Lawrence Berkeley National Laboratory

Title
Reconstructing the Scattering Matrix from Scanning Electron Diffraction Measurements Alone

Permalink
https://escholarship.org/uc/item/84w777qt

Authors
Pelz, Philipp M
Brown, Hamish G
Ciston, Jim
et al.

Reconstructing the Scattering Matrix from Scanning Electron Diffraction Measurements Alone

Philipp M Pelz,[1,2,*] Hamish G Brown,[2] Jim Ciston,[2] Scott D
Findlay,[3] Yaqian Zhang,[1] Mary Scott,[1,2] and Colin Ophus[2,†]

[1]*Department of Materials Science and Engineering,
University of California Berkeley, Berkeley, CA 94720*

[2]*NCEM, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720*

[3]*School of Physics and Astronomy, Monash University, Clayton VIC 3800, Australia*

(Dated: August 31, 2020)

Three-dimensional phase contrast imaging of multiply-scattering samples in X-ray and electron microscopy is extremely challenging, due to small numerical apertures, the unavailability of wavefront shaping optics, and the highly nonlinear inversion required from intensity-only measurements. In this work, we present a new algorithm using the scattering matrix formalism to solve the scattering from a non-crystalline medium from scanning diffraction measurements, and recover the illumination aberrations. Our method will enable 3D imaging and materials characterization at high resolution for a wide range of materials.

I. INTRODUCTION

Phase contrast imaging is widely used in light [1, 2], x-ray [3, 4], and electron microscopy [5, 6], due to its high efficiency and resolution. By using coherent radiation to illuminate a sample, we can resolve very small changes in a sample's local index of refraction through the interference of the illumination wave fronts that the accumulated phase shifts produce [7]. However, because we can only directly measure the probability density of a illumination wave function (given by the wave intensity, or amplitude squared), phase contrast imaging is a fundamentally nonlinear measurement process: we must indirectly infer the underlying relative phase shifts induced by the sample [8].

Various approximations can make phase contrast microscopy data easier to interpret. The first is by assuming that the sample is a pure phase object, i.e. it does not modulate the illumination wave function amplitude directly, and so any variations in the measured intensity can be directly ascribed to changes in the sample's index of refraction [9]. However this assumption does not guarantee uniqueness in all cases, due the possibly of phase wrapping [10]. An even stronger assumption is the weak phase object approximation (WPOA), where the sample's transmission function is assumed to be a small imaginary perturbation a known carrier wave [11]. When the WPOA holds, the linear relation implied between specimen potential and measured intensity allow constructive and unambiguous solution. Another commonly used simplification in phase contrast microscopy is the projection approximation

(PA), where all scattering is assumed to originate from an infinitesimally thin 2D plane [12, 13]. The various different approximations above hold for a wide range of samples of interest and are therefore very useful in practice [14].

However, phase contrast imaging of many samples cannot be approximated by any of the above assumptions. Transmission electron microscopy (TEM) in particular often violates these assumptions, due to high scattering cross section of electrons with matter [15]. Instead, these scattering processes can typically only be modeled by a framework that includes multiple scattering [16]. The equations describing multiple scattering for a paraxial wave function can be approximately solved with the multislice algorithm [17], which has also been used as a model for inverse scattering in many experimental configurations in light, X-ray- and electron microscopy. While the inverse multislice model has been successfully applied to image thick, multiply scattering specimens in light microscopy [18–21], its use in X-ray [22–26] and electron microscopy [27–30] has been limited to proof-of-principle demonstrations with less than 10 slices or weakly scattering samples. This is mainly due to the fact that the optical systems in X-ray and electron microscopy have relatively small numerical apertures, such that the information recorded from a single view covers only a small fraction of reciprocal space [31–33]. This problem can be overcome either by enforcing strong prior knowledge about the underlying scattering potential in the form of sparsity constraints or the proper choice of slice separation [30], or by performing tomographic experiments [34–36].

Another framework that incorporates multiple scattering is the scattering matrix ($\mathcal{S}$-matrix) formalism [37, 38]. In electron microscopy, the $\mathcal{S}$-matrix formalism has been used to efficiently calculate diffraction results

with single crystals [38] and for scanning TEM (STEM) experiments [39, 40], and to retrieve projected potentials of strongly scattering samples in a two-step approach. First, the $\mathcal{S}$-matrix is retrieved from a series of intensity measurements. Second, the projected structure is retrieved. The proposed experimental methods for retrieval of the $\mathcal{S}$-matrix from intensity measurements range from measurements with different crystal thicknesses and sample tilts [41], different sample tilts alone [42–44], wavelength variation [45], large-angle rocking beam diffraction [46], and scanning diffraction with a convergent beam [47]. Only the last two of these approaches have been experimentally demonstrated [46, 48], and only on single-crystal structures.

In the visible light wavelengths, $\mathcal{S}$-matrix retrieval and subsequent singular value decomposition allows the identification of transmission eigenchannels [49] in strongly scattering materials and maximization of energy transport [50] through the system. Phase retrieval of the $\mathcal{S}$-matrix is performed by real-space phase- [51] or amplitude-modulation [52, 53], 4-phase interferometry [49], or full-field Mach-Zehnder interferometry [54] with input- and output channels in the plane-wave basis. The input and output channels of the $\mathcal{S}$-matrix are often represented in real-space, achieved by imaging the output plane with a CCD camera.

Our contribution in this work is three-fold: first, we develop the measurement operator to calculate scanning diffraction intensities of arbitrary samples from a given $\mathcal{S}$-matrix and derive its adjoint operator. Second, we formulate a phase retrieval algorithm that retrieves the $\mathcal{S}$-matrix of arbitrary samples from a series of scanning diffraction measurements with different modulations of the illumination aperture (e.g. a defocus series). Third, we formulate a relaxation of the phase retrieval algorithm for samples that do not require the full $\mathcal{S}$-matrix to be reconstructed.

II. RECONSTRUCTING THE $\mathcal{S}$-MATRIX

A. Theory of phase contrast imaging

Phase contrast microscopy with coherent light or matter waves defined by the wavefunction $|\psi\rangle_\mathbf{r}$ typically uses a series of interferometric measurements to invert a partial differential equation of the form

$$i\left[a\,\nabla_\perp^2 + b\,V(\mathbf{r}')\right]|\psi\rangle_{\mathbf{r}'} = \frac{\partial\,|\psi\rangle_{\mathbf{r}'}}{\partial z}, \quad (1)$$

where i is the imaginary constant, $\nabla_\perp^2$ is the two-dimensional Laplace operator, $V(\mathbf{r}')$ is the three-dimensional potential over the real space coordinates $\mathbf{r}' = (\mathbf{r}, z)$, and a and b are real-valued constant prefactors. The formal operator solution to this equation

for a wave function that has propagated a distance Δz through the potential is given by [55],

$$|\psi\rangle_{(\mathbf{r},z+\Delta z)} = \exp\left[ia\,\Delta z\nabla_\perp^2 + ib\,V_{\Delta z}(\mathbf{r},z)\right]|\psi\rangle_{\mathbf{r}'}. \quad (2)$$

In the scattering matrix formalism, the entire process of multiple scattering is modeled by multiplication with the complex-valued linear operator $\mathcal{S}$,

$$|\psi\rangle_{out} = \mathcal{S}\,|\psi\rangle_{in}. \quad (3)$$

The $\mathcal{S}$-matrix formalism has a wide range of applications in describing the interaction of coherent waves with multiply scattering objects [56].

All the previously discussed methods for $\mathcal{S}$-matrix-retrieval at high resolution have in common that they require a crystalline sample to solve for either the scattering matrix or the structure factors. The interferometric methods developed for light optics rely on the ability to precisely manipulate phases and/or amplitudes of the $\mathcal{S}$-matrix input channels and such precise control of the electron and X-ray optics is not yet feasible. In the following section, we describe our iterative reconstruction scheme from scanning diffraction measurements for $\mathcal{S}$-matrix-retrieval.

B. A real-space $\mathcal{S}$-matrix measurement model

Previous work for retrieving the $\mathcal{S}$-matrix from scanning diffraction measurements modeled the formation of the diffraction pattern intensity in the far-field of the sample, given a coherent probe $|\psi\rangle$ at position $\boldsymbol{\rho}$,

$$|\psi\rangle_{\mathbf{r}-\boldsymbol{\rho}} = \sum_{|\mathbf{h}|<h_{max}} \Psi(\mathbf{h})e^{2\pi i\mathbf{h}\cdot(\mathbf{r}-\boldsymbol{\rho})}, \quad (4)$$

with an intensity measurement given by [47]

$$I(\mathbf{q},\boldsymbol{\rho},\Psi) = \left|\sum_{|\mathbf{h}|<h_{max}} \mathcal{S}_{\mathbf{q},\mathbf{h}}\Psi(\mathbf{h})e^{-2\pi i\mathbf{h}\cdot\boldsymbol{\rho}}\right|^2. \quad (5)$$

In this work, we use the approximation that the wave function has a finite support after propagating through the specimen potential. To use this approximation as a constraint in an inversion algorithm, we need to represent the $\mathcal{S}$-matrix in real space:

$$I(\mathbf{q},\boldsymbol{\rho},\Psi) = \left|\mathcal{F}_\mathbf{r}\left[\sum_{|\mathbf{h}|<h_{max}} \mathcal{S}_{\mathbf{r},\mathbf{h}}\Psi(\mathbf{h})e^{-2\pi i\mathbf{h}\cdot\boldsymbol{\rho}}\right]\right|^2. \quad (6)$$

