Phase recovery and lensless imaging by iterative methods in optical, X-ray and electron diffraction

By J. C. H. Spence¹, U. Weierstall¹ and M. Howells²

¹Department of Physics and Astronomy, Arizona State University, Tempe, AZ 85287-1504, USA
²Advanced Light Source, Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

Published online 28 March 2002

Thomas Young’s quantitative analysis of interference effects provided the confidence needed to revive the wave theory of light, and firmly established the concept of phase in optics. Phase plays a similarly fundamental role in matter–wave interferometry, for which the field-emission electron microscope provides ideal instrumentation. The wave–particle duality is vividly demonstrated by experimental ‘Young’s fringes’ using coherent electron beams under conditions in which the flight time is less than the time between particle emission. A brief historical review is given of electron interferometry and holography, including the Aharonov–Bohm effect and the electron Sagnac interferometer. The simultaneous development of phase-contrast imaging at subnanometre spatial resolution has greatly deepened our understanding of atomic processes in biology, materials science and condensed-matter physics, while electron holography has become a routine tool for the mapping of electrostatic and magnetic fields in materials on a nanometre scale. The encoding of phase information in scattered far-field intensities is discussed, and non-interferometric, non-crystallographic methods for phase retrieval are reviewed in relationship to electron holography. Examples of phase measurement and diffraction-limited imaging using the hybrid input-output iterative algorithm are given, including simulations for soft X-ray imaging, and new experimental results for coherent electron and visible-light scattering. Image reconstruction is demonstrated from experimental electron and visible-light Fraunhofer diffraction patterns. The prospects this provides for lensless imaging using particles for which no lenses exist (such as neutrons, condensates, coherent atom beams and X-rays) are discussed. These new interactions can be expected to provide new information, perhaps, for example, in biology, with the advantage of less damage to samples.

Keywords: diffraction; phase recovery; lensless imaging; phase problem; coherent imaging

One contribution of 15 to a special Theme Issue ‘Interference 200 years after Thomas Young’s discoveries’.

1. Introduction

It was the quantitative analysis of interference effects which distinguished the work of the great polymath Thomas Young from that of his predecessors, and so gave Fresnel and others the confidence to later develop the wave theory of light fully (Young 1804). Whereas Huygens and Grimaldi had earlier advocated wave theories against the authority of Newton, Young was the first to analyse both his new, and Newton’s old, experimental data, using the correct path differences at slits and edges, and to deduce from them both the wavelength and frequency of light. (Young used Roemer’s remarkably accurate estimate of the speed of light, made a century earlier.) In this way the concept of phase was firmly established within optics.

But the wave–particle duality remained. In 1909, G. I. Taylor wrote ‘Sir J. J. Thomson has suggested that if the light in a diffraction experiment were so greatly reduced that only a few ‘indivisible units of energy” should occur on a Huygens zone at once, then the ordinary phenomena of diffraction would be modified’. Taylor’s experiments, aimed at understanding this dualism, examined the shadow-edges of a needle (Taylor 1909), using a gas flame illuminating a slit as a coherent source, smoked-glass attenuators and exposure times of up to three months. The modern equivalent of this experiment for matter waves has been shown by Tonomura (1993) and can also be found as figure 3 of the accompanying article by Lichte (2002). The statistical build-up of an electron ‘two-slit’ interference pattern (actually obtained using an electrostatic biprism) is observed, under conditions in which each electron arrives at one pixel of the detector long before the next leaves the source. Dirac’s comment that ‘each photon interferes only with itself’ clearly also applies here to electrons under these experimental conditions. Localization of the particle at the source prevents a determination of which hole the electron went through, consistent with both the uncertainty principle and the Van Cittert–Zernike theorem. This theorem ensures that the spatial coherence width span the two holes. In this case, uncertainty is synonymous with coherence. As discussed further below, these two concepts are also linked through the definition of beam degeneracy.

This extension of Thomas Young’s work on interference of light to interference of massive particles lies at the heart of much modern physics, since it requires the assignment of a phase to the matter–wave wavefunction. That idea played a crucial role in the development of quantum mechanics, with the argument that, for a single-valued wave function, phase may change only by $2\pi n$ around any closed loop. This argument was used to justify the quantization of electron orbitals in Bohr’s atomic theory, the quantitative success of which (for hydrogen spectra) again gave confidence to the developers of formal quantum mechanics. It arises in Dirac’s quantization condition for magnetic monopole strength (Dirac 1931) and in the quantization of magnetic flux in superconductors amongst other places, and so may be considered to be one of the more important ideas of 20th-century physics.

