PHYSICAL REVIEW B **82**, 121104(R) (2010)

Phase retrieval based on wave-front relay and modulation

Fucai Zhang* and J. M. Rodenburg
Department of Electronic & Electrical Engineering, University of Sheffield, Sheffield S1 3JD, United Kingdom
(Received 15 July 2010; published 20 September 2010)

We report and demonstrate experimentally an approach to retrieving the phase of a general complex-valued wave field from a single diffraction pattern. The approach employs a modulator in its data acquisition, which greatly reduces the dynamic range requirement of the detector and also greatly facilitates the inverse calculation. The new algorithm, involving a nonlinear modulus constraint, is free from ambiguities and robust to noise; it converges rapidly even with a rather loose support constraint. This approach provides a practical solution to coherent imaging with a broad range of radiations and at all wavelengths.

DOI: 10.1103/PhysRevB.82.121104 PACS number(s): 42.30.Rx, 42.30.Kq, 42.30.Va

Coherent diffraction imaging (CDI) has emerged as a promising technique for providing structural information of materials at nanometre length scales and femtosecond time scales.[1–5] A key issue in CDI is the retrieval of the phase information of the diffracted wave from its intensity measurement.[6,7] Conventional algorithms are mostly plagued with inherent ambiguities and are confined to isolated specimens. For objects that have soft edges or a large phase variance, their convergence is relatively slow and may even run into stagnation.[8,9] In some particular situations, the convergence could be improved by applying additional constraints in the real space. It has also been shown that using curved illumination with a known curvature is advantageous for the reconstruction.[10] However, reconstructing a general complex-valued object reliably and rapidly from a single diffraction pattern is far from being routine.

Another big drawback with current CDI techniques is the extremely high dynamic range requirement for detectors. A beam stop is usually required to block the direct beam, resulting in the "missing data" problem that further increases the difficulty in inverting the data.[11] By combining data recorded with different exposure times, it is possible to build a diffraction pattern of high dynamic range; but this requires a stable setup and will considerably increase the total experimental time and the sample damage.

Here we propose an approach that is free from the above drawbacks. The method converges rapidly to the solution for general complex-valued objects and has a less stringent dynamic range requirement on detectors.

Figure 1 shows the experimental setup. Three planes are defined in this configuration: the modulator, the detector, and a plane called the "entrance plane" in which the wave front is known to have a finite extent. The modulator has a known transmission function that is designed, in one respect, to diffract the incident wave into a large solid angle, yielding a flattened diffraction intensity pattern. Consequently, the dynamic range requirement of the detector is reduced.

The phase recovery process starts with an initial random guess of the entrance wave φ_0 and proceeds as follows:

(1) Apply the support constraint,

$$\varphi_{n+1} = \varphi'_n S + \beta(\varphi'_n - \varphi_n)(1 - S), \tag{1}$$

where φ_n and φ'_n are the current and the updated estimates of the entrance wave in the nth $(n=0,1,2,\ldots)$ iteration; S denotes the support area where the entrance wave is assumed to have significant values. The parameter β alters the strength of feedback and takes values in the range of [0.4, 0.9].

(2) Propagate φ_{n+1} to the modulator plane. (3) Modulate with the function of the modulator. (4) Propagate the wave field to the detector plane, yielding $\varphi^D_{n+1}=A_{n+1}\exp(i\phi_{n+1})$, where A_{n+1} and ϕ_{n+1} denote the amplitude and the phase of the diffracted wave, respectively.

(5) Apply the magnitude constraint,

$$\varphi'^D_{n+1} = I^\gamma \exp(i\phi_{n+1}), \tag{2}$$

where I is the recorded diffraction intensity and a new parameter γ is introduced. We have found that a large value of γ leads to a much more rapid convergence. Here we first run n_1 iterations with a large value of γ in the range of [0.5, 2] and then finish with $\gamma=0.5$.

(6) Backpropagate the wave to the modulator plane. (7) Remove the effect of the modulator. (8) Backpropagate to the entrance plane, yielding an updated estimate of the entrance wave φ'_{n+1}.

The steps 1–8 are iterated until the improvement between sequential estimates of the entrance wave becomes sufficiently small. The update formula of Eq. (1) is similar to that in the HIO algorithm[6] but takes a different form. The new form is selected to comply with the introduction of the parameter γ. In step 5, the parameter γ is changed stepwise. In fact we have found that gradually reducing γ to the value 0.5 as the iteration proceeds gives a better overall rate of convergence. The advantageous effect of using a nonlinear modulus constraint has already been observed in Ref. 12. This should not be surprising in view of the nonlinear nature of the phase retrieval problem itself. Detailed investigation of this will be the subject of further work. For the results shown below, the parameter β and the starting value of γ are set to 0.6 and 1.6, respectively.

