Acta Cryst. (1998). A**54**, 61–73

Electron Ptychography. II. Theory of Three-Dimensional Propagation Effects

T. PLAMANN† AND J. M. RODENBURG*

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, England. E-mail: jr43@phy.cam.ac.uk

(Received 24 March 1997; accepted 23 July 1997)

Abstract

The interpretation of image reconstructions obtained by ptychography in a scanning transmission electron microscope (STEM) is investigated. In this technique, a number of coherent electron diffraction patterns are obtained from adjacent points of a crystalline sample by scanning a focused electron beam. The phase problem is solved by using an interference phenomenon, thus obtaining a very high resolution (0.136 nm) reconstruction of the specimen. Simulations presented here assist experiments described in a previous paper [Nellist & Rodenburg (1998). *Acta Cryst.* A**54**, 49–60]. In this paper, the simulation method which is best suited for the calculation of ptychographical reconstructions for perfect crystals is described. Simulation results are presented for the simplest case of diffraction beams with single overlap and it is shown that the reconstruction does not match the exit-surface wave function for axial plane-wave illumination but instead resembles the image wave function with a fixed focus setting on the middle section of the specimen. A second reconstruction method is described, in which a greater angular diameter of the aperture is used and every beam is re-interfered with the central beam. The variation of the amplitude and phase of all reconstructed beams with crystal thickness is very similar up to a thickness of about 25 nm. It is argued that such a simple relationship can only be expected for projected structures containing identical atomic columns.

1. Introduction

Ptychography is a name given to a technique invented by Hoppe (Hoppe, 1969; Hegerl & Hoppe, 1970; Hoppe, 1982) that aims to solve the diffraction-pattern phase problem by interfering adjacent Bragg reflections coherently and thus determining their relative phase. In the original formulation, Hoppe envisaged that such interference could be effected by placing a very narrow aperture in the plane of the specimen so that each reciprocal-lattice point would be spread out and thus overlap with one another. The name ptychography, from the Greek $\pi\tau\upsilon\xi$, which means fold, derives from this optical configuration; each reciprocal-lattice point is convoluted with some function, and thus made to interfere with its neighbours. In fact, measuring only the intensities of interfering adjacent diffracted beams still leads to an ambiguity of two possible complex conjugates for each underlying complex diffraction amplitude and so in its original formulation ptychography is equivalent to the well known theorem that for a finite specimen (that is one delineated by a narrow aperture, sometimes known as a 'finite support') the one-dimensional phase problem is soluble to within an ambiguity of 2^N, where N is the number of Fourier components that make up the specimen (Rodenburg, 1989). However, such ambiguities may be resolved by changing the phase, profile or position of the illuminating beam in some way (Hoppe, 1969). The fact that we must measure not only the intensities of the diffracted beams but also the intensities lying midway between the beams, where the convoluted Bragg beams interfere, is similarly an alternative statement of the Nyquist sampling theorem for components of diffracted intensity, which in general have twice the frequency (in reciprocal space) of their underlying complex amplitudes.

Ptychography has one very practical experimental advantage: the degree of coherence required in the illuminating beam is rather modest; it is only necessary that adjacent diffracted beams can interfere with one another significantly. In electron reference-beam techniques, such as conventional bright-field imaging (Spence, 1988), holography (Lichte, 1992) and focal-series reconstructions (Van Dyck, Op de Beeck & Coene, 1993) in transition electron microscopy, the most profound limitation to spatial resolution arises from the narrow width of the coherence envelope (typically 15–30 mrad). In the previous paper (Nellist & Rodenburg, 1998) (paper I), we applied ptychography experimentally to obtain an image of a silicon crystal at several times the resolution determined by these conventional limits. The strength of the technique derives from the fact that all pertinent wave interferences occur in the vicinity of the specimen itself and not across macroscopic distances within a microscope or diffractometer. The method is much less affected by lens aberrations than conventional imaging methods and can be thought of as a simplification of a much more general imaging method that can cope with non-crystalline objects. Ptychography, or its derivatives, may therefore provide a solution to the long-

† Now at H. H. Wills Laboratory, University of Bristol, Bristol BS8 1TL, England.

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

ISSN 0108-7673 © 1998

standing resolution problem in electron imaging (or any coherence-limited imaging technique).

In this paper, we start by extending the thin-specimen approximation of ptychography in order to account for three-dimensional scattering effects in the scanning transmission electron microscope (STEM) mode. These turn out to be important; they clearly express themselves in the experimental data presented in the preceding paper. However, we will show that the reconstructed-object wave function (ROWF) we obtain by phasing the diffraction pattern while supposing it to be a two-dimensional grating still gives a good representation of the exit-surface wave function (ESWF) obtained under axial plane-wave illumination, albeit apart from a slight modification due to a free-space defocus propagation, provided the specimen is not too thick. In particular, the calculations presented here suggest that the direct interpretation of experimental 'super-resolution' images (Nellist, McCallum & Rodenburg, 1995, Nellist & Rodenburg, 1998) is justified.

2. Practical optical configurations for ptychography

The ideal ptychographical experiment would consist of the optical configuration shown in Fig. 1. A single plane wave is incident upon the specimen, which, for simplicity, we first assume is crystalline. In the far field, we have a diffraction pattern consisting of a number of reflections. Ignoring experimental details, let us say we are able to interfere each pair of adjacent reflections and thus measure the relative phase between each pair of beams. If we arbitrarily assign zero phase to the zero-order beam (this is of no consequence, being the absolute phase of the entire wavefield at a particular time), we may then phase any particular beam by choosing a path between it and the zero-order beam *via* a number of intermediate beams. The phase of the particular beam is then given by adding the phase differences between each of the intermediate pairs of beams. For a two-dimensional object, there may be a number of different routes to a particular reflection, in which case we can test the accuracy of our measurements by phase closure. Having assigned a complex amplitude (modulus and phase) to all the diffracted beams, we may then perform a back Fourier transform on the resulting function to produce the ROWF. Assuming we can do this for every point in reciprocal space, the ROWF should be identical to the ESWF. The ESWF may then be related to the atomic structure, say by assuming, in the case of electrons, that the phase of the ESWF is proportional to the projected atomic potential, or by more sophisticated inverse methods which account for propagation and inelastic effects.

The configuration in Fig. 1 could be realized experimentally by tilting both the specimen and the illuminating beam simultaneously in a conventional transmission electron microscope (CTEM) while using a lens as an interferometer. If we form an image (a set of interference fringes) from each pair of beams, we do not even encounter the phase ambiguity alluded to above because we can measure the absolute position of the fringes relative to the optic axis of the microscope. However, achieving this form of set-up is experimentally non-trivial, especially when we consider the mundane constraints of specimen drift, damage, contamination and the difficulty of controllably altering the beam alignment and specimen tilt without introducing variations in the lens parameters (say due to variation in the specimen height, of shifts due to coma or other non-axial aberrations).

In contrast with these difficulties, a much more straightforward ptychographical configuration arises in the STEM mode illustrated in Fig. 2(*a*). An aperture in the back-focal plane of a lens defines a range of incident vectors upon the specimen. If the lens is perfect and the specimen lies in the focal plane, a 'probe' function in the form of any Airy disc is incident upon the specimen. In the far field, *i.e.* in the so-called microdiffraction plane, each diffraction spot is now conveniently convoluted with a disc-shaped function. For a well chosen aperture size, each disc overlaps with its nearest neighbours, leading to the classic definition of ptychography. The phase ambiguity may be resolved by either moving the probe or by introducing some other known phase distribution across the aperture function: a probe shift is equivalent to a phase ramp across the aperture function, a defocus is equivalent to a parabolic phase distribution.

