\frac{1}{4}}\,\lambda^{\frac{3}{4}} = 0.66 C_s^{\frac{1}{4}}\,\lambda^{\frac{3}{4}} \qquad (28.41)$$

You will often see this expression with different values for the constant (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 $\left(C_s\lambda^3\right)^{\frac{1}{4}}$ 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 $T_{eff}(\mathbf{u})$.

$$T_{eff}(\mathbf{u}) = T(\mathbf{u})E_cE_a \qquad (28.42)$$

The effect of the envelope functions is to impose a virtual aperture in the back focal plane of the objective

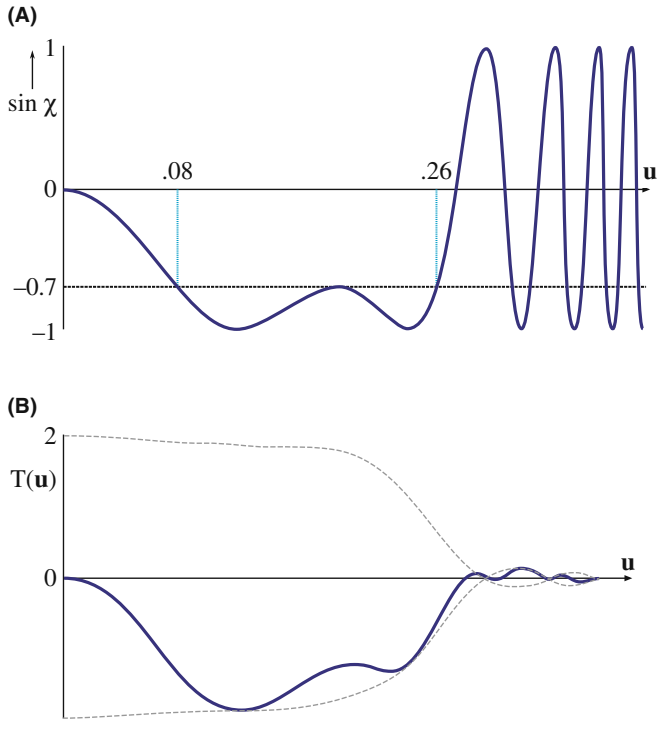

FIGURE 28.7. (A) $\sin\chi(\mathbf{u})$ versus $\mathbf{u}$ without damping of the higher spatial frequencies. (B) $T(\mathbf{u})$ versus $\mathbf{u}$ modified by the damping envelope (dashed line); $\Delta f = -100$ nm, $C_s = 2.2$ nm.

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.

$$\sin \chi(\mathbf{u}) = 0.3 \tag{28.43}$$

or

$$\Delta f_{MC} = -0.44(C_s\lambda)^{1/2} \tag{28.44}$$

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,

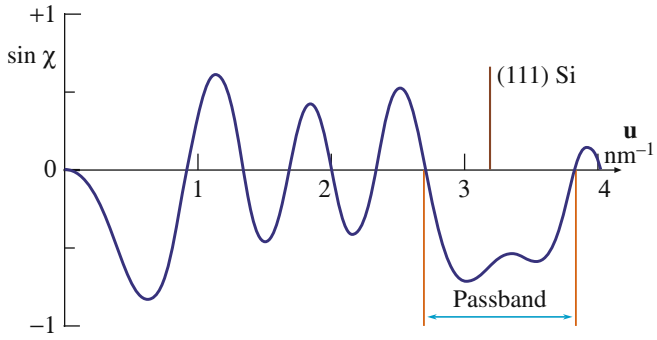

FIGURE 28.8. Special settings of the CTF to make use of passbands or 'windows' in the transfer function, here optimized to image Si(111).

over a range of **u** which includes the reflection of interest. These passbands occur periodically with underfocus at values set by

$$\Delta f_p^n = -\{([8n+3]/2)(C_s\lambda)\}^{1/2} \tag{28.45}$$

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

$$\Delta f_{AFF} \cong \{2(4m \pm 0.23)C_s\lambda^3/d^4\}(d^2/2\lambda) \tag{28.46}$$

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 LaB_6 filament and a small condenser aperture (unless you have an FEG: see later).
- Perform current and voltage centering of the objective lens routinely and frequently at high magnification.
- Work in thin, flat, and clean regions of the specimen.
- Orient the specimen using small SAD apertures or bend contours in the image, so the beam is aligned along a zone axis.
- Correct the astigmatism, using optical diffractograms if necessary, but ideally on-line (Chapter 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 **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

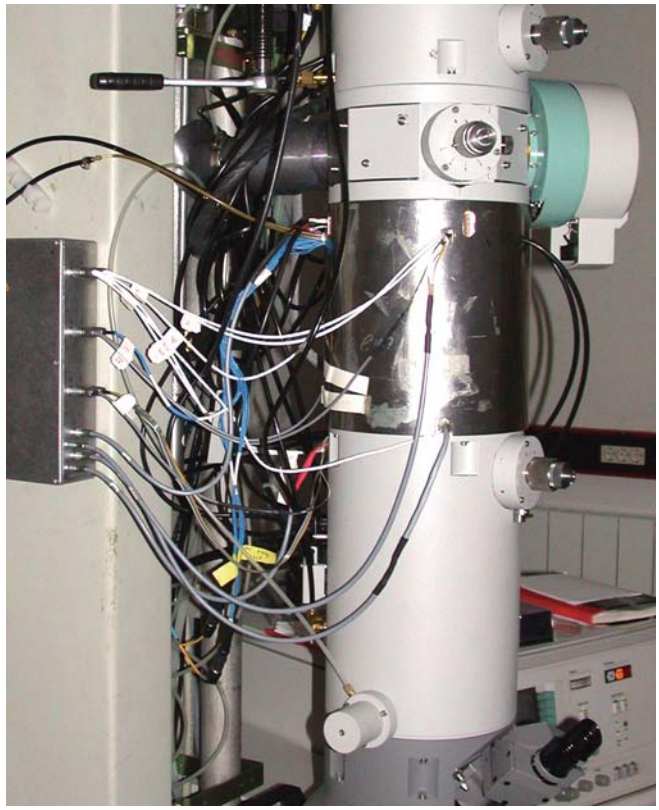

FIGURE 28.9. The post-objective lens corrector system as incorporated into a 200-kV JEOL TEM; it looks sleeker in the finished product. (See also Figure 6.12B.)

