Structure retrieval with fast electrons using segmented detectors

H. G. Brown,[1] A. J. D'Alfonso,[1] Z. Chen,[2] A. J. Morgan,[3] M. Weyland,[4,5] C. Zheng,[2,6] M. S. Fuhrer,[2] S. D. Findlay,[2] and L. J. Allen[1,*]

[1]*School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia*
[2]*School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia*
[3]*Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany*
[4]*Monash Centre for Electron Microscopy, Monash University, Clayton, Victoria 3800, Australia*
[5]*Department of Materials Science and Engineering, Monash University, Clayton, Victoria 3800, Australia*
[6]*Department of Civil Engineering, Monash University, Clayton, Victoria 3800, Australia*

(Received 3 February 2016; revised manuscript received 13 April 2016; published 27 April 2016)

We introduce an algorithm for the reconstruction of the complex transmission function of a specimen using segmented detectors in scanning transmission electron microscopy geometry. The phase of the transmission function can be related to magnetic and electric fields within the specimen and is sensitive to lighter elements. The technique is demonstrated for simulated data and also using experimental datasets taken from a MoS_2 monolayer and a $SrTiO_3$ crystal. We present an extension to the algorithm to account for uncertainties in the illuminating probe. The algorithm can be implemented using fast Fourier transforms, and this provides the possibility of reconstructing specimen transmission functions in real time.

DOI: 10.1103/PhysRevB.93.134116

I. INTRODUCTION

Aberration-corrected electron microscopy has ushered in a new era of imaging at atomic resolution, and this has become an essential tool for the characterization of functional nanostructures. Often it is the phase of the wave after the probe has transited the specimen, i.e., the exit-surface wave, that encodes significant physical properties of the specimen. For example, the phase is related to electric and magnetic fields within the specimen. Consequently, much effort has been devoted to reconstructing the complex exit surface wave from one or more images or diffraction patterns, which themselves do not directly manifest the phase information.

Deterministic retrieval of the exit-surface wave for a CeO_2 nanoparticle from the diffraction pattern formed using a defocused coherent probe was recently demonstrated [1]. An independent retrieval of the exit-surface wave with the probe at a second position such that the region of illumination on the specimen overlapped the first yielded consistency in the region of overlap. However, there are decided advantages in scanning a coherent probe, as is done in scanning transmission electron microscopy (STEM), to obtain a four-dimensional (4D) dataset (where a 2D diffraction pattern is obtained for each probe position while scanning the probe in two dimensions). When treated as a whole, such 4D datasets are overdetermined if there are regions of overlap of the area of the specimen illuminated for different probe positions. The quantity that it is then convenient to work with in such ptychographic approaches is the transmission function of the specimen rather than the exit-surface wave, as has been done in several applications along these lines [2–4]. The transmission function is defined as the quantity that multiplies the incident probe wave function to give the exit-surface wave function. If the specimen is thin enough, then a transmission function in a single plane will represent the effect of the specimen on the probe in a consistent way for overlapping probe positions, and the transmission function will also represent the structure of the specimen. This transmission function is the same as the exit-surface wave when the specimen is illuminated by a plane wave, assumed to be unity everywhere as it illuminates the specimen.

The advent of fast-readout 2D electron cameras has greatly facilitated the acquisition of 4D datasets in STEM [5]. The use of such a rapid detector enabled Pennycook *et al.* [6,7] to demonstrate efficient phase contrast imaging in STEM using a pixelated detector and to explore the optimization of imaging conditions by synthesizing different detectors in terms of the smaller pixels on a charge-coupled device (CCD) [8]. Diffraction contrast imaging using virtual apertures has also been explored [9]. Such results can also be compared to those obtained using fixed-configuration segmented detectors. For example, Shibata *et al.* [10–13] and Lazić *et al.* [14] use fixed-configuration detectors for differential phase contrast imaging.

In this paper, we present an implementation of segmented detector ptychography [15–19] (SDP) for the reconstruction of a transmission function of a specimen. The present approach is akin to the generalized holography method outlined in Ref. [4]. Measurements recorded in the diffraction plane using a focused coherent atomic scale probe are used to construct images corresponding to segmented detectors. These are used, in conjunction with the known illumination, to reconstruct the transmission function. The approach is deterministic, it can be implemented in a memory-efficient way in terms of fast Fourier transforms, and it can efficiently handle very large 4D datasets. However, we note that a 4D dataset taken using a fast-readout 2D electron camera is not a requirement to apply the SDP algorithm. The technique is also applicable to datasets recorded using segmented detectors [10,14,20,21]. The technique is demonstrated for simulated data and also using experimental datasets recorded for a MoS_2 monolayer and a $SrTiO_3$ crystal for two different thicknesses.

*lja@unimelb.edu.au

II. THEORETICAL FRAMEWORK

We assume that we have a specimen that is sufficiently thin to represent the scattering of the incident electrons by a transmission function $T(\mathbf{r})$ in a single plane, which, for what follows, will be convenient to write in the form

$$T(\mathbf{r}) = 1 + O(\mathbf{r}), \tag{1}$$

where we will refer to $O(\mathbf{r})$ as the object function. An atomic-scale, focused coherent electron probe placed at position $\mathbf{R}$ on the specimen is denoted by $\psi_{\text{illum}}(\mathbf{r},\mathbf{R})$. The exit-surface wave after scattering by the specimen is then given by

$$\begin{aligned}\psi_{\text{exit}}(\mathbf{r},\mathbf{R}) &= T(\mathbf{r})\psi_{\text{illum}}(\mathbf{r},\mathbf{R}) \\ &= [1 + O(\mathbf{r})]\psi_{\text{illum}}(\mathbf{r},\mathbf{R}).\end{aligned} \tag{2}$$

The exit-surface wave in the far-field diffraction plane can be written as

$$\begin{aligned}\hat{\psi}_{\text{exit}}(\mathbf{q},\mathbf{R}) &= \mathcal{F}_{\mathbf{r}\to\mathbf{q}}\{\psi_{\text{illum}}(\mathbf{r},\mathbf{R})[1 + O(\mathbf{r})]\} \\ &= \hat{\psi}_{\text{illum}}(\mathbf{q},\mathbf{R}) + \hat{\psi}_{\text{illum}}(\mathbf{q},\mathbf{R}) * \hat{O}(\mathbf{q}) \\ &= \hat{\psi}_{\text{illum}}(\mathbf{q},\mathbf{R}) + \sum_{\mathbf{g}} \hat{\psi}_{\text{illum}}(\mathbf{q}-\mathbf{g},\mathbf{R})\hat{O}(\mathbf{g}),\end{aligned} \tag{3}$$

where $\mathcal{F}_{\mathbf{r}\to\mathbf{q}}$ denotes a Fourier transform from real space to the reciprocal (diffraction) space, and the hat symbol $\hat{}$ emphasizes a reciprocal-space quantity. The convolution of $\hat{O}(\mathbf{q})$ and $\hat{\psi}_{\text{illum}}(\mathbf{q},\mathbf{R})$, denoted by the operator $*$, is a result of the Fourier convolution theorem, and a reciprocal-space coordinate $\mathbf{g}$ has been introduced when explicitly writing out this convolution in discretized form in the final line of the equation.

Assume now that we have acquired a 4D dataset using a pixelated detector. We can synthesize detectors in terms of sets of pixels that define each detector, and we can construct an image as a function of probe position $\mathbf{R}$ for each such detector j in the following way:

$$I_j(\mathbf{R}) = \int |\hat{\psi}_{\text{exit}}(\mathbf{q},\mathbf{R})|^2 D_j(\mathbf{q}) d\mathbf{q}, \tag{4}$$

where $D_j(\mathbf{q})$ is a binary function defining the region of the diffraction pattern spanned by the synthesized detector.