The use of a modulator and the corresponding wave propagation between the entrance plane and the modulator are essential when comparing the approach with other CDI schemes. As mentioned above, conventional CDI techniques are accompanied with inherent ambiguities like the translation and the Hermite conjugation. The competition among the correct and these ambiguous solutions accounts for their slow convergence and possible stagnation.[9] In the proposed

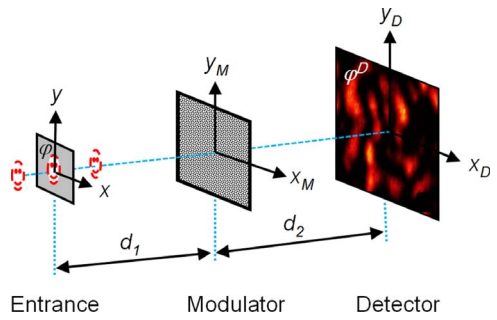

FIG. 1. (Color online) Schematic of the experimental setup; the smiley symbol shows the possible object positions; the wave at the entrance plane is known to have a finite extent.

method, these ambiguous components still occur at the plane just behind the modulator. However, after applying the division operation in step 7, their relationships with the correct component break down; when they are propagated to the entrance plane, most of their energy falls outside the support region. Therefore the effectiveness of the support constraint is substantially increased in refining the estimate of the entrance wave.

The performance of the method was first investigated using simulated intensity data generated for various object waves. As an example, we show results of retrieving the exit wave of an extended object illuminated by a curved beam. Figure 2(a) shows the magnitude transmission of object, scaled to [0, 1]. The object phase had the same spatial distribution as the magnitude but rotated by 90° clockwise and scaled to $[0,8\pi]$. Figure 2(b) shows the probe amplitude at the object. Conventional CDI methods would face grave difficulties when dealing with this situation because of the soft probe and the large range of phase variations.

Unlike lens-based imaging techniques where the lenses must be of a particular form in order to provide a good image, there is great flexibility in the design of the modulator. A number of different modulator designs have been successfully tested, including phase gratings, and aberrated lenses. Here, we show results for a phase plate consisting an array of pixels that randomly take values of either 0 or π with the same probability.

Let us assume that the detector has $N \times N$ pixels, each of

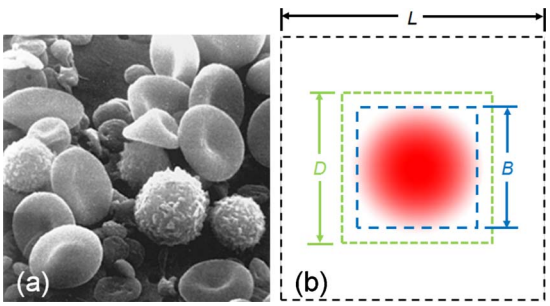

FIG. 2. (Color online) Test object and illumination probe (a) map for the object magnitude transmission scaled to [0, 1] and the phase retardance, rotated, and scaled to $[0,8\pi]$; and (b) illumination probe amplitude with a soft boundary, wherein B, D, and L are the side length of object, support constraint, and field of view.

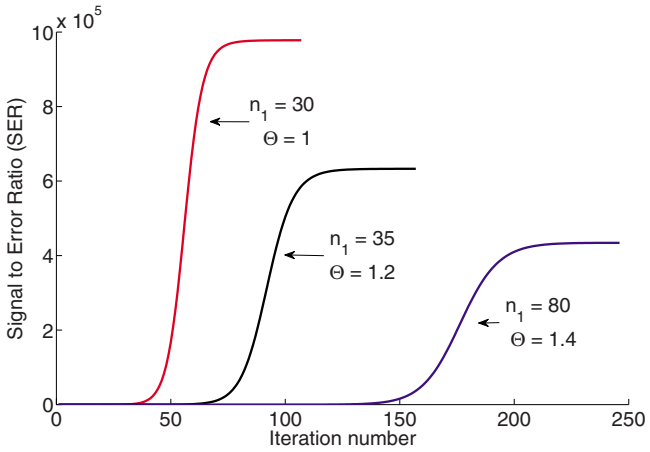

FIG. 3. (Color online) Convergence dependence on the support looseness Θ defined as the area ratio of the support and the object extent.

which are square and Δx_D across; according to the Fresnel propagation algorithm, the sampling intervals at the modulator Δx_M and at the entrance plane Δx are

$$\Delta x_M = \lambda d_2 / N \Delta x_D, \tag{3}$$

$$\Delta x = \lambda d_1 / N \Delta x_M = (d_1/d_2)\Delta x_D, \tag{4}$$

where λ is the wavelength of the radiation employed; d_1 and d_2 the distances between the three planes (see Fig. 1). For the far-field recording geometry as in the experiments with x rays and electrons, the Fresnel propagator reduces to the Fourier propagator, and Eqs. (3) and (4) still hold. A diffraction pattern was calculated with the following parameters: λ=635 nm; d_1=9.7 mm; d_2=47.7 mm; N=256; and Δx_D =7.4 μm in agreement with the experimental conditions. It was then quantized to 2^{12} levels to simulate the finite dynamic range of a detector.