of 1 in 10^4 to give atomic resolution. When C_s is zero, the specimen resolution limit will be determined by C_c

$$d_{C_s=0} \approx \left[\left(\frac{\Delta E}{E}\right)\lambda C_c\right]^{\frac{1}{2}} \tag{28.47}$$

If $C_c = 2\,\text{mm}$ and $\Delta E \sim 0.3\,\text{eV}$, a 200-kV FEG TEM could achieve a resolution of 0.8 Å. 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, $C_s = 3\,\text{mm}$ and the resolution limit is 0.28 Å for this 200-kV FEG TEM. Other lens defects will limit this to ~0.5 Å 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 **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).

$$E_T(\mathbf{u}) = E_c(\mathbf{u})E_s(\mathbf{u})E_d(\mathbf{u})E_v(\mathbf{u})E_D(\mathbf{u}) \qquad (28.48)$$

The individual envelope functions in equation 28.48 are

$E_c(\mathbf{u})$: for chromatic aberration.
$E_s(\mathbf{u})$: for the source dependence due to the small spread of angles from the probe.
$E_d(\mathbf{u})$: for specimen drift.
$E_v(\mathbf{u})$: for specimen vibration.
$E_D(\mathbf{u})$: for the detector.

As you can see, some of these envelope functions are new, some are old. We won't discuss all 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

$$E_c(\mathbf{u}) = \exp\left[-\frac{1}{2}(\pi\lambda\delta)^2 u^4\right] \qquad (28.49)$$

where c reminds us that this is a chromatic aberration and δ is the defocus spread due to this aberration

$$\delta = C_c\left[4\left(\frac{\Delta I_{obj}}{I_{obj}}\right)^2 + \left(\frac{\Delta E}{V_{acc}}\right)^2 + \left(\frac{\Delta V_{acc}}{V_{acc}}\right)^2\right]^{\frac{1}{2}} \qquad (28.50)$$

The terms $\Delta V_{acc}/V_{acc}$ and $\Delta I_{obj}/I_{obj}$ represent the instabilities in the high-voltage (accelerator) supply

and the objective lens current. $\Delta E/V_{acc}$ is the intrinsic energy spread in the electron gun. Notice that ΔE and ΔV are different: ΔV depends on how well we can control the voltage supply whereas ΔE depends on our choice of electron source (see Chapter 5). If we neglect any other contributions to the envelope function, then we can define an information limit due to instrument chromatic aberrations by ρ_c

$$\rho_c = \left(\frac{\pi\lambda\delta}{\sqrt{2\ln(s)}}\right)^{\frac{1}{2}} \tag{28.51}$$

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

$$\rho_c = \left(\frac{\pi\lambda\delta}{2}\right)^{\frac{1}{2}} \tag{28.52}$$

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

$$\begin{aligned} E_s(\mathbf{u}) &= \exp\left[\left(\frac{\alpha}{2\lambda}\right)^2\left(\frac{\partial\chi(\mathbf{u})}{\partial u}\right)^2\right] \\ &= \exp\left[-\left(\frac{\pi\alpha}{\lambda}\right)^2\left(C_s\lambda^3u^3+\lambda u\right)^2\right] \end{aligned} \tag{28.53}$$

Here α is the semi-angle characterizing the Gaussian distribution. What this equation tells us is that if α is too large ($\leq$1 mrad) it can limit the information limit. If we say that u must lie between u and some maximum value u_{max}, we can maximize the argument of the exponential in equation 28.53 to give an optimum focus

$$\Delta f_{opt} = -\frac{3}{4}C_s\lambda^2u_{max}^2 = -\frac{3}{4}\frac{C_s\lambda^2}{\rho_i^2} \tag{28.54}$$

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

$$\rho_\alpha = \left(\frac{6\pi\alpha a}{\lambda\sqrt{\ln(s)}}\rho_s^4\right)^{\frac{1}{3}} \tag{28.55}$$

The envelope functions for the drift/translation and vibration of the specimen represent a new method for

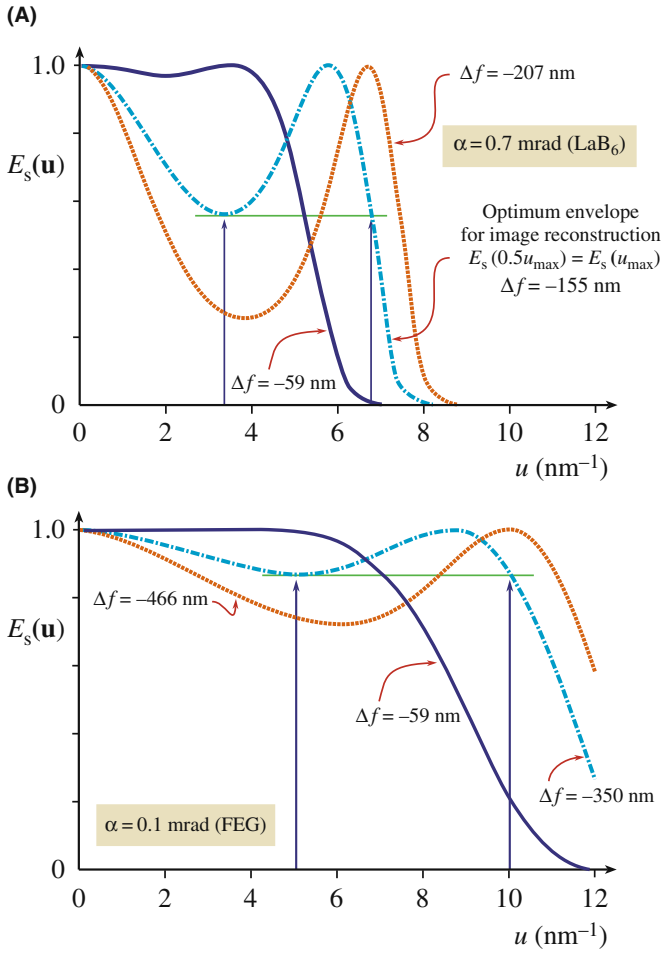

FIGURE 28.10. Variations in the envelope function, $E_s(\mathbf{u})$, for different objective lens defocus: (A) LaB_6 source, (B) FEG.

