Inverse Multislice Ptychography by Layer-Wise Optimisation and Sparse Matrix Decomposition

Arya Bangun, Oleh Melnyk, Benjamin März, Benedikt Diederichs, Alexander Clausen, Dieter Weber, Frank Filbir, and Knut Müller-Caspary

***Abstract*—We propose algorithms based on an optimisation method for inverse multislice ptychography in, e.g. electron microscopy. The multislice method is widely used to model the interaction between relativistic electrons and thick specimens. Since only the intensity of diffraction patterns can be recorded, the challenge in applying inverse multislice ptychography is to uniquely reconstruct the electrostatic potential in each slice up to some ambiguities. In this conceptual study, we show that a unique separation of atomic layers for simulated data is possible when considering a low acceleration voltage. We also introduce an adaptation for estimating the illuminating probe. For the sake of practical application, we finally present slice reconstructions using experimental 4D scanning transmission electron microscopy (STEM) data.**

***Index Terms*—Ptychography, Phase Retrieval, Multislice Method, Electron Microscopy.**

Manuscript received 23 April 2022; revised 11 August 2022; accepted 18 October 2022. Date of publication 2 November 2022; date of current version 16 November 2022. The work of Knut Müller-Caspary and Benjamin März was supported in part by the Helmholtz Association under Contract VH-NG 1317 (*more*STEM), and in part by the Deutsche Forschungsgemeinschaft under DFG Grant EXC 2089/1-390776260. The work of Knut Müller-Caspary, Dieter Weber, Arya Bangun, Alexander Clausen, Benjamin März, Oleh Melnyk, Benedikt Diederichs and Frank Filbir was supported was supported by the Helmholtz Association under Contract ZT-I-0025 (Ptychography 4.0). The work of Benedikt Diederichs, Frank Filbir, and Knut Müller-Caspary was supported by the Helmholtz Association under Grant ZT-I-PF-5-28 (EDARTI). The work of Frank Filbir and Oleh Melnyk was supported by the Helmholtz Association under Project BRELMM ZT-I-PF-4-024. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Doga Gursoy. *(Corresponding author: Arya Bangun.)*

Arya Bangun, Alexander Clausen, and Dieter Weber are with the Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich, 52425 Jülich, Germany (e-mail: a.bangun@fz-juelich.de; a.clausen@fz-juelich.de; d.weber@fz-juelich.de).

Oleh Melnyk is with the Mathematical Imaging and Data Analysis, Helmholtz Center Munich, 85764 Munich, Germany, and also with the Department of Mathematics, Technical University of Munich, 85746 Munich, Germany (e-mail: oleh.melnyk@helmholtz-muenchen.de).

Benjamin März is with the Department of Chemistry and Center for NanoScience, Ludwig-Maximilians- University of Munich (LMU), 81377 Münich, Germany (e-mail: benjamin.maerz@cup.lmu.de).

Benedikt Diederichs is with the Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich, 52425 Jülich, Germany, and also with the Mathematical Imaging and Data Analysis, Helmholtz Center Munich, 85764 Munich, Germany (e-mail: b.diederichs@fz-juelich.de).

Frank Filbir is with the Mathematical Imaging and Data Analysis, Helmholtz Center Munich, 85764 Munich, Germany (e-mail: filbir@helmholtz-muenchen.de).

Knut Müller-Caspary is with the Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich, 52425 Jülich, Germany, and also with the Department of Chemistry and Center for NanoScience, Ludwig-Maximilians- University of Munich (LMU), 81377 Münich, Germany (e-mail: k.mueller-caspary@cup.lmu.de).

Digital Object Identifier 10.1109/TCI.2022.3218993

I. Introduction

One of the fundamental challenges in electron microscopy is dealing with phase retrieval from the intensity of diffraction patterns. The reason for this problem is that current detectors used in electron microscopy are unable to record phase information, which is necessary for example to improve the image resolution, to understand the interaction of electrons and atoms within a material, and in particular to recover the electrostatic potential of the specimen.

Several methodologies and approaches have been developed for solving the phase problem. One of the most prevalent techniques is ptychography. Instead of only exploiting the intensity of a single diffraction pattern, ptychography takes advantage of a large set of subsequently recorded diffraction patterns stemming from multiple, partly overlapping illuminations of the object. The illumination, that is the incident electron wave, is also referred to as 'probe'. In essence, the acquisition of diffracted intensities from adjacent scanning positions provides additional information enabling to solve the phase problem [1], [2], [3]. Following on from these original approaches, various contributions to the phase retrieval from a single diffraction pattern have led to the introduction of new algorithms, i.e. enhanced methods adapted for ptychographic reconstructions. For instance, classical alternating projection-based algorithms like Gerchberg Saxton (GS) [4] and Fienup Hybrid Input-Output (HIO) [5] inspired Ptychographic Iterative Engine (PIE) algorithms [6], [7], [8]. Another approach for solving the phase problem by direct inversion has been proposed in refs. [9], [10]. It utilises the property of the ambiguity function, sometimes also called Wigner distribution function, which naturally appears by reformulating the equation for deriving the intensity in terms of the probe and object transfer functions.

In addition to the aforementioned methods, further approaches for modelling ptychography as an optimisation problem have been developed over the last few years. As the phase retrieval problem is generally non-convex, there is no certainty that the global optimum can be attained. However, several contributions [11], [12], [13] manage to achieve the convergence to a local optimum. The crucial fundamental assumption for most studies is the single multiplicative approximation used for modelling the interaction between the electron beam and a thin specimen. However, this assumption does not necessarily hold when investigating thick specimens due to strong dynamical electron scattering effects [14], [15]. For thick specimens, one should take into account the theory of multiple scattering and

propagation when solving the phase problem, e.g. via the multislice approach [14], [16], Bloch waves [17], [18], i.e. scattering matrix-based formulations [19].

Several attempts have been made to adapt the phase retrieval model for thick specimens by incorporating the scattering matrix, as discussed in [19], [20], [21]. In [19], the authors developed an iterative projection algorithm called $N - phaser$ for estimating the scattering matrix. The key idea stems from the specific eigenvalue structure of the scattering matrix. It can therefore be used for estimating the object transfer function from a thick specimen while eliminating the unwanted scattering artefacts in the recorded diffraction patterns. Another approach has been proposed in [21], where the authors used optimisation methods, e.g. Alternating Direction Method of Multipliers (ADMM) and block coordinate descent, in order to estimate both the scattering matrix as well as the probe. A similar approach to estimate object and probe is presented in [22], where the reconstruction is done iteratively by a modified Gauss–Newton method.

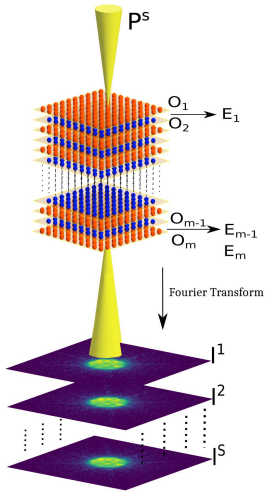


Fig. 1. Illustration of the multislice method showing a probe focused on a specimen and the resulting diffraction intensity $\mathbf{I}^s$ at scanning position s recorded in the far field. After each slice $\mathbf{X}_M$ of the specimen an exit wave $\mathbf{E}_M^s$ is produced.

A. Related Work

Implementations of inverse multislice ptychography have been applied for instance in [23], [24], [25]. The key idea in these studies bears a strong resemblance to extending the established algorithms, such as the extended PIE. As three-dimensional specimens were investigated, these algorithms were named 3PIE. The forward model deals with the propagation of the specimen entrance wave (probe) and calculates the complex wave function for the observed specimen at a specific thickness. The backward model constructs an estimation of the entrance wave by applying an inverse Fourier transform to the product of the estimated phase and the intensity of diffraction patterns acquired in the measurement. However, mentioned works focus on the reconstruction of visible light and x-ray data sets. The same algorithm has been applied to reconstruct images from the LED array microscope data in [26]. Another approach are gradient-based methods, where the gradients are calculated over the whole multislice model all at once. Examples can be found in [27], [28], [29], [30].

In order to address the inverse multislice ptychography problem for electron microscopy data sets we present two different approaches, an adaptation from the Amplitude Flow method and a matrix decomposition, respectively. Amplitude Flow is a gradient-based method, which has been analysed for a randomised one-dimensional phase retrieval [11]. This analysis was later enhanced for arbitrary measurements and in particular for ptychography [31]. In the second approach, a matrix factorisation technique adopted from the field of optimisation and dictionary learning [32] is incorporated into the estimation of the matrix from intensity measurements. Apart from proposing different techniques to solve inverse multislice ptychography we have also reformulated the forward multislice model. This adjustment enables separation of the effect of the illuminating probe from interaction with the specimen, which in turn yields only the construction of a thick object transmission function in respect of a single matrix. Additionally, we outline a methodology to reconstruct each atomic plane by applying the proposed algorithm to synthetic data simulated for a low acceleration voltage.

B. Summary of Contributions

- The forward multislice model is reformulated to disentangle the effect of the probe from the object transfer function at any thickness. This allows modelling a thick object as a matrix consisting only of the product of phase gratings and Fresnel propagators.
- Two approaches for estimating the slices of a thick object are proposed, namely layer-wise optimisation and sparse matrix decomposition. In the first approach we cycle over slices applying the Amplitude Flow algorithm, only optimising with respect to a single slice. The matrix which represents the phase gratings and the Fresnel propagations is recovered by applying a second algorithm. Further factorisation of this matrix is then carried out in order to extract the slices of the object. Additionally, the proposed algorithms are compared numerically to 3PIE [23] in terms of reconstruction and computational time.
- Simulations of diffraction data of specimens (MoS_2, $SrTiO_3$ and GaAs) with different crystal structures are carried out. These simulations are performed for different energies of the incoming electrons, i.e. different wavelengths. They serve as the ground truth for determining the error of the reconstructions, which are carried out using the proposed algorithms. Since the depth resolution in an electron microscope is limited, we investigate the impact of the slice thickness and under what conditions a unique reconstruction is possible. We also adapt the algorithm to estimate the illuminating probe and present the probe reconstruction.

- To highlight the practical use of our proposed methods we provide results for the first applications of the algorithms in respect of experimental data, notably in our reconstructions of the object transfer function of a MoS_2 specimen using a four-dimensional data set acquired by scanning transmission electron microscopy (STEM).

C. Notations

Vectors are written in bold small-cap letters $\mathbf{x} \in \mathbb{C}^L$ and matrices are written as a bold big-cap letter $\mathbf{A} \in \mathbb{C}^{K\times L}$ for a complex field $\mathbb{C}$ and for a real field $\mathbb{R}$. A matrix can also be written by indexing its elements, $\mathbf{A}=(a_{k\ell})$, where $k\in[K], \ell\in[L]$. The set of integers is written as $[N] := \{1, 2, \dots, N\}$ and calligraphic letters are used to define functions $\mathcal{A} : \mathbb{C} \to \mathbb{C}$. Specifically, we denote the discrete two-dimensional Fourier transform by $\mathcal{F}$. For both matrices and vectors, the notation $\circ$ is used to represent element-wise or Hadamard product. The $\mathbf{A}^H$ is used to represent conjugate transpose. For a vector $\mathbf{x} \in \mathbb{C}^L$, the ℓ_p-norm is given by $\|\mathbf{x}\|_p = (\sum_{\ell=1}^{L} |x_\ell|^p)^{1/p}$, $1 \le p < \infty$ and for $p = \infty$ we have $\max_{\ell\in[L]} |x_\ell|$. For a matrix $\mathbf{X} \in \mathbb{C}^{K\times L}$, the Frobenius norm is denoted by $\|\mathbf{X}\|_F := \sqrt{\sum_{k=1}^{K}\sum_{\ell=1}^{L} |X_{k\ell}|^2}$ and the spectral norm is given by $\|\mathbf{X}\| = \max_{\|\mathbf{v}\|_2=1} \|\mathbf{X}\mathbf{v}\|_2$. The trace operator $\mathrm{Tr}(.)$ is the operator to sum all elements in the diagonal of square matrices. Indices are wrapped around, so that $\mathbf{x}_{-p} = \mathbf{x}_{N-p}$.

II. Problem Statement

A. Forward Multislice Model

The forward multislice model is based on the idea that a thick object can be approximated by multiple thin slices stacked on top of each other. For each slice of the specimen the interaction between the electron wave incident on this slice and the potential of the slice can be modelled by a multiplicative approximation in analogy to the standard model in ptychography. Furthermore, as the illumination progresses through the object, the exit wave of the previous slice propagates through potential-free space to the subsequent slice, where it acts as the new illumination for this slice, as schematically shown in Fig. 1.

