Acta Cryst. (1998). **A54**, 49-60

Electron Ptychography. I. Experimental Demonstration Beyond the Conventional Resolution Limits

P. D. NELLIST* AND J. M. RODENBURG

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, England. E-mail: pdn10@cam.ac.uk

(Received 24 March 1997; accepted 23 July 1997)

Abstract

A solution to the phase problem of electron diffraction is described which allows an aberration-free atomic resolution image of Si⟨110⟩, showing the expected dumbbell contrast, to be reconstructed at a resolution of 3 times the point resolution and 2.5 times the information limit of the scanning transmission electron microscope (STEM) used. The data set required consists of coherent microdiffraction patterns recorded as a function of illuminating probe position and the method of image reconstruction beyond the conventional resolution limits using this data set is described. Using the inherent redundancy in the experimental data set, the accuracy of the reconstructed image is examined and the experimental imperfections that affect it are identified. It is found that aperture charging, compounded by distortions in the detection system, are the major sources of error. As an additional application of this method of phase retrieval, the diffracted-beam phases of electrons that have lost energy by exciting a plasmon are compared with those of elastically scattered electrons in a specimen of graphite. Within the limits of this approach, it is found that there is no difference in the beam phases, supporting the view that electrons that have undergone multiple elastic and inelastic scattering dominate the plasmon-loss scattering at higher angles.

1. Introduction

Although the resolution of the transmission electron microscope is usually limited by the aberrations of the image-forming lens, here we show how a solution to the phase problem provides an opportunity to overcome this limit. In conventional weak-phase-object imaging, the point resolution is the highest periodicity that can be directly interpreted in terms of the specimen structure [see Spence (1988) for a review of resolution limits in TEM]. Beyond this limit, the information is corrupted and contrast reversals occur. Various techniques have been suggested in order to solve for the phase of the image-plane wavefunction, such as off-axis holography (Lichte, 1991; Orchowski, Rau & Lichte, 1995) and focal-series reconstruction (Van Dyck, Op de Beeck & Coene, 1993). Knowledge of the image phase information allows the lens aberrations to be deconvolved, thus extending the resolution. However, the partial coherence of the electron wave in the microscope provides a further limit, known as the information limit, determined by the inability of high spatial frequencies to express themselves at all upon the image contrast.

An alternative technique of phase retrieval has been proposed by Rodenburg & Bates (1992), which allows the lens aberrations to be deconvolved and uses a method of aperture synthesis to reconstruct the phase information for scattering beyond the conventional information limit. Optical configurations for such a method are discussed more fully in the following paper (Plamann & Rodenburg, 1998) (paper II). The data set used experimentally in this work is the intensity of coherent microdiffraction patterns recorded in a scanning transmission electron microscope (STEM) as a function of the illuminating probe position. By reciprocity (Cowley, 1969; Zeitler & Thomson, 1970), this data set is equivalent to the set of conventional TEM images collected as a function of the illuminating beam tilt. In the case of a weakly scattering object, application of the aperture synthesis method can separate aberration-free information up to twice the point resolution and a deconvolution of lens aberrations is not required (Rodenburg, McCallum & Nellist, 1993). The image information for a crystalline specimen can be similarly separated (McCallum & Rodenburg, 1993*a*). In this case, only a line scan of probe positions is required and, in principle, a resolution improvement of many times the conventional limits can be achieved.

For a crystalline specimen, the aperture synthesis method is similar to an approach proposed earlier by Hoppe (1969) and Hoppe & Strube (1969), by which a reduction in the illuminated area of the specimen causes the diffracted orders to overlap and coherently interfere, thus revealing their phase relationships; this approach has been given the name 'ptychography' (Hegerl & Hoppe, 1970). Such a configuration is realized in the STEM and use of the coherent diffraction information has been proposed as a source of resolution improvement (Cowley & Jap, 1976; Nathan, 1976; Spence, 1978). High-angle coherent diffraction information has been utilized experimentally to accurately located atom positions in Si⟨110⟩ by Konnert, D'Antonio, Cowley, Higgs & Ou (1989), but their approach was to match the experimental

© 1998 International Union of Crystallography
Printed in Great Britain – all rights reserved

patterns with forward calculations from a model structure. Here we describe the details of the direct ptychographic image reconstruction of Si⟨110⟩, which has already been reported briefly elsewhere (Nellist, McCallum & Rodenburg, 1995), and examine the accuracy of the reconstructed phase information. Since the recorded data set has higher dimensionality than the two-dimensional reconstructed image, it contains redundancy that can be used as a self-measure of the accuracy of the reconstruction. This property is used to detect any experimental imperfections, and here we clearly detect the effects of aperture blade charging. As an alternative application of this solution to the diffracted-beam phase problem, we also present data recorded from a specimen of graphite to compare the scattering of the zero energy-loss and plasmon-loss electrons. Paper II (Plamann & Rodenburg, 1998) considers how dynamical scattering and the breakdown of the two-dimensional specimen approximation affects the ptychographical reconstruction used here.

2. Aperture synthesis image of Si⟨110⟩

The microscope used to collect all the experimental data presented here was a VG Microscopes HB501 scanning transmission electron microscope (STEM). It has an accelerating voltage of 100 kV and an objective lens designed for mainly analytical purposes with a coefficient of spherical aberration, C_S, of 3.1 mm. Thus, the conventional phase-contrast point resolution is 0.42 nm and the information limit for axial bright-field imaging is about 0.33 nm. The microdiffraction patterns were energy filtered then detected using a YAG scintillator coupled *via* a fibre optic to a Photek ICCD 125 image-intensifying TV-rate charge-coupled-device (CCD) camera. Coherent microdiffraction patterns were collected using a 256 × 256 array of CCD pixels over a single line scan of 64 probe positions with a spacing of 0.1 nm. The beam was incident along the [110] direction of a 7 nm thick specimen of silicon. We can form the conventional incoherent convergent-beam electron diffraction (CBED) pattern (Fig. 1) by summing the data over all probe positions.

Consider a microdiffraction pattern in which the diffracted discs represented by the $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$ reciprocal-lattice vectors overlap (Fig. 2). If the illuminating convergent beam is sufficiently coherent (the exact condition will become apparent below), interference can occur in the overlap region. The form of the intensity in the overlap will therefore be dependent on the phase variation across each diffracted disc. Each diffracted beam will have a complex amplitude given by the corresponding Fourier component of the object function; the complex amplitude of the $\mathbf{G}$ beam we will label as $\Psi_{\mathbf{G}}$ and similarly for $\mathbf{G}+\mathbf{H}$. In addition, there will be a phase change across a diffracted disc due to the objective-lens aberration. These aberrations cause a phase shift of the illuminating convergent waves as a function of angle, which results in a phase variation across each of the diffracted discs. In the absence of astigmatism, the phase shift is given by

$$\chi(\mathbf{K}) = \pi z\lambda|\mathbf{K}|^2 + \tfrac{1}{2}\pi C_S\lambda^3|\mathbf{K}|^4, \quad (1)$$

where $\mathbf{K}$ is a reciprocal-space vector corresponding to the angle of the illuminating wave, z is the lens defocus, λ is the electron wavelength and C_S is the coefficient of spherical aberration. We must also include the probe-position information, which, being a shift in real space, is a linear phase term across each diffracted disc in reciprocal space. Thus, the intensity in the overlap region is

