

Accurate structure determination from image reconstruction in ADF STEM

P. D. NELLIST* & S. J. PENNYCOOK†

**Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK*

†*Oak Ridge National Laboratory, Solid State Division, PO Box 2008, Oak Ridge, TN 37831-6030, U.S.A.*

Key words. ADF, image reconstruction, incoherent imaging, probe reconstruction, STEM, structure determination.

Summary

Annular dark-field (ADF) imaging in a scanning transmission electron microscope results in direct structure images of the atomic configuration of the specimen. Since such images are almost perfectly incoherent they can be treated as a convolution between a point-spread function, which is simply the intensity of the illuminating electron probe, and a sharply peaked object function that represents the projected structure of the specimen. Knowledge of the object function for an image region of perfect crystal allows the point-spread function to be directly determined for that image. We examine how the object function for an image can then be reconstructed using a Wiener filter, the CLEAN algorithm and a maximum entropy reconstruction. Prior information is required to perform a reconstruction, and we discuss what nature of prior information is suitable for ADF imaging.

1. Introduction

The straightforward way that structural information is codified in atomic-resolution annular dark-field (ADF) images formed in a scanning transmission electron microscope (STEM) means that there are many opportunities for image processing in order to perform quantitative image analysis. ADF imaging is close to being perfect incoherent imaging, the principles of which as applied to light optics were first discussed more than 100 years ago by Lord Rayleigh (1896). He noted that incoherent imaging gave a doubling in resolving power over the coherent case for the same imaging lens configuration. More importantly, however, he discussed how interference effects from spatially separated parts of the specimen were suppressed, which is very important in transmission electron microscopy (TEM)

because it suppresses the complicated effects that dynamical scattering can have on the image.

Conventional high-resolution TEM (HRTEM) can be regarded as an almost perfectly coherent mode of imaging, where the specimen is illuminated by a nearly plane-wave source. Interference effects between multiply scattered beams as the electron wavefront propagates through the specimen, described by dynamical diffraction theory, mean that in general the images cannot be directly interpreted in terms of the structure of the specimen. At certain thickness and focus combinations, structure images of perfect crystals can be formed (for example, see Spence *et al.*, 1977), but it is not clear whether the maxima or minima of intensity should be associated with the atomic-columns, and at defects such stationary points can be significantly displaced from the atomic columns (Saxton & Smith, 1985; Bourret *et al.*, 1988). In spite of this, there has been a growing interest in quantitative image analysis in HRTEM. Accurate structure determinations can be performed by iteratively matching trial models of the structure to the experimental image by simulating the imaging process (for example Möbus, 1996; Möbus & Dehm, 1996). These procedures require a dynamical diffraction calculation for each iteration and are therefore relatively expensive in computer time. Quantitative chemical mapping techniques have also been developed (Baumann *et al.*, 1995), but are limited to perfect crystals of known structure. In general, these techniques rely on having a reasonably high degree of prior knowledge about the specimen before the refinement procedure starts.

Since ADF imaging does not codify the specimen information in such a complicated way, it lends itself to direct structure determination and does not require such a high degree of prior knowledge. Structure determinations have been performed on a variety of specimens (for example,

see McGibbon *et al.*, 1994, 1995; Chisholm *et al.*, 1994), often leading to hitherto unpredicted structures. The relatively high angles of scatter used to form an ADF image also lead to a strong dependence of the image intensities on the atomic number of the atoms being imaged (for example Nellist & Pennycook, 1996), hence the name *Z*-contrast.

Here we make use of the important attribute of incoherent imaging that there is no phase problem because the image intensity is a convolution between the intensity of the illuminating STEM probe and an object function that is directly related to the projected structure of the specimen. From an image region of a specimen of known structure, the probe intensity profile can be directly determined. Using this we examine various approaches to how accurate atomic-column positions may be determined. Because of the resolution limit of the microscope, information is lost and an unambiguous reconstruction is impossible without prior knowledge. We examine to what extent the various approaches used depend on such prior knowledge of the specimen.

2. The principles of incoherent imaging using an ADF detector in an STEM

The concept of incoherent imaging does not mean that the partial waves propagating through the image-forming lens are incoherent with respect to each other, since in this case no image contrast would be formed, rather that there is no interference between waves emitted from spatially separated parts of the specimen. This is obviously the case for an extended self-luminous object, but Lord Rayleigh (1896) also discussed that illuminating the specimen with a large, incoherent radiation source could achieve the same effect, and he also suggested that this was an important function of the condenser lens in the illumination system.

In practical terms it is found that incoherent imaging is more easily attained in the STEM than in the conventional transmission electron microscope (CTEM). Imaging in the STEM is intimately related to imaging in the CTEM by the principle of reciprocity (Cowley, 1969; Zeitler & Thomson, 1970) which essentially states that source and detector points in an optical system can be interchanged and the same intensity will be measured. Thus STEM imaging can be thought of as a source in the image plane that can be effectively scanned over the image plane by means of scan coils, followed by the image-forming lens, which physically acts to focus an image of the source to form a small electron probe at the specimen. Detectors are now placed in the far-field beyond the specimen, and the detected intensity recorded as a function of the scanned probe position to form an image. By reciprocity we can see that the detector in an STEM is equivalent to the illuminating electron source

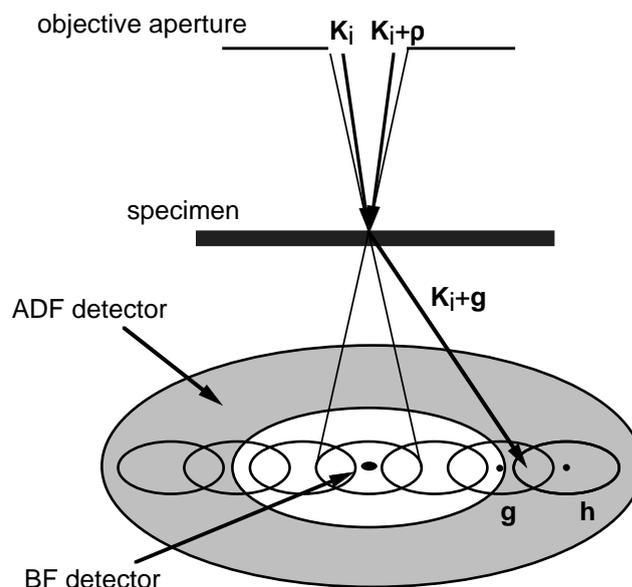


Fig. 1. The scattering geometry for ADF imaging in the STEM. The wavevectors indicate possible incident and scattered partial plane waves that can contribute to the ρ image spatial frequency.

in a CTEM; for example, a small axial detector in an STEM is equivalent to almost plane-wave illumination in the CTEM, and in principle gives an identical image. It is now clear that if we wish to achieve the STEM equivalent of a large incoherent source in CTEM, all we need to do is use a large detector that integrates the intensity scattered over a large range of angles.

