Title

Scattering Matrix Determination in Crystalline Materials from 4D Scanning Transmission Electron Microscopy at a Single Defocus Value

Journal

Microscopy and Microanalysis, 27(4)

ISSN

1431-9276

Authors

Findlay, Scott D

Brown, Hamish G

Pelz, Philipp M

et al.

Publication Date

2021-08-01

DOI

10.1017/s1431927621000490

Peer reviewed

Research Paper

Scattering matrix determination in crystalline materials from 4D scanning transmission electron microscopy at a single defocus value

Scott D. Findlay[1], Hamish G. Brown[2,3], Philipp M. Pelz[2,4], Colin Ophus[2], Jim Ciston[2] and Leslie J. Allen[5]

[1]School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia

[2]National Center for Electron Microscopy Facility, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[3]Now at: Ian Holmes Imaging Center, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3052, Australia

[4]Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA

[5]School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia

Abstract

Recent work has revived interest in the scattering matrix formulation of electron scattering in transmission electron microscopy as a stepping stone towards atomic-resolution structure determination in the presence of multiple scattering. We discuss ways of visualising the scattering matrix that make its properties clear. Through a simulation-based case study incorporating shot noise, we shown how regularising on this continuity enables the scattering matrix to be reconstructed from 4D scanning transmission electron microscopy measurements from a single defocus value. Intriguingly, for crystalline samples this process also yields the sample thickness to nanometer accuracy with no *a priori* knowledge about the sample structure. The reconstruction quality is gauged by using the reconstructed scattering matrix to simulate STEM images at defocus values different to that of the data from which it was reconstructed.

Key words: scattering matrix, 4D STEM, phase retrieval

()

Introduction

The recent development of fast-readout electron pixel detectors suitable for scanning transmission electron microscopy (STEM) has spurred the development and application of a range of phase retrieval / structure determination methods, including ptychography (Yang et al., 2016; Gao et al., 2017; Jiang et al., 2018; Schloz et al., 2020; Chen et al., 2020) and differential phase contrast (Müller et al., 2014; Lazić et al., 2016; Chen et al., 2016). At atomic resolution, the phase object approximation on which many such techniques are based starts to quantitatively break down for samples with thickness greater than a few nanometers (Close et al., 2015; Müller-Caspary et al., 2017; Winkler et al., 2020). Attention has thus turned to developing approaches to handle multiple electron scattering in thicker samples, with methods including optical sectioning (Yang et al., 2016; Bosch & Lazić, 2019; Brown et al., 2020), 3D ptychography / inverse multislice (Maiden et al., 2012; Van den Broek & Koch, 2012, 2013; Gao et al., 2017; Schloz et al., 2020; Chen et al., 2020), and scattering matrix inversion (Brown et al., 2018) all showing promise.

This paper concerns the scattering matrix approach, and will focus solely on reconstructing the scattering matrix from 4D STEM measurements. Pixel detectors in 4D STEM imaging are sometimes referred to as universal detectors (Tate et al., 2016; Hachtel et al., 2018) since post-experiment one can synthesize any desired detector geometry (constrained by the extent of the detector). Reconstructing the (complex) scattering matrix elements allows a further degree of generalisation: having reconstructed the scattering matrix, it can be used to simulate completely different imaging modes, constrained by the range of scattering matrix elements reconstructed and provided they result from elastic scattering. For instance, in STEM one could synthesize images for a different probe defocus (or indeed any other desired combination of aberrations), or for an annular probe-forming aperture. A more novel example is the parallax method of Ophus and coworkers which probes the 3D structure via a kind of optical sectioning (Ophus et al., 2019; Brown et al., 2020).

Most recent discussions of reconstructing the scattering matrix regard doing so as an intermediate step towards determining the structure (specifically the specimen electrostatic potential) (Brown et al., 2018;

Author for correspondence: Scott D. Findlay, Email: scott.findlay@monash.edu

Donatelli & Spence, 2020; Pelz et al., 2020). When that can be done[a] there would be little reason to synthesize other imaging modes. However, in cases where at present the structure can only be determined to limited accuracy, reconstructing the scattering matrix could, via synthesizing other imaging modes, offer insights not readily visible in the measured data.

The previous, experimentally-realised scattering matrix reconstructions (Brown et al., 2018; Pelz et al., 2020; Brown et al., 2020) involved 4D STEM datasets (2D diffraction patterns at each probe position in a 2D raster scan) from a few defocus values. Having more defocus values improves the reliability of phase retrieval, making it more robust to noise in each 4D STEM dataset, but also increases the complexity of the data acquisition, increases the potential for beam damage, and necessitates careful alignment between datasets. Through simulation, we demonstrate that 4D STEM data from a single defocus value is sufficient to reconstruct a reasonable estimate of the scattering matrix. Additionally, if the sample is crystalline — specifically, if it has small periodicity along the beam direction — we show that its thickness can be estimated to high accuracy with no additional assumptions (i.e. beyond periodicity) about the structure.

This paper is structured as follows. The scattering matrix formulation is first introduced, including an overview of methods for calculating the scattering matrix and an explanation of the so-called antidiagonal symmetry property it possesses when the sample is periodic. Ways of visualising the scattering matrix are then considered, establishing where the continuity resides. This continuity is then used to regularise the reconstruction of the scattering matrix from measurements at a single defocus value, where the antidiagonal symmetry property is also shown to determine the sample thickness. Discussion follows.

The scattering matrix formulation in electron microscopy

Representations of the scattering matrix

The scattering matrix operator $\mathcal{S}$ is such that when applied to the entrance-surface wavefield ψ_{in} the result is the (elastically scattered) exit-surface wavefield ψ_{out} (Sturkey, 1962):

$$\psi_{\mathrm{out}} = \mathcal{S}\psi_{\mathrm{in}} \ . \tag{1}$$

Having chosen the basis/bases in which to work, this can be considered as a matrix-vector product. In high-resolution S/TEM, the most common choices of basis are:

[a]Note that there are methods to determine the structure in the presence of multiple scattering that do not involve determining the scattering matrix as an intermediate step (Van den Broek & Koch, 2012, 2013; Ren et al., 2020; Schloz et al., 2020; Chen et al., 2020).

- real space – discrete 2D array of spatial coordinates, for instance corresponding to detector pixel locations for a detector in the image plane; and
- reciprocal space – discrete 2D array of spatial frequencies / Fourier coefficients, for instance corresponding to detector pixel locations for a detector in the diffraction plane.

There being no intrinsic requirement for the entrance-surface and exit-surface wavefields to be represented in the same basis, the following matrix representations of the operator are both valid:

$$\psi_{\mathrm{out}}(\mathbf{r}) = \sum_{\mathbf{g}} \mathcal{S}_{\mathbf{r},\mathbf{g}} \psi_{\mathrm{in}}(\mathbf{g}) \tag{2a}$$

$$\psi_{\mathrm{out}}(\mathbf{h}) = \sum_{\mathbf{g}} \mathcal{S}_{\mathbf{h},\mathbf{g}} \psi_{\mathrm{in}}(\mathbf{g}) \ , \tag{2b}$$

where $\mathbf{r}$ is a 2D real space vector and $\mathbf{h}$ and $\mathbf{g}$ are 2D reciprocal space vectors, all orthogonal to the optic axis. We will rely solely on which vectors are invoked to indicate in which space(s) ψ_{in}, ψ_{out} and $\mathcal{S}$ reside.[b] Transformation between these bases is accomplished by (discrete) Fourier transform:

$$\mathcal{S}_{\mathbf{r},\mathbf{g}} = \sum_{\mathbf{h}} \mathcal{S}_{\mathbf{h},\mathbf{g}} e^{2\pi i \mathbf{h}\cdot\mathbf{r}} \ . \tag{3}$$

If the incident wave is a plane wave with transverse wavevector $\mathbf{g}_1$ (i.e. $\psi_{\mathrm{in}}(\mathbf{g}) = \delta_{\mathbf{g},\mathbf{g}_1}$) then it follows from Eq. (2a) that $\mathcal{S}_{\mathbf{r},\mathbf{g}_1}$ is the complex amplitude of the exit surface wavefield, and from Eq. (2b) that $\mathcal{S}_{\mathbf{h},\mathbf{g}_1}$ is the complex amplitude of the diffraction beam $\mathbf{h}$. The corresponding intensities are simply the modulus square of these complex amplitudes. Because experiments only measure intensities, we cannot *directly* measure the complex elements of the $\mathcal{S}$ matrix – electron microscopy's perennial phase problem. However, there are several proposals (Spence, 1998; Allen et al., 2000; Findlay, 2005; Brown et al., 2018; Donatelli & Spence, 2020; Pelz et al., 2020; Brown et al., 2020), and some proof-of-principle experiments (Brown et al., 2018, 2020), on how to reconstruct the (complex) $\mathcal{S}$ matrix elements from measured data.

Simulating the scattering matrix via multislice

Since $\mathcal{S}_{\mathbf{r},\mathbf{g}}$ can be regarded as the exit surface wavefield $\psi_{\mathrm{out}}(\mathbf{r})$ obtained when the incident wavefield is $\psi_{\mathrm{in}}(\mathbf{r}) = e^{2\pi i \mathbf{g}\cdot\mathbf{r}}$, it can be calculated via the multislice method for as many $\mathbf{g}$ values as desired. Ophus and coworkers have shown that this approach is particularly suitable for large, non-periodic specimens (Ophus,

[b] Eq. (2) could be augmented by two further equations involving $\psi_{\mathrm{in}}(\mathbf{r})$, but we seldom find cause to use them in the $\mathcal{S}$ matrix formulation of STEM.

2017), and that there is some structure / predictability in the rate at which $\mathcal{S}_{\mathbf{r},\mathbf{g}}$ varies with $\mathbf{g}$ that can be used to limit the number of distinct $\mathbf{g}$ values for which the calculation need be performed (Pelz et al., 2020).

Simulating the scattering matrix via Bloch wave in crystalline materials

In the paraxial approximation and the projected potential (zero-order Laue zone) approximation, the $\mathcal{S}$ matrix is related to the so-called structure matrix $\mathcal{A}$ via

$$\mathcal{S}(t) = \exp\left(\frac{i\pi t}{K}\mathcal{A}\right) \ , \tag{4}$$

where t is the sample thickness, K is the (relativistically corrected) electron wavevector, and, adopting the reciprocal space basis, the elements of $\mathcal{A}$ may be written

$$\mathcal{A}_{\mathbf{h},\mathbf{g}} = \begin{cases} -(\mathbf{k}_t + \mathbf{h})^2 + iU'_{\mathbf{0}} & \mathbf{h} = \mathbf{g} \\ U_{\mathbf{h}-\mathbf{g}} + iU'_{\mathbf{h}-\mathbf{g}} & \mathbf{h} \neq \mathbf{g} \end{cases} \ , \tag{5}$$

in which $U_{\mathbf{g}}$ and $U'_{\mathbf{g}}$ are the Fourier coefficients of the elastic and absorptive scattering potential, and $\mathbf{k}_t$ is the transverse component (relative to the sample zone axis) of the incident wavevector.[c] In atomic-resolution STEM, $\mathbf{k}_t$ is generally close to zero, and in what follows we chose our coordinate system such that it is identically zero.

