Error tolerance of an iterative phase retrieval algorithm for moveable illumination microscopy

H.M.L. Faulkner, J.M. Rodenburg*

Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S1 3JD, UK

Received 21 August 2004; received in revised form 16 November 2004; accepted 22 November 2004

Abstract

This paper examines the behaviour of the new Ptychographical Iterative Engine (PIE) algorithm when part of the initial information it requires is inaccurately known. This could be the parameters describing the illuminating wavefunction, the precise location of the specimen relative to the illuminating wavefunction, or other information that is assumed about the physical system. The tolerance of the algorithm for unavoidable problems such as noise and source incoherence is also investigated, leading to the conclusion that this approach to phase retrieval is very robust. It can not only tolerate errors in the assumed parameters, but can often be used as a method of characterising the parameters more accurately.

© 2004 Elsevier B.V. All rights reserved.

PACS: 42.30.Rx; 42.25.Fx; 42.30.–d; 42.30.Kq; 42.30.Va; 42.30.Wb

Keywords: Phase retrieval; Iterative algorithms; Electron microscopy; X-ray microscopy; Ptychography

1. Introduction

Iterative phase retrieval methods have been known for many years [1,2], but have recently come to particular prominence in the context of high resolution microscopy. This is for two reasons. Firstly, it is increasingly important to be able to find the phase of a wavefunction in order to obtain high resolution information about the specimen being examined. Secondly, the iterative methods offer a particularly appealing approach to solving the phase problem. This is because they come in a variety of forms allowing many different experimental approaches to their implementation. When considering the iterative phase retrieval approach, particular attention must be paid to the popular Fienup algorithm [2] (a refinement of the Gerchberg Saxton algorithm [1]) which has been experimentally implemented by a number of
researchers [3–5]. Although the Fienup method is experimentally possible, it suffers from several limitations. These include the fact that the specimen examined must be relatively small in order to satisfy the Nyquist rate information requirement, and also that one cannot examine any particular region of interest, being limited to the exact region exposed by the support function. It is also important to note that the Fienup approach, along with most methods of phase retrieval, does not handle the situation where the wave incident onto the specimen is highly curved [6].

Another approach to phase retrieval is the method known as ptychography [7,8]. This method involves moving a scanning probe, across the specimen, to every position in an \(xy\) grid, and recording the diffraction patterns thus obtained. The result is a four-dimensional data set that can be processed to solve for the phase of the specimen transmission function. One advantage of ptychography is that the diffraction pattern information gathered can be measured out to high angles, allowing high resolution reconstruction of the specimen wavefunction. However ptychography is hampered by the difficulty of measuring the very large amount of data required to form the data set.

A new method of phase retrieval, which we call the PIE (Ptychographical Iterative Engine) algorithm has recently been described by the authors [9,10]. The approach combines the ideas behind ptychography and iterative phase retrieval to produce a powerful new technique of retrieving the phase of a wavefunction in situations where the incident radiation can be moved laterally, relative to the specimen. The early version of this algorithm worked well only when the exit function could be passed through a sharp aperture, giving a wavefunction with clearly defined zero regions. However the algorithm has since been dramatically improved, and is now effective for a much wider range of experimental situations. This algorithm surmounts the problems faced by the Fienup approach while retaining the benefits of using an iterative approach. In particular the PIE method works very well with any incident radiation, including that which is highly focused or curved, or any aberrated lens system. This means that the method is easily applied to the Scanning Transmission Electron Microscopy (STEM) configuration.

In the PIE approach, measured images or diffraction patterns taken at different positions of the incident radiation beam, are used to perform a phase retrieval. The result is a recovery of the complex transmission function of the specimen being examined, which can then be used to discover information about the structure of the specimen, such as its projected potential. The use of diffraction plane measurements means that high frequency information can be measured, and thus this method can be used to achieve super-resolution, or resolution beyond the information limit given by the transfer functions of the lenses in the microscope system. This is because an objective lens is not used to produce a focused image of the post-specimen wavefield. The information is instead inferred from diffraction data.