To analyse in detail the properties of an image formed using a large detector in an STEM, we start by considering the incident cone of partial plane-waves that are focused by the objective lens to form the illuminating probe (Fig. 1). Each partial plane-wave has a complex amplitude given by the function $A(\mathbf{K}_i)$, where \mathbf{K}_i is the transverse component of the partial plane-wave's wavevector. The phase of $A(\mathbf{K}_i)$ is controlled by the lens aberration terms, such as defocus and spherical aberration, and the magnitude of $A(\mathbf{K}_i)$ is a circular top-hat function defined by the position and size of the objective aperture present. The complex amplitude of the probe is given by integrating over all the partial plane-waves,

$$P(\mathbf{R}) = \int A(\mathbf{K}_i) \exp[-i2\pi\mathbf{K}_i \cdot \mathbf{R}] d\mathbf{K}_i, \quad (1)$$

which is simply the Fourier transform of $A(\mathbf{K}_i)$. It is also clear that multiplying $A(\mathbf{K}_i)$ by a phase factor $\exp[i2\pi\mathbf{K}_i \cdot \mathbf{R}_0]$ in Eq. (1) gives $P(\mathbf{R} - \mathbf{R}_0)$, which is simply a shift of the probe to the position \mathbf{R}_0 .

Let us now consider that there is an elastic scattering process that scatters the incident partial plane-wave, \mathbf{K}_i , into a final plane-wave, \mathbf{K}_f , changing the amplitude and

phase of the wave by the complex multiplier $\Psi(\mathbf{K}_f, \mathbf{K}_i)$. In the far-field that is the STEM detector plane, \mathbf{K}_f defines a position, at which the measured intensity is

$$I(\mathbf{K}_f, \mathbf{R}_0) = \left| \int A(\mathbf{K}_i) \exp[i2\pi\mathbf{K}_i \cdot \mathbf{R}_0] \Psi(\mathbf{K}_f, \mathbf{K}_i) d\mathbf{K}_i \right|^2 \quad (2)$$

where we have integrated over the partial plane waves forming the probe. Expanding the modulus squared gives a double integral,

$$I(\mathbf{K}_f, \mathbf{R}_0) = \iint A(\mathbf{K}_i) A^*(\mathbf{K}'_i) \exp[i2\pi(\mathbf{K} - \mathbf{K}'_i) \cdot \mathbf{R}_0] \times \Psi(\mathbf{K}_f, \mathbf{K}_i) \Psi^*(\mathbf{K}_f, \mathbf{K}'_i) d\mathbf{K}_i d\mathbf{K}'_i. \quad (3)$$

However, this expression can be reduced to a single integral by taking the Fourier transform of Eq. (3) with respect to \mathbf{R}_0 to give an entirely reciprocal-space expression as a function of image spatial frequency, $\boldsymbol{\rho}$,

$$i(\mathbf{K}_f, \boldsymbol{\rho}) = \int A(\mathbf{K}_i) A^*(\mathbf{K}_i + \boldsymbol{\rho}) \Psi(\mathbf{K}_f, \mathbf{K}_i) \Psi^*(\mathbf{K}_f, \mathbf{K}_i + \boldsymbol{\rho}) d\mathbf{K}_i. \quad (4)$$

Thus the contribution to an image at the spatial frequency, $\boldsymbol{\rho}$, comes from interference between pairs of partial plane-waves in the illuminating convergent beam with wavevectors, \mathbf{K}_i , separated by $\boldsymbol{\rho}$ (Fig. 1). In the case of only Bragg scattering by a crystal, the detector plane will contain a coherent convergent-beam electron diffraction pattern, and the image contrast arises from overlapping discs (Spence & Cowley, 1978) separated by the reciprocal space vector, $\boldsymbol{\rho}$.

The Fourier transform of the ADF image, $i_{\text{ADF}}(\boldsymbol{\rho})$, is then found by integrating Eq. (4) over some detector function, $D_{\text{ADF}}(\mathbf{K}_f)$, with respect to \mathbf{K}_f . For the moment we will make the assumption that the specimen is thin and neglect three-dimensional propagation effects within the crystal. The scattering factor, Ψ , can then be written as a function of $\mathbf{K}_f - \mathbf{K}_i$, so the Fourier transform of the ADF image is

$$i_{\text{ADF}}(\boldsymbol{\rho}) = \iint D_{\text{ADF}}(\mathbf{K}_f) A(\mathbf{K}_i) A^*(\mathbf{K}_i + \boldsymbol{\rho}) \Psi(\mathbf{K}_f - \mathbf{K}_i) \times \Psi^*(\mathbf{K}_f - \mathbf{K}_i - \boldsymbol{\rho}) d\mathbf{K}_i d\mathbf{K}_f, \quad (5)$$

which may be rewritten

$$i_{\text{ADF}}(\boldsymbol{\rho}) = \iint A(\mathbf{K}_i) A^*(\mathbf{K}_i + \boldsymbol{\rho}) D_{\text{ADF}}(\mathbf{K}'_f + \mathbf{K}_i) \times \Psi(\mathbf{K}'_f) \Psi^*(\mathbf{K}'_f - \boldsymbol{\rho}) d\mathbf{K}_i d\mathbf{K}'_f, \quad (6)$$

where we have substituted \mathbf{K}'_f for $\mathbf{K}_f - \mathbf{K}_i$. The range of integration over \mathbf{K}_i is limited by the product of the aperture functions, A , to a shape given by the overlap of two discs with the same radius as the objective aperture, separated by $\boldsymbol{\rho}$. If the detector function, D_{ADF} is much larger than the objective aperture, we can neglect \mathbf{K}_i in its argument and

separate the integrals in Eq. (6), thus

$$i_{\text{ADF}}(\boldsymbol{\rho}) = \int A(\mathbf{K}_i) A^*(\mathbf{K}_i + \boldsymbol{\rho}) d\mathbf{K}_i \times \int D_{\text{ADF}}(\mathbf{K}'_f) \Psi(\mathbf{K}'_f) \Psi^*(\mathbf{K}'_f - \boldsymbol{\rho}) d\mathbf{K}'_f = t(\boldsymbol{\rho}) o(\boldsymbol{\rho}) \quad (7)$$

where $t(\boldsymbol{\rho})$ is the function representing the integral over \mathbf{K}_i and $o(\boldsymbol{\rho})$ represents the integral over \mathbf{K}'_f . Physically, this approximation neglects any overlap regions that are intersected by the inner edge of the ADF detector, and assumes that the whole overlaps detected by the large area of the detector dominate. For the case of Fig. 1 this approximation is exact since no overlaps are intersected by the inner radius of the detector. Finally, we note that $t(\boldsymbol{\rho})$ is the autocorrelation of A , and since the Fourier transform of the modulus squared of a function is the autocorrelation of the Fourier transform of that function, Eq. (7) can be written in real space as a convolution,

$$I_{\text{ADF}}(\mathbf{R}_0) = |P(-\mathbf{R}_0)|^2 \otimes O(\mathbf{R}_0) = \int |P(\mathbf{R} - \mathbf{R}_0)|^2 O(\mathbf{R}) d\mathbf{R} \quad (8)$$

where $O(\mathbf{R}_0)$ is the inverse Fourier transform of $o(\boldsymbol{\rho})$. Equation (8) is an extremely important equation; it is the definition of incoherent imaging when points in the specimen separated with a component transverse to the electron beam are almost perfectly incoherent and do not interfere, and therefore whose contributions to the image are added in intensity. An STEM image can be written in this form only when all the partial plane-waves in the illuminating cone can contribute with a similar weight to the image-forming process. Here this is seen to occur for a thin specimen that can be expressed by a complex transmission function and a large detector. If we now consider the electron wave propagation within the specimen, we have to use dynamical diffraction theory. It is found (Pennycook & Jesson, 1990) that the scattering to a high-angle ADF detector is mostly from the 1s-type states that are tightly bound to the atomic columns. Since these states are highly nondispersive, the approximation of $\Psi(\mathbf{K}_f - \mathbf{K}_i)$ for $\Psi(\mathbf{K}_f, \mathbf{K}_i)$ can still be made, and incoherent imaging applies.

