Image reconstruction from electron and X-ray diffraction patterns using iterative algorithms: experiment and simulation

U. Weierstall\textsuperscript{a,*}, Q. Chen\textsuperscript{a}, J.C.H. Spence\textsuperscript{a}, M.R. Howells\textsuperscript{b}, M. Isaacson\textsuperscript{c}, R.R. Panepucci\textsuperscript{c}

\textsuperscript{a} Department of Physics and Astronomy, Arizona State University, Tempe, AZ 85287-1507, USA
\textsuperscript{b} Lawrence Berkeley National Laboratory, Cyclotron Rd, Berkeley, CA 94720, USA
\textsuperscript{c} Cornell Nanofabrication Facility, Cornell University, USA

Received 19 December 2000; received in revised form 17 July 2001

Abstract

The hybrid input–output iterative algorithm, which solves the phase problem for scattering from non-periodic objects, is reviewed for application to X-ray and electron diffraction data. Desirable convex constraints, including the sign of the scattering potential for electrons, and compact support, are discussed. The cases of complex and real exit-face wavefunctions, strong and weak phase objects, various supports, and the use of coherent focussed radiation are reviewed. Reconstruction of general complex objects requires accurate knowledge of the support, which should consist of two holes or a triangle in an opaque mask. The support boundaries should be as sharp as possible. Strong phase objects without absorption can be recovered if the support consists of one hole, is accurately known and has sufficiently sharp boundaries. Real and weak phase objects with absorption can be recovered without accurate knowledge of the support area if the support boundaries are sufficiently sharp and the support consists of one or more holes. A sign constraint on the scattering potential is used to recover weak phase objects. The experimental realization of theoretically desirable support conditions is discussed. A two-stage method of finding the support for complex objects is proposed. Experimental results from applying the Gerchberg–Saxton–Fienup HiO-algorithm to coherent electron diffraction patterns are presented, using specially made e-beam lithographed support structures. Images with a resolution of about 5 nm are thus recovered from the intensities alone in coherent electron diffraction patterns from non-periodic objects. Limitations of the present experiments are identified and suggestions made for development of both X-ray and electron work. © 2002 Elsevier Science B.V. All rights reserved.

1. Introduction

Solutions to the phase problem for non-periodic objects have developed steadily over the past 30 years (see Ref. [1]\textsuperscript{1} for a review). Gerchberg and Saxton [3] first described an algorithm, which iterates between diffraction and real space to solve this problem, by introducing known information (amplitudes and/or phases) at each step. Two crucial issues—uniqueness and stagnation (the global search problem) had to be understood before the non-linear Fourier series equations (for sampled data) linking these spaces could hope to be solved. The need for compact support

\textsuperscript{1}Abstracts from a recent workshop on the topic of this paper can be found in http://www-esg.lbl.gov/esg/meetings/phasing/index.html which is also summarised in Ref. [2].
(sometimes referred to as oversampling, for an isolated object) to assist in establishing uniqueness was understood in the early 1980s [4], while the connection with the theory of projection onto convex sets has clarified the local minima problem [5]. By identifying desirable convex constraints Fienup has developed a successful hybrid input–output (HIO) algorithm, which solves the phase problem for two- and three-dimensional real objects given suitable supports [6]. Convergence is found to improve with the number of dimensions. (The support is the region of the object space in which the object is non-zero.) For real objects, this support boundary does not need to be known accurately; for complex objects we find that it does, and must correspond to a physical boundary in the object. The list of useful convex constraints, which help avoid local minima, is steadily increasing, and includes symmetry, the known sign of the electron charge density in matter, that the total charge is zero, atomicity for point atoms if their potentials do not overlap (otherwise it is non-convex), compact support, and retrieval from phase or real or imaginary parts [7].

The extension to complex objects (meaning that the exit-face wavefunction is complex, as discussed below) has also been demonstrated [8], and uniqueness established in the sense that “a multiplicity of solutions is pathologically rare” [9]. Constraints based on the sign of the real and imaginary parts of the X-ray scattering factors have also been used recently [10] to successfully reconstruct an image of lithographed letters from a soft X-ray diffraction pattern. However to date no non-periodic images have been reconstructed from either electron or X-ray diffracted intensities with better resolution than can be obtained with lenses or zone plates.

The importance of oversampling and support can readily be understood as follows. For a two-dimensional noise-free non-periodic object consisting of a set of separated sharp peaks, it has been shown that the autocorrelation function of the object can be inverted to provide a unique reconstruction of the object, thus solving the phase problem [11]. Consider a two-dimensional complex object \( f(\mathbf{r}) \) of width \( W \) with continuous Fourier Transform \( F(\mathbf{u}) \). The continuous transform \( P(\mathbf{r}) \) of \( I(\mathbf{u}) = |F(\mathbf{u})|^2 \) has width \( 2W \), being the autocorrelation or Patterson function of \( f(\mathbf{r}) \). The optimum Nyquist sampling of the diffracted intensities \( I(\mathbf{u}) \) needed to reconstruct all of \( P(\mathbf{r}) \) is therefore \( 1/2W \), not \( 1/W \). Assume \( f(\mathbf{r}) \) is bandlimited, and is optimally sampled at \( N^2 \) (complex) points. Taking \( N^2 \) samples of intensity \( I(\mathbf{u}) \) in the diffraction plane at intervals of \( 1/W = \Delta u \), however, does not allow solution of the phase problem. The known diffracted intensities \( I(\mathbf{u}) \) define \( N^2 \) non-linear Fourier Series equations linking them to \( 2N^2 \) unknown real-space complex amplitudes \( f(\mathbf{r}) \). These equations are insoluble unless the number of unknowns is reduced by requiring that at least half the pixels in the image are known (e.g. have zero value outside the support). This can be achieved (for an isolated object) by sampling the diffraction pattern more finely, thus putting the same sized object into a bigger frame of width \( 2W \), which accommodates the Patterson function. (For crystals, the connection with density modification is evident, since this requires that a molecule occupy less than half the volume of the cell.) For a square object the pattern must be oversampled by \( 2^{1/2} \) in each direction, so that, in the previous example, one then has \( 2N^2 \) equations and diffracted intensities and a similar number of unknowns. While this support requirement does not establish uniqueness (some equations may not be independent), the method of analytic continuation has been used [9] to show that, under these conditions, multiple solutions are very rare. Given uniqueness and convex constraints (which guarantee the absence of local minima [5], an algorithm, which iterates between two constrained sets (finding the most similar member of each at each iteration) can then be expected to converge to a unique solution [5]. In our work, convex constraints (positivity, compact support) are applied to the real-space set of data, but known amplitudes in Fourier space constitute a non-convex constraint [7], therefore convergence of the iterative algorithm is not assured. Nevertheless, the use of “feedback” in the HIO-algorithm greatly improves convergence and reduces stagnation in practice. The idea of oversampling has its origins in work by Sayre [12], and has been extended to crystals, using either local (pseudo)
symmetry [13], or solvent flattening, in which the water jacket around a molecule in a crystal is taken to be a form of known support function. The disjoint support around each atom, which results from the fact that atomic charge densities are sharply peaked, is the atomicity constraint on which direct methods are based. The use of projection onto convex sets to invert multiple-scattering electron diffraction data is described in Ref. [14].