In this paper we are concerned with the encoding and retrieval of the phase of electron and optical wavefields from simple scattering experiments, for the purpose of image formation. Because of the bright, coherent, field-emission electron sources available, the sophisticated design, flexibility and convenience of modern transmission electron microscopes (TEMs), much of the most significant matter–wave interferometry has been undertaken using the high-quality, coherent electron beam available in a TEM. Under this heading we should include modern phase-contrast atomic-
resolution imaging, the probe-formation process in scanning electron microscopy, and electron crystallography. For example, the principle of coherent detection originally suggested by Schiske in 1968 has since been extensively applied to allow atomic-resolution images to be synthesized from a through-focal series of electron micrographs (for a historical review, see Saxton (1994)). In this way, by efficiently deconvolving the instrument response, controlling noise and compensating for loss of information due to partial coherence, resolution has been extended beyond the conventional Scherzer point resolution limit out to the information limit set by electronic instabilities (Coene et al. 1992). This limit currently stands at slightly below 1 Å (O’Keefe et al. 2001), which will soon open up new possibilities in materials science for understanding atomic processes in matter, especially at interfaces and in glasses, and for the discovery of useful new nanostructures. Atomic-resolution TEM provided the first images of the fullerene ‘buckyballs’ (Iijima 1980) and the discovery of carbon nanotubes (Iijima 1991). For biology, although resolution may be limited by other factors such as variations in the conformation of molecules, radiation damage and the bending of monomolecular layers, the impact of this approach has been considerable. Using two-dimensional crystals, it has revealed the detailed mechanism of photosynthesis in cell membranes (Henderson et al. 1990; see also Subramaniam et al. 2002). For macromolecules such as the ribosome, responsible for protein production (‘life itself’), the cryomicroscopy method was the first to reveal the shape (at 1 nm resolution) of large molecules, which cannot be crystallized (Frank 1996). Despite recent limited success in crystallizing membrane proteins for X-ray work, electron cryomicroscopy remains vital for assisting in phasing X-ray data, confirming the heavy-atom positions assumed in X-ray multi-wavelength anomalous diffraction analysis, and in showing how smaller proteins, of known structure, fit together and interact.

2. Electron interferometry and holography

More directly related to Young’s work is modern electron interferometry and holography. Following the 1927 discovery of electron diffraction (Davisson & Germer 1927; Thomson & Reid 1927) in confirmation of de Broglie’s hypothesis, observations of helium-atom diffraction by crystals soon after provided further confirmation of the matter–wave hypothesis (Esterman & Stern 1930). The observation of Fresnel fringes in an electron microscope (EM) by Boersch (1943) provided further evidence of wave-like properties—and the realization that these patterns encoded phase information, which might be reconstructed—led to the birth of holography (Gabor 1949). (Since Gabor was unable to solve the twin-image problem presented by these in-line holograms, Boersch, with some justification, remained convinced that he had invented ‘holography’, and viewed the term as an unnecessary acronym for an interference pattern (T. Mulvey 1996, personal communication).) Early, partly successful experiments (Haine & Mulvey 1952) used the ‘transmission method’, based on the realization (currently being rediscovered by the electron-microscopy community) that an out-of-focus image is an in-line hologram. Gabor’s original aim was to improve the resolution of EM images, using holography to expose the phase of the wavefront aberrations, and so to deblur the images. Reconstructions were obtained in 1951 using a monochromatized arc lamp at 1 nm resolution (before the invention of laser or useful computers). However, EM images were limited in resolution by mechanical

Phil. Trans. R. Soc. Lond. A (2002)
and electronic instabilities rather than aberrations until the early 1960s. Only in 1995 was ‘Gabor’s dream’ fully realized, when off-axis EM holograms were first used to extend resolution beyond the limit of 0.198 nm, set by aberrations, toward the ‘super-resolution’ limit, set by electronic instabilities, of 0.1 nm (Orchowski et al. 1995).

The earliest beam-splitters for electrons used either Bragg diffraction in a thin crystal in the transmission geometry (Marton 1952), or an electrostatic biprism (Moellenstedt & Duker 1955), which has become the favoured choice today. The subsequent development of the field-emission electron source (Crewe et al. 1968) at the end of the 1960s was of the greatest importance for the field, since it provided a source of nanometre dimensions and sub-eV energy spread, with a brightness per unit bandwidth greater than current-generation synchrotrons (Qian et al. 1993). By combining Moellenstedt’s biprism with Crewe’s field emission gun, modern EM instruments, using beam energies of a few hundred keV, are capable of recording, in a few seconds, images showing hundreds of Fresnel fringes, with a lateral coherence width of perhaps 0.1 µm and a longitudinal coherence length of several hundred angstroms. They thus provide for interferometry a beam of extremely high quality by the standards of accelerator physics.

