Phase Object Reconstruction for 4D-STEM using Deep Learning

Thomas Friedrich, Chu-Ping Yu, Johan Verbeeck, Sandra Van Aert

Microscopy and Microanalysis, 2023, **29**, 395–407
https://doi.org/10.1093/micmic/ozac002
Advance access publication 12 January 2023
Original Article

Microscopy AND Microanalysis

Phase Object Reconstruction for 4D-STEM using Deep Learning

Thomas Friedrich[1,2], Chu-Ping Yu[1,2], Johan Verbeeck[1,2], and Sandra Van Aert[1,2,*]

[1]*Electron Microscopy for Materials Science (EMAT), University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium*
[2]*NANOlab Center of Excellence, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium*

Corresponding author: Sandra Van Aert, E-mail: sandra.vanaert@uantwerpen.be

Abstract

In this study, we explore the possibility to use deep learning for the reconstruction of phase images from 4D scanning transmission electron microscopy (4D-STEM) data. The process can be divided into two main steps. First, the complex electron wave function is recovered for a convergent beam electron diffraction pattern (CBED) using a convolutional neural network (CNN). Subsequently, a corresponding patch of the phase object is recovered using the phase object approximation. Repeating this for each scan position in a 4D-STEM dataset and combining the patches by complex summation yields the full-phase object. Each patch is recovered from a kernel of 3×3 adjacent CBEDs only, which eliminates common, large memory requirements and enables live processing during an experiment. The machine learning pipeline, data generation, and the reconstruction algorithm are presented. We demonstrate that the CNN can retrieve phase information beyond the aperture angle, enabling super-resolution imaging. The image contrast formation is evaluated showing a dependence on the thickness and atomic column type. Columns containing light and heavy elements can be imaged simultaneously and are distinguishable. The combination of super-resolution, good noise robustness, and intuitive image contrast characteristics makes the approach unique among live imaging methods in 4D-STEM.

Key words: electron microscopy, inverse problem, low-dose imaging, machine learning, phase retrieval, 4D-STEM

Introduction

Scanning transmission electron microscopy (STEM) is among the most widely used techniques for the visualization, characterization, and quantification of atomic structures in material- and nanoscience. Images are acquired by scanning a sample with an electron probe over a two dimensional (2D) grid. Traditionally, STEM data are collected using annular detectors which measure an integrated intensity over the detector area. This results directly in 2D images where the pixel intensities are proportional to the angular integrated scattered intensity of the electron beam at the corresponding probe position. These images are intuitively interpretable and can further be used for the quantification of atomic structures (Van Aert et al., 2013). However, conventional STEM imaging has its limitations, particularly in terms of dose efficiency, resolution, and the ability to image light and heavy elements simultaneously. Indeed, in the case of annular dark field (ADF) imaging, the scattering cross section of an atom scales with the atomic number Z roughly $Z^{1.7}$ (Krivanek et al., 2010; Yamashita et al., 2018), which often leads to heavy atoms obscuring nearby light elements. On the other hand, imaging very thin or radiation-sensitive samples may even be impossible due to the high dose requirements of ADF. This is highly relevant especially for investigations of 2D nanostructures and organic compounds.

To overcome these limitations much effort has been devoted towards the development of fast pixelated electron detectors which can record an entire convergent beam electron diffraction pattern (CBED) in a reasonable amount of time (Ballabriga et al., 2011; Tate et al., 2016; Ciston et al., 2019; MacLaren et al., 2020; Haas et al., 2021; Jannis et al., 2021*a*). This enables a set of new imaging modalities, such as ptychography, the calculation of phase contrast (Lazić et al., 2016), and true center of mass (COM) imaging (Müller et al., 2014). These imaging modalities can all be considered phase retrieval algorithms in a wider sense (Close et al., 2015), as they all aim to retrieve the projected electrostatic potential of a sample, which directly affects the phase of the transmitted electron wave. Ptychographic methods have been of particular interest for their super-resolution capabilities and the possibility to determine/correct for microscope aberrations as well (Rodenburg & Bates, 1992; Jiang et al., 2018; Rodenburg et al., 1993). However, the computational cost and memory requirements for these algorithms are considerable. Iterative algorithms like ePIE (Maiden & Rodenburg, 2009; Chen et al., 2020) use optimization algorithms to fit an object to a given dataset such that the estimated phase corresponds to the intensity observations across the entire dataset. This is a computationally intensive task and the result is influenced by optimization parameters and convergence criteria. Other noniterative methods typically use only the bright-field disc for phase reconstructions which imposes limitations on the maximum achievable resolution (Rodenburg et al., 1993). They rely on taking Fourier transform with respect to the probe positions in real space, which means that for conventional fast Fourier transform (FFT) algorithms, entire datasets (or at least substantial parts of them) have to fit into computer memory, which is becoming increasingly restrictive considering the growing size of 4D-datasets. For these reasons, ptychography has found many useful applications, mainly as a postexperiment data processing analysis step in specialized studies, but has not become a mainstream imaging

Received: February 25, 2022. **Revised:** October 5, 2022. **Accepted:** October 5, 2022

modality so far. However, there is an increasing interest in using these algorithms interactively during experiments. To that end, live imaging using single sideband (SSB) ptychography was recently implemented and demonstrated by Strauch et al. (2021) and Pelz et al. (2022), as well as live COM imaging by Yu et al. (2021*a*).

In the present work, we explore the possibility to use machine learning (ML) for dose-efficient phase object (PO) reconstructions with super-resolution in (near) real-time. We show that using a convolutional neural network (CNN) enables fast exit wave retrieval for a given CBED, by using only a 3×3 kernel of adjacent diffraction patterns at a time. The method allows the retrieval of exit waves, with a resolution theoretically only limited by the Nyquist frequency of the detector and thus enables super-resolution imaging at sufficiently high doses. Using only nine CBEDs per probe position in a 4D-STEM dataset implies that the dataset can practically be of arbitrary size and the reconstruction can be performed live during the experiment with appropriate accelerator hardware, such as a modern, single GPU. In this paper, the character and capability of the proposed method are discussed in detail and demonstrated on both simulated and experimental data. Comparisons are also made with other possible live processing methods.

To the best of our knowledge, the only reconstruction methods that go beyond utilizing the traditional annular detector and enable live imaging are SSB (Strauch et al., 2021; Pelz et al., 2022), integrated differential phase contrast (iDPC), and integrated center of mass (iCOM) (Yu et al., 2021*a*). This is why we focus our analysis on comparing the results of our proposed method to those methods.

Materials and Methods

Theoretical Framework

The interaction of fast electrons with thin specimens can be conveniently described with the phase object approximation (POA). As an electron passes through a positive electrostatic potential its wavelength λ is temporarily altered which is equivalent to shifting the phase of the electron (Kirkland, 2010). For cases where the specimen is extremely thin, the propagation of the wave as it goes through the material can be neglected as a reasonable approximation and the real space 3D electrostatic potential of the atomic structure $V_s(\vec{r}, z)$ can be expressed as its integral along the optical axis z, resulting in the projected electrostatic potential $v_z(\vec{r}) = \int V_s(\vec{r}, z)\, dz$, which is a function of the vector $\vec{r}$ that spans the remaining two dimensions. With this approximation the exit wave $\psi_{\text{out}}(\vec{r})$ is simply the product of the incident wave $\psi_{\text{in}}(\vec{r})$ and the object, which can be described by the transmission function $T(\vec{r})$:

$$\psi_{\text{out}}(\vec{r}) = \psi_{\text{in}}(\vec{r}) T(\vec{r}) \tag{1}$$

where $T(\vec{r}) = \exp(i\sigma v_z(\vec{r}))$ and σ is an interaction parameter (see Kirkland, 2010 for a more detailed derivation). However, a direct solution of the transmission function according to equation (1) is only possible if both incident and exit wave are known, while in practice neither of them are known a priori. The incident wave $\psi_{\text{in}}(\vec{r})$, can be fairly well approximated as the Fourier transform ($\mathcal{F}$) of the product of the aperture function $A(\vec{k})$ and an aberration-function ($\chi(\vec{k})$)-dependent phase shift (Zuo & Spence, 2016).

$$\psi_{\text{in}}(\vec{r}) = \mathcal{F}[A(\vec{k}) \exp[i\chi(\vec{k})]] \tag{2}$$

Here, $\vec{r}$ describes a 2D space at the object plane and $\vec{k}$ describes the reciprocal space. The function $\chi(\vec{k})$, considering only the spherical aberration C_s and defocus Δf, is given by

$$\chi(\vec{k}) = \pi\lambda k^2 (0.5 C_s \lambda^2 k^2 - \Delta f) \tag{3}$$

Assuming that at least the low order aberration parameters of $\chi(\vec{k})$ are known, equations (2) and (3) can be used to estimate $\psi_{\text{in}}(\vec{k})$. The other piece of missing information is then the exit wave $\psi_{\text{out}}(\vec{r})$. From a 4D STEM experiment, only the intensity $|\psi_{\text{out}}(\vec{k})|^2$ can be measured (Fig. 1-②), and thus a method to retrieve the exit waveform on the sample plane based on the information accessible from the experiment is required to solve equation (1) to get the transmission function $T(\vec{r})$.

Retrieving the phase of $\psi_{\text{out}}(\vec{k})$ is a common inverse problem, but is severely complicated in 4D-STEM by the presence of noise to a level that makes even the estimation of $|\psi_{\text{out}}(\vec{k})|$ a challenging task in its own.