Inserting Eq. (3) into Eq. (4), we obtain an expression for the segmented detector signal,

$$\begin{aligned}I_j(\mathbf{R}) = &\int |\hat{\psi}_{\text{illum}}(\mathbf{q})|^2 D_j(\mathbf{q}) d\mathbf{q} \\ &+ \sum_{\mathbf{g}} \hat{O}(\mathbf{g}) e^{2\pi i \mathbf{g}\cdot\mathbf{R}} \int \hat{\psi}^*_{\text{illum}}(\mathbf{q})\hat{\psi}_{\text{illum}}(\mathbf{q}-\mathbf{g}) D_j(\mathbf{q}) d\mathbf{q} \\ &+ \sum_{\mathbf{g}} \hat{O}^*(\mathbf{g}) e^{-2\pi i \mathbf{g}\cdot\mathbf{R}} \int \hat{\psi}_{\text{illum}}(\mathbf{q})\hat{\psi}^*_{\text{illum}}(\mathbf{q}-\mathbf{g}) D_j(\mathbf{q}) d\mathbf{q} \\ &+ \int \left| \sum_{\mathbf{g}} \hat{\psi}_{\text{illum}}(\mathbf{q}-\mathbf{g}) e^{-2\pi i (\mathbf{q}-\mathbf{g})\cdot\mathbf{R}} \hat{O}(\mathbf{g}) \right|^2 D_j(\mathbf{q}) d\mathbf{q},\end{aligned} \tag{5}$$

where we have used the decomposition $\hat{\psi}_{\text{illum}}(\mathbf{q},\mathbf{R}) = \hat{\psi}_{\text{illum}}(\mathbf{q}) e^{-2\pi i \mathbf{q}\cdot\mathbf{R}}$. Examining Eq. (5), we can identify the first term as the wave component that is not scattered by the specimen. The next two terms are cross terms (and are the complex conjugate of one another) that are linear with respect to the function $\hat{O}(\mathbf{g})$. The final term, nonlinear with respect to the function $\hat{O}(\mathbf{g})$, is assumed to be negligible and is ignored in subsequent equations. This is equivalent to making a weak phase approximation, but it is possible to correct for this after an initial solution, without the nonlinear term, has been obtained [22]. With this approximation, and defining

$$\hat{\phi}_j(\mathbf{g}) = \int \hat{\psi}^*_{\text{illum}}(\mathbf{q})\hat{\psi}_{\text{illum}}(\mathbf{q}-\mathbf{g}) D_j(\mathbf{q}) d\mathbf{q}, \tag{6}$$

we can rewrite Eq. (5) in the form

$$\begin{aligned}b_j(\mathbf{R}) &= I_j(\mathbf{R}) - \int |\hat{\psi}_{\text{illum}}(\mathbf{q})|^2 D_j(\mathbf{q}) d\mathbf{q} \\ &\approx \sum_{\mathbf{g}} \hat{O}(\mathbf{g}) e^{2\pi i \mathbf{g}\cdot\mathbf{R}} \hat{\phi}_j(\mathbf{g}) + \hat{O}^*(\mathbf{g}) e^{-2\pi i \mathbf{g}\cdot\mathbf{R}} \hat{\phi}^*_j(\mathbf{g}) \\ &= 2\sum_{\mathbf{g}} \text{Re}\{\hat{O}(\mathbf{g}) e^{2\pi i \mathbf{g}\cdot\mathbf{R}} \hat{\phi}_j(\mathbf{g})\},\end{aligned} \tag{7}$$

where the quantities $b_j(\mathbf{R})$ are constructed from the synthesized images and the illumination. The illumination is assumed known, but if it is inadequately characterized at the outset, the Appendix suggests a scheme whereby corrections can be made to a reasonable guess of the illumination function. Equation (7) is a set of linear equations for the unknown Fourier coefficients $\hat{O}(\mathbf{g})$, a fact that is made more manifest by the following matrix formulation:

$$\begin{bmatrix} [2\,\text{Re}\{e^{2\pi i \mathbf{g}\cdot\mathbf{R}}\hat{\phi}_1(\mathbf{g})\}] & [-2\,\text{Im}\{e^{2\pi i \mathbf{g}\cdot\mathbf{R}}\hat{\phi}_1(\mathbf{g})\}] \\ [2\,\text{Re}\{e^{2\pi i \mathbf{g}\cdot\mathbf{R}}\hat{\phi}_2(\mathbf{g})\}] & [-2\,\text{Im}\{e^{2\pi i \mathbf{g}\cdot\mathbf{R}}\hat{\phi}_2(\mathbf{g})\}] \\ \vdots & \vdots \\ [2\,\text{Re}\{e^{2\pi i \mathbf{g}\cdot\mathbf{R}}\hat{\phi}_n(\mathbf{g})\}] & [-2\,\text{Im}\{e^{2\pi i \mathbf{g}\cdot\mathbf{R}}\hat{\phi}_n(\mathbf{g})\}] \end{bmatrix} \times \begin{bmatrix} [\text{Re}\{\hat{O}(\mathbf{g})\}] \\ [\text{Im}\{\hat{O}(\mathbf{g})\}] \end{bmatrix} = \begin{bmatrix} [b_1(\mathbf{R})] \\ [b_2(\mathbf{R})] \\ \vdots \\ [b_n(\mathbf{R})] \end{bmatrix}. \tag{8}$$

This is a set of linear equations of the form $\mathbf{Ax} = \mathbf{b}$. For each of the n_D detectors, there is a "row" of two subblocks written in terms of real and imaginary parts in the coefficient matrix $\mathbf{A}$. Within the subblocks of $\mathbf{A}$, the columns vary with the reciprocal-space coordinate $\mathbf{g}$ and the rows by the probe position $\mathbf{R}$. Each subblock has dimension $n_{\mathbf{R}}$, the number of probe positions, by $n_{\mathbf{g}}$, the number of Fourier coefficients $\hat{O}_{\mathbf{g}}$, which is half the dimension of the vector of unknowns $\mathbf{x}$ (since we are solving for the real and imaginary parts of $\hat{O}_{\mathbf{g}}$ separately). The vector $\mathbf{b}$ contains the different $b_j(\mathbf{R})$ from Eq. (7) and has dimension $n_D \times n_{\mathbf{R}}$. To have an overdetermined set of equations, we require $n_D \times n_{\mathbf{R}} > 2n_{\mathbf{g}}$, which is ensured by sampling above the Nyquist frequency where $n_{\mathbf{R}} > 2n_{\mathbf{g}}$. It may be possible to overdetermine the problem with sampling less than the Nyquist frequency by using a sufficient number of detectors.

The details of the probe, which define the matrix on the left-hand side of Eq. (8), via Eq. (6), are important in determining contrast. The probe aperture should contain the significant Fourier coefficients pertinent to the structure of the object we are seeking to retrieve. However, making the probe much

narrower than the features of interest is likely to reduce the sensitivity of the technique. In the extreme case of a δ-function probe, perfect phase objects give no contrast, irrespective of the segmentation of the detector.

Equation (8) can be solved using conventional approaches such as QR or LU decomposition or singular value decomposition (SVD) [23]. However, there can be both memory and speed advantages in using the conjugate gradient least-squares (CGLS) algorithm. Furthermore, the CGLS algorithm has some favorable properties for regularization in the presence of noise. The CGLS method proceeds by constructing a series of vectors (in a so-called Krylov subspace) that are dependent on both the matrix $\mathbf{A}$ and the experimental data $\mathbf{b}$. As a result, the Krylov subspace is able to adapt itself optimally to the input data, i.e., it is more accommodating of noise [24].

The CGLS algorithm [23] begins with the following initialization:

$$\mathbf{x}_0 = \mathbf{0},\ \mathbf{d}_0 = \mathbf{b},\ \mathbf{r}_0 = A^T\mathbf{b},\ \mathbf{p}_0 = \mathbf{r}_0,\ \mathbf{t}_0 = A\mathbf{p}_0. \tag{9}$$

Here the initial guess is $\mathbf{x}_0 = \mathbf{0}$, but any starting guess produces the same least-squares solution within a number that is equal to the number of unknowns ($2 \times n_{\mathbf{g}}$). For each iteration i, the intermediate quantities that are constructed in the CGLS algorithm are defined by

$$\begin{aligned}
&\alpha_i = \|\mathbf{r}_{i-1}\|^2/\|\mathbf{t}_{i-1}\|^2, \quad \mathbf{x}_i = \mathbf{x}_{i-1} + \alpha_i\mathbf{p}_{i-1},\\
&\mathbf{d}_i = \mathbf{d}_{i-1} - \alpha_i\mathbf{t}_{i-1}, \quad \mathbf{r}_i = \mathbf{A}^T\mathbf{d}_i,\\
&\beta_i = \|\mathbf{r}_i\|^2/\|\mathbf{r}_{i-1}\|^2, \quad \mathbf{p}_i = \mathbf{r}_i + \beta_i\mathbf{p}_{i-1}, \text{ and}\\
&\mathbf{t}_i = \mathbf{A}\mathbf{p}_i.
\end{aligned} \tag{10}$$