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 $E_v(\mathbf{u})$

$$\rho_d = \frac{\pi d}{\sqrt{6\ln(s)}} \tag{28.56}$$

and

$$\rho_v = \frac{\pi v}{\sqrt{\ln(s)}} \tag{28.57}$$

In these equations, d is the total specimen drift during the exposure, t_{exp}, so $d = v_d t_{exp}$ for a drift velocity v_d; v is the amplitude of the vibration.

The detector envelope function, $E_D(\mathbf{u})$, is something we never worried about with film, but CCD cameras have a limited number of pixels, i.e., we only have a limited number of resolved image points. This envelope function results from two effects

- Delocalization of the information in the image.
- The finite pixel size.

DELOCALIZATION

Image delocalization depends on **u**. It is large when $\partial\chi(\mathbf{u})/\partial u$ oscillates rapidly as it does for large **u**, i.e., where we are placing the information limit.

The 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

$$\alpha C_s \lambda^3 u_{max}^3 = R_w \tag{28.58}$$

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

$$R_w = \frac{1}{2} ND \tag{28.59}$$

The information limit due to the detector (i.e., the cross-over values of the detector envelope functions $E_D(\mathbf{u})$) is

$$\rho_D = \left(\frac{12\sqrt{2}\pi a}{N\sqrt{\ln(s)}} \right)^{\frac{1}{4}} \rho_s \tag{28.60}$$

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.

$$\alpha \leq \frac{\lambda}{6\pi a \rho_s} \left(\frac{\rho_c}{\rho_s} \right)^3 \tag{28.61}$$

$$N \geq 12\sqrt{2}\pi a \left(\frac{\rho_s}{\rho_c} \right)^4 \tag{28.62}$$

$$d \leq \frac{\sqrt{6}}{\pi} \rho_c;\ 0.8\rho_c \tag{28.63}$$

$$u \leq \frac{1}{\pi} \rho_c;\ 0.3\rho_c \tag{28.64}$$

Table 28.1 gives some numerical examples of what these equations mean.

To see whether we will ever reach the information limit, we have to consider the effect of the noise. We know that the signal-to-noise ratio is proportional to $\beta^{1/2}$, where β is the brightness of the electron gun. The smallest image element we need to examine has an area ρ_i^2. Then the background signal I_0 is given by

$$I_0 = D\rho_i^2 = \beta\pi\alpha^2 t \rho_i^2 \tag{28.65}$$

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

$$DEF\rho_i^2 = kI_0^{1/2} = k\rho_i D^{1/2} \tag{28.66}$$

Therefore for $\alpha = 1$ mrad and $t = 1$ second, we can express the signal-to-noise ratio as

$$s_0 = 443\rho_i \frac{F}{k} \beta^{1/2} \tag{28.67}$$

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/m^2 sr for a LaB_6 gun and 10^{13} A/m^2 sr for a Schottky FEG. You can show that for $\rho_i = 0.15$ nm (a LaB_6 gun) $\ln_e s_0$ is 1.2–2.2 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.

$$\alpha_{opt} = \frac{1}{k_s\sqrt{2}} \left(\frac{\rho_i}{\rho_c} \right)^3 = \frac{1}{6\pi a\sqrt{2}} \frac{\lambda}{\rho_s} \left(\frac{\rho_i}{\rho_s} \right)^3 \tag{28.68}$$

$$t_{opt} = \frac{1}{2k_d} \left(\frac{\rho_i}{\rho_c} \right) \simeq 0.39 \left(\frac{\rho_i}{v_d} \right) \tag{28.69}$$

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)

	α (mrad)		N (pix)	
ρ_s/ρ_c	ε_0	ε_{opt}	ε_0	ε_{opt}
1	0.58	2.3	53	13
1.5	0.17	0.69	270	67
2	0.07	0.30	853	213
2.5	0.04	0.15	2082	521
3	0.02	0.09	4320	1080

ε_0: Gaussian focus, ε_{opt}: optimum focus; $\lambda = 0.011\ \rho_s$ (de Jong and Van Dyck 1993).

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 ~1500°C. It gives the same high brightness, but a larger maximum current because of the larger emitting area. So what are the difficulties?

- The emitter area may be so small that we have to 'fan' the beam in order to illuminate the area used in TEM. This fanning may actually increase the effect of coma aberration (a radial aberration as noted in Section 28.10). If a CFEG has a source size of ~3 nm, we can study ~15 nm with a 5× magnification. The Schottky source has a source diameter ~10× greater, and the price you pay for this is a decrease in spatial coherence, and a larger energy spread.
- Correcting astigmatism is very tricky with an FEG. As 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