In this paper we apply the HIO-algorithm to experimental electron diffraction patterns. Electron-beam lithography was used to make physical supports (holes in an opaque membrane), within which an unknown sample can be placed. We evaluate the use of the phase-grating approximation for a weak and strong phase object, in which a unknown sample can be placed. We now do this for the case of visible light, electron beams and X-rays. Both refractive and dissipative (inelastic) processes may occur. In each case one must consider whether an iconal or projection approximation may be made, and the question of whether three-dimensional (tomographic) information may be extracted.

The simplest case for each radiation is that in which the projection approximation holds, in which case the image $\Psi(r)$ may be treated as a simple projection of some property of the sample, taken in the beam direction. Then, if a transmission sample of thickness $t$ is illuminated by a plane-wave,

$$\psi(r) = \exp(-2\pi i \Delta n_p(r)/\lambda).$$

Here $\Delta n_p$ is proportional to the complex refractive index of the sample for the radiation concerned. We see that a real (**"mask-like"**) object can only be obtained if the real part of $\Delta n_p$ is independent of $r$, and all structural information is contained in the imaginary part. These experimental conditions (pure **"absorption"** contrast) can only be obtained at relatively low spatial resolution (under incoherent conditions) for both electrons and X-rays.

For X-rays

$$\Delta n_p(r) = \int_0^t (\delta(R) - i\beta(R)) \, dz,$$

where $\delta$ is a positive quantity, and $R$ a three-dimensional vector [16]. In terms of mean values, the complex index of refraction for X-rays is $n = 1-\Delta n = (1-\delta)+i\beta$, where $\delta$ describes refraction and $\beta$ absorption (mainly the photo-electric effect, important at absorption edges). The linear absorption coefficient is $\mu = 4\pi \beta / \lambda$. The dependence of $\Delta n(r)$ on the real and imaginary parts of the atomic scattering factors $f$ and $f'$ is given by $\delta = (r_e \lambda^2 / 2\pi) n_a f$, and $\beta = (r_e \lambda^2 / 2\pi) n_a f'$, with $n_a$ atoms per unit volume and $r_e$ the classical electron radius. Away from absorption edges, the electronic charge density (excluding the nuclear contribution) is

$$\rho(R) = \left( \frac{2\pi}{r_e \lambda^2} \right) \delta(R).$$

If small bonding effects are ignored, $\rho(R)$ is obtainable from tabulated X-ray scattering factors for neutral atoms.

Since $\delta$ is about $10^{-3}$ at 6 kV for light materials, a thickness of about 0.3 $\mu$m of sapphire is needed to obtain a phase shift of $\pi/2$, allowing a first-order expansion of Eq. (1). Then the diffracted amplitudes are simply proportional to the Fourier transform of the projected charge density of the object.
For an electron beam of kinetic energy $eV_0$:

$$\Delta\eta_p(\mathbf{r}) = \int_0^t \frac{V_c(\mathbf{R})}{2V_0} \, dz,$$

(4)

where $V_c(\mathbf{R})$ is the complex “Optical” potential for high energy electrons [17] and the mean refractive index for electrons is $n = 1 + \Delta n$. The real part of this, for high energies, is the positive electrostatic or Coulomb potential, also obtainable from X-ray scattering factors using Poisson’s equation [12]. The imaginary part (typically about a tenth of the real part) accounts for depletion of the elastic scattering factors using Poisson’s equation [12]. The imaginary part (typically about a tenth of the real part) accounts for depletion of the elastic wavefield by inelastic scattering events such as plasmon, inner-shell and phonon scattering. The average value of $\text{Re}(V_c)$ is about $12\text{eV}$ for light materials, and is positive. Both real and imaginary parts of the potential $V_c(\mathbf{R})$ are positive if the mean inner potential is included, since electron beams are attracted predominantly to the positive nuclei [18]. Electron diffraction in the transmission geometry is not, however, sensitive to the mean potential, which produces only a constant phase shift. The mean inner potential can only be meaningfully defined for a finite object with zero total charge. Although the potential due to an (unphysical) isolated negative ion may have small negative excursions, the mean value depends on the volume assumed, and, in a real crystal, any long-range ionic potential is also screened [19].

In summary, if inelastic processes are neglected (away from absorption edges), at high energies, transmission samples of thickness $t$ are phase objects for electrons and X-rays, for which the refractive index is proportional to the electron density for X-rays and to the total electrostatic potential, including the nuclear contribution, for electrons. Poisson’s equation relates these. The magnitudes of these quantities are such that a first-order expansion of Eq. (1) (weak phase object approximation) is justified $(2\pi n_0(r)/\lambda < \pi/2)$ if $t < 20\text{nm}$ for electrons (light elements, $V_0 = 200\text{kV}$) but $t < 0.3\mu\text{m}$ for X-rays (light inorganic material, $6\text{kV}$). Higher order terms in the expansion of Eq. (1) correspond to the multiple-scattering terms of the Born series in a “flat Ewald sphere” approximation.

This “flat sphere approximation”, on which Eq. (1) is based (the projection or iconal approximation) depends on the ratio of wavelength to smallest detail $d$ of interest (with spatial frequency $u = 1/d$), and on the thickness of the sample. Scattering kinematics restrict elastic scattering to regions of reciprocal space near the energy and momentum-conserving Ewald sphere of radius $1/\lambda$. The projection approximation holds if the “excitation error” distance $S_u \approx \lambda u^2/2$ from this sphere onto a plane in reciprocal space normal to the beam (passing through the origin) is less than either $1/t$ or $1/\xi_u$, whichever is the smallest. ($\xi_u$ is a multiple-scattering extinction distance for spatial frequency $u$. Thus samples never look thicker than $\xi_u$, to diffracting radiation.) Hence we require

$$\frac{\lambda u^2}{2} < \frac{1}{t} \text{ or } \frac{1}{\xi_u} \quad \text{or}$$

$$\left(\frac{\lambda \cdot t}{2}\right)^{1/2} < d$$

For visible light, $\Delta\eta_p(r)$ has a similar interpretation as for X-rays (with $n$ given by the square root of the complex dielectric constant), however $n > 1$ Thus visible light and electrons ($n > 1$) are bent toward the normal on entering a denser medium, while X-ray undergo total external reflection, with $n < 1$.

In summary, if inelastic processes are neglected (away from absorption edges), at high energies, transmission samples of thickness $t$ are phase objects for electrons and X-rays, for which the refractive index is proportional to the electron density for X-rays and to the total electrostatic potential, including the nuclear contribution, for electrons. Poisson’s equation relates these. The magnitudes of these quantities are such that a first-order expansion of Eq. (1) (weak phase object approximation) is justified $(2\pi n_0(r)/\lambda < \pi/2)$ if $t < 20\text{nm}$ for electrons (light elements, $V_0 = 200\text{kV}$) but $t < 0.3\mu\text{m}$ for X-rays (light inorganic material, $6\text{kV}$). Higher order terms in the expansion of Eq. (1) correspond to the multiple-scattering terms of the Born series in a “flat Ewald sphere” approximation.