The PIE algorithm requires accurate knowledge of the incident radiation, which in the STEM case is the probe, in order to retrieve the specimen transmission function. The relative positions the probe is solved to must also be known so that the algorithm can function correctly. However in experimental practice it is likely that the probe positions will not be known with complete accuracy. It is also likely that some of the parameters that characterise the probe will be inaccurately measured. Other experimental problems will also arise, such as the inevitable Poisson noise and incoherence caused by the finite size of the source. These will also impact the success of the algorithm. It is important to understand the effects of all of these variables on the algorithm’s behaviour, in order to judge which parameters are the most important ones to control and characterise exactly.

2. The PIE algorithm

2.1. Description of the algorithm

In what follows, the example of a probe incident onto a specimen (as in STEM, for example) is used to illustrate the experimental arrangement (Fig. 1), however it is important to note that this is only
one of a great number of possible arrangements of apparatus. Let \( O(\mathbf{r}) \) and \( P(\mathbf{r}) \) represent two-dimensional complex functions. In what follows, \( O(\mathbf{r}) \) will physically represent an exit wave that would emanate from an object function which is illuminated by a plane wave, or equivalently the transmission function of the specimen. In the case of electron scattering considered here, \( O(\mathbf{r}) \) represents the phase and amplitude alteration introduced into an incident wave as a result of passing through the specimen examined.

\( P(\mathbf{r}) \) represents either an illumination function such as that generated by a caustic or an illumination profile formed by a lens or other optical component (e.g., as is shown in Fig. 1), it being understood that \( P(\mathbf{r}) \) is the complex stationary value of this wavefield calculated at the plane of the object function.

We assume that \( O(\mathbf{r}) \) or \( P(\mathbf{r}) \) can be moved relative to one another by various distances \( \mathbf{R} \). The notation we adopt is written in terms of moving \( P(\mathbf{r}) \), although equivalently we could instead move \( O(\mathbf{r}) \) relative to \( P(\mathbf{r}) \). In both situations, the complex value of \( O(\mathbf{r}) \) is altered by forming the product of \( O(\mathbf{r}) \) with \( P(\mathbf{r} - \mathbf{R}) \) to give a total exit wave function of \( \psi(\mathbf{r}) \), i.e.,

\[
\psi(\mathbf{r}, \mathbf{R}) = O(\mathbf{r})P(\mathbf{r} - \mathbf{R}).
\]  

This is will generally be satisfied when the specimen is sufficiently thin, or the energy of the illuminating electrons is sufficiently high.

We note that there are very few practical restrictions on either the object function or the probe function. Neither function may be a plane wave, or periodic, in the \( xy \) plane, with a repeat distance that is a multiple of the difference between different values for \( \mathbf{R} \). This is because the algorithm requires several measurements that are different in order to work. In experimental practice these criteria are easy to satisfy. Since this paper concentrates on simulating the experimental situation in STEM, we will assume the probe function has the form of a STEM probe, formed by taking the Fourier transform of an aperture corresponding to the probe-forming lens.

The algorithm works to find the phase and intensity of the complex object transmission function \( O(\mathbf{r}) \). It requires as input knowledge of the function \( P(\mathbf{r} - \mathbf{R}) \), and one or more (preferably several) measurements of the intensity of the wavefunction in a plane which is different to that containing the specimen. In this case we use data measured in the diffraction plane, which is related to the specimen plane by the Fourier transform. The measured input data is the intensities of the diffraction patterns at one or more probe positions. Using diffraction data has several advantages, including ease of collection, no requirement for focussing the exit wavefunction into an image, and the increase of resolution achieved by measuring data at high angles.

The algorithm proceeds as follows:

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 \( n \)th 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 wavefunction for position \( \mathbf{R} \),

\[
\psi_{g,n}(\mathbf{r}, \mathbf{R}) = O_{g,n}(\mathbf{r})P(\mathbf{r} - \mathbf{R}).
\]  

3. Fourier transform \( \psi_{g,n}(\mathbf{r}, \mathbf{R}) \) to obtain the corresponding wavefunction in the diffraction space plane, for that position \( \mathbf{R} \).