The idea in this study is to leverage the multislice formalism, incorporating the calculation of electrostatic atomic potentials (Lobato & Van Dyck, 2014) and the frozen phonon approximation (Van Dyck, 2009), as a forward model to generate a large synthetic dataset. This dataset can then be used to train a CNN to retrieve an estimate of ψ_{out} for any given experiment within the boundaries of the validity of the used forward model and within the parameter space of the training data, which will be discussed in section "Training Data Generation" and is given concisely in Table 1.

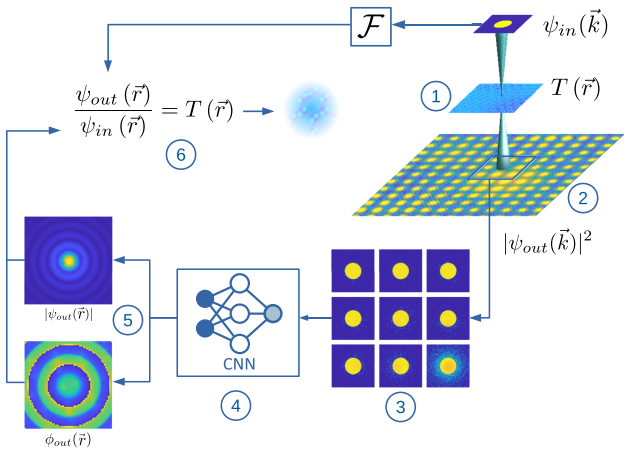

Fig. 1. General workflow: A patch of the phase object ① of a 4D-STEM dataset ② is reconstructed by extracting a 3×3 kernel of adjacent CBEDs ③, using a CNN ④ to reconstruct the amplitude ($|\psi_{\text{out}}(\vec{r})|$) and phase ($\phi_{\text{out}}(\vec{r})$) of the exit wave of the central CBED ⑤ and using the phase object approximation to reconstruct the object patch ⑥ from the reconstructed exit wave and an estimated probe function $\psi_{\text{in}}(\vec{r})$. Patches are then stitched together by complex addition to yield a reconstruction of the full-phase object.

General Workflow

The general concept of the proposed reconstruction method is schematically illustrated in Figure 1. The workflow to retrieve the object $T(\vec{r})$ (Fig. 1-①) of a 4D-STEM dataset (Fig. 1-②) can be divided into two main steps: first, a neural network, trained to solve the inverse problem as outlined in section "Neural

Table 1. Simulation Parameters for the Training Dataset.

Description	Value
Acceleration voltage	$\in$ {30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200, 300} kV
Step size ($\vec{r}$)	0.0167…0.33 Å
Convergence angle	5 . . . 30 mrad
Spherical aberration	0.001 mm
Defocus	Scherzer defocus
Frozen phonons	30
Atom rmsd	0.08 Å
Orientation	$\in$ {[1 1 0], [0 1 1], [1 0 1], [0 0 1], [1 0 0], [0 1 0], [1 1 1]}
Thickness	<30 Å
Dose	$3 \ldots 3 \times 10^9$ e/CBED
Structures	126,335
Samples	742,688

Network Implementation," reconstructs the phase ϕ_{out} and amplitude $|\psi_{\text{out}}|$ of the exit wave, based on the intensity measurements of a 3×3 kernel of adjacent diffraction patterns, as depicted in Figure 1-③–⑤. Second, a patch of the object is retrieved from the previously reconstructed exit wave according to equation (1), as shown in Figure 1-⑥.

In order to take the relative position of each individual object patch into account, a phase factor is included to the approximated $\psi_{\text{in}}(\vec{k})$ and the predicted $\psi_{\text{out}}(\vec{k})$ so that their real space counterpart $\psi_{\text{in}}(\vec{r})$ and $\psi_{\text{out}}(\vec{r})$ are found at the right position. This phase factor ϕ_{shift} is a function of the displacement of the probe position in the x and y directions (Δx and Δy) with regard to the center of an array of size $N_x \times N_y$.

$$\phi_{\text{shift}}(\Delta x, \Delta y) = \exp\left[2\pi \mathrm{i}\left(\frac{\Delta x}{N_x} + \frac{\Delta y}{N_y}\right)\right] \tag{4}$$

The real space wave functions considering the phase factor are

$$\begin{aligned} \psi_{\text{in},n}(\vec{r}) &= \mathcal{F}\{\psi_{\text{in}}(\vec{k}) \times \phi_{\text{shift},n}\} \\ \psi_{\text{out},n}(\vec{r}) &= \mathcal{F}\{\psi_{\text{out}}(\vec{k}) \times \phi_{\text{shift},n}\} \end{aligned} \tag{5}$$

In equation (5), $\mathcal{F}$ is Fourier transformation, and the amount of displacement Δx and Δy is absorbed into n, which specifies the nth probe position.

The POA assumes that the retrieved object patch should have a homogeneous amplitude of 1 with a phase distribution reflecting the projected potential of the imaged material. However, since the transmitting electron probe carries information mostly from a specific region of the examined sample at the probe position, the retrieved object patches are given a weighting function ω_n according to the nth incident probe intensity distribution, and the accordingly weighted object patch $\psi_{\text{patch},n}$ is expressed as

$$T_{\text{patch},n}(\vec{r}) = \frac{\psi_{\text{out},n}(\vec{r})/\psi_{\text{in},n}(\vec{r})}{|\psi_{\text{out},n}(\vec{r})/\psi_{\text{in},n}(\vec{r})|} \times \omega_n(\vec{r}) \tag{6}$$

with the weighting function ω_n as

$$\omega_n(\vec{r}) = \begin{cases} \frac{|\psi_{\text{in},n}(\vec{r})|^2}{\sum_{\vec{r}} |\psi_{\text{in},n}(\vec{r})|^2} & \text{if} |\psi_{\text{in,n}}(\vec{\text{r}})|^2 > \frac{1}{10}\max\left(|\psi_{\text{in,n}}(\vec{\text{r}})|^2\right) \\ 0 & \text{if otherwise} \end{cases} \tag{7}$$

This procedure is repeated for all real space coordinates in the 4D-STEM dataset and the object patches are combined into the final phase object T by complex addition over n probe positions.

$$T(\vec{r}) = \sum_n T_{\text{patch},n}(\vec{r}) \tag{8}$$

The object patch estimations coming from the CNN are not perfect but carry some errors. However, the full object is the combined result of the predictions made at multiple probe positions (equation (8)). Even if one particular prediction is very inaccurate, its impact on the final result is limited as long as it is outweighed by the contributions from neighboring probe positions, which is the case when a significant probe overlap is established by a dense scanning raster.

Training Data Generation

We created a large synthetic dataset, using atomic structures extracted from the materials project (Jain et al., 2013) database. Based on the unit cell definitions we created bulk specimens in one of the low-index zone axis orientations given in Table 1. Each sample consists of a 3×3 kernel of simulated CBED patterns as features and the corresponding exit wave in real-space as label. The simulations were performed using the multislice formalism and the frozen phonon approximation. In the implementation given by Lobato & Van Dyck (2015), Lobato et al. (2016), following the derivation of Van Dyck (2009) the CBED intensity can be expressed as the sum of coherent and incoherent intensities.

$$\langle|\psi_{\text{out}}(\vec{k})|^2\rangle = |\langle\psi_{\text{out}}(\vec{k})\rangle|^2 + \langle|\delta(\vec{k}, t)|^2\rangle \tag{9}$$

where $\psi_{\text{out}}(\vec{k})$ is the exit wave, δ is a phonon configuration(t)-dependent difference and $\langle\rangle$ denotes the average over t. This formalism gives us access to the average, coherent, complex wave function. We use this wave function of the central CBED of the 3×3 kernel as labels (i.e. the ground truth training target) and the CBED intensities of all patterns in the kernel, as given in equation (9) as features (i.e. the CNN input). Only low order aberration parameters Δf and C_s of $\chi(\vec{k})$ (equation (3)) are considered as they are unavoidable and typically have the strongest influence. We assume that including these effects phenomenologically with a constant, small C_s and corresponding Scherzer defocus is sufficient. Temporal and spatial incoherence are also not taken into account. This reduces the parameter space considerably but also implies that the method is (for now) limited to aberration corrected, well-adjusted microscopes.

Further, the dataset includes the CBED-size in $Å^{-1}$, the aperture size and the acceleration voltage, which allows the computation of the probe function $\psi_{in}(\vec{k})$ within the data preprocessing pipeline during the training using equations (2) and (3). Also, the effect of finite electron dose is applied as a data augmentation step during the training, assuming only Poisson noise. To accommodate the possibility that there may be no specimen interacting with the beam, another augmentation step replaces the CBEDs with the probe function intensity with a 3% chance. An example of the resulting training sample inputs and labels is illustrated in Figure 2 in the "input" and "ground truth" panels, respectively. All parameters describing the dataset are summarized in Table 1. Visualizations of the parameter distributions are shown in Supplementary Figure S1. The data generation code was published open source under https://github.com/ThFriedrich/ap_data_generation, as well as the training dataset used (Friedrich et al., 2022*b*).

Neural Network Implementation

CNN Architecture

The complexity of the inverse problem practically dictates the use of deep neural networks in this study. The U-Net architecture (Ronneberger et al., 2015) is one of the most popular choices for deep learning applications, since it allows to expand the number of parameters considerably, while maintaining strong backpropagation. Each depth level in the "U-shape" reduces the filter map size which accommodates the retrieval of features of different sizes. This makes the U-NET an extremely versatile and easy to train CNN and hence suitable as a starting point for this project as well (Friedrich et al., 2021). However, since the training target is naturally complex valued, a generic pixel-to-pixel mapping, as commonly employed for CNNs in image processing, cannot be used. Two popular ways of dealing with this exist. First, both, phase and amplitude retrieval problems can essentially be treated separately by defining two outputs and optimizing for dedicated loss functions on the phase and amplitude components of the wave function (Wu et al., 2021). This treats the complex wave as two real valued images, which in practice has the advantage that common, highly optimized AI tools can be readily employed. Another approach that naturally lends itself to this kind of problem is using complex valued neural networks (Trabelsi et al., 2018); an approach that has found relatively few applications so far. However, the theoretical framework for complex CNNs is established (Trabelsi et al., 2018) and implementations have been showing promising results (Virtue et al., 2018; Munshi et al., 2022). We implemented the U-NET architecture for both CNN types to test the main ideas. The complex networks delivered reconstruction results and accuracies of predicted phases, which are, for all intents and purposes, equivalent to the real valued CNN, while decreasing the speed performance considerably. Since live imaging is an envisioned application, inference speed is a critical concern and the faster real valued CNN was chosen in the study accordingly.