Considering an aberration-free probe, the two-dimensional probe $\mathbf{P} \in \mathbb{C}^{N\times N}$ can be described for different aperture sizes $q_{\max}$ with entries given by

$$p_{y,x} = \pi q_{\max}^2 \left(\frac{2J_1\left(2\pi q_{\max}\,|\mathbf{r}|\right)}{2\pi q_{\max}\,|\mathbf{r}|} \right), \text{where}$$

$$y, x \in [N],\ |\mathbf{r}| = \sqrt{x^2+y^2},$$

where J_1 is a Bessel function of the first kind of order 1. The intensity of this function is called Airy disk. In general, this function can be derived analytically by applying a two-dimensional inverse Fourier transform to a circular aperture. The probe shifted to the scanning position (x_s, y_s), $s \in [S]$ is denoted by a matrix $\mathbf{P}^s \in \mathbb{C}^{N\times N}$ with entries $p^s_{x,y} = p_{x-x_s, y-y_s}$. In general aberrations exist and affect the probe formation. In this study the focus is on the aberration-free condition when generating the simulated data. For a more general treatment of this subject please refer to [15].

The interaction between the probe $\mathbf{P}^s \in \mathbb{C}^{N\times N}$ at scanning point $s \in [S]$ and the first slice $\mathbf{X}_1 \in \mathbb{C}^{N\times N}$ is given by the element-wise product and in turn produces an exit wave of slice 1

$$\mathbf{E}^s_1 = \mathbf{P}^s \circ \mathbf{X}_1.$$

After passing through the first slice the propagation of the exit wave between the slices is modelled by the Fresnel transform $\mathcal{V}_z$ which is given by

$$\mathcal{V}_z\left(\mathbf{E}\right) := \mathcal{F}^{-1}\left(\mathcal{F}\left(\mathbf{E}\right) \circ \mathbf{H}_m\right),$$

where $\mathcal{F}$ is the Fourier operator and $\mathbf{H}_m \in \mathbb{C}^{N\times N}$ is the Fresnel propagator matrix with entries

$$h_{y,x} := e^{-\pi i \Delta_m \lambda \left(\left(q_y^2+q_x^2\right) + 2\left(q_x \frac{\sin\theta_x}{\lambda} + q_y \frac{\sin\theta_y}{\lambda}\right)\right)}, \quad y, x \in [N]. \tag{1}$$

The parameters q_y, q_x denote the discrete grid in the reciprocal space and hence represent spatial frequencies, Δ_m is the distance of the wave propagation, and θ_x, θ_y are the two-dimensional tilt angles. In this article, the illumination was set to be perpendicular to the object surface along a major crystallographic axis, i.e. tilt angles are zero.

As the beam reaches the second slice, it is described by $\mathcal{V}_z(\mathbf{E}^s_1)$ and the next exit wave is given by

$$\mathbf{E}^s_2 = \mathcal{V}_z\left(\mathbf{E}^s_1\right) \circ \mathbf{X}_2.$$

Consequently, the general representation of the m-th observed exit wave is written as

$$\mathbf{E}^s_m = \mathcal{V}_z\left(\mathbf{E}^s_{m-1}\right) \circ \mathbf{X}_m \quad \text{for} \quad m \in [M], m \neq 1,$$

Finally, the intensity of the Fraunhofer diffraction pattern that is recorded by a detector in the far field is given by

$$\mathbf{I}^s = \left|\mathcal{F}\left(\mathbf{E}^s_M\right)\right|^2. \tag{2}$$

In scanning transmission electron microscopy (STEM) the illuminating probe is rastered across the specimen. Therefore, the illumination is varied to yield a set of S diffraction pattern intensities collected throughout an experiment. This four-dimensional data set is then subjected to phase retrieval by multislice ptychography.

B. Reformulation of Multislice Ptychography

The measurement model in (2) can be further reformulated in order to separate the probe- and the object-related terms. This reformulation is based on the following property of the Hadamard product. For matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{N\times N}$ the Hadamard product $\mathbf{A} \circ \mathbf{B}$ can be written according to

$$\begin{aligned} \mathrm{vec}\left(\mathbf{A} \circ \mathbf{B}\right) &= \mathrm{diag}\left(\mathrm{vec}\left(\mathbf{A}\right)\right) \mathrm{vec}\left(\mathbf{B}\right) \\ &= \mathrm{diag}\left(\mathrm{vec}\left(\mathbf{B}\right)\right) \mathrm{vec}\left(\mathbf{A}\right). \end{aligned} \tag{3}$$

The notation $\mathrm{vec} : \mathbb{C}^{N\times N} \to \mathbb{C}^{N^2}$ is an operator that vectorises the matrix and $\mathrm{diag} : \mathbb{C}^{N^2} \to \mathbb{C}^{N^2\times N^2}$ constructs a diagonal matrix by placing the elements of the given vector on the main diagonal. Additionally, the second equality in (3) is valid from

the commutative property of the Hadamard product. By using (3) the first exit wave for s-th position of the probe can be rewritten as

$$\operatorname{vec}\left(\mathbf{E}_1^s\right) = \operatorname{vec}\left(\mathbf{P}^s \circ \mathbf{X}_1\right) = \operatorname{diag}\left(\operatorname{vec}\left(\mathbf{X}_1\right)\right) \operatorname{vec}\left(\mathbf{P}^s\right).$$

For convenience, the following notations are introduced

$$\mathbf{o}_m := \operatorname{vec}\left(\mathbf{X}_m\right) \in \mathbb{C}^{N^2},$$

$$\mathbf{O}_m := \operatorname{diag}\left(\mathbf{o}_m\right) \in \mathbb{C}^{N^2 \times N^2} \text{ for } m \in [M],$$

and

$$\mathbf{p}^s := \operatorname{vec}\left(\mathbf{P}^s\right) \in \mathbb{C}^{N^2} \text{ for } s \in [S].$$

As a result, the first exit wave can be simplified as

$$\operatorname{vec}\left(\mathbf{E}_1^s\right) = \mathbf{O}_1 \mathbf{p}^s.$$

For the second exit wave the action of the Fresnel propagator is required. Let $\mathbf{F}_{2D}, \mathbf{F}_{2D}^{-1} \in \mathbb{C}^{N^2 \times N^2}$ denote the vectorized representation of the two-dimensional Fourier and inverse Fourier transforms as matrices, respectively. The term one-dimensional Fourier matrix represents the discrete implementation of the Fourier basis, i.e., when the Fourier basis is sampled and stored as a matrix. Likewise, the two-dimensional Fourier matrix can be constructed by using the Kronecker product of two one-dimensional Fourier matrices. Consequently, for a two-dimensional object $\mathbf{V} \in \mathbb{C}^{N \times N}$ the product of $\mathbf{F}_{2D}$ with $\operatorname{vec}(\mathbf{V})$ corresponds to the vectorization of the two-dimensional Fourier transform of $\mathbf{V}$, i.e., $\mathbf{F}_{2D}\operatorname{vec}(\mathbf{V}) = \operatorname{vec}(\mathcal{F}(\mathbf{V}))$. By facilitating (3), the second exit wave is given by

$$\begin{aligned}
\operatorname{vec}\left(\mathcal{V}_z\left(\mathbf{E}_1^{\mathbf{s}}\right)\right) &= \operatorname{vec}\left(\mathcal{F}^{-1}\left(\mathcal{F}\left(\mathbf{E}_1^{\mathbf{s}}\right) \circ \mathbf{H}_1\right)\right) \\
&= \mathbf{F}_{2D}^{-1} \operatorname{vec}\left(\left(\mathcal{F}\left(\mathbf{E}_1^{\mathbf{s}}\right) \circ \mathbf{H}_1\right)\right) \\
&= \mathbf{F}_{2D}^{-1} \operatorname{diag}\left(\operatorname{vec}\left(\mathbf{H}_1\right)\right) \operatorname{vec}\left(\mathcal{F}\left(\mathbf{E}_1^{\mathbf{s}}\right)\right) \\
&= \mathbf{F}_{2D}^{-1} \operatorname{diag}\left(\operatorname{vec}\left(\mathbf{H}_1\right)\right) \mathbf{F}_{2D} \operatorname{vec}\left(\mathbf{E}_1^{\mathbf{s}}\right).
\end{aligned}$$

Hence, the Fresnel propagator is a multiplication of the vectorised exit wave $\mathbf{E}_m^s$ with a matrix

$$\mathbf{G}_m := \mathbf{F}_{2D}^{-1} \operatorname{diag}\left(\operatorname{vec}\left(\mathbf{H}_m\right)\right) \mathbf{F}_{2D} \in \mathbb{C}^{N^2 \times N^2},\ m \in [M-1].$$

Substituting the obtained representation to the second exit wave results in

$$\begin{aligned}
\operatorname{vec}\left(\mathbf{E}_{\mathbf{2}}^{\mathbf{s}}\right) &= \operatorname{vec}\left(\mathcal{V}_z\left(\mathbf{E}_1^s\right) \circ \mathbf{X}_2\right) \\
&= \operatorname{diag}\left(\operatorname{vec}\left(\mathbf{X}_2\right)\right) \operatorname{vec}\left(\mathcal{V}_z\left(\mathbf{E}_1^s\right)\right) = \mathbf{O}_2 \mathbf{G}_1 \mathbf{O}_1 \mathbf{p}^s.
\end{aligned}$$

Consequently, the M-th exit wave is given by

$$\operatorname{vec}\left(\mathbf{E}_M^s\right) = \left(\mathbf{O}_M \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m\right) \mathbf{p}^s =: \mathbf{A}_M \mathbf{p}^s \tag{4}$$

and the resulting far-field vectorised intensity is

$$\begin{aligned}
\mathbf{i}^s := \operatorname{vec}\left(\mathbf{I}^s\right) &= \operatorname{vec}\left(\left|\mathcal{F}\left(\mathbf{E}_M^s\right)\right|^2\right) = \left|\operatorname{vec}\left(\mathcal{F}\left(\mathbf{E}_M^s\right)\right)\right|^2 \\
&= \left|\mathbf{F}_{2D} \operatorname{vec}\left(\mathbf{E}_M^s\right)\right|^2 = \left|\mathbf{F}_{2D} \mathbf{A}_M \mathbf{p}^s\right|^2.
\end{aligned}$$

By combining all vectorised intensities $\mathbf{i}^s$ and probes $\mathbf{p}^s$ as columns of the matrices

$$\mathbf{I} := \left(\mathbf{i}^1, \mathbf{i}^2, \ldots, \mathbf{i}^S\right) \in \mathbb{C}^{N^2 \times S} \text{ and}$$

$$\mathbf{P} := \left(\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^S\right) \in \mathbb{C}^{N^2 \times S},$$

respectively, can be simplified to

$$\mathbf{I} = \left|\mathbf{F}_{2D} \mathbf{A}_M \mathbf{P}\right|^2. \tag{5}$$

There are at least two benefits of this reformulation. Firstly, the object transfer function for an arbitrary thickness M is now represented by the matrix $\mathbf{A}_M \in \mathbb{C}^{N^2 \times N^2}$. Note that $\mathbf{A}_M$ purely represents the properties of the thick specimen without being affected by the probe, contrary to the model (2), where the probe's illumination is entangled with the slices. Secondly, the matrix $\mathbf{A}_M$ decomposes into the product according to (4) and each slice $\mathbf{O}_m$ can therefore be separated from other multipliers, i.e. Fresnel propagator $\mathbf{G}_m$, which will be convenient in the next section where the recovery of a thick object is discussed.

With this reformulation the matrices now have an ambient dimension of $N^2 \times N^2$ which increases the computational complexity for processing data of this form. However, because most of these matrices are diagonal matrices they may therefore allow a more efficient treatment and storage in comparison to, e.g. the Bloch wave method, in which an eigenvalue decomposition is directly performed on a scattering matrix of dimension $N^2 \times N^2$.

III. Methods and Algorithms

This section considers the recovery of a thick object and the probe from intensity measurements (5) in diffraction space. Firstly it is posed as a constrained optimisation problem, then two algorithms are proposed for solving the optimisation problem under the assumption that the probes are known. Finally concepts for incorporating the probe estimation into the suggested methods are provided.

A. Inverse Multislice Ptychography as an Optimisation Problem

One of the standard approaches for the recovery of the object from intensity measurements is the optimisation of the data fidelity, which is represented by the least squares problem

$$\begin{aligned}
&\underset{\mathbf{P}, \mathbf{O}_m, m \in [M]}{\text{minimize}} && \left\| \sqrt{\mathbf{I}} - \left|\mathbf{F}_{2D} \mathbf{A}_M \mathbf{P}\right| \right\|_F^2 \\
&\text{subject to} && \mathbf{A}_M = \mathbf{O}_M \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m.
\end{aligned} \tag{6}$$

The challenge in obtaining the minimiser of (6) is the non-convexity of the objective function, which results from the absolute value and product representation of the matrix $\mathbf{A}_M$ as well as the multiplication with the probes $\mathbf{P}$. In general, non-convex functions are known to require non-polynomial time to find the global optima [33].

One common method for tackling a non-convex minimisation is the alternating projections method [34], [35]. It is based on minimising the objective function with respect to a single

selected unknown at a time, while other unknowns remain fixed. This usually results in simpler intermediate subproblems which can be solved efficiently. Afterwards, the next unknown is chosen for optimisation and this process is continued until the minimisation with respect to any of the variables does not improve further. Whereas the alternating minimisation is well-understood for convex functions [34], it often acts as a heuristic for non-convex ones. Nevertheless, in applications such as ptychography [36], the estimates obtained by alternating minimisation are quite accurate, which motivates applying this technique to inverse multislice ptychography.