Here $\mathcal{S}_{\mathbf{r},\mathbf{h}}$ is the $\mathcal{S}$-matrix that maps Fourier-space input coefficents at wave-vectors $\mathbf{h}$ (we refer to these as the "beams" of the $\mathcal{S}$-matrix) to real-space output coefficients at positions $\mathbf{r}$.
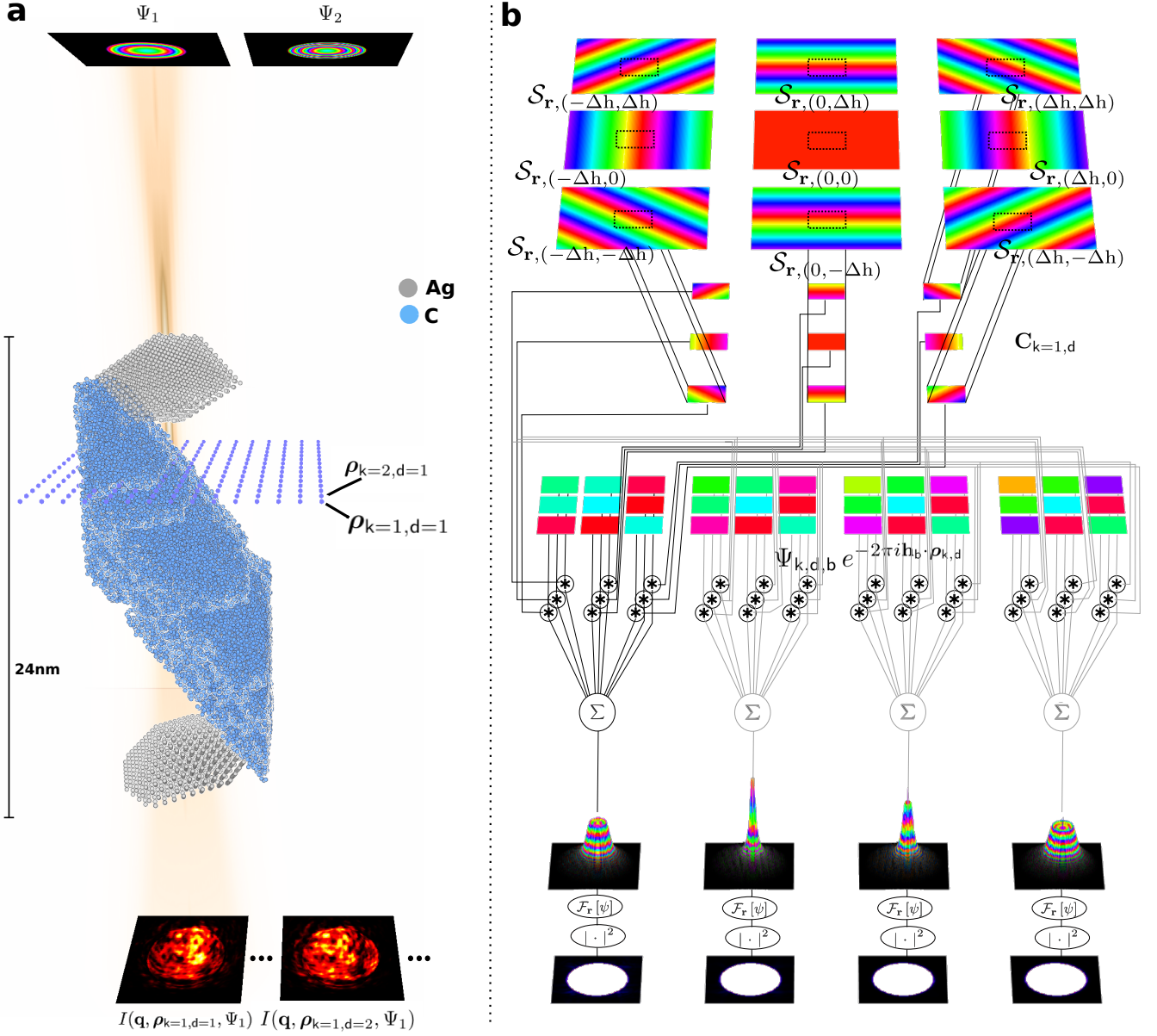
FIG. 1. Measurement scheme for $\mathcal{S}$-matrix inversion. (a) A scanning diffraction series of a strongly scattering sample at atomic resolution, where the phases Ψ_d of the probe-forming aperture are varied after each scan, here by changing the defocus. (b) Computational graph of the $\mathcal{S}$-matrix measurement operator for $D = 4$ different defocus aberrations. For each scan position, a patch with the size of the diffraction detector ($M_1 \times M_2$) is cropped out of each $\mathcal{S}$-matrix beam. Then, each cropped beam is multiplied by the corresponding complex phase factor (indicated by the * operation), depending on the phase and amplitude of beam the illumination aperture $\Psi_{d,b}$, and the scanning phase $e^{-2\pi i \mathbf{h}_b \cdot \boldsymbol{\rho}_{k,d}}$ of the current position. Subsequently, all phase-shifted beams are coherently summed (the Σ operator) to form an exit wave. Then the exit wave is propagated to the far-field ($\mathcal{F}_\mathbf{r}$ operation) and measured on the detector.

A previous experiment [48] used Eq. 5 and a series of defocus modulations to retrieve the phases of $\mathcal{S}_{\mathbf{q},\mathbf{h}}$ for a set of $\mathbf{h}$ vectors separately, and then used symmetry relations of $\mathcal{S}_{\mathbf{q},\mathbf{h}}$ to find the relative phases between the different $\mathcal{S}$-matrix columns. Whereas that approach is only valid for crystalline samples, we use only self-consistency in the measured data and retrieve all amplitudes and phases of $\mathcal{S}_{\mathbf{r},\mathbf{h}}$ simultaneously. We

also introduce a real-space compactness constraint on the scattered probes produced by the scattering matrix, equivalent to the method of Fourier-interpolating the $\mathcal{S}$-matrix [39]. We introduce the cropping operator,

$$\mathbf{C}_{\boldsymbol{\rho},\boldsymbol{\Delta}}(\mathbf{r}) = \begin{cases} 1 & \text{if } r_x - \rho_x \le \Delta_x/2 \text{ and } r_y - \rho_y \le \Delta_y/2 \\ 0 & \text{otherwise} \end{cases}$$

(7)

a two dimensional rectangular function of width $\boldsymbol{\Delta}$ centered about each probe scan position $\boldsymbol{\rho}$, which transforms Eq. 5 into

$$I(\mathbf{q}, \boldsymbol{\rho}, \Psi) = \left| \mathcal{F}_{\mathbf{r}} \left[\sum_{|\mathbf{h}| < h_{\max}} [\mathbf{C}_{\boldsymbol{\rho}, \boldsymbol{\Delta}}(\mathbf{r}) \mathcal{S}_{\mathbf{r}, \mathbf{h}}] \Psi(\mathbf{h}) e^{-2\pi i \mathbf{h} \cdot \boldsymbol{\rho}} \right] \right|^2. \tag{8}$$

The fact that the cropping operator acts on all $\mathcal{S}$-matrix beams equally leads to a self-consistent solution when measurements are taken with overlapping probe positions.

C. Phase retrieval of the $\mathcal{S}$-matrix

We now describe an algorithm to retrieve all amplitudes and phases of $\mathcal{S}_{\mathbf{r}, \mathbf{h}}$ simultaneously, given a set of phase modulations $\{\chi_\mathsf{d}(\mathbf{h})\}_{\mathsf{d}=1,\dots,\mathsf{D}}$ of the probe-forming aperture, using only self-consistency in the measured data. Let the detector be sampled with $\mathsf{M}_1 \times \mathsf{M}_2$ pixels. We perform a scan with K positions and D different probes and label a single position with k and a single defocus with d. Then the measured intensities have the dimension $\mathbf{I} \in \mathbb{R}^{\mathsf{K} \cdot \mathsf{D} \cdot \mathsf{M}_1 \cdot \mathsf{M}_2}$. For ease of notation, we enumerate all B samples in $|\mathbf{h}| < h_{max}$ with indices $\mathsf{b} = 1, \dots, \mathsf{B}$. The $\mathcal{S}$-matrix measurement operator maps the B beams of the $\mathcal{S}$-matrix of sampled on a discrete grid of $\mathsf{N}_1 \times \mathsf{N}_2$ pixels and the D probes to $\mathsf{K} \cdot \mathsf{D}$ diffraction patterns of size $\mathsf{M}_1 \times \mathsf{M}_2$. $\mathcal{A} : \mathbb{C}^{\mathsf{B} \times \mathsf{N}_1 \times \mathsf{N}_2} \times \mathbb{C}^{\mathsf{D} \times \mathsf{M}_1 \times \mathsf{M}_2} \to \mathbb{C}^{\mathsf{KDM}_1 \mathsf{M}_2}$. For better readability, we first define the measurement operator for position k and probe d: $\mathcal{A}_{\mathsf{k}, \mathsf{d}} : \mathbb{C}^{\mathsf{B} \times \mathsf{N}_1 \times \mathsf{N}_2} \times \mathbb{C}^{\mathsf{M}_1 \times \mathsf{M}_2} \to \mathbb{C}^{\mathsf{M}_1 \cdot \mathsf{M}_2}$:

$$\mathcal{A}_{\mathsf{k}, \mathsf{d}}(\mathcal{S}, \Psi_\mathsf{d}) := \left[\mathcal{F}_{\mathbf{r}} \left[\sum_{\mathsf{b}=1}^{\mathsf{B}} \Psi_{\mathsf{d}, \mathsf{b}} \, e^{-2\pi i \mathbf{h}_\mathsf{b} \cdot \boldsymbol{\rho}_{\mathsf{k}, \mathsf{d}}} \left[\mathbf{C}_{\mathsf{k}, \mathsf{d}} \mathcal{S} \right]_\mathsf{b} \right] \right]^V, \tag{9}$$

where $[\cdot]^V$ is a vectorization from 2D to 1D. We have also introduced the linear cropping operator $\mathbf{C}_{\mathsf{k}, \mathsf{d}} := \mathbf{C}_{\boldsymbol{\rho}_{\mathsf{k}, \mathsf{d}}} : \mathbb{C}^{\mathsf{B} \times \mathsf{N}_1 \times \mathsf{N}_2} \to \mathbb{C}^{\mathsf{B} \times \mathsf{M}_1 \times \mathsf{M}_2}$, which extracts a real-space patch of size $\mathsf{M}_1 \times \mathsf{M}_2$ from each beam of a given $\mathcal{S}$-matrix at the position with index k for the phase modulation d. The measurement operator for the full experiment is just the operators for each probe and position stacked on top of each other: $\mathcal{A}(\mathcal{S}, \Psi) = [\mathcal{A}_{1,1}(\mathcal{S}, \Psi_1), \mathcal{A}_{2,1}(\mathcal{S}, \Psi_1), \dots, \mathcal{A}_{\mathsf{K}, \mathsf{D}}(\mathcal{S}, \Psi_\mathsf{D})]^T$ We can then write the forward model for the measured intensities of a series of D scanning diffraction experiments taken with different probes as

$$\mathbf{y} = |\mathcal{A}(\mathcal{S}, \Psi)|^2. \tag{10}$$

Given this forward model and a set of intensity measurements $\mathbf{I}$ we can formulate the phase retrieval problem for blind $\mathcal{S}$-matrix inversion as