This “flat sphere approximation”, on which Eq. (1) is based (the projection or iconal approximation) depends on the ratio of wavelength to smallest detail $d$ of interest (with spatial frequency $u = 1/d$), and on the thickness of the sample. Scattering kinematics restrict elastic scattering to regions of reciprocal space near the energy and momentum-conserving Ewald sphere of radius $1/\lambda$. The projection approximation holds if the “excitation error” distance $S_u \approx \lambda u^2/2$ from this sphere onto a plane in reciprocal space normal to the beam (passing through the origin) is less than either $1/t$ or $1/\xi_u$, whichever is the smallest. ($\xi_u$ is a multiple-scattering extinction distance for spatial frequency $u$. Thus samples never look thicker than $\xi_u$, to diffracting radiation.) Hence we require

$$\frac{\lambda u^2}{2} < \frac{1}{t} \text{ or } \frac{1}{\xi_u} \quad \text{or}$$

$$\left(\frac{\lambda \cdot t}{2}\right)^{1/2} < d$$

For visible light, $\Delta\eta_p(r)$ has a similar interpretation as for X-rays (with $n$ given by the square root of the complex dielectric constant), however $n > 1$ Thus visible light and electrons ($n > 1$) are bent toward the normal on entering a denser medium, while X-ray undergo total external reflection, with $n < 1$.

In summary, if inelastic processes are neglected (away from absorption edges), at high energies, transmission samples of thickness $t$ are phase objects for electrons and X-rays, for which the refractive index is proportional to the electron density for X-rays and to the total electrostatic potential, including the nuclear contribution, for electrons. Poisson’s equation relates these. The magnitudes of these quantities are such that a first-order expansion of Eq. (1) (weak phase object approximation) is justified $(2\pi n_0(r)/\lambda < \pi/2)$ if $t < 20\text{nm}$ for electrons (light elements, $V_0 = 200\text{kV}$) but $t < 0.3\mu\text{m}$ for X-rays (light inorganic material, $6\text{kV}$). Higher order terms in the expansion of Eq. (1) correspond to the multiple-scattering terms of the Born series in a “flat Ewald sphere” approximation.
for the validity of Eq. (1). Since the width of the first Fresnel fringe due to propagation over distance \( t \) is approximately \( w = (\lambda t/2)^{1/2} \), this condition requires that the spreading of the wavefield due to free-space propagation over a distance equal to the thickness of the sample be small compared to the resolution required.

The failure of Eq. (1) may require either single or multiple scattering treatments, depending on the strength of the interaction and sample thickness. We note that Eq. (1) is an exact solution which sums the Born series, including all multiple scattering effects, in the limit of vanishing wavelength. For X-ray tomography, use of Eq. (1) greatly simplifies the collection of data for three-dimensional reconstruction. However, the use of a curved Ewald sphere does not prevent three-dimensional reconstruction in the single scattering approximation [21], since data may be collected for various object orientations and, in each case assigned to points on the sphere, until reciprocal space is filled. Use of the Fienup algorithm with three-dimensional Fourier Transform iterations can then take advantage of the improved convergence in three dimensions [22].

3. Phase retrieval from electron diffraction patterns

In the following we investigate the possibility of exit wave reconstruction in transmission electron microscopy (TEM) from the modulus of the diffracted wave amplitude. In the TEM, a Fraunhofer diffraction pattern of a selected sample area is recorded by imaging the back focal plane of the objective lens onto the detector using diffraction lenses. For coherent imaging, if we ignore the aberrations of these diffraction lenses (taken to have unit magnification), then the wave amplitude \( \phi(u) \) at the detector is given by the product of the Fourier transform of the object exit wave \( \Psi(r) \) and the transfer function of the objective lens \( T(u) \)

\[
\Phi(u) = FT(\Psi(r))T(u),
\]

where \( T(u) \) is given as

\[
T(u) = A(u) \exp(i\chi(u))
\]

with the aperture function \( A(u) \) and the wave aberration of the objective lens

\[
\chi(u) = \pi\Delta\lambda|u|^2 + \frac{\pi}{2} C_2\lambda^2|u|^4.
\]

The recorded intensity of the diffraction pattern is given as

\[
I(u) = \Phi(u)\Phi(u)^* = |FT(\Psi(r))|^2,
\]

where the phase factor due to objective lens aberrations cancels out and the aperture function has been omitted since it represents only a bandwidth limit. If in the projection approximation the effect of a thin object on the incident wave can be represented by multiplication by a transmission function \( q(r) \), an incident wave of amplitude unity gives an exit wave equal to the transmission function.

Only the magnitude of the Fourier transform of \( \Psi(r) \) can be measured, the phase is lost, which constitutes the phase problem. If the phase of the Fourier transform \( FT(\Psi(r)) \) could be recovered, than the aberration-free complex exit wave \( \Psi(r) \) could be reconstructed. Aberrations of the diffraction lenses which magnify the diffraction pattern (astigmatism, distortion, etc.) could alter the intensity distribution \( I(u) \) and so could make image recovery difficult.

For non-periodic images the phase problem can be solved, if the image has a finite support i.e. the image is non-zero only within a finite region of image space (which is equivalent to oversampling Fourier space) and if the image is real and non-negative [6]. Most phase-recovery algorithms for far-field scattering are based on Fienup’s modification [6] of the Gerchberg–Saxton algorithm [3]. This algorithm is one member of a large number of iterative solution methods used to solve problems like differential equations and matrix inversion [23].

Since the exit wave in TEM is in general complex, phase-recovery by the above-mentioned algorithm is much more difficult, since the powerful constraint of non-negativity is lost. However, if the support of the object is sufficiently well known (is sufficiently tight), has sufficiently sharp boundaries, and is one of several types, it is still possible to recover the Fourier phase for complex objects [8]. In the following, it is assumed, that
\[ FT(\Psi(x,y)) \] has been measured, and that the support \( S(x,y) \) is known. \( S(x,y) \) is the region outside which the image is known to be zero. The iterations start with an initial estimate \( G_0(u,v) = FT(\Psi(x,y)) \exp(i\theta_0(u,v)) \) of the spectrum. \( \theta_0(u,v) \) is chosen to be an array of independent pseudo-random real numbers distributed between 0 and 2\( \pi \). The iterative Fourier-transform algorithm consists of the following steps (with subscript \( k \) labeling quantities at the \( k \)th iteration):

1. Inverse Fourier transform \( \tilde{G}_k(u,v) \) to obtain the image \( \tilde{g}_k(x,y) \):

\[
g_{k+1}(x,y) = \begin{cases} 
\tilde{g}_k(x,y) & \text{if } (x,y) \in S(x,y), \\
g_k(x,y) - \beta \tilde{g}_k(x,y) & \text{if } (x,y) \notin S(x,y).
\end{cases}
\]  

(11)