B. Phase Retrieval via Amplitude Flow

As observed throughout this section, the minimisation of the objective function in (6) with respect to a single unknown, either $\mathbf{O}_m, \mathbf{A}_M$ or $\mathbf{P}$, leads to the phase retrieval problem. It concerns the recovery of an unknown vector $\mathbf{z} \in \mathbb{C}^L$ from the measurements of the form

$$\mathbf{y} = |\mathbf{Q}\mathbf{z}|^2 \in \mathbb{R}^K,$$

with the measurement matrix $\mathbf{Q} \in \mathbb{C}^{K\times L}$. One popular approach for the reconstruction of $\mathbf{z}$ is the Amplitude Flow algorithm [11], [31]. It applies the gradient descent in order to minimise the least squares objective

$$\mathcal{A}(\mathbf{z}) = \|\sqrt{\mathbf{y}} - |\mathbf{Q}\mathbf{z}|\|_2^2. \tag{7}$$

The generalised Wirtinger gradient of the function $\mathcal{A}$ is given by

$$\nabla\mathcal{A}(\mathbf{z}) = \mathbf{Q}^H\left(\mathbf{Q}\mathbf{z} - \frac{\mathbf{Q}\mathbf{z}}{|\mathbf{Q}\mathbf{z}|}\circ\sqrt{\mathbf{y}}\right),$$

where each element $k \in [K]$ in the fraction $(\frac{\mathbf{Q}\mathbf{z}}{|\mathbf{Q}\mathbf{z}|})_k$ is set to 0 whenever $(\mathbf{Q}\mathbf{z})_k = 0$. Then, starting from a position $\mathbf{z}^0$, the t-th iteration is obtained via the gradient step

$$\mathbf{z}^t = \mathbf{z}^{t-1} - \mu\nabla\mathcal{A}\left(\mathbf{z}^{t-1}\right), \tag{8}$$

with step size $\mu = \|\mathbf{Q}\|^{-2}$ given by the squared inverse of the spectral norm of the matrix $\mathbf{Q}$. In this way, the value of the objective function (7) does not increase and the algorithm converges to a critical point of $\mathcal{A}$ [31].

C. Layer-Wise Optimisation

For the layer-wise optimisation, the alternating minimisation was adopted so as to optimise (6) with respect to a single slice $\mathbf{O}_m$ at a time, while keeping all other slices unchanged. Note that the objective in the optimisation problem (6) can be understood as a sum of errors for each scanning point s,

$$\underset{\mathbf{O}_m, m\in[M]}{\text{minimize}} \quad \sum_{s=1}^{S}\left\|\sqrt{\mathbf{i}^s} - \left|\mathbf{F}_{2D}\mathbf{O}_M\prod_{m=1}^{M-1}\mathbf{G}_m\mathbf{O}_m\mathbf{p}^s\right|\right\|_2^2 \tag{9}$$

Using initial guesses of the object transfer functions $\mathbf{O}_1^0, \ldots, \mathbf{O}_M^0$, the alternating minimisation technique is employed in order to optimise with respect to a single slice $\mathbf{O}_\ell$, $\ell \in [M]$ by solving

$$\mathbf{O}_\ell^{t+1} = \underset{\mathbf{O}_\ell}{\arg\min}\sum_{s\in[S]}\left\|\sqrt{\mathbf{i}^s} - \left|\mathbf{F}_{2D}\mathbf{R}_\ell^t\mathbf{O}_\ell\mathbf{S}_\ell^t\mathbf{p}^s\right|\right\|_2^2, \tag{10}$$

where the supporting prefix and suffix matrices are given by

$$\mathbf{R}_\ell^t := \prod_{m=\ell+1}^{M}\mathbf{O}_m^t\mathbf{G}_{m-1} \text{ and } \mathbf{S}_\ell^t := \prod_{m=1}^{\ell-1}\mathbf{G}_m\mathbf{O}_m^{t+1}. \tag{11}$$

Once an estimate for the ℓ-th slice is produced, the algorithm continues with the $\ell+1$-th slice. After the L-th slice is estimated, the estimation process is repeated from the first slice until a desired stopping criterion is reached. By applying (3), the intensity measurement for a single probe $\mathbf{p}^s, s \in [S]$ can be rearranged according to

$$\begin{aligned}\left|\mathbf{F}_{2D}\mathbf{R}_\ell^t\mathbf{O}_\ell\mathbf{S}_\ell^t\mathbf{p}^s\right| &= \left|\mathbf{F}_{2D}\mathbf{R}_\ell^t\,\mathrm{diag}\left(\mathbf{o}_\ell\right)\mathbf{S}_\ell^t\mathbf{p}^s\right| \\ &= \left|\mathbf{F}_{2D}\mathbf{R}_\ell^t\,\mathrm{diag}\left(\mathbf{S}_\ell^t\mathbf{p}^s\right)\mathbf{o}_\ell\right|.\end{aligned}$$

The optimisation problem (10) is equivalent to the phase retrieval problem

$$\mathbf{o}_\ell^{t+1} = \underset{\mathbf{o}_\ell}{\mathrm{argmin}}\left\|\sqrt{\mathbf{y}} - \left|\mathbf{Q}_\ell^t\mathbf{o}_\ell\right|\right\|_2^2, \tag{12}$$

with the measurement matrix and the measurements given by

$$\mathbf{Q}_\ell^t := \begin{bmatrix}\mathbf{F}_{2D}\mathbf{R}_\ell^t\,\mathrm{diag}\left(\mathbf{S}_\ell^t\mathbf{p}^1\right)\\ \vdots \\ \mathbf{F}_{2D}\mathbf{R}_\ell^t\,\mathrm{diag}\left(\mathbf{S}_\ell^t\mathbf{p}^S\right)\end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix}\mathbf{i}^1\\ \vdots \\ \mathbf{i}^S\end{bmatrix}, \tag{13}$$

respectively. The problem in (12) can be solved by running the gradient descent method as discussed in Section III-B for a fixed number of iterations.

Overall, it grants us the Algorithm 1 summarised below.

Since each gradient step in Layer-wise Optimisation does not increase the value of the objective function in (9), Algorithm 1

Algorithm 1: Layer-wise Optimisation.

1: **Initialisation:**
- Initial object transfer functions $\mathbf{O}_1^0, \ldots, \mathbf{O}_M^0$, are all equal to identity matrices.
- Intensity measurement $\mathbf{I} \in \mathbb{R}^{N^2\times S}$.
- Number of iterations T and number of gradient steps K.

2: **for** each iteration $t \in [T]$ **do**
3: **for** each layer $\ell \in [M]$ **do**
4: Compute prefix and suffix matrices $\hat{\mathbf{R}}_\ell^t$ and $\hat{\mathbf{S}}_\ell^t$ via (11).
5: Construct the measurement matrix $\mathbf{Q}_\ell^t$ and measurement $\mathbf{y}$ as in (13).
6: Produce estimate $\mathbf{O}_\ell^{t+1} = \mathrm{diag}(\mathbf{o}_\ell^{t+1})$ as in (12) by performing K gradient steps (8) with starting point $\mathbf{o}_\ell^t$ corresponding to the diagonal elements of $\mathbf{O}_\ell^t$.
7: **end for**
8: If convergence criteria is reached $\rightarrow$ Stop
9: **end for**

will always converge in a sense that it will stop at a certain value of the objective.

D. Sparse Matrix Decomposition

Another approach for solving the optimisation problem (6) is separating it into two subproblems. At first, the matrix $\mathbf{A}_M \in \mathbb{C}^{N^2\times N^2}$, which represents the complex object transfer function of a thick specimen is estimated from the measurements. In the second step, this estimated matrix $\mathbf{A}_M$ is to be decomposed in order to determine the slices $\mathbf{O}_m^0$, $m \in [M]$, which is sparse in the sense that only diagonal elements are non-zero. Both steps are then reiterated. The detailed procedure is described below.

In the first step, the object transfer function is estimated by solving the optimisation problem

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \quad \frac{1}{2}\left\| \sqrt{\mathbf{I}} - |\mathbf{F}_{2D}\mathbf{A}\mathbf{P}| \right\|_F^2. \tag{14}$$

In case the matrix $\mathbf{A}$ is vectorised, (14) can be treated as a phase retrieval problem of the form (7), which gives the gradient step

$$\mathbf{A}^{t+1} = \mathbf{A}^t - \mu \mathbf{F}_{2D}^H \left(\left(\left| \mathbf{F}_{2D}\mathbf{A}^t\mathbf{P} \right| - \sqrt{\mathbf{I}} \right) \circ \frac{\mathbf{F}_{2D}\mathbf{A}^t\mathbf{P}}{|\mathbf{F}_{2D}\mathbf{A}^t\mathbf{P}|} \right) \mathbf{P}^H,$$

with the step size $\mu = \frac{1}{\|\mathbf{P}\|^2\|\mathbf{F}_{2D}\|^2} = \frac{1}{N^2\|\mathbf{P}\|^2}$. The initial guess of the matrix $\mathbf{A}$ is

$$\mathbf{A}^0 = \mathbf{O}_M^0 \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m^0, \tag{15}$$

with $\mathbf{O}_m^0, m \in [M]$ being the intialisations for each slice. Once the matrix $\mathbf{A}$ is estimated using $\hat{\mathbf{A}}$, its sparse decomposition [32] is achieved by solving the following problem

$$\underset{\substack{\lambda,\, \mathbf{O}_m, m\in[M] \\ \mathbf{O}_m - \text{diagonal} \\ \|\mathbf{O}_m\|_F = 1}}{\text{minimize}} \frac{1}{2}\left\| \hat{\mathbf{A}} - \lambda \mathbf{O}_M \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m \right\|_F^2.$$

In view of the fact that for any set of multipliers $\{\alpha_m \mathbf{O}_m, m \in [M]\}$ such that $\prod_{m=1}^{M} \alpha_m = 1$ the slices $\alpha_m \mathbf{O}_m$ will generate the same $\mathbf{A}_M$, we address this ambiguity by normalising $\mathbf{O}_m$ during the optimisation and by introducing the data fidelity parameter λ. Simultaneous minimisation with respect to all unknowns is cumbersome. Instead the alternating minimisation technique is employed.

Starting with initial guesses $\mathbf{O}_m^0, m \in [M]$ as used in (15) and $\lambda^0 = 1$ for the ℓ-th slice, $\ell \in [M]$, the new estimate is obtained by minimising

$$\mathbf{O}_\ell^{t+1} = \underset{\substack{\mathbf{O}_\ell - \text{diagonal} \\ \|\mathbf{O}_\ell\|_F = 1}}{\arg\min} \frac{1}{2}\left\| \hat{\mathbf{A}} - \lambda^t \mathbf{R}_\ell^t \mathbf{O}_\ell \mathbf{S}_\ell^t \right\|_F^2,$$

where the objective is reformulated in terms of the prefix and suffix (11) matrices. Thereby, proximal gradient descent methods [37] can be applied, which grants an update of the form

$$\mathbf{O}_\ell^{t+1} = \mathcal{P}\left(\mathbf{O}_\ell^t - \mu\lambda^t \left(\mathbf{R}_\ell^t\right)^H \left(\lambda^t \mathbf{R}_\ell^t \mathbf{O}_\ell \mathbf{S}_\ell^t - \hat{\mathbf{A}} \right) \left(\mathbf{S}_\ell^t\right)^H \right), \tag{16}$$

Algorithm 2: Sparse Matrix Decomposition.

1: **Initialisation:**
- Initial matrix $\mathbf{A}^0 = \mathbf{O}_M^0 \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m^0$, where the object is initialised with an identity matrix.
- Intensity measurement $\mathbf{I} \in \mathbb{R}^{N^2\times S}$
- Number of iterations T and regularization λ^0

2: **for** each iteration $t \in [T]$ **do**
3: Estimate matrix $\hat{\mathbf{A}}$ with initial $\mathbf{A}^t$ by solving (14).
4: **for** each layer $\ell \in [M]$ **do**
5: Compute prefix matrix : $\mathbf{R}_\ell^t$
6: Compute suffix matrix : $\mathbf{S}_\ell^t$
7: Estimate : $\mathbf{O}_\ell^{t+1}$ by solving (16)
8: **end for**
9: Set $\tilde{\mathbf{A}} = \mathbf{O}_M^{t+1} \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m^{t+1}$
10: Update λ^{t+1} via (17)
11: Update $\mathbf{A}^{t+1} = \lambda^{t+1}\mathbf{O}_M^{t+1} \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m^{t+1}$
12: If convergence criteria is reached $\rightarrow$ Stop
13: **end for**
14: **return** $(\lambda^T)^{\frac{1}{M}}\mathbf{O}_1^T, \ldots, (\lambda^T)^{\frac{1}{M}}\mathbf{O}_M^T$

with the projection operator $\mathcal{P}$ acting onto the space of diagonally normalised matrices given by

$$(\mathcal{P}(\mathbf{X}))_{k,j} = \begin{cases} \frac{\mathbf{X}_{k,k}}{\sqrt{\sum_{n=1}^{N^2} |\mathbf{X}_{n,n}|^2}}, & k = j, \\ 0, & k \neq j, \end{cases} \quad k, j \in [N^2],$$

and the step size $\mu = \frac{1}{c}$ where $c > (\lambda^t \|\mathbf{R}_\ell^t\| \|\mathbf{S}_\ell^t\|)^2$, as discussed in [32].