There are many ways of evaluating the matrix exponential in Eq. (4) (Moler & Van Loan, 2003), but the most common in the electron microscopy literature is via spectral decomposition of the structure matrix:

$$\mathcal{A} = \mathcal{C}\left[\lambda^i\right]_D \mathcal{C}^{-1} \ , \tag{6}$$

where $\left[\lambda^i\right]_D$ is a diagonal matrix whose diagonal elements are the eigenvalues λ^i of $\mathcal{A}$, and $\mathcal{C}$ is a matrix whose columns are the eigenvectors of $\mathcal{A}$. Since the latter eigenvectors represent the Bloch states, we regard this as a variant of the Bloch wave method. It follows from Eq. (4) that the spectral decomposition of the $\mathcal{S}$ matrix is given by

$$\mathcal{S}(t) = \mathcal{C}\left[\exp\left(\frac{i\pi t}{K}\lambda^i\right)\right]_D \mathcal{C}^{-1} \ . \tag{7}$$

Numerical evaluation of Eq. (7) by solving Eq. (6) as an eigenvalue/eigenvector problem requires using a square structure ($\mathcal{A}$) matrix of finite order. Unlike the multislice approach for evaluating $\mathcal{S}_{\mathbf{r},\mathbf{g}}$, where calculations are only needed for those $\mathbf{g}$ of interest (e.g. within the probe-forming aperture in STEM), one

[c] In STEM, where the convergent probe can be considered to contain many incident wavevectors, it is conventional to associate $\mathbf{k}_t$ with the wavevector parallel to the optic axis.

generally needs a larger range of spatial frequencies $\mathbf{g}$ in Eqs. (6) and (7) for a converged calculation. This ultimately makes the Bloch wave approach poorly suited to non-periodic structures and large fields of view, but it is quite tractable for crystalline materials with relatively small lattice parameters. Using a limited number of spatial frequencies $\mathbf{h}$ and $\mathbf{g}$ introduces 'truncation artifacts', but fortunately these are usually confined to the vicinity of the highest spatial frequencies in the $\mathcal{S}$ matrix.

For the case $\mathbf{k}_t = \mathbf{0}$, it follows from Eq. (5) that

$$\mathcal{A}_{\mathbf{h},\mathbf{g}} = \mathcal{A}_{-\mathbf{g},-\mathbf{h}} \ . \tag{8}$$

This property together with Eq. (4) ensures the $\mathcal{S}$ matrix has the same property,

$$\mathcal{S}_{\mathbf{h},\mathbf{g}} = \mathcal{S}_{-\mathbf{g},-\mathbf{h}} \ , \tag{9}$$

which we call antidiagonal symmetry, though using the row/column ordering which makes that description accurate (Allen et al., 2000) is not required for Eqs. (8) and (9) to hold. We stress that this property follows from the projected potential / zero-order Laue zone approximation alone, which is known to be a good approximation in high energy electron microscopy for periodic crystals in low order zone axis orientations (provided the repeat distance along the zone axis is sub-nanometer); it does not require the sample to posses any further symmetry properties. This is proven in the Appendix, where it is also shown why this property ceases to hold in samples that do not satisfy the projected potential approximation for their entire thickness.

If the specimen is crystalline then (Findlay et al., 2003)

$$\mathcal{A}_{\mathbf{q}_1+\mathbf{H},\mathbf{q}_2+\mathbf{G}} = \begin{cases} -(\mathbf{k}_t + \mathbf{q}_1 + \mathbf{H})^2 + iU'_{\mathbf{0}} & \mathbf{q}_1 = \mathbf{q}_2,\ \mathbf{H} = \mathbf{G} \\ U_{\mathbf{H}-\mathbf{G}} + iU'_{\mathbf{H}-\mathbf{G}} & \mathbf{q}_1 = \mathbf{q}_2,\ \mathbf{H} \neq \mathbf{G} \\ 0 & \mathbf{q}_1 \neq \mathbf{q}_2 \end{cases} , \tag{10}$$

which is in block-diagonal form, with the different blocks characterised by different values of $\mathbf{q}_i$ in the first Brillouin zone, and we have used upper case $\mathbf{H}$, $\mathbf{G}$ to denote reciprocal lattice vectors. Because the matrix exponential of a block diagonal matrix is also block diagonal, it follows that for periodic crystals $\mathcal{S}_{\mathbf{h},\mathbf{g}} = 0$ unless $\mathbf{h} - \mathbf{g}$ is a reciprocal lattice vector. Thus for periodic crystals we often write the (non-zero) $\mathcal{S}$ matrix elements as $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}}$.

Approximations and limits

In the limit of zero potential ($U = U' = 0$ in Eq. (5)), the $\mathcal{A}$ and $\mathcal{S}$ matrices are purely diagonal in the Fourier space form of Eq. (2b), with the diagonal elements of the $\mathcal{S}$ matrix describing Fresnel free-space

propagation:

$$\mathcal{S}_{\mathbf{h},\mathbf{g}}(t) = \delta_{\mathbf{h},\mathbf{g}} e^{-i\pi\lambda t h^2} \ . \tag{11}$$

If we instead neglect the $-(\mathbf{k}_t + \mathbf{h})^2$ terms on the diagonal in Eq. (5) then we obtain the so-called phase object approximation:

$$\mathcal{S}_{\mathbf{r},\mathbf{g}} = Q(\mathbf{r}) e^{2\pi i \mathbf{g}\cdot\mathbf{r}} \tag{12a}$$

$$\mathcal{S}_{\mathbf{h},\mathbf{g}} = Q(\mathbf{h}-\mathbf{g}) \ , \tag{12b}$$

where $Q(\mathbf{r}) = e^{i\sigma[V(\mathbf{r})+iV'(\mathbf{r})]t}$ is the transmission function, and $Q(\mathbf{h})$ its Fourier transform.

A variation that is accurate to second order in sample thickness (Van Dyck, 1985; Plamann & Rodenburg, 1998) is to apply the multiplicative phase object at the specimen mid-plane, propagating a distance $t/2$ in free-space before and after:[d]

$$\mathcal{S}_{\mathbf{r},\mathbf{g}} = \hat{\mathcal{P}}_{t/2}\left[Q(\mathbf{r}) e^{2\pi i \mathbf{g}\cdot\mathbf{r}} e^{-i\pi\lambda t g^2/2}\right] \tag{13a}$$

$$\mathcal{S}_{\mathbf{h},\mathbf{g}} = e^{-i\pi\lambda t h^2/2} Q(\mathbf{h}-\mathbf{g}) e^{-i\pi\lambda t g^2/2} \ , \tag{13b}$$

where $\hat{\mathcal{P}}_t = \mathcal{F}^{-1}_{\mathbf{k}\to\mathbf{r}}\left\{e^{-\pi i\lambda t k^2}\mathcal{F}_{\mathbf{r}\to\mathbf{k}}\left[\cdot\right]\right\}$ is the free-space propagation operator for a real-space wavefield.

Visualising the scattering matrix

Matrix as array

Regarding a matrix as a 2D array, the $\mathcal{S}$ matrix could simply be displayed as an image (up to choosing how to represent complex numbers). Figure 1 shows an example for the reciprocal space form $\mathcal{S}_{\mathbf{h},\mathbf{g}}$, where hue represents phase and saturation represents the square root of the modulus (reducing the contrast range in this way improves the visibility of smaller values). This calculation included only 89 beams (i.e. unique values of $\mathbf{g}$), ordered such that the zero beam is in the centre, $\mathbf{g}$ and $-\mathbf{g}$ pairs are placed symmetrically about the centre, and the magnitude of the beams increase away from the centre. The combination of thin sample and compressed modulus scale means that the diagonal elements appear with near-uniform saturation, and their

[d]For small, aperiodic structures, recent work suggests this sort of approximation can be improved by taking the superposition of the scattering from each atom as though in isolation (i.e. propagating to its correct depth, applying a multiplicative transmission function for that single atom, and propagating to the specimen exit plane), amounting to neglecting multiple scattering between atoms and under which circumstances one can solve for the 3D structure (Gureyev et al., 2020a,b).

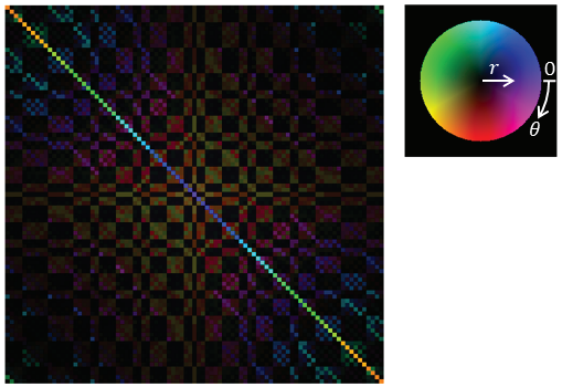

Fig. 1. Reciprocal space form of the scattering matrix $\mathcal{S}_{\mathbf{h},\mathbf{g}}$ displayed with each matrix element as a pixel in the image. The calculation used a 3×3 supercell of $SrTiO_3$ in [001] orientation, a thickness of 39 Å and an accelerating voltage of 300 kV. As per the colour wheel in the top right, complex numbers are displayed by using hue to represent the phase (all phases are displayed between $-\pi$ and π) and saturation to represent the square root (to improve visibility of smaller values) of the modulus. This calculation included only 89 beams $\mathbf{g}$, equivalent to a cut-off at 9 mrad, ordered such that the zero beam is in the centre, $\mathbf{g}$ and $-\mathbf{g}$ pairs are placed symmetrically about the centre, and the magnitude g of the beams increase away from the centre.

phases are a good approximation to those of the free-space propagator. Off-diagonal structure is also discernible and there is some geometric regularity in features. It is tempting to think a more judicious ordering of beams might yield a clearer pattern. However, there is no 'optimum' index ordering of a set of 2D vectors $\{\mathbf{g}\}$, and as such we think representing the $\mathcal{S}$ matrix in the manner of Fig. 1 obscures meaningful structure. Recognising that $\mathcal{S}_{\mathbf{h},\mathbf{g}}$ is better regarded as a four-dimensional quantity, let us explore more informative representations.