Numerically, $\mathbf{x}$ might be solved by explicit construction of $\mathbf{A}$ and $\mathbf{b}$ and by following the algorithm in Eq. (10). Alternatively, in this case, due to the structure of $\mathbf{A}$, explicit construction of $\mathbf{A}$ can be avoided by casting the operations $\mathbf{A}\mathbf{p}_i$ and $\mathbf{A}^T\mathbf{d}_i$ in terms of Fourier transforms, as was done in Ref. [1]. We note that multiplication by the matrix $\mathbf{A}$ maps a complex object, $O(\mathbf{g})$ in Eq. (8), to a set of real images for each detector j. Multiplication by the matrix's transpose $\mathbf{A}^T$ maps a set of real images back to a complex object. Consider the multiplication of $\mathbf{p}_{i,j}$, the portion of vector $\mathbf{p}_i$ corresponding to $\hat{\phi}_j(\mathbf{r})$, by the relevant subblock of matrix $\mathbf{A}$ in Eq. (10). From Eq. (7), the multiplication by this subblock of matrix $\mathbf{A}$ may be expressed as

$$\begin{aligned}
\mathbf{A}\mathbf{p}_{i,j} &= \sum_{\mathbf{g}} \hat{\mathbf{p}}_{i,j}(\mathbf{g})e^{2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\phi}_j(\mathbf{g}) + \hat{\mathbf{p}}^*_{i,j}(\mathbf{g})e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\phi}^*_j(\mathbf{g}),\\
&= 2\,\mathrm{Re}\{\mathcal{F}_{\mathbf{g}\to\mathbf{R}}[\hat{\mathbf{p}}_{ij}(\mathbf{g})\hat{\phi}_i(\mathbf{g})]\}.
\end{aligned} \tag{11}$$

With reference to Eq. (8), the transpose operation of $\mathbf{A}$ may be written as

$$\begin{aligned}
\mathbf{A}^T\mathbf{d}_{i,j} &= \sum_j \begin{bmatrix} 2\,\mathrm{Re}\{e^{2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\phi}_j(\mathbf{g})\} \\ -2\,\mathrm{Im}\{e^{2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\phi}_j(\mathbf{g})\} \end{bmatrix}\big[d_{i,j}(\mathbf{R})\big]\\
&= \begin{bmatrix} \mathrm{Re}\{\mathbf{r}_i(\mathbf{g})\} \\ \mathrm{Im}\{\mathbf{r}_i(\mathbf{g})\} \end{bmatrix}.
\end{aligned} \tag{12}$$

The rows of the matrix in Eq. (12) now vary with respect to $\mathbf{g}$, and the columns now vary with $\mathbf{R}$. Expanding the matrix multiplication and noting that $d_{i,j}(\mathbf{R})$ is a purely real quantity,

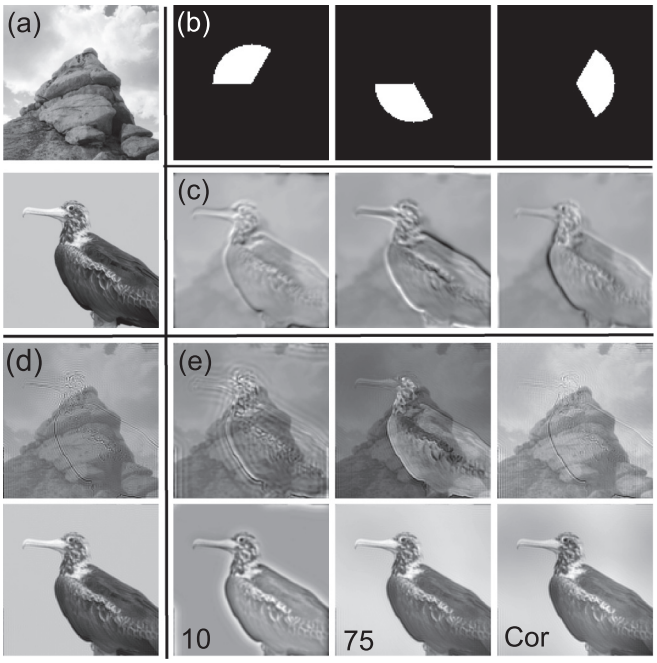

FIG. 1. Reconstruction of a model transmission function using a focused coherent probe for the imaging conditions described in the text. (a) Model transmission function intensity (top) and phase (bottom). (b) Segments in the bright field are used to construct the images in (c). (d) The band-pass-limited transmission function intensity (top) and phase (bottom) can be compared with the retrieved transmission function in (e) after 10 and 75 CGLS iterations and also with the latter as corrected for neglect of the nonlinear term in Eq. (5).

the matrix multiplication may be expressed explicitly as

$$\begin{aligned}
\mathbf{r}_i(\mathbf{g}) &= 2\sum_{j,\mathbf{R}} \mathrm{Re}[e^{2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\phi}_j(\mathbf{g})d_{i,j}(\mathbf{R})]\\
&\quad - i\,\mathrm{Im}[e^{2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\phi}_j(\mathbf{g})d_{i,j}(\mathbf{R})]\\
&= 2\sum_j \hat{\phi}^*_j(\mathbf{g})\sum_{\mathbf{R}} e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}d_{i,j}(\mathbf{R})\\
&= 2\sum_j \hat{\phi}^*_j(\mathbf{g})\hat{d}_{i,j}(\mathbf{g}).
\end{aligned} \tag{13}$$

As is well known, consideration of Eq. (5) shows that the maximum resolution in the image $I_j(\mathbf{R})$ is twice the magnitude of the maximum reciprocal-space frequency, $g_{\max}$, allowed by the probe-forming aperture. Then both $\hat{\phi}_j(\mathbf{g})$ and $b_j(\mathbf{R})$ will be band-pass-limited by $2g_{\max}$. The object function $\hat{O}(\mathbf{g})$ will be calculated from $b_j(\mathbf{R})$ through successive convolution and correlation operations with $\hat{\phi}_j(\mathbf{g})$ according to Eqs. (11) and (13), which will preserve this band-pass limit. The maximum reciprocal space resolution of the retrieved transmission function will also be $2g_{\max}$.

Before applying the algorithm to experimental data, let us demonstrate how this algorithm works using simulated data that are free of noise. An arbitrarily chosen complex transmission function $T(\mathbf{r})$ is shown in Fig. 1(a) in terms of its intensity (above) and phase (below). We scale this to have typical values representative of elastic scattering potentials in condensed matter, and we assume we are imaging using a

200 keV probe with a convergence semiangle of 15 mrad, a field of view of 53.5 Å, and a defocus of 100 Å. Data from pixel detectors offer much flexibility in the detector configurations used to synthesize segmented detector STEM images. The averaging implied in integrating over larger detector segments reduces noise effects, but it leads to fewer constraining images. Using smaller segments leads to more constraints, but it increases the dimensions of the **A** matrix and makes the data more sensitive to noise. A detailed analysis of how to optimize the contrast transfer function in the presence of noise has been provided by Yang *et al.* [8]. For the present test, however, we choose to partition the bright-field disk into three wedges as shown in Fig. 1(b), the minimal number of segments needed such that the contrast transfer functions adequately sample the 2D object [19]. (Increasing the number of segments does not substantially alter the results.)

Using the procedure described above, the model transmission function was retrieved from the images in Fig. 1(c). There is a band-pass limit imposed by the probe-forming aperture, and we can thus only expect to retrieve the band-pass-limited transmission function shown in Fig. 1(d). After only 10 iterations, the basic phase structure is evident in the reconstruction. By 75 iterations, convergence has been reached in the phase. However, the intensity contains a larger admixture of phase information than what is present in the band-pass-limited transmission function in Fig. 1(d). This is a consequence of the reconstruction algorithm neglecting the nonlinear terms in Eq. (5). We seek to correct for this following the procedure described in Ref. [22]. The values of $\hat{O}(\mathbf{g})$ obtained from the reconstruction can be used to estimate the nonlinear term in Eq. (5). This nonlinear term is then subtracted from the right-hand side of Eq. (8), and a second inversion is performed. The results, labeled "Cor" in Fig. 1(e), show a much improved reconstruction of the amplitude as well as the phase. The nonlinear correction procedure can be further iterated, though our experience is that this is not generally convergent. Note, however, that the specimen information of most interest in electron microscopy is primarily contained in the phase, which was successfully reconstructed here without need of the nonlinear correction procedure.