Find $\quad \mathcal{S} \in \mathbb{C}^{\mathsf{B} \times \mathsf{N}_1 \times \mathsf{N}_2} \quad$ and $\quad \Psi \in \mathbb{C}^{\mathsf{D} \times \mathsf{M}_1 \times \mathsf{M}_2}$
Subject to $\quad |\mathcal{A}(\mathcal{S}, \Psi)|^2 = \mathbf{I}.$

If the wave functions Ψ are known, the problem of finding $\mathcal{S}$ from a set of measurements $\mathbf{I}$ is a classical phase retrieval problem. There is a rich history of a algorithmic developments to solve the phase retrieval problem. Historically the first were algorithms based on alternating projections onto non-convex constraint sets [57–59]. Since these algorithms lack theoretical convergence guarantees, more recently convex relaxations were developed [60, 61] which provide a convergence guarantee, but use a prohibitive amount of memory. More recently, Bayesian accelerated gradient methods [62] and methods based on the alternating direction method of multipliers (ADMM) [63] have become popular. Since the wave functions Ψ_d are usually not known precisely in advance, the problem turns into multi-objective optimization. Additionally, in the presence of noise, it is beneficial to the reconstruction quality to include the noise model of the detector in the optimization. Since most advanced detectors in X-ray and electron microscopy are counting detectors, the noise statistics follow a Poisson distribution: $\mathbf{I} \sim \text{Poisson}(\mathbf{y})$. Here we choose an amplitude-based cost function as an approximation to the Poisson likelihood, due to its better convergence behaviour and divergence-free derivative [64, 65]:

$$\mathcal{D}(\mathbf{y}, \mathbf{I}) := \left\| \mathbf{y} - \sqrt{\mathbf{I}} \right\|_2, \tag{11}$$

where $\| \cdot \|_2$ is the l_2 norm and $\mathbf{y}$ are far-field amplitudes of the current model. We use the ADMM algorithm [66] to solve the joint optimization problem of $\mathcal{S}$ and Ψ. The augmented Langrangian of the $\mathcal{S}$-matrix retrieval problem is

$$\mathcal{L}_\beta(\mathcal{S}, \Psi, \mathbf{z}, \boldsymbol{\Lambda}) = \mathcal{D}(|\mathbf{z}|) + \text{Re}\{\boldsymbol{\Lambda}^\dagger (\mathcal{A}(\mathcal{S}, \Psi) - \mathbf{z})\} + \frac{\beta}{2} \|\mathcal{A}(\mathcal{S}, \Psi) - \mathbf{z}\|_2^2, \tag{12}$$

where we have introduced the auxiliary variables $\mathbf{z} \in \mathbb{C}^{\mathsf{K} \cdot \mathsf{D} \cdot \mathsf{M}_1 \cdot \mathsf{M}_2}$ and $\boldsymbol{\Lambda} \in \mathbb{C}^{\mathsf{K} \cdot \mathsf{D} \cdot \mathsf{M}_1 \cdot \mathsf{M}_2}$, which link the data-loss term with the model-loss term. We seek to solve for $\mathcal{S}$ and Ψ such that $\mathcal{L}(\mathcal{S}, \Psi, \mathbf{z}, \boldsymbol{\Lambda})$ is minimized:

$$(\mathcal{S}^*, \Psi^*, \mathbf{z}^*, \boldsymbol{\Lambda}^*) = \arg \max_{\boldsymbol{\Lambda}} \arg \min_{\mathcal{S}, \Psi, \mathbf{z}} \mathcal{L}(\mathcal{S}, \Psi, \mathbf{z}, \boldsymbol{\Lambda}) \tag{13}$$

ADMM decouples the joint problem into subproblems and solves them step by step:

1. $\Psi^{\mathsf{l}+1} = \arg \min_\Psi \mathcal{L}_\beta^\Psi := \arg \min_\Psi \mathcal{L}_\beta(\mathcal{S}^\mathsf{l}, \Psi, z^\mathsf{l}, \boldsymbol{\Lambda}^\mathsf{l})$

2. $\mathcal{S}^{\mathsf{l}+1} = \arg \min_\mathcal{S} \mathcal{L}_\beta^\mathcal{S} := \arg \min_\mathcal{S} \mathcal{L}_\beta(\mathcal{S}, \Psi^{\mathsf{l}+1}, z^\mathsf{l}, \boldsymbol{\Lambda}^\mathsf{l})$

3. $z^{\mathsf{l}+1} = \arg \min_z \mathcal{L}_\beta(\mathcal{S}^{\mathsf{l}+1}, \Psi^{\mathsf{l}+1}, z, \boldsymbol{\Lambda}^\mathsf{l})$

4. $\boldsymbol{\Lambda}^{\mathsf{l}+1} = \boldsymbol{\Lambda}^\mathsf{l} + \beta(z^{\mathsf{l}+1} - \mathcal{A}(\mathcal{S}^{\mathsf{l}+1}, \Psi^{\mathsf{l}+1}))$

D. Subproblems w.r.t. Ψ and $\mathcal{S}$

The subproblems with respect to Ψ and $\mathcal{S}$ both involve the adjoint of the measurement operator $\mathcal{A}$, which for a single measurement is given by $\mathcal{A}_{k,d}^{\mathcal{S}\,\dagger} : \mathbb{C}^{M_1 M_2} \to \mathbb{C}^{B \times N_1 \times N_2}$

$$\mathcal{A}_{k,d}^{\mathcal{S}_b\,\dagger}(\mathbf{z}) = \mathbf{C}_{k,d}^T \left[\Psi_{d,b}^* e^{2\pi i \mathbf{h}_b \cdot \boldsymbol{\rho}_{k,d}} \mathcal{F}_{\mathbf{q}}^\dagger [\mathbf{z}_{k,d}] \right] \qquad (14)$$

for a fixed Ψ, and $\mathcal{A}_{k,d}^{\Psi_{d,b}\,\dagger} : \mathbb{C}^{M_1 \cdot M_2} \to \mathbb{C}^{M_1 \times M_2}$

$$\mathcal{A}_{k,d}^{\Psi_{d,b}\,\dagger}(\mathbf{z}) = \frac{1}{M_1 M_2} \sum_{m_1}^{M_1} \sum_{m_2}^{M_2}$$
$$\left[\sum_{k=1}^{K} [\mathbf{C}_{k,d}\mathcal{S}]_b^* e^{2\pi i \mathbf{h}_b \cdot \boldsymbol{\rho}_{k,d}} \mathcal{F}_{\mathbf{q}}^\dagger [\mathbf{z}_{k,d}] \right]_{m_1, m_2} \qquad (15)$$

for a fixed $\mathcal{S}_b$. We solve the subproblems with respect to Ψ and $\mathcal{S}$ with gradient descent.

$$\Psi^{l+1} = \Psi^l + \gamma_1 \frac{\partial \mathcal{L}_\beta^\Psi}{\partial \Psi} \qquad (16)$$

$$\mathcal{S}^{l+1} = \mathcal{S}^l + \gamma_2 \frac{\partial \mathcal{L}_\beta^{\mathcal{S}}}{\partial \mathcal{S}}, \qquad (17)$$

where $\gamma_1, \gamma_2 \in \mathbb{R}$ are gradient descent step sizes. We found that one gradient step per iteration is usually enough for fast convergence. The gradient is given by

$$\frac{\partial \mathcal{L}_\beta^\Psi}{\partial \Psi_{d,b}} = \beta \, \mathcal{A}_{k,d}^{\Psi_{d,b}\,\dagger}(\mathbf{z}^l - \mathcal{A}_{k,d}(\mathcal{S}, \Psi_d) - \frac{\Lambda^l}{\beta}) \qquad (18)$$

$$\frac{\partial \mathcal{L}_\beta^{\mathcal{S}}}{\partial \mathcal{S}_b} = \beta \sum_{k=1}^{K} \sum_{d=1}^{D} \mathcal{A}_{k,d}^{\mathcal{S}_b\,\dagger}(\mathbf{z}^l - \mathcal{A}_{k,d}(\mathcal{S}, \Psi_d) - \frac{\Lambda^l}{\beta}). \quad (19)$$

See the Appendix B for a detailed derivation.