Real space visualisation

The real-space representation $\mathcal{S}_{\mathbf{r},\mathbf{g}}$ amounts to representing the exit-surface wavefield for an incident plane wave with transverse wavevector $\mathbf{g}$. Ophus and coworkers have discussed this approach in detail (Pelz et al., 2020; Brown et al., 2020), but we briefly review it here.

Figure 2(a) shows a tableau of 'images' of $\mathcal{S}_{\mathbf{r},\mathbf{g}}$, with each tile corresponding to a different parametric $\mathbf{g}$ value, for a 39 Å thick, [001]-oriented $SrTiO_3$ specimen assuming 300 keV electrons. These complex quantities are displayed using hue to represent the phase and saturation to represent the modulus, as per the colour wheel at the bottom right of Fig. 2. Each tile corresponds to a single unit cell, with Sr columns in

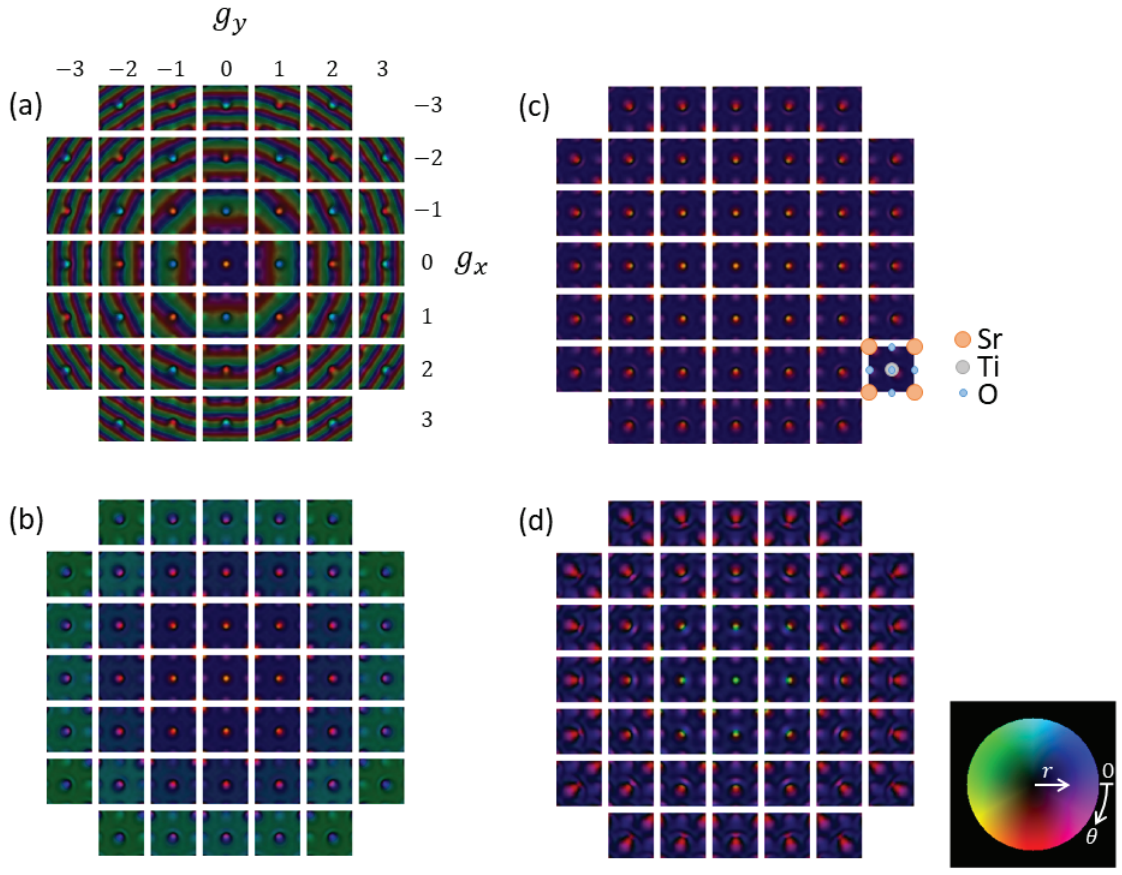

Fig. 2. Tableaux for images (coordinate $\mathbf{r}$ for discrete parametric $\mathbf{g}$ values) of $\mathcal{S}_{\mathbf{r},\mathbf{g}}$-related quantities for [001]-oriented $SrTiO_3$ assuming 300 keV electrons. The calculation included 197 reciprocal lattice vectors $\mathbf{G}$ (equivalent to a cut-off at 2.0 $Å^{-1}$ or 40 mrad). For a 39 Å thick sample we show (a) $\mathcal{S}_{\mathbf{r},\mathbf{g}}$, (b) a transformed version of (a) that removes phase ramps of the form present in Eq. (12a), and (c) a transformed version of (a) that removes phase ramps and propagation factors of the form present in Eq. (13a), with structure schematic inset. (d) As per (c) but for a 78 Å thick sample. Complex numbers are displayed by using hue to represent the phase and saturation to represent the modulus (now without the square root used in Fig. 1) as per the colour wheel.

the corner, the mixed titanium-oxygen column in the centre, and pure oxygen columns at the middle of each edge. The atomistic structure is visible, but somewhat obscured by phase ramp structure on each tile.

Such phase ramps are predicted by the phase object approximation of Eq. (12a). Figure 2(b) shows that tableau with the phase ramps divided out, i.e. $\mathcal{S}_{\mathbf{r},\mathbf{g}}e^{-2\pi i\mathbf{g}\cdot\mathbf{r}}$, making the atomistic structure more clearly visible. The central tile, $\mathbf{g}=\mathbf{0}$, looks very much like the projected potential of $SrTiO_3$. The tiles adjacent to it look very similar, but for $\mathbf{g}$ values further from the origin the phase, and to a lesser extent structure, varies

appreciably. This is not consistent with Eq. (12a): even in this thin sample the phase object approximation is breaking down.

As per Eq. (13a), a better approximation should be to assume the multiplicative phase object at the specimen mid-plane, with free-space propagation before and after. Figure 2(c) shows the tableau from Fig. 2(a) transformed to remove both the phase ramp and propagation factors in Eq. (13a). This succeeds in making the tiles much more similar, suggesting Eq. (13a) is a better approximation, although some slight asymmetry is perceptible at the tiles for $\mathbf{g}$ values further from the origin that is not present in the $\mathbf{g} = \mathbf{0}$ tile. To underscore this, Fig. 2(d) is the result for a 78 Å thick sample after removing both the phase ramp and propagation factors, showing more variation across the different tiles as evidence of scattering effects not accounted for in Eq. (13a).

Reciprocal space visualisation

Figure 1 showed that representing $\mathcal{S}_{\mathbf{h},\mathbf{g}}$ as a two-dimensional image based on matrix element indexing was of limited use. It is possible to take the approach of Fig. 2 and display a tableau of 'images' of $\mathcal{S}_{\mathbf{h},\mathbf{g}}$ (with $\mathbf{h}$ describing the coordinate locations within each tile and each tile corresponding to a different parametric $\mathbf{g}$ value): these would be the diffraction-plane wavefields for incident plane waves with different transverse wavevectors. As per Fig. 2, we would expect a gradual variation with $\mathbf{g}$, but, because the sample considered is crystalline, the diffraction wavefields would consist of discrete Bragg peaks with no real continuity from peak to peak. In mathematical terms, $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}}$ is not continuously varying across (discrete) reciprocal lattice vectors $\mathbf{H}$.

To better clarify the underlying continuity, we can instead visualise $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ as a mosaic of tiles where each tile corresponding to a different parametric $\mathbf{G}$ value and $\mathbf{h}$ describing the coordinate locations within each tile. Thus each tile corresponds to a rocking curve for the Bragg diffracted beam $\mathbf{G}$. Such visualisations are shown in Fig. 3 for three different sample thicknesses (rows in the figure) of $SrTiO_3$ in [001] orientation. For clarity we now plot the modulus (left column, again without the square root used in Fig. 1) and phase (middle column) separately.

The modulus, or rather the intensity that is the square thereof, is precisely what is measured in large-angle rocking-beam electron diffraction, and the present simulations are reminiscent of experimental patterns reported by Koch and coworkers (Koch, 2011; Wang et al., 2016). They are seen to be continuous and reasonably smooth, though finer features become evident at higher thicknesses.