To compare Eq. (8) with bright-field (BF) imaging in a CTEM we must examine what happens to Eq. (5) when we assume that we are using a small axial detector in an STEM. Assuming a CTEM with perfect coherence, we can substitute $\delta(\mathbf{K}_f)$ for the $D_{\text{ADF}}(\mathbf{K}_f)$ function. Equation (5) is no longer separable as the product of two integrals, rather it shows that the BF image can be written

$$I_{\text{BF}}(\mathbf{R}_0) = |P(-\mathbf{R}_0) \otimes \psi(\mathbf{R}_0)|^2 = \left| \int P(\mathbf{R} - \mathbf{R}_0) \psi(\mathbf{R}) d\mathbf{R} \right|^2 \quad (9)$$

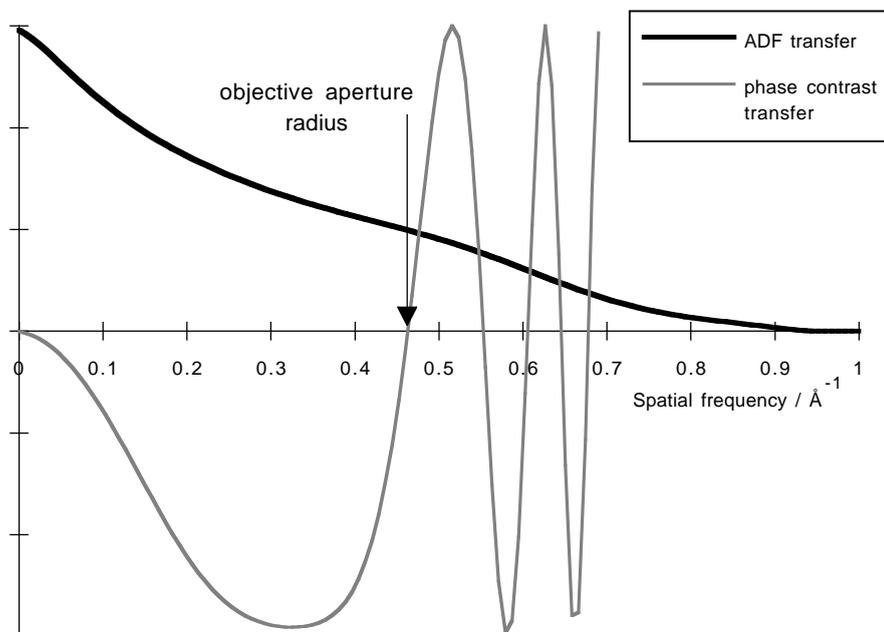


Fig. 2. A comparison of the weak-phase contrast transfer function for CTEM imaging with the incoherent transfer function for ADF STEM imaging. The accelerating voltage is 300 kV, $C_s = 1$ mm, and the defocus was set to -40 nm. The ADF transfer assumes the presence of an objective aperture with a radius as marked at the first zero of the weak-phase contrast transfer function.

where $\psi(\mathbf{R}_0)$ is the inverse Fourier transform of Ψ assuming still a multiplicative object. This equation assumes that all spatially separated points in the specimen remain coherent. The first point to make comparing the last two equations is that Eq. (8) transfers spatial frequencies from the specimen going up to twice the maximum found for Eq. (9) for the same $P(\mathbf{R}_0)$. The convolution with $P(\mathbf{R}_0)$ in Eq. (9) limits the transfer at the spatial frequency corresponding to the objective aperture radius, whereas $|P(\mathbf{R}_0)|^2$ contains spatial frequencies up to twice this limit. Figure 2 demonstrates this effect by comparing the incoherent transfer function, $t(\rho)$, with the conventional coherent phase-contrast transfer function for the same imaging parameters, which are those for a VG Microscopes HB603U 300-kV STEM ($C_s = 1$ mm). Similarly, for the Bragg scattering shown in Fig. 1, the conventional BF image will not show any contrast because the lattice spacing is beyond the resolution limit so there is no overlap and interference between the diffracted discs at the detector on the optical axis. An ADF detector will detect the disc-overlap interference at higher angles and will therefore resolve these lattice planes. It must be conceded that the presence of noise combined with the smoothly decaying form of the transfer function will impose a practical resolution limit slightly worse than twice the aperture radius; nevertheless spacings of 0.13 nm are routinely resolved in a microscope with a conventional weak-phase point resolution of 0.19 nm.

The important attribute of Eq. (8), which is the basis of this paper, is that the ADF image intensity is itself a convolution between an object function and a real, positive point spread function (PSF) that is simply the intensity of the illuminating STEM probe. If we know, or can determine, the

profile of the probe then we can simply deconvolve it from Eq. (8) resulting in the restoration of the object function. In contrast, Eq. (9) shows how the phase of the convolution between the probe's complex amplitude and the specimen function is lost in conventional HRTEM by taking the modulus squared. To deconvolve the transfer function, the phase problem must first be solved using, for example, holography (Orchowski *et al.*, 1995) or focal-series reconstruction (Coene *et al.*, 1992).

In this paper we do not wish to consider in detail the derivation of the ADF object function, $O(\mathbf{R}_0)$. We have already mentioned that it can also be shown that there is negligible interference between neighbouring columns (see also Jesson & Pennycook, 1993). However, the detector geometry alone is not effective at breaking the coherence between atoms within the same atomic column that is aligned with the beam. Since the scattering is relatively high-angle, most of the electrons detected have also undergone phonon scattering. Although this has little effect on the transverse coherence, which has already been broken by the detector geometry, it is effective at reducing coherent effects within a single column (Jesson & Pennycook, 1995). The crucial point for the analysis used later in this paper is that the object function consists of highly localized sources at the positions of the atomic columns, with a strength that approximates to the square of the atomic number of the atomic species present in that column at the limit of high inner radius.

3. Probe intensity profile reconstruction

One of the major strengths of atomic-resolution microscopy is that it can image the structure of defects, such as