The convergence of the algorithm is monitored using the signal-to-error ratio (SER) defined as

$$\mathrm{SER} = \sum |\varphi_{test}|^2 \Big/ \sum (|\varphi_n| - |\varphi_{test}|)^2, \tag{5}$$

where φ_{test} is the known test entrance wave. Compared to the normalized rms error metric commonly used in the literature, the SER is adapted to account for the introduction of the parameter γ and is able to represent the convergence behavior of the algorithm more effectively.

In practice, it may be difficult to locate the true boundary of the entrance wave. A factor accounting for the degree of looseness of the support constraint is introduced,

$$\Theta = (D/B)^2, \tag{6}$$

where B and D are the linear dimensions of the entrance wave extent and the support as depicted in Fig. 2(b).

Figure 3 shows the convergence curves for three Θ values. One can see the algorithm converges rapidly even with considerable uncertainty in the *a priori* knowledge of support. In the case of Θ=1.4, we had B=116 and D=138; the support was 11 pixels broader than the real boundary of the entrance wave on all sides. Note that the smallest square

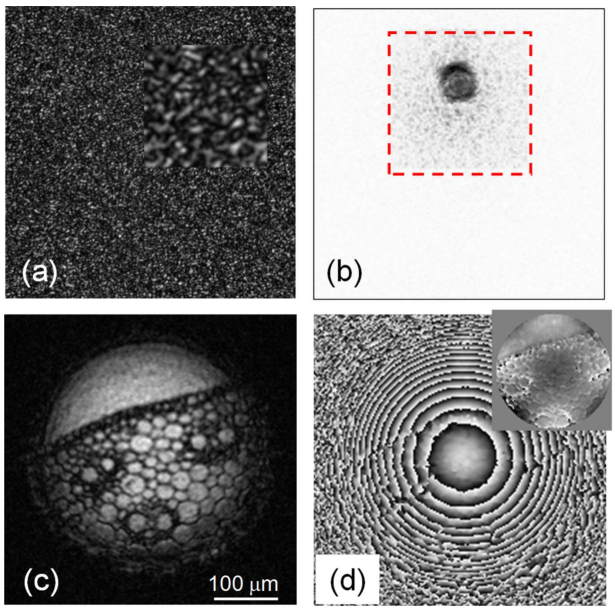

FIG. 4. (Color online) Reconstruction of a monocotyledon sample; (a) recorded diffraction pattern, inset is a closeup view; (b) amplitude at the plane where the support constraint was applied (in reversed gray color map); the dashed square shows the boundary of support constraint; (c) amplitude; and (d) phase of the sample exit wave, inset shows the phase retardance due to the sample.

surrounding the entrance wave was used for specifying its extent. If the actual extent were used, the value of Θ would be even larger. We also found that applying the support constraint at a plane where the wave front is mostly localized in space gave better convergence and image quality even if the size of support constraint (in pixels) remains the same.

In principle, the approach is applicable to a broad range of radiations and wavelengths, such as high-energy electrons and x-ray photons. Here, we experimentally demonstrate it using a diode laser with a wavelength of $\lambda=635$ nm. A lens ($f=50$ mm) was used to form a convergent illumination. A phase plate made of silica glass was placed about 18.5 mm behind the illumination focus. The plate consisted of 1100×1100 square pixels with a size of 16 μm, which deliver 0 or π phase shift for the employed wavelength.[13] A 14-bits charge-coupled-device camera consisting of square pixels of size 7.4 μm was placed 70 mm further downstream from the plate. The test sample, a microscopic monocotyledon specimen, was situated about 1.5 mm before the illumination focus. Figure 4(a) shows a part of the entire recorded diffraction pattern, of which the central 376×376 samples were cut out and used in the reconstruction. The number 376 was calculated according to Eq. (3). Figures 4(c) and 4(d) show the reconstructed amplitude and phase images of the sample exit wave after 50 iterations. The illumination wave on sample was reconstructed in the same way from a diffraction pattern recorded without the sample. Subtraction of the illumination phase from the phase of the sample exit wave yields the sample phase as shown in the inset in (d). The residual phase that one could see is due to the 1-mm-thick sample plate that altered the actual probe on the sample. A very loose support was used in the reconstruction. The dashed square in Fig. 4(b) indicates the size of support constraint applied at the illumination focus plane. No support refinement algorithm was used here, such as the shrink-wrap algorithm.[14] Adoption of such algorithms may lead to even more rapid convergence.

