A phase retrieval algorithm for shifting illumination

J. M. Rodenburg and H. M. L. Faulkner[a)]
Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 3JD, United Kingdom

(Received 16 June 2004; accepted 17 September 2004)

We propose a method of iterative phase retrieval that uses measured intensities in the diffraction plane to solve the phase problem in a way that bypasses the problem of lens aberration, leading to greatly improved spatial resolution. This method is stable, easy to implement experimentally, and can be used to view a large area of the specimen when that is desired. © *2004 American Institute of Physics*. [DOI: 10.1063/1.1823034]

Short wavelength transmission microscopy (electron and x-ray) may well be revolutionized by recent advances in diffractive imaging. If we dispose of the lens, but instead arrange for the object to be small, then the diffracted intensity in the far-field can be used to solve for the phase of the scattered radiation via iterative methods.[1–5] In this way, resolution is no longer limited by the transfer function of the lens. Unfortunately, the requirement of an isolated object function is hard to arrange experimentally. In this letter we consider the scattering geometry shown in Fig. 1. Part of an extended object is illuminated by a substantially confined illumination function, formed using a poor or aberrated lens. We model specifically a defocused beam crossover (i.e., a probe) as might be found in a scanning transmission electron microscope (STEM). Under these conditions, we have found that conventional iterative methods do not work, because curved wave fronts incident upon on the object lead to an ambiguity in the defocus value of the reconstruction.

The geometry in Fig. 1 has the key advantage that the illumination (or object) can be moved laterally to many positions, and so we can record diffraction patterns that contain varying sets of information about the object. This shifting process itself leads to a solution of the diffraction phase problem using a method called ptychography.[6,7] This letter presents an iterative method for solving for the object function at high resolution (predicated by the angular size of the detector) using only a few diffraction patterns, recorded with the illumination beam in overlapping positions. The algorithm can be extended to any number of probe positions, which can span a very large field of view, and can be updated in real time. In this way, the lens is used merely to define a current area of interest in the object, whereas the high resolution data are being extracted from the diffraction plane. The method is therefore a hybrid of low-resolution scanning imaging and very high resolution diffractive imaging. The algorithm is applicable to solving a number of different phase retrieval problems where a known multiplicative function can be moved relative to an unknown function of interest.

We now describe the phase retrieval algorithm. Let $O(\mathbf{r})$ and $P(\mathbf{r})$ represent two-dimensional complex functions. $O(\mathbf{r})$ is the transmission function of the specimen. $P(\mathbf{r})$ is the illumination function, which in this case is the complex STEM probe at the entry surface of the specimen. We assume that $O(\mathbf{r})$ or $P(\mathbf{r})$ can be moved relative to one another by various distances $\mathbf{R}$. We refer herein to moving $P(\mathbf{r})$, although $O(\mathbf{r})$ can equally well be moved.

We now form the product of $O(\mathbf{r})$ with $P(\mathbf{r}-\mathbf{R})$ to produce the exit wave function of $\psi(\mathbf{r})$, i.e.,

$$\psi(\mathbf{r},\mathbf{R}) = O(\mathbf{r})P(\mathbf{r}-\mathbf{R}). \tag{1}$$

This will generally be accurate for a thin object.

The algorithm works to find the phase and modulus of the complex function $O(\mathbf{r})$. It assumes knowledge of the function $P(\mathbf{r}-\mathbf{R})$. In some situations this will be known to high accuracy already, because the parameters that effect the formation of the incident beam are known. In other situations it will be necessary to determine the phase of the beam using methods such as the iterative through focal series algorithm,[8] or to use some other method of accurately characterizing the incident probe. If the incident beam is inaccurately known, the phase retrieval algorithm will still work effectively, though with a less accurate end result.[9]

Several measurements of the intensity of the wave function in some plane other than that containing the specimen are also required. We usually use the diffraction plane, which is related to the specimen plane by the Fourier transform. In that case the measured input data are the intensities of the diffraction patterns at the different probe positions. Using diffraction data has several advantages, including ease of collection, no requirement for focusing the exit wave function into an image, and the increase of resolution achieved by measuring data at high angles. In this letter we will therefore use the Fourier example. However it is important to note that the algorithm is not restricted to use of the Fourier transform. One of many possible alternative transforms is the Fresnel propagator.

The algorithm proceeds as follows.

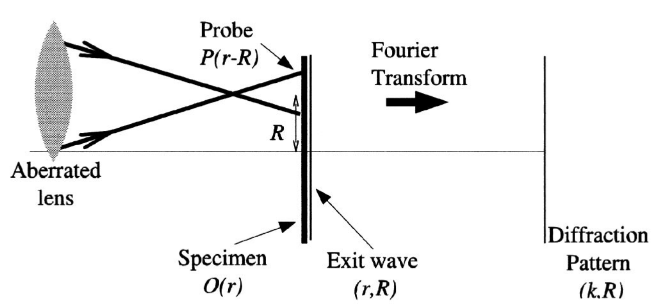

FIG. 1. STEM probe incident onto a specimen.