$$\Delta R = \lambda u\left(\Delta f + C_s\lambda^2 u^2\right) \tag{28.70}$$

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

$$\Delta f_{opt} = -MC_s\lambda^2 u_{max}^2 \tag{28.71}$$

where M is a factor between 0.75 and 1. The value for ΔR_{min} is then close to

$$\Delta R_{min} = \frac{1}{4}C_s\lambda^3 u_{max}^3 \tag{28.72}$$

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

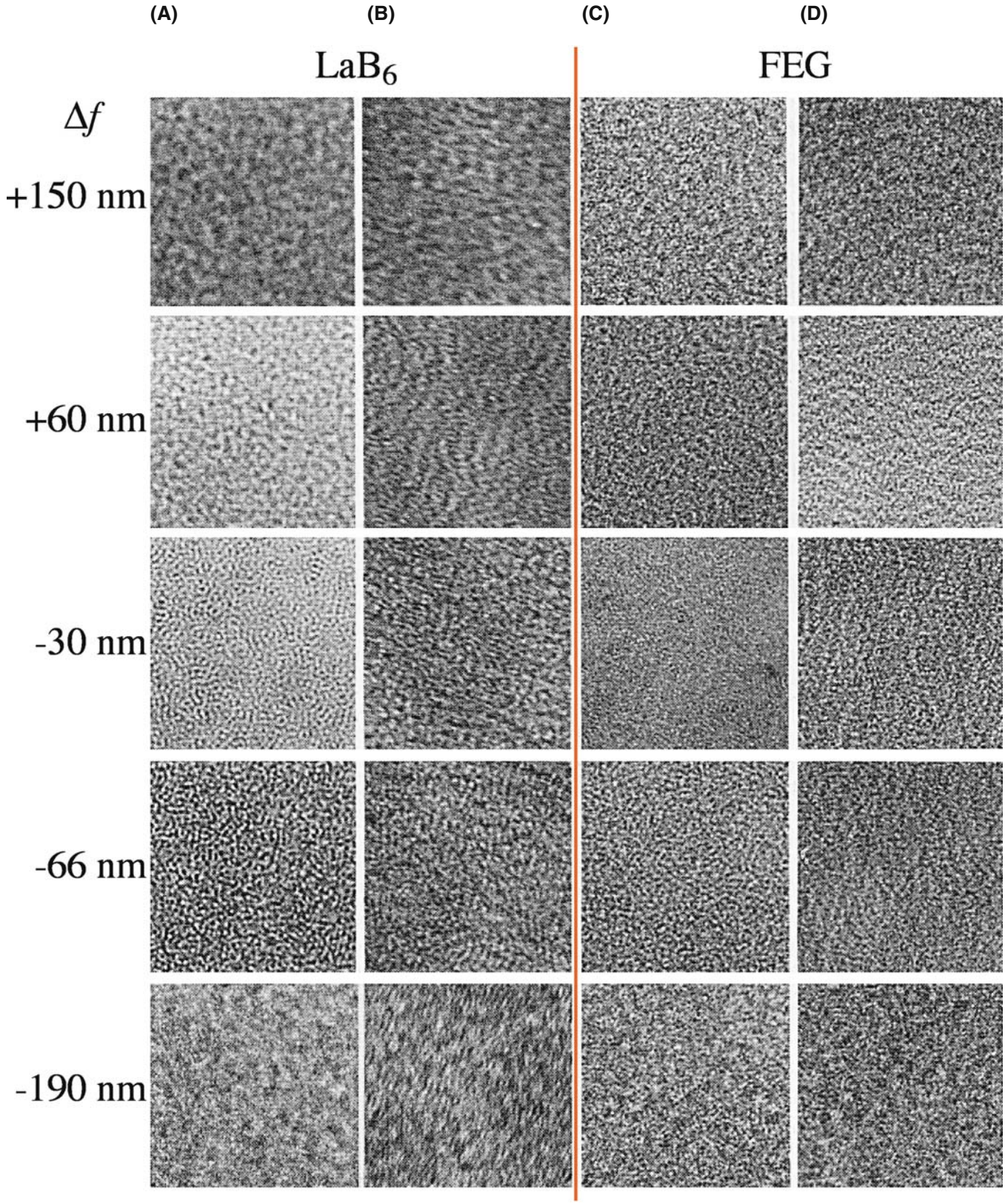

FIGURE 28.11. A tableau of images from an amorphous film at varying defocus. (columns A and B) LaB_6; (columns C and D) FEG. (columns A and C) Without astigmatism; (columns B and D) with astigmatism. With LaB_6 you can easily see the astigmatism, while with an FEG, you can't.

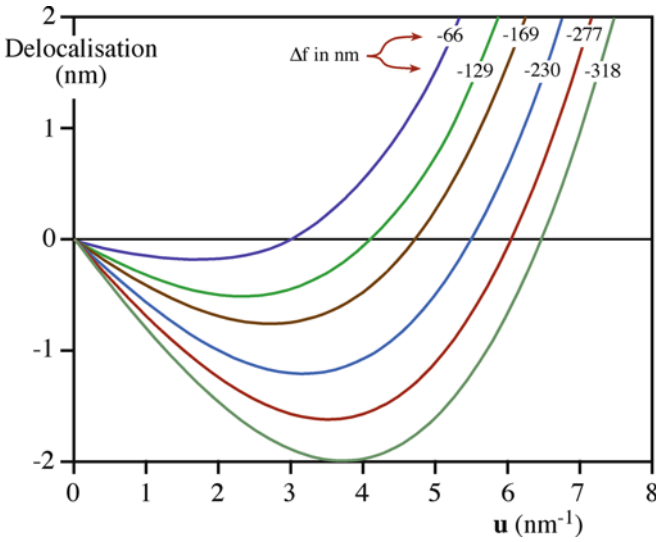

FIGURE 28.12. Image delocalization plotted against **u** as Δf is changed for a Philips CM20 FEG with $C_s = 1.2$ nm.