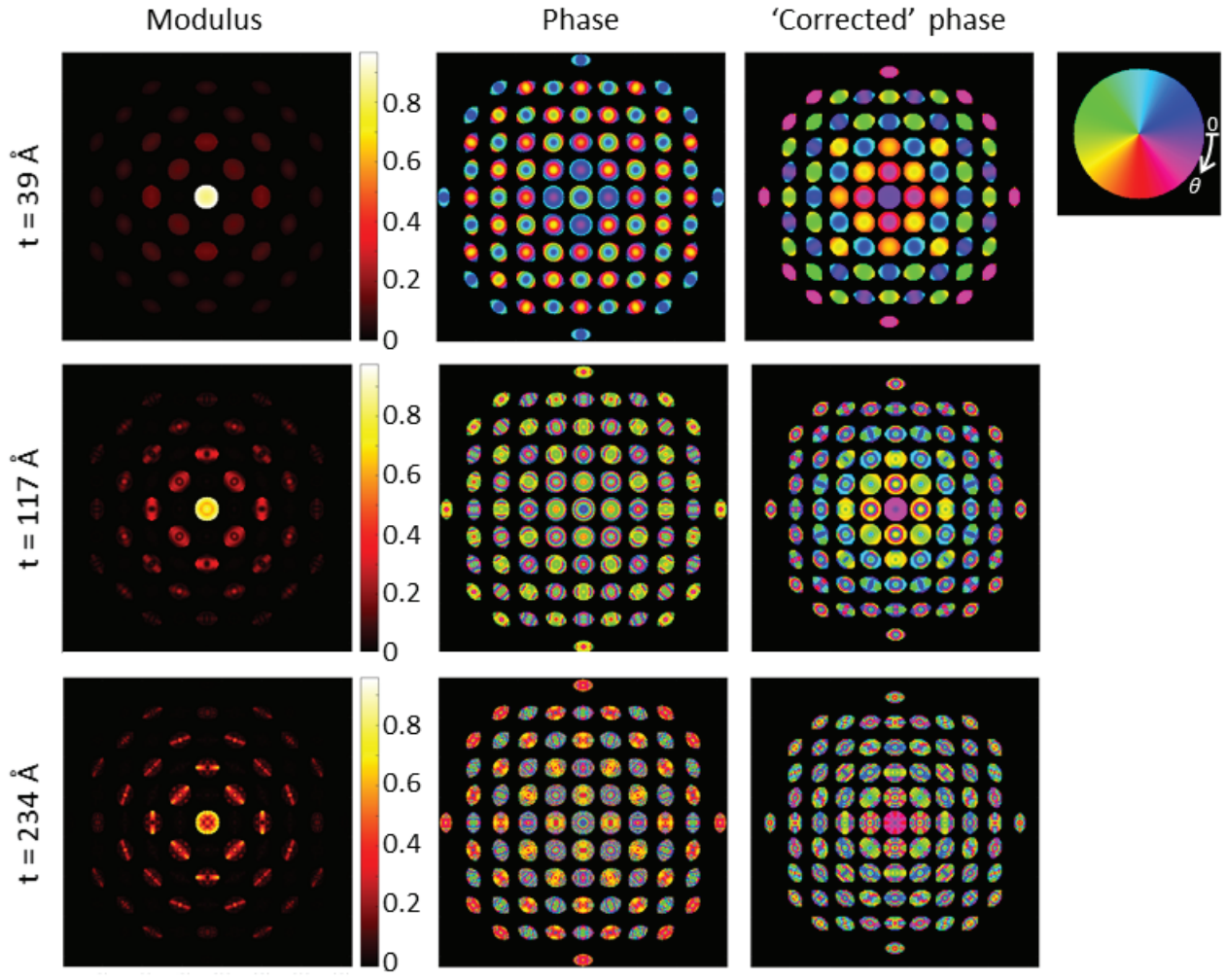

Fig. 3. Visualisation of the $\mathcal{S}$ matrix elements $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$, in which each 'disk' in the tableau corresponds to a different value of $\mathbf{G}$ (with location conforming to the geometry of those reciprocal space lattice points), while the structure internal to each 'disk' shows continuous variation with coordinate $\mathbf{h}$. The calculation used a 3×3 supercell of $SrTiO_3$ in [001] orientation, an accelerating voltage of 300 kV, and included 709 beams $\mathbf{h}$ (81 reciprocal lattice vectors $\mathbf{G}$), equivalent to a cut-off at 25 mrad. The three different rows correspond to different thicknesses as labeled. From left to right, the columns are: modulus (i.e. $|\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}|$), phase (i.e. $\arg\left[\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}\right]$, with colours as per the colour wheel at the top right), and 'corrected' phase (i.e. $\arg\left[e^{i\pi\lambda t(\mathbf{h}+\mathbf{G})^2/2}\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}e^{i\pi\lambda t h^2/2}\right]$). The off-roundness of the outermost 'disks' is an artifact resulting from calculations with a maximum spatial frequency $f_{\max}$: modulus and phase values of $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ are only included for which $|\mathbf{h}+\mathbf{G}| < f_{\max}$. All 'disks' could be extended further by increasing the value of $f_{\max}$ used in the calculation.

Moreover, the phase also varies in a smooth, continuous fashion within each 'disk'. The main features in the phase structure in each 'disk' seems to be near-circular phase contours, with value decreasing from the centre of each 'disk'. The right column shows the phase obtained after correcting for the propagation factors

in Eq. (13b) (which correction does not change the modulus). The central $\mathbf{G} = \mathbf{0}$ 'disk' at each thickness then becomes almost uniform in phase, which implies that when traversing the sample the phase accumulation of the forward-scattered beam is dominated by free-space propagation. The other 'disks' contain more detailed structure, though that too has been somewhat simplified after correcting for the propagation factors. When seeking to reconstruct the (complex) $\mathcal{S}$ matrix elements from measured intensities, this expectation of continuity and smoothness can be applied as a form of regularisation.

Reconstructing the scattering matrix from 4D STEM measurements

Previous approaches

Our discussions focus on STEM approaches, but in passing we note the longer history of proposals for how $\mathcal{S}$ matrices might be reconstructed from measurements in conventional TEM at a careful series of incident beam tilts (see Allen et al. (2000); Donatelli & Spence (2020) and references therein), which has not to our knowledge yet been experimentally achieved.

The intensity in a 4D STEM dataset from a crystalline material is given by

$$I(\mathbf{q}+\mathbf{H},\mathbf{R},\Delta f) = \left| \sum_{\mathbf{G}} \mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}} T_{\Delta f}(\mathbf{q}+\mathbf{G}) e^{-2\pi i(\mathbf{q}+\mathbf{G})\cdot\mathbf{R}} \right|^2 , \tag{14}$$

where $T_{\Delta f}(\mathbf{q}+\mathbf{G})$ is the lens transfer function, comprising both the aperture pupil function and the lens aberration function, though for simplicity we have only explicitly indicated dependence on Δf, the defocus. From intensity measurements $I(\mathbf{h},\mathbf{R},\Delta f)$ across multiple defocus values, for each parametric value of $\mathbf{h} = \mathbf{q}+\mathbf{H}$ (a detector pixel location in 4D STEM) through-focal series phase retrieval allows the elements $\mathcal{S}_{\mathbf{h},\mathbf{g}}$ of the $\mathbf{h}$ row of the $\mathcal{S}$ matrix to be determined. Using a suitable series of detector pixel locations, one can retrieve all columns of an N-beam $\mathcal{S}$ matrix. That the separate retrievals for different rows will not necessarily be correctly phased with respect to one another can, for crystal samples, be overcome by using the expected antidiagonal symmetry of the $\mathcal{S}$ matrix, Eq. (9) (as proposed by Allen et al. (2000) in the context of conventional TEM). Experimental proof-of-principle was shown by Brown et al. (2018) using four defocus values .

To move beyond periodic samples, Pelz et al. (2020) solved for the $\mathcal{S}$ matrix in the form $\mathcal{S}_{\mathbf{r},\mathbf{g}}$ via optimisation based on the measured constraints of a 4D STEM dataset for multiple defocus values. Dealing with

non-periodic samples meant the antidiagonal symmetry constraint did not apply (see Appendix for a justification), but it was found that applying a compact support constraint on the exit surface wavefield produced reliable reconstructions.

Reconstructing the scattering matrix using 4D STEM measurements from a single defocus value

Using joint optimisation to solve for both the $\mathcal{S}$ matrix and the probe wavefield, Pelz et al. (2020) showed that two defocus values may suffice but that more defocus values improved the convergence properties. Let us here relinquish the generality of seeking to simultaneously solve for the probe wavefield (instead assuming it to have been adequately characterised separately) and explore whether it is possible to determine the $\mathcal{S}$ matrix from 4D STEM data from a single defocus value.

Because STEM imaging involves a hard probe-forming aperture, Eq. (14) has the form of a much studied phase retrieval problem: retrieving the phase of an object *with known compact support* (effected by the probe-forming aperture function implicit in $T_{\Delta f}(\mathbf{q}+\mathbf{G})$) from a measurement of the intensity of the Fourier transform of the object. In 2D phase retrieval problems of this kind, 'non-trivial' ambiguities are known to be rare (Bendory et al., 2017). Therefore, provided both 'trivial' ambiguities and noise can be managed, if a solution can be found we would have confidence in its correctness.

Previous phase retrieval work in electron microscopy had the compact support in the real-space image plane and the measured intensity in the Fourier-space diffraction plane (Morishita et al., 2008): the coherent diffractive imaging problem. In contrast, Eq. (14) has the measurements in the real-space synthesised STEM images and compact support in the Fourier space of the coordinates of scattering matrix $\mathcal{S}_{\mathbf{h},\mathbf{g}}$. This reversal has significant consequences. In coherent diffractive imaging, the unknown wavefield within the compact support can be assumed continuous and smooth and the measured diffraction pattern intensity generally drops off with increasing scattering angle. Sufficiently over-determining that problem is then a matter of sampling the diffraction pattern sufficiently finely. However, in the phase problem of Eq. (14) the measured STEM image intensity for any given parametric value of $\mathbf{q}+\mathbf{H}$ is both periodic (assuming a crystalline sample) and bandwidth limited (by the physics of STEM imaging (Dwyer, 2010)), and so contains only a fixed amount of information no matter how finely it is sampled. Furthermore, for parametrically fixed $\mathbf{q}$ and $\mathbf{H}$ the function $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}}$ within the compact support is neither continuous nor smooth. These properties mean the over-sampling ratio, a measure of the degree of over-determination, is effectively capped at around

five. While comfortably above the minimum over-sampling ratio of two needed for a well-determined problem, this is still relatively modest. In addition, one of the 'trivial' ambiguities in phase problems of the form of Eq. (14) is that when $\mathbf{q} = \mathbf{0}$ the complex conjugate with $\mathbf{G} \rightarrow -\mathbf{G}$ of the desired solution is also a solution (Fienup, 1987). This is known to complicate phase retrieval, especially when the compact support is symmetric, which the circular probe-forming aperture is.

It is instructive at this point to make conceptual comparison with ptychography on 4D STEM datasets. For concreteness we envisage the ptychographical iterative engine (PIE) approach and its variants, as pioneered by Rodenburg and co-workers (Faulkner & Rodenburg, 2004; Maiden & Rodenburg, 2009). PIE does not require datasets from multiple defocus values, and its robustness is attributed to the requirement of consistency between overlapping probe positions being effective for avoiding the ambiguities that often stymie coherent diffractive imaging from a single probe position (Rodenburg & Maiden, 2019). It is important to appreciate that most ptychographic methods (though not all — see Maiden et al. (2012); Schloz et al. (2020); Chen et al. (2020)) assume the phase object approximation, which we wish to move beyond. However, let us for a moment consider Eq. (14) if we make the phase object approximation of Eq. (12b), giving[e]

$$I(\mathbf{q}+\mathbf{H},\mathbf{R}) = \left| \sum_{\mathbf{G}} Q(\mathbf{H}-\mathbf{G}) A(\mathbf{q}+\mathbf{G}) e^{-2\pi i(\mathbf{q}+\mathbf{G})\cdot\mathbf{R}} \right|^2 , \tag{15}$$

where for simplicity we have dropped the lens aberrations and replaced the lens transfer function T with the aperture function A. Each point $\mathbf{q}+\mathbf{H}$ gives rise to a 2D STEM image which can be regarded as a compact-support-type phase retrieval problem. The problems are distinct in so much as the STEM images $I(\mathbf{q}+\mathbf{H},\mathbf{R})$ for each different $\mathbf{q}+\mathbf{H}$ value involve distinct pixels on the detector, but interrelated in so much as they all draw on the same Fourier frequencies $Q(\mathbf{H})$ (or at least varying subsets thereof, since which frequencies are included inside the aperture varies between the different $\mathbf{q}+\mathbf{H}$). Suppose we regard them as independent problems and try to solve them separately via standard techniques. Suppose that for some $\mathbf{q}+\mathbf{H}$ problems the solutions are fairly reliable while for others they are poor. Because these separately-treated problems are seeking to determine the same underlying set of Fourier frequencies $Q(\mathbf{H})$, we could then average the corresponding frequencies from the different reconstructions, perhaps preferentially weighting those solutions for which the error metric relative to measured intensities was smaller. The result could then be taken as a new starting guess and the process repeated. Though methodologically different to PIE,

[e]Though phase retrieval strategies have been based on Eq. (15) (Chapman, 1997), it does not describe PIE per se, which is generally formulated with the transmission function Q in real space.

this approach seems to be comparably effective at avoiding ambiguities (simulations not shown — they are effectively a limiting case of more general results shown below).