This constitutes the hybrid input–output (HIO) version of the algorithm. \( \beta \) is a constant chosen between 0.5 and 1. In the error-reduction (ER) version of the algorithm this step is replaced by:

\[
g_{k+1}(x,y) = \begin{cases} 
\tilde{g}_k(x,y) & \text{if } (x,y) \in S(x,y), \\
0 & \text{if } (x,y) \notin S(x,y).
\end{cases}
\]  

(12)

3. Fourier transform \( g_{k+1}(x,y) \) to obtain \( G_{k+1}(u,v) \);

4. Define new Fourier domain function \( \tilde{G}_{k+1}(u,v) \) using the known Fourier modulus \( |FT(\Psi(x,y))| \) with the computed phase: \( \tilde{G}_{k+1}(u,v) = |FT(\Psi(x,y))| \exp(i\theta_{k+1}(u,v)) \);

5. Go to step 1 with \( k \) replaced by \( (k+1) \).

To monitor the progress of the algorithm, the image space error metric \( e_k \) is calculated during each iteration:

\[
e_k = \frac{\sum_{(x,y) \notin S} |\tilde{g}_k(x,y)|^2 }{\sum_{(x,y) \in S} |\tilde{g}_k(x,y)|^2}.
\]  

(13)

\( e_k \) is the amount by which the reconstructed image violates the image-space constraints. Initially we used a combination of ER and HIO algorithm as described by Fienup [8] with 20 ER and 50 HIO cycles with \( \beta = 0.7 \) repeated until the error \( e_k \) drops below a certain level. Subsequent trails with the HIO algorithm alone confirmed that it performs as well as the combination of ER and HIO. Other authors have suggested, that combining the HIO with the ER algorithm actually reduces the effectiveness of the HIO algorithm and its ability to emerge from a local minima [24,25]. Therefore for all calculations shown here we used only the HIO algorithm.

When experimental data are analyzed, the physical support in the object may not be known accurately. (In our case this support is the hole in our opaque mask.) To avoid confusion we call the support present in the experiment the “physical support” and the support used in the iterations the “computational support”. The term “loose support” describes in the following the situation where the computational support is bigger than the physical support. “Tight support” means the physical support is the same as the experimental support.

4. Simulations

4.1. The complex object

All our calculations have been done using the script language of Digital Micrograph Software from Gatan, which uses fast Fourier transforms. According to Fienup [8], reconstruction of a complex-valued object from the modulus of its Fourier transform is possible for certain shapes of the support function. These include supports having sufficiently separated parts, or one triangle. Simple symmetric supports such as circles and squares do not work well. Therefore our first simulated object consists of two irregular shaped holes in an opaque plane filled with a 10 nm thick Si(100) crystal (see Fig. 1a). The complex exit wave for plane wave illumination with 200 keV electrons was calculated using a multislice program for dynamical electron scattering [26]. The modulus of the Fourier transform (Fig. 1b) from this object is calculated and used together with random phases as the input for the HIO-algorithm. The support is assumed to be known exactly (physical support=computational support) and is
used in the iteration as the image space constraint. No other image space constraint is used. After about 100 iterations the complex exit wave is reconstructed almost perfectly as shown in Fig. 1d. Fig. 2 shows the image space error $e_k$ as a function of iteration number—when this is zero the image intensity equals its known value of zero everywhere outside the support. This and all following simulations and experimental reconstructions have been repeated many times with different starting phases, since the problem is non-convex and therefore conclusions cannot be drawn from one trial. Although the set of all diffraction patterns with known modulus is non-convex, and therefore the overall problem is one of projection onto non-convex sets [5], for the problems shown here the HIO-algorithm converged always to the same minima despite different starting phases.

Since the boundary of an experimental physical support will not be as sharp as the simulated support edge in Fig. 1, simulations have been done with blurred support edges, as shown in Fig. 3. If
the physical support edge is blurred, and the computational support has the same shape and sharp edges and is larger enough to contain the blurred part of the physical support, the algorithm still converges, but much more iterations are needed. Convergence time also depends on the amount of fine detail in the image. If the diffraction pattern contains only low frequency information and the physical support is blurred, convergence is very fast. If on the other hand the diffraction pattern contains high frequency information and the physical support is blurred (Fig. 3), the error falls slowly and more than 1000 iterations can be necessary to reconstruct the image (Fig. 4). This is consistent with the observation [27] that the HIO-algorithm usually reconstructs the phase of $G(u, v)$ outwards from the middle of Fourier space. This implies, that only low spatial frequency information is recovered during early iterations.

Because both images and diffraction patterns are readily observed in electron microscopes, so that an image could, for example, be used to obtain the support function, phase recovery for TEM diffraction patterns should be easier than for X-ray patterns. (In this work we have not used...
TEM images to obtain supports.) The Gerchberg–Saxton algorithm makes use of both image and diffraction pattern intensities to reconstruct the complex image wave, which is distorted by the lens aberrations. Since our objective is to reconstruct the unaberrated object exit wave, we cannot use the image intensity in the iteration, but we can make use of the fact, that the shape of the physical support is known from the TEM image. Nevertheless the exact magnification and orientation of the support corresponding to the recorded diffraction pattern is not known. One way of finding the correct size and orientation of the support is to compute the autocorrelation \( FT^{-1}(|FT(h(r))|^2) \) from the measured diffraction pattern and adjust the TEM image of the support so that its autocorrelation resembles the experimental one in size and orientation. Another way is to try to reconstruct the image with a computational support, which is large enough to contain both physical support holes (the size can be determined from the size of the diffraction pattern autocorrelation). During this trial, a modulus constraint imposed in image space then makes reconstruction easier and the physical support does not have to be known very accurately, as shown in Fig. 8, where a much bigger computational triangular support has been used. The reconstructed image is indistinguishable from the initial image except that it is inverted.

Since \( g(x,y) \), \( g^*(-x-a_1,-y-a_2) \exp(i\theta) \) and \( g(x-a_1,y-a_2) \exp(i\theta) \) all have the same Fourier modulus, this ambiguity cannot be resolved. In our experience reconstruction with a much bigger computational support works only in cases where the physical support has sharp edges. As soon as the edges are blurred a tighter support is needed for the algorithm to converge.

4.3. The phase object

4.3.1. Strong-phase object

If the projection approximation holds (see Eq. (6)), the transmission function of the specimen inside the support holes, the support holes are not very well reconstructed with this method, as shown in Fig. 6.

Coherent plane-wave illumination in a TEM produces low current on the sample. To increase the intensity of the diffraction pattern, illumination with a focused coherent probe could be used. The convergent spherical wave incident on the sample has then a quadratic phase factor, which adds to the sample phase shift. The HIO-algorithm will recover the sum of those phases. Fig. 7 shows a calculation with a simulated probe and blurred support edges. We note that any use of such a probe formed by a lens thus converts a real object into the more difficult case of a complex object.

Simulations show also that the area outside the support does not need to be zero (opaque support) or constant, but can have any values as long as these values are known.