E. Subproblem w.r.t. z

The subproblem w.r.t. $\mathbf{z}$ was solved elsewhere [67]. The solution is

$$\mathbf{z}^{l+1} = \frac{\text{sgn}(\hat{\mathbf{z}}) \left[\sqrt{\mathbf{I}} + \beta |\hat{\mathbf{z}}| \right]}{(1+\beta)}. \qquad (20)$$

The full ADMM algorithm is then given as:

Algorithm 1 Joint $\mathcal{S}$-matrix and probe retrieval via ADMM

Input:
measured intensities $\mathbf{I} \in \mathbb{R}^{K \times D \times M_1 \times M_2}$
scan positions $\boldsymbol{\rho} \in \mathbb{R}^{K \times D \times 2}$
initial Fourier space probe phases $\boldsymbol{\chi}^0 \in \mathbb{C}^{D \times B}$
step sizes $\gamma_1, \gamma_2, \beta \in \mathbb{R}$
Initialize:
set $(N_1, N_2) = \lceil \frac{\max(\mathbf{r}_s) + M}{M} \rceil \cdot M$ such that the plane waves $e^{i\mathbf{h} \cdot \mathbf{r}}$ have periodic boundary conditions
calculate $\mathbf{I}^{mean} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{I}_k$ and
$a_{max} = \max\{||\mathbf{I}_k||_1 \forall k = \{1, ..., K\}\}$
$\Psi^0 \leftarrow \frac{a_{max}}{\sqrt{||\mathbf{I}^{mean}||_1}} \mathbf{I}^{mean} e^{i\boldsymbol{\chi}^0}$
$\mathcal{S}_b^0 \leftarrow e^{i\mathbf{h}_b \cdot \mathbf{r}}, \mathcal{S} \in \mathbb{C}^{B \times N_1 \times N_2}$
$\Lambda = 0, \mathbf{z} = 0$

Run:
1: **for** $l = 0$ to L **do**
2: $\quad \hat{\mathbf{z}} = \mathbf{z}^l + \frac{\Lambda^l}{\beta}$
3: $\quad \Psi^{l+1} \leftarrow \Psi^l + \gamma_1 \cdot \frac{\partial \mathcal{L}_\beta^\Psi}{\partial \Psi} \left(\mathcal{S}^l, \Psi^l, \hat{\mathbf{z}} \right)$
4: $\quad \mathcal{S}^{l+1} \leftarrow \mathcal{S}^l + \gamma_2 \cdot \frac{\partial \mathcal{L}_\beta^{\mathcal{S}}}{\partial \mathcal{S}} \left(\mathcal{S}^l, \Psi^{l+1}, \hat{\mathbf{z}} \right)$
5: $\quad \hat{\mathbf{z}} = \mathcal{A}(\mathcal{S}^{l+1}, \Psi^{l+1}) - \frac{\Lambda^l}{\beta}$
6: $\quad \mathbf{z}^{l+1} \leftarrow \frac{\text{sgn}(\hat{\mathbf{z}}) \left[\sqrt{\mathbf{I}} + \beta |\hat{\mathbf{z}}| \right]}{(1+\beta)}$
7: $\quad \Lambda^{l+1} \leftarrow \Lambda^l + \beta(\mathbf{z}^{l+1} - \mathcal{A}(\mathcal{S}^{l+1}, \Psi^{l+1}))$
8: **end for**
Output: $\mathcal{S}^* = \mathcal{S}^L$

III. SIMULATED $\mathcal{S}$-MATRIX PHASE RETRIEVAL

In this section, we use forward simulations to validate our $\mathcal{S}$-matrix phase retrieval algorithm. We also examine the algorithm dependence on the sampling density and calibration.

A. Sampling and calibration dependence

To demonstrate that our algorithm can reconstruct $\mathcal{S}$-matrices of realistic samples, we simulate a 4D-STEM focal series of the sample shown in Fig. 1 a), as it may appear in a tomography experiment. The sample contains two decahedral Ag nanoparticles of 3.3 nm diameter, placed on the top and bottom sides of an amorphous carbon substrate, tilted by 67°, giving it an axial extent of 24 nm. The probe convergence angle is chosen as 26 mrad and the electron energy as 300 kV, resulting in a depth of focus (DOF) of 5.8 nm and a sample depth of $4.1\times$ DOF. The detector was set to record diffraction signal up to 40 mrad, resulting in a sampling grid with steps of 25 pm. The field of view was scanned with 129×129 positions on a 2D grid with the half-period resolution. The reconstruction shown in Fig. 2 used 6 defoci with a step of 4.6 nm, with the first
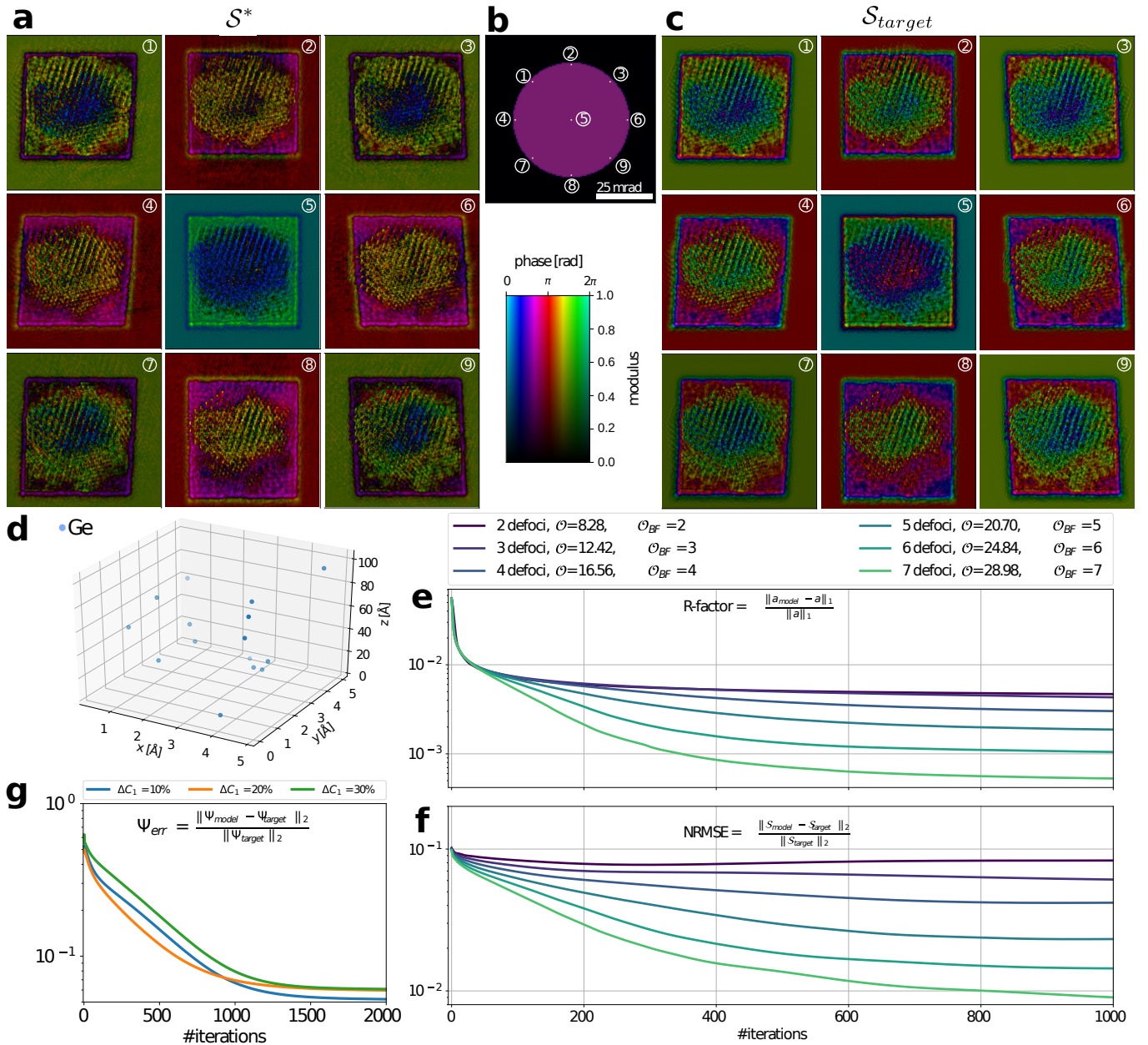
FIG. 2. (a) Simulated experiment with (b) the probe-forming aperture used for simulating the experiment shown in Fig. 1 (a). Selected beams numbered in (b) are shown from the reconstructed $\mathcal{S}$-matrix in (a) and the ground-truth $\mathcal{S}$-matrix in (c). The skew effect of the exit waves in different beams comes from the three-dimensional structure, and is a parallax effect of the different propagation directions of the beams. (d) Test sample of randomly distributed Germanium atoms. (e) R-factor vs number of iterations for different numbers of defoci and oversampling rates used in the simulations. (f) Normalized root mean square error of the model $\mathcal{S}$-matrix vs number of iterations for different numbers of defoci and oversampling rates. h) Mean probe error vs number of iterations for defocus miscalibration levels of 10 %, 20 % and 30 % of the defocus step and random higher order aberrations.

defocus at the top of the sample. The detector size was set to 128×128 pixels, yielding an angular resolution of $0.31\,\mu\text{rad}$ and $\mathcal{S}$-matrix dimensions of $\mathcal{S} \in \mathbb{C}^{5973 \times 256 \times 256}$.

We ran Algorithm 1 for 500 iterations, utilizing 48 NVIDIA V-100 GPUs. After 200 minutes, the reconstruction converged to an normalized root mean square error (NRMSE) of 4 % and an R-factor of 0.1 %. Nine selected $\mathcal{S}$-matrix beams from the reconstruction are shown in Fig. 2 a, and the ground-truth $\mathcal{S}$-matrix is shown in 2c. To investigate the convergence properties under varying number of measurements and calibration errors, we used a smaller test sample, consisting of 16 randomly distributed Germanium atoms in a volume of

7

5 Å × 5 Å × 100 Å, shown in Fig. 2d. The convergence angle for the following tests was chosen as 30 mrad, with a detector spanning 60 mrad, and the diffraction patterns were sampled on a 20 × 20 pixel detector, yielding $\mathcal{S}$-matrix dimensions of $\mathcal{S} \in \mathbb{C}^{177\times60\times60}$, and the defocus step was chosen as 2 nm.