$$\begin{aligned}|M(\boldsymbol{\mu}', \boldsymbol{\rho})|^2 = |&\Psi_{\mathbf{G}} \exp\{i[\chi(\boldsymbol{\mu}' - \mathbf{G}) + 2\pi\boldsymbol{\rho}\cdot(\boldsymbol{\mu}' - \mathbf{G})]\} \\ &+ \Psi_{\mathbf{G}+\mathbf{H}} \exp\{i[\chi(\boldsymbol{\mu}' - \mathbf{G} - \mathbf{H}) \\ &+ 2\pi\boldsymbol{\rho}\cdot(\boldsymbol{\mu}' - \mathbf{G} - \mathbf{H})]\}|^2, \end{aligned} \quad (2)$$

where $\boldsymbol{\rho}$ is the real-space probe-position vector and $\boldsymbol{\mu}'$ is the position vector in the microdiffraction plane (in previous papers, this vector has been referred to as $\mathbf{r}'$, following the image-processing convention). Expanding (2) gives

$$\begin{aligned}|M(\boldsymbol{\mu}', \boldsymbol{\rho})|^2 = {}&|\Psi_{\mathbf{G}}|^2 + |\Psi_{\mathbf{G}+\mathbf{H}}|^2 \\ &+ 2|\Psi_{\mathbf{G}}\Psi_{\mathbf{G}+\mathbf{H}}| \cos\{(\alpha_{\mathbf{G}} - \alpha_{\mathbf{G}+\mathbf{H}}) \\ &+ [\chi(\boldsymbol{\mu} - \mathbf{G}) - \chi(\boldsymbol{\mu}' - \mathbf{G} - \mathbf{H})] \\ &+ 2\pi\boldsymbol{\rho}\cdot\mathbf{H}\}, \end{aligned} \quad (3)$$

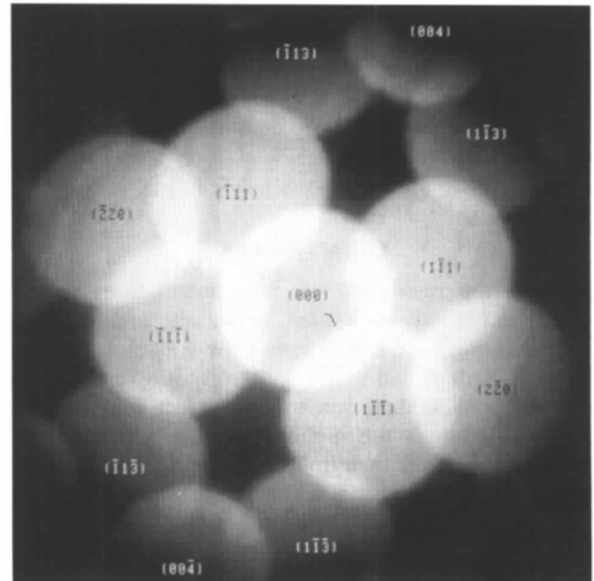


Fig. 1. An indexed CBED pattern from silicon oriented with the beam incident along the [110] direction. In practice, this figure has been formed by summing the coherent microdiffraction patterns for every probe position. The logarithm of the intensity is shown, to reduce the dynamic range of the figure.

where $\alpha_{\mathbf{G}}$ and $\alpha_{\mathbf{G+H}}$ are the phases of the beams $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$, respectively.

The first term of the cosine argument in (3) shows how the interference contains the phase difference between the interfering adjacent diffracted beams. Having determined the phase difference of all neighbouring beams in a diffraction pattern, it is a simple matter to determine the phases of all the recorded beams by a process of 'stepping out'. We start by assigning an arbitrary phase of 0° to the zero-order beam, which then allows us to determine the phases of any diffracted beams whose discs overlap with the zero-order beam. We can now use the phases of these diffracted beams together with their overlap with higher-angle diffracted beams to assign phases to these higher-angle beams and continue this procedure for all available overlap regions. The magnitude of the diffracted beams can be calculated by taking the square root of the intensity in non-overlapping parts of the diffracted discs. Finally, an image can be computed by taking the inverse Fourier transform of the magnitude and phase of the recorded diffracted beams.

Implicit here is the assumption that the phase of a diffracted beam is not dependent on the illumination tilt, so that different parts of a diffracted disc can be assigned with the same diffracted-beam phase. This assumption is valid for very thin specimens, where the effects of wavefunction propagation within the specimen are negligible. In the kinematical approximation, the beam phase will vary across a disc due to propagation effects. At the centre of a disc overlap region, however, the excitation errors for the two diffracted beams will be identical and in this case also there will be no phase error. The limits to thickness and methods of data interpretation for full dynamical theory are discussed extensively in paper II (Plamann & Rodenburg, 1998).

The second term in the cosine argument of (3) shows that the intensity in an overlap region can vary across the overlap due to the effect of lens aberrations. If we substitute using (1) and expand the defocus terms, then the second term of the cosine argument is

$$\begin{aligned}\chi(\boldsymbol{\mu}' - \mathbf{G}) - \chi(\boldsymbol{\mu}' - \mathbf{G} - \mathbf{H}) \\ = \pi z\lambda[2\boldsymbol{\mu}'\cdot\mathbf{H} - 2\mathbf{G}\cdot\mathbf{H} - |\mathbf{H}|^2] \\ + \tfrac{1}{2}\pi C_S\lambda^3[|\boldsymbol{\mu}' - \mathbf{G}|^4 - |\boldsymbol{\mu}' - \mathbf{G} - \mathbf{H}|^4]. \quad (4)\end{aligned}$$

The defocus terms is linear in $\boldsymbol{\mu}'$ and therefore defocus will cause the disc overlap regions of a coherent microdiffraction pattern to show straight fringes perpendicular to the relative reciprocal-lattice vector connecting the discs [see for example Vincent, Midgley & Spellward (1993) who also noted that astigmatism can rotate the direction of the fringes]. However, the spherical aberration term has a cubic dependence on $\boldsymbol{\mu}'$ and will distort the uniformity of the fringes. It is important to note that for any centrosymmetric aberration, such as spherical aberration or astigmatism, setting $\boldsymbol{\mu}' = \mathbf{G} + \mathbf{H}/2$ (which is the position of the centre of the overlap region) sets (4) to zero and therefore aberration-free information can be obtained. The physical reason for this is that the beams interfering at this point have passed through diametrically

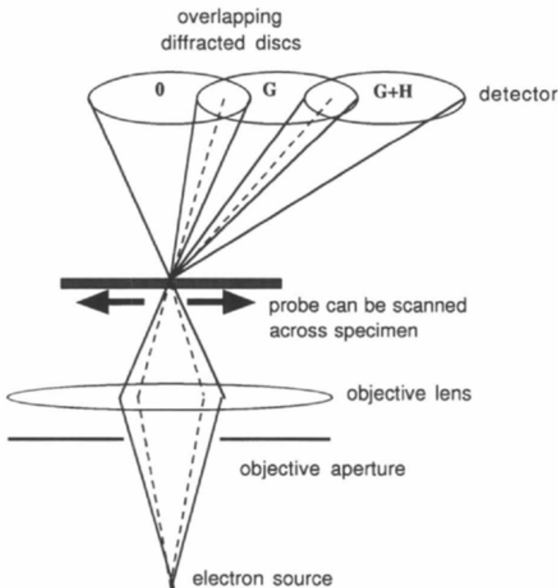


Fig. 2. The formation of overlapping discs in a scanning transmission electron microscope (STEM). A lens focuses a convergent beam of electrons to form a probe that can be scanned over a specimen. In the far field, a CBED pattern is formed and interference can occur in the disc-overlap regions. The dashed lines indicate paths of the electrons that interfere at the centres of the overlaps. They pass through diametrically opposite parts of the objective lens and therefore experience identical phase shifts due to the lens aberrations.