Upon using the principle of reciprocity (Pogany & Turner, 1968), we can reverse the ray propagation in Fig. 2(*a*) and we arrive at the configuration shown in Fig. 2(*b*), which were refer to as the CTEM configuration. The configurations in Fig. 2(*a*) and Fig. 2(*b*) are

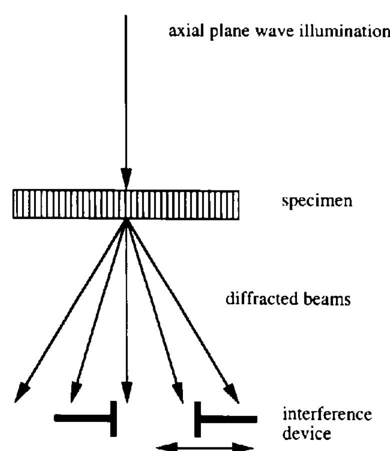

Fig. 1. Optical configuration of an ideal ptychographical experiment. A single axial plane wave interacts with the specimen and the relative phase between each pair of diffracted beams is measured by an interference device, which can be moved. The exit surface wave function (ESWF) can be reconstructed.

equivalent. Different illumination tilts in Fig. 2(*b*) correspond to different points in the microdiffraction plane of Fig. 2(*a*), and different probe positions in Fig. 2(*a*) correspond to different position vectors in the image plane of Fig. 2(*b*). It is clear that, as we change the illumination angle in Fig. 2(*b*), the ESWF we are attempting to measure will change if the specimen is substantially thick. As a consequence, neither Fig. 2(*a*) nor Fig. 2(*b*) correspond to the ideal configuration in Fig. 1, except that, for thin specimens and small scattering angles, these geometries are approximately equivalent. In what follows, we will be largely concerned with how the ROWFs obtained in the STEM configuration depart from the true ESWF that we would observe in the ideal plane-wave configuration of Fig. 1.

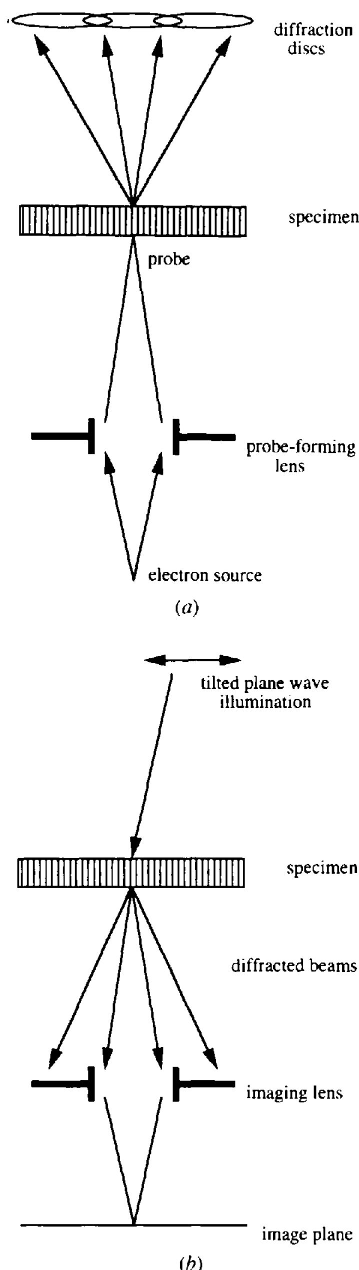

Fig. 2. (*a*) Ptychography in the STEM configuration. A lens forms a probe onto the specimen; in the so-called microdiffraction plane, overlapping diffraction discs yield the phase information. (*b*) The equivalent CTEM configuration. Different illumination tilts correspond to different points in the microdiffraction plane in (*a*). Neither (*a*) nor (*b*) are equivalent to Fig. 1 and the reconstructed-object wave function (ROWF) is different from the ESWF of Fig. 1 except for suitably thin specimens and small scattering angles.

In the STEM configuration, we collect a four-dimensional data set of microdiffraction patterns $[|M(\boldsymbol{\mu}', \boldsymbol{\rho})|^2]$ recorded as a function of the probe position $\boldsymbol{\rho}$. In this notation, reciprocal-space coordinates are denoted by a prime, $\boldsymbol{\mu}'$ being a two-dimensional vector describing a position in the microdiffraction plane (*i.e.* a certain scattering angle) and $\boldsymbol{\rho}$ being a two-dimensional vector describing the position of the probe in the specimen plane. If and only if the interaction of the electron wave with the specimen is multiplicative is the wave function in the microdiffraction plane a convolution of the specimen function with the aperture function in reciprocal space such that the intensity distribution is given by

$$|M(\boldsymbol{\mu}', \boldsymbol{\rho})|^2 = \iint A(\boldsymbol{\mu}' - \mathbf{a}')A^*(\boldsymbol{\mu}' - \mathbf{b}')\Psi(\mathbf{a}')\Psi^*(\mathbf{b}') \times \exp[2\pi i\boldsymbol{\rho}(\mathbf{b}' - \mathbf{a}')]\,\mathrm{d}\mathbf{a}'\,\mathrm{d}\mathbf{b}'. \quad (1)$$

We shall proceed by taking the Fourier transform with respect to the probe position $\boldsymbol{\rho}$ to obtain a four-dimensional reciprocal-space function

$$G(\boldsymbol{\mu}', \boldsymbol{\rho}') = \int |M(\boldsymbol{\mu}', \boldsymbol{\rho})|^2 \exp(2\pi i\boldsymbol{\rho} \cdot \boldsymbol{\rho}')\,\mathrm{d}\boldsymbol{\rho}. \quad (2)$$

We can interpret $G(\boldsymbol{\mu}',\boldsymbol{\rho}')$ as the set of all fringes of a given spatial frequency for each point $\boldsymbol{\mu}'$ in the microdiffraction plane. If we assume a perfectly crystalline object and if the size of the objective aperture is such that the diffraction discs have just single overlap, the intensity in the overlap region between the discs $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$ can easily be shown to be

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

where $\alpha_{\mathbf{G}}$ and $\alpha_{\mathbf{G+H}}$ are the phases of the beams $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$, respectively. The intensity in the overlap region is thus seen to vary sinusoidally as the probe is scanned parallel to the disc separation vector $\mathbf{H}$.

After the Fourier transform has been taken with respect to $\boldsymbol{\rho}$, $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ has considerable magnitude only for values of $\boldsymbol{\mu}'$ within the overlap region between the two discs when $\boldsymbol{\rho}'$ is equal to $\mathbf{H}$ or $-\mathbf{H}$. For $\boldsymbol{\rho}' = \mathbf{H}$, we have

$$G(\boldsymbol{\mu}', \mathbf{H}) = |\Psi_{\mathbf{G}}||\Psi_{\mathbf{G+H}}| \exp\big(i(\alpha_{\mathbf{G+H}} - \alpha_{\mathbf{G}}) + \{\chi[\boldsymbol{\mu}' - (\mathbf{G}+\mathbf{H})] - \chi(\boldsymbol{\mu}' - \mathbf{G})\}\big). \quad (4)$$

The magnitude of $G(\boldsymbol{\mu}', \mathbf{H})$ is thus given by the amplitude product of the re-interfering beams and its phase by their phase difference (including the aberrations introduced by the lens). It should be noted and it will be assumed in the following that this is true in the presence of dynamical scattering including inelastic scattering (see Nellist & Rodenburg, 1998, for a discussion of the effects of the lens aberrations).

When we know all the relative phases of each pair of adjacent discs, we can phase all the diffraction orders, starting from the origin in diffraction space. The reconstructed wave function in reciprocal space can be Fourier transformed to give the ROWF in real space. The ROWF has a higher resolution than conventional images recorded in the same optical configuration since these would have a spatial frequency cut-off given by the radius of the aperture of the probe-forming lens. The ROWF has a spatial frequency cut-off defined by the highest-order disc which has been phased. Therefore, ptychography overcomes the resolution problem of conventional imaging and we will refer to it also as super-resolution imaging, which can easily be extended to the reconstruction of aperiodic structures from the four-dimensional microdiffraction data set (Rodenburg & Bates, 1992).