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

\( \mathbf{k} \) 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 wavefunction 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

$$\Psi_{g,n}(\mathbf{k}, \mathbf{R}) = |\Psi_{g,n}(\mathbf{k}, \mathbf{R})|e^{i\theta_{g,n}(\mathbf{k}, \mathbf{R})},$$  \hspace{1cm} (4)

where $|\Psi_{g,n}(\mathbf{k}, \mathbf{R})|$ is the (guessed—probably incorrect) wavefunction 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 wavefunction to the known values.

$$\Psi_{c,n}(\mathbf{k}, \mathbf{R}) = |\Psi(\mathbf{k}, \mathbf{R})|e^{i\theta_{c,n}(\mathbf{k}, \mathbf{R})},$$  \hspace{1cm} (5)

where $|\Psi(\mathbf{k}, \mathbf{R})|$ is the known diffraction space modulus, and the subscript $c$ represents the corrected wavefunction.

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

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

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

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

where the parameters $z$ and $\beta$ 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. Note that this “next” position will eventually be one of the previously-visited positions, since the algorithm cycles repeatedly through a fixed number of positions.

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

$$\text{SSE} = \frac{(|\Psi(\mathbf{k}, \mathbf{R})|^2 - |\Psi_{g,n}(\mathbf{k}, \mathbf{R})|^2)^2}{N},$$  \hspace{1cm} (8)

where $N$ is the number of pixels in the array representing the wavefunction.

The update function used in step 6 is crucial to the success of the algorithm, since it makes the effective deconvolution that occurs possible. As explained in detail in Ref. [9], the value $z$ is used to prevent a divide-by-zero occurring if $|P(\mathbf{r} - \mathbf{R})| = 0$ and the constant $\beta$ controls the amount of feedback in the algorithm.

The expression

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

maximises the effect of regions where $|P(\mathbf{r} - \mathbf{R})|$ is large favouring the influence of those areas of the specimen which have been strongly illuminated and attenuating the high errors which otherwise arise where the illumination was weak. This part of the update function is particularly important in the case of an illuminating STEM probe, since the probe is a soft function.

The algorithm works in a manner similar to other iterative algorithms that make use of multiple data sets. Each step requires the guessed object function to conform to the diffraction space criteria at that probe position, resulting in a steadily reducing SSE. The overlapping regions of the probes at different positions play an important part in fast convergence of the algorithm, since they force the results obtained at each step to fit in with those obtained in previous steps that share an overlapping region with the current probe position. The feedback parameter $\beta$ controls the amount of importance given to the result of previous steps in the algorithm. The result is a powerful method of phase retrieval that works very well for a wide variety of situations, including of course the STEM configuration modelled in this paper.

### 2.2. Example simulation with correct input information

We now demonstrate the algorithm working normally in idealised situations. Fig. 2 shows the result of a simulation where there is no error either
in the recorded data on in the known probe characteristics and positions.

The object transmission function has been created using the functions shown in Fig. 2a and e for its intensity and phase. The intensity ranges from 0.1 to 1.0, and the phase from −0.5 to 0.5. The object has a sidelength of 379 Å.

The probe, shown in Fig. 2b and f, has been created using a simple aperture of size 8.0 mrad, with an incident energy of 100 keV (which corresponds to a wavelength of 0.037 Å). A defocus of 7000 Å has been applied, and for this simulation we assume that spherical and other higher order aberrations are not present. Higher order aberrations are easily applied to simulations using our algorithm, producing similar success rates as simulations that do not include such high order terms, however for simplicity we will only consider defocus and spherical aberration in this paper.

The object is next multiplied by the probe, which is moved through a set of four different probe positions, given as follows on a 256 × 256 grid: (120, 120), (160, 160), (120, 160) and (160, 120). These produce recorded diffraction patterns such as that shown in Fig. 2c. The region covered by the probe positions is indicated in Fig. 2g.