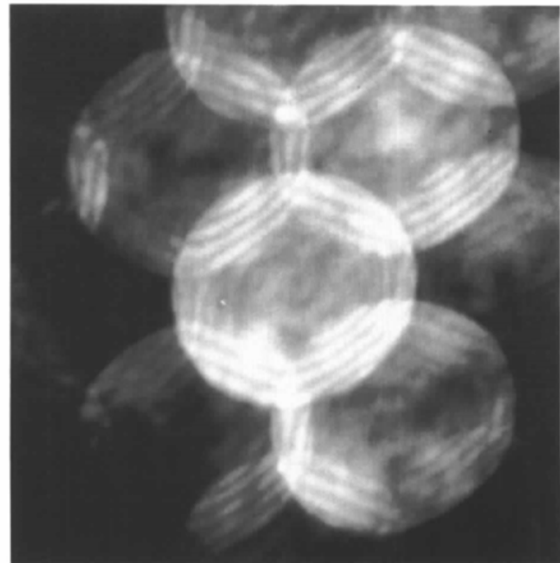


Fig. 3. An example of a heavily defocused coherent microdiffraction pattern from a relatively thick Si⟨110⟩ crystal. Note the interference fringes in the overlap regions.

opposite parts of the objective lens (Fig. 2) and the aberrations cancel. Fig. 3 is an example of a defocused coherent microdiffraction pattern from a relatively thick Si⟨110⟩ crystal. For ptychographical aperture synthesis, it is advantageous to work with a focused probe, which reduces the effects of detector incoherence (Nellist & Rodenburg, 1994) and improves the accuracy of the technique, as will be discussed later. Using a focused probe involves using underfocus to counteract the effects of spherical aberration, which results in (4) being a much flatter function of $\boldsymbol{\mu}'$.

The final term in the argument of the cosine function in (3) shows that the intensity in a disc-overlap region varies sinusoidally as a function of probe position. This effect is simply STEM lattice imaging (Spence & Cowley, 1978). Such an intensity variation is observed experimentally, as is shown in Fig. 4, where the intensity at points in various disc overlaps has been plotted. It is important to note that the spatial frequency of the intensity variation depends only on the difference of the two scattering vectors involved, not on the absolute angle of diffraction. Since there are only three different relative reciprocal-lattice vectors for the disc overlaps in Fig. 1, there are only three possible spatial frequencies of intensity variation. Equation (3) shows that the frequency of intensity variation along a line scan is given by the component of the relative reciprocal-lattice vector, $\mathbf{H}$, parallel to the scan direction. By selecting an appropriate scan direction, we can conveniently separate the three spatial frequencies. A full two-dimensional probe scan is not required. We proceed by taking the Fourier transform of the intensity data with respect to the probe position to form a data set we call $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ [see Plamann & Rodenburg (1998) for the full expression], where $\boldsymbol{\rho}'$ is the reciprocal-space vector conjugate to $\boldsymbol{\rho}$. The plane $\boldsymbol{\rho}' = \mathbf{0}$ of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ is simply the CBED pattern shown in Fig. 1. We also observe significant magnitude in $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ at three other frequencies, as shown in Fig. 5, correspond-