For the minimisation with respect to λ, estimates of the M-th slice are combined according to $\tilde{\mathbf{A}} = \mathbf{O}_M^{t+1} \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m^{t+1}$ and λ is updated by minimising it according to the one-parameter least squares problem

$$\lambda^{t+1} = \arg\min_{\lambda} \frac{1}{2}\left\| \hat{\mathbf{A}} - \lambda \tilde{\mathbf{A}} \right\|_F^2$$

.

Therefore, the update for λ^{t+1} is given by

$$\lambda^{t+1} = \frac{\text{Tr}\left(\hat{\mathbf{A}}^H \tilde{\mathbf{A}} \right)}{\text{Tr}\left(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} \right)}, \tag{17}$$

which concludes the second step of our method. These two steps are reiterated by using the new initialisation for the first step $\mathbf{A}^{t+1} = \lambda^{t+1}\mathbf{O}_M^{t+1} \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m^{t+1}$. The summary of the procedure is given in Algorithm 2.

E. Probe Reconstruction

After estimating the object i.e. the phase gratings of a specimen the optimisation method can be adapted related to Amplitude Flow in (7) in order to estimate the centered probe $\mathbf{p}^c \in \mathbb{C}^{N^2}$, by utilising the intensity of diffraction patterns at the same position, i.e. $\mathbf{i}^c \in \mathbb{R}^{N^2}$,

$$\underset{\mathbf{p}^c}{\text{minimize}} \qquad \mathcal{A}\left(\mathbf{p}^c\right) := \frac{1}{2}\left\| \sqrt{\mathbf{i}^c} - \left| \mathbf{F}_{2D}\hat{\mathbf{A}}\mathbf{p}^c \right| \right\|_2^2. \tag{18}$$

Additionally, the gradient update for the t-th iteration is similar to (8), where there is

$$\mathbf{p}^t = \mathbf{p}^{t-1} - \mu \nabla \mathcal{A}\left(\mathbf{p}^{t-1}\right). \tag{19}$$

In this case, the learning rate μ is calculated by using the spectral norm of the estimated matrix $\mathbf{F}_{2D}\hat{\mathbf{A}}$, i.e. $\|\mathbf{F}_{2D}\hat{\mathbf{A}}\|$.

IV. Computational Complexity and Implementation

In this section we discuss the computational complexity and shortcuts, which could be implemented in order to speed up the proposed algorithms. The naïve implementation via matrix vectorisation as well as the construction of a Fourier matrix is straightforward. However the experimental setting requires to process huge data and, thus, a fast and efficient implementation is required.

A. Layer-Wise Optimisation

First, we consider the implementation of Layer-wise Optimisation summarized in Algorithm 1. The algorithm is presented via construction of the measurement matrix $\mathbf{Q}_\ell^t$ given by (13). The gradient step (8) only requires to compute the multiplication with $\mathbf{Q}_\ell^t$ and its conjugate transpose $(\mathbf{Q}_\ell^t)^H$, which can be done efficiently on the basis of Fast Fourier transform.

A multiplication of the prefix and suffix matrices $\mathbf{S}_\ell^t$ and $\mathbf{R}_\ell^t$ given by (11) with an arbitrary vector in $\mathbb{C}^{N^2}$ consists of the two repeated operations: $\mathbf{O}_m v$ and $\mathbf{G}_m v$ for some $v \in \mathbb{C}^{N^2}$. As $\mathbf{O}_m$ is a diagonal matrix, only N^2 operations are required. $\mathbf{G}_m$ is a product of two Fourier matrices and a diagonal matrix. Hence, when Fast Fourier transform is used, the second product requires $\mathcal{O}(N^2 \log N)$ operations. Therefore, a multiplication with $\mathbf{S}_\ell^t$ and $\mathbf{R}_\ell^t$ can be performed in $\mathcal{O}((\ell-1)N^2 \log N)$ and $\mathcal{O}((M-\ell)N^2 \log N)$ operations, respectively. Then, a multiplication with a single row block of $\mathbf{Q}_\ell^t$ is done in $\mathcal{O}(MN^2 \log N)$ and for a multiplication with the whole matrix $\mathbf{Q}_\ell^t$ $\mathcal{O}(SMN^2 \log N)$ operations are needed. The same argument applies for a multiplication with the conjugate transpose of $\mathbf{Q}_\ell^t$.

In addition to the computation of the gradient, also the step size given by $\|\mathbf{Q}_\ell^t\|^{-2}$ needs to be evaluated. The spectral norm can be accurately computed by performing J iterations of the power method, which are given by

$$v_{j+1}^{t,\ell} = \frac{(\mathbf{Q}_\ell^t)\, \mathbf{Q}_\ell^t v_j^{t,\ell}}{\| (\mathbf{Q}_\ell^t)\, \mathbf{Q}_\ell^t v_j^{t,\ell}\|_2}.$$

Then, $\|\mathbf{Q}_\ell^t v_j^{t,\ell}\|_2$ is a good estimate of the spectral norm of $\mathbf{Q}_\ell^t$. Again, each iteration is a product of $v_j^{t,\ell}$ with $\mathbf{Q}_\ell^t$ and its conjugate and, thus, the power method requires $\mathcal{O}(JSMN^2 \log N)$ operations. For a faster convergence $v_0^{t+1,\ell} = v_J^{t,\ell}$ is initialised as the top eigenvectors of $\mathbf{Q}_\ell^{t+1}$ and $\mathbf{Q}_\ell^t$ are not expected to differ much, especially for later iterations. It was shown that in Algorithm 1 for each iteration and each slice the spectral norm is estimated and then K gradient steps are performed, which grants the total complexity of

$$\mathcal{O}\left(T(K+J)SM^2\, N^2 \log N\right).$$

B. Sparse Matrix Decomposition

Algorithm 2 consists of two steps, the estimation of $\mathbf{A}_t$ and its decomposition. In the first step, $\mathbf{A}_t$ is treated as an $N^2 \times N^2$ matrix without any additional information about its structure. Hence, for K_1 gradient steps for (14) $\mathcal{O}(K_1(N^2 S \log(N^2 S) + N^4 S)$. operations are performed. The step size stays the same for all $t \in [T]$ and can be computed via the power method for $\mathbf{P}$, which is $\mathcal{O}(J_1\, N^2 S)$ operations for some J_1 denoting the number of power method iterations.

Turning to the decomposition step, the update (16) is investigated at first. The matrix $\mathbf{Z} = \lambda^t \mathbf{R}_\ell^t \mathbf{O}_\ell \mathbf{S}_\ell^t$ has to be computed. This can be done by evaluating its columns as $\mathbf{Z}\mathbf{e}_1$, …, $\mathbf{Z}\mathbf{e}_{N^2}$, where $\{\mathbf{e}_j,\ j \in [N^2]\}$ are standard basis vectors. This can be done with $\mathcal{O}(MN^4 \log N)$ operations. Furthermore, the difference with $\hat{\mathbf{A}}$ gives additional $\mathcal{O}(N^4)$ operations. Next, the product with conjugate transposes has complexity of $\mathcal{O}(MN^4 \log N)$, which implies that the total complexity of a single iteration is given by $\mathcal{O}(MN^4 \log N)$. The step size once again determined by performing J_2 iterations of power method for both $\mathbf{R}_\ell^t$ and $\mathbf{S}_\ell^t$, which requires additional $\mathcal{O}(J_2\, MN^2 \log N)$ operations. Therefore, the estimation of slices by performing update (16) K_2 times for each slice has a complexity of $\mathcal{O}((K_2\, N^4 + J_2\, N^2) M^2 \log N)$.

At last, the update for λ^{t+1} can be computed in $\mathcal{O}(MN^4 \log N)$ as $\hat{\mathbf{A}}$ is again a product of diagonal and Fourier matrices. Combining everything together results in a total complexity of

$$\mathcal{O}\left(J_1\, N^2 S + \left[K_1\, N^2 S \log\left(N^2 S\right) + N^4 S\right] T \right. \\ \left. +\ \left(K_2\, N^4 + J_2\, N^2\right) M^2\, T \log N\right)$$

for Algorithm 2.

For numerical trials below, two implementations were tested. The first is based on the Fast Fourier transform as described above. The second uses matrix multiplication in the decomposition step, specifically, sparse matrix products to improve the numerical computation by using the fact that each slice object $\mathbf{O}_l$ is a diagonal matrix, and randomized Singular Value Decomposition (SVD) for computation of the eigenvalues. The latter turned out to be faster, which may be caused either by the large constants hidden in $\mathcal{O}$ notation or a small N.

C. Comparison to 3PIE

The computational complexity of 3PIE [23] is given by

$$\mathcal{O}\left(TKSMN^2 \log N\right),$$

where K denotes the number of iterations. This is M times smaller than the complexity of Layer-Wise optimisation as it updates all slices simultaneously. The use of the full matrix $\mathbf{A}$ in Sparse Decomposition leads to notably larger computation time.

V. Simulation and Experimental Details

In this section, information regarding the simulated dataset of a specimen used as the ground truth, including its type and crystal structure, is given. Furthermore, the microscope and a

TABLE I
PARAMETERS FOR GENERATING SIMULATED DATA SETS TAKEN FROM GAAS [38], MoS_2 [39], $SrTiO_3$ [40]

Parameters	**GaAs**	**$SrTiO_3$**	**MoS_2**
Unitcell (a,b,c) (nm)	$(0.56533, 0.56533, 0.56533)$	$(0.3905, 0.3905, 0.3905)$	$(0.3161, 0.54750, 1.2295)$
Supercell (Na,Nb)	$(2, 2)$	$(2, 2)$	$(2, 2)$
Semiconv. angle (mrad)	32	32	32
Accel. voltage (keV)	$80, 200$	$80, 200$	$80, 200$
Scan and detector size	$(40, 40, 40, 40)$	$(40, 40, 40, 40)$	$(40, 40, 40, 40)$
Fresnel distance monolayer/3 slices (nm)	$(0.1413, 0.1413, 0.1413)$	$(0.1952, 0.1952, 0.1952)$	$(0.1561, 0.1561, 0.1561)$

In the simulated data, the hexagonal unit cell was transformed into an orthogonal cell.

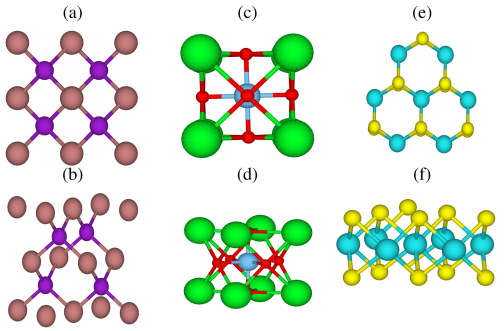


Fig. 2. Structure of materials. (a) 2D projection of GaAs. (b) 3D structure of GaAs. (c) 2D projection of $SrTiO_3$. (d) 3D structure of $SrTiO_3$. (e) 2D projection of MoS_2. (f) 3D structure of MoS_2. Structural parameters were taken from GaAs [38], MoS_2 [39], $SrTiO_3$ [40]. In the simulated data, the hexagonal cell was transformed into an orthogonal cell.

description of the experimental conditions used for obtaining actual experimental diffraction data are provided.

A. Simulated Data Sets

Intensities of simulated diffraction patterns from Molybdenum disulfide (MoS_2), Strontium titanate ($SrTiO_3$), and Gallium arsenide (GaAs) specimens with elevated thicknesses were generated by using a forward multislice algorithm. In Fig. 2, their 3D structural representations as well as 2D projections along [0 0 1] of the unit cells are shown. The structural as well as simulation parameters are given in Table I. In addition, the parameter unit cell presents the most simple repeated lattice point in the crystal. The collection of several unit cells is called a supercell. At last, the semi-convergence angle represents the semi-angle that appears in a cone shape when a convergent electron beam illuminates a specimen.

B. Experimental Dataset

Besides simulated datasets, numerical evaluations of experimental datasets are also performed. From a bulk crystal of 2H-MoS_2, sheets were exfoliated by using a poly-dimethylsiloxane elastomeric film supported on a glass slide and transferred onto a holey silicon nitride membrane for the use in transmission electron microscopy (TEM). Experimental data of MoS_2 was acquired using a probe corrected Hitachi HF5000 field emission microscope in STEM mode and with an acceleration voltage of 200 keV as well as a beam current of about 7.4 pA. Intensities of diffraction patterns were recorded by using a Medipix3 Merlin4EM camera with 256×256 pixels. The distance between neighbouring scan points was 26.5 pm in x, i.e. horizontal, and y, i.e. vertical, scanning directions. In addition, the acquisition time per diffraction pattern was 0.5 ms and data was acquired using a dynamic range of 6 b. The position-averaged convergent beam electron diffraction pattern (PACBED) is depicted in Fig. 3, where the intensity of all diffraction patterns from 128×128 scanning points is averaged.