The algorithm is run for 200 iterations, producing an absolute error per pixel of $1.105 \times 10^{-5}$, in diffraction space. This results in the recovered object intensity (d) and phase (h). Naturally, only the region covered by the combined probe positions is recovered accurately. It is often more useful to consider a scaled error value that may be easily compared to other simulations. For the purposes of this paper, we will measure the error achieved by the algorithm compared with the initial error produced by guessing the object to be a constant plane wave with intensity 1.0 everywhere. This first guess produces an error of 25006. Therefore the scaled error after 200 iterations, for the simulation shown in Fig. 2 is $0.00001105/25006 = 4.419 \times 10^{-10}$. This very small value after relatively few iterations clearly demonstrates the effectiveness of the algorithm in the ideal case simulated. A typical calculation such as this takes about 4 min to run on the author’s notebook computer.

Fig. 3 shows some variations on this ideal simulation. One of the strengths of the algorithm is that it can be used to examine as much or as little
of the object as is desired, by varying the number and values of the different probe positions that are used. Fig. 3a and e shows the result of the algorithm after 200 iterations when only one probe position has been used. In this situation the PIE algorithm is mathematically similar to the Fienup algorithm, though it uses a more sophisticated weighting function during the update step of the iteration. Obviously the area covered by the single probe position is smaller than that covered by the four positions used in the simulation for Fig. 2. What is more interesting is to note that even in the area of focus, the object is recovered less well. This is partly because with only one probe position, the algorithm is working with less input information, and is therefore expected to work less effectively.

Fig. 3b and f show the results of a simulation where two non-intersecting probe positions have been used. This is mathematically like running two separate single-probe-position versions of the PIE algorithm. While information about a larger area is recovered, the regions covered by the two probes are still not well retrieved.

In Fig. 3c and g, the two probe positions have been moved so they intersect. It is immediately apparent that allowing the probe positions to overlap improves the performance of the PIE algorithm significantly. This is because the overlapping region allows information from both the probe positions to be connected and compared. This behaviour is one of the great strengths of the PIE algorithm.

Fig. 3d and h shows the result when a number of probe positions are tiled out over a larger area of interest in the object. Nine overlapping probe positions have been used in this case. Nearly the entire object is accurately retrieved, showing the ability of the PIE algorithm to retrieve large regions when desired. It is important to note that, in a real experiment, edge effects would occur if probe positions were tiled out to the edges of the image as shown here. In practice, a large image size would be required to retrieve a very large area of the object, however this simulation demonstrates that the PIE algorithm is certainly not limited to a small number of probe positions.
3. Analysis of algorithm when assumed input information is incorrect

We now begin investigating the effect of various problems that may occur in an experiment based on this algorithm. This is done by simulating the effect of several different sources of error in the recorded diffraction patterns (noise and incoherence) and in the probe (incorrect probe parameters).

3.1. Noise occurs in experiment

The one “error” that is unavoidable in experimental reality is the noise created by the probabilistic nature of the electron wave used to illuminate the object. The recorded diffraction patterns will vary from the ideal values according to the Poisson distribution, which varies depending on the intensity of the recorded wave (i.e., the probability of each electron hitting a specific detector pixel). This effect has been simulated, with the results shown in Fig. 4. The parameter varied was the simulated exposure time of the CCD camera that would record the diffraction pattern. The electron beam current was 10 pA, with exposure times varying from 0.01 to 1.0 s. These correspond to electron counts varying from \(6.24 \times 10^5\) to \(6.24 \times 10^7\). As could be expected, the accuracy of the retrieved object is lower for the lower exposure times, because the variation in the number of actual electron counts is relatively greater. As the exposure time is increased, the accuracy of the retrieval visually improves, and the scaled error value decreases roughly in proportion to the exposure time.

For all subsequent simulations in this paper, the exposure time of 0.1 s was used. This was chosen because it is consistent with the times that could be practically used in an actual experiment.