The refractive index for electrons in a magnetic field was first derived, using Fermat’s principle, in 1948 (Ehrenberg & Siday 1948). The Aharonov–Bohm (AB) effect (Aharonov & Bohm 1959) subsequently demonstrated observable effects along optical paths in which fields (but not potentials) are zero, suggesting that potentials are more fundamental than fields. (This idea had apparently already been suggested by Maxwell.) More recent electron-interferometry experiments include, for example, the observation of flux quantization using the AB effect (Tonomura 1988), the scalar (electrostatic) AB effect with electrons (Matteucci et al. 1992), and the electron Sagnac interferometer (Hasselbach & Nicklaus 1993). From this has grown the field of electron holography, now widely used to map out, on a nanometre scale, internal magnetic and electrostatic fields in materials, from digital recording media to ferroelectrics, semiconductors and superconducting vortices (for reviews, see Voelkl 1998; Lichte 1988; Tonomura 1993).

A different class of electron-interference experiments depends on the chromatic coherence of electron sources and their second-order coherence properties. For the free particle beams used for electron interferometry, a degeneracy parameter δ may be defined to indicate the degree of crowding in phase space; a unity value indicates two electrons of opposite spin per cell, with cell dimensions defined by the uncertainty principle. (Components of $\Delta K$ are defined by energy spread and beam divergence; components of $\Delta r$ define the coherence patch and longitudinal coherence length.) It is readily shown (Spence et al. 1994) that δ is equal to the ratio of actual source brightness to Langmuir’s theoretical maximum value, and that δ may be measured from the dimensions of the coherence patch and the source energy spread. For a cold tungsten nano-tip field emitter, we measure $\delta = 10^{-4}$ (Spence et al. 1994), whereas a value of about unity is found for undulators at synchrotrons. (For an optical laser, in which any number of indistinguishable bosons may occupy a state, $\delta = 10^{15}$.) Coulomb repulsion between electrons limits the value of δ. Nevertheless, the brightness of electron field emitters exceeds that of current synchrotron undulators (Qian et al. 1993). The effects of photon bunching, seen in the Hanbury Brown and Twiss experiment (Loudon 1976), and the suggestion that electrons should ‘antibunch’ for

large values of $\delta$, were made at an early stage by Purcell (1956). In a quantum picture, we imagine wave packets overlapping coherently along the propagation axis by chance, giving non-classical (non-random) results for the probability of detecting particles between wave packets. For this to occur, however, the temporal coherence of the source must exceed the time between particle emissions, so that the particles are in the same quantum state. This requires the source degeneracy $\delta$ for electrons to approach unity. It is therefore not surprising that, despite considerable effort in our own laboratory and elsewhere, antibunching has yet to be observed in free coherent electron beams. In condensed matter, however, these conditions are readily achieved, and fermion antibunching has recently been observed for the first time in a two-dimensional electron gas in the quantum Hall regime (Henny et al. 1999). A review of other novel effects expected in multiparticle fermion interferometry can be found in Silverman (1993).

3. Phase in imaging: iterative methods for phase measurement

An image is a two-dimensional near-field wavefield, which may be real or complex. It may be obtained by Fourier transform from far-field scattered amplitudes if the phases of these are known. Holography encodes this phase information efficiently through the use of a reference wave, and Fourier transform (FT) holography (which uses a point reference in the same plane, and to one side, of the object) is particularly effective in separating the real and virtual twin images after the reconstruction process. This involves only illumination of the hologram with the original reference wave. Phase information is encoded in the displacements of the ‘Young’s fringes’ which result in this geometry. (The method has been extended to allow a ‘complicated or extended reference’ (Szoke 1997), such as a known fragment of a molecule, and thus may be considered as an extension of the ‘heavy-atom’ method of crystallography.) Since this experimental inversion (reconstruction) of FT holograms to images produces a unique result, it should not surprise us that an iterative search algorithm, the hybrid input–output (HIO) algorithm (Gerchberg & Saxton 1972; Fienup 1982) applied to the same geometry, should perform well. Remarkably, however, it has been found that the algorithm also converges for complex objects consisting of two separated parts, neither of which is known, but which generate an appreciable fringe modulation in the scattering pattern. A knowledge of the ‘support’ of the object (and of the sign of the scattering density) is found to be sufficient for inversion; the support is the area within which the object transmission function $T(r)$ is non-zero (or unknown), and so may take the form, for example, of the Young’s fringe pinhole mask. Different ‘unknown’ objects fill each hole. The inversion, given finite known support, has been shown to be ‘almost always’ unique (Bruck & Sodin 1979). An error metric, which depends only on the known values of the support and Fourier modulus, monitors progress toward this unique solution. In summary, we have recently come to understand that this experimental geometry of Thomas Young now offers the possibility of lensless imaging using any of the radiations for which lenses do not exist, such as coherent atom beams, X-rays and neutrons. The complementary arrangement—two compact, complex objects lying on a transparent membrane—also works. For a different non-interferometric approach to phase measurement akin to the through-focus method mentioned in the introduction, and based on the ‘transport of intensity’ equations see Paganin & Nugent (1998). This
method, which requires intensity measurements across two planes downstream of the object, has been used, for example, to consider a non-interferometric variant of the AB effect (Paganin 2001). (For a phase object, a single plane is sufficient, since unit intensity can be assumed everywhere in the object plane.) The phase vortices, which occur at zeros of intensity, have also attracted considerable interest (Allen et al. 2001). So far it has been applied to cases using collimated illumination only, so that resolution is limited by detector pixel size; however, there seems no reason not to apply it to the case of diverging illumination, which provides magnification, and a resolution limit imposed by the source size.