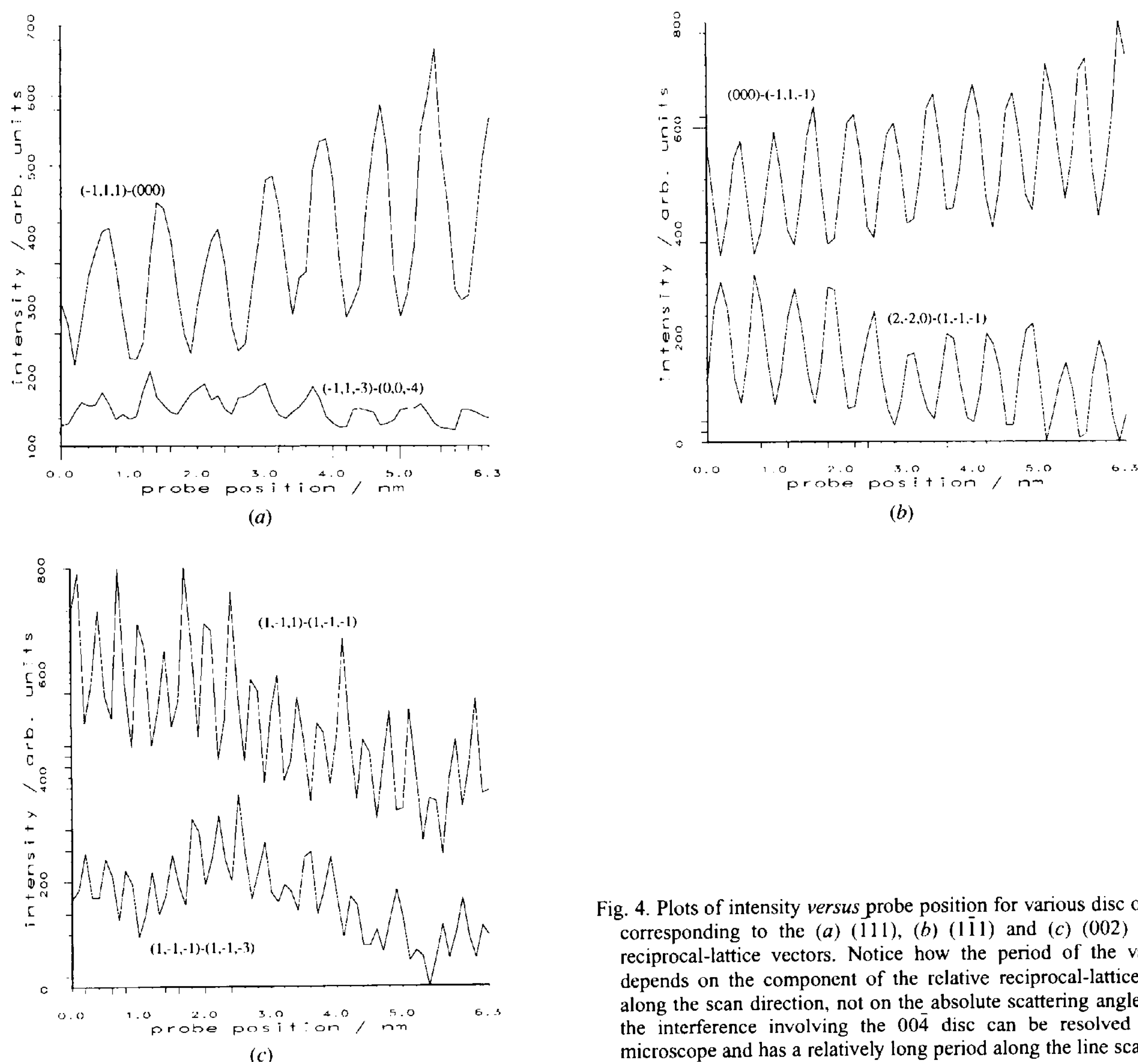


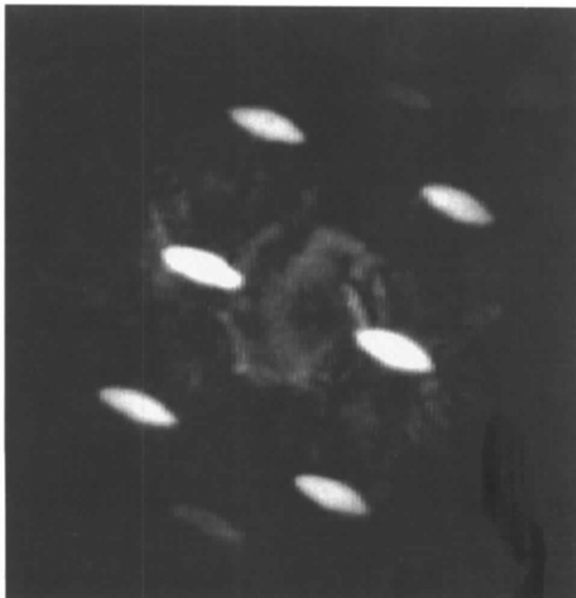
Fig. 4. Plots of intensity *versus* probe position for various disc overlaps corresponding to the (*a*) $(\bar{1}11)$, (*b*) $(1\bar{1}1)$ and (*c*) (002) relative reciprocal-lattice vectors. Notice how the period of the variation depends on the component of the relative reciprocal-lattice vector along the scan direction, not on the absolute scattering angle. Thus, the interference involving the $00\bar{4}$ disc can be resolved by the microscope and has a relatively long period along the line scan used.

ing to the relative reciprocal-lattice vectors for the three different types of overlap. All the overlaps of a particular type show interference at one frequency, independent of the absolute scattering angle. For the crystalline case considered here, the overlaps are all separated and (3) shows that the magnitude of a disc-overlap feature in $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ is given by the product of the magnitudes of the two diffracted beams and the phase of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ is given by the first two terms of the cosine argument in (3). Had we not moved the probe, there would have been an ambiguity of sign for the phase difference between the diffracted beams owing to the cosine being an even function. Moving the probe solves the sign ambiguity and allows us to separate the Fourier components of the image. We can therefore take the phase of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ at the centre of an overlap, which is the aberration-free point, to be the phase difference of the interfering beams and use these for the stepping-out procedure.

By thinking of the intensity variation in an overlap as lattice imaging, it can now be seen that the microscope only requires the stability and coherence required to observe the fringes corresponding to the largest relative reciprocal-lattice vector, $\mathbf{H}$, which in this experiment corresponds to a spacing of 0.27 nm. There is no

(*a*)

(*b*)

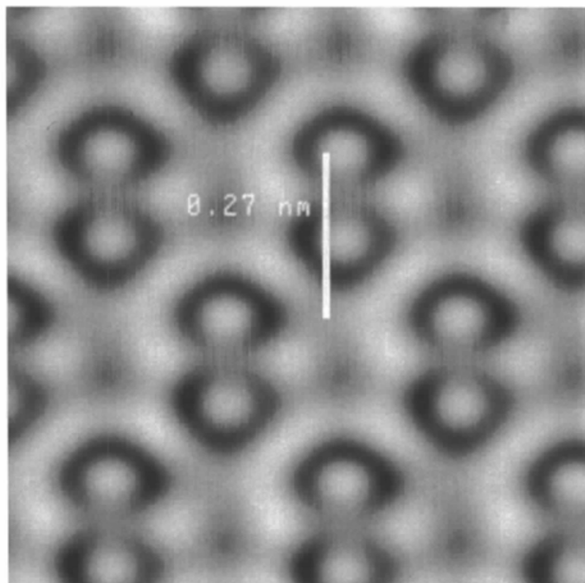


(*c*)

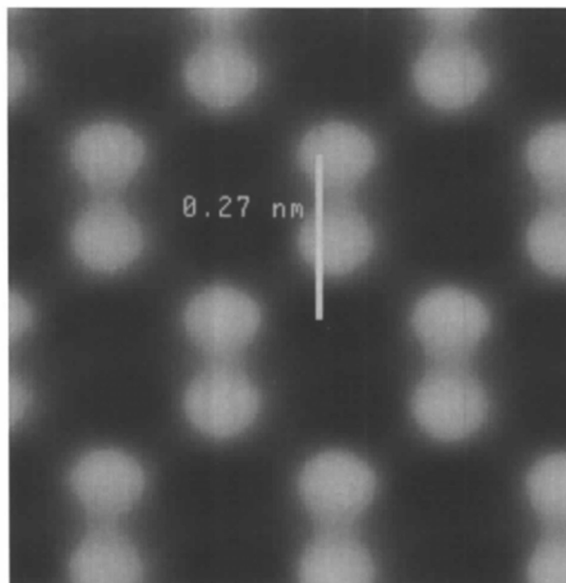
Fig. 5. A map of the magnitude of the Fourier transform, with respect to the probe-position vector, of the recorded intensities at each detector pixel for the spatial frequencies corresponding to the component of the (*a*) $(\bar{1}11)$, (*b*) $(1\bar{1}1)$ and (*c*) (002) reciprocal-lattice vectors along the probe line-scan direction. This figure therefore indicates where the detected intensity varies at the quoted frequency as the probe is scanned. Note how the intensities in all the overlap regions corresponding to the same relative reciprocal-lattice vector vary at the same spatial frequency.

coherence limit on the absolute scattering angle that can be used, thus it is the stepping-out procedure that allows image reconstruction as a resolution well beyond the information limit. Indeed, Hoppe (1982) realized that the conventional resolution limits did not apply to ptychography.

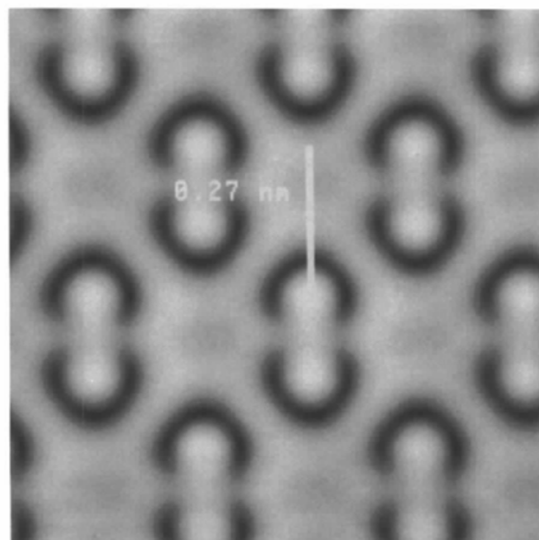
Although we have three relative reciprocal-lattice vectors available in this experiment, we only need two to span the two-dimensional diffraction space. For the data presented here, we shall see later that aperture charging results in an asymmetric χ function, which prevents the use of all three vectors, so stepping out was performed using the $(\bar{1}11)$ and the (002) relative reciprocal-lattice vectors only. The $\bar{1}11$ and $1\bar{1}\bar{1}$ beams could therefore be phased directly from the 000 beam, and these two beams were used to phase the $\bar{1}1\bar{1}$ and $1\bar{1}1$ beams using the (002) relative reciprocal-lattice vector overlaps. From these $\{111\}$ beams, the (002)-type overlaps were again used to phase the $\{113\}$ beams, which could then be used to phase the $\{004\}$ beams using the $(\bar{1}11)$ relative reciprocal-lattice vector overlaps. The phases of diffracted beams out as far as the $\{004\}$