The structure of the U-NET is modified to account for physical considerations. The aim of the neural network is to model the electron-specimen interaction. Adding skip connections from the input probe function to the output exit waves (essentially enabling global residual learning) isolates the specimen interaction contributions from the probe function contributions to the exit wave. The CNN, therefore, does not need to learn to actually model the electron probe. The skip connections have the additional benefit of providing a common template during inference. The training is done on isolated scan points of only 9 CBEDs per specimen, while during inference the probe function should be consistent for the entire dataset, which is a requirement that cannot be captured by any metric during the training. Providing an estimated probe function greatly promotes this consistency during inference. Global residual learning also enhances noise robustness because the probe function serves as template, which is hardly altered if the input is merely noise.

On the input, the dynamic range of the CBEDs is being scaled by taking them to the power of 0.1 in a preprocessing step, which puts more relative weight on the dark field scattered electrons to support exit wave reconstructions beyond the convergence angle. Subsequently, each pattern is scaled, according to equation (10) depending on its distance from the central beam position, where ζ_{xy} and ζ_d correspond to the CBED weights adjacent to the central CBED along x and y and on the diagonal, respectively. This is a straightforward way to include the step size Δs into the workflow.

$$\zeta_{xy} = \frac{1}{\Delta s * 50}, \quad \zeta_d = \frac{1}{\sqrt{2 * \Delta s^2} * 50} \tag{10}$$

The constant factor of 50 accounts for the range of step sizes in the training datasets such that all ζ are between 0 and 1. The effect of this weighting can be seen in the "Input" panel of Figure 2, evident by the higher mean intensity of the central CBED.

Other notable differences as compared to the original U-NET implementation (Ronneberger et al., 2015) are different map sizes, the use of Swish-Activation functions (Ramachandran et al., 2017) and the use of strided convolutional layers instead of max-pooling layers to avoid any loss of information when feature map sizes are reduced. Trainable scaling factors on the two output layers for the phase and amplitude were added to scale between the batch-normalized feature maps (with standard deviations of 1 and means of zero) in the CNN and the relatively small target value distributions, which correspond to the difference between exit waves and probe functions. The variables were initialized accordingly with small values of 1×10^{-4} and optimized during training. The amplitude output layer includes a regularization which penalizes integrated exit wave intensities >1. The phase output layer penalizes values larger than π and smaller than $-\pi$. These penalties are added to the loss during training. The resulting overall architecture is depicted in Figure 3. The tensorflow implementations of the models and individual layers are available open source on Github at https://github.com/ThFriedrich/airpi. During the training the CNN can process >750,000 sample points in about 6 min on a single Nvidia RTX 3080 GPU, indicating that the model itself may be well suited for live processing at rates >2 kHz, if the pre and post-processing pipelines are well optimized.

Loss Function

To facilitate the learning of a general representation for both the phase and amplitude we designed a multi-objective loss function as a linear combination of $\mathcal{L}_2$-losses on the phase and the amplitude in fourier space and in real space. Enforcing the correspondence between $\vec{r}$ and $\vec{k}$-space during

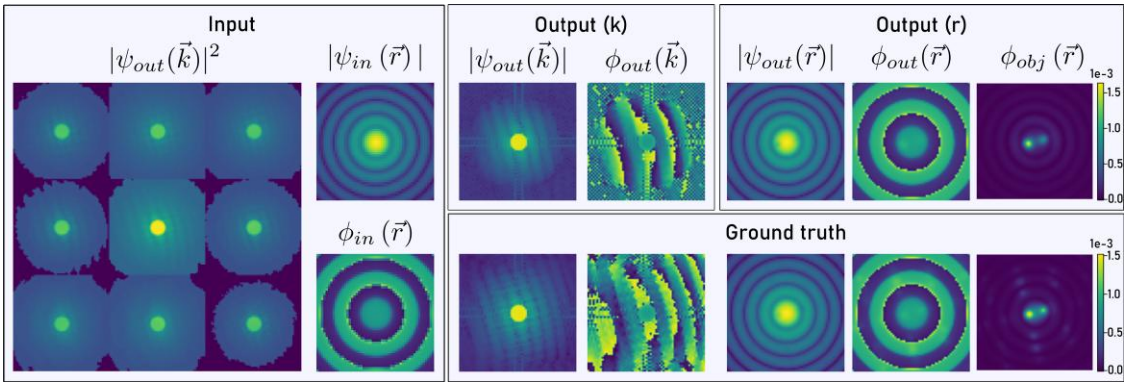

Fig. 2. Example of an exit wave reconstruction taken from the validation dataset, illustrating the inputs and outputs of the CNN, as well as the Fourier transforms of the (real space) exit waves. Intensities and amplitudes are depicted in log scale.

the training encourages the CNN to abide to physical constraints. It was also observed that the decomposition of the phase into its sin and cos components facilitates better convergence, compared to just optimizing for the phase directly. This is presumably related to the decompositions being smooth so the CNN does not have to account for phase wrapping effects. Since the probe function is an input to the CNN the object can be solved directly and used in the loss function too. The optimization of the $\mathcal{L}_2$ error of the phase of the object $T(\vec{r})$ directly promotes an agreement of the object with the transmission function, which is practically the most meaningful metric. However, a good quantitative match may be impossible to achieve in certain scenarios (e.g. very low dose). To promote at least a visual match, the object phase is also optimized for its cross correlation xc as given in equation (14). Further it was empirically determined that a higher weight on the phase in $(\vec{k})$-space leads to faster convergence and overall better results. Putting all terms into a sum, for an exit wave ψ with phase ϕ and an object phase ϕ_{obj} the loss function is given by

$$\begin{aligned}\mathcal{L} = {} & \mathcal{L}_2(|\psi(\vec{k})|^2) + (\mathcal{L}_2(\sin\phi(\vec{k})) + \mathcal{L}_2(\cos\phi(\vec{k}))) * 3 \\ & + \mathcal{L}_2(|\psi(\vec{r})|^2) + \mathcal{L}_2(\sin\phi(\vec{r})) + \mathcal{L}_2(\cos\phi(\vec{r})) \\ & + \mathcal{L}_2(\phi_{\text{obj}}(\vec{r})) + \mathcal{L}_{xc}(\phi_{\text{obj}}(\vec{r}))\end{aligned} \tag{11}$$

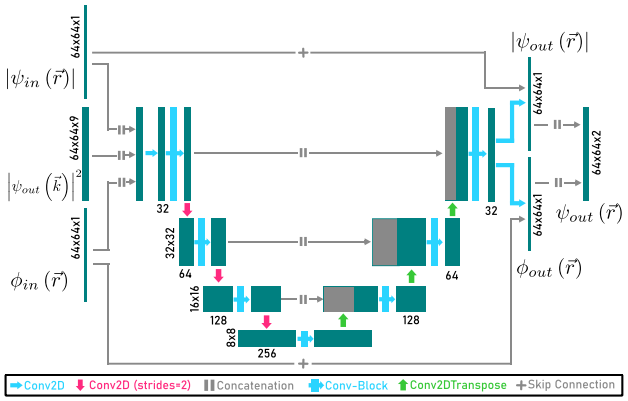

Fig. 3. The CNN architecture used in this study is a modified U-NET with separate, real valued phase and amplitude outputs. The model leverages global residual learning through added skip connections of the probe function to the output. Each "convolution layer" is composed of a 2D convolution layer, batch normalization and a swish-activation function. Each "convolution block" consists of three consecutive convolution layers.

with

$$\mathcal{L}_2(x) = (x_{\text{true}} - x_{\text{predicted}})^2 \tag{12}$$

$$\mathcal{L}_{xc}(x) = (1 - xc(x_{\text{true}}, x_{\text{predicted}}))/2 \tag{13}$$

where

$$xc(x, y) = \frac{\sum_{i,j}[(y(i,j) - \overline{y})(x(i,j) - \overline{x})}{\sqrt{\sum_{i,j}[y(i,j) - \overline{y}]^2}\sqrt{\sum_{i,j}[x(i,j) - \overline{x}]^2}} \tag{14}$$

where x and y correspond to pixel values at the locations (i, j) and $\overline{x}$ and $\overline{y}$ are the mean values, respectively.

Training

The training was performed using the Adam optimizer with a learning rate of 5×10^{-4}, a batch size of 256 and a momentum setting of 0.9. The learning rate was decreased by a factor of 0.5 when the validation loss did not decrease for 3 epochs. Convergence was reached after ≈50 epochs. After convergence the training was resumed for another 10 epochs with a weighting factor of 10 applied to the $\mathcal{L}_2(\phi_{\text{obj}}(\vec{r}))$ term of equation (11), which leads to a further small decrease on the object error. This step does not alter the reconstruction results considerably but improves the quantitative match between reconstructed objects and transmission functions somewhat.