3. The assumption of a multiplicative interaction between probe and specimen

When we describe dynamical diffraction effects for high-energy electrons, because of its forward-scattering character, all vectors are usually separated into a lateral vector and a z-dependent part:

$$\mathbf{r} = (\mathbf{R}, z); \quad \mathbf{k} = (\mathbf{K}_t, k_z); \quad \nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) = \left(\nabla_{\mathbf{R}}, \frac{\partial}{\partial z}\right). \tag{5}$$

A wave function $\Psi_0(\mathbf{R}) = \Psi(\mathbf{R}, z = 0)$ of high-energy electrons is incident on the entrance of surface of a crystal and interacts with a potential distribution $V(\mathbf{r})$ inside the specimen of constant thickness t. The wave function at the exit surface of the crystal will be called $\Psi_t(\mathbf{R}) = \Psi(\mathbf{R}, z = t)$. Only elastic scattering is considered and we can make use of the time-independent Schrödinger equation

$$[\nabla^2 + 4\pi^2\mathbf{k}^2 + (2m/\hbar^2)V(\mathbf{r})]\Psi(\mathbf{r})\exp(2\pi i\mathbf{k}\cdot\mathbf{r}) = 0. \tag{6}$$

Equation (6) is sufficiently accurate in the case of high-energy electron diffraction (HEED) provided that the relativistic electron mass $m = m_0(1 - v^2/c^2)^{-1/2}$ is used (Fujiwara, 1961). Using the high-energy approximation, in which the second-order derivative with respect to z is neglected, and the projection approximation, in which the potential variation along the z direction is neglected [$V(\mathbf{R}, z) = V(\mathbf{R})$], we obtain the fundamental equation of high-energy electron diffraction (see Bird, 1989; Van Dyck, 1985):

$$\left[4\pi i k_z \frac{\partial}{\partial z} + \nabla_{\mathbf{R}^2} + (2m/\hbar^2)V(\mathbf{R})\right]\Psi(\mathbf{R}, z) = 0. \tag{7}$$

The formal solution of (7) is given by

$$\Psi_t(\mathbf{R}) = \exp[2\pi i\mathbf{A}(\mathbf{R})t]\Psi_0(\mathbf{R}), \tag{8}$$

where $\mathbf{A}$ is an operator defined as

$$\mathbf{A}(\mathbf{R}) = [(1/8\pi^2 k_z)\nabla_{\mathbf{R}^2} + (\sigma/2\pi)V(\mathbf{R})] \tag{9}$$

and the interaction constant σ is defined as

$$\sigma = m/2\pi k_z \hbar^2. \tag{10}$$

In the well known phase-object approximation (POA), the propagation term in (9) is neglected and we obtain

$$\Psi_t(\mathbf{R}) = \exp[i\sigma V(\mathbf{R})t]\Psi_0(\mathbf{R}). \tag{11}$$

In this approximation, the effect of the specimen on the incident wave is represented by multiplying the incident wave function $\Psi_0(\mathbf{R})$ by a transmission function whose phase is proportional to the projected potential. The POA represents the limiting case for high voltages, for which the wavelength tends to zero and the interaction constant tends to a constant value. The terms appearing in the expansion of (11) can be identified with those of the Born series and therefore represent the multiple scattering of the electrons, although neglecting their propagation. In practice, the POA has only a limited range of validity as it is only correct up to first order in thickness and is suited only to rather low resolution. However, we can obtain an improvement to the POA by splitting the propagation term symmetrically:

$$\Psi_t(\mathbf{R}) = \{\exp[(t/2)(i/4\pi k_z)\nabla_{\mathbf{R}^2}]\exp[i\sigma V(\mathbf{R})t] \times \exp[(t/2)(i/4\pi k_z)\nabla_{\mathbf{R}^2}]\}\Psi_0(\mathbf{r}). \tag{12}$$

We will call (12) the improved POA (IPOA), which we have obtained by separating propagation and interaction effects into three terms. Equation (12) is correct to second order in t and yields an important extension of the POA, which holds for greater thicknesses and higher resolution. According to the IPOA, the electron wave travels through free space before and after the central plane of the crystal, where it interacts multiplicatively with the projected potential. Van Dyck (1983) used the same expression for the discussion of improvements to slice functions in multislice electron diffraction calculations In the case of a general primary wave function $\Psi_0(\mathbf{r})$ such as a nanoprobe, we realize that the wave function

$$\Psi_m(\mathbf{R}) = \{\exp[(t/2)(i/4\pi k_z)\nabla_{\mathbf{R}^2}]\}\Psi_0(\mathbf{R}) \tag{13}$$

interacts multiplicatively with the usual phase-object function instead of $\Psi_0(\mathbf{R})$; $\Psi_m(\mathbf{R})$ is the wave function that would arrive at the middle section of the specimen after free-space propagation. Broeckx, Op de Beeck & Van Dyck (1995) used channelling theory for the discussion of the factorization of the STEM wave function into the probe function at the entrance surface of the specimen and a specimen-dependent part. However, (12) implies that this assumption is too strict. The plane of interaction of the probe with the specimen can be thought of as the middle section of the specimen. The simulations described below serve as a good demonstration of this property.

4. Calculation of ptychographical reconstructions

We now turn to the question of the calculation of ptychographical reconstructions. If we wanted to simulate the probe propagation through the crystal for each probe position, we would have to make great demands on the size and the sampling of the supercell used. The computer time required would be proportional to $N_{\mu'}^2 N_{\rho'}^2 \log N_{\mu'}$, where $N_{\mu'}$ is the number of sampling points in the microdiffraction plane (that is, the number of beams included in the calculation) and $N_{\rho'}$ the number of probe positions. This would be identical to the simulation of ADF images (Kirkland, Loane & Silcox, 1987) which is notoriously slow. However, the propagation of the probe does not have to be calculated, at least not in the case of a perfect crystal, because in the simplest reconstruction methods the relative phase between diffraction discs is measured at one point of their overlap region (typically the midpoint), in which only two beams from opposite points in the illuminating convergent beam interfere (Fig. 3). For instance, we measure the relative phase of the discs $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$, where $\mathbf{H}$ is the disc separation vector, at a point into which beams with wave vectors $-\mathbf{H}/2$ and $\mathbf{H}/2$ are scattered. These beams propagate through the crystal independently of all other scattered beams and can thus be modelled by (tilted) plane-wave illumination conditions in ordinary multislice programs (Goodman & Moodie, 1974; Self, O'Keefe, Buseck & Spargo, 1983). The advantage is that after the calculation of two beams of interest we can look at their amplitude product and relative phase as a function of thickness similar to a *Pendellösung* plot. However, in the case of aperiodic specimens, mixing of all the beams in the scattering process occurs and the full propagation of the probe has to be simulated.

Illumination tilts $\mathbf{K}_t$ of the incoming plane wave can be modelled by the alteration of the propagation function in multislice programs according to

$$P(\mathbf{K}) = \exp[\pi i \lambda \Delta z (\mathbf{K} + \mathbf{K}_t)^2], \tag{14}$$

which corresponds to a slight shear of the specimen such that every successive slice is shifted by a small amount. In order to calculated complex products of beam amplitudes, we have the choice between two calculation methods (Fig. 4). According to the theorem of reciprocity, the rocking curve of a beam from a structure that has a mirror plane coinciding with the plane $z = t/2$, *e.g.* if the projection approximation holds, is symmetric with respect to the Bragg condition (Pogany & Turner, 1968), that is:

$$\Psi_{\mathbf{G}}(-\mathbf{G}/2 + \Delta\mathbf{K}_t) = \Psi_{\mathbf{G}}(-\mathbf{G}/2 - \Delta\mathbf{K}_t). \tag{15}$$

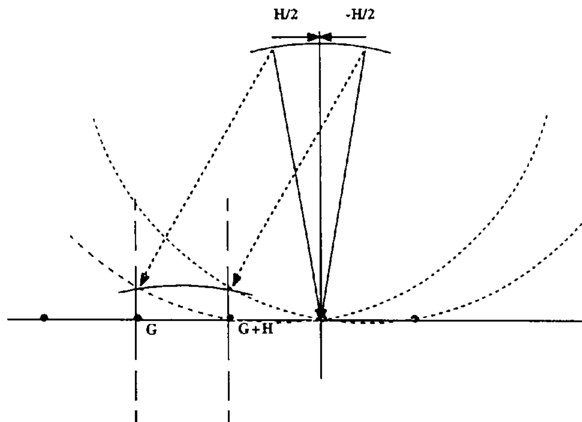

Fig. 3. The Ewald-sphere construction for a microdiffaction experiment. For each beam from the convergent beam, one Ewald sphere must be drawn and all the resulting diffraction patterns should be summed coherently. To find the relative phase between the reflections $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$, beams scattered from diametrically opposite points in the incoming convergent beam (beam tilts $\mathbf{H}/2$ and $-\mathbf{H}/2$) are re-interfered. The two scattered waves cross the same point of the shape transform (*i.e.* the excitation errors of the two reflections are equal).