With the inversion cast in terms of fast Fourier transforms, the results shown are obtained in seconds using less than a MB of memory (working on a 128×128 pixel grid). The fidelity of the retrieved transmission function with respect to the band-pass-limited transmission function in Fig. 1(d) is shown as a function of iteration number in Fig. 2 [without any correction for the neglect of the nonlinear term in Eq. (5)]. A cross indicates the fidelity of the object after correction for the nonlinear term that is displayed in Fig. 1(e). The fidelity is defined as

$$\varepsilon = \frac{\sum \|T - T'\|^2}{\sum \|T'\|^2}, \tag{14}$$

where T' is the comparison transmission function and T is the retrieved transmission function. For the perfect data of our test case, Fig. 2 shows the fidelity to continually decrease with increasing iteration number. As we shall see later, in the presence of noise the CGLS algorithm is "semiconvergent,"

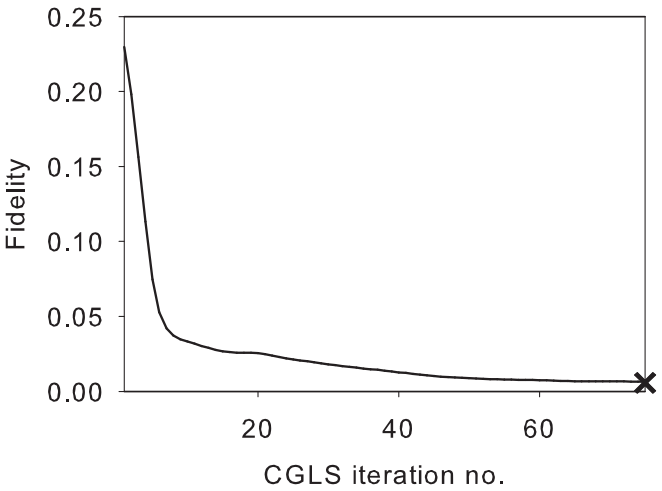

FIG. 2. Fidelity as a function of CGLS iteration number for the results shown in Fig. 1(e). The "×" indicates the fidelity after correction for neglect of the nonlinear term.

meaning that beyond a certain number of iterations the fidelity metric increases again [24].

Lastly, in this section we consider a practical issue that will be pertinent when applying SDP to experimental datasets, namely spatial incoherence due to a finite source size. The effect of the spatial incoherence may be represented by the convolution of a STEM image with a Gaussian distribution, $G(\mathbf{R})$,

$$\begin{aligned} I'_j(\mathbf{R}) &= G(\mathbf{R}) * I_j(\mathbf{R}) \\ &= \int G(\mathbf{R} - \mathbf{R}') I_j(\mathbf{R}') d\mathbf{R}'. \end{aligned} \tag{15}$$

This may be integrated into the theoretical framework by multiplying the matrix subblocks on the left- and right-hand sides of Eq. (10) by the Toeplitz matrix,

$$\mathbf{G} = \begin{bmatrix} G(\mathbf{0}) & G(\mathbf{R}_1) & \cdots & G(\mathbf{R}_n) \\ G(-\mathbf{R}_1) & G(\mathbf{0}) & \cdots & G(\mathbf{R}_n - \mathbf{R}_1) \\ \vdots & \vdots & \ddots & \vdots \\ G(-\mathbf{R}_n) & G(\mathbf{R}_1 - \mathbf{R}_n) & \cdots & G(\mathbf{0}) \end{bmatrix}, \tag{16}$$

where the set $\{\mathbf{0}, \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_n\}$ represents the probe positions in the image. We assume here periodic boundary conditions on the real-space grid on which $G(\mathbf{R})$ is evaluated, that is, $G(-\mathbf{R}_i) = G(\mathbf{R}_n - \mathbf{R}_i)$. Rather than inverting the matrix **A**, which describes the case of perfect spatial coherence, we instead invert the matrix product **GA**. Then we are able to correct for spatial incoherence to the extent allowed by the information transfer. This is implemented by replacing $\hat{\phi}_j(\mathbf{g})$ in Eqs. (12) and (13) with $\hat{G}(\mathbf{g})\hat{\phi}_j(\mathbf{g})$.

III. RETRIEVAL OF THE TRANSMISSION FUNCTION OF A MoS_2 MONOLAYER

A 4D dataset was taken on a thin MoS_2 sample (one monolayer) down the [001] axis using a 300 keV STEM probe with a 17.1 mrad convergence angle at the National Centre for Electron Microscopy (NCEM) in Berkeley, CA. Diffraction patterns were recorded for each probe position using a Gatan

K2-IS direct detection camera. A finite source size of a half-width at half-maximum (HWHM) of approximately 0.45 Å was deduced by comparing a high-angle annular dark field (HAADF) image that was recorded simultaneous with the 4D dataset with forward simulations using the freely available μSTEM software package [25]. Similarly, a probe defocus of −20 Å (underfocus) was deduced by comparing images synthesized using different detector configurations from the 4D dataset with equivalent images simulated with the μSTEM package for a range of different defocus values. Since the MoS_2 sample is a true monolayer, the multiplicative approximation holds—since all scattering occurs within one spatial plane—and the weak phase approximation also holds—since electrons focused to a point on the specimen will scatter off at most one molybdenum atom or two sulfur atoms. Therefore, we expect the linear approximation of Eq. (7) to hold.

The bright-field disk was segmented as shown in Fig. 3(a) and images synthesized using each of these detectors. To correct for scan drift and distortion in the synthesized images, the simultaneous ADF image was used to identify atomic peaks at the vertices of unit cells. The deviation of these peaks from their known positions was used to determine the transformation that was then applied to the synthesized images. Since this specimen was known to be periodic, we have averaged over the unit cells in the scan to obtain the results shown in Fig. 3(b). The reconstruction algorithm works equally well on the raw data. However, increasing the signal-to-noise ratio and decreasing scan distortion via the averaging procedure facilitates comparison with the simulated transmission function. Faster acquisition speed, either via fixed-configuration segmented detectors or the next generation of 4D pixel detector technology, would reduce distortions and render this correction procedure unnecessary. The phase of the transmission function reconstructed from the segmented detector images is shown in Fig. 3(c) with a schematic of the MoS_2 unit cell overlaid on the 2D map. Importantly, both molybdenum and sulfur atoms are clearly visible in the retrieved phase. While molybdenum atoms are clearly visible in each of the segmented detector images of Fig. 3(b), the sulfur atoms are barely evident. This is consistent with the prediction that phase imaging is sensitive to light atoms.

A plot of the fidelity, using the calculated transmission function displayed in Fig. 3(c) as the comparison transmission function, is shown in Fig. 3(d) as a function of CGLS iteration. The best match with the calculated transmission function (best fidelity) is achieved after six CGLS iterations. For subsequent iterations the quality of the retrieval declines, and this is consistent with the semiconvergence of the CGLS algorithm observed for cases in which the **b** vectors contain errors due to noisy experimental measurements. The corresponding L-curve is also displayed in Fig. 3(d). The sixth CGLS iteration is found just before the subsequent iterations cause rapid growth of $||\mathbf{x}||^2$. Figure 3(d) shows how selecting a number of CGLS iterations just prior to this point of inflection on the L-curve can be used as a criterion to select an appropriate number of CGLS iterations. Such an approach is necessary because in practice a reliable comparison transmission function T' will not be available from which to evaluate the fidelity in Eq. (14).