Experiments and Simulations

The demonstrations of the reconstruction methods are performed on both experimental and simulated datasets. For the experiments, probe corrected Thermo Fisher Titan (X-Ant-TEM) and Themis (Advan-TEM) were used. The former is equipped with a MerlinEM direct electron detector (Ballabriga et al., 2011) and the latter with a custom-made Timepix3 detector (Poikela et al., 2014). For the experimental datasets of an Au crystal and a $SrTiO_3$ focused ion beam lamella, which can be found in the online repository (Yu et al., 2021*b*), the acceleration voltage is set at 300 kV, the semi-convergence angle of the electron beam is 20 mrad, and the scanning step size 0.2 and 0.185 Å, respectively. The USY-zeolite dataset, which can be found in Jannis et al. (2021*b*), is collected at 200 keV, with 12 mrad convergence angle and 0.15 Å step size. The simulated twisted bilayer graphene dataset is generated with an acceleration voltage 200 kV, a convergence angle of 25 mrad, and a scan step size 0.2 Å. The twisted bilayer MoS_2 dataset was simulated

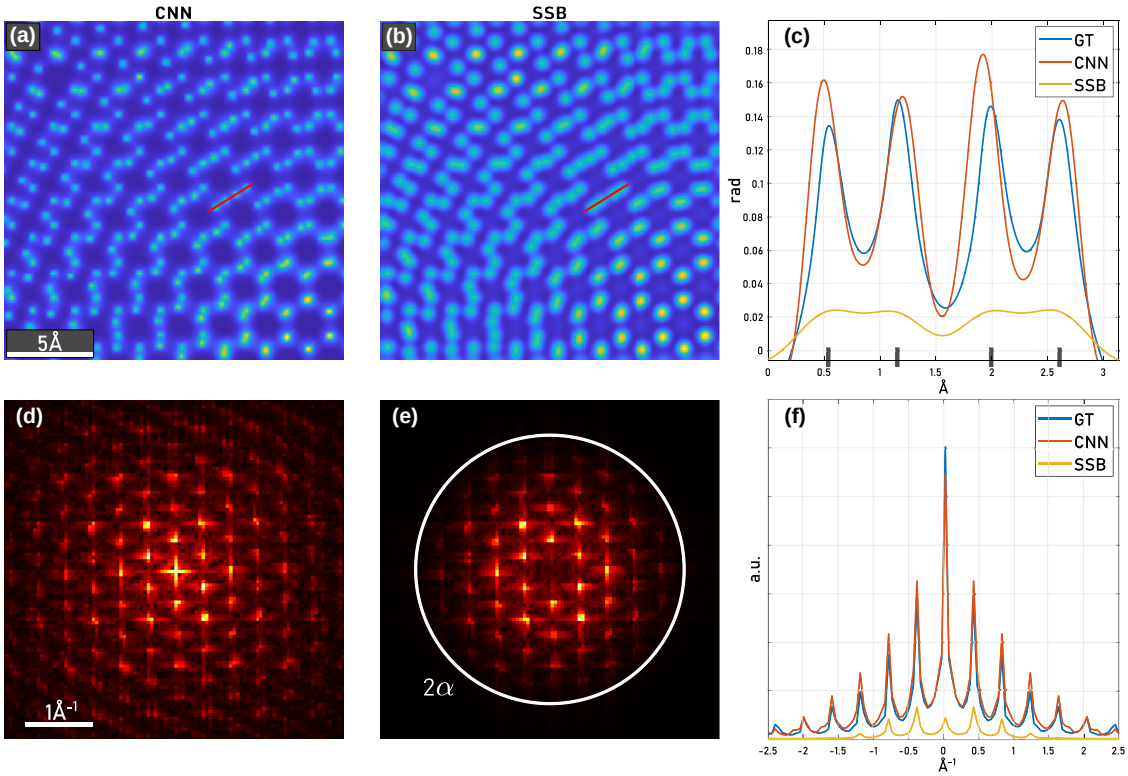

Fig. 4. Demonstration of super-resolution capabilities on simulated datasets with infinite dose. Compared are CNN reconstructions and standard SSB ptychography (**b**). Their corresponding Fourier transformed (FT) intensities show the maximal spatial frequency achieved by each method (**d**,**e**). For the FT image of the SSB result, a circle indicating twice the convergence angle is added, which corresponds to the maximal spatial frequency of the method. (**c**) Shows the intensity along the line profile drawn in each image. Markers on the *x*-axis indicate atom positions. (**f**) Depicts the integrated intensity of the FT images along *y*-axis.

with the settings: acceleration voltage 300 kV, convergence angle 20 mrad, and scan step size 0.1 Å. The MgO dataset was created with an acceleration voltage of 300 kV, a convergence angle of 20 mrad, and a scan step size 0.05 Å. All of the simulated datasets are generated with the MULTEM software (Lobato & Van Dyck, 2015).

Results and Discussion

Super-Resolution

The reconstruction of the proposed method is based on solving equation (1) for the object using the incident and the exit wave functions, and therefore the resolution of the method is not explicitly limited by neither the optical conditions of the imaging system nor the sampling density of the electron probe. By reciprocity, the object plane is sampled with a maximum resolution determined by the maximum scattering angle the detector can reach, or the highest angle at which the exit wave can be accurately retrieved, which can potentially result in a higher resolution than permitted by the former two limitations.

This super-resolution granted by the knowledge of the exit wave at higher scattering angles is demonstrated by the reconstruction of a simulated dataset of a twisted bilayer graphene sample at infinite dose (Friedrich et al., 2022*a*). The result from the CNN reconstruction is shown in Figure 4a, and compared with a SSB reconstruction in Figure 4b.

To analyze the spatial frequency achieved by each method, the Fourier transformed (FT) images are presented as well (Figs. 4d, 4e). The circle in the FT SSB image indicates twice the range of the convergence angle (α), which is the upper limit of the spatial frequency of this reconstruction method (Rodenburg et al., 1993), and therefore all the frequency components beyond are eliminated. The reconstruction of the CNN successfully retrieves components beyond this limitation, also reflected in the ability to distinguish atoms with very short spacing in between, as can be seen in the atom pairs profile (Fig. 4c).

The improved resolution capability of the method, as well as its dependency on the individual per-CBED-dose (more thoroughly discussed in section "Step Size") are also demonstrated on an experimental USY-zeolite dataset (Jannis et al., 2021*b*). This dataset has a fiarly low dose and small step size. To increase the dose of each individual diffraction pattern, the CBED at each probe position is replaced by a summation of CBEDs within a 5×5 box, while the reconstruction is done with a step size twice as large as originally. This repetition in data usage increases the effective dose in the dataset, as each individual CBED now contains 25/4-times more electrons and greatly increases the accuracy of the neural network prediction. The actual dose that inflicts damage while interacting with the material, on the other hand, remains the same. For comparison, SSB is performed on the original dataset and a dataset with the same data repetition strategy applied. In Figure 5, it is shown that the last three reconstructions successfully build a clear image of the zeolite crystal structure with atomic-level resolution. The CNN reconstruction based on the original dataset (Fig. 5a) does not showcase a similar quality, since the dose for individual CBEDs is too low to make accurate predictions of the exit waves, but after data repetition is

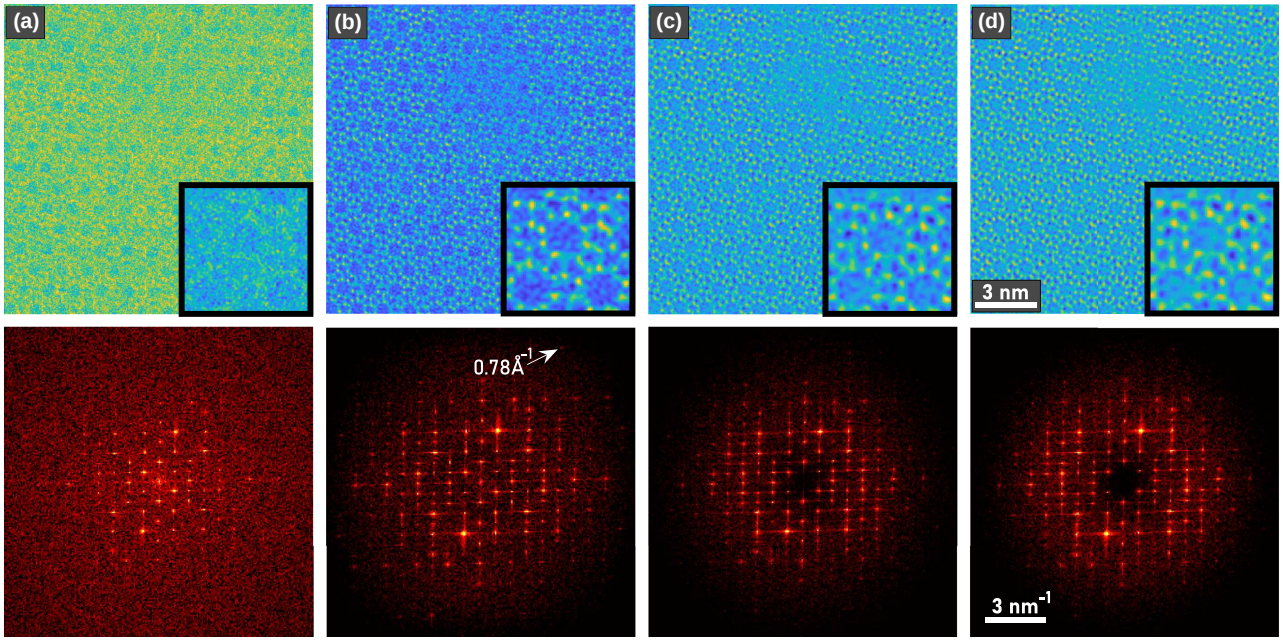

Fig. 5. Reconstruction results of three different approaches. (**a**,**b**) Neural Network performed on datasets without and with data repetition, respectively. (**c**,**d**) SSB reconstruction done on datasets without and with data repetition, respectively. The Fourier transforms of the reconstructed images are shown below. Notice that in (**a**) vertical streaks of very strong intensity can be found, which originate from an unknown defect of the detector, also reported by Jannis et al. (2021*a*).

applied, the neural network gives results that capture details of the material (Fig. 5b). SSB, on the other hand, does neither benefit nor suffer from this repetition, at least not at a noticeable level. From the Fourier transform of the three images one can estimate the resolution limits of the methods by comparing the most distant frequency component. The neural network reconstruction shows a maximum frequency component at 0.78 Å^{-1}, which according to the Raleigh criterion:

$$d = \frac{0.61\lambda}{\sin\alpha} \tag{15}$$

is equivalent to the resolving power of an un-aberrated perfect optical system of convergence angle 12 mrad, which is the same as the aperture size used in the experiment, at the electron wavelength (λ) of 0.02 Å. As most of the microscopes, even ones equipped with probe corrector, cannot achieve the resolving power given by the Raleigh criterion, the presented method shows the ability to overcome the effect of remaining aberrations, shot noise, and other imperfection in the system to reach a higher resolution.