4.2. The real and positive object

Although the electron exit wave leaving a sample is in general complex, both the X-ray and electron cases are equivalent to real objects in the weak phase-object approximation. Simulations have therefore been done for positive real images. The additional modulus constraint imposed in image space then makes reconstruction easier and the physical support does not have to be known very accurately, as shown in Fig. 8, where a much bigger computational triangular support has been used. The reconstructed image is indistinguishable from the initial image except that it is inverted.

Since \( g(x,y) \), \( g^*(-x-a_1,-y-a_2) \exp(i\theta) \) and \( g(x-a_1,y-a_2) \exp(i\theta) \) all have the same Fourier modulus, this ambiguity cannot be resolved. In our experience reconstruction with a much bigger computational support works only in cases where the physical support has sharp edges. As soon as the edges are blurred a tighter support is needed for the algorithm to converge.
is given from (1) as

$$q(x, y) = \exp(-i\sigma \phi(x, y))$$

(14)

(and similar for the X-ray case) with the projected specimen potential $\phi(x, y) = \int V_c(R) \, dz$ and $\sigma = \pi/\lambda V_0$. Neglecting the imaginary part of the optical potential $V_c(R)$ which accounts for inelastic scattering and is small for a thin sample, the transmission function (14) is that of a pure phase object.

Although a pure phase object has still a complex transmission function, its reconstruction from the modulus of its Fourier transform should be easier than for the general complex object, because it is determined by a single number, the phase shift $\sigma \phi(x, y)$. To test this idea, step 2 of the HIO-algorithm was modified and an additional image

Fig. 5. Finding the size and orientation of the physical support from the diffraction pattern modulus for a weak phase-amplitude object (phase shift $\phi = \pi/2$, transparency $0.9 < t < 1$): The image is reconstructed with a larger triangular computational support. (a) Modulus of complex exit wave with support consisting of two irregular shaped holes. (b) Fourier transform of (a), modulus used as input for the HIO-algorithm. (c) Triangular computational support (loose support) used together with modulus constraint in image space for reconstruction. (d) Real result after 100 HIO-iterations. Since the original image is complex and a modulus constraint was used, the algorithm did not converge, but the image reconstruction is good enough to see the shape of the physical support holes and define a tighter support for complex reconstruction. If no modulus constraint is used the two holes are not recovered.
space constraint was introduced: during each iteration, the modulus of $g_k(x, y)$ inside $S$ is set to 1 and the phase is preserved:

$$g_{k+1}(x, y) = \begin{cases} 
\text{phase}(\tilde{g}_k(x, y)) & \text{if } (x, y) \in S(x, y), \\
g_k(x, y) - \beta \tilde{g}_k(x, y) & \text{if } (x, y) \notin S(x, y). 
\end{cases} \quad (15)$$

Fig. 9 shows a simulation with this modified algorithm for a pure phase object. The TEM image intensity from MgO crystals was modified in the computer to create a pure phase object: the intensities were converted to phase shifts between $-\pi$ and $\pi$. Fig. 9d shows the reconstruction of this object after 300 iterations. The computational support used is identical to the physical support and is shown in Fig. 9c. Fig. 10 shows the image space error for this reconstruction.

These simulations show that for a strong-phase object the complex exit wave can be reconstructed with a tight support consisting of a single hole, unlike the general complex object. Therefore it might be possible to recover images of small biological molecules (phase objects) spanning a
hole in an otherwise opaque film, from their diffraction pattern. For the general complex object with amplitude and phase the support needs to consist of two holes or a triangle. With a loose support the phase object algorithm did not converge to a solution.

4.3.2. Weak-phase object

For sufficiently thin samples, perhaps containing only light atoms, it is possible to make a first order expansion of Eq. (1) for either X-rays or electrons. For the X-ray case this expansion becomes possible at high X-ray energies for samples hundreds of micrometer thick, or tens of nanometers thick for soft-X-rays (<1 kV). For the electron case (Eq. (14) with $\sigma(x, y) < \pi/2$) we then have

$$q(x, y) = 1 - i\sigma\phi(x, y)$$

which is known as the weak-phase object approximation (WPOA). In this approximation the real part of the exit wave is 1 and the imaginary part
\[ \sigma \varphi(x, y) \text{ is positive since } \text{Re}(V_c) > 0. \]

These conditions can be used as image plane constraints. With a loose support constraint (computational support is bigger than physical support), the real part of \( \hat{g}_k(x, y) \) cannot be set to 1 during the iterations, because parts of the image outside the physical support which are still inside the computational support have \( \text{Re}(\hat{g}(x, y)) = 0. \) Therefore the image space constraints used are positivity of both the real and imaginary part inside the computational support. This includes objects having additional amplitude modulations (weak-phase amplitude objects). A simulation with the modified HIO-algorithm is shown in Fig. 11. The fiber inside the physical support is a weak phase object with \( 0 < \sigma \varphi(x, y) < \pi/4. \) Reconstruction with positivity constraint on real and imaginary part and with a larger physical support is shown in Fig. 11c. The reconstructed fiber is indistinguishable from the initial image. For comparison the algorithm was started without positivity constraint on real and imaginary part and with the same computational support, which is shown in Fig. 11d, and did not converge to a solution (see error Fig. 12).

The weak phase object algorithm did not converge with blurred physical support edges and
a loose triangle support. We note that weak phase objects and real objects with a two or more hole support are easier to reconstruct (faster convergence) than with a one hole support. No knowledge of the support shape is necessary, if a larger computational support is used, which contains all physical support holes.

4.4. Simulations for X-ray objects

Successful reconstructions of prepared lithographed objects with simple transmission functions from soft X-ray diffraction patterns have recently been described [10], and similar reconstructions using weak phase objects and hard X-rays are in progress. Experiments using random pinhole arrays as possible holographic reference objects for X-ray Fourier holography are under way, as reported elsewhere [21]. The pinholes were produced by etching damage tracks made in opaque mica plates with high energy particles [28]. Simulations have been done to examine if the random pinhole array could be reconstructed from its soft-X-ray diffraction pattern. If so, these holes