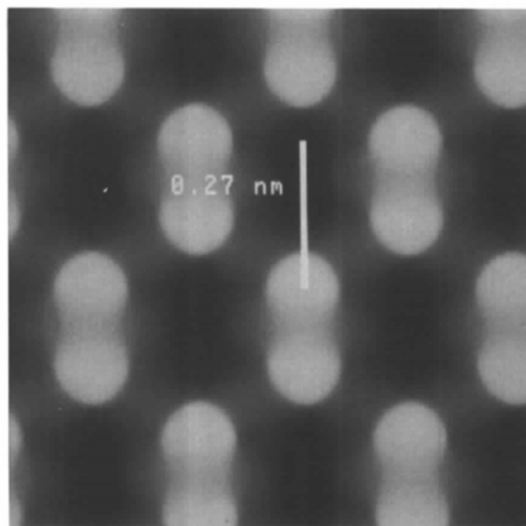
(*a*)



(*b*)



(*c*)



(*d*)

Fig. 6. Reconstructed images for the (*a*), (*b*) 4.5 nm thick specimen and (*c*), (*d*) 7 nm thick specimen. The magnitudes are shown in (*a*) and (*c*): the grey scale spans the range 91 (black) to 239 (white) arbitrary units for (*a*) and 24 (black) to 317 (white) arbitrary units for (*b*). The phases are shown in (*b*) and (*d*); the grey scale spans the range −0.67 to +1.81 rad for (*b*) and −1.00 to +2.85 rad for (*d*).

Table 1. *The reconstructed phases of all beams up to 004 for approximately 4.5 nm and 7 nm thick specimens of Si⟨110⟩*

The theoretical phases using a multislice calculation are also presented for various thicknesses. For the experimental results, the arbitrary origin of real space has been set by minimizing the phase differences between the four {111}-type symmetry-related beams.

Beam	Experimental phase for 4.5 nm (°)	Experimental phase for 7 nm (°)	Calculated phase for 3 nm (°)	Calculated phase for 6 nm (°)	Calculated phase for 9 nm (°)
$1\bar{1}1$	+112	+138	+109	+127	+156
$1\bar{1}\bar{1}$	+120	+138	+109	+127	+156
$\bar{1}11$	+120	+137	+109	+127	+156
$\bar{1}1\bar{1}$	+128	+138	+109	+127	+156
$2\bar{2}0$	+128	−169	+124	+153	−166
$\bar{2}20$	+171	+171	+124	+153	−166
$1\bar{1}3$	−49	+6	−53	−23	+13
$1\bar{1}\bar{3}$	−51	−1	−53	−23	+13
$\bar{1}13$	−59	−9	−53	−23	+13
$\bar{1}1\bar{3}$	−26	−3	−53	−23	+13
004	0	+50	−43	−1	+53
$00\bar{4}$	−17	+38	−43	−1	+53

reflections were determined. Note that, at the thicknesses used, the kinematically forbidden {002} and {222} beams were not present in any significance, they do not show strong interference features in Fig. 5 and therefore were not included in the reconstruction. The magnitudes of the diffracted beams were determined from the non-overlapping parts of the discs in the CBED pattern (Fig. 1). In general, however, a reconstructed image is usually much more sensitive to the phases of its Fourier components rather that the magnitudes (Ramachandran & Srinivasen, 1970). The reconstructed image for the approximately 7 nm thick crystal is shown in Fig. 6, along with a similar reconstruction for a crystal approximately 4.5 nm thick. These images are aberration free and contain periodicities as small as 0.136 nm, thus resolving the atomic columns. An improvement of a factor of 3 over the point resolution and a factor of 2.5 over the conventional information limit has been achieved.

It can be seen from Fig. 6 that the distance between the phase maxima for a dumbbell pair is slightly larger than the 0.136 nm expected. The sharp truncation of reciprocal space beyond the 004 reflections results in the atomic locations being convolved with a point spread function (PSF). Addition of the PSF's for the adjacent atoms has the effect of apparently widening the spacing slightly. However, this image is perfectly periodic and the quality of the reconstruction is therefore more easily considered in reciprocal space.

3. The accuracy of the reconstructed phases

We have shown how it is possible to solve the phase problem and reconstruct images beyond the information limit using aperture synthesis. However, at this point we should examine how accurately we have determined the phases in the experimental data. To determine the quality of the reconstruction, we compare the experimentally determined phases with those expected theoretically. Table 1 shows the reconstructed phases for the two specimen thicknesses. We have arbitrarily shifted the origin of real space to a centre of symmetry in the crystal so that the experimental phases of the Fourier components can be compared with theory. The position of the centre of symmetry was chosen by minimizing the differences between the four $\{1\bar{1}1\}$ beams. Any errors in the determined phases of these beams, and therefore in determining the centre of symmetry, will manifest themselves as errors in the phases of the higher-order beams. We can immediately see that the phases have deviated considerably from the ±90° expected from kinematical theory. So, in Table 1, we also present the phases determined from dynamical calculations (Plamann & Rodenburg, 1998) for various thicknesses. Propagation effects will in general cause a further variation in phase over a diffracted disc, and thus the phases determined here by stepping out also deviate from the Fourier components of the exit-surface wavefunction for axial illumination.

Although the phases in Table 1 are generally within the range defined by the calculations for thicknesses closest to the experimental ones, it can be seen that there is a degree of variation in the determined phases for symmetry-related beams, especially for the thinner specimen. There are many possible reasons for this and we will examine some here.

It is possible that there are multiple paths of stepping out to reach any given reciprocal-lattice point. Both paths must give the same result since the phase of the diffracted beam must be single valued. Equivalently, the phase difference round any closed path of relative reciprocal-lattice vectors must sum to zero, which we call 'phase closure' and is a useful check on the self-consistency of the experimental data. The results of some phase-closure

triangle and parallelogram loops are shown in Fig. 7. Notice how the triangle phase closures are about as far from 0° as it is possible to be, whereas the parallelogram phase closures are much closer to 0°. This result is consistent with the phase of the aperture function, χ, not being centrosymmetric. An asymmetric aperture-function phase will cause the second term of the cosine argument in (3) to be non-zero at the centre of the overlap and its value will be a function of the relative reciprocal-lattice vector. Thus, for each different relative reciprocal-lattice vector used, the phase difference measured will include a phase offset. In the case of a parallelogram closure, only two different relative reciprocal-lattice vectors are used, both of which are used in both directions, and therefore the phase offsets will cancel. The triangle loops use three different reciprocal-lattice vectors and will therefore not close in the case of an asymmetric aperture phase function. Thus, only two relative reciprocal-lattice vectors were used for the reconstructions in §2, since using all three would have given rise to inconsistencies. The phase offsets for the two relative reciprocal-lattice vectors used add a linear phase ramp across the complex reconstructed Fourier spectrum, which is simply a shift in the origin of the reconstructed image.