[a)]Electronic mail: h.faulkner@sheffield.ac.uk

(1) Start with a guess at the object function $O_{g,n}(\mathbf{r})$, where the subscript g,n represents a guessed function at the nth iteration of the algorithm. This function is in real space.

(2) Multiply the current guess at the object function by the illumination function at the current position $\mathbf{R}$, $P(\mathbf{r}-\mathbf{R})$. This produces the guessed exit wave function for position $\mathbf{R}$,

$$\psi_{g,n}(\mathbf{r},\mathbf{R}) = O_{g,n}(\mathbf{r})P(\mathbf{r}-\mathbf{R}). \tag{2}$$

(3) Transform $\psi_{g,n}(\mathbf{r},\mathbf{R})$ to obtain the corresponding wave function in the diffraction space plane, for that position $\mathbf{R}$,

$$\Psi_{g,n}(\mathbf{k},\mathbf{R}) = \mathcal{F}[\psi_{g,n}(\mathbf{r},\mathbf{R})]. \tag{3}$$

$\mathbf{k}$ is the usual reciprocal space coordinate. It is important to note that $\Psi_{g,n}(\mathbf{k},\mathbf{R})$ is a "guessed" version of the actual wave function in diffraction space, since it has been produced by the guessed object function $O_{g,n}(\mathbf{r})$. Successive iterations of the algorithm will produce increasingly accurate versions of $\Psi_{g,n}(\mathbf{k},\mathbf{R})$. We can of course write $\Psi_{g,n}(\mathbf{k},\mathbf{R})$ as

$$\Psi_{g,n}(\mathbf{k},\mathbf{R}) = |\Psi_{g,n}(\mathbf{k},\mathbf{R})|e^{i\theta_{g,n}(\mathbf{k},\mathbf{R})}, \tag{4}$$

where $|\Psi_{g,n}(\mathbf{k},\mathbf{R})|$ is the (guessed—probably incorrect) wave function amplitude and $\theta_{g,n}(\mathbf{k},\mathbf{R})$ is the (guessed—probably incorrect) phase in diffraction space at iteration n, for position $\mathbf{R}$.

(4) Correct the intensities of the guessed diffraction space wave function to the known values,

$$\Psi_{c,n}(\mathbf{k},\mathbf{R}) = |\Psi(\mathbf{k},\mathbf{R})|e^{i\theta_{g,n}(\mathbf{k},\mathbf{R})}, \tag{5}$$

where $|\Psi(\mathbf{k},\mathbf{R})|$ is the known diffraction space modulus.

(5) Inverse transform back to real space to obtain a new and improved guess at the exit wave function

$$\psi_{c,n}(\mathbf{r},\mathbf{R}) = \mathcal{F}^{-1}[\Psi_{c,n}(\mathbf{k},\mathbf{R})]. \tag{6}$$

(6) Update the guessed object wave function in the area covered by the aperture or probe, using the update function

$$O_{g,n+1}(\mathbf{r}) = O_{g,n}(\mathbf{r}) + \frac{|P(\mathbf{r}-\mathbf{R})|}{|P_{\max}(\mathbf{r}-\mathbf{R})|}\frac{P^*(\mathbf{r}-\mathbf{R})}{(|P(\mathbf{r}-\mathbf{R})|^2+\alpha)} \times \beta(\psi_{c,n}(\mathbf{r},\mathbf{R}) - \psi_{g,n}(\mathbf{r},\mathbf{R})), \tag{7}$$

where the parameters β and α are appropriately chosen, and $|P_{\max}(\mathbf{r}-\mathbf{R})|$ is the maximum value of the amplitude of $P(\mathbf{r})$.

(7) Move to the next position $\mathbf{R}$, for which the illumination in part overlaps that of a previous position.

(8) Repeat (2)–(7) until the sum squared error (SSE) is sufficiently small. The SSE is measured in the diffraction plane as

$$\mathrm{SSE} = \frac{(|\Psi(\mathbf{k},\mathbf{R})|^2 - |\Psi_{g,n}(\mathbf{k},\mathbf{R})|^2)^2}{N}, \tag{8}$$

where N is the number of pixels in the array representing the wave function.

The update function used in step (6) is crucial to the success of the algorithm, since it makes the effective deconvolution that occurs possible. The value α is used to prevent a divide-by-zero occurring if $|P(\mathbf{r}-\mathbf{R})|=0$. This is effectively a Wiener filter. The constant β controls the amount of feedback in the algorithm, and may be varied between roughly 0.5 and 1. Lower values of β increase the importance of the newest estimate of the object function, whereas higher values increase the importance of the previous estimate. The expression

$$\frac{|P(\mathbf{r}-\mathbf{R})|}{|P_{\max}(\mathbf{r}-\mathbf{R})|} \tag{9}$$

maximizes the effect of regions where $|P(\mathbf{r}-\mathbf{R})|$ is large. The function favors the influence of those areas of the specimen which have been strongly illuminated and attenuates the high errors which otherwise arise where the illumination was weak.