Fig. 9. Reconstruction of strong pure phase object from the modulus of its Fourier transform is possible with one support area and a tight computational support constraint: (a) Pure phase object with phase between \(-\pi\) and \(\pi\). The object was generated in the computer from a TEM image of MgO-crystals, the image intensities were converted to phases. (b) Fourier transform of (a), modulus used as input for the HIO-algorithm. (c) Computational support used in iteration together with phase object constraint in image space. (d) Reconstructed phase image after 300 iterations.
could later be filled with structures of interest for reconstruction. Because the pinholes have to be illuminated coherently by the X-ray beam in order to reconstruct them by phase recovery, the area of the pinholes has to be limited to the coherence width of the beam by a 1\(\mu\)m aperture. This aperture will not be exactly in the plane of the pinholes, therefore the wave front at the pinholes will have undergone Fresnel diffraction at the aperture and will be complex. For the simulations a mask with a random array of holes has been generated in the computer and multiplied by a wave amplitude which has been Fresnel propagated (Fig. 13b) a distance of 200nm downstream beyond a 1\(\mu\)m aperture (X-ray wave length 2.5nm). The pinholes are assumed to be 10nm in diameter and are idealized as having no thickness (no diffraction effects inside the holes). The dark area between the holes has a transmission of \(10^{-5}\) to account for some X-ray transmission through the mica. The modulus of the complex wave amplitude behind the pinholes is shown in Fig. 13a. The modulus of the Fourier transform of this image, which is a speckled Airy’s disc (Fig. 13c), was used as the input for the HIO-algorithm. A triangular computational support with zero transmission outside the triangle was used. The triangle is large enough to contain the physical support consisting of the defocused circular 1\(\mu\)m aperture (Fig. 13b). The image reconstruction after 150 iterations of HIO algorithm is shown in Fig. 13d. Apart from being inverted, the pinholes have been reconstructed perfectly, because they constitute a real binary object, whereas the phase of the complex wave diffracted by the aperture and partly transmitted through the pinholes and the mica has not been reconstructed faithfully. This shows, that for this special object, reconstruction of the features of interest (pinholes) is possible, although the object exit wave is complex and the physical support consists of only one area and is circular. The convergence of the algorithm seems to depend on the shape of the larger computational support, i.e. if a square computational support is used rather than a triangle during the iterations the algorithm fails to reconstruct all the pinholes. Instead, only one pinhole is reconstructed. Soft X-ray imaging of weak phase and amplitude objects filling a hole would be expected to reconstruct correctly if positivity constraints were applied to them subsequently.

For X-ray experiments the transverse coherence of the beam \(X_c\), determined by the physical X-ray source size, must exceed the total width \(W\) of the object and its surround. Since the coherent power is proportional to \(\lambda^3\) [29] the spatial coherence becomes difficult to achieve for hard X-rays except for the smallest object. In contrast, one has the simplification that for hard X-rays most objects are pure real at energies far from absorption edges. Since at the unapertured diffraction limit (\(\Theta = 90^\circ\)) the resolution is approximately equal to the wavelength and about two pixels are required per resolution element, a total of about \((4X_c/\lambda)^2\) image pixels would be needed for a coherence width \(X_c\) and oversampling factor 2.

The temporal coherence length \(L_c\) is also important. For a field of view \(W\) (where the first oversampling point occurs at \(1/W\)) and finest (bandlimited) object spatial frequency \(d^{-1}\), the optical path difference between points on opposite sides of the object and a distant detector point is \(W \sin \Theta = W \lambda/d\), which should not exceed \(L_c = \lambda E/\Delta E\). Hence the fractional energy spread allow-able in the beam to record spatial frequency \(d^{-1}\) is \(E/\Delta E > W/d = N\), where \(d\) is the sampling interval in the object and \(N\) the number of pixels.
needed to sample the object space in the HIO algorithm. Thus the requirement on longitudinal coherence is
\[ E/\Delta E > N. \] (17)

4.5. Periodic objects

All our simulations have been performed using 512 \times 512 element arrays which sample the object space and represent it by a Fourier series. This series generates a periodic continuation of the density inside the array. Hence all our results can be applied to crystals, in which the unit cell corresponds to the outer boundaries of the real-space computational array, and the first sampling point in reciprocal space corresponds to the first order Bragg reflection. For hard X-ray diffraction from crystals we may treat the objects as pure real if absorption edges are avoided or produce weak effects.

It follows that the phase problem for X-ray diffraction from crystals may be solved using the HIO algorithm if the molecule within each cell
occupies a fraction $\sigma^{-1}$ of less than half the volume of the cell (as is often the case in protein crystallography, where molecules are surrounded by a solvent jacket), if the molecular shape is roughly known, and if the sign constraint is used. Since the objects are weak phase objects, an approximate knowledge of the support, as might be obtained from a Patterson function, should suffice. (With $\sigma = 2$, the Patterson provides the true autocorrelation function of the molecule, rather than that of the crystal). This approach bears a close similarity to the method of density modification or solvent flattening [30], however that method also uses statistical estimates of the sum of three phases based on the theory of Direct Methods. The HIO algorithm operates in a very similar way to the Shake and Bake algorithm [31], which iterates between real and reciprocal space, applying Direct methods in one space and fragment identification in the other, in addition to other optimizations. In summary, if low resolution techniques such as TEM or AFM imaging were able to provide images of the approximate shape of the molecules within a crystal (for which $\sigma = 2$), the phase problem could then be solved using HIO without recourse to Direct Methods. Alternatively, an estimate of the support could be based on the Patterson function.

5. Experiments

Since reconstruction of general complex objects has been shown to be possible in simulations using a support consisting of two separated parts, we have attempted to implement this scheme experimentally in coherent TEM, using a special support mask made by electron beam lithography.

5.1. Support preparation

According to Feinup [8], the most favorable support functions are triangles with very sharp edges and supports consisting of two holes spanned by the coherence width of the beam, with one or both holes containing an unknown object. The area outside the support needs to be completely opaque. The other alternative, to avoid the need for two holes, would be to use a transparent film and deposit particles (e.g. MgO, graphite) on it. Only the central pixel value in the Fourier transform modulus, which is the average of the image intensity, is than changed. By scaling the central pixel of the diffraction pattern, the particle images could still be recovered with a loose triangular computational support, assuming zero intensity outside this area, if the object is real or a weak phase object. Experimentally, the intensity of the center pixel of the diffraction pattern is smeared out over several pixels, which could make scaling difficult. Since no material is completely transparent for electrons, ultrathin carbon film of uniform thickness may the best choice to support the objects, but even the thinnest carbon film causes some non-uniform phase shift. This would result in a phase contribution to the diffraction pattern from an area outside the computational support and would therefore prevent convergence. Therefore in our first experiments we chose to make an opaque support consisting of two holes.

First we tried to use holey carbon film, which was coated with 200 nm of gold to make it opaque for electrons [32]. The field-emission electron beam energy was reduced to 40 keV to make the gold film opaque. The size and distribution of the holes was not found to be controllable by this method, therefore support films were fabricated by electron beam lithography.
The holes used in this experiment were defined by 100 keV electron beam lithography on 160 nm of polymethylmethacrylate (PMMA) deposited over thin 60 nm thick Si$_3$N$_4$ membranes. The membranes were fabricated by photolithography of 2 × 2 mm windows and subsequent KOH etching of the supporting silicon wafer. After the development of the exposed PMMA in isopropanol: methylisobutylketone solution, the holes in the membranes were etched by CF$_4$: H$_2$ reactive ion etching. Following the removal of the PMMA resist, the samples were loaded on an electron beam evaporator and 5 nm of Ti followed by 75 nm of Au were deposited on each side of the membrane to make it opaque to 40 keV electrons. (The membrane alone was transparent at 200 keV.) Due to the evaporation conditions, the size of the holes was reduced to between 10 and 25 nm.

Fig. 14 shows a low magnification TEM-image of the membranes. The membrane has pairs of approximately elliptically shaped holes, whose short diagonal ranges from 15 to 50 nm, separated...
by a distance equal to the sum of the short diagonals of the two holes. The distance between the hole pairs is 1 μm. The hole size ranges from 10 to 25 nm.