Despite many simulations, few experimental tests of the HIO algorithm have appeared. The algorithm iterates between real and reciprocal space (related by a Fourier transform), imposing known information in each domain. The moduli of the scattered amplitudes $I(u)$ are assumed known, and, in real space, the object function $T(r)$ is set to its known values outside the support. (Here $r$ and $u$ are two-dimensional vectors. $T(r)$ is the ratio of the complex exit wavefield (the in-focus image) to the incident wave function, often assumed planar.) The support need not always be known very accurately, and an estimate can be obtained from the 50% contour of the autocorrelation function, which is given by the Fourier transform of $I(u)$. Finally, knowledge of the sign of the real (and possibly imaginary) parts of $T$ may be used. The introduction of a small amount of ‘feedback’ (Fienup 1987) distinguishes the HIO algorithm from the error-reduction algorithm (Gerchberg & Saxton 1972). Convergence is found to be improved further in three-dimensions (Miao et al. 2001). For real objects (such as the charge density which diffracts X-rays), a single object is sufficient and the computational boundary used in the iterations may be a different shape from the physical support, which it includes. When viewed as an optimization process, the algorithm appears to be capable of finding the thousands of unknown parameters (the phases in a typical image) in the presence of noise without the usual problems of local minima and stagnation. It thus appears to solve a very large global optimization problem, despite the application of a non-convex constraint (the known Fourier moduli). (A convex set of solutions is one for which all points lying on any line joining any two points within the $N$-dimensional set lie inside the set; a kidney-shaped set in two dimensions is non-convex. $N$ is the number of pixels in the image. If a unique solution can be shown to exist (as in this case), iteration between the boundaries of two convex sets guarantees convergence to this solution without stagnation at local minima. Hence convex constraints, such as symmetry, known support or sign, are sought in optimization problems.) While inversion from known phase information is a convex process leading to a unique result (Lindaas et al. 1998), inversion from intensity alone is non-convex, so that local minima may formally be expected in approaches based on numerical optimization (Stark 1987). The HIO algorithm has been analysed in considerable detail (Fienup 1982; Paganin 2001), including treatments based on Bregman projections onto convex sets (Stark 1987; Bauschke et al. 2002) and those, going back to Sayre (1952), which start from the realization that Bragg sampling undersamples the autocorrelation function of a molecule in a crystal (Miao et al. 2001). The number of Fourier series equations is then compared with the number of unknowns as the support is made a progressively smaller fraction of the total diffracting area. The lowest spatial frequency in the diffraction pattern corresponds to a distance wider than the support. Using this ‘oversampling’ approach, soft X-ray images of lithographed symbols were recently

reconstructed from far-field scattered intensities in a lensless scattering experiment (Miao et al. 1999). Spatial resolution is limited only by noise and wavelength in these experiments. For technical reasons, however, the reconstruction was assisted by low-resolution optical images, which provided the low-order spatial frequencies, and the object transmission function was of a special mask-like form, having only two (complex) transmission values, whose signs were known. We have therefore conducted both simulations and experiments for phase measurement in scattering of X-rays, coherent visible light and coherent electron beams, to evaluate this HIO algorithm for more general objects.

4. Simulations for X-rays

Simulations have been undertaken for objects consisting of an array of randomly positioned pinholes, to determine if this image could be reconstructed from its soft X-ray diffraction pattern. If so, these holes could later be filled with structures of interest for reconstruction. The area occupied by pinholes has been limited to the coherence width of the beam by a 1 μm aperture. This aperture will not fall exactly in the plane of the pinholes, therefore the wavefront at the pinholes will have undergone Fresnel diffraction at the aperture and so will be complex. Figure 1 shows the results of these simulations assuming a wavelength of 2.5 nm for the incident soft X-ray beam. The pinholes are 10 nm in diameter. After 150 iterations, the HIO algorithm has converged without stagnation to an image which is related to the object by inversion through the centre. The modulus of the estimated object function was taken at each iteration (modulus constraint), since the object is assumed real. Only a rough estimate of the size of the aperture is needed: the computational support condition applied in the iterations is a triangle larger than the aperture. The unimportant inversion ambiguity cannot be resolved by the HIO algorithm, since an inverted image, our real computational support, and the circular aperture all have the same Fourier modulus. The Fresnel propagation makes this strictly a complex object; however, we obtain an accurate reconstruction of the holes in the film, despite the use of a positivity (modulus) constraint in object space. The phase of the complex wave transmitted through the pinholes is not reconstructed due to the imposed positivity.