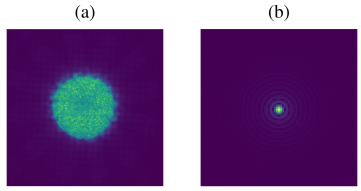


Fig. 3. (a) PACBED of an experimental data set of MoS_2 acquired using a Hitachi HF5000 microscope in STEM mode with a Medipix3 Merlin camera. (b) Amplitude of the probe initialisation, or the so-called Airy disk, which is generated by taking the absolute value of the two-dimensional inverse Fourier transform of the circular aperture generated from the PACBED.

VI. NUMERICAL RESULTS

Several numerical evaluations that measure performance of the proposed algorithms as well as comparisons to the 3PIE algorithm [23] in terms of reconstruction quality and computation time are presented in this section. In contrast to the proposed algorithms, there is no theoretical choice of the step size parameter α for the 3PIE algorithm. Hence, three different step size parameters $\alpha \in \{0.1, 0.5, 1.0\}$, where $\alpha = 1.0$ comes from the original implementation [23], were used.

For all algorithms, the object slices $\mathbf{O}_m^0$ were initialised with an identity matrices. The probe was initialised by constructing a binary circular aperture from the aggregated diffraction patterns and appling the inverse Fourier transform.

Throughout this section the abbreviation GT for ground truth, LW for Layer-wise Optimisation, SD for Sparse Matrix Decomposition, and 3PIE_α for step size $\alpha \in \{0.1, 0.5, 1.0\}$ are used.

A. Error Metrics for Evaluation of the Algorithm

The error metric used for evaluating the reconstruction of each slice is calculated as the mean square error

$$\frac{1}{M}\sum_{m=1}^{M}\frac{\left\|\mathbf{O}_m - \gamma_m\hat{\mathbf{O}}_m\right\|_F^2}{\|\mathbf{O}_m\|_F^2}, \tag{20}$$

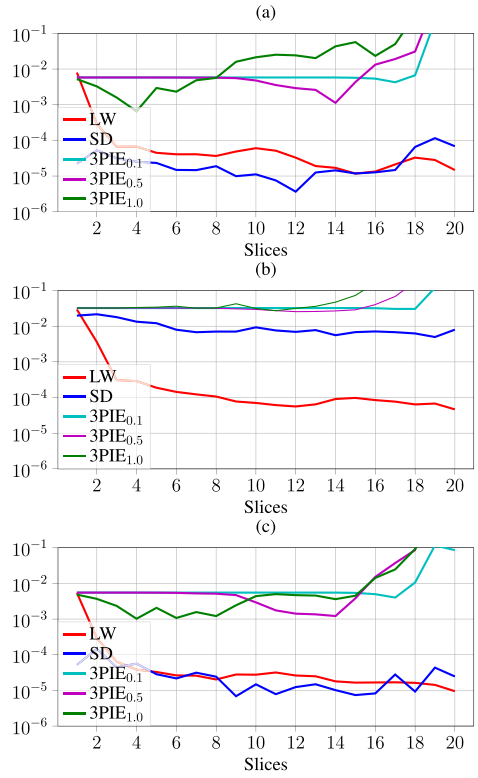


Fig. 4. Per slice relative reconstruction error of 20 slices with distance 1 nm for specimen: (a) GaAs. (b) MoS_2. (c) $SrTiO_3$.

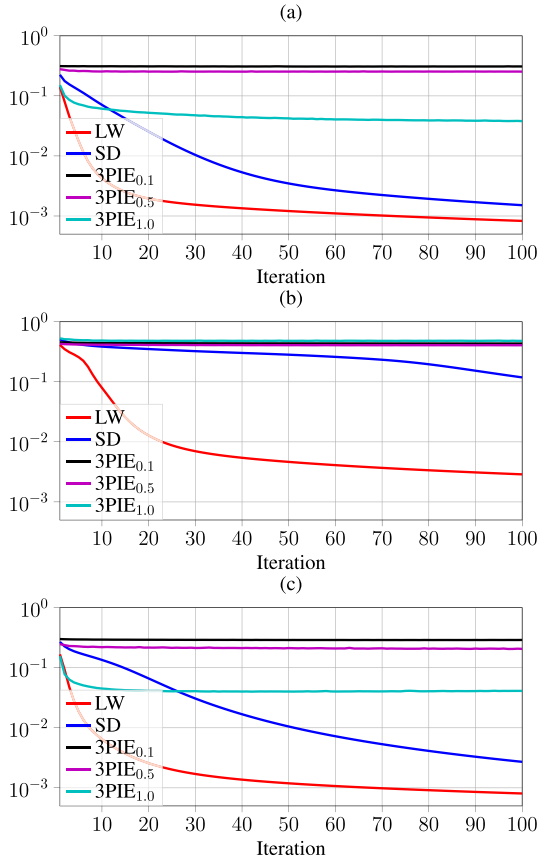


Fig. 5. Relative measurement error as in (21) of 20 slices with distance 1 nm for specimen: (a) GaAs. (b) MoS_2. (c) $SrTiO_3$.

with $\mathbf{O}_m$ and $\hat{\mathbf{O}}_m$ being the ground truth and the estimated object at the m-th slice, respectively. The global phase factors γ_m are given by $\mathrm{Tr}(\hat{\mathbf{O}}_m^H \mathbf{O}_m)/|\mathrm{Tr}(\hat{\mathbf{O}}_m^H \mathbf{O}_m)|$.

The objective of the presented optimisation problem is to minimise the error between measured and estimated intensities of diffraction patterns. Accordingly, it is necessary to introduce an additional error metric

$$\frac{\left\| \sqrt{\mathbf{I}} - \left| \mathbf{F}_{2D} \hat{\mathbf{A}}_M \mathbf{P} \right| \right\|_F}{\left\| \sqrt{\mathbf{I}} \right\|_F}, \tag{21}$$

with $\hat{\mathbf{A}}_M$ being the total estimated object transfer function at slice M. In (20) the error metric is referred to as the relative reconstruction error whereas, the error metric in (21) is referred to as the relative measurement error.

B. Reconstruction of Arbitrary Slice Thickness

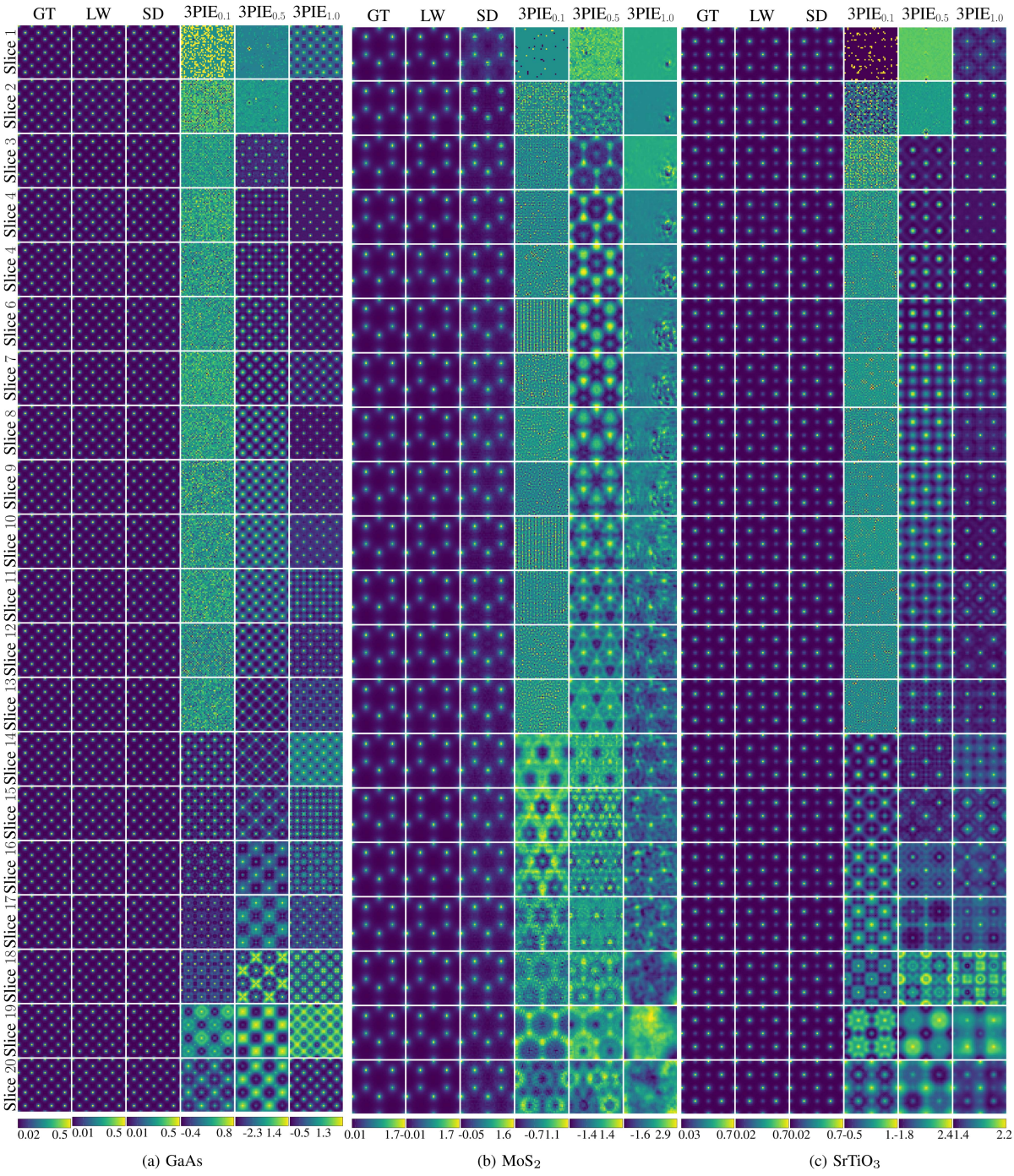
In the first experiment, recovery of accumulated atom positions of a phase grating were considered. Intensities of diffraction patterns of GaAs, MoS_2 and $SrTiO_3$ were simulated from phase gratings of 20 nm thickness using the forward multislice model (5). The simulation parameters can be found in Table I. For the reconstruction, it was assumed that as in experimental setup the precise distances are unknown and the crystals are decomposed into slices separated by a 1 nm Fresnel propagation distance. Thus, one can heuristically approximate the correct reconstruction by using an initial, comparably large slice thickness and evaluating the atomic positions.

Fig. 5 shows the relative measurement error (21) of the proposed algorithms and 3PIE for different step size parameters α. It turned out that the choice of α has a strong impact on the convergence of the relative measurement error as presented in Fig. 5. For different specimens with 20 slices, the 3PIE algorithm converged slower compared to our proposed algorithms. Regarding the relative measurement error, the layer-wise optimisation converged faster than the sparse matrix decomposition and 3PIE.

However, the relative measurement error only describes the characteristic of an algorithm with respect to the measurement data. Since in general the correctness of the reconstructed object should be measured with respect to the object ground truth, also the relative reconstruction error (20) was evaluated. Fig. 4 shows that the sparse matrix decomposition shows lower relative reconstruction error (20) on datasets, of GaAs and $SrTiO_3$, while for MoS_2 the Layer-wise Optimisation outperforms other algorithms. The individual relative reconstruction error for each slice is presented to observe the deviation of the reconstructed object from the ground truth. The average can be observed in Table II, where Sparse Matrix Decomposition performed better in terms of relative reconstruction error than other algorithms.

The complete phase reconstructions are depicted in Fig. 6. In general, the 3PIE algorithm shows unstable reconstruction for 20 slices for all chosen step sizes α.

The experiment was repeated for specimens with only 5 slices and the distance between slices increased to 4 nm. The obtained



(a) GaAs (b) MoS_2 (c) $SrTiO_3$

Fig. 6. Phase reconstruction in radian of 20 slices for arbitrary slice thickness with Fresnel propagation distance 1 nm. The acceleration voltage of the datasets is 200 keV and the reconstruction is observed at 100 iterations.

reconstructions are shown in Fig. 7. In contrast to Fig. 6, 3PIE algorithm was able to perform better and the phase reconstruction for $\alpha = 1$ showed the best reconstruction among three choices of α. These observation suggested, that the proposed algorithms are more robust in terms of larger number of slices than 3PIE.

C. Atomic Plane Decomposition

The reconstruction of each atomic plane, i.e., direct reconstruction of each slice of the crystals as in Fig. 2, was also examined. This approach proved to be quite challenging due to arising ambiguities concerning the atom positions along the direction of the electron beam. These ambiguities can be observed in Fig. 8 by comparing the reconstructed slice for one of the algorithms and the ground truth column. Atom positions were combined and a decomposition of the layers was not successful.