A line scan taken along the (red) arrow, over columns of molybdenum and sulfur atoms, is shown on the right-hand side of Fig. 3(c). This plot further shows the phase predicted from a MoS_2 monolayer with the lower plane of sulfur atoms removed, assuming the independent atom approximation. The excellent agreement shown in Fig. 3(c) for the phase of the

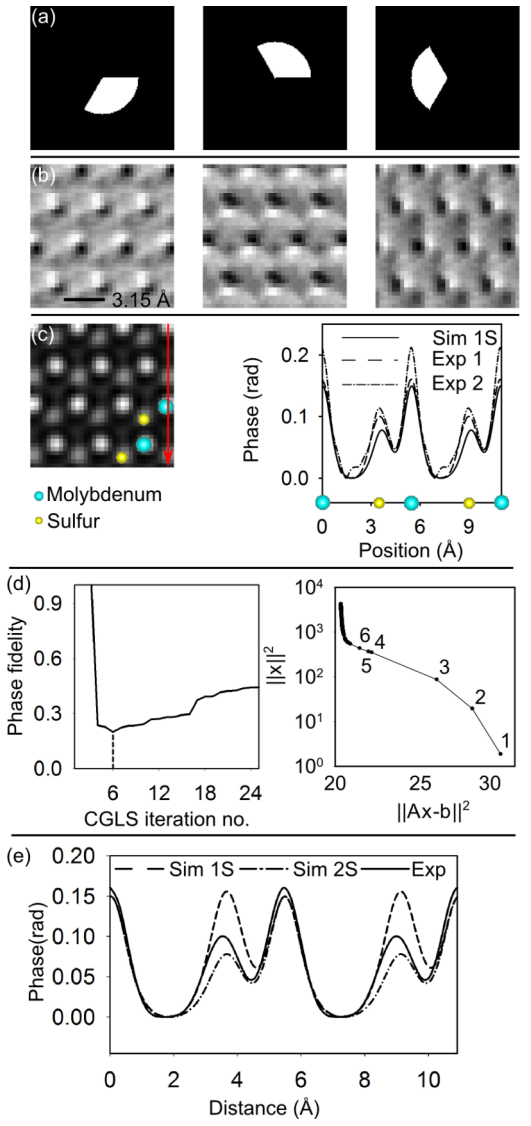

FIG. 3. Results for the MoS_2 sample. (a) Detector segments in the bright field used to synthesize the images in (b). (c) Retrieved phase of the transmission function. The method to deconvolve the blur associated with spatial incoherence (estimated to be 0.45 Å) described in Sec. II was employed to achieve the phase map shown in (c). The adjacent plot compares a line scan taken from the phase map (Exp 1), as indicted by the (red) arrow, with a line scan from a result (not shown) where spatial incoherence is not taken into account (Exp 2) and a transmission function simulated using μSTEM with one sulfur atom per unit cell (Sim). Both Exp 1 and Sim were convolved with the Gaussian associated with the spatial incoherence for comparison with Exp 2. (d) Plot of the fidelity, where the reference transmission function is Sim, and the L-curve as a function of CGLS iteration. (e) Plot of the phase of the transmission function calculated in μSTEM assuming both 1 (Sim 1S) and 2 (Sim 2S) sulfur atoms per primitive cell to that retrieved in (c) convolved with a Gaussian associated with the spatial incoherence for the sake of comparison.

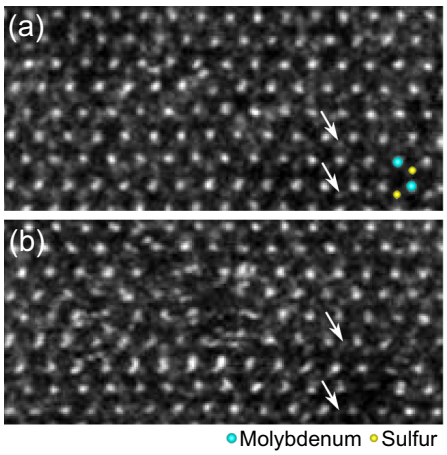

FIG. 4. (a) Before and (b) during ADF images for the MoS_2 sample. Arrows in images (a) and (b) indicate sites of sulfur atoms. For the uppermost arrow, the sulfur intensity is not appreciably reduced between images (a) and (b), while for the bottommost arrow, the change in intensity indicates sputtering of one or both sulfur atoms.

transmission function that is retrieved from the experimental data and the simulated transmission function may only be had if a plane of sulfur atoms is assumed to be missing. This is demonstrated clearly in Fig. 3(e), which compares the experimental result with the phase of simulated transmission functions assuming two sulfur atoms per primitive cell (2S) and one sulfur atom per primitive cell (1S). This apparent deviation from the expected stoichiometry of the manufactured MoS_2 sample is attributed to damage caused by the electron irradiation. This interpretation is reinforced in two ways.

First, an estimate can be made of the probability of the electron beam knocking a sulfur atom out of the monolayer specimen. Komsa *et al.* [26] predict the cross section for sputter of a sulfur atom from MoS_2 for 300 keV incident electrons to be about 150 barns. With four sulfur atoms per 1.7×10^{-19} m^2 projected unit cell, the probability of an electron incident upon one unit cell sputtering a sulfur atom is 3.4×10^{-7}. The 48 pA beam current and 2.5 ms dwell time mean that 7.5×10^5 electrons are incident upon the sample per probe position. As a crude estimate, this would lead us to expect that on average 0.26 sulfur atoms were sputtered per probe position. The probe is not uniformly distributed across the unit cell—for most probe positions the irradiation of the sulfur locations will be small—but this crude estimate nevertheless makes it likely that significant sputtering of sulfur atoms is taking place due to electron irradiation.

Second, compare the rapid ADF scan shown in Fig. 4(a) obtained prior to recording the 4D dataset with the ADF scan shown in Fig. 4(b) obtained during the 4D data acquisition at a much slower scan rate (specifically, a dwell time of 2.5 ms per probe position). While the latter shows some drift distortion due to the slower scan, it also shows clear evidence of localized beam damage. Closer inspection reveals that the intensity at the sulfur sites is, by and large, smaller in Fig. 4(b) than Fig. 4(a). Note, though, that there is much site-to-site variation, suggesting that in some cases both sulfur atoms might be sputtered, in others one might be sputtered, while in others both may remain intact. It should be reiterated that the reconstruction in Fig. 3(c) was based on repeat unit (Fourier) averaged data: the fact that the reconstructed phases is consistent with one rather than two sulfur atoms at the sulfur site should be understood to be an average, and the fact that this average is close to unity is coincidental rather than exact.

IV. RETRIEVAL OF THE TRANSMISSION FUNCTION OF A $SrTiO_3$ CRYSTAL

A 4D dataset was taken on a $SrTiO_3$ sample down the [001] axis using a 300 keV STEM probe with a 21.3 mrad convergence angle at NCEM, using the same apparatus as in Sec. III.

Position-averaged convergent beam electron diffraction (PACBED) patterns were synthesized for two scan areas of interest by adding the recorded diffraction patterns together. Comparison with simulated PACBED patterns enables the thickness of a specimen to be determined with an accuracy better than 10% [27]. PACBED patterns were simulated using μSTEM [25], and the L_2 norm was used to decide which thickness provided the best fit to the experimental data. The thickness of the specimen was found to be 35.1 Å for the first region and 78.1 Å for the second region. This suggests that this dataset will be a good test of the validity of the approximation, made in Sec. II, that the scattering from the specimen can be represented by a transmission function in a single plane. (In the context of differential phase-contrast imaging, the breakdown of this approximation with thickness has recently been systematically explored through simulation by Close *et al.* [12], though the manner in which it breaks down need not be the same when analyzed via SDP.)

A finite source size of a HWHM of 0.5 Å was deduced from an independently taken HAADF image. As with the dataset in Sec. III, a defocus of −30 Å for the 35.1-Å-thick sample and a defocus of −20 Å for the 78.1-Å-thick sample was deduced by comparing the data with the results of μSTEM simulations for different defocus values and detector geometries.

The results for the 35.1-Å-thick sample are shown in Figs. 5(b)–5(d), and the results for the 78.1-Å-thick sample are shown in Figs. 5(e)–5(g). The bright-field disk was segmented as shown in Fig. 5(a). Images synthesized using each detector are shown in (b) and (e). The phase of the transmission function obtained from these images is shown in (c) and (f) with a schematic of the $SrTiO_3$ unit cell overlaid. A line scan taken along the (red) arrow, over columns of strontium and oxygen atoms, is shown on the right-hand side. This line scan also shows the phase for the projected transmission function of the relevant structure for comparison, as calculated using the μSTEM [25] software package. Finally, for comparison, Figs. 5(d) and 5(g) show for both thicknesses of $SrTiO_3$ the bright-field STEM images that were synthesized from the same experimental data (a schematic of the $SrTiO_3$ unit cell is also overlaid on these images). In both cases, oxygen atoms are not visible in the bright-field image but are clearly visible in the retrieved phases shown in Figs. 5(c) and 5(f).