The algorithm clearly works in a similar way to other iterative phase retrieval algorithms. For the situation where $\beta=1$, and $\alpha=0$, and the function $P(\mathbf{r}-\mathbf{R})$ is a mask, or support function, the algorithm is very similar to the well-known Fienup algorithm. If in this situation, only one position $\mathbf{R}$ is used, then the algorithm reduces to being mathematically identical to the basic Fienup algorithm. However where more than one position $\mathbf{R}$ is used, the algorithm has considerable advantages over the Fienup method, including the fact that it does not suffer from problems with uniqueness, since the use of multiple probe positions effectively breaks the symmetry, which can result in a nonunique solution. Another advantage is that a wider field of view may be imaged. The end resolution achieved by the algorithm will depend on the sampling frequency in diffraction space. In cases where the diffraction

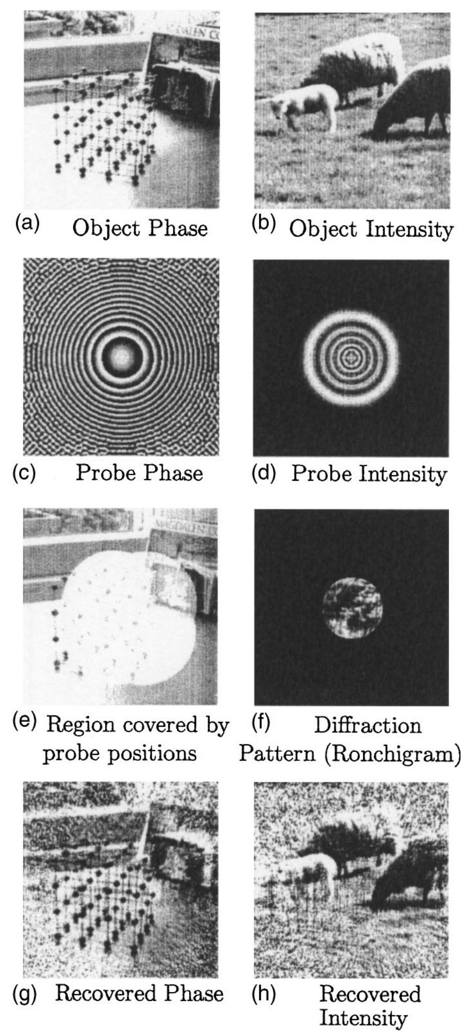

FIG. 2. Simulation of new phase retrieval method for STEM.

patterns may be measured to high angles, the algorithm will achieve higher resolution results than algorithms which depend on the use of images.

We have found that the algorithm works for a large variety of simulated objects. As an example, Fig. 2 shows the result of a simulated phase retrieval using the moving beam algorithm. The object transmission function phase [Fig. 2(a)] and intensity [Fig. 2(b)] are two unrelated data sets, creating a difficult problem where a complicated complex wave function must be retrieved. The STEM probe used is shown in Figs. 2(c) and 2(d). The probe was moved to four different positions, which together cover the input data in the region highlighted in Fig. 2(e), which is therefore the region we expect to be recovered accurately. At each position, a diffraction pattern such as that shown in Fig. 2(f) is produced. The bright disc seen is an electron Ronchigram, caused by the illuminating aperture of the lens. These diffraction patterns form the input data for the algorithm.

The algorithm was run using the parameters $\beta=1$ and $\alpha=0.0001$. After 200 iterations the object was retrieved as shown in Figs. 2(g) and 2(h), with a SSE of 1.105×10^{-5}. Clearly, the phase and intensity have been recovered accurately in the region covered by the four probe positions. The rest of the object has been recovered less well, especially in the intensity, as is expected since the probe positions used do not cover that region.

In conclusion, we have defined and demonstrated an algorithm for phase retrieval that relies only on measurements taken in diffraction space, and which is successful when the incident wave field is highly curved. In this letter we have not commented on the effect of noise and other errors on the data, although initial tests imply that the algorithm is robust in most situations. More detailed calculations will be presented elsewhere. The main benefit of the method is that the range of possible experimental implementations is large. By allowing the use of a moving illumination function, instead of a well-defined support, iterative methods can be applied to extended to objects, yet still achieve wavelength-limited resolution.

The authors are grateful for financial support from EPSRC (Grant No. GR/R75076/02).

[1] H. M. L. Faulkner and J. M. Rodenburg, Phys. Rev. Lett. **93**, 023903 (2004).
[2] J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
[3] J. M. Zuo, I. Vartanyants, M. Gao, R. Zhang, and L. A. Nagahara, Science **300**, 1419 (2003).
[4] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, Nature (London) **400**, 342 (1999).
[5] J. C. H. Spence, U. Weierstall, and M. Howells, Philos. Trans. R. Soc. London **360**, 875 (2002).
[6] P. D. Nellist, B. C. McCallum, and J. M. Rodenburg, Nature (London) **374**, 630 (1995).
[7] P. D. Nellist and J. M. Rodenburg, Acta Crystallogr., Sect. A: Found. Crystallogr. **54**, 49 (1998).
[8] L. J. Allen and M. P. Oxley, Opt. Commun. **199**, 67 (2001).
[9] H. M. L. Faulkner and J. M. Rodenburg, in Proceedings of the 13th European Microscopy Conference, 2004, Vol. 1, pp. 431–432.