Step Size

Since the proposed reconstruction method is based on retrieving individual object patches, which are commonly sampled finer than the step the electron beam takes to scan the sample, a rather coarse scanning grid can produce high-quality images, as long as a good prediction of the exit wave, and hence the object patch, can be made. This character of the proposed reconstruction method is demonstrated on a simulated MgO particle with different step sizes of 0.1, 0.4, 0.8, and 1.6 Å. The "ratio" values shown in Figure 6 refer to the ratio of the diameter of the incident probe function (1.2 Å) and the step size, where the probe size for a given convergence angle α in Å^{-1} is defined by the first root of the Bessel function of the first kind and first order:

$$d = 2 * \frac{3.8317}{\alpha\pi} \tag{16}$$

Figures in the left column are generated with infinite dose and therefore the neural network has very detailed knowledge of the amplitude of the exit wave to make accurate predictions. In this case, the difference between results of overlap ratio of 12 and overlap ratio of 3 is barely noticeable. By further reducing the scan density, probe positions reach a distance where the weighting function forbids any overlap, as shown in equation (7). Despite the weighting function cutoff, which creates blank spaces between the object patches, the actual probe positions used to generate the data overlap with each other just enough, making exit wave predictions possible to maintain the crystal structure to a certain level in the reconstructed image. As the step size reaches 1.6 Å and the ratio drops below 1, the retrieved object patches deteriorate severely and no longer reflect any crystal periodicity. This failure shows that the neural network follows certain physical and mathematical constraints, such as necessary probe overlap for accurate exit wave retrieval, and that it would fail rather then making false predictions that continue to resemble atoms or the crystal. This failure can be identified by the user not only based on the deviation of the resulting image from the expected appearance of the object, but also by the wide blanks left between the object patches, indicating insufficient probe overlap.

The images in the right column of Figure 6 were generated with the same dose per area. As mentioned in the previous section, the accuracy of the retrieved object patch is not directly related to the total dose in the dataset, but rather to the dose per CBED. By this consideration, it follows that larger step sizes work better for the neural network, since this would mean fewer probe positions in the same area and a higher dose at every individual CBED. On the other hand, a certain

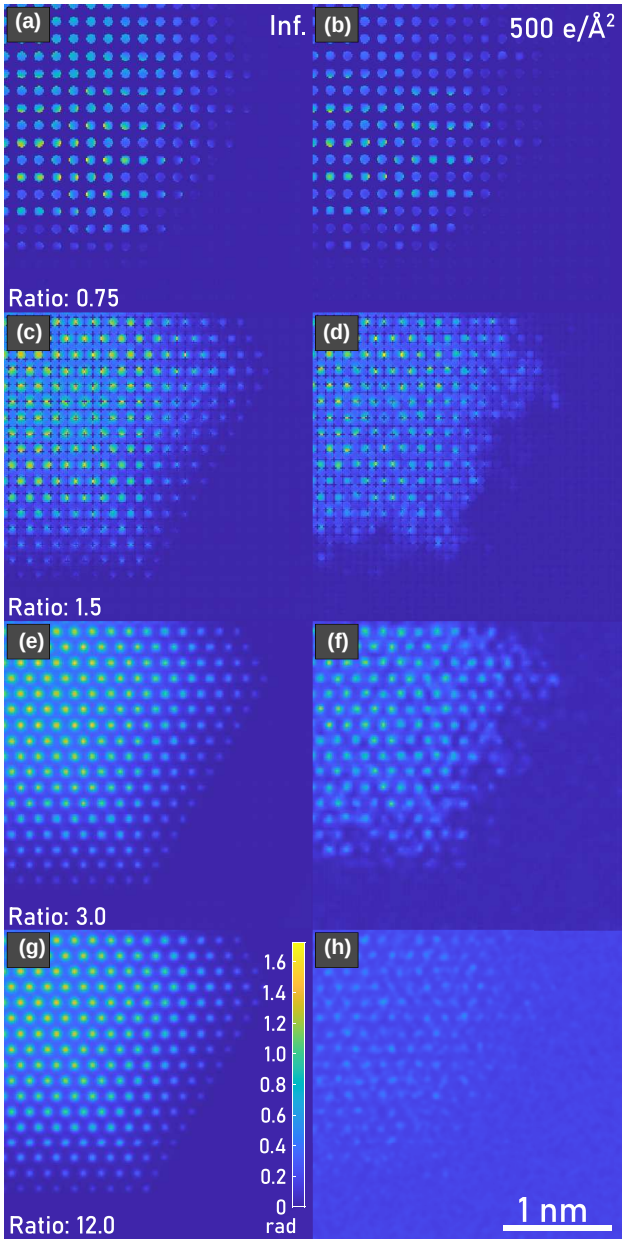

Fig. 6. Reconstruction results of simulated MgO particle. In the left column (**a**,**c**,**e**,**g**) the images are generated with infinite dose, and in the right (**b**,**d**,**f**,**h**) the dose is set to be 500 electron per Å^2. Each row of images is constructed with the same step size, as well as the same step/probe-width ratio. The colorbar in the bottom left panel applies to all images in the figure.

level of probe overlap is also required for accurate predictions. Therefore, not only the total dose per area, but also the scanning strategy is an important consideration for the proposed method. A balanced scan density will generate better results as compared to a very fine scan grid, even if the total dose per area would be the same. This behavior is illustrated in the images in the right column of Figure 6. The noise level is lowered significantly as the step/probe-width ratio drops from 12 to 3 in Figures 6h and 6f. Ptychographic methods in comparison offer more flexibility in this regard as shown in Figures 5c and 5d.

The noise created by inaccurate predictions also creates different features as the step size changes. As the training is

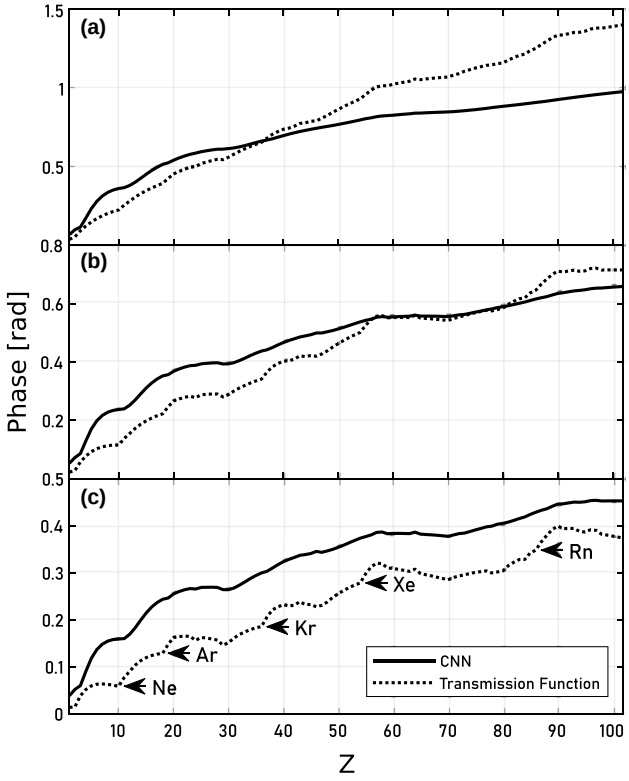

Fig. 7. Phase response of the CNN compared to ground truth transmission functions for simulations of single atoms throughout the periodic table at infinite dose for (**a**) peak intensity, (**b**) mean over 3 × 3 pixels around the atomic position, and (**c**) mean over 5 × 5 pixels around the atomic position.

exclusively done on crystalline materials in zone axis orientations, the predicted object patches may show atomic-scale features, even if it the input is merely noise. In other words, the frequency transfer function of each object patch is highest at the spatial frequency that would compose an image of an atom. This somewhat dangerous behavior of the neural network is compensated by the stitching of the patches, since atomic-scale features of the noise would not remain sharp as multiple object patches contribute to the same area, and would thus contribute to a cloudy, low-intensity background, as seen in Figure 6h. However, when the degree of overlap is reduced, such that the phase value is completely determined by 1 or few object patches, the risk to observe a false, atom-like feature, such as the ones seen in Figures 6f and 6d, greatly increases. Awareness of this effect is, therefore, important when using the method and large step-size scanning patterns should be treated with extra caution. The other effect of noise is that the phase value drops in cases of very few electrons per CBED. This is essentially related to the failure of the CNN at making accurate phase object predictions, but it also shows that as long as the sufficient dose is present in the 3 × 3 CBED input set, the phase value retrieved is not strongly related to step size.