As well as Poisson statistical noise, an experiment would normally suffer from other sources of noise as a result of imperfections in the physical processes occurring, or external environmental influences. For the purpose of this investigation, we have modelled such noise very simply, using a random distribution. For a given percentage value, \(v\%\) of noise, the recorded diffraction intensity at

Fig. 4. Effect of Poisson noise on recovered object: (a) original object intensity, (b) rec. intensity, exposure = 0.01 s, current = 10 pA, scaled error of \(2.051 \times 10^{-2}\), (c) rec. intensity, exposure = 0.1 s, current = 10 pA, scaled error of \(2.846 \times 10^{-3}\), (d) rec. intensity, exposure = 1.0 s, current = 10 pA, scaled error of \(8.934 \times 10^{-4}\), (e) original object phase, (f) rec. phase exposure = 0.01 s, current = 10 pA, (g) rec. phase exposure = 0.1 s, current = 10 pA, (h) rec. phase, exposure = 1.0 s, current = 10 pA.
Each pixel, \( I \) was altered by a value varying randomly through \( \pm \frac{1}{100} \times I \). Thus the mean contribution of the noise would be half of \( v \% \).

The effect of adding this kind of noise is shown in Figs. 5 and 6. Fig. 5 shows the recovered object intensity and phase for \( v = 5, 10, 20 \) and 50\%. While the smaller values make little or no observable difference to the accuracy of the recovered object, it is clear that the larger values of noise do degrade the performance of the algorithm. This is as expected. It is worth noting that even for the relatively large value of 20\% noise, the structure of the object is accurately recovered, clearly demonstrating the high degree of noise tolerance of this algorithm.

Fig. 6 shows a plot of the error values achieved by the algorithm after 200 iterations. The error values have been scaled relative to the initial error found by the algorithm when there is no noise included and the initial guess that the object is a plane wave with intensity of 1.0 has been used to calculate the error in the guessed diffraction pattern. The last four points on the plot correspond to the four cases shown in Fig. 5, allowing a comparison of what a particular scaled error value means in terms of retrieving the features present in the object.

![Error with random noise](image)

**Fig. 6.** Error with varying random noise. The noise varies randomly within \( \pm \) the stated value. Thus the mean noise is half the stated value.

---

**Fig. 5.** Effect of random noise on recovered object: (a) rec. object intensity \( \pm 5\% \) noise, (b) rec. object intensity \( \pm 10\% \) noise, (c) rec. object intensity \( \pm 20\% \) noise, (d) rec. object intensity \( \pm 50\% \) noise, (e) rec. object phase \( \pm 5\% \) noise, (f) rec. object phase \( \pm 10\% \) noise, (g) rec. object phase \( \pm 20\% \) noise, (h) rec. object phase \( \pm 5\% \) noise.
3.2. Probe is incoherent

The final error that is likely to be present in an experiment is some amount of incoherence in the probe. We have considered the performance of the algorithm in the presence of two different types of incoherence: finite source size in the STEM configuration and finite detector pixel size.

The presence of a finite source was modelled as follows. Several forward calculations were performed to estimate the measured diffraction pattern for a number of probe positions centred close to the nominal probe position coordinate. The intensities of these forward calculations were summed together in the detector plane. The relative contribution of each of the diffraction patterns obtained from each of the probe positions was weighted according to its distance from the nominal probe position, according to a Gaussian distribution. In real space, this has the effect of modelling in intensity a convolution of an image of the source shape function with the STEM probe. For the purposes of calculation we have assumed a Gaussian profile with a standard deviation equal to half of the simulated radius of the source probe.

The results of this simulation are shown in Fig. 7a, which shows the scaled error with changing source size radius. The most interesting point about these results is that the scaled error decreases when the source size increases slightly, before increasing again as would be expected. One possible explanation for this effect is that high-resolution data (which is most affected by source incoherence) resides at higher scattering angles than the radius of the Ronchigram. These data are also most affected by the Poisson statistics as we have modelled them. Consequently, the measure we used to estimate the error (least squares over the whole of the diffraction plane) is, ironically, slightly improved by a small degree of source incoherence.