For the following investigations we fix the scan step to Nyquist sampling. First we investigate the converge behaviour with respect to the number of measured defoci. Fig. 2e and f show the R-factor, and the NRMSE as a function of iterations and number of defoci measured. We define the oversampling factor as

$$\mathcal{O} = \frac{\#\text{ nonzero measurements}}{\#\text{ variables in }\mathcal{S}\text{-matrix}}, \qquad (21)$$

and the bright-field oversampling factor as

$$\mathcal{O}_{BF} = \frac{\#\text{ nonzero measurements in bright-field}}{\#\text{ variables in }\mathcal{S}\text{-matrix}}. \quad (22)$$

One can see that for 2 defocus measurements, the NRMSE diverges slowly, and for 3 measurements the NRMSE does not converge monotonously with the R-factor. While the oversampling factor $\mathcal{O}$ lies above the number 4 typically needed for successful phase retrieval, the number of phase modulations that each beam receives, $\mathcal{O}_{BF}$, is below the threshold. For this case, a more heterogeneous sample than the crystalline objects considered in previous work, the reconstruction does not stably converge in these cases. This could be due to the small defocus steps used and will be investigated in the future.

We also investigate the dependence of the probe refinement on the level of defocus miscalibration and residual uncorrected probe aberrations. Fig. 2 h) shows the mean errors of 30 reconstructions performed with defocus errors ΔC_1 drawn form a normal distribution with a standard deviation of 10 %, 20 % and 30 % of the defocus step, axial coma with a standard deviation of 100 nm, three-fold astigmatism with a standard deviation of 20 nm, spherical aberration with a standard deviation of 4 μm, and star aberration with a standard deviation of 4 μm. Although convergence takes roughly twice as many iterations $\mathcal{S}$-matrix-reconstruction with miscalibrated aberrations, for all miscalibration values a probe reconstruction error of less than 10 % was achieved.

B. Reconstructing the projected $\mathcal{S}$-matrix

Consider the scattering matrix for a phase object, which is a valid approximation for a thin and weakly scattering sample [14], with specimen potential $V(\mathbf{r})$. The analytic expression for each component will be,

$$\mathcal{S}_{\mathbf{r},\mathbf{h}} = e^{i\sigma V(\mathbf{r})-2\pi i\mathbf{h}\cdot\mathbf{r}}. \qquad (23)$$

So every $\mathcal{S}$-matrix component will be the same except for the multiplicative phase ramp of $e^{-2\pi i\mathbf{h}\cdot\mathbf{r}}$. As we consider thicker, more strongly scattering objects we would expect each component of the $\mathcal{S}$-matrix to be increasingly different and we consider the similarity or lack thereof of each of the $\mathcal{S}$-matrix components to be an indication of the degree of strong multiple scattering of a sample. When reconstructing the $\mathcal{S}$-matrix from a 4D-STEM dataset the automatic choice for choosing the sampling of beams, the set of $\mathbf{h}$ vectors, is to match it to the number of pixels within the bright-field disk or aperture function of the STEM probe in the diffraction patterns. For a fine diffraction space sampling of an object that does not exhibit much multiple scattering this sampling of beams might be highly redundant and we might improve our reconstruction by forcing a more sparse sampling of beams and increasing the ratio of experimental measurements to unknown parameters in our reconstruction. On the other hand, very thick and strongly scattering samples might require very high sampling of the diffraction patterns for an accurate reconstruction of the $\mathcal{S}$-matrix. While the latter case can only be solved with better sampling in the diffraction plane, for the former case in this section we outline a strategy for choosing a sparser sampling of the input beams $\mathbf{h}$ that involves partitioning of the bright-field disk into separate "tiles".

Shown in Fig. 3 are the complex values of a subset of $\mathcal{S}$-matrix components for a) 7.3 Å, b) 36.5 Å and c) 109.5 Å thicknesses of an ScAlO$_3$ crystal. All beams have been multiplied by the conjugate of the phase ramp that appears in Eq. 23, $e^{2\pi i\mathbf{h}\cdot r}$. For Fig. 3(a) the difference between beams is minimal so a phase object approximation would be appropriate for this thickness. For Fig. 3 b)-c) we see increasing variation between beams as the object becomes thicker. A partitioning system aims to group these $\mathcal{S}$-matrix components by similarity and visual comparison of the $\mathcal{S}$-matrix montages with the Fresnel propagator, $\mathcal{P}(h) = \exp(-i\lambda\pi h^2 t)$, for free-space of equivalent thickness of the crystal (shown to the right of each subfigure) suggests that a criterion based on phase variation of a Fresnel free-space propagator might be an effective way of doing this. We partition the bright-field disk into annular regions $\sqrt{(i-1)\Delta\phi/\lambda\pi t} < h_i < \sqrt{i\Delta\phi/\lambda\pi t}$ where $i \in \mathbb{N}$ and $\Delta\phi$, the Fresnel propagator phase variance, is a predetermined criterion (we use $\Delta\phi = \pi/4$ in this work). These regions are further divided azimuthally, with an arclength equal to the radius of the inner-most partition, $\sqrt{\Delta\phi/\lambda\pi t}$. We represent these partitions with the map $\tau : \{0,...,\mathsf{B}\} \to \{0,...,\mathsf{B}_{\mathsf{tile}}\}$ from beams to beam tiles. Partitioning according to his criterion is shown to the right where each different color indicates a separate partition of the bright-field disk for each of the thicknesses in Fig. 3(a)-(c). We note finally that this is an approximate criterion only since thickness of an uncharacterised object can only be guessed at based on the intuition of the microscopist and the Fresnel criterion

does not take into account the scattering strength per unit volume of the object. For example we might expect samples containing a high density of heavy (large Z) elements to exhibit greater beam to beam variation of the $\mathcal{S}$-matrix components than materials with only small Z numbers.

Partitioning reduces the number of needed measurements by a factor B_{tile}/B. The reconstructed $\mathcal{S}$-matrix then has the reduced dimension $\mathcal{S} \in \mathbb{C}^{B_{\text{tile}} \times N_1 \times N_2}$, and the beam-dependent beam-tilt $e^{-2\pi i \mathbf{h} \cdot \boldsymbol{r}}$ has to be separated from the $\mathcal{S}$-matrix variable to allow the beam-averaging. We define therefore the un-tilted $\mathcal{S}$-matrix $\mathcal{S}^t_{\mathbf{r},\mathbf{h}} := \mathcal{S}_{\mathbf{r},\mathbf{h}} e^{-2\pi i \mathbf{h} \cdot \boldsymbol{r}}$, which is the new latent-variable in the projected $\mathcal{S}$-matrix problem. The forward operator becomes

$$\mathcal{A}^t_{\mathsf{k},\mathsf{d}}(\mathcal{S}^t, \Psi_{\mathsf{d}}) :=$$

$$\left[\mathcal{F}_{\mathbf{r}} \left[\sum_{\mathsf{b}=1}^{B} \Psi_{\mathsf{k},\mathsf{d},\mathsf{b}} \, e^{-2\pi i \mathbf{h}_\mathsf{b} \cdot \boldsymbol{\rho}_{\mathsf{k},\mathsf{d}}} \left[\mathbf{C}_{\mathsf{k},\mathsf{d}} \mathcal{S}^t \right]_{\tau(\mathsf{b})} e^{2\pi i \mathbf{h} \cdot \boldsymbol{r}} \right] \right]^V,$$

and the gradient with respect to $\mathcal{S}^t$ is

$$\frac{\partial \mathcal{L}^{\mathcal{S}^t}_\beta}{\partial \mathcal{S}^t_\mathsf{b}} = \frac{\beta}{|Q_\mathsf{b}|} \sum_{\mathsf{q} \in Q_\mathsf{b}} \sum_{\mathsf{k}=1}^{K} \sum_{\mathsf{d}=1}^{D}$$

$$\mathcal{A}^{\mathcal{S}_\mathsf{b} \, \dagger}_{\mathsf{k},\mathsf{d}} (\mathbf{z}^{\mathsf{l}} - \mathcal{A}^t_{\mathsf{k},\mathsf{d}}(\mathcal{S}^t, \Psi_\mathsf{d}) - \frac{\boldsymbol{\Lambda}^{\mathsf{l}}}{\beta}) \; (24)$$

where we have introduced the set $Q_\mathsf{b} = \{n \,|\, \tau(n) = b \,\forall\, n \in \{1, ..., B\}\}$ of beams that belong to tile b and $|Q_\mathsf{b}|$ is the cardinality of Q_b.

IV. CONCLUSION AND OUTLOOK

We have introduced a new method for $\mathcal{S}$-matrix retrieval, that converges without any regularization for samples which span 4 depths of focus and more, and numerical apertures which are experimentally accessible, and can recover aberration miscalibrations of up to 30 %. We have also introduced a simplified model, projected $\mathcal{S}$-matrix inversion, for the case when the sample is thin enough that not every beam that is measured on the detector has to be included in the model. In future work, we will compare projected $\mathcal{S}$-matrix inversion to mixed-state ptychography and multi-slice ptychography, since both offer alternative methods for moving beyond the simple model of single-mode ptychography.