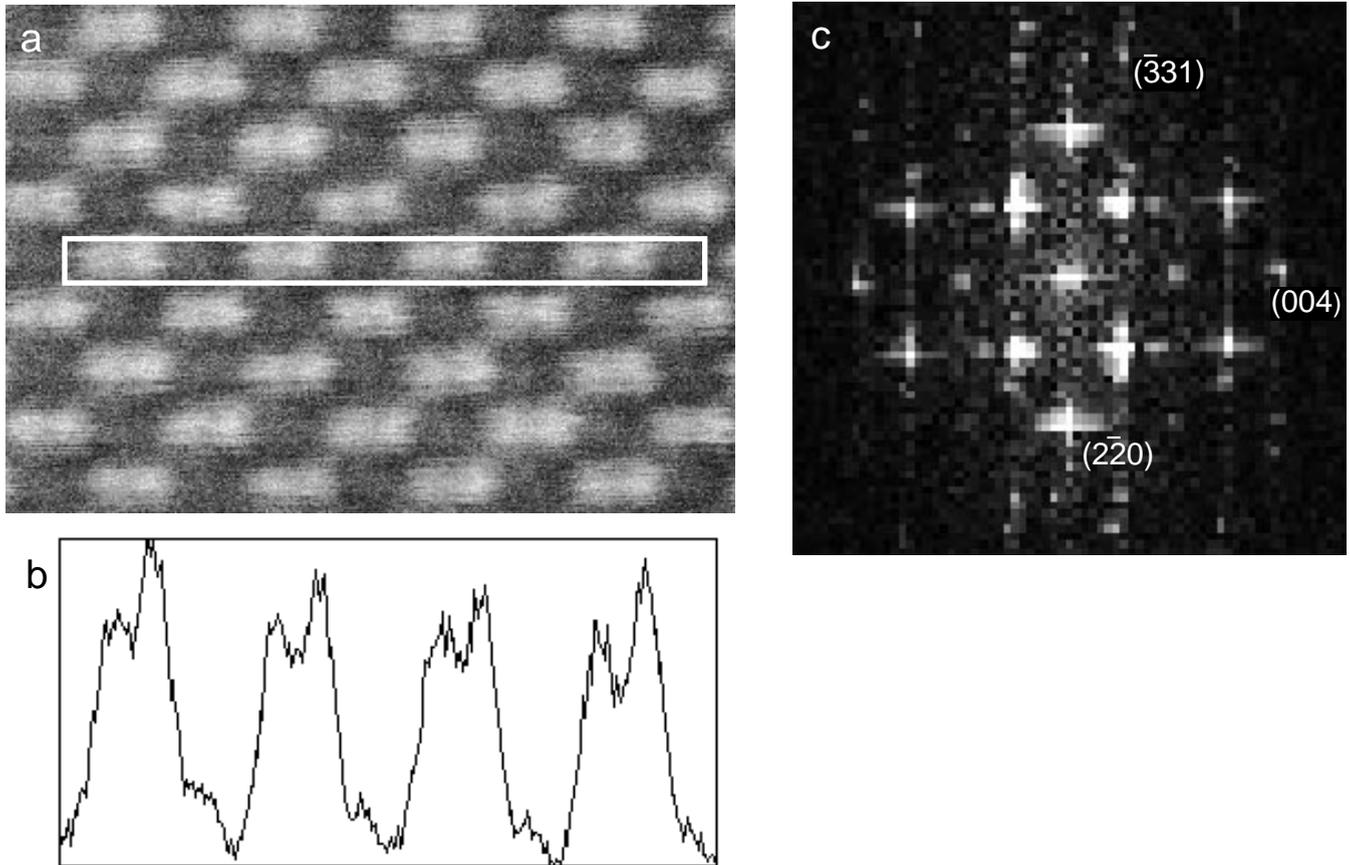


Fig. 3. (a) An ADF image of GaAs<110>. The 0.14-nm dumbbell spacing is resolved. (b) An intensity profile plot of the marked region of the image. Note the obvious polarity of the lattice. (c) The magnitude of the Fourier transform of the image showing spots out as far as weak {311}-type spots, demonstrating a 0.128-nm resolution.

a dislocation core or an interface, within a crystalline matrix. Often the defect area of interest in the image is surrounded by large regions of image of the perfectly crystalline matrix which is usually of a known structure. For such an image, there exists the opportunity of using the image region of the known crystalline structure to determine something about the microscope function, which can then be used to analyse in greater detail the defect area of interest.

For incoherent imaging, Eq. (7) suggests that if the object function is known for a region of the image, then it can be deconvolved resulting in a direct determination of the PSF. In practice, the PSF cannot be uniquely determined from an image of a crystal, but this is most easily illustrated by considering an example. Figure 3(a) shows a region of an image of GaAs<110> in the VG HB603U STEM and can be seen to be a direct structure image of the material, resolving the 0.14-nm dumbbell spacing and showing clear chemical or Z-contrast (Fig. 3b). Taking the Fourier transform of this image (Fig. 3c) gives spots in reciprocal space, the magnitude of which is given by the product of the magnitude of the Fourier transform of the object function

at that spot and the strength of the transfer function $t(\rho)$ at that spatial frequency. Thus to determine $t(\rho)$ we need an estimate of the object function. Let us assume that the object function consists of δ -functions at the column sites weighted by Z^2 of the atomic species of that column, we can now estimate $t(\rho)$ for the spatial frequencies present (Fig. 4). The fact that $t(\rho)$ calculated from the (002) spot, which is present purely from the asymmetry in the dumbbell pair, fits nicely with the rest of the data suggests that our Z^2 approximation is reasonable. Although we now have values for $t(\rho)$ at six different spatial frequencies going out as far as the {331} components, nothing is known about the transfer function between these points, illustrating how the PSF is not uniquely determined from a periodic image. We need to introduce *a priori* knowledge, which in HRTEM is usually done by parameterizing the microscope function and performing some kind of best-fit search (see for example Möbus & Dehm, 1996). However, Fig. 2 shows that the incoherent imaging transfer function is usually a smooth, monotonically decaying, real, positive function, and therefore we can reasonably approximate the transfer function by smoothly joining the data points in Fig. 4. Linear

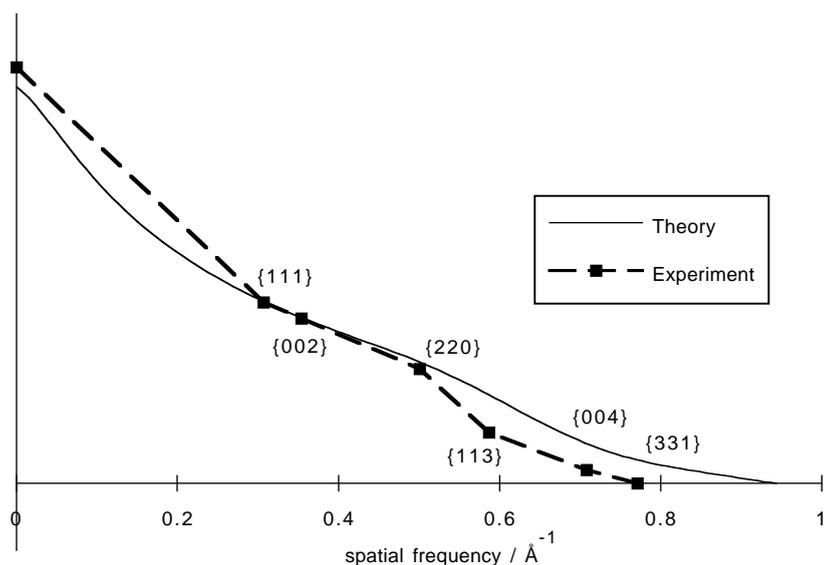


Fig. 4. The reconstructed transfer function using the experimental data at the frequencies shown, compared with the theoretically optimum transfer function copied from Fig. 2. The reduction in the experimental transfer at higher spatial frequencies compared to the optimum transfer function is a result of a slight broadening of the probe, due to the finite size of the electron source, microscope instabilities or a slight overfocusing of the probe.

interpolation between the points is sufficient, as illustrated by Fig. 5 which shows the reconstructed probe intensity profile, which is simply the inverse Hankel transform of Fig. 4. The reconstructed probe intensity profile plotted in Fig. 5, which has a full width at half-maximum of 0.14 nm, is remarkably close to the optimum diffraction-limited probe for the machine used. The slight broadening of the probe is probably due to residual instabilities in the microscope, the finite size of the electron source or a slight overfocus. Two other approximations have also been made in this approach. We have assumed that the PSF has circular symmetry so that the transfer function is a one-dimensional function of $|\rho|$, and where there are different spots in Fig. 3(c) with the same $|\rho|$ we have used the mean transfer function determined from these spots. The scatter in the experimental transfer values for the same $|\rho|$ was in the range 10–20%, and in a more complex analysis this scatter could be used to estimate the two-dimensional profile of the probe, thus taking account of noncircularly symmetric aberrations such as astigmatism. Finally, there is often an unknown constant background present in the images which prevents determination of the transfer function for $|\rho| = 0$. In practice this is found not to be important for the fine detail of the PSE, and we fit $t(0)$ by assuming that the first minimum of the PSF just reaches zero.