- 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

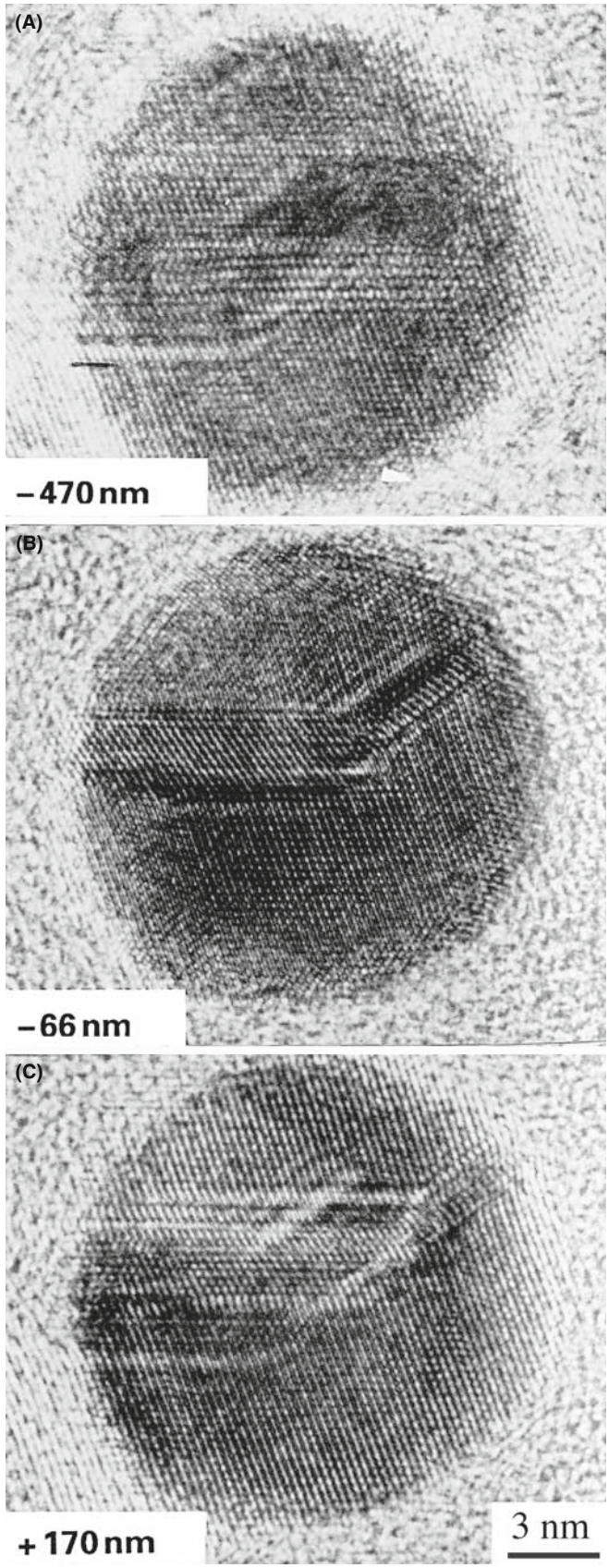

FIGURE 28.13. Experimental images showing delocalization in HRTEM images of an Au particle: (A) underfocus, (B) Scherzer focus, (C) overfocus.

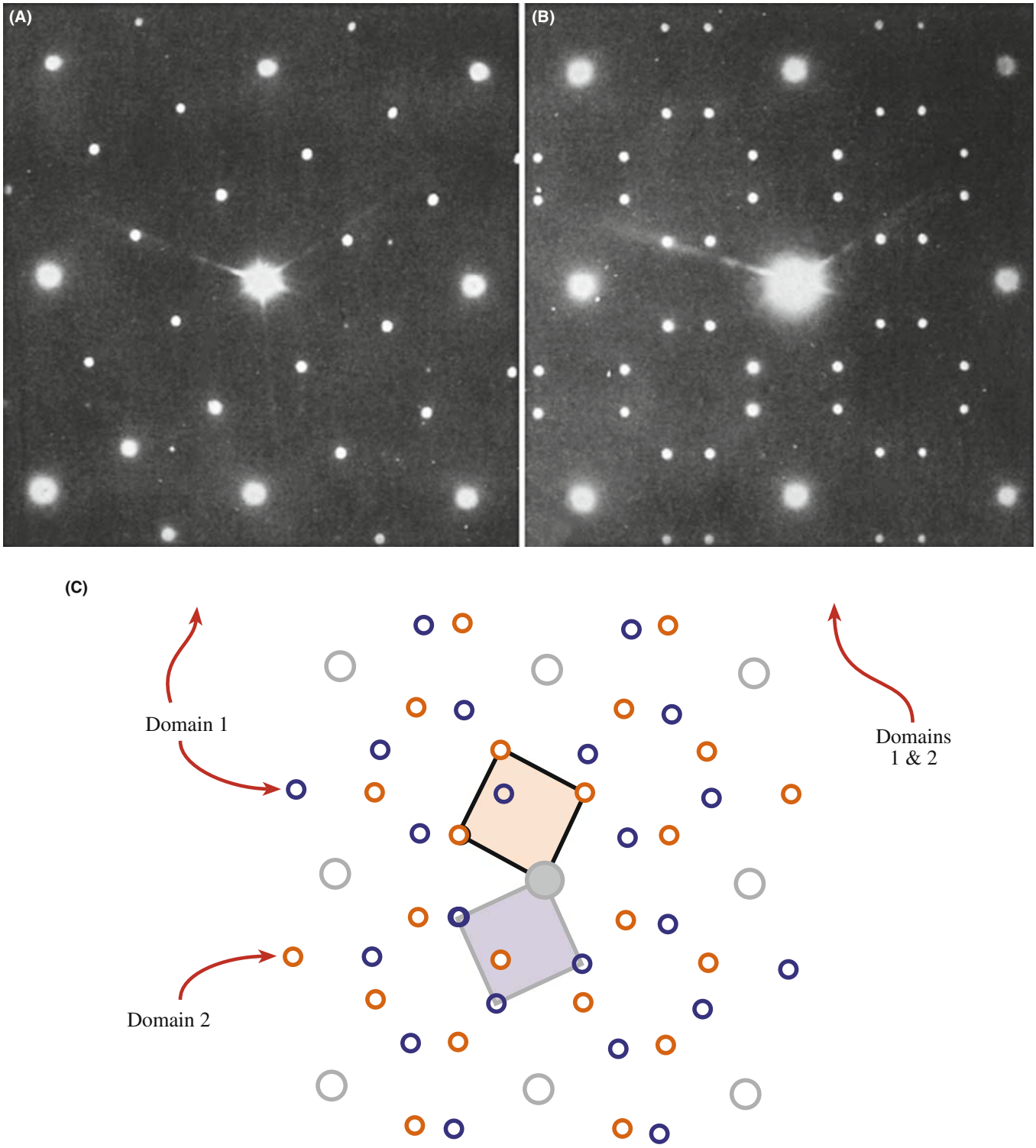

FIGURE 28.14. Two [001] DPs from an ordered alloy of Au_4Mn with schematics. (A) One domain. (B) Two symmetry-related domains. (C) Schematic diagram showing how (B) arises from the relative rotation of the two domains.

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.