The $\mathcal{S}$-matrix-retrieval methods developed here could be used for a number of advancements in imaging through and with strongly scattering materials in X-ray and electron microscopy. In combination with adaptive electron optics [68], selective focusing through crystalline materials may become possible in a similar vein to light
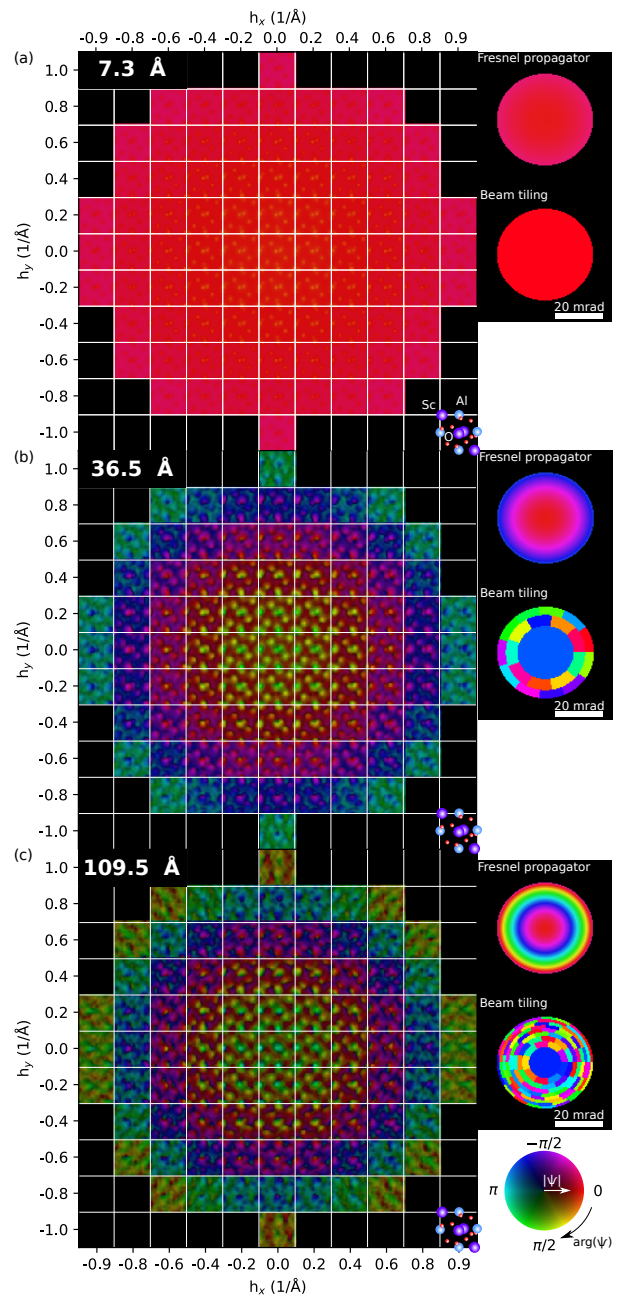

FIG. 3. Partitioning of the $\mathcal{S}$-matrix beams into separate tiles according to the expected degree of departure from the phase object approximation. In each subplot a subset of the complex components of the scattering matrix (color hue is phase and color saturation is amplitude according to inset colorwheel) are shown for a $ScAlO_3$ crystal with (a) 7.3 Å, (b) 36.5 Å and (c) 109.5 Å thickness. With increasing thickness there is increasing variability between the different components. The beam partitioning suggested by a phase variance of the Fresnel freespace propagator of $\pi/4$ described in the text is inset on the top right for each of the thicknesses in (a)-(c) where each colour in the bright-field disk corresponds to a seperate B_{tile}

optical experiments [69].

The retrieved $\mathcal{S}$-matrix can be used for depth-sectioning which is robust against multiple scattering. $\mathcal{S}$-matrix-retrieval may also form the basis of inverse multi-slice algorithms for phase-contrast tomography in scanning diffraction microscopes. The angular decomposition in the $\mathcal{S}$-matrix may be useful for ab-initio angular and transverse alignment of different tilt angles for phase-contrast tomography. This approach may be experimentally more feasible than end-to-end tomographic reconstruction algorithms. Finally, one could think about characterizing amorphous materials from their $\mathcal{S}$-matrix.

To allow optimal image quality, future refinements of the algorithm could include experimental uncertainties like position errors, and modeling of nuisance parameters like spatial and temporal incoherence, similar to their treatment in ptychographic reconstruction algorithms [70–74].

ACKNOWLEDGMENTS

Appendix A: Complexity analysis

Both the forward calculation and the backward calculation have the following computational complexity: per diffraction pattern the forward pass has a complexity of $O(\mathsf{M_1}^2\mathsf{M_2}^2\mathsf{B}\log(\mathsf{M_1M_2}))$ where the factor $O(\mathsf{M_1M_2}\log(\mathsf{M_1M_2}))$ comes from the fast Fourier transform operation. The forward and backward calculation on the full dataset then have a complexity of $O(\mathsf{KDM_1}^2\mathsf{M_2}^2\mathsf{B}\log(\mathsf{M_1M_2}))$. Since the number of beams scales quadratically with the size of the detector, the overall complexity scales with $O(\mathsf{KDM}^4\log(\mathsf{M}^2))$ for a square detector of size M. While this might seem intractable for currently available large detectors, it is offset by the fact that KDM^2 of these computations are embarrassingly parallel batched complex matrix multiplications and can be carried out very efficiently on commonly available hardware accelerators.

[1] M. Pluta. *Advanced light microscopy*, volume 1. Elsevier Amsterdam, 1988.

[2] A. Clarke and C. N. Eberhardt. *Microscopy techniques for materials science*. Woodhead Publishing, 2002.

[3] J. Kirz, C. Jacobsen, and M. Howells. Soft X-ray microscopes and their biological applications. *Quarterly reviews of biophysics*, 28(1):33–130, 1995.

[4] S. C. Mayo, A. W. Stevenson, and S. W. Wilkins. In-line phase-contrast X-ray imaging and tomography for materials science. *Materials*, 5(5):937–965, 2012.

[5] J.C.H. Spence. The future of atomic resolution electron microscopy for materials science. *Materials Science and Engineering: R: Reports*, 26(1-2):1–49, 1999.

[6] R. M. Glaeser. Invited review article: Methods for imaging weak-phase objects in electron microscopy. *Review of Scientific Instruments*, 84(11):312, 2013.

[7] F Zernike. Phase contrast. *Z. Tech. Physik.*, 16:454, 1935.

[8] L. E. Ballentine. The statistical interpretation of quantum mechanics. *Reviews of Modern Physics*, 42(4):358, 1970.

[9] R Barer. A vector theory of phase contrast and interference contrast. i. positive phase contrast. *Journal of the Royal Microscopical Society*, 72(1):10–30, 1952.

[10] S. Maretzke. A uniqueness result for propagation-based phase contrast imaging from a single measurement. *Inverse Problems*, 31(6):065003, 2015.

[11] M. Born and E. Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light.* Elsevier, 2013.

[12] H. A. Cohen, M. F. Schmid, and W. Chiu. Estimates of validity of projection approximation for three-dimensional reconstructions at high resolution. *Ultramicroscopy*, 14(3):219–226, 1984.

[13] A. Burvall, U. Lundström, P.A.C. Takman, D. H. Larsson, and H. M. Hertz. Phase retrieval in X-ray phase-contrast imaging suitable for tomography. *Optics express*, 19(11):10359–10376, 2011.

[14] M. Vulović, L. M. Voortman, L. J. van Vliet, and B. Rieger. When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM. *Ultramicroscopy*, 136:61–66, 2014.

[15] A. V. Crewe and T. Groves. Thick specimens in the cem and stem. i. contrast. *Journal of Applied Physics*, 45(8):3662–3672, 1974.

[16] H. Bethe. The theory of electron diffraction on crystals. *Annals of Physics*, 392(17):55–129, 1928.

[17] J. M. Cowley and A. F. Moodie. The scattering of electrons by atoms and crystals. I. a new theoretical approach. *Acta Crystallographica*, 10(10):609–619, 1957.

[18] S. Chowdhury, M. Chen, R. Eckert, D. Ren, F. Wu, N. Repina, and L. Waller. High-resolution 3d refractive index microscopy of multiple-scattering samples from

intensity images. *Optica*, 6(9):12111219, Sep 2019. ISSN 2334-2536. doi:10/ggzqdc.

[19] T. M. Godden, R. Suman, M. J. Humphry, J. M. Rodenburg, and A. M. Maiden. Ptychographic microscope for three-dimensional imaging. *Optics Express*, 22(10):1251312523, May 2014. ISSN 1094-4087. doi:10/ggzqfn.

[20] U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, and D. Psaltis. Learning approach to optical tomography. *Optica*, 2(6):517–522, 2015.

[21] Peng Li and Andrew Maiden. Multi-slice ptychographic tomography. *Scientific Reports*, 8(1):2049, Feb 2018. ISSN 2045-2322. doi:10/gfz5vw.

[22] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg. Ptychographic transmission microscopy in three dimensions using a multi-slice approach. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.*, 29(8):160614, Aug 2012. ISSN 1520-8532.

[23] A. Suzuki, S. Furutaku, K. Shimomura, K. Yamauchi, Y. Kohmura, T. Ishikawa, and Y. Takahashi. High-resolution multislice x-ray ptychography of extended thick objects. *Physical review letters*, 112(5):053903, 2014.

[24] K. Shimomura, A. Suzuki, M. Hirose, and Y. Takahashi. Precession x-ray ptychography with multislice approach. *Physical Review B*, 91(21):214114, Jun 2015. doi:10/gfz5vs.

[25] E. H. R. Tsai, I. Usov, A. Diaz, A. Menzel, and M. Guizar-Sicairos. X-ray ptychography with extended depth of field. *Optics Express*, 24(25):2908929108, Dec 2016. ISSN 1094-4087. doi:10/gc5qcg.

[26] H. Öztürk, H. Yan, Y. He, M. Ge, Z. Dong, M. Lin, E. Nazaretski, I. K. Robinson, Y. S. Chu, and X. Huang. Multi-slice ptychography with large numerical aperture multilayer laue lenses. *Optica*, 5(5):601607, May 2018. ISSN 2334-2536. doi:10/gf2twh.