The function of the modulator and the geometric setup parameters are the only *a priori* information required in the technique. In our experiment, the phase plate was characterized by Ptychography[15] using the ePIE algorithm.[16] Therefore, there is no stringent accuracy requirement on the modulator fabrication. In practice, a large feature size of the modulator is desirable in order to facilitate easy manufacture. If the coherence performance of the radiation source is not a limitation, the plate feature size can be freely selected by changing the distance d_1 and d_2. In fact, the feature size of plate can be selected to be much larger than Δx_M. We have tested binning the modulator pixels. The same amplitude and phase maps as in Fig. 2(a) but resized to one third of the whole view ($B/L=1/3$), were used as test entrance wave. For a modulator pixel size of $4\times$ and $6\times$ of Δx_M, the required number of iterations to get an SER value of 100 for $\Theta=1$ was 44 and 223, respectively. For a weakly diffractive object illuminated by a curved beam, it is possible to use an even bigger modulator pixel size. Consider a possible experiment using 8 keV x rays, with a desired resolution $\Delta x=10$ nm, and with $d_2=8$ m, $N=1024$, $\Delta x_D=24$ μm then according to Eqs. (3) and (4), the distance d_1 is calculated to be 0.6 mm and Δx_M to be 50 nm. In view of the binning calculation above, modulator with a feature size of about 300 nm, easily achievable with current fabrication techniques, can be used in this technique.

In conclusion, we have proposed a general approach to the phase measurement of a complex-valued wave field. The approach resolves many of the outstanding problems with current CDI methods: the extremely high dynamic range requirement of detectors and the stringent requirement on samples for which a reconstruction can be reliably obtained. The new algorithm has shown rapid convergence with real data recorded with a normal detector and using a rather loose support constraint. This approach provides a robust solution to the "phase problem" in all disciplines; in particular, it would allow for routine coherent imaging for biological and material science.

The work was supported by EPSRC, Basic Technology Grant (Grant No. EP/E034055/1), Ultimate Microscopy. The phase plate was manufactured at the Institute for Applied Optics, University of Stuttgart, Germany.

*Corresponding author; fucaizhang@gmail.com

[1] H. N. Chapman, Nature Mater. **8**, 299 (2009).
[2] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, Nature (London) **400**, 342 (1999).
[3] J. M. Zuo, I. Vartanyants, M. Gao, R. Zhang, and L. A. Nagahara, Science **300**, 1419 (2003).
[4] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. M. Neiman, and D. Sayre, Proc. Natl. Acad. Sci. U.S.A. **102**, 15343 (2005).
[5] M. A. Pfeifer, G. J. Williams, I. A. Vartanyants, R. Harder, and I. K. Robinson, Nature (London) **442**, 63 (2006).
[6] J. R. Fienup, Opt. Lett. **3**, 27 (1978).
[7] J. Miao, D. Sayre, and H. N. Chapman, J. Opt. Soc. Am. A **15**, 1662 (1998).
[8] J. R. Fienup and C. C. Wackerman, J. Opt. Soc. Am. A **3**, 1897 (1986).
[9] J. R. Fienup, J. Opt. Soc. Am. A **4**, 118 (1987).
[10] G. J. Williams, H. M. Quiney, B. B. Dhal, C. Q. Tran, K. A. Nugent, A. G. Peele, D. Paterson, and M. D. de Jonge, Phys. Rev. Lett. **97**, 025506 (2006).
[11] J. Miao, Y. Nishino, Y. Kohmura, B. Johnson, C. Song, S. H. Risbud, and T. Ishikawa, Phys. Rev. Lett. **95**, 085503 (2005).
[12] F. Zhang, G. Pedrini, and W. Osten, Phys. Rev. A **75**, 043805 (2007).
[13] C. Kohler, F. Zhang, and W. Osten, Appl. Opt. **48**, 4003 (2009).
[14] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, Phys. Rev. B **68**, 140101 (2003).
[15] H. M. L. Faulkner and J. M. Rodenburg, Phys. Rev. Lett. **93**, 023903 (2004).
[16] A. M. Maiden and J. M. Rodenburg, Ultramicroscopy **109**, 1256 (2009).