1) Increasing the Fresnel Propagation Distance: A huge advantage of conducting a conceptual study is the possibility to set the Fresnel propagation distance in the forward multislice model to any desired value within the simulation in order to study

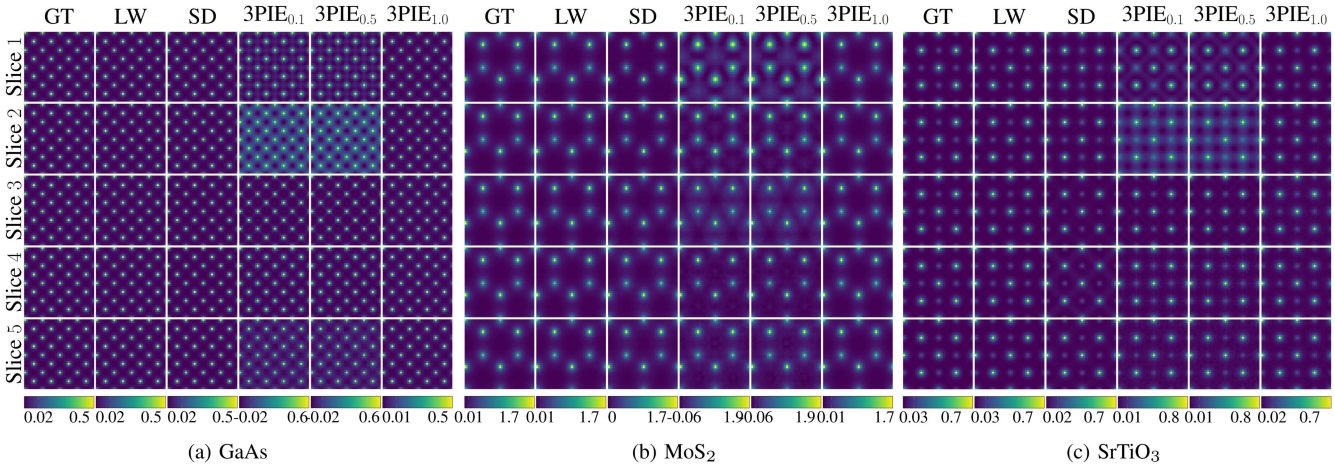


Fig. 7. Phase reconstruction in radian of 5 slices for arbitrary slice thick with Fresnel propagation distance 4 nm. The acceleration voltage of the datasets is 200 keV and the reconstruction is observed at 100 iterations.

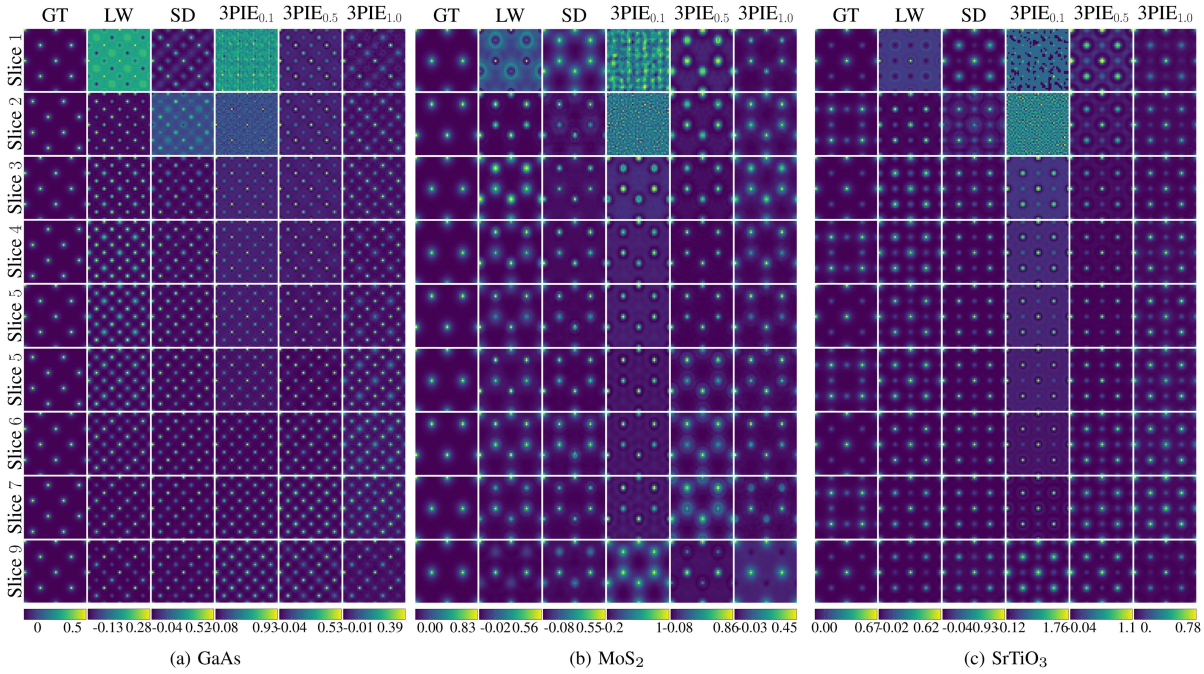


Fig. 8. Phase reconstruction in radian of 9 slices for atomic plane decomposition with given Fresnel propagation distance as in Table I. The acceleration voltage of the datasets is 200 keV and the reconstruction is observed at 100 iterations.

TABLE II
RELATIVE RECONSTRUCTION ERROR AS IN (20) OF THE ALGORITHMS FOR DIFFERENT SPECIMENS WITH 20 SLICES

Algorithms	**GaAs**	**MoS_2**	**$SrTiO_3$**
Layer-wise	4.4×10^{-4}	1.7×10^{-3}	3.1×10^{-4}
Sparse decomp.	2.7×10^{-5}	9.5×10^{-3}	2.95×10^{-5}
3PIE ($\alpha = 0.1$)	5.3×10^{-2}	4.2×10^{-2}	1.5×10^{-2}
3PIE ($\alpha = 0.5$)	6.1×10^{-1}	79.8	1.3
3PIE ($\alpha = 1.0$)	1×10^{4}	9.2×10^{8}	2×10^{1}

the performance of the reconstruction algorithms. In the next experiment, the Fresnel propagation distance was deliberately increased for both the forward and the inverse processes.

The relative reconstruction errors (20) for all algorithms are shown in Fig. 10, where the first index represents the original Fresnel propagation distance as listed in Table I. It is notable that both Layer-wise Optimisation and Sparse Decomposition benefited from a larger Fresnel distance. Fig. 9 shows phase reconstruction at the Fresnel propagation distance of 4 nm. Compared to the original Fresnel propagation distance, the decomposition of each atomic plane was improved. Due to the fact that with the increase of the distance, the Fresnel matrix is less similar to the identity what consequently prevented ambiguities related to the atomic positions for different atomic planes to appear.

Increasing the Fresnel propagation distance would correspond to artificially increasing the lattice parameter in electron beam

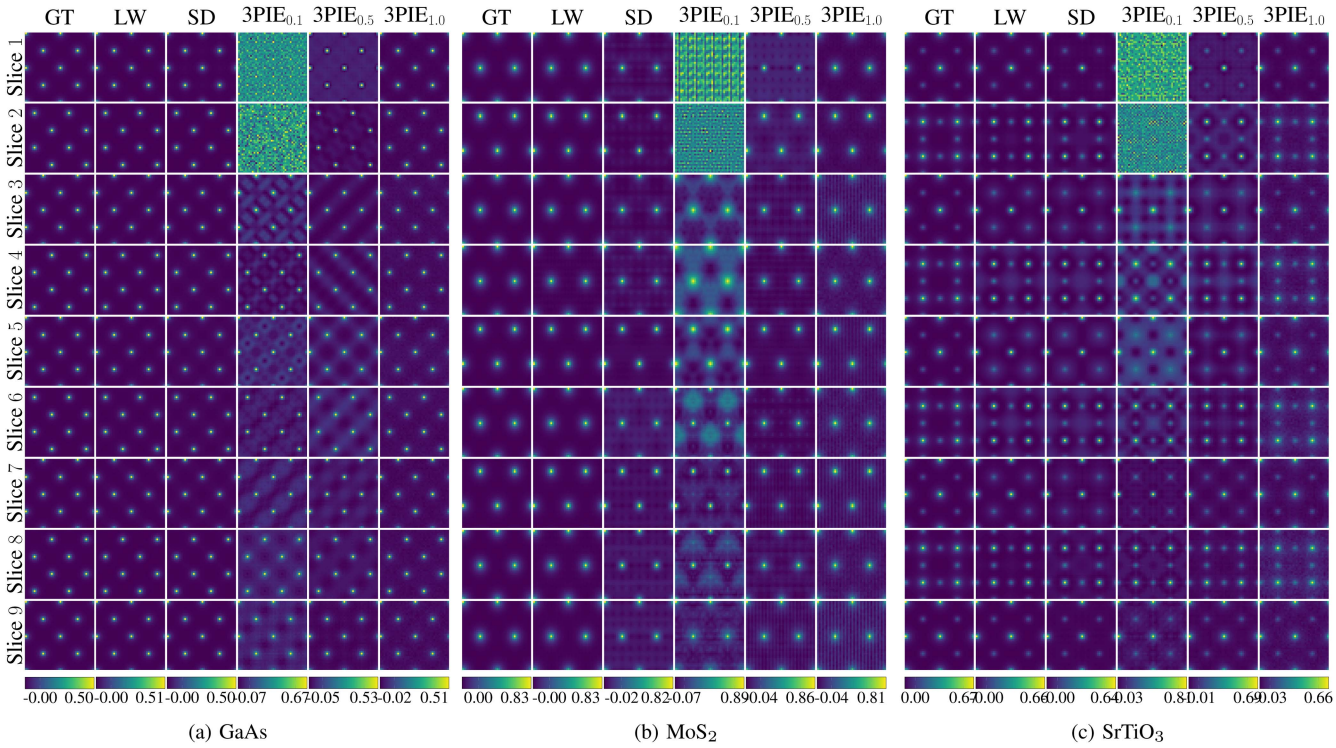


Fig. 9. Phase reconstruction in radian of 9 slices for atomic plane decomposition with increasing Fresnel propagation distance 4 nm. The acceleration voltage of the datasets is 200 keV and the reconstruction is observed at 100 iterations.

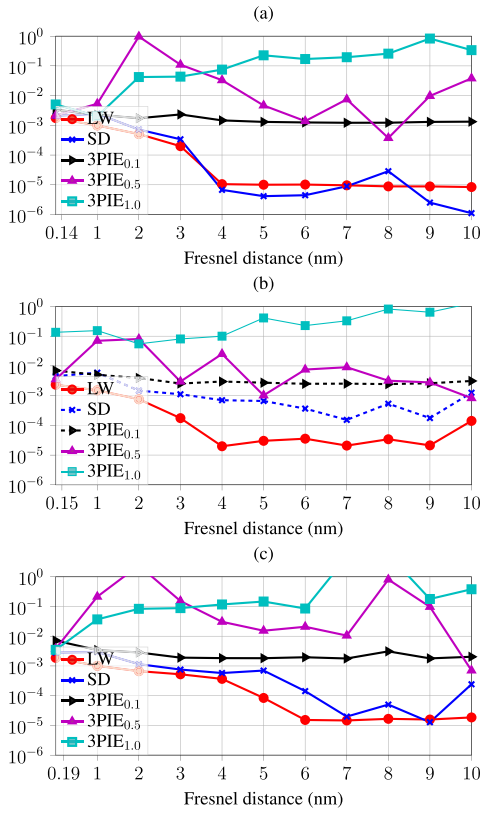
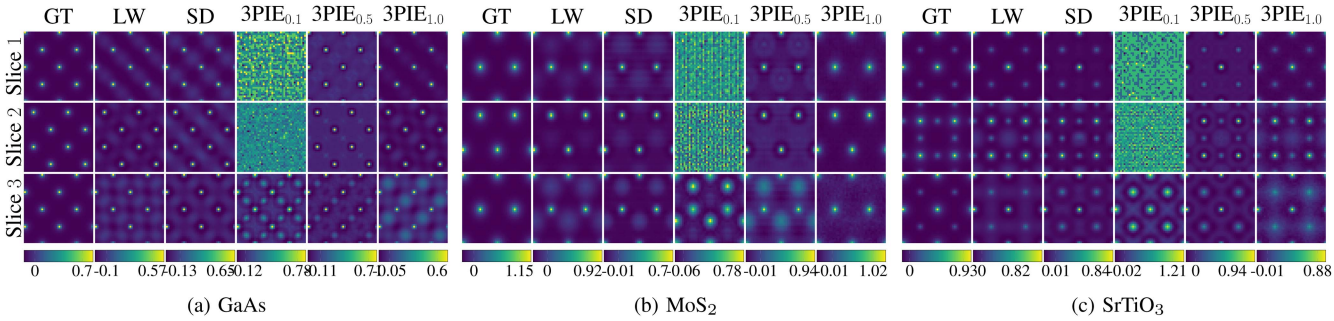


Fig. 10. Relative reconstruction error of atomic plane decomposition with 9 slices of specimens (a) GaAs. (b) MoS_2. (c) $SrTiO_3$. The error is observed at 100 iterations.