[27] W. Van den Broek and C. T Koch. Method for retrieval of the three-dimensional object potential by inversion of dynamical electron scattering. *Physical review letters*, 109(24):245502, 2012.

[28] W. Van den Broek and C. T. Koch. General framework for quantitative three-dimensional reconstruction from arbitrary detection geometries in tem. *Phys. Rev. B*, 87(18):184108, May 2013. doi:10.1103/PhysRevB.87.184108.

[29] Si Gao, Peng Wang, Fucai Zhang, Gerardo T. Martinez, Peter D. Nellist, Xiaoqing Pan, and Angus I. Kirkland. Electron ptychographic microscopy for three-dimensional imaging. *Nature Communications*, 8, 2017. doi:10.1038/s41467-017-00150-1.

[30] M. Schloz, T. C. Pekin, Z. Chen, W. Van den Broek, D. A. Muller, and C. T. Koch. Overcoming information reduced data and experimentally uncertain parameters in ptychography with regularized optimization. *arXiv:2005.01530 [eess]*, May 2020. URL http://arxiv.org/abs/2005.01530.

[31] Esther H. R. Tsai, Federica Marone, and Manuel Guizar-Sicairos. Gridrec-ms: an algorithm for multi-slice tomography. *Optics Letters*, 44(9):21812184, May 2019. ISSN 1539-4794. doi:10/gfz5vr.

[32] C. Jacobsen. Relaxation of the crowther criterion in multislice tomography. *Optics Letters*, 43(19):48114814, Oct 2018. doi:10/gfz5vx.

[33] H. L. Xin and D. A. Muller. Aberration-corrected adf-stem depth sectioning and prospects for reliable 3d imaging in s/tem. *Journal of Electron Microscopy*, 58(3): 157165, Jun 2009. ISSN 0022-0744. doi:10/fndbg5.

[34] M. A. Gilles, Y. S. G. Nashed, M. Du, C. Jacobsen, and S. M. Wild. 3d x-ray imaging of continuous objects beyond the depth of focus limit. *Optica*, 5(9):10781086, Sep 2018. ISSN 2334-2536. doi:10/gfmhzc.

[35] M. Du, Y. S. G. Nashed, S. Kandel, D. Grsoy, and C. Jacobsen. Three dimensions, two microscopes, one code: Automatic differentiation for x-ray nanotomography beyond the depth of focus limit. *Science Advances*, 6(13):eaay3700, Mar 2020. doi: 10.1126/sciadv.aay3700.

[36] D. Ren, C. Ophus, M. Chen, and L. Waller. A multiple scattering algorithm for three dimensional phase contrast atomic electron tomography. *Ultramicroscopy*, 208:112860, 2020.

[37] F. Fujimoto. Dynamical theory of electron diffraction in Laue-case, I. general theory. *Journal of the Physical Society of Japan*, 14(11):1558–1568, 1959.

[38] L. Sturkey. The calculation of electron diffraction intensities. *Proceedings of the Physical Society*, 80(2): 321, 1962.

[39] C. Ophus. A fast image simulation algorithm for scanning transmission electron microscopy. *Advanced structural and chemical imaging*, 3(1):13, 2017.

[40] H. G. Brown, J. Ciston, and C. Ophus. Linear-scaling algorithm for rapid computation of inelastic transitions in the presence of multiple electron scattering. *Physical Review Research*, 1(3):033186, 2019.

[41] J.C.H. Spence. Direct inversion of dynamical electron diffraction patterns to structure factors. *Acta Crystallographica Section A: Foundations of Crystallography*, 54(1):7–18, 1998.

[42] L. J. Allen, T. W. Josefsson, and H. Leeb. Obtaining the crystal potential by inversion from electron scattering intensities. *Acta Crystallographica Section A: Foundations of Crystallography*, 54(44):388398, Jul 1998. ISSN 0108-7673. doi:10/bg6bgk.

[43] L. J. Allen, H. M. L. Faulkner, and H. Leeb. Inversion of dynamical electron diffraction data including absorption. *Acta Crystallographica Section A: Foundations of Crystallography*, 56(2):119–126, 2000.

[44] J. J. Donatelli and J. C. H. Spence. Inversion of many-beam bragg intensities for phasing by iterated projections: Removal of multiple scattering artifacts from diffraction data. *Physical Review Letters*, 125(6): 065502, Aug 2020. doi:10.1103/PhysRevLett.125.065502.

[45] P. Rez. Schemes to determine the crystal potential under dynamical conditions using voltage variation. *Acta Crystallographica Section A: Foundations of Crystallography*, 55(2):160–167, 1999.

[46] F. Wang, R. S. Pennington, and C. T. Koch. *Physical Review Letters*, (1):015501, Jun 2016.

[47] S. D. Findlay. Quantitative structure retrieval using scanning transmission electron microscopy. *Acta Crystallographica Section A: Foundations of Crystallography*, 61(4):397–404, 2005.

[48] H. G. Brown, Z. Chen, M. Weyland, C. Ophus, J. Ciston, L. J. Allen, and S. D. Findlay. Structure retrieval at atomic resolution in the presence of multiple scattering of the electron probe. *Physical Review Letters*, 121(26):

266102, 2018.

[49] S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan. Measuring the transmission matrix in optics: An approach to the study and control of light propagation in disordered media. *Physical Review Letters*, 104(10):100601, Mar 2010. ISSN 0031-9007, 1079-7114. doi:10/b6hg8x.

[50] Moonseok Kim, Youngwoon Choi, Changhyeong Yoon, Wonjun Choi, Jaisoon Kim, Q.-Han Park, and Wonshik Choi. Maximal energy transport through disordered media with the implementation of transmission eigenchannels. *Nature Photonics*, 6(9):581585, Sep 2012. ISSN 1749-4893. doi:10/gbbsw9.

[51] C. A. Metzler, M. K. Sharma, S. Nagesh, R. G. Baraniuk, O. Cossairt, and A. Veeraraghavan. Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset. In *2017 IEEE International Conference on Computational Photography (ICCP)*, page 116, May 2017. doi:10/ggh89w.

[52] A. Drémeau, A. Liutkus, D. Martina, O. Katz, C. Schülke, F. Krzakala, S. Gigan, and L. Daudet. Reference-less measurement of the transmission matrix of a highly scattering material using a DMD and phase retrieval techniques. *Optics Express*, 23(9):1189811911, May 2015. ISSN 1094-4087. doi:10/ggm6g2.

[53] B. Rajaei, E. W. Tramel, S. Gigan, F. Krzakala, and L. Daudet. Intensity-only optical compressive imaging using a multiply scattering material and a double phase retrieval approach. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 40544058, Mar 2016. doi:10/ggm6g3.

[54] H. Yu, T. R. Hillman, W. Choi, J. O. Lee, M. S. Feld, R. R. Dasari, and Y. Park. Measuring large optical transmission matrices of disordered media. *Physical Review Letters*, 111(15):153902, Oct 2013. doi:10/f5mq6f.

[55] E. J. Kirkland. *Advanced Computing in Electron Microscopy, 3rd Edition*. Springer, 2020.

[56] S. Rotter and S. Gigan. Light fields in complex media: Mesoscopic scattering meets wave control. *Reviews of Modern Physics*, 89(1):015005, Mar 2017. doi:10/f9t5fk.

[57] J. Miao, D. Sayre, and H. N. Chapman. Phase retrieval from the magnitude of the fourier transforms of nonperiodic objects. *JOSA A*, 15(6):16621669, Jun 1998. ISSN 1520-8532. doi:10/fw83fw.

[58] J R Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21(15):275869, Aug 1982. ISSN 0003-6935.

[59] Y. Shechtman, Y.C. Eldar, O. Cohen, H.N. Chapman, Jianwei Miao, and M. Segev. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87109, May 2015. ISSN 1053-5888. doi:10.1109/MSP.2014.2352673.

[60] E. J. Candes, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[61] I. Waldspurger, A. dAspremont, and S. Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1):4781, Feb 2015. ISSN 1436-4646. doi:10.1007/s10107-013-0738-9.

[62] E. Bostan, M. Soltanolkotabi, D. Ren, and L. Waller. Accelerated wirtinger flow for multiplexed fourier ptychographic microscopy. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3823–3827, 2018.

[63] V. Nikitin, S. Aslan, Y. Yao, T. Biçer, S. Leyffer, R. Mokso, and D. Gürsoy. Photon-limited ptychography of 3d objects via bayesian reconstruction. *OSA Continuum*, 2(10):2948–2968, 2019.

[64] L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller. Experimental robustness of fourier ptychography phase retrieval algorithms. *Optics Express*, 23(26):33214, Dec 2015. ISSN 1094-4087. doi:10/gc5pwb.

[65] A. Fannjiang and T. Strohmer. The numerics of phase retrieval. *arXiv:2004.05788 [cs, eess, math]*, Apr 2020. URL http://arxiv.org/abs/2004.05788.

[66] N. Parikh and S. Boyd. Proximal algorithms. *Found. Trends Optim.*, 1(3):127239, Jan 2014. ISSN 2167-3888. doi:10.1561/2400000003.

[67] Z. Wen, C. Yang, X. Liu, and S. Marchesini. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems*, 28(11):115010, Oct 2012. ISSN 0266-5611. doi:10/gf3q42.

[68] J. Verbeeck, A. Béché, K. Müller-Caspary, G. Guzzinati, M. A. Luong, and M. Den Hertog. Demonstration of a 22 programmable phase plate for electrons. *Ultramicroscopy*, 190:5865, Jul 2018. ISSN 0304-3991. doi:10/gdpbjd.