Unlike the line scan for Fig. 3(c), i.e., the case of the MoS_2 monolayer, line scans in Fig. 5 comparing the retrieved phase of SDP with a calculated transmission function do not show reasonable quantitative agreement. This is partly due to the fact that both the weak phase approximation [see Eq. (5)] and the

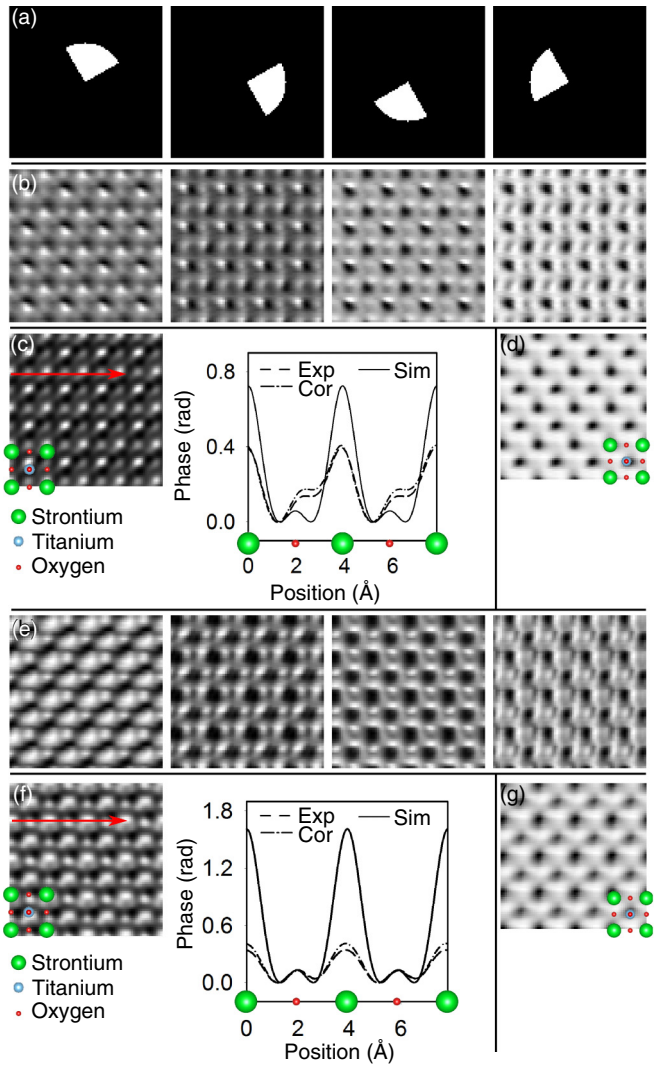

FIG. 5. Results for 35.1- and 78.1-Å-thick $SrTiO_3$ samples. (a) Detector segments in the bright field used to synthesize the images in (b) for the 35.1-Å-thick $SrTiO_3$ sample. (c) Retrieved phase with a line scan as indicated by the (red) arrow. Sim is the simulated projected transmission function, Exp is the retrieved transmission function, and Cor is the retrieved transmission function corrected for neglect of the nonlinear term. The method to deconvolve the blur associated with spatial incoherence (estimated to be 0.5 Å) described in Sec. II was employed to achieve the phase map shown in (c). The results in (c) are convolved with this blur for the purposes of comparing to the simulated transmission function in the line scan. (d) A bright-field image synthesized from the same dataset using a detector with the same radius as the bright field disk. The results displayed in (e)–(g) are the same as those displayed in (b)–(d) except that they are for the 78.1-Å-thick sample.

multiplicative approximation [see Eq. (2)] are no longer valid for either case. Note that agreement between the simulated and experimentally retrieved transmission functions is generally better in Fig. 5(c), which is a thinner sample and so represents a less serious breach of both approximations, than it is in Fig. 5(f). Regardless, the SDP phase map is a useful tool that very clearly identifies the presence of oxygen—an element that is not easily inferred from any of the raw segmented detector images or the bright-field images in Figs. 5(d) and 5(g).

V. ACCOUNTING FOR PROPAGATION THROUGH THE SPECIMEN—THE *M*-PLANE INVERSION

The multiplicative approximation that was employed to write Eq. (2) is correct in the limit of an infinitesimally thin imaging specimen. We note that factorization into a probe and object function for realistic thicknesses has been discussed in Ref. [28]. A focused coherent electron probe will evolve rapidly for propagation through thick specimens, and it is therefore necessary to investigate methods to better account for thickness in the retrieval.

Scattering by a thick object may be modeled using the multislice method [29,30] by projecting the scattering potential into a finite number of slices, $T_i(\mathbf{r})$, and propagating the resulting elastically scattered wave between these slices. A weak phase (linear) approximation at each slice may also be assumed [31–34]. Scattering through a single slice is calculated by integration with the function,

$$\begin{aligned} S_i(\mathbf{r},\mathbf{r}') &= P(\mathbf{r}'-\mathbf{r})T_i(\mathbf{r}') \\ &= P(\mathbf{r}'-\mathbf{r})[1+O_i(\mathbf{r}')], \end{aligned} \tag{17}$$

which is the multiplication by the transmission function followed by a convolution with the propagator $P(\mathbf{r})$. Scattering through M layers of the sample would entail integration of $\psi_{\text{illum}}(\mathbf{r})$ with $S_i(\mathbf{r},\mathbf{r}')$ M times,

$$\begin{aligned} \psi_{\text{exit}}(\mathbf{r}) = \int\cdots\int & S_M(\mathbf{r},\mathbf{r}_{M-1})\cdots S_2(\mathbf{r}_2,\mathbf{r}_1)S_1(\mathbf{r}_1,\mathbf{r}') \\ & \times\psi_{\text{illum}}(\mathbf{r}')d\mathbf{r}_{M-1}\cdots d\mathbf{r}_1 d\mathbf{r}'. \end{aligned} \tag{18}$$

Within the paraxial approximation, the propagator may be written in reciprocal space as

$$\hat{P}(\mathbf{g}) = \exp\left[-\pi i g^2 \lambda \Delta z\right], \tag{19}$$

where λ is the wavelength of the electron wave function and Δz is the slice thickness. With a linear approximation for the object functions, Eq. (18) may be expanded to give

$$\begin{aligned} \psi_{\text{exit}}(\mathbf{r},\mathbf{R}) =& \psi_{\text{illum}}(\mathbf{r},\mathbf{R}) * P^M(\mathbf{r}) + \sum_{i=0}^{M-1}\{[\psi_{\text{illum}}(\mathbf{r},\mathbf{R}) \\ & * P^{(i)}(\mathbf{r})]O_i(\mathbf{r})\} * P^{(M-i)}(\mathbf{r}). \end{aligned} \tag{20}$$

In Fourier space, this is

$$\begin{aligned} \hat{\psi}_{\text{exit}}(\mathbf{q},\mathbf{R}) =& \hat{\psi}_{\text{illum}}(\mathbf{q})e^{-2\pi i \mathbf{g}\cdot\mathbf{R}}\hat{P}^M(\mathbf{q}) \\ & + \sum_{i=0}^{M-1}\sum_{\mathbf{g}}\hat{\psi}_{\text{illum}}(\mathbf{q}-\mathbf{g})e^{-2\pi i(\mathbf{q}-\mathbf{g})\cdot\mathbf{R}} \\ & \times \hat{P}^{(i)}(\mathbf{q}-\mathbf{g})\hat{O}_i(\mathbf{g})\hat{P}^{(M-i)}(\mathbf{q}). \end{aligned} \tag{21}$$

For a sample that is periodic in the z direction, i.e., $O_i(\mathbf{r}) = O(\mathbf{r})$, and if $\hat{\phi}(\mathbf{g})$ is redefined,

$$\begin{aligned} \hat{\phi}_j(\mathbf{g}) = \sum_{i=0}^{M-1}\int & \hat{\psi}_{\text{illum}}(\mathbf{q}-\mathbf{g})\hat{P}^{(i)}(\mathbf{q}-\mathbf{g}) \\ & \times \hat{P}^{*(i)}(\mathbf{q})\hat{\psi}^*_{\text{illum}}(\mathbf{q})D_j(\mathbf{q})d\mathbf{q}, \end{aligned} \tag{22}$$