5. Experiments with light

Coherent laser optical phase retrieval experiments have been performed using the HIO algorithm for several objects. The phase was successfully retrieved for several different support constraints and object types. So far, only opaque support constraints (apertures) have been used, since these present fewer problems with excessive central beam intensity (blooming). The source was a frequency-stabilized He–Ne laser (λ = 632.8 nm), collimated to provide the planar wavefronts needed at the sample for real objects (figure 2). Diffracted light was focused with a lens onto a liquid-nitrogen-cooled charge-coupled device (CCD) camera, and the digital image reconstructed using the HIO algorithm in the Digital Micrograph script language.

The first object was a resolution test object with transparent numbers and lines in an opaque chrome film on glass. The illuminated area was limited by a thin 300 μm diameter aperture in contact with the chrome film. Assuming the glass is flat to within less than λ/10 across the illuminated area, the chrome mask acts as a
Figure 1. Reconstruction of simulated pinholes illuminated by a Fresnel diffracted X-ray wavefield (out-of-focus physical support). (a) Simulated random array of pinholes illuminated by wave amplitude, which has been Fresnel propagated 200 nm from an aperture to the pinholes. Wavelength $\lambda = 2.5$ nm, pinhole size 10 nm, aperture diameter 1 $\mu$m. (b) 200 nm out-of-focus wavefield from 1 $\mu$m aperture incident on pinholes. (c) The Fourier transform of (a) is a speckled Airy’s disc. Modulus used as input for the HIO algorithm. (d) Reconstructed pinholes after 150 iterations with modulus constraint. (To enforce positivity in real space, only the modulus of the pixel values inside the support area is retained during each iteration.) A large triangular computational support was used. The image is inverted; this is an ambiguity which cannot be resolved by the HIO algorithm.

real binary object so that the diffraction pattern should be symmetric. A coherent optical image was also recorded with the camera moved into the image plane of the lens. This is shown in figure 3a. The camera was then moved to the back focal plane of the lens and the diffraction pattern intensity $I(u,v)$ was recorded (figure 3b). The autocorrelation FT $I(u,v)$ was used to determine the approximate size of the object and a circular support half the size of the autocorrelation was used in the
computer. The HIO algorithm was used to recover the phases of the object exit wave and the algorithm always converged after about 50 iterations with different initial random phase sets (figure 4a). A larger triangular support constraint was also used to recover the image, as shown in figure 4b. The normalized root-mean-squared error in the object domain (Fienup 1987) (a measure of difference between the current image estimate and the known pixel values outside the support) did not drop below 0.2, probably due to imperfections in the diffractogram, e.g. noise and light reflections from the entrance glass surface of the CCD camera, which create asymmetry in the diffraction pattern (see figure 4b above centre). Despite these problems the reconstructed image looks almost identical to the coherent image in figure 3a.

A strong phase object (introducing phase shifts greater than \( \pi/2 \)) was made using a support consisting of two separated areas (two apertures of 300 µm diameter sep-
J. C. H. Spence, U. Weierstall and M. Howells

Figure 4. (a) Modulus of reconstructed image wave using a tight circular support constraint. (b) Modulus of reconstructed image wave using a loose triangular support constraint. A positivity constraint was used in real space.

Figure 5. (a) Coherent laser optical image of phase object with two-hole support. The edges of two mica plates are visible across the hole on the right. (b) Laser optical diffraction pattern obtained by merging three patterns with different exposure times to avoid saturation of the central beam.

arated by 350 μm). One hole was covered with two 5 μm thick mica sheets which overlap in one region. The expected phase shift is calculated to be 9π. The coherent optical image is shown in figure 5a. The mica is transparent except for the edges where Fresnel contrast effects are visible.

The resulting diffraction pattern is shown in figure 5b. The central beam intensity in the diffraction pattern from this transparent object was too strong to be recorded simultaneously with the weaker high spatial frequencies, so that several exposures were needed. All exposures were merged to one pattern in which saturated pixels were replaced by scaled pixels from the lower exposure patterns. The computational
support size and orientation were derived from the autocorrelation. The complex HIO-phase retrieval algorithm was used to recover the image wave phase and modulus shown in figure 6. The phase shift by the mica plates in the reconstruction is \( \pi \), which, since the algorithm reconstructs phase modulo \( 2\pi \), is in good agreement with our independent measurement of \( 9\pi \) based on the measured thickness and known refractive index of mica. Thus the HIO algorithm is capable of quantitative phase measurement.