Contrast Analysis

As outlined in section "Theoretical Framework," the phase of the object is proportional to the electrostatic potential. By this relation atomic species should, at least within the boundaries of the POA, be distinguishable. To verify whether this

requirement holds true for the CNN reconstructions, 4D STEM datasets of isolated, single atoms for each species in the periodic table up to $Z = 103$ are simulated individually with a step size of 0.2 Å, a simulation box size of 3×3 Å, an aperture angle of 20 mrad, and a collection angle 60 mrad. The frozen phonon approximation was used with a root-mean-square-displacement of 0.08 Å and 100 phonon configurations for each simulation. The retrieved phase objects of these datasets are compared to the ground truth transmission function, which is based on the parameterization by Lobato & Van Dyck (2014). The comparison is carried out by taking averaged phase values from pixels within various ranges. In Figure 7, from top to bottom, the curves show phase values at the peak only (Fig. 7a), averaged phase over 3×3 pixels around the atomic position (Fig. 7b), and the average over 5×5 pixels (Fig. 7c). Both, the curves of the ground truth transmission function and the CNN predictions, generally increase against atomic number, with the exception of certain dips at larger averaging ranges. This effect stems from different electron orbital distributions in the radial direction, and thus only the phase value averages at larger ranges are sensitive to this difference. The CNN predictions obviously also preserve these sub-atomic level details to some extent, as the shape of the curves bear strong resemblance to the ground truth curves. Although the reconstructed phase values and the transmission functions do not match exactly, the predictions are accurate enough, such that the phase values of the reconstructed objects are indeed useful as an indicator for different atomic species, potentially even allowing semi-quantitatively predicting the exact atomic species.

For thick samples, the method is not expected to yield results in quantitative agreement with projected potentials, because even if the neural network would retrieve the correct exit wave, the reconstruction algorithm is still based on the POA and inherits its limitations. The analysis of thicker samples presented here is, therefore, done in a more qualitative/empirical manner and comparisons are made against ADF imaging and SSB. ADF images are well known for their strong contrast related to the scattering power of the imaged object, and thus are suitable for examining the thickness variation of the sample (De Backer et al., 2016) and local elemental compositions (Pennycook & Boatner, 1988). The contrast of SSB reconstructions is not as strong as ADF (Yang et al., 2016), yet the method is often used to study crystals containing elements of a wide range of atomic numbers due to its ability to image heavy and light atomic columns at the same time with distinguishable contrast (Lozano et al., 2018). Albeit a quantitative match can hardly be expected, it is important to verify whether reconstructed phase images still reflect the relative projected potentials of thicker samples to avoid misinterpretations. To that end an experimental dataset of a $SrTiO_3$ FIB-lamella was analyzed. The reconstruction results are presented and compared to ADF and SSB images in Figure 8.

For ADF imaging the contrast difference between the Sr and O columns is too large, making it difficult to locate the O columns without the help of the profiling. On the other hand, the SSB reconstruction does successfully image both atomic column types. While the peak intensities of the columns are ambiguous, they can still be distinguished by their corresponding size. The O columns are sharper than the Sr ones, indicating that an integrated signal from the area of each column could still be used as a reference of the local projected potential. The CNN reconstruction exhibits the advantage of both: while the light atom columns are observable, both the intensity and size differences are large enough to distinguish their types. This confirms, that the Z-contrast sensitivity is preserved for thicker specimen.

To further investigate the thickness dependence of the retrieved phase signal, an experimental dataset of the tip of an Au nanorod was used. As shown in Figure 9, the intensity recorded by the virtual low angle ADF (LAADF) detector (20–30 mrad) and HAADF detector (45 mrad and beyond) increases from the top to the bottom of the image. Based on statistical analysis of the HAADF signal to retrieve atom counts in each column (De Backer et al., 2016), the thickest part in the image is about 9 nm. A line profile is then drawn for each imaging method presented in the figure. For the two ADF imaging methods at different collection angle, the profiles show monotonic increase at different pace against thickness, while the SSB profile only shows locations but the intensity is not correlated to the thickness of the atomic column. Compared to these profiles, the CNN reconstruction appears to be qualitatively most similar to the one of HAADF, and correlates with the estimated thickness accordingly. It should be noted that the maximum thickness of ≈9 nm is well outside the parameter range of the training data (<3 nm). Also the $SrTiO_3$-sample, being a FIB-lamella, should be well beyond 3 nm thickness. This means that these examples also demonstrate the extrapolation capabilities of the CNN, which albeit being quantitatively arguably inaccurate, may still provide very useful reconstructions for imaging purposes. The strong resemblance with HAADF images at larger thicknesses may in fact be a very desirable characteristic, as it aligns well with many microscopists' experience and intuition.

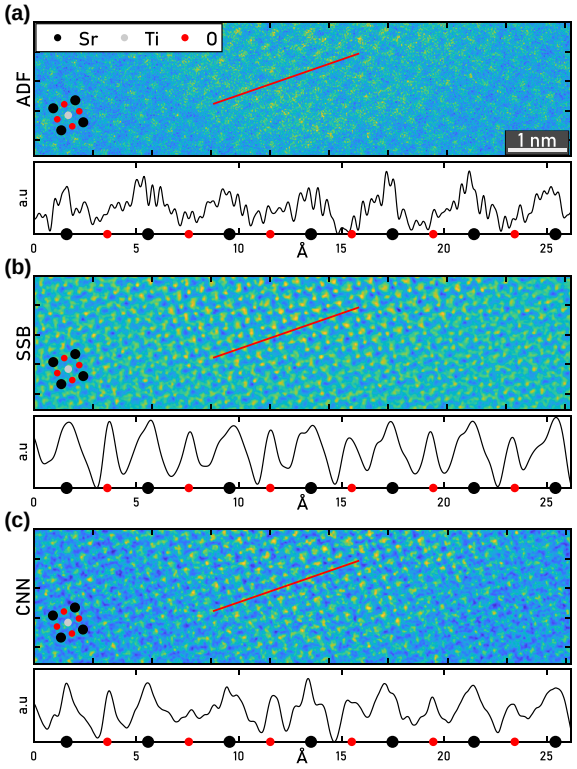

Fig. 8. Reconstructed images of $SrTiO_3$ with (**a**) virtual ADF detector, (**b**) SSB, and (**c**) CNN. Line profiles are drawn to illustrate contrast between the heavier Sr columns and the lighter O columns, which are also indicated, respectively, with black and red dots under the curves.

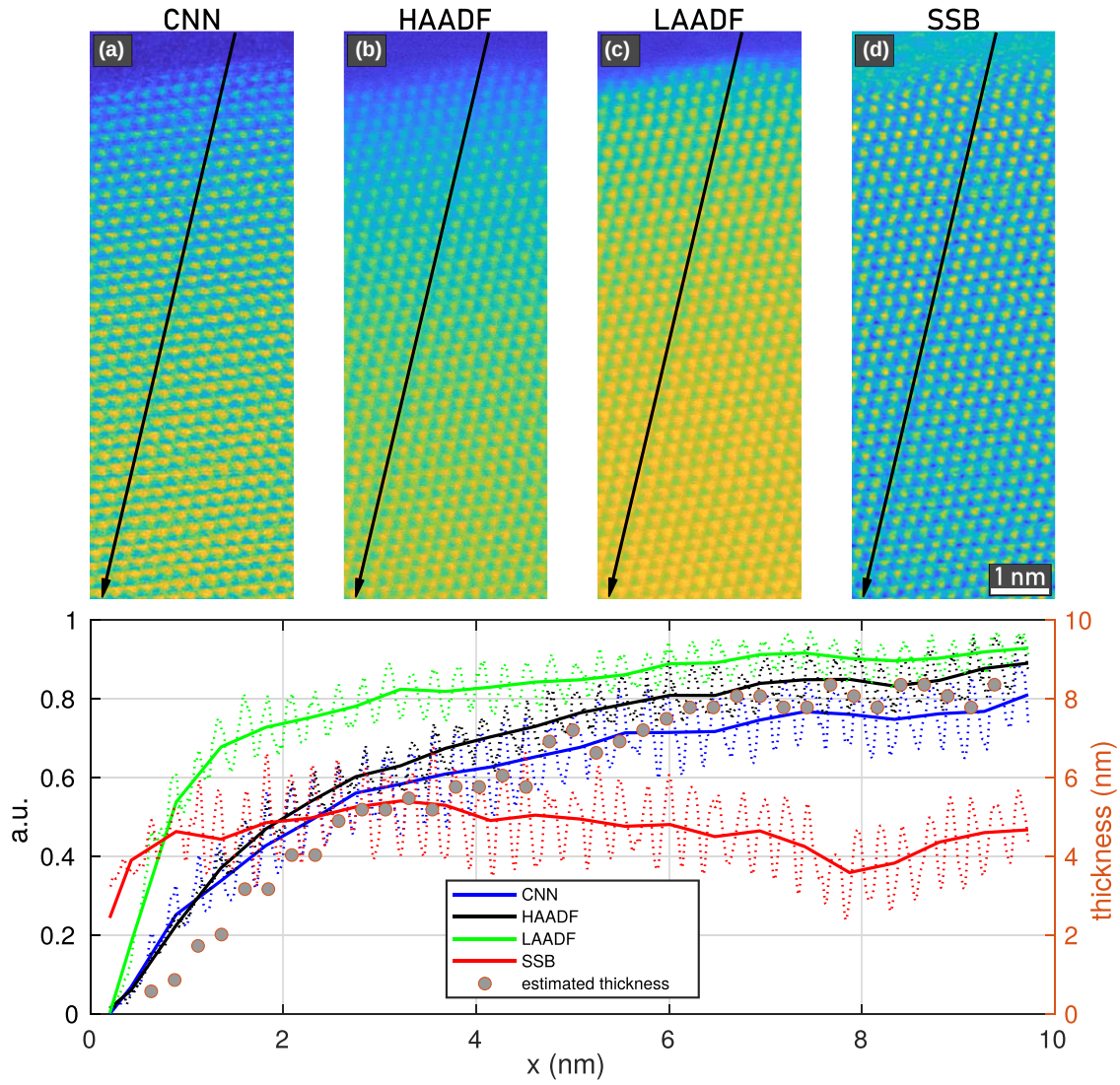

Fig. 9. Reconstructed images of an edge of a Au crystal using (**a**) CNN reconstruction, (**b**) HAADF, (**c**) LAADF, and (**d**) SSB. Line profiles across the nanorod illustrate the thickness dependence of the corresponding signals.