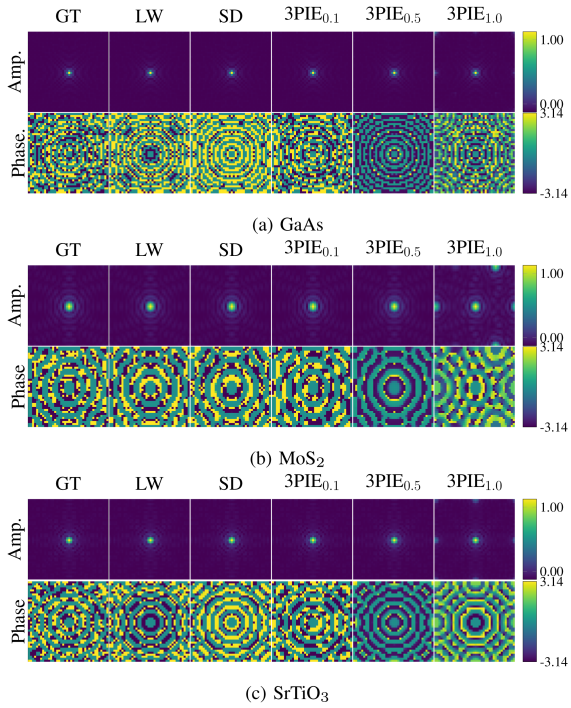
direction. This is only possible in a simulation study. However, the conceptual insight is that the phase fronts of electron waves with 200 keV energy would only change significantly after propagating 4 nm for the algorithms to separate the Fresnel propagation from the interaction with the Coulomb potential of the slices.

2) Low Electron Energy: The Fresnel propagator on (1) contains the product of the wavelength and the propagation distance as the governing parameters. Therefore, a realistic reconstruction with atomic layer sensitivity needs to use larger wavelengths to compansate for atomic spacings in the range of 0.1 nm Furthermore, the same potential in a given specimen leads to a larger phase change of low-energy electrons as compared to high energies due to the effect of the so-called interaction constant on the phase grating. The relation between electron acceleration voltage U and wavelength λ is given by $\lambda = \frac{hc}{\sqrt{e^2U^2+2eUmc^2}}$, with c being the speed of light, h the Planck constant, m the electron mass, and e the elementary charge. For 200 keV electrons, the wavelength is approximately 2.5 pm, whereas it increases to 4.18 pm for electrons with 80 keV energy. The interaction constant increases by about 40%. Note that both acceleration voltages, 200 and 80 kV, are common settings in STEM such that atomic resolution can be obtained readily in aberration corrected machines. Fig. 11 shows a reconstruction of the atomic planes using an acceleration voltage of 80 keV. Despite the low-signal artefacts related to the location of the atoms in the different slices, the exact location of the atoms in each slice can now be correctly determined. However, for 3PIE algorithm the step size α is a crucial factor to improve the reconstruction. In this



(a) GaAs (b) MoS_2 (c) $SrTiO_3$

Fig. 11. Phase reconstruction in radian of 3 slices for atomic plane decomposition with acceleration voltage 80 keV. The reconstruction is observed at 2000 iterations.



(a) GaAs (b) MoS_2 (c) $SrTiO_3$

Fig. 12. Probe reconstruction after estimating 9 slices of atomic plane decomposition as in Fig. 8. The amplitude is normalized and the phase is in radian. The reconstruction is observed at 100 iterations.

case, the $\alpha = 1$ improve the reconstruction compared to $\alpha = 0.1$ and $\alpha = 0.5$. After estimating the object, it is also possible to reconstruct the probe as discussed in the following section.

D. Probe Reconstruction

The reconstruction of the probe using the optimisation problem in (18) can be performed after the estimation of the matrix $\mathbf{A}_M$. As discussed in the forward multislice model, the estimated matrix $\mathbf{A}_M$ can be generated by calculating the product of each slice of the object $\mathbf{O}_m$ and the Fresnel matrix $\mathbf{G}_m$, $\mathbf{A}_M = \mathbf{O}_M \prod_{m=1}^{M-1} \mathbf{G}_m \mathbf{O}_m$. In Fig. 12, the reconstructed probes observed at 100 iterations are shown after estimating the slices in Fig. 8. It can be seen that for 3PIE algorithm for specimen MoS_2 at $\alpha = 1.0$, we have artifacts on the amplitude since the Airy disk appears not only at the center but also at different positions on the left and top right.

E. Experimental Data

Numerical evaluations on diffraction intensity measurements acquired experimentally from a MoS_2 specimen are presented below. The thickness of the experimental specimen was determined by comparison with simulated position-averaged diffraction patterns to be approximately 35 nm. Further experimental details are described in Section V.

Applying the same algorithms to experimental instead of simulated data is a crucial aspect to demonstrate the practical usefulness of the methods.

Real data is affected by additional parameters that are difficult or even impossible to include in simulations and in the algorithmic setups above. For example, the recording inherently contains Poissonian counting noise, the camera has a modulation transfer function which leads to a blurring of diffraction space features, and the projection system of the microscope can cause geometrical distortions of the diffraction patterns. Furthermore, the scan positions of the STEM probe usually deviate slightly from the ideal regular raster due to instabilities of the scan engine. Hence, reconstructing based on experimental data is a crucial checkpoint in the evaluation of the proposed algorithms.

Due to this a detailed analysis of the performance of layer-wise optimisation and sparse matrix decomposition algorithms in dependence of the experimental conditions were set aside for a future task. Instead this preliminary evaluation focuses on demonstrating the principal applicability by targeting the qualitative reconstruction of the MoS_2 structure using a relatively small number of five slices, similar to the example in Fig. 7. In particular, this was necessary due to computational efficiency and the much higher dimensionality of the experimental data as compared to the simulations, i.e., the large number of probe positions and camera pixels.

1) Slice Reconstruction: Consequently, the phases of the individual slice reconstructions in Fig. 13 were not expected to quantitatively represent the actual phase gratings on the one hand. On the other hand, they were supposed to resemble the atomic structure of the specimen in the respective slices, taking a large portion of the dynamical scattering into account.

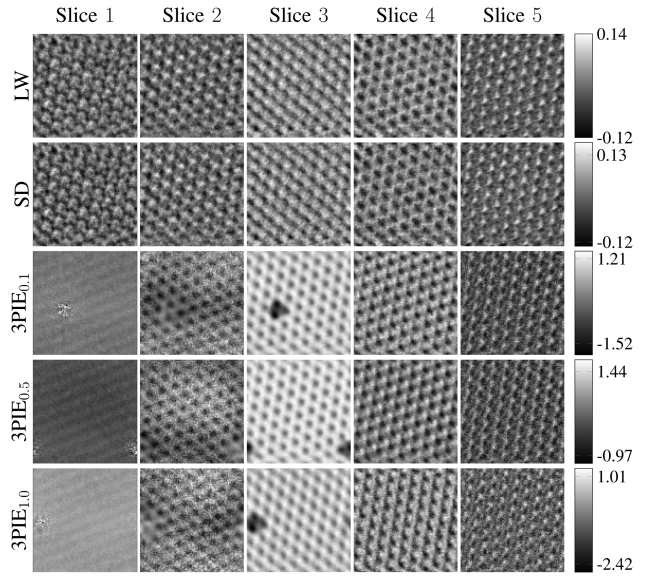


Fig. 13. Phase reconstruction in radian of 5 slices from experimental data MoS_2 with Fresnel propagation distance 7.377 nm. The reconstruction is observed after 50 iterations.

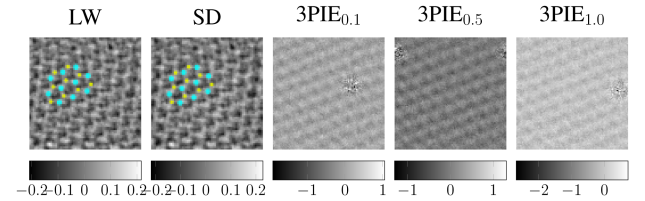


Fig. 14. Projection of the phase reconstruction, in radian, of MoS_2.

Indeed, the atomic structure is consistently visible in all slices, opposite to single-slice models for which evaluations at thicknesses of tens of nanometers are by far out of range. The dynamic range of the phase is comparably low, most probably because more slices would be needed to disentangle the slice potentials and Fresnel propagation between the slices completely. Fig. 13 shows the reconstructions for all considered algorithms. In contrast to Layer-wise optimization and Sparse Matrix Decomposition, the 3PIE algorithms struggle to reconstruct slices 1 to 3.

Furthermore, the slices were combined into the object transfer function matrix $\mathbf{A}$ and its two-dimensional projection represening all atoms at once is presented in Fig. 14. A coloured overlay of the Molybdenum (Mo) and Sulfur (S) atoms were added to the figure in order to better visualise the reconstructed atomic arrangement.

2) Probe Reconstruction: A direct implementation of the Amplitude Flow as in (7) was adopted in order to reconstruct the illuminating probe after estimating the object transfer function matrix $\mathbf{A}$. At this setting the focus was solely on the intensity of the diffraction patterns acquired at the center position of the illuminated area on the specimen. The resulting reconstructions of amplitude and phase of the probe are presented in Fig. 15. It shows that the data was taken with a well-focused probe as indicated by the sharp peak in the amplitude and a flat phase except for the noise. The reconstruction of the probe is in general

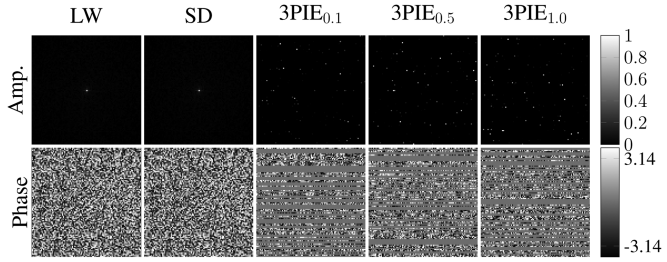


Fig. 15. Probe reconstruction after estimating the specimen's slice of experimental data MoS_2. The reconstruction is observed after 50 iterations.

a robust check whether the algorithm and the parameters used for the reconstruction were suitable to separate illumination and specimen. In the presented case it can be concluded that the large slice thickness did not affect this, because no specimen details are visible in the reconstructed probe of both, reconstructed data from sparse matrix decomposition and layer-wise optimisation.

F. Numerical Computations

Besides theoretical computation complexity, the numerical computation time for all algorithms was evaluated. Since the parameters that affect the computation are the number of slices and dimension of the datasets, numerical computation for both aspects are presented. For all algorithms the evaluation was performed individually on the same workstation. It was equipped with AMD EPYC 7543P with 32 CPU (64 threads) operating at a base frequency of 2.80 GHz, and 512 GB DDR4 RAM with operating frequency of 3.2 GHz. It should be noted that we performed the simulation for each algorithm without noise background, i.e., no other processes or algorithms were running during the evaluation. Fig. 16 shows the wall time of all algorithms for increasing dimension and number of slices. In general, 3PIE requires smaller computation time per iteration compared to the proposed algorithms.

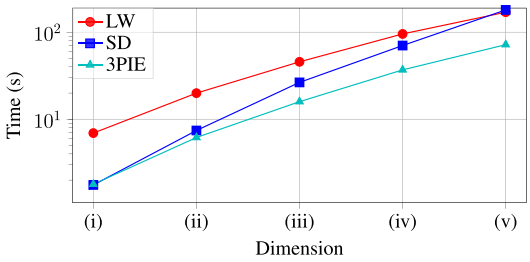
For Sparse Matrix Decomposition the practical implementation, as discussed in Section IV, can significantly improve the computation time, as for both scenarios the Sparse Matrix Decomposition required less computation time than the Layer-wise Optimisation. While the computation complexity, as discussed in Section IV, showed that Sparse Matrix Decomposition has the worst complexity among the algorithms presented in this article, it could be optimised for fast and efficient implementation by reshaping the product of $\mathbf{AP} \in \mathbb{C}^{N^2 \times S}$ into a stack of two-dimensional and incorporating the fast Fourier transform. Furthermore, if $\mathbf{P}_s$ is given by shifts of a probe $\mathbf{P}$, that is $\mathbf{P}_s = \mathbf{S}_s\mathbf{P}$ with a shift operator $\mathbf{S}_s$, then the computational speed of the product $\mathbf{AP}$ was increased. For a k-th row of $\mathbf{AP}$ the identity

$$(\mathbf{AP})_{(k)} = N^2\mathbf{F}_{2D}\left[\mathbf{F}_{2D}^{-1}\mathbf{A}_{(k)} \circ \mathbf{F}_{2D}^{-1}\mathbf{RP}\right]$$

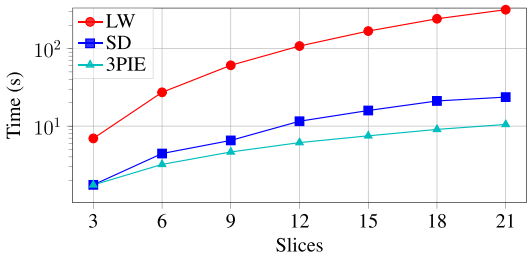
holds, where $\mathbf{R}$ is a coordinate reversal. This would lead to the computational complexity of

$$\mathcal{O}\left(N^4 \log N^2 + SN^2\right)$$

operations for the product $\mathbf{AP}$ instead of $\mathcal{O}(N^4 S)$.