Experiments with a transparent physical support have been attempted, using a limiting aperture upstream of the sample, but have not been successful so far because of the high intensity in the centre of the diffraction pattern, which causes blooming in the CCD chip. This problem persists for three-dimensional objects; however, a three-quadrant mask in the object plane has been proposed to address this difficulty (Sayre et al. 1998). (Beam stops cannot be placed near enough to the CCD to avoid generating additional fringes.) Experiments using the known intensity envelope of a focused beam as the support have also not been successful; for such complex objects there appears to be a need for sharp physical support edges, perhaps required to generate reference edge waves. (For real objects, there is no such requirement.) However, a coherent probe spanning two ‘known’ holes works well in simulation.

Previous experiments have been based on objects of simple type, such as binary
objects, which contain much less information than greyscale objects. Experiments with human cheek cells, which serve as more complicated greyscale phase objects, have therefore been performed. The cheek cells were positioned on a glass slide inside two 300 μm apertures separated by 300 μm. The apertures act as the opaque support consisting of two separated parts as needed for reconstruction of complex objects. Figure 7a shows the coherent optical image of the slightly defocused cheek cells as formed with a lens. Only Fresnel edge contrast is visible, while the nuclei produce some absorption. Figure 7b shows the optical diffraction pattern merged from four different exposures. The size and orientation of the support were estimated from the autocorrelation. The complex reconstruction of phase and amplitude after about 50 HIO iterations, using the support constraint, is shown in figure 8. The cells are visible with large-area phase contrast in the phase image and the phase shift of the cells is about π. In the reconstructed amplitude only edge contrast is visible, as in the optical image.

For a three-dimensional object extended along the optical axis, the question arises as to which plane in the object is reconstructed, given that the detected intensity lies at infinity and contains contributions from scattering at every point along the axis in the object space. A simulation to clarify this is shown in figure 9. The simulated object consists of two planes separated by 1 mm in the z-direction. Plane 1 has a hole containing a complex object (the number 1), while plane 2 has two holes containing complex objects (numbers 2 and 3). The object exit wave calculated on plane 2
Figure 8. (a) Modulus and (b) phase of reconstructed image wave. The modulus shows only edge contrast, whereas the phase shows the cells in large-area phase contrast. The phase shift due to the cells is about $\pi$. No modulus or sign constraint.

(figure 9a) shows plane 1 out of focus. The modulus of the Fourier transform (Fourier modulus) of this image wave is used as input for the HIO algorithm. In the inversion, different shapes of the supports used at each plane, shown inset, can distinguish the two object planes. If a plane is assumed unknown, a larger support hole is used that contains the physical support. Depending on which support plane is assumed as known (inset) in the iteration (plane 1 support known and plane 2 support unknown or vice versa), the other plane is always reconstructed out of focus by $+1$ or $-1$ mm (figure 9c, d). That means that in the case of known support on plane 1 the algorithm reconstructs the wave as seen through an imaging lens with out-of-focus images of the holes on plane 2. As expected, the depth of focus for interactive inversion is given by the expression for an ideal lens, which imposes a diffraction limit equal to that of the highest spatial frequency recorded in the diffraction pattern.

To verify this experimentally, a pin was mounted 1.2 mm upstream of a pair of 300 $\mu$m apertures. The coherent optical image is shown in figure 10a. A Fresnel fringe surrounds the pin, since the lens is focused on the apertures. The image was recovered from the recorded diffraction pattern intensity (figure 10b). Again three diffraction patterns with different exposure times had to be merged to avoid pixel saturation and blooming. The reconstructed image modulus is shown in figure 11a, where the
Figure 9. (a) Schematic view of a simulated object consisting of two planes separated by 1 mm in the direction of light propagation. Plane 1 has a hole containing a complex object (number 1) and plane 2 has two holes containing complex objects (numbers 2 and 3). (b) Modulus of complex exit wave on plane 2. (c) Iterative reconstruction with plane 2 support assumed known (the support shape used is shown in the top right-hand corner). Plane 1 is out of focus by 1 mm. (d) Iterative reconstruction with plane 1 support assumed known (the support shape is shown in the top right-hand corner). Plane 2 is out of focus by -1 mm.

pin appears out of focus (Fresnel fringe barely visible) because the aperture shape was imposed as support. The pin can be brought into focus by using the appropriate Fresnel propagator with Δz = 1.2 mm and λ = 632.8 nm, taking the apertures out of focus (see figure 11b). Thus, as expected, the reconstructed plane is always that plane in which the support condition is imposed, where the algorithm finds a unique solution subject to the given boundary condition. Propagation to other planes may be performed subsequently, since the complex wavefield is obtained, subject to the depth-of-focus limitation. The relative sideways shift of the pin inside the aperture between figure 10a and figure 11a is due to a slight unintentional tilt of the object (pin plus pinhole) between recording the image and the diffraction pattern.
6. Experiments with coherent electrons