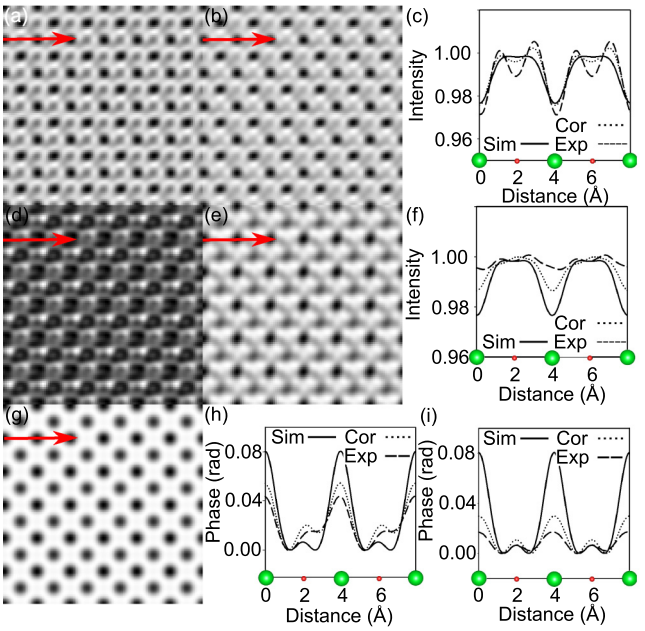

FIG. 6. Application of the M-plane inversion to the $SrTiO_3$ dataset. For the 35.1-Å-thick sample, the retrieved amplitude (a) without using the M-plane inversion and (b) using the M-plane inversion. A line scan comparing the retrieved intensity to the intensity of the simulated transmission function for a layer of $SrTiO_3$ with a thickness of one unit cell (3.905 Å) is shown in (c). The results without M-plane inversion are labeled "Exp" and the results with M-plane inversion are labeled "Cor." Parts (d)–(f) are the same as (a)–(c) except that they are for the 78.1 Å case. The intensity of the simulated transmission function for a layer of $SrTiO_3$ with a thickness of one unit cell is shown in (g) for comparison. The effect of the M-plane inversion on the retrieved phase is displayed in (h) for the 35.1-Å-thick sample of $SrTiO_3$ and in (i) for the 78.1-Å-thick sample of $SrTiO_3$.

then the equation solved to retrieve $O(\mathbf{g})$ will be exactly the same as Eq. (8) in Sec. II. This method is dubbed the M-plane inversion since the retrieval assumes scattering of the probe by M distinct planes.

To test this modification to the algorithm, the M-plane inversion is applied to the $SrTiO_3$ datasets explored in Sec. IV. The number M of planes used in the retrieval was set equal to the thickness in terms of $SrTiO_3$ unit cells for each specimen. It was found that applying the M-plane inversion generally improved agreement with the simulated transmission function for both the intensity and phase of the retrieved transmission function. The intensity results are displayed in Fig. 6 for the 35.1 Å sample in (a) without the M-plane inversion and (b) with the M-plane inversion. A line scan that compares both results with the simulated transmission function intensity for a layer of $SrTiO_3$ with a thickness of one unit cell is shown in Fig. 6(c). Figures 6(d)–6(f) show the same results as Figs. 6(a)–6(c) except for the 78.1-Å-thick $SrTiO_3$ sample. Figure 6(g) shows the transmission function intensity for a layer of $SrTiO_3$ with a thickness of one unit cell for comparison with the experimental retrievals. It can be seen that the M-plane inversion improves qualitative and quantitative agreement between the experimental retrieval and the simulated transmission function intensity for both thicknesses of $SrTiO_3$, but the improvement is most marked for the 78.1-Å-thick $SrTiO_3$ sample. The phase retrieved for the transmission function of a $SrTiO_3$ sample with a thickness of a single unit cell is shown in Fig. 6(h) for the 35.1 Å sample and Fig. 6(i) for the 78.1 Å sample using both the approach described in Sec. II ("Exp") and the M-plane inversion ("Cor"), and this is compared with the simulated transmission function ("Sim"). In both cases, there is a very small change in the phase associated with the oxygen column and a larger increase in the phase associated with the strontium column such that the overall agreement with the simulated transmission function is improved.

A different approach to that described in this paragraph would be to parametrize the channeling with a limited number of structural parameters, such as in the S-state model [28].

VI. CONCLUSION

We have demonstrated a method for phase retrieval based on diffraction patterns acquired as a function of probe position in STEM using fast readout detectors (so-called 4D datasets). The approach allows the complex transmission function of a thin specimen to be retrieved from images synthesized from the diffraction patterns taken as a function of probe position for assumed virtual segmented detectors in the diffraction plane. This approach provides maximum flexibility in the choice defining the images. However, the approach introduced here can also be applied to data obtained from "fixed-configuration" segmented detectors, where each segment is a bucket detector rather than being defined in terms of a number of pixels on a fast-readout 2D electron camera [10,14,20,21]. In cases in which the full 4D dataset is recorded, the relative simplicity of SDP—a small number of Fourier transform operations is applied to a reduced 2D projection of the dataset—can complement other ptychographic approaches such as ePIE [3] and Wigner distribution deconvolution [35].

We have demonstrated SDP using model data and three experimental 4D datasets. The first application was to a MoS_2 monolayer. The retrieved phase of the transmission function was in good agreement with that simulated from within an isolated atom approximation, assuming that damage under the beam had led to approximately 50% of the sulfur atoms being removed from the specimen. The second application was to two separate regions of a $SrTiO_3$ sample, with different thicknesses. As expected, better agreement was obtained between theory and experiment for the thinner region. Importantly, sulfur and oxygen atoms, which have relatively low atomic numbers and are therefore difficult to see in conventional STEM images, were easily visible in the retrieved phase maps.

For all of the simulated and experimental cases, the complex transmission function can be retrieved using the signal from a small number of segmented detectors. Only two fast Fourier transforms need to be performed per CGLS iteration, and this means that real-time phase retrieval is feasible.

ACKNOWLEDGMENTS

The authors wish to thank Peter Ercius and Jim Ciston for their assistance during the experiments, and Peter Nellist

and Hao Yang for helpful discussions. This research was supported under the Australian Research Council's Discovery Projects funding scheme (Projects No. DP110102228, No. DP140102538, and No. DP150103837), the Australian Laureate Fellowships funding scheme (FL120100038), and the Discovery Early Career Researcher Awards funding scheme (Projects No. DE130100739 and No. DE140101555). Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, Office of the US Department of Energy under Contract No. DE-AC02-05CH11231. The Monash Centre for Electron Microscopy is acknowledged for providing facilities for sample preparation, initial microscopy, and data handling.

APPENDIX: SIMULTANEOUS RECONSTRUCTION OF THE ILLUMINATION FUNCTION

We have assumed that the probing illumination is well known. If refinement of our knowledge of the probe is required, then we can proceed as follows. We consider the scattering of the illumination $\psi_{\text{illum}}(\mathbf{r}) = \psi(\mathbf{r}) + \psi^{\delta}(\mathbf{r})$. The quantity $\psi(\mathbf{r})$ is our initial guess of the probe wave function, and $\psi^{\delta}(\mathbf{r})$ is a correction. A first estimate of $T(\mathbf{r})$ is retrieved using $\psi(\mathbf{r})$ and is used to calculate $\psi^{\delta}(\mathbf{r})$. For scattering of $\psi_{\text{illum}}(\mathbf{r})$ by $T(\mathbf{r})$, the exit surface wave is

$$\psi_{\text{exit}}(\mathbf{r},\mathbf{R}) = T(\mathbf{r})[\psi(\mathbf{r}-\mathbf{R}) + \psi^{\delta}(\mathbf{r}-\mathbf{R})]. \tag{A1}$$

In Fourier space, we have

$$\hat{\psi}_{\text{exit}}(\mathbf{q},\mathbf{R}) = [\hat{\psi}(\mathbf{q})e^{-2\pi i\mathbf{q}\cdot\mathbf{R}}] * \hat{T}(\mathbf{q}) + [\hat{\psi}^{\delta}(\mathbf{q})e^{-2\pi i\mathbf{q}\cdot\mathbf{R}}] * \hat{T}(\mathbf{q}). \tag{A2}$$

For brevity, we introduce the reference wave

$$\hat{\psi}_{\text{ref}}(\mathbf{q},\mathbf{R}) = [\hat{\psi}(\mathbf{q})e^{-2\pi i\mathbf{q}\cdot\mathbf{R}}] * \hat{T}(\mathbf{q}). \tag{A3}$$

The convolution with $\hat{\psi}^{\delta}(\mathbf{q})$ in Eq. (A1) is written explicitly as

$$\hat{\psi}_{\text{exit}}(\mathbf{q},\mathbf{R}) = \hat{\psi}_{\text{ref}}(\mathbf{q},\mathbf{R}) + \sum_{\mathbf{g}} \hat{\psi}^{\delta}(\mathbf{g})e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{T}(\mathbf{q}-\mathbf{g}). \tag{A4}$$