A Philips CM200 electron microscope with field emission gun (FEG) was used to record the experimental electron diffraction patterns. The Schottky FEG with source size 0.5 μm and energy width δE = 0.5 eV provides the required coherent source. An electron energy of 40 keV was used to avoid transmission through the support.

Microscope alignment and astigmatism correction at 40 keV were obtained at a nearby area. The condenser lens was adjusted so that the central spot in the diffraction pattern without a sample was minimized. This is the condition where the incident beam is almost parallel (beam divergence = 0.02 mrad). The sample area was then moved into the beam. The sample height was adjusted to focus the holes. The selected area aperture was used to confine the diffraction pattern to an area, which included two holes. The extent of the illuminated area was checked in diffraction mode by changing the condenser lens current. The condenser lens current was then changed back to parallel beam condition and the diffraction pattern recorded.

A CCD camera with 1024 × 1024 24 μm square pixels and 14-bit dynamic range was used for recording the diffraction pattern. At the camera lengths used, this ensured adequate oversampling of the data by a factor of about 12. Since the beam current passing through the holes for parallel beam illumination is very low, the diffraction pattern intensity from the two holes is very weak. Exposure times currently are limited to about 20 s due to instabilities (lens currents of magnifying lenses, AC fields), therefore we cannot make use of the full dynamic range of the CCD-camera.

5.2. Experimental results

Fig. 15a shows the experimental diffraction pattern from two empty holes. Young’s fringes from the two holes are visible in the diffraction pattern. The Schottky FEG with source size 0.5 μm and energy width δE = 0.5 eV provides the required coherent source. An electron energy of 40 keV was used to avoid transmission through the support.

Microscope alignment and astigmatism correction at 40 keV were obtained at a nearby area. The condenser lens was adjusted so that the central spot in the diffraction pattern without a sample was minimized. This is the condition where the incident beam is almost parallel (beam divergence = 0.02 mrad). The sample area was then moved into the beam. The sample height was adjusted to focus the holes. The selected area aperture was used to confine the diffraction pattern to an area, which included two holes. The extent of the illuminated area was checked in diffraction mode by changing the condenser lens current. The condenser lens current was then changed back to parallel beam condition and the diffraction pattern recorded.

A CCD camera with 1024 × 1024 24 μm square pixels and 14-bit dynamic range was used for recording the diffraction pattern. At the camera lengths used, this ensured adequate oversampling of the data by a factor of about 12. Since the beam current passing through the holes for parallel beam illumination is very low, the diffraction pattern intensity from the two holes is very weak. Exposure times currently are limited to about 20 s due to instabilities (lens currents of magnifying lenses, AC fields), therefore we cannot make use of the full dynamic range of the CCD-camera.

5.2. Experimental results

Fig. 15a shows the experimental diffraction pattern from two empty holes. Young’s fringes from the two holes are visible in the diffraction pattern. The Schottky FEG with source size 0.5 μm and energy width δE = 0.5 eV provides the required coherent source. An electron energy of 40 keV was used to avoid transmission through the support.
frequency information from the hole edge are missing in the experimental diffraction pattern, as seen when this is compared with the Fourier transform of the TEM image of the two holes, shown in Fig. 15d. The information here extends out to much higher spatial frequencies. Possible reasons could be insufficient coherence of the incident electron wave, and the low count rate in the diffraction pattern. To simulate the effect of limited spatial coherence on the diffraction pattern, the TEM image of two holes (Fig. 15c) has been multiplied by a phase factor \( \exp\left(2\pi i k_\theta \cdot \mathbf{r}\right) \) with \( |k_\theta| = \theta/\lambda \) the projection of incident wave vector \( \mathbf{k} \) onto the sample plane and \( \theta \) the angle of the incident illumination ranging from 0.001 to 0.02 mrad. The resulting shifted Fourier intensities are added to simulate incoherent superposition of all illumination angles. The resulting Fourier modulus is shown in Fig. 16a. Comparison with the perfectly coherent pattern of Fig. 15d shows an overall loss of contrast in the interference fringes for all spatial frequencies, which is clearly visible in
the linescans (Fig. 16b and c). Since the contrast is reduced for all frequencies this does not explain the absence of high frequency information. The linescan through the experimental diffraction pattern modulus (Fig. 16d) shows similarity to the simulated data with partial coherence (Fig. 16b), but the signal level at higher spatial frequency is too low to distinguish it from the noise. The maximum count in the experimental diffraction pattern modulus is about 200, whereas the dynamic range (14-bit) of the CCD-camera allows for $2^{14} = 16384$ counts. This means that regions of the diffraction pattern with low intensity will have considerable statistical noise. Since the stability of the microscope did not allow longer exposure times, the two holes could be illuminated by a coherent probe instead of a plane wave to increase the intensity (see paragraph 4.1).

The question arises whether the dynamic range of the CCD camera is high enough to record all the fringes assuming perfect coherence and stability. Fig. 16e shows a logarithmic plot of the intensity linescan from the perfect coherent pattern (Fig. 15d). With a dynamic range of $10^4$ all 35 fringes visible in the linescan could theoretically be recorded. The experimental pattern shows only 12 fringes.

Fig. 16. Simulation of partial spatial coherence: (a) Modulus of incoherent superposition of 400 different Fourier intensities calculated from Fig. 15c with illumination angle ranging from 0.001 to 0.02 mrad and azimuth $0 - 2\pi$. Overall contrast of interference fringes is reduced (compare to Fig. 15d). Line indicates position of linescan (b). (b) Linescan through (a) shows reduced contrast for all spatial frequencies compared to linescan (c). (c) Linescan through perfectly coherent Fourier modulus (Fig. 15d). (d) Linescan through experimental diffraction pattern modulus (Fig. 15a) looks similar to (b) but high frequency information is buried in noise. (e) Logarithmic plot of the intensity linescan from the perfect coherent pattern (Fig. 15d). With a dynamic range of the CCD detector of $10^4$ all 35 fringes visible in the linescan could theoretically be recorded. The experimental pattern shows only 12 fringes.
much lower than for simulated data, where sometimes more than 1000 iterations are needed. The simulations have been done for perfect data going out to high spatial frequencies. The low spatial frequencies in the simulations are recovered after less than 100 iterations, but the high frequencies need more iterations. This is because the HIO algorithm recovers low spatial frequencies first. Our experimental patterns on the other hand consist only of low spatial frequency information, which explains the small number of iterations.

Noise in the measurement may also cause difficulties in obtaining convergence in the phase recovery. Strategies have been devised for phase retrieval from noisy Fourier intensities [27] and the HIO algorithm seems not to be extremely sensitive to noise [33].