4. Object function reconstruction

4.1. The problem

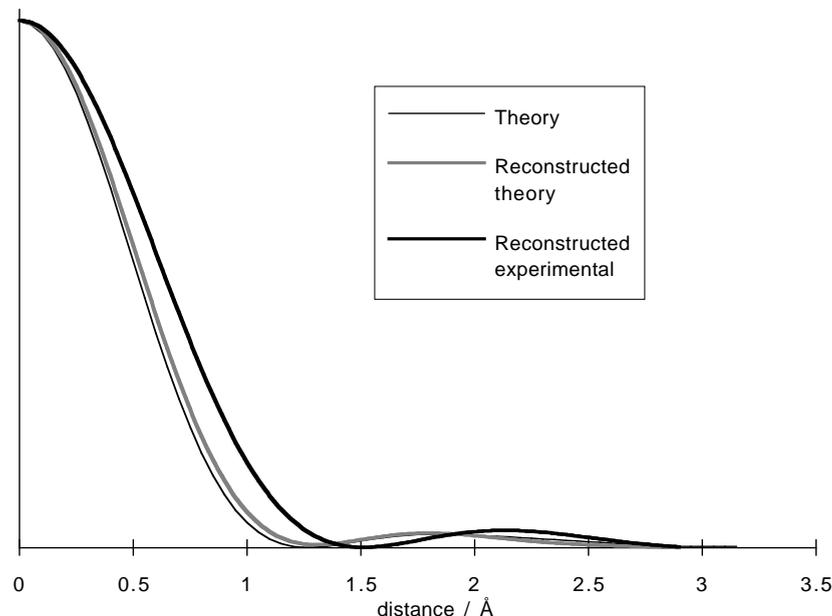
To illustrate the possibilities and associated problems for the object function reconstruction we first pose a simulated problem. An image of Si<110> is simulated by assuming that the object function consists of δ -functions placed in an

arrangement corresponding to the classic dumbbell structure of Si when projected along <110>, with the columns forming the dumbbell pairs being separated by 0.136 nm. In practice, any blurring of the δ -functions by thermal lattice vibration or mechanical vibration of the whole specimen will be taken account of in the probe reconstruction by a slight probe broadening since δ -function sources are assumed for that reconstruction. Convolution of the object function with a simulated probe function using the parameters for the VG HB603U microscope, but with a relatively large underfocus of –60 nm, results in Fig. 6. In practice we find that human operators can achieve a much more accurate focus condition than the one simulated here, focusing being a relatively straightforward procedure in incoherent imaging. However, this focus condition is of interest because the distance between the peak pixel positions in one dumbbell pair is 36 pixels which is 12% higher than the 32-pixel dumbbell spacing in the object function. Even in this extreme case the absolute error in the peak positions is only 0.017 nm, which experimentally will be close to the accuracy limits imposed by the blurring of the object because of the vibration of the specimen, but this is still an illustration of how a resolution-limited system can move peak intensities from the exact column positions, as is also observed in conventional HRTEM imaging (Saxton & Smith, 1985). The aim of an object function reconstruction method should be to restore the object function to allow accurate structure determination.

4.2. Multiplicative deconvolution

The obvious next step is to divide out the reconstructed transfer function from Eq. (7), resulting in a reconstructed object function. Care must be taken to avoid divide by zero

Fig. 5. The reconstructed probe intensity profile from the experimental data compared with the theoretical profile calculated assuming the optimum lens parameters for the microscope used. Also shown is a profile (labelled Reconstructed theory) reconstructed in a similar way to the experimentally determined one, but using the theoretically optimum transfer values at the spatial frequencies available in an image of GaAs<110> to check the effect of a linear interpolation in reconstructing the transfer function.



problems when the transfer function goes to zero, which can be implemented using a Wiener filter (for a discussion of Wiener filters see for example Bates & McDonnell, 1986), thus

$$\tilde{o}(\boldsymbol{\rho}) = \frac{t^*(\boldsymbol{\rho})i_{\text{ADF}}(\boldsymbol{\rho})}{|t(\boldsymbol{\rho})|^2 + \epsilon} \quad (10)$$

where $\tilde{o}(\boldsymbol{\rho})$ is an estimate of $o(\boldsymbol{\rho})$ and ϵ is a small constant. The effect of performing this on the simulated Si data is shown in Fig. 7. Although the peak intensity positions for the dumbbells are now separated by 34 pixels, only 6% larger than the actual spacing, the object function is far from being a dumbbell array of δ -functions. The atomic sites are indicated by a large diffuse region of intensity, but there are also weaker, artefactual regions of intensity in the 'tunnels' between the dumbbells. However, this reconstructed object function, when convolved with the PSF, will still give an image very similar to the original, and thus is a valid object function for these data.

The important conclusion to draw therefore is that the object function is not uniquely defined by the image data. This is obvious when we consider the transfer function of the microscope in reciprocal space (Fig. 4) which falls to zero at the incoherent resolution limit of the microscope. Information in the object function at frequencies beyond this limit cannot be transferred by the microscope, and therefore any object function with the correct Fourier spectrum within the resolution limit can be convolved with the PSF to give the correct image data. The Wiener filter sets the object function to zero for spatial frequencies above the resolution limit, resulting in an effective transfer function that is quite sharply truncated. In real-space it is now clear that the Wiener deconvolved object function is actually a

convolution between the original array of δ -functions and a new PSF which is the Fourier transform of the truncated effective transfer function which has relatively large side lobes giving rise to the artefacts in the object function. A general point is thus raised that if we want accurate atomic column locations, which is equivalent to reconstructing an object function with δ -function like peaks, then we need a further constraint than just the experimental data.

4.3. The CLEAN algorithm

It is striking that incoherent imaging in TEM has many similarities to imaging in astronomy, albeit at a different length scale. Images in both modes often consist of a uniform background with discrete localized sources forming the detail in the object, which are for example stars in astronomy and atomic columns in TEM, but then convolved by a PSF that is restricted in Fourier space. It therefore seems sensible to try techniques that have been successfully applied in astronomy. One such technique is called the CLEAN algorithm, and was initially proposed by Högbom (1974). It is a subtractive deconvolution technique that, in the implementation used here, follows the iterative loop:

- 1 locate the pixel in raw data that has the maximum intensity;
- 2 transfer a fraction, γ , known as the loop gain, of the maximum intensity to the CLEANed reconstructed object function at that pixel location;
- 3 subtract a PSF from the raw data, centred at the peak pixel position and with a height of γ times that intensity;
- 4 test the raw data to see if the contrast, measured here simply by the variance of the intensity data, has fallen below some previously specified criterion.