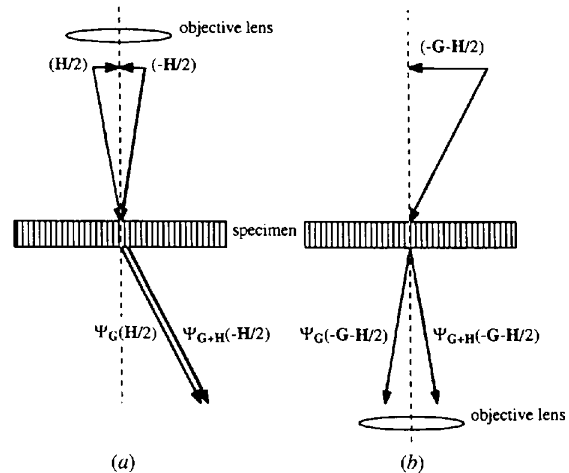

Fig. 4. Two possible methods to calculate the interference of the beams $\mathbf{G}$ and $\mathbf{G}+\mathbf{H}$ in the middle of the overlap between the two beams. (*a*) STEM geometry: two beam tilts $(-\mathbf{H}/2)$ and $(\mathbf{H}/2)$ are introduced and the complex product $\Psi_{\mathbf{G}+\mathbf{H}}(-\mathbf{H}/2)\Psi_{\mathbf{G}}(\mathbf{H}/2)$ is calculated. (*b*) CTEM geometry: one beam tilt $(-\mathbf{G}-\mathbf{H}/2)$ is introduced and the complex product $\Psi_{\mathbf{G}+\mathbf{H}}(-\mathbf{G}-\mathbf{H}/2)\Psi_{\mathbf{G}}^*(-\mathbf{G}-\mathbf{H}/2)$ is calculated.

In this notation, the beam tilt is put in brackets and $\Psi_{\mathbf{G}}(-\mathbf{G}/2)$ corresponds to the Bragg condition for the reflection **G**. In particular, we get the identity

$$\begin{aligned}\Psi_{\mathbf{G}+\mathbf{H}}(-\mathbf{H}/2)\Psi^{*}_{\mathbf{G}}(\mathbf{H}/2)\\ = \Psi_{\mathbf{G}+\mathbf{H}}(-\mathbf{G}-\mathbf{H}/2)\Psi^{*}_{\mathbf{G}}(-\mathbf{G}-\mathbf{H}/2). \qquad (16)\end{aligned}$$

The two sides of (16) correspond to the STEM and the CTEM geometry, respectively, and we find that the phase information in a ptychographical experiment can be simulated either way. Note that the propagator functions applied in the two cases are different, but the amplitudes and relative phases calculated by the two methods turn out to be almost the same, which gives us confidence in the simulation method. In contrast to conventional lattice imaging, we do not have to calculate the effect of the lens aberrations, as we usually choose diffraction information from the centre of overlap between the discs, which is not affected by the lens transfer function, provided that the objective lens is perfectly aligned (Nellist & Rodenburg, 1998). This is very different from image simulations in high-resolution electron microscopy (Spence, 1988): the defocus is no free parameter in the calculations. Instead, all sections of a thick specimen are in a sense imaged in focus and we will discuss in the next section how this is represented in the image reconstructions.

5. Ptychography from partly overlapping discs: reconstruction method I

In this paper, we use silicon as an example in order to examine the two simplest reconstruction methods of ptychography. The characteristic dumbbell structure showing pairs of Si atoms with an apparent spacing of 0.136 nm is seen in the ⟨110⟩ orientation, which is often used as a practical resolution test for high-resolution microscopes. The {002} and {222} beams are kinematically forbidden and only weakly excited for small crystal thicknesses. In order to resolve the atom pairs at a resolution of 0.136 nm, all 19 inner reflections up to the {004} beams have to be included in the imaging process. However, if fewer beams contribute, images can be taken that apparently show up the atomic positions but cannot be regarded as structure images (Hutchison, 1982; Hutchison & Waddington, 1988). Izui, Furono & Otsu (1977) were the first to observe the dumbbell contrast in a microscope with a point resolution of only 0.29 nm. Following this experimental success, simulations were carried out using the multislice (Spence, O'Keefe & Kolar, 1977) and the Bloch-wave methods (Desseaux, Renault & Bourret, 1977), which showed that the characteristic image contrast is periodic with respect to both thickness and defocus. A detailed study regarding conventional bright-field imaging of elemental semiconductors including effects of the lens aberrations and of thickness variations is given in Glaisher, Spargo & Smith (1989). Other techniques to image Si⟨110⟩ at atomic resolution are holography (Orchowski, Rau & Lichte, 1995) and ADF imaging (McGibbon & Pennycook, 1994).

The simplest situation for ptychography occurs when diffraction discs partly overlap. In this case, the object wave function in reciprocal space can be readily reconstructed using a technique that we will refer to as reconstruction method I (RMI). The amplitudes of the ROWF are taken from the centres of the diffraction discs and its phases are successively determined from the set of relative phases between adjacent discs. To find the relative phase between the reflections **G** and **G** + **H**, we use beams from diametrically opposite points in the incoming convergent beam that correspond to beam tilts **H**/2 and −**H**/2, respectively. In order to simulate the experiment on silicon ⟨110⟩ (electron energy 100 keV) presented in Nellist & Rodenburg (1998), we proceeded as follows: First, we calculated amplitudes and phases of all beams including the {004} reflections for axial illumination, that is $\mathbf{K}_t = 0$. This produced the ESWF *via* one Fourier transformation, but also the beam amplitudes for the ptychographical reconstruction, which are extracted from the centres of the diffraction discs. Second, we calculated phase differences between pertinent beams by applying tilted plane-wave illumination conditions. All beams were phased successively, using the shortest routes available. After reconstructing the complex values of 19 beams, we took a Fourier transform to obtain a periodic function in real space, which is our ROWF. The multislice simulations were carried out using the software package *CERIUS* (developed by Molecular Simulations) on a Silicon Graphics computer. Electron scattering factors by Doyle & Turner (1968) were used. For Si⟨110⟩, a slice of thickness of 0.384 nm was chosen, which corresponds to the inclusion of zero-layer reflections only (projection approximation). The Debye–Waller factor was set to 0.45 (Saunders, private communication).

Let us consider the case of a crystal with a thickness of 7 nm (*cf.* Fig. 6 in Nellist & Rodenburg, 1998). Figs. 5(*a*), (*b*) show the magnitude and phase of the ESWF, respectively. Figs. 5(*c*), (*d*) show the same image wave function with a focus setting on the middle section of the structure, *i.e.* the complex amplitudes of the scattered beams obtained under axial illumination conditions have been multiplied with a defocus term corresponding to a focus defect of $t/2$ before taking the Fourier transform. Figs. 5(*e*), (*f*) show the magnitude and phase of the ROWF, which clearly demonstrate the match with the 'defocused' wave function in Figs. 5(*c*), (*d*). It is the phase of the ROWF that mirrors the projected structure, and we can explain this easily using the IPOA in (12): The propagation of $\Psi_0(\mathbf{r})$ to the middle section of the specimen does not appear in the ROWF because the reconstruction is insensitive to defocus. The other propagation term in (12) corresponds to a phase factor

in diffraction space that cannot be measured. Hence, the ROWF is closer to a pure phase object than the ESWF because the ESWF does involve propagation through the specimen. Indeed, the phase change on top of the atomic columns is about 25% greater in the ROWF than in the ESWF. Similarly, the standard variation of the magnitude of the ROWF is 30% smaller than that of the ESWF. We cannot write the ROWF in a closed form which accounts for terms higher than second order in t, but the 'thick phase grating' formula by Cowley & Moodie (1962) yields the correct amplitudes of the reconstructed beams – which are taken from the centres of diffraction discs –