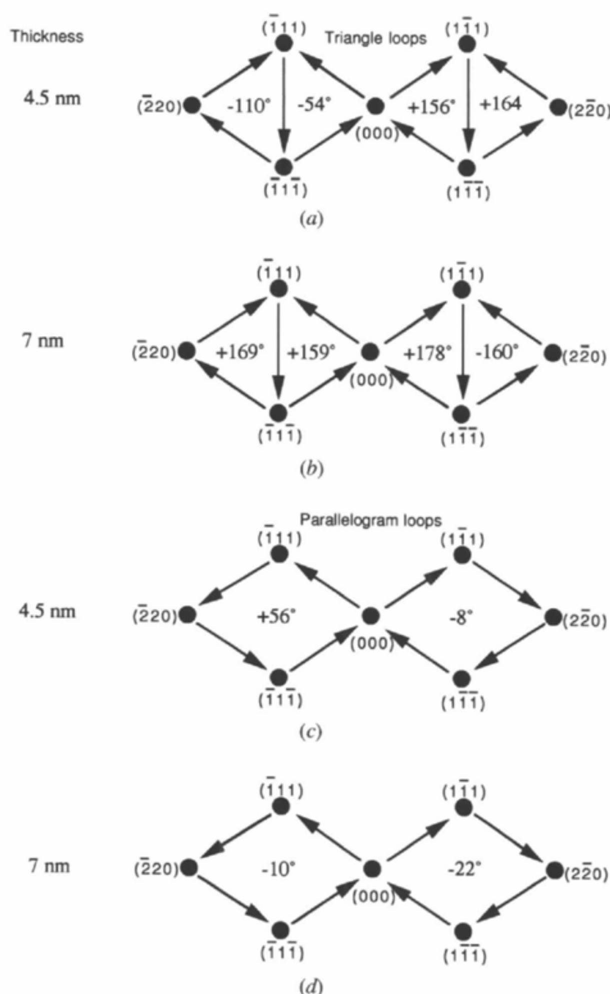


Fig. 7. The sum of beam phase differences around (*a*), (*b*) triangle and (*c*), (*d*) parallelogram closed loops for the data from the (*a*), (*c*) 4.5 nm and (*b*), (*d*) 7 nm thick specimens. Ideally, the sum round a closed loop should be zero. Notice how the parallelogram loops are much closer to zero than the triangles.

The asymmetry in the phase of the aperture function, χ, could be caused by objective aperture misalignment or by contamination on the aperture charging and causing a phase shift to the electron waves in addition to the lens aberrations. However, we can use the phase of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ to examine the phase variation over an overlap region. Fig. 8 shows the phase of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ for the overlap regions corresponding to the $(1\bar{1}1)$ relative reciprocal-lattice vector, and indeed its distorted shape, which cannot be explained by aperture misalignment, indicates that aperture charging is occurring.

High-resolution electron microscopy is often performed without a physical objective aperture being present. Conditions of partial coherence are relied upon to destroy the interference with the high-angle aberrated beams. Nellist & Rodenburg (1994) have shown that conditions of partial coherence imposed by the finite size of the detector pixels lead to the existence of an 'effective aperture' in aperture synthesis. A physical aperture is not required to separate the interference between different

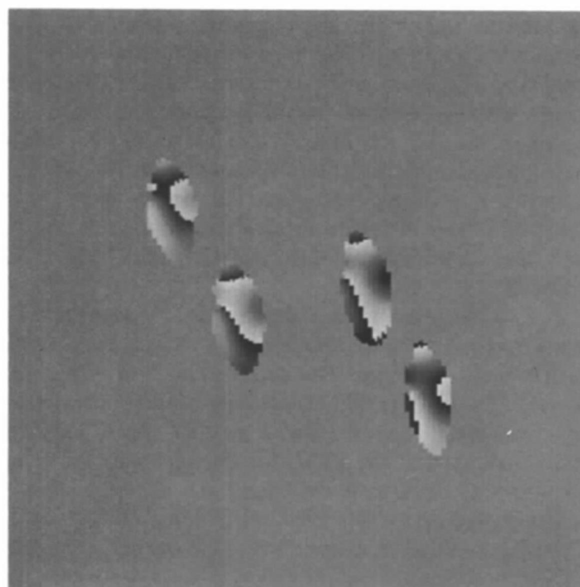


Fig. 8. The phase of the Fourier transform, with respect to the probe position, of the recorded intensities at the spatial frequency corresponding to the component of the $(1\bar{1}1)$ reciprocal-lattice vector along the probe line-scan direction. This is the phase corresponding to the magnitude in Fig. 5(*b*). For clarity, we have used Fig. 5(*b*) as a mask so that only the phases in the overlap regions are shown. Note how the phase variation is distorted towards a point on the left of each overlap region; this indicates that the aperture blade is charging at this point.

beams. In practice, it is best to use a physical aperture that is significantly larger than the coherence envelope, such that it does not interfere with the imaging process but does prevent too many high-angle convergent beams from contributing to the incoherent background. In the experiment described here, the effective aperture has a radius of approximately 11 mrad, so the experiment was repeated with a physical objective aperture with a radius of about 16 mrad. The CBED pattern, formed in a similar way to Fig. 1, is shown in Fig. 9 and shows that there is now multiple overlap between the diffracted discs. Determination of the magnitudes of the diffracted beams now becomes more difficult, but still may be possible. Here, we will limit ourselves to the accuracy of the phases. A relevant plane of the complex $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ function is shown in Fig. 10, and indicates how the effective aperture has separated all the regions of interference. The phases of the interference regions in Fig. 10 are all more symmetric and flatter than Fig. 8. Phase-closure loops for these data are shown in Fig. 11 and show a dramatic improvement in the triangle phase-closure loops, demonstrating a big improvement in the symmetry of the aperture-function phase.

There is still some deviation from zero in the phase-closure loops shown in Fig. 11 and we should consider other factors that can affect the phase accuracy. One experimental difficulty encountered was the selection of the midpoint of the overlap region. It is important that the midpoint is used in all overlaps to avoid the effects of the lens aberrations. Even if the phase of the aperture function is asymmetric, we must still use the same point in all of the overlaps corresponding to a particular relative reciprocal-lattice vector so that the same phase offset always applies. The electron-optical configuration of the spectrometer used to energy filter the images gives rise to distortions in the recorded microdiffraction patterns. The distortions are complicated and, as yet, we have not been able to correct for them, so the overlap midpoint positions were determined by hand. For a well focused probe, the phase of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ in the overlap regions should be reasonably flat, especially at the centre of the overlap region. However, aperture charging, defocus and astigmatism can all conspire to increase the phase gradient within an overlap region, as is

Fig. 9. A CBED pattern from Si⟨110⟩ formed using a convergence semi-angle of approximately 16 mrad. This figure has been formed by summing the coherent microdiffraction patterns for every probe position collected. The logarithm of the intensity is shown, to reduce the dynamic range of the figure.

(004)
(Ī13)
(1Ī3)
(Ī11)
(2̄20)
(1Ī1)
(000)
(ĪĪĪ)
(22̄0)
(1ĪĪ)
(Ī1̄3̄)
(11̄3̄)
(004̄)

(*a*)

(*b*)

Fig. 10. The (*a*) magnitude and (*b*) phase of the Fourier transform, with respect to the probe position, of the recorded intensities at the spatial frequency corresponding to the component of the $(1\bar{1}1)$ reciprocal-lattice vector along the probe line-scan direction. These figures are the large-aperture equivalents of Figs. 5(*b*) and 8. Although the data set was collected with a beam convergence semi-angle of approximately 16 mrad, giving multiple disc overlap, the effective aperture still separates the various interference regions.

demonstrated by aperture charging in Fig. 8. For the data presented in §2, an error of one detector pixel in selecting the overlap region midpoint of the (002) relative reciprocal-lattice-vector regions can lead to a change of up to 29° in the determined phase. However, in this case, the point is close to the aperture edges and therefore the phase gradient is high due to aperture charging. For the data collected using an effective aperture to separate the interference regions, the error reduces to about 10° for the (002) relative reciprocal-lattice vector.