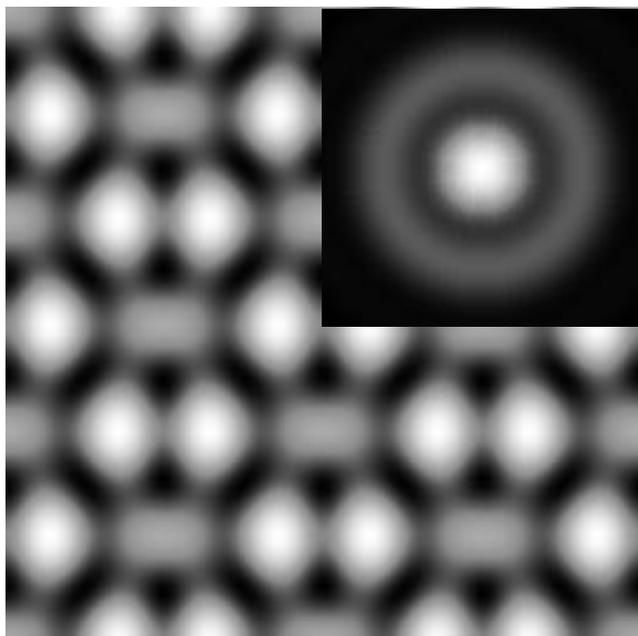


Fig. 6. A simulated image of Si<110> using an underfocused probe (shown in the inset). Each pixel is 0.00425 nm wide.

This algorithm has been run on the simulated image data described in Section 4.1, using $\gamma=0.1$ and allowing the variance to fall to 1/100 of its original value, resulting in the reconstructed object function shown in Fig. 8. Although this simulation did not include the effects of noise in the original image, it is still instructive to examine its effect. The reconstructed object function is now much more localized, indicating that it has been reconstructed to spatial frequencies higher than the resolution limit. However, the peak pixels of the dumbbell pair are still spaced at 36 pixels, as per the peaks in the raw data, which is perhaps not surprising since the largest single contribution to the CLEANed object for an iteration will be from the peak pixels in the raw data. As the algorithm proceeds, pixels with a closer spacing are excited to compensate, leading to a streaked object function with some excited pixels only 26 pixels apart.

Thus it can be seen that the CLEAN algorithm has reconstructed much more localized sources in the object function, which are consistent with the raw image, but it has not reconstructed the initial object function exactly. Högbom (1974) suggested that the CLEAN algorithm was suitable for reconstructing well-separated point sources, but the question remains as to how it chooses which object function to reconstruct from the range of possible solutions. The analysis by Tan (1986) shows that the choice is rather arbitrary, and depends on the nature of the object function and the loop gain in a complicated way.

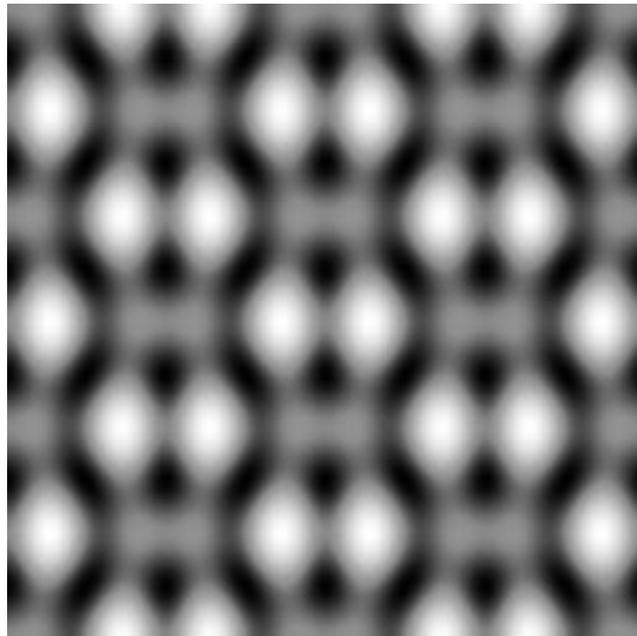


Fig. 7. An object function formed by deconvolving the PSF from the simulated image using a Wiener filter.

4.4. Maximum entropy reconstruction

Since the experimental data cannot uniquely identify the correct object function, the best we can hope to do is to try and determine a probability distribution for reconstructed object functions based on the experimental data. It has been discussed many times (see for example Sivia *et al.*, 1993) how Bayes' theorem can be used, which for this problem can be written

$$p(\text{object function} | \text{image data}) = \frac{p(\text{image data} | \text{object function})p(\text{object function})}{p(\text{image data})} \quad (11)$$

where $p(A|B)$ means the probability of A given B . Given the PSF we can compute $p(\text{image data} | \text{object function})$, which is the likelihood of getting the measured data for the trial object function, and $p(\text{image data})$ is independent of the object function. However, we do need to assign a so-called prior probability distribution, $p(\text{object function})$, to the range of object functions. A prior that was originally applied to radio astronomy (Gull & Daniell, 1978), but has since found many applications (see for examples Buck & Macaulay, 1991) is that of entropy

$$p(\text{object function}) \propto \exp(\alpha S) \quad (12)$$

where

$$S = - \sum_i f_i \log f_i, \quad (13)$$

is the entropy, f_i are the pixel values for the object function and α is a constant. A maximum entropy prior biases the

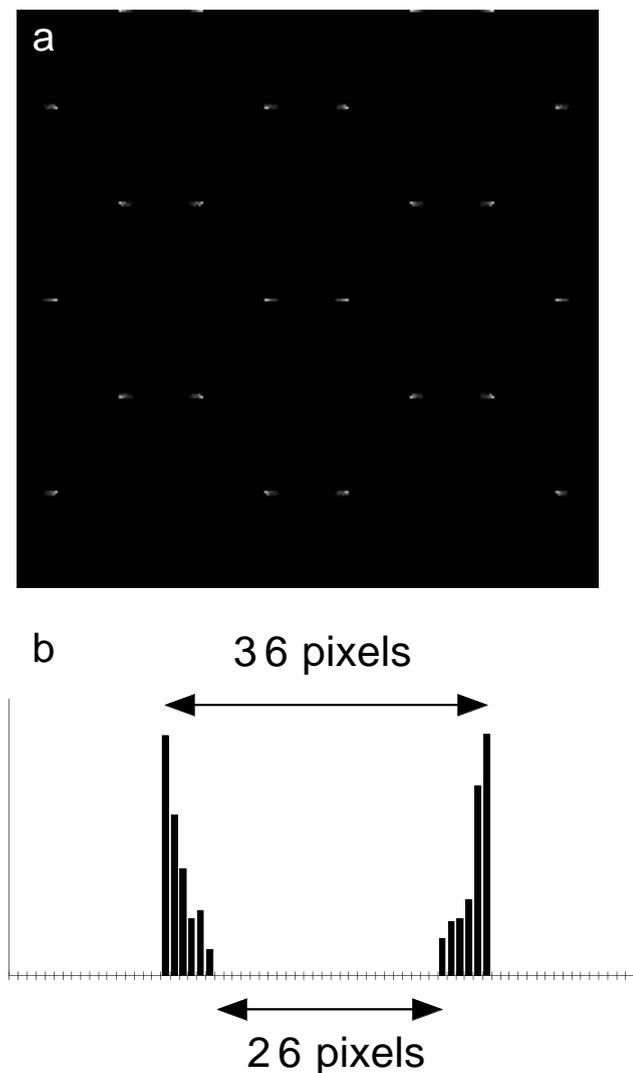


Fig. 8. (a) A reconstructed object function formed by applying the CLEAN algorithm to the simulated data. (b) A profile plot of pixel intensities across the central dumbbell pair.

reconstructed object function towards smoothness; the object function with the highest entropy is a uniform image. In this way an object function is reconstructed with the minimum structure that is consistent with the experimental image data.