(a) Average computation time of increasing dimension of four-dimensional datasets per iteration for 3 slices. The evaluation is performed at 100 iterations. (i). $(40, 40, 40, 40)$, (ii). $(60, 60, 60, 60)$, (iii) $(80, 80, 80, 80)$, (iv). $(100, 100, 100, 100)$, (v). $(120, 120, 120, 120)$.



(b) Average computation time of increasing the number of slices per iteration for dimension $(40, 40, 40, 40)$. The evaluation is performed at 100 iterations.

Fig. 16. CPU Wall Time of all algorithms both for increasing dimension and number of slices.

When the number of slices was increased, the computation time of Layer-wise Optimisation deteriorated in contrast to 3PIE, which is explained by the slice-by-slice optimisation approach. That Layer-wise Optimisation can be further accelerated by a parallelized computation of the gradient, which was not used in the conducted numerical experiments.

VII. Discussion

Numerical observations showed that the proposed algorithms performed robust reconstruction in terms of increasing number of slices for arbitrary slice thickness compared to the state of the art algorithm 3PIE. Additionally, it was shown that for experimental data sets, the proposed algorithms provided better reconstruction than the 3PIE algorithm. However, one disadvantage of the proposed methods is that the computation time required to perform the evaluation is larger than for the state of the art algorithm. Especially for Layer-wise Optimisation when a larger number of slices is to be estimated. The possible improvement for both proposed algorithms should be investigated further in a future study.

Theoretically it would be interesting to examine the fundamental limit of the algorithms with respect to the number of slices needed to disentangle interaction and Fresnel propagation sufficiently well. Moreover, a systematic study addressing the impact of the probe semi-convergence angle, the electron energy, aberrations of the electron-optical system and coherence effects could shed light on the robustness of the presented methodology in respect to the multitude of experimental parameters in real measurements. If one wants to work with low dose data, where Poisson noise is dominant, one has to modify the objective function. Poisson maximum likelihood has been used with success in case of single slice ptychography [41] and can be adapted to our layer-wise estimation. For the sparse matrix decomposition, one can even use a Poisson phase retrieval method [42] in the first step, estimating $\mathbf{A}$, without changing the decomposition method at all.

Since the reformulation of the forward multislice model in this article yielded purely a matrix representing the transfer function of a thick specimen, it would be possible to relate such a matrix to a scattering matrix constructed from the Bloch wave method and observe the differences between both approaches. The latter would require an intensive computational effort of eigenvalue decomposition for huge scattering matrices.

It should be possible to estimate the specimen thickness directly from the algorithms. One possibility could be to incorporate information from the high-angle intensity of diffraction patterns, or to start from a coarse slicing first with large slice thicknesses, and then increase the number of slices subsequently. In case that a sufficiently high total thickness is assumed, empty slices should emerge, indicating that the specimen is actually compact along the electron beam direction. In general, a suitable regularisation should be developed and applied in future works to improve the reconstruction as well as to optimally estimate the Fresnel propagation distance. An interesting approach is a suitable sparsity model, as applied in [43] for the case of single slice ptychography.

Finally, the reformulation of the multislice scheme as a simple, though large, one-step matrix multiplication can be used to easily characterise the specimen transfer function. Hence, investigating the effect of different probe settings could be done by a matrix multiplication without re-running the propagation from the first slice. In that respect, studying the capabilities and performance of the reformulation is not only relevant for solving inverse problems, but also interesting with respect to conventional forward simulations.

VIII. Conclusion and Summary

We proposed reformulation of the forward multislice method such that the transfer function of a thick specimen can be directly determined. In combination with the ptychographic approach we presented two optimisation models for solving the inverse multislice ptychography problem for both arbitrary thickness and atomic plane decomposition. In the first case, given the intensity of diffraction patterns, several atomic planes were jointly processed into a single reconstruction in order to show the total potential. In the atomic plane decomposition each slice was reconstructed at its atomic plane and given only the intensity of the diffraction patterns the results showed the unique atomic positions in each slice.

Our observations showed that the proposed algorithms outperform the 3PIE algorithm for a larger number of slices in the arbitrary slice thickness scenario as well as for experimental data sets. In contrast to the 3PIE algorithm that highly depends on the selection of an optimal step size, our proposed method can automatically determine the step size, in terms of the

maximum singular value. In the atomic plane decomposition, since we have small Fresnel propagation distance, it is in general difficult to uniquely decompose the atom positions for each slice. Nevertheless, some potential approaches such as using data acquired at a lower acceleration voltage, namely 80 keV, can improve the reconstruction.

We also supported our numerical observations with the reconstruction of the object given the intensity of diffraction patterns acquired from the experimental data set of MoS_2. It was shown that both algorithms can reconstruct the phase of the specimen. Additionally, by using the reformulation of the forward multislice method, a matrix was constructed that represented the thick object transfer function, i.e. scattering matrix. This reformulation can be used to directly generate the two-dimensional projection of the atomic arrangement.

REFERENCES

[1] W. Hoppe, "Beugung im inhomogenen primärstrahlwellenfeld. I. Prinzip einer phasenmessung von elektronenbeungungsinterferenzen," *Acta Crystallographica Sect. A*, vol. 25, no. 4, pp. 495–501, Jul. 1969.

[2] W. Hoppe and G. Strube, "Beugung in inhomogenen primärstrahlenwellenfeld. II. Lichtoptische analogieversuche zur phasenmessung von gitterinterferenzen," *Acta Crystallographica Sect. A*, vol. 25, no. 4, pp. 502–507, Jul. 1969.

[3] W. Hoppe, "Beugung im inhomogenen primärstrahlwellenfeld. III. Amplituden- und phasenbestimmung bei unperiodischen objekten," *Acta Crystallographica Sect. A*, vol. 25, no. 4, pp. 508–514, Jul. 1969.

[4] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[5] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.

[6] J. M. Rodenburg and H. M. L. Faulkner, "A phase retrieval algorithm for shifting illumination," *Appl. Phys. Lett.*, vol. 85, no. 20, pp. 4795–4797, 2004.

[7] A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," *Ultramicroscopy*, vol. 109, no. 10, pp. 1256–1262, 2009.

[8] H. Yang et al., "Electron ptychographic phase imaging of light elements in crystalline materials using Wigner distribution deconvolution," *Ultramicroscopy*, vol. 180, pp. 173–179, 2017.

[9] J.M. Rodenburg and R.H.T. Bates, "The theory of super-resolution electron microscopy via Wigner-distribution deconvolution," *Philos. Trans. Roy. Soc. London. Ser. A: Phys. Eng. Sci.*, vol. 339, no. 1655, pp. 521–553, 1992.

[10] P. Li, T. B. Edo, and J. M. Rodenburg, "Ptychographic inversion via Wigner distribution deconvolution: Noise suppression and probe design," *Ultramicroscopy*, vol. 147, pp. 106–113, 2014.

[11] G. Wang, G. B. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 773–794, Feb. 2018.

[12] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.

[13] Y. S. Tan and R. Vershynin, "Phase retrieval via randomized Kaczmarz: Theoretical guarantees," *Inf. Inference: A J. IMA*, vol. 8, no. 1, pp. 97–123, 2019.

[14] J. M. Cowley and A. F. Moodie, "The scattering of electrons by atoms and crystals. I. A new theoretical approach," *Acta Crystallographica*, vol. 10, no. 10, pp. 609–619, 1957.

[15] E. J. Kirkland, *Advanced Computing in Electron Microscopy*. Berlin, Germany:Springer, 1998.

[16] P. Goodman and A. F. Moodie, "Numerical evaluations of N-beam wave functions in electron scattering by the multi-slice method," *Acta Crystallographica Sect. A: Cryst. Phys., Diffraction, Theor. Gen. Crystallogr.*, vol. 30, no. 2, pp. 280–290, 1974.

[17] H. Bethe, "Theorie der beugung von elektronen an kristallen," *Annalen der Physik*, vol. 392, no. 17, pp. 55–129, 1928.

[18] A. Howie and M. J. Whelan, "Diffraction contrast of electron microscope images of crystal lattice defects-II. The development of a dynamical theory," *Proc. Roy. Soc. London. Ser. A Math. Phys. Sci.*, vol. 263, no. 1313, pp. 217–237, 1961.

[19] 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," *Phys. Rev. Lett.*, vol. 125, no. 6, 2020, Art. no. 065502.

[20] H. G. Brown et al., "A three-dimensional reconstruction algorithm for scanning transmission electron microscopy data from a single sample orientation," *Microsc. Microanal.*, vol. 28, no. 5, pp. 1632–1640, 2022.

[21] P. M. Pelz et al., "Phase-contrast imaging of multiply-scattering extended objects at atomic resolution by reconstruction of the scattering matrix," *Phys. Rev. Res.*, vol. 3, no. 2, 2021, Art. no. 023159.

[22] J. Zhong, L. Tian, P. Varma, and L. Waller, "Nonlinear optimization algorithm for partially coherent phase retrieval and source recovery," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 310–322, Sep. 2016.

[23] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg, "Ptychographic transmission microscopy in three dimensions using a multi-slice approach," *J. Opt. Soc. Amer. A*, vol. 29, no. 8, pp. 1606–1614, 2012.

[24] P. Li and A. Maiden, "Multi-slice ptychographic tomography," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, 2018.

[25] M. Kahnt et al., "Multi-slice ptychography enables high-resolution measurements in extended chemical reactors," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021.

[26] L. Tian and L. Waller, "3-D intensity and phase imaging from light field measurements in an led array microscope," *Optica*, vol. 2, no. 2, pp. 104–111, 2015.

[27] W. V. den Broek and C. T. Koch, "Method for retrieval of the three-dimensional object potential by inversion of dynamical electron scattering," *Phys. Rev. Lett.*, vol. 109, no. 24, 2012, Art. no. 245502.

[28] W. V. den Broek and C. T. Koch, "General framework for quantitative three-dimensional reconstruction from arbitrary detection geometries in TEM," *Phys. Rev. B*, vol. 87, no. 18, 2013, Art. no. 184108.

[29] M. Schloz et al., "Overcoming information reduced data and experimentally uncertain parameters in ptychography with regularized optimization," *Opt. Exp.*, vol. 28, no. 19, pp. 28306–28323, 2020.

[30] Z. Chen et al., "Electron ptychography achieves atomic-resolution limits set by lattice vibrations," *Science*, vol. 372, no. 6544, pp. 826–831, 2021.

[31] R. Xu et al., "Accelerated Wirtinger flow: A fast algorithm for ptychography," 2018, *arXiv:1806.05546*.

[32] L. L. Magoarou and R. Gribonval, "Flexible multilayer sparse approximations of matrices and applications," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 688–700, Jun. 2016.

[33] P. M. Pardalos and S. A. Vavasis, "Quadratic programming with one negative eigenvalue is NP-hard," *J. Glob. Optim.*, vol. 1, pp. 15–22, 1991.

[34] A. Beck, "On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes," *SIAM J. Optim.*, vol. 25, no. 1, pp. 185–209, 2015.

[35] S. Lu, I. Tsaknakis, and M. Hong, "Block alternating optimization for non-convex min-max problems: Algorithms and applications in signal processing and communications," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 4754–4758.

[36] H. Chang, P. Enfedaque, and S. Marchesini, "Blind ptychographic phase retrieval via convergent alternating direction method of multipliers," *SIAM J. Imag. Sci.*, vol. 12, no. 1, pp. 153–185, 2019.

[37] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[38] R. W. G. Wyckoff, *Crystal Structures*, vol. 1. New York, NY, USA: Interscience, 1963.

[39] B. Schönfeld, J. J. Huang, and S. C. Moss, "Anisotropic mean-square displacements (MSD) in single-crystals of 2H-and 3R-MoS_2," *Acta Crystallographica Sect. B: Struct. Sci.*, vol. 39, no. 4, pp. 404–407, 1983.

[40] K. Tsuda and M. Tanaka, "Refinement of crystal structure parameters using convergent-beam electron diffraction: The low-temperature phase of $SrTiO_3$," *Acta Crystallographica Sect. A: Foundations Crystallogr.*, vol. 51, no. 1, pp. 7–19, 1995.

[41] L. Bian et al., "Fourier ptychographic reconstruction using poisson maximum likelihood and truncated Wirtinger gradient," *Sci. Rep.*, vol. 6, no. 1, pp. 1–10, 2016.

[42] Z. Li, K. Lange, and J. A. Fessler, "Poisson phase retrieval in very low-count regimes," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 838–850, 2022.

[43] G. Jagatap, Z. Chen, S. Nayer, C. Hegde, and N. Vaswani, "Sample efficient Fourier ptychography for structured data," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 344–357, 2020.