A further source of error is that not all of the energy expressed in $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ at a $\boldsymbol{\rho}'$ value corresponding to a crystal periodicity actually comes from diffraction by that periodicity. There is always an amorphous oxide layer that, being non-periodic, will scatter into all spatial frequencies. Similarly, it is observed that the beam current fluctuates in a random manner and will also add spurious energy at all $\boldsymbol{\rho}'$ values. Finally, variations in specimen thickness will cause a non-periodic variation in the diffracted-beam intensities and therefore in the magnitude of the interference effects. The effects of all this non-periodic energy can be seen in the curves in Fig. 4, which are not perfectly sinusoidal, and in the observation of magnitude in the non-overlapping parts of the discs in Fig. 5. This energy will have a phase in $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ that is not related to the crystal structure and thus its addition to the periodic interference in the overlap regions will cause error. It must be remembered, however, that the approach taken here is a simplification of a method that can deal with a general non-periodic specimen (Rodenburg & Bates, 1992). In principle, we can perform a deconvolution in the $\boldsymbol{\mu}'$ direction to solve for the entire specimen function, including the non-periodic parts such as the amorphous oxide layer and the thickness variations.

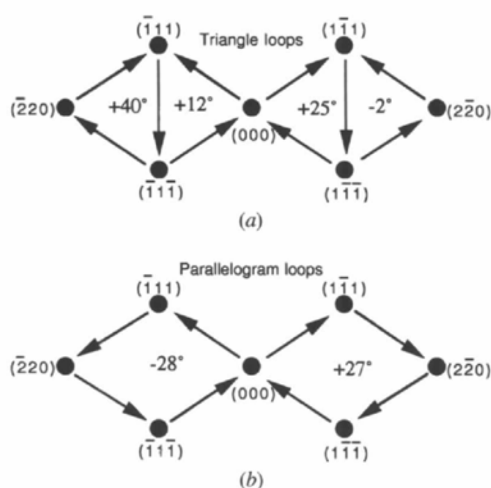


Fig. 11. The sum of beam phase differences around (*a*) triangle and (*b*) parallelogram closed loops for the data collected using a beam-convergence semi-angle of approximately 16 mrad. Note how the triangle closures are much improved compared with Fig. 7.

4. Aperture synthesis using inelastically scattered electrons

For the experiments described above, a spectrometer was used to ensure that only the zero-energy-loss electrons were detected. If desired, this experimental arrangement can also be used to collect electrons at a specific energy loss. The coherent microdiffraction patterns and shadow images collected using electrons that have lost energy by exciting a plasmon also display coherent interference effects. Here, we present the results of an aperture-synthesis experiment performed using the plasmon-loss electrons scattered by a graphite test specimen.

The energy-filtering spectrometer was set to pass electrons that had lost energy by exciting a single plasmon in a graphite, with a filter window of approximately 5 eV centred at an energy loss of ~25 eV. Shadow images (with no objective aperture present) were collected over 32 probe positions at a probe spacing of 0.05 nm. A typical shadow image is shown in Fig. 12. As expected, significant magnitude is observed in $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ at the $\boldsymbol{\rho}'$ value corresponding to the 0.34 nm lattice spacing of graphite. The magnitude and phase of this plane of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ are show in Fig. 13, which shows that the effective aperture has separated the interference regions. For comparison, the experiment was repeated using the zero-loss electrons. Two data sets were collected for each energy whilst trying to maintain the probe scan over the same region of specimen. The phases extracted from the centres of the overlap in $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ for all four experiments are shown in Table 2. The values shown are therefore the phase differences between adjacent beams for the five diffracted beams detected.

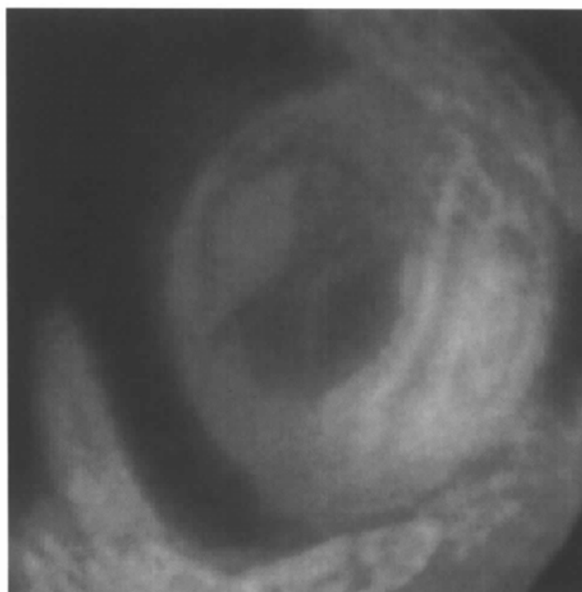


Fig. 12. A coherent shadow image from graphite formed using electrons that have lost energy by exciting a plasmon. Note that coherent interference regions can be observed.

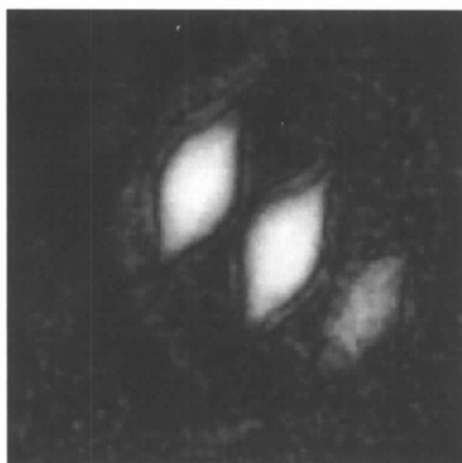
It can be observed that even for one energy there is a significant variation in the measured phases, which may be due to the scan sampling a slightly different region of specimen with a slightly different tilt and thickness. Since the phase offset was calculated by setting the first-order beam phases to be identical, any asymmetry in the diffracted-beam phases will be added to the second-order phases. In addition, the second of the zero-loss experiments did show some crystal tilt and the $00\bar{2}$–$00\bar{4}$ overlap region was found to be dominated by the non-crystalline background described earlier.

The sample of graphite used for these experiments was relatively thick and may well have contained strongly bent planes, hence the large deviation of the determined phases from the ±90° expected for single scattering theory. The interesting result from the data, however, is that, within the limits of the experimental error, there is no significant difference between the phases measured using the plasmon-loss electrons and the phases measured using the zero-loss electrons. This result is consistent with the idea that the scattering is due to a plasmon scattering event combined with at least one elastic scattering event. Indeed, phase-contrast images of graphite have already been observed by detecting the plasmon-loss electrons with an axial collector aperture in a STEM, and this scattering mechanism was proposed by Craven & Colliex (1977). In that work, it was noted that the range of defocus over which fringes could be observed was reduced in the plasmon-loss case, which was explained by the slight angular spread of the plasmon scattering leading to an increased effective collection angle. In the present work, the effect is to increase the effective detector pixel size slightly, leading to worsened spatial coherence (Nellist & Rodenburg, 1994), possibly compounded by the dispersion of the spectrometer over the energy window used. In principle, it should be possible to measure the angular spread by observing the visibility of the interference fringes in the shadow image as a function of the illuminating probe defocus. As the defocus is increased, the observed fringes become finer and are therefore more susceptible to being washed out by the angular spread of the plasmon.