By comparing the reconstructed images from the Au crystal and $SrTiO_3$, one would notice that SSB recovers contrast of higher spatial frequencies, such as the intensity and shape of atomic columns, but it does not recover long-range features induced by e.g. thickness variation. This is due to the band-pass characteristics of the method (Yang et al., 2015; O'Leary et al., 2021), which in practice has a cutoff for high frequencies and a strong tendency to attenuate low frequencies. Long-range features are built with low-frequency components, and thus for reconstruction methods that filter out, or cannot utilize signals that fall in the low-frequency end, these features are lost. For the CNN reconstruction, the object patches are also localized and no information beyond one probe position away are shared among the predictions of the exit waves. Therefore, the existence of thickness-related contrast variations can only be attributed to a reasonably good prediction accuracy of the CNN, also for the low-frequency components.

Noise Robustness

The performance of the method under different dose conditions is demonstrated and analyzed on a simulated dataset of a twisted MoS_2 bilayer. The dose used for the reconstruction ranges from 500 to 1×10^5 e/Å^2, and the dataset is processed by the proposed method, SSB, and iDPC. The methods reconstructions are illustrated in Figure 10 and compared against the ground truth transmission function in terms of their normalized cross correlations (equation (14)).

It is evident that the method is strong in all three dose conditions, as can be seen visually in Figure 10, and is confirmed by a cross correlations higher than those of the corresponding SSB and iDPC reconstructions. The higher cross correlation is not only due to a lower noise level, but also the generally better matched atom shape and phase value with respect to the ground truth, as indicated in the line profiles drawn under each dose condition (Figs. 10j–10l). The sharper atom shape is due to the superior resolution, which is confirmed by the better distinction of close atom pairs, as well as by the higher frequency component found in the Fourier transform of the image (Fig. S2). In Figure 10l, the atom profile of the CNN reconstruction at high dose almost strikes a perfect match with the transmission function. Note that the CNN- and SSB-lines are not normalized or shifted, indicating that a very accurate wave retrieval is achieved by the CNN.

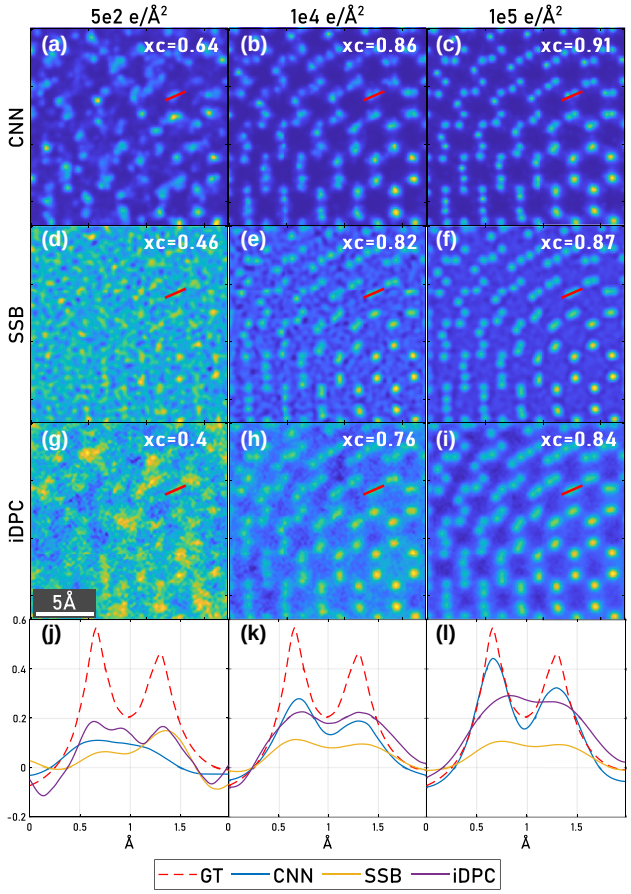

Fig. 10. The dose dependency of the proposed reconstruction method is demonstrated for a simulated dataset of twisted bilayer MoS_2 (top row panels) and compared to the corresponding SSB (second row) and iDPC (third row) reconstructions. Cross correlation values *xc* are given with respect to the ground truth phase object. In (j, k, l) line profiles averaged in the perpendicular direction over 1.6 Å are drawn across a Mo–S_2 pair in all the images, and shown with the ground truth. iDPC values were normalized by the maximum value of the transmission function.

Additionally, the contrast of the CNN results allows to distinguish Mo and S_2, which is more difficult with other methods, as their difference is much smaller. At low dose, the line profiles suggest a stronger low-dose robustness may be found in iDPC, as the two peaks are preserved in the reconstructed image. However, the signal almost completely falls into the noise level, as confirmed by the Fourier transform, and thus this seemingly better low-dose performance could very likely be a coincidental noise distribution. The given example highlights the potential of the proposed method for low-dose imaging. As pointed out in sections "Super-Resolution" and "Step Size" and illustrated in Figures 5 and 6, respectively, the noise robustness may depend substantially on the step size and the effective dose per CBED. To gain an advantage over other methods in this regard this context needs to be taken into account and scanning strategies adapted accordingly.

Conclusion

This paper presents a new computational imaging method, leveraging a CNN to retrieve complex exit wave functions from CBEDs and an algorithm to reconstruct the phase object from the predictions of the neural network. Since the exit waves are retrieved for each real-space coordinate in a 4D-STEM dataset, based only on a small kernel of adjacent diffraction patterns, the method can be employed in a sequential manner, thus enabling live imaging during an experiment. The ML system is based on a well-established model but streamlined to the task at hand and adapted to account for physical constraints and considerations. The model was trained on a large synthetic dataset of multislice simulations. Large and higher-order aberrations, as well as CBED distortions, like noncentricity, geometric distortions and hot/dead pixels are not considered in the training data. Therefore, experimental data may require a preprocessing step. The range of practical conditions for which the method works reliably is, therefore, arguably limited accordingly to aberration corrected, well-adjusted instruments at this stage. The trained model, code, and training data are publicly available as summarized in the supplementary information. In the discussion, multiple unique characteristics and advantages of the method are demonstrated. The CNN-based reconstruction is shown to enable higher resolutions than any other live-imaging-capable method considered, on simulated, as well as on experimental data, provided that a sufficiently high dose-per-CBED is maintained. In correspondence to this consideration, the effect of the step size is analyzed. While a better estimation of the exit wave is obtained with the electron dose-per-area distributed across fewer probe positions, some probe overlap is necessary to insure the accuracy of the exit wave retrieval. Hence, the method is most suitably applied at a balanced scan density. If these considerations are taken into account the reconstruction method can be very dose efficient.

The Z-contrast was analyzed on single atom-simulations across the periodic table. The phase signal of the reconstructions could indeed be linked qualitatively to atomic properties and a semi-quantitative analysis of thin specimen within the limits of the POA, was shown to be possible. We confirmed the contrast sensitivity to atomic species and sample thickness on experimental datasets of a $SrTiO_3$ FIB-lamella and an Au nanorod, respectively. The observed monotonic increase of the phase signal with thickness and nearly monotonic increase with atomic number indicates that quantitative analyses based on the reconstruction results may be feasible.

Generally, we believe the proposed method presents an attractive imaging modality for its super-resolution capability, high noise robustness, and the feasibility of qualitative or even quantitative contrast analysis. While further studies would be necessary to obtain a more detailed view on the model performance over the entire parameter space (and beyond), we could already show that the method is robust for a wide range of practically meaningful applications, even exhibiting reasonably good extrapolation behavior well beyond the maximum sample thickness of the training data. The fact that none of the examples shown in this study exist in those exact configurations in the training data, further indicates that the system generalizes well within the parameter interpolation range as well.

Availability of data and materials

The training data, the trained model, all implementations, scripts, and the data generation codes are publicly available under their respective license terms as summarized in the supplementary information.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1093/micmic/ozac002.

Acknowledgments

We acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 770887 PICOMETRICS) and funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 823717 ESTEEM3. J.V. and S.V.A acknowledge funding from the University of Antwerp through a TOP BOF project. The direct electron detector (Merlin, Medipix3, Quantum Detectors) was funded by the Hercules fund from the Flemish Government. This work was supported by the FWO and FNRS within the 2Dto3D project of the EOS program (grant number 30489208).

Conflict of interest

The authors declare that they have no competing interest.

References

Ballabriga R, Campbell M, Heijne E, Llopart X, Tlustos L & Wong W (2011). Medipix3: A 64 k pixel detector readout chip working in single photon counting mode with improved spectrometric performance. *Nucl Instrum Methods Phys Res A* **633**, S15–S18.

Chen Z, Odstrcil M, Jiang Y, Han Y, Chiu MH, Li LJ & Muller DA (2020). Mixed-state electron ptychography enables sub-angstrom resolution imaging with picometer precision at low dose. *Nat Commun* **11**, 2994.