Fig. 17 shows another experimental diffraction pattern from two holes and the recovered image modulus and phase after 100 iterations with triangle support and weak-phase object constraint (complex reconstruction with positivity constraint on real and imaginary part). Fig. 17d shows the TEM image of the two holes. The smaller hole is not reconstructed very well, but the bigger hole has similar shape in the TEM image and reconstruction. One would expect the phase inside the hole to be constant, since the hole is empty. Instead the phase has a wedge shaped profile. This phase profile depends on the choice of the origin in Fourier space. Shifting the diffraction pattern prior to reconstruction produced different phase profiles inside the hole, in a similar way to image reconstruction in off-axis electron holography where inaccurate centering of the sideband produces a phase wedge [34]. Different iteration trials with the same origin always reproduced approximately the same phase wedge inside the hole. The diffraction pattern has no inversion symmetry, which is due to the fact that the object exit wave is complex but does not have inversion symmetry. Absence of inversion symmetry in the diffraction patterns could also result from misalignment of the microscope or aberrations (astigmatism, distortions) of the magnifying lenses after the objective lens (intermediate lenses and projector). To minimize these effects, the central spot of the diffraction pattern without any sample was made as small and circular as possible prior to recording the diffraction pattern from the sample. We have also tried to deconvolute the diffraction pattern with this point-spread function, but this did not result in any improvement of the experimental reconstructions. Reconstruction efforts to recover the complex exit wave for holes filled with MgO crystals (strong phase objects) with a tighter support generated from the TEM image were not yet successful, which is not surprising in light of the noisy data and missing high spatial frequencies in the diffraction pattern. Another problem could arise from the fact that the physical support consists of a long tunnel (30 nm hole in 220 nm thick film), therefore diffraction at the tunnel walls (similar to RHEED) could make the exit wave complex even in the absence of an object.

6. Conclusion

Recovery of the aberrated complex image wave in TEM using image and diffraction pattern information has been possible since Gerchberg and Saxton introduced their iterative algorithm 30 years ago [3]. Here an attempt has been made to use Fienup’s HIO-algorithm to recover the un-aberrated complex exit wave from TEM diffraction data alone by using a support constraint. This would provide diffraction-limited super-resolution images. Our preliminary experimental results show the reconstruction of holes in an opaque film with a resolution of about 5 nm. Further experiments are under way with weak phase objects.

Several possible scenarios are proposed for exit wave phase recovery from the diffraction pattern modulus in TEM:

(a) For general complex objects an opaque two-hole support made by lithography can be used. A loose support and real non-negativity constraint in image space can be used to find the approximate magnification and orientation of the support. Alternatively the autocorrelation calculated from the diffraction pattern can be used. A tighter support can then be defined from these results combined with the TEM image of the support. Subsequent complex reconstruction with a tight
support constraint should recover the complex exit wave. Many pixels are needed to record the Young’s interference fringes between the two holes (low spatial frequencies) and the Bragg spots of the crystal inside the hole (high spatial frequencies) in the same diffraction pattern, e.g. $2048 \times 2048$ pixels are necessary to record 0.1 nm Bragg spots and Young’s fringes from 2 holes 50 nm apart. The use of a small selected area aperture in a conventional TEM to confine the beam to the two support holes has the disadvantage that the actual contributing area of the sample depends on the diffraction angles $\Theta$. The area contributing to a high order diffraction beam is displaced by $y = C_s \Theta^3$ [35] with the diffraction angle $\Theta$ and $C_s$ the spherical aberration constant. This effect can be appreciable for high order Bragg reflections and small selected area apertures. If the support holes are filled with a crystal from which the exit wave will be reconstructed, the high order reflections will be weak or missing due to the spherical aberration error (the area of the sample which

Fig. 17. Image reconstruction from experimental diffraction pattern: (a) experimental diffraction pattern from two empty holes in the film shown in Fig. 14. Microscope conditions the same as in Fig. 15. (b) Modulus of the reconstructed image. 100 HIO-iteration with positivity constraint on real and imaginary part in image space and larger triangular computational support. The hole shape agrees well with the TEM image in (c). Small hole is not well reconstructed. (c) Phase of the reconstructed image. The phase inside the hole is expected to be constant, since the hole is empty. Different centering of the diffraction pattern creates different phase profiles inside the hole. (d) TEM image of holes that created diffraction pattern (a).
contributes to the high order reflections does not include the crystal). With the Koehler illumination system [36], which has been adapted for electron microscopy [37], all diffracted beams come from the same sample area, within a much smaller error. In Koehler illumination, the contributing area is controlled by the condenser aperture. Therefore Koehler illumination would be ideally suited as an illumination system for phase recovery. The problem of low intensity in the diffraction pattern could be reduced by using coherent probe or Koehler mode illumination. Simulations aimed at using a coherent probe of known form as the support were not successful, probably because the support edge is not well defined. We note that coherent nanodiffraction patterns from small crystals can readily be obtained showing diffraction out to sub-Angstrom spacings with high intensity [20].

(b) The results from the pure phase object simulation suggest the use of a small selected area aperture as the support for a phase object since one support area is enough in this case. For the above-mentioned reasons this scheme could only work with Koehler illumination. A tight computational support is needed in the reconstruction.

(c) The weak-phase object needs only a loose computational support and a physical support with one area (e.g. single hole) is possible, but two or more holes are preferable. The computational support has to be large enough to contain all physical support holes. Therefore reconstruction from experiments should be much easier. The object could be stretched across a hole in an opaque film and the beam confined to the area of the hole. Parallel illumination is necessary; probe illumination would alter the exit wave due to the convergence of the wave front and create strong phase shifts so that the real and imaginary part of the exit wave would no longer be positive. Use of a transparent support is possible theoretically, however in practice, most thin films are non-homogeneous weak phase objects. The image could be recovered with a computational support, which is zero outside by scaling the central pixel of the diffraction pattern. This scaling involves removing the intensity, which is due to the transparent area around the object.

(d) A weak phase object with complex optical potential (or charge density for X-rays) behaves similarly to a weak phase object with real potential, provided that the sign constraint can be applied to both the real and imaginary (absorptive) parts of the potential. Thus, for phase shifts of less than $\pi/2$ and similar restrictions on absorption, the shape of the object (support function) need not be accurately known, and a triangular computational support may be used.

For the soft X-ray case, our simulations suggest the use of one or several pinholes in otherwise opaque mica, containing a weak phase and amplitude object of interest, and an aperture of about 1 μm diameter, placed as close as possible to the mica. The aperture diameter should equal the coherence width of the beam, and serves to exclude other holes. The positivity constraint can be applied for the weak phase objects in the center of the aperture, where Fresnel diffraction from the aperture edge is avoided. The computational support is a triangle larger than the 1-μm aperture, outside of which the intensity is set to zero. The simulations have shown in all cases, that the boundary of the physical support area should be as sharp as possible. Reconstructions for real or weak objects where the physical support is not accurately known (one hole support) and a loose computational support is used converge only if the physical support boundary is abrupt. To reconstruct a general complex object, accurate knowledge of the physical support is needed (two area support) and the boundary should be as sharp as possible.

**Acknowledgements**

We are grateful to Dr. J. Miao for useful discussions. This work was supported by ARO award DAAD190010500. This work was performed in part at the Cornell Nanofabrication Facility (a member of the National Nanofabrication Users Network), which is supported by the National Science Foundation under Grant ECS-9731293, Cornell University and industrial affiliates.
References