Table 2. *The phase of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ (°) at the centres of the overlaps, at $|\boldsymbol{\rho}'| = (0.34\ \text{nm})^{-1}$, for the two zero-loss and two plasmon-loss experiments*

The errors shown are the largest deviation in phase found in the pixels adjacent to the one selected. The phase offset due to the arbitrary origin of the scan has been calculated by assuming that the (000)–($00\bar{2}$) and (002)–(000) phases are equal and opposite, which is equivalent to assuming that the two first-order diffracted beams have equal phase.

Beams	$00\bar{2}$–$00\bar{4}$	000–$00\bar{2}$	002–000	004–002
Plasmon exp. 1	50 (6)	102 (2)	−102 (2)	−49 (17)
Plasmon exp. 2	45 (9)	103 (3)	−103 (3)	−48 (7)
Zero loss exp. 1	46 (6)	98 (4)	−98 (3)	−37 (3)
Zero loss exp. 2	−6 (21)	107 (3)	−107 (7)	−42 (3)



(*a*)

(*b*)

Fig. 13. The (*a*) magnitude and (*b*) phase of the Fourier transform, with respect to the probe position, of the plasmon-loss shadow-image-intensity data set from graphite at a spatial frequency of $(0.34\ \text{nm})^{-1}$. The usual interference features expected in the zero-loss data can also be observed here.

5. Conclusions

By solving the phase problem for beams diffracted by crystalline specimens, we have demonstrated that images with a resolution well beyond the conventional point and information limits can be formed in transmission electron microscopy. Using a microscope with a relatively poor point resolution (0.45 nm) and information limit (0.33 nm), we have examined the accuracy of the phase retrieval in a ptychographically reconstructed image that

resolves the 0.136 nm 'dumbbell' spacing in silicon viewed along ⟨110⟩. It is found that the effects of aperture charging can be detected using the redundancy of the data set and, compounded by distortions in the detection system, may slightly affect the accuracy of the reconstruction. However, the images display the expected dumbbell contrast and it should be remembered that the phase errors described here are much smaller than the phase shifts introduced into conventional bright-field imaging by the lens aberrations. As another example of an application of the phase information, we have used the phase information to compare the Bragg scattering of the zero-energy-loss and plasmon-loss electrons and found that the reconstructed beam phases are within the experimental error of each other, supporting the view that the scattering of the plasmon-loss electrons is dominated by multiple elastic–inelastic scattering. The method, as it has been described here, can be applied to the study of any perfectly periodic structures. Since the probe scan need only include a few unit cells, the method is very suited to the structural study of nanocrystals or tiny precipitates. Any application of conventional bright-field imaging to a perfectly periodic structure is also a candidate for aperture synthesis.

When the specimen is thick enough for dynamical scattering effects to become important, the reconstructed beam phases are not equal to those that would be scattered by plane-wave illumination (Plamann & Rodenburg, 1998). In such cases, however, the conventional bright-field images are usually compared with forward simulations of trial structures. A similar approach could be applied here. Indeed, the higher-resolution information included in the data and the lack of lens aberrations create a much more constrained, and therefore probably more accurate, comparison.

It is, however, important to re-emphasise that the method described here is a simplification of a general method that does not require perfect periodicity. For a non-periodic specimen, the scattering becomes a continuous distribution and is no longer a set of discrete beams. For a given spatial frequency of interference, $\boldsymbol{\rho}'$, the data set $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ contains a convolution in $\boldsymbol{\mu}'$ of the disc-overlap function with a continuous function representing the interference between all parts of the specimen scattering separated by the reciprocal-space vector $\boldsymbol{\rho}'$. In principle, we can deconvolve the disc-overlap function, even when the lens parameters are not accurately known (McCallum & Rodenburg, 1993*b*) and apply a continuous object function reconstruction, which will include stepping out to form images of non-periodic specimens at resolutions many times the conventional limit.

This work was supported by the EPSRC. The authors also gratefully acknowledge The Royal Society, St John's College Cambridge and the Cambridge Philosophical Society for financial support.

References

Cowley, J. M. (1969). *Appl. Phys. Lett.* **15**, 58–59.

Cowley, J. M. & Jap, B. K. (1976). *Scanning Electron Microscopy*, Vol. 1, edited by O. Johari, pp. 377–380. Chicago: IITRI.

Craven, A. J. & Colliex, C. (1977). *Proceedings of EMAG77, Glasgow, Scotland. Inst. Phys. Conf. Ser.* No. 36, pp. 271–273.

Hegerl, R. & Hoppe, W. (1970). *Ber. Bunsenges. Phys. Chem.* **74**, 1148–1154.

Hoppe, W. (1969). *Acta Cryst.* A**25**, 495–501, 508–514.

Hoppe, W. (1982). *Ultramicroscopy*, **10**, 187–198.

Hoppe, W. & Strube, G. (1969). *Acta Cryst.* A**25**, 502–507.

Konnert, J., D'Antonio, P., Cowley, J. M., Higgs, A. & Ou, J. J. (1989). *Ultramicroscopy*, **30**, 371–384.

Lichte, H. (1991). *Adv. Opt. Electron Microsc.* **12**, 25–91.

McCallum, B. C. & Rodenburg, J. M. (1993*a*). *Ultramicroscopy*, **52**, 85–99.

McCallum, B. C. & Rodenburg, J. M. (1993*b*). *J. Opt. Soc. Am.* A**10**, 231–239.

Nathan, R. (1976). *Digital Processing of Biomedical Images*, edited by K. Preston & M. Onoe, pp. 75–88. New York: Plenum.

Nellist, P. D., McCallum, B. C. & Rodenburg, J. M. (1995). *Nature (London)*, **374**, 630–632.

Nellist, P. D. & Rodenburg, J. M. (1994). *Ultramicroscopy*, **54**, 61–74.

Orchowski, A., Rau, W. D. & Lichte, H. (1995). *Phys. Rev. Lett.* **74**, 399–401.

Plamann, T. & Rodenburg, J. M. (1998). *Acta Cryst.* A**54**, 61–73.

Ramachandran, G. N. & Srinivasen, R. (1970). *Fourier Methods in Crystallography*. New York: Wiley-Interscience.

Rodenburg, J. M. & Bates, R. H. T. (1992). *Philos. Trans. R. Soc. London Ser. A*, **339**, 521–553.

Rodenburg, J. M., McCallum, B. C. & Nellist, P. D. (1993). *Ultramicroscopy*, **48**, 303–314.

Spence, J. C. H. (1978). *Scanning Electron Microscopy*, Vol. 1, edited by O. Johari, pp. 61–68. AMF O'Hare: SEM Inc.

Spence, J. C. H. (1988). *Experimental High-Resolution Electron Microscopy*. Oxford University Press.

Spence, J. C. H. & Cowley, J. M. (1978). *Optik (Stuttgart)*, **50**, 129–142.

Van Dyck, D., Op de Beeck, M. & Coene, W. (1993). *Optik (Stuttgart)*, **93**, 103–107.

Vincent, R., Midgley, P. A. & Spellward, P. (1993). *Proceedings of EMAG93, Liverpool, England. Inst. Phys. Conf. Ser.* No. 138, pp. 119–124.

Zeitler, E. & Thomson, M. G. R. (1970). *Optik (Stuttgart)*, **31**, 258–280, 359–366.