Let us move beyond the phase object approximation to return to the question of determining the $\mathcal{S}$ matrix via phase retrieval applied to Eq. (14). This is potentially a harder problem since the Fourier frequencies $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}}$ for different $\mathbf{q}+\mathbf{H}$ are no longer the same. If they could be arbitrarily different then for different values of $\mathbf{q}+\mathbf{H}$ Eq. (14) would be a set of strictly independent phase retrieval problems. However, as per the discussion of Fig. 3, we expect $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ to vary smoothly with $\mathbf{h}$. Provided the separate phase retrievals carried out on STEM images corresponding to different values of $\mathbf{q}+\mathbf{H}$ are more successful than not, we propose to use the smoothness of $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ as a regularisation constraint to generate improved starting guesses for refining the phase retrievals and to ameliorate the effects of noise on the reconstructions.

Case study

We take as test case $SrTiO_3$ in [001] orientation, assuming an accelerating voltage of 300 kV. We simulated $\mathcal{S}$ matrices with 197 reciprocal lattice vectors (implying a maximum spatial frequency of 2.0 Å^{-1} or, expressed as a scattering angle, 40 mrad) and a 10×10 sampling of points $\mathbf{q}$ within the Brillouin zone (i.e. a diffraction plane sampling of 0.5 mrad per pixel) for a range of thicknesses. The STEM probe-forming aperture was set to 20 mrad, and the probe assumed to be aberration-free (implying it is focused on the specimen entrance surface, a reasonable approximation to the focal condition for maximum contrast in annular dark field imaging). STEM images were then generated for square synthetic detectors of side length 2.5 mrad[f] within the bright field region at centres corresponding to the 2.5 mrad half-spacing between reciprocal lattice points (there are 56 distinct reciprocal lattice vectors $\mathbf{H}$ for which $\mathbf{q}+\mathbf{H}$ falls in the bright field region), amounting to a 2×2 sampling of the Brillouin zone. STEM images were sampled at 32×32 pixels per unit cell over a 2×2 unit cell field of view. Shot noise was simulated assuming a Poisson distribution based on 930 electrons per probe position, implying a dose of 100 C/cm^2 (or 6.3×10^4 e/Å^2), a typical 'high-dose' condition (Yang et al., 2015). Algorithm 1 summarises the analysis process, which will now be described in detail.

[f]Using extended detectors rather than 'point' detectors improves signal-to-noise within each synthesised STEM image, but, amounting to an integration over $\mathbf{q}+\mathbf{H}$ in Eq. (14), potentially introduces inconsistency into the phase retrieval problems if the diffraction pattern intensity is non-uniform across the extent of the detector. However, for the case study presented here, noise proves to be the greater limitation.

Algorithm 1: $\mathcal{S}$ matrix determination from a single defocus value

Input:

4D STEM dataset $I(\mathbf{q}+\mathbf{H},\mathbf{R})$

STEM probe lens transfer function $T_{\Delta f}(\mathbf{q}+\mathbf{G})$

Hyperparameters for the hybrid input-output algorithm with randomised overrelaxation, and for TGV regularisation

Initialise:

$\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}} \quad \delta_{\mathbf{H},\mathbf{G}}$

$n_{\text{loops}} \quad 5$

Run:

for $i \leftarrow 1$ **to** n_{loops} **do**

  foreach $\mathbf{q}+\mathbf{H}$ *in bright field disk* **do**

    From input $I(\mathbf{q}+\mathbf{H},\mathbf{R})$ *solve Eq. (14) for* $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}}$ *via the hybrid input-output algorithm with randomised overrelaxation*

  end

  Rearrange $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}}$ *into* $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ *format, apply TGV regularisation, then rearrange back again*

end

Define function of t: $\mathcal{M}_{\mathbf{q},\mathbf{H},\mathbf{G}}(t) = e^{-i\pi\lambda t(\mathbf{q}+\mathbf{H})^2}\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}}$

Now apply antidiagonal symmetry: $t \quad \underset{t}{\operatorname{argmin}} \sum_{\mathbf{q},\mathbf{H},\mathbf{G}} \left|\mathcal{M}_{\mathbf{q},\mathbf{H},\mathbf{G}}(t) - \mathcal{M}_{-\mathbf{q},-\mathbf{G},-\mathbf{H}}(t)\right|^2$

Output:

$\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}} = \mathcal{M}_{\mathbf{q},\mathbf{H},\mathbf{G}}(t)$

If the sample thickness were known in advance, a natural choice for initialising the phase retrieval for each synthesised $\mathbf{q}+\mathbf{H}$ STEM image would be to set it to the free-space propagation limit of Eq. (11). Since in practice we do not expect to have reliable foreknowledge of the sample thickness we instead initialise to Eq. (11) sans the propagation factor, i.e. initialise $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}}$ to $\delta_{\mathbf{H},\mathbf{G}}$. We shall presently see that to an excellent approximation we can deduce the thickness from the phase retrieval results and the expected antidiagonal symmetry (Eq. (9)).

For each synthesised $\mathbf{q}+\mathbf{H}$ STEM image, phase retrieval was carried out using the hybrid input-output algorithm with randomised overrelaxation of Köhl et al. (2012), with modifications following Martin et al.

(2012) to improve robustness to noise. We set $\beta = 0.8$ for hybrid input-output and $\nu = 0.5$ for overrelaxation following Köhl et al. (2012) (see that reference for further details about these hyperparameters). The compact support applied was based on the aperture cutoff alone, i.e. we did not enforce the periodicity expected of the 2×2 unit cell field of view. Each phase retrieval cycle comprised two lots of 10 iterations of hybrid input-output with overrelaxation followed by 8 iterations of error reduction (fewer per cycle than Köhl et al. (2012) but we use more cycles; the general reliability of the reconstruction was not very sensitive to the particular values chosen). We ran five such cycles in total, there being little further improvement beyond that point. After each cycle we recast the current estimate for the wavefields into $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ form (similar to Fig. 3) and applied a smoothness regularisation. Specifically, we applied second-order total generalized variation (TGV) (Bredies et al., 2010) (as implemented in the CCPi-regularisation toolkit (Kazantsev et al., 2019)), which approximates the 'disks' via piecewise-affine functions — a reasonable assumption given the resolution of the structure in the 'disks' and one that avoids the staircasing artefacts common in first-order total variation. We set $\alpha_1 = 0.05$ and $\alpha_0 = 0.1$ following Bredies et al. (2010), and found that $\lambda = 0.15$ and 120 Primal-Dual iterations gave smoothing that was perceptible but not overly aggressive, i.e. gave minimal signal spread outside the 'disk' regions (see Bredies et al. (2010); Kazantsev et al. (2019) for further details about these hyperparameters). We note in passing that good results were also obtained when using a 2D Savitsky-Golay filter (Krumm, 2001) (which can be efficiently implemented as a convolution but has the effect of replacing each point with the value predicted by a low-order polynomial fitted across a local region of pixels) as a simpler alternative to TGV regularisation. We present the TGV results since they gives slightly better clarity in the weaker-intensity 'disks'.

Figure 4 shows the results of these reconstructions for the four sample thicknesses 39 Å, 78 Å, 156 Å and 234 Å. Results are shown in two representations. The first is as a tableau in real space after removing the phase ramps, i.e. $e^{2\pi i(\mathbf{q}+\mathbf{H})\cdot\mathbf{R}}\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{R}}$ (reminiscent of Fig. 2, though note that the indices are reversed such that these are STEM 'images' rather than conventional TEM exit-surface wavefields). The second is via the $S_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ representation in reciprocal space (similar to Fig. 3, though now with modulus and phase information represented in a single image). The images in the left column are split into left and right half-panes, where the former shows results after one cycle and prior to TGV regularisation, while the latter shows results after five cycles and TGV regularisation. In the real space case, the entire central column of tiles is included in both half-panes to allow direct comparison. Close inspection shows that in the real space representation the central ($\mathbf{q} + \mathbf{H} = \mathbf{0}$) tile is not sensibly reconstructed after one cycle for the thinner samples, and that

the nearby tiles also show the most pronounced asymmetries due to noise and under-convergence of the phase retrieval. At the same point in the reconstruction, the results for the thicker samples show clearer structure, the greater contrast in those STEM images having made the phase retrieval more effective. In the reciprocal space representation, the consequences of noise are most evident in the weaker-intensity 'disks'. The results after five cycles (the right half-panes in the left column of images) show a perceptible improvement, especially in the central tiles of the real space images, though some irregularities resulting from noise remain evident in the reciprocal space representation.

The left half-pane of the left column is particularly significant. To that point in the phase retrieval, the STEM images for different $\mathbf{q}+\mathbf{H}$ values were treated as completely independent phase retrieval problems. Since one trivial ambiguity of such phase problems is that phase is only defined up to an additive constant (in the real space representation), had we seeded the reconstructions with random starting phases we might expect no clear phase relation between different tiles in the real space representation (or points within each 'disk' in the reciprocal space representation). Instead, we have the strong visual impression of a phase relation: the mean phase in all real space tiles appears very similar (implied by the images all having very similar colours) and the central 'disk' in the reciprocal space representation has nearly constant phase (uniform colour). The latter is consistent with minimal variation of the central 'disk' away from the $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}} = \delta_{\mathbf{H},\mathbf{G}}$ initialisation. This overall smoothness of the $S_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ representation gives confidence that TGV regularisation is appropriate, that localised imperfections primarily reflect noise and inadequate convergence in the phase retrieval to that point.

Having largely treated the phase retrievals for the different $\mathbf{q}+\mathbf{H}$ in Eq. (14) as distinct problems, the different rows of the retrieved $\mathcal{S}$ matrix will not necessarily be correctly phased with respect to one another. As mentioned earlier, for crystal samples this can be overcome by using the expected antidiagonal symmetry of the $\mathcal{S}$ matrix, Eq. (9) (Allen et al., 2000). One possibility is to fit for arbitrary phase factors between the rows to best satisfy Eq. (9) (Brown et al., 2018), but we will take a slightly different approach. The $\mathcal{S}_{\mathbf{q}+\mathbf{H},\mathbf{q}+\mathbf{G}} = \delta_{\mathbf{H},\mathbf{G}}$ initialisation corresponds, in the $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ visualisation, to a single central 'disk' with uniform amplitude and phase. The reconstructions in the left column in Fig. 4 introduce additional 'disks' but largely maintain the uniformity of the central 'disk'. By contrast, in discussing the the $\mathcal{S}_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ visualisation in Fig. 3 we found that for the (correctly phased) $\mathcal{S}$ matrix the central 'disk' closely resembled that expected for free-space propagation through the thickness of the sample. Consequently, in seeking to enforce Eq. (9)

we will optimise for the single parameter t assuming a missing phase factor of the form $e^{-i\pi\lambda t(\mathbf{q}+\mathbf{H})^2}$ across all the different $\mathbf{q}+\mathbf{H}$ images, precisely that omitted from our initialisation of the phase retrievals.