Finite detector pixel size is not a physical manifestation of incoherence, but can be considered as an incoherence term because the detailed changes of intensity across the detector plane are integrated, or binned, into a coarser measurement. Using the principle of reciprocity, finite detector size in STEM is equivalent to finite beam convergence (i.e., incoherence in the illuminating beam) in a TEM tilt-series reconstruction, and so this nomenclature is reasonable. To test the effect of this we perform the forward calculation on a large unit cell in real space, equivalent to very fine sampling (‘fine’ pixels) in the STEM detector plane. Intensity from this calculation is then binned into ‘real’ detector pixels which are a factor of \( N \) larger in linear dimensions than the fine pixels (therefore each containing \( N^2 \) fine pixels). The reconstruction is calculated using the ‘fine’ detector pixel density (and equivalent field of

![Variation with probe incoherence](image1)

![Variation with detector pixel size](image2)

Fig. 7. Error with increasing incoherence: (a) error with increasing source size, (b) error with increasing detector pixel size (binning factor).
view in real space). The underlying phase of these small pixels is retained during the iteration. The real (forward calculation) experimental data is used to update the moduli of the fine pixels as follows. At step 4 of the iteration, the present estimate of the intensity in a particular real pixel is calculated by summing the squares of each fine pixel within the real pixel, to give a current estimated intensity for that pixel, which we call $I_e$. The moduli of all the fine pixels within this real pixel are then scaled by the square root of $I_d/I_e$, where $I_d$ is the actual intensity measured over that real pixel. The algorithm is otherwise identical to that described in Section 2.

The results from the simulation of finite detector pixel size are shown in Fig. 7b for binning factors, $N$, of up to 5. It is clear that the final error increases dramatically as the binning factor increases, which is unsurprising, given that the number of ‘fine’ pixels included in each ‘real’ pixel is increasing as $N^2$. It seems that binning factors of up to $N = 2$ or $3$ will be tolerable in practice, but larger values of $N$ may badly affect the success of the algorithm.

### 3.3. Probe has incorrect defocus, spherical aberration or aperture size

We will now explore the behaviour of the algorithm when the probe parameters are assumed to have values that have been incorrectly characterised. Despite all care, this could happen in an experimental situation. It is important to know what the result will be in terms of the success or otherwise of the algorithm. One important question is whether it is possible to use the algorithm to identify parameters that have been incorrectly measured, and to characterise those parameters more accurately. The results presented in Fig. 8 show the effect on the accuracy of the phase retrieval when one of three probe parameters (defocus, spherical aberration and size of the probe-forming aperture) is inaccurately known. The algorithm was only permitted to run to 200 iterations, although greater accuracy could be achieved with more iterations.

It is clear that inaccuracy of any of these parameters has a detrimental effect on the algorithm, since the phase retrieval is working with incorrect information. However it can also be seen that using an improved value of any of these parameters in general decreases the resulting error in the algorithm. Therefore it is possible to use small variations in the starting parameters to test the accuracy of the known parameters. Where one is incorrect, this will become quickly apparent, and the initial guess can be improved immediately, thus increasing the success of that and future experiments using the same probe parameters.

This is particularly true of the defocus of the probe, shown in Fig. 8a, which is the most important probe parameter in this experiment. The minimum error is obtained for the correct defocus value of 7000 Å, and the error increases in a monotonic fashion as we move away from this correct value.

The behaviour of the algorithm when the spherical aberration is inaccurate, as seen in

![Fig. 8. Error results of iterative algorithm (after 200 iterations) when varying known probe parameters away from the correct value: (a) error with varying de-focus. Correct defocus was 4000 Å, (b) error with varying $C_s$. Correct $C_s$ was 0.5 mm, (c) error with varying aperture size. Correct aperture was 8.0 mrad.](image-url)
Fig. 8(b), is similar, though the variation in error is not so large. The two plots on the graph show the error varying with the coefficient of spherical aberration ($C_s$) for two different sizes of the probe-forming aperture. We expect to see more error due to inaccurate characterisation of the spherical aberration when the aperture is larger, and that is indeed the case. This is because the higher order nature of spherical aberration (quartic function) means it has a much greater effect at the extremes of the probe-forming lens. The other important thing to note about Fig. 8 is that when the spherical aberration is approximately known, there is very little difference in the performance of the algorithm. As with the defocus, some improvement of the characterisation of the $C_s$ parameter may be possible, however this is likely to be less accurate because the error varies less steeply. In general the algorithm is less sensitive to having incorrectly measured spherical aberration in the probe than it is to having incorrectly measured probe defocus.