[69] F. Kong, R. H. Silverman, L. Liu, P. V. Chitnis, K. K. Lee, and Y. C. Chen. Photoacoustic-guided convergence of light through optically diffusive media. *Optics Letters*, 36(11):20532055, Jun 2011. ISSN 1539-4794. doi:10/dhr9j8.

[70] A. M. Maiden, M. J. Humphry, M. C. Sarahan, B. Kraus, and J. M. Rodenburg. An annealing algorithm to correct positioning errors in ptychography. *Ultramicroscopy*, 120:6472, Sep 2012. ISSN 0304-3991. doi:10.1016/j.ultramic.2012.06.001.

[71] M. Odstril, A. Menzel, and M. Guizar-Sicairos. Iterative least-squares solver for generalized maximum-likelihood ptychography. *Optics Express*, 26(3):31083123, Feb 2018. ISSN 1094-4087. doi:10/gcx53m.

[72] A. Rana, J. Zhang, M. Pham, A. Yuan, Yuan-Hung Lo, H. Jiang, S. J. Osher, and J. Miao. Potential of attosecond coherent diffractive imaging. *Physical Review Letters*, 125(8):086101, Aug 2020. doi:10.1103/PhysRevLett.125.086101.

[73] P. Thibault and A. Menzel. Reconstructing state mixtures from diffraction measurements. *Nature*, 494(7435):6871, Feb 2013. ISSN 1476-4687. doi:10.1038/nature11806.

[74] Z. Chen, M. Odstrcil, Y. Jiang, Y. Han, M.-H. Chiu, L.-J. Li, and D. A. Muller. Mixed-state electron ptychography enables sub-angstrom resolution imaging with picometer precision at low dose. *Nature Communications*, 11(11): 2994, Jun 2020. ISSN 2041-1723. doi:10/gg2fbw.

Appendix B: Derivation of the adjoint S-matrix measurement operator

We derive the adjoint operators $\mathcal{A}_{k,d}^{\mathcal{S}_b\,\dagger}(\mathbf{z})$ and $\mathcal{A}_{k,d}^{\Psi_{d,b}\,\dagger}(\mathbf{z})$ with matrix algebra. In matrix notation, the $\mathcal{S}$-matrix forward model to generate $\mathbf{I} \in \mathbb{R}^{KDM}$ with $M = M_1 \cdot M_2$ from $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{BN}$ with $N = N_1 \cdot N_2$ can be written as $\mathcal{A}_{\mathcal{S}} : \mathbb{C}^{BN} \to \mathbb{C}^{KDM}$

$$\mathbf{I} = |\mathbf{F}\boldsymbol{\Sigma}\mathbf{C}\boldsymbol{\mathcal{S}}|^2 = |\mathcal{A}_{\mathcal{S}}\boldsymbol{\mathcal{S}}|^2, \tag{B1}$$

with $\mathbf{F} \in \mathbb{C}^{KDM \times KDM}$ a block-diagonal matrix representing a batched Fourier transform acting on KD exit waves, $\mathbf{C} \in \mathbb{R}^{KDBM \times BN}$ the cropping matrix that extracts KD patches centered at the scanning positions out of the B beams of the $\mathcal{S}$-matrix, and $\boldsymbol{\Sigma} \in \mathbb{C}^{KDM \times KDBM}$ the coherent summation operator over all beams. Written out in block matrices with diagonal entries, $\boldsymbol{\Sigma}$ is shown in Fig. 4. The adjoint (hermitian transpose) operator $\mathcal{A}_{\mathcal{S}}^{\dagger}$ is then

$$\mathcal{A}_{\mathcal{S}}^{\dagger} = \mathbf{C}^T \boldsymbol{\Sigma}^{\dagger} \mathbf{F}^{\dagger}, \tag{B2}$$

which, written out for a single diffraction pattern with defocus index d and position index k is

$$\mathcal{A}_{k,d}^{\mathcal{S}_b\,\dagger}(\mathbf{z}) = \mathbf{C}_{k,d}^T \left[\Psi_{d,b}^* e^{2\pi i \mathbf{h}_b \cdot \boldsymbol{\rho}_{k,d}} \mathcal{F}_{\mathbf{q}}^{\dagger} [\mathbf{z}_{k,d}] \right] \tag{B3}$$

In the same vein, the forward model to generate $\mathbf{I} \in \mathbb{R}^{KDM}$ from $\boldsymbol{\Psi} \in \mathbb{C}^{DB}$ can be written as $\mathcal{A}_{\boldsymbol{\Psi}} : \mathbb{C}^{DB} \to \mathbb{C}^{KDM}$

$$\mathbf{I} = |\mathbf{F}\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}\boldsymbol{\Psi}|^2 = |\mathcal{A}_{\boldsymbol{\Psi}}\boldsymbol{\Psi}|^2 \tag{B4}$$

The adjoint operator $\mathcal{A}_{\boldsymbol{\Psi}}^{\dagger}$ is then

$$\mathcal{A}_{\boldsymbol{\Psi}}^{\dagger} = \boldsymbol{\Sigma}_{\boldsymbol{\Psi}}^{\dagger} \mathbf{F}^{\dagger}, \tag{B5}$$

which, written out for a single diffraction pattern with defocus index d and position index k is

$$\mathcal{A}_{k,d}^{\Psi_{d,b}\,\dagger}(\mathbf{z}) = \frac{1}{M_1 M_2} \sum_{m_1}^{M_1} \sum_{m_2}^{M_2} \left[\sum_{k=1}^{K} [\mathbf{C}_{k,d}\boldsymbol{\mathcal{S}}]_b^* e^{2\pi i \mathbf{h}_b \cdot \boldsymbol{\rho}_{k,d}} \mathcal{F}_{\mathbf{q}}^{\dagger} [\mathbf{z}_{k,d}] \right]_{m_1, m_2} \tag{B6}$$

$$\boldsymbol{\Sigma}_{[:,1:\mathsf{BM}]} :=$$

$$
\begin{array}{cccc}
\mathsf{b}=1 & \mathsf{b}=2 & & \mathsf{b}=\mathsf{B} \\
\end{array}
$$

$$
\left[
\begin{array}{cccc}
\left[\mathrm{diag}(\Psi_{1,1}e^{-2\pi i \mathbf{h}_1\cdot\boldsymbol{\rho}_{1,1}})\right] & \left[\mathrm{diag}(\Psi_{1,2}e^{-2\pi i \mathbf{h}_2\cdot\boldsymbol{\rho}_{1,1}})\right] & \cdots & \left[\mathrm{diag}(\Psi_{1,\mathsf{B}}e^{-2\pi i \mathbf{h}_{\mathsf{B}}\cdot\boldsymbol{\rho}_{1,1}})\right] \\
\left[\mathrm{diag}(\Psi_{1,1}e^{-2\pi i \mathbf{h}_1\cdot\boldsymbol{\rho}_{1,1}})\right] & \left[\mathrm{diag}(\Psi_{1,2}e^{-2\pi i \mathbf{h}_2\cdot\boldsymbol{\rho}_{1,1}})\right] & \cdots & \left[\mathrm{diag}(\Psi_{1,\mathsf{B}}e^{-2\pi i \mathbf{h}_{\mathsf{B}}\cdot\boldsymbol{\rho}_{1,1}})\right] \\
\vdots & \vdots & \ddots & \vdots \\
\left[\mathrm{diag}(\Psi_{2,1}e^{-2\pi i \mathbf{h}_1\cdot\boldsymbol{\rho}_{1,2}})\right] & \left[\mathrm{diag}(\Psi_{2,2}e^{-2\pi i \mathbf{h}_2\cdot\boldsymbol{\rho}_{1,2}})\right] & \cdots & \left[\mathrm{diag}(\Psi_{2,\mathsf{B}}e^{-2\pi i \mathbf{h}_{\mathsf{B}}\cdot\boldsymbol{\rho}_{1,2}})\right] \\
\left[\mathrm{diag}(\Psi_{2,1}e^{-2\pi i \mathbf{h}_1\cdot\boldsymbol{\rho}_{2,2}})\right] & \left[\mathrm{diag}(\Psi_{2,2}e^{-2\pi i \mathbf{h}_2\cdot\boldsymbol{\rho}_{2,2}})\right] & \cdots & \left[\mathrm{diag}(\Psi_{2,\mathsf{B}}e^{-2\pi i \mathbf{h}_{\mathsf{B}}\cdot\boldsymbol{\rho}_{2,2}})\right] \\
\vdots & \vdots & & \vdots \\
\left[\mathrm{diag}(\Psi_{\mathsf{D},1}e^{-2\pi i \mathbf{h}_1\cdot\boldsymbol{\rho}_{\mathsf{K},\mathsf{D}}})\right] & \left[\mathrm{diag}(\Psi_{\mathsf{D},2}e^{-2\pi i \mathbf{h}_2\cdot\boldsymbol{\rho}_{\mathsf{K},\mathsf{D}}})\right] & \cdots & \left[\mathrm{diag}(\Psi_{\mathsf{D},\mathsf{B}}e^{-2\pi i \mathbf{h}_{\mathsf{B}}\cdot\boldsymbol{\rho}_{\mathsf{K},\mathsf{D}}})\right] \\
\end{array}
\right]
\begin{array}{l}
\mathsf{k}=1\ \mathsf{d}=1 \\[3em]
\mathsf{k}=2\ \mathsf{d}=1 \\[4em]
\mathsf{k}=1\ \mathsf{d}=2 \\[2em]
\mathsf{k}=2\ \mathsf{d}=2 \\[3em]
\mathsf{k}=\mathsf{K}\ \mathsf{d}=\mathsf{D}
\end{array}
$$

$$\in \mathbb{C}^{\mathsf{KDM}\times\mathsf{BM}}$$

FIG. 4. The first BM columns of $\boldsymbol{\Sigma}$ written out explicitly with diagonal matrix blocks of size $\mathsf{M}\times\mathsf{M}$.