We now define

$$\begin{aligned}\hat{\chi}_j(\mathbf{g},\mathbf{R}) &= \int \hat{\psi}^{*}_{\text{ref}}(\mathbf{q},\mathbf{R})\hat{T}(\mathbf{q}-\mathbf{g})D_j(\mathbf{q})d\mathbf{q} \\ &= [\hat{\psi}^{*}_{\text{ref}}(\mathbf{g},\mathbf{R})D_j(\mathbf{g})] * \hat{T}(-\mathbf{g}).\end{aligned} \tag{A5}$$

We now follow the work in Sec. II. If Eq. (A4) is substituted into Eq. (4) and the final nonlinear term is ignored, the quantity $b_j(\mathbf{R})$ may be written as

$$\begin{aligned}b_j(\mathbf{R}) &= I_j(\mathbf{R}) - \int |\hat{\psi}_{\text{ref}}(\mathbf{q},\mathbf{R})|^2 D_j(\mathbf{q})d\mathbf{q} \\ &= 2\sum_{\mathbf{g}} \text{Re}[\hat{\psi}^{\delta}(\mathbf{g})e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\chi}_j(\mathbf{g},\mathbf{R})].\end{aligned} \tag{A6}$$

If it is assumed that an estimate of the transmission function, $\hat{T}(\mathbf{q})$, is known, then we can solve for $\hat{\psi}^{\delta}(\mathbf{g})$. To illustrate how this proceeds, Eq. (A6) is expressed as a matrix equation,

$$\begin{bmatrix} [2\,\text{Re}[e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\chi}_1(\mathbf{g},\mathbf{R})]] & [-2\,\text{Im}[e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\chi}_1(\mathbf{g},\mathbf{R})]] \\ [2\,\text{Re}[e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\chi}_2(\mathbf{g},\mathbf{R})]] & [-2\,\text{Im}[e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\chi}_2(\mathbf{g},\mathbf{R})]] \\ \vdots & \vdots \\ [2\,\text{Re}[e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\chi}_n(\mathbf{g},\mathbf{R})]] & [-2\,\text{Im}[e^{-2\pi i\mathbf{g}\cdot\mathbf{R}}\hat{\chi}_3(\mathbf{g},\mathbf{R})]] \end{bmatrix} \times \begin{bmatrix} \text{Re}[\hat{\psi}^{\delta}(\mathbf{g})] \\ \text{Im}[\hat{\psi}^{\delta}(\mathbf{g})] \end{bmatrix} = \begin{bmatrix} [b_1(\mathbf{R})] \\ [b_2(\mathbf{R})] \\ \vdots \\ [b_n(\mathbf{R})] \end{bmatrix}. \tag{A7}$$

Once again, this is of the form $\mathbf{Ax} = \mathbf{b}$. However, we are prevented from an efficient implementation of the CGLS method using Fourier transforms, similar to that in Sec. II, by the dependence of $\hat{\chi}(\mathbf{g},\mathbf{R})$ on both $\mathbf{g}$ and $\mathbf{R}$.

[1] A. J. Morgan, A. J. D'Alfonso, P. Wang, H. Sawada, A. I. Kirkland, and L. J. Allen, Phys. Rev. B **87**, 094115 (2013).
[2] C. T. Putkunz, A. J. D'Alfonso, A. J. Morgan, M. Weyland, C. Dwyer, L. Bourgeois, J. Etheridge, A. Roberts, R. E. Scholten, K. A. Nugent, and L. J. Allen, Phys. Rev. Lett. **108**, 073901 (2012).
[3] M. J. Humphry, B. Kraus, A. C. Hurst, A. M. Maiden, and J. M. Rodenburg, Nat. Commun. **3**, 730 (2012).
[4] A. J. D'Alfonso, A. J. Morgan, A. W. C. Yan, P. Wang, H. Sawada, A. I. Kirkland, and L. J. Allen, Phys. Rev. B **89**, 064101 (2014).
[5] C. Ophus, P. Ercius, M. Sarahan, C. Czarnik, and J. Ciston, Microsc. Microanal. **20**, 62 (2014).
[6] T. J. Pennycook, A. R. Lupini, H. Yang, M. F. Murfitt, L. Jones, and P. D. Nellist, Ultramicroscopy **151**, 160 (2015).
[7] H. Yang, L. Jones, H. Ryll, M. Simson, H. Soltau, Y. Kondo, R. Sagawa, H. Banba, I. MacLaren, and P. D. Nellist, J. Phys.: Conf. Ser. **644**, 012032 (2015).
[8] H. Yang, T. J. Pennycook, and P. D. Nellist, Ultramicroscopy **151**, 232 (2015).
[9] C. Gammer, V. B. Ozdol, C. H. Liebscher, and A. M. Minor, Ultramicroscopy **155**, 1 (2015).
[10] N. Shibata, Y. Kohno, S. D. Findlay, H. Sawada, Y. Kondo, and Y. Ikuhara, J. Electron Microsc. **59**, 473 (2010).
[11] N. Shibata, S. D. Findlay, Y. Kohno, H. Sawada, Y. Kondo, and Y. Ikuhara, Nat. Phys. **8**, 611 (2012).
[12] R. Close, Z. Chen, N. Shibata, and S. D. Findlay, Ultramicroscopy **159**, 124 (2015).
[13] N. Shibata, S. D. Findlay, H. Sasaki, T. Matsumoto, H. Sawada, Y. Kohno, S. Otomo, R. Minato, and Y. Ikuhara, Sci. Rep. **5**, 10040 (2015).
[14] I. Lazić, E. G. T. Bosch, and S. Lazar, Ultramicroscopy **160**, 265 (2016).
[15] H. Rose, Ultramicroscopy **2**, 251 (1977).
[16] R. E. Burge and P. Van Toorn, Inst. Phys. Conf. Ser. **52**, 249 (1980).
[17] R. E. Burge and P. Van Toorn, Scanning Electron Microsc. **I**, 81 (1980).
[18] M. N. Landauer, B. C. McCallum, and J. M. Rodenburg, Optik **100**, 37 (1995).

[19] B. C. McCallum, M. N. Landauer, and J. M. Rodenburg, Optik **101**, 53 (1995).
[20] J. N. Chapman, R. Ploessl, and D. M. Donnet, Ultramicroscopy **47**, 331 (1992).
[21] M. Haider, A. Epstein, P. Jarron, and C. Boulin, Ultramicroscopy **54**, 41 (1994).
[22] A. W. Yan, A. J. D'Alfonso, A. J. Morgan, C. T. Putkunz, and L. J. Allen, Microsc. Microanal. **20**, 1090 (2014).
[23] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[24] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms* (Society for Industrial and Applied Mathematics, Philadelphia, 2010), Vol. 7.
[25] L. J. Allen, A. J. D'Alfonso, and S. D. Findlay, Ultramicroscopy **151**, 11 (2015), software available at http://tcmp.ph.unimelb.edu.au/mustem/muSTEM.html.
[26] H.-P. Komsa, J. Kotakoski, S. Kurasch, O. Lehtinen, U. Kaiser, and A. V. Krasheninnikov, Phys. Rev. Lett. **109**, 035503 (2012).
[27] J. M. LeBeau, S. D. Findlay, L. J. Allen, and S. Stemmer, Ultramicroscopy **110**, 118 (2010).
[28] J. Broeckx, M. O. de Beeck, and D. Van Dyck, Ultramicroscopy **60**, 71 (1995).
[29] J. M. Cowley and A. F. Moodie, Acta Crystallogr. **10**, 609 (1957).
[30] W. Coene and D. Van Dyck, Ultramicroscopy **33**, 261 (1990).
[31] D. Misell, J. Phys. D **9**, 1849 (1976).
[32] B. K. Jap and R. M. Glaeser, Acta Crystallogr. Sect. A **36**, 57 (1980).
[33] F. Li and D. Tang, Acta Crystallogr. Sect. A **41**, 376 (1985).
[34] M. Vulović, L. M. Voortman, L. J. van Vliet, and B. Rieger, Ultramicroscopy **136**, 61 (2014).
[35] P. Li, T. B. Edo, and J. M. Rodenburg, Ultramicroscopy **147**, 106 (2014).