The behaviour of the algorithm when the size of the probe-forming aperture is inaccurately known is shown in Fig. 8. The general pattern is the same as that for the defocus, in that the error consistently increases as the aperture size is changed from the correct value of 8.0 mrad. However the curve is much more uneven, showing a pattern of clear steps in error. This effect is due to pixellation in the data arrays used, since the apertures are small enough for the discrete pixel sizes to have a large impact on the number of pixels included within the aperture region.

3.4. Probe positions are incorrect

Another possible inaccuracy that could occur in an experiment is not knowing the actual probe positions used when recording the different diffraction patterns. Note that only the relative probe positions matter. If all positions are translated by the same amount, the recovered object will be a translated version of the original object, with no effect on the recovered structure. What creates a problem is when one or more probe positions are incorrect relative to the others.

The plot in Fig. 9 shows the algorithm's behaviour when one of four probe positions is incorrectly known. The error increases sharply when the probe position is incorrect, indicating that the algorithm is fairly sensitive to incorrect probe positions. However it is easy to identify the correct probe position, since the error decreases monotonically in each direction around the correct location.

4. Conclusions

This paper has demonstrated the operation and behaviour of the PIE algorithm in a number of different simulated situations, all of which mimic various problems that could occur in an experiment. The results allow us to conclude that the new algorithm is tolerant of both noise and incorrect initial assumptions about the probe parameters.

The first part of the investigation explored the behaviour of the PIE algorithm when statistical and random noise are introduced into the simulated input data. When Poisson statistics or general random noise are introduced into the simulation, the algorithm continues to recover the details of the object function for relatively high values of such noise. When the Poisson statistics are varied by changing the exposure time of the detector, the algorithm behaves less well with a shorter exposure, as would be expected in any
experiment. For an exposure of 0.1 s, a satisfactory result is obtained. This value is well within possible experimental parameters, suggesting that the PIE algorithm will not be badly affected by whatever Poisson noise is present in an actual experiment. The inclusion of random noise in the simulation suggests that experimental noise will not be a significant or limiting problem with this approach, since noise values of up to 20% still allow the main details of the object to be recovered.

Additionally, the effects of incoherence caused by finite source size and detector pixel size were investigated. The PIE algorithm copes well with these problems, retrieving the phase accurately for an experimentally feasible source size. The detector pixel size has a more dramatic effect, and we conclude that the binning factor should be kept below 3 for accurate results.

The second part of this investigation focussed on the behaviour of the algorithm when the incident probe or its positions are incorrectly known. The effect of varying the defocus, spherical aberration and aperture size away from its correct value is explored. In each of these cases it is discovered that the scaled diffraction space error produced by the algorithm increases nearly monotonically as the parameter is moved further from the correct value. Thus we conclude that it would be possible to more accurately characterise a wrong parameter by varying it slightly and observing whether the algorithm produced a smaller error after the same number of iterations. This could be a useful aspect of the PIE approach.

The PIE algorithm turns out to be fairly insensitive to variation in the parameters of the incident beam. It is particularly insensitive to spherical aberration, especially when the probe-forming aperture is small. It is somewhat more sensitive to wrongly characterised defocus, though not to a degree that would affect the success of the algorithm in an experimental situation. The PIE algorithm is most sensitive to an incorrectly characterised size of the probe forming aperture. Fortunately, the size of this aperture can be accurately known in most experimental situations, since other microscopic techniques can be used to measure the aperture first.

We can therefore conclude that none of the errors examined in this paper will be a significant barrier to the success of the PIE algorithm in STEM experiments. This method, translated into experimental reality, will be a significant step towards achieving wavelength-limited resolution in electron microscopy.

References