Using a special support mask made by electron-beam lithography, experiments have also been undertaken using the HIO algorithm to retrieve the phases from experimental transmission electron diffraction patterns (Weierstall et al. 2002). Images might then be reconstructed that are free of electron lens aberrations. The support mask contains hole pairs of various sizes, with pairs separated by 1 μm. The separation between holes of a pair is ca. 50 nm, less than the coherence width of the illumination beam. The holes, irregular in shape, form the object we wish to reconstruct; in future work we plan to introduce nanoscale objects into the holes. Figure 12a below shows the experimental diffraction pattern from two holes shown in figure 12c obtained...
Figure 12. (a) Experimental electron diffraction pattern from two nanoscale holes in an opaque mask (shown in (c)), recorded using a field-emission electron microscope (CM200 FEG at 40 kV, 1k × 1k CCD with 24 μm pixels, camera length 6.11 m). Young’s fringes from the two holes are visible; these fringes encode phase information. (b) Reconstruction of image of holes from the diffraction pattern (a) after 100 iterations of the HIO algorithm. Modulus constraint applied in image space with large triangular computational support. The hole shape is comparable with the TEM image in (c). (c) Conventional TEM image of holes that created the diffraction pattern (a).

with a Philips CM 200 field-emission electron microscope at 40 kV. The accelerating voltage was lowered to 40 kV to make the multi-layered support opaque. Modulated Young’s fringes are clearly seen.

The resulting image after 100 HIO iterations is shown in figure 12b. The holes reconstructed from the diffraction pattern modulus look similar to the TEM image of the holes in figure 12c, but the resolution of the reconstruction is lower than the normal TEM image resolution. High spatial frequency information from the hole edge is missing in the experimental diffraction pattern, due, perhaps, to drift of the imaging lenses during the exposure time.

Table 1 shows a summary chart of all the conditions so far examined with HIO sim-
Table 1. Summary chart showing different conditions for phase recovery with HIO algorithm

(‘CA’ denotes the complex algorithm (uses diffracted intensity and support constraint in real space); ‘RIP’ denotes the complex algorithm with positivity constraint of Re\(\psi\), Im\(\psi\) (uses diffracted intensity and support constraint in real space, negative Re\(\psi\) and Im\(\psi\) inside the support area are set positive); ‘M’ denotes the modulus constraint on \(\psi\) in real space (uses diffracted intensity and support constraint in real space, complex numbers inside the support area are replaced by their modulus to enforce positivity); ‘P’ denotes the phase constraint on \(\psi\) in real space (uses diffracted intensity and support constraint in real space, the phase is extracted and retained from complex numbers inside the support area, modulus is extracted and set to 1 inside the support area).)

<table>
<thead>
<tr>
<th></th>
<th>general complex object, (\psi = A \exp(i\phi))</th>
<th>pure strong phase object, (\psi = \exp(i\phi(r)))</th>
<th>pure ‘weak’ phase object, (\phi &lt; \pi/2)</th>
<th>real (amplitude) object</th>
<th>weak-phase object with absorption</th>
</tr>
</thead>
<tbody>
<tr>
<td>known two-hole physical support</td>
<td>CA (laser experiment: pin, cheek cells)</td>
<td>CA, P (laser experiment: mica)</td>
<td>RIP, CA</td>
<td>M, RIP, CA</td>
<td>RIP, CA</td>
</tr>
<tr>
<td>known one-hole physical support</td>
<td>CA, only for certain support shapes (e.g. triangle with sharp edges)</td>
<td>CA, slow convergence, P works better</td>
<td>RIP, CA</td>
<td>M, RIP, CA (laser experiment: numbers)</td>
<td>RIP, CA</td>
</tr>
<tr>
<td>unknown two-or-more-hole physical support, loose (e.g. triangle computational support)</td>
<td>—</td>
<td>—</td>
<td>RIP</td>
<td>M (TEM experiment: holes) (laser experiment: numbers) (planned: X-ray, mica pinholes)</td>
<td>RIP (Miao X-ray experiment: transparent support)</td>
</tr>
<tr>
<td>unknown one-hole physical support (e.g. triangle computational support)</td>
<td>—</td>
<td>—</td>
<td>RIP, only for hole with low symmetry and sharp edges, slow convergence</td>
<td>M, RIP, only for hole with low symmetry and sharp edges, slow convergence</td>
<td>RIP, only for hole with low symmetry and sharp edges, slow convergence</td>
</tr>
</tbody>
</table>
ulations and experiments. The effectiveness of different constraints in real space for different physical and computational supports is compared. The physical support is the support which exists in the experiment and which may or may not be accurately known. The computational support is the support used in the HIO iterations. The table shows experiments that have been performed in italics. The first column shows different possibilities for the physical support, the most favourable at the top, the worst at the bottom. The computational support and known physical support are identical. The computational support used with the unknown physical support is larger than the physical support. Reconstructions with one-hole physical support and loose computational support are only possible with weak or real objects, and support holes with sharp edges and low symmetry and convergence are very slow in simulations. A known two-or-more-hole support proved to be most favourable in all cases.