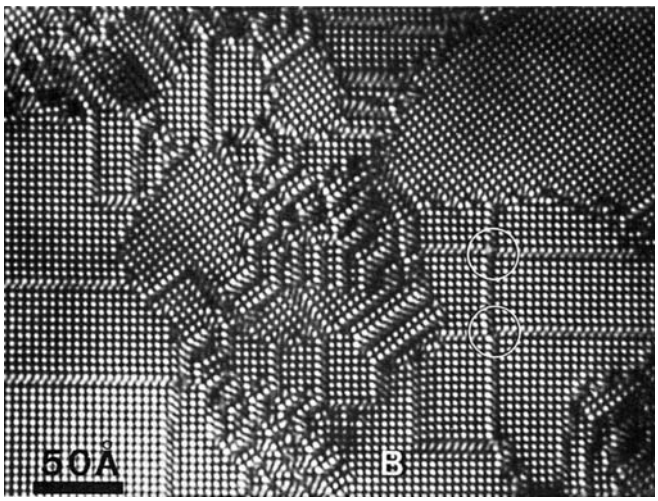

FIGURE 28.15. DF lattice image of Au_4Mn using an objective diaphragm to exclude all the fcc reflections. The two differently oriented domains correspond to the two orientations in Figure 28.14C.

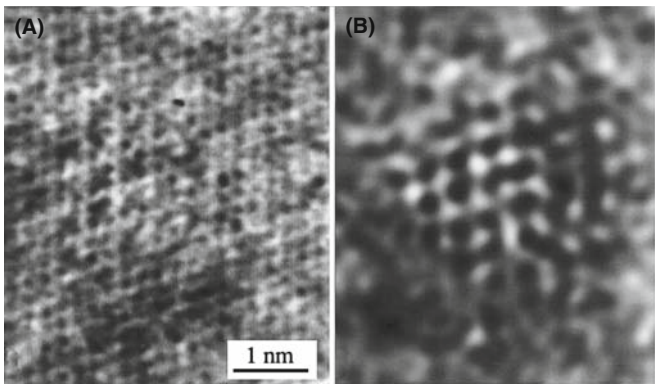

FIGURE 28.16. (A) The experimental image of a small spinel particle in a NiO matrix. The NiO is thicker and dominates the image. (B) After filtering out the NiO contribution (its lattice parameter is twice that of the spinel), we can see the spinel particle and estimate its size. See also Figures 31.2 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.

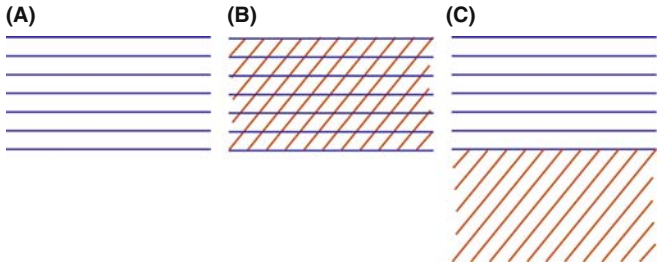

FIGURE 28.17. Schematic HRTEM images of grain boundaries showing (A) one set of fringes in one grain; (B) specimen tilted to give crossing fringes in one grain; (C) one set of fringes in each grain boundary remains parallel to the beam.

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

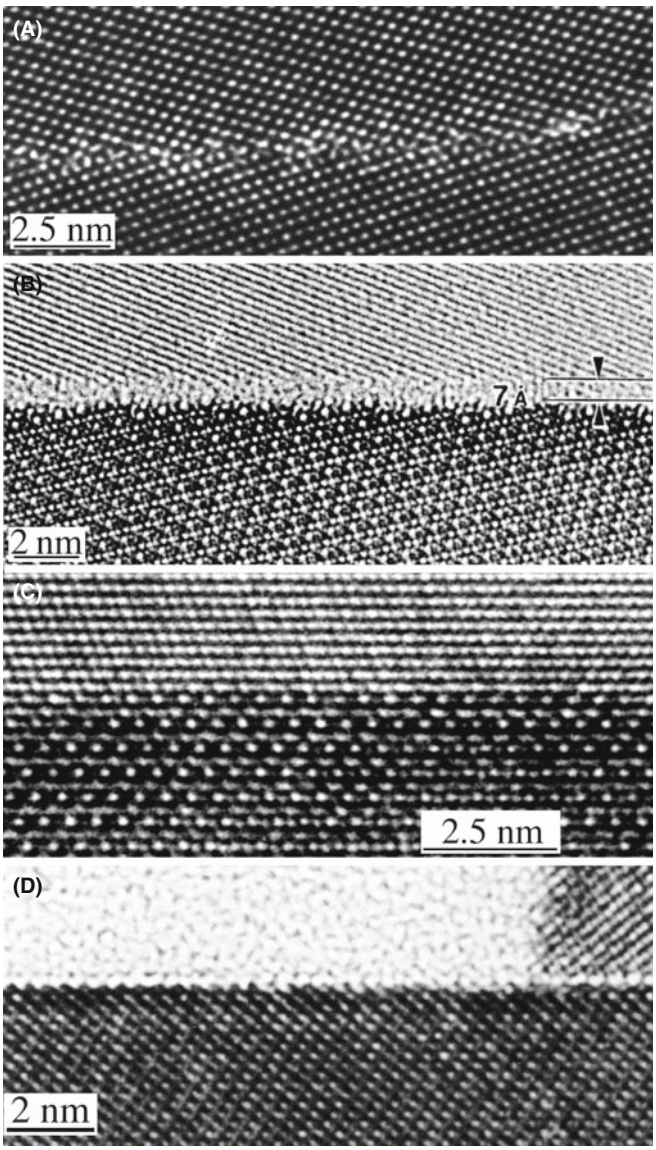

FIGURE 28.18. Examples of HRTEM images of planar interfaces. (A) Grain boundary in Ge; (B) grain boundary in Si_3N_4 with a layer of glass along the interface; (C) phase boundary separating NiO and $NiAl_2O_4$; (D) profile images of the (0001) surface of Fe_2O_3.

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-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!