Ciston J, Johnson IJ, Draney BR, Ercius P, Fong E, Goldschmidt A, Joseph JM, Lee JR, Mueller A, Ophus C, Selvarajan A, Skinner DE, Stezelberger T, Tindall CS, Minor AM & Denes P (2019). The 4D camera: Very high speed electron counting for 4D-STEM. *Microsc Microanal* **25**, 1930–1931.

Close R, Chen Z, Shibata N & Findlay SD (2015). Towards quantitative, atomic-resolution reconstruction of the electrostatic potential via differential phase contrast using electrons. *Ultramicroscopy* **159**, 124–137. https://doi.org/10.1016/j.ultramic.2015.09.002

De Backer A, Van den Bos K, Van den Broek W, Sijbers J & Van Aert S (2016). Statstem: An efficient approach for accurate and precise model-based quantification of atomic resolution electron microscopy images. *Ultramicroscopy* **171**, 104–116.

Friedrich T, Yu CP, Verbeek J, Pennycook T & Aert SV (2021). Phase retrieval from 4-dimensional electron diffraction datasets. In *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 3453–3457. IEEE.

Friedrich T, Yu CP, Verbeeck J & Van Aert S (2022*a*). Phase object reconstruction for 4D-STEM using deep learning (4D-STEM Example Data). Available at https://doi.org/10.5281/zenodo.7034879.

Friedrich T, Yu CP, Verbeeck J & Van Aert S (2022*b*). Phase object reconstruction for 4D-STEM using deep learning (4D-STEM Training Data). Available at https://doi.org/10.5281/zenodo.6971200.

Haas B, Mittelberger A, Meyer C, Plotkin-Swing B, Dellby N, Krivanek O, Lovejoy T & Koch C (2021). High-fidelity 4D-STEM enabled by live processing at 15,000 detector frames per second. *Microsc Microanal* **27**, 994–997.

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G & Persson KA (2013). Commentary: The materials project: A materials genome approach to accelerating materials innovation. Available at https://doi.org/10.1063/1.4812323.

Jannis D, Hofer C, Gao C, Xie X, Béché A, Pennycook TJ & Verbeeck J (2021*a*). Event driven 4D STEM acquisition with a Timepix3 detector: Microsecond dwell time and faster scans for high precision and low dose applications. *Ultramicroscopy* **233**, 113423.

Jannis D, Hofer C, Gao C, Xie X, Béché A, Pennycook TJ & Verbeeck J (2021*b*). Event driven 4D STEM acquisition with a Timepix3 detector: Microsecond dwelltime and faster scans for high precision and low dose applications. Available at https://doi.org/10.5281/zenodo.5068510.

Jiang Y, Chen Z, Han Y, Deb P, Gao H, Xie S, Purohit P, Tate MW, Park J, Gruner SM, Elser V & Muller DA (2018). Electron ptychography of 2D materials to deep sub-ångström resolution. *Nature* **559**, 343–349.

Kirkland EJ (2010). *Advanced Computing in Electron Microscopy*. New York, NY: Springer Science & Business Media.

Krivanek OL, Chisholm MF, Nicolosi V, Pennycook TJ, Corbin GJ, Dellby N, Murfitt MF, Own CS, Szilagyi ZS, Oxley MP, Pantelides ST & Pennycook SJ (2010). Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. *Nature* **464**, 571–574.

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

Lobato I, van Aert S & Verbeeck J (2016). Progress and new advances in simulating electron microscopy datasets using MULTEM. *Ultramicroscopy* **168**, 17–27.

Lobato I & Van Dyck D (2014). An accurate parameterization for scattering factors, electron densities and electrostatic potentials for neutral atoms that obey all physical constraints. *Acta Crystallogr A: Found Adv* **70**, 636–649.

Lobato I & Van Dyck D (2015). MULTEM: A new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA. *Ultramicroscopy* **156**, 9–17.

Lozano JG, Martinez GT, Jin L, Nellist PD & Bruce PG (2018). Low-dose aberration-free imaging of Li-rich cathode materials at various states of charge using electron ptychography. *Nano Lett* **18**, 6850–6855.

MacLaren I, MacGregor TA, Allen CS & Kirkland AI (2020). Detectors—The ongoing revolution in scanning transmission electron microscopy and why this important to material characterization. *APL Mater* **8**, 110901.

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

Müller K, Krause FF, Béché A, Schowalter M, Galioit V, Löffler S, Verbeeck J, Zweck J, Schattschneider P & Rosenauer A (2014). Atomic electric fields revealed by a quantum mechanical approach to electron picodiffraction. *Nat Commun* **5**, 1–8.

Munshi J, Rakowski A, Savitzky BH, Zeltmann SE, Ciston J, Henderson M, Cholia S, Minor AM, Chan MK & Ophus C (2022). Disentangling multiple scattering with deep learning: Application to strain mapping from electron diffraction patterns. arXiv, preprint arXiv:220200204.

O'Leary CM, Martinez GT, Liberti E, Humphry MJ, Kirkland AI & Nellist PD (2021). Contrast transfer and noise considerations in focused-probe electron ptychography. *Ultramicroscopy* **221**, 113189.

Pelz PM, Johnson I, Ophus C, Ercius P & Scott MC (2022). Real-time interactive 4D-STEM phase-contrast imaging from electron event representation data: Less computation with the right representation. *IEEE Signal Process Mag* **39**, 25–31.

Pennycook S & Boatner L (1988). Chemically sensitive structure-imaging with a scanning transmission electron microscope. *Nature* **336**, 565–567.

Poikela T, Plosila J, Westerlund T, Campbell M, De Gaspari M, Llopart X, Gromov V, Kluit R, Van Beuzekom M, Zappon F, Zivkovic V, Brezina C, Desch K, Fu Y & Kruth A (2014). Timepix3: A 65 k

channel hybrid pixel readout chip with simultaneous toa/tot and sparse readout. *J Instrum* **9**, C05013.
Ramachandran P, Zoph B & Le QV (2017). Searching for activation functions. arXiv, preprint arXiv:171005941.
Rodenburg J, McCallum B & Nellist P (1993). Experimental tests on double-resolution coherent imaging via stem. *Ultramicroscopy* **48**, 304–314.
Rodenburg JM & Bates R (1992). The theory of super-resolution electron microscopy via Wigner-distribution deconvolution. *Philos Trans R Soc A* **339**, 521–553.
Ronneberger O, Fischer P & Brox T (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Navab N, Hornegger J, Wells WM, Frangi Af (Eds.), pp. 234–241. Cham: Springer International Publishing.
Strauch A, Weber D, Clausen A, Lesnichaia A, Bangun A, März B, Lyu FJ, Chen Q, Rosenauer A, Dunin-Borkowski R & Müller-Caspary K (2021). Live processing of momentum-resolved stem data for first moment imaging and ptychography. *Microsc Microanal* **27**, 1078–1092.
Tate MW, Purohit P, Chamberlain D, Nguyen KX, Hovden R, Chang CS, Deb P, Turgut E, Heron JT, Schlom DG, Ralph DC, Fuchs GD, Shanks KS, Philipp HT, Muller DA & Gruner SM (2016). High dynamic range pixel array detector for scanning transmission electron microscopy. *Microsc Microanal* **22**, 237–249.
Trabelsi C, Bilaniuk O, Zhang Y, Serdyuk D, Subramanian S, Santos JF, Mehri S, Rostamzadeh N, Bengio Y & Pal CJ (2018). Deep complex networks. In *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*. Available at http://github.com/ChihebTrabelsi/deep˙complex˙.
Van Aert S, De Backer A, Martinez GT, Goris B, Bals S, Van Tendeloo G & Rosenauer A (2013). Procedure to count atoms with trustworthy single-atom sensitivity. *Phys Rev B Condens Matter Phys* **87**, 1–6.
Van Dyck D (2009). Is the frozen phonon model adequate to describe inelastic phonon scattering? *Ultramicroscopy* **109**, 677–682.
Virtue P, Yu SX & Lustig M (2018). Better than real: Complex-valued neural nets for MRI fingerprinting. In *Proceedings - International Conference on Image Processing, ICIP*, vol. 2017-September, pp. 3953–3957.
Wu L, Juhas P, Yoo S & Robinson I (2021). Complex imaging of phase domains by deep neural networks. *IUCrJ* **8**, 12–21.
Yamashita S, Kikkawa J, Yanagisawa K, Nagai T, Ishizuka K & Kimoto K (2018). Atomic number dependence of Z contrast in scanning transmission electron microscopy. *Sci Rep* **8**, 1–7.
Yang H, Pennycook TJ & Nellist PD (2015). Efficient phase contrast imaging in stem using a pixelated detector. Part II: Optimisation of imaging conditions. *Ultramicroscopy* **151**, 232–239.
Yang H, Rutte R, Jones L, Simson M, Sagawa R, Ryll H, Huth M, Pennycook T, Green M, Soltau H, Kondo Y, Davis BG & Nellist PD (2016). Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures. *Nat Commun* **7**, 1–8.
Yu CP, Friedrich T, Jannis D, Van Aert S & Verbeeck J (2021*a*). Real-time integration center of mass (riCOM) reconstruction for 4D STEM. *Microsc Microanal* **28**, 1526–1537.
Yu CP, Friedrich T, Jannis D, Xiaobin X, Van Aert S & Verbeeck J (2021*b*). Real time integration center of mass (riCOM) reconstruction for 4D-STEM. Available at https://doi.org/10.5281/zenodo.5572123.
Zuo JM & Spence JC (2016). *Advanced Transmission Electron Microscopy: Imaging and Diffraction in Nanoscience*. New York, NY: Springer.