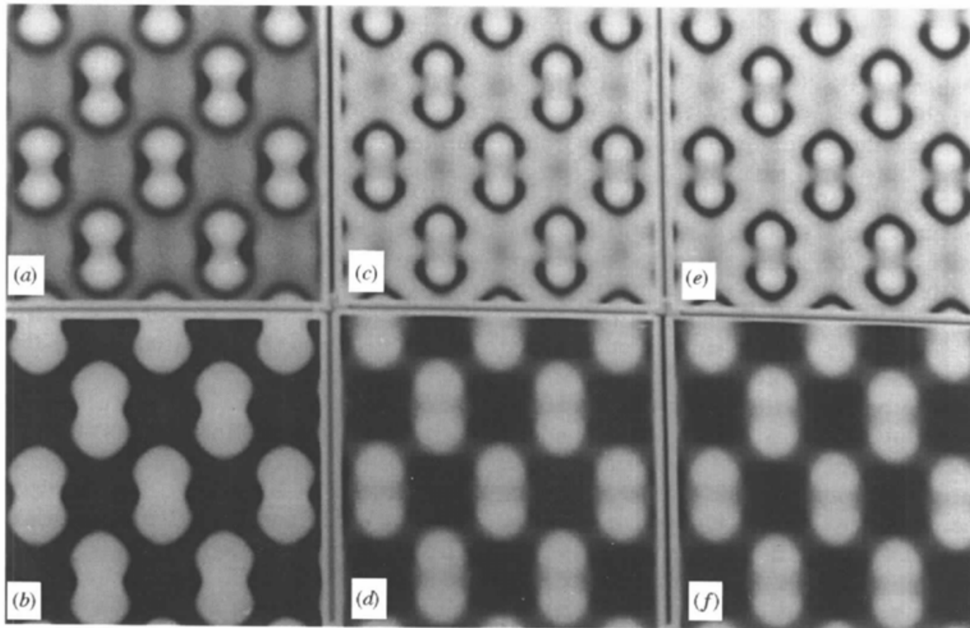

Fig. 5. Object wavefunctions from Si⟨110⟩ (thickness of 7 nm): (*a*) magnitude and (*b*) phase of the ESWF; (*c*), (*d*) 'back-propagated' wavefunction in magnitude and phase with a focus setting on the middle section of the structure; (*e*), (*f*) magnitude and phase of the ROWF, which clearly matches the 'defocused' wavefunction (*c*), (*d*).

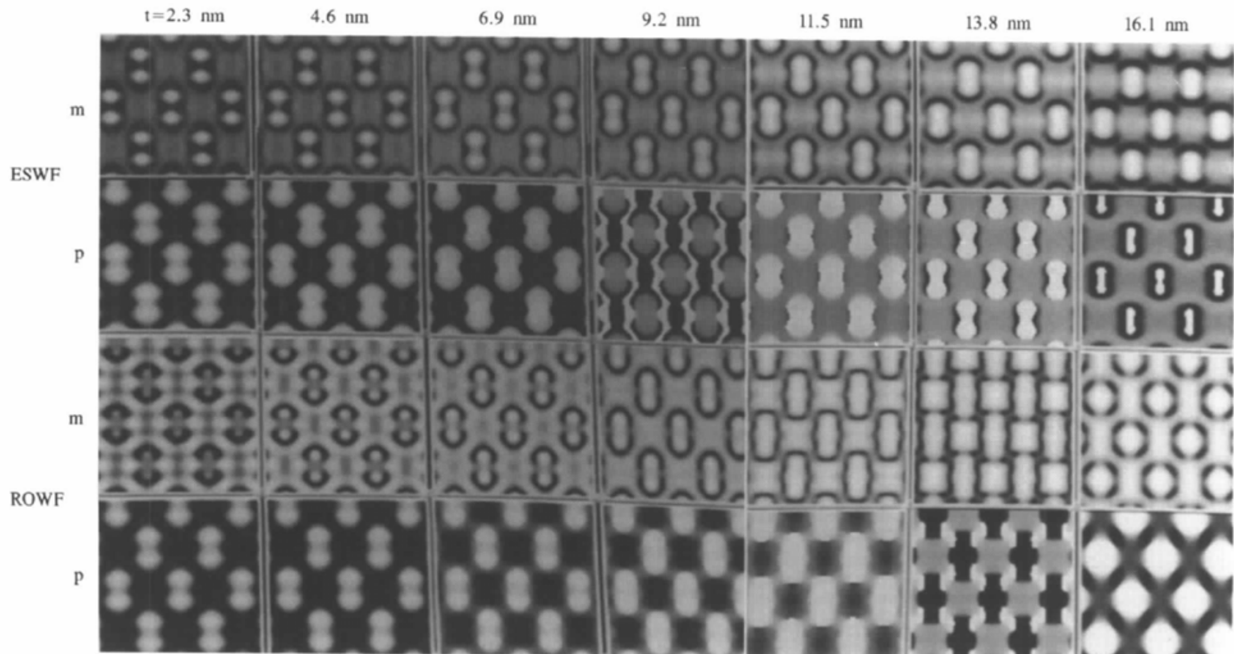

Fig. 6. Comparison between ESWF and ROWF of Si⟨110⟩ in magnitude (*m*) and phase (*p*) for thickness steps of 2.3 nm.

up to third order in t:

$$\Psi_{\rm ROWF}(\mathbf{R}) = \exp\left\{i\sigma \sum_{\mathbf{G}} V_{\mathbf{G}}[\sin(\pi s_{\mathbf{G}} t)/\pi s_{\mathbf{G}}] \exp(2\pi i \mathbf{G}\mathbf{R})\right\}. \quad (17)$$

Ptychography is one of several methods currently under investigation that aim at the reconstruction of a complex-valued specimen function. In the first group of approaches, the specimen is axially illuminated and phase information in the image is gained by focus variation in the objective lens (Van Dyck, Op de Beeck & Coene, 1993) or interference of the electron wave with a holographic reference wave (Lichte, 1992). In the second class of reconstruction methods, tilt series are employed (Kirkland, Saxton, Chau, Tsuno & Kawasaki, 1995) or different portions of the microdiffraction plane are recorded as a probe is scanned across the specimen [ptychography or quadrant detector imaging (Landauer, McCallum & Rodenburg, 1995)], both schemes being related to one another by reciprocity. In the former two cases, the aim is to reconstruct the wave function at the exit surface of the specimen up to the information limit of the microscope. However, it should be noted that the absolute defocus value required for the deconvolution will rarely be accurately known and will usually be determined by trial and error. In the latter two cases, the reconstruction can exceed the information limit. The relevant scattering geometry is one in which the beams that are being re-interfered pass through diametrically opposite regions of the objective lens and their path lengths during the propagation through the specimen are equal. The consequence is that, in this class of reconstruction methods, the defocus is not a parameter as in conventional imaging and the reconstructed wave corresponds to a *fixed* focus setting on the middle section of the specimen. Furthermore, deconvolution schemes (Rodenburg & Bates, 1992) should assume a probe profile corresponding to the middle section of the specimen.

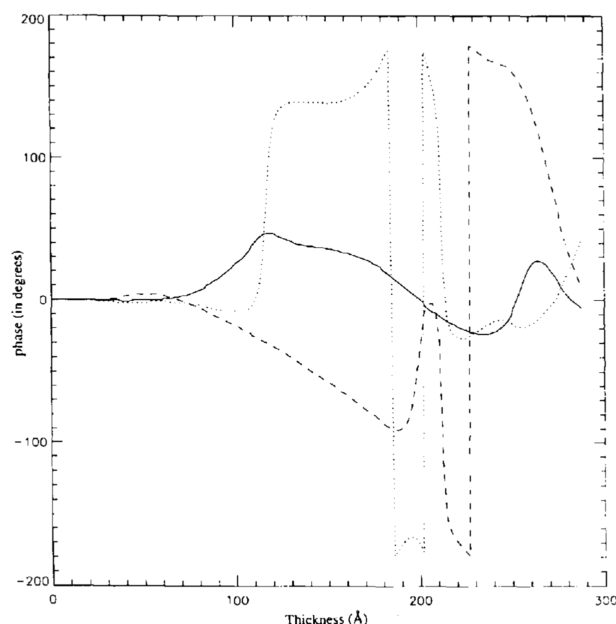

Fig. 7. Phase closures as a function of thickness: (000) → (1$\bar{1}$1) → (002) → (000) (solid line); (1$\bar{1}$1) → (002) → (1$\bar{1}$3) → (1$\bar{1}$1) (dashed line); (002) → (1$\bar{1}$3) → (004) → (002) (dotted line).

Fig. 6 shows a comparison between ESWF and ROWF for thickness steps of 2.3 nm. We note that the phase of the ROWF shows the projected structure up to a thickness of about 10 nm. Beyond this thickness, its resemblance to the 'defocused' ESWF also vanishes. For thicknesses exceeding 10 nm, we face increasing discrepancies in the phases assigned to the reconstructed beams depending on the route that we choose to successively phase the diffraction discs in the microdiffraction plane. This is well expressed in the phase-closure diagrams (Fig. 7). Considering the form of the microdiffraction patterns from silicon [*cf.* Fig. 1 in Nellist & Rodenburg (1998)], we see that the phases assigned to three overlapping discs that are arranged in a close triangle are uniquely defined if the three pairwise phase differences add to zero. It is clear that in the case of triangles which consist of symmetry-related reflections the phase closure is zero for all thicknesses, *e.g.* (000) → (1$\bar{1}$1) → (1$\bar{1}\bar{1}$) → (000). Fig. 7 shows the phase closures (measured in degrees) of (000) → (1$\bar{1}$1) → (002) → (000) (solid line), (1$\bar{1}$1) → (002) → (1$\bar{1}$3) → (1$\bar{1}$1) (dashed line) and (002) → (1$\bar{1}$3) → (004) → (002) (dotted line), respectively. The phase closure including the central beam (solid line in Fig. 7) does not exceed 50°. This corresponds to the validity of the assumption of a multiplicative interaction of the probe with the specimen at low resolution (the 002 reflection corresponds to a fringe spacing of 0.272 nm). However, the phase closures of the higher-order beams depart rather abruptly from zero for thicknesses greater than 10 nm. The notion of a uniquely defined ROWF becomes meaningless at this thickness but it should be noted that a higher accelerating voltage would increase this thickness value by some amount.