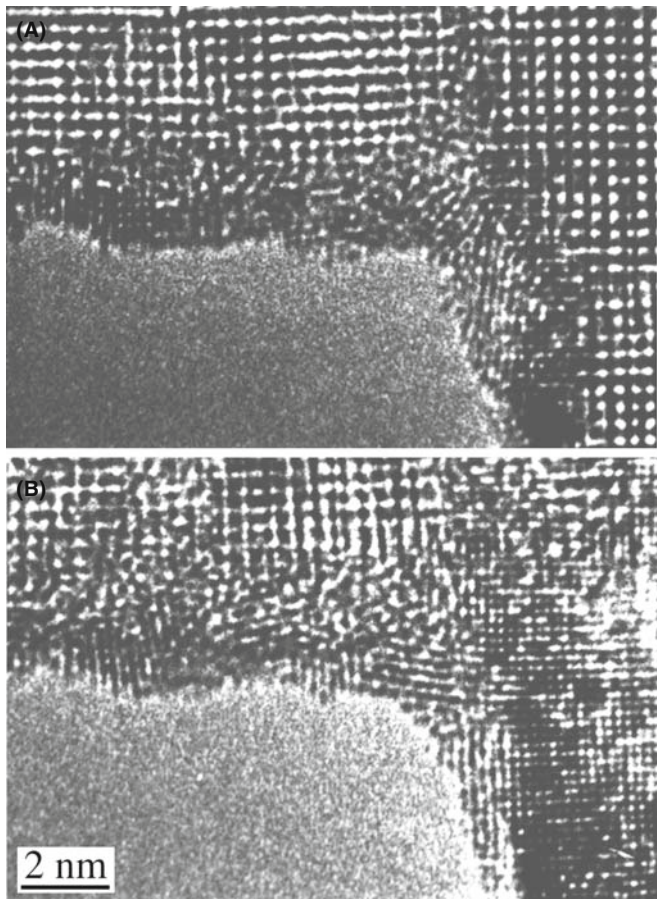

FIGURE 28.19. Reduction of Nb oxide to the metal by beam-induced loss of oxygen during observation of the edge of the foil. The reduction increases with time.

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.

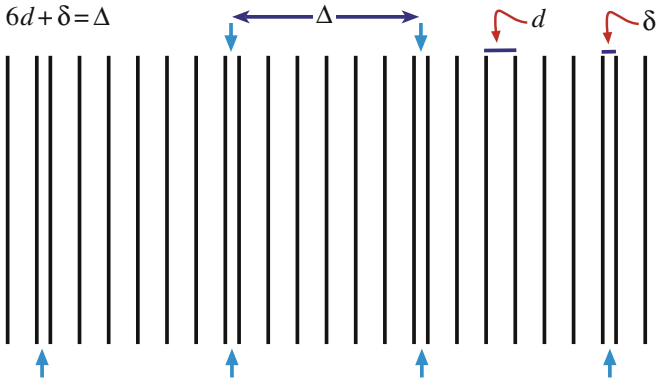

FIGURE 28.20. An incommensurate structure formed by inserting a planar defect after every seventh layer to expand the lattice by δ every seventh plane.

- 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 $Bi_2Sr_2Ca_nCu_{n+1}O_{2n+\delta}$ so for $n = 1$ we have a sequence of layers (planes) described as BiO-SrO-CuO_2-Ca-CuO_2-SrO-BiO, as shown in Figure 28.21. The DPs depend on the particular value of n in the chemical formula and show rows of satellite reflections due to the modulation of the 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