7. Discussion and conclusions: diffractive imaging with other particles

In this paper we have reviewed the implications of Thomas Young’s experiment for matter waves and, in particular, for electron interferometry. By taking advantage of the very high quality of the coherent electron beam available in the modern field-emission electron microscope, a number of important results (such as the Aharonov–Bohm effect and the electron Sagnac effect) have been established, and the field of atomic-resolution transmission electron microscopy, phase-contrast imaging, and electron holography opened up. All these methods continue to produce a large scientific pay-off in microstructural materials characterization and discovery, and in biology. A new lensless imaging method is then reviewed, which may be understood in part by analogy with electron holography. This HIO algorithm uses an iterative approach to assign phases to diffraction pattern intensities. To evaluate its performance, we give experimental examples of image reconstruction with both coherent electrons and light, together with simulations for the soft X-ray case. Thus we have shown experimentally that the HIO algorithm is capable of solving the phase problem for scattering from non-periodic objects, and of making quantitative measurements of phase (modulo $2\pi$) in good agreement with independent measurements. The beamsplitting arrangement of an interferometer is not required, and diffraction-limited lensless imaging seems possible as a result. This may open the way for imaging with the many particles for which no lenses exist, such as neutrons, ballistic electrons in solids at low temperature, or neutral atoms and ions, either as free coherent beams or condensates. (Bright, coherent sources have recently appeared for many of these (Pauly 2000).) The differing interactions of these particles may provide useful new probes for chemistry, materials science and biology. For example, the reduced radiation damage effects of imaging organic material with low-energy electrons, neutrons and helium atoms have recently been compared (Spence et al. 1999), and HIO simulations have been completed for neutron imaging of a diatom (Spence et al. 2001). The very small refractive index of organic matter for neutrons limits resolution and contrast for thin samples. For low-energy ions or metastable helium, strong sample–beam interactions must be considered. Damage from neutral helium-atom imaging of molecules should be negligible; however, penetration is also negligible (producing a profile image). Detectors for coherent atom beams still present serious problems, since area detectors are inefficient. For biology the weak atom interaction is desirable.
Phase recovery and lensless imaging

893 to avoid damage; however, a strong interaction at the detector is also needed, so that ionization before detection must be arranged.

For real objects, we have seen that the resolution in images recovered by the HIO algorithm is largely independent of the accuracy with which the support is known. The addition of noise affects the probability that the unique solution will be found (as measured by the error metric), rather than the resolution, which is near diffraction-limited. For complex objects, our experience indicates that a much more accurate knowledge of the physical support boundary is needed, approaching the scale of the resolution sought. The tight support and two-hole geometry (see Fienup (1987) for other geometries) needed for complex objects (complex refractive index) will become increasingly accessible with the development of new lithographic techniques for support construction. This can be combined with independent characterization of the support shape and the scaling information provided by the known autocorrelation of the support. Experimentally favourable cases for imaging with nanoscale resolution at present require loose support, which, table 1 shows, greatly favours real objects, such as the charge density which diffracts hard X-rays at energies well away from absorption edges, or very thin samples for high-energy electrons. Unfortunately, beam coherence is more readily obtained for soft X-rays.

Using modern lithographic techniques, it is becoming increasingly possible to establish the experimental conditions of known support required by the HIO algorithm, both for free particle beams and for electrons in solids at low temperatures. Thus, for example, we have seen the algorithm used to determine the phase of wave functions for electrons in a one-dimensional YBCO Josephson junction, using TEM images of the structure to define the required support (boundary conditions) (Carmody et al. 2000). Given the fundamental role that phase has played historically in physics, and which Thomas Young was amongst the first to appreciate, any technique capable of exposing phase can be expected to yield many surprises in physics in the future.

Supported by ARO award DAAD190010500. We thank Dr Q. Chen for the use of figure 12 (from Weierstall et al. 2002).

References


Phil. Trans. R. Soc. Lond. A (2002)
Dirac, P. 1931 Quantum singularities in the electromagnetic field. Proc. R. Soc. Lond. A 133, 60.
Esterman, I. & Stern, O. 1930 Z. Phys. 61, 95.

Phil. Trans. R. Soc. Lond. A (2002)
Phase recovery and lensless imaging

895


Pauly, H. 2000 Atom, molecule and cluster beams. Springer.


Young, T. 1804 The Bakerian Lecture [1803]. Experiments and calculations relative to physical objects. Phil. Trans. R. Soc. Lond. 94, 1–16.