6. Relation between different $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ values

If the structure is sufficiently thin such that inelastic scattering can be neglected, every electron from the incoming wave is contained in each collected microdiffraction pattern and we have

$$\int |M(\boldsymbol{\mu}', \boldsymbol{\rho})|^2 \, d\boldsymbol{\mu}' = 1, \quad (18)$$

assuming that the wave function has been suitably normalized. If we take the Fourier transform with respect to the probe position $\boldsymbol{\rho}$ and exchange the order of integration, we obtain

$$\int G(\boldsymbol{\mu}', \boldsymbol{\rho}')\,d\boldsymbol{\mu}' = \delta(\boldsymbol{\rho}'). \quad (19)$$

Equation (19) signifies that fringes in amplitude and phase of a given spatial frequency cancel out over the entire microdiffraction plane. This applies to any form of incident wave and thus in theory also to one comprising only the beams $\mathbf{H}/2$ and $-\mathbf{H}/2$. As a consequence, according to (19), the sum of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ over all the conjugate points in the interference regions in each of the planes $\boldsymbol{\rho}' =$ constant [*e.g.* the interference regions in Fig. 5 in Nellist & Rodenburg (1998)] should be zero if we include only elastic scattering in the calculation. The reason for this constraint is that the sum of complex products of beam pairs is equal to the scalar product of two columns of the scattering matrix (Sturkey, 1962), which must vanish because of the unitarity of the scattering matrix. Fig. 8 shows the magnitude of the sums of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ over different sets of conjugate points in the interference regions comprising:

(*a*) only the first-order interference region of the beam pair $(1\bar{1}1)/(000)$ (solid line); for small thickness values, this is just the kinematical value of the beam $(1\bar{1}1)$, which increases linearly with thickness;

(*b*) the opposite first-order interference regions $(1\bar{1}1)/(000)$ and $(000)/(\bar{1}1\bar{1})$ (dotted line); a quadratic increase for small t results;

(*c*) all interference regions up to $\{004\}$ (dashed line); its deviation from zero is caused by the interference of beams with large wave vectors;

(*d*) the same regions as in (*c*) but with inelastic scattering included in the calculation according to an absorption potential $V_a(\mathbf{r}) = 0.1V(\mathbf{r})$, where $V(\mathbf{r})$ is the real potential used (dash-dotted line); (*d*) is always greater than (*c*), as expected.

It is interesting to note that (*d*) can be directly related to the contrast in ADF images since it is the deficit of electrons in the low-angle interference regions that corresponds to the redistribution of electrons scattered into high angles and that defines the contrast of the $\{111\}$ fringes in ADF images. The plateau in (*d*) expresses the well known absence of contrast reversal (Pennycook & Jesson, 1991).

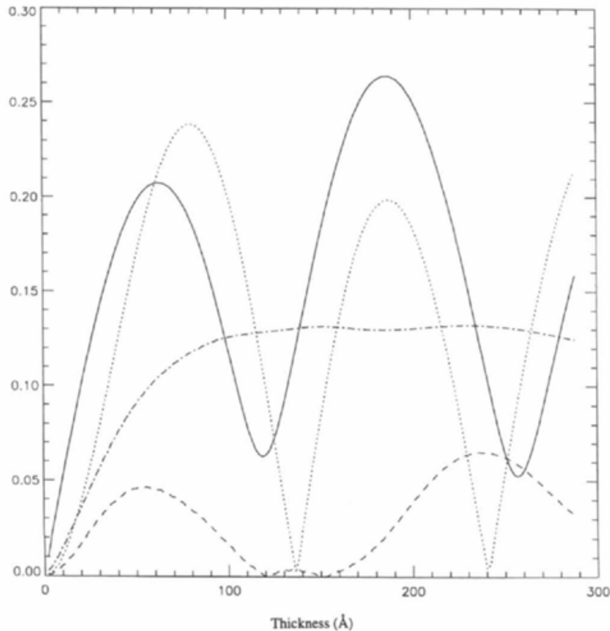

Fig. 8. Magnitudes of the sums of $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ over different sets of conjugate points in the interference regions in a plane $\boldsymbol{\rho}' =$ constant [corresponding to the disc separation vector $(1\bar{1}1)$]; see text.

7. Reconstructions with a bigger aperture: reconstruction method II

Another simple approach to reconstruct the object wave function becomes apparent if we increase the radius of the aperture such that there is overlap between several diffraction orders and we consider the intensity variation midway in the overlap between the central disc and each diffracted disc. Fig. 9 shows simulated microdiffraction patterns from Si$\langle 110\rangle$ for two probe positions. The first thing to note is that due to the multiple interference of many discs the adjustment of lens parameters, particularly the defocus, is not as straightforward as in the case

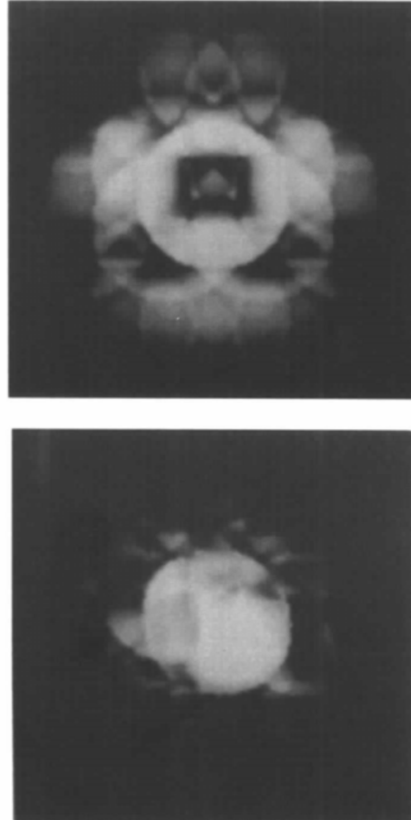

Fig. 9. Simulated microdiffraction patterns from Si$\langle 110\rangle$, using a bigger aperture, for two probe positions.

of single overlap. After Fourier transformation with respect to the probe position, we obtain interference regions of the form depicted in Fig. 10. In reconstruction method II (RMII), the amplitude and phase of each reconstructed beam are taken from the centre of the overlap between the central disc and the pertinent diffracted disc (these points are indicated by arrows in Fig. 10). The Fourier components for the ROWF are thus given by the products of beam amplitudes of the form $\Psi_{\mathbf{H}}(-\mathbf{H}/2)\Psi_0^*(\mathbf{H}/2)$, where $\mathbf{H}$ is not fixed but represents all the different reflections. This corresponds to the set of points in $G(\boldsymbol{\mu}', \boldsymbol{\rho}')$ for which $\boldsymbol{\mu}' = \boldsymbol{\rho}'/2$. [See Plamann & Rodenburg (1994) for the extension of RMII to a general imaging technique, applicable to aperiodic structures.] In the case of a strongly scattering crystal, we have to ensure that the centre of overlap between the central disc and each diffracted disc does not coincide with any part of the overlap region of any two other discs, as this would lead to mixing of the interference regions. However, as Fig. 10 indicates, we can find an aperture size for which the centres of the interference regions are not obscured.