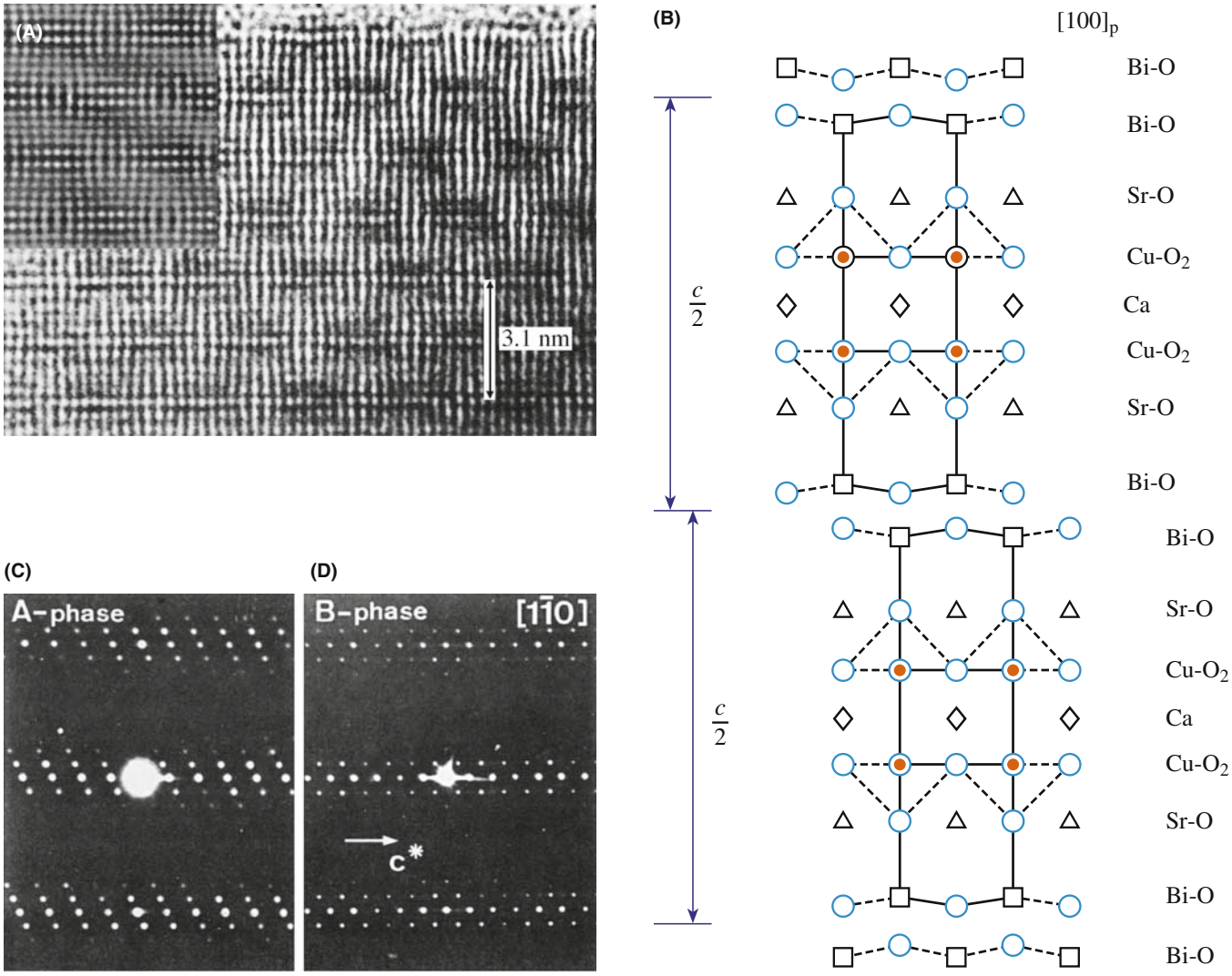

FIGURE 28.21. (A) HRTEM image of the superconductor $Bi_2Sr_2Ca_nCu_{n+1}O_{2n+\delta}$; this simulation (inset) assumes that the lattice relaxation occurs in the Bi-O layer. (B) For $n = 1$, the structure is built up from blocks which are shifted relative to one another. The DPs are from (C) the $n = 0$ phase and (D) the $n = 1.2$ phase. Notice that the spacings of the spots (the satellite sequence) are different.

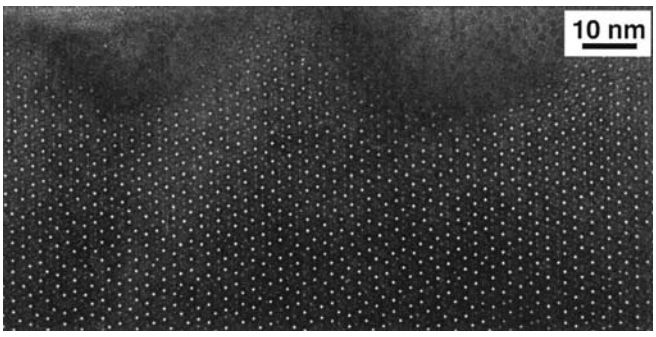

FIGURE 28.22. Tenfold symmetry in a decagonal Al-Mn-Pd quasicrystal.

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

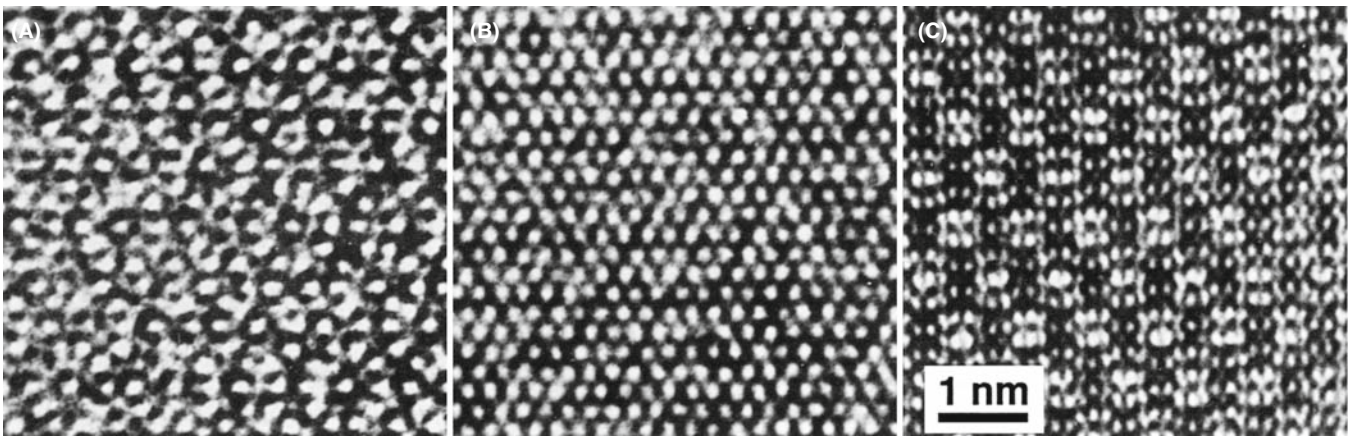

FIGURE 28.23. Images of (A) fivefold, (B) threefold, and (C) twofold projections of an Al-Cu-Li quasicrystal; Δf –27 nm.

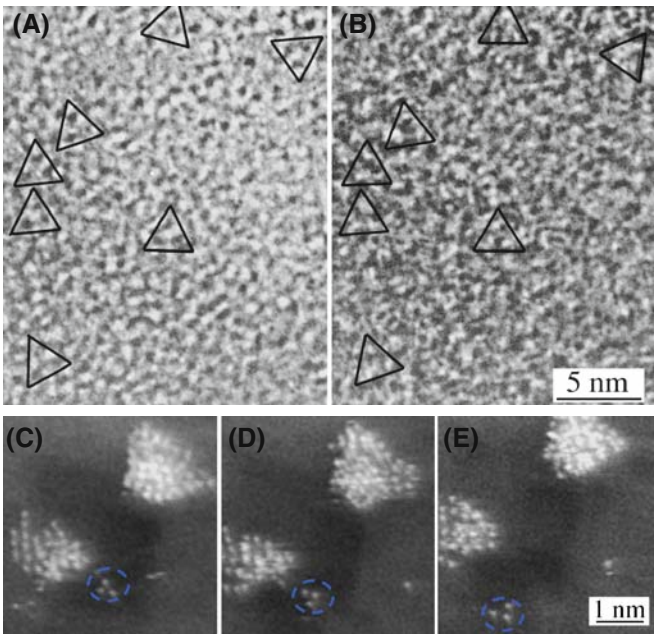

FIGURE 28.24. Images of atoms. (A, B) Triangular arrangements of uranium atoms at different values of defocus in conventional TEM. (C–E) STEM images of Pt/Fe catalyst supported on MgO; time between images is 10–15 seconds. These images have been processed to reduce noise and optimize visibility of the 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, Müler, H, Uhlemann, S., Zach, J, Loebau, U and Hoeschen, R 2008 *Prerequisites for a Cc/Cs-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 $Ti_2Nb_{10}O_{27}$ 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 $Ti_2Nb_{10}O_{27}$.
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,. . .,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.