Using the VG Microscopes MaxEnt software routine on a PC equipped with an accelerator card, which is an implementation of the Skilling & Bryan (1984) algorithm, we performed a reconstruction on the simulated experimental data described in Section 4.1 which resulted in the object function in Fig. 9(a). In this case the atom sites are indicated by only two neighbouring pixels being excited with one of them dominant, so once again the high-frequency part of the object function that was not transferred by the microscope has been reconstructed.

However, the distance between the dumbbell peaks in this object function is 36 pixels, as in the original data. The explanation for this effect is that our simulated data are artificially noise-free. The algorithm used assumes a Gaussian error in the experimental values from the counting statistics and therefore uses the value of χ^2 to compute $p(\text{image data} \mid \text{object function})$. As noted by Gull & Daniell (1978), the algorithm does not attempt to fit the data exactly, rather it attempts to let χ^2 approach a value based on an assumed noise scaling that can be automatically estimated from the experimental data, thereby avoiding the introduction of artefacts by the over-fitting of the data. Although this simulated noise-free data could be fitted exactly, the algorithm has assumed that it has been noise corrupted, avoided the perfect fit and failed to converge properly. To test this hypothesis, Gaussian distributed noise was added to the simulated data to give a signal-to-noise ratio of ≈ 3 . Although the reconstructed object function (Fig. 9b) is now not nearly as localized, the clusters of points are now centred on the correctly spaced column positions demonstrating that the algorithm is now making an estimate much closer to the true object function. In practice, ADF image data are relatively noisy and the maximum entropy algorithm is found to be an extremely powerful technique for ADF image analysis (McGibbon *et al.*, 1994, 1995). Some recent analysis (Dickey *et al.*, 1997) has suggested that the degree of relaxation at an interface can be determined to an accuracy approaching ± 0.01 nm using an ADF image combined with a maximum entropy reconstruction using the PSF reconstructed from the experimental data.

4.5. Parameterization

In Section 4.4 we assumed that there was no further information available apart from the experimental image and the PSF, but in practice we know that the object function consists of discrete, localized sources that can be assumed to be δ -function-like in the object function. Often the projected structure of the specimen can be determined intuitively from the raw image data (Chisholm *et al.*, 1994) since a focused probe gives minimal artefacts and the 0.13-nm resolution of the VG HB603U in ADF imaging resolves all the projected spacings for many specimens. If this is the case then all that is required is a simple procedure to match accurately the atomic-column positions to the experimental data using, for instance, a least-squares fit.

The process of starting with a model for the structure and then refining it to fit the experimental data is simply reducing the number of unknowns to a few parameters. In the case described above, we have reduced our problem to a few parameters of atomic-column location coordinates. The danger in this approach is of over-parameterization, in

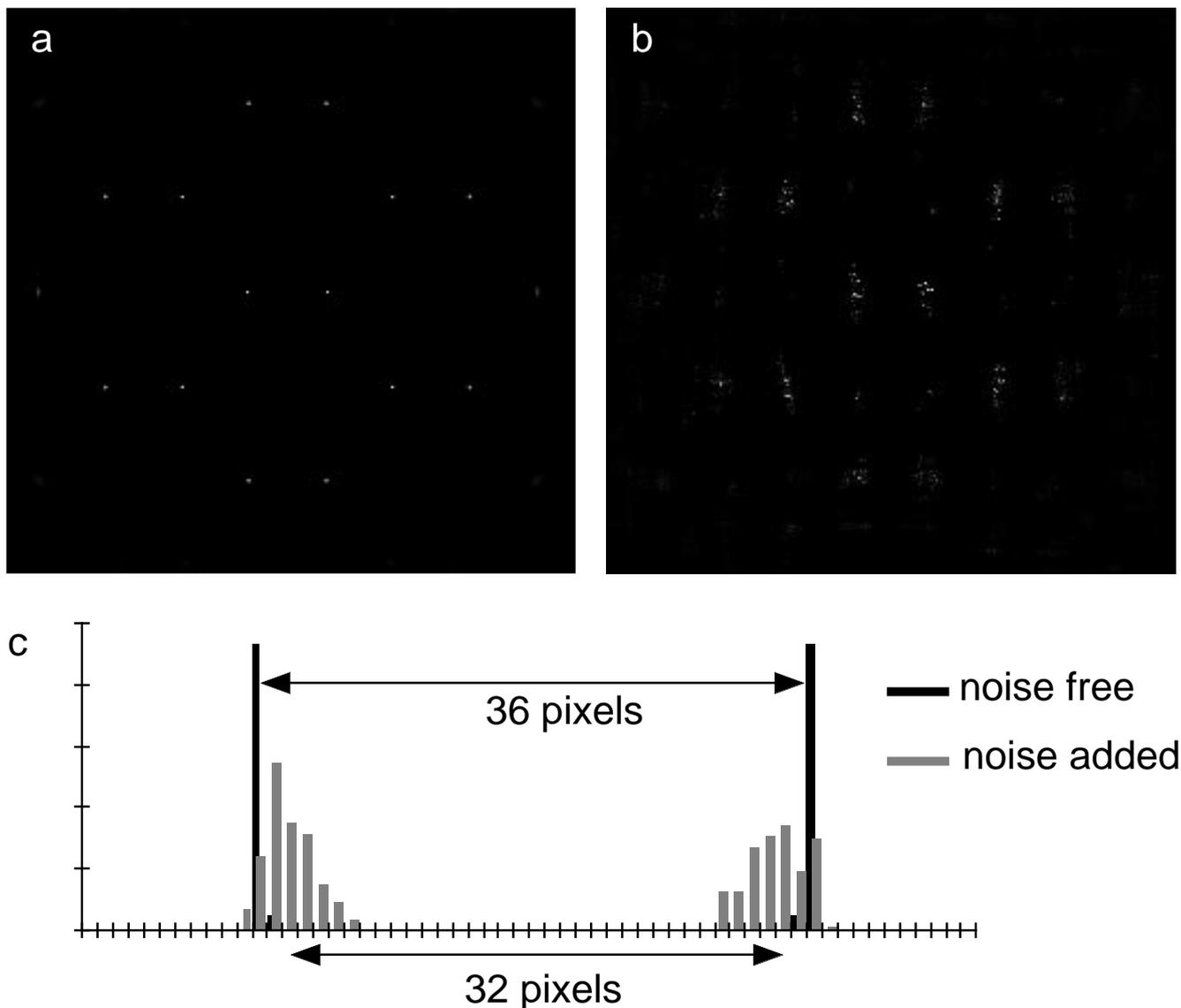


Fig. 9. Reconstructed object functions formed by a maximum entropy reconstruction of the object function: (a) with no noise added to the image, (b) with Gaussian noise added to the image data to give a signal-to-noise ratio of ≈ 3 . (c) Profile plots across the central dumbbell pair for both cases, the intensities having been summed perpendicular to the plot axis.

which the constraints imposed by the parameterization exclude the reconstruction of the correct object function which may be an unexpected structure. For example, we may try to fit one source to a region of intensity in an image, where in reality it was two sources that were unresolved by the microscope, for instance where a single atomic column has become split into two, half occupied, closely spaced columns. It has been shown once again (Sivia *et al.*, 1993) that Bayes' theorem offers a way of finding a probability distribution for the number of sources that are unresolved in a peak of intensity, and this approach may well be of use for ADF imaging.