Table 1 compares the values of t that optimise the $\mathcal{S}$ matrices having antidiagonal symmetry to the reference sample thicknesses input to generate the simulated data. The agreement is striking, with all reconstructed values agreeing with the true values to less than a nanometer. (These results are for one noise realisation; the specific reconstructed thickness values vary somewhat with noise realisation, but the similarity to the reference thickness values remains.) We emphasise that the only assumption about the structure that has gone into retrieving the sample thickness has been the assumption of crystallinity required for the antidiagonal symmetry to hold (as Eq. (9) only follows from Eq. (8) when Eq. (4) correctly describes scattering through the full sample). Everything else came from the iterative phase retrieval applied to the 4D STEM intensity. This accuracy of thickness determination is comparable to that obtained from position-averaged convergent beam electron diffraction (LeBeau et al., 2010), but does not require a known structure with which to perform reference simulations.

Table 1. Comparison (in angstrom units) of the reference sample thickness input to the simulation with that obtained through optimisation for the antidiagonal symmetry constraint on the $\mathcal{S}$ matrices determined from phase retrieval.

Reference	39	78	117	156	195	234
Reconstructed	42	78	115	150	188	239

The right column in Fig. 4 compares the final reconstruction (i.e. after having enforced the antidiagonal symmetry constraint) with the expected results (i.e. the $\mathcal{S}_{\mathbf{h},\mathbf{g}}$ used in Eq. (14) to produce the simulated 4D STEM dataset on which this phase retrieval was based). Close agreement between the left and right half-panes in the right column of Fig. 4 would indicate a successful reconstruction. The agreement is fairly good, especially for the thinner samples. Closer inspection shows some differences. For the thinner samples, these include distortions in the near-centre tiles in the real space representation and some asymmetric features within the 'disks' in the reciprocal space representation. For the largest thickness, especially 234 Å, there are more perceptible differences in both structure and phase values, though the overall pattern remains quite similar.

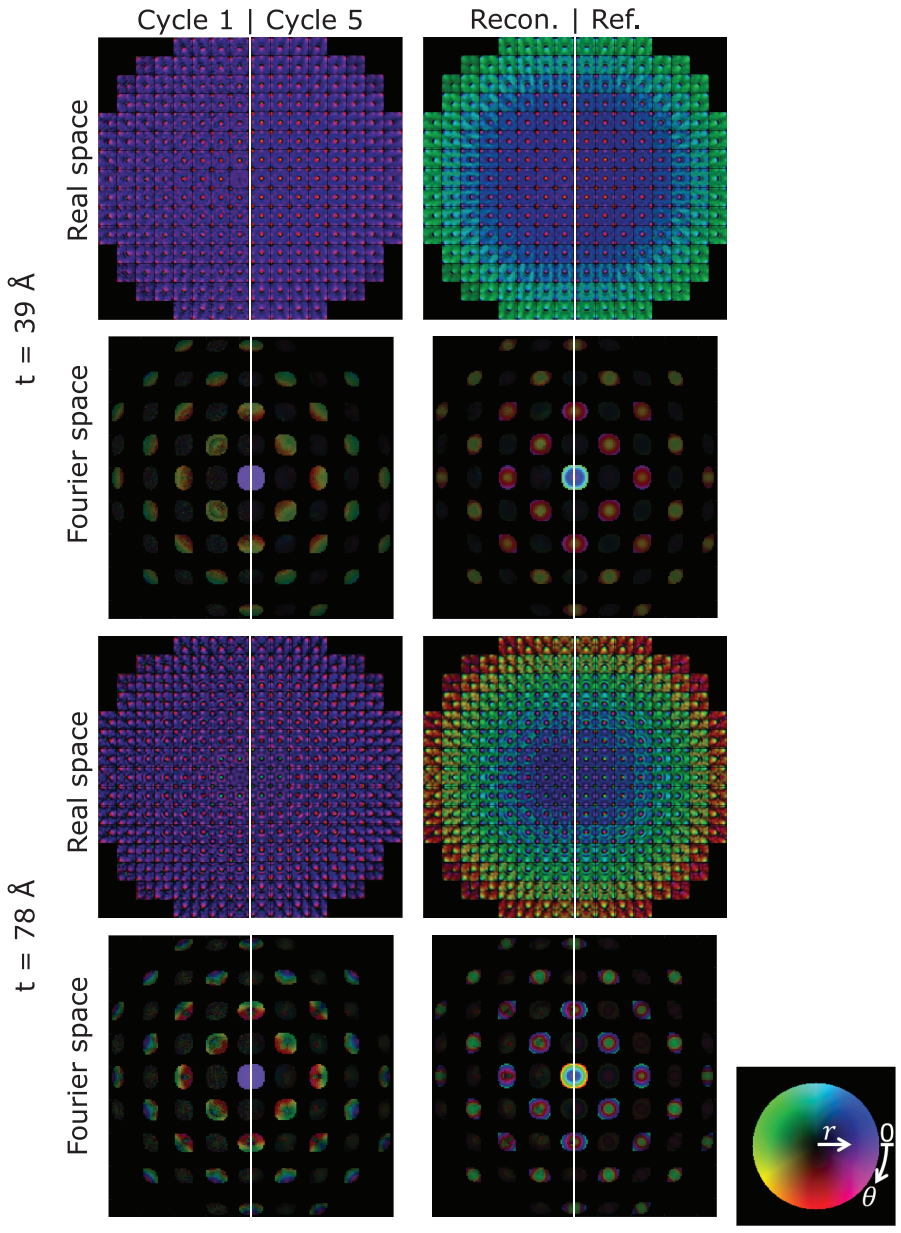

Figure continued on next page...

Discussion

The comparisons in Fig. 4 of the reconstructed $\mathcal{S}$ matrix against its expected value appear favourable for the most part, but such visual comparison does not convey the consequences of the perceptible errors. To that end, Fig. 5 compares a defocus series of annular bright field (detector spanning 10-20 mrad) and central bright field (detector spanning 0-10 mrad) STEM images simulated from the reconstructed $\mathcal{S}$ matrices for the 39 Å and 156 Å thick samples against those of the reference $\mathcal{S}$ matrix (i.e. those on which the 4D STEM simulations were based). We stress that the 4D STEM dataset used for the reconstructions are based was

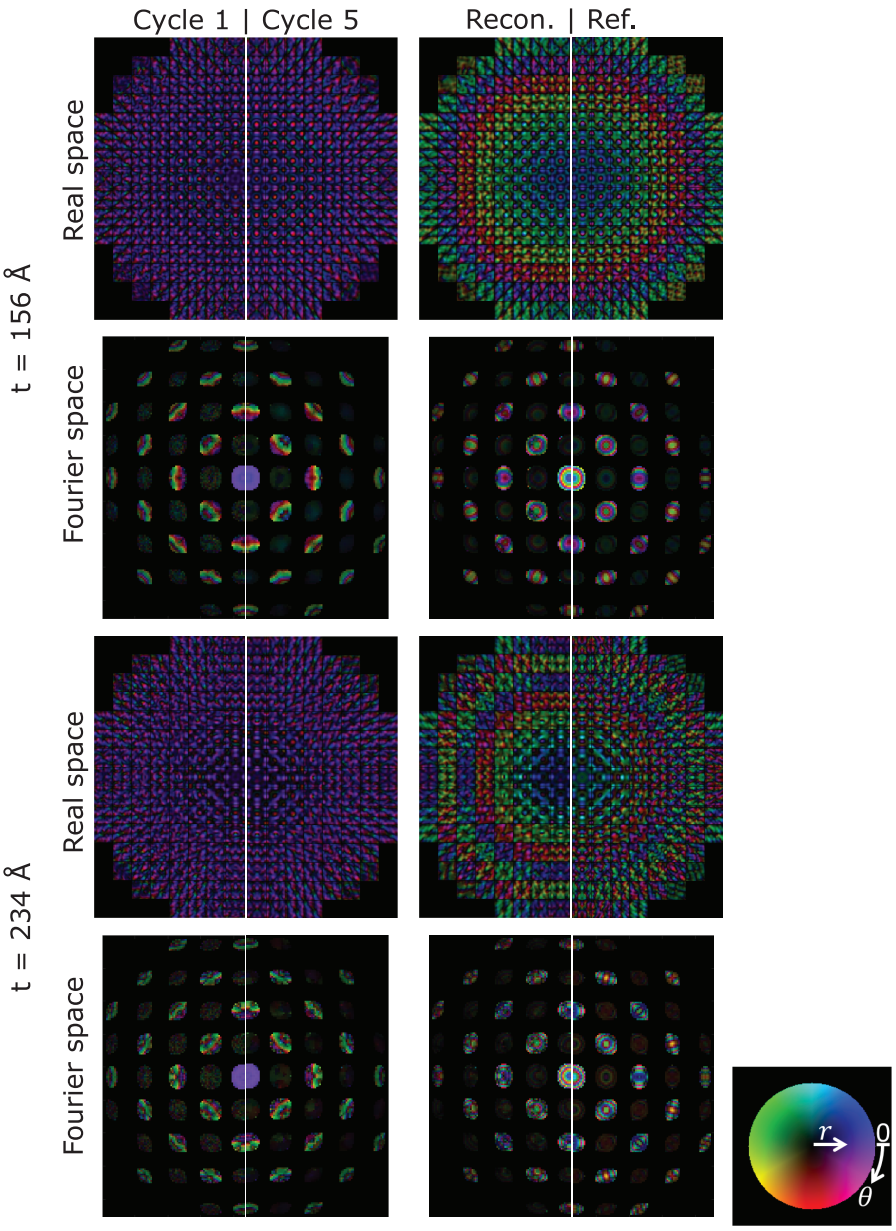

Fig. 4. $\mathcal{S}$ matrix reconstructions from simulated 4D STEM data including shot noise. Results are shown for the four different sample thicknesses 39 Å, 78 Å, 156 Å and 234 Å, both as a tableau in real space (after removing the phase ramps) and in $S_{\mathbf{h}+\mathbf{G},\mathbf{h}}$ representation in reciprocal space. The left column compares results after one cycle (but prior to TGV regularisation) with results after five cycles (post TGV regularisation) in a half-pane format. The right column compares the final reconstruction ('Recon.'; after enforcing the antidiagonal symmetry constraint) with the expected result ('Ref.'; that on which the 4D STEM simulations were based), again in a half-pane format. In the real space case, the entire central column of tiles is included in both half-panes to allow direct comparison. As per the colour wheel, complex numbers are displayed by using hue to represent the phase and saturation to represent the modulus for the real space representation and the square root of the modulus for the reciprocal space representation, the latter to render the fainter 'disks' more visible.

purely for zero defocus. Simulating STEM images at different defocus values is not possible by applying sythetic detectors to the original dataset but only as a consequence of having reconstructed the $\mathcal{S}$ matrix.