Two properties of this scattering geometry should be noted: Firstly, a sector detector such as a quadrant detector would integrate over the interference regions in Fig. 10 such that the image contrast in reconstructions from quadrant detector images (Landauer, McCallum & Rodenburg, 1995) is expected to be rather similar to those presented here. Secondly, the Bragg condition is satisfied in the centre of the overlap of the central disc and each diffracted disc (Fig. 11). In the kinematical theory, we find

$$\Psi_{\mathbf{H}}(-\mathbf{H}/2)\Psi_0^*(\mathbf{H}/2) = i\sigma V_{\mathbf{H}}t, \tag{20}$$

which is linearly related to the structure factor of the reflection $\mathbf{H}$. No excitation error appears in (20), such that the imaginary part of the ROWF in real space is equal to the projected potential without any thickness effects. [Similarly, we can reconstruct a double-resolution image that is unaberrated and has infinite depth of focus for a general three-dimensional object, see Plamann & Rodenburg (1994).] However, here we are interested in the non-kinematical case, which we encounter in practice. Because of the vanishing excitation errors, the ROWF for RMII can be described by the POA to a better approximation than for RMI:

$$\Psi_{\text{ROWF}}(\mathbf{R}) = \exp[i\sigma V(\mathbf{R})t]. \tag{21}$$

Equation (21) includes the kinematical theory and multiple scattering up to second order. The reconstruction technique thus ensures the best possible range of applicability of the POA.

Here we present simulations for the same structure as above but choose an electron energy of 300 keV because

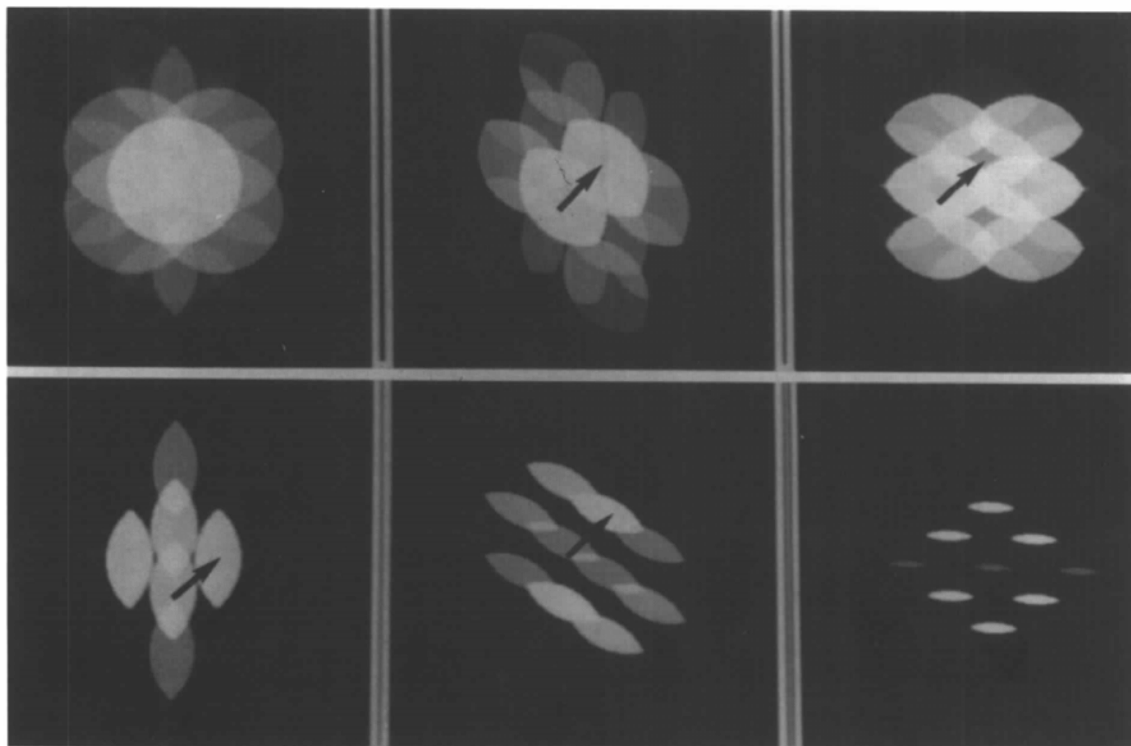

Fig. 10. Interference regions for the case of a bigger aperture; in reconstruction method II, amplitudes and phases of the reconstructed beams are taken from the centre of the overlap between the central disc and the pertinent diffracted disc, indicated by arrows.

we believe that it should be possible to perform an equivalent experiment in a 300 keV STEM, which has a sufficient degree of coherence. In Fig. 12, we present magnitude and phase of the ROWF for different thicknesses. The phase shows the projected structure for all thicknesses below 25 nm. However, for large thicknesses, the spacing of the dumbbells is slightly wrong. For small thicknesses, the imaginary part also shows the projected structure but it was found that for larger thicknesses there is an additional phase factor. Indeed, dumbbell features can be seen in the imaginary part of the ROWF after the ROWF has been multiplied by a constant phase factor (Fig. 13). (The same effect is obtained if the phase of the reconstructed central beam is not set to zero. It should be noted that the imaginary part of the ROWF for a phase offset of $-\pi/2$ is the same as the real part of the ROWF without phase offset.) The phase factor is approximately -0.3 rad per 2.9 nm thickness step. We can understand this remarkable phenomenon when we look at the variation with crystal thickness of the amplitudes and phases of the reconstructed beams (Figs. 14 and 15). The amplitude of the kinematically forbidden $\{002\}$ beams remains very small for thicknesses below 25 nm. Apart from a slight oscillation of the $\{111\}$ beams, the variation with thickness of all beam amplitudes ($\{220\}$, $\{004\}$ and $\{113\}$ are also shown) is very similar up to a thickness of about 25 nm. Similarly, the phases of the reconstructed

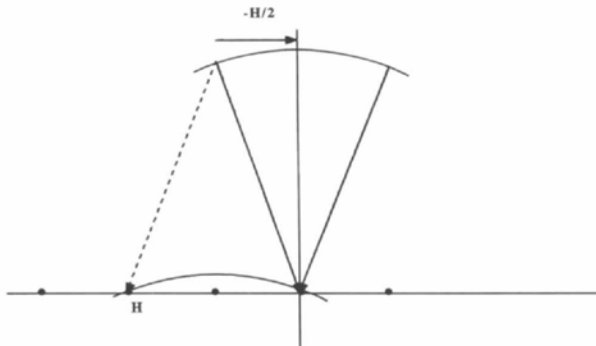

Fig. 11. Interferences between the central beam and the scattered beams in reconstruction method II occur in a symmetric scattering geometry in which the Bragg condition for each beam is satisfied.

Fig. 12. Magnitude (m) and phase (p) of the ROWF (reconstruction technique II) for different thicknesses.

beams increase at the same rate in this thickness range. It is apparent that after the multiplication of these complex beam amplitudes with a pertinent phase we obtain a function in reciprocal space that is approximately proportional to the Fourier transform of the projected potential. The POA offers a plausible explanation for this behaviour. If the crystal structure is composed of atomic columns containing the same elements, such as in the case of Si⟨110⟩, the square of the projected potential is approximately proportional to the projected potential itself, which is the basis for the Sayre equation (Sayre, 1952). By induction, this applies to all positive powers of the projected potential and we have

$$[\sigma V(\mathbf{R})]^n = \alpha^{n-1} V(\mathbf{R}), \quad (22)$$

where α is a suitable constant. In the POA, we obtain

$$\begin{aligned}\Psi(\mathbf{R}) &= \exp[i\sigma V(\mathbf{R})t] \\ &= 1 + \sum_{n=1}^{\infty}[i\sigma V(\mathbf{R})t]^n/n! \\ &= 1 + [V(\mathbf{R})/\alpha]\sum_{n=1}^{\infty}(i\alpha t)^n/n! \\ &= 1 + [V(\mathbf{R})/\alpha][\exp(i\alpha t) - 1] \\ &= 1 + (2i/\alpha)\exp(i\alpha t/2)\sin(\alpha t/2)V(\mathbf{R}). \quad (23)\end{aligned}$$

The amplitudes of the scattered beams are thus approximately given by

$$\Psi_{\mathbf{G}} = (2i/\alpha)\exp(i\alpha t/2)\sin(\alpha t/2)V_{\mathbf{G}}. \quad (24)$$

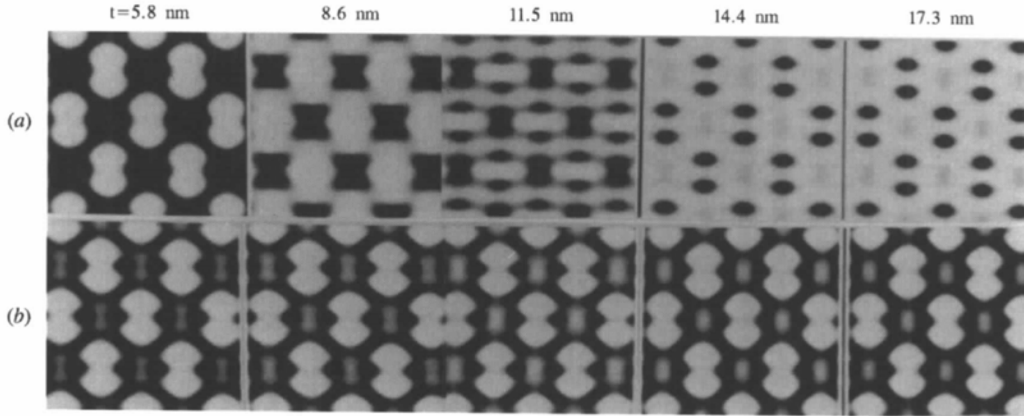

Fig. 13. Imaginary part of the ROWF: (*a*) without phase offset; (*b*) with phase offset in real space (see text).