5. Conclusions

Although we have been discussing quantitative approaches to image analysis in ADF imaging in this paper, one of the main strengths of incoherent imaging is that the structural information of the specimen is codified in the image in a much more straightforward way than it is for coherent HRTEM imaging, peaks of intensity in the image being directly related to the presence of an atomic column. In many cases, these structure images are all that is required to solve for the structure of the specimen. However, as in any resolution-limited system, there are ambiguities

remaining as to the correct object function, and peaks in the image intensity may well be displaced from the exact atomic-column locations, though in practice such displacements are of the order of 0.01 nm for ADF imaging.

A second important strength of incoherent imaging is that there is no phase problem, the image intensity being a convolution between the probe intensity and an object function consisting of very localized peaks. We have shown how the probe intensity profile can be directly determined from an image region of a specimen of known structure, and found that the VG HB603U at Oak Ridge is operating close to its theoretical diffraction-limited optimum performance, and resolving spacings as small as 0.128 nm. Even though the probe function can be determined, a simple deconvolution of it from the image data does not dramatically improve our ability to interpret the data because it does not attempt to reconstruct the information that was lost by the limited resolution. Bayes' theorem shows how some prior probability distribution of object functions is required to perform this reconstruction. The CLEAN algorithm and a maximum entropy reconstruction are both found to reconstruct object functions with highly localized sources, but the CLEAN algorithm is subject to the criticism in that the prior basis to the reconstruction cannot be adequately characterized.

The extreme case is to take the approach used in accurate atomic-column location in HRTEM: to use a starting model and then allow the columns to move to fit the experimental data in some iterative fitting routine. In ADF imaging this is a lot simpler than for HRTEM since a full dynamical calculation would not be required for each iteration. In this way we have parameterized the object function to a few numbers describing the coordinates of the columns. The constraints that this approach imposes may be dangerous because they preclude the determination of an unexpected structure. Perhaps the best compromise approach is to include the prior information that the sources are discrete and localized to single pixels, but to use a Bayesian approach to determine a probability distribution for their number and position in the object function.

Acknowledgments

We thank Dr R. Vincent for bringing their attention to the CLEAN algorithm. This work was partially supported by Oak Ridge National Laboratory, managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract number DE-AC05-96OR22464, and through an appointment to the Oak Ridge National Laboratory Postdoctoral Program administered by ORISE. Additional support was received from Magdalene College, Cambridge.

References

- Bates, R.H.T. & McDonnell, M.J. (1986) *Image Restoration and Reconstruction*. Oxford Engineering Science Series, Oxford.
- Baumann, F.H., Griebelyuk, M., Kim, Y., Kisielowski, C., Maurice, J.L., Rau, W.D., Rentschler, J.A., Schwander, P. & Ourmazd, A. (1995) High-resolution composition profiles of multilayers. *Phys. Status Solidi A*, **150**, 31–50.
- Bourret, A., Rouviere, J.L. & Penisson, J.M. (1988) Structure determination of planar defects in crystals of germanium and molybdenum by HREM. *Acta Crystallogr. A*, **44**, 838–847.
- Buck, B. & Macaulay, V.A. (1991) *Maximum Entropy in Action*. Oxford University Press, Oxford.
- Chisholm, M.F., Pennycook, S.J., Jevasinski, R. & Mantl, S. (1994) New interface structure for A-type CoSi₂/Si (111). *Appl. Phys. Lett.* **64**, 2409–2411.
- Coene, W., Janssen, G., Op de Beeck, M. & Van Dyck, D. (1992) Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscopy. *Phys. Rev. Lett.* **69**, 3743–3746.
- Cowley, J.M. (1969) Image contrast in a transmission scanning electron microscope. *Appl. Phys. Lett.* **15**, 58–59.
- Dickey, E.C., Dravid, V.P., Nellist, P.D., Wallis, D.J. & Pennycook, S.J. (1997) *Acta Mater.*, in press.
- Gull, S.F. & Daniell, G.J. (1978) Image reconstruction from incomplete and noisy data. *Nature*, **272**, 686–690.
- Högbom, J.A. (1974) Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron. Astrophys. Suppl.* **15**, 417–426.
- Jesson, D.E. & Pennycook, S.J. (1993) Incoherent imaging of thin specimens using coherently scattered electrons. *Proc. R. Soc. (London) A*, **441**, 261–281.
- Jesson, D.E. & Pennycook, S.J. (1995) Incoherent imaging of crystals using thermally scattered electrons. *Proc. R. Soc. (London) A*, **449**, 273–293.
- McGibbon, M.M., Browning, N.D., Chisholm, M.F., McGibbon, A.J., Pennycook, S.J., Ravikumar, V. & Dravid, V.P. (1994) Direct determination of grain boundary atomic structure in SrTiO₃. *Science*, **266**, 102–104.
- McGibbon, A.J., Pennycook, S.J. & Angelo, J.E. (1995) Direct observation of dislocation core structures in CdTe/GaAs (001). *Science*, **269**, 519–521.
- Möbus, G. (1996) Retrieval of crystal defect structures from HREM images by simulated evolution. I. Basic technique. *Ultramicroscopy*, **65**, 205–216.
- Möbus, G. & Dehm, G. (1996) Retrieval of crystal defect structures from HREM images by simulated evolution. II. Experimental image evaluation. *Ultramicroscopy*, **65**, 217–228.
- Nellist, P.D. & Pennycook, S.J. (1996) Direct Imaging of the atomic configuration of ultradispersed catalysts. *Science*, **274**, 413–415.
- Orchowski, A., Rau, W.D. & Lichte, H. (1995) Electron holography surmounts resolution limit of electron microscopy. *Phys. Rev. Lett.* **74**, 399–401.
- Pennycook, S.J. & Jesson, D.E. (1990) High-resolution incoherent imaging of crystals. *Phys. Rev. Lett.* **64**, 938–941.
- Lord Rayleigh (1896) On the theory of optical images with special reference to the microscope. *Philos. Mag.* **42**(5), 167–195.
- Saxton, W.O. & Smith, D.J. (1985) The determination of atomic

- positions in high-resolution electron micrographs. *Ultramicroscopy*, **18**, 39–48.
- Sivia, D.S., David, W.I.F., Knight, K.S. & Gull, S.F. (1993) An introduction to Bayesian model selection. *Physica D*, **66**, 234–242.
- Skilling, J. & Bryan, R.K. (1984) Maximum entropy image reconstruction: general algorithm. *Monthly Notices R. Astronom. Soc.* **211**, 111–124.
- Spence, J.C.H. & Cowley, J.M. (1978) Lattice imaging in STEM. *Optik*, **50**, 129–142.
- Spence, J.C.H., O’Keefe, M.A. & Kolar, H. (1977) High resolution image interpretation in crystalline germanium. *Optik*, **49**, 307–323.
- Tan, S.M. (1986) An analysis of the properties of CLEAN and smoothness stabilised CLEAN – some warnings. *Monthly Notices R. Astronom. Soc.* **220**, 971–1001.
- Zeitler, E. & Thomson, M.G.R. (1970) Scanning transmission electron microscopy. *Optik*, **31**, 258–280, 359–366.