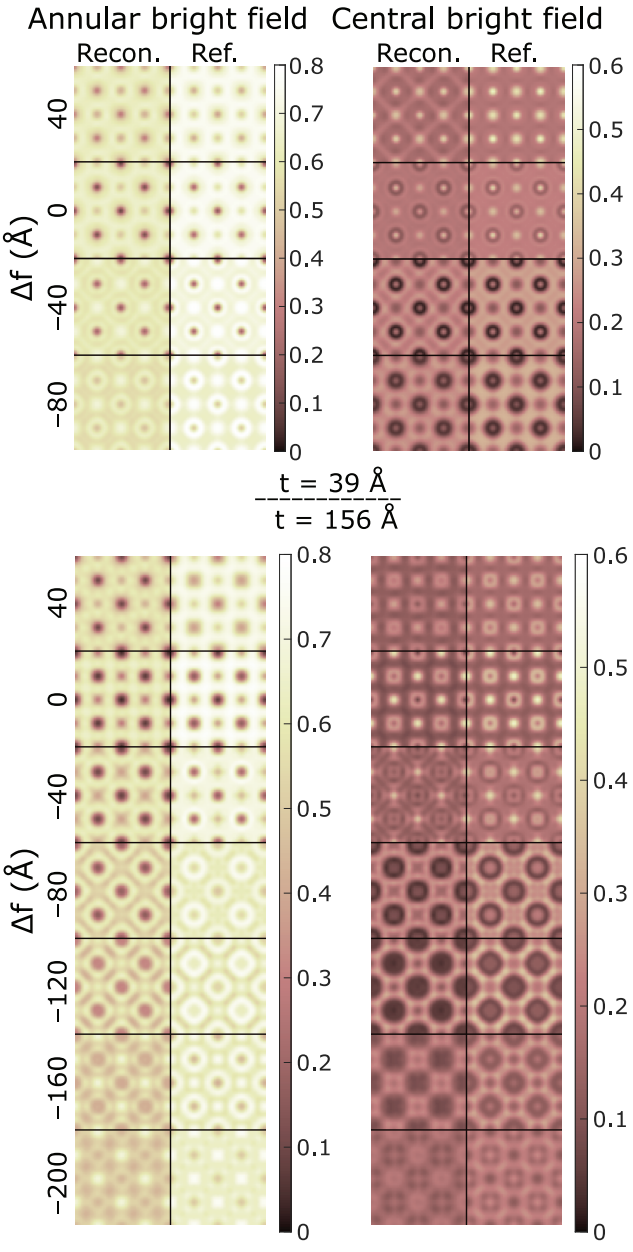

Fig. 5. Comparison of the reconstructed ('Recon.') and reference ('Ref.') $\mathcal{S}$ matrix simulations of annular bright field images (detector spanning 10-20 mrad) and central bright field images (detector spanning 0-10 mrad) for the 39 Å thick sample (upper portion) and 156 Å thick sample (lower portion) each for a range of defocus values (overfocus being positive). The agreement is good for the 39 Å thick sample. For the 156 Å sample, the agreement is fair for the defocus values near zero, but discrepancies become increasingly evident for increasingly large underfocus values. There is also a systematic slight underestimate of intensity in all images from the reconstructed $\mathcal{S}$ matrix.

For the 39 Å thick sample, the qualitative appearance of the simulated images at different defocus values are in excellent agreement with the reference simulations, though quantitatively the average intensities in the images synthesised from the reconstructed $\mathcal{S}$ matrix are systematically slightly lower than the reference simulations. We attribute this primarily to noise impacting both phase and amplitude of the higher spatial frequencies, which then causes some cancellation in the averaging-like regularisation process. The finite detector size used in processing the 4D STEM data may also be a contributing factor, especially for ABF since the detector extent encompassing some of the dark field region for points near the edge of the bright field disk effectively reduced the intensity at the outer edges of the bright field disk.

The qualitative appearance of the simulated images for the 156 Å thick sample is less good than for the 39 Å sample: still fair for most defocus values, but becoming notably poorer for the largest underfocus values. These are the cases for which the phase gradient across $\psi_{\mathrm{in}}(\mathbf{g})$ is steepest and so accurate scattering calculations are most sensitive to phase errors in the reconstructed $\mathcal{S}$ matrix. Results for the 234 Å case (not shown) show more pronounced adverse consequences of the larger reconstruction errors for that thicker sample. Both cases can be improved substantially by using 4D STEM datasets from two different defocus values (not shown), which makes the phase retrieval more robust, consistent with previous work (Pelz et al., 2020).

The simulations in this paper have included shot noise and finite detector size, the former being the dominant effect of the parameters chosen. On 'perfect' (i.e. simulated) data without these effects, near-perfect reconstruction is possible, so ameliorating these effects — for instance by using higher dose if the sample tolerates it, or using a larger field of view if the sample is crystalline — should improve the reconstruction. Effects such as residual probe aberrations, spatial and temporal incoherence, scan distortion, phonon and plasmon scattering backgrounds and amorphous surface scattering have not been considered. All would be expected to erode the quality of the data and therefore the reliability of the reconstruction, though several can be ameliorated by careful instrument characterisation and preprocessing. The success of proof-of-principle $\mathcal{S}$-matrix reconstructions from experimental 4D STEM datasets at multiple defocus values (Brown et al., 2018; Pelz et al., 2020; Brown et al., 2020) implies that none of these effects are overly prohibitive, but further work is needed to establish the quality of the experimental data and the degree of instrument characterisation necessary to realise $\mathcal{S}$ matrix reconstruction using 4D STEM measurements from a single defocus value in practice.

Conclusion

Through simulation, we have shown that an atomic resolution 4D STEM dataset from a single defocus value suffices to reconstruct the scattering matrix when regularised by the continuity expected therein. Shot noise and finite detector size effects included in the simulated 4D STEM datasets limit the accuracy of the reconstructed scattering matrices, which, for instance, limit the reliability with which they could be used to simulate annular bright field STEM images at defocus values significantly different (tens of nanometers) from that at which the data was generated.

For periodic crystals, a particularly significant by-product of reconstructing the scattering matrix is an accurate (to within a nanometer or two) determination of the sample thickness, without requiring any reference to simulation or any other assumptions about the sample structure.

Acknowledgements. This research was supported under the Discovery Projects funding scheme of the Australian Research Council (Project No. FT190100619). Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. JC acknowledges support from the Presidential Early Career Award for Scientists and Engineers (PECASE) through the US Department of Energy. CO acknowledges support from the US DOE early career research award program. PMP acknowledges financial support from STROBE: A National Science Foundation Science & Technology Center under Grant No. DMR 1548924. We thank Dr Zhen Chen for initial algorithm development towards this work, and Drs Zhen Chen, Timothy Petersen and Matthew Weyland for helpful discussions.

References

Allen, L.J., Faulkner, H.M.L. & Leeb, H. (2000). Inversion of dynamical electron diffraction data including absorption, *Acta Crystallographica Section A: Foundations of Crystallography* **56**, 119–126.

Bendory, T., Beinert, R. & Eldar, Y.C. (2017). Fourier phase retrieval: Uniqueness and algorithms, *Compressed Sensing and its Applications*, 55–91, Springer.

Bosch, E.G.T. & Lazić, I. (2019). Analysis of depth-sectioning stem for thick samples and 3D imaging, *Ultramicroscopy* **207**, 112831.

Bredies, K., Kunisch, K. & Pock, T. (2010). Total generalized variation, *SIAM Journal on Imaging Sciences* **3**, 492–526.

Brown, H.G., Chen, Z., Weyland, M., Ophus, C., Ciston, J., Allen, L.J. & Findlay, S.D. (2018). Structure retrieval at atomic resolution in the presence of multiple scattering of the electron probe, *Physical Review Letters* **121**, 266102.

Brown, H.G., Pelz, P.M., Hsu, S.L., Zhang, Z., Ramesh, R., Inzani, K., Sheridan, E., Griffin, S.M., Findlay, S.D., Allen, L.J., Scott, M.C., Ophus, C. & Ciston, J. (2020). A three-dimensional reconstruction algorithm for scanning transmission electron microscopy data, *arXiv preprint arXiv:201107652*.

Chapman, H.N. (1997). Phase-retrieval x-ray microscopy by Wigner-distribution deconvolution: signal processing, *Scanning Microscopy* **11**, 67–80.

Chen, Z., Jiang, Y., Shao, Y.T., Holtz, M.E., Odstrčil, M., Guizar-Sicairos, M., Hanke, I., Ganschow, S., Schlom, D.G. & Muller, D.A. (2020). Electron ptychography achieves atomic-resolution limits set by lattice vibrations, *arXiv preprint arXiv:210100465*.

Chen, Z., Weyland, M., Ercius, P., Ciston, J., Zheng, C., Fuhrer, M.S., D'Alfonso, A.J., Allen, L.J. & Findlay, S.D. (2016). Practical aspects of diffractive imaging using an atomic-scale coherent electron probe, *Ultramicroscopy* **169**, 107–121.

Close, R., Chen, Z., Shibata, N. & Findlay, S.D. (2015). Towards quantitative, atomic-resolution reconstruction of the electrostatic potential via differential phase contrast using electrons, *Ultramicroscopy* **159**, 124–137.

Donatelli, J.J. & Spence, J.C.H. (2020). Inversion of many-beam Bragg intensities for phasing by iterated projections: Removal of multiple scattering artifacts from diffraction data, *Physical Review Letters* **125**, 065502.

Dwyer, C. (2010). Simulation of scanning transmission electron microscope images on desktop computers, *Ultramicroscopy* **110**, 195–198.

Faulkner, H.M.L. & Rodenburg, J.M. (2004). Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm, *Physical Review Letters* **93**, 023903.

Fienup, J.R. (1987). Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint, *Journal of the Optical Society of America A* **4**, 118–123.

Findlay, S.D. (2005). Quantitative structure retrieval using scanning transmission electron microscopy, *Acta Crystallographica Section A: Foundations of Crystallography* **61**, 397–404.