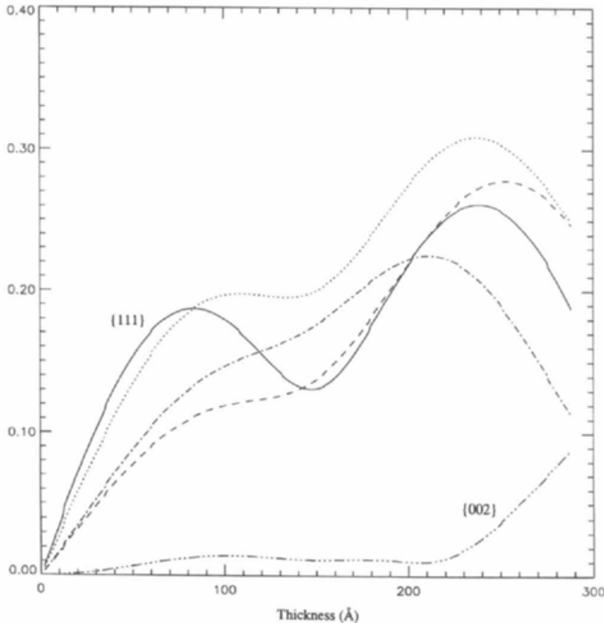

Fig. 14. The variation with crystal thickness of the amplitudes of the reconstructed beams (reconstruction method II).

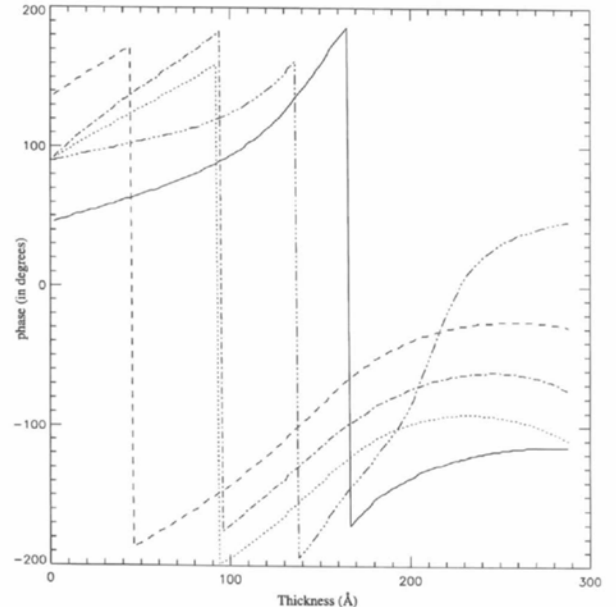

Fig. 15. The variation with crystal thickness of the phases (up to multiples of 360°) of the reconstructed beams (reconstruction method II).

We see that the POA would at least explain the observed linear increase in phase and amplitude of the reconstructed beams as a function of thickness for small thicknesses.

8. Conclusions

We have investigated the interpretation of ptychographical reconstructions from perfect crystals. In the case of a perfect crystal, the phase determination between diffraction discs relies on the interference of only two beams from opposite points in the illuminating convergent beam, whose propagation through the specimen can easily be calculated in multislice programs by applying different plane-wave illumination tilts. Two simple ptychographical reconstruction methods have been introduced and simulations have been presented using Si⟨110⟩ as an example. The first reconstruction method applies to the case of discs with single overlap. It has been shown that the reconstructed wave function corresponds to a fixed focus setting on the middle section of the specimen. For Si⟨110⟩, the projected structure can be reconstructed up to a thickness of about 10 nm. For greater thicknesses, the phases assigned to the reconstructed beams are not unique. The second reconstruction method applies to discs with multiple overlap and uses only diffraction information from the central disc. For Si⟨110⟩, the dumbbells can be reconstructed for all thicknesses smaller than 25 nm. For perfect crystals, the first reconstruction method requires only a line scan of the probe, while for the second reconstruction method a two-dimensional area scan has to be performed, and therefore considerably more data have to be recorded. A very practical simplification for the latter case could be the use of a quadrant detector, which should produce very similar image reconstructions from only four real-space images. It would be interesting to see if the silicon dumbbells could be reconstructed using quadrant detector imaging for rather thick crystals, as our simulations suggest.

TP acknowledges support from BASF. JMR is grateful to the Royal Society and the EPSRC for financial support.

References

Bird, D. M. (1989). *J. Electron. Microsc. Tech.* **13**, 77–97.

Broeckx, J., Op de Beeck, M. & Van Dyck, D. (1995). *Ultramicroscopy*, **60**, 71–80.

Cowley, J. M. & Moodie, A. F. (1962). *J. Phys. Soc. Jpn*, Suppl. B2, 86–92.

Desseaux, J., Renault, A. & Bourret, A. (1977). *Philos. Mag.* **35**, 357.

Doyle, P. A. & Turner, P. S. (1968). *Acta Cryst.* **24**, 390–397.

Fujiwara, K. (1961). *J. Phys. Soc. Jpn*, **16**, 2226–2229.

Glaisher, R. W., Spargo, A. E. C. & Smith, D. J. (1989). *Ultramicroscopy*, **27**, 19–34, 35–53.

Goodman, P. & Moodie, A. F. (1974). *Acta Cryst.* **30**, 280–290.

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

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

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

Hutchison, J. L. (1982). *Ultramicroscopy*, **9**, 191–196.

Hutchison, J. L. & Waddington, W. G. (1988). *Ultramicroscopy*, **25**, 93–95.

Izui, K., Furono, S. & Otsu, H. (1977). *J. Electron. Microsc.* **26**, 129–132.

Kirkland, A. I., Saxton, W. P., Chau, K. L., Tsuno, K. & Kawasaki, M. (1995). *Ultramicroscopy*, **57**, 355–365.

Kirkland, E. J., Loane, R. F. & Silcox, J. (1987). *Ultramicroscopy*, **23**, 77–96.

Landauer, M. N., McCallum, B. C. & Rodenburg, J. M. (1995). *Optik (Stuttgart)*, **100**, 37–46.

Lichte, H. (1992). *Ultramicroscopy*, **47**, 223–230.

McGibbon, A. J. & Pennycook, S. J. (1994). Proceedings ICEM 13, Paris, Vol. 1, pp. 477–478.

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

Nellist, P. D. & Rodenburg, J. M. (1998). *Acta Cryst.* A**54**, 49–60.

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

Pennycook, S. J. & Jesson, D. E. (1991). *Ultramicroscopy*, **37**, 14–38.

Plamann, T. & Rodenburg, J. M. (1994). *Optik (Stuttgart)*, **96**, 31–36.

Pogany, A. P. & Turner, P. S. (1968). *Acta Cryst.* A**24**, 103–109.

Rodenburg, J. M. (1989). *Ultramicroscopy*, **27**, 413–422.

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

Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

Self, P. G., O'Keefe, M. A., Buseck, P. R. & Spargo, A. E. C. (1983). *Ultramicroscopy*, **11**, 35–52.

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

Spence, J. C. H., O'Keefe, M. A. & Kolar, H. (1977). *Optik (Stuttgart)*, **49**, 307–323.

Sturkey, L. (1962). *Proc. Phys. Soc.* **80**, 321–354.

Van Dyck, D. (1983). *J. Microsc.* **132**, 31–42.

Van Dyck, D. (1985). *Advances in Electronics and Electron Physics*, Vol. 65, pp. 295–354. New York: Academic Press.

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