Findlay, S.D., Allen, L.J., Oxley, M.P. & Rossouw, C.J. (2003). Lattice-resolution contrast from a focused coherent electron probe. Part II, *Ultramicroscopy* **96**, 65–81.

Gao, S., Wang, P., Zhang, F., Martinez, G.T., Nellist, P.D., Pan, X. & Kirkland, A.I. (2017). Electron ptychographic microscopy for three-dimensional imaging, *Nature Communications* **8**, 163.

Gureyev, T.E., Quiney, H.M., Kozlov, A. & Allen, L.J. (2020a). Relative roles of multiple scattering and fresnel diffraction in the imaging of small molecules using electrons, *Ultramicroscopy* **218**, 113094.

Gureyev, T.E., Quiney, H.M., Kozlov, A., Paganin, D.M., Schmalz, G. & Allen, L.J. (2020b). Relative roles of multiple scattering and fresnel diffraction in the imaging of small molecules using electrons, Part II: Differential holographic tomography, *arXiv preprint arXiv:201207012*.

Hachtel, J.A., Idrobo, J.C. & Chi, M. (2018). Sub-ångstrom electric field measurements on a universal detector in a scanning transmission electron microscope, *Advanced Structural and Chemical Imaging* **4**, 10.

Jiang, Y., Chen, Z., Han, Y., Deb, P., Gao, H., Xie, S., Purohit, P., Tate, M.W., Park, J., Gruner, S.M. et al. (2018). Electron ptychography of 2D materials to deep sub-ångström resolution, *Nature* **559**, 343–349.

Kazantsev, D., Pasca, E., Turner, M.J. & Withers, P.J. (2019). Ccpi-regularisation toolkit for computed tomographic image reconstruction with proximal splitting algorithms, *SoftwareX* **9**.

Koch, C.T. (2011). Aberration-compensated large-angle rocking-beam electron diffraction, *Ultramicroscopy* **111**, 828–840.

Köhl, M., Minkevich, A.A. & Baumbach, T. (2012). Improved success rate and stability for phase retrieval by including randomized overrelaxation in the hybrid input output algorithm, *Optics Express* **20**, 17093–17106.

Krumm, J. (2001). SavitzkyâĂŞGolay filters for 2D Images, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_C [Accessed: 28-October-2020].

Lazić, I., Bosch, E.G.T. & Lazar, S. (2016). Phase contrast STEM for thin samples: Integrated differential phase contrast, *Ultramicroscopy* **160**, 265–280.

LeBeau, J.M., Findlay, S.D., Allen, L.J. & Stemmer, S. (2010). Position averaged convergent beam electron diffraction: Theory and applications, *Ultramicroscopy* **110**, 118–125.

Maiden, A.M., Humphry, M.J. & Rodenburg, J.M. (2012). Ptychographic transmission microscopy in three dimensions using a multi-slice approach, *J Opt Soc Am A* **29**, 1606–1614.

Maiden, A.M. & Rodenburg, J.M. (2009). An improved ptychographical phase retrieval algorithm for diffractive imaging, *Ultramicroscopy* **109**, 1256–1262.

Martin, A.V., Wang, F., Loh, N.D., Ekeberg, T., Maia, F.R., Hantke, M., van der Schot, G., Hampton, C.Y., Sierra, R.G., Aquila, A. et al. (2012). Noise-robust coherent diffractive imaging with a single diffraction pattern, *Optics Express* **20**, 16650–16661.

Moler, C. & Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later, *SIAM review* **45**, 3–49.

Morishita, S., Yamasaki, J., Nakamura, K., Kato, T. & Tanaka, N. (2008). Diffractive imaging of the dumbbell structure in silicon by spherical-aberration-corrected electron diffraction, *Applied Physics Letters* **93**, 183103.

Müller, K., Krause, F.F., Béché, A., Schowalter, M., Galioit, V., Löffler, S., Verbeeck, J., Zweck, J., Schattschneider, P. & Rosenauer, A. (2014). Atomic electric fields revealed by a quantum mechanical approach to electron picodiffraction, *Nature Communications* **5**, 5653.

Müller-Caspary, K., Krause, F.F., Grieb, T., Löffler, S., Schowalter, M., Béché, A., Galioit, V., Marquardt, D., Zweck, J., Schattschneider, P. et al. (2017). Measurement of atomic electric fields and charge densities from average momentum transfers using scanning transmission electron microscopy, *Ultramicroscopy* **178**, 62–80.

Ophus, C. (2017). A fast image simulation algorithm for scanning transmission electron microscopy, *Advanced Structural and Chemical Imaging* **3**, 13.

Ophus, C., Harvey, T.R., Yasin, F.S., Brown, H.G., Pelz, P.M., Savitzky, B.H., Ciston, J. & McMorran, B.J. (2019). Advanced phase reconstruction methods enabled by four-dimensional scanning transmission electron microscopy, *Microscopy and Microanalysis* **25 (Suppl 2)**, 10–11.

Pelz, P.M., Brown, H.G., Ciston, J., Findlay, S.D., Zhang, Y., Scott, M. & Ophus, C. (2020). Reconstructing the scattering matrix from scanning electron diffraction measurements alone, *arXiv preprint arXiv:200812768*.

Plamann, T. & Rodenburg, J.M. (1998). Electron ptychography. II. Theory of three-dimensional propagation effects, *Acta Crystallographica Section A: Foundations of Crystallography* **54**, 61–73.

Ren, D., Ophus, C., Chen, M. & Waller, L. (2020). A multiple scattering algorithm for three dimensional phase contrast atomic electron tomography, *Ultramicroscopy* **208**, 112860.

Rodenburg, J. & Maiden, A. (2019). Ptychography, *Springer Handbook of Microscopy*, 819–904, Springer.

Schloz, M., Pekin, T.C., Chen, Z., Broek, W.V.d., Muller, D.A. & Koch, C.T. (2020). Overcoming information reduced data and experimentally uncertain parameters in ptychography with regularized optimization, *Optics Express* **28**, 28306–28323.

Spence, J.C.H. (1998). Direct inversion of dynamical electron diffraction patterns to structure factors, *Acta Crystallographica Section A: Foundations of Crystallography* **54**, 7–18.

Sturkey, L. (1962). The calculation of electron diffraction intensities, *Proceedings of the Physical Society* **80**, 321.

Tate, M.W., Purohit, P., Chamberlain, D., Nguyen, K.X., Hovden, R., Chang, C.S., Deb, P., Turgut, E., Heron, J.T., Schlom, D.G. et al. (2016). High dynamic range pixel array detector for scanning transmission electron microscopy, *Microscopy and Microanalysis* **22**, 237–249.

Van den Broek, W. & Koch, C.T. (2012). Method for retrieval of the three-dimensional object potential by inversion of dynamical electron scattering, *Physical Review Letters* **109**, 245502.

Van den Broek, W. & Koch, C.T. (2013). General framework for quantitative three-dimensional reconstruction from arbitrary detection geometries in TEM, *Physical Review B* **87**, 184108.

Van Dyck, D. (1985). Image calculations in high-resolution electron microscopy: problems, progress, and prospects, *Advances in Electronics and Electron Physics* **65**, 295–355.

Wang, F., Pennington, R.S. & Koch, C.T. (2016). Inversion of dynamical scattering from large-angle rocking-beam electron diffraction patterns, *Physical Review Letters* **117**, 015501.

Winkler, F., Barthel, J., Dunin-Borkowski, R.E. & Mueller-Caspary, K. (2020). Direct measurement of electrostatic potentials at the atomic scale: A conceptual comparison between electron holography and scanning transmission electron microscopy, *Ultramicroscopy* **210**, 112926.

Yang, H., Pennycook, T.J. & Nellist, P.D. (2015). Efficient phase contrast imaging in stem using a pixelated detector. Part II: Optimisation of imaging conditions, *Ultramicroscopy* **151**, 232–239.

Yang, H., Rutte, R.N., Jones, L., Simson, M., Sagawa, R., Ryll, H., Huth, M., Pennycook, T.J., Green, M.L.H., Soltau, H., Kondo, Y., Davis, B.G. & Nellist, P.D. (2016). Simultaneous atomic-resolution electron ptychography and z-contrast imaging of light and heavy elements in complex nanostructures, *Nature Communications* **7**.

Appendix: Antidiagonal symmetry of the $\mathcal{S}$ matrix – mathematical derivation

In indexing the elements of the structure matrix $\mathcal{A}_{\mathbf{h},\mathbf{g}}$ suppose we choose an ordering such that $\mathbf{g} = \mathbf{0}$ is in the middle, and for which, for $\mathbf{g} \neq \mathbf{0}$, vectors $\mathbf{g}$ and $-\mathbf{g}$ are symmetrically spaced about that middle. Under this ordering, the symmetry properties of Eq. (8) – which assumes $\mathbf{k}_t = \mathbf{0}$ – can be written in integer-index labelling as (Allen et al., 2000)

$$\mathcal{A}_{i,j} = \mathcal{A}_{N-j+1,N-i+1} \; . \tag{1}$$

This corresponds to symmetry across the anti-diagonal of the matrix. Mathematicians call such matrices persymmetric, and this property can be written as

$$\mathcal{A}\mathcal{J} = \mathcal{J}\mathcal{A}^T \tag{2}$$

where $\mathcal{J}$ is an $N \times N$ exchange matrix:

$$\mathcal{J}_{i,j} = \begin{cases} 1 & j = N - i + 1 \\ 0 & j \neq N - i + 1 \end{cases} \; . \tag{3}$$

Note that $\mathcal{J}^{2n} = \mathcal{I}$ (i.e. the identity matrix) for integer n.

The product of persymmetric matrices is persymmetric if and only if they commute. Eq. (5) by definition means that

$$\mathcal{S}(t) = \exp\left(\frac{i\pi t}{K}\mathcal{A}\right) = \sum^{\infty} \frac{1}{n!}\left(\frac{i\pi t}{K}\right)^n \mathcal{A}^n \tag{4}$$

Consequently, since the $\mathcal{S}$ matrix comprises a sum of powers of persymmetric structure matrices (which necessarily commute with themselves), the $\mathcal{S}$ matrix is also persymmetric,

$$\mathcal{S}_{i,j} = \mathcal{S}_{N-j+1,N-i+1} \ , \tag{5}$$

which upon rewriting using vector subscript notation yields Eq. (9).

This derivation further makes clear why this property is largely restricted to perfect crystals (more specifically to samples for which the projected potential approximation holds for the entire thickness): in a case where the $\mathcal{S}$ matrix for the sample needs to be expressed as the product of the $\mathcal{S}$ matrices for each of a series of structurally-distinct slices, though the individual $\mathcal{S}$ matrices are persymmetric their product is not since the $\mathcal{S}$ matrices from structurally